ds3_nematoda_limpieza_de_datos

December 14, 2020

Limpieza de datos

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

1 Declaración de variables

[2]:
```
organismo ="nematoda"
dataset = 3
nombre = ("ds" + str(dataset) + "_" + str(organismo))
nombre2 = (str(organismo)+ " dataset " + str(dataset))
r2 = ("Datos/resultados/"+ str(organismo) + "/" + str(nombre) + "/
 ↪transformaciones/sin_filtrar")
r3 = ("Datos/resultados/"+ str(organismo) + "/" + str(nombre) + "/
 ↪transformaciones/sin_atipicos")

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

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

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

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

2 Composición de aminoácidos (AAC)

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

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

    if etiq == "efectores":
        df=AAC_efec

    if etiq == "no_efectores":
        df=AAC_no_efec

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

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

```
efectores
Composición de aminoácidos (AAC) efectores nematoda dataset 3, con valores
atípicos.
Valores del documento csv.

         X0      X1     X2     X3     X4     X5     X6      X7     X8     X9  \
0     7.426   5.941  3.465  6.436  3.960  6.436  5.446   3.465  2.970  4.950
1     9.901  16.832  2.970  3.960  2.970  5.941  4.950   1.980  1.980  3.960
2     7.500  10.000  0.000  2.500  0.000  5.000  5.000   0.000  0.000  5.000
3     4.286  14.286  1.429  2.857  1.429  5.714  4.286   2.857  1.429  4.286
4     5.941   4.950  6.931  6.931  0.990  3.960  2.970   1.980  0.990  1.980
..      …      …      …      …      …      …      …       …      …      …
495   7.036   3.412  3.838  5.330  0.853  6.823  3.838  11.940  1.493  3.838
496   6.027   4.932  6.027  6.575  0.822  7.945  3.288   2.740  2.740  3.288
497  12.000   5.333  2.667  2.667  2.222  3.556  2.667   5.333  0.444  7.111
498   8.750   6.528  4.306  6.528  2.500  6.250  2.222   6.111  3.889  4.444
```

```
499   7.179   7.179  6.667  6.154  1.026  9.231  3.590   5.641  2.051  5.641

     …     X11     X12     X13    X14     X15     X16    X17    X18     X19  \
0    …   8.911   3.960   2.970  3.960   5.941   1.980  1.485  3.465   6.931
1    …   6.931   1.980   3.960  3.960  10.891   3.960  1.980  1.980   4.950
2    …   5.000  12.500   2.500  7.500  12.500   2.500  0.000  0.000  12.500
3    …   2.857   4.286  10.000  2.857   1.429  10.000  2.857  7.143   7.143
4    …   6.931   5.941   1.980  3.960  11.881   7.921  1.980  7.921   2.970
..   …     …     …      …     …      …     …    …     …      …
495  …   7.036   1.706   4.051  7.036   6.183   7.463  0.853  3.838   7.249
496  …  10.685   1.644   2.192  4.658  12.877   6.849  0.274  2.740   5.753
497  …   2.222   4.444   8.000  3.111  10.667   4.000  1.778  3.111   5.778
498  …   4.722   3.056   2.917  3.611   6.111   4.028  0.833  2.500   8.889
499  …   6.154   2.564   3.077  4.615   5.128   1.538  1.026  3.077   6.154

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

[500 rows x 21 columns]


Composición de aminoácidos (AAC) efectores nematoda dataset 3, con valores
atípicos.
Estadísticas.

                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     6.809128    6.039334    4.410460    5.031770    2.338910    6.072492
std      2.533559    2.652233    2.039699    2.498164    1.938216    3.017880
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      5.336250    4.462750    3.052000    3.544500    1.159500    4.042000
50%      6.511000    5.777000    4.305500    4.962000    1.951000    5.882000
75%      8.000000    7.347000    5.556000    6.306000    2.972250    7.411750
max     18.531000   21.053000   12.346000   28.788000   17.647000   16.854000

                X6          X7          X8          X9         X10         X11  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     3.923442    5.736838    2.394836    5.687022    8.872892    5.875948
```

```
std      2.395484    3.346760    1.442926    2.429213    2.962644    3.011283
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      2.564000    3.844750    1.446000    4.185750    7.071000    3.964250
50%      3.618500    5.371000    2.353000    5.401500    8.730000    5.579000
75%      4.786000    6.898250    3.078750    7.150500   10.620000    7.390500
max     29.412000   32.090000   13.187000   21.667000   20.408000   26.316000

               X12         X13         X14         X15         X16         X17  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     2.956944    4.624396    4.853238    7.737420    5.610934    1.229042
std      1.507993    2.492863    3.493567    2.994639    2.067805    1.015367
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      1.978500    3.073000    3.113250    5.820250    4.348000    0.495000
50%      2.694500    4.334000    4.397000    7.374000    5.508000    1.101500
75%      3.695750    5.848250    5.618000    9.168000    6.859500    1.686750
max     12.500000   29.487000   49.231000   19.192000   14.118000    6.667000

               X18        X19
count  500.000000  500.00000
mean     3.278126    6.51681
std      1.898936    2.41929
min      0.000000    0.00000
25%      2.068750    5.04000
50%      3.108500    6.25000
75%      4.214500    7.92500
max     14.407000   19.50200
```

no_efectores

Composición de aminoácidos (AAC) no_efectores nematoda dataset 3, con valores atípicos.

Valores del documento csv.

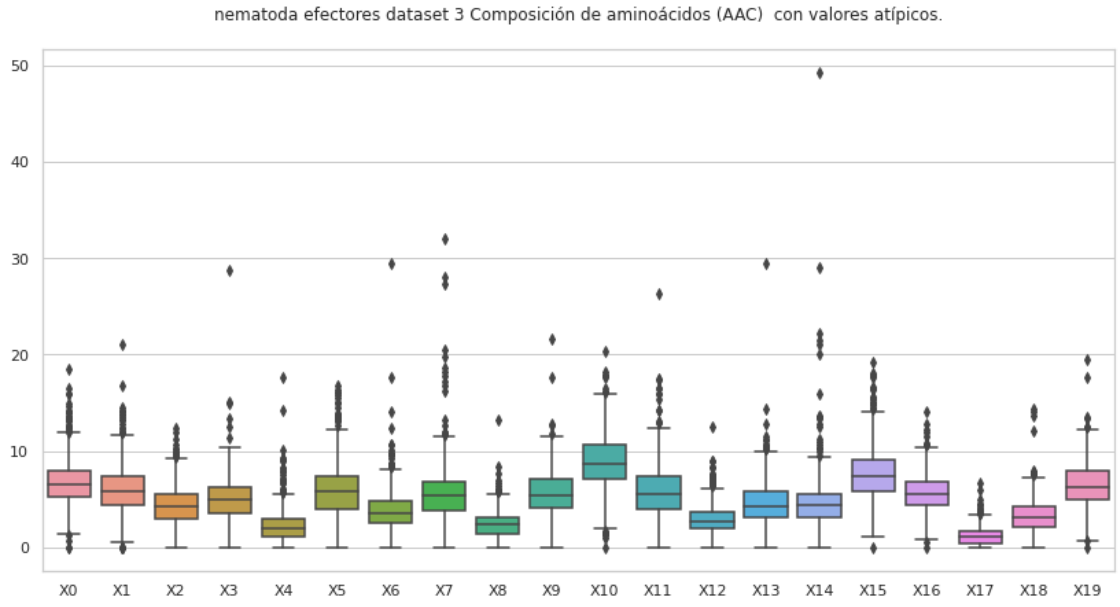
```
         X0     X1     X2      X3     X4      X5     X6     X7     X8      X9  \
0    12.857  7.143  7.143   4.286  5.714   4.286  1.429  1.429  0.000   5.714
1     7.143  3.439  4.497  10.053  1.587  11.111  2.381  5.820  2.116   7.937
2     7.552  5.469  5.729   5.729  1.042   7.812  5.469  6.771  2.865   4.167
3     7.557  5.290  3.778   3.023  2.519   2.519  4.786  5.038  2.519  11.335
4     5.852  4.647  3.442   6.196  1.721   8.950  4.819  3.959  2.410   5.680
..      …      …    …       …      …       …      …      …      …
495  11.765  5.882  3.676   5.147  1.471   8.824  2.941  2.451  2.451   7.843
496   6.852  3.640  3.212   3.854  5.782  11.991  2.784  6.210  1.499   4.711
497   4.697  6.849  4.697   6.262  4.305   4.697  4.110  4.305  1.957   7.241
498   3.518  8.543  4.020   7.035  3.015   4.523  3.518  6.030  2.010   8.040
499   6.276  4.184  5.858   2.929  2.092   3.766  2.510  3.347  1.674   6.695

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

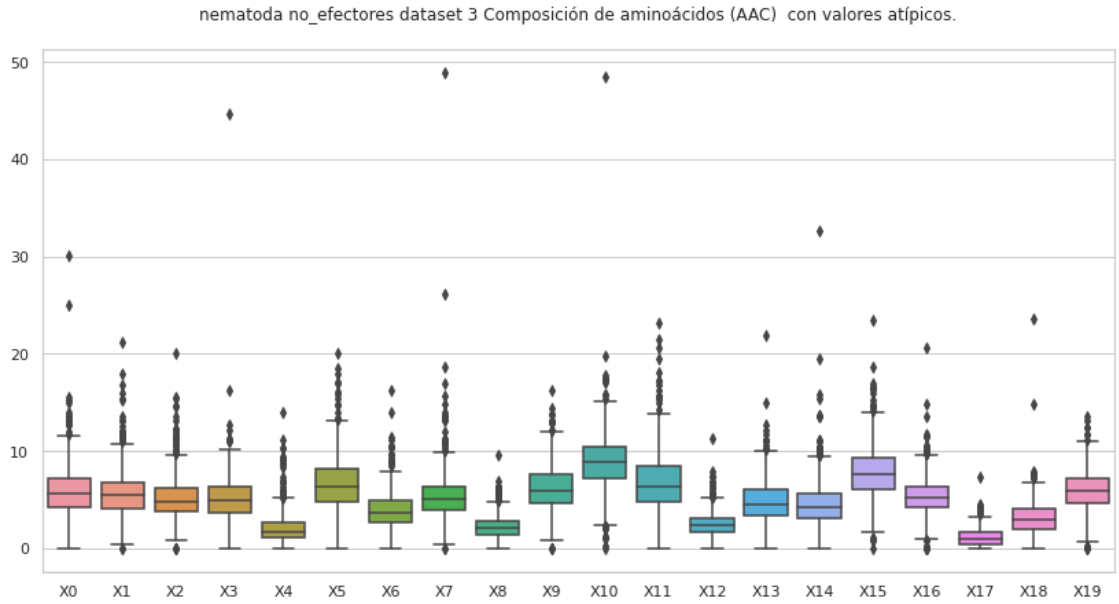
```
0    …   8.571  4.286  4.286  7.143  7.143  2.857  2.857  2.857  4.286
1    …  11.905  1.058  2.910  3.439  3.704  3.175  0.794  1.852  6.349
2    …   6.771  2.865  4.948  5.208  7.812  4.427  0.260  2.083  4.427
3    …   3.023  3.023  8.816  2.771  7.809  3.526  1.259  3.023  6.297
4    …   9.639  2.582  3.959  4.991  6.540  6.540  0.516  2.754  6.713
..   …     …    …    …    …    …    …    …    …    …
495  …   9.314  2.451  3.186  1.961  7.598  3.431  0.245  3.431  6.127
496  …  10.064  2.355  3.212  4.283  8.779  7.066  0.857  2.141  6.210
497  …   3.131  1.761  3.523  3.718  9.589  5.284  0.196  3.131  5.479
498  …   4.020  2.513  5.528  7.035  5.025  4.020  1.005  3.015  8.543
499  …   5.439  4.603  5.858  4.603  5.858  9.623  2.092  3.766  5.439

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

[500 rows x 21 columns]
```

Composición de aminoácidos (AAC) no_efectores nematoda dataset 3, con valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count   500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean      6.006338    5.600768    5.262376    5.079808    2.128404    6.714734
std       2.866247    2.631884    2.407633    2.797731    1.817490    2.969434
min       0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%       4.306750    4.054000    3.826500    3.620750    1.062250    4.828000
50%       5.659000    5.464500    4.849500    5.008000    1.731500    6.305000
75%       7.220750    6.718000    6.209750    6.340750    2.703000    8.186500
max      30.120000   21.212000   20.134000   44.715000   14.013000   20.134000

                X6          X7          X8          X9         X10         X11  \
count   500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean      3.887828    5.429862    2.238290    6.213740    8.956788    6.860034
std       2.007169    3.248521    1.292904    2.491856    3.390382    3.250935
min       0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%       2.663500    3.907250    1.466250    4.681250    7.216750    4.800000
```

```
50%      3.622000    5.080000    2.105500    5.958000    8.887500    6.341000
75%      4.890500    6.359250    2.819000    7.622250   10.469250    8.548500
max     16.320000   48.892000    9.639000   16.239000   48.449000   23.200000

              X12         X13         X14         X15         X16         X17  \
count  500.000000  500.000000  500.00000  500.000000  500.000000  500.000000
mean     2.515828    4.873424    4.70791    7.915898    5.378348    1.174434
std      1.292354    2.354156    2.76313    2.778729    2.147377    0.966822
min      0.000000    0.000000    0.00000    0.000000    0.000000    0.000000
25%      1.649500    3.430750    3.12300    6.131000    4.222250    0.476250
50%      2.354000    4.564000    4.21300    7.692000    5.285500    1.011000
75%      3.059750    6.085250    5.70100    9.350750    6.395750    1.629750
max     11.364000   21.944000   32.71000   23.507000   20.619000    7.333000

              X18         X19
count  500.000000  500.000000
mean     3.096562    5.958668
std      1.898575    2.104613
min      0.000000    0.000000
25%      2.017750    4.623000
50%      2.997000    5.958500
75%      4.070000    7.242250
max     23.585000   13.514000
```





2.1 Composición de aminoácidos (AAC), sin valores atípicos

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

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


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

    if etiq == "efectores":
        df=AAC_efec

    if etiq == "no_efectores":
        df=AAC_no_efec

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

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

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



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

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

```
efectores
Composición de aminoácidos (AAC) efectores nematoda dataset 3, sin valores
atípicos.
Valores del documento csv.

         X0      X1     X2     X3     X4     X5     X6      X7     X8     X9  \
0     7.426   5.941  3.465  6.436  3.960  6.436  5.446   3.465  2.970  4.950
4     5.941   4.950  6.931  6.931  0.990  3.960  2.970   1.980  0.990  1.980
5     2.817   8.451  5.634  5.634  4.225  7.042  0.000   7.042  4.225  4.225
6    11.905  10.714  7.143  3.571  2.381  2.381  0.000   8.333  0.000  8.333
8     6.024   8.434  9.639  3.614  6.024  3.614  6.024   3.614  2.410  6.024
..      …       …      …      …      …      …      …       …      …      …
495   7.036   3.412  3.838  5.330  0.853  6.823  3.838  11.940  1.493  3.838
496   6.027   4.932  6.027  6.575  0.822  7.945  3.288   2.740  2.740  3.288
497  12.000   5.333  2.667  2.667  2.222  3.556  2.667   5.333  0.444  7.111
498   8.750   6.528  4.306  6.528  2.500  6.250  2.222   6.111  3.889  4.444
499   7.179   7.179  6.667  6.154  1.026  9.231  3.590   5.641  2.051  5.641

     …     X11    X12    X13    X14     X15    X16    X17    X18    X19  \
0    …   8.911  3.960  2.970  3.960   5.941  1.980  1.485  3.465  6.931
4    …   6.931  5.941  1.980  3.960  11.881  7.921  1.980  7.921  2.970
5    …   1.408  5.634  4.225  4.225   5.634  9.859  0.000  7.042  4.225
6    …   3.571  2.381  4.762  3.571   5.952  8.333  2.381  3.571  8.333
8    …   8.434  2.410  0.000  4.819   9.639  3.614  0.000  6.024  2.410
..   …       …      …      …      …       …      …      …      …      …
```

```
495  …   7.036  1.706  4.051  7.036   6.183  7.463  0.853  3.838  7.249
496  …  10.685  1.644  2.192  4.658  12.877  6.849  0.274  2.740  5.753
497  …   2.222  4.444  8.000  3.111  10.667  4.000  1.778  3.111  5.778
498  …   4.722  3.056  2.917  3.611   6.111  4.028  0.833  2.500  8.889
499  …   6.154  2.564  3.077  4.615   5.128  1.538  1.026  3.077  6.154

           X20
0    efectores
4    efectores
5    efectores
6    efectores
8    efectores
..         …
495  efectores
496  efectores
497  efectores
498  efectores
499  efectores

[417 rows x 21 columns]
```

Composición de aminoácidos (AAC) efectores nematoda dataset 3, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  417.000000  417.000000  417.000000  417.000000  417.000000  417.000000
mean     6.756734    6.092110    4.485398    5.109012    2.217194    6.127089
std      2.259148    2.304366    1.815248    2.022324    1.530661    2.603057
min      1.471000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      5.385000    4.552000    3.289000    3.760000    1.235000    4.464000
50%      6.436000    5.839000    4.342000    5.078000    1.961000    6.069000
75%      7.865000    7.246000    5.556000    6.349000    2.920000    7.407000
max     14.286000   13.861000   10.286000   12.500000    8.000000   14.554000

                X6          X7          X8          X9         X10         X11  \
count  417.000000  417.000000  417.000000  417.000000  417.000000  417.000000
mean     3.799592    5.481242    2.415161    5.841242    9.092602    5.893772
std      1.812777    2.239726    1.199805    2.138719    2.599760    2.415326
min      0.000000    0.000000    0.000000    0.000000    1.980000    0.000000
25%      2.667000    3.922000    1.586000    4.375000    7.347000    4.337000
50%      3.604000    5.381000    2.410000    5.571000    8.929000    5.741000
75%      4.667000    6.829000    3.077000    7.246000   10.784000    7.287000
max     10.740000   13.265000    6.667000   12.613000   17.647000   14.286000

               X12         X13         X14         X15         X16         X17  \
count  417.000000  417.000000  417.000000  417.000000  417.000000  417.000000
```

mean	2.845444	4.715727	4.534384	7.844583	5.616564	1.221964
std	1.245925	2.048263	1.927563	2.573204	1.859148	0.896744
min	0.429000	0.000000	0.000000	1.639000	0.503000	0.000000
25%	1.983000	3.352000	3.247000	5.970000	4.412000	0.573000
50%	2.691000	4.484000	4.367000	7.438000	5.525000	1.111000
75%	3.472000	5.882000	5.426000	9.231000	6.818000	1.684000
max	7.407000	11.465000	13.514000	16.516000	11.594000	4.000000

	X18	X19
count	417.000000	417.000000
mean	3.308285	6.601897
std	1.557814	2.186805
min	0.000000	0.000000
25%	2.198000	5.263000
50%	3.193000	6.439000
75%	4.206000	8.000000
max	7.921000	13.453000

no_efectores
Composición de aminoácidos (AAC) no_efectores nematoda dataset 3, sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9	\
0	12.857	7.143	7.143	4.286	5.714	4.286	1.429	1.429	0.000	5.714	
1	7.143	3.439	4.497	10.053	1.587	11.111	2.381	5.820	2.116	7.937	
2	7.552	5.469	5.729	5.729	1.042	7.812	5.469	6.771	2.865	4.167	
3	7.557	5.290	3.778	3.023	2.519	2.519	4.786	5.038	2.519	11.335	
4	5.852	4.647	3.442	6.196	1.721	8.950	4.819	3.959	2.410	5.680	
..	…	…	…	…	…	…	…	…	…	…	
495	11.765	5.882	3.676	5.147	1.471	8.824	2.941	2.451	2.451	7.843	
496	6.852	3.640	3.212	3.854	5.782	11.991	2.784	6.210	1.499	4.711	
497	4.697	6.849	4.697	6.262	4.305	4.697	4.110	4.305	1.957	7.241	
498	3.518	8.543	4.020	7.035	3.015	4.523	3.518	6.030	2.010	8.040	
499	6.276	4.184	5.858	2.929	2.092	3.766	2.510	3.347	1.674	6.695	

	…	X11	X12	X13	X14	X15	X16	X17	X18	X19	\
0	…	8.571	4.286	4.286	7.143	7.143	2.857	2.857	2.857	4.286	
1	…	11.905	1.058	2.910	3.439	3.704	3.175	0.794	1.852	6.349	
2	…	6.771	2.865	4.948	5.208	7.812	4.427	0.260	2.083	4.427	
3	…	3.023	3.023	8.816	2.771	7.809	3.526	1.259	3.023	6.297	
4	…	9.639	2.582	3.959	4.991	6.540	6.540	0.516	2.754	6.713	
..	…	…	…	…	…	…	…	…	…		
495	…	9.314	2.451	3.186	1.961	7.598	3.431	0.245	3.431	6.127	
496	…	10.064	2.355	3.212	4.283	8.779	7.066	0.857	2.141	6.210	
497	…	3.131	1.761	3.523	3.718	9.589	5.284	0.196	3.131	5.479	

```
498  …   4.020  2.513  5.528  7.035  5.025  4.020  1.005  3.015  8.543
499  …   5.439  4.603  5.858  4.603  5.858  9.623  2.092  3.766  5.439

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

[411 rows x 21 columns]
```

Composición de aminoácidos (AAC) no_efectores nematoda dataset 3, sin valores atípicos.
Estadísticas.

