Nest count estimation

Tomo Eguchi

August 9, 2018

## Introduction

Nest counts are ofren used as an index of adult female abundance of a marine turtle nesting population. Because adult female turtles return to their nesting beaches in somewhat regular intervals (depends on the productivity at their foraging grounds) and they lay similar numbers of clutches per individual per season, nest counts correlate with the total abundance of adult females.

Nesting beaches of marine turtles often exist in remote areas, where access to beaches can be difficult. Nocturnal nesting also adds difficulty in collecting accurate data. Logistical issues have sometimes result in missing data for some periods during a nesting season. To estimate an annual abundance of nests, hence an index of adult female abundance, missing values during a nesting season need to be estimated.

In this report, a statistical time-series analysis is used to estimate nest counts for months that lack data. The analysis uses a parametric model to estimate missing counts in a time series. The model takes advantage of a temporal pattern of nest counts at a nesting beach.

This analysis is solely designed to fill in data gaps and not used to estimate population growth rates or any other demographic parameters. Estimated data are used in the subsequent separate analyses to determine those parameters.

## Methods

### Data

Data for this analysis came from two leatherback nesting beaches in Indonesia (Jamursba-Medi and Wermon). Raw data for the analysis are the recorded number of nests per month. We treat these two beaches separately.

### Jamursba-Medi

At Jamursba-Medi, some data have been collected since 1981. However, there were many years of no data between the mid 1980s and late 1990s. A somewhat consistent effort starts in early 2000s and continue to recent years (Figure 1). The observed nest count in 1981 was 4000 and declined to less than 1000 since 2000.

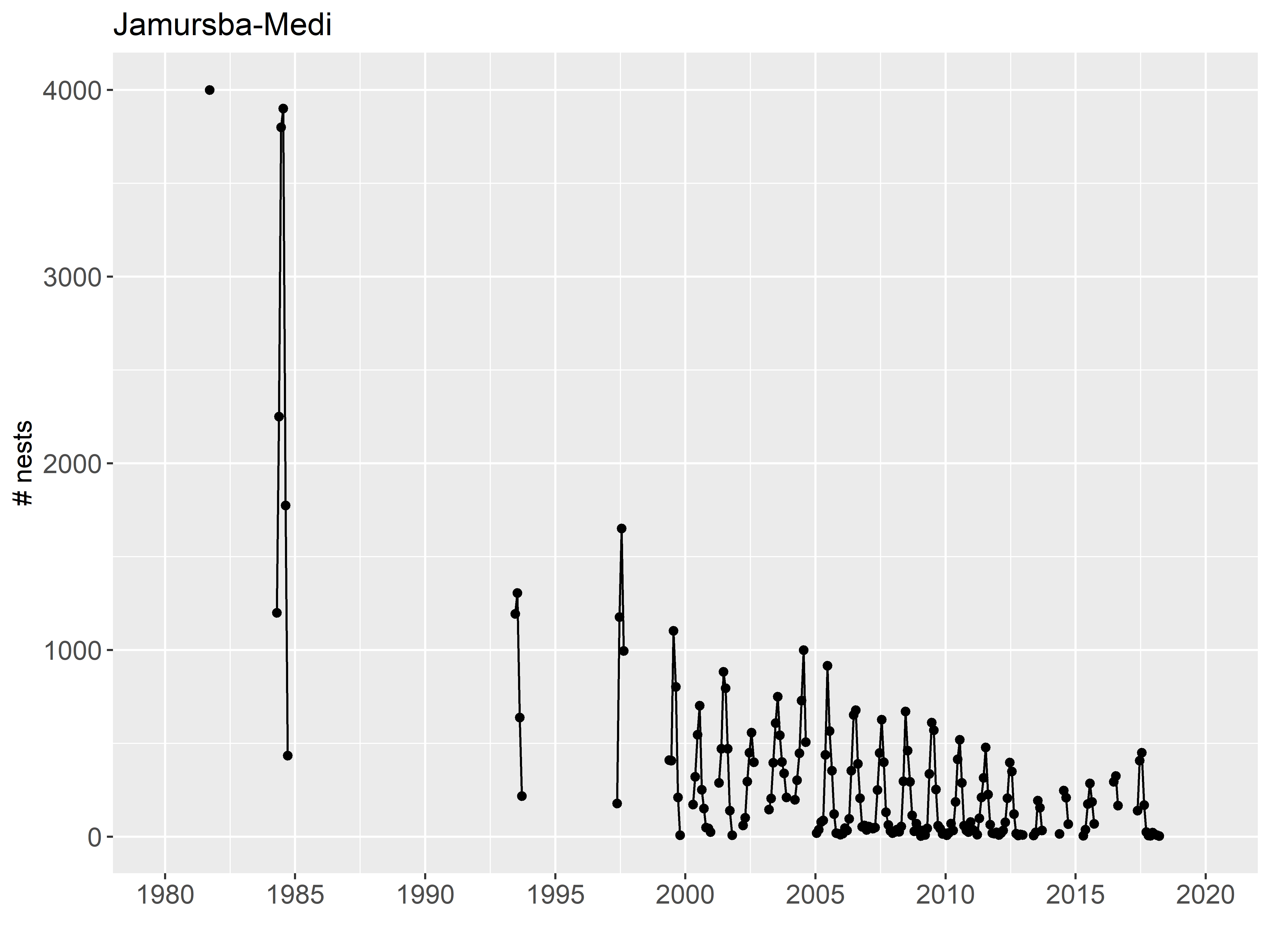


Figure 1. Nest counts at Jamursba-Medi

No data were collected in 1998. Since then at least some data were collected within each year. After analyzing datasets with different starting years, we found that starting at 1999 resulted in the longest dataset and good convergence of MCMC. Consequently, we decided to use data since 1999. The cyclical nature or seaonal fluctuations of counts is obvious in the raw counts (Figure 2).

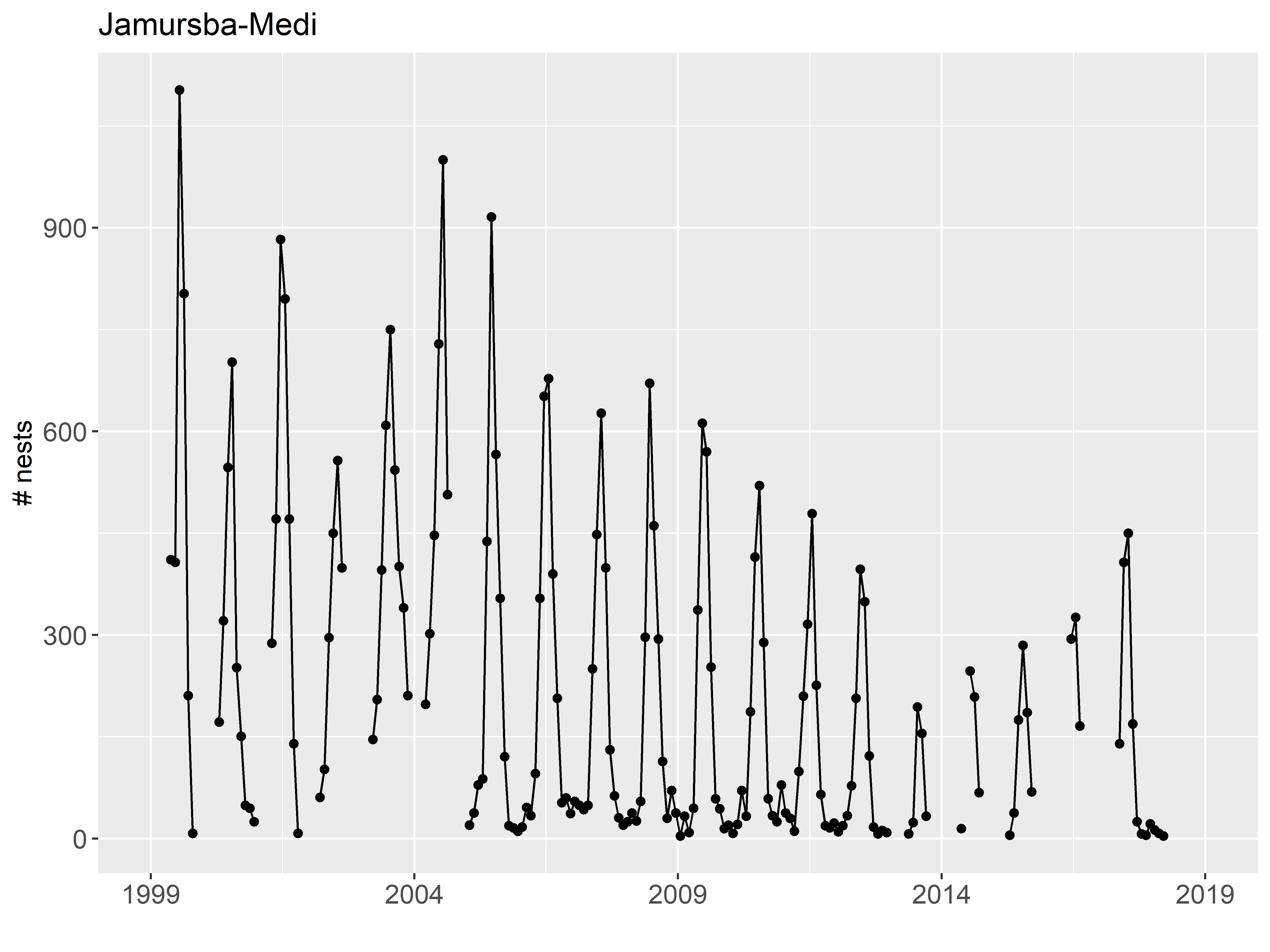


Figure 2. Nest counts at Jamursba-Medi since 1999

When the number of nests is plotted by month, the seasonal fluctations become more obvious (Figure 3). In general, high counts within a year are seen during summer months (approximately from April to September), whereas low counts are found during winter months (October through March).

