ds4_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 = 4
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 4, con valores
atípicos.
Valores del documento csv.

        X0      X1     X2     X3     X4      X5     X6      X7     X8     X9  \
0    4.950   9.901  3.960  0.990  0.000   7.921  5.941  11.881  3.960  5.941
1    3.750   7.083  2.917  7.083  3.750  12.500  4.583   4.167  2.500  2.917
2    4.583   7.917  3.750  3.333  3.750   5.833  3.750   5.000  2.917  7.500
3    5.696   6.329  5.063  6.962  2.532   5.696  1.266   2.532  4.430  4.430
4    5.069  10.138  4.608  5.069  0.922   3.687  5.991   2.765  3.226  4.608
..     …       …      …      …      …       …      …       …      …      …
995  3.306   3.306  4.132  7.438  3.306   8.264  4.959   3.306  4.132  2.479
996  4.717   5.660  4.717  7.547  5.660   7.547  5.660   3.774  2.830  8.491
997  3.020   8.054  4.698  7.383  2.013   9.732  6.711   3.691  1.342  7.383
998  6.800   6.448  3.869  4.572  3.048   4.689  4.103   5.744  2.931  4.924
999  4.953   6.604  4.717  6.840  3.066   3.302  5.425   4.953  1.887  5.896
```

```
     …     X11    X12    X13     X14    X15     X16    X17    X18    X19  \
0    …   5.941  1.980  4.950   3.960  9.901   4.950  0.990  2.970  3.960
1    …  12.083  2.500  3.333   5.000  5.833   1.667  1.667  5.417  4.583
2    …   4.583  5.833  5.000   4.583  5.833   4.583  1.250  2.500  7.500
3    …   5.696  3.165  5.696   5.063  6.962   3.797  0.633  6.329  6.329
4    …   6.452  1.843  2.304   6.912  7.834  12.442  0.461  0.922  6.912
..   …     …    …    …     …     …     …    …    …    …
995  …   7.438  2.479  3.306  10.744  7.438   7.438  0.000  0.826  9.917
996  …   3.774  8.491  4.717   0.943  5.660   3.774  0.943  2.830  3.774
997  …   8.389  3.356  2.685   3.691  6.711   6.376  0.000  2.349  5.034
998  …   5.158  3.634  4.220   4.455  7.151   5.041  0.469  3.634  9.261
999  …   4.009  1.651  4.953   5.425  8.491   7.075  1.887  3.066  6.368

            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 4, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      6.935465     6.050466     4.346126     5.208663     2.257850
std       2.693499     2.628683     1.982489     2.216592     1.866405
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       5.239500     4.406500     3.137000     3.862000     1.075000
50%       6.743000     5.825000     4.156500     5.162000     1.890000
75%       8.303250     7.347750     5.330250     6.313250     2.905000
max      25.850000    16.949000    16.949000    25.373000    15.824000

               X5           X6           X7           X8           X9  \
count  1000.00000  1000.000000  1000.000000  1000.000000  1000.000000
mean      6.41109     3.855057     5.676309     2.415440     5.558891
std       3.00636     2.167077     3.005864     1.441167     2.304102
```

min	0.00000	0.000000	0.000000	0.000000	0.000000
25%	4.36550	2.560750	3.849000	1.449000	4.054000
50%	6.19650	3.621500	5.305500	2.199500	5.497500
75%	8.11525	4.878000	6.944000	3.090500	7.029750
max	23.07700	27.692000	31.395000	10.417000	15.190000

	X10	X11	X12	X13	X14 \
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	8.938482	5.919579	2.822676	4.287162	4.968626
std	2.885885	3.050026	1.382538	2.099511	3.051091
min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	7.002500	3.987750	1.869000	2.857000	3.239000
50%	8.883000	5.452000	2.632000	4.000000	4.425000
75%	10.744000	7.348250	3.521000	5.460750	5.943750
max	20.482000	26.087000	8.772000	21.311000	25.253000

	X15	X16	X17	X18	X19
count	1000.00000	1000.000000	1000.000000	1000.000000	1000.000000
mean	7.85782	5.646030	1.162292	3.091467	6.590505
std	2.86729	2.176537	1.033532	1.653598	2.378249
min	1.38900	0.000000	0.000000	0.000000	0.000000
25%	6.00000	4.348000	0.423500	2.015250	5.017250
50%	7.50000	5.512000	1.022000	2.913000	6.410000
75%	9.43175	6.780750	1.653000	3.916750	7.921000
max	24.71900	25.843000	9.195000	11.667000	17.073000

no_efectores
Composición de aminoácidos (AAC) no_efectores nematoda dataset 4, con valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9 \
0	8.989	6.742	1.124	15.730	0.000	6.742	2.247	5.618	3.371	2.247
1	1.460	2.190	9.489	7.299	1.460	6.569	0.730	8.759	6.569	13.139
2	3.791	7.820	4.739	4.976	0.948	10.427	3.555	2.607	2.370	5.687
3	3.953	6.324	3.953	3.162	2.372	3.953	3.557	1.581	0.791	7.115
4	5.025	5.025	5.813	5.025	2.266	8.473	3.842	4.138	2.266	8.177
..	…	…	…	…	…	…	…	…	…	
995	6.098	6.098	4.878	7.317	0.000	6.098	1.220	2.439	3.659	7.317
996	4.127	4.444	1.270	4.127	1.905	3.492	7.937	18.413	1.270	4.444
997	5.592	5.428	2.961	5.592	2.961	5.757	3.618	7.401	3.618	6.579
998	2.254	3.099	5.070	9.859	2.535	6.197	1.972	3.099	2.254	7.606
999	4.167	1.389	5.556	5.556	0.000	12.500	6.944	2.778	1.389	9.722

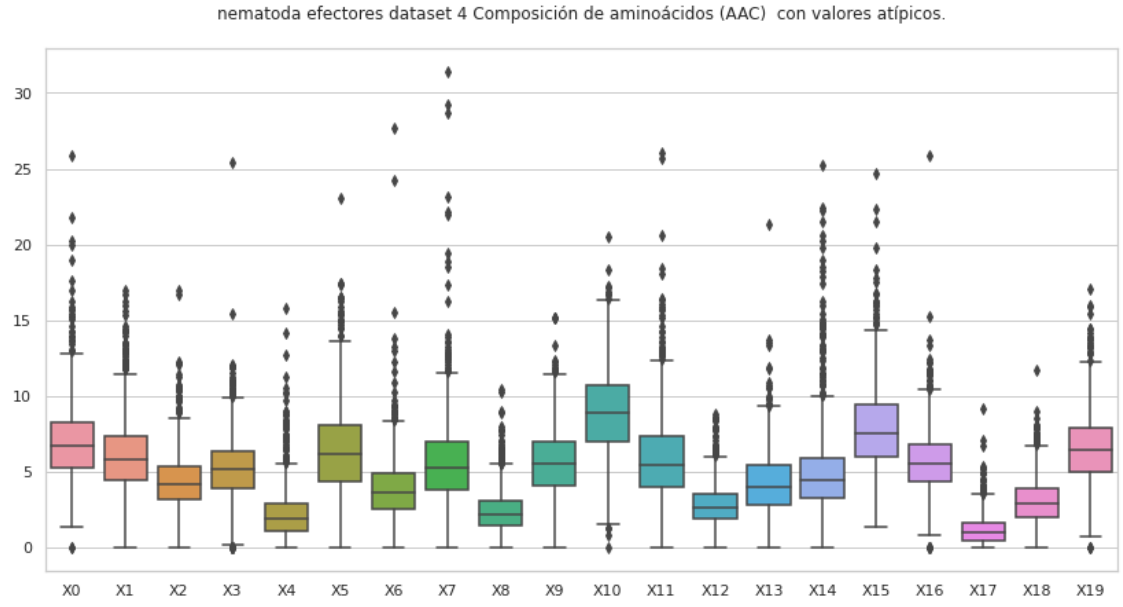
	…	X11	X12	X13	X14	X15	X16	X17	X18	X19 \
0	…	3.371	1.124	1.124	6.742	25.843	1.124	0.000	2.247	2.247

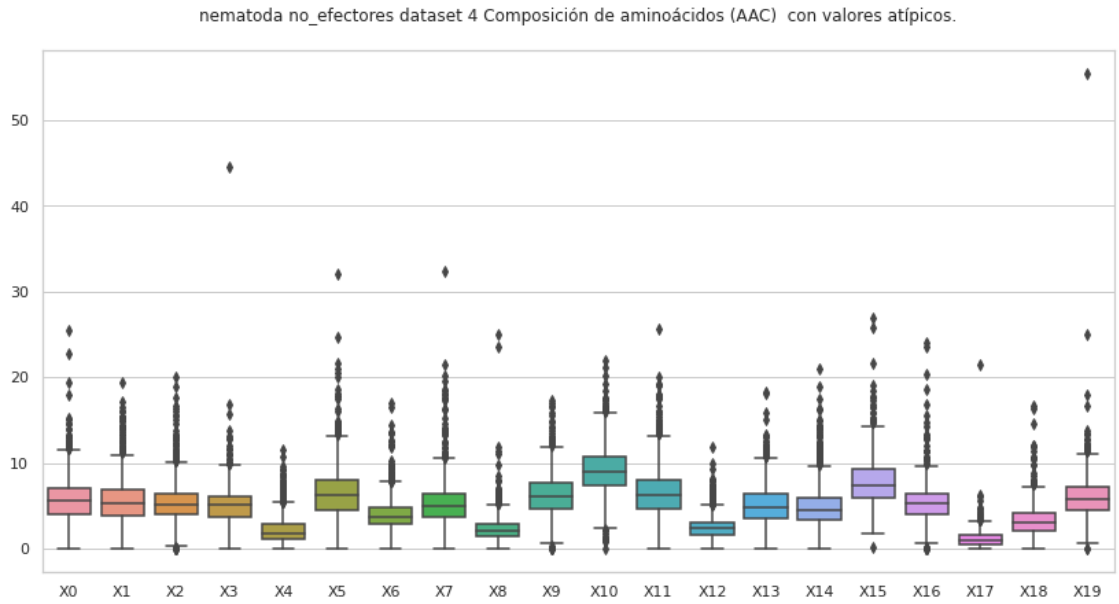
```
1    …  1.460  6.569   4.380   1.460   5.839  5.839  1.460  5.839  2.920
2    …  6.398  3.318   4.976   5.213   6.872  5.450  0.474  2.844  5.687
3    …  3.557  4.348  10.672   1.581   9.091  6.324  0.791  5.534  9.881
4    …  6.995  2.463   6.404   3.350   7.783  3.842  0.788  1.773  4.433
..   …    …    …     …     …     …     …    …    …    …    …
995  …  6.098  3.659   2.439   6.098   9.756  8.537  1.220  3.659  3.659
996  …  1.905  1.270   4.444  20.952   5.079  3.492  0.000  3.175  4.444
997  …  6.250  1.974   4.934   4.605   6.743  3.783  1.151  3.783  6.908
998  …  5.634  3.380   7.042   4.225   5.634  8.451  1.127  5.070  7.042
999  …  6.944  2.778   4.167   2.778   5.556  2.778  2.778  4.167  6.944

