

# Chignik Early Run Modeling

*Sarah Power*

*2018-09-24*

## Recipients

Heather Finkle, Birch Foster, Kevin Schaberg.

## Chignik Sockeye Early Run Timing

## Background

For over a decade genetic samples of sockeye (*Oncorhynchus nerka*) have been collected at the Chignik weir which has distinguished the early run, which mostly spawns in Black Lake from later runs. The majority of these later runs spawn in Chignik Lake. Currently these two stocks are managed separately. The processing of genetics samples is expensive, and so 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 will examine 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 to those that are determined with the additional aid of genetics to see how viable this method might be. To determine which model is most appropriate and how viable this method is, similarity to the mean and standard deviation of the early run and number of sockeye in the early run as determined with the additional aid of genetics will be used .

Currently the run is the sum of the escapement at the weir, and catch from various multiple management districts, some of which are quite far away.

## Methods

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, which 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 run timing distributions from 2006 to present there appears to be three distributions. 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 we will take is that there is a biological reason for three distributions to be presented in the data, which were modeled in the analysis.

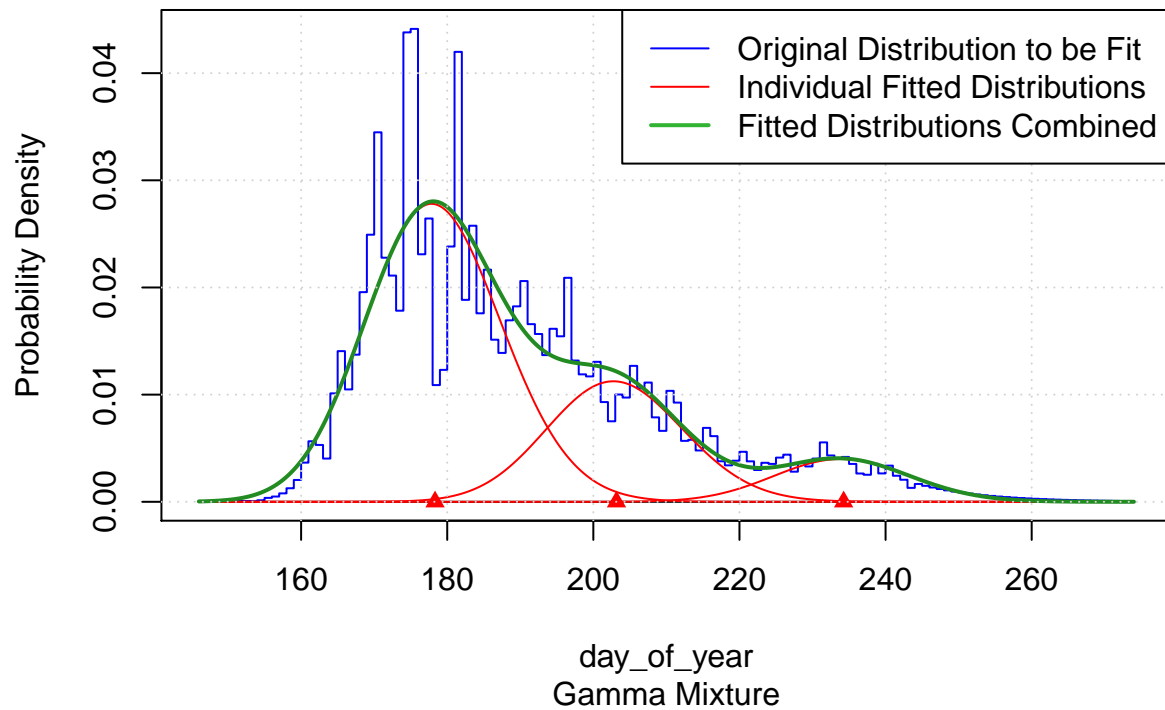
I used maximum likelihood methods, including the EM algorithm and a Newton-type algorithm to parse out various distributions using the ‘mixdist’ package in R. In initial investigations Normal and Weibull distributions were fit. Also Normal distributions with certain constraints were considered in order 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 a particular mean or standard error. The fixed means and standard errors were the median values from the early distribution determined using genetics.

It was determined that the most appropriate fitting model was a three distribution normally-distributed model with the constraint of equal standard deviations for each of the distributions. This was determined as

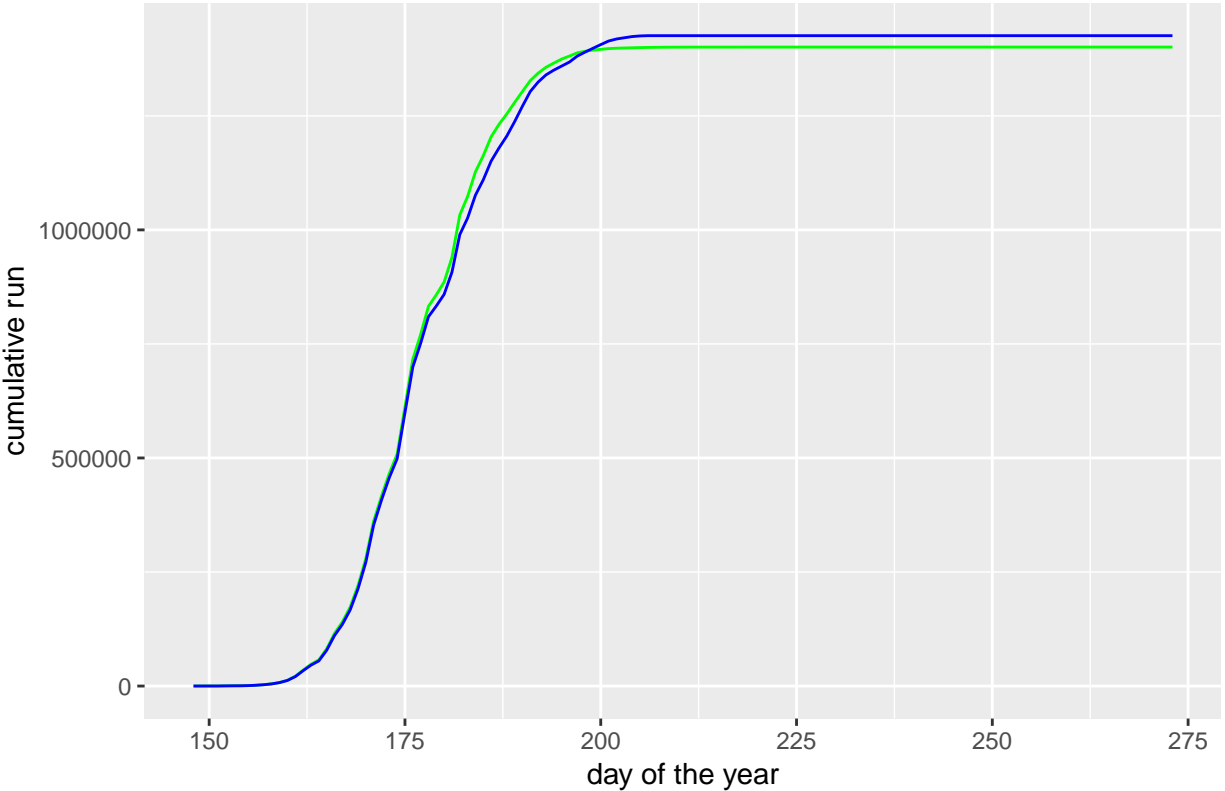
this model had an early run with a mean and standard deviation most like the early run as determined with genetics.

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 runtiming alone were within 10% of the counts using genetic information. The three best years included 2010, 2013, and 2014, where early run counts using runtiming information alone differed by no more than 3% of the counts using additional genetic information.