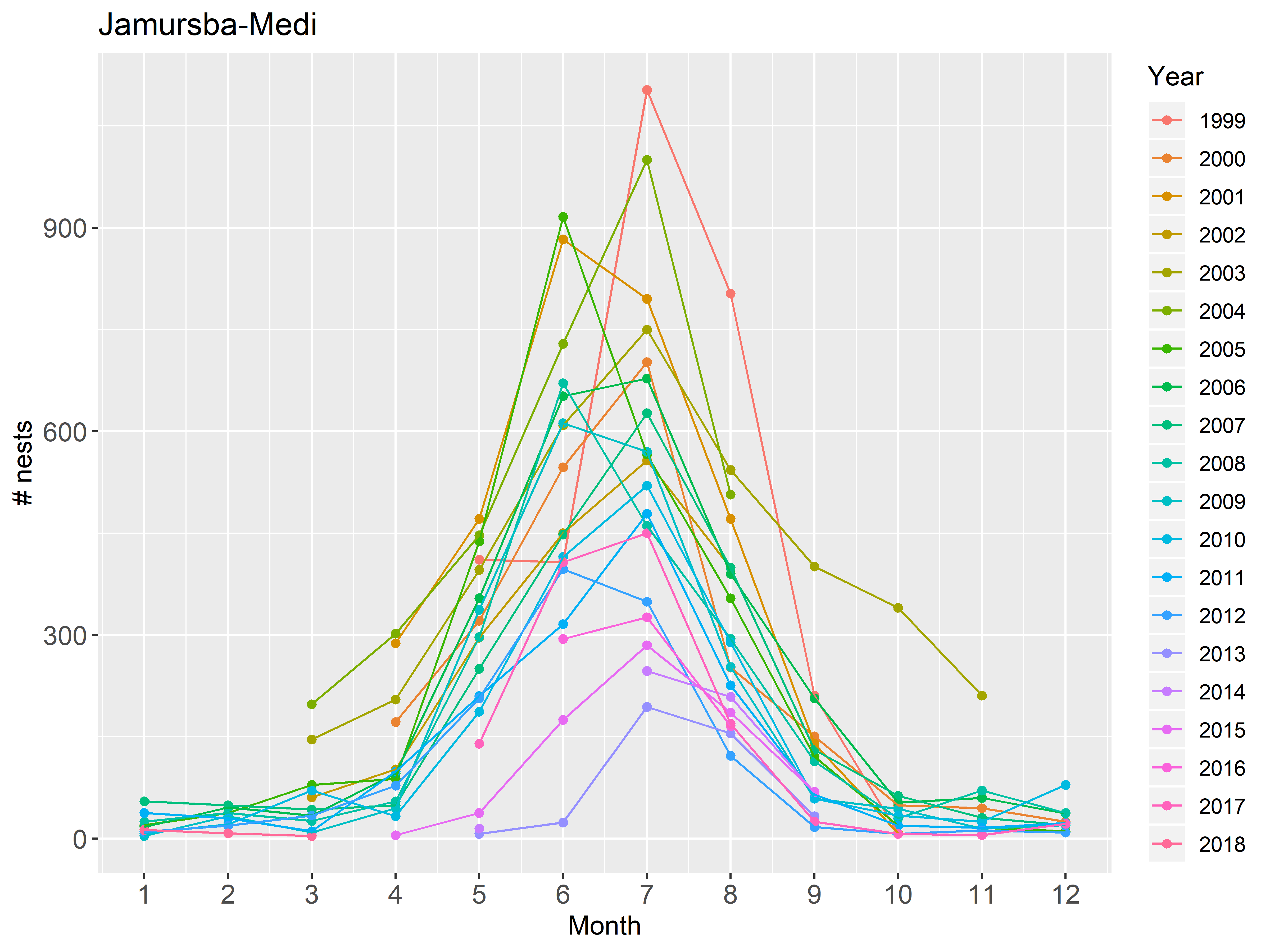


Figure 3. Monthly nest counts at Jamursba-Medi

### Wermon

At Wermon, some data have been collected since 2004. However, no data were collected between July 2013 and December 2015. The largeset nest count was 494 in 2005 (Figure 4).

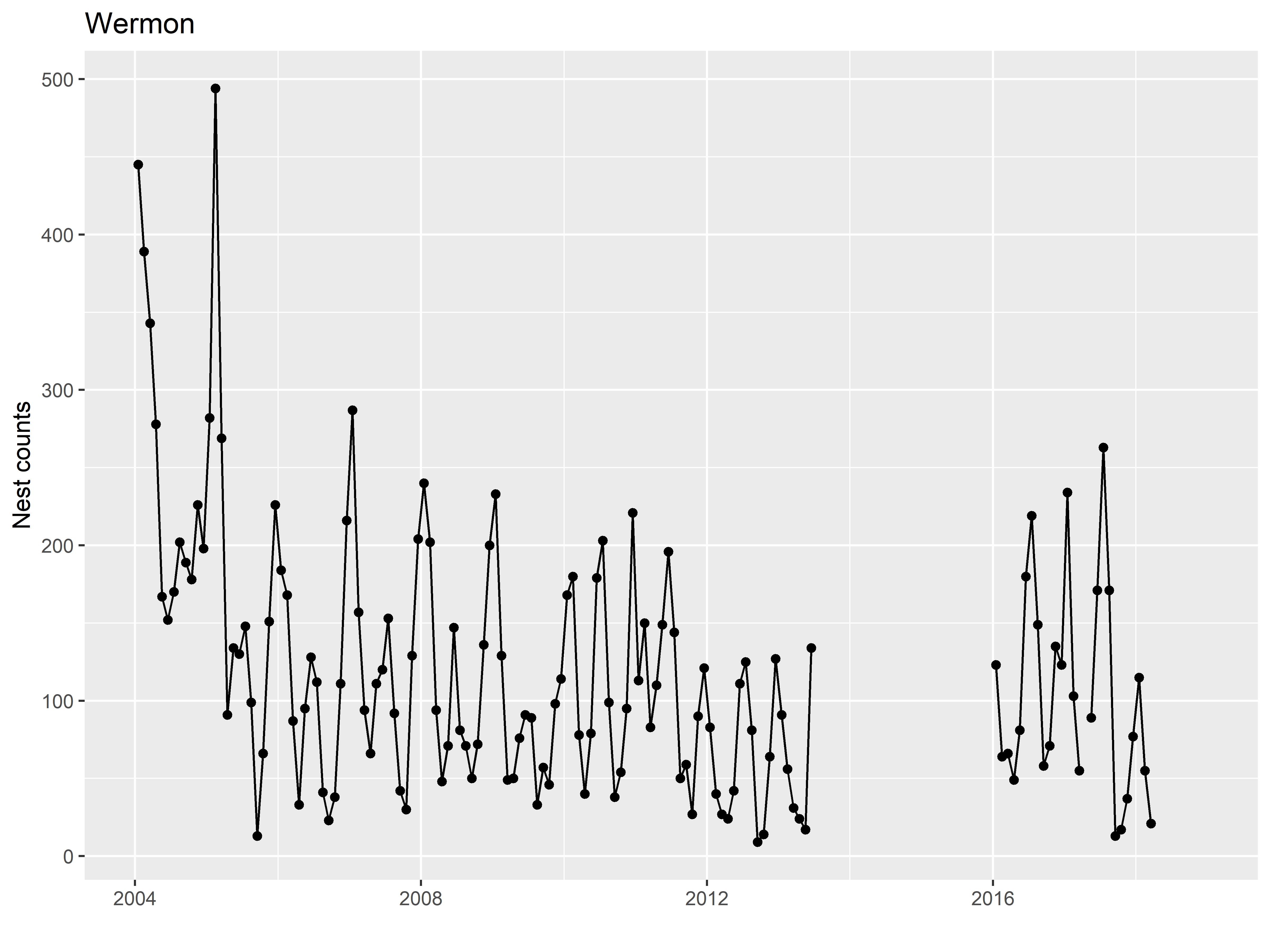


Figure 4. Nest counts at Wermon.

The temporal nest count pattern at Wermon is different from that of Jamursba-Medi (Figure 3 vs. Figure 5). At Wermon, there are two peaks annually: higher values are found between May and September and between October and April.

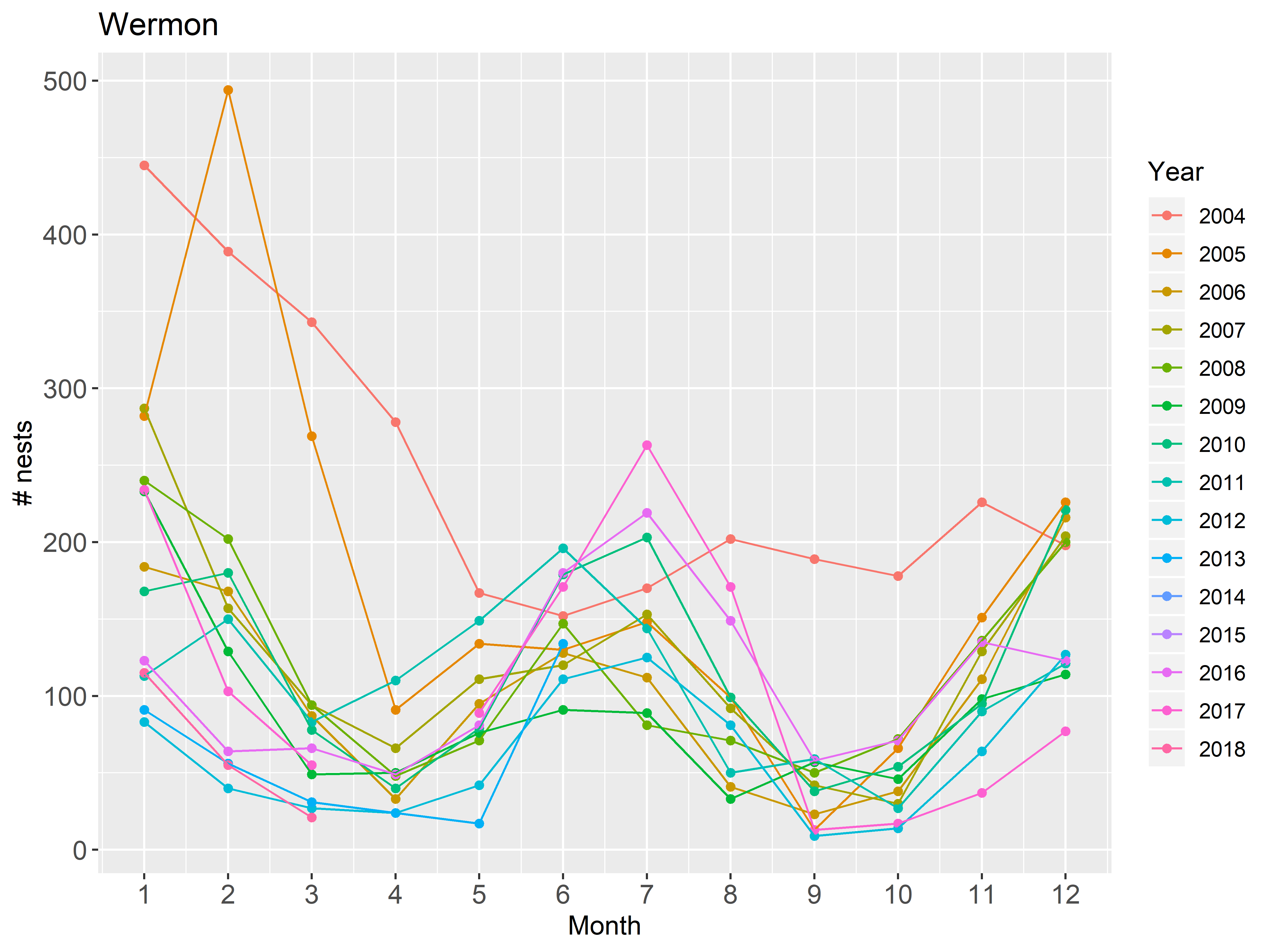


Figure 5. Monthly nest counts at Wermon.

