Tracking multi-generational colonization of the breeding grounds by monarch butterflies in eastern North America

D. T. Tyler Flockhart1, Leonard I. Wassenaar2†, Tara G. Martin3, Keith A. Hobson2, Michael B. Wunder4 and D. Ryan Norris1

1Department of Integrative Biology, University of Guelph, Guelph, Ontario, Canada N1G 2W1
2Environment Canada, Saskatoon, Saskatchewan, Canada S7N 3H5
3Ecosystem Sciences, CSIRO, GPO 2583, Brisbane, Queensland 4001, Australia
4Department of Integrative Biology, University of Colorado Denver, PO Box 173364, Denver, CO 80217, USA

Insect migration may involve movements over multiple breeding generations at continental scales, resulting in formidable challenges to their conservation and management. Using distribution models generated from citizen scientist occurrence data and stable-carbon and -hydrogen isotope measurements, we tracked multi-generational colonization of the breeding grounds of monarch butterflies (Danaus plexippus) in eastern North America. We found that monarch breeding occurrence was best modelled with geographical and climatic variables resulting in an annual breeding distribution of greater than 12 million km² that encompassed 99% occurrence probability. Combining occurrence models with stable isotope measurements to estimate natal origin, we show that butterflies which overwintered in Mexico came from a wide breeding distribution, including southern portions of the range. There was a clear northward progression of monarchs over successive generations from May until August when reproductive butterflies began to change direction and moved south. Fifth-generation individuals breeding in Texas in the late summer/autumn tended to originate from northern breeding areas rather than regions further south. Although the Midwest was the most productive area during the breeding season, monarchs that re-colonized the Midwest were produced largely in Texas, suggesting that conserving breeding habitat in the Midwest alone is insufficient to ensure long-term persistence of the monarch butterfly population in eastern North America.

1. Introduction

Migratory species typically form complex networks consisting of multiple breeding and non-breeding populations that are demographically linked through individual migratory movements [1,2]. Thus, the conservation and management of migratory species at continental scales requires information on how different phases of the annual cycle are geographically connected [3]. Tracking the movement of individuals can be technologically challenging for small, short-lived species, for instance insects that colonize large geographical areas over multiple breeding generations in a single year [4–8].

Monarch butterflies (Danaus plexippus) are legendary for a complex, long-distance migration that traverses three countries over successive breeding generations [9,10]. Documented population declines [11] are thought to be linked to multiple threats occurring throughout the annual cycle such as habitat loss on the breeding grounds and habitat loss and degradation on the wintering grounds [11–13]. Seven decades of butterfly tagging efforts ([14], http://www.monarchwatch.org/tagmig/index.htm) have provided insight regarding how monarchs are geographically connected between these different periods of the annual cycle. While invaluable, mark–recovery outcomes are spatially biased to the location of human efforts and may be disconnected from landscape patterns of productivity that make it difficult to assess how these
spatial movements might influence population-level responses to these threats. Intrinsic markers, for instance stable isotopes, can be used to estimate natal origin, because wing chitin is metabolically inert after it has completed growth and monarchs form their wings during the relatively immobile immature stages of development [15]. Thus, stable isotopes can provide a fingerprint of natal origin regardless of how far an adult has migrated since it eclosed [15]. We used spatially explicit Bayesian assignment models [16] to combine two stable isotopes ($\delta^{13}H, \delta^{13}C$) sampled from wing tissue, and a species distribution model that served as a spatial prior to reveal how successive generations of monarch butterflies colonized the breeding grounds in eastern North America over a seven-month period.

We tested three hypotheses regarding the geographical distribution of monarchs and used the best models to produce distribution maps of probability of occurrence for each month of the breeding season. We then used these monthly occurrence distribution models as spatially explicit priors in a dual-isotope assignment model to describe how monarch butterflies in eastern North America were spatially connected throughout the breeding season. Previous isotopic studies to describe migratory connectivity during the breeding season in eastern monarchs sampled individuals during specific windows of the breeding period at restricted geographical locations [17,18]. In this study, we described connectivity throughout the entire breeding season and sampled monarchs across the entire eastern breeding range. In doing so, we addressed three long-standing questions in monarch breeding biology: (i) does movement into northern breeding areas continue throughout the breeding season or is there a single re-colonization pulse into northern breeding distributions in early summer followed by local recruitment [4,18,19]? (ii) is the re-colonization of southern areas during early autumn a result of local or long-distance dispersal [20,21]? and (iii) do monarch butterflies breeding in the south in the early autumn produce offspring that successfully migrate to Mexico [21,22]?

