Simulation Update

Michael W Belitz

July 23, 2019

Unimodal Distribution

Below are the simulated distributions of three hypothetical species.

The distributions all have a mean of 200. The standard deviations are 10 [Fig. 1], 20 [Fig. 2], and 40 [Fig. 3]

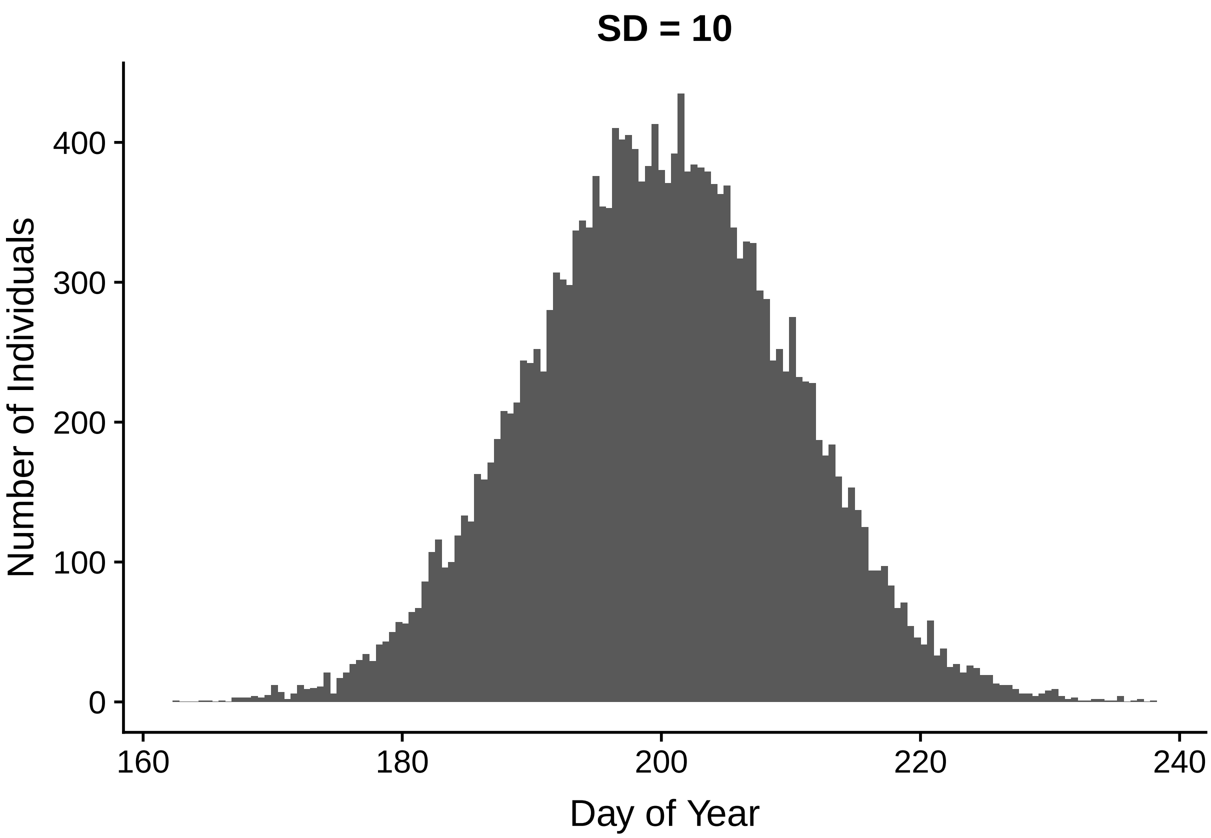


Fig. 1 Distribution of number of adult individual sof a species that has a single peak and the most narrow flight length, although it is still 76 days from onset to offset (onset = 162; 10% = 186.99; 50% = 199.89; 90% = 212.76; max = 238.1).

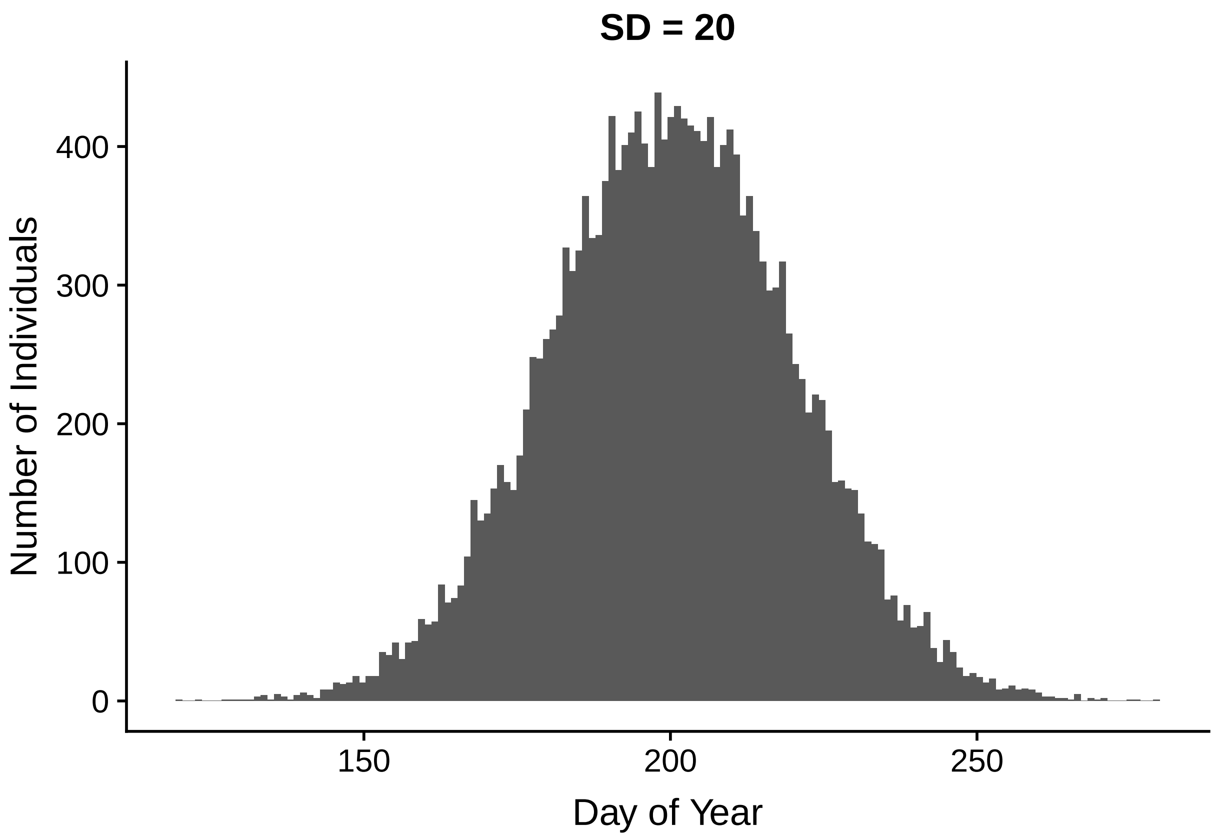


Fig. 2 Distribution of number of adult individuals of a species that has a single peak and has the middle flight length. In total the flight length is 159 days (onset = 119.68; 10% = 174.82; 50% = 200.19; 90% = 225.60; max = 279.35).

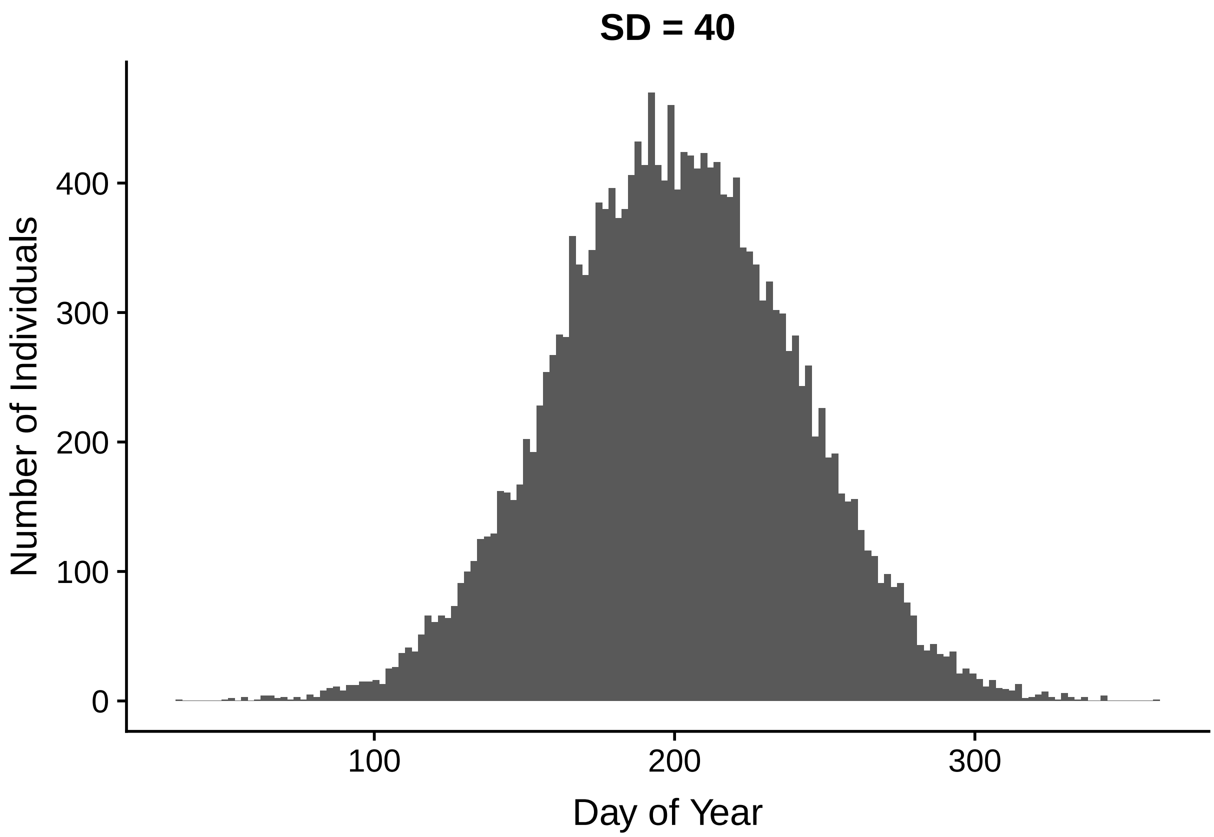


Fig. 3 Distribution of number of adult individuals of a species that has a single peak and the longest flight length of the three single peak distributions we are examing. The flight length lasts almost the entiire year (onset = 35.88112, 10% = 149.30459; 50% = 199.80307; 90% = 251.79269; offset = 361.37887).

Results of unimodal simulation

Onset

For onset and offset estimates, I compared estimates derived from my biased-corrected weibull estimator to Will Pearse's R package. My estimator is similar in the MLE estimate but has smaller CIs. The Pearse estimator is much faster than my estimator because it has a analytical solution, which is its big advantage. However, we show that onset and offset may not be the pheno-metrics that can be most reliably estimated.