              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 4, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      5.769184     5.568575     5.399405     5.000658     2.164999
std       2.707888     2.648280     2.481629     2.396545     1.662164
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       4.064500     3.902250     3.972750     3.659750     1.106750
50%       5.559000     5.282500     5.087500     5.063000     1.818000
75%       7.074750     6.824000     6.438750     6.133250     2.817000
max      25.378000    19.328000    20.000000    44.444000    11.594000

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      6.464175     3.959061     5.403243     2.335024     6.286842
std       3.199501     2.024180     2.891121     1.685084     2.573315
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       4.500000     2.804000     3.657250     1.416750     4.704000
50%       6.283000     3.704000     5.026500     2.157000     6.129500
```

```
75%       7.967000     4.820000     6.443750     2.930250     7.606250
max      32.000000    17.021000    32.277000    25.000000    17.284000

                X10          X11          X12          X13          X14  \
count   1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean       9.089971     6.518292     2.495102     5.110248     4.882022
std        2.868457     3.104524     1.279194     2.407218     2.582190
min        0.000000     0.000000     0.000000     0.000000     0.000000
25%        7.401000     4.564750     1.599250     3.492250     3.315000
50%        8.994500     6.231000     2.339000     4.762000     4.407500
75%       10.804250     8.018250     3.012750     6.338000     5.882000
max       21.875000    25.685000    11.905000    18.182000    20.952000

                X15          X16          X17          X18          X19
count   1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean       7.789351     5.396881     1.124958     3.277297     5.964716
std        2.857718     2.279301     1.113348     1.874500     2.792640
min        0.083000     0.000000     0.000000     0.000000     0.000000
25%        5.878750     4.064250     0.464500     2.098000     4.412000
50%        7.427000     5.284000     1.002000     3.064500     5.814000
75%        9.357750     6.411750     1.575000     4.192750     7.160250
max       26.882000    24.087000    21.429000    16.667000    55.379000
```





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 4, sin valores
atípicos.
Valores del documento csv.

        X0     X1     X2      X3     X4      X5     X6      X7     X8     X9  \
0    4.950  9.901  3.960   0.990  0.000   7.921  5.941  11.881  3.960  5.941
1    3.750  7.083  2.917   7.083  3.750  12.500  4.583   4.167  2.500  2.917
2    4.583  7.917  3.750   3.333  3.750   5.833  3.750   5.000  2.917  7.500
3    5.696  6.329  5.063   6.962  2.532   5.696  1.266   2.532  4.430  4.430
5    7.652  5.013  3.430   6.332  1.583   5.805  2.902   8.971  2.375  5.013
..     …      …      …       …      …       …      …       …      …      …
994  5.946  7.568  3.784  10.270  0.541  12.432  2.162   4.324  0.541  2.162
995  3.306  3.306  4.132   7.438  3.306   8.264  4.959   3.306  4.132  2.479
997  3.020  8.054  4.698   7.383  2.013   9.732  6.711   3.691  1.342  7.383
998  6.800  6.448  3.869   4.572  3.048   4.689  4.103   5.744  2.931  4.924
999  4.953  6.604  4.717   6.840  3.066   3.302  5.425   4.953  1.887  5.896

     …     X11    X12    X13     X14     X15    X16    X17    X18    X19  \
0    …   5.941  1.980  4.950   3.960   9.901  4.950  0.990  2.970  3.960
1    …  12.083  2.500  3.333   5.000   5.833  1.667  1.667  5.417  4.583
2    …   4.583  5.833  5.000   4.583   5.833  4.583  1.250  2.500  7.500
3    …   5.696  3.165  5.696   5.063   6.962  3.797  0.633  6.329  6.329
5    …   7.124  2.902  5.541   3.958   6.069  7.124  0.264  3.166  8.443
..   …     …      …      …       …       …      …      …      …      …
994  …   6.486  1.081  3.243   4.324  16.216  5.405  0.541  0.000  4.865
```

```
995  …   7.438  2.479  3.306  10.744   7.438  7.438  0.000  0.826  9.917
997  …   8.389  3.356  2.685   3.691   6.711  6.376  0.000  2.349  5.034
998  …   5.158  3.634  4.220   4.455   7.151  5.041  0.469  3.634  9.261
999  …   4.009  1.651  4.953   5.425   8.491  7.075  1.887  3.066  6.368

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

[828 rows x 21 columns]
```

Composición de aminoácidos (AAC) efectores nematoda dataset 4, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count   828.000000  828.000000  828.000000  828.000000  828.000000  828.000000
mean      6.849407    6.113682    4.341314    5.277011    2.146353    6.506301
std       2.185701    2.320507    1.665788    1.898150    1.443847    2.618050
min       1.389000    0.000000    0.000000    0.000000    0.000000    0.000000
25%       5.331000    4.612500    3.279000    4.061750    1.162250    4.697250
50%       6.778500    5.948500    4.211000    5.263000    1.914000    6.299000
75%       8.191750    7.356750    5.297000    6.313250    2.834000    8.085000
max      14.286000   13.542000   10.000000   11.538000    7.801000   15.294000

                X6          X7          X8          X9         X10         X11  \
count   828.000000  828.000000  828.000000  828.000000  828.000000  828.00000
mean      3.795597    5.529287    2.407272    5.722039    9.164636    5.87701
std       1.670327    2.263853    1.200317    2.060882    2.612213    2.55766
min       0.000000    0.000000    0.000000    0.000000    1.235000    0.00000
25%       2.646000    3.998000    1.560500    4.348000    7.365250    4.07950
50%       3.645500    5.332500    2.258500    5.661000    9.091000    5.49400
75%       4.682750    6.824000    3.077000    7.078500   10.758000    7.32375
max      10.256000   13.924000    6.604000   12.371000   17.241000   14.28600

               X12         X13         X14         X15         X16         X17  \
count   828.000000  828.000000  828.000000  828.000000  828.000000  828.000000
mean      2.761006    4.348926    4.605068    7.874240    5.662042    1.152662
```

std	1.193515	1.833414	2.120689	2.554762	1.850077	0.854991
min	0.000000	0.000000	0.000000	1.429000	0.000000	0.000000
25%	1.938250	3.039750	3.277000	6.170750	4.478500	0.516750
50%	2.639000	4.117000	4.331500	7.591000	5.578500	1.048000
75%	3.448000	5.535750	5.635750	9.431750	6.749000	1.648000
max	6.767000	10.526000	13.944000	16.273000	11.712000	4.211000

	X18	X19
count	828.000000	828.000000
mean	3.192725	6.673436
std	1.532586	2.141940
min	0.000000	0.000000
25%	2.183500	5.283250
50%	3.030000	6.498500
75%	3.969500	7.937250
max	8.040000	13.699000

no_efectores
Composición de aminoácidos (AAC) no_efectores nematoda dataset 4, sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9	\
2	3.791	7.820	4.739	4.976	0.948	10.427	3.555	2.607	2.370	5.687	
3	3.953	6.324	3.953	3.162	2.372	3.953	3.557	1.581	0.791	7.115	
4	5.025	5.025	5.813	5.025	2.266	8.473	3.842	4.138	2.266	8.177	
5	7.392	6.623	5.795	5.145	2.129	6.150	5.677	4.140	3.312	5.086	
8	4.905	4.632	7.084	2.452	2.997	1.907	4.087	3.542	1.362	11.444	
..	…	…	…	…	…	…	…	…	…		
994	5.000	9.167	3.333	8.750	1.250	7.500	3.750	4.583	3.750	2.083	
995	6.098	6.098	4.878	7.317	0.000	6.098	1.220	2.439	3.659	7.317	
997	5.592	5.428	2.961	5.592	2.961	5.757	3.618	7.401	3.618	6.579	
998	2.254	3.099	5.070	9.859	2.535	6.197	1.972	3.099	2.254	7.606	
999	4.167	1.389	5.556	5.556	0.000	12.500	6.944	2.778	1.389	9.722	

	…	X11	X12	X13	X14	X15	X16	X17	X18	X19	\
2	…	6.398	3.318	4.976	5.213	6.872	5.450	0.474	2.844	5.687	
3	…	3.557	4.348	10.672	1.581	9.091	6.324	0.791	5.534	9.881	
4	…	6.995	2.463	6.404	3.350	7.783	3.842	0.788	1.773	4.433	
5	…	4.258	2.839	3.134	6.742	7.865	5.322	1.419	3.489	5.855	
8	…	4.632	2.180	8.174	2.997	7.084	5.995	2.452	3.815	5.177	
..	…	…	…	…	…	…	…	…	…		
994	…	2.917	0.833	1.667	7.083	15.000	3.750	0.000	0.833	9.583	
995	…	6.098	3.659	2.439	6.098	9.756	8.537	1.220	3.659	3.659	
997	…	6.250	1.974	4.934	4.605	6.743	3.783	1.151	3.783	6.908	
998	…	5.634	3.380	7.042	4.225	5.634	8.451	1.127	5.070	7.042	

```
999  …  6.944  2.778   4.167  2.778   5.556  2.778  2.778  4.167  6.944

              X20
2    no_efectores
3    no_efectores
4    no_efectores
5    no_efectores
8    no_efectores
..            …
994  no_efectores
995  no_efectores
997  no_efectores
998  no_efectores
999  no_efectores

