University of Nevada, Reno

# Using species accumulation curves to study change through time in a diverse butterfly fauna along an elevational gradient

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Statistics and Data Science

by

Alejandro Pacheco

Anna K. Panorska, Ph.D., Thesis Advisor

May, 2023

University of Nevada, Reno

## THE GRADUATE SCHOOL

We recommend that the thesis

prepared under our supervision by

## ALEJANDRO PACHECO

entitled

# Using species accumulation curves to study change through time in a diverse butterfly fauna along an elevational gradient

be accepted in partial fulfillment of the

requirements for the degree of

## MASTER OF SCIENCE

Anna K. Panorska, Ph.D., Advisor

Tomasz J. Kozubowski, Ph.D., Committee Member

Matthew Forister, Ph.D., Committee Member

Frederick C. Harris, Jr., Ph.D., Graduate School Representative

Markus Kemmelmeier, Ph.D., Dean, Graduate School

May, 2023

# ABSTRACT

The motivation for this thesis comes from ecological questions about the variability in the population dynamics of butterfly species across geographically diverse locations within Northern California. The goal of this thesis can be summarized in the following parts: i) to parameterize and fit the skewed log-logistic model to the observed species accumulation curves at each location; ii) to develop associations between the estimated parameters of the accumulation curves (response) and the weather variables (predictors) for each site, and iii) to analyze the fit of the models and interpret the findings in statistical and ecological terms. Ten locations were analyzed. Annual butterfly species data were available for the period from 1973 to 2016 with small site to site variation. Weather variables considered for the models were seasonal and annual precipitation totals and maximum/minimum seasonal temperatures. We found that a majority of inter-annual variation in weather was explained by variation in precipitation. Associations between the parameters of the species accumulation curves were modeled using linear and polynomial regression tools. Fit was assessed using the mean squared error. Models were developed using each SLL parameter as the response variable and the seasonal weather variables as the predictors. Through the use of step-wise regression model selection, an optimal model was developed from this initial model for each site analyzed. All computing work was done using R software.

# ACKNOWLEDGEMENTS

I would like to offer my sincere gratitude to my thesis advisor, Dr. Anna Panorska, Professor, Department of Mathematics and Statistics: the University of Nevada Reno, for all of the support throughout each step of the completion of my MS thesis. Dr. Panorska's excellent guidance, outstanding patience, and profound knowledge of Statistics greatly contributed to my success in this research. I would also like to thank Dr. Matthew Forister for his incredible insight and guidance regarding the ecological aspects of this research. His incredible knowledge was crucial in conducting this analysis in an efficient manner. In addition I would also like to thank my committee members: Dr. Matthew Forister, Dr. Anna K Panorska, Dr. Tomasz Kozubowki, and Dr. Frederick Harris, Jr. It was an unbelievable experience and privilege to work and study under their guidance.

I am beyond grateful to my supportive parents Consuelo and Luis Pacheco for their love and inspiration, for without them I would not be where I am today.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

## 1.1 Organization of the thesis.

This thesis is organized as follows. Chapter 1 contains the introduction. Chapter 2 provides information on the weather and butterfly data that were used for analysis in this project. Chapter 3 demonstrates the statistical methodology that was employed to conduct the analysis in the research. Lastly, chapter 4 provides our analysis and findings. The findings are divided into the following three sections. Section 1 demonstrates the impact in changes of the parameters on the skewed log-logistic species accumulation curves. Section 2 explains (1) the approach taken for computing the parameters of the SLL curves (nls.LM R function in R), (2) the details of the fitted species accumulation curves for each of the sites researched, and (3) analysis of the goodness of fit. Section 3 contains our results on modeling the relationships between the SLL parameters and the weather variables for each site. The final chapter is devoted to a summary and discussion of findings, limitations and future research directions.

## 1.2 Background: Weather and butterflies

To gain an understanding of the annual changes in weather and the response of a certain species to these changes is a fundamental area of focus in ecology. With that goal in mind, we examine a location over a long period of time to observe and understand the variation in weather and how it affects the wildlife at that location.

There are several aspects of weather variability to consider when conducting an analysis of this type for different geographic regions. For example, seasonal precipitation may be more impactful for a dryer location, whereas high variations in temperature can be more significant for other locations. By examining the changes in weather variables across ten different locations, we can better understand the impact of these changes on insects, specifically the various butterfly species that inhabit these locations. According to Forister et al. (2018) there are both positive and negative effects on the insect populations when conditions are warmer and drier. Observing a location over a long time period allows us to record data on the variability of weather and the corresponding changes in the observed cumulative number of butterfly species. From previous studies at these locations we understand that a majority of populations at the lower elevation sites have been in decline since the mid-1990s, while the populations at the higher elevation sites have remained relatively unchanged over time, with the exception of a few species (Forister et al., 2018). This variation in the change of the number of species at high and low elevation causes challenges for the analysis and interpretation of the relationship between changes in weather and the corresponding change in the number of butterfly species at a certain location. There are also other, often unique, factors impacting the change in the number of species of butterflies at a given location. Due to the varying importance of different weather variables at each location, the challenge arises in creating a common model that fits for every location being studied. Therefore, it will be equally beneficial to understand the individual models for each location in relation to the species being studied and determine any significant relationships between our models and the weather variables.

Once we have an understanding of the weather data and the variables to be used in our analysis, we can shift our focus to the other crucial information: the butterfly species data. For over 45 years (1973 - 2016), 10 study sites, in the United States across Northern California, have been monitored and surveyed by Dr. Art Shapiro (UC Davis) for the number of species of butterflies that call these locations home. The objective of surveying these sites was to gain a better understanding of how insects respond to a changing world within a region that encompasses much of the climatological, geologic, and floristic diversity of this area (Forister et al., 2010). These long term surveying efforts resulted in data on the number of butterfly species dating back (in certain locations) to the 1970's. These data make studies like ours possible. The data we have obtained is comprised of the various species of butterflies observed at a given location each year at each visit. For each site the same fixed path was monitored biweekly during suitable insect conditions to record the number of unique butterfly species observed on that given day. From these data sets we are able to compute the cumulative number of butterfly species that have been observed at a given site on a given observation day. Overall, we have the cumulative number of species dating back 45 years in certain locations. Our analysis begins with fitting skewed log-logistic model (SLL) to the species accumulation curves. Fitting these models means estimating model parameters. Once the parameters for each year are estimated, we analyze their (potential) relationships with the weather data. This step aims at better understanding the impact of weather patterns on the number of butterfly species at a given location. We now turn to the species accumulation curves and the SLL model.

## 1.3 Species accumulation curves

Using species accumulation curves is a common practice in understanding the diversity of a given species in a specific region of study. A species accumulation curve shows the cumulative number of species observed at a given site at a given point in time.

This allows researchers to assess and compare diversity across populations or to evaluate the need/benefits of additional sampling (Deng et al., 2015). There have been various models proposed for these curves, such as the general compound Poisson mixture model used in Deng et al. (2015). In our research we apply the SLL model, which has been used in epidemiology (Giri, 2021) and other fields. Using this method allows us the flexibility of having several parameters to use in our analysis when determining a relationship with the weather variables. The SLL model was built in Sastry and Bhati (2016). The development of this model and its extended uses, including the estimation of parameters, are discussed in Arellano-Valle et al. (2005), Fernandez and Steel (1998) and Sastry and Bhati (2016). We will expand further on the skewed log-logistic model in chapter 3, Methodology.

To create the annual butterfly species accumulation curves, we implemented the skewed log-logistic model for each year of observations at a given site. The benefit of this model pertains to the number of parameters used to create each curve allowing us to find potential relationships between each of these parameters and the various weather variables. The SLL parameters are as follows: skewness coefficient, scale parameter, inflection point, and carrying capacity. We begin our analysis with developing an understanding of the change in the SLL curve caused by the change in each parameter (while all other parameters remain fixed). Our analysis of these parameters can be found in section 1 of chapter 4, Analysis and Findings.

The fitting of the SLL curves to the cumulative number of species data were completed using the R package **minpack.lm** and specifically the **nls.LM** function. The **nls.LM** function estimates parameters of a given function based on the data. We fitted an SLL curve to each year of species accumulation. An important point to note here is that the accuracy of these parameter estimates relies greatly on the amount of data we have for a given year. A more sparse data set (fewer site visits) for a given year results in a fitted SLL curve being less accurate. We evaluate the goodness of fit of the fitted SLL curves using the mean squared error and mean absolute error for each year. Our process of computing and verifying the fitted curves for each year is presented in section 2 of chapter 4.

Each fitted SLL curve has 4 parameters which are then used in the analysis of a potential association to the weather variables. Our intention is to discover a single parameter that will effectively provide a relationship with a varying combination of weather variables across all of the sites in our analysis. Due to the uniqueness of each site and how weather effects each region, we will begin by creating individual models for each site. Once the models are completed we will then analyze them for any patterns or relationships between the models themselves. Our findings for each of the sites can be found in section 3 of chapter 4.

Figure 1.1: Sample species accumulation curve

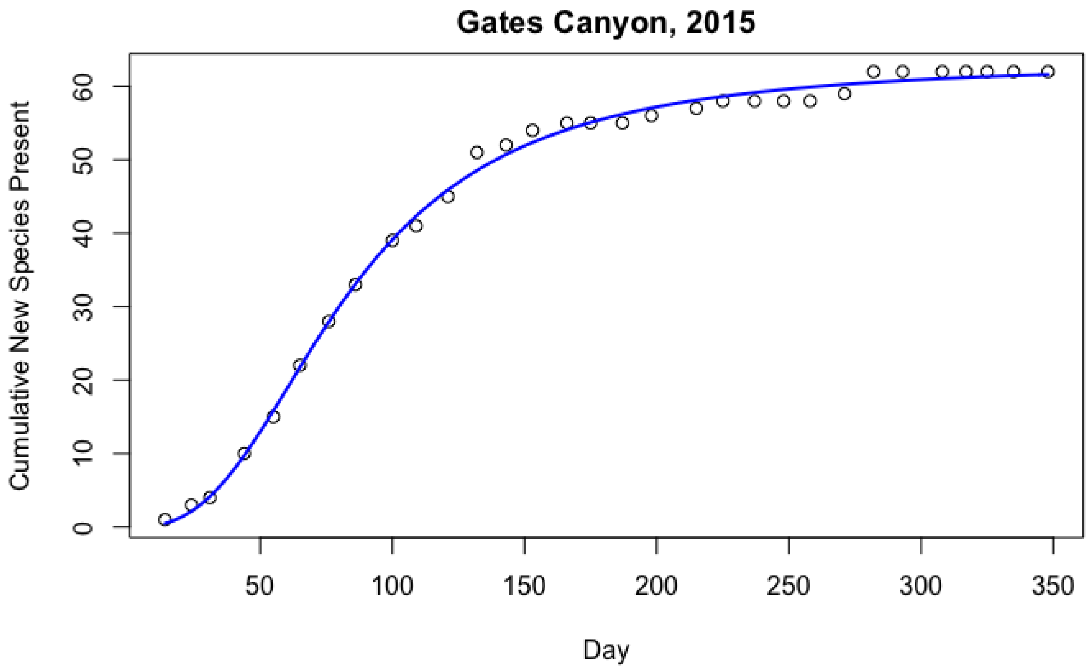



Figure 1.1 is an example of the species accumulation curve utilized within this thesis.

## 1.4 Weather variables and the cumulative number of butterfly species

We took into consideration various weather variables including precipitation and temperature data when exploring a relationship between the yearly cumulative number of butterfly species and the impact these weather variables have upon them. The significance of understanding this relationship according to Parmesan (2000) is that more detailed climatological analyses to biological processes can aid in the identification of the impacts of specific facets of climate on natural systems. The challenge arises in the uniqueness of each location due to many factors including varying habitats, plant communities, micro-climates, elevation and topography, just to name a few, and the way weather variables may affect the number of butterfly species at each location. For the purposes of this thesis we focus on studying possible relationships between the weather variables and their impact on the number of species of butterflies observed at each location. Each of the additional factors could be an area of study in itself when partitioning the sites and analyzing changes for a specific species. For our purposes we will differentiate between sites using only their elevation and classifying each site by either high or low elevation. We describe the weather variables we used as well as the SLL parameter we selected which was the parameter showing the strongest relationship to the weather variables for each site studied. Our analysis and findings are described in their entirety in chapter 4.

## 1.5 Objectives of this thesis

The objectives of this thesis can be summarized as follows.

- To develop an understanding of the impact of the change in the parameters on the skewed log-logistic species accumulation curve.
- To develop (parameterize) species accumulation curves and quantify their goodness of fit.
- To identify any possible association between the SLL accumulation curves' parameters and the weather data for each site researched in this analysis.

# Chapter 2

# Data

This chapter is committed to the description of the sources of data we used for this analysis. There are two primary types of data that we compiled.

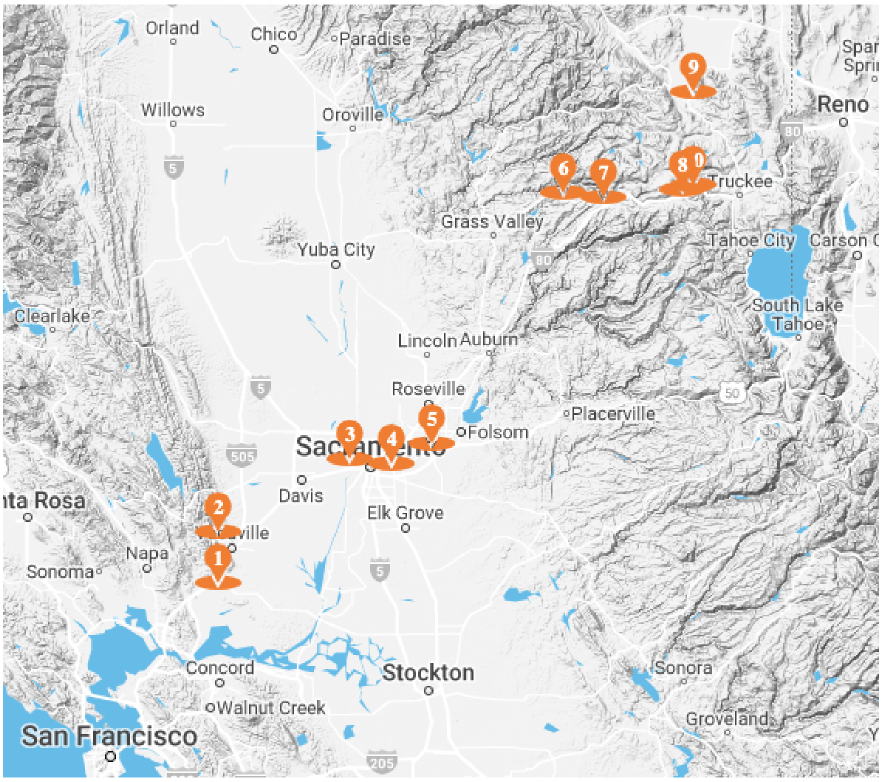
- Data on yearly butterfly species counts for each site.
- Yearly weather data for each site under study.

## 2.1 Butterfly data.

The butterfly data for this thesis was collected during biweekly visits by Dr. Art Shapiro (UC Davis) to the ten study sites over a period of 29 - 45 years, depending on the site, when the weather conditions were suitable for observing butterflies. For each site, a fixed walking route was taken and the various species of butterflies were observed and noted. This data can be found at https://sites.google.com/view/westernbutterflies/data. The data set for each site contains the visit date, butterfly species and ordinal day of visit for that year. This raw data set must be converted to a different format for further analysis. The conversion was done using a python script to track the number of unique species observed per ordinal day for each site in a given year. With the number of unique species observed, the script maintains a running total of the accumulated number of unique species per ordinal day. From this transformation, our data set now contains the following variables: year, ordinal day, number of new species, and cumulative number of new species. The ordinal day represents the day the site was visited to record butterfly data. The number of new species is the number of unique species recorded on that ordinal day. Finally, the cumulative number of new species is the running total of unique species recorded for the given year. This transformed data set will be used in further analysis. In particular, these data will be used to fit the species accumulation curves with ordinal day running across the x-axis and cumulative number of new species on the y-axis. The sites that will be analyzed are as follows:

- **High Elevation Sites:** Castle Peak, Donner Pass, Lang Crossing, Sierra Valley, Washington.
- **Low Elevation Sites:** Gates Canyon, Rancho Cordova, West Sacramento, North Sacramento, Suisun Marsh.

Figure 2.1: Map of sites used in the analysis




The map in Figure 2.1 depicts the study sites for this research. 1: Suisun Marsh. 2: Gates Canyon. 3: West Sacramento. 4: North Sacramento. 5: Rancho Cordova. 6: Washington. 7: Lang Crossing. 8: Donner Pass. 9: Sierra Valley. 10: Castle Peak.

## 2.2 Weather data

The second data set for this study is comprised of the weather variables for each of the ten sites spanning a time period of 29 - 45 years. The data were compiled using the PRISM system (Parameter-Elevation Relationships on Independent Slopes Model, PRISM weather Group; http://prism.oregonstate.edu) by utilizing the latitude and longitude coordinates at the center of each site route. The weather variables are as follows:

- **Max temperature:** Average daily maximum temperatures per site measured in degrees Celsius. This variable is recorded on a seasonal basis and as a result there are four separate variables per year (spring, summer, fall and winter).

- **Min temperature:** Average daily minimum temperatures per site measured in degrees Celsius. This variable is recorded on a seasonal basis and as a result there are four separate variables per year (spring, summer, fall and winter).

- **Precipitation:** Precipitation per site per year was compiled on a seasonal and annual basis measured in millimeters (mm). As a result we have spring, summer, fall, winter, and annual precipitation.

We have obtained the annual weather data from the PRISM system for each site corresponding to the available butterfly data for each site. As a result, the time period being analyzed varied from site to site depending on the availability of the butterfly data. For the sites of North Sacramento, West Sacramento and Washington, the years analyzed will be 1988 - 2016. Castle Peak will be analyzed for the time period of 1989-2016. Donner Pass will be analyzed from 1973 to 2016, Lang Crossing from 1974 to 2016, Rancho Cordova from 1975 to 2016, Suisun Marsh from 1973 to 2016, Sierra Valley for 1982-2016 and Gates Canyon for 1976 - 2016.

As an example, Table 2.1 shows our weather data for the Castle Peak site from 1989 to 2016. The values of the variables were rounded to the nearest tenth. The following abbreviations were used to fit the table on one page. S: spring, SS: summer, F: fall, W: winter, t-: minimum temperature, t+: maximum temperature, ppt: precipitation.

Table 2.1: Weather variables

| Year | t+ S | t+ SS | t+ F | t+ W | t- S | t- SS | t- F | t- W | ppt S | ppt SS | ppt F | ppt W | ppt ann. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1989 | 8.8 | 20.7 | 15.1 | 1.8 | -3.7 | 4.4 | 0.2 | -10.7 | 132.4 | 34.0 | 98.1 | 93.8 | 1397.1 |
| 1990 | 8.9 | 21.4 | 13.4 | 3.4 | -3.9 | 5.3 | -1.2 | -10.6 | 83.4 | 10.3 | 101.4 | 98.6 | 926.8 |
| 1991 | 5.2 | 20.5 | 14.9 | 4.3 | -6.0 | 4.6 | -0.6 | -10.1 | 154.8 | 30.6 | 25.8 | 44.3 | 1285.4 |
| 1992 | 10.8 | 20.8 | 15.9 | 3.8 | -2.9 | 5.1 | 0.8 | -8.1 | 49.1 | 35.3 | 52.7 | 108.7 | 1182.9 |
| 1993 | 7.6 | 19.4 | 14.5 | 0.3 | -4.1 | 3.8 | -0.4 | -10.3 | 81.6 | 18.9 | 38.8 | 313.0 | 1627.1 |
| 1994 | 8.9 | 22.2 | 14.0 | 3.1 | -4.3 | 4.7 | -1.1 | -9.4 | 53.5 | 1.8 | 53.9 | 106.2 | 1172.2 |
| 1995 | 5.5 | 19.7 | 12.2 | 3.2 | -4.6 | 4.7 | -1.4 | -7.0 | 230.0 | 29.9 | 93.7 | 228.3 | 2184.4 |
| 1996 | 7.9 | 22.3 | 16.1 | 4.1 | -4.5 | 6.0 | 0.2 | -6.5 | 131.7 | 13.8 | 5.7 | 284.2 | 2613.7 |
| 1997 | 9.5 | 19.7 | 12.9 | 3.4 | -3.7 | 5.1 | -0.3 | -7.0 | 31.5 | 28.4 | 77.8 | 286.0 | 1446.0 |
| 1998 | 4.5 | 19.8 | 13.0 | 2.0 | -5.3 | 5.1 | 0.4 | -7.6 | 137.2 | 19.0 | 80.7 | 260.1 | 2161.3 |
| 1999 | 6.0 | 19.7 | 11.5 | 1.8 | -5.2 | 4.6 | 0.6 | -8.8 | 74.4 | 18.4 | 96.5 | 182.4 | 1514.2 |
| 2001 | 9.1 | 21.4 | 11.6 | 2.7 | -3.9 | 5.8 | -1.0 | -9.0 | 52.6 | 3.2 | 58.4 | 96.9 | 1368.0 |
| 2002 | 7.2 | 21.3 | 14.3 | 3.1 | -4.8 | 6.1 | 0.7 | -8.3 | 84.9 | 4.8 | 81.8 | 167.2 | 1388.3 |
| 2003 | 6.5 | 21.8 | 13.3 | 4.0 | -3.9 | 7.1 | 0.0 | -6.7 | 112.4 | 27.0 | 54.0 | 179.0 | 1511.4 |
| 2004 | 9.6 | 21.3 | 13.8 | 2.3 | -2.8 | 6.8 | 1.0 | -6.9 | 38.2 | 3.8 | 33.2 | 216.7 | 1262.3 |
| 2006 | 5.9 | 21.9 | 12.8 | 3.5 | -3.8 | 6.9 | 0.4 | -7.1 | 171.0 | 1.8 | 37.9 | 295.3 | 1859.5 |
| 2007 | 9.0 | 21.9 | 12.6 | 2.4 | -3.5 | 6.8 | 0.1 | -9.5 | 47.8 | 4.8 | 31.6 | 132.3 | 1117.1 |
| 2008 | 7.0 | 21.5 | 12.5 | 1.5 | -4.6 | 6.6 | 0.3 | -8.6 | 19.2 | 0.3 | 37.6 | 163.0 | 1252.8 |
| 2009 | 7.8 | 20.8 | 14.1 | 2.9 | -3.6 | 6.1 | 1.0 | -8.6 | 95.3 | 12.8 | 48.4 | 135.3 | 1513.1 |
| 2010 | 5.6 | 20.6 | 13.3 | 1.8 | -5.3 | 5.5 | 0.6 | -7.6 | 121.2 | 4.0 | 48.7 | 177.9 | 2156.8 |
| 2011 | 5.3 | 20.0 | 13.3 | 2.9 | -4.7 | 5.1 | 0.9 | -7.9 | 131.3 | 22.6 | 116.1 | 204.2 | 1457.1 |
| 2012 | 7.9 | 21.5 | 13.6 | 4.2 | -3.9 | 5.7 | 0.7 | -9.5 | 77.8 | 4.9 | 40.7 | 90.9 | 1846.2 |
| 2013 | 9.5 | 21.9 | 14.6 | 2.7 | -3.1 | 6.5 | 1.4 | -9.6 | 46.1 | 21.9 | 73.5 | 136.0 | 561.8 |
| 2014 | 9.5 | 22.0 | 13.4 | 4.8 | -2.8 | 6.8 | 0.2 | -8.1 | 55.4 | 15.3 | 36.5 | 86.1 | 1384.8 |
| 2015 | 9.5 | 22.2 | 14.9 | 6.1 | -2.9 | 7.3 | 2.1 | -5.6 | 50.0 | 17.6 | 52.4 | 161.8 | 1078.7 |
| 2016 | 9.0 | 22.0 | 14.2 | 4.1 | -2.7 | 6.0 | 1.4 | -7.0 | 125.1 | 6.5 | 113.3 | 177.8 | 2022.8 |

# Chapter 3

# Methodology

In this chapter, we will provide information about the model chosen to create the annual species accumulation curves, the skewed log-logistic model. Our fitting strategy follows the methodology used in Giri (2021) for cumulative number of COVID-19 cases. However, for our purposes we have adapted this methodology to be used in fitting the accumulated number of butterfly species. We will provide information on how the parameters of the skewed log-logistic model interact as well as the effects on the overall curve with changes to each parameter. In addition, we will discuss the R script utilized to fit the skewed log-logistic curve. Lastly, we will provide a brief discussion of the goodness of fit of each curve using the mean squared error and mean absolute error.

## 3.1 Modeling the accumulated number of species

In the field of ecology, a primary goal is to gain an understanding of how many species of a certain type of taxon occur in a given area or areas of study. Within the study of species accumulation curves there have been various models used for analysis. In Deng et al. (2015), a general compound Poisson mixture model was used for the sampling process. In Good and Toulmin (1956) a general multinomial model was used to predict the accumulation curves. In a comparison of a climate index and weather variables on population growth, the climate index-based model was used by Matter and Roland (2017). To demonstrate the drying process of corn over time, Siatkowski et al. (2010) used several growth curve functions including: exponential, Gompertz, logistic, modified logistic, log-logistic, and Weibull. What is unique about this thesis is the use of the skewed log-logistic curve in modeling the cumulative number of butterfly species. There are several factors to consider when analyzing populations in a given area over time, such as resources or density dependence, that can affect the growth of a species in that area (Ugland et al., 2003). For the purposes of this research, we will use the SLL model already used in epidemiology for the cumulative number of COVID-19 cases. The SLL model is an S-shaped asymmetric curve which is similar to a cumulative probability distribution function for continuous random variables (Giri, 2021). There are three focus areas when analyzing these adjusted curves: the initial increase, the inflection point at which the total number of species begins to level off, and the data after the point of inflection (Vandermeer, 2010).

While the model provides us with information on the accumulated number of butterfly species for a given year and location, it does not provide us with any information regarding the causes of growth or decline for this period of time. What this model does provide however are the parameters used to create the SLL curve, which we will use to find a potential relationship with the weather variables. The number of parameters in this model provides us with a higher number of response variables to be able to find optimal models for each site. Within the studies of Gottschalk and Dunn (2005), they found that a 5 parameter logistic function can dramatically improve their accuracy of asymmetric assays. Using the SLL curve parameters provides an opportunity to find a potential relationship between the parameters of the species accumulation curve and the weather variables. As we have previously stated, the difficulty in this type of research is that each location is unique and as a result the challenge is determining a relationship between all of the models developed for different locations. However, thanks to the number of parameters being used to develop the SLL curves we have more data to use in the development of these models. While studies on this butterfly data have been done in the past in (Nice et al., 2019) and studies conducted on various other butterfly data sets (Nieto-Sanchez, 2015), the driving motivation behind this thesis is the use of the skew log-logistic model upon a butterfly data set.

## 3.2 The skew log-logistic model

In this section we will present the parameters utilized in the skewed log-logistic model and the method for modeling the species accumulation curves, the methodology follows Giri (2021). The equation for the skewed log-logistic curve is presented in eq. (3.1):

$$F(t) = d\left\{\frac{c}{d} + \left(1 - \frac{c}{d}\right) G(\log t - \log \mu)\right\}, \tag{3.1}$$

In eq. (3.1), $G(x)$ is the cumulative distribution function (CDF) of a logistic distribution with $\sigma$ as the scale parameter . $G(x)$ has the following equation:

$$G(x) = \frac{1}{1 + e^{-\sigma x}}, \quad x \in \mathbb{R}. \tag{3.2}$$

This logistic distribution with the CDF $G(x)$ in eq. (3.2) is symmetric around x = 0, therefore its probability density function (PDF) $g(x)$ given in eq. (3.3) is symmetric around zero. The PDF $g(x)$ is provided below in eq. (3.3):

$$g(x) = \frac{\sigma e^{-\sigma x}}{[1 + e^{-\sigma x}]^2}, \quad x \in \mathbb{R}. \tag{3.3}$$

Next, we consider the function $F(t)$ in eq. (3.1) with $G(\cdot)$ provided in eq. (3.2). $F(t)$ is a symmetric generalized logistic curve. In order to add skewness to the generalized logistic model, we have multiple methods at our disposal. We continue with the method provided in Fernandez and Steel (1998), with $k > 0$ and the inverse $1/k > 0$, are used to derive a skewed PDF $\tilde{g}(\cdot)$ eq. (3.4) from the symmetric $g(\cdot)$ .

$$\tilde{g}(x) = \frac{2\kappa}{\kappa^2 + 1} \begin{cases} g(x\kappa) & \text{for } x \geq 0, \\ \\ g\left(\frac{x}{\kappa}\right) & \text{for } x < 0. \end{cases} \tag{3.4}$$

The PDF $\tilde{g}(x)$ and CDF $\tilde{G}(x)$ are presented in eq. (3.5) and eq. (3.6).

$$\tilde{g}(x) = \frac{2\sigma\kappa}{\kappa^2 + 1} \begin{cases} \frac{e^{-\sigma\kappa x}}{[1+e^{-\sigma\kappa x}]^2} & \text{for } x \geq 0, \\ \\ \frac{e^{-\sigma x/\kappa}}{\left[1+e^{-\sigma x/\kappa}\right]^2} & \text{for } x < 0, \end{cases} \tag{3.5}$$

and

$$\tilde{G}(x) = \begin{cases} \frac{\kappa^2-1}{\kappa^2+1} + \frac{2}{\kappa^2+1}\frac{1}{1+e^{-\sigma\kappa x}} & \text{for } x \geq 0, \\ \\ \frac{2\kappa^2}{\kappa^2+1}\frac{e^{-\sigma x/\kappa}}{1+e^{-\sigma x/\kappa}} & \text{for } x < 0. \end{cases} \tag{3.6}$$

The analysis of the skewed generalized logistic distribution is provided by Arellano-Valle et al. (2005) and Sastry and Bhati (2016). The special case when $k = 1$ is the symmetric generalized logistic model. We can then replace $G(\cdot)$ in eq. (3.1) with the $\tilde{G}(x)$ to obtain the skewed logistic growth curve provided in eq. (3.7). This methodology was presented in Giri (2021) and has been adapted for the use in the species accumulation curves of butterflies.

$$F(t) = d\left\{\frac{c}{d} + \left(1 - \frac{c}{d}\right)\tilde{G}(\log t - \log\mu)\right\}. \tag{3.7}$$

The parameters of interest in eq. (3.7) are defined below:

- **Parameter $c$:** Initial number of species of butterflies per year per site. For our purposes this research $c$ will be set to zero at the beginning of each year.
- **Parameter $d$:** The carrying capacity $d$ represents the maximum total number of species per site.
- **Parameter $\mu$:** The inflection point of the skewed log-logistic growth curve. It is the day, where the total number of observed new species is maximum, so that the rate of change of the cumulative number of species changes and the number of new species observed on subsequent days begins to decrease for that year.
- **Parameter $\sigma$:** The scale parameter. This parameter provides information about the scale of the growth curve.
- **Parameter $k$:** (hidden in the formula for $\tilde{G}$). Skewness parameter. This parameter provides information about the yearly rates of growth of the accumulation of butterfly species before and after the inflection point.

