Evaluating the effectiveness of matching the timing of occurrences and environmental data in ecological niche models: Insights for low-dispersing species

Gonzalo Pinilla-Buitrago¹

¹City University of New York The Graduate Center

March 30, 2024

Abstract

Ecological niche models, crucial for estimating species' potential distribution under global change, can face reduced accuracy when the timing of occurrence data does not align with the environmental data. One solution is to ensure a close temporal match between the environment and the observation date. While this approach is typically recommended for highly mobile species, a few findings support its use for species with limited mobility, whose distributions may be responding to climate change via local population changes. Additionally, it remains unclear what specific temporal resolution could improve model performance. This study assesses the effectiveness of temporal matching for a species with low mobility, the Mexican small-eared shrew (Cryptotis mexicanus), by evaluating different temporal resolutions (one-, five-, and ten-year averaged environmental data) against the standard method (30-year). Occurrences between 1971 and 2000 were used for model training and cross-validation, while those outside this range were used for external evaluation. Based on the omission rate of the external evaluation occurrences, the approach that matched environmental data using the prior ten-year resolution performed better than the standard 30-year average approach, while the rest of evaluation metrics (for any temporal resolution) were not different. Visual inspection indicated that the geographic prediction resulting from a ten-year resolution was as realistic as the one from the standard 30-year approach. In contrast, the shorter temporal resolutions (one and five years) resulted in unrealistic estimates. Therefore, matching the timing of occurrences and environmental data for other species with low mobility may also improve model performance and geographic predictions. Additionally, this correlative approach identifies a potential time lag between climatic changes and population responses in this species. Studies can select the optimal temporal resolution by exploring several or using available information about population responses to climate change.

Introduction

Ecological niche models (ENMs) are commonly used to estimate the potential range of species based on the association between occurrence sites and their environmental conditions (Peterson 2001). Under the traditional ENM framework, occurrences are associated with a single long-term environmental average without considering the temporal match of the observation date and the period of the environmental baseline (Ingenloff and Peterson 2021, Stewart et al. 2021). This temporal occurrence-environment mismatch may result in less accurate distribution models when occurrences do not align with the temporal range of environmental data, thereby impacting estimations of how biodiversity will respond to environmental change (Roubicek et al. 2010, Nadeau et al. 2017). While constraining the temporal range of the occurrence data may help address this issue (Fig. 1a), practical applications are challenging due to several factors. Examples of these factors include the usually limited availability of occurrence data within long-term temporal ranges (Roubicek et al. 2010), the need for methodologies that incorporate temporal variability in model building (Milanesi et al. 2020, Ingenloff and Peterson 2021, Pang et al. 2022, Dobson et al. 2023), and the lack of guidelines to obtain environmental variables reflecting non-traditional temporal ranges. ENMs commonly rely on climatic variables, such as temperature and precipitation, usually derived from the interpolation of monthly data obtained by weather stations (e.g., Climatic Research Unit dataset; Harris et al. 2020). Although these data are available for more than the past 50 years, climatic variables are typically summarized as a single period that corresponds to a climatological standard normal, a 30-year average in which the last year of the period ends with a zero (e.g., 1971-2000). Bioclimatic variables in global climatic datasets are created using these long-term periods as reference (Karger et al. 2017, Fick and Hijmans 2017, Abatzoglou et al. 2018). Unfortunately, using a single average can hide trends and variability in climate (Zimmermann et al. 2009, Ingenloff and Peterson 2021, Perez-Navarro et al. 2022, Pinilla-Buitrago 2023) or could fail to represent the current climate due to recent changes in temperature and precipitation (Livezey et al. 2007, Arguez and Vose 2011, Wilks and Livezey 2013). One approach to address these limitations is to incorporate explicit temporal variables, such as extremes during the reference period (Zimmermann et al. 2009, Moran-Ordonez et al. 2018, Stewart et al. 2021) and inter-annual variability variables (Zimmermann et al. 2009, Brodie et al. 2021, Gardner et al. 2021). However, this may not completely solve the issue of occurrence-environment mismatch since occurrences are associated with the climatic variability or extreme values that occurred after their observation date, especially for records from the beginning of the reference period (Fig. 1a).

ENMs that account for temporal mismatches have improved model performance by aligning the weather or climatic variability with the precise location and time of the occurrence data. These frameworks, known as dynamic ENMs (Milanesi et al. 2020, Dobson et al. 2023) or time-matched ENMs (Peterson et al. 2005, Ingenloff and Peterson 2021), establish correlations between occurrences and atmospheric conditions over temporal resolutions ranging from days (Abrahms et al. 2019), months (Reside et al. 2010, Welch et al. 2018, Andrew and Fox 2020, Ingenloff and Peterson 2021), seasons (Williams et al. 2017), years (Bateman et al. 2012, VanDerWal et al. 2013), or even a decade (Smith et al. 2019), and are designed to match the species' response time to the changing variable (referred to in this study as a time-matched approach; Fig. 1b). For example, Abrahms et al. (2019) used tracked satellite data to create daily suitability maps of blue whales. In another study, monthly to yearly variability in temperature and precipitation improved predictions of geographic ranges in highly mobile birds (Reside et al. 2010, VanDerWal et al. 2013, Williams et al. 2017, Andrew and Fox 2020, Ingenloff and Peterson 2021).

