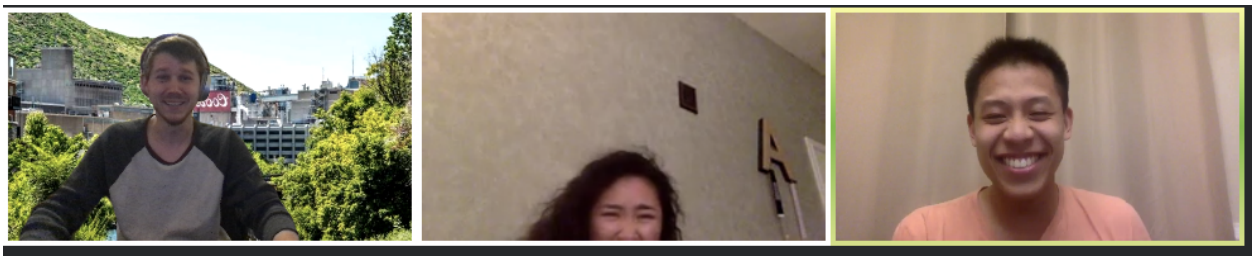
City of Boulder

Exploratory Data Analysis

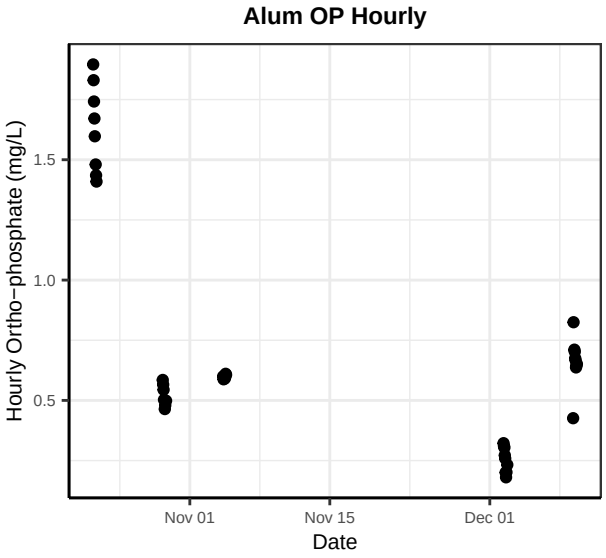


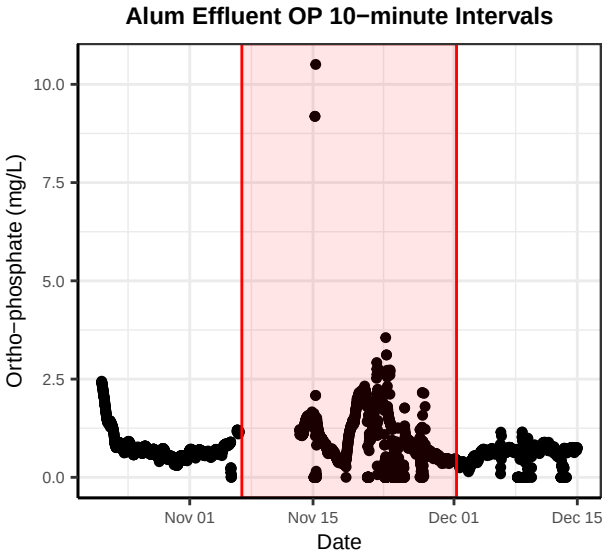
Justin Nichols Arlyn Alcid Ocean Wave

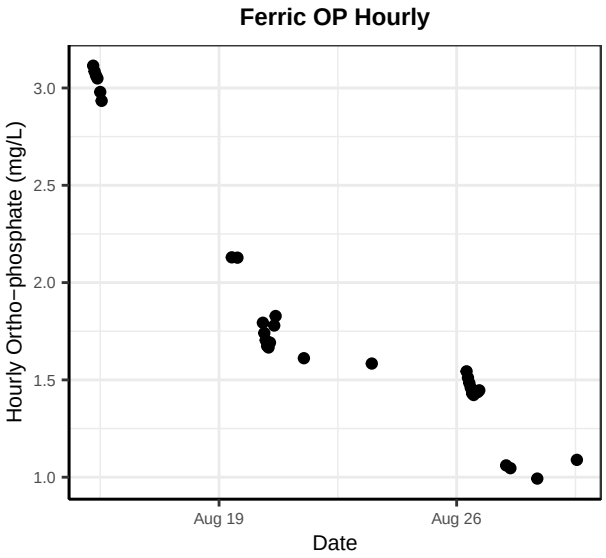
We begin by looking at how many observations we have based on the coagulant:

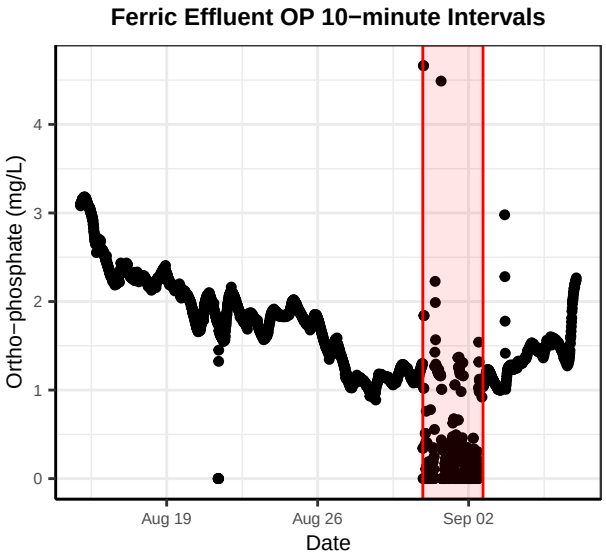
	Alum	Ferric	None
0	40	32	4

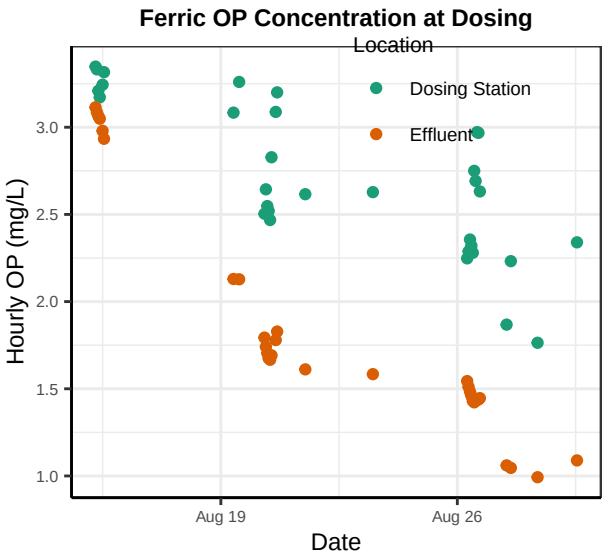
The following plots provide visuals for the effluent OP. They are split up by ferric or alum, as well as a comparison of the hourly data after removal with the original 10-minute intervals.