We would like to note that the CDFs $G(\cdot)$ and $\tilde{G}(\cdot)$ may differ in symmetry and in the calculated value of the function at the point of inflection. However, the point of inflection for both equations $G(\cdot)$ and $\tilde{G}(\cdot)$ is $x = 0$, we see that in eq. (3.8) $G(0) = 1/2$, but

$$\tilde{G}(0) = \frac{\kappa^2}{\kappa^2 + 1}, \tag{3.8}$$

this results in any value within the interval $(0, 1)$. We also note that, while the derivative of $G(\cdot)$ is symmetric at the point of inflection, the derivative of $\tilde{G(\cdot)}$ is not. Thus, we have eq. (3.9)

$$\tilde{g}(x) = \frac{d}{dt}\tilde{G}(t)\bigg|_{t=x} = \tilde{g}\left(-x\kappa^{\text{sign}(x)}\right) = \frac{d}{dt}\tilde{G}(t)\bigg|_{t=-x\kappa^{\text{sign}(x)}}, \tag{3.9}$$

here $sign(x)$ is the *signum* function, which is equal to 1 for $x < 0$, $-1$ for $x > 0$, and zero for $x = 0$. As a result this curve can therefore be increasing at a rate that is faster prior to the point of inflection than the rate it decreases after the inflection point and vice versa which allows for more flexibility in our curve.

In order to estimate the parameters of the skewed log-logistic curve given by eq. (3.7) we call upon an R function nls.lm, which fits a non-linear curve using the Levenberg-Marquardt algorithm. This R function selects parameters that minimize the mean squared error between the SLL model and the provided data. To execute the nls.lm function we must first provide positive initial values for all parameters with their lower bounds set to zero.

# Chapter 4

# Analysis and Findings

In this chapter we will present our analysis and findings. This chapter will be divided into the following sections:

- Section 4.1 will demonstrate the impact of each parameter on the skewed log-logistic species accumulation curve.
- Section 4.2 will provide the fitted SLL curves as well as an analysis of the measure of fit between the observed data and the SLL curve for each year at each site.
- Section 4.3 provides an analysis of the relationships between the SLL parameters and the weather variables for each of our research sites.

In this first section we will analyze the impact of the value of each parameter on the SLL (skewed log-logistic) species accumulation curve. We will then proceed to fit the SLL species accumulation curves for each of the sites for each of the years with data. For each site we present yearly plots of the species accumulation curve. The plot will contain the empirical number of species observed as well as the fitted accumulation curve for that year. In order to fit the curve we will call upon the nls.LM function from R's minpack.lm package. In each plot we shall report the estimated parameters of the skew log-logistic model. To provide evidence of the goodness of fit, we will also provide the mean squared error and mean absolute error for each plot.

Next, we will describe the weather variables that were chosen to compare with our SLL parameters and their relationships for each site. Due to the geographical complexity of each site, various combinations of weather variables will be used in the search for the optimal model for each site. Finally, we will analyze the association between our SLL parameters and the weather variables chosen in the optimal model derived for each site.

## 4.1 Impact of the parameters on the SLL species curve

We will now describe the dependence of the parameters on the skewed log-logistic curve and the methods used for inputting the data based initial values as our parameters.

### 4.1.1 Carrying capacity ($d$)

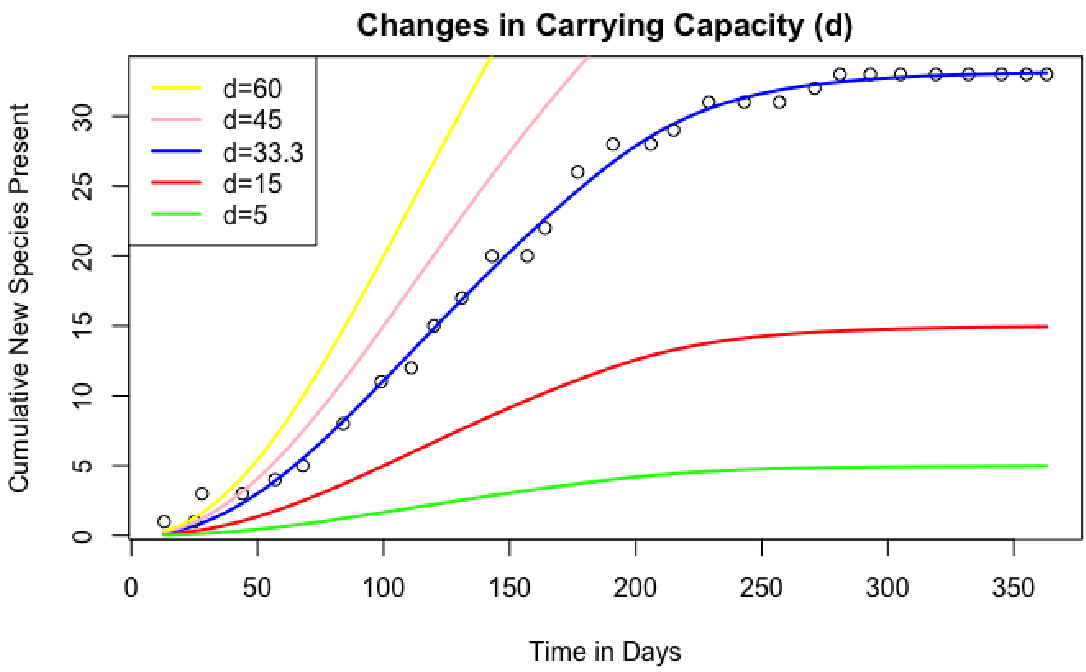



Figure 4.1: Change in the species accumulation curve with respect to the carrying capacity ($d$)

The carrying capacity represents the total number of accumulated species recorded at the end of the year. For our purposes this reflects the final number of accumulated species recorded at the site for a given year. Graphically, this parameter is the y-value corresponding to the last day of observation for a given year. While this parameter in theory represents a whole number, due to the use of the nls.LM function this parameter is calculated as an estimate and as a result can be a decimal. It is rounded to the nearest hundredth decimal.

In Figure 4.1 we used the data of the Suisun Marsh site for the year 1999. The plot represents the change in the carrying capacity parameter, $d$, with all other parameters remaining fixed. The skewness $k$ parameter is fixed to 1.758, scale $\sigma$ is fixed to 3.780, the inflection point $\mu$ is 179.910, and the initial value of the carrying capacity (d) is 33.259. It can be observed that with a higher carrying capacity the plot reaches higher values of the cumulative number of species, whereas a smaller carrying capacity results in smaller values of the cumulative number of species.

### 4.1.2 Inflection point ($\mu$)

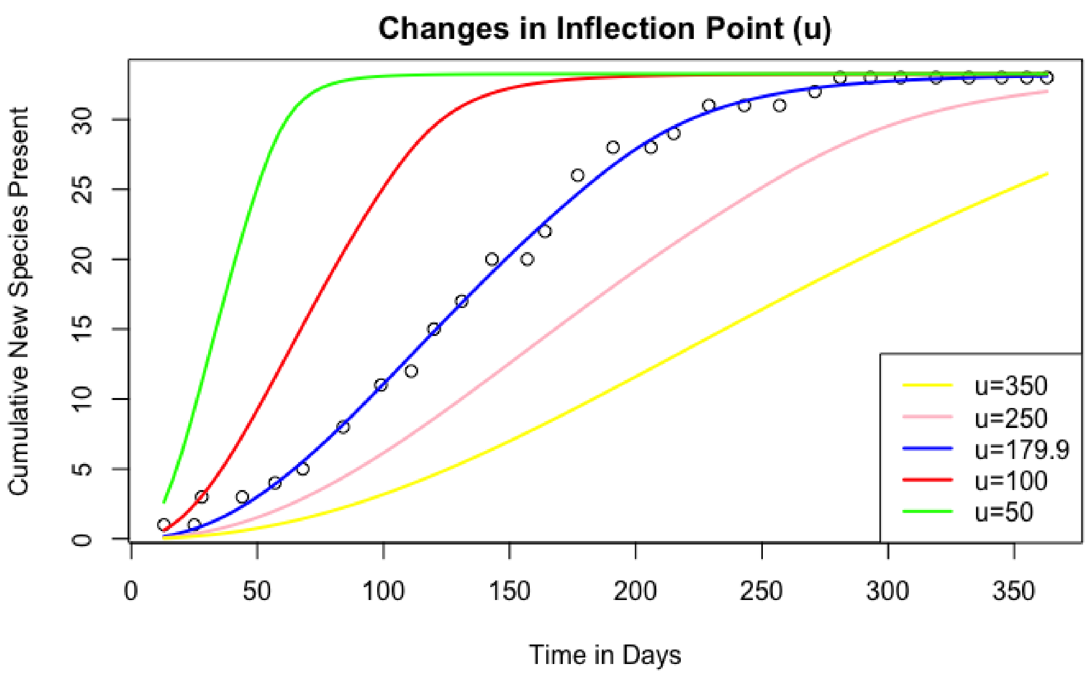



Figure 4.2: Change in the species accumulation curve with respect to the inflection point ($\mu$)

The inflection point of the species accumulation curve is the point where the daily observed number of new butterfly species attains its maximum and the new daily observed number of species begins to decline. Graphically speaking, this is the point on the x-axis where the shape of the curve changes from concave upward to concave downward. This parameter is of particular interest in our analysis as we will attempt to connect it to the weather variables. We will discuss this parameter further in the next section.

In Figure 4.2 we used the data of the Suisun Marsh site for the year 1999. The plot represents the change of the curve in response to the change in the inflection point, $\mu$, with all other parameters remaining fixed. The carrying capacity $d$ is 33.259, the skewness parameter $k$ is 1.758, the scale $\sigma$ is fixed to 3.780, and the initial value of the inflection point $\mu$ is 179.910. From a graphical perspective we can see that the curve is moving towards the left on the x-axis with a decrease in the value of the inflection point.

### 4.1.3 Skewness coefficient $(k)$

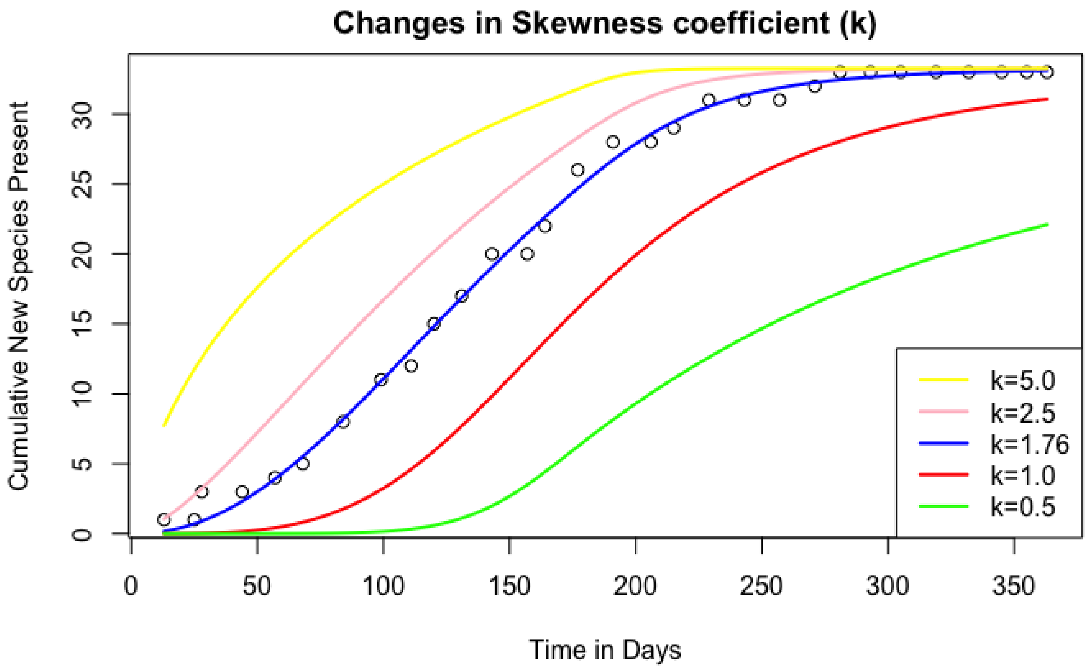



Figure 4.3: Change in the species accumulation curve with respect to the skewness coefficient $(k)$

The skewness coefficient $k$ is one of our parameters of interest in this study as it determines the rate of growth of the cumulative number of species, and can vary before and after the point of inflection. The greater the $k$, the faster the growth before the inflection point and as a result the slower the growth after the inflection point. In the next section we will discuss the importance of $k$ and the dependence of the accumulation curve on this parameter.

In the Figure 4.3 we used the data of the Suisun Marsh site for the year 1999. The plot represents the change in the skewness coefficient parameter, $k$, with all other parameters remaining fixed. The carrying capacity $d$ is 33.259, the scale $\sigma$ is fixed to 3.780, the inflection point $\mu$ is 179.910 and the initial value of the skewness coefficient $k$ is 1.758. As we can see from the plot, the greater the value of the skewness coefficient, $k$, the faster the change from the initial growth to leveling off. Graphically speaking, the larger the k-value, the sharper the curvature of the line will be and vice-versa. With a smaller skewness coefficient the line has a softer curvature when leveling off.

To further understand this parameter we will examine two cases where $k > 1$ and $k < 1$. We will begin this analysis of the impact of $k$ on the rate of the species accumulation curve by looking at the derivative $\tilde{g}(x)$ of the cumulative growth function $\tilde{G}(x)$ which is presented in eq. (3.9). Let us look at the first case and suppose that $k > 1$ and let $x > 0$. Thus, we have

$$\tilde{g}(x) = \tilde{g}(-xk^{sign(x)}) = \tilde{g}(-kx).$$

Note that $\tilde{g}(x)$ is increasing for $x < 0$ and $-kx < -x$, therefore $\tilde{g}(-x) > \tilde{g}(x)$. As a result, when $k > 1$ the increase in daily cumulative number of species prior to the inflection point (maximum daily number of new butterfly species) is faster than the decrease of daily number of new butterfly species after the inflection point. This shows that the rate of number of new butterfly species observed before the inflection point is greater than the speed of leveling off or after the inflection point. We now look at the second case where $k < 1$ and we note that the reverse process is true. The rate of number of new species observed is slower before the inflection point versus after the inflection point.

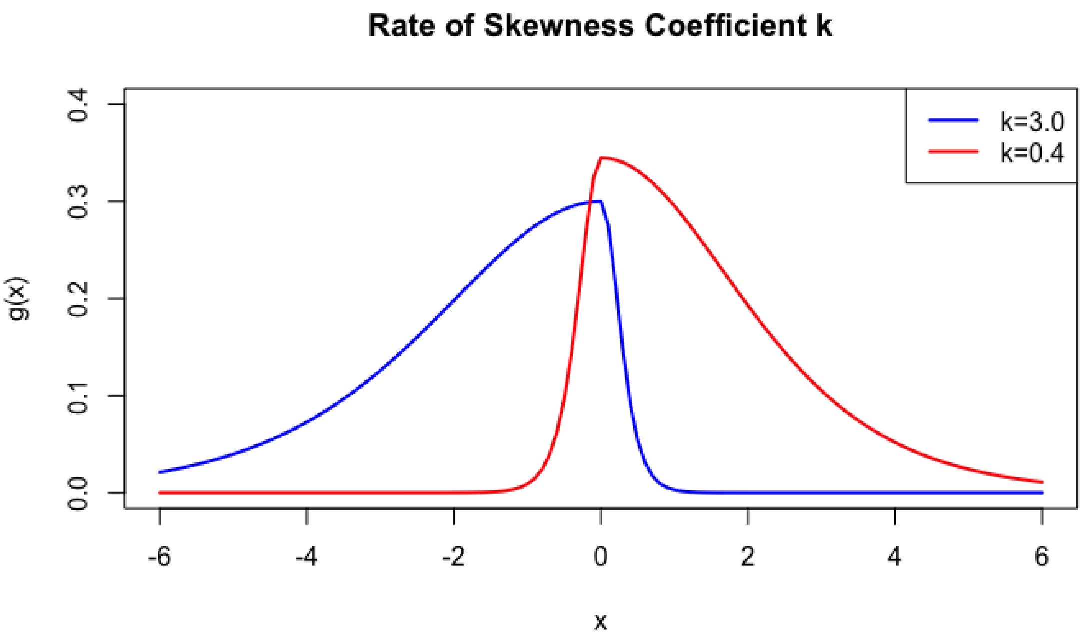



Figure 4.4: Change in rate of skewness coefficient ($k$)

We provide Figure 4.4 to provide a visualization of the two cases and to understand the impact of the skewness coefficient. This plot shows the two cases with $k < 1$ represented in red and $k \geq 1$ represented in blue. This plot demonstrates the changes in the rate of number of new butterfly species observed before and after the inflection point.

### 4.1.4 Scale parameter ($\sigma$)

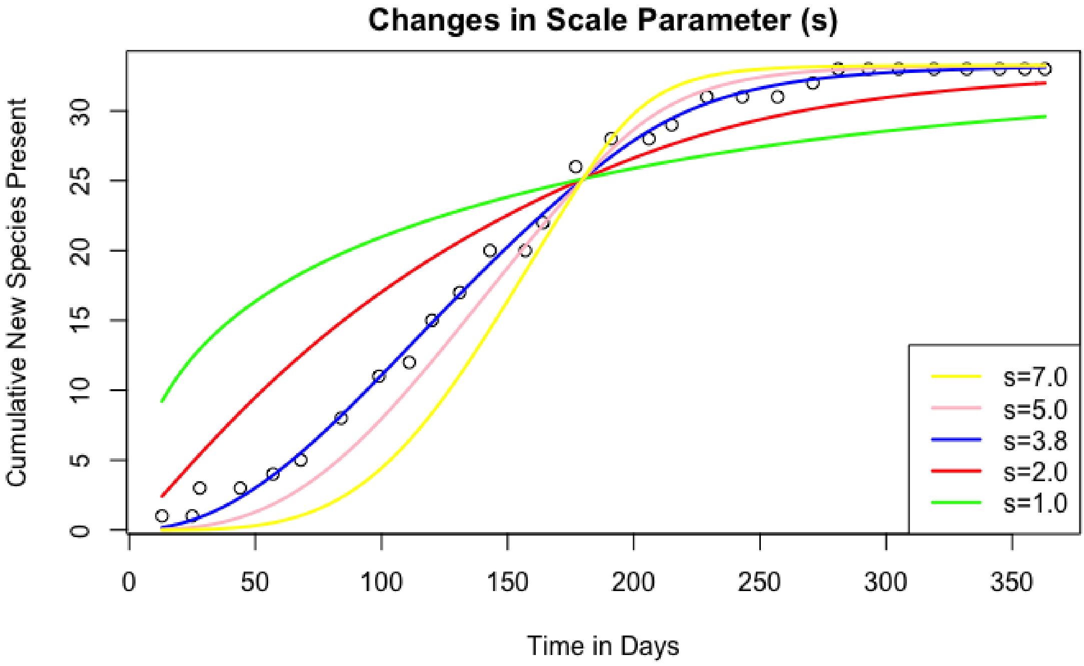



Figure 4.5: Change in the species accumulation curve with respect to scale ($\sigma$)

The scale parameter $\sigma$ controls the scale of the species accumulation curve. This parameter is calculated by utilizing the initial values of the other parameters, $d$, $\mu$, and $k$. This parameter is of significant importance to our analysis, as we will attempt to use it as a possible response variable when modeling relationships with weather variables. Further discussion of this parameter is provided in the next section.

In Figure 4.5 we used the data of the Suisun Marsh site for the year 1999. The plot represents the change in the scale parameter, $\sigma$, with all other parameters remaining fixed. The carry capacity $d$ is 33.259, the skewness coefficient $k$ is fixed at 1.758, the inflection point $\mu$ is 179.910 and the initial value of the the scale $\sigma$ is 3.780. The larger the scale parameter, the steeper vertical climb of the accumulation curve.

## 4.2 Fitting the SLL species accumulation curves

In this section we will discuss the methods and results of fitting the SLL species accumulation curves. We begin with the selection of the initial values used for each parameter in the computation of the nls.lm function in R. The selection of the initial values is crucial in the speed and accuracy of the algorithm returning the estimates of our parameters (Wicklin, 2014). Next, we will discuss the computed parameters for each site for the time period of recorded butterfly data we have for that given site. Finally, we will present the fitted curves for each site for a sample number of years along with a discussion of the goodness of fit of the SLL accumulation curves for each site in terms of the mean absolute error (MAE) and mean squared error (MSE).

### 4.2.1 Computation of initial values

**Initial value for the carrying capacity $d$:** The cumulative number of species on the last day of recorded observation for that year was used as the initial value for carrying capacity.

**Initial value for $\mu$:** In order to find the inflection point of the species accumulation curve, we used the Extremum Distance Estimator (EDE) contained in the R function "ede" to calculate this point for each year and each site, (Christopoulos, 2014).

**Initial value for $k$:** Once the day (x-value) of the inflection point $\mu$ was obtained, we can then compute the corresponding y-value which is the cumulative number of species on that day, $\hat{F}(\mu)$. Finally, we substitute the initial value of $d$ to eq. (3.7) and obtain:

$$\hat{F}(\mu) = \frac{dk^2}{k^2 + 1},$$

here $\hat{F}(\mu)$ and $d$ are known and we can solve for $k$ to obtain eq. (4.1):

$$k = \sqrt{\frac{\hat{F}(\mu)}{d - \hat{f}(\mu)}}. \tag{4.1}$$

**Initial value for $\sigma$.** In order to derive the initial value for $\sigma$ we solved eq. (3.7) for $\sigma$

at each value of $x$ in eq. (4.2):

$$\sigma(x) = \begin{cases} \frac{-log(d(\frac{\frac{2k^2}{k^2+1}}{F(x)})-1)k}{log(x)-log(\mu)} & \text{for } x \geq 0, \\ \\ \frac{-log(\frac{2d}{(F(x)(k^2+1)-(k^2-1)d)}-1)}{k(log(x)-log(\mu))} & \text{for } x < 0. \end{cases} \tag{4.2}$$

From here, we then take the average of the results to obtain our initial value for $\sigma$. The methods and R script for this analysis were developed in Giri (2021) and were adapted and modified to fit our purposes of creating species accumulation curves.

### 4.2.2 Goodness of fit analysis

We measured goodness of fit of the SLL model to the observed number of accumulated species using two methods: the mean squared error (MSE) of the fitted curve and mean absolute error (MAE) of the fitted curve. Using the observed values, $F(t)$, and the fitted, $\hat{F}(t)$, values of the species accumulation curve, the mean squared error was defined as follows:

$$MSE = (1/n)\sum_{t=1}^{n}(F(t) - \hat{F}(t))^2, \text{ where } t = 1, 2, \ldots, n, \tag{4.3}$$

and $n$ is the number of observations (site visits) for a given year.

Using the observed values, $F(t)$, and the fitted, $\hat{F}(t)$, values of the species accumulation curve, the mean absolute error was defined as follows:

$$MAE = (1/n)\sum_{t=1}^{n}\left|F(t) - \hat{F}(t)\right|, \text{ where } t = 1, 2, \ldots, n, \tag{4.4}$$

and $n$ is the number of observations (site visits) for a given year.

In the tables in the next section we provide the computed mean squared error (MSE) and mean absolute error (MAE) for each site for each year. While the range of computed errors varies from site to site, we see an overall relatively good fit for the SLL curves computed for all years and sites. While the MSE is relative for each site, we have chosen to provide the MAE as well to provide another calculation of error which is more consistent when analyzing the fit from site to site. The MAE represents the average absolute value of the difference between observed and estimated (by the SLL curve) cumulative number of species. It is the average absolute error of the cumulative number of species estimated using the SLL curve. By analyzing the goodness of fit for each site, we can see that the use of the SLL curve for fitting our data points provided a reasonably good fit.

### 4.2.3 Fitted species accumulation curves by site

In this subsection, we will present the results of the estimation of parameters used in the fitting the species accumulation curves of the skew log-logistic model for each year of the given site to the butterfly data. For each site, the data is presented on an annual basis for the following parameters: skewness coefficient $k$, scale parameter $\sigma$, carrying capacity $d$, and inflection point $\mu$ with the fitted SLL species accumulation curves. The SLL species accumulation curve model was fit to the data using R's minpack.lm package and the nlsLM function. In table (4.1) we have provided the range of SLL parameters for each site as well as the number of visits per year per site, although this parameter is not used in the creation of the SLL curve, it is relevant in understanding the data and accuracy of modeling for each year. For each location we will provide one SLL curve for each decade of recorded data. Each curve will contain the ordinal number of days along the x-axis and the cumulative number of species along the y-axis. The fitted SLL curve is provided for each plot presented as well as the values of each parameter in the caption under each graph.

Table 4.1: Summary of the range of fitted parameters for all sites

| Site | $k$ | $\sigma$ | $d$ | $\mu$ | Visits |
|---|---|---|---|---|---|
| GC | (0.11, 4.12) | (2.36, 48.12) | (45.19, 66.4) | (38.08, 201.03) | (7, 32) |
| SM | (0.05, 5.35) | (0.92, 38.23) | (29.03, 85.48) | (22.69, 357.67) | (17, 35) |
| WS | (0.37, 9.01) | (1.44, 12.03) | (28.7, 56.18) | (46.98, 469.36) | (18, 35) |
| NS | (0.27, 4.2) | (2.13, 7.81) | (33.06, 45.58) | (49.63, 226.41) | (12, 33) |
| RC | (0.05, 4.2) | (0.75, 46.73) | (31.75, 204.34) | (42.26, 298.3) | (9, 38) |
| W | (0.09, 2.84) | (2.89, 9.86) | (58.4, 87.7) | (44.28, 208.95) | (13, 25) |
| LC | (0.11, 2.56) | (4.89, 71.04) | (47.12, 84.21) | (111.1, 214.24) | (6, 20) |
| SV | (0.05, 3.94) | (4, 29.22) | (44.08, 69.27) | (80.4, 236.62) | (5, 20) |
| DP | (0.05, 1.64) | (6.12, 264.7) | (63.47, 91.73) | (127.03, 194.53) | (7, 23) |
| CP | (0.04, 1.76) | (9.76, 349.19) | (44.13, 79.1) | (123.37, 226.0) | (4, 10) |

The following abbreviations where used for each site in table (4.1) : GC: Gates Canyon, SM: Suisun Marsh, WS: West Sacramento, NS: North Sacramento, RC: Rancho Cordova, W: Washington, LC: Lang Crossing, SV: Sierra Valley, DP: Donner Pass, CP: Castle Peak.

We will now present the selected SLL curves for each site as well as a table that contains the estimated parameters, number of visits, mean squared error (MSE) and mean absolute error (MAE) for all years used in the analysis of each site. We will present the sites in order of geographic location moving from west to east beginning with Gates Canyon and ending with Castle Peak.

**Fitted parameters and the species accumulation curves for Gates Canyon**

Tables 4.2 and 4.2 are comprised of the fitted parameters for the Gates Canyon site as well as the MSE and MAE for each year to demonstrate the goodness of fit for each year. The tables are formatted on a yearly basis displaying the parameters for the years 1976 - 2016. Please note that for the years 1999 and 2006 the nlsLM function was unable to provide an accurate estimate of the skewness coefficient and scale parameters. This is due to repetition of the number of observed species at and around the point of inflection causing no curvature and a sharp plateau. We have chosen to provide one SLL curve from each decade during the observed period for this site to demonstrate various goodness of fit.

Table 4.2: Gates Canyon fitted parameters

| Year | $k$ | $\sigma$ | $d$ | $\mu$ | Visits | MSE | MAE |
|---|---|---|---|---|---|---|---|
| 1976 | 1.99 | 2.36 | 61.94 | 126.64 | 20 | 4.39 | 1.65 |
| 1977 | 0.11 | 48.12 | 45.19 | 75.45 | 7 | 0.79 | 0.53 |
| 1978 | 1.70 | 4.78 | 59.23 | 118.02 | 24 | 0.71 | 0.69 |
| 1979 | 1.81 | 3.76 | 55.77 | 119.21 | 14 | 0.28 | 0.41 |
| 1980 | 1.19 | 3.41 | 57.00 | 105.09 | 16 | 0.69 | 0.67 |
| 1981 | 1.29 | 3.52 | 55.70 | 101.03 | 12 | 0.65 | 0.69 |
| 1982 | 1.47 | 3.41 | 54.64 | 124.24 | 14 | 1.37 | 0.87 |
| 1983 | 2.35 | 5.06 | 52.12 | 158.35 | 16 | 1.85 | 0.95 |
| 1984 | 0.96 | 3.09 | 60.19 | 79.11 | 18 | 1.18 | 0.82 |
| 1985 | 1.68 | 3.69 | 57.63 | 115.95 | 17 | 2.33 | 1.32 |
| 1986 | 1.30 | 3.08 | 57.61 | 103.28 | 18 | 2.49 | 1.08 |
| 1987 | 1.68 | 3.97 | 58.33 | 113.82 | 21 | 6.35 | 1.93 |
| 1988 | 0.95 | 3.36 | 60.41 | 61.14 | 22 | 0.46 | 0.62 |

Table 4.3: Gates Canyon fitted parameters continued

| Year | $k$ | $\sigma$ | $d$ | $\mu$ | Visits | MSE | MAE |
|---|---|---|---|---|---|---|---|
| 1989 | 1.45 | 3.99 | 57.40 | 112.59 | 21 | 3.88 | 1.47 |
| 1990 | 1.66 | 3.31 | 56.72 | 111.04 | 27 | 5.44 | 1.85 |
| 1991 | 0.13 | 8.75 | 66.40 | 38.08 | 23 | 5.86 | 1.63 |
| 1992 | 1.14 | 3.21 | 63.50 | 90.19 | 26 | 1.51 | 0.91 |
| 1993 | 0.11 | 21.40 | 61.29 | 54.32 | 22 | 1.67 | 1.03 |
| 1994 | 2.07 | 4.50 | 61.51 | 129.29 | 21 | 1.23 | 0.95 |
| 1995 | 1.87 | 3.42 | 60.26 | 201.03 | 25 | 0.90 | 0.70 |
| 1996 | 1.58 | 3.45 | 60.94 | 116.03 | 25 | 1.49 | 0.76 |
| 1997 | 1.02 | 4.34 | 61.93 | 83.19 | 28 | 1.57 | 0.82 |
| 1998 | 1.40 | 3.27 | 55.02 | 142.57 | 26 | 1.77 | 0.87 |
| 1999 | - | - | 54.40 | 175.70 | 28 | 5.03 | 1.64 |
| 2000 | 1.60 | 3.51 | 54.76 | 123.46 | 28 | 0.77 | 0.73 |
| 2001 | 1.61 | 3.60 | 59.62 | 117.39 | 25 | 1.92 | 1.04 |
| 2002 | 1.46 | 3.28 | 58.32 | 133.23 | 25 | 3.30 | 1.35 |
| 2003 | 4.12 | 6.54 | 60.55 | 193.15 | 26 | 2.43 | 1.24 |
| 2004 | 0.55 | 5.26 | 58.27 | 66.33 | 27 | 3.24 | 1.13 |
| 2005 | 1.30 | 2.92 | 59.50 | 115.40 | 30 | 2.87 | 1.27 |
| 2006 | - | - | - | - | 27 | 5.44 | 1.69 |
| 2007 | 1.11 | 2.77 | 60.62 | 82.77 | 27 | 2.22 | 1.15 |
| 2008 | 0.97 | 3.26 | 61.13 | 94.19 | 23 | 0.61 | 0.59 |
| 2009 | 1.55 | 2.94 | 58.75 | 147.37 | 27 | 1.72 | 1.04 |
| 2010 | 1.53 | 3.61 | 59.38 | 139.26 | 28 | 0.67 | 0.59 |
| 2011 | 1.82 | 2.90 | 60.24 | 162.72 | 28 | 2.35 | 1.07 |
| 2012 | 1.98 | 2.95 | 59.62 | 147.21 | 32 | 2.58 | 1.24 |
| 2013 | 1.22 | 3.03 | 57.35 | 106.85 | 31 | 0.86 | 0.67 |
| 2014 | 1.83 | 3.00 | 50.69 | 130.89 | 29 | 1.52 | 0.91 |
| 2015 | 0.96 | 2.63 | 63.21 | 80.39 | 31 | 1.13 | 0.89 |
| 2016 | 1.96 | 3.61 | 55.38 | 131.93 | 29 | 2.02 | 1.10 |

