A Closer Look at Particle Counts

30/06/2020

Background

Expt D in the Transfer studies shows high variability and much higher transfer ratio, T_r, values than the others, especially with wool. On closer inspection it appears that it (as well as expt B to some extent) shows quite high values for Photo 1, P_1, - median ~100) - when the A,C & E are much near to zero.

With that information can simulate or reproduce the behaviour on simulated data. As all the data are raw counts of independent events it is reasonable to assume a Poisson distribution. Poisson distributions are dependent on two parameters k, the number of occurrences, and λ, the mean (which is also equal to the variance).

The hypothesis being are relatively high P_1 counts causative of high Transfer Ratios independently of other photo counts?

So we can model the counts for photos P_1, P_2 and P_3 assuming a Poisson distribution and taking equivalent means from the Transfer experiment. Photos P_4 and P_5 are a function of these starting counts based on the amount of transfer. A 35% transfer of particles is consistent with the data. A random normal sampling around this mean with a standard deviation of 4 was used.

$n = 20$ for the number of times each transfer 'experiment' was simulated.

Loss and 'declumping' of particles was not simulated.

Simulation

```
# function to produce the starting data
makeData <- function(p1 = 45, p2 = 6, p3 = 120, n = 20, transfer.pc = 35) {
    # create donor background data
    p1.dist = rpois(n, p1)

    # low background on receiver
    p2.dist = rpois(n, p2)
    # add some particles
    p3.dist = rpois(n, p3)

    # assume 30-40 percent transfer
    trans = round(rnorm(n, transfer.pc, sd = 4))/100

    # create modelled data for particulate transfer
    df = data.frame(p1.dist, p2.dist, p3.dist, p4 = p1.dist + round(p3.dist * trans),
        p5 = p1.dist + round(p3.dist * (1 - trans)))

    return(df)
}
```

With the function for creating the distributions defined a range of counts was parameterised:

- $P_1 = 5, 30, 55, 80, 105$

- $P_2 = 5, 30, 55, 80, 105$
- $P_3 = 100, 150, 200, 250, 300$

```
df.dat = data.frame()
for (p1 in seq(5, 105, 25)) {
    for (p2 in seq(5, 105, 25)) {
        for (p3 in seq(100, 300, 50)) {
            df = makeData(p1, p2, p3)
            if (nrow(df.dat) > 0) {
                tmp = data.frame(p1.m = as.character(p1), p2.m = as.character(p2),
                  p3.m = as.character(p3), df)
                df.dat = rbind(df.dat, tmp)
            } else {
                df.dat = data.frame(p1.m = as.character(p1), p2.m = as.character(p2),
                  p3.m = as.character(p3), df)
            }
        }
    }
}
```

The Transfer Ratio was calculated as previous described:

$$t_r = (P_5 - P_2)/(P_3 - P_1)$$

```
# calc transfer ratio
df.dat$tr = (df.dat$p5 - df.dat$p2.dist)/(df.dat$p3.dist - df.dat$p1.dist)
```

Note: If $P_1 = P_3$ the denominator is 0 which results in a non-finite T_r.

Results

If we fix $P_2 = 5$ and simulate P_1 and P_3 we can visualise the effect on T_r as a function of both counts.

```
df.dat %>% filter(p2.m == "5") %>% ggplot(aes(x = factor(p1.m, levels = c(seq(5,
    105, 25))), y = tr)) + labs(title = "Simulation of Transfer Ratio as a Function of P1 & P3 Counts",
    subtitle = "n = 20, Mean P2 counts = 5, % particles transferred = 35% (+/- 4 s.d.)",
    x = "Mean P1 Counts", y = "Transfer Ratio", caption = "Y-axes are not the same") +
    geom_boxplot() + facet_wrap(p3.m ~ ., scales = "free_y")
```
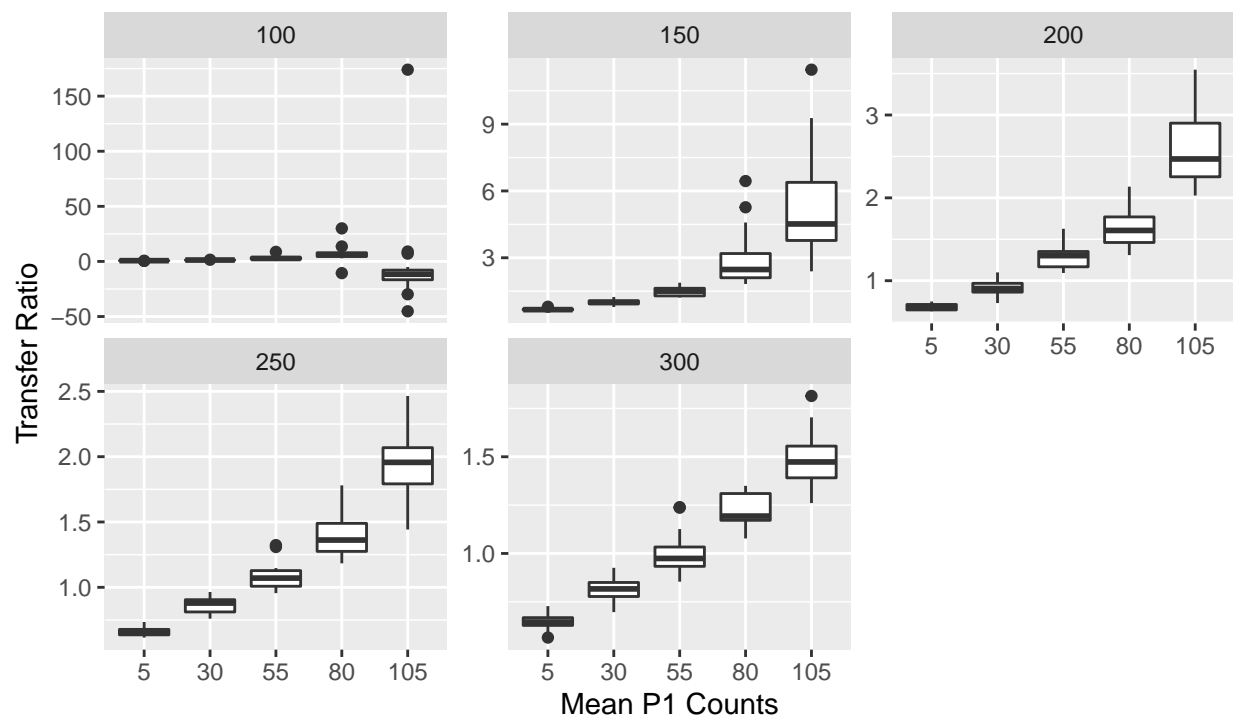
```
# ggsave('p1_count_simulation.png')
```

