A novel approach to detecting changes in dynamic rules governing population fluctuations

C.A. Bahlai And E.F. Zipkin

**Abstract**

**Introduction**

In this paper, we develop a generalizable tool for detecting shifts in dynamic regime in population time series data. The regime shift model is based on a script initially developed in (Bahlai, vander Werf, et al. 2015) to understand changes governing the internal dynamics of population regulation in the invasion process of *Harmonia axyridis* in the two decades following their arrival in southwestern Michigan. This model was able to detect the precise timing external factors affecting the internal drivers of population cycling (namely, the invasion and then subsequent pesticidal control of a preferred prey item), suggesting that the model itself may have broader applicability in understanding drivers of population dynamics for other species important in conservation and ecosystem management.

This paper describes how this tool was developed as a suite of functions for examining population time series data for the presence, location, and magnitude of shifts in dynamic regime.

**Methods**

A model script, simulations, and two case studies using population data from insect population monitoring were scripted and run in R Version 3.3.3 “Another Canoe” (R Development Core Team 2017) run within the RStudio Integrated development environment 1.0.136 (RStudio Team 2015). All data manipulations, analyses and figure scripts, including the complete development history, are publicly available in a Github repository at <https://github.com/cbahlai/monarch_regime>.

*The model*

For our purposes, we use the Ricker model to describe the dynamics within a system (Turchin 2003).

The Ricker model is a single-variable discrete time model in which N(t+1), the population at time t+1, is a function of N(t), where N(t) is the measure of population size in year t, and the parameters K and r (the carrying capacity and per capita yearly rate of increase respectively) are estimated for the population during the time period modeled. The Ricker model was selected for our purposes because 1) it does not rely on any external information, other than population data over time, to be fit; 2) only two parameters need to be estimated, and those parameters have ecologically meaningful interpretations, and 3) in our experience, the model simply fits insect population data well.

We used an iterative, model-selection based process to determine if, and when, shifts in dynamic regime had occurred within a given time series. To achieve this, first, the Ricker model was fit to the entire population time series, then the population time series was subdivided into all possible combinations of 2, 3, …, n subsets of sequential data points (hereafter, ‘break point combination’, for example, a twelve year series with a break point combination of 4, 8 would be broken into subset 1 = 1, , 2, 3, 4; subset 2 = 5, 6, 7, 8; subset 3 = 9, 10, 11, 12) and the Ricker model was fitted to each of the subsets produced for each break point combination. Break point combinations were constrained to only include subsets with more than three sequential data points to avoid over-fitting.

After fitting each subset for a given break point combination, the Akaike Information Criteria for each subset were summed together, providing an overall AIC for the fit. To further account for the effect of small series sizes, we calculated AICc (AIC correction for small samples) by using the total number of parameters estimated for the fit (where nparameters = 3 (nbreaks + 1), because for each fit, r, K and error were estimated), and this factor was added to the total AIC for each break point combination. Then, AICc values were used to rank fits for each break point combination, and fits for break point combinations with lower AICc values were considered to have better performance. When AICc values differed by two units or less, model performance was considered equivalent (Burnham and Anderson 2002), so when models were found with equivalent performance, the simplest model (ie: the one with fewest parameters and break points) was selected for further analysis. If the top-ranked models had the same number of parameters, the one with the numerically lowest AICc was considered best-ranked for the purpose of further analysis.

*Technical implementation*

The ‘regime shift detector’ was implemented as a series of R functions to enable a user starting with a standard data frame of population observations at a standard time intervals to quickly generate a report on the fit of the model described above to their own data. Herein we summarize the role of each function, but we encourage the reader to download the script file directly from https://github.com/cbahlai/monarch\_regime/blob/master/regime\_shift\_detector.R for the details of implementation, included as line-by-line comments in the script.

addNt1- takes a raw data frame with columns for year, population abundance measure, and converts it to a three column data frame with year, population in year, population in the next year, and gives the data consistent column names for use in downstream functions. Most of the functions described below require this function to be used to transform data prior to use, unless otherwise specified.

