ds3_nematoda_limpieza_de_datos

February 1, 2021

Limpieza de datos

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

1 Declaración de variables

[2]:
```
organismo ="nematoda"
dataset = 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     6.158  5.083   4.399  7.234  1.466  8.798  4.985  2.639  2.053   5.767
1     6.406  8.541   2.491  4.270  1.423  4.982  1.779  5.338  2.847   9.609
2     6.435  6.178   5.920  5.148  1.544  7.979  5.534  4.891  3.089   6.306
3     4.255  9.574   4.255  6.383  0.000  9.574  5.319  3.191  0.000   5.319
4     3.200  4.800  10.400  3.200  4.000  4.800  6.400  5.600  1.600  11.200
..      …      …      …      …      …      …      …      …      …       …
995  10.236  3.150   0.787  1.575  3.937  7.087  4.724  6.299  3.937   5.512
996   2.041  9.184   5.102  4.082  6.122  5.102  1.020  3.061  1.020   9.184
997   5.966  3.835   6.108  3.267  1.136  4.688  7.244  8.239  3.835   5.398
998   5.389  3.593   2.395  4.790  7.186  3.593  3.593  6.587  2.395   3.593
999   4.324  7.027   2.703  6.486  2.162  4.865  4.865  4.324  2.162   5.405
```

```
     …    X11    X12    X13    X14     X15    X16    X17    X18     X19  \
0    …  8.895  2.346  3.324  2.835  10.557  3.519  0.684  2.835   5.376
1    …  4.626  2.491  6.050  3.559   3.915  5.338  2.135  5.694   7.473
2    …  5.663  4.118  5.277  2.059   7.207  4.762  1.030  2.317   6.049
3    …  6.383  1.064  4.255  3.191  14.894  2.128  3.191  2.128  10.638
4    …  4.800  4.000  6.400  2.400   8.000  4.000  0.800  4.000   3.200
..   …    …    …    …    …     …    …    …    …     …
995  …  2.362  2.362  3.937  5.512  10.236  8.661  1.575  1.575   8.661
996  …  2.041  0.000  6.122  4.082  13.265  1.020  4.082  5.102   8.163
997  …  4.688  4.261  3.835  6.818   8.097  4.830  0.426  3.267   7.528
998  …  5.988  1.796  4.192  5.988  10.778  5.988  1.796  4.192   9.581
999  …  4.324  2.162  3.784  6.486  10.811  4.865  0.541  2.162   8.108

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

[1000 rows x 21 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      7.037242     6.094491     4.293374     5.295819     2.349846
std       2.901479     2.642230     1.923025     2.334473     2.078743
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       5.318250     4.456500     3.096750     3.846000     1.157750
50%       6.663500     5.882000     4.172500     5.183500     1.954500
75%       8.488750     7.563000     5.263000     6.548000     2.968500
max      27.830000    21.951000    17.021000    29.787000    29.762000

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      6.414085     3.923620     5.850920     2.319747     5.612410
std       3.318146     2.021391     3.148009     1.523753     2.388701
```

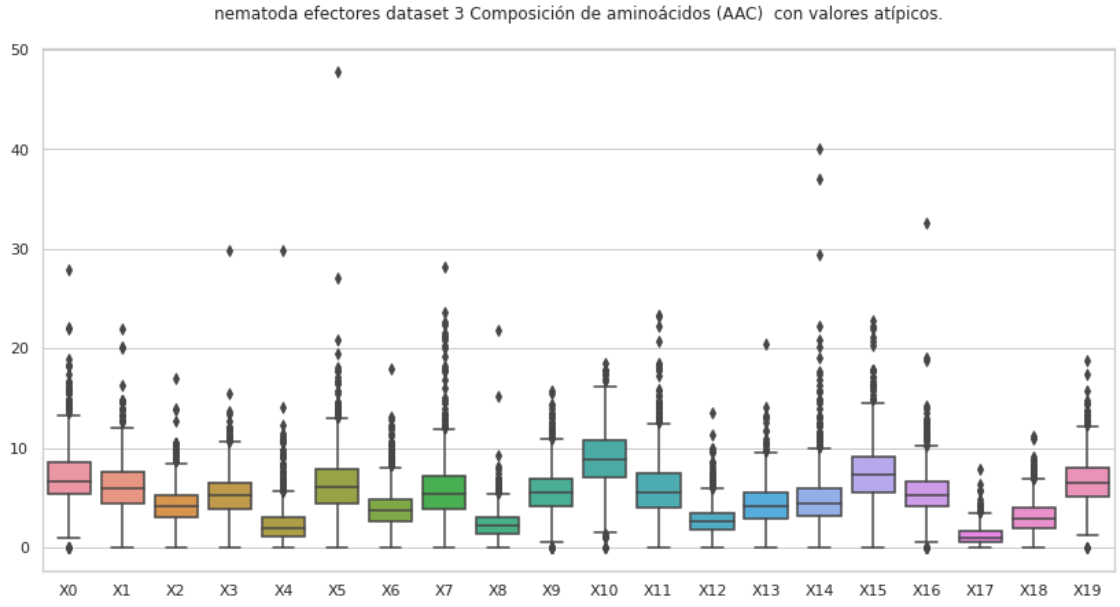
```
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       4.443250     2.594250     3.895250     1.429000     4.109000
50%       6.134000     3.704000     5.428500     2.195500     5.464000
75%       7.877500     4.832750     7.118250     3.049000     6.832250
max      47.727000    18.000000    28.125000    21.739000    15.686000

               X10          X11          X12          X13          X14  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      8.856904     5.997663     2.786084     4.347575     4.924925
std       2.972290     3.085017     1.502186     2.241033     3.183913
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       7.021000     4.030750     1.821750     2.857000     3.204000
50%       8.819500     5.577500     2.604500     4.167000     4.465000
75%      10.732750     7.432750     3.464250     5.556000     5.912250
max      18.421000    23.313000    13.559000    20.436000    40.000000

               X15          X16          X17          X18          X19
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      7.540928     5.498052     1.187814     3.043303     6.625255
std       3.074833     2.395614     1.013934     1.642589     2.311302
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       5.485750     4.102500     0.488250     1.969000     5.112750
50%       7.246000     5.242000     0.990000     2.920000     6.496000
75%       9.160750     6.558250     1.681750     3.947000     7.948250
max      22.787000    32.609000     7.843000    11.111000    18.750000



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     7.857  5.714  5.000  5.000   0.714  5.000  3.571   5.000  2.857   7.143
1    10.526  4.276  2.961  3.289   1.316  3.618  2.632   3.618  1.645   9.539
2     4.839  8.065  4.839  6.452   1.613  9.677  6.452   0.000  0.000   4.839
3     5.785  3.306  1.653  1.653   4.959  1.653  5.785  18.182  0.000   4.132
4     6.952  6.061  3.922  4.991   1.070  6.061  4.813   4.991  3.209   7.308
..      …      …      …      …       …      …      …       …      …       …
995   4.863  3.875  9.195  3.495   2.356  6.231  5.547   4.179  1.824   7.827
996   4.310  1.724  6.034  1.724   5.172  1.724  4.310   2.586  0.862  10.345
997   8.000  1.000  0.000  1.000  10.000  0.000  2.000   6.000  1.000   4.000
998   3.750  6.250  1.875  5.000   1.875  7.500  1.875   6.875  7.500   5.000
999   1.538  0.000  9.231  3.077   1.538  6.154  6.154   3.077  0.000  15.385

     …    X11    X12     X13     X14     X15    X16    X17    X18     X19  \
0    …  6.429  1.429   4.286   4.286   7.857  6.429  0.000  2.143   7.143
```

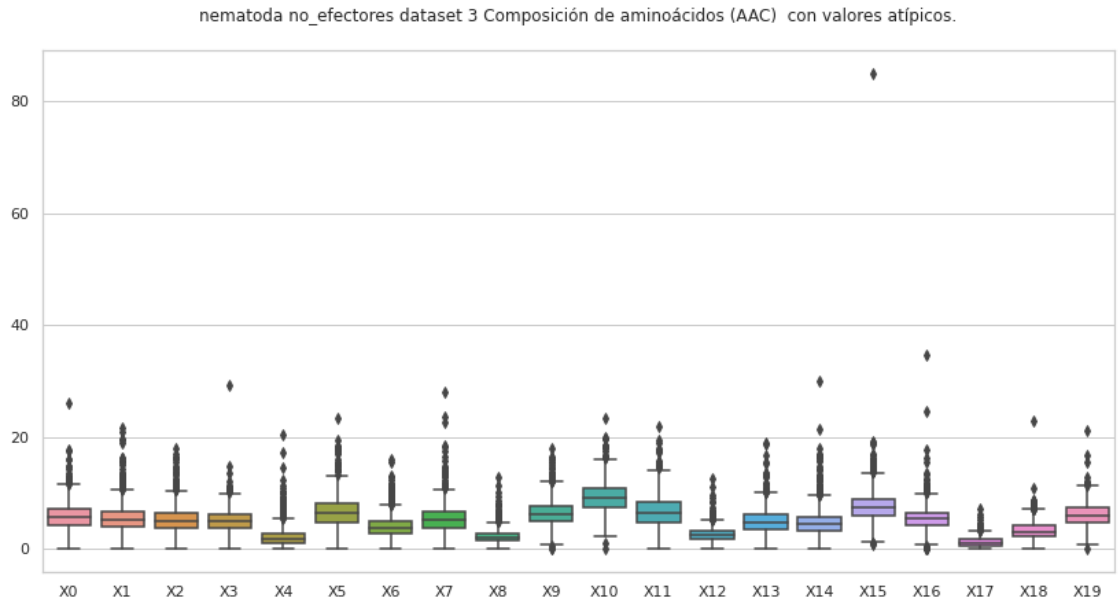
```
1    …  2.632  4.934   8.553   2.961   6.908  4.934  0.987  5.592   8.553
2    …  8.065  3.226   0.000   6.452  12.903  4.839  1.613  1.613   6.452
3    …  2.479  8.264   1.653   6.612  14.050  6.612  0.826  4.132   1.653
4    …  5.348  1.961   3.387   6.061   7.843  3.565  2.139  3.030   5.704
..   …    …    …     …     …      …      …    …      …      …
995  …  6.535  1.140   5.927   2.660   8.891  6.383  1.216  2.660   3.799
996  …  6.897  2.586  12.931   2.586   2.586  1.724  0.000  2.586   6.034
997  …  1.000  5.000   5.000  30.000   4.000  5.000  0.000  0.000  12.000
998  …  9.375  2.500   3.750   4.375   8.125  6.875  3.125  3.750   2.500
999  …  7.692  6.154   6.154   4.615   7.692  3.077  0.000  7.692   1.538

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

[1000 rows x 21 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      5.859227     5.470492     5.308098     5.013980     2.247808
std       2.539876     2.686948     2.458007     2.101092     1.910416
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       4.265000     3.846000     3.784750     3.720750     1.087000
50%       5.704500     5.152500     5.000000     5.011000     1.841500
75%       7.177250     6.570250     6.452000     6.215250     2.817000
max      26.027000    21.495000    17.910000    29.104000    20.280000

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      6.576440     3.926137     5.380749     2.242613     6.405793
std       3.014497     1.920454     2.660380     1.335953     2.526401
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       4.632000     2.731500     3.771250     1.464250     4.813250
50%       6.495000     3.682500     5.078500     2.083000     6.140500
```

```
75%       8.108000     4.809000     6.521500     2.797000     7.711750
max      23.377000    15.948000    28.090000    12.658000    17.857000

                X10          X11          X12          X13          X14  \
count   1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean       9.110186     6.689107     2.548199     4.980177     4.707768
std        2.852802     3.013867     1.340516     2.251895     2.539456
min        0.000000     0.000000     0.000000     0.000000     0.000000
25%        7.293500     4.555500     1.674250     3.509000     3.209250
50%        9.062000     6.390500     2.389500     4.665000     4.376000
75%       10.787500     8.436500     3.094000     6.099250     5.721000
max       23.276000    21.930000    12.500000    18.868000    30.000000

                X15          X16          X17          X18          X19
count   1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean       7.663172     5.462035     1.168077     3.206100     6.033930
std        3.663443     2.265946     0.911494     1.724189     2.100048
min        0.830000     0.000000     0.000000     0.000000     0.000000
25%        5.879750     4.167000     0.545000     2.128000     4.664500
50%        7.339000     5.290500     1.010000     3.000500     5.980000
75%        8.958500     6.454250     1.641250     4.095750     7.309250
max       84.906000    34.653000     7.018000    22.785000    21.127000
```





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     6.158  5.083  4.399  7.234  1.466   8.798  4.985  2.639  2.053  5.767
1     6.406  8.541  2.491  4.270  1.423   4.982  1.779  5.338  2.847  9.609
2     6.435  6.178  5.920  5.148  1.544   7.979  5.534  4.891  3.089  6.306
3     4.255  9.574  4.255  6.383  0.000   9.574  5.319  3.191  0.000  5.319
5     9.050  9.050  2.715  2.262  1.810  12.217  4.977  2.715  3.167  6.787
..      …      …      …      …      …       …      …      …      …      …
995  10.236  3.150  0.787  1.575  3.937   7.087  4.724  6.299  3.937  5.512
996   2.041  9.184  5.102  4.082  6.122   5.102  1.020  3.061  1.020  9.184
997   5.966  3.835  6.108  3.267  1.136   4.688  7.244  8.239  3.835  5.398
998   5.389  3.593  2.395  4.790  7.186   3.593  3.593  6.587  2.395  3.593
999   4.324  7.027  2.703  6.486  2.162   4.865  4.865  4.324  2.162  5.405

     …    X11    X12    X13    X14     X15    X16    X17    X18     X19  \
0    …  8.895  2.346  3.324  2.835  10.557  3.519  0.684  2.835   5.376
1    …  4.626  2.491  6.050  3.559   3.915  5.338  2.135  5.694   7.473
2    …  5.663  4.118  5.277  2.059   7.207  4.762  1.030  2.317   6.049
3    …  6.383  1.064  4.255  3.191  14.894  2.128  3.191  2.128  10.638
5    …  6.335  3.620  5.430  1.357   6.787  3.167  0.452  2.715   3.620
..   …    …      …      …      …       …      …      …      …       …
995  …  2.362  2.362  3.937  5.512  10.236  8.661  1.575  1.575   8.661
```

```
996  …  2.041  0.000  6.122  4.082  13.265  1.020  4.082  5.102   8.163
997  …  4.688  4.261  3.835  6.818   8.097  4.830  0.426  3.267   7.528
998  …  5.988  1.796  4.192  5.988  10.778  5.988  1.796  4.192   9.581
999  …  4.324  2.162  3.784  6.486  10.811  4.865  0.541  2.162   8.108

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

[837 rows x 21 columns]
```

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

```
                X0          X1          X2          X3          X4          X5  \
count  837.000000  837.000000  837.000000  837.000000  837.000000  837.000000
mean     6.904681    6.109865    4.317993    5.421723    2.203542    6.447545
std      2.376814    2.230210    1.624458    2.005618    1.488940    2.667070
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      5.376000    4.671000    3.218000    4.104000    1.213000    4.592000
50%      6.660000    5.980000    4.296000    5.385000    1.961000    6.256000
75%      8.333000    7.491000    5.263000    6.593000    2.878000    7.906000
max     15.500000   13.889000    9.890000   11.976000    8.403000   15.663000

                X6          X7          X8          X9         X10         X11  \
count  837.000000  837.000000  837.000000  837.000000  837.000000  837.000000
mean     3.846622    5.584763    2.379225    5.726935    9.200393    5.990250
std      1.721915    2.310498    1.216204    2.063740    2.681604    2.551705
min      0.000000    0.000000    0.000000    0.515000    1.905000    0.758000
25%      2.655000    3.974000    1.547000    4.369000    7.356000    4.271000
50%      3.667000    5.388000    2.299000    5.590000    9.091000    5.735000
75%      4.737000    6.818000    3.089000    6.857000   10.948000    7.273000
max      9.717000   14.706000    6.667000   12.583000   17.742000   14.458000

               X12         X13         X14         X15         X16         X17  \
count  837.000000  837.000000  837.000000  837.000000  837.000000  837.000000
mean     2.721393    4.424142    4.697296    7.531799    5.496897    1.192302
```

std	1.259906	1.923734	2.119481	2.622479	2.022046	0.898894
min	0.000000	0.000000	0.000000	1.010000	0.000000	0.000000
25%	1.852000	3.125000	3.356000	5.769000	4.235000	0.570000
50%	2.597000	4.242000	4.478000	7.328000	5.338000	1.044000
75%	3.390000	5.556000	5.830000	9.111000	6.529000	1.688000
max	7.285000	10.687000	14.000000	16.667000	12.676000	4.225000

	X18	X19
count	837.000000	837.000000
mean	3.098026	6.704642
std	1.441613	2.109892
min	0.000000	0.000000
25%	2.174000	5.314000
50%	3.017000	6.614000
75%	3.939000	8.000000
max	7.937000	13.415000

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	7.857	5.714	5.000	5.000	0.714	5.000	3.571	5.000	2.857	7.143	
1	10.526	4.276	2.961	3.289	1.316	3.618	2.632	3.618	1.645	9.539	
2	4.839	8.065	4.839	6.452	1.613	9.677	6.452	0.000	0.000	4.839	
4	6.952	6.061	3.922	4.991	1.070	6.061	4.813	4.991	3.209	7.308	
5	7.317	5.066	4.128	5.441	0.750	7.880	2.064	4.128	2.251	5.253	
..	…	…	…	…	…	…	…	…	…	…	
991	4.132	3.719	7.438	7.438	3.719	6.198	2.893	4.132	3.719	5.785	
992	7.778	10.000	3.333	4.444	1.111	5.556	7.778	6.667	2.222	3.333	
993	7.538	6.533	3.518	8.040	2.513	5.025	1.508	6.030	3.518	4.020	
994	2.996	4.869	8.240	4.120	0.000	10.487	0.749	6.367	0.749	9.738	
995	4.863	3.875	9.195	3.495	2.356	6.231	5.547	4.179	1.824	7.827	

	…	X11	X12	X13	X14	X15	X16	X17	X18	X19	\
0	…	6.429	1.429	4.286	4.286	7.857	6.429	0.000	2.143	7.143	
1	…	2.632	4.934	8.553	2.961	6.908	4.934	0.987	5.592	8.553	
2	…	8.065	3.226	0.000	6.452	12.903	4.839	1.613	1.613	6.452	
4	…	5.348	1.961	3.387	6.061	7.843	3.565	2.139	3.030	5.704	
5	…	10.882	3.002	3.189	6.942	8.443	4.128	1.313	3.002	6.754	
..	…	…	…	…	…	…	…	…	…		
991	…	4.132	2.066	4.959	3.719	7.438	4.545	1.653	1.240	5.785	
992	…	4.444	1.111	3.333	0.000	8.889	10.000	2.222	1.111	7.778	
993	…	6.030	1.508	4.523	3.518	7.538	7.538	0.503	2.010	11.558	
994	…	7.865	0.749	5.618	4.494	7.116	5.993	0.375	3.745	5.993	

```
995  …   6.535  1.140  5.927  2.660   8.891   6.383  1.216  2.660   3.799

              X20
0     no_efectores
1     no_efectores
2     no_efectores
4     no_efectores
5     no_efectores
..             …
991   no_efectores
992   no_efectores
993   no_efectores
994   no_efectores
995   no_efectores

