ds1_Meloidogyne_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 ="Meloidogyne"
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 Meloidogyne dataset 1, con valores
atípicos.
Valores del documento csv.

        X0     X1     X2      X3      X4      X5     X6      X7     X8  \
0    6.897  4.023  4.023   2.299   7.471   2.874  3.448   9.770  2.299
1    1.266  2.532  8.861   2.532   1.266   2.532  1.266  31.646  0.000
2    7.429  1.714  7.429   1.714   2.857   7.429  3.429   8.571  1.143
3    6.934  6.569  0.000   6.204   0.365  20.438  0.000   2.190  1.825
4    6.190  1.429  4.286   2.857   5.714  11.429  1.429   1.905  1.429
..     …      …      …       …       …      …      …       …      …
95   5.000  0.556  7.222   6.111   2.778   3.333  3.889   6.667  1.667
96   6.289  0.629  3.459  12.579   4.088  21.069  4.088   4.717  0.943
97   6.512  0.465  9.767   4.186   0.000   6.977  3.256   6.977  2.326
98   0.000  1.042  3.125   3.125   2.083  11.458  0.000  13.542  0.000
99  10.145  0.000  8.696   5.797  11.594   0.000  7.246   1.449  0.000
```

```
        X9  …     X11    X12     X13    X14     X15     X16    X17    X18  \
0   12.069  …   5.172  1.149   3.448  9.195   6.897   2.299  0.575  3.448
1    3.797  …   3.797  1.266   7.595  1.266  16.456   0.000  1.266  3.797
2    8.571  …   9.714  2.286   4.571  2.857   5.714  10.286  0.000  1.143
3    4.745  …  20.803  0.730   0.365  4.745   4.380  13.139  0.000  0.365
4    6.190  …  10.476  1.429   3.810  5.238   7.143  15.238  0.952  5.238
..     …  …     …    …     …    …     …    …    …    …
95   6.667  …  10.556  3.889   4.444  5.556   7.778   6.667  0.000  2.222
96   2.201  …  16.352  1.258   1.258  2.516   6.604   2.830  0.314  0.629
97   6.977  …  10.233  1.860   8.837  3.256   4.186   3.256  0.465  4.651
98   3.125  …  14.583  3.125  10.417  4.167   5.208   3.125  0.000  2.083
99  11.594  …  10.145  1.449   2.899  4.348  10.145   1.449  0.000  1.449

      X19         X20
0   2.874   efectores
1   2.532   efectores
2   5.714   efectores
3   4.015   efectores
4   3.333   efectores
..    …         …
95  9.444   efectores
96  2.516   efectores
97  6.047   efectores
98  8.333   efectores
99  1.449   efectores

[100 rows x 21 columns]


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

                X0          X1          X2          X3         X4          X5  \
count  100.000000  100.000000  100.000000  100.000000  100.00000  100.00000
mean     5.567740    2.934040    6.294810    4.758190    2.95682    6.53192
std      2.939291    2.413183    3.110864    2.788327    3.40403    5.92157
min      0.000000    0.000000    0.000000    0.000000    0.00000    0.00000
25%      3.303750    1.200250    4.017250    2.775500    0.63825    2.77800
50%      5.204000    2.213500    6.222500    4.456000    2.15100    4.44450
75%      7.333750    4.074500    8.107000    6.009750    3.51250    8.38950
max     14.000000   13.600000   14.667000   13.253000   18.34900   34.94000

                X6          X7          X8          X9         X10         X11  \
count  100.000000  100.000000  100.000000  100.000000  100.000000  100.000000
mean     3.482180    9.018310    1.534920    6.713010    8.200210    8.908050
std      2.717604    6.948871    1.257462    2.933336    3.559774    4.174185
```

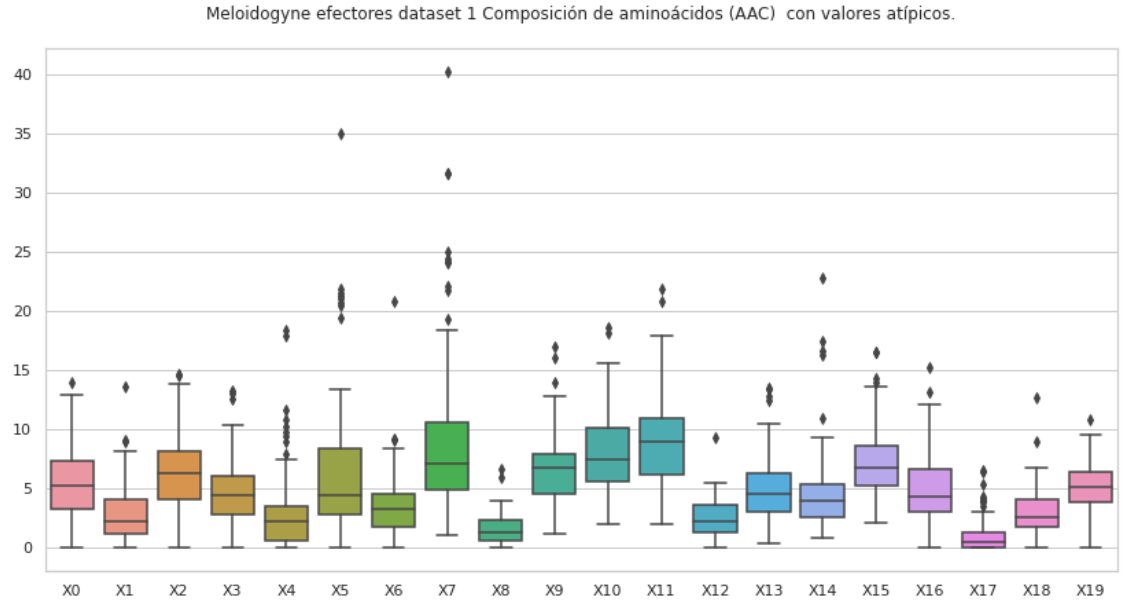
```
min      0.000000    1.031000    0.000000    1.149000    1.961000    1.961000
25%      1.740750    4.895250    0.634500    4.571250    5.634000    6.103500
50%      3.288500    7.060000    1.315500    6.726500    7.407500    8.929000
75%      4.586750   10.632250    2.290750    7.844000   10.069250   10.918750
max     20.803000   40.196000    6.593000   17.021000   18.557000   21.858000

              X12         X13         X14         X15         X16         X17  \
count  100.000000  100.000000  100.00000  100.000000  100.000000  100.000000
mean     2.509160    4.994650    4.59916    7.081730    5.019710    0.984940
std      1.573719    2.893151    3.49829    2.904576    2.977821    1.432658
min      0.000000    0.365000    0.82600    2.128000    0.000000    0.000000
25%      1.266000    3.061000    2.51975    5.184500    3.007750    0.000000
50%      2.260000    4.494500    3.90600    6.742500    4.348000    0.458500
75%      3.571000    6.250000    5.33900    8.578750    6.667000    1.317250
max      9.302000   13.514000   22.78900   16.456000   15.238000    6.494000

              X18         X19
count  100.000000  100.000000
mean     2.867530    5.042900
std      2.022923    2.228547
min      0.000000    0.000000
25%      1.724000    3.790750
50%      2.521500    5.120500
75%      4.040000    6.377250
max     12.644000   10.811000



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

       X0     X1      X2      X3     X4     X5      X6     X7     X8      X9  \
0   0.000  2.439   2.439   0.000  2.439  4.878   7.317  2.439  2.439  19.512
1   0.000  2.174  10.870   6.522  0.000  8.696   0.000  6.522  2.174   6.522
2   2.128  2.128   3.191   6.383  6.383  5.319   2.128  3.191  1.064   9.574
3   8.462  6.154   5.385  13.846  1.538  7.692   3.846  5.385  0.769   4.615
4   7.843  2.941   4.902   2.941  0.000  5.882   9.804  5.882  0.000   7.843
..    …      …       …       …      …      …       …      …      …
95  3.618  4.605   5.592   7.237  1.316  4.934   7.237  8.224  1.645   6.908
96  0.000  3.333   8.333   0.000  0.000  0.000   0.000  0.000  1.667  15.000
97  0.000  2.913  14.563   2.913  2.913  5.825  11.650  0.971  1.942   8.738
98  5.245  4.895   3.846   5.245  0.350  8.042   6.643  3.846  3.147   6.643
99  6.487  3.006   3.639   2.373  1.582  3.323   2.215  6.329  4.272   9.810

    …     X11    X12     X13    X14     X15    X16    X17    X18    X19  \
0   …  14.634  4.878   2.439  0.000   0.000  2.439  0.000  0.000  9.756
```

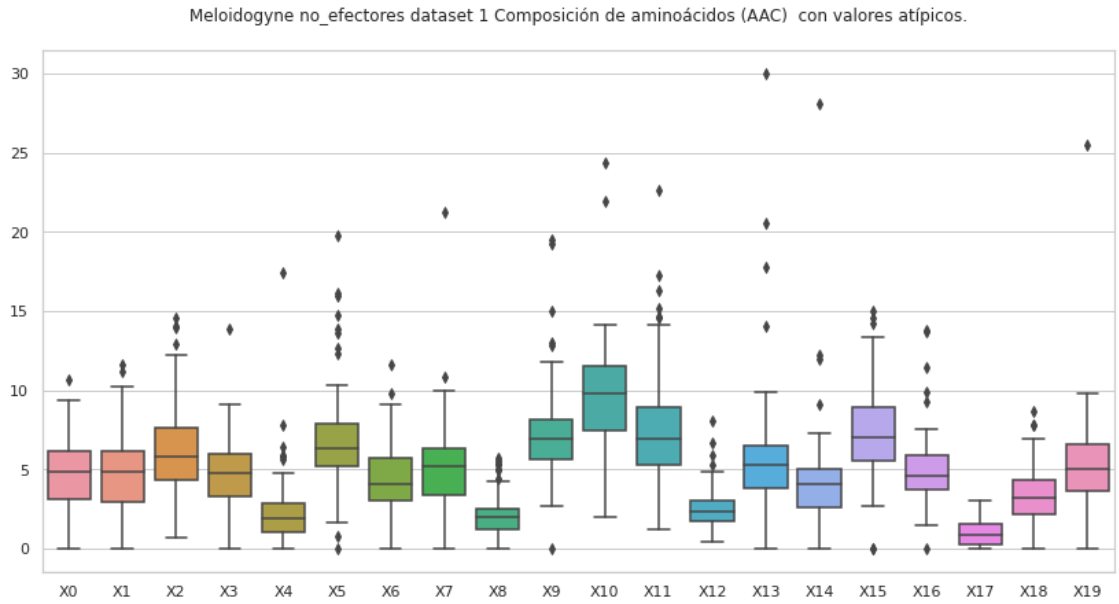
```
1   …  15.217  2.174   2.174  4.348   6.522  4.348  0.000  8.696  0.000
2   …   7.447  3.191   6.383  5.319   8.511  5.319  2.128  6.383  4.255
3   …   3.846  1.538   0.769  1.538   9.231  2.308  0.000  3.077  6.923
4   …   4.902  1.961   8.824  4.902  10.784  3.922  1.961  0.980  1.961
..  …     …    …     …    …     …     …    …    …    …
95  …   5.263  3.618   5.592  4.276   8.224  7.566  0.329  3.289  5.592
96  …   1.667  1.667  30.000  1.667  13.333  1.667  0.000  1.667  8.333
97  …   7.767  0.971   1.942  2.913  11.650  5.825  0.000  1.942  3.883
98  …  11.189  2.098   2.797  4.196   8.741  3.846  1.049  2.797  9.441
99  …   4.430  3.323   8.861  3.956   8.703  4.747  2.373  3.956  6.487

              X20
0    no_efectores
1    no_efectores
2    no_efectores
3    no_efectores
4    no_efectores
..            …
95   no_efectores
96   no_efectores
97   no_efectores
98   no_efectores
99   no_efectores

[100 rows x 21 columns]
```

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

```
                X0          X1         X2          X3          X4          X5  \
count   100.000000  100.000000  100.00000  100.000000  100.000000  100.000000
mean      4.671980    4.823100    6.18562    4.737090    2.220050    6.871710
std       2.449523    2.299882    2.69853    2.122221    2.141266    3.181203
min       0.000000    0.000000    0.67100    0.000000    0.000000    0.000000
25%       3.109000    2.974000    4.33950    3.311000    1.081250    5.236250
50%       4.832000    4.841000    5.78450    4.785500    1.945500    6.308500
75%       6.126250    6.141750    7.66775    6.011250    2.871250    7.895000
max      10.681000   11.628000   14.56300   13.846000   17.391000   19.792000

                X6          X7          X8          X9         X10         X11  \
count   100.000000  100.000000  100.000000  100.000000  100.000000  100.000000
mean      4.376980    5.143050    2.086760    7.280340    9.737850    7.546370
std       2.029462    2.692629    1.250005    2.871879    3.241566    3.730272
min       0.000000    0.000000    0.000000    0.000000    1.976000    1.227000
25%       3.069250    3.380750    1.227000    5.649000    7.461250    5.263000
50%       4.033000    5.188000    2.005500    6.915500    9.808000    6.930500
```

```
75%      5.759500    6.319000    2.493750    8.180250   11.550000    8.958250
max     11.650000   21.277000    5.747000   19.512000   24.359000   22.667000

              X12         X13         X14         X15         X16         X17  \
count  100.000000  100.000000  100.000000  100.000000  100.000000  100.000000
mean     2.543410    5.711070    4.127260    7.377280    4.924870    0.991500
std      1.288321    3.833061    3.232869    2.869861    2.171599    0.818243
min      0.395000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      1.700000    3.807250    2.599000    5.588250    3.728000    0.270000
50%      2.371500    5.263000    4.088000    7.059500    4.584000    0.886000
75%      3.008500    6.490250    5.018750    8.973250    5.933250    1.599000
max      8.065000   30.000000   28.063000   14.972000   13.780000    3.067000

              X18         X19
count  100.00000  100.000000
mean     3.32695    5.316760
std      1.70421    2.997238
min      0.00000    0.000000
25%      2.17075    3.627500
50%      3.17400    5.034500
75%      4.31575    6.621000
max      8.69600   25.503000
```





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 Meloidogyne dataset 1, sin valores
atípicos.
Valores del documento csv.

        X0     X1     X2      X3      X4      X5     X6      X7     X8  \
0    6.897  4.023  4.023   2.299   7.471   2.874  3.448   9.770  2.299
2    7.429  1.714  7.429   1.714   2.857   7.429  3.429   8.571  1.143
3    6.934  6.569  0.000   6.204   0.365  20.438  0.000   2.190  1.825
5    6.322  1.724  5.747   8.621   2.874   2.299  1.149   5.747  0.575
6   11.218  0.641  4.808   8.654   0.000   7.372  3.846   4.808  1.603
..     …    …    …     …       …       …    …     …      …
95   5.000  0.556  7.222   6.111   2.778   3.333  3.889   6.667  1.667
96   6.289  0.629  3.459  12.579   4.088  21.069  4.088   4.717  0.943
97   6.512  0.465  9.767   4.186   0.000   6.977  3.256   6.977  2.326
98   0.000  1.042  3.125   3.125   2.083  11.458  0.000  13.542  0.000
99  10.145  0.000  8.696   5.797  11.594   0.000  7.246   1.449  0.000

        X9  …     X11    X12     X13    X14     X15     X16    X17    X18  \
0   12.069  …   5.172  1.149   3.448  9.195   6.897   2.299  0.575  3.448
2    8.571  …   9.714  2.286   4.571  2.857   5.714  10.286  0.000  1.143
3    4.745  …  20.803  0.730   0.365  4.745   4.380  13.139  0.000  0.365
5    7.471  …  10.345  5.172   5.747  2.874   7.471  12.069  0.000  1.724
6    3.526  …  17.949  5.128   1.603  8.654   5.449   4.487  0.000  2.564
..     …  …     …    …     …       …    …     …      …
95   6.667  …  10.556  3.889   4.444  5.556   7.778   6.667  0.000  2.222
```

```
96   2.201  …  16.352  1.258   1.258  2.516   6.604   2.830  0.314  0.629
97   6.977  …  10.233  1.860   8.837  3.256   4.186   3.256  0.465  4.651
98   3.125  …  14.583  3.125  10.417  4.167   5.208   3.125  0.000  2.083
99  11.594  …  10.145  1.449   2.899  4.348  10.145   1.449  0.000  1.449

      X19        X20
0   2.874  efectores
2   5.714  efectores
3   4.015  efectores
5   5.747  efectores
6   4.808  efectores
..    …        …
95  9.444  efectores
96  2.516  efectores
97  6.047  efectores
98  8.333  efectores
99  1.449  efectores

[78 rows x 21 columns]
```

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

```
              X0         X1         X2         X3         X4         X5  \
count  78.000000  78.000000  78.000000  78.000000  78.000000  78.000000
mean    5.854167   2.754821   6.823974   5.031615   2.783462   6.867513
std     2.739625   1.987811   2.863793   2.551107   2.647225   5.065803
min     0.000000   0.000000   0.000000   0.000000   0.000000   0.000000
25%     4.201500   1.186000   5.362750   3.501000   0.695750   3.146250
50%     5.505500   2.075000   6.999000   5.000000   2.701000   5.648500
75%     7.397250   4.057000   8.174500   6.040000   3.551500   8.628750
max    14.000000   8.145000  14.667000  13.016000  11.594000  21.515000

              X6         X7         X8         X9        X10        X11  \
count  78.000000  78.000000  78.000000  78.000000  78.000000  78.000000
mean    3.431167   7.424885   1.642115   6.849231   8.257487   9.468756
std     2.046629   4.527084   1.017198   2.583696   3.504864   3.811352
min     0.000000   1.031000   0.000000   1.905000   2.190000   2.174000
25%     1.786000   4.728250   0.893000   4.808750   5.773500   6.740750
50%     3.419000   6.667000   1.635000   7.032500   7.546000   9.569000
75%     4.545000   9.372500   2.319250   8.377000  10.119750  11.111000
max     9.044000  25.000000   3.933000  13.978000  18.557000  20.803000

             X12        X13        X14        X15        X16        X17  \
count  78.000000  78.000000  78.000000  78.000000  78.000000  78.000000
mean    2.434192   5.050910   4.113769   6.774718   5.435000   0.838769
```

std	1.393912	2.908362	1.871240	2.219453	2.757104	1.131489
min	0.000000	0.365000	0.826000	2.532000	0.985000	0.000000
25%	1.261000	3.081000	2.708250	5.236000	3.275250	0.000000
50%	2.292500	4.130500	3.909500	6.620000	4.785000	0.458500
75%	3.563250	6.227000	5.257250	7.768000	6.861750	1.268000
max	5.369000	13.514000	9.278000	13.636000	13.139000	4.082000

	X18	X19
count	78.000000	78.000000
mean	2.801936	5.361513
std	1.583119	2.187300
min	0.000000	1.075000
25%	1.786000	3.953500
50%	2.573500	5.477000
75%	3.972250	6.772250
max	6.731000	10.811000

no_efectores

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

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9	…	\
2	2.128	2.128	3.191	6.383	6.383	5.319	2.128	3.191	1.064	9.574	…	
4	7.843	2.941	4.902	2.941	0.000	5.882	9.804	5.882	0.000	7.843	…	
5	7.675	2.632	4.825	2.412	3.070	3.509	3.070	6.360	1.535	8.333	…	
7	5.325	4.438	6.213	5.325	0.888	5.917	2.959	3.846	2.071	9.172	…	
8	3.374	6.135	3.374	3.681	2.454	5.828	3.067	4.908	1.840	5.828	…	
..	…	…	…	…	…	…	…	…	…	…		
93	5.807	3.116	6.657	4.391	1.983	5.099	1.841	7.649	1.133	6.657	…	
94	4.992	5.304	6.864	4.680	1.404	4.056	5.928	5.304	2.496	5.616	…	
95	3.618	4.605	5.592	7.237	1.316	4.934	7.237	8.224	1.645	6.908	…	
98	5.245	4.895	3.846	5.245	0.350	8.042	6.643	3.846	3.147	6.643	…	
99	6.487	3.006	3.639	2.373	1.582	3.323	2.215	6.329	4.272	9.810	…	

	X11	X12	X13	X14	X15	X16	X17	X18	X19	\
2	7.447	3.191	6.383	5.319	8.511	5.319	2.128	6.383	4.255	
4	4.902	1.961	8.824	4.902	10.784	3.922	1.961	0.980	1.961	
5	3.509	3.947	7.018	5.263	4.605	5.921	1.316	3.947	8.333	
7	7.988	0.888	7.692	5.030	7.101	4.734	1.775	1.479	6.805	
8	3.988	3.067	6.135	6.135	8.282	7.362	2.147	4.601	4.294	
..	…	…	…	…	…	…	…	…		
93	4.249	4.108	8.215	4.249	7.507	5.099	2.125	4.108	6.232	
94	7.176	2.340	5.928	4.680	6.084	3.900	2.340	5.148	4.992	
95	5.263	3.618	5.592	4.276	8.224	7.566	0.329	3.289	5.592	
98	11.189	2.098	2.797	4.196	8.741	3.846	1.049	2.797	9.441	

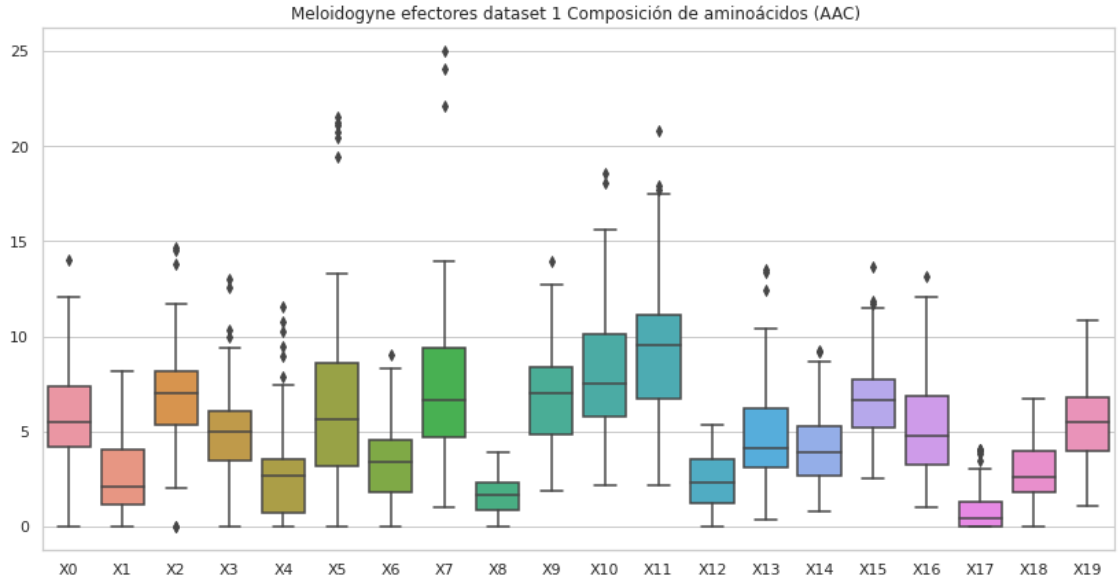
```
99   4.430  3.323  8.861  3.956   8.703  4.747  2.373  3.956  6.487

             X20
2   no_efectores
4   no_efectores
5   no_efectores
7   no_efectores
8   no_efectores
..           …
93  no_efectores
94  no_efectores
95  no_efectores
98  no_efectores
99  no_efectores

[82 rows x 21 columns]
```