Observed counts for 2004 and 2005 seem to be outliers with respect to within-year patterns. Because this analysis is to estimate missing values from observed patterns, we remove these years from the data for the following analysis.

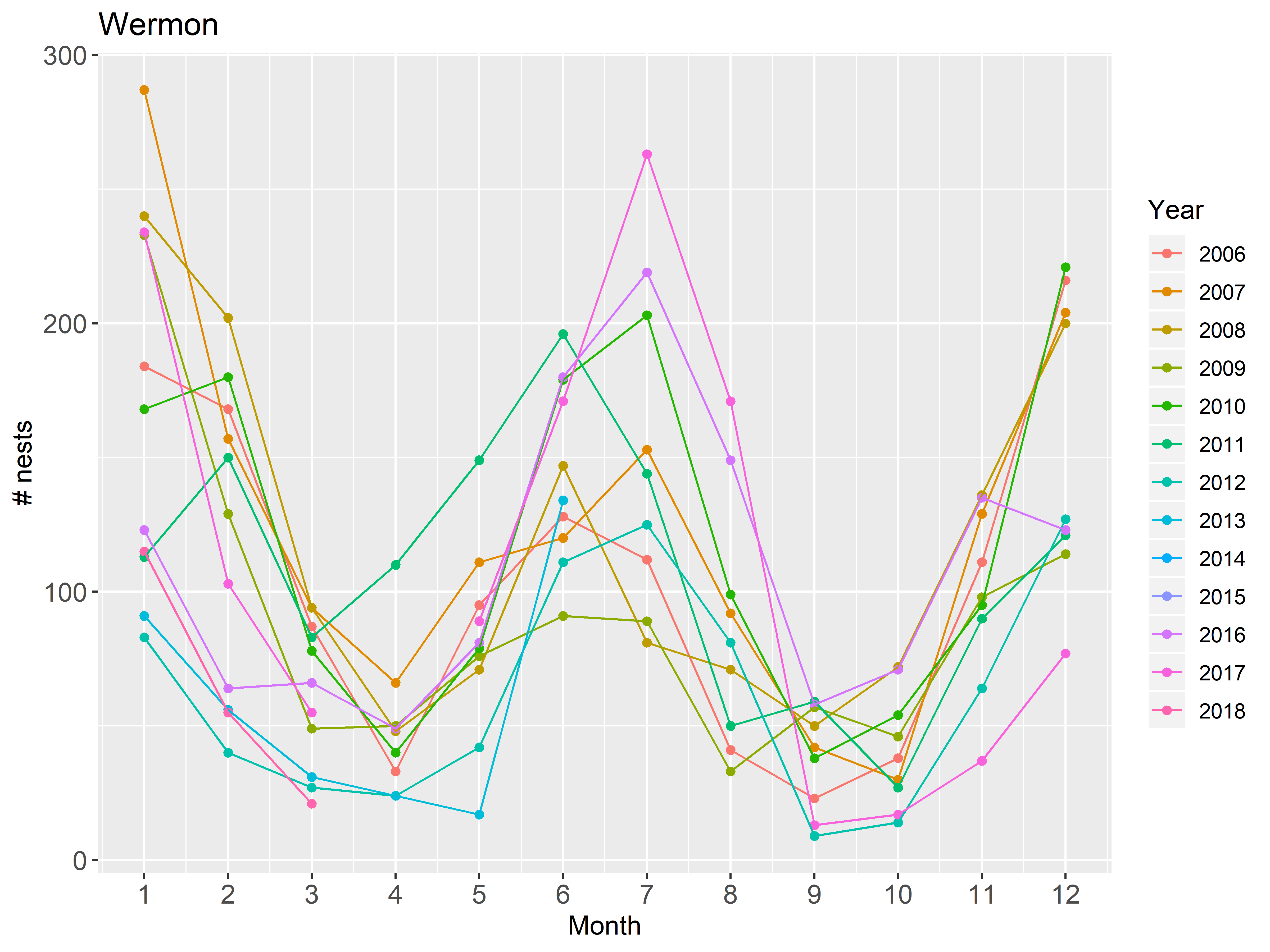


Figure 6. Monthly nest counts at Wermon since 2006.

The two datasets (counts since 1999 for Jamursba-Medi and since 2006 for Wermon) are used in the following analysis for estimating missing counts.

# Statistical model

Missing data are imputed using auto-regressive time-series analysis with 1 month of lag (AR(1)). It is a Bayesian state-space model with two variances and two slope parameters in the state model. The two variances correspond to high and low counts, whereas two slope parameters correspond to increasing and decreasing phases of counts.

The mean predicted count in month t (mu[t]) is a function of the true count of a month prior:

mu[t] = theta[t] \* X[t-1]

The process (the true number of nests) at time t, X[t], is assumed to be distributed with a truncated normal distribution with mean mu[t] and standard deviation sigma.pro1 or sigma.pro2:

X[t] ~ N(mu[t], sigma.pro1)T(0,), if season1

X[t] ~ N(mu[t], sigma.pro2)T(0,), if season2

where N() indicates Normal distribution and T(0,) indicates the truncation function for X[t] > 0. Season1 and season2 correspond to high and low count months. This is explained in subsections for each nesting beach in the following sections.

Observed counts, y[t], is also modeled with a truncated normal distribution with the mean X[t] and standard deviation sigma.obs, which was assumed to be time-invariant.

y[t] ~ N(X[t], sigma.obs)T(0,)

Vague priors were used for all parameters.

sigma.pro1 ~ UNIF(0, 600)

sigma.pro2 ~ UNIF(0, 200)

sigma.obs ~ UNIF(0, 100)

theta[t] ~ N(0, 0.01)

JAGS code for these analyses can be found in Appendix.

In this analysis, missing values are treated as unknown parameters. Consequently, posterior distributions of the missing values are used to make inference about the missing values. The estimated number of nests (X[t]) were used to compute the total annual number of nests. These estimated counts and their credible intervals were used to provide estimated annual counts and their uncertainty.

Analysis was conducted JAGS (REF) and rjags package (REF) within the R statistical environment.

# Results

## Jamursba-Medi

For Jamusba-Medi, we explored two sets of two variances of nest counts (sigma.pro1 and sigma.pro2). One set corresponded to summer (April through September as season1) and winter (October through March as season2), whereas the other set corresponded to the visual determination of two variances (Figure 3): May through August as season1 and September through April as season2. Convergence of MCMC was used to determine which set was better with respect to parameter estimation. For the slope parameter, two slopes were for increasing (Jan to Jul) and decreasing (Aug to Dec) phases of each year. Gelman-Rubin statistic and trace plots indicated that the visual separation of the variance parameter (May through August and September through April) resulted in better MCMC convergence. So, we will use the model for the following analysis.

The multi-variate potential scale reduction factor (Gelman diagnostic statistic) was 1.002, indicating an adequate convergence. Visual inspection of MCMC chains also indicated that a convergence was reached for all parameters (Figure 7).

