

# Chignik Sockeye Early Run Modeling

*Sarah Power*

*2018-12-19*

## 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 run timing distributions to determine how well early run stocks can be discerned from later stocks on run timing alone. The results will be compared to the run timing distributions that are determined with the additional aid of genetics to examine the viability of this method.

Currently the run consists of the escapement as captured by the weir, and catches from various management districts, some of which are distant from Chignik river.

## 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. An initial review of run timing across years indicates that there are typically three district run timing modes, not four. (Appendix A, original distributions to be fit.) 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.

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

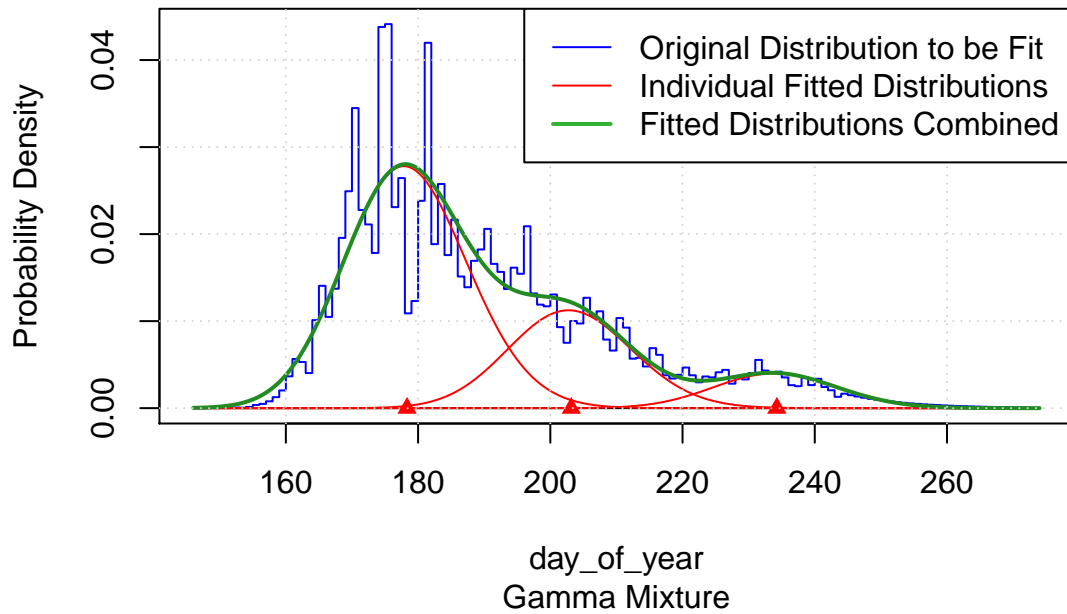
Distributions of the early run were estimated using run timing and genetic assignment for each year. This run timing estimate was used as the baseline for comparison of the models evaluated sans genetics.

In initial investigations Normal and Weibull distributions were fit to the multimodal run timing data, additionally 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.

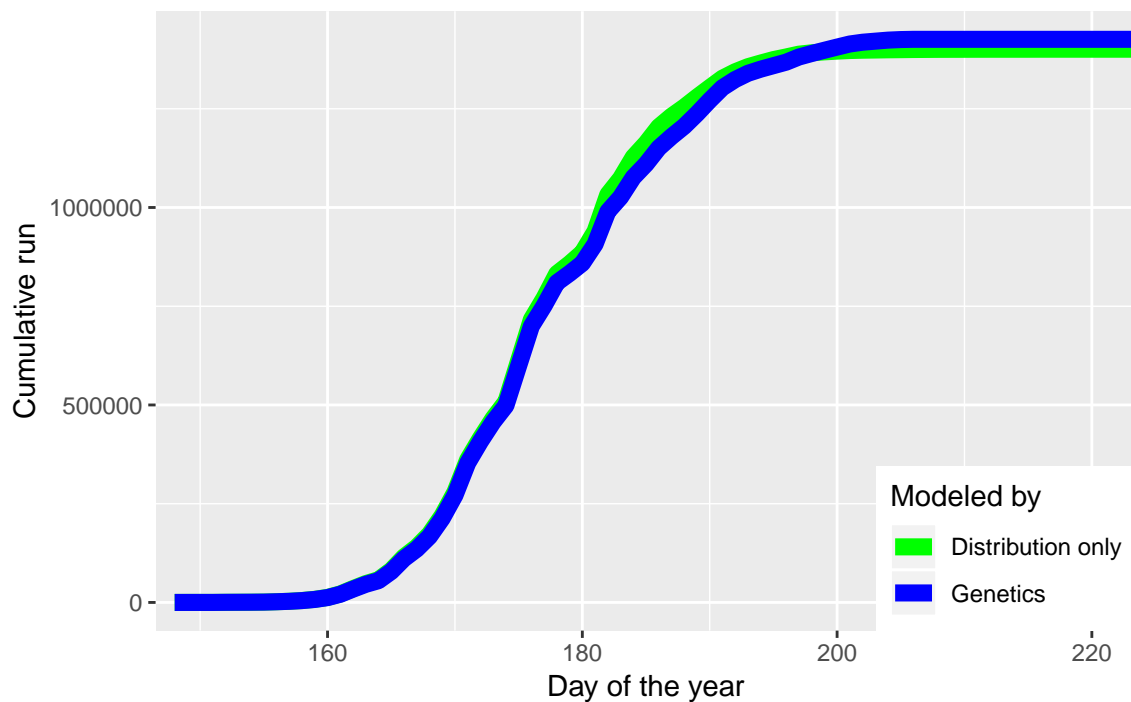
The most appropriate model was a trimodal, normally distributed model with the constraint of equal standard deviations for each of the distributions. This model had an early run timing with means and standard deviations most like the estimated early run timing distributions determined with genetics.

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)