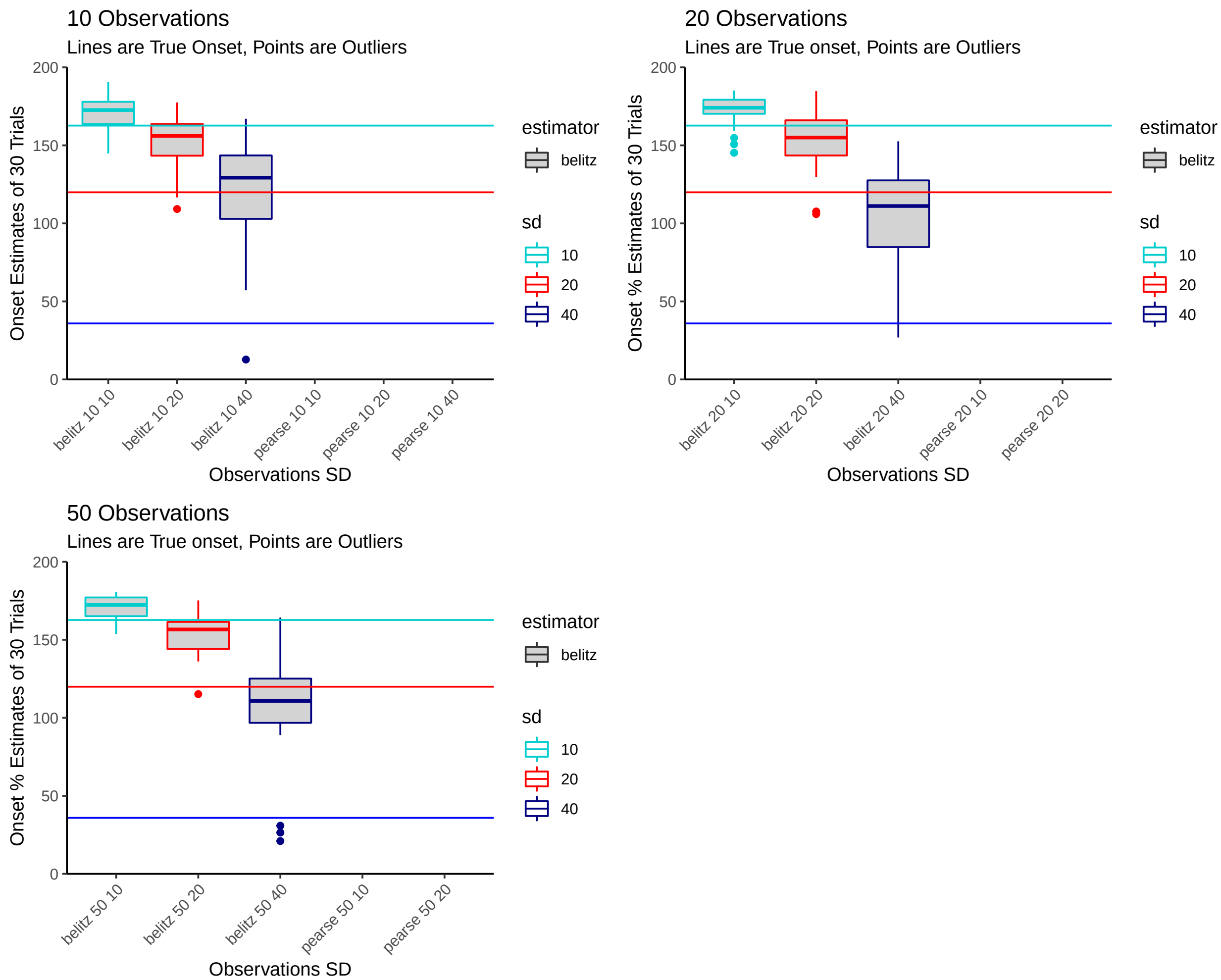
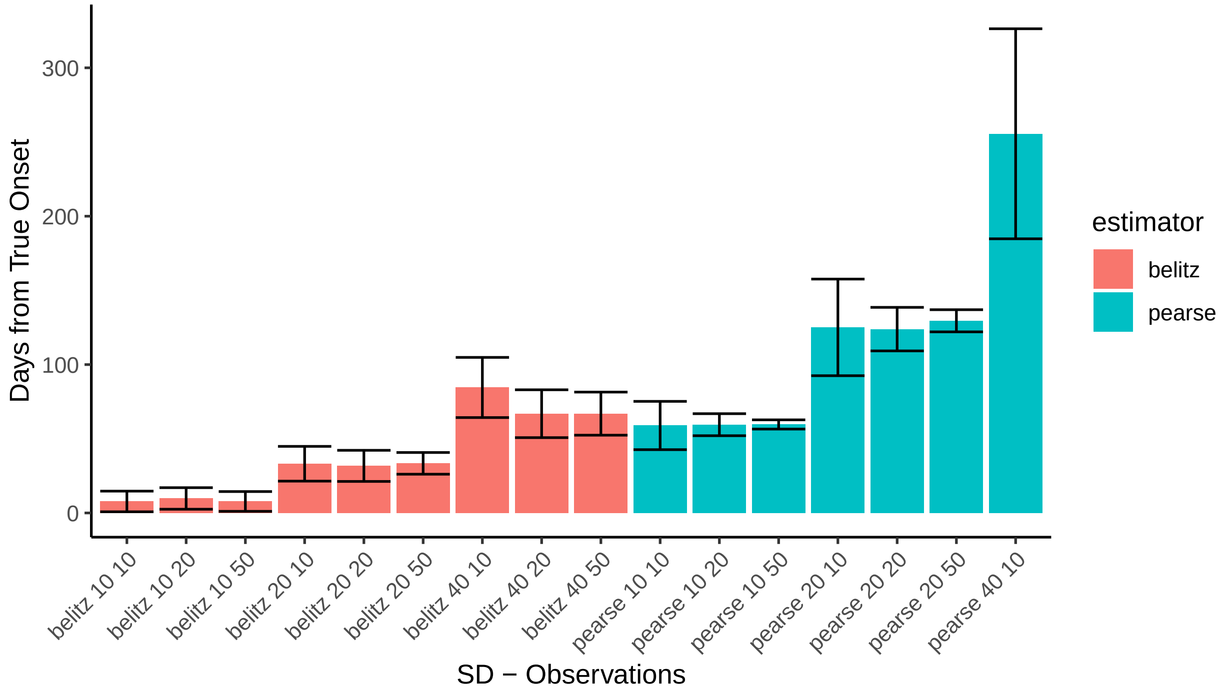


Fig. 4 Box plots showing the estimates of 30 random observation iterations per observation-sd combination. The line within the boxplot is median estimate of the 30 iterations, and the ends of the box show the upper and lower quartiles. The extreme lines show the highest and lowest estimates, excluding outliers which are displayed with points. The true onset for each distribution (10 SD, 20 SD, and 40 SD) is represented by the colored horizontal lines. In general, the estimators over-estimated the onset. The smaller the SD in the distribution, the closer the estimates are to the true onset. Increasing the number of observations marginally improves the estimates.

I calculated "days from true onset" by taking the absolute value of the onset estimate and subtracting by the true onset. Although both estimators perform similar in how many days away they estimate from the "true onset", Pearse's CI include the "true onset" more often because of the much larger CI.



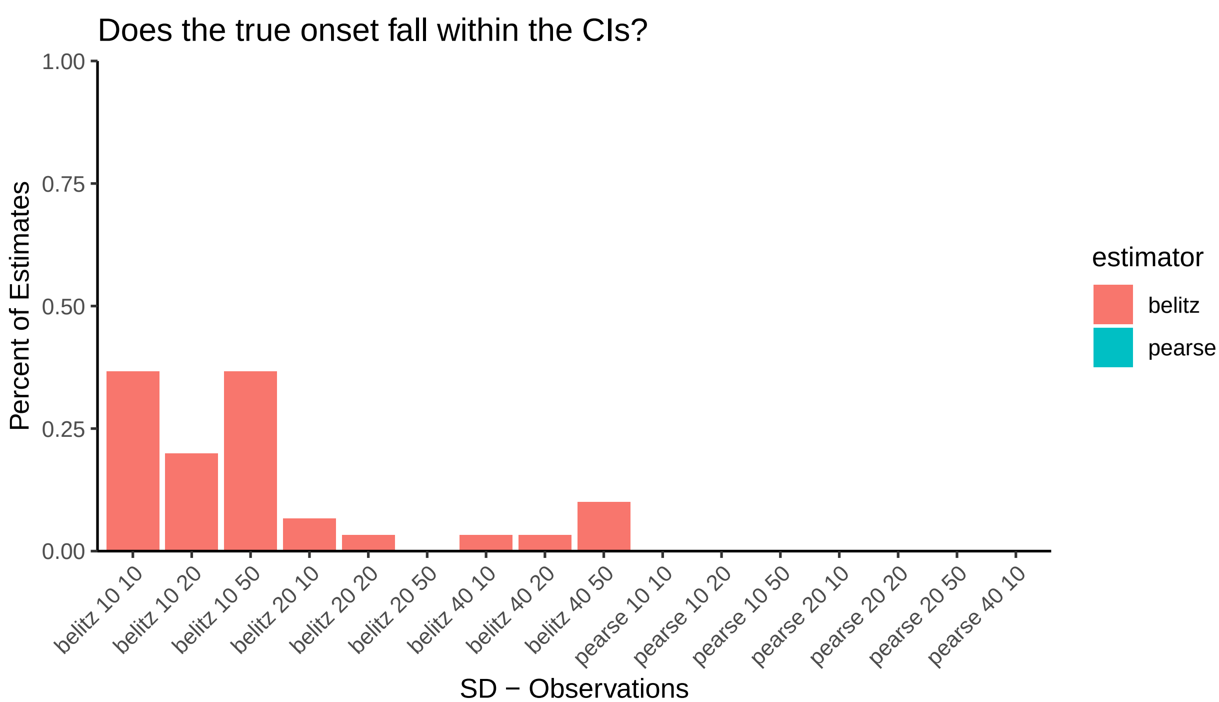


Fig. 5 Average calculated days from true onset metric and average 95% confidence interval shown for the 30 iterations representing each Observation-SD combination. Increasing the SD increased the days from true onset and increasing the number of observations decreased the days from the true onset. However, increasing the number of observations actually decreased the percent of estimates that fall within the CIs because the CI decrease with the increasing number of observations.

Tenth

My estimator does better at estimatin the 10 percentile of the distribution than estimating onset (min).

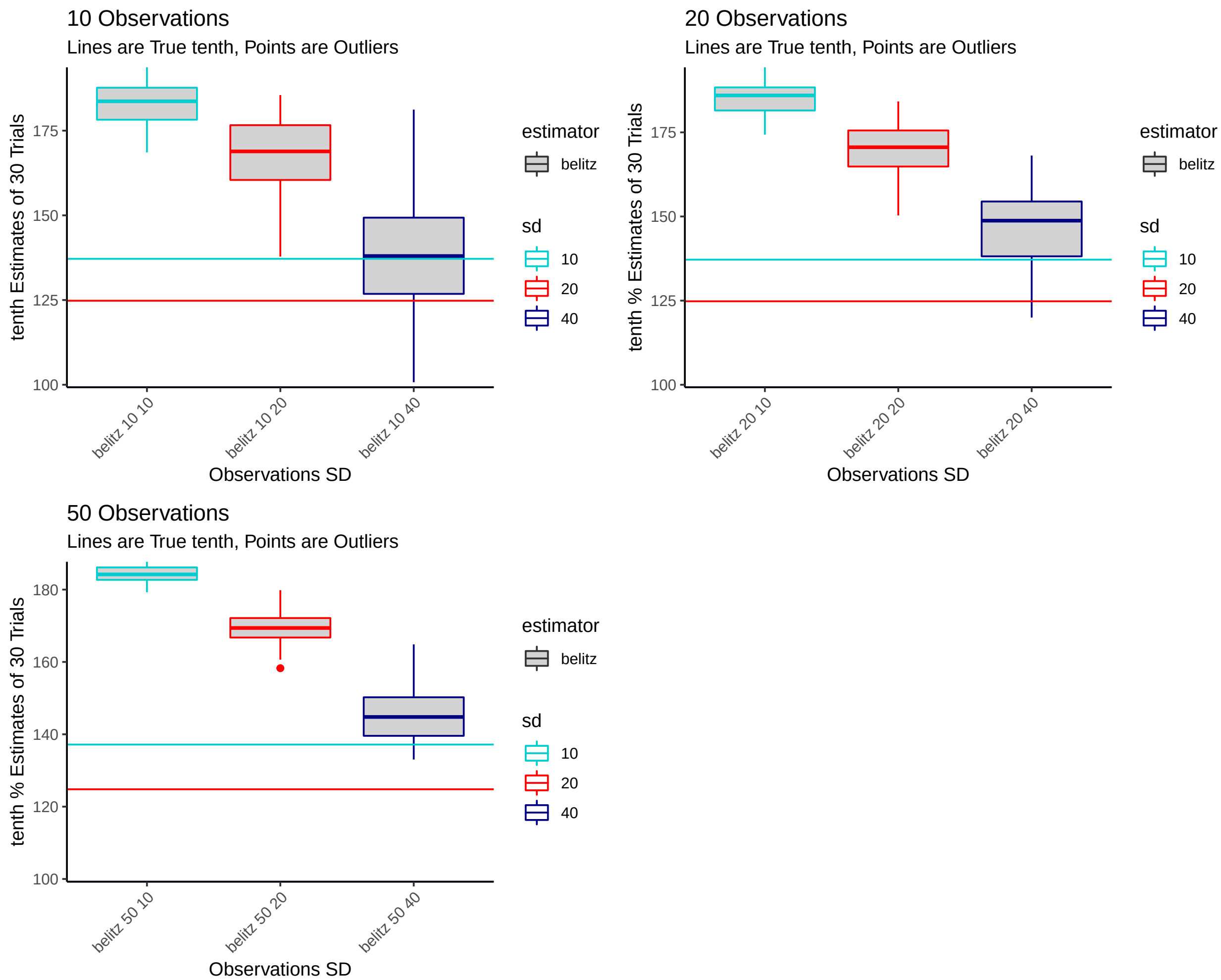


Fig. 6 The estimates of the 10th percentile for 10 SD are around the true tenth percentile. Increasing SD decreases the accurracy of the estimates. Increasing observations slightly increases the precision of the estimates.

