Population models and forecasts for a recently delisted riparian forb *Oenothera coloradensis*

*Manuscript draft prepared for Biological Conservation or equivalent journal by Tyson Wepprich*

# Introduction

A quantitative population viability analysis (PVA) links measured vital rates with environmental covariates to model response functions that can forecast population trends and inform management or regulatory action (Morris and Doak 2002). When data are available, PVA provides one form of “objective, measurable criteria” (ESA) that may guide recovery plans of listed species (Boor 2014). However, others advocate not to use a PVA in isolation as it may limit the scope of recovery or promote overconfidence in modeled extinction thresholds based in part on the analyst’s subjective choices (Wolf et al. 2015). In place of calculating extinction risk, we provide a focused PVA on two aspects of a rare species’ population regulation that may impact its post-delisting monitoring plans: accounting for potential effects of climate change on population growth and planning how many years of monitoring required to detect trends in highly variable counts.

We undertake this study in the context of regulatory decisions for the Colorado butterfly plant (*Oenothera coloradensis* (Rydberg) W.L. Wagner & Hoch; syn. *Gaura neomexicana* Woot. ssp. *coloradensis* (Rydb.) Raven & Gregory). Previously, *O. coloradensis* was listed as Threatened under the Endangered Species Act (USFWS 2000). It was recently delisted due to discovery of additional populations since its listing and due to observations of the resiliency of the populations in rebounding from low numbers (USFWS 2019). Long-term viability under climate change was considered in its biological assessment but ruled unthreatening due to the uncertainty of how climate impacts the species as a whole. Resilience to environmental stochasticity, and presumably climate change, comes from prolonged dormancy in the seedbank and in the ability of rosettes to delay flowering until favorable years (USFWS 2019). We assess the response of population growth to climate by testing which weather variables are important during the multiyear lifecycle of individual plants and projecting population growth into the future under scenarios that show how climate may change viability.

After delisting, the Recovery Team will monitor trends for at least five years at a sample of populations spanning the species’ distribution and compare the range, mean, and median of population counts with historical data to assess the recovery of *O. coloradensis* (USFWS 2017). Monitoring will note new environmental stressors, such as rare herbivory events that have impacted populations previously (Heidel et al. 2011). After trends are assessed, the species will be considered secure without ESA protection if no monitored population has a decline of more than 50% (USFWS 2017). The significance of population trends is difficult to judge for *O. coloradensis* because only the conspicuous flowering life stage is counted and annual fluctuations in flowering are high (Figure 1). Given these limitations, our PVA will test how many years it may take to detect a significant change in a population with flower monitoring to inform the confidence the Recovery Team may place in their assessment after five years.

Chart, line chart

Description automatically generated

Figure 1: Total flowering census by year for three populations on Warren Air Force Base with significant ESA decisions noted in text boxes. Blue lines are the nonlinear fits from generalized additive models (knots = 10) with automatic cross-validation to select the degree of nonlinearity.

# Methods

## Study system

Long-term monitoring along three creeks on F.E. Warren Air Force Base (WAFB) near Cheyenne, Wyoming provides a population census of the flowering life stage each year. With 32 years of continuous monitoring, we can analyze annual demographic responses to a wide range of environmental variability to predict how the populations may respond to weather variables. Differences between populations segments in a diversity of microclimates may increase the resiliency of the species on WAFB, so we analyze both the populations at each of three creeks as well as further subdivisions into population segments grouped by local environmental differences.

## Population models

We used linear mixed-effects models to estimate annual population growth rates log(*Nt*/*Nt-1*) with the *lme4* package in R 4.0.3 (Bates et al. 2015, R Core Team 2020). Population growth rates were assumed to result from the population-specific intrinsic rate (intercept), density-dependence based on the previous year’s count (fixed effect), environmental conditions (fixed effect of weather covariates), and annual stochasticity shared across segments. The form of density dependent population regulation may either increase or decrease population resiliency under future conditions (Jaatinen et al. 2021). We compared two models and selected Gompertz density-dependence over Ricker density-dependence. We fit models with density-dependence (natural logarithm of previous year’s flower count) and scaled weather covariates identified as important as described below. Each year and population segment varied with random intercepts to account for non-independent observations.

## Environmental covariates

We used monthly temperature, precipitation, and snowfall data from the Cheyenne Municipal Airport weather station (481675) from the Western Region Climate Center (wrcc.dri.edu). We used the monthly Palmer Drought Severity Index (PDSI) from the Lower Platte Climatological division in Wyoming from the NOAA Climate Divisional Database (Vose et al. 2014). Stream discharge data from Crow Creek downstream of the study area was available for 1994-2017 from the USGS site 06755960 (waterdata.usgs.gov/wy/nwis).

