

Assessing the forecastability & forecast skill of models to predict sex ratios of *Valeriana edulis*

Yifei Zhang: Independent research and course

Sydney Petersen: Independent research and course/ REU

Mentors: Dr. Will Petry & Dr. Kailen Mooney

Assessing the forecastability & forecast skill of models to predict sex ratios of *Valeriana edulis*

Yifei Zhang & Sydney Petersen

Mentors: Dr. Will Petry & Dr. Kailen Mooney

ABSTRACT:

Anthropogenic actions are rapidly changing ecological systems, and the species specific responses to this change are expected to have great impacts on ecological organization, natural resource abundance, and human health. Scientists urgently need accurate forecasts to help inform social, political, and environmental decisions. However, from past attempts at ecological forecasts, it is unclear whether existing data and approaches are sufficient to meet this challenge. This study addresses this knowledge gap by forecasting the population sex ratio of the long-lived, dioecious perennial *Valeriana edulis* in which biased sex ratios increase the likelihood of extinction. We constructed forecasting models with varying biological complexity, trained each model with historic data, and assessed the forecast skill of each model in predicting the sex ratio observed in 16 populations in 2020. Despite the prior expectation that male frequency is increasing in populations over time, our linear trend forecasts did not produce a high forecast skill. In contrast, our climate-informed models produced forecasts with the highest skill, outperforming the null model for 14 out of 16 populations. Ultimately, these results indicate that sex ratios might be predictable given our current understanding of species biology. On a larger scale, this study presents an optimistic outlook for ecological forecasting to inform biodiversity policy and inform more sustainable management decisions under ongoing climate change.

INTRODUCTION:

Although climate change is affecting ecological systems in many ways, it is clear that the conditions and resources necessary for the survival of many species are being altered by anthropogenic actions (Clark 2001; Dietze 2018; Petry et al. 2016). Species are unlikely to respond identically to these changes: some may thrive under the new conditions imposed by a changing climate, others may go extinct (Field et al. 2014). The reshuffling of species in space and time and of their abundance is expected to have profound impacts on all levels of ecological organization. Regardless of a species' specific reaction to climate change, their responses will impact the dynamics of other species and whole ecosystems. Accurately predicting how different species will respond to climate change will be necessary for managers to maintain biodiversity, which may be an important component of preventing rapid fluctuations in the chemical and biological cycles that maintain life (Clark 2001). Climate change therefore demands that ecologists provide accurate, actionable predictions of the future states of ecological systems.

Historically, researchers have most often employed a retrospective approach to understand ecological systems, focusing on explanation of what *has* happened rather than forecasting what *will* happen. This pragmatic, hypothesis-driven method has emphasized the

importance of treatment-based experimentation and has been successful in unraveling complex ecological processes (Clark 2001). In principle, simply pivoting a retrospective approach to produce forecasts could be successful, but even a very precise understanding of ecological mechanisms does not guarantee that accurate forecasting is possible. Firstly, forecasts that assume a continuation of past conditions in the future are, by definition, much more likely to fail when the environment is changing because they are over-fit for past conditions and cannot predict drivers that may impact ecosystem dynamics (Pennekamp et al. 2019). Secondly, it may be impossible to predict the drivers of an ecosystem's dynamics, rendering it impossible to apply a detailed, mechanistic understanding of the driver effects to forecast the future. For example, imagine that a retrospective, experimental manipulation has revealed that cumulative monthly rainfall determines the timing of flowering in a plant species. No matter how precisely the underlying mechanisms of the plant flowering are understood, a retrospective-based forecast is almost impossible because rainfall can only be forecasted a few days in advance. Without the ability to accurately predict the variable (monthly rainfall) that drives the ecological process of interest (plant flowering), the ability of a retrospective-based understanding to improve the forecast is very limited (Pennekamp et al. 2019). Finally, we cannot measure the parameters of forecast models without error. Therefore, adding the many parameters needed to capture biological mechanisms can inflate the total forecast error to a level that the forecast becomes meaningless (e.g., a forecast in which every possible future state is equally likely; Dietze 2017). We need to identify tools and approaches that complement what we have learned from retrospective studies and pivot them towards better understanding and controlling of the future dynamics of ecological systems (Dietze 2017).

Ecological forecasting—the process of generating model-based, quantitative predictions of the future state and dynamics of an ecological system—could provide a strategic complement to the standard scientific method in a changing world (Clark 2001). Ecological forecasting is in its infancy, and many of the most fundamental questions about the forecastability of ecological systems and forecast skill of different modeling approaches remain open. Scientists are pressed to figure out if ecological systems are *even* forecastable, and if so, where, when, and why forecast errors appear (Pennekamp et al. 2019). A big knowledge gap in ecological forecasting surrounds the range of model complexity and inclusion of biological mechanisms: Do forecasting models need detailed biological information to make accurate predictions? In an early test of this question, Tredennick et al. (2016) compared the ability of two models to forecast changes in plant community structure: one based on statistical patterns in the change of each species' cover and the second based on a biologically-informed, complex model (an integral projection model or IPM). Although the second model contained vastly more biological detail, the more naive forecast model had nearly equal forecast skill (Tredennick et al. 2016). Similarly, a forecasting competition focused on penguin population size found that biologically-ignorant models made by non-biologist outperformed a model detailed by experts (Humphries et al. 2018), countering the intuition that attempting to include the full complexity of ecological systems is necessary. While there are *still* no definitive conclusions on which model type outperforms others or if ecology is *even* forecastable, it will be important for ecologists to implement a wide range of models that cover the spectrum of biologically-naive statistical approaches to detailed statistical models (Humphries et al. 2018; Tredennick et al. 2016).