[834 rows x 21 columns]
```

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

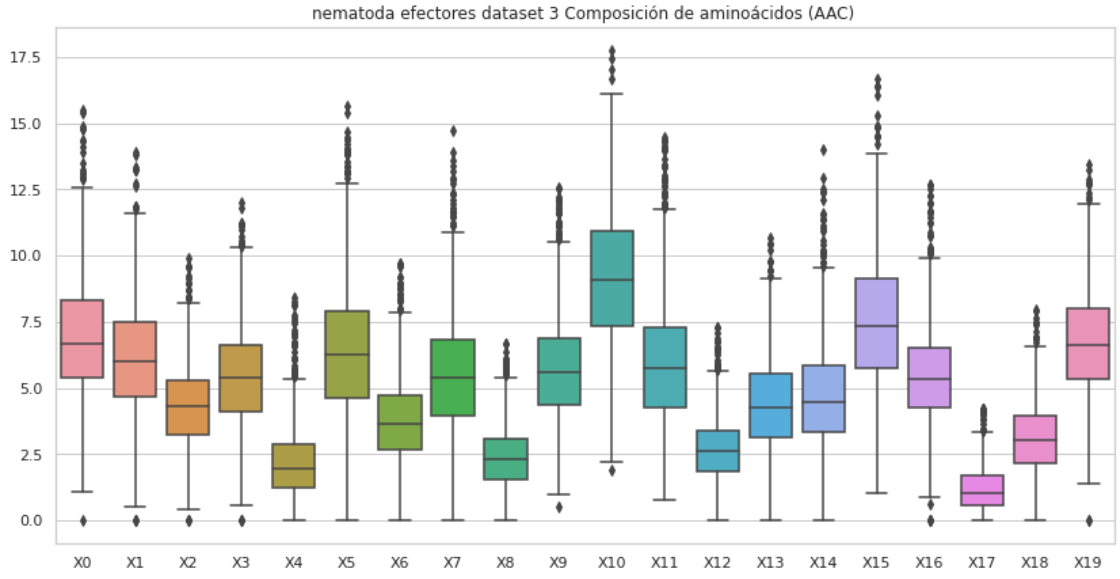
	X0	X1	X2	X3	X4	X5	\
count	834.000000	834.000000	834.000000	834.000000	834.000000	834.000000	
mean	5.916528	5.460528	5.253360	5.188987	2.077936	6.698145	
std	2.175797	2.202242	2.078724	1.768496	1.350442	2.596553	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	4.483250	3.944000	3.874250	4.101000	1.149250	5.000000	
50%	5.794500	5.272500	4.992000	5.263000	1.834500	6.690500	
75%	7.176500	6.547750	6.410000	6.303250	2.686500	8.182000	
max	13.043000	13.450000	12.626000	11.029000	7.895000	15.493000	

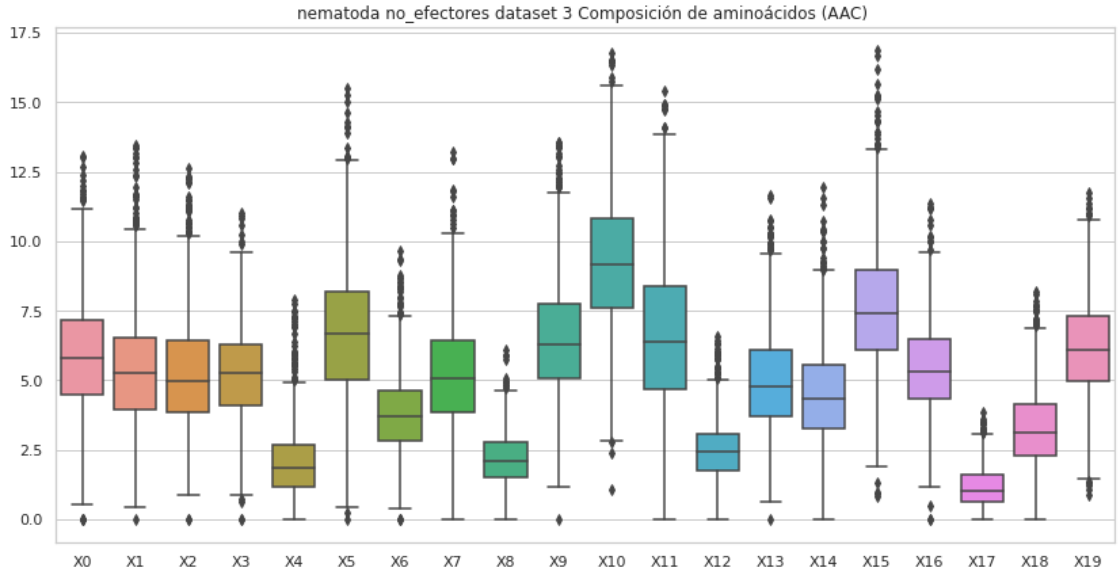
	X6	X7	X8	X9	X10	X11	\
count	834.000000	834.000000	834.000000	834.000000	834.000000	834.000000	
mean	3.838223	5.214043	2.193339	6.510498	9.275797	6.662221	
std	1.513221	2.044830	1.017575	2.230285	2.484832	2.700217	
min	0.000000	0.000000	0.000000	0.000000	1.053000	0.000000	
25%	2.836750	3.846000	1.505250	5.075250	7.620000	4.656000	
50%	3.697000	5.064500	2.118000	6.263500	9.183500	6.400000	
75%	4.640500	6.424250	2.778000	7.773250	10.840500	8.378000	
max	9.659000	13.208000	6.107000	13.571000	16.774000	15.419000	

	X12	X13	X14	X15	X16	X17	\
count	834.000000	834.000000	834.000000	834.000000	834.000000	834.000000	
mean	2.521470	5.009463	4.517198	7.636287	5.450000	1.158139	
std	1.105295	1.881566	1.821912	2.508849	1.712462	0.771714	
min	0.000000	0.000000	0.000000	0.830000	0.000000	0.000000	
25%	1.733000	3.714500	3.281500	6.074000	4.320500	0.635250	

50%	2.441500	4.767500	4.358500	7.426500	5.318000	1.047000
75%	3.068750	6.090750	5.556000	8.969750	6.477000	1.625750
max	6.566000	11.659000	11.940000	16.859000	11.364000	3.846000

	X18	X19
count	834.000000	834.000000
mean	3.287482	6.130394
std	1.480259	1.860922
min	0.000000	0.893000
25%	2.273000	4.979000
50%	3.096000	6.072500
75%	4.152000	7.322750
max	8.176000	11.765000





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    0.023907  0.005692  0.028081  0.034152  0.012902  0.010246  0.007969
1    0.019405  0.004312  0.012936  0.015093  0.018327  0.016171  0.008624
2    0.040165  0.009640  0.032132  0.049805  0.032935  0.030526  0.019279
3    0.028460  0.000000  0.042691  0.064036  0.028460  0.021345  0.000000
4    0.032758  0.040948  0.032758  0.049137  0.065516  0.057327  0.016379
..        …         …         …         …         …         …         …
995  0.046365  0.017833  0.007133  0.032099  0.017833  0.028532  0.017833
996  0.028307  0.084920  0.056613  0.070767  0.084920  0.042460  0.014153
997  0.027903  0.005315  0.015280  0.021924  0.017938  0.038533  0.017938
998  0.029072  0.038763  0.025842  0.019382  0.022612  0.035533  0.012921
999  0.028896  0.014448  0.043344  0.032508  0.025284  0.028896  0.014448

          X7        X8        X9  …       X74       X75       X76  \
0    0.022389  0.034532  0.042880  …  0.003755  0.024647  0.001884
1    0.029107  0.014015  0.033419  …  0.019313  0.012042 -0.017673
2    0.039362  0.035345  0.053018  … -0.002341 -0.005898  0.002718
3    0.035575  0.042691  0.028460  … -0.121403  0.004466 -0.040339
4    0.114653  0.049137  0.073706  …  0.154405  0.133459 -0.047994
..        …         …         …  …         …         …         …
995  0.024966  0.010700  0.035665  … -0.000649  0.011690  0.003123
996  0.127380  0.028307  0.141533  … -0.182459 -0.142556  0.104872
997  0.025246  0.021924  0.030561  … -0.002502  0.008896  0.029719
998  0.019382  0.032303  0.035533  …  0.036424  0.002265  0.060961
999  0.036120  0.028896  0.083076  … -0.014320 -0.016939  0.007312

          X77       X78       X79       X80       X81       X82         X83
0   -0.004195  0.017409 -0.001178  0.005691  0.023933 -0.002358  efectores
1    0.003517 -0.004335 -0.010706  0.008729 -0.005049  0.009505  efectores
2   -0.013452 -0.012503  0.011419  0.024259  0.038549  0.011367  efectores
3    0.061751  0.039764 -0.036660  0.002825  0.062832 -0.083430  efectores
4   -0.036092 -0.053136 -0.019673  0.027115  0.055705  0.022773  efectores
..        …         …         …         …         …         …         …
995  0.036822  0.001191  0.024378 -0.016153 -0.031573  0.010834  efectores
996  0.116104 -0.058696 -0.069131  0.044660 -0.074274  0.071741  efectores
997  0.024624  0.012464  0.028572  0.006242  0.003222  0.031870  efectores
```

```
998  0.039246  0.002981  0.044399 -0.009072 -0.034714  0.032755  efectores
999  0.018541  0.005545  0.014848 -0.017802  0.025070  0.005581  efectores

[1000 rows x 84 columns]
```

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

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.035777	0.012930	0.028466	0.032511	0.023776	
std	0.038616	0.023188	0.032316	0.047128	0.062447	
min	-0.889309	-0.444655	-0.594386	-0.891579	-1.778618	
25%	0.023136	0.004690	0.014821	0.018158	0.011242	
50%	0.032569	0.009643	0.026177	0.030215	0.020487	
75%	0.046120	0.017683	0.039192	0.045167	0.033321	
max	0.286369	0.173500	0.212718	0.157582	0.223504	

	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.028198	0.013172	0.031695	0.031629	0.048524	…	
std	0.045562	0.024199	0.036843	0.054673	0.075182	…	
min	-0.891579	-0.594386	-0.594386	-1.485965	-1.783158	…	
25%	0.017082	0.005380	0.016067	0.016468	0.027464	…	
50%	0.026268	0.010798	0.027071	0.027800	0.043751	…	
75%	0.038670	0.017834	0.040870	0.043488	0.064804	…	
max	0.223504	0.212672	0.265898	0.233622	0.381826	…	

	X73	X74	X75	X76	X77	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.014439	-0.001742	0.003930	0.005677	0.003068	
std	0.100724	0.108150	0.100921	0.154845	0.064726	
min	-0.378442	-2.967414	-2.350050	-4.774132	-0.409558	
25%	-0.000020	-0.011075	-0.006319	-0.000062	-0.010636	
50%	0.011132	0.002850	0.007031	0.011507	0.003363	
75%	0.022762	0.015261	0.019123	0.022514	0.016465	
max	3.054546	1.166241	1.846453	0.192782	1.547349	

	X78	X79	X80	X81	X82
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.007530	0.010622	0.004875	0.011006	0.011474
std	0.048875	0.074498	0.122013	0.115472	0.094333
min	-0.374317	-0.316934	-0.652667	-0.410520	-1.543584
25%	-0.004358	-0.001149	-0.010724	-0.004857	0.000185
50%	0.006754	0.010226	0.003168	0.006792	0.011685
75%	0.019256	0.022129	0.015271	0.021086	0.022480

```
max       0.802747     2.166014     2.577534     3.283625     2.411079

[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    0.039380  0.003580  0.025060  0.025060  0.021480  0.025060  0.014320
1    0.042482  0.005310  0.013276  0.014603  0.034517  0.014603  0.006638
2    0.017606  0.005869  0.023474  0.035211  0.000000  0.000000  0.000000
3    0.013293  0.011394  0.003798  0.003798  0.003798  0.041777  0.000000
4    0.049698  0.007646  0.035681  0.043326  0.024212  0.035681  0.022937
..        …         …         …         …         …         …
995  0.030917  0.014975  0.022221  0.039612  0.037680  0.026569  0.011594
996  0.010325  0.012390  0.004130  0.004130  0.030974  0.006195  0.002065
997  0.011116  0.013894  0.001389  0.000000  0.006947  0.008337  0.001389
998  0.019648  0.009824  0.026198  0.039297  0.019648  0.036022  0.039297
999  0.011346  0.011346  0.022692  0.045384  0.045384  0.022692  0.000000

           X7        X8        X9  …       X74       X75       X76  \
0    0.035800  0.032220  0.060860  …  0.009192 -0.010200  0.021438
1    0.038499  0.010620  0.042482  … -0.000093 -0.004417  0.006653
2    0.017606  0.029343  0.029343  …  0.077400  0.090879  0.026966
3    0.009495  0.005697  0.015192  … -0.005021 -0.002275  0.015934
4    0.052246  0.038229  0.082830  … -0.016260 -0.016881  0.025643
..        …         …         …  …       …         …         …
995  0.049757  0.041545  0.072461  … -0.002133 -0.002341  0.002694
996  0.024779  0.016519  0.055753  …  0.049524  0.015590  0.000224
997  0.005558  0.001389  0.006947  …  0.013450  0.004184  0.029319
998  0.026198  0.049121  0.042571  … -0.043459 -0.013213 -0.001455
999  0.113460  0.056730  0.068076  … -0.011940 -0.000427  0.007857

           X77       X78       X79       X80       X81       X82           X83
0   -0.019998  0.012164 -0.007450  0.002467 -0.002881  0.008534  no_efectores
1    0.021220 -0.003873  0.008060  0.026096  0.005420 -0.008731  no_efectores
2   -0.047855 -0.023920  0.001745 -0.017778 -0.018997 -0.015167  no_efectores
3   -0.003046  0.001546  0.049661 -0.008175 -0.003793  0.016891  no_efectores
4   -0.005908 -0.007224  0.017529  0.004350  0.020206  0.019026  no_efectores
..        …         …         …         …         …         …
995  0.001019 -0.001565  0.015322 -0.002850  0.000128  0.010163  no_efectores
996  0.039809  0.025036  0.004213  0.024824  0.013738  0.003775  no_efectores
997  0.017308  0.006629  0.029622  0.016309  0.004395  0.027184  no_efectores
998  0.038729  0.041624  0.007020  0.019968  0.027155 -0.026545  no_efectores
```

```
999  0.003412 -0.077809 -0.032986  0.024204 -0.005873 -0.029974  no_efectores

[1000 rows x 84 columns]
```

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

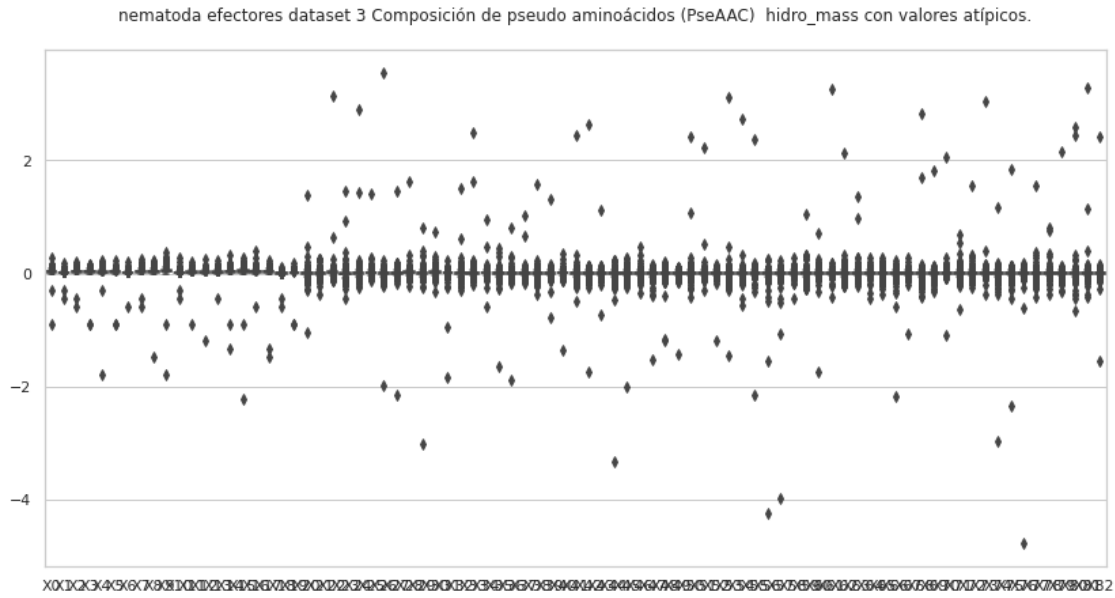
	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.031730	0.015235	0.028750	0.036826	0.029560	
std	0.047034	0.089300	0.047078	0.048921	0.049201	
min	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.018880	0.005052	0.014540	0.019183	0.014559	
50%	0.027297	0.008952	0.024011	0.031855	0.022956	
75%	0.038687	0.016009	0.037356	0.046826	0.035885	
max	1.403405	2.806811	1.403405	1.403405	1.403405	

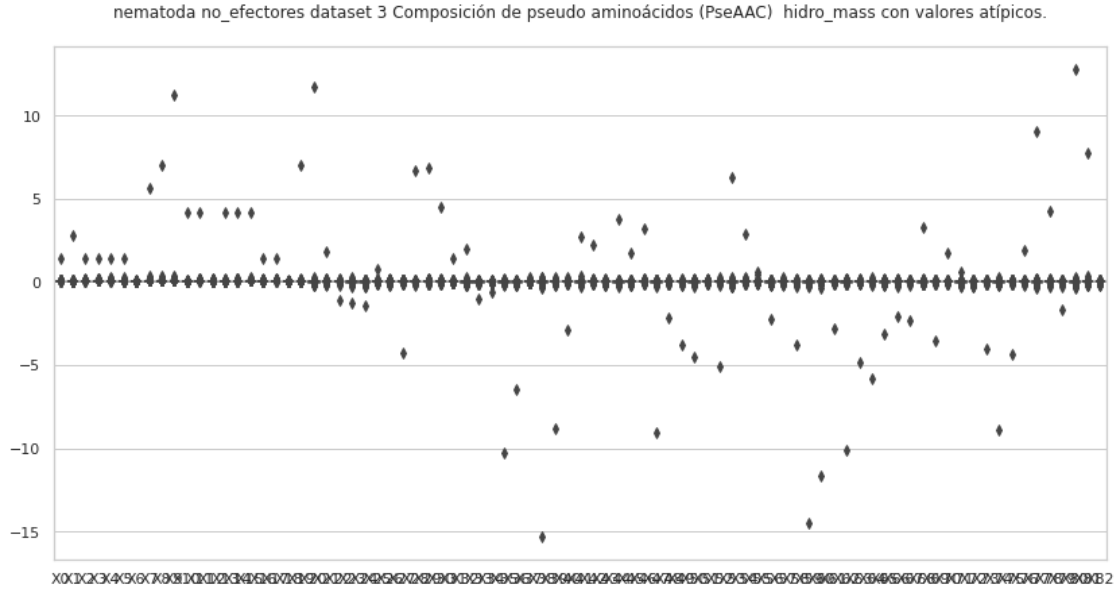
	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.029136	0.012673	0.041364	0.043354	0.061470	…	
std	0.046934	0.010902	0.178480	0.222345	0.354988	…	
min	0.000000	0.000000	0.000000	0.000000	0.000000	…	
25%	0.016452	0.005595	0.019192	0.020523	0.029631	…	
50%	0.024747	0.009705	0.030769	0.031101	0.043262	…	
75%	0.035228	0.016826	0.044719	0.047376	0.063833	…	
max	1.403405	0.102333	5.613621	7.017026	11.227242	…	

	X73	X74	X75	X76	X77	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.004052	-0.005998	0.003776	0.009974	0.010355	
std	0.128666	0.283167	0.139843	0.062849	0.287495	
min	-3.993520	-8.894949	-4.339521	-0.167551	-0.363486	
25%	-0.001050	-0.008644	-0.002887	-0.000349	-0.008329	
50%	0.008030	0.003108	0.007986	0.008493	0.003715	
75%	0.018762	0.014404	0.018860	0.018568	0.014208	
max	0.124968	0.281828	0.147650	1.871067	9.031102	

	X78	X79	X80	X81	X82
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.011145	0.006496	0.014577	0.015068	0.008406
std	0.135867	0.057848	0.405459	0.246383	0.023463
min	-0.262553	-1.680335	-0.368189	-0.214888	-0.226852
25%	-0.003147	-0.000729	-0.008201	-0.002787	-0.000360
50%	0.007384	0.007785	0.004031	0.007747	0.008767
75%	0.019114	0.017940	0.015055	0.018876	0.018396
max	4.219770	0.124551	12.778342	7.742278	0.138950

[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	0.023907	0.005692	0.028081	0.034152	0.012902	0.010246	0.007969	
1	0.019405	0.004312	0.012936	0.015093	0.018327	0.016171	0.008624	
2	0.040165	0.009640	0.032132	0.049805	0.032935	0.030526	0.019279	
3	0.028460	0.000000	0.042691	0.064036	0.028460	0.021345	0.000000	
4	0.032758	0.040948	0.032758	0.049137	0.065516	0.057327	0.016379	
..	…	…	…	…	…	…	…	
994	0.047835	0.015945	0.039862	0.043848	0.047835	0.059793	0.023917	
995	0.046365	0.017833	0.007133	0.032099	0.017833	0.028532	0.017833	
997	0.027903	0.005315	0.015280	0.021924	0.017938	0.038533	0.017938	
998	0.029072	0.038763	0.025842	0.019382	0.022612	0.035533	0.012921	
999	0.028896	0.014448	0.043344	0.032508	0.025284	0.028896	0.014448	

	X7	X8	X9	…	X74	X75	X76	\
0	0.022389	0.034532	0.042880	…	0.003755	0.024647	0.001884	
1	0.029107	0.014015	0.033419	…	0.019313	0.012042	-0.017673	
2	0.039362	0.035345	0.053018	…	-0.002341	-0.005898	0.002718	
3	0.035575	0.042691	0.028460	…	-0.121403	0.004466	-0.040339	
4	0.114653	0.049137	0.073706	…	0.154405	0.133459	-0.047994	
..	…	…	…	…	…	…	…	
994	0.039862	0.031890	0.063780	…	0.050533	0.021509	0.027930	
995	0.024966	0.010700	0.035665	…	-0.000649	0.011690	0.003123	
997	0.025246	0.021924	0.030561	…	-0.002502	0.008896	0.029719	
998	0.019382	0.032303	0.035533	…	0.036424	0.002265	0.060961	
999	0.036120	0.028896	0.083076	…	-0.014320	-0.016939	0.007312	

	X77	X78	X79	X80	X81	X82	X83
0	-0.004195	0.017409	-0.001178	0.005691	0.023933	-0.002358	efectores
1	0.003517	-0.004335	-0.010706	0.008729	-0.005049	0.009505	efectores
2	-0.013452	-0.012503	0.011419	0.024259	0.038549	0.011367	efectores
3	0.061751	0.039764	-0.036660	0.002825	0.062832	-0.083430	efectores
4	-0.036092	-0.053136	-0.019673	0.027115	0.055705	0.022773	efectores
..	…	…	…	…	…	…	
994	-0.013931	-0.018149	-0.000457	-0.008729	-0.000469	0.028501	efectores
995	0.036822	0.001191	0.024378	-0.016153	-0.031573	0.010834	efectores
997	0.024624	0.012464	0.028572	0.006242	0.003222	0.031870	efectores
998	0.039246	0.002981	0.044399	-0.009072	-0.034714	0.032755	efectores
999	0.018541	0.005545	0.014848	-0.017802	0.025070	0.005581	efectores

[941 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	941.000000	941.000000	941.000000	941.000000	941.000000	941.000000	
mean	0.034248	0.012104	0.027626	0.032336	0.023041	0.028254	
std	0.017190	0.011084	0.017031	0.019918	0.016576	0.015548	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.022706	0.004568	0.014492	0.017601	0.010864	0.017045	
50%	0.031690	0.009258	0.025423	0.029299	0.019570	0.025863	
75%	0.043948	0.016676	0.037532	0.043388	0.031509	0.037242	
max	0.134344	0.072120	0.105238	0.143935	0.101837	0.095957	

	X6	X7	X8	X9	…	X73	\
count	941.000000	941.000000	941.000000	941.000000	…	941.000000	
mean	0.012268	0.029091	0.030136	0.046540	…	0.011194	
std	0.009735	0.019394	0.018650	0.028138	…	0.019420	
min	0.000000	0.000000	0.000000	0.000000	…	-0.100075	
25%	0.005279	0.015621	0.016172	0.026899	…	0.000590	
50%	0.010485	0.025871	0.027132	0.042077	…	0.011017	
75%	0.016867	0.037537	0.040734	0.060511	…	0.022264	
max	0.068888	0.130647	0.144235	0.179125	…	0.096382	

	X74	X75	X76	X77	X78	X79	\
count	941.000000	941.000000	941.000000	941.000000	941.000000	941.000000	
mean	0.001893	0.006745	0.011288	0.002255	0.006848	0.010444	
std	0.028392	0.024126	0.020325	0.028003	0.023516	0.020025	
min	-0.163407	-0.104182	-0.082692	-0.133870	-0.106641	-0.076362	
25%	-0.009679	-0.004835	0.000328	-0.009629	-0.003269	0.000033	
50%	0.003224	0.007447	0.011499	0.003569	0.006815	0.010562	
75%	0.014849	0.018920	0.022168	0.015841	0.018922	0.022034	
max	0.154405	0.133459	0.092712	0.105375	0.112678	0.087619	

	X80	X81	X82
count	941.000000	941.000000	941.000000
mean	0.002088	0.007707	0.011424
std	0.026189	0.022215	0.020177
min	-0.145150	-0.108400	-0.097890
25%	-0.009877	-0.003700	0.000825
50%	0.003179	0.006840	0.011839
75%	0.014543	0.020567	0.022229
max	0.091669	0.095483	0.093116