Figure 4.6: SLL species accumulation curve - Gates Canyon 1978

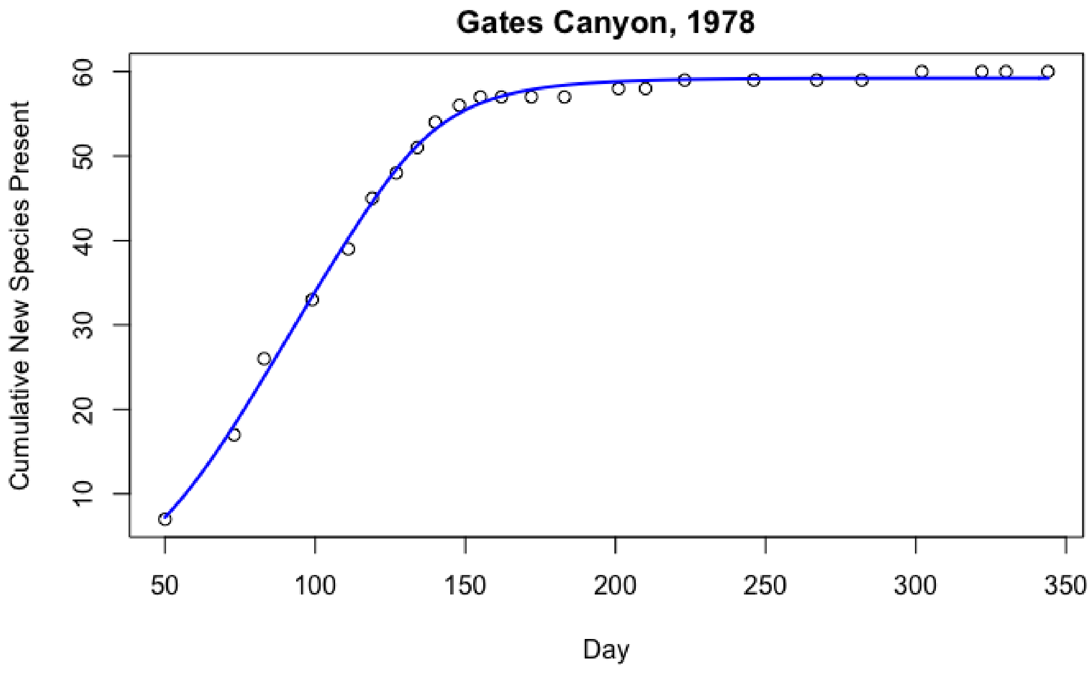



Figure 4.6 shows the SLL curve for Gates Canyon for 1978. The parameters for that year are: $k = 1.70$, $\sigma = 4.78$, $d = 59.23$ and $\mu = 118.02$. MSE = 0.709 and MAE = 0.69

Figure 4.7: SLL species accumulation curve - Gates Canyon 1984

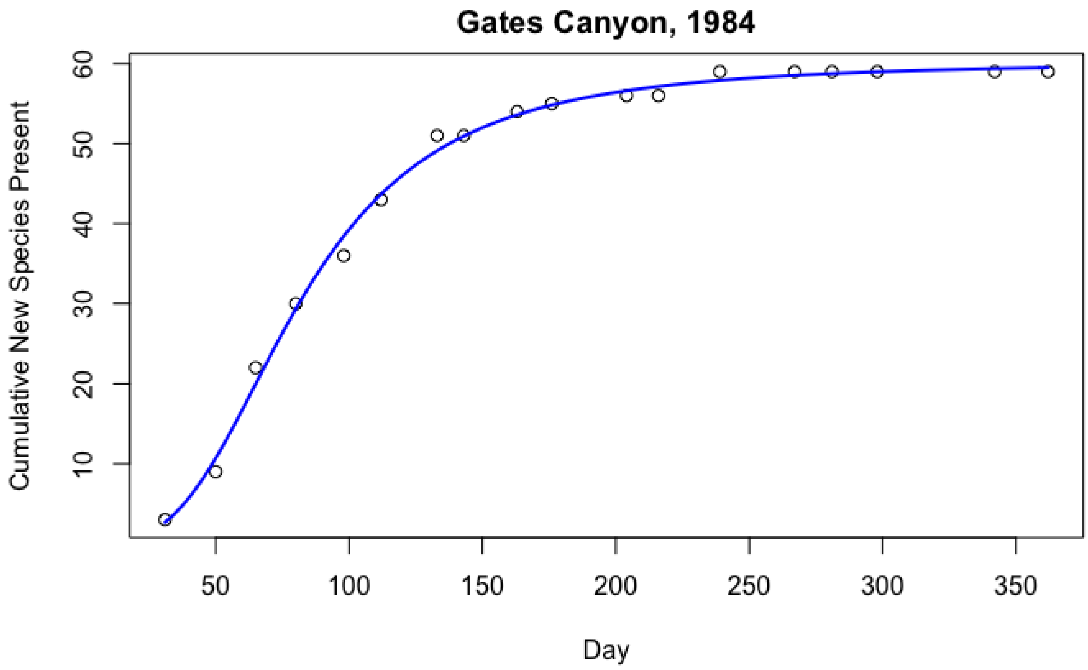



Figure 4.7 shows the SLL curve for Gates Canyon for 1984. The parameters for that year are: $k = 0.96$, $\sigma = 3.09$, $d = 60.19$ and $\mu = 79.11$. MSE = 1.183 and MAE = 0.82

Figure 4.8: SLL species accumulation curve - Gates Canyon 1997

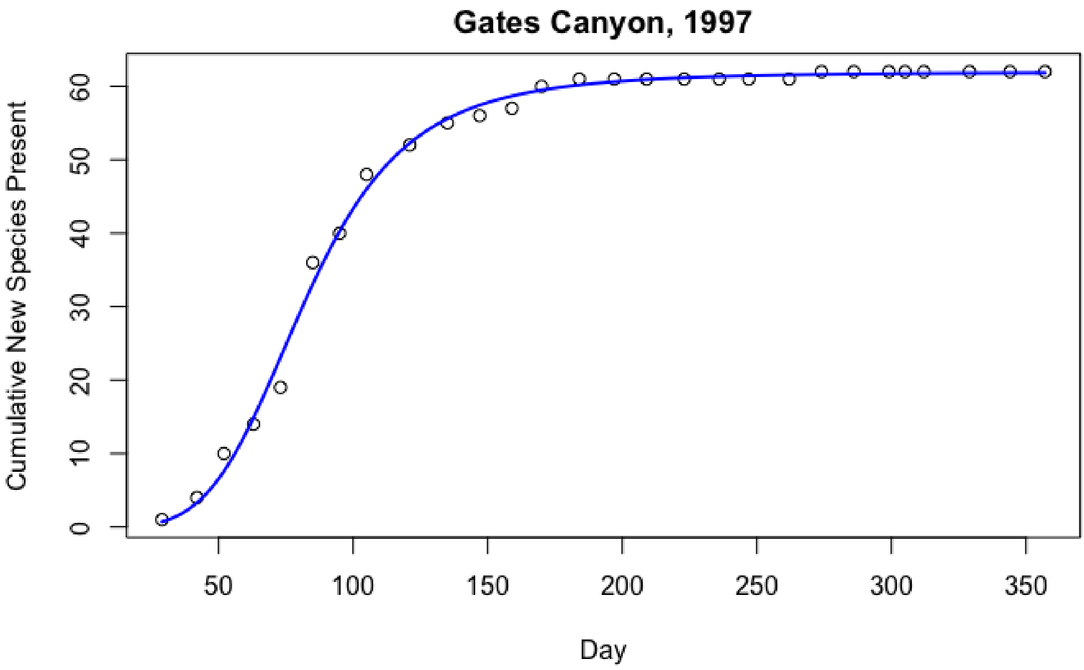



Figure 4.8 shows the SLL curve for Gates Canyon for 1997. The parameters for that year are: $k = 1.02$, $\sigma = 4.34$, $d = 61.93$ and $\mu = 83.19$. MSE = 1.571 and MAE = 0.82

Figure 4.9: SLL species accumulation curve - Gates Canyon 2007

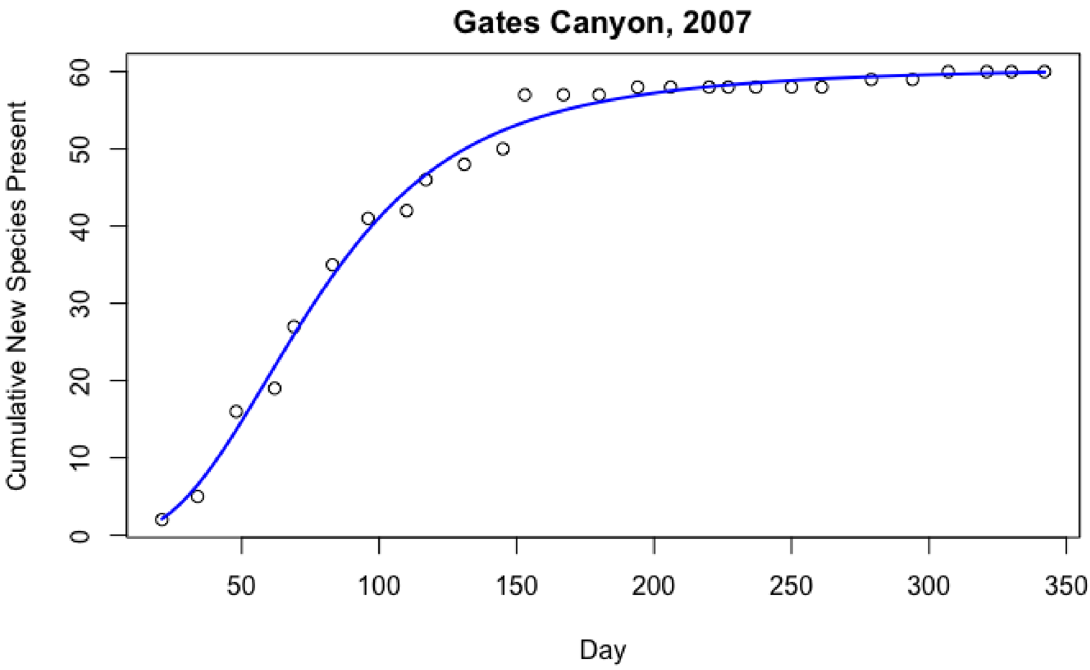



Figure 4.9 shows the SLL curve for Gates Canyon for 2007. The parameters for that year are: $k = 1.11$, $\sigma = 2.77$, $d = 60.62$ and $\mu = 82.77$. MSE = 2.221 and MAE = 1.15

Figure 4.10: SLL species accumulation curve - Gates Canyon 2015

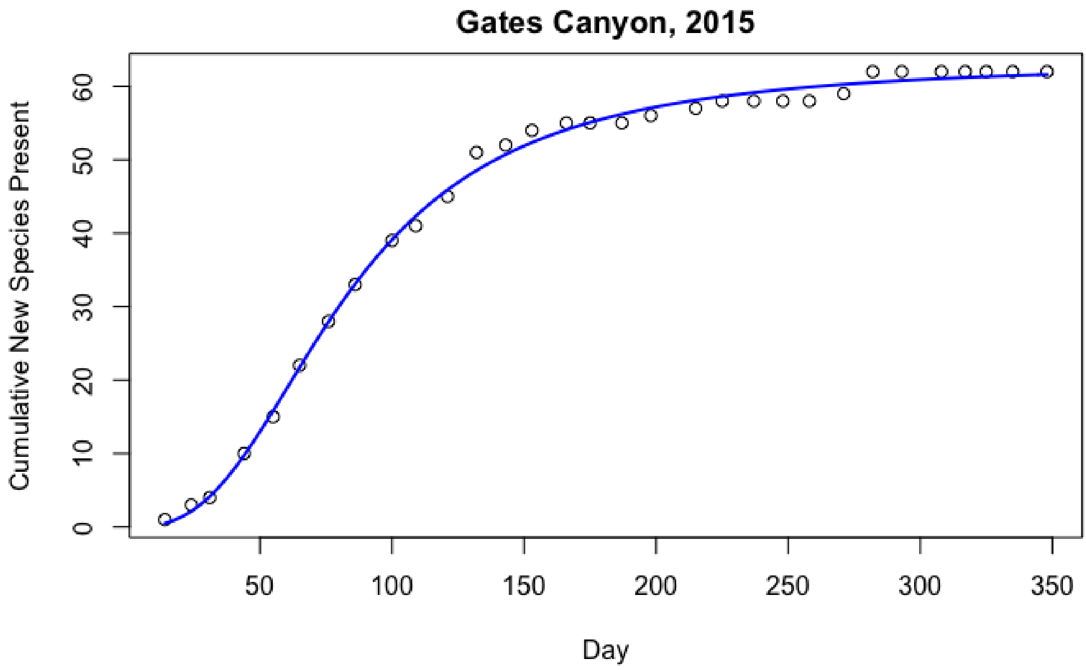



Figure 4.10 shows the SLL curve for Gates Canyon for 2015. The parameters for that year are: $k = 0.96$, $\sigma = 2.63$, $d = 63.21$ and $\mu = 80.39$. MSE = 1.13 and MAE = 0.89

We have chosen to select these years as examples to provide examples of years with a skewness parameter greater than 1 and less than 1. Figure 4.7 has a skewness parameter equal to 0.96, while Figure 4.9 has $k = 1.11$. The range for the mean absolute error for all years of Gates Canyon is 0.41 to 1.93. This demonstrates a goodness of fit where the maximum error occurred in the year 1987 showing a difference of 2 cumulative species between the estimated SLL curve and our observed number of species. This low MAE demonstrates the effectiveness of the use of the SLL curve in modeling the species accumulation curve for Gates Canyon.

**Fitted parameters and the species accumulation curves for Suisun Marsh**

Tables 4.4 and 4.5 contain the fitted parameters for the Suisun Marsh site as well as the MSE and MAE for each year to demonstrate the goodness of fit for each year. The tables are formatted on a yearly basis and displays the parameters for each year for the

Suisun Marsh site from 1973 - 2016. We have chosen to provide one SLL curve from each decade during the observed period for this site to demonstrate various goodness of fit.

Table 4.4: Suisun Marsh fitted parameters

| Year | $k$ | $\sigma$ | $d$ | $\mu$ | Visits | MSE | MAE |
|---|---|---|---|---|---|---|---|
| 1973 | 0.91 | 2.42 | 44.01 | 138.99 | 18 | 1.50 | 0.98 |
| 1974 | 0.72 | 1.47 | 53.88 | 88.83 | 32 | 1.56 | 1.01 |
| 1975 | 3.64 | 4.96 | 36.1 | 243.57 | 26 | 0.64 | 0.58 |
| 1976 | 0.12 | 5.32 | 57.18 | 22.69 | 30 | 3.37 | 1.47 |
| 1977 | 0.05 | 38.23 | 40.17 | 50.89 | 26 | 1.50 | 0.96 |
| 1978 | 1.38 | 3.13 | 38.31 | 133.47 | 34 | 1.44 | 0.90 |
| 1979 | 0.09 | 16.58 | 38.4 | 53.56 | 17 | 0.97 | 0.84 |
| 1980 | 2.33 | 3.74 | 31.88 | 199.54 | 19 | 0.90 | 0.76 |
| 1981 | 3.74 | 7.84 | 32.98 | 154.17 | 19 | 0.62 | 0.47 |
| 1982 | 3.54 | 4.51 | 32.12 | 242.02 | 18 | 0.70 | 0.65 |
| 1983 | 2.75 | 5.52 | 30.85 | 219.42 | 18 | 1.76 | 1.00 |
| 1984 | 0.12 | 12.71 | 37.04 | 47.62 | 19 | 1.03 | 0.77 |
| 1985 | 0.94 | 1.74 | 49.38 | 133.47 | 18 | 1.56 | 1.00 |
| 1986 | 1.39 | 2.53 | 38.86 | 150.45 | 19 | 0.83 | 0.73 |
| 1987 | 1.54 | 2.39 | 37.73 | 151.18 | 21 | 1.15 | 0.78 |
| 1988 | 0.65 | 2.15 | 42.45 | 53.73 | 22 | 1.16 | 0.86 |
| 1989 | 1.29 | 3.2 | 37.2 | 98.38 | 24 | 1.83 | 0.96 |

Table 4.5: Suisun Marsh fitted parameters continued

| Year | $k$ | $\sigma$ | $d$ | $\mu$ | Visits | MSE | MAE |
|---|---|---|---|---|---|---|---|
| 1990 | 1.19 | 1.81 | 38.74 | 110.72 | 25 | 1.49 | 0.93 |
| 1991 | 3.46 | 4.25 | 46.32 | 357.67 | 28 | 1.10 | 0.91 |
| 1992 | 1.38 | 2.62 | 41.05 | 125.61 | 24 | 0.30 | 0.42 |
| 1993 | 1.28 | 2.69 | 38.54 | 140.07 | 26 | 1.33 | 0.90 |
| 1994 | 0.84 | 2.88 | 36.62 | 82.15 | 21 | 1.39 | 0.84 |
| 1995 | 2.39 | 3.64 | 40.64 | 314.84 | 22 | 0.59 | 0.59 |
| 1996 | 4.03 | 4.19 | 42.16 | 281.05 | 25 | 2.15 | 1.00 |
| 1997 | 0.51 | 2.98 | 42.71 | 55.89 | 26 | 2.21 | 1.21 |
| 1998 | 3.86 | 5.31 | 40.39 | 345.04 | 26 | 1.41 | 0.89 |
| 1999 | 1.76 | 3.78 | 33.26 | 179.91 | 30 | 0.54 | 0.54 |
| 2000 | 3.33 | 3.78 | 34.25 | 250.21 | 30 | 0.98 | 0.70 |
| 2001 | 0.51 | 1.77 | 45.31 | 54.74 | 29 | 1.15 | 0.82 |
| 2002 | 0.49 | 1.59 | 52.4 | 62.44 | 29 | 0.68 | 0.63 |
| 2003 | 0.53 | 0.92 | 85.48 | 100.1 | 35 | 0.74 | 0.73 |
| 2004 | 0.6 | 3.79 | 34.65 | 69.42 | 28 | 1.55 | 0.94 |
| 2005 | 1.15 | 2.26 | 41.57 | 155.24 | 29 | 2.54 | 1.25 |
| 2006 | 5.35 | 7.71 | 29.03 | 282.62 | 31 | 0.97 | 0.73 |
| 2007 | 0.75 | 2.89 | 34.45 | 60.53 | 31 | 2.44 | 1.07 |
| 2008 | 0.5 | 2.52 | 43.02 | 66.5 | 28 | 0.51 | 0.58 |
| 2009 | 0.8 | 1.61 | 41.8 | 96.12 | 28 | 1.55 | 0.99 |
| 2010 | 1.04 | 2.09 | 33.16 | 119.01 | 29 | 1.73 | 1.07 |
| 2011 | 1.73 | 3.15 | 30.68 | 190.94 | 28 | 0.62 | 0.58 |
| 2012 | 1.41 | 1.87 | 36.37 | 112.66 | 30 | 0.53 | 0.52 |
| 2013 | 0.62 | 1.73 | 44.78 | 63.48 | 30 | 0.45 | 0.52 |
| 2014 | 2.22 | 2.8 | 34.11 | 158.24 | 31 | 0.78 | 0.57 |
| 2015 | 0.57 | 1.89 | 39.75 | 49.62 | 29 | 0.61 | 0.59 |
| 2016 | 2.47 | 3.25 | 35.13 | 192.91 | 28 | 1.84 | 0.95 |

Figure 4.11: SLL species accumulation curve - Suisun Marsh 1978

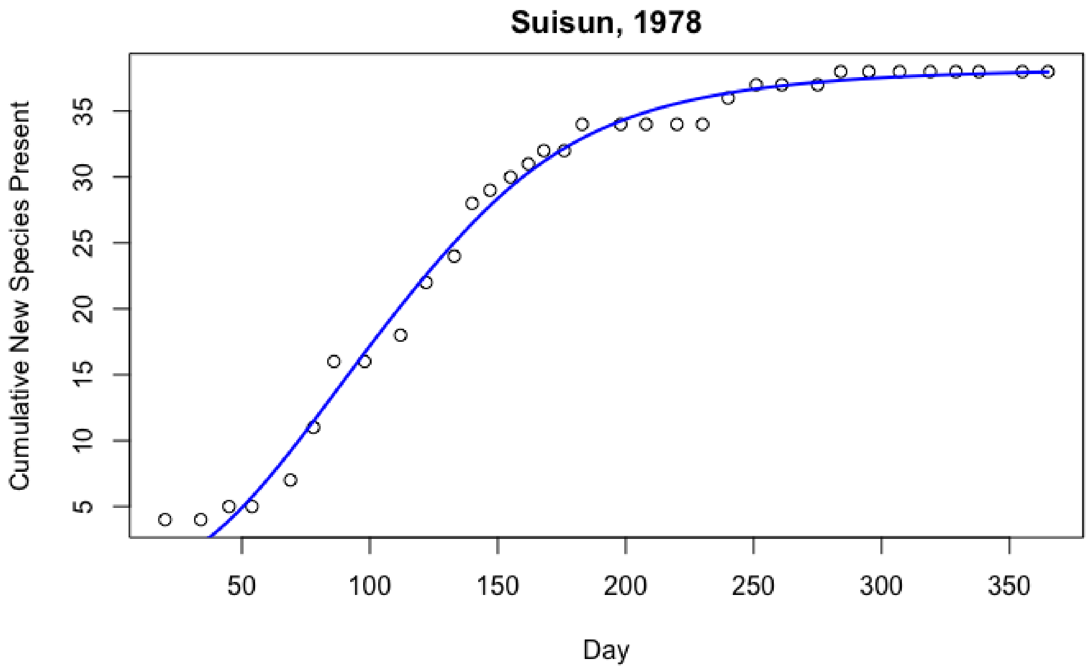



Figure 4.11 shows the SLL curve for Suisun Marsh for 1978. The parameters for that year are: $k = 1.38$, $\sigma = 3.13$, $d = 38.31$ and $\mu = 133.47$. MSE = 1.444 and MAE = 0.90

Figure 4.12: SLL species accumulation curve - Suisun Marsh 1988

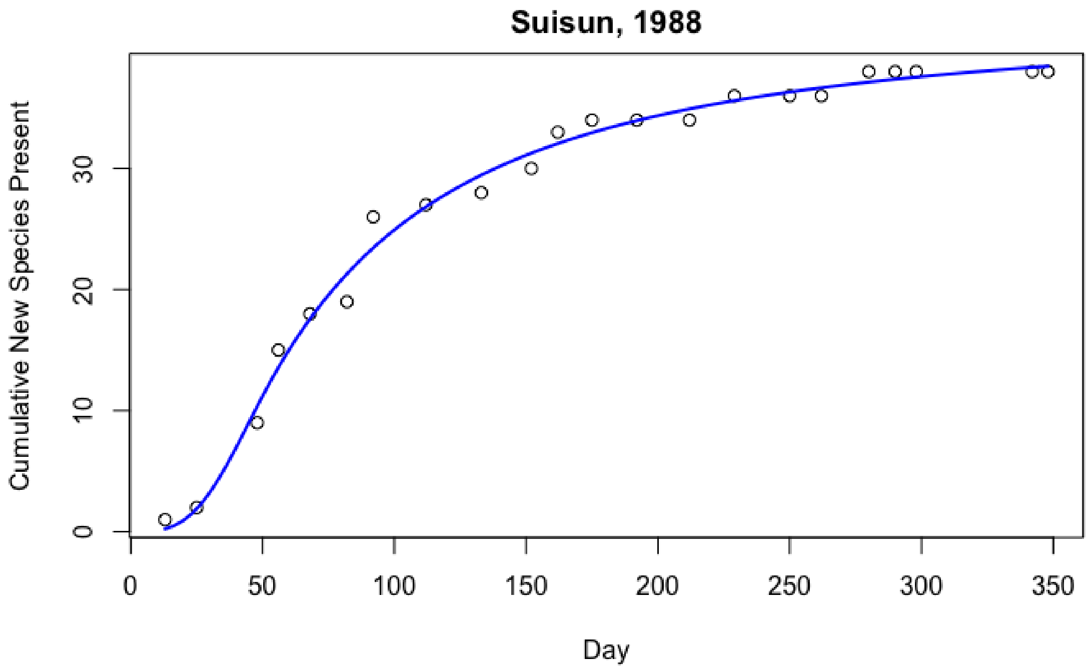



Figure 4.12 shows the SLL curve for Suisun Marsh for 1988. The parameters for that year are: $k = 0.65$, $\sigma = 2.15$, $d = 42.45$ and $\mu = 53.73$. MSE = 1.155 and MAE = 0.86

Figure 4.13: SLL species accumulation curve - Suisun Marsh 1993

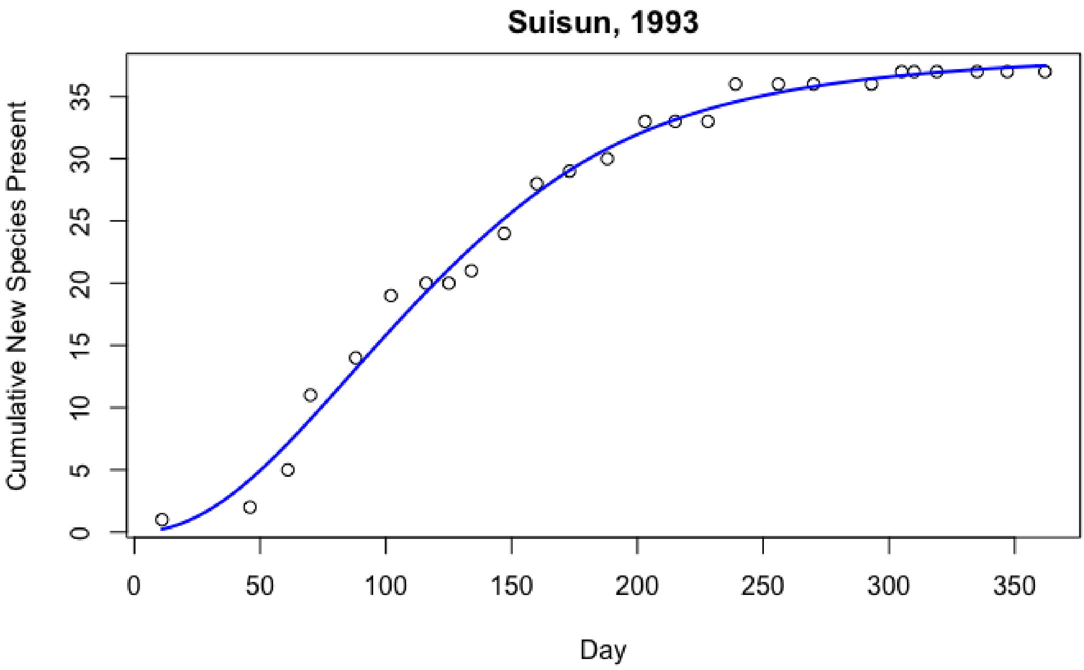



Figure 4.13 shows the SLL curve for Suisun Marsh for 1993. The parameters for that year are: $k = 1.28$, $\sigma = 2.69$, $d = 38.54$ and $\mu = 140.07$. MSE = 1.334 and MAE = 0.90

Figure 4.14: SLL species accumulation curve - Suisun Marsh 2004

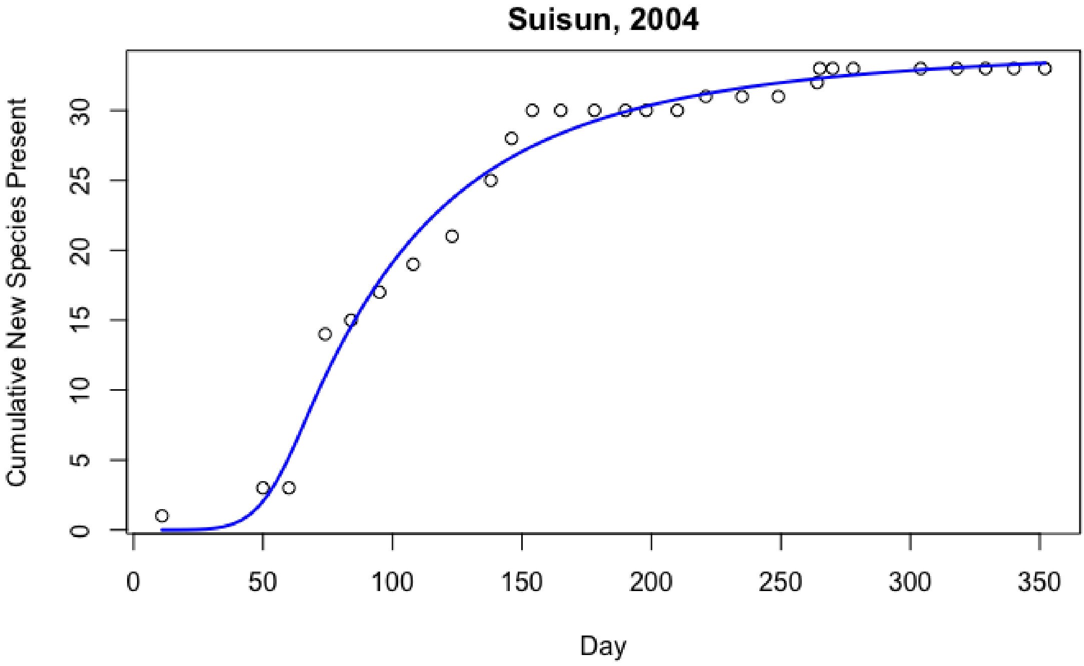



Figure 4.14 shows the SLL curve for Suisun Marsh for 2004. The parameters for that year are: $k = 0.60$, $\sigma = 3.79$, $d = 34.65$ and $\mu = 69.42$. MSE = 1.552 and MAE = 0.94

Figure 4.15: SLL species accumulation curve - Suisun Marsh 2013

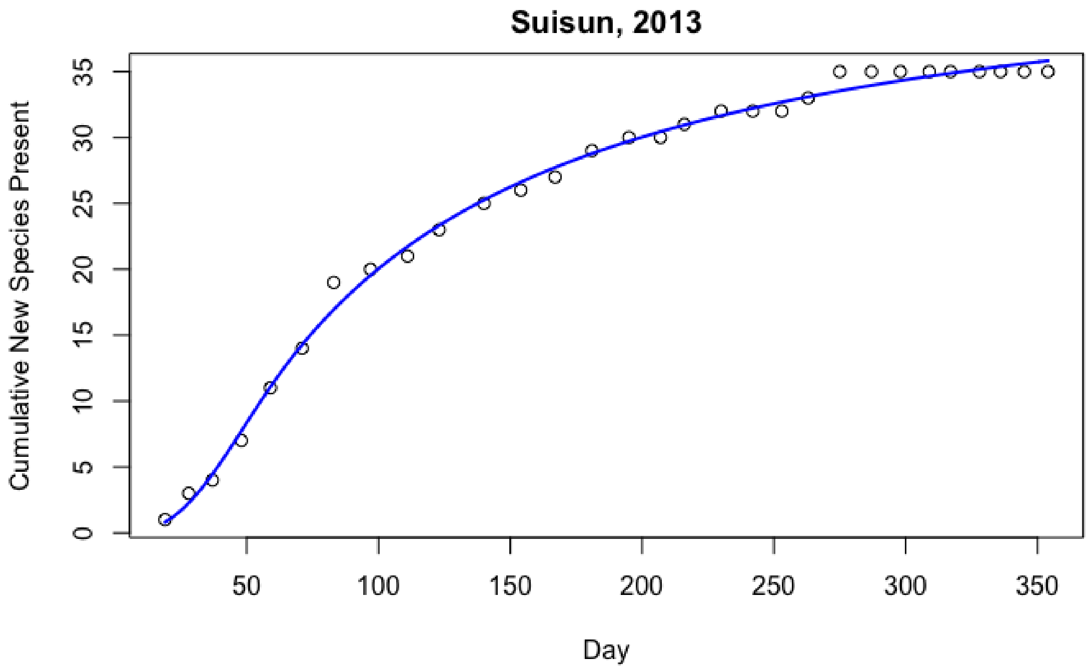



Figure 4.15 shows the SLL curve for Suisun Marsh for 2013. The parameters for that year are: $k = 0.62$, $\sigma = 1.73$, $d = 44.78$ and $\mu = 63.48$. MSE = 0.45 and MAE = 0.52

The years chosen provide examples with a skewness parameter greater than 1 and less than 1. Figure 4.14 has a skewness parameter equal to 0.60, while Figure 4.11 has $k = 1.38$. The range for the calculated mean absolute error for Suisun Marsh is 0.42 to 1.47. This relatively low range demonstrates the effectiveness of using the SLL curve in estimating the species accumulation of this site.