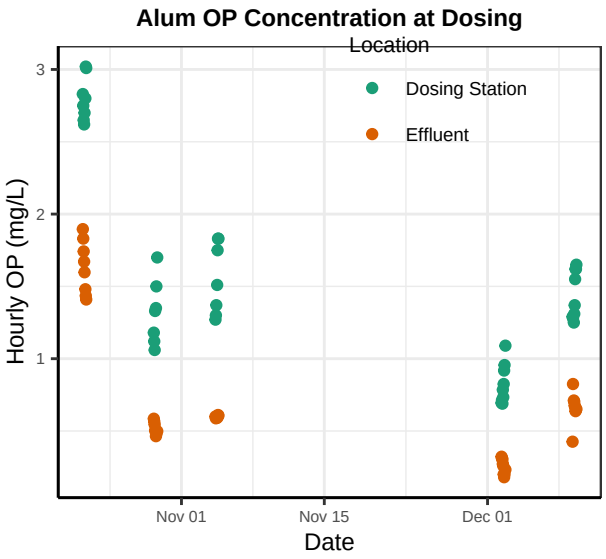


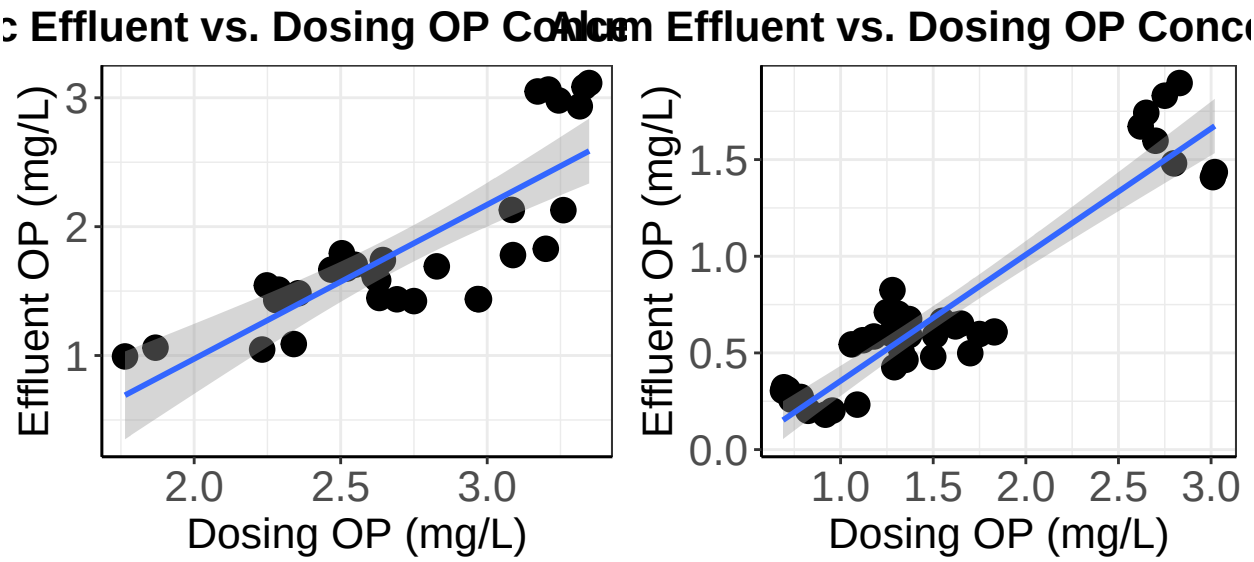


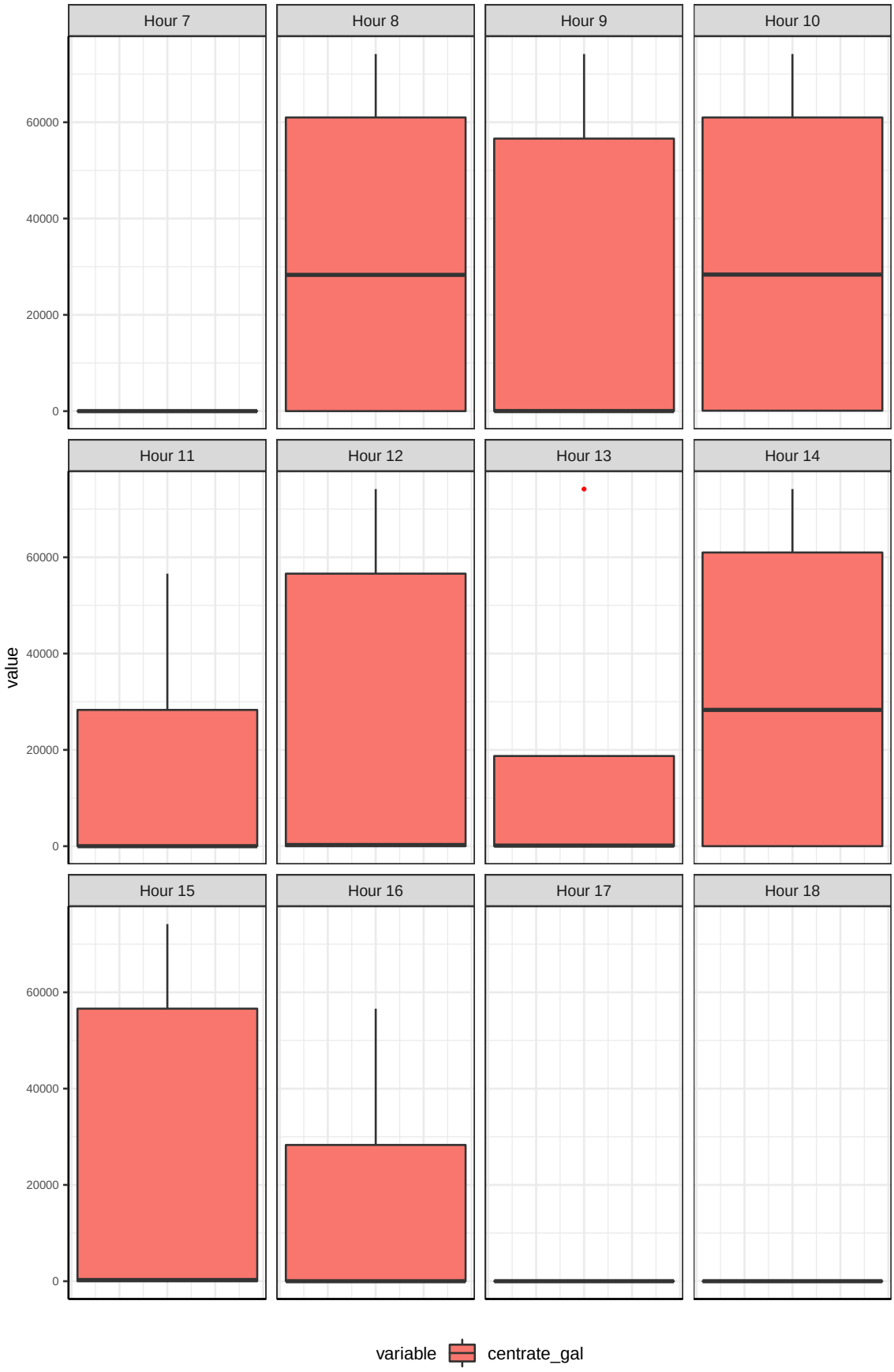


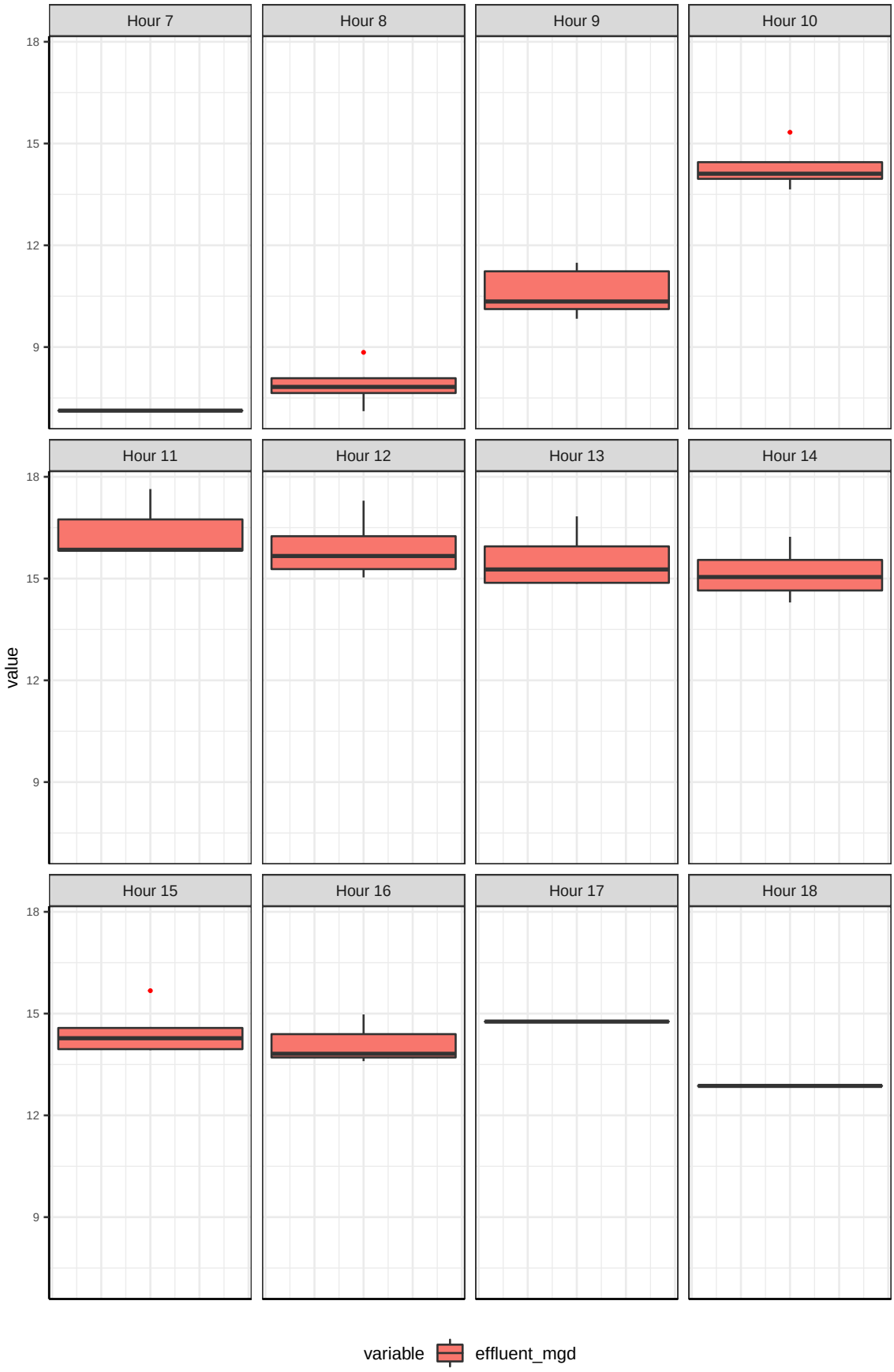


