Chignik Sockeye Early Run Modeling

Sarah Power

2018-09-28

Recipient

Heather Finkle.

Background

For over a decade genetic samples of sockeye (Oncorhynchus nerka) have been collected at the Chignik weir to distinguish the early run, which mostly spawns in Black Lake from later runs. The majority of these later runs spawn in Chignik Lake. Currently these stocks are managed separately. The processing of genetics samples is expensive, and therefore alternative methods for distinguishing between the stocks are being examined. In the years prior to genetic assignment July 4th was used as a cut off date between the early and late runs with the understanding that some early run fish arrive later and some later run fish arrive earlier.

This analysis examines the run timing distributions to determine how well early run stocks can be discerned from the later ones on run timing alone. We will compare this to the run timing distributions that are determined with the additional aid of genetics to see how viable this method might be.

Currently the run consists of the escapement as captured by the weir, as well as catch from various managment districts, some of which are quite far away.

Method

Creelman (2011) hypothesized that the early run and later runs are comprised of various smaller runs with their own run timing and run locations. She found that run timing played an important role in genetic variation. This makes distinguishing run based on run timing a possibility. She grouped collections into four genetic groups: Black Lake, Chignik Lake, Chignik Lake October, and Chignik River. Looking at the distributions across the years there appears to be three modes. Noting that data from the weir is collected from about late may until the end of September, and that Creelman noted an October Chignik Lake run the assumption is that there is a biological reason for three modes to be presented in the data.

I used maximum likelihood methods including the EM algorithm and a Newton-type algorithm to parse out various distributions using the mixdist package in R.

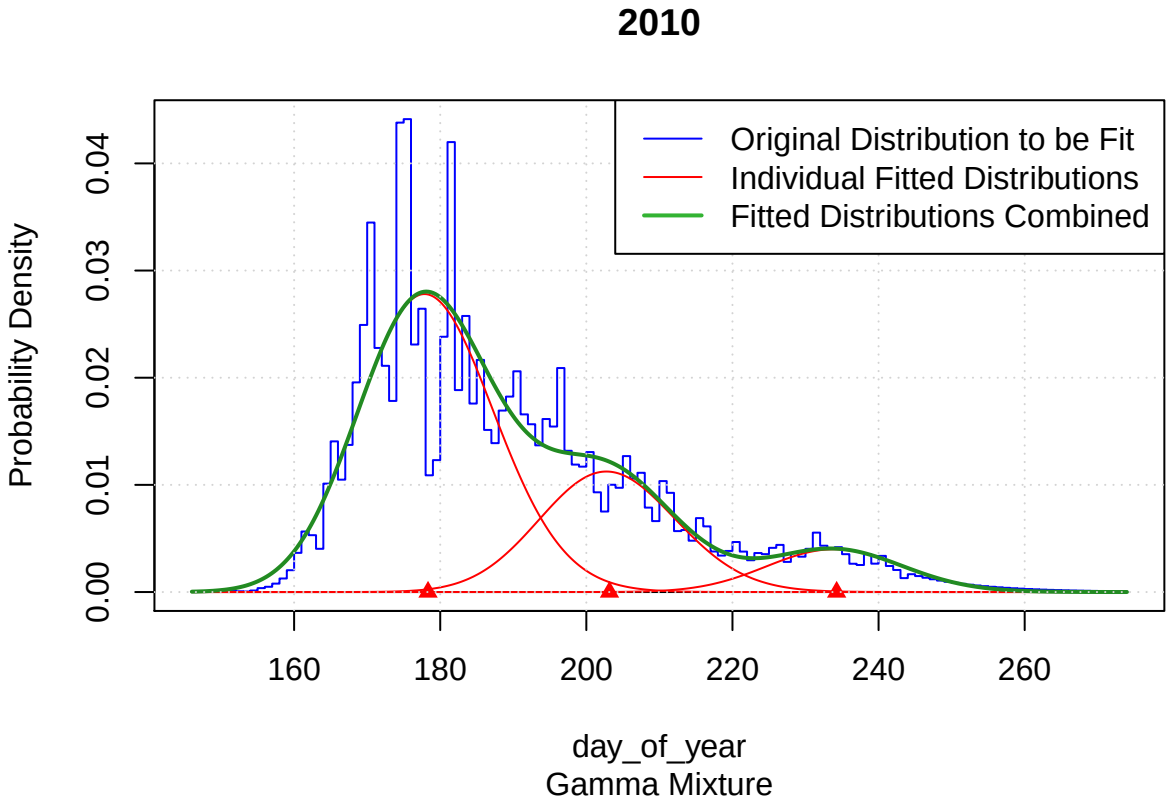
Firstly the distribution of the early run was estimated using run timing and genetic assignment for each year. All other models that used run timing exclusively were compared to these distributions.

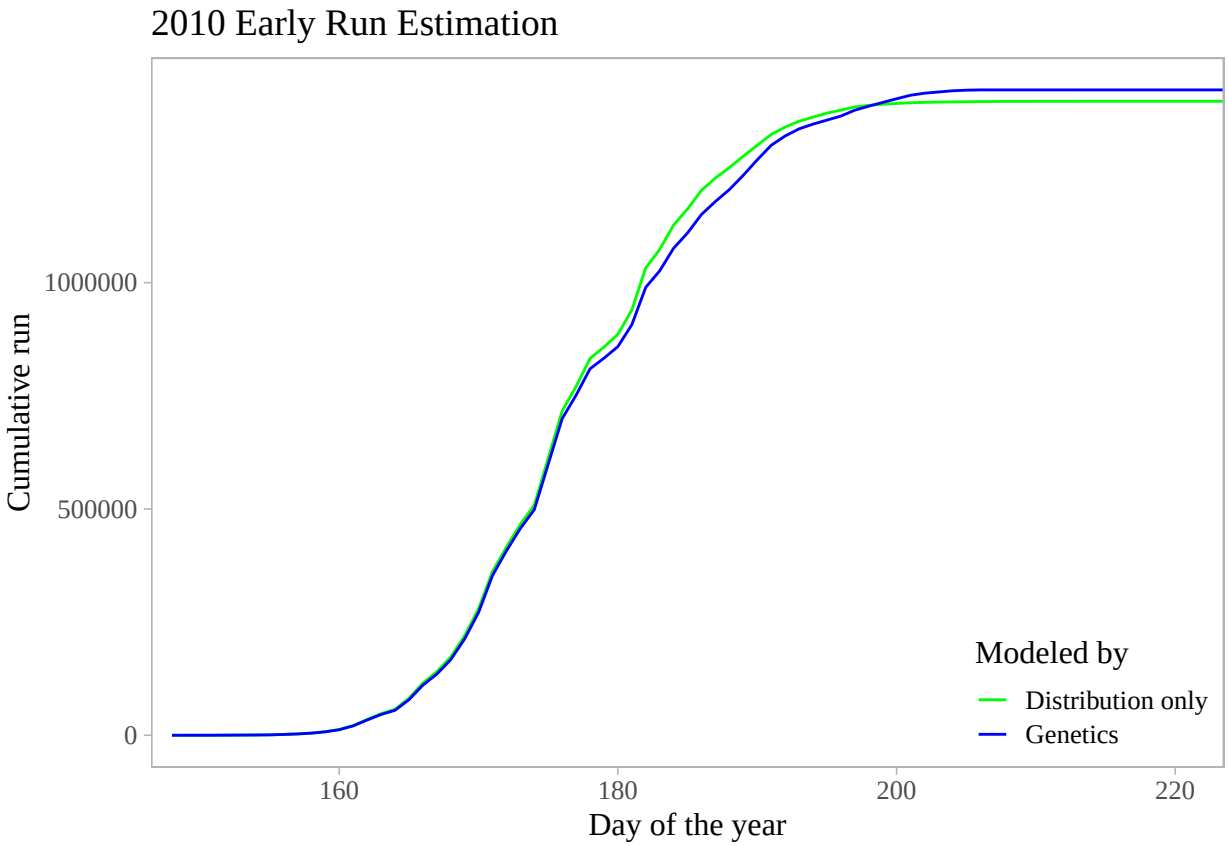
In initial investigations Normal and Weibull distributions were fit to the multimodal run timing data, also certain constraints were considered in order to find the best fit. Those constraints included equal standard errors; means equally spaced; and having the mean or standard error of the early distribution fixed. The fixed means and standard errors were the medians values from the early distribution determined using genetics.

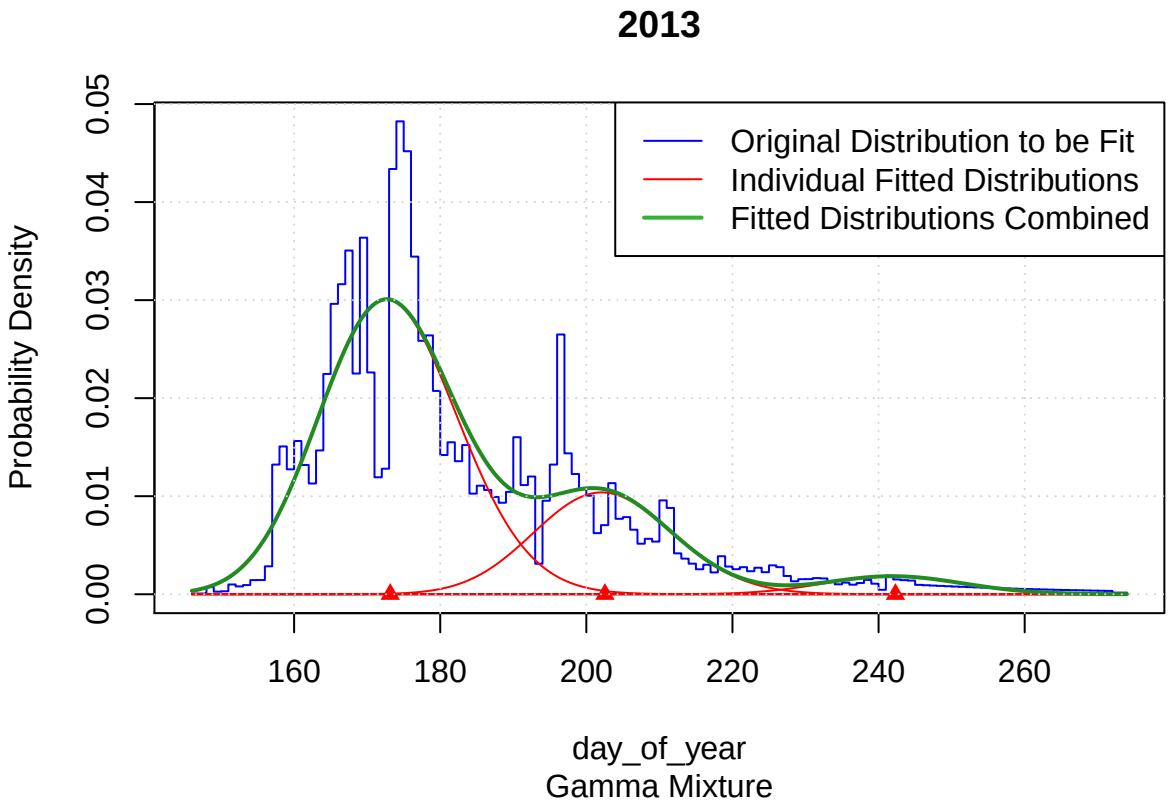
It was determined that the most appropriate fitting model was a trimodal normally distributed model with the constraint of equal standard deviations for each of the distributions. This was determined as this model had an early run with means and standard deviations most like the estimated early run distributions as determined with genetics.

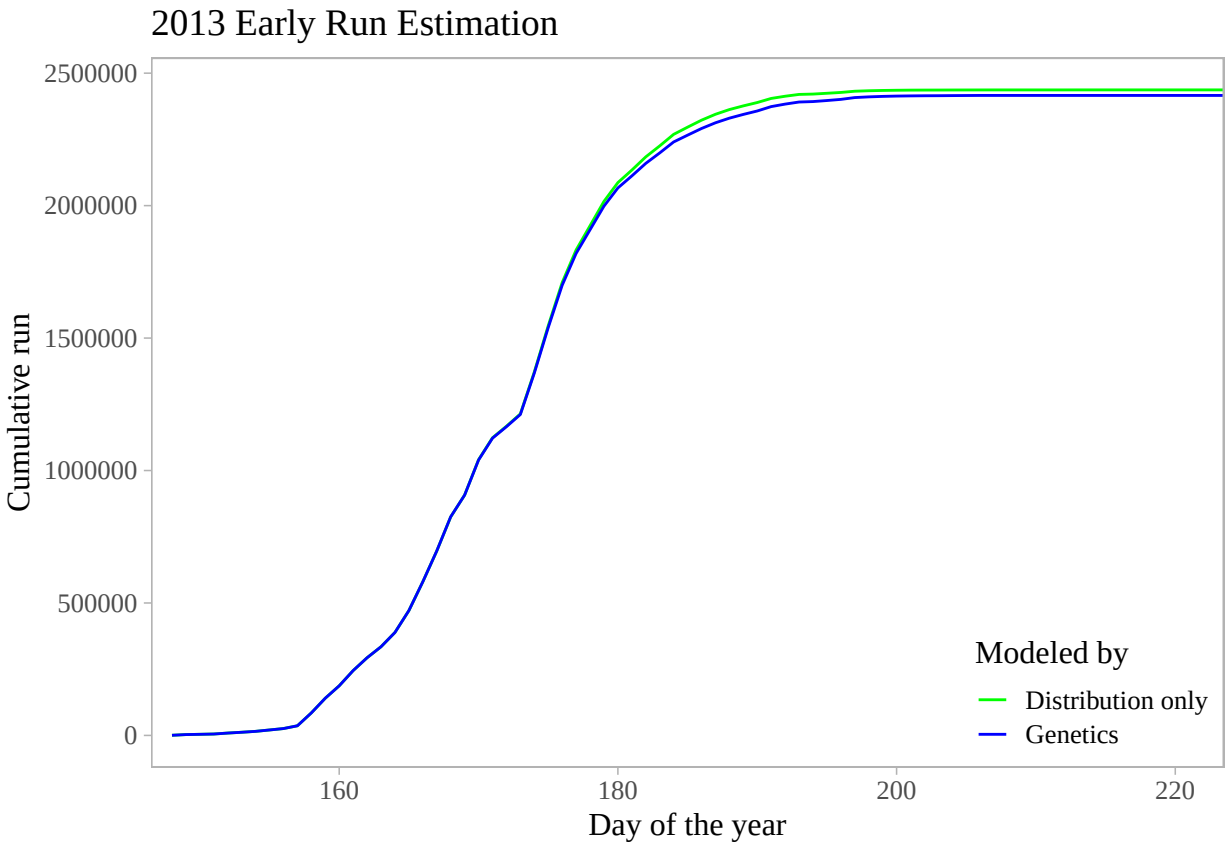
With this model for 8 of the 11 years early run counts using run timing alone were with 10% of the counts using additional genetic information. The three best years included 2010, 2013, and 2014, where early run counts using run timing information alone differed by no more than 3% of the counts using additional genetic information.