Figure 1 shows that for $P_3 > 150$ there is a predictable, linear relationship between P_1 and P_3 counts. However, with $P_3 \leq 150$ and mean P_1 increasing the T_r relationship gets very unstable - high positive and negative values. This suggests that it's not just high P_1 counts that are the problem, but it's when $P_1 \approx P_3$ then T_r is extreme.

As a comparison if we look at the P_2 (Figure 2), the receiver background, relationship with T_r there is no issue either at high counts or when $P_2 \approx P_3$.

```
df.dat %>% filter(p1.m == 5) %>% ggplot(aes(x = factor(p2.m, levels = c(seq(5, 105,
    25))), y = tr)) + labs(title = "Simulation of Transfer Ratio as a Function of P2 Counts",
    subtitle = "n = 20, Mean P1 counts = 5, % particles transferred = 35% (+/- 4 s.d.)",
    x = "Mean P2 Counts", y = "Transfer Ratio", caption = "Y-axes are the same") +
    geom_boxplot() + facet_wrap(p3.m ~ .)
```
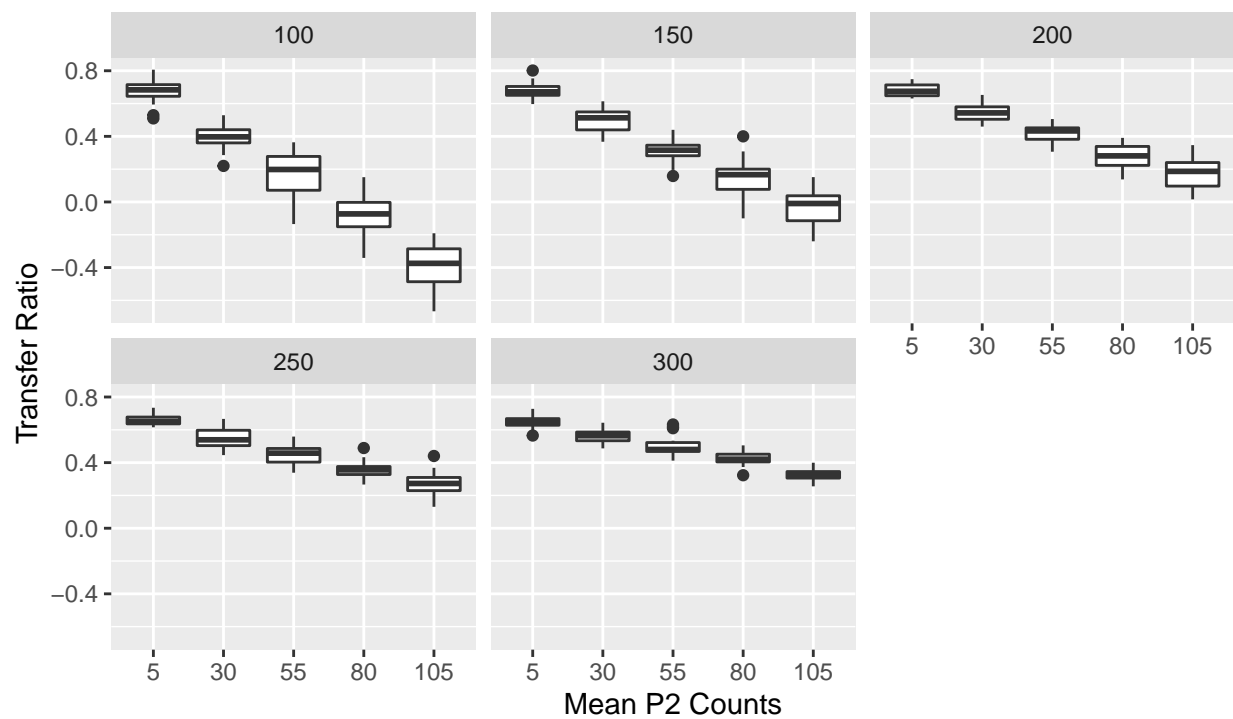
2

Figure 1: Simulation of Transfer Ratio as a Function of P1 & P3 Counts

Figure 2: Simulation of Transfer Ratio as a Function of P2 Counts

Conclusion

High P_1 - background receiver counts - are not the cause of the observed issues with T_r in expt D, but it appears to be when $P_1 \approx P_3$ that results in very high or negative T_r values. This is not the case for expt D which suggests that the hypothesis that high P_1 counts are causative of high Transfer Ratios is incorrect.