[820 rows x 21 columns]
```

Composición de aminoácidos (AAC) no_efectores nematoda dataset 4, sin valores atípicos.
Estadísticas.

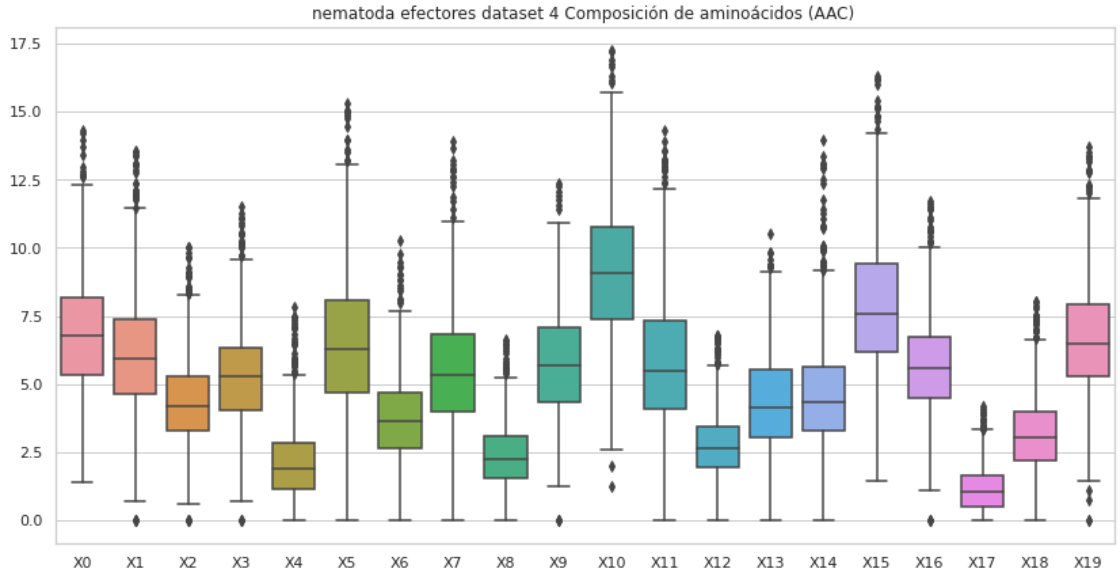
	X0	X1	X2	X3	X4	X5	\
count	820.000000	820.000000	820.000000	820.000000	820.000000	820.000000	
mean	5.853722	5.624951	5.258298	5.097593	2.129406	6.543130	
std	2.250470	2.175520	1.989703	1.762598	1.398313	2.597261	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	4.308500	4.246750	4.054000	3.971750	1.170750	4.916500	
50%	5.671500	5.471000	5.066500	5.193500	1.846500	6.505500	
75%	7.050750	6.831000	6.284750	6.237250	2.794000	7.984250	
max	13.333000	13.333000	12.717000	11.650000	7.143000	14.806000	

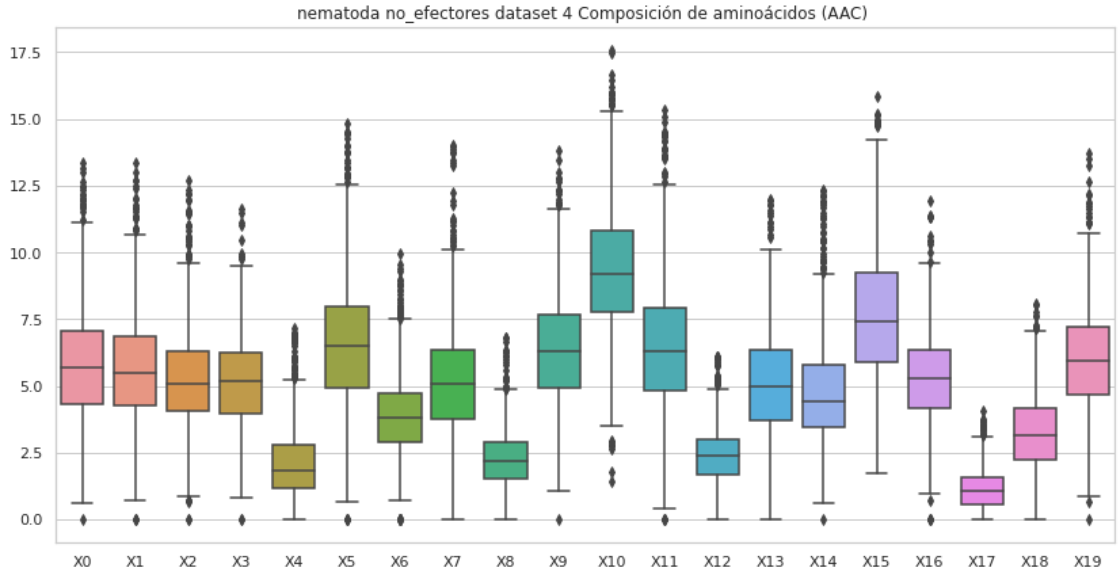
	X6	X7	X8	X9	X10	X11	\
count	820.000000	820.000000	820.000000	820.000000	820.000000	820.000000	
mean	3.932596	5.252243	2.289234	6.409205	9.295437	6.483835	
std	1.635330	2.218764	1.126966	2.240189	2.461624	2.541404	
min	0.000000	0.000000	0.000000	0.000000	1.408000	0.000000	
25%	2.895000	3.774000	1.539500	4.915500	7.746000	4.802500	
50%	3.785000	5.058500	2.209000	6.306000	9.182000	6.288000	
75%	4.740000	6.339500	2.874000	7.653750	10.818000	7.937750	
max	9.942000	14.000000	6.818000	13.793000	17.568000	15.313000	

	X12	X13	X14	X15	X16	X17	\
count	820.000000	820.000000	820.000000	820.000000	820.000000	820.000000	
mean	2.455400	5.166263	4.776984	7.653154	5.344323	1.126422	
std	1.062375	2.092408	2.049805	2.461135	1.746279	0.797236	
min	0.000000	0.000000	0.000000	1.714000	0.000000	0.000000	
25%	1.670750	3.719750	3.473250	5.882000	4.178750	0.559750	

50%	2.379000	4.988500	4.428500	7.410000	5.281500	1.056000
75%	2.994750	6.364500	5.769750	9.260750	6.347500	1.590750
max	6.114000	12.000000	12.360000	15.862000	11.950000	4.054000

	X18	X19
count	820.000000	820.000000
mean	3.257866	6.049949
std	1.485022	2.046256
min	0.000000	0.000000
25%	2.237250	4.651000
50%	3.144000	5.955000
75%	4.167000	7.193500
max	8.065000	13.699000





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 4, con valores atípicos.
Valores del documento csv.

          X0        X1        X2        X3        X4        X5        X6  \
0    0.018852  0.000000  0.003770  0.030164  0.018852  0.045245  0.015082
1    0.012446  0.012446  0.023508  0.041485  0.011063  0.013828  0.008297
2    0.048190  0.039428  0.035047  0.061333  0.052571  0.052571  0.030666
3    0.095548  0.042466  0.116781  0.095548  0.095548  0.042466  0.074315
4    0.019380  0.003524  0.019380  0.014094  0.008809  0.010571  0.012332
..        …         …         …         …         …         …         …
995  0.013638  0.013638  0.030686  0.034096  0.013638  0.013638  0.017048
996  0.090534  0.108641  0.144855  0.144855  0.090534  0.072427  0.054321
997  0.009199  0.006133  0.022487  0.029641  0.008177  0.011243  0.004088
998  0.048135  0.021578  0.032366  0.033196  0.029877  0.040665  0.020748
999  0.022843  0.014141  0.031545  0.015229  0.022843  0.022843  0.008702

          X7        X8        X9  …       X74       X75       X76  \
0    0.022623  0.022623  0.018852  …  0.003215  0.009183 -0.013096
1    0.009680  0.040102  0.022125  … -0.016390  0.006968 -0.008464
2    0.078856  0.048190  0.105141  … -0.034653 -0.025890  0.005489
3    0.074315  0.095548  0.191097  …  0.071672  0.007057 -0.030409
4    0.017618  0.024665  0.029950  …  0.023275  0.014709  0.012552
..        …         …         …  …         …         …         …
995  0.010229  0.030686  0.023867  …  0.035085  0.036673  0.018915
996  0.162962  0.072427  0.162962  … -0.274759 -0.204198  0.072263
997  0.022487  0.025553  0.022487  …  0.001847  0.008843 -0.000388
998  0.034856  0.036516  0.069712  …  0.012688  0.008423  0.022342
999  0.027194  0.018492  0.043510  …  0.010938  0.020885  0.005044

          X77       X78       X79       X80       X81       X82        X83
0   -0.009180  0.006397  0.001124 -0.003546 -0.000833  0.019458  efectores
1    0.020816  0.036688  0.000013  0.016295  0.029330 -0.021494  efectores
2    0.044522 -0.012468 -0.031458  0.053428  0.033045  0.009121  efectores
3    0.095003  0.066062  0.038301  0.024208 -0.023460  0.078180  efectores
4    0.004566  0.002701 -0.006572  0.032666  0.025636  0.008896  efectores
..        …         …         …         …         …         …         …
995  0.030811  0.040713  0.016654  0.007881  0.013588  0.026099  efectores
996 -0.171894 -0.002146  0.003256  0.062657  0.034120  0.080387  efectores
997  0.008376  0.020320  0.003284 -0.003215  0.026897  0.000754  efectores
```

```
998 -0.001564 -0.016723  0.033757  0.029997  0.006639  0.011114  efectores
999 -0.003319  0.001563  0.012370  0.009302  0.023247  0.010487  efectores

[1000 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro_mass efectores nematoda
dataset 4, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.023892	0.008100	0.023723	0.031284	0.022350	
std	0.435980	0.145818	0.202307	0.154086	0.101385	
min	-13.661442	-4.553814	-6.071752	-4.553814	-3.035876	
25%	0.022580	0.004356	0.015124	0.018679	0.011121	
50%	0.032008	0.009327	0.025309	0.030151	0.019270	
75%	0.044730	0.016240	0.037124	0.043888	0.032063	
max	1.320454	0.256461	1.650567	1.320454	0.384692	

	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.028025	0.011714	0.023548	0.028203	0.033385	…	
std	0.071507	0.100576	0.243613	0.100430	0.586475	…	
min	-1.517938	-3.035876	-7.589690	-3.035876	-18.215256	…	
25%	0.016575	0.005699	0.015346	0.016339	0.027336	…	
50%	0.025032	0.010552	0.026404	0.026875	0.042786	…	
75%	0.036707	0.018233	0.041215	0.040648	0.064401	…	
max	0.990340	0.660227	0.258722	0.206978	2.310794	…	

	X73	X74	X75	X76	X77	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.009178	-0.002279	-0.004896	-0.002466	0.005337	
std	0.069277	0.123106	0.347934	0.291676	0.143278	
min	-1.743205	-3.398943	-10.037314	-8.117730	-2.565091	
25%	-0.000243	-0.011898	-0.004930	-0.000060	-0.008500	
50%	0.010740	0.003026	0.007083	0.010131	0.004819	
75%	0.023254	0.015378	0.018102	0.022339	0.016237	
max	0.473984	1.311434	2.766761	0.650430	2.731854	

	X78	X79	X80	X81	X82
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.009252	0.009190	0.013932	0.017092	-0.002746
std	0.094696	0.226715	0.434488	0.439973	0.363193
min	-0.742069	-6.130323	-2.206495	-2.216999	-11.411002
25%	-0.003701	0.000269	-0.009194	-0.003874	-0.000911
50%	0.007779	0.011677	0.003893	0.008089	0.009765
75%	0.020256	0.022489	0.017728	0.020990	0.022683