Improving the skill of ecological forecasts will require study species that are amenable to a wide range of forecast models and that allow the resulting forecasts to be assessed rigorously. A high availability of complementary data sources will be essential to meet the first criterion, and we argue that species with life history complexity are ideal for evaluating forecast skill at the population level. Forecasts that analyze long-term data sets tend to have a higher forecast skill because they offer more opportunity to train the model about a wider range of past conditions. The higher quality and quantity of a historical dataset, the more information there is to estimate the parameters of our forecast model, and more precise parameter estimates reduce the overall uncertainty in the resulting forecast (Humphries et al. 2018; Pennekamp et al. 2019). Although forecasting a system with a complex life history introduces more variables into forecasts of population dynamics, this ultimately makes it easier to distinguish candidate models by forcing each to *simultaneously* predict multiple aspects of the population state (e.g., total size and breakdowns by individual age, stage, size, or sex). By analogy, two weather forecasting models might do equally well when predicting whether a city such as San Diego will be sunny or rainy. However, one of the models could predict “sunny and warm temperatures” while the other predicts “sunny and below freezing temperatures. By adding another dimension to the forecast, we are better able to analyze forecast skill. Moreover, most species have some aspect of life cycle complexity, whether it is multiple growth stages, separate sexes, or longevity (Salguero-Gómez et al. 2016), each of which may differ in its response to climate change, yet all of which depend on one another to complete their life cycle. A forecast that accounts for more biological complexity will be more comparable to natural processes and ultimately exportable to other species (Salguero-Gómez et al. 2018).

In this study, we forecasted the population sex ratios of *Valeriana edulis* (hereafter, *Valeriana*) by assessing the forecast skill of models with varying amounts of biological detail with the end goal of determining the ideal range of forecast model complexity. *Valeriana* is an ideal subject for forecasting: it has ample historical data and is experiencing climate conditions that deviate from the historical levels of variance (Petry et al. 2016). *Valeriana* is a long-lived perennial, dioecious herb that grows in the Rocky Mountains of western North America. Through four decades of climate change, and in response to experimental manipulations to simulate future climate change, there is evidence that population sex ratios and demographic rates of *Valeriana* respond to climate over its contemporary elevation range (Petry et al. 2016; Petry et al. unpublished). It is imperative that we can predict the shifts in sex ratios, because if *Valeriana*’s sex ratios become too biased, it will be unable to reproduce effectively, ultimately leading to devastating impacts on the individual species and the ecological organization (Petry et al. 2013; Petry et al. 2016). We specifically seek to answer the following questions about the forecastability of *Valeriana* populations under a changing climate:

- 1) **Can we build forecast models for *Valeriana edulis* sex ratios that outperform a naive, null model?**
- 2) **How much biological detail needs to be included in the model to make a good forecast?**

In addressing these questions, this study provides a unique opportunity to explore the necessary biological detail that needs to be incorporated into a successful forecast while also allowing us to investigate *Valeriana*'s sex specific responses to climate change.

METHODS:

Forecast target variables

For this project, we focused on predicting *Valeriana edulis* sex ratios observed in 2020. We visited the population during peak flowering, walking transects to count flowering males and females separately until all individuals (at least 200 in very large populations) have been sexed. We characterized operational sex ratio (i.e. of flowering individuals), a population parameter that has the potential to vary substantially over time based upon the relative flowering of males and females.

Training and evaluation dataset

Our training dataset consists of sex ratio data that has been collected annually from 2011-2019 in 16 populations of *Valeriana* in Gunnison County, Colorado (Fig. 1). These populations span the plant's ca. 1800m elevation range.