While most studies using time-matched ENMs focus on species with high mobility (the ability of individuals to move or spread in the environment), only a few studies have addressed its utility for models of species ranges with limited dispersal capacity. This is particularly relevant for non-volant small mammals, which may exhibit greater sensitivity to environmental variation due to their slower ability to track suitable conditions (Schloss et al. 2012). For instance, Smith et al. (2019) used a time-matched approach of ten years to generate models of American Pika distribution. Similarly, Bateman et al. (2012) used a five-year matching to uncover competitive interactions between two bettong species, small marsupials in Australia, in which the time-matched approach was able to define better range edges. These studies are some of the few examples in which a temporal resolution longer than 12 months is used. Of these, Bateman et al. (2012) is the only study that compares against a long-term average. Despite the potential advances in time-matched ENMs, it remains unclear whether and in which kinds of cases different temporal resolutions of the environment (ranging from one year to less than 30 years) could perform better or as accurately as traditional ENMs that use long-term averages.

This study aims to gain insights into the effectiveness of matching the timing of occurrences and environmental data in the modeling framework in a species with low mobility. It evaluates the performance and ecological plausibility of different temporal resolutions of occurrence-environment matching to predict the potential distribution of a non-volant small mammal, the Mexican small-eared shrew *Cryptotis mexicanus* , in the current time. Additionally, the results of this time-matching method were compared with a standard 30-year average model and differences in the prediction of the current distribution of *C. mexicanus* are explored.

Methods

In this study, I used the Mexican small-eared shrew, *Cryptotis mexicanus*, to compare the model performance of the time-matched approach against the standard 30-year approach (STA). The workflow consisted of five steps: (i) collection of occurrence and background data, (ii) generation of bioclimatic variables, (iii) comparison of environmental values, (iv) training and selection of models, and (v) comparison between models (Fig. 2). *Cryptotis mexicanus* is a cloud forest specialist that inhabits the Sierra Madre Oriental and the mountainous areas of Oaxaca State in Mexico. It is a good model species for exploring predictions of environmental suitability using several temporal resolutions for three reasons. First, its geographical distribution seems largely constrained by moisture and temperature, which allows for informative ENMs based on a few bioclimatic variables (Guevara et al. 2018, Guevara 2020). Second, shrews are considered one of the mammalian groups with the lowest individual dispersal capabilities (Schloss et al. 2012, Urban et al. 2013), which enables me to test the reliability of time-matched ENMs in a species with low mobility. Third, it has a fairly short, estimated generation time of 330 days, given its body size (Pacifici et al. 2013), which allows me to explore the existence of different time lag relationships between population dynamics and the temporal resolutions of the environmental data (one, five, and ten years prior to the records).

Input data: Occurrences

To train the models, georeferenced and taxonomically verified occurrences of *C. mexicanus* were obtained from a taxonomic study of the genus (Guevara and Sánchez-Cordero 2018). Metadata for the occurrences used in that study did not include the date of observation, which is necessary to filter the occurrences points by date. Therefore, I retrieved the capture date of each specimen from the *Sistema Nacional de Información sobre Biodiversidad de México* (SNIB) (Sánchez-Cordero and Guevara 2016) using the museum collection name and catalog number. To compare models between the 30-year average and different temporal resolutions (one, five, and ten years), occurrences points collected between January 1971 and December 2000 were used in model training, to match the months used to build the standard climatic normal (Fig. 3a). Occurrences points were selected per cell per month (with "month" defined as any of the 360 months between 1971 and 2000; 30 years \times 12 months). If two observations occurred in the same locality in different months (e.g., November 2000 and December 2000), both localities were retained for model training (see Supplemental material, Table S1).

I used occurrences reported outside the temporal range of the model training to evaluate the models. This second set consisted primarily of occurrences with observation dates recorded before 1971 and after 2000 from two published studies (Sánchez-Cordero and Guevara 2016, Guevara and Sánchez-Cordero 2018). Additionally, occurrences from recent literature (Hernández-Flores and Rojas-Martínez 2010, Mayen-Zaragoza et al. 2019) or recent fieldwork by colleagues and myself were incorporated into this withheld dataset. To ensure that the withheld dataset did not include occurrences near those used in model training, the withheld occurrences needed to be at a distance greater than 10 km away from a training occurrence (see Supplemental material, Table S2). Finally, a spatial thinning of 10 km was applied to reduce the spatial bias of the withheld data. Information on the environmental values assigned to the withheld dataset is explained later in the model comparison section.

