ds1_nematoda_limpieza_de_datos

December 14, 2020

Limpieza de datos

[1]:
```
import pandas as pd
import seaborn as sns
import numpy as np
import os
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings("ignore")
%matplotlib inline
from mlxtend.preprocessing import standardize
from scipy import stats
```

1 Declaración de variables

[2]:
```
organismo ="nematoda"
dataset = 1
nombre = ("ds" + str(dataset) + "_" + str(organismo))
nombre2 = (str(organismo)+ " dataset " + str(dataset))
r2 = ("Datos/resultados/"+ str(organismo) + "/" + str(nombre) + "/
 ↪transformaciones/sin_filtrar")
r3 = ("Datos/resultados/"+ str(organismo) + "/" + str(nombre) + "/
 ↪transformaciones/sin_atipicos")

nom1 = ("/ds" + str(dataset) + "_AAC_efectores_" + str(organismo) + ".txt")
nom2 = ("/ds" + str(dataset) + "_ACC_hidro_mass_efectores_" + str(organismo) +␣
 ↪".txt")
nom3 = ("/ds" + str(dataset) + "_ACC_mass_efectores_" + str(organismo) + ".txt")
nom4 = ("/ds" + str(dataset) + "_ACC_hidro_efectores_" + str(organismo) + ".
 ↪txt")
nom5 = ("/ds" + str(dataset) + "_PseAAC_hidro_mass_efectores_" + str(organismo)␣
 ↪+ ".txt")
nom6 = ("/ds" + str(dataset) + "_PseAAC_mass_efectores_" + str(organismo) + ".
 ↪txt")
nom7 = ("/ds" + str(dataset) + "_PseAAC_hidro_efectores_" + str(organismo) + ".
 ↪txt")
```

```
nom8 = ("/ds" + str(dataset) + "_AAC_no_efectores_" + str(organismo) + ".txt")
nom9 = ("/ds" + str(dataset) + "_ACC_hidro_mass_no_efectores_" + str(organismo)␣
 ↪+ ".txt")
nom10 = ("/ds" + str(dataset) + "_ACC_mass_no_efectores_" + str(organismo) + ".
 ↪txt")
nom11 = ("/ds" + str(dataset) + "_ACC_hidro_no_efectores_" + str(organismo) + ".
 ↪txt")
nom12 = ("/ds" + str(dataset) + "_PseAAC_hidro_mass_no_efectores_" +␣
 ↪str(organismo) + ".txt")
nom13 = ("/ds" + str(dataset) + "_PseAAC_mass_no_efectores_" + str(organismo) +␣
 ↪".txt")
nom14 = ("/ds" + str(dataset) + "_PseAAC_hidro_no_efectores_" + str(organismo)␣
 ↪+ ".txt")

#Efectores
AAC_efec= pd.read_csv(str(r2) + str(nom1), header=None,prefix='X',sep=',')
ACC_hidro_mass_efec = pd.read_csv(str(r2) + str(nom2),␣
 ↪header=None,prefix='X',sep=',')
ACC_mass_efec = pd.read_csv(str(r2) + str(nom3), header=None,prefix='X',sep=',')
ACC_hidro_efec = pd.read_csv(str(r2) + str(nom4),␣
 ↪header=None,prefix='X',sep=',')
PseAAC_hidro_mass_efec = pd.read_csv(str(r2) +str(nom5),␣
 ↪header=None,prefix='X',sep=',')
PseAAC_mass_efec = pd.read_csv(str(r2) + str(nom6),␣
 ↪header=None,prefix='X',sep=',')
PseAAC_hidro_efec = pd.read_csv(str(r2) + str(nom7),␣
 ↪header=None,prefix='X',sep=',')

#No efectores
AAC_no_efec= pd.read_csv(str(r2) + str(nom8), header=None,prefix='X',sep=',')
ACC_hidro_mass_no_efec =pd.read_csv(str(r2) + str(nom9),␣
 ↪header=None,prefix='X',sep=',')
ACC_mass_no_efec =pd.read_csv(str(r2) + str(nom10),␣
 ↪header=None,prefix='X',sep=',')
ACC_hidro_no_efec =pd.read_csv(str(r2) + str(nom11),␣
 ↪header=None,prefix='X',sep=',')
PseAAC_hidro_mass_no_efec =pd.read_csv(str(r2) + str(nom12),␣
 ↪header=None,prefix='X',sep=',')
PseAAC_mass_no_efec =pd.read_csv(str(r2) + str(nom13),␣
 ↪header=None,prefix='X',sep=',')
PseAAC_hidro_no_efec =pd.read_csv(str(r2) + str(nom14),␣
 ↪header=None,prefix='X',sep=',')
```

2 Composición de aminoácidos (AAC)

[3]:
```
transf = "Composición de aminoácidos (AAC) "
etiq="efectores "
estado = "con valores atípicos.\n"
df=""

for etiq in "efectores", "no_efectores":
    titulo = (str(transf) + str(etiq) + " " + str(nombre2) + ", " +str(estado))
    print (str(etiq))

    if etiq == "efectores":
        df=AAC_efec

    if etiq == "no_efectores":
        df=AAC_no_efec

    #del df['X20']
    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")

    #Gráfica de caja y bigotes
    sns.set(style="whitegrid")
    fig , ax = plt.subplots(figsize=(14,7))
    ax = sns.boxplot(data=df)
    ax.set_title(organismo +' '+str(etiq) +" dataset "+ str(dataset)+"␣
 ↪"+str(transf)+" "+str(estado))
```

```
efectores
Composición de aminoácidos (AAC) efectores nematoda dataset 1, con valores
atípicos.
Valores del documento csv.

         X0     X1     X2     X3     X4     X5     X6      X7     X8      X9  \
0     4.457  8.357  3.621  6.407  2.228  5.571  4.178  16.713  1.950   4.457
1    10.423  4.560  3.583  3.257  0.977  3.583  1.954   4.235  1.629   8.795
2    11.526  2.804  4.984  5.296  4.050  5.296  2.804   3.738  2.492   5.607
3     3.462  6.154  4.615  3.846  3.462  6.154  4.615   9.231  3.077   4.231
4     4.202  7.563  5.882  2.521  2.521  6.723  8.403   8.403  2.521   3.361
..      …      …      …      …      …      …      …       …      …       …
495   7.226  5.128  3.030  5.711  1.515  7.692  3.147   6.061  2.448   5.245
496   8.929  5.357  5.357  7.143  1.786  5.357  0.000   1.786  3.571   5.357
497   4.044  5.699  4.596  6.985  1.654  7.537  6.618   4.412  2.206   6.250
498   6.757  2.703  8.108  4.054  0.000  6.757  1.351   1.351  0.000  12.162
```

```
499   2.041  5.102  0.000  8.163  1.020  10.204  2.041   4.082  8.163   6.122

     …     X11    X12    X13     X14     X15    X16    X17    X18     X19  \
0    …   5.014  1.114  1.671  15.042   4.735  3.621  0.557  2.507   4.457
1    …   3.257  2.932  7.166   3.909   7.166  6.840  2.280  4.560   8.795
2    …   6.542  2.181  4.673   4.361   6.542  9.657  1.558  2.181   5.607
3    …   5.769  3.846  3.462   3.846   7.692  8.077  1.923  1.154   6.923
4    …   4.202  3.361  5.042   4.202   3.361  6.723  1.681  8.403   3.361
..   …     …    …    …     …       …      …    …      …      …
495  …   8.042  2.098  3.497   5.478   5.944  5.944  2.214  3.380   7.226
496  …   3.571  3.571  5.357   8.929  10.714  3.571  0.000  3.571  10.714
497  …   4.963  2.574  5.147   4.228   5.331  4.044  0.735  5.699   6.066
498  …  10.811  2.703  5.405   6.757   4.054  6.757  2.703  4.054   1.351
499  …   5.102  5.102  1.020   3.061   6.122  5.102  2.041  4.082  10.204

           X20
0     efectores
1     efectores
2     efectores
3     efectores
4     efectores
..          …
495   efectores
496   efectores
497   efectores
498   efectores
499   efectores

[500 rows x 21 columns]
```

Composición de aminoácidos (AAC) efectores nematoda dataset 1, con valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     7.162798    6.049354    4.348568    5.287820    2.373692    6.364204
std      2.928374    2.623532    2.054059    2.298071    2.075644    2.867897
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      5.300000    4.382500    3.070000    3.867750    1.119500    4.491000
50%      6.931000    5.867000    4.091000    5.191000    1.905000    6.227500
75%      8.661750    7.551000    5.405000    6.386500    3.059000    7.705500
max     23.226000   19.355000   12.195000   13.772000   17.391000   20.000000

                X6          X7          X8          X9         X10         X11  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     3.864656    5.455146    2.519896    5.483726    9.022420    5.874908
```

```
std      2.125270    3.291619    2.193690    2.374137    2.978185    2.878414
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      2.590750    3.662000    1.493000    4.022250    7.143000    3.983750
50%      3.560500    4.895000    2.324000    5.349000    9.033500    5.544000
75%      4.721000    6.667000    3.132000    6.798250   11.017750    7.306750
max     15.942000   38.824000   37.931000   15.625000   27.778000   19.048000

              X12         X13         X14         X15         X16         X17  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     2.797884    4.108908    5.066796    7.761982    5.703904    1.244034
std      1.493566    2.145608    3.222189    3.066539    2.734419    1.060433
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      1.816750    2.652000    3.125000    5.737000    4.292750    0.532000
50%      2.548500    4.025500    4.439500    7.529000    5.420500    1.097000
75%      3.407750    5.271000    6.173750    9.375000    6.708000    1.727750
max     10.000000   20.513000   22.741000   21.739000   37.931000   10.390000

              X18         X19
count  500.000000  500.000000
mean     2.949314    6.560136
std      1.679278    2.392036
min      0.000000    0.000000
25%      1.905000    5.028000
50%      2.796500    6.466500
75%      3.815500    7.777000
max     12.941000   17.391000
```

no_efectores

Composición de aminoácidos (AAC) no_efectores nematoda dataset 1, con valores atípicos.

Valores del documento csv.

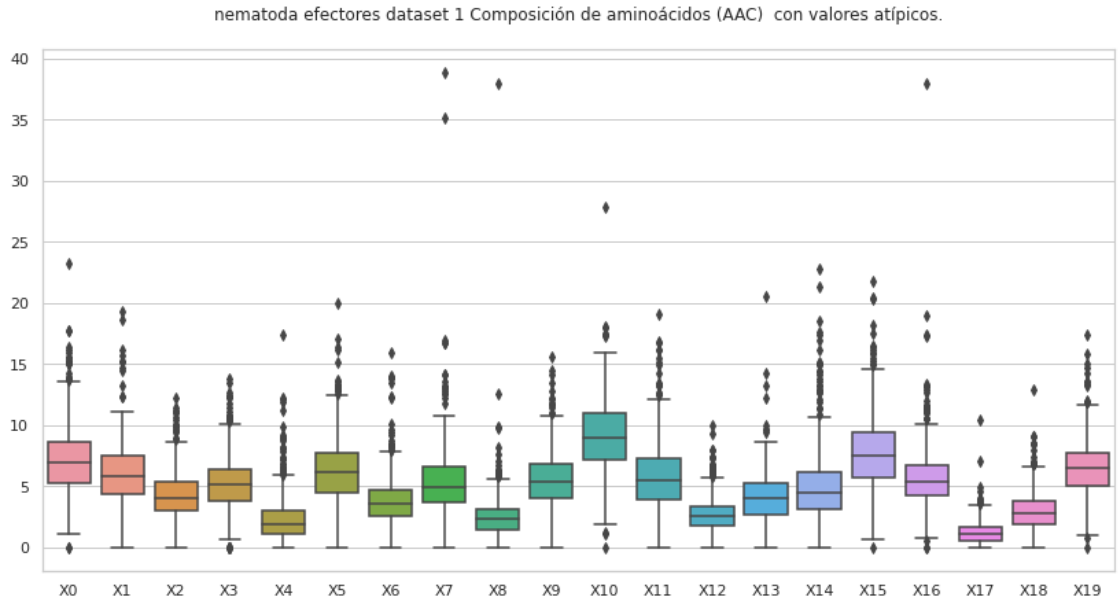
```
        X0     X1     X2     X3     X4     X5     X6      X7     X8     X9  \
0    5.023  8.447  3.425  8.447  0.685  9.132  3.881   5.479  3.653  5.251
1    6.804  3.338  3.915  6.932  0.963  7.060  3.659   5.199  1.348  4.108
2    4.494  6.742  4.494  5.056  2.809  8.989  1.124   6.742  6.180  4.494
3    4.412  5.882  5.882  7.353  0.000  7.353  7.353   1.471  1.471  8.824
4    7.661  4.839  4.032  5.645  2.419  2.823  4.435  10.484  2.419  6.048
..     …      …      …      …      …      …      …       …      …      …
495  7.463  5.970  4.478  1.493  1.493  4.478  2.985   1.493  0.000  2.985
496  4.805  4.204  1.201  3.303  3.303  6.006  3.904   8.408  5.405  6.607
497  8.115  5.759  3.403  4.188  1.571  4.712  2.880   7.853  4.188  3.403
498  6.667  5.390  5.106  6.950  0.993  9.504  3.972   5.816  2.270  4.965
499  7.517  4.784  4.100  3.872  1.708  4.214  2.506   6.492  2.278  6.948

     …     X11    X12    X13    X14    X15    X16    X17    X18    X19  \
```

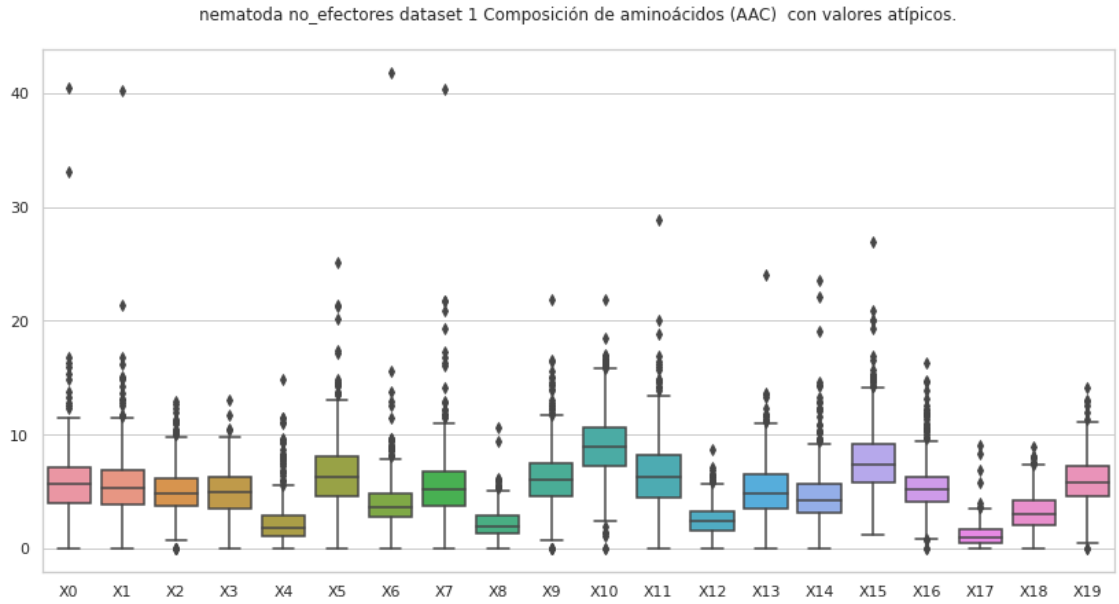
```
0    …   7.991  2.511  3.425  2.968  4.338  3.653  2.055  3.196  5.251
1    …   9.884  2.246  3.659  6.675  6.611  9.243  0.770  2.696  6.739
2    …   7.865  1.124  6.180  5.056  7.865  2.247  2.809  3.933  5.056
3    …  10.294  4.412  4.412  5.882  2.941  7.353  0.000  4.412  4.412
4    …   6.048  2.823  3.226  4.435  6.452  6.855  0.403  3.629  7.258
..   …     …    …    …    …    …    …    …    …    …
495  …  10.448  4.478  8.955  4.478  7.463  4.478  1.493  8.955  7.463
496  …   5.706  5.105  5.706  2.703  3.003  4.204  3.303  4.505  9.009
497  …   5.497  1.309  7.330  6.806  8.115  5.236  0.785  3.403  5.497
498  …   6.241  2.695  2.979  4.681  8.511  4.681  1.135  2.411  5.957
499  …   3.645  3.645  6.036  5.239  8.200  6.948  0.569  2.733  8.884

                X20
0     no_efectores
1     no_efectores
2     no_efectores
3     no_efectores
4     no_efectores
..             …
495   no_efectores
496   no_efectores
497   no_efectores
498   no_efectores
499   no_efectores

[500 rows x 21 columns]
```

Composición de aminoácidos (AAC) no_efectores nematoda dataset 1, con valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count   500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean      5.880718    5.660326    5.185124    4.949422    2.277748    6.454864
std       3.207325    3.144684    2.186123    1.988428    1.962305    3.159521
min       0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%       4.027750    3.880500    3.784750    3.510500    1.099000    4.615500
50%       5.642000    5.268500    4.829000    5.000000    1.828000    6.250000
75%       7.148000    6.941250    6.213500    6.239500    2.893000    8.088000
max      40.506000   40.268000   12.963000   13.043000   14.873000   25.175000

                X6          X7          X8          X9         X10         X11  \
count   500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean      4.024360    5.619326    2.139274    6.233534    9.059778    6.618006
std       2.620694    3.252335    1.309969    2.696002    3.097425    3.190380
min       0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%       2.789250    3.787250    1.328750    4.623500    7.207500    4.509500
```

```
50%      3.612500    5.168500    1.981500    6.044000    8.898500    6.250000
75%      4.824500    6.737500    2.859000    7.467250   10.652500    8.173750
max     41.772000   40.278000   10.638000   21.875000   21.875000   28.859000

              X12         X13         X14         X15         X16         X17  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     2.543830    5.153670    4.637594    7.780076    5.445070    1.216930
std      1.271262    2.511457    2.664344    3.090818    2.175657    1.060047
min      0.000000    0.000000    0.000000    1.266000    0.000000    0.000000
25%      1.620500    3.537250    3.125000    5.800250    4.141000    0.499750
50%      2.409000    4.853000    4.179500    7.359500    5.228500    0.980000
75%      3.269250    6.537250    5.630250    9.172750    6.334250    1.726500
max      8.642000   24.038000   23.558000   26.923000   16.288000    9.091000

              X18         X19
count  500.000000  500.000000
mean     3.184118    5.936098
std      1.676698    2.155638
min      0.000000    0.000000
25%      2.034000    4.574000
50%      2.960500    5.842000
75%      4.169000    7.267750
max      8.955000   14.141000
```





2.1 Composición de aminoácidos (AAC), sin valores atípicos

[4]:
```
transf = "Composición de aminoácidos (AAC) "
estado = "sin valores atípicos.\n"
transf2="AAC"

out = (str(r3) + '/ds' + str(dataset) + '_' + str(transf2) + '_' +␣
 ↪str(organismo) + '.csv')
os.makedirs(str(r3), exist_ok=True)
df=""
df_out = pd.DataFrame()


for etiq in "efectores", "no_efectores":
    titulo = (str(transf) + str(etiq) + " " + str(nombre2) + ", " +str(estado))
    print (str(etiq))

    if etiq == "efectores":
        df=AAC_efec

    if etiq == "no_efectores":
        df=AAC_no_efec

    del df['X20']
    #Se eliminan todas las filas que tengan valores atípicos en al menos una de␣
 ↪sus columnas.
    df = (df[(np.abs(stats.zscore(df)) < 3).all(axis=1)])
```

```
    df['X20'] = etiq
    df_out = pd.concat([df_out,df])

    #Guarda la lista csv sin valores atípicos.
    df_out.to_csv(str(out), index=False, header=False)



    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")

    #Gráfica de caja y bigotes
    sns.set(style="whitegrid")
    fig , ax = plt.subplots(figsize=(14,7))
    ax = sns.boxplot(data=df)
    ax.set_title(organismo +' '+str(etiq) +" dataset "+ str(dataset)+"␣
↪"+str(transf))
```

```
efectores
Composición de aminoácidos (AAC) efectores nematoda dataset 1, sin valores
atípicos.
Valores del documento csv.

         X0     X1     X2      X3     X4      X5     X6     X7     X8      X9  \
1    10.423  4.560  3.583   3.257  0.977   3.583  1.954  4.235  1.629   8.795
2    11.526  2.804  4.984   5.296  4.050   5.296  2.804  3.738  2.492   5.607
3     3.462  6.154  4.615   3.846  3.462   6.154  4.615  9.231  3.077   4.231
5     6.050  8.897  3.203   8.185  0.356   6.762  4.270  3.915  1.423   6.050
6     4.938  8.642  3.704  11.111  7.407   6.173  3.704  8.642  0.000   4.938
..      …      …      …       …      …       …      …      …      …       …
495   7.226  5.128  3.030   5.711  1.515   7.692  3.147  6.061  2.448   5.245
496   8.929  5.357  5.357   7.143  1.786   5.357  0.000  1.786  3.571   5.357
497   4.044  5.699  4.596   6.985  1.654   7.537  6.618  4.412  2.206   6.250
498   6.757  2.703  8.108   4.054  0.000   6.757  1.351  1.351  0.000  12.162
499   2.041  5.102  0.000   8.163  1.020  10.204  2.041  4.082  8.163   6.122

     …     X11    X12    X13    X14     X15    X16    X17    X18     X19  \
1    …   3.257  2.932  7.166  3.909   7.166  6.840  2.280  4.560   8.795
2    …   6.542  2.181  4.673  4.361   6.542  9.657  1.558  2.181   5.607
3    …   5.769  3.846  3.462  3.846   7.692  8.077  1.923  1.154   6.923
5    …   8.541  2.847  3.559  5.338   7.117  6.050  1.068  4.626   3.559
6    …   6.173  2.469  7.407  0.000   8.642  2.469  2.469  0.000   2.469
..   …     …      …      …      …       …      …      …      …       …
```

```
495  …   8.042  2.098  3.497  5.478   5.944  5.944  2.214  3.380   7.226
496  …   3.571  3.571  5.357  8.929  10.714  3.571  0.000  3.571  10.714
497  …   4.963  2.574  5.147  4.228   5.331  4.044  0.735  5.699   6.066
498  …  10.811  2.703  5.405  6.757   4.054  6.757  2.703  4.054   1.351
499  …   5.102  5.102  1.020  3.061   6.122  5.102  2.041  4.082  10.204

           X20
1    efectores
2    efectores
3    efectores
5    efectores
6    efectores
..         …
495  efectores
496  efectores
497  efectores
498  efectores
499  efectores

[406 rows x 21 columns]
```

Composición de aminoácidos (AAC) efectores nematoda dataset 1, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  406.000000  406.000000  406.000000  406.000000  406.000000  406.000000
mean     7.182010    5.979507    4.453337    5.375135    2.223722    6.316554
std      2.441505    2.170607    1.782697    1.963377    1.581209    2.461876
min      1.143000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      5.493000    4.481500    3.310250    4.094500    1.177250    4.688750
50%      7.055000    5.898500    4.208500    5.279500    1.905000    6.234000
75%      8.573750    7.443250    5.451500    6.360250    2.894750    7.615250
max     15.504000   13.265000   10.145000   11.786000    8.235000   13.699000

                X6          X7          X8          X9         X10         X11  \
count  406.000000  406.000000  406.000000  406.000000  406.000000  406.000000
mean     3.769616    5.364635    2.456621    5.622788    9.228094    5.876039
std      1.615663    2.336691    1.330294    2.085747    2.528721    2.365016
min      0.000000    0.000000    0.000000    0.000000    2.469000    0.000000
25%      2.724750    3.867250    1.590500    4.241500    7.435750    4.269500
50%      3.611500    4.949000    2.381500    5.487000    9.141000    5.615000
75%      4.638250    6.638500    3.103250    6.915750   11.049750    7.205250
max     10.127000   14.118000    8.163000   12.222000   17.391000   14.286000

               X12         X13         X14         X15         X16         X17  \
count  406.000000  406.000000  406.000000  406.000000  406.000000  406.000000
```

mean	2.749012	4.217131	4.754904	7.778131	5.668278	1.284222
std	1.294455	1.811892	2.330157	2.648766	2.007198	0.888599
min	0.000000	0.000000	0.000000	0.685000	0.752000	0.000000
25%	1.870500	2.971000	3.213000	5.915500	4.421250	0.686250
50%	2.546000	4.142000	4.416500	7.605500	5.437000	1.162000
75%	3.400500	5.396000	5.820750	9.364500	6.653250	1.777500
max	6.818000	10.000000	14.516000	16.434000	13.380000	3.977000

	X18	X19
count	406.000000	406.000000
mean	3.040103	6.660340
std	1.468583	2.090448
min	0.000000	0.813000
25%	2.109250	5.342500
50%	2.921000	6.604000
75%	3.889000	7.791000
max	7.937000	13.636000

no_efectores
Composición de aminoácidos (AAC) no_efectores nematoda dataset 1, sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9	\
0	5.023	8.447	3.425	8.447	0.685	9.132	3.881	5.479	3.653	5.251	
1	6.804	3.338	3.915	6.932	0.963	7.060	3.659	5.199	1.348	4.108	
3	4.412	5.882	5.882	7.353	0.000	7.353	7.353	1.471	1.471	8.824	
4	7.661	4.839	4.032	5.645	2.419	2.823	4.435	10.484	2.419	6.048	
5	8.705	5.469	4.241	6.920	0.670	8.036	5.022	7.366	1.451	5.692	
..	…	…	…	…	…	…	…	…	…	…	
494	4.969	9.006	3.106	7.764	1.242	5.590	3.416	6.832	3.106	5.280	
496	4.805	4.204	1.201	3.303	3.303	6.006	3.904	8.408	5.405	6.607	
497	8.115	5.759	3.403	4.188	1.571	4.712	2.880	7.853	4.188	3.403	
498	6.667	5.390	5.106	6.950	0.993	9.504	3.972	5.816	2.270	4.965	
499	7.517	4.784	4.100	3.872	1.708	4.214	2.506	6.492	2.278	6.948	

	…	X11	X12	X13	X14	X15	X16	X17	X18	X19	\
0	…	7.991	2.511	3.425	2.968	4.338	3.653	2.055	3.196	5.251	
1	…	9.884	2.246	3.659	6.675	6.611	9.243	0.770	2.696	6.739	
3	…	10.294	4.412	4.412	5.882	2.941	7.353	0.000	4.412	4.412	
4	…	6.048	2.823	3.226	4.435	6.452	6.855	0.403	3.629	7.258	
5	…	8.147	2.121	3.237	3.125	5.246	5.580	0.112	1.674	9.263	
..	…	…	…	…	…	…	…	…	…	…	
494	…	7.143	4.348	3.416	3.727	5.901	4.658	1.242	4.037	5.590	
496	…	5.706	5.105	5.706	2.703	3.003	4.204	3.303	4.505	9.009	
497	…	5.497	1.309	7.330	6.806	8.115	5.236	0.785	3.403	5.497	