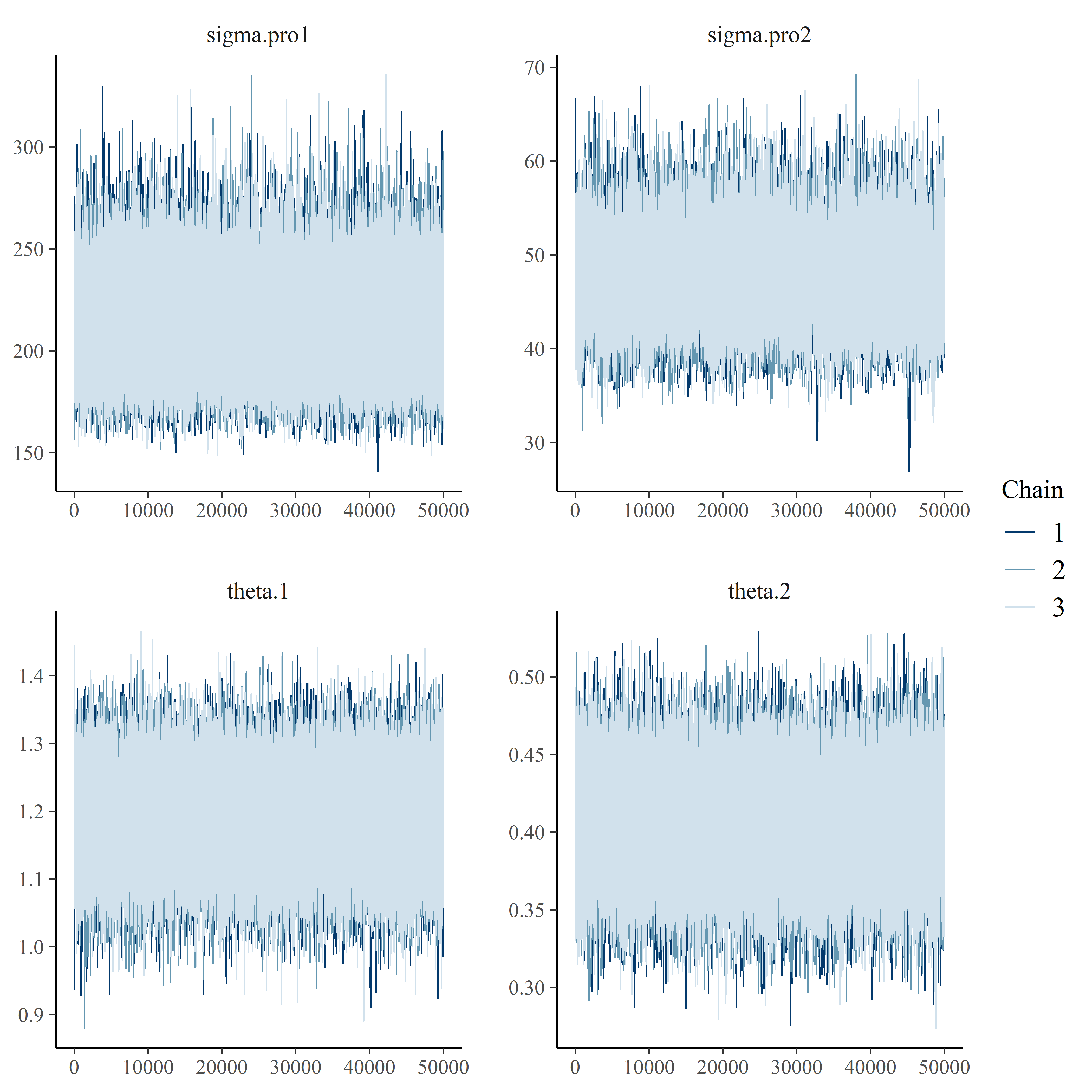


Figure 7. MCMC trace plots for Jamursba-Medi dataset.

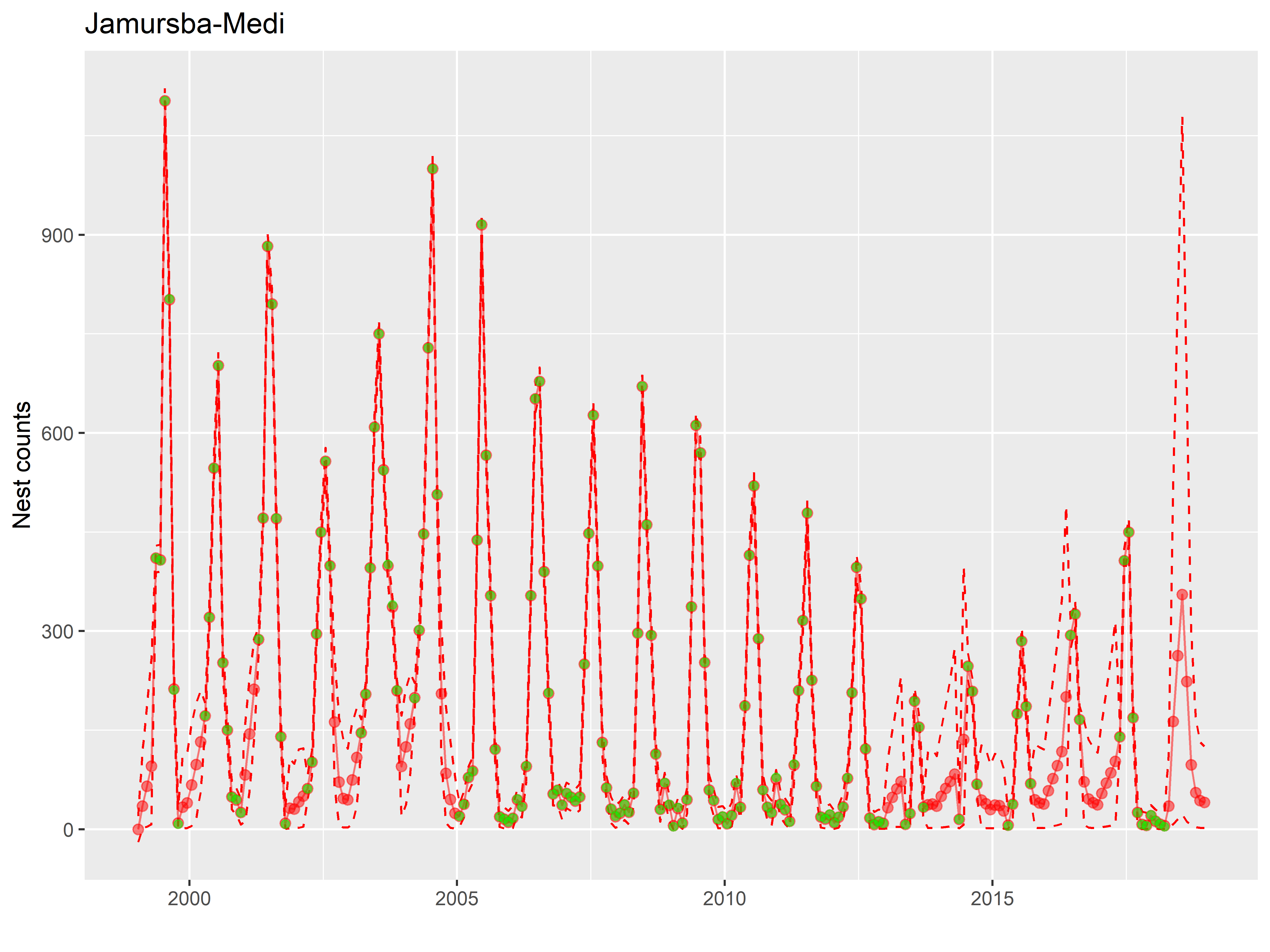


Figure 8. Predicted counts of nests for Jamursba-Medi dataset. Red is the estimated values and green are observed counts. 5 and 95% confidence limits are shown in dashed lines.

To estimate annual nest counts, median estimated counts were summed from April 1 to March 31 over two calendar years. Credible intervals were computed by summing lower and upper limits over the same time periods (Figure 9).

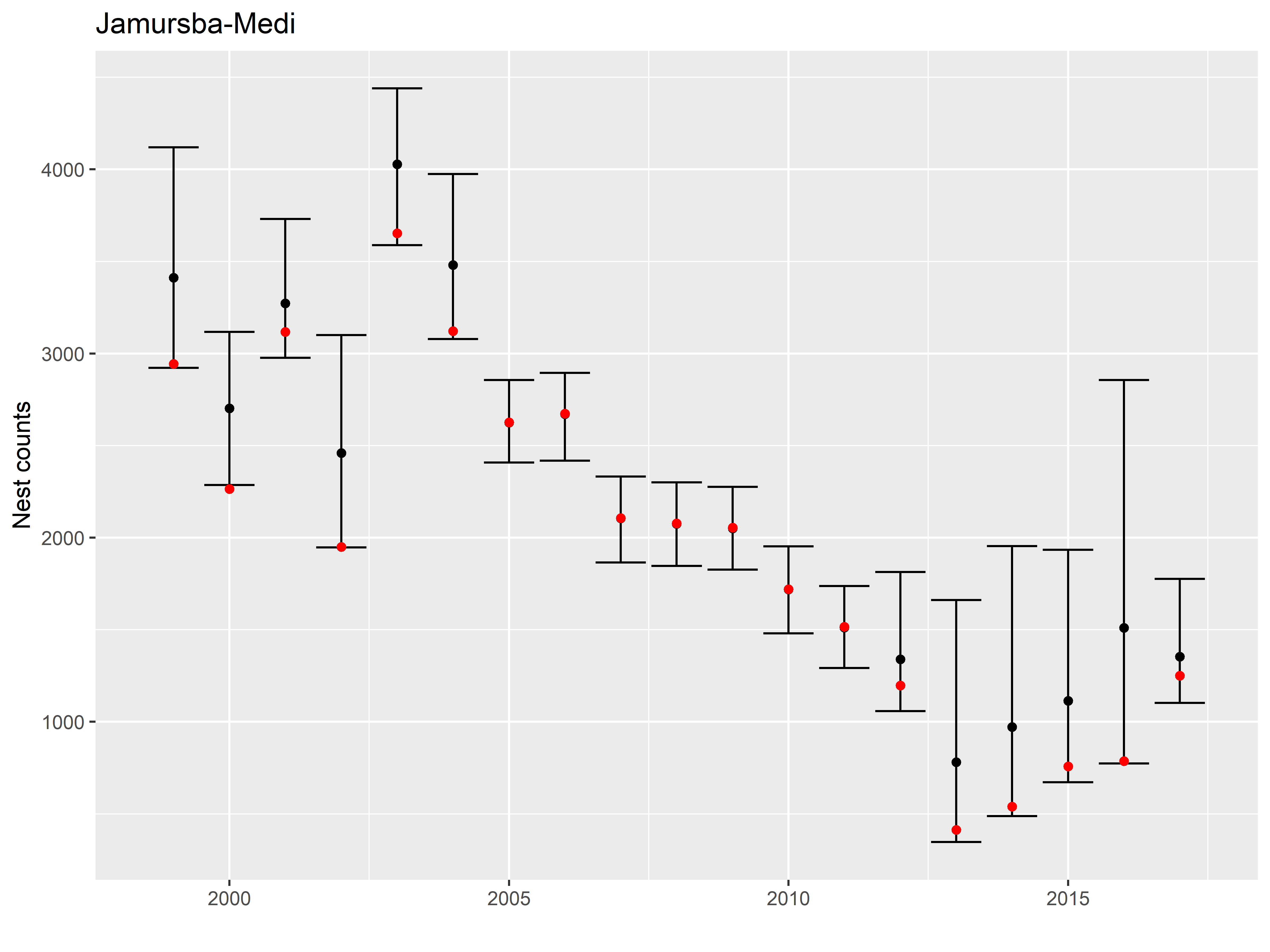


Figure 9. Total annual counts of nests for Jamursba-Medi dataset. Each sampling season starts April 1 and ends March 31. Black dots indicate medians, error bars indicate 95% credible intervals, and red dots indicate observed counts.

