SIDE BLOTCHED LIZARDS

AN EVOLUTIONARY GAME THEORY MODEL

BUSINESS PROBLEM

A comprehensive look into the curious mating habits of Side Blotched Lizards. This project aims to shine a light on the driving forces behind mate selection, creating a model to simulate population changes over time according to evolutionary game theory principles. Further, a time series analysis will be completed to detect patterns within the generational changes.

BACKGROUND

Side blotched lizards, native to the northern coast of California, exhibit very peculiar mating behavior. There exist three distinct color variations: orange, blue and yellow. Each coloration not only has unique color profiles around their face and throat, but also exhibits vastly different personalities and mating strategies. This lends itself very well to game theory modelling, as each strategy strictly dominates another.

The orange lizards tend to be quite aggressive due to higher levels of testosterone. This causes them to seek out and defend larger territories, as well as leading them to establishing harems of female lizards. This mating strategy strictly dominates the blue lizards' strategy, as blue lizards struggle to defend their mates leaving them vulnerable. The yellow lizards, on the other hand, do not defend territories. Rather, they infiltrate the large harems of the orange lizards and steal mates without being detected. This strategy strictly dominates the large, sprawling harem of the orange lizard. The blue lizards, due to having fewer mates to protect, are typically able to defend their female counterparts from the yellow lizards, making their strategy strictly dominant over the yellow lizards (Sinervo).

The success of each mating strategy is based on the breakdown of the population dynamics. If there is a high proportion of orange lizards, blue lizards will have a poor mating season, orange lizards will have an average mating season, and yellow lizards will have a great mating season (Networks). These payoffs are demonstrated below:

	Orange	Yellow	Blue
Orange	(1, 1)	(0, 2)	(2, 0)
Yellow	(2, 0)	(1, 1)	(0, 2)
Blue	(0, 2)	(2, 0)	(1, 1)

DATA

Using the principles of evolutionary game theory, we will create a model that calculates the success of each generation's mating strategies based on the above payoffs. The model will take three inputs: the starting total population, the starting proportion of each color variation, and the number of generations to simulate.

A few assumptions will be made in order to create the simulation. The first assumption is that there is an even number of male and female lizards. The second assumption is that each individual is a rational actor who is maximizing their own payoff. The third assumption is that the lizard inherits the coloration of its father with a small chance of mutation (true heritability is h=0.96). The final assumption is that there is some form of environmental constraint that limits the population from increasing exponentially (Edwards).

Once the model takes the inputs, it will then compute the statistics for the following generation. It will, for each actor, simulate their mating success which will either end in "Success" or "Failure". Success is quantified as finding a mate and producing an offspring. The model will then calculate the number of successes for each color variation and update the demographic information, saving each generation into a row in the dataset. The final dataset will take the shape of 12 columns containing demographic information, and **x** rows, where **x** is the number of generations input to the initial model.