```
498  …   6.241  2.695  2.979  4.681  8.511  4.681  1.135  2.411  5.957
499  …   3.645  3.645  6.036  5.239  8.200  6.948  0.569  2.733  8.884

              X20
0     no_efectores
1     no_efectores
3     no_efectores
4     no_efectores
5     no_efectores
..             …
494   no_efectores
496   no_efectores
497   no_efectores
498   no_efectores
499   no_efectores

[418 rows x 21 columns]
```

Composición de aminoácidos (AAC) no_efectores nematoda dataset 1, sin valores atípicos.
Estadísticas.

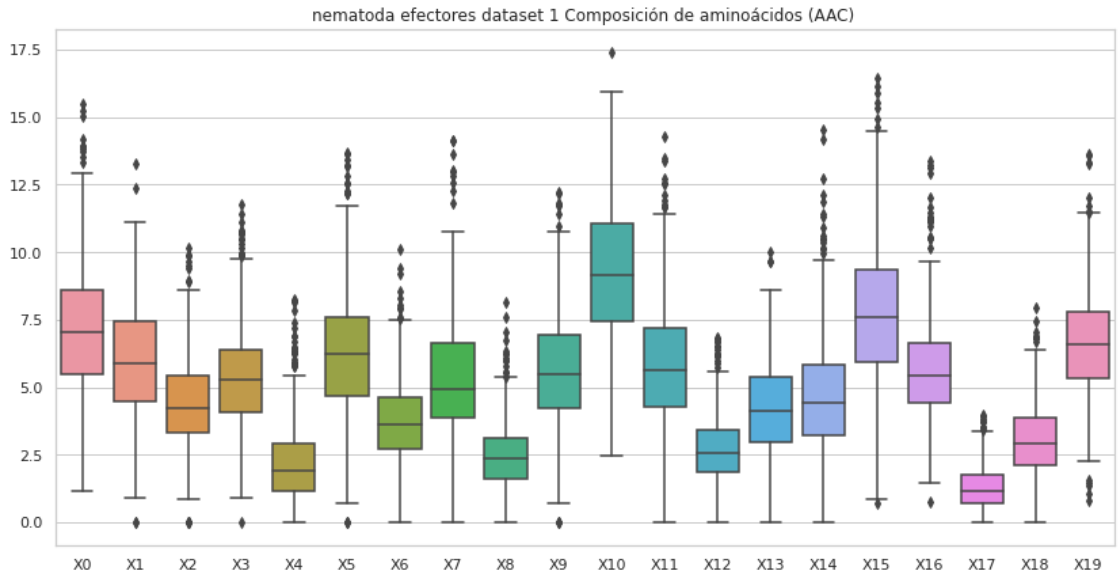
	X0	X1	X2	X3	X4	X5	\
count	418.000000	418.000000	418.000000	418.000000	418.000000	418.000000	
mean	5.832153	5.618761	5.224895	5.128873	2.131852	6.595340	
std	2.121239	2.329147	2.002058	1.805220	1.527511	2.626689	
min	0.000000	0.000000	0.000000	0.885000	0.000000	0.000000	
25%	4.417000	4.141750	3.907500	3.823250	1.120250	4.954500	
50%	5.797000	5.299500	4.884500	5.192000	1.814500	6.400000	
75%	7.133250	6.831250	6.168500	6.294250	2.828000	8.109500	
max	13.793000	15.033000	11.404000	10.384000	8.088000	14.857000	

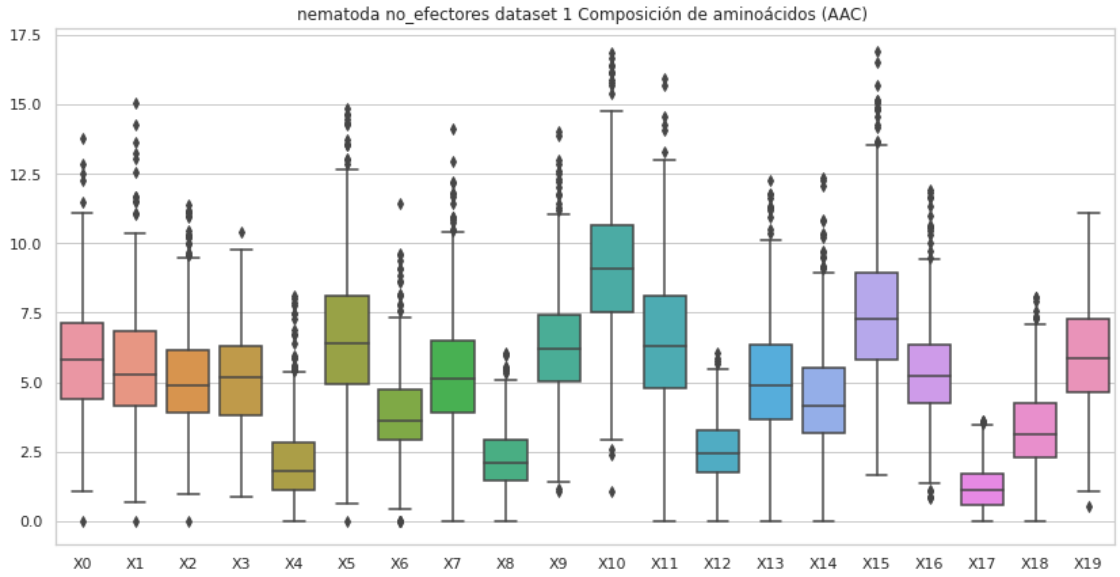
	X6	X7	X8	X9	X10	X11	\
count	418.000000	418.000000	418.000000	418.000000	418.000000	418.000000	
mean	3.910486	5.355617	2.251935	6.343115	9.277411	6.580156	
std	1.743073	2.219753	1.139171	2.152265	2.651369	2.586627	
min	0.000000	0.000000	0.000000	1.079000	1.064000	0.000000	
25%	2.941000	3.891000	1.471250	5.009250	7.547000	4.812250	
50%	3.613500	5.147500	2.116000	6.196000	9.084500	6.297500	
75%	4.762000	6.504750	2.930500	7.450750	10.665500	8.120750	
max	11.450000	14.130000	6.061000	14.035000	16.854000	15.929000	

	X12	X13	X14	X15	X16	X17	\
count	418.000000	418.000000	418.000000	418.000000	418.000000	418.000000	
mean	2.546847	5.182227	4.475189	7.608538	5.419565	1.217971	
std	1.119827	2.133744	2.022155	2.622622	1.849107	0.862309	
min	0.000000	0.000000	0.000000	1.639000	0.847000	0.000000	

25%	1.773250	3.685750	3.171250	5.822000	4.268750	0.578500
50%	2.462000	4.904500	4.163500	7.282500	5.248500	1.109500
75%	3.258750	6.347000	5.535250	8.946250	6.336750	1.731500
max	6.061000	12.281000	12.355000	16.901000	11.905000	3.614000

	X18	X19
count	418.000000	418.000000
mean	3.320447	5.978404
std	1.572577	1.962500
min	0.000000	0.543000
25%	2.284750	4.657500
50%	3.141000	5.882000
75%	4.269500	7.296000
max	8.073000	11.111000





3 Composición de pseudo aminoácidos (PseAAC) hidro_mass

[5]:
```
#hidro_mass
transf = "Composición de pseudo aminoácidos (PseAAC) "
transf2 = "PseAAC"
estado = "con valores atípicos.\n"
comp = "hidro_mass"
df=""

for etiq in "efectores", "no_efectores":
    titulo = (str(transf)+" "+ str(comp)+" "+ str(etiq) + " "+ str(nombre2) +",␣
 ↪" + str(estado))
    print (str(etiq))

    if etiq == "efectores":
        df=PseAAC_hidro_mass_efec

    if etiq == "no_efectores":
        df=PseAAC_hidro_mass_no_efec

    #del df['X83']
    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")
```

```
    #Gráfica de caja y bigotes
    sns.set(style="whitegrid")
    fig , ax = plt.subplots(figsize=(14,7))
    ax = sns.boxplot(data=df)
    ax.set_title(organismo +' '+str(etiq)+" dataset "+str(dataset)+"␣
↪"+str(transf)+" "+str(comp)+" "+str(estado))
```

```
efectores
Composición de pseudo aminoácidos (PseAAC)  hidro_mass efectores nematoda
dataset 1, con valores atípicos.
Valores del documento csv.
```

	X0	X1	X2	X3	X4	X5	X6	\
0	0.007836	0.003918	0.011265	0.009795	0.002939	0.029386	0.003428	
1	0.030037	0.002816	0.009387	0.010325	0.020651	0.012203	0.004693	
2	0.057628	0.020248	0.026478	0.026478	0.023363	0.018690	0.012460	
3	0.015188	0.015188	0.016876	0.027001	0.015188	0.040502	0.013501	
4	0.055468	0.033281	0.033281	0.088749	0.066562	0.110936	0.033281	
..	…	…	…	…	…	…	…	
495	0.039120	0.008203	0.030917	0.041644	0.018929	0.032810	0.013250	
496	0.041172	0.008234	0.032938	0.024703	0.024703	0.008234	0.016469	
497	0.040104	0.016406	0.069271	0.074739	0.051042	0.043750	0.021875	
498	0.095446	0.000000	0.057268	0.095446	0.076357	0.019089	0.000000	
499	0.013326	0.006663	0.053305	0.066631	0.006663	0.026652	0.053305	

	X7	X8	X9	…	X74	X75	X76	\
0	0.007836	0.008816	0.005877	…	-0.003822	0.008165	0.013440	
1	0.025344	0.009387	0.029098	…	0.013568	-0.001269	0.015437	
2	0.028035	0.032708	0.040495	…	0.000905	0.004293	0.005864	
3	0.018564	0.025314	0.037127	…	0.011856	0.006461	0.031532	
4	0.044374	0.055468	0.099842	…	-0.010891	-0.042273	0.019204	
..	…	…	…	…	…	…	…	
495	0.028394	0.043537	0.048584	…	0.031890	0.044387	0.007738	
496	0.024703	0.016469	0.024703	…	-0.014249	-0.023521	0.037052	
497	0.061979	0.049219	0.111198	…	0.028267	0.015835	0.007139	
498	0.171803	0.152714	0.171803	…	0.060882	0.038789	0.035966	
499	0.039979	0.033315	0.073294	…	0.092049	0.053991	0.007647	

	X77	X78	X79	X80	X81	X82	X83
0	0.002100	0.007574	0.027159	0.008853	0.011629	0.039347	efectores
1	0.026391	-0.001450	0.007185	0.012212	0.008466	0.008232	efectores
2	-0.005351	0.013373	0.009126	-0.006274	0.005327	0.022124	efectores
3	-0.000603	0.004717	-0.002630	0.003535	0.023471	0.005184	efectores
4	-0.014848	-0.027603	0.021124	0.013133	0.104520	-0.002720	efectores
..	…	…	…	…	…	…	…
495	-0.016023	-0.003318	0.004836	-0.013647	0.001471	-0.001451	efectores
496	0.000154	0.002367	-0.012882	0.066041	0.056630	0.027313	efectores

```
497 -0.026377  0.009688  0.006002 -0.034426 -0.010780  0.011688  efectores
498  0.060624  0.072183  0.113628  0.152977  0.061196  0.095336  efectores
499 -0.025882  0.015781  0.019194  0.026852  0.016847  0.005103  efectores

[500 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro_mass efectores nematoda
dataset 1, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.036247	0.012691	0.026727	0.034890	0.019477	0.026629	
std	0.039134	0.035178	0.062180	0.027432	0.088042	0.036819	
min	-0.633179	-0.633179	-1.266359	0.000000	-1.899538	-0.633179	
25%	0.022821	0.004301	0.014597	0.017524	0.010722	0.015270	
50%	0.032630	0.009271	0.024824	0.029459	0.019199	0.023514	
75%	0.048012	0.017654	0.037416	0.044810	0.031133	0.035574	
max	0.236953	0.292058	0.236953	0.331735	0.162254	0.284344	

	X6	X7	X8	X9	…	X73	\
count	500.000000	500.000000	500.000000	500.000000	…	500.000000	
mean	0.013135	0.027246	0.026651	0.043486	…	0.010373	
std	0.034373	0.089990	0.117551	0.176722	…	0.042869	
min	-0.633179	-1.899538	-2.532718	-3.799076	…	-0.269312	
25%	0.005630	0.015163	0.016301	0.027520	…	-0.001864	
50%	0.010508	0.026315	0.025861	0.041969	…	0.011033	
75%	0.019046	0.041262	0.040814	0.065488	…	0.022964	
max	0.257783	0.259607	0.284344	0.426516	…	0.656000	

	X74	X75	X76	X77	X78	X79	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	-0.006985	0.003655	0.015445	0.021039	0.013432	0.018759	
std	0.192540	0.089862	0.058754	0.401805	0.132915	0.184419	
min	-4.230861	-1.884372	-0.200167	-0.153310	-0.287175	-0.255803	
25%	-0.008780	-0.003976	0.001191	-0.010086	-0.003091	0.001161	
50%	0.003918	0.006223	0.011926	0.003433	0.008672	0.011075	
75%	0.015434	0.019194	0.023769	0.016530	0.020067	0.022263	
max	0.156006	0.200352	1.008012	8.959838	2.911246	4.075306	

	X80	X81	X82
count	500.000000	500.000000	500.000000
mean	0.004560	0.012960	0.012185
std	0.095847	0.132586	0.036354
min	-0.504104	-0.304992	-0.274814
25%	-0.007929	-0.004449	-0.000209
50%	0.004595	0.007354	0.011507

```
75%      0.016627    0.019857    0.021877
max      1.873755    2.866911    0.471393

[8 rows x 83 columns]
```

```
no_efectores
Composición de pseudo aminoácidos (PseAAC)  hidro_mass no_efectores nematoda
dataset 1, con valores atípicos.
Valores del documento csv.
```

	X0	X1	X2	X3	X4	X5	X6	\
0	0.022996	0.003136	0.038675	0.041810	0.015679	0.025086	0.016724	
1	0.021685	0.003069	0.022094	0.022503	0.011661	0.016570	0.004296	
2	0.036451	0.022782	0.041008	0.072903	0.050120	0.054677	0.050120	
3	0.022273	0.000000	0.037121	0.037121	0.022273	0.007424	0.007424	
4	0.033648	0.010626	0.024793	0.012397	0.014168	0.046044	0.010626	
..	…	…	…	…	…	…	…	
495	0.088607	0.017721	0.017721	0.053164	0.106328	0.017721	0.000000	
496	0.019271	0.013249	0.013249	0.024089	0.022884	0.033724	0.021680	
497	0.045966	0.008897	0.023725	0.026690	0.041518	0.044483	0.023725	
498	0.026735	0.003982	0.027872	0.038111	0.011945	0.023322	0.009101	
499	0.021636	0.004917	0.011146	0.012129	0.017375	0.018686	0.006556	

	X7	X8	X9	…	X74	X75	X76	\
0	0.024041	0.036584	0.051218	…	-0.001373	0.036804	-0.012604	
1	0.013093	0.031504	0.025981	…	0.002710	0.010393	0.018582	
2	0.036451	0.063790	0.054677	…	-0.014363	0.049078	0.002598	
3	0.044545	0.051969	0.029697	…	0.047895	0.030388	-0.016403	
4	0.026564	0.026564	0.035419	…	0.013011	-0.018878	0.052896	
..	…	…	…	…	…	…	…	
495	0.035443	0.124050	0.106328	…	-0.071858	-0.131607	0.065814	
496	0.026498	0.022884	0.038542	…	-0.014993	-0.001859	0.013378	
497	0.019276	0.031138	0.056346	…	-0.006907	-0.004928	0.012238	
498	0.019909	0.025028	0.036405	…	-0.006250	0.007351	0.006748	
499	0.019997	0.010490	0.027865	…	0.014148	0.008245	0.010554	

	X77	X78	X79	X80	X81	X82	X83
0	0.003639	0.021300	-0.019104	-0.008210	0.022874	0.003490	no_efectores
1	0.014726	0.020501	0.019256	-0.000283	0.010150	0.017498	no_efectores
2	-0.010550	0.011699	-0.047440	-0.015560	0.032229	0.002253	no_efectores
3	-0.036120	0.012045	0.032978	-0.002671	0.047340	0.015817	no_efectores
4	-0.002155	0.001760	-0.004354	-0.004789	0.004096	0.028776	no_efectores
..	…	…	…	…	…	…	
495	-0.143312	-0.034928	0.115736	-0.056532	-0.056335	-0.104012	no_efectores
496	0.006040	0.011009	0.001940	0.003994	0.005009	-0.005480	no_efectores
497	-0.006599	-0.032492	0.028730	-0.010642	0.005263	0.004303	no_efectores

```
498  0.002176  0.013962  0.009825  0.020966  0.040198  0.008565  no_efectores
499  0.014542  0.011297  0.014365  0.005172  0.004127  0.014795  no_efectores

[500 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores nematoda
dataset 1, con valores atípicos.
Estadísticas.

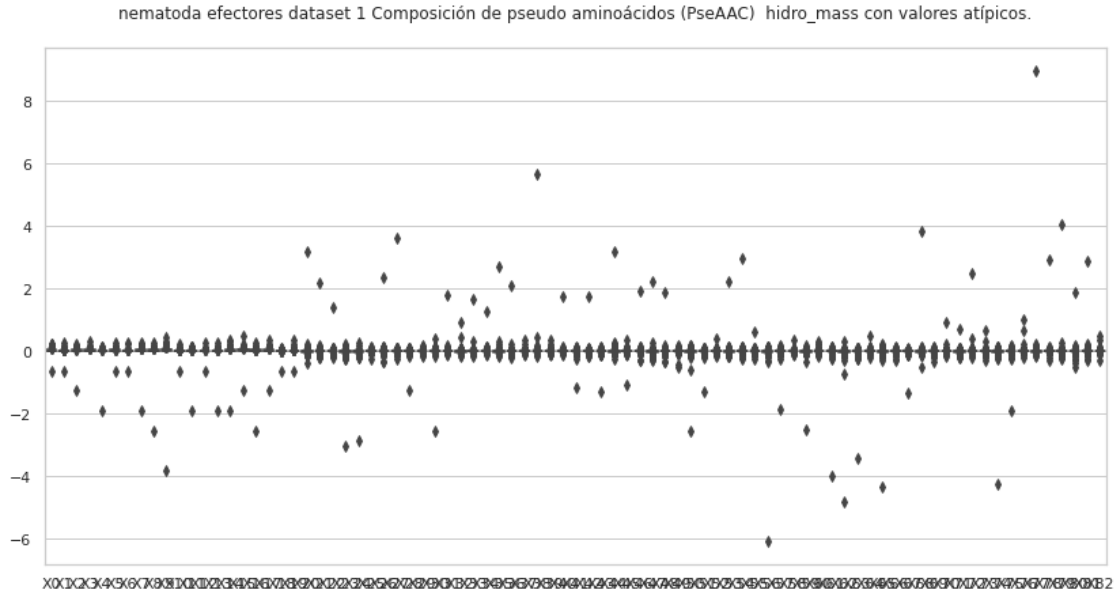
	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.031063	0.013290	0.028628	0.038728	0.032986	0.030709	
std	0.027361	0.015091	0.028123	0.059271	0.075074	0.037380	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.018151	0.004381	0.013634	0.017854	0.014314	0.016628	
50%	0.026860	0.009356	0.022901	0.031629	0.024406	0.025053	
75%	0.039381	0.015693	0.037623	0.047337	0.038328	0.037464	
max	0.462706	0.129099	0.462706	1.156764	1.619469	0.694058	

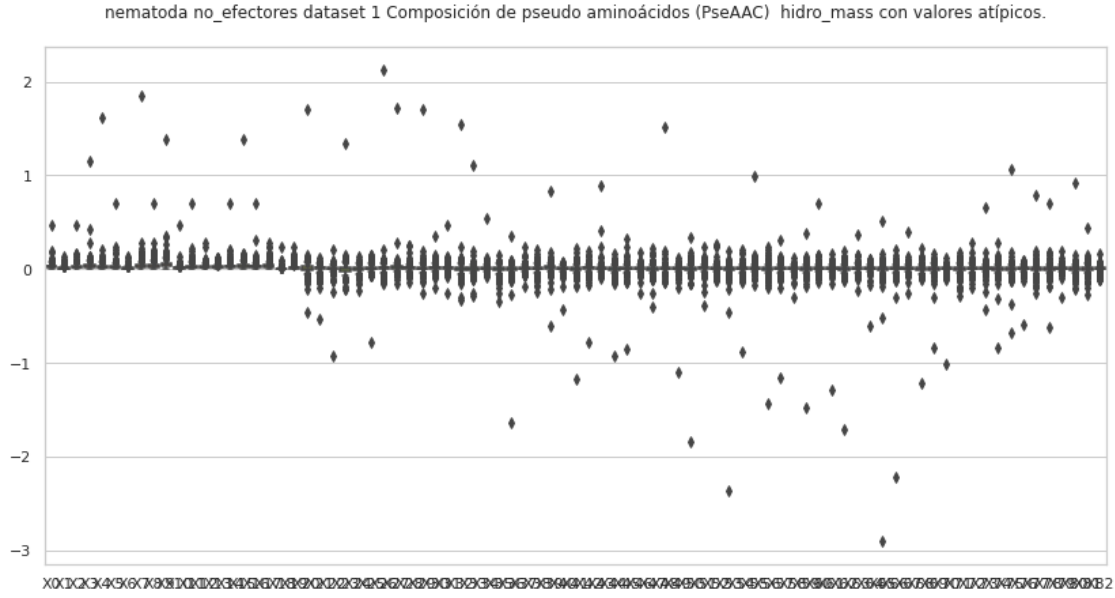
	X6	X7	X8	X9	…	X73	\
count	500.000000	500.000000	500.000000	500.000000	…	500.000000	
mean	0.012534	0.039202	0.038238	0.054286	…	0.008773	
std	0.013103	0.086040	0.041577	0.070951	…	0.044002	
min	0.000000	0.000000	0.000000	0.000000	…	-0.429280	
25%	0.004901	0.018324	0.018687	0.028950	…	-0.000887	
50%	0.009381	0.029929	0.030790	0.044371	…	0.008133	
75%	0.016084	0.044430	0.047786	0.063297	…	0.019200	
max	0.129099	1.850822	0.694058	1.388117	…	0.663836	

	X74	X75	X76	X77	X78	X79	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.000929	0.006908	0.006388	0.001639	0.005468	0.007438	
std	0.051192	0.064621	0.034152	0.050862	0.052029	0.025170	
min	-0.841451	-0.682570	-0.589221	-0.264286	-0.618957	-0.300893	
25%	-0.008908	-0.004081	-0.001992	-0.010564	-0.004751	-0.001069	
50%	0.003412	0.008044	0.007810	0.003790	0.006589	0.006954	
75%	0.015427	0.019157	0.018491	0.014809	0.018698	0.017408	
max	0.274352	1.067132	0.083806	0.784630	0.704303	0.185660	

	X80	X81	X82
count	500.000000	500.000000	500.000000
mean	0.003644	0.006761	0.006304
std	0.051836	0.037733	0.024062
min	-0.217362	-0.278185	-0.118340
25%	-0.007490	-0.004520	-0.000912
50%	0.004473	0.006464	0.006811
75%	0.015649	0.019691	0.017656

```
max       0.916834     0.442391     0.162602

[8 rows x 83 columns]
```





3.1 Composición de pseudo aminoácidos (PseAAC) hidro_mass, sin valores atípicos

[6]:
```
#hidro_mass
transf = "Composición de pseudo aminoácidos (PseAAC) "
transf2 = "PseAAC"
estado = "sin valores atípicos.\n"
comp = "hidro_mass"
df=""

out = (str(r3) + '/ds' + str(dataset) + '_' + str(transf2) + '_' + str(comp) +␣
↪'_' + str(organismo) + '.csv')
os.makedirs(str(r3), exist_ok=True)
df_out = pd.DataFrame()


for etiq in "efectores", "no_efectores":
    titulo = (str(transf)+" "+ str(comp)+" "+ str(etiq) + " "+ str(nombre2) +",␣
↪" + str(estado))
    print (str(etiq))

    if etiq == "efectores":
        df=PseAAC_hidro_mass_efec

    if etiq == "no_efectores":
        df=PseAAC_hidro_mass_no_efec

    del df['X83']
    #Se eliminan todas las filas que tengan valores atípicos en al menos una de␣
↪sus columnas.
    df = (df[(np.abs(stats.zscore(df)) < 3).all(axis=1)])
    df['X83'] = etiq
    df_out = pd.concat([df_out,df])

    #Guarda la lista csv sin valores atípicos.
    df_out.to_csv(str(out), index=False, header=False)


    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")

    #Gráfica de caja y bigotes
    sns.set(style="whitegrid")
    fig , ax = plt.subplots(figsize=(14,7))
```

```
    ax = sns.boxplot(data=df)
    ax.set_title(organismo +' '+str(etiq)+" dataset "+str(dataset)+"␣
↪"+str(transf)+" "+str(comp))
```