Detailed knowledge of the species life history has pointed to the potential impact of temperature, precipitation, and stream discharge on population growth. However, the specific life stages affected may alter the time lags and time windows for when these environmental conditions are important. For example, springtime germination may be a key vital rate affected by weather that would not be observed in flowering counts for two to three years later. Some conditions may have a cumulative effect of longer than one season, such as prolonged drought. We used the *climwin* R package to search for the optimal model (linear or quadratic) and time windows with a minimum of 3 months duration of up to three years before the August when flower monitoring occurs (Bailey and Pol 2016). The base model for this search included density-dependence and random intercepts for each site as predictors of population growth rates. Model outputs were used to select covariates to use in multiple regression models of population growth rates. Important variables were selected primarily based on model selection (weight based on AICc) and visualizing potential time windows of impact beyond the top model.

## Forecasts

We used the model of population growth rates to project flowering counts in 2021-2040 using the 2020 observed counts as the start. Density-dependence, the selected environmental covariates, site intercepts, and variation from annual intercepts and residual error drawn from Normal distributions determined the annual population growth rate (*r*). The simulated flower count in the following year was drawn from a Poisson distribution with *λ = Nt-1er*. This process was repeated for each of 100 simulations for each scenario.

Nine scenarios were projected with climate covariates. To simply incorporate the temporal correlation between years in weather, we used the observed weather in previous decades to project population counts forward. These scenarios used every combination of the 1990-1999, 2000-2009, and 2010-2019 decadal weather for nine 20-year forecasts. The 2000s were a decade with considerable drought, and we expected the 20-year scenario with the 2000s repeated twice would have the most pessimistic outcome.

We forecasted populations at the segment level, but aggregated results to the three creek populations and total WAFB population for the results. We compare the median and range of observed counts over 32 years to those of the forecasted scenarios. To simplify the results, we present three of the nine scenarios that span the range of outcomes: 1990s, 2000s, and 2010s weather repeated for 2021-2040.

## Trend detection

We estimated retrospectively when shifts in population growth rate and carrying capacity had occurred across the long-term monitoring data. We used a model comparison algorithm that fits population time-series with all possible breakpoints and selects the most supported shifts in population demographic rates (Bahlai and Zipkin 2020). We ran this algorithm on the full length of monitoring for each creek, and then shortened the dataset systematically to determine how many years of monitoring would have been needed to detect the significant breakpoints in the monitoring trends.

# Results and discussion

## Environmental covariates

The best time window and time lag for covariates and the form of the relationship (linear or quadratic) varied. We interpret the results of the top models in light of the lifecycle, but the associations in the best-fit models are unlikely to be the sole time window in which the environment affects population growth rates.

Stream discharge for Crow Creek, assumed to correlate with conditions across all segments, has the most supported time window (weight 11%) 23-7 months before the August census (red region in Figure 2A). The quadratic model suggests that both low and high flow in the two years prior to flowering counts decreases the annual population growth rates, while growth rates are maximized at intermediate flow (Figure 2B).

Chart, scatter chart

Description automatically generated

Figure 2: The timing of stream discharge associated with models of annual population growth rates has its best fits in the red region (A) and shows maximum growth rates with intermediate discharge within the time window (B). Data points show the growth rates of each segment and year, with the blue color lightening as years are more recent.

A negative linear model of temperature’s effect was supported over a quadratic model (Figure 3B). The best model (69% weight) includes fall-winter of two years before census (23-19 months ago)(Figure 3A). The small region of yellow in the lower left region shows that temperature in the months immediately preceding the August census also impact the estimates of population growth rates, with higher temperatures lowering the growth rate (not shown).

Chart, scatter chart

Description automatically generated

Figure 3: The timing of temperature associated with models of annual population growth rates has its best fits in the red region (A) and shows decreasing growth rates with higher temperatures within the time window (B). Data points show the growth rates of each segment and year, with the blue color lightening as years are more recent.

For precipitation, there is strong support (69% weight) for a longer window starting two years before (24-2 months ago) (Figure 4A). The quadratic model is strongly supported over linear (ΔAICc = -31.0) with higher population growth rates at intermediate values (Figure 4B).