[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    0.039380  0.003580  0.025060  0.025060  0.021480  0.025060  0.014320
1    0.042482  0.005310  0.013276  0.014603  0.034517  0.014603  0.006638
2    0.017606  0.005869  0.023474  0.035211  0.000000  0.000000  0.000000
3    0.013293  0.011394  0.003798  0.003798  0.003798  0.041777  0.000000
4    0.049698  0.007646  0.035681  0.043326  0.024212  0.035681  0.022937
..        …         …         …         …         …         …         …
995  0.030917  0.014975  0.022221  0.039612  0.037680  0.026569  0.011594
996  0.010325  0.012390  0.004130  0.004130  0.030974  0.006195  0.002065
997  0.011116  0.013894  0.001389  0.000000  0.006947  0.008337  0.001389
998  0.019648  0.009824  0.026198  0.039297  0.019648  0.036022  0.039297
999  0.011346  0.011346  0.022692  0.045384  0.045384  0.022692  0.000000

           X7        X8        X9  …       X74       X75       X76  \
0    0.035800  0.032220  0.060860  …  0.009192 -0.010200  0.021438
1    0.038499  0.010620  0.042482  … -0.000093 -0.004417  0.006653
2    0.017606  0.029343  0.029343  …  0.077400  0.090879  0.026966
3    0.009495  0.005697  0.015192  … -0.005021 -0.002275  0.015934
4    0.052246  0.038229  0.082830  … -0.016260 -0.016881  0.025643
..        …         …         …  …         …         …         …
995  0.049757  0.041545  0.072461  … -0.002133 -0.002341  0.002694
996  0.024779  0.016519  0.055753  …  0.049524  0.015590  0.000224
997  0.005558  0.001389  0.006947  …  0.013450  0.004184  0.029319
998  0.026198  0.049121  0.042571  … -0.043459 -0.013213 -0.001455
999  0.113460  0.056730  0.068076  … -0.011940 -0.000427  0.007857

           X77       X78       X79       X80       X81       X82           X83
0   -0.019998  0.012164 -0.007450  0.002467 -0.002881  0.008534  no_efectores
1    0.021220 -0.003873  0.008060  0.026096  0.005420 -0.008731  no_efectores
2   -0.047855 -0.023920  0.001745 -0.017778 -0.018997 -0.015167  no_efectores
3   -0.003046  0.001546  0.049661 -0.008175 -0.003793  0.016891  no_efectores
4   -0.005908 -0.007224  0.017529  0.004350  0.020206  0.019026  no_efectores
..        …         …         …         …         …         …           …
995  0.001019 -0.001565  0.015322 -0.002850  0.000128  0.010163  no_efectores
996  0.039809  0.025036  0.004213  0.024824  0.013738  0.003775  no_efectores
997  0.017308  0.006629  0.029622  0.016309  0.004395  0.027184  no_efectores
998  0.038729  0.041624  0.007020  0.019968  0.027155 -0.026545  no_efectores
999  0.003412 -0.077809 -0.032986  0.024204 -0.005873 -0.029974  no_efectores

[916 rows x 84 columns]
```

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

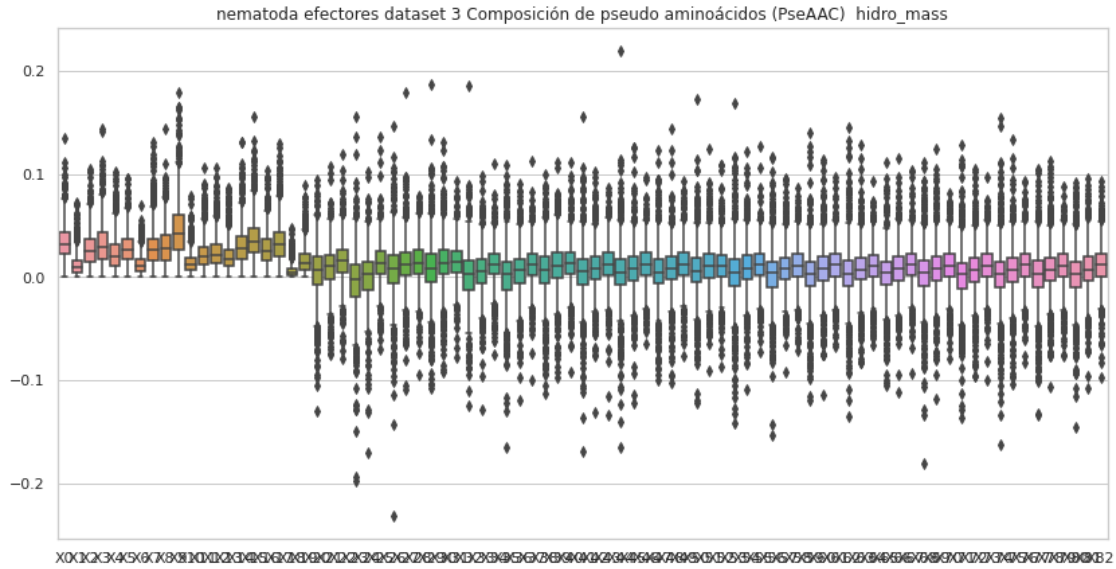
	X0	X1	X2	X3	X4	X5	\
count	916.000000	916.000000	916.000000	916.000000	916.000000	916.000000	
mean	0.028426	0.011077	0.025712	0.033437	0.025003	0.026006	
std	0.014566	0.009473	0.014725	0.019907	0.015923	0.013914	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.018354	0.004934	0.013889	0.018740	0.013923	0.016058	
50%	0.026389	0.008394	0.023406	0.030638	0.021638	0.023725	
75%	0.036240	0.014758	0.036073	0.044993	0.033017	0.033956	
max	0.107876	0.072520	0.102103	0.170351	0.136662	0.098923	

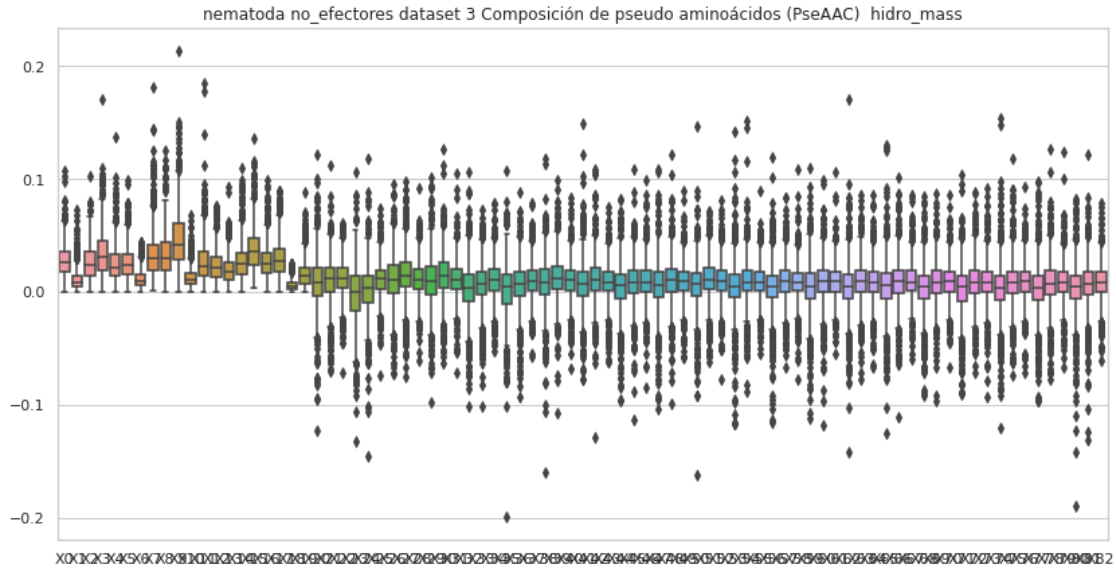
	X6	X7	X8	X9	…	X73	\
count	916.000000	916.000000	916.000000	916.000000	…	916.000000	
mean	0.011456	0.032976	0.033119	0.046256	…	0.008852	
std	0.008431	0.020721	0.019518	0.025442	…	0.018067	
min	0.000000	0.001188	0.000000	0.000000	…	-0.093031	
25%	0.005461	0.018889	0.019422	0.029023	…	-0.000279	
50%	0.009248	0.029666	0.029536	0.041755	…	0.008208	
75%	0.015736	0.041776	0.044170	0.060294	…	0.018197	
max	0.044281	0.181707	0.124924	0.213386	…	0.092281	

	X74	X75	X76	X77	X78	X79	\
count	916.000000	916.000000	916.000000	916.000000	916.000000	916.000000	
mean	0.002896	0.008403	0.009234	0.002755	0.007845	0.009245	
std	0.023124	0.019138	0.016320	0.022887	0.020190	0.016751	
min	-0.120184	-0.073770	-0.071264	-0.096960	-0.077809	-0.066609	
25%	-0.007033	-0.001695	0.000409	-0.007147	-0.002277	0.000468	
50%	0.003271	0.008139	0.009078	0.003890	0.007613	0.008096	
75%	0.013977	0.018415	0.018387	0.013945	0.018791	0.017533	
max	0.153711	0.118474	0.093181	0.097651	0.126881	0.124551	

	X80	X81	X82
count	916.000000	916.000000	916.000000
mean	0.002593	0.007577	0.008831
std	0.023425	0.020897	0.016041
min	-0.189990	-0.130946	-0.060307
25%	-0.006455	-0.001810	0.000157
50%	0.004241	0.007793	0.008678
75%	0.014539	0.018363	0.017747
max	0.082463	0.122013	0.078735

[8 rows x 83 columns]


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


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    0.047290  0.011260  0.055547  0.067558  0.025522  0.020267  0.015763
1    0.052683  0.011707  0.035122  0.040976  0.049756  0.043902  0.023415
2    0.051306  0.012313  0.041045  0.063619  0.042071  0.038992  0.024627
3    0.049935  0.000000  0.074903  0.112355  0.049935  0.037452  0.000000
4    0.028906  0.036133  0.028906  0.043360  0.057813  0.050586  0.014453
..        …         …         …         …         …         …         …
995  0.043197  0.016614  0.006646  0.029906  0.016614  0.026583  0.016614
996  0.017802  0.053407  0.035604  0.044506  0.053407  0.026703  0.008901
```

```
997  0.028573  0.005442  0.015647  0.022450  0.018368  0.039458  0.018368
998  0.029701  0.039602  0.026401  0.019801  0.023101  0.036302  0.013201
999  0.029900  0.014950  0.044850  0.033638  0.026163  0.029900  0.014950

          X7        X8        X9  …       X32       X33       X34  \
0    0.044288  0.068308  0.084822  …  0.015353  0.016306  0.001734
1    0.079024  0.038049  0.090732  … -0.008533  0.044940  0.026279
2    0.050280  0.045149  0.067724  …  0.004519  0.009955  0.003741
3    0.062419  0.074903  0.049935  …  0.082926  0.024585 -0.035718
4    0.101172  0.043360  0.065039  … -0.044478 -0.036112  0.019557
..        …         …         …  …         …         …         …
995  0.023260  0.009969  0.033229  …  0.019816  0.022574  0.032475
996  0.080110  0.017802  0.089011  …  0.024939 -0.013647  0.003600
997  0.025852  0.022450  0.031294  …  0.030730  0.035839  0.015877
998  0.019801  0.033001  0.036302  … -0.002073  0.025723  0.005358
999  0.037375  0.029900  0.085963  …  0.017226  0.019465  0.030846

          X35       X36       X37       X38       X39       X40        X41
0    0.018370  0.005278  0.011608  0.003726 -0.002331 -0.004665  efectores
1    0.013698 -0.006474 -0.016471 -0.047983 -0.029066  0.025806  efectores
2    0.006746  0.016475  0.001861  0.003471  0.014587  0.014519  efectores
3   -0.069769 -0.003784 -0.057243 -0.070777 -0.064323 -0.146382  efectores
4   -0.020337 -0.062311 -0.042894 -0.042351 -0.017360  0.020095  efectores
..        …         …         …         …         …         …         …
995  0.024532  0.043465  0.018222  0.002910  0.022713  0.010094  efectores
996  0.018955 -0.004804  0.014022  0.065955 -0.043477  0.045118  efectores
997  0.021590  0.028861  0.013474  0.030432  0.029258  0.032635  efectores
998  0.011858  0.010322  0.029245  0.062280  0.045359  0.033463  efectores
999  0.035161  0.047340 -0.017775  0.007566  0.015364  0.005775  efectores

[1000 rows x 42 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.045936     0.016315     0.037549     0.046365     0.031948
std       0.018028     0.014302     0.021561     0.035161     0.022857
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       0.034188     0.006789     0.023215     0.026085     0.016965
50%       0.044343     0.013324     0.034744     0.040874     0.027982
75%       0.055054     0.021742     0.048427     0.059277     0.041273
max       0.137765     0.114247     0.144856     0.588807     0.224514

                X5           X6           X7           X8           X9  …  \
```

```
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
mean      0.037332     0.016702     0.040454     0.042913     0.063181  …
std       0.017324     0.014034     0.023983     0.027868     0.031244  …
min       0.000000     0.000000     0.000000     0.000000     0.000000  …
25%       0.025990     0.008615     0.024557     0.023326     0.041626  …
50%       0.035861     0.014569     0.036830     0.037739     0.061506  …
75%       0.045583     0.022229     0.051290     0.055306     0.082670  …
max       0.171874     0.215671     0.171874     0.180256     0.260655  …

               X31          X32          X33          X34          X35  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.013971     0.012427     0.012324     0.012641     0.013876
std       0.032663     0.029583     0.034681     0.028949     0.028965
min      -0.479462    -0.233296    -0.415917    -0.221939    -0.190566
25%       0.000708    -0.000808    -0.000621    -0.000396     0.001114
50%       0.016577     0.015679     0.016779     0.015194     0.015955
75%       0.029945     0.029522     0.029750     0.028408     0.029412
max       0.272789     0.165852     0.139726     0.175527     0.224796

               X36          X37          X38          X39          X40
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.013420     0.013622     0.013316     0.011907     0.012415
std       0.028889     0.031193     0.035159     0.031380     0.031045
min      -0.278742    -0.385574    -0.184184    -0.273415    -0.304350
25%       0.000261    -0.000020     0.000082    -0.002050     0.000246
50%       0.015899     0.016006     0.015541     0.014607     0.017050
75%       0.028948     0.028932     0.028622     0.028921     0.029997
max       0.137235     0.188764     0.602637     0.167122     0.118188

[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    0.049636  0.004512  0.031586  0.031586  0.027074  0.031586  0.018049
1    0.093203  0.011650  0.029126  0.032038  0.075727  0.032038  0.014563
2    0.045586  0.015195  0.060782  0.091172  0.000000  0.000000  0.000000
3    0.014270  0.012232  0.004077  0.004077  0.004077  0.044849  0.000000
4    0.048885  0.007521  0.035097  0.042618  0.023816  0.035097  0.022562
..        …         …         …         …         …         …         …
995  0.033583  0.016267  0.024138  0.043029  0.040930  0.028861  0.012594
996  0.039520  0.047424  0.015808  0.015808  0.118559  0.023712  0.007904
997  0.015030  0.018787  0.001879  0.000000  0.009394  0.011272  0.001879
```

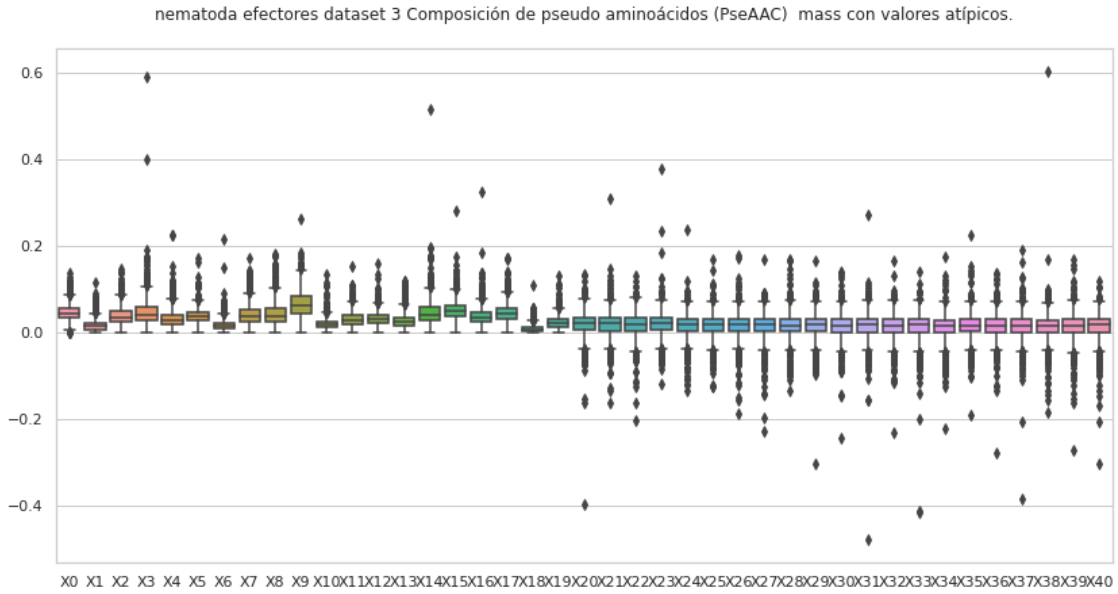
```
998  0.032863  0.016432  0.043817  0.065726  0.032863  0.060249  0.065726
999  0.015177  0.015177  0.030354  0.060709  0.060709  0.030354  0.000000

           X7        X8        X9  …       X32       X33       X34  \
0    0.045124  0.040611  0.076710  … -0.002785  0.045530 -0.003960
1    0.084465  0.023301  0.093203  … -0.012417 -0.000416 -0.017124
2    0.045586  0.075977  0.075977  …  0.077431  0.000641  0.044619
3    0.010193  0.006116  0.016309  …  0.027534  0.027554  0.046213
4    0.051392  0.037604  0.081475  …  0.022920 -0.017770  0.022243
..        …         …         …  …       …         …         …
995  0.054048  0.045128  0.078711  …  0.009474  0.008129  0.017868
996  0.094847  0.063231  0.213406  … -0.014938 -0.029635  0.027766
997  0.007515  0.001879  0.009394  …  0.043592  0.035331  0.040513
998  0.043817  0.082158  0.071203  …  0.047326  0.035726 -0.013603
999  0.151772  0.075886  0.091063  … -0.002057  0.042291  0.004403

          X35       X36       X37       X38       X39       X40           X41
0    0.034887  0.004385  0.021985  0.027021 -0.009391  0.010756  no_efectores
1    0.022825  0.032943 -0.025137  0.014596  0.017682 -0.019156  no_efectores
2    0.021048  0.005441  0.039721  0.069823  0.004519 -0.039271  no_efectores
3    0.025706  0.042154  0.029178  0.017106  0.053313  0.018133  no_efectores
4    0.004922  0.041555  0.014731  0.025223  0.017243  0.018715  no_efectores
..        …         …         …         …         …         …           …
995  0.007551  0.010218  0.011764  0.002926  0.016644  0.011040  no_efectores
996  0.044335 -0.003712 -0.017283  0.000858  0.016124  0.014448  no_efectores
997  0.037320  0.040302  0.035373  0.039643  0.040054  0.036757  no_efectores
998 -0.034407  0.027151  0.021952 -0.002433  0.011742 -0.044397  no_efectores
999  0.011168 -0.027148  0.010992  0.010511 -0.044125 -0.040095  no_efectores

[1000 rows x 42 columns]


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

                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.041250     0.016591     0.037615     0.050661     0.038743
std       0.016921     0.013981     0.019283     0.030100     0.024974
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       0.031072     0.007668     0.024225     0.028970     0.022790
50%       0.040959     0.013530     0.036153     0.046679     0.033782
75%       0.050170     0.021305     0.048153     0.065612     0.048997
max       0.113498     0.106706     0.197836     0.226996     0.258574

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

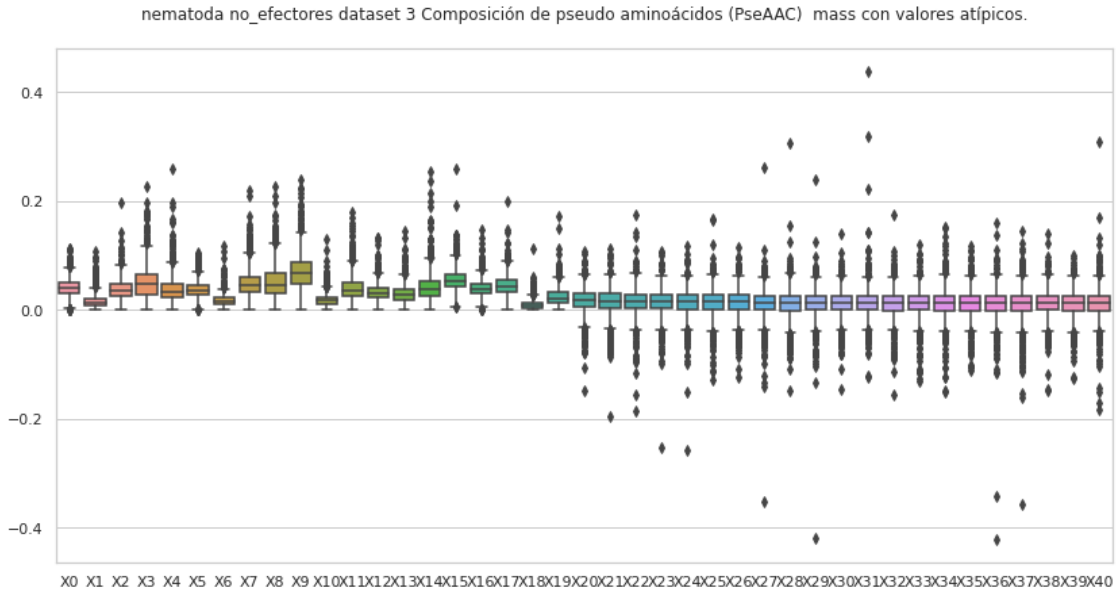
mean	0.037279	0.017015	0.049140	0.051535	0.069315	…
std	0.015139	0.012142	0.027068	0.030389	0.032159	…
min	0.000000	0.000000	0.000000	0.000000	0.000000	…
25%	0.027605	0.009685	0.031633	0.030672	0.047671	…
50%	0.035589	0.014780	0.044498	0.045491	0.067052	…
75%	0.045290	0.021750	0.060995	0.066968	0.086570	…
max	0.104838	0.116165	0.218601	0.226996	0.239884	…

	X31	X32	X33	X34	X35	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.013241	0.010554	0.011246	0.011543	0.011351	
std	0.030481	0.025368	0.025344	0.027652	0.025699	
min	-0.123032	-0.157143	-0.130863	-0.151441	-0.110232	
25%	0.000804	-0.001170	-0.000337	-0.000958	-0.000693	
50%	0.013301	0.012523	0.013567	0.013163	0.012983	
75%	0.025159	0.024316	0.025314	0.025480	0.025232	
max	0.437662	0.175088	0.118820	0.151408	0.116531	

	X36	X37	X38	X39	X40
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.010630	0.010123	0.011822	0.010653	0.011461
std	0.030846	0.030650	0.025587	0.024644	0.029299
min	-0.421818	-0.356811	-0.147888	-0.126069	-0.182432
25%	-0.000998	-0.001629	-0.000597	-0.000941	-0.000758
50%	0.013223	0.012929	0.013501	0.012921	0.013095
75%	0.025607	0.025447	0.025909	0.025470	0.025089
max	0.158661	0.145718	0.139386	0.099760	0.308358

[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   0.047290  0.011260  0.055547  0.067558  0.025522  0.020267  0.015763
1   0.052683  0.011707  0.035122  0.040976  0.049756  0.043902  0.023415
2   0.051306  0.012313  0.041045  0.063619  0.042071  0.038992  0.024627
5   0.062078  0.012416  0.015519  0.083805  0.037247  0.018623  0.021727
```

```
6    0.047756  0.017056  0.057990  0.051168  0.044345  0.047756  0.017056
..        …         …         …         …         …         …
995  0.043197  0.016614  0.006646  0.029906  0.016614  0.026583  0.016614
996  0.017802  0.053407  0.035604  0.044506  0.053407  0.026703  0.008901
997  0.028573  0.005442  0.015647  0.022450  0.018368  0.039458  0.018368
998  0.029701  0.039602  0.026401  0.019801  0.023101  0.036302  0.013201
999  0.029900  0.014950  0.044850  0.033638  0.026163  0.029900  0.014950

           X7        X8        X9  …       X32       X33       X34  \
0    0.044288  0.068308  0.084822  …  0.015353  0.016306  0.001734
1    0.079024  0.038049  0.090732  … -0.008533  0.044940  0.026279
2    0.050280  0.045149  0.067724  …  0.004519  0.009955  0.003741
5    0.046558  0.043455  0.080701  … -0.003358  0.011427  0.040162
6    0.075046  0.071635  0.061401  … -0.033452  0.017745  0.002757
..        …         …         …  …         …         …         …
995  0.023260  0.009969  0.033229  …  0.019816  0.022574  0.032475
996  0.080110  0.017802  0.089011  …  0.024939 -0.013647  0.003600
997  0.025852  0.022450  0.031294  …  0.030730  0.035839  0.015877
998  0.019801  0.033001  0.036302  … -0.002073  0.025723  0.005358
999  0.037375  0.029900  0.085963  …  0.017226  0.019465  0.030846

           X35       X36       X37       X38       X39       X40         X41
0    0.018370  0.005278  0.011608  0.003726 -0.002331 -0.004665  efectores
1    0.013698 -0.006474 -0.016471 -0.047983 -0.029066  0.025806  efectores
2    0.006746  0.016475  0.001861  0.003471  0.014587  0.014519  efectores
5    0.019669  0.018045  0.015806 -0.007418  0.017606  0.037562  efectores
6    0.011605  0.016232  0.031644  0.013686  0.001506 -0.011410  efectores
..        …         …         …         …         …         …
995  0.024532  0.043465  0.018222  0.002910  0.022713  0.010094  efectores
996  0.018955 -0.004804  0.014022  0.065955 -0.043477  0.045118  efectores
997  0.021590  0.028861  0.013474  0.030432  0.029258  0.032635  efectores
998  0.011858  0.010322  0.029245  0.062280  0.045359  0.033463  efectores
999  0.035161  0.047340 -0.017775  0.007566  0.015364  0.005775  efectores

[820 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  820.000000  820.000000  820.000000  820.000000  820.000000  820.000000
mean     0.044100    0.014366    0.034603    0.041667    0.028393    0.035739
std      0.014600    0.010526    0.017355    0.023955    0.017077    0.013328
min      0.004427    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.034173    0.006668    0.022691    0.024954    0.016102    0.026036
50%      0.043269    0.012645    0.033129    0.037628    0.026530    0.035047
```