```
efectores
Composición de pseudo aminoácidos (PseAAC)  hidro_mass efectores nematoda
dataset 1, sin valores atípicos.
Valores del documento csv.
```

	X0	X1	X2	X3	X4	X5	X6	\
0	0.007836	0.003918	0.011265	0.009795	0.002939	0.029386	0.003428	
1	0.030037	0.002816	0.009387	0.010325	0.020651	0.012203	0.004693	
2	0.057628	0.020248	0.026478	0.026478	0.023363	0.018690	0.012460	
3	0.015188	0.015188	0.016876	0.027001	0.015188	0.040502	0.013501	
5	0.028093	0.001653	0.038008	0.031398	0.016525	0.018178	0.006610	
..	…	…	…	…	…	…	…	
494	0.045026	0.019297	0.033591	0.030732	0.026444	0.032161	0.015009	
495	0.039120	0.008203	0.030917	0.041644	0.018929	0.032810	0.013250	
496	0.041172	0.008234	0.032938	0.024703	0.024703	0.008234	0.016469	
497	0.040104	0.016406	0.069271	0.074739	0.051042	0.043750	0.021875	
499	0.013326	0.006663	0.053305	0.066631	0.006663	0.026652	0.053305	

	X7	X8	X9	…	X74	X75	X76	\
0	0.007836	0.008816	0.005877	…	-0.003822	0.008165	0.013440	
1	0.025344	0.009387	0.029098	…	0.013568	-0.001269	0.015437	
2	0.028035	0.032708	0.040495	…	0.000905	0.004293	0.005864	
3	0.018564	0.025314	0.037127	…	0.011856	0.006461	0.031532	
5	0.028093	0.039661	0.038008	…	0.020864	0.039389	0.005895	
..	…	…	…	…	…	…	…	
494	0.049314	0.029303	0.045026	…	0.006299	0.015705	0.018816	
495	0.028394	0.043537	0.048584	…	0.031890	0.044387	0.007738	
496	0.024703	0.016469	0.024703	…	-0.014249	-0.023521	0.037052	
497	0.061979	0.049219	0.111198	…	0.028267	0.015835	0.007139	
499	0.039979	0.033315	0.073294	…	0.092049	0.053991	0.007647	

	X77	X78	X79	X80	X81	X82	X83
0	0.002100	0.007574	0.027159	0.008853	0.011629	0.039347	efectores
1	0.026391	-0.001450	0.007185	0.012212	0.008466	0.008232	efectores
2	-0.005351	0.013373	0.009126	-0.006274	0.005327	0.022124	efectores
3	-0.000603	0.004717	-0.002630	0.003535	0.023471	0.005184	efectores
5	-0.006153	0.008173	-0.001833	-0.039462	-0.003576	0.000831	efectores
..	…	…	…	…	…	…	
494	0.005296	0.011382	0.013957	0.016266	0.011452	0.014433	efectores
495	-0.016023	-0.003318	0.004836	-0.013647	0.001471	-0.001451	efectores
496	0.000154	0.002367	-0.012882	0.066041	0.056630	0.027313	efectores
497	-0.026377	0.009688	0.006002	-0.034426	-0.010780	0.011688	efectores
499	-0.025882	0.015781	0.019194	0.026852	0.016847	0.005103	efectores

[469 rows x 84 columns]

Composición de pseudo aminoácidos (PseAAC) hidro_mass efectores nematoda
dataset 1, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	469.000000	469.000000	469.000000	469.000000	469.000000	469.000000	
mean	0.034679	0.012246	0.026727	0.031453	0.021390	0.025676	
std	0.018003	0.011934	0.016515	0.018697	0.015619	0.014862	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.022334	0.004161	0.014082	0.017235	0.010281	0.015146	
50%	0.031898	0.008805	0.023866	0.028755	0.018089	0.022662	
75%	0.045026	0.016663	0.035571	0.041733	0.028697	0.034572	
max	0.111190	0.110084	0.088568	0.111640	0.098934	0.098934	

	X6	X7	X8	X9	…	X73	\
count	469.000000	469.000000	469.000000	469.000000	…	469.000000	
mean	0.012739	0.028005	0.028729	0.046266	…	0.010254	
std	0.009926	0.017822	0.018519	0.027825	…	0.019386	
min	0.000000	0.000000	0.000000	0.000000	…	-0.084917	
25%	0.005466	0.014916	0.015683	0.027045	…	-0.000577	
50%	0.010254	0.025019	0.025314	0.039896	…	0.011181	
75%	0.017113	0.039573	0.038200	0.061988	…	0.022785	
max	0.055042	0.100205	0.152617	0.190191	…	0.063714	

	X74	X75	X76	X77	X78	X79	\
count	469.000000	469.000000	469.000000	469.000000	469.000000	469.000000	
mean	0.002668	0.007808	0.012682	0.003271	0.007848	0.011213	
std	0.027414	0.022714	0.020763	0.024596	0.020362	0.019554	
min	-0.191763	-0.080164	-0.076564	-0.129125	-0.077077	-0.066709	
25%	-0.008133	-0.003581	0.001629	-0.009174	-0.002723	0.001954	
50%	0.004104	0.006277	0.011733	0.003447	0.008173	0.011032	
75%	0.015135	0.018927	0.023151	0.015807	0.019247	0.021715	
max	0.150140	0.145271	0.101765	0.105016	0.071438	0.103145	

	X80	X81	X82
count	469.000000	469.000000	469.000000
mean	0.002817	0.007816	0.011406
std	0.028115	0.024765	0.018623
min	-0.147139	-0.109029	-0.066681
25%	-0.006871	-0.004107	0.000640
50%	0.004483	0.007370	0.011552
75%	0.015775	0.019586	0.021489
max	0.094918	0.101166	0.082259

[8 rows x 83 columns]

no_efectores
Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores nematoda
dataset 1, sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.022996	0.003136	0.038675	0.041810	0.015679	0.025086	0.016724	
1	0.021685	0.003069	0.022094	0.022503	0.011661	0.016570	0.004296	
2	0.036451	0.022782	0.041008	0.072903	0.050120	0.054677	0.050120	
3	0.022273	0.000000	0.037121	0.037121	0.022273	0.007424	0.007424	
4	0.033648	0.010626	0.024793	0.012397	0.014168	0.046044	0.010626	
..	…	…	…	…	…	…	…	
494	0.033632	0.008408	0.052550	0.037836	0.023122	0.046244	0.021020	
496	0.019271	0.013249	0.013249	0.024089	0.022884	0.033724	0.021680	
497	0.045966	0.008897	0.023725	0.026690	0.041518	0.044483	0.023725	
498	0.026735	0.003982	0.027872	0.038111	0.011945	0.023322	0.009101	
499	0.021636	0.004917	0.011146	0.012129	0.017375	0.018686	0.006556	

	X7	X8	X9	…	X74	X75	X76	\
0	0.024041	0.036584	0.051218	…	-0.001373	0.036804	-0.012604	
1	0.013093	0.031504	0.025981	…	0.002710	0.010393	0.018582	
2	0.036451	0.063790	0.054677	…	-0.014363	0.049078	0.002598	
3	0.044545	0.051969	0.029697	…	0.047895	0.030388	-0.016403	
4	0.026564	0.026564	0.035419	…	0.013011	-0.018878	0.052896	
..	…	…	…	…	…	…	…	
494	0.035734	0.048346	0.065162	…	-0.034757	0.019152	-0.021115	
496	0.026498	0.022884	0.038542	…	-0.014993	-0.001859	0.013378	
497	0.019276	0.031138	0.056346	…	-0.006907	-0.004928	0.012238	
498	0.019909	0.025028	0.036405	…	-0.006250	0.007351	0.006748	
499	0.019997	0.010490	0.027865	…	0.014148	0.008245	0.010554	

	X77	X78	X79	X80	X81	X82	X83
0	0.003639	0.021300	-0.019104	-0.008210	0.022874	0.003490	no_efectores
1	0.014726	0.020501	0.019256	-0.000283	0.010150	0.017498	no_efectores
2	-0.010550	0.011699	-0.047440	-0.015560	0.032229	0.002253	no_efectores
3	-0.036120	0.012045	0.032978	-0.002671	0.047340	0.015817	no_efectores
4	-0.002155	0.001760	-0.004354	-0.004789	0.004096	0.028776	no_efectores
..	…	…	…	…	…	…	
494	-0.027727	0.000486	0.014019	-0.014456	0.005927	0.028552	no_efectores
496	0.006040	0.011009	0.001940	0.003994	0.005009	-0.005480	no_efectores
497	-0.006599	-0.032492	0.028730	-0.010642	0.005263	0.004303	no_efectores
498	0.002176	0.013962	0.009825	0.020966	0.040198	0.008565	no_efectores
499	0.014542	0.011297	0.014365	0.005172	0.004127	0.014795	no_efectores

[448 rows x 84 columns]

Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores nematoda
dataset 1, sin valores atípicos.
Estadísticas.

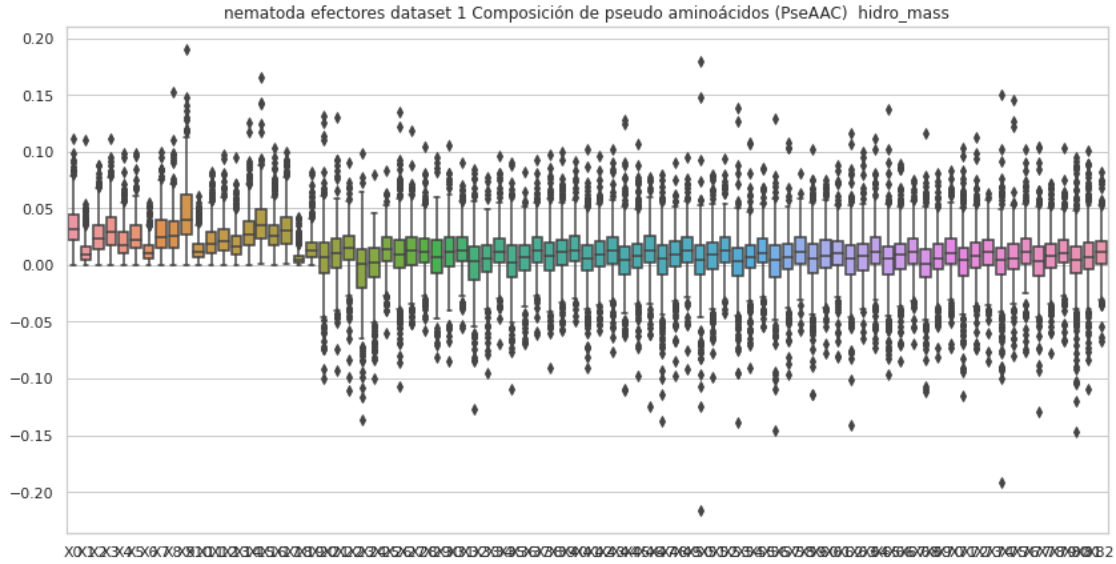
	X0	X1	X2	X3	X4	X5	\
count	448.000000	448.000000	448.000000	448.000000	448.000000	448.000000	
mean	0.027929	0.010911	0.025164	0.031337	0.025639	0.026435	
std	0.014437	0.009685	0.015316	0.018935	0.016539	0.014733	
min	0.000000	0.000000	0.000601	0.000000	0.000000	0.000000	
25%	0.017910	0.004147	0.013354	0.017142	0.013603	0.016110	
50%	0.026489	0.008931	0.022355	0.029624	0.023075	0.024164	
75%	0.036476	0.014448	0.034434	0.041958	0.034626	0.034114	
max	0.095764	0.052863	0.088549	0.099266	0.082783	0.096599	

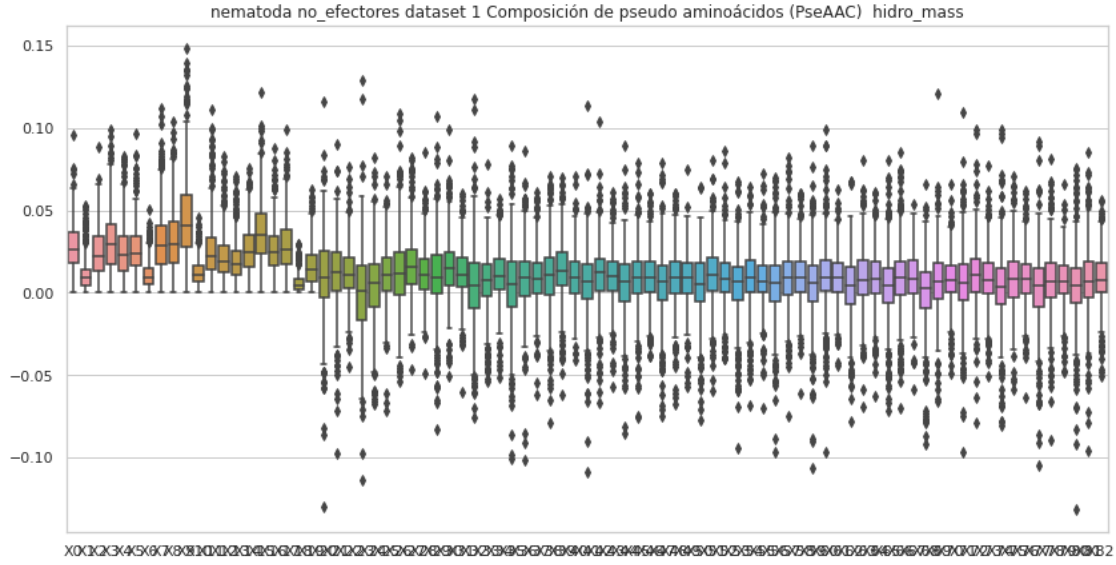
	X6	X7	X8	X9	…	X73	\
count	448.000000	448.000000	448.000000	448.000000	…	448.000000	
mean	0.010941	0.031328	0.032494	0.045875	…	0.008496	
std	0.008306	0.019326	0.019483	0.026936	…	0.016510	
min	0.000000	0.000000	0.000000	0.000000	…	-0.050564	
25%	0.004901	0.017513	0.018081	0.027791	…	-0.000833	
50%	0.009004	0.028252	0.029529	0.040778	…	0.007801	
75%	0.015147	0.040715	0.043432	0.059164	…	0.017699	
max	0.050120	0.112124	0.103460	0.148276	…	0.076868	

	X74	X75	X76	X77	X78	X79	\
count	448.000000	448.000000	448.000000	448.000000	448.000000	448.000000	
mean	0.003636	0.008286	0.008721	0.002373	0.007222	0.007453	
std	0.022399	0.018196	0.016429	0.022645	0.019541	0.014896	
min	-0.080489	-0.058831	-0.063883	-0.105171	-0.089131	-0.048621	
25%	-0.006949	-0.002624	-0.000302	-0.008220	-0.002671	-0.000729	
50%	0.003779	0.008451	0.008013	0.004337	0.006823	0.006840	
75%	0.014958	0.019082	0.017296	0.014606	0.018291	0.016433	
max	0.098612	0.070725	0.061750	0.091724	0.080832	0.065542	

	X80	X81	X82
count	448.000000	448.000000	448.000000
mean	0.003403	0.007825	0.008556
std	0.022394	0.020947	0.015977
min	-0.131552	-0.096003	-0.051012
25%	-0.006237	-0.003106	0.000599
50%	0.004323	0.006856	0.007473
75%	0.014599	0.019285	0.017741
max	0.075238	0.085169	0.056429

[8 rows x 83 columns]


nematoda efectores dataset 1 Composición de pseudo aminoácidos (PseAAC) hidro_mass
0.20
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
−0.20


nematoda no_efectores dataset 1 Composición de pseudo aminoácidos (PseAAC) hidro_mass
0.15
0.10
0.05
0.00
−0.05
−0.10

4 Composición de pseudo aminoácidos (PseAAC) mass

[7]:
```
#mass
transf = "Composición de pseudo aminoácidos (PseAAC) "
transf2 = "PseAAC"
estado = "con valores atípicos.\n"
comp = "mass"
df=""

for etiq in "efectores", "no_efectores":
    titulo = (str(transf)+" "+ str(comp)+" "+ str(etiq) + " "+ str(nombre2) +",␣
 ↪" + str(estado))
    print (str(etiq))

    if etiq == "efectores":
        df=PseAAC_mass_efec

    if etiq == "no_efectores":
        df=PseAAC_mass_no_efec

    #del df['X41']
    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")

    #Gráfica de caja y bigotes
    sns.set(style="whitegrid")
    fig , ax = plt.subplots(figsize=(14,7))
    ax = sns.boxplot(data=df)
    ax.set_title(organismo +' '+str(etiq)+" dataset "+str(dataset)+"␣
 ↪"+str(transf)+" "+str(comp)+" "+str(estado))
```

```
efectores
Composición de pseudo aminoácidos (PseAAC)  mass efectores nematoda dataset 1,
con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.011033  0.005516  0.015859  0.013791  0.004137  0.041372  0.004827
1    0.060499  0.005672  0.018906  0.020797  0.041593  0.024578  0.009453
2    0.062182  0.021848  0.028570  0.028570  0.025209  0.020167  0.013445
3    0.019429  0.019429  0.021588  0.034540  0.019429  0.051811  0.017270
4    0.048666  0.029200  0.029200  0.077866  0.058399  0.097332  0.029200
..        …         …         …         …         …         …         …
495  0.053155  0.011145  0.042009  0.056584  0.025720  0.044581  0.018004
```

```
496  0.040665  0.008133  0.032532  0.024399  0.024399  0.008133  0.016266
497  0.042758  0.017492  0.073855  0.079685  0.054419  0.046645  0.023322
498  0.059908  0.000000  0.035945  0.059908  0.047926  0.011982  0.000000
499  0.017834  0.008917  0.071335  0.089168  0.008917  0.035667  0.071335

          X7        X8        X9  …       X32       X33       X34  \
0    0.011033  0.012412  0.008274  …  0.007463  0.029439  0.062371
1    0.051046  0.018906  0.058609  …  0.007793  0.006805  0.042261
2    0.030251  0.035293  0.043695  …  0.031274  0.023233 -0.004625
3    0.023747  0.032382  0.047493  …  0.004200  0.033201  0.041191
4    0.038933  0.048666  0.087599  … -0.091650 -0.094441  0.017338
..        …         …         …  …         …         …         …
495  0.038580  0.059156  0.066015  …  0.014909  0.001893  0.022667
496  0.024399  0.016266  0.024399  …  0.005814  0.046303  0.018681
497  0.066080  0.052476  0.118556  …  0.006700 -0.034040 -0.024519
498  0.107834  0.095852  0.107834  … -0.029015  0.013332  0.102599
499  0.053501  0.044584  0.098085  … -0.030616  0.006462  0.054738

          X35       X36       X37       X38       X39       X40        X41
0    0.019566  0.035132  0.056898  0.018922  0.038238  0.055396  efectores
1    0.008482  0.025237  0.017945  0.031093  0.014473  0.016580  efectores
2    0.025357  0.016493  0.031986  0.006328  0.009847  0.023872  efectores
3    0.011525  0.050805  0.008623  0.040336 -0.003365  0.006631  efectores
4    0.011851  0.067464  0.112432  0.016849  0.018534 -0.002387  efectores
..        …         …         …         …         …         …         …
495  0.017614  0.007179  0.015244  0.010513  0.006571 -0.001972  efectores
496  0.071518  0.014114  0.004701  0.036596 -0.012723  0.026977  efectores
497 -0.007300  0.005906  0.002113  0.007611  0.006399  0.012462  efectores
498  0.003598 -0.030060 -0.081045  0.022575  0.071320  0.059839  efectores
499  0.012525  0.054801  0.027826  0.010234  0.025686  0.006829  efectores

[500 rows x 42 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass efectores nematoda dataset 1, con valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.046593    0.016615    0.037598    0.045943    0.029812    0.034736
std      0.018653    0.015674    0.022164    0.028834    0.020605    0.016342
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.035464    0.006641    0.022430    0.025996    0.015015    0.024564
50%      0.045210    0.012878    0.033550    0.040391    0.025416    0.033996
75%      0.056387    0.022361    0.048219    0.059882    0.040997    0.042146
max      0.141234    0.144685    0.129558    0.162626    0.139928    0.140465
```

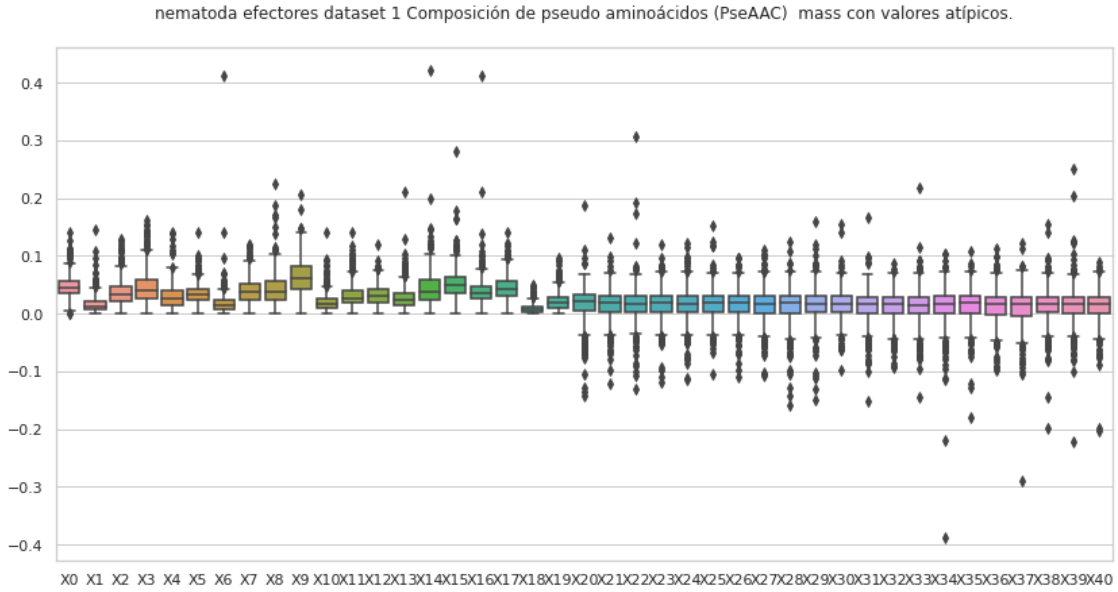
```
              X6          X7          X8          X9  …         X31  \
count  500.000000  500.000000  500.000000  500.000000  …  500.000000
mean     0.018376    0.038712    0.042338    0.063795  …    0.012242
std      0.022376    0.021414    0.027951    0.030435  …    0.028767
min      0.000000    0.000000    0.000000    0.000000  …   -0.151436
25%      0.008581    0.023107    0.022804    0.041661  …    0.000688
50%      0.015328    0.036927    0.037834    0.061037  …    0.016653
75%      0.022719    0.051750    0.056119    0.083188  …    0.028434
max      0.413017    0.118831    0.224386    0.207488  …    0.165676

              X32         X33         X34         X35         X36         X37  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.013409    0.012051    0.013065    0.014109    0.013781    0.011235
std      0.025695    0.029412    0.035374    0.028555    0.027776    0.030987
min     -0.093502   -0.143473   -0.387864   -0.179867   -0.098651   -0.289419
25%      0.000169   -0.000685    0.001112    0.000948   -0.001260   -0.003275
50%      0.017916    0.014096    0.017414    0.018105    0.015958    0.015871
75%      0.029625    0.028406    0.031676    0.030184    0.029154    0.028511
max      0.087643    0.217920    0.104202    0.107837    0.112576    0.122935

              X38         X39         X40
count  500.000000  500.000000  500.000000
mean     0.014415    0.013828    0.012377
std      0.029715    0.032357    0.028855
min     -0.199251   -0.222575   -0.202682
25%      0.001839    0.001483   -0.000346
50%      0.016603    0.016113    0.016438
75%      0.029588    0.028880    0.029150
max      0.154107    0.250419    0.088584