Input data: Background

To correct for geographical and temporal sampling biases of the occurrence data, I created a weighted target-group background (Fig. 3b; sensu Anderson 2003) using data from the Global Biodiversity Information Facility (GBIF.org 2022). Target-group method uses species observed with similar methods (i.e., Sherman and pitfall traps) as background, assuming that points have similar sampling bias as the focal species (Phillips and Dudík 2008, Phillips et al. 2009), and therefore cancelling out any spatial and temporal bias in the occurrence data (Reside et al. 2010, VanDerWal et al. 2013). To ensure similar sampling as the occurrences, I used occurrence records of preserved specimens from the same collections where the training occurrence data were obtained. The Mexican collections used were: *Colección Nacional de Mamíferos, Instituto de Biología* (CNMA); *Museo de Zoología 'Alfonso L. Herrera', Facultad de Ciencias* (MZFC); and *Colección de Mamíferos, Universidad Autónoma Metropolitana Unidad Iztapalapa*(UAMI). From these collections, I used records of Mexican small non-volant mammals captured between January 1971 and December 2000 (same

date range as the occurrence data) from the orders Soricomorpha, Rodentia (only families Cricetidae and Heteromyidae), and Didelphimorphia (only genus *Marmosa*) as background points. Background points were removed if they were not located within a one-degree point buffer around the training occurrences (between 102 to 106 km at the latitude of occurrences). Finally, I used the CoordinateCleaner R package to identify and remove background points mistakenly georeferenced in the ocean (Zizka et al. 2019).

Bioclimatic variables

Climate data were obtained from the Terraclimate dataset, which provides monthly temperature and precipitation data at a resolution of 5 arcmin (Abatzoglou et al. 2018). The ClimateR R package (*htt-ps://github.com/mikejohnson51/climateR*) was used to download the climate data in R. Four bioclimatic variables – maximum temperature of the warmest month (bio05), minimum temperature of the coldest month (bio06), precipitation of the wettest month (bio13), and precipitation of the driest month (bio14) – were selected for model training. These variables were chosen based on previous studies of the species, which have established their usefulness when capturing the averages of the extreme temperature and precipitation conditions (Guevara et al. 2018).

To analyse the impact of time-matched approaches in comparison to models without time-matching, I created bioclimatic variables of different temporal resolutions. First, a single set of the standard 30-year bioclimatic average variables (STA; Fig. 1a) was created using monthly data from January 1971 to December 2000, based on a modified version of the biovars() function of the dismo R package (Hijmans et al. 2021) that use terra R package (Hijmans 2022) for faster processing. Multiple bioclimatic variables were then generated using a one-forward month moving window to explore predictions using several temporal resolutions of occurrence-environment matching. For instance, the one-year occurrence-environment matching (T01) used bioclimatic values generated within the 12 months preceding the occurrence observation, including the month of observation. Similarly, the five-year temporal resolution (T05) used 60 months, and the ten-year temporal resolution (T10; Fig. 1b) used 120 months. Environmental values for both occurrence and background points were obtained by aligning the collection date with the last month of each window period. For example, if an occurrence record was observed in January 1971, the bioclimatic values used for ten-year matching corresponded to the average of 120 months between February 1961 and January 1971.

Comparing environmental values before model training

Before building models or comparing their performance, it was essential to evaluate whether the environmental values between the standard approach and the temporal resolutions differed. Only if the environmental values are different is it expected that the models will also differ. Here, time-matched bioclimatic values of the occurrences (i.e., T01, T05, T10) were compared against those of the standard 30-year average values (STA). The differences between values were evaluated qualitatively with Bland-Altman plots (Bland and Altman 1986, 1995) and quantitatively with paired Wilcoxon signed-rank tests (Wilcoxon 1945). The Bland-Altman plots help to visualize the differences in environmental values between the 30-year average and the respective temporal resolutions per occurrence. In the X-axis, each occurrence is represented by the mean of two values (30-year and one temporal resolution), and the Y-axis denotes the difference between the same values. If the difference between the two bioclimatic variables at a particular site of occurrence is larger than zero, it means that the 30-year average value is higher than the average value obtained by the time-matched approach; on the other hand, a negative value means that the 30-year average value is lower than the time-matched approach. For example, if a reported difference is -20 mm for precipitation, the standard approach was lower by 20 mm compared to the time-matched approach. For the Wilcoxon signed-rank tests (12 total), a Bonferroni adjustment was applied separately to each of the three temporal resolution comparisons against the standard approach (i.e., STA vs. T01, STA vs. T05, STA vs. T10 each were considered a family of comparisons) to reduce the probability of obtaining false positives when performing multiple comparisons. Each comparison group (e.g., STA vs. T01) consisted of the four paired Wilcoxon signed-rank tests corresponding to each of the four bioclimatic variables.

Model training and model selection

Maxent v3.4.3 (Phillips et al. 2017) was used through ENMeval v2 (Kass et al. 2021) to train the models for this study. Before model comparisons, I first selected the optimal model complexity for each approach (STA, T01, T05, T10) by running and testing 500 combinations of feature classes (linear, linear-quadratic, hinge, linear-quadratic-hinge, linear-quadratic-hinge-product) and regularization multipliers (0.1 to 5, steps of 0.1). The selection of the best model in each approach was based on the lowest corrected Akaike Information Criterion (AICc), as it provides a measure to balance model complexity and predictability (Warren and Seifert 2011). To account for the elongated distribution of the occurrences along the Sierra Madre Oriental, occurrence and background data were divided into four spatial bins for cross-validation using a latitudinal block partition. Validation metrics, specifically the average Area under the Curve (AUC), the continuous Boyce Index (CBI), and the 10th percentile of training omission rate (OR), were used to quantify the performance of the selected model in each approach. The AUC values range from 0 to 1, with values closer to 1 indicating better model performance. Similarly, the CBI values range from -1 to 1, with values closer to 1 indicating a better model performance (Hirzel et al. 2006). On the other hand, the OR ranges from 0 to 1, and a lower value indicates a better prediction of the occurrences.