```
max       2.016337     3.196991    13.498553    13.641814     0.246002

[8 rows x 83 columns]
```

no_efectores
Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores nematoda
dataset 4, con valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.012784  0.000000  0.022372  0.009588  0.001598  0.007990  0.004794
1    0.012079  0.012079  0.060394  0.054354  0.036236  0.072472  0.054354
2    0.019306  0.004827  0.025339  0.053092  0.025339  0.013273  0.012066
3    0.013986  0.008392  0.011189  0.013986  0.037763  0.005594  0.002797
4    0.032614  0.014708  0.032614  0.054996  0.041567  0.026859  0.014708
..        …         …         …         …         …         …
995  0.050851  0.000000  0.061022  0.050851  0.020341  0.020341  0.030511
996  0.006384  0.002946  0.006384  0.005402  0.006875  0.028482  0.001964
997  0.036994  0.019585  0.036994  0.038082  0.032641  0.048962  0.023937
998  0.022994  0.025868  0.100598  0.063233  0.071855  0.031616  0.022994
999  0.046861  0.000000  0.062481  0.140582  0.046861  0.031240  0.015620

           X7        X8        X9  …       X74       X75       X76  \
0    0.003196  0.004794  0.004794  …  0.004043  0.031730  0.025487
1    0.108709  0.012079  0.054354  … -0.012906 -0.022176  0.077653
2    0.028959  0.032579  0.060331  … -0.001590 -0.005772 -0.001139
3    0.025175  0.012588  0.040560  …  0.022027  0.011449  0.003057
4    0.053078  0.045404  0.078657  …  0.011815  0.023175  0.002102
..        …         …         …  …         …         …
995  0.061022  0.050851  0.081362  …  0.059203  0.044589 -0.009659
996  0.006875  0.002946  0.005893  …  0.004809  0.002699  0.024026
997  0.043522  0.041346  0.068547  …  0.006683  0.009414  0.035469
998  0.077604  0.057484  0.086226  …  0.007248 -0.001851  0.012467
999  0.109341  0.078101  0.124962  … -0.079663 -0.003604 -0.025589

           X77       X78       X79       X80       X81       X82           X83
0   -0.005968  0.007762  0.018848  0.008035  0.024457  0.018588  no_efectores
1   -0.038668 -0.020675  0.009482 -0.006433 -0.055350 -0.041453  no_efectores
2   -0.003268  0.031285  0.006815 -0.002185  0.016269  0.001487  no_efectores
3    0.003000 -0.012508 -0.003551  0.015268  0.010233 -0.005471  no_efectores
4    0.017207  0.018374  0.016568  0.004448  0.007392  0.016943  no_efectores
..        …         …         …         …         …         …
995 -0.047767  0.054968  0.031909  0.037867  0.046766  0.001472  no_efectores
996  0.002008  0.002229  0.034085  0.001699 -0.001231  0.041304  no_efectores
997  0.002397 -0.004034  0.027561 -0.004066 -0.002040 -0.004968  no_efectores
998  0.051915  0.074318 -0.012246 -0.037723 -0.037396 -0.008766  no_efectores
```

```
999 -0.025006 -0.033081  0.061361 -0.081078 -0.036926  0.053611  no_efectores

[1000 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores nematoda
dataset 4, con valores atípicos.
Estadísticas.

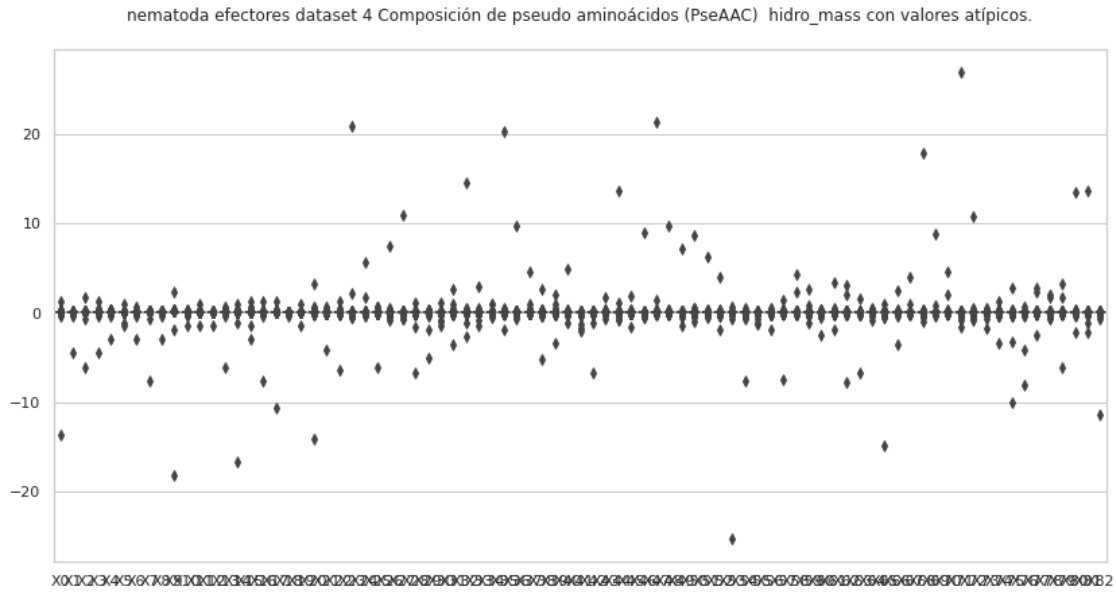
	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.030306	0.012579	0.028798	0.036035	0.031188	
std	0.022577	0.016808	0.024970	0.031591	0.033203	
min	-0.077314	-0.009198	-0.051543	-0.051543	-0.128857	
25%	0.018129	0.004703	0.014085	0.018796	0.014097	
50%	0.026973	0.009178	0.024231	0.031404	0.024215	
75%	0.038353	0.015899	0.036754	0.046078	0.038148	
max	0.362714	0.241809	0.342014	0.604524	0.428090	

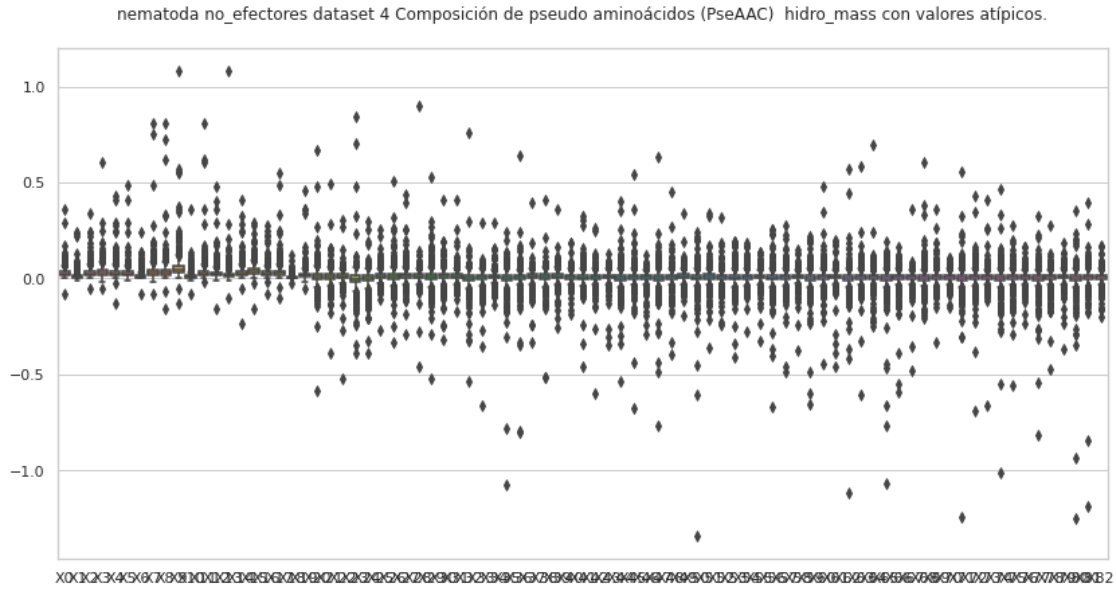
	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.029024	0.014005	0.038212	0.038353	0.054702	…	
std	0.027687	0.016208	0.045970	0.047831	0.057708	…	
min	-0.009198	-0.077314	-0.077314	-0.154628	-0.128857	…	
25%	0.016133	0.005712	0.018144	0.018176	0.028378	…	
50%	0.025024	0.010608	0.029952	0.030495	0.044641	…	
75%	0.035724	0.017187	0.045483	0.046877	0.064652	…	
max	0.483619	0.241809	0.809987	0.809987	1.079982	…	

	X73	X74	X75	X76	X77	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.008458	0.000608	0.006250	0.007886	-0.000223	
std	0.034643	0.054248	0.037343	0.025358	0.048430	
min	-0.662683	-1.008080	-0.556242	-0.233398	-0.813809	
25%	-0.000624	-0.009291	-0.002285	-0.001375	-0.008249	
50%	0.008139	0.003820	0.007624	0.007945	0.002990	
75%	0.018742	0.015474	0.019287	0.018723	0.014458	
max	0.411050	0.468321	0.273578	0.164560	0.325015	

	X78	X79	X80	X81	X82
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.005249	0.009485	-0.000306	0.005274	0.008587
std	0.034043	0.027459	0.062842	0.056630	0.025423
min	-0.468817	-0.367631	-1.250313	-1.183774	-0.200593
25%	-0.003177	0.000457	-0.009842	-0.004121	-0.001083
50%	0.007385	0.009876	0.003342	0.007343	0.008737
75%	0.017673	0.020051	0.015688	0.020181	0.019146
max	0.278378	0.182474	0.353869	0.392139	0.174543

[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 4, sin valores atípicos.
Valores del documento csv.
```

	X0	X1	X2	X3	X4	X5	X6	\
0	0.018852	0.000000	0.003770	0.030164	0.018852	0.045245	0.015082	
1	0.012446	0.012446	0.023508	0.041485	0.011063	0.013828	0.008297	
2	0.048190	0.039428	0.035047	0.061333	0.052571	0.052571	0.030666	
3	0.095548	0.042466	0.116781	0.095548	0.095548	0.042466	0.074315	
4	0.019380	0.003524	0.019380	0.014094	0.008809	0.010571	0.012332	
..	…	…	…	…	…	…	…	
995	0.013638	0.013638	0.030686	0.034096	0.013638	0.013638	0.017048	
996	0.090534	0.108641	0.144855	0.144855	0.090534	0.072427	0.054321	
997	0.009199	0.006133	0.022487	0.029641	0.008177	0.011243	0.004088	
998	0.048135	0.021578	0.032366	0.033196	0.029877	0.040665	0.020748	
999	0.022843	0.014141	0.031545	0.015229	0.022843	0.022843	0.008702	

	X7	X8	X9	…	X74	X75	X76	\
0	0.022623	0.022623	0.018852	…	0.003215	0.009183	-0.013096	
1	0.009680	0.040102	0.022125	…	-0.016390	0.006968	-0.008464	
2	0.078856	0.048190	0.105141	…	-0.034653	-0.025890	0.005489	
3	0.074315	0.095548	0.191097	…	0.071672	0.007057	-0.030409	
4	0.017618	0.024665	0.029950	…	0.023275	0.014709	0.012552	
..	…	…	…	…	…	…	…	
995	0.010229	0.030686	0.023867	…	0.035085	0.036673	0.018915	
996	0.162962	0.072427	0.162962	…	-0.274759	-0.204198	0.072263	
997	0.022487	0.025553	0.022487	…	0.001847	0.008843	-0.000388	
998	0.034856	0.036516	0.069712	…	0.012688	0.008423	0.022342	
999	0.027194	0.018492	0.043510	…	0.010938	0.020885	0.005044	