[8 rows x 41 columns]
```

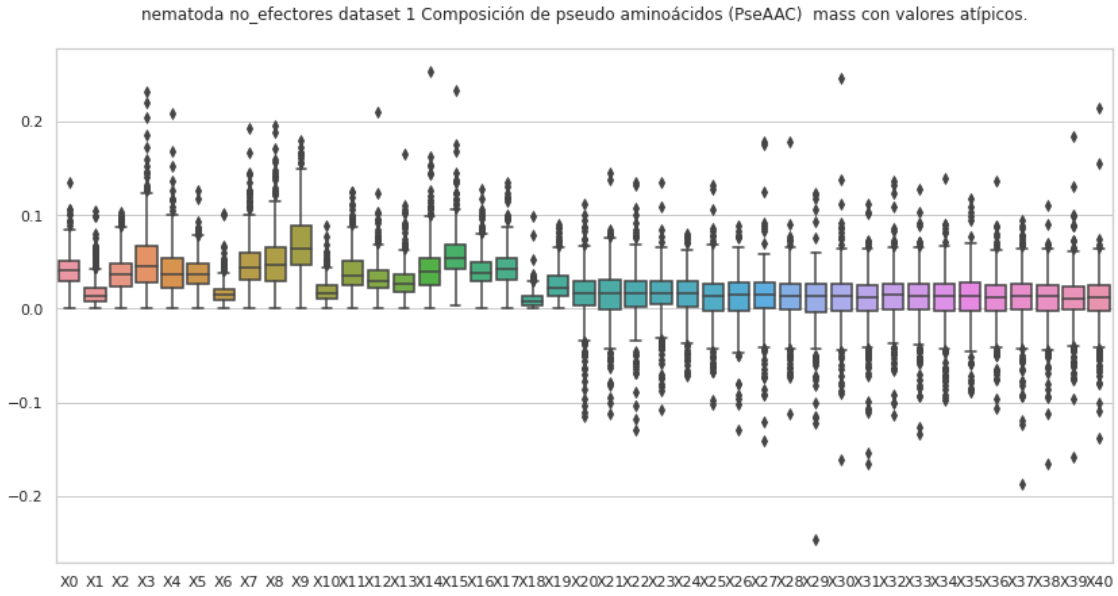
no_efectores
Composición de pseudo aminoácidos (PseAAC) mass no_efectores nematoda dataset 1, con valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.046867  0.006391  0.078821  0.085212  0.031954  0.051127  0.034085
1    0.034036  0.004816  0.034678  0.035320  0.018302  0.026009  0.006743
2    0.047187  0.029492  0.053086  0.094374  0.064882  0.070781  0.064882
3    0.048744  0.000000  0.081239  0.081239  0.048744  0.016248  0.016248
4    0.034217  0.010805  0.025213  0.012606  0.014407  0.046823  0.010805
..        …         …         …         …         …         …         …
495  0.067395  0.013479  0.013479  0.040437  0.080874  0.013479  0.000000
496  0.045237  0.031100  0.031100  0.056546  0.053719  0.079165  0.050892
```

```
497  0.048433  0.009374  0.024997  0.028122  0.043746  0.046870  0.024997
498  0.040255  0.005995  0.041968  0.057385  0.017986  0.035116  0.013704
499  0.036520  0.008300  0.018813  0.020473  0.029327  0.031540  0.011067

          X7        X8        X9  …       X32       X33       X34  \
0    0.048997  0.074560  0.104385  … -0.012831 -0.002051 -0.015210
1    0.020550  0.049449  0.040779  …  0.016827  0.018390  0.017018
2    0.047187  0.082578  0.070781  … -0.048229 -0.020324  0.061618
3    0.097487  0.113735  0.064991  …  0.000215  0.023396 -0.075293
4    0.027013  0.027013  0.036018  …  0.012175  0.034159  0.037583
..        …         …         …  …       …         …         …
495  0.026958  0.094353  0.080874  … -0.033062  0.127413 -0.094878
496  0.062201  0.053719  0.090474  …  0.001299  0.003340 -0.014217
497  0.020310  0.032809  0.059369  …  0.027369  0.015093  0.019315
498  0.029977  0.037686  0.054816  …  0.029809  0.029170  0.015489
499  0.033753  0.017707  0.047033  …  0.032240  0.030289  0.019110

          X35       X36       X37       X38       X39       X40           X41
0   -0.001257 -0.008981 -0.013941 -0.025688 -0.038934  0.007114  no_efectores
1    0.024981  0.028569  0.022347  0.029166  0.030225  0.027464  no_efectores
2    0.017833  0.020669  0.062108  0.003363 -0.061412  0.002917  no_efectores
3    0.035712  0.040634 -0.065294 -0.035898  0.072172  0.034615  no_efectores
4    0.005221  0.029115  0.024639  0.053791 -0.004428  0.029262  no_efectores
..        …         …         …         …         …         …           …
495  0.002159  0.089282 -0.123598  0.050058  0.088029 -0.079112  no_efectores
496 -0.001907 -0.030387  0.029314  0.031404  0.004555 -0.012863  no_efectores
497  0.011900  0.037273  0.033286  0.012894  0.030272  0.004534  no_efectores
498  0.018491  0.020406  0.020571  0.010161  0.014794  0.012897  no_efectores
499  0.015165  0.020317  0.023192  0.017815  0.024247  0.024973  no_efectores

[500 rows x 42 columns]


Composición de pseudo aminoácidos (PseAAC)  mass no_efectores nematoda dataset
1, con valores atípicos.
Estadísticas.

                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.041121    0.017027    0.037616    0.050437    0.040716    0.038735
std      0.017758    0.015138    0.019635    0.032492    0.026412    0.016636
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.029674    0.007224    0.023116    0.028407    0.022348    0.027104
50%      0.040849    0.013326    0.036343    0.045322    0.035985    0.036182
75%      0.051402    0.021809    0.048652    0.067080    0.053727    0.047643
max      0.135045    0.104275    0.102428    0.230963    0.208315    0.126465

                X6          X7          X8          X9  …         X31  \
```

```
count  500.000000  500.000000  500.000000  500.000000  …  500.000000
mean     0.016602    0.048117    0.050929    0.069275  …    0.009560
std      0.012533    0.026981    0.030703    0.032611  …    0.028958
min      0.000000    0.000000    0.000000    0.000000  …   -0.166165
25%      0.008822    0.030648    0.029567    0.047017  …   -0.001998
50%      0.014543    0.044334    0.046211    0.064684  …    0.012062
75%      0.021265    0.059734    0.065497    0.088839  …    0.024691
max      0.101615    0.192514    0.195426    0.179273  …    0.111747

              X32         X33         X34         X35         X36         X37  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.012521    0.011278    0.010859    0.011398    0.009937    0.010540
std      0.026769    0.026969    0.027467    0.026238    0.025641    0.029268
min     -0.113139   -0.134209   -0.097648   -0.088878   -0.105802   -0.187765
25%     -0.000919   -0.000658   -0.001978   -0.001623   -0.001850   -0.001424
50%      0.014345    0.014090    0.013656    0.013898    0.011663    0.012989
75%      0.026091    0.026332    0.025858    0.027640    0.024469    0.025891
max      0.136017    0.127413    0.138927    0.117343    0.136065    0.093805

              X38         X39         X40
count  500.000000  500.000000  500.000000
mean     0.009549    0.010079    0.009370
std      0.027266    0.028017    0.028884
min     -0.165516   -0.158021   -0.138064
25%     -0.002609   -0.001597   -0.001978
50%      0.012882    0.010442    0.011670
75%      0.024896    0.024138    0.025230
max      0.110309    0.183541    0.213801

[8 rows x 41 columns]
```





4.1 Composición de pseudo aminoácidos (PseAAC) mass, sin valores atípicos

[8]:
```
#mass
transf = "Composición de pseudo aminoácidos (PseAAC) "
transf2 = "PseAAC"
estado = "sin valores atípicos.\n"
comp = "mass"
```

```
df=""

out = (str(r3) + '/ds' + str(dataset) + '_' + str(transf2) + '_' + str(comp) +␣
 ↪'_' + str(organismo) + '.csv')
os.makedirs(str(r3), exist_ok=True)
df_out = pd.DataFrame()

for etiq in "efectores", "no_efectores":
    titulo = (str(transf)+" "+ str(comp)+" "+ str(etiq) + " "+ str(nombre2) +",␣
 ↪" + str(estado))

    if etiq == "efectores":
        df=PseAAC_mass_efec

    if etiq == "no_efectores":
        df=PseAAC_mass_no_efec

    del df['X41']
    df = (df[(np.abs(stats.zscore(df)) < 3).all(axis=1)])
    df['X41'] = etiq
    df_out = pd.concat([df_out,df])

    #Guarda la lista csv sin valores atípicos.
    df_out.to_csv(str(out), index=False, header=False)

    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")

    #Gráfica de caja y bigotes
    sns.set(style="whitegrid")
    fig , ax = plt.subplots(figsize=(14,7))
    ax = sns.boxplot(data=df)
    ax.set_title(organismo +' '+str(etiq)+" dataset "+str(dataset)+"␣
 ↪"+str(transf)+" "+str(comp))
```

```
Composición de pseudo aminoácidos (PseAAC)  mass efectores nematoda dataset 1,
sin valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.011033  0.005516  0.015859  0.013791  0.004137  0.041372  0.004827
1    0.060499  0.005672  0.018906  0.020797  0.041593  0.024578  0.009453
2    0.062182  0.021848  0.028570  0.028570  0.025209  0.020167  0.013445
3    0.019429  0.019429  0.021588  0.034540  0.019429  0.051811  0.017270
```

```
5    0.046636  0.002743  0.063095  0.052122  0.027433  0.030176  0.010973
..        …         …         …         …         …         …
494  0.047831  0.020499  0.035684  0.032647  0.028091  0.034165  0.015944
495  0.053155  0.011145  0.042009  0.056584  0.025720  0.044581  0.018004
496  0.040665  0.008133  0.032532  0.024399  0.024399  0.008133  0.016266
497  0.042758  0.017492  0.073855  0.079685  0.054419  0.046645  0.023322
499  0.017834  0.008917  0.071335  0.089168  0.008917  0.035667  0.071335

           X7        X8        X9  …       X32       X33       X34  \
0    0.011033  0.012412  0.008274  …  0.007463  0.029439  0.062371
1    0.051046  0.018906  0.058609  …  0.007793  0.006805  0.042261
2    0.030251  0.035293  0.043695  …  0.031274  0.023233 -0.004625
3    0.023747  0.032382  0.047493  …  0.004200  0.033201  0.041191
5    0.046636  0.065839  0.063095  … -0.005931  0.003467  0.037311
..        …         …         …  …         …         …         …
494  0.052387  0.031128  0.047831  …  0.005972  0.019853  0.015686
495  0.038580  0.059156  0.066015  …  0.014909  0.001893  0.022667
496  0.024399  0.016266  0.024399  …  0.005814  0.046303  0.018681
497  0.066080  0.052476  0.118556  …  0.006700 -0.034040 -0.024519
499  0.053501  0.044584  0.098085  … -0.030616  0.006462  0.054738

           X35       X36       X37       X38       X39       X40         X41
0    0.019566  0.035132  0.056898  0.018922  0.038238  0.055396  efectores
1    0.008482  0.025237  0.017945  0.031093  0.014473  0.016580  efectores
2    0.025357  0.016493  0.031986  0.006328  0.009847  0.023872  efectores
3    0.011525  0.050805  0.008623  0.040336 -0.003365  0.006631  efectores
5    0.021028  0.019397 -0.015652  0.009786 -0.003042  0.001379  efectores
..        …         …         …         …         …         …         …
494  0.032051  0.026653 -0.001474  0.019989  0.014826  0.015333  efectores
495  0.017614  0.007179  0.015244  0.010513  0.006571 -0.001972  efectores
496  0.071518  0.014114  0.004701  0.036596 -0.012723  0.026977  efectores
497 -0.007300  0.005906  0.002113  0.007611  0.006399  0.012462  efectores
499  0.012525  0.054801  0.027826  0.010234  0.025686  0.006829  efectores

[417 rows x 42 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass efectores nematoda dataset 1, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  417.000000  417.000000  417.000000  417.000000  417.000000  417.000000
mean     0.044975    0.014831    0.035033    0.041079    0.026801    0.033011
std      0.015393    0.011872    0.018295    0.022499    0.016387    0.012524
min      0.005063    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.035504    0.006507    0.021853    0.024342    0.014248    0.024467
50%      0.044195    0.012295    0.032103    0.037680    0.024413    0.033227
```

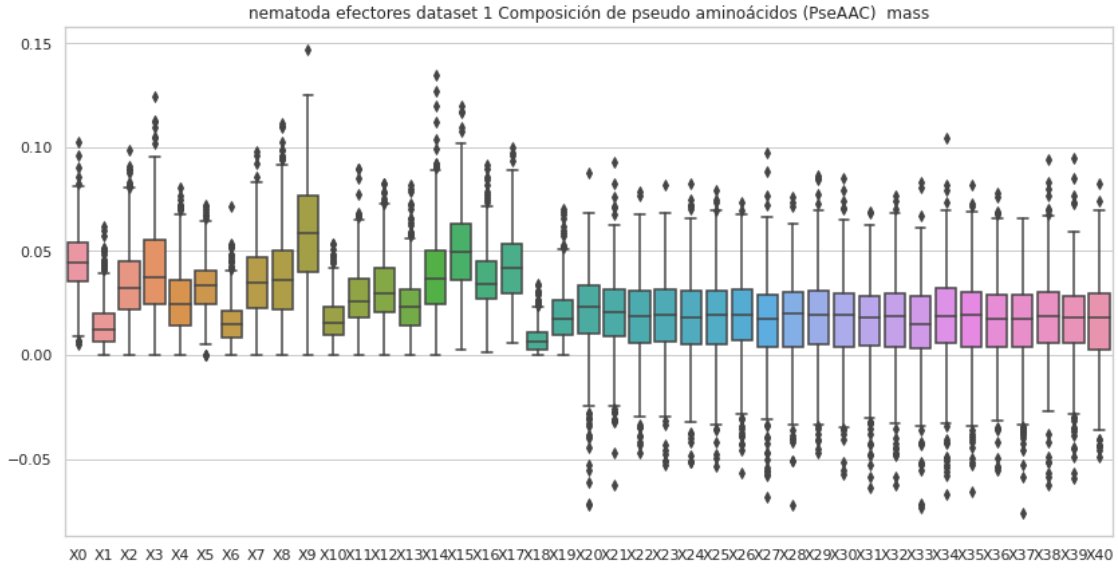
```
75%      0.054011    0.020139    0.045361    0.055151    0.036163    0.040547
max      0.102433    0.062117    0.098496    0.124390    0.080765    0.072309

               X6          X7          X8          X9  …         X31  \
count  417.000000  417.000000  417.000000  417.000000  …  417.000000
mean     0.016131    0.036098    0.038129    0.058839  …    0.015737
std      0.010749    0.018678    0.021949    0.026300  …    0.020669
min      0.000000    0.000000    0.000000    0.000000  …   -0.063695
25%      0.008348    0.022370    0.022023    0.039731  …    0.004807
50%      0.014593    0.034850    0.035829    0.058476  …    0.018119
75%      0.021450    0.047223    0.050356    0.076508  …    0.028575
max      0.071335    0.097900    0.111300    0.146906  …    0.068700

               X32         X33         X34         X35         X36         X37  \
count  417.000000  417.000000  417.000000  417.000000  417.000000  417.000000
mean     0.015457    0.013973    0.017316    0.017133    0.016411    0.014816
std      0.021930    0.021643    0.023049    0.022382    0.020960    0.021727
min     -0.062748   -0.073825   -0.066973   -0.065928   -0.055602   -0.076086
25%      0.003538    0.003184    0.005526    0.003999    0.003824    0.003544
50%      0.018856    0.014864    0.018812    0.019566    0.017106    0.017099
75%      0.029555    0.028361    0.031920    0.030422    0.028976    0.028858
max      0.076587    0.082975    0.104202    0.081855    0.077960    0.065914

               X38         X39         X40
count  417.000000  417.000000  417.000000
mean     0.017632    0.015959    0.015618
std      0.021680    0.021012    0.020976
min     -0.062627   -0.059063   -0.049090
25%      0.005506    0.005541    0.002494
50%      0.018808    0.017882    0.017887
75%      0.030333    0.028369    0.029869
max      0.094200    0.095021    0.082648

[8 rows x 41 columns]
```

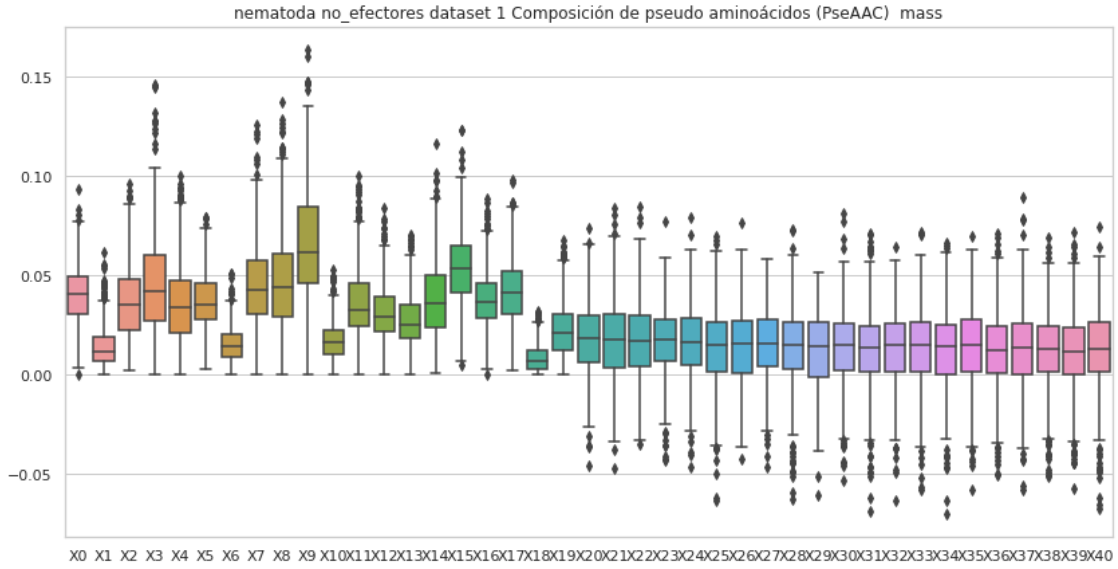
Composición de pseudo aminoácidos (PseAAC) mass no_efectores nematoda dataset 1, sin valores atípicos.
Valores del documento csv.

```
          X0        X1        X2        X3        X4        X5        X6  \
0    0.046867  0.006391  0.078821  0.085212  0.031954  0.051127  0.034085
1    0.034036  0.004816  0.034678  0.035320  0.018302  0.026009  0.006743
4    0.034217  0.010805  0.025213  0.012606  0.014407  0.046823  0.010805
5    0.042032  0.003233  0.033410  0.038798  0.015627  0.035565  0.007005
6    0.041102  0.022606  0.055488  0.073983  0.045212  0.026716  0.020551
..        …         …         …         …         …         …         …
```

```
494  0.044725  0.011181  0.069883  0.050316  0.030749  0.061497  0.027953
496  0.045237  0.031100  0.031100  0.056546  0.053719  0.079165  0.050892
497  0.048433  0.009374  0.024997  0.028122  0.043746  0.046870  0.024997
498  0.040255  0.005995  0.041968  0.057385  0.017986  0.035116  0.013704
499  0.036520  0.008300  0.018813  0.020473  0.029327  0.031540  0.011067

          X7        X8        X9  …       X32       X33       X34  \
0    0.048997  0.074560  0.104385  … -0.012831 -0.002051 -0.015210
1    0.020550  0.049449  0.040779  …  0.016827  0.018390  0.017018
4    0.027013  0.027013  0.036018  …  0.012175  0.034159  0.037583
5    0.027482  0.039337  0.038260  …  0.026102  0.021754  0.023069
6    0.047267  0.104810  0.067818  …  0.011729  0.019343  0.008454
..        …         …         …  …         …         …
494  0.047521  0.064293  0.086655  … -0.018654  0.028866  0.016814
496  0.062201  0.053719  0.090474  …  0.001299  0.003340 -0.014217
497  0.020310  0.032809  0.059369  …  0.027369  0.015093  0.019315
498  0.029977  0.037686  0.054816  …  0.029809  0.029170  0.015489
499  0.033753  0.017707  0.047033  …  0.032240  0.030289  0.019110

          X35       X36       X37       X38       X39       X40           X41
0   -0.001257 -0.008981 -0.013941 -0.025688 -0.038934  0.007114  no_efectores
1    0.024981  0.028569  0.022347  0.029166  0.030225  0.027464  no_efectores
4    0.005221  0.029115  0.024639  0.053791 -0.004428  0.029262  no_efectores
5    0.023416  0.017135  0.021407  0.027436  0.034564  0.023520  no_efectores
6   -0.000162 -0.004924  0.002814  0.009534 -0.000191  0.016415  no_efectores
..        …         …         …         …         …         …           …
494  0.010787  0.005589  0.035855 -0.028080  0.018643  0.037970  no_efectores
496 -0.001907 -0.030387  0.029314  0.031404  0.004555 -0.012863  no_efectores
497  0.011900  0.037273  0.033286  0.012894  0.030272  0.004534  no_efectores
498  0.018491  0.020406  0.020571  0.010161  0.014794  0.012897  no_efectores
499  0.015165  0.020317  0.023192  0.017815  0.024247  0.024973  no_efectores

[405 rows x 42 columns]


Composición de pseudo aminoácidos (PseAAC)  mass no_efectores nematoda dataset
1, sin valores atípicos.
Estadísticas.

                X0          X1          X2          X3          X4          X5  \
count   405.000000  405.000000  405.000000  405.000000  405.000000  405.000000
mean      0.040464    0.014814    0.036243    0.045448    0.036729    0.037660
std       0.014298    0.011214    0.017732    0.026025    0.021070    0.013807
min       0.000000    0.000000    0.002196    0.000000    0.000000    0.003109
25%       0.030908    0.006829    0.022734    0.027336    0.021037    0.027809
50%       0.040847    0.011922    0.035344    0.042339    0.034239    0.035614
75%       0.049729    0.019182    0.047958    0.060420    0.047826    0.046164
max       0.093542    0.061704    0.095807    0.146183    0.099827    0.079469
```

```
              X6          X7          X8          X9  …         X31  \
count  405.000000  405.000000  405.000000  405.000000  …  405.000000
mean     0.015527    0.045333    0.046819    0.066274  …    0.012907
std      0.009431    0.022201    0.025219    0.029900  …    0.020361
min      0.000000    0.000000    0.000000    0.000000  …   -0.068399
25%      0.009264    0.030602    0.029286    0.046029  …    0.001330
50%      0.014222    0.043015    0.043877    0.061808  …    0.013474
75%      0.020662    0.057581    0.061076    0.084402  …    0.024646
max      0.050892    0.125680    0.137445    0.163550  …    0.071179

              X32         X33         X34         X35         X36         X37  \
count  405.000000  405.000000  405.000000  405.000000  405.000000  405.000000
mean     0.013918    0.013527    0.012328    0.013726    0.012001    0.013490
std      0.019387    0.019573    0.020568    0.019899    0.019732    0.020587
min     -0.063224   -0.058032   -0.069913   -0.057997   -0.050361   -0.057674
25%      0.001474    0.001502    0.000222    0.001785    0.001012    0.000393
50%      0.015349    0.015049    0.014155    0.015325    0.012725    0.014066
75%      0.026054    0.026929    0.025261    0.027611    0.024531    0.025587
max      0.064600    0.071633    0.066547    0.070050    0.071315    0.089100

              X38         X39         X40
count  405.000000  405.000000  405.000000
mean     0.012131    0.011414    0.012668
std      0.020458    0.019502    0.021371
min     -0.050794   -0.056943   -0.067068
25%      0.001863    0.000507    0.001908
50%      0.013272    0.011799    0.013169
75%      0.024698    0.023752    0.026477
max      0.069053    0.072084    0.074227

[8 rows x 41 columns]
```





5 Composición de pseudo aminoácidos (PseAAC) hidro

[9]:
```
#hidro
transf = "Composición de pseudo aminoácidos (PseAAC) "
transf2 = "PseAAC"
estado = "con valores atípicos.\n"
comp = "hidro"
df=""
```

```
for etiq in "efectores", "no_efectores":
    titulo = (str(transf)+" "+ str(comp)+" "+ str(etiq) + " "+ str(nombre2) +",␣
 ↪" + str(estado))
    print (str(etiq))

    if etiq == "efectores":
        df=PseAAC_hidro_efec

    if etiq == "no_efectores":
        df=PseAAC_hidro_no_efec

    #del df['X62']
    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")

    #Gráfica de caja y bigotes
    sns.set(style="whitegrid")
    fig , ax = plt.subplots(figsize=(14,7))
    ax = sns.boxplot(data=df)
    ax.set_title(organismo +' '+str(etiq)+" dataset "+str(dataset)+"␣
 ↪"+str(transf)+" "+str(comp)+" "+str(estado))
```

```
efectores
Composición de pseudo aminoácidos (PseAAC)  hidro efectores nematoda dataset 1,
con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.016832  0.008416  0.024196  0.021040  0.006312  0.063121  0.007364
1    0.037941  0.003557  0.011857  0.013042  0.026084  0.015413  0.005928
2    0.100537  0.035324  0.046193  0.046193  0.040758  0.032607  0.021738
3    0.023116  0.023116  0.025685  0.041096  0.023116  0.061644  0.020548
4    0.046992  0.028195  0.028195  0.075187  0.056390  0.093984  0.028195
..        …         …         …         …         …         …         …
495  0.048572  0.010184  0.038387  0.051706  0.023503  0.040738  0.016452
496  0.091767  0.018353  0.073414  0.055060  0.055060  0.018353  0.036707
497  0.038059  0.015570  0.065738  0.070928  0.048439  0.041519  0.020760
498  0.116485  0.000000  0.069891  0.116485  0.093188  0.023297  0.000000
499  0.014713  0.007357  0.058852  0.073565  0.007357  0.029426  0.058852

           X7        X8        X9  …        X53       X54       X55  \
0    0.016832  0.018936  0.012624  …   0.001304  0.008445  0.013798
1    0.032013  0.011857  0.036755  … -0.007831  0.011100 -0.005046
```

```
2    0.048910  0.057062  0.070648  …  0.006062  0.017863  0.002463
3    0.028253  0.038527  0.056507  …  0.005163  0.013533  0.012017
4    0.037594  0.046992  0.084585  …  0.008991  0.014513 -0.041273
..        …         …        …  …        …         …
495  0.035254  0.054056  0.060323  … -0.010753  0.014232  0.021963
496  0.055060  0.036707  0.055060  … -0.000876 -0.034641 -0.049125
497  0.058819  0.046709  0.105528  … -0.035465 -0.009834  0.002999
498  0.209673  0.186376  0.209673  …  0.099808 -0.058381  0.001244
499  0.044139  0.036783  0.080922  … -0.042523 -0.010360  0.052927

          X56       X57       X58       X59       X60       X61        X62
0   -0.008210  0.017539  0.004511  0.016270  0.019016  0.024980  efectores
1    0.017138 -0.001603  0.033335 -0.001832  0.015425  0.010694  efectores
2    0.001579  0.007489 -0.009335  0.023331 -0.010945  0.009294  efectores
3    0.018044  0.009833 -0.000918  0.007179  0.005381  0.035723  efectores
4   -0.009227 -0.035813 -0.012579 -0.023385  0.011126  0.088548  efectores
..        …         …         …         …         …         …
495  0.039595  0.055112 -0.019894 -0.004120 -0.016945  0.001826  efectores
496 -0.031759 -0.052426  0.000344  0.005275  0.147196  0.126221  efectores
497  0.026826  0.015028 -0.025032  0.009194 -0.032671 -0.010230  efectores
498  0.074302  0.047339  0.073987  0.088094  0.186697  0.074686  efectores
499  0.101629  0.059610 -0.028576  0.017423  0.029647  0.018601  efectores

[500 rows x 63 columns]


Composición de pseudo aminoácidos (PseAAC)  hidro efectores nematoda dataset 1,
con valores atípicos.
Estadísticas.

               X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.054585    0.019522    0.039717    0.045780    0.031101    0.040921
std      0.035579    0.023926    0.025689    0.026693    0.021974    0.029649
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.028650    0.006218    0.020500    0.026776    0.016014    0.019824
50%      0.045684    0.013021    0.035870    0.043534    0.027519    0.032609
75%      0.073293    0.025376    0.053602    0.061154    0.042849    0.057736
max      0.213069    0.230031    0.187034    0.260310    0.239935    0.223123

               X6          X7          X8          X9  …         X52  \
count  500.000000  500.000000  500.000000  500.000000  …  500.000000
mean     0.019194    0.042098    0.042306    0.069653  …   -0.002132
std      0.018989    0.029320    0.029625    0.047169  …    0.047870
min      0.000000    0.000000    0.000000    0.000000  …   -0.483428
25%      0.007995    0.021572    0.023570    0.039202  …   -0.018435
50%      0.015009    0.037558    0.038089    0.062874  …    0.001454
75%      0.025832    0.055216    0.054802    0.090161  …    0.019224
```

```
max      0.244237    0.239935    0.319914    0.479871  …    0.184512

              X53         X54         X55         X56         X57         X58  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.005369    0.000019    0.007269    0.003243    0.009444    0.001408
std      0.036569    0.043421    0.039948    0.050446    0.037474    0.066427
min     -0.290454   -0.340591   -0.314496   -0.305220   -0.143594   -1.131739
25%     -0.006708   -0.017069   -0.006820   -0.012099   -0.006021   -0.013643
50%      0.007905    0.005243    0.010427    0.005887    0.009270    0.005103
75%      0.024163    0.020391    0.026542    0.021385    0.025346    0.021232
max      0.116897    0.175600    0.307548    0.534410    0.240457    0.339083

              X59         X60         X61
count  500.000000  500.000000  500.000000
mean     0.008168    0.001318    0.008513
std      0.037564    0.055117    0.044119
min     -0.367727   -0.395567   -0.362127
25%     -0.004646   -0.014856   -0.007282
50%      0.011658    0.006248    0.009322
75%      0.025062    0.022283    0.027758
max      0.194314    0.249462    0.221044

[8 rows x 62 columns]
```

no_efectores
Composición de pseudo aminoácidos (PseAAC) hidro no_efectores nematoda dataset 1, con valores atípicos.
Valores del documento csv.