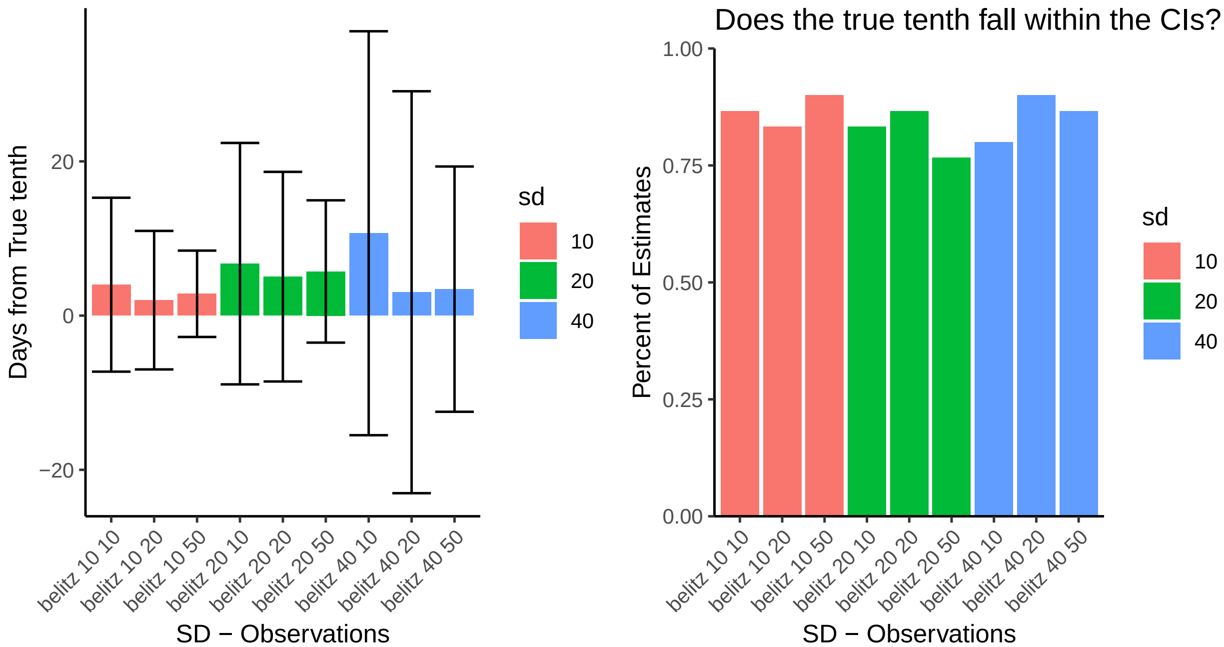


Fig.7 On average, even with 40 SD, the estimate is within 10 days from the true tenth. However, at 10 obs and 40 SD, the estimate can be 40 days off. The CI of the 10 percentile estimates does capture the true tenth much frequently than the onset estimates.

Median

Median Estimates are even better.

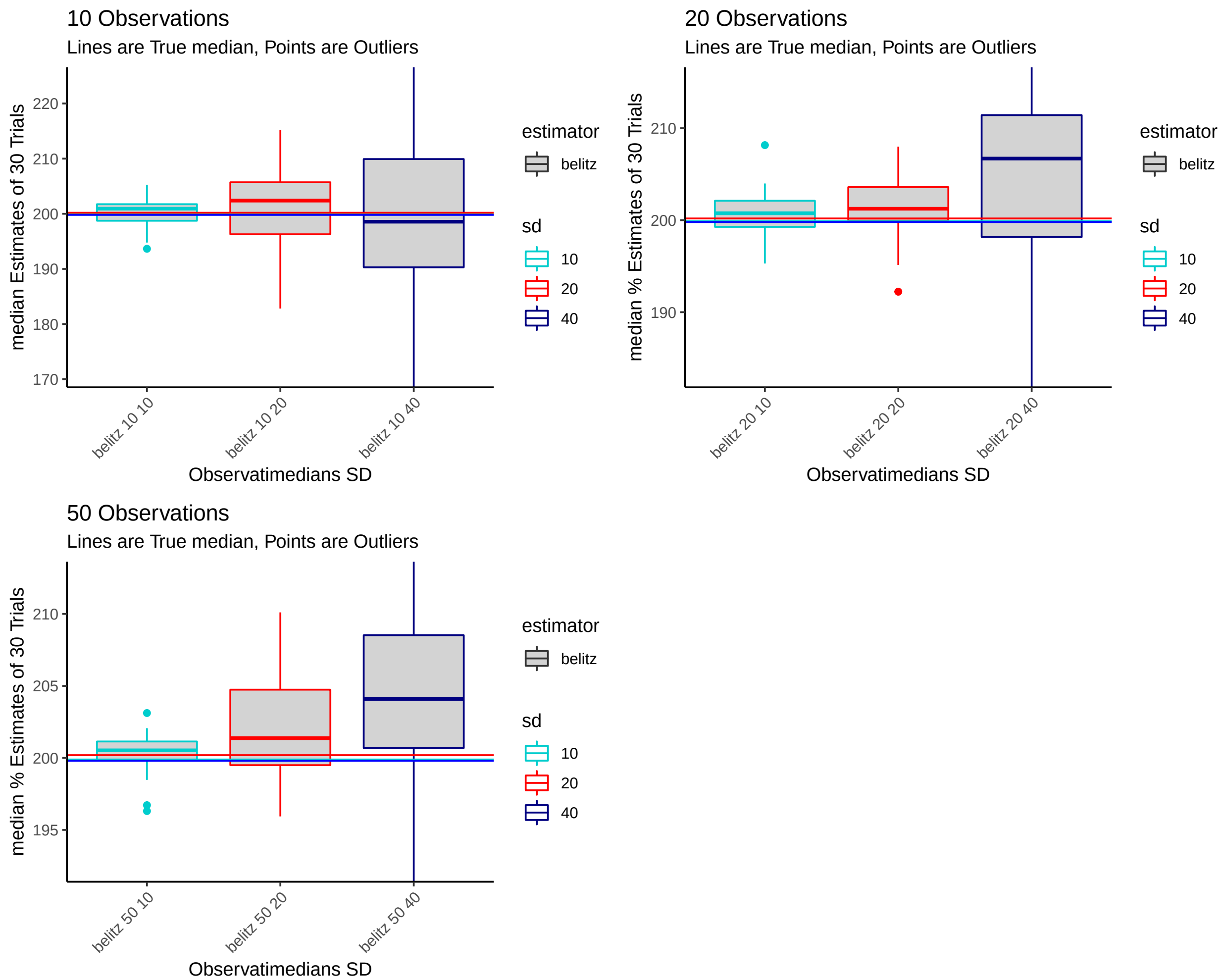


Fig.8 True median is ~200 for all three distributions. Similar trends to the other boxplots.

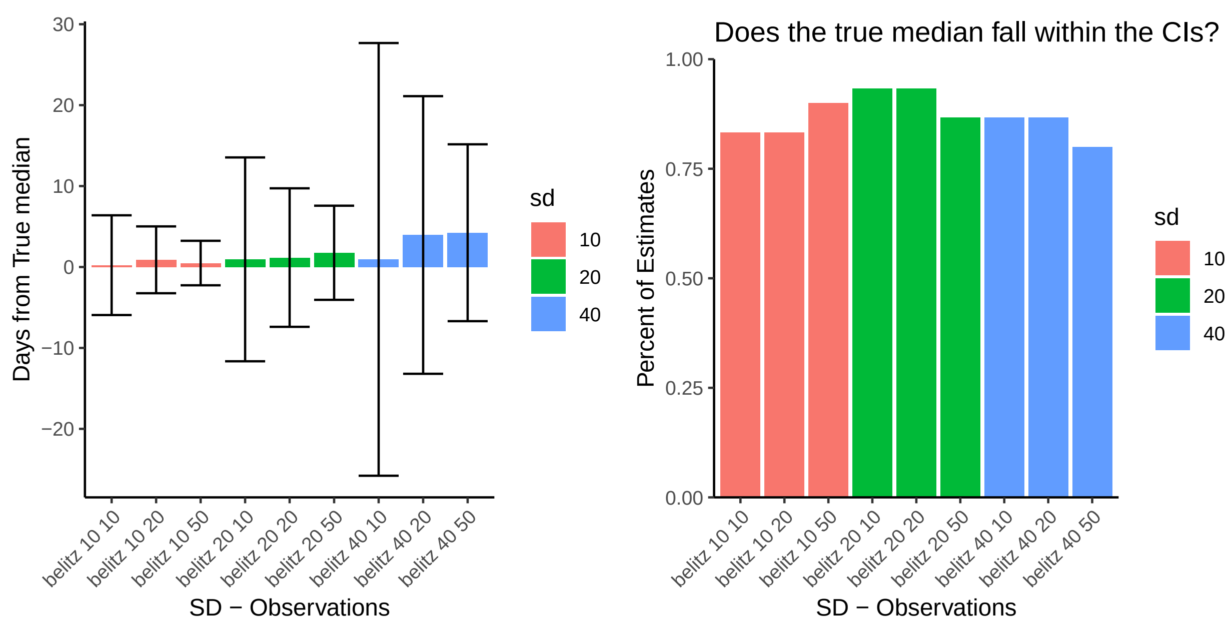


Fig. 9 Similar trend as 10 percentile estimates with slightly improved estimates.

Ninety

Ninety was performing even better than 10% which, I thought they would be the same. This made me dig back to the code that I wrote to write the estimator function and realized a small error that could lead to the estimator being slightly biased towards later DOY estimates. So I fixed the error. Used a different estimator function for the bimodal distributions.

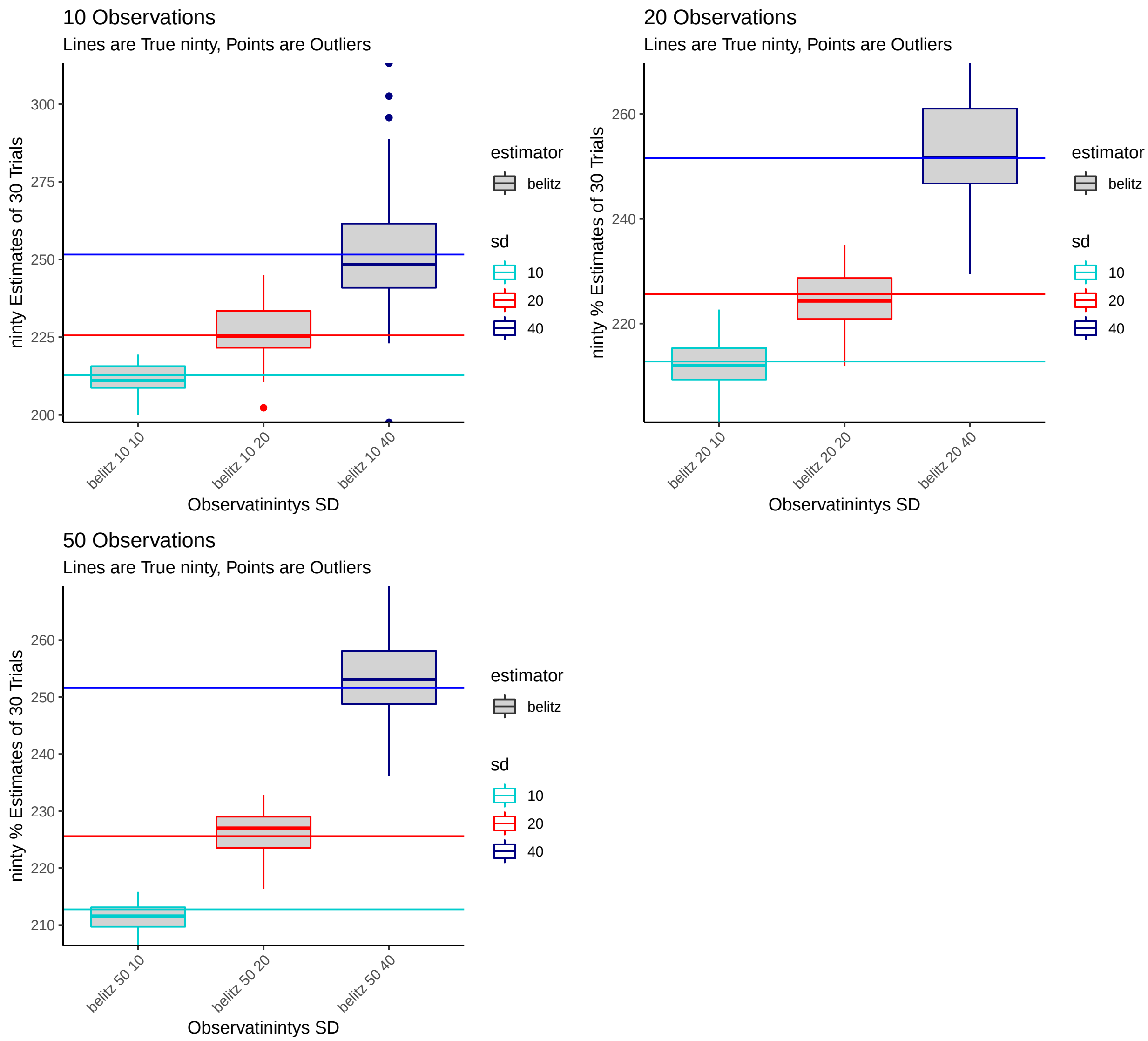


Fig. 10 Estimates of 90 percentile compared to true 90th percentile of distributions.

