ds1_nematoda_limpieza_de_datos

February 1, 2021

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    6.373   5.882  4.412  4.412  6.863  9.314  2.451  3.431  1.471   3.431
1    6.333  10.000  3.667  5.333  3.667  3.667  2.333  6.333  4.333   2.333
2    6.734   7.407  2.694  7.071  1.684  8.754  5.051  4.040  2.020   6.061
3    6.548   5.060  3.125  6.548  3.274  7.589  2.827  7.143  2.083   3.571
4    5.579   5.992  5.165  5.165  5.992  4.339  6.198  6.198  3.099   7.025
..     …       …      …      …      …      …      …      …      …       …
995  4.682   2.676  7.692  7.023  3.344  6.689  6.355  3.010  2.676   3.010
996  5.732   7.962  3.822  5.096  1.911  5.732  2.866  5.414  3.185   5.096
997  6.557   9.836  4.098  5.738  4.918  9.016  9.016  6.557  1.639   4.918
998  4.040   2.020  4.040  1.010  7.071  4.040  4.040  8.081  2.020  10.101
999  9.239   4.348  1.630  4.348  2.174  2.717  1.087  6.522  1.087   9.783
```

```
     …    X11    X12    X13     X14     X15    X16    X17    X18    X19  \
0    …  8.333  3.922  5.392   2.941   6.863  4.902  0.000  2.451  6.863
1    …  4.333  1.000  4.000  10.667  10.000  7.333  0.000  1.333  4.333
2    …  3.704  4.377  4.040   2.357   8.081  4.040  0.673  2.694  6.734
3    …  6.250  1.935  4.911   4.613   5.804  4.762  1.637  3.571  8.185
4    …  3.306  3.099  5.992   2.273   7.025  3.306  2.066  2.893  6.612
..   …    …    …    …     …       …    …    …    …    …
995  …  9.365  1.338  2.676   3.679  12.709  4.682  1.672  4.348  5.017
996  …  5.732  3.185  5.096   4.140   6.688  4.777  1.592  4.140  7.006
997  …  2.459  0.820  2.459   3.279   4.918  3.279  2.459  3.279  7.377
998  …  3.030  3.030  7.071   3.030   7.071  7.071  1.010  6.061  8.081
999  …  2.174  5.435  7.609   2.717   6.522  8.696  1.630  3.261  9.239

            X20
0     efectores
1     efectores
2     efectores
3     efectores
4     efectores
..          …
995   efectores
996   efectores
997   efectores
998   efectores
999   efectores

[1000 rows x 21 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      6.910065     6.038946     4.379460     5.196223     2.370344
std       2.606999     2.631675     1.942608     2.242583     1.973012
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       5.302500     4.430250     3.125000     3.881250     1.111000
50%       6.599000     5.816000     4.255000     5.112000     1.921000
75%       8.284250     7.308750     5.382000     6.345750     3.062000
max      25.987000    18.462000    17.073000    29.000000    12.857000

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      6.348358     3.828983     5.771930     2.521280     5.526219
std       3.225991     1.975915     3.306165     1.802421     2.227420
```

```
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       4.522500     2.610500     3.903750     1.518000     4.142500
50%       6.147000     3.685500     5.233500     2.282500     5.405000
75%       7.792000     4.762000     6.773250     3.196250     6.851750
max      55.000000    21.905000    34.021000    36.111000    14.286000

                X10          X11          X12          X13          X14  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      8.999158     5.878929     2.783987     4.258335     4.996510
std       3.128231     2.838078     1.575352     2.075743     3.076653
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       7.028750     4.029500     1.769250     2.896750     3.358000
50%       8.883000     5.579500     2.542000     4.061000     4.545000
75%      10.825750     7.205000     3.518750     5.307000     5.840250
max      24.675000    32.836000    19.588000    17.143000    36.000000

                X15          X16          X17          X18          X19
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      7.843990     5.560362     1.206274     3.160803     6.419940
std       3.228594     2.476802     1.003387     1.863218     2.351317
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       5.807750     4.223500     0.477500     2.043250     4.900750
50%       7.473500     5.277000     1.043500     2.937000     6.383000
75%       9.468750     6.571500     1.744750     4.054000     7.820250
max      22.581000    36.111000     6.818000    17.582000    16.234000
```

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

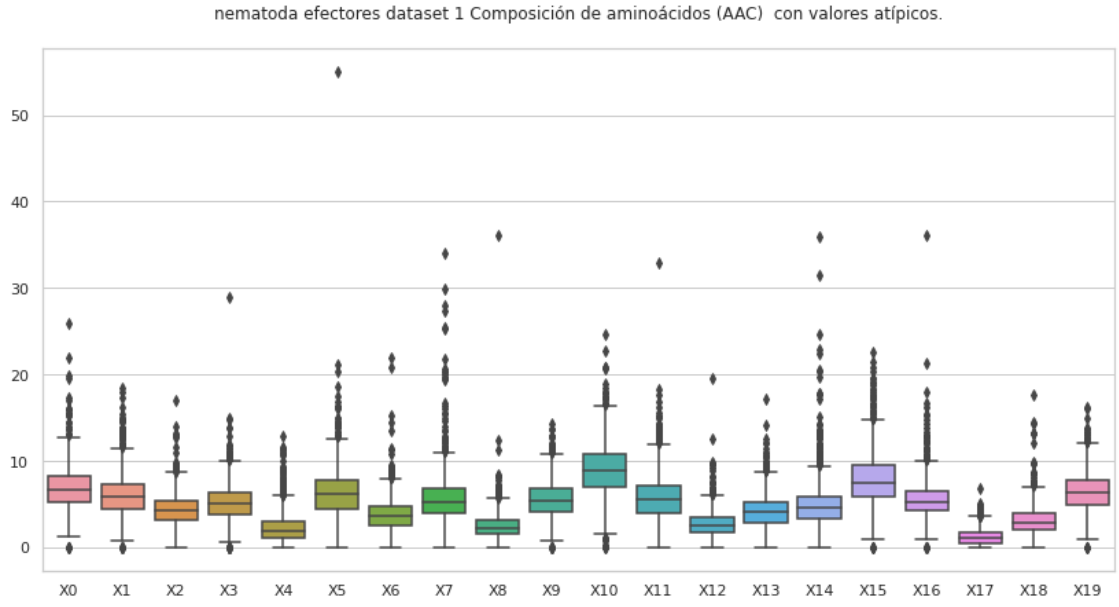
```
        X0     X1     X2     X3     X4     X5     X6     X7     X8     X9  \
0    5.144  5.144  5.446  5.446  1.967  6.505  3.631  2.723  4.539  7.867
1    4.967  4.305  3.642  2.649  2.980  1.325  2.980  3.642  1.987  8.940
2    7.303  7.210  5.993  4.401  1.404  7.584  4.682  5.618  2.154  6.742
3    9.417  2.691  4.484  6.726  0.448  9.865  1.794  1.345  0.000  1.345
4    4.955  6.532  4.505  5.856  3.604  5.405  2.252  5.180  1.351  4.054
..     …      …      …      …      …      …      …      …      …      …
995  3.614  8.434  1.205  6.024  2.410  9.639  1.205  9.639  2.410  7.229
996  7.595  4.219  4.430  5.696  1.688  4.641  3.376  7.806  3.165  9.072
997  6.162  7.563  4.482  5.602  2.241  7.003  3.361  6.443  2.241  5.322
998  5.327  4.843  4.116  5.811  3.874  4.358  2.179  8.717  3.148  3.390
999  6.278  4.933  3.587  4.036  3.139  3.587  3.139  7.175  0.000  8.969

     …     X11    X12    X13    X14     X15     X16    X17    X18     X19  \
0    …   4.690  2.269  4.992  4.236   8.321   4.085  1.059  3.631   6.959
```

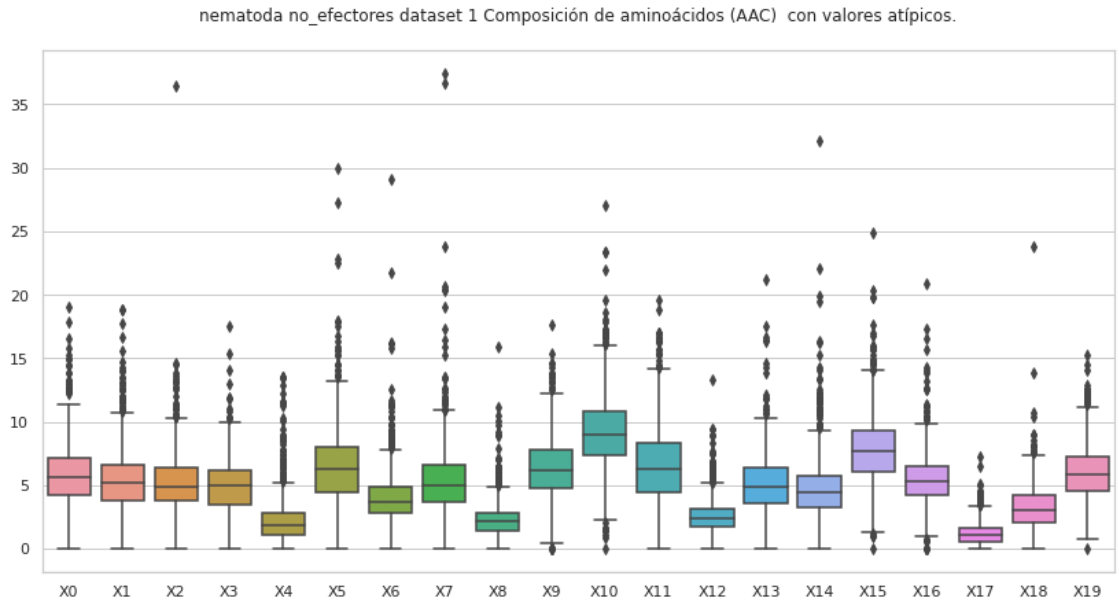
```
1    …   3.974  2.980  9.272  4.967   7.616   6.623  2.318  3.311   5.629
2    …   6.461  1.966  5.805  3.745   6.086   3.464  1.030  2.715   5.243
3    …  15.695  1.794  0.897  8.072  11.659  15.695  0.000  0.448   4.484
4    …   5.856  2.477  6.757  5.631   9.234   6.982  0.676  2.928   7.207
..   …     …    …    …    …     …     …    …    …     …
995  …   7.229  2.410  0.000  3.614   4.819   7.229  0.000  3.614   7.229
996  …   6.329  2.321  3.586  3.797   9.283   6.329  0.633  2.743   6.540
997  …   6.443  2.521  3.361  5.042   6.162   7.003  1.401  1.961   5.602
998  …   3.874  1.211  5.811  3.390  11.622   6.780  1.453  1.937  10.169
999  …   3.139  3.139  6.278  2.691   6.726   5.381  0.897  5.381   7.175

                X20
0     no_efectores
1     no_efectores
2     no_efectores
3     no_efectores
4     no_efectores
..             …
995   no_efectores
996   no_efectores
997   no_efectores
998   no_efectores
999   no_efectores

[1000 rows x 21 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      5.773885     5.396313     5.259419     4.910922     2.208191
std       2.532854     2.505410     2.480557     2.070800     1.861347
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       4.203500     3.811750     3.762000     3.499500     1.086250
50%       5.611500     5.232500     4.916500     5.012000     1.836500
75%       7.113500     6.607250     6.383000     6.196750     2.771000
max      19.040000    18.868000    36.478000    17.518000    13.542000

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      6.449910     4.011710     5.351272     2.258771     6.402505
std       3.104874     2.198608     2.965810     1.428379     2.439076
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       4.459750     2.778000     3.643500     1.413250     4.762000
50%       6.298000     3.679500     4.979000     2.150000     6.132500
```

```
75%       7.967000     4.834250     6.554750     2.830000     7.805250
max      29.952000    29.101000    37.421000    15.897000    17.647000

                X10          X11          X12          X13          X14  \
count   1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean       9.213277     6.687088     2.571816     5.124371     4.733937
std        2.965895     3.057318     1.335290     2.360689     2.582514
min        0.000000     0.000000     0.000000     0.000000     0.000000
25%        7.365250     4.483250     1.694000     3.601000     3.216250
50%        8.968000     6.274500     2.384000     4.878000     4.387500
75%       10.843750     8.382250     3.105000     6.383000     5.722500
max       27.053000    19.626000    13.333000    21.201000    32.143000

                X15          X16          X17          X18          X19
count   1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean       7.849852     5.415434     1.182808     3.232627     5.965926
std        2.757948     2.052391     0.930097     1.807367     2.136017
min        0.000000     0.000000     0.000000     0.000000     0.000000
25%        6.064750     4.167000     0.529000     2.037000     4.526500
50%        7.670500     5.336000     1.030500     3.070500     5.804500
75%        9.266500     6.466000     1.676750     4.176000     7.211250
max       24.868000    20.913000     7.292000    23.762000    15.278000
```





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  \
0    6.373   5.882  4.412  4.412  6.863  9.314  2.451  3.431  1.471   3.431
1    6.333  10.000  3.667  5.333  3.667  3.667  2.333  6.333  4.333   2.333
2    6.734   7.407  2.694  7.071  1.684  8.754  5.051  4.040  2.020   6.061
3    6.548   5.060  3.125  6.548  3.274  7.589  2.827  7.143  2.083   3.571
4    5.579   5.992  5.165  5.165  5.992  4.339  6.198  6.198  3.099   7.025
..     …       …      …      …      …      …      …      …      …       …
995  4.682   2.676  7.692  7.023  3.344  6.689  6.355  3.010  2.676   3.010
996  5.732   7.962  3.822  5.096  1.911  5.732  2.866  5.414  3.185   5.096
997  6.557   9.836  4.098  5.738  4.918  9.016  9.016  6.557  1.639   4.918
998  4.040   2.020  4.040  1.010  7.071  4.040  4.040  8.081  2.020  10.101
999  9.239   4.348  1.630  4.348  2.174  2.717  1.087  6.522  1.087   9.783

     …    X11    X12    X13     X14     X15    X16    X17    X18    X19  \
0    …  8.333  3.922  5.392   2.941   6.863  4.902  0.000  2.451  6.863
1    …  4.333  1.000  4.000  10.667  10.000  7.333  0.000  1.333  4.333
2    …  3.704  4.377  4.040   2.357   8.081  4.040  0.673  2.694  6.734
3    …  6.250  1.935  4.911   4.613   5.804  4.762  1.637  3.571  8.185
4    …  3.306  3.099  5.992   2.273   7.025  3.306  2.066  2.893  6.612
..   …    …      …      …       …       …      …      …      …      …
995  …  9.365  1.338  2.676   3.679  12.709  4.682  1.672  4.348  5.017
```

```
996  …  5.732  3.185  5.096   4.140   6.688  4.777  1.592  4.140  7.006
997  …  2.459  0.820  2.459   3.279   4.918  3.279  2.459  3.279  7.377
998  …  3.030  3.030  7.071   3.030   7.071  7.071  1.010  6.061  8.081
999  …  2.174  5.435  7.609   2.717   6.522  8.696  1.630  3.261  9.239

          X20
0    efectores
1    efectores
2    efectores
3    efectores
4    efectores
..         …
995  efectores
996  efectores
997  efectores
998  efectores
999  efectores

[838 rows x 21 columns]
```

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

```
                X0          X1          X2          X3          X4          X5  \
count   838.000000  838.000000  838.000000  838.000000  838.000000  838.000000
mean      6.844282    6.082461    4.376922    5.295431    2.328171    6.411549
std       2.181932    2.334980    1.637783    1.866854    1.641902    2.490785
min       0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%       5.420750    4.577000    3.259500    4.112000    1.251250    4.842000
50%       6.630000    5.845500    4.285000    5.187000    1.973000    6.266500
75%       8.162500    7.304250    5.317750    6.394750    3.045750    7.792000
max      14.545000   13.855000    9.701000   11.765000    8.046000   15.988000

                X6          X7          X8          X9         X10         X11  \
count   838.000000  838.000000  838.000000  838.000000  838.000000  838.000000
mean      3.804019    5.521097    2.480014    5.661537    9.281913    5.862840
std       1.582392    2.234317    1.242023    1.981311    2.710222    2.407767
min       0.000000    0.000000    0.000000    0.000000    1.176000    0.000000
25%       2.741250    4.034000    1.597000    4.321750    7.407000    4.194500
50%       3.704000    5.259500    2.310500    5.529500    9.186000    5.715500
75%       4.713500    6.662750    3.175000    6.931250   10.951250    7.114500
max       9.449000   14.970000    7.143000   12.146000   18.367000   14.286000

               X12         X13         X14         X15         X16         X17  \
count   838.000000  838.000000  838.000000  838.000000  838.000000  838.000000
mean      2.717525    4.404038    4.721414    7.822829    5.422687    1.216870
```

std	1.259068	1.805874	2.162029	2.819201	1.865467	0.889642
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	1.819750	3.176500	3.405500	5.909000	4.294750	0.588000
50%	2.574500	4.200500	4.505000	7.584000	5.279500	1.084000
75%	3.443250	5.385750	5.670750	9.402000	6.462750	1.752750
max	7.432000	10.471000	14.074000	17.460000	12.353000	4.145000

	X18	X19
count	838.000000	838.000000
mean	3.141105	6.603385
std	1.508996	2.082745
min	0.000000	0.885000
25%	2.158000	5.211500
50%	3.013000	6.565000
75%	4.017250	7.871750
max	8.537000	13.402000

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	5.144	5.144	5.446	5.446	1.967	6.505	3.631	2.723	4.539	7.867	
1	4.967	4.305	3.642	2.649	2.980	1.325	2.980	3.642	1.987	8.940	
2	7.303	7.210	5.993	4.401	1.404	7.584	4.682	5.618	2.154	6.742	
4	4.955	6.532	4.505	5.856	3.604	5.405	2.252	5.180	1.351	4.054	
5	5.732	5.732	3.822	3.503	2.548	7.325	5.414	5.096	1.911	5.732	
..	…	…	…	…	…	…	…	…	…		
995	3.614	8.434	1.205	6.024	2.410	9.639	1.205	9.639	2.410	7.229	
996	7.595	4.219	4.430	5.696	1.688	4.641	3.376	7.806	3.165	9.072	
997	6.162	7.563	4.482	5.602	2.241	7.003	3.361	6.443	2.241	5.322	
998	5.327	4.843	4.116	5.811	3.874	4.358	2.179	8.717	3.148	3.390	
999	6.278	4.933	3.587	4.036	3.139	3.587	3.139	7.175	0.000	8.969	

	…	X11	X12	X13	X14	X15	X16	X17	X18	X19	\
0	…	4.690	2.269	4.992	4.236	8.321	4.085	1.059	3.631	6.959	
1	…	3.974	2.980	9.272	4.967	7.616	6.623	2.318	3.311	5.629	
2	…	6.461	1.966	5.805	3.745	6.086	3.464	1.030	2.715	5.243	
4	…	5.856	2.477	6.757	5.631	9.234	6.982	0.676	2.928	7.207	
5	…	4.459	1.911	8.599	2.548	8.917	7.006	1.911	1.911	6.369	
..	…	…	…	…	…	…	…	…	…		
995	…	7.229	2.410	0.000	3.614	4.819	7.229	0.000	3.614	7.229	
996	…	6.329	2.321	3.586	3.797	9.283	6.329	0.633	2.743	6.540	
997	…	6.443	2.521	3.361	5.042	6.162	7.003	1.401	1.961	5.602	
998	…	3.874	1.211	5.811	3.390	11.622	6.780	1.453	1.937	10.169	

```
999  …  3.139  3.139  6.278  2.691   6.726  5.381  0.897  5.381   7.175

              X20
0     no_efectores
1     no_efectores
2     no_efectores
4     no_efectores
5     no_efectores
..             …
995   no_efectores
996   no_efectores
997   no_efectores
998   no_efectores
999   no_efectores

[830 rows x 21 columns]
```

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

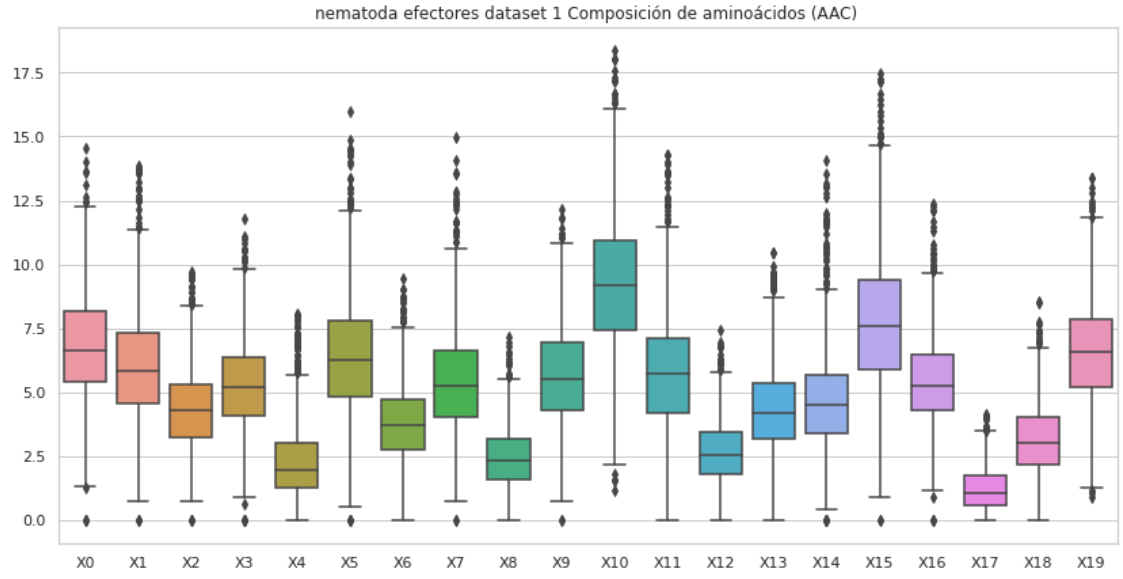
	X0	X1	X2	X3	X4	X5	\
count	830.000000	830.000000	830.000000	830.000000	830.000000	830.000000	
mean	5.732123	5.426178	5.234108	5.057927	2.124254	6.474536	
std	2.157069	2.032952	2.033000	1.819289	1.361142	2.546889	
min	0.000000	0.388000	0.000000	0.000000	0.000000	0.000000	
25%	4.354750	4.040000	3.909250	3.913500	1.188500	4.810500	
50%	5.621500	5.328000	4.989500	5.104000	1.878500	6.449000	
75%	7.018000	6.557000	6.319500	6.237250	2.755000	8.009750	
max	13.022000	12.658000	12.676000	10.995000	7.735000	15.738000	