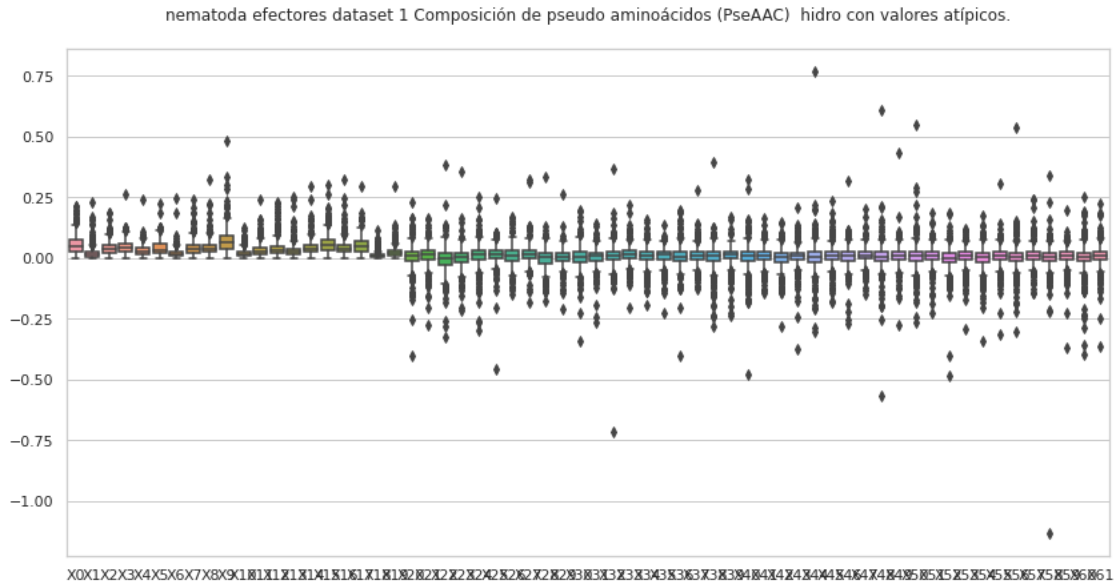
```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.023777  0.003242  0.039988  0.043230  0.016211  0.025938  0.017292
1    0.031814  0.004502  0.032414  0.033014  0.017107  0.024310  0.006303
2    0.035098  0.021936  0.039485  0.070196  0.048260  0.052647  0.048260
3    0.021254  0.000000  0.035424  0.035424  0.021254  0.007085  0.007085
4    0.073817  0.023311  0.054392  0.027196  0.031081  0.101013  0.023311
..        …         …         …         …         …         …         …
495  0.101545  0.020309  0.020309  0.060927  0.121854  0.020309  0.000000
496  0.019763  0.013587  0.013587  0.024704  0.023469  0.034586  0.022234
497  0.074458  0.014411  0.038430  0.043234  0.067252  0.072056  0.038430
498  0.036280  0.005403  0.037824  0.051719  0.016210  0.031649  0.012351
499  0.031114  0.007071  0.016029  0.017443  0.024986  0.026872  0.009429

           X7        X8        X9  …       X53       X54       X55  \
0    0.024857  0.037826  0.052957  …  0.017613 -0.010963 -0.004968
1    0.019208  0.046220  0.038116  …  0.025055 -0.011673  0.005956
2    0.035098  0.061421  0.052647  …  0.044978  0.044108  0.014685
```

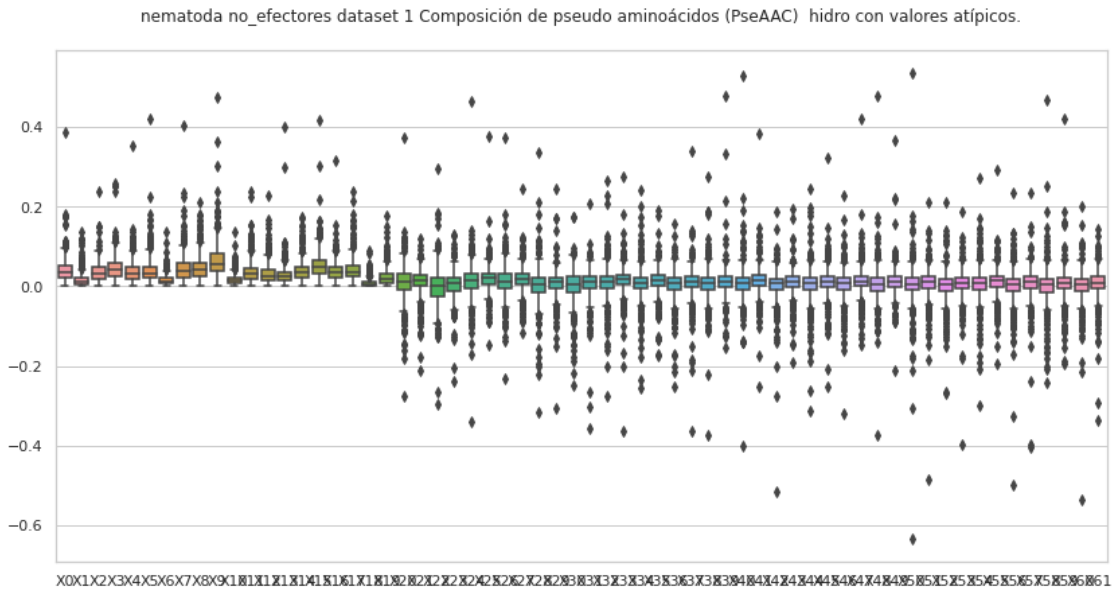
```
3    0.042509  0.049593  0.028339  … -0.003379  0.029454  0.094346
4    0.058277  0.058277  0.077702  …  0.025436  0.034689  0.020357
..        …         …         …  …         …         …
495  0.040618  0.142163  0.121854  …  0.188072 -0.144484 -0.093615
496  0.027175  0.023469  0.039527  …  0.019433  0.004938  0.013562
497  0.031224  0.050439  0.091271  … -0.024218  0.040317  0.023094
498  0.027017  0.033965  0.049403  …  0.003848  0.002384  0.012367
499  0.028757  0.015086  0.040072  …  0.011906  0.021035  0.008489

          X56       X57       X58       X59       X60       X61           X62
0   -0.001419  0.038054  0.003762  0.022023 -0.008489  0.023650  no_efectores
1    0.003976  0.015248  0.021604  0.030078 -0.000416  0.014891  no_efectores
2   -0.013829  0.047256 -0.010159  0.011265 -0.014982  0.031032  no_efectores
3    0.045705  0.028998 -0.034468  0.011494 -0.002549  0.045176  no_efectores
4    0.028544 -0.041415 -0.004727  0.003861 -0.010506  0.008987  no_efectores
..        …         …         …         …         …         …           …
495 -0.082351 -0.150824 -0.164238 -0.040028 -0.064787 -0.064561  no_efectores
496 -0.015376 -0.001907  0.006194  0.011290  0.004097  0.005137  no_efectores
497 -0.011188 -0.007982 -0.010689 -0.052631 -0.017238  0.008525  no_efectores
498 -0.008482  0.009976  0.002953  0.018947  0.028452  0.054550  no_efectores
499  0.020346  0.011857  0.020912  0.016246  0.007438  0.005935  no_efectores

[500 rows x 63 columns]


Composición de pseudo aminoácidos (PseAAC)  hidro no_efectores nematoda dataset
1, con valores atípicos.
Estadísticas.

                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.040419    0.016743    0.034534    0.044373    0.036287    0.039590
std      0.030696    0.018292    0.023405    0.031290    0.027567    0.033676
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.021838    0.005355    0.017722    0.024605    0.019249    0.020612
50%      0.035014    0.011939    0.031467    0.040226    0.031682    0.032076
75%      0.051984    0.020646    0.046928    0.058406    0.048236    0.049752
max      0.386874    0.135455    0.237592    0.258831    0.354125    0.419918

                X6          X7          X8          X9  …         X52  \
count  500.000000  500.000000  500.000000  500.000000  …  500.000000
mean     0.015174    0.043993    0.045106    0.064036  …    0.000117
std      0.013743    0.033899    0.030004    0.044041  …    0.037951
min      0.000000    0.000000    0.000000    0.000000  …   -0.268133
25%      0.006348    0.022486    0.023443    0.037385  …   -0.013180
50%      0.012165    0.037924    0.040183    0.056860  …    0.003833
75%      0.020512    0.057000    0.058302    0.081431  …    0.017371
max      0.135455    0.404714    0.210227    0.475184  …    0.211968
```

	X53	X54	X55	X56	X57	X58	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.006258	0.005584	0.011131	0.001027	0.006529	0.001078	
std	0.035942	0.041074	0.032864	0.044715	0.041219	0.047058	
min	-0.398920	-0.299672	-0.195753	-0.500544	-0.406032	-0.242604	
25%	-0.006330	-0.009438	-0.001452	-0.011811	-0.005601	-0.014346	
50%	0.008410	0.007850	0.013037	0.004749	0.010256	0.004327	
75%	0.022376	0.021284	0.024797	0.018674	0.024749	0.018379	
max	0.188072	0.272108	0.293012	0.235824	0.233347	0.466744	

	X59	X60	X61
count	500.000000	500.000000	500.000000
mean	0.007071	0.002093	0.006621
std	0.039599	0.043819	0.037854
min	-0.194630	-0.534784	-0.337981
25%	-0.006941	-0.011539	-0.006047
50%	0.008391	0.005032	0.009027
75%	0.022283	0.019277	0.024276
max	0.418960	0.200481	0.143819

[8 rows x 62 columns]





5.1 Composición de pseudo aminoácidos (PseAAC) hidro, sin valores atípicos

[10]:
```
#hidro
transf = "Composición de pseudo aminoácidos (PseAAC) "
transf2 = "PseAAC"
estado = "sin valores atípicos.\n"
comp = "hidro"
df=""

out = (str(r3) + '/ds' + str(dataset) + '_' + str(transf2) + '_' + str(comp) +␣
 ↪'_' + str(organismo) + '.csv')
os.makedirs(str(r3), exist_ok=True)
df_out = pd.DataFrame()


for etiq in "efectores", "no_efectores":
    titulo = (str(transf) +" "+ str(etiq) + " " + str(nombre2) + ", " +␣
 ↪str(estado))
    print (str(etiq))

    if etiq == "efectores":
        df=PseAAC_hidro_efec

    if etiq == "no_efectores":
        df=PseAAC_hidro_no_efec

    del df['X62']
```

```
    #Se eliminan todas las filas que tengan valores atípicos en al menos una de␣
↪sus columnas.
    df = (df[(np.abs(stats.zscore(df)) < 3).all(axis=1)])
    df['X62'] = etiq
    df_out = pd.concat([df_out,df])

    #Guarda la lista csv sin valores atípicos.
    df_out.to_csv(str(out), index=False, header=False)


    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")

    #Gráfica de caja y bigotes
    sns.set(style="whitegrid")
    fig , ax = plt.subplots(figsize=(14,7))
    ax = sns.boxplot(data=df)
    ax.set_title(organismo +' '+str(etiq)+" dataset "+str(dataset)+"␣
↪"+str(transf)+" "+str(comp))
```

```
efectores
Composición de pseudo aminoácidos (PseAAC)  efectores nematoda dataset 1, sin
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.016832  0.008416  0.024196  0.021040  0.006312  0.063121  0.007364
1    0.037941  0.003557  0.011857  0.013042  0.026084  0.015413  0.005928
2    0.100537  0.035324  0.046193  0.046193  0.040758  0.032607  0.021738
3    0.023116  0.023116  0.025685  0.041096  0.023116  0.061644  0.020548
5    0.032592  0.001917  0.044094  0.036426  0.019172  0.021089  0.007669
..        …         …         …         …         …         …         …
492  0.053098  0.000000  0.019912  0.046461  0.006637  0.000000  0.006637
494  0.069302  0.029701  0.051701  0.047301  0.040701  0.049501  0.023101
495  0.048572  0.010184  0.038387  0.051706  0.023503  0.040738  0.016452
497  0.038059  0.015570  0.065738  0.070928  0.048439  0.041519  0.020760
499  0.014713  0.007357  0.058852  0.073565  0.007357  0.029426  0.058852

           X7        X8        X9  …       X53       X54       X55  \
0    0.016832  0.018936  0.012624  …  0.001304  0.008445  0.013798
1    0.032013  0.011857  0.036755  … -0.007831  0.011100 -0.005046
2    0.048910  0.057062  0.070648  …  0.006062  0.017863  0.002463
3    0.028253  0.038527  0.056507  …  0.005163  0.013533  0.012017
5    0.032592  0.046012  0.044094  …  0.040915  0.003552  0.035499
```

```
..        …       …       …  …       …       …       …
492  0.000000  0.046461  0.006637  …  0.000314 -0.008526  0.019925
494  0.075902  0.045101  0.069302  …  0.024843  0.026541 -0.005335
495  0.035254  0.054056  0.060323  … -0.010753  0.014232  0.021963
497  0.058819  0.046709  0.105528  … -0.035465 -0.009834  0.002999
499  0.044139  0.036783  0.080922  … -0.042523 -0.010360  0.052927

          X56       X57       X58       X59       X60       X61        X62
0   -0.008210  0.017539  0.004511  0.016270  0.019016  0.024980  efectores
1    0.017138 -0.001603  0.033335 -0.001832  0.015425  0.010694  efectores
2    0.001579  0.007489 -0.009335  0.023331 -0.010945  0.009294  efectores
3    0.018044  0.009833 -0.000918  0.007179  0.005381  0.035723  efectores
5    0.024205  0.045697 -0.007138  0.009482 -0.045781 -0.004149  efectores
..        …       …       …       …       …       …       …
492 -0.006933  0.055058  0.000527  0.071969  0.041818  0.067504  efectores
494  0.009695  0.024172  0.008151  0.017518  0.025036  0.017626  efectores
495  0.039595  0.055112 -0.019894 -0.004120 -0.016945  0.001826  efectores
497  0.026826  0.015028 -0.025032  0.009194 -0.032671 -0.010230  efectores
499  0.101629  0.059610 -0.028576  0.017423  0.029647  0.018601  efectores

[416 rows x 63 columns]


Composición de pseudo aminoácidos (PseAAC)  efectores nematoda dataset 1, sin
valores atípicos.
Estadísticas.

                X0          X1          X2          X3          X4          X5  \
count  416.000000  416.000000  416.000000  416.000000  416.000000  416.000000
mean     0.047308    0.015282    0.035115    0.041369    0.027638    0.036183
std      0.027580    0.013692    0.020130    0.021734    0.017271    0.024203
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.027116    0.005780    0.019068    0.025242    0.014374    0.018759
50%      0.041817    0.011465    0.032951    0.038692    0.025121    0.029607
75%      0.062144    0.021713    0.048821    0.056151    0.038883    0.048623
max      0.158566    0.088041    0.108712    0.116440    0.087947    0.125066

                X6          X7          X8          X9  …         X52  \
count  416.000000  416.000000  416.000000  416.000000  …  416.000000
mean     0.016287    0.037056    0.037239    0.060196  …    0.002430
std      0.011382    0.021793    0.019495    0.032899  …    0.029610
min      0.000000    0.000000    0.000000    0.003877  …   -0.109345
25%      0.007664    0.020187    0.022858    0.036579  …   -0.012360
50%      0.014068    0.033920    0.034716    0.056488  …    0.002802
75%      0.022500    0.050891    0.051306    0.078787  …    0.018977
max      0.058852    0.113241    0.103438    0.183906  …    0.090115

               X53         X54         X55         X56         X57         X58  \
```

```
count  416.000000  416.000000  416.000000  416.000000  416.000000  416.000000
mean     0.009698    0.002901    0.010666    0.004141    0.010409    0.004163
std      0.025577    0.029540    0.025295    0.028833    0.023620    0.029226
min     -0.100741   -0.114294   -0.076239   -0.100948   -0.072872   -0.151623
25%     -0.002662   -0.010504   -0.002984   -0.010122   -0.001872   -0.009533
50%      0.009328    0.005984    0.011900    0.006634    0.011281    0.005416
75%      0.024239    0.019913    0.026451    0.020500    0.024893    0.019480
max      0.106881    0.099160    0.097356    0.109453    0.075921    0.097602

              X59         X60         X61
count  416.000000  416.000000  416.000000
mean     0.009996    0.004850    0.011483
std      0.024373    0.032190    0.027889
min     -0.090506   -0.140660   -0.088974
25%     -0.002174   -0.007528   -0.004145
50%      0.012000    0.007128    0.010746
75%      0.023969    0.020760    0.027093
max      0.092348    0.115306    0.099493

[8 rows x 62 columns]
```

no_efectores

Composición de pseudo aminoácidos (PseAAC) no_efectores nematoda dataset 1, sin valores atípicos.

Valores del documento csv.

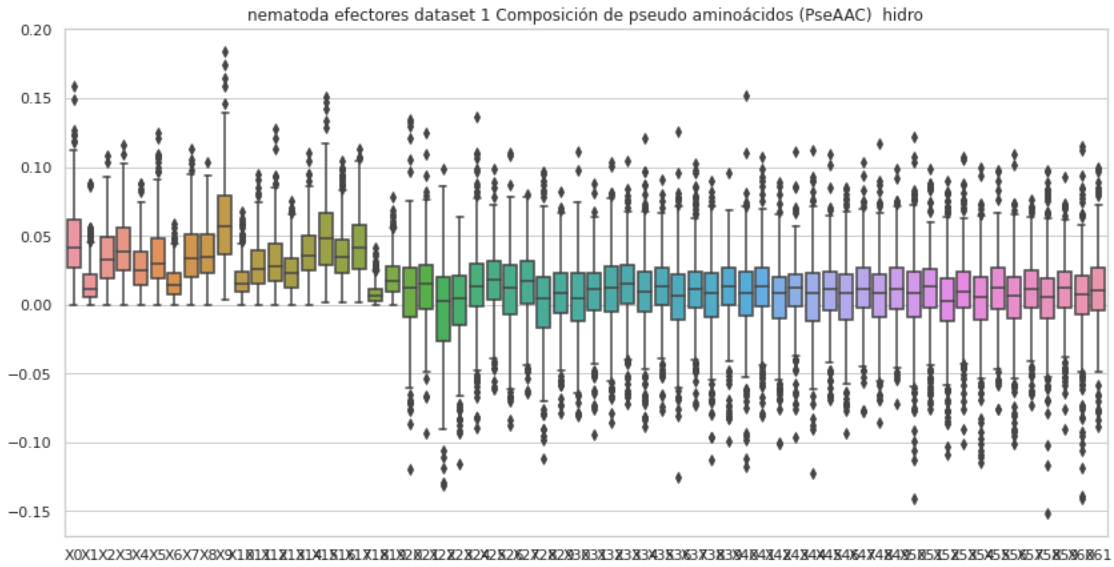
```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.023777  0.003242  0.039988  0.043230  0.016211  0.025938  0.017292
1    0.031814  0.004502  0.032414  0.033014  0.017107  0.024310  0.006303
2    0.035098  0.021936  0.039485  0.070196  0.048260  0.052647  0.048260
3    0.021254  0.000000  0.035424  0.035424  0.021254  0.007085  0.007085
4    0.073817  0.023311  0.054392  0.027196  0.031081  0.101013  0.023311
..        …         …         …         …         …         …         …
494  0.036364  0.009091  0.056819  0.040909  0.025000  0.050000  0.022727
496  0.019763  0.013587  0.013587  0.024704  0.023469  0.034586  0.022234
497  0.074458  0.014411  0.038430  0.043234  0.067252  0.072056  0.038430
498  0.036280  0.005403  0.037824  0.051719  0.016210  0.031649  0.012351
499  0.031114  0.007071  0.016029  0.017443  0.024986  0.026872  0.009429

           X7        X8        X9  …       X53       X54       X55  \
0    0.024857  0.037826  0.052957  …  0.017613 -0.010963 -0.004968
1    0.019208  0.046220  0.038116  …  0.025055 -0.011673  0.005956
2    0.035098  0.061421  0.052647  …  0.044978  0.044108  0.014685
3    0.042509  0.049593  0.028339  … -0.003379  0.029454  0.094346
4    0.058277  0.058277  0.077702  …  0.025436  0.034689  0.020357
..        …         …         …  …         …         …         …
```

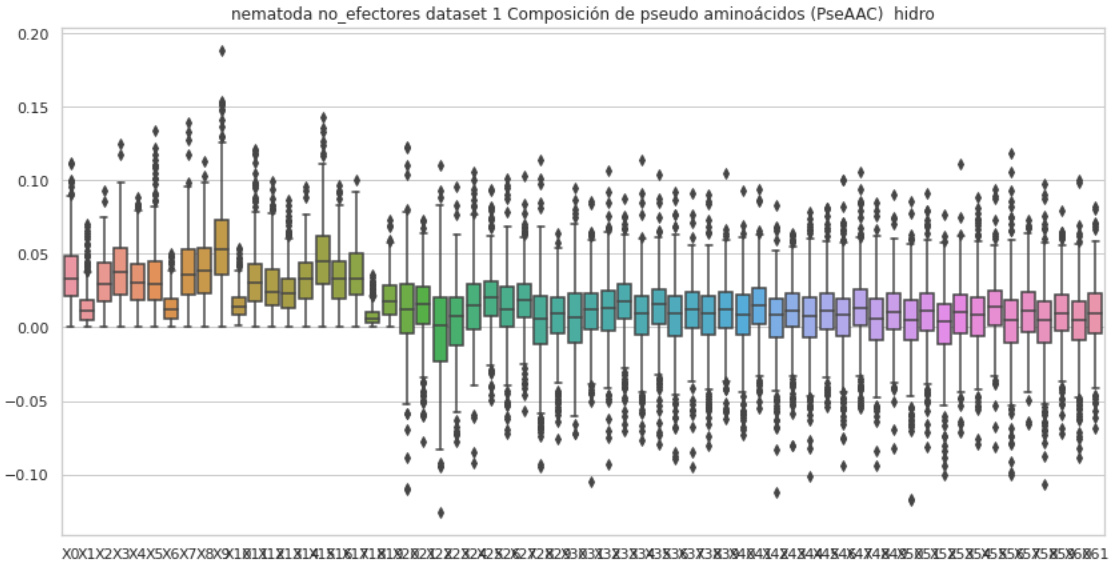
```
494  0.038637  0.052273  0.070455  … -0.006105 -0.042100 -0.020596
496  0.027175  0.023469  0.039527  …  0.019433  0.004938  0.013562
497  0.031224  0.050439  0.091271  … -0.024218  0.040317  0.023094
498  0.027017  0.033965  0.049403  …  0.003848  0.002384  0.012367
499  0.028757  0.015086  0.040072  …  0.011906  0.021035  0.008489

          X56       X57       X58       X59       X60       X61           X62
0   -0.001419  0.038054  0.003762  0.022023 -0.008489  0.023650  no_efectores
1    0.003976  0.015248  0.021604  0.030078 -0.000416  0.014891  no_efectores
2   -0.013829  0.047256 -0.010159  0.011265 -0.014982  0.031032  no_efectores
3    0.045705  0.028998 -0.034468  0.011494 -0.002549  0.045176  no_efectores
4    0.028544 -0.041415 -0.004727  0.003861 -0.010506  0.008987  no_efectores
..        …         …         …         …         …         …           …
494 -0.037581  0.020708 -0.029980  0.000525 -0.015630  0.006409  no_efectores
496 -0.015376 -0.001907  0.006194  0.011290  0.004097  0.005137  no_efectores
497 -0.011188 -0.007982 -0.010689 -0.052631 -0.017238  0.008525  no_efectores
498 -0.008482  0.009976  0.002953  0.018947  0.028452  0.054550  no_efectores
499  0.020346  0.011857  0.020912  0.016246  0.007438  0.005935  no_efectores

[432 rows x 63 columns]


Composición de pseudo aminoácidos (PseAAC)  no_efectores nematoda dataset 1, sin
valores atípicos.
Estadísticas.

                X0          X1          X2          X3          X4          X5  \
count  432.000000  432.000000  432.000000  432.000000  432.000000  432.000000
mean     0.036611    0.013966    0.031522    0.039184    0.031851    0.034553
std      0.021076    0.012593    0.017988    0.021836    0.018137    0.022378
min      0.000000    0.000000    0.000634    0.000000    0.000000    0.000000
25%      0.021061    0.005041    0.017227    0.022469    0.018697    0.018889
50%      0.033458    0.011087    0.029434    0.037755    0.030752    0.029722
75%      0.048504    0.018269    0.043618    0.053735    0.043154    0.045250
max      0.112020    0.070725    0.093440    0.124671    0.088200    0.133906

                X6          X7          X8          X9  …         X52  \
count  432.000000  432.000000  432.000000  432.000000  …  432.000000
mean     0.013724    0.039432    0.040381    0.057115  …    0.001638
std      0.009995    0.022723    0.021787    0.030728  …    0.025046
min      0.000000    0.000000    0.000000    0.000000  …   -0.100323
25%      0.006296    0.022273    0.023126    0.036050  …   -0.011774
50%      0.011761    0.036102    0.038561    0.052990  …    0.004109
75%      0.019391    0.052643    0.054305    0.073245  …    0.016148
max      0.050648    0.138992    0.112751    0.188260  …    0.076217

               X53         X54         X55         X56         X57         X58  \
count  432.000000  432.000000  432.000000  432.000000  432.000000  432.000000
```

mean	0.008952	0.007179	0.012907	0.004478	0.010210	0.003167
std	0.021841	0.024259	0.021203	0.026584	0.020749	0.026164
min	-0.070877	-0.079208	-0.081640	-0.100696	-0.066238	-0.106643
25%	-0.004374	-0.005893	0.001400	-0.010190	-0.003058	-0.010631
50%	0.010147	0.008445	0.013674	0.005260	0.011299	0.005360
75%	0.022376	0.020645	0.024680	0.018226	0.024186	0.017655
max	0.111342	0.088447	0.094346	0.118180	0.073474	0.097532

	X59	X60	X61
count	432.000000	432.000000	432.000000
mean	0.009157	0.004244	0.009534
std	0.021870	0.025574	0.022680
min	-0.072113	-0.088561	-0.068674
25%	-0.002689	-0.008457	-0.003792
50%	0.009872	0.005032	0.009885
75%	0.022006	0.018080	0.023204
max	0.079493	0.100031	0.082377