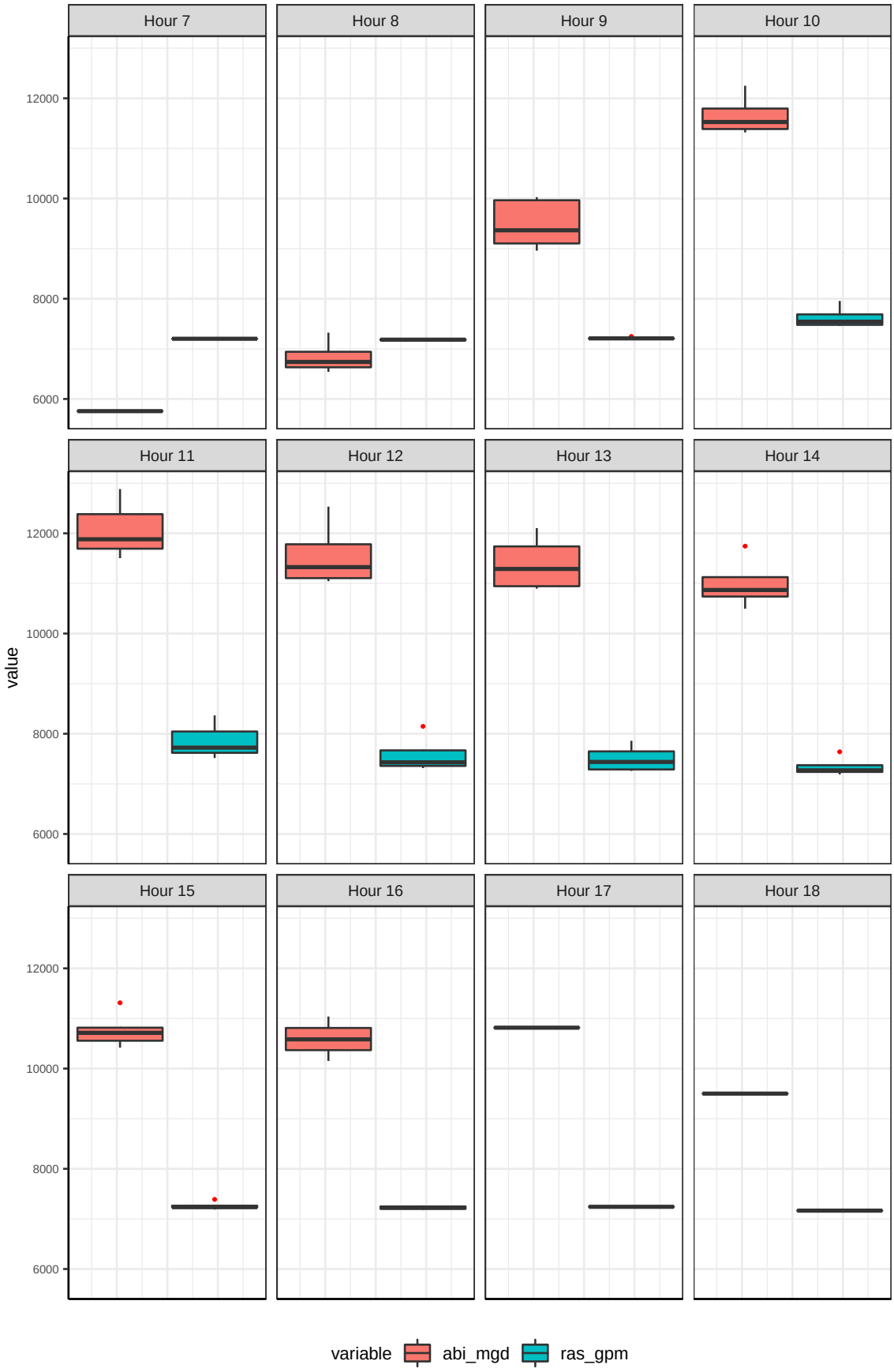

Ferric OP Concentration at Dosing
Location
Dosing Station
Effluent
Hourly OP (mg/L)
3.0
2.5
2.0
1.5
1.0
Aug 19
Aug 26
Date