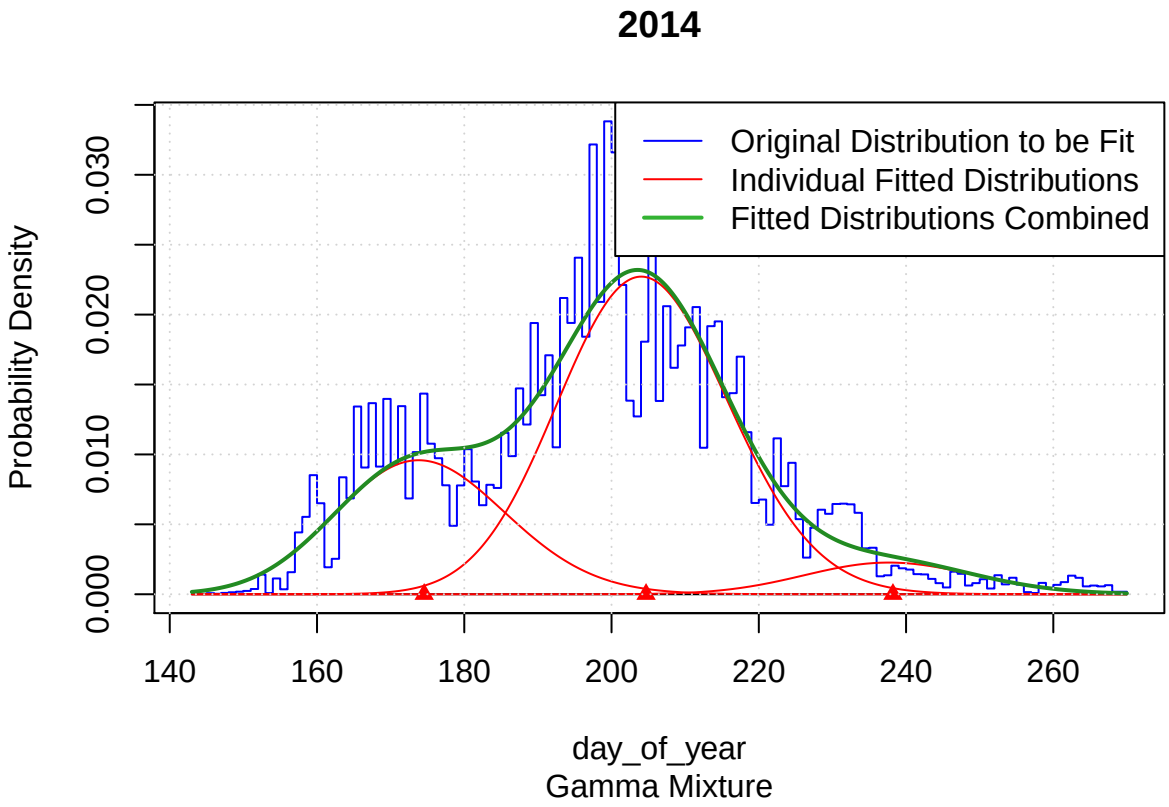
Once this model was determined to be most appropriate the run counts of each of the years were examined in order to determine if the model could be useful. For 8 of the 11 years between 2006 and 2017 early run counts using run timing alone were within 10% of the counts using genetic information. The three best years included 2010, 2013, and 2014, where early run counts using run timing information alone differed by no more than 3% of the counts using additional genetic information. (Shown below)

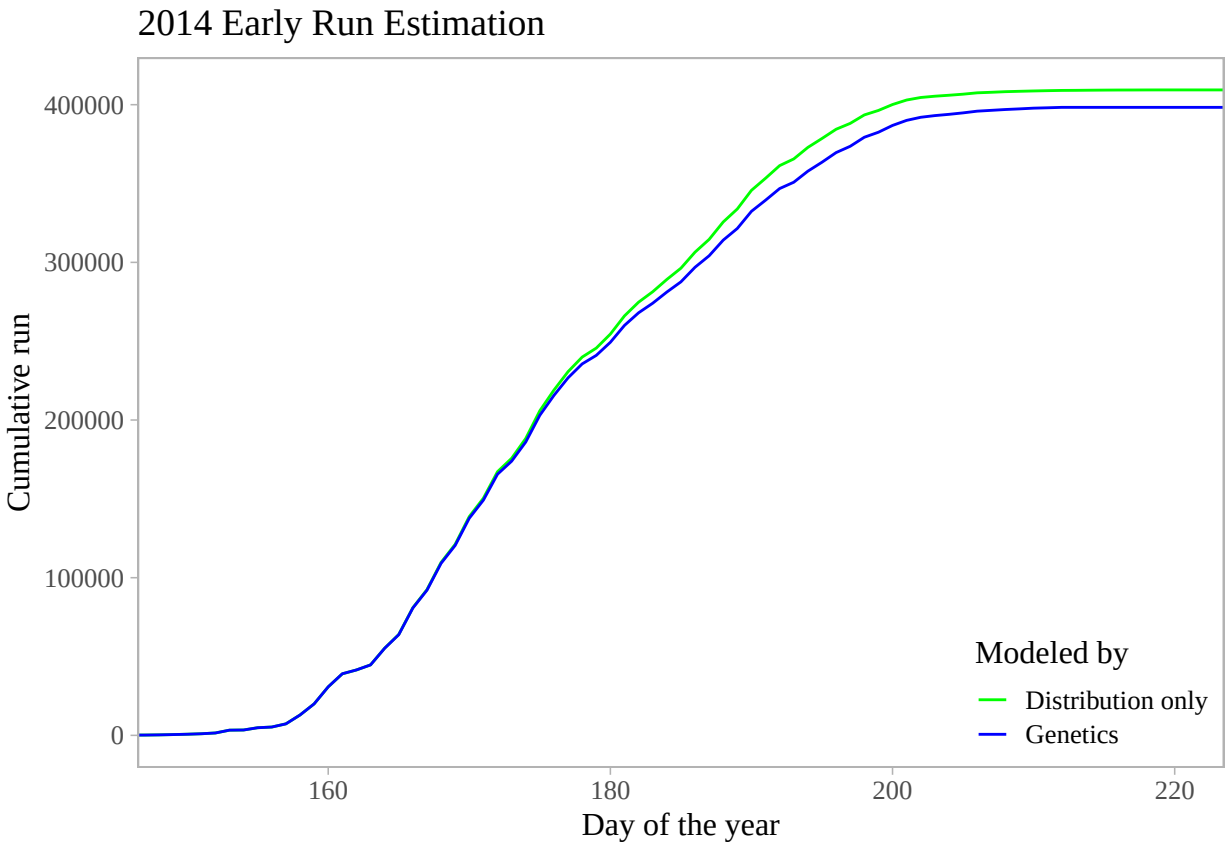



2010 Early Run Estimation
Cumulative run
1000000
500000
0
160
180
200
220
Day of the year
Modeled by
Distribution only
Genetics

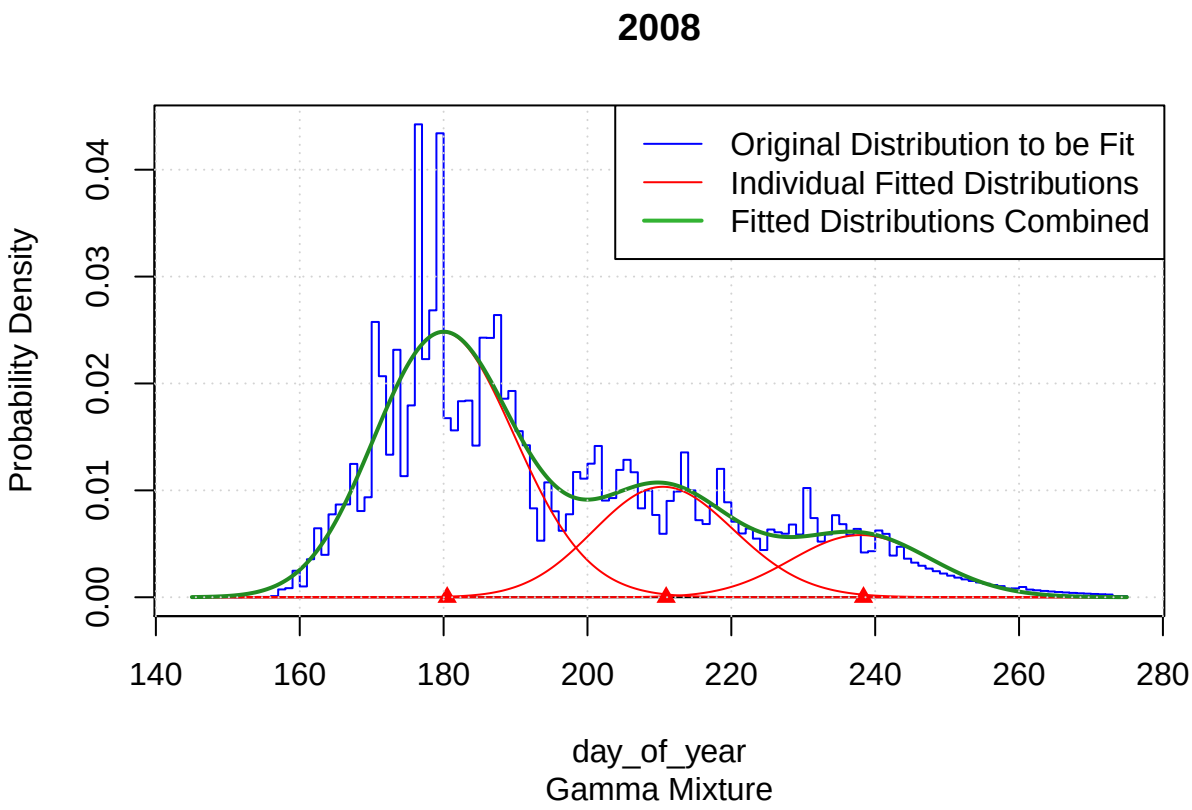

2013
Original Distribution to be Fit
Individual Fitted Distributions
Fitted Distributions Combined
Probability Density
0.00
0.01
0.02
0.03
0.04
0.05
160
180
200
220
240
260
day_of_year
Gamma Mixture

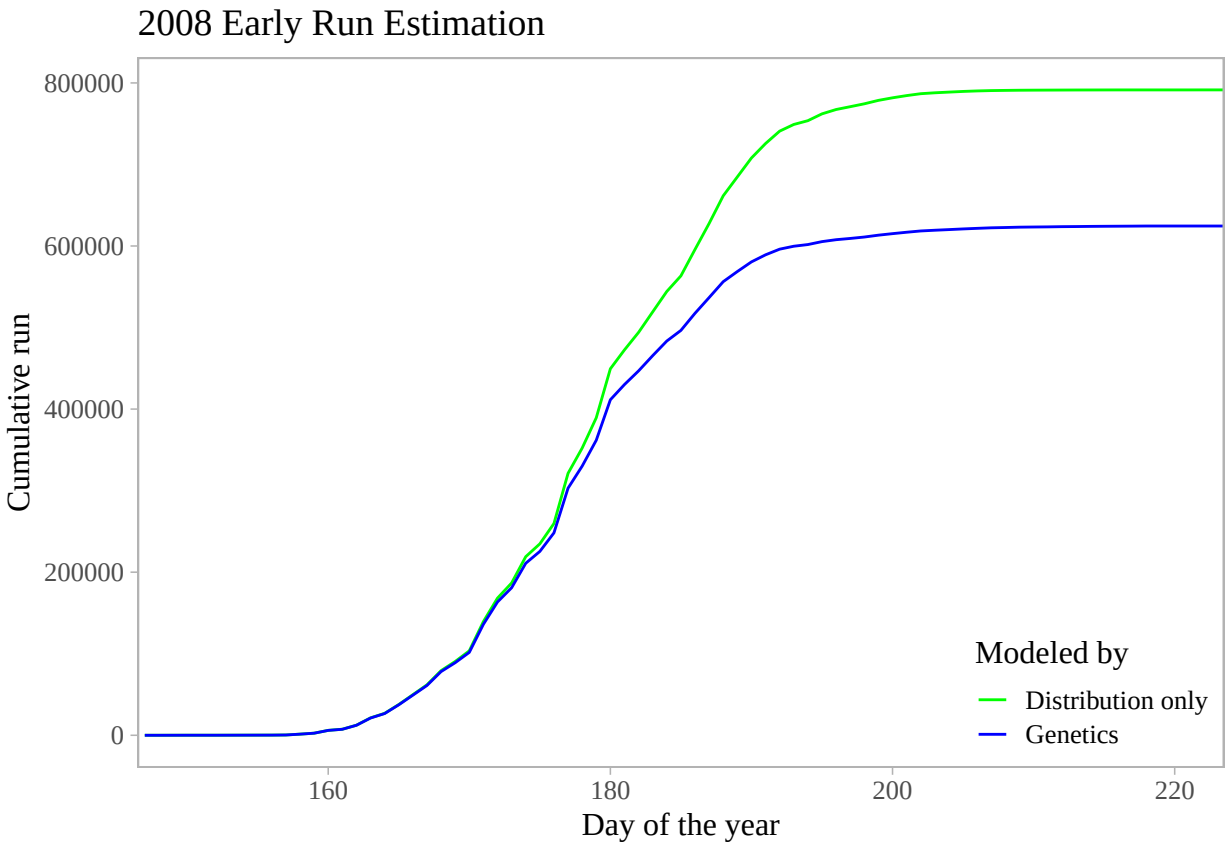

2013 Early Run Estimation
2500000
2000000
1500000
1000000
500000
0
Cumulative run
160
180
200
220
Day of the year
Modeled by
Distribution only
Genetics

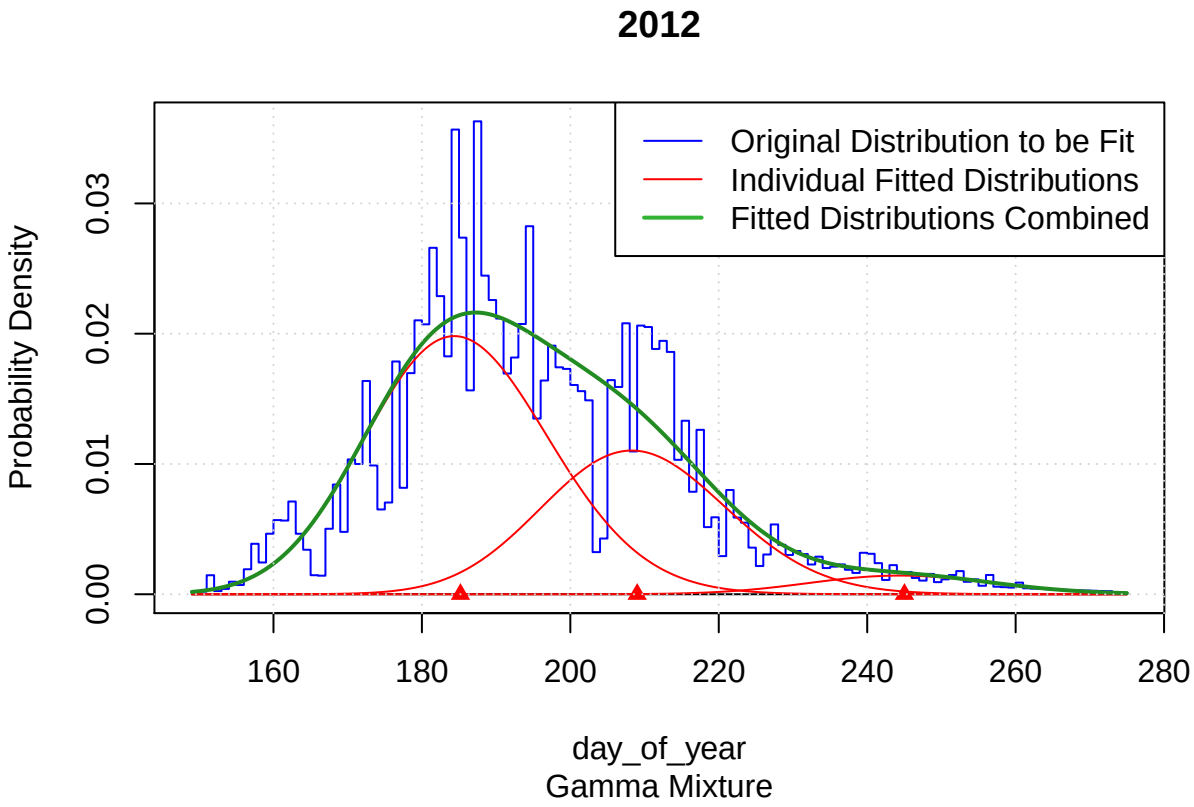

2014
Original Distribution to be Fit
Individual Fitted Distributions
Fitted Distributions Combined
Probability Density
0.000
0.010
0.020
0.030
140
160
180
200
220
240
260
day_of_year
Gamma Mixture