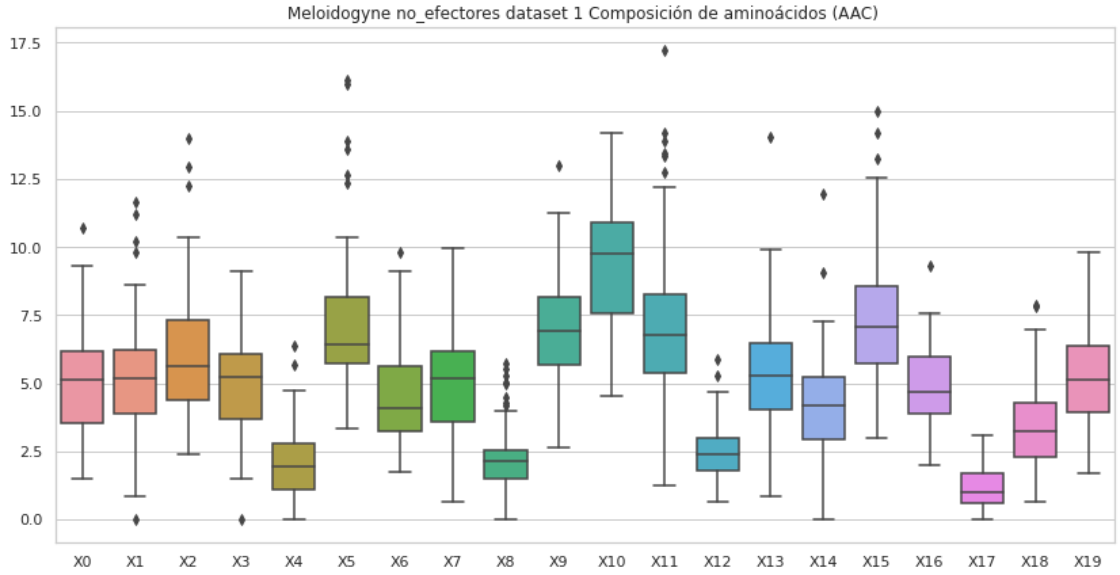
Composición de aminoácidos (AAC) no_efectores Meloidogyne dataset 1, sin valores atípicos.
Estadísticas.

```
              X0         X1         X2         X3         X4         X5  \
count  82.000000  82.000000  82.000000  82.000000  82.000000  82.000000
mean    5.072110   5.283854   6.026390   5.022841   2.022207   7.139683
std     1.987082   2.174013   2.327751   1.684025   1.290951   2.614023
min     1.493000   0.000000   2.381000   0.000000   0.000000   3.323000
25%     3.536000   3.900750   4.401500   3.701500   1.098000   5.713250
50%     5.120500   5.197500   5.623500   5.224500   1.945500   6.412500
75%     6.160750   6.225750   7.340750   6.079000   2.784000   8.186750
max    10.681000  11.628000  14.000000   9.121000   6.383000  16.126000

              X6         X7         X8         X9        X10        X11  \
count  82.000000  82.000000  82.000000  82.000000  82.000000  82.000000
mean    4.460256   5.007659   2.202049   7.028695   9.418683   7.137329
std     1.681217   1.811966   1.215915   1.915589   2.335870   2.937887
min     1.754000   0.649000   0.000000   2.649000   4.525000   1.227000
25%     3.246000   3.584250   1.491250   5.682500   7.564750   5.362750
50%     4.085500   5.153000   2.117000   6.915500   9.754000   6.787000
75%     5.608000   6.185750   2.551750   8.143750  10.907500   8.238250
max     9.804000   9.934000   5.747000  12.981000  14.179000  17.219000

             X12        X13       X14        X15        X16        X17  \
count  82.000000  82.000000  82.00000  82.000000  82.000000  82.000000
mean    2.497573   5.493878   4.15972   7.390012   4.889451   1.114939
std     1.057698   1.996134   1.89291   2.407261   1.419507   0.787278
min     0.658000   0.838000   0.00000   2.979000   2.000000   0.000000
25%     1.805750   4.041750   2.92525   5.747000   3.900000   0.590500
```

50%	2.407500	5.299000	4.17550	7.059500	4.692500	1.000000
75%	2.989000	6.459500	5.20850	8.578000	5.957750	1.697750
max	5.882000	14.062000	11.94000	14.972000	9.302000	3.067000

	X18	X19
count	82.000000	82.000000
mean	3.376500	5.256085
std	1.608071	1.971044
min	0.649000	1.683000
25%	2.271250	3.908250
50%	3.223000	5.111000
75%	4.267250	6.367250
max	7.843000	9.804000





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 Meloidogyne
dataset 1, con valores atípicos.
Valores del documento csv.
```

	X0	X1	X2	X3	X4	X5	X6	\
0	0.018897	0.020471	0.006299	0.007874	0.009448	0.026770	0.006299	
1	0.001621	0.001621	0.003243	0.003243	0.009729	0.040537	0.000000	
2	0.027096	0.010422	0.006253	0.027096	0.016674	0.031265	0.004169	
3	0.005799	0.000305	0.005189	0.017092	0.000305	0.001831	0.001526	
4	0.021134	0.019508	0.009754	0.039017	0.013006	0.006503	0.004877	
..	…	…	…	…	…	…	…	
95	0.020909	0.011616	0.025555	0.013939	0.018585	0.027878	0.006970	
96	0.004702	0.003056	0.009403	0.015751	0.000940	0.003526	0.000705	
97	0.039954	0.000000	0.025685	0.042808	0.054223	0.042808	0.014269	
98	0.000000	0.002731	0.004097	0.015021	0.013655	0.017752	0.000000	
99	0.029341	0.033533	0.016766	0.000000	0.008383	0.004192	0.000000	

	X7	X8	X9	…	X74	X75	X76	X77	\
0	0.033069	0.014173	0.026770	…	0.005232	0.001630	0.032979	0.023946	
1	0.004864	0.004864	0.008107	…	-0.005145	0.001681	0.039137	-0.006813	
2	0.031265	0.035433	0.027096	…	0.015019	0.018419	0.024560	0.012424	
3	0.003968	0.017397	0.001831	…	0.020590	0.035614	0.002425	0.009132	
4	0.021134	0.035766	0.014631	…	-0.005189	-0.005877	0.006352	-0.013282	
..	…	…	…	…	…	…	…	…	
95	0.027878	0.044140	0.023232	…	0.019263	0.029422	0.018502	-0.001569	
96	0.001646	0.012224	0.004231	…	0.009058	0.032233	0.003119	0.006304	
97	0.042808	0.062784	0.059931	…	0.009246	0.013373	0.003259	-0.006087	
98	0.004097	0.019117	0.015021	…	0.005937	0.008705	0.013505	0.009508	
99	0.033533	0.029341	0.029341	…	0.010907	-0.005243	0.011315	0.015574	

	X78	X79	X80	X81	X82	X83
0	0.017041	0.023175	0.014647	0.012149	0.007834	efectores
1	-0.005309	0.023253	-0.006938	0.003013	0.037321	efectores
2	-0.006581	0.035818	-0.028596	-0.012169	0.021654	efectores
3	0.029067	-0.001383	0.008384	0.028439	-0.001011	efectores
4	-0.011849	0.014591	-0.000959	0.002831	0.007688	efectores
..	…	…	…	…	…	
95	0.009509	0.021101	0.017714	0.026309	0.028042	efectores
96	0.032769	0.001198	0.008117	0.032827	0.000697	efectores
97	0.020874	0.019769	-0.033489	-0.035663	0.025476	efectores

```
98  0.009133  0.004499  0.003172  0.005083 -0.007894  efectores
99  0.025122  0.017542 -0.001451 -0.015658  0.027093  efectores

[100 rows x 84 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	
mean	0.019270	0.010919	0.018609	0.021691	0.020699	0.028094	
std	0.017840	0.012906	0.016664	0.023106	0.022861	0.022491	
min	-0.082991	-0.020748	-0.020748	-0.062243	-0.062243	-0.082991	
25%	0.008584	0.001145	0.007208	0.008526	0.008044	0.014674	
50%	0.018681	0.007059	0.012735	0.016512	0.016469	0.028012	
75%	0.027134	0.017122	0.025707	0.027891	0.024652	0.036329	
max	0.083706	0.053948	0.077366	0.125559	0.132534	0.104632	

	X6	X7	X8	X9	…	X73	\
count	100.000000	100.000000	100.000000	100.000000	…	100.000000	
mean	0.006629	0.025442	0.031468	0.034053	…	0.016251	
std	0.008574	0.027478	0.026012	0.034033	…	0.021489	
min	-0.020748	-0.165982	-0.103739	-0.124487	…	-0.086561	
25%	0.001319	0.012437	0.017470	0.014744	…	0.002076	
50%	0.004159	0.025066	0.028315	0.025405	…	0.017137	
75%	0.009667	0.036513	0.044278	0.048428	…	0.029813	
max	0.051111	0.087619	0.132534	0.125559	…	0.084253	

	X74	X75	X76	X77	X78	X79	\
count	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	
mean	0.001514	0.006323	0.012776	0.001227	0.003721	0.015296	
std	0.033968	0.035260	0.026379	0.044536	0.044969	0.023408	
min	-0.110165	-0.061263	-0.075722	-0.082236	-0.122913	-0.115578	
25%	-0.007435	-0.007343	0.002913	-0.010093	-0.008052	0.002701	
50%	0.003147	0.008289	0.010937	0.000074	0.001027	0.015312	
75%	0.010923	0.020239	0.026189	0.007687	0.015752	0.028889	
max	0.226319	0.239278	0.181337	0.388708	0.388600	0.065022	

	X80	X81	X82
count	100.000000	100.000000	100.000000
mean	0.000834	0.005563	0.017972
std	0.035749	0.032044	0.040482
min	-0.066864	-0.062812	-0.040528
25%	-0.019623	-0.007945	0.001303
50%	0.002580	0.003034	0.015245
75%	0.013729	0.018546	0.028876

```
max      0.254767    0.203697    0.358751

[8 rows x 83 columns]
```

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.000000	0.010767	0.000000	0.021534	0.010767	0.010767	0.010767	
1	0.000000	0.000000	0.024738	0.032984	0.008246	0.024738	0.008246	
2	0.027376	0.082127	0.082127	0.068439	0.082127	0.041063	0.013688	
3	0.023702	0.004309	0.038785	0.021547	0.002155	0.015083	0.002155	
4	0.053496	0.000000	0.020061	0.040122	0.060183	0.040122	0.000000	
..	…	…	…	…	…	…	…	
95	0.020329	0.007392	0.040658	0.027722	0.031418	0.046203	0.009241	
96	0.000000	0.000000	0.000000	0.000000	0.038709	0.000000	0.002150	
97	0.000000	0.019159	0.019159	0.038318	0.012773	0.006386	0.012773	
98	0.028969	0.001931	0.028969	0.044419	0.015450	0.021244	0.017381	
99	0.021294	0.005194	0.007790	0.010907	0.029084	0.020775	0.014023	

	X7	X8	X9	…	X74	X75	X76	X77	\
0	0.086136	0.064602	0.096903	…	0.012368	-0.035435	0.014761	0.032928	
1	0.024738	0.057722	0.049476	…	-0.043658	0.022491	-0.003875	-0.034769	
2	0.123190	0.095814	0.123190	…	0.036434	-0.059022	-0.067367	0.081901	
3	0.012928	0.010774	0.036631	…	0.019149	0.015768	-0.005927	-0.005821	
4	0.053496	0.033435	0.080244	…	-0.004963	-0.010490	0.056910	0.044671	
..	…	…	…	…	…	…	…	…	
95	0.038810	0.029570	0.027722	…	0.007145	-0.003367	0.022315	0.003470	
96	0.019354	0.002150	0.015053	…	0.020753	0.022860	0.007886	0.030559	
97	0.057477	0.051091	0.070250	…	0.063442	0.047391	0.000350	0.071597	
98	0.036694	0.061801	0.032832	…	0.012403	0.034451	0.008253	-0.004520	
99	0.032201	0.014542	0.033239	…	0.017821	0.005531	0.009075	0.010625	

	X78	X79	X80	X81	X82	X83
0	0.003735	-0.000388	0.069991	0.054803	0.009256	no_efectores
1	-0.053898	0.006316	-0.007485	-0.009578	0.009507	no_efectores
2	0.026905	-0.102559	-0.008703	-0.032984	0.029318	no_efectores
3	0.012026	0.007474	-0.020082	0.000567	-0.002711	no_efectores
4	-0.004226	0.107305	0.017937	0.009063	0.009968	no_efectores
..	…	…	…	…	…	…
95	-0.019952	0.031754	0.007861	0.008541	0.014522	no_efectores
96	0.027779	0.005721	0.023201	0.015169	-0.008566	no_efectores
97	0.044387	0.011437	0.033746	-0.019844	-0.005951	no_efectores
98	0.010058	-0.001571	0.002078	0.028342	0.005033	no_efectores

```
99  0.005327  0.012142  0.011757  0.003110  0.009496  no_efectores

[100 rows x 84 columns]
```

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

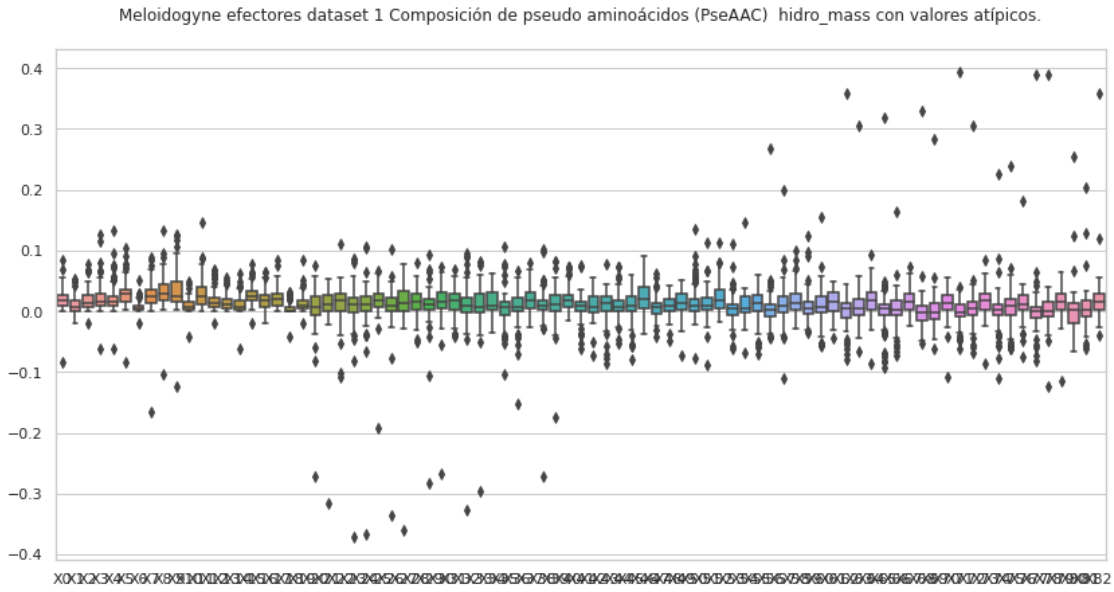
	X0	X1	X2	X3	X4	X5	\
count	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	
mean	0.025085	0.012905	0.027323	0.037054	0.032684	0.026748	
std	0.015790	0.013959	0.017563	0.020033	0.027044	0.014009	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.015285	0.004744	0.013834	0.023850	0.015597	0.016577	
50%	0.023780	0.009482	0.024628	0.034062	0.026259	0.024791	
75%	0.034648	0.016491	0.040545	0.049003	0.043468	0.038117	
max	0.074008	0.082127	0.082127	0.107533	0.215176	0.069266	

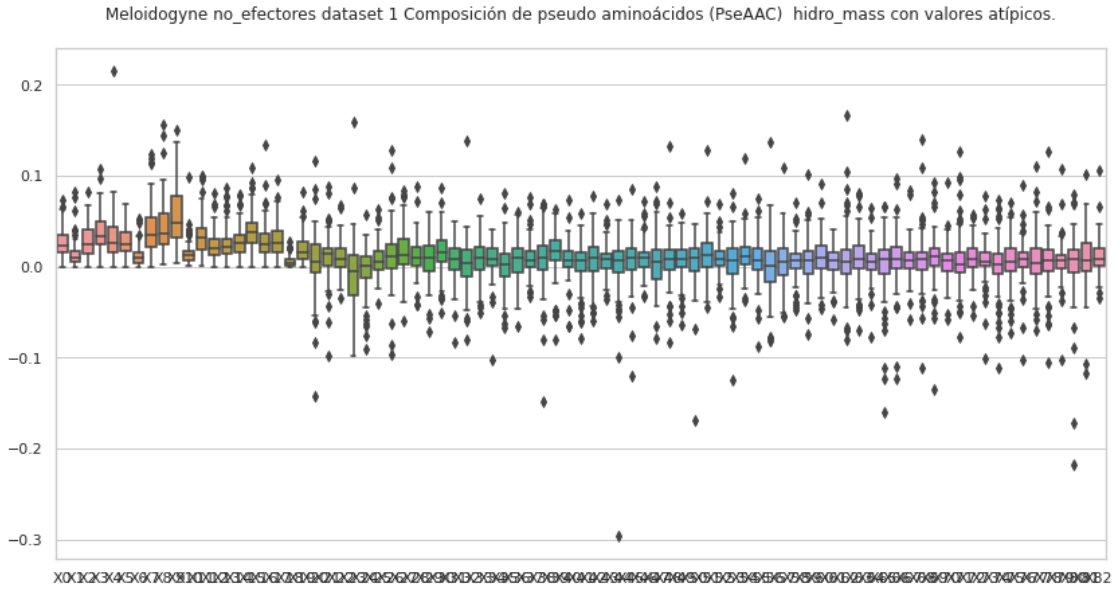
	X6	X7	X8	X9	…	X73	\
count	100.000000	100.000000	100.000000	100.000000	…	100.000000	
mean	0.012115	0.041042	0.042881	0.053953	…	0.006441	
std	0.010207	0.025486	0.028804	0.030077	…	0.021038	
min	0.000000	0.000000	0.002150	0.003777	…	-0.100727	
25%	0.004462	0.022317	0.024846	0.031297	…	0.000568	
50%	0.009934	0.034607	0.036989	0.047827	…	0.005633	
75%	0.016200	0.054780	0.057987	0.077256	…	0.016114	
max	0.053283	0.123190	0.156492	0.150458	…	0.078423	

	X74	X75	X76	X77	X78	X79	\
count	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	
mean	0.000332	0.005357	0.007291	0.005327	0.006269	0.006744	
std	0.028034	0.024556	0.021202	0.028002	0.028047	0.021311	
min	-0.110976	-0.073302	-0.102082	-0.070398	-0.105779	-0.102559	
25%	-0.007651	-0.004572	-0.000862	-0.007660	-0.004813	-0.001550	
50%	0.001915	0.005000	0.008521	0.003664	0.007479	0.006341	
75%	0.013898	0.019494	0.015348	0.020600	0.018475	0.013412	
max	0.072773	0.070119	0.089752	0.110581	0.126092	0.107305	

	X80	X81	X82
count	100.000000	100.000000	100.000000
mean	0.004069	0.008308	0.010332
std	0.038506	0.028283	0.018097
min	-0.218392	-0.117785	-0.034952
25%	-0.006888	-0.005065	0.000755
50%	0.008434	0.006693	0.008361
75%	0.019077	0.026263	0.019912
max	0.078625	0.102095	0.105515

[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 Meloidogyne
dataset 1, sin valores atípicos.
Valores del documento csv.
```

	X0	X1	X2	X3	X4	X5	X6	\
0	0.018897	0.020471	0.006299	0.007874	0.009448	0.026770	0.006299	
1	0.001621	0.001621	0.003243	0.003243	0.009729	0.040537	0.000000	
2	0.027096	0.010422	0.006253	0.027096	0.016674	0.031265	0.004169	
3	0.005799	0.000305	0.005189	0.017092	0.000305	0.001831	0.001526	
4	0.021134	0.019508	0.009754	0.039017	0.013006	0.006503	0.004877	
..	…	…	…	…	…	…	…	
95	0.020909	0.011616	0.025555	0.013939	0.018585	0.027878	0.006970	
96	0.004702	0.003056	0.009403	0.015751	0.000940	0.003526	0.000705	
97	0.039954	0.000000	0.025685	0.042808	0.054223	0.042808	0.014269	
98	0.000000	0.002731	0.004097	0.015021	0.013655	0.017752	0.000000	
99	0.029341	0.033533	0.016766	0.000000	0.008383	0.004192	0.000000	

	X7	X8	X9	…	X74	X75	X76	X77	\
0	0.033069	0.014173	0.026770	…	0.005232	0.001630	0.032979	0.023946	
1	0.004864	0.004864	0.008107	…	-0.005145	0.001681	0.039137	-0.006813	
2	0.031265	0.035433	0.027096	…	0.015019	0.018419	0.024560	0.012424	
3	0.003968	0.017397	0.001831	…	0.020590	0.035614	0.002425	0.009132	
4	0.021134	0.035766	0.014631	…	-0.005189	-0.005877	0.006352	-0.013282	
..	…	…	…	…	…	…	…	…	
95	0.027878	0.044140	0.023232	…	0.019263	0.029422	0.018502	-0.001569	
96	0.001646	0.012224	0.004231	…	0.009058	0.032233	0.003119	0.006304	
97	0.042808	0.062784	0.059931	…	0.009246	0.013373	0.003259	-0.006087	
98	0.004097	0.019117	0.015021	…	0.005937	0.008705	0.013505	0.009508	
99	0.033533	0.029341	0.029341	…	0.010907	-0.005243	0.011315	0.015574	

	X78	X79	X80	X81	X82	X83
0	0.017041	0.023175	0.014647	0.012149	0.007834	efectores
1	-0.005309	0.023253	-0.006938	0.003013	0.037321	efectores
2	-0.006581	0.035818	-0.028596	-0.012169	0.021654	efectores
3	0.029067	-0.001383	0.008384	0.028439	-0.001011	efectores
4	-0.011849	0.014591	-0.000959	0.002831	0.007688	efectores
..	…	…	…	…	…	
95	0.009509	0.021101	0.017714	0.026309	0.028042	efectores
96	0.032769	0.001198	0.008117	0.032827	0.000697	efectores
97	0.020874	0.019769	-0.033489	-0.035663	0.025476	efectores
98	0.009133	0.004499	0.003172	0.005083	-0.007894	efectores
99	0.025122	0.017542	-0.001451	-0.015658	0.027093	efectores

[76 rows x 84 columns]

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

	X0	X1	X2	X3	X4	X5	\
count	76.000000	76.000000	76.000000	76.000000	76.000000	76.000000	
mean	0.018196	0.008867	0.015771	0.019557	0.015636	0.024764	
std	0.010589	0.010456	0.012656	0.014241	0.012331	0.013804	
min	0.000000	0.000000	0.000000	0.000000	0.000305	0.001831	
25%	0.009089	0.000720	0.006913	0.008630	0.007390	0.014596	
50%	0.018277	0.004914	0.011466	0.016512	0.013412	0.027521	
75%	0.026008	0.011975	0.023411	0.026181	0.020807	0.031820	
max	0.040409	0.049391	0.067215	0.066990	0.055115	0.061023	

	X6	X7	X8	X9	…	X73	X74	\
count	76.000000	76.000000	76.000000	76.000000	…	76.000000	76.000000	
mean	0.005607	0.022641	0.029015	0.026132	…	0.015700	0.004511	
std	0.005772	0.014918	0.017063	0.021144	…	0.015548	0.014043	
min	0.000000	0.001006	0.001350	0.001350	…	-0.019232	-0.041699	
25%	0.001601	0.007195	0.017395	0.009560	…	0.002076	-0.003661	
50%	0.003779	0.022975	0.026986	0.022864	…	0.017137	0.005410	
75%	0.007987	0.031624	0.037811	0.030163	…	0.028342	0.011850	
max	0.029389	0.058777	0.069798	0.088044	…	0.061959	0.037492	

	X75	X76	X77	X78	X79	X80	\
count	76.000000	76.000000	76.000000	76.000000	76.000000	76.000000	
mean	0.010758	0.014211	0.000156	0.005316	0.016244	-0.000627	
std	0.017369	0.014629	0.015269	0.016766	0.015575	0.016622	
min	-0.037523	-0.023311	-0.060083	-0.037520	-0.018925	-0.049454	
25%	0.001343	0.004254	-0.006971	-0.003574	0.004905	-0.008412	
50%	0.011061	0.011196	0.002377	0.004318	0.016788	0.003498	
75%	0.022366	0.025379	0.008769	0.017274	0.026305	0.012764	
max	0.055671	0.045388	0.031005	0.039354	0.062402	0.029720	

	X81	X82
count	76.000000	76.000000
mean	0.005447	0.015548
std	0.018638	0.015784
min	-0.052154	-0.025315
25%	-0.004721	0.002823
50%	0.004089	0.021160
75%	0.021073	0.027772
max	0.034333	0.048968

[8 rows x 83 columns]

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

          X0        X1        X2        X3        X4        X5        X6  \
0   0.000000  0.010767  0.000000  0.021534  0.010767  0.010767  0.010767
3   0.023702  0.004309  0.038785  0.021547  0.002155  0.015083  0.002155
5   0.020672  0.008269  0.006497  0.009450  0.018901  0.017129  0.004134
6   0.024645  0.016803  0.015683  0.002240  0.006721  0.025205  0.002240
7   0.024517  0.004086  0.024517  0.027241  0.035414  0.017707  0.009534
..       …         …         …         …         …         …
94  0.046178  0.012988  0.043292  0.037520  0.054836  0.049064  0.023089
95  0.020329  0.007392  0.040658  0.027722  0.031418  0.046203  0.009241
96  0.000000  0.000000  0.000000  0.000000  0.038709  0.000000  0.002150
98  0.028969  0.001931  0.028969  0.044419  0.015450  0.021244  0.017381
99  0.021294  0.005194  0.007790  0.010907  0.029084  0.020775  0.014023

          X7        X8        X9  …       X74       X75       X76       X77  \
0   0.086136  0.064602  0.096903  …  0.012368 -0.035435  0.014761  0.032928
3   0.012928  0.010774  0.036631  …  0.019149  0.015768 -0.005927 -0.005821
5   0.022444  0.009450  0.034257  …  0.014439  0.002376  0.013047  0.003058
6   0.011202  0.015123  0.010642  … -0.006208 -0.000506  0.034728 -0.005686
7   0.042224  0.036776  0.047672  …  0.000278 -0.004137  0.016281  0.023366
..       …         …         …  …         …         …         …         …
94  0.051950  0.066381  0.099571  … -0.007513 -0.015262  0.012498 -0.018919
95  0.038810  0.029570  0.027722  …  0.007145 -0.003367  0.022315  0.003470
96  0.019354  0.002150  0.015053  …  0.020753  0.022860  0.007886  0.030559
98  0.036694  0.061801  0.032832  …  0.012403  0.034451  0.008253 -0.004520
99  0.032201  0.014542  0.033239  …  0.017821  0.005531  0.009075  0.010625

          X78       X79       X80       X81       X82           X83
0   0.003735 -0.000388  0.069991  0.054803  0.009256  no_efectores
3   0.012026  0.007474 -0.020082  0.000567 -0.002711  no_efectores
5  -0.004150  0.010876  0.012735  0.000237  0.000628  no_efectores
6   0.000298  0.024108  0.002061  0.004410  0.029678  no_efectores
7   0.019434  0.010290  0.009526  0.009054  0.005132  no_efectores
..       …         …         …         …         …           …
94 -0.007742 -0.001876 -0.007982  0.026135 -0.010388  no_efectores
95 -0.019952  0.031754  0.007861  0.008541  0.014522  no_efectores
96  0.027779  0.005721  0.023201  0.015169 -0.008566  no_efectores
98  0.010058 -0.001571  0.002078  0.028342  0.005033  no_efectores
99  0.005327  0.012142  0.011757  0.003110  0.009496  no_efectores

[78 rows x 84 columns]
```