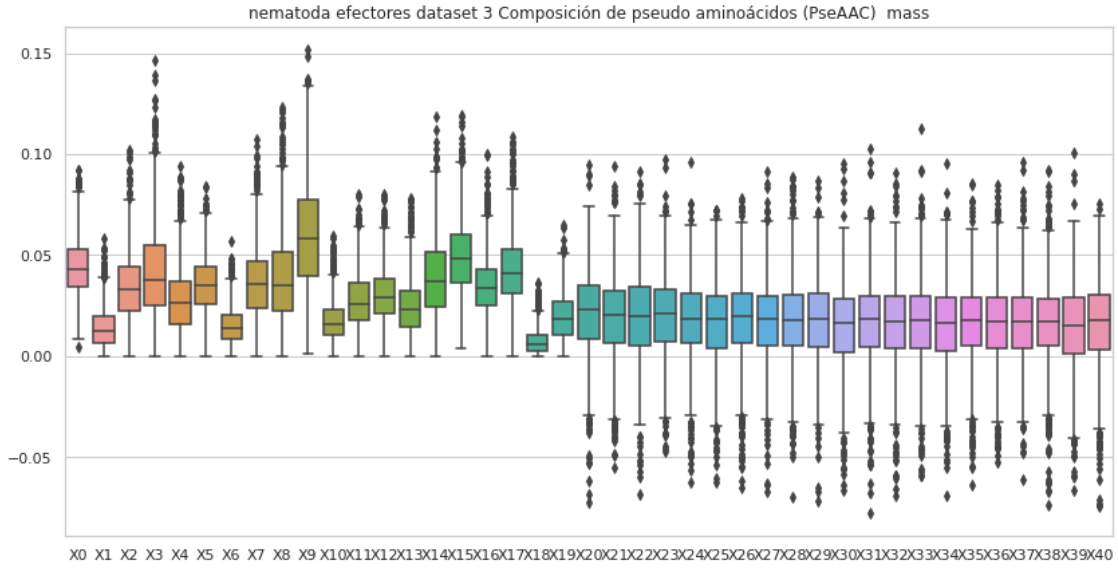
```
75%      0.053153    0.019574    0.044593    0.055211    0.037067    0.044153
max      0.092361    0.058063    0.101919    0.146324    0.094134    0.084204

               X6          X7          X8          X9  …         X31  \
count  820.000000  820.000000  820.000000  820.000000  …  820.000000
mean     0.015125    0.036668    0.038804    0.059524  …    0.016740
std      0.009363    0.018704    0.022352    0.027320  …    0.021903
min      0.000000    0.000000    0.000000    0.001513  …   -0.077635
25%      0.008558    0.023494    0.022699    0.040005  …    0.004466
50%      0.014139    0.035534    0.035062    0.058288  …    0.018176
75%      0.020588    0.046911    0.051321    0.077697  …    0.029936
max      0.056875    0.107246    0.123045    0.151691  …    0.103009

               X32         X33         X34         X35         X36         X37  \
count  820.000000  820.000000  820.000000  820.000000  820.000000  820.000000
mean     0.016211    0.016181    0.015872    0.015770    0.016346    0.016415
std      0.021748    0.021854    0.021427    0.021203    0.021259    0.020687
min     -0.068800   -0.059078   -0.069030   -0.064018   -0.052694   -0.060975
25%      0.003989    0.003870    0.002844    0.005141    0.003725    0.004061
50%      0.017132    0.017634    0.016349    0.017500    0.016976    0.017376
75%      0.029799    0.029680    0.028935    0.029348    0.029034    0.028791
max      0.090991    0.112673    0.095688    0.085616    0.084810    0.095850

               X38         X39         X40
count  820.000000  820.000000  820.000000
mean     0.016234    0.014262    0.015645
std      0.021991    0.022784    0.022801
min     -0.073604   -0.066216   -0.074633
25%      0.005446    0.001166    0.003151
50%      0.016968    0.015350    0.017982
75%      0.028646    0.029156    0.030172
max      0.091822    0.100776    0.075833

[8 rows x 41 columns]
```

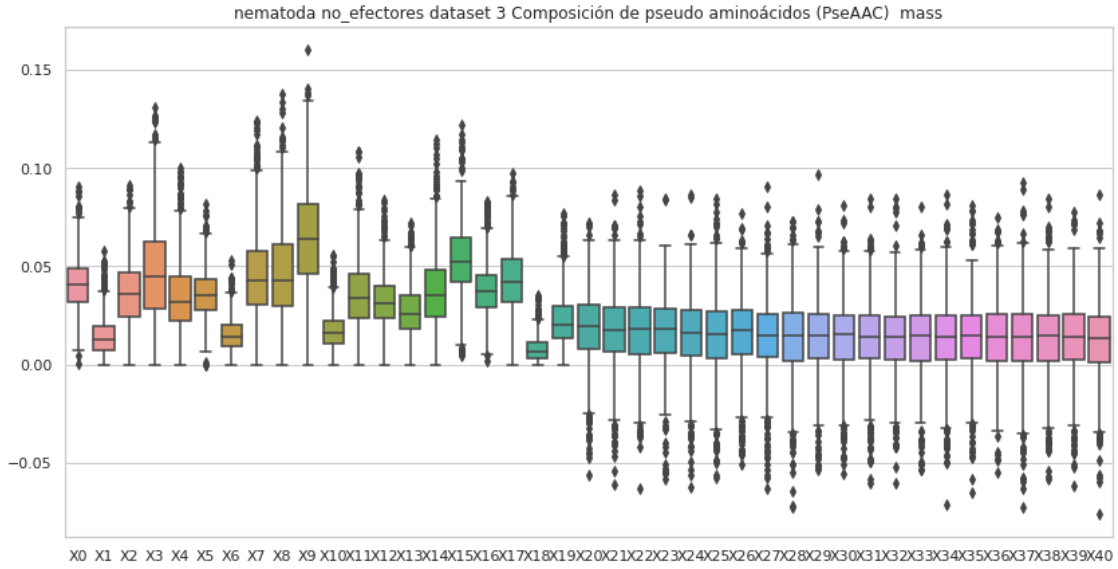
Composición de pseudo aminoácidos (PseAAC) mass no_efectores nematoda dataset 3, sin valores atípicos.
Valores del documento csv.

```
          X0        X1        X2        X3        X4        X5        X6  \
0    0.049636  0.004512  0.031586  0.031586  0.027074  0.031586  0.018049
2    0.045586  0.015195  0.060782  0.091172  0.000000  0.000000  0.000000
3    0.014270  0.012232  0.004077  0.004077  0.004077  0.044849  0.000000
4    0.048885  0.007521  0.035097  0.042618  0.023816  0.035097  0.022562
5    0.051008  0.005232  0.037929  0.054931  0.022234  0.028774  0.015695
..        …         …         …         …         …         …         …
```

```
992  0.043439  0.006206  0.024822  0.031028  0.018617  0.037233  0.012411
993  0.048912  0.016304  0.052173  0.032608  0.029347  0.039130  0.022826
994  0.023577  0.000000  0.032418  0.082518  0.044206  0.050100  0.005894
995  0.033583  0.016267  0.024138  0.043029  0.040930  0.028861  0.012594
997  0.015030  0.018787  0.001879  0.000000  0.009394  0.011272  0.001879

           X7        X8        X9  …       X32       X33       X34  \
0    0.045124  0.040611  0.076710  … -0.002785  0.045530 -0.003960
2    0.045586  0.075977  0.075977  …  0.077431  0.000641  0.044619
3    0.010193  0.006116  0.016309  …  0.027534  0.027554  0.046213
4    0.051392  0.037604  0.081475  …  0.022920 -0.017770  0.022243
5    0.036621  0.075858  0.056239  …  0.011532 -0.007029  0.022222
..        …         …         …  …         …         …
992  0.018617  0.024822  0.049644  … -0.008942  0.030395 -0.012775
993  0.026086  0.039130  0.045651  …  0.011198  0.019660  0.005994
994  0.076624  0.061889  0.076624  …  0.018465  0.031435 -0.010630
995  0.054048  0.045128  0.078711  …  0.009474  0.008129  0.017868
997  0.007515  0.001879  0.009394  …  0.043592  0.035331  0.040513

          X35       X36       X37       X38       X39       X40           X41
0    0.034887  0.004385  0.021985  0.027021 -0.009391  0.010756  no_efectores
2    0.021048  0.005441  0.039721  0.069823  0.004519 -0.039271  no_efectores
3    0.025706  0.042154  0.029178  0.017106  0.053313  0.018133  no_efectores
4    0.004922  0.041555  0.014731  0.025223  0.017243  0.018715  no_efectores
5    0.035223 -0.000193  0.012078  0.015544  0.023310  0.002560  no_efectores
..        …         …         …         …         …         …           …
992  0.003851  0.018654  0.032965  0.002961  0.020361 -0.025803  no_efectores
993  0.019694  0.020667  0.005570  0.050021 -0.024072  0.026681  no_efectores
994 -0.015735  0.020920  0.013664 -0.025242 -0.008363  0.010961  no_efectores
995  0.007551  0.010218  0.011764  0.002926  0.016644  0.011040  no_efectores
997  0.037320  0.040302  0.035373  0.039643  0.040054  0.036757  no_efectores

[813 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  813.000000  813.000000  813.000000  813.000000  813.000000  813.000000
mean     0.040530    0.014898    0.036396    0.047162    0.034967    0.036184
std      0.013910    0.010214    0.016534    0.024598    0.018476    0.012428
min      0.000443    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.031572    0.007521    0.024586    0.028418    0.021990    0.027803
50%      0.040754    0.013048    0.035780    0.044760    0.032022    0.035134
75%      0.048885    0.019691    0.046766    0.062703    0.044593    0.043385
max      0.090552    0.057524    0.091375    0.130939    0.100242    0.081971
```

```
              X6          X7          X8          X9  …         X31  \
count  813.000000  813.000000  813.000000  813.000000  …  813.000000
mean     0.015546    0.045704    0.046867    0.064526  …    0.013747
std      0.008901    0.022319    0.024825    0.025980  …    0.018795
min      0.000000    0.000000    0.000000    0.000000  …   -0.060529
25%      0.009454    0.030415    0.029696    0.046439  …    0.003056
50%      0.014295    0.042946    0.042925    0.063806  …    0.014215
75%      0.020457    0.057735    0.061182    0.081748  …    0.024901
max      0.052683    0.123772    0.137271    0.160063  …    0.084485

              X32         X33         X34         X35         X36         X37  \
count  813.000000  813.000000  813.000000  813.000000  813.000000  813.000000
mean     0.013089    0.013305    0.012979    0.013765    0.013375    0.013093
std      0.018521    0.018493    0.019967    0.019462    0.018399    0.020674
min     -0.060312   -0.053761   -0.071356   -0.065196   -0.054544   -0.072261
25%      0.002660    0.002127    0.002178    0.003130    0.001714    0.001486
50%      0.013802    0.014743    0.014379    0.014539    0.014142    0.014063
75%      0.024428    0.024858    0.025157    0.025260    0.025362    0.025793
max      0.084559    0.080513    0.086465    0.081131    0.074847    0.092410

              X38         X39         X40
count  813.000000  813.000000  813.000000
mean     0.013326    0.013356    0.012517
std      0.019739    0.019269    0.019353
min     -0.058395   -0.061352   -0.075993
25%      0.001486    0.002397    0.000841
50%      0.014546    0.014167    0.013485
75%      0.024977    0.025642    0.024518
max      0.084498    0.077765    0.086611

[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    0.027084  0.006449  0.031813  0.038692  0.014617  0.011608  0.009028
1    0.020763  0.004614  0.013842  0.016149  0.019609  0.017302  0.009228
2    0.047741  0.011458  0.038193  0.059199  0.039148  0.036283  0.022916
3    0.025900  0.000000  0.038849  0.058274  0.025900  0.019425  0.000000
4    0.036789  0.045986  0.036789  0.055183  0.073577  0.064380  0.018394
..        …         …         …         …         …         …         …
995  0.122134  0.046975  0.018790  0.084554  0.046975  0.075159  0.046975
996  0.035518  0.106554  0.071036  0.088795  0.106554  0.053277  0.017759
997  0.056813  0.010821  0.031112  0.044638  0.036522  0.078455  0.036522
998  0.051857  0.069143  0.046095  0.034571  0.040333  0.063381  0.023048
999  0.041174  0.020587  0.061761  0.046321  0.036027  0.041174  0.020587

           X7        X8        X9  …       X53       X54       X55  \
0    0.025365  0.039122  0.048580  …  0.020197  0.002767  0.022901
1    0.031144  0.014995  0.035758  …  0.004720  0.005339  0.012743
2    0.046787  0.042012  0.063019  …  0.028249 -0.013051  0.014121
```

```
3    0.032374  0.038849  0.025900  …  0.036252 -0.031668  0.034457
4    0.128760  0.055183  0.082774  … -0.012090  0.048865 -0.012283
..        …         …         …  …         …         …
995  0.065764  0.028185  0.093949  … -0.015629  0.039588  0.008535
996  0.159832  0.035518  0.177591  … -0.231344  0.079750 -0.049997
997  0.051402  0.044638  0.062223  … -0.015452 -0.074874 -0.022161
998  0.034571  0.057619  0.063381  … -0.010559  0.049554  0.001634
999  0.051467  0.041174  0.118375  … -0.036648 -0.069239  0.035593

          X56       X57       X58       X59       X60       X61         X62
0    0.004254  0.027924 -0.004752  0.019724  0.006447  0.027114  efectores
1    0.020664  0.012885  0.003763 -0.004638  0.009340 -0.005403  efectores
2   -0.002783 -0.007011 -0.015990 -0.014861  0.028835  0.045821  efectores
3   -0.110479  0.004064  0.056195  0.036186  0.002571  0.057179  efectores
4    0.173402  0.149879 -0.040533 -0.059674  0.030451  0.062559  efectores
..        …         …         …         …         …         …
995 -0.001710  0.030794  0.096996  0.003138 -0.042551 -0.083169  efectores
996 -0.228943 -0.178873  0.145683 -0.073650  0.056038 -0.093197  efectores
997 -0.005094  0.018113  0.050136  0.025378  0.012710  0.006561  efectores
998  0.064969  0.004039  0.070004  0.005317 -0.016183 -0.061919  efectores
999 -0.020405 -0.024136  0.026419  0.007901 -0.025366  0.035722  efectores

[1000 rows x 63 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.052833     0.018690     0.038806     0.044151     0.034106
std       0.038768     0.027339     0.039196     0.052285     0.034976
min      -0.459553    -0.459553    -0.919105    -1.378658    -0.459553
25%       0.029540     0.006422     0.021607     0.026524     0.016038
50%       0.045887     0.013180     0.037483     0.043599     0.029412
75%       0.069378     0.024464     0.053111     0.059436     0.044643
max       0.305904     0.251342     0.244723     0.194590     0.367085

                X5           X6           X7           X8           X9  …  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
mean      0.043081     0.017498     0.043087     0.042143     0.066480  …
std       0.056009     0.034820     0.046321     0.082444     0.101702  …
min      -1.378658    -0.919105    -0.919105    -2.297763    -2.757315  …
25%       0.021913     0.007272     0.022900     0.023697     0.038068  …
50%       0.037369     0.014394     0.037499     0.039052     0.060887  …
75%       0.058308     0.024487     0.055766     0.056466     0.089217  …
max       0.305904     0.267549     0.428266     0.428266     0.611808  …
```

```
              X52          X53          X54          X55          X56  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.000564     0.004669    -0.001015     0.007432     0.002998
std       0.103735     0.051371     0.059594     0.047129     0.077040
min      -0.900457    -0.585647    -0.976217    -0.415998    -0.353382
25%      -0.014805    -0.006891    -0.018335    -0.008386    -0.016198
50%       0.005112     0.010652     0.002936     0.009904     0.003835
75%       0.022084     0.025599     0.019138     0.025086     0.021065
max       2.623964     0.182076     0.412139     0.422375     1.803371

              X57          X58          X59          X60          X61
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.008733     0.002487     0.008542     0.002608     0.008247
std       0.101624     0.056717     0.056429     0.139412     0.075069
min      -0.485235    -0.320707    -0.542256    -0.740875    -0.675334
25%      -0.008602    -0.015380    -0.006177    -0.015748    -0.007332
50%       0.009753     0.005298     0.009847     0.004270     0.009780
75%       0.026201     0.022760     0.024874     0.022095     0.026891
max       2.855189     0.702958     1.175906     3.985667     1.775759

[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.

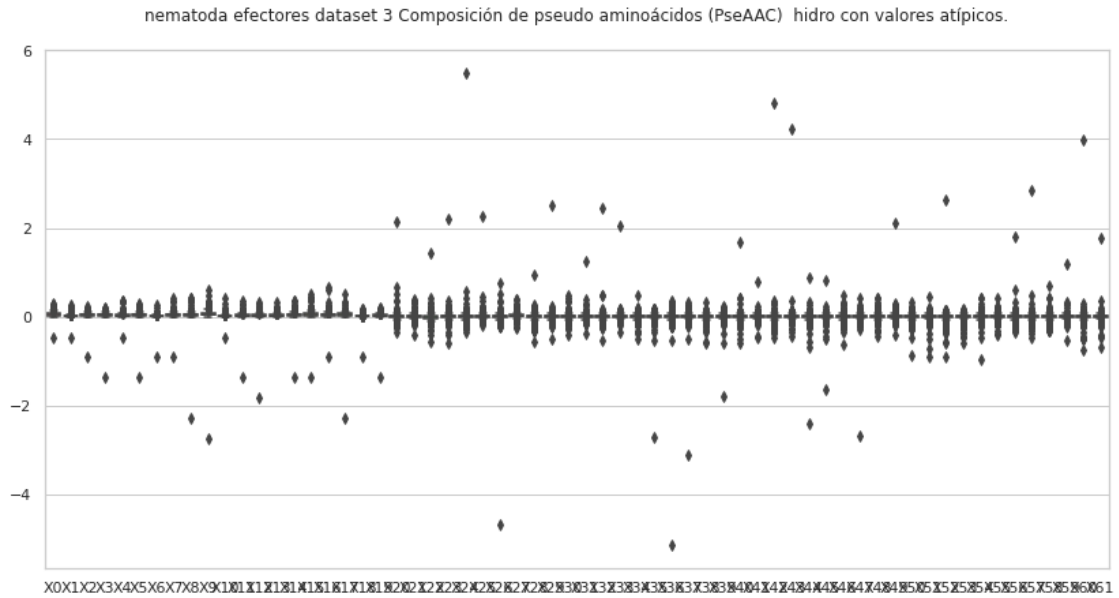
```
          X0        X1        X2        X3        X4        X5        X6  \
0    0.055636  0.005058  0.035404  0.035404  0.030347  0.035404  0.020231
1    0.044823  0.005603  0.014007  0.015408  0.036418  0.015408  0.007004
2    0.018008  0.006003  0.024011  0.036016  0.000000  0.000000  0.000000
3    0.044565  0.038199  0.012733  0.012733  0.012733  0.140063  0.000000
4    0.071174  0.010950  0.051099  0.062049  0.034675  0.051099  0.032850
..        …         …         …         …         …         …         …
995  0.043233  0.020941  0.031073  0.055392  0.052690  0.037153  0.016212
996  0.010554  0.012665  0.004222  0.004222  0.031662  0.006332  0.002111
997  0.027832  0.034790  0.003479  0.000000  0.017395  0.020874  0.003479
998  0.021217  0.010608  0.028289  0.042434  0.021217  0.038898  0.042434
999  0.011462  0.011462  0.022923  0.045846  0.045846  0.022923  0.000000