**Fitted parameters and the species accumulation curves for West Sacramento**

Table 4.6 is comprised of the fitted parameters for the West Sacramento site as well as the MSE and MAE for each year to demonstrate the goodness of fit for each year. The table is formatted on a yearly basis displaying the parameters for each year for the West Sacramento site from 1988 - 2016. We have chosen to provide one SLL curve from each decade during the observed period for this site to demonstrate various goodness of fit.

Table 4.6: West Sacramento fitted parameters

| Year | $k$ | $\sigma$ | $d$ | $\mu$ | Visits | MSE | MAE |
|---|---|---|---|---|---|---|---|
| 1988 | 0.61 | 2.07 | 42.58 | 49.47 | 18 | 0.86 | 0.77 |
| 1989 | 1.08 | 2.97 | 39.04 | 97.78 | 26 | 1.37 | 0.99 |
| 1990 | 1.08 | 3.35 | 40.01 | 86.85 | 29 | 1.94 | 0.92 |
| 1991 | 3.37 | 5.45 | 33.07 | 232.83 | 26 | 0.88 | 0.71 |
| 1992 | 0.76 | 3.01 | 45.28 | 83.63 | 26 | 2.14 | 1.15 |
| 1993 | 2.09 | 3.64 | 39.17 | 184.68 | 25 | 3.21 | 1.15 |
| 1994 | 1.12 | 3.19 | 39.03 | 103.5 | 22 | 0.67 | 0.60 |
| 1995 | 9.01 | 12.03 | 56.18 | 469.36 | 26 | 1.52 | 0.98 |
| 1996 | 2.24 | 2.9 | 34.35 | 179.48 | 27 | 2.43 | 0.98 |
| 1997 | 1.37 | 3.19 | 36.6 | 121.22 | 26 | 0.80 | 0.64 |
| 1998 | 1.84 | 3.04 | 32.52 | 166.5 | 27 | 2.55 | 1.11 |
| 1999 | 1.3 | 3.04 | 32.46 | 122.4 | 26 | 0.92 | 0.62 |
| 2000 | 1.29 | 3.67 | 36.94 | 118.35 | 29 | 1.19 | 0.77 |
| 2001 | 1.75 | 2.89 | 35.07 | 127.51 | 29 | 0.60 | 0.57 |
| 2002 | 0.92 | 2.78 | 35.97 | 99.83 | 31 | 0.55 | 0.61 |
| 2003 | 0.64 | 2.36 | 33.86 | 70.72 | 31 | 1.82 | 0.98 |
| 2004 | 0.45 | 3 | 42.05 | 64.51 | 28 | 2.06 | 1.04 |
| 2005 | 0.4 | 2.69 | 36.77 | 50.09 | 31 | 1.16 | 0.85 |
| 2006 | 2.95 | 4.81 | 31.14 | 246.79 | 30 | 1.40 | 0.89 |
| 2007 | 0.86 | 3.55 | 33.32 | 70.71 | 27 | 0.56 | 0.52 |
| 2008 | 1.05 | 3.59 | 28.7 | 96.38 | 29 | 1.00 | 0.77 |
| 2009 | 1.09 | 1.82 | 35.1 | 122.8 | 31 | 1.61 | 1.01 |
| 2010 | 0.55 | 2.44 | 35.36 | 82.97 | 31 | 0.76 | 0.69 |
| 2011 | 0.37 | 3.72 | 30.7 | 67.81 | 28 | 1.37 | 0.99 |
| 2012 | 0.8 | 1.44 | 43.32 | 72.01 | 35 | 1.08 | 0.84 |
| 2013 | 0.96 | 2.76 | 36.57 | 82.12 | 32 | 1.15 | 0.92 |
| 2014 | 1.08 | 1.7 | 41.41 | 76.52 | 30 | 1.73 | 0.93 |
| 2015 | 0.52 | 2.15 | 41.51 | 46.98 | 31 | 1.44 | 0.94 |
| 2016 | 0.56 | 3.09 | 37.19 | 55.74 | 29 | 1.05 | 0.85 |

Figure 4.16: SLL species accumulation curve - West Sacramento 1989

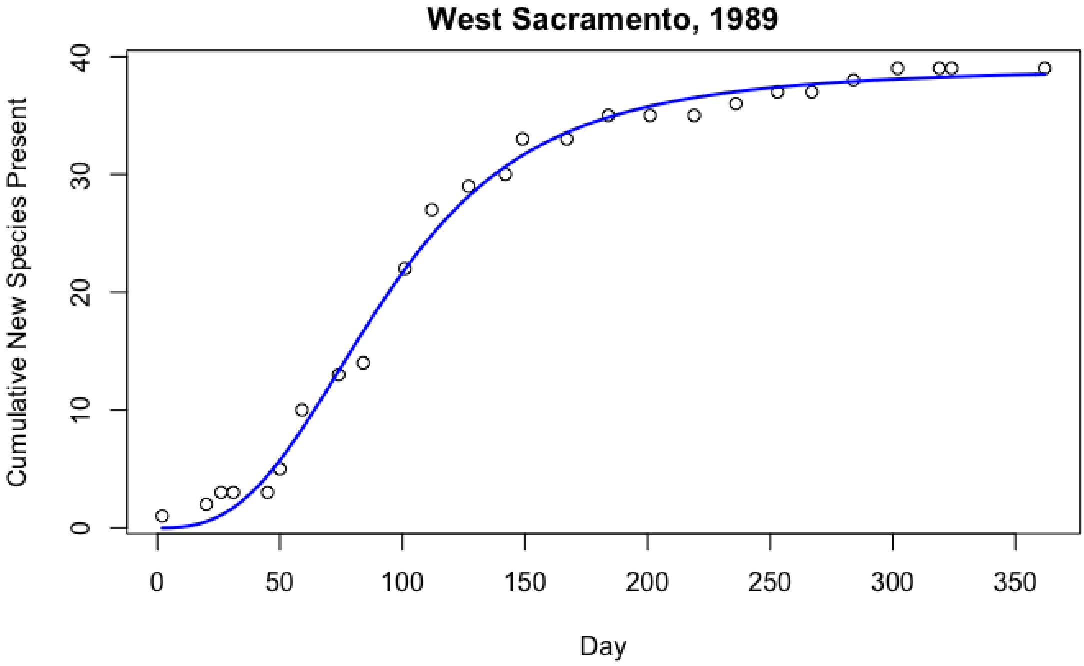



Figure 4.16 shows the SLL curve for West Sacramento for 1989. The parameters for that year are: $k = 1.08$, $\sigma = 2.97$, $d = 39.04$ and $\mu = 97.78$. MSE = 1.374 and MAE = 0.99

Figure 4.17: SLL species accumulation curve - West Sacramento 1997

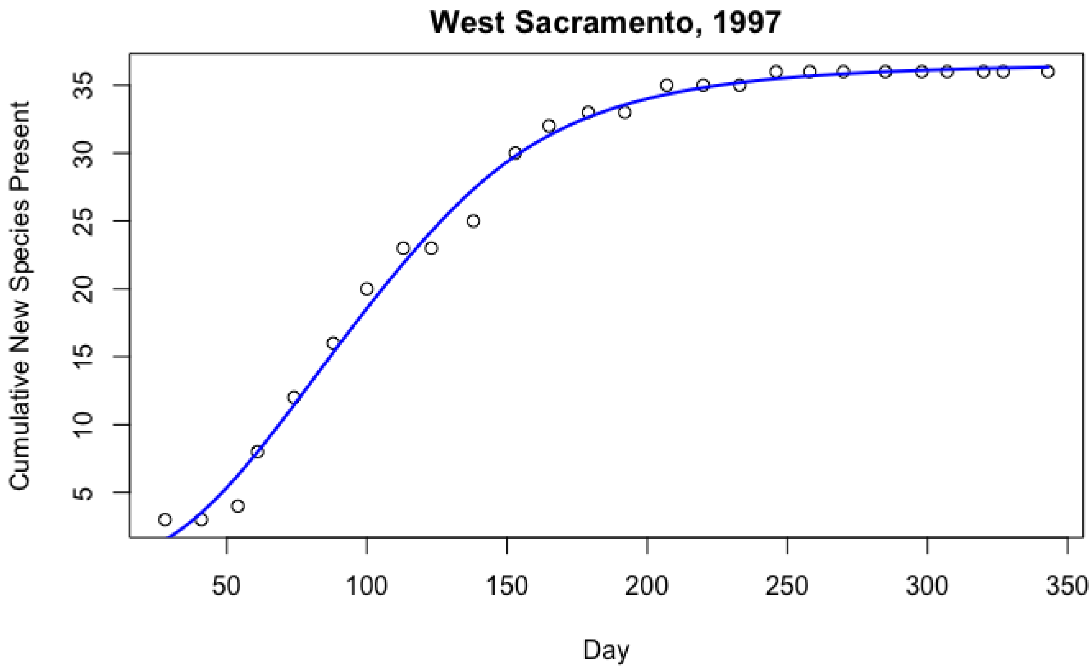



Figure 4.17 shows the SLL curve for West Sacramento for 1997. The parameters for that year are: $k = 1.37$, $\sigma = 3.19$, $d = 36.60$ and $\mu = 121.22$. MSE = 0.802 and MAE = 0.64

Figure 4.18: SLL species accumulation curve - West Sacramento 2008

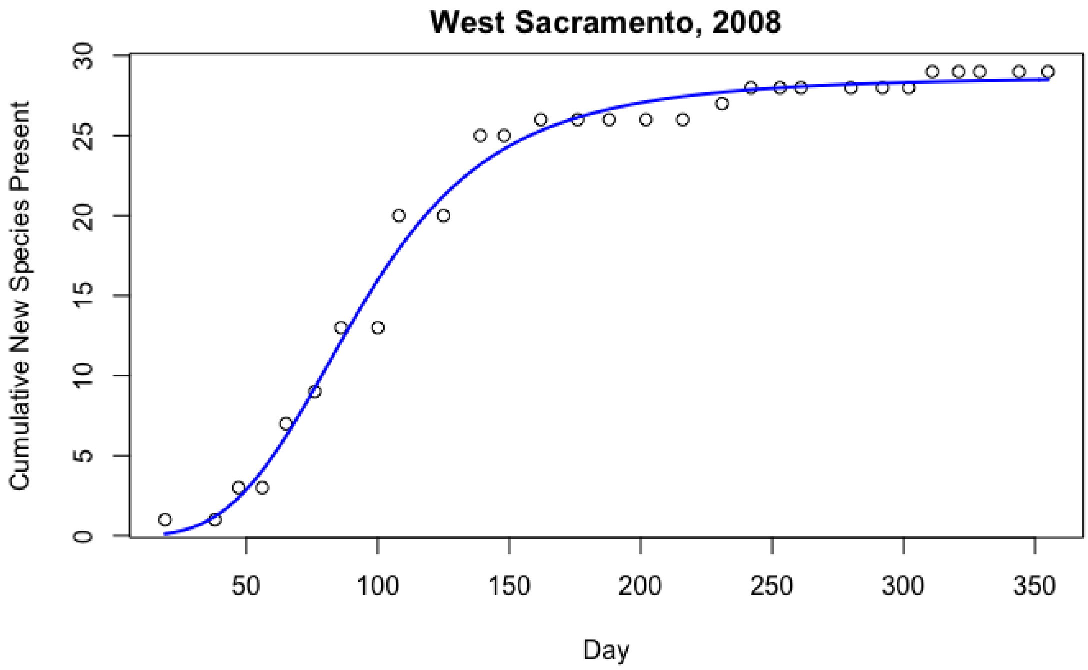



Figure 4.18 shows the SLL curve for West Sacramento for 2008. The parameters for that year are: $k = 1.05$, $\sigma = 3.59$, $d = 28.70$ and $\mu = 96.38$. MSE = 1.003 and MAE = 0.77

Figure 4.19: SLL species accumulation curve - West Sacramento 2013

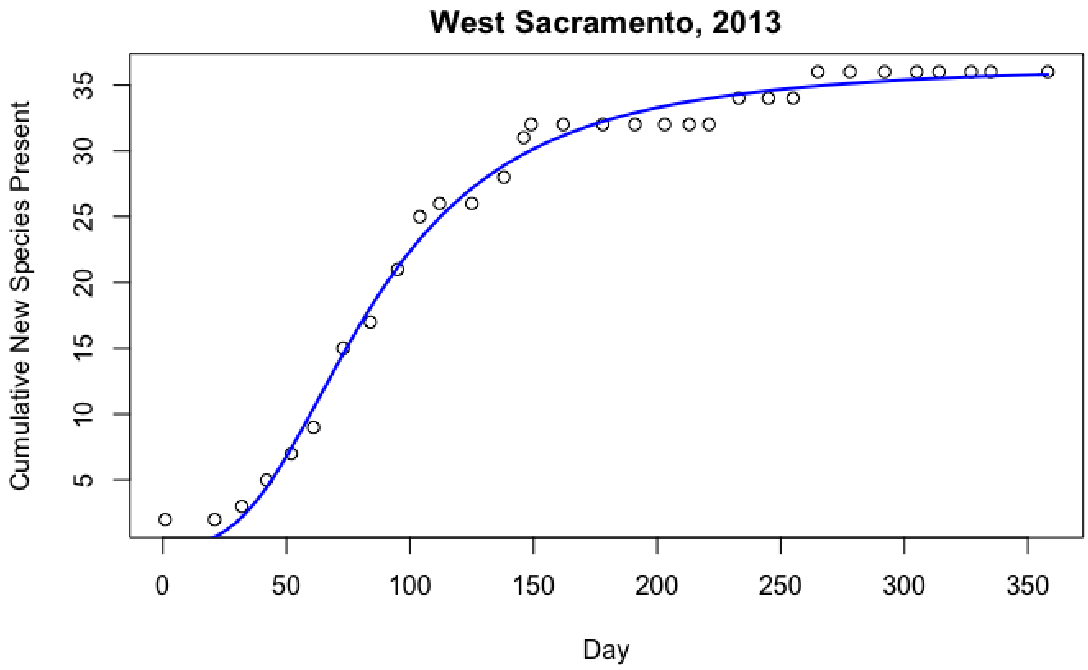



Figure 4.19 shows the SLL curve for West Sacramento for 2013. The parameters for that year are: $k = 0.96$, $\sigma = 2.76$, $d = 36.57$ and $\mu = 82.12$. MSE = 1.145 and MAE = 0.92

These figures demonstrate years with a $k$ greater and less than 1. Figure 4.19 has a $k$ equal to 0.96, while Figure 4.17 has $k = 1.37$. The range of MAE for all years in the West Sacramento site is 0.52 to 1.15 showing a low error with at most 1.15 species variance between the estimated SLL curve and the observed number of species.

**Fitted parameters and the species accumulation curves for North Sacramento**

Table 4.7 is comprised of the fitted parameters for the North Sacramento site as well as the MSE and MAE for each year to demonstrate the goodness of fit for each year. The table is formatted on a yearly basis displaying the parameters for each year for the North Sacramento site from 1988 - 2016. We have chosen to provide one SLL curve from each decade during the observed period for this site to demonstrate various goodness of fit.

Table 4.7: North Sacramento fitted parameters

| Year | $k$ | $\sigma$ | $d$ | $\mu$ | Visits | MSE | MAE |
|---|---|---|---|---|---|---|---|
| 1988 | 0.83 | 2.69 | 41.75 | 71.06 | 12 | 0.47 | 0.51 |
| 1989 | 1.14 | 4.04 | 42.84 | 95.19 | 20 | 0.40 | 0.54 |
| 1990 | 2.26 | 4.9 | 45.58 | 122.73 | 24 | 3.83 | 1.41 |
| 1991 | 2.3 | 4.25 | 36.84 | 173.5 | 25 | 1.02 | 0.73 |
| 1992 | 0.81 | 3.91 | 43.25 | 82.45 | 25 | 1.45 | 0.91 |
| 1993 | 1.6 | 4.16 | 40.87 | 140.19 | 24 | 0.75 | 0.68 |
| 1994 | 1.26 | 3.71 | 43.52 | 120.69 | 21 | 0.55 | 0.58 |
| 1995 | 3.03 | 6.48 | 38 | 214.69 | 24 | 1.45 | 0.81 |
| 1996 | 0.68 | 3.4 | 41.74 | 77.38 | 26 | 2.09 | 1.00 |
| 1997 | 0.89 | 3.27 | 39.99 | 82.84 | 25 | 0.89 | 0.77 |
| 1998 | 4.2 | 7.81 | 39.85 | 226.41 | 27 | 1.49 | 0.95 |
| 1999 | 1 | 3.63 | 36.67 | 103.77 | 28 | 0.58 | 0.54 |
| 2000 | 2.34 | 5.21 | 37.91 | 177.64 | 27 | 2.78 | 1.15 |
| 2001 | 0.81 | 2.57 | 38.88 | 77.48 | 29 | 2.11 | 1.17 |
| 2002 | 0.83 | 2.94 | 37.07 | 86.67 | 28 | 0.48 | 0.56 |
| 2003 | 0.66 | 2.13 | 39.81 | 61.85 | 30 | 2.60 | 1.11 |
| 2004 | 0.34 | 4.41 | 38.82 | 56.82 | 25 | 3.11 | 1.43 |
| 2005 | 1.43 | 2.97 | 34.75 | 141.81 | 27 | 0.64 | 0.57 |
| 2006 | 2.15 | 4.71 | 34.48 | 204.66 | 31 | 1.70 | 0.98 |
| 2007 | 0.88 | 3.07 | 38.5 | 86.25 | 30 | 1.05 | 0.76 |
| 2008 | 0.27 | 4.47 | 45.31 | 49.63 | 28 | 1.19 | 0.81 |
| 2009 | 0.96 | 2.98 | 39 | 106.64 | 30 | 2.17 | 1.16 |
| 2010 | 0.62 | 3.05 | 36.9 | 85.73 | 30 | 0.86 | 0.71 |
| 2011 | 2.38 | 5.26 | 33.06 | 193.65 | 32 | 1.43 | 0.81 |
| 2012 | 1.06 | 2.18 | 41.47 | 93.43 | 33 | 1.67 | 1.04 |
| 2013 | 1.18 | 3.24 | 38.24 | 107.65 | 31 | 0.79 | 0.75 |
| 2014 | 1.94 | 3.39 | 35.06 | 129.77 | 32 | 0.71 | 0.55 |
| 2015 | 1.07 | 2.49 | 40.25 | 88.18 | 31 | 0.69 | 0.59 |
| 2016 | 1.67 | 3.73 | 41.78 | 143.1 | 31 | 1.47 | 0.86 |

Figure 4.20: SLL species accumulation curve - North Sacramento 1988

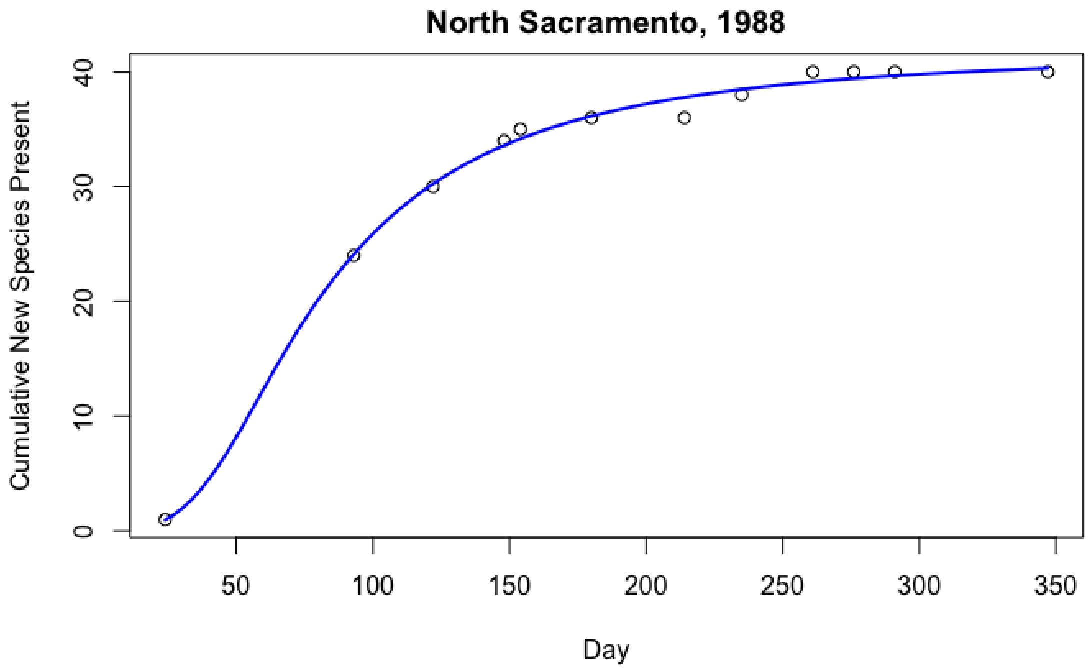



Figure 4.20 shows the SLL curve for North Sacramento for 1988. The parameters for that year are: $k = 0.83$, $\sigma = 2.69$, $d = 41.75$ and $\mu = 71.06$. MSE = 0.474 and MAE = 0.51

Figure 4.21: SLL species accumulation curve - North Sacramento 1996

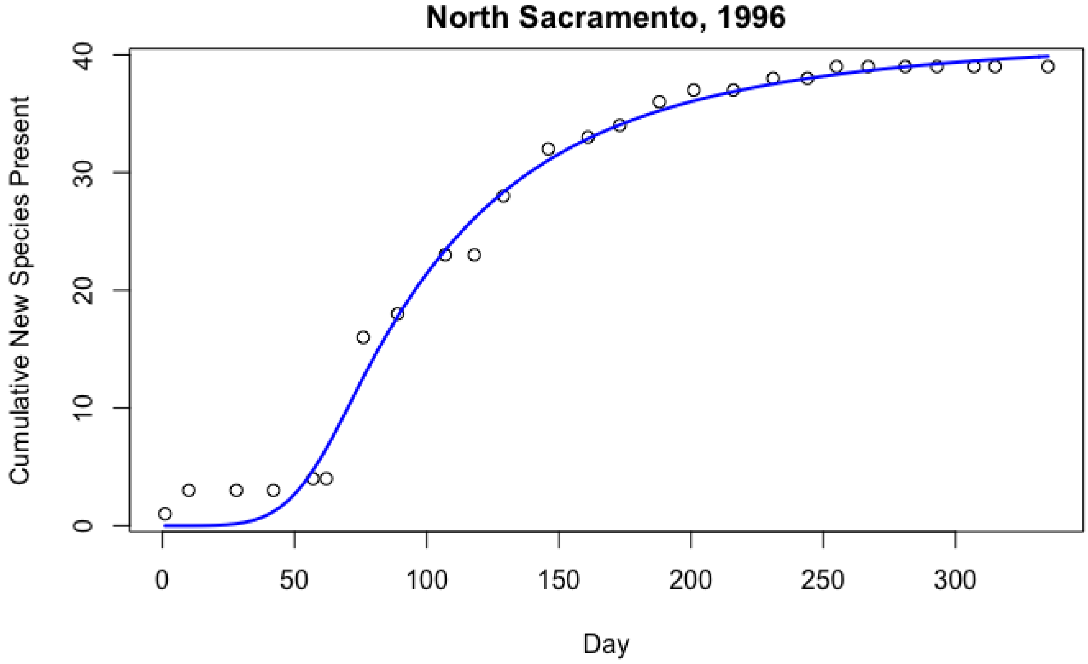



Figure 4.21 shows the SLL curve for North Sacramento for 1996. The parameters for that year are: $k = 0.68$, $\sigma = 3.40$, $d = 41.74$ and $\mu = 77.38$. MSE = 2.094 and MAE = 1.00

Figure 4.22: SLL species accumulation curve - North Sacramento 2007

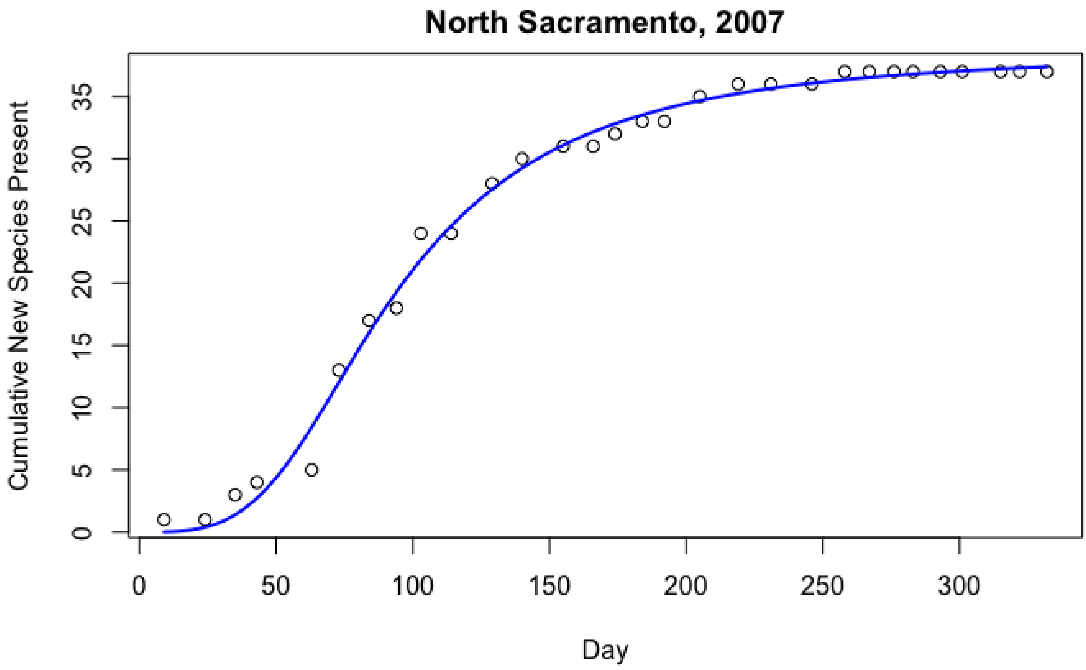



Figure 4.22 shows the SLL curve for North Sacramento for 2007. The parameters for that year are: $k = 0.88$, $\sigma = 3.07$, $d = 38.50$ and $\mu = 86.25$. MSE = 1.052 and MAE = 0.76

Figure 4.23: SLL species accumulation curve - North Sacramento 2012

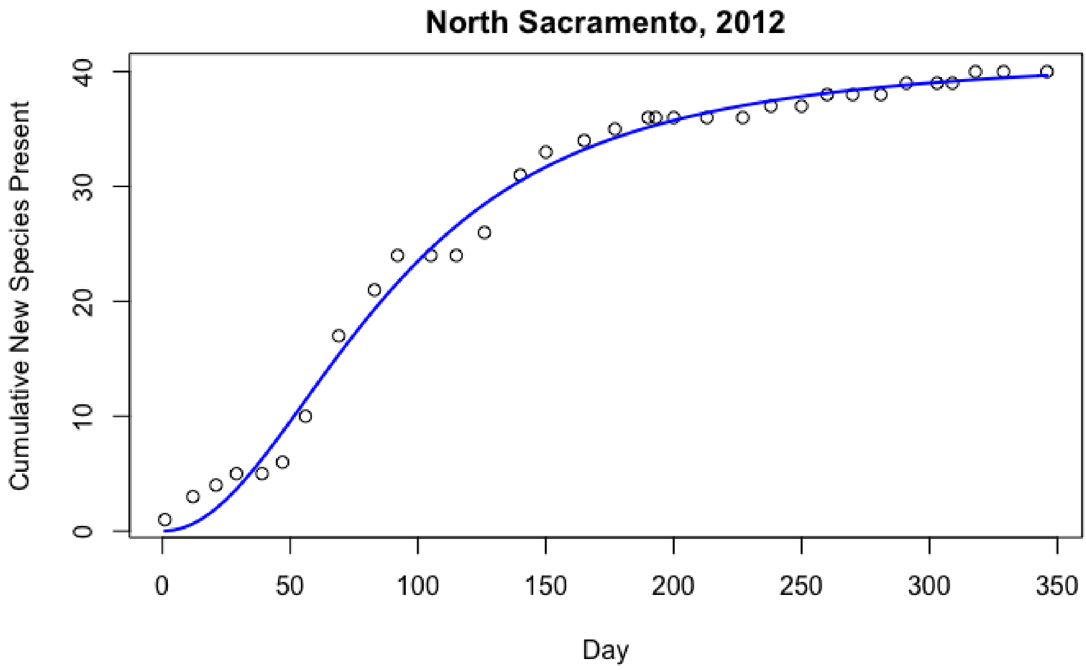



Figure 4.23 shows the SLL curve for North Sacramento for 2012. The parameters for that year are: $k = 1.06$, $\sigma = 2.18$, $d = 41.47$ and $\mu = 93.43$. MSE = 1.674 and MAE = 1.04

The figures provided show examples of a skewness parameter $k$ greater and less than 1. Figure 4.21 has a skewness parameter equal to 0.68, while Figure 4.23 has $k = 1.06$. The range of MAE for all years for this site is 0.51 to 1.43. This demonstrates the effectiveness of our fit through the use of the SLL curve. The MAE shows that the error between our observed number of species and our estimated SLL curve is at most a difference of 1.43 species for this site.

**Fitted parameters and the species accumulation curves for Rancho Cordova**

Tables 4.8 and 4.9 are comprised of the fitted parameters for the Rancho Cordova site as well as the MSE and MAE for each year to demonstrate the goodness of fit for each year. The tables are formatted on a yearly basis with the parameters of the SLL curve for the years 1975 - 2016 as well as the number of visits for each year. We have chosen to provide one SLL curve from each decade during the observed period for this site to demonstrate various goodness of fit.