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

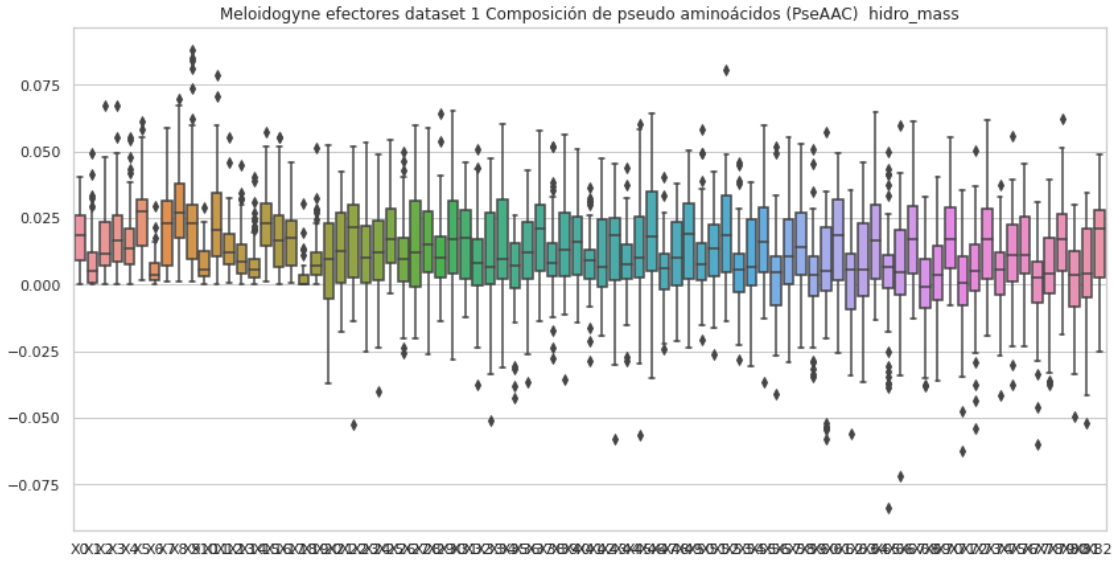
	X0	X1	X2	X3	X4	X5	\
count	78.000000	78.000000	78.000000	78.000000	78.000000	78.000000	
mean	0.023386	0.009757	0.024380	0.033125	0.026155	0.023802	
std	0.013141	0.006979	0.015796	0.015800	0.015252	0.012554	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.015439	0.004519	0.011705	0.022901	0.015335	0.014473	
50%	0.022983	0.008718	0.022801	0.031873	0.022783	0.022933	
75%	0.029179	0.014800	0.034375	0.044225	0.035372	0.031398	
max	0.063046	0.028141	0.067081	0.074363	0.060507	0.054295	

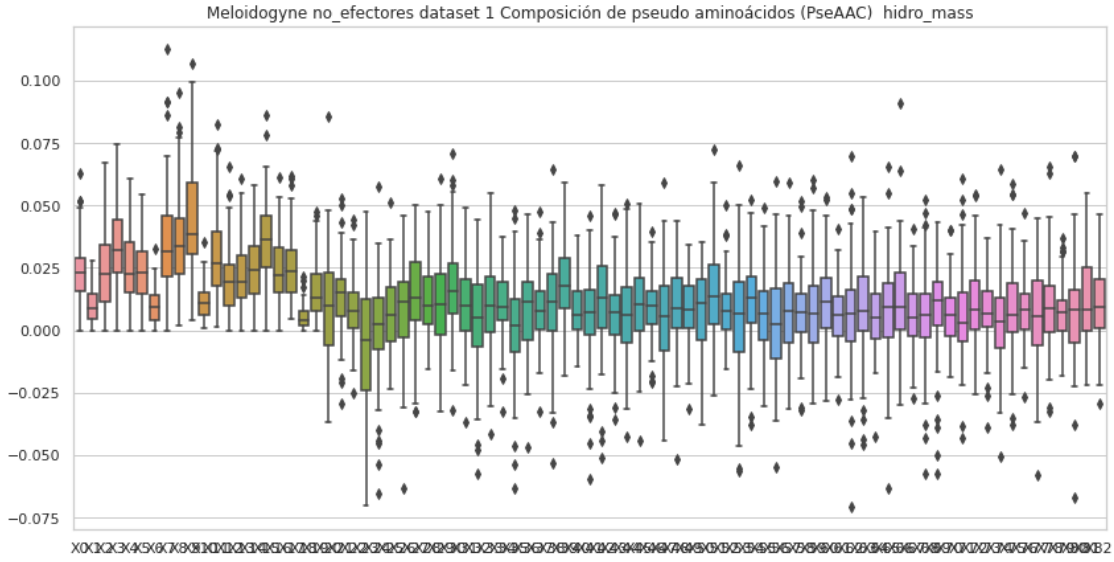
	X6	X7	X8	X9	…	X73	X74	\
count	78.000000	78.000000	78.000000	78.000000	…	78.000000	78.000000	
mean	0.009865	0.036068	0.035602	0.045684	…	0.007261	0.002126	
std	0.006724	0.020867	0.021101	0.023196	…	0.012961	0.017925	
min	0.000000	0.000000	0.002150	0.003777	…	-0.038688	-0.050333	
25%	0.004273	0.021534	0.022370	0.030265	…	0.001244	-0.007351	
50%	0.009161	0.031727	0.033563	0.038377	…	0.006605	0.003514	
75%	0.013881	0.046050	0.044885	0.059068	…	0.015864	0.013027	
max	0.032853	0.112637	0.095308	0.106720	…	0.036692	0.064247	

	X75	X76	X77	X78	X79	X80	\
count	78.000000	78.000000	78.000000	78.000000	78.000000	78.000000	
mean	0.008507	0.008470	0.005739	0.009383	0.007512	0.007693	
std	0.018804	0.011993	0.019159	0.017946	0.011336	0.021002	
min	-0.037822	-0.026609	-0.057936	-0.032675	-0.018088	-0.066856	
25%	-0.000880	0.000380	-0.005787	-0.000697	-0.000385	-0.005217	
50%	0.006052	0.008402	0.005790	0.008735	0.007003	0.008414	
75%	0.019061	0.014744	0.020031	0.018042	0.012013	0.016121	
max	0.058848	0.046871	0.044898	0.065631	0.036919	0.069991	

	X81	X82
count	78.000000	78.000000
mean	0.011205	0.009893
std	0.017537	0.014120
min	-0.021846	-0.029252
25%	-0.000060	0.000693
50%	0.008514	0.009373
75%	0.025426	0.020389
max	0.054803	0.046587

[8 rows x 83 columns]





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))
```

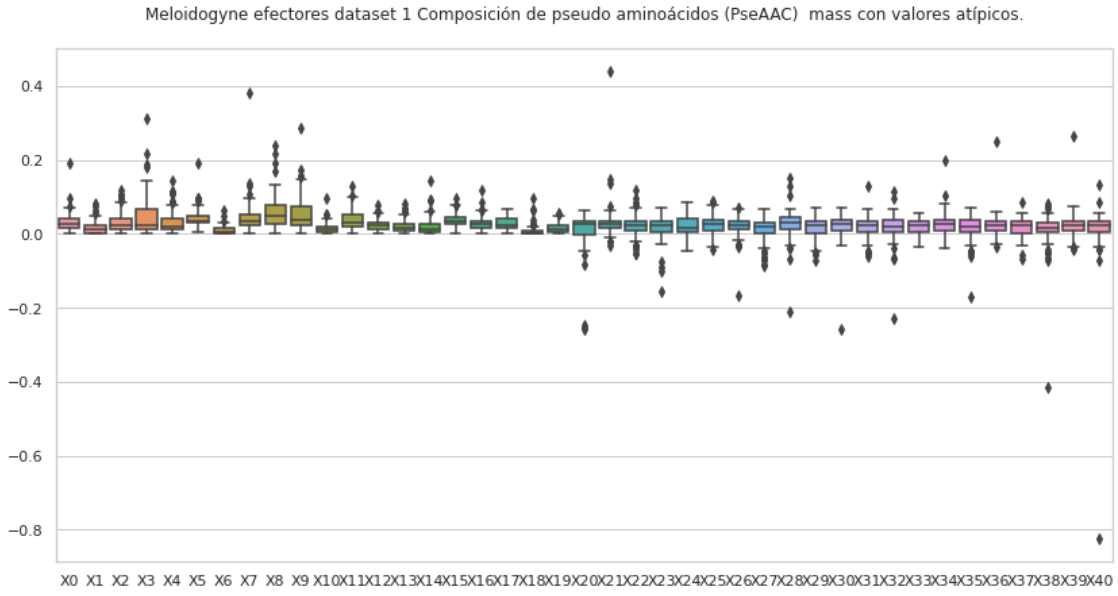
```
efectores
Composición de pseudo aminoácidos (PseAAC)  mass efectores Meloidogyne dataset
1, con valores atípicos.
Valores del documento csv.

          X0        X1        X2        X3        X4        X5        X6  \
0   0.026234  0.028421  0.008745  0.010931  0.013117  0.037165  0.008745
1   0.001686  0.001686  0.003371  0.003371  0.010114  0.042142  0.000000
2   0.028961  0.011139  0.006683  0.028961  0.017822  0.033417  0.004456
3   0.063467  0.003340  0.056787  0.187062  0.003340  0.020042  0.016702
4   0.034106  0.031482  0.015741  0.062965  0.020988  0.010494  0.007871
..       …         …         …         …         …         …         …
95  0.022473  0.012485  0.027467  0.014982  0.019976  0.029964  0.007491
96  0.043220  0.028093  0.086439  0.144786  0.008644  0.032415  0.006483
97  0.050265  0.000000  0.032313  0.053855  0.068216  0.053855  0.017952
98  0.000000  0.015036  0.022554  0.082698  0.075180  0.097734  0.000000
99  0.042656  0.048750  0.024375  0.000000  0.012187  0.006094  0.000000

          X7        X8        X9  …       X32       X33       X34       X35  \
0   0.045910  0.019676  0.037165  …  0.015633  0.035683  0.041154  0.026938
```

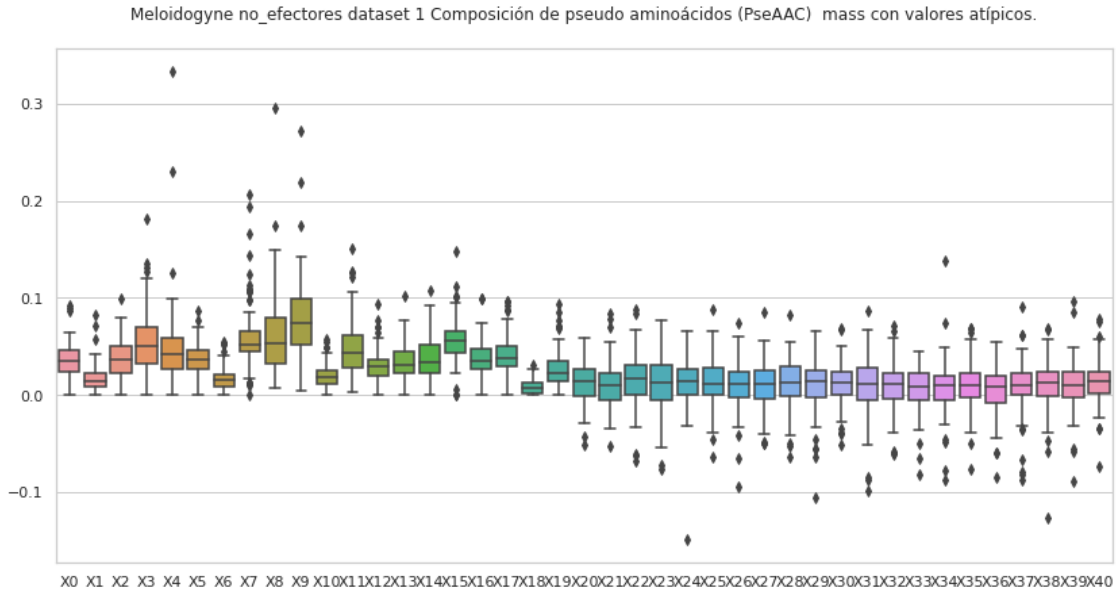
```
1   0.005057  0.005057  0.008428  …  0.037724  0.051397  0.035094  0.046170
2   0.033417  0.037872  0.028961  …  0.019347  0.023954  0.032441  0.040017
3   0.043425  0.190402  0.020042  … -0.018888 -0.005890  0.008221  0.001944
4   0.034106  0.057718  0.023612  …  0.004755  0.027853  0.040369  0.033898
..       …         …        …  …        …         …         …         …
95  0.029964  0.047443  0.024970  …  0.008507  0.034435  0.022843  0.018698
96  0.015127  0.112371  0.038898  …  0.013794  0.010895  0.008410  0.017363
97  0.053855  0.078987  0.075397  …  0.004978  0.032413  0.079784 -0.000944
98  0.022554  0.105252  0.082698  …  0.064501 -0.012669 -0.008301  0.021432
99  0.048750  0.042656  0.042656  …  0.028027  0.055538  0.023332  0.003887

          X36       X37       X38       X39       X40        X41
0   0.029655  0.036531  0.045785  0.032174  0.010875  efectores
1   0.037128  0.036655  0.040687  0.024174  0.038798  efectores
2   0.021038  0.021369  0.026250  0.038284  0.023144  efectores
3   0.014151 -0.006848  0.026541 -0.015133 -0.011069  efectores
4   0.027578 -0.000016  0.010250  0.023547  0.012407  efectores
..       …         …         …         …         …         …
95  0.042529  0.032691  0.019887  0.022680  0.030140  efectores
96  0.023562  0.012281  0.028669  0.011013  0.006403  efectores
97  0.021068 -0.014733  0.004100  0.024870  0.032051  efectores
98  0.053555  0.001265  0.074351  0.024768 -0.043462  efectores
99  0.022592  0.026047  0.016450  0.025503  0.039388  efectores

[100 rows x 42 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  100.000000  100.000000  100.000000  100.000000  100.000000  100.000000
mean     0.031555    0.015067    0.030606    0.045072    0.031409    0.041753
std      0.024048    0.016509    0.025130    0.052296    0.028689    0.024460
min      0.000000    0.000000    0.000000    0.000000    0.001479    0.006094
25%      0.017935    0.003238    0.012059    0.011530    0.012050    0.029452
50%      0.027386    0.011022    0.025213    0.024859    0.021133    0.034592
75%      0.040714    0.022588    0.042885    0.065671    0.040463    0.049833
max      0.190589    0.081043    0.118478    0.312351    0.142942    0.190589

                X6          X7          X8          X9  …         X31  \
count  100.000000  100.000000  100.000000  100.000000  …  100.000000
mean     0.009989    0.043057    0.057227    0.054430  …    0.017774
std      0.010744    0.043124    0.043666    0.046608  …    0.028584
min      0.000000    0.002973    0.001692    0.001692  …   -0.059173
25%      0.002495    0.023280    0.028824    0.024630  …    0.004371
50%      0.006693    0.034297    0.048172    0.036550  …    0.024305
```

```
75%      0.015612    0.053866    0.078996    0.075571  …    0.033487
max      0.063459    0.381179    0.238237    0.285884  …    0.129813

              X32         X33         X34         X35         X36         X37  \
count  100.000000  100.000000  100.000000  100.000000  100.000000  100.000000
mean     0.018779    0.018844    0.025265    0.017648    0.021945    0.019712
std      0.036920    0.021673    0.029617    0.030834    0.029798    0.023517
min     -0.227387   -0.036745   -0.038888   -0.168867   -0.036384   -0.069442
25%      0.006812    0.005404    0.007166    0.006287    0.007535    0.003048
50%      0.020323    0.023697    0.026654    0.020814    0.022450    0.021919
75%      0.037180    0.034441    0.039665    0.037744    0.034632    0.035021
max      0.114928    0.056602    0.198028    0.070558    0.248838    0.086040

              X38         X39         X40
count  100.000000  100.000000  100.000000
mean     0.011288    0.021758    0.009772
std      0.050643    0.034324    0.088392
min     -0.416441   -0.042773   -0.823872
25%      0.005212    0.007634    0.003552
50%      0.016750    0.023982    0.025135
75%      0.029309    0.035184    0.033205
max      0.080786    0.265426    0.133728

[8 rows x 41 columns]



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

          X0        X1        X2        X3        X4        X5        X6  \
0   0.000000  0.024252  0.000000  0.048503  0.024252  0.024252  0.024252
1   0.000000  0.000000  0.063487  0.084649  0.021162  0.063487  0.021162
2   0.023596  0.070788  0.070788  0.058990  0.070788  0.035394  0.011798
3   0.039380  0.007160  0.064441  0.035800  0.003580  0.025060  0.003580
4   0.047741  0.000000  0.017903  0.035806  0.053709  0.035806  0.000000
..       …         …         …         …         …         …         …
95  0.022631  0.008229  0.045261  0.030860  0.034975  0.051433  0.010287
96  0.000000  0.000000  0.000000  0.000000  0.332833  0.000000  0.018491
97  0.000000  0.025473  0.025473  0.050946  0.016982  0.008491  0.016982
98  0.043065  0.002871  0.043065  0.066032  0.022968  0.031581  0.025839
99  0.042693  0.010413  0.015619  0.021867  0.058312  0.041651  0.028115

          X7        X8        X9  …        X32       X33       X34       X35  \
0   0.194012  0.145509  0.218264  … -0.008517  0.014627  0.019280  0.034938
1   0.063487  0.148136  0.126974  …  0.005314 -0.031304  0.137634  0.017065
```

```
2   0.106183  0.082587  0.106183  … -0.035834 -0.049550  0.002747  0.067958
3   0.021480  0.017900  0.060861  …  0.044261  0.018818  0.012838  0.042601
4   0.047741  0.029838  0.071612  … -0.012857 -0.009685  0.005832  0.028608
..       …         …         …  …         …         …         …         …
95  0.043204  0.032917  0.030860  …  0.013278  0.023638  0.034497  0.011212
96  0.166417  0.018491  0.129435  …  0.039421 -0.081481 -0.048421 -0.001104
97  0.076419  0.067928  0.093400  … -0.000602  0.023975  0.016004  0.015593
98  0.054548  0.091871  0.048807  … -0.016955 -0.025333  0.034914 -0.022543
99  0.064560  0.029156  0.066642  …  0.021475  0.023712  0.009791  0.021505

         X36       X37       X38       X39       X40           X41
0   0.024876  0.014962  0.033248 -0.000874  0.020848  no_efectores
1  -0.059004 -0.081990 -0.009944  0.016208  0.024399  no_efectores
2  -0.005399 -0.086821 -0.058067 -0.088400  0.025270  no_efectores
3   0.013886  0.038237 -0.009847  0.012418 -0.004505  no_efectores
4   0.030168  0.012169  0.050788  0.095762  0.008896  no_efectores
..       …         …         …         …         …           …
95  0.008548  0.032017  0.024841  0.035348  0.016166  no_efectores
96 -0.059256 -0.013733  0.067807  0.049190 -0.073650  no_efectores
97  0.016174  0.028666  0.000465  0.015206 -0.007913  no_efectores
98 -0.015805  0.031309  0.012269 -0.002335  0.007481  no_efectores
99  0.028894  0.006089  0.018194  0.024345  0.019039  no_efectores

[100 rows x 42 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  100.000000  100.000000  100.000000  100.000000  100.000000  100.000000
mean     0.034453    0.016691    0.037656    0.055690    0.047534    0.037334
std      0.018469    0.013932    0.020076    0.032060    0.040978    0.015566
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.024230    0.007966    0.022693    0.032749    0.026024    0.027037
50%      0.035359    0.014867    0.037142    0.050867    0.041592    0.036947
75%      0.046473    0.022882    0.050023    0.069326    0.058315    0.045854
max      0.091975    0.082805    0.099499    0.180645    0.332833    0.086993

                X6          X7          X8          X9  …         X31  \
count  100.000000  100.000000  100.000000  100.000000  …  100.000000
mean     0.016639    0.059473    0.062795    0.078889  …    0.009127
std      0.011207    0.033471    0.043066    0.039133  …    0.030501
min      0.000000    0.000000    0.007573    0.004496  …   -0.098824
25%      0.008968    0.044879    0.032829    0.051114  …   -0.005813
50%      0.015395    0.052116    0.053084    0.073391  …    0.011083
75%      0.021484    0.065611    0.080099    0.099086  …    0.028112
```

```
max      0.054191    0.206078    0.295775    0.272244  …    0.085953

              X32         X33         X34         X35         X36         X37  \
count  100.000000  100.000000  100.000000  100.000000  100.000000  100.000000
mean     0.008084    0.005099    0.007106    0.010011    0.004917    0.008055
std      0.024558    0.023311    0.027272    0.023545    0.024538    0.026587
min     -0.060512   -0.081481   -0.087090   -0.075636   -0.084404   -0.086821
25%     -0.003912   -0.005149   -0.004912   -0.002553   -0.007943    0.000779
50%      0.011437    0.009260    0.009815    0.010319    0.008511    0.009653
75%      0.023126    0.022222    0.019856    0.022393    0.020217    0.022468
max      0.071754    0.044238    0.137634    0.068068    0.054588    0.090681

              X38         X39         X40
count  100.000000  100.000000  100.000000
mean     0.010559    0.009524    0.013606
std      0.025964    0.024984    0.022734
min     -0.126865   -0.088400   -0.073650
25%     -0.001283   -0.001846    0.001328
50%      0.012152    0.010439    0.013798
75%      0.023921    0.024157    0.024402
max      0.067807    0.095762    0.078491

[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 Meloidogyne dataset
1, sin valores atípicos.
Valores del documento csv.