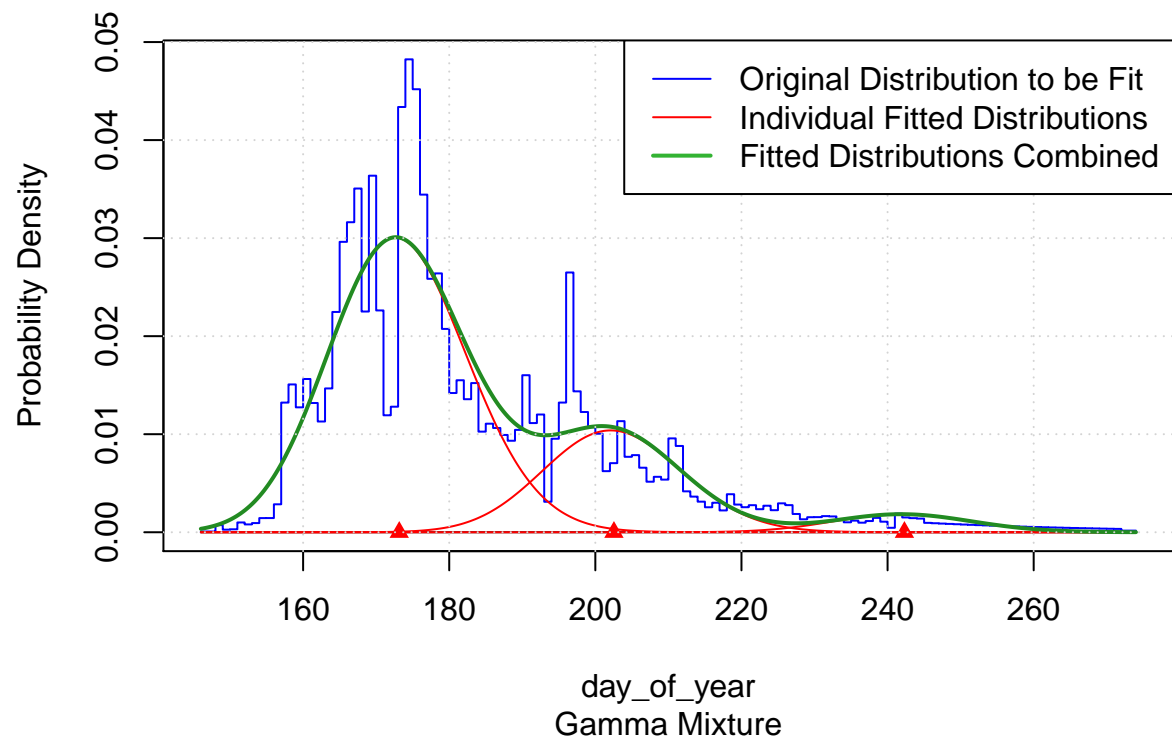
## 2010

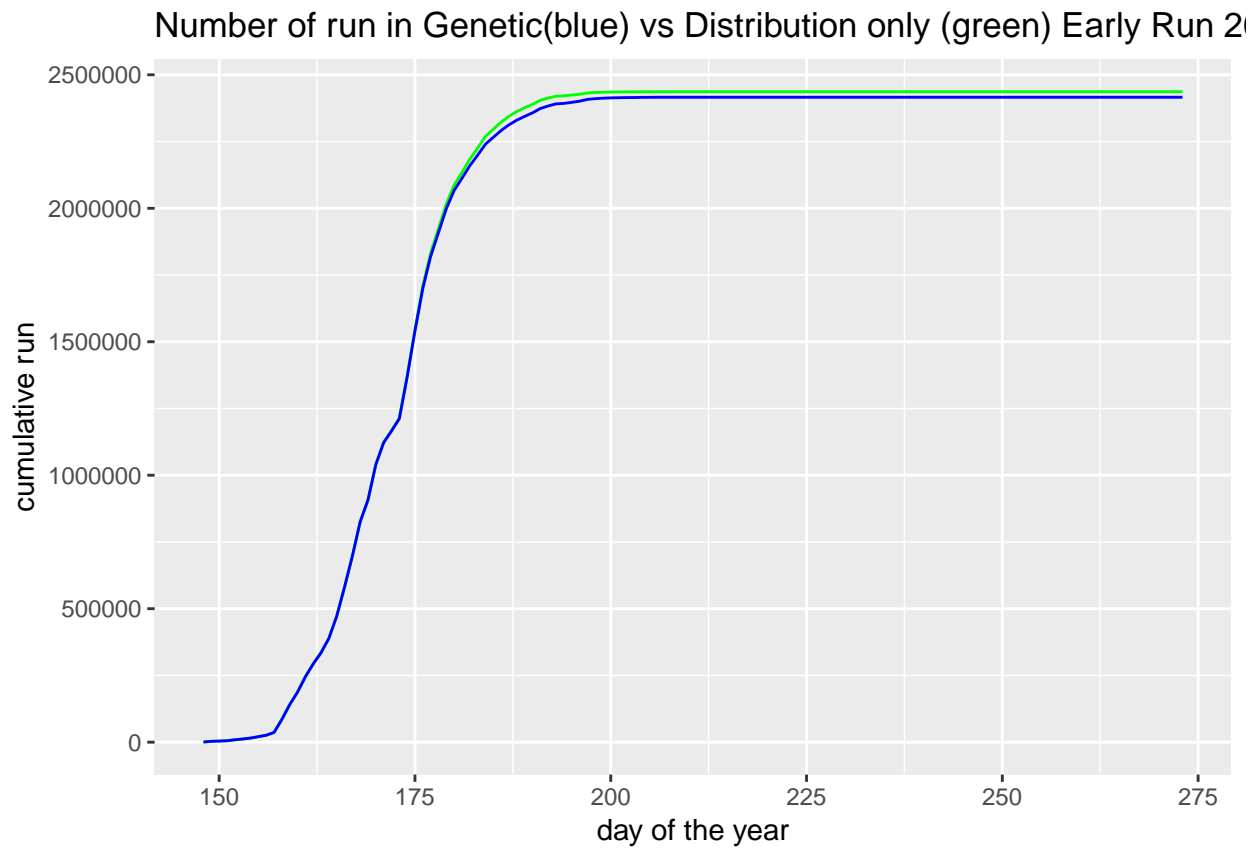


Number of run in Genetic(blue) vs Distribution only (green) Early Run 2

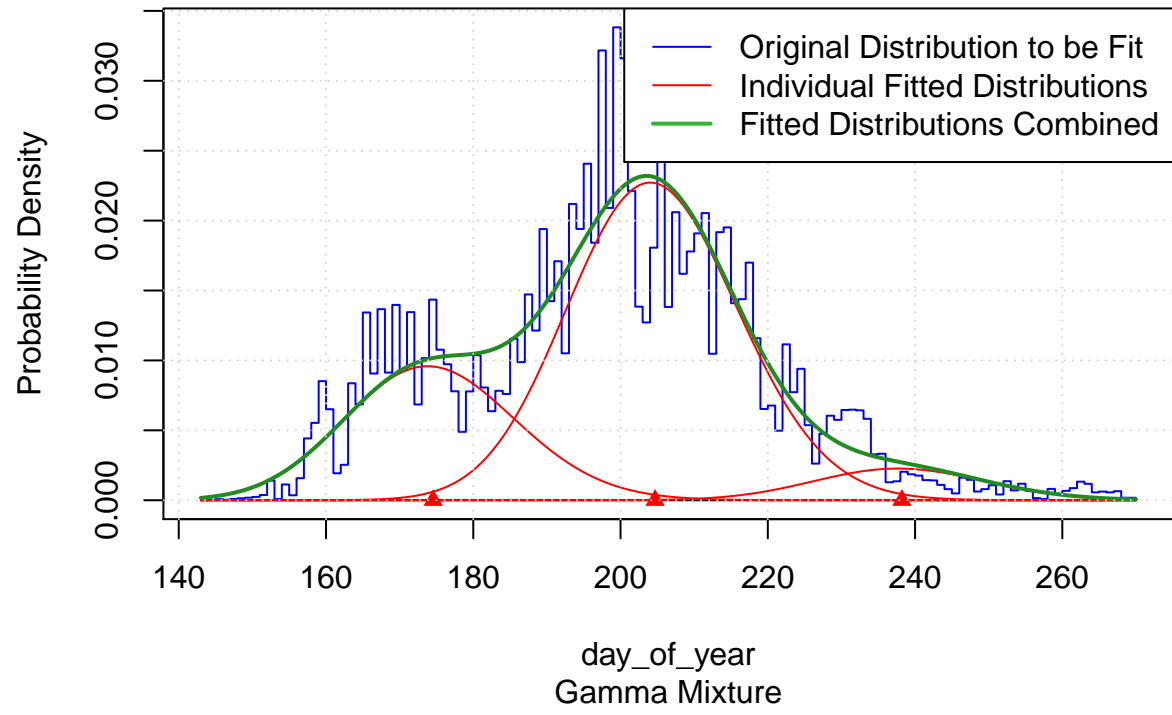


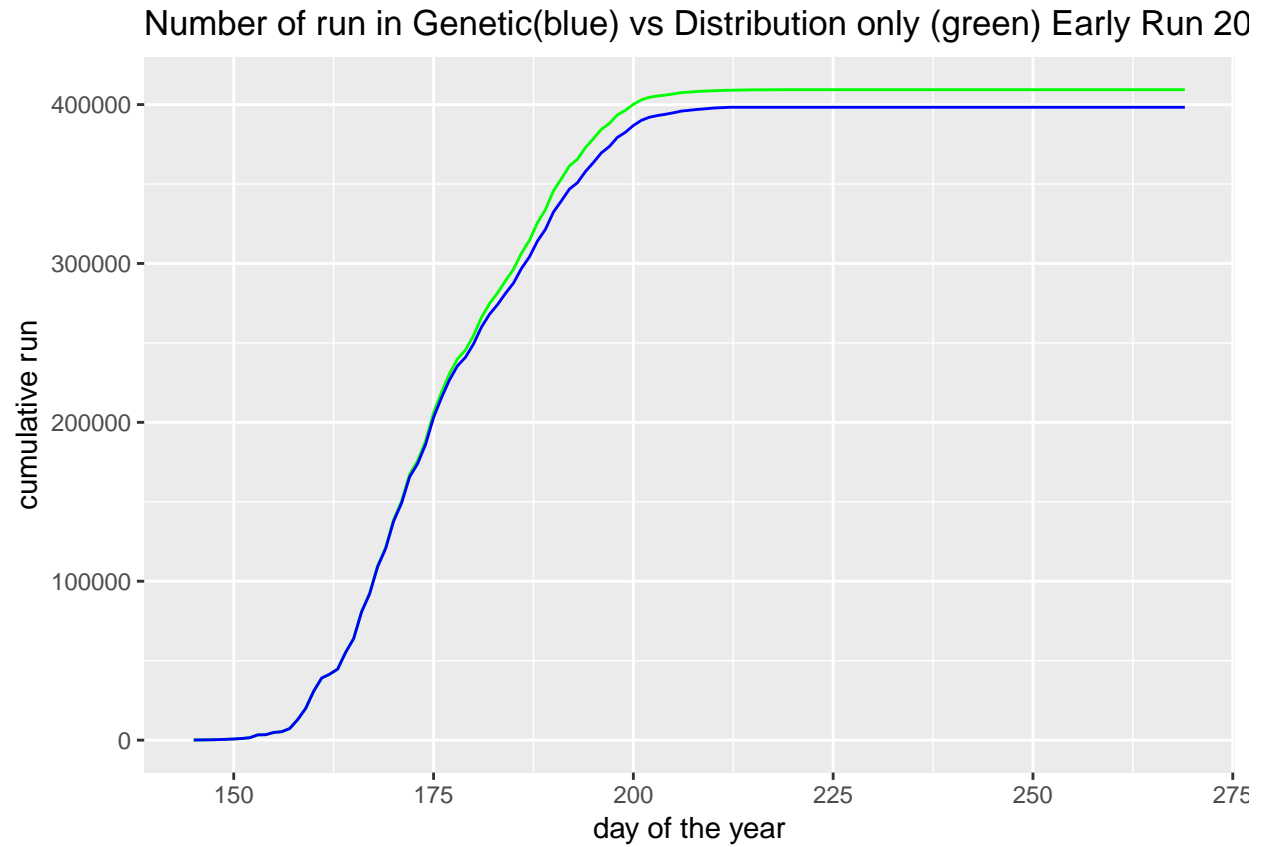
2013





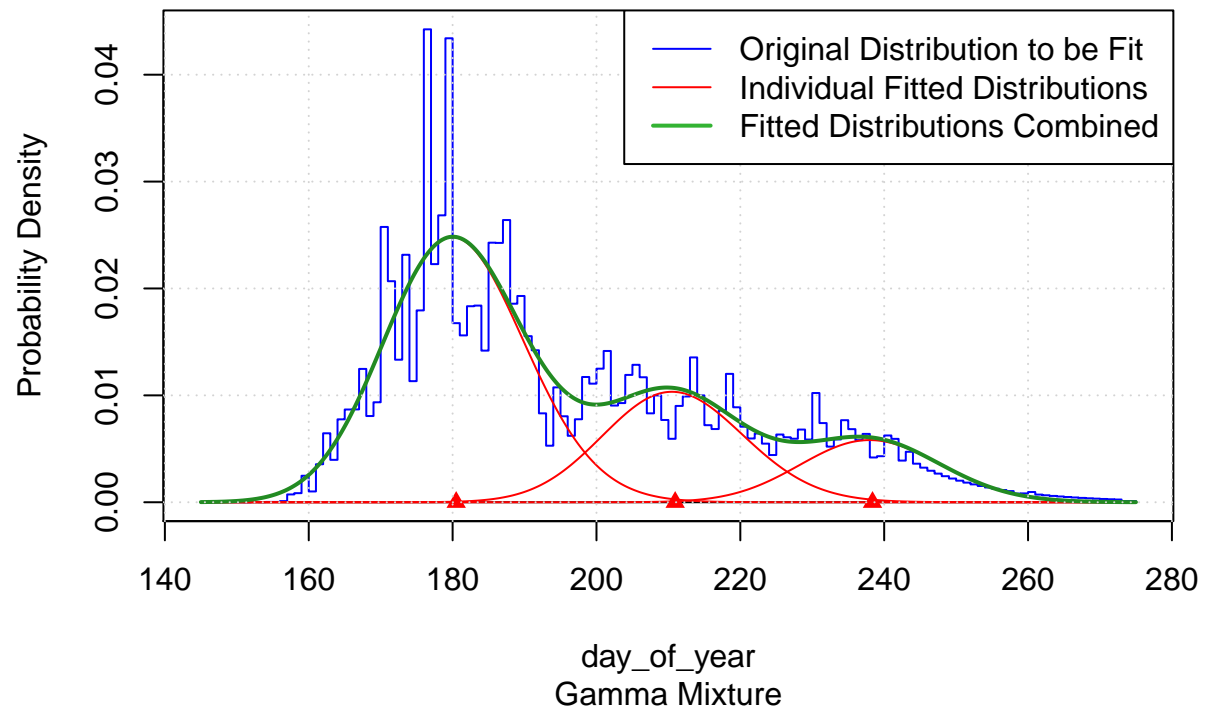
2014



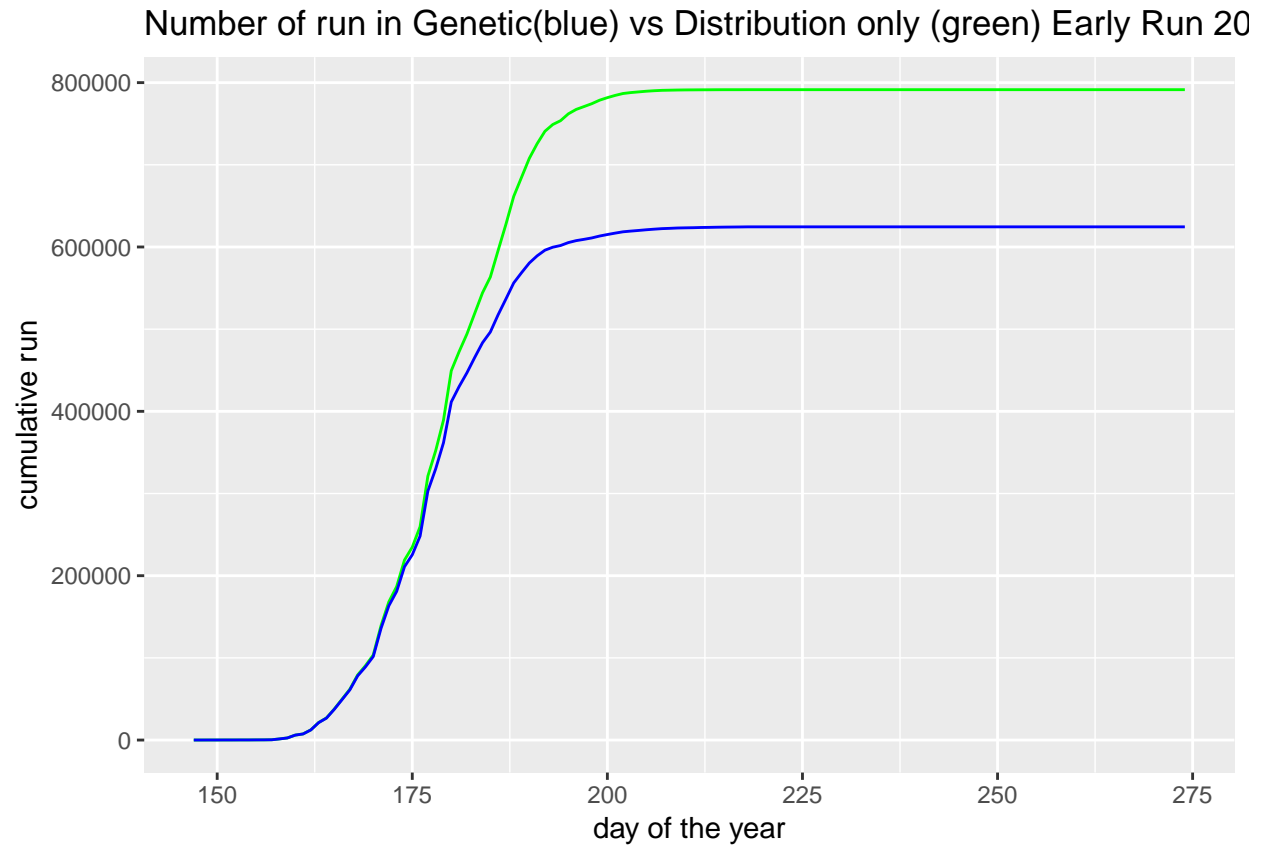


The three years that runtiming distributions were off by more than 10% included 2008, which overestimated the early run by 27%, 2015 where it underestimated the early run by 25% and 2012 where is over estimated the run by a whopping 67%.