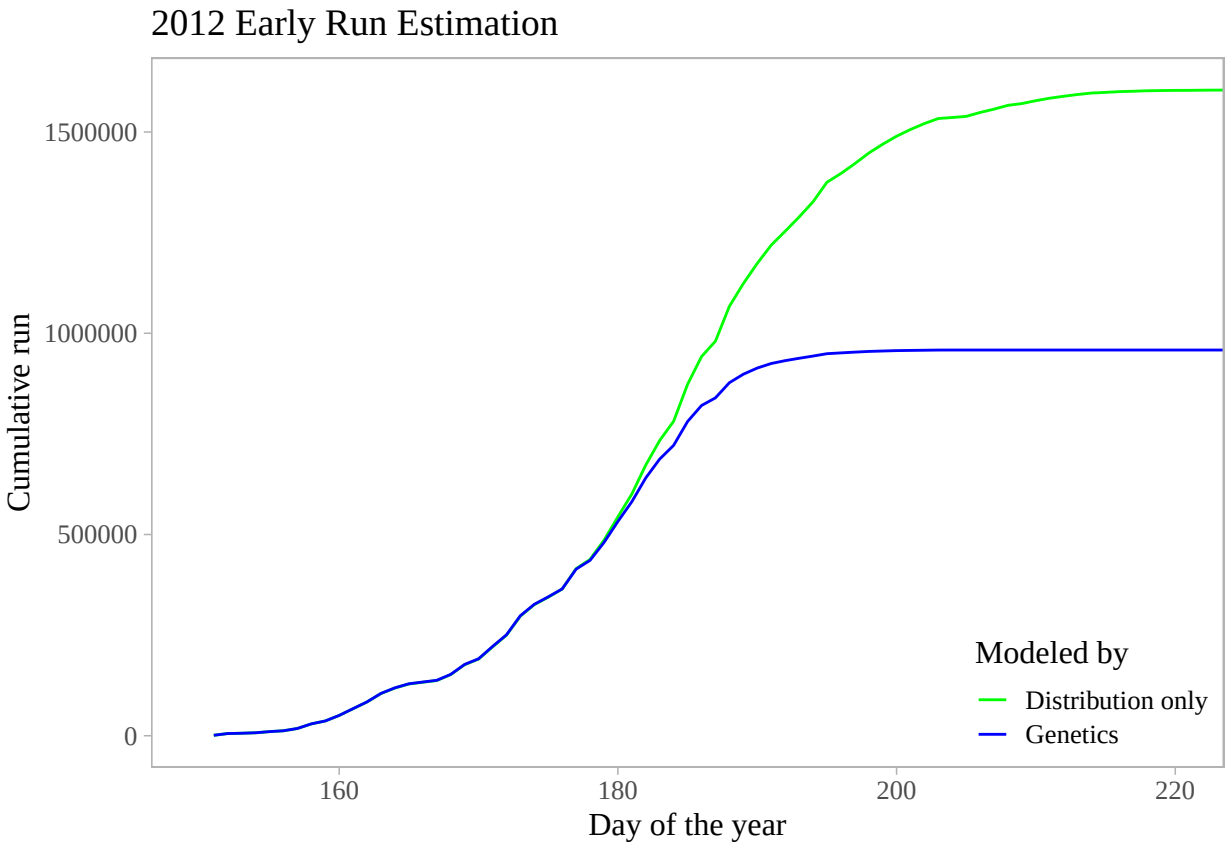


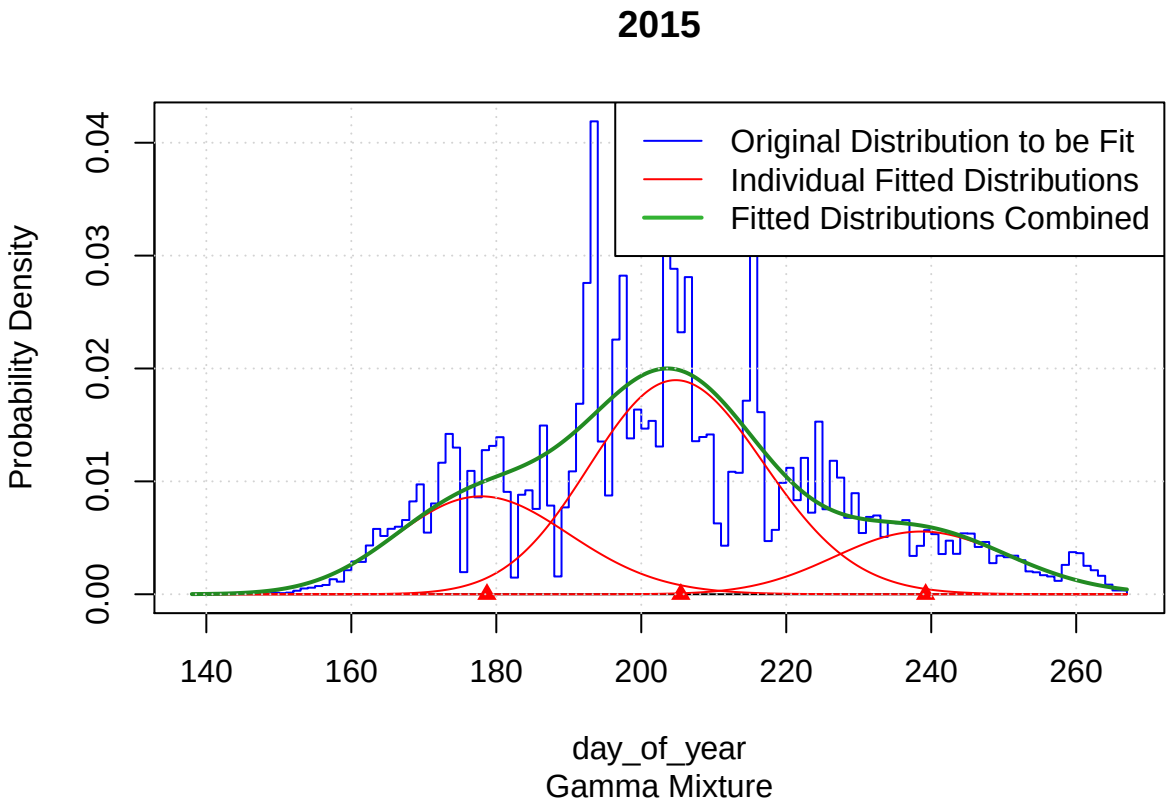
The three years that run timing distributions were off by more than 10% included 2008, which overestimated the early run by 27%, 2015 where it underestimated the early run by 25% and 2012 where is over estimated the run by a whopping 67%.

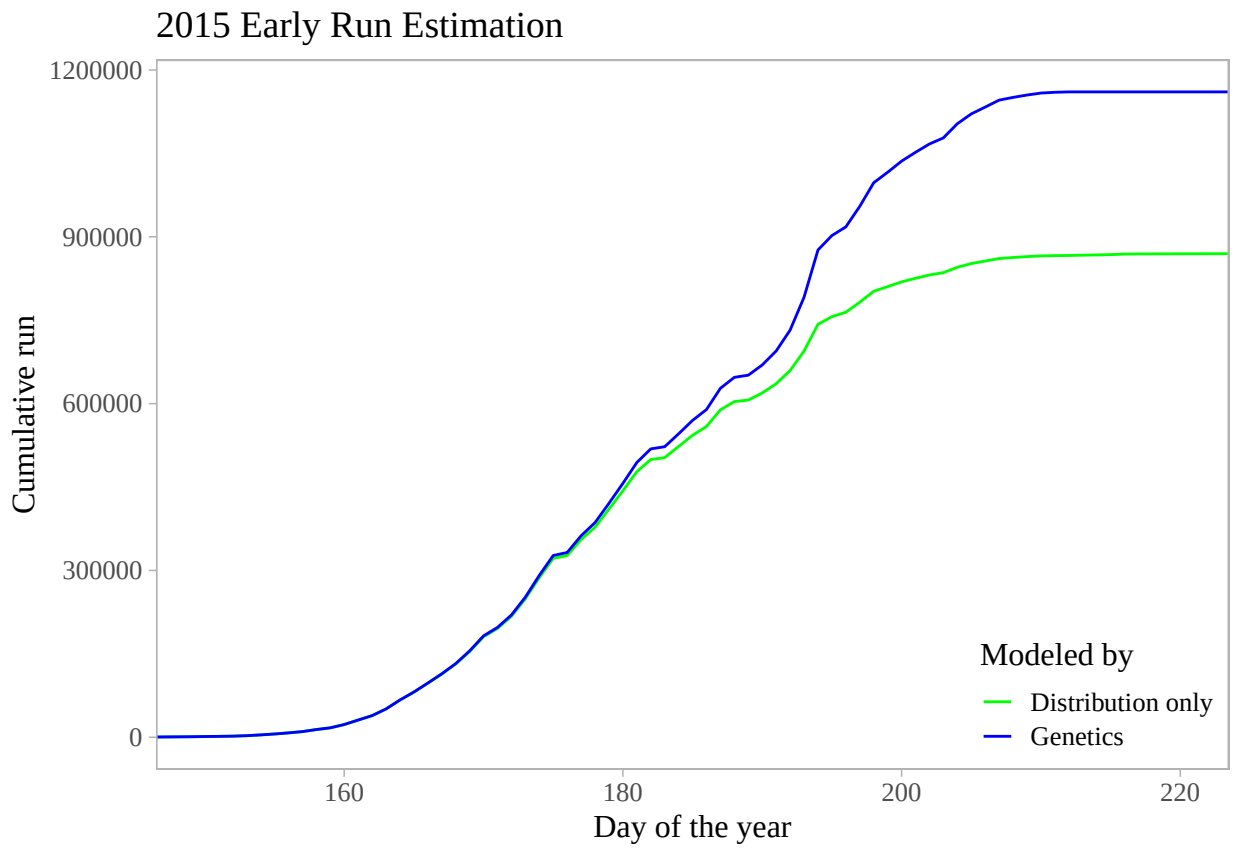

2008
Original Distribution to be Fit
Individual Fitted Distributions
Fitted Distributions Combined
Probability Density
0.00
0.01
0.02
0.03
0.04
140
160
180
200
220
240
260
280
day_of_year
Gamma Mixture


2008 Early Run Estimation
800000
600000
400000
200000
0
Cumulative run
160
180
200
220
Day of the year
Modeled by
Distribution only
Genetics

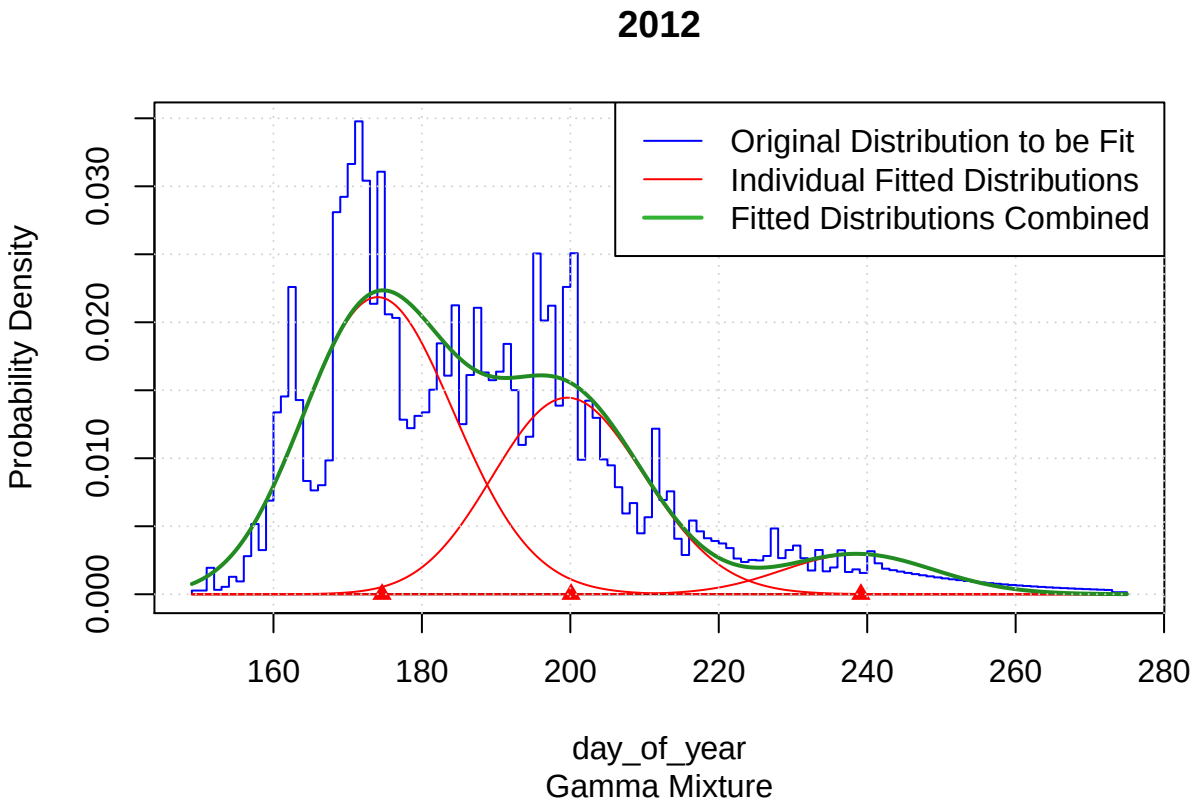

2012
Original Distribution to be Fit
Individual Fitted Distributions
Fitted Distributions Combined
Probability Density
0.00
0.01
0.02
0.03
160
180
200
220
240
260
280
day_of_year
Gamma Mixture