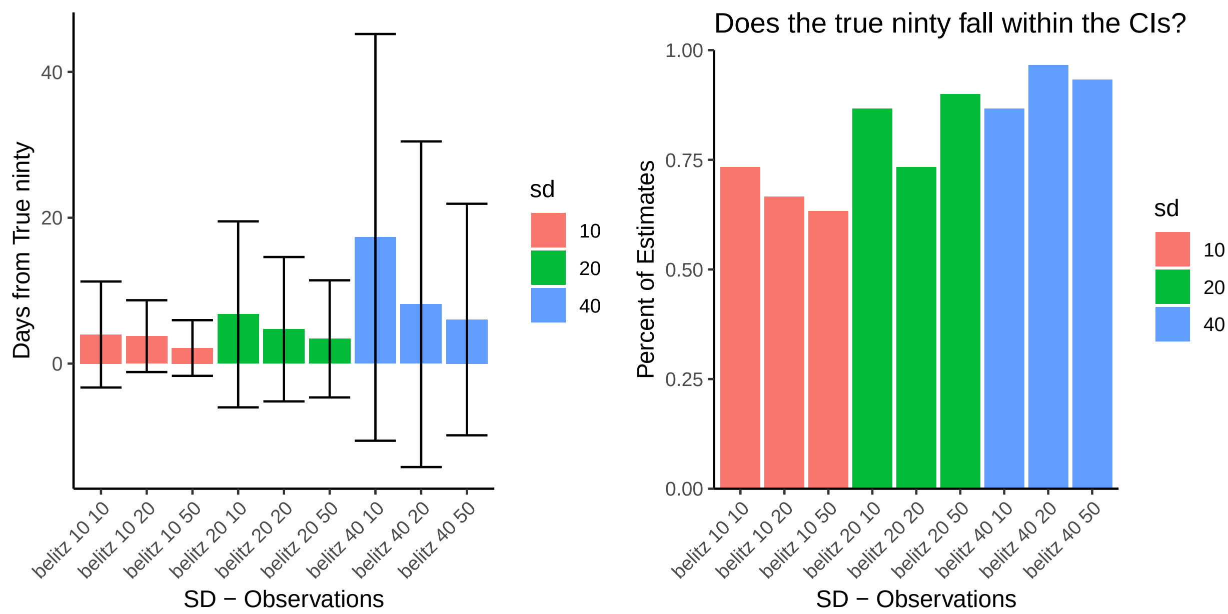


Fig. 11 More prominent display of increasing precision with increasing observations.

Offset

Offset again proves more challenging to estimate than 10, 50, or 90 percentiles. Also, my estimator outperformed Pearse's, which also helped convince me something was wrong in my function code.

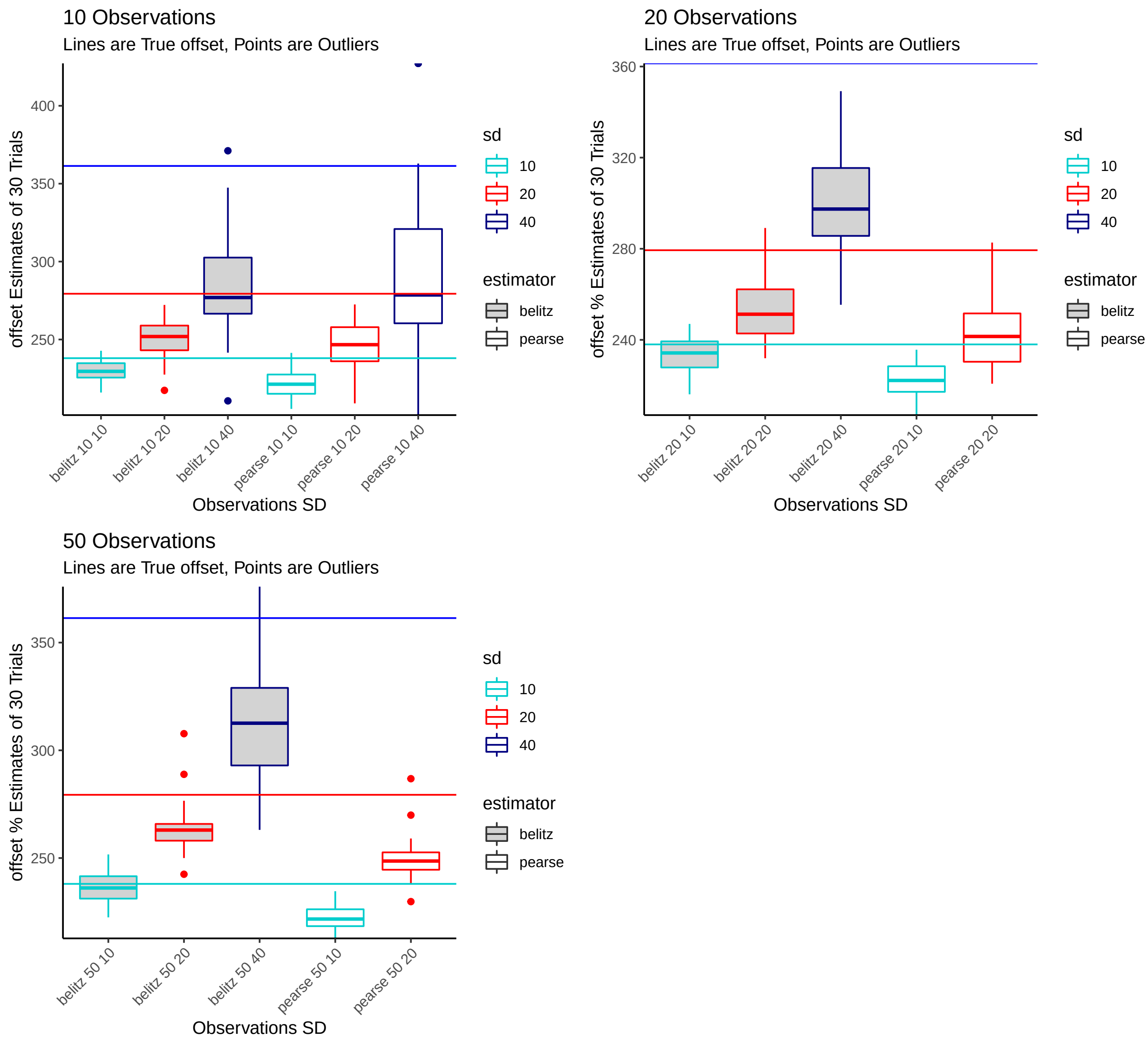
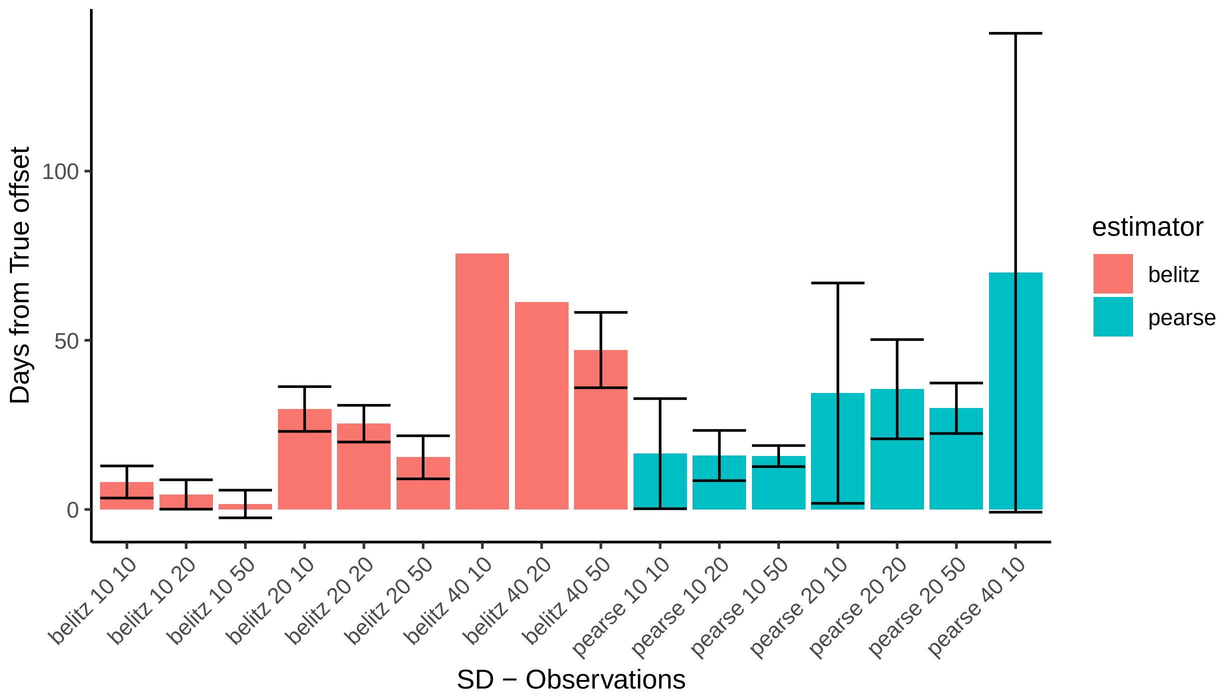


Fig. 12 Offset estimates



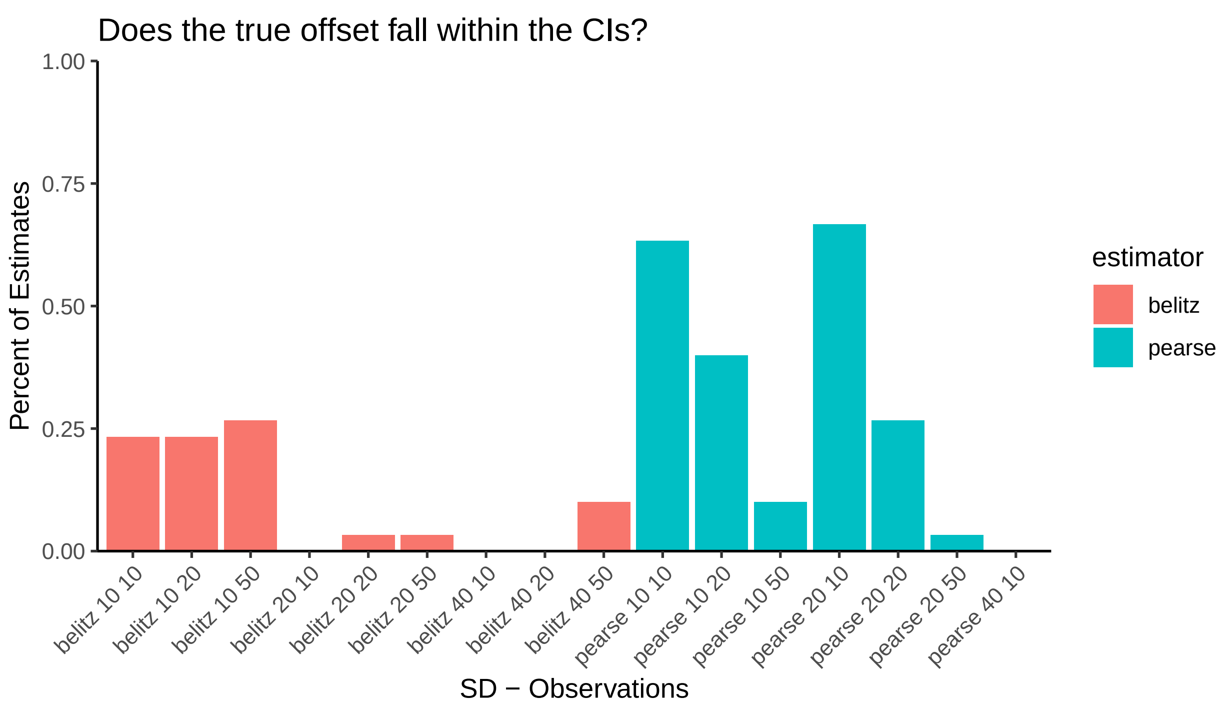


Fig. 13 Notice how an increase in observations decreases days from true onset (in most cases), but decreases the percent of estimates that capture the true offset within the CIs.

Bimodal Distribution

Below are the simulated distributions of two hypothetical species.

The first represents a species that is bi-voltine with 2/3 of its population density occurring in the second flight peak and overall a narrow flight length per brood that has little overlap (Mean 150 and 220, 10 SD).