Table 4.8: Rancho Cordova fitted parameters

| Year | $k$ | $\sigma$ | $d$ | $\mu$ | Visits | MSE | MAE |
|---|---|---|---|---|---|---|---|
| 1975 | 1.14 | 5.96 | 32.77 | 131.48 | 9 | 0.12 | 0.29 |
| 1976 | 1.27 | 2.11 | 42.51 | 90.53 | 38 | 1.02 | 0.79 |
| 1977 | 0.06 | 42.77 | 40.54 | 52.66 | 32 | 1.93 | 1.20 |
| 1978 | 1.13 | 3.75 | 43.88 | 108.44 | 30 | 1.32 | 0.91 |
| 1979 | 1.07 | 5.41 | 43.08 | 89.02 | 20 | 1.43 | 1.06 |
| 1980 | 0.89 | 3.45 | 44.84 | 99.81 | 19 | 1.08 | 0.83 |
| 1981 | 0.86 | 2.84 | 40.00 | 74.79 | 19 | 0.75 | 0.66 |
| 1982 | 1.23 | 2.60 | 39.93 | 119.58 | 19 | 0.89 | 0.71 |
| 1983 | 1.91 | 3.30 | 41.27 | 166.22 | 18 | 1.93 | 1.11 |
| 1984 | 0.86 | 2.85 | 44.21 | 78.71 | 20 | 2.50 | 1.22 |
| 1985 | 1.30 | 4.37 | 45.78 | 118.22 | 17 | 1.91 | 1.01 |
| 1986 | 0.48 | 3.20 | 46.71 | 59.16 | 19 | 0.90 | 0.76 |
| 1987 | 1.20 | 3.27 | 42.86 | 85.64 | 21 | 4.51 | 1.83 |
| 1988 | 0.70 | 3.33 | 43.32 | 51.38 | 19 | 0.34 | 0.43 |
| 1989 | 0.78 | 3.23 | 40.56 | 66.03 | 20 | 1.69 | 0.97 |
| 1990 | 1.20 | 3.78 | 40.61 | 88.85 | 24 | 4.03 | 1.66 |

Table 4.9: Rancho Cordova fitted parameters continued

| Year | $k$ | $\sigma$ | $d$ | $\mu$ | Visits | MSE | MAE |
|---|---|---|---|---|---|---|---|
| 1991 | 0.61 | 3.77 | 37.81 | 78.49 | 21 | 2.64 | 1.24 |
| 1992 | 0.68 | 4.21 | 36.72 | 69.67 | 22 | 0.39 | 0.48 |
| 1993 | 0.62 | 3.01 | 45.34 | 80.52 | 22 | 1.30 | 0.92 |
| 1994 | 0.89 | 2.96 | 42.78 | 86.09 | 21 | 0.99 | 0.83 |
| 1995 | 4.20 | 5.88 | 39.26 | 298.30 | 25 | 0.80 | 0.75 |
| 1996 | 1.25 | 2.94 | 41.91 | 106.54 | 24 | 3.57 | 1.51 |
| 1997 | 0.69 | 3.14 | 42.97 | 84.57 | 28 | 1.88 | 1.18 |
| 1998 | 0.59 | 2.88 | 42.07 | 77.26 | 26 | 2.44 | 1.15 |
| 1999 | 1.07 | 3.71 | 37.39 | 107.99 | 28 | 2.02 | 1.17 |
| 2000 | 0.49 | 3.99 | 42.27 | 74.52 | 25 | 2.81 | 1.41 |
| 2001 | 0.54 | 4.84 | 33.83 | 62.92 | 24 | 1.50 | 0.99 |
| 2002 | 0.59 | 3.25 | 35.63 | 68.44 | 24 | 0.70 | 0.62 |
| 2003 | 0.61 | 4.11 | 31.75 | 66.21 | 28 | 1.69 | 1.00 |
| 2004 | 0.05 | 46.73 | 32.80 | 52.77 | 24 | 1.65 | 0.98 |
| 2005 | 0.44 | 2.66 | 43.64 | 62.31 | 27 | 1.21 | 0.88 |
| 2006 | 1.50 | 2.41 | 34.89 | 163.04 | 27 | 1.54 | 1.08 |
| 2007 | 1.06 | 3.37 | 39.94 | 93.28 | 27 | 2.13 | 1.19 |
| 2008 | 0.08 | 24.51 | 38.06 | 55.12 | 27 | 2.37 | 1.24 |
| 2009 | 0.72 | 3.99 | 34.22 | 67.49 | 26 | 1.36 | 0.97 |
| 2010 | 0.07 | 18.81 | 38.89 | 42.26 | 26 | 2.65 | 1.30 |
| 2011 | 0.24 | 0.75 | 204.34 | 93.41 | 31 | 2.53 | 1.27 |
| 2012 | 0.83 | 1.08 | 45.70 | 77.49 | 31 | 1.23 | 0.87 |
| 2013 | 0.41 | 2.73 | 46.64 | 53.08 | 28 | 0.91 | 0.68 |
| 2014 | 2.00 | 2.77 | 34.21 | 158.81 | 30 | 0.37 | 0.44 |
| 2015 | 0.55 | 2.49 | 41.42 | 48.90 | 30 | 0.61 | 0.59 |
| 2016 | 0.76 | 2.74 | 39.75 | 77.28 | 30 | 1.12 | 0.81 |

Figure 4.24: SLL species accumulation curve - Rancho Cordova 1978

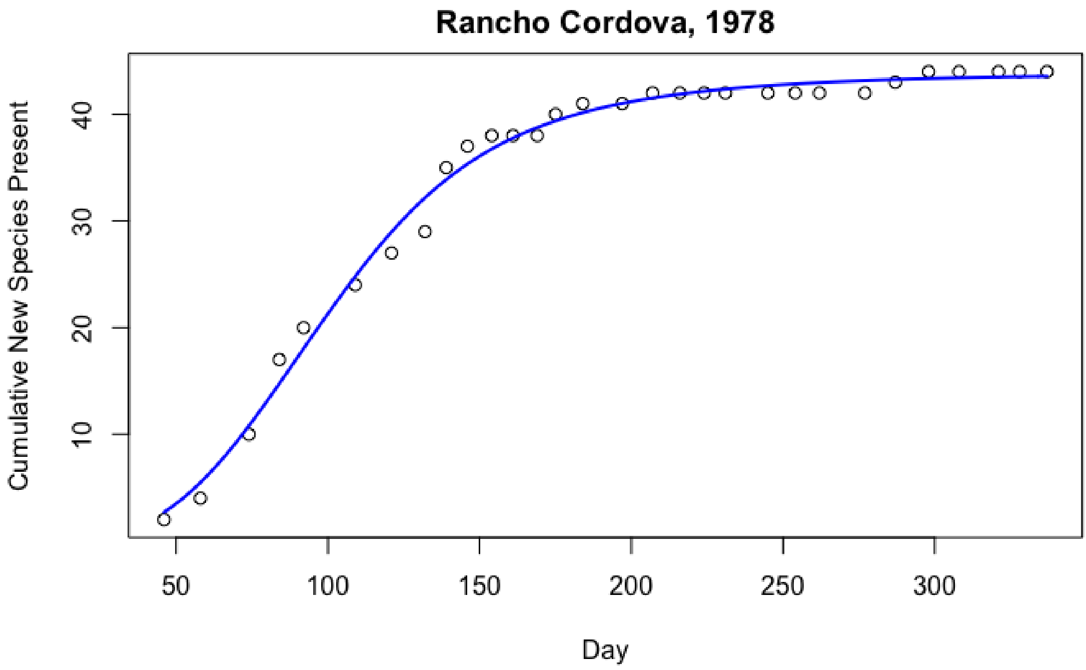



Figure 4.24 shows the SLL curve for Rancho Cordova for 1978. The parameters for that year are: $k = 1.13$, $\sigma = 3.75$, $d = 43.88$ and $\mu = 108.44$. MSE = 1.319 and MAE = 0.91

Figure 4.25: SLL species accumulation curve - Rancho Cordova 1985

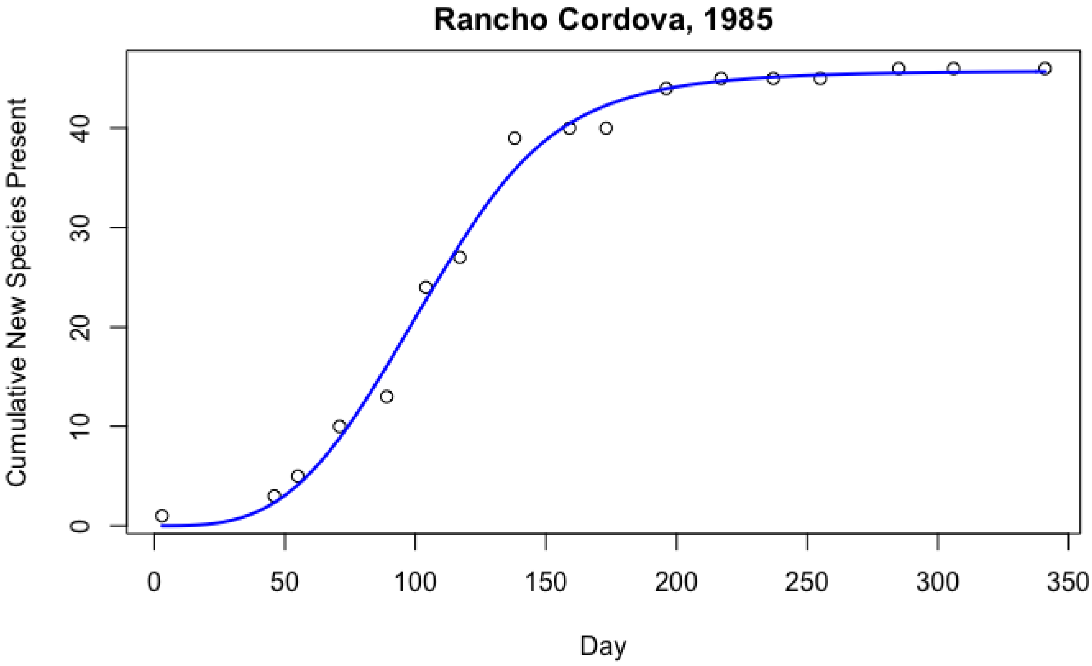



Figure 4.25 shows the SLL curve for Rancho Cordova for 1985. The parameters for that year are: $k = 1.30$, $\sigma = 4.37$, $d = 45.78$ and $\mu = 118.22$. MSE = 1.906 and MAE = 1.01

Figure 4.26: SLL species accumulation curve - Rancho Cordova 1996

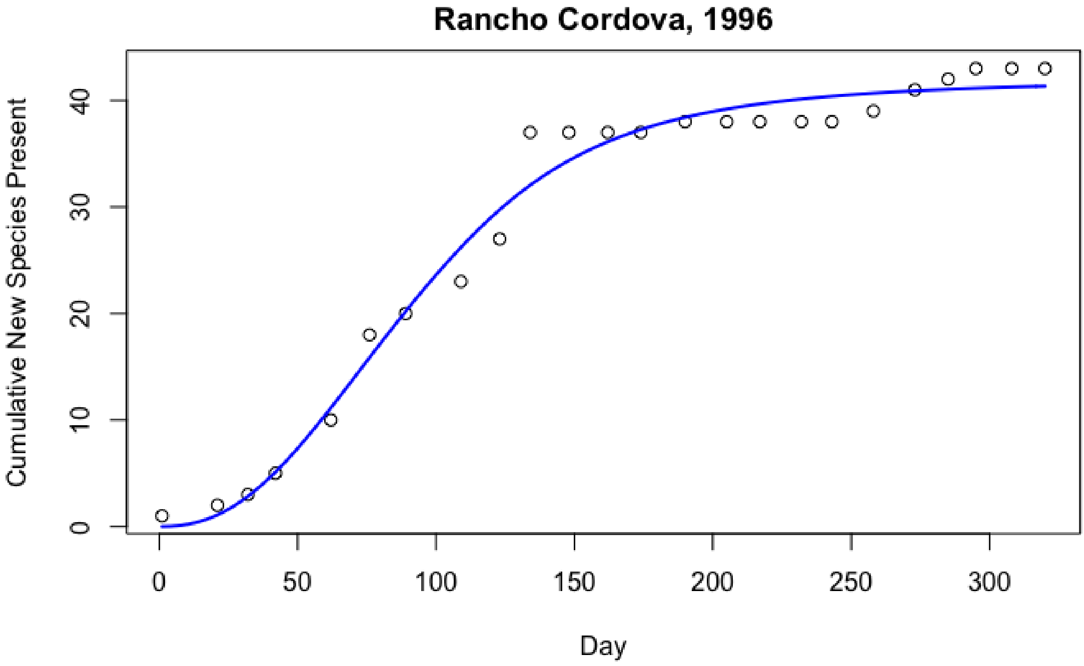



Figure 4.26 shows the SLL curve for Rancho Cordova for 1996. The parameters for that year are: $k = 1.25$, $\sigma = 2.94$, $d = 41.91$ and $\mu = 106.54$. MSE = 3.567 and MAE = 1.51

Figure 4.27: SLL species accumulation curve - Rancho Cordova 2007

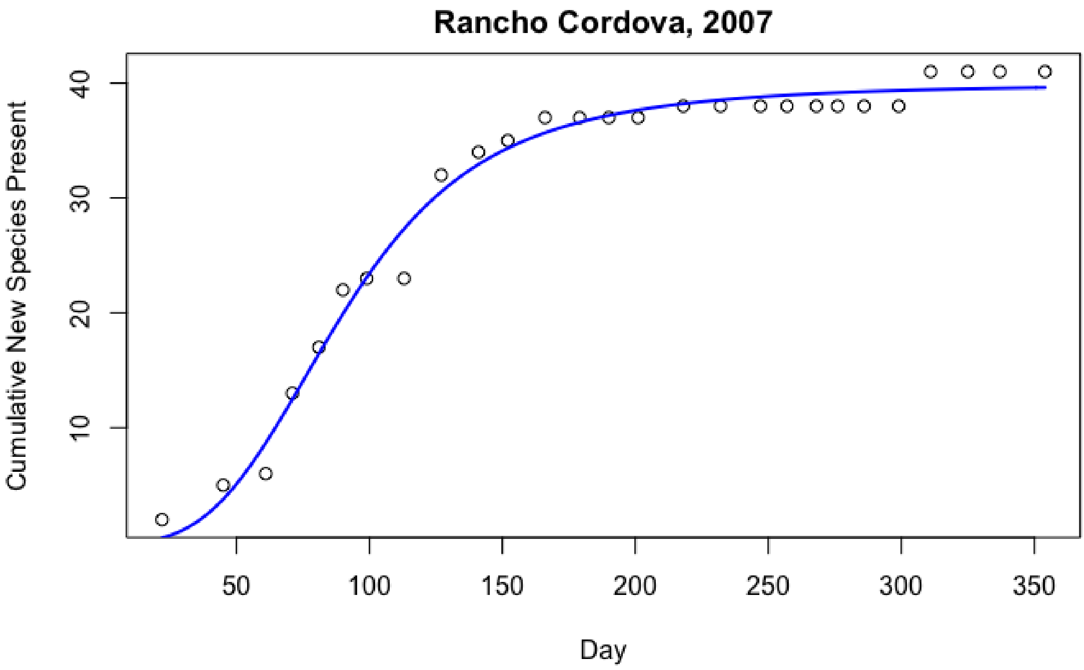



Figure 4.27 shows the SLL curve for Rancho Cordova for 2007. The parameters for that year are: $k = 1.06$, $\sigma = 3.37$, $d = 39.94$ and $\mu = 93.28$. MSE = 2.126 and MAE = 1.19

Figure 4.28: SLL species accumulation curve - Rancho Cordova 2015

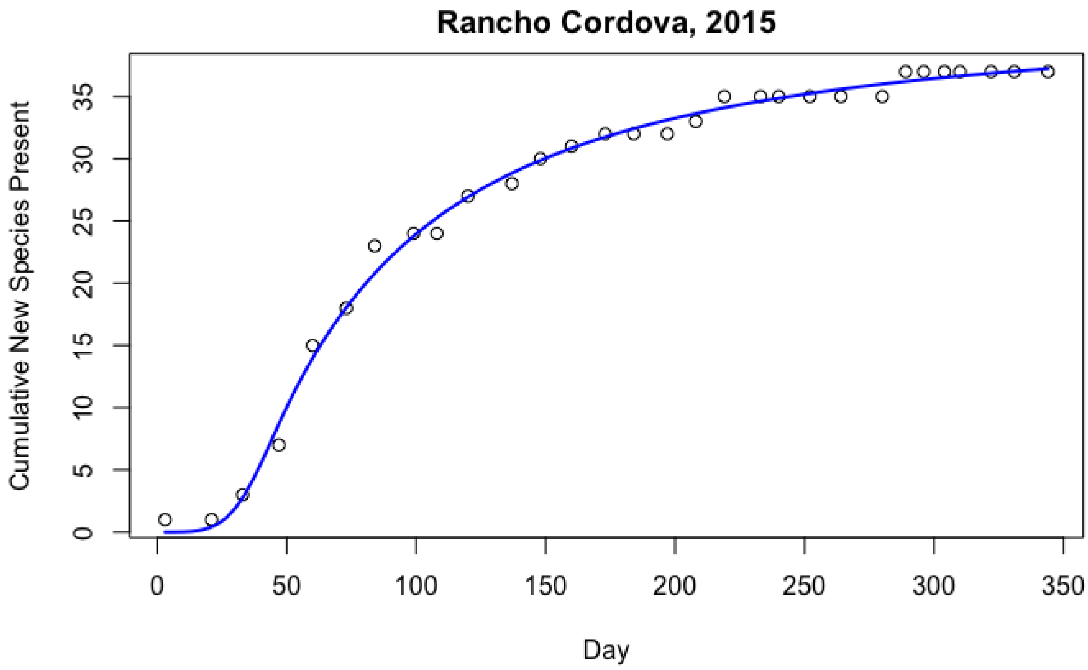



Figure 4.28 shows the SLL curve for Rancho Cordova for 2015. The parameters for that year are: $k = 0.55$, $\sigma = 2.49$, $d = 41.42$ and $\mu = 48.90$. MSE = 0.611 and MAE = 0.59

The figures provided show examples of a skewness parameter $k$ greater than and less than 1. Figure 4.28 has a skewness parameter equal to 0.55 and Figure 4.26 has $k = 1.25$. The range of MAE for all years for the Rancho Cordova site is 0.29 to 1.83. This max error of 1.83 demonstrates a relatively small error when comparing the estimated number of cumulative species in the SLL curve to the observed number of species.

**Fitted parameters and the species accumulation curves for Washington**

Table 4.10 is comprised of the fitted parameters for the Washington site as well as the MSE and MAE for each year to demonstrate the goodness of fit for each year. The table is formatted on a yearly basis displaying the SLL parameters for the years 1988 - 2016. We have chosen to provide one SLL curve from each decade during the observed period for this site to demonstrate various goodness of fit.

Table 4.10: Washington fitted parameters

| Year | $k$ | $\sigma$ | $d$ | $\mu$ | Visits | MSE | MAE |
|---|---|---|---|---|---|---|---|
| 1988 | 0.09 | 9.86 | 85.46 | 44.28 | 13 | 3.60 | 1.48 |
| 1989 | 1.26 | 2.97 | 74.58 | 155.24 | 15 | 3.67 | 1.49 |
| 1990 | 0.52 | 4.13 | 82.09 | 91.94 | 14 | 3.59 | 1.45 |
| 1991 | 0.87 | 4.98 | 64.19 | 170.10 | 13 | 0.79 | 0.68 |
| 1992 | 1.21 | 4.20 | 78.63 | 114.58 | 20 | 1.44 | 0.88 |
| 1993 | 0.57 | 3.39 | 80.20 | 103.72 | 19 | 2.30 | 1.27 |
| 1994 | 0.89 | 4.70 | 77.13 | 113.28 | 20 | 4.73 | 1.69 |
| 1995 | 2.48 | 7.18 | 67.41 | 208.95 | 21 | 4.47 | 1.61 |
| 1996 | 0.59 | 3.83 | 87.70 | 89.20 | 21 | 2.19 | 1.16 |
| 1997 | 1.14 | 3.52 | 73.09 | 125.75 | 22 | 3.41 | 1.29 |
| 1998 | 1.32 | 4.90 | 65.23 | 189.49 | 20 | 6.75 | 1.93 |
| 1999 | 0.80 | 4.37 | 66.41 | 114.19 | 20 | 1.48 | 1.06 |
| 2000 | 0.50 | 4.67 | 87.62 | 79.45 | 20 | 3.80 | 1.45 |
| 2001 | 2.81 | 7.91 | 69.05 | 174.01 | 25 | 3.09 | 1.44 |
| 2002 | 1.23 | 5.10 | 68.93 | 154.39 | 19 | 4.54 | 1.64 |
| 2003 | 2.84 | 8.02 | 72.06 | 208.78 | 23 | 13.36 | 2.89 |
| 2004 | 1.16 | 3.87 | 73.68 | 132.68 | 17 | 4.37 | 1.56 |
| 2005 | 1.47 | 4.31 | 60.09 | 168.28 | 18 | 3.26 | 1.45 |
| 2006 | 0.88 | 8.94 | 61.26 | 140.31 | 20 | 4.65 | 1.67 |
| 2007 | 1.46 | 4.13 | 62.75 | 141.03 | 19 | 1.26 | 0.78 |
| 2008 | 1.30 | 4.13 | 61.78 | 146.46 | 20 | 2.14 | 1.04 |
| 2009 | 1.31 | 5.80 | 65.01 | 149.45 | 20 | 3.56 | 1.51 |
| 2010 | 1.64 | 5.10 | 66.14 | 173.11 | 18 | 1.32 | 0.91 |
| 2011 | 1.96 | 7.36 | 62.14 | 191.68 | 19 | 2.12 | 0.93 |
| 2012 | 2.13 | 5.26 | 65.20 | 180.49 | 23 | 2.32 | 0.97 |
| 2013 | 1.16 | 4.25 | 63.55 | 126.05 | 21 | 1.61 | 0.92 |
| 2014 | 1.73 | 5.40 | 62.76 | 155.51 | 21 | 2.74 | 1.29 |
| 2015 | 0.76 | 2.89 | 58.40 | 83.30 | 20 | 1.03 | 0.81 |
| 2016 | 0.93 | 3.74 | 64.46 | 121.82 | 21 | 1.05 | 0.84 |

Figure 4.29: SLL species accumulation curve - Washington 1989

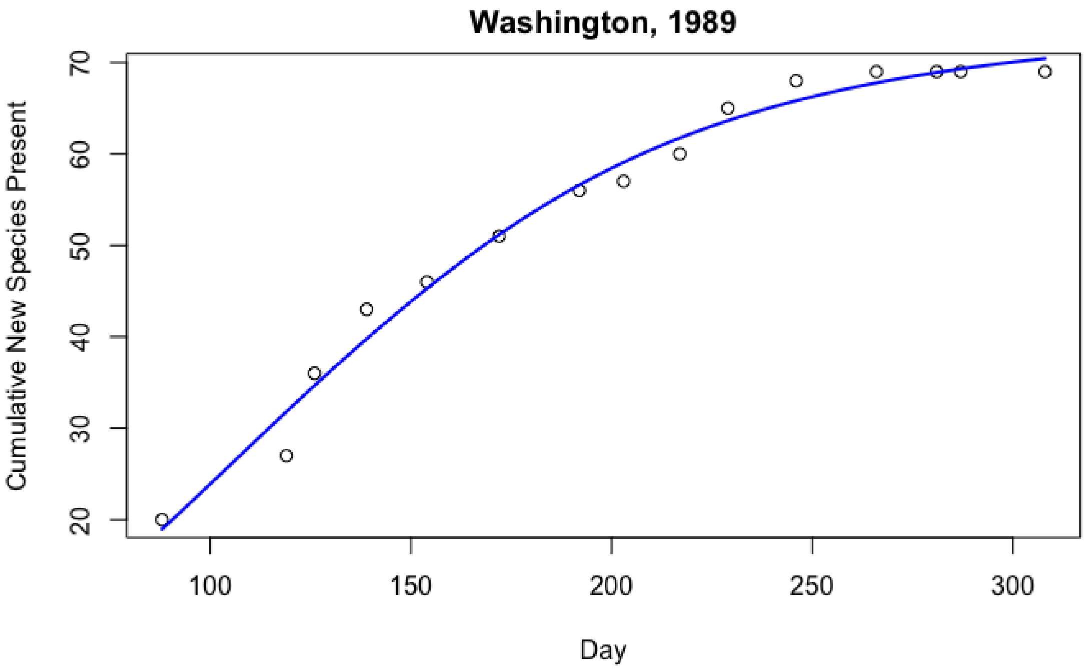



Figure 4.29 shows the SLL curve for Washington for 1989. The parameters for that year are: $k = 1.26$, $\sigma = 2.97$, $d = 74.58$ and $\mu = 155.24$. MSE = 3.666 and MAE = 1.49

Figure 4.30: SLL species accumulation curve - Washington 1997

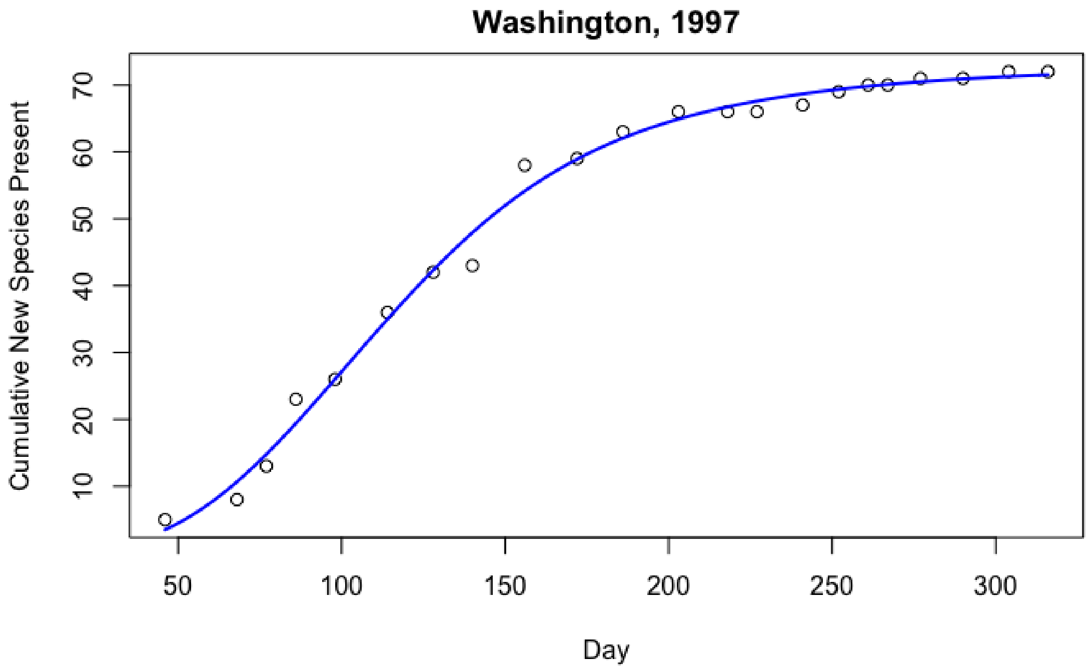



Figure 4.30 shows the SLL curve for Washington for 1997. The parameters for that year are: $k = 1.14$, $\sigma = 3.52$, $d = 73.09$ and $\mu = 125.75$. MSE = 3.412 and MAE = 1.29

Figure 4.31: SLL species accumulation curve - Washington 2009

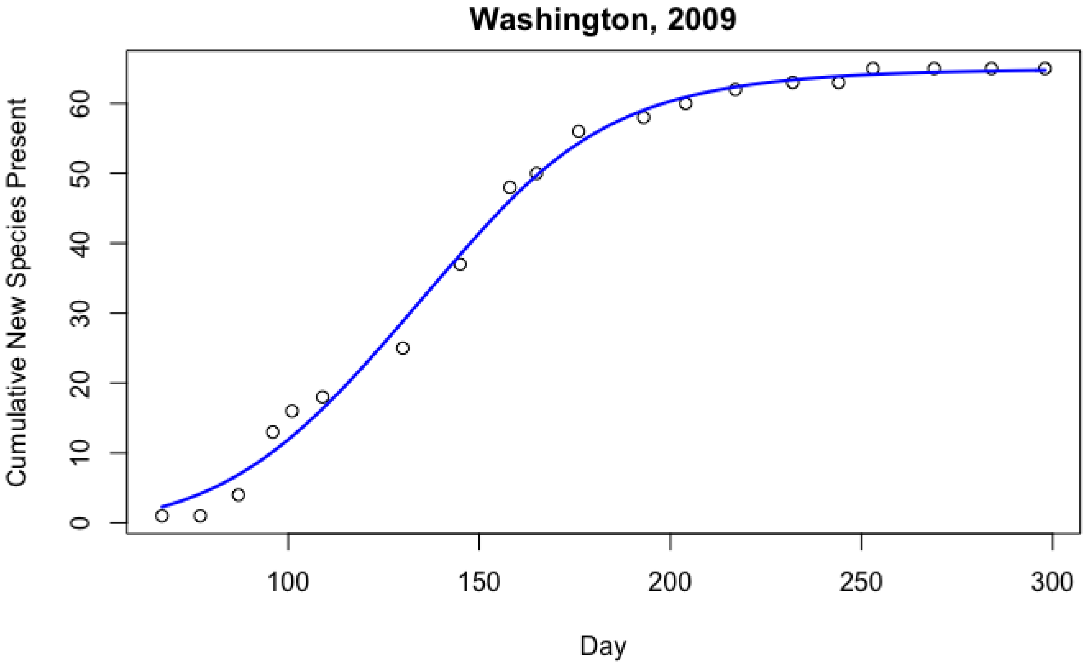



Figure 4.31 shows the SLL curve for Washington for 2009. The parameters for that year are: $k = 1.31$, $\sigma = 5.80$, $d = 65.01$ and $\mu = 149.45$. MSE = 3.564 and MAE = 1.51

Figure 4.32: SLL species accumulation curve - Washington 2014

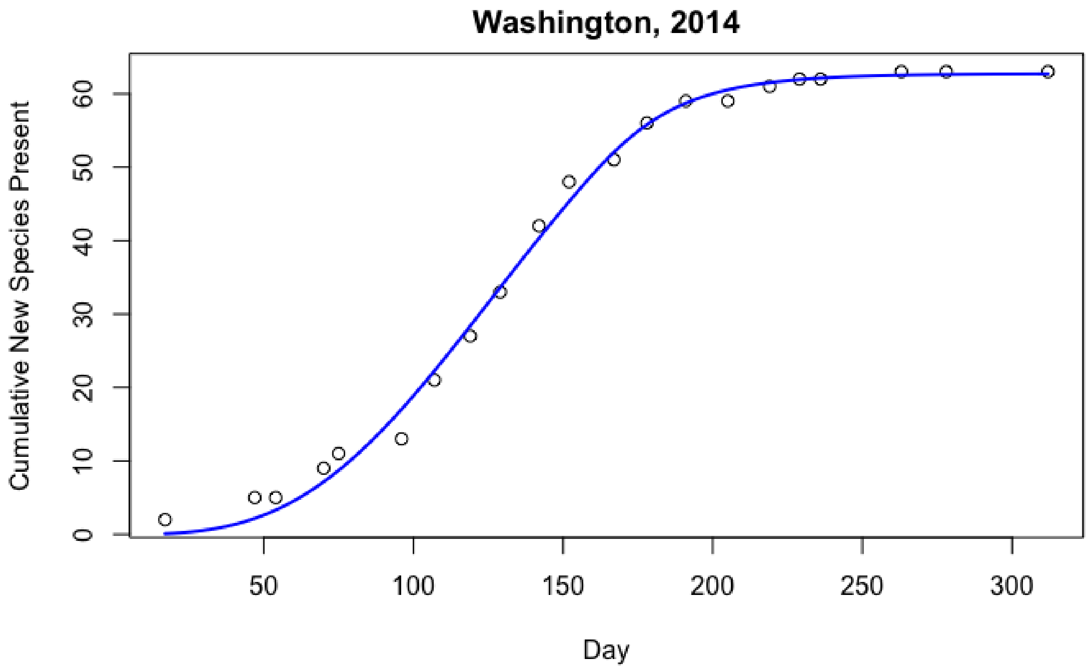



Figure 4.32 shows the SLL curve for Washington for 2014. The parameters for that year are: $k = 1.726$, $\sigma = 5.40$, $d = 62.76$ and $\mu = 155.51$. MSE = 2.736 and MAE = 1.29

The range of MAE for all years for the Washington site is 0.68 to 2.89. The interpretation of MAE shows that at most, the difference in number of cumulative species between the SLL curve and the observed number of species is equal to 2.89 species. This range of MAE shows the effectiveness in the fit of our SLL curve in modeling the accumulated number of species for the Washington site.