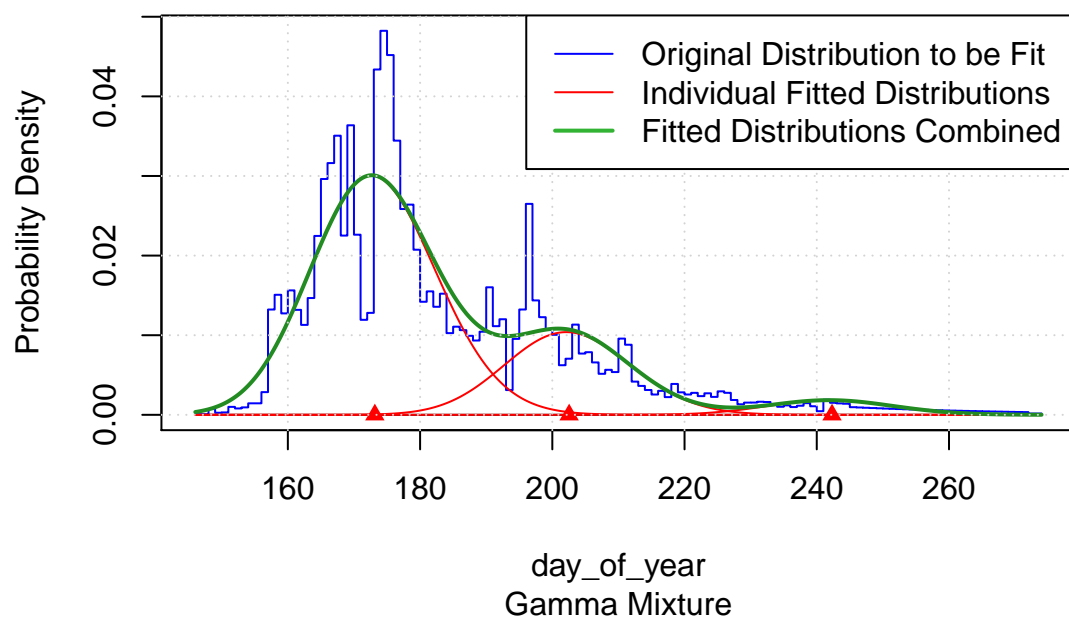
## 2010



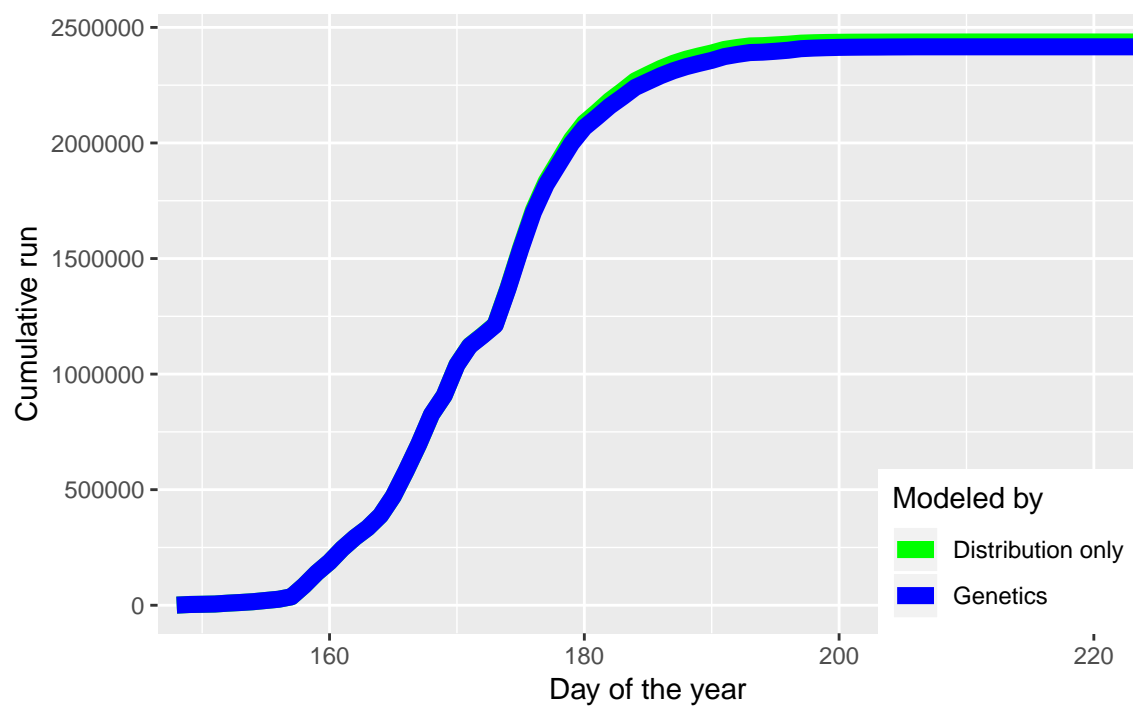
## 2010 Early Run Estimation



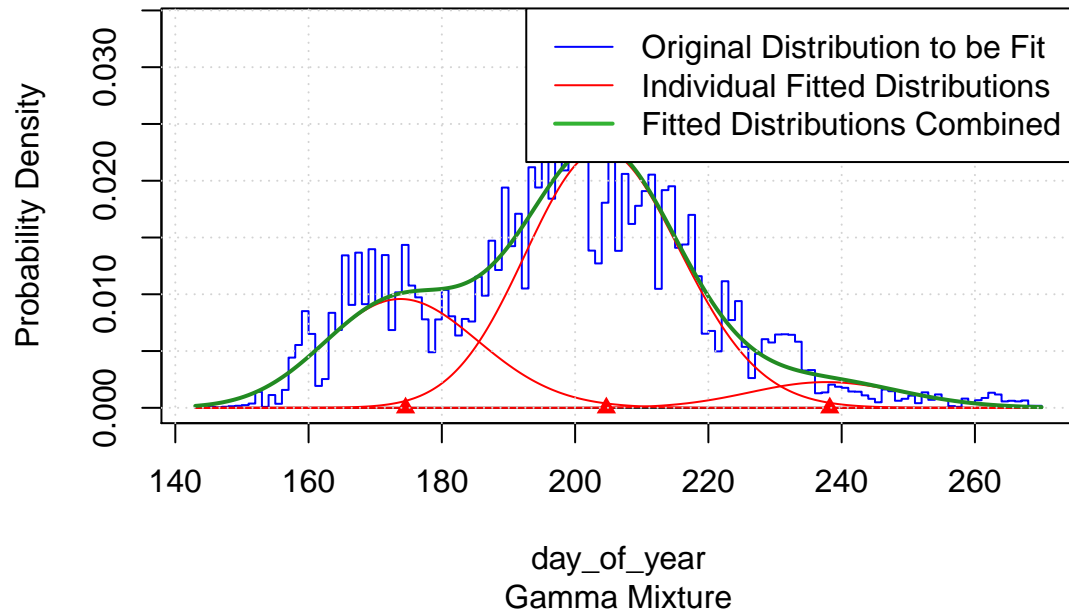
**2013**



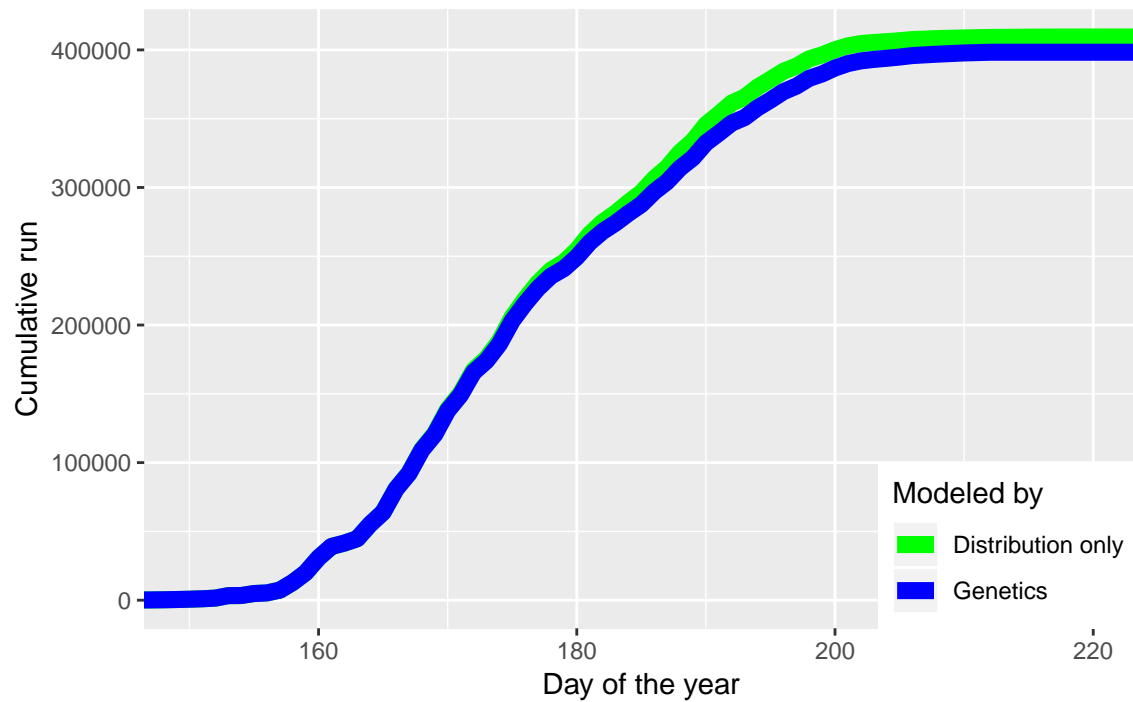
**2013 Early Run Estimation**



2014

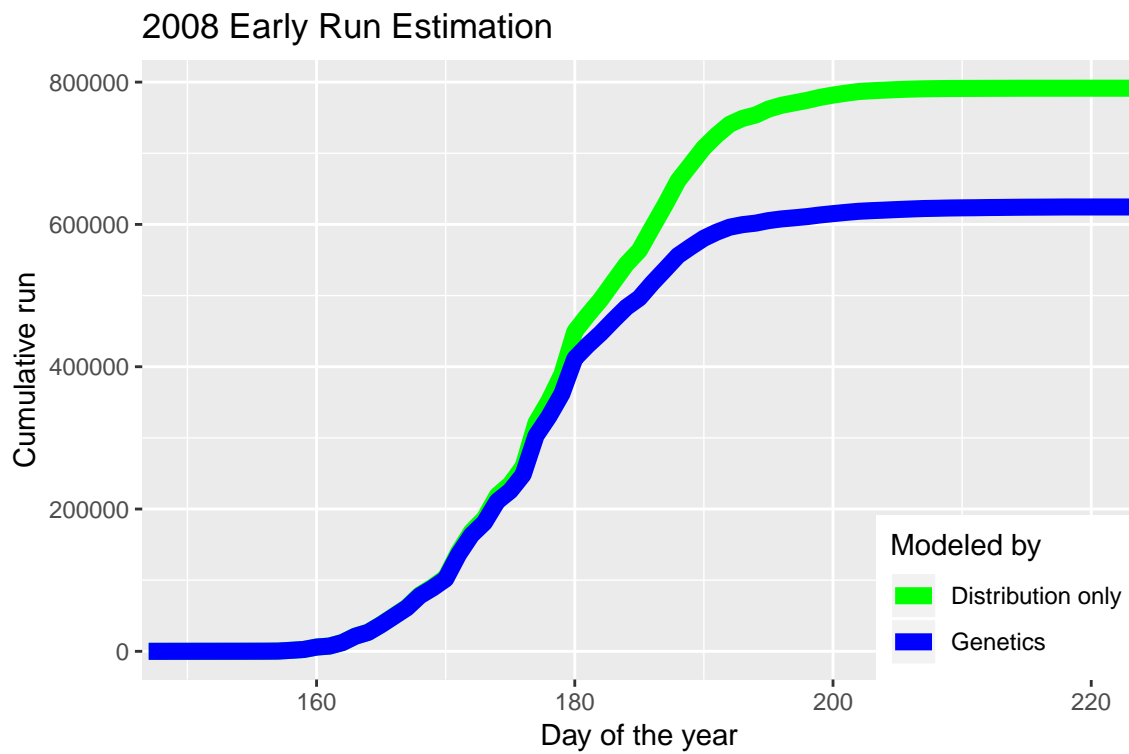
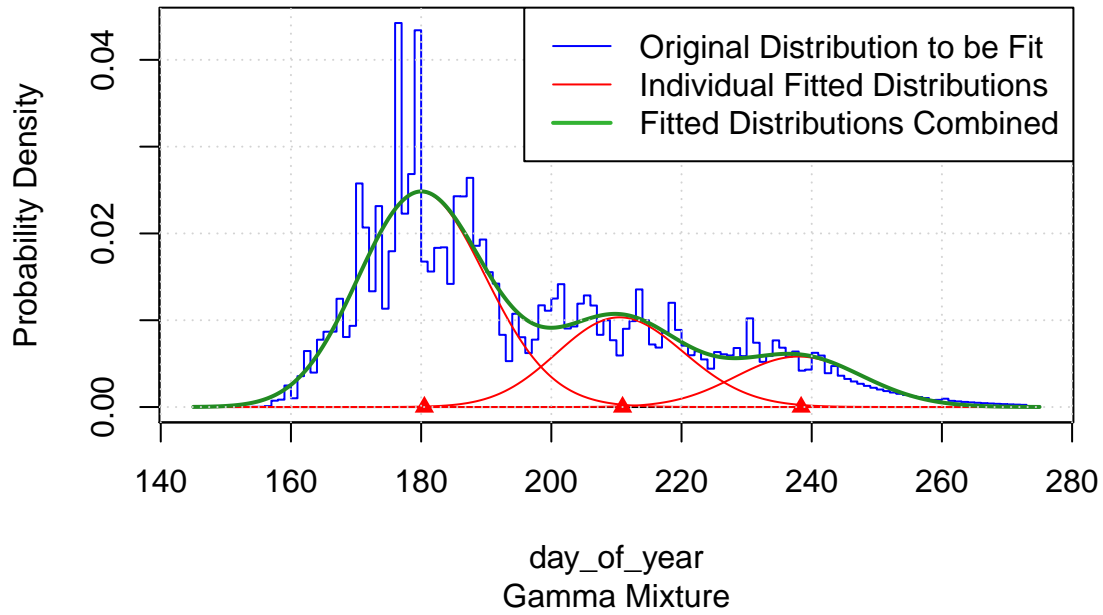


2014 Early Run Estimation

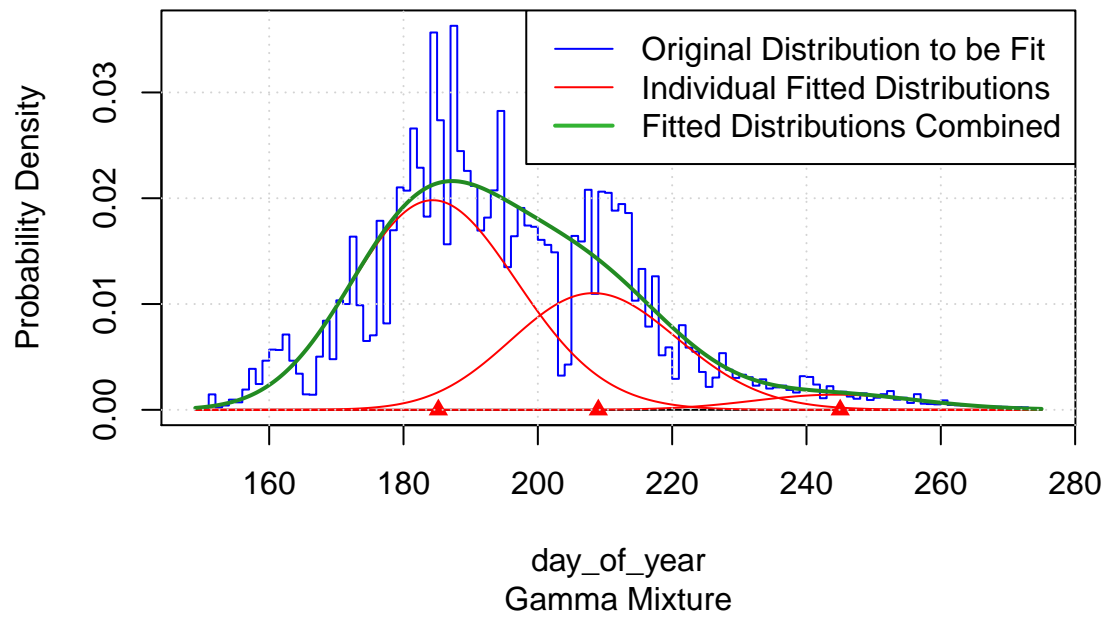


The three years that run timing distributions were off by more than 10% included 2008, which overestimated the early run by 27%, 2012 which overestimated by 67% and 2015 which underestimated by 25% and 2012.

