

Lab 5: Network Assignment and Validation

Hayden Atchley, Steven Goodsell, Tristan Parker, Austin Nichalson

November 05, 2021

This lab is focused on determining and discussing the differences between the model's predicted traffic volumes and actual observed volumes.

Calculating Error

After calibrating the model in our last lab we could run the trip assignment section in Cube. The outputs from this run gave us a loaded network file and highway link summary that we could further analyze to determine which classes of roads are performing the best in our model. We calculated the percent root-mean-squared error (pRMSE), grouped by facility type, area type, screenline, and volume group:

```
links <- read_csv("data/RMSE_Links.csv") %>%
  mutate(error = Volume - Count,
         pcterror = error/Count * 100)

links %<>%
  mutate(`Fac Type Group` = factor(
    `Fac Type Group`, labels = c(
      "1 - Highway", "2 - Arterial", "3 - Collector", "4 - Local")),
    `Area Type` = factor(
      `Area Type`, labels = c(
        "2 - Urban", "3 - Suburban", "4 - Rural Suburban", "5 - Rural"))
  )

links %>%
  group_by(`Fac Type Group`) %>%
  summarise(prmse = prmse(Volume, Count) %>% percent(accuracy = 0.1, suffix = "\\%")) %>%
  my_kbl(digits = 1)
```

Fac Type Group	prmse
1 - Highway	20.9%
2 - Arterial	36.2%
3 - Collector	42.4%
4 - Local	70.4%

Area Type	pRMSE
2 - Urban	46.5%
3 - Suburban	47.0%
4 - Rural Suburban	38.8%
5 - Rural	28.0%

Screenline	pRMSE
1	54.3%
2	17.9%
3	42.4%
4	40.9%
98	41.8%

Volume Group	pRMSE
1 - 5000	56.3%
5000 - 10000	32.6%
10000 - 15000	32.8%
15000 - 20000	23.8%
20000 - 30000	15.4%
30000 +	26.8%

Facility type 1 (highways) in our model seems closer to the actual traffic count data than any other facility type. It is also less common than all the other facility type groups, so there may be less variation in traffic volume than the larger facility type groups. Facility type 4 (local roads) has the worst prediction in our trip assignment model and a higher PRMSE than any other class. This could be because there are more of this kind of link than any other so there is more variation in counts across the city.

Comparing PRMSE analysis by area type, we find area type 5 (rural) is the closest to approximating actual values from our model. The other error values are relatively low for this class analysis. Values between 20% and 50% are considered ok. Overall, the model is better at estimating rural suburban and rural areas in the region (types 4 and 5).

The volume group PRMSE values indicate that sorting the data by volume group gives very accurate results for high volume roads in the group 20000 - 30000. However, there aren't very many links of that category in the system. Estimating traffic on low volume roads in the group 1 - 5000 is very difficult with this grouping, resulting in 56.3% PRMSE.

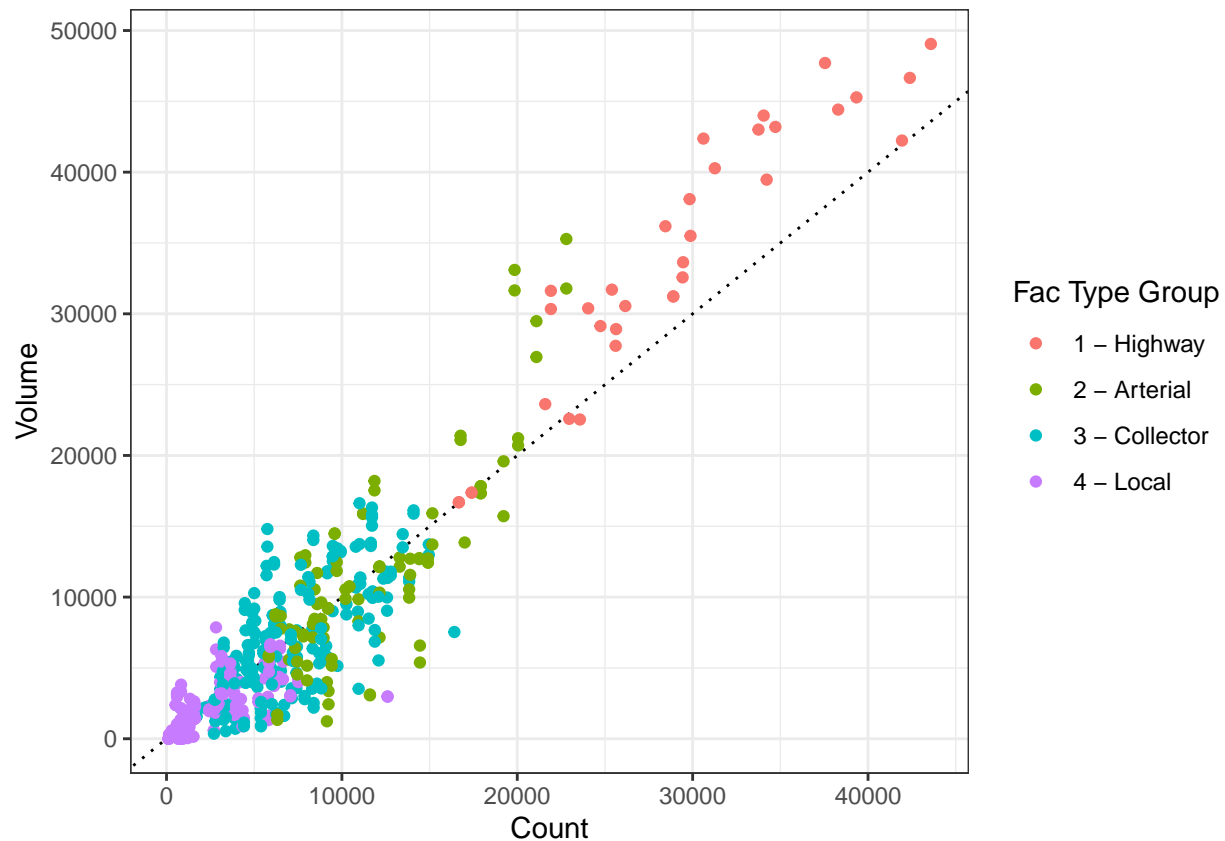
Screenline 2 is very accurate with a PRMSE of 17.9%. The other screenlines, however, provide mediocre results, ranging from 40.9% to 54.3%, all skirting around the general value of all links not on a screenline, labeled as screenline 98. Screenlines generally don't seem to be a very reliable way to predict traffic volumes.