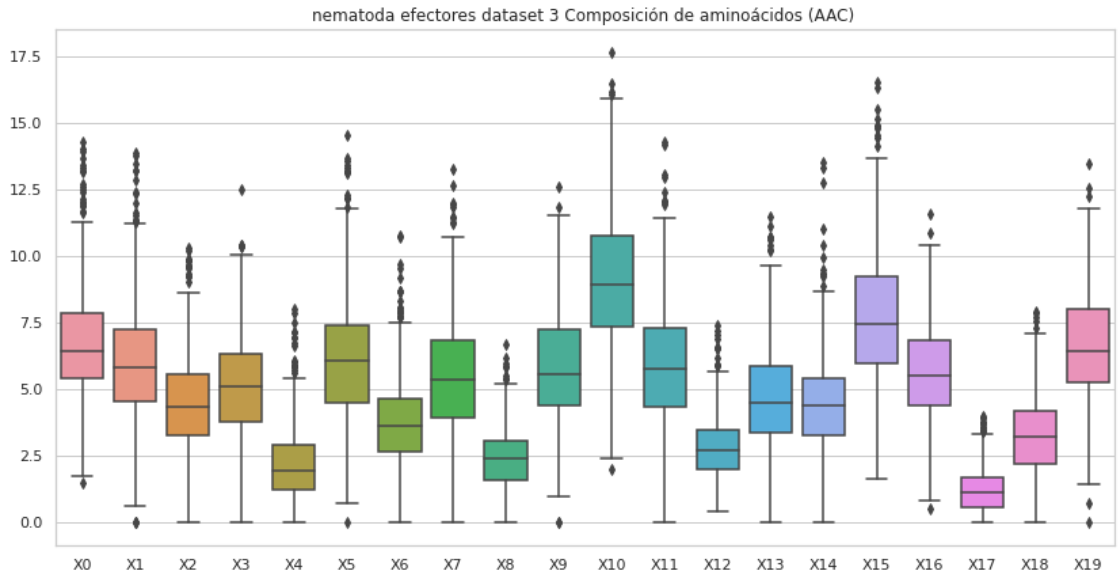
	X0	X1	X2	X3	X4	X5
count	411.000000	411.000000	411.000000	411.000000	411.000000	411.000000
mean	6.008696	5.565051	5.126611	5.122662	2.060540	6.688328
std	2.219507	2.029366	1.913439	1.908449	1.340975	2.489281
min	0.331000	0.606000	0.000000	0.000000	0.000000	0.000000
25%	4.557500	4.279000	3.914000	4.143500	1.177000	5.000000
50%	5.837000	5.505000	4.847000	5.110000	1.770000	6.331000
75%	7.205500	6.695500	6.041000	6.347000	2.604000	7.985500
max	14.004000	13.103000	12.322000	12.088000	7.353000	14.773000

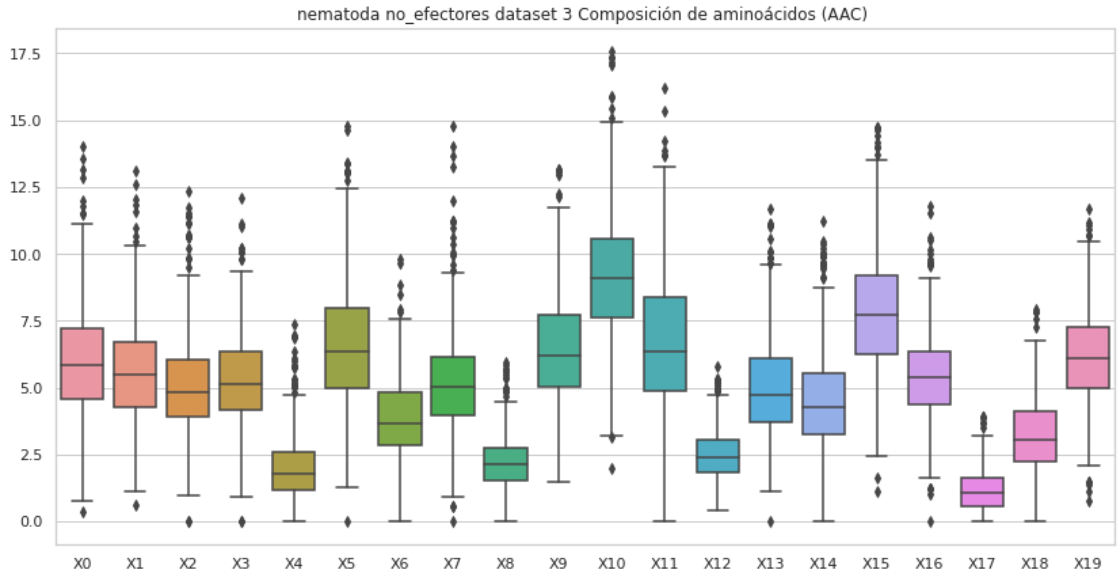
	X6	X7	X8	X9	X10	X11
count	411.000000	411.000000	411.000000	411.000000	411.000000	411.000000
mean	3.859718	5.178109	2.203440	6.412630	9.231277	6.668224
std	1.624005	2.084957	1.083721	2.126471	2.515513	2.656513
min	0.000000	0.000000	0.000000	1.471000	1.961000	0.000000
25%	2.824500	3.949500	1.507500	5.009500	7.595000	4.881000
50%	3.664000	5.024000	2.135000	6.216000	9.091000	6.366000
75%	4.823000	6.130500	2.733000	7.712000	10.560000	8.395000
max	9.790000	14.783000	5.941000	13.174000	17.568000	16.201000

	X12	X13	X14	X15	X16	X17
count	411.000000	411.000000	411.000000	411.000000	411.000000	411.000000
mean	2.499637	4.999630	4.601190	7.879365	5.437543	1.173905
std	1.011244	1.977021	1.980622	2.341358	1.733373	0.817733
min	0.405000	0.000000	0.000000	1.111000	0.000000	0.000000

25%	1.818000	3.726000	3.260500	6.267500	4.378000	0.569500
50%	2.397000	4.724000	4.260000	7.733000	5.370000	1.047000
75%	3.031000	6.074000	5.543000	9.185500	6.356500	1.644000
max	5.797000	11.688000	11.200000	14.740000	11.765000	3.922000

	X18	X19
count	411.000000	411.000000
mean	3.165229	6.118253
std	1.466506	1.786737
min	0.000000	0.746000
25%	2.252000	4.962500
50%	3.030000	6.098000
75%	4.091000	7.241000
max	7.918000	11.667000





3 Composición de pseudo aminoácidos (PseAAC) hidro_mass

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

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

    if etiq == "efectores":
        df=PseAAC_hidro_mass_efec

    if etiq == "no_efectores":
        df=PseAAC_hidro_mass_no_efec

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

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

efectores
Composición de pseudo aminoácidos (PseAAC) hidro_mass efectores nematoda
dataset 3, con valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.052053  0.027761  0.045112  0.045112  0.020821  0.024291  0.020821
1    0.030081  0.009024  0.012032  0.018049  0.012032  0.006016  0.006016
2    0.307491  0.000000  0.102497  0.204994  0.102497  0.000000  0.000000
3    0.016563  0.005521  0.011042  0.022084  0.038647  0.011042  0.005521
4    0.015532  0.002589  0.018120  0.010355  0.005177  0.005177  0.002589
..        …         …         …         …         …         …         …
495  0.023742  0.002878  0.017987  0.023023  0.013670  0.040290  0.005036
496  0.014907  0.002033  0.016262  0.019650  0.005421  0.006776  0.006776
497  0.038569  0.007142  0.008571  0.011428  0.025713  0.017142  0.001428
498  0.050283  0.014367  0.037513  0.035917  0.016761  0.035118  0.022348
499  0.039732  0.005676  0.034056  0.051084  0.017028  0.031218  0.011352

           X7        X8        X9  …       X74       X75       X76  \
0    0.034702  0.062463  0.069404  …  0.019534  0.008682  0.014497
1    0.012032  0.021057  0.012032  …  0.049216  0.029254 -0.005171
2    0.204994  0.204994  0.409989  …  1.043890  0.827113  0.117325
3    0.016563  0.011042  0.033126  …  0.018686  0.040416 -0.014304
4    0.005177  0.018120  0.028475  … -0.020667 -0.015108  0.037667
..        …         …         …  …       …         …         …
495  0.012950  0.023742  0.020865  … -0.007354  0.002107  0.018654
496  0.008131  0.026425  0.019650  …  0.019950  0.025965  0.007675
497  0.022856  0.007142  0.041426  … -0.004229 -0.007440  0.005023
498  0.025541  0.027137  0.067842  … -0.027646 -0.011185  0.021757
499  0.031218  0.034056  0.068112  …  0.007740  0.027655 -0.025516

           X77       X78       X79       X80       X81       X82         X83
0   -0.051407  0.003311 -0.009435  0.023001  0.056030  0.027411  efectores
1    0.016928  0.019320  0.004039 -0.014436 -0.009909  0.016525  efectores
2   -0.843008 -0.480917  0.304078 -1.308953 -0.778083  0.337291  efectores
3    0.000827  0.009116 -0.031933  0.003914  0.002056  0.025715  efectores
4    0.005643  0.019018  0.014884  0.015680  0.016220 -0.011417  efectores
..        …         …         …         …         …         …         …
495  0.007867  0.000920  0.027213  0.006271  0.010150  0.025487  efectores
496  0.012852  0.014462  0.016175  0.012098  0.014030  0.003432  efectores
```

```
497  0.015193  0.000576  0.006853  0.002293 -0.007782  0.006205  efectores
498 -0.001925 -0.011077  0.024603 -0.001352 -0.011683  0.023092  efectores
499 -0.002094  0.035896  0.017444  0.019893  0.027209 -0.015606  efectores

[500 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro_mass efectores nematoda
dataset 3, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.042021	0.013996	0.030731	0.035922	0.030039	0.031060	
std	0.082741	0.020411	0.048549	0.049893	0.062825	0.034287	
min	-0.022074	-0.044147	0.000000	-0.044147	-0.044147	-0.066221	
25%	0.021666	0.004705	0.013642	0.017907	0.012978	0.016852	
50%	0.032691	0.009502	0.025962	0.029426	0.022135	0.026426	
75%	0.044136	0.017223	0.037552	0.044398	0.034216	0.039353	
max	1.656900	0.331380	0.994140	0.994140	1.325520	0.662760	

	X6	X7	X8	X9	…	X73	\
count	500.000000	500.000000	500.000000	500.000000	…	500.000000	
mean	0.015654	0.036866	0.036926	0.060461	…	0.006242	
std	0.031938	0.078813	0.077462	0.166520	…	0.072129	
min	0.000000	-0.044147	-0.022074	-0.022074	…	-1.310063	
25%	0.005768	0.017095	0.017660	0.027869	…	-0.001467	
50%	0.011037	0.027404	0.028399	0.044461	…	0.008766	
75%	0.018991	0.040331	0.043217	0.064369	…	0.021878	
max	0.662760	1.656900	1.656900	3.645181	…	0.358918	

	X74	X75	X76	X77	X78	X79	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.002781	0.005870	0.004612	0.000172	0.004986	0.007976	
std	0.072809	0.076589	0.083868	0.079473	0.043964	0.041561	
min	-0.635076	-1.239796	-1.567496	-0.843008	-0.480917	-0.355593	
25%	-0.010305	-0.003274	-0.002935	-0.008981	-0.004706	0.000122	
50%	0.004239	0.007774	0.008922	0.003762	0.007074	0.009774	
75%	0.017953	0.020805	0.021739	0.016347	0.018589	0.023382	
max	1.043890	0.827113	0.202576	0.918723	0.268089	0.304078	

	X80	X81	X82
count	500.000000	500.000000	500.000000
mean	0.003142	0.008409	0.008832
std	0.123365	0.080029	0.067329
min	-1.308953	-0.778083	-0.871451
25%	-0.010623	-0.004615	-0.002677
50%	0.003532	0.005711	0.008385

```
75%      0.014693    0.019701    0.021976
max      2.162058    1.215805    0.969783

[8 rows x 83 columns]
```

```
no_efectores
Composición de pseudo aminoácidos (PseAAC)  hidro_mass no_efectores nematoda
dataset 3, con valores atípicos.
Valores del documento csv.
```

	X0	X1	X2	X3	X4	X5	X6	\
0	0.061971	0.027543	0.020657	0.020657	0.020657	0.006886	0.000000	
1	0.025673	0.005705	0.036133	0.039937	0.010460	0.020919	0.007607	
2	0.035252	0.004862	0.026743	0.036468	0.023096	0.031606	0.013372	
3	0.025967	0.008656	0.010387	0.008656	0.030295	0.017311	0.008656	
4	0.030051	0.008839	0.031819	0.045960	0.020329	0.020329	0.012374	
..	…	…	…	…	…	…	…	
495	0.069606	0.008701	0.030453	0.052205	0.018852	0.014501	0.014501	
496	0.015717	0.013261	0.008841	0.027504	0.007367	0.014243	0.003438	
497	0.042182	0.038666	0.056242	0.042182	0.031636	0.038666	0.017576	
498	0.019611	0.016809	0.039221	0.025214	0.030817	0.033618	0.011206	
499	0.024945	0.008315	0.011641	0.014967	0.023282	0.013304	0.006652	

	X7	X8	X9	…	X74	X75	X76	\
0	0.027543	0.041314	0.027543	…	0.062627	0.066320	-0.017575	
1	0.028526	0.042789	0.031379	…	-0.000996	0.029110	0.013344	
2	0.019450	0.031606	0.040115	…	-0.014937	-0.000080	0.011092	
3	0.038951	0.010387	0.041547	…	-0.000571	-0.001567	0.017045	
4	0.029167	0.049496	0.041541	…	-0.005633	0.006233	0.002548	
..	…	…	…	…	…	…	…	
495	0.046404	0.055105	0.058005	…	0.019002	0.034381	0.008386	
496	0.010805	0.023084	0.010314	…	-0.000506	0.017297	0.005288	
497	0.065030	0.028121	0.135332	…	-0.031165	-0.031659	0.006509	
498	0.044824	0.022412	0.050427	…	-0.019625	-0.005213	-0.012063	
499	0.026608	0.021619	0.053217	…	0.031658	0.022053	0.006913	

	X77	X78	X79	X80	X81	X82	X83
0	-0.012191	0.014652	0.007137	-0.054621	-0.013160	-0.024573	no_efectores
1	-0.011936	0.028765	0.008884	0.007259	0.020409	0.007852	no_efectores
2	-0.009026	-0.001697	0.005605	0.017371	0.044306	0.001026	no_efectores
3	0.035694	0.015288	0.008107	0.016652	0.009317	0.002080	no_efectores
4	-0.009640	0.021503	0.015348	-0.005744	0.012837	0.005744	no_efectores
..	…	…	…	…	…	…	…
495	0.012967	0.019396	0.028571	-0.007252	-0.001416	0.025443	no_efectores
496	0.011878	0.028791	0.017527	0.006630	0.017697	0.009892	no_efectores
497	-0.035886	-0.023151	0.017147	-0.044206	-0.051800	0.039261	no_efectores

```
498  0.014968 -0.001909  0.000544  0.034152  0.030578  0.034490  no_efectores
499  0.007720  0.011554  0.022396 -0.003920 -0.013582 -0.005735  no_efectores

[500 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores nematoda
dataset 3, con valores atípicos.
Estadísticas.

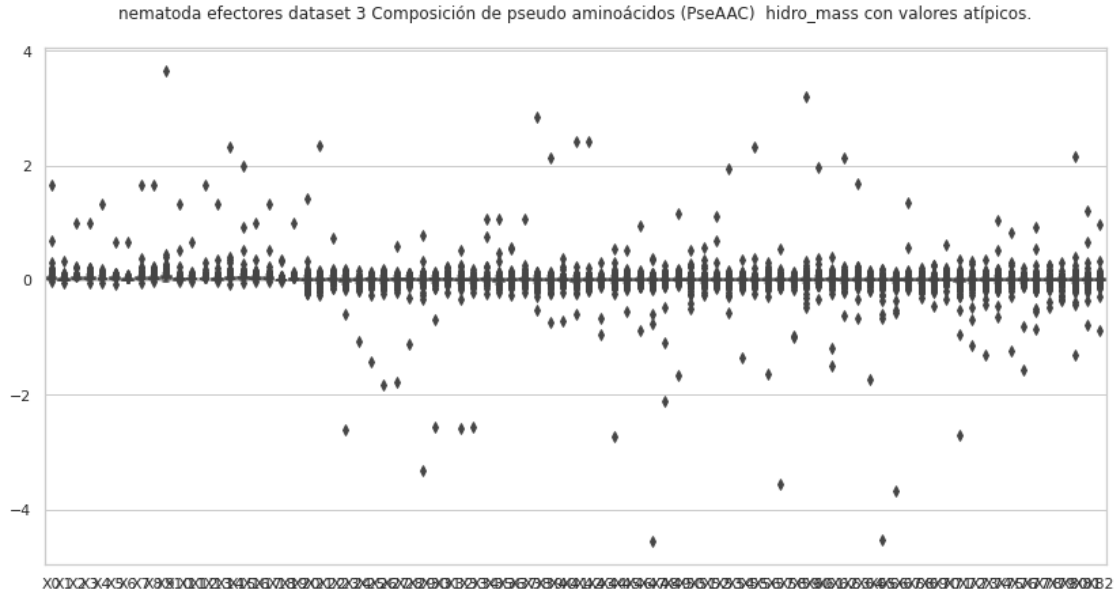
	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.049592	0.040447	0.026439	0.054113	0.057025	0.046761	
std	0.431477	0.648313	0.017384	0.431464	0.648054	0.431590	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.019330	0.004201	0.013919	0.020438	0.013937	0.015900	
50%	0.028640	0.008921	0.023866	0.032024	0.023636	0.025482	
75%	0.037608	0.015040	0.035104	0.044520	0.034583	0.034861	
max	9.670514	14.505771	0.155152	9.670514	14.505771	9.670514	

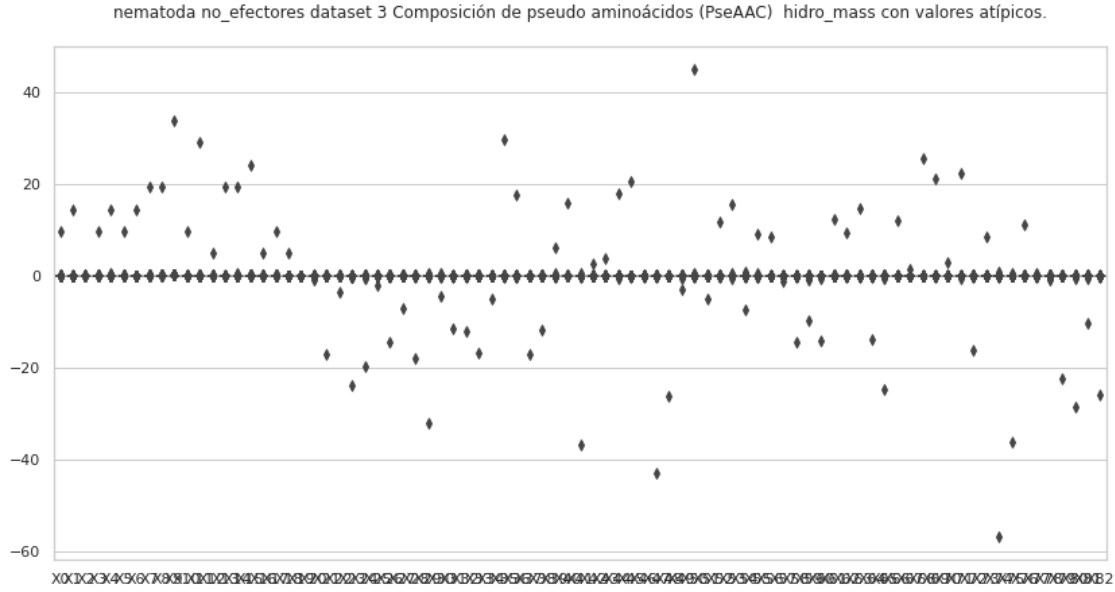
	X6	X7	X8	X9	…	X73	\
count	500.000000	500.000000	500.000000	500.000000	…	500.000000	
mean	0.041170	0.073175	0.074597	0.116998	…	0.024904	
std	0.648247	0.863867	0.863786	1.511862	…	0.376307	
min	0.000000	0.000000	0.000000	0.000000	…	-0.143745	
25%	0.005649	0.018381	0.020587	0.028779	…	-0.000085	
50%	0.010299	0.031244	0.031580	0.043058	…	0.008088	
75%	0.016343	0.044639	0.044537	0.060289	…	0.017684	
max	14.505771	19.341028	19.341028	33.846799	…	8.408614	

	X74	X75	X76	X77	X78	X79	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	-0.109468	-0.062795	0.032061	0.002357	0.005744	-0.035517	
std	2.544864	1.622886	0.504143	0.036962	0.057601	0.997647	
min	-56.892261	-36.271328	-0.113304	-0.268797	-1.033456	-22.294511	
25%	-0.007937	-0.001600	-0.000240	-0.008475	-0.002702	-0.000429	
50%	0.003956	0.008354	0.008323	0.003256	0.007760	0.008679	
75%	0.015594	0.020844	0.018996	0.013316	0.019964	0.018216	
max	0.725381	0.430034	11.270435	0.496420	0.377490	0.121025	

	X80	X81	X82
count	500.000000	500.000000	500.000000
mean	-0.056255	-0.013640	-0.042675
std	1.280240	0.467672	1.155750
min	-28.612730	-10.420946	-25.826075
25%	-0.008907	-0.003432	-0.000162
50%	0.004167	0.007568	0.008344
75%	0.015344	0.019375	0.018840

```
max       0.145884     0.239124     0.354207

[8 rows x 83 columns]
```





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

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

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


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

    if etiq == "efectores":
        df=PseAAC_hidro_mass_efec

    if etiq == "no_efectores":
        df=PseAAC_hidro_mass_no_efec

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

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


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

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

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