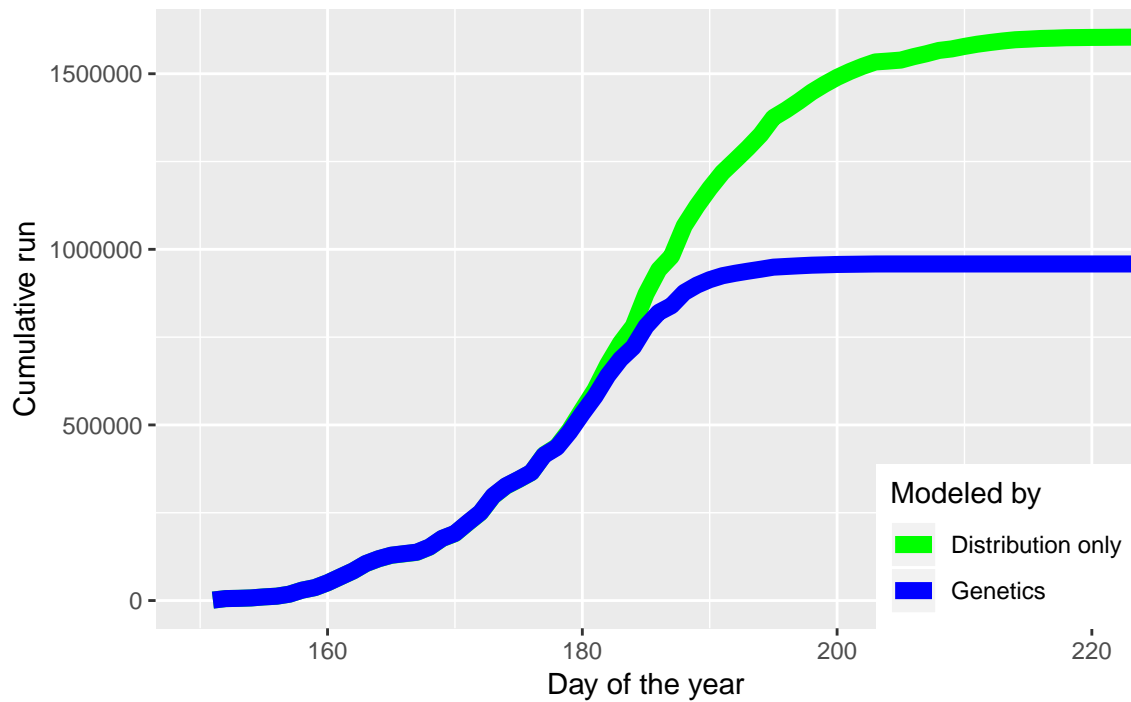
## 2008



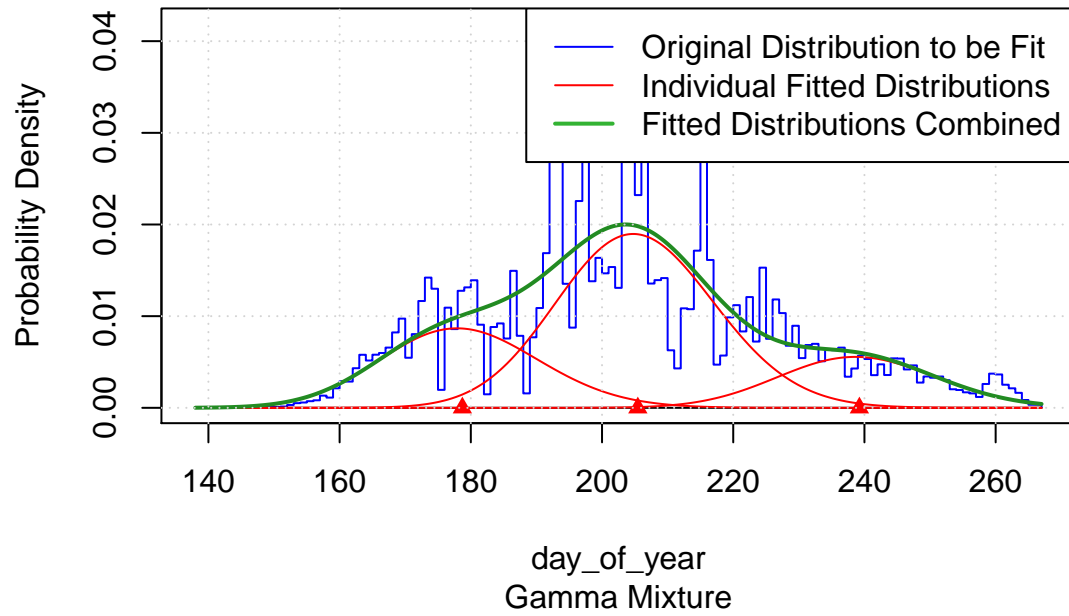
**2012**



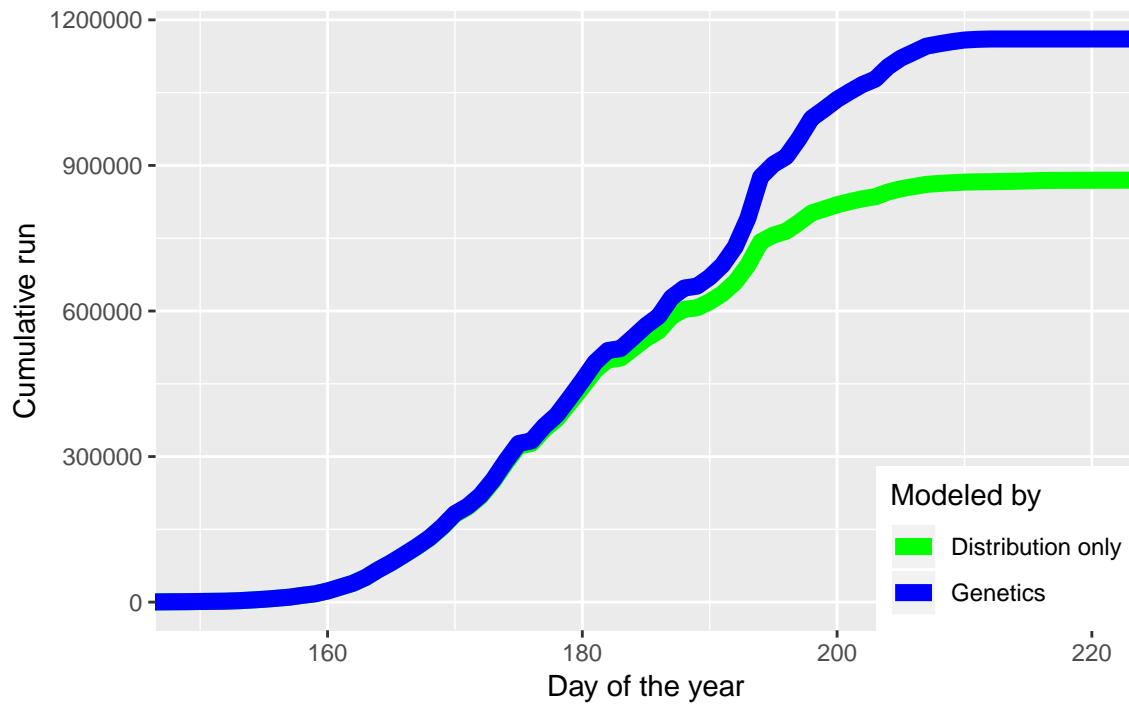
**2012 Early Run Estimation**



2015



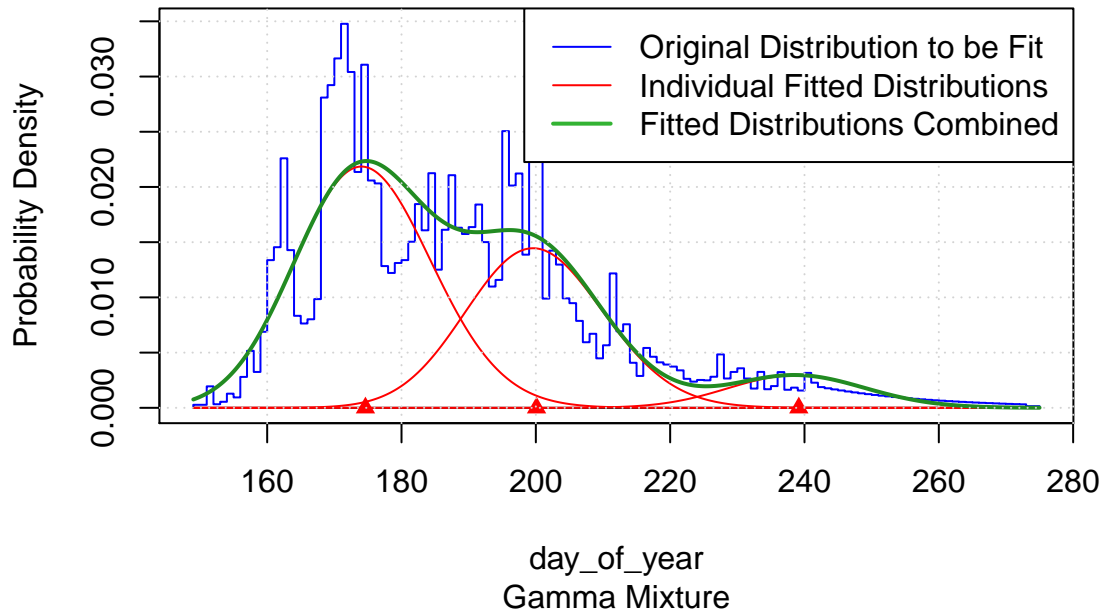
2015 Early Run Estimation



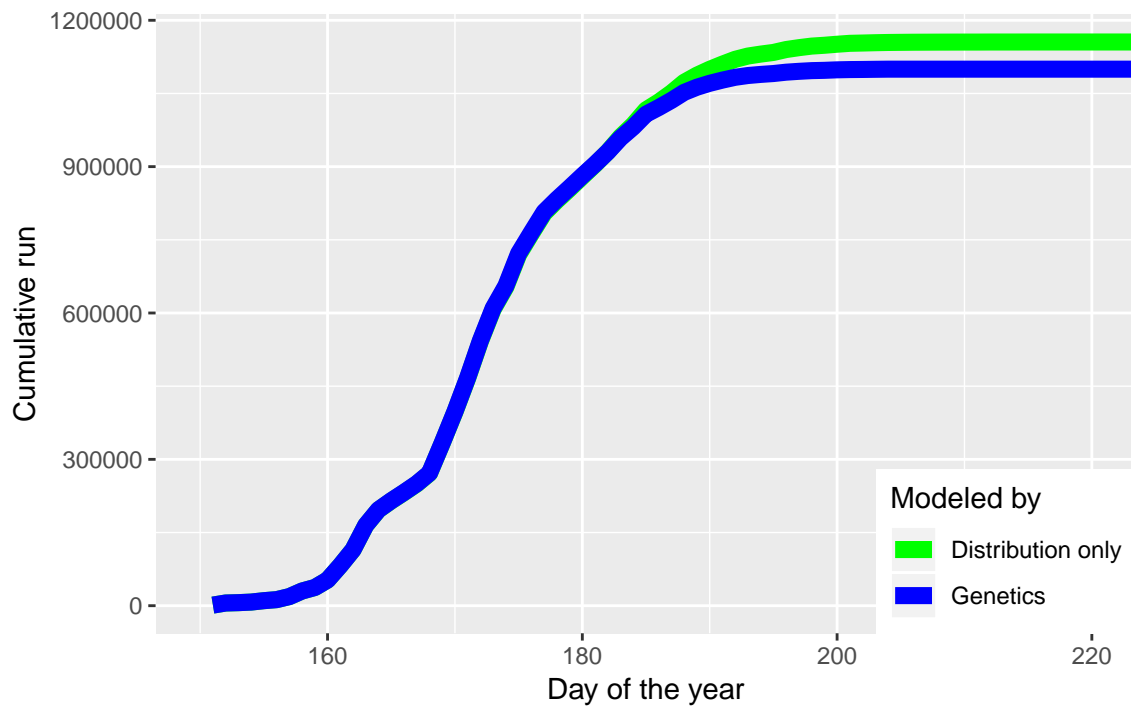
It was suggested by Heather Finkle that using catch information from farther flung management districts might reduce the ability to accurately model run timing since those districts would contain harvest from non-Chignik stocks. The analysis was reexamined this time using only harvests that were within a two day fish travel time from Chignik weir. This change improved the model significantly for 2012. The run timing only model previously over estimated the genetics model by 67% after the management area correction this dropped to a 5% over estimate. The 2008 and 2015 estimates were similar to those found under the original model with a 23% overestimate and 25% underestimate respectively.



2012



2012 Early Run Estimation



## 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 for 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, relative to the run-timing distribution established using genetic information. Heather Finkle noted that harvest and run size may have been different. These were some of the smaller early runs examined. Examining which harvest to include temporally and spatially in runtime estimates may improve the model. Should this approach be accepted for run timing determination it would also be necessary to determine whether estimates would be used only to distinguish the percentage of early vs late run stocks, or if the methods for determining harvest need be re-examined. This is an important point since the model provided excludes harvest from some management areas and therefore has a different definition of what is a run than has previously been used.

## 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>

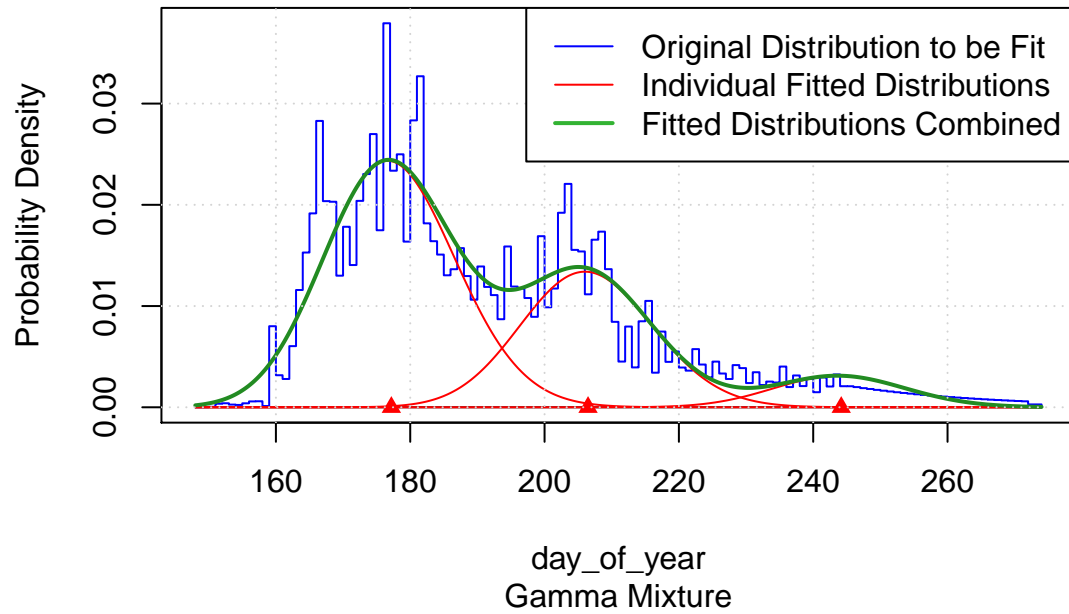
## R Session Info

```
## R version 3.5.1 (2018-07-02)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 7 x64 (build 7601) Service Pack 1
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United States.1252
## [2] LC_CTYPE=English_United States.1252
## [3] LC_MONETARY=English_United States.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.1252
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods    base
##
## other attached packages:
## [1] bindrcpp_0.2.2   gridExtra_2.3    lubridate_1.7.4
## [4] data.table_1.11.6 mixdist_0.5-5     forcats_0.3.0
## [7] stringr_1.3.1    dplyr_0.7.8      purrr_0.2.5
## [10] readr_1.2.1      tidyr_0.8.2      tibble_1.4.2
## [13] ggplot2_3.1.0    tidyverse_1.2.1
##
## loaded via a namespace (and not attached):
## [1] tidyselect_0.2.5 reshape2_1.4.3  haven_1.1.2      lattice_0.20-35
## [5] colorspace_1.3-2 htmltools_0.3.6  yaml_2.2.0       rlang_0.3.0.1
## [9] pillar_1.3.0     glue_1.3.0       withr_2.1.2      modelr_0.1.2
## [13] readxl_1.1.0     bindr_0.1.1      plyr_1.8.4       munsell_0.5.0
## [17] gtable_0.2.0     cellranger_1.1.0 rvest_0.3.2      evaluate_0.11
## [21] labeling_0.3     knitr_1.20       broom_0.5.0      Rcpp_1.0.0
## [25] scales_0.5.0     backports_1.1.2  jsonlite_1.5     hms_0.4.2
## [29] digest_0.6.15    stringi_1.1.7    grid_3.5.1       rprojroot_1.3-2
## [33] cli_1.0.0        tools_3.5.1      magrittr_1.5     lazyeval_0.2.1
## [37] crayon_1.3.4     pkgconfig_2.0.1  xml2_1.2.0       assertthat_0.2.0
## [41] rmarkdown_1.10   httr_1.3.1       rstudioapi_0.7   R6_2.2.2
## [45] nlme_3.1-137     compiler_3.5.1
```