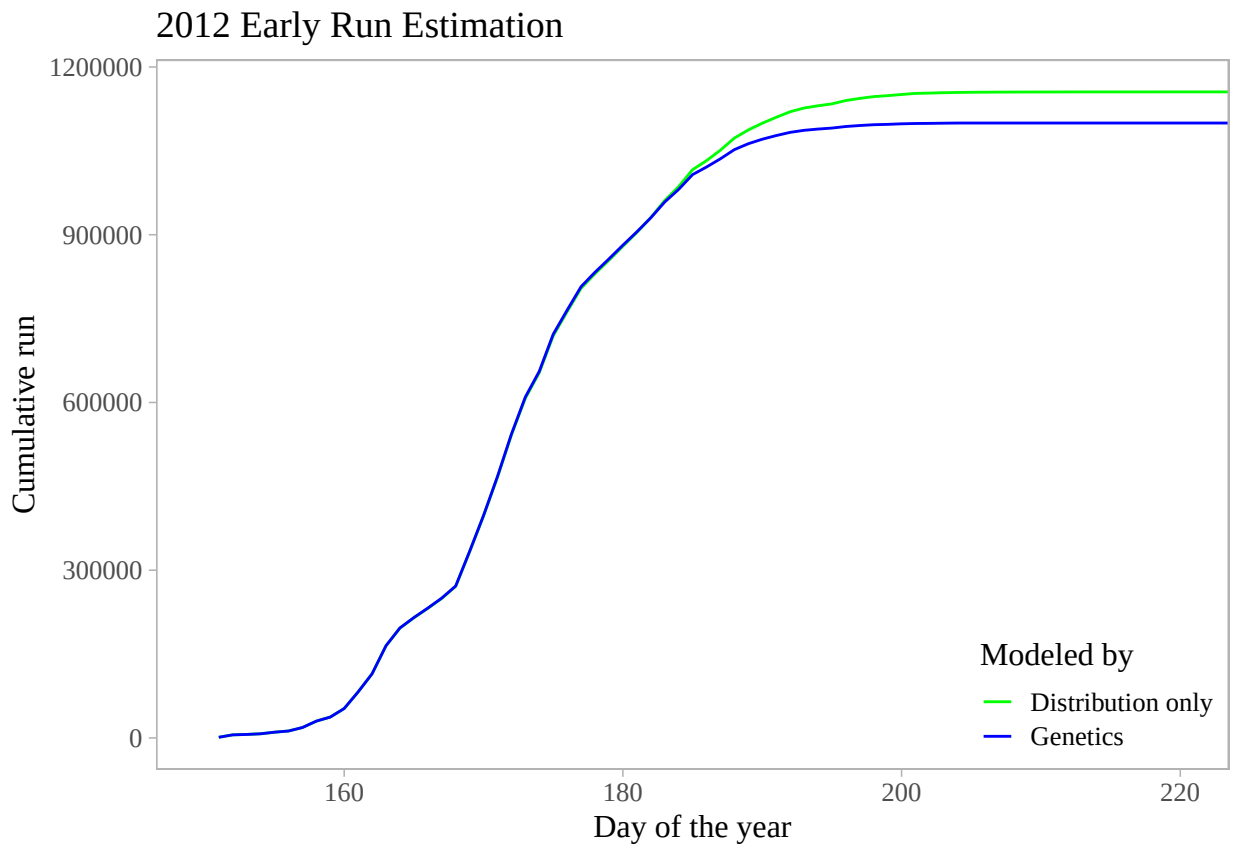

2012 Early Run Estimation
Cumulative run
1500000
1000000
500000
0
160
180
200
220
Day of the year
Modeled by
Distribution only
Genetics


2015
Original Distribution to be Fit
Individual Fitted Distributions
Fitted Distributions Combined
Probability Density
0.00
0.01
0.02
0.03
0.04
140
160
180
200
220
240
260
day_of_year
Gamma Mixture



It was suggested by Heather Finkle that using catch information from farther flung management districts might add noise to the distributions since those districts would contain harvest from non-Chignik stocks. The analysis was reexamined this time using only harvests that were a day or twos fish-swim away.This change improved the model. The most significant improvement was for 2012 which changed from an overestimate of 67% to a overestimate of 5%. The 2008 and 2015 estimates were similar to those found under the original model with a 23% overestimate and 25% underestimate respectively.


2012
Original Distribution to be Fit
Individual Fitted Distributions
Fitted Distributions Combined
Probability Density
0.000
0.010
0.020
0.030
160
180
200
220
240
260
280
day_of_year
Gamma Mixture



Results

For most years the model using run timing distributions was useful in simulating the early run (as estimated with the use of genetics). Most specifically of the 11 years, 8 were within 5%, one more within 10%, and the remaining two years within 25% of what was estimated with the use of genetics.

Recommendations/Discussion

Future analysis could include examining if there are any factors that explain why 2008 and 2015 performed poorly with the model. Initial investigations note that the mean day-of-year for early run timing is shifted to the right 3 days for 2008 and 2 days for 2015, indicating that the estimated distribution is pulled to the right of what it would be using additional genetic information. Heather Finkle notes that harvest and run size may have been different. These were amoungst some of the smaller early runs examined. Also examining what time and area harvest to include in runtime estimates may improve the model. In terms of this model it is important to note that the definition of the 'run' changed from what is currently used. Harvests farther from the Chignik area were excluded. If this model were used in the future it should be determined if the run estimates for this model are only to distinguish the percentage of early vs late run stocks (recommended for near term), or if the methods behind estimates of harvest should be reexamined, (recommended for the future).

Citation

Elisabeth K. Creelman, Lorenz Hauser, Ryan K. Simmons, William D. Templin & Lisa W. Seeb (2011):Temporal and Geographic Genetic Divergence: Characterizing Sockeye Salmon Populations in the Chignik Watershed, Alaska, Using Single-Nucleotide Polymorphisms, Transactions of the American Fisheries Society, 140:3, 749-762

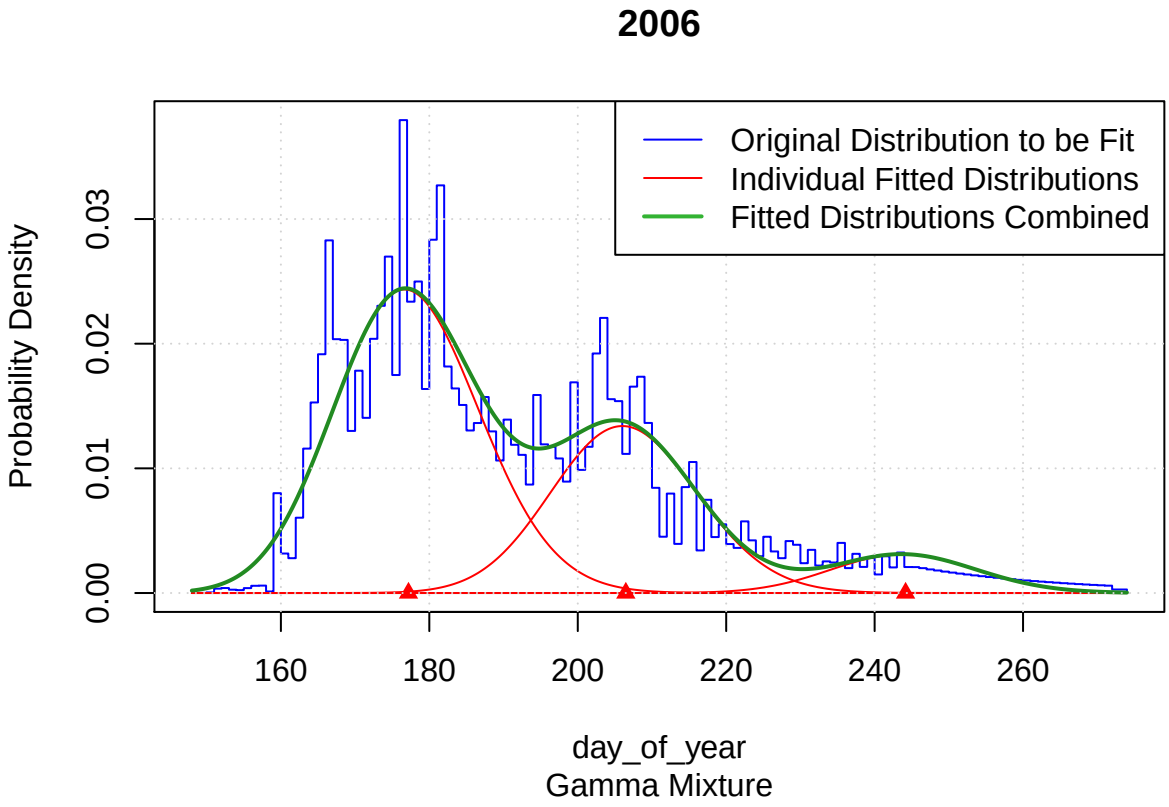
Link to this article: http://dx.doi.org/10.1080/00028487.2011.584494

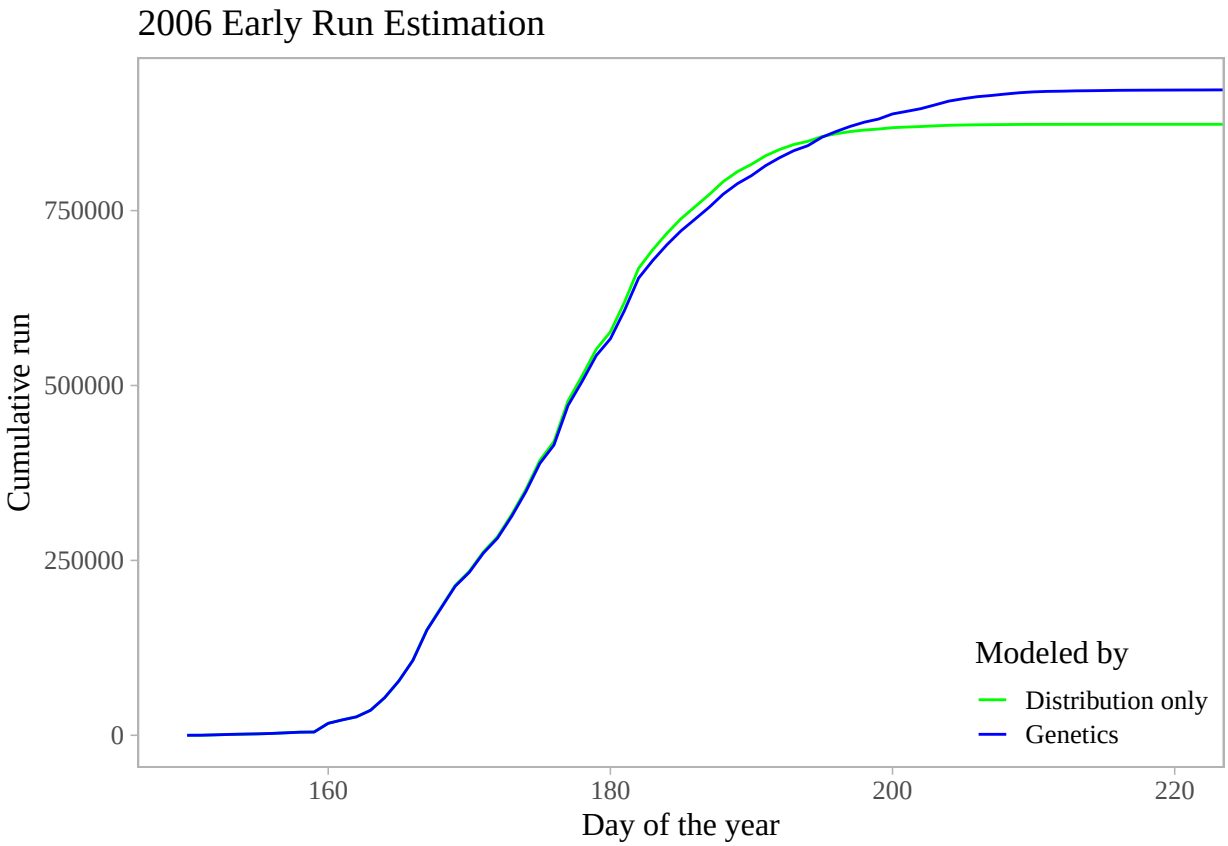
R Session Info

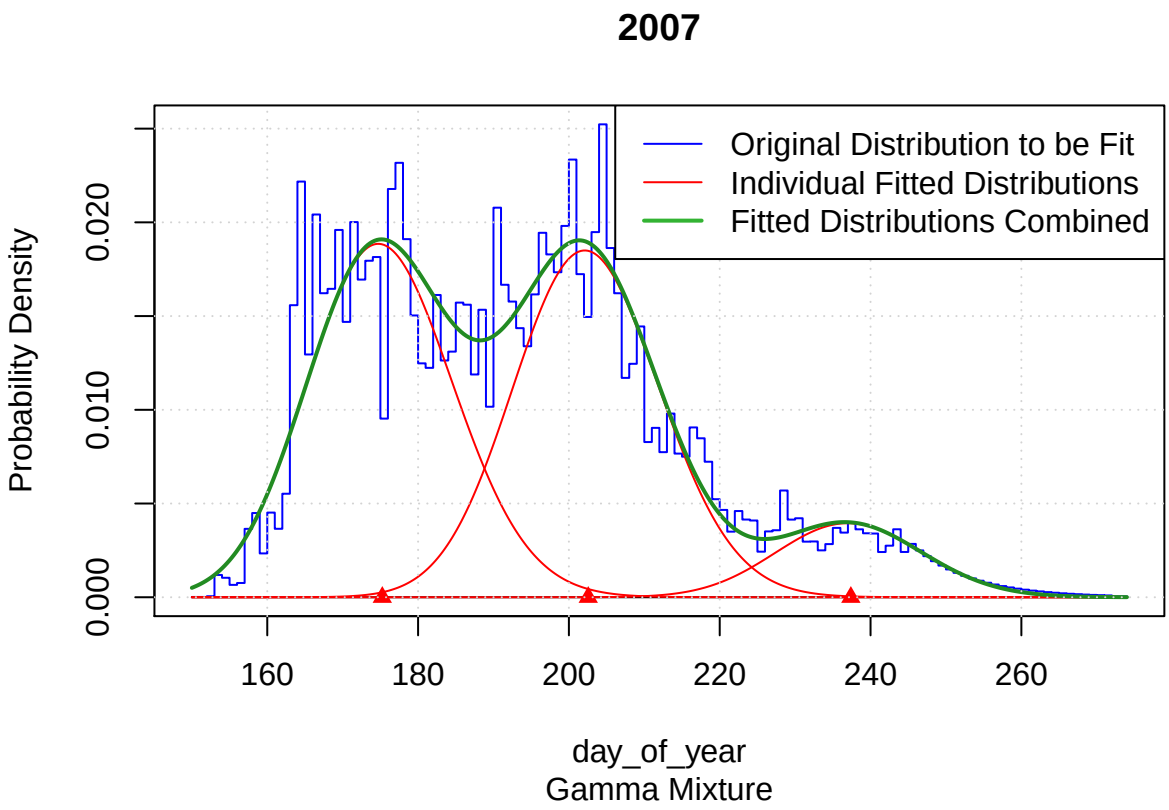
sessionInfo()