Our table with PRMSE values for every type in the four classes can be used as a validation report determining what cases our model is best for estimating traffic volumes. Overall, the best estimation for traffic volumes are highways in the rural areas for a high volume group and crossing screenline 2. Our model is not performing well at estimating local roads in the suburban areas for low volumes and crossing screenline 1. We can now take this validation information and examine what might be causing our model to be less accurate in certain areas and types.

Plotting Observed v. Modeled

We created a plot showing the observed vs modeled volumes, as well as a line showing the ideal distribution (where modeled and observed volumes match):

```
ggplot(links,
  aes(x= Count, y = Volume, color = `Fac Type Group`)) +
  geom_abline(slope = 1, intercept = 0, lty = "dotted") +
  geom_point() + theme_bw()
```



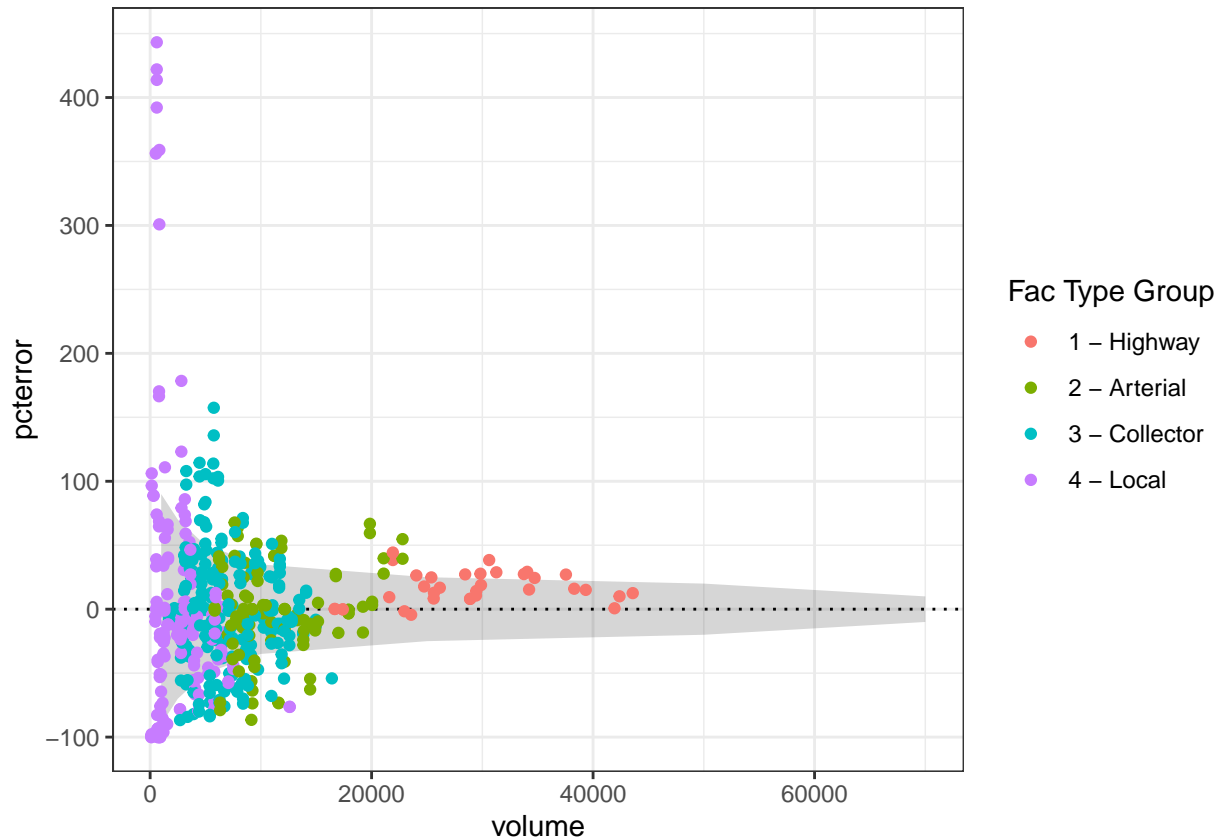
Additionally, the MDD plot in NCHRP 765 identifies a “target zone” that percent errors should fall into:

```
mdd <- tibble(
  volume = c(1000, 2500, 5000, 10000, 25000, 50000, 70000),
  mdd = c(0.9, 0.7, 0.5, 0.35, 0.25, 0.2, 0.10) * 100
)
```