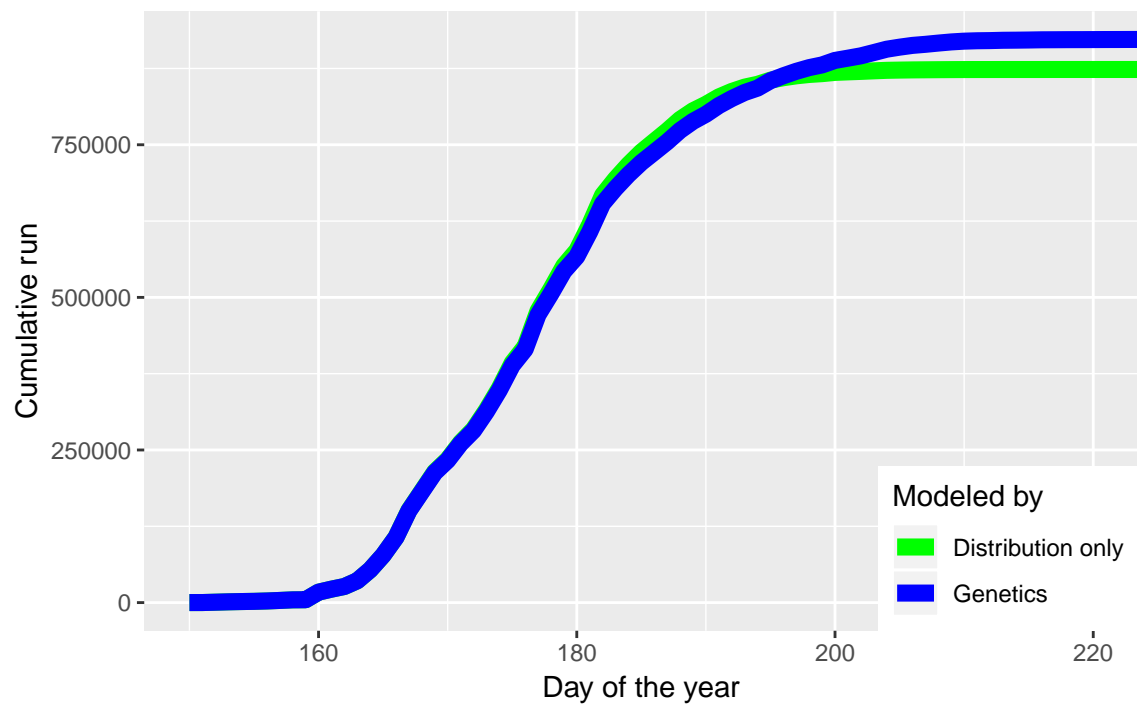
## Appendix A

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

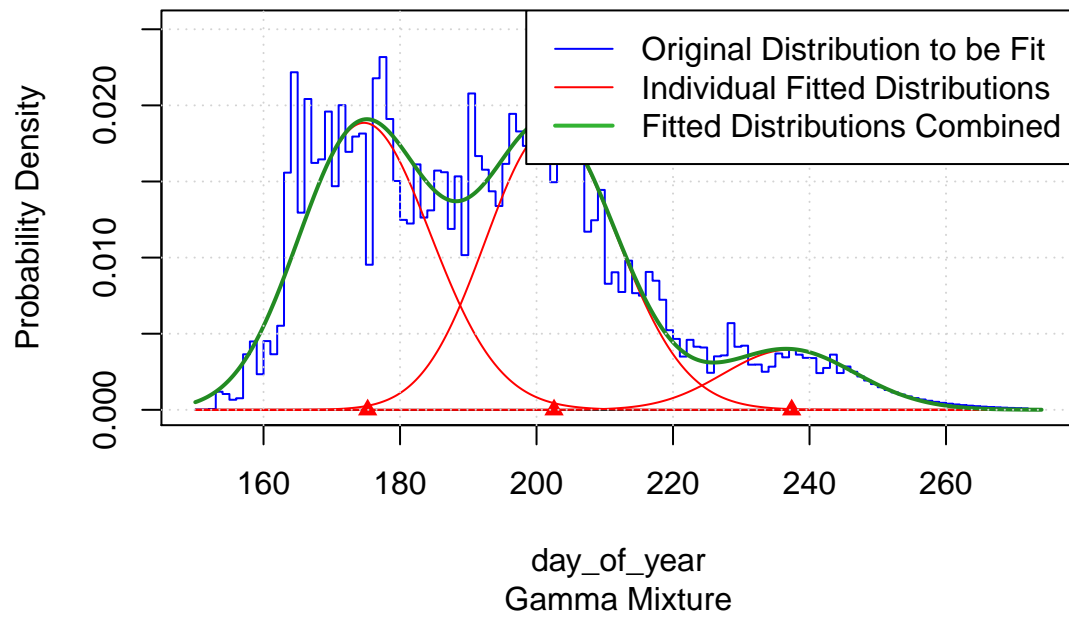
**2006**



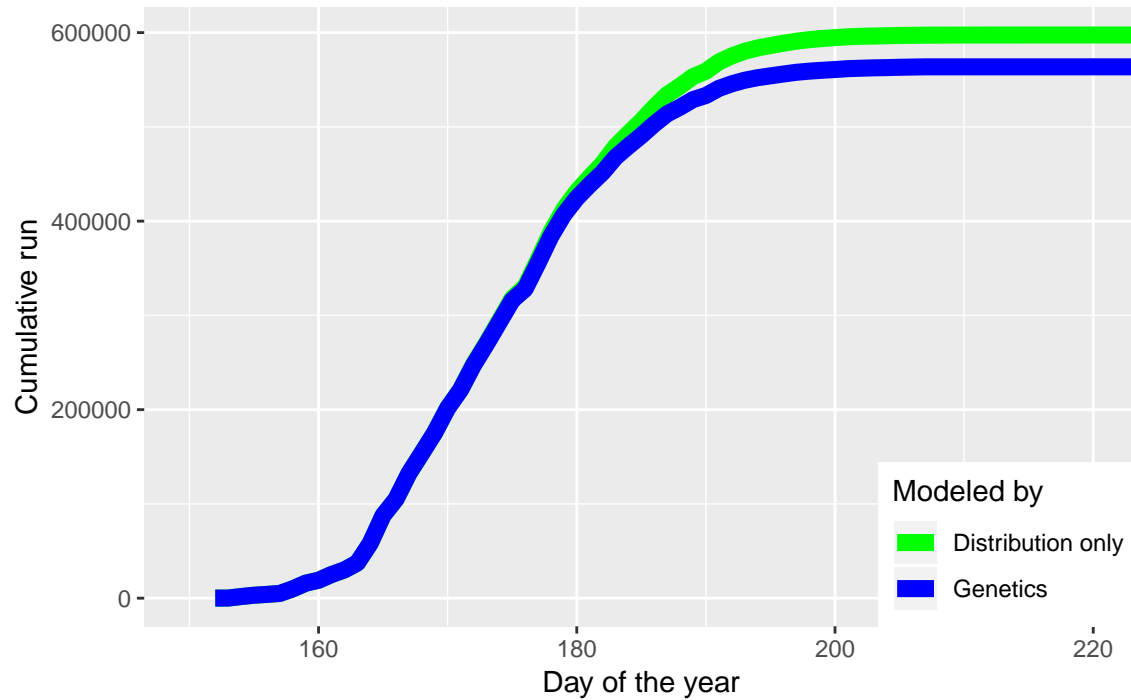
**2006 Early Run Estimation**



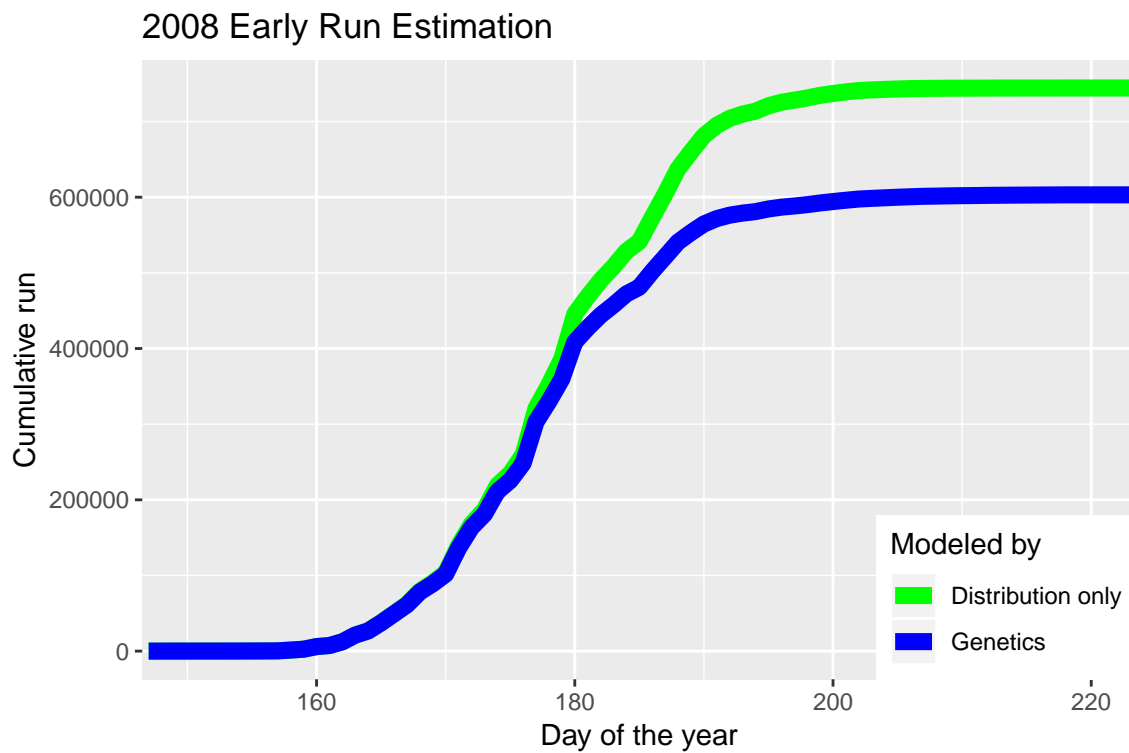
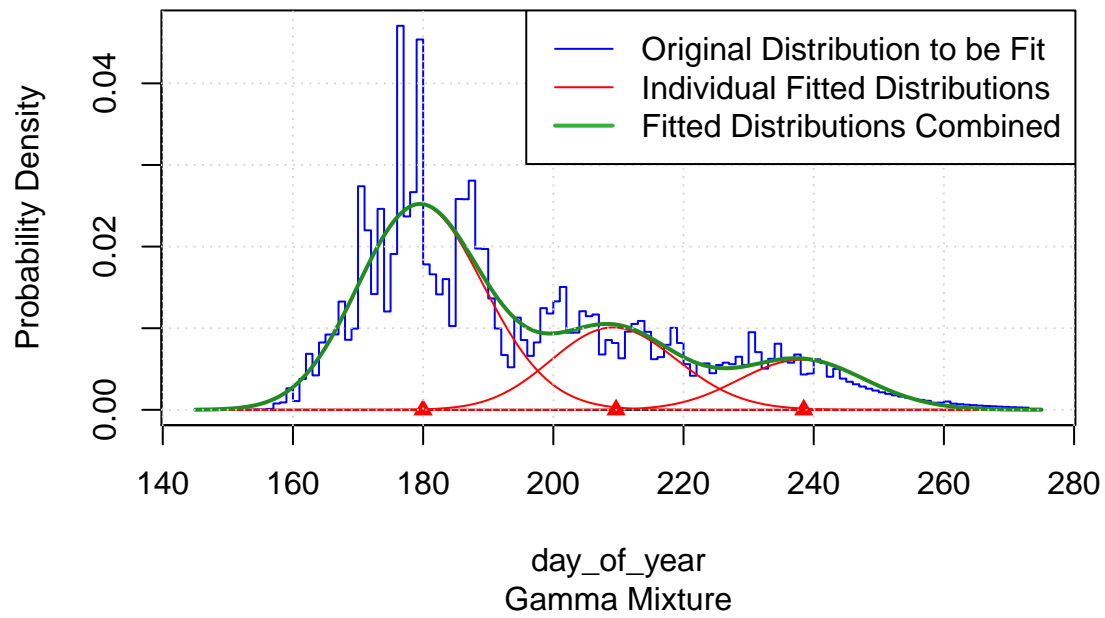
2007