	X77	X78	X79	X80	X81	X82	X83
0	-0.009180	0.006397	0.001124	-0.003546	-0.000833	0.019458	efectores
1	0.020816	0.036688	0.000013	0.016295	0.029330	-0.021494	efectores
2	0.044522	-0.012468	-0.031458	0.053428	0.033045	0.009121	efectores
3	0.095003	0.066062	0.038301	0.024208	-0.023460	0.078180	efectores
4	0.004566	0.002701	-0.006572	0.032666	0.025636	0.008896	efectores
..	…	…	…	…	…	…	
995	0.030811	0.040713	0.016654	0.007881	0.013588	0.026099	efectores
996	-0.171894	-0.002146	0.003256	0.062657	0.034120	0.080387	efectores
997	0.008376	0.020320	0.003284	-0.003215	0.026897	0.000754	efectores
998	-0.001564	-0.016723	0.033757	0.029997	0.006639	0.011114	efectores
999	-0.003319	0.001563	0.012370	0.009302	0.023247	0.010487	efectores

[975 rows x 84 columns]

Composición de pseudo aminoácidos (PseAAC) hidro_mass efectores nematoda
dataset 4, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	975.000000	975.000000	975.000000	975.000000	975.000000	975.000000	
mean	0.034895	0.012284	0.027745	0.033090	0.023544	0.028236	
std	0.018589	0.012434	0.017905	0.021100	0.017860	0.016920	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.022229	0.004378	0.015027	0.018516	0.011092	0.016481	
50%	0.031832	0.009228	0.024809	0.029908	0.018977	0.024845	
75%	0.044108	0.015768	0.036616	0.043338	0.031275	0.035838	
max	0.150047	0.130856	0.144855	0.161287	0.118515	0.114081	

	X6	X7	X8	X9	…	X73	\
count	975.000000	975.000000	975.000000	975.000000	…	975.000000	
mean	0.013518	0.030242	0.030275	0.048470	…	0.011192	
std	0.012004	0.020958	0.019610	0.030555	…	0.024187	
min	0.000000	0.000000	0.000000	0.001148	…	-0.142476	
25%	0.005694	0.015249	0.016264	0.027211	…	0.000157	
50%	0.010425	0.026093	0.026295	0.042309	…	0.010789	
75%	0.017821	0.040063	0.039837	0.062493	…	0.023116	
max	0.087626	0.164533	0.138249	0.225325	…	0.148730	

	X74	X75	X76	X77	X78	X79	\
count	975.000000	975.000000	975.000000	975.000000	975.000000	975.000000	
mean	0.001572	0.006615	0.011591	0.003224	0.008034	0.010758	
std	0.031461	0.024635	0.021644	0.035143	0.028339	0.023270	
min	-0.274759	-0.204198	-0.116259	-0.262794	-0.167480	-0.130581	
25%	-0.010398	-0.004500	0.000750	-0.007948	-0.002751	0.000422	
50%	0.003214	0.007147	0.010513	0.004914	0.007925	0.011623	
75%	0.015385	0.017931	0.022599	0.016200	0.020199	0.021770	
max	0.156137	0.129652	0.117181	0.226590	0.185331	0.179683	

	X80	X81	X82
count	975.000000	975.000000	975.000000
mean	0.003184	0.008671	0.009575
std	0.032017	0.026742	0.022393
min	-0.177300	-0.155940	-0.117308
25%	-0.008431	-0.003183	-0.000510
50%	0.004019	0.008318	0.009788
75%	0.017629	0.020995	0.022393
max	0.161270	0.141363	0.095720

[8 rows x 83 columns]

```
no_efectores
Composición de pseudo aminoácidos (PseAAC)  hidro_mass no_efectores nematoda
dataset 4, sin valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.012784  0.000000  0.022372  0.009588  0.001598  0.007990  0.004794
1    0.012079  0.012079  0.060394  0.054354  0.036236  0.072472  0.054354
2    0.019306  0.004827  0.025339  0.053092  0.025339  0.013273  0.012066
3    0.013986  0.008392  0.011189  0.013986  0.037763  0.005594  0.002797
4    0.032614  0.014708  0.032614  0.054996  0.041567  0.026859  0.014708
..        …         …         …         …         …         …
993  0.033896  0.009684  0.026148  0.049391  0.029053  0.017432  0.015495
994  0.012933  0.003233  0.022633  0.019399  0.004311  0.011855  0.009700
996  0.006384  0.002946  0.006384  0.005402  0.006875  0.028482  0.001964
997  0.036994  0.019585  0.036994  0.038082  0.032641  0.048962  0.023937
998  0.022994  0.025868  0.100598  0.063233  0.071855  0.031616  0.022994

           X7        X8        X9  …       X74       X75       X76  \
0    0.003196  0.004794  0.004794  …  0.004043  0.031730  0.025487
1    0.108709  0.012079  0.054354  … -0.012906 -0.022176  0.077653
2    0.028959  0.032579  0.060331  … -0.001590 -0.005772 -0.001139
3    0.025175  0.012588  0.040560  …  0.022027  0.011449  0.003057
4    0.053078  0.045404  0.078657  …  0.011815  0.023175  0.002102
..        …         …         …  …       …         …         …
993  0.050359  0.052296  0.069728  …  0.008931  0.027202 -0.003461
994  0.005389  0.007544  0.023710  … -0.006942  0.010234  0.011349
996  0.006875  0.002946  0.005893  …  0.004809  0.002699  0.024026
997  0.043522  0.041346  0.068547  …  0.006683  0.009414  0.035469
998  0.077604  0.057484  0.086226  …  0.007248 -0.001851  0.012467

           X77       X78       X79       X80       X81       X82            X83
0   -0.005968  0.007762  0.018848  0.008035  0.024457  0.018588  no_efectores
1   -0.038668 -0.020675  0.009482 -0.006433 -0.055350 -0.041453  no_efectores
2   -0.003268  0.031285  0.006815 -0.002185  0.016269  0.001487  no_efectores
3    0.003000 -0.012508 -0.003551  0.015268  0.010233 -0.005471  no_efectores
4    0.017207  0.018374  0.016568  0.004448  0.007392  0.016943  no_efectores
..        …         …         …         …         …         …
993 -0.029269  0.006711  0.016369 -0.002866  0.000601  0.015150  no_efectores
994  0.001334  0.015076 -0.000211  0.020234  0.011651  0.012497  no_efectores
996  0.002008  0.002229  0.034085  0.001699 -0.001231  0.041304  no_efectores
997  0.002397 -0.004034  0.027561 -0.004066 -0.002040 -0.004968  no_efectores
998  0.051915  0.074318 -0.012246 -0.037723 -0.037396 -0.008766  no_efectores

[889 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores nematoda
dataset 4, sin valores atípicos.
Estadísticas.

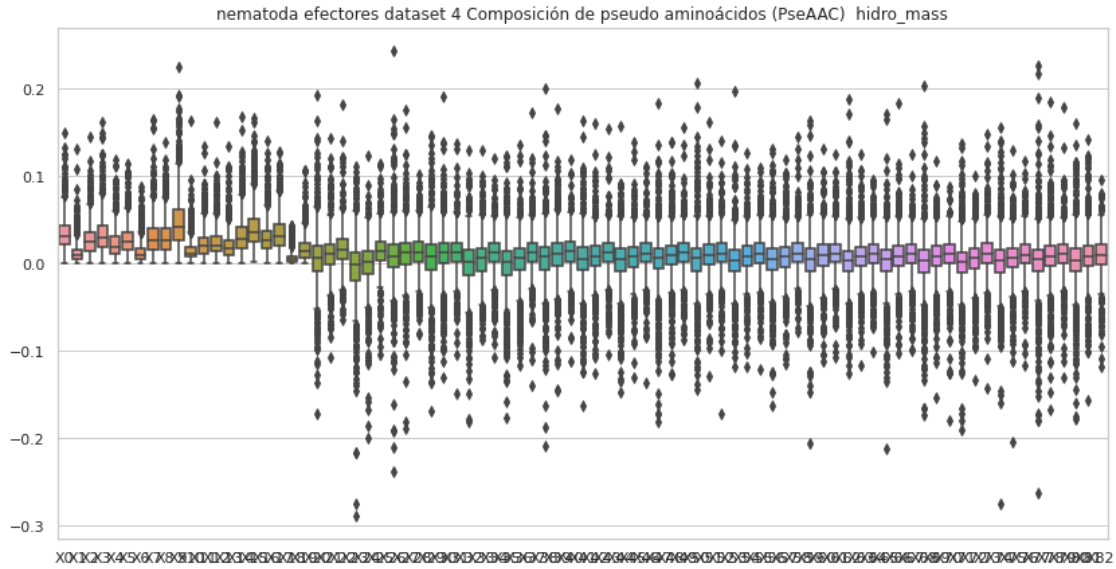
	X0	X1	X2	X3	X4	X5	\
count	889.000000	889.000000	889.000000	889.000000	889.000000	889.000000	
mean	0.027511	0.010732	0.025022	0.031812	0.025400	0.025457	
std	0.013890	0.008579	0.015145	0.018579	0.017020	0.013269	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.017561	0.004749	0.013252	0.017892	0.013327	0.015548	
50%	0.026033	0.008880	0.022463	0.029385	0.022784	0.023696	
75%	0.035613	0.014910	0.034488	0.042162	0.033341	0.033838	
max	0.087795	0.060953	0.100598	0.114229	0.111157	0.085151	

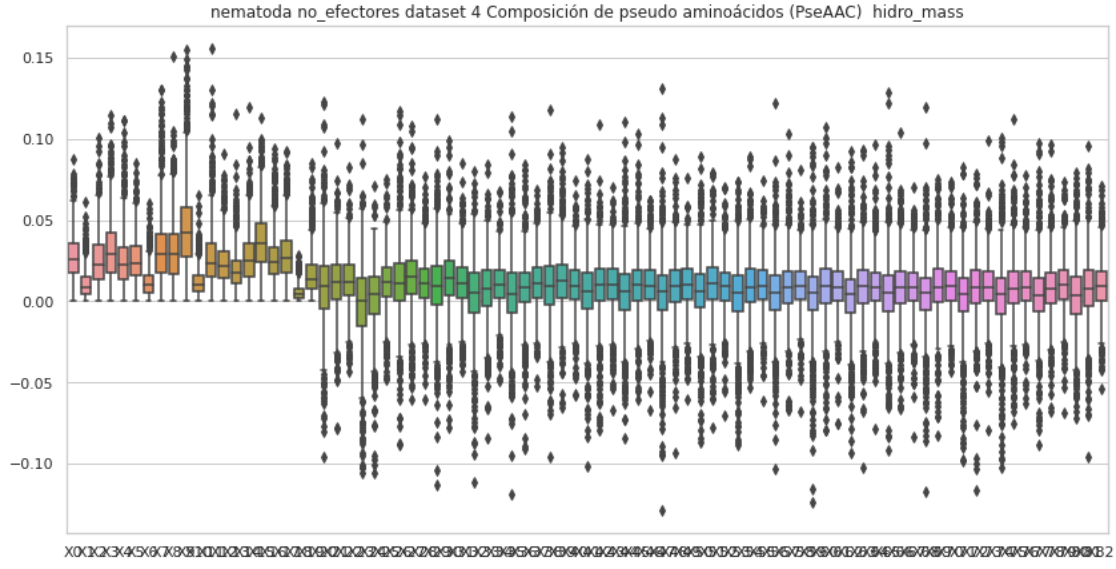
	X6	X7	X8	X9	…	X73	\
count	889.000000	889.000000	889.000000	889.000000	…	889.000000	
mean	0.011608	0.032072	0.031607	0.045634	…	0.008999	
std	0.008667	0.020905	0.019423	0.026086	…	0.016271	
min	0.000000	0.000000	0.000000	0.000000	…	-0.079348	
25%	0.005386	0.017178	0.017039	0.027125	…	0.000522	
50%	0.009787	0.028849	0.028997	0.041831	…	0.008324	
75%	0.015508	0.041400	0.041717	0.058218	…	0.018335	
max	0.060357	0.130529	0.150635	0.154463	…	0.099156	

	X74	X75	X76	X77	X78	X79	\
count	889.000000	889.000000	889.000000	889.000000	889.000000	889.000000	
mean	0.002600	0.008059	0.009211	0.003782	0.008054	0.010347	
std	0.022077	0.019349	0.016519	0.020704	0.017651	0.015870	
min	-0.102886	-0.075673	-0.063216	-0.088630	-0.068111	-0.068546	
25%	-0.007767	-0.001138	0.000019	-0.006377	-0.001536	0.001185	
50%	0.004020	0.007636	0.008184	0.003725	0.007548	0.010072	
75%	0.014414	0.018436	0.018223	0.014186	0.017354	0.019457	
max	0.100587	0.112507	0.077653	0.097409	0.096083	0.073868	

	X80	X81	X82
count	889.000000	889.000000	889.000000
mean	0.003706	0.007783	0.009740
std	0.021043	0.020342	0.015790
min	-0.071933	-0.095528	-0.053673
25%	-0.007628	-0.002633	-0.000044
50%	0.003869	0.007453	0.009048
75%	0.015451	0.019016	0.018611
max	0.079334	0.095401	0.070771

[8 rows x 83 columns]





4 Composición de pseudo aminoácidos (PseAAC) mass

[7]:
```
#mass
transf = "Composición de pseudo aminoácidos (PseAAC) "
transf2 = "PseAAC"
```