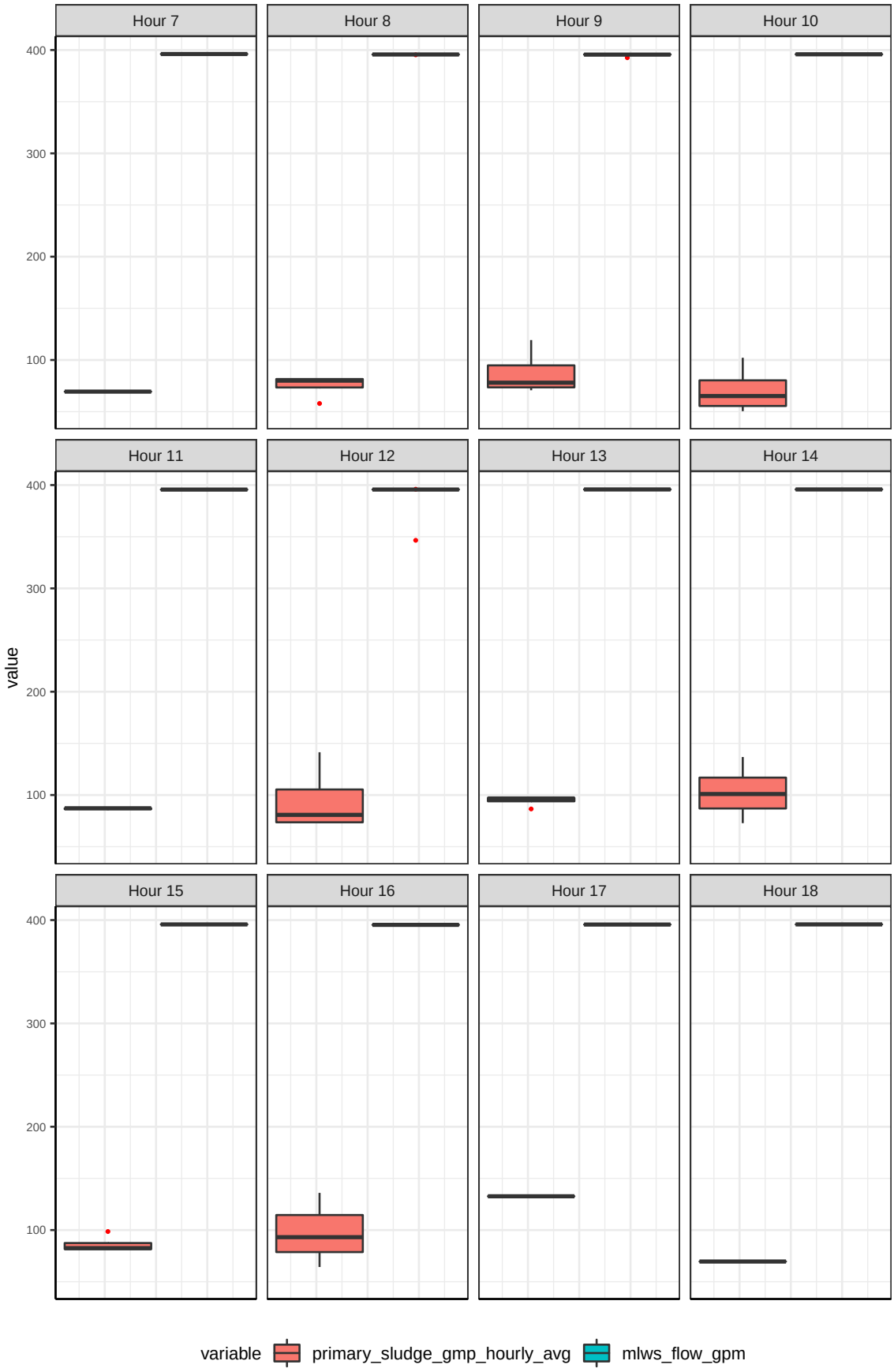

Alum OP Concentration at Dosing
Location
Dosing Station
Effluent
Hourly OP (mg/L)
3
2
1
Nov 01
Nov 15
Dec 01
Date

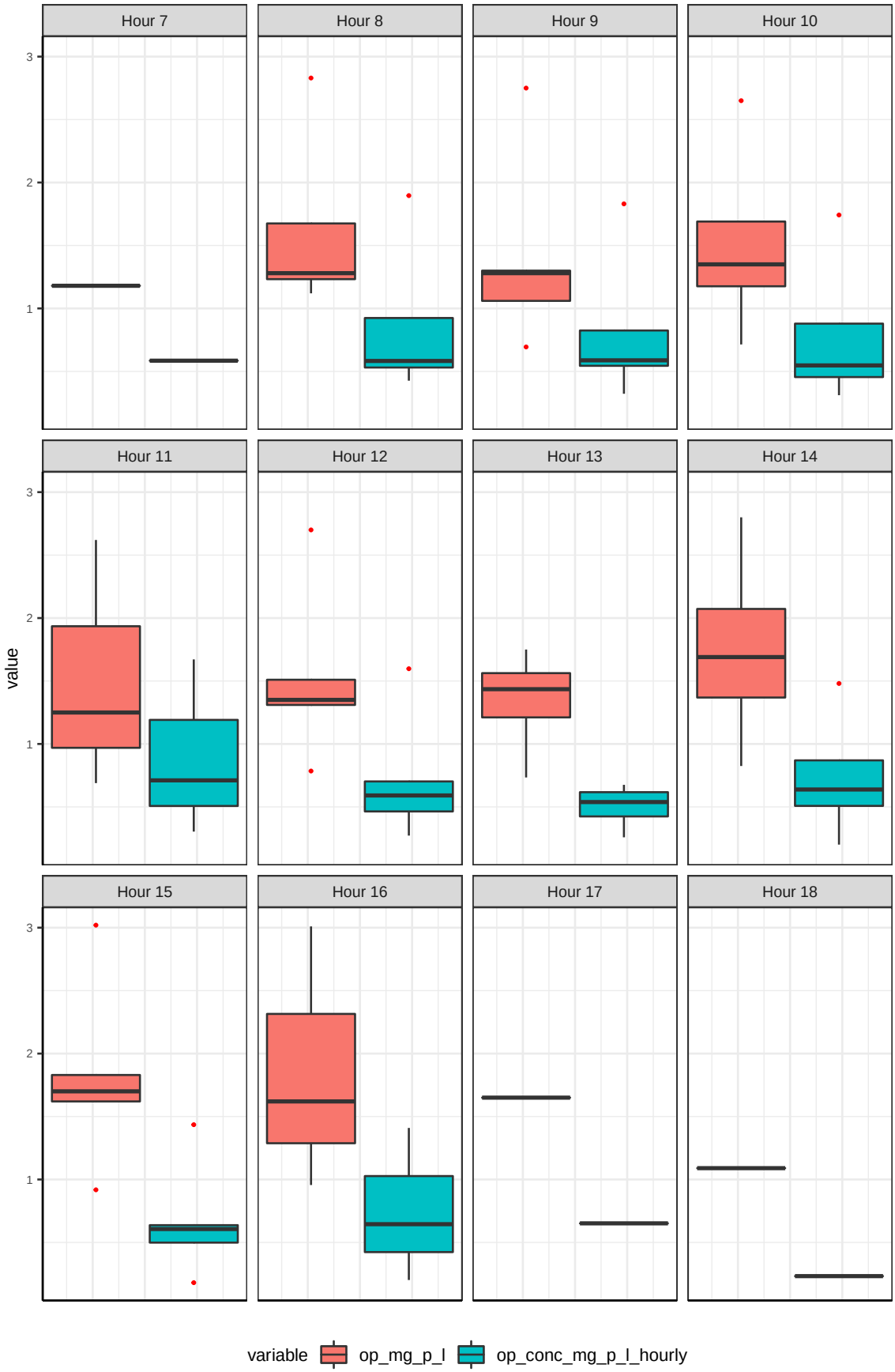

c Effluent vs. Dosing OP Conce
Alum Effluent vs. Dosing OP Conc
Effluent OP (mg/L)
3
2
1
2.0
2.5
3.0
Dosing OP (mg/L)
Effluent OP (mg/L)
1.5
1.0
0.5
0.0
1.0
1.5
2.0
2.5
3.0
Dosing OP (mg/L)