**Fitted parameters and the species accumulation curves for Lang Crossing**

Tables 4.11 and 4.12 are comprised of the fitted parameters for the Lang Crossing site as well as the MSE and MAE for each year to demonstrate the goodness of fit for each year. The tables are formatted on a yearly basis displaying the SLL parameters for the years 1974 - 2016. Please note that for the year 1979 the nlsLM function was unable to provide an accurate estimate of the SLL parameters. This is due to the linear shape of the data with no curvature and no point of inflection. We have chosen to provide one SLL curve from each decade during the observed period for this site to demonstrate various goodness of fit.

Table 4.11: Lang Crossing fitted parameters

| Year | $k$ | $\sigma$ | $d$ | $\mu$ | Visits | MSE | MAE |
|---|---|---|---|---|---|---|---|
| 1974 | 0.47 | 10.75 | 82.19 | 117.88 | 16 | 2.61 | 1.06 |
| 1975 | 0.68 | 10.13 | 71.88 | 150.18 | 15 | 4.43 | 1.81 |
| 1976 | 1.88 | 6.95 | 47.12 | 187.50 | 6 | 0.10 | 0.25 |
| 1977 | 1.74 | 4.89 | 67.82 | 182.28 | 12 | 5.36 | 1.83 |
| 1978 | 1.28 | 8.01 | 71.87 | 160.25 | 15 | 9.71 | 2.18 |
| 1979 | - | - | - | - | 7 | 11.52 | 2.69 |
| 1980 | 2.26 | 8.51 | 74.30 | 197.96 | 12 | 1.47 | 1.05 |
| 1981 | 1.96 | 8.65 | 66.38 | 172.66 | 7 | 0.12 | 0.30 |
| 1982 | 0.93 | 9.18 | 69.61 | 156.83 | 12 | 0.85 | 0.69 |
| 1983 | 0.11 | 71.04 | 61.37 | 141.90 | 10 | 2.28 | 1.14 |
| 1984 | 0.69 | 6.91 | 72.98 | 128.69 | 14 | 7.60 | 2.11 |
| 1985 | 1.32 | 6.73 | 77.37 | 162.06 | 16 | 1.53 | 0.98 |
| 1986 | 1.26 | 7.22 | 69.93 | 154.15 | 14 | 6.51 | 1.70 |
| 1987 | 2.00 | 7.46 | 63.03 | 191.73 | 12 | 0.69 | 0.58 |
| 1988 | 2.56 | 5.27 | 84.21 | 214.24 | 19 | 6.18 | 1.78 |
| 1989 | 0.92 | 5.82 | 77.67 | 142.29 | 17 | 7.48 | 2.07 |
| 1990 | 1.66 | 5.20 | 73.22 | 182.75 | 16 | 5.51 | 1.78 |
| 1991 | 1.27 | 10.45 | 71.39 | 174.85 | 15 | 2.45 | 1.13 |
| 1992 | 0.55 | 11.06 | 83.34 | 111.10 | 16 | 2.69 | 1.32 |
| 1993 | 0.45 | 9.38 | 72.85 | 129.11 | 17 | 2.38 | 1.16 |
| 1994 | 0.84 | 6.98 | 75.91 | 134.70 | 19 | 3.47 | 1.39 |
| 1995 | 0.51 | 10.45 | 76.14 | 150.64 | 17 | 6.55 | 2.06 |
| 1996 | 0.68 | 7.61 | 79.50 | 128.93 | 17 | 2.67 | 1.16 |
| 1997 | 0.92 | 5.59 | 69.27 | 129.22 | 15 | 6.67 | 1.86 |
| 1998 | 1.33 | 9.44 | 63.29 | 198.07 | 15 | 3.61 | 1.48 |
| 1999 | 1.10 | 9.53 | 70.39 | 161.56 | 16 | 3.06 | 1.21 |

Table 4.12: Lang Crossing fitted parameters continued

| Year | $k$ | $\sigma$ | $d$ | $\mu$ | Visits | MSE | MAE |
|---|---|---|---|---|---|---|---|
| 2000 | 1.32 | 7.36 | 70.05 | 162.72 | 19 | 0.84 | 0.71 |
| 2001 | 0.74 | 8.86 | 75.98 | 134.90 | 19 | 3.69 | 1.53 |
| 2002 | 1.24 | 9.80 | 68.90 | 171.61 | 16 | 1.55 | 0.88 |
| 2003 | 0.82 | 9.89 | 71.39 | 153.18 | 20 | 5.23 | 1.40 |
| 2004 | 1.30 | 7.61 | 74.08 | 150.08 | 17 | 1.21 | 0.72 |
| 2005 | 2.00 | 9.83 | 68.18 | 190.71 | 13 | 4.08 | 1.49 |
| 2006 | 1.03 | 9.70 | 63.91 | 158.42 | 18 | 3.09 | 1.15 |
| 2007 | 0.95 | 8.13 | 73.22 | 140.98 | 17 | 4.22 | 1.47 |
| 2008 | 1.33 | 6.77 | 62.76 | 165.40 | 14 | 1.18 | 0.79 |
| 2009 | 1.06 | 8.47 | 66.28 | 150.93 | 15 | 3.93 | 1.33 |
| 2010 | 1.22 | 10.67 | 67.17 | 178.63 | 17 | 2.64 | 1.14 |
| 2011 | 1.24 | 10.23 | 71.99 | 186.87 | 17 | 3.38 | 1.32 |
| 2012 | 0.94 | 7.27 | 77.44 | 147.66 | 20 | 1.89 | 1.08 |
| 2013 | 0.86 | 6.35 | 74.12 | 123.18 | 18 | 3.95 | 1.59 |
| 2014 | 1.17 | 6.82 | 56.14 | 140.97 | 18 | 1.88 | 1.06 |
| 2015 | 2.17 | 5.84 | 59.87 | 171.05 | 18 | 2.14 | 1.10 |
| 2016 | 1.22 | 10.67 | 67.17 | 178.63 | 16 | 2.64 | 1.14 |

Figure 4.33: SLL species accumulation curve - Lang Crossing 1975

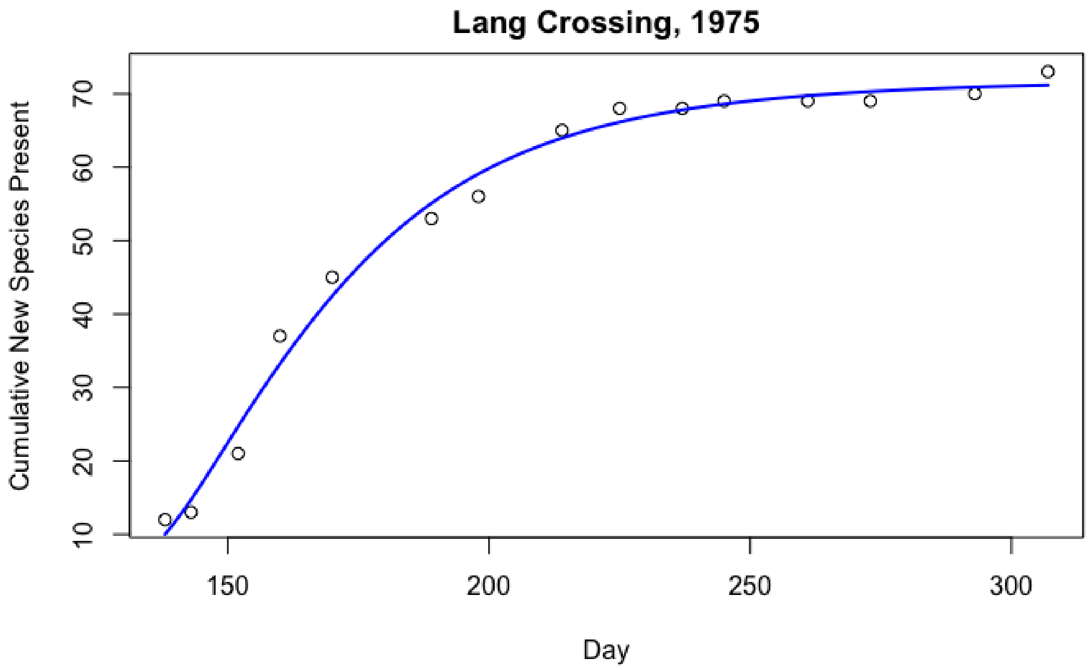



Figure 4.33 shows the SLL curve for Lang Crossing for 1975. The parameters for that year are: $k = 0.68$, $\sigma = 10.13$, $d = 71.88$ and $\mu = 150.18$. MSE = 4.434 and MAE = 1.81

Figure 4.34: SLL species accumulation curve - Lang Crossing 1984

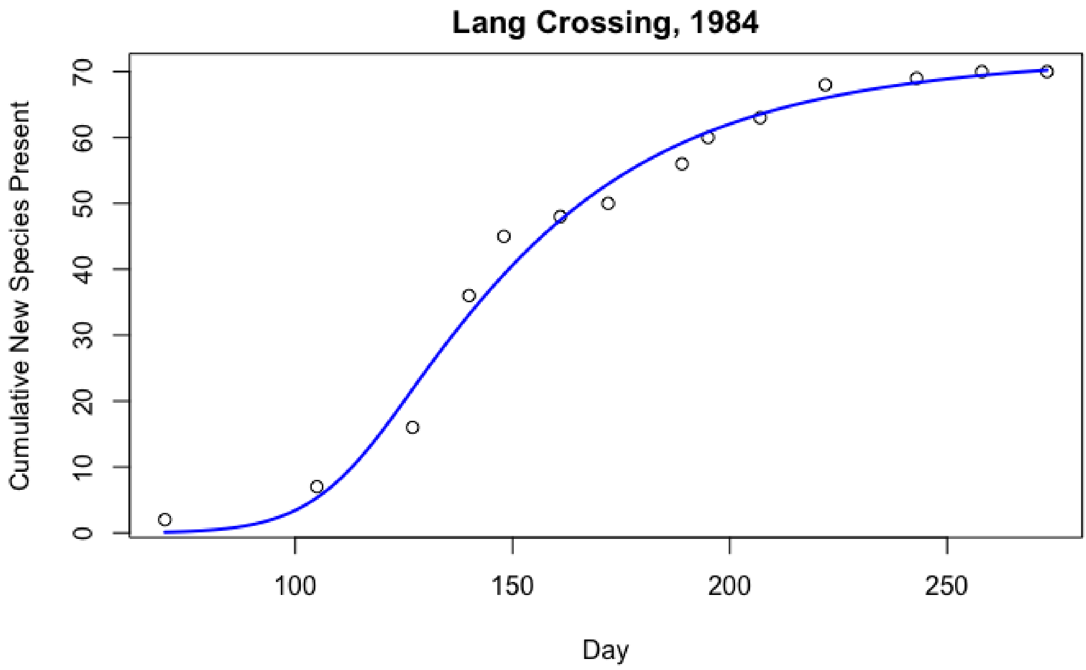



Figure 4.34 shows the SLL curve for Lang Crossing for 1984. The parameters for that year are: $k = 0.69$, $\sigma = 6.91.17$, $d = 72.98$ and $\mu = 128.69$. MSE = 7.599 and MAE = 2.11

Figure 4.35: SLL species accumulation curve - Lang Crossing 1999

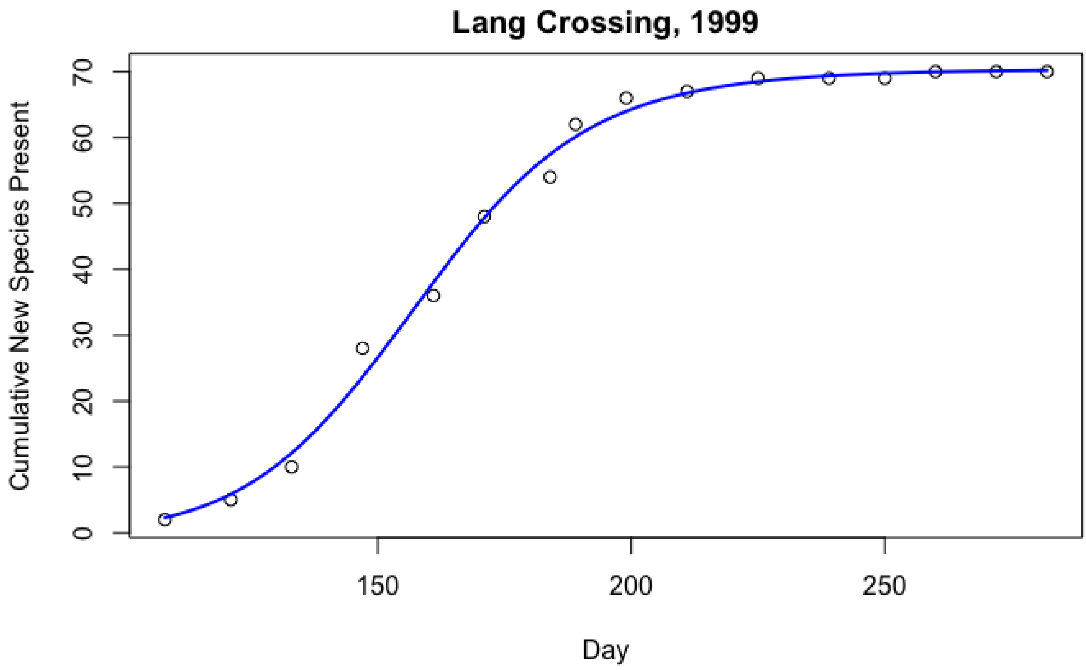



Figure 4.35 shows the SLL curve for Lang Crossing for 1999. The parameters for that year are: $k = 1.10$, $\sigma = 9.53$, $d = 70.39$ and $\mu = 161.56$. MSE = 3.055 and MAE = 1.21

Figure 4.36: SLL species accumulation curve - Lang Crossing 2003

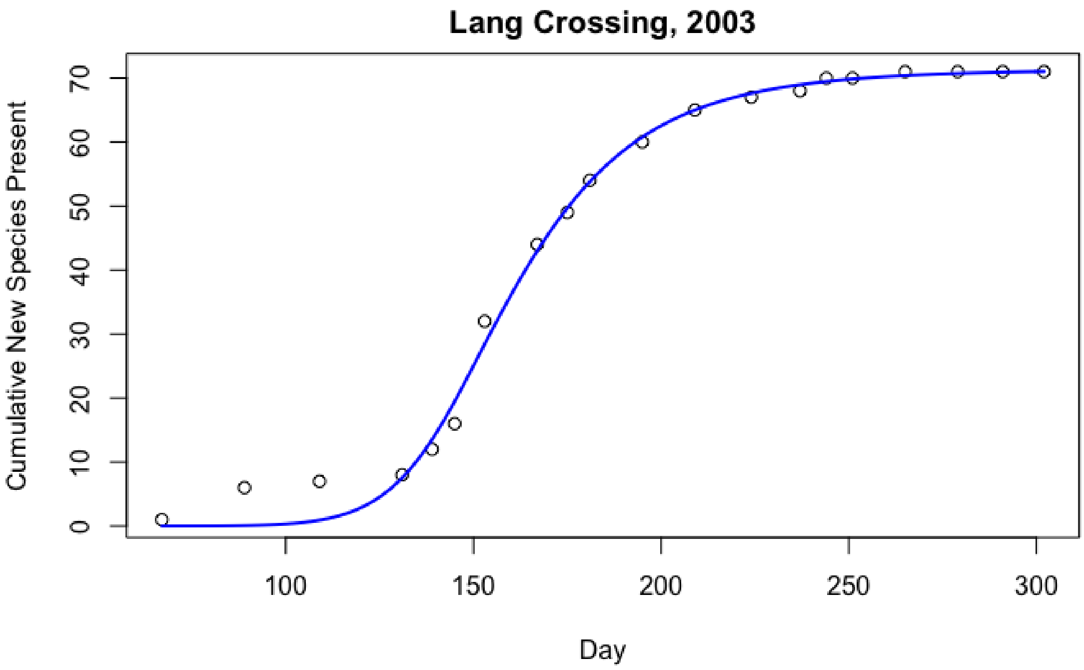



Figure 4.36 shows the SLL curve for Lang Crossing for 2003. The parameters for that year are: $k = 0.82$, $\sigma = 9.89$, $d = 71.39$ and $\mu = 153.18$. MSE = 5.225 and MAE = 1.40

Figure 4.37: SLL species accumulation curve - Lang Crossing 2014

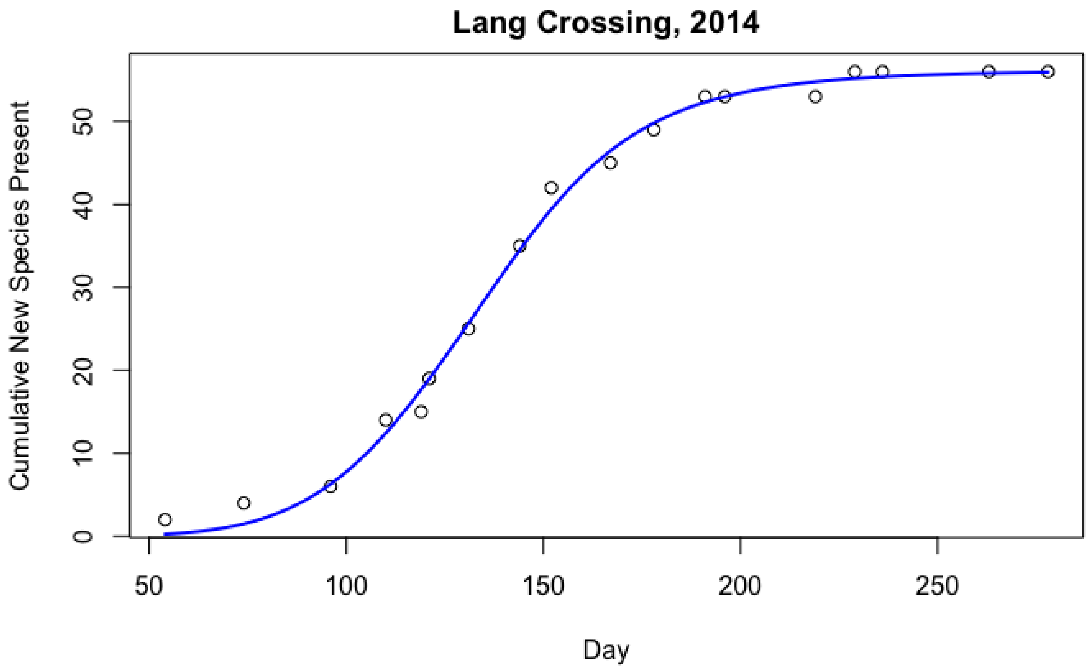



Figure 4.37 shows the SLL curve for Lang Crossing for 2014. The parameters for that year are: $k = 1.17$, $\sigma = 6.82$, $d = 56.14$ and $\mu = 140.97$. MSE = 1.884 and MAE = 1.06

We have chosen to select these years as examples to provide examples of years with a skewness parameter greater than and less than 1. Figure 4.34 has a skewness parameter equal to 0.69, while Figure 4.37 has $k = 1.17$. The range of MAE for the Lang Crossing site is 0.25 to 2.69. This relatively low range shows the effectiveness of the fit of the SLL curve in estimating the accumulated number of species for this site.

**Fitted parameters and the species accumulation curves for Sierra Valley**

Tables 4.13 and 4.14 are comprised of the fitted parameters for the Sierra Valley site as well as the MSE and MAE for each year to demonstrate the goodness of fit for each year. The tables are formatted on a yearly basis displaying the SLL parameters for the years 1982 - 2016. Please note that for the year 1995 the nlsLM function was unable to provide an accurate estimate of the skewness coefficient and scale parameters. This is due to repetition of the number of observed species at and beyond the point of inflection causing no curvature but a sharp plateau. We have chosen to provide one SLL curve from each decade during the observed period for this site to demonstrate various goodness of fit.

Table 4.13: Sierra Valley fitted parameters

| Year | $k$ | $\sigma$ | $d$ | $\mu$ | Visits | MSE | MAE |
|---|---|---|---|---|---|---|---|
| 1982 | 3.49 | 8.95 | 48.07 | 236.62 | 9 | 0.04 | 0.13 |
| 1983 | 0.20 | 29.21 | 44.08 | 143.82 | 7 | 0.04 | 0.17 |
| 1984 | 0.05 | 29.22 | 54.58 | 80.40 | 5 | 0.65 | 0.71 |
| 1985 | 2.00 | 5.48 | 64.19 | 195.25 | 10 | 3.82 | 1.38 |
| 1986 | 3.94 | 9.30 | 61.97 | 216.80 | 10 | 6.83 | 1.67 |
| 1987 | 0.83 | 4.00 | 65.11 | 144.87 | 13 | 0.97 | 0.88 |
| 1988 | 0.51 | 5.23 | 69.27 | 102.17 | 14 | 1.53 | 0.98 |
| 1989 | 0.61 | 5.49 | 59.29 | 116.95 | 13 | 1.78 | 1.08 |
| 1990 | 0.56 | 4.69 | 68.57 | 108.34 | 19 | 4.17 | 1.59 |
| 1991 | 0.80 | 5.90 | 60.33 | 143.87 | 17 | 1.79 | 0.99 |
| 1992 | 0.73 | 6.84 | 58.45 | 116.18 | 18 | 1.34 | 0.81 |
| 1993 | 0.68 | 7.37 | 62.64 | 137.39 | 17 | 2.59 | 1.23 |
| 1994 | 0.49 | 6.03 | 66.54 | 103.29 | 16 | 1.25 | 0.99 |
| 1995 | - | - | 51.10 | 220.70 | 14 | 8.85 | 2.45 |
| 1996 | 0.86 | 5.98 | 66.43 | 134.70 | 16 | 2.48 | 1.23 |
| 1997 | 1.52 | 4.89 | 62.52 | 168.44 | 14 | 2.36 | 1.10 |
| 1998 | 2.78 | 10.55 | 61.65 | 203.60 | 16 | 1.02 | 0.80 |
| 1999 | 1.08 | 6.35 | 58.98 | 155.72 | 17 | 1.04 | 0.84 |
| 2000 | 1.20 | 4.78 | 67.93 | 156.89 | 19 | 3.85 | 1.51 |
| 2001 | 0.73 | 8.31 | 62.52 | 130.37 | 17 | 1.10 | 0.84 |
| 2002 | 0.95 | 6.09 | 68.16 | 139.78 | 19 | 2.28 | 1.18 |
| 2003 | 1.55 | 8.63 | 64.15 | 171.52 | 17 | 0.76 | 0.58 |
| 2004 | 1.67 | 5.53 | 65.53 | 170.86 | 19 | 7.56 | 2.01 |

Table 4.14: Sierra Valley fitted parameters continued

| Year | $k$ | $\sigma$ | $d$ | $\mu$ | Visits | MSE | MAE |
|---|---|---|---|---|---|---|---|
| 2005 | 1.22 | 6.45 | 59.68 | 170.81 | 16 | 0.87 | 0.74 |
| 2006 | 1.17 | 8.26 | 56.28 | 162.09 | 17 | 3.91 | 1.28 |
| 2007 | 1.80 | 7.22 | 56.82 | 167.06 | 16 | 2.23 | 1.06 |
| 2008 | 0.69 | 5.39 | 59.80 | 131.39 | 17 | 0.69 | 0.66 |
| 2009 | 0.92 | 4.97 | 56.46 | 152.19 | 18 | 1.25 | 0.95 |
| 2010 | 2.27 | 11.64 | 54.76 | 197.25 | 18 | 2.21 | 1.04 |
| 2011 | 1.23 | 7.46 | 56.57 | 178.93 | 17 | 1.68 | 1.01 |
| 2012 | 0.65 | 5.79 | 61.62 | 127.78 | 19 | 3.48 | 1.49 |
| 2013 | 1.18 | 4.04 | 49.93 | 162.82 | 18 | 2.83 | 1.37 |
| 2014 | 1.21 | 6.25 | 58.99 | 150.32 | 20 | 1.60 | 0.93 |
| 2015 | 1.67 | 5.43 | 47.58 | 171.61 | 20 | 4.72 | 1.71 |
| 2016 | 1.90 | 7.30 | 52.85 | 181.86 | 19 | 3.75 | 1.26 |

Figure 4.38: SLL species accumulation curve - Sierra Valley 1988

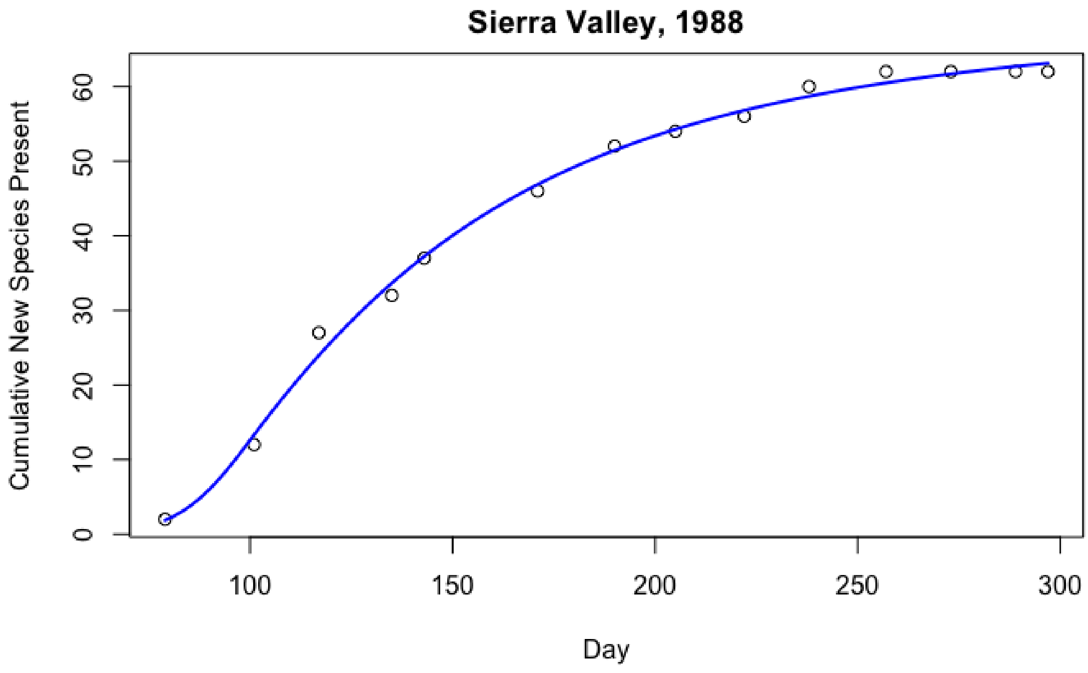



Figure 4.38 shows the SLL curve for Sierra Valley for 1988. The parameters for that year are: $k = 0.51$, $\sigma = 5.23$, $d = 69.27$ and $\mu = 102.17$. MSE = 1.525 and MAE = 0.98

Figure 4.39: SLL species accumulation curve - Sierra Valley 1992

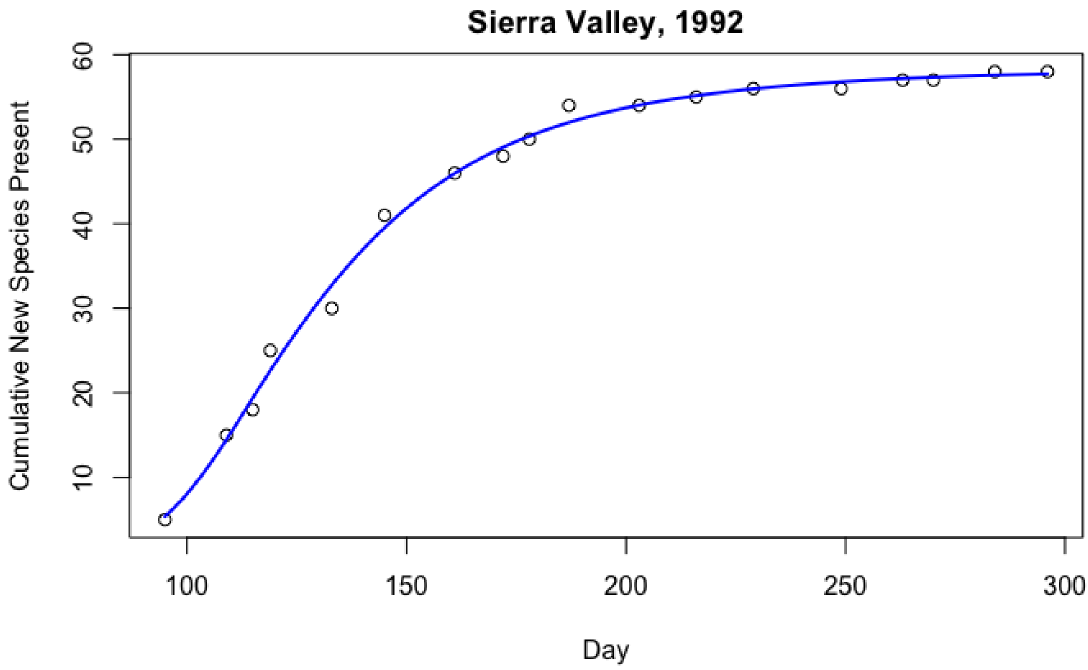



Figure 4.39 shows the SLL curve for Sierra Valley for 1992. The parameters for that year are: $k = 0.73$, $\sigma = 6.84$, $d = 58.45$ and $\mu = 116.18$. MSE = 1.336 and MAE = 0.81

Figure 4.40: SLL species accumulation curve - Sierra Valley 2005

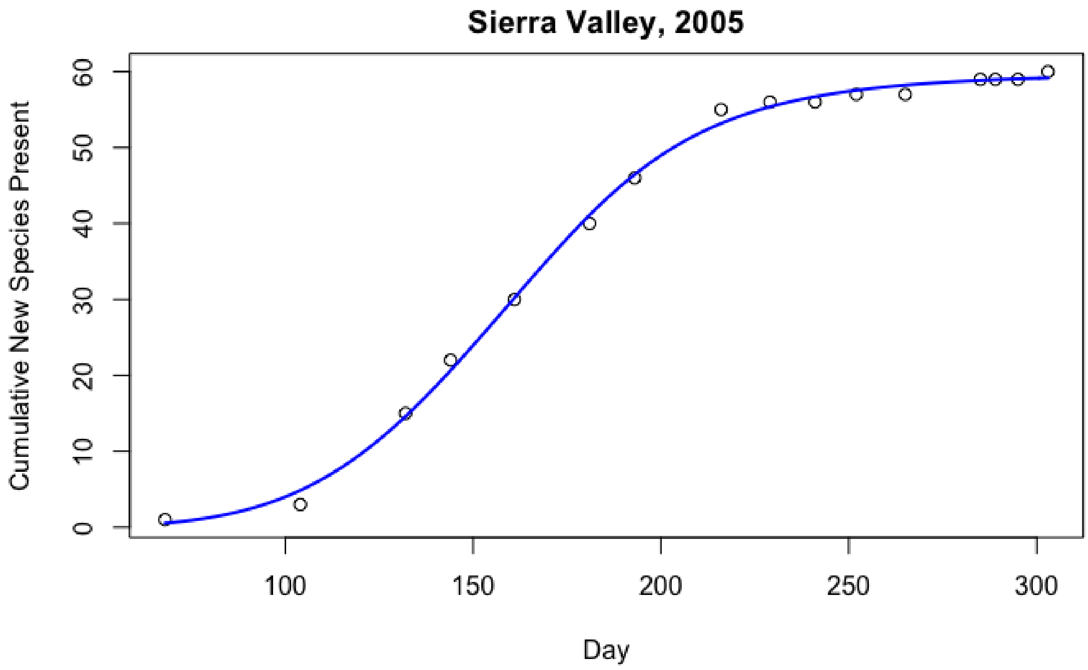



Figure 4.40 shows the SLL curve for Sierra Valley for 2005. The parameters for that year are: $k = 1.22$, $\sigma = 6.45$, $d = 59.68$ and $\mu = 170.81$. MSE = 0.869 and MAE = 0.74

Figure 4.41: SLL species accumulation curve - Sierra Valley 2014

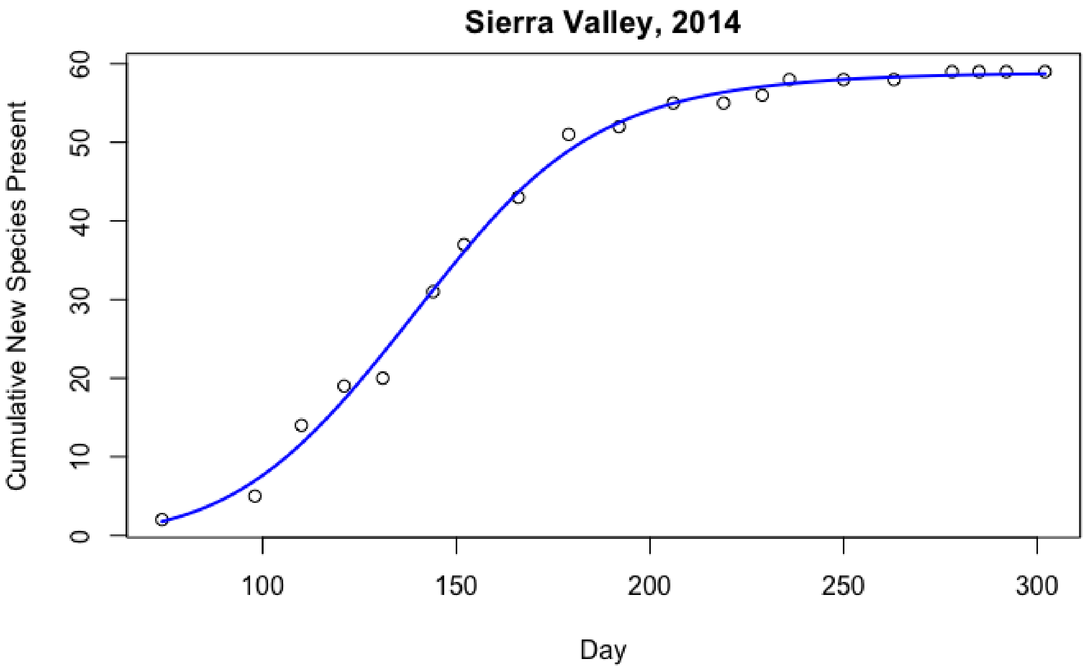



Figure 4.41 shows the SLL curve for Sierra Valley for 2014. The parameters for that year are: $k = 1.21$, $\sigma = 6.25$, $d = 58.99$ and $\mu = 150.32$. MSE = 1.596 and MAE = 0.93

We have chosen to select these years as examples to provide examples of years with a skewness parameter greater than 1 and less than 1. Figure 4.38 has a skewness parameter equal to 0.51, while Figure 4.41 has $k = 1.21$. The MAE range for this site is 0.13 to 2.45. The MAE can be interpreted as the maximum difference between the estimated number of accumulated species and the observed number of species for this site to be 2.45. This relatively low error shows the effectiveness of the fit of the SLL curve in modeling the accumulated number of species for this site.