AICcorrection- takes a data series and the number of breaks used in a given fit to calculate the AICc correction factor to be added to the total AIC for the fit

rickerfit- fits the Ricker model using the Levenberg-Marquart nonlinear least squares method for nonlinear model fitting to a given data frame. This function calls the nlsLM function from minpack.lm (Elzhov et al. 2016). To aide in model convergence, the function also computes a realistic starting value for K by calculating the mean of the population (based on the assumption that a population being fit to the Ricker model is likely fluctuating around its carrying capacity). The starting value for r was set at 1.5. The function outputs a vector containing the AIC, the estimate for r, its standard error, and the estimate for K and its standard error.

splitnfit- takes a given data fame, fits the complete data series with the rickerfit function, then subsets it in two by creating a break point three years after the start of the series, calls the rickerfit function to fit the data from each subset produced there. Then the function walks through the data, increasing the break point by one time step each iteration, and compiles the AICs and break pointss used for each fit, resulting in a data frame of break point combinations and respective AICs.

findbreakable- examines the output from the splitnfit function to determine if any of the break point combinations produced might be further subdivided (ie: has enough points to not violate the rule we set to only fit series with greater than 4 points).

subsequentsplit- used output from findbreakable function to identify cases where data can further be subsetted using the splitnfit function, feeds those cases in, and compiles results together with that produced by simpler break point combinations produced by splitnfit.

nbreaker- uses splitnfit, findbreakable, and subsequentsplit, combined with input data, to create a data frame consisting of a column of all possible break point combinations, and the respective AICs of the resultant fits. This function uses an iterative approach to allow simpler functions that break a data into two parts to be used to find an unlimited number of break points (within constraints of series length).

AICtally- takes data in, subjects it to nbreaker, pulls out the AICs produced by nbreaker, adds them together and counts the number of fits performed, number of breaks in the data, computes the corrected AICc using AICcorrection. and returns these values as a data frame.

allfits- appends the results of nbreaker and AICtally together into a single data frame.

equivalentfit- takes in data, feeds it to allfits, and uses the output from allfits to pull out the subset of all equivalently-performing breakpoint combination fits (here, within 2 units of AICc), and outputs these fits as a data frame.

bestfit- feeds data to the equivalentfit function to get a data frame describing equivalent fits, and uses our decision rules (in this case, select the break point combination requiring the fewest parameter estimates from the list of equivalent models, and if there are multiple equivalent models still, select the one with the numerically lowest AICc) to output the specifics of that break point combination as a data frame.

bestmodel- feeds data to the bestfit function to identify the best break point combination, and then use that information to create a data frame describing the parameter estimates (r, K and standard error for each) for fitting the Ricker model to each of the subsets of timeseries, allowing a user to quantify how the dynamic rule found by fitting the model changes at each break point.

modelspecification- in case a user wants to investigate specification of similarly ranked models, this function takes data in the format produced by the bestfit function and produced a data frame describing the model specification of the given break point combination.

RSdetector- uses the raw time series data to produce a report, calling all the previous functions, either directly, or through other functions, with short explanatory text preceding each result. First, a simple plot of population over time is produced (N(t) by t), then data is fed to the addNt1 function, and the resultant N(t), N(t+1) data is plotted to visualize the potential for the data to conform to a Ricker curve. Then, the data is fed through the allfits function, producing a complete list of all break point combinations tested and their respective fit statistics. The data is subsequently fed through the equivalentfit and bestfit functions so that a user can assess how the decision rules specified impacted the selection of the best model. Finally, the data is fed through the bestmodel function to produce the set of regression parameters for each time series subset produced by the best break point combination found.

*Simulations*

A function was created to simulate time series data following Ricker dynamics under set break point combination conditions. The function takes values for of start year, number of years to simulate, % noise, a starting population N, starting values for K and r, a break point combination, a % change in K and a % change in r to be simulated for each break point. Percent noise was included as a means to simulate sampling error that would be observed in a real sampling plan, and was simulated by creating a continuous interval from 100%-noise to 100%+noise, randomly sampling from within that interval, and multiplying the predicted observation N(t+1) by the resultant value. Change of K and r at each break point were randomly selected by the script to either increase or decrease by the given % change.