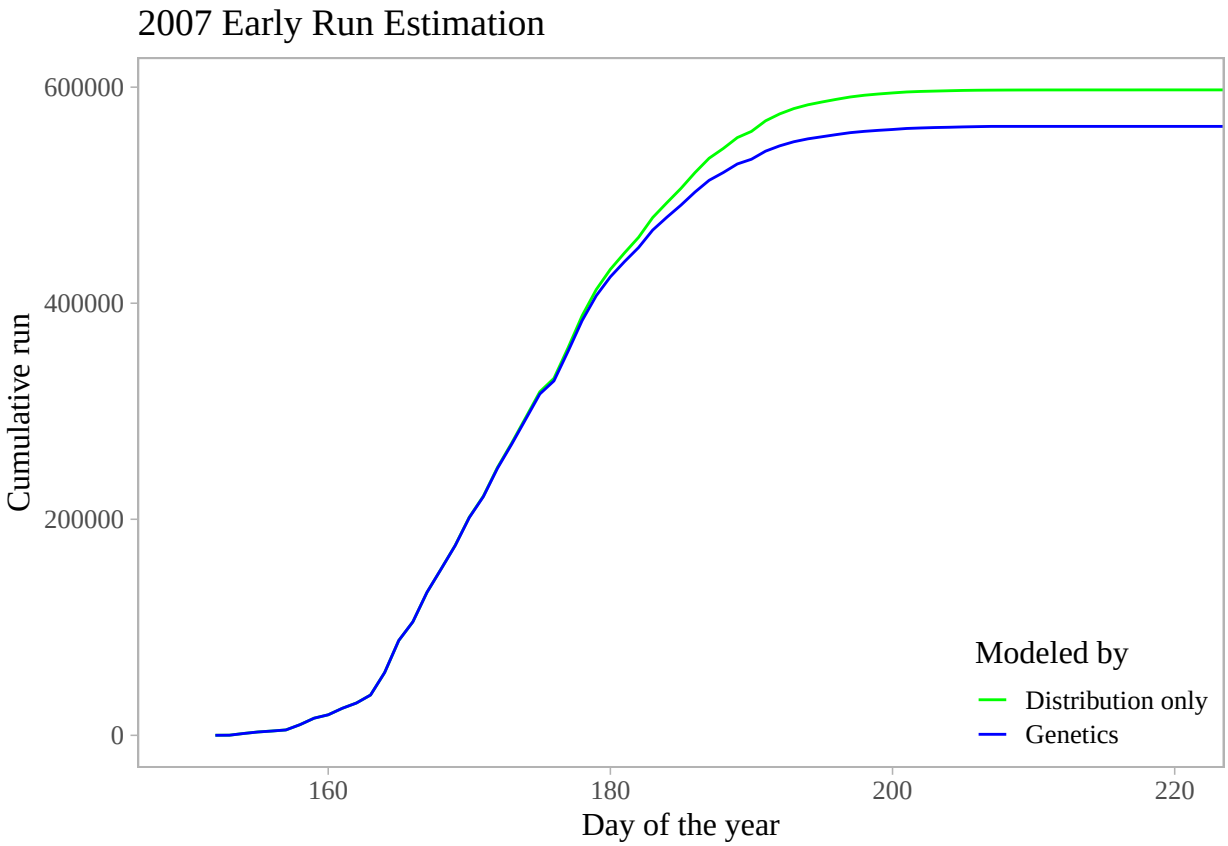
Appendix A

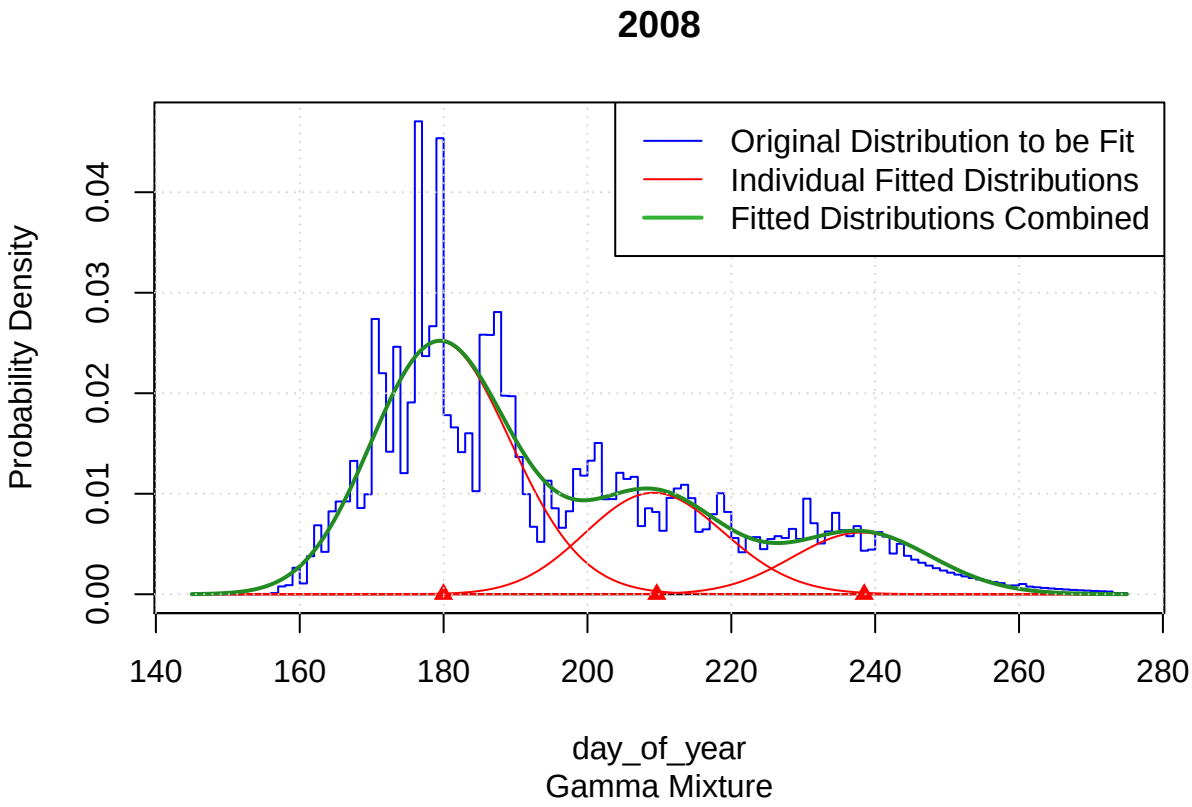
Below are the graphs for all the years using model with the more conservative data set:

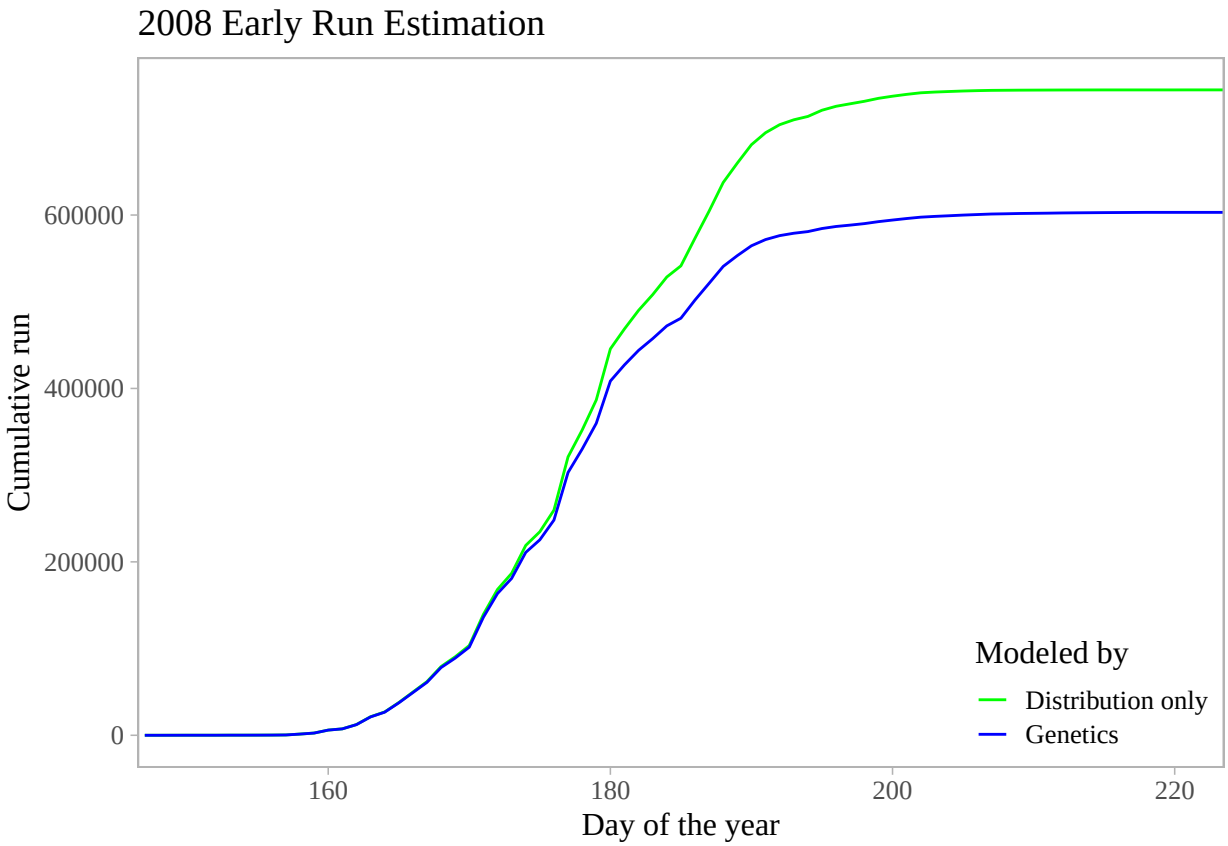


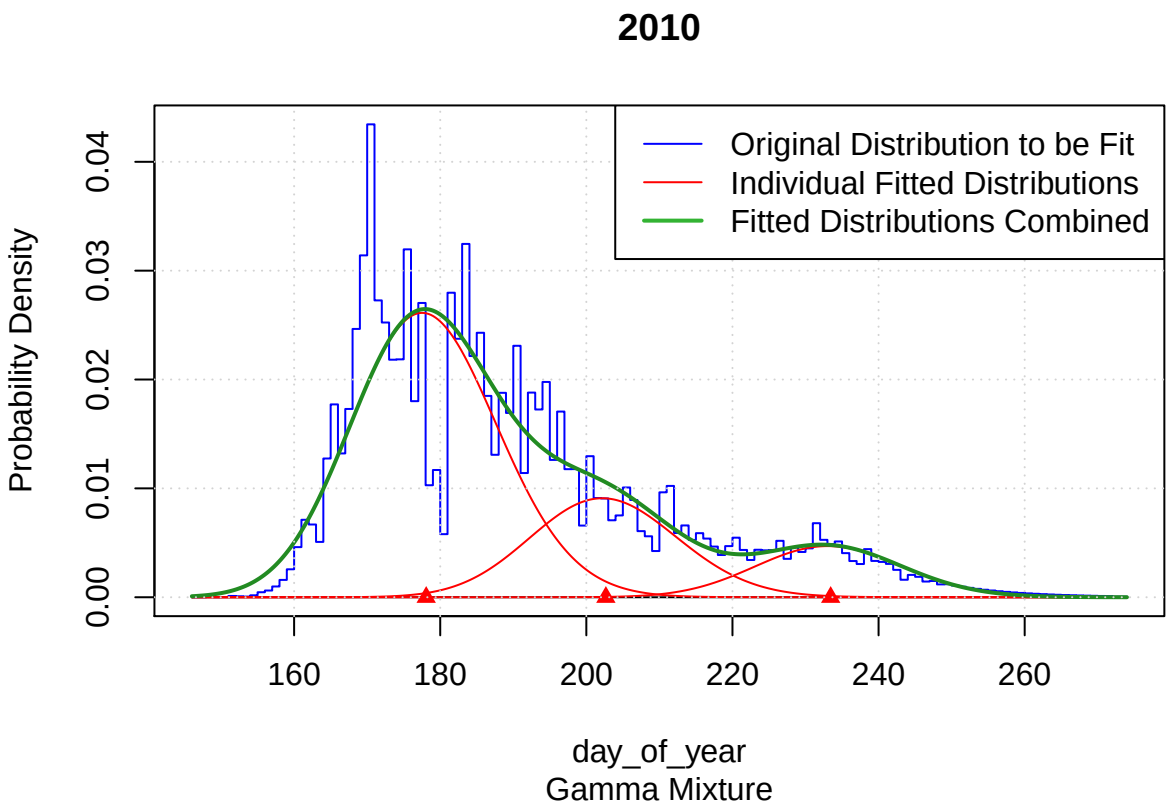

2006 Early Run Estimation
Cumulative run
750000
500000
250000
0
160
180
200
220
Day of the year
Modeled by
Distribution only
Genetics

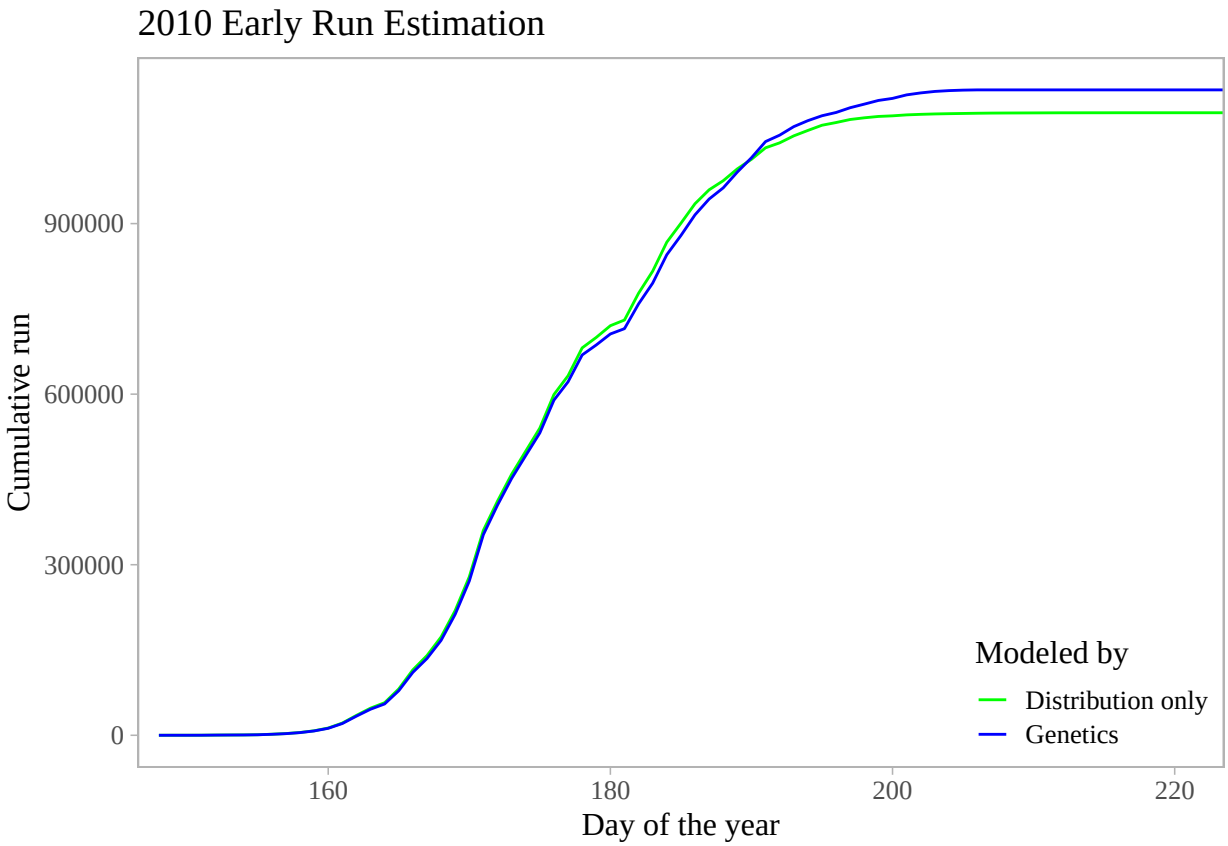

2007
Original Distribution to be Fit
Individual Fitted Distributions
Fitted Distributions Combined
Probability Density
0.000
0.010
0.020
160
180
200
220
240
260
day_of_year
Gamma Mixture