```
efectores
Composición de pseudo aminoácidos (PseAAC)  hidro_mass efectores nematoda
dataset 3, sin valores atípicos.
Valores del documento csv.
```

	X0	X1	X2	X3	X4	X5	X6	\
0	0.052053	0.027761	0.045112	0.045112	0.020821	0.024291	0.020821	
1	0.030081	0.009024	0.012032	0.018049	0.012032	0.006016	0.006016	
3	0.016563	0.005521	0.011042	0.022084	0.038647	0.011042	0.005521	
4	0.015532	0.002589	0.018120	0.010355	0.005177	0.005177	0.002589	
6	0.056178	0.011236	0.016854	0.011236	0.022471	0.039325	0.000000	
..	…	…	…	…	…	…	…	
495	0.023742	0.002878	0.017987	0.023023	0.013670	0.040290	0.005036	
496	0.014907	0.002033	0.016262	0.019650	0.005421	0.006776	0.006776	
497	0.038569	0.007142	0.008571	0.011428	0.025713	0.017142	0.001428	
498	0.050283	0.014367	0.037513	0.035917	0.016761	0.035118	0.022348	
499	0.039732	0.005676	0.034056	0.051084	0.017028	0.031218	0.011352	

	X7	X8	X9	…	X74	X75	X76	\
0	0.034702	0.062463	0.069404	…	0.019534	0.008682	0.014497	
1	0.012032	0.021057	0.012032	…	0.049216	0.029254	-0.005171	
3	0.016563	0.011042	0.033126	…	0.018686	0.040416	-0.014304	
4	0.005177	0.018120	0.028475	…	-0.020667	-0.015108	0.037667	
6	0.039325	0.016854	0.011236	…	0.005518	0.016937	0.013528	
..	…	…	…	…	…	…	…	
495	0.012950	0.023742	0.020865	…	-0.007354	0.002107	0.018654	
496	0.008131	0.026425	0.019650	…	0.019950	0.025965	0.007675	
497	0.022856	0.007142	0.041426	…	-0.004229	-0.007440	0.005023	
498	0.025541	0.027137	0.067842	…	-0.027646	-0.011185	0.021757	
499	0.031218	0.034056	0.068112	…	0.007740	0.027655	-0.025516	

	X77	X78	X79	X80	X81	X82	X83
0	-0.051407	0.003311	-0.009435	0.023001	0.056030	0.027411	efectores
1	0.016928	0.019320	0.004039	-0.014436	-0.009909	0.016525	efectores
3	0.000827	0.009116	-0.031933	0.003914	0.002056	0.025715	efectores
4	0.005643	0.019018	0.014884	0.015680	0.016220	-0.011417	efectores
6	0.045805	0.034776	0.001843	-0.030432	-0.032987	0.048108	efectores
..	…	…	…	…	…	…	
495	0.007867	0.000920	0.027213	0.006271	0.010150	0.025487	efectores
496	0.012852	0.014462	0.016175	0.012098	0.014030	0.003432	efectores
497	0.015193	0.000576	0.006853	0.002293	-0.007782	0.006205	efectores
498	-0.001925	-0.011077	0.024603	-0.001352	-0.011683	0.023092	efectores
499	-0.002094	0.035896	0.017444	0.019893	0.027209	-0.015606	efectores

[469 rows x 84 columns]

Composición de pseudo aminoácidos (PseAAC) hidro_mass efectores nematoda
dataset 3, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	469.000000	469.000000	469.000000	469.000000	469.000000	469.000000	
mean	0.034230	0.012302	0.026551	0.031414	0.024368	0.028279	
std	0.017782	0.011439	0.016067	0.019842	0.017166	0.015975	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.021391	0.004708	0.013455	0.017471	0.012751	0.016564	
50%	0.031208	0.009399	0.025372	0.028683	0.021424	0.025417	
75%	0.042554	0.016217	0.036415	0.042226	0.032149	0.037811	
max	0.116530	0.072990	0.097263	0.129212	0.129477	0.111866	

	X6	X7	X8	X9	…	X73	\
count	469.000000	469.000000	469.000000	469.000000	…	469.000000	
mean	0.012944	0.029948	0.030431	0.047402	…	0.011353	
std	0.010570	0.019375	0.019328	0.028979	…	0.021133	
min	0.000000	0.000000	0.000000	0.000000	…	-0.068752	
25%	0.005774	0.016563	0.017081	0.027293	…	-0.000019	
50%	0.010615	0.026560	0.027214	0.042867	…	0.009386	
75%	0.016891	0.038124	0.040709	0.061216	…	0.021616	
max	0.079268	0.142992	0.120586	0.220112	…	0.157556	

	X74	X75	X76	X77	X78	X79	\
count	469.000000	469.000000	469.000000	469.000000	469.000000	469.000000	
mean	0.002559	0.007003	0.008553	0.001276	0.005950	0.010748	
std	0.031206	0.023592	0.021254	0.029544	0.022227	0.020602	
min	-0.183074	-0.123452	-0.121497	-0.165190	-0.121887	-0.082430	
25%	-0.008598	-0.003033	-0.001897	-0.008452	-0.003904	0.000761	
50%	0.004600	0.007706	0.008946	0.003768	0.007024	0.009849	
75%	0.017409	0.020537	0.021221	0.015387	0.017833	0.022667	
max	0.129358	0.097635	0.080528	0.111304	0.083651	0.098468	

	X80	X81	X82
count	469.000000	469.000000	469.000000
mean	0.003985	0.008551	0.009358
std	0.029726	0.024179	0.020973
min	-0.142507	-0.070095	-0.120307
25%	-0.008655	-0.003518	-0.001555
50%	0.003914	0.006043	0.008803
75%	0.014534	0.019464	0.021606
max	0.250434	0.216015	0.078145

[8 rows x 83 columns]

no_efectores
Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores nematoda
dataset 3, sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.061971	0.027543	0.020657	0.020657	0.020657	0.006886	0.000000	
1	0.025673	0.005705	0.036133	0.039937	0.010460	0.020919	0.007607	
2	0.035252	0.004862	0.026743	0.036468	0.023096	0.031606	0.013372	
3	0.025967	0.008656	0.010387	0.008656	0.030295	0.017311	0.008656	
4	0.030051	0.008839	0.031819	0.045960	0.020329	0.020329	0.012374	
..	…	…	…	…	…	…	…	
495	0.069606	0.008701	0.030453	0.052205	0.018852	0.014501	0.014501	
496	0.015717	0.013261	0.008841	0.027504	0.007367	0.014243	0.003438	
497	0.042182	0.038666	0.056242	0.042182	0.031636	0.038666	0.017576	
498	0.019611	0.016809	0.039221	0.025214	0.030817	0.033618	0.011206	
499	0.024945	0.008315	0.011641	0.014967	0.023282	0.013304	0.006652	

	X7	X8	X9	…	X74	X75	X76	\
0	0.027543	0.041314	0.027543	…	0.062627	0.066320	-0.017575	
1	0.028526	0.042789	0.031379	…	-0.000996	0.029110	0.013344	
2	0.019450	0.031606	0.040115	…	-0.014937	-0.000080	0.011092	
3	0.038951	0.010387	0.041547	…	-0.000571	-0.001567	0.017045	
4	0.029167	0.049496	0.041541	…	-0.005633	0.006233	0.002548	
..	…	…	…	…	…	…	…	
495	0.046404	0.055105	0.058005	…	0.019002	0.034381	0.008386	
496	0.010805	0.023084	0.010314	…	-0.000506	0.017297	0.005288	
497	0.065030	0.028121	0.135332	…	-0.031165	-0.031659	0.006509	
498	0.044824	0.022412	0.050427	…	-0.019625	-0.005213	-0.012063	
499	0.026608	0.021619	0.053217	…	0.031658	0.022053	0.006913	

	X77	X78	X79	X80	X81	X82	X83
0	-0.012191	0.014652	0.007137	-0.054621	-0.013160	-0.024573	no_efectores
1	-0.011936	0.028765	0.008884	0.007259	0.020409	0.007852	no_efectores
2	-0.009026	-0.001697	0.005605	0.017371	0.044306	0.001026	no_efectores
3	0.035694	0.015288	0.008107	0.016652	0.009317	0.002080	no_efectores
4	-0.009640	0.021503	0.015348	-0.005744	0.012837	0.005744	no_efectores
..	…	…	…	…	…	…	…
495	0.012967	0.019396	0.028571	-0.007252	-0.001416	0.025443	no_efectores
496	0.011878	0.028791	0.017527	0.006630	0.017697	0.009892	no_efectores
497	-0.035886	-0.023151	0.017147	-0.044206	-0.051800	0.039261	no_efectores
498	0.014968	-0.001909	0.000544	0.034152	0.030578	0.034490	no_efectores
499	0.007720	0.011554	0.022396	-0.003920	-0.013582	-0.005735	no_efectores

[479 rows x 84 columns]

Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores nematoda
dataset 3, sin valores atípicos.
Estadísticas.

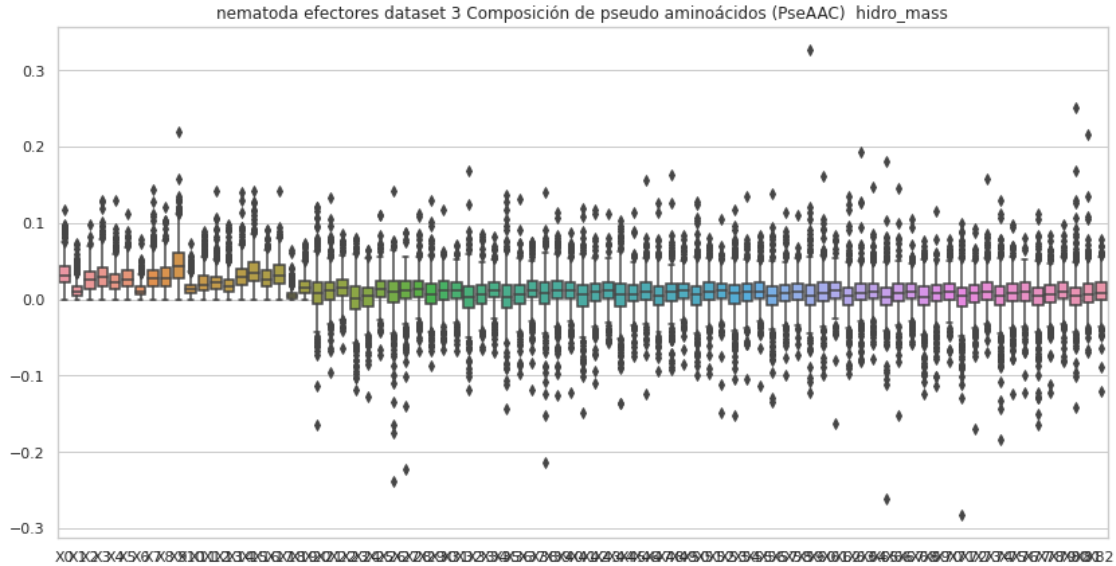
	X0	X1	X2	X3	X4	X5	\
count	479.000000	479.000000	479.000000	479.000000	479.000000	479.000000	
mean	0.029120	0.010985	0.025145	0.033273	0.025461	0.025912	
std	0.014525	0.010055	0.014287	0.018146	0.016912	0.014326	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.019038	0.004237	0.013837	0.019929	0.013692	0.015503	
50%	0.028257	0.008701	0.023650	0.031140	0.023054	0.024581	
75%	0.036974	0.014344	0.034082	0.043128	0.033331	0.033528	
max	0.084823	0.076900	0.076750	0.107660	0.107660	0.123040	

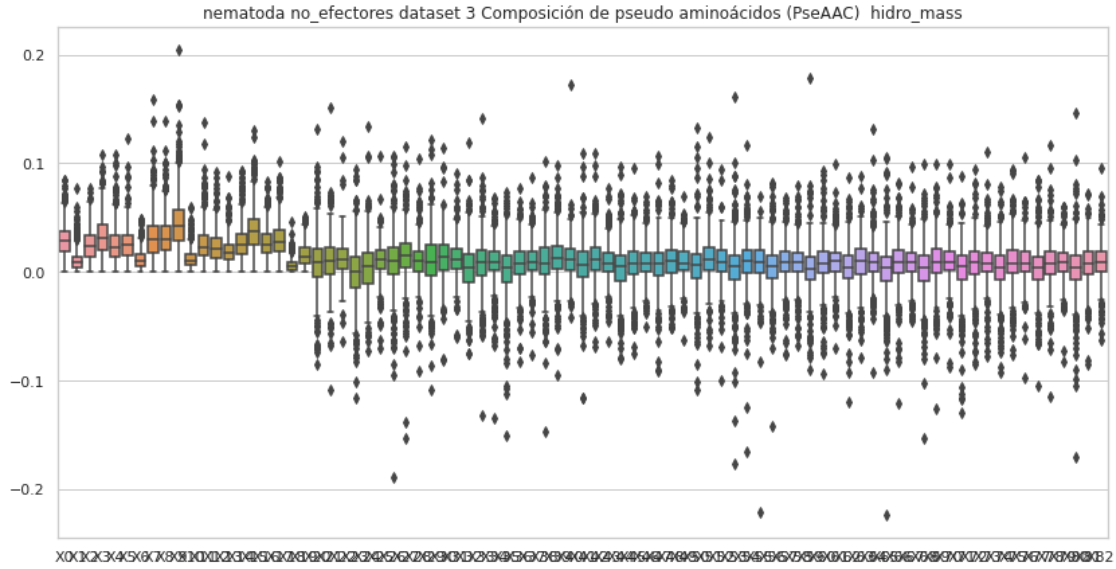
	X6	X7	X8	X9	…	X73	\
count	479.000000	479.000000	479.000000	479.000000	…	479.000000	
mean	0.011462	0.032510	0.033529	0.046402	…	0.008553	
std	0.008394	0.020214	0.019657	0.027208	…	0.018401	
min	0.000000	0.000000	0.000000	0.000000	…	-0.077979	
25%	0.005579	0.017873	0.020300	0.028437	…	0.000462	
50%	0.010144	0.030518	0.030449	0.042545	…	0.008016	
75%	0.015927	0.042892	0.042073	0.057436	…	0.017419	
max	0.048847	0.159163	0.138420	0.204335	…	0.110175	

	X74	X75	X76	X77	X78	X79	\
count	479.000000	479.000000	479.000000	479.000000	479.000000	479.000000	
mean	0.002587	0.009293	0.009519	0.002952	0.008200	0.009429	
std	0.022003	0.020457	0.017672	0.021833	0.021114	0.016499	
min	-0.092573	-0.076354	-0.097885	-0.105025	-0.115076	-0.083056	
25%	-0.007399	-0.001471	-0.000167	-0.007560	-0.001874	0.000154	
50%	0.003686	0.008348	0.008223	0.003541	0.007805	0.008761	
75%	0.014590	0.020466	0.018940	0.013323	0.019274	0.018055	
max	0.078385	0.094227	0.105836	0.085039	0.116015	0.060179	

	X80	X81	X82
count	479.000000	479.000000	479.000000
mean	0.002459	0.008054	0.009209
std	0.025576	0.019694	0.017227
min	-0.170566	-0.084513	-0.062947
25%	-0.006887	-0.002414	0.000270
50%	0.004462	0.008071	0.008436
75%	0.015212	0.019449	0.018450
max	0.145884	0.071951	0.095723

[8 rows x 83 columns]


nematoda efectores dataset 3 Composición de pseudo aminoácidos (PseAAC) hidro_mass
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3


nematoda no_efectores dataset 3 Composición de pseudo aminoácidos (PseAAC) hidro_mass
0.2
0.1
0.0
−0.1
−0.2

4 Composición de pseudo aminoácidos (PseAAC) mass

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

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

    if etiq == "efectores":
        df=PseAAC_mass_efec

    if etiq == "no_efectores":
        df=PseAAC_mass_no_efec

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

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

```
efectores
Composición de pseudo aminoácidos (PseAAC)  mass efectores nematoda dataset 3,
con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.067505  0.036003  0.058504  0.058504  0.027002  0.031502  0.027002
1    0.104985  0.031495  0.041994  0.062991  0.041994  0.020997  0.020997
2    0.064027  0.000000  0.021342  0.042685  0.021342  0.000000  0.000000
3    0.051200  0.017067  0.034134  0.068267  0.119468  0.034134  0.017067
4    0.038406  0.006401  0.044807  0.025604  0.012802  0.012802  0.006401
..        …         …         …         …         …         …         …
495  0.029647  0.003594  0.022460  0.028749  0.017070  0.050310  0.006289
```

```
496  0.036780  0.005015  0.040124  0.048483  0.013375  0.016718  0.016718
497  0.077319  0.014318  0.017182  0.022909  0.051546  0.034364  0.002864
498  0.050251  0.014357  0.037489  0.035893  0.016750  0.035096  0.022334
499  0.065953  0.009422  0.056531  0.084796  0.028265  0.051820  0.018844

          X7        X8        X9  …       X32       X33       X34  \
0    0.045003  0.081006  0.090007  …  0.020947  0.007597  0.020320
1    0.041994  0.073489  0.041994  … -0.010842 -0.074997  0.021876
2    0.042685  0.042685  0.085370  …  0.019953  0.081352  0.034221
3    0.051200  0.034134  0.102401  … -0.097785 -0.137502  0.008863
4    0.012802  0.044807  0.070411  … -0.034038 -0.005718  0.078032
..        …         …         …  …         …         …         …
495  0.016171  0.029647  0.026054  …  0.032995  0.025333  0.030230
496  0.020062  0.065201  0.048483  …  0.028380  0.023596  0.015893
497  0.045819  0.014318  0.083046  …  0.004456  0.007167  0.002740
498  0.025524  0.027119  0.067799  …  0.026802  0.031055  0.017780
499  0.051820  0.056531  0.113061  …  0.027086  0.019852  0.003929

          X35       X36       X37       X38       X39       X40         X41
0    0.032476 -0.008824  0.016551  0.018800 -0.012236  0.035548  efectores
1    0.035255 -0.022839  0.011010 -0.018049  0.014095  0.057675  efectores
2    0.003770 -0.067455 -0.090486  0.024430  0.063317  0.070232  efectores
3   -0.113023  0.034421 -0.020420 -0.044218 -0.098714  0.079490  efectores
4    0.066743 -0.040882 -0.054970  0.093141  0.036804 -0.028232  efectores
..        …         …         …         …         …         …         …
495  0.013312  0.025662  0.020048  0.023294  0.033981  0.031825  efectores
496  0.009017  0.011056  0.011813  0.018938  0.039910  0.008468  efectores
497  0.055300  0.032826  0.036349  0.010069  0.013737  0.012439  efectores
498  0.024007  0.032422  0.013844  0.021743  0.024587  0.023078  efectores
499  0.020987  0.007451 -0.048003 -0.042354  0.028955 -0.025904  efectores

[500 rows x 42 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass efectores nematoda dataset 3, con valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.046889    0.016417    0.036258    0.044001    0.034315    0.036522
std      0.020758    0.017071    0.020718    0.028761    0.025410    0.020441
min     -0.085005   -0.170009    0.000000   -0.170009   -0.170009   -0.255014
25%      0.035171    0.007226    0.022727    0.025611    0.019028    0.027191
50%      0.044512    0.013534    0.034131    0.041664    0.029886    0.035758
75%      0.056244    0.023030    0.046549    0.056560    0.045516    0.045165
max      0.155546    0.115622    0.162217    0.167659    0.198587    0.109705
```

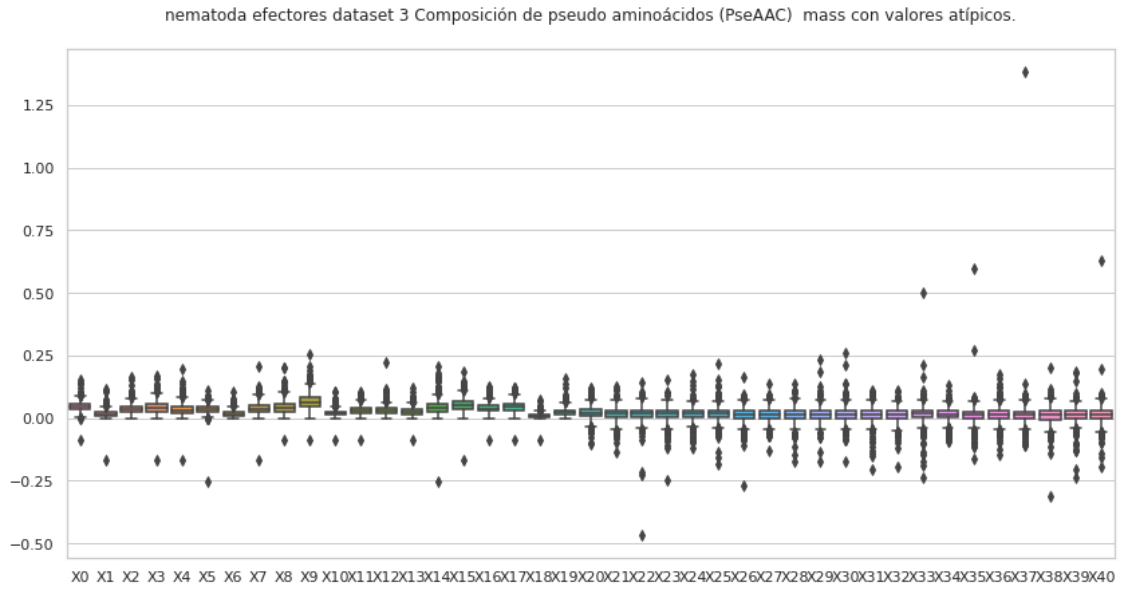
```
              X6          X7          X8          X9  …         X31  \
count  500.000000  500.000000  500.000000  500.000000  …  500.000000
mean     0.017772    0.041002    0.043223    0.065435  …    0.013163
std      0.012856    0.024718    0.028018    0.033970  …    0.031909
min      0.000000   -0.170009   -0.085005   -0.085005  …   -0.204085
25%      0.009102    0.025854    0.024504    0.044376  …    0.000209
50%      0.015628    0.038111    0.038485    0.060825  …    0.015072
75%      0.024092    0.052996    0.056339    0.082313  …    0.030041
max      0.105477    0.205476    0.200546    0.255238  …    0.112055

              X32         X33         X34         X35         X36         X37  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.011063    0.014569    0.012414    0.011448    0.013481    0.013836
std      0.029110    0.040855    0.027305    0.040667    0.031415    0.068326
min     -0.196488   -0.235444   -0.091046   -0.161173   -0.146336   -0.110642
25%     -0.003466    0.001585    0.000860   -0.000880   -0.001124   -0.002270
50%      0.014995    0.017446    0.015675    0.013467    0.015614    0.013538
75%      0.028404    0.030906    0.027752    0.027467    0.028210    0.027121
max      0.107836    0.498330    0.132626    0.596673    0.176573    1.382175

              X38         X39         X40
count  500.000000  500.000000  500.000000
mean     0.010698    0.011504    0.011597
std      0.033598    0.034287    0.042335
min     -0.314112   -0.237512   -0.193957
25%     -0.004901    0.000324   -0.004412
50%      0.012801    0.014638    0.012922
75%      0.027832    0.027753    0.028984
max      0.201780    0.184176    0.627360

[8 rows x 41 columns]
```

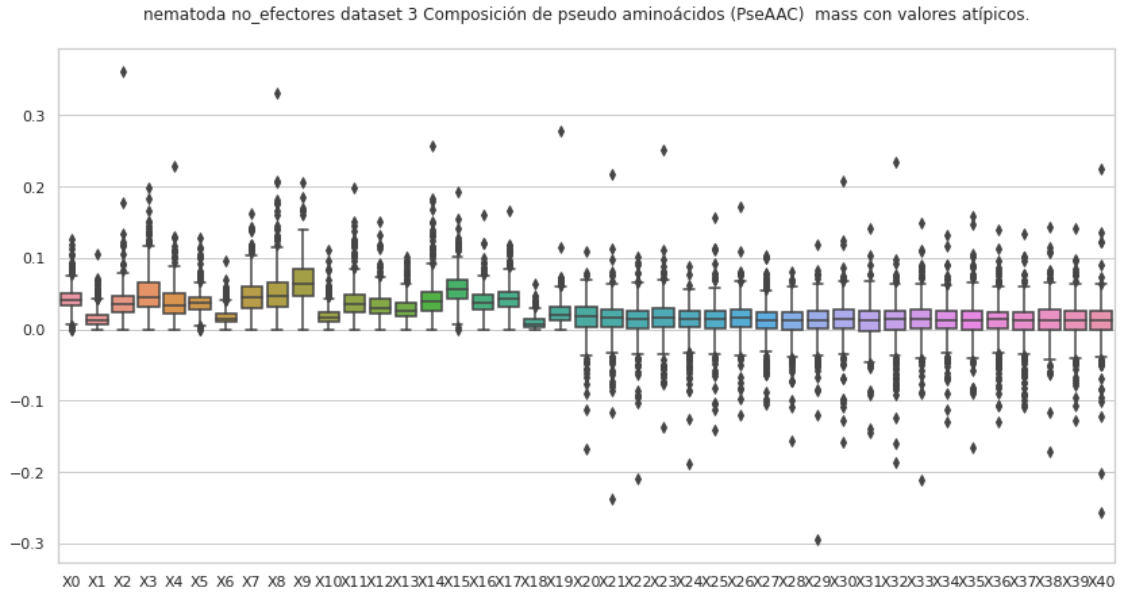
no_efectores
Composición de pseudo aminoácidos (PseAAC) mass no_efectores nematoda dataset 3, con valores atípicos.
Valores del documento csv.