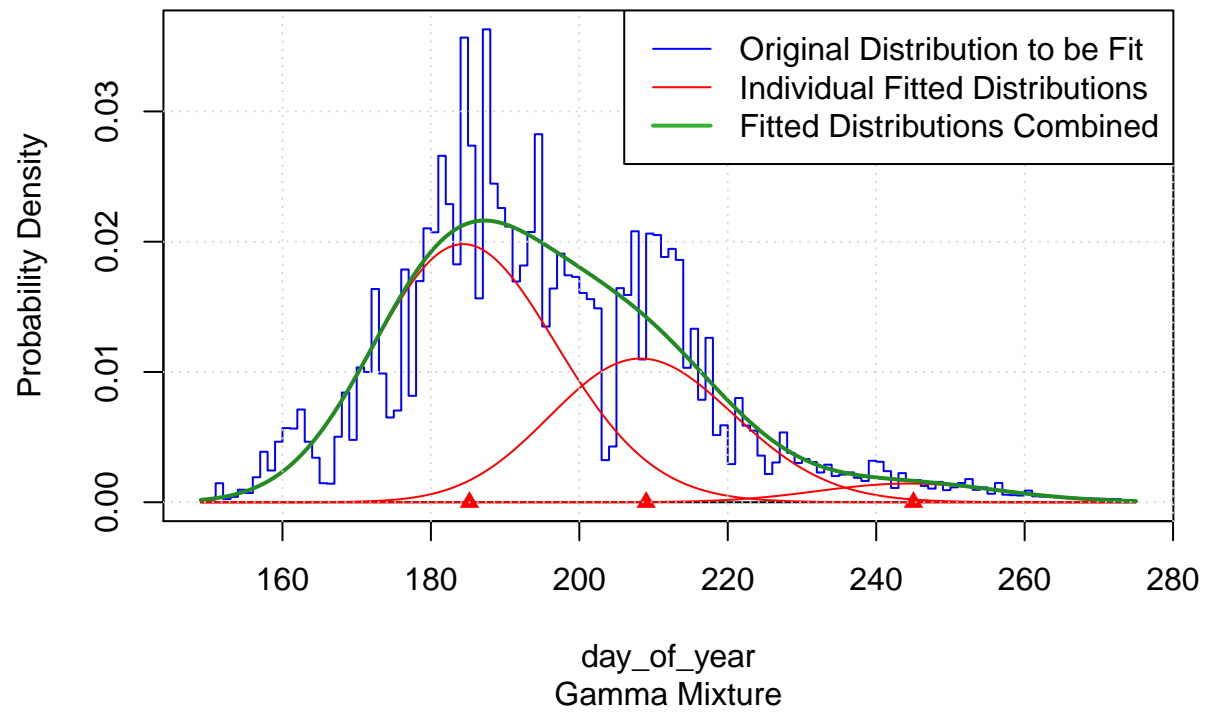
2008

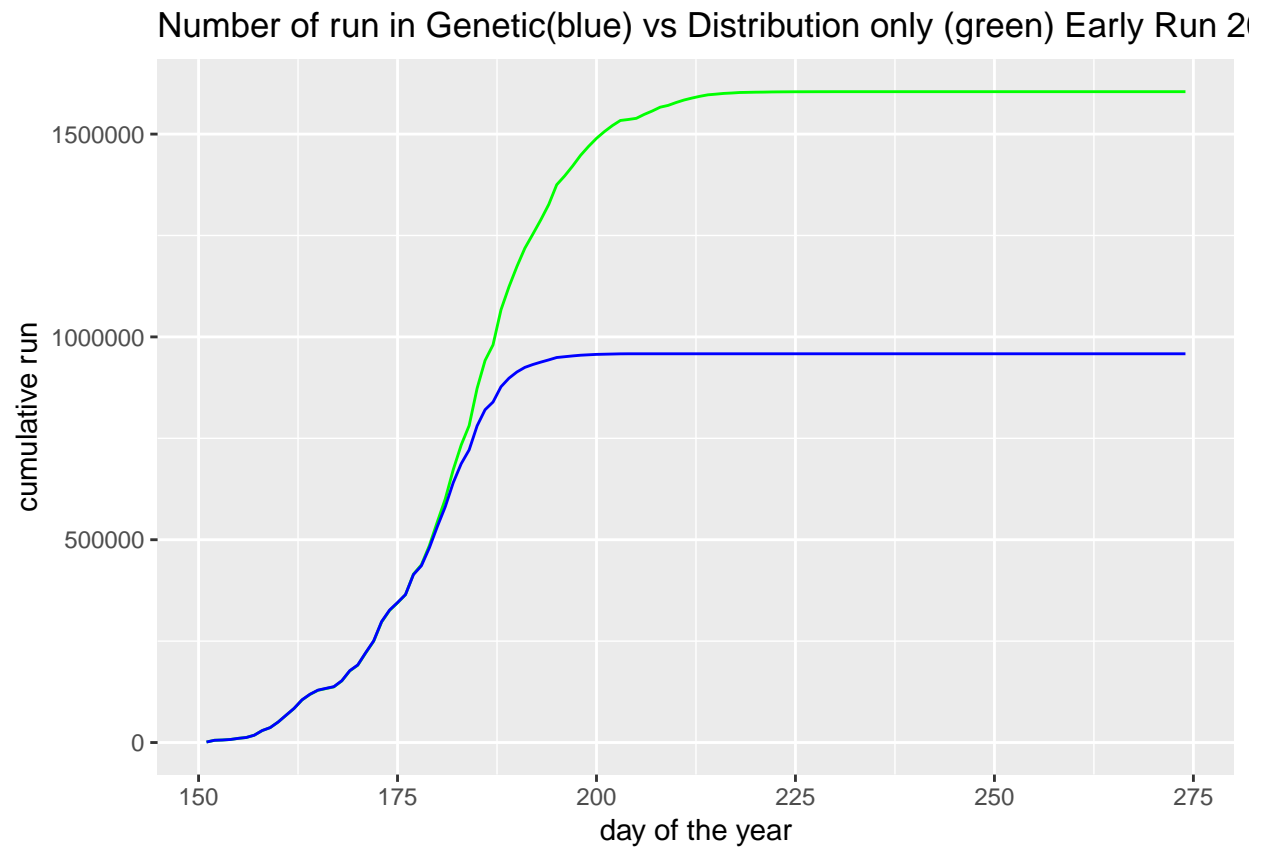




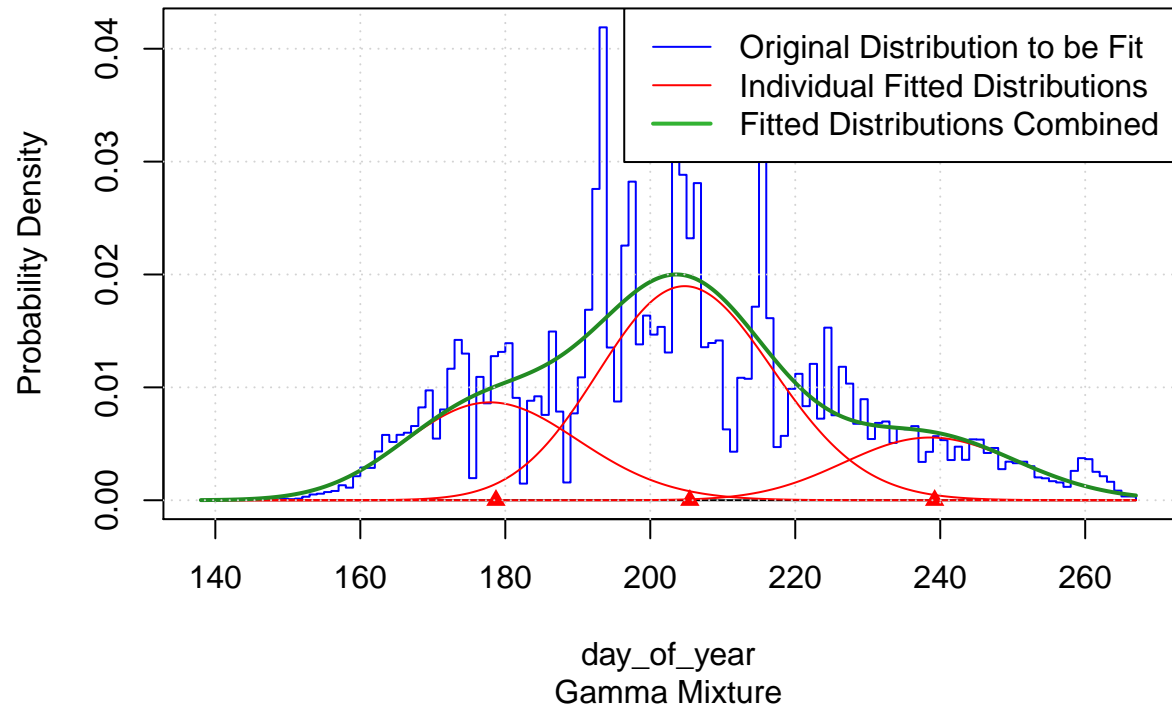


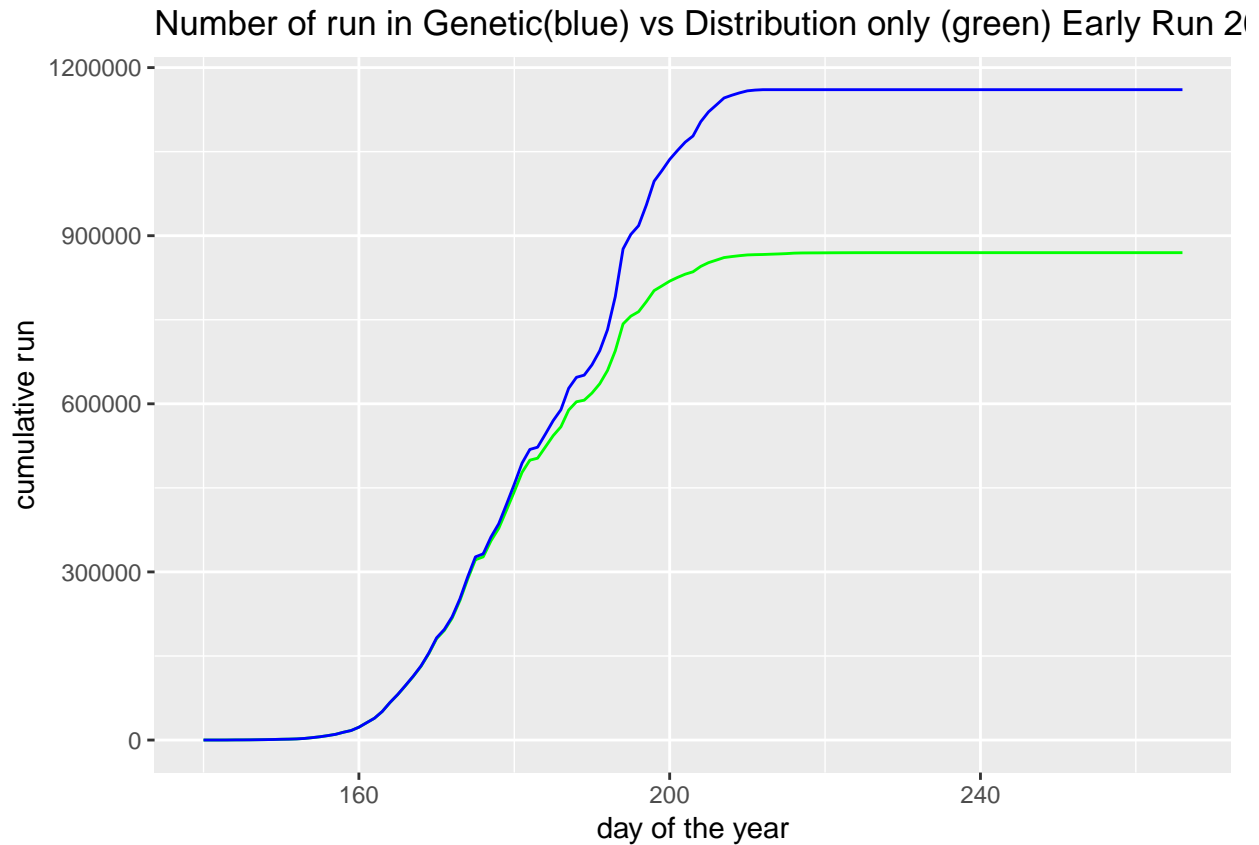
2012





2015

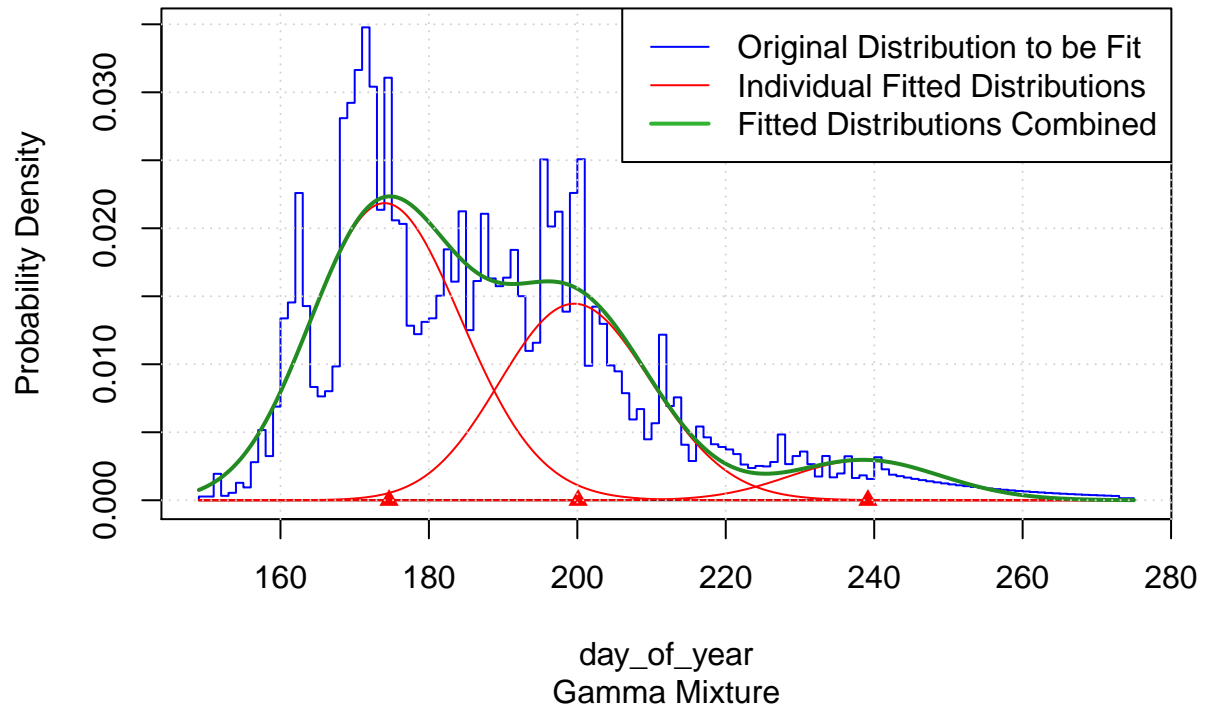


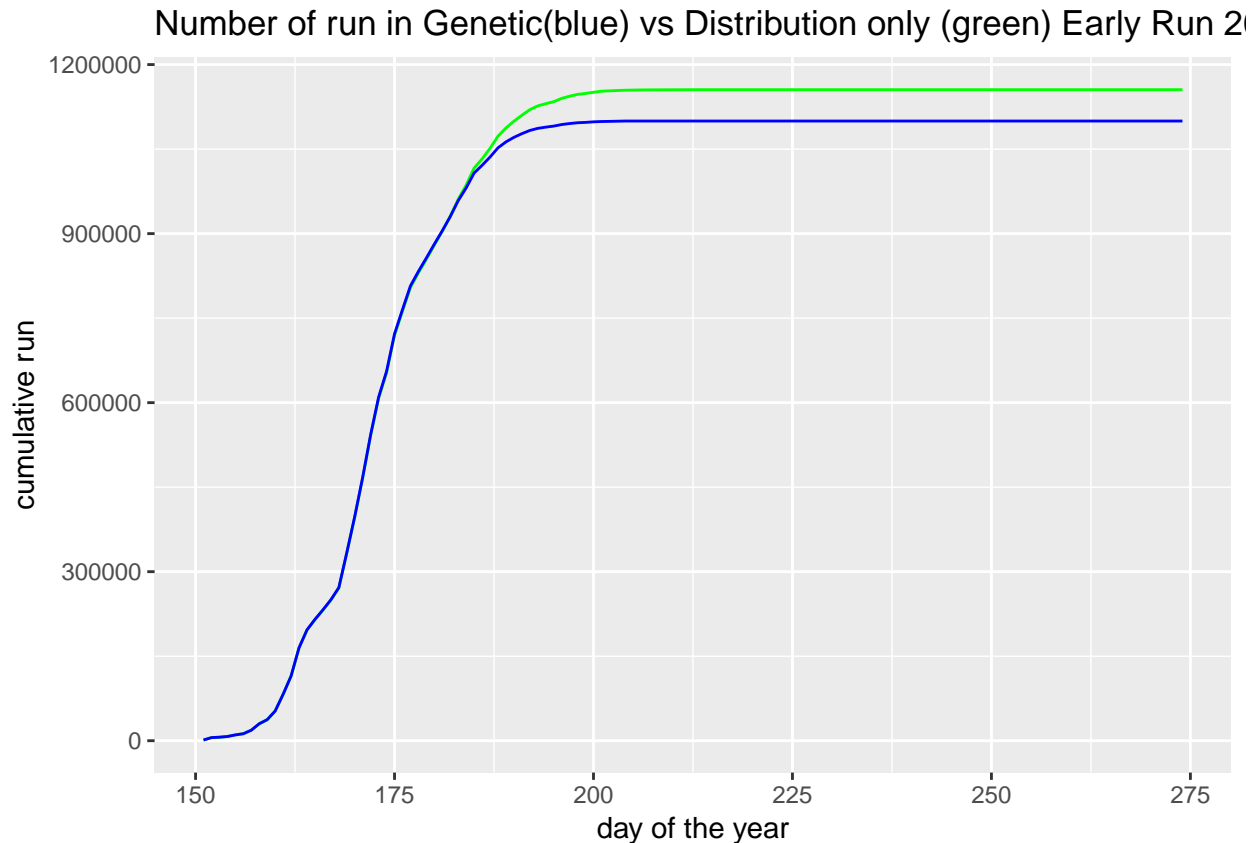


It was suggested by Heather Finkle that using catch information from farther flung districts might add noise to the distributions since those districts surely would contain harvest from non-Chignik stocks. The analysis was reexamined this time using only harvests that were a day or twos fish-swim away.