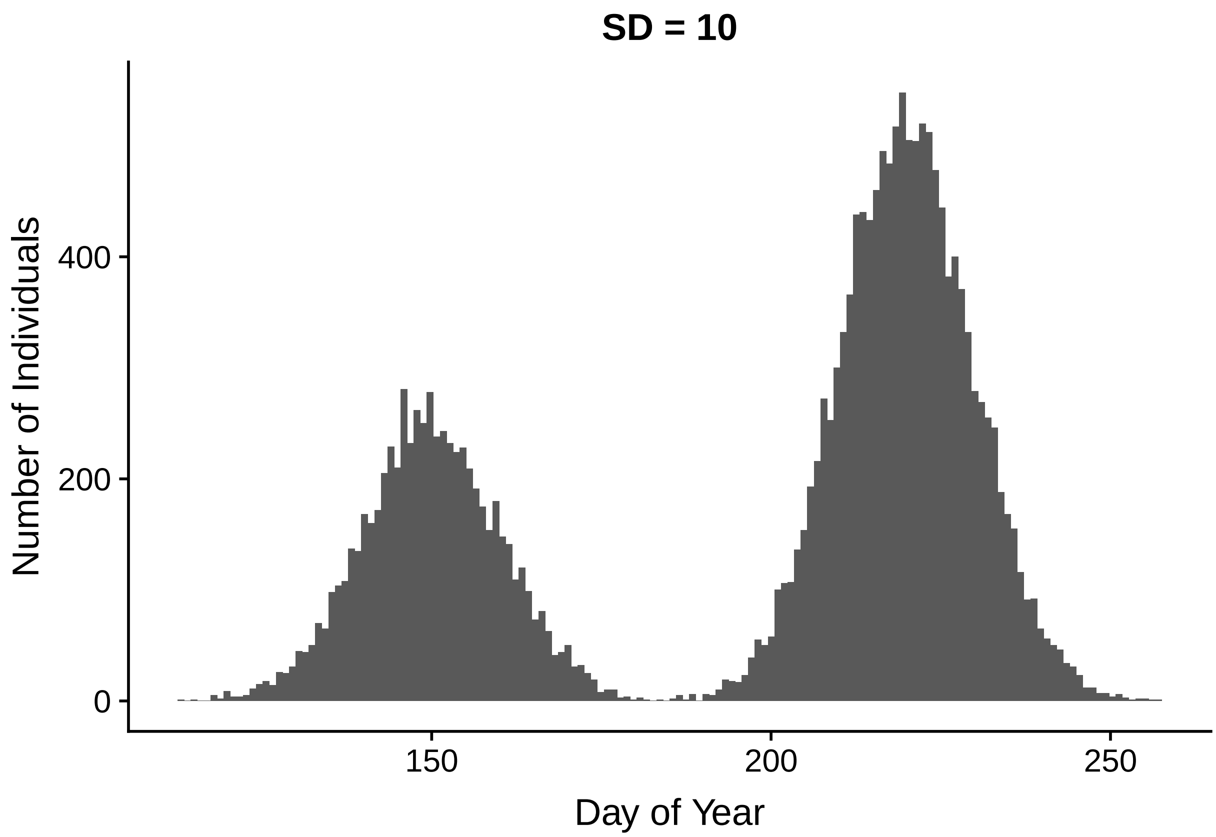


Fig. 14 Onset = 113.29, 10% = 144.65, 50% = 213.21, 90% = 230.27, 100% = 257.28

The next represents a species with a longer flight length, so more overlap between the two broods (Mean 150 and 220, 20 SD).

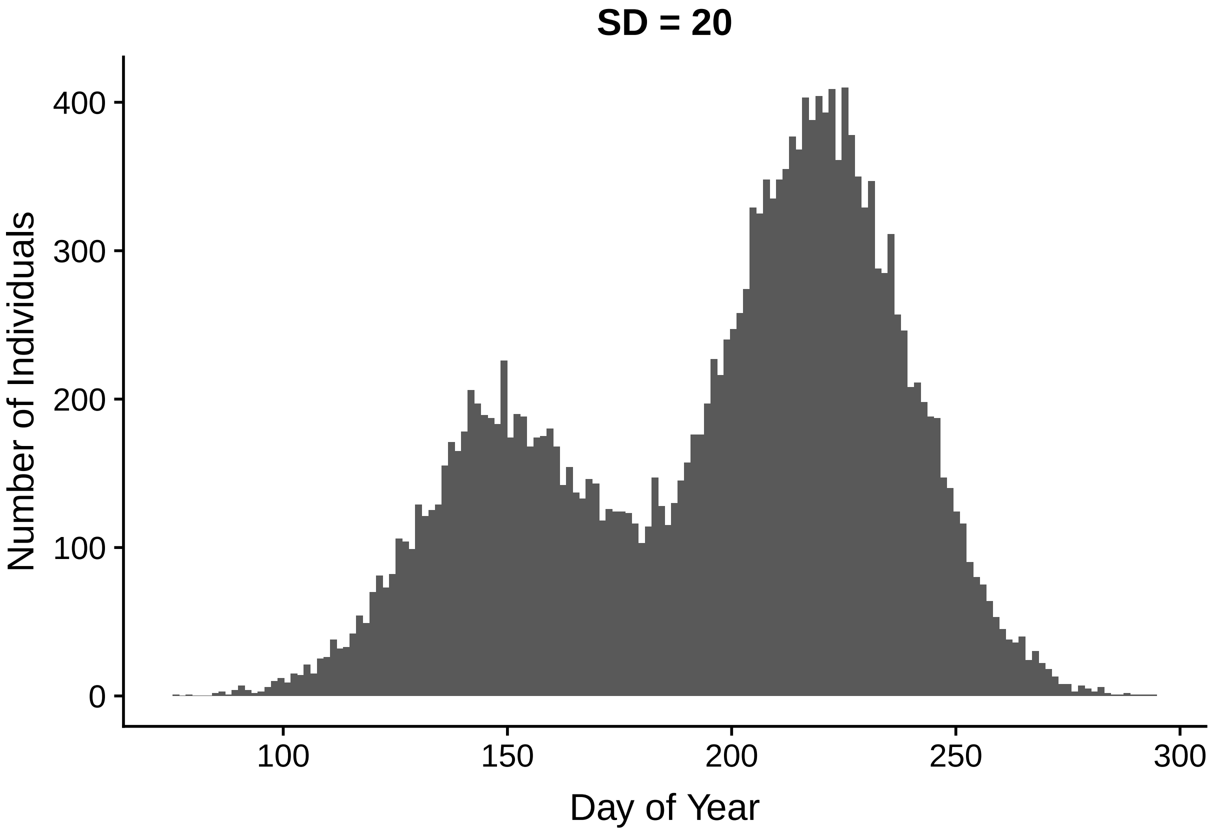


Fig. 15 onset = 76.58, 10% = 139.29, 50% = 206.48, 90% = 240.55, 100% = 294.56

Results of Bimodal simulation

Onset

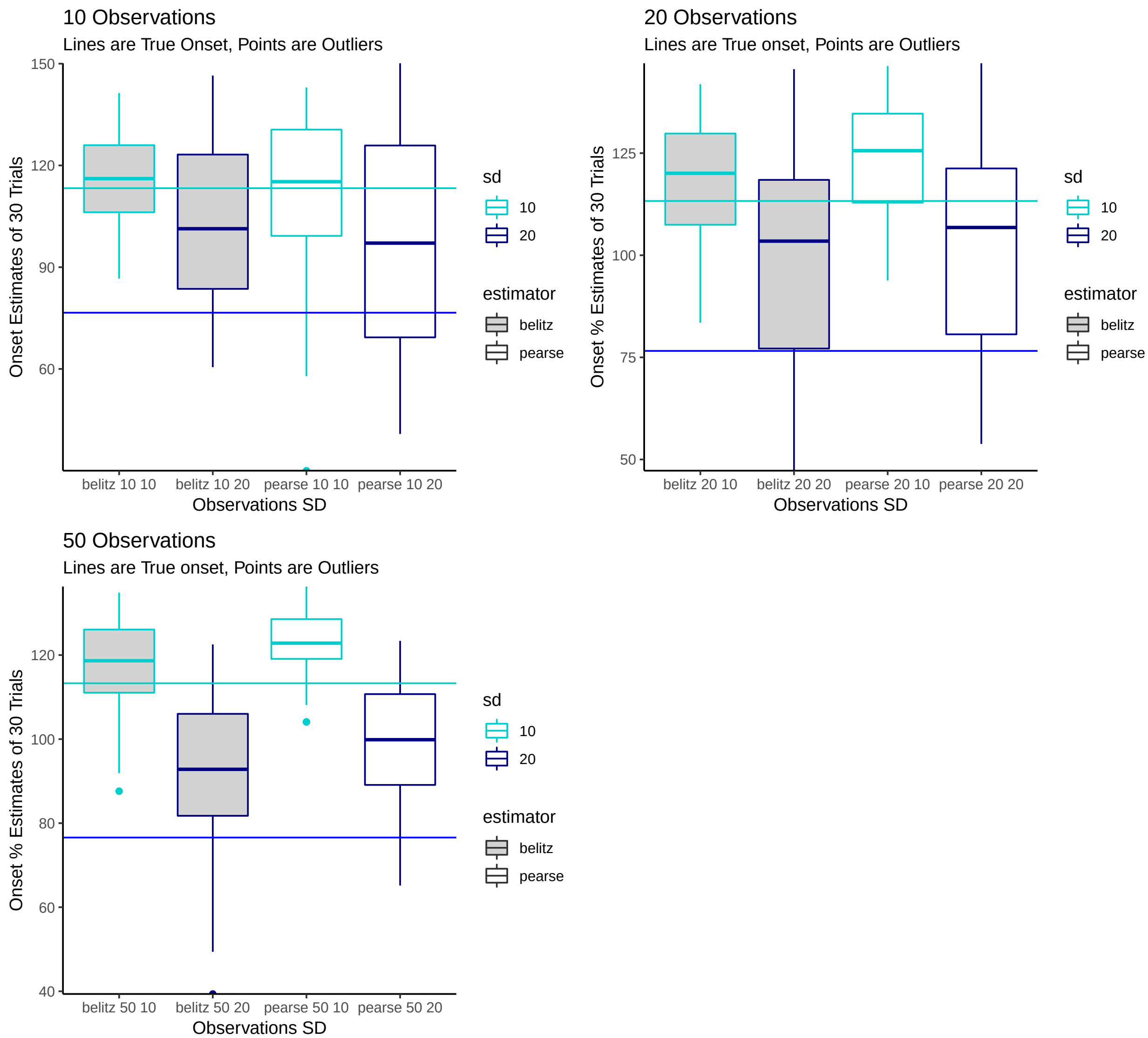


Fig. 16 For onset, it appears that onset can be estimated with greater accurracy when there are the two seperate flight periods (10 SD).

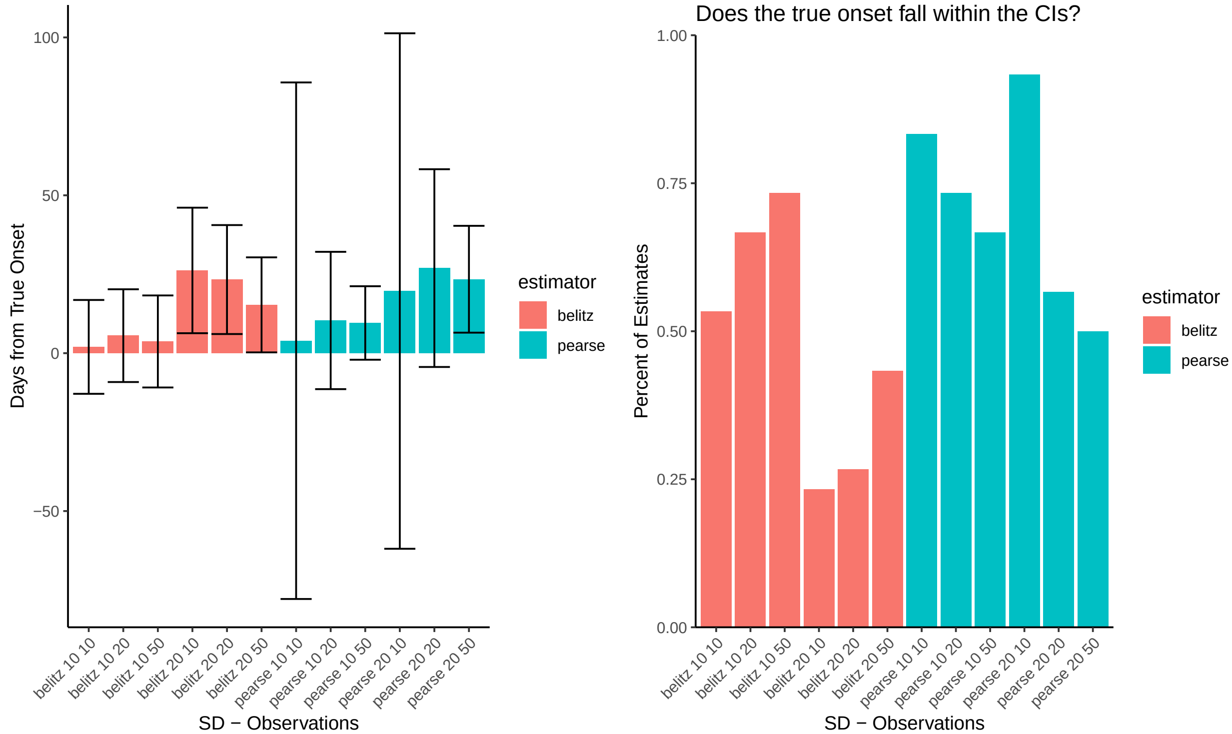


Fig. 17 My estimator does a similar job predicting onset as Pearse's estimator, but Pearse how much larger CI. Therefore, his estimator encapsulates the true onset more often. Interestingly because the precision increases with more observations with the Pearse estimator, but less so with the Belitz estimator, an increase in observations decreases the percent of estimates that encapsulate the true onset with Pearse's estimator but increases with the Belitz estimator.

Tenth

The tenth percentile estimates are noticeably better than the onset estimates for the overlapping broods (20 SD), but the improvement (if any) is less clear with the 10 SD distribution.