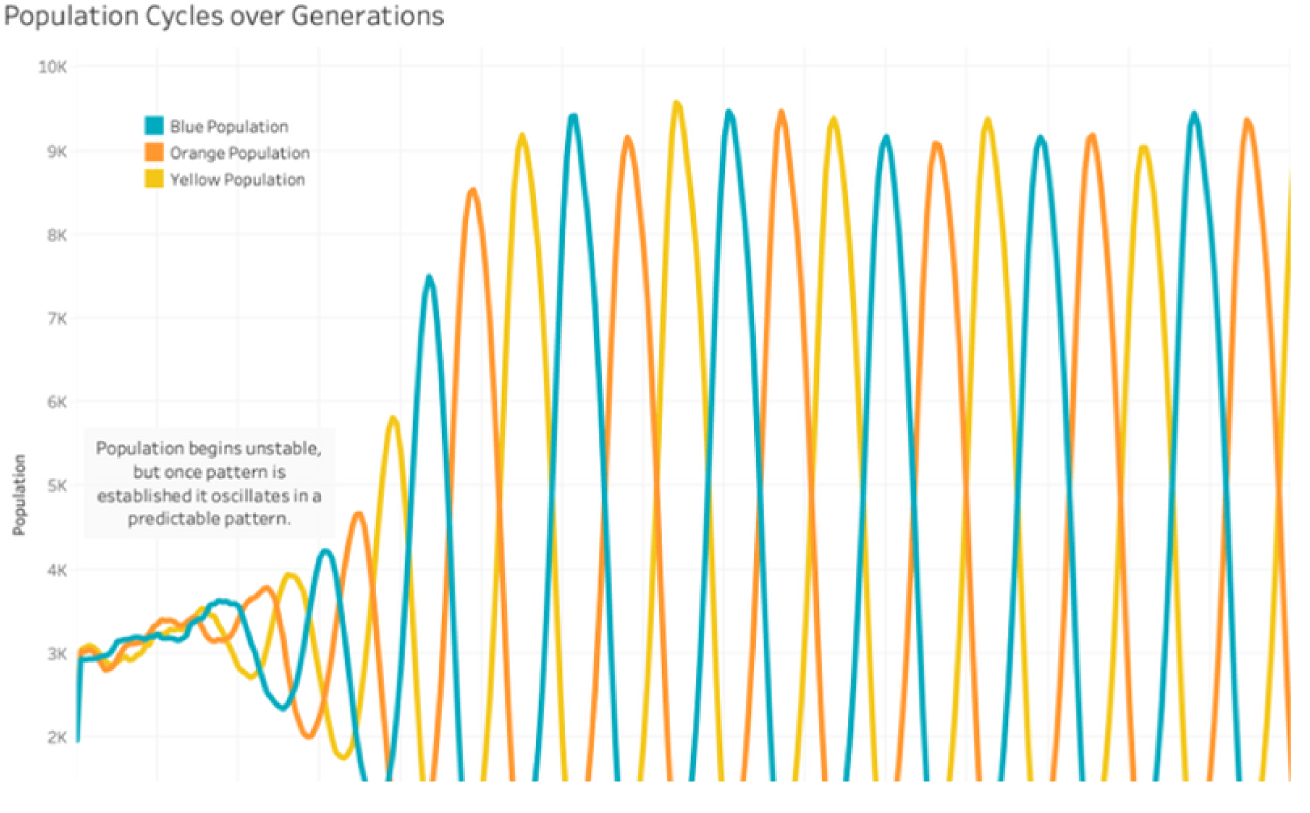
The columns included in the final dataset are as follows:

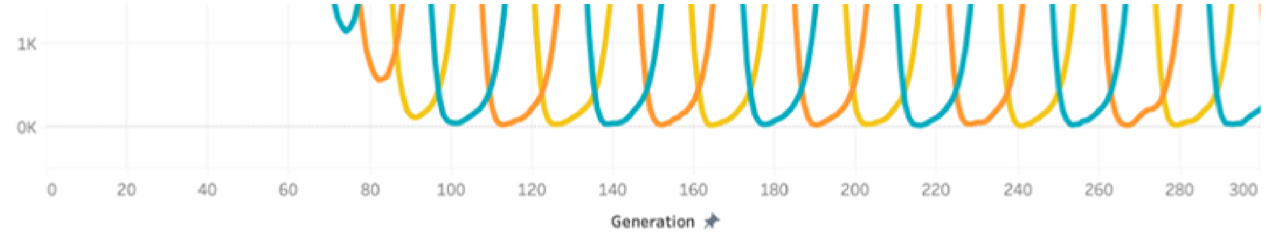
- NewGenPopulation – the total number of offspring produced
- NewGenOrange – the number of newly born orange offspring
- NewGenYellow
- NewGenBlue
- TotalPop – the total population, including the newly born generation, their parents, and the grandparent generation who are ineligible to produce children
- TotalOrange – the total number of orange lizards in the population
- TotalYellow
- TotalBlue
- ShareOrange – the share of the population that is comprised of orange lizards
- ShareYellow
- ShareBlue
- Best Strategy – the mating strategy (orange, yellow or blue) that successfully produced the most offspring this generation

METHODS

The final dataset contains information over the course of 1000 generations in the experimental environment simulated by our model. We will visualize this information and then create a time series model with the goal of predicting future generations' genetic makeup.