This change improved the model. The most significant change being that of the year 2012 which changed from an over estimate of 67% to an over estimate of 5%, which is much more reasonable. The 2008 and 2015 estimates were similar with a 25% over estimate and 26% underestimate respectively.

2012





#### # Recommendations/Discussion

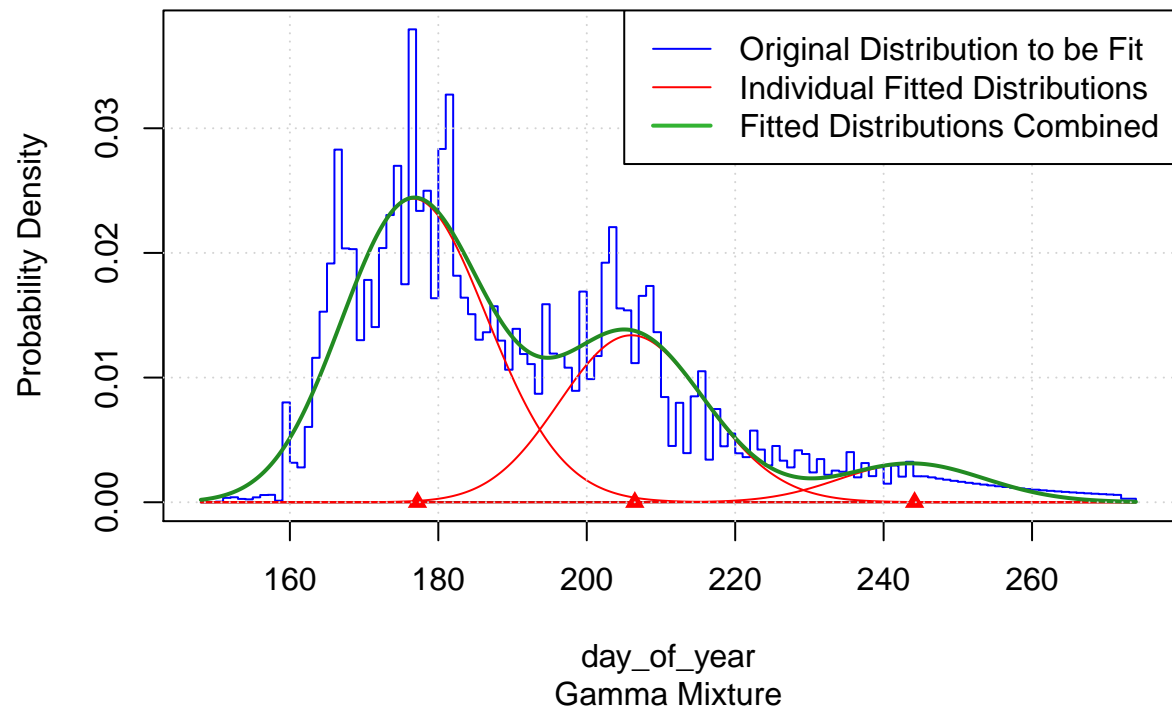
Future analysis could include examining if there are any factors that explain why 2008 and 2015 performed poorly with the model. Initial investigations note that the mean day-of-year for early run timing is shifted to the right 3 days for 2008 and 2 days for 2015, indicating that the estimated distribution is pulled to the right of what it would be using additional genetic information. Biologists note that harvest and runsize may have been different. Also examining time and area harvest to include in runtime estimates may improve the model.

It is important to note that the definition of the run, at least in terms of this model, changed from what is currently used. Harvest farther from the Chignik area were excluded. If this model, or one like it 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).