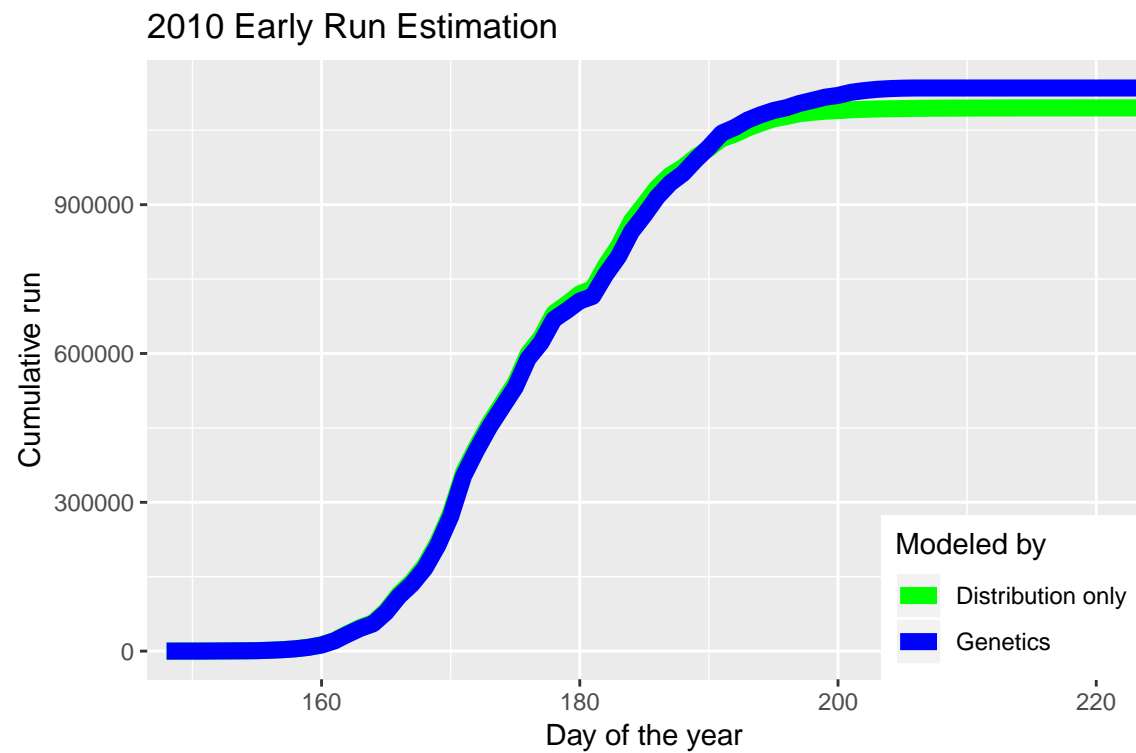
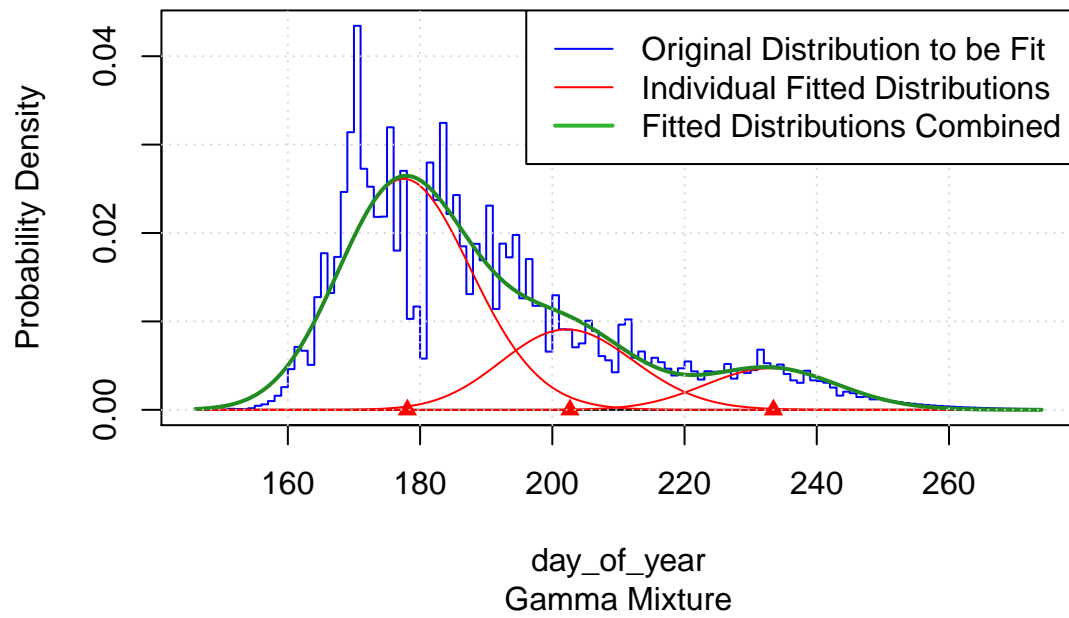
2007 Early Run Estimation