Over the course of generations, the population oscillates in a fairly predictable pattern:

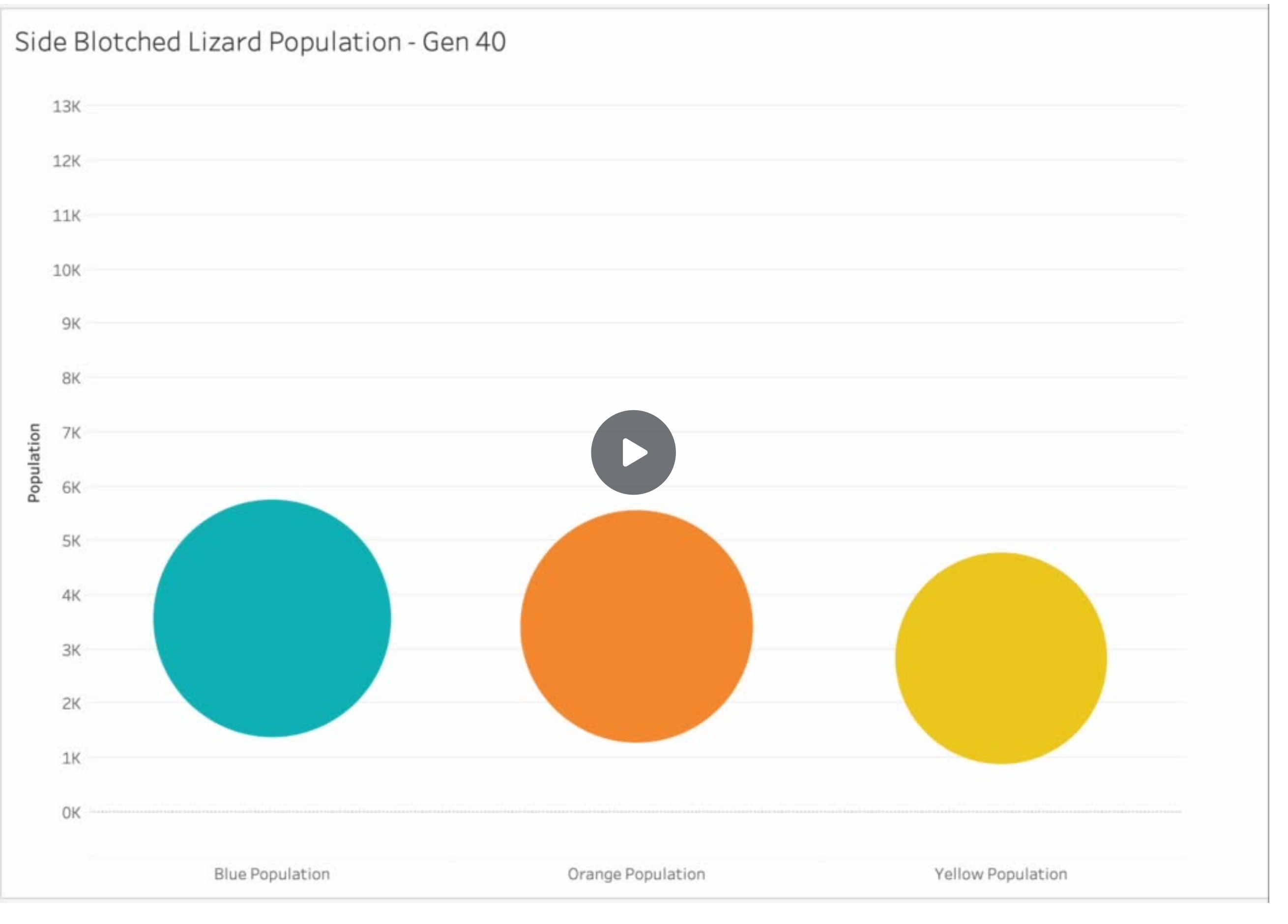




At the beginning, all parties have equal opportunities for payoffs. This results in population changing in an unstable way. Once one color variation has a clear numbers advantage, the clear dominance of strategies begins to emerge. This looks to occur with the blue color variation leaping ahead around generation 40.

Once the blue variation population rises, the orange lizards gain a significant advantage and are able to essentially steal the mates of their blue counterparts. This leads to a spike in orange population. Once the orange population rises, the yellow lizards then gain the advantage. They are able to sneak into the orange harems and steal mates from their territory. This leads to a third spike, this time in the yellow population. This the leads to the blue lizards gaining the advantage, engaging in monogamous relationships that leave the yellow lizards high and dry. This cycle continues as the population continues to grow.