Model comparison

After selecting the best model for each approach, the standard 30-year average model (STA) was compared against each time-matched approach (T01, T05, T10). A repeated random k -fold cross-validation (Berrar 2019) was used to obtain 1000 model iterations (4 folds \times 250 repeats) using the same feature classes and regularization multipliers of the best model per approach. At each iteration, the same training and validation folds were used for comparing pairwise the standard approach against each of the three temporal resolutions (T01, T05, T10), using AUC, CBI, and OR metrics (each separately). A 10th percentile of training omission rate of the fully withheld data (OR-W) was also utilized for model evaluation. As described above, these withheld data were outside the temporal range used for model training (i.e., outside the 1971 to 2000 period). Nevertheless, for simplicity, environmental values from the standard 30-year average were assigned to the withheld occurrences to ensure consistency in interpreting and comparing all geographic predictions.

Significant differences in each metric were identified using a correlated t-test (Bouckaert 2003, Nadeau and Bengio 2003). This test accounts for deviations from the standard t-test assumption of data independence between iterations, which is invalid in the present analysis because the occurrences are shared between the k-folds. A Bonferroni adjustment was used to correct multiple comparisons for each of the three validation metrics (AUC, CBI, OR) and one evaluation metric (OR.W); i.e. each metric constituted a family. Each family consisted of three correlated t-tests, each obtained by comparing the time-matched approach (T01, T05, T10) against the 30-year standard (STA). Finally, the geographic predictions were visually inspected to evaluate ecological plausibility.

Results

Comparing environmental values before model training

Following the temporal filtering, the analysis included 53 records observed between January 1971 and December 2000 (see Supplemental material, Table S1). The paired Wilcoxon signed-rank tests revealed significant differences in the environmental values of the occurrences when comparing the standard 30-year average approach (STA) to each of the temporal resolutions (T01, T05, T10) for the maximum temperature of the warmest month (bio05) and the precipitation of the driest month (bio14). In addition, the precipitation of the wettest month (bio13) showed a significant difference when comparing the 30-year standard average (STA) against the one-year temporal resolution (T01). The minimum temperature of the coldest month (bio06) did not present significant differences between the approaches. The Bland-Altman plots also confirmed the quantitative results, showing the difference between the standard average and the temporal resolutions (Fig. 4). Specifically, significant differences in environmental values were observed when the mean values of the subtraction between the 30-year average and the temporal resolution (Fig. 4). Specifically from the y-axis zero. Bio06 indicated a temporal pattern where older records (dark orange) had lower temperatures when doing a time-matched approach, while recent occurrences (light oran-

ge) had higher temperatures. This temporal pattern was not evident for precipitation-based variables (bio13 and bio14) and the other temperature variable (bio06). Moreover, the magnitude of the difference decreases for all the variables as the temporal resolution is lower (i.e., from one-year to ten-year). For example, the difference between standard and time-matched approaches in some occurrences was as high as 2°C when comparing a one-year temporal resolution to the 30-year average but decreased to 0.5°C when comparing against the ten-year temporal resolution.

Model selection

Along with the 53 occurrences, 1401 background points were selected and used for model training. The best models selected by the delta.AIC for each of the four approaches shared a similar combination of the feature classes (Linear, Quadratic, and Hinge), without hinge features in the one-year temporal resolution (parameters and validation metrics of each model are shown in Fig. 5). Selected regularization multipliers range from 0.3 to 2.2. The standard approach yielded the highest validation AUC, although it had the lowest CBI compared to the time-matched approaches. The validation omission rate (OR) of the 30-year average was slightly higher for the standard approach than for the other time-matched approaches. In terms of the geographic predictions of each approach, shorter temporal resolutions resulted in more extensive predictions (Fig. 5).

Model comparison

After conducting repeated random k-fold cross-validation, I observed that the ten-year resolution approach outperformed the standard 30-year average approach in terms of the omission rate for withheld data, while the other resolutions performed the same or worse based on some validation metrics. First, the AUC validation values of the standard approach were significantly higher than those of the one-year and five-year timematched approaches (Fig. 6a). When comparing with the ten-year time-matched approach, the AUC values of the standard approach were still higher, but the difference was not significant. The Continuous Boyce Index (Fig. 6b) and the 10th percentile of training omission rate (Fig. 6c) showed similar (non-significant) values between the standard and time-matched approaches. However, the omission rate based on the 18 fully withheld occurrences (OR.W) indicated a higher omission rate for the 30-year standard approach than for each of the three temporal resolutions (Fig. 6d).

Visual inspection showed that the standard 30-year average and the ten-year time-matched approaches provide more accurate predictions than the other two approaches (Fig. 7 and Fig. S1). The one-year matching approach overpredicts suitability in the northern and southern parts of the study area, while the predictions of the five-year matching approach are narrower than the one-year approach but still contain some unrealistic predictions of suitable conditions to the west of the Sierra Madre Oriental between 20°N and 22°N. In contrast, the standard and ten-year matching approaches have similar predictions. However, there is a larger prediction at the north of the potential distribution for the ten-year matching approaches. A more detailed comparison of ecological plausibility between these two approaches is found in Supplemental material (Fig. S2). Moreover, the number of times the withheld data were predicted as suitable (represented by black circles in Fig. 7) shows that the time-matched approaches have higher prediction rates than the standard approach (which supports the quantitative test).