```
          X0        X1        X2        X3        X4        X5        X6  \
0    0.126917  0.056407  0.042306  0.042306  0.042306  0.014102  0.000000
1    0.054236  0.012052  0.076332  0.084367  0.022096  0.044192  0.016070
2    0.050012  0.006898  0.037940  0.051737  0.032767  0.044839  0.018970
3    0.052189  0.017396  0.020876  0.017396  0.060887  0.034793  0.017396
4    0.043437  0.012776  0.045992  0.066433  0.029384  0.029384  0.017886
..        …         …         …         …         …         …         …
495  0.081805  0.010226  0.035790  0.061354  0.022155  0.017043  0.017043
496  0.038242  0.032266  0.021511  0.066923  0.017926  0.034656  0.008365
```

```
497  0.033894  0.031069  0.045192  0.033894  0.025420  0.031069  0.014122
498  0.026967  0.023114  0.053934  0.034672  0.042376  0.046229  0.015410
499  0.042662  0.014221  0.019909  0.025597  0.039818  0.022753  0.011376

           X7        X8        X9  …       X32       X33       X34  \
0    0.056407  0.084611  0.056407  … -0.020299  0.083577 -0.046302
1    0.060262  0.090394  0.066289  …  0.020083  0.005932 -0.008716
2    0.027593  0.044839  0.056911  …  0.018030 -0.009399 -0.003369
3    0.078283  0.020876  0.083502  …  0.036929 -0.004607  0.007353
4    0.042160  0.071543  0.060045  …  0.012087  0.002022  0.006735
..        …         …         …  …         …         …         …
495  0.054537  0.064762  0.068171  …  0.007923 -0.003907  0.032467
496  0.026291  0.056167  0.025096  …  0.013434  0.025588  0.039001
497  0.052253  0.022596  0.108743  … -0.003053 -0.008427  0.016080
498  0.061639  0.030819  0.069343  …  0.004002  0.016111  0.029623
499  0.045506  0.036974  0.091012  …  0.033901  0.008392  0.029503

          X35       X36       X37       X38       X39       X40           X41
0    0.085951 -0.018426  0.101324 -0.035994  0.014616 -0.050325  no_efectores
1    0.018439  0.030751 -0.001900  0.028189  0.018768  0.016588  no_efectores
2    0.003707  0.002654  0.008913  0.015737  0.007951  0.001455  no_efectores
3    0.020635  0.006097  0.019653  0.034257  0.016294  0.004180  no_efectores
4    0.017913  0.004489  0.004747  0.003683  0.022184  0.008302  no_efectores
..        …         …         …         …         …         …           …
495  0.023797  0.001323  0.005770  0.009856  0.033578  0.029902  no_efectores
496  0.019524  0.012043  0.023243  0.012866  0.042646  0.024069  no_efectores
497  0.009719  0.014500  0.000304  0.005230  0.013778  0.031547  no_efectores
498 -0.007499 -0.042395  0.028881 -0.016588  0.000748  0.047428  no_efectores
499  0.018599  0.007987  0.025699  0.011822  0.038302 -0.009808  no_efectores

[500 rows x 42 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.042016    0.015478    0.037469    0.050671    0.037504    0.037331
std      0.017169    0.013338    0.024627    0.028765    0.023678    0.016306
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.032639    0.006819    0.023931    0.030992    0.021069    0.028526
50%      0.040683    0.012515    0.035244    0.045136    0.033992    0.036245
75%      0.050380    0.021043    0.047359    0.066340    0.050078    0.044074
max      0.126917    0.105113    0.361219    0.198404    0.228594    0.127575

                X6          X7          X8          X9  …         X31  \
```

```
count  500.000000  500.000000  500.000000  500.000000  …  500.000000
mean     0.016844    0.047054    0.052459    0.066792  …    0.010983
std      0.011378    0.025346    0.033354    0.030419  …    0.025863
min      0.000000    0.000000    0.000000    0.000000  …   -0.144284
25%      0.009698    0.029135    0.030860    0.046664  …   -0.002625
50%      0.015114    0.044440    0.046769    0.063211  …    0.012774
75%      0.022088    0.059601    0.065434    0.084900  …    0.026733
max      0.096411    0.163022    0.331095    0.206264  …    0.141489

              X32         X33         X34         X35         X36         X37  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.010503    0.013449    0.011003    0.011469    0.011719    0.010572
std      0.030222    0.027577    0.025823    0.026883    0.026798    0.027189
min     -0.185954   -0.212065   -0.129516   -0.165789   -0.129807   -0.109203
25%     -0.000309    0.000217    0.000392   -0.001179    0.000979   -0.000128
50%      0.013651    0.014674    0.012706    0.012603    0.013854    0.013357
75%      0.025563    0.027261    0.025294    0.025817    0.024846    0.024382
max      0.233705    0.148430    0.132073    0.157568    0.138868    0.133929

              X38         X39         X40
count  500.000000  500.000000  500.000000
mean     0.012899    0.011941    0.010868
std      0.027028    0.025887    0.030950
min     -0.170891   -0.127773   -0.255780
25%     -0.000463   -0.001062   -0.000471
50%      0.012846    0.013207    0.011996
75%      0.027636    0.025292    0.025315
max      0.142407    0.142258    0.224199

[8 rows x 41 columns]
```





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

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

```
df=""

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

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

    if etiq == "efectores":
        df=PseAAC_mass_efec

    if etiq == "no_efectores":
        df=PseAAC_mass_no_efec

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

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

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

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

```
Composición de pseudo aminoácidos (PseAAC)  mass efectores nematoda dataset 3,
sin valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.067505  0.036003  0.058504  0.058504  0.027002  0.031502  0.027002
4    0.038406  0.006401  0.044807  0.025604  0.012802  0.012802  0.006401
6    0.042308  0.008462  0.012692  0.008462  0.016923  0.029616  0.000000
9    0.041647  0.011713  0.052059  0.059868  0.041647  0.031235  0.020824
```

```
10   0.046494  0.030996  0.023247  0.054243  0.015498  0.046494  0.030996
..        …         …         …         …         …         …
495  0.029647  0.003594  0.022460  0.028749  0.017070  0.050310  0.006289
496  0.036780  0.005015  0.040124  0.048483  0.013375  0.016718  0.016718
497  0.077319  0.014318  0.017182  0.022909  0.051546  0.034364  0.002864
498  0.050251  0.014357  0.037489  0.035893  0.016750  0.035096  0.022334
499  0.065953  0.009422  0.056531  0.084796  0.028265  0.051820  0.018844

           X7        X8        X9  …       X32       X33       X34  \
0    0.045003  0.081006  0.090007  …  0.020947  0.007597  0.020320
4    0.012802  0.044807  0.070411  … -0.034038 -0.005718  0.078032
6    0.029616  0.012692  0.008462  …  0.015178  0.025320  0.037479
9    0.045552  0.055963  0.067677  …  0.024101  0.011145 -0.007021
10   0.023247  0.069741  0.054243  …  0.009008  0.010477 -0.034241
..        …         …         …  …         …         …         …
495  0.016171  0.029647  0.026054  …  0.032995  0.025333  0.030230
496  0.020062  0.065201  0.048483  …  0.028380  0.023596  0.015893
497  0.045819  0.014318  0.083046  …  0.004456  0.007167  0.002740
498  0.025524  0.027119  0.067799  …  0.026802  0.031055  0.017780
499  0.051820  0.056531  0.113061  …  0.027086  0.019852  0.003929

           X35       X36       X37       X38       X39       X40         X41
0    0.032476 -0.008824  0.016551  0.018800 -0.012236  0.035548  efectores
4    0.066743 -0.040882 -0.054970  0.093141  0.036804 -0.028232  efectores
6    0.039796  0.031826  0.011495  0.010188  0.001388  0.036231  efectores
9   -0.006411  0.023441 -0.018522 -0.009457 -0.007564  0.033872  efectores
10   0.032674  0.003535  0.096727 -0.009247 -0.046054 -0.027300  efectores
..        …         …         …         …         …         …         …
495  0.013312  0.025662  0.020048  0.023294  0.033981  0.031825  efectores
496  0.009017  0.011056  0.011813  0.018938  0.039910  0.008468  efectores
497  0.055300  0.032826  0.036349  0.010069  0.013737  0.012439  efectores
498  0.024007  0.032422  0.013844  0.021743  0.024587  0.023078  efectores
499  0.020987  0.007451 -0.048003 -0.042354  0.028955 -0.025904  efectores

[407 rows x 42 columns]


Composición de pseudo aminoácidos (PseAAC)  mass efectores nematoda dataset 3,
sin valores atípicos.
Estadísticas.

                X0          X1          X2          X3          X4          X5  \
count  407.000000  407.000000  407.000000  407.000000  407.000000  407.000000
mean     0.044645    0.014695    0.034715    0.042208    0.030730    0.036828
std      0.015276    0.010928    0.016478    0.023890    0.016980    0.013249
min      0.004896    0.000000    0.000000    0.000000    0.000000    0.005280
25%      0.035167    0.006844    0.023396    0.025745    0.018552    0.027921
50%      0.043280    0.012657    0.033704    0.040592    0.028145    0.035663
```

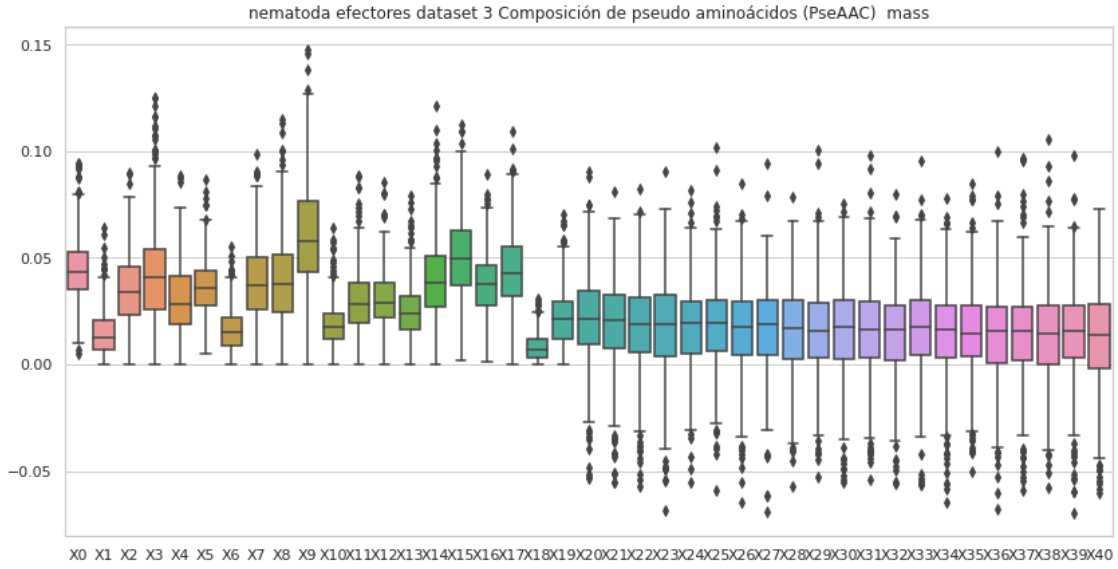
```
75%      0.053083    0.020621    0.045718    0.054023    0.041369    0.043902
max      0.094509    0.063942    0.089645    0.124813    0.088905    0.086401

                X6          X7          X8          X9  …         X31  \
count   407.000000  407.000000  407.000000  407.000000  …  407.000000
mean      0.016186    0.038642    0.039784    0.060728  …    0.016073
std       0.009755    0.018308    0.021248    0.026347  …    0.022044
min       0.000000    0.000000    0.000000    0.000000  …   -0.054133
25%       0.008980    0.025597    0.024297    0.043334  …    0.002861
50%       0.014855    0.037187    0.037722    0.058029  …    0.016340
75%       0.021853    0.050463    0.051751    0.076791  …    0.029383
max       0.055166    0.098632    0.115201    0.147622  …    0.097950

                X32         X33         X34         X35         X36         X37  \
count   407.000000  407.000000  407.000000  407.000000  407.000000  407.000000
mean      0.013599    0.016248    0.014508    0.014268    0.013786    0.015208
std       0.021328    0.021268    0.021905    0.020700    0.021584    0.022605
min      -0.056241   -0.056715   -0.064498   -0.050241   -0.068056   -0.058873
25%       0.001614    0.004451    0.002945    0.003729    0.000435    0.001914
50%       0.016079    0.017852    0.016589    0.014700    0.015765    0.015384
75%       0.027604    0.030319    0.027432    0.027523    0.027294    0.027183
max       0.079879    0.095447    0.078032    0.084533    0.100160    0.096727

                X38         X39         X40
count   407.000000  407.000000  407.000000
mean      0.013559    0.014606    0.012246
std       0.022776    0.022344    0.023026
min      -0.058063   -0.069783   -0.060632
25%       0.000136    0.003220   -0.001800
50%       0.014304    0.015784    0.013858
75%       0.027601    0.027770    0.028041
max       0.105319    0.097819    0.072713

[8 rows x 41 columns]
```

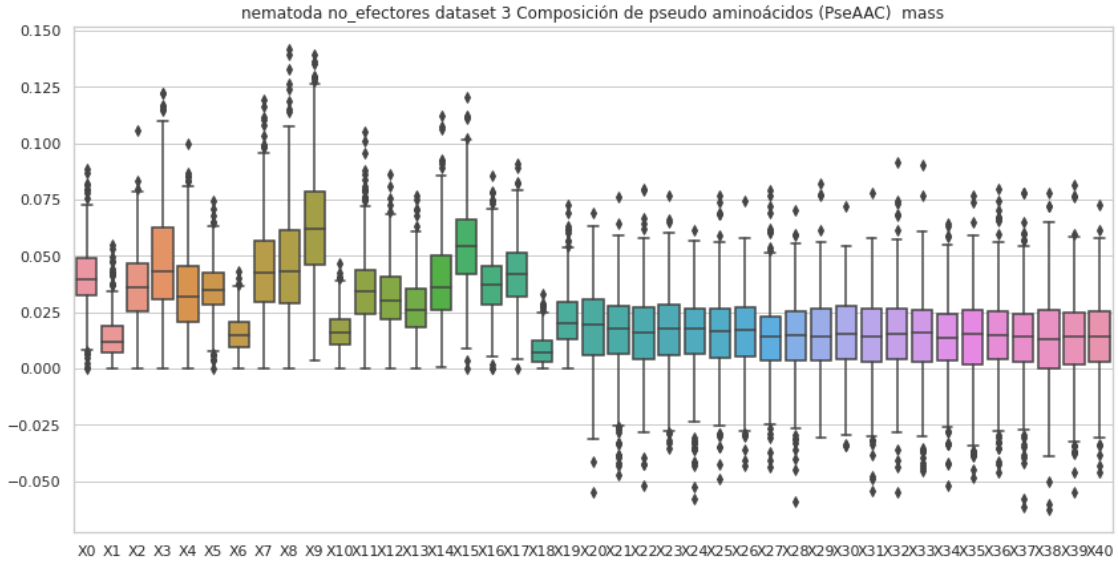
Composición de pseudo aminoácidos (PseAAC) mass no_efectores nematoda dataset 3, sin valores atípicos.
Valores del documento csv.

```
          X0        X1        X2        X3        X4        X5        X6  \
1   0.054236  0.012052  0.076332  0.084367  0.022096  0.044192  0.016070
2   0.050012  0.006898  0.037940  0.051737  0.032767  0.044839  0.018970
3   0.052189  0.017396  0.020876  0.017396  0.060887  0.034793  0.017396
4   0.043437  0.012776  0.045992  0.066433  0.029384  0.029384  0.017886
5   0.040570  0.025356  0.060855  0.045641  0.070998  0.030428  0.025356
..       …         …         …         …         …         …         …
```

```
495  0.081805  0.010226  0.035790  0.061354  0.022155  0.017043  0.017043
496  0.038242  0.032266  0.021511  0.066923  0.017926  0.034656  0.008365
497  0.033894  0.031069  0.045192  0.033894  0.025420  0.031069  0.014122
498  0.026967  0.023114  0.053934  0.034672  0.042376  0.046229  0.015410
499  0.042662  0.014221  0.019909  0.025597  0.039818  0.022753  0.011376

           X7        X8        X9  …       X32       X33       X34  \
1    0.060262  0.090394  0.066289  …  0.020083  0.005932 -0.008716
2    0.027593  0.044839  0.056911  …  0.018030 -0.009399 -0.003369
3    0.078283  0.020876  0.083502  …  0.036929 -0.004607  0.007353
4    0.042160  0.071543  0.060045  …  0.012087  0.002022  0.006735
5    0.086211  0.065926  0.111568  …  0.009213 -0.006795  0.002367
..        …         …         …  …         …         …
495  0.054537  0.064762  0.068171  …  0.007923 -0.003907  0.032467
496  0.026291  0.056167  0.025096  …  0.013434  0.025588  0.039001
497  0.052253  0.022596  0.108743  … -0.003053 -0.008427  0.016080
498  0.061639  0.030819  0.069343  …  0.004002  0.016111  0.029623
499  0.045506  0.036974  0.091012  …  0.033901  0.008392  0.029503

           X35       X36       X37       X38       X39       X40           X41
1    0.018439  0.030751 -0.001900  0.028189  0.018768  0.016588  no_efectores
2    0.003707  0.002654  0.008913  0.015737  0.007951  0.001455  no_efectores
3    0.020635  0.006097  0.019653  0.034257  0.016294  0.004180  no_efectores
4    0.017913  0.004489  0.004747  0.003683  0.022184  0.008302  no_efectores
5   -0.027653  0.054465  0.034048 -0.012052 -0.007379  0.014136  no_efectores
..        …         …         …         …         …         …           …
495  0.023797  0.001323  0.005770  0.009856  0.033578  0.029902  no_efectores
496  0.019524  0.012043  0.023243  0.012866  0.042646  0.024069  no_efectores
497  0.009719  0.014500  0.000304  0.005230  0.013778  0.031547  no_efectores
498 -0.007499 -0.042395  0.028881 -0.016588  0.000748  0.047428  no_efectores
499  0.018599  0.007987  0.025699  0.011822  0.038302 -0.009808  no_efectores

[404 rows x 42 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  404.000000  404.000000  404.000000  404.000000  404.000000  404.000000
mean     0.040314    0.013975    0.036349    0.047388    0.034148    0.035230
std      0.014276    0.010110    0.016408    0.023599    0.018480    0.011814
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.032639    0.007052    0.025579    0.030605    0.020730    0.028496
50%      0.039812    0.012034    0.035794    0.043071    0.031740    0.034778
75%      0.048810    0.018582    0.046847    0.062688    0.045200    0.042594
max      0.088514    0.054655    0.105430    0.122341    0.099809    0.074550
```

```
              X6          X7          X8          X9  …         X31  \
count  404.000000  404.000000  404.000000  404.000000  …  404.000000
mean     0.015299    0.044991    0.047133    0.063875  …    0.013104
std      0.008274    0.021940    0.024780    0.027006  …    0.019351
min      0.000000    0.000000    0.000000    0.003702  …   -0.054480
25%      0.009681    0.029270    0.029016    0.046123  …    0.003112
50%      0.014860    0.042447    0.042855    0.061857  …    0.014229
75%      0.020542    0.056749    0.061429    0.078583  …    0.026529
max      0.043362    0.119275    0.141705    0.139595  …    0.077662

              X32         X33         X34         X35         X36         X37  \
count  404.000000  404.000000  404.000000  404.000000  404.000000  404.000000
mean     0.015108    0.014017    0.012735    0.013208    0.014852    0.013287
std      0.018451    0.019372    0.017639    0.018476    0.018751    0.019254
min     -0.054850   -0.045350   -0.051849   -0.048372   -0.045922   -0.061652
25%      0.004363    0.003136    0.003391    0.001709    0.004029    0.002970
50%      0.015551    0.015664    0.013515    0.015478    0.014763    0.014356
75%      0.026249    0.026157    0.024230    0.025866    0.025317    0.024361
max      0.091282    0.090468    0.064361    0.077008    0.079663    0.077833

              X38         X39         X40
count  404.000000  404.000000  404.000000
mean     0.012236    0.013579    0.013475
std      0.020454    0.019584    0.017548
min     -0.062683   -0.055095   -0.046301
25%     -0.000054    0.001952    0.002642
50%      0.012846    0.013925    0.014022
75%      0.025964    0.024836    0.025315
max      0.077744    0.081595    0.072635

[8 rows x 41 columns]
```





5 Composición de pseudo aminoácidos (PseAAC) hidro

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

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

    if etiq == "efectores":
        df=PseAAC_hidro_efec

    if etiq == "no_efectores":
        df=PseAAC_hidro_no_efec

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

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

```
efectores
Composición de pseudo aminoácidos (PseAAC)  hidro efectores nematoda dataset 3,
con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.055978  0.029855  0.048514  0.048514  0.022391  0.026123  0.022391
1    0.029570  0.008871  0.011828  0.017742  0.011828  0.005914  0.005914
2    1.034003  0.000000  0.344668  0.689335  0.344668  0.000000  0.000000
3    0.015582  0.005194  0.010388  0.020776  0.036358  0.010388  0.005194
4    0.018123  0.003020  0.021143  0.012082  0.006041  0.006041  0.003020
..        …         …         …         …         …         …         …
495  0.044246  0.005363  0.033520  0.042906  0.025475  0.075085  0.009386
496  0.017703  0.002414  0.019313  0.023336  0.006438  0.008047  0.008047
497  0.046888  0.008683  0.010420  0.013893  0.031259  0.020839  0.001737
498  0.087598  0.025028  0.065351  0.062570  0.029199  0.061180  0.038933
499  0.041780  0.005969  0.035812  0.053718  0.017906  0.032827  0.011937

           X7        X8        X9  …       X53       X54       X55  \
0    0.037318  0.067173  0.074637  …  0.048459 -0.000277 -0.027318
1    0.011828  0.020699  0.011828  …  0.027624  0.056985  0.031360
```

```
2    0.689335  0.689335  1.378671  …  0.851209 -3.164018 -2.302631
3    0.015582  0.010388  0.031164  … -0.016302 -0.045539  0.013393
4    0.006041  0.021143  0.033225  … -0.004579 -0.011408  0.018592
..        …         …         …  …         …         …
495  0.024134  0.044246  0.038883  …  0.011441  0.015151  0.021451
496  0.009656  0.031383  0.023336  …  0.015727  0.009994  0.022773
497  0.027786  0.008683  0.050361  … -0.005450 -0.029997 -0.023161
498  0.044494  0.047275  0.118188  …  0.000862  0.009738  0.035002
499  0.032827  0.035812  0.071623  …  0.051547  0.013507  0.018238

          X56       X57       X58       X59       X60       X61         X62
0    0.021007  0.009337 -0.055284  0.003561  0.024735  0.060255  efectores
1    0.048379  0.028757  0.016641  0.018991 -0.014191 -0.009741  efectores
2    3.510292  2.781337 -2.834788 -1.617181 -4.401623 -2.616463  efectores
3    0.017579  0.038021  0.000778  0.008576  0.003682  0.001934  efectores
4   -0.024115 -0.017628  0.006584  0.022190  0.018295  0.018925  efectores
..        …         …         …         …         …         …         …
495 -0.013705  0.003927  0.014660  0.001714  0.011687  0.018916  efectores
496  0.023692  0.030836  0.015264  0.017175  0.014368  0.016662  efectores
497 -0.005141 -0.009045  0.018470  0.000700  0.002788 -0.009461  efectores
498 -0.048161 -0.019486 -0.003353 -0.019298 -0.002355 -0.020354  efectores
499  0.008139  0.029081 -0.002202  0.037746  0.020919  0.028611  efectores

[500 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro efectores nematoda dataset 3, con valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.076465    0.017905    0.038853    0.053295    0.056816    0.052244
std      0.481385    0.022976    0.029570    0.194373    0.479174    0.194042
min     -0.117113   -0.234227    0.000000   -0.234227   -0.234227   -0.351340
25%      0.027960    0.006197    0.019139    0.025540    0.018642    0.020749
50%      0.044466    0.013209    0.033850    0.040496    0.029850    0.035515
75%      0.064697    0.022539    0.050031    0.057949    0.045089    0.059148
max     10.725089    0.137482    0.344668    4.290036   10.725089    4.290036

                X6          X7          X8          X9  …         X52  \
count  500.000000  500.000000  500.000000  500.000000  …  500.000000
mean     0.023242    0.057904    0.053197    0.092798  …    0.009144
std      0.096605    0.289650    0.194663    0.482678  …    0.244610
min      0.000000   -0.234227   -0.117113   -0.117113  …   -1.023363
25%      0.007627    0.024086    0.023183    0.037954  …   -0.014803
50%      0.014838    0.036775    0.038127    0.059800  …    0.004468
75%      0.025271    0.052509    0.057121    0.090092  …    0.022566
```

```
max      2.145018    6.435053    4.290036   10.725089  …    5.240764

              X53         X54         X55         X56         X57         X58  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.015102   -0.023582   -0.020596    0.003605   -0.002784   -0.012372
std      0.138691    0.363320    0.450119    0.192714    0.353462    0.277910
min     -0.188512   -7.369123   -9.726807   -2.157024   -7.265702   -5.303105
25%     -0.005365   -0.017826   -0.009985   -0.015244   -0.005309   -0.012622
50%      0.009743    0.005887    0.010095    0.006055    0.010129    0.005437
75%      0.024242    0.020143    0.024412    0.023400    0.026317    0.021380
max      2.829359    0.208974    0.238095    3.510292    2.781337    0.771557

              X59         X60         X61
count  500.000000  500.000000  500.000000
mean     0.003454   -0.038149   -0.006595
std      0.088130    0.758666    0.313579
min     -1.617181  -16.303808   -6.185048
25%     -0.007636   -0.014161   -0.007349
50%      0.009543    0.004870    0.008249
75%      0.023925    0.019119    0.024403
max      0.275106    1.058852    1.620887

[8 rows x 62 columns]
```

no_efectores
Composición de pseudo aminoácidos (PseAAC) hidro no_efectores nematoda dataset 3, con valores atípicos.
Valores del documento csv.