The simulated data were fed into a function that tested if the regime shift detector model was able to identify the breaks as set for the simulation by comparing the input conditions to those output by the bestmodel function in the regime shift detector script file. Results of comparing the input to the output were encoded as follows:

1. script was successful at detecting all break points and simulation conditions
2. script identified all simulated breaks, but also found one or more ‘extra’ breaks
3. script missed one of the simulated breaks, but all others found were correct
4. script identified the correct number of breaks, but one or more breaks were mismatched
5. no correct breaks were identified by the script, or breaks were identified in a no-break scenario

A base scenario was constructed, with start year =1, number of years = 25, a starting population of 3000, a sampling error of up to 5%, a starting value for K = 2000, a starting value for r =2, a % change at each break point of 40 and 20% for K and r respectively, and a set of 0, 1, 2, or 3 break points randomly selected from within the possible values defined by start year and number of years.  
  
This script was used to test the frequency of a given result under a variety of different scenarios by changing the value of one parameter at a time from the base scenario. Specifically, the impact of different sizes of regime shift was tested by creating scenarios where the % change in r and K were individually modified at 10% intervals from 0 to 90%, the impact of length of time series was examined by extending the length of the time series by 2 year intervals from 25 to 33, and the impact of sampling error was tested at 1, 5, 10%, and every 10% interval thereafter to 90%. Each of these scenarios was run on 0, 1, 2 and 3 break point combinations, and each scenario by break point combination was iterated 500 times with newly simulated data. Results from the simulations were compiled at intervals of 50 iterations and stored as csv files within the project repository for further analysis. The complete script which simulates the data, applies the regime shift detector, and compiles the results is available at: <https://github.com/cbahlai/monarch_regime/blob/master/simulations.R>

*Case studies*

We used two case studies to test the performance of the regime shift detector script on population time series data produced under natural conditions as parts of observational experiments. Both case studies involve approximately two decades of observations of economically or culturally important insect species, however, one case examines an invasion process, and another examines a population decline, both occurring over the same time period in recent history.

*Harmonia axyridis* in southwestern Michigan

The 1994 invasion of *Harmonia axyridis* to Michigan State University’s Kellogg Biological Station in southwestern Michigan, United States was captured as part of an observational experiment to monitor agriculturally-important Coccinellidae (ladybeetles) in landscapes dominated by field crops. The coccinellid monitoring experiment was initiated in 1989 as part of the Long Term Ecological Research project, a National Science Foundation funded network of sites devoted to the maintenance of similar experiments. Population density of ladybeetles is monitored in 10 plant communities weekly over the growing season using yellow sticky card glue traps: detailed sampling methodology is available in previous work (Bahlai et al. 2013, Bahlai, Colunga-Garcia, et al. 2015, Bahlai, vander Werf, et al. 2015). The invasion process observed for *H. axyridis* initially inspired the need for a regime shift detector (Bahlai, vander Werf, et al. 2015); herein, we revisit these data with a refined tool and two years of additional observations (1994-2015).

Raw sampling data documenting the captures of adult *H. axyridis* at each sampling point, during each sampling week were extracted from the database, and these raw data are available here: <https://github.com/cbahlai/monarch_regime/blob/master/casestudydata/kbs_harmonia94-15.csv>. Dates were converted to day-of-year format, and then data were culled at day-of-year 240 to minimize the effect of variation in sampling period between sampling years (Bahlai, vander Werf, et al. 2015). From these data, the average number of *H. axyridis* adults captured per trap, across all traps deployed within a sampling year were computed, providing a data frame in the format required by the RSdetector function. The RSdetector function was then used on this 1994-2015 data frame, and then the analysis was repeated using data culled at 1994-2013, to allow for direct comparison with the outputs from the original conception of the regime shift detector model.

Monarch butterflies in Mexican overwintering grounds

North American Monarch butterflies are migratory, with the majority of their population overwintering in the Mexican highlands in large aggregations on Oyamel fir trees within the transvolcanic mountains in the central region of the country (Urquhart and Urquhart 1978, Wassenaar and Hobson 1998). Monarchs are highly dispersed over their breeding season, occupying landscapes throughout south and central North America (Flockhart et al. 2017), so overwintering population measures provide a convenient integration of total population size in a large portion of the species’ North American range. Since the 1994-1995 overwintering season, various groups have monitored the total area occupied by overwintering monarch colonies each season as a proxy for raw population counts, to minimize disturbance to the butterfly aggregations themselves. We used data documenting observations of area occupied from the winter of 1995 to the winter of 2017, compiled from these surveys by MonarchWatch.org. As these data are proprietary, they are not reproduced here, but are available directly from MonarchWatch (Lovett 2017).