```
estado = "con valores atípicos.\n"
comp = "mass"
df=""

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

    if etiq == "efectores":
        df=PseAAC_mass_efec

    if etiq == "no_efectores":
        df=PseAAC_mass_no_efec

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

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

```
efectores
Composición de pseudo aminoácidos (PseAAC)  mass efectores nematoda dataset 4,
con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.031156  0.000000  0.006231  0.049849  0.031156  0.074774  0.024925
1    0.050142  0.050142  0.094713  0.167141  0.044571  0.055714  0.033428
2    0.041527  0.033976  0.030201  0.052852  0.045302  0.045302  0.026426
3    0.062386  0.027727  0.076249  0.062386  0.062386  0.027727  0.048522
4    0.031586  0.005743  0.031586  0.022972  0.014357  0.017229  0.020100
..        …         …         …         …         …         …         …
995  0.019764  0.019764  0.044469  0.049410  0.019764  0.019764  0.024705
996  0.046791  0.056149  0.074865  0.074865  0.046791  0.037432  0.028074
997  0.027262  0.018175  0.066641  0.087845  0.024233  0.033320  0.012117
998  0.042355  0.018987  0.028480  0.029211  0.026290  0.035783  0.018257
999  0.034776  0.021528  0.048024  0.023184  0.034776  0.034776  0.013248

           X7        X8        X9  …       X32       X33       X34  \
```

```
0    0.037387  0.037387  0.031156  …  0.013461 -0.050976 -0.040106
1    0.039000  0.161570  0.089142  … -0.037704 -0.031762 -0.024292
2    0.067953  0.041527  0.090604  …  0.015305  0.002807  0.008122
3    0.048522  0.062386  0.124771  …  0.007061 -0.013580 -0.024570
4    0.028715  0.040201  0.048815  …  0.032962  0.016822 -0.002421
..        …       …       …  …       …        …       …
995  0.014823  0.044469  0.034587  …  0.013729  0.056828  0.015408
996  0.084223  0.037432  0.084223  … -0.017083  0.016550 -0.018610
997  0.066641  0.075728  0.066641  … -0.017891  0.026216  0.018816
998  0.030671  0.032132  0.061342  …  0.013008  0.019915  0.022802
999  0.041400  0.028152  0.066240  …  0.017184  0.012892 -0.007289

          X35       X36       X37       X38       X39       X40        X41
0    0.053950  0.027920  0.007150 -0.021643  0.001857  0.032156  efectores
1    0.000186  0.049938 -0.038859 -0.034102  0.000051 -0.086598  efectores
2    0.017960 -0.016695 -0.003150  0.004730 -0.027108  0.007860  efectores
3    0.008181  0.006137 -0.016276 -0.019855  0.025008  0.051045  efectores
4    0.040818  0.011321  0.023756  0.020459 -0.010711  0.014499  efectores
..        …       …       …       …       …       …        …
995  0.026292  0.020417  0.000930  0.027410  0.024134  0.037820  efectores
996 -0.043991  0.043330  0.040644  0.037347  0.001683  0.041546  efectores
997 -0.019502  0.019903 -0.002420 -0.001149  0.009732  0.002235  efectores
998  0.024948  0.021923  0.024412  0.019659  0.029704  0.009779  efectores
999  0.036855  0.014223  0.018561  0.007679  0.018833  0.015965  efectores

[1000 rows x 42 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass efectores nematoda dataset 4, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.045739     0.015782     0.036767     0.046119     0.031391
std       0.017835     0.014289     0.020603     0.028885     0.021062
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       0.034412     0.006589     0.023093     0.025861     0.016268
50%       0.043932     0.012609     0.034468     0.041382     0.027641
75%       0.055080     0.020760     0.046755     0.060994     0.042327
max       0.163294     0.122091     0.199508     0.200334     0.190582

                X5           X6           X7           X8           X9  …  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
mean      0.036476     0.017544     0.039731     0.042498     0.063869  …
std       0.016386     0.013109     0.022754     0.028618     0.031296  …
min       0.000000     0.000000     0.000000     0.000000     0.000000  …
25%       0.025970     0.008917     0.023907     0.023424     0.041470  …
```

```
50%       0.034694     0.014749     0.036232     0.037071     0.060757  …
75%       0.044121     0.023382     0.051979     0.054440     0.082152  …
max       0.136611     0.110407     0.182332     0.238055     0.200151  …

               X31          X32          X33          X34          X35  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.013930     0.014120     0.013311     0.011673     0.012621
std       0.028245     0.029224     0.030750     0.031498     0.029240
min      -0.141271    -0.223163    -0.261012    -0.281287    -0.129549
25%       0.001079     0.002398     0.002313    -0.000286     0.000606
50%       0.016377     0.016170     0.016651     0.015221     0.015270
75%       0.029441     0.029264     0.029484     0.028550     0.028168
max       0.184333     0.225982     0.203956     0.177667     0.278621

               X36          X37          X38          X39          X40
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.011599     0.013195     0.013116     0.012054     0.010542
std       0.031831     0.030546     0.033249     0.031912     0.030682
min      -0.202045    -0.159993    -0.256829    -0.318849    -0.257629
25%      -0.000693    -0.000290     0.000074     0.000583    -0.001784
50%       0.015287     0.015998     0.014812     0.016457     0.015056
75%       0.028714     0.028481     0.028715     0.028790     0.027539
max       0.361795     0.182073     0.418590     0.196648     0.111079

[8 rows x 41 columns]
```

no_efectores
Composición de pseudo aminoácidos (PseAAC) mass no_efectores nematoda dataset 4, con valores atípicos.
Valores del documento csv.

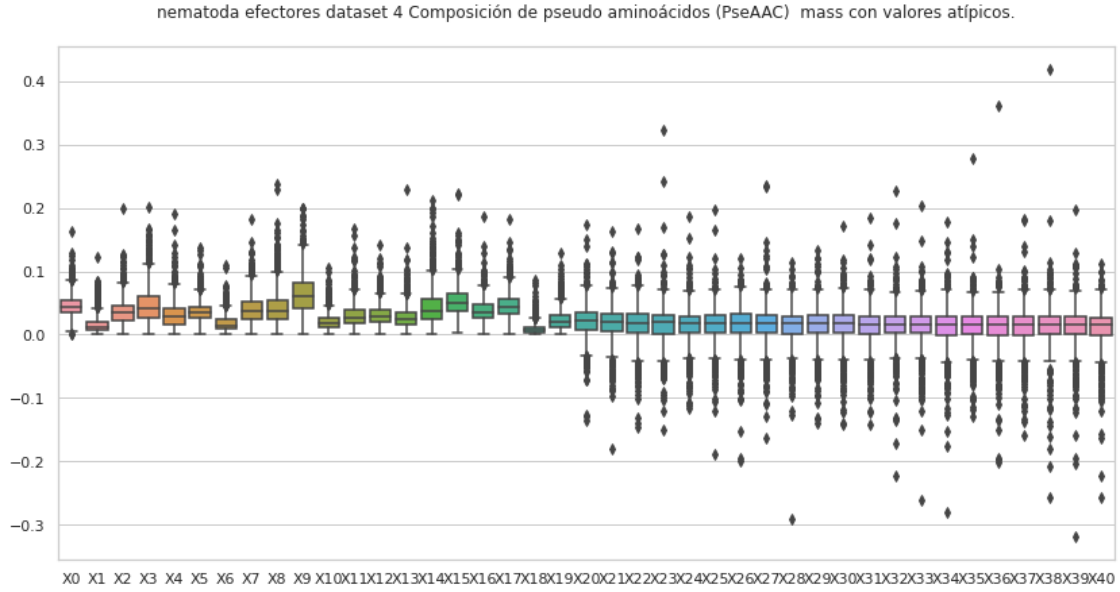
```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.024460  0.000000  0.042805  0.018345  0.003058  0.015288  0.009173
1    0.012041  0.012041  0.060204  0.054183  0.036122  0.072245  0.054183
2    0.032104  0.008026  0.042137  0.088287  0.042137  0.022072  0.020065
3    0.036051  0.021630  0.028840  0.036051  0.097337  0.014420  0.007210
4    0.039495  0.017812  0.039495  0.066600  0.050337  0.032525  0.017812
..        …         …         …         …         …         …         …
995  0.054011  0.000000  0.064813  0.054011  0.021604  0.021604  0.032407
996  0.006892  0.003181  0.006892  0.005832  0.007423  0.030751  0.002121
997  0.038256  0.020253  0.038256  0.039381  0.033755  0.050633  0.024754
998  0.021627  0.024330  0.094618  0.059474  0.067584  0.029737  0.021627
999  0.042489  0.000000  0.056652  0.127467  0.042489  0.028326  0.014163

           X7        X8        X9  …       X32       X33       X34  \
0    0.006115  0.009173  0.009173  …  0.053870  0.034065  0.015348
```