We plotted our model’s percent errors along with this “target zone”:

```
ggplot() +
  # add the MDD ribbon, with positive upside and negative downside
  geom_ribbon(
    data = mdd %>% mutate(mdd1 = -1 * mdd),
    aes(x = volume, ymax = mdd, ymin = mdd1), alpha = 0.2) +
  # add a horizontal line
```

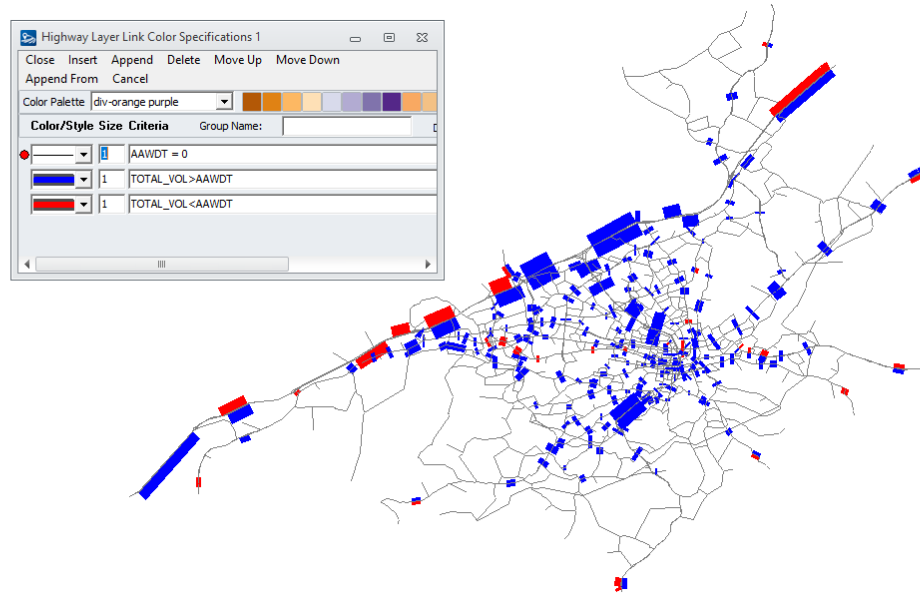
```
geom_hline(yintercept = 0, lty = "dotted") +
# add the link percent error
geom_point(data = links, aes (x = Count, y = pcterror, color = `Fac Type Group`)) +
theme_bw()
```



It's clear that we are generally overpredicting Highway volumes, but tend to be mostly okay on the other facility types. The smaller the volume on the link, the less accurate our model becomes, but since most of the data falls within the MDD "target zone", we conclude that our model is decent enough for the purposes of this class, but would probably require some further calibration if used in a more professional setting.

Geographic Distribution of Error

The following network depicts the geographic distribution of the Link Error in the Roanoke model in CUBE:



To accurately depict the Link Error in the region, we used the following three formulas: $AAWDT = 0$, $TOTAL_VOL > AAWDT$, and $TOTAL_VOL < AAWDT$. Using the multi-bandwidth tool in CUBE, we are able to show the areas where the Annual Average Weekday Traffic is greater than or less than the Total Volume associated with the model. The blue links on the map indicate where the AAWDT is less than the Total Volume and the red links indicate where the AAWDT is greater than the Total Volume assigned to this particular model.

From this map, we can identify that the blue links are where the model is overestimating traffic volume and the red links are where it is underestimating. In general, we observe that the model has a tendency to overestimate traffic in the Roanoke area more so than underestimating it. Specifically, areas that are on or near the interstate, as well as arterials that connect to the downtown areas in Roanoke are typically overestimated.