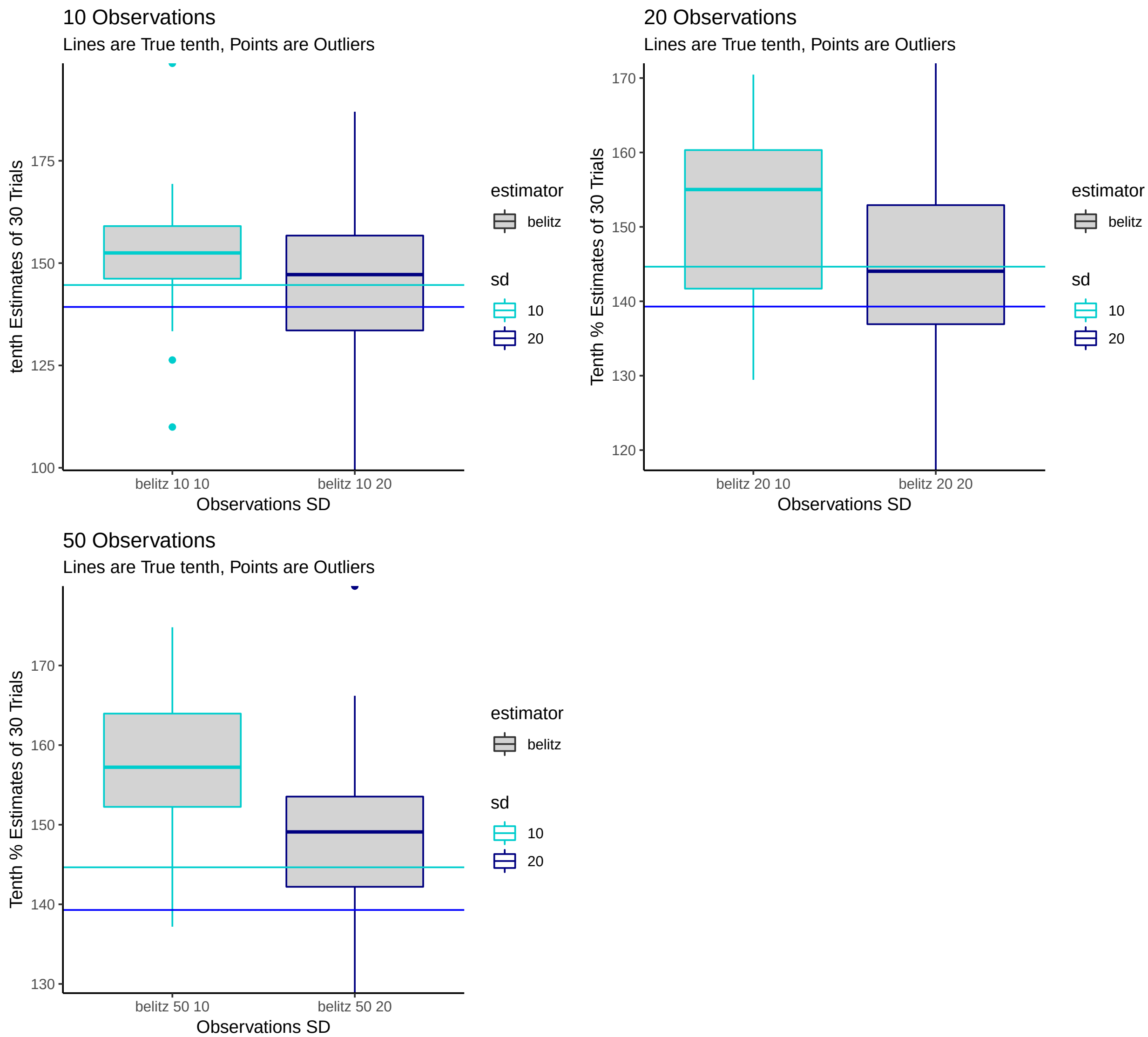


Fig. 18 It appears that when using the 10th percentile, having a a bimodal distribution of more overlap may produce more reliable estimates.

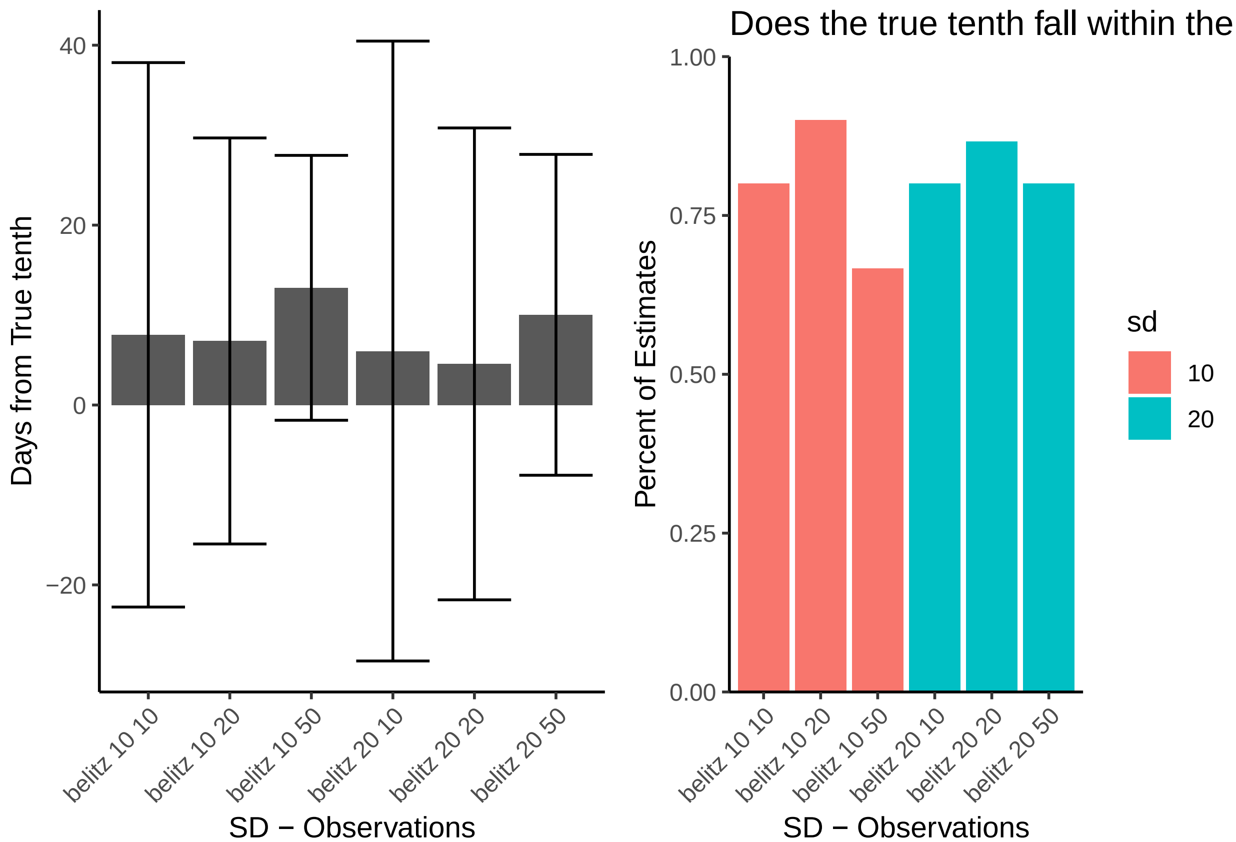


Fig. 19 With this figure, it suggests that the overall accuracy of the 10th percentile estimates improve upon the onset estimates.

Median

Median Estimates

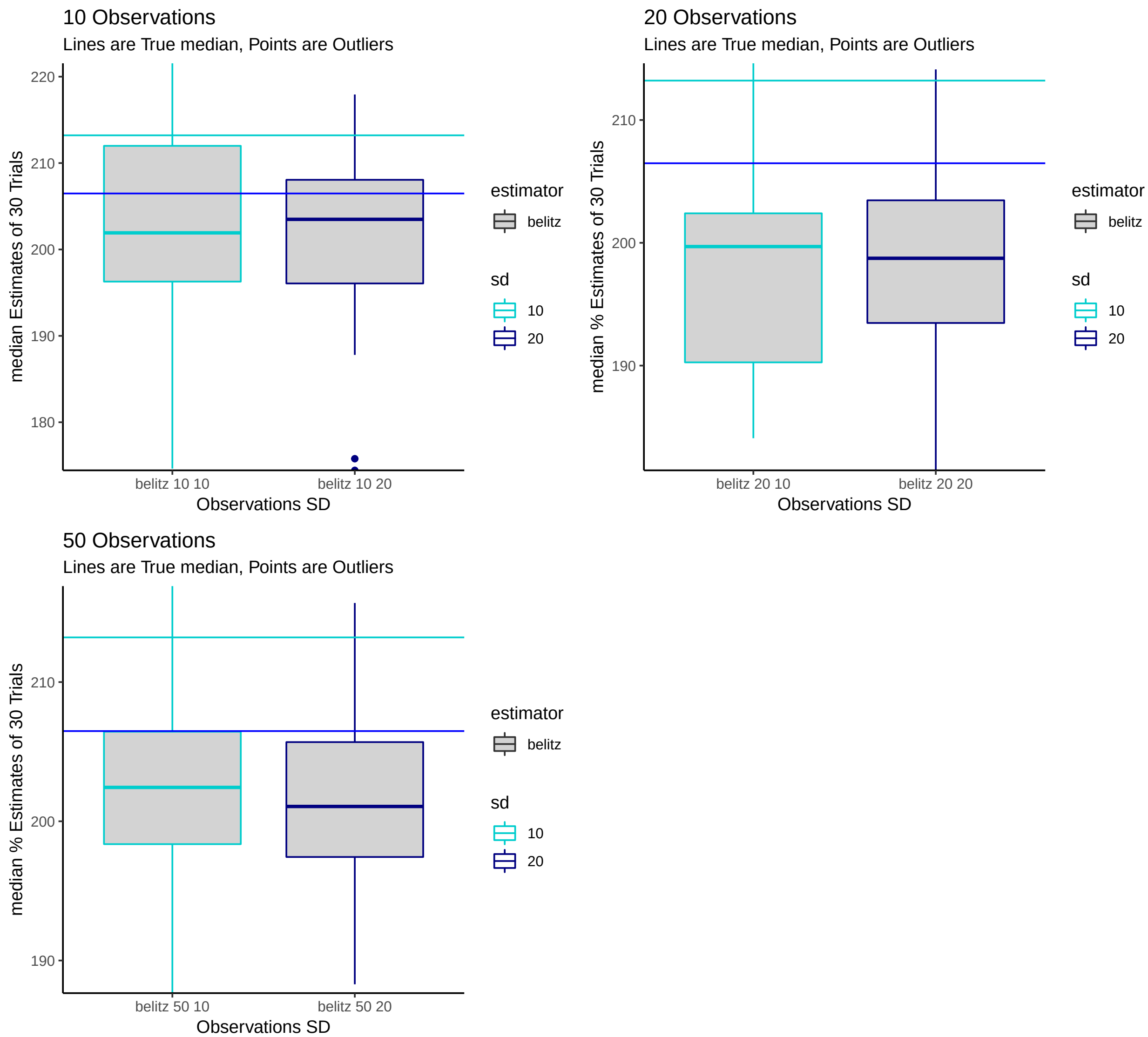


Fig. 20 Median is less reliable than 10%, especially when there is two distinct peaks (10 SD).

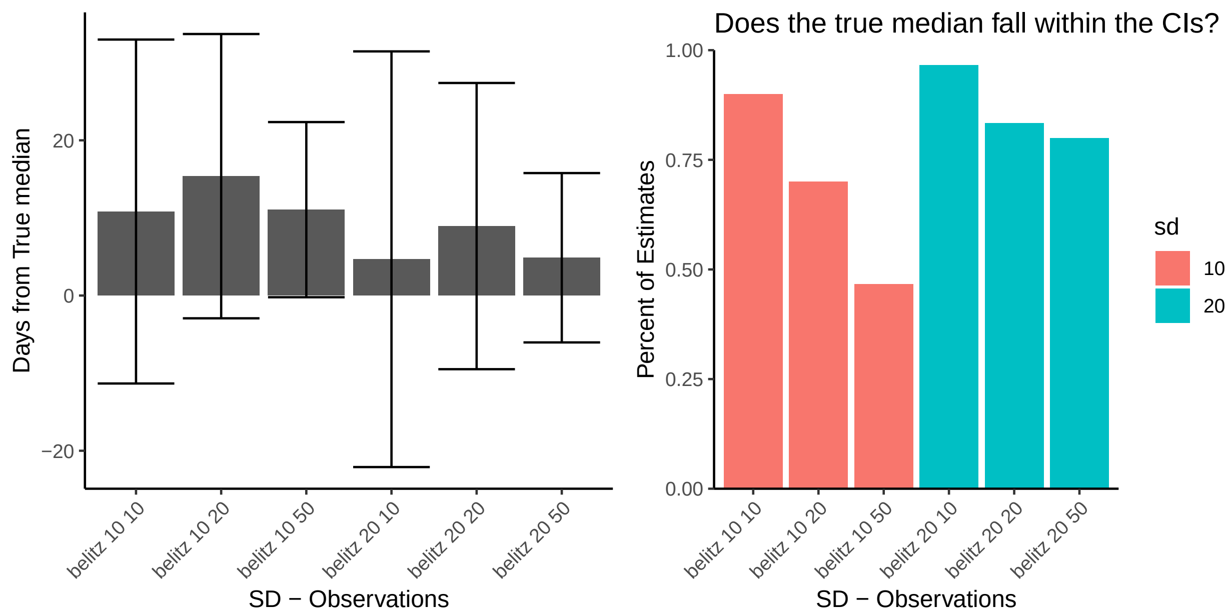


Fig. 21 Due to the distrubtion being back-weighted with more individuals in the second flight period, more observations, just seems to decrease the CI but keep the estimates at around the same DOY, which is consistently below the true median. # Ninety

The ninety percentile estimates are more precise and accurrate than the other estimates tried so far for the bi-modal distribution. The extra-weighting of more individuals being in the second flight periods contribute to this phenomena.