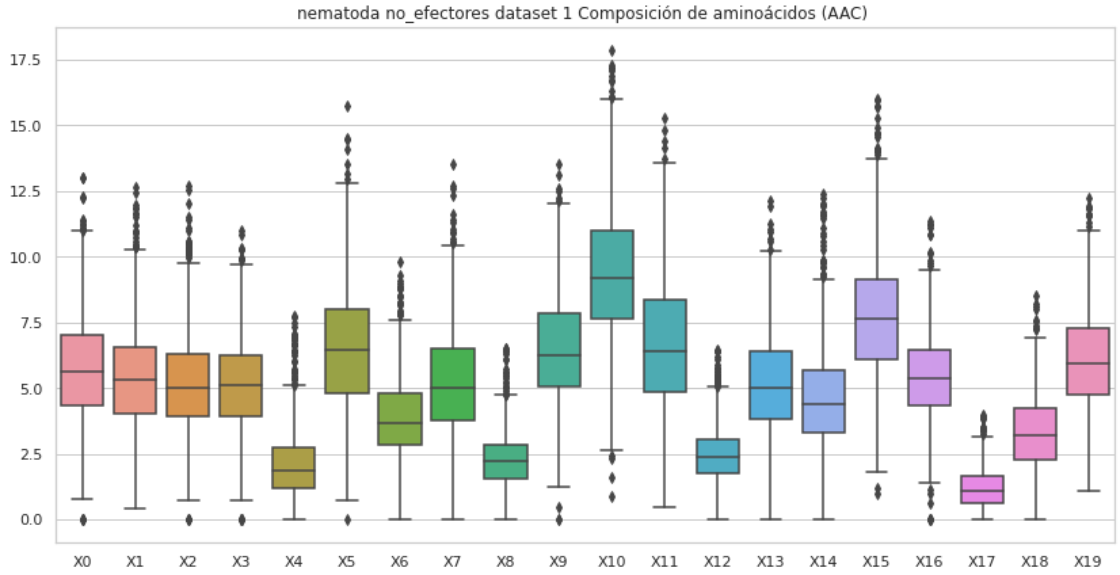
	X6	X7	X8	X9	X10	X11	\
count	830.000000	830.000000	830.000000	830.000000	830.000000	830.000000	
mean	3.909735	5.220066	2.256324	6.483959	9.385778	6.693933	
std	1.606365	2.075656	1.091596	2.176687	2.595882	2.647326	
min	0.000000	0.000000	0.000000	0.000000	0.901000	0.463000	
25%	2.863750	3.774750	1.550500	5.047250	7.636000	4.851250	
50%	3.679500	5.016500	2.206000	6.229000	9.201500	6.381000	
75%	4.801750	6.481250	2.830000	7.841250	10.990750	8.371500	
max	9.815000	13.514000	6.522000	13.538000	17.857000	15.278000	

	X12	X13	X14	X15	X16	X17	\
count	830.000000	830.000000	830.000000	830.000000	830.000000	830.000000	
mean	2.517878	5.179778	4.600270	7.758182	5.411986	1.183387	
std	1.104598	2.009027	1.993751	2.450892	1.670523	0.801335	
min	0.000000	0.000000	0.000000	0.990000	0.000000	0.000000	
25%	1.743250	3.817000	3.297000	6.078500	4.330250	0.626750	

50%	2.398500	4.990500	4.378000	7.634000	5.364500	1.069000
75%	3.069750	6.392000	5.653000	9.157750	6.428750	1.676000
max	6.478000	12.121000	12.381000	16.000000	11.343000	3.960000

	X18	X19
count	830.000000	830.000000
mean	3.295143	6.054467
std	1.517134	1.943580
min	0.000000	1.087000
25%	2.267750	4.726750
50%	3.184000	5.936500
75%	4.222000	7.288500
max	8.511000	12.230000





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	0.042333	0.045590	0.029308	0.061872	0.035821	0.022795	0.009769	
1	0.020101	0.011638	0.016927	0.011638	0.012696	0.020101	0.013753	
2	0.053493	0.013373	0.056168	0.069541	0.032096	0.032096	0.016048	
3	0.037273	0.018636	0.037273	0.043203	0.027955	0.040661	0.011860	
4	0.048213	0.051784	0.044641	0.037499	0.051784	0.053570	0.026785	
..	…	…	…	…	…	…	…	
995	0.015642	0.011173	0.023463	0.022346	0.008938	0.010056	0.008938	
996	0.053082	0.017694	0.047184	0.053082	0.047184	0.050133	0.029490	
997	0.050704	0.038028	0.044366	0.069717	0.019014	0.050704	0.012676	
998	0.013247	0.023183	0.003312	0.013247	0.023183	0.026494	0.006624	
999	0.023822	0.005605	0.011210	0.007006	0.019618	0.016815	0.002803	

	X7	X8	X9	…	X74	X75	X76	\
0	0.022795	0.055359	0.068385	…	-0.004548	0.002450	-0.007173	
1	0.007406	0.013753	0.028565	…	0.002364	0.004978	0.019006	
2	0.048144	0.029421	0.093613	…	-0.005760	0.034441	0.028391	
3	0.020331	0.035579	0.060145	…	0.020617	0.029349	0.010519	
4	0.060712	0.028570	0.074997	…	0.007009	0.014278	0.017343	
..	…	…	…	…	…	…	…	
995	0.010056	0.031284	0.024580	…	0.002789	0.018070	0.003471	
996	0.047184	0.053082	0.100265	…	-0.051324	-0.046206	-0.052154	
997	0.038028	0.019014	0.057041	…	-0.031645	0.013858	-0.001271	
998	0.033118	0.009935	0.026494	…	0.019924	0.018435	0.027706	
999	0.025223	0.005605	0.025223	…	0.035700	0.005919	0.001106	

	X77	X78	X79	X80	X81	X82	X83
0	-0.033810	-0.016785	-0.016594	-0.002852	0.024732	-0.014855	efectores
1	0.042515	0.026823	0.011857	0.011770	0.017982	0.016416	efectores
2	-0.026229	-0.036511	0.000087	0.005488	0.048497	0.006364	efectores
3	-0.008706	-0.003725	0.015890	-0.008288	0.010693	0.021357	efectores
4	0.010937	0.016341	-0.001054	-0.001891	0.017455	-0.039274	efectores
..	…	…	…	…	…	…	
995	0.010063	0.012035	0.010404	0.017296	0.023560	0.009196	efectores
996	0.044805	-0.014629	-0.013945	-0.071063	-0.026025	0.034249	efectores
997	0.005842	0.018322	0.030594	0.073357	0.051615	0.031191	efectores

```
998 -0.005833  0.009686  0.029465 -0.011932 -0.023130  0.020445  efectores
999  0.003572 -0.007286  0.016550  0.027447  0.002521 -0.006769  efectores

[1000 rows x 84 columns]
```

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

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.039566	0.014193	0.029767	0.036607	0.026759	
std	0.079465	0.017923	0.023814	0.078559	0.044717	
min	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.023116	0.004749	0.015850	0.019729	0.012040	
50%	0.033076	0.010064	0.025631	0.030955	0.020613	
75%	0.044284	0.017907	0.037112	0.043598	0.032423	
max	2.421161	0.242157	0.263911	2.421161	1.210581	

	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.031869	0.015131	0.034520	0.035942	0.058184	…	
std	0.043380	0.018069	0.079803	0.116918	0.193561	…	
min	0.000000	0.000000	0.000000	0.000000	0.000000	…	
25%	0.017028	0.006434	0.017273	0.018183	0.029119	…	
50%	0.026398	0.011427	0.026933	0.027553	0.044415	…	
75%	0.039410	0.018762	0.040054	0.040300	0.065630	…	
max	1.210581	0.306720	2.421161	3.631742	6.052903	…	

	X73	X74	X75	X76	X77	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.006475	0.011148	0.011085	0.011398	-0.014488	
std	0.182325	0.329923	0.158546	0.059121	0.454278	
min	-5.642736	-0.349700	-0.244850	-0.490266	-14.314873	
25%	0.000733	-0.010177	-0.004404	0.000511	-0.011918	
50%	0.011503	0.003325	0.006940	0.011740	0.003045	
75%	0.022861	0.016112	0.018783	0.022793	0.014684	
max	0.398767	10.361763	4.924181	1.419031	0.288155	

	X78	X79	X80	X81	X82
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	-0.003675	0.013272	-0.007910	-0.000407	0.011108
std	0.325701	0.063893	0.332780	0.199038	0.056252
min	-10.221416	-0.249748	-10.457505	-6.215630	-0.553980
25%	-0.004762	0.000745	-0.010042	-0.004779	0.000284
50%	0.007269	0.010665	0.003689	0.006939	0.011226
75%	0.018369	0.023172	0.016516	0.019235	0.022442

```
max       0.605885     1.658738     0.290152     0.260903     1.386178

[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	0.034942	0.013360	0.036997	0.044191	0.033914	0.018499	0.030831	
1	0.019224	0.011535	0.010253	0.005126	0.035885	0.014098	0.007690	
2	0.046541	0.008950	0.028044	0.048331	0.036994	0.035801	0.013724	
3	0.013281	0.000632	0.009487	0.013914	0.001265	0.001897	0.000000	
4	0.018787	0.013663	0.022203	0.020495	0.025618	0.019641	0.005124	
..	…	…	…	…	…	…	…	
995	0.016997	0.011331	0.028328	0.045324	0.000000	0.045324	0.011331	
996	0.033760	0.007502	0.025320	0.020631	0.015942	0.034698	0.014067	
997	0.033427	0.012155	0.030388	0.037985	0.018233	0.034946	0.012155	
998	0.020455	0.014877	0.022315	0.016736	0.022315	0.033472	0.012087	
999	0.031862	0.015931	0.020483	0.018207	0.031862	0.036414	0.000000	

	X7	X8	X9	…	X74	X75	X76	\
0	0.053440	0.031859	0.077077	…	0.021514	0.023186	0.009211	
1	0.034604	0.015379	0.061518	…	0.025929	0.004952	0.003185	
2	0.042961	0.041171	0.066231	…	-0.005979	-0.001500	0.015318	
3	0.001897	0.022136	0.004427	…	0.007802	0.026940	0.010089	
4	0.015371	0.022203	0.032450	…	0.002108	0.009335	0.009532	
..	…	…	…	…	…	…	…	
995	0.033993	0.033993	0.056656	…	0.074907	0.055246	0.025444	
996	0.040325	0.028133	0.030009	…	0.016178	0.007421	0.035764	
997	0.028869	0.034946	0.054699	…	-0.009633	-0.002077	0.015497	
998	0.013017	0.014877	0.030683	…	0.009360	0.009500	0.041088	
999	0.045517	0.015931	0.072827	…	0.016572	-0.012865	-0.015040	

	X77	X78	X79	X80	X81	X82	X83
0	-0.005970	0.022459	0.006132	0.009940	0.007352	-0.006890	no_efectores
1	0.004183	-0.010631	0.008917	0.002072	0.006741	-0.002440	no_efectores
2	0.020368	0.024249	0.005605	0.002571	0.005653	0.020460	no_efectores
3	0.008391	0.022068	0.009018	0.007809	0.020255	0.009647	no_efectores
4	-0.002106	-0.000860	0.011369	0.004957	-0.003299	-0.000336	no_efectores
..	…	…	…	…	…	…	…
995	0.034272	-0.004513	0.041941	-0.008633	-0.006541	-0.042622	no_efectores
996	0.013449	0.002625	0.031013	-0.003737	-0.005780	0.013363	no_efectores
997	0.030736	0.021168	0.025517	0.027540	0.025306	0.003154	no_efectores
998	0.011359	0.007678	0.031896	-0.015465	-0.007455	0.033793	no_efectores

```
999 -0.014904 -0.039498  0.035253  0.021098 -0.012380  0.000502  no_efectores

[1000 rows x 84 columns]
```

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

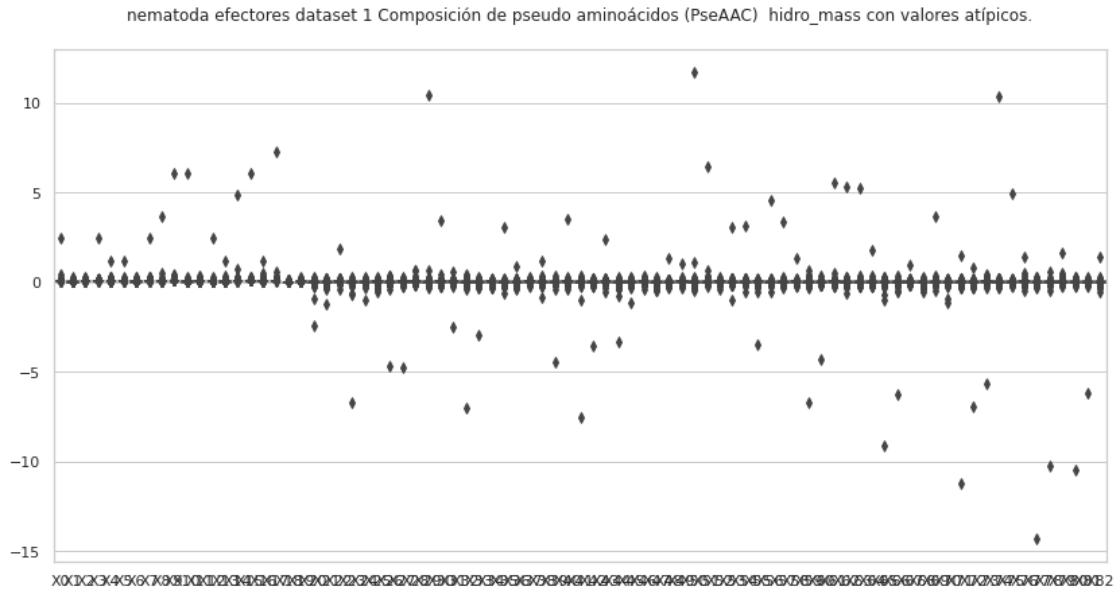
	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.029053	0.012110	0.027271	0.033382	0.028133	
std	0.048296	0.037370	0.025074	0.074607	0.076802	
min	-1.312987	-0.875324	-0.437662	-2.188311	-2.188311	
25%	0.018403	0.004477	0.014704	0.018880	0.014657	
50%	0.027396	0.009152	0.023780	0.031683	0.024029	
75%	0.038259	0.015828	0.037291	0.046991	0.037713	
max	0.435111	0.543889	0.132740	0.227931	0.398219	

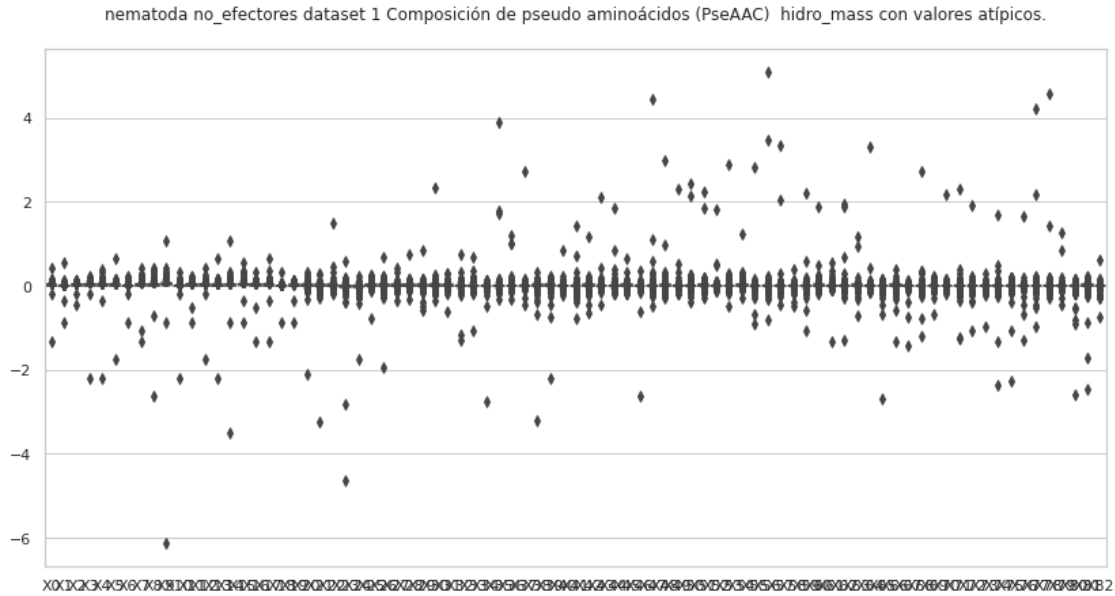
	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.026823	0.012598	0.034772	0.034885	0.047112	…	
std	0.062270	0.033299	0.062524	0.093600	0.204236	…	
min	-1.750649	-0.875324	-1.312987	-2.625973	-6.127271	…	
25%	0.016051	0.005382	0.019624	0.018836	0.029736	…	
50%	0.024299	0.010243	0.031196	0.031892	0.043460	…	
75%	0.036077	0.016809	0.045263	0.048391	0.065838	…	
max	0.652667	0.217556	0.435111	0.435111	1.087779	…	

	X73	X74	X75	X76	X77	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.005995	-0.001141	0.002955	0.007245	0.006836	
std	0.041003	0.108145	0.085892	0.074672	0.158595	
min	-0.984093	-2.363936	-2.272592	-1.278297	-0.960347	
25%	-0.001228	-0.009861	-0.004503	-0.000369	-0.007383	
50%	0.007910	0.002953	0.006929	0.008297	0.003747	
75%	0.018031	0.013290	0.018016	0.017303	0.015568	
max	0.284727	1.697514	0.211720	1.651628	4.231847	

	X78	X79	X80	X81	X82
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.012821	0.010053	-0.002765	0.003521	0.006219
std	0.154346	0.055265	0.098395	0.104093	0.041146
min	-0.280525	-0.440189	-2.583829	-2.468017	-0.753894
25%	-0.002286	0.000016	-0.008033	-0.003415	-0.002023
50%	0.007697	0.008120	0.003985	0.007965	0.007529
75%	0.019603	0.019137	0.014976	0.021061	0.016941
max	4.586586	1.258091	0.150174	0.240124	0.619529

[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	0.042333	0.045590	0.029308	0.061872	0.035821	0.022795	0.009769	
1	0.020101	0.011638	0.016927	0.011638	0.012696	0.020101	0.013753	
2	0.053493	0.013373	0.056168	0.069541	0.032096	0.032096	0.016048	
3	0.037273	0.018636	0.037273	0.043203	0.027955	0.040661	0.011860	
4	0.048213	0.051784	0.044641	0.037499	0.051784	0.053570	0.026785	
..	…	…	…	…	…	…	…	
995	0.015642	0.011173	0.023463	0.022346	0.008938	0.010056	0.008938	
996	0.053082	0.017694	0.047184	0.053082	0.047184	0.050133	0.029490	
997	0.050704	0.038028	0.044366	0.069717	0.019014	0.050704	0.012676	
998	0.013247	0.023183	0.003312	0.013247	0.023183	0.026494	0.006624	
999	0.023822	0.005605	0.011210	0.007006	0.019618	0.016815	0.002803	

	X7	X8	X9	…	X74	X75	X76	\
0	0.022795	0.055359	0.068385	…	-0.004548	0.002450	-0.007173	
1	0.007406	0.013753	0.028565	…	0.002364	0.004978	0.019006	
2	0.048144	0.029421	0.093613	…	-0.005760	0.034441	0.028391	
3	0.020331	0.035579	0.060145	…	0.020617	0.029349	0.010519	
4	0.060712	0.028570	0.074997	…	0.007009	0.014278	0.017343	
..	…	…	…	…	…	…	…	
995	0.010056	0.031284	0.024580	…	0.002789	0.018070	0.003471	
996	0.047184	0.053082	0.100265	…	-0.051324	-0.046206	-0.052154	
997	0.038028	0.019014	0.057041	…	-0.031645	0.013858	-0.001271	
998	0.033118	0.009935	0.026494	…	0.019924	0.018435	0.027706	
999	0.025223	0.005605	0.025223	…	0.035700	0.005919	0.001106	

	X77	X78	X79	X80	X81	X82	X83
0	-0.033810	-0.016785	-0.016594	-0.002852	0.024732	-0.014855	efectores
1	0.042515	0.026823	0.011857	0.011770	0.017982	0.016416	efectores
2	-0.026229	-0.036511	0.000087	0.005488	0.048497	0.006364	efectores
3	-0.008706	-0.003725	0.015890	-0.008288	0.010693	0.021357	efectores
4	0.010937	0.016341	-0.001054	-0.001891	0.017455	-0.039274	efectores
..	…	…	…	…	…	…	
995	0.010063	0.012035	0.010404	0.017296	0.023560	0.009196	efectores
996	0.044805	-0.014629	-0.013945	-0.071063	-0.026025	0.034249	efectores
997	0.005842	0.018322	0.030594	0.073357	0.051615	0.031191	efectores
998	-0.005833	0.009686	0.029465	-0.011932	-0.023130	0.020445	efectores
999	0.003572	-0.007286	0.016550	0.027447	0.002521	-0.006769	efectores

[938 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	938.000000	938.000000	938.000000	938.000000	938.000000	938.000000	
mean	0.034334	0.012174	0.026890	0.032129	0.022707	0.028075	
std	0.016452	0.010871	0.015853	0.018228	0.015730	0.015253	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.022885	0.004646	0.015484	0.019171	0.011486	0.016790	
50%	0.032346	0.009460	0.024697	0.030055	0.019867	0.025537	
75%	0.042715	0.016777	0.036069	0.041937	0.030193	0.037822	
max	0.123112	0.064365	0.097550	0.112741	0.130967	0.091038	

	X6	X7	X8	X9	…	X73	\
count	938.000000	938.000000	938.000000	938.000000	…	938.000000	
mean	0.013290	0.029558	0.029812	0.047548	…	0.011477	
std	0.010240	0.019054	0.017734	0.027783	…	0.019709	
min	0.000000	0.000000	0.000000	0.000000	…	-0.099965	
25%	0.006249	0.016702	0.017580	0.028182	…	0.001368	
50%	0.010848	0.026331	0.026912	0.042655	…	0.011489	
75%	0.017425	0.038707	0.038449	0.062814	…	0.022266	
max	0.059223	0.157247	0.123416	0.215773	…	0.105345	

	X74	X75	X76	X77	X78	X79	\
count	938.000000	938.000000	938.000000	938.000000	938.000000	938.000000	
mean	0.002812	0.007402	0.011905	0.001691	0.007426	0.011805	
std	0.027045	0.023864	0.020545	0.026251	0.022218	0.020713	
min	-0.242174	-0.206772	-0.109631	-0.131146	-0.100263	-0.158107	
25%	-0.009137	-0.003658	0.001860	-0.010676	-0.003065	0.001566	
50%	0.003715	0.007279	0.012230	0.003467	0.007737	0.010894	
75%	0.015928	0.018648	0.022784	0.014616	0.018350	0.022898	
max	0.162451	0.179700	0.140790	0.234313	0.126684	0.136321	

	X80	X81	X82
count	938.000000	938.000000	938.000000
mean	0.002687	0.006809	0.011559
std	0.025748	0.021914	0.020817
min	-0.131883	-0.114480	-0.115236
25%	-0.008705	-0.004098	0.000976
50%	0.003770	0.007329	0.011372
75%	0.015864	0.019230	0.022069
max	0.148573	0.100987	0.126418

[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    0.034942  0.013360  0.036997  0.044191  0.033914  0.018499  0.030831
1    0.019224  0.011535  0.010253  0.005126  0.035885  0.014098  0.007690
2    0.046541  0.008950  0.028044  0.048331  0.036994  0.035801  0.013724
3    0.013281  0.000632  0.009487  0.013914  0.001265  0.001897  0.000000
4    0.018787  0.013663  0.022203  0.020495  0.025618  0.019641  0.005124
..        …         …         …         …         …         …         …
995  0.016997  0.011331  0.028328  0.045324  0.000000  0.045324  0.011331
996  0.033760  0.007502  0.025320  0.020631  0.015942  0.034698  0.014067
997  0.033427  0.012155  0.030388  0.037985  0.018233  0.034946  0.012155
998  0.020455  0.014877  0.022315  0.016736  0.022315  0.033472  0.012087
999  0.031862  0.015931  0.020483  0.018207  0.031862  0.036414  0.000000

           X7        X8        X9  …       X74       X75       X76  \
0    0.053440  0.031859  0.077077  …  0.021514  0.023186  0.009211
1    0.034604  0.015379  0.061518  …  0.025929  0.004952  0.003185
2    0.042961  0.041171  0.066231  … -0.005979 -0.001500  0.015318
3    0.001897  0.022136  0.004427  …  0.007802  0.026940  0.010089
4    0.015371  0.022203  0.032450  …  0.002108  0.009335  0.009532
..        …         …         …  …         …         …         …
995  0.033993  0.033993  0.056656  …  0.074907  0.055246  0.025444
996  0.040325  0.028133  0.030009  …  0.016178  0.007421  0.035764
997  0.028869  0.034946  0.054699  … -0.009633 -0.002077  0.015497
998  0.013017  0.014877  0.030683  …  0.009360  0.009500  0.041088
999  0.045517  0.015931  0.072827  …  0.016572 -0.012865 -0.015040

           X77       X78       X79       X80       X81       X82           X83
0   -0.005970  0.022459  0.006132  0.009940  0.007352 -0.006890  no_efectores
1    0.004183 -0.010631  0.008917  0.002072  0.006741 -0.002440  no_efectores
2    0.020368  0.024249  0.005605  0.002571  0.005653  0.020460  no_efectores
3    0.008391  0.022068  0.009018  0.007809  0.020255  0.009647  no_efectores
4   -0.002106 -0.000860  0.011369  0.004957 -0.003299 -0.000336  no_efectores
..        …         …         …         …         …         …           …
995  0.034272 -0.004513  0.041941 -0.008633 -0.006541 -0.042622  no_efectores
996  0.013449  0.002625  0.031013 -0.003737 -0.005780  0.013363  no_efectores
997  0.030736  0.021168  0.025517  0.027540  0.025306  0.003154  no_efectores
998  0.011359  0.007678  0.031896 -0.015465 -0.007455  0.033793  no_efectores
999 -0.014904 -0.039498  0.035253  0.021098 -0.012380  0.000502  no_efectores

[956 rows x 84 columns]
```