Discussion

This study found differences in the performance and ecological plausibility of ENMs that incorporated temporal occurrence-environment matching compared to models that used a standard 30-year average. Using *Cryptotis mexicanus* as a use-case, the results indicated that the ten-year temporal resolution performed similar to the standard 30-year average approach based on cross-validation AUC, CBI and OR (with no significant difference). Moreover, the results even showed an improvement in performance when considering the fully withheld omission rate, with an average rate that is lower by 0.1 than the standard approach. However, given the significantly lower AUC values of other shorter temporal resolutions (one and five years), this result may not be generalized to all temporal resolutions. The ecological plausibility interpretation supported the findings of model performance, showing that the ten-year resolution had as much of a realistic

geographical prediction of suitability as the standard 30-year approach for almost all the known shrew range, while the shorter temporal resolutions resulted in unrealistic estimates of the potential shrew distribution.

Time-matched approaches have proven successful in obtaining reliable model predictions for other species of small non-volant mammals (with low vagility), but none of these studies compared the performance and ecological plausibility of different temporal resolutions, as presented in this study. For instance, time-specific models of American Pika (Smith et al. 2019) and Australian Bettongs (Bateman et al. 2012) achieved favourable results using resolutions of ten and five years, respectively. The observed differences in model performance across the three temporal resolutions here suggest that using an uninformed time-matched approach will not translate into better models for small non-volant mammals or other species with low mobility. Ideally, the temporal resolution selection should be based on the species' biology, specifically the time lag in which a population could be impacted by the change in the variables used in the model training (Levin 1992, Nadeau et al. 2017). For instance, shorter-term resolutions, such as yearly, monthly, or even daily resolutions (Reside et al. 2010, Abrahms et al. 2019), might be more useful for highly mobile species (Ingenloff and Peterson 2021) than for species of low mobility, such as C. mexicanus. Unfortunately, most species worldwide lack the necessary demographic information to establish the time lag of population response, although often information could be borrowed from taxonomically related species with similar body sizes and natural history information. As an alternative, time-matched models should explore several temporal resolutions and tune them empirically.

For the ten-year temporal resolution, the modified version of the target-group background obtained realistic predictions of the shrew's distribution, except for the northern region. This large area predicted in the north is not expected, as it corresponds to xeric shrublands and mesic lowland areas in which populations of *C. mexicanus* never have been found. This could be due to the minimal representation of the background data in the northern area (as seen in Fig. 3b), which may fail to capture the full range of environmental conditions due to under-sampling by biologists. In such cases, the target-group background tends to overcompensate for the least sampled areas by increasing the number of false positives (Syfert et al. 2013, Barber et al. 2022). This tendency can be observed even in the standard 30-year average when used to correct for spatial bias, but it increases when also correcting the temporal bias in occurrence-environment approaches. Furthermore, while the model performance may be affected by occurrences removed because they fell outside the temporal range between 1971 and 2000, the time-matched approaches allow the use of occurrences traditionally discarded in model training, such as occurrences from the exact location but captured at different times, which could experience different environmental conditions. This situation can benefit species with a restricted distribution or where spatial thinning (Boria et al. 2014) may significantly reduce the occurrence data available for modelling.

Further directions

Time-matched models have potential benefits beyond current estimations of biodiversity distributions. One such benefit is the ability to estimate how species are responding and will respond to climate change, based on fewer years of future climatic projections. Occurrence-environment matching approaches can capture extreme weather events hidden in long-term averages, events which could drastically affect population growth or even drive local extirpation depending on the species (Frederiksen et al. 2008, Bauerfeind and Fischer 2014, Marcelino et al. 2020). Studies have shown that even in cases where models that incorporate the frequency of extreme events have similar geographic predictions to standard average models in the current time, they could differ considerably in future geographic estimations (Moran-Ordonez et al. 2018). Given that the frequency of extreme events is becoming higher due to global warming, studies that incorporate extreme events intrinsically in model training (as in this study) or by using extreme-derived variables (e.g., Zimmermann et al. 2009, Feldmeier et al. 2018) should be considered in any conservation study that accounts for climate change.

Alongside investigating the differences between time-matched and standard approaches in climate change impacts on species ranges, further studies should explore the benefits of using time-matched remote sensing data. Derived remote-sensing variables such as tree forest cover (Hansen et al. 2013) or the Normalized Difference Vegetation Index may be used together with temperature and precipitation at the exact same temporal resolution to get almost real-time estimations of habitat suitability (Crego et al. 2022, Paz et al. 2022). For instance, based on the positive results using a ten-year resolution for estimating the *C. mexicanus* distribution, time-matching of traditional bioclimatic variables could be combined with the two decades of data captured by remote sensing instruments. When sufficient recent occurrence data exist, this represents a viable pathway that holds advantages over the common alternative of building a climate-only model and post-processing it by masking out regions no longer holding suitable habitat (e.g., based on remote sensing; Merow et al. 2022). This line of further research integrating disparate environmental data streams and temporal matching could enable us to obtain better model predictions and forecast how species respond to the changing environment.