          X0        X1        X2        X3        X4        X5        X6  \
0   0.026234  0.028421  0.008745  0.010931  0.013117  0.037165  0.008745
1   0.001686  0.001686  0.003371  0.003371  0.010114  0.042142  0.000000
2   0.028961  0.011139  0.006683  0.028961  0.017822  0.033417  0.004456
4   0.034106  0.031482  0.015741  0.062965  0.020988  0.010494  0.007871
5   0.027198  0.012363  0.037088  0.009890  0.024725  0.024725  0.002473
..       …         …         …         …         …         …         …
95  0.022473  0.012485  0.027467  0.014982  0.019976  0.029964  0.007491
96  0.043220  0.028093  0.086439  0.144786  0.008644  0.032415  0.006483
97  0.050265  0.000000  0.032313  0.053855  0.068216  0.053855  0.017952
98  0.000000  0.015036  0.022554  0.082698  0.075180  0.097734  0.000000
99  0.042656  0.048750  0.024375  0.000000  0.012187  0.006094  0.000000

          X7        X8        X9  …       X32       X33       X34       X35  \
0   0.045910  0.019676  0.037165  …  0.015633  0.035683  0.041154  0.026938
1   0.005057  0.005057  0.008428  …  0.037724  0.051397  0.035094  0.046170
2   0.033417  0.037872  0.028961  …  0.019347  0.023954  0.032441  0.040017
4   0.034106  0.057718  0.023612  …  0.004755  0.027853  0.040369  0.033898
5   0.032143  0.044505  0.027198  …  0.031362  0.038289  0.029519  0.018008
..       …         …         …  …         …         …         …         …
95  0.029964  0.047443  0.024970  …  0.008507  0.034435  0.022843  0.018698
96  0.015127  0.112371  0.038898  …  0.013794  0.010895  0.008410  0.017363
97  0.053855  0.078987  0.075397  …  0.004978  0.032413  0.079784 -0.000944
98  0.022554  0.105252  0.082698  …  0.064501 -0.012669 -0.008301  0.021432
99  0.048750  0.042656  0.042656  …  0.028027  0.055538  0.023332  0.003887
```

```
         X36       X37       X38       X39       X40        X41
0   0.029655  0.036531  0.045785  0.032174  0.010875  efectores
1   0.037128  0.036655  0.040687  0.024174  0.038798  efectores
2   0.021038  0.021369  0.026250  0.038284  0.023144  efectores
4   0.027578 -0.000016  0.010250  0.023547  0.012407  efectores
5   0.023249  0.021528  0.010985  0.004958  0.027443  efectores
..       …         …         …         …         …         …
95  0.042529  0.032691  0.019887  0.022680  0.030140  efectores
96  0.023562  0.012281  0.028669  0.011013  0.006403  efectores
97  0.021068 -0.014733  0.004100  0.024870  0.032051  efectores
98  0.053555  0.001265  0.074351  0.024768 -0.043462  efectores
99  0.022592  0.026047  0.016450  0.025503  0.039388  efectores

[87 rows x 42 columns]
```

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

```
              X0         X1         X2         X3         X4         X5  \
count  87.000000  87.000000  87.000000  87.000000  87.000000  87.000000
mean    0.029819   0.013853   0.027224   0.036762   0.027030   0.039306
std     0.017987   0.013674   0.021658   0.037476   0.022918   0.017702
min     0.000000   0.000000   0.000000   0.000000   0.001479   0.006094
25%     0.018012   0.003799   0.011207   0.011219   0.011320   0.029927
50%     0.027405   0.011139   0.022900   0.021854   0.020582   0.034489
75%     0.040001   0.021031   0.038309   0.057411   0.032188   0.047615
max     0.097944   0.062016   0.095416   0.180483   0.113193   0.097734

              X6         X7         X8         X9  …        X31        X32  \
count  87.000000  87.000000  87.000000  87.000000  …  87.000000  87.000000
mean    0.009112   0.035539   0.049657   0.045729  …   0.017966   0.024309
std     0.008307   0.022621   0.033437   0.035107  …   0.026137   0.021445
min     0.000000   0.002973   0.001692   0.001692  … -0.059173  -0.029379
25%     0.002721   0.020360   0.024832   0.021091  …   0.008228   0.009279
50%     0.006483   0.032717   0.041224   0.030536  …   0.026700   0.021669
75%     0.015453   0.045193   0.076278   0.066142  …   0.034519   0.038179
max     0.029341   0.130607   0.168211   0.174142  …   0.068542   0.096380

             X33        X34        X35        X36        X37        X38  \
count  87.000000  87.000000  87.000000  87.000000  87.000000  87.000000
mean    0.020188   0.024193   0.022875   0.022538   0.022482   0.019145
std     0.021852   0.023869   0.021492   0.017875   0.019963   0.022258
min    -0.036745  -0.020031  -0.045788  -0.036384  -0.030777  -0.048699
25%     0.006287   0.007606   0.010704   0.013809   0.011454   0.008261
50%     0.025772   0.027336   0.023107   0.025093   0.022932   0.021435
75%     0.035093   0.038923   0.039514   0.036303   0.035677   0.030360
```

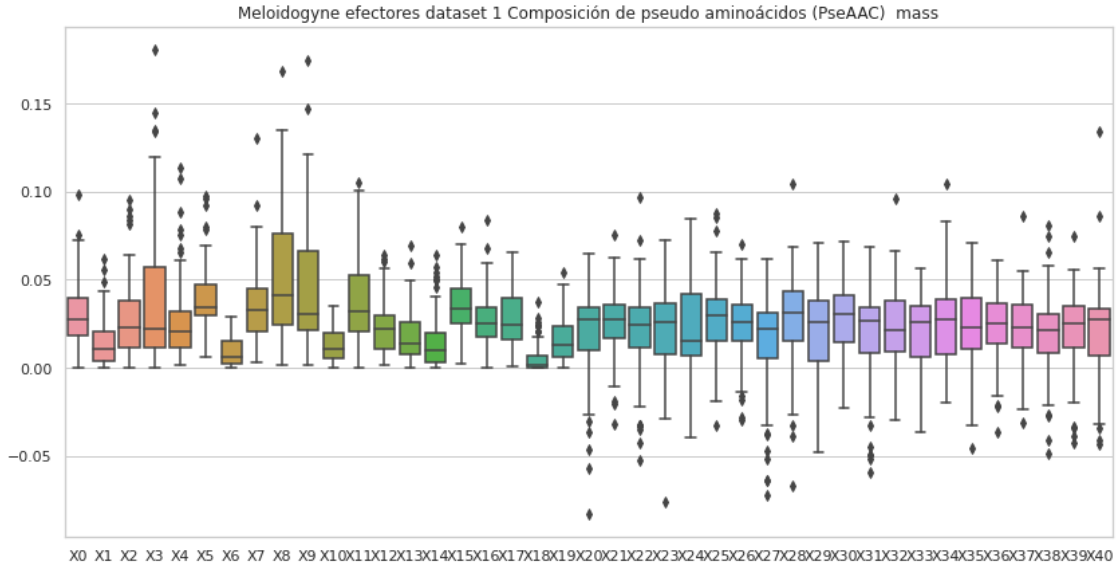
```
max     0.056602   0.104173   0.070558   0.061125   0.086040   0.080786

             X39        X40
count  87.000000  87.000000
mean    0.021387   0.020639
std     0.022265   0.025834
min    -0.042773  -0.043462
25%     0.011268   0.007132
50%     0.024852   0.027250
75%     0.035362   0.033673
max     0.074633   0.133728

[8 rows x 41 columns]
```

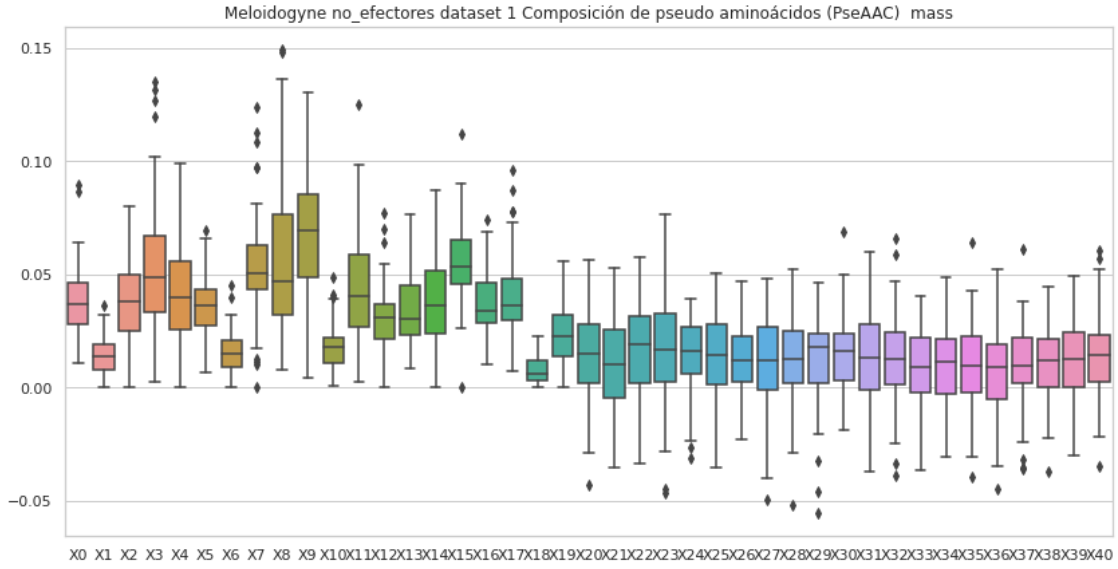
Composición de pseudo aminoácidos (PseAAC) mass no_efectores Meloidogyne
dataset 1, sin valores atípicos.
Valores del documento csv.

```
          X0        X1        X2        X3        X4        X5        X6  \
3   0.039380  0.007160  0.064441  0.035800  0.003580  0.025060  0.003580
5   0.044088  0.017635  0.013856  0.020154  0.040309  0.036530  0.008818
6   0.024692  0.016836  0.015713  0.002245  0.006734  0.025253  0.002245
7   0.037914  0.006319  0.037914  0.042127  0.054765  0.027383  0.014744
8   0.026468  0.019250  0.028874  0.045718  0.048124  0.038499  0.014437
..       …         …         …         …         …         …
93  0.041230  0.014078  0.031174  0.036202  0.058325  0.054302  0.008045
94  0.042251  0.011883  0.039610  0.034329  0.050173  0.044892  0.021125
95  0.022631  0.008229  0.045261  0.030860  0.034975  0.051433  0.010287
98  0.043065  0.002871  0.043065  0.066032  0.022968  0.031581  0.025839
99  0.042693  0.010413  0.015619  0.021867  0.058312  0.041651  0.028115

          X7        X8        X9  …       X32       X33       X34       X35  \
3   0.021480  0.017900  0.060861  …  0.044261  0.018818  0.012838  0.042601
5   0.047867  0.020154  0.073060  …  0.004229  0.040445  0.023739  0.001049
6   0.011224  0.015152  0.010663  …  0.024666  0.040551  0.021636  0.040489
7   0.065297  0.056871  0.073722  …  0.024291  0.001124  0.024958  0.003920
8   0.045718  0.031281  0.105873  … -0.001484  0.039693 -0.000963  0.001679
..       …         …         …  …         …         …         …         …
93  0.047263  0.030168  0.069386  …  0.016283  0.016740  0.018084  0.015149
94  0.047532  0.060736  0.091103  …  0.003579  0.024921 -0.004265  0.009738
95  0.043204  0.032917  0.030860  …  0.013278  0.023638  0.034497  0.011212
98  0.054548  0.091871  0.048807  … -0.016955 -0.025333  0.034914 -0.022543
99  0.064560  0.029156  0.066642  …  0.021475  0.023712  0.009791  0.021505

         X36       X37       X38       X39       X40           X41
3   0.013886  0.038237 -0.009847  0.012418 -0.004505  no_efectores
```

```
5   0.008475  0.028007  0.027825  0.023195  0.001339  no_efectores
6   0.030385  0.036763  0.034795  0.024154  0.029735  no_efectores
7   0.013262  0.001654  0.025178  0.015913  0.007937  no_efectores
8   0.009014  0.009423 -0.016440  0.013322  0.018724  no_efectores
..       …       …       …       …       …           …
93 -0.005590  0.012769  0.003745  0.010246  0.006371  no_efectores
94  0.011279  0.002562  0.011435 -0.001717 -0.009504  no_efectores
95  0.008548  0.032017  0.024841  0.035348  0.016166  no_efectores
98 -0.015805  0.031309  0.012269 -0.002335  0.007481  no_efectores
99  0.028894  0.006089  0.018194  0.024345  0.019039  no_efectores

[73 rows x 42 columns]


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

                X0          X1          X2          X3          X4          X5  \
count   73.000000   73.000000   73.000000   73.000000   73.000000   73.000000
mean     0.038008    0.014087    0.038332    0.053848    0.040079    0.036267
std      0.014572    0.008786    0.017763    0.027599    0.019834    0.012387
min      0.011020    0.000000    0.000000    0.002245    0.000000    0.006666
25%      0.028088    0.007785    0.025064    0.033001    0.025850    0.027242
50%      0.036615    0.013624    0.037914    0.048741    0.039991    0.036530
75%      0.046424    0.019012    0.049629    0.066715    0.055721    0.043449
max      0.089465    0.036007    0.080020    0.135075    0.099176    0.069494

                X6          X7          X8          X9  …         X31         X32  \
count   73.000000   73.000000   73.000000   73.000000  …   73.000000   73.000000
mean     0.016103    0.052592    0.055343    0.069112  …    0.011822    0.013090
std      0.009218    0.022382    0.032507    0.025099  …    0.020836    0.019266
min      0.000000    0.000000    0.007573    0.004496  …   -0.037004   -0.039127
25%      0.009294    0.043204    0.032270    0.048883  …   -0.001205    0.001448
50%      0.015213    0.050274    0.047001    0.069386  …    0.013335    0.012287
75%      0.021125    0.063024    0.076250    0.085556  …    0.027944    0.024291
max      0.044960    0.123546    0.149200    0.130099  …    0.059816    0.065539

               X33         X34         X35         X36         X37         X38  \
count   73.000000   73.000000   73.000000   73.000000   73.000000   73.000000
mean     0.007555    0.009764    0.009468    0.008162    0.009703    0.010940
std      0.018687    0.018385    0.019298    0.018330    0.017314    0.016963
min     -0.036310   -0.030840   -0.039358   -0.044780   -0.035696   -0.037298
25%     -0.002500   -0.002634   -0.002181   -0.005146    0.001654    0.000060
50%      0.009219    0.011212    0.009738    0.009014    0.009423    0.012035
75%      0.021994    0.021195    0.022351    0.018969    0.022069    0.021426
max      0.040551    0.048768    0.064135    0.052386    0.060966    0.044348
```

```
              X39         X40
count   73.000000   73.000000
mean     0.011883    0.014484
std      0.016399    0.018748
min     -0.030306   -0.034740
25%      0.000156    0.002626
50%      0.012418    0.014593
75%      0.024345    0.022983
max      0.049231    0.060632

[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 Meloidogyne dataset
1, con valores atípicos.
Valores del documento csv.

          X0        X1        X2        X3        X4        X5        X6  \
0   0.034128  0.036972  0.011376  0.014220  0.017064  0.048348  0.011376
1   0.009757  0.009757  0.019514  0.019514  0.058543  0.243928  0.000000
2   0.063137  0.024283  0.014570  0.063137  0.038853  0.072850  0.009713
3   0.005844  0.000308  0.005229  0.017225  0.000308  0.001846  0.001538
4   0.029282  0.027030  0.013515  0.054060  0.018020  0.009010  0.006757
..       …         …         …         …         …         …         …
95  0.042865  0.023814  0.052391  0.028577  0.038102  0.057154  0.014288
96  0.004867  0.003164  0.009735  0.016305  0.000973  0.003650  0.000730
97  0.048801  0.000000  0.031372  0.052287  0.066230  0.052287  0.017429
98  0.000000  0.002876  0.004315  0.015820  0.014382  0.018697  0.000000
99  0.048791  0.055761  0.027880  0.000000  0.013940  0.006970  0.000000

          X7        X8        X9  …       X53       X54       X55       X56  \
0   0.059724  0.025596  0.048348  … -0.016044 -0.017970 -0.013730  0.009448
1   0.029271  0.029271  0.048786  … -0.042781 -0.046443  0.004865 -0.030958
2   0.072850  0.082563  0.063137  …  0.013279 -0.002846  0.002789  0.034997
3   0.003999  0.017533  0.001846  …  0.032102  0.011078  0.032594  0.020751
4   0.029282  0.049555  0.020272  … -0.001753  0.002109  0.008915 -0.007189
..       …         …         …  …         …         …         …         …
95  0.057154  0.090493  0.047628  …  0.011462 -0.012116  0.013920  0.039491
96  0.001704  0.012655  0.004381  …  0.034925  0.007429  0.034583  0.009377
97  0.052287  0.076687  0.073202  … -0.023311  0.016121 -0.022957  0.011293
98  0.004315  0.020135  0.015820  …  0.009668  0.001671  0.007717  0.006253
99  0.055761  0.048791  0.048791  … -0.007733  0.005612  0.000205  0.018137

         X57       X58       X59       X60       X61        X62
0   0.002943  0.043247  0.030776  0.026452  0.021941  efectores
1   0.010114 -0.040999 -0.031946 -0.041748  0.018131  efectores
2   0.042919  0.028950 -0.015335 -0.066631 -0.028355  efectores
3   0.035892  0.009204  0.029294  0.008450  0.028661  efectores
4  -0.008143 -0.018403 -0.016418 -0.001328  0.003922  efectores
..       …         …         …         …         …         …
95  0.060320 -0.003217  0.019495  0.036315  0.053937  efectores
96  0.033368  0.006526  0.033924  0.008403  0.033983  efectores
97  0.016334 -0.007435  0.025496 -0.040905 -0.043560  efectores
```

```
98  0.009169  0.010014  0.009619  0.003341  0.005354  efectores
99 -0.008718  0.025897  0.041776 -0.002413 -0.026038  efectores

[100 rows x 63 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	
mean	0.032215	0.018662	0.028036	0.030296	0.031139	0.055443	
std	0.032432	0.020514	0.022477	0.029173	0.030050	0.055895	
min	-0.180299	-0.045075	-0.045075	-0.135224	-0.135224	-0.180299	
25%	0.012416	0.001831	0.012267	0.015345	0.017199	0.017922	
50%	0.030105	0.015176	0.024079	0.027946	0.029545	0.049950	
75%	0.049674	0.028122	0.043370	0.041656	0.041696	0.073536	
max	0.101589	0.082428	0.099445	0.141710	0.149583	0.244697	

	X6	X7	X8	X9	…	X52	\
count	100.000000	100.000000	100.000000	100.000000	…	100.000000	
mean	0.010000	0.038357	0.047281	0.049886	…	-0.000890	
std	0.012378	0.047211	0.039278	0.048364	…	0.077740	
min	-0.045075	-0.360598	-0.225374	-0.270448	…	-0.094341	
25%	0.002294	0.024241	0.027146	0.028351	…	-0.028634	
50%	0.007770	0.042829	0.050154	0.048567	…	-0.004073	
75%	0.014885	0.060331	0.069227	0.064879	…	0.014273	
max	0.048061	0.101589	0.149583	0.174153	…	0.715199	

	X53	X54	X55	X56	X57	X58	\
count	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	
mean	0.004279	0.004419	0.010401	0.005041	0.010896	0.005783	
std	0.066951	0.089677	0.071084	0.061679	0.064882	0.090757	
min	-0.077318	-0.085583	-0.076342	-0.117104	-0.124497	-0.094272	
25%	-0.020315	-0.018299	-0.014094	-0.013850	-0.010412	-0.018630	
50%	-0.001916	-0.002956	0.008739	0.004341	0.011888	0.000402	
75%	0.022480	0.014870	0.025233	0.022205	0.033513	0.011759	
max	0.614241	0.855159	0.661286	0.491680	0.519834	0.844470	

	X59	X60	X61
count	100.000000	100.000000	100.000000
mean	0.005805	0.001305	0.008003
std	0.089897	0.066096	0.054342
min	-0.140903	-0.092203	-0.086615
25%	-0.015364	-0.030410	-0.014781
50%	0.002289	0.003869	0.005992
75%	0.020136	0.020018	0.024560

```
max      0.844235    0.553482    0.442533

[8 rows x 62 columns]
```

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.000000	0.010794	0.000000	0.021588	0.010794	0.010794	0.010794	
1	0.000000	0.000000	0.024996	0.033328	0.008332	0.024996	0.008332	
2	0.024302	0.072905	0.072905	0.060755	0.072905	0.036453	0.012151	
3	0.034947	0.006354	0.057185	0.031770	0.003177	0.022239	0.003177	
4	0.095268	0.000000	0.035726	0.071451	0.107177	0.071451	0.000000	
..	…	…	…	…	…	…	…	
95	0.030638	0.011141	0.061277	0.041779	0.047350	0.069632	0.013926	
96	0.000000	0.000000	0.000000	0.000000	0.038222	0.000000	0.002123	
97	0.000000	0.021155	0.021155	0.042309	0.014103	0.007052	0.014103	
98	0.032932	0.002195	0.032932	0.050496	0.017564	0.024150	0.019759	
99	0.025672	0.006261	0.009392	0.013149	0.035064	0.025046	0.016906	