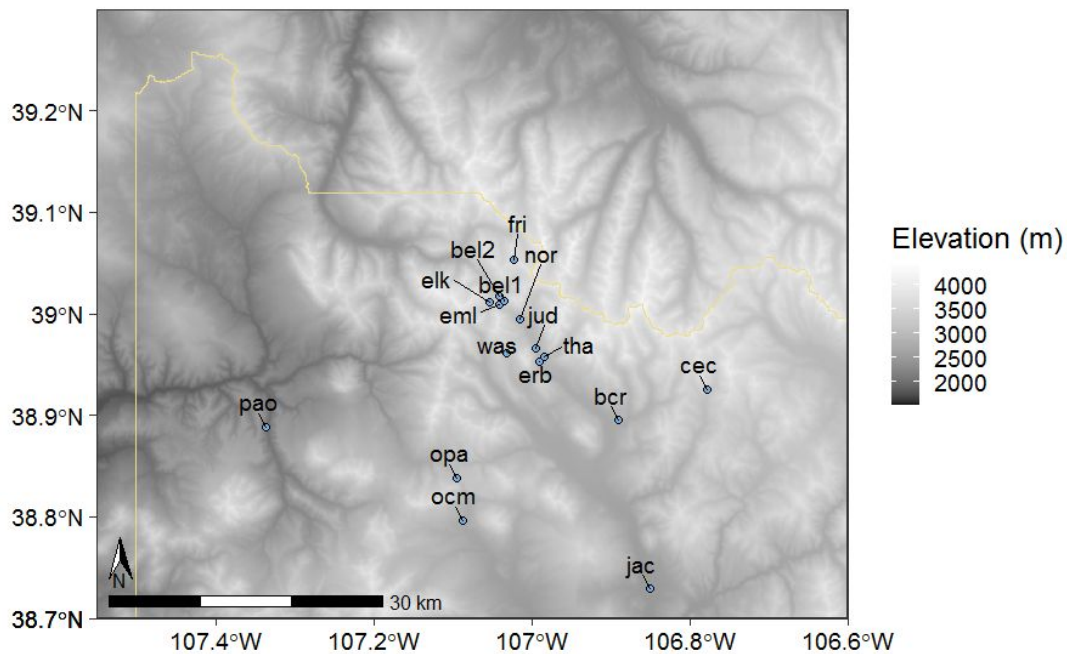


Fig. 1 An elevation map with labelled dots for each of the 16 populations.

We created a forecast evaluation dataset by repeating the same field protocol in each population in 2020. Importantly, these data were held out from the forecast model fitting procedures (Fig. 2) (described below) As such, our 2020 sex ratio data represent a true out-of-sample evaluation dataset.

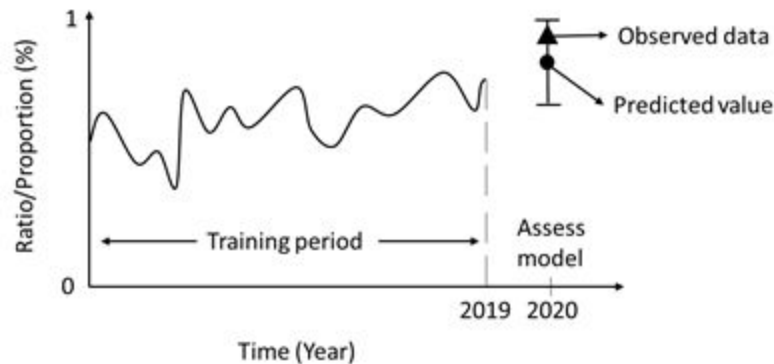


Fig. 2 Conceptual schematic illustrating the process of constructing and evaluating a sex ratio forecasting model. All available historic data up to and including the year 2019 was used to train the model. Probabilistic forecasts of the sex ratio in 2020 (triangle with error bars) were generated from the trained model, and compared to the observed sex ratio (circle). See “Assessing forecast skill” for model scoring methodology.

Candidate forecast model construction

In order to determine whether *Valeriana* sex ratios are forecastable (Q1) and how model complexity influences forecast skill (Q2), we built a series of models with increasing detail on the biological processes that drive changes in population sex ratios. As in weather forecasts, measures of forecast uncertainty are essential to the usefulness of ecological forecasts. We used a Bayesian approach for model construction and analysis, which allows us to fully incorporate uncertainties in the measurement of the current population sex ratio (i.e., counts of males and females rather than direct measurement of the sex ratio) and in how much driver variables will change the population composition into a probability distribution of the future population composition (Dietze, 2017). Because sex ratios cannot be less than 0 or greater than 1, our models used a logit link function and serially autocorrelated, binomially-distributed errors to ensure these bounds were not crossed.

We fitted five candidate forecasting models, detailed below in order of increasing biological complexity:

(1 - Random walk model) The first model simply forecasts that the sex ratio will stay the same with any deviation caused by process and observation error. During fitting, the model learns only about the magnitude of these deviations (i.e., how much the sex ratio can change in just one year). Although the most likely future sex ratio will be identical to the last observed sex ratio (i.e., the 2020 forecast will be equal to the 2019 value with an addition of estimated observation

error), forecasts from a random walk model can drift to 0, 1, or any value in between if deviations tend to be large or if projected far into the future. For the purpose of our study, we used this random walk model as the baseline null model against which we measured the performance of all other models.

(2 - Constant mean model) The second model adds only one constraint on sex ratios that vaguely approximates biological processes: that there is some unknown mechanism(s) that keep the population composition constant through time. Any deviations from this constant mean are purely the result of observation error and stochasticity arising from finite sample sizes. This model will contain two parameters that will be estimated from the data: a constant mean and constant variance (analogous to the “deviations” in the random walk model) for the entire observed time series. The 2020 forecast will therefore be a Beta distribution arising from a normal distribution with the estimated mean and variance that has been back-transformed to the probability scale.