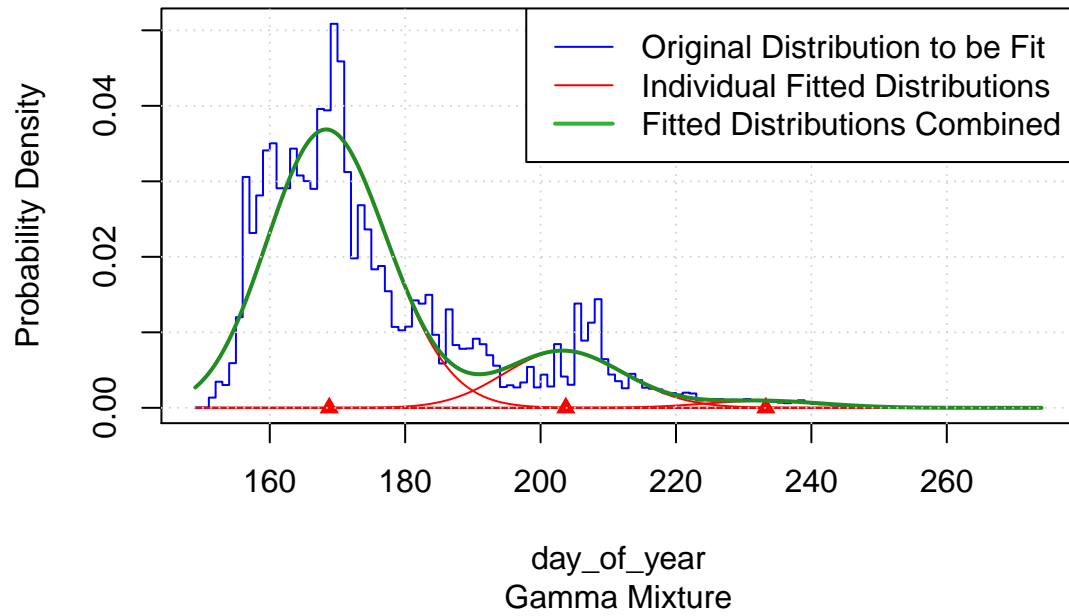
2008



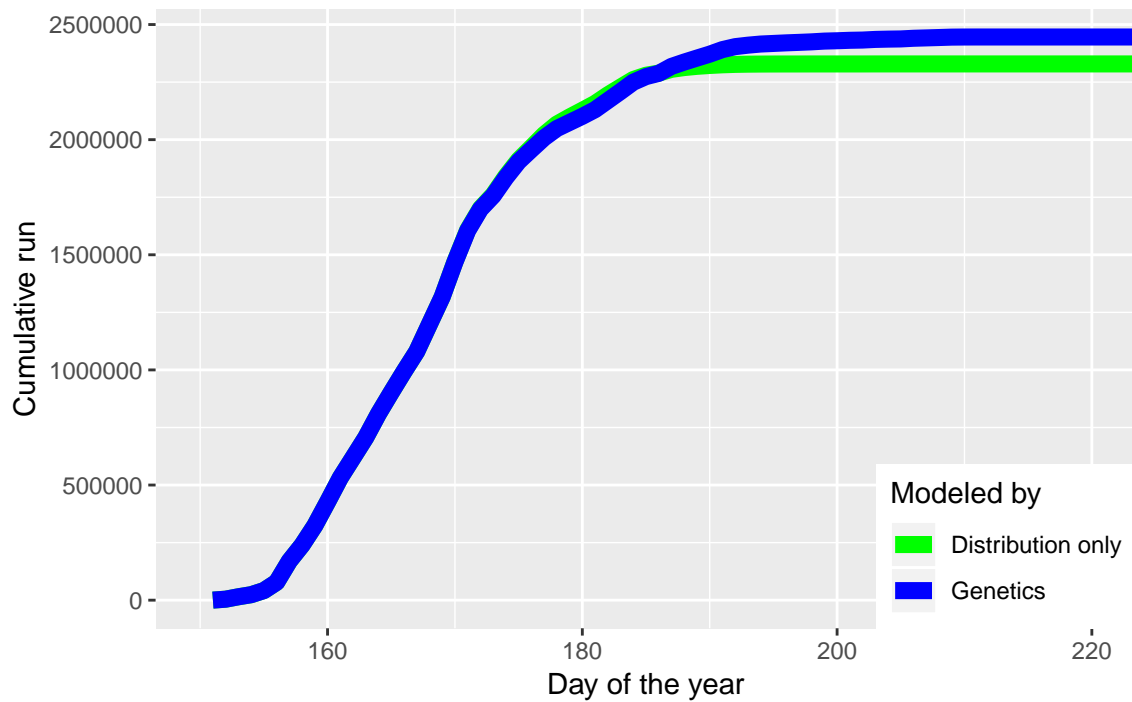
2010



2011

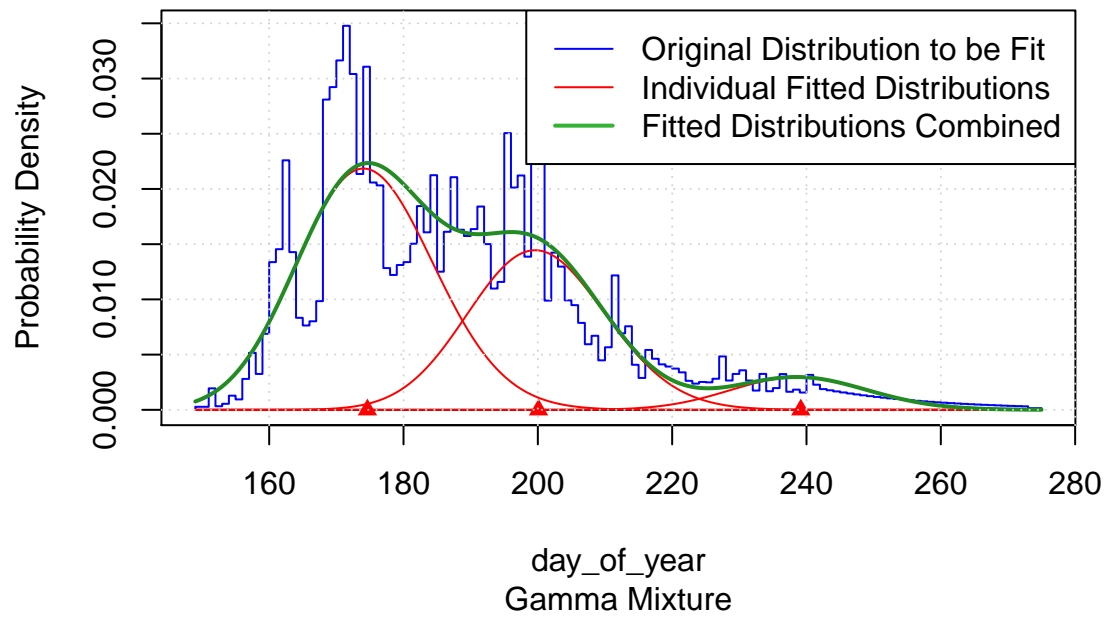


2011 Early Run Estimation

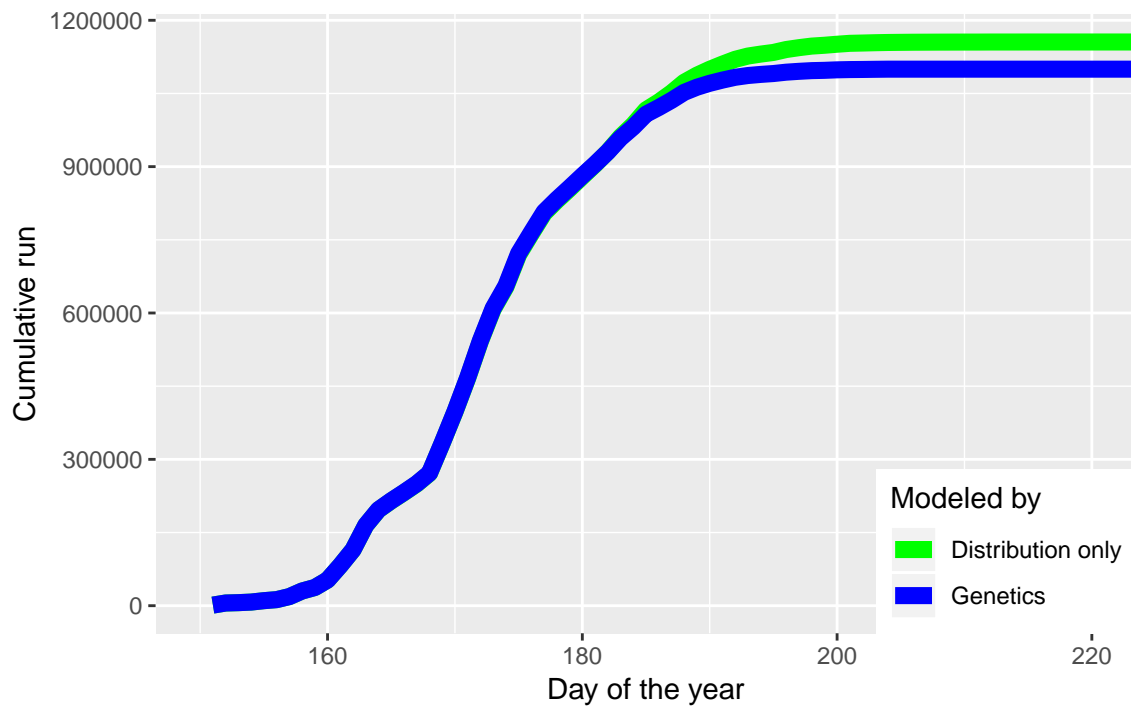




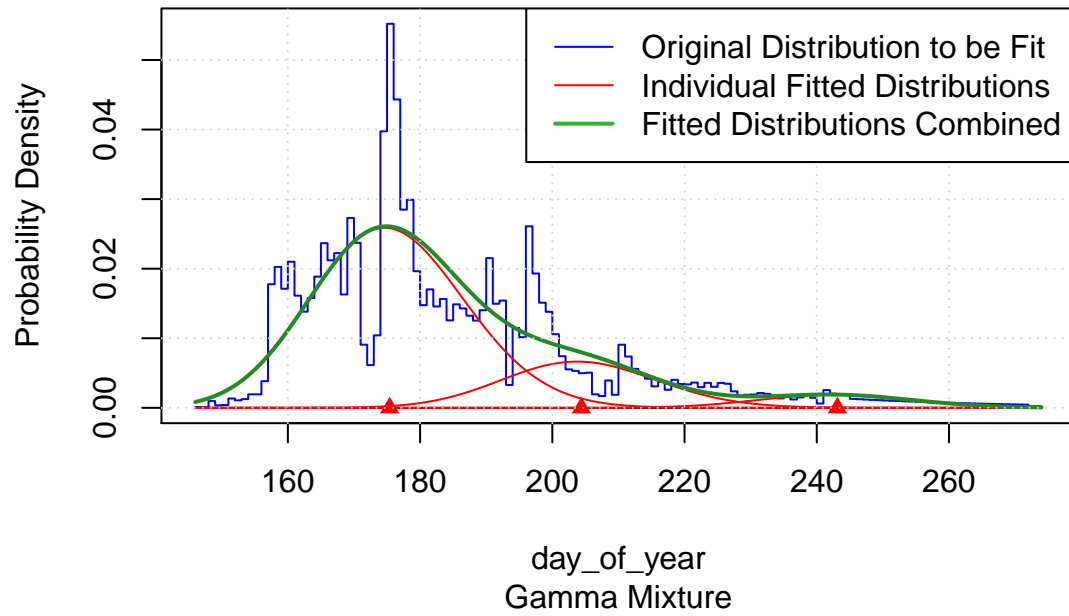