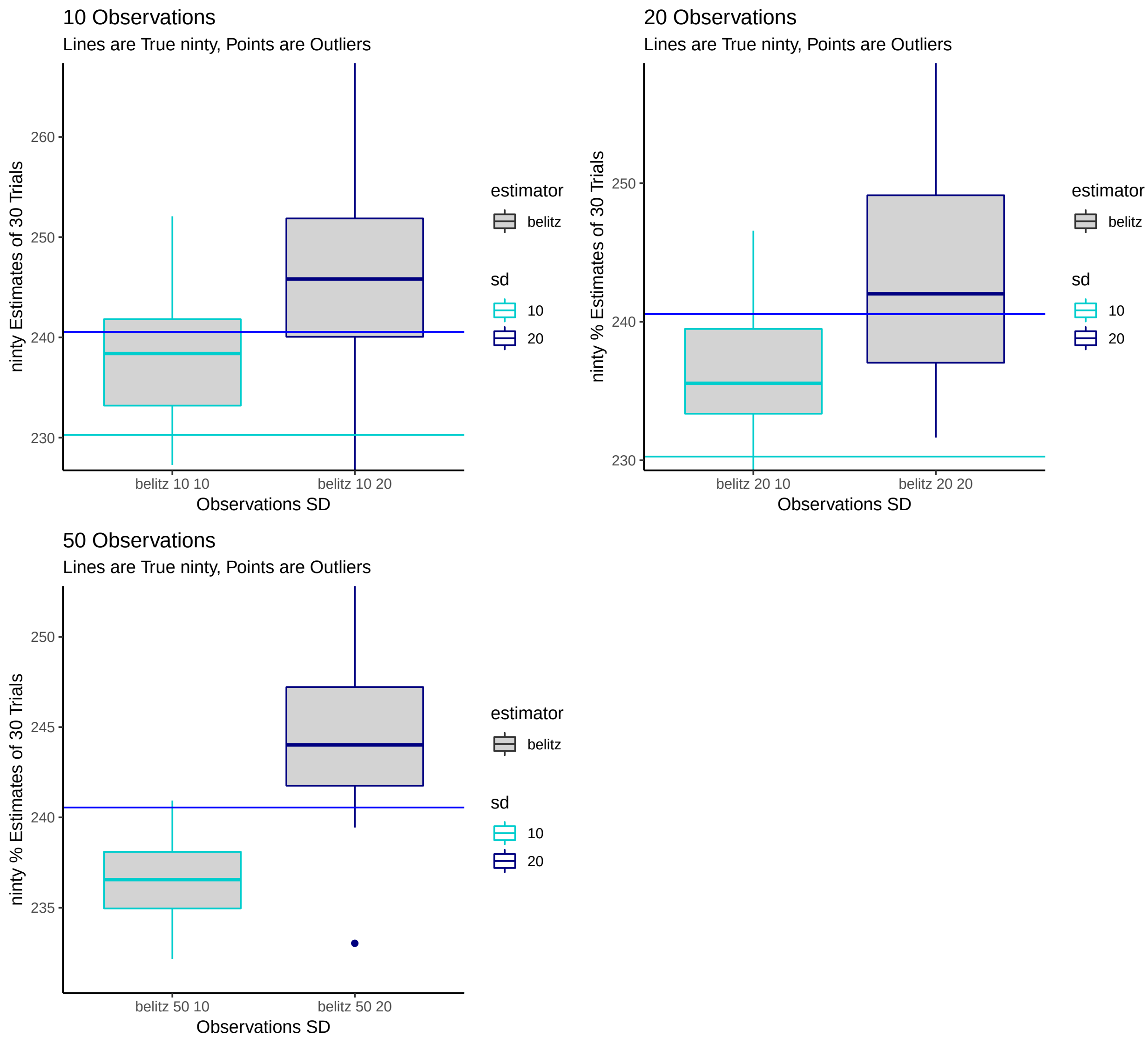


Fig. 22 Ninty percentile estimates

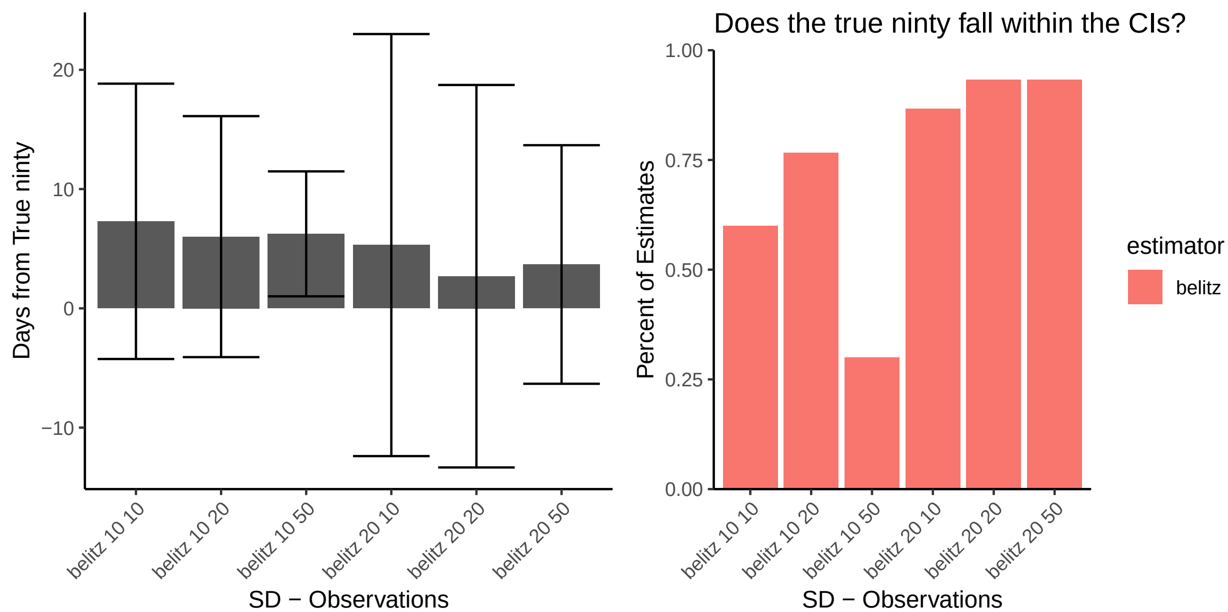


Fig. 23 The ninty percentile estimate has larger CI in the 20 SD, making more estimates encapsulate the true ninty number, even thought the estiamtes are only slightly more accurate.

Offset

Offset does worse than 90%

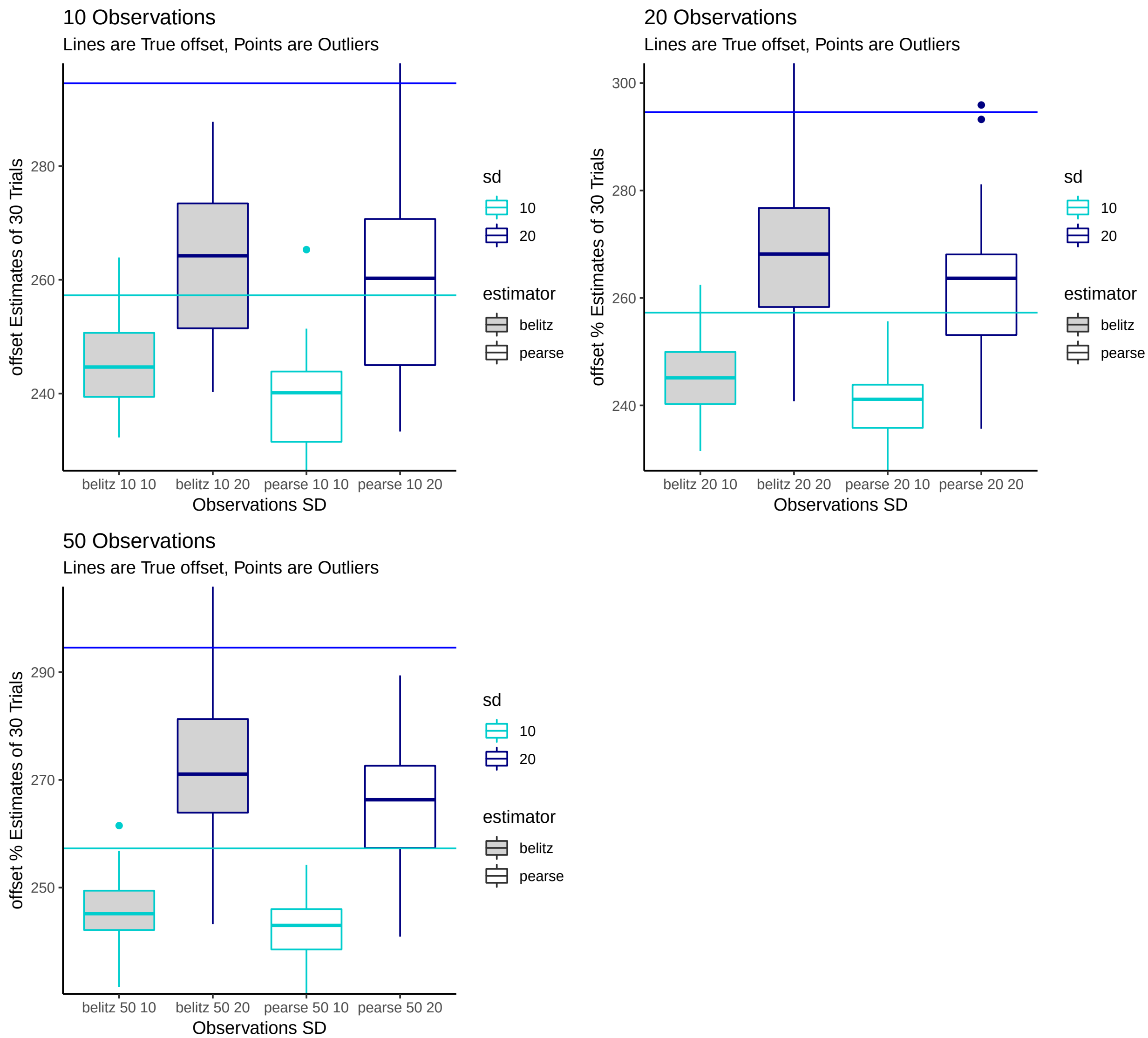


Fig. 24 Offset estimates for bimodal distributions

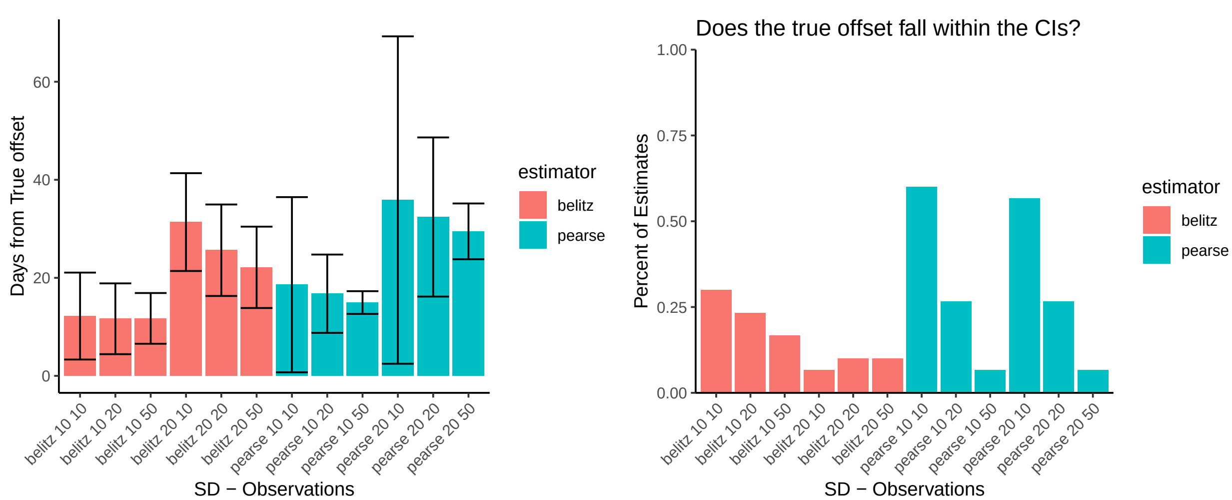


Fig. 25 Offset estimates do not do a good job of predicting true offset compared other pheno-metrics we

estimated for the bimodal distributions.