This model assumes that the calculated mean value of sex ratio will remain constant as time changes. Although this project focuses on year+1 forecasts, it can be helpful to recognize that the constant mean model would make identical forecasts for all year+n forecasts (i.e., the Beta-distributed forecast model trained up to 2019 would not change for 2021, 2022, ...). Therefore, this model assumes that all variation in observed sex ratios is from noise or uncertainty caused by observation error, which does not affect the underlying process itself (Dietze, 2017).

(3 - Linear trend model) We know from previous research that male frequency has increased over time due to climate change in *Valeriana* populations (Soule, 1981; Petry et al., 2016). A linear relationship is the simplest model that allows for directional change in sex ratios. Therefore, to incorporate our prior understanding of the species into a forecast model, we added a slope parameter to the constant mean model that would allow for a constantly *changing* mean sex ratio. The complexity of this model increases as we add in the rate of change over time as another parameter, now apart from the uncertainty caused by observation error, we will also need to consider parameter uncertainty (Dietze, 2017). Although incorporating our knowledge of *Valeriana*'s long-term sex ratio change might be helpful to construct a better forecasting model, adding in more parameters will also bring in more uncertainties, because we are not sure if our calculated rate of change based on past data is exactly accurate and actually constant, and whether it will remain the same in the future.

(4 - Climate driver models) The last two models add a further layer of biological complexity including the key climate drivers that are changing *Valeriana* population composition. Petry et al. (2016; unpubl.) have shown that the date of snowmelt and previous growing season temperature can drive rapid changes in the performance of males and females. Demographic projections could explain the observed changes in population sex ratios over the species elevation range and in response to recent climate change. However, both candidate drivers show strong fluctuations from year to year in addition to directional change. We tested these

climate drivers separately. Each employs a single climate driver, and as such they are equivalent in terms of biological complexity.

(4a - Climate: snowmelt model) For this model, we included the date of snowmelt as a covariate in place of the single linear trend used in the previous model. We recorded the first date of bare ground at each *Valeriana* population from 3m-resolution Planet satellite imagery (“Daily Satellite Imagery and Insights” 2019). The observation error and parameter uncertainties still exist and will be carried on in the process.

(4b - Climate: temperature model) Following the same idea of the previous model, we tested the effect of previous growing season temperature on *Valeriana* population sex ratios. We replaced the date of snowmelt with the previous growing season temperature as covariate in place of the single linear trend model. We estimated growing season temperatures from a high resolution climatological interpolation (PRISM Climate Group 2014). The observation error and parameter uncertainties still exist and will be carried on in the process.

Assessing forecast skill

For each forecast model, we compared the probability distribution of the predicted population composition (sex ratio) with the actual observed data collected in the field (Fig. 2). We repeated the process for all the populations, and assessed the performance of each model based on the accuracy of its predicted value using the root mean square error (RMSE). This commonly used metric for scoring forecasts is essentially the standard deviation of the distribution of residuals (Dietze, 2017); in other words, RMSE tells us, on average, how far the forecasted value is from the observed data. Lower RMSE values indicate a model with higher forecast skill.

To address our study questions, (Q1) we compared the RMSE of our more complex models to the RMSE values of the null (random walk) model, (Q2) and measured how forecast skill changes as a function of model complexity.

RESULTS:

We successfully fit each candidate model to the observed sex ratio data in all 16 populations, and we found strong differences in forecast error (RMSE) both among and within populations (Fig. 3). Among-population differences in forecast error revealed that some populations were more forecastable (i.e., lower RMSE across all models) than others. For example, at Paonia B, all models had very small RMSE (the mean RMSE of the five models range from 0.0068 to 0.0512); but at Frigid Air Pass, the RMSE were all large (the mean RMSE of the five models range from 0.1685 to 0.2823). There were also consistent patterns in model performance within populations (Fig. 3).

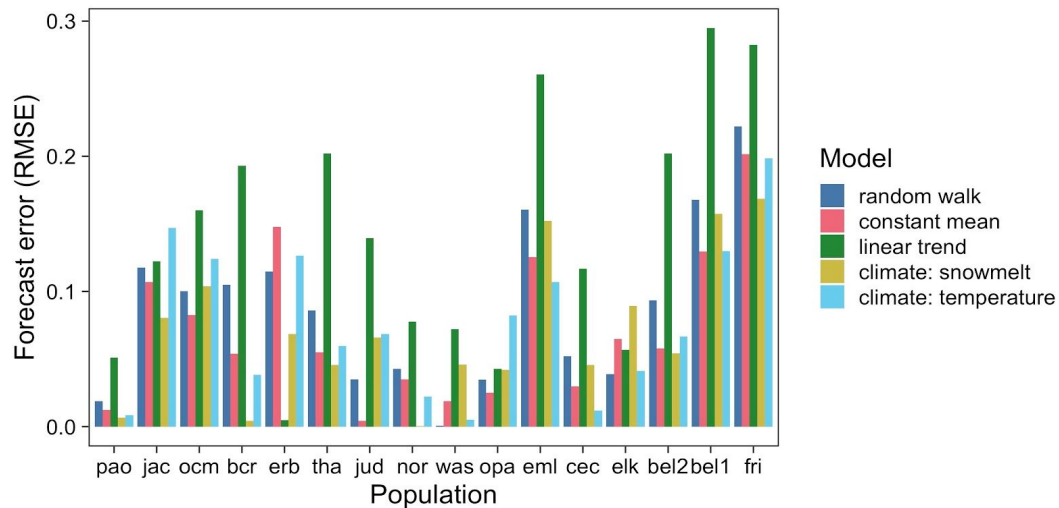


Fig. 3 Bar plot of root mean square error (RMSE) for all five sex ratio forecast models (color) at each of the 16 study populations (bar groups) The forecastability of sex ratios varied among populations and the relative forecast skill of each candidate model differed within populations.