## Wermon

For Wermon, two variances corresponded to high nest count months (March, April, September, and October) and low nest count months (January, February, May, June, July, August, November and December). Two slopes corresponded to increasing (May, June, November, and December) and decreasing (January through April and July through October) phases of each year.

The multi-variate potential scale reduction factor (Gelman diagnostic statistic) was 1.009, indicating an adequate convergence. Visual inspection of MCMC chains also indicated a convergence was reached for all parameters (Figure 10).

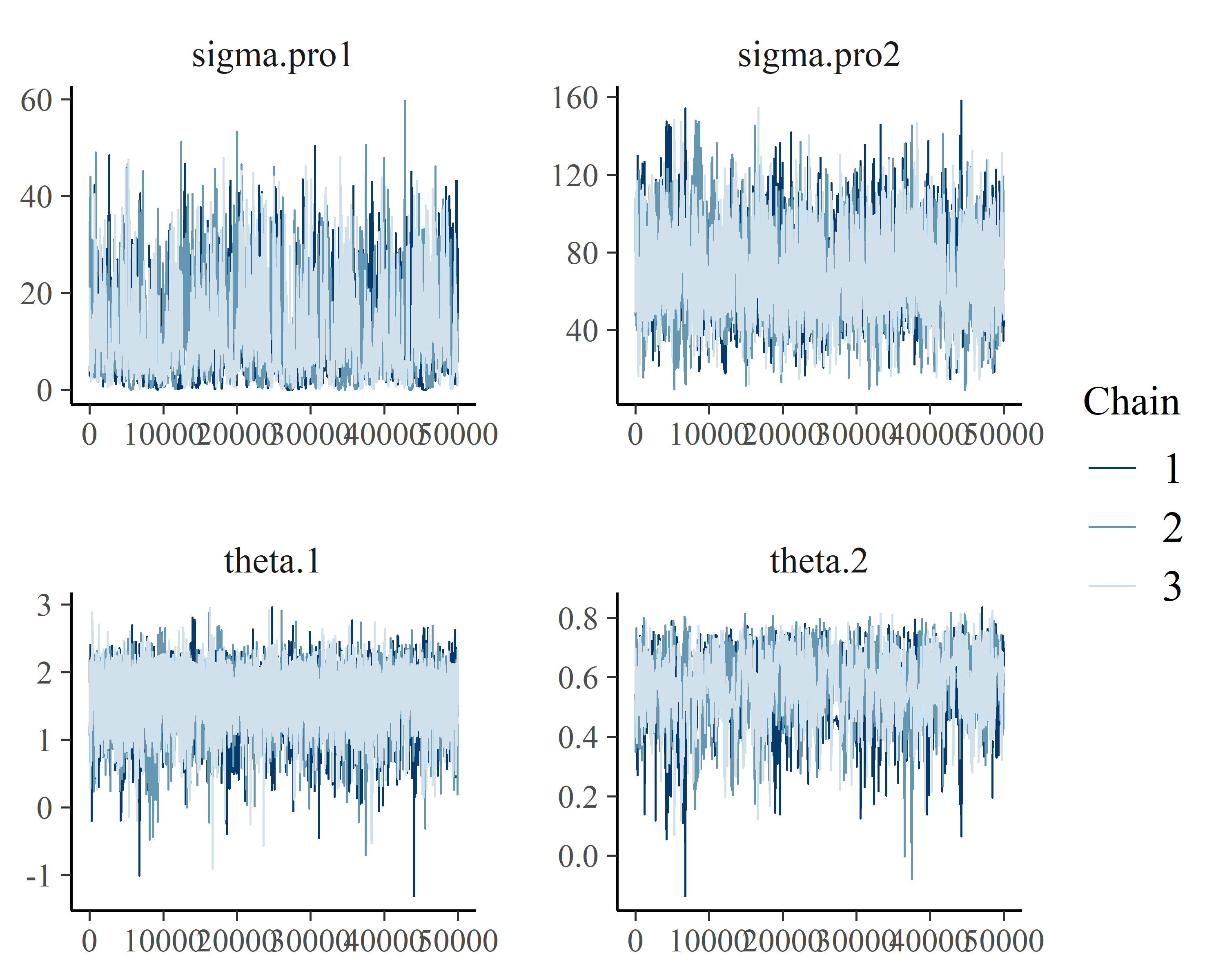


Figure 10. MCMC trace plots for Wermon dataset.

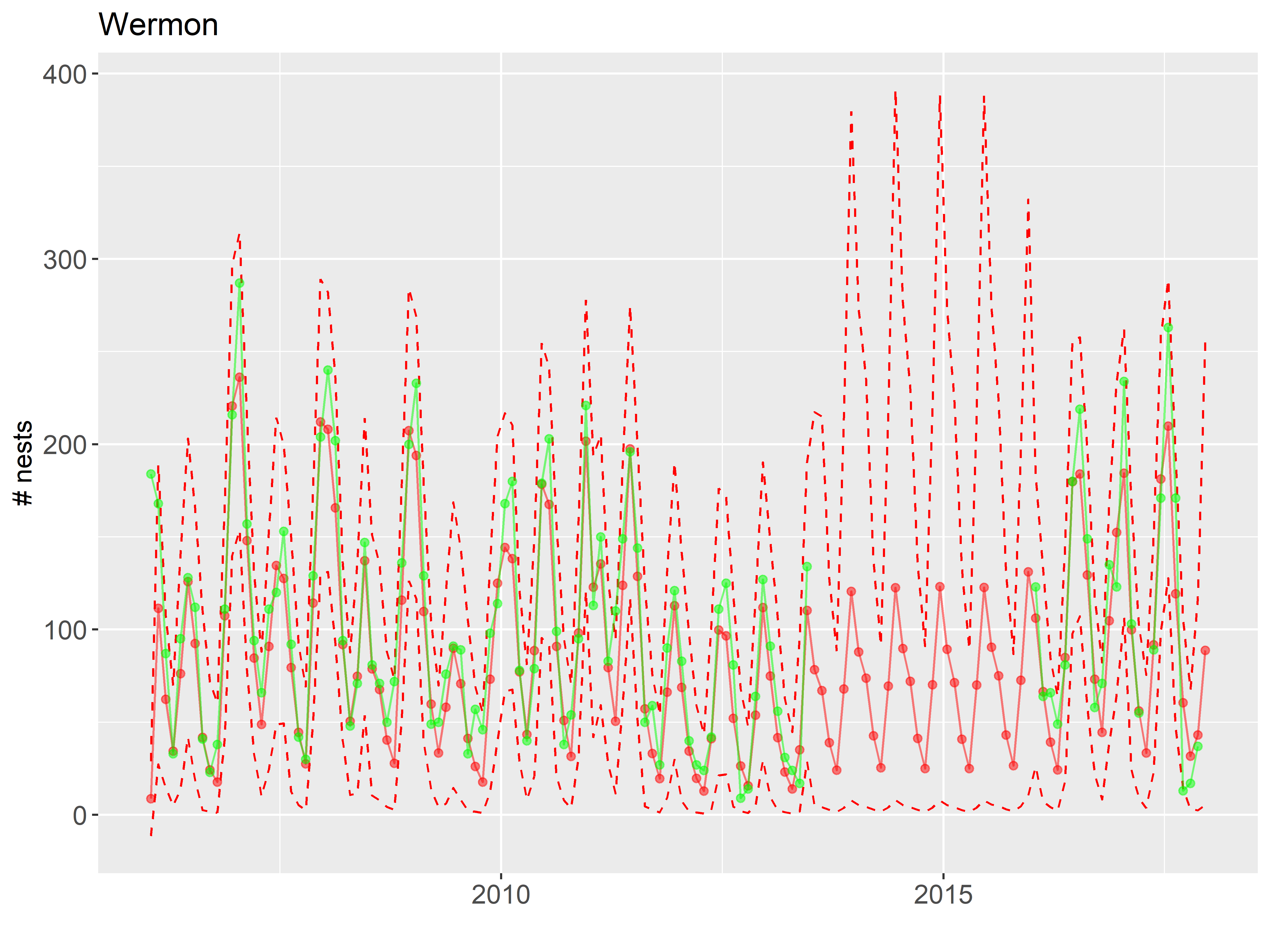


Figure 11. Predicted nest counts for Wermon dataset.

It is not useful to fill in those years with no data whatsoever as we have no idea what was happening (Figure 11). For other years, however, the estimated true counts (red) seem reasonable. We will use all years except 2014 and 2015. I pool estimated counts annually.