**Fitted parameters and the species accumulation curves for Donner Pass**

Tables 4.15 and 4.16 are comprised of the fitted parameters for the Donner Pass site as well as the MSE and MAE for each year to demonstrate the goodness of fit for each year. The tables are formatted on a yearly basis displaying the SLL parameters for the years 1973 - 2016. We have chosen to provide one SLL curve from each decade during the observed period for this site to demonstrate various goodness of fit.

Table 4.15: Donner Pass fitted parameters

| Year | $k$ | $\sigma$ | $d$ | $\mu$ | Visits | MSE | MAE |
|---|---|---|---|---|---|---|---|
| 1973 | 0.39 | 17.10 | 84.08 | 171.55 | 14 | 5.47 | 1.74 |
| 1974 | 1.26 | 9.75 | 74.21 | 185.95 | 16 | 1.69 | 0.96 |
| 1975 | 1.03 | 8.30 | 77.01 | 194.43 | 15 | 6.92 | 1.95 |
| 1976 | 0.74 | 7.94 | 81.95 | 147.49 | 23 | 4.00 | 1.46 |
| 1977 | 0.42 | 13.74 | 83.99 | 147.52 | 13 | 4.92 | 1.70 |
| 1978 | 1.09 | 11.01 | 73.62 | 191.27 | 13 | 4.44 | 1.74 |
| 1979 | 0.22 | 21.23 | 84.14 | 139.99 | 7 | 0.81 | 0.73 |
| 1980 | 1.14 | 11.73 | 78.11 | 192.17 | 11 | 2.54 | 1.11 |
| 1981 | 0.07 | 152.55 | 79.34 | 148.29 | 12 | 10.73 | 3.04 |
| 1982 | 0.07 | 95.90 | 77.23 | 154.72 | 10 | 4.76 | 1.72 |
| 1983 | 0.58 | 13.09 | 66.51 | 184.01 | 11 | 2.40 | 1.24 |
| 1984 | 1.25 | 10.82 | 68.45 | 181.20 | 12 | 6.23 | 2.03 |
| 1985 | 0.97 | 13.27 | 85.91 | 161.24 | 16 | 1.91 | 1.23 |
| 1986 | 0.63 | 11.95 | 79.64 | 155.04 | 11 | 1.21 | 0.89 |
| 1987 | 0.68 | 7.63 | 67.10 | 149.32 | 12 | 1.45 | 0.99 |
| 1988 | 0.92 | 6.12 | 91.73 | 164.99 | 18 | 7.53 | 2.04 |
| 1989 | 0.72 | 10.86 | 86.21 | 155.83 | 15 | 3.22 | 1.27 |
| 1990 | 1.64 | 10.68 | 80.24 | 192.32 | 22 | 8.06 | 1.96 |
| 1991 | 0.43 | 13.93 | 89.36 | 154.81 | 18 | 1.08 | 0.87 |
| 1992 | 0.54 | 13.62 | 89.01 | 127.03 | 20 | 1.95 | 0.90 |
| 1993 | 0.38 | 19.88 | 75.59 | 168.06 | 16 | 1.52 | 0.90 |
| 1994 | 0.79 | 9.77 | 70.53 | 154.27 | 13 | 2.35 | 1.13 |

Table 4.16: Donner Pass fitted parameters continued

| Year | $k$ | $\sigma$ | $d$ | $\mu$ | Visits | MSE | MAE |
|---|---|---|---|---|---|---|---|
| 1995 | 0.69 | 17.87 | 72.68 | 194.53 | 17 | 3.36 | 1.21 |
| 1996 | 0.07 | 126.11 | 76.11 | 144.55 | 13 | 2.15 | 0.98 |
| 1997 | 0.60 | 11.36 | 78.33 | 141.47 | 15 | 0.87 | 0.64 |
| 1998 | 0.52 | 18.75 | 78.17 | 176.61 | 15 | 0.95 | 0.78 |
| 1999 | 0.10 | 103.83 | 81.10 | 159.70 | 15 | 4.50 | 1.74 |
| 2000 | 0.11 | 58.56 | 78.73 | 138.75 | 17 | 1.07 | 0.87 |
| 2001 | 0.54 | 15.03 | 72.60 | 138.77 | 18 | 3.74 | 1.26 |
| 2002 | 0.16 | 75.67 | 69.37 | 154.94 | 13 | 10.24 | 2.69 |
| 2003 | 0.40 | 22.21 | 76.45 | 155.70 | 13 | 0.83 | 0.72 |
| 2004 | 1.22 | 9.67 | 70.20 | 170.82 | 13 | 0.85 | 0.72 |
| 2005 | 0.07 | 102.37 | 82.05 | 152.60 | 12 | 6.15 | 1.65 |
| 2006 | 0.05 | 195.39 | 75.70 | 157.73 | 13 | 6.24 | 2.00 |
| 2007 | 0.57 | 16.73 | 70.06 | 144.34 | 14 | 3.54 | 1.62 |
| 2008 | 0.06 | 174.39 | 79.87 | 151.35 | 17 | 5.78 | 1.95 |
| 2009 | 0.68 | 9.49 | 77.46 | 162.99 | 17 | 1.87 | 1.02 |
| 2010 | 0.10 | 98.09 | 70.94 | 167.07 | 14 | 0.44 | 0.52 |
| 2011 | 0.06 | 264.70 | 67.67 | 180.34 | 12 | 4.18 | 1.41 |
| 2012 | 0.89 | 10.63 | 64.60 | 159.60 | 19 | 0.36 | 0.47 |
| 2013 | 1.05 | 6.29 | 68.93 | 153.38 | 17 | 8.19 | 1.84 |
| 2014 | 0.78 | 10.65 | 73.94 | 144.83 | 18 | 1.21 | 0.80 |
| 2015 | 1.40 | 6.65 | 63.47 | 162.62 | 18 | 8.20 | 1.83 |
| 2016 | 0.05 | 167.54 | 65.78 | 145.67 | 14 | 5.54 | 1.89 |

Figure 4.42: SLL species accumulation curve - Donner Pass 1976

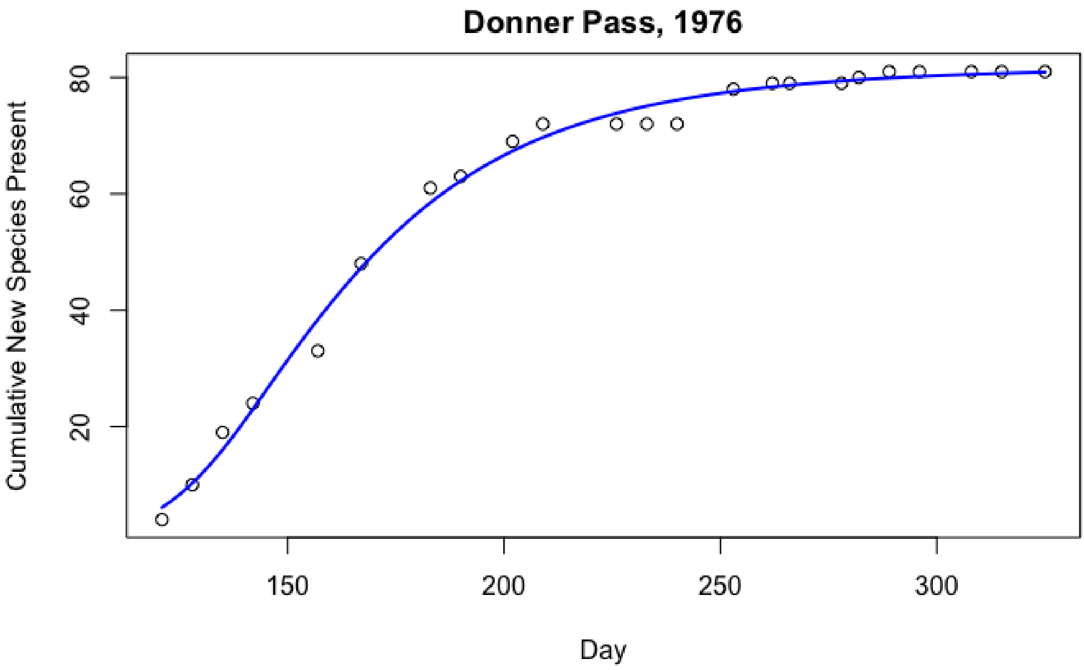



Figure 4.42 shows the SLL curve for Donner Pass for 1976. The parameters for that year are: $k = 0.74$, $\sigma = 7.94$, $d = 81.95$ and $\mu = 147.49$. MSE = 4.00 and MAE = 1.46

Figure 4.43: SLL species accumulation curve - Donner Pass 1985

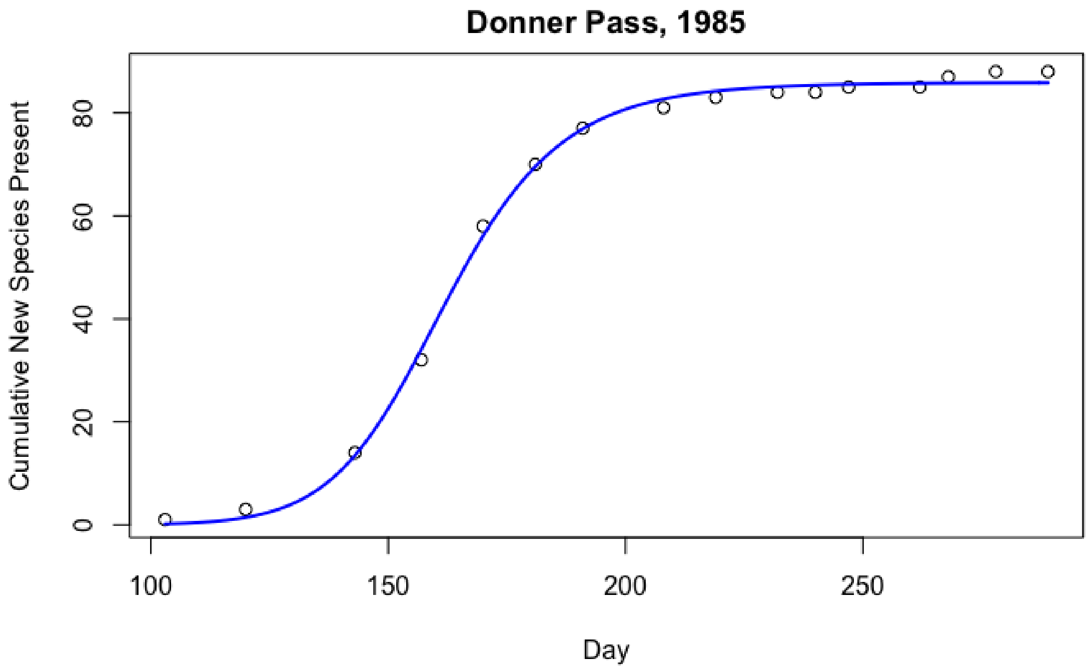



Figure 4.43 shows the SLL curve for Donner Pass for 1985. The parameters for that year are: $k = 0.97$, $\sigma = 13.27$, $d = 85.91$ and $\mu = 161.24$. MSE = 1.909 and MAE = 1.23

Figure 4.44: SLL species accumulation curve - Donner Pass 1998

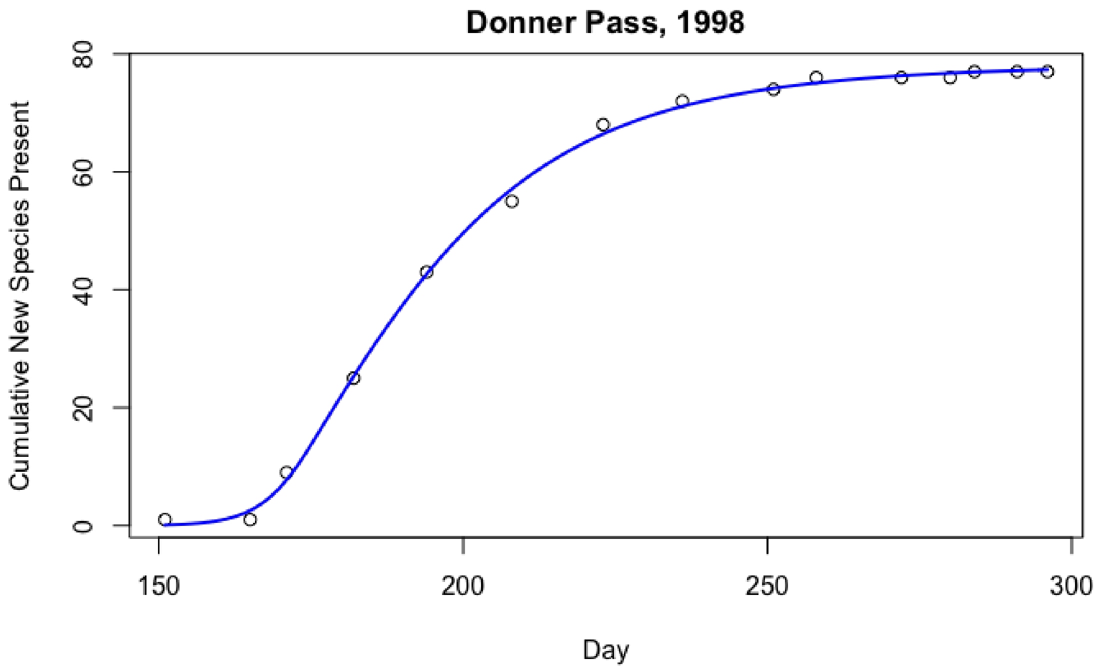



Figure 4.44 shows the SLL curve for Donner Pass for 1998. The parameters for that year are: $k = 0.52$, $\sigma = 18.75$, $d = 78.17$ and $\mu = 176.61$. MSE $= 0.953$ and MAE $= 0.78$

Figure 4.45: SLL species accumulation curve - Donner Pass 2009

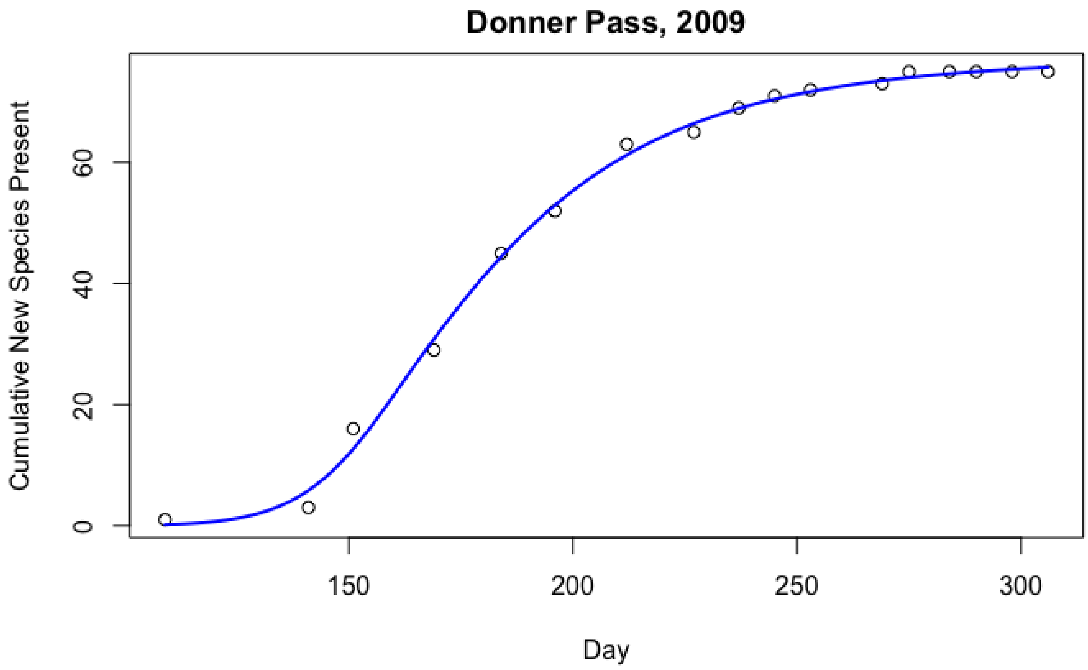



Figure 4.45 shows the SLL curve for Donner Pass for 2009. The parameters for that year are: $k = 0.68$, $\sigma = 9.49$, $d = 77.46$ and $\mu = 162.99$. MSE = 1.871 and MAPE = 1.02

Figure 4.46: SLL species accumulation curve - Donner Pass 2014

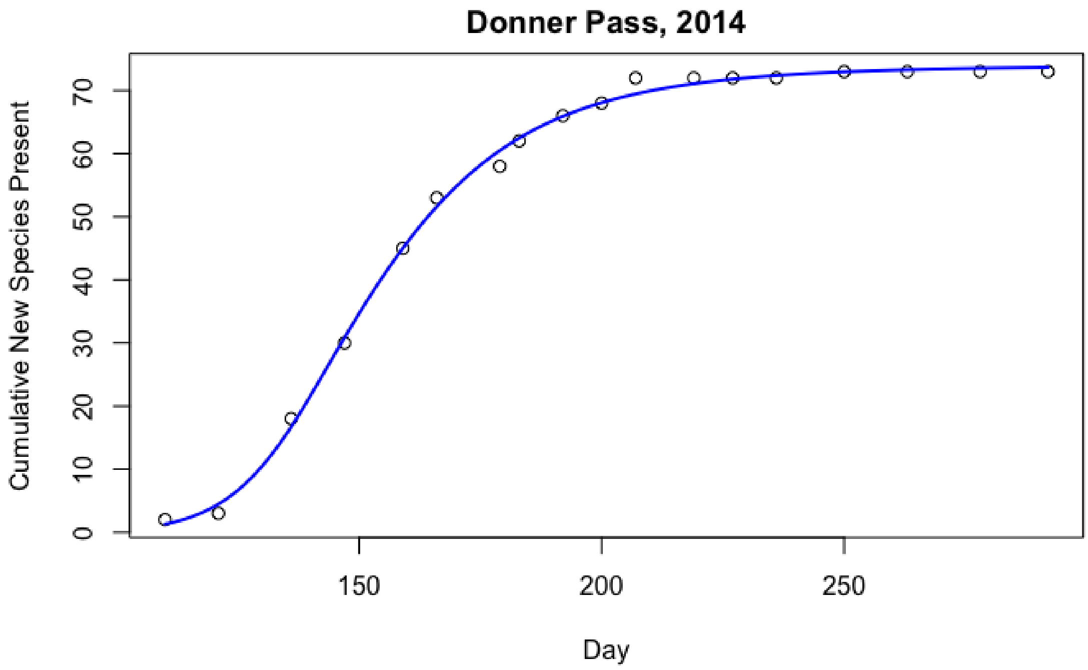



Figure 4.46 shows the SLL curve for Donner Pass for 2014. The parameters for that year are: $k$ = 0.78, $\sigma$ = 10.65, $d$ = 73.94 and $\mu$ = 144.83. MSE = 1.205 and MAE = 0.80

The MAE range for this site is 0.47 to 3.04 for all years analyzed. This relatively low error shows the effectiveness of the fit of the SLL curve in modeling the accumulated number of species for the Donner Pass site.

**Fitted parameters and the species accumulation curves for Castle Peak**

Table 4.17 is comprised of the fitted parameters for the Castle Peak site as well as the MSE and MAE for each year to demonstrate the goodness of fit for each year. The table is formatted on a yearly basis displaying the SLL parameters for the years 1989 - 2016. Please note that for the years 2000 and 2005 the nlsLM function was unable to provide an accurate estimate of the skewness coefficient and scale parameters. This is due to the linearity of data for the year 2005 and few data points for the year 2000 resulting in a lack of curvature at the point of inflection. We have chosen to provide one SLL curve from each decade during the observed period for this site to demonstrate various goodness of fit.

Table 4.17: Castle Peak fitted parameters

| Year | $k$ | $\sigma$ | $d$ | $\mu$ | Visits | MSE | MAE |
|---|---|---|---|---|---|---|---|
| 1989 | 0.09 | 141.22 | 66.79 | 170.35 | 10 | 9.82 | 2.59 |
| 1990 | 1.76 | 20.55 | 59.04 | 188.75 | 7 | 0.01 | 0.05 |
| 1991 | 0.09 | 112.47 | 66.20 | 163.15 | 10 | 1.83 | 1.09 |
| 1992 | 0.32 | 13.73 | 79.10 | 129.57 | 8 | 2.48 | 1.42 |
| 1993 | 0.11 | 78.18 | 59.33 | 164.01 | 8 | 2.68 | 1.30 |
| 1994 | 0.14 | 53.30 | 66.37 | 145.88 | 8 | 3.39 | 1.75 |
| 1995 | 1.74 | 20.04 | 56.59 | 226.00 | 7 | 1.85 | 1.06 |
| 1996 | 1.68 | 9.76 | 60.02 | 204.89 | 5 | 0.06 | 0.21 |
| 1997 | 0.10 | 91.67 | 67.55 | 155.34 | 9 | 4.56 | 1.61 |
| 1998 | 0.08 | 215.19 | 63.40 | 192.44 | 7 | 1.10 | 0.89 |
| 1999 | 0.09 | 226.17 | 52.06 | 188.11 | 8 | 0.30 | 0.36 |
| 2000 | - | - | - | - | 6 | 19.02 | 3.73 |
| 2001 | 0.45 | 38.33 | 66.55 | 163.17 | 9 | 5.00 | 1.98 |
| 2002 | 0.10 | 179.93 | 59.30 | 172.26 | 7 | 1.52 | 1.14 |
| 2003 | 0.07 | 257.54 | 61.30 | 172.96 | 8 | 5.24 | 2.11 |
| 2004 | 0.18 | 32.75 | 59.78 | 148.53 | 6 | 1.74 | 1.16 |
| 2005 | - | - | 67.01 | 247.99 | 6 | 0.13 | 0.26 |
| 2006 | 0.27 | 56.01 | 55.13 | 182.62 | 4 | 0.50 | 0.60 |
| 2007 | 0.18 | 52.97 | 59.51 | 146.79 | 6 | 0.22 | 0.41 |
| 2008 | 0.77 | 21.23 | 55.35 | 172.29 | 8 | 3.77 | 1.34 |
| 2009 | 0.07 | 130.80 | 59.78 | 168.82 | 7 | 3.64 | 1.64 |
| 2010 | 0.18 | 89.66 | 53.97 | 183.75 | 6 | 0.47 | 0.56 |
| 2011 | 0.06 | 349.19 | 44.13 | 203.17 | 5 | 5.21 | 1.95 |
| 2012 | 0.22 | 40.03 | 58.35 | 149.92 | 7 | 0.36 | 0.50 |
| 2013 | 1.33 | 12.11 | 59.07 | 174.34 | 7 | 1.84 | 1.15 |
| 2014 | 0.04 | 325.69 | 61.10 | 145.28 | 9 | 5.71 | 1.99 |
| 2015 | 0.18 | 23.63 | 60.37 | 123.37 | 7 | 3.50 | 1.63 |
| 2016 | 0.16 | 92.62 | 52.19 | 170.67 | 5 | 1.30 | 0.95 |

Figure 4.47: SLL species accumulation curve - Castle Peak 1989

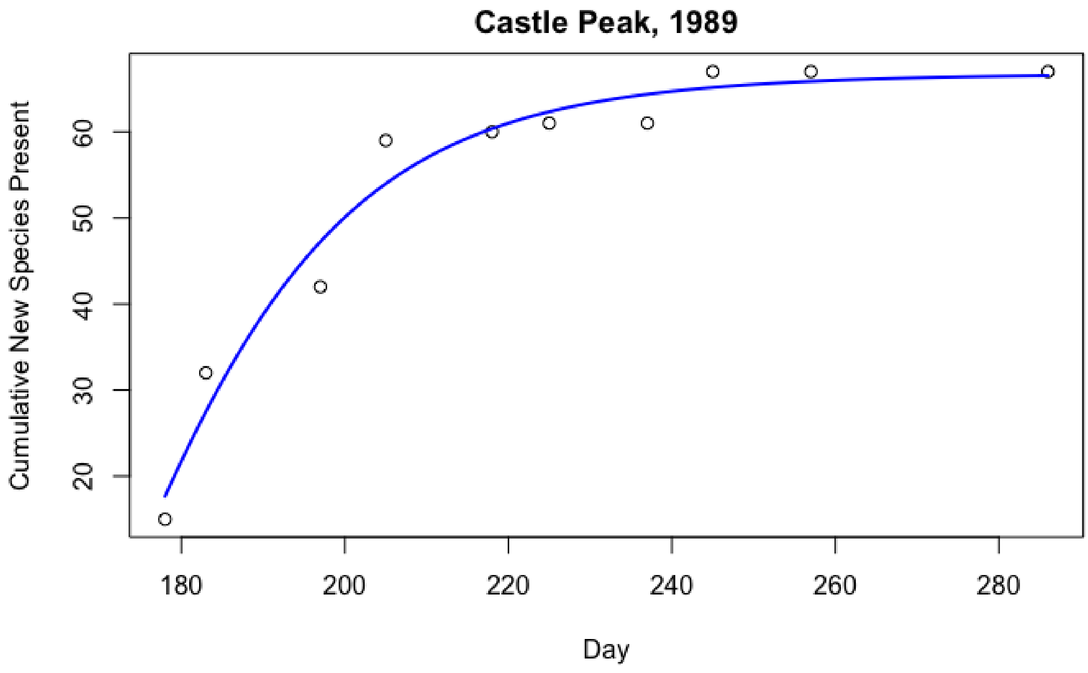



Figure 4.47 shows the SLL curve for Castle Peak for 1989. The parameters for that year are: $k = 0.09$, $\sigma = 141.22$, $d = 66.79$ and $\mu = 170.35$. MSE = 9.825 and MAE = 2.59

Figure 4.48: SLL species accumulation curve - Castle Peak 1999

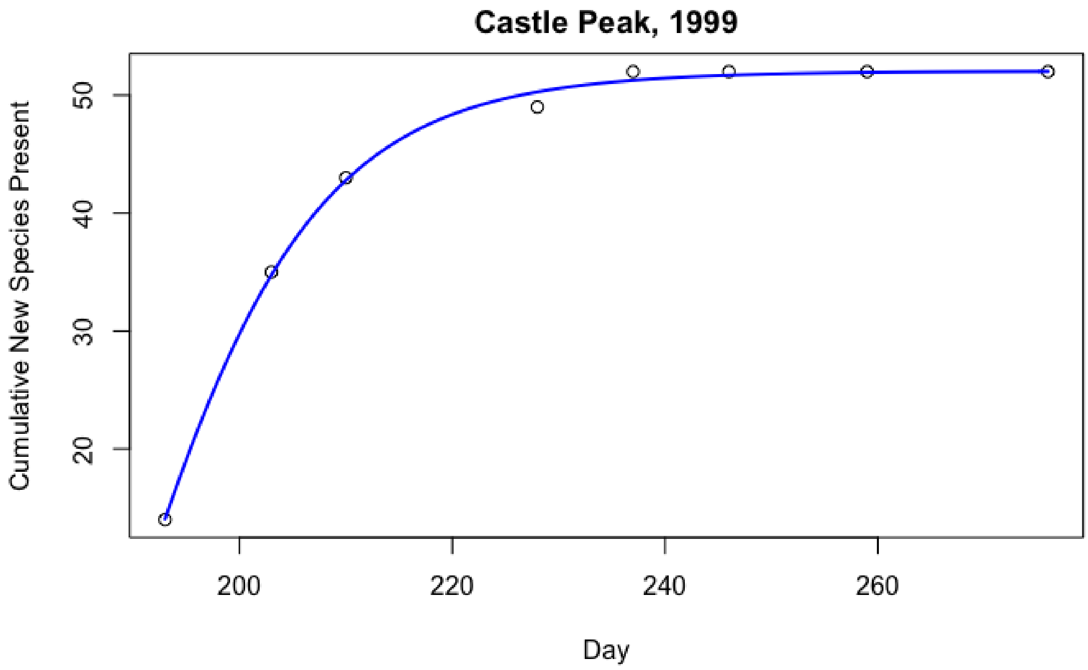



Figure 4.48 shows the SLL curve for Castle Peak for 1999. The parameters for that year are: $k = 0.09$, $\sigma = 226.17$, $d = 52.06$ and $\mu = 188.11$. MSE = 0.298 and MAE = 0.36