At the majority of our populations ($n = 14$ out of 16) at least one of our models outperformed the null model (Fig. 4). The random walk model mostly ranked fourth in the 16 populations, meaning at those populations three of the other four models all made better forecasts than the random walk model (Fig. 4). Surprisingly, the only model that generally performed worse than the null model is the linear trend model. It only outperformed the random walk model once and ranked fifth in 12 populations.

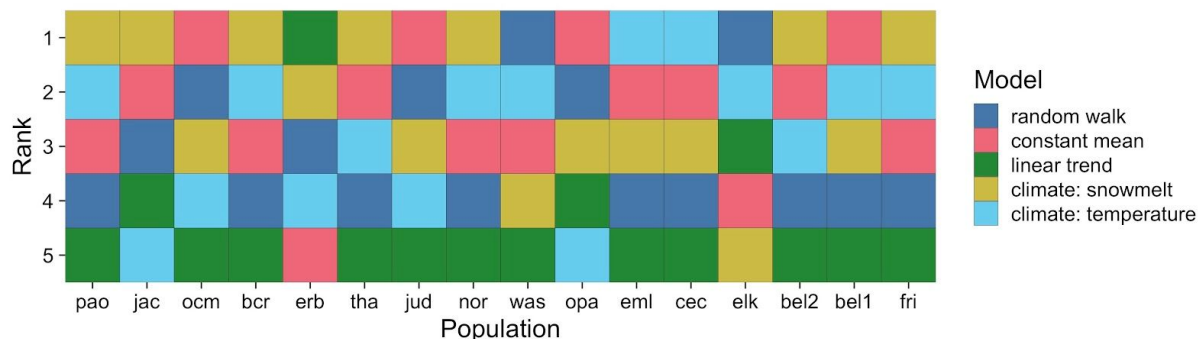


Fig. 4 Model forecast skill ranks (1 = highest skill/lowest error) for all five models at each of the 16 populations, sorted in order of increasing elevation. The climate driver models tended to be ranked highest whereas the random walk and linear trend models had the poorest performance.

Among our five forecast models, those with the most biological details tended to have lower forecast error (Fig. 5). The two climate driver models were the only models that identified the underlying biological drivers, and the fact that overall the climate models did better than the simpler models means adding biological details to the model can aid the forecasting of sex ratios.

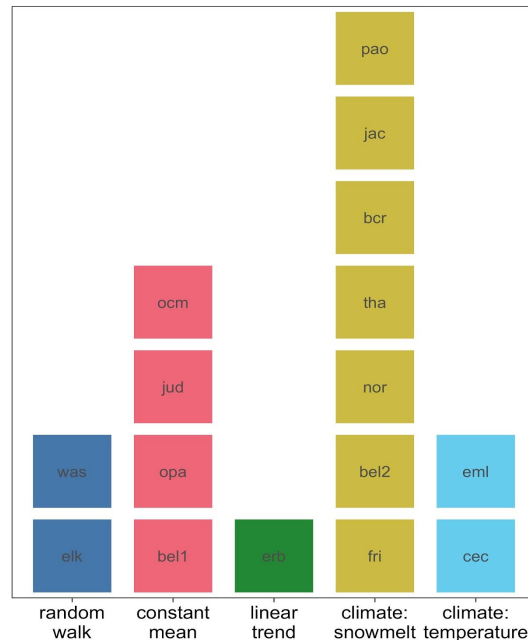


Fig. 5 Labelled qualitative plot of the best model in 16 of the populations. The models are arranged in the order of model complexity following the same color scheme as Figs. 3 and 4. It shows complex models with biological details tend to make more better forecasts.

DISCUSSION:

We showed that climate models generally outperformed the null model when forecasting *Valeriana* population sex ratios. Our abstract simplification of climate change effects on sex ratios (the linear trend model) based on a previous study performed poorly, often worse than the null model. Instead, the best-performing models tended to include the highest level of mechanistic detail we considered (snow melt and temperature climate drivers), despite the additional uncertainty that each parameter adds. Below we explore why these patterns likely arose and discuss the prospects for both improving forecasts of *Valeriana* sex ratios and for extending ecological forecasting to anticipate the responses of ecological systems to ongoing climate change.