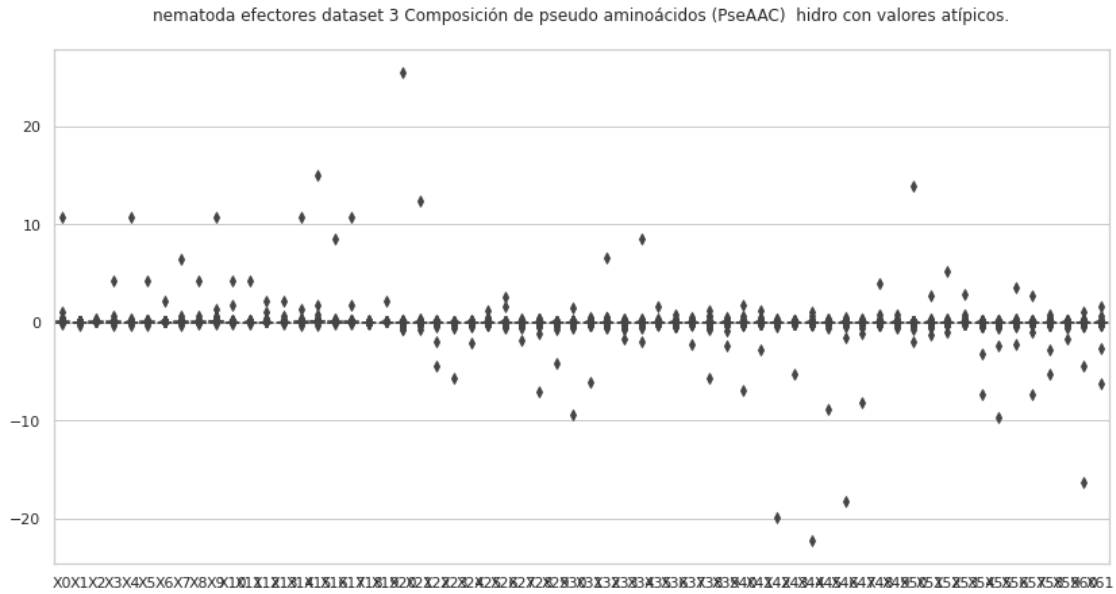
```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.062363  0.027717  0.020788  0.020788  0.020788  0.006929  0.000000
1    0.028975  0.006439  0.040779  0.045072  0.011805  0.023609  0.008585
2    0.046268  0.006382  0.035100  0.047864  0.030314  0.041482  0.017550
3    0.030691  0.010230  0.012277  0.010230  0.035807  0.020461  0.010230
4    0.036572  0.010757  0.038724  0.055934  0.024740  0.024740  0.015059
..        …         …         …         …         …         …         …
495  0.093964  0.011746  0.041109  0.070473  0.025449  0.019576  0.019576
496  0.019205  0.016204  0.010803  0.033608  0.009002  0.017404  0.004201
497  0.064537  0.059159  0.086049  0.064537  0.048403  0.059159  0.026890
498  0.023619  0.020245  0.047238  0.030367  0.037116  0.040490  0.013497
499  0.030693  0.010231  0.014323  0.018416  0.028647  0.016370  0.008185

           X7        X8        X9  …       X53       X54       X55  \
0    0.027717  0.041575  0.027717  …  0.022755 -0.027772 -0.040186
1    0.032194  0.048291  0.035414  …  0.026001 -0.003569  0.036975
2    0.025527  0.041482  0.052650  …  0.004726  0.022228  0.020023
```

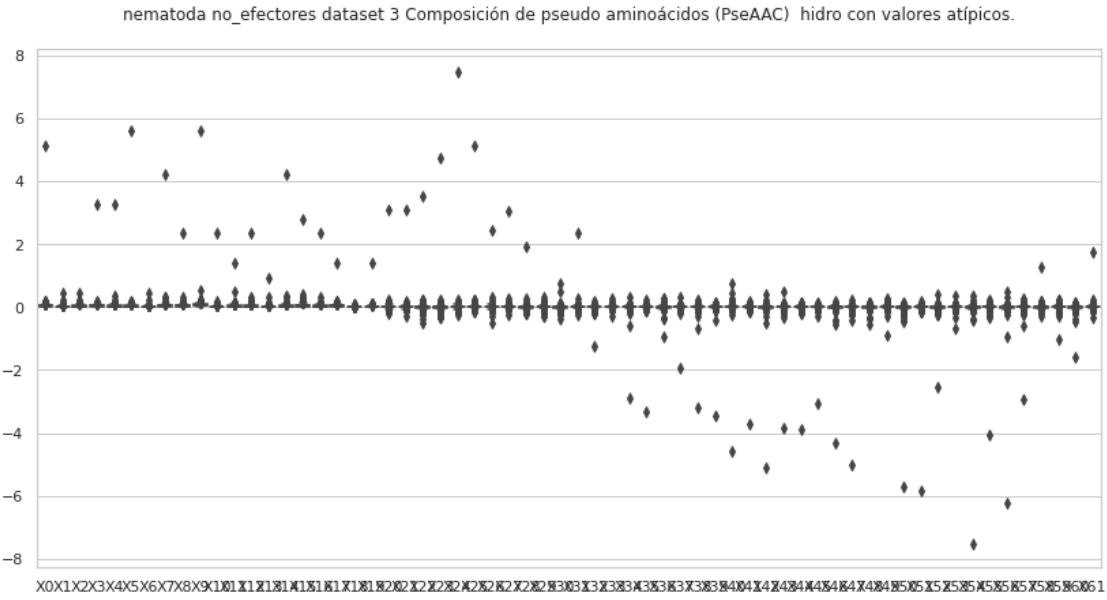
```
3    0.046037  0.012277  0.049106  … -0.004789 -0.007523  0.008218
4    0.035497  0.060237  0.050556  …  0.031751 -0.014461  0.014197
..        …         …         …  …         …         …
495  0.062643  0.074388  0.078303  …  0.045966 -0.049441 -0.002661
496  0.013203  0.028207  0.012603  …  0.023483  0.013542  0.024293
497  0.099495  0.043025  0.207056  …  0.000413  0.036280  0.002561
498  0.053986  0.026993  0.060735  … -0.021884  0.020317  0.005298
499  0.032739  0.026600  0.065478  …  0.017829  0.033015  0.023070

          X56       X57       X58       X59       X60       X61           X62
0    0.063023  0.066739 -0.012268  0.014745 -0.054966 -0.013244  no_efectores
1   -0.001124  0.032853 -0.013471  0.032464  0.008192  0.023034  no_efectores
2   -0.019605 -0.000105 -0.011847 -0.002227  0.022799  0.058150  no_efectores
3   -0.000675 -0.001852  0.042188  0.018069  0.019682  0.011012  no_efectores
4   -0.006856  0.007585 -0.011732  0.026169 -0.006991  0.015623  no_efectores
..        …         …         …         …         …         …           …
495  0.025651  0.046412  0.017504  0.026184 -0.009789 -0.001911  no_efectores
496 -0.000618  0.021135  0.014514  0.035181  0.008102  0.021625  no_efectores
497 -0.047682 -0.048437 -0.054905 -0.035420 -0.067635 -0.079254  no_efectores
498 -0.023636 -0.006279  0.018027 -0.002299  0.041133  0.036828  no_efectores
499  0.038952  0.027134  0.009499  0.014217 -0.004823 -0.016712  no_efectores

[500 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro no_efectores nematoda dataset 3, con valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.051268    0.016348    0.035238    0.050270    0.041526    0.048356
std      0.229809    0.027258    0.030066    0.146368    0.147588    0.250519
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.022376    0.005535    0.018084    0.027236    0.019748    0.019349
50%      0.035581    0.011821    0.032561    0.042454    0.030503    0.031999
75%      0.050996    0.019931    0.046870    0.056666    0.044637    0.046664
max      5.141557    0.467414    0.467414    3.271900    3.271900    5.608971

                X6          X7          X8          X9  …         X52  \
count  500.000000  500.000000  500.000000  500.000000  …  500.000000
mean     0.016959    0.052206    0.049546    0.074212  …   -0.002838
std      0.025468    0.188894    0.106613    0.251847  …    0.120328
min      0.000000    0.000000    0.000000    0.000000  …   -2.538232
25%      0.007081    0.025003    0.026770    0.039036  …   -0.011879
50%      0.012784    0.039812    0.041014    0.057708  …    0.005974
75%      0.020981    0.055720    0.056522    0.078214  …    0.018767
max      0.467414    4.206729    2.337071    5.608971  …    0.412736
```

```
              X53         X54         X55         X56         X57         X58  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.007993   -0.012604    0.000743   -0.009710    0.003971    0.004053
std      0.048877    0.339661    0.185569    0.283626    0.139823    0.066584
min     -0.690517   -7.525995   -4.052485   -6.206825   -2.954282   -0.280320
25%     -0.002213   -0.011181   -0.004450   -0.011223   -0.002171   -0.011758
50%      0.011098    0.006667    0.011372    0.005608    0.010866    0.004554
75%      0.025442    0.019514    0.023762    0.018448    0.025150    0.016302
max      0.341654    0.364084    0.239491    0.469021    0.278054    1.264647

              X59         X60         X61
count  500.000000  500.000000  500.000000
mean     0.007217   -0.002351    0.012114
std      0.059143    0.084595    0.085769
min     -1.011706   -1.606068   -0.337749
25%     -0.003443   -0.010664   -0.004153
50%      0.010307    0.005188    0.010037
75%      0.025588    0.019094    0.024282
max      0.244080    0.144160    1.765819

[8 rows x 62 columns]
```





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

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

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


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

    if etiq == "efectores":
        df=PseAAC_hidro_efec

    if etiq == "no_efectores":
        df=PseAAC_hidro_no_efec

    del df['X62']
```

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

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


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

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

```
efectores
Composición de pseudo aminoácidos (PseAAC)  efectores nematoda dataset 3, sin
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.055978  0.029855  0.048514  0.048514  0.022391  0.026123  0.022391
1    0.029570  0.008871  0.011828  0.017742  0.011828  0.005914  0.005914
3    0.015582  0.005194  0.010388  0.020776  0.036358  0.010388  0.005194
4    0.018123  0.003020  0.021143  0.012082  0.006041  0.006041  0.003020
5    0.043883  0.065825  0.087766  0.109708  0.065825  0.109708  0.065825
..        …         …         …         …         …         …         …
495  0.044246  0.005363  0.033520  0.042906  0.025475  0.075085  0.009386
496  0.017703  0.002414  0.019313  0.023336  0.006438  0.008047  0.008047
497  0.046888  0.008683  0.010420  0.013893  0.031259  0.020839  0.001737
498  0.087598  0.025028  0.065351  0.062570  0.029199  0.061180  0.038933
499  0.041780  0.005969  0.035812  0.053718  0.017906  0.032827  0.011937

           X7        X8        X9  …       X53       X54       X55  \
0    0.037318  0.067173  0.074637  …  0.048459 -0.000277 -0.027318
1    0.011828  0.020699  0.011828  …  0.027624  0.056985  0.031360
3    0.015582  0.010388  0.031164  … -0.016302 -0.045539  0.013393
4    0.006041  0.021143  0.033225  … -0.004579 -0.011408  0.018592
5    0.065825  0.021942  0.131649  …  0.098492  0.208974 -0.035171
```

```
..        …       …       …  …       …       …       …
495  0.024134  0.044246  0.038883  …  0.011441  0.015151  0.021451
496  0.009656  0.031383  0.023336  …  0.015727  0.009994  0.022773
497  0.027786  0.008683  0.050361  … -0.005450 -0.029997 -0.023161
498  0.044494  0.047275  0.118188  …  0.000862  0.009738  0.035002
499  0.032827  0.035812  0.071623  …  0.051547  0.013507  0.018238

          X56       X57       X58       X59       X60       X61        X62
0    0.021007  0.009337 -0.055284  0.003561  0.024735  0.060255  efectores
1    0.048379  0.028757  0.016641  0.018991 -0.014191 -0.009741  efectores
3    0.017579  0.038021  0.000778  0.008576  0.003682  0.001934  efectores
4   -0.024115 -0.017628  0.006584  0.022190  0.018295  0.018925  efectores
5   -0.027953 -0.081525  0.088316  0.130404 -0.149197 -0.168224  efectores
..        …       …       …       …       …       …       …
495 -0.013705  0.003927  0.014660  0.001714  0.011687  0.018916  efectores
496  0.023692  0.030836  0.015264  0.017175  0.014368  0.016662  efectores
497 -0.005141 -0.009045  0.018470  0.000700  0.002788 -0.009461  efectores
498 -0.048161 -0.019486 -0.003353 -0.019298 -0.002355 -0.020354  efectores
499  0.008139  0.029081 -0.002202  0.037746  0.020919  0.028611  efectores

[473 rows x 63 columns]


Composición de pseudo aminoácidos (PseAAC)  efectores nematoda dataset 3, sin
valores atípicos.
Estadísticas.

                X0          X1          X2          X3          X4          X5  \
count  473.000000  473.000000  473.000000  473.000000  473.000000  473.000000
mean     0.049298    0.016153    0.036556    0.042431    0.033110    0.042204
std      0.030525    0.014527    0.021972    0.024901    0.022305    0.030938
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.027317    0.006209    0.019151    0.025463    0.018449    0.020437
50%      0.043365    0.012706    0.033541    0.040026    0.029097    0.034623
75%      0.061856    0.021393    0.048953    0.056525    0.044192    0.056657
max      0.201605    0.084885    0.109161    0.195194    0.167309    0.212486

                X6          X7          X8          X9  …         X52  \
count  473.000000  473.000000  473.000000  473.000000  …  473.000000
mean     0.018249    0.041402    0.041101    0.065063  …    0.001476
std      0.015046    0.026566    0.025136    0.038279  …    0.044271
min      0.000000    0.000000    0.000000    0.001661  …   -0.290300
25%      0.007864    0.024128    0.022934    0.037353  …   -0.013138
50%      0.014707    0.036026    0.037014    0.058827  …    0.004629
75%      0.024100    0.050880    0.054040    0.084773  …    0.022270
max      0.093301    0.193236    0.158102    0.253829  …    0.224969

               X53         X54         X55         X56         X57         X58  \
```

count	473.000000	473.000000	473.000000	473.000000	473.000000	473.000000
mean	0.007578	0.000596	0.005831	0.002871	0.009561	0.002486
std	0.031696	0.043237	0.036063	0.043596	0.032426	0.043741
min	-0.188512	-0.256201	-0.285538	-0.244358	-0.145975	-0.265219
25%	-0.004402	-0.016683	-0.008791	-0.013554	-0.004371	-0.010859
50%	0.009890	0.006147	0.010309	0.006343	0.010405	0.005853
75%	0.024185	0.020026	0.024222	0.023376	0.026306	0.021170
max	0.124168	0.208974	0.161968	0.190750	0.243353	0.234014

	X59	X60	X61
count	473.000000	473.000000	473.000000
mean	0.007736	0.003153	0.008247
std	0.029933	0.043655	0.031350
min	-0.147334	-0.298681	-0.168224
25%	-0.005571	-0.012126	-0.006058
50%	0.010024	0.005155	0.008461
75%	0.023697	0.019104	0.024022
max	0.130404	0.294175	0.158422

[8 rows x 62 columns]

no_efectores

Composición de pseudo aminoácidos (PseAAC) no_efectores nematoda dataset 3, sin valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.062363	0.027717	0.020788	0.020788	0.020788	0.006929	0.000000	
1	0.028975	0.006439	0.040779	0.045072	0.011805	0.023609	0.008585	
2	0.046268	0.006382	0.035100	0.047864	0.030314	0.041482	0.017550	
3	0.030691	0.010230	0.012277	0.010230	0.035807	0.020461	0.010230	
4	0.036572	0.010757	0.038724	0.055934	0.024740	0.024740	0.015059	
..	…	…	…	…	…	…	…	
495	0.093964	0.011746	0.041109	0.070473	0.025449	0.019576	0.019576	
496	0.019205	0.016204	0.010803	0.033608	0.009002	0.017404	0.004201	
497	0.064537	0.059159	0.086049	0.064537	0.048403	0.059159	0.026890	
498	0.023619	0.020245	0.047238	0.030367	0.037116	0.040490	0.013497	
499	0.030693	0.010231	0.014323	0.018416	0.028647	0.016370	0.008185	

	X7	X8	X9	…	X53	X54	X55	\
0	0.027717	0.041575	0.027717	…	0.022755	-0.027772	-0.040186	
1	0.032194	0.048291	0.035414	…	0.026001	-0.003569	0.036975	
2	0.025527	0.041482	0.052650	…	0.004726	0.022228	0.020023	
3	0.046037	0.012277	0.049106	…	-0.004789	-0.007523	0.008218	
4	0.035497	0.060237	0.050556	…	0.031751	-0.014461	0.014197	
..	…	…	…	…	…	…	…	

```
495  0.062643  0.074388  0.078303  …  0.045966 -0.049441 -0.002661
496  0.013203  0.028207  0.012603  …  0.023483  0.013542  0.024293
497  0.099495  0.043025  0.207056  …  0.000413  0.036280  0.002561
498  0.053986  0.026993  0.060735  … -0.021884  0.020317  0.005298
499  0.032739  0.026600  0.065478  …  0.017829  0.033015  0.023070

          X56       X57       X58       X59       X60       X61           X62
0    0.063023  0.066739 -0.012268  0.014745 -0.054966 -0.013244  no_efectores
1   -0.001124  0.032853 -0.013471  0.032464  0.008192  0.023034  no_efectores
2   -0.019605 -0.000105 -0.011847 -0.002227  0.022799  0.058150  no_efectores
3   -0.000675 -0.001852  0.042188  0.018069  0.019682  0.011012  no_efectores
4   -0.006856  0.007585 -0.011732  0.026169 -0.006991  0.015623  no_efectores
..        …         …         …         …         …         …           …
495  0.025651  0.046412  0.017504  0.026184 -0.009789 -0.001911  no_efectores
496 -0.000618  0.021135  0.014514  0.035181  0.008102  0.021625  no_efectores
497 -0.047682 -0.048437 -0.054905 -0.035420 -0.067635 -0.079254  no_efectores
498 -0.023636 -0.006279  0.018027 -0.002299  0.041133  0.036828  no_efectores
499  0.038952  0.027134  0.009499  0.014217 -0.004823 -0.016712  no_efectores

[477 rows x 63 columns]
```

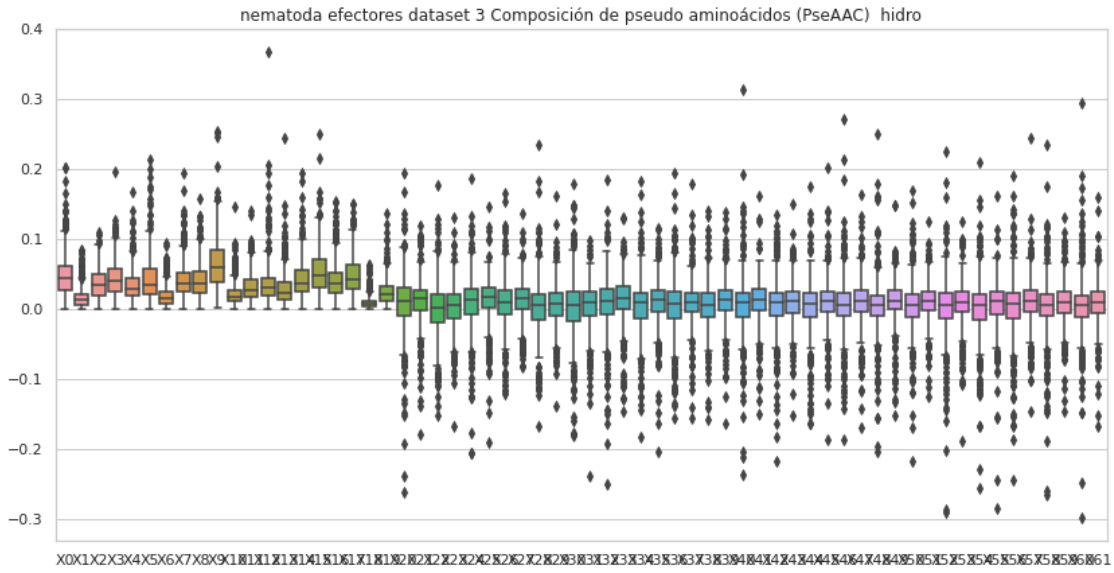
Composición de pseudo aminoácidos (PseAAC) no_efectores nematoda dataset 3, sin valores atípicos.
Estadísticas.

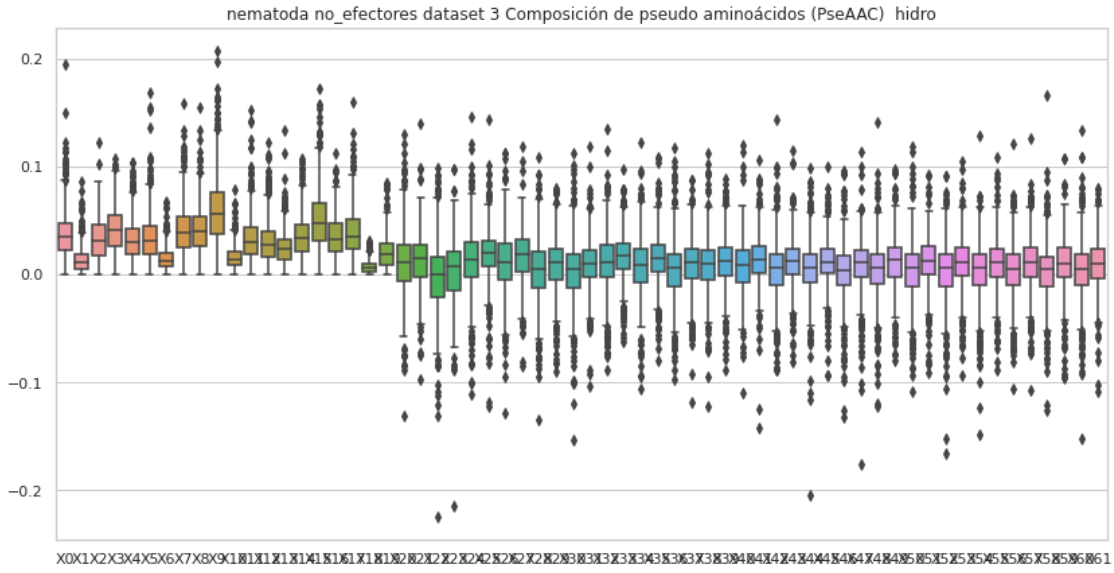
```
                X0          X1          X2          X3          X4          X5  \
count  477.000000  477.000000  477.000000  477.000000  477.000000  477.000000
mean     0.038896    0.013754    0.033015    0.041787    0.032019    0.035386
std      0.024163    0.011904    0.019033    0.020508    0.019591    0.023437
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.022051    0.005512    0.017996    0.026340    0.018771    0.018497
50%      0.034575    0.011200    0.031644    0.041837    0.029706    0.031336
75%      0.048096    0.019077    0.045710    0.054669    0.042065    0.044729
max      0.194395    0.085746    0.122018    0.107262    0.103246    0.168948

                X6          X7          X8          X9  …         X52  \
count  477.000000  477.000000  477.000000  477.000000  …  477.000000
mean     0.015251    0.041285    0.042121    0.059213  …    0.002719
std      0.011378    0.023805    0.022416    0.032820  …    0.029513
min      0.000000    0.000000    0.000000    0.000000  …   -0.165311
25%      0.007085    0.024808    0.026600    0.037238  …   -0.010955
50%      0.012737    0.039144    0.040528    0.055930  …    0.006050
75%      0.020426    0.053895    0.053840    0.076336  …    0.018359
max      0.066898    0.158563    0.154894    0.207056  …    0.091483

               X53         X54         X55         X56         X57         X58  \
count  477.000000  477.000000  477.000000  477.000000  477.000000  477.000000
```

mean	0.010631	0.003591	0.010349	0.004046	0.011068	0.002671
std	0.024691	0.027738	0.024272	0.026193	0.024653	0.027765
min	-0.092859	-0.148573	-0.089081	-0.105365	-0.106846	-0.126204
25%	-0.001215	-0.009619	-0.002672	-0.010202	-0.001852	-0.011161
50%	0.011268	0.006902	0.011630	0.005696	0.011008	0.004718
75%	0.025015	0.019189	0.023551	0.018322	0.024429	0.016077
max	0.104376	0.128666	0.109119	0.121140	0.126364	0.165482

	X59	X60	X61
count	477.000000	477.000000	477.000000
mean	0.009389	0.002794	0.009599
std	0.025635	0.030040	0.025174
min	-0.095959	-0.151764	-0.108193
25%	-0.002541	-0.009566	-0.003287
50%	0.010404	0.005432	0.010356
75%	0.024790	0.018648	0.024256
max	0.107120	0.132980	0.080119

[8 rows x 62 columns]





6 Covarianza de auto cruzamiento (ACC) hidro_mass

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

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

    if etiq == "efectores":
        df=ACC_hidro_mass_efec

    if etiq == "no_efectores":
        df=ACC_hidro_mass_no_efec

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

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

```
efectores
Covarianza de auto cruzamiento (ACC)  hidro_mass efectores nematoda dataset 3,
con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.067574  0.014875  0.035373 -0.044539 -0.081174 -0.002364  0.026290
1   -0.038948 -0.021504 -0.063170 -0.002248  0.041282  0.032201  0.005037
2    0.032394 -0.166128 -0.030822 -0.140125 -0.140222 -0.057108  0.168507
3    0.034751 -0.121128 -0.018064  0.151695  0.029766 -0.007470  0.081695
4    0.038887 -0.178362  0.081188  0.217353  0.051034 -0.073401  0.031043
..        …       …         …         …         …         …         …
495 -0.040382  0.063172 -0.017651 -0.128119  0.045118  0.051010  0.018887
496  0.045805  0.001007 -0.026085 -0.056956  0.053699  0.035992 -0.011649
497  0.027568  0.007965  0.007640 -0.043602 -0.019084  0.056352 -0.099296
498  0.027346 -0.063347 -0.007162  0.006551  0.012856 -0.041870  0.027789
499  0.095074  0.009598  0.064270 -0.069404 -0.057066 -0.021500  0.059777

           X7        X8        X9       X10       X11       X12        X13
0   -0.051971  0.048886 -0.074416 -0.009161 -0.036376  0.031173  efectores
1   -0.053610  0.029868  0.059712  0.041194 -0.188499 -0.019537  efectores
2    0.135215 -0.012155 -0.032203 -0.014851 -0.101480  0.025777  efectores
3   -0.002985 -0.051724  0.040106  0.128700 -0.263257 -0.183772  efectores
4    0.154368  0.045239  0.042857  0.227094  0.038654 -0.108649  efectores
..        …       …         …         …         …         …         …
495  0.052714  0.020648 -0.022387  0.016803 -0.022150  0.030789  efectores
496  0.070088 -0.027904  0.021147  0.016688  0.015516  0.042576  efectores
497 -0.104812 -0.155987  0.016726 -0.042135 -0.001740 -0.061822  efectores
498  0.000744 -0.012621  0.021992 -0.009248 -0.002035  0.029154  efectores
499 -0.065315 -0.049809  0.047032 -0.031616 -0.053430  0.041541  efectores

[500 rows x 14 columns]


Covarianza de auto cruzamiento (ACC)  hidro_mass efectores nematoda dataset 3,
con valores atípicos.
Estadísticas.