Monarch overwintering population data were subjected to the RSdetector function. Because the time series data also suggested the possibility of a simple linear decline in K (i.e. a linear decline in the mean population) data were also modelled this way, and the information criteria produced from this simpler model was used to compare to the performance of the RS detector.

**Results**

*Simulations*

Simulations were conducted by modifying one critical parameter at a time from a base scenario to determine how modifying each parameter affected the findings of the RSdetector model. When varied sampling error was simulated as ‘noise’ (Fig. noise\_sim), the script’s ability to detect starting conditions generally dropped as percent noise increased, with the exception of no-break scenarios, which were generally correctly identified at a rate of approximately 60%, regardless of simulated sampling error (Fig. noise\_sim A). Outcomes involving the script finding extra breaks were most common in scenarios initiated with only one break (Fig. noise\_sim B), while outcomes where one break was missed by the script only occurred in scenarios initiated with three breaks, increasing with sampling error, and then plateauing at about 20% of outcomes above 30% sampling error (Fig. noise\_sim C). Outcomes identifying the correct number of breaks but misidentifying one break’s location peaked at around 20% sampling error in scenarios initiated with three break points, and at approximately 40% sampling error in scenarios initiated with two break points (Fig. noise\_sim D), and total failure to identify initial conditions generally increased with sampling error, with the exception of scenarios intiated with no break points (Fig. noise\_sim E).

Modifying the length of time series that a scenario was initiated with affected the ability of the script to identify the starting conditions (Fig. Nyears), with model performance decreasing slightly with length of time series (Fig. Nyears A). In general, increasing time series length increased the probability that the script would correctly identify the break points from the initial conditions, but also ‘find’ an additional break points (Fig. Nyears B) or find a break in a scenario that was not initiated with any breaks (Fig. Nyears E). Other erroneous results were rare (Figs. Nyears C, D).

The effect of modifying regime shift size on the script’s ability to detect conditions with which the scenarios were initiated was examined by modifying the % change in r and K at the given break point combination (Figs. changeK, changeR). The script was best able to identify initial conditions when the value for K was shifted by approximately 40% (Fig. changeK A) with extra breaks more frequently detected in scenarios initiated with larger changes of K at break points (Fig. changeK B). Complete failure to identify break points was most common in scenarios with small shifts in K (Fig. changeK E); missed breaks occurred rarely in 3 break scenarios regardless of the shift in K (Fig. changeK C) and misidentified breaks occurred occasionally in scenarios with 2 or 3 breaks and very large or very small shifts in K (Fig. changeK D.) The efficiency of the script responded differently to modifications of the size of shifts in r: instead of an intermediate optimum shift as observed for K, smaller shifts involving changes in r were most easily detected by the script (Fig. changeR A). The script was more likely to erroneously find additional breaks, miss breaks, or misidentify breaks as shifts in r increased (Fig. changeR B, C,D). Complete failure to identify correct break combinations increased slightly with increases in r for scenarios initiated with 1, 2, or 3 break points, but error rates remained constant regardless of shift in r in the zero-break scenarios (Fig. changeR E)

To aide in the interpretation of regime shift detector script outputs in a situation where the conditions under which the data were produced are unknown (i.e. any ‘real’ population data) we also examined the scenarios where sampling error was varied in the converse way- by the proportion of input scenarios resulting in a given observed outcome (Fig. obs\_outcomes). When the regime shift detector script indicated that it had found no breaks in the data, this result generally reflected input scenarios with more than 80% accuracy when sampling error was below 50% (Fig. obs\_outcomes A). When the script identified scenarios with one two breaks, sampling error affected the accuracy of outcomes more negatively, with accuracy dropping to approximately 60% at levels of sampling error approaching 25% (Figs. obs\_outcomes B, C). A similar pattern was observed for scenarios identified to have three breaks, however, accuracy was generally higher with this output, with >80% accuracy observed even at a sampling error rate of 25%.

