

clip-max-ignore-zeros-asymmetric

August 13, 2020

```
[1]: # rank the obtained results using the *.log files
import os
import pandas as pd
import numpy as np
```

```
[2]: source = "7b"
targetdir = '../..data/' + source + "/"
filelist = sorted(os.listdir(targetdir))
```

```
[3]: filelist
```

```
[3]: ['1.data', '2.data', '3.data', '4.data', '5.data', '6.data', '7.data']
```

```
[4]: # Create dataframe from files
df = pd.DataFrame()

for file in filelist:
    filename = targetdir+file
    col_name = [file]
    temp_df = pd.read_csv(filename,names=col_name)
    df = pd.concat([df, temp_df], axis=1)

# Look at the data
df.head()
```

```
[4]:
```

	1.data	2.data	3.data	4.data	5.data	6.data	7.data
0	1.0	0.85	1.00	1.00	0.64	1.00	1.0
1	-1.0	-1.00	-1.00	-1.00	-1.00	-1.00	-1.0
2	-1.0	-1.00	-1.00	-1.00	-1.00	-1.00	-1.0
3	0.8	0.11	0.09	0.12	0.07	0.07	1.0
4	-1.0	0.11	0.09	0.12	0.07	0.07	-1.0

```
[5]: # Clip values > 1 with 1 and ignore 0s
df.mask(df > 1, 1, inplace=True)
df.mask(df <= 0, np.NaN, inplace=True)
```

```
[6]: # Count NaN values
df.isna().sum()
```

```
[6]: 1.data    369
     2.data    315
     3.data    351
     4.data    342
     5.data    324
     6.data    342
     7.data    369
     dtype: int64
```

```
[7]: df
```

```
[7]:
```

	1.data	2.data	3.data	4.data	5.data	6.data	7.data
0	1.0	0.85	1.00	1.00	0.64	1.00	1.0
1	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	0.8	0.11	0.09	0.12	0.07	0.07	1.0
4	NaN	0.11	0.09	0.12	0.07	0.07	NaN
..
724	1.0	1.00	NaN	NaN	1.00	NaN	1.0
725	NaN	NaN	NaN	NaN	NaN	NaN	NaN
726	1.0	0.85	1.00	1.00	0.64	0.58	1.0
727	1.0	0.85	1.00	1.00	0.64	0.58	1.0
728	1.0	0.85	1.00	1.00	0.64	1.00	1.0

```
[729 rows x 7 columns]
```

```
[8]: # Save processed dataframe as csv file
df.to_csv("../data/processed/asymmetric/" + source + ".csv", index=False)
```

```
[9]: # Creating ranked dataframe
ranked_df = pd.DataFrame()
stats_df = pd.DataFrame()
```

```
[10]: # going through every column
for column in df:
    wwtp = column[0]

    # In every column, drop na values
    asym_column = df[column].dropna()

    # and calculate individual tao
    tao = len(asym_column)

    # calculate mean
```

```

avg_eff = round(asm_column.mean(),3)

# calculate max
max_eff = round(asm_column.max(),3)

# calculate min
min_eff = round(asm_column.min(),3)

# calculate amplitude
amplitude = round((max_eff - min_eff)*100,2)

amp_str = "Amplitude (max-min)(%)"

# print stats results
print("WWTP",wwtp,
      "\nMean =",avg_eff,"Maximum =",max_eff,"Minimum =",min_eff,
      ↪amp_str,"=",amplitude)

stats_df = stats_df.append({ 'WWTP': wwtp,
                             "Mean": avg_eff, "Maximum" : max_eff, "Minimum": min_eff, amp_str:
      ↪amplitude},ignore_index=True)

# Calculating Sk sum of factors
Sk = round(asm_column.sum(),3)

# Calculating ek sum of factors of 1 (or above if errors in calculation)
ek = asm_column >= 1
ek = ek.sum()

# Calculating R1k ek/tao
R1k = round(ek/tao,3)

# Calculate R2k
if tao != ek:
    R2k = (Sk - ek)/(tao - ek)
elif R1k == 1:
    R2k = 0

R2k = round(R2k,3)

# Printing results
print("tao =",tao,"| ek =",ek,"| R1k =",R1k, "| Sk =",Sk, "| R2k
      ↪=",R2k,"\n")

# Populate ranking dataframe using pd.df.append
# Using unicode to name columns with super and subscripts
R1k_col = 'R\u00B9\u2096\u2080'

```

```
R2k_col = 'R\u00B2\u2096\u2080'
ranked_df = ranked_df.append({ R2k_col:R2k, R1k_col: R1k, 'WWTP':\u2192wwtp},ignore_index=True)
```

WWTP 1

Mean = 0.731 Maximum = 1.0 Minimum = 0.08 Amplitude (max-min)(%) = 92.0
 tao = 360 | ek = 188 | R1k = 0.522 | Sk = 263.26 | R2k = 0.438

WWTP 2

Mean = 0.477 Maximum = 1.0 Minimum = 0.01 Amplitude (max-min)(%) = 99.0
 tao = 414 | ek = 95 | R1k = 0.229 | Sk = 197.51 | R2k = 0.321

