Methods

Data Analysis

To best compare across the effects of each climate variable, we scaled all of the predictors and used a z-score following the binary predictor approach (Gelman & Hill, 2006). To control for spatial autocorrelation and to account for spatially structured processes independent from our regional predictors of false springs, we generate an additional 'space' parameter for the model. To generate our space parameter we first extracted spatial eigenvectors corresponding to our analyses' units and selected the subset that minimizes spatial autocorrelation of the residuals of a model including all predictors except for the space parameter (Diniz-Filho *et al.*, 2012; Bauman *et al.*, 2017) (see supplement 'Methods: Spatial parameter' for more details). We then took the eigenvector subset determined from the minimization of Moran's I in the residuals (MIR approach) and regressed them against the above residuals—i.e. number of false springs *vs.* climatic and geographical factors. Finally we used the fitted values of that regression as our space parameter, which, by definition, represents the portion of the variation in false springs that is both spatially structured and independent from all other predictors in the model (e.g. average spring temperature, elevation, etc. Griffith & Peres-Neto, 2006; Morales-Castilla *et al.*, 2012).

To estimate the probability of false spring risk across species and across our multivariate predictors we used a Bayesian modeling approach. By including all parameters in the model, as well as species levels groups, we were able to distinguish the strongest contributing factors to false spring risk while eliminating artifacts due to data availability or data distribution for specific species. We fit a Bernoulli distribution model (also know as a logistic regression) using mean spring temperature (written as MST in the model equation), NAO, elevation, distance from the coast, space, and climate change as predictors and all two-way interactions (fixed effects) and species as two-way interactions to simulate modeled groups on the main effects (Equation 1), using the brms package (Bürkner, 2017), version 2.3.1, in R (R Development Core Team, 2017), version 3.3.1, and was written as follows:

$$logit(p) = \alpha_{[i]} + \beta_{MST_{[i]}} + \beta_{DistanceCoast_{[i]}} + \beta_{Elevation_{[i]}} + \beta_{NAO_{[i]}} + \beta_{Space_{[i]}}\beta_{ClimateChange_{[i]}}$$

$$+ \beta_{MST \times Species_{[i]}} + \beta_{DistanceCoast \times Species_{[i]}} + \beta_{Elevation \times Species_{[i]}} + \beta_{NAO \times Species_{[i]}}$$

$$+ \beta_{Space \times Species_{[i]}} + \beta_{ClimateChange \times Species_{[i]}} + \beta_{MST \times ClimateChange_{[i]}}$$

$$+ \beta_{DistanceCoast \times ClimateChange_{[i]}} + \beta_{Elevation \times ClimateChange_{[i]}}$$

$$+ \beta_{NAO \times ClimateChange_{[i]}} + \beta_{Space \times ClimateChange_{[i]}}$$

$$y_i \sim Binomial(1, p) \tag{1}$$

We ran four chains, each with 2,500 warm-up iterations and 4,000 sampling iterations for a total of 6,000 posterior samples for each predictor. We evaluated our model performance based on $\hat{R}$ values that were close to one and low ratios of effective sample size estimates to total sample size (most parameters were below 1.05, with two parameters above 1.1). We additionally assessed chain convergence visually and posterior predictive checks. Due to the large number of observations in the data we used the FASRC Cannon cluster supported by the FAS Division of Science Research Computing Group at Harvard University to run the model.

The model output estimates are for *Aesculus hippocastanum* as species were used as two-way interactions to simulate modeled groups on the main effects. For the model estimate of change plots (Figure 3, Figure S3 and Figure S4), we used average predictive comparisons to determine the sum of the posteriors for each species and then calculated the mean from the summed posteriors for each predictor. For the interactions of each predictor with climate change, we simply used the main model output since these were not modeled as simulated groups (i.e., we did not have three-way interactions in the model). These model estimates were on the logit scale and were converted to probability percentages for easier interpretation by using the 'divide by 4' rule (Gelman & Hill, 2006) and then back converted to the original scale by multiplying by two standard deviations.

We additionally ran three simpler models—following the same equation with varying responses—to assess the effects of climate change on budburst, minimum temperatures between budburst and leafout and the number of false springs across species (Equation 2).

$$y_i = \alpha_{[i]} + \beta_{ClimateChange_{[i]}} + \beta_{Species_{[i}} + \beta_{ClimateChange \times Species_{[i]}} + \epsilon_{[i]}$$

$$\epsilon_i \sim Normal(0, \sigma^2) \tag{2}$$

References

Bauman D, Drouet T, Dray S, Vleminckx J (2017) Disentangling good from bad practices in the selection of spatial or phylogenetic eigenvectors. *Ecography*, **0**. doi:10.1111/ecog.03380.

Bürkner PC (2017) brms: An R Package for Bayesia Multilevel Models. *Journal of Statistical Software*, **80**, 1–28.

Diniz-Filho JAF, Bini LM, Rangel TF, Morales-Castilla I, Olalla-Tárraga MÁ, Rodríguez MÁ, Hawkins BA (2012) On the selection of phylogenetic eigenvectors for ecological analyses. *Ecography*, **35**, 239–249.

Gelman A, Hill J (2006) *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.

Griffith DA, Peres-Neto PR (2006) Spatial modeling in ecology: the flexibility of eigenfunction spatial analyses. *Ecology*, **87**, 2603–2613.

Morales-Castilla I, Olalla-Tarraga MA, Purvis A, Hawkins BA, Rodriguez MA (2012) The imprint of cenozoic migrations and evolutionary history on the biogeographic gradient of body size in new world mammals. *The American Naturalist*, **180**, 246–256.

R Development Core Team (2017) R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*.