*Case study- Harmonia axyridis*

The sampling error (in the form of standard error of the mean) for population samples of *H. axyridis* was estimated at about 6% from the raw data.

When the regime shift detector script was run using the *H. axyridis* population data from 1994-2013, as was used in the previous study, the script produced identical results to the previous implementation (Bahlai, vander Werf, et al. 2015) when break point combinations were ranked by AICc and AIC. Only one break point combination was identified by the regime shift detector: no break point combinations with equivalent fit were identified. Two break points- one occurring after 2000, and one occurring after 2005 were observed in this ‘best’ break point combination model (Fig. harmonia\_fit A). In this truncated data, the shift from ‘phase A’ to ‘phase B’ at the year 2000 was characterized by substantial increases in the fitted values for K and r (Table 1, ‘original’ data structure), followed by a return to parameter estimates nearly identical to those observed for ‘phase A’ in the post-2005 shift from ‘phase B’ to ‘phase C’ (Table 1 ‘original’ data structure, Fig. harmonia\_fit B).  
  
However, when the regime shift detector was applied to updated *H. axyridis* population data, which included two additional sampling years, the results were strikingly different. The two new observations, but 2015 observation, in particular, broke from the trend in dynamics observed in 2006 and after (Fig. harmonia\_fit A, data to the left of black vertical dashed line), and the regime shift detector script only located the post-2000 break in these data (Table 1, ‘updated’ data structure, Fig. harmonia\_fit B). In this case, the resultant regression parameters estimated for the period combining all apparent phases from 2001 on were intermediate in value, with greater standard error, than those estimated for phases B and C from the fits resulting from the shorter time series (Table 1, ‘updated’ data structure , Fig. harmonia fit B, dashed curve).

**Table 1:** Regression parameters Ricker model fits for each phase between break points resulting from fitting population data of *Harmonia axyridis* from Kellogg Biological Station, 1994-2015. Regression parameters r represent the per capita yearly intrinsic rate of increase and K the carrying capacity, based on population numbers expressed as average number of adult *H. axyridis* captured per trap, per year. Analyses were performed on a subset of the data, from 1994-2013 to compare to previous use of this approach (Bahlai, vander Werf, et al. 2015), and then again on the updated data including two additional sampling years. Note that the information criteria cannot be compared between the two data structures here, as they represent two different sets of independent variables; these criteria represent the ‘best’ of those that were used to rank competing break point combinations tested within the given data structures. The ‘Phase’ column gives a shorthand for referring to the data subsetting structure under the most complex scenario represented here.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Data**  **structure** | **AIC** | **AICc** | **Phase** | **Years in subset** | **r ( ± SE)** | **K ( ± SE)** |
| Original, to 2013 | -13.5 | -7.5 | A | 1994-2000 | 1.3 ± 0.3 | 0.33 ± 0.03 |
|  |  |  | B | 2001-2005 | 2.2 ± 0.2 | 0.46 ± 0.02 |
|  |  |  | C | 2006-2013 | 1.5 ± 0.2 | 0.29 ± 0.02 |
|  |  |  |  |  |  |  |
| Complete, to 2015 | -36.8 | -16.8 | A | 1994-2000 | 1.3 ± 0.3 | 0.33 ± 0.03 |
|  |  |  | B+C | 2001-2015 | 1.6 ± 0.4 | 0.43 ± 0.05 |
|  |  |  |  |  |  |  |

*Case study- Monarch butterflies*

Sampling error could not be estimated for this population measure as it is only reported as a single value- total area occupied by overwintering monarchs.

The regime shift detector script found three different break point combinations that were deemed to have equivalent performance by their respective AICcs, two models with a single break after 2003 and 2006 respectively, and a third with breaks at 2003 and 2008. However, when ranked by AIC, the two-break model substantially out-ranked both of the single-break models. Similarly, the population dynamic was modelled as a linear decline in carrying capacity K produced a fit that was ranked best of all scenarios tested by AICc, but second best after the two break point model by AIC (Table 2). The break point combination as ranked by AIC, the two break model is represented graphically by the solid lines in Fig. monarch\_fit (A, B) but the fit of the one break model is also given by the dashed line in Fig. monarch\_fit B.