WWTP 3

Mean = 0.486 Maximum = 1.0 Minimum = 0.01 Amplitude (max-min)(%) = 99.0
 tao = 378 | ek = 122 | R1k = 0.323 | Sk = 183.74 | R2k = 0.241

WWTP 4

Mean = 0.496 Maximum = 1.0 Minimum = 0.01 Amplitude (max-min)(%) = 99.0
 tao = 387 | ek = 131 | R1k = 0.339 | Sk = 191.82 | R2k = 0.238

WWTP 5

Mean = 0.411 Maximum = 1.0 Minimum = 0.01 Amplitude (max-min)(%) = 99.0
 tao = 405 | ek = 86 | R1k = 0.212 | Sk = 166.54 | R2k = 0.252

WWTP 6

Mean = 0.436 Maximum = 1.0 Minimum = 0.01 Amplitude (max-min)(%) = 99.0
 tao = 387 | ek = 107 | R1k = 0.276 | Sk = 168.69 | R2k = 0.22

WWTP 7

Mean = 0.768 Maximum = 1.0 Minimum = 0.1 Amplitude (max-min)(%) = 90.0
 tao = 360 | ek = 241 | R1k = 0.669 | Sk = 276.45 | R2k = 0.298

0.1 Ranking of WWTP

```
[11]: # Reorder columns to be usable as a results table
ranked_df = ranked_df.reindex(columns=['WWTP',R1k_col, R2k_col])
```

```
[12]: ranked_df
```

```
[12]:
```

	WWTP	R ¹	R ²
0	1	0.522	0.438
1	2	0.229	0.321
2	3	0.323	0.241
3	4	0.339	0.238
4	5	0.212	0.252

```

5    6    0.276    0.220
6    7    0.669    0.298

```

```

[13]: import os

path = "../../results/" + source + "/asymmetric"

# Save rankings dataframe as csv file

try:
    ranked_df.to_csv(path + "/ranking.csv",index=False)
    print("Save succesful")
except:
    print("Creating folder and saving")
    os.mkdir(path)
    ranked_df.to_csv(path + "/ranking.csv",index=False)

```

Save succesful

0.2 Calculate Descriptive Statistics

```

[14]: # Calculate the mean of every column
mean_mean = round(stats_df.Mean.mean(),3)
mean_max = round(stats_df.Maximum.mean(),3)
mean_min = round(stats_df.Minimum.mean(),3)
mean_amp = round(stats_df[amp_str].mean(),3)

```

```

[15]: # Add means to stats dataframe
stats_df = stats_df.append({ 'WWTP': "Mean", "Mean" : mean_mean, "Maximum" : ↵
↵mean_max,
                             "Minimum" : mean_min, amp_str : ↵
↵mean_amp},ignore_index=True)

```

```

[16]: # Calculate the standard deviation of every column
sd_mean = round(stats_df.Mean.std(),3)
sd_max = round(stats_df.Maximum.std(),3)
sd_min = round(stats_df.Minimum.std(),3)
sd_amp = round(stats_df[amp_str].std(),3)

```

```

[17]: # Add means to stats dataframe
stats_df = stats_df.append({ 'WWTP': "SD", "Mean" : sd_mean, "Maximum" : sd_max,
                             "Minimum" : sd_min, amp_str : ↵
↵sd_amp},ignore_index=True)

```

```

[18]: # Reorder columns

```

```
stats_df = stats_df.reindex(columns=["WWTP", "Mean", "Maximum", "Minimum", "Amplitude (max-min) (%)", "SD"])
```

```
[19]: stats_df
```

```
[19]:
```

	WWTP	Mean	Maximum	Minimum	Amplitude (max-min) (%)
0	1	0.731	1.0	0.080	92.000
1	2	0.477	1.0	0.010	99.000
2	3	0.486	1.0	0.010	99.000
3	4	0.496	1.0	0.010	99.000
4	5	0.411	1.0	0.010	99.000
5	6	0.436	1.0	0.010	99.000
6	7	0.768	1.0	0.100	90.000
7	Mean	0.544	1.0	0.033	96.714
8	SD	0.133	0.0	0.037	3.653

```
[20]: # Save statistics dataframe as csv file
stats_df.to_csv(path + "/statistics.csv", index=False)
```

```
[21]: # Convert Jupyter Notebook to PDF LaTeX file
!jupyter-nbconvert --to pdf "clip-max-ignore-zeros-asymmetric" --output-dir "../results/7b/asymmetric"
```

```
[NbConvertApp] Converting notebook clip-max-ignore-zeros-asymmetric.ipynb to pdf
[NbConvertApp] Writing 44325 bytes to ./notebook.tex
[NbConvertApp] Building PDF
[NbConvertApp] Running xelatex 3 times: ['xelatex', './notebook.tex', '-quiet']
[NbConvertApp] Running bibtex 1 time: ['bibtex', './notebook']
[NbConvertApp] WARNING | bibtex had problems, most likely because there were no citations
[NbConvertApp] PDF successfully created
[NbConvertApp] Writing 47217 bytes to ../../results/7a/asymmetric/clip-max-ignore-zeros-asymmetric.pdf
```

```
[ ]:
```