We used maximum likelihood to estimate monthly models for the spatial distribution of the probability of monarch occurrence [29] using data from the citizen scientist programme Journey North (www.learner.org/jnorth/; [30]). Journey North engages citizen scientists in a global study of wildlife migration, including monarch butterflies. We compiled observational location and date of first sighting for adults in spring (March–July) and autumn (August–October) between 1997 and 2011. Our goal in using this information as a prior probability in the assignment tests was simply to circumscribe the boundary of isotope-based assignment to areas in which monarchs are likely to occur at a given time interval (i.e. month) and not to model the main cohort or mean passage of monarchs that would have constrained our assignments to a more restricted area. We limited records of mainland observations east of the continental divide and excluded records of overwintering colonies in central Mexico and records from Arizona because individuals here may overwinter locally [9]. The two major assumptions of species distribution modelling using presence-only data are constant detection probability and a spatially random sampling design [29,31]. There are no published estimates of detection probability for monarchs, so our assumption of the citizen scientist database was that the reporting rate was constant. The assumption that observations come from a spatially random design was impossible to fulfill using online datasets from citizen scientists not randomly located across the study area [31]. Thus, we accounted for monarch butterfly observation records clumped around urban areas by including human population density as a partial explanatory factor and smoothing for this variable when generating predicted models of monarch distribution.

We used Akaike information criterion to select among competing models that best described monarch distribution. For each month, we started with global models containing all of the explanatory variables (see the electronic supplementary material, table S1) and removed variables stepwise to all simpler combinations of the variables. We first did this for each hypothesis (habitat suitability, geographical limits and physiological constraints), then compared the top models of each hypothesis in a similar manner by starting with all hypothesis-specific models as the global model and conducting stepwise removal of each hypothesis to arrive at a final model for each month that described monarch distribution. The relationship between each explanatory variable and occurrence was based on monarch biology (see the electronic supplementary material, table S1), and we considered only additive models to reduce the number of models and the difficulty of explaining complex interactions. As we were interested in using distribution as an informative prior to estimate natal origin, we generated a predicted breeding occurrence map across eastern North America by overlaying the monthly occurrence probability and taking the maximum monthly value of each cell in the landscape up to the focal month.

2. Material and methods

(a) Breeding distribution models

As species are rarely surveyed completely, species distribution models can be used to predict the probability of occupancy for a given geographical location. These models use observations of the focal species and environmental variables (e.g. temperature and resource availability) at observation locations to extrapolate landscape-level occupancy patterns based on geographical gradients of environmental variables. In Bayesian assignment models, the probability of occupancy for a given location over the course of the breeding season can serve as a statistical prior that spatially delineates the boundary of where butterflies were expected to occur, and hence to constrain the range of possible origins.

Monarch breeding distribution has been described by three non-mutually exclusive hypotheses (see the electronic supplementary material, table S1): (i) habitat suitability characterized by land cover (per cent trees, per cent herbaceous and per cent bare ground) and monthly normalized difference vegetation index (NDVI; [23,24]); (ii) geographical limits dictated by the timing and extent of migration characterized by latitude, longitude, altitude and slope [19,25,26]; and (iii) physiological constraints on development and movement imposed by weather conditions characterized by temperature and precipitation [19,25,27,28].
Table 1. Models explaining monthly probability of occupancy for monarch butterflies in eastern North America. (For each month, the top model is presented including the model weight (w) and number of parameters (K). Individual variables are log-transformed human population density (logPop); monthly normalized difference vegetation index (NDVI), per cent trees (Tree), per cent herbaceous (Herb) and per cent bare (Bare) land cover; latitude (Lat), longitude (Long), Altitude (Alt) and slope (Slope); mean (Mean), minimum (Min) and maximum (Max) monthly temperatures and monthly precipitation (Precip). The combination of terms in the model derives from the top model for each respective hypothesis to explain monarch distribution (see the electronic supplementary material, table S4.)