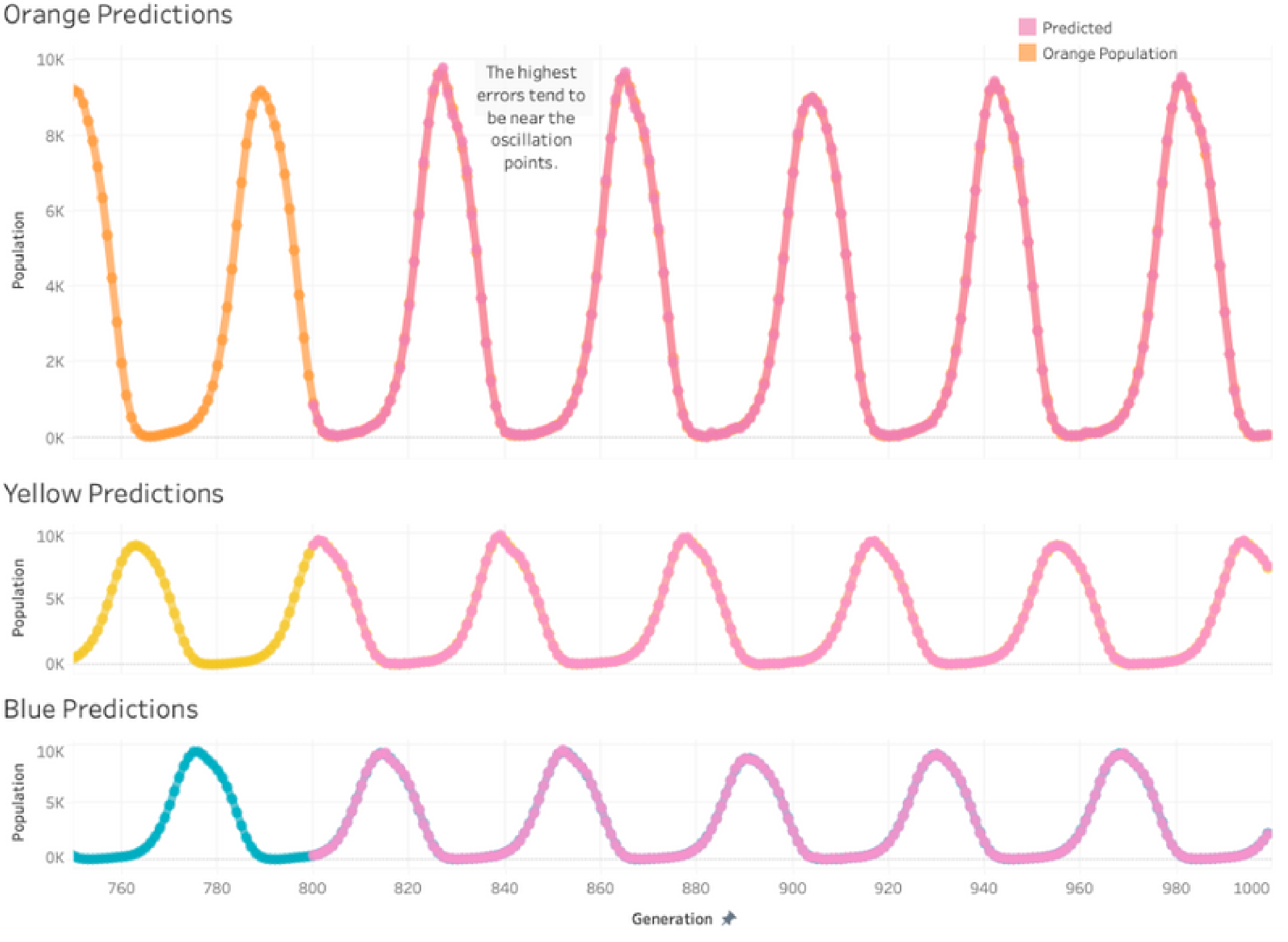
Here is a time lapse showing the population changes over time:



To predict the future population cycles, a time series model has been created to understand the timing and nuances of the population cycles. An ARIMA model, or an autoregressive integrated moving average model, which is designed to forecast future values in a time series by conferring with the most recent time periods and forecast errors (Mau).

This ARIMA model studies the cyclical trends of the offspring production and forecasts the future demographic breakdowns. This model has parameters of p=5, d=1, q=0. Meaning that it analyzes 5 autoregressive terms, 1 degree of differencing, and 0 lagged forecast errors. This establishes a model that takes into account the 5 most recent values for produced offspring to predict the value for the next generation.

The predictions are displayed below. The model summary, containing descriptive statistics for the model are in **Appendix B**.