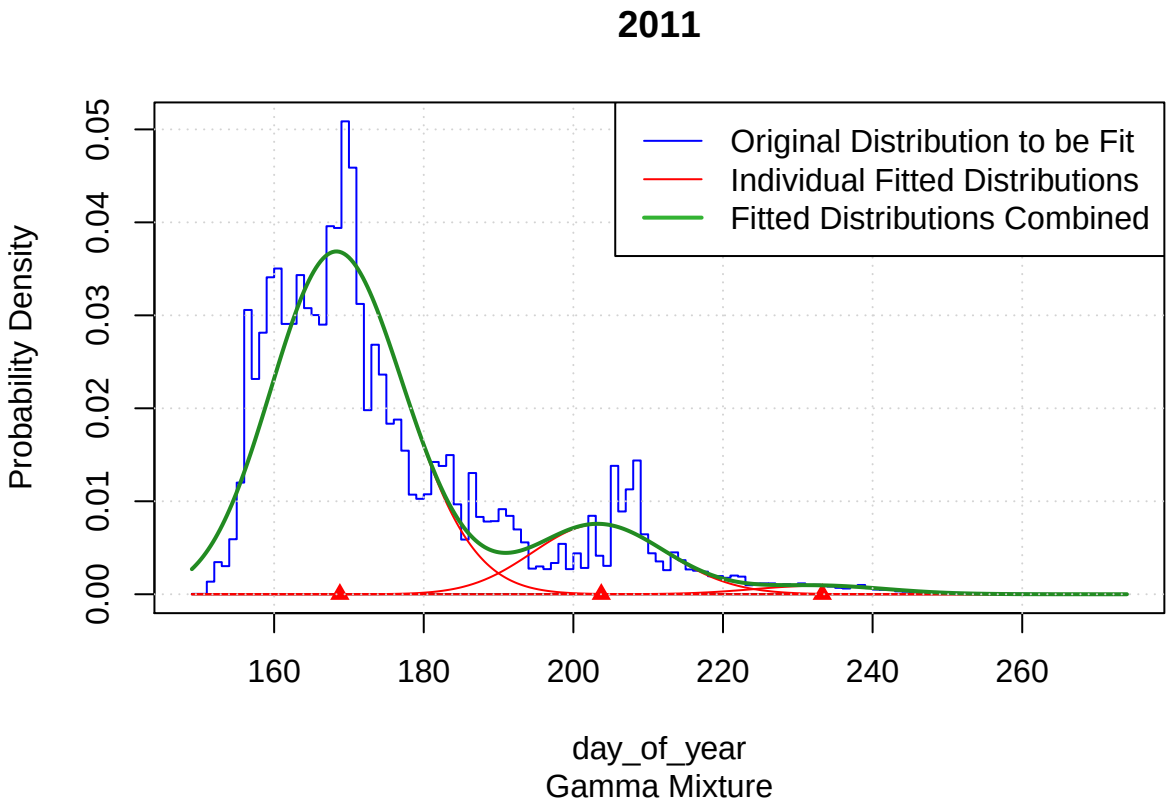

2007 Early Run Estimation
600000
400000
200000
0
Cumulative run
160
180
200
220
Day of the year
Modeled by
Distribution only
Genetics

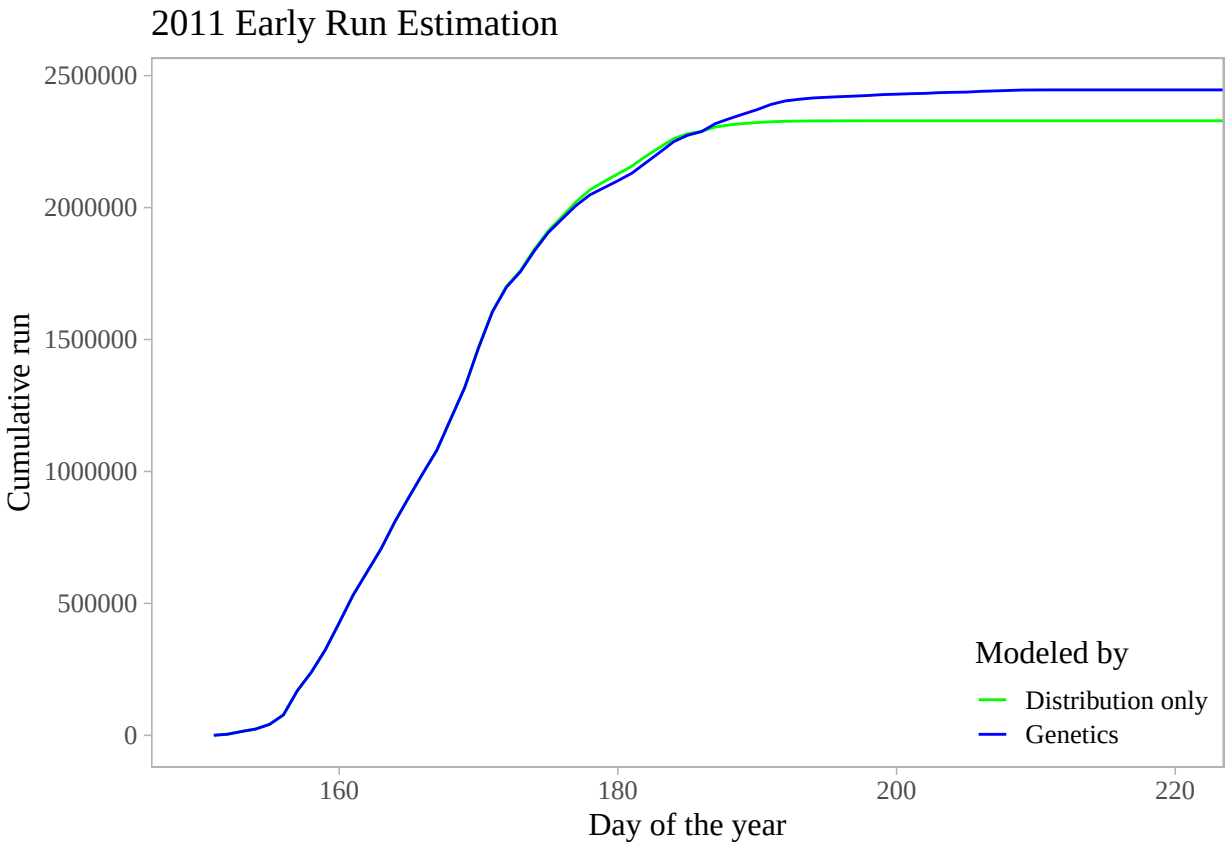

2008
Original Distribution to be Fit
Individual Fitted Distributions
Fitted Distributions Combined
Probability Density
0.00
0.01
0.02
0.03
0.04
140
160
180
200
220
240
260
280
day_of_year
Gamma Mixture

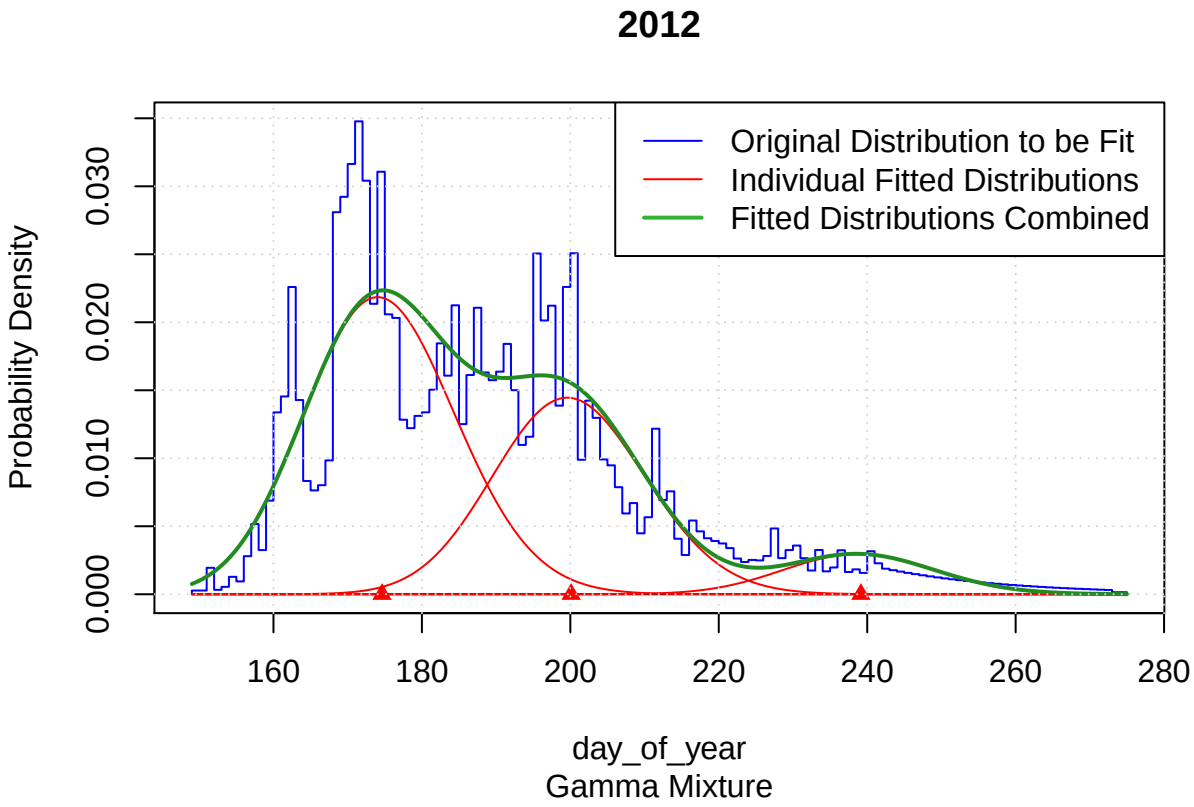

2008 Early Run Estimation
Cumulative run
600000
400000
200000
0
160
180
200
220
Day of the year
Modeled by
Distribution only
Genetics

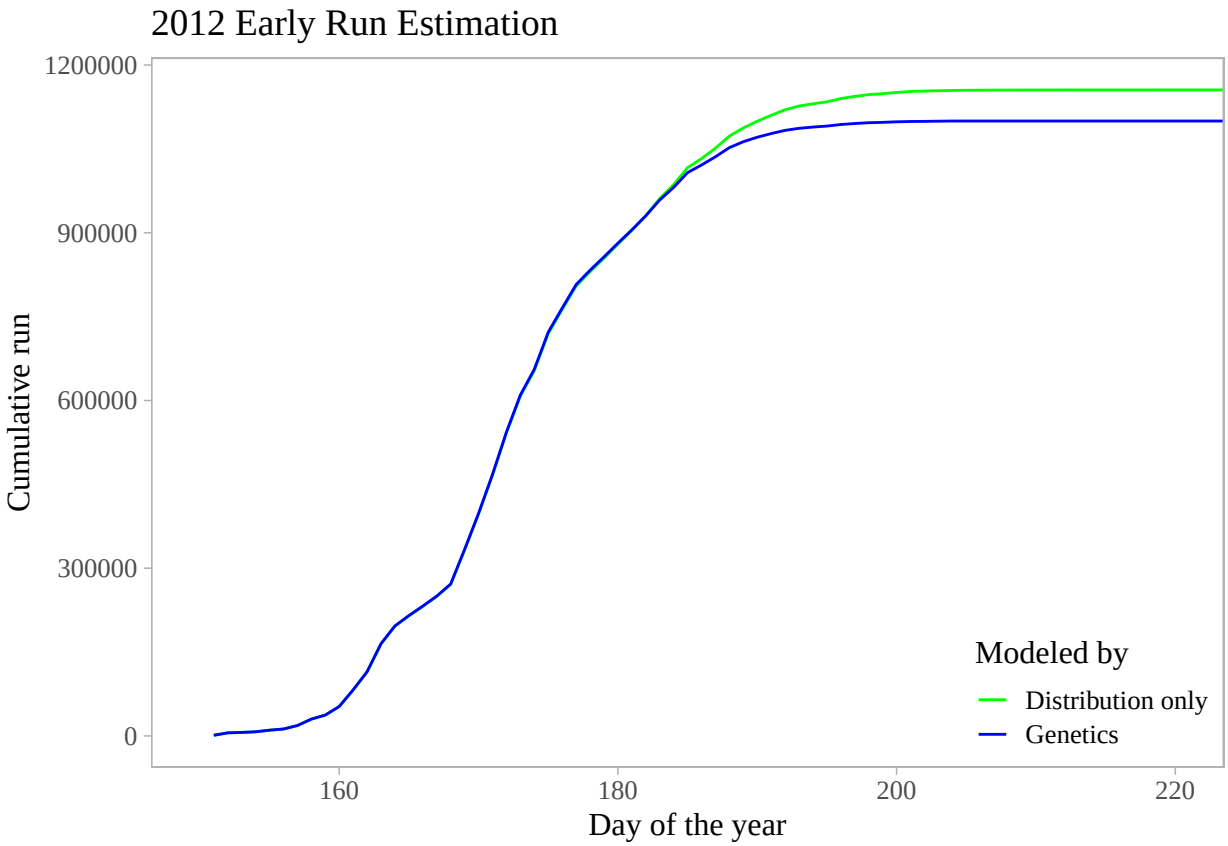

2010
Original Distribution to be Fit
Individual Fitted Distributions
Fitted Distributions Combined
Probability Density
0.00
0.01
0.02
0.03
0.04
160
180
200
220
240
260
day_of_year
Gamma Mixture

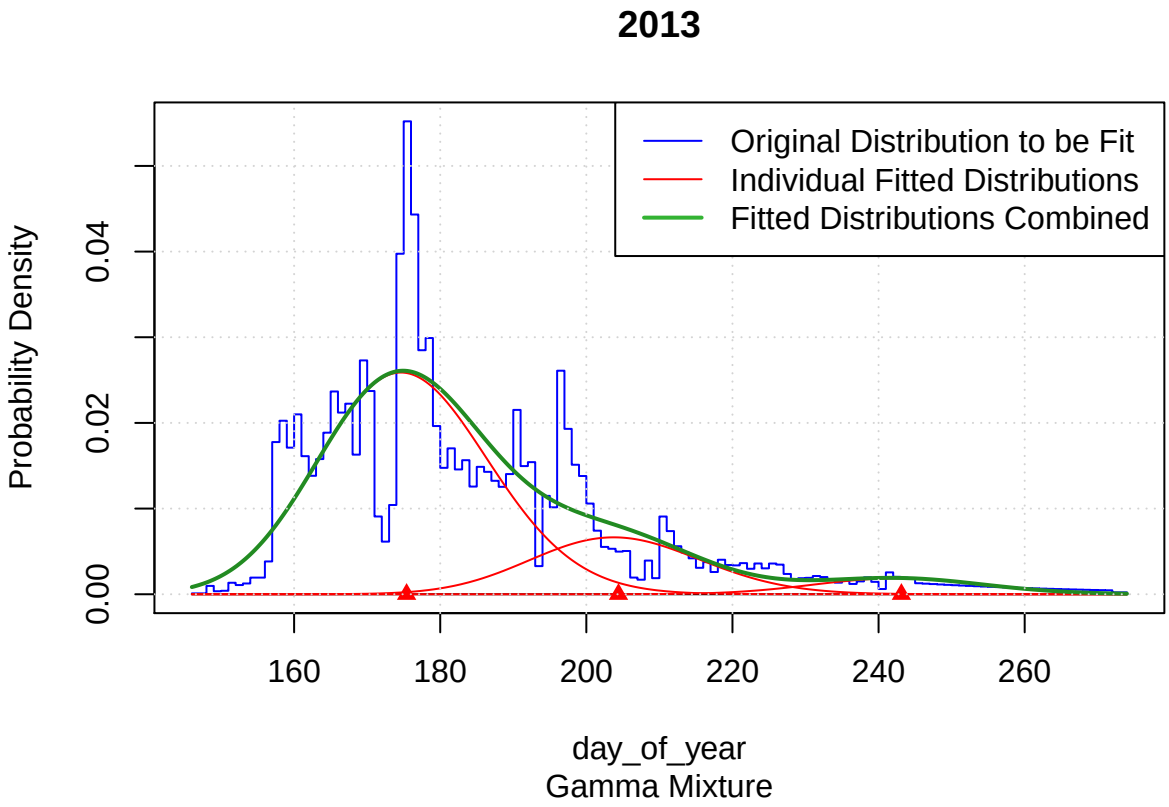

2010 Early Run Estimation
Cumulative run
900000
600000
300000
0
160
180
200
220
Day of the year
Modeled by
Distribution only
Genetics