<table>
<thead>
<tr>
<th>month</th>
<th>model</th>
<th>w_i</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mar</td>
<td>logPop + NDVI + Tree + Lat^2 + Long^2 + Alt + Mean^2 + Precip^2 + Min</td>
<td>0.83</td>
<td>14</td>
</tr>
<tr>
<td>Apr</td>
<td>logPop + Lat^2 + Long^2 + Alt + Mean^2 + Precip^2 + Min</td>
<td>0.85</td>
<td>12</td>
</tr>
<tr>
<td>May</td>
<td>logPop + NDVI + Tree + Herb + Lat^2 + Long^2 + Alt + Slope + Mean^2 + Precip^2 + Min</td>
<td>0.94</td>
<td>16</td>
</tr>
<tr>
<td>June</td>
<td>logPop + NDVI + Tree + Herb + Lat^2 + Long^2 + Alt + Slope + Mean^2 + Precip^2 + Min</td>
<td>1</td>
<td>16</td>
</tr>
<tr>
<td>July</td>
<td>logPop + Lat^2 + Long^2 + Alt + Slope</td>
<td>0.62</td>
<td>8</td>
</tr>
<tr>
<td>Aug</td>
<td>logPop + NDVI + Tree + Bare + Lat^2 + Long^2 + Alt + Slope + Mean^2 + Precip^2 + Max</td>
<td>0.98</td>
<td>16</td>
</tr>
<tr>
<td>Sep</td>
<td>logPop + NDVI + Tree + Herb + Lat^2 + Long^2 + Alt + Slope + Mean^2 + Precip^2 + Max</td>
<td>1</td>
<td>16</td>
</tr>
<tr>
<td>Oct</td>
<td>logPop + NDVI + Tree + Herb + Bare + Lat^2 + Long^2 + Alt + Slope + Mean^2 + Precip^2 + Max</td>
<td>1</td>
<td>17</td>
</tr>
</tbody>
</table>

(ii) Stable isotope analysis

Wing tissue membrane was washed twice in 2:1 chloroform:methanol solution to remove surface oils and contaminants. Wing chitin subamples (1.0 ± 0.1 mg) for δ¹³C were loaded into 8.0 × 5.0 mm pressed tin capsules and analysed using continuous-flow isotope-ratio mass spectrometry (CF-IRMS). Wing chitin subamples (0.35 ± 0.02 mg) for δ²H isotopes were loaded into 4.0 × 3.2 mm pressed silver capsules and analysed using flash pyrolysis using CF-IRMS. Non-exchangeable δ²H values were obtained using the Comparative Equilibrium procedure [32] and normalized to the Vienna Standard Mean Ocean Water—Standard Light Antarctic Precipitation scales. Precision of laboratory keratin control standards was better than ±0.2% for δ¹³C and ±1.6% for δ²H. Laboratory standards and their assigned values for hydrogen isotopes were EC1 and EC2 with δ²H values of −197‰ and −54‰, respectively. For carbon isotopes, the laboratory standards were BWBII and PUGEL with assigned δ¹³C values of −18.5‰ and −13.6‰ versus the Vienna Pee Dee Belemnite standard.

(iii) Geospatial natal assignments

We assumed a bivariate normal distribution for the error term in our isotope model for assigning probability of natal origin. For each butterfly, we were interested in calculating a probability based on the correspondence between the measured δ²H and δ¹³C values and the predicted monarch δ²H and δ¹³C values of each geographically indexed cell in the landscape. The probability density of individual $i$ having location $j$ as the natal origin is $Y_j \sim N(\mu_j, \Sigma)$, where $Y_j$ is a vector of observed δ²H and δ¹³C values, $\mu_j$ is a vector of predicted δ²H and δ¹³C values derived from previously calibrated isoscapes [33,34] (see the electronic supplementary material, figure S2) and $\Sigma$ is the positive–definite variance–covariance matrix of δ²H and δ¹³C. Here, $\Sigma$ was assumed to be constant across the entire isoscape and was estimated based on all values from known-location butterflies from data in [33]. Explicitly incorporating the variance–covariance in our models acknowledges the inherent variation in isotopic measurements that influence conditional probability of origin and allows us to draw more robust inference [35]. We applied Bayes’ rule to invert the conditional probabilities of natal origin based on isotopes using a prior described by the model for probability of occurrence at time $M$, where $m$ indexes the month of capture, for location $j$ as follows:

$$f_{Y | X, M}(J = j | Y = y, X = x) = \frac{f_{X | Y}(Y = y | X = x) f_{M}(J = j | M = m)}{\sum_{j} f_{X | Y}(Y = y | X = x) f_{M}(J = j | M = m)}.$$

where $f_{Y | X, M}$ is the spatially explicit posterior probability density function for location $j$ as the true origin of individual with measured isotope value $y$ collected in month $m$, given the measured isotope values $y_j$ for locations $x_j$. The function $f_{X | Y}$ represents the conditional distribution on $Y_j$ from above. The function $f_{M}$ is the probability of occurrence for locations $j$, as described above, for the month prior to capture, $M$. Simply put, butterflies captured in June, say, were expected to result from occurrence patterns modelled for May, given that development from egg to eclosion is approximately one month [27]. The exception was butterflies with high wing wear scores (see the electronic supplementary material, figure S1) captured in April (wing wear greater than or equal to 3) and May (wing wear greater than or equal to 4) that were assumed to have overwintered in Mexico and were expected to come from a wide spatial distribution from the prior year’s breeding season [10,14,36]. For these individuals, we used the occurrence model estimate from October that included the entire annual breeding distribution.

To geographically quantify areas of production for monarchs, we determined the odds that a given assigned origin was correct relative to the odds that it was incorrect as 2:1 and coded the upper 33% of the assignment surface for each butterfly as a binary surface [37]. The odds ratio constitutes the compromise between having sufficient geographical structure in the assignments while correctly assigning the natal origin of an individual [38] and is akin to choosing a type I error rate (e.g. $\alpha = 0.05$) in a traditional statistical test to determine significance. We then summed the layers by the month of capture, which we used to roughly represent generations, except for butterflies captured in April and May that were differentiated into overwintered or first-generation individuals as above. We used the ‘raster’ package [39] in program R [40] to conduct all spatial interpretations and statistical analysis. The raw isotope data used in the analysis are presented in the electronic supplementary material, table S3.

3. Results

(a) Breeding distribution models

Human population density was significant in all models (table 1) implying that our citizen scientists’ observations were biased towards urban populated areas and model-based predictions that did not smooth for human population density would have been misleading in describing monarch occurrence (see the electronic supplementary material, figure S3). In all cases except July, more than one hypothesis was included in the
top model to describe the distribution of monarch butterflies (table 1). Variables that best described each hypothesis were remarkably similar throughout the breeding season (see the electronic supplementary material, table S4). Geographical variables occurred in all models and were consistently described by altitude and quadratic terms of latitude and longitude. Climate was found in all but two of the top models and was best described by quadratic forms of monthly mean temperature and precipitation in addition to either minimum (March–June) or maximum (August–October) temperature. Vegetation occurred in nine of the top 14 models and was best described by NDVI and several combinations of land cover types.

Monthly species distribution maps showed an increasing northward movement between March and June with notable expansion towards western and northeastern areas in June and July (see the electronic supplementary material, figure S4). Low occurrence probabilities occurred in the northern and western portion of the continent (figure 1) but the isoscape for δ13C was obtained through kriging, and therefore geographically constrained by insufficient sample collection locations (see the electronic supplementary material, figure S2). Subsequently, we excluded probabilities less than 0.05 and reclassified the distributional probability maps at lower cutoffs of 0.5, 0.25, 0.1 and 0.05 (see the electronic supplementary material, figure S4). Overall, breeding season occupancy of monarch butterflies was $4.73 \times 10^6$, $7.33 \times 10^6$, $9.92 \times 10^6$, $11.16 \times 10^6$ and $12.33 \times 10^6$ km$^2$ at the occurrence probabilities of 0.5, 0.25, 0.1, 0.05 and 0.01 (see the electronic supplementary material, figure S5).