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

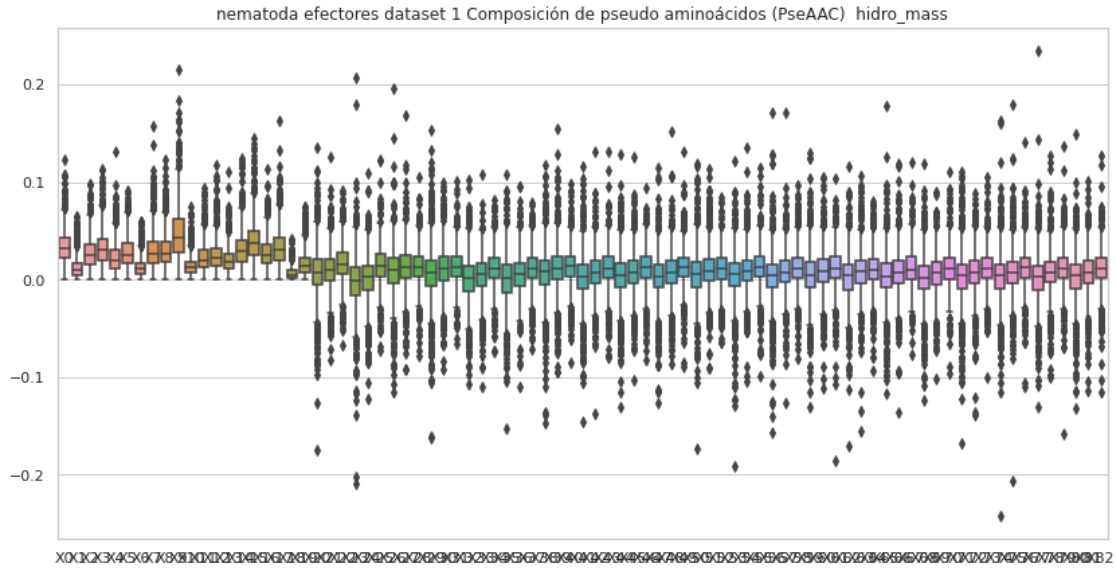
	X0	X1	X2	X3	X4	X5	\
count	956.000000	956.000000	956.000000	956.000000	956.000000	956.000000	
mean	0.029255	0.011836	0.026844	0.034033	0.027757	0.026847	
std	0.015821	0.011526	0.017162	0.021109	0.019401	0.014985	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.018390	0.004474	0.014534	0.018714	0.014473	0.015951	
50%	0.026983	0.009005	0.023618	0.030954	0.023518	0.024051	
75%	0.037059	0.015164	0.036117	0.045542	0.036366	0.035218	
max	0.154054	0.119550	0.097512	0.214134	0.127069	0.118549	

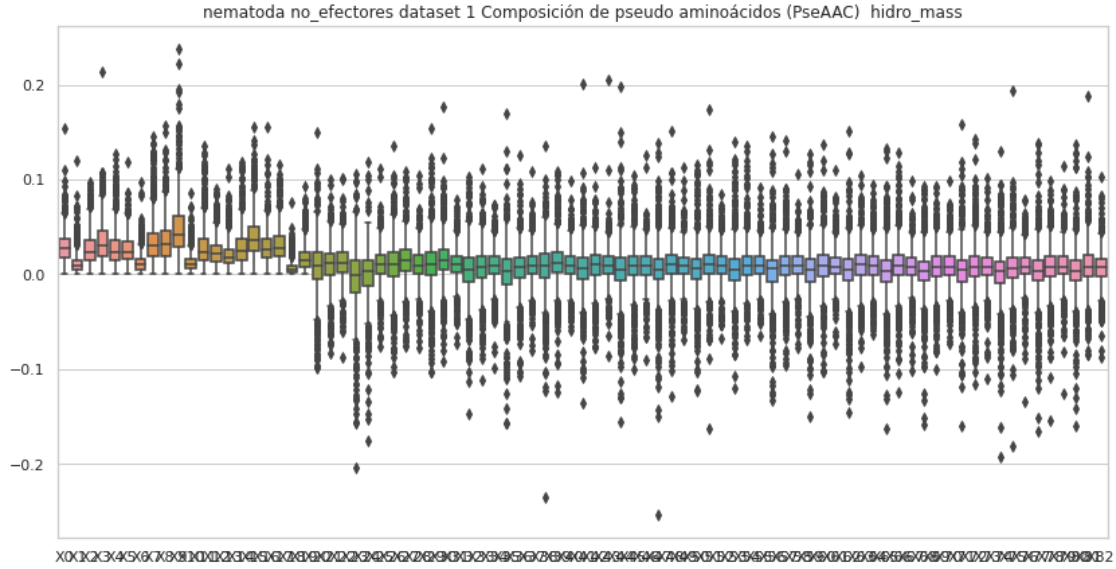
	X6	X7	X8	X9	…	X73	\
count	956.000000	956.000000	956.000000	956.000000	…	956.000000	
mean	0.012256	0.034735	0.035193	0.049077	…	0.008112	
std	0.010133	0.022666	0.022238	0.029720	…	0.018715	
min	0.000000	0.000000	0.000000	0.000000	…	-0.109650	
25%	0.005481	0.019387	0.018687	0.029507	…	-0.000486	
50%	0.010136	0.030395	0.031252	0.042384	…	0.008162	
75%	0.016377	0.043702	0.046436	0.062377	…	0.018031	
max	0.097880	0.145210	0.157032	0.237616	…	0.101814	

	X74	X75	X76	X77	X78	X79	\
count	956.000000	956.000000	956.000000	956.000000	956.000000	956.000000	
mean	0.000695	0.006284	0.008492	0.002830	0.007228	0.008880	
std	0.026799	0.024099	0.018386	0.026824	0.022545	0.018825	
min	-0.192869	-0.181172	-0.134122	-0.166049	-0.153383	-0.125678	
25%	-0.009037	-0.004176	0.000017	-0.006955	-0.001917	0.000349	
50%	0.002953	0.006924	0.008444	0.003888	0.007652	0.008313	
75%	0.012967	0.017534	0.017176	0.015242	0.019080	0.018817	
max	0.129922	0.193627	0.078180	0.138122	0.098398	0.114826	

	X80	X81	X82
count	956.000000	956.000000	956.000000
mean	0.003426	0.009031	0.007188
std	0.025880	0.023239	0.018292
min	-0.159991	-0.085113	-0.086697
25%	-0.006841	-0.002777	-0.001437
50%	0.004187	0.008048	0.007568
75%	0.014976	0.020577	0.016796
max	0.136469	0.187732	0.102692

[8 rows x 83 columns]


nematoda efectores dataset 1 Composición de pseudo aminoácidos (PseAAC) hidro_mass
0.2
0.1
0.0
−0.1
−0.2


nematoda no_efectores dataset 1 Composición de pseudo aminoácidos (PseAAC) hidro_mass
0.2
0.1
0.0
−0.1
−0.2

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    0.054842  0.059060  0.037967  0.080153  0.046405  0.029530  0.012656
1    0.029843  0.017278  0.025131  0.017278  0.018848  0.029843  0.020419
2    0.053838  0.013460  0.056530  0.069990  0.032303  0.032303  0.016151
3    0.044806  0.022403  0.044806  0.051934  0.033605  0.048879  0.014256
4    0.047503  0.051022  0.043984  0.036947  0.051022  0.052781  0.026390
..        …         …         …         …         …         …         …
995  0.038280  0.027343  0.057420  0.054686  0.021874  0.024609  0.021874
996  0.050625  0.016875  0.045000  0.050625  0.045000  0.047812  0.028125
```

```
997  0.074084  0.055563  0.064823  0.101865  0.027781  0.074084  0.018521
998  0.021015  0.036777  0.005254  0.021015  0.036777  0.042030  0.010508
999  0.056741  0.013351  0.026701  0.016688  0.046728  0.040052  0.006675

           X7        X8        X9  …       X32       X33       X34  \
0    0.029530  0.071716  0.088590  …  0.016578 -0.034561 -0.002222
1    0.010995  0.020419  0.042409  …  0.013675  0.033557  0.038197
2    0.048454  0.029611  0.094217  … -0.000478 -0.008089 -0.014269
3    0.024440  0.042769  0.072301  …  0.004967  0.005752  0.025124
4    0.059818  0.028150  0.073893  … -0.021090 -0.011271  0.039584
..        …         …         …  …        …         …         …
995  0.024609  0.076560  0.060154  …  0.012683 -0.001648 -0.000094
996  0.045000  0.050625  0.095624  … -0.010375 -0.001971  0.021210
997  0.055563  0.027781  0.083344  … -0.083440 -0.066658 -0.038510
998  0.052538  0.015761  0.042030  …  0.023093  0.011598 -0.003939
999  0.060078  0.013351  0.060078  …  0.004595  0.016086 -0.017987

          X35       X36       X37       X38       X39       X40         X41
0    0.001526  0.022612  0.003088 -0.009292 -0.021497 -0.019244  efectores
1    0.032502  0.036195  0.035619  0.028217  0.017604  0.024372  efectores
2    0.033033  0.019642 -0.000427  0.028574  0.000088  0.006405  efectores
3    0.028174  0.018713 -0.000267  0.012645  0.019102  0.025674  efectores
4    0.001885  0.024593  0.029267  0.017088 -0.001038 -0.038696  efectores
..        …         …         …         …         …         …         …
995  0.013500  0.013125  0.015227  0.008494  0.025462  0.022506  efectores
996  0.076494  0.012607 -0.021901 -0.049740 -0.013299  0.032664  efectores
997 -0.026402 -0.017788  0.020764 -0.001856  0.044702  0.045574  efectores
998  0.004995  0.033323  0.007503  0.043952  0.046743  0.032433  efectores
999  0.003811  0.050070  0.052640  0.002634  0.039421 -0.016122  efectores

[1000 rows x 42 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.045910     0.016466     0.036946     0.045386     0.031003
std       0.018121     0.014862     0.021519     0.029144     0.021185
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       0.034976     0.007126     0.022983     0.026842     0.016982
50%       0.043816     0.013236     0.034746     0.040649     0.027112
75%       0.054854     0.021307     0.046855     0.058078     0.039807
max       0.147360     0.116410     0.231839     0.439694     0.256509

                X5           X6           X7           X8           X9  …  \
```

```
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
mean      0.036907     0.018189     0.039221     0.042048     0.064000  …
std       0.016528     0.017652     0.021076     0.027109     0.031208  …
min       0.000000     0.000000     0.000000     0.000000     0.000000  …
25%       0.026437     0.009344     0.024195     0.024491     0.042586  …
50%       0.035300     0.015212     0.036465     0.037129     0.060681  …
75%       0.044662     0.023930     0.050799     0.053689     0.081950  …
max       0.139013     0.411947     0.160023     0.337785     0.221521  …

               X31          X32          X33          X34          X35  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.013946     0.013729     0.013861     0.011263     0.013166
std       0.030206     0.027546     0.029893     0.029912     0.027679
min      -0.246998    -0.174704    -0.227958    -0.196005    -0.142672
25%       0.002055     0.001320     0.000928    -0.001806    -0.000538
50%       0.016862     0.016438     0.016241     0.013665     0.014895
75%       0.029759     0.029892     0.029653     0.027439     0.028941
max       0.161066     0.156268     0.169602     0.118668     0.132105

               X36          X37          X38          X39          X40
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.012140     0.013782     0.013205     0.013181     0.012536
std       0.031160     0.029750     0.032921     0.030432     0.030934
min      -0.294499    -0.145765    -0.216701    -0.194196    -0.291116
25%      -0.001600     0.001001     0.001159     0.001182     0.000621
50%       0.014724     0.016574     0.016572     0.015950     0.015496
75%       0.029435     0.028707     0.028582     0.028170     0.028107
max       0.189369     0.214874     0.431515     0.262686     0.150590

[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    0.044418  0.016983  0.047031  0.056175  0.043111  0.023515  0.039192
1    0.044979  0.026988  0.023989  0.011994  0.083962  0.032985  0.017992
2    0.053661  0.010319  0.032334  0.055724  0.042653  0.041277  0.015823
3    0.036320  0.001730  0.025943  0.038050  0.003459  0.005189  0.000000
4    0.029859  0.021716  0.035288  0.032573  0.040717  0.031216  0.008143
..        …         …         …         …         …         …
995  0.021648  0.014432  0.036080  0.057728  0.000000  0.057728  0.014432
996  0.034360  0.007636  0.025770  0.020998  0.016226  0.035315  0.014317
997  0.042075  0.015300  0.038250  0.047812  0.022950  0.043987  0.015300
```

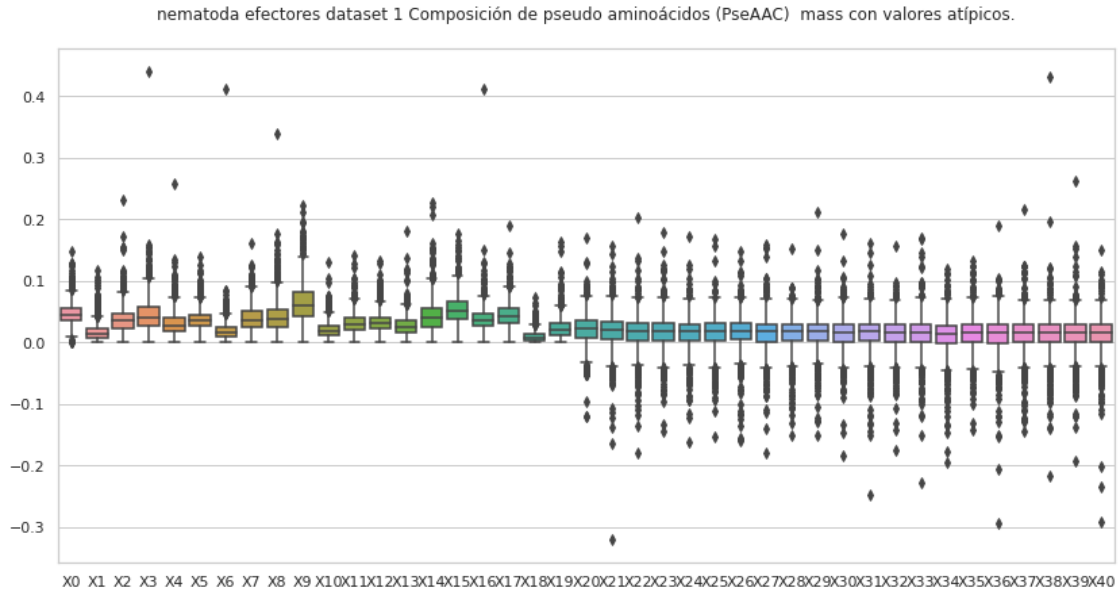
```
998  0.021937  0.015955  0.023932  0.017949  0.023932  0.035898  0.012963
999  0.037416  0.018708  0.024053  0.021381  0.037416  0.042762  0.000000

          X7        X8        X9  …       X32       X33       X34  \
0    0.067933  0.040499  0.097980  …  0.011240  0.002346  0.013670
1    0.080963  0.035984  0.143934  …  0.015118 -0.006434  0.026619
2    0.049533  0.047469  0.076363  …  0.011222  0.020248  0.024191
3    0.005189  0.060534  0.012107  …  0.028190  0.035078  0.024535
4    0.024430  0.035288  0.051575  …  0.018808  0.019919  0.017405
..        …         …         …  …       …         …         …
995  0.043296  0.043296  0.072161  …  0.053140  0.062042  0.047942
996  0.041042  0.028634  0.030543  …  0.013756  0.016528  0.010018
997  0.036337  0.043987  0.068849  …  0.013949  0.000477  0.016541
998  0.013960  0.015955  0.032906  …  0.038736  0.030882  0.034561
999  0.053452  0.018708  0.085523  …  0.010939 -0.008883  0.014012

          X35       X36       X37       X38       X39       X40           X41
0    0.022872  0.015902 -0.007222  0.011709  0.007796 -0.008758  no_efectores
1   -0.006747  0.012018  0.015650  0.007453  0.020864 -0.005709  no_efectores
2    0.017334 -0.007309  0.026904  0.017661  0.006462  0.023590  no_efectores
3    0.026828  0.027208  0.027548  0.027591  0.024662  0.026380  no_efectores
4    0.001548  0.018006  0.012648  0.015150  0.018069 -0.000535  no_efectores
..        …         …         …         …         …         …           …
995  0.057951  0.020071  0.027246  0.032407  0.053419 -0.054287  no_efectores
996  0.034235  0.034217  0.026785  0.036400  0.031564  0.013601  no_efectores
997 -0.004361  0.019700  0.018372  0.019506  0.032118  0.003970  no_efectores
998  0.027489  0.018041  0.031967  0.044065  0.034207  0.036242  no_efectores
999  0.002857  0.026225 -0.019339 -0.017662  0.041398  0.000589  no_efectores

[1000 rows x 42 columns]


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

                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.041007     0.016516     0.037228     0.050055     0.039923
std       0.017636     0.014473     0.019543     0.031136     0.024843
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       0.029831     0.007448     0.023239     0.028866     0.022695
50%       0.039715     0.013505     0.035658     0.044727     0.035268
75%       0.049892     0.020851     0.048712     0.067111     0.051903
max       0.139798     0.108731     0.163484     0.316959     0.271868