**Table 2.** Model performance of top-ranked models of differing structures fit to population data documenting the area occupied by overwintering Monarch butterflies in their winter habitat in the Mexico, 1995-2016. One break and two break models are for best break point combinations selected by regime shift detector script, while ‘linear K’ model assumes a linear decline of carrying capacity K and a single constant intrinsic rate of increase r. The ‘Phase’ column gives a shorthand for referring to the data subsetting structure under the most complex scenario represented here.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model structure** | **AIC** | **AICc** | **Phase** | **Years in subset** | **r ( ± SE)** | **K ( ± SE)** |
| One break | 114.6 | 120.2 | A | 1995-2003 | 1.0 ± 0.5 | 10.1 ± 1.9 |
|  |  |  | B+C | 2001-2015 | 0.8 ± 0.3 | 4.1 ± 0.7 |
|  |  |  |  |  |  |  |
| Two break | 106.9 | 121.9 | A | 1995-2003 | 1.0 ± 0.5 | 10.1 ± 1.9 |
|  |  |  | B | 2004-2008 | 1.6 ± 0.2 | 5.6 ± 0.3 |
|  |  |  | C | 2009-2016 | 1.2 ± 0.4 | 2.8 ± 0.5 |
|  |  |  |  |  |  |  |
| Linear K | 112.9 | 118.5 | A+B+C | 1995-2016 | 1.3 ± 0.3 | - |
|  |  |  |  |  |  |  |

**Discussion**

*Regime shift model structure*

The implementation of the model described here used the Ricker function because it presented an ideal compromise of simplicity and fit for the populations we wished to model. However, the method presented here could easily be adapted to population processes better described by other models, and incorporating other dependent variables which may be available (for example, if a population had a known response to temperature or another environmental variable). Similarly, this approach is not necessarily limited to population processes: a regime shift detector script could be developed to identify changes in any ecological dynamic with a well-defined internal rule governing its fluctuations. The sensitivity and precision of the approach could also be adjusted in these future implementations by adjustment of decision rules regarding selecting models of equivalent and best fit.

AICc was the information criterion used to rank break-point combination models, with all models ranked within two units of the lowest AICc considered to have equivalent performance, however, just a single ‘best’ model from the set of equivalent models was used for comparison in the simulations. The decision to only include one ‘best’ fit represented a compromise between accuracy, simplicity of script outputs, and computational intensity when running many simulation iterations. Overall script performance would likely have a higher rate of detecting all initial conditions if the set of all equivalently fitting models, instead of just the top-ranked, had been considered when comparing the performance of the script to the input conditions. In response to this observation, we developed the ‘modelspecification’ function so that a user may manually produce regression statistics associated with similarly ranked fits and interpret those values in the context of the known biology of the species under evaluation.

With regards to selection of information criteria, AICc was used for decision-making in the regime shift detector script rather than AIC because it allowed for a more conservative selection of break-point combinations while minimizing overfitting in higher sampling error scenarios- essentially by down weighting the selection criterion for models with many break points. However, this more conservative approach negatively affected the script’s ability to detect higher break frequency in low sampling error scenarios, particularly for one and two break input scenarios, because the penalty term for increasing the complexity of the model dramatically increases with AICc. Thus, if it is reasonable to assume that the population data being subjected to the regime shift detector script has a low associated sampling error, a user may wish to use less conservative information criteria (i.e. AIC) to rank break point combination models.

Because the model uses a single datum to represent the population in a given year, the model had to be constrained to avoid over-fitting to short time series. Unfortunately, this limitation means that shifts in dynamic regime occurring less than four time steps apart will not be detected by this modelling approach. In populations undergoing rapid change in their environments or internal dynamics, thus, the results of the script should be interpreted with caution, because a single-variable discrete time step model like the Ricker may not fully leverage available information. In these cases, using a model that allows, for example, within season dynamics to be measured may be more useful.

Regardless of model used to form the basis of the regime shift detector script, it is important that the model’s fitting function is set with some understanding of the data’s structure to prevent fitting or convergence issues. For example in our case, the rickerfit function was set to have a starting value of r at 1.5. For populations with dynamics that are expected to deviate from this value dramatically, setting this value to one closer to the expected value will aide in model convergence.