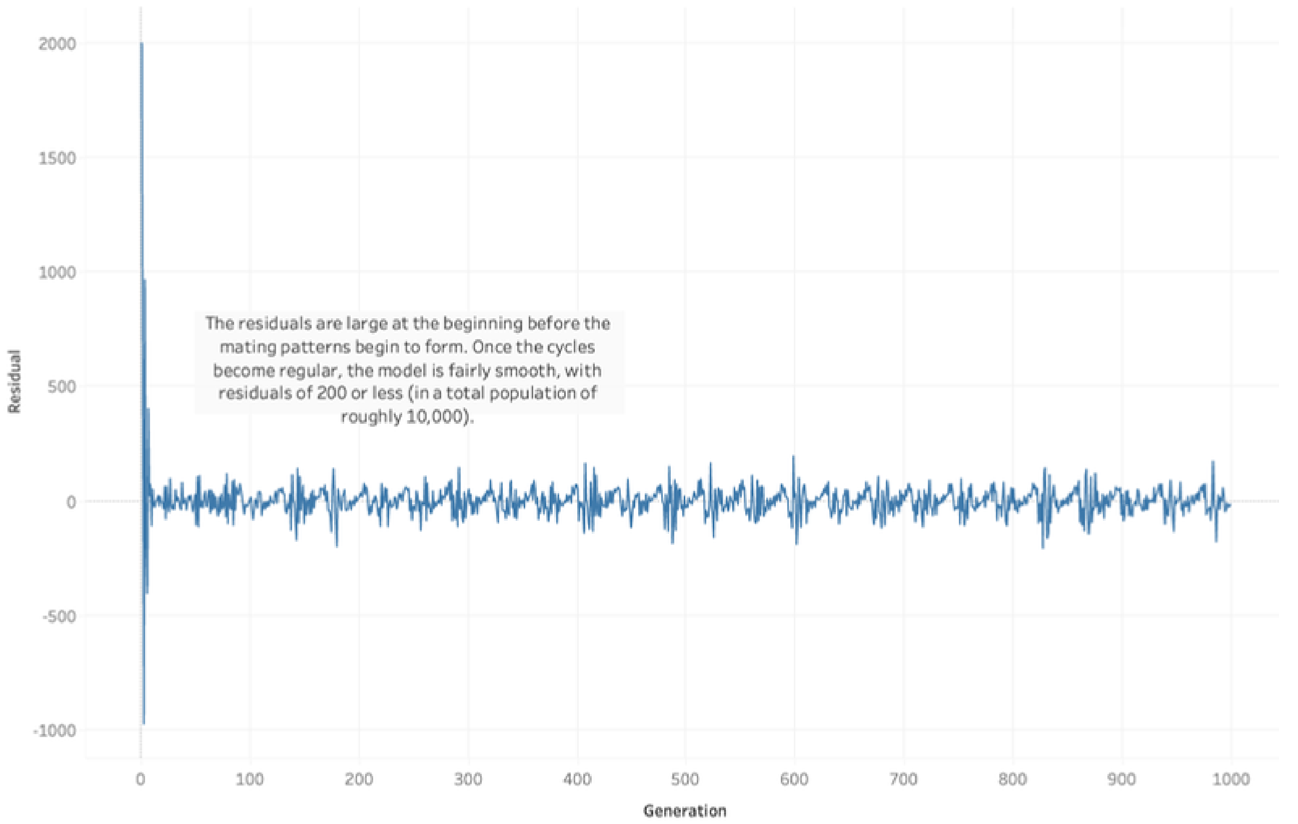


The pink lines represent the predicted offspring for the next 200 generations. We can see that the predictions follow the oscillating pattern very closely, mimicking the demographic changes that were produced by the simulator.

ANALYSIS

We will first take a look at the residuals for the model on the existing data. This shows us the difference between what the model expects and the true values. This number represents how closely the offspring of each new generation match the model.