                X5           X6           X7           X8           X9  …  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
```

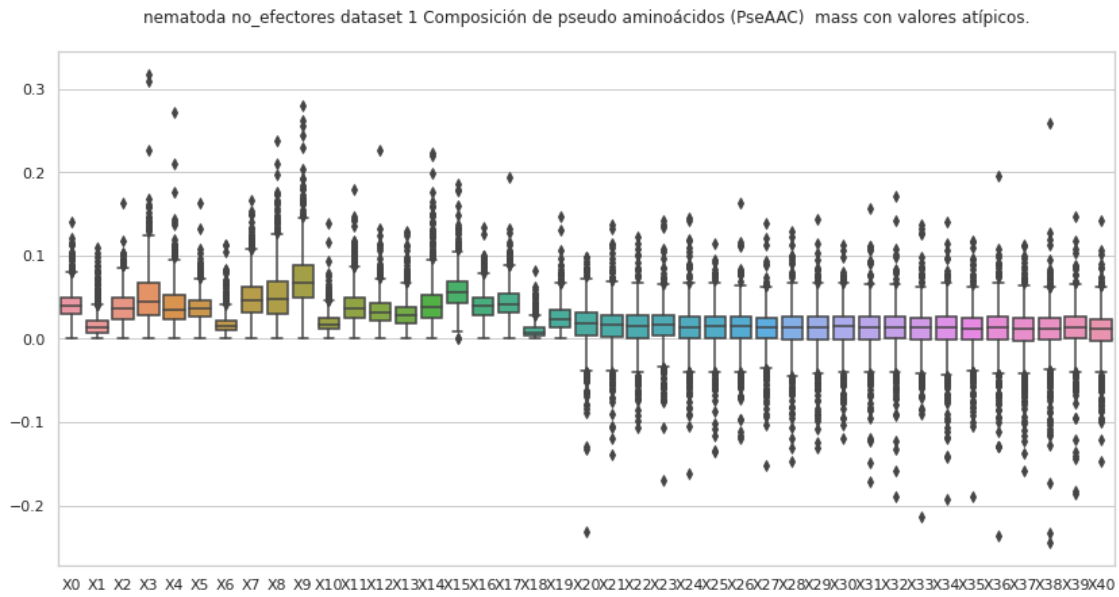
mean	0.037047	0.017291	0.049304	0.052074	0.070675	…
std	0.015820	0.012850	0.025489	0.031623	0.033646	…
min	0.000000	0.000000	0.000000	0.000000	0.000000	…
25%	0.026984	0.009510	0.031945	0.029190	0.048914	…
50%	0.035968	0.015603	0.045364	0.046953	0.066994	…
75%	0.045312	0.022340	0.062464	0.068005	0.087898	…
max	0.163366	0.112043	0.166045	0.237786	0.279352	…

	X31	X32	X33	X34	X35	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.011186	0.012030	0.011775	0.010776	0.010065	
std	0.026576	0.026348	0.025542	0.027740	0.025227	
min	-0.171323	-0.188633	-0.213547	-0.193132	-0.189900	
25%	-0.000600	-0.000162	-0.001302	-0.001621	-0.001045	
50%	0.013350	0.013749	0.013090	0.014212	0.012295	
75%	0.026110	0.026381	0.025636	0.026303	0.024588	
max	0.156702	0.170272	0.136129	0.139702	0.117675	

	X36	X37	X38	X39	X40
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.011204	0.009596	0.010172	0.010901	0.008740
std	0.028060	0.027087	0.030030	0.027271	0.026490
min	-0.237207	-0.158073	-0.244610	-0.185414	-0.146873
25%	-0.000956	-0.002044	-0.000575	-0.000124	-0.003132
50%	0.013577	0.012670	0.012617	0.013246	0.012008
75%	0.026874	0.025103	0.024483	0.026287	0.023427
max	0.194523	0.111759	0.258240	0.146301	0.142125

[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   0.054842  0.059060  0.037967  0.080153  0.046405  0.029530  0.012656
1   0.029843  0.017278  0.025131  0.017278  0.018848  0.029843  0.020419
2   0.053838  0.013460  0.056530  0.069990  0.032303  0.032303  0.016151
3   0.044806  0.022403  0.044806  0.051934  0.033605  0.048879  0.014256
```

```
4    0.047503  0.051022  0.043984  0.036947  0.051022  0.052781  0.026390
..        …         …         …         …         …         …
994  0.039459  0.021523  0.046633  0.025110  0.035872  0.050220  0.032285
995  0.038280  0.027343  0.057420  0.054686  0.021874  0.024609  0.021874
996  0.050625  0.016875  0.045000  0.050625  0.045000  0.047812  0.028125
998  0.021015  0.036777  0.005254  0.021015  0.036777  0.042030  0.010508
999  0.056741  0.013351  0.026701  0.016688  0.046728  0.040052  0.006675

           X7        X8        X9  …       X32       X33       X34  \
0    0.029530  0.071716  0.088590  …  0.016578 -0.034561 -0.002222
1    0.010995  0.020419  0.042409  …  0.013675  0.033557  0.038197
2    0.048454  0.029611  0.094217  … -0.000478 -0.008089 -0.014269
3    0.024440  0.042769  0.072301  …  0.004967  0.005752  0.025124
4    0.059818  0.028150  0.073893  … -0.021090 -0.011271  0.039584
..        …         …         …  …         …         …         …
994  0.032285  0.021523  0.050220  …  0.021287  0.028400 -0.005855
995  0.024609  0.076560  0.060154  …  0.012683 -0.001648 -0.000094
996  0.045000  0.050625  0.095624  … -0.010375 -0.001971  0.021210
998  0.052538  0.015761  0.042030  …  0.023093  0.011598 -0.003939
999  0.060078  0.013351  0.060078  …  0.004595  0.016086 -0.017987

           X35       X36       X37       X38       X39       X40         X41
0    0.001526  0.022612  0.003088 -0.009292 -0.021497 -0.019244  efectores
1    0.032502  0.036195  0.035619  0.028217  0.017604  0.024372  efectores
2    0.033033  0.019642 -0.000427  0.028574  0.000088  0.006405  efectores
3    0.028174  0.018713 -0.000267  0.012645  0.019102  0.025674  efectores
4    0.001885  0.024593  0.029267  0.017088 -0.001038 -0.038696  efectores
..        …         …         …         …         …         …
994  0.032072  0.028891  0.018935 -0.004516  0.043958  0.028222  efectores
995  0.013500  0.013125  0.015227  0.008494  0.025462  0.022506  efectores
996  0.076494  0.012607 -0.021901 -0.049740 -0.013299  0.032664  efectores
998  0.004995  0.033323  0.007503  0.043952  0.046743  0.032433  efectores
999  0.003811  0.050070  0.052640  0.002634  0.039421 -0.016122  efectores

[818 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  818.000000  818.000000  818.000000  818.000000  818.000000  818.000000
mean     0.044024    0.014873    0.033950    0.041673    0.027988    0.034979
std      0.014470    0.011206    0.016652    0.022333    0.015752    0.012750
min      0.008405    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.034963    0.006882    0.022131    0.026150    0.016529    0.026125
50%      0.042870    0.012469    0.032867    0.039338    0.025866    0.034245
```

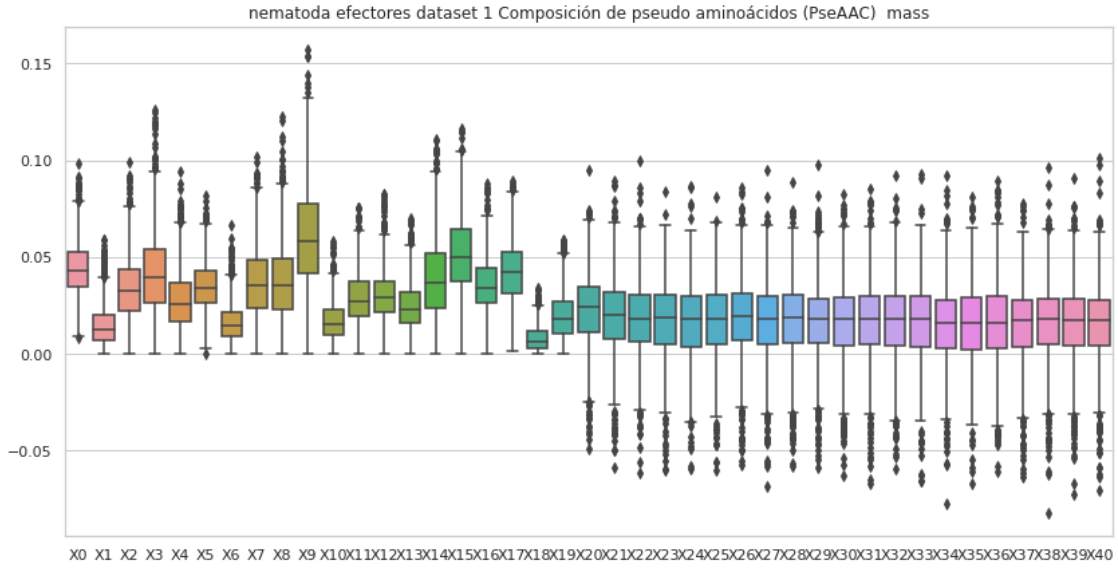
```
75%      0.052653    0.019955    0.044021    0.053900    0.037025    0.042735
max      0.098274    0.059060    0.098923    0.125896    0.094080    0.082057

               X6          X7          X8          X9  …         X31  \
count  818.000000  818.000000  818.000000  818.000000  …  818.000000
mean     0.016388    0.037073    0.037889    0.060045  …    0.017161
std      0.010408    0.018431    0.020061    0.026539  …    0.020996
min      0.000000    0.000000    0.000000    0.000000  …   -0.067053
25%      0.008986    0.023415    0.023217    0.041353  …    0.005080
50%      0.014574    0.035432    0.035650    0.057967  …    0.018359
75%      0.021704    0.048367    0.049338    0.077628  …    0.029745
max      0.066651    0.101836    0.122403    0.157184  …    0.085280

               X32         X33         X34         X35         X36         X37  \
count  818.000000  818.000000  818.000000  818.000000  818.000000  818.000000
mean     0.016516    0.016169    0.014802    0.015404    0.015961    0.015357
std      0.021227    0.021088    0.021309    0.021186    0.022371    0.020878
min     -0.059536   -0.065500   -0.077235   -0.067372   -0.060953   -0.063228
25%      0.004294    0.003786    0.002995    0.002173    0.002933    0.003560
50%      0.017875    0.017797    0.015828    0.016160    0.016233    0.017230
75%      0.029970    0.030060    0.027982    0.029159    0.029759    0.028088
max      0.092359    0.093102    0.092280    0.081037    0.089184    0.077787

               X38         X39         X40
count  818.000000  818.000000  818.000000
mean     0.016346    0.015814    0.015930
std      0.020885    0.021361    0.021556
min     -0.082262   -0.072314   -0.070489
25%      0.004736    0.004442    0.004399
50%      0.018272    0.017399    0.017686
75%      0.028649    0.028364    0.028139
max      0.096349    0.090844    0.101247

[8 rows x 41 columns]
```

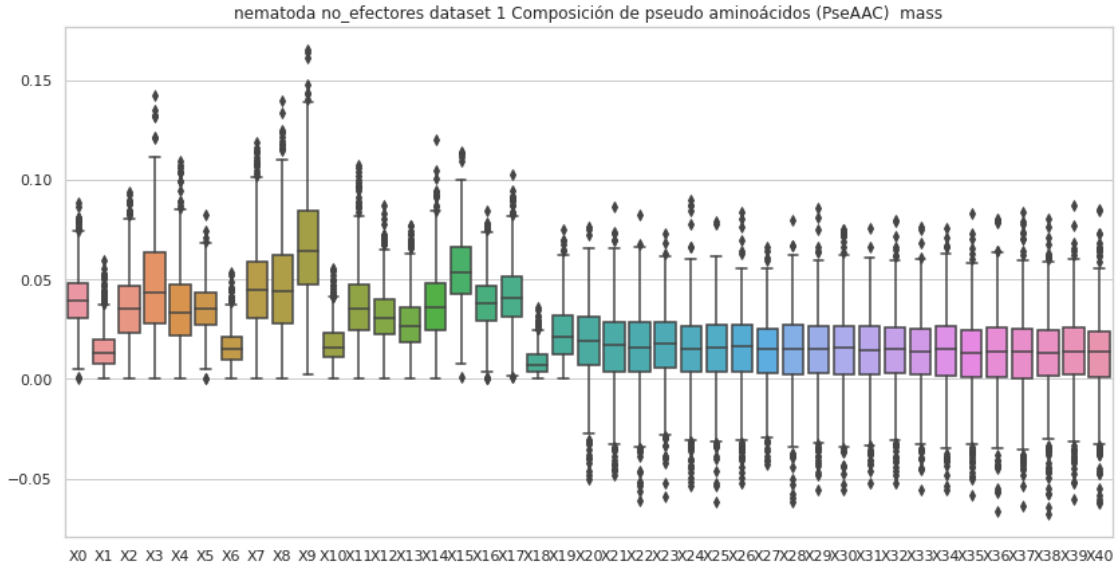
Composición de pseudo aminoácidos (PseAAC) mass no_efectores nematoda dataset 1, sin valores atípicos.
Valores del documento csv.

```
          X0        X1        X2        X3        X4        X5        X6  \
0    0.044418  0.016983  0.047031  0.056175  0.043111  0.023515  0.039192
1    0.044979  0.026988  0.023989  0.011994  0.083962  0.032985  0.017992
2    0.053661  0.010319  0.032334  0.055724  0.042653  0.041277  0.015823
3    0.036320  0.001730  0.025943  0.038050  0.003459  0.005189  0.000000
4    0.029859  0.021716  0.035288  0.032573  0.040717  0.031216  0.008143
..        …         …         …         …         …         …         …
```

```
994  0.058488  0.025995  0.064986  0.038992  0.045491  0.064986  0.025995
996  0.034360  0.007636  0.025770  0.020998  0.016226  0.035315  0.014317
997  0.042075  0.015300  0.038250  0.047812  0.022950  0.043987  0.015300
998  0.021937  0.015955  0.023932  0.017949  0.023932  0.035898  0.012963
999  0.037416  0.018708  0.024053  0.021381  0.037416  0.042762  0.000000

           X7        X8        X9  …       X32       X33       X34  \
0    0.067933  0.040499  0.097980  …  0.011240  0.002346  0.013670
1    0.080963  0.035984  0.143934  …  0.015118 -0.006434  0.026619
2    0.049533  0.047469  0.076363  …  0.011222  0.020248  0.024191
3    0.005189  0.060534  0.012107  …  0.028190  0.035078  0.024535
4    0.024430  0.035288  0.051575  …  0.018808  0.019919  0.017405
..        …         …         …  …       …         …         …
994  0.048740  0.068236  0.058488  …  0.033482  0.030122 -0.026405
996  0.041042  0.028634  0.030543  …  0.013756  0.016528  0.010018
997  0.036337  0.043987  0.068849  …  0.013949  0.000477  0.016541
998  0.013960  0.015955  0.032906  …  0.038736  0.030882  0.034561
999  0.053452  0.018708  0.085523  …  0.010939 -0.008883  0.014012

           X35       X36       X37       X38       X39       X40           X41
0    0.022872  0.015902 -0.007222  0.011709  0.007796 -0.008758  no_efectores
1   -0.006747  0.012018  0.015650  0.007453  0.020864 -0.005709  no_efectores
2    0.017334 -0.007309  0.026904  0.017661  0.006462  0.023590  no_efectores
3    0.026828  0.027208  0.027548  0.027591  0.024662  0.026380  no_efectores
4    0.001548  0.018006  0.012648  0.015150  0.018069 -0.000535  no_efectores
..        …         …         …         …         …         …           …
994  0.007060  0.011770  0.031444  0.000385 -0.025558  0.021216  no_efectores
996  0.034235  0.034217  0.026785  0.036400  0.031564  0.013601  no_efectores
997 -0.004361  0.019700  0.018372  0.019506  0.032118  0.003970  no_efectores
998  0.027489  0.018041  0.031967  0.044065  0.034207  0.036242  no_efectores
999  0.002857  0.026225 -0.019339 -0.017662  0.041398  0.000589  no_efectores

[812 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  812.000000  812.000000  812.000000  812.000000  812.000000  812.000000
mean     0.040267    0.014806    0.036020    0.046530    0.036518    0.035986
std      0.014411    0.010526    0.017370    0.025204    0.019715    0.012312
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.030517    0.007471    0.023305    0.028048    0.022074    0.027367
50%      0.039677    0.012914    0.035039    0.043336    0.033476    0.035574
75%      0.048185    0.019554    0.046569    0.063289    0.047520    0.043700
max      0.088712    0.059447    0.093929    0.142384    0.109473    0.082514
```

```
              X6          X7          X8          X9  …         X31  \
count  812.000000  812.000000  812.000000  812.000000  …  812.000000
mean     0.015769    0.046595    0.047091    0.066338  …    0.013728
std      0.008856    0.022038    0.024957    0.027650  …    0.018645
min      0.000000    0.000000    0.000000    0.002518  …   -0.052295
25%      0.009608    0.030834    0.027970    0.047127  …    0.002451
50%      0.015108    0.044545    0.044395    0.064322  …    0.014433
75%      0.020849    0.059090    0.062418    0.084198  …    0.026298
max      0.053596    0.118623    0.139674    0.165264  …    0.075499

              X32         X33         X34         X35         X36         X37  \
count  812.000000  812.000000  812.000000  812.000000  812.000000  812.000000
mean     0.013854    0.013314    0.013335    0.011685    0.013092    0.012423
std      0.019168    0.018577    0.019353    0.018958    0.019853    0.019954
min     -0.052469   -0.055482   -0.055844   -0.058637   -0.066384   -0.063706
25%      0.003123    0.002092    0.001745    0.000967    0.001229    0.000352
50%      0.014777    0.013960    0.015297    0.013067    0.013881    0.013794
75%      0.026071    0.025552    0.026344    0.024231    0.026172    0.025103
max      0.079673    0.076596    0.075402    0.083251    0.080202    0.083498

              X38         X39         X40
count  812.000000  812.000000  812.000000
mean     0.012513    0.013041    0.011759
std      0.019497    0.019171    0.019986
min     -0.067735   -0.060038   -0.062444
25%      0.001543    0.002495    0.000944
50%      0.013416    0.013554    0.013600
75%      0.024541    0.026125    0.023603
max      0.080213    0.086878    0.085177

[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    0.047438  0.051087  0.032842  0.069333  0.040140  0.025544  0.010947
1    0.031221  0.018076  0.026292  0.018076  0.019719  0.031221  0.021362
2    0.066801  0.016700  0.070141  0.086841  0.040081  0.040081  0.020040
3    0.050547  0.025274  0.050547  0.058589  0.037910  0.055142  0.016083
4    0.056767  0.060972  0.052562  0.044152  0.060972  0.063074  0.031537
..        …         …         …         …         …         …         …
995  0.016902  0.012073  0.025353  0.024146  0.009658  0.010866  0.009658
996  0.060496  0.020165  0.053774  0.060496  0.053774  0.057135  0.033609
997  0.046567  0.034926  0.040746  0.064030  0.017463  0.046567  0.011642
998  0.018992  0.033236  0.004748  0.018992  0.033236  0.037984  0.009496
999  0.028427  0.006689  0.013377  0.008361  0.023410  0.020066  0.003344

           X7        X8        X9  …       X53       X54       X55  \
0    0.025544  0.062035  0.076631  …  0.049744  0.018666  0.054729
1    0.011503  0.021362  0.044367  …  0.035516  0.028507  0.021577
2    0.060121  0.036741  0.116902  … -0.007083  0.043157  0.055618
```

```
3    0.027571  0.048249  0.081565  …  0.001456 -0.016140 -0.005577
4    0.071484  0.033639  0.088304  … -0.000746 -0.022340 -0.013037
..        …         …         …  …         …         …
995  0.010866  0.033804  0.026561  …  0.012370  0.015061  0.021423
996  0.053774  0.060496  0.114269  …  0.078236  0.016585  0.049369
997  0.034926  0.017463  0.052388  … -0.016453  0.019008 -0.025247
998  0.047481  0.014244  0.037984  … -0.011660 -0.035196 -0.019995
999  0.030099  0.006689  0.030099  …  0.017446  0.025319  0.000680

          X56       X57       X58       X59       X60       X61        X62
0   -0.005097  0.002745 -0.037887 -0.018809 -0.003196  0.027714  efectores
1    0.003671  0.007731  0.066035  0.041662  0.018281  0.027929  efectores
2   -0.007193  0.043009 -0.032754 -0.045594  0.006854  0.060562  efectores
3    0.027959  0.039802 -0.011806 -0.005052 -0.011240  0.014501  efectores
4    0.008253  0.016811  0.012878  0.019240 -0.002227  0.020552  efectores
..        …         …         …         …         …         …
995  0.003014  0.019526  0.010873  0.013004  0.018689  0.025458  efectores
996 -0.058493 -0.052660  0.051063 -0.016672 -0.080989 -0.029660  efectores
997 -0.029064  0.012728  0.005366  0.016827  0.067373  0.047404  efectores
998  0.028565  0.026430 -0.008363  0.013886 -0.017107 -0.033160  efectores
999  0.042602  0.007063  0.004262 -0.008695  0.032753  0.003008  efectores

[1000 rows x 63 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.054381     0.020949     0.040097     0.046252     0.034140
std       0.054247     0.039705     0.027072     0.039476     0.031764
min      -0.267192     0.000000     0.000000    -0.267192    -0.133596
25%       0.030655     0.006667     0.022475     0.028477     0.017376
50%       0.046328     0.014154     0.036931     0.042679     0.028756
75%       0.065607     0.025044     0.051247     0.059157     0.044042
max       1.171166     0.780777     0.260259     0.910907     0.520518

                X5           X6           X7           X8           X9  …  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
mean      0.047057     0.020503     0.044331     0.043864     0.071089  …
std       0.055607     0.023459     0.046606     0.048650     0.063808  …
min      -0.133596     0.000000    -0.267192    -0.400788    -0.667980  …
25%       0.022077     0.008641     0.023793     0.024994     0.040005  …
50%       0.037475     0.016070     0.037324     0.038396     0.062930  …
75%       0.056983     0.025861     0.055928     0.054339     0.089765  …
max       1.122703     0.374234     0.748469     1.194865     1.041036  …
```

```
              X52          X53          X54          X55          X56  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean     -0.002415     0.003634     0.003962     0.007096    -0.000116
std       0.089972     0.073487     0.076810     0.066398     0.067539
min      -1.917188    -1.401903    -0.743122    -0.992239    -1.143493
25%      -0.015581    -0.008834    -0.014790    -0.007814    -0.016171
50%       0.002391     0.009908     0.005465     0.009376     0.004393
75%       0.020539     0.025515     0.021113     0.025409     0.021130
max       0.450916     0.360520     1.238046     1.077847     0.446365

              X57          X58          X59          X60          X61
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.005685    -0.001202     0.006533     0.004319     0.007737
std       0.054364     0.091441     0.086340     0.067229     0.048888
min      -0.927038    -1.614761    -1.606986    -0.806997    -0.720980
25%      -0.006547    -0.016349    -0.006531    -0.013550    -0.007413
50%       0.010029     0.004279     0.010350     0.005393     0.009975
75%       0.024832     0.020069     0.024937     0.021481     0.025114
max       0.224627     1.579746     1.128004     1.154058     0.685938

[8 rows x 62 columns]
```

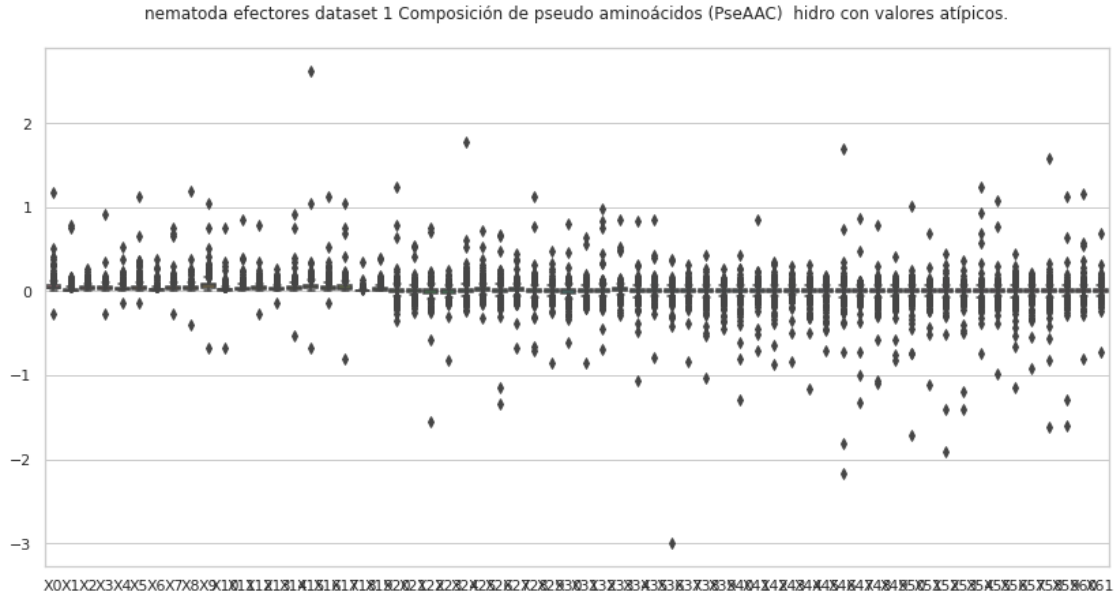
```
no_efectores
Composición de pseudo aminoácidos (PseAAC)  hidro no_efectores nematoda dataset
1, con valores atípicos.
Valores del documento csv.

          X0        X1        X2        X3        X4        X5        X6  \
0    0.039144  0.014967  0.041446  0.049505  0.037993  0.020723  0.034539
1    0.020033  0.012020  0.010684  0.005342  0.037394  0.014691  0.008013
2    0.060448  0.011625  0.036424  0.062773  0.048048  0.046498  0.017824
3    0.017129  0.000816  0.012235  0.017945  0.001631  0.002447  0.000000
4    0.025050  0.018218  0.029605  0.027327  0.034159  0.026189  0.006832
..        …         …         …         …         …         …         …
995  0.024809  0.016539  0.041348  0.066157  0.000000  0.066157  0.016539
996  0.073077  0.016239  0.054808  0.044658  0.034509  0.075107  0.030449
997  0.044691  0.016251  0.040628  0.050785  0.024377  0.046722  0.016251
998  0.045299  0.032945  0.049417  0.037063  0.049417  0.074126  0.026768
999  0.048572  0.024286  0.031225  0.027756  0.048572  0.055511  0.000000

          X7        X8        X9  …       X53       X54       X55  \
0    0.059867  0.035690  0.086347  … -0.004810  0.011998  0.011158
1    0.036059  0.016026  0.064104  …  0.018758  0.027402  0.014377
2    0.055798  0.053473  0.086022  …  0.011375  0.018788 -0.000442
3    0.002447  0.028549  0.005710  …  0.028716  0.007513  0.028755
```