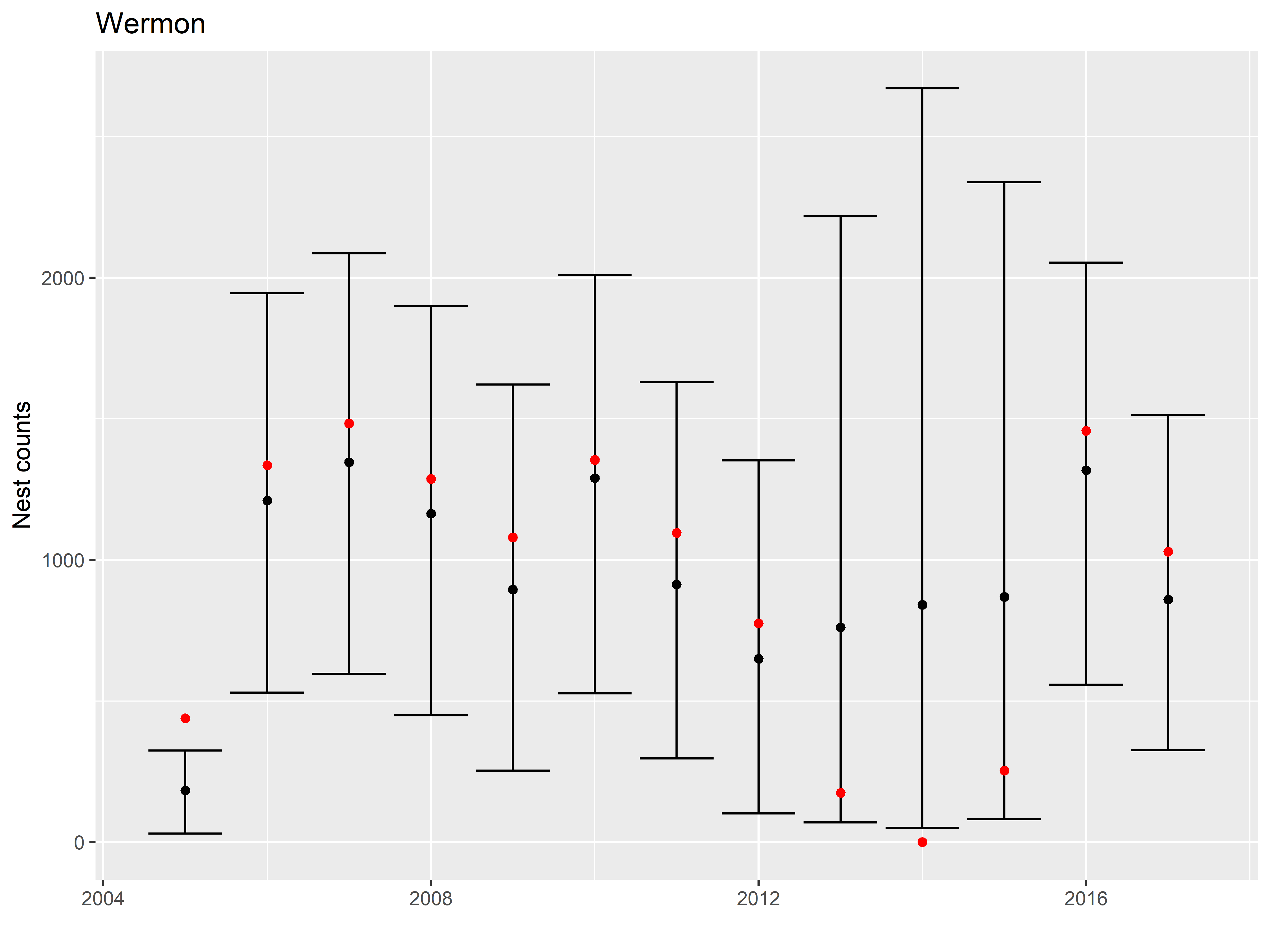


Figure 12. Predicted annual nest counts for Wermon.

# Summer vs winter nesting

In order to determine the summer and winter nest counts, we added Jamursba-Medi and Wermon counts for each season. Summer is defined as April 1 - September 30, whereas winter is from October 1 to March 31.

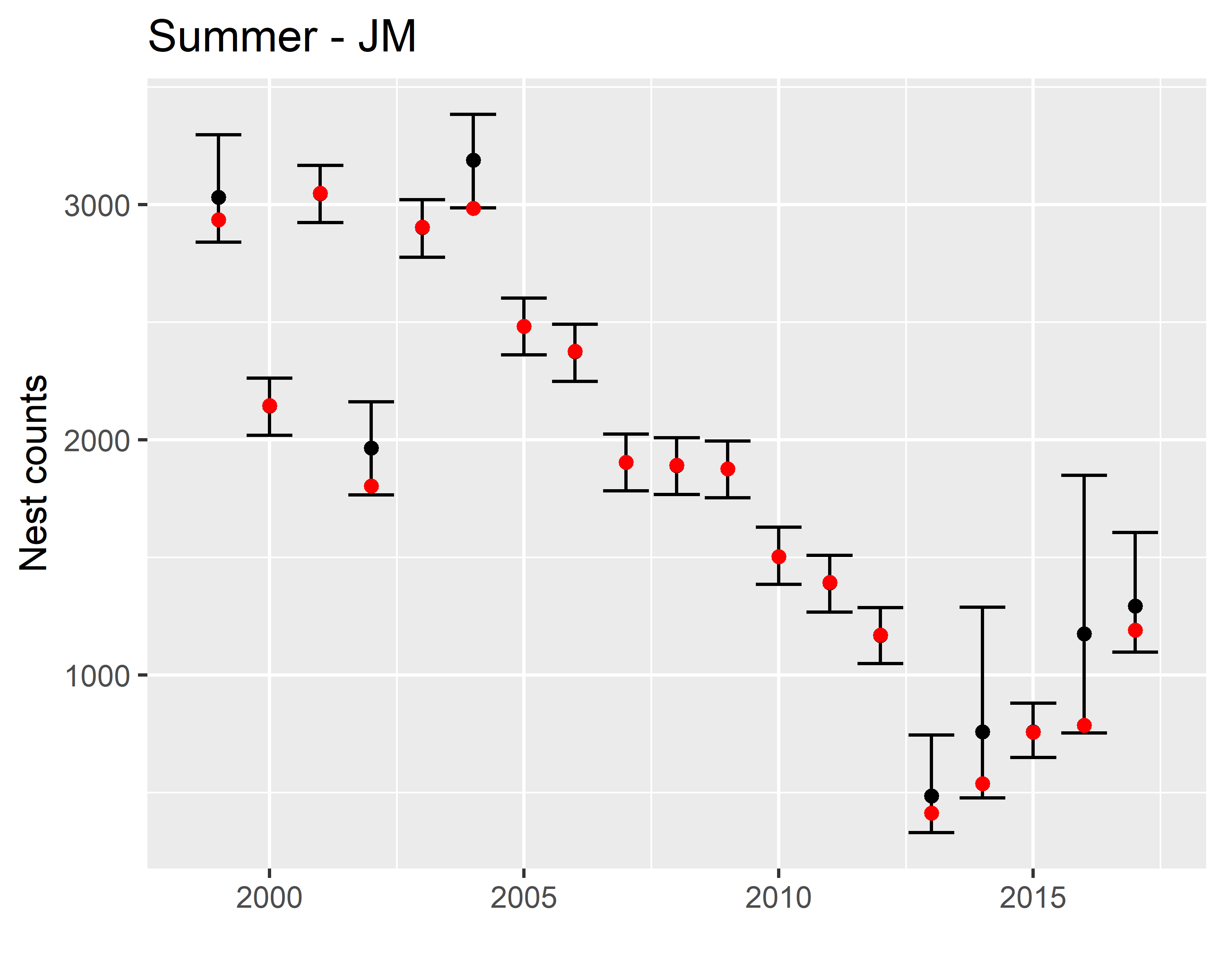


Figure 13. Predicted and observed annual nest counts during summer months for Jamursba-Medi.

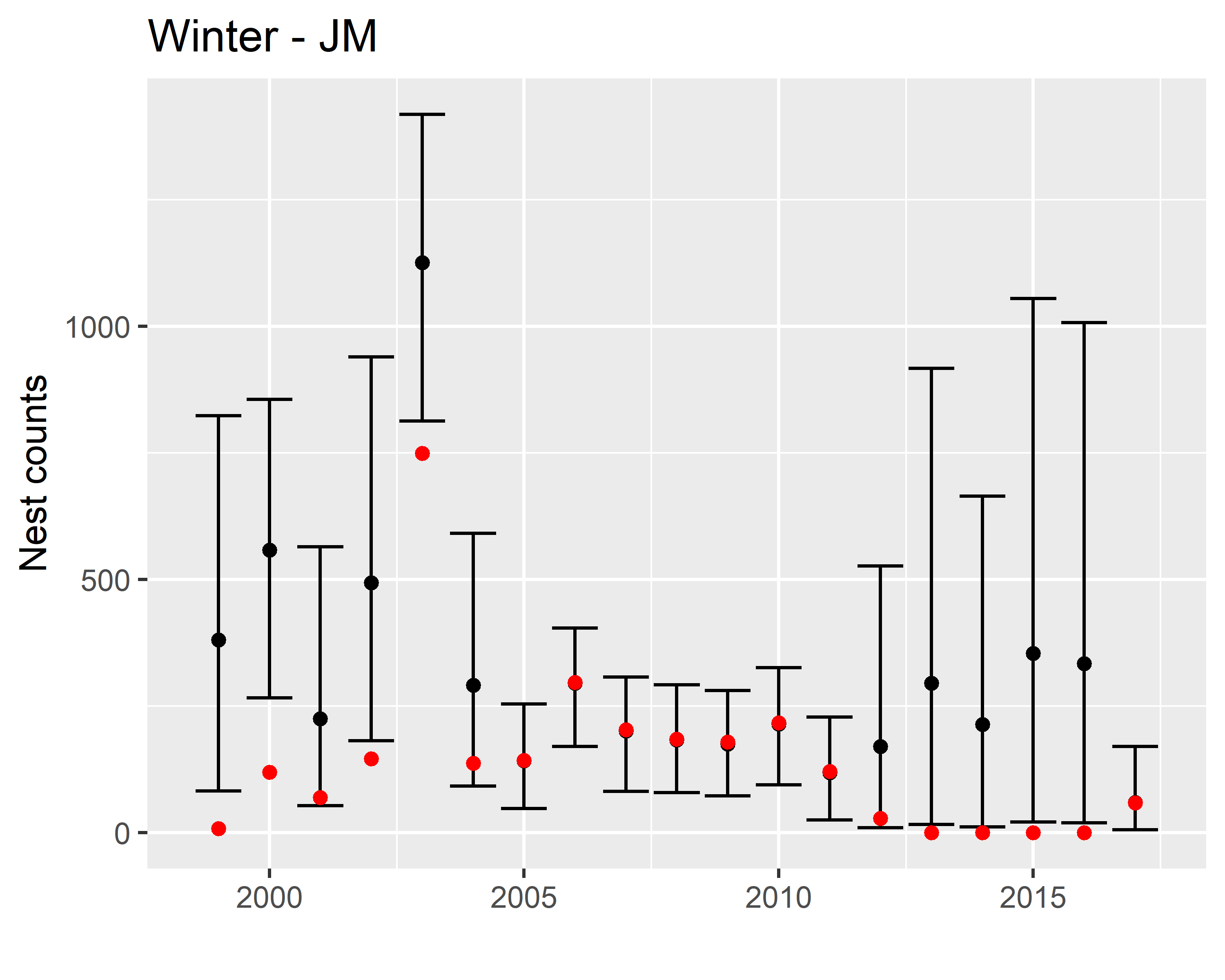


Figure 14. Predicted and observed annual nest counts during winter months for Jamursba-Medi.