          X7        X8        X9  …       X53       X54       X55  \
0    0.050578  0.045520  0.085982  …  0.025241  0.055316  0.029865
1    0.040621  0.011206  0.044823  …  0.001905  0.007265  0.006030
2    0.018008  0.030014  0.030014  …  0.017805  0.105567  0.089432
3    0.031832  0.019099  0.050932  …  0.026237  0.046950  0.022341
```

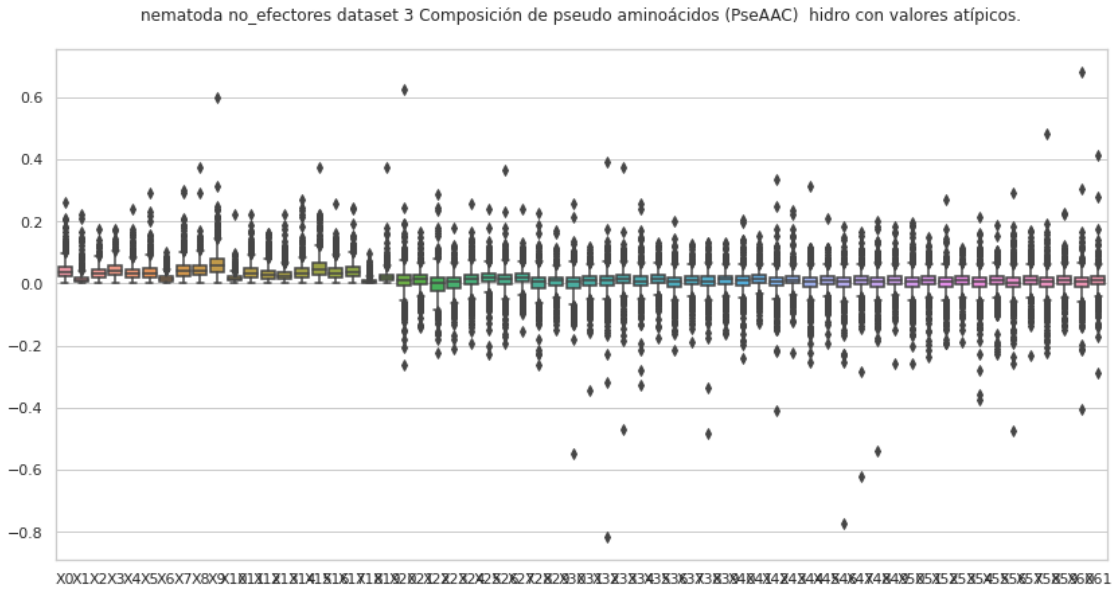
```
4    0.074824  0.054749  0.118623  …  0.012749  0.009738  0.017477
..        …         …         …  …         …         …
995  0.069577  0.058094  0.101326  … -0.027506 -0.019470 -0.001337
996  0.025329  0.016886  0.056991  …  0.010317  0.020464  0.008256
997  0.013916  0.003479  0.017395  … -0.000652  0.018421  0.002692
998  0.028289  0.053042  0.045970  … -0.009995  0.053334  0.073227
999  0.114616  0.057308  0.068769  …  0.007476  0.028679  0.060531

          X56       X57       X58       X59       X60       X61           X62
0    0.012986 -0.014410 -0.028253  0.017184  0.003485 -0.004071  no_efectores
1   -0.000098 -0.004660  0.022390 -0.004087  0.027534  0.005719  no_efectores
2    0.079170  0.092957 -0.048949 -0.024467 -0.018185 -0.019431  no_efectores
3   -0.016833 -0.007629 -0.010212  0.005184 -0.027407 -0.012716  no_efectores
4   -0.023287 -0.024176 -0.008461 -0.010345  0.006230  0.028938  no_efectores
..        …         …         …         …         …         …           …
995 -0.002982 -0.003274  0.001425 -0.002188 -0.003985  0.000179  no_efectores
996  0.050624  0.015936  0.040693  0.025592  0.025375  0.014043  no_efectores
997  0.033677  0.010475  0.043336  0.016597  0.040836  0.011004  no_efectores
998 -0.046928 -0.014267  0.041821  0.044947  0.021562  0.029322  no_efectores
999 -0.012062 -0.000432  0.003447 -0.078601  0.024450 -0.005933  no_efectores

[1000 rows x 63 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.040890     0.016343     0.034786     0.043865     0.034762
std       0.027939     0.019176     0.021116     0.024379     0.023682
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       0.022280     0.006352     0.019407     0.026023     0.018958
50%       0.035058     0.011510     0.031764     0.041713     0.030228
75%       0.052870     0.020942     0.046912     0.058363     0.044944
max       0.260644     0.220706     0.175653     0.176848     0.238458

                X5           X6           X7           X8           X9  …  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
mean      0.038072     0.016023     0.044846     0.045220     0.063728  …
std       0.028389     0.012852     0.030109     0.029146     0.040624  …
min       0.000000     0.000000     0.000000     0.000000     0.000000  …
25%       0.019292     0.006887     0.025484     0.026343     0.038464  …
50%       0.031326     0.013010     0.039529     0.039749     0.056548  …
75%       0.049127     0.022064     0.056842     0.058916     0.081701  …
max       0.293225     0.100183     0.299333     0.374166     0.598665  …
```

```
                X52           X53           X54           X55           X56  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.001279     0.008743     0.001974     0.009061     0.002298
std       0.035208     0.027922     0.039319     0.033018     0.041399
min      -0.194449    -0.189108    -0.374725    -0.229489    -0.474301
25%      -0.011318    -0.003398    -0.011449    -0.002748    -0.011601
50%       0.005229     0.010795     0.006022     0.011904     0.004032
75%       0.018340     0.023666     0.018598     0.024672     0.018437
max       0.269280     0.116401     0.212840     0.187345     0.291024

                X57           X58           X59           X60           X61
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.009003     0.002206     0.008518     0.002502     0.008825
std       0.031215     0.040432     0.031055     0.043782     0.035318
min      -0.231394    -0.223589    -0.150040    -0.405652    -0.289052
25%      -0.004362    -0.011060    -0.004105    -0.011866    -0.003725
50%       0.010326     0.005035     0.009671     0.005036     0.010106
75%       0.023431     0.018052     0.023919     0.019342     0.023305
max       0.170681     0.481561     0.225009     0.681374     0.412838

[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    0.027084  0.006449  0.031813  0.038692  0.014617  0.011608  0.009028
1    0.020763  0.004614  0.013842  0.016149  0.019609  0.017302  0.009228
2    0.047741  0.011458  0.038193  0.059199  0.039148  0.036283  0.022916
3    0.025900  0.000000  0.038849  0.058274  0.025900  0.019425  0.000000
4    0.036789  0.045986  0.036789  0.055183  0.073577  0.064380  0.018394
..        …         …         …         …         …         …         …
994  0.076119  0.025373  0.063433  0.069776  0.076119  0.095149  0.038060
995  0.122134  0.046975  0.018790  0.084554  0.046975  0.075159  0.046975
997  0.056813  0.010821  0.031112  0.044638  0.036522  0.078455  0.036522
998  0.051857  0.069143  0.046095  0.034571  0.040333  0.063381  0.023048
999  0.041174  0.020587  0.061761  0.046321  0.036027  0.041174  0.020587

           X7        X8        X9  …       X53       X54       X55  \
0    0.025365  0.039122  0.048580  …  0.020197  0.002767  0.022901
1    0.031144  0.014995  0.035758  …  0.004720  0.005339  0.012743
2    0.046787  0.042012  0.063019  …  0.028249 -0.013051  0.014121
3    0.032374  0.038849  0.025900  …  0.036252 -0.031668  0.034457
4    0.128760  0.055183  0.082774  … -0.012090  0.048865 -0.012283
..        …         …         …  …         …         …         …
```

```
994  0.063433  0.050746  0.101492  … -0.045398  0.014356  0.036395
995  0.065764  0.028185  0.093949  … -0.015629  0.039588  0.008535
997  0.051402  0.044638  0.062223  … -0.015452 -0.074874 -0.022161
998  0.034571  0.057619  0.063381  … -0.010559  0.049554  0.001634
999  0.051467  0.041174  0.118375  … -0.036648 -0.069239  0.035593

          X56       X57       X58       X59       X60       X61         X62
0    0.004254  0.027924 -0.004752  0.019724  0.006447  0.027114  efectores
1    0.020664  0.012885  0.003763 -0.004638  0.009340 -0.005403  efectores
2   -0.002783 -0.007011 -0.015990 -0.014861  0.028835  0.045821  efectores
3   -0.110479  0.004064  0.056195  0.036186  0.002571  0.057179  efectores
4    0.173402  0.149879 -0.040533 -0.059674  0.030451  0.062559  efectores
..        …         …         …         …         …         …         …
994  0.080413  0.034227 -0.022168 -0.028881 -0.013891 -0.000746  efectores
995 -0.001710  0.030794  0.096996  0.003138 -0.042551 -0.083169  efectores
997 -0.005094  0.018113  0.050136  0.025378  0.012710  0.006561  efectores
998  0.064969  0.004039  0.070004  0.005317 -0.016183 -0.061919  efectores
999 -0.020405 -0.024136  0.026419  0.007901 -0.025366  0.035722  efectores

[918 rows x 63 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  918.000000  918.000000  918.000000  918.000000  918.000000  918.000000
mean     0.049090    0.016893    0.037479    0.043337    0.030887    0.040609
std      0.027858    0.015882    0.021124    0.022824    0.020478    0.026617
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.028566    0.006428    0.020709    0.026068    0.015449    0.020991
50%      0.044237    0.012580    0.036072    0.042225    0.028170    0.035951
75%      0.065390    0.023070    0.050508    0.058282    0.041530    0.053041
max      0.168054    0.099618    0.125986    0.167986    0.107979    0.151620

                X6          X7          X8          X9  …         X52  \
count  918.000000  918.000000  918.000000  918.000000  …  918.000000
mean     0.016760    0.039605    0.040355    0.062776  …    0.004467
std      0.012913    0.024445    0.022586    0.034977  …    0.034965
min      0.000000    0.000000    0.000000    0.000000  …   -0.184497
25%      0.007228    0.022521    0.022889    0.036554  …   -0.010946
50%      0.013921    0.035794    0.038003    0.058426  …    0.006317
75%      0.022917    0.051694    0.055017    0.083360  …    0.021753
max      0.116392    0.168083    0.137802    0.224073  …    0.140409

               X53         X54         X55         X56         X57         X58  \
count  918.000000  918.000000  918.000000  918.000000  918.000000  918.000000
```

mean	0.010271	0.001637	0.009147	0.003307	0.009411	0.003454
std	0.029522	0.033645	0.029979	0.035205	0.029707	0.036632
min	-0.143138	-0.158827	-0.112606	-0.156227	-0.147206	-0.162806
25%	-0.004681	-0.014777	-0.005861	-0.013551	-0.005594	-0.012759
50%	0.011977	0.003829	0.010347	0.004436	0.010762	0.005606
75%	0.025940	0.019024	0.024532	0.020550	0.026175	0.021965
max	0.148946	0.148664	0.147064	0.173402	0.149879	0.152035

	X59	X60	X61
count	918.000000	918.000000	918.000000
mean	0.009232	0.002709	0.009663
std	0.028645	0.034949	0.029609
min	-0.108853	-0.193280	-0.153447
25%	-0.004413	-0.012663	-0.005064
50%	0.010471	0.004590	0.010372
75%	0.024288	0.021087	0.026687
max	0.116024	0.143735	0.109033

[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	0.055636	0.005058	0.035404	0.035404	0.030347	0.035404	0.020231	
1	0.044823	0.005603	0.014007	0.015408	0.036418	0.015408	0.007004	
2	0.018008	0.006003	0.024011	0.036016	0.000000	0.000000	0.000000	
4	0.071174	0.010950	0.051099	0.062049	0.034675	0.051099	0.032850	
5	0.039539	0.004055	0.029401	0.042581	0.017235	0.022304	0.012166	
..	…	…	…	…	…	…	…	
994	0.016801	0.000000	0.023101	0.058803	0.031501	0.035702	0.004200	
995	0.043233	0.020941	0.031073	0.055392	0.052690	0.037153	0.016212	
996	0.010554	0.012665	0.004222	0.004222	0.031662	0.006332	0.002111	
998	0.021217	0.010608	0.028289	0.042434	0.021217	0.038898	0.042434	
999	0.011462	0.011462	0.022923	0.045846	0.045846	0.022923	0.000000	

	X7	X8	X9	…	X53	X54	X55	\
0	0.050578	0.045520	0.085982	…	0.025241	0.055316	0.029865	
1	0.040621	0.011206	0.044823	…	0.001905	0.007265	0.006030	
2	0.018008	0.030014	0.030014	…	0.017805	0.105567	0.089432	
4	0.074824	0.054749	0.118623	…	0.012749	0.009738	0.017477	
5	0.028387	0.058802	0.043594	…	0.020587	0.006030	0.035202	
..	…	…	…	…	…	…	…	
994	0.054603	0.044102	0.054603	…	-0.004683	0.000709	0.043994	

```
995  0.069577  0.058094  0.101326  … -0.027506 -0.019470 -0.001337
996  0.025329  0.016886  0.056991  …  0.010317  0.020464  0.008256
998  0.028289  0.053042  0.045970  … -0.009995  0.053334  0.073227
999  0.114616  0.057308  0.068769  …  0.007476  0.028679  0.060531

          X56       X57       X58       X59       X60       X61           X62
0    0.012986 -0.014410 -0.028253  0.017184  0.003485 -0.004071  no_efectores
1   -0.000098 -0.004660  0.022390 -0.004087  0.027534  0.005719  no_efectores
2    0.079170  0.092957 -0.048949 -0.024467 -0.018185 -0.019431  no_efectores
4   -0.023287 -0.024176 -0.008461 -0.010345  0.006230  0.028938  no_efectores
5   -0.001771  0.016394 -0.018341  0.011329 -0.001754  0.017463  no_efectores
..        …         …         …         …         …         …            …
994 -0.022584 -0.010904 -0.011095  0.004093  0.008126  0.001432  no_efectores
995 -0.002982 -0.003274  0.001425 -0.002188 -0.003985  0.000179  no_efectores
996  0.050624  0.015936  0.040693  0.025592  0.025375  0.014043  no_efectores
998 -0.046928 -0.014267  0.041821  0.044947  0.021562  0.029322  no_efectores
999 -0.012062 -0.000432  0.003447 -0.078601  0.024450 -0.005933  no_efectores

[810 rows x 63 columns]
```

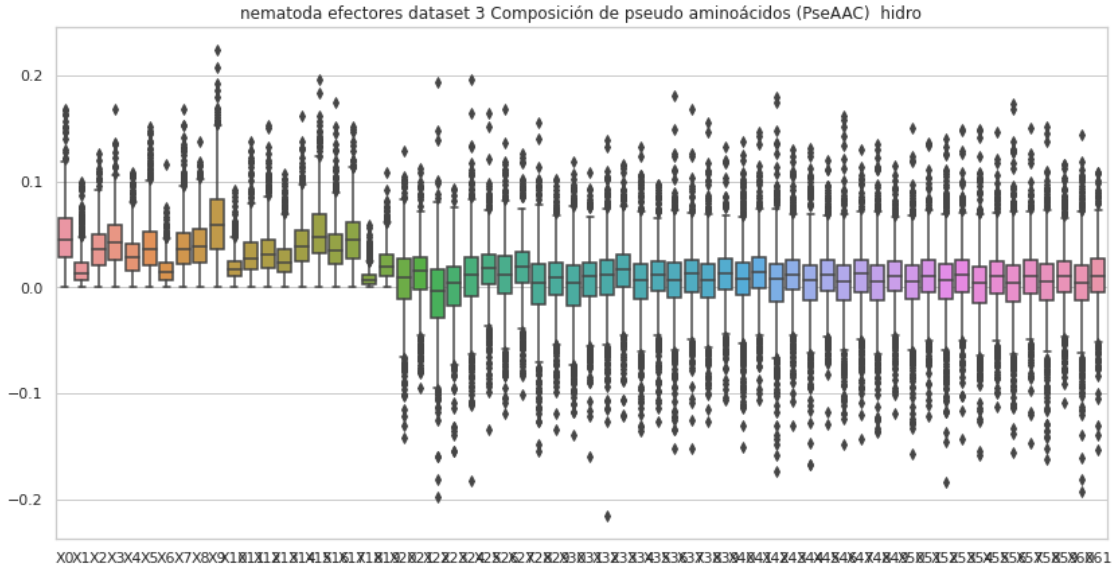
Composición de pseudo aminoácidos (PseAAC) no_efectores nematoda dataset 3, sin valores atípicos.
Estadísticas.

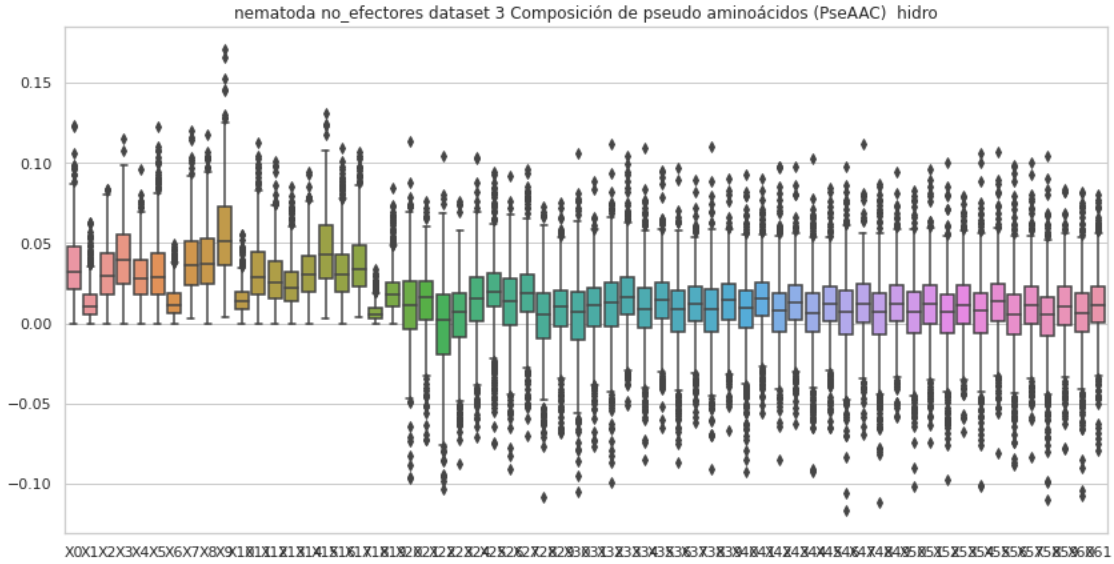
```
                X0          X1          X2          X3          X4          X5  \
count  810.000000  810.000000  810.000000  810.000000  810.000000  810.000000
mean     0.035437    0.013065    0.031648    0.040364    0.029505    0.032507
std      0.020052    0.010497    0.017251    0.020609    0.015699    0.019876
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.020835    0.005799    0.018052    0.024848    0.017853    0.018022
50%      0.031629    0.010356    0.029433    0.039067    0.027696    0.028434
75%      0.047882    0.017984    0.043677    0.055034    0.039290    0.043528
max      0.123070    0.062950    0.083927    0.115250    0.095940    0.122267

                X6          X7          X8          X9  …         X52  \
count  810.000000  810.000000  810.000000  810.000000  …  810.000000
mean     0.013813    0.038872    0.039617    0.055329  …    0.004702
std      0.009539    0.020604    0.020575    0.026919  …    0.022401
min      0.000000    0.002575    0.000000    0.003442  …   -0.097746
25%      0.006443    0.023862    0.024824    0.035852  …   -0.006611
50%      0.011539    0.036168    0.037009    0.050866  …    0.006672
75%      0.018917    0.051382    0.052998    0.072389  …    0.017936
max      0.049243    0.120185    0.117253    0.170711  …    0.100201

               X53         X54         X55         X56         X57         X58  \
count  810.000000  810.000000  810.000000  810.000000  810.000000  810.000000
mean     0.011560    0.005732    0.012530    0.004607    0.010705    0.003594
```

std	0.019753	0.023415	0.020948	0.023635	0.020426	0.023902
min	-0.067457	-0.101854	-0.081360	-0.088430	-0.083126	-0.109741
25%	-0.000186	-0.006633	0.001517	-0.007109	-0.000569	-0.007982
50%	0.011586	0.007606	0.013591	0.005189	0.011229	0.005396
75%	0.023465	0.018441	0.024584	0.017745	0.022645	0.016562
max	0.078960	0.105567	0.106674	0.098090	0.099739	0.104333

	X59	X60	X61
count	810.000000	810.000000	810.000000
mean	0.009480	0.005352	0.011065
std	0.021002	0.023000	0.020922
min	-0.078601	-0.107436	-0.079069
25%	-0.000861	-0.005730	0.000081
50%	0.010112	0.006237	0.011013
75%	0.022614	0.018514	0.022957
max	0.083403	0.081942	0.080081

[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))
```