The relatively poor performance of the (null) random walk model in comparison to the climate informed models suggests that population sex ratios of *Valeriana* are forecastable. The random walk model is a biologically-unrealistic model. It's more likely that such an ecologically-important aspect of a population as sex ratio doesn't purely drift around. However, this lack of realism is a useful foil for our other models—if the other models did equally well as the RW, we would be forced to accept that our biological understanding (represented by the other models) doesn't add much value. At the same time, we're not "stacking the deck" by including the RW as a model we know will be doomed to fail. Remember that in *Paonia B*, there's very little fluctuation from year to year in the sex ratio. When asked to forecast 1 year ahead, the RW is going to be pretty confident that the observed sex ratio will be very similar to the previous year. That makes it a very difficult model to improve upon. An analogy can be that it is almost always sunny in San Diego, so it's hard to beat a model that almost always predicts

the weather will be sunny. And yet, we did do better.

Despite the results of a previous study suggesting a consistent increase in male frequency in response to recent climate change (Petry et al. 2016), forecasts from the linear trend model tended to perform poorly (high RMSE values) relative to other models including the (null) random walk. Considering these findings, it came as a surprise that only one population produced a linear trend forecast with a low forecast error and that it was often outperformed by even the (null) random walk model. This mismatch suggests that population sex ratios are not changing in a pattern as linearly as originally predicted. Our results suggest that the common practice of assuming that ecological systems will respond linearly to climate change is overly naive and may produce poor predictions, at least in the near-term. While we can only speculate as to why the linear trend forecasts had the highest error, the influence of climate drivers likely plays a role. Climate variables often fluctuate from year to year, even before the most recent onset of climate change. As such, trends (whether linear or otherwise) tend to only be observable over long time periods. For some variables, climate change is amplifying these fluctuations. Therefore, we speculate that near-term ecological forecasts will perform better when they incorporate fluctuations in the climate drivers themselves rather than (or in addition to) long-term trends.

Although the intermediate-complexity linear trend model generally failed to outperform the baseline random-walk model, our most complex models that integrated climate drivers had higher forecast skills than we had expected. Past forecasting studies have investigated the intuition that biologically-informed models will outperform more abstract, phenomenological models. However, in each of these tests, the biologically informed models that incorporated demographic and climate data failed to outperform the simpler models on both occasions (Humphries et al. 2018; Tredennick et al. 2016). To our knowledge, our climate driver models are the first ecological forecasts to outperform their component naive models. Past forecasting studies produced unsuccessful biological models, and while our complex models proved to be successful with low RMSE values, there is a possibility that we incorporated less parameters into our models, making them more susceptible to overlook error. A recent climate change experiment lends support to the two candidate climate drivers that tended to produce the best sex ratio forecasts in our study. The experiment (Petry et al. unpubl. data) simulated future climate experience by *Valeriana* by advancing snowmelt and increasing temperatures at one of our study populations ("nor"). From individual-level demographic rates measured from plants under these manipulations, population sex ratios were projected using an Integral Projection Model. The investigators found that accelerated snowmelt caused a decrease in male frequency whereas warming increased male frequency. Following from these independent data, we expected that we would see negative coefficients for the snowmelt driver and positive coefficients for the temperature driver in our forecast models in the populations where these models performed best ($n = 7$ and $n = 2$, respectively; Fig. 5).

Comparing these expectations to our estimated coefficients, we found that the mean coefficients for the snowmelt model tended to be negative (5 of 7 cases where the snowmelt model had the highest forecast skill), however, the coefficients for the temperature driver model failed to match the expectation (0 of two cases where the temperature model had the highest forecast skill).

Table. 1 Mean climate factors (snowmelt date and previous growing season temperature) coefficient values for 16 of the populations with lower (2.5%) and upper (97.5%) distribution boundaries. A positive mean value indicates a positive effect of the climate factor on the sex ratio, and a negative mean value indicates a negative effect. Populations highlighted are which the snowmelt model (top) and temperature model (bottom) did the best.

Climate factors coefficient							
snowmelt date							
population	mean	2.50%	97.50%	population	mean	2.50%	97.50%
bel1	0.00209	-0.01488	0.01527	jac	0.00050	-0.04042	0.03250
bel2	-0.00113	-0.01520	0.01063	jud	-0.00219	-0.01628	0.00950
bcr	-0.00756	-0.03426	0.01820	nor	0.00127	-0.01561	0.01704
cec	0.00197	-0.01437	0.01876	ocm	0.00286	-0.01893	0.02137
erb	-0.00270	-0.02065	0.01518	opa	0.00376	-0.01028	0.01648
elk	-0.00216	-0.01621	0.00991	pao	-0.00187	-0.02977	0.02339
eml	0.00491	-0.00977	0.01642	tha	-0.00638	-0.02247	0.01036
fri	-0.00060	-0.01455	0.01343	was	-0.00690	-0.02117	0.00711
temperature							
population	mean	2.50%	97.50%	population	mean	2.50%	97.50%
bel1	0.03024	-0.19470	0.21808	jac	0.01429	-0.15867	0.18387
bel2	0.01949	-0.22488	0.23623	jud	0.06742	-0.10560	0.22429
bcr	-0.08353	-0.26337	0.09523	nor	0.00672	-0.19295	0.20423
cec	-0.01599	-0.19626	0.16071	ocm	0.04463	-0.14044	0.21333
erb	-0.02880	-0.22762	0.15533	opa	0.04909	-0.11326	0.19220
elk	-0.02919	-0.21039	0.14261	pao	-0.03206	-0.15806	0.09497
eml	-0.04283	-0.29441	0.20404	tha	-0.07910	-0.26483	0.11318
fri	-0.02899	-0.28201	0.22786	was	0.01091	-0.18731	0.21012