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

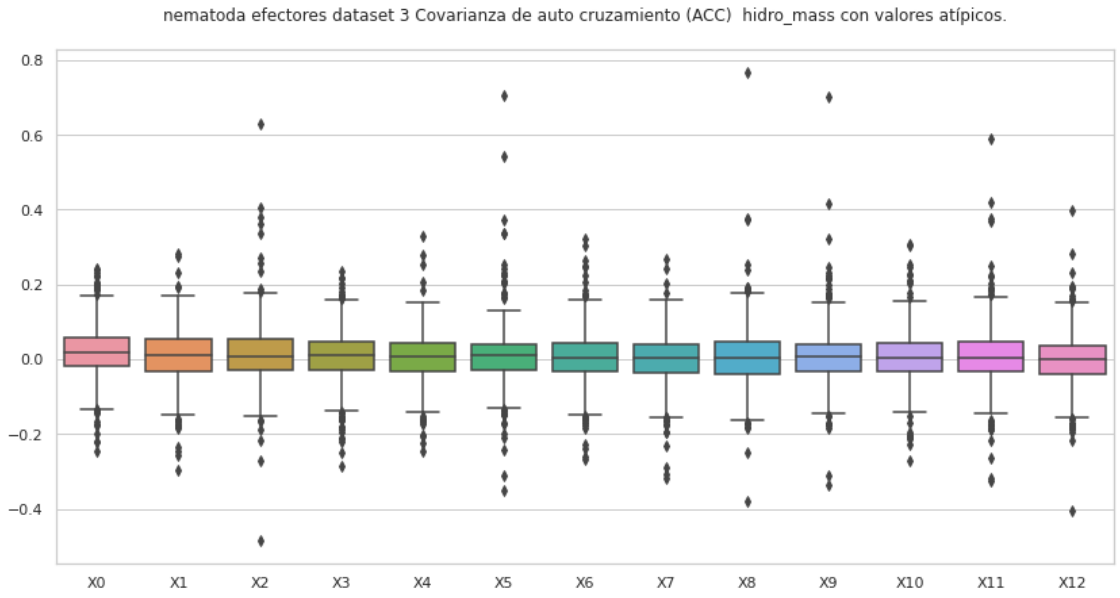
```
mean     0.019660    0.008109    0.013582    0.007004    0.005833    0.010581
std      0.071147    0.073303    0.083608    0.071025    0.067429    0.083164
min     -0.245955   -0.298363   -0.485052   -0.284438   -0.247287   -0.352898
25%     -0.019716   -0.031548   -0.029756   -0.027688   -0.031639   -0.029165
50%      0.016492    0.010240    0.008611    0.011593    0.007622    0.009733
75%      0.057163    0.053576    0.053927    0.046652    0.041791    0.039709
max      0.242964    0.280443    0.630979    0.235888    0.330367    0.705028

                X6          X7          X8          X9         X10         X11  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.003799    0.000524    0.005211    0.007145    0.004338    0.007213
std      0.074315    0.071398    0.083589    0.079034    0.072246    0.083414
min     -0.267772   -0.319206   -0.380071   -0.335842   -0.271925   -0.327207
25%     -0.033101   -0.037720   -0.040085   -0.034437   -0.033826   -0.033710
50%      0.005090    0.004349    0.002168    0.005941    0.004491    0.004258
75%      0.044660    0.041310    0.046363    0.041352    0.044328    0.047744
max      0.320432    0.267904    0.766804    0.702911    0.309136    0.588851

               X12
count  500.000000
mean    -0.001829
std      0.072951
min     -0.406405
25%     -0.041158
50%      0.000453
75%      0.036659
max      0.398449



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

            X0        X1        X2        X3        X4        X5        X6  \
0   -0.080318  0.415328 -0.230321  0.083258 -0.274482 -0.242999 -0.193973
1   -0.001729 -0.023643 -0.046531 -0.024700  0.019356 -0.021432 -0.015418
2    0.061067  0.069078 -0.024183  0.105004  0.002223 -0.012671  0.012263
3   -0.032425  0.046906 -0.034851  0.043032 -0.030402  0.048499  0.051594
4    0.017289  0.057714 -0.020004  0.054341 -0.058852 -0.012305 -0.033200
..        …         …         …         …         …         …         …
495 -0.041017 -0.029259  0.000538 -0.026881  0.071518  0.018035  0.041053
496 -0.001644  0.006011  0.034200 -0.016577  0.029214 -0.036656  0.005452
497  0.012145 -0.048386  0.032361  0.018159  0.005582  0.006625 -0.015101
498 -0.051507 -0.059683  0.055028  0.077163 -0.064524  0.129438  0.075755
499  0.110203 -0.017338  0.003435  0.005879 -0.027668  0.027413  0.052665
```

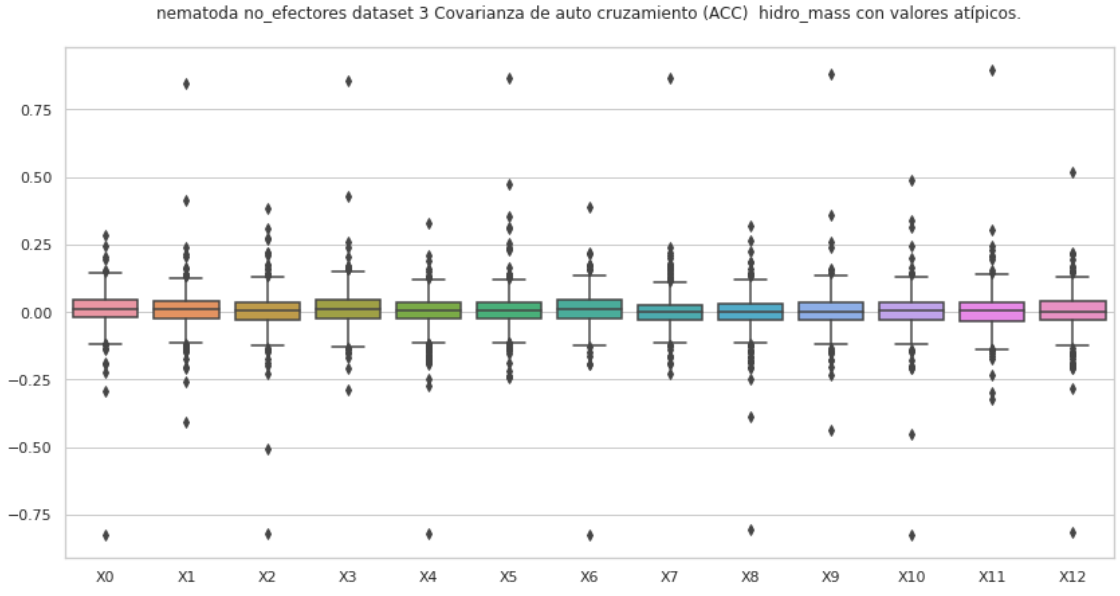
```
           X7        X8        X9       X10       X11       X12           X13
0   -0.037777 -0.019144  0.047755 -0.069682  0.139663 -0.057241  no_efectores
1   -0.027785 -0.018700  0.016540 -0.033451 -0.008687  0.014901  no_efectores
2   -0.014634  0.025758 -0.106538  0.060003  0.055937  0.001520  no_efectores
3   -0.057391  0.098589 -0.018735  0.021412 -0.021453  0.076114  no_efectores
4    0.003182  0.030371  0.067385 -0.005745  0.060547  0.008385  no_efectores
..        …         …         …         …         …         …
495  0.027128 -0.067596  0.022308  0.010472  0.010027 -0.011926  no_efectores
496 -0.023508 -0.010705  0.014787  0.018378 -0.087119 -0.027442  no_efectores
497 -0.001780  0.009955 -0.024945  0.011112  0.004385 -0.052488  no_efectores
498 -0.001550  0.021319  0.014881 -0.010143  0.068720 -0.003104  no_efectores
499  0.044254  0.071150 -0.073051  0.014066  0.001662  0.081498  no_efectores

[500 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores nematoda dataset 3, con valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.010041    0.007967    0.004467    0.012789   -0.000627    0.007570
std      0.068832    0.073274    0.077439    0.074485    0.071224    0.077780
min     -0.824744   -0.405688   -0.822580   -0.290487   -0.820578   -0.242999
25%     -0.019782   -0.022986   -0.028450   -0.024903   -0.026458   -0.025067
50%      0.010915    0.008452    0.005190    0.011391    0.004826    0.006426
75%      0.045574    0.039256    0.036523    0.045041    0.033344    0.033700
max      0.284696    0.847939    0.384854    0.857303    0.329170    0.864414

                X6          X7          X8          X9         X10         X11  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.008370    0.003639   -0.000849    0.005538    0.005861    0.002915
std      0.069103    0.069536    0.071382    0.074395    0.075869    0.074883
min     -0.825343   -0.229665   -0.805523   -0.437411   -0.825913   -0.321234
25%     -0.025439   -0.029541   -0.027618   -0.031478   -0.027611   -0.032269
50%      0.010231   -0.000880    0.001016    0.002460    0.007077    0.003245
75%      0.043422    0.027233    0.031719    0.037526    0.036517    0.037688
max      0.387990    0.869286    0.319127    0.883810    0.485997    0.896302

               X12
count  500.000000
mean    -0.000529
std      0.077085
min     -0.814735
25%     -0.029889
50%      0.001695
75%      0.038198
```

```
max       0.520689
```


nematoda efectores dataset 3 Covarianza de auto cruzamiento (ACC) hidro_mass con valores atípicos.
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12


nematoda no_efectores dataset 3 Covarianza de auto cruzamiento (ACC) hidro_mass con valores atípicos.
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12

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

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

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

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

    if etiq == "efectores":
        df=ACC_hidro_mass_efec

    if etiq == "no_efectores":
        df=ACC_hidro_mass_no_efec

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

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

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

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

efectores
Covarianza de auto cruzamiento (ACC) hidro_mass efectores nematoda dataset 3,
sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	-0.067574	0.014875	0.035373	-0.044539	-0.081174	-0.002364	0.026290	
1	-0.038948	-0.021504	-0.063170	-0.002248	0.041282	0.032201	0.005037	
2	0.032394	-0.166128	-0.030822	-0.140125	-0.140222	-0.057108	0.168507	
6	0.231971	0.115592	0.136653	0.120364	0.125282	0.251953	0.223327	
8	-0.037751	0.101132	0.235923	-0.039544	0.044689	-0.049065	0.069713	
..	…	…	…	…	…	…	…	
495	-0.040382	0.063172	-0.017651	-0.128119	0.045118	0.051010	0.018887	
496	0.045805	0.001007	-0.026085	-0.056956	0.053699	0.035992	-0.011649	
497	0.027568	0.007965	0.007640	-0.043602	-0.019084	0.056352	-0.099296	
498	0.027346	-0.063347	-0.007162	0.006551	0.012856	-0.041870	0.027789	
499	0.095074	0.009598	0.064270	-0.069404	-0.057066	-0.021500	0.059777	

	X7	X8	X9	X10	X11	X12	X13
0	-0.051971	0.048886	-0.074416	-0.009161	-0.036376	0.031173	efectores
1	-0.053610	0.029868	0.059712	0.041194	-0.188499	-0.019537	efectores
2	0.135215	-0.012155	-0.032203	-0.014851	-0.101480	0.025777	efectores
6	-0.071779	0.237496	0.216809	0.057060	-0.053500	0.029196	efectores
8	0.009519	-0.152949	-0.013519	-0.072980	-0.136514	-0.082855	efectores
..	…	…	…	…	…	…	…
495	0.052714	0.020648	-0.022387	0.016803	-0.022150	0.030789	efectores
496	0.070088	-0.027904	0.021147	0.016688	0.015516	0.042576	efectores
497	-0.104812	-0.155987	0.016726	-0.042135	-0.001740	-0.061822	efectores
498	0.000744	-0.012621	0.021992	-0.009248	-0.002035	0.029154	efectores
499	-0.065315	-0.049809	0.047032	-0.031616	-0.053430	0.041541	efectores

[462 rows x 14 columns]

Covarianza de auto cruzamiento (ACC) hidro_mass efectores nematoda dataset 3,
sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	462.000000	462.000000	462.000000	462.000000	462.000000	462.000000	
mean	0.019135	0.009654	0.010052	0.006980	0.005876	0.007203	
std	0.063627	0.063583	0.063042	0.063220	0.059104	0.060692	
min	-0.178700	-0.181927	-0.217473	-0.197299	-0.175029	-0.208822	
25%	-0.017882	-0.029280	-0.028373	-0.027212	-0.029631	-0.028275	
50%	0.016081	0.010240	0.008251	0.012035	0.007874	0.009733	
75%	0.053493	0.050550	0.050286	0.045019	0.040788	0.038187	
max	0.231971	0.194721	0.255835	0.201013	0.206744	0.251953	

```
                X6          X7          X8          X9          X10          X11  \
count  462.000000  462.000000  462.000000  462.000000  462.000000  462.000000
mean     0.003725    0.001889    0.002628    0.001890    0.002905    0.003927
std      0.064588    0.061580    0.066038    0.061507    0.059738    0.065117
min     -0.184715   -0.197066   -0.183040   -0.178554   -0.208197   -0.189706
25%     -0.030917   -0.034950   -0.036949   -0.035234   -0.032212   -0.033415
50%      0.005344    0.006001    0.001500    0.004372    0.003844    0.002170
75%      0.043767    0.040037    0.044863    0.035520    0.041116    0.044974
max      0.223327    0.203857    0.252124    0.231388    0.203805    0.223539

               X12
count  462.000000
mean     0.000224
std      0.062096
min     -0.195419
25%     -0.034989
50%      0.002688
75%      0.036587
max      0.190101
```

no_efectores
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores nematoda dataset 3, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
1   -0.001729 -0.023643 -0.046531 -0.024700  0.019356 -0.021432 -0.015418
2    0.061067  0.069078 -0.024183  0.105004  0.002223 -0.012671  0.012263
3   -0.032425  0.046906 -0.034851  0.043032 -0.030402  0.048499  0.051594
4    0.017289  0.057714 -0.020004  0.054341 -0.058852 -0.012305 -0.033200
5   -0.047089  0.022403 -0.073033  0.007530 -0.029163 -0.059911  0.092467
..        …         …         …         …         …         …
495 -0.041017 -0.029259  0.000538 -0.026881  0.071518  0.018035  0.041053
496 -0.001644  0.006011  0.034200 -0.016577  0.029214 -0.036656  0.005452
497  0.012145 -0.048386  0.032361  0.018159  0.005582  0.006625 -0.015101
498 -0.051507 -0.059683  0.055028  0.077163 -0.064524  0.129438  0.075755
499  0.110203 -0.017338  0.003435  0.005879 -0.027668  0.027413  0.052665

           X7        X8        X9       X10       X11       X12           X13
1   -0.027785 -0.018700  0.016540 -0.033451 -0.008687  0.014901  no_efectores
2   -0.014634  0.025758 -0.106538  0.060003  0.055937  0.001520  no_efectores
3   -0.057391  0.098589 -0.018735  0.021412 -0.021453  0.076114  no_efectores
4    0.003182  0.030371  0.067385 -0.005745  0.060547  0.008385  no_efectores
5   -0.045124 -0.009771 -0.052347 -0.017109 -0.030457  0.003225  no_efectores
..        …         …         …         …         …         …
495  0.027128 -0.067596  0.022308  0.010472  0.010027 -0.011926  no_efectores
```

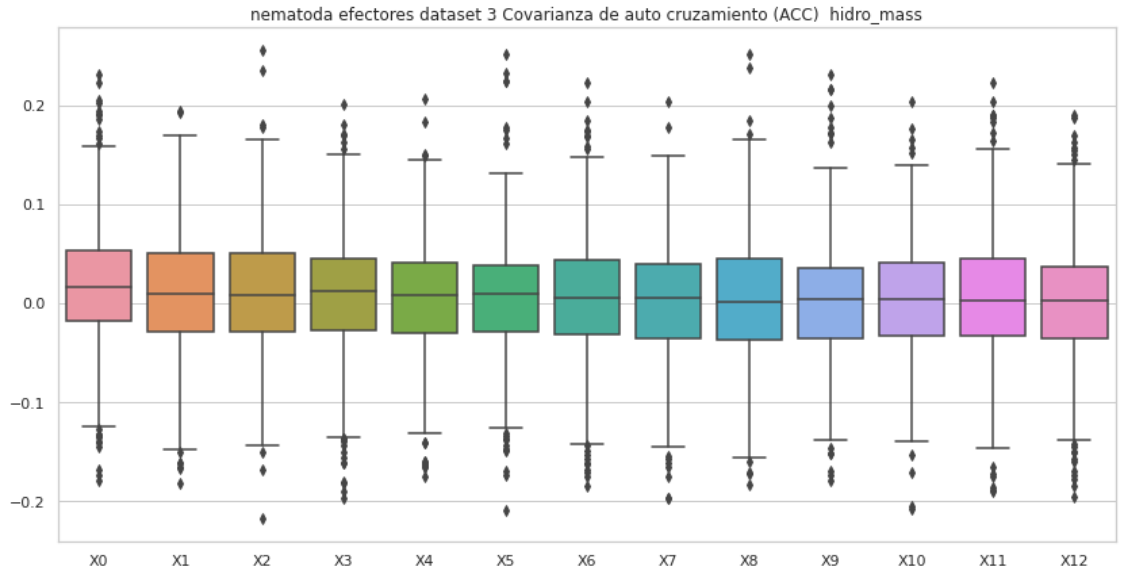
```
496 -0.023508 -0.010705  0.014787  0.018378 -0.087119 -0.027442  no_efectores
497 -0.001780  0.009955 -0.024945  0.011112  0.004385 -0.052488  no_efectores
498 -0.001550  0.021319  0.014881 -0.010143  0.068720 -0.003104  no_efectores
499  0.044254  0.071150 -0.073051  0.014066  0.001662  0.081498  no_efectores

[471 rows x 14 columns]
```

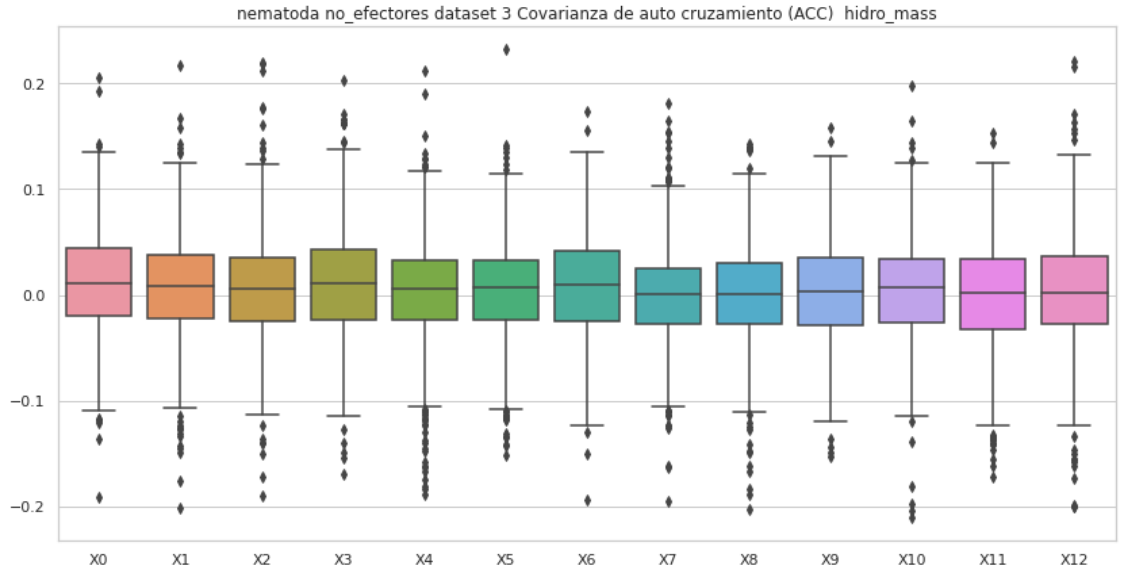
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores nematoda dataset 3, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	471.000000	471.000000	471.000000	471.000000	471.000000	471.000000	
mean	0.012467	0.006045	0.006419	0.009549	0.002967	0.004528	
std	0.050160	0.052273	0.054690	0.055055	0.054975	0.049559	
min	-0.190985	-0.201155	-0.189637	-0.169785	-0.188606	-0.151625	
25%	-0.019132	-0.021773	-0.025381	-0.023955	-0.023477	-0.023206	
50%	0.011223	0.008464	0.005494	0.010819	0.005582	0.006510	
75%	0.043925	0.037789	0.035510	0.042720	0.032825	0.033031	
max	0.204710	0.216139	0.219266	0.202914	0.211820	0.231979	

	X6	X7	X8	X9	X10	X11	\
count	471.000000	471.000000	471.000000	471.000000	471.000000	471.000000	
mean	0.008320	0.000615	0.000199	0.004243	0.006200	-0.001056	
std	0.051058	0.049946	0.049870	0.051017	0.051992	0.052709	
min	-0.193573	-0.194222	-0.202908	-0.152836	-0.210159	-0.171367	
25%	-0.024913	-0.027925	-0.026807	-0.029094	-0.026531	-0.031968	
50%	0.009604	0.000371	0.000771	0.003157	0.006807	0.002066	
75%	0.041138	0.025570	0.030264	0.035642	0.034334	0.033746	
max	0.172919	0.180820	0.143154	0.157962	0.197494	0.152817	

	X12
count	471.000000
mean	0.001571
std	0.058013
min	-0.200238
25%	-0.027388
50%	0.002086
75%	0.036905
max	0.221148





7 Covarianza de auto cruzamiento (ACC) mass

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

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

    if etiq == "efectores":
        df=ACC_mass_efec

    if etiq == "no_efectores":
        df=ACC_mass_no_efec

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

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