2012



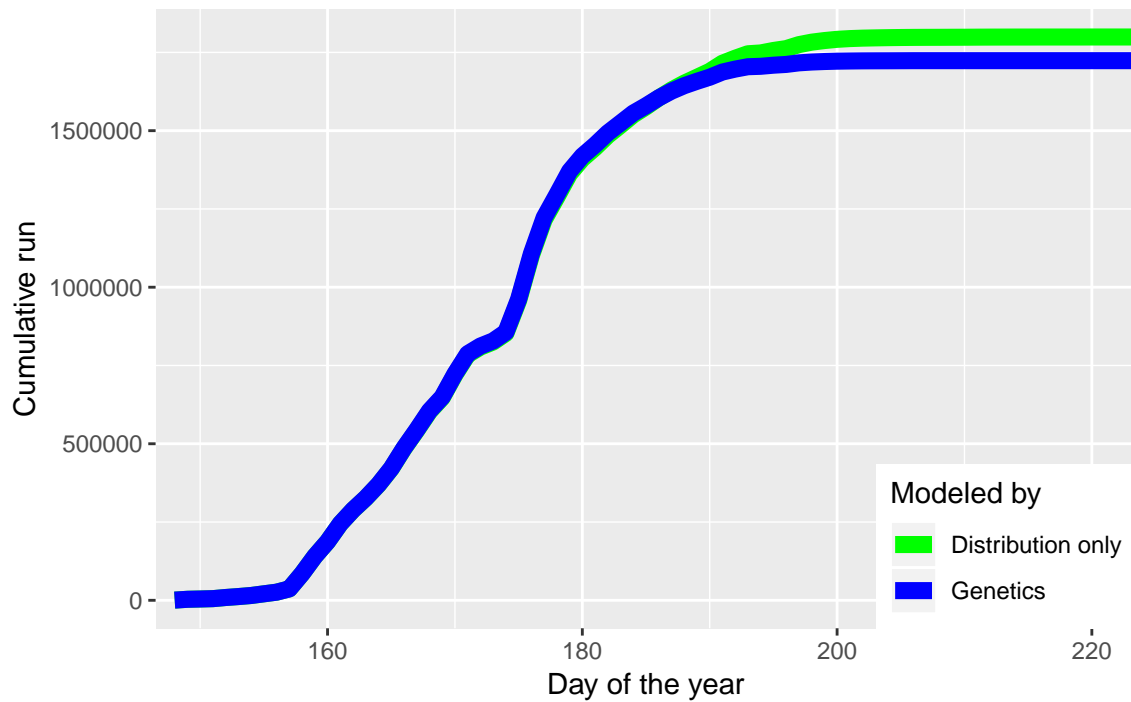
2012 Early Run Estimation



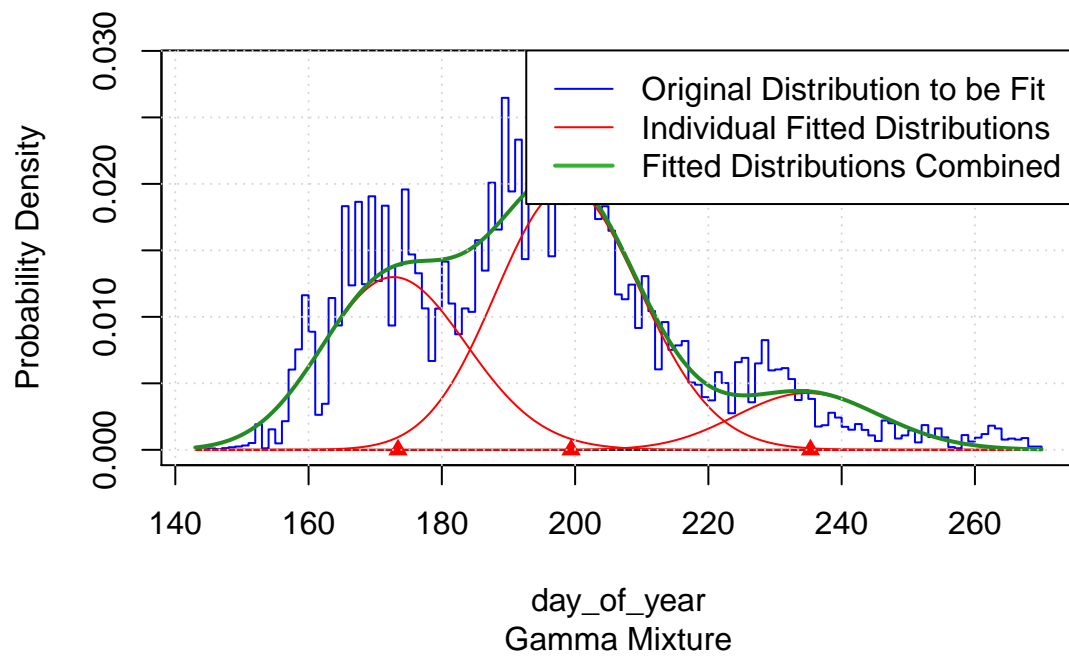
2013



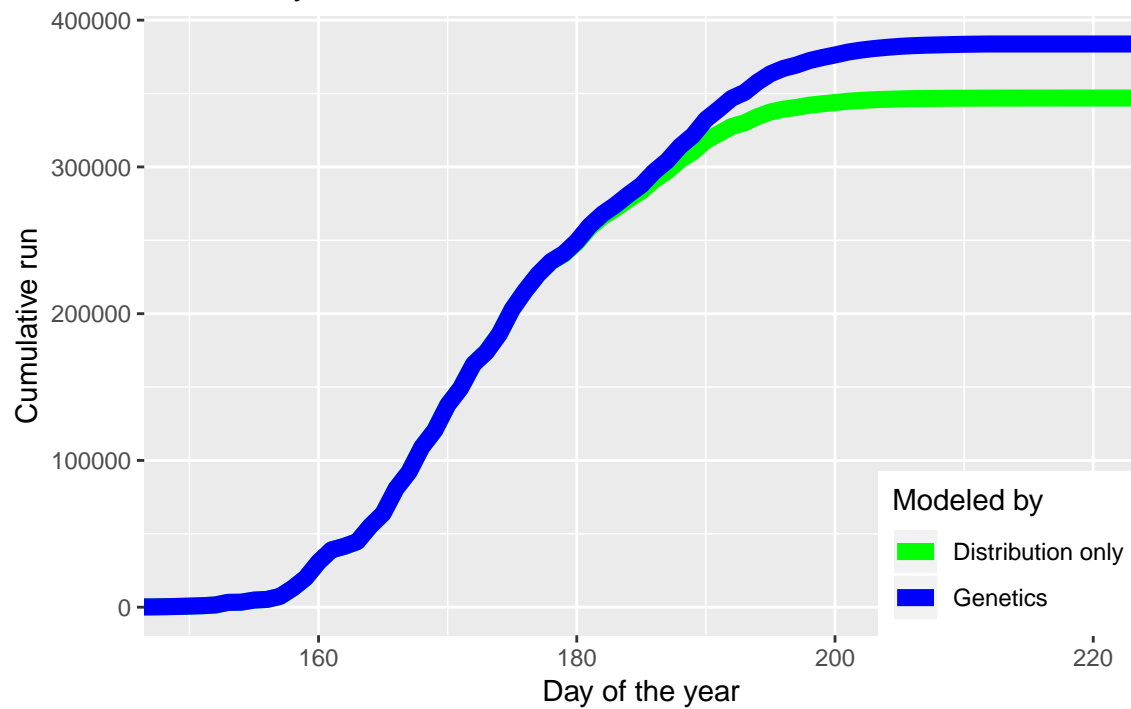
2013 Early Run Estimation



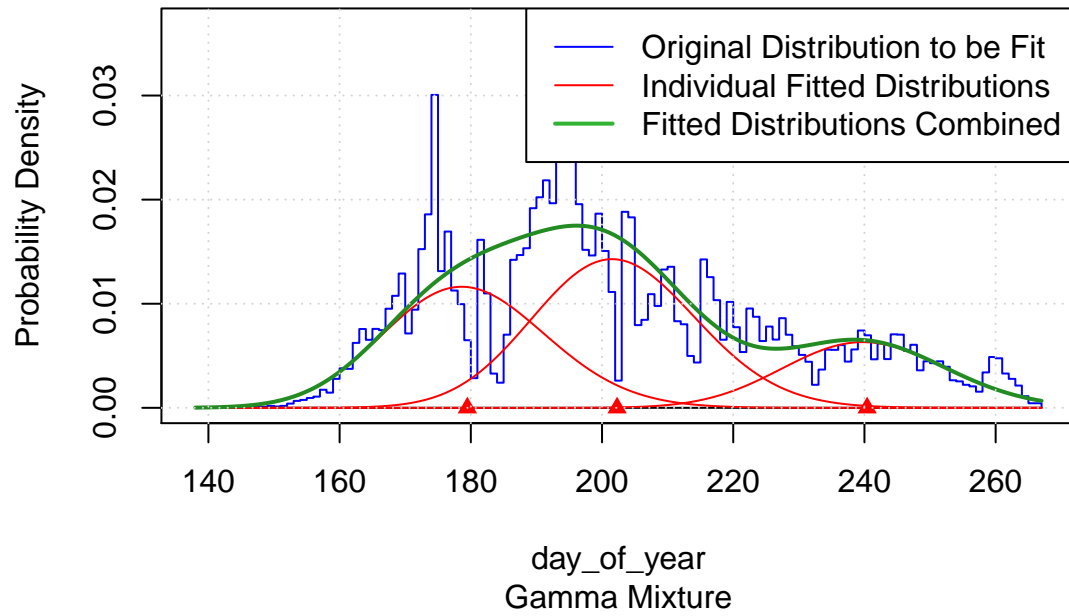