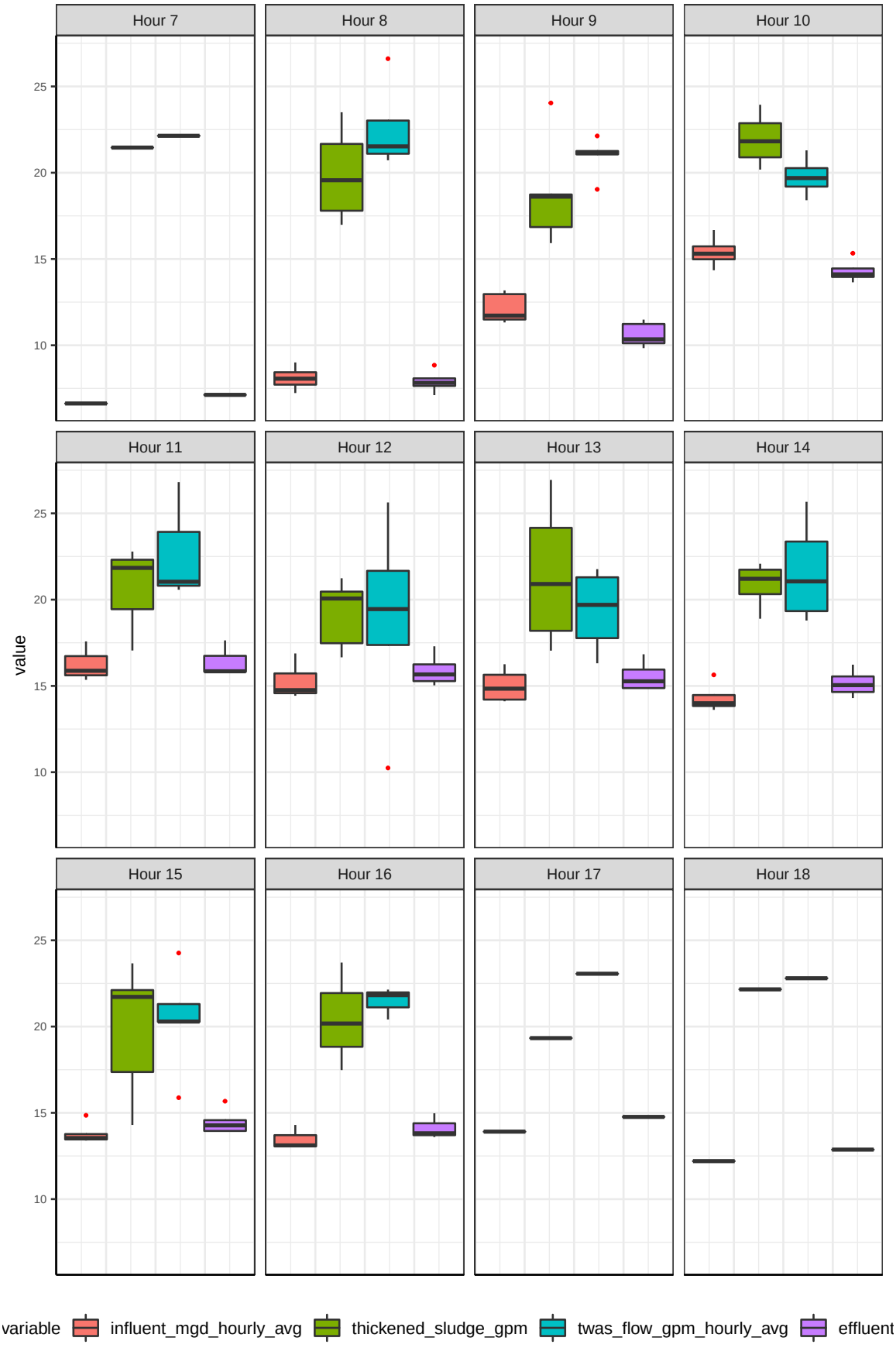

Hour 7
Hour 8
Hour 9
Hour 10
Hour 11
Hour 12
Hour 13
Hour 14
Hour 15
Hour 16
Hour 17
Hour 18
value
0
20000
40000
60000
variable
centrate_gal

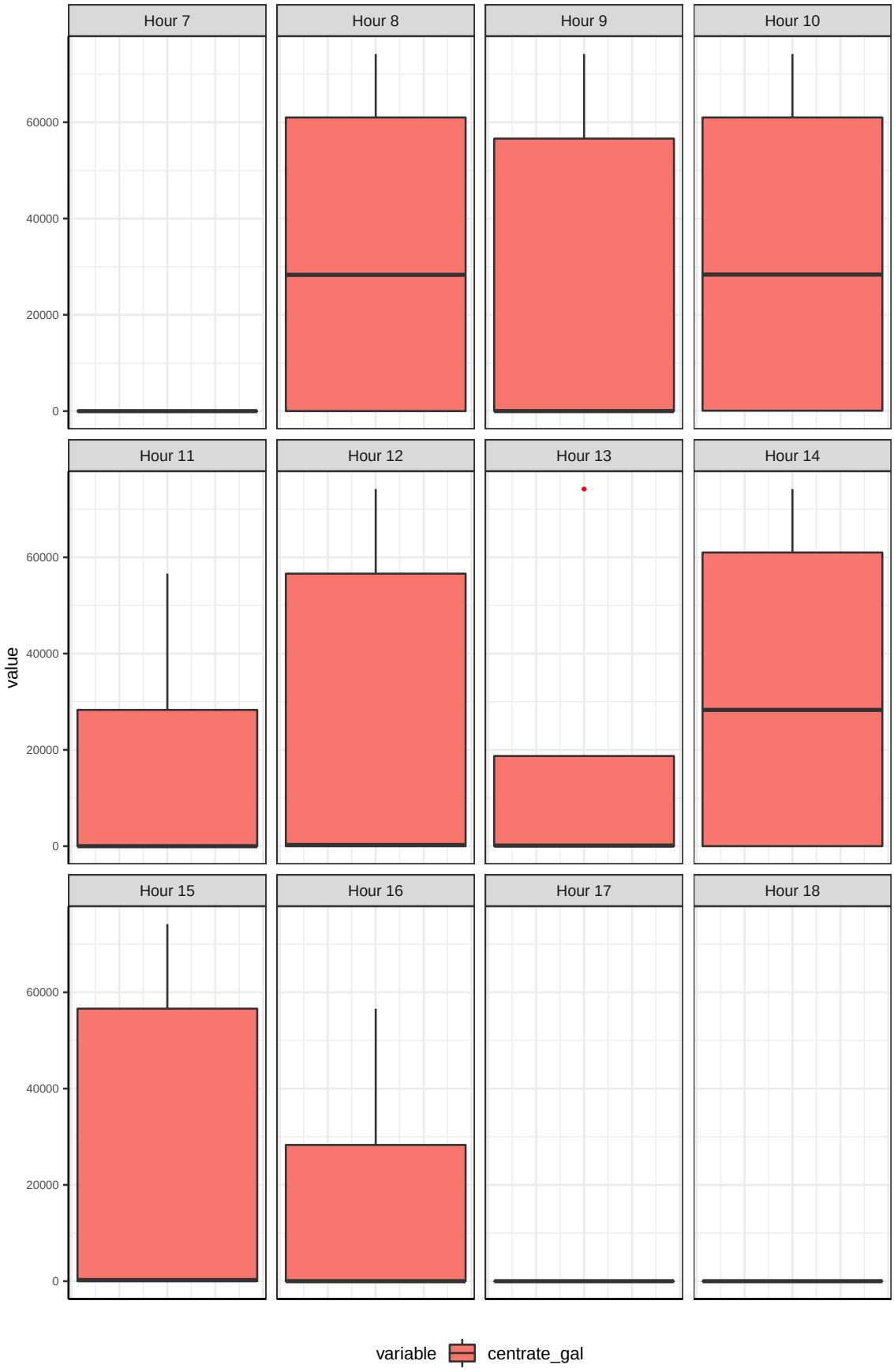

Hour 7
Hour 8
Hour 9
Hour 10
Hour 11
Hour 12
Hour 13
Hour 14
Hour 15
Hour 16
Hour 17
Hour 18
value
18
15
12
9
variable
effluent_mgd