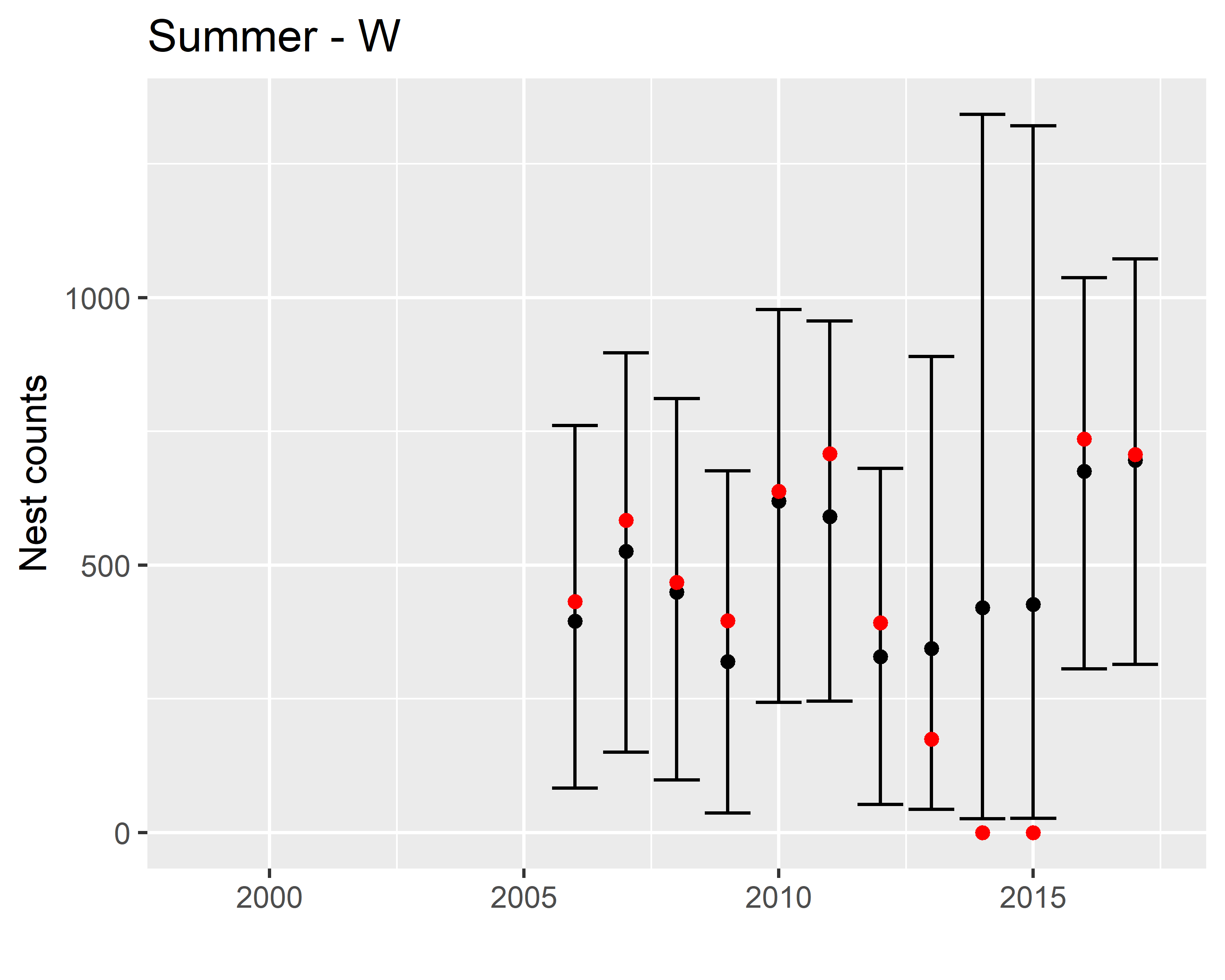


Figure 15. Predicted and observed annual nest counts during summer months for Wermon.

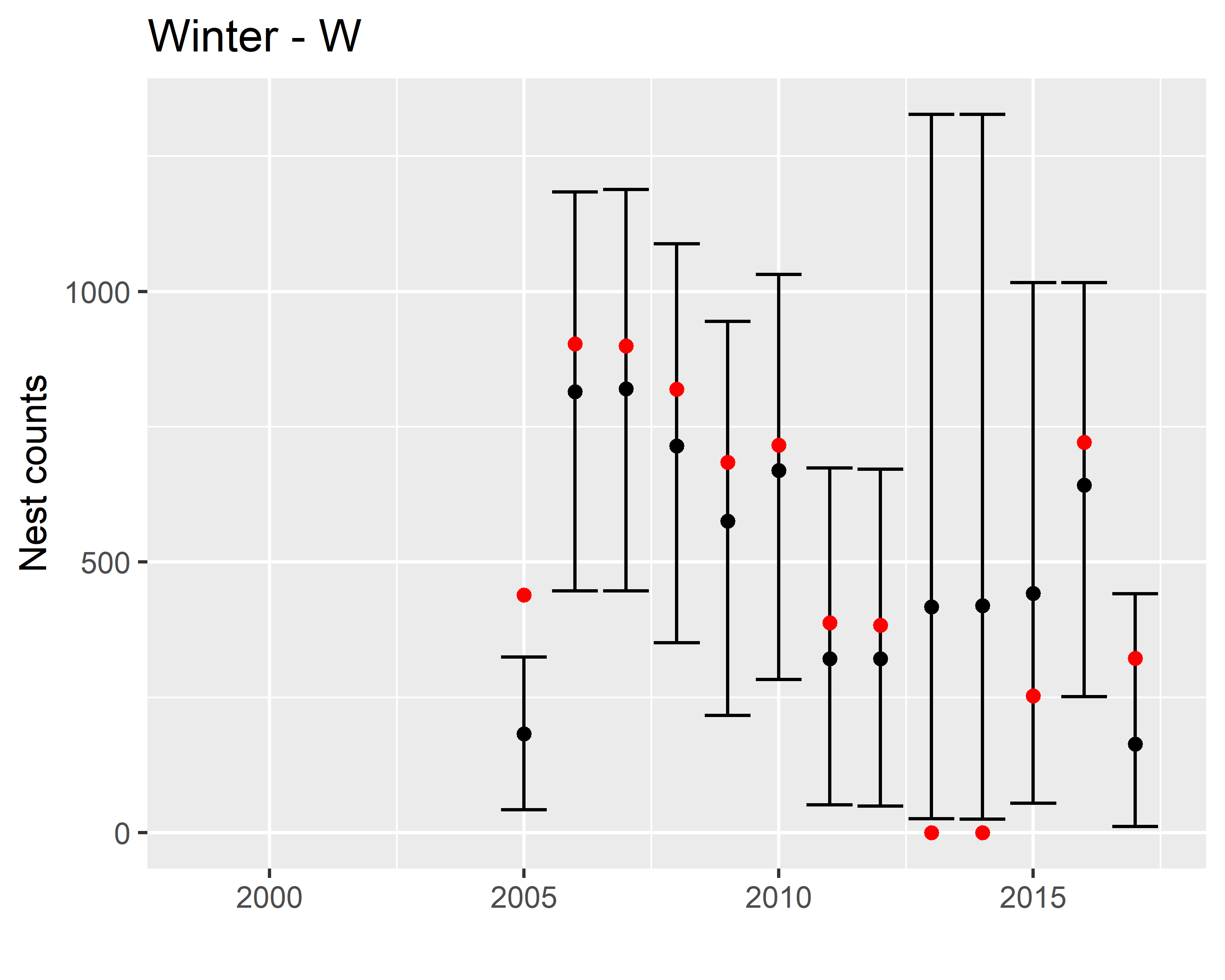


Figure 16. Predicted and observed annual nest counts during winter months for Wermon.

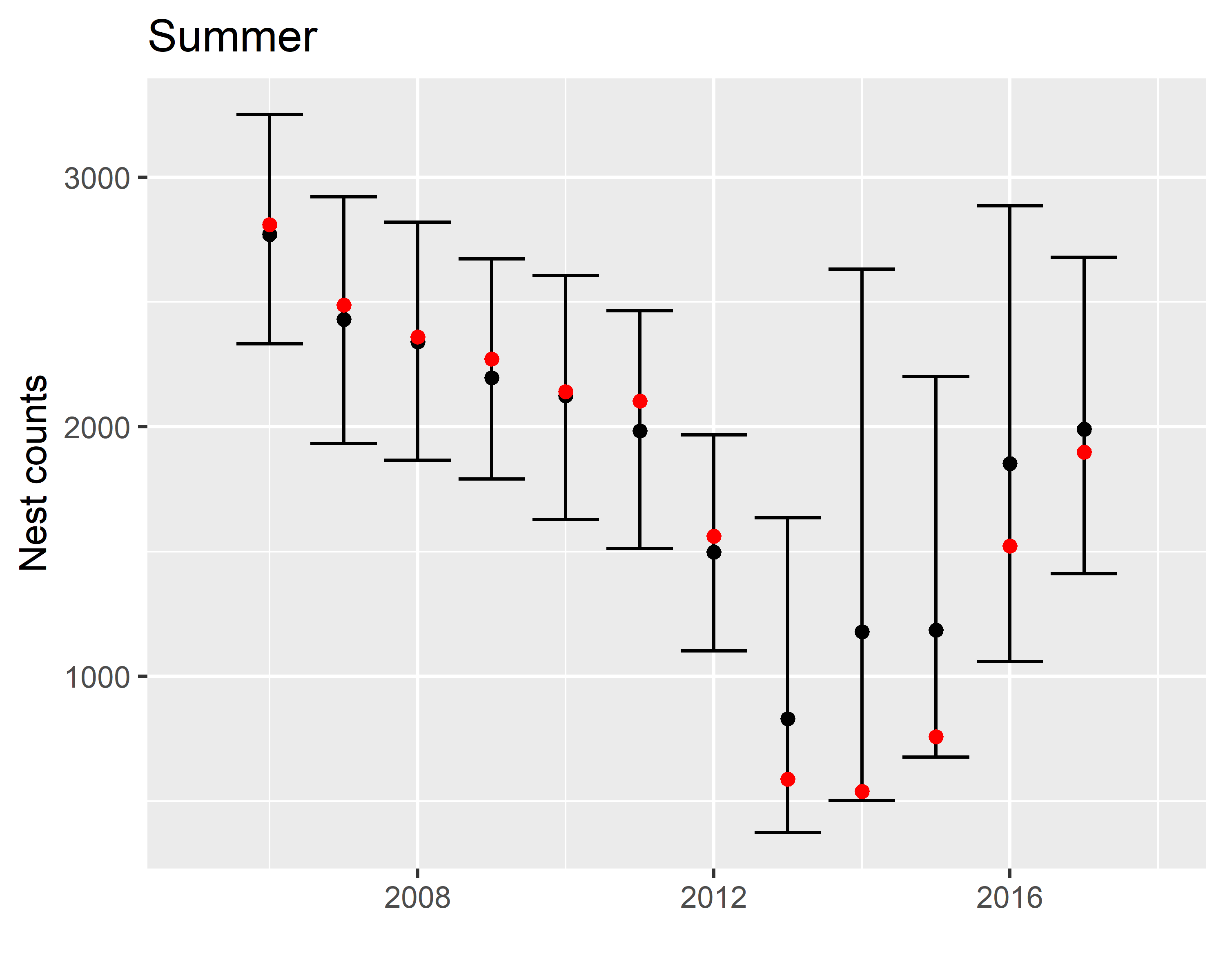


Figure 17. Predicted and observed annual nest counts during summer months at Jamursba-Medi and Wermon.