*Simulations*

Using the decision rules as set, simulations were performed to understand how changing various inputs affected the likelihood of the regime shift detector script identifying the conditions under which the data were produced. Simulations indicated that the performance of the regime shift detector script declined rapidly with increasing levels of sampling error (Fig. noise\_sim), a behavior that is, in general, expected of any statistical tool. Nevertheless, whenever possible, the sampling error of the data subjected to the script should be quantified to help evaluate the script’s results in the context of variation within the data due to sampling error. The error rate in detecting initial conditions varies with output, but in low-sampling-error scenarios, an output of zero or three or more break points by the script is generally approaching 90% accuracy, while outputs of one or two breaks have a lower rate of accurately detecting input conditions, at just under 80%.

Other input conditions also effected the performance of the regime shift detector script. The effect of changing the magnitude of the shift was dependent on which parameter was changed and by how much (Figs. change, changeR). Although larger shifts in regression parameters would, intuitively, lead to a higher likelihood of detection, these larger shifts would also be more likely to induce chaotic dynamics in the years immediately following the shift, potentially making the timing of shifts more difficult to pinpoint. Similarly, longer time series yielded regime shift detector script results that were more error prone (Fig, Nyears A), likely because, firstly, there were simply more possible break-point combinations for the model to select from, and secondly, because the penalty for increasing parameterization (AICc) would decrease as sample sizes grew, leading to increasing likelihood of identifying extra breaks (Fig. Nyears B).

*Case studies*

Our case studies represent two different biological processes- invasion and a population decline, in two very well studied insect species.

**Conclusions**

Information criteria used and decision rules for cutoff have a dramatic impact on the results- and should be considered critically before drawing any conclusions by the use of this tool.

Critical understanding of biology needed to interpret statistical results

**References**

**Bahlai, C. A., M. Colunga-Garcia, S. H. Gage, and D. A. Landis**. **2013**. Long term functional dynamics of an aphidophagous coccinellid community are unchanged in response to repeated invasion. PLoS One. 8: e83407.

**Bahlai, C. A., W. vander Werf, M. O’Neal, L. Hemerik, and D. A. Landis**. **2015**. Shifts in dynamic regime of an invasive lady beetle are linked to the invasion and insecticidal management of its prey. Ecological Applications.

**Bahlai, C., M. Colunga-Garcia, S. Gage, and D. Landis**. **2015**. The role of exotic ladybeetles in the decline of native ladybeetle populations: evidence from long-term monitoring. Biol Invasions. 17: 1005–1024.

**Burnham, K. P., and D. R. Anderson**. **2002**. Model selection and multimodal inference: a practical information-theoretic approach, 2nd ed. Springer Science + Business Media, LLC, New York.

**Elzhov, T. V., K. M. Mullen, A.-N. Spiess, and B. Bolker**. **2016**. minpack. lm: R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in MINPACK, R package version 1.2-1.

**Flockhart, D. T. T., L. P. Brower, M. I. Ramirez, K. A. Hobson, L. I. Wassenaar, S. Altizer, and D. R. Norris**. **2017**. Regional climate on the breeding grounds predicts variation in the natal origin of monarch butterflies overwintering in Mexico over 38 years. Glob Change Biol. 23: 2565–2576.

**Lovett, J.** **2017**. Monarch Population Status.

**R Development Core Team**. **2017**. R: A Language and Environment for Statistical Computing 3.3.3. R Foundation for Statistical Computing.

**RStudio Team**. **2015**. RStudio: Integrated Development for R (RStudio, Inc., Boston, MA, 2015). URL: https://www. rstudio. com/products/rstudio.

**Turchin, P.** **2003**. Complex population dynamics: a theoretical/empirical synthesis. Princeton University Press.

**Urquhart, F. A., and N. R. Urquhart**. **1978**. Autumnal migration routes of the eastern population of the monarch butterfly (Danaus p. plexippus L.; Danaidae; Lepidoptera) in North America to the overwintering site in the Neovolcanic Plateau of Mexico. Can. J. Zool. 56: 1759–1764.

**Wassenaar, L. I., and K. A. Hobson**. **1998**. Natal origins of migratory monarch butterflies at wintering colonies in Mexico: New isotopic evidence. Proceedings of the National Academy of Sciences. 95: 15436–15439.