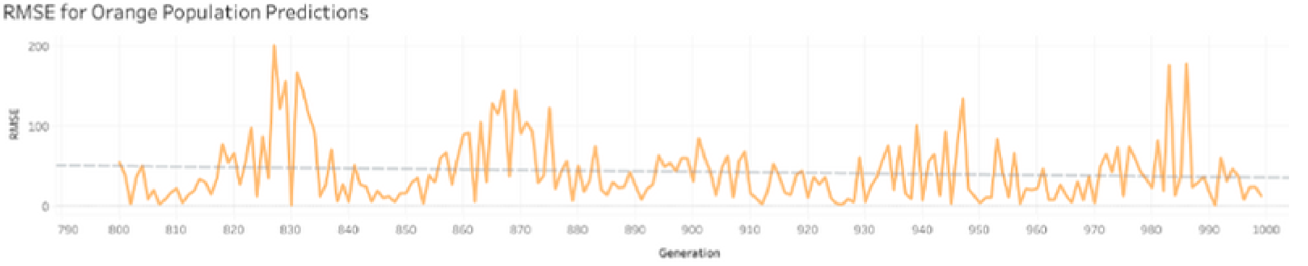
Residuals



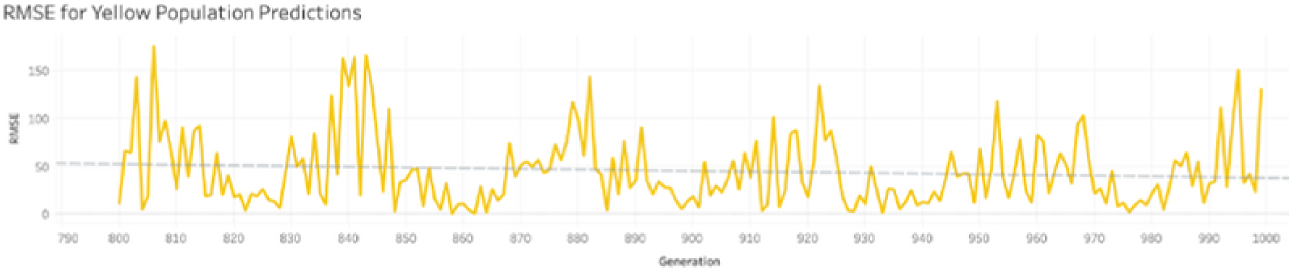
The residuals begin very large, as our population in the beginning is very volatile and unpredictable. Once the pattern is established and the population cycles become regular, the model's residuals are consistently below 200. The population caps at 10,000 lizards, so this leads to fairly accurate conclusions.

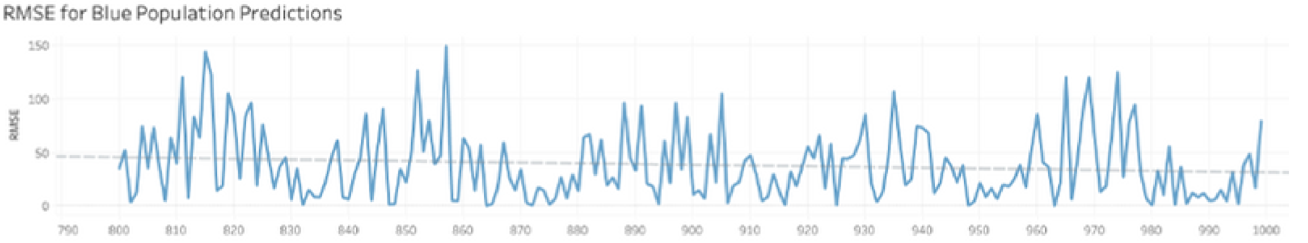
To assess the model's forecasts, we will calculate the RMSE, or Root Mean Squared Error. This metric looks at the difference between the predicted value and the actual value, putting more weight on larger errors. The RMSE for each of our three populations is displayed below:

RMSE for Orange Population Predictions



RMSE for Yellow Population Predictions





We can see a slight downward trend in RMSE as we continue through the generations as the model learns the patterns over time. There are also noticeable spikes in RMSE around the high and low points in the oscillation. This indicates that the model does not do as well when the ecological momentum is shifting. This is intuitive as the model is unable to see the shifts in the other populations, which is the driving factor for the quick momentum changes for breeding of a particular color variation.

CONCLUSION

The three strategy, strictly dominating Side Blotched Lizard mating game creates a fascinating vessel to explore environmental mating simulations. It is beautifully simple, allowing us to create a model that can replicate the population dynamics that exist in the real world. By simulating the strategies and seeing how they affect each individual population, we can get a greater insight into how and why they work. We can measure success, or lack thereof, and see the consequences on a large scale.

Simulating the environment provides us freedom that we do not have in real life, as it would take many years to observe each full cycle which lasts dozens of generations. We can use models such as this in an educational setting to demonstrate not only biological concepts, but also introductory game theory concepts as well. It is a fantastic educational tool that provides users the ability to gain a deeper understanding of mating strategies and evolutionary decision-making.