Chart, scatter chart

Description automatically generated

Figure 4: The timing of precipitation associated with models of annual population growth rates has its best fits in the red region (A) and shows higher growth rates at intermediate values within the time window (B). Data points show the growth rates of each segment and year, with the blue color lightening as years are more recent.

Snowfall had a surprising window with the highest impact. The best models start 4-6 months before census (Feb-April) and end within 1-3 months of the census (Figure 5A). Since the endpoint is during the summer with very little snow, the takeaway should be that late season snow has a larger impact that early winter snow. This is new snowfall, and an analysis of accumulated snowfall and how long it persists in the spring could be important as well. The quadratic model is strongly supported over linear (ΔAICc = -31.7) with higher population growth rates at intermediate values (Figure 5B).

Chart, scatter chart

Description automatically generated

Figure 5: The timing of snowfall associated with models of annual population growth rates has its best fits in the red region (A) and shows higher growth rates at intermediate values within the time window (B). Data points show the growth rates of each segment and year, with the blue color lightening as years are more recent.

Water availability (PDSI) had no clear top model (best weight 10%) and the top linear models performed slightly better than top quadratic models (ΔAICc = -5.02). The linear model had the strongest support in the lower left corner (Figure 6A, 7-4 month before census) which suggests greater water availability in the spring has a slight positive affect on population growth rates (Figure 6B). Regions of support in the upper left and upper right of figure 6A suggest that cumulative water availability over the last three years is important, as is water availability two years previous.

