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 = "8b"
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	1.00	0.92	1.00	1.00	0.68	1.00	1.0
1	1.02	0.92	0.00	0.00	0.68	0.00	0.0
2	1.05	0.92	0.00	0.00	0.68	0.00	0.0
3	0.99	0.84	0.96	0.91	0.63	0.96	1.0
4	0.99	0.84	0.96	0.91	0.63	0.00	0.0

[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      0
2.data      0
3.data    324
4.data    324
5.data      0
6.data    297
7.data    324
dtype: int64
```

[7]:
```
df
```

[7]:

	1.data	2.data	3.data	4.data	5.data	6.data	7.data
0	1.00	0.92	1.00	1.00	0.68	1.00	1.0
1	1.00	0.92	NaN	NaN	0.68	NaN	NaN
2	1.00	0.92	NaN	NaN	0.68	NaN	NaN
3	0.99	0.84	0.96	0.91	0.63	0.96	1.0
4	0.99	0.84	0.96	0.91	0.63	NaN	NaN
..	…	…	…	…	…	…	…
724	1.00	0.86	NaN	NaN	0.64	0.94	1.0
725	1.00	0.86	NaN	NaN	0.64	NaN	NaN
726	1.00	0.79	1.00	1.00	0.58	0.54	1.0
727	1.00	0.79	1.00	1.00	0.58	0.54	1.0
728	1.00	0.79	1.00	1.00	0.58	0.54	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(asym_column.mean(),3)

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

    # calculate min
    min_eff = round(asym_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(asym_column.sum(),3)

    # Calculating ek sum of factors of 1 (or above if errors in calculation)
    ek = asym_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':␣
↪wwtp},ignore_index=True)
```

```
WWTP 1
Mean = 0.996 Maximum = 1.0 Minimum = 0.98 Amplitude (max-min)(%) = 2.0
tao = 729 | ek = 513 | R1k = 0.704 | Sk = 726.39 | R2k = 0.988

WWTP 2
Mean = 0.854 Maximum = 0.95 Minimum = 0.77 Amplitude (max-min)(%) = 18.0
tao = 729 | ek = 0 | R1k = 0.0 | Sk = 622.62 | R2k = 0.854

WWTP 3
Mean = 0.961 Maximum = 1.0 Minimum = 0.88 Amplitude (max-min)(%) = 12.0
tao = 405 | ek = 108 | R1k = 0.267 | Sk = 389.28 | R2k = 0.947

WWTP 4
Mean = 0.933 Maximum = 1.0 Minimum = 0.83 Amplitude (max-min)(%) = 17.0
tao = 405 | ek = 108 | R1k = 0.267 | Sk = 377.91 | R2k = 0.909

WWTP 5
Mean = 0.637 Maximum = 0.71 Minimum = 0.57 Amplitude (max-min)(%) = 14.0
tao = 729 | ek = 0 | R1k = 0.0 | Sk = 464.22 | R2k = 0.637

WWTP 6
Mean = 0.745 Maximum = 1.0 Minimum = 0.53 Amplitude (max-min)(%) = 47.0
tao = 432 | ek = 56 | R1k = 0.13 | Sk = 322.03 | R2k = 0.708

WWTP 7
Mean = 0.997 Maximum = 1.0 Minimum = 0.99 Amplitude (max-min)(%) = 1.0
tao = 405 | ek = 291 | R1k = 0.719 | Sk = 403.86 | R2k = 0.99
```

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	0.704	0.988
1	2	0.000	0.854
2	3	0.267	0.947
3	4	0.267	0.909
4	5	0.000	0.637

```
5    6  0.130  0.708
6    7  0.719  0.990
```

[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",␣
↪amp_str])
```

[19]:
```
stats_df
```

[19]:

	WWTP	Mean	Maximum	Minimum	Amplitude (max-min)(%)
0	1	0.996	1.000	0.980	2.000
1	2	0.854	0.950	0.770	18.000
2	3	0.961	1.000	0.880	12.000
3	4	0.933	1.000	0.830	17.000
4	5	0.637	0.710	0.570	14.000
5	6	0.745	1.000	0.530	47.000
6	7	0.997	1.000	0.990	1.000
7	Mean	0.875	0.951	0.793	15.857
8	SD	0.128	0.100	0.170	14.177

[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/8b/asymmetric"
```

```
[NbConvertApp] Converting notebook clip-max-ignore-zeros-asymmetric.ipynb to pdf
[NbConvertApp] Writing 44348 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 47130 bytes to ../../results/8a/asymmetric/clip-max-
ignore-zeros-asymmetric.pdf
```

[ ]:
```

```