[8 rows x 62 columns]





6 Covarianza de auto cruzamiento (ACC) hidro_mass

[11]:
```
#hidro_mass
transf = "Covarianza de auto cruzamiento (ACC) "
transf2 = "ACC"
estado = "con valores atípicos.\n"
comp = "hidro_mass"
df=""

for etiq in "efectores", "no_efectores":
    titulo = (str(transf)+" "+ str(comp)+" "+ str(etiq) + " "+ str(nombre2) +",␣
 ↪" + str(estado))
    print (str(etiq))

    if etiq == "efectores":
        df=ACC_hidro_mass_efec

    if etiq == "no_efectores":
        df=ACC_hidro_mass_no_efec

    #del df['X13']
    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")
```

```
    #Gráfica de caja y bigotes
    sns.set(style="whitegrid")
    fig , ax = plt.subplots(figsize=(14,7))
    ax = sns.boxplot(data=df)
    ax.set_title(organismo +' '+str(etiq)+" dataset "+str(dataset)+"␣
↪"+str(transf)+" "+str(comp)+" "+str(estado))
```

```
efectores
Covarianza de auto cruzamiento (ACC)  hidro_mass efectores nematoda dataset 1,
con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.025211 -0.078576  0.435216 -0.017247 -0.049976  0.414317 -0.026065
1   -0.011797 -0.009725  0.084725  0.136624 -0.033562  0.097687 -0.038754
2   -0.012463 -0.014299 -0.011442  0.056073 -0.058643  0.068880 -0.024983
3    0.056212  0.029135 -0.017337  0.000754  0.019356  0.031176  0.050343
4    0.092383 -0.114493 -0.021033 -0.161981 -0.195468 -0.115002 -0.112935
..        …       …       …       …       …       …       …
495 -0.021151 -0.012916  0.025724 -0.004460  0.019319 -0.000540  0.015006
496 -0.022709  0.126096  0.111002  0.118960  0.057710 -0.056092  0.152603
497 -0.008882 -0.014505  0.001263  0.008876 -0.024333 -0.009379  0.035413
498  0.000425 -0.087246 -0.247833  0.080331  0.167989  0.061285 -0.160020
499  0.061341  0.084701 -0.003269  0.034905  0.039719  0.073065 -0.055073

           X7        X8        X9       X10       X11       X12        X13
0   -0.091376  0.311555 -0.014858  0.061172  0.272843 -0.128816  efectores
1   -0.012509  0.039248  0.015632  0.019566 -0.005492 -0.020774  efectores
2   -0.008127  0.038218  0.010554 -0.017068  0.005978  0.031625  efectores
3    0.059253  0.014641 -0.072541 -0.030492 -0.013226 -0.050396  efectores
4    0.058062  0.126407 -0.032723  0.038422  0.066819 -0.158610  efectores
..        …       …       …       …       …       …        …
495  0.026149  0.004811  0.027502 -0.018389 -0.031828  0.004571  efectores
496 -0.041177  0.097274  0.015525  0.071043  0.035047 -0.032244  efectores
497 -0.043695 -0.000724 -0.001322 -0.041151  0.002160  0.010295  efectores
498 -0.115440 -0.021992  0.249278  0.116749 -0.139009 -0.070152  efectores
499 -0.071602 -0.034598 -0.072024 -0.134974 -0.048590 -0.072310  efectores

[500 rows x 14 columns]


Covarianza de auto cruzamiento (ACC)  hidro_mass efectores nematoda dataset 1,
con valores atípicos.
Estadísticas.

                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
```

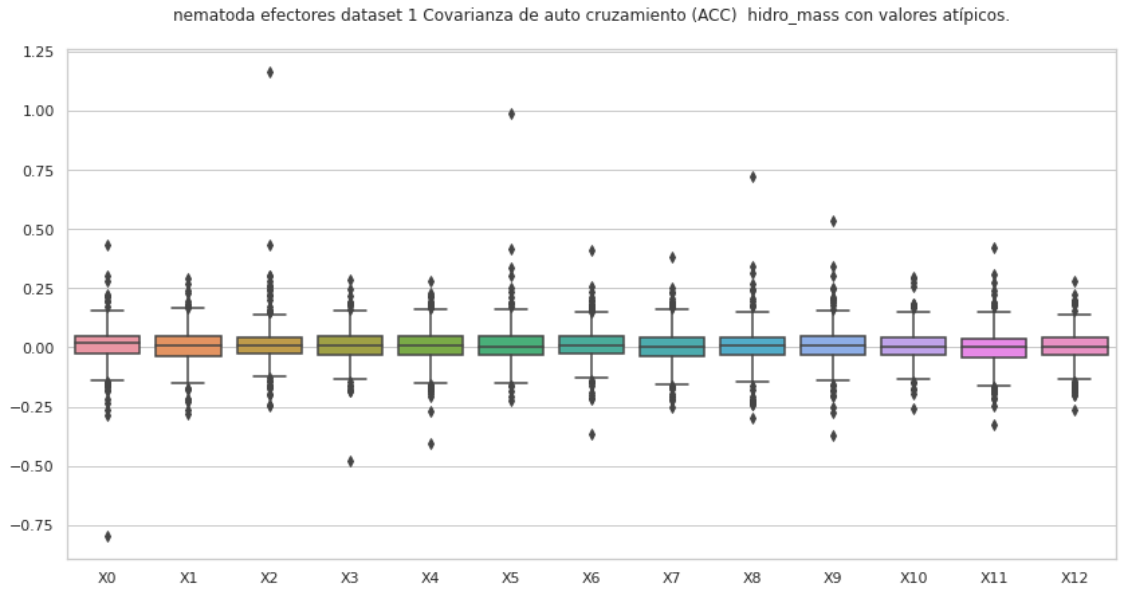
```
mean     0.008293    0.006226    0.012406    0.009666    0.005315    0.012384
std      0.082713    0.071897    0.088742    0.071130    0.072178    0.083455
min     -0.796466   -0.280452   -0.247833   -0.481303   -0.406789   -0.223372
25%     -0.027233   -0.035431   -0.028245   -0.029114   -0.033098   -0.032898
50%      0.016870    0.006580    0.007808    0.006134    0.006884    0.005189
75%      0.048568    0.046564    0.039495    0.046326    0.045247    0.047751
max      0.433674    0.291440    1.163856    0.285492    0.277449    0.985809

                X6          X7          X8          X9          X10         X11  \
count   500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean      0.009583    0.002260    0.006413    0.008848    0.007602    0.000451
std       0.068762    0.075231    0.080024    0.076167    0.066835    0.074882
min      -0.363852   -0.253643   -0.298026   -0.368876   -0.259720   -0.326409
25%      -0.025604   -0.038445   -0.034145   -0.031644   -0.031368   -0.041401
50%       0.010212    0.003801    0.006692    0.005511    0.003951    0.002629
75%       0.045489    0.042069    0.041852    0.045304    0.043775    0.038897
max       0.411898    0.380252    0.720522    0.535666    0.298103    0.421571

               X12
count   500.000000
mean      0.002565
std       0.067195
min      -0.264768
25%      -0.031034
50%       0.003773
75%       0.039637
max       0.277625



no_efectores
Covarianza de auto cruzamiento (ACC)  hidro_mass no_efectores nematoda dataset
1, con valores atípicos.
Valores del documento csv.

            X0        X1        X2        X3        X4        X5        X6  \
0     0.115441  0.038070  0.030851  0.022599  0.046505  0.028404 -0.038752
1     0.063993  0.038335  0.044468  0.043469  0.023921  0.009412  0.016765
2     0.026215 -0.154273  0.074887 -0.043504 -0.049558 -0.028002 -0.094115
3     0.050824 -0.011756  0.012374  0.055829 -0.129675 -0.021034  0.083787
4    -0.064920 -0.070168 -0.096588 -0.003268  0.022062  0.073161 -0.043712
..         …         …         …         …         …         …         …
495 -0.041672 -0.046088  0.203464  0.015147 -0.035698  0.031145 -0.035094
496 -0.022277 -0.008190  0.092112  0.049676 -0.049231  0.061640  0.075976
497  0.000781  0.035087  0.025331  0.017103  0.080298 -0.044145 -0.012913
498  0.015630  0.036722  0.011245  0.044062  0.001443  0.022589 -0.034586
499  0.058914 -0.034206  0.018069  0.077425 -0.011081  0.038940  0.049779
```

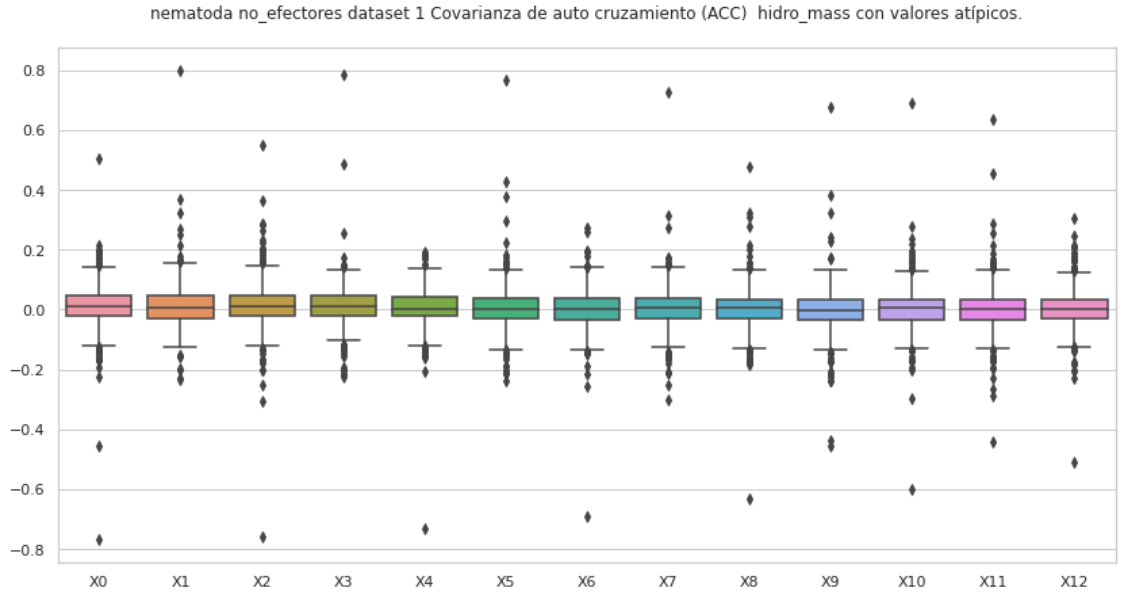
```
           X7        X8        X9       X10       X11       X12           X13
0    0.028387  0.010034  0.042835  0.029551  0.013673 -0.028602  no_efectores
1    0.038850  0.006178  0.042438  0.010026 -0.001328  0.001493  no_efectores
2   -0.143690 -0.046976  0.031613  0.139037  0.074984 -0.095019  no_efectores
3   -0.167587 -0.009602  0.134049 -0.051341 -0.196695  0.000594  no_efectores
4   -0.007955  0.092660  0.080603 -0.076962  0.052785 -0.070343  no_efectores
..        …         …         …         …         …         …           …
495 -0.022387 -0.075223  0.121284  0.130548 -0.089992 -0.073631  no_efectores
496 -0.032396 -0.032674 -0.025782  0.004230 -0.093376 -0.002864  no_efectores
497 -0.026529 -0.005030 -0.024967 -0.042186 -0.066716  0.030267  no_efectores
498  0.035446 -0.020920 -0.002452  0.011123  0.039312  0.050999  no_efectores
499  0.001646  0.034517  0.023128 -0.001730 -0.045006  0.045527  no_efectores

[500 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores nematoda dataset 1, con valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.009956    0.010884    0.012652    0.013821    0.007417    0.005228
std      0.079540    0.075256    0.079632    0.070260    0.065515    0.076149
min     -0.767623   -0.234745   -0.758319   -0.225384   -0.733605   -0.237639
25%     -0.021972   -0.029890   -0.022490   -0.018631   -0.022453   -0.031708
50%      0.009503    0.007426    0.012026    0.011271    0.003909    0.004265
75%      0.045717    0.045323    0.045454    0.046310    0.042872    0.036328
max      0.505626    0.798790    0.548220    0.784768    0.192236    0.769448

                X6          X7          X8          X9         X10         X11  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.001699    0.005524    0.002175   -0.000410    0.002443    0.000518
std      0.067517    0.070978    0.071806    0.075937    0.075702    0.075866
min     -0.690465   -0.303143   -0.632800   -0.456386   -0.601473   -0.439466
25%     -0.032499   -0.027869   -0.031678   -0.032244   -0.033071   -0.033650
50%      0.002496    0.006063    0.005012   -0.000516    0.004649    0.000584
75%      0.037465    0.040050    0.034725    0.035594    0.032693    0.033282
max      0.275249    0.726801    0.475577    0.676298    0.688794    0.634251

               X12
count  500.000000
mean     0.004057
std      0.067576
min     -0.511735
25%     -0.028454
50%      0.003867
75%      0.034910
```

```
max       0.303997
```


nematoda efectores dataset 1 Covarianza de auto cruzamiento (ACC) hidro_mass con valores atípicos.
1.25
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12


nematoda no_efectores dataset 1 Covarianza de auto cruzamiento (ACC) hidro_mass con valores atípicos.
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12

6.1 Covarianza de auto cruzamiento (ACC) hidro_mass, sin valores atípicos

[12]:
```
#hidro_mass
transf = "Covarianza de auto cruzamiento (ACC) "
transf2 = "ACC"
estado = "sin valores atípicos.\n"
comp = "hidro_mass"
df=""

out = (str(r3) + '/ds' + str(dataset) + '_' + str(transf2) + '_' + str(comp) +␣
↪'_' + str(organismo) + '.csv')
os.makedirs(str(r3), exist_ok=True)
df_out = pd.DataFrame()

for etiq in "efectores", "no_efectores":
    titulo = (str(transf)+" "+ str(comp)+" "+ str(etiq) + " "+ str(nombre2) +",␣
↪" + str(estado))
    print (str(etiq))

    if etiq == "efectores":
        df=ACC_hidro_mass_efec

    if etiq == "no_efectores":
        df=ACC_hidro_mass_no_efec

    del df['X13']
    #Se eliminan todas las filas que tengan valores atípicos en al menos una de␣
↪sus columnas.
    df = (df[(np.abs(stats.zscore(df)) < 3).all(axis=1)])
    df['X13'] = etiq
    df_out = pd.concat([df_out,df])

    #Guarda la lista csv sin valores atípicos.
    df_out.to_csv(str(out), index=False, header=False)

    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")

    #Gráfica de caja y bigotes
    sns.set(style="whitegrid")
    fig , ax = plt.subplots(figsize=(14,7))
    ax = sns.boxplot(data=df)
    ax.set_title(organismo +' '+str(etiq)+" dataset "+str(dataset)+"␣
↪"+str(transf)+" "+str(comp))
```

efectores
Covarianza de auto cruzamiento (ACC) hidro_mass efectores nematoda dataset 1,
sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
1	-0.011797	-0.009725	0.084725	0.136624	-0.033562	0.097687	-0.038754	
2	-0.012463	-0.014299	-0.011442	0.056073	-0.058643	0.068880	-0.024983	
3	0.056212	0.029135	-0.017337	0.000754	0.019356	0.031176	0.050343	
4	0.092383	-0.114493	-0.021033	-0.161981	-0.195468	-0.115002	-0.112935	
5	0.022282	0.061606	0.019112	0.042486	-0.015739	0.028194	0.079772	
..	…	…	…	…	…	…	…	
494	0.027374	0.050105	-0.029213	-0.006676	0.009277	0.030228	0.018824	
495	-0.021151	-0.012916	0.025724	-0.004460	0.019319	-0.000540	0.015006	
496	-0.022709	0.126096	0.111002	0.118960	0.057710	-0.056092	0.152603	
497	-0.008882	-0.014505	0.001263	0.008876	-0.024333	-0.009379	0.035413	
499	0.061341	0.084701	-0.003269	0.034905	0.039719	0.073065	-0.055073	

	X7	X8	X9	X10	X11	X12	X13
1	-0.012509	0.039248	0.015632	0.019566	-0.005492	-0.020774	efectores
2	-0.008127	0.038218	0.010554	-0.017068	0.005978	0.031625	efectores
3	0.059253	0.014641	-0.072541	-0.030492	-0.013226	-0.050396	efectores
4	0.058062	0.126407	-0.032723	0.038422	0.066819	-0.158610	efectores
5	0.055807	0.037695	0.018709	0.066338	-0.068009	-0.021688	efectores
..	…	…	…	…	…	…	
494	-0.069008	0.014486	-0.010295	0.002849	0.068722	-0.030837	efectores
495	0.026149	0.004811	0.027502	-0.018389	-0.031828	0.004571	efectores
496	-0.041177	0.097274	0.015525	0.071043	0.035047	-0.032244	efectores
497	-0.043695	-0.000724	-0.001322	-0.041151	0.002160	0.010295	efectores
499	-0.071602	-0.034598	-0.072024	-0.134974	-0.048590	-0.072310	efectores

[458 rows x 14 columns]

Covarianza de auto cruzamiento (ACC) hidro_mass efectores nematoda dataset 1,
sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	458.000000	458.000000	458.000000	458.000000	458.000000	458.000000	
mean	0.010067	0.006711	0.008386	0.008135	0.003197	0.004578	
std	0.065678	0.060617	0.058933	0.061605	0.062410	0.060957	
min	-0.218895	-0.179127	-0.196178	-0.185760	-0.195468	-0.207338	
25%	-0.026445	-0.030565	-0.026839	-0.028289	-0.032877	-0.035125	
50%	0.016870	0.006771	0.007413	0.004081	0.004899	0.002775	
75%	0.048183	0.045153	0.037283	0.043785	0.037923	0.041617	
max	0.226307	0.176356	0.259005	0.191580	0.211224	0.188744	

```
                X6          X7          X8          X9          X10          X11  \
count  458.000000  458.000000  458.000000  458.000000  458.000000  458.000000
mean     0.010932    0.002111    0.005276    0.006205    0.005595   -0.001492
std      0.056801    0.065657    0.060532    0.061029    0.057219    0.063171
min     -0.187746   -0.217181   -0.230922   -0.202308   -0.151435   -0.220121
25%     -0.023659   -0.034876   -0.031396   -0.031605   -0.031191   -0.038805
50%      0.012034    0.003929    0.006811    0.004152    0.003064    0.002223
75%      0.045461    0.040618    0.039155    0.041496    0.042838    0.037131
max      0.214592    0.195436    0.240317    0.210826    0.183193    0.220926

               X12
count  458.000000
mean     0.002063
std      0.059410
min     -0.187481
25%     -0.030222
50%      0.002629
75%      0.036065
max      0.197275
```

no_efectores
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores nematoda dataset 1, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.115441  0.038070  0.030851  0.022599  0.046505  0.028404 -0.038752
1    0.063993  0.038335  0.044468  0.043469  0.023921  0.009412  0.016765
2    0.026215 -0.154273  0.074887 -0.043504 -0.049558 -0.028002 -0.094115
3    0.050824 -0.011756  0.012374  0.055829 -0.129675 -0.021034  0.083787
4   -0.064920 -0.070168 -0.096588 -0.003268  0.022062  0.073161 -0.043712
..        …         …         …         …         …         …
495 -0.041672 -0.046088  0.203464  0.015147 -0.035698  0.031145 -0.035094
496 -0.022277 -0.008190  0.092112  0.049676 -0.049231  0.061640  0.075976
497  0.000781  0.035087  0.025331  0.017103  0.080298 -0.044145 -0.012913
498  0.015630  0.036722  0.011245  0.044062  0.001443  0.022589 -0.034586
499  0.058914 -0.034206  0.018069  0.077425 -0.011081  0.038940  0.049779

           X7        X8        X9       X10       X11       X12           X13
0    0.028387  0.010034  0.042835  0.029551  0.013673 -0.028602  no_efectores
1    0.038850  0.006178  0.042438  0.010026 -0.001328  0.001493  no_efectores
2   -0.143690 -0.046976  0.031613  0.139037  0.074984 -0.095019  no_efectores
3   -0.167587 -0.009602  0.134049 -0.051341 -0.196695  0.000594  no_efectores
4   -0.007955  0.092660  0.080603 -0.076962  0.052785 -0.070343  no_efectores
..        …         …         …         …         …         …
495 -0.022387 -0.075223  0.121284  0.130548 -0.089992 -0.073631  no_efectores
```

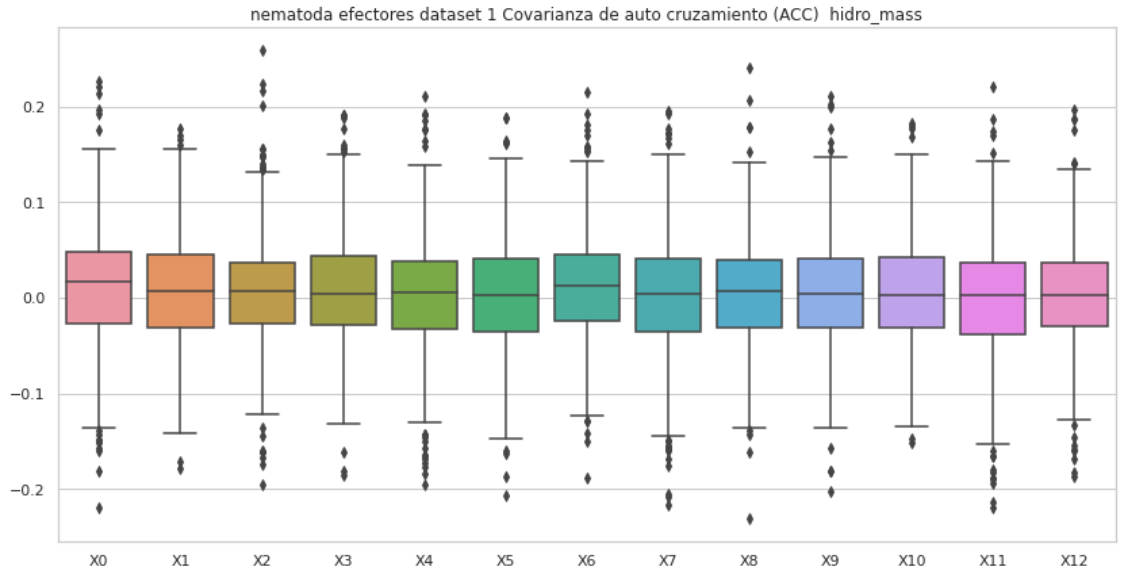
```
496 -0.032396 -0.032674 -0.025782  0.004230 -0.093376 -0.002864  no_efectores
497 -0.026529 -0.005030 -0.024967 -0.042186 -0.066716  0.030267  no_efectores
498  0.035446 -0.020920 -0.002452  0.011123  0.039312  0.050999  no_efectores
499  0.001646  0.034517  0.023128 -0.001730 -0.045006  0.045527  no_efectores

[466 rows x 14 columns]
```

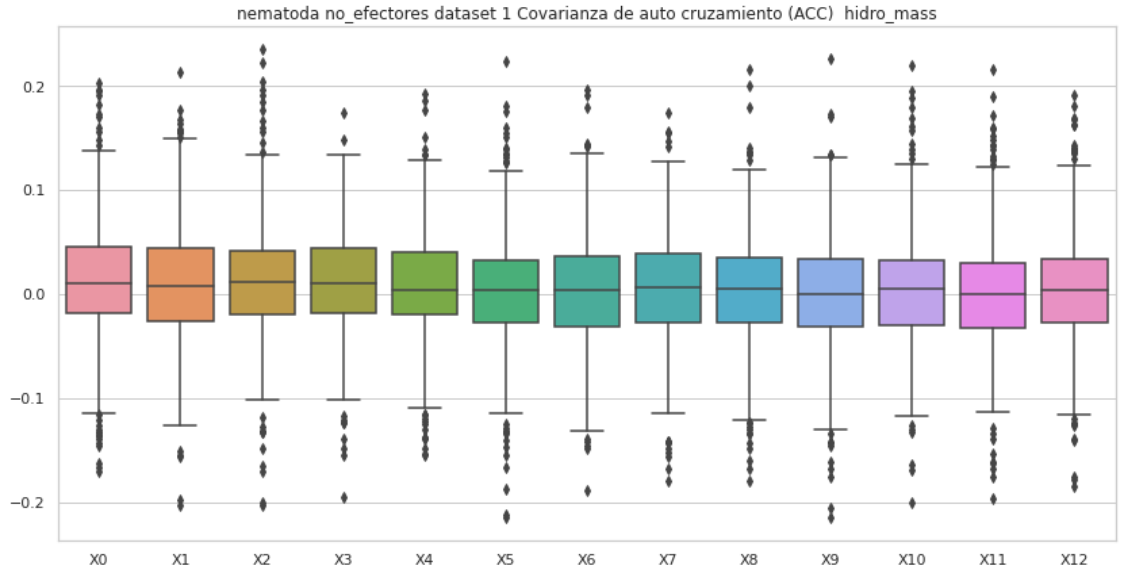
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores nematoda dataset 1, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	466.000000	466.000000	466.000000	466.000000	466.000000	466.000000	
mean	0.012972	0.008753	0.012063	0.011904	0.009222	0.002866	
std	0.059027	0.058148	0.058446	0.050343	0.052219	0.058002	
min	-0.171005	-0.203026	-0.203242	-0.194721	-0.155227	-0.214488	
25%	-0.018549	-0.026026	-0.019863	-0.018042	-0.019905	-0.027906	
50%	0.010216	0.007689	0.011814	0.010437	0.004064	0.003924	
75%	0.045434	0.044182	0.041977	0.044092	0.039939	0.032886	
max	0.202255	0.213158	0.235150	0.174097	0.192236	0.223670	

	X6	X7	X8	X9	X10	X11	\
count	466.000000	466.000000	466.000000	466.000000	466.000000	466.000000	
mean	0.003664	0.005219	0.002407	-0.000257	0.003527	0.000056	
std	0.053808	0.053936	0.055505	0.055742	0.056915	0.056720	
min	-0.189048	-0.179247	-0.179117	-0.214619	-0.200023	-0.196695	
25%	-0.030670	-0.026971	-0.028010	-0.031262	-0.030477	-0.033054	
50%	0.003275	0.006142	0.005379	-0.000230	0.005484	0.000546	
75%	0.036545	0.038433	0.034668	0.034322	0.032088	0.030254	
max	0.196240	0.174620	0.215787	0.226556	0.219540	0.215220	

	X12
count	466.000000
mean	0.003759
std	0.055163
min	-0.184259
25%	-0.027775
50%	0.003341
75%	0.033318
max	0.190759





7 Covarianza de auto cruzamiento (ACC) mass

[13]:
```
#mass
transf = "Covarianza de auto cruzamiento (ACC) "
transf2 = "ACC"
estado = "con valores atípicos.\n"
comp = "mass"
df=""
```