	X7	X8	X9	…	X53	X54	X55	X56	\
0	0.086354	0.064765	0.097148	…	0.024714	0.061028	0.024219	0.012399	
1	0.024996	0.058325	0.049993	…	0.031245	-0.050645	-0.044105	-0.044114	
2	0.109358	0.085056	0.109358	…	0.058748	-0.018585	-0.016823	0.032344	
3	0.019062	0.015885	0.054008	…	0.011754	-0.002075	0.011835	0.028233	
4	0.095268	0.059543	0.142903	…	0.010302	-0.042312	0.013502	-0.008839	
..	…	…	…	…	…	…	…	…	
95	0.058491	0.044565	0.041779	…	0.020033	0.015128	0.010053	0.010768	
96	0.019111	0.002123	0.014864	…	0.012160	0.017881	0.011464	0.020492	
97	0.063464	0.056413	0.077567	…	0.001427	0.002871	0.003838	0.070051	
98	0.041714	0.070255	0.037323	…	0.015279	-0.000528	0.009260	0.014100	
99	0.038821	0.017532	0.040073	…	0.008449	0.010641	0.000762	0.021485	

	X57	X58	X59	X60	X61	X62
0	-0.035525	0.033011	0.003745	0.070168	0.054942	no_efectores
1	0.022726	-0.035132	-0.054461	-0.007563	-0.009678	no_efectores
2	-0.052395	0.072705	0.023884	-0.007726	-0.029281	no_efectores
3	0.023249	-0.008582	0.017731	-0.029609	0.000835	no_efectores
4	-0.018681	0.079553	-0.007526	0.031943	0.016140	no_efectores
..	…	…	…	…	…	…
95	-0.005075	0.005229	-0.030069	0.011847	0.012873	no_efectores
96	0.022573	0.030175	0.027430	0.022910	0.014979	no_efectores
97	0.052328	0.079055	0.049010	0.037262	-0.021911	no_efectores
98	0.039164	-0.005138	0.011435	0.002362	0.032220	no_efectores

```
99  0.006668  0.012809  0.006422  0.014175  0.003749  no_efectores

[100 rows x 63 columns]
```

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

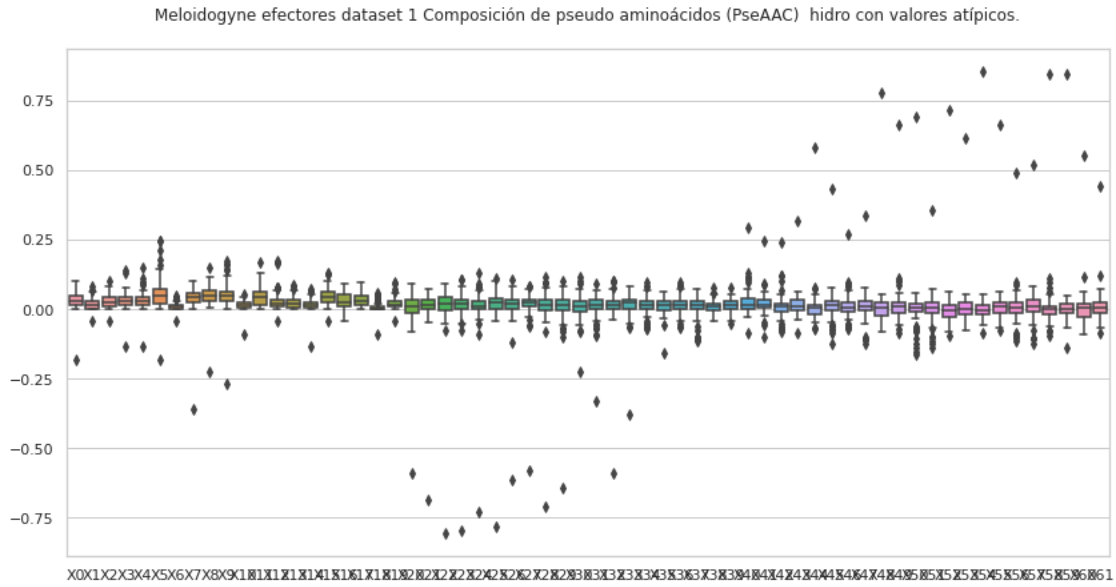
	X0	X1	X2	X3	X4	X5	\
count	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	
mean	0.032539	0.016928	0.033762	0.044027	0.040819	0.035117	
std	0.023161	0.020169	0.022032	0.020164	0.045669	0.022389	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.017505	0.005773	0.018383	0.031354	0.020984	0.018648	
50%	0.030689	0.010916	0.031330	0.042202	0.031894	0.030857	
75%	0.039751	0.021224	0.051712	0.056480	0.052169	0.048022	
max	0.109048	0.121153	0.116484	0.105408	0.427107	0.119388	

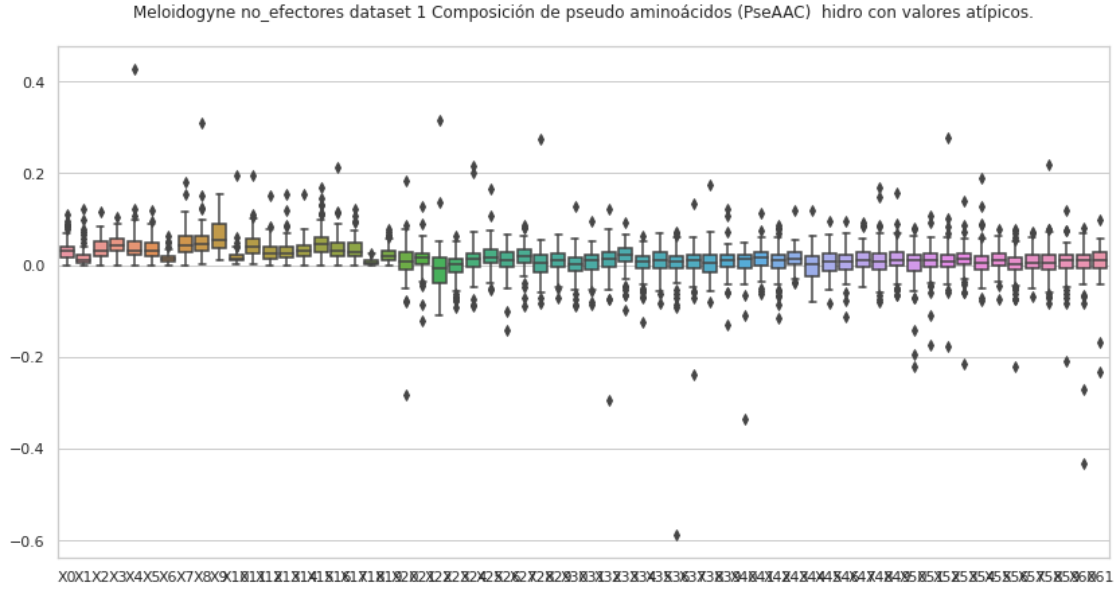
	X6	X7	X8	X9	…	X52	\
count	100.000000	100.000000	100.000000	100.000000	…	100.000000	
mean	0.014488	0.050152	0.051859	0.065446	…	0.009437	
std	0.011342	0.031400	0.037722	0.034717	…	0.042166	
min	0.000000	0.000000	0.002123	0.010762	…	-0.176122	
25%	0.006972	0.027706	0.032290	0.037172	…	-0.005273	
50%	0.012244	0.041841	0.046979	0.054182	…	0.008676	
75%	0.019544	0.064141	0.062487	0.090031	…	0.022098	
max	0.062583	0.181729	0.310623	0.155312	…	0.278095	

	X53	X54	X55	X56	X57	X58	\
count	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	
mean	0.011564	0.007057	0.011977	-0.000762	0.006822	0.006412	
std	0.034558	0.035080	0.023531	0.036787	0.027194	0.037611	
min	-0.214181	-0.078152	-0.074770	-0.220279	-0.068480	-0.083887	
25%	0.001006	-0.009228	-0.001518	-0.009485	-0.007122	-0.010799	
50%	0.013068	0.003804	0.010290	0.002485	0.006055	0.006147	
75%	0.024263	0.018541	0.026444	0.015264	0.023181	0.023330	
max	0.138644	0.191068	0.077344	0.077588	0.071217	0.219494	

	X59	X60	X61
count	100.000000	100.000000	100.000000
mean	0.005321	0.002164	0.007895
std	0.036353	0.059169	0.038345
min	-0.209962	-0.433490	-0.233793
25%	-0.006929	-0.008074	-0.005565
50%	0.009520	0.011777	0.009608
75%	0.022418	0.023761	0.028839
max	0.120496	0.081950	0.097564

[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 Meloidogyne dataset 1, sin
valores atípicos.
Valores del documento csv.
```

	X0	X1	X2	X3	X4	X5	X6	\
0	0.034128	0.036972	0.011376	0.014220	0.017064	0.048348	0.011376	
2	0.063137	0.024283	0.014570	0.063137	0.038853	0.072850	0.009713	
3	0.005844	0.000308	0.005229	0.017225	0.000308	0.001846	0.001538	
4	0.029282	0.027030	0.013515	0.054060	0.018020	0.009010	0.006757	
5	0.046135	0.020971	0.062912	0.016776	0.041941	0.041941	0.004194	
..	…	…	…	…	…	…	…	
95	0.042865	0.023814	0.052391	0.028577	0.038102	0.057154	0.014288	
96	0.004867	0.003164	0.009735	0.016305	0.000973	0.003650	0.000730	
97	0.048801	0.000000	0.031372	0.052287	0.066230	0.052287	0.017429	
98	0.000000	0.002876	0.004315	0.015820	0.014382	0.018697	0.000000	
99	0.048791	0.055761	0.027880	0.000000	0.013940	0.006970	0.000000	

	X7	X8	X9	…	X53	X54	X55	X56	\
0	0.059724	0.025596	0.048348	…	-0.016044	-0.017970	-0.013730	0.009448	
2	0.072850	0.082563	0.063137	…	0.013279	-0.002846	0.002789	0.034997	
3	0.003999	0.017533	0.001846	…	0.032102	0.011078	0.032594	0.020751	
4	0.029282	0.049555	0.020272	…	-0.001753	0.002109	0.008915	-0.007189	
5	0.054523	0.075494	0.046135	…	0.000517	-0.017433	0.010427	-0.051057	
..	…	…	…	…	…	…	…	…	
95	0.057154	0.090493	0.047628	…	0.011462	-0.012116	0.013920	0.039491	
96	0.001704	0.012655	0.004381	…	0.034925	0.007429	0.034583	0.009377	
97	0.052287	0.076687	0.073202	…	-0.023311	0.016121	-0.022957	0.011293	
98	0.004315	0.020135	0.015820	…	0.009668	0.001671	0.007717	0.006253	
99	0.055761	0.048791	0.048791	…	-0.007733	0.005612	0.000205	0.018137	

	X57	X58	X59	X60	X61	X62
0	0.002943	0.043247	0.030776	0.026452	0.021941	efectores
2	0.042919	0.028950	-0.015335	-0.066631	-0.028355	efectores
3	0.035892	0.009204	0.029294	0.008450	0.028661	efectores
4	-0.008143	-0.018403	-0.016418	-0.001328	0.003922	efectores
5	-0.020423	-0.056095	-0.025394	-0.004867	0.005011	efectores
..	…	…	…	…	…	…
95	0.060320	-0.003217	0.019495	0.036315	0.053937	efectores
96	0.033368	0.006526	0.033924	0.008403	0.033983	efectores
97	0.016334	-0.007435	0.025496	-0.040905	-0.043560	efectores
98	0.009169	0.010014	0.009619	0.003341	0.005354	efectores
99	-0.008718	0.025897	0.041776	-0.002413	-0.026038	efectores

```
[82 rows x 63 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	82.000000	82.000000	82.000000	82.000000	82.000000	82.000000	
mean	0.031563	0.016199	0.026670	0.029872	0.029013	0.047343	
std	0.021666	0.017049	0.019321	0.018328	0.020811	0.038565	
min	0.000000	0.000000	0.000000	0.000000	0.000308	0.001846	
25%	0.011949	0.000315	0.011803	0.015581	0.016337	0.016196	
50%	0.028615	0.014499	0.021784	0.027917	0.028009	0.042543	
75%	0.047056	0.023904	0.042988	0.039622	0.037998	0.064976	
max	0.083817	0.068314	0.080793	0.075241	0.098048	0.210455	

	X6	X7	X8	X9	…	X52	X53	\
count	82.000000	82.000000	82.000000	82.000000	…	82.000000	82.000000	
mean	0.009176	0.039301	0.048564	0.045813	…	-0.003904	0.003048	
std	0.008982	0.022097	0.026685	0.028818	…	0.025334	0.023787	
min	0.000000	0.001013	0.004974	0.001846	…	-0.062589	-0.046626	
25%	0.003123	0.024880	0.027001	0.023374	…	-0.024787	-0.014982	
50%	0.007588	0.039937	0.049694	0.046707	…	-0.001239	0.003023	
75%	0.013691	0.057997	0.068324	0.059471	…	0.014916	0.024100	
max	0.040128	0.083817	0.117657	0.143686	…	0.063435	0.054273	

	X54	X55	X56	X57	X58	X59	\
count	82.000000	82.000000	82.000000	82.000000	82.000000	82.000000	
mean	-0.001148	0.008466	0.005908	0.011306	-0.001113	0.003800	
std	0.024815	0.023290	0.030481	0.034520	0.026137	0.024695	
min	-0.085583	-0.061558	-0.093812	-0.105953	-0.076047	-0.068114	
25%	-0.012038	-0.005868	-0.009382	-0.006023	-0.014301	-0.009596	
50%	0.001890	0.009910	0.006276	0.014669	0.002766	0.007335	
75%	0.015351	0.027995	0.022962	0.034124	0.011523	0.022231	
max	0.052728	0.060562	0.086135	0.084067	0.078807	0.046444	

	X60	X61
count	82.000000	82.000000
mean	-0.003522	0.004096
std	0.032666	0.030344
min	-0.092203	-0.086615
25%	-0.027513	-0.015052
50%	0.006158	0.006056
75%	0.019199	0.025103
max	0.063835	0.072234

[8 rows x 62 columns]

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

```
          X0        X1        X2        X3        X4        X5        X6  \
0   0.000000  0.010794  0.000000  0.021588  0.010794  0.010794  0.010794
3   0.034947  0.006354  0.057185  0.031770  0.003177  0.022239  0.003177
5   0.025826  0.010331  0.008117  0.011806  0.023613  0.021399  0.005165
7   0.030129  0.005022  0.030129  0.033477  0.043520  0.021760  0.011717
8   0.022617  0.016449  0.024673  0.039066  0.041122  0.032898  0.012337
..       …         …         …         …         …         …         …
95  0.030638  0.011141  0.061277  0.041779  0.047350  0.069632  0.013926
96  0.000000  0.000000  0.000000  0.000000  0.038222  0.000000  0.002123
97  0.000000  0.021155  0.021155  0.042309  0.014103  0.007052  0.014103
98  0.032932  0.002195  0.032932  0.050496  0.017564  0.024150  0.019759
99  0.025672  0.006261  0.009392  0.013149  0.035064  0.025046  0.016906

          X7        X8        X9  …       X53       X54       X55       X56  \
0   0.086354  0.064765  0.097148  …  0.024714  0.061028  0.024219  0.012399
3   0.019062  0.015885  0.054008  …  0.011754 -0.002075  0.011835  0.028233
5   0.028040  0.011806  0.042798  …  0.007111  0.018852 -0.001492  0.018038
7   0.051889  0.045194  0.058585  …  0.002436 -0.003341  0.007254  0.000342
8   0.039066  0.026730  0.090469  … -0.008133 -0.010993 -0.010220 -0.035804
..       …         …         …  …       …         …         …         …
95  0.058491  0.044565  0.041779  …  0.020033  0.015128  0.010053  0.010768
96  0.019111  0.002123  0.014864  …  0.012160  0.017881  0.011464  0.020492
97  0.063464  0.056413  0.077567  …  0.001427  0.002871  0.003838  0.070051
98  0.041714  0.070255  0.037323  …  0.015279 -0.000528  0.009260  0.014100
99  0.038821  0.017532  0.040073  …  0.008449  0.010641  0.000762  0.021485

         X57       X58       X59       X60       X61           X62
0  -0.035525  0.033011  0.003745  0.070168  0.054942  no_efectores
3   0.023249 -0.008582  0.017731 -0.029609  0.000835  no_efectores
5   0.002968  0.003820 -0.005184  0.015910  0.000296  no_efectores
7  -0.005084  0.028715  0.023883  0.011707  0.011127  no_efectores
8  -0.023350  0.044828  0.004970  0.051726  0.035325  no_efectores
..       …         …         …         …         …           …
95 -0.005075  0.005229 -0.030069  0.011847  0.012873  no_efectores
96  0.022573  0.030175  0.027430  0.022910  0.014979  no_efectores
97  0.052328  0.079055  0.049010  0.037262 -0.021911  no_efectores
98  0.039164 -0.005138  0.011435  0.002362  0.032220  no_efectores
99  0.006668  0.012809  0.006422  0.014175  0.003749  no_efectores

[80 rows x 63 columns]
```

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

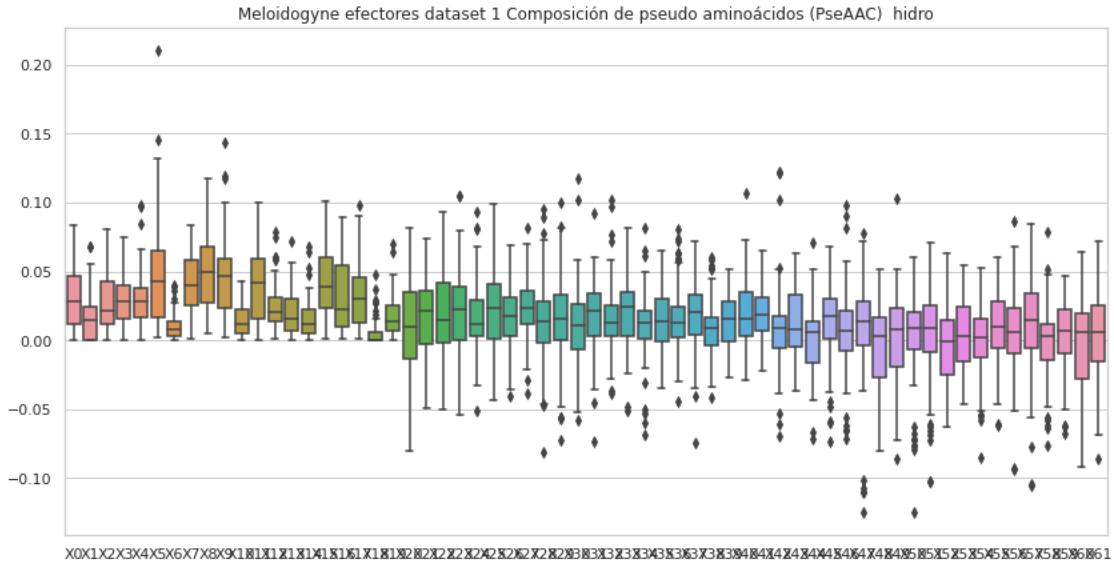
	X0	X1	X2	X3	X4	X5	\
count	80.000000	80.000000	80.000000	80.000000	80.000000	80.000000	
mean	0.028389	0.012425	0.030511	0.041728	0.033497	0.030478	
std	0.016912	0.009057	0.018944	0.017675	0.019248	0.019140	
min	0.000000	0.000000	0.000000	0.000000	0.001756	0.000000	
25%	0.017307	0.005538	0.015786	0.029915	0.019945	0.015755	
50%	0.029124	0.010746	0.029849	0.041877	0.030449	0.026510	
75%	0.035351	0.017869	0.040179	0.054782	0.044384	0.043717	
max	0.078160	0.039616	0.078579	0.075830	0.098638	0.094552	

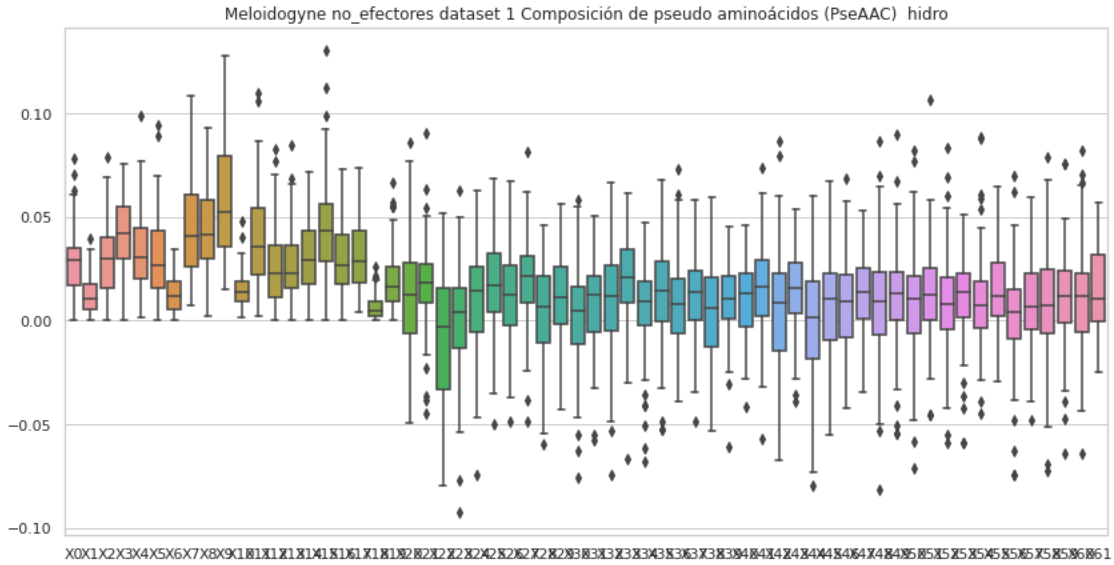
	X6	X7	X8	X9	…	X52	X53	\
count	80.000000	80.000000	80.000000	80.000000	…	80.000000	80.000000	
mean	0.012877	0.044986	0.043288	0.058038	…	0.007974	0.011619	
std	0.008294	0.023606	0.022422	0.028104	…	0.025394	0.021006	
min	0.000000	0.007245	0.002123	0.014864	…	-0.058927	-0.059128	
25%	0.005351	0.026192	0.030204	0.035929	…	-0.004401	0.001725	
50%	0.012139	0.040697	0.041469	0.052158	…	0.008187	0.013825	
75%	0.018727	0.060617	0.058235	0.079520	…	0.021177	0.022600	
max	0.034393	0.108485	0.092866	0.127775	…	0.083594	0.050934	

	X54	X55	X56	X57	X58	X59	\
count	80.000000	80.000000	80.000000	80.000000	80.000000	80.000000	
mean	0.009347	0.014602	0.000966	0.007295	0.008083	0.011247	
std	0.023625	0.019103	0.025191	0.023098	0.025951	0.023440	
min	-0.044963	-0.029095	-0.074496	-0.048028	-0.072140	-0.063888	
25%	-0.003762	0.002187	-0.008976	-0.004301	-0.006221	-0.000753	
50%	0.007446	0.011650	0.003920	0.006682	0.007296	0.011799	
75%	0.018646	0.028067	0.015073	0.022719	0.024584	0.024109	
max	0.088447	0.064353	0.070051	0.059429	0.079055	0.075511	

	X60	X61
count	80.000000	80.000000
mean	0.011152	0.013625
std	0.025248	0.021102
min	-0.064392	-0.024593
25%	-0.005654	-0.000440
50%	0.011889	0.010632
75%	0.023119	0.031613
max	0.081950	0.057183

[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 Meloidogyne dataset
1, con valores atípicos.
Valores del documento csv.

