ListofParams

WB

11/21/2019

Figuring out which parameters we want to include for the feature importance tests

Preliminaries

```
#load the column headers
params <-  c("TempMean",
             "SpCondMean",
             "pHMean",
             "ODOMean",
             "ChlMean",
             "TurbMean",
             "DeltaDOMean",
             "DeltaTempMean",
             "PCMean1",
             "PCMean2",
             "PCMean3",
             "PCMean4",
             "PCMean5",
             "PCMean6",
             "PCMean7",
             "ws",
             "wd",
             "solar",
             "temp",
             "discharge")
```

Set 1

```
set1_0 <- c(params[2:5], params[8:9], params[16:20])
set1_1 <- c(params[2:5], params[8], params[10], params[16:20])
set1_2 <- c(params[2:5], params[8], params[11], params[16:20])
set1_3 <- c(params[2:5], params[8], params[12], params[16:20])
set1_4 <- c(params[2:5], params[8], params[13], params[16:20])
set1_5 <- c(params[2:5], params[8], params[14], params[16:20])
set1_6 <- c(params[2:5], params[8], params[15], params[16:20])
```

Set 1 includes: *SpCondMean, pHMean, ODOMean, ChlMean, DeltaTempMean, ws, wd, solar, temp, discharge*

Set 1 does not include:

- *TempMean* (water temp) –> going to use air temp instead
- *TurbMean* (turbidity) –> too closely correlated with PCMeans
- *DeltaDOMean* (dissolved oxygen stratification) –> going to use deltaTempMean instead which is a better proxy for stratification (DeltaDO is affected by biological activity too much)

The subsequent runs in **Set 1** have the lag times incorporated for PC:

- Predicting PC for that day (Set1_0): *SpCondMean, pHMean, ODOMean, ChlMean, DeltaTempMean, PCMean1, ws, wd, solar, temp, discharge*
- Predicting PC 1 day out (Set1_1): *SpCondMean, pHMean, ODOMean, ChlMean, DeltaTempMean, PCMean2, ws, wd, solar, temp, discharge*
- Predicting PC 2 days out (Set1_2): *SpCondMean, pHMean, ODOMean, ChlMean, DeltaTempMean, PCMean3, ws, wd, solar, temp, discharge*
- Predicting PC 3 days out (Set1_3): *SpCondMean, pHMean, ODOMean, ChlMean, DeltaTempMean, PCMean4, ws, wd, solar, temp, discharge*
- Predicting PC 4 days out (Set1_4): *SpCondMean, pHMean, ODOMean, ChlMean, DeltaTempMean, PCMean5, ws, wd, solar, temp, discharge*
- Predicting PC 5 days out (Set1_5): *SpCondMean, pHMean, ODOMean, ChlMean, DeltaTempMean, PCMean6, ws, wd, solar, temp, discharge*
- Predicting PC 6 days out (Set1_6): *SpCondMean, pHMean, ODOMean, ChlMean, DeltaTempMean, PCMean7, ws, wd, solar, temp, discharge*

Set 2

```
set2_0 <- c(params[3:5], params[8:9], params[16], params[18:20])
set2_1 <- c(params[3:5], params[8], params[10], params[16], params[18:20])
set2_2 <- c(params[3:5], params[8], params[11], params[16], params[18:20])
set2_3 <- c(params[3:5], params[8], params[12], params[16], params[18:20])
set2_4 <- c(params[3:5], params[8], params[13], params[16], params[18:20])
set2_5 <- c(params[3:5], params[8], params[14], params[16], params[18:20])
set2_6 <- c(params[3:5], params[8], params[15], params[16], params[18:20])
```

Set 2 includes: *pHMean, ODOMean, ChlMean*, *DeltaTempMean*, *ws*, *solar, temp, discharge*

Set 2 does not include:

- *SpCondMean* (specific conductivity) –> i don't think this actually affects PC and it might be giving a false importance signal
- *wd* (wind direction) –> i'm not sure it's useful to have wind direction in as a continuous variable so it might cloud other relationships...
- *TempMean*(water temp) –> going to use air temp instead
- *TurbMean* (turbidity) –> too closely correlated with PCMeans
- *DeltaDOMean* (dissolved oxygen stratification) –> going to use deltaTempMean instead which is a better proxy for stratification (DeltaDO is affected by biological activity too much)

Runs in **Set 2** have the lag times incorporated for PC:

- Predicting PC for that day (Set2_0): *pHMean, ODOMean, ChlMean, DeltaTempMean, PCMean1, ws, solar, temp, discharge*
- Predicting PC 1 day out (Set1_1): *pHMean, ODOMean, ChlMean, DeltaTempMean, PCMean2, ws, solar, temp, discharge*
- Predicting PC 2 days out (Set1_2): *pHMean, ODOMean, ChlMean, DeltaTempMean, PCMean3, ws, solar, temp, discharge*
- Predicting PC 3 days out (Set1_3): *pHMean, ODOMean, ChlMean, DeltaTempMean, PCMean4, ws, solar, temp, discharge*
- Predicting PC 4 days out (Set1_4): *pHMean, ODOMean, ChlMean, DeltaTempMean, PCMean5, ws, solar, temp, discharge*
- Predicting PC 5 days out (Set1_5): *pHMean, ODOMean, ChlMean, DeltaTempMean, PCMean6, ws, solar, temp, discharge*
- Predicting PC 6 days out (Set1_6): *pHMean, ODOMean, ChlMean, DeltaTempMean, PCMean7, ws, solar, temp, discharge*