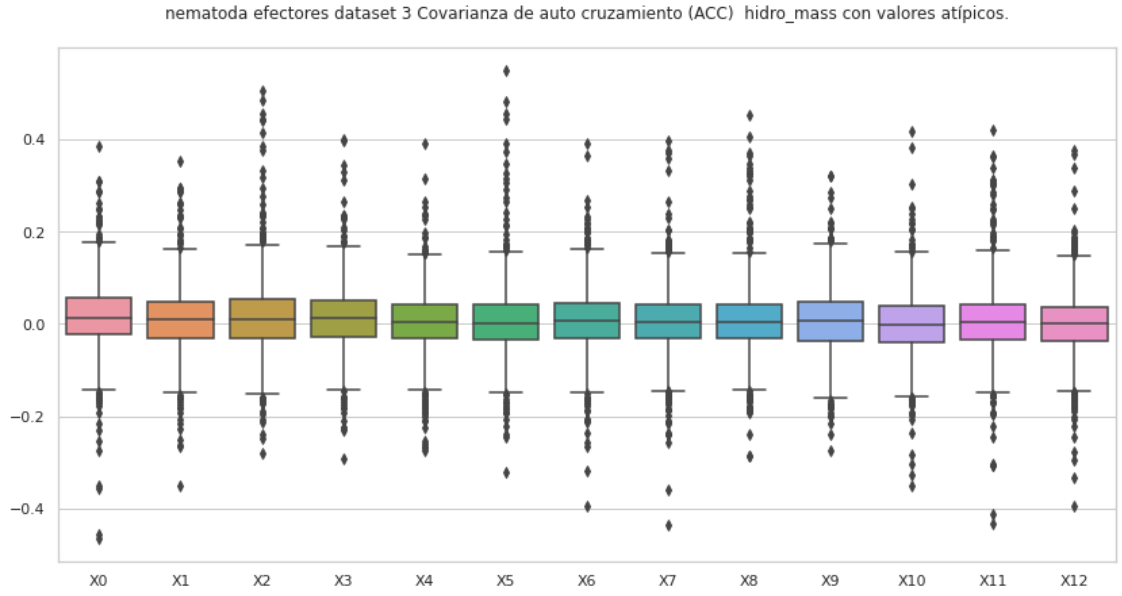
```
efectores
Covarianza de auto cruzamiento (ACC)  hidro_mass efectores nematoda dataset 3,
con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.079293  0.063434  0.048679  0.005100  0.005796 -0.021472 -0.003543
1    0.062142  0.000370 -0.000905  0.014367  0.001473  0.160242  0.116996
2    0.071002  0.080575  0.018343 -0.046646  0.022020 -0.029680  0.055163
3    0.148351  0.033734 -0.035410  0.023460 -0.148762 -0.087082 -0.118124
4    0.130044  0.121923  0.021932  0.078560  0.055482  0.117875  0.067911
..        …         …         …         …         …         …         …
995 -0.041904 -0.018839  0.094401 -0.046842  0.080568 -0.062324  0.056777
996 -0.166697  0.173928  0.003227  0.156165 -0.077882  0.004215  0.023281
997  0.010353  0.021485  0.067200  0.028902  0.023099 -0.011267  0.018219
998  0.018160 -0.020017  0.037025  0.003775  0.130536 -0.071240 -0.116582
999 -0.017526 -0.015652 -0.032574 -0.080518 -0.035784 -0.108690  0.018563

           X7        X8        X9       X10       X11       X12        X13
0    0.008483  0.010540 -0.037415 -0.006514  0.011490  0.016921  efectores
1    0.059421 -0.046531  0.074896 -0.021676 -0.036108 -0.023796  efectores
2    0.023533 -0.015681  0.004301 -0.055699 -0.010974 -0.004525  efectores
3    0.083843  0.276603  0.116391 -0.000555 -0.106871  0.132156  efectores
4   -0.002103 -0.008577 -0.053130  0.025964 -0.006805 -0.111036  efectores
..        …         …         …         …         …         …         …
995  0.005412 -0.035754  0.040496  0.017676 -0.000239 -0.029119  efectores
996  0.019182 -0.034819  0.067893 -0.045003 -0.095281  0.056648  efectores
997  0.001931 -0.000979 -0.036352  0.097764  0.003273  0.050528  efectores
998 -0.094705  0.061620 -0.104453 -0.050190  0.009778 -0.095075  efectores
999  0.136964 -0.019033  0.020948 -0.055601  0.009024  0.006125  efectores

[1000 rows x 14 columns]


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

                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.015663     0.009166     0.013628     0.014323     0.003515
```

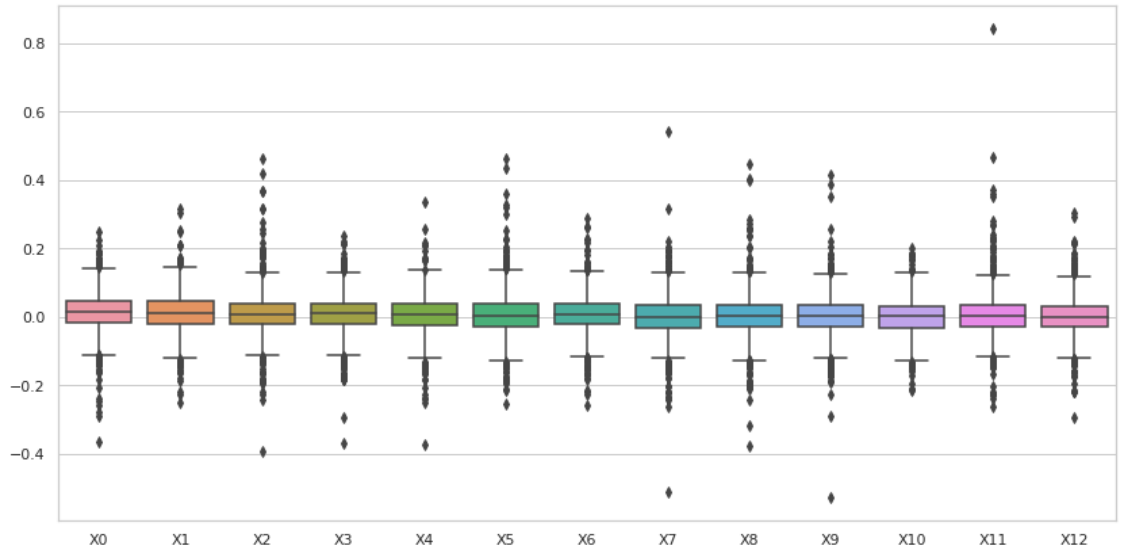
```
std       0.077600     0.073663     0.082006     0.071561     0.070453
min      -0.464282    -0.350415    -0.278409    -0.291165    -0.272926
25%      -0.022603    -0.029953    -0.030301    -0.027466    -0.031896
50%       0.013718     0.009169     0.010367     0.014178     0.004801
75%       0.057494     0.048088     0.052750     0.051131     0.041976
max       0.385999     0.352358     0.503779     0.399967     0.390112

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.006756     0.007027     0.004310     0.008204     0.005271
std       0.078907     0.072302     0.075138     0.079515     0.071512
min      -0.321258    -0.394679    -0.435204    -0.285921    -0.273369
25%      -0.035340    -0.031513    -0.032048    -0.031470    -0.037268
50%       0.002662     0.007262     0.004321     0.003923     0.006606
75%       0.042259     0.046230     0.043129     0.043221     0.046760
max       0.548424     0.390188     0.395076     0.452031     0.320611

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.000050     0.006681     0.001138
std       0.072099     0.078274     0.074129
min      -0.350280    -0.431778    -0.395298
25%      -0.039157    -0.034153    -0.035969
50%      -0.000817     0.005398     0.001902
75%       0.039268     0.043528     0.037795
max       0.416941     0.418933     0.376127



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    0.061866  0.018128 -0.043203  0.046485  0.036258 -0.037896 -0.070244
1   -0.007655  0.027260 -0.045591 -0.040476  0.024649  0.037387  0.040304
2   -0.061851  0.042982  0.054552 -0.125135  0.003719  0.013687 -0.086801
3   -0.018009 -0.029244  0.087914 -0.184167  0.053417  0.223143  0.156171
4   -0.011995  0.049566 -0.022653  0.014553  0.001848 -0.002898  0.010011
..        …         …         …         …         …         …         …
995  0.029236  0.051410  0.040596  0.051009 -0.009188  0.014661  0.008969
996  0.048233 -0.100025 -0.007612 -0.039678  0.073642 -0.025328 -0.058726
997  0.008117  0.119898  0.039441  0.043064  0.053708  0.091717  0.116420
998  0.038425 -0.026574  0.000759 -0.042966 -0.052579 -0.060460  0.029184
999 -0.076480  0.001755  0.093136  0.042996 -0.009441 -0.041141 -0.035826

           X7        X8        X9       X10       X11       X12           X13
```

```
0    0.061221  0.036713 -0.007302 -0.082094 -0.070626 -0.065722  no_efectores
1    0.024465 -0.063922 -0.053530 -0.041946 -0.048256 -0.058116  no_efectores
2   -0.080023 -0.092710 -0.016815 -0.120135 -0.031099  0.155410  no_efectores
3   -0.136739  0.104865 -0.092145 -0.039062  0.038934 -0.042937  no_efectores
4   -0.052516 -0.023240  0.011489  0.064084  0.015504  0.034445  no_efectores
..        …       …       …       …       …       …           …
995  0.038133  0.037658  0.019979 -0.014609  0.033729  0.001256  no_efectores
996  0.042079  0.014723 -0.004514  0.024921  0.042729 -0.039386  no_efectores
997  0.090945  0.069188  0.027335  0.069766  0.004311  0.116166  no_efectores
998 -0.029150 -0.007797  0.183314  0.003609  0.095368  0.104616  no_efectores
999 -0.027932  0.014509  0.082549 -0.104665  0.107159 -0.012294  no_efectores

[1000 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  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.012315     0.010289     0.009435     0.007773     0.006280
std       0.060932     0.060238     0.066026     0.058949     0.060385
min      -0.363665    -0.250625    -0.393581    -0.367323    -0.373411
25%      -0.017355    -0.021750    -0.022635    -0.022018    -0.025404
50%       0.014051     0.010053     0.007276     0.010444     0.007789
75%       0.045742     0.045045     0.038409     0.039496     0.039655
max       0.246475     0.317092     0.460662     0.234807     0.334583

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.006026     0.006708    -0.000359     0.002683     0.001809
std       0.067014     0.060001     0.064807     0.068961     0.063767
min      -0.254112    -0.258973    -0.510418    -0.376682    -0.527435
25%      -0.028299    -0.022717    -0.031018    -0.029868    -0.028585
50%       0.004505     0.008173     0.000168     0.001357     0.002908
75%       0.039167     0.039493     0.033346     0.034789     0.033187
max       0.460816     0.286504     0.539659     0.444341     0.412963

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.000586     0.006159     0.000499
std       0.056521     0.070905     0.058514
min      -0.215076    -0.262492    -0.295124
25%      -0.032278    -0.026772    -0.030599
50%       0.001945     0.003273     0.000105
75%       0.032406     0.033599     0.029614
max       0.202329     0.841763     0.302482
```


nematoda efectores dataset 3 Covarianza de auto cruzamiento (ACC) hidro_mass 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_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

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	0.079293	0.063434	0.048679	0.005100	0.005796	-0.021472	-0.003543	
1	0.062142	0.000370	-0.000905	0.014367	0.001473	0.160242	0.116996	
2	0.071002	0.080575	0.018343	-0.046646	0.022020	-0.029680	0.055163	
4	0.130044	0.121923	0.021932	0.078560	0.055482	0.117875	0.067911	
5	0.097107	0.021204	-0.048569	-0.033159	0.045475	0.067008	0.044175	
..	…	…	…	…	…	…	…	
995	-0.041904	-0.018839	0.094401	-0.046842	0.080568	-0.062324	0.056777	
996	-0.166697	0.173928	0.003227	0.156165	-0.077882	0.004215	0.023281	
997	0.010353	0.021485	0.067200	0.028902	0.023099	-0.011267	0.018219	
998	0.018160	-0.020017	0.037025	0.003775	0.130536	-0.071240	-0.116582	
999	-0.017526	-0.015652	-0.032574	-0.080518	-0.035784	-0.108690	0.018563	

	X7	X8	X9	X10	X11	X12	X13
0	0.008483	0.010540	-0.037415	-0.006514	0.011490	0.016921	efectores
1	0.059421	-0.046531	0.074896	-0.021676	-0.036108	-0.023796	efectores
2	0.023533	-0.015681	0.004301	-0.055699	-0.010974	-0.004525	efectores
4	-0.002103	-0.008577	-0.053130	0.025964	-0.006805	-0.111036	efectores
5	0.044293	0.063129	0.035886	0.092834	0.012453	-0.012802	efectores
..	…	…	…	…	…	…	
995	0.005412	-0.035754	0.040496	0.017676	-0.000239	-0.029119	efectores
996	0.019182	-0.034819	0.067893	-0.045003	-0.095281	0.056648	efectores
997	0.001931	-0.000979	-0.036352	0.097764	0.003273	0.050528	efectores
998	-0.094705	0.061620	-0.104453	-0.050190	0.009778	-0.095075	efectores
999	0.136964	-0.019033	0.020948	-0.055601	0.009024	0.006125	efectores

[906 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	906.000000	906.000000	906.000000	906.000000	906.000000	906.000000	
mean	0.016690	0.007298	0.007895	0.011851	0.005655	0.003361	
std	0.063576	0.061032	0.063887	0.059386	0.058740	0.060929	
min	-0.216679	-0.207806	-0.213517	-0.192653	-0.191352	-0.221407	
25%	-0.019556	-0.026238	-0.029266	-0.026691	-0.028899	-0.033167	
50%	0.014244	0.009169	0.009231	0.013524	0.005263	0.002087	
75%	0.055439	0.044655	0.048187	0.046339	0.041103	0.039546	
max	0.231853	0.192957	0.233618	0.225383	0.198025	0.242088	

```
                X6           X7           X8           X9          X10          X11  \
count   906.000000   906.000000   906.000000   906.000000   906.000000   906.000000
mean      0.006621     0.003325     0.002948     0.004460    -0.000930     0.003297
std       0.060527     0.061110     0.062962     0.062845     0.059622     0.060775
min      -0.208429    -0.214316    -0.190182    -0.192592    -0.205639    -0.221652
25%      -0.029308    -0.030048    -0.031152    -0.036267    -0.037043    -0.033564
50%       0.007262     0.004196     0.002947     0.006000    -0.001297     0.004089
75%       0.044195     0.039038     0.039656     0.043766     0.036419     0.039424
max       0.219249     0.203233     0.220660     0.219346     0.207101     0.225869

               X12
count   906.000000
mean      0.001258
std       0.063519
min      -0.220762
25%      -0.032342
50%       0.001671
75%       0.035952
max       0.202440
```

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  \
0    0.061866  0.018128 -0.043203  0.046485  0.036258 -0.037896 -0.070244
1   -0.007655  0.027260 -0.045591 -0.040476  0.024649  0.037387  0.040304
2   -0.061851  0.042982  0.054552 -0.125135  0.003719  0.013687 -0.086801
4   -0.011995  0.049566 -0.022653  0.014553  0.001848 -0.002898  0.010011
5   -0.016728 -0.021714 -0.022491  0.010269 -0.005957 -0.019291  0.035690
..        …         …         …         …         …         …         …
995  0.029236  0.051410  0.040596  0.051009 -0.009188  0.014661  0.008969
996  0.048233 -0.100025 -0.007612 -0.039678  0.073642 -0.025328 -0.058726
997  0.008117  0.119898  0.039441  0.043064  0.053708  0.091717  0.116420
998  0.038425 -0.026574  0.000759 -0.042966 -0.052579 -0.060460  0.029184
999 -0.076480  0.001755  0.093136  0.042996 -0.009441 -0.041141 -0.035826

           X7        X8        X9       X10       X11       X12           X13
0    0.061221  0.036713 -0.007302 -0.082094 -0.070626 -0.065722  no_efectores
1    0.024465 -0.063922 -0.053530 -0.041946 -0.048256 -0.058116  no_efectores
2   -0.080023 -0.092710 -0.016815 -0.120135 -0.031099  0.155410  no_efectores
4   -0.052516 -0.023240  0.011489  0.064084  0.015504  0.034445  no_efectores
5   -0.018538  0.015128  0.057819  0.048646  0.035941 -0.003161  no_efectores
..        …         …         …         …         …         …           …
995  0.038133  0.037658  0.019979 -0.014609  0.033729  0.001256  no_efectores
```

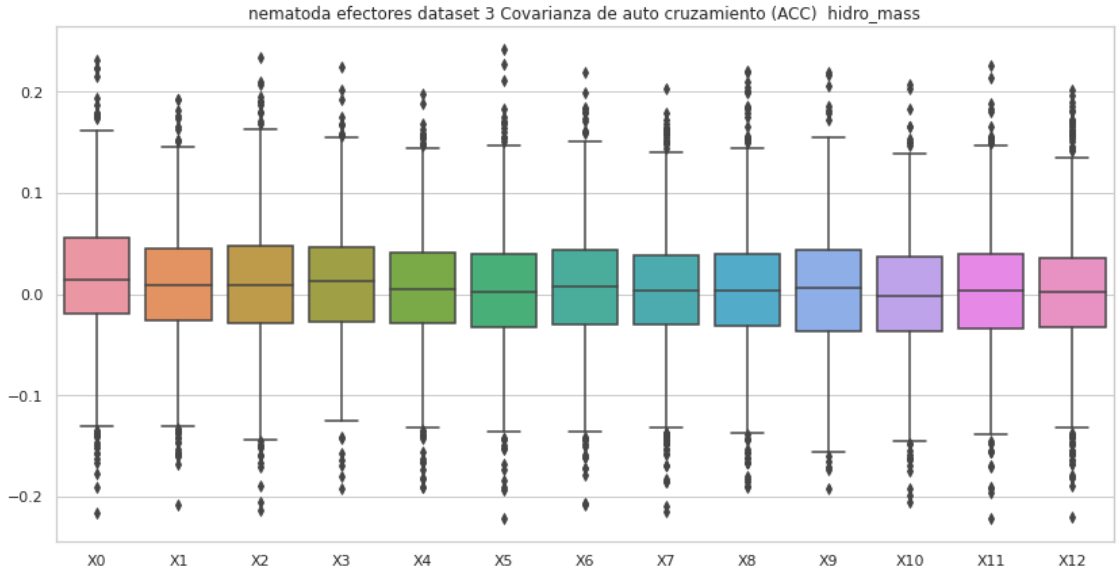
```
996  0.042079  0.014723 -0.004514  0.024921  0.042729 -0.039386  no_efectores
997  0.090945  0.069188  0.027335  0.069766  0.004311  0.116166  no_efectores
998 -0.029150 -0.007797  0.183314  0.003609  0.095368  0.104616  no_efectores
999 -0.027932  0.014509  0.082549 -0.104665  0.107159 -0.012294  no_efectores

[911 rows x 14 columns]
```

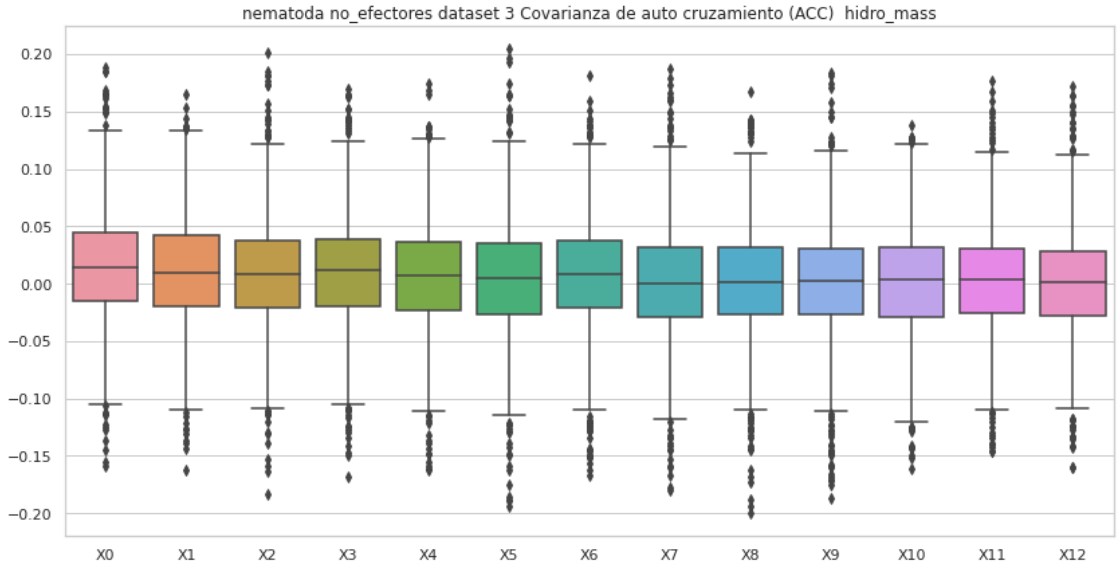
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores nematoda dataset 3, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	911.000000	911.000000	911.000000	911.000000	911.000000	911.000000	
mean	0.014115	0.009835	0.009255	0.009290	0.006704	0.004249	
std	0.051218	0.050591	0.050356	0.050182	0.050497	0.054162	
min	-0.159155	-0.162656	-0.183301	-0.168292	-0.161848	-0.194402	
25%	-0.015346	-0.019796	-0.020828	-0.019877	-0.023225	-0.026451	
50%	0.014471	0.010006	0.007963	0.011330	0.007555	0.004437	
75%	0.044640	0.042106	0.037781	0.038327	0.036621	0.034791	
max	0.188336	0.164560	0.201420	0.169091	0.174673	0.204632	

	X6	X7	X8	X9	X10	X11	\
count	911.000000	911.000000	911.000000	911.000000	911.000000	911.000000	
mean	0.006569	0.000677	0.001131	0.001425	0.001872	0.002647	
std	0.051409	0.052335	0.052110	0.051954	0.050365	0.049363	
min	-0.166997	-0.180302	-0.199800	-0.187172	-0.160727	-0.145866	
25%	-0.021053	-0.029155	-0.026415	-0.026435	-0.029024	-0.025805	
50%	0.007792	0.000570	0.001310	0.002785	0.003319	0.003092	
75%	0.038039	0.031401	0.031108	0.030813	0.031772	0.031075	
max	0.180850	0.187730	0.167174	0.183314	0.137953	0.176499	

	X12
count	911.000000
mean	0.001548
std	0.049187
min	-0.160543
25%	-0.027750
50%	0.001101
75%	0.028751
max	0.172542





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    0.079293  0.063434  0.048679  0.005100  0.005796 -0.021472 -0.003543
1    0.062142  0.000370 -0.000905  0.014367  0.001473  0.160242  0.116996
2    0.071002  0.080575  0.018343 -0.046646  0.022020 -0.029680  0.055163
3    0.148351  0.033734 -0.035410  0.023460 -0.148762 -0.087082 -0.118124
4    0.130044  0.121923  0.021932  0.078560  0.055482  0.117875  0.067911
..        …         …         …         …         …         …         …
995 -0.041904 -0.018839  0.094401 -0.046842  0.080568 -0.062324  0.056777
996 -0.166697  0.173928  0.003227  0.156165 -0.077882  0.004215  0.023281
997  0.010353  0.021485  0.067200  0.028902  0.023099 -0.011267  0.018219
998  0.018160 -0.020017  0.037025  0.003775  0.130536 -0.071240 -0.116582
999 -0.017526 -0.015652 -0.032574 -0.080518 -0.035784 -0.108690  0.018563

           X7        X8        X9       X10       X11       X12        X13
0    0.008483  0.010540 -0.037415 -0.006514  0.011490  0.016921  efectores
1    0.059421 -0.046531  0.074896 -0.021676 -0.036108 -0.023796  efectores
2    0.023533 -0.015681  0.004301 -0.055699 -0.010974 -0.004525  efectores
```

```
3    0.083843  0.276603  0.116391 -0.000555 -0.106871  0.132156  efectores
4   -0.002103 -0.008577 -0.053130  0.025964 -0.006805 -0.111036  efectores
..        …       …       …       …       …       …        …
995  0.005412 -0.035754  0.040496  0.017676 -0.000239 -0.029119  efectores
996  0.019182 -0.034819  0.067893 -0.045003 -0.095281  0.056648  efectores
997  0.001931 -0.000979 -0.036352  0.097764  0.003273  0.050528  efectores
998 -0.094705  0.061620 -0.104453 -0.050190  0.009778 -0.095075  efectores
999  0.136964 -0.019033  0.020948 -0.055601  0.009024  0.006125  efectores

[1000 rows x 14 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.015663     0.009166     0.013628     0.014323     0.003515
std       0.077600     0.073663     0.082006     0.071561     0.070453
min      -0.464282    -0.350415    -0.278409    -0.291165    -0.272926
25%      -0.022603    -0.029953    -0.030301    -0.027466    -0.031896
50%       0.013718     0.009169     0.010367     0.014178     0.004801
75%       0.057494     0.048088     0.052750     0.051131     0.041976
max       0.385999     0.352358     0.503779     0.399967     0.390112

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.006756     0.007027     0.004310     0.008204     0.005271
std       0.078907     0.072302     0.075138     0.079515     0.071512
min      -0.321258    -0.394679    -0.435204    -0.285921    -0.273369
25%      -0.035340    -0.031513    -0.032048    -0.031470    -0.037268
50%       0.002662     0.007262     0.004321     0.003923     0.006606
75%       0.042259     0.046230     0.043129     0.043221     0.046760
max       0.548424     0.390188     0.395076     0.452031     0.320611

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.000050     0.006681     0.001138
std       0.072099     0.078274     0.074129
min      -0.350280    -0.431778    -0.395298
25%      -0.039157    -0.034153    -0.035969
50%      -0.000817     0.005398     0.001902
75%       0.039268     0.043528     0.037795
max       0.416941     0.418933     0.376127
```