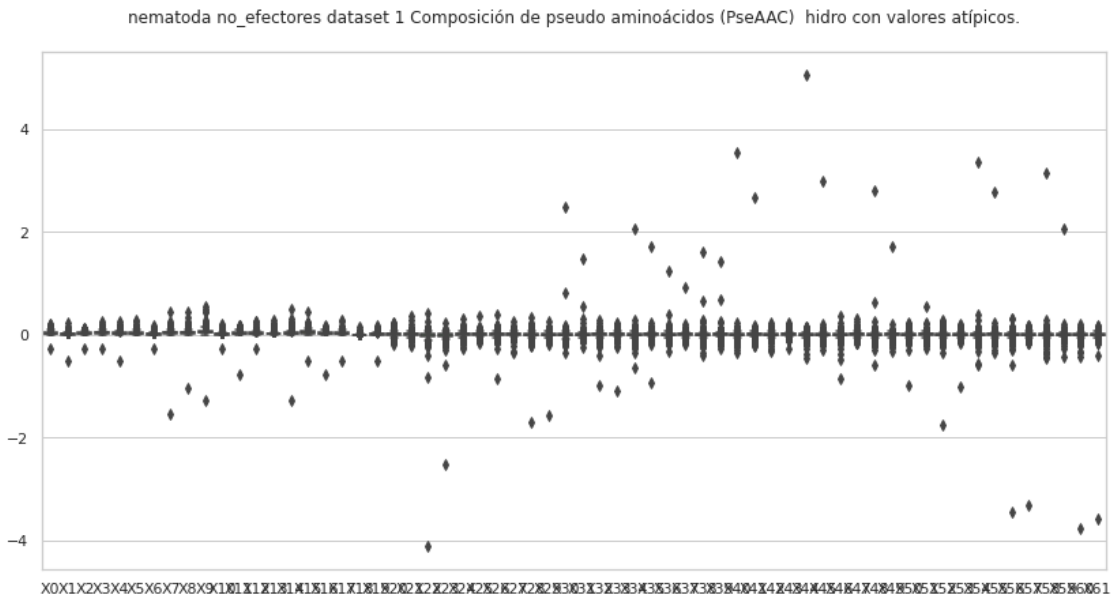
```
4    0.020496  0.029605  0.043268  … -0.003505 -0.001872  0.014656
..        …         …         …  …         …         …
995  0.049617  0.049617  0.082696  … -0.004177 -0.001503  0.016168
996  0.087287  0.060898  0.064957  …  0.032675  0.011854 -0.013190
997  0.038597  0.046722  0.073130  …  0.038928  0.032462  0.034395
998  0.028827  0.032945  0.067949  …  0.034055 -0.023772 -0.018708
999  0.069389  0.024286  0.111022  … -0.043421  0.003734 -0.015310

          X56       X57       X58       X59       X60       X61           X62
0    0.024101  0.025974 -0.006688  0.025160  0.011136  0.008236  no_efectores
1    0.027019  0.005160  0.004359 -0.011078  0.002159  0.007024  no_efectores
2   -0.007766 -0.001948  0.026455  0.031494  0.003339  0.007343  no_efectores
3    0.010063  0.034745  0.010822  0.028461  0.010071  0.026123  no_efectores
4    0.002811  0.012448 -0.002809 -0.001146  0.006609 -0.004399  no_efectores
..        …         …         …         …         …         …           …
995  0.109335  0.080639  0.050024 -0.006587 -0.012601 -0.009547  no_efectores
996  0.035018  0.016062  0.029111  0.005682 -0.008088 -0.012511  no_efectores
997 -0.012880 -0.002777  0.041093  0.028301  0.036820  0.033834  no_efectores
998  0.020728  0.021038  0.025156  0.017004 -0.034247 -0.016509  no_efectores
999  0.025263 -0.019612 -0.022721 -0.060213  0.032164 -0.018872  no_efectores

[1000 rows x 63 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.039326     0.015926     0.034260     0.043293     0.036320
std       0.026978     0.025155     0.023418     0.027748     0.032196
min      -0.255403    -0.510805    -0.255403    -0.255403    -0.510805
25%       0.022293     0.005999     0.018257     0.025716     0.019744
50%       0.034230     0.011785     0.031365     0.041315     0.031378
75%       0.051276     0.020949     0.047042     0.057487     0.047225
max       0.201897     0.246729     0.134619     0.257156     0.259887

                X5           X6           X7           X8           X9  …  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
mean      0.037643     0.016363     0.043660     0.045007     0.064644  …
std       0.028575     0.018720     0.058936     0.047181     0.063041  …
min       0.000000    -0.255403    -1.532415    -1.021610    -1.277013  …
25%       0.019182     0.007017     0.025513     0.025334     0.038178  …
50%       0.031060     0.012869     0.039304     0.040426     0.056087  …
75%       0.049689     0.021695     0.057357     0.059854     0.082597  …
max       0.296075     0.191085     0.450023     0.450023     0.555598  …
```

```
                X52           X53           X54           X55           X56  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.001508     0.008304     0.004992     0.011057    -0.003466
std       0.068454     0.044871     0.116132     0.095163     0.117505
min      -1.753785    -1.003416    -0.573155    -0.341421    -3.430838
25%      -0.009923    -0.003407    -0.010306    -0.003541    -0.012895
50%       0.005104     0.010253     0.006150     0.011214     0.003731
75%       0.017474     0.023081     0.018254     0.023733     0.016876
max       0.282202     0.216635     3.345990     2.779923     0.318427

                X57           X58           X59           X60           X61
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.004435     0.003909     0.009697    -0.001396     0.006475
std       0.109510     0.110363     0.073647     0.125611     0.118938
min      -3.298267    -0.447361    -0.415895    -3.749973    -3.581893
25%      -0.005925    -0.009951    -0.003250    -0.010420    -0.004480
50%       0.009701     0.005317     0.009914     0.005312     0.011000
75%       0.022373     0.018504     0.023784     0.019294     0.025611
max       0.196159     3.153662     2.057299     0.222684     0.171314

[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    0.047438  0.051087  0.032842  0.069333  0.040140  0.025544  0.010947
1    0.031221  0.018076  0.026292  0.018076  0.019719  0.031221  0.021362
2    0.066801  0.016700  0.070141  0.086841  0.040081  0.040081  0.020040
3    0.050547  0.025274  0.050547  0.058589  0.037910  0.055142  0.016083
4    0.056767  0.060972  0.052562  0.044152  0.060972  0.063074  0.031537
..        …         …         …         …         …         …         …
995  0.016902  0.012073  0.025353  0.024146  0.009658  0.010866  0.009658
996  0.060496  0.020165  0.053774  0.060496  0.053774  0.057135  0.033609
997  0.046567  0.034926  0.040746  0.064030  0.017463  0.046567  0.011642
998  0.018992  0.033236  0.004748  0.018992  0.033236  0.037984  0.009496
999  0.028427  0.006689  0.013377  0.008361  0.023410  0.020066  0.003344

           X7        X8        X9  …       X53       X54       X55  \
0    0.025544  0.062035  0.076631  …  0.049744  0.018666  0.054729
1    0.011503  0.021362  0.044367  …  0.035516  0.028507  0.021577
2    0.060121  0.036741  0.116902  … -0.007083  0.043157  0.055618
3    0.027571  0.048249  0.081565  …  0.001456 -0.016140 -0.005577
4    0.071484  0.033639  0.088304  … -0.000746 -0.022340 -0.013037
..        …         …         …  …       …         …         …
```

```
995  0.010866  0.033804  0.026561  …  0.012370  0.015061  0.021423
996  0.053774  0.060496  0.114269  …  0.078236  0.016585  0.049369
997  0.034926  0.017463  0.052388  … -0.016453  0.019008 -0.025247
998  0.047481  0.014244  0.037984  … -0.011660 -0.035196 -0.019995
999  0.030099  0.006689  0.030099  …  0.017446  0.025319  0.000680

          X56       X57       X58       X59       X60       X61        X62
0   -0.005097  0.002745 -0.037887 -0.018809 -0.003196  0.027714  efectores
1    0.003671  0.007731  0.066035  0.041662  0.018281  0.027929  efectores
2   -0.007193  0.043009 -0.032754 -0.045594  0.006854  0.060562  efectores
3    0.027959  0.039802 -0.011806 -0.005052 -0.011240  0.014501  efectores
4    0.008253  0.016811  0.012878  0.019240 -0.002227  0.020552  efectores
..        …         …         …         …         …         …
995  0.003014  0.019526  0.010873  0.013004  0.018689  0.025458  efectores
996 -0.058493 -0.052660  0.051063 -0.016672 -0.080989 -0.029660  efectores
997 -0.029064  0.012728  0.005366  0.016827  0.067373  0.047404  efectores
998  0.028565  0.026430 -0.008363  0.013886 -0.017107 -0.033160  efectores
999  0.042602  0.007063  0.004262 -0.008695  0.032753  0.003008  efectores

[920 rows x 63 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  920.000000  920.000000  920.000000  920.000000  920.000000  920.000000
mean     0.048889    0.017301    0.036719    0.043562    0.030508    0.040829
std      0.026697    0.016448    0.020534    0.021707    0.018524    0.026503
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.030135    0.006337    0.021659    0.028126    0.016912    0.021375
50%      0.044121    0.013358    0.035291    0.042302    0.027564    0.035454
75%      0.062244    0.023549    0.048628    0.056534    0.041268    0.053755
max      0.201396    0.111482    0.113210    0.132059    0.107492    0.198777

                X6          X7          X8          X9  …         X52  \
count  920.000000  920.000000  920.000000  920.000000  …  920.000000
mean     0.018035    0.040030    0.040254    0.064250  …    0.001289
std      0.012726    0.023305    0.021435    0.034196  …    0.033441
min      0.000000    0.000000    0.000000    0.000000  …   -0.195140
25%      0.008538    0.023451    0.024427    0.038769  …   -0.013811
50%      0.015616    0.036536    0.037285    0.060210  …    0.002820
75%      0.024906    0.052696    0.052115    0.084769  …    0.019136
max      0.074850    0.175095    0.128002    0.253110  …    0.143462

               X53         X54         X55         X56         X57         X58  \
count  920.000000  920.000000  920.000000  920.000000  920.000000  920.000000
```

mean	0.007922	0.003609	0.008325	0.003667	0.009501	0.002295
std	0.027553	0.034743	0.029403	0.034936	0.028746	0.033134
min	-0.106391	-0.171583	-0.189195	-0.153501	-0.126343	-0.222515
25%	-0.007731	-0.010810	-0.006324	-0.012937	-0.004486	-0.013687
50%	0.010130	0.006175	0.009949	0.005047	0.010614	0.005119
75%	0.025046	0.020740	0.025135	0.020617	0.024773	0.019227
max	0.108536	0.115478	0.112665	0.193949	0.146557	0.137804

	X59	X60	X61
count	920.000000	920.000000	920.000000
mean	0.009954	0.003466	0.008184
std	0.027176	0.032932	0.028135
min	-0.122538	-0.149113	-0.120238
25%	-0.004176	-0.011866	-0.005484
50%	0.010809	0.005393	0.010327
75%	0.024561	0.020348	0.024826
max	0.111990	0.163378	0.124489

[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	0.039144	0.014967	0.041446	0.049505	0.037993	0.020723	0.034539	
1	0.020033	0.012020	0.010684	0.005342	0.037394	0.014691	0.008013	
2	0.060448	0.011625	0.036424	0.062773	0.048048	0.046498	0.017824	
3	0.017129	0.000816	0.012235	0.017945	0.001631	0.002447	0.000000	
4	0.025050	0.018218	0.029605	0.027327	0.034159	0.026189	0.006832	
..	…	…	…	…	…	…	…	
995	0.024809	0.016539	0.041348	0.066157	0.000000	0.066157	0.016539	
996	0.073077	0.016239	0.054808	0.044658	0.034509	0.075107	0.030449	
997	0.044691	0.016251	0.040628	0.050785	0.024377	0.046722	0.016251	
998	0.045299	0.032945	0.049417	0.037063	0.049417	0.074126	0.026768	
999	0.048572	0.024286	0.031225	0.027756	0.048572	0.055511	0.000000	

	X7	X8	X9	…	X53	X54	X55	\
0	0.059867	0.035690	0.086347	…	-0.004810	0.011998	0.011158	
1	0.036059	0.016026	0.064104	…	0.018758	0.027402	0.014377	
2	0.055798	0.053473	0.086022	…	0.011375	0.018788	-0.000442	
3	0.002447	0.028549	0.005710	…	0.028716	0.007513	0.028755	
4	0.020496	0.029605	0.043268	…	-0.003505	-0.001872	0.014656	
..	…	…	…	…	…	…	…	
995	0.049617	0.049617	0.082696	…	-0.004177	-0.001503	0.016168	

```
996  0.087287  0.060898  0.064957  …  0.032675  0.011854 -0.013190
997  0.038597  0.046722  0.073130  …  0.038928  0.032462  0.034395
998  0.028827  0.032945  0.067949  …  0.034055 -0.023772 -0.018708
999  0.069389  0.024286  0.111022  … -0.043421  0.003734 -0.015310

          X56       X57       X58       X59       X60       X61           X62
0    0.024101  0.025974 -0.006688  0.025160  0.011136  0.008236  no_efectores
1    0.027019  0.005160  0.004359 -0.011078  0.002159  0.007024  no_efectores
2   -0.007766 -0.001948  0.026455  0.031494  0.003339  0.007343  no_efectores
3    0.010063  0.034745  0.010822  0.028461  0.010071  0.026123  no_efectores
4    0.002811  0.012448 -0.002809 -0.001146  0.006609 -0.004399  no_efectores
..        …         …         …         …         …         …           …
995  0.109335  0.080639  0.050024 -0.006587 -0.012601 -0.009547  no_efectores
996  0.035018  0.016062  0.029111  0.005682 -0.008088 -0.012511  no_efectores
997 -0.012880 -0.002777  0.041093  0.028301  0.036820  0.033834  no_efectores
998  0.020728  0.021038  0.025156  0.017004 -0.034247 -0.016509  no_efectores
999  0.025263 -0.019612 -0.022721 -0.060213  0.032164 -0.018872  no_efectores

[888 rows x 63 columns]
```

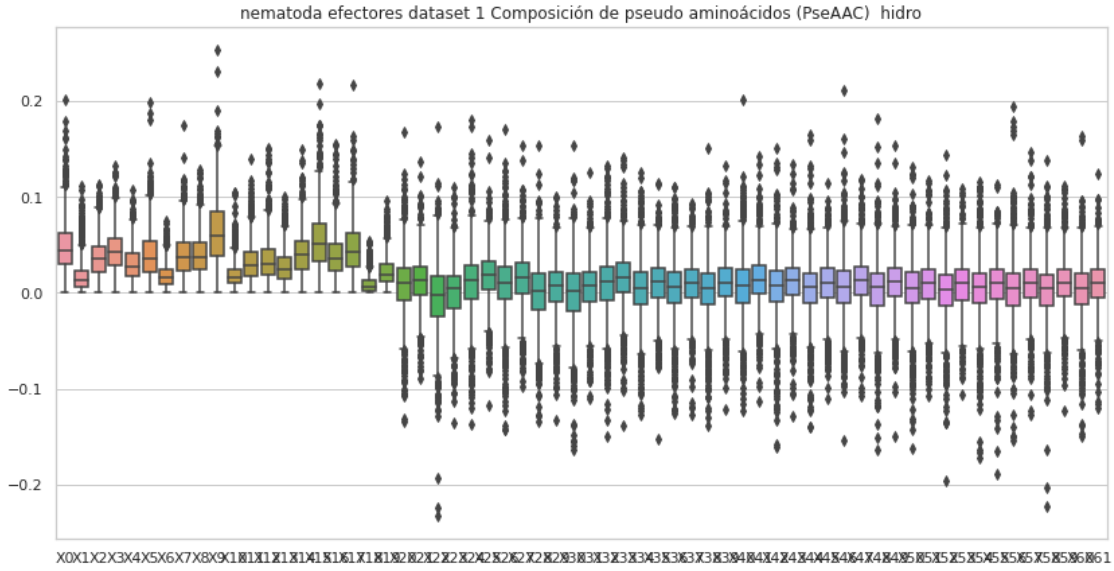
Composición de pseudo aminoácidos (PseAAC) no_efectores nematoda dataset 1, sin valores atípicos.
Estadísticas.

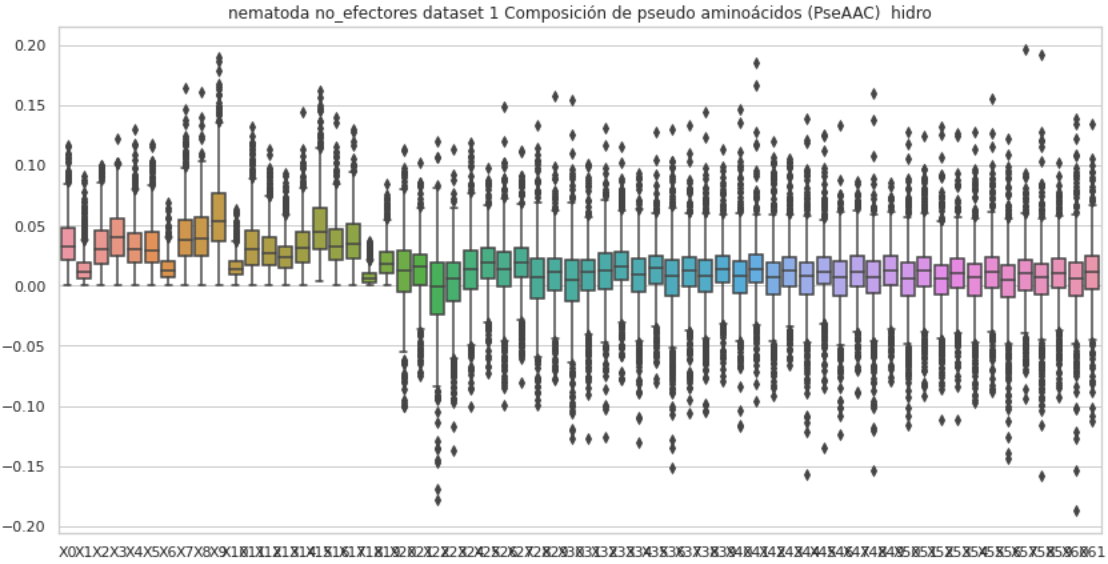
```
                X0          X1          X2          X3          X4          X5  \
count  888.000000  888.000000  888.000000  888.000000  888.000000  888.000000
mean     0.036634    0.014190    0.032841    0.040704    0.032722    0.033537
std      0.020347    0.012895    0.019465    0.021256    0.019831    0.020481
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.021848    0.005862    0.018008    0.024702    0.019090    0.018633
50%      0.032726    0.011082    0.030110    0.039988    0.029961    0.029076
75%      0.047611    0.018577    0.045314    0.055144    0.043579    0.044627
max      0.116877    0.090904    0.099823    0.122326    0.129396    0.117345

                X6          X7          X8          X9  …         X52  \
count  888.000000  888.000000  888.000000  888.000000  …  888.000000
mean     0.014377    0.041974    0.042232    0.059459  …    0.004781
std      0.010205    0.024275    0.023170    0.031632  …    0.025068
min      0.000000    0.000000    0.000000    0.000000  …   -0.111047
25%      0.006829    0.024784    0.024731    0.037142  …   -0.007742
50%      0.012355    0.037713    0.038885    0.053383  …    0.005630
75%      0.020056    0.054375    0.056896    0.076572  …    0.016896
max      0.068440    0.164257    0.160590    0.189462  …    0.132559

               X53         X54         X55         X56         X57         X58  \
count  888.000000  888.000000  888.000000  888.000000  888.000000  888.000000
mean     0.009988    0.004323    0.010695    0.001730    0.008579    0.004601
```

std	0.023134	0.026162	0.024787	0.028659	0.024914	0.028364
min	-0.111885	-0.097211	-0.088739	-0.143092	-0.093666	-0.158183
25%	-0.002118	-0.008782	-0.001913	-0.009826	-0.003817	-0.007332
50%	0.010710	0.006454	0.011771	0.004270	0.010251	0.006404
75%	0.022583	0.017603	0.023599	0.016375	0.021787	0.018275
max	0.126706	0.128149	0.155352	0.122525	0.196159	0.191279

	X59	X60	X61
count	888.000000	888.000000	888.000000
mean	0.009019	0.004379	0.010493
std	0.024277	0.028806	0.025523
min	-0.092376	-0.186778	-0.113013
25%	-0.001583	-0.008369	-0.003422
50%	0.010037	0.005821	0.011669
75%	0.022565	0.018602	0.025095
max	0.102501	0.138834	0.133894

[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))
```

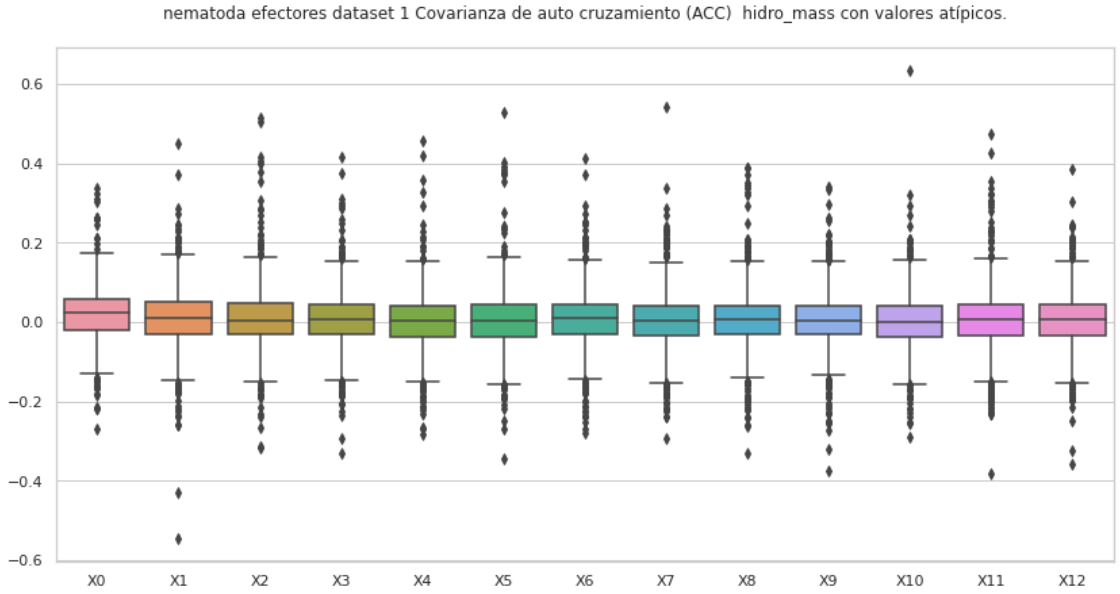
```
efectores
Covarianza de auto cruzamiento (ACC)  hidro_mass efectores nematoda dataset 1,
con valores atípicos.
Valores del documento csv.