Through our experimentation, we have discovered the patterns that arise during Side Blotched Lizard mating seasons. We can visualize how an increase in blue lizard populations reduces the effectiveness of yellow lizard mating strategies, and how that detrimentally affects the yellow populations. Modelling the behavior through the three-actor game allows us to predict future demographic patterns in real populations based on their current genetic makeup. This information is invaluable to researchers and students studying the species.

ASSUMPTIONS

There were several assumptions made in the creation of the simulator in order to establish a stable experimental environment.

1. There is an even number of male and female lizards
2. Each individual is a rational actor who is maximizing their own payoff

3. Offspring inherit the coloration of its father with a small chance of mutation (true heritability is h=0.96)
4. There is some form of environmental constraint that limits the population from increasing exponentially

Each of these assumptions is critical in ensuring that the environment runs as intended. In nature, there are occasionally actors that act out of their self-interest, though this is difficult to quantify and would not be feasible to add to the model. There are also many intricacies when it comes to mutation, however, for this model we assume the true heritability and consider any mutation to simply be the lizard presenting features of a different coloration.

LIMITATIONS

The SARIMA model only takes into account the past behavior of a single variable when forecasting future values. In this particular scenario, the offspring produced by a certain generation is influenced by not only the population of the color variant, but that of the neighboring variants as well. Without being able to see the recent values of the other populations, our model is left with lower accuracy around the points where the mating momentum changes in the opposite direction.

Multivariate time series analysis is possible in the form of VAR models, or Vector Autoregressive Models. These models consider additional variables as well as their lag and errors. This may bolster the model to be more accurate around the oscillation points. However, the model has a very high accuracy already, and the computational cost that it would take to create a VAR would exceed the accuracy boost we would receive from its creation.

CHALLENGES

A few challenges arose as I was creating the evolutionary simulator. The simulator assumes that three generations are currently alive, with the oldest generation "dying" each cycle. Even though we were removing a chunk of the population, it was growing at an alarming rate (due to each male lizard having the potential to have up to 2 children) with no external bounds.

This led me to add the environmental constraints, capping the population so that it could not grow infinitely. In hitting the population cap, the lizards that were unable to survive (due to lack of food, predators, whichever storyline you find most compelling) were evenly distributed across all three color variants. This allowed the dynamics to remain the same while the number of lizard offspring did not become unmanageable.

RECOMMENDATIONS & IMPLEMENTATION

This simulator can be used for educational purposes for audiences to learn more about the intricacies of the Side Blotched Lizard. Simulations can be customized to have any number of generations, starting proportions and population sizes. By changing the settings, users can see how differently the populations shape up over the course of generations. This allows students to understand how quickly the dynamics change and learn when different mating strategies are successful.

It is also a helpful tool to provide an overview of introductory game theory. The simulator models a basic game with three strategies, each one being strictly dominant over another. The simplest version of this game is referred to as the "Rock, Paper, Scissors" game. The simulator takes this game and adds some evolutionary elements to it (mutation, environmental limitations) in order to as closely as possible mimic true mating habits. Running the simulator allows students to see how the different strategies play out over time as the genetic makeup changes.

ETHICAL ASSESSMENT

All data used has been derived from a simulator based on an evolutionary game. There is no risk in obtainment or improper use of this dataset. The most pressing ethical concern is that the assumptions are clearly laid out in advance of users experimenting with the simulator. It is close to impossible to perfectly mimic the conditions of nature, so assumptions must be made in order to create the most realistic environment. Quantifying a variable such as "population cap" would, in real-life scenarios, require an understanding of the ease of acquiring resources, seasonal changes, predators, migration and many other factors. It is unfortunately not feasible to gather all this information for a period of 10,000 generations. The assumptions made have been clearly outlined throughout the paper so the audience is aware and can draw any conclusions with clarity.

REFERENCES

Edwards, W. J. (n.d.). Nature news. https://www.nature.com/scitable/knowledge/library/populationlimiting-factors-17059572/

Mau. (n.d.). *Introduction to ARIMA Models*. Fuqua School of Business. https://people.duke.edu/~rnau/411arim.htm

Networks. Side Blotched Lizard and Evolutionary Game Theory : Networks Course blog for INFO 2040/CS 2850/Econ 2040/SOC 2090. (n.d.). https://blogs.cornell.edu/info2040/2016/11/26/side-blotched-lizard-and-evolutionary-game-theory/#:~:text=Under%20these%20conditions%2C%20the%20payoff,the%20payoffs%20of%20blue%20lizards

Sinervo, B., & Lively, C. M. (1996). The rock-paper-scissors game and the evolution of alternative male strategies. Nature, 380(6571), 240–243. https://doi.org/10.1038/380240a0