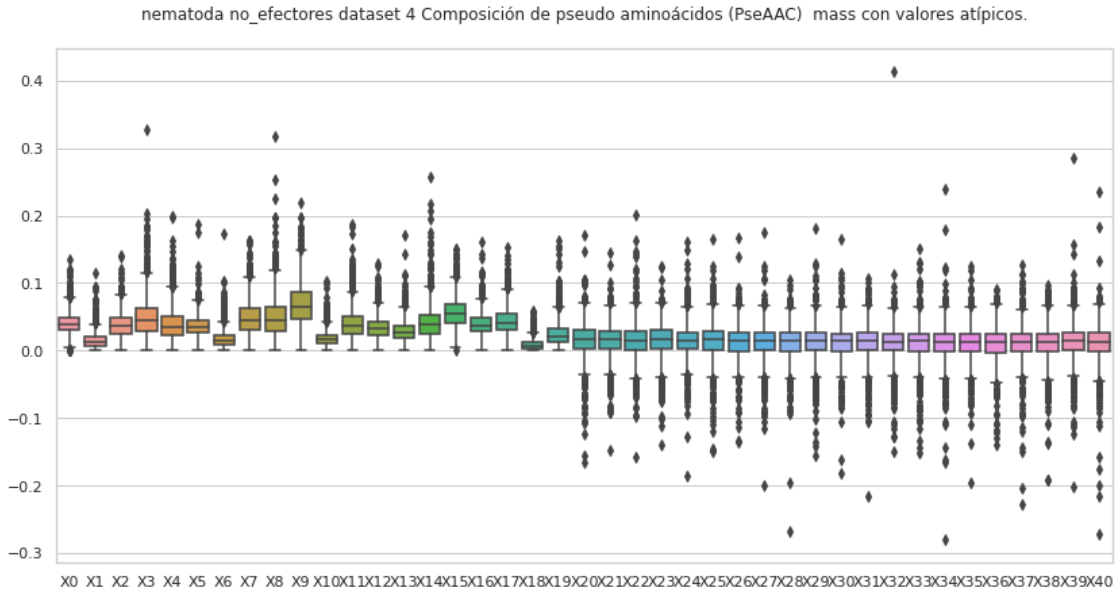
```
1    0.108367  0.012041  0.054183  … -0.014241 -0.015462 -0.005373
2    0.048157  0.054176  0.100326  … -0.002488  0.005162  0.028672
3    0.064891  0.032446  0.104547  …  0.006065 -0.005452  0.012767
4    0.064276  0.054983  0.095253  …  0.002039  0.004466 -0.003310
..        …         …        …  …        …         …
995  0.064813  0.054011  0.086418  … -0.012624 -0.022590  0.050122
996  0.007423  0.003181  0.006362  …  0.026215  0.036391  0.057871
997  0.045007  0.042757  0.070886  … -0.014994  0.025809  0.022068
998  0.072991  0.054067  0.081101  … -0.030876 -0.001083 -0.025974
999  0.099141  0.070815  0.113304  … -0.024856 -0.003808  0.044766

          X35       X36       X37       X38       X39       X40           X41
0    0.045237  0.025150  0.022296  0.048764  0.036062  0.035565  no_efectores
1    0.038207 -0.005957 -0.005611  0.077409  0.009452 -0.041323  no_efectores
2    0.002176  0.000264 -0.001477 -0.001893  0.011332  0.002472  no_efectores
3   -0.026082  0.014609 -0.027286  0.007881 -0.009153 -0.014103  no_efectores
4    0.010452  0.009896  0.010060  0.002546  0.020063  0.020518  no_efectores
..        …         …         …         …         …         …           …
995 -0.002732  0.043736 -0.019212 -0.010259  0.033892  0.001563  no_efectores
996  0.028853  0.037738  0.052648  0.025941  0.036800  0.044595  no_efectores
997  0.000374  0.032875 -0.002323  0.036680  0.028501 -0.005137  no_efectores
998  0.007404  0.008298  0.000241  0.011726 -0.011518 -0.008244  no_efectores
999  0.007168 -0.035436 -0.042281 -0.023202  0.055636  0.048609  no_efectores

[1000 rows x 42 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass no_efectores nematoda dataset 4, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.040229     0.015888     0.037230     0.049450     0.039796
std       0.017681     0.013428     0.019501     0.031592     0.025623
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       0.030101     0.007320     0.023888     0.028567     0.022202
50%       0.038417     0.012720     0.036083     0.045338     0.034968
75%       0.049574     0.020441     0.047961     0.063561     0.051180
max       0.134336     0.115918     0.141495     0.327330     0.200041

                X5           X6           X7           X8           X9  …  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
mean      0.037095     0.017762     0.047953     0.050263     0.068871  …
std       0.016973     0.013832     0.026348     0.032113     0.032300  …
min       0.000000     0.000000     0.000000     0.000000     0.000000  …
25%       0.026061     0.009406     0.030067     0.029644     0.046587  …
50%       0.035478     0.015416     0.044059     0.045502     0.065540  …
```

```
75%       0.045496     0.022788     0.062155     0.065597     0.087902  …
max       0.187616     0.173129     0.162765     0.317073     0.219889  …

                X31          X32          X33          X34          X35  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.012840     0.011045     0.011762     0.009656     0.010149
std       0.026484     0.028173     0.027612     0.029730     0.025821
min      -0.216685    -0.150072    -0.151238    -0.279985    -0.195425
25%       0.000474    -0.000030    -0.000302    -0.002233    -0.001700
50%       0.014869     0.013448     0.013809     0.012602     0.012796
75%       0.027446     0.025333     0.025533     0.024569     0.025283
max       0.106469     0.413610     0.150698     0.239577     0.125729

                X36          X37          X38          X39          X40
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.010133     0.010200     0.010161     0.013102     0.010823
std       0.026396     0.028295     0.027768     0.028589     0.030554
min      -0.140426    -0.227496    -0.192211    -0.201918    -0.272083
25%      -0.003899    -0.001161    -0.001967     0.000845    -0.002028
50%       0.012784     0.012390     0.012656     0.014273     0.013464
75%       0.025568     0.024931     0.025624     0.027332     0.026343
max       0.091220     0.127354     0.096632     0.285281     0.234348

[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 4,
sin valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.031156  0.000000  0.006231  0.049849  0.031156  0.074774  0.024925
2    0.041527  0.033976  0.030201  0.052852  0.045302  0.045302  0.026426
3    0.062386  0.027727  0.076249  0.062386  0.062386  0.027727  0.048522
4    0.031586  0.005743  0.031586  0.022972  0.014357  0.017229  0.020100
5    0.030884  0.006390  0.025560  0.023430  0.022365  0.036209  0.009585
..        …         …         …         …         …         …         …
994  0.028783  0.002617  0.049716  0.060182  0.015700  0.020933  0.002617
995  0.019764  0.019764  0.044469  0.049410  0.019764  0.019764  0.024705
997  0.027262  0.018175  0.066641  0.087845  0.024233  0.033320  0.012117
998  0.042355  0.018987  0.028480  0.029211  0.026290  0.035783  0.018257
999  0.034776  0.021528  0.048024  0.023184  0.034776  0.034776  0.013248

           X7        X8        X9  …       X32       X33       X34  \
0    0.037387  0.037387  0.031156  …  0.013461 -0.050976 -0.040106
2    0.067953  0.041527  0.090604  …  0.015305  0.002807  0.008122
3    0.048522  0.062386  0.124771  …  0.007061 -0.013580 -0.024570
4    0.028715  0.040201  0.048815  …  0.032962  0.016822 -0.002421
5    0.020235  0.028754  0.025560  …  0.033607  0.016426  0.029547
..        …         …         …  …         …         …         …
994  0.010467  0.031400  0.039249  …  0.023333  0.002736  0.030491
995  0.014823  0.044469  0.034587  …  0.013729  0.056828  0.015408
997  0.066641  0.075728  0.066641  … -0.017891  0.026216  0.018816
998  0.030671  0.032132  0.061342  …  0.013008  0.019915  0.022802
999  0.041400  0.028152  0.066240  …  0.017184  0.012892 -0.007289
```

```
          X35       X36       X37       X38       X39       X40        X41
0    0.053950  0.027920  0.007150 -0.021643  0.001857  0.032156  efectores
2    0.017960 -0.016695 -0.003150  0.004730 -0.027108  0.007860  efectores
3    0.008181  0.006137 -0.016276 -0.019855  0.025008  0.051045  efectores
4    0.040818  0.011321  0.023756  0.020459 -0.010711  0.014499  efectores
5    0.026880  0.022912  0.032158  0.007978  0.027589  0.031042  efectores
..        …         …         …         …         …         …
994  0.024826  0.037938  0.024477  0.023036  0.015702  0.028502  efectores
995  0.026292  0.020417  0.000930  0.027410  0.024134  0.037820  efectores
997 -0.019502  0.019903 -0.002420 -0.001149  0.009732  0.002235  efectores
998  0.024948  0.021923  0.024412  0.019659  0.029704  0.009779  efectores
999  0.036855  0.014223  0.018561  0.007679  0.018833  0.015965  efectores

[809 rows x 42 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass efectores nematoda dataset 4, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  809.000000  809.000000  809.000000  809.000000  809.000000  809.000000
mean     0.043600    0.013740    0.034090    0.040814    0.027851    0.034974
std      0.014367    0.010280    0.016662    0.022042    0.016615    0.012554
min      0.004918    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.034249    0.006472    0.022324    0.024288    0.015023    0.026584
50%      0.042858    0.011880    0.032736    0.038722    0.025411    0.034375
75%      0.052829    0.018741    0.043957    0.054710    0.037076    0.042026
max      0.091436    0.058186    0.097396    0.127222    0.092334    0.076088

                X6          X7          X8          X9  …         X31  \
count  809.000000  809.000000  809.000000  809.000000  …  809.000000
mean     0.015245    0.036951    0.037740    0.058286  …    0.017311
std      0.009708    0.018758    0.021706    0.025860  …    0.020537
min      0.000000    0.000000    0.000000    0.001828  …   -0.065918
25%      0.008312    0.023509    0.022134    0.039433  …    0.005675
50%      0.013482    0.035468    0.033593    0.056484  …    0.017977
75%      0.020307    0.048394    0.048031    0.075159  …    0.029803
max      0.056039    0.107502    0.126833    0.139250  …    0.084530

               X32         X33         X34         X35         X36         X37  \
count  809.000000  809.000000  809.000000  809.000000  809.000000  809.000000
mean     0.016566    0.016613    0.015428    0.015587    0.016169    0.016325
std      0.019847    0.020924    0.022062    0.020116    0.020846    0.022147
min     -0.059881   -0.070819   -0.079634   -0.071745   -0.081527   -0.076991
25%      0.005929    0.005656    0.005144    0.005181    0.004006    0.003153
50%      0.017268    0.018132    0.017515    0.016590    0.017408    0.017466
75%      0.029239    0.029477    0.029547    0.028165    0.028998    0.029000
```

```
max      0.090852    0.081733    0.091013    0.089515    0.103948    0.091384

              X38         X39         X40
count  809.000000  809.000000  809.000000
mean     0.015738    0.015900    0.015402
std      0.020428    0.020431    0.021198
min     -0.075116   -0.074608   -0.059502
25%      0.004009    0.004714    0.003297
50%      0.016629    0.017829    0.016904
75%      0.028624    0.029280    0.029111
max      0.102421    0.080387    0.085727

[8 rows x 41 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass no_efectores nematoda dataset 4, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.024460  0.000000  0.042805  0.018345  0.003058  0.015288  0.009173
1    0.012041  0.012041  0.060204  0.054183  0.036122  0.072245  0.054183
2    0.032104  0.008026  0.042137  0.088287  0.042137  0.022072  0.020065
3    0.036051  0.021630  0.028840  0.036051  0.097337  0.014420  0.007210
4    0.039495  0.017812  0.039495  0.066600  0.050337  0.032525  0.017812
..        …         …         …         …         …         …         …
993  0.048922  0.013978  0.037740  0.071286  0.041933  0.025160  0.022364
994  0.023775  0.005944  0.041607  0.035663  0.007925  0.021794  0.017832
996  0.006892  0.003181  0.006892  0.005832  0.007423  0.030751  0.002121
997  0.038256  0.020253  0.038256  0.039381  0.033755  0.050633  0.024754
998  0.021627  0.024330  0.094618  0.059474  0.067584  0.029737  0.021627

           X7        X8        X9  …       X32       X33       X34  \
0    0.006115  0.009173  0.009173  …  0.053870  0.034065  0.015348
1    0.108367  0.012041  0.054183  … -0.014241 -0.015462 -0.005373
2    0.048157  0.054176  0.100326  … -0.002488  0.005162  0.028672
3    0.064891  0.032446  0.104547  …  0.006065 -0.005452  0.012767
4    0.064276  0.054983  0.095253  …  0.002039  0.004466 -0.003310
..        …         …         …  …       …         …         …
993  0.072683  0.075479  0.100639  … -0.014828 -0.016504 -0.009594
994  0.009906  0.013869  0.043588  …  0.028070  0.043984  0.020006
996  0.007423  0.003181  0.006362  …  0.026215  0.036391  0.057871
997  0.045007  0.042757  0.070886  … -0.014994  0.025809  0.022068
998  0.072991  0.054067  0.081101  … -0.030876 -0.001083 -0.025974

          X35       X36       X37       X38       X39       X40           X41
0    0.045237  0.025150  0.022296  0.048764  0.036062  0.035565  no_efectores
```

```
1    0.038207 -0.005957 -0.005611  0.077409  0.009452 -0.041323  no_efectores
2    0.002176  0.000264 -0.001477 -0.001893  0.011332  0.002472  no_efectores
3   -0.026082  0.014609 -0.027286  0.007881 -0.009153 -0.014103  no_efectores
4    0.010452  0.009896  0.010060  0.002546  0.020063  0.020518  no_efectores
..        …         …         …         …         …         …
993  0.016542  0.006462  0.002558 -0.004996  0.023625  0.021867  no_efectores
994  0.002405  0.016667  0.021092  0.020864 -0.000388  0.022974  no_efectores
996  0.028853  0.037738  0.052648  0.025941  0.036800  0.044595  no_efectores
997  0.000374  0.032875 -0.002323  0.036680  0.028501 -0.005137  no_efectores
998  0.007404  0.008298  0.000241  0.011726 -0.011518 -0.008244  no_efectores

[816 rows x 42 columns]
```