```
efectores
Covarianza de auto cruzamiento (ACC)  mass efectores nematoda dataset 3, con
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.067574  0.014875  0.035373 -0.044539 -0.081174 -0.002364  0.026290
1   -0.038948 -0.021504 -0.063170 -0.002248  0.041282  0.032201  0.005037
2    0.032394 -0.166128 -0.030822 -0.140125 -0.140222 -0.057108  0.168507
3    0.034751 -0.121128 -0.018064  0.151695  0.029766 -0.007470  0.081695
4    0.038887 -0.178362  0.081188  0.217353  0.051034 -0.073401  0.031043
..        …         …         …         …         …         …         …
495 -0.040382  0.063172 -0.017651 -0.128119  0.045118  0.051010  0.018887
496  0.045805  0.001007 -0.026085 -0.056956  0.053699  0.035992 -0.011649
497  0.027568  0.007965  0.007640 -0.043602 -0.019084  0.056352 -0.099296
498  0.027346 -0.063347 -0.007162  0.006551  0.012856 -0.041870  0.027789
499  0.095074  0.009598  0.064270 -0.069404 -0.057066 -0.021500  0.059777

           X7        X8        X9       X10       X11       X12        X13
0   -0.051971  0.048886 -0.074416 -0.009161 -0.036376  0.031173  efectores
1   -0.053610  0.029868  0.059712  0.041194 -0.188499 -0.019537  efectores
```

```
2    0.135215 -0.012155 -0.032203 -0.014851 -0.101480  0.025777  efectores
3   -0.002985 -0.051724  0.040106  0.128700 -0.263257 -0.183772  efectores
4    0.154368  0.045239  0.042857  0.227094  0.038654 -0.108649  efectores
..        …         …         …         …         …         …
495  0.052714  0.020648 -0.022387  0.016803 -0.022150  0.030789  efectores
496  0.070088 -0.027904  0.021147  0.016688  0.015516  0.042576  efectores
497 -0.104812 -0.155987  0.016726 -0.042135 -0.001740 -0.061822  efectores
498  0.000744 -0.012621  0.021992 -0.009248 -0.002035  0.029154  efectores
499 -0.065315 -0.049809  0.047032 -0.031616 -0.053430  0.041541  efectores

[500 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) mass efectores nematoda dataset 3, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.019660	0.008109	0.013582	0.007004	0.005833	0.010581	
std	0.071147	0.073303	0.083608	0.071025	0.067429	0.083164	
min	-0.245955	-0.298363	-0.485052	-0.284438	-0.247287	-0.352898	
25%	-0.019716	-0.031548	-0.029756	-0.027688	-0.031639	-0.029165	
50%	0.016492	0.010240	0.008611	0.011593	0.007622	0.009733	
75%	0.057163	0.053576	0.053927	0.046652	0.041791	0.039709	
max	0.242964	0.280443	0.630979	0.235888	0.330367	0.705028	

	X6	X7	X8	X9	X10	X11	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.003799	0.000524	0.005211	0.007145	0.004338	0.007213	
std	0.074315	0.071398	0.083589	0.079034	0.072246	0.083414	
min	-0.267772	-0.319206	-0.380071	-0.335842	-0.271925	-0.327207	
25%	-0.033101	-0.037720	-0.040085	-0.034437	-0.033826	-0.033710	
50%	0.005090	0.004349	0.002168	0.005941	0.004491	0.004258	
75%	0.044660	0.041310	0.046363	0.041352	0.044328	0.047744	
max	0.320432	0.267904	0.766804	0.702911	0.309136	0.588851	

	X12
count	500.000000
mean	-0.001829
std	0.072951
min	-0.406405
25%	-0.041158
50%	0.000453
75%	0.036659
max	0.398449

no_efectores
Covarianza de auto cruzamiento (ACC) mass no_efectores nematoda dataset 3, con valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	-0.080318	0.415328	-0.230321	0.083258	-0.274482	-0.242999	-0.193973	
1	-0.001729	-0.023643	-0.046531	-0.024700	0.019356	-0.021432	-0.015418	
2	0.061067	0.069078	-0.024183	0.105004	0.002223	-0.012671	0.012263	
3	-0.032425	0.046906	-0.034851	0.043032	-0.030402	0.048499	0.051594	
4	0.017289	0.057714	-0.020004	0.054341	-0.058852	-0.012305	-0.033200	
..	…	…	…	…	…	…	…	
495	-0.041017	-0.029259	0.000538	-0.026881	0.071518	0.018035	0.041053	
496	-0.001644	0.006011	0.034200	-0.016577	0.029214	-0.036656	0.005452	
497	0.012145	-0.048386	0.032361	0.018159	0.005582	0.006625	-0.015101	
498	-0.051507	-0.059683	0.055028	0.077163	-0.064524	0.129438	0.075755	
499	0.110203	-0.017338	0.003435	0.005879	-0.027668	0.027413	0.052665	

	X7	X8	X9	X10	X11	X12	X13
0	-0.037777	-0.019144	0.047755	-0.069682	0.139663	-0.057241	no_efectores
1	-0.027785	-0.018700	0.016540	-0.033451	-0.008687	0.014901	no_efectores
2	-0.014634	0.025758	-0.106538	0.060003	0.055937	0.001520	no_efectores
3	-0.057391	0.098589	-0.018735	0.021412	-0.021453	0.076114	no_efectores
4	0.003182	0.030371	0.067385	-0.005745	0.060547	0.008385	no_efectores
..	…	…	…	…	…	…	…
495	0.027128	-0.067596	0.022308	0.010472	0.010027	-0.011926	no_efectores
496	-0.023508	-0.010705	0.014787	0.018378	-0.087119	-0.027442	no_efectores
497	-0.001780	0.009955	-0.024945	0.011112	0.004385	-0.052488	no_efectores
498	-0.001550	0.021319	0.014881	-0.010143	0.068720	-0.003104	no_efectores
499	0.044254	0.071150	-0.073051	0.014066	0.001662	0.081498	no_efectores

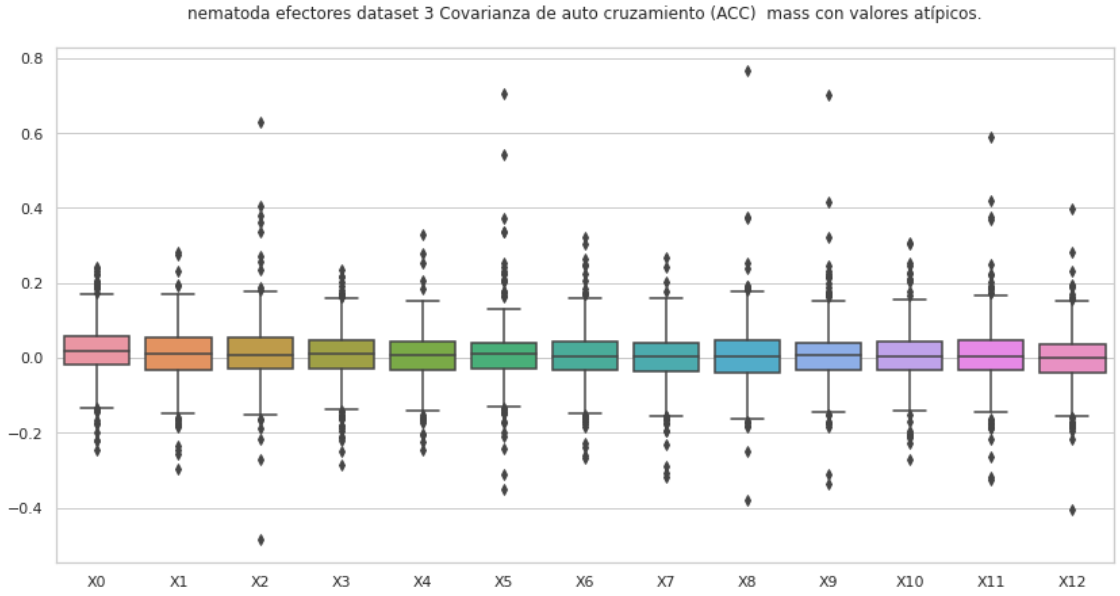
[500 rows x 14 columns]

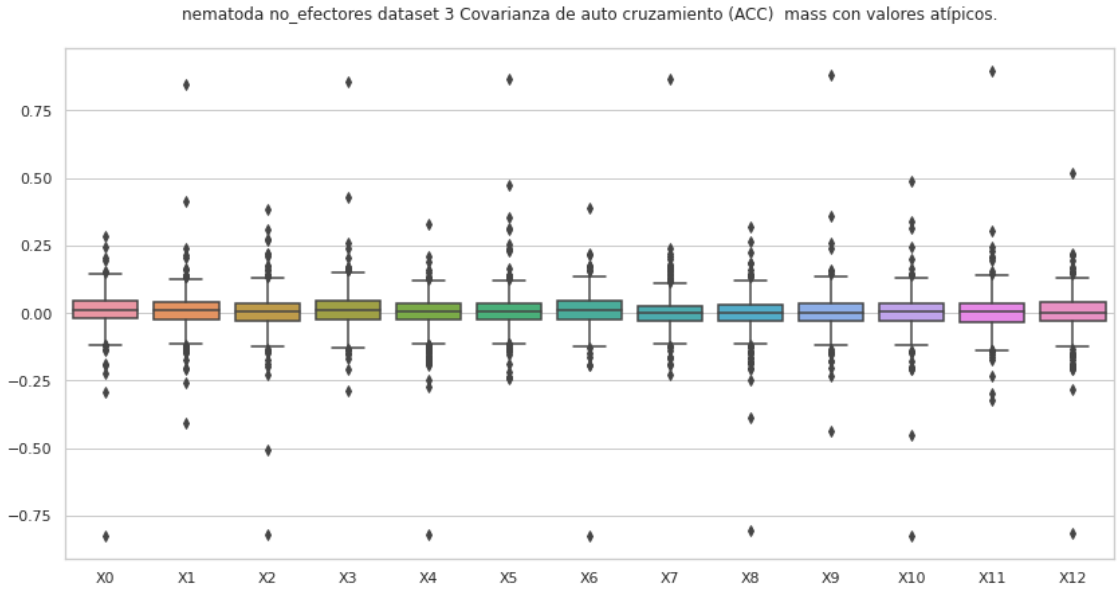
Covarianza de auto cruzamiento (ACC) mass no_efectores nematoda dataset 3, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.010041	0.007967	0.004467	0.012789	-0.000627	0.007570	
std	0.068832	0.073274	0.077439	0.074485	0.071224	0.077780	
min	-0.824744	-0.405688	-0.822580	-0.290487	-0.820578	-0.242999	
25%	-0.019782	-0.022986	-0.028450	-0.024903	-0.026458	-0.025067	
50%	0.010915	0.008452	0.005190	0.011391	0.004826	0.006426	
75%	0.045574	0.039256	0.036523	0.045041	0.033344	0.033700	
max	0.284696	0.847939	0.384854	0.857303	0.329170	0.864414	

	X6	X7	X8	X9	X10	X11	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.008370	0.003639	-0.000849	0.005538	0.005861	0.002915	
std	0.069103	0.069536	0.071382	0.074395	0.075869	0.074883	
min	-0.825343	-0.229665	-0.805523	-0.437411	-0.825913	-0.321234	
25%	-0.025439	-0.029541	-0.027618	-0.031478	-0.027611	-0.032269	
50%	0.010231	-0.000880	0.001016	0.002460	0.007077	0.003245	
75%	0.043422	0.027233	0.031719	0.037526	0.036517	0.037688	
max	0.387990	0.869286	0.319127	0.883810	0.485997	0.896302	

	X12
count	500.000000
mean	-0.000529
std	0.077085
min	-0.814735
25%	-0.029889
50%	0.001695
75%	0.038198
max	0.520689





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

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

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

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

    if etiq == "efectores":
        df=ACC_mass_efec

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

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

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

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

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

```
Covarianza de auto cruzamiento (ACC)  mass efectores nematoda dataset 3, sin
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.067574  0.014875  0.035373 -0.044539 -0.081174 -0.002364  0.026290
1   -0.038948 -0.021504 -0.063170 -0.002248  0.041282  0.032201  0.005037
2    0.032394 -0.166128 -0.030822 -0.140125 -0.140222 -0.057108  0.168507
6    0.231971  0.115592  0.136653  0.120364  0.125282  0.251953  0.223327
8   -0.037751  0.101132  0.235923 -0.039544  0.044689 -0.049065  0.069713
..        …         …         …         …         …         …         …
495 -0.040382  0.063172 -0.017651 -0.128119  0.045118  0.051010  0.018887
496  0.045805  0.001007 -0.026085 -0.056956  0.053699  0.035992 -0.011649
497  0.027568  0.007965  0.007640 -0.043602 -0.019084  0.056352 -0.099296
498  0.027346 -0.063347 -0.007162  0.006551  0.012856 -0.041870  0.027789
499  0.095074  0.009598  0.064270 -0.069404 -0.057066 -0.021500  0.059777

           X7        X8        X9       X10       X11       X12        X13
0   -0.051971  0.048886 -0.074416 -0.009161 -0.036376  0.031173  efectores
1   -0.053610  0.029868  0.059712  0.041194 -0.188499 -0.019537  efectores
2    0.135215 -0.012155 -0.032203 -0.014851 -0.101480  0.025777  efectores
6   -0.071779  0.237496  0.216809  0.057060 -0.053500  0.029196  efectores
8    0.009519 -0.152949 -0.013519 -0.072980 -0.136514 -0.082855  efectores
..        …         …         …         …         …         …         …
```

```
495  0.052714  0.020648 -0.022387  0.016803 -0.022150  0.030789  efectores
496  0.070088 -0.027904  0.021147  0.016688  0.015516  0.042576  efectores
497 -0.104812 -0.155987  0.016726 -0.042135 -0.001740 -0.061822  efectores
498  0.000744 -0.012621  0.021992 -0.009248 -0.002035  0.029154  efectores
499 -0.065315 -0.049809  0.047032 -0.031616 -0.053430  0.041541  efectores

[462 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) mass efectores nematoda dataset 3, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	462.000000	462.000000	462.000000	462.000000	462.000000	462.000000	
mean	0.019135	0.009654	0.010052	0.006980	0.005876	0.007203	
std	0.063627	0.063583	0.063042	0.063220	0.059104	0.060692	
min	-0.178700	-0.181927	-0.217473	-0.197299	-0.175029	-0.208822	
25%	-0.017882	-0.029280	-0.028373	-0.027212	-0.029631	-0.028275	
50%	0.016081	0.010240	0.008251	0.012035	0.007874	0.009733	
75%	0.053493	0.050550	0.050286	0.045019	0.040788	0.038187	
max	0.231971	0.194721	0.255835	0.201013	0.206744	0.251953	

	X6	X7	X8	X9	X10	X11	\
count	462.000000	462.000000	462.000000	462.000000	462.000000	462.000000	
mean	0.003725	0.001889	0.002628	0.001890	0.002905	0.003927	
std	0.064588	0.061580	0.066038	0.061507	0.059738	0.065117	
min	-0.184715	-0.197066	-0.183040	-0.178554	-0.208197	-0.189706	
25%	-0.030917	-0.034950	-0.036949	-0.035234	-0.032212	-0.033415	
50%	0.005344	0.006001	0.001500	0.004372	0.003844	0.002170	
75%	0.043767	0.040037	0.044863	0.035520	0.041116	0.044974	
max	0.223327	0.203857	0.252124	0.231388	0.203805	0.223539	

	X12
count	462.000000
mean	0.000224
std	0.062096
min	-0.195419
25%	-0.034989
50%	0.002688
75%	0.036587
max	0.190101

Covarianza de auto cruzamiento (ACC) mass no_efectores nematoda dataset 3, sin valores atípicos.
Valores del documento csv.

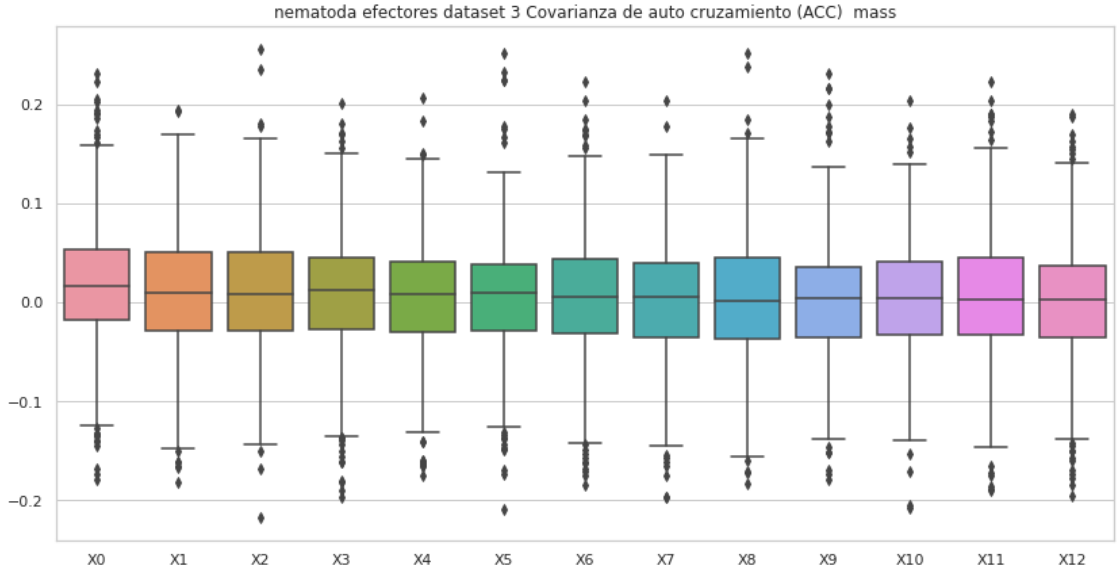
```
           X0        X1        X2        X3        X4        X5        X6  \
1   -0.001729 -0.023643 -0.046531 -0.024700  0.019356 -0.021432 -0.015418
2    0.061067  0.069078 -0.024183  0.105004  0.002223 -0.012671  0.012263
3   -0.032425  0.046906 -0.034851  0.043032 -0.030402  0.048499  0.051594
4    0.017289  0.057714 -0.020004  0.054341 -0.058852 -0.012305 -0.033200
5   -0.047089  0.022403 -0.073033  0.007530 -0.029163 -0.059911  0.092467
..        …         …         …         …         …         …
495 -0.041017 -0.029259  0.000538 -0.026881  0.071518  0.018035  0.041053
496 -0.001644  0.006011  0.034200 -0.016577  0.029214 -0.036656  0.005452
497  0.012145 -0.048386  0.032361  0.018159  0.005582  0.006625 -0.015101
498 -0.051507 -0.059683  0.055028  0.077163 -0.064524  0.129438  0.075755
499  0.110203 -0.017338  0.003435  0.005879 -0.027668  0.027413  0.052665

           X7        X8        X9       X10       X11       X12           X13
1   -0.027785 -0.018700  0.016540 -0.033451 -0.008687  0.014901  no_efectores
2   -0.014634  0.025758 -0.106538  0.060003  0.055937  0.001520  no_efectores
3   -0.057391  0.098589 -0.018735  0.021412 -0.021453  0.076114  no_efectores
4    0.003182  0.030371  0.067385 -0.005745  0.060547  0.008385  no_efectores
5   -0.045124 -0.009771 -0.052347 -0.017109 -0.030457  0.003225  no_efectores
..        …         …         …         …         …         …
495  0.027128 -0.067596  0.022308  0.010472  0.010027 -0.011926  no_efectores
496 -0.023508 -0.010705  0.014787  0.018378 -0.087119 -0.027442  no_efectores
497 -0.001780  0.009955 -0.024945  0.011112  0.004385 -0.052488  no_efectores
498 -0.001550  0.021319  0.014881 -0.010143  0.068720 -0.003104  no_efectores
499  0.044254  0.071150 -0.073051  0.014066  0.001662  0.081498  no_efectores

[471 rows x 14 columns]
```

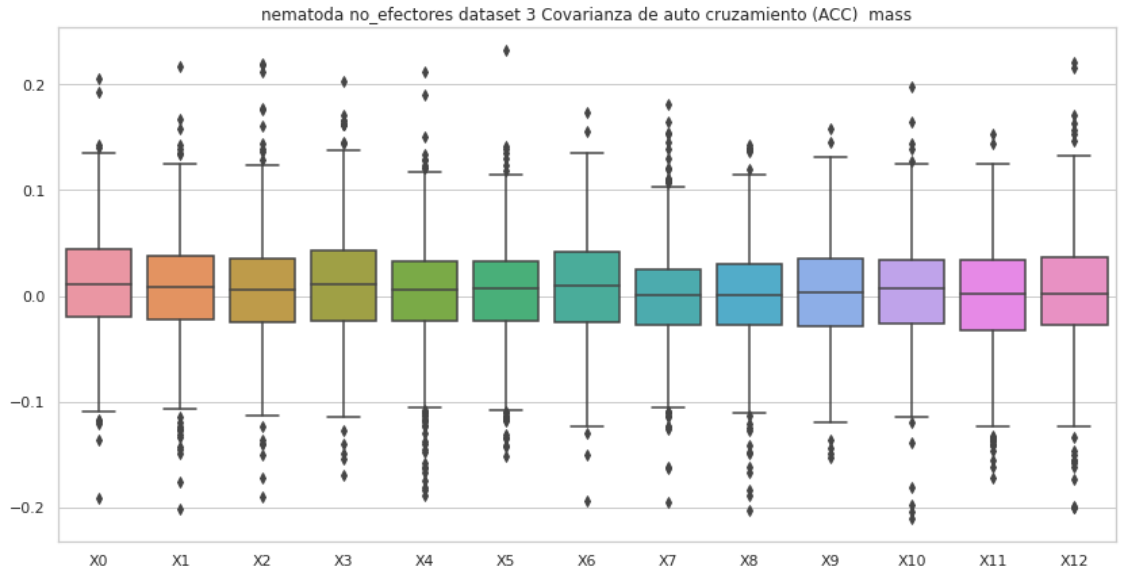
Covarianza de auto cruzamiento (ACC) mass no_efectores nematoda dataset 3, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  471.000000  471.000000  471.000000  471.000000  471.000000  471.000000
mean     0.012467    0.006045    0.006419    0.009549    0.002967    0.004528
std      0.050160    0.052273    0.054690    0.055055    0.054975    0.049559
min     -0.190985   -0.201155   -0.189637   -0.169785   -0.188606   -0.151625
25%     -0.019132   -0.021773   -0.025381   -0.023955   -0.023477   -0.023206
50%      0.011223    0.008464    0.005494    0.010819    0.005582    0.006510
75%      0.043925    0.037789    0.035510    0.042720    0.032825    0.033031
max      0.204710    0.216139    0.219266    0.202914    0.211820    0.231979

                X6          X7          X8          X9         X10         X11  \
count  471.000000  471.000000  471.000000  471.000000  471.000000  471.000000
mean     0.008320    0.000615    0.000199    0.004243    0.006200   -0.001056
std      0.051058    0.049946    0.049870    0.051017    0.051992    0.052709
```

min	-0.193573	-0.194222	-0.202908	-0.152836	-0.210159	-0.171367
25%	-0.024913	-0.027925	-0.026807	-0.029094	-0.026531	-0.031968
50%	0.009604	0.000371	0.000771	0.003157	0.006807	0.002066
75%	0.041138	0.025570	0.030264	0.035642	0.034334	0.033746
max	0.172919	0.180820	0.143154	0.157962	0.197494	0.152817

	X12
count	471.000000
mean	0.001571
std	0.058013
min	-0.200238
25%	-0.027388
50%	0.002086
75%	0.036905
max	0.221148