          X0        X1        X2        X3        X4        X5        X6  \
0    0.019625 -0.001848 -0.008108  0.016969  0.037711  0.013584  0.017178
1    0.038273  0.006226  0.081118  0.054531 -0.004516 -0.035899  0.086913
2    0.109991 -0.015873 -0.035930 -0.017213 -0.077442 -0.060394 -0.005183
3   -0.011817 -0.029095 -0.039041  0.000039 -0.013657 -0.036385 -0.018370
4    0.045998  0.014915 -0.020015 -0.018419 -0.047657 -0.001625 -0.063503
..        …         …         …         …         …         …         …
995  0.036302 -0.006523  0.052501  0.002684 -0.023637 -0.054717 -0.030591
996 -0.019978  0.053566 -0.046486  0.043441 -0.048772 -0.019229  0.094234
997  0.043840 -0.059896  0.007442 -0.071917 -0.079505  0.053620  0.212745
998  0.114792  0.053308  0.003530 -0.116544  0.040821  0.063930  0.065471
999  0.060660 -0.058119 -0.021007  0.083183 -0.022137 -0.067753  0.016109

          X7        X8        X9       X10       X11       X12         X13
0    0.036823 -0.038163  0.024723  0.036541 -0.060687  0.012481  efectores
1    0.002178 -0.036401 -0.014205 -0.036856 -0.029526 -0.030152  efectores
2    0.059509  0.105795  0.079493 -0.034332 -0.009044 -0.019448  efectores
3   -0.030519  0.019919  0.036059  0.039881  0.045897 -0.030677  efectores
4   -0.010114  0.011296 -0.022164  0.020798  0.045351 -0.073713  efectores
..        …         …         …         …         …         …         …
995 -0.021354 -0.038319  0.020939 -0.051076 -0.023962  0.003693  efectores
996 -0.111275  0.048079  0.014268  0.028765 -0.028096 -0.037747  efectores
997 -0.002601 -0.135568 -0.052740 -0.134199 -0.015835 -0.167305  efectores
998  0.118267 -0.047932 -0.070228  0.045694  0.101246  0.033825  efectores
999 -0.022026 -0.037645 -0.007081 -0.021282 -0.091118 -0.065639  efectores

[1000 rows x 14 columns]


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

                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.020739     0.009894     0.009366     0.006711     0.003008
```

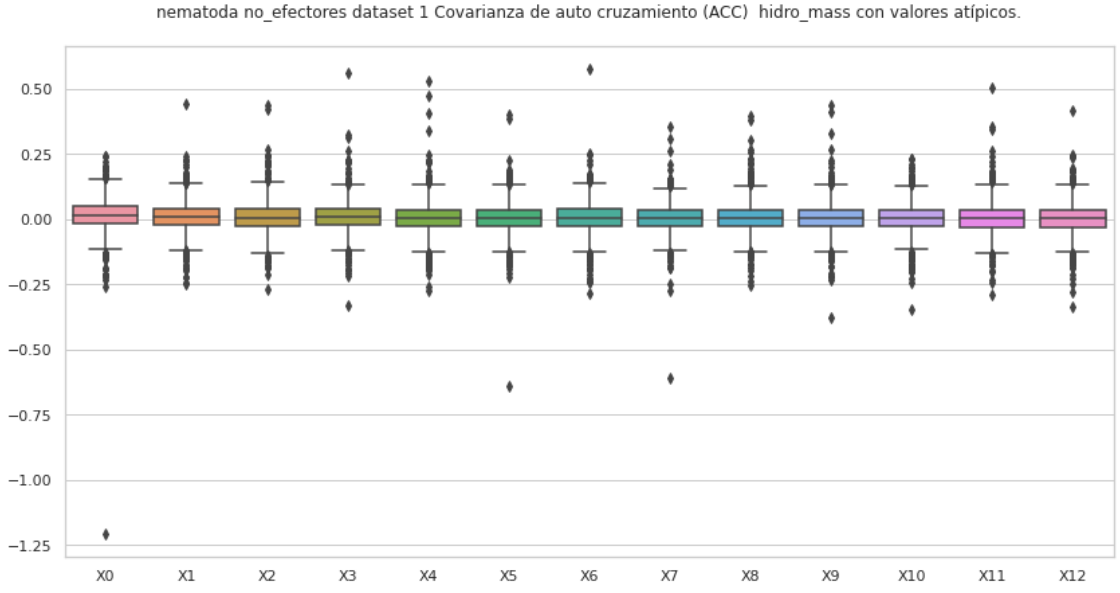
```
std       0.069948     0.076227     0.077688     0.070445     0.070899
min      -0.267594    -0.545313    -0.315913    -0.329086    -0.282758
25%      -0.019785    -0.029117    -0.032091    -0.031932    -0.035989
50%       0.023906     0.011160     0.003319     0.007803     0.002669
75%       0.057970     0.050888     0.047875     0.045440     0.042105
max       0.338158     0.450677     0.515261     0.416225     0.457174

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.005562     0.007968     0.004163     0.006573     0.004508
std       0.076218     0.072424     0.074537     0.075388     0.071772
min      -0.342657    -0.281087    -0.294338    -0.330694    -0.374668
25%      -0.035776    -0.029610    -0.035656    -0.031854    -0.031202
50%       0.004332     0.009656     0.004903     0.007112     0.004294
75%       0.044823     0.045614     0.042544     0.042315     0.042468
max       0.528023     0.411799     0.543545     0.390329     0.340530

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.001420     0.006962     0.003540
std       0.072039     0.079497     0.071830
min      -0.291168    -0.381983    -0.358366
25%      -0.038155    -0.032640    -0.034902
50%      -0.000428     0.005645     0.005406
75%       0.039916     0.045923     0.043169
max       0.633831     0.473414     0.384113



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    0.017580  0.022179  0.027164  0.017104  0.001275 -0.022052 -0.003837
1    0.009096  0.013476 -0.100321  0.023707 -0.023600 -0.039430 -0.078163
2    0.054927  0.015489  0.008992 -0.008384  0.012210  0.014689 -0.020932
3    0.028351  0.024664  0.076025  0.049284  0.051564 -0.019353  0.051728
4    0.044265  0.051431  0.077956 -0.031474 -0.006130 -0.030696  0.017136
..        …         …         …         …         …         …         …
995  0.102852 -0.068763  0.044283  0.129520 -0.041985 -0.020252 -0.195257
996 -0.045558  0.078998 -0.040740 -0.053890  0.005129  0.028930  0.001145
997 -0.004655 -0.070958 -0.049885  0.017624  0.064193  0.034228 -0.027500
998 -0.039260  0.024520  0.029100  0.059829 -0.076523  0.037623  0.043185
999  0.078712  0.066046  0.146760  0.045262  0.028705  0.013218 -0.007974

           X7        X8        X9       X10       X11       X12           X13
```

```
0   -0.052146 -0.011817 -0.005375 -0.021543  0.027961  0.012654  no_efectores
1   -0.023510  0.087389 -0.012996  0.025569 -0.014514  0.020138  no_efectores
2   -0.003877 -0.003104  0.024400  0.019663 -0.025867  0.010056  no_efectores
3   -0.037745 -0.000412  0.052082 -0.024103 -0.021581  0.003359  no_efectores
4    0.022518  0.082939  0.096112  0.039332  0.040911  0.017739  no_efectores
..        …         …         …         …         …         …           …
995 -0.273295  0.062250  0.034087 -0.046227 -0.089696  0.125439  no_efectores
996  0.014863  0.076227  0.052230  0.047585 -0.015317 -0.048790  no_efectores
997  0.049648  0.029439 -0.021985  0.039596  0.020699  0.000186  no_efectores
998 -0.064152  0.020724 -0.004462  0.018598 -0.024517  0.071289  no_efectores
999  0.060898  0.138351 -0.086351  0.069640  0.029354 -0.012334  no_efectores

[1000 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  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.013700     0.008944     0.007611     0.009210     0.003388
std       0.073655     0.061430     0.064271     0.061907     0.066652
min      -1.207743    -0.249510    -0.270250    -0.330647    -0.277006
25%      -0.018343    -0.023703    -0.027666    -0.023980    -0.028903
50%       0.013658     0.009170     0.005814     0.008877     0.001972
75%       0.050774     0.040686     0.041697     0.039439     0.036780
max       0.240550     0.443358     0.439010     0.561189     0.532099

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.004275     0.004896     0.001385     0.003986     0.002819
std       0.063043     0.061169     0.059404     0.062337     0.062425
min      -0.639343    -0.287167    -0.610041    -0.252014    -0.377434
25%      -0.027914    -0.028084    -0.028156    -0.028390    -0.029387
50%       0.006206     0.005125     0.003683     0.003338     0.003438
75%       0.036702     0.040456     0.032620     0.035724     0.035136
max       0.398753     0.575589     0.352564     0.394870     0.434437

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.002762     0.003387     0.003836
std       0.059815     0.062796     0.062240
min      -0.346253    -0.288228    -0.336845
25%      -0.028336    -0.030877    -0.029794
50%       0.004935     0.002568     0.002884
75%       0.035056     0.035948     0.036834
max       0.232778     0.505038     0.417907
```


nematoda efectores dataset 1 Covarianza de auto cruzamiento (ACC) hidro_mass con valores atípicos.
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
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.
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.25
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	\
0	0.019625	-0.001848	-0.008108	0.016969	0.037711	0.013584	0.017178	
1	0.038273	0.006226	0.081118	0.054531	-0.004516	-0.035899	0.086913	
2	0.109991	-0.015873	-0.035930	-0.017213	-0.077442	-0.060394	-0.005183	
3	-0.011817	-0.029095	-0.039041	0.000039	-0.013657	-0.036385	-0.018370	
4	0.045998	0.014915	-0.020015	-0.018419	-0.047657	-0.001625	-0.063503	
..	…	…	…	…	…	…	…	
995	0.036302	-0.006523	0.052501	0.002684	-0.023637	-0.054717	-0.030591	
996	-0.019978	0.053566	-0.046486	0.043441	-0.048772	-0.019229	0.094234	
997	0.043840	-0.059896	0.007442	-0.071917	-0.079505	0.053620	0.212745	
998	0.114792	0.053308	0.003530	-0.116544	0.040821	0.063930	0.065471	
999	0.060660	-0.058119	-0.021007	0.083183	-0.022137	-0.067753	0.016109	

	X7	X8	X9	X10	X11	X12	X13
0	0.036823	-0.038163	0.024723	0.036541	-0.060687	0.012481	efectores
1	0.002178	-0.036401	-0.014205	-0.036856	-0.029526	-0.030152	efectores
2	0.059509	0.105795	0.079493	-0.034332	-0.009044	-0.019448	efectores
3	-0.030519	0.019919	0.036059	0.039881	0.045897	-0.030677	efectores
4	-0.010114	0.011296	-0.022164	0.020798	0.045351	-0.073713	efectores
..	…	…	…	…	…	…	…
995	-0.021354	-0.038319	0.020939	-0.051076	-0.023962	0.003693	efectores
996	-0.111275	0.048079	0.014268	0.028765	-0.028096	-0.037747	efectores
997	-0.002601	-0.135568	-0.052740	-0.134199	-0.015835	-0.167305	efectores
998	0.118267	-0.047932	-0.070228	0.045694	0.101246	0.033825	efectores
999	-0.022026	-0.037645	-0.007081	-0.021282	-0.091118	-0.065639	efectores

[916 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	916.000000	916.000000	916.000000	916.000000	916.000000	916.000000	
mean	0.019764	0.010070	0.005418	0.005618	0.001994	0.002102	
std	0.060977	0.062001	0.060012	0.058327	0.058990	0.060524	
min	-0.182771	-0.217918	-0.191644	-0.188767	-0.201347	-0.217727	
25%	-0.016350	-0.027207	-0.031262	-0.028921	-0.032646	-0.033589	
50%	0.023499	0.010795	0.002813	0.007509	0.001944	0.003247	
75%	0.055781	0.047041	0.043892	0.043314	0.039136	0.041505	
max	0.212228	0.210899	0.218139	0.203243	0.214412	0.192435	

```
                X6           X7           X8           X9          X10          X11  \
count   916.000000   916.000000   916.000000   916.000000   916.000000   916.000000
mean      0.008337     0.001814     0.005505     0.006245     0.000508     0.003110
std       0.060418     0.062952     0.061302     0.058238     0.060201     0.064395
min      -0.199514    -0.217032    -0.215229    -0.185995    -0.194576    -0.231162
25%      -0.027349    -0.034203    -0.027775    -0.028122    -0.036098    -0.031275
50%       0.009837     0.003303     0.007025     0.004617    -0.000564     0.004200
75%       0.044614     0.039091     0.039786     0.040370     0.037675     0.043226
max       0.213983     0.222613     0.207965     0.217524     0.203472     0.226267

               X12
count   916.000000
mean      0.004721
std       0.062847
min      -0.198887
25%      -0.029628
50%       0.006276
75%       0.042293
max       0.205097
```

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    0.017580  0.022179  0.027164  0.017104  0.001275 -0.022052 -0.003837
1    0.009096  0.013476 -0.100321  0.023707 -0.023600 -0.039430 -0.078163
2    0.054927  0.015489  0.008992 -0.008384  0.012210  0.014689 -0.020932
3    0.028351  0.024664  0.076025  0.049284  0.051564 -0.019353  0.051728
4    0.044265  0.051431  0.077956 -0.031474 -0.006130 -0.030696  0.017136
..        …         …         …         …         …         …         …
994 -0.052807 -0.074447 -0.014823  0.040332  0.031054  0.007968  0.054998
996 -0.045558  0.078998 -0.040740 -0.053890  0.005129  0.028930  0.001145
997 -0.004655 -0.070958 -0.049885  0.017624  0.064193  0.034228 -0.027500
998 -0.039260  0.024520  0.029100  0.059829 -0.076523  0.037623  0.043185
999  0.078712  0.066046  0.146760  0.045262  0.028705  0.013218 -0.007974

           X7        X8        X9       X10       X11       X12           X13
0   -0.052146 -0.011817 -0.005375 -0.021543  0.027961  0.012654  no_efectores
1   -0.023510  0.087389 -0.012996  0.025569 -0.014514  0.020138  no_efectores
2   -0.003877 -0.003104  0.024400  0.019663 -0.025867  0.010056  no_efectores
3   -0.037745 -0.000412  0.052082 -0.024103 -0.021581  0.003359  no_efectores
4    0.022518  0.082939  0.096112  0.039332  0.040911  0.017739  no_efectores
..        …         …         …         …         …         …           …
994  0.015326 -0.088767 -0.015535 -0.013625 -0.039201  0.053529  no_efectores
```

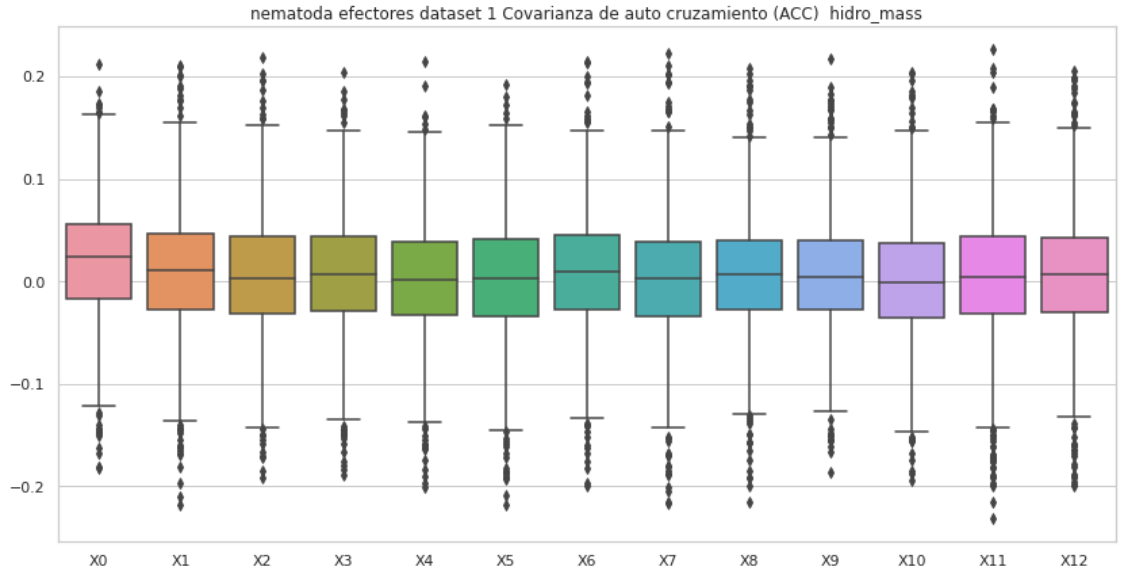
```
996  0.014863  0.076227  0.052230  0.047585 -0.015317 -0.048790  no_efectores
997  0.049648  0.029439 -0.021985  0.039596  0.020699  0.000186  no_efectores
998 -0.064152  0.020724 -0.004462  0.018598 -0.024517  0.071289  no_efectores
999  0.060898  0.138351 -0.086351  0.069640  0.029354 -0.012334  no_efectores

[903 rows x 14 columns]
```

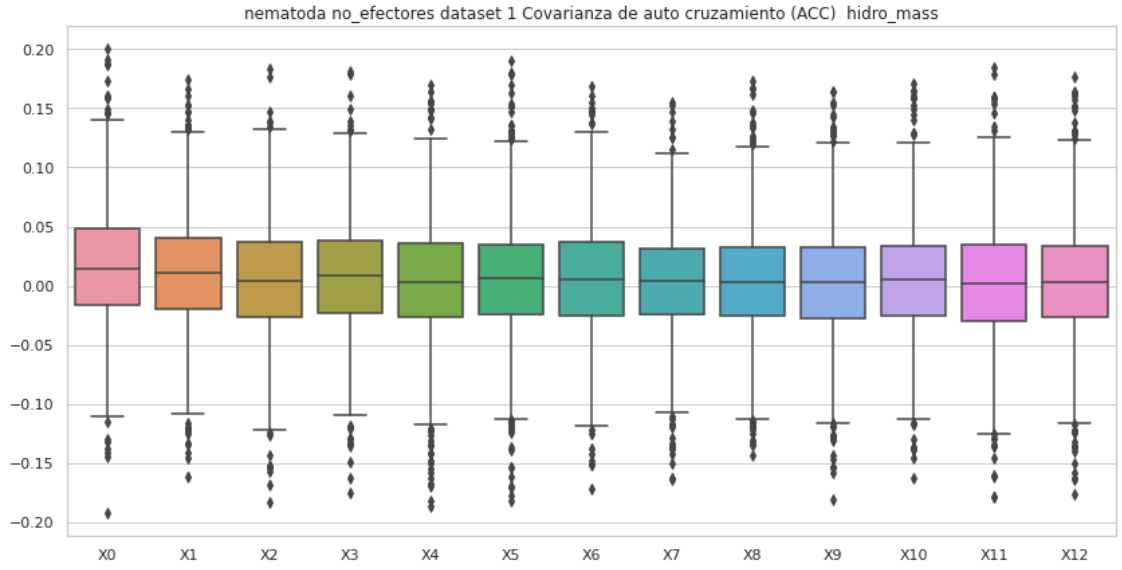
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores nematoda dataset 1, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	903.000000	903.000000	903.000000	903.000000	903.000000	903.000000	
mean	0.016628	0.009514	0.004911	0.008136	0.003169	0.004757	
std	0.053955	0.050370	0.052405	0.049302	0.052327	0.052263	
min	-0.191986	-0.161682	-0.183043	-0.175318	-0.186126	-0.182155	
25%	-0.015955	-0.019903	-0.026590	-0.022592	-0.025934	-0.024148	
50%	0.014389	0.010387	0.004190	0.008866	0.002668	0.006231	
75%	0.047902	0.040245	0.037549	0.038315	0.035597	0.035047	
max	0.200134	0.174136	0.183294	0.181502	0.169424	0.189785	

	X6	X7	X8	X9	X10	X11	\
count	903.000000	903.000000	903.000000	903.000000	903.000000	903.000000	
mean	0.005052	0.002284	0.003619	0.001887	0.004329	0.001545	
std	0.048853	0.047503	0.048698	0.049211	0.049063	0.051039	
min	-0.171543	-0.163581	-0.142937	-0.180740	-0.162078	-0.178355	
25%	-0.025820	-0.024627	-0.025832	-0.027160	-0.025424	-0.029358	
50%	0.005018	0.003898	0.003021	0.003227	0.005666	0.002067	
75%	0.036894	0.030907	0.032092	0.032283	0.034009	0.034431	
max	0.168846	0.155019	0.173393	0.163790	0.170948	0.183812	

	X12
count	903.000000
mean	0.004418
std	0.051469
min	-0.175885
25%	-0.025992
50%	0.003163
75%	0.034034
max	0.176976