References

Abatzoglou, J. T. et al. 2018. TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958-2015. - Sci Data 5: 170191.

Abrahms, B. et al. 2019. Dynamic ensemble models to predict distributions and anthropogenic risk exposure for highly mobile species. - Diversity and Distributions 25: 1182–1193.

Anderson, R. P. 2003. Real vs. artefactual absences in species distributions: tests for Oryzomys albigularis (Rodentia: Muridae) in Venezuela. - J. Biogeogr. 30: 591–605.

Andrew, M. E. and Fox, E. 2020. Modelling species distributions in dynamic landscapes: The importance of the temporal dimension. - J. Biogeogr. 159: 2533.

Arguez, A. and Vose, R. S. 2011. The Key to Deriving Alternative Climate Normals. - Bull. Am. Meteorol. Soc. 92: 699–704.

Barber, R. A. et al. 2022. Target-group backgrounds prove effective at correcting sampling bias in Maxent models. - Divers. Distrib. 28: 128–141.

Bateman, B. L. et al. 2012. Nice weather for bettongs: using weather events, not climate means, in species distribution models. - Ecography 35: 306–314.

Bauerfeind, S. S. and Fischer, K. 2014. Simulating climate change: temperature extremes but not means diminish performance in a widespread butterfly. - Popul. Ecol. 56: 239–250.

Berrar, D. 2019. Cross-Validation. - In: Ranganathan, S. et al. (eds), Encyclopedia of Bioinformatics and Computational Biology. Academic Press, pp. 542–545.

Bland, J. M. and Altman, D. G. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. - Lancet 1: 307–310.

Bland, J. M. and Altman, D. G. 1995. Comparing methods of measurement: why plotting difference against standard method is misleading. - Lancet 346: 1085–1087.

Boria, R. A. et al. 2014. Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. - Ecol. Modell. 275: 73–77.

Bouckaert, R. R. 2003. Choosing between two learning algorithms based on calibrated tests. - Proceedings of the Twentieth International Conference on International Conference on Machine Learning 3: 51–58.

Brodie, S. et al. 2021. Exploring timescales of predictability in species distributions. - Ecography 44: 832–844.

Crego, R. D. et al. 2022. Implementation of species distribution models in Google Earth Engine. - Divers. Distrib. 28: 904–916.

Dobson, R. et al. 2023. dynamicSDM : An R package for species geographical distribution and abundance modelling at high spatiotemporal resolution. - Methods Ecol. Evol. in press.

Feldmeier, S. et al. 2018. Climate versus weather extremes: Temporal predictor resolution matters for future rather than current regional species distribution models. - Diversity and Distributions 24: 1047–1060.

Fick, S. E. and Hijmans, R. J. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. - Int. J. Climatol. 37: 4302–4315.

Frederiksen, M. et al. 2008. The demographic impact of extreme events: stochastic weather drives survival and population dynamics in a long-lived seabird. - J. Anim. Ecol. 77: 1020–1029.

Gardner, A. S. et al. 2021. Accounting for inter-annual variability alters long-term estimates of climate suitability. - J. Biogeogr. 48: 1960–1971.

GBIF.org 2022. GBIF. - GBIF Occurrence Download

Guevara, L. 2020. Altitudinal, latitudinal and longitudinal responses of cloud forest species to Quaternary glaciations in the northern Neotropics. - Biol. J. Linn. Soc. Lond. 130: 615–625.

Guevara, L. and Sanchez-Cordero, V. 2018. Patterns of morphological and ecological similarities of smalleared shrews (Soricidae, Cryptotis) in tropical montane cloud forests from Mesoamerica. - System. Biodivers. 16: 551–564.

Guevara, L. et al. 2018. Toward ecologically realistic predictions of species distributions: A cross-time example from tropical montane cloud forests. - Glob. Chang. Biol. 24: 1511–1522.

Hansen, M. C. et al. 2013. High-resolution global maps of 21st-century forest cover change. - Science 342: 850–853.

Harris, I. et al. 2020. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. - Sci Data 7: 109.

Hernandez-Flores, S. D. and Rojas-Martinez, A. E. 2010. Lista actualizada y estado de conservacion de los mamiferos del Parque Nacional El Chico, Hidalgo, Mexico. - Acta Zool. Mex. 26: 563–583.

Hijmans, R. J. 2022. Spatial Data Analysis [R package terra version 1.5-17]. in press.

Hijmans, R. J. et al. 2021. Species distribution modeling [R package dismo version 1.3-5]. in press.

Hirzel, A. H. et al. 2006. Evaluating the ability of habitat suitability models to predict species presences. - Ecol. Modell. 199: 142–152.

Ingenloff, K. and Peterson, A. T. 2021. Incorporating time into the traditional correlational distributional modelling framework: A proof-of-concept using the Wood Thrush Hylocichla mustelina. - Methods Ecol. Evol. 12: 311–321.

Karger, D. N. et al. 2017. Climatologies at high resolution for the earth's land surface areas. - Sci Data 4: 170122.

Kass, J. M. et al. 2021. ENMeval 2.0: Redesigned for customizable and reproducible modeling of species' niches and distributions. - Methods Ecol. Evol. 12: 1602–1608.