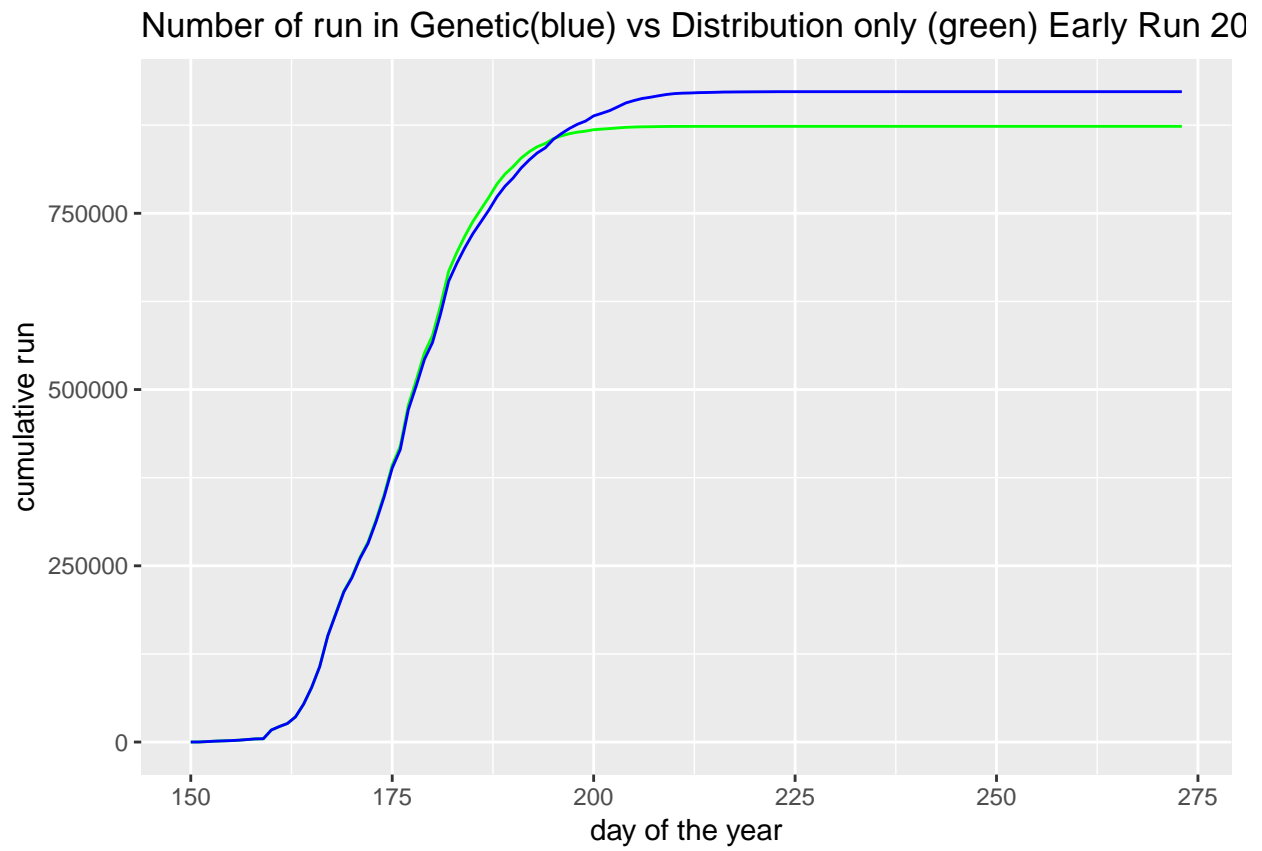
## Results

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

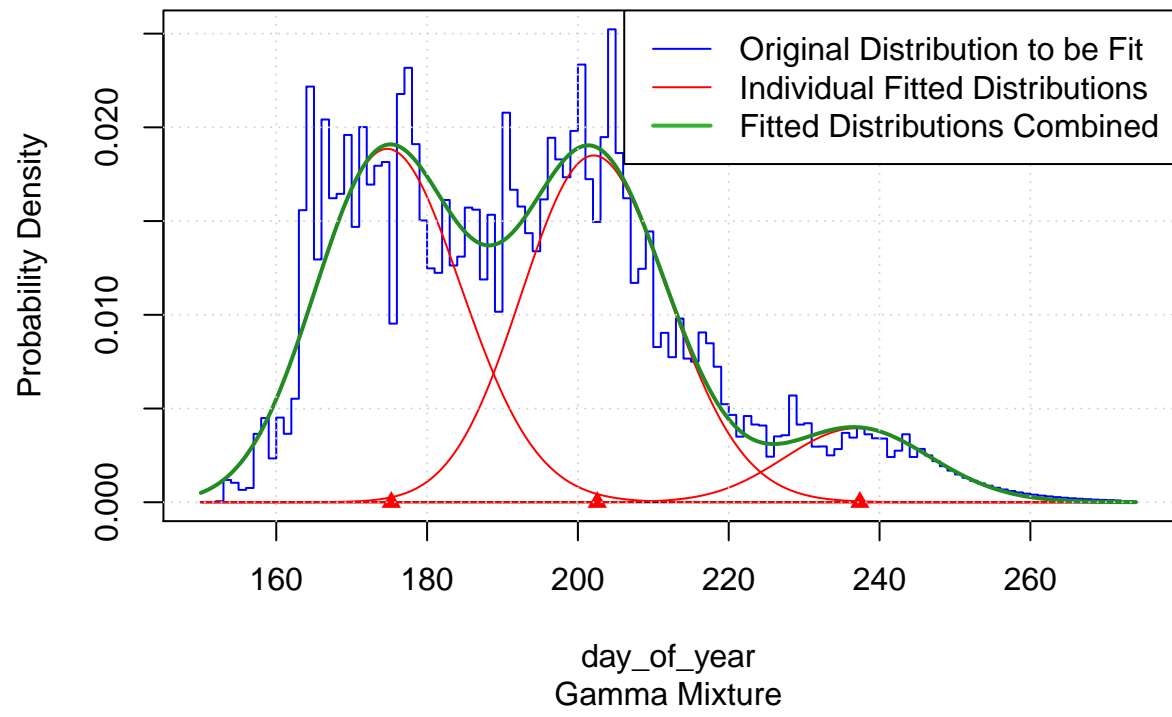
2006

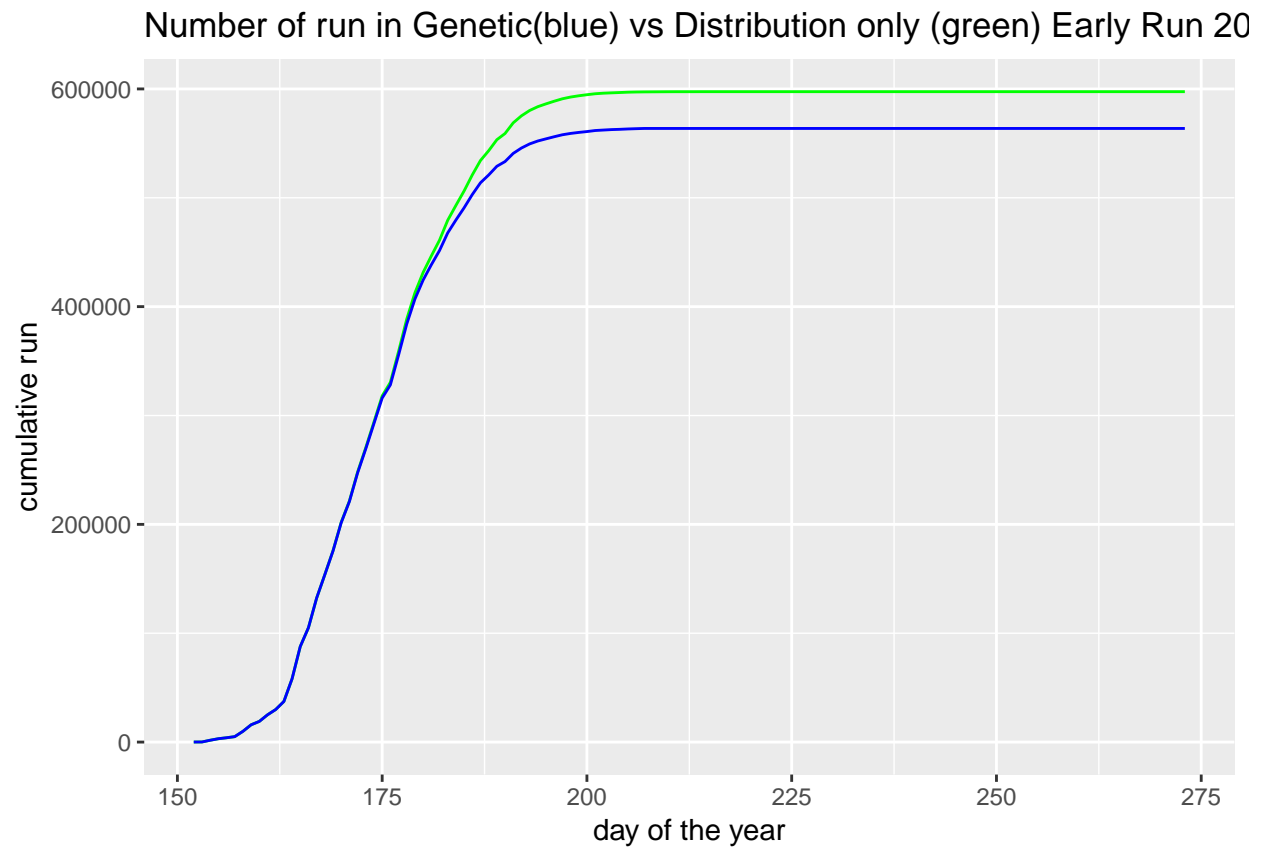




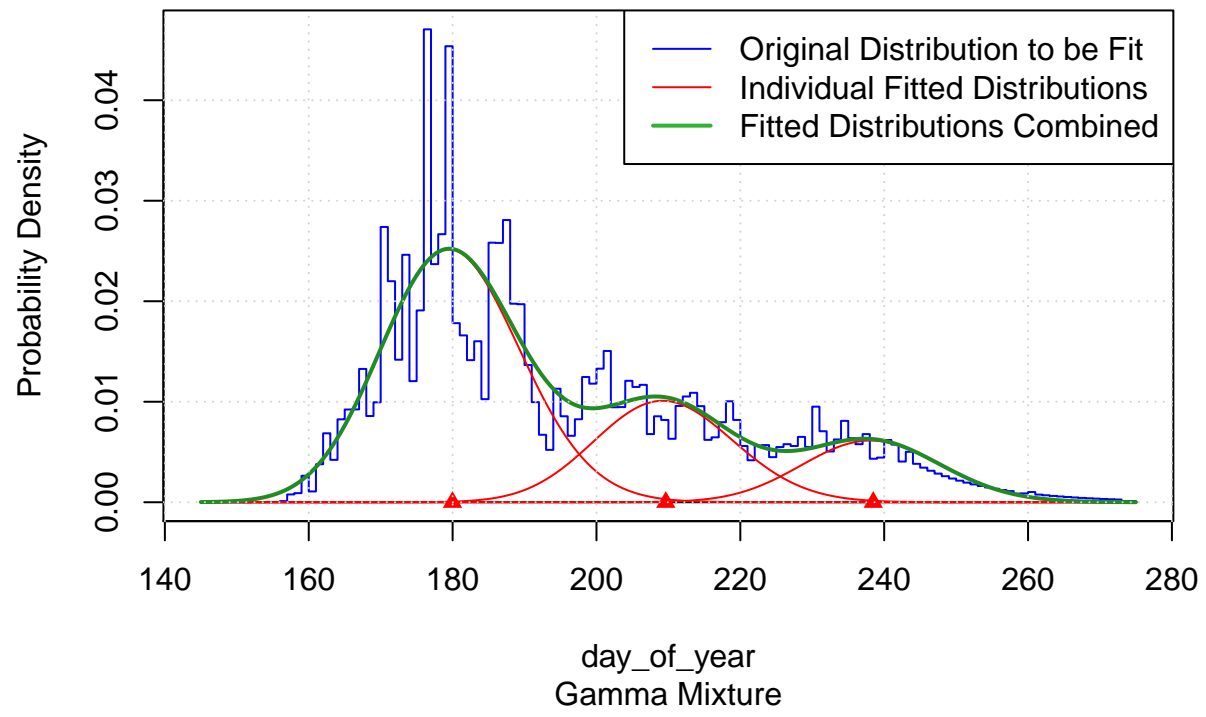


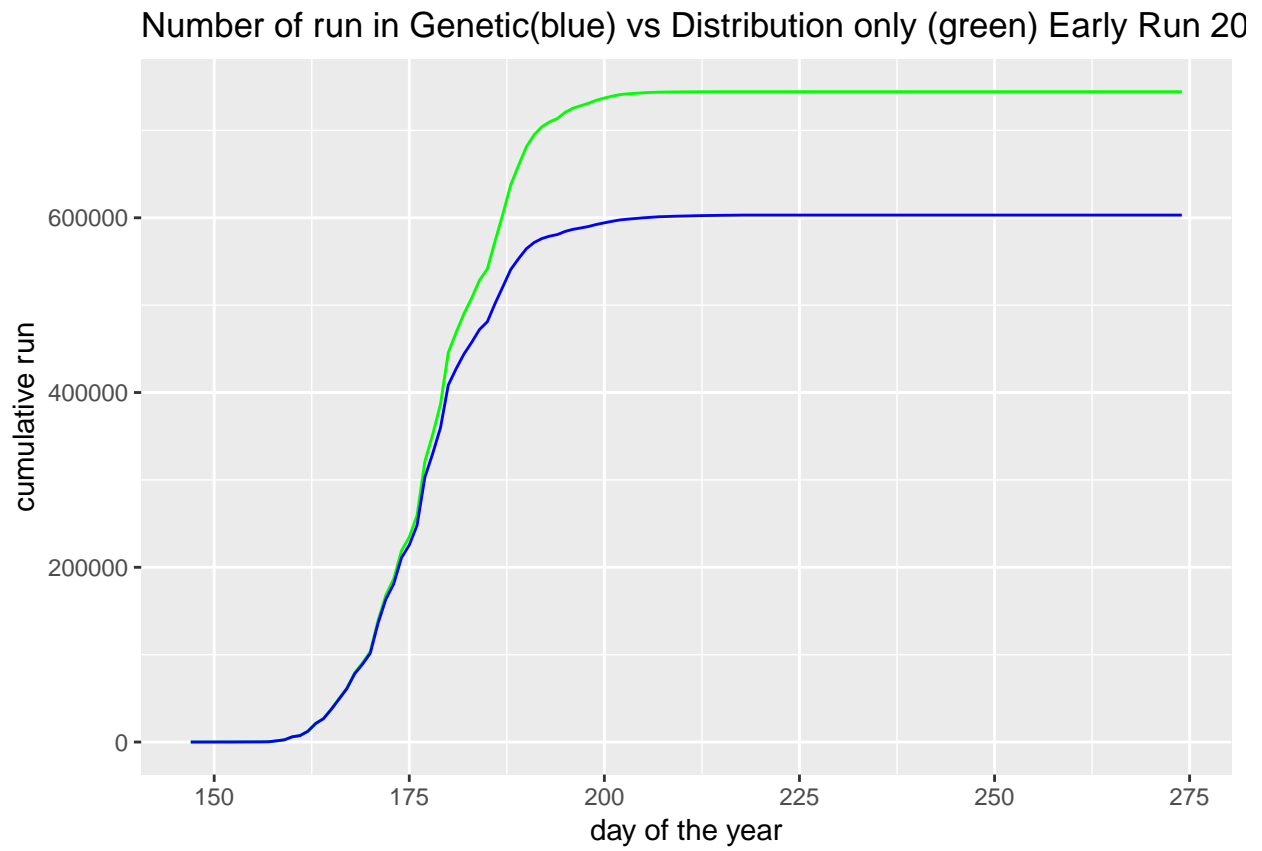
2007



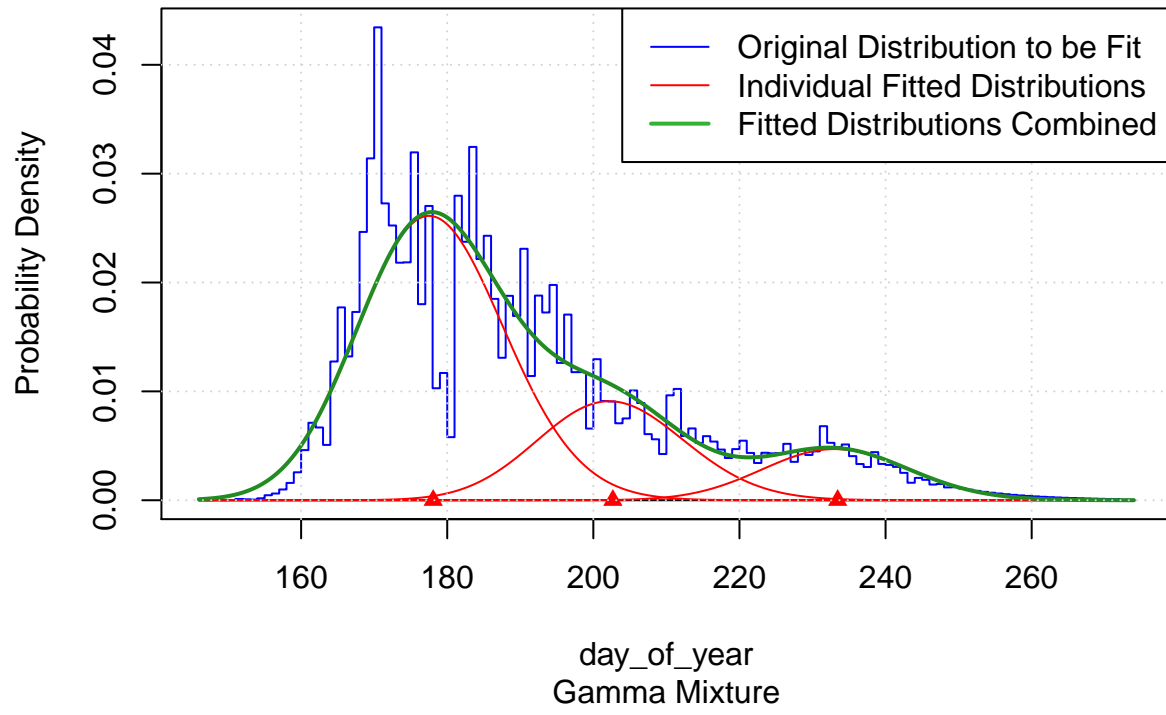


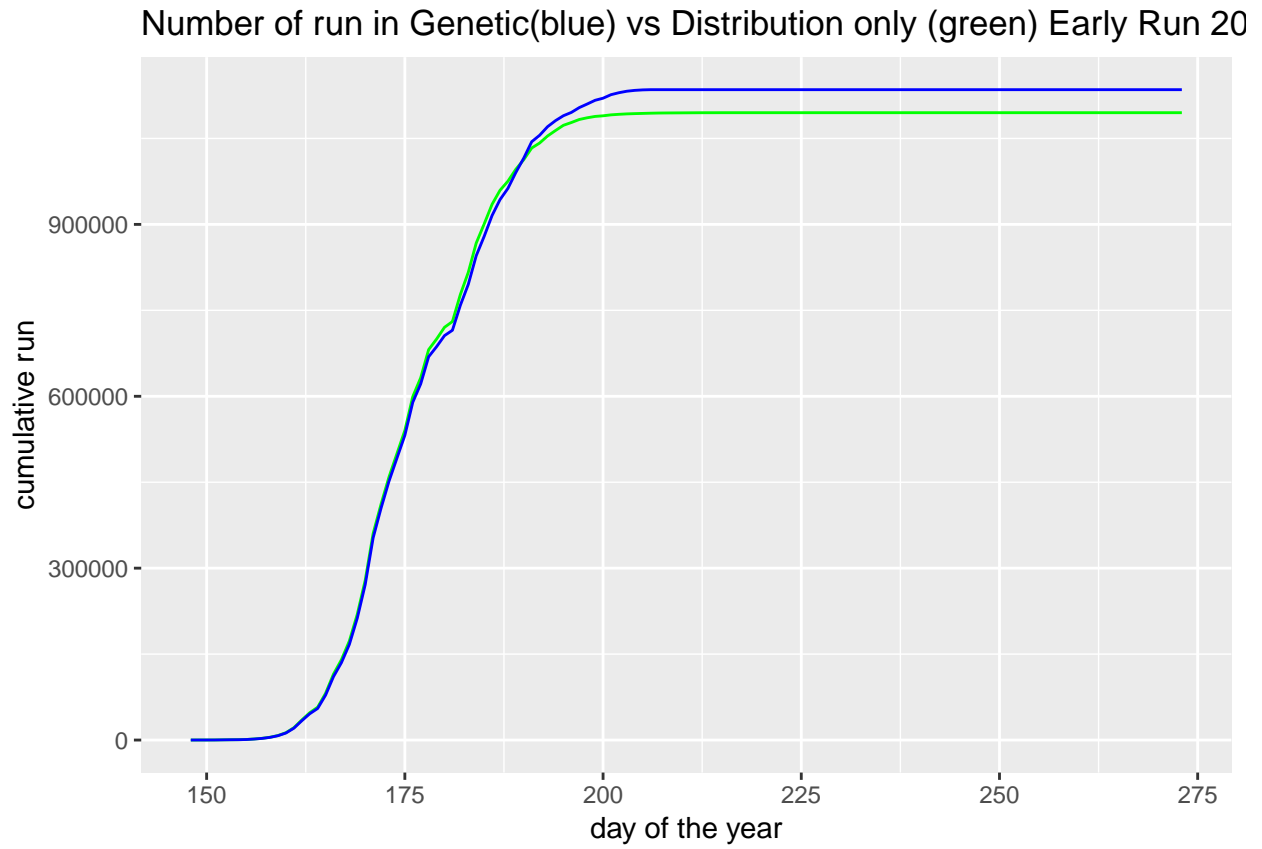
2008



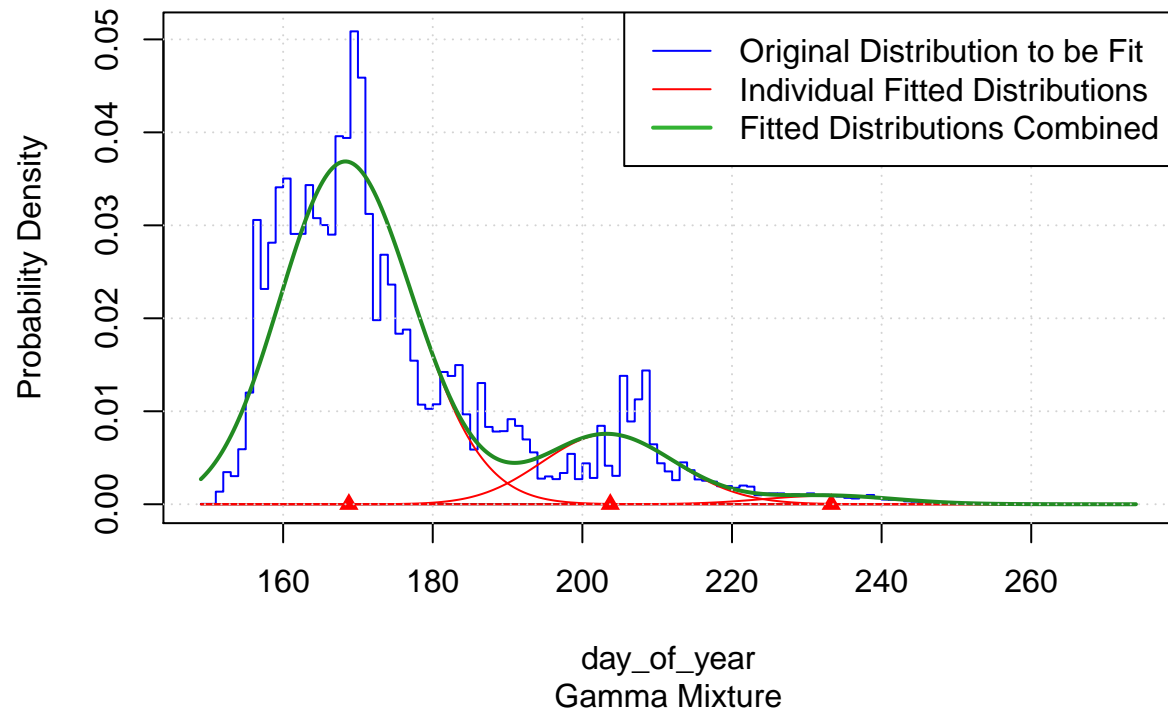


2010

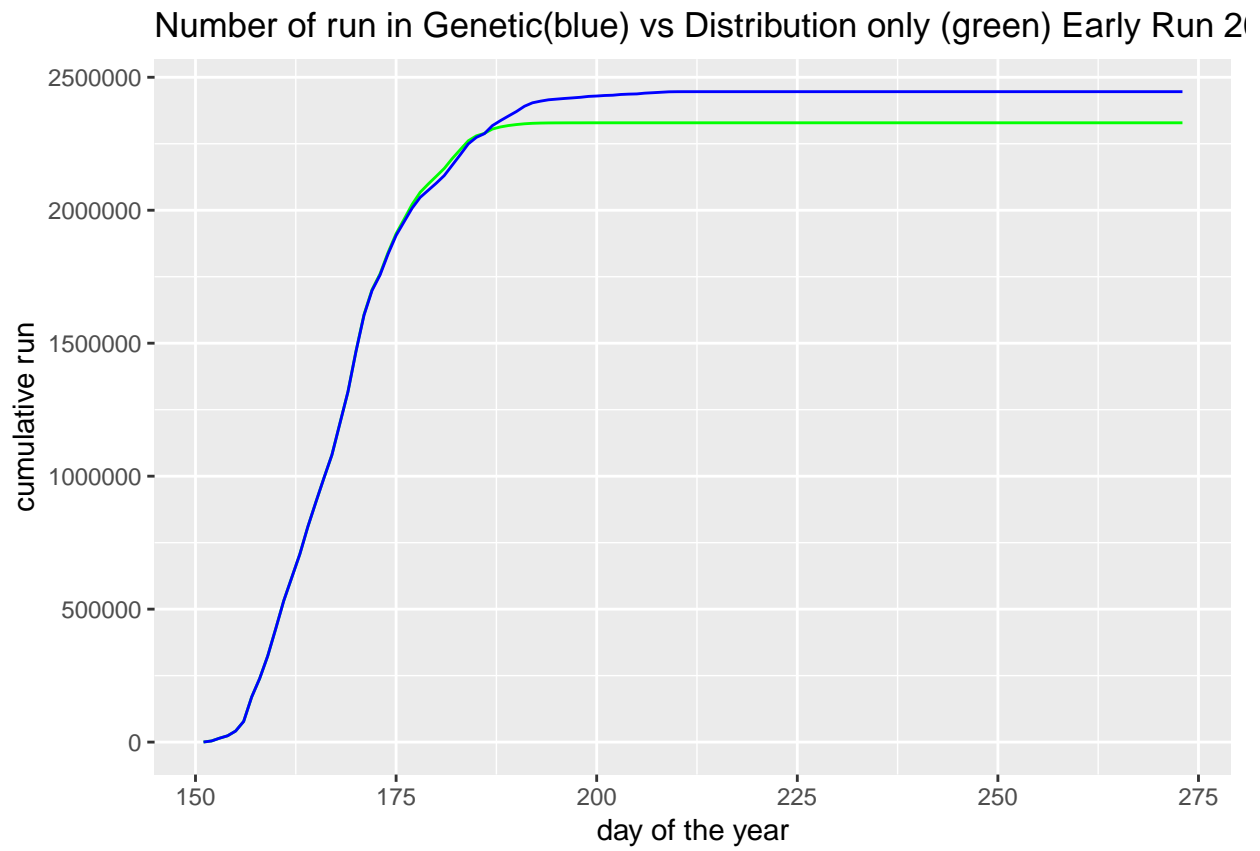




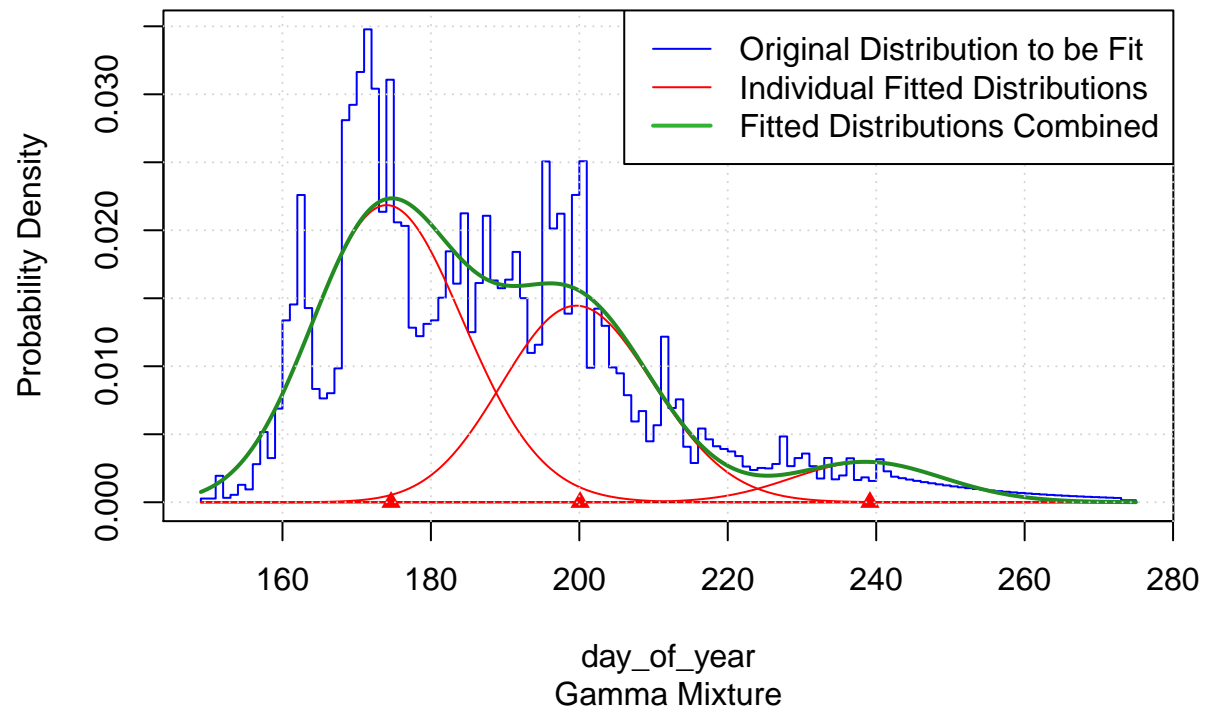
2011

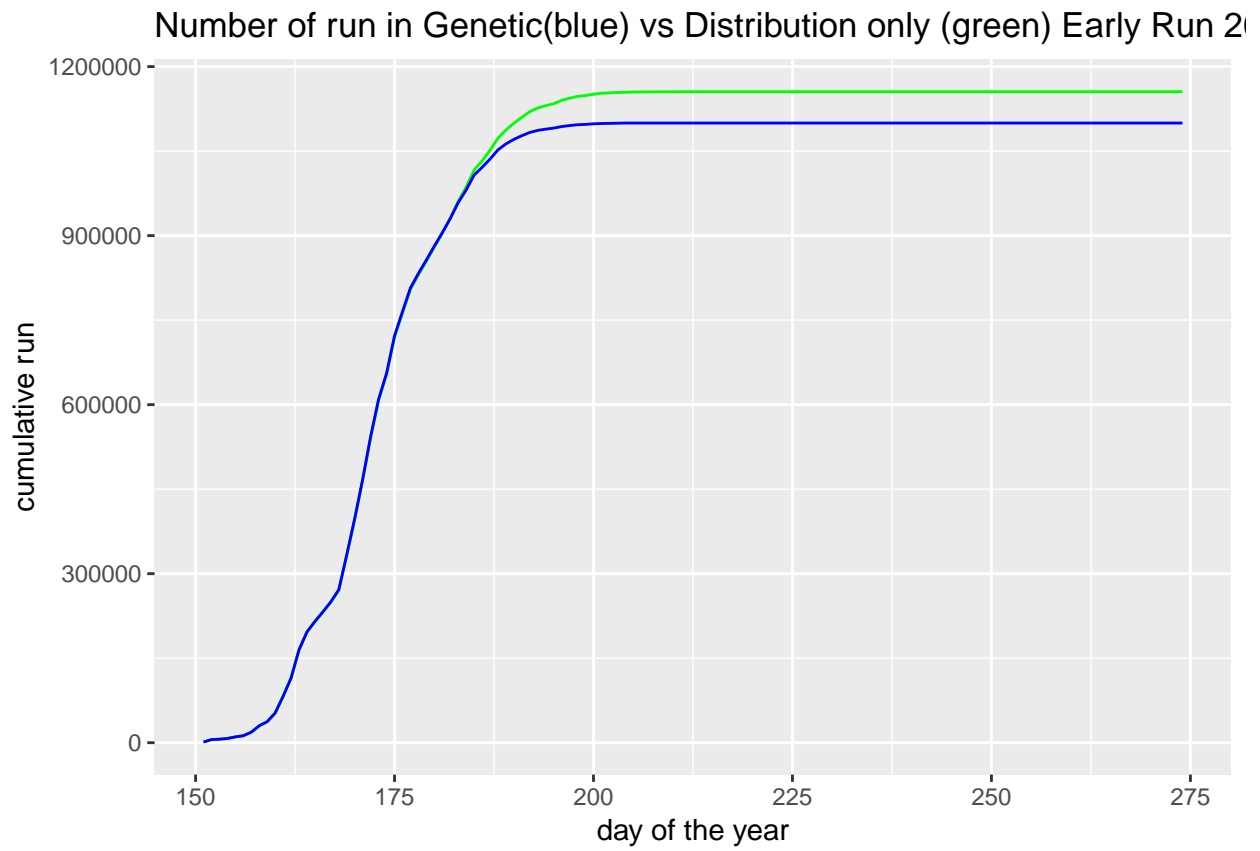