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    0.019625 -0.001848 -0.008108  0.016969  0.037711  0.013584  0.017178
1    0.038273  0.006226  0.081118  0.054531 -0.004516 -0.035899  0.086913
2    0.109991 -0.015873 -0.035930 -0.017213 -0.077442 -0.060394 -0.005183
3   -0.011817 -0.029095 -0.039041  0.000039 -0.013657 -0.036385 -0.018370
4    0.045998  0.014915 -0.020015 -0.018419 -0.047657 -0.001625 -0.063503
..        …         …         …         …         …         …         …
995  0.036302 -0.006523  0.052501  0.002684 -0.023637 -0.054717 -0.030591
996 -0.019978  0.053566 -0.046486  0.043441 -0.048772 -0.019229  0.094234
997  0.043840 -0.059896  0.007442 -0.071917 -0.079505  0.053620  0.212745
998  0.114792  0.053308  0.003530 -0.116544  0.040821  0.063930  0.065471
999  0.060660 -0.058119 -0.021007  0.083183 -0.022137 -0.067753  0.016109

           X7        X8        X9       X10       X11       X12         X13
0    0.036823 -0.038163  0.024723  0.036541 -0.060687  0.012481  efectores
1    0.002178 -0.036401 -0.014205 -0.036856 -0.029526 -0.030152  efectores
2    0.059509  0.105795  0.079493 -0.034332 -0.009044 -0.019448  efectores
```

```
3   -0.030519  0.019919  0.036059  0.039881  0.045897 -0.030677  efectores
4   -0.010114  0.011296 -0.022164  0.020798  0.045351 -0.073713  efectores
..        …       …        …        …         …        …        …
995 -0.021354 -0.038319  0.020939 -0.051076 -0.023962  0.003693  efectores
996 -0.111275  0.048079  0.014268  0.028765 -0.028096 -0.037747  efectores
997 -0.002601 -0.135568 -0.052740 -0.134199 -0.015835 -0.167305  efectores
998  0.118267 -0.047932 -0.070228  0.045694  0.101246  0.033825  efectores
999 -0.022026 -0.037645 -0.007081 -0.021282 -0.091118 -0.065639  efectores

[1000 rows x 14 columns]
```

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

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.020739	0.009894	0.009366	0.006711	0.003008	
std	0.069948	0.076227	0.077688	0.070445	0.070899	
min	-0.267594	-0.545313	-0.315913	-0.329086	-0.282758	
25%	-0.019785	-0.029117	-0.032091	-0.031932	-0.035989	
50%	0.023906	0.011160	0.003319	0.007803	0.002669	
75%	0.057970	0.050888	0.047875	0.045440	0.042105	
max	0.338158	0.450677	0.515261	0.416225	0.457174	

	X5	X6	X7	X8	X9	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.005562	0.007968	0.004163	0.006573	0.004508	
std	0.076218	0.072424	0.074537	0.075388	0.071772	
min	-0.342657	-0.281087	-0.294338	-0.330694	-0.374668	
25%	-0.035776	-0.029610	-0.035656	-0.031854	-0.031202	
50%	0.004332	0.009656	0.004903	0.007112	0.004294	
75%	0.044823	0.045614	0.042544	0.042315	0.042468	
max	0.528023	0.411799	0.543545	0.390329	0.340530	

	X10	X11	X12
count	1000.000000	1000.000000	1000.000000
mean	0.001420	0.006962	0.003540
std	0.072039	0.079497	0.071830
min	-0.291168	-0.381983	-0.358366
25%	-0.038155	-0.032640	-0.034902
50%	-0.000428	0.005645	0.005406
75%	0.039916	0.045923	0.043169
max	0.633831	0.473414	0.384113

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	0.017580	0.022179	0.027164	0.017104	0.001275	-0.022052	-0.003837	
1	0.009096	0.013476	-0.100321	0.023707	-0.023600	-0.039430	-0.078163	
2	0.054927	0.015489	0.008992	-0.008384	0.012210	0.014689	-0.020932	
3	0.028351	0.024664	0.076025	0.049284	0.051564	-0.019353	0.051728	
4	0.044265	0.051431	0.077956	-0.031474	-0.006130	-0.030696	0.017136	
..	…	…	…	…	…	…	…	
995	0.102852	-0.068763	0.044283	0.129520	-0.041985	-0.020252	-0.195257	
996	-0.045558	0.078998	-0.040740	-0.053890	0.005129	0.028930	0.001145	
997	-0.004655	-0.070958	-0.049885	0.017624	0.064193	0.034228	-0.027500	
998	-0.039260	0.024520	0.029100	0.059829	-0.076523	0.037623	0.043185	
999	0.078712	0.066046	0.146760	0.045262	0.028705	0.013218	-0.007974	

	X7	X8	X9	X10	X11	X12	X13
0	-0.052146	-0.011817	-0.005375	-0.021543	0.027961	0.012654	no_efectores
1	-0.023510	0.087389	-0.012996	0.025569	-0.014514	0.020138	no_efectores
2	-0.003877	-0.003104	0.024400	0.019663	-0.025867	0.010056	no_efectores
3	-0.037745	-0.000412	0.052082	-0.024103	-0.021581	0.003359	no_efectores
4	0.022518	0.082939	0.096112	0.039332	0.040911	0.017739	no_efectores
..	…	…	…	…	…	…	…
995	-0.273295	0.062250	0.034087	-0.046227	-0.089696	0.125439	no_efectores
996	0.014863	0.076227	0.052230	0.047585	-0.015317	-0.048790	no_efectores
997	0.049648	0.029439	-0.021985	0.039596	0.020699	0.000186	no_efectores
998	-0.064152	0.020724	-0.004462	0.018598	-0.024517	0.071289	no_efectores
999	0.060898	0.138351	-0.086351	0.069640	0.029354	-0.012334	no_efectores

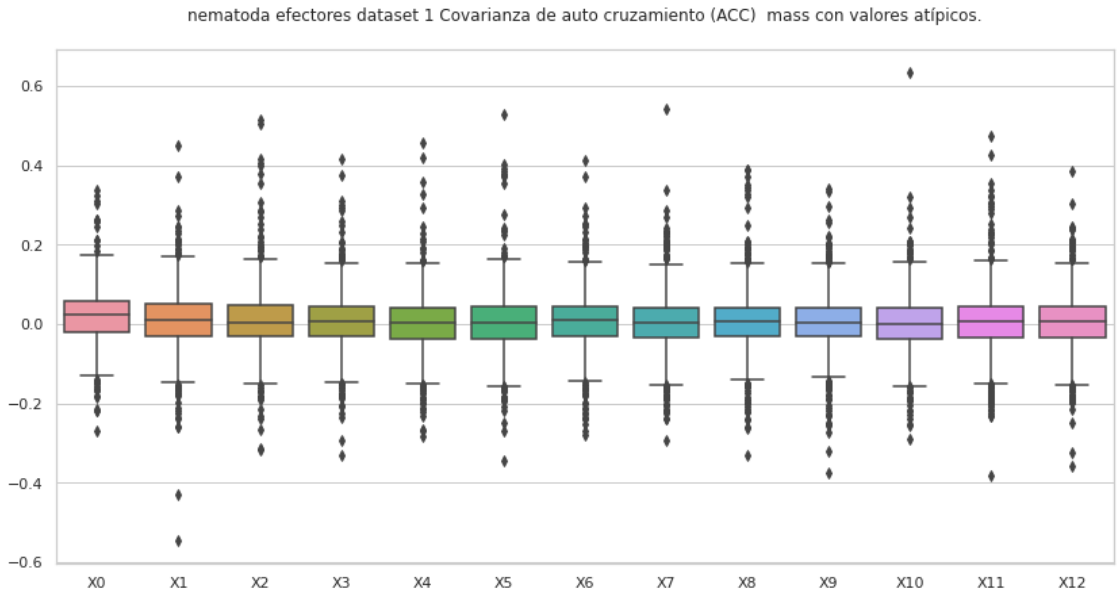
[1000 rows x 14 columns]

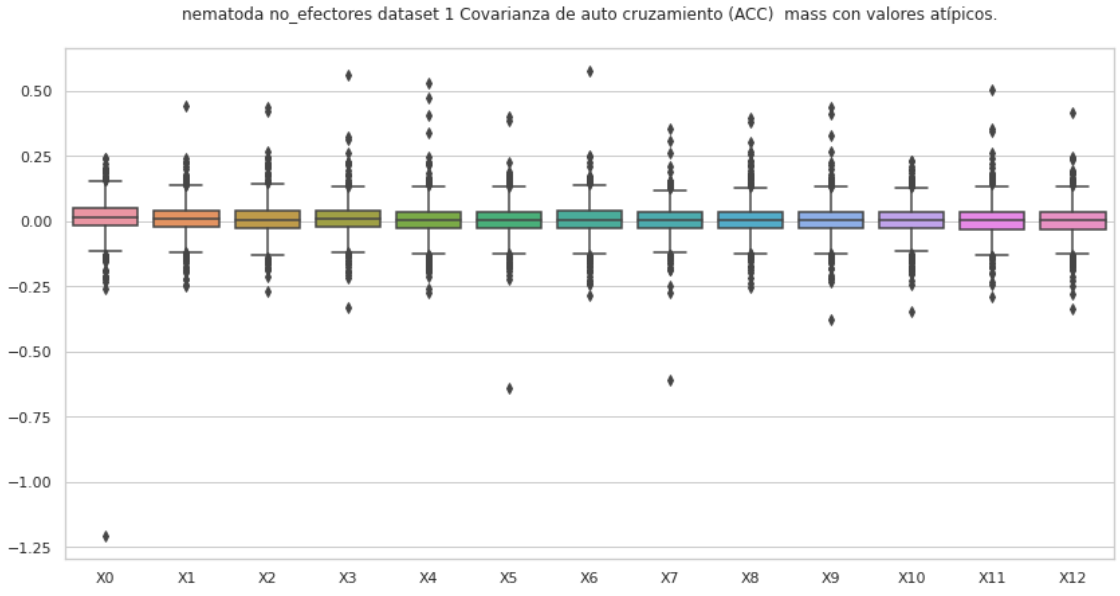
Covarianza de auto cruzamiento (ACC) mass no_efectores nematoda dataset 1, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.013700	0.008944	0.007611	0.009210	0.003388	
std	0.073655	0.061430	0.064271	0.061907	0.066652	
min	-1.207743	-0.249510	-0.270250	-0.330647	-0.277006	
25%	-0.018343	-0.023703	-0.027666	-0.023980	-0.028903	
50%	0.013658	0.009170	0.005814	0.008877	0.001972	
75%	0.050774	0.040686	0.041697	0.039439	0.036780	
max	0.240550	0.443358	0.439010	0.561189	0.532099	

	X5	X6	X7	X8	X9	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.004275	0.004896	0.001385	0.003986	0.002819	
std	0.063043	0.061169	0.059404	0.062337	0.062425	
min	-0.639343	-0.287167	-0.610041	-0.252014	-0.377434	
25%	-0.027914	-0.028084	-0.028156	-0.028390	-0.029387	
50%	0.006206	0.005125	0.003683	0.003338	0.003438	
75%	0.036702	0.040456	0.032620	0.035724	0.035136	
max	0.398753	0.575589	0.352564	0.394870	0.434437	

	X10	X11	X12
count	1000.000000	1000.000000	1000.000000
mean	0.002762	0.003387	0.003836
std	0.059815	0.062796	0.062240
min	-0.346253	-0.288228	-0.336845
25%	-0.028336	-0.030877	-0.029794
50%	0.004935	0.002568	0.002884
75%	0.035056	0.035948	0.036834
max	0.232778	0.505038	0.417907





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  \
0    0.019625 -0.001848 -0.008108  0.016969  0.037711  0.013584  0.017178
1    0.038273  0.006226  0.081118  0.054531 -0.004516 -0.035899  0.086913
2    0.109991 -0.015873 -0.035930 -0.017213 -0.077442 -0.060394 -0.005183
3   -0.011817 -0.029095 -0.039041  0.000039 -0.013657 -0.036385 -0.018370
4    0.045998  0.014915 -0.020015 -0.018419 -0.047657 -0.001625 -0.063503
..        …         …         …         …         …         …         …
995  0.036302 -0.006523  0.052501  0.002684 -0.023637 -0.054717 -0.030591
996 -0.019978  0.053566 -0.046486  0.043441 -0.048772 -0.019229  0.094234
997  0.043840 -0.059896  0.007442 -0.071917 -0.079505  0.053620  0.212745
998  0.114792  0.053308  0.003530 -0.116544  0.040821  0.063930  0.065471
999  0.060660 -0.058119 -0.021007  0.083183 -0.022137 -0.067753  0.016109

           X7        X8        X9       X10       X11       X12        X13
0    0.036823 -0.038163  0.024723  0.036541 -0.060687  0.012481  efectores
1    0.002178 -0.036401 -0.014205 -0.036856 -0.029526 -0.030152  efectores
2    0.059509  0.105795  0.079493 -0.034332 -0.009044 -0.019448  efectores
3   -0.030519  0.019919  0.036059  0.039881  0.045897 -0.030677  efectores
4   -0.010114  0.011296 -0.022164  0.020798  0.045351 -0.073713  efectores
..        …         …         …         …         …         …         …
995 -0.021354 -0.038319  0.020939 -0.051076 -0.023962  0.003693  efectores
```

```
996 -0.111275  0.048079  0.014268  0.028765 -0.028096 -0.037747  efectores
997 -0.002601 -0.135568 -0.052740 -0.134199 -0.015835 -0.167305  efectores
998  0.118267 -0.047932 -0.070228  0.045694  0.101246  0.033825  efectores
999 -0.022026 -0.037645 -0.007081 -0.021282 -0.091118 -0.065639  efectores

[916 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	916.000000	916.000000	916.000000	916.000000	916.000000	916.000000	
mean	0.019764	0.010070	0.005418	0.005618	0.001994	0.002102	
std	0.060977	0.062001	0.060012	0.058327	0.058990	0.060524	
min	-0.182771	-0.217918	-0.191644	-0.188767	-0.201347	-0.217727	
25%	-0.016350	-0.027207	-0.031262	-0.028921	-0.032646	-0.033589	
50%	0.023499	0.010795	0.002813	0.007509	0.001944	0.003247	
75%	0.055781	0.047041	0.043892	0.043314	0.039136	0.041505	
max	0.212228	0.210899	0.218139	0.203243	0.214412	0.192435	

	X6	X7	X8	X9	X10	X11	\
count	916.000000	916.000000	916.000000	916.000000	916.000000	916.000000	
mean	0.008337	0.001814	0.005505	0.006245	0.000508	0.003110	
std	0.060418	0.062952	0.061302	0.058238	0.060201	0.064395	
min	-0.199514	-0.217032	-0.215229	-0.185995	-0.194576	-0.231162	
25%	-0.027349	-0.034203	-0.027775	-0.028122	-0.036098	-0.031275	
50%	0.009837	0.003303	0.007025	0.004617	-0.000564	0.004200	
75%	0.044614	0.039091	0.039786	0.040370	0.037675	0.043226	
max	0.213983	0.222613	0.207965	0.217524	0.203472	0.226267	

	X12
count	916.000000
mean	0.004721
std	0.062847
min	-0.198887
25%	-0.029628
50%	0.006276
75%	0.042293
max	0.205097

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

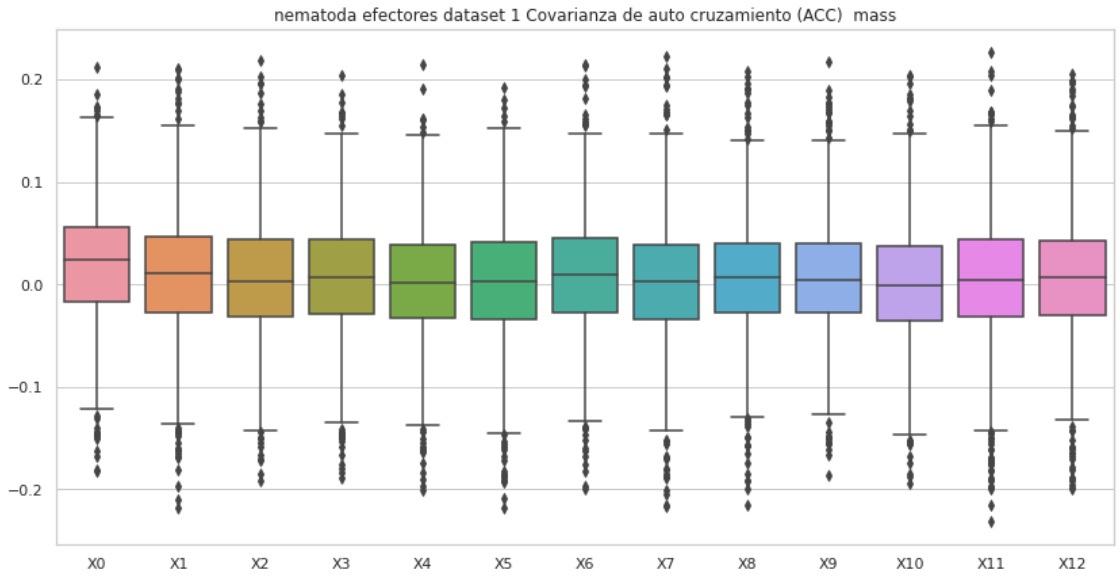
```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.017580  0.022179  0.027164  0.017104  0.001275 -0.022052 -0.003837
1    0.009096  0.013476 -0.100321  0.023707 -0.023600 -0.039430 -0.078163
2    0.054927  0.015489  0.008992 -0.008384  0.012210  0.014689 -0.020932
3    0.028351  0.024664  0.076025  0.049284  0.051564 -0.019353  0.051728
4    0.044265  0.051431  0.077956 -0.031474 -0.006130 -0.030696  0.017136
..        …         …         …         …         …         …
994 -0.052807 -0.074447 -0.014823  0.040332  0.031054  0.007968  0.054998
996 -0.045558  0.078998 -0.040740 -0.053890  0.005129  0.028930  0.001145
997 -0.004655 -0.070958 -0.049885  0.017624  0.064193  0.034228 -0.027500
998 -0.039260  0.024520  0.029100  0.059829 -0.076523  0.037623  0.043185
999  0.078712  0.066046  0.146760  0.045262  0.028705  0.013218 -0.007974

           X7        X8        X9       X10       X11       X12           X13
0   -0.052146 -0.011817 -0.005375 -0.021543  0.027961  0.012654  no_efectores
1   -0.023510  0.087389 -0.012996  0.025569 -0.014514  0.020138  no_efectores
2   -0.003877 -0.003104  0.024400  0.019663 -0.025867  0.010056  no_efectores
3   -0.037745 -0.000412  0.052082 -0.024103 -0.021581  0.003359  no_efectores
4    0.022518  0.082939  0.096112  0.039332  0.040911  0.017739  no_efectores
..        …         …         …         …         …         …
994  0.015326 -0.088767 -0.015535 -0.013625 -0.039201  0.053529  no_efectores
996  0.014863  0.076227  0.052230  0.047585 -0.015317 -0.048790  no_efectores
997  0.049648  0.029439 -0.021985  0.039596  0.020699  0.000186  no_efectores
998 -0.064152  0.020724 -0.004462  0.018598 -0.024517  0.071289  no_efectores
999  0.060898  0.138351 -0.086351  0.069640  0.029354 -0.012334  no_efectores

[903 rows x 14 columns]
```

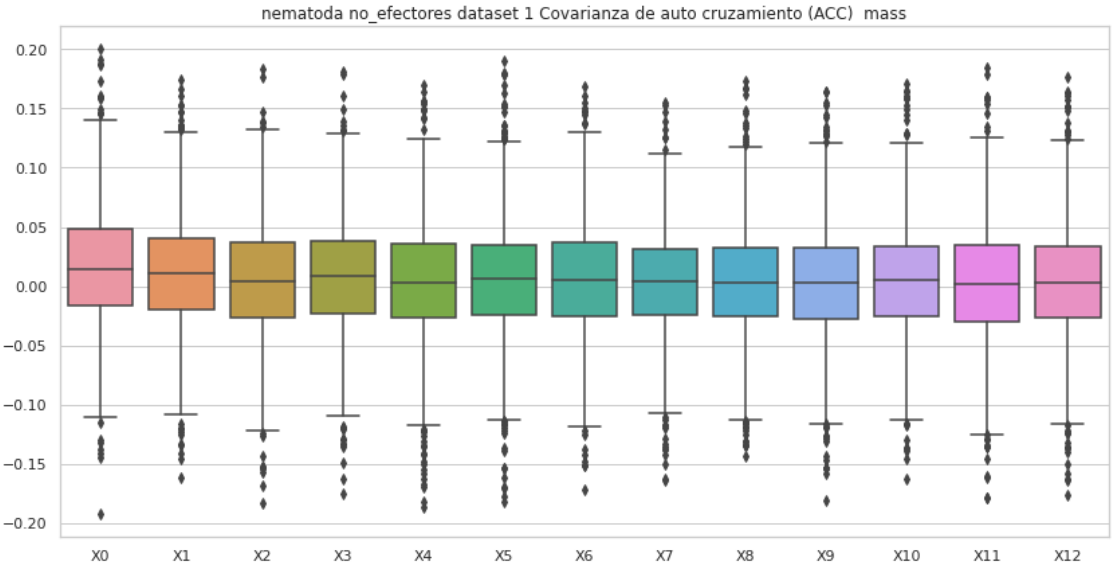
Covarianza de auto cruzamiento (ACC) mass no_efectores nematoda dataset 1, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  903.000000  903.000000  903.000000  903.000000  903.000000  903.000000
mean     0.016628    0.009514    0.004911    0.008136    0.003169    0.004757
std      0.053955    0.050370    0.052405    0.049302    0.052327    0.052263
min     -0.191986   -0.161682   -0.183043   -0.175318   -0.186126   -0.182155
25%     -0.015955   -0.019903   -0.026590   -0.022592   -0.025934   -0.024148
50%      0.014389    0.010387    0.004190    0.008866    0.002668    0.006231
75%      0.047902    0.040245    0.037549    0.038315    0.035597    0.035047
max      0.200134    0.174136    0.183294    0.181502    0.169424    0.189785

                X6          X7          X8          X9         X10         X11  \
count  903.000000  903.000000  903.000000  903.000000  903.000000  903.000000
mean     0.005052    0.002284    0.003619    0.001887    0.004329    0.001545
std      0.048853    0.047503    0.048698    0.049211    0.049063    0.051039
min     -0.171543   -0.163581   -0.142937   -0.180740   -0.162078   -0.178355
```

25%	-0.025820	-0.024627	-0.025832	-0.027160	-0.025424	-0.029358
50%	0.005018	0.003898	0.003021	0.003227	0.005666	0.002067
75%	0.036894	0.030907	0.032092	0.032283	0.034009	0.034431
max	0.168846	0.155019	0.173393	0.163790	0.170948	0.183812

	X12
count	903.000000
mean	0.004418
std	0.051469
min	-0.175885
25%	-0.025992
50%	0.003163
75%	0.034034
max	0.176976