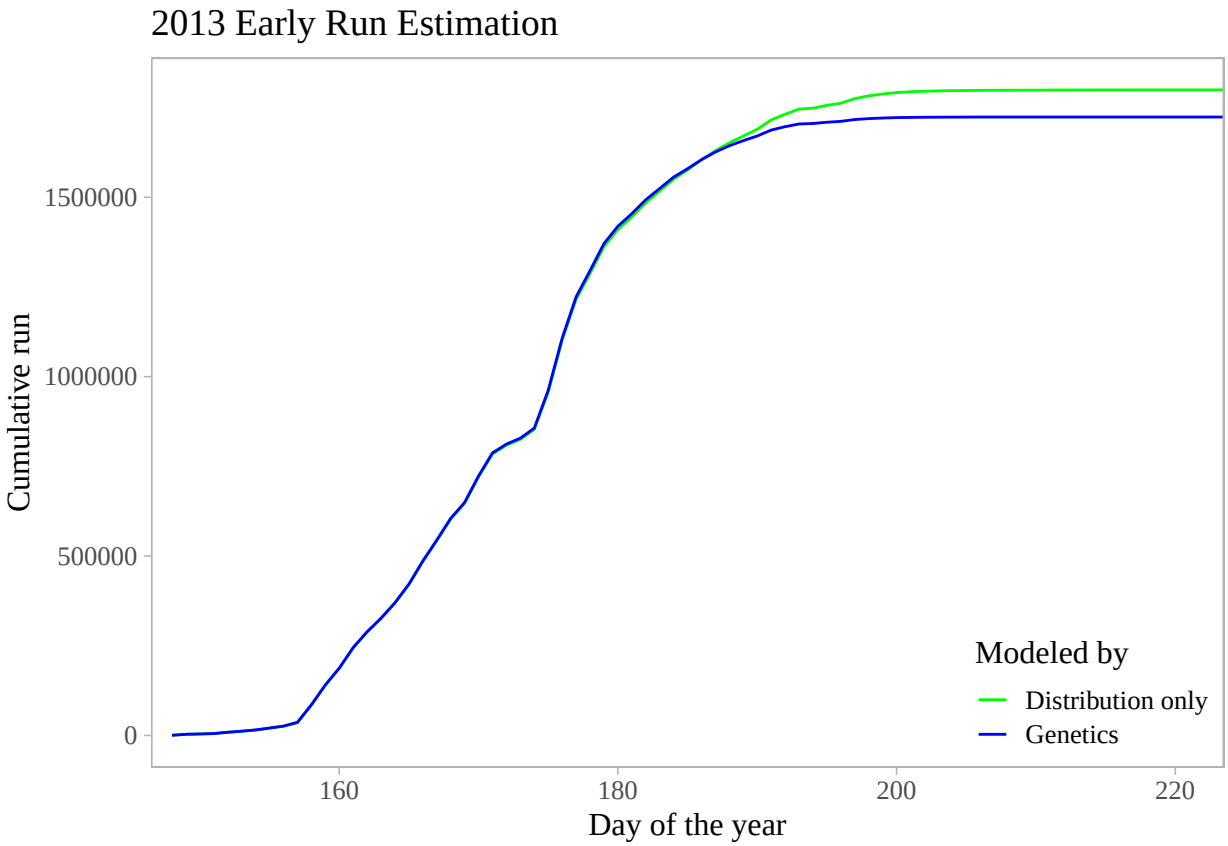

2011
Original Distribution to be Fit
Individual Fitted Distributions
Fitted Distributions Combined
Probability Density
0.00
0.01
0.02
0.03
0.04
0.05
160
180
200
220
240
260
day_of_year
Gamma Mixture

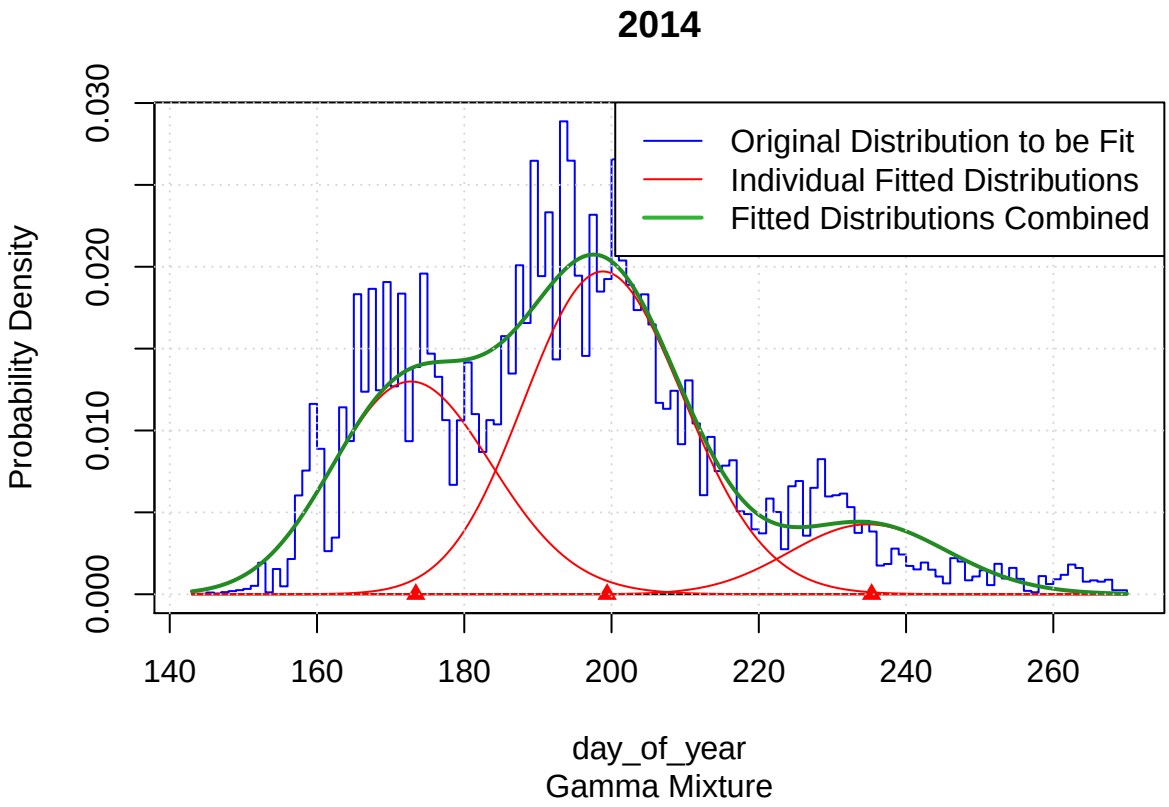

2011 Early Run Estimation
Cumulative run
2500000
2000000
1500000
1000000
500000
0
160
180
200
220
Day of the year
Modeled by
Distribution only
Genetics

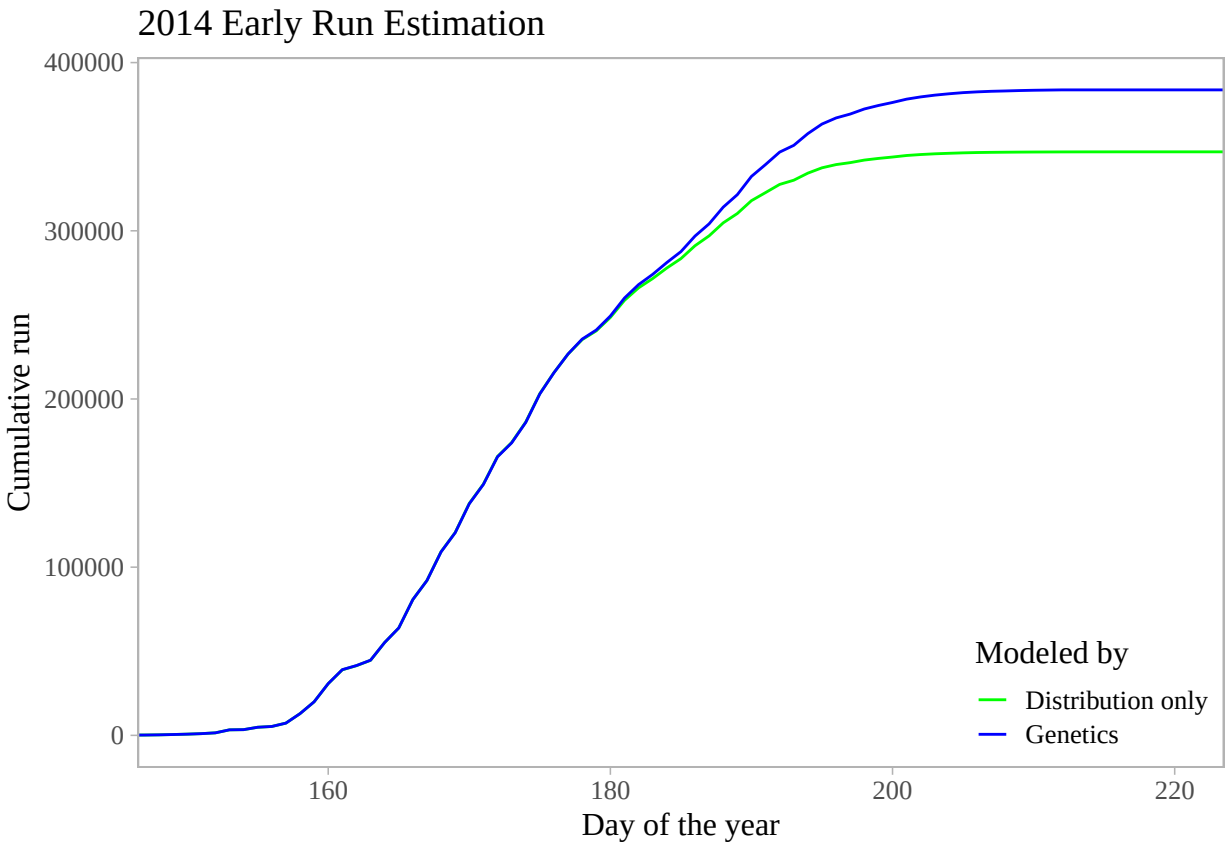

2012
Original Distribution to be Fit
Individual Fitted Distributions
Fitted Distributions Combined
Probability Density
0.000
0.010
0.020
0.030
160
180
200
220
240
260
280
day_of_year
Gamma Mixture

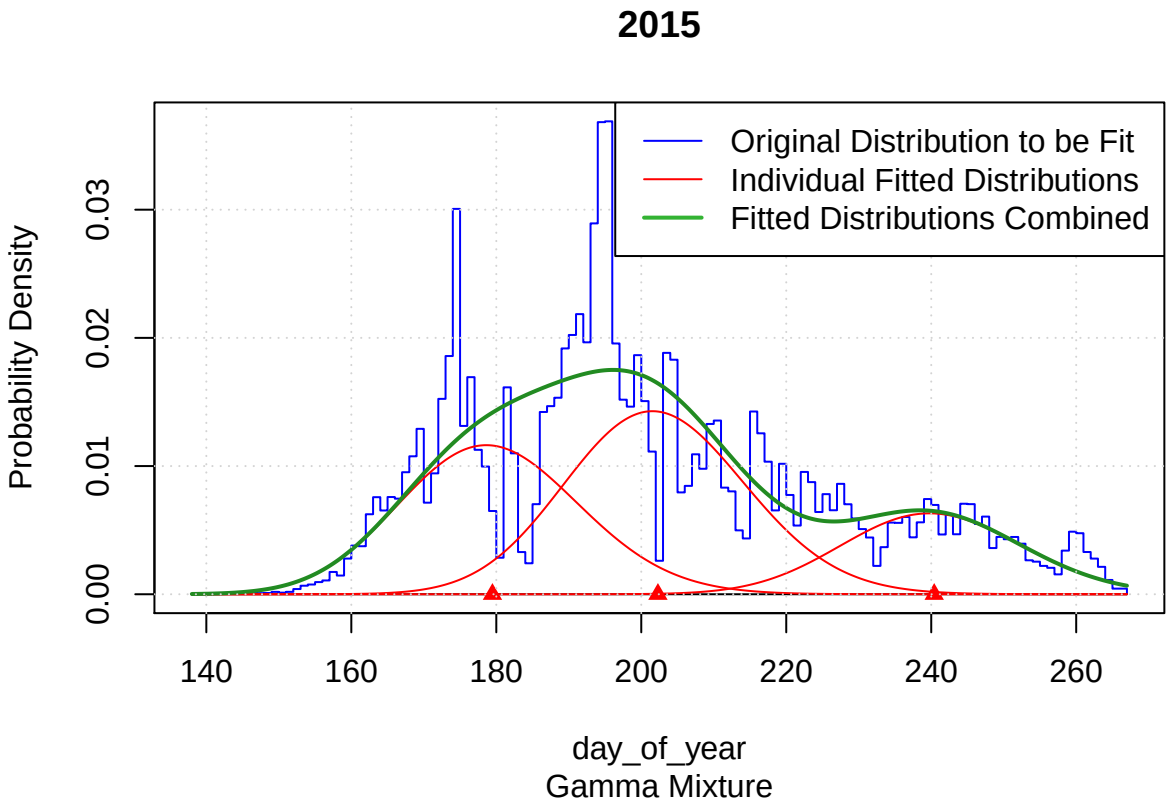

2012 Early Run Estimation
1200000
900000
600000
300000
0
Cumulative run
160
180
200
220
Day of the year
Modeled by
Distribution only
Genetics

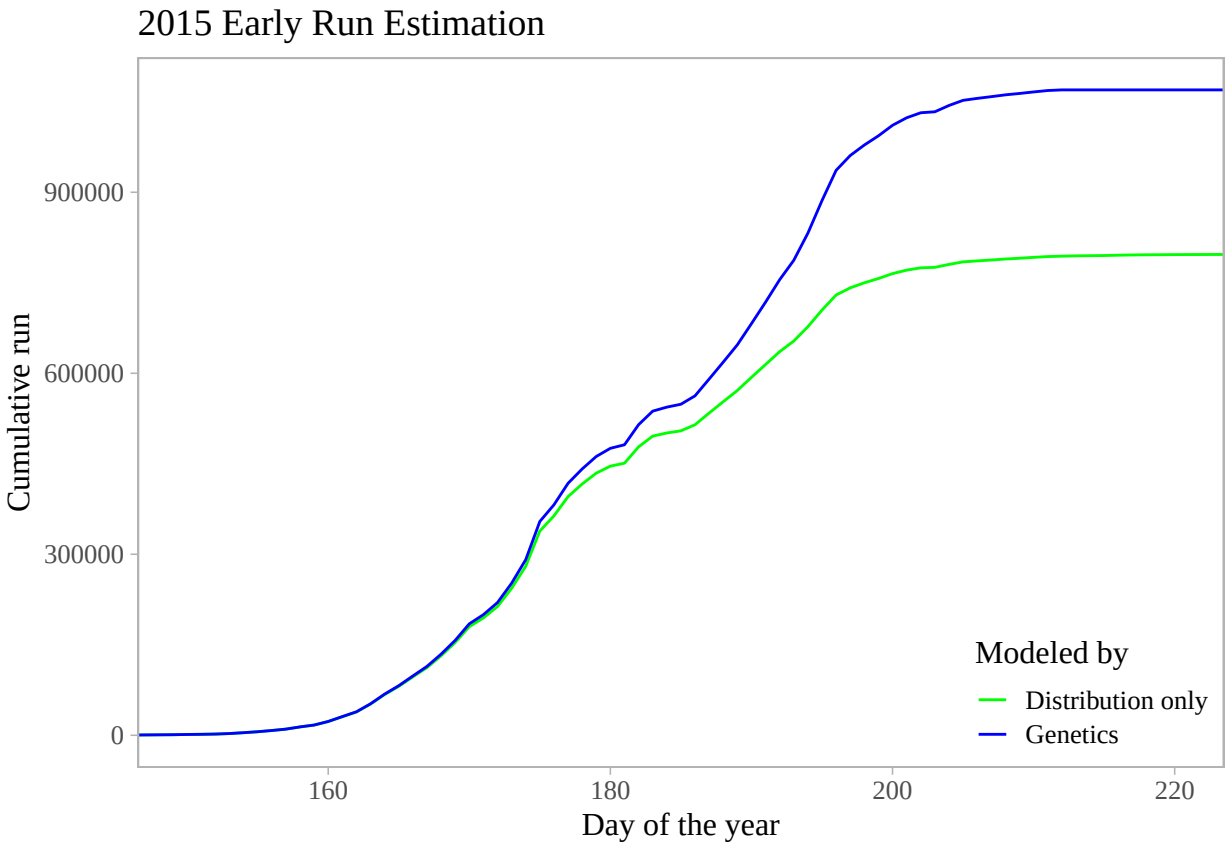

2013
Original Distribution to be Fit
Individual Fitted Distributions
Fitted Distributions Combined
Probability Density
0.00
0.02
0.04
160
180
200
220
240
260
day_of_year
Gamma Mixture