Despite the lack of direct correspondence between the two independent studies, we conjecture that improvements to the near-term sex ratio forecasts will come from more deeply incorporating information on climate driver(s). One likely reason for the apparent mismatch is that our forecasting study was focused on short-term (1 year), potentially transient changes in sex ratio instead of the long-term equilibrium sex ratio projected under the climate manipulation experiment. If this was the case, this might explain why we produced successful biologically informed forecasts even though the results from the two independent experiments are not entirely congruent. Moreover, it is possible that *Valeriana* sex ratio responses to climate drivers vary among populations, perhaps as a result of local adaptation to the differences in climate across their elevation range that cause them to respond differently to climate drivers (Moran et al.2017). If this is the case, the lack of relation of the results in the two independent experiments could be attributed to differences in the *Valeriana* populations. However, there is a possibility that our climate forecasts produced a high forecast skill through chance, amplified by the relatively short historical snowmelt data that we were able to extract from satellite imagery to incorporate into the model. Since we were only able to use snowmelt data from 2017-2019, there was very little data in which to train the model, and when there is a lower quantity and quality of historical data in which the model is trained, there is a higher chance for error to

appear in the forecast (Humphries et al. 2018; Pennekamp et al. 2019).

To our knowledge, this is the first study that has, to a degree of success, used models to generate quantitative forecasts of *Valeriana* sex ratios. Since ecological forecasting is well-suited for improvement through iteration, we advocate using our study as a foundation for further improvement rather than an endpoint. We see several clear avenues for refinement.

First, we suggest expanding the training dataset by continuing to collect sex ratio data and by filling gaps in the existing data. Even though the earliest *Valeriana* sex ratio data dates back to the late 1970's (Soule 1981), there is a significant time gap before data collection resumed in 2011 and there are occasional gaps for some populations in the intervening years. A nine year dataset is a long dataset compared to many other ecological datasets. However, forecasts that analyze long-term data tend to have a higher forecast skill because they offer more opportunity to train the model about a wider range of past conditions. Ultimately, a higher quality and quantity of data allows us to minimize the overall uncertainty when generating an estimate of the parameters of our forecast (Humphries et al. 2018; Pennekamp et al. 2019). It would be productive for both the field of ecological forecasting as well the study of *Valeriana* populations to compare the forecast skill of future models that are trained with more historical data to the skill of our models from 2020. Furthermore, since the technological capability of satellites has only recently allowed us to accurately record the date of snowmelt at each *Valeriana* population, there is only useful snowmelt data from approximately 2017-2020 that informed these forecasts. Since the snowmelt climate driver model was the overwhelmingly dominant model, it would be useful to train the model with more snowmelt data to see if it is still as successful for future forecasts as it was in 2020.

Second, we think that combining the predictions of our models (i.e., model ensembling) could improve forecast skill. Results from past forecasting attempts suggest that ensembling techniques can produce forecasts with higher skill than individual models (Humphries et al. 2018; Tredennick et al. 2016). By combining models in ensembles, the prevailing “signal” can be strengthened by capturing the consistencies in the predictions of the component models while simultaneously eliminating or “averaging out” the idiosyncratic errors in each forecast (Dietze 2017). A data science competition aimed at forecasting penguin population size used ensemble techniques found that this approach outperformed every other component model in the study (Humphries et al. 2018). Since our climate driver models were so successful, it would be interesting to see if an ensemble approach would gain extra predictive power. Regardless of the result, this is a future direction that needs to be investigated with the intent of learning more about *Valeriana* populations closing knowledge gaps in ecological forecasting,

Lastly, we urge future sex ratio forecasters to fit models that combine multiple climate drivers and/or include more biologically-based models of sex ratio change (e.g., demographic models). In addition to historic and long-term sex-ratio data, there are parallel demographic datasets from several of the same populations we focused on in this study. These demographic data have tracked plant performance related to the sex, size, and reproductive status of individual plants. In future years, it would be informative to create models that incorporate this demographic data, or a combination of climate and demographic data. A model that incorporates aspects of the life history of a species will introduce more variables into the forecast, which could ultimately increase forecast error. However, by incorporating more

variables, there is more opportunity to see where forecast error might arise, which is an important aspect in the process of refining a model and improving forecasts (Salguero-Gómez et al. 2018).

Prospectus and conclusion

From first principles, we expect that there will be an optimal forecast model at intermediate model complexity (Fig. 6). Such a model would include enough biological detail in the form of model parameters to closely approximate the real system's dynamics while balancing the inflation of forecast error that comes from adding each imperfectly estimated parameter. Our candidate models revealed a generally decreasing relationship between model complexity (=number of parameters) and forecast error. This presents an optimistic outlook for the future of *Valeriana* sex ratio forecasting—the optimal model is likely yet to be found. Forecasting is an iterative process, it is important that we keep refining these models until we find this range. Although *Valeriana* is just one of many millions of species that will be subjected to ongoing climate change, we anticipate that the lessons learned from forecasting will become essential tools for managing the future of biodiversity. Ultimately, a step forward in forecasting is also a step towards generating a more sustainable future in which we can predict the chemical and biological cycles that impact not only ecological systems, but also human health and public policy.