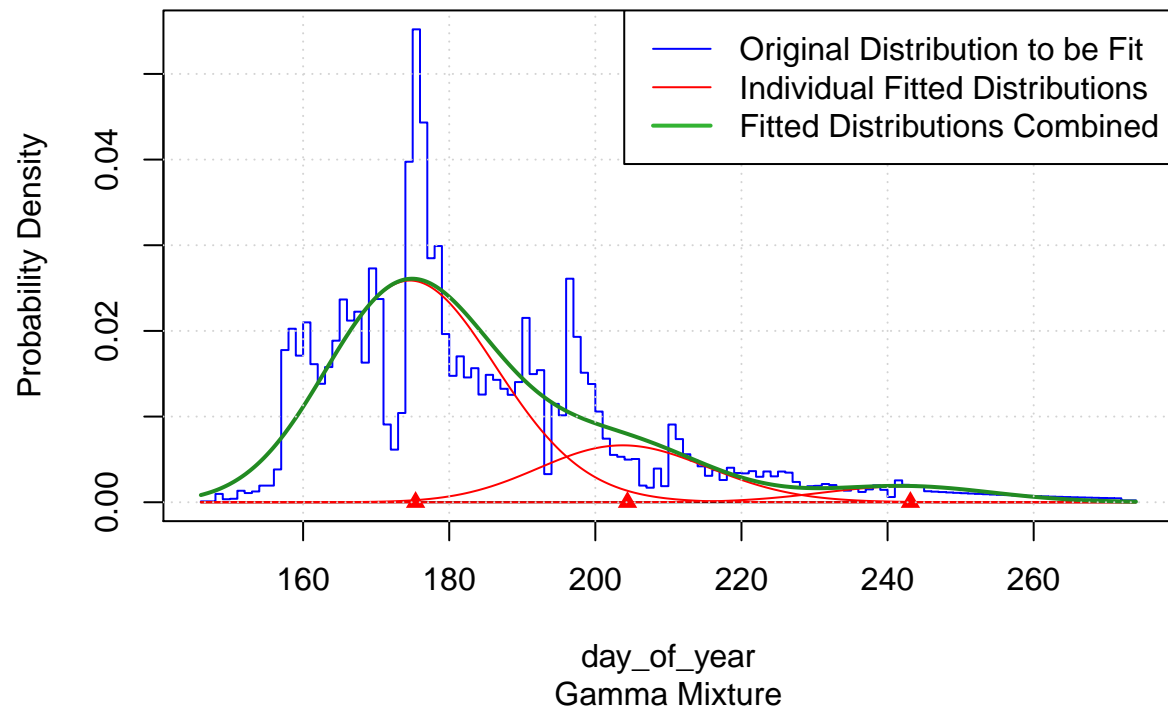


2012

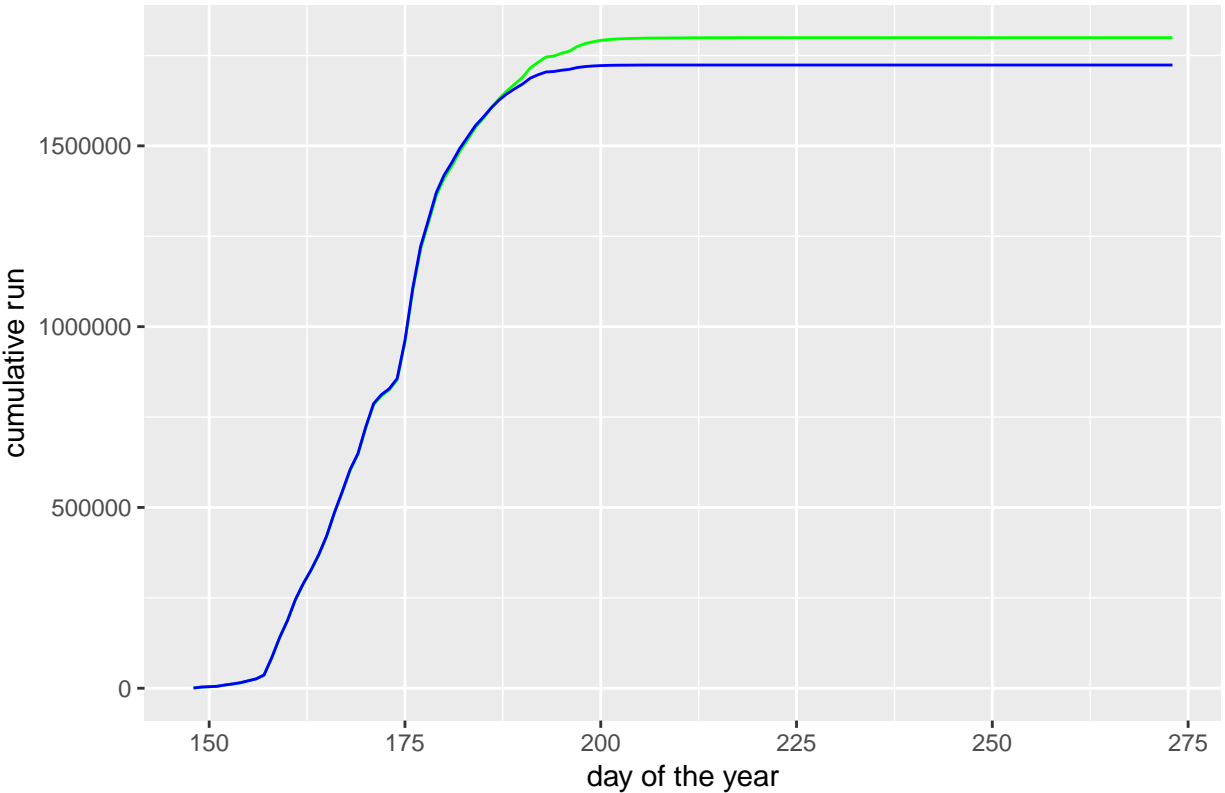




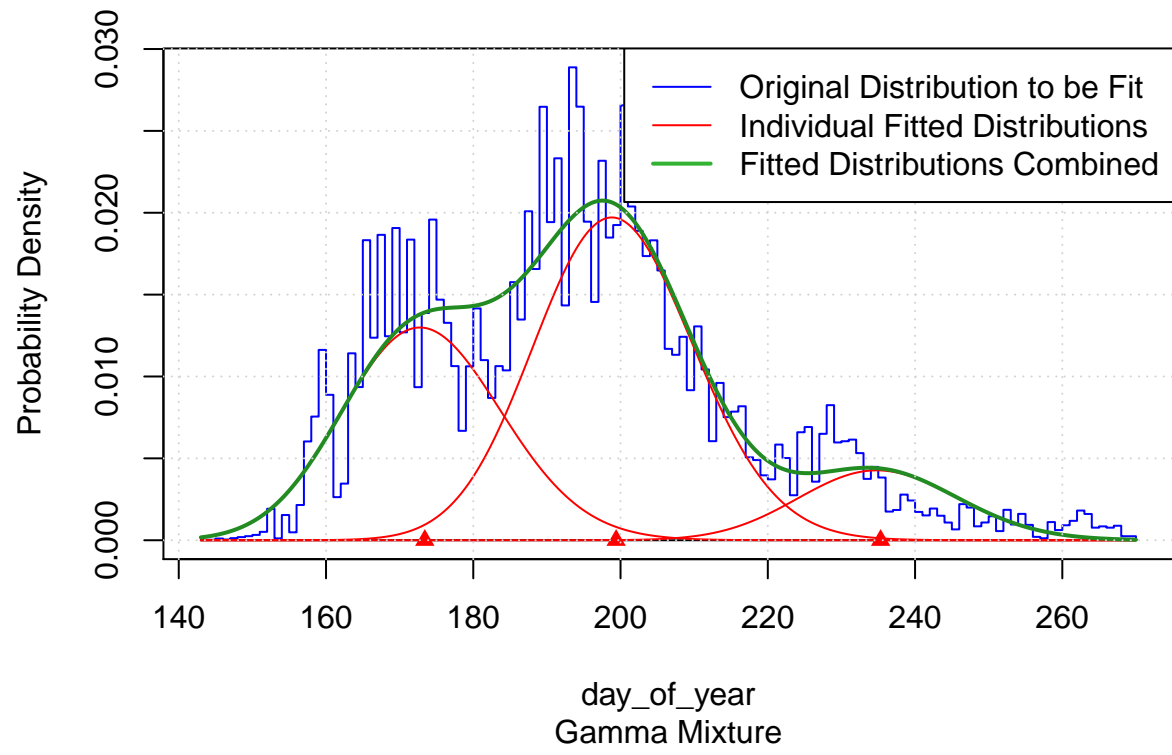
2013

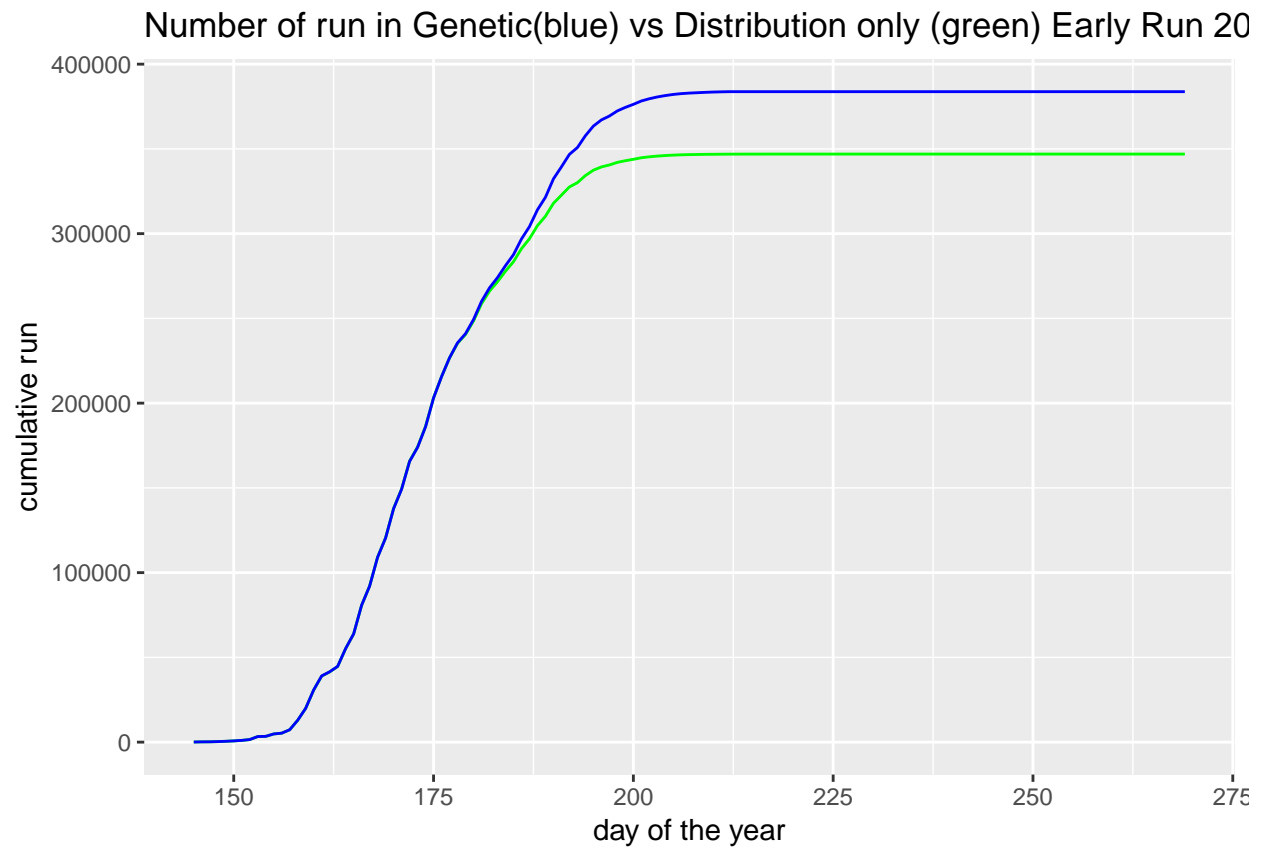


Number of run in Genetic(blue) vs Distribution only (green) Early Run 2

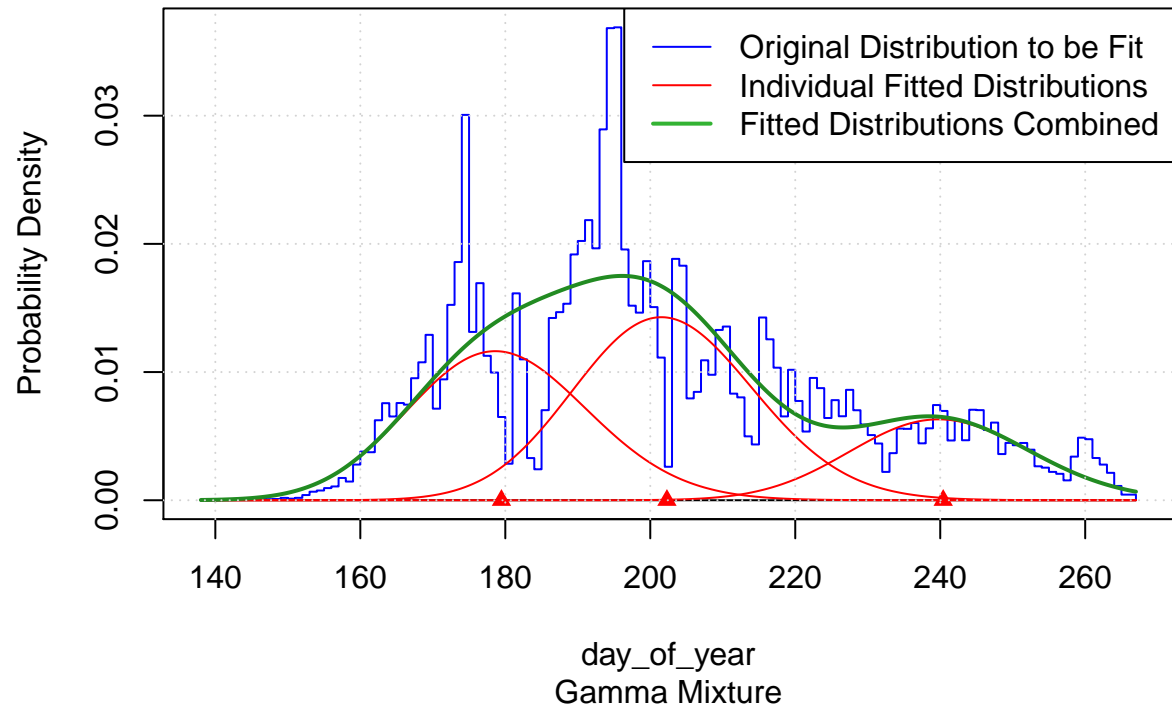


2014

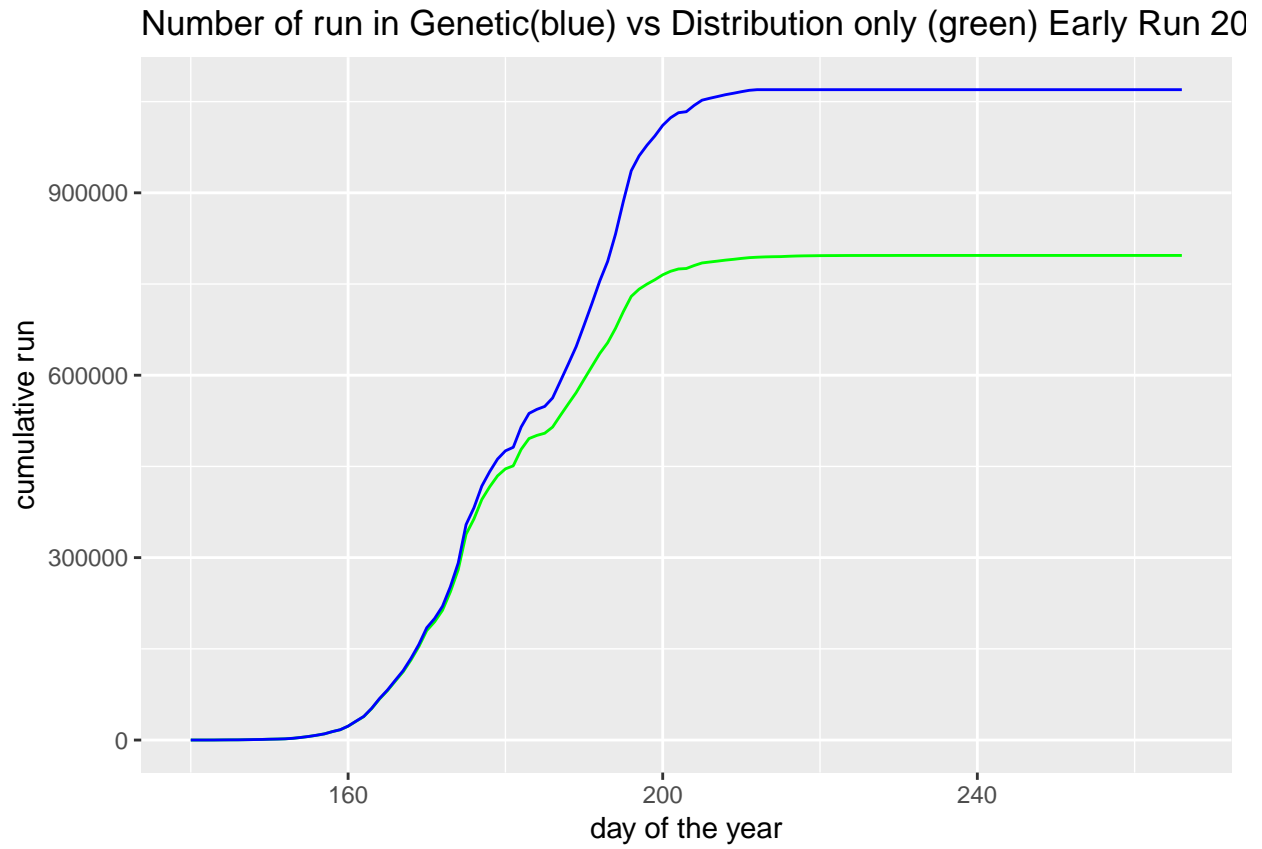




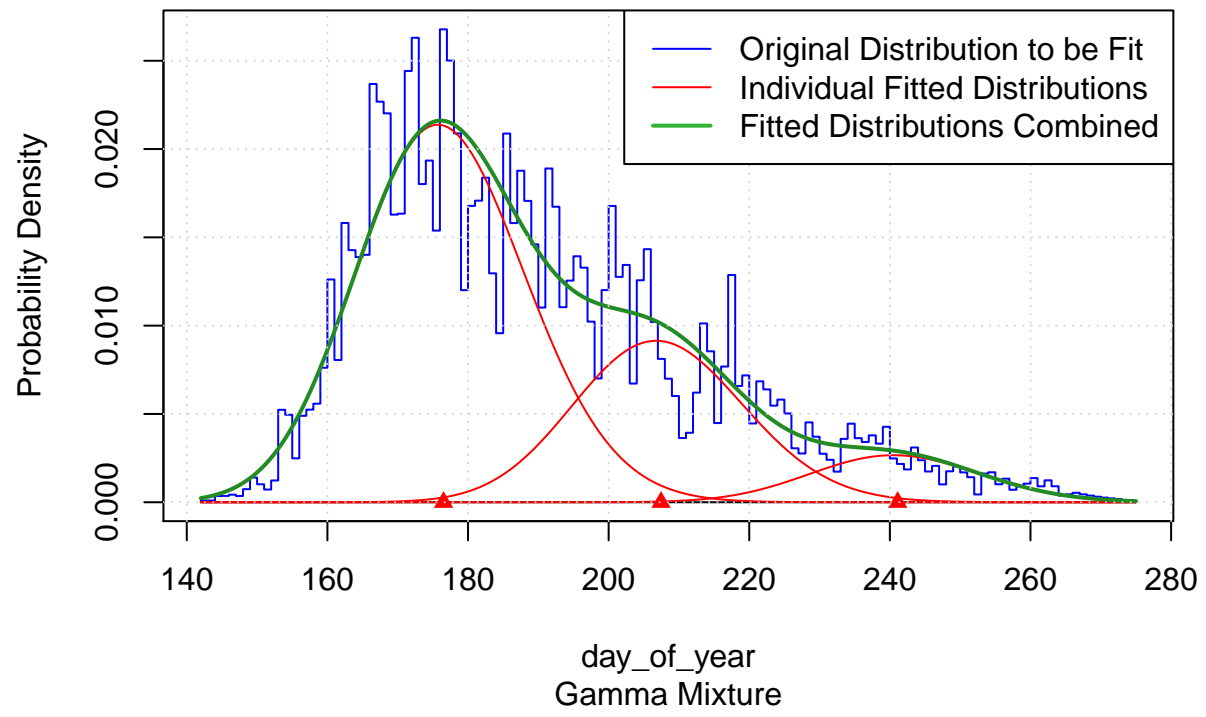
**2015**

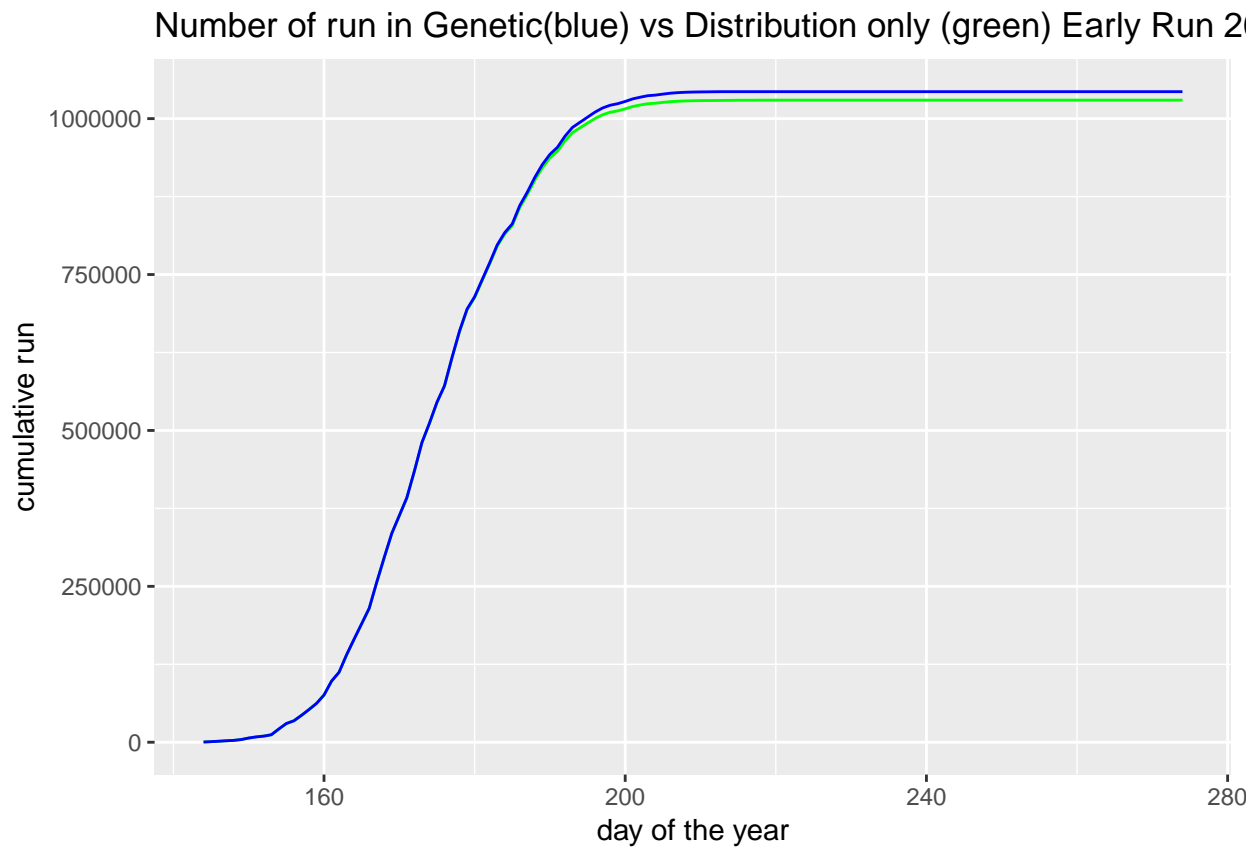




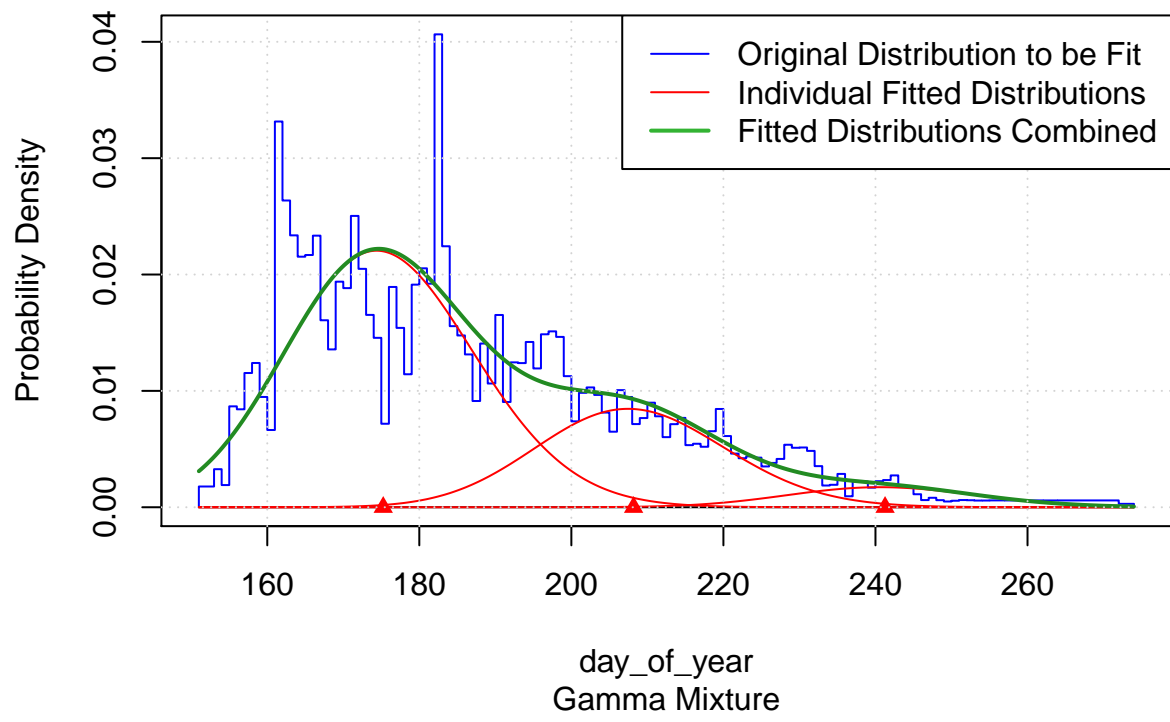


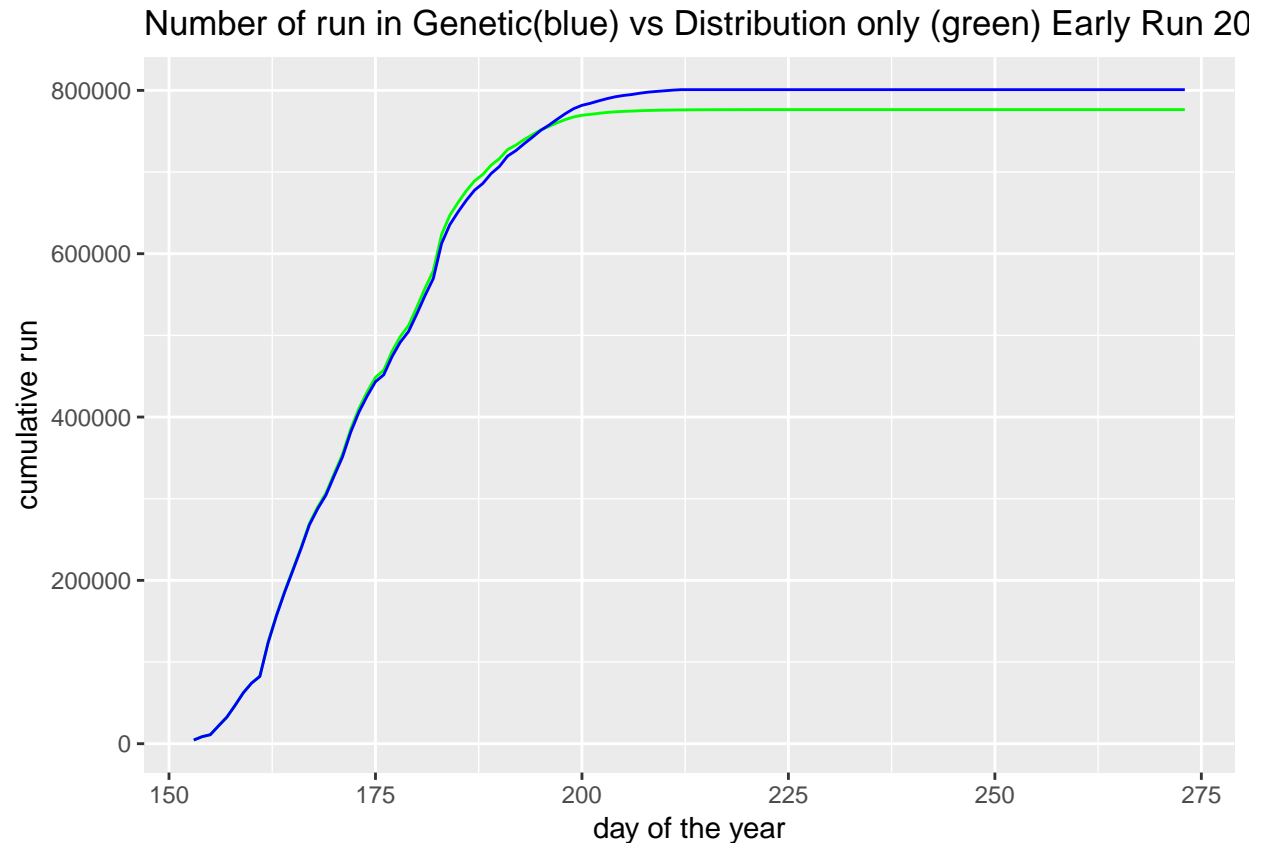
2016





2017