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	0.061866	0.018128	-0.043203	0.046485	0.036258	-0.037896	-0.070244	
1	-0.007655	0.027260	-0.045591	-0.040476	0.024649	0.037387	0.040304	
2	-0.061851	0.042982	0.054552	-0.125135	0.003719	0.013687	-0.086801	
3	-0.018009	-0.029244	0.087914	-0.184167	0.053417	0.223143	0.156171	
4	-0.011995	0.049566	-0.022653	0.014553	0.001848	-0.002898	0.010011	
..	…	…	…	…	…	…	…	
995	0.029236	0.051410	0.040596	0.051009	-0.009188	0.014661	0.008969	
996	0.048233	-0.100025	-0.007612	-0.039678	0.073642	-0.025328	-0.058726	
997	0.008117	0.119898	0.039441	0.043064	0.053708	0.091717	0.116420	
998	0.038425	-0.026574	0.000759	-0.042966	-0.052579	-0.060460	0.029184	
999	-0.076480	0.001755	0.093136	0.042996	-0.009441	-0.041141	-0.035826	

	X7	X8	X9	X10	X11	X12	X13
0	0.061221	0.036713	-0.007302	-0.082094	-0.070626	-0.065722	no_efectores
1	0.024465	-0.063922	-0.053530	-0.041946	-0.048256	-0.058116	no_efectores
2	-0.080023	-0.092710	-0.016815	-0.120135	-0.031099	0.155410	no_efectores
3	-0.136739	0.104865	-0.092145	-0.039062	0.038934	-0.042937	no_efectores
4	-0.052516	-0.023240	0.011489	0.064084	0.015504	0.034445	no_efectores
..	…	…	…	…	…	…	…
995	0.038133	0.037658	0.019979	-0.014609	0.033729	0.001256	no_efectores
996	0.042079	0.014723	-0.004514	0.024921	0.042729	-0.039386	no_efectores
997	0.090945	0.069188	0.027335	0.069766	0.004311	0.116166	no_efectores
998	-0.029150	-0.007797	0.183314	0.003609	0.095368	0.104616	no_efectores
999	-0.027932	0.014509	0.082549	-0.104665	0.107159	-0.012294	no_efectores

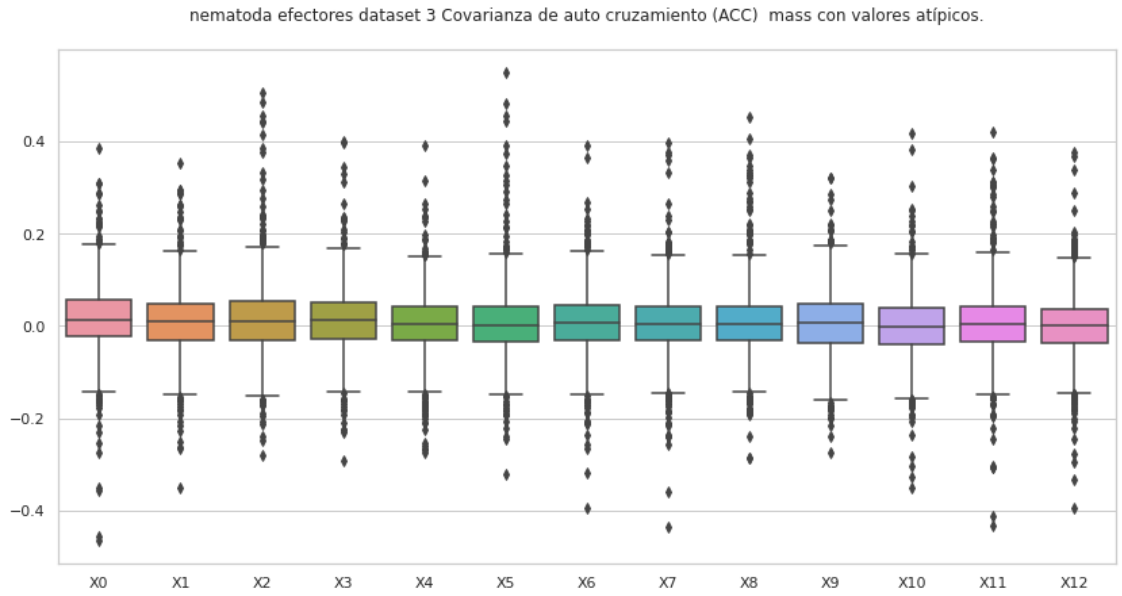
[1000 rows x 14 columns]

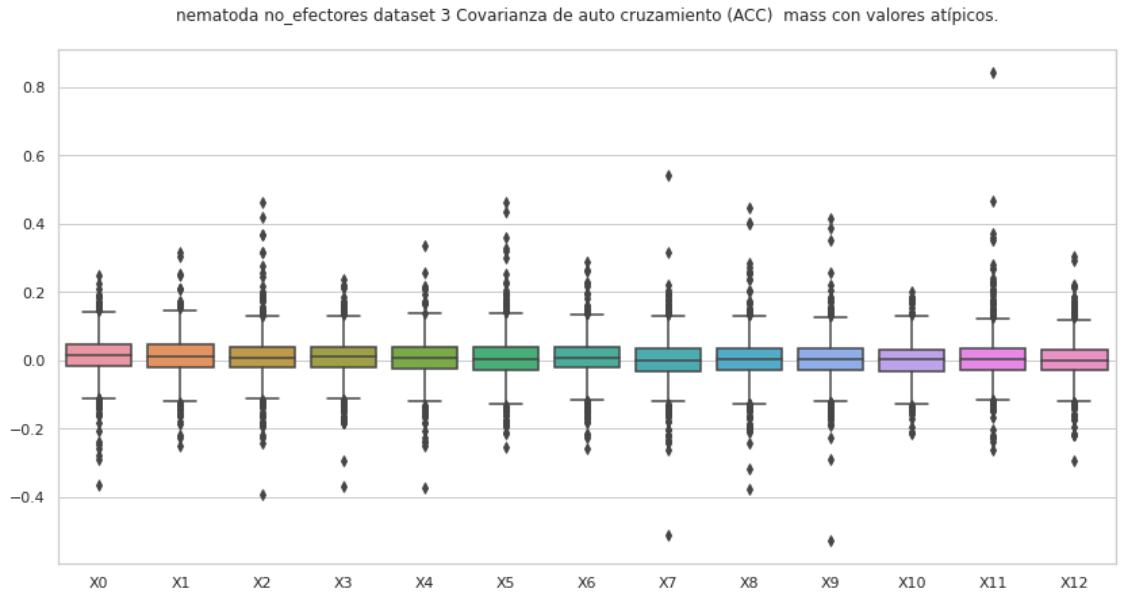
Covarianza de auto cruzamiento (ACC) mass no_efectores nematoda dataset 3, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.012315	0.010289	0.009435	0.007773	0.006280	
std	0.060932	0.060238	0.066026	0.058949	0.060385	
min	-0.363665	-0.250625	-0.393581	-0.367323	-0.373411	
25%	-0.017355	-0.021750	-0.022635	-0.022018	-0.025404	
50%	0.014051	0.010053	0.007276	0.010444	0.007789	
75%	0.045742	0.045045	0.038409	0.039496	0.039655	
max	0.246475	0.317092	0.460662	0.234807	0.334583	

	X5	X6	X7	X8	X9	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.006026	0.006708	-0.000359	0.002683	0.001809	
std	0.067014	0.060001	0.064807	0.068961	0.063767	
min	-0.254112	-0.258973	-0.510418	-0.376682	-0.527435	
25%	-0.028299	-0.022717	-0.031018	-0.029868	-0.028585	
50%	0.004505	0.008173	0.000168	0.001357	0.002908	
75%	0.039167	0.039493	0.033346	0.034789	0.033187	
max	0.460816	0.286504	0.539659	0.444341	0.412963	

	X10	X11	X12
count	1000.000000	1000.000000	1000.000000
mean	0.000586	0.006159	0.000499
std	0.056521	0.070905	0.058514
min	-0.215076	-0.262492	-0.295124
25%	-0.032278	-0.026772	-0.030599
50%	0.001945	0.003273	0.000105
75%	0.032406	0.033599	0.029614
max	0.202329	0.841763	0.302482





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    0.079293  0.063434  0.048679  0.005100  0.005796 -0.021472 -0.003543
1    0.062142  0.000370 -0.000905  0.014367  0.001473  0.160242  0.116996
2    0.071002  0.080575  0.018343 -0.046646  0.022020 -0.029680  0.055163
4    0.130044  0.121923  0.021932  0.078560  0.055482  0.117875  0.067911
5    0.097107  0.021204 -0.048569 -0.033159  0.045475  0.067008  0.044175
..        …         …         …         …         …         …         …
995 -0.041904 -0.018839  0.094401 -0.046842  0.080568 -0.062324  0.056777
996 -0.166697  0.173928  0.003227  0.156165 -0.077882  0.004215  0.023281
997  0.010353  0.021485  0.067200  0.028902  0.023099 -0.011267  0.018219
998  0.018160 -0.020017  0.037025  0.003775  0.130536 -0.071240 -0.116582
999 -0.017526 -0.015652 -0.032574 -0.080518 -0.035784 -0.108690  0.018563

           X7        X8        X9       X10       X11       X12        X13
0    0.008483  0.010540 -0.037415 -0.006514  0.011490  0.016921  efectores
1    0.059421 -0.046531  0.074896 -0.021676 -0.036108 -0.023796  efectores
2    0.023533 -0.015681  0.004301 -0.055699 -0.010974 -0.004525  efectores
4   -0.002103 -0.008577 -0.053130  0.025964 -0.006805 -0.111036  efectores
5    0.044293  0.063129  0.035886  0.092834  0.012453 -0.012802  efectores
..        …         …         …         …         …         …         …
995  0.005412 -0.035754  0.040496  0.017676 -0.000239 -0.029119  efectores
```

```
996  0.019182 -0.034819  0.067893 -0.045003 -0.095281  0.056648  efectores
997  0.001931 -0.000979 -0.036352  0.097764  0.003273  0.050528  efectores
998 -0.094705  0.061620 -0.104453 -0.050190  0.009778 -0.095075  efectores
999  0.136964 -0.019033  0.020948 -0.055601  0.009024  0.006125  efectores

[906 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	906.000000	906.000000	906.000000	906.000000	906.000000	906.000000	
mean	0.016690	0.007298	0.007895	0.011851	0.005655	0.003361	
std	0.063576	0.061032	0.063887	0.059386	0.058740	0.060929	
min	-0.216679	-0.207806	-0.213517	-0.192653	-0.191352	-0.221407	
25%	-0.019556	-0.026238	-0.029266	-0.026691	-0.028899	-0.033167	
50%	0.014244	0.009169	0.009231	0.013524	0.005263	0.002087	
75%	0.055439	0.044655	0.048187	0.046339	0.041103	0.039546	
max	0.231853	0.192957	0.233618	0.225383	0.198025	0.242088	

	X6	X7	X8	X9	X10	X11	\
count	906.000000	906.000000	906.000000	906.000000	906.000000	906.000000	
mean	0.006621	0.003325	0.002948	0.004460	-0.000930	0.003297	
std	0.060527	0.061110	0.062962	0.062845	0.059622	0.060775	
min	-0.208429	-0.214316	-0.190182	-0.192592	-0.205639	-0.221652	
25%	-0.029308	-0.030048	-0.031152	-0.036267	-0.037043	-0.033564	
50%	0.007262	0.004196	0.002947	0.006000	-0.001297	0.004089	
75%	0.044195	0.039038	0.039656	0.043766	0.036419	0.039424	
max	0.219249	0.203233	0.220660	0.219346	0.207101	0.225869	

	X12
count	906.000000
mean	0.001258
std	0.063519
min	-0.220762
25%	-0.032342
50%	0.001671
75%	0.035952
max	0.202440

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

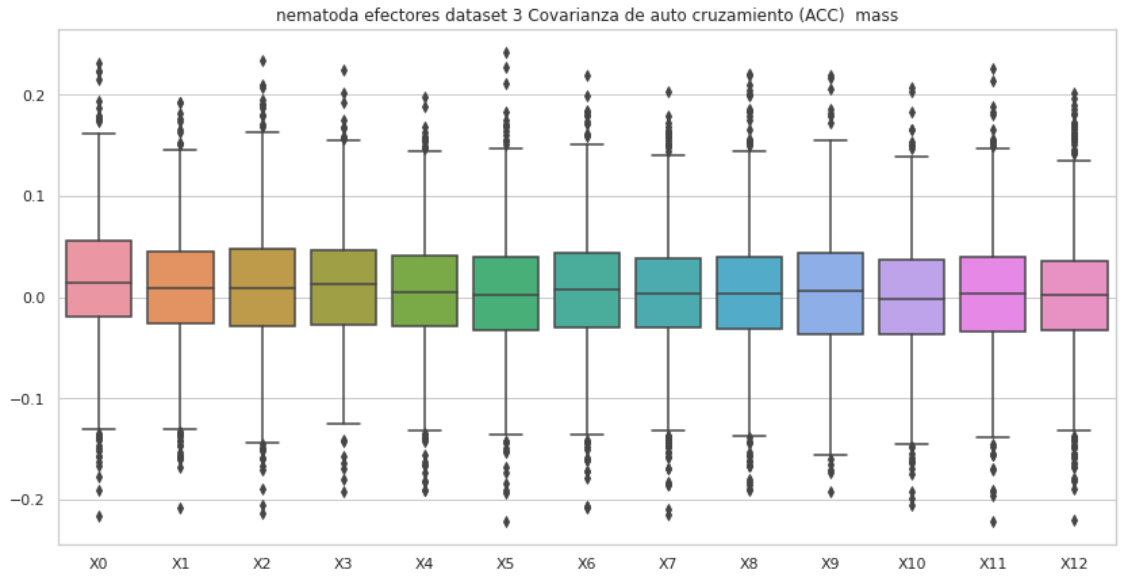
```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.061866  0.018128 -0.043203  0.046485  0.036258 -0.037896 -0.070244
1   -0.007655  0.027260 -0.045591 -0.040476  0.024649  0.037387  0.040304
2   -0.061851  0.042982  0.054552 -0.125135  0.003719  0.013687 -0.086801
4   -0.011995  0.049566 -0.022653  0.014553  0.001848 -0.002898  0.010011
5   -0.016728 -0.021714 -0.022491  0.010269 -0.005957 -0.019291  0.035690
..        …         …         …         …         …         …
995  0.029236  0.051410  0.040596  0.051009 -0.009188  0.014661  0.008969
996  0.048233 -0.100025 -0.007612 -0.039678  0.073642 -0.025328 -0.058726
997  0.008117  0.119898  0.039441  0.043064  0.053708  0.091717  0.116420
998  0.038425 -0.026574  0.000759 -0.042966 -0.052579 -0.060460  0.029184
999 -0.076480  0.001755  0.093136  0.042996 -0.009441 -0.041141 -0.035826

           X7        X8        X9       X10       X11       X12           X13
0    0.061221  0.036713 -0.007302 -0.082094 -0.070626 -0.065722  no_efectores
1    0.024465 -0.063922 -0.053530 -0.041946 -0.048256 -0.058116  no_efectores
2   -0.080023 -0.092710 -0.016815 -0.120135 -0.031099  0.155410  no_efectores
4   -0.052516 -0.023240  0.011489  0.064084  0.015504  0.034445  no_efectores
5   -0.018538  0.015128  0.057819  0.048646  0.035941 -0.003161  no_efectores
..        …         …         …         …         …         …
995  0.038133  0.037658  0.019979 -0.014609  0.033729  0.001256  no_efectores
996  0.042079  0.014723 -0.004514  0.024921  0.042729 -0.039386  no_efectores
997  0.090945  0.069188  0.027335  0.069766  0.004311  0.116166  no_efectores
998 -0.029150 -0.007797  0.183314  0.003609  0.095368  0.104616  no_efectores
999 -0.027932  0.014509  0.082549 -0.104665  0.107159 -0.012294  no_efectores

[911 rows x 14 columns]
```

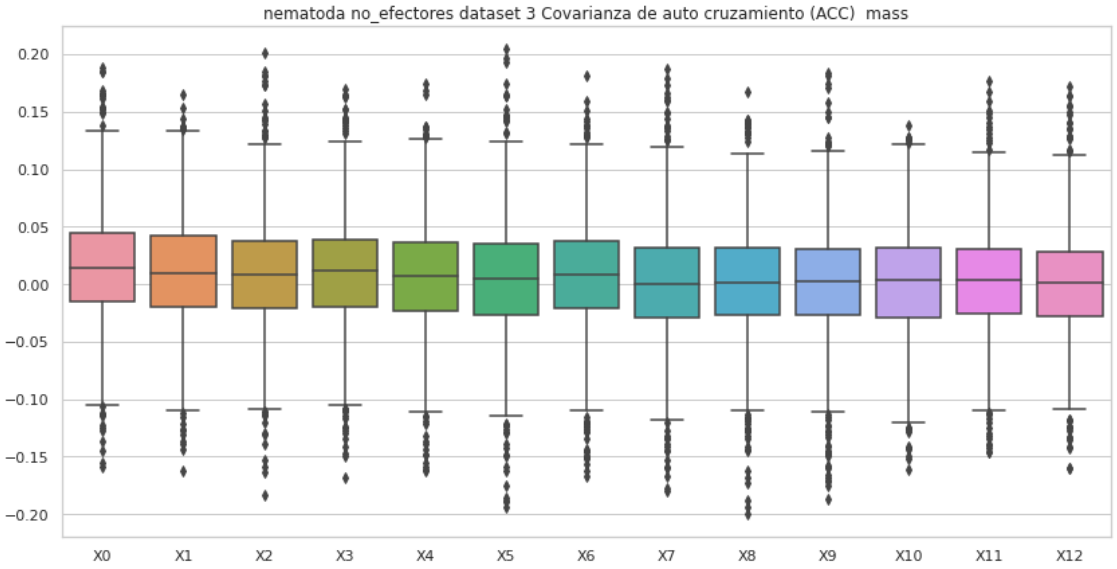
Covarianza de auto cruzamiento (ACC) mass no_efectores nematoda dataset 3, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  911.000000  911.000000  911.000000  911.000000  911.000000  911.000000
mean     0.014115    0.009835    0.009255    0.009290    0.006704    0.004249
std      0.051218    0.050591    0.050356    0.050182    0.050497    0.054162
min     -0.159155   -0.162656   -0.183301   -0.168292   -0.161848   -0.194402
25%     -0.015346   -0.019796   -0.020828   -0.019877   -0.023225   -0.026451
50%      0.014471    0.010006    0.007963    0.011330    0.007555    0.004437
75%      0.044640    0.042106    0.037781    0.038327    0.036621    0.034791
max      0.188336    0.164560    0.201420    0.169091    0.174673    0.204632

                X6          X7          X8          X9         X10         X11  \
count  911.000000  911.000000  911.000000  911.000000  911.000000  911.000000
mean     0.006569    0.000677    0.001131    0.001425    0.001872    0.002647
std      0.051409    0.052335    0.052110    0.051954    0.050365    0.049363
min     -0.166997   -0.180302   -0.199800   -0.187172   -0.160727   -0.145866
```

25%	-0.021053	-0.029155	-0.026415	-0.026435	-0.029024	-0.025805
50%	0.007792	0.000570	0.001310	0.002785	0.003319	0.003092
75%	0.038039	0.031401	0.031108	0.030813	0.031772	0.031075
max	0.180850	0.187730	0.167174	0.183314	0.137953	0.176499

	X12
count	911.000000
mean	0.001548
std	0.049187
min	-0.160543
25%	-0.027750
50%	0.001101
75%	0.028751
max	0.172542





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))
```