(b) Natal origin
Natal origin of monarch butterflies showed a clear seasonal progression spanning successive generations (figure 2). Monarchs that overwintered in Mexico (captured in April and May, and had high wing wear scores) came from a broad spatial distribution spanning the northeastern coast of North America to western Texas (figure 2a). Of these 115 overwintered butterflies, 47% had southerly origins that included areas within Texas, but there were no differences in wing wear scores between individuals with or without Texas isotopic assignments (Wilcoxon test: $W = 1532$, $p = 0.47$). Patterns of geographical origin were robust when considering more restrictive aging classifications (see the electronic supplementary material, figure S6).

Most first-generation monarchs captured in April and May originated from eastern Texas and southern Oklahoma, with fewer individuals from Arkansas and Missouri or Virginia (figure 2b). Few captured individuals originated from the Gulf Coast states of Louisiana, Mississippi, Alabama, Georgia and Florida (figure 2b). By contrast, the natal locations of most second-generation monarchs captured in June were from two areas: a southerly area including northern Texas and southern Oklahoma and a more restricted northerly location centred on Illinois (figure 2c). Butterflies captured in July showed a larger range of natal origin in the Midwest that extended from eastern Ohio to western Missouri (figure 2d).

Natal origins of butterflies collected in August showed even wider distribution than June or July captures, including areas in the northeast, east coast, Midwest and western portions of the range (figure 2e). By contrast, butterflies collected in September in Texas, Ontario and Minnesota all had northern origins (figure 2f) indicating that butterflies breeding in Texas had migrated long distances while reproductively active.

The natal origins of all 839 monarch butterflies showed a broad spatial distribution that encompassed the entire breeding range in eastern North America (figure 3). However, there was a preponderance of individuals that originated between northern Texas to western Ohio, a region that extended from the southern Great Plains through the Corn Belt. There were few indications of natal origins from Mississippi, Alabama, Georgia and Florida despite the fact that areas located north of these locations were sampled extensively. Few individuals were assigned to areas in the upper Midwest including Nebraska, Iowa, Minnesota, Wisconsin and Michigan (figure 3).

4. Discussion
Using citizen science data and a novel dual-isotope continuous-surface natal assignment model, we describe, to our knowledge for the first time, the patterns of migratory connectivity of monarch butterflies over the entire breeding season in eastern North America. At a minimum 50% occurrence probability, we estimate that monarchs occupied a breeding area more than 4.5 million km$^2$. The best predictors of monarch occurrence were geographical attributes, climatic variables and to a lesser extent vegetation characteristics [19,23–25]. Overwintered butterflies that hatched in the previous autumn came from a wide geographical distribution, confirming that the discrete colonies that form in Mexico originate from over a broad range on the breeding grounds [10,14,36,41]. In general, we found a clear northward progression of natal origin over successive summer months, but during each month there were a small number of individuals that appeared to move in different patterns compared with the majority. Reproductive butterflies began to change direction in August and moved south, presumably to
encounter suitable environmental conditions for breeding at the end of the season [20,21]. The offspring of these individuals may comprise a large number of individuals in the overwintering colonies in Mexico.

(a) Occurrence

The geographical limits and physiological constraint hypotheses held the most support to describe seasonal monarch distribution patterns [19,25,28]. Geographical limitations imply migration timing and extent are largely predictable, regulated events that could be summarized in a deterministic equation of movement rates based upon static geographical features [19]. However, the physiological constraint hypothesis held nearly as much support, which suggests that distribution limits are also a function of stochastic weather patterns which predicts distributional range shifts of organisms in response to changing weather [25,42]. While land cover has a strong effect on host plant distribution and abundance [13,43,44], it had less influence on butterfly occurrence, which supports the notion that monarch butterflies are generalists that use diverse habitats during the breeding season [23,24].

Traditional approaches of estimating the probability of species occurrence rely on sufficient presence–absence data with suitable sampling designs [31]. However, these types of datasets are rarely available and recent advances towards applying presence-only datasets collected by citizen science provide an alternative means to estimate important state variables, for instance occurrence [29]. Monarch butterflies are the subject of several long-term citizen science observational programmes and these data can be applied to understand their population dynamics [13,14,24–26,30]. Most citizen science programmes rely on observations that contain bias associated with imperfect detection probability which influences direct estimation of occurrence [45]. The assumption of a constant detection probability in our study is important because monarch observations were spatially biased to areas with higher human population density that would have produced strongly biased priors used to inform probabilistic assignments using stable isotopes.