```
for etiq in "efectores", "no_efectores":
    titulo = (str(transf)+" "+ str(comp)+" "+ str(etiq) + " "+ str(nombre2) +",␣
 ↪" + str(estado))
    print (str(etiq))

    if etiq == "efectores":
        df=ACC_mass_efec

    if etiq == "no_efectores":
        df=ACC_mass_no_efec

    #del df['X13']
    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")

    #Gráfica de caja y bigotes
    sns.set(style="whitegrid")
    fig , ax = plt.subplots(figsize=(14,7))
    ax = sns.boxplot(data=df)
    ax.set_title(organismo +' '+str(etiq)+" dataset "+str(dataset)+"␣
 ↪"+str(transf)+" "+str(comp)+" "+str(estado))
```

```
efectores
Covarianza de auto cruzamiento (ACC)  mass efectores nematoda dataset 1, con
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.025211 -0.078576  0.435216 -0.017247 -0.049976  0.414317 -0.026065
1   -0.011797 -0.009725  0.084725  0.136624 -0.033562  0.097687 -0.038754
2   -0.012463 -0.014299 -0.011442  0.056073 -0.058643  0.068880 -0.024983
3    0.056212  0.029135 -0.017337  0.000754  0.019356  0.031176  0.050343
4    0.092383 -0.114493 -0.021033 -0.161981 -0.195468 -0.115002 -0.112935
..        …         …         …         …         …         …         …
495 -0.021151 -0.012916  0.025724 -0.004460  0.019319 -0.000540  0.015006
496 -0.022709  0.126096  0.111002  0.118960  0.057710 -0.056092  0.152603
497 -0.008882 -0.014505  0.001263  0.008876 -0.024333 -0.009379  0.035413
498  0.000425 -0.087246 -0.247833  0.080331  0.167989  0.061285 -0.160020
499  0.061341  0.084701 -0.003269  0.034905  0.039719  0.073065 -0.055073

           X7        X8        X9       X10       X11       X12        X13
0   -0.091376  0.311555 -0.014858  0.061172  0.272843 -0.128816  efectores
1   -0.012509  0.039248  0.015632  0.019566 -0.005492 -0.020774  efectores
```

```
2   -0.008127  0.038218  0.010554 -0.017068  0.005978  0.031625  efectores
3    0.059253  0.014641 -0.072541 -0.030492 -0.013226 -0.050396  efectores
4    0.058062  0.126407 -0.032723  0.038422  0.066819 -0.158610  efectores
..        …         …         …         …         …         …
495  0.026149  0.004811  0.027502 -0.018389 -0.031828  0.004571  efectores
496 -0.041177  0.097274  0.015525  0.071043  0.035047 -0.032244  efectores
497 -0.043695 -0.000724 -0.001322 -0.041151  0.002160  0.010295  efectores
498 -0.115440 -0.021992  0.249278  0.116749 -0.139009 -0.070152  efectores
499 -0.071602 -0.034598 -0.072024 -0.134974 -0.048590 -0.072310  efectores

[500 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) mass efectores nematoda dataset 1, con valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.008293    0.006226    0.012406    0.009666    0.005315    0.012384
std      0.082713    0.071897    0.088742    0.071130    0.072178    0.083455
min     -0.796466   -0.280452   -0.247833   -0.481303   -0.406789   -0.223372
25%     -0.027233   -0.035431   -0.028245   -0.029114   -0.033098   -0.032898
50%      0.016870    0.006580    0.007808    0.006134    0.006884    0.005189
75%      0.048568    0.046564    0.039495    0.046326    0.045247    0.047751
max      0.433674    0.291440    1.163856    0.285492    0.277449    0.985809

                X6          X7          X8          X9         X10         X11  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.009583    0.002260    0.006413    0.008848    0.007602    0.000451
std      0.068762    0.075231    0.080024    0.076167    0.066835    0.074882
min     -0.363852   -0.253643   -0.298026   -0.368876   -0.259720   -0.326409
25%     -0.025604   -0.038445   -0.034145   -0.031644   -0.031368   -0.041401
50%      0.010212    0.003801    0.006692    0.005511    0.003951    0.002629
75%      0.045489    0.042069    0.041852    0.045304    0.043775    0.038897
max      0.411898    0.380252    0.720522    0.535666    0.298103    0.421571

               X12
count  500.000000
mean     0.002565
std      0.067195
min     -0.264768
25%     -0.031034
50%      0.003773
75%      0.039637
max      0.277625
```

no_efectores
Covarianza de auto cruzamiento (ACC) mass no_efectores nematoda dataset 1, con valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.115441	0.038070	0.030851	0.022599	0.046505	0.028404	-0.038752	
1	0.063993	0.038335	0.044468	0.043469	0.023921	0.009412	0.016765	
2	0.026215	-0.154273	0.074887	-0.043504	-0.049558	-0.028002	-0.094115	
3	0.050824	-0.011756	0.012374	0.055829	-0.129675	-0.021034	0.083787	
4	-0.064920	-0.070168	-0.096588	-0.003268	0.022062	0.073161	-0.043712	
..	…	…	…	…	…	…	…	
495	-0.041672	-0.046088	0.203464	0.015147	-0.035698	0.031145	-0.035094	
496	-0.022277	-0.008190	0.092112	0.049676	-0.049231	0.061640	0.075976	
497	0.000781	0.035087	0.025331	0.017103	0.080298	-0.044145	-0.012913	
498	0.015630	0.036722	0.011245	0.044062	0.001443	0.022589	-0.034586	
499	0.058914	-0.034206	0.018069	0.077425	-0.011081	0.038940	0.049779	

	X7	X8	X9	X10	X11	X12	X13
0	0.028387	0.010034	0.042835	0.029551	0.013673	-0.028602	no_efectores
1	0.038850	0.006178	0.042438	0.010026	-0.001328	0.001493	no_efectores
2	-0.143690	-0.046976	0.031613	0.139037	0.074984	-0.095019	no_efectores
3	-0.167587	-0.009602	0.134049	-0.051341	-0.196695	0.000594	no_efectores
4	-0.007955	0.092660	0.080603	-0.076962	0.052785	-0.070343	no_efectores
..	…	…	…	…	…	…	…
495	-0.022387	-0.075223	0.121284	0.130548	-0.089992	-0.073631	no_efectores
496	-0.032396	-0.032674	-0.025782	0.004230	-0.093376	-0.002864	no_efectores
497	-0.026529	-0.005030	-0.024967	-0.042186	-0.066716	0.030267	no_efectores
498	0.035446	-0.020920	-0.002452	0.011123	0.039312	0.050999	no_efectores
499	0.001646	0.034517	0.023128	-0.001730	-0.045006	0.045527	no_efectores

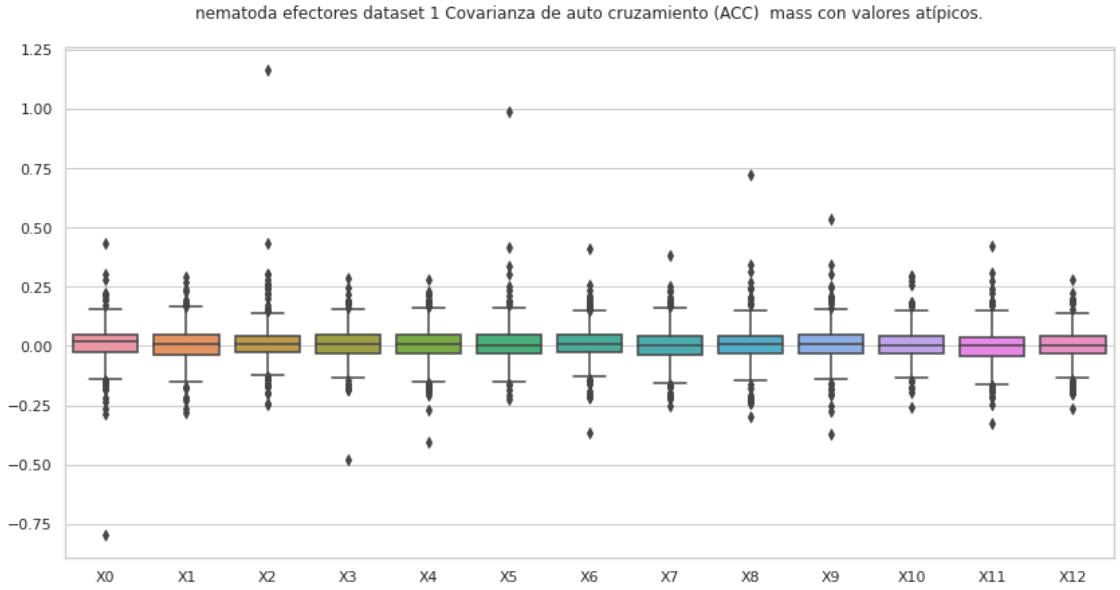
[500 rows x 14 columns]

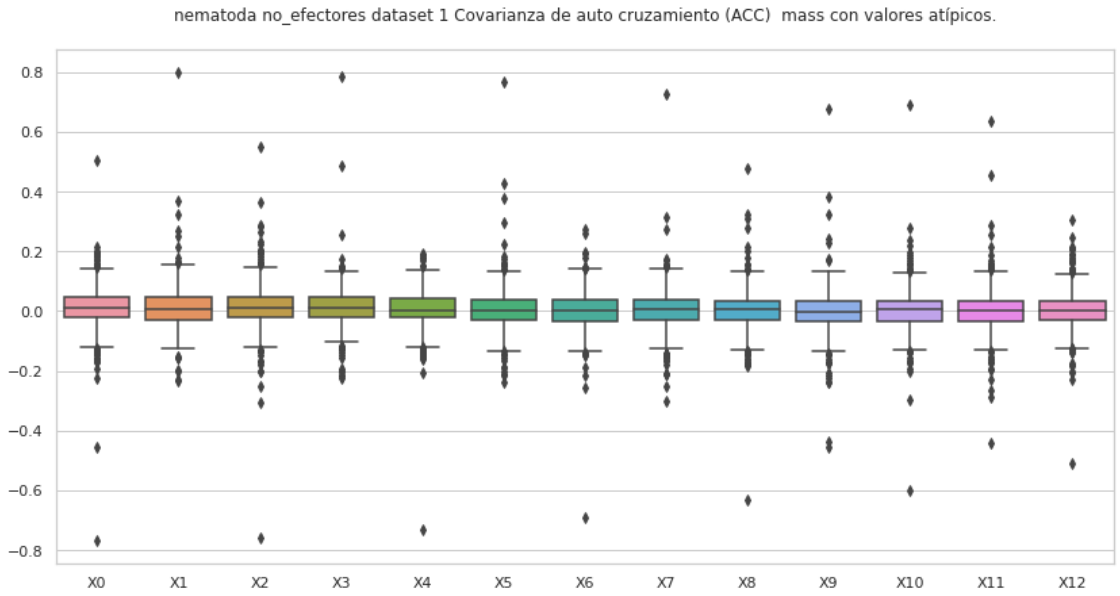
Covarianza de auto cruzamiento (ACC) mass no_efectores nematoda dataset 1, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.009956	0.010884	0.012652	0.013821	0.007417	0.005228	
std	0.079540	0.075256	0.079632	0.070260	0.065515	0.076149	
min	-0.767623	-0.234745	-0.758319	-0.225384	-0.733605	-0.237639	
25%	-0.021972	-0.029890	-0.022490	-0.018631	-0.022453	-0.031708	
50%	0.009503	0.007426	0.012026	0.011271	0.003909	0.004265	
75%	0.045717	0.045323	0.045454	0.046310	0.042872	0.036328	
max	0.505626	0.798790	0.548220	0.784768	0.192236	0.769448	

	X6	X7	X8	X9	X10	X11	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.001699	0.005524	0.002175	-0.000410	0.002443	0.000518	
std	0.067517	0.070978	0.071806	0.075937	0.075702	0.075866	
min	-0.690465	-0.303143	-0.632800	-0.456386	-0.601473	-0.439466	
25%	-0.032499	-0.027869	-0.031678	-0.032244	-0.033071	-0.033650	
50%	0.002496	0.006063	0.005012	-0.000516	0.004649	0.000584	
75%	0.037465	0.040050	0.034725	0.035594	0.032693	0.033282	
max	0.275249	0.726801	0.475577	0.676298	0.688794	0.634251	

	X12
count	500.000000
mean	0.004057
std	0.067576
min	-0.511735
25%	-0.028454
50%	0.003867
75%	0.034910
max	0.303997





7.1 Covarianza de auto cruzamiento (ACC) mass, sin valores atípicos

[14]:
```
#mass
transf = "Covarianza de auto cruzamiento (ACC) "
transf2 = "ACC"
estado = "sin valores atípicos.\n"
comp = "mass"
df=""

#Se eliminan todas las filas que tengan valores atípicos en al menos una de sus␣
 ↪columnas.
out = (str(r3) + '/ds' + str(dataset) + '_' + str(transf2) + '_' + str(comp) +␣
 ↪'_' + str(organismo) + '.csv')
os.makedirs(str(r3), exist_ok=True)
df=""
df_out = pd.DataFrame()

for etiq in "efectores", "no_efectores":
    titulo = (str(transf)+" "+ str(comp)+" "+ str(etiq) + " "+ str(nombre2) +",␣
 ↪" + str(estado))

    if etiq == "efectores":
        df=ACC_mass_efec

    if etiq == "no_efectores":
        df=ACC_mass_no_efec
```

```
    del df['X13']
    #Se eliminan todas las filas que tengan valores atípicos en al menos una de␣
↪sus columnas.
    df = (df[(np.abs(stats.zscore(df)) < 3).all(axis=1)])
    df['X13'] = etiq
    df_out = pd.concat([df_out,df])

    #Guarda la lista csv sin valores atípicos.
    df_out.to_csv(str(out), index=False, header=False)

    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")

    #Gráfica de caja y bigotes
    sns.set(style="whitegrid")
    fig , ax = plt.subplots(figsize=(14,7))
    ax = sns.boxplot(data=df)
    ax.set_title(organismo +' '+str(etiq)+" dataset "+str(dataset)+"␣
↪"+str(transf)+" "+str(comp))
```

```
Covarianza de auto cruzamiento (ACC)  mass efectores nematoda dataset 1, sin
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
1   -0.011797 -0.009725  0.084725  0.136624 -0.033562  0.097687 -0.038754
2   -0.012463 -0.014299 -0.011442  0.056073 -0.058643  0.068880 -0.024983
3    0.056212  0.029135 -0.017337  0.000754  0.019356  0.031176  0.050343
4    0.092383 -0.114493 -0.021033 -0.161981 -0.195468 -0.115002 -0.112935
5    0.022282  0.061606  0.019112  0.042486 -0.015739  0.028194  0.079772
..        …         …         …         …         …         …         …
494  0.027374  0.050105 -0.029213 -0.006676  0.009277  0.030228  0.018824
495 -0.021151 -0.012916  0.025724 -0.004460  0.019319 -0.000540  0.015006
496 -0.022709  0.126096  0.111002  0.118960  0.057710 -0.056092  0.152603
497 -0.008882 -0.014505  0.001263  0.008876 -0.024333 -0.009379  0.035413
499  0.061341  0.084701 -0.003269  0.034905  0.039719  0.073065 -0.055073

           X7        X8        X9       X10       X11       X12        X13
1   -0.012509  0.039248  0.015632  0.019566 -0.005492 -0.020774  efectores
2   -0.008127  0.038218  0.010554 -0.017068  0.005978  0.031625  efectores
3    0.059253  0.014641 -0.072541 -0.030492 -0.013226 -0.050396  efectores
4    0.058062  0.126407 -0.032723  0.038422  0.066819 -0.158610  efectores
5    0.055807  0.037695  0.018709  0.066338 -0.068009 -0.021688  efectores
..        …         …         …         …         …         …         …
```

```
494 -0.069008  0.014486 -0.010295  0.002849  0.068722 -0.030837  efectores
495  0.026149  0.004811  0.027502 -0.018389 -0.031828  0.004571  efectores
496 -0.041177  0.097274  0.015525  0.071043  0.035047 -0.032244  efectores
497 -0.043695 -0.000724 -0.001322 -0.041151  0.002160  0.010295  efectores
499 -0.071602 -0.034598 -0.072024 -0.134974 -0.048590 -0.072310  efectores

[458 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) mass efectores nematoda dataset 1, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	458.000000	458.000000	458.000000	458.000000	458.000000	458.000000	
mean	0.010067	0.006711	0.008386	0.008135	0.003197	0.004578	
std	0.065678	0.060617	0.058933	0.061605	0.062410	0.060957	
min	-0.218895	-0.179127	-0.196178	-0.185760	-0.195468	-0.207338	
25%	-0.026445	-0.030565	-0.026839	-0.028289	-0.032877	-0.035125	
50%	0.016870	0.006771	0.007413	0.004081	0.004899	0.002775	
75%	0.048183	0.045153	0.037283	0.043785	0.037923	0.041617	
max	0.226307	0.176356	0.259005	0.191580	0.211224	0.188744	

	X6	X7	X8	X9	X10	X11	\
count	458.000000	458.000000	458.000000	458.000000	458.000000	458.000000	
mean	0.010932	0.002111	0.005276	0.006205	0.005595	-0.001492	
std	0.056801	0.065657	0.060532	0.061029	0.057219	0.063171	
min	-0.187746	-0.217181	-0.230922	-0.202308	-0.151435	-0.220121	
25%	-0.023659	-0.034876	-0.031396	-0.031605	-0.031191	-0.038805	
50%	0.012034	0.003929	0.006811	0.004152	0.003064	0.002223	
75%	0.045461	0.040618	0.039155	0.041496	0.042838	0.037131	
max	0.214592	0.195436	0.240317	0.210826	0.183193	0.220926	

	X12
count	458.000000
mean	0.002063
std	0.059410
min	-0.187481
25%	-0.030222
50%	0.002629
75%	0.036065
max	0.197275

Covarianza de auto cruzamiento (ACC) mass no_efectores nematoda dataset 1, sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.115441	0.038070	0.030851	0.022599	0.046505	0.028404	-0.038752	
1	0.063993	0.038335	0.044468	0.043469	0.023921	0.009412	0.016765	
2	0.026215	-0.154273	0.074887	-0.043504	-0.049558	-0.028002	-0.094115	
3	0.050824	-0.011756	0.012374	0.055829	-0.129675	-0.021034	0.083787	
4	-0.064920	-0.070168	-0.096588	-0.003268	0.022062	0.073161	-0.043712	
..	…	…	…	…	…	…	…	
495	-0.041672	-0.046088	0.203464	0.015147	-0.035698	0.031145	-0.035094	
496	-0.022277	-0.008190	0.092112	0.049676	-0.049231	0.061640	0.075976	
497	0.000781	0.035087	0.025331	0.017103	0.080298	-0.044145	-0.012913	
498	0.015630	0.036722	0.011245	0.044062	0.001443	0.022589	-0.034586	
499	0.058914	-0.034206	0.018069	0.077425	-0.011081	0.038940	0.049779	

	X7	X8	X9	X10	X11	X12	X13
0	0.028387	0.010034	0.042835	0.029551	0.013673	-0.028602	no_efectores
1	0.038850	0.006178	0.042438	0.010026	-0.001328	0.001493	no_efectores
2	-0.143690	-0.046976	0.031613	0.139037	0.074984	-0.095019	no_efectores
3	-0.167587	-0.009602	0.134049	-0.051341	-0.196695	0.000594	no_efectores
4	-0.007955	0.092660	0.080603	-0.076962	0.052785	-0.070343	no_efectores
..	…	…	…	…	…	…	…
495	-0.022387	-0.075223	0.121284	0.130548	-0.089992	-0.073631	no_efectores
496	-0.032396	-0.032674	-0.025782	0.004230	-0.093376	-0.002864	no_efectores
497	-0.026529	-0.005030	-0.024967	-0.042186	-0.066716	0.030267	no_efectores
498	0.035446	-0.020920	-0.002452	0.011123	0.039312	0.050999	no_efectores
499	0.001646	0.034517	0.023128	-0.001730	-0.045006	0.045527	no_efectores

[466 rows x 14 columns]

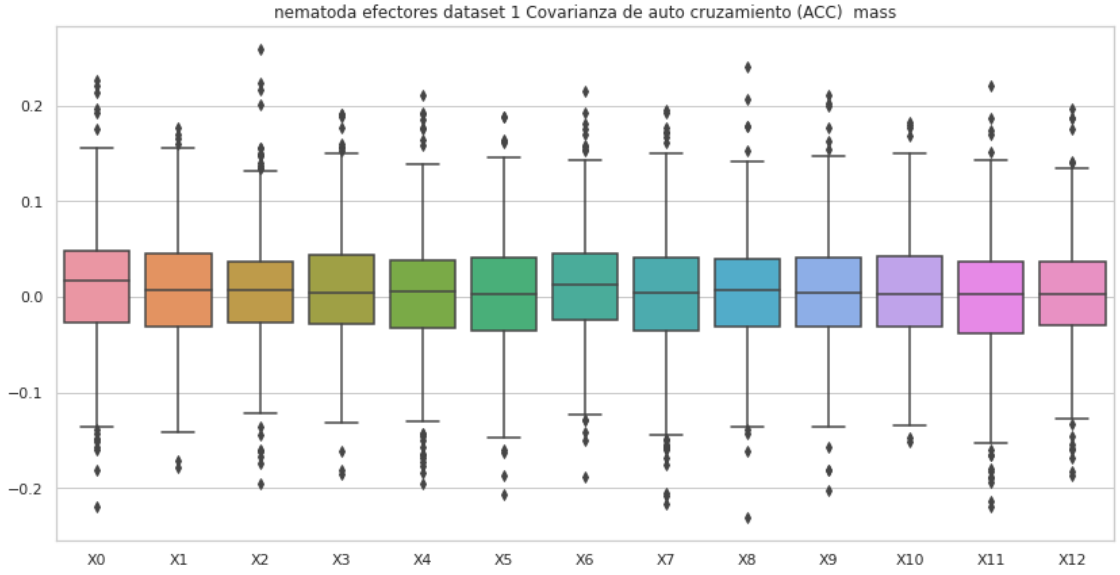
Covarianza de auto cruzamiento (ACC) mass no_efectores nematoda dataset 1, sin valores atípicos.
Estadísticas.