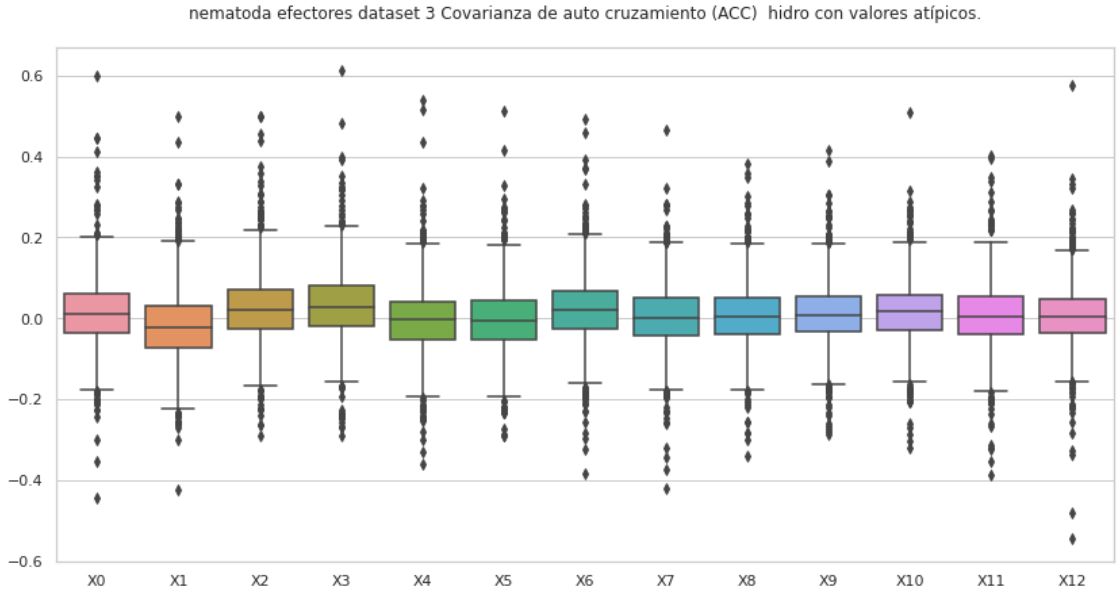
```
efectores
Covarianza de auto cruzamiento (ACC)  hidro efectores nematoda dataset 3, con
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.022621 -0.054573 -0.024733  0.021480  0.041785  0.119181 -0.028403
1    0.109106  0.006863  0.093888  0.135450  0.091751 -0.033773  0.073244
2   -0.078946 -0.071251 -0.036030  0.100100 -0.010995  0.006796  0.020599
3   -0.007631 -0.064787  0.074098 -0.121072 -0.165068  0.017188 -0.160806
4   -0.092814 -0.009529 -0.091094  0.101692 -0.070141 -0.100347  0.101009
..        …         …         …         …         …         …         …
995 -0.130955 -0.046789 -0.076787  0.032969 -0.018424 -0.070920  0.112522
996 -0.076022  0.286885 -0.064062 -0.011440 -0.153276 -0.130346  0.082968
997  0.023002 -0.071761  0.017088  0.054920  0.083202 -0.062007  0.037778
998  0.121520 -0.007196  0.029746  0.033286 -0.111176 -0.054503 -0.054766
999 -0.012251 -0.176094  0.087300 -0.008534 -0.063393  0.011561  0.046467

           X7        X8        X9       X10       X11       X12         X13
0    0.007862  0.025040  0.058704  0.029494  0.080642  0.049721  efectores
1    0.096818  0.041748  0.023701  0.126412  0.094129  0.065181  efectores
2    0.015595  0.003201  0.007418  0.045705  0.031637 -0.021779  efectores
3    0.011772  0.164905 -0.113203  0.111599  0.226474 -0.018040  efectores
4    0.007192 -0.061431 -0.184858  0.017881  0.143599 -0.091281  efectores
..        …         …         …         …         …         …         …
995 -0.108332  0.058022 -0.008071  0.026399  0.005106 -0.041289  efectores
996 -0.229670  0.358040  0.017517  0.083367 -0.001341  0.092912  efectores
997  0.049954  0.002458 -0.062539  0.001327  0.048768 -0.004326  efectores
998 -0.099849 -0.054014 -0.021885  0.140258  0.029983 -0.081170  efectores
999  0.072475 -0.014381 -0.108784  0.043180  0.118970 -0.065968  efectores

[1000 rows x 14 columns]


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

                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.013795    -0.018975     0.025626     0.030154    -0.001500
```

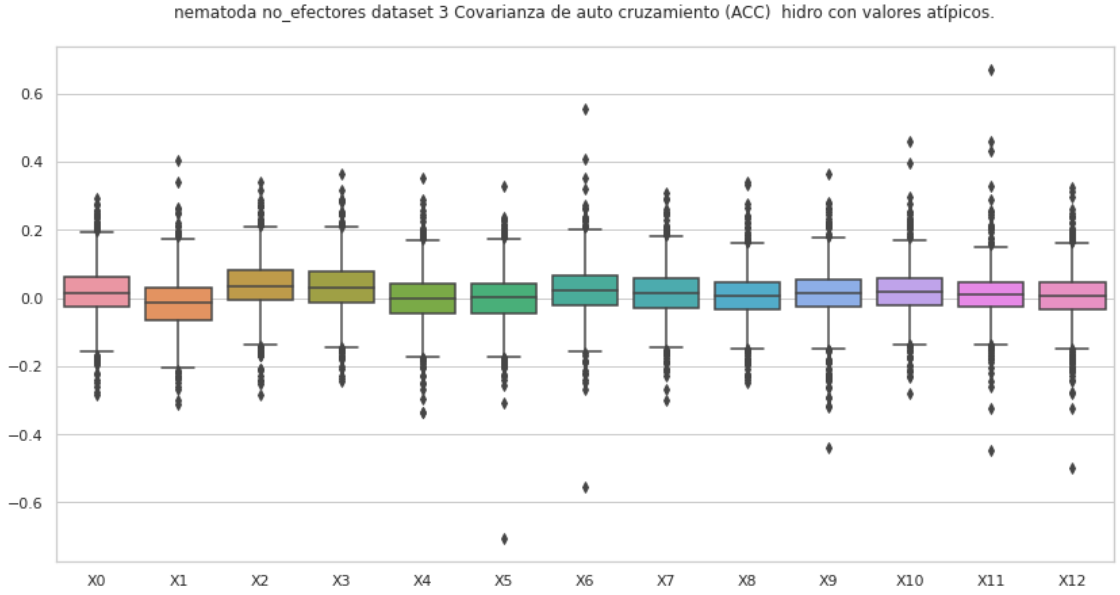
```
std       0.089912     0.093014     0.089385     0.089538     0.089591
min      -0.443458    -0.424697    -0.289157    -0.290954    -0.360247
25%      -0.035549    -0.073465    -0.026001    -0.018405    -0.052044
50%       0.011372    -0.022300     0.019727     0.027645    -0.003948
75%       0.060486     0.032054     0.071761     0.080686     0.042131
max       0.598384     0.500180     0.500164     0.612327     0.539216

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean     -0.002500     0.020492     0.003387     0.006184     0.010658
std       0.086993     0.088190     0.083202     0.082053     0.081100
min      -0.292019    -0.385510    -0.421137    -0.339646    -0.286300
25%      -0.054548    -0.027934    -0.042193    -0.039869    -0.032862
50%      -0.005431     0.019799     0.001160     0.004850     0.007803
75%       0.043952     0.066500     0.049374     0.050543     0.054590
max       0.512620     0.492234     0.466391     0.381776     0.414649

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.015619     0.006303     0.007012
std       0.081701     0.086407     0.084318
min      -0.320766    -0.386000    -0.544771
25%      -0.030107    -0.040964    -0.034997
50%       0.016584     0.005124     0.003720
75%       0.058920     0.053137     0.046362
max       0.509884     0.400312     0.574676



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   -0.020404 -0.053257  0.011565  0.003321  0.069203 -0.097557  0.008840
1    0.063791  0.022715  0.088888  0.035832  0.028094 -0.017482 -0.033476
2    0.006031 -0.210807  0.067938 -0.207992 -0.334380 -0.007471  0.167258
3   -0.015330 -0.068761 -0.011889  0.017598  0.022257  0.021424 -0.031142
4    0.015292 -0.072735 -0.044196  0.033263 -0.001761 -0.026991 -0.018477
..        …       …       …       …       …       …       …
995 -0.034820 -0.050290 -0.016702 -0.016791 -0.004308 -0.018242 -0.004243
996 -0.004687  0.051385 -0.011443 -0.030967 -0.107011  0.017471 -0.012289
997  0.013404  0.017955 -0.008811 -0.056886 -0.035024 -0.072479 -0.048617
998 -0.043220 -0.114652  0.106121  0.079009 -0.089256 -0.098366  0.029520
999  0.043002  0.061845  0.137073 -0.127840  0.025935  0.011094  0.036467

           X7        X8        X9       X10       X11       X12           X13
```

```
0    0.136759 -0.050590  0.008372 -0.091070  0.009445  0.047297  no_efectores
1    0.009922  0.022545 -0.071375 -0.030974 -0.081379  0.006693  no_efectores
2    0.100380  0.332513 -0.026919  0.064595  0.102997 -0.276807  no_efectores
3    0.022900  0.128080 -0.017030  0.011243  0.041952  0.037237  no_efectores
4   -0.082702 -0.015843 -0.029310  0.047563  0.051515 -0.041295  no_efectores
..        …       …       …       …       …       …           …
995 -0.001702 -0.032401  0.015217  0.042429  0.047680 -0.014590  no_efectores
996 -0.031146  0.119650 -0.005811 -0.116934 -0.018177  0.051946  no_efectores
997  0.002698 -0.005163  0.061708  0.026008  0.026813  0.034581  no_efectores
998 -0.078641 -0.013639 -0.010675  0.042416 -0.097726  0.066928  no_efectores
999  0.162807  0.037544  0.092748 -0.033703 -0.110642 -0.047257  no_efectores

[1000 rows x 14 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.016975    -0.015874     0.034500     0.031903    -0.001080
std       0.077985     0.082373     0.077666     0.077031     0.077019
min      -0.284679    -0.313901    -0.285411    -0.245039    -0.337473
25%      -0.026526    -0.064979    -0.007100    -0.012297    -0.044471
50%       0.014120    -0.013346     0.033021     0.031314    -0.001086
75%       0.062690     0.031099     0.080150     0.076225     0.041981
max       0.293495     0.403721     0.339765     0.366168     0.353605

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean     -0.000291     0.023207     0.015248     0.007616     0.011842
std       0.078669     0.080204     0.074125     0.071919     0.077367
min      -0.706073    -0.555538    -0.300002    -0.250288    -0.439802
25%      -0.043527    -0.023413    -0.027623    -0.031758    -0.027100
50%       0.001013     0.021414     0.012753     0.008067     0.013854
75%       0.043557     0.067175     0.056457     0.046204     0.055883
max       0.328761     0.555077     0.309765     0.339040     0.364417

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.018086     0.010102     0.005316
std       0.072737     0.076048     0.075822
min      -0.280037    -0.445515    -0.498239
25%      -0.020371    -0.026977    -0.031985
50%       0.018397     0.008246     0.006757
75%       0.056510     0.045723     0.046612
max       0.461789     0.670491     0.323526
```


nematoda 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


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	0.022621	-0.054573	-0.024733	0.021480	0.041785	0.119181	-0.028403	
1	0.109106	0.006863	0.093888	0.135450	0.091751	-0.033773	0.073244	
2	-0.078946	-0.071251	-0.036030	0.100100	-0.010995	0.006796	0.020599	
3	-0.007631	-0.064787	0.074098	-0.121072	-0.165068	0.017188	-0.160806	
4	-0.092814	-0.009529	-0.091094	0.101692	-0.070141	-0.100347	0.101009	
..	…	…	…	…	…	…	…	
994	0.058663	-0.053529	0.021721	-0.047356	-0.104359	-0.005383	-0.063202	
995	-0.130955	-0.046789	-0.076787	0.032969	-0.018424	-0.070920	0.112522	
997	0.023002	-0.071761	0.017088	0.054920	0.083202	-0.062007	0.037778	
998	0.121520	-0.007196	0.029746	0.033286	-0.111176	-0.054503	-0.054766	
999	-0.012251	-0.176094	0.087300	-0.008534	-0.063393	0.011561	0.046467	

	X7	X8	X9	X10	X11	X12	X13
0	0.007862	0.025040	0.058704	0.029494	0.080642	0.049721	efectores
1	0.096818	0.041748	0.023701	0.126412	0.094129	0.065181	efectores
2	0.015595	0.003201	0.007418	0.045705	0.031637	-0.021779	efectores
3	0.011772	0.164905	-0.113203	0.111599	0.226474	-0.018040	efectores
4	0.007192	-0.061431	-0.184858	0.017881	0.143599	-0.091281	efectores
..	…	…	…	…	…	…	
994	-0.100800	-0.028426	0.019176	-0.105824	0.148346	-0.111380	efectores
995	-0.108332	0.058022	-0.008071	0.026399	0.005106	-0.041289	efectores
997	0.049954	0.002458	-0.062539	0.001327	0.048768	-0.004326	efectores
998	-0.099849	-0.054014	-0.021885	0.140258	0.029983	-0.081170	efectores
999	0.072475	-0.014381	-0.108784	0.043180	0.118970	-0.065968	efectores

[919 rows x 14 columns]

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

	X0	X1	X2	X3	X4	X5	\
count	919.000000	919.000000	919.000000	919.000000	919.000000	919.000000	
mean	0.011474	-0.022423	0.022752	0.027752	-0.003473	-0.005461	
std	0.074011	0.081207	0.073715	0.076149	0.077069	0.075965	
min	-0.208382	-0.271539	-0.225922	-0.237099	-0.249475	-0.234106	
25%	-0.033749	-0.072021	-0.023452	-0.017401	-0.051571	-0.052453	
50%	0.010419	-0.024216	0.019286	0.024213	-0.004557	-0.007422	
75%	0.058572	0.026766	0.067699	0.076504	0.038080	0.040961	
max	0.282332	0.243484	0.265264	0.279640	0.266447	0.243373	

```
                X6           X7           X8           X9          X10          X11  \
count   919.000000   919.000000   919.000000   919.000000   919.000000   919.000000
mean      0.019854     0.003300     0.004100     0.008539     0.014945     0.003764
std       0.075338     0.071620     0.070537     0.071349     0.071538     0.074227
min      -0.230691    -0.245448    -0.208084    -0.198095    -0.207271    -0.222467
25%      -0.025191    -0.039645    -0.038659    -0.032117    -0.028094    -0.039892
50%       0.019771     0.001662     0.003405     0.007281     0.015754     0.003928
75%       0.063267     0.045991     0.046644     0.050808     0.056365     0.048266
max       0.277249     0.229175     0.247649     0.252707     0.256146     0.263536

               X12
count   919.000000
mean      0.007057
std       0.070619
min      -0.234220
25%      -0.033390
50%       0.004440
75%       0.044881
max       0.244081
```

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  \
0    -0.020404 -0.053257  0.011565  0.003321  0.069203 -0.097557  0.008840
1     0.063791  0.022715  0.088888  0.035832  0.028094 -0.017482 -0.033476
3    -0.015330 -0.068761 -0.011889  0.017598  0.022257  0.021424 -0.031142
4     0.015292 -0.072735 -0.044196  0.033263 -0.001761 -0.026991 -0.018477
5     0.062127 -0.014926 -0.023398  0.031076 -0.006118  0.052388  0.006291
..         …        …         …         …         …         …         …
995  -0.034820 -0.050290 -0.016702 -0.016791 -0.004308 -0.018242 -0.004243
996  -0.004687  0.051385 -0.011443 -0.030967 -0.107011  0.017471 -0.012289
997   0.013404  0.017955 -0.008811 -0.056886 -0.035024 -0.072479 -0.048617
998  -0.043220 -0.114652  0.106121  0.079009 -0.089256 -0.098366  0.029520
999   0.043002  0.061845  0.137073 -0.127840  0.025935  0.011094  0.036467

            X7        X8        X9       X10       X11       X12           X13
0     0.136759 -0.050590  0.008372 -0.091070  0.009445  0.047297  no_efectores
1     0.009922  0.022545 -0.071375 -0.030974 -0.081379  0.006693  no_efectores
3     0.022900  0.128080 -0.017030  0.011243  0.041952  0.037237  no_efectores
4    -0.082702 -0.015843 -0.029310  0.047563  0.051515 -0.041295  no_efectores
5     0.094241  0.076127  0.055865  0.041638 -0.005134 -0.011842  no_efectores
..         …         …         …         …         …         …           …
995  -0.001702 -0.032401  0.015217  0.042429  0.047680 -0.014590  no_efectores
```

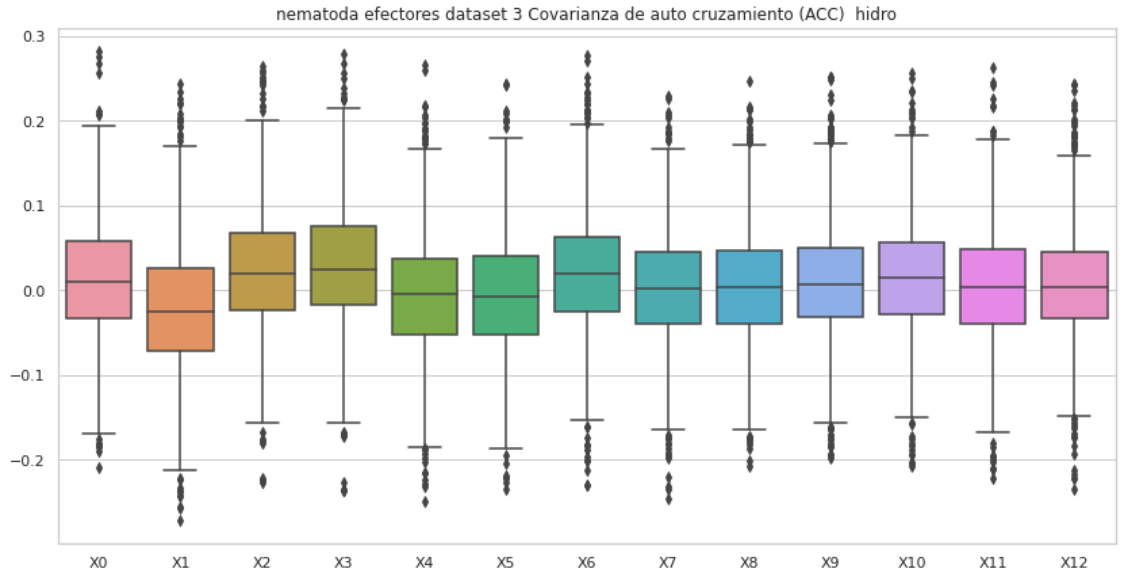
```
996 -0.031146  0.119650 -0.005811 -0.116934 -0.018177  0.051946  no_efectores
997  0.002698 -0.005163  0.061708  0.026008  0.026813  0.034581  no_efectores
998 -0.078641 -0.013639 -0.010675  0.042416 -0.097726  0.066928  no_efectores
999  0.162807  0.037544  0.092748 -0.033703 -0.110642 -0.047257  no_efectores

[901 rows x 14 columns]
```

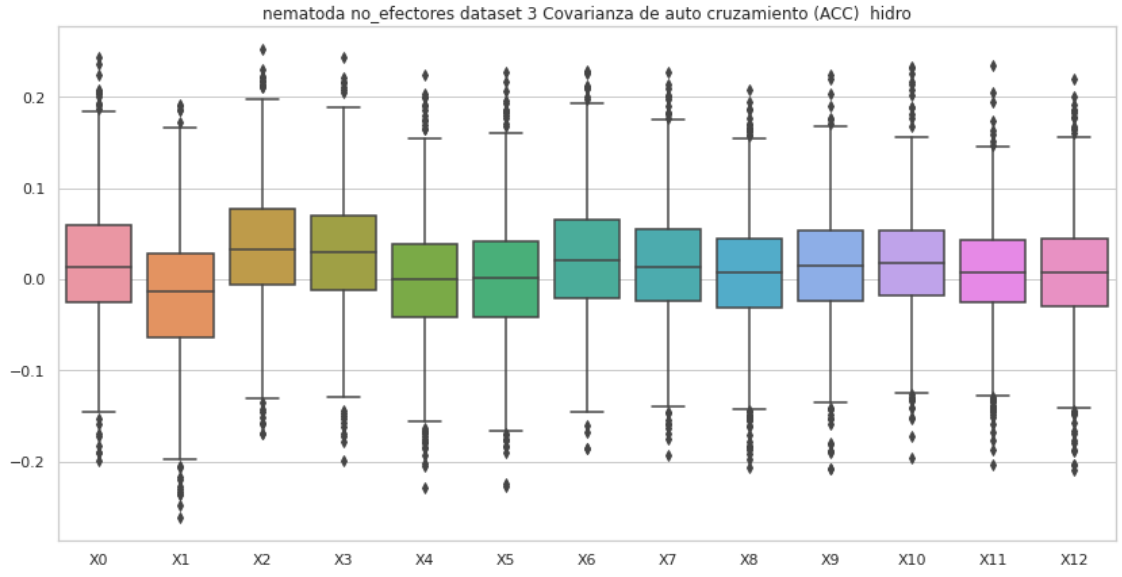
Covarianza de auto cruzamiento (ACC) no_efectores nematoda dataset 3, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	901.000000	901.000000	901.000000	901.000000	901.000000	901.000000	
mean	0.016286	-0.018244	0.034311	0.030118	-0.000842	0.000587	
std	0.068697	0.074627	0.067325	0.068487	0.068053	0.067879	
min	-0.198616	-0.261592	-0.170034	-0.198791	-0.229027	-0.227993	
25%	-0.024847	-0.063433	-0.005756	-0.011596	-0.041712	-0.041828	
50%	0.013194	-0.013763	0.032706	0.029948	-0.000683	0.001134	
75%	0.058942	0.028728	0.077724	0.069715	0.037969	0.041895	
max	0.243910	0.191391	0.252648	0.243716	0.225245	0.227135	

	X6	X7	X8	X9	X10	X11	\
count	901.000000	901.000000	901.000000	901.000000	901.000000	901.000000	
mean	0.022770	0.016253	0.006291	0.013576	0.018005	0.007949	
std	0.065401	0.065032	0.063390	0.061784	0.060931	0.059491	
min	-0.185635	-0.193003	-0.206977	-0.208624	-0.195487	-0.204033	
25%	-0.020073	-0.023937	-0.030758	-0.023525	-0.017894	-0.025925	
50%	0.021193	0.013829	0.007690	0.014325	0.018419	0.007662	
75%	0.066061	0.055655	0.044039	0.053108	0.053795	0.043010	
max	0.228936	0.227142	0.207838	0.224168	0.233469	0.235186	

	X12
count	901.000000
mean	0.005865
std	0.061580
min	-0.208906
25%	-0.030085
50%	0.006766
75%	0.044686
max	0.220125


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
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12