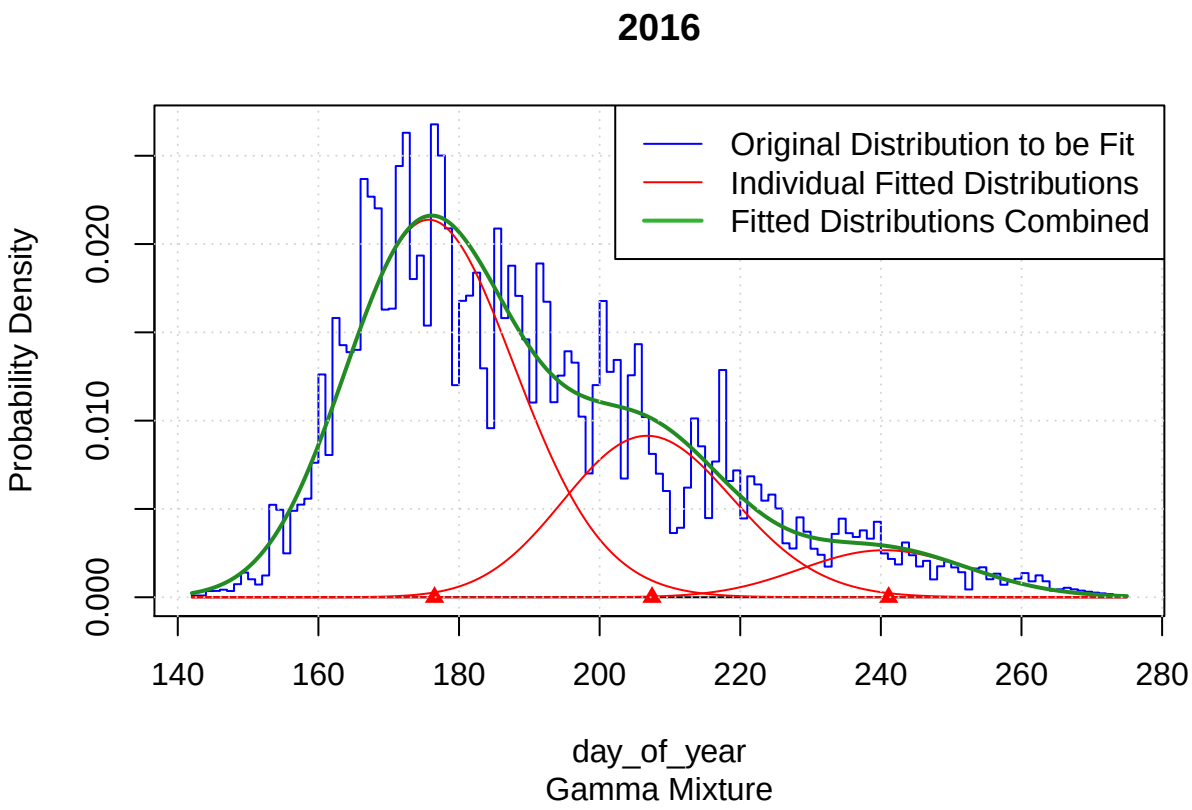

2013 Early Run Estimation
Cumulative run
1500000
1000000
500000
0
160
180
200
220
Day of the year
Modeled by
Distribution only
Genetics

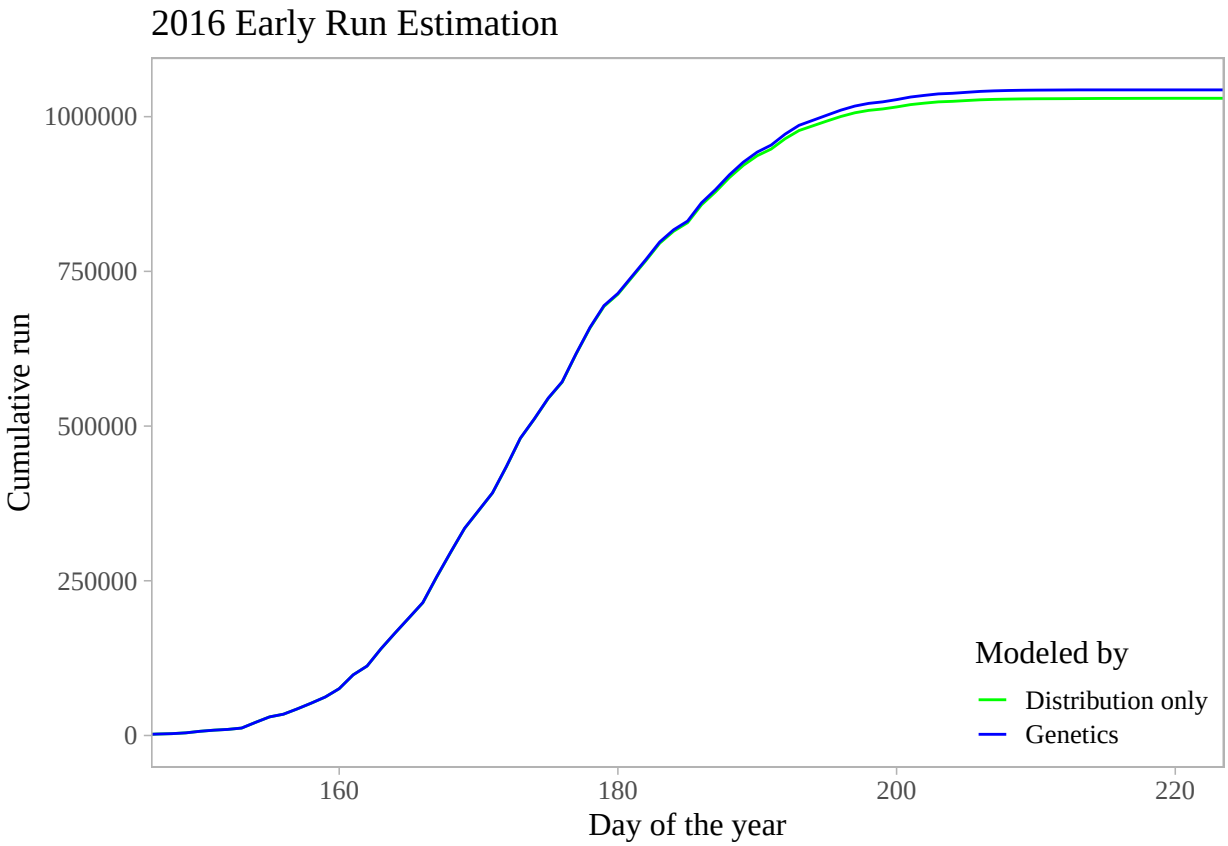

2014
Probability Density
0.000
0.010
0.020
0.030
140
160
180
200
220
240
260
Original Distribution to be Fit
Individual Fitted Distributions
Fitted Distributions Combined
day_of_year
Gamma Mixture

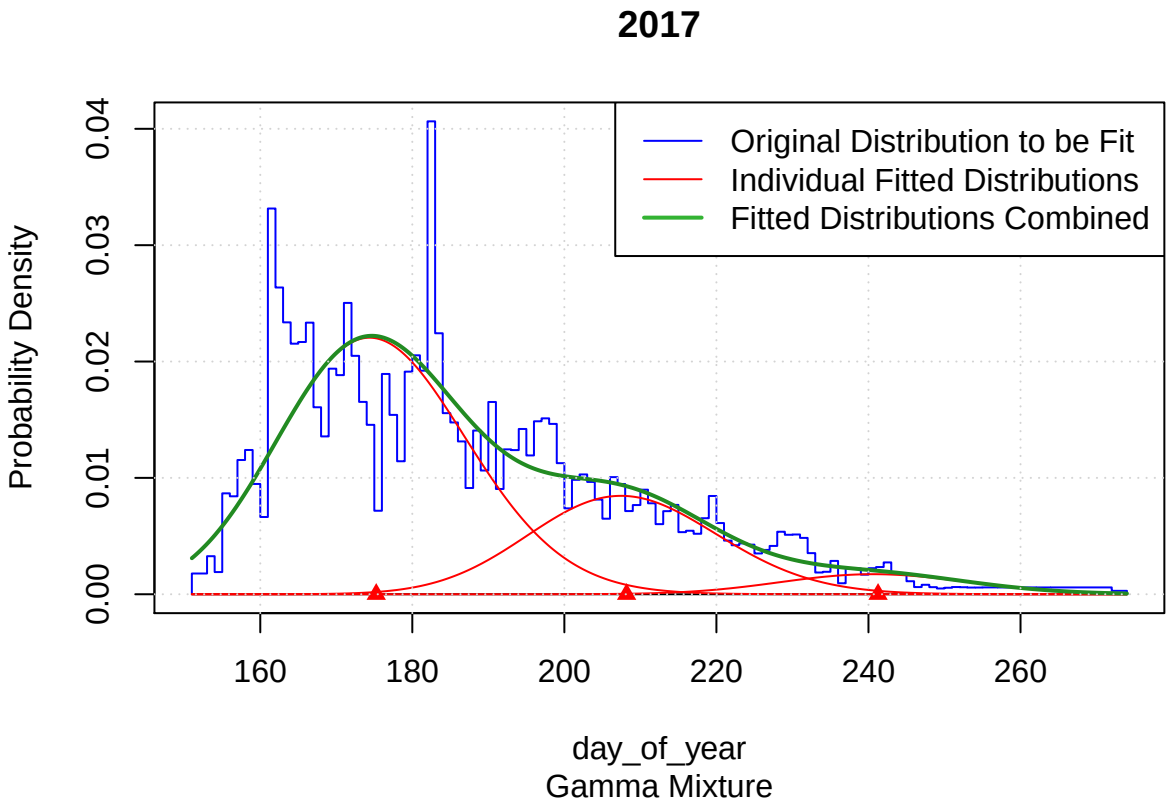

2014 Early Run Estimation
400000
300000
200000
100000
0
Cumulative run
160
180
200
220
Day of the year
Modeled by
Distribution only
Genetics

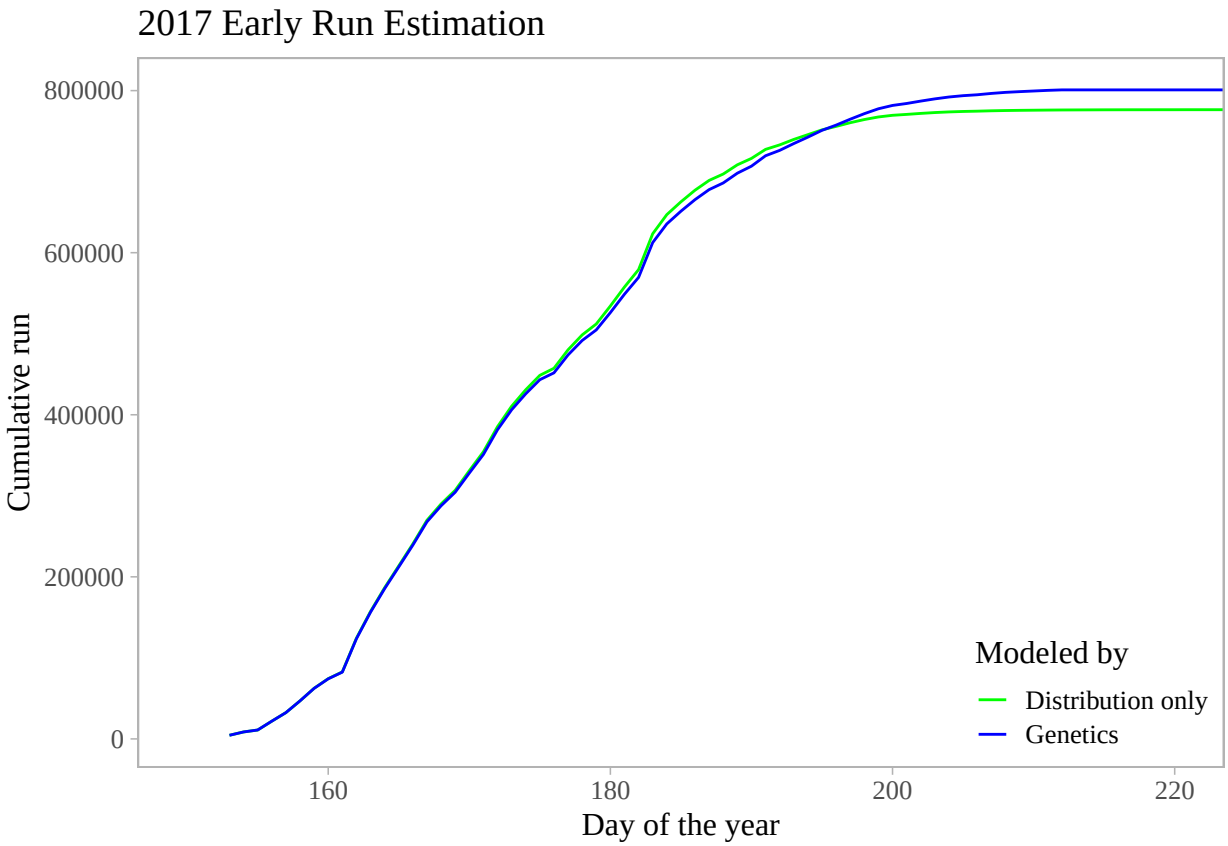

2015
Original Distribution to be Fit
Individual Fitted Distributions
Fitted Distributions Combined
Probability Density
0.00
0.01
0.02
0.03
140
160
180
200
220
240
260
day_of_year
Gamma Mixture


2015 Early Run Estimation
Cumulative run
900000
600000
300000
0
160
180
200
220
Day of the year
Modeled by
Distribution only
Genetics


2016
Original Distribution to be Fit
Individual Fitted Distributions
Fitted Distributions Combined
Probability Density
0.000
0.010
0.020
140
160
180
200
220
240
260
280
day_of_year
Gamma Mixture


2016 Early Run Estimation
Cumulative run
1000000
750000
500000
250000
0
160
180
200
220
Day of the year
Modeled by
Distribution only
Genetics


2017
Original Distribution to be Fit
Individual Fitted Distributions
Fitted Distributions Combined
Probability Density
0.00
0.01
0.02
0.03
0.04
160
180
200
220
240
260
day_of_year
Gamma Mixture


2017 Early Run Estimation
800000
600000
400000
200000
0
Cumulative run
160
180
200
220
Day of the year
Modeled by
Distribution only
Genetics