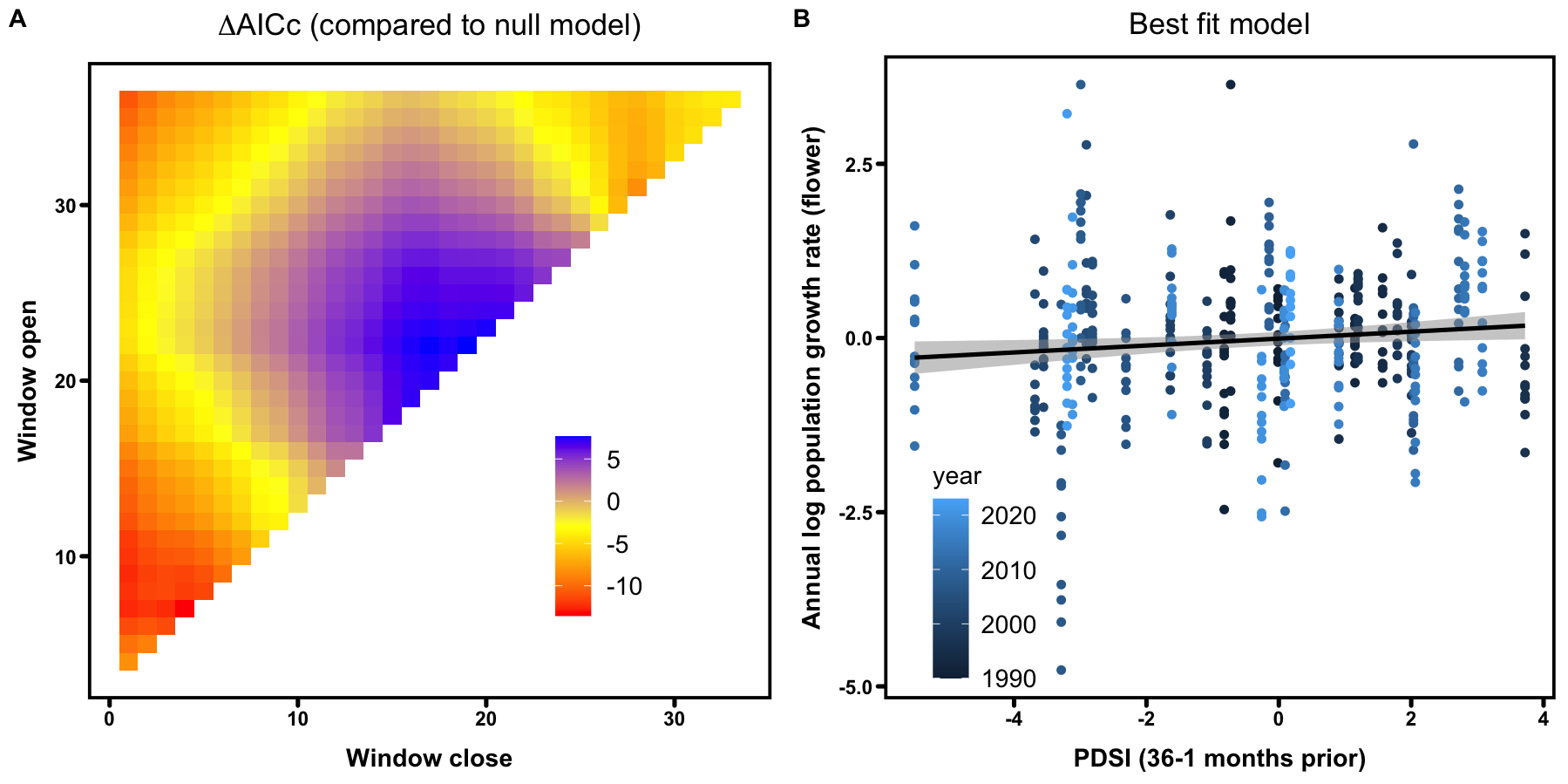


Figure 6: The timing of water availability (PDSI) associated with models of annual population growth rates has its best fits in the red region (A) and shows increasing growth rates with higher water availability within the time window (B). Data points show the growth rates of each segment and year, with the blue color lightening as years are more recent.

After the best time windows were selected for environmental covariates, we tested for correlations between them. Water availability (PDSI) (7-4 months prior) and stream discharge (23-7 months prior) were strongly correlated (Pearsons’s *r* = 0.72), so discharge was removed as it had years of missing data. Using all years for the other variables, paired correlation coefficients were all less than or equal to 0.31.

## Population models

Annual population growth rates show strong negative density-dependence, with lower growth following years with high counts. All environmental covariates identified in the above analysis and their quadratic terms (if selected) were added to the density-dependence population model (Table 1). Variables were removed if not significant (p > 0.05). The quadratic effect of snow and the linear effect of PDSI were not significant in multiple regression and were removed for the simulations. Even though these variables may improve model fit when tested in isolation in the time window analysis above, when included with other variables they did not improve the model fit. The final model, used for population simulations, included temperature, precipitation (linear and quadratic effects), and snowfall as important drivers of population growth rates.

Table 1: Two mixed-effects models of annual population growth rates with density-dependence only or including the selected climate covariates.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Density-dependence** | | | **\_+ climate covariates** | | |
| *Predictors* | *Estimates* | *CI* | *p* | *Estimates* | *CI* | *p* |
| (Intercept) | 1.99 | 1.38 – 2.59 | **<0.001** | 1.35 | 0.67 – 2.04 | **<0.001** |
| Log(Nt-1) | -0.40 | -0.48 – -0.32 | **<0.001** | -0.40 | -0.47 – -0.32 | **<0.001** |
| Scaled temperature  (23-19 months prior) |  |  |  | -0.27 | -0.42 – -0.12 | **<0.001** |
| Scaled precipitation  (24-2 months prior) |  |  |  | -0.00 | -0.16 – 0.15 | 0.964 |
| Scaled precipitation2 |  |  |  | -7.05 | -10.00 – -4.10 | **<0.001** |
| Scaled snowfall  (4-1 months prior) |  |  |  | 0.74 | 0.28 – 1.21 | **0.002** |
| **Random Effects** | | | | | | |
| σ2 | 0.52 | | | 0.52 | | |
| Std Dev | 0.36 year | | | 0.13 year | | |
|  | 0.54 segment | | | 0.52 segment | | |
| N | 13 segment | | | 13 segment | | |
|  | 31 year | | | 31 year | | |
| Observations | 403 | | | 403 | | |
| Marginal R2 / Conditional R2 | 0.304 / 0.744 | | | 0.425 / 0.745 | | |

## Forecasts

When density-dependence is strong as in our population models, annual population growth rates increase at lower counts and generally prevent the population from declining to extinction. Unlike previous PVA of *O. coloradensis* where segments and creeks were ranked by extinction probability, we compare the range and median of forecasted counts to emulate the assessments to be performed under the post-delisting monitoring plan (USFWS 2017).

Under a density-dependent model without environmental covariates, the WAFB creek populations and total population are expected to increase from the 2020 numbers and fluctuate around the statis carrying capacity estimated in this model. Simulations vary based on annual stochasticity and sampling, so the 25th-75th percentile of counts vary substantially, with expectations of the total WAFB count between about 4,000 and 11,000 flowers counted each year (Figure 7).