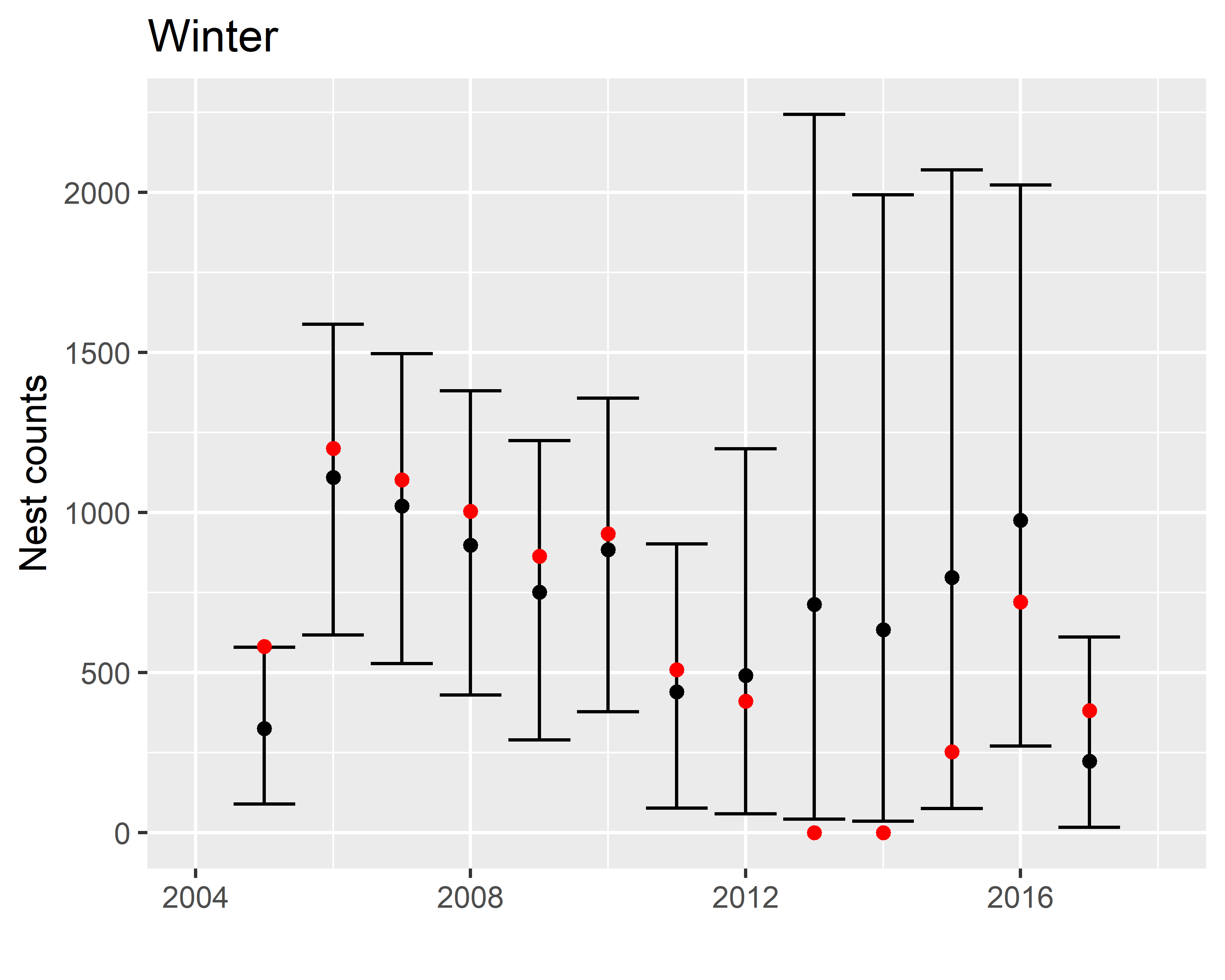


Figure 18. Predicted and observed annual nest counts during winter months at Jamursba-Medi and Wermon.

# Appendix

JAGS code for this analysis

## Jamursba-Medi

results.JM\_SSAR1\_month\_var\_theta$jm

## JAGS model:  
##   
## # simple state space AR1 model for turtle nesting   
## # original code from Lab 7, Introduction to Bayesian Time-Series  
## # analysis using Jags. Univ Washington:   
## # https://www.scribd.com/document/373707632/Lab-7-Fitting-models-with-JAGS-pdf  
##   
## # Also can be found here:  
## # https://nwfsc-timeseries.github.io/atsa-labs/sec-jags-uss.html  
## #  
## # For this model, I added another theta. Two thetas are used for increasing  
## # and decreasing months within each year.   
##   
## model{  
## for (t in 2:T){  
## # process  
## theta[t] <- ifelse(m[t] < 8, theta.1, theta.2)  
## predX[t] <- theta[t] \* X[t-1]  
##   
## tau.pro[t] <- ifelse(m[t] < 9 && m[t] > 4, tau.pro1, tau.pro2)  
## X[t] ~ dnorm(predX[t], tau.pro[t])T(0,)  
##   
## # observation  
## predY[t] <- X[t]  
## y[t] ~ dnorm(X[t], tau.obs)T(0,)  
## }  
##   
## X[1] <- mu  
## predY[1] <- X[1]  
## y[1] ~ dnorm(X[1], tau.obs)T(0,)  
##   
## mu ~ dnorm(0, 0.01)  
## #tau.pro1 ~ scaled.gamma(100, 2)  
## sigma.pro1 ~ dunif(0, 600)  
## tau.pro1 <- 1/(sigma.pro1 \* sigma.pro1)  
## #sigma.pro1 <- 1/sqrt(tau.pro1)  
##   
## #tau.pro2 ~ dscaled.gamma(50, 2)  
## #sigma.pro2 <- 1/sqrt(ta.pro2)  
## sigma.pro2 ~ dunif(0, 100)  
## tau.pro2 <- 1/(sigma.pro2 \* sigma.pro2)  
##   
## #tau.obs ~ dscaled.gamma(10, 2)  
## #sigma.obs <- 1/sqrt(tau.obs)  
## sigma.obs ~ dunif(0, 100)  
## tau.obs <- 1/(sigma.obs \* sigma.obs)  
##   
## theta.1 ~ dnorm(0, 0.010)  
## theta.2 ~ dnorm(0, 0.010)  
## }  
## Fully observed variables:  
## T m   
## Partially observed variables:  
## y

## Wermon

results.W\_SSAR1\_month\_var\_theta$jm

## JAGS model:  
##   
## # simple state space AR1 model for turtle nesting   
## # original code from Lab 7, Introduction to Bayesian Time-Series  
## # analysis using Jags. Univ Washington:   
## # https://www.scribd.com/document/373707632/Lab-7-Fitting-models-with-JAGS-pdf  
##   
## # Also can be found here:  
## # https://nwfsc-timeseries.github.io/atsa-labs/sec-jags-uss.html  
## #  
## # For this model, I added another theta. Two thetas are used for increasing  
## # and decreasing months within each year.   
##   
## model{  
## for (t in 2:T){  
## # process  
## theta[t] <- ifelse(m[t] ==5 || m[t] == 6 || m[t] == 11 || m[t] == 12, theta.1, theta.2)  
## predX[t] <- theta[t] \* X[t-1]  
##   
## tau.pro[t] <- ifelse(m[t] == 3 || m[t] == 4 || m[t] == 9 || m[t] == 10, tau.pro1, tau.pro2)  
## X[t] ~ dnorm(predX[t], tau.pro[t])T(0,)  
##   
## # observation  
## predY[t] <- X[t]  
## y[t] ~ dnorm(X[t], tau.obs)T(0,)  
##   
## }  
##   
## X[1] <- mu  
## predY[1] <- X[1]  
## y[1] ~ dnorm(X[1], tau.obs)T(0,)  
##   
## mu ~ dnorm(0, 0.01)  
## #tau.pro1 ~ scaled.gamma(100, 2)  
## sigma.pro1 ~ dunif(0, 600)  
## tau.pro1 <- 1/(sigma.pro1 \* sigma.pro1)  
## #sigma.pro1 <- 1/sqrt(tau.pro1)  
##   
## #tau.pro2 ~ dscaled.gamma(50, 2)  
## #sigma.pro2 <- 1/sqrt(ta.pro2)  
## sigma.pro2 ~ dunif(0, 600)  
## tau.pro2 <- 1/(sigma.pro2 \* sigma.pro2)  
##   
## #tau.obs ~ dscaled.gamma(10, 2)  
## #sigma.obs <- 1/sqrt(tau.obs)  
## sigma.obs ~ dunif(0, 200)  
## tau.obs <- 1/(sigma.obs \* sigma.obs)  
##   
## theta.1 ~ dnorm(0, 0.010)  
## theta.2 ~ dnorm(0, 0.010)  
## }  
## Fully observed variables:  
## T m   
## Partially observed variables:  
## y