## 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: <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  lubridate_1.7.4 mixdist_0.5-5   forcats_0.3.0
## [5] stringr_1.3.1  dplyr_0.7.6     purrr_0.2.5     readr_1.1.1
## [9] tidyr_0.8.1    tibble_1.4.2    ggplot2_3.0.0   tidyverse_1.2.1
## [13] zoo_1.8-3
##
## loaded via a namespace (and not attached):
## [1] Rcpp_0.12.18    cellranger_1.1.0 compiler_3.5.1  pillar_1.3.0
## [5] plyr_1.8.4      bindr_0.1.1     tools_3.5.1     digest_0.6.15
## [9] jsonlite_1.5    evaluate_0.11   nlme_3.1-137    gtable_0.2.0
## [13] lattice_0.20-35 pkgconfig_2.0.1 rlang_0.2.1     cli_1.0.0
## [17] rstudioapi_0.7  yaml_2.2.0      haven_1.1.2     withr_2.1.2
## [21] xml2_1.2.0      httr_1.3.1      knitr_1.20      hms_0.4.2
## [25] rprojroot_1.3-2 grid_3.5.1      tidyselect_0.2.4 glue_1.3.0
## [29] R6_2.2.2        readxl_1.1.0    rmarkdown_1.10  modelr_0.1.2
## [33] magrittr_1.5    backports_1.1.2 scales_0.5.0    htmltools_0.3.6
## [37] rvest_0.3.2     assertthat_0.2.0 colorspace_1.3-2 labeling_0.3
## [41] stringi_1.1.7   lazyeval_0.2.1  munsell_0.5.0   broom_0.5.0
## [45] crayon_1.3.4
```