          X0        X1        X2        X3        X4        X5        X6  \
0  -0.009618  0.018338  0.018008  0.101040 -0.081912 -0.050435  0.010749
1   0.265555  0.337582 -0.136483 -0.138220  0.028433 -0.005262  0.277734
2   0.019018 -0.046037 -0.076364  0.038581  0.113805  0.027334  0.024965
3   0.010457  0.041014  0.026605 -0.044688  0.001339 -0.006004  0.000650
4   0.020956  0.081031  0.016139  0.033608  0.010271  0.082927  0.085404
..       …         …         …         …         …         …         …
95  0.010029 -0.024230  0.041947 -0.023538 -0.058212  0.007205 -0.010750
96  0.009290  0.048523  0.090097 -0.020052  0.009021 -0.000889 -0.000265
97 -0.000649  0.025176 -0.114403 -0.061001 -0.130957  0.056214  0.014240
98 -0.064560 -0.058733 -0.081470 -0.064559  0.057397  0.115181 -0.097361
99  0.008618 -0.077600  0.041785 -0.008346 -0.038076 -0.005824  0.129545

          X7        X8        X9       X10       X11       X12        X13
0   0.086237 -0.087873  0.082352  0.064974 -0.033856 -0.060928  efectores
```

```
1   0.186487  0.331994  0.148957  0.260076  0.193246  0.068512  efectores
2   0.000714 -0.015229  0.071540  0.134466  0.014104 -0.019980  efectores
3  -0.042737  0.083362  0.047332 -0.078357 -0.004925 -0.052892  efectores
4   0.103706  0.041430  0.086995  0.005580 -0.001361 -0.020023  efectores
..       …       …       …       …       …       …       …
95 -0.021706  0.095961  0.036574 -0.018507  0.080866 -0.068077  efectores
96  0.011530 -0.009703  0.030045 -0.013366 -0.052270  0.004336  efectores
97  0.115503  0.060589 -0.109096 -0.073740 -0.023275 -0.015019  efectores
98 -0.043737 -0.012124 -0.153420  0.042288 -0.041851  0.131486  efectores
99  0.004862 -0.056304  0.010018  0.054083 -0.052566  0.009773  efectores

[100 rows x 14 columns]
```

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

```
                X0          X1          X2          X3          X4          X5  \
count   100.000000  100.000000  100.000000  100.000000  100.000000  100.000000
mean     -0.003043    0.038740    0.020301    0.013636    0.015350    0.037593
std       0.137824    0.101201    0.108151    0.086700    0.102013    0.103611
min      -0.568520   -0.341363   -0.170590   -0.267092   -0.177306   -0.136557
25%      -0.038868   -0.013518   -0.024807   -0.025291   -0.049862   -0.018496
50%       0.010243    0.018375    0.014999    0.013177    0.006489    0.018984
75%       0.057455    0.056850    0.045584    0.057086    0.072971    0.071407
max       0.366084    0.421241    0.595207    0.251137    0.398630    0.424122

                X6          X7          X8          X9         X10         X11  \
count   100.000000  100.000000  100.000000  100.000000  100.000000  100.000000
mean      0.029261   -0.002673    0.055870    0.011285    0.040582    0.018197
std       0.078483    0.097604    0.107336    0.092542    0.088739    0.107474
min      -0.169134   -0.314314   -0.198825   -0.239777   -0.236089   -0.266382
25%      -0.021321   -0.033898   -0.007559   -0.046569   -0.018807   -0.030954
50%       0.012931    0.001064    0.046848    0.018323    0.037763    0.021404
75%       0.070870    0.030448    0.097163    0.067351    0.080371    0.065388
max       0.277734    0.366508    0.378650    0.323957    0.400308    0.523388

               X12
count   100.000000
mean      0.020297
std       0.089523
min      -0.210603
25%      -0.027984
50%       0.009178
75%       0.068892
max       0.342163
```

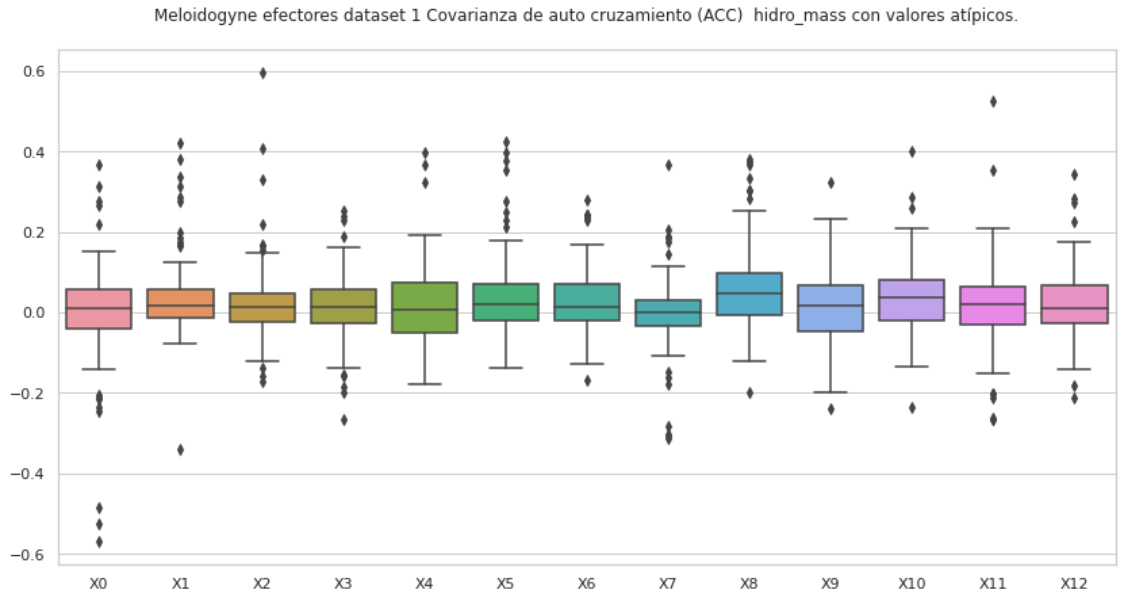
no_efectores
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores Meloidogyne
dataset 1, con valores atípicos.
Valores del documento csv.

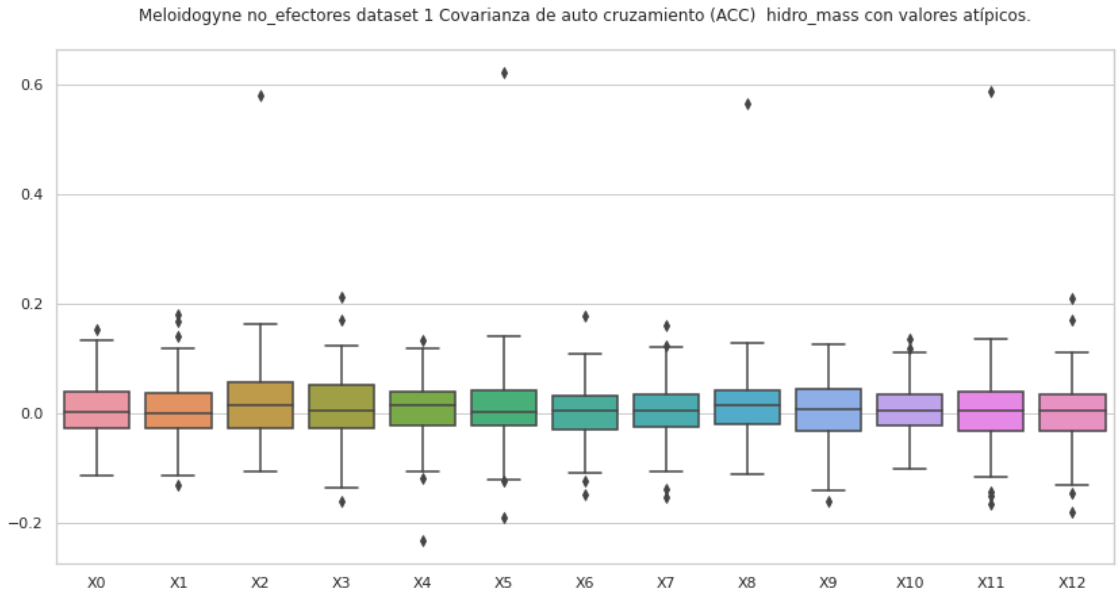
```
          X0        X1        X2        X3        X4        X5        X6  \
0   0.051259  0.000690 -0.048400 -0.015754 -0.052978  0.017801 -0.032664
1  -0.048316  0.014728 -0.062590 -0.159565  0.132780 -0.012115  0.083942
2   0.099394  0.090723 -0.016915 -0.135168  0.035190  0.101196 -0.080053
3   0.046510  0.138997 -0.083147  0.171023  0.023618  0.099237 -0.057729
4   0.133884  0.019570  0.046151  0.049495 -0.092208 -0.059080 -0.079340
..       …         …         …         …         …         …
95 -0.042570  0.048552 -0.038856 -0.015031  0.037405 -0.003761 -0.047638
96 -0.044376  0.013869 -0.005466  0.053364 -0.031360 -0.036567  0.004094
97 -0.033934 -0.017214  0.015474 -0.027739 -0.011371  0.013692 -0.028193
98  0.015381 -0.105875  0.059435  0.034510 -0.047557  0.082131  0.033396
99  0.031584  0.071183  0.029517 -0.028686 -0.005184  0.069613 -0.064414

          X7        X8        X9       X10       X11       X12           X13
0  -0.049770 -0.055617  0.036160 -0.070329 -0.082556 -0.020838  no_efectores
1   0.056576 -0.019316  0.125752 -0.086119 -0.045727  0.011244  no_efectores
2  -0.016365 -0.027642  0.071447  0.044712 -0.020866 -0.068651  no_efectores
3   0.105175 -0.006188  0.080656  0.001820  0.072216  0.111836  no_efectores
4   0.009409 -0.055605 -0.071736  0.024170  0.000796 -0.095368  no_efectores
..       …         …         …         …         …         …
95 -0.000306  0.028478  0.023346  0.023350 -0.057852 -0.008086  no_efectores
96  0.090458 -0.090443 -0.055358  0.116809 -0.035596  0.071084  no_efectores
97  0.014250 -0.004228  0.018025 -0.073557 -0.002365 -0.018636  no_efectores
98  0.019521  0.038926  0.045048 -0.016221  0.030569 -0.059675  no_efectores
99  0.011805  0.008870 -0.000037  0.035674 -0.005309  0.035080  no_efectores

[100 rows x 14 columns]
```

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

```
              X0          X1          X2          X3          X4          X5  \
count  100.000000  100.000000  100.000000  100.000000  100.000000  100.000000
mean     0.009447    0.002587    0.020732    0.006549    0.007572    0.011896
std      0.049477    0.057712    0.078990    0.063288    0.052791    0.082161
min     -0.113027   -0.132325   -0.105551   -0.159565   -0.232271   -0.190051
25%     -0.026637   -0.028358   -0.027732   -0.027976   -0.023610   -0.023442
50%      0.002353   -0.000031    0.013908    0.005123    0.015645    0.002181
75%      0.039686    0.037777    0.055332    0.050462    0.038040    0.042225
```

```
max      0.152136    0.180164    0.579399    0.210638    0.132780    0.621072

                X6          X7          X8          X9         X10         X11  \
count   100.000000  100.000000  100.000000  100.000000  100.000000  100.000000
mean      0.002898    0.005071    0.013253    0.001222    0.006275    0.008531
std       0.052459    0.053877    0.074861    0.058090    0.044709    0.083632
min      -0.148124   -0.152661   -0.110437   -0.161313   -0.101257   -0.165787
25%      -0.029071   -0.023875   -0.020668   -0.033513   -0.021744   -0.031811
50%       0.003829    0.004823    0.013644    0.006411    0.005119    0.005693
75%       0.030739    0.033754    0.041013    0.045225    0.033621    0.039449
max       0.178332    0.160266    0.564627    0.125752    0.136189    0.587675

               X12
count   100.000000
mean     -0.000787
std       0.062946
min      -0.180013
25%      -0.032691
50%       0.004320
75%       0.034008
max       0.210077
```





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 Meloidogyne dataset
1, sin valores atípicos.
Valores del documento csv.

          X0        X1        X2        X3        X4        X5        X6  \
0  -0.009618  0.018338  0.018008  0.101040 -0.081912 -0.050435  0.010749
2   0.019018 -0.046037 -0.076364  0.038581  0.113805  0.027334  0.024965
3   0.010457  0.041014  0.026605 -0.044688  0.001339 -0.006004  0.000650
4   0.020956  0.081031  0.016139  0.033608  0.010271  0.082927  0.085404
5   0.095727 -0.024230 -0.024181  0.010596  0.090147  0.038821  0.032303
..       …         …         …         …         …         …         …
95  0.010029 -0.024230  0.041947 -0.023538 -0.058212  0.007205 -0.010750
96  0.009290  0.048523  0.090097 -0.020052  0.009021 -0.000889 -0.000265
97 -0.000649  0.025176 -0.114403 -0.061001 -0.130957  0.056214  0.014240
98 -0.064560 -0.058733 -0.081470 -0.064559  0.057397  0.115181 -0.097361
99  0.008618 -0.077600  0.041785 -0.008346 -0.038076 -0.005824  0.129545

          X7        X8        X9       X10       X11       X12        X13
0   0.086237 -0.087873  0.082352  0.064974 -0.033856 -0.060928  efectores
2   0.000714 -0.015229  0.071540  0.134466  0.014104 -0.019980  efectores
3  -0.042737  0.083362  0.047332 -0.078357 -0.004925 -0.052892  efectores
4   0.103706  0.041430  0.086995  0.005580 -0.001361 -0.020023  efectores
5   0.030398  0.083257  0.036961  0.060210  0.057526  0.050491  efectores
..       …         …         …         …         …         …         …
95 -0.021706  0.095961  0.036574 -0.018507  0.080866 -0.068077  efectores
```

```
96  0.011530 -0.009703  0.030045 -0.013366 -0.052270  0.004336  efectores
97  0.115503  0.060589 -0.109096 -0.073740 -0.023275 -0.015019  efectores
98 -0.043737 -0.012124 -0.153420  0.042288 -0.041851  0.131486  efectores
99  0.004862 -0.056304  0.010018  0.054083 -0.052566  0.009773  efectores

[84 rows x 14 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	84.000000	84.000000	84.000000	84.000000	84.000000	84.000000	
mean	0.014326	0.022744	-0.002117	0.014598	-0.000289	0.018958	
std	0.076734	0.062037	0.068489	0.066372	0.080139	0.073552	
min	-0.211447	-0.077600	-0.170590	-0.155161	-0.177306	-0.136557	
25%	-0.031026	-0.015875	-0.036457	-0.023776	-0.052919	-0.025997	
50%	0.013860	0.016230	0.008591	0.009663	0.000958	0.013172	
75%	0.055894	0.048745	0.034872	0.051962	0.040105	0.055826	
max	0.275770	0.312105	0.168145	0.251137	0.193364	0.275986	

	X6	X7	X8	X9	X10	X11	\
count	84.000000	84.000000	84.000000	84.000000	84.000000	84.000000	
mean	0.015966	0.004571	0.032357	0.011648	0.030028	0.013707	
std	0.061761	0.069297	0.076737	0.074454	0.070514	0.068456	
min	-0.128069	-0.282952	-0.120373	-0.199405	-0.136223	-0.213021	
25%	-0.024080	-0.022060	-0.015609	-0.026259	-0.019110	-0.027727	
50%	0.010818	0.001064	0.032082	0.018323	0.024039	0.015937	
75%	0.036948	0.030223	0.075510	0.056282	0.066129	0.050366	
max	0.242701	0.206014	0.304270	0.182667	0.287288	0.209694	

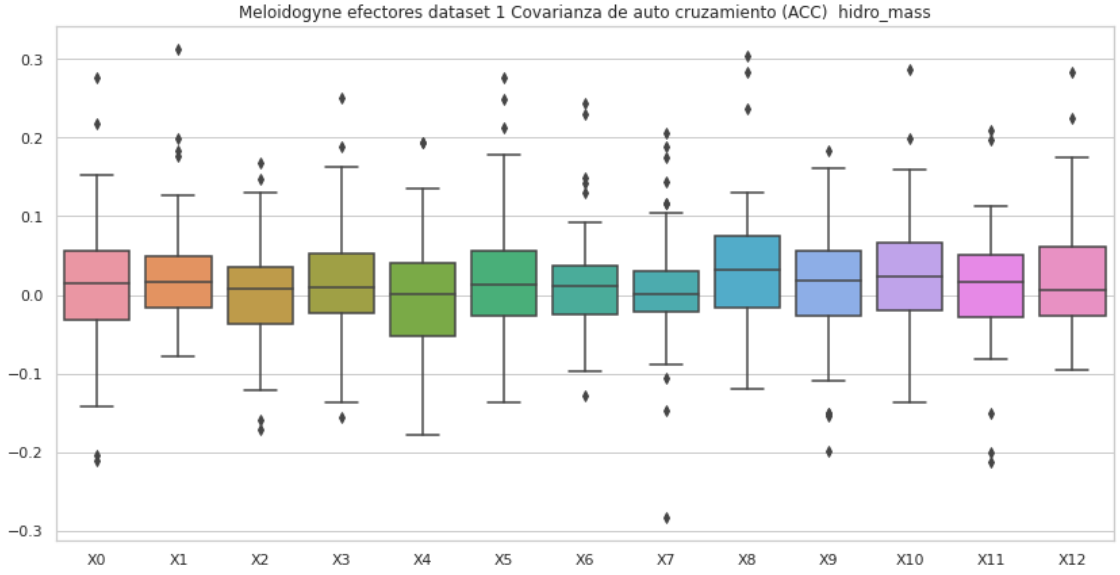
	X12
count	84.000000
mean	0.019290
std	0.069218
min	-0.095788
25%	-0.027115
50%	0.005510
75%	0.060535
max	0.283589

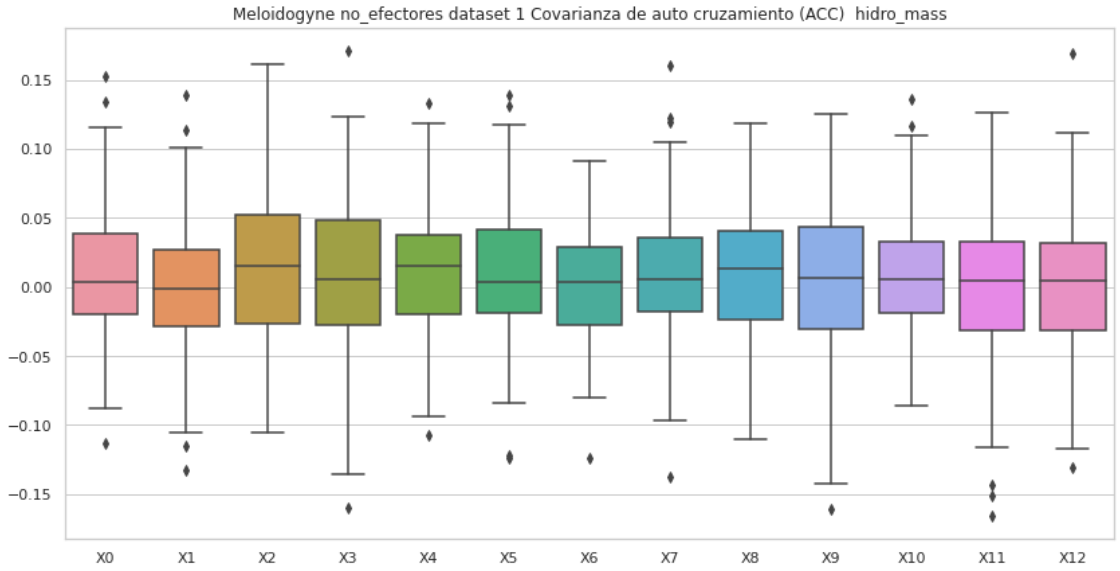
no_efectores
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores Meloidogyne dataset 1, sin valores atípicos.
Valores del documento csv.

```
          X0        X1        X2        X3        X4        X5        X6  \
0   0.051259  0.000690 -0.048400 -0.015754 -0.052978  0.017801 -0.032664
1  -0.048316  0.014728 -0.062590 -0.159565  0.132780 -0.012115  0.083942
2   0.099394  0.090723 -0.016915 -0.135168  0.035190  0.101196 -0.080053
3   0.046510  0.138997 -0.083147  0.171023  0.023618  0.099237 -0.057729
4   0.133884  0.019570  0.046151  0.049495 -0.092208 -0.059080 -0.079340
..       …         …         …         …         …         …
95 -0.042570  0.048552 -0.038856 -0.015031  0.037405 -0.003761 -0.047638
96 -0.044376  0.013869 -0.005466  0.053364 -0.031360 -0.036567  0.004094
97 -0.033934 -0.017214  0.015474 -0.027739 -0.011371  0.013692 -0.028193
98  0.015381 -0.105875  0.059435  0.034510 -0.047557  0.082131  0.033396
99  0.031584  0.071183  0.029517 -0.028686 -0.005184  0.069613 -0.064414

          X7        X8        X9       X10       X11       X12           X13
0  -0.049770 -0.055617  0.036160 -0.070329 -0.082556 -0.020838  no_efectores
1   0.056576 -0.019316  0.125752 -0.086119 -0.045727  0.011244  no_efectores
2  -0.016365 -0.027642  0.071447  0.044712 -0.020866 -0.068651  no_efectores
3   0.105175 -0.006188  0.080656  0.001820  0.072216  0.111836  no_efectores
4   0.009409 -0.055605 -0.071736  0.024170  0.000796 -0.095368  no_efectores
..       …         …         …         …         …         …
95 -0.000306  0.028478  0.023346  0.023350 -0.057852 -0.008086  no_efectores
96  0.090458 -0.090443 -0.055358  0.116809 -0.035596  0.071084  no_efectores
97  0.014250 -0.004228  0.018025 -0.073557 -0.002365 -0.018636  no_efectores
98  0.019521  0.038926  0.045048 -0.016221  0.030569 -0.059675  no_efectores
99  0.011805  0.008870 -0.000037  0.035674 -0.005309  0.035080  no_efectores

[94 rows x 14 columns]


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

              X0         X1         X2         X3         X4         X5  \
count  94.000000  94.000000  94.000000  94.000000  94.000000  94.000000
mean    0.010888  -0.002042   0.016391   0.005964   0.010021   0.009642
std     0.048177   0.050913   0.054791   0.058924   0.044461   0.050390
min    -0.113027  -0.132325  -0.105551  -0.159565  -0.107675  -0.123967
25%    -0.019384  -0.028569  -0.026880  -0.027699  -0.019810  -0.018704
50%     0.003669  -0.001343   0.015068   0.005123   0.015645   0.003483
75%     0.038871   0.026966   0.052661   0.048902   0.037694   0.041340
max     0.152136   0.138997   0.161712   0.171023   0.132780   0.139449