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   -0.015879 -0.227110  0.041703  0.235716 -0.052305 -0.045074  0.058961
1    0.010432  0.052530  0.055032  0.063511  0.023262  0.003188  0.013872
2   -0.105986 -0.105110 -0.012023  0.109602  0.009184  0.067531  0.057841
3    0.063876 -0.112935  0.052471  0.041460 -0.015919 -0.029323  0.013991
4    0.076748  0.009897  0.025720 -0.082987 -0.051708 -0.064936 -0.092319
..        …         …         …         …         …         …         …
995  0.096076  0.042401 -0.011171  0.096414  0.111949  0.143337  0.071515
996  0.030539 -0.072813 -0.034217 -0.121507 -0.107095 -0.023468  0.083599
997 -0.121834 -0.134590 -0.005868 -0.028540  0.240343  0.023227  0.151660
998  0.025105  0.020409 -0.092511  0.003513  0.060886 -0.009475  0.105048
999  0.125635  0.002760  0.038263  0.132525  0.026728 -0.028473 -0.061868

           X7        X8        X9       X10       X11       X12         X13
0    0.070757 -0.175323 -0.052719  0.025820 -0.047692  0.036330  efectores
1    0.010695 -0.152736  0.066296 -0.035506 -0.090709 -0.043982  efectores
2   -0.131059 -0.007901 -0.026685  0.080164 -0.097253 -0.089864  efectores
3   -0.046375  0.012867  0.020937 -0.012449 -0.021056 -0.002420  efectores
4   -0.026142  0.057181  0.087239  0.031185 -0.017453  0.001573  efectores
..        …         …         …         …         …         …         …
995 -0.011190  0.074856  0.107583  0.048024  0.105940 -0.024704  efectores
996  0.066431  0.104389  0.047752 -0.036114 -0.103704 -0.103224  efectores
997 -0.109650 -0.158080  0.142666  0.106810  0.029089 -0.248484  efectores
998 -0.001678 -0.004229 -0.053298  0.002519  0.000472  0.066354  efectores
999 -0.012575  0.035294 -0.033653 -0.003241  0.044868 -0.036476  efectores

[1000 rows x 14 columns]


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

                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.013236    -0.020560     0.025510     0.019164    -0.006741
```

```
std       0.079991     0.085813     0.085565     0.083007     0.080410
min      -0.319572    -0.328009    -0.437658    -0.396283    -0.376306
25%      -0.032063    -0.075622    -0.021892    -0.029680    -0.053879
50%       0.010448    -0.019586     0.025304     0.017543    -0.008132
75%       0.055814     0.030726     0.074947     0.067322     0.042606
max       0.373769     0.290860     0.378514     0.358347     0.346224

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean     -0.005901     0.017945     0.005390     0.002087     0.002636
std       0.080077     0.083842     0.079595     0.078184     0.079848
min      -0.288412    -0.451203    -0.256661    -0.337866    -0.330964
25%      -0.056659    -0.028813    -0.041582    -0.041302    -0.042931
50%      -0.006159     0.018720     0.006596     0.002158     0.004943
75%       0.039955     0.069005     0.050309     0.047927     0.049227
max       0.356724     0.356758     0.385286     0.544768     0.378794

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.016395     0.010731     0.002759
std       0.082196     0.079149     0.083172
min      -0.248835    -0.284340    -0.417123
25%      -0.029888    -0.034660    -0.041032
50%       0.011171     0.011028     0.002870
75%       0.059210     0.053860     0.048878
max       0.447349     0.383714     0.447980
```

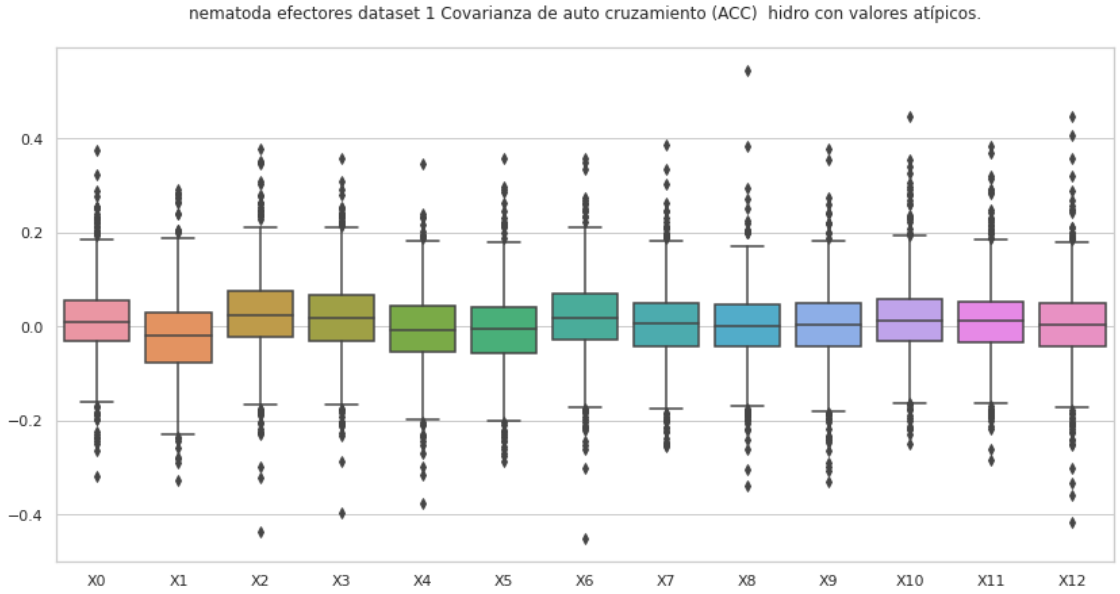
no_efectores
Covarianza de auto cruzamiento (ACC) hidro no_efectores nematoda dataset 1, con valores atípicos.
Valores del documento csv.

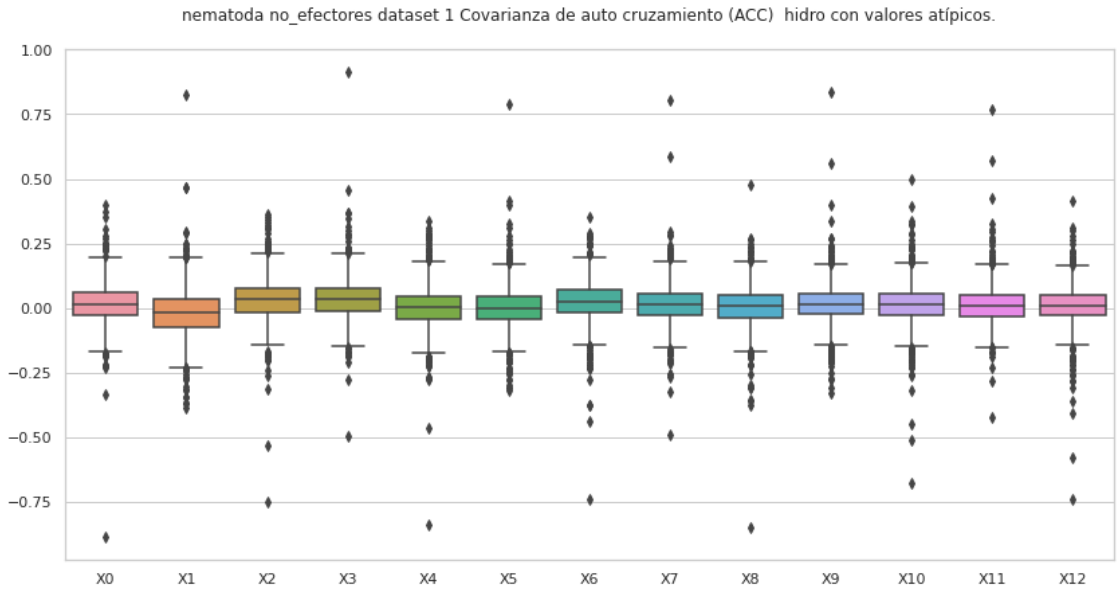
```
           X0        X1        X2        X3        X4        X5        X6  \
0   -0.008800 -0.090361  0.000426  0.004647  0.012586  0.012021 -0.027270
1    0.043484 -0.018800 -0.039152  0.074500  0.086162  0.001562 -0.028151
2    0.053034 -0.089265  0.011553  0.007399  0.002885 -0.045284 -0.025541
3    0.117177  0.021792  0.070165  0.083552  0.008154  0.020945  0.082580
4    0.121314 -0.000484 -0.003454  0.045097  0.069876  0.097343  0.086205
..        …       …         …         …         …         …         …
995  0.168907  0.044849  0.134128 -0.071104 -0.109648  0.051198  0.075702
996  0.004084 -0.143751 -0.025218  0.047390 -0.031197 -0.054521 -0.052160
997 -0.024063 -0.033709  0.051080 -0.083193 -0.000176 -0.079151  0.130902
998 -0.042356  0.084443  0.035541 -0.069032 -0.027521 -0.064881  0.012255
999  0.126724  0.083856  0.109494  0.057561  0.046918  0.022956 -0.063334

           X7        X8        X9       X10       X11       X12           X13
```

```
0    0.060355 -0.002245  0.032358 -0.022482  0.033734  0.012359  no_efectores
1   -0.033270 -0.001962  0.030205  0.024899 -0.029829  0.005693  no_efectores
2    0.035846  0.071050 -0.046408  0.005012 -0.026854 -0.004936  no_efectores
3    0.036637  0.088414  0.126459  0.065041 -0.012315  0.006733  no_efectores
4    0.049286  0.055767  0.119392  0.067979  0.042105  0.045791  no_efectores
..        …       …       …       …       …       …           …
995 -0.109478 -0.104500  0.135657 -0.139982 -0.084147 -0.050374  no_efectores
996  0.044737  0.003461 -0.023499 -0.020293 -0.012525  0.002612  no_efectores
997 -0.101615 -0.009979 -0.014860  0.006032  0.047263 -0.081721  no_efectores
998  0.031348  0.014213  0.054482 -0.014831  0.039383  0.026504  no_efectores
999 -0.135519 -0.074895 -0.105516 -0.101347 -0.075728 -0.136789  no_efectores

[1000 rows x 14 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.016998    -0.018837     0.031950     0.033826     0.002194
std       0.084094     0.096191     0.085172     0.083907     0.086549
min      -0.886490    -0.386752    -0.752495    -0.495935    -0.840153
25%      -0.029806    -0.072588    -0.016121    -0.013636    -0.042213
50%       0.015573    -0.018343     0.032362     0.032608     0.001422
75%       0.062318     0.033921     0.075552     0.076628     0.047056
max       0.400733     0.823115     0.361222     0.913339     0.337297

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.001126     0.024317     0.013473     0.006107     0.013600
std       0.085272     0.082245     0.082242     0.083284     0.080128
min      -0.319254    -0.743267    -0.491289    -0.851732    -0.328833
25%      -0.042084    -0.016760    -0.027604    -0.037368    -0.024135
50%      -0.001155     0.022868     0.012112     0.009987     0.012507
75%       0.043756     0.069101     0.057516     0.049247     0.054355
max       0.786487     0.354065     0.804329     0.474297     0.836100

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.012756     0.012760     0.007546
std       0.083569     0.079703     0.081507
min      -0.678470    -0.426001    -0.738911
25%      -0.028040    -0.031445    -0.029958
50%       0.012021     0.006753     0.006796
75%       0.053036     0.048596     0.048057
max       0.497299     0.766591     0.412903
```


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.
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

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	-0.015879	-0.227110	0.041703	0.235716	-0.052305	-0.045074	0.058961	
1	0.010432	0.052530	0.055032	0.063511	0.023262	0.003188	0.013872	
2	-0.105986	-0.105110	-0.012023	0.109602	0.009184	0.067531	0.057841	
3	0.063876	-0.112935	0.052471	0.041460	-0.015919	-0.029323	0.013991	
4	0.076748	0.009897	0.025720	-0.082987	-0.051708	-0.064936	-0.092319	
..	…	…	…	…	…	…	…	
994	-0.024666	-0.028229	0.017673	-0.011255	-0.092907	-0.013107	0.004820	
995	0.096076	0.042401	-0.011171	0.096414	0.111949	0.143337	0.071515	
996	0.030539	-0.072813	-0.034217	-0.121507	-0.107095	-0.023468	0.083599	
998	0.025105	0.020409	-0.092511	0.003513	0.060886	-0.009475	0.105048	
999	0.125635	0.002760	0.038263	0.132525	0.026728	-0.028473	-0.061868	

	X7	X8	X9	X10	X11	X12	X13
0	0.070757	-0.175323	-0.052719	0.025820	-0.047692	0.036330	efectores
1	0.010695	-0.152736	0.066296	-0.035506	-0.090709	-0.043982	efectores
2	-0.131059	-0.007901	-0.026685	0.080164	-0.097253	-0.089864	efectores
3	-0.046375	0.012867	0.020937	-0.012449	-0.021056	-0.002420	efectores
4	-0.026142	0.057181	0.087239	0.031185	-0.017453	0.001573	efectores
..	…	…	…	…	…	…	
994	0.002552	0.017123	-0.093487	-0.053149	0.025865	0.007259	efectores
995	-0.011190	0.074856	0.107583	0.048024	0.105940	-0.024704	efectores
996	0.066431	0.104389	0.047752	-0.036114	-0.103704	-0.103224	efectores
998	-0.001678	-0.004229	-0.053298	0.002519	0.000472	0.066354	efectores
999	-0.012575	0.035294	-0.033653	-0.003241	0.044868	-0.036476	efectores

[914 rows x 14 columns]

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

	X0	X1	X2	X3	X4	X5	\
count	914.000000	914.000000	914.000000	914.000000	914.000000	914.000000	
mean	0.012294	-0.022005	0.023147	0.019336	-0.005601	-0.006557	
std	0.068332	0.076252	0.075308	0.075963	0.072719	0.070143	
min	-0.224558	-0.257166	-0.222794	-0.226025	-0.233395	-0.243473	
25%	-0.029899	-0.073615	-0.020846	-0.026580	-0.052548	-0.053072	
50%	0.010059	-0.018927	0.024029	0.017456	-0.007447	-0.006177	
75%	0.050796	0.027086	0.070324	0.065856	0.041422	0.036425	
max	0.250162	0.205703	0.262331	0.253546	0.216541	0.220056	

```
                X6           X7           X8           X9          X10          X11  \
count   914.000000   914.000000   914.000000   914.000000   914.000000   914.000000
mean      0.017393     0.005527     0.001100     0.003157     0.012256     0.007172
std       0.073482     0.069383     0.067170     0.069756     0.072233     0.069451
min      -0.221902    -0.213685    -0.205233    -0.234076    -0.218751    -0.216645
25%      -0.026538    -0.037179    -0.038110    -0.040926    -0.029819    -0.034495
50%       0.018661     0.006613     0.001428     0.004831     0.009693     0.009389
75%       0.064388     0.048175     0.044635     0.046240     0.053621     0.049969
max       0.268745     0.215054     0.205660     0.241310     0.260790     0.247100

               X12
count   914.000000
mean      0.002560
std       0.070534
min      -0.242036
25%      -0.037031
50%       0.002870
75%       0.047375
max       0.247583
```

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   -0.008800 -0.090361  0.000426  0.004647  0.012586  0.012021 -0.027270
1    0.043484 -0.018800 -0.039152  0.074500  0.086162  0.001562 -0.028151
2    0.053034 -0.089265  0.011553  0.007399  0.002885 -0.045284 -0.025541
3    0.117177  0.021792  0.070165  0.083552  0.008154  0.020945  0.082580
4    0.121314 -0.000484 -0.003454  0.045097  0.069876  0.097343  0.086205
..        …         …         …         …         …         …         …
995  0.168907  0.044849  0.134128 -0.071104 -0.109648  0.051198  0.075702
996  0.004084 -0.143751 -0.025218  0.047390 -0.031197 -0.054521 -0.052160
997 -0.024063 -0.033709  0.051080 -0.083193 -0.000176 -0.079151  0.130902
998 -0.042356  0.084443  0.035541 -0.069032 -0.027521 -0.064881  0.012255
999  0.126724  0.083856  0.109494  0.057561  0.046918  0.022956 -0.063334

           X7        X8        X9       X10       X11       X12           X13
0    0.060355 -0.002245  0.032358 -0.022482  0.033734  0.012359  no_efectores
1   -0.033270 -0.001962  0.030205  0.024899 -0.029829  0.005693  no_efectores
2    0.035846  0.071050 -0.046408  0.005012 -0.026854 -0.004936  no_efectores
3    0.036637  0.088414  0.126459  0.065041 -0.012315  0.006733  no_efectores
4    0.049286  0.055767  0.119392  0.067979  0.042105  0.045791  no_efectores
..        …         …         …         …         …         …           …
995 -0.109478 -0.104500  0.135657 -0.139982 -0.084147 -0.050374  no_efectores
```

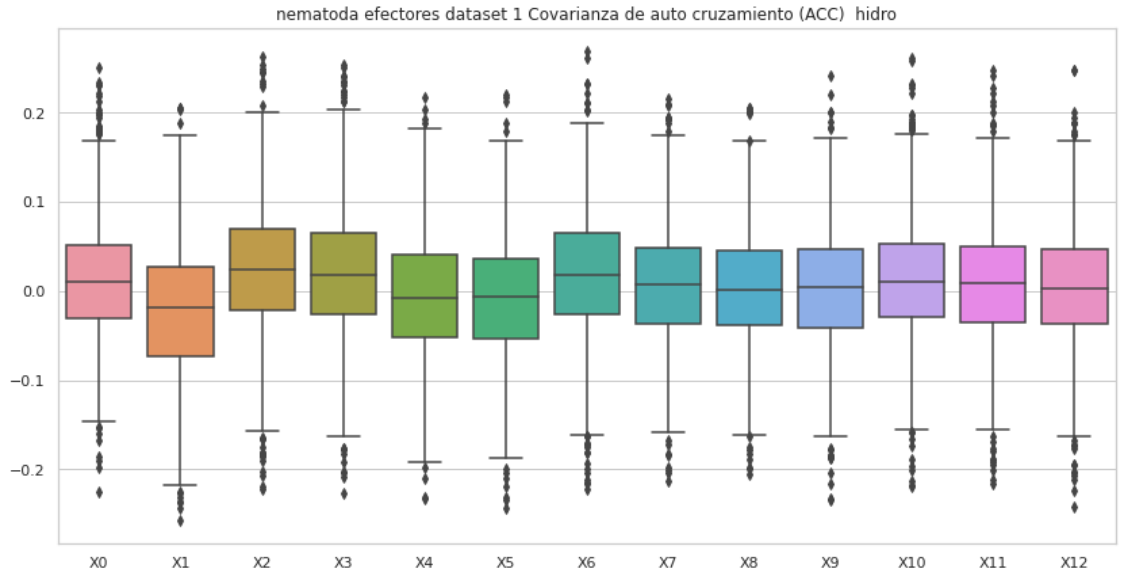
```
996  0.044737  0.003461 -0.023499 -0.020293 -0.012525  0.002612  no_efectores
997 -0.101615 -0.009979 -0.014860  0.006032  0.047263 -0.081721  no_efectores
998  0.031348  0.014213  0.054482 -0.014831  0.039383  0.026504  no_efectores
999 -0.135519 -0.074895 -0.105516 -0.101347 -0.075728 -0.136789  no_efectores

[926 rows x 14 columns]
```

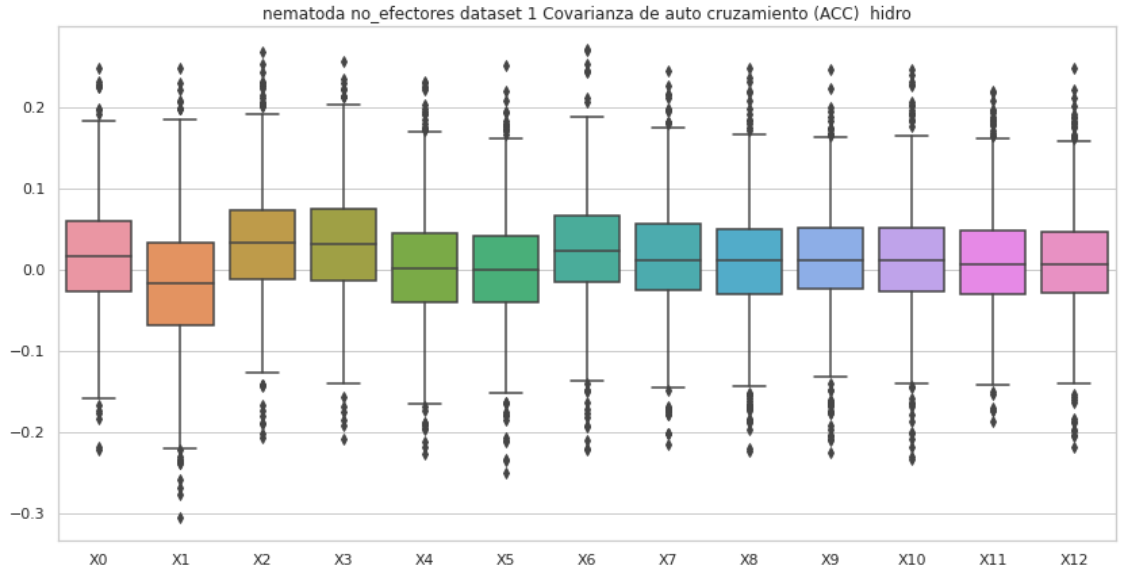
Covarianza de auto cruzamiento (ACC) no_efectores nematoda dataset 1, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	926.000000	926.000000	926.000000	926.000000	926.000000	926.000000	
mean	0.017329	-0.018333	0.031362	0.030548	0.001224	0.000268	
std	0.070717	0.081318	0.069125	0.068532	0.070845	0.068012	
min	-0.222590	-0.305200	-0.206945	-0.208678	-0.227393	-0.249679	
25%	-0.027577	-0.068896	-0.012790	-0.013101	-0.040604	-0.040330	
50%	0.016110	-0.017636	0.032196	0.031097	0.001024	-0.001333	
75%	0.059472	0.032814	0.071945	0.074826	0.043736	0.041333	
max	0.247009	0.248272	0.267358	0.255666	0.231425	0.250189	

	X6	X7	X8	X9	X10	X11	\
count	926.000000	926.000000	926.000000	926.000000	926.000000	926.000000	
mean	0.024854	0.012146	0.009919	0.011417	0.012142	0.009864	
std	0.067961	0.067591	0.068953	0.064813	0.066959	0.064282	
min	-0.221199	-0.215120	-0.223088	-0.225250	-0.233000	-0.186996	
25%	-0.014981	-0.025804	-0.030831	-0.023867	-0.026730	-0.030298	
50%	0.022868	0.011025	0.011400	0.011519	0.011310	0.005999	
75%	0.066530	0.055664	0.049159	0.051026	0.051430	0.047133	
max	0.270703	0.243910	0.246909	0.245152	0.246185	0.219315	

	X12
count	926.000000
mean	0.008820
std	0.064897
min	-0.218380
25%	-0.028655
50%	0.006157
75%	0.046513
max	0.247499


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