Figure 4.49: SLL species accumulation curve - Castle Peak 2002

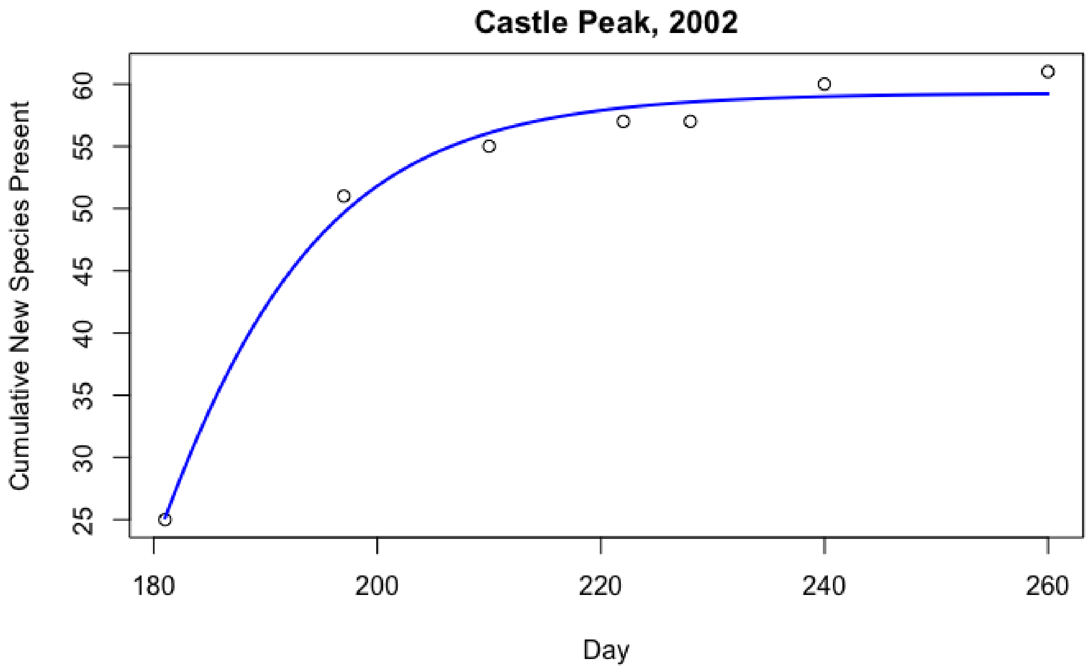



Figure 4.49 shows the SLL curve for Castle Peak for 2002. The parameters for that year are: $k = 0.10$, $\sigma = 179.93$, $d = 59.30$ and $\mu = 172.26$. MSE = 1.52 and MAE = 1.14

Figure 4.50: SLL species accumulation curve - Castle Peak 2014

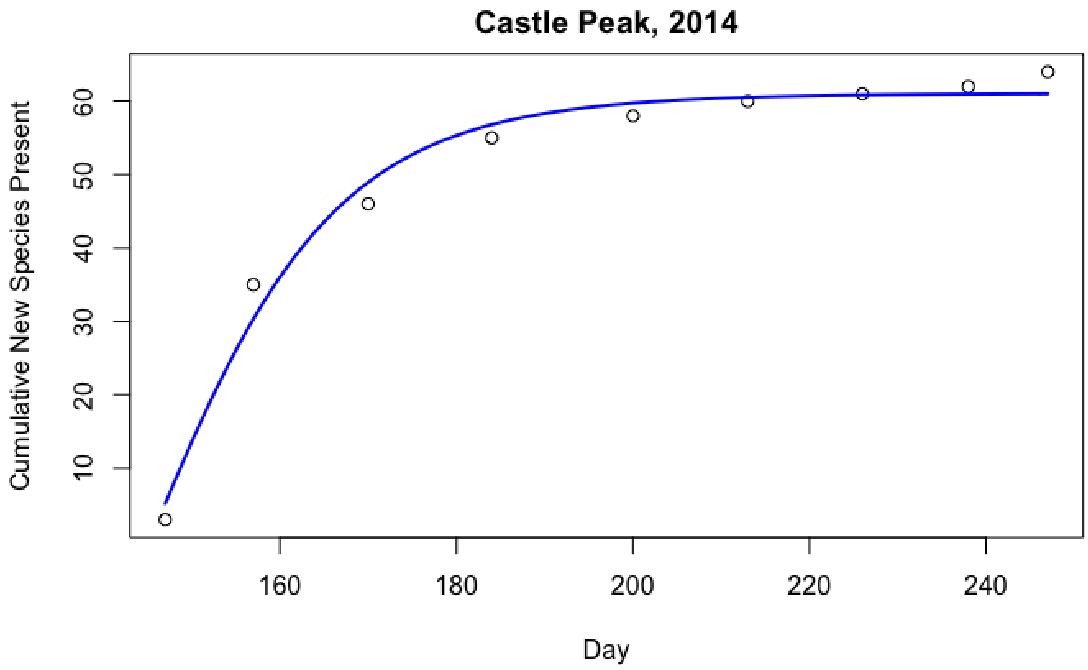



Figure 4.50 shows the SLL curve for Castle Peak for 2014. The parameters for that year are: $k = 0.04$, $\sigma = 325.69$, $d = 61.10$ and $\mu = 145.28$. MSE = 5.71 and MAE = 1.99

The range of mean absolute error for this site is 0.05 to 3.73. While this site has the highest maximum in the range of MAE this is still a relatively low error and therefore provides a good fit for the SLL curve. The MAE can be interpreted as the maximum difference between the estimated number of accumulated species and the observed number of species for this site to be 3.73, again demonstrating the effectiveness of estimating the species accumulation curves using the skewed log-logistic model.

## 4.3 SLL parameters' relationships with weather

In this section we will discuss the various regression models that were built using each SLL parameter as a response to the change in (annual) weather conditions. In our analysis, we attempted to discover a model that would work effectively for every site being analyzed. However, in addition to the weather variables there are many other environmental factors that impact the fit of the models being used. Each location selected is unique and chosen to capture the habitat and butterfly diversity that can be found in Northern California. The common challenge that we faced when conducting this research was due to the varying factors of the sites that make each one unique. The weather variables available for modeling were: the average daily maximum temperature per site, the average daily minimum temperature per site, and the precipitation per site. Maximum and minimum temperature were recorded on a seasonal basis providing us with data for spring, summer, fall and winter. The precipitation data were recorded on a seasonal and annual basis. Thus, for precipitation we had spring, summer, fall, winter, and annual precipitation per year available for modeling. We included the year for each model as a proxy for the passing of time as well as the number of visits per year as a measure of effort. According to Halsch et al. (2021), conditioning on year strengthens the inference of causation and population dynamics should be investigated using weather at both seasonal and annual scales. It is with this in mind that we have chosen to use these variables in building the regression models below.

To begin our analysis we first standardized (z transformed) all of our weather variables to be able to interpret and compare the coefficients of the regression models in our analysis. We first developed two full models using the seasonal maximum and minimum temperatures, year, and one with the seasonal precipitation and the other with annual precipitation as our predictor variables with each SLL parameter individually as the response variable. From there we conducted a stepwise regression to iteratively add and remove the weather predictors in order to find the subset of weather variables that would result in the best performing model. Once the stepwise regression resulted in a final model, we inspected this model to find if the regression was significant. That was assessed using the standard "significance of regression test". The significance of regression test has the following hypothesis: $H_0$: all parameters = 0 and $H_a$: at least one parameter $\neq 0$. We use the significance level of 0.05. A model was statistically "significant" if the p-value for this test was smaller than 0.05 (the significance level). Out of the models that were significant, we selected those with the largest adjusted R-squared, which adjusts for the number of predictors in the model. The higher the adjusted R-squared the better the model fits the data. In our analysis the model with the highest adjusted R-squared was deemed to be the best model for that site. During the analysis we found that for each site, the response variable that resulted in the optimal model was the inflection point ($mu$) along with a varying combination of the weather variables as the predictor variables. We conducted a step-wise regression analysis for each SLL parameter for each site and while we found successful models using the other SLL parameters, the optimal models for each site used the inflection point as the response parameter due to the adjusted R-squared. In addition, the inflection point parameter was the only parameter analyzed that created a successful model for all of the sites analyzed. We will now provide the optimal regression model for each site.

Table 4.18: Gates Canyon linear regression model output with $\mu$

| Coefficients: | Estimate | Std. Error | t value | Pr($>\|t\|$) |
|---|---|---|---|---|
| (Intercept) | -55.324 | 27.418 | -2.018 | 0.05234 |
| year | 0.028 | 0.014 | 2.018 | 0.05233 |
| tmax.spring | -0.638 | 0.214 | -2.976 | 0.00562 |
| tmin.spring | 0.348 | 0.228 | 1.530 | 0.13627 |
| tmin.fall | -0.369 | 0.156 | -2.362 | 0.02464 |
| tmin.winter | 0.470 | 0.182 | 2.589 | 0.01452 |
| ppt.fall | 0.321 | 0.143 | 2.253 | 0.03149 |
| ppt.winter | -0.416 | 0.198 | -2.099 | 0.04410 |
| Adjusted R-squared: | | | | 0.4162 |
| p-value: | | | | 0.0008601 |

**Gates Canyon:** The combination of weather variables that resulted in the optimal regression model for Gates Canyon were the seasonal minimum temperatures for spring, fall and winter, the seasonal maximum temperature for spring, and the seasonal precipitation for fall and winter. This model was determined to be the best model for this site due to having the highest adjusted R-squared value of all the models that passed the significance of regression test. Table 4.18 contains the result of our regression model showing the p-values for partial t-tests for each variable, computed adjusted R-squared, the estimated regression coefficients for each predictor variable and the p-value for the test of significance of regression. While interpreting this regression it can be seen that the weather variable that is deemed most impactful is the average daily maximum temperature for spring due to having the lowest individual p-value.

Table 4.19: Suisun Marsh linear regression model output with $\mu$

| Coefficients: | Estimate | Std. Error | t value | $\Pr(>\|t\|)$ |
|---|---|---|---|---|
| (Intercept) | -34.538 | 21.422 | -1.612 | 0.11563 |
| year | 0.017 | 0.011 | 1.612 | 0.11563 |
| tmax.spring | -0.226 | 0.145 | -1.552 | 0.12942 |
| tmax.fall | 0.314 | 0.119 | 2.648 | 0.01194 |
| tmin.summer | -0.201 | 0.137 | -1.467 | 0.15116 |
| tmin.fall | -0.207 | 0.136 | -1.517 | 0.13793 |
| ppt.spring | 0.468 | 0.140 | 3.331 | 0.00201 |
| ppt.winter | 0.325 | 0.109 | 2.978 | 0.00516 |
| Adjusted R-squared: | | | | 0.5769 |
| p-value: | | | | 1.446e-06 |

**Suisun Marsh:** With the inflection point as the response variable, the following weather predictors yielded the optimal regression model for the Suisun Marsh site: seasonal maximum temperatures for spring and fall, seasonal minimum temperatures for summer and fall, and seasonal precipitation for spring and winter. Table 4.19 is an output of our regression model showing the p-values for partial t-tests for each variable, the computed adjusted R-squared, the estimated regression coefficients for each predictor variable and the p-value for the test of significance of regression. By analyzing the individual p-values of each predictor variable, we can conclude that the seasonal precipitations for spring and winter had the strongest relationships with the response variable of the inflection point for this site.

Table 4.20: West Sacramento linear regression model output with $\mu$

| Coefficients: | Estimate | Std. Error | t value | $\Pr(>\|t\|)$ |
|---|---|---|---|---|
| (Intercept) | -8.492428 | 38.457389 | -0.221 | 0.8271 |
| year | 0.004242 | 0.019209 | 0.221 | 0.82709 |
| tmin.fall | -0.466096 | 0.171928 | -2.711 | 0.0122 |
| ppt.spring | 0.478683 | 0.127821 | 3.745 | 0.001 |
| ppt.winter | 0.404749 | 0.133023 | 3.043 | 0.00561 |
| Adjusted R-squared: | | | | 0.5889 |
| p-value: | | | | 3.21e-05 |

**West Sacramento:** Through our step-wise regression, we found that the optimal regression model for West Sacramento contained the following weather variables as the predictors: seasonal minimum temperature for fall, seasonal precipitation for spring and winter with the inflection point as our response variable. Table 4.20 contains our regression model showing the p-values for partial t-tests for each variable, computed adjusted R-squared, the estimated regression coefficients for each predictor variable and the p-value for the test of significance of regression. Upon analysis of this regression model, due to the low individual p-value, we can determine that the seasonal precipitation for spring and winter had the strongest relationships with our response SLL parameter $\mu$, the inflection point.

Table 4.21: North Sacramento linear regression model output with $\mu$

| Coefficients: | Estimate | Std. Error | t value | Pr(> $\|t\|$) |
|---|---|---|---|---|
| (Intercept) | -29.03535 | 56.12466 | -0.517 | 0.611587 |
| year | 0.0145 | 0.02803 | 0.517 | 0.611586 |
| tmax.spring | -1.138 | 0.24663 | -4.614 | 0.000247 |
| tmax.summer | -0.34408 | 0.19522 | -1.762 | 0.095958 |
| tmax.fall | 0.3618 | 0.21792 | 1.66 | 0.115203 |
| tmax.winter | 0.50887 | 0.2023 | 2.515 | 0.022232 |
| tmin.spring | 0.91175 | 0.27276 | 3.343 | 0.003857 |
| tmin.fall | -0.54709 | 0.22171 | -2.468 | 0.024517 |
| tmin.winter | -0.34007 | 0.2287 | -1.487 | 0.155333 |
| ppt.summer | -0.25337 | 0.17199 | -1.473 | 0.158985 |
| ppt.fall | 0.38122 | 0.15493 | 2.461 | 0.024872 |
| ppt.winter | 0.3898 | 0.21735 | 1.793 | 0.09071 |
| Adjusted R-Squared: | R-squared | | | 0.6099 |
| p-value: | | | | 0.001656 |

**North Sacramento:** The combination of weather variables that resulted in the optimal regression model for North Sacramento through the step-wise regression model selection were: the seasonal maximum temperature for spring, summer, fall and winter, the seasonal minimum temperatures for spring, fall and winter and the seasonal precipitation for summer, fall and winter. Table 4.21 contains our regression model output showing the p-values for partial t-tests for each variable, computed adjusted R-squared, the estimated regression coefficients for each predictor variable and the p-value for the test of significance of regression. Through the analysis of individual p-values for each predictor variable we determined that the average daily maximum temperature for spring resulted in the strongest relationship with the SLL parameter $d$, the inflection point.

Table 4.22: Rancho Cordova linear regression model output with $\mu$

| Coefficients: | Estimate | Std. Error | t value | Pr(> $\|t\|$) |
|---|---|---|---|---|
| (Intercept) | -7.35769 | 45.484824 | -0.162 | 0.872794 |
| year | 0.003675 | 0.02272 | 0.162 | 0.872793 |
| tmin.fall | -0.049857 | 0.199435 | -0.25 | 0.804636 |
| ppt.spring | 0.647465 | 0.156502 | 4.137 | 0.000348 |
| Adjusted R-squared: | | | | 0.3662 |
| p-value: | | | | 0.002296 |

**Rancho Cordova:** Using the inflection point as the response variable, the following weather predictors yielded the optimal regression model for the Rancho Cordova site: seasonal minimum temperature for fall and seasonal precipitation for spring. Table 4.22 provides an output of our optimal regression model showing the p-values for each variable as well as the computed adjusted R-squared with the inflection point parameter ($\mu$) as the response variable. By comparing the individual p-values for each predictor weather variable, we can determine that the seasonal precipitation for spring resulted in the strongest relationship with our response variable, inflection point.

Table 4.23: Washington linear regression model output with $\mu$

| Coefficients: | Estimate | Std. Error | t value | Pr(> $|t|$) |
|---|---|---|---|---|
| (Intercept) | -101.0942 | 44.8791 | -2.253 | 0.035108 |
| year | 0.0505 | 0.0224 | 2.253 | 0.035107 |
| tmax.summer | -0.5877 | 0.2049 | -2.868 | 0.009209 |
| tmax.winter | -0.2356 | 0.1477 | -1.595 | 0.125662 |
| tmin.summer | 0.5111 | 0.2372 | 2.154 | 0.042971 |
| tmin.fall | -0.3534 | 0.1615 | -2.189 | 0.040072 |
| ppt.spring | 0.5412 | 0.1405 | 3.851 | 0.000927 |
| ppt.winter | -0.3905 | 0.1462 | -2.671 | 0.014297 |
| Adjusted R-squared: | | | | 0.594 |
| p-value: | | | | 0.0002645 |

**Washington:** Through our step-wise regression model selection, we found that the optimal regression model for Washington contained the following weather variables as the predictors: seasonal maximum temperature for summer and winter, seasonal minimum temperature for summer and fall, seasonal precipitation for spring and winter. Table 4.23 provides the output for the optimal regression model showing the p-values for each variable as well as the computed adjusted R-squared with the inflection point ($\mu$) as the response variable. Through the interpretation of the regression output we can determine that the weather variable that has the most impactful relationship to the inflection point is the seasonal precipitation for spring due to having the lowest individual p-value.

Table 4.24: Lang Crossing linear regression model output with $\mu$

| Coefficients: | Estimate | Std. Error | t value | Pr(> $\|t\|$) |
|---|---|---|---|---|
| (Intercept) | 9.4146 | 25.4736 | 0.37 | 0.7139 |
| year | -0.0047 | 0.0128 | -0.37 | 0.7139 |
| tmax.spring | -0.3858 | 0.1671 | -2.309 | 0.027 |
| tmax.summer | -0.3158 | 0.2320 | -1.361 | 0.1822 |
| tmin.winter | 0.4203 | 0.2093 | 2.008 | 0.0524 |
| ppt.summer | -0.3714 | 0.1904 | -1.951 | 0.0591 |
| ppt.winter | -0.5728 | 0.2254 | -2.541 | 0.0156 |
| Adjusted R-squared: | | | | 0.1631 |
| p-value: | | | | 0.05342 |

**Lang Crossing:** With the inflection point as the response variable, the following weather predictors yielded the optimal regression model for the Lang Crossing site: seasonal maximum temperatures for spring and summer, seasonal minimum temperature for winter, and seasonal precipitation for summer and winter. Table 4.24 provides the optimal regression model output containing the p-values for partial t-tests for each variable, computed adjusted R-squared, the estimated regression coefficients for each predictor variable and the p-value for the test of significance of regression. By comparing the individual p-values for each weather predictors, we determined that the variables that provides the strongest relationship with the inflection point parameter was the average daily maximum temperatures for spring and the seasonal precipitation for winter.

Table 4.25: Sierra Valley linear regression model output with $\mu$

| Coefficients: | Estimate | Std. Error | t value | $\Pr(>\|t\|)$ |
|---|---|---|---|---|
| (Intercept) | -132.922 | 54.418 | -2.443 | 0.02199 |
| year | 0.066 | 0.027 | 2.443 | 0.02199 |
| tmax.spring | 1.538 | 0.580 | 2.652 | 0.01369 |
| tmax.summer | -0.589 | 0.285 | -2.065 | 0.04945 |
| tmax.fall | -0.501 | 0.166 | -3.024 | 0.0057 |
| tmin.spring | -1.169 | 0.452 | -2.589 | 0.0158 |
| tmin.summer | 0.586 | 0.331 | 1.769 | 0.08908 |
| ppt.spring | 0.920 | 0.298 | 3.091 | 0.00485 |
| ppt.winter | 0.310 | 0.170 | 1.822 | 0.08037 |
| Adjusted R-squared: | | | | 0.3566 |
| p-value: | | | | 0.01061 |

**Sierra Valley:** From the step-wise regression model selection, the combination of weather variables that resulted in the optimal regression model for Sierra Valley were: the seasonal maximum temperature for spring, summer, and fall, the seasonal minimum temperatures for spring and summer and the seasonal precipitation for summer and winter. Table 4.25 provides the regression output showing the p-values for partial t-tests for each variable, computed adjusted R-squared, the estimated regression coefficients for each predictor variable and the p-value for the test of significance of regression utilizing the inflection point $\mu$ as the response variable. Through the analysis of individual p-values for each weather predictor variable, we can conclude that the average daily maximum temperature for fall along with the seasonal precipitation for spring resulted in the most impactful variables to our response variable $\mu$.

Table 4.26: Donner Pass linear regression model output with $\mu$

| Coefficients: | Estimate | Std. Error | t value | Pr(> $\|t\|$) |
|---|---|---|---|---|
| (Intercept) | 62.8567 | 31.9165 | 1.969 | 0.056 |
| year | -0.0315 | 0.0160 | -1.969 | 0.056 |
| tmax.spring | -0.6929 | 0.2205 | -3.142 | 0.0032 |
| tmin.spring | 0.4548 | 0.2826 | 1.609 | 0.1156 |
| ppt.fall | 0.2106 | 0.1346 | 1.564 | 0.1259 |
| Adjusted R-squared: | | | | 0.299 |
| p-value: | | | | 0.001182 |

**Donner Pass:** Using the inflection point as the response variable, the following weather predictors yielded the optimal regression model for the Donner Pass site: seasonal maximum and minimum temperatures for spring and seasonal precipitation for fall. Table 4.26 provides the output for the regression model containing the p-values for partial t-tests for each variable, computed adjusted R-squared, the estimated regression coefficients for each predictor variable and the p-value for the test of significance of regression. By the analysis of individual p-values for each weather predictor, the average daily maximum temperature for spring was determined to have the strongest relationship to the inflection point.

Table 4.27: Castle Peak linear regression model output with $\mu$

| Coefficients: | Estimate | Std. Error | t value | Pr(> $\|t\|$) |
|---|---|---|---|---|
| (Intercept) | -49.26467 | 48.65010 | -1.013 | 0.32631 |
| year | 0.02460 | 0.02429 | 1.013 | 0.3263 |
| tmax.spring | 0.66650 | 0.43418 | 1.535 | 1.44E-01 |
| tmax.winter | -0.33028 | 0.15095 | -2.188 | 4.39E-02 |
| tmin.spring | -0.82358 | 0.36388 | -2.263 | 0.03788 |
| tmin.summer | 0.45341 | 0.20833 | 2.176 | 0.04485 |
| tmin.fall | -0.23155 | 0.15254 | -1.518 | 0.14854 |
| ppt.spring | 0.78400 | 0.22811 | 3.437 | 3.39E-03 |
| ppt.fall | 0.39137 | 0.12843 | 3.047 | 7.68E-03 |
| ppt.winter | 0.27964 | 0.12922 | 2.164 | 0.04592 |
| Adjusted R-squared: | | | | 0.7125 |
| p-value: | | | | 0.0002057 |

**Castle Peak:** The combination of weather variables that resulted in the optimal regression model with the inflection point as the response variable for Castle Peak were: the seasonal minimum temperatures for spring, summer and fall, the seasonal maximum temperatures for spring and winter, and the seasonal precipitation for spring, fall and winter. Table 4.27 provides the output for the optimal regression model showing the p-values for partial t-tests for each variable, computed adjusted R-squared, the estimated regression coefficients for each predictor variable and the p-value for the test of significance of regression. Through the interpretation of the regression output we can determine that the weather variables that have the most impactful relationships to the inflection point are the seasonal precipitations for spring and fall due to having the lowest individual p-values.

Table 4.28: Summary of models for each site

| Site | SLL Parameter | Weather Variables | Avg. $\mu$ |
|---|---|---|---|
| Gates Canyon | $\mu$ | tmax spring | 114.16 |
| Suisun Marsh | $\mu$ | ppt spring, ppt winter | 142.27 |
| West Sacramento | $\mu$ | ppt spring, ppt winter | 118.95 |
| North Sacramento | $\mu$ | tmax spring | 117.31 |
| Rancho Cordova | $\mu$ | ppt spring | 90.17 |
| Washington | $\mu$ | ppt spring | 139.57 |
| Lang Crossing | $\mu$ | tmax spring, ppt winter | 158.75 |
| Sierra Valley | $\mu$ | ppt spring, tmax fall | 153.88 |
| Donner Pass | $\mu$ | tmax spring | 160.91 |
| Castle Peak | $\mu$ | ppt spring, ppt fall | 169.48 |

**Summary:** We began the analysis by creating a full model utilizing all of the seasonal weather variables available to us as predictors. Starting with that full model we ran multiple step-wise regression models using the individual SLL parameters as our response variables. For each SLL parameter we were able to create models using various combinations of the weather variables for each site. Once these models were complete we analyzed them to determine the optimal model for each site. Through this analysis we found that by using the inflection point as our response variable we were able to produce models for each site that passed the significance of regression test and achieved the highest adjusted R-squared of the models for all the SLL parameters.

Once the optimal models were determined for each site we began to analyze the model for each site individually to determine if there were any significant relationships with the weather predictors. Through this analysis we discovered that the weather predictors for spring continued to have the strongest relationship with the inflection point for each of the models. In some instances that weather predictor was the average daily maximum or minimum temperatures for spring or the seasonal precipitation for spring. Essentially, the inflection point for each of our models had a significant relationship with the weather variables for spring.

If we look at the season of spring in terms of the ordinal days in our analysis, the range of days for spring is determined to be (60 - 152). By analyzing the average inflection point for each site we see that this average falls well within our range or close to the tail end of the range for certain sites. The inflection point for each site provides us with ordinal day where the rate of change of the cumulative number of species begins to decline. Therefore, from our analysis it can be determined that since the inflection point on average falls within our spring range that the weather variables for spring would have a strong relationship on this parameter. To summarize our modeling and analysis results we provide table 4.28 with the most impactful weather variables for each site along with the average inflection point for all years for each site.

# Chapter 5

# Discussion and Conclusion

We now recall the objectives of this research and summarize our results below.

- **To develop an understanding of the impact of the parameters on the skewed log-logistic species accumulation curve.** In this thesis we have provided an analysis and discussion on the impact of the various SLL parameters on the properties of the species accumulation curve. We demonstrated the effects on the curve resulting from changing one parameter, while all others remained fixed.

- **Develop and assess the fit of the species accumulation curves using the nls.LM function to compute the parameters.** We were able to effectively develop annual species accumulation curves for each of the sites researched using the **nls.LM** function in R. The goodness of fit of our curves was then assessed using the mean squared error and mean absolute error between the observed data points and the fitted curve for each year. We observed that the skewed log-logistic model provided a relatively good fit for the butterfly species data measured by the MAE and MSE for each site.

- **To identify and model associations between the SLL accumulation curve parameters and the weather variables for each site.** For each site researched in this analysis, we were able to provide a regression model that reasonably explains the association between the SLL inflection point $\mu$ and the weather variables. The optimal models were chosen through the use of the step-wise regression model selection. The goodness of fit in the regression models was demonstrated through the adjusted R-squared. In addition, we were able to gain an understanding of the impact of the weather variables on the butterfly species accumulation curves at various locations. Our analysis showed that the weather during spring resulted in the strongest relationship with the inflection point of the SLL curves.

**Limitations:** Through our analysis we were able to identify the following limitations.

1. A limitation of this thesis arises in the use of the optimal model for each site and not further analyzing the additional models for the other SLL parameters with respect to weather.

2. We find replication of this research to be a limitation as we are unable to record butterfly species in the past. At best we could obtain butterfly data from another source, but that would not be consistent with our original source data.

**Examples of areas for further study:**

1. Perform a more comprehensive study of the fit of the skew log-logistic model to the butterfly species data per site of interest.

2. Continue to develop and expand upon the R script to estimate the 5 parameters of the SLL species accumulation curves.

3. Consider natural history differences among groups of butterfly species to be able to partition species into groups with informative and different responses to weather.

4. Continue to examine various regression models between weather and the inflection point of the SLL butterfly species curves for additional areas that are being monitored.

5. Discover additional factors of influence on butterfly species that interact with weather that could potentially have a relationship to the SLL curve, specifically inflection point.

# Chapter 6

# Bibliography

[1] Arellano-Valle, R., Bolfarine, H., and Lachos, V. (2005). Skew-normal linear mixed models. 3, 415-438. https://doi.org/10.6339/JDS.2005.03(4).238.

[2] Christopoulos, D.T. (2014). Developing methods for identifying the inflection point of a convex/concave curve. National and Kapodistrian University of Athens: Department of Economics, https://arxiv.org/pdf/1206.5478.pdf.

[3] Deng, C., Daley, T., and Smith, A. D. (2015). Applications of species accumulation curves in large-scale biological data analysis. Quantitative Biology (Beijing, China), 3(3), 135–144. https://doi.org/10.1007/s40484-015-0049-7.

[4] Fernandez, C. and Steel, M. (1998). On bayesian modeling of fat tails and skewness. Journal of The American Statistical Association, 93 (March 1, 1998): 359–71. https://doi.org/10.1080/01621459.1998.10474117.

[5] Forister, M. L., McCall, A. C., Sanders, N. J., Fordyce, J. A., Thorne, J. H., O'Brien, J., Waetjen, D. P., and Shapiro, A. M. (2010). Compounded effects of climate change and habitat alteration shift patterns of butterfly diversity. Proceedings of the National Academy of Sciences, 107(5), 2088–2092. https://doi.org/10.1073/pnas.0909686107.

[6] Forister, M. L., Fordyce, J. A., Nice, C. C., Thorne, J. H., Waetjen, D. P., and Shapiro, A. M. (2018). Impacts of a millennium drought on butterfly faunal dy-

namics. Climate Change Responses, 5(1), 3. https://doi.org/10.1186/s40665-018-0039-x.

[7] Giri, A. (2021). Exploring the COVID-19 Cases around the World [M.S., University of Nevada, Reno]. In ProQuest Dissertations and Theses. https://www.proquest.com/pqdtlocal1006038/docview/2584002478.

[8] Gottschalk, P.G. and Dunn, J.R. (2005). The five-parameter logistic: a characterization and comparison with the four-parameter logistic. PubMed.gov, 1, 54-65, DOI:10.1016/j.ab.2005.04.035. PMID:15953581.

[9] Good, I. J., and Toulmin, G. H. (1956). The number of new species, and the increase in population coverage, when sample is increased. Biometrika, 43(1–2), 45–63. https://doi.org/10.1093/biomet/43.1-2.45.

[10] Halsch, C. A., Shapiro, A. M., Fordyce, J. A., Nice, C. C., Thorne, J. H., Waetjen, D. P., and Forister, M. L. (2021). Insects and recent climate change. Proceedings of the National Academy of Sciences, 118(2), e2002543117. https://doi.org/10.1073/pnas.2002543117.

[11] Matter, S. F., and Roland, J. (2017). Climate and extreme weather independently affect population growth, but neither is a consistently good predictor. Ecosphere, 8(5), e01816. https://doi.org/10.1002/ecs2.1816.

[12] Nice, C. C., Forister, M. L., Harrison, J. G., Gompert, Z., Fordyce, J. A., Thorne, J. H., Waetjen, D. P., and Shapiro, A. M. (2019). Extreme heterogeneity of population response to climatic variation and the limits of prediction. Global Change Biology, 25(6), 2127–2136. https://doi.org/10.1111/gcb.14593.

[13] Nieto-Sánchez, S., Gutiérrez, D., and Wilson, R. J. (2015). Long-term change and spatial variation in butterfly communities over an elevational gradient: Driven by climate, buffered by habitat. Diversity and Distributions, 21(8), 950–961. https://doi.org/10.1111/ddi.12316.

[14] Parmesan, C., Root, T. L., and Willig, M. R. (2000). Impacts of Extreme Weather and Climate on Terrestrial Biota. Bulletin of the Amer-

ican Meteorological Society, 81(3), 443–450. https://doi.org/10.1175/1520-0477(2000)081¡0443:IOEWAC¿2.3.CO;2.

[15] Sastry, D. V. S., and Bhati, D. (2016). A new skew logistic distribution: Properties and applications. Brazilian Journal of Probability and Statistics, 30(2), 248–271. https://doi.org/10.1214/14-BJPS278.

[16] Siatkowski, M., Weres, J., Kujawa, S., Szabelska, A., and Zyprych-Walczak, J. (2010). Growth curve functions in modeling the thin-layer drying of corn.

[17] Ugland, K. I., Gray, J. S., and Ellingsen, K. E. (2003). The species–accumulation curve and estimation of species richness. Journal of Animal Ecology, 72(5), 888–897. https://doi.org/10.1046/j.1365-2656.2003.00748.x.

[18] Vandermeer, J. (2010). How Populations Grow: The Exponential and Logistic Equations. Nature Education Knowledge, 10(3).

[19] Wicklin, R. (2014). How to find an initial guess for an optimization, SAS, https://blogs.sas.com/content/iml/2014/06/11/initial-guess-for-optimization.html.