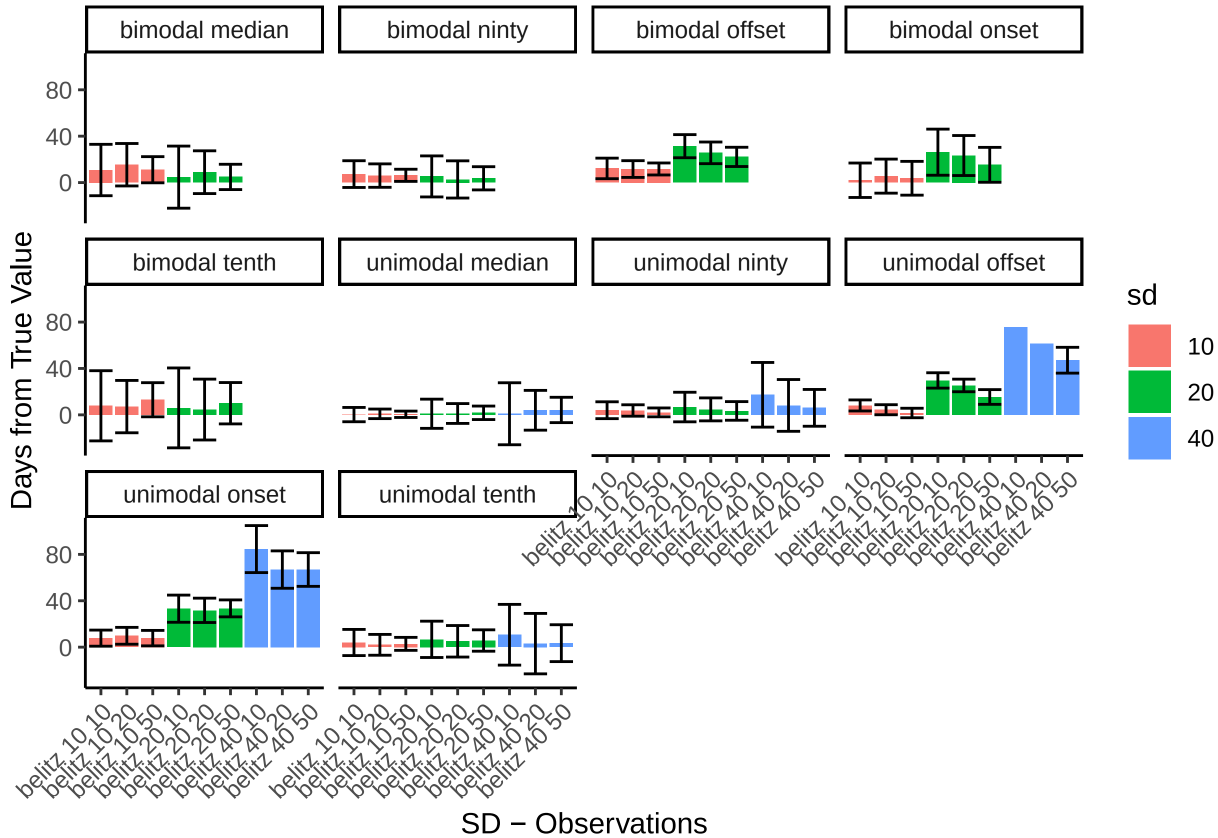
Total Summaries

```
total_pass_sum <- rbind(um_tenth_pass_sum, um_onset_pass_sum, um_median_pass_sum,
                        um_ninty_pass_sum, um_offset_pass_sum, bm_onset_pass_sum,
                        bm_tenth_pass_sum, bm_median_pass_sum, bm_ninty_pass_sum,
                        bm_offset_pass_sum)
```

```
## Warning in bind_rows_(x, .id): Unequal factor levels: coercing to character

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector
```

```
total_pass_sum_nopearse <- total_pass_sum %>%
  dplyr::filter(estimator != "pearse")

total_dis <- ggplot(total_pass_sum_nopearse, aes(x = uid, y = abs(mean_dis))) +
  geom_bar(stat = "identity", aes(fill = sd)) +
  geom_errorbar(aes(ymin = abs(mean_dis) - mean_ci/2, ymax = abs(mean_dis) + mean_ci/2)) +
  labs(x = "SD - Observations", y = "Days from True Value") +
  theme_classic() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  facet_wrap(~ perc)

total_dis
```

```
## Warning: Removed 2 rows containing missing values (geom_errorbar).
```




```
total_pass_sum <- rbind(um_tenth_pass_sum, um_onset_pass_sum, um_median_pass_sum,
                        um_ninty_pass_sum, um_offset_pass_sum, bm_onset_pass_sum,
                        bm_tenth_pass_sum, bm_median_pass_sum, bm_ninty_pass_sum,
                        bm_offset_pass_sum)
```

```
## Warning in bind_rows_(x, .id): Unequal factor levels: coercing to character

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector
```

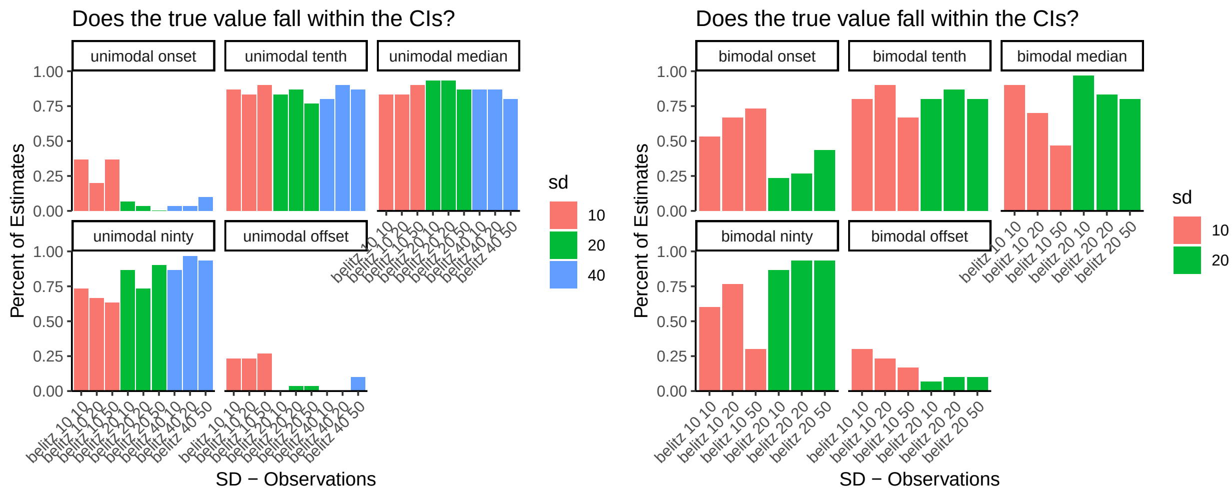
```
total_pass_sum_nopearse <- total_pass_sum %>%
  dplyr::filter(estimator != "pearse") %>%
```

```
  dplyr::mutate(perc = factor(perc, levels = c("unimodal onset", "unimodal tenth",
                              "unimodal median", "unimodal ninty", "unimodal offset",
                              "bimodal onset", "bimodal tenth", "bimodal median", "bimodal ninty", "bimo

total_corr <- ggplot(total_pass_sum_nopearse, aes(x = uid, y = abs(percent_right))) +
  geom_bar(stat = "identity", aes(fill = sd)) +
  labs(x = "SD - Observations", y = "Percent of Estimates") +
  ggtitle("Does the true value fall within the CIs?") +
  scale_y_continuous(limits = c(0,1), expand = c(0,0)) +
  theme_classic() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  facet_wrap(~perc)
```

```
unimodal_pass_sum <- rbind(um_tenth_pass_sum, um_onset_pass_sum, um_median_pass_sum,
                        um_ninty_pass_sum, um_offset_pass_sum)
bimodal_pass_sum <- rbind(bm_onset_pass_sum, bm_tenth_pass_sum, bm_median_pass_sum, bm_ninty_pass_sum,
                        bm_offset_pass_sum)

um_pass_sum_nopearse <- unimodal_pass_sum %>%
  dplyr::filter(estimator != "pearse") %>%
  dplyr::mutate(perc = factor(perc, levels = c("unimodal onset", "unimodal tenth",
                              "unimodal median", "unimodal ninty", "unimodal offset")))

bm_pass_sum_nopearse <- bimodal_pass_sum %>%
  dplyr::filter(estimator != "pearse") %>%
  dplyr::mutate(perc = factor(perc, levels = c("bimodal onset", "bimodal tenth",
                              "bimodal median", "bimodal ninty", "bimodal offset")))

um_corr <- ggplot(um_pass_sum_nopearse, aes(x = uid, y = abs(percent_right))) +
  geom_bar(stat = "identity", aes(fill = sd)) +
  labs(x = "SD - Observations", y = "Percent of Estimates") +
  ggtitle("Does the true value fall within the CIs?") +
  scale_y_continuous(limits = c(0,1), expand = c(0,0)) +
  theme_classic() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  facet_wrap(~perc)

bm_corr <- ggplot(bm_pass_sum_nopearse, aes(x = uid, y = abs(percent_right))) +
  geom_bar(stat = "identity", aes(fill = sd)) +
  labs(x = "SD - Observations", y = "Percent of Estimates") +
  ggtitle("Does the true value fall within the CIs?") +
  scale_y_continuous(limits = c(0,1), expand = c(0,0)) +
  scale_fill_manual(values = c("#F8766D", "#00BA38")) +
  theme_classic() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  facet_wrap(~perc)

total_corr <- cowplot::plot_grid(um_corr, bm_corr)
```

```
## Warning: Removed 2 rows containing missing values (position_stack).
```

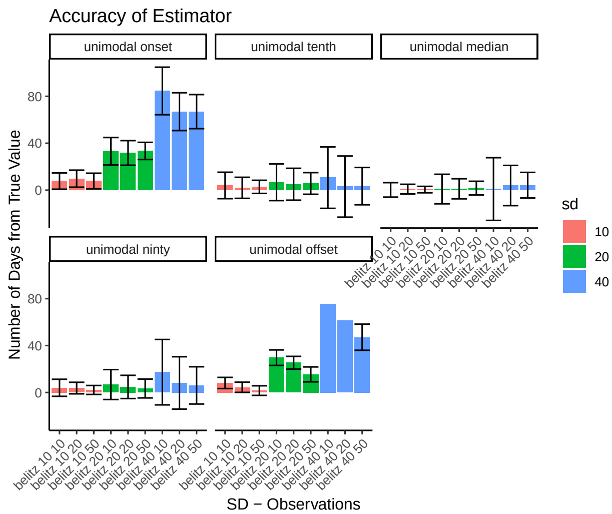
```
total_corr
```



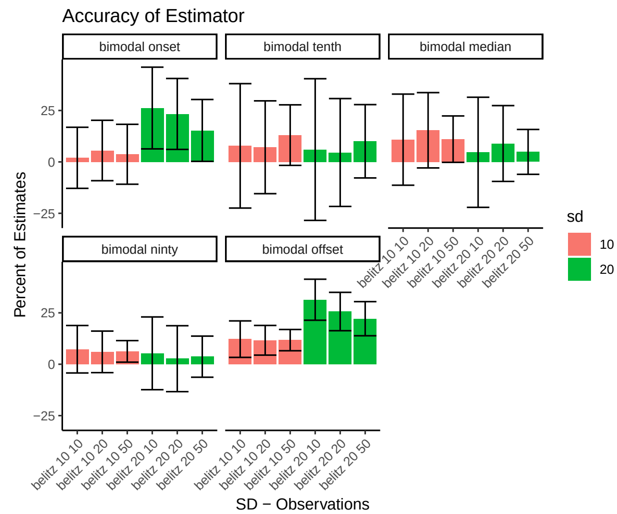

```
um_dis <- ggplot(um_pass_sum_nopearse, aes(x = uid, y = abs(mean_dis))) +
  geom_bar(stat = "identity", aes(fill = sd)) +
  labs(x = "SD - Observations", y = "Number of Days from True Value") +
  geom_errorbar(aes(ymin = abs(mean_dis) - mean_ci/2, ymax = abs(mean_dis) + mean_ci/2)) +
  ggtitle("Accuracy of Estimator") +
  theme_classic() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  facet_wrap(~perc)

bm_dis <- ggplot(bm_pass_sum_nopearse, aes(x = uid, y = abs(mean_dis))) +
  geom_bar(stat = "identity", aes(fill = sd)) +
  geom_errorbar(aes(ymin = abs(mean_dis) - mean_ci/2, ymax = abs(mean_dis) + mean_ci/2)) +
  labs(x = "SD - Observations", y = "Percent of Estimates") +
  ggtitle("Accuracy of Estimator") +
  scale_fill_manual(values = c("#F8766D", "#00BA38")) +
  theme_classic() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  facet_wrap(~perc)

total_dis <- cowplot::plot_grid(um_dis, bm_dis)
```

```
## Warning: Removed 2 rows containing missing values (geom_errorbar).
```

```
total_dis
```



Accuracy of Estimator
unimodal onset
unimodal tenth
unimodal median
unimodal ninty
unimodal offset
Number of Days from True Value
SD − Observations
sd
10
20
40



Accuracy of Estimator
bimodal onset
bimodal tenth
bimodal median
bimodal ninty
bimodal offset
Percent of Estimates
SD − Observations
sd
10
20