(b) Migratory connectivity

Our results suggest that migratory connectivity during the breeding season is strongly temporally dependent for monarch
Given the diverse data available for monarch butterflies, further information could be applied as Bayesian priors to inform assignment. For example, previous studies of birds have applied monitoring data of abundance [35,48] and orientation vectors from mark–recapture [49] to better inform isotope assignments. Experimental and field data that combine individual age, temperature-dependent development rates [27] and flight distance [19] could eventually lead to a predictive spatially dependent model of movement–distance that can inform assignment of butterflies over continuous space and time, and thereby account for overlapping generations of monarchs.

(c) Population dynamics and conservation

Our findings add to a growing body of evidence that indicates the agriculturally intensive ‘Corn Belt’ region of the United States Midwest is the most important area in terms of monarch productivity during the breeding season [10,13,24]. This is important because the adoption of genetically modified corn and soya bean crops are suggested to lead to decreasing milkweed abundance in fields [44] and has been implicated as one of the leading causes of population declines of monarch butterflies [11,13]. However, because butterflies that re-colonize the Midwest come largely from Texas and movement patterns are dynamic, conserving breeding habitat in the Midwest alone will be insufficient to ensure long-term persistence of the monarch butterfly population in eastern North America [3].

The results of our study provide new scientific information to estimate year-round movement patterns of monarch butterflies. Tagging data [14] coupled with studies of chemical fingerprint techniques [10,36] have previously shown that the overwintering colonies in Mexico are panmictic and comprising individuals from across a large portion of the breeding distribution. Our results, plus studies of butterflies during the spring breeding period [4,18], suggest that the first generation of monarchs produced in the Gulf Coast comprise the majority of individuals that re-colonize the northern breeding distribution. Our research provides clear evidence that there is dynamic movement of individuals throughout the breeding season, including breeding individuals that move south to breed in the autumn and their offspring contribute to overwintering populations. There is also evidence that migratory monarchs from northern areas may recruit into year-round breeding populations in Florida [50] or migrate to Cuba and be lost from the eastern breeding population [51].

Movement patterns of monarch butterflies are also likely to vary between years [30]. Similar to previous studies, our data were collected in one year, preventing us from addressing the factors that may influence interannual variation in movement rates. However, our results agree with Malcolm et al. [4] and Miller et al. [18], who studied the relative proportion of individuals moving between the Gulf Coast and the Great Lakes during spring in different years. Overall, annual variation in movement rates is probably driven by climate and weather [9] with warmer years predicted to have increased northward expansion [25]. Distribution of occurrence studies like ours conducted periodically across multiple years would be a convenient means of evaluating how variable these movement patterns are.

In eastern North American monarch butterflies, we are now able to estimate connectivity throughout the annual cycle that
could be applied in year-round population models. Recent work to link spatio-temporal environmental conditions to population dynamics across the breeding range [13,26] are hinged upon the assumption that different geographical locations are linked by transition of butterflies among study areas. As monarchs face multiple threats throughout the annual cycle [11–13], decisions regarding which conservation actions are likely to be most successful require incorporating movement patterns [3] and population dynamics. Conservation planning of long-distance migratory animals therefore requires incorporating reproduction, survival and movement into spatio-temporal population models that link changing dynamic landscapes to ensure population persistence at continental scales [1].

Acknowledgements: We gratefully acknowledge E. Howard and the thousands of volunteers who reported their observations to Journey North in addition to those who collected butterflies for this study: D. Brooks, A. Buckley, P. Cherubini, D. Clark, D. Davis, J. Ellis, C. Goodwin, D. Jackson, B. Krewoski, L. LaPlante, M. Miller, N. Miller, B. Patterson-Weber, P. Ryan, G. Stefly and K. Van Camp. Hospitality in the field was provided by K. Kiphart. Thanks to D. Fabiano, D. Davis, L. Flockhart and especially D. Flockhart for assistance in the field. Advice on spatial analyses came from T. Avgar, A. Bonnycastle and T. Lewitzky. G. Betini, G. Otis, K. Matsuura and three anonymous reviewers provided helpful feedback to improve the manuscript.


References