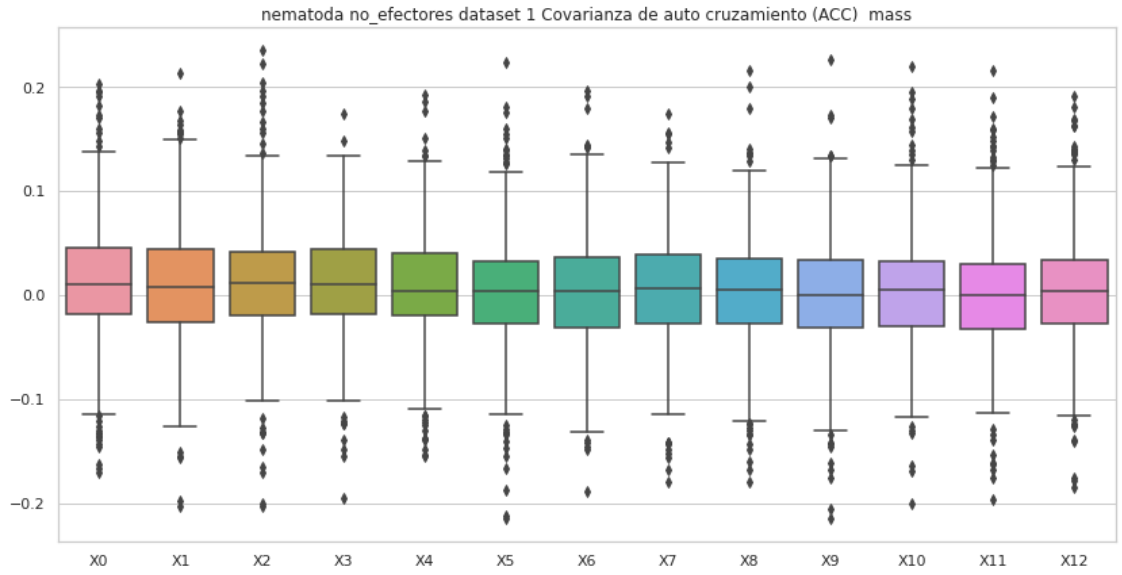
	X0	X1	X2	X3	X4	X5	\
count	466.000000	466.000000	466.000000	466.000000	466.000000	466.000000	
mean	0.012972	0.008753	0.012063	0.011904	0.009222	0.002866	
std	0.059027	0.058148	0.058446	0.050343	0.052219	0.058002	
min	-0.171005	-0.203026	-0.203242	-0.194721	-0.155227	-0.214488	
25%	-0.018549	-0.026026	-0.019863	-0.018042	-0.019905	-0.027906	
50%	0.010216	0.007689	0.011814	0.010437	0.004064	0.003924	
75%	0.045434	0.044182	0.041977	0.044092	0.039939	0.032886	
max	0.202255	0.213158	0.235150	0.174097	0.192236	0.223670	

	X6	X7	X8	X9	X10	X11	\
count	466.000000	466.000000	466.000000	466.000000	466.000000	466.000000	
mean	0.003664	0.005219	0.002407	-0.000257	0.003527	0.000056	
std	0.053808	0.053936	0.055505	0.055742	0.056915	0.056720	

min	-0.189048	-0.179247	-0.179117	-0.214619	-0.200023	-0.196695
25%	-0.030670	-0.026971	-0.028010	-0.031262	-0.030477	-0.033054
50%	0.003275	0.006142	0.005379	-0.000230	0.005484	0.000546
75%	0.036545	0.038433	0.034668	0.034322	0.032088	0.030254
max	0.196240	0.174620	0.215787	0.226556	0.219540	0.215220

	X12
count	466.000000
mean	0.003759
std	0.055163
min	-0.184259
25%	-0.027775
50%	0.003341
75%	0.033318
max	0.190759





8 Covarianza de auto cruzamiento (ACC) hidro

[15]:
```
#hidro
transf = "Covarianza de auto cruzamiento (ACC) "
transf2 = "ACC"
estado = "con valores atípicos.\n"
comp = "hidro"
df=""

for etiq in "efectores", "no_efectores":
    titulo = (str(transf)+" "+ str(comp)+" "+ str(etiq) + " "+ str(nombre2) +",␣
 ↪" + str(estado))
    print (str(etiq))

    if etiq == "efectores":
        df=ACC_hidro_efec

    if etiq == "no_efectores":
        df=ACC_hidro_no_efec

    #del df['X13']
    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")
```

```
    #Gráfica de caja y bigotes
    sns.set(style="whitegrid")
    fig , ax = plt.subplots(figsize=(14,7))
    ax = sns.boxplot(data=df)
    ax.set_title(organismo +' '+str(etiq)+" dataset "+str(dataset)+"␣
↪"+str(transf)+" "+str(comp)+" "+str(estado))
```

```
efectores
Covarianza de auto cruzamiento (ACC)  hidro efectores nematoda dataset 1, con
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.039602 -0.006943  0.010522 -0.006757  0.006104  0.073556 -0.038334
1    0.097353  0.083211  0.040260 -0.000851  0.083741  0.056525  0.058509
2   -0.021780 -0.092868  0.106863  0.036129 -0.108825  0.055048  0.036694
3   -0.057560 -0.046682  0.035556 -0.108565  0.077530 -0.000855 -0.015301
4   -0.217609 -0.083398 -0.071533  0.103998 -0.079001 -0.082309 -0.089237
..        …       …       …       …       …       …       …
495  0.055506 -0.024264 -0.015286  0.036777 -0.009313 -0.058018  0.001881
496 -0.102627 -0.089111  0.021388 -0.149801  0.047335  0.119571 -0.053517
497  0.064801 -0.149521  0.036341  0.130045  0.016815  0.014176  0.014433
498 -0.100908 -0.235718  0.115631 -0.211028  0.051729  0.164477 -0.024101
499 -0.003238 -0.002208  0.168277 -0.009803 -0.043817 -0.035921  0.170966

           X7        X8        X9       X10       X11       X12        X13
0   -0.008109  0.034402  0.039644 -0.005990  0.035869  0.096251  efectores
1    0.013658  0.046935  0.019543 -0.000939  0.038776 -0.034636  efectores
2    0.028259 -0.063695 -0.030577  0.058093 -0.066462 -0.019728  efectores
3    0.092049  0.018906 -0.030218 -0.064440 -0.121905  0.134820  efectores
4    0.097177  0.050795 -0.040866  0.042670  0.191818 -0.067929  efectores
..        …       …       …       …       …       …        …
495 -0.016713  0.069589  0.045729 -0.007767  0.022425  0.043526  efectores
496  0.116782 -0.265777  0.104351  0.103688 -0.329488  0.281082  efectores
497 -0.044347  0.014839  0.064797  0.010776  0.051691 -0.028054  efectores
498 -0.100276  0.044623 -0.228611  0.049175  0.015905 -0.203940  efectores
499  0.006737 -0.077104 -0.113391  0.123615 -0.106208 -0.008380  efectores

[500 rows x 14 columns]


Covarianza de auto cruzamiento (ACC)  hidro efectores nematoda dataset 1, con
valores atípicos.
Estadísticas.

                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
```

mean	0.012385	-0.019580	0.023957	0.023430	-0.006758	-0.001024
std	0.088914	0.089830	0.080964	0.085076	0.082191	0.083075
min	-0.395330	-0.392887	-0.283854	-0.211869	-0.282300	-0.481026
25%	-0.042754	-0.080113	-0.021290	-0.029489	-0.053929	-0.048620
50%	0.010712	-0.021558	0.019415	0.020265	-0.005082	0.000006
75%	0.064000	0.034950	0.071969	0.074766	0.038098	0.043826
max	0.329436	0.279877	0.321327	0.344437	0.270897	0.306535

	X6	X7	X8	X9	X10	X11
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	0.019238	0.010646	0.005732	0.008684	0.009846	0.000783
std	0.085501	0.078117	0.081502	0.080673	0.085914	0.082002
min	-0.337496	-0.259197	-0.310661	-0.282351	-0.308475	-0.329488
25%	-0.028771	-0.033813	-0.042783	-0.033430	-0.040176	-0.039863
50%	0.021412	0.011640	0.007037	0.009890	0.010717	0.004402
75%	0.063800	0.059072	0.047372	0.054523	0.058534	0.045219
max	0.354071	0.259717	0.428387	0.337783	0.471826	0.308247

	X12
count	500.000000
mean	0.003989
std	0.086447
min	-0.373657
25%	-0.045248
50%	0.004207
75%	0.049205
max	0.412324

no_efectores
Covarianza de auto cruzamiento (ACC) hidro no_efectores nematoda dataset 1, con valores atípicos.
Valores del documento csv.

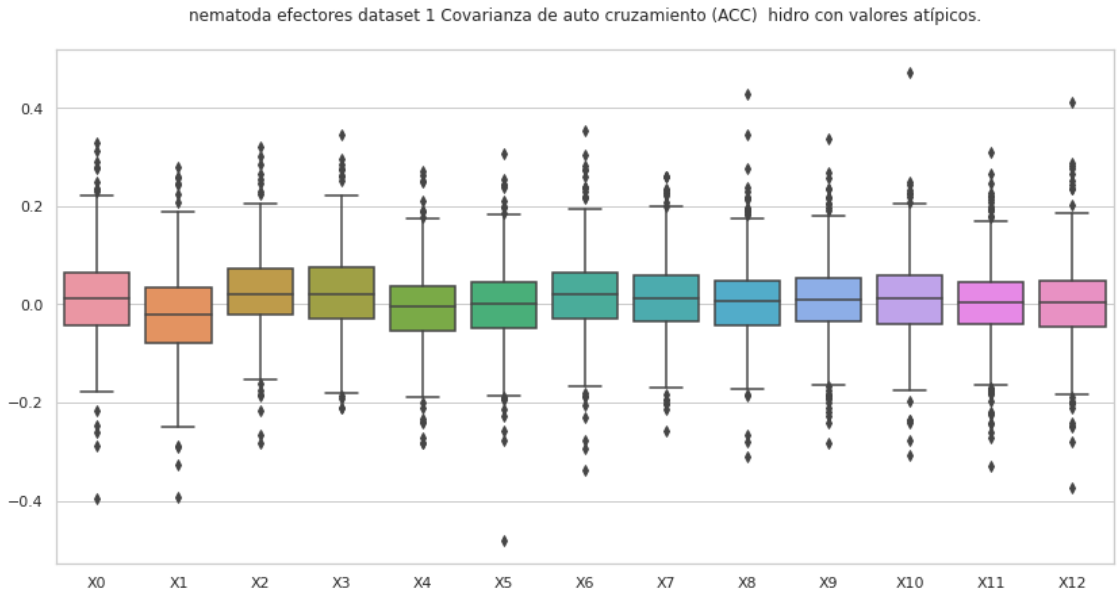
	X0	X1	X2	X3	X4	X5	X6
0	0.015655	-0.111021	-0.003149	-0.025243	-0.018548	0.020666	0.024036
1	-0.009899	-0.064337	0.178934	0.061817	-0.066707	0.072254	0.124070
2	0.081609	0.117402	-0.017617	-0.074565	-0.126035	-0.052885	0.002530
3	-0.046131	0.003154	0.028563	0.127672	-0.067983	0.179148	-0.047978
4	-0.016471	-0.059399	-0.045822	-0.005434	-0.038284	0.029427	-0.037409
..	…	…	…	…	…	…	…
495	0.113085	-0.047895	0.178723	-0.021863	-0.045848	0.043614	0.010294
496	0.135422	-0.016367	0.149417	0.166231	0.069276	0.048291	0.090520
497	0.011288	-0.040811	0.060764	0.005519	-0.111573	-0.047331	0.059826
498	-0.064340	0.001938	0.024195	-0.018060	0.083935	-0.048083	0.065347
499	0.073582	-0.014514	0.041324	0.066169	0.053110	0.020287	0.020976

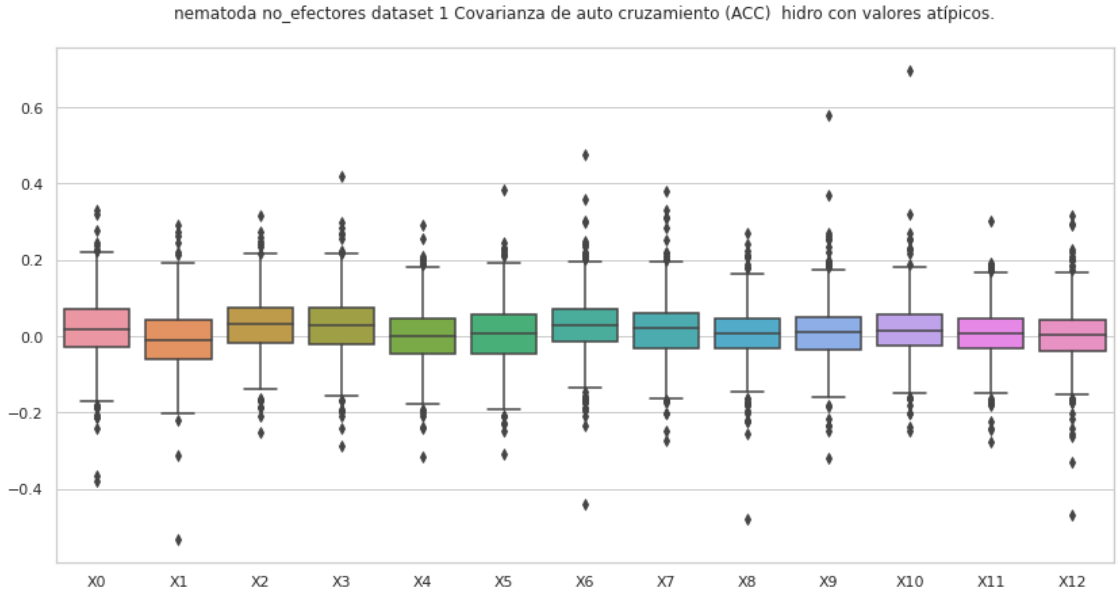
```
           X7        X8        X9       X10       X11       X12           X13
0    0.072053  0.054913  0.047996  0.038510  0.006609 -0.047500  no_efectores
1   -0.001212 -0.011319  0.092662  0.019381 -0.001725  0.040249  no_efectores
2    0.030542 -0.046670 -0.185778  0.006777 -0.100782 -0.043358  no_efectores
3   -0.000849  0.091311  0.151828 -0.235838  0.105138 -0.147720  no_efectores
4   -0.001363 -0.129475  0.000894  0.015796  0.058101 -0.057155  no_efectores
..        …         …         …         …         …         …             …
495 -0.201924 -0.223088  0.173884  0.016841 -0.012427 -0.049079  no_efectores
496  0.069875  0.074576  0.062442  0.038133  0.078990  0.097567  no_efectores
497  0.073510 -0.053768  0.051479  0.125655 -0.064709  0.044535  no_efectores
498  0.065987  0.022268 -0.030397 -0.041820  0.022487  0.004442  no_efectores
499  0.044911  0.114064 -0.005864  0.064465  0.024751  0.023931  no_efectores

[500 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro no_efectores nematoda dataset 1, con valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.020649   -0.009011    0.033474    0.028300    0.000542    0.006096
std      0.087462    0.085021    0.080306    0.082837    0.078333    0.082805
min     -0.378123   -0.532280   -0.252049   -0.289012   -0.315730   -0.309814
25%     -0.026817   -0.061185   -0.018050   -0.021519   -0.045840   -0.044648
50%      0.016685   -0.010427    0.032230    0.027879    0.001879    0.006037
75%      0.072851    0.043026    0.075901    0.073565    0.045986    0.056234
max      0.329943    0.292552    0.316472    0.419258    0.291893    0.383558

                X6          X7          X8          X9         X10         X11  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.028129    0.017857    0.006194    0.010733    0.019911    0.006515
std      0.082437    0.079347    0.077707    0.081795    0.079004    0.069751
min     -0.439602   -0.274360   -0.479478   -0.320457   -0.247640   -0.276923
25%     -0.014432   -0.031002   -0.032767   -0.034701   -0.025600   -0.032757
50%      0.027568    0.022073    0.007192    0.009804    0.016368    0.007895
75%      0.071158    0.059925    0.047405    0.049478    0.058534    0.047735
max      0.476140    0.380475    0.268556    0.578384    0.696132    0.301447

               X12
count  500.000000
mean     0.003946
std      0.080077
min     -0.469583
25%     -0.037488
50%      0.003076
75%      0.044399
```

```
max       0.314813
```


nematoda efectores dataset 1 Covarianza de auto cruzamiento (ACC) hidro con valores atípicos.
0.4
0.2
0.0
−0.2
−0.4
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12


nematoda no_efectores dataset 1 Covarianza de auto cruzamiento (ACC) hidro con valores atípicos.
0.6
0.4
0.2
0.0
−0.2
−0.4
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12

8.1 Covarianza de auto cruzamiento (ACC) hidro, sin valores atípicos

[16]:
```
#hidro
transf = "Covarianza de auto cruzamiento (ACC) "
transf2 = "ACC"
estado = "sin valores atípicos.\n"
comp = "hidro"
df=""

out = (str(r3) + '/ds' + str(dataset) + '_' + str(transf2) + '_' + str(comp) +␣
↪'_' + str(organismo) + '.csv')
os.makedirs(str(r3), exist_ok=True)
df_out = pd.DataFrame()

for etiq in "efectores", "no_efectores":
    titulo = (str(transf) +" "+ str(etiq) + " " + str(nombre2) + ", " +␣
↪str(estado))
    print (str(etiq))

    if etiq == "efectores":
        df=ACC_hidro_efec

    if etiq == "no_efectores":
        df=ACC_hidro_no_efec

    del df['X13']
    #Se eliminan todas las filas que tengan valores atípicos en al menos una de␣
↪sus columnas.
    df = (df[(np.abs(stats.zscore(df)) < 3).all(axis=1)])
    df['X13'] = etiq
    df_out = pd.concat([df_out,df])

    #Guarda la lista csv sin valores atípicos.
    df_out.to_csv(str(out), index=False, header=False)

    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")

    #Gráfica de caja y bigotes
    sns.set(style="whitegrid")
    fig , ax = plt.subplots(figsize=(14,7))
    ax = sns.boxplot(data=df)
    ax.set_title(organismo +' '+str(etiq)+" dataset "+str(dataset)+"␣
↪"+str(transf)+" "+str(comp))
```

efectores
Covarianza de auto cruzamiento (ACC) efectores nematoda dataset 1, sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.039602	-0.006943	0.010522	-0.006757	0.006104	0.073556	-0.038334	
1	0.097353	0.083211	0.040260	-0.000851	0.083741	0.056525	0.058509	
2	-0.021780	-0.092868	0.106863	0.036129	-0.108825	0.055048	0.036694	
3	-0.057560	-0.046682	0.035556	-0.108565	0.077530	-0.000855	-0.015301	
4	-0.217609	-0.083398	-0.071533	0.103998	-0.079001	-0.082309	-0.089237	
..	…	…	…	…	…	…	…	
494	-0.018124	-0.005536	-0.028179	-0.059781	0.008017	-0.081402	-0.002145	
495	0.055506	-0.024264	-0.015286	0.036777	-0.009313	-0.058018	0.001881	
497	0.064801	-0.149521	0.036341	0.130045	0.016815	0.014176	0.014433	
498	-0.100908	-0.235718	0.115631	-0.211028	0.051729	0.164477	-0.024101	
499	-0.003238	-0.002208	0.168277	-0.009803	-0.043817	-0.035921	0.170966	

	X7	X8	X9	X10	X11	X12	X13
0	-0.008109	0.034402	0.039644	-0.005990	0.035869	0.096251	efectores
1	0.013658	0.046935	0.019543	-0.000939	0.038776	-0.034636	efectores
2	0.028259	-0.063695	-0.030577	0.058093	-0.066462	-0.019728	efectores
3	0.092049	0.018906	-0.030218	-0.064440	-0.121905	0.134820	efectores
4	0.097177	0.050795	-0.040866	0.042670	0.191818	-0.067929	efectores
..	…	…	…	…	…	…	
494	-0.031475	0.035397	0.078941	0.004399	0.007375	0.034099	efectores
495	-0.016713	0.069589	0.045729	-0.007767	0.022425	0.043526	efectores
497	-0.044347	0.014839	0.064797	0.010776	0.051691	-0.028054	efectores
498	-0.100276	0.044623	-0.228611	0.049175	0.015905	-0.203940	efectores
499	0.006737	-0.077104	-0.113391	0.123615	-0.106208	-0.008380	efectores

[457 rows x 14 columns]

Covarianza de auto cruzamiento (ACC) efectores nematoda dataset 1, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	457.000000	457.000000	457.000000	457.000000	457.000000	457.000000	
mean	0.010473	-0.020093	0.020746	0.019627	-0.005537	-0.001392	
std	0.078178	0.078507	0.072778	0.077858	0.073005	0.073243	
min	-0.246250	-0.250944	-0.187078	-0.211869	-0.242059	-0.228936	
25%	-0.039900	-0.071286	-0.021413	-0.026438	-0.049563	-0.046264	
50%	0.009104	-0.020961	0.016850	0.018805	-0.003603	0.000074	
75%	0.060621	0.032688	0.067193	0.065569	0.037290	0.042276	
max	0.275825	0.245717	0.254100	0.276263	0.210041	0.243251	

```
                X6           X7           X8           X9          X10          X11  \
count   457.000000   457.000000   457.000000   457.000000   457.000000   457.000000
mean      0.016770     0.010220     0.005152     0.007085     0.007255     0.001472
std       0.072949     0.070504     0.071311     0.071649     0.076346     0.071360
min      -0.230313    -0.203453    -0.185265    -0.228611    -0.236070    -0.224529
25%      -0.028750    -0.031701    -0.041235    -0.032874    -0.039648    -0.038637
50%       0.020964     0.011887     0.007008     0.008470     0.008756     0.004545
75%       0.059268     0.055981     0.044501     0.050086     0.052675     0.041358
max       0.272343     0.234760     0.237950     0.236489     0.248610     0.226513

               X12
count   457.000000
mean      0.001447
std       0.075275
min      -0.248598
25%      -0.042692
50%       0.004022
75%       0.044510
max       0.250723
```

no_efectores
Covarianza de auto cruzamiento (ACC) no_efectores nematoda dataset 1, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.015655 -0.111021 -0.003149 -0.025243 -0.018548  0.020666  0.024036
1   -0.009899 -0.064337  0.178934  0.061817 -0.066707  0.072254  0.124070
2    0.081609  0.117402 -0.017617 -0.074565 -0.126035 -0.052885  0.002530
4   -0.016471 -0.059399 -0.045822 -0.005434 -0.038284  0.029427 -0.037409
5    0.001134 -0.068719  0.090238  0.023882  0.025835 -0.044588  0.040712
..        …         …         …         …         …         …         …
495  0.113085 -0.047895  0.178723 -0.021863 -0.045848  0.043614  0.010294
496  0.135422 -0.016367  0.149417  0.166231  0.069276  0.048291  0.090520
497  0.011288 -0.040811  0.060764  0.005519 -0.111573 -0.047331  0.059826
498 -0.064340  0.001938  0.024195 -0.018060  0.083935 -0.048083  0.065347
499  0.073582 -0.014514  0.041324  0.066169  0.053110  0.020287  0.020976

           X7        X8        X9       X10       X11       X12           X13
0    0.072053  0.054913  0.047996  0.038510  0.006609 -0.047500  no_efectores
1   -0.001212 -0.011319  0.092662  0.019381 -0.001725  0.040249  no_efectores
2    0.030542 -0.046670 -0.185778  0.006777 -0.100782 -0.043358  no_efectores
4   -0.001363 -0.129475  0.000894  0.015796  0.058101 -0.057155  no_efectores
5    0.087729  0.015197  0.002999  0.010216  0.019113 -0.003511  no_efectores
..        …         …         …         …         …         …           …
495 -0.201924 -0.223088  0.173884  0.016841 -0.012427 -0.049079  no_efectores
```

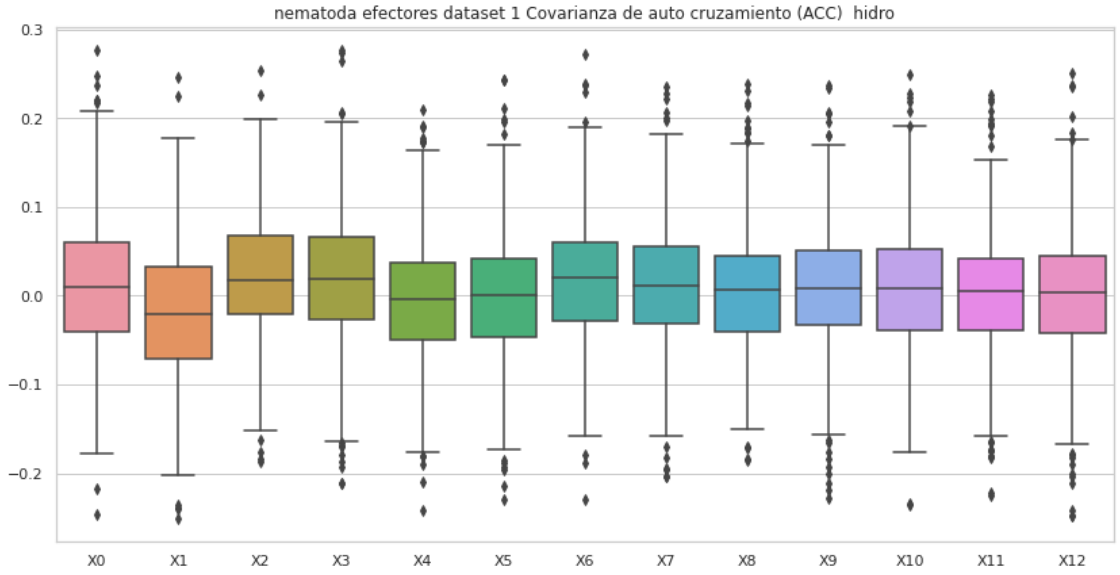
```
496  0.069875  0.074576  0.062442  0.038133  0.078990  0.097567  no_efectores
497  0.073510 -0.053768  0.051479  0.125655 -0.064709  0.044535  no_efectores
498  0.065987  0.022268 -0.030397 -0.041820  0.022487  0.004442  no_efectores
499  0.044911  0.114064 -0.005864  0.064465  0.024751  0.023931  no_efectores

[457 rows x 14 columns]
```

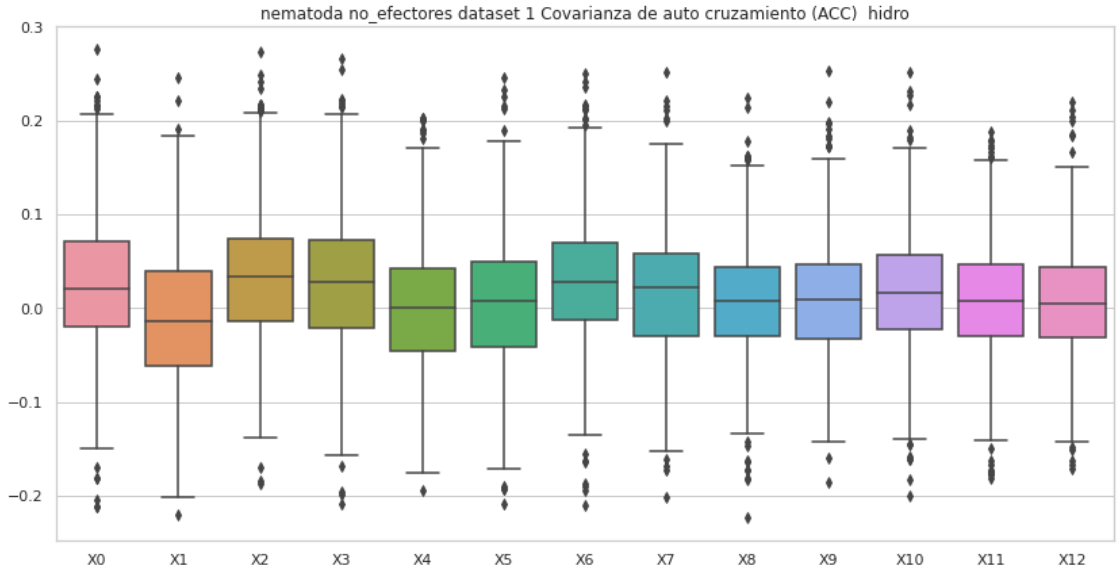
Covarianza de auto cruzamiento (ACC) no_efectores nematoda dataset 1, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	457.000000	457.000000	457.000000	457.000000	457.000000	457.000000	
mean	0.025233	-0.011285	0.035009	0.028538	-0.000439	0.007717	
std	0.076039	0.074785	0.074317	0.074398	0.070587	0.072195	
min	-0.211629	-0.220033	-0.187707	-0.208964	-0.194298	-0.208729	
25%	-0.019985	-0.061217	-0.014483	-0.020738	-0.045776	-0.040936	
50%	0.020682	-0.014514	0.033428	0.028605	0.000209	0.007878	
75%	0.071148	0.039529	0.074361	0.072000	0.042800	0.048960	
max	0.275880	0.245294	0.273130	0.266109	0.203129	0.246204	

	X6	X7	X8	X9	X10	X11	\
count	457.000000	457.000000	457.000000	457.000000	457.000000	457.000000	
mean	0.027936	0.016218	0.006298	0.008393	0.017840	0.007129	
std	0.070301	0.068004	0.065871	0.066561	0.065152	0.063211	
min	-0.209917	-0.201924	-0.223088	-0.185778	-0.200440	-0.181736	
25%	-0.012417	-0.029242	-0.030233	-0.032699	-0.021906	-0.030138	
50%	0.027633	0.021680	0.007129	0.009219	0.015796	0.007704	
75%	0.069945	0.058289	0.043330	0.046642	0.056660	0.046200	
max	0.249562	0.252089	0.223580	0.252596	0.252136	0.188621	

	X12
count	457.000000
mean	0.006793
std	0.064018
min	-0.171411
25%	-0.031451
50%	0.004528
75%	0.043350
max	0.219762


nematoda efectores dataset 1 Covarianza de auto cruzamiento (ACC) hidro
0.3
0.2
0.1
0.0
−0.1
−0.2
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12


nematoda no_efectores dataset 1 Covarianza de auto cruzamiento (ACC) hidro
0.3
0.2
0.1
0.0
−0.1
−0.2
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12