2014



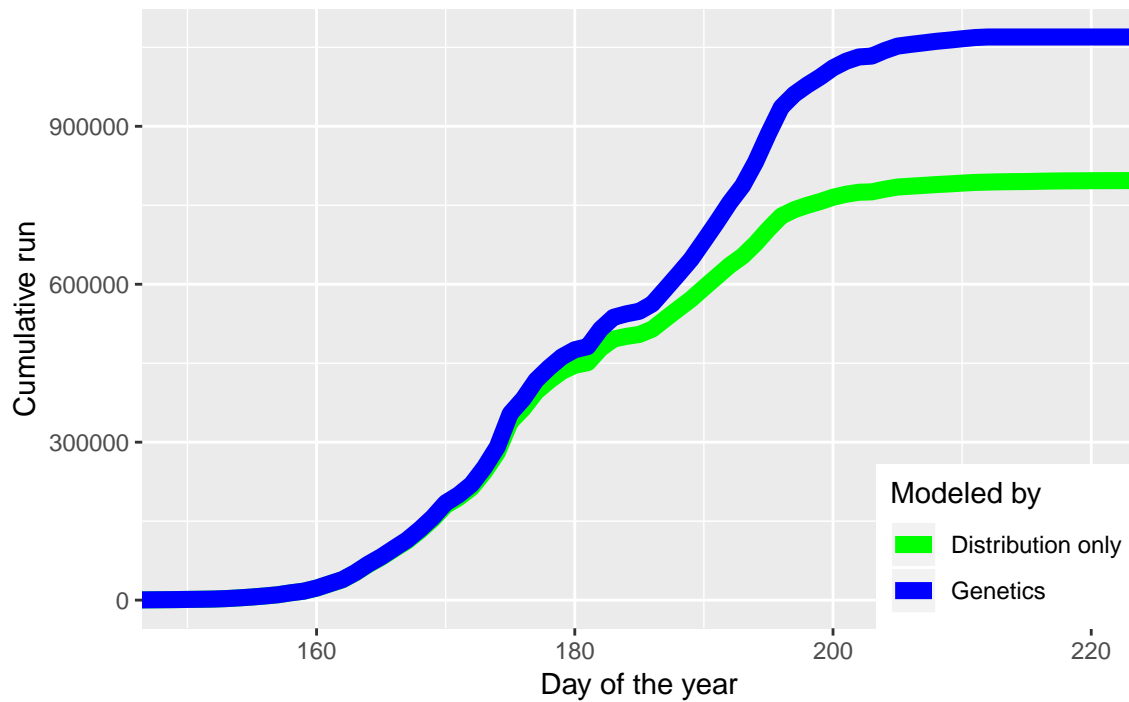
2014 Early Run Estimation



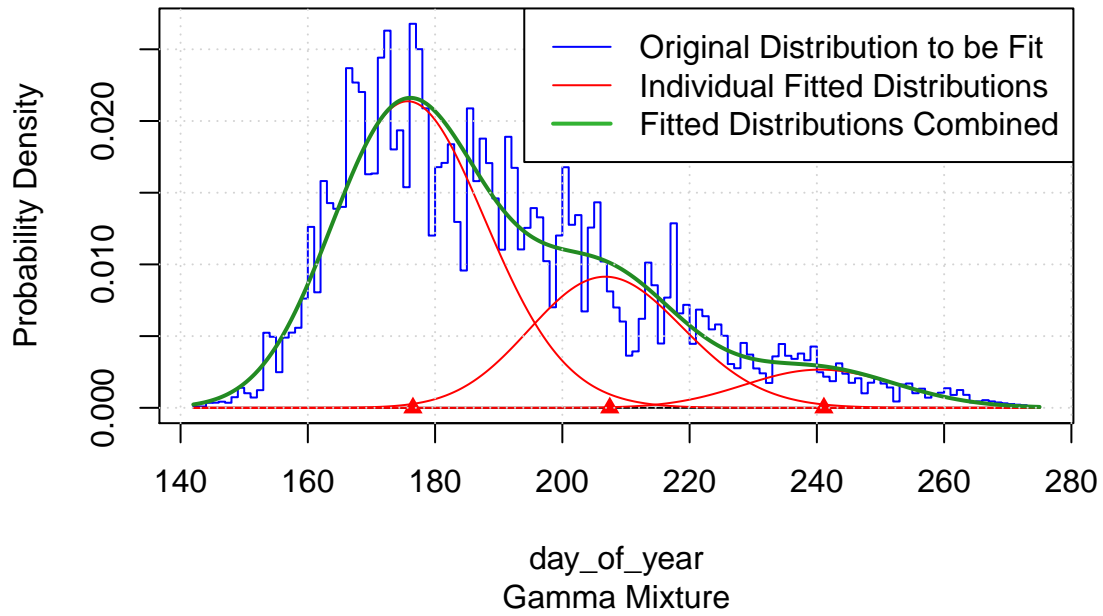
2015



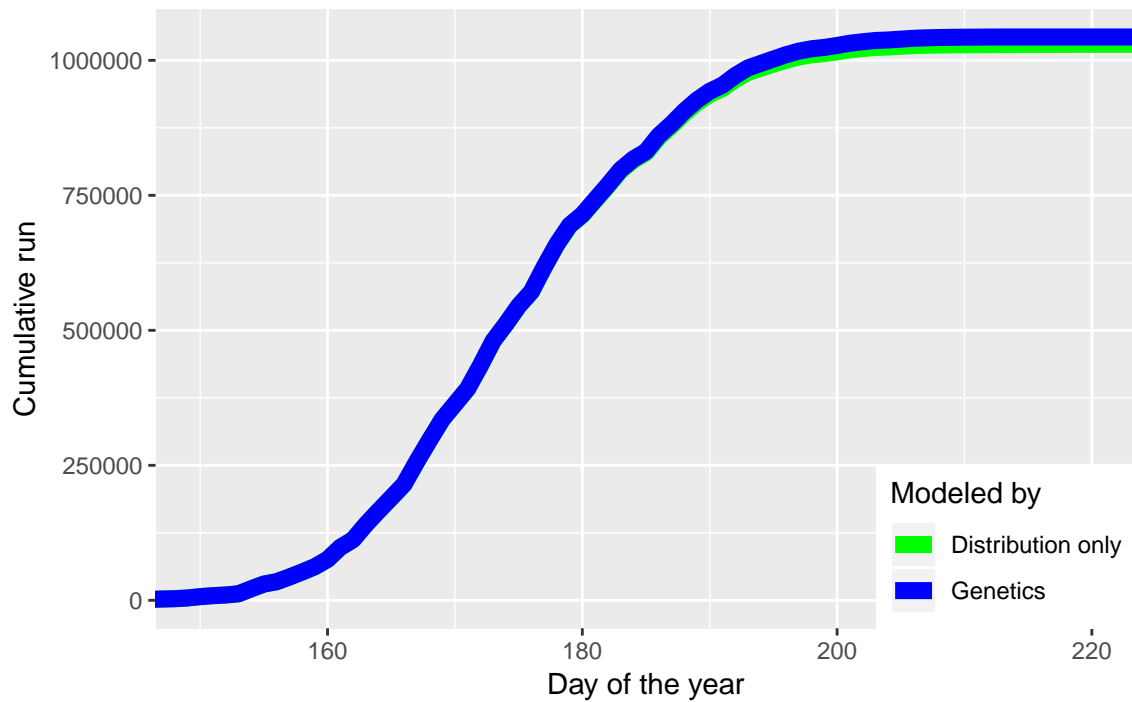
2015 Early Run Estimation



2016



2016 Early Run Estimation



2017