              X6         X7         X8         X9        X10        X11  \
count  94.000000  94.000000  94.000000  94.000000  94.000000  94.000000
mean    0.001715   0.008022   0.006574   0.001347   0.007345   0.000256
std     0.045021   0.049845   0.049615   0.055648   0.043487   0.058487
```

min	-0.123432	-0.137968	-0.110437	-0.161313	-0.086119	-0.165787
25%	-0.027784	-0.018217	-0.023373	-0.030893	-0.019192	-0.031906
50%	0.003606	0.005878	0.013039	0.006411	0.005119	0.004955
75%	0.028681	0.035961	0.040359	0.043925	0.032665	0.032706
max	0.091412	0.160266	0.118471	0.125752	0.136189	0.126662

	X12
count	94.000000
mean	0.000322
std	0.055787
min	-0.130800
25%	-0.031100
50%	0.005023
75%	0.032084
max	0.168676





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 Meloidogyne dataset 1, con
valores atípicos.
Valores del documento csv.

          X0        X1        X2        X3        X4        X5        X6  \
0  -0.009618  0.018338  0.018008  0.101040 -0.081912 -0.050435  0.010749
1   0.265555  0.337582 -0.136483 -0.138220  0.028433 -0.005262  0.277734
2   0.019018 -0.046037 -0.076364  0.038581  0.113805  0.027334  0.024965
3   0.010457  0.041014  0.026605 -0.044688  0.001339 -0.006004  0.000650
4   0.020956  0.081031  0.016139  0.033608  0.010271  0.082927  0.085404
..       …         …         …         …         …         …         …
95  0.010029 -0.024230  0.041947 -0.023538 -0.058212  0.007205 -0.010750
96  0.009290  0.048523  0.090097 -0.020052  0.009021 -0.000889 -0.000265
97 -0.000649  0.025176 -0.114403 -0.061001 -0.130957  0.056214  0.014240
98 -0.064560 -0.058733 -0.081470 -0.064559  0.057397  0.115181 -0.097361
99  0.008618 -0.077600  0.041785 -0.008346 -0.038076 -0.005824  0.129545

          X7        X8        X9       X10       X11       X12         X13
0   0.086237 -0.087873  0.082352  0.064974 -0.033856 -0.060928  efectores
1   0.186487  0.331994  0.148957  0.260076  0.193246  0.068512  efectores
2   0.000714 -0.015229  0.071540  0.134466  0.014104 -0.019980  efectores
3  -0.042737  0.083362  0.047332 -0.078357 -0.004925 -0.052892  efectores
4   0.103706  0.041430  0.086995  0.005580 -0.001361 -0.020023  efectores
..       …         …         …         …         …         …          …
95 -0.021706  0.095961  0.036574 -0.018507  0.080866 -0.068077  efectores
96  0.011530 -0.009703  0.030045 -0.013366 -0.052270  0.004336  efectores
97  0.115503  0.060589 -0.109096 -0.073740 -0.023275 -0.015019  efectores
98 -0.043737 -0.012124 -0.153420  0.042288 -0.041851  0.131486  efectores
99  0.004862 -0.056304  0.010018  0.054083 -0.052566  0.009773  efectores

[100 rows x 14 columns]


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

                X0          X1          X2          X3          X4          X5  \
count   100.000000  100.000000  100.000000  100.000000  100.000000  100.000000
mean     -0.003043    0.038740    0.020301    0.013636    0.015350    0.037593
```

```
std      0.137824    0.101201    0.108151    0.086700    0.102013    0.103611
min     -0.568520   -0.341363   -0.170590   -0.267092   -0.177306   -0.136557
25%     -0.038868   -0.013518   -0.024807   -0.025291   -0.049862   -0.018496
50%      0.010243    0.018375    0.014999    0.013177    0.006489    0.018984
75%      0.057455    0.056850    0.045584    0.057086    0.072971    0.071407
max      0.366084    0.421241    0.595207    0.251137    0.398630    0.424122

                X6          X7          X8          X9         X10         X11  \
count   100.000000  100.000000  100.000000  100.000000  100.000000  100.000000
mean      0.029261   -0.002673    0.055870    0.011285    0.040582    0.018197
std       0.078483    0.097604    0.107336    0.092542    0.088739    0.107474
min      -0.169134   -0.314314   -0.198825   -0.239777   -0.236089   -0.266382
25%      -0.021321   -0.033898   -0.007559   -0.046569   -0.018807   -0.030954
50%       0.012931    0.001064    0.046848    0.018323    0.037763    0.021404
75%       0.070870    0.030448    0.097163    0.067351    0.080371    0.065388
max       0.277734    0.366508    0.378650    0.323957    0.400308    0.523388

               X12
count   100.000000
mean      0.020297
std       0.089523
min      -0.210603
25%      -0.027984
50%       0.009178
75%       0.068892
max       0.342163



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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.051259  0.000690 -0.048400 -0.015754 -0.052978  0.017801 -0.032664
1   -0.048316  0.014728 -0.062590 -0.159565  0.132780 -0.012115  0.083942
2    0.099394  0.090723 -0.016915 -0.135168  0.035190  0.101196 -0.080053
3    0.046510  0.138997 -0.083147  0.171023  0.023618  0.099237 -0.057729
4    0.133884  0.019570  0.046151  0.049495 -0.092208 -0.059080 -0.079340
..        …         …         …         …         …         …         …
95  -0.042570  0.048552 -0.038856 -0.015031  0.037405 -0.003761 -0.047638
96  -0.044376  0.013869 -0.005466  0.053364 -0.031360 -0.036567  0.004094
97  -0.033934 -0.017214  0.015474 -0.027739 -0.011371  0.013692 -0.028193
98   0.015381 -0.105875  0.059435  0.034510 -0.047557  0.082131  0.033396
99   0.031584  0.071183  0.029517 -0.028686 -0.005184  0.069613 -0.064414

           X7        X8        X9       X10       X11       X12           X13
```

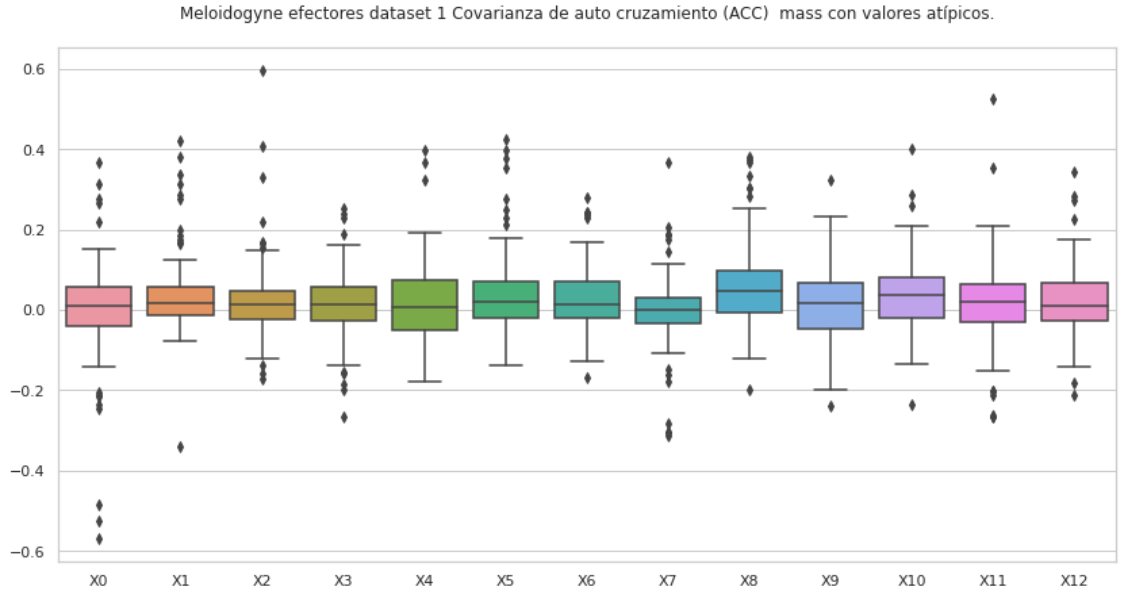
```
0  -0.049770 -0.055617  0.036160 -0.070329 -0.082556 -0.020838  no_efectores
1   0.056576 -0.019316  0.125752 -0.086119 -0.045727  0.011244  no_efectores
2  -0.016365 -0.027642  0.071447  0.044712 -0.020866 -0.068651  no_efectores
3   0.105175 -0.006188  0.080656  0.001820  0.072216  0.111836  no_efectores
4   0.009409 -0.055605 -0.071736  0.024170  0.000796 -0.095368  no_efectores
..       …       …       …       …       …       …           …
95 -0.000306  0.028478  0.023346  0.023350 -0.057852 -0.008086  no_efectores
96  0.090458 -0.090443 -0.055358  0.116809 -0.035596  0.071084  no_efectores
97  0.014250 -0.004228  0.018025 -0.073557 -0.002365 -0.018636  no_efectores
98  0.019521  0.038926  0.045048 -0.016221  0.030569 -0.059675  no_efectores
99  0.011805  0.008870 -0.000037  0.035674 -0.005309  0.035080  no_efectores

[100 rows x 14 columns]
```

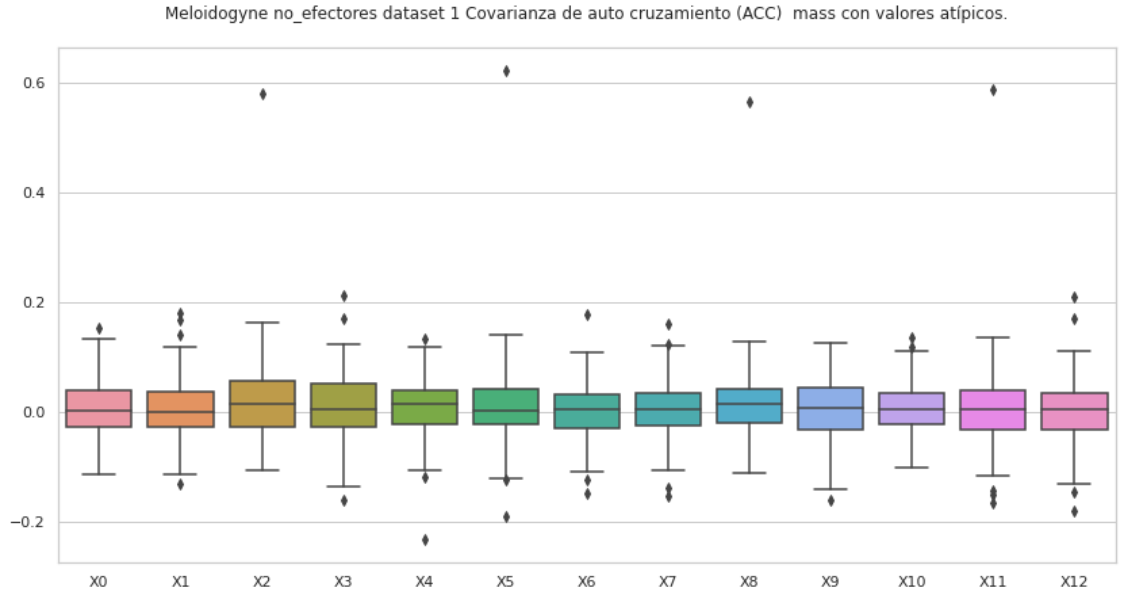
Covarianza de auto cruzamiento (ACC) mass no_efectores Meloidogyne dataset 1, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	
mean	0.009447	0.002587	0.020732	0.006549	0.007572	0.011896	
std	0.049477	0.057712	0.078990	0.063288	0.052791	0.082161	
min	-0.113027	-0.132325	-0.105551	-0.159565	-0.232271	-0.190051	
25%	-0.026637	-0.028358	-0.027732	-0.027976	-0.023610	-0.023442	
50%	0.002353	-0.000031	0.013908	0.005123	0.015645	0.002181	
75%	0.039686	0.037777	0.055332	0.050462	0.038040	0.042225	
max	0.152136	0.180164	0.579399	0.210638	0.132780	0.621072	

	X6	X7	X8	X9	X10	X11	\
count	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	
mean	0.002898	0.005071	0.013253	0.001222	0.006275	0.008531	
std	0.052459	0.053877	0.074861	0.058090	0.044709	0.083632	
min	-0.148124	-0.152661	-0.110437	-0.161313	-0.101257	-0.165787	
25%	-0.029071	-0.023875	-0.020668	-0.033513	-0.021744	-0.031811	
50%	0.003829	0.004823	0.013644	0.006411	0.005119	0.005693	
75%	0.030739	0.033754	0.041013	0.045225	0.033621	0.039449	
max	0.178332	0.160266	0.564627	0.125752	0.136189	0.587675	

	X12
count	100.000000
mean	-0.000787
std	0.062946
min	-0.180013
25%	-0.032691
50%	0.004320
75%	0.034008
max	0.210077


Meloidogyne efectores dataset 1 Covarianza de auto cruzamiento (ACC) 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


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

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 Meloidogyne dataset 1, sin
valores atípicos.
Valores del documento csv.
```

	X0	X1	X2	X3	X4	X5	X6	\
0	-0.009618	0.018338	0.018008	0.101040	-0.081912	-0.050435	0.010749	
2	0.019018	-0.046037	-0.076364	0.038581	0.113805	0.027334	0.024965	
3	0.010457	0.041014	0.026605	-0.044688	0.001339	-0.006004	0.000650	
4	0.020956	0.081031	0.016139	0.033608	0.010271	0.082927	0.085404	
5	0.095727	-0.024230	-0.024181	0.010596	0.090147	0.038821	0.032303	
..	…	…	…	…	…	…	…	
95	0.010029	-0.024230	0.041947	-0.023538	-0.058212	0.007205	-0.010750	
96	0.009290	0.048523	0.090097	-0.020052	0.009021	-0.000889	-0.000265	
97	-0.000649	0.025176	-0.114403	-0.061001	-0.130957	0.056214	0.014240	
98	-0.064560	-0.058733	-0.081470	-0.064559	0.057397	0.115181	-0.097361	
99	0.008618	-0.077600	0.041785	-0.008346	-0.038076	-0.005824	0.129545	

	X7	X8	X9	X10	X11	X12	X13
0	0.086237	-0.087873	0.082352	0.064974	-0.033856	-0.060928	efectores
2	0.000714	-0.015229	0.071540	0.134466	0.014104	-0.019980	efectores
3	-0.042737	0.083362	0.047332	-0.078357	-0.004925	-0.052892	efectores
4	0.103706	0.041430	0.086995	0.005580	-0.001361	-0.020023	efectores
5	0.030398	0.083257	0.036961	0.060210	0.057526	0.050491	efectores
..	…	…	…	…	…	…	…
95	-0.021706	0.095961	0.036574	-0.018507	0.080866	-0.068077	efectores
96	0.011530	-0.009703	0.030045	-0.013366	-0.052270	0.004336	efectores
97	0.115503	0.060589	-0.109096	-0.073740	-0.023275	-0.015019	efectores
98	-0.043737	-0.012124	-0.153420	0.042288	-0.041851	0.131486	efectores
99	0.004862	-0.056304	0.010018	0.054083	-0.052566	0.009773	efectores

```
[84 rows x 14 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	84.000000	84.000000	84.000000	84.000000	84.000000	84.000000	
mean	0.014326	0.022744	-0.002117	0.014598	-0.000289	0.018958	
std	0.076734	0.062037	0.068489	0.066372	0.080139	0.073552	
min	-0.211447	-0.077600	-0.170590	-0.155161	-0.177306	-0.136557	
25%	-0.031026	-0.015875	-0.036457	-0.023776	-0.052919	-0.025997	
50%	0.013860	0.016230	0.008591	0.009663	0.000958	0.013172	
75%	0.055894	0.048745	0.034872	0.051962	0.040105	0.055826	

```
max     0.275770   0.312105   0.168145   0.251137   0.193364   0.275986

              X6         X7         X8         X9        X10        X11  \
count  84.000000  84.000000  84.000000  84.000000  84.000000  84.000000
mean    0.015966   0.004571   0.032357   0.011648   0.030028   0.013707
std     0.061761   0.069297   0.076737   0.074454   0.070514   0.068456
min    -0.128069  -0.282952  -0.120373  -0.199405  -0.136223  -0.213021
25%    -0.024080  -0.022060  -0.015609  -0.026259  -0.019110  -0.027727
50%     0.010818   0.001064   0.032082   0.018323   0.024039   0.015937
75%     0.036948   0.030223   0.075510   0.056282   0.066129   0.050366
max     0.242701   0.206014   0.304270   0.182667   0.287288   0.209694

             X12
count  84.000000
mean    0.019290
std     0.069218
min    -0.095788
25%    -0.027115
50%     0.005510
75%     0.060535
max     0.283589
```

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

```
          X0        X1        X2        X3        X4        X5        X6  \
0   0.051259  0.000690 -0.048400 -0.015754 -0.052978  0.017801 -0.032664
1  -0.048316  0.014728 -0.062590 -0.159565  0.132780 -0.012115  0.083942
2   0.099394  0.090723 -0.016915 -0.135168  0.035190  0.101196 -0.080053
3   0.046510  0.138997 -0.083147  0.171023  0.023618  0.099237 -0.057729
4   0.133884  0.019570  0.046151  0.049495 -0.092208 -0.059080 -0.079340
..       …         …         …         …         …         …
95 -0.042570  0.048552 -0.038856 -0.015031  0.037405 -0.003761 -0.047638
96 -0.044376  0.013869 -0.005466  0.053364 -0.031360 -0.036567  0.004094
97 -0.033934 -0.017214  0.015474 -0.027739 -0.011371  0.013692 -0.028193
98  0.015381 -0.105875  0.059435  0.034510 -0.047557  0.082131  0.033396
99  0.031584  0.071183  0.029517 -0.028686 -0.005184  0.069613 -0.064414

          X7        X8        X9       X10       X11       X12           X13
0  -0.049770 -0.055617  0.036160 -0.070329 -0.082556 -0.020838  no_efectores
1   0.056576 -0.019316  0.125752 -0.086119 -0.045727  0.011244  no_efectores
2  -0.016365 -0.027642  0.071447  0.044712 -0.020866 -0.068651  no_efectores
3   0.105175 -0.006188  0.080656  0.001820  0.072216  0.111836  no_efectores
4   0.009409 -0.055605 -0.071736  0.024170  0.000796 -0.095368  no_efectores
..       …         …         …         …         …         …           …
```

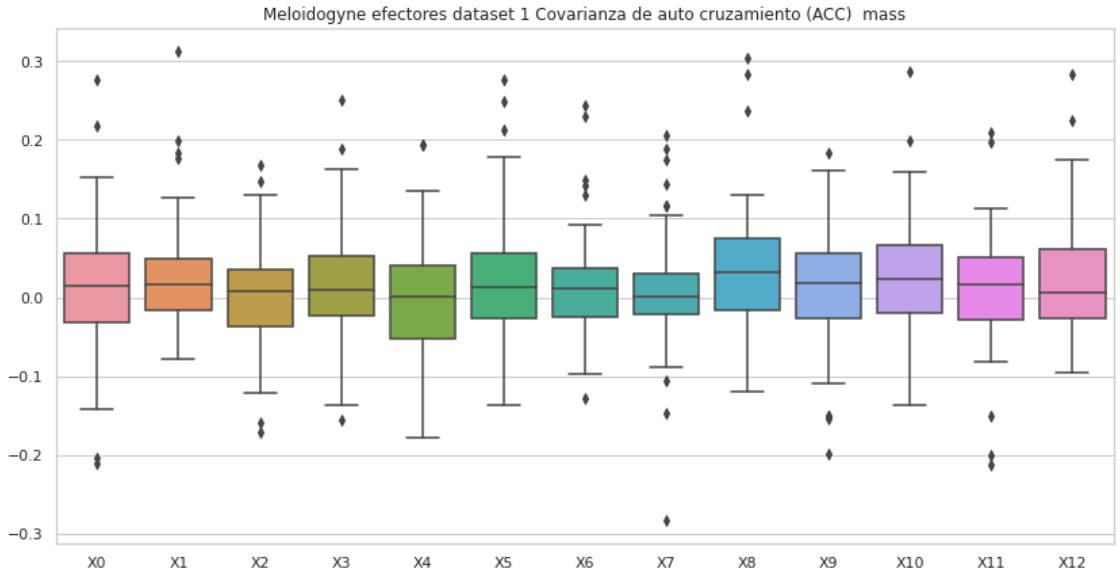
```
95 -0.000306  0.028478  0.023346  0.023350 -0.057852 -0.008086  no_efectores
96  0.090458 -0.090443 -0.055358  0.116809 -0.035596  0.071084  no_efectores
97  0.014250 -0.004228  0.018025 -0.073557 -0.002365 -0.018636  no_efectores
98  0.019521  0.038926  0.045048 -0.016221  0.030569 -0.059675  no_efectores
99  0.011805  0.008870 -0.000037  0.035674 -0.005309  0.035080  no_efectores

[94 rows x 14 columns]
```

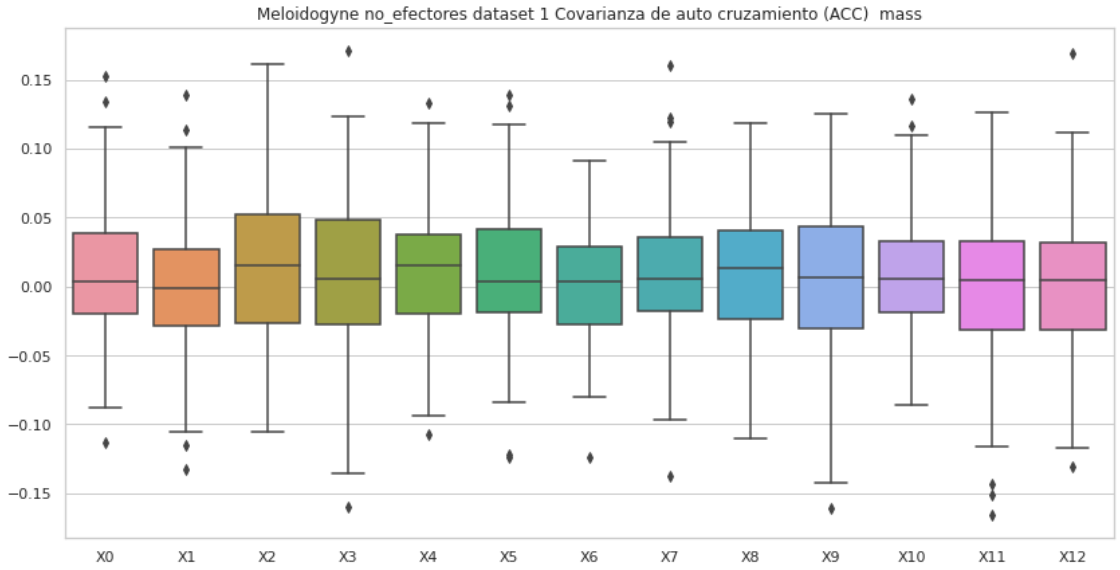
Covarianza de auto cruzamiento (ACC) mass no_efectores Meloidogyne dataset 1, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	94.000000	94.000000	94.000000	94.000000	94.000000	94.000000	
mean	0.010888	-0.002042	0.016391	0.005964	0.010021	0.009642	
std	0.048177	0.050913	0.054791	0.058924	0.044461	0.050390	
min	-0.113027	-0.132325	-0.105551	-0.159565	-0.107675	-0.123967	
25%	-0.019384	-0.028569	-0.026880	-0.027699	-0.019810	-0.018704	
50%	0.003669	-0.001343	0.015068	0.005123	0.015645	0.003483	
75%	0.038871	0.026966	0.052661	0.048902	0.037694	0.041340	
max	0.152136	0.138997	0.161712	0.171023	0.132780	0.139449	

	X6	X7	X8	X9	X10	X11	\
count	94.000000	94.000000	94.000000	94.000000	94.000000	94.000000	
mean	0.001715	0.008022	0.006574	0.001347	0.007345	0.000256	
std	0.045021	0.049845	0.049615	0.055648	0.043487	0.058487	
min	-0.123432	-0.137968	-0.110437	-0.161313	-0.086119	-0.165787	
25%	-0.027784	-0.018217	-0.023373	-0.030893	-0.019192	-0.031906	
50%	0.003606	0.005878	0.013039	0.006411	0.005119	0.004955	
75%	0.028681	0.035961	0.040359	0.043925	0.032665	0.032706	
max	0.091412	0.160266	0.118471	0.125752	0.136189	0.126662	

	X12
count	94.000000
mean	0.000322
std	0.055787
min	-0.130800
25%	-0.031100
50%	0.005023
75%	0.032084
max	0.168676





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 Meloidogyne dataset 1, con
valores atípicos.
Valores del documento csv.