With climate scenarios, population simulations diverge depending on the weather experienced. Scenarios including the 1990s weather had higher average population level 2-3 times that of those with 2000s or 2010s weather (Figure 8). Annual variation was more correlated across the 100 simulations and the median value fluctuating more widely than in the model without climate included. The simulations demonstrate the quick population growth that can occur with multiple years of positive effects of weather with the example of the 1990s scenario that nearly doubles within 3 years in 2028-2030 and 2038-2040, corresponding to the observed weather in 1997-1999 (Figure 8). One surprise is that weather from the 2010s is not much better than the 2000s that included a major drought. These scenarios show that population size, while fluctuating between years, will average much lower with weather similar to that of the last two decades compared to the higher growth years in the 1990s.

We show the simulated values in three scenarios to the whole time-series of observed counts since 1988 (Figure 9). Crow Creek is unlikely to rebound to its higher populations levels unless weather is extremely favorable. Diamond Creek and Unnamed Creek populations may average lower than the observations in 2010 with weather similar to either the 2000s or the 2010s. The ranges of the observed counts (1988-2020) and the simulated counts overlap considerably, which will make any comparison during post-delisting assessments difficult (Figures 10 and 11). The median value of the simulations was generally lower than the median of the observed counts. Approximately, in the event of two decades of weather similar to the 1990s the population size may double. However, the forecasted median is lower than the long-term monitoring median for either the 2000s (60% of observed) or 2010s (85% of observed) weather scenarios.

Climate change that increases temperature and decreases snowfall would likely lower population growth rates further. The impact of precipitation, as a quadratic relationship, would predict decreased population growth rates at either extreme. Density-dependence is a key factor that increase population resilience after years of low growth and would need to be investigated further to know if its impact will change with climate.

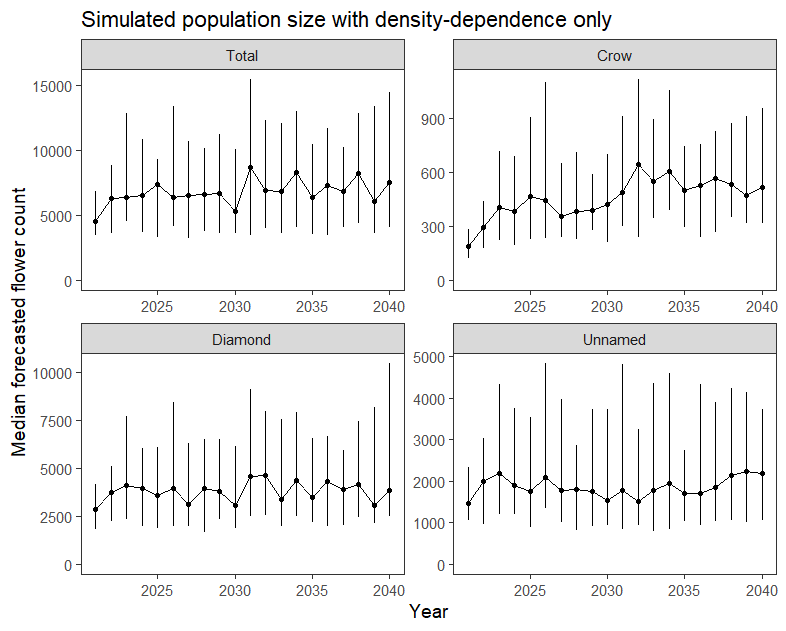


Figure 7: Simulated variation in counts under a model with density-dependence and annual stochasticity, but without weather covariates (Table 1, left model). Data points show the median simulated value with 25th-75th percentile error bars.