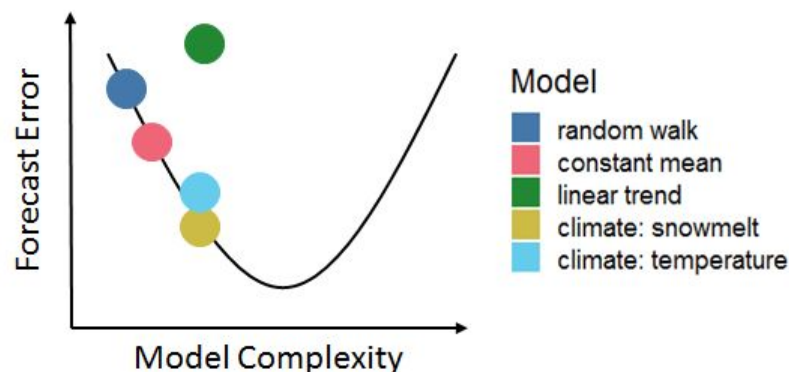


Fig. 6 Conceptual relationship between forecast error and model complexity. The approximate location of each of our models is shown as colored points. In theory, the relationship curve should be a U-shaped curve, since we've shown that increasing model complexity can reduce forecast error, but at the same time it will also bring in uncertainties, so eventually the uncertainties caused by increasing complexity will lead to increase in forecast error.

ACKNOWLEDGEMENTS:

We would like to express our most sincere gratitude to Dr. Will Petry and Dr. Kailen Mooney for mentoring us through the research process, as well as previous students and technicians for assisting in sex ratio data collection: Jackson Snow, Val Watson, Laurel Schumm, Peter Innes, Su'ad Yoon, Ana Chicas-Mosier, and Kayla Perry. Furthermore, we

would like to thank the NSF REU program for supporting Sydney Petersen on this project, as well as the Rocky Mountain Biological Laboratory and University of California Irvine for providing us with facilities and equipment to make this research possible.

REFERENCES:

Clark, J.S. (2001). Ecological Forecasts: An Emerging Imperative. *Science*, 293, 657–660.

Daily Satellite Imagery and Insights. (2019). Planet. Available at: <https://planet.com/>.

Dietze, M.C. (2017). *Ecological forecasting*. Princeton University Press, Princeton.

Dietze, M.C., Fox, A., Beck-Johnson, L.M., Betancourt, J.L., Hooten, M.B., Jarnevich, C.S., *et al.* (2018). Iterative near-term ecological forecasting: Needs, opportunities, and challenges. *Proc. Natl. Acad. Sci.*, 115, 1424–1432.

Field, C.B., Barros, V.R. & Intergovernmental Panel on Climate Change (Eds.). (2014). *Climate change 2014: impacts, adaptation, and vulnerability: Working Group II contribution to the fifth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, New York, NY.

Humphries, G.R.W., Che-Castaldo, C., Bull, P.J., Lipstein, G., Ravia, A., Carrión, B., *et al.* (2018). Predicting the future is hard and other lessons from a population time series data science competition. *Ecol. Inform.*, 48, 1–11.

Moran, E., Reid, A., Levine, J. A. (2017). Population genetics and adaptation to climate along elevation gradients in invasive *Solidago canadensis*. *Plos One*. 12, 1-24.

Pennekamp, F., Iles, A.C., Garland, J., Brennan, G., Brose, U., Gaedke, U., *et al.* (2019). The intrinsic predictability of ecological time series and its potential to guide forecasting. *Ecol. Monogr.*, 89.

Petry, W.K., Soule, J.D., Iler, A.M., Chicas-Mosier, A., Inouye, D.W., Miller, T.E.X., *et al.* (2016). Sex-specific responses to climate change in plants alter population sex ratio and performance. *Science*, 353, 69–71.

PRISM Climate Group. 2014. PRISM. Oregon State University.

Salguero-Gómez, R., Jones, O.R., Jongejans, E., Blomberg, S.P., Hodgson, D.J., Mbeau-Ache, C., *et al.* (2016). Fast–slow continuum and reproductive strategies structure plant life-history variation worldwide. *Proc. Natl. Acad. Sci.*, 113, 230–235.

Salguero-Gómez, R., Violle, C., Gimenez, O. & Childs, D. (2018). Delivering the promises of trait-based approaches to the needs of demographic approaches, and *vice versa*. *Funct. Ecol.*, 32, 1424–1435.

Soule, J. D. (1981). Ecological consequences of dioecism in plants: A case study of sex differences, sex ratios and population dynamics of *Valeriana edulis* Nutt. (Ph.D.). Michigan State University, East Lansing, Michigan, USA.

Tredennick, A.T., Hooten, M.B. & Adler, P.B. (2017). Do we need demographic data to forecast plant population dynamics? *Methods Ecol. Evol.*, 8, 541–551.