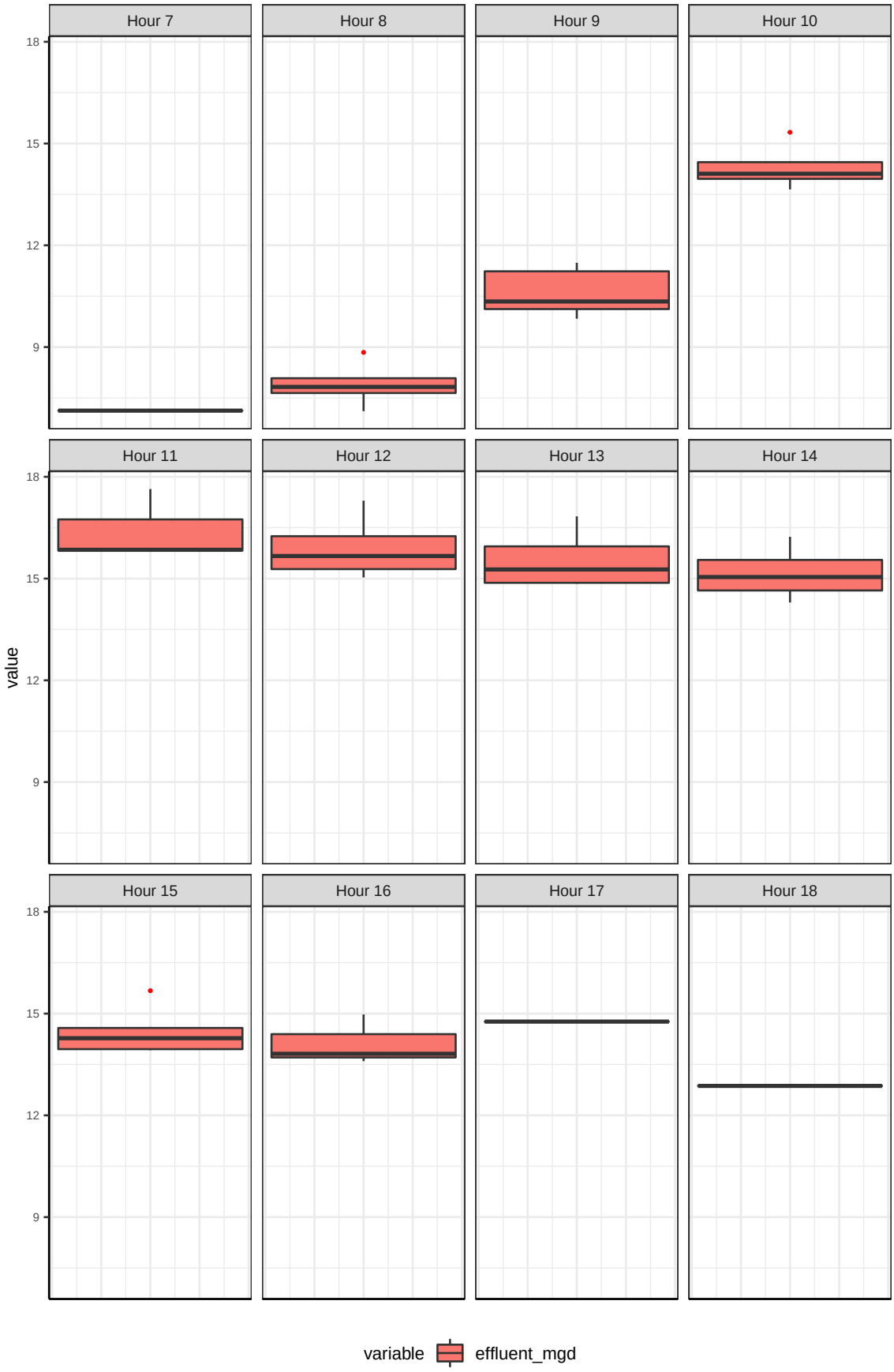

Hour 7
Hour 8
Hour 9
Hour 10
Hour 11
Hour 12
Hour 13
Hour 14
Hour 15
Hour 16
Hour 17
Hour 18
value
12000
10000
8000
6000
variable
abi_mgd
ras_gpm

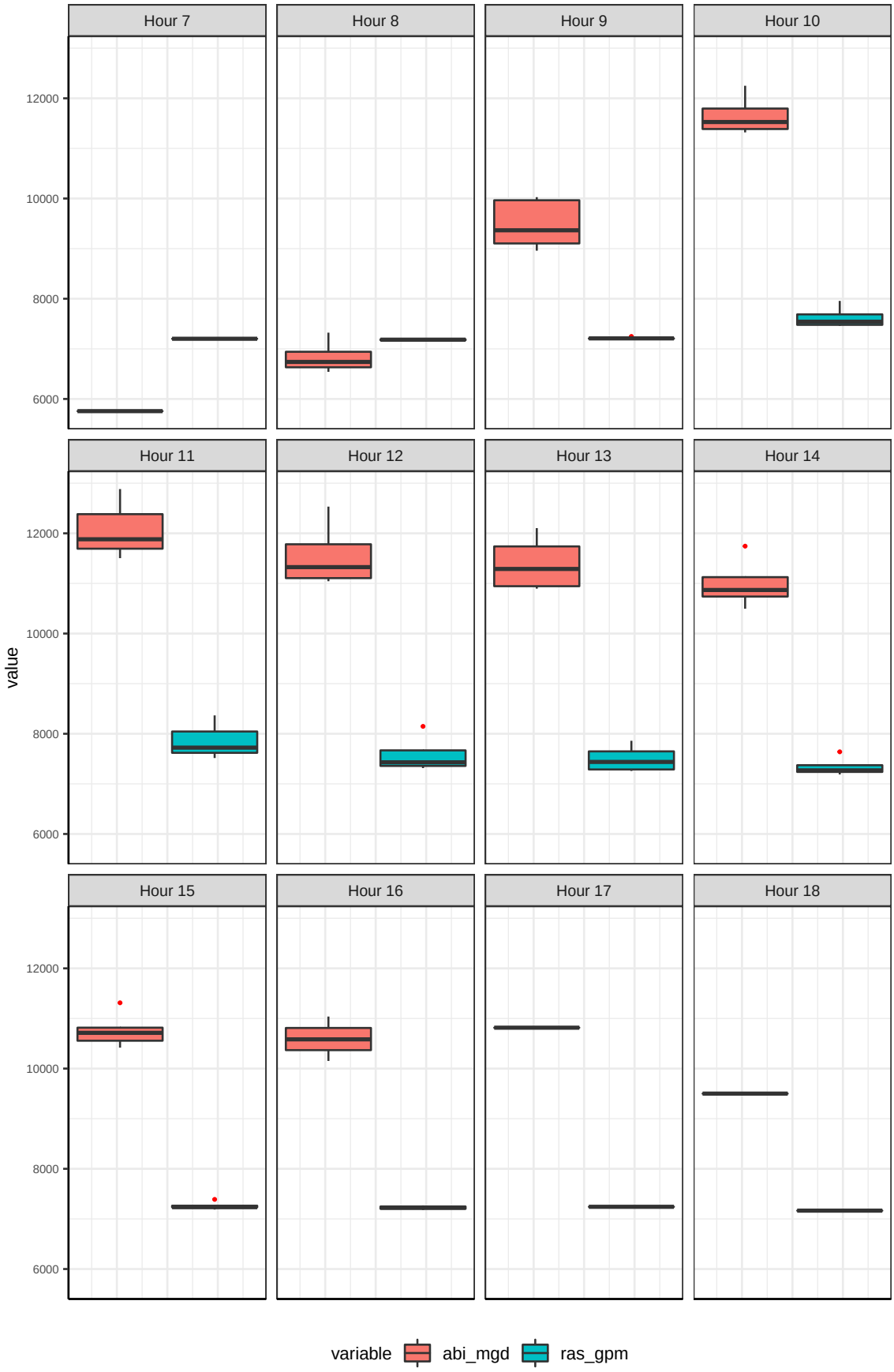

Hour 7
Hour 8
Hour 9
Hour 10
Hour 11
Hour 12
Hour 13
Hour 14
Hour 15
Hour 16
Hour 17
Hour 18
value
400
300
200
100
variable
primary_sludge_gmp_hourly_avg
mlws_flow_gpm