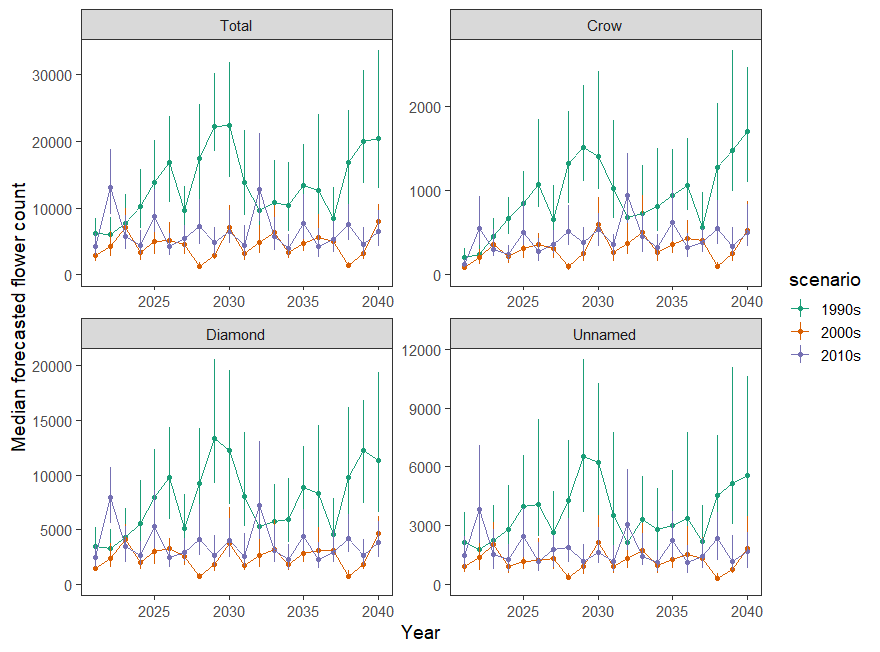


Figure 8: Simulated variation in counts under a model with density-dependence, annual stochasticity, and weather covariates (Table 1, right model). Data points show the median simulated value with 25th-75th percentile error bars. Scenarios included weather covariates from the listed decade, repeated twice over 20 years in order.



Figure 9: Simulated variation in counts under a model with density-dependence, annual stochasticity, and weather covariates (Table 1, right model). Data points show the median simulated value, with observed monitoring counts from 1988-2020. Scenarios included weather covariates from the listed decade, repeated twice over 20 years in order.

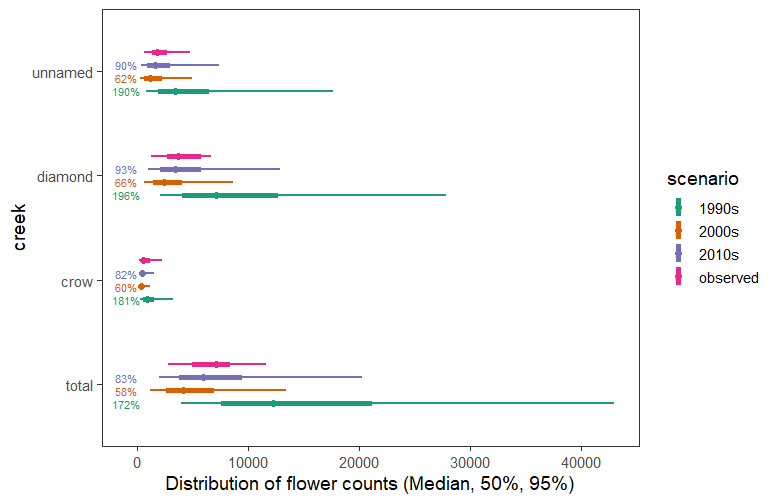


Figure 10: Comparison of median and range of observed (1988-2020) versus simulated (2021-2040) counts under a model with density-dependence, annual stochasticity, and weather covariates (Table 1, right model). Data points show the median simulated value, thick lines the 25th-75th percentiles, and thin lines the 5th-95th percentile. Scenarios included weather covariates from the listed decade, repeated twice over 20 years in order.