          X0        X1        X2        X3        X4        X5        X6  \
0   0.117608  0.061154 -0.006290  0.128834  0.108760  0.046244 -0.083620
1   0.006116  0.040256  0.085573  0.009836  0.095392  0.037092  0.138060
2   0.006297  0.103350 -0.020410  0.011157  0.011883  0.110948  0.019616
3   0.128914  0.070383  0.037067  0.023371  0.123119  0.104615  0.114291
4   0.052548 -0.039601  0.144974  0.222654 -0.004322  0.035412  0.102958
..       …       …       …       …       …       …       …
95 -0.016993  0.113756 -0.081239  0.026624 -0.127232  0.040674 -0.017737
96  0.369266  0.335450  0.290589  0.277131  0.262565  0.271424  0.316538
97 -0.027047 -0.079922  0.106599  0.183933 -0.023468 -0.122658  0.067190
98  0.606838  0.523570  0.615178  0.569969  0.598898  0.508538  0.474328
99 -0.064544 -0.001975  0.147000  0.047178 -0.014108  0.145410  0.079621

          X7        X8        X9       X10       X11       X12        X13
0   0.053393 -0.014573  0.113240  0.064022 -0.041749 -0.067803  efectores
1  -0.069523  0.049469  0.028341 -0.025499  0.087526  0.000556  efectores
2  -0.085058  0.044598 -0.062720 -0.022593  0.041408  0.020963  efectores
```

```
3  -0.004265  0.194290  0.143070 -0.007102  0.039075 -0.013278  efectores
4   0.113608 -0.065268  0.171744  0.001858 -0.049987  0.045358  efectores
..       …       …       …       …       …       …       …
95 -0.010933  0.135464  0.035428  0.098815 -0.055571  0.050810  efectores
96  0.296736  0.343878  0.287837  0.310060  0.287351  0.279077  efectores
97 -0.041757 -0.007865  0.013158  0.065139  0.007978  0.101401  efectores
98  0.482565  0.357388  0.327567  0.256904  0.179603  0.398254  efectores
99 -0.010421  0.095491  0.170737 -0.048994 -0.084257  0.205339  efectores

[100 rows x 14 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	
mean	0.065805	0.073121	0.080853	0.078551	0.054273	0.047635	
std	0.134958	0.154316	0.135988	0.115554	0.139850	0.111698	
min	-0.141160	-0.190564	-0.121116	-0.106794	-0.183891	-0.168304	
25%	-0.017606	-0.020152	-0.016953	0.010876	-0.028434	-0.016414	
50%	0.025328	0.047097	0.057195	0.050734	0.036100	0.039599	
75%	0.119325	0.116934	0.152420	0.129356	0.111748	0.091576	
max	0.647257	0.751686	0.740489	0.569969	0.605469	0.508538	

	X6	X7	X8	X9	X10	X11	\
count	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	
mean	0.065069	0.041965	0.038773	0.036822	0.043183	0.024770	
std	0.115078	0.104745	0.113575	0.090478	0.091666	0.101368	
min	-0.157831	-0.128874	-0.346655	-0.121852	-0.256946	-0.318202	
25%	-0.014556	-0.011694	-0.023286	-0.021089	-0.011640	-0.036197	
50%	0.050092	0.021485	0.033481	0.021949	0.023094	0.010854	
75%	0.115384	0.083178	0.094601	0.079293	0.084069	0.085098	
max	0.474328	0.482565	0.394873	0.341708	0.323615	0.333918	

	X12
count	100.000000
mean	0.019454
std	0.113268
min	-0.419976
25%	-0.029486
50%	0.019555
75%	0.081229
max	0.398254

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.221055	-0.252138	0.030118	0.098138	0.062300	-0.021324	0.023311	
1	-0.048871	-0.137078	0.196053	-0.040163	0.084065	0.144761	0.100809	
2	-0.153624	-0.036566	-0.092275	-0.042491	0.017350	0.092534	-0.082706	
3	0.202551	-0.087901	0.087729	0.131130	-0.020374	0.052696	0.075656	
4	-0.035255	0.004265	0.103618	-0.148085	-0.117038	-0.078830	0.151140	
..	…	…	…	…	…	…	…	
95	-0.051669	-0.127815	-0.025792	-0.078221	-0.014507	-0.046316	0.104070	
96	0.016788	-0.130706	0.167419	0.161753	-0.080679	-0.037425	0.058198	
97	0.083649	-0.151580	0.059474	0.109872	-0.132089	-0.076389	0.038836	
98	-0.015382	-0.093697	-0.066900	0.031255	0.003852	-0.033690	-0.043951	
99	-0.016186	-0.001635	0.049383	0.004812	-0.000601	0.065026	0.025671	

	X7	X8	X9	X10	X11	X12	X13
0	-0.069250	-0.092079	0.082621	0.115757	-0.049889	-0.003391	no_efectores
1	-0.313036	0.037430	0.189067	-0.052357	0.565713	0.158591	no_efectores
2	-0.078277	-0.000614	0.009347	0.199358	0.092371	-0.076173	no_efectores
3	0.104127	0.028334	-0.041008	-0.004003	0.002813	-0.117091	no_efectores
4	-0.104214	-0.098319	-0.068373	-0.085521	0.005185	0.061230	no_efectores
..	…	…	…	…	…	…	…
95	-0.038682	-0.024221	0.075686	-0.035940	-0.072626	-0.131122	no_efectores
96	-0.010663	0.015414	0.065993	0.070547	0.092405	0.046765	no_efectores
97	0.017910	-0.077732	-0.053833	0.024591	0.064328	-0.079908	no_efectores
98	0.075498	-0.072482	-0.047965	0.035564	0.013590	0.141468	no_efectores
99	0.058432	0.005974	0.032826	-0.007544	0.040453	-0.021775	no_efectores

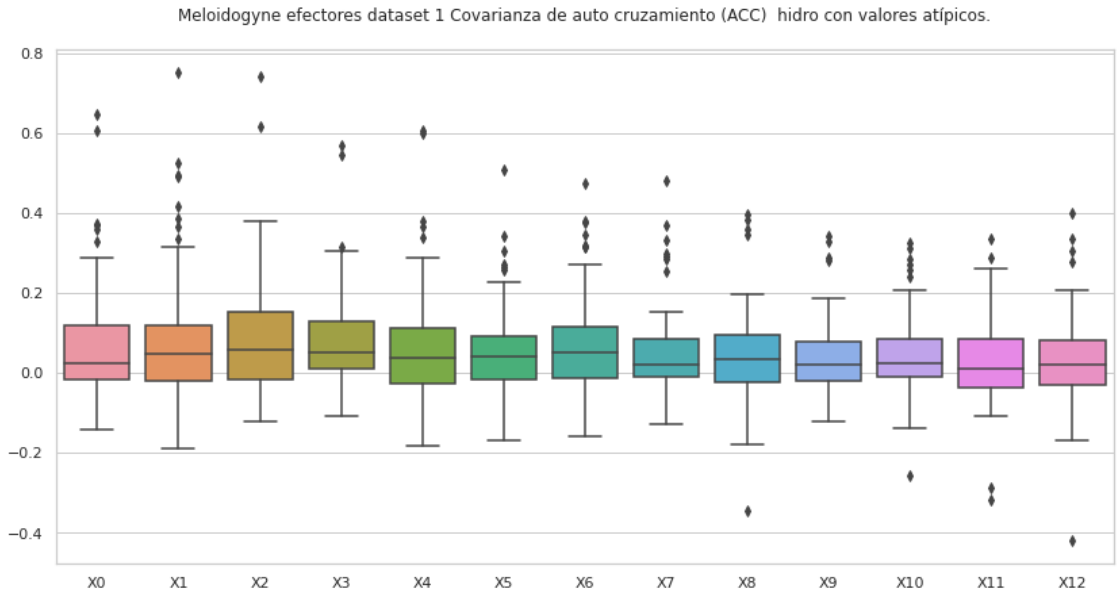
[100 rows x 14 columns]

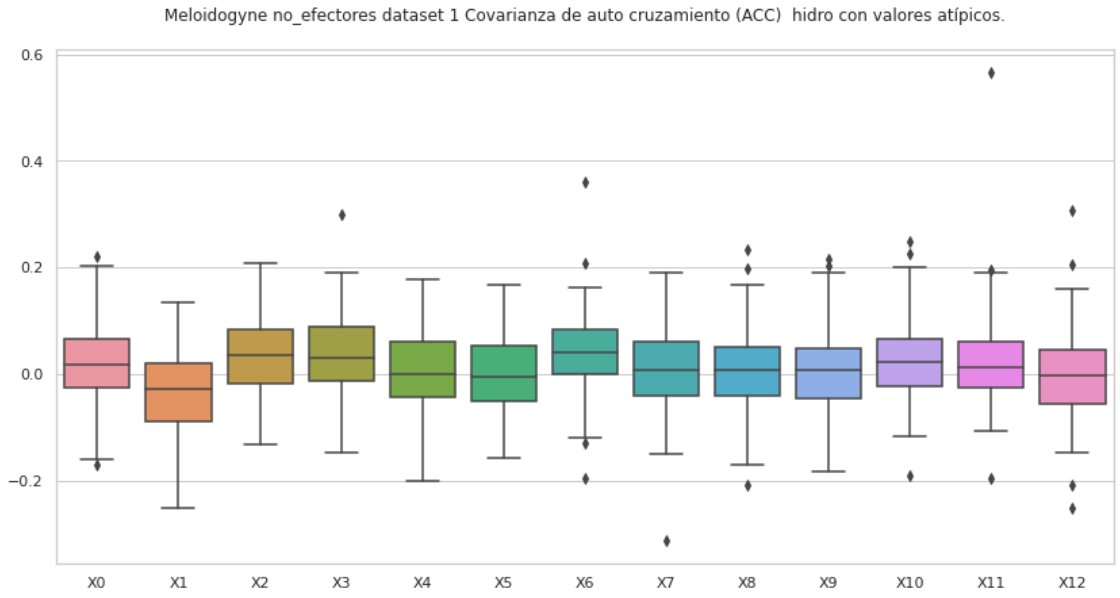
Covarianza de auto cruzamiento (ACC) hidro no_efectores Meloidogyne dataset 1,
con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	
mean	0.018337	-0.034059	0.036404	0.033543	0.002478	0.000733	
std	0.074303	0.081516	0.078508	0.079490	0.070951	0.070145	
min	-0.171476	-0.252138	-0.133177	-0.148085	-0.201305	-0.156886	
25%	-0.026427	-0.089350	-0.018111	-0.012893	-0.044907	-0.050943	
50%	0.016749	-0.029211	0.033785	0.030842	-0.001616	-0.005588	
75%	0.066680	0.020449	0.083588	0.089325	0.059550	0.053266	
max	0.221055	0.132885	0.207184	0.299224	0.177508	0.167001	

	X6	X7	X8	X9	X10	X11	\
count	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	
mean	0.038147	0.007870	0.005761	0.004944	0.023646	0.020682	
std	0.077515	0.076064	0.075748	0.075343	0.072507	0.085170	
min	-0.195906	-0.313036	-0.208055	-0.183839	-0.192197	-0.196647	
25%	-0.001824	-0.040537	-0.042370	-0.045121	-0.024377	-0.027241	
50%	0.040902	0.007389	0.006334	0.005801	0.021055	0.011646	
75%	0.082338	0.061145	0.051350	0.048723	0.066001	0.060078	
max	0.360647	0.190775	0.231953	0.214581	0.248296	0.565713	

	X12
count	100.000000
mean	-0.003421
std	0.081854
min	-0.250919
25%	-0.055624
50%	-0.002599
75%	0.045375
max	0.306082





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 Meloidogyne dataset 1, sin
valores atípicos.
Valores del documento csv.

          X0        X1        X2        X3        X4        X5        X6  \
0   0.117608  0.061154 -0.006290  0.128834  0.108760  0.046244 -0.083620
1   0.006116  0.040256  0.085573  0.009836  0.095392  0.037092  0.138060
2   0.006297  0.103350 -0.020410  0.011157  0.011883  0.110948  0.019616
3   0.128914  0.070383  0.037067  0.023371  0.123119  0.104615  0.114291
4   0.052548 -0.039601  0.144974  0.222654 -0.004322  0.035412  0.102958
..       …       …       …       …       …       …       …
94 -0.141160  0.128452 -0.035979  0.026444  0.032158  0.016447  0.118664
95 -0.016993  0.113756 -0.081239  0.026624 -0.127232  0.040674 -0.017737
96  0.369266  0.335450  0.290589  0.277131  0.262565  0.271424  0.316538
97 -0.027047 -0.079922  0.106599  0.183933 -0.023468 -0.122658  0.067190
99 -0.064544 -0.001975  0.147000  0.047178 -0.014108  0.145410  0.079621

          X7        X8        X9       X10       X11       X12        X13
0   0.053393 -0.014573  0.113240  0.064022 -0.041749 -0.067803  efectores
1  -0.069523  0.049469  0.028341 -0.025499  0.087526  0.000556  efectores
2  -0.085058  0.044598 -0.062720 -0.022593  0.041408  0.020963  efectores
3  -0.004265  0.194290  0.143070 -0.007102  0.039075 -0.013278  efectores
4   0.113608 -0.065268  0.171744  0.001858 -0.049987  0.045358  efectores
..       …       …       …       …       …       …       …
94 -0.028063  0.016116 -0.061112  0.010552  0.067207 -0.022610  efectores
```

```
95 -0.010933  0.135464  0.035428  0.098815 -0.055571  0.050810  efectores
96  0.296736  0.343878  0.287837  0.310060  0.287351  0.279077  efectores
97 -0.041757 -0.007865  0.013158  0.065139  0.007978  0.101401  efectores
99 -0.010421  0.095491  0.170737 -0.048994 -0.084257  0.205339  efectores

[94 rows x 14 columns]
```

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

```
              X0         X1         X2         X3         X4         X5  \
count  94.000000  94.000000  94.000000  94.000000  94.000000  94.000000
mean    0.050064   0.052103   0.062618   0.064175   0.038053   0.031668
std     0.099871   0.124507   0.098348   0.088386   0.112536   0.090521
min    -0.141160  -0.190564  -0.121116  -0.106794  -0.183891  -0.168304
25%    -0.018833  -0.034307  -0.018021   0.010313  -0.029488  -0.033202
50%     0.023122   0.040502   0.052706   0.040777   0.032853   0.036252
75%     0.108581   0.104154   0.149771   0.116814   0.106777   0.082328
max     0.369266   0.493866   0.379847   0.301968   0.377851   0.271424

              X6         X7         X8         X9        X10        X11  \
count  94.000000  94.000000  94.000000  94.000000  94.000000  94.000000
mean    0.054611   0.026507   0.034795   0.027796   0.040098   0.024852
std     0.097174   0.081889   0.089058   0.076150   0.076718   0.081904
min    -0.106601  -0.128874  -0.179100  -0.121852  -0.137637  -0.110051
25%    -0.016676  -0.015822  -0.019491  -0.023928  -0.009049  -0.035941
50%     0.045850   0.018890   0.033481   0.021445   0.023094   0.010854
75%     0.101689   0.071927   0.089826   0.066503   0.077116   0.082106
max     0.377863   0.330375   0.343878   0.287837   0.310060   0.287351

             X12
count  94.000000
mean    0.013849
std     0.089650
min    -0.169147
25%    -0.033692
50%     0.017390
75%     0.079198
max     0.279077
```

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

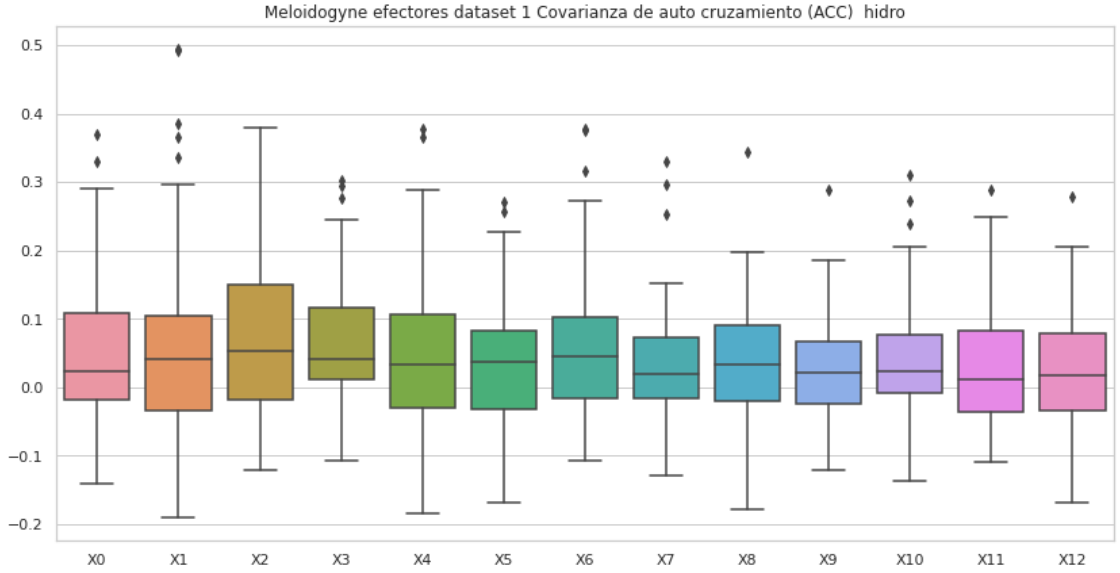
```
          X0        X1        X2        X3        X4        X5        X6  \
0   0.221055 -0.252138  0.030118  0.098138  0.062300 -0.021324  0.023311
2  -0.153624 -0.036566 -0.092275 -0.042491  0.017350  0.092534 -0.082706
3   0.202551 -0.087901  0.087729  0.131130 -0.020374  0.052696  0.075656
4  -0.035255  0.004265  0.103618 -0.148085 -0.117038 -0.078830  0.151140
5   0.066243  0.013390  0.007320  0.025889 -0.009701  0.069010  0.027550
..       …         …         …         …         …         …         …
95 -0.051669 -0.127815 -0.025792 -0.078221 -0.014507 -0.046316  0.104070
96  0.016788 -0.130706  0.167419  0.161753 -0.080679 -0.037425  0.058198
97  0.083649 -0.151580  0.059474  0.109872 -0.132089 -0.076389  0.038836
98 -0.015382 -0.093697 -0.066900  0.031255  0.003852 -0.033690 -0.043951
99 -0.016186 -0.001635  0.049383  0.004812 -0.000601  0.065026  0.025671

          X7        X8        X9       X10       X11       X12           X13
0  -0.069250 -0.092079  0.082621  0.115757 -0.049889 -0.003391  no_efectores
2  -0.078277 -0.000614  0.009347  0.199358  0.092371 -0.076173  no_efectores
3   0.104127  0.028334 -0.041008 -0.004003  0.002813 -0.117091  no_efectores
4  -0.104214 -0.098319 -0.068373 -0.085521  0.005185  0.061230  no_efectores
5   0.042111  0.025991  0.030363 -0.002409  0.008538 -0.009277  no_efectores
..       …         …         …         …         …         …           …
95 -0.038682 -0.024221  0.075686 -0.035940 -0.072626 -0.131122  no_efectores
96 -0.010663  0.015414  0.065993  0.070547  0.092405  0.046765  no_efectores
97  0.017910 -0.077732 -0.053833  0.024591  0.064328 -0.079908  no_efectores
98  0.075498 -0.072482 -0.047965  0.035564  0.013590  0.141468  no_efectores
99  0.058432  0.005974  0.032826 -0.007544  0.040453 -0.021775  no_efectores

[92 rows x 14 columns]
```

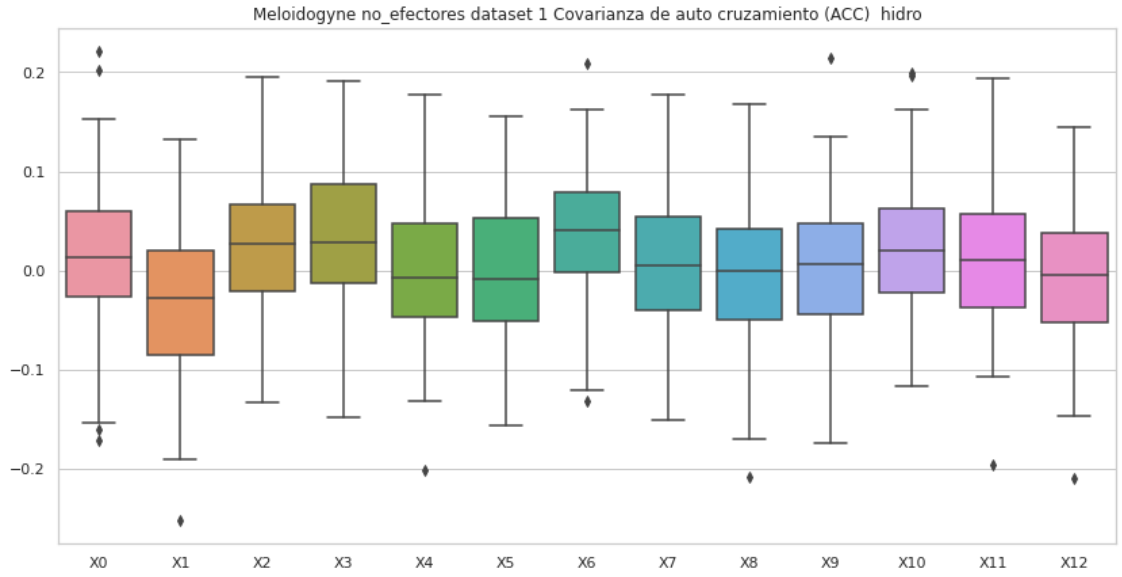
Covarianza de auto cruzamiento (ACC) no_efectores Meloidogyne dataset 1, sin valores atípicos.
Estadísticas.

```
              X0         X1         X2         X3         X4         X5  \
count  92.000000  92.000000  92.000000  92.000000  92.000000  92.000000
mean    0.015938  -0.033197   0.028485   0.028835  -0.002103  -0.003063
std     0.073394   0.081411   0.073855   0.072654   0.067604   0.066131
min    -0.171476  -0.252138  -0.133177  -0.148085  -0.201305  -0.156886
25%    -0.026427  -0.085545  -0.021347  -0.012893  -0.046494  -0.050943
50%     0.012920  -0.028076   0.027045   0.028647  -0.006704  -0.009281
75%     0.059186   0.020449   0.066441   0.086805   0.047441   0.052610
max     0.221055   0.132885   0.195380   0.190445   0.177508   0.155830

              X6         X7         X8         X9        X10        X11  \
count  92.000000  92.000000  92.000000  92.000000  92.000000  92.000000
mean    0.036566   0.007437  -0.002434   0.003876   0.022458   0.013813
std     0.068065   0.066383   0.069897   0.064261   0.062891   0.066690
```

min	-0.131226	-0.151207	-0.208055	-0.173536	-0.116724	-0.196647
25%	-0.001824	-0.040537	-0.049532	-0.043673	-0.022930	-0.036706
50%	0.039994	0.004557	-0.000794	0.005801	0.020629	0.010848
75%	0.079201	0.054675	0.042331	0.047244	0.062486	0.056437
max	0.208795	0.177193	0.167209	0.214581	0.199358	0.193848

	X12
count	92.000000
mean	-0.007397
std	0.066828
min	-0.209557
25%	-0.052711
50%	-0.004281
75%	0.038507
max	0.145156




Meloidogyne no_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