Levin, S. A. 1992. The problem of pattern and scale in ecology: The Robert H. macarthur award lecture. - Ecology 73: 1943–1967.

Livezey, R. E. et al. 2007. Estimation and Extrapolation of Climate Normals and Climatic Trends. - J. Appl. Meteorol. Climatol. 46: 1759–1776.

Marcelino, J. et al. 2020. Extreme events are more likely to affect the breeding success of lesser kestrels than average climate change. - Sci. Rep. 10: 7207.

Mayen-Zaragoza, M. et al. 2019. First record of shrews (Eulipotyphla, Soricidae) in the Sierra de Otontepec, an isolated mountain in Veracruz, Mexico. - Therya 10: 59.

Merow, C. et al. 2022. Operationalizing expert knowledge in species' range estimates using diverse data types. - Frontiers of Biogeography 14: e53589.

Milanesi, P. et al. 2020. Integrating dynamic environmental predictors and species occurrences: Toward true dynamic species distribution models. - Ecol. Evol. 10: 1087–1092.

Moran-Ordonez, A. et al. 2018. Modelling species responses to extreme weather provides new insights into constraints on range and likely climate change impacts for Australian mammals. - Ecography 41: 308–320.

Nadeau, C. and Bengio, Y. 2003. Inference for the Generalization Error. - Mach. Learn. 52: 239–281.

Nadeau, C. et al. 2017. Coarse climate change projections for species living in a fine-scaled world. - Glob. Chang. Biol. 23: 12–24.

Pacifici, M. et al. 2013. Generation length for mammals. - Nature Conservation 5: 89–94.

Pang, S. E. H. et al. 2022. Occurrence-habitat mismatching and niche truncation when modelling distributions affected by anthropogenic range contractions. - Divers. Distrib. 28: 1327–1343.

Paz, A. et al. 2022. A framework for near-real time monitoring of diversity patterns based on indirect remote sensing, with an application in the Brazilian Atlantic rainforest. - PeerJ 10: e13534.

Perez-Navarro, M. A. et al. 2022. Comparing climatic suitability and niche distances to explain populations responses to extreme climatic events. - Ecography 2022: e06263.

Peterson, A. T. 2001. Predicting species' geographic distributions based on ecological niche modeling. - Condor 103: 599.

Peterson, A. T. et al. 2005. Time-specific ecological niche modeling predicts spatial dynamics of vector insects and human dengue cases. - Trans. R. Soc. Trop. Med. Hyg. 99: 647–655.

Phillips, S. J. and Dudik, M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. - Ecography 31: 161–175.

Phillips, S. J. et al. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. - Ecol. Appl. 19: 181–197.

Phillips, S. J. et al. 2017. Opening the black box: an open-source release of Maxent. - Ecography 40: 887–893.

Pinilla-Buitrago, G. E. 2023. Predicting potential range shifts using climatic time series and niche models: A Neotropical montane shrew's case. - Ecol. Inform. 77: 102212.

Reside, A. E. et al. 2010. Weather, not climate, defines distributions of vagile bird species. - PLoS One 5: e13569.

Roubicek, A. J. et al. 2010. Does the choice of climate baseline matter in ecological niche modelling? - Ecol. Modell. 221: 2280–2286.

Sanchez-Cordero, V. and Guevara, L. 2016. Modelado de la distribución potencial de las musaranas (Mammalia, Soricidae). - Instituto de Biologia. Universidad Nacional Autonoma de Mexico. Comisión nacional para el conocimiento y uso de la biodiversidad.

Schloss, C. A. et al. 2012. Dispersal will limit ability of mammals to track climate change in the Western Hemisphere. - Proc. Natl. Acad. Sci. U. S. A. 109: 8606–8611.

Smith, A. B. et al. 2019. Alternatives to genetic affinity as a context for within-species response to climate. - Nat. Clim. Chang. 9: 787–794.

Stewart, S. B. et al. 2021. Climate extreme variables generated using monthly time-series data improve predicted distributions of plant species. - Ecography 44: 626–639.

Syfert, M. M. et al. 2013. The Effects of Sampling Bias and Model Complexity on the Predictive Performance of MaxEnt Species Distribution Models. - PLoS One 8: e55158.

Urban, M. C. et al. 2013. Moving forward: dispersal and species interactions determine biotic responses to climate change. - Ann. N. Y. Acad. Sci. 1297: 44–60.

VanDerWal, J. et al. 2013. Focus on poleward shifts in species' distribution underestimates the fingerprint of climate change. - Nat. Clim. Chang. 3: 239–243.

Warren, D. L. and Seifert, S. N. 2011. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. - Ecol. Appl. 21: 335–342.

Welch, H. et al. 2018. Using temporally explicit habitat suitability models to assess threats to mobile species and evaluate the effectiveness of marine protected areas. - J. Nat. Conserv. 41: 106–115.

Wilcoxon, F. 1945. Individual Comparisons by Ranking Methods. - Biometrics Bulletin 1: 80-83.

Wilks, D. S. and Livezey, R. E. 2013. Performance of Alternative "Normals" for Tracking Climate Changes, Using Homogenized and Nonhomogenized Seasonal U.S. Surface Temperatures. - J. Appl. Meteorol. Climatol. 52: 1677–1687.