APPENDIX A

Displayed below are the first 20 rows of the experimental environment full simulation dataset. The full dataset can be found at this link: https://github.com/alissatrujillo/Portfolio/blob/main/Side%20Blotched%20Lizards%20-%20Evolutionary%20Simulator%20%26%20Time%20Series%20Analysis/pop.csv

	NewGenPopulation	NewGenOrange	NewGenYellow	NewGenBlue	TotalPop	TotalOrange	TotalYellow	TotalBlue	ShareOrange	ShareYellow	ShareBlue	BestStrategy
0	2972	1002	1011	959	5972	2002.0	2011.0	1959.0	0.335231	0.336738	0.328031	Yellow
1	3027	1025	1027	975	8999	3027.0	3038.0	2934.0	0.336371	0.337593	0.326036	Yellow
2	3011	991	1029	991	9010	3018.0	3067.0	2925.0	0.334961	0.340400	0.324639	Yellow
3	3037	1029	1040	968	9075	3045.0	3096.0	2934.0	0.335537	0.341157	0.323306	Yellow
4	2988	1002	1009	977	9036	3022.0	3078.0	2936.0	0.334440	0.340637	0.324923	Yellow
5	2926	944	987	995	8951	2975.0	3036.0	2940.0	0.332365	0.339180	0.328455	Blue
6	2896	932	971	993	8810	2878.0	2967.0	2965.0	0.326674	0.336776	0.336549	Blue
7	2859	925	950	984	8681	2801.0	2908.0	2972.0	0.322659	0.334984	0.342357	Blue
8	2920	956	948	1016	8675	2813.0	2869.0	2993.0	0.324265	0.330720	0.345014	Blue
9	2975	965	957	1053	8754	2846.0	2855.0	3053.0	0.325109	0.326137	0.348755	Blue
10	3082	1015	992	1075	8977	2936.0	2897.0	3144.0	0.327058	0.322714	0.350228	Blue
11	3048	1037	985	1026	9105	3017.0	2934.0	3154.0	0.331356	0.322241	0.346403	Orange
12	3089	1035	988	1066	9219	3087.0	2965.0	3167.0	0.334852	0.321618	0.343530	Blue
13	3063	1039	946	1078	9200	3111.0	2919.0	3170.0	0.338152	0.317283	0.344565	Blue
14	3089	1042	1001	1046	9241	3116.0	2935.0	3190.0	0.337193	0.317606	0.345201	Blue
15	3153	1042	1041	1070	9305	3123.0	2988.0	3194.0	0.335626	0.321118	0.343256	Blue
16	3084	1054	961	1069	9326	3138.0	3003.0	3185.0	0.336479	0.322003	0.341518	Blue
17	3206	1085	1095	1026	9443	3181.0	3097.0	3165.0	0.336863	0.327968	0.335169	Yellow
18	3259	1106	1044	1109	9549	3245.0	3100.0	3204.0	0.339826	0.324641	0.335533	Blue
19	3290	1128	1099	1063	9755	3319.0	3238.0	3198.0	0.340236	0.331932	0.327832	Orange

APPENDIX B

```
                               SARIMAX Results
==============================================================================
Dep. Variable:             TotalOrange   No. Observations:                 1000
Model:                  ARIMA(5, 1, 0)   Log Likelihood               -5570.594
Date:                 Fri, 08 Sep 2023   AIC                          11153.189
Time:                         11:17:50   BIC                          11182.629
Sample:                              0   HQIC                         11164.379
                                - 1000
Covariance Type:                   opg
```

```
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          2.0531      0.014    148.404      0.000       2.026       2.080
ar.L2         -0.9213      0.021    -43.865      0.000      -0.962      -0.880
ar.L3         -0.9221      0.016    -56.611      0.000      -0.954      -0.890
ar.L4          1.1278      0.017     66.016      0.000       1.094       1.161
ar.L5         -0.3819      0.008    -46.914      0.000      -0.398      -0.366
sigma2      4061.4586     60.482     67.151      0.000    3942.916    4180.001
===================================================================================
Ljung-Box (L1) (Q):                  27.55   Jarque-Bera (JB):             37048.20
Prob(Q):                              0.00   Prob(JB):                         0.00
Heteroskedasticity (H):               0.49   Skew:                             1.80
Prob(H) (two-sided):                  0.00   Kurtosis:                        32.61
===================================================================================
```