8 Covarianza de auto cruzamiento (ACC) hidro

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

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

    if etiq == "efectores":
        df=ACC_hidro_efec

    if etiq == "no_efectores":
        df=ACC_hidro_no_efec

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

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

```
efectores
Covarianza de auto cruzamiento (ACC)  hidro efectores nematoda dataset 3, con
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.045981  0.098141  0.106728 -0.060690 -0.083583 -0.059933  0.038655
1   -0.064242 -0.111209  0.002145 -0.092619 -0.108017  0.006212  0.173279
2   -0.125122 -0.101022 -0.020189  0.041991 -0.120725  0.056036 -0.129436
3    0.158284  0.049502  0.225241  0.205561  0.042274 -0.026775  0.296503
4    0.158192  0.229680  0.317186  0.181821  0.113494  0.204116  0.200391
..        …         …         …         …         …         …         …
495  0.022047 -0.025041  0.001914 -0.045175  0.025893 -0.064922  0.115761
496  0.117091  0.045035  0.060781  0.056875  0.002347 -0.012986  0.000166
497  0.201939  0.032856  0.129623  0.086441  0.130097  0.160020  0.046970
498 -0.006665 -0.001886  0.053790 -0.008732 -0.041052 -0.077728 -0.032580
499  0.034034 -0.127440  0.068490  0.117965 -0.087883 -0.028590  0.181336

           X7        X8        X9       X10       X11       X12        X13
0   -0.096047  0.041216  0.142697 -0.055478  0.101260  0.019191  efectores
1   -0.050172 -0.086263  0.170115  0.025290  0.019953 -0.090582  efectores
2   -0.041673  0.064656  0.067765  0.161517  0.032432  0.044532  efectores
3    0.233072 -0.047139 -0.092498  0.303524  0.182702 -0.034421  efectores
4    0.091718  0.308495  0.251691  0.245739  0.214974  0.151067  efectores
..        …         …         …         …         …         …         …
495  0.069273 -0.016226  0.057860 -0.002810  0.058104  0.025849  efectores
496 -0.007123  0.090613  0.154763  0.060774  0.089672 -0.021026  efectores
497 -0.037663  0.058029  0.128414  0.107324  0.030105 -0.001787  efectores
498 -0.002603 -0.024867  0.007672  0.016852 -0.036210 -0.044216  efectores
499 -0.057070 -0.118728 -0.033217  0.136889 -0.045001 -0.015093  efectores

[500 rows x 14 columns]


Covarianza de auto cruzamiento (ACC)  hidro efectores nematoda dataset 3, con
valores atípicos.
Estadísticas.

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

```
mean     0.018561   -0.011022    0.021932    0.017985   -0.003495   -0.000566
std      0.087253    0.090266    0.086915    0.079700    0.081272    0.088639
min     -0.359508   -0.265240   -0.337171   -0.286149   -0.235185   -0.387759
25%     -0.031974   -0.070691   -0.027334   -0.027404   -0.051504   -0.056298
50%      0.013693   -0.011916    0.026794    0.018822   -0.005571   -0.000999
75%      0.064545    0.039527    0.067824    0.062990    0.043409    0.045644
max      0.414262    0.444901    0.394350    0.316122    0.316366    0.354060

               X6          X7          X8          X9         X10         X11  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.020979    0.009744    0.008633    0.012390    0.015126    0.013174
std      0.093179    0.083346    0.084668    0.085585    0.083088    0.077669
min     -0.452320   -0.293683   -0.338523   -0.501900   -0.339056   -0.251446
25%     -0.028603   -0.040096   -0.039382   -0.034906   -0.030969   -0.034673
50%      0.020259    0.013636    0.006426    0.006135    0.012363    0.007858
75%      0.074332    0.058240    0.046991    0.058593    0.060607    0.054881
max      0.304988    0.297161    0.424582    0.336555    0.377558    0.321936

              X12
count  500.000000
mean     0.006906
std      0.079718
min     -0.273013
25%     -0.039533
50%      0.004579
75%      0.052113
max      0.488485
```

no_efectores
Covarianza de auto cruzamiento (ACC) hidro no_efectores nematoda dataset 3, con valores atípicos.
Valores del documento csv.

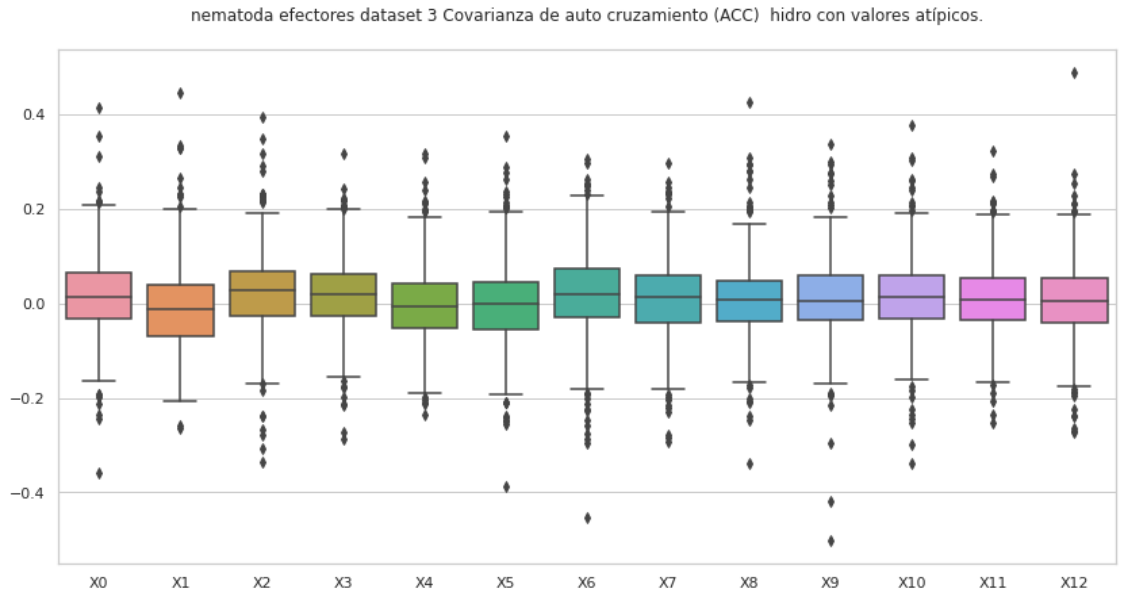
```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.024149 -0.146056  0.032968  0.325623  0.162893 -0.011170  0.115122
1    0.060115 -0.020551  0.023964  0.015334  0.008796 -0.022788  0.031953
2   -0.036624 -0.027357 -0.034069  0.080519 -0.013715  0.041791 -0.008960
3    0.059941 -0.003337  0.014720  0.044888  0.077855  0.003319 -0.014804
4   -0.016187 -0.107803  0.022860 -0.021203 -0.043060  0.065203 -0.056146
..        …         …         …         …         …         …         …
495 -0.022327 -0.128425  0.052709  0.100502 -0.075212 -0.065055  0.027728
496  0.114522  0.135757  0.099504  0.117890  0.117097  0.114322  0.055933
497 -0.064704 -0.060712  0.002162 -0.012883  0.008811 -0.043261 -0.033871
498 -0.032299  0.026906 -0.037384  0.023450 -0.008593 -0.014521  0.147196
499  0.166789  0.102483  0.105533  0.146961  0.043878  0.017471  0.054146
```

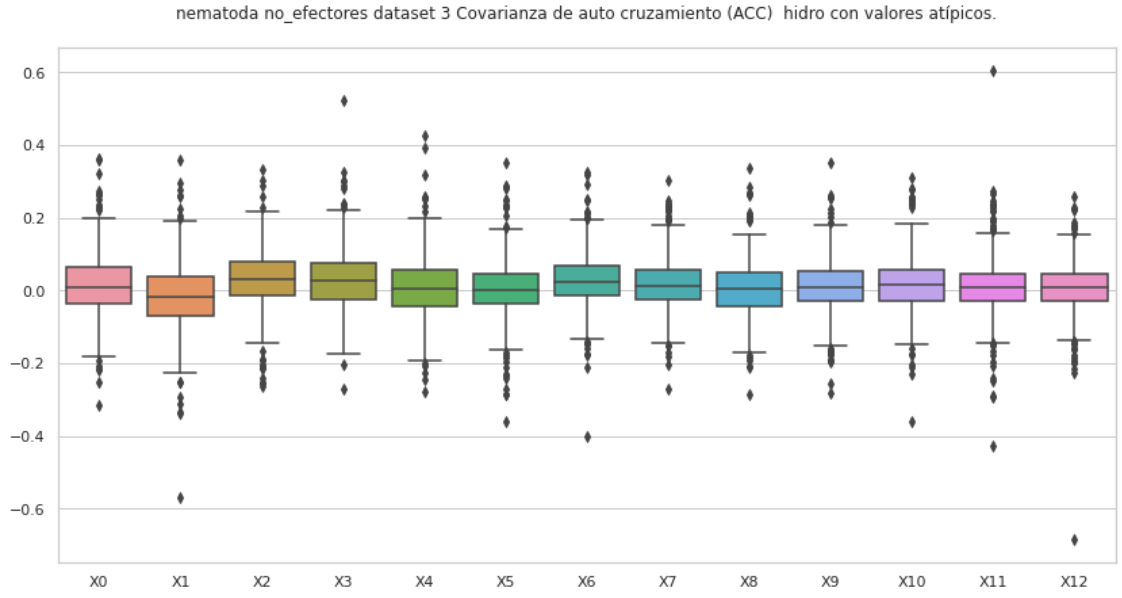
```
           X7        X8        X9       X10       X11       X12           X13
0    0.237448  0.065763 -0.159074  0.185205  0.274330  0.045060  no_efectores
1    0.055365  0.007469 -0.016994 -0.005724 -0.053762  0.078226  no_efectores
2    0.040142  0.047290 -0.090938 -0.070794 -0.009754  0.048348  no_efectores
3   -0.004206  0.146802  0.027236  0.031583  0.084101  0.004933  no_efectores
4    0.013895  0.022633  0.006470  0.048265  0.040571 -0.002068  no_efectores
..        …         …         …         …         …         …             …
495 -0.055039 -0.074413 -0.018605  0.034252  0.059632 -0.024313  no_efectores
496  0.151266  0.129212  0.066404  0.065705  0.071109  0.073378  no_efectores
497  0.031993 -0.043157 -0.007131  0.015809  0.011016  0.011190  no_efectores
498  0.022592  0.069077  0.000445  0.120563 -0.037872 -0.127962  no_efectores
499 -0.090259 -0.003287  0.035221  0.035646 -0.080555 -0.000291  no_efectores

[500 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro no_efectores nematoda dataset 3, con valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.013904   -0.014951    0.030323    0.028310    0.007203    0.005756
std      0.088265    0.094191    0.083394    0.081831    0.086473    0.082243
min     -0.315646   -0.568926   -0.263771   -0.272396   -0.276394   -0.358927
25%     -0.036339   -0.068256   -0.015484   -0.025777   -0.044813   -0.037404
50%      0.009359   -0.018783    0.032421    0.026995    0.005401   -0.000052
75%      0.064478    0.038812    0.077860    0.074816    0.056880    0.046615
max      0.364036    0.358161    0.332514    0.521675    0.425006    0.352649

                X6          X7          X8          X9         X10         X11  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.027201    0.019172    0.003770    0.011406    0.017803    0.010355
std      0.077194    0.073557    0.077206    0.074473    0.076221    0.082838
min     -0.399369   -0.271266   -0.286537   -0.283461   -0.359140   -0.425886
25%     -0.014003   -0.024674   -0.041908   -0.029700   -0.028954   -0.030308
50%      0.024072    0.013987    0.004300    0.009806    0.014619    0.009347
75%      0.069962    0.057655    0.048122    0.054839    0.057182    0.045932
max      0.327096    0.302730    0.338130    0.351968    0.310558    0.604416

               X12
count  500.000000
mean     0.003864
std      0.075805
min     -0.684136
25%     -0.028476
50%      0.008231
75%      0.044163
```

```
max       0.257356
```


nematoda efectores dataset 3 Covarianza de auto cruzamiento (ACC) hidro con valores atípicos.
0.4
0.2
0.0
−0.2
−0.4
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12


nematoda no_efectores dataset 3 Covarianza de auto cruzamiento (ACC) hidro 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

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

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

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

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

    if etiq == "efectores":
        df=ACC_hidro_efec

    if etiq == "no_efectores":
        df=ACC_hidro_no_efec

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

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

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

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

efectores
Covarianza de auto cruzamiento (ACC) efectores nematoda dataset 3, sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	-0.045981	0.098141	0.106728	-0.060690	-0.083583	-0.059933	0.038655	
1	-0.064242	-0.111209	0.002145	-0.092619	-0.108017	0.006212	0.173279	
2	-0.125122	-0.101022	-0.020189	0.041991	-0.120725	0.056036	-0.129436	
5	-0.008230	0.018932	-0.102613	-0.076962	-0.068001	-0.151432	0.064673	
6	-0.069047	-0.005825	-0.027232	-0.079459	-0.100788	-0.173063	0.129590	
..	…	…	…	…	…	…	…	
495	0.022047	-0.025041	0.001914	-0.045175	0.025893	-0.064922	0.115761	
496	0.117091	0.045035	0.060781	0.056875	0.002347	-0.012986	0.000166	
497	0.201939	0.032856	0.129623	0.086441	0.130097	0.160020	0.046970	
498	-0.006665	-0.001886	0.053790	-0.008732	-0.041052	-0.077728	-0.032580	
499	0.034034	-0.127440	0.068490	0.117965	-0.087883	-0.028590	0.181336	

	X7	X8	X9	X10	X11	X12	X13
0	-0.096047	0.041216	0.142697	-0.055478	0.101260	0.019191	efectores
1	-0.050172	-0.086263	0.170115	0.025290	0.019953	-0.090582	efectores
2	-0.041673	0.064656	0.067765	0.161517	0.032432	0.044532	efectores
5	-0.216429	-0.028035	0.093326	-0.034284	0.014136	0.091195	efectores
6	-0.171413	0.162279	0.070822	-0.032743	-0.106873	-0.060606	efectores
..	…	…	…	…	…	…	…
495	0.069273	-0.016226	0.057860	-0.002810	0.058104	0.025849	efectores
496	-0.007123	0.090613	0.154763	0.060774	0.089672	-0.021026	efectores
497	-0.037663	0.058029	0.128414	0.107324	0.030105	-0.001787	efectores
498	-0.002603	-0.024867	0.007672	0.016852	-0.036210	-0.044216	efectores
499	-0.057070	-0.118728	-0.033217	0.136889	-0.045001	-0.015093	efectores

[457 rows x 14 columns]

Covarianza de auto cruzamiento (ACC) efectores nematoda dataset 3, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	457.000000	457.000000	457.000000	457.000000	457.000000	457.000000	
mean	0.017140	-0.014445	0.022072	0.017933	-0.007194	-0.004900	
std	0.075965	0.081194	0.072003	0.069448	0.072706	0.078343	
min	-0.236676	-0.265240	-0.184651	-0.212452	-0.212498	-0.255569	
25%	-0.030107	-0.071110	-0.024122	-0.025031	-0.051151	-0.056015	
50%	0.013374	-0.011850	0.026936	0.019875	-0.008399	-0.004568	
75%	0.060434	0.035241	0.064312	0.061176	0.039894	0.041068	
max	0.237340	0.246523	0.280830	0.218733	0.239859	0.237765	

```
                X6           X7           X8           X9          X10          X11  \
count   457.000000   457.000000   457.000000   457.000000   457.000000   457.000000
mean      0.021597     0.010565     0.004031     0.010811     0.013223     0.010802
std       0.081006     0.073395     0.070379     0.070168     0.070939     0.069544
min      -0.245844    -0.216429    -0.210370    -0.214901    -0.224822    -0.190010
25%      -0.025586    -0.037663    -0.039288    -0.033217    -0.030279    -0.032651
50%       0.020750     0.014195     0.005234     0.005655     0.010360     0.006214
75%       0.069262     0.055376     0.042506     0.054851     0.057413     0.051297
max       0.261944     0.237866     0.244563     0.259437     0.262151     0.217509

               X12
count   457.000000
mean      0.006535
std       0.065457
min      -0.191041
25%      -0.036554
50%       0.004320
75%       0.048409
max       0.227271
```

no_efectores
Covarianza de auto cruzamiento (ACC) no_efectores nematoda dataset 3, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
1    0.060115 -0.020551  0.023964  0.015334  0.008796 -0.022788  0.031953
2   -0.036624 -0.027357 -0.034069  0.080519 -0.013715  0.041791 -0.008960
3    0.059941 -0.003337  0.014720  0.044888  0.077855  0.003319 -0.014804
4   -0.016187 -0.107803  0.022860 -0.021203 -0.043060  0.065203 -0.056146
5    0.069376 -0.158192 -0.114478 -0.202364  0.087203  0.129728 -0.011743
..        …         …         …         …         …         …
495 -0.022327 -0.128425  0.052709  0.100502 -0.075212 -0.065055  0.027728
496  0.114522  0.135757  0.099504  0.117890  0.117097  0.114322  0.055933
497 -0.064704 -0.060712  0.002162 -0.012883  0.008811 -0.043261 -0.033871
498 -0.032299  0.026906 -0.037384  0.023450 -0.008593 -0.014521  0.147196
499  0.166789  0.102483  0.105533  0.146961  0.043878  0.017471  0.054146

           X7        X8        X9       X10       X11       X12           X13
1    0.055365  0.007469 -0.016994 -0.005724 -0.053762  0.078226  no_efectores
2    0.040142  0.047290 -0.090938 -0.070794 -0.009754  0.048348  no_efectores
3   -0.004206  0.146802  0.027236  0.031583  0.084101  0.004933  no_efectores
4    0.013895  0.022633  0.006470  0.048265  0.040571 -0.002068  no_efectores
5   -0.096524  0.011332 -0.008499  0.096552  0.116052 -0.190655  no_efectores
..        …         …         …         …         …         …
495 -0.055039 -0.074413 -0.018605  0.034252  0.059632 -0.024313  no_efectores
```

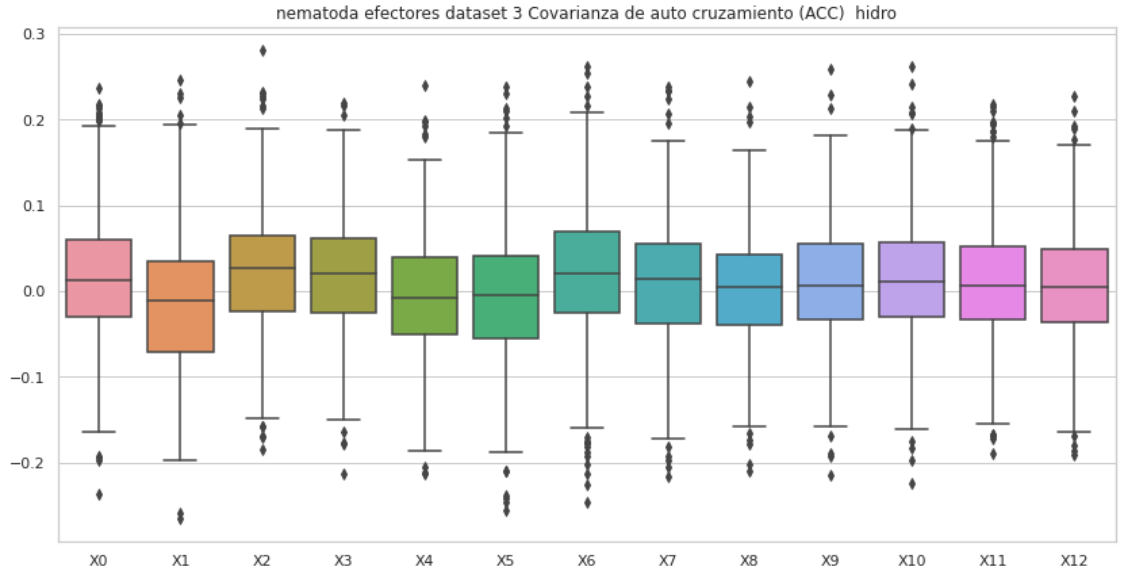
```
496  0.151266  0.129212  0.066404  0.065705  0.071109  0.073378  no_efectores
497  0.031993 -0.043157 -0.007131  0.015809  0.011016  0.011190  no_efectores
498  0.022592  0.069077  0.000445  0.120563 -0.037872 -0.127962  no_efectores
499 -0.090259 -0.003287  0.035221  0.035646 -0.080555 -0.000291  no_efectores

[458 rows x 14 columns]
```

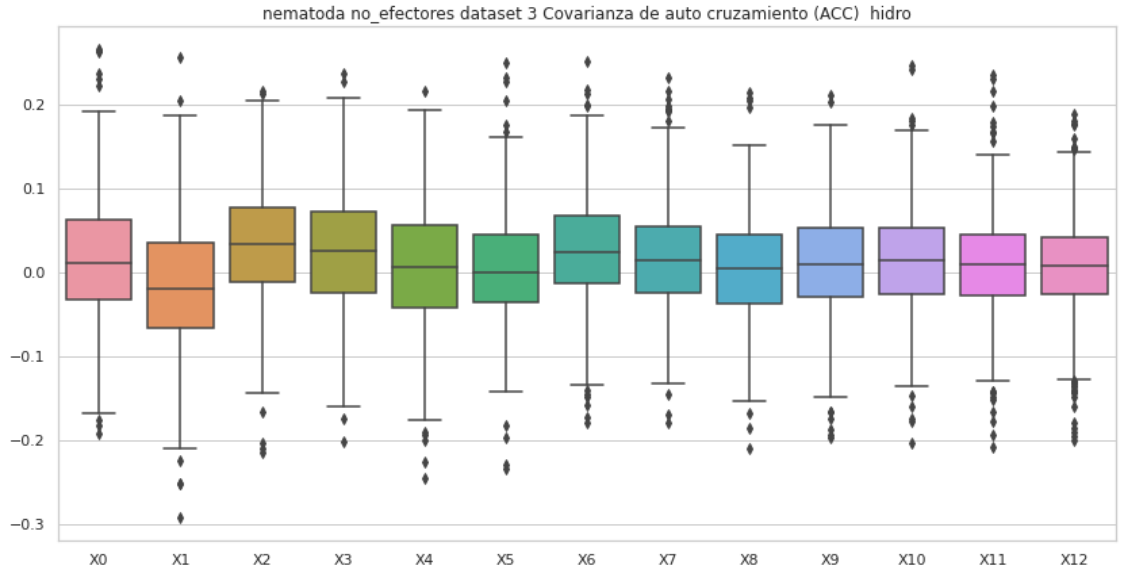
Covarianza de auto cruzamiento (ACC) no_efectores nematoda dataset 3, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	458.000000	458.000000	458.000000	458.000000	458.000000	458.000000	
mean	0.013989	-0.015805	0.032490	0.024814	0.005979	0.006033	
std	0.077244	0.079555	0.071971	0.069109	0.074961	0.066975	
min	-0.192993	-0.292593	-0.215000	-0.202364	-0.245092	-0.234868	
25%	-0.032327	-0.066590	-0.011647	-0.024919	-0.041797	-0.035041	
50%	0.010925	-0.019506	0.033348	0.026202	0.005982	-0.000101	
75%	0.063402	0.035275	0.077529	0.072405	0.055409	0.044595	
max	0.265980	0.256476	0.215412	0.237269	0.216027	0.250189	

	X6	X7	X8	X9	X10	X11	\
count	458.000000	458.000000	458.000000	458.000000	458.000000	458.000000	
mean	0.024818	0.016679	0.003227	0.009509	0.015338	0.009233	
std	0.066147	0.065792	0.066209	0.064447	0.064077	0.064239	
min	-0.179242	-0.179686	-0.210528	-0.197118	-0.204403	-0.207857	
25%	-0.013779	-0.024469	-0.037359	-0.029148	-0.025850	-0.027445	
50%	0.023928	0.013726	0.004681	0.009400	0.014268	0.009676	
75%	0.067678	0.055036	0.045157	0.052961	0.052396	0.044120	
max	0.251842	0.232048	0.215061	0.211827	0.246127	0.235198	

	X12
count	458.000000
mean	0.004896
std	0.062749
min	-0.200919
25%	-0.026738
50%	0.007605
75%	0.041879
max	0.189026


nematoda efectores dataset 3 Covarianza de auto cruzamiento (ACC) hidro
0.3
0.2
0.1
0.0
−0.1
−0.2
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12


nematoda no_efectores dataset 3 Covarianza de auto cruzamiento (ACC) hidro
0.2
0.1
0.0
−0.1
−0.2
−0.3
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12