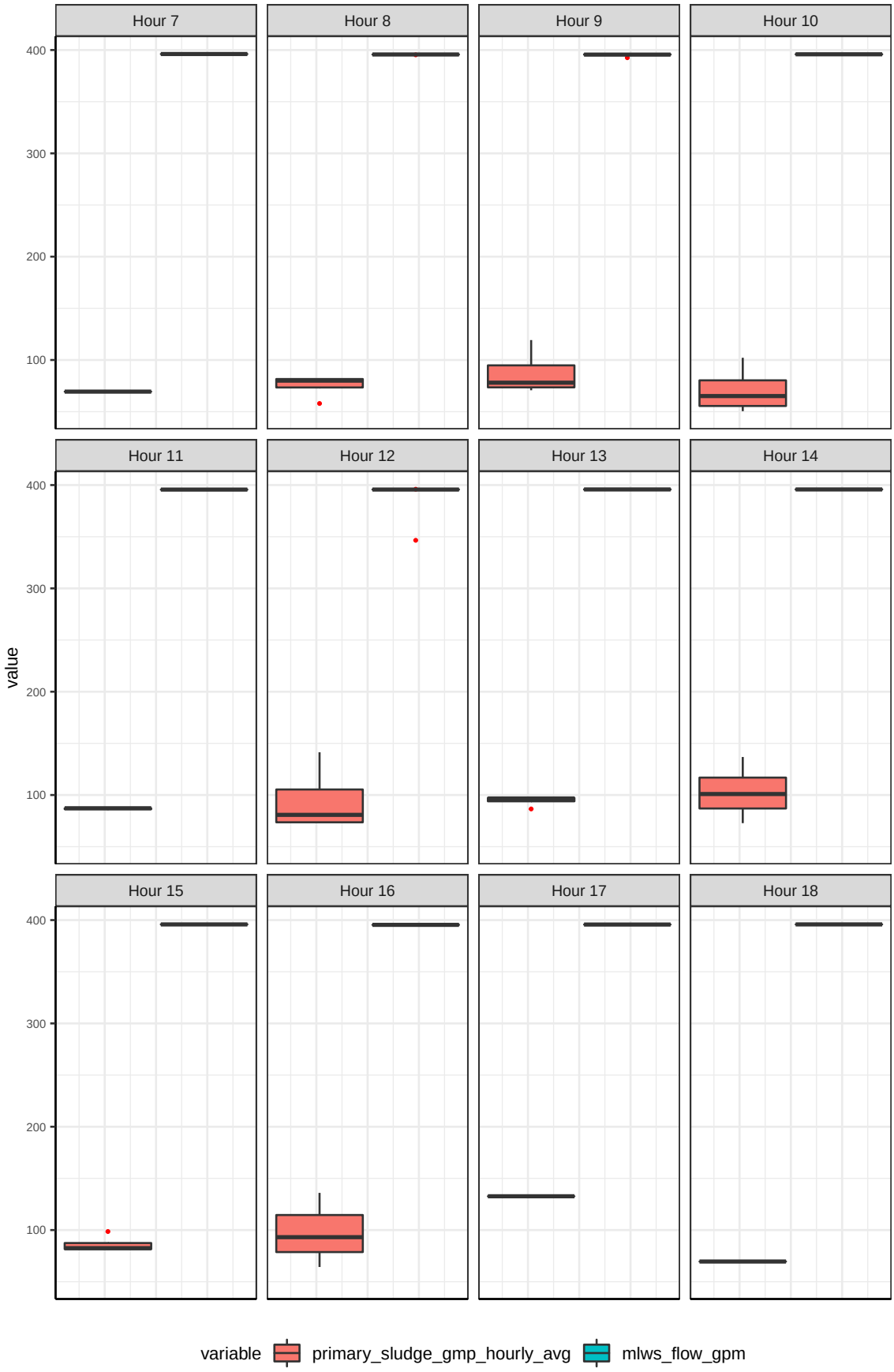

Hour 7
Hour 8
Hour 9
Hour 10
Hour 11
Hour 12
Hour 13
Hour 14
Hour 15
Hour 16
Hour 17
Hour 18
value
1
2
3
variable
op_mg_p_l
op_conc_mg_p_l_hourly

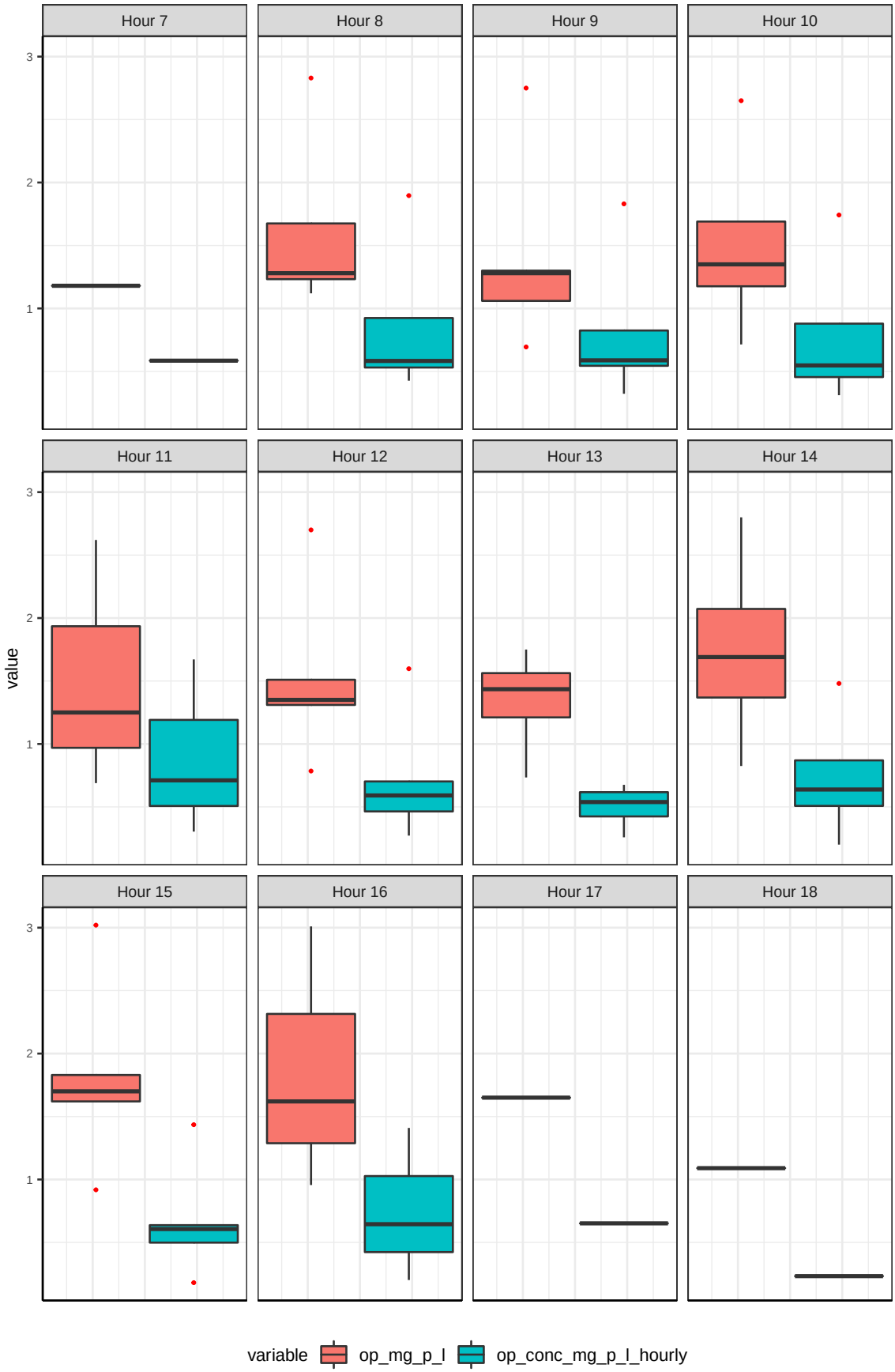

Hour 7
Hour 8
Hour 9
Hour 10
Hour 11
Hour 12
Hour 13
Hour 14
Hour 15
Hour 16
Hour 17
Hour 18
value
25
20
15
10
variable
influent_mgd_hourly_avg
thickened_sludge_gpm
twas_flow_gpm_hourly_avg
effluent