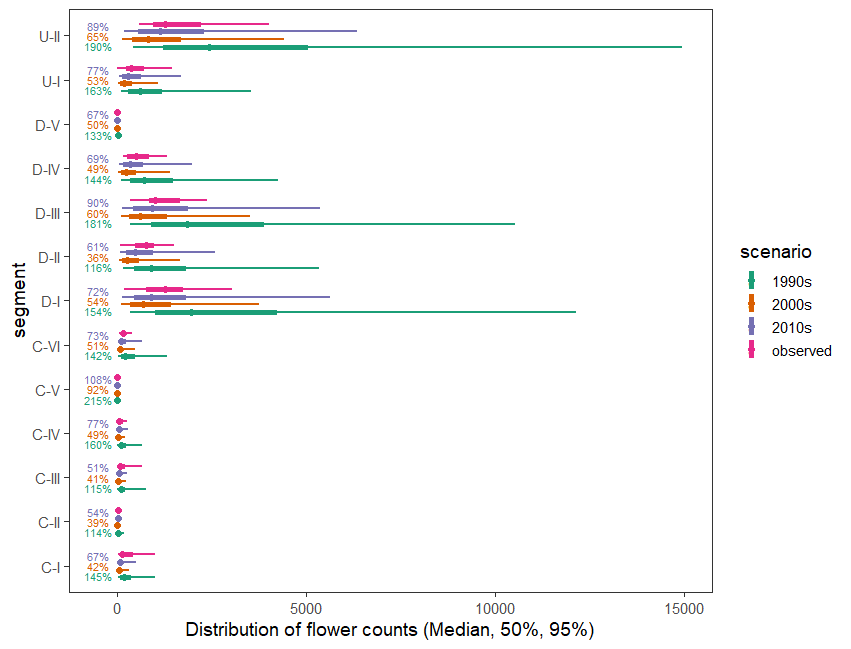


Figure 11: Comparison of median and range of observed (1988-2020) versus simulated (2021-2040) counts under a model with density-dependence, annual stochasticity, and weather covariates (Table 1, right model). Data points show the median simulated value, thick lines the 25th-75th percentiles, and thin lines the 5th-95th percentile. Scenarios included weather covariates from the listed decade, repeated twice over 20 years in order.

## Trend detection

Detecting trends within a five-year monitoring window, as proposed, is unlikely if the detection of past trends is an indicator. Automatic detection of changes in the monitoring time-series did not detect the significant breakpoint in the population trend until 4-8 years after the change had occurred.

The selected breakpoint years in the population trends change depending on the scale of population (total on WAFB, creek population, or smaller segment subdivisions). If summing the entire WAFB population together, no decline during the 2000-2006 drought would be detected (Figure 12). When split into the three creek populations, we see how the decline on Crow Creek was masked by the relative stability on Diamond and Unnamed Creeks.

The carrying capacity of the Crow Creek population has approximately halved twice during the 32-year monitoring period. Across these two trend breakpoints, carrying capacity was estimated at 1,585 in 1988-1997, 841 in 1998-2001, and 346 in 2002-2019 (Figure 12). The decline in carrying capacity would be a key goal for post-delisting monitoring to detect. The breakpoint detection algorithm was carried out with shorter time-series from 1988 to varying endpoints from 1995-2020 to see when changes were detected. The changes on Crow Creek were not detected until after 2005, at a lag of either 4 or 8 years depending on whether the breakpoint in 1997 or 2001 is the basis. Either 1997 or 2001 were considered as the best single breakpoint as the time-series length grew until the final identification of two distinct breakpoints in 1998 and 2001 solidified in 2016.

For an example of the algorithm performance with an increasing population, the Unnamed Creek carrying capacity increased from 1731 in 1988-1994 to 2503 in 1995-2019. The algorithm was carried out with time-series from 1988 to varying endpoints from 1995-2020 to see when changes were detected. Including a breakpoint for 1994 was not considered in the best models until after 2003. It was consistently included for all longer time-series. For a couple years of analysis, an additional breakpoint in 2012 was estimated as the start of another increase in carrying capacity but was no longer supported with more data.

We show the breakpoints estimated for each segment but did not retrospectively test when they could have been detected (Figure 13). This analysis and a breakpoint detection for the population simulations with climate models would give greater insight into how frequently trend detection algorithms would give accurate results.

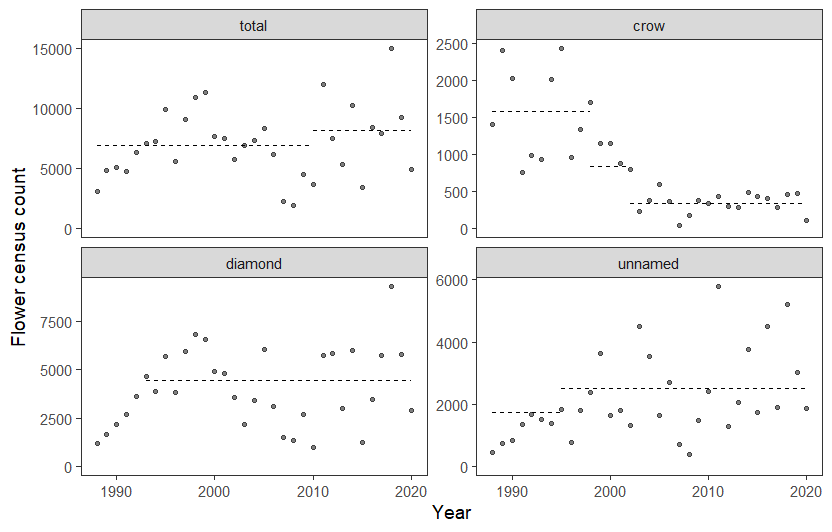


Figure 12: Breakpoints were estimated for the monitoring time-series, and the modeled carrying capacity for each distinct era is plotted as dashed lines. If lines are missing, the algorithm gave unrealistic results that we exclude here. Data points are flower counts observed during monitoring.