Williams, H. M. et al. 2017. A temporally explicit species distribution model for a long distance avian migrant, the common cuckoo. - J. Avian Biol. 48: 1624–1636.

Zimmermann, N. E. et al. 2009. Climatic extremes improve predictions of spatial patterns of tree species. -Proceedings of the National Academy of Sciences 106: 19723–19728.

Zizka, A. et al. 2019. CoordinateCleaner : Standardized cleaning of occurrence records from biological collection databases. - Methods Ecol. Evol. 10: 744–751.



Figure 1. Examples of environmental averages and their estimation. a) 30-year standard average between January 1971 and December 2000 (360 months or 30 years). b) Matching the timing of occurrence data and the environmental average using 120 months (ten years) before the date of observation. Black points represent the date of observation at each of ten occurrence points. Greyline intervals represent the period for which bioclimatic averages are calculated.



Figure 2. Methods flowchart. Occurrence data for *Cryptotis mexicanus* and other small non-volant mammal localities for the target-group (TG) background are obtained (yellow). Monthly temperature and precipitation are used to create bioclimatic variables (ENVS) at several temporal resolutions (blue). Darker colours correspond to the standard 30-year average approach (STA), while lighter colours correspond to time-matched approaches (T01, T05, T10). Bioclimatic values are extracted to create a sample with data (SWD) file to run a MaxEnt model to select the best model for each approach (green). Occurrence data are partitioned using repeated cross-validation to obtain 1000 model iterations for each approach (purple). Model comparisons of each temporal resolution are performed against the standard 30-year average models (dashed lines).



Figure 3. (a) Occurrences of *Cryptotis mexicanus* and (b) target-group background used in model training. Z-axis in (a) and the White-Purple palette represent the date of capture of specimens (The darker colour represents recent dates). Green represents lower elevation and brown higher elevation. The polygon in (b) is the 1-degree point buffer of occurrences employed to select background points.



Figure 4. Bland-Altman plots representing the differences in environmental values between the standard 30year average approach (STA) and each of the three temporal resolutions (T01, one-year; T05, five-year; T10, ten-year). The temperature-based variables—the maximum temperature of the warmest month (BIO05) and the minimum temperature of the coldest month (BIO06) are represented in orange. Precipitation-based variables—the maximum precipitation of the wettest month (BIO13) and the minimum precipitation of the driest month (BIO14)—are represented in blue. The values in darker colours represent older observation occurrences, while lighter colours represent recent ones. The horizontal black lines represent the mean values of the subtraction, and the red dashed lines correspond to the 5th (lower) and 95th (upper) percentile of the data. Occurrences with no difference between approaches are represented by zero values on the y-axis. Positive y-axis values represent occurrences where the standard 30-year averages are higher than the time matched. In contrast, negative y-axis values correspond to cases with higher time matched values than the standard approach. Asterisks in the top-right of each plot represent significant differences obtained by a Wilcoxon signed-rank test.



Figure 5. Thresholded geographic predictions of *Cryptotis mexicanus* using the standard 30-year average approach (STA) and three temporal resolutions (T01, one-year; T05, five-year; T10, ten-year). Black areas indicate suitable environments, while grey areas represent unsuitable environments. Red points indicate occurrences used in model training (and cross-validation), while green points represent fully withheld evaluation occurrences from other time periods (not used for selecting the best model). The dashed line outlines the one-degree buffer around the training occurrences. The maps also display average validation metrics of the latitudinal block partition at the top-right corner, including the Area under the Curve (AUC), Continuous Boyce Index (CBI), and 10th percentile of training omission rate (OR). At the bottom-left corner of each map, the feature classes (fc) and regularization multiplier (rm) of the selected model are reported.



Figure 6. Difference between validation metrics for the standard 30-year average (STA) against each of the three temporal resolutions (T01, one-year; T05, five-year; T10, ten-year). Each panel shows the pairwise difference between validation metrics of 1000 iterations (4 folds \times 250 repeats) based on repeated k -fold cross-validation. The histograms show the distribution of the pairwise differences of the area under the curve (AUC; shown in A), the Continuous Boyce Index (CBI; shown in B), the 10th percentile of training omission rate of the validation folds (OR; shown in C), and the withheld data (OR-W; shown in D) between the standard and each time-matched approach. The black dashed lines represent the expected mean values if the standard and time-matched matching approaches are the same. The red lines in the histograms indicate the mean value of the difference. Mean values above the dashed line indicate higher values for the standard approach. Higher values in AUC and CBI represent a better performance of the standard approach. In contrast, for the omission rates (OR and OR-W), lower values represent a better performance of the standard approach. Asterisks in the top-right of each plot indicate significant differences obtained by the correlated t-tests.



Figure 7. Maps showing the summation of thresholded suitability maps based on 1000 iterations (4 folds x 250 repeats) of *Cryptotis mexicanus* for each of the four approaches: standard 30-year average (STA), and three temporal resolutions (T01, one-year; T05, five-year; T10, ten-year). Areas with no suitable predictions are indicated in blue, while those where most iterations predict suitability are shown in red. Occurrences that were fully withheld during model training and cross-validation are indicated with a white cross mark, while the number of times an occurrence was predicted as suitable is represented the size of a semi-transparent black circle, with larger circles indicating a greater number of suitable predictions.