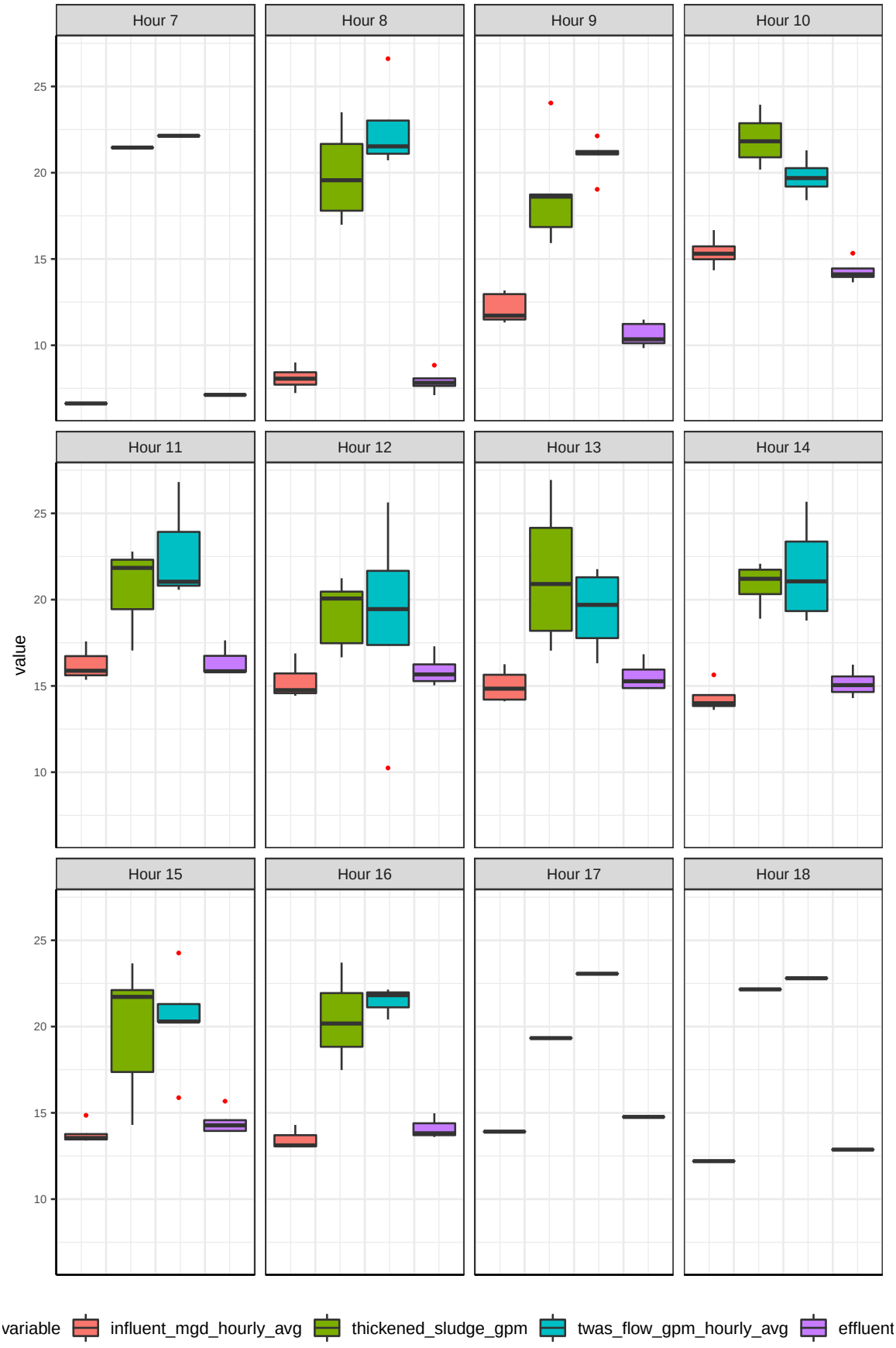

Hour 7
Hour 8
Hour 9
Hour 10
Hour 11
Hour 12
Hour 13
Hour 14
Hour 15
Hour 16
Hour 17
Hour 18
value
0
20000
40000
60000
variable
centrate_gal


Hour 7
Hour 8
Hour 9
Hour 10
Hour 11
Hour 12
Hour 13
Hour 14
Hour 15
Hour 16
Hour 17
Hour 18
value
18
15
12
9
variable
effluent_mgd


Hour 7
Hour 8
Hour 9
Hour 10
Hour 11
Hour 12
Hour 13
Hour 14
Hour 15
Hour 16
Hour 17
Hour 18
value
12000
10000
8000
6000
variable
abi_mgd
ras_gpm


Hour 7
Hour 8
Hour 9
Hour 10
Hour 11
Hour 12
Hour 13
Hour 14
Hour 15
Hour 16
Hour 17
Hour 18
value
400
300
200
100
variable
primary_sludge_gmp_hourly_avg
mlws_flow_gpm


Hour 7
Hour 8
Hour 9
Hour 10
Hour 11
Hour 12
Hour 13
Hour 14
Hour 15
Hour 16
Hour 17
Hour 18
value
1
2
3
variable
op_mg_p_l
op_conc_mg_p_l_hourly


Hour 7
Hour 8
Hour 9
Hour 10
Hour 11
Hour 12
Hour 13
Hour 14
Hour 15
Hour 16
Hour 17
Hour 18
value
25
20
15
10
variable
influent_mgd_hourly_avg
thickened_sludge_gpm
twas_flow_gpm_hourly_avg
effluent

Percent Change in Alum OP:

```
 [1] 33.00311 33.43780 34.26500 36.20696 40.83457 47.13253 52.47718 53.17811
 [9] 50.42695 49.50094 48.65866 62.18977 65.61906 68.02027 70.66944 52.75772
[17] 54.78610 56.80516 60.86528 65.85033 66.63732 66.86018 53.56720 56.47498
[25] 55.87934 65.33614 64.77530 75.68154 80.34528 78.90321 78.66838 66.97313
[33] 35.55792 43.13317 46.38196 50.71036 57.01887 60.67348 60.18477 60.50797
```

Percent Change in Ferric OP:

```
 [1]  6.974935  7.407288  4.515576  3.876140  8.157696 11.522592 30.938932
 [8] 34.725112 28.370414 34.159254 33.107300 33.556065 32.477654 40.178866
[15] 42.377844 42.869229 38.402886 39.726303 31.313979 33.963210 37.040344
[22] 37.107716 37.277368 48.279045 46.643583 51.663055 51.545022 45.043066
[29] 43.218335 53.120027 43.714583 53.457158
```

We perform a t-test to determine if they're performing differently

```
    Welch Two Sample t-test

data:  alum_diff_pct and ferr_diff_pct
t = 6.8612, df = 61.462, p-value = 3.881e-09
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 15.98862 29.13849
sample estimates:
mean of x mean of y
 56.52389  33.96033
```