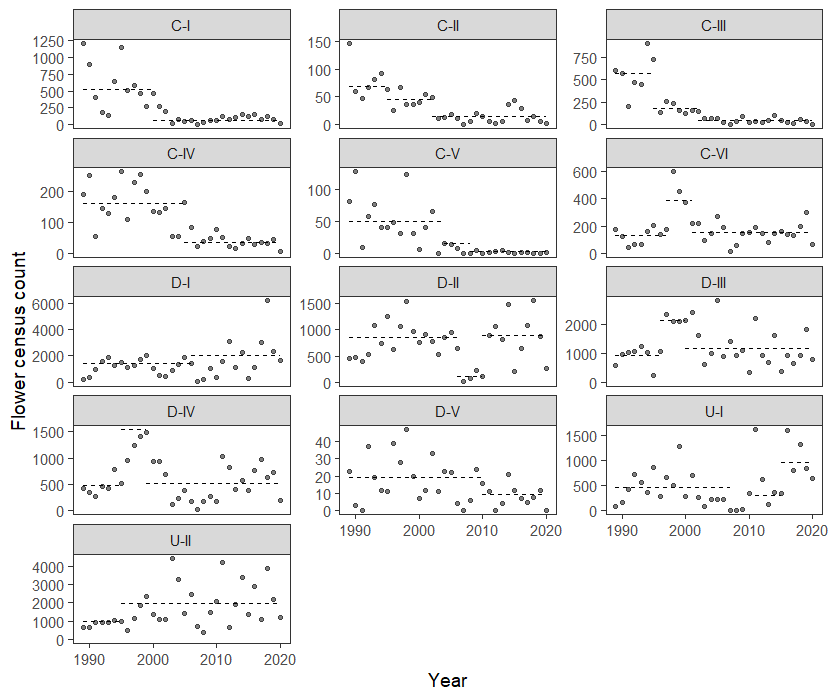


Figure 13: Breakpoints were estimated for the monitoring time-series, and the modeled carrying capacity for each distinct era is plotted as dashed lines. If lines are missing, the algorithm gave unrealistic results that we exclude here. Data points are flower counts observed during monitoring.

# References

Bahlai, C. A., and E. F. Zipkin. 2020. The Dynamic Shift Detector: An algorithm to identify changes in parameter values governing populations. PLOS Computational Biology 16:e1007542.

Bailey, L. D., and M. van de Pol. 2016. climwin: An R Toolbox for Climate Window Analysis. PLOS ONE 11:e0167980.

Bates, D., M. Mächler, B. Bolker, and S. Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software 67:1–48.

Boor, G. K. H. 2014. A Framework for Developing Objective and Measurable Recovery Criteria for Threatened and Endangered Species. Conservation Biology 28:33–43.

Heidel, B., L. Tronstad, and J. Handley. 2011. Flea Beetle ( *Altica* spp.) Herbivory on a Threatened Plant, F.E. Warren Air Force Base, Wyoming. Natural Areas Journal 31:283–287.

Jaatinen, K., M. Westerbom, A. Norkko, O. Mustonen, and D. N. Koons. 2021. Detrimental impacts of climate change may be exacerbated by density-dependent population regulation in blue mussels. Journal of Animal Ecology 90:562–573.

Morris, W. F., and D. F. Doak. 2002. Quantitative conservation biology. Sinauer, Sunderland, Massachusetts, USA.

R Core Team. 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

USFWS. 2017. Post-Delisting Monitoring Plan for Colorado butterfly plant (Oenothera coloradensis, formerly Gaura neomexicana subsp. coloradensis).

USFWS. 2019. Endangered and Threatened Wildlife and Plants; Removing Oenothera coloradensis (Colorado Butterfly Plant) From the Federal List of Endangered and Threatened Plants. Federal Register 84:59570–59588.

Vose, R. S., S. Applequist, M. Squires, I. Durre, M. J. Menne, C. Williams, and D. Arndt. 2014. NOAA’s Gridded Climate Divisional Dataset (CLIMDIV). NOAA National Climatic Data Center.

Wolf, S., B. Hartl, C. Carroll, M. C. Neel, and D. N. Greenwald. 2015. Beyond PVA: Why Recovery under the Endangered Species Act Is More than Population Viability. BioScience 65:200–207.