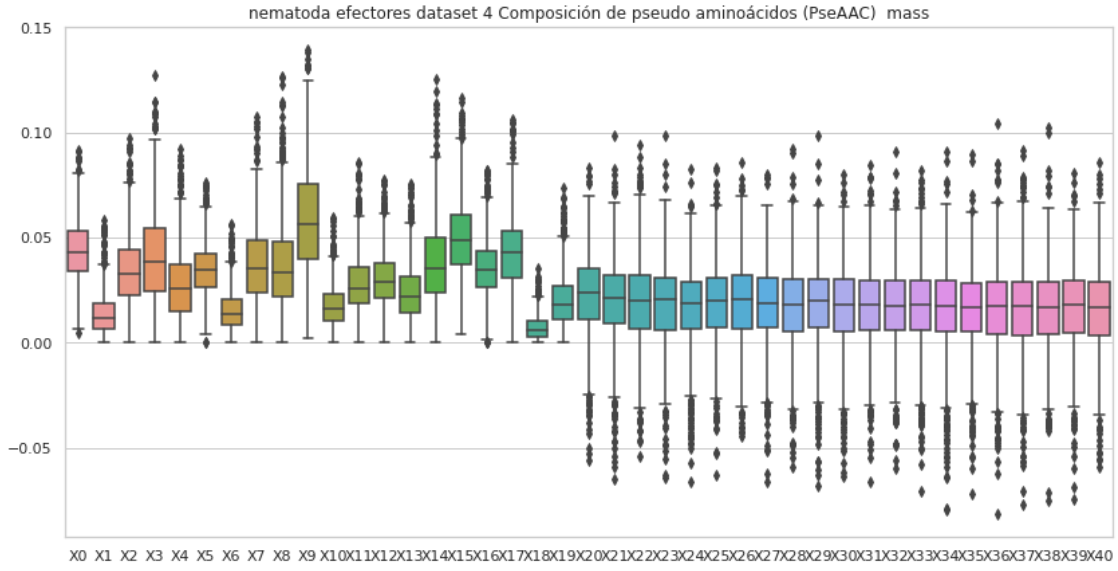
Composición de pseudo aminoácidos (PseAAC) mass no_efectores nematoda dataset 4, sin valores atípicos.
Estadísticas.

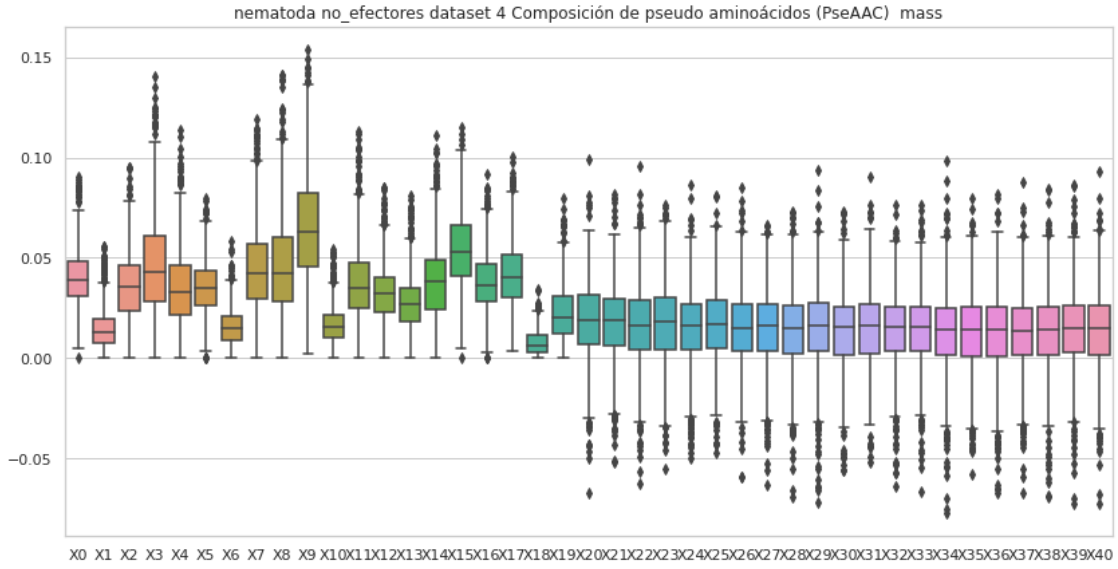
	X0	X1	X2	X3	X4	X5	\
count	816.000000	816.000000	816.000000	816.000000	816.000000	816.000000	
mean	0.039560	0.014949	0.035777	0.045731	0.035376	0.035634	
std	0.014617	0.010526	0.016828	0.024807	0.019478	0.012907	
min	0.000131	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.030677	0.007562	0.023931	0.028036	0.021591	0.026417	
50%	0.038930	0.012644	0.035583	0.043057	0.032701	0.034917	
75%	0.048415	0.019845	0.046509	0.060650	0.046395	0.043481	
max	0.090455	0.055641	0.095109	0.140412	0.113882	0.079757	

	X6	X7	X8	X9	…	X31	\
count	816.000000	816.000000	816.000000	816.000000	…	816.000000	
mean	0.015848	0.044731	0.045523	0.064613	…	0.014888	
std	0.009126	0.022084	0.024862	0.027514	…	0.019554	
min	0.000000	0.000263	0.000000	0.002244	…	-0.051860	
25%	0.009223	0.029760	0.028411	0.045820	…	0.002353	
50%	0.014888	0.042525	0.042626	0.063025	…	0.016298	
75%	0.021212	0.057262	0.060587	0.082531	…	0.027271	
max	0.058520	0.119082	0.140853	0.153658	…	0.090130	

	X32	X33	X34	X35	X36	X37	\
count	816.000000	816.000000	816.000000	816.000000	816.000000	816.000000	
mean	0.014135	0.014555	0.012666	0.012546	0.012755	0.012733	
std	0.018671	0.019246	0.020621	0.019865	0.020744	0.019749	
min	-0.063855	-0.066231	-0.077118	-0.058001	-0.066800	-0.067056	
25%	0.003479	0.003385	0.001594	0.001252	0.000738	0.001681	
50%	0.015396	0.015436	0.014346	0.013968	0.014296	0.013791	
75%	0.025841	0.025806	0.025243	0.025740	0.025806	0.025221	
max	0.076383	0.076387	0.098602	0.079356	0.081393	0.087595	

```
              X38          X39          X40
count  816.000000   816.000000   816.000000
mean     0.012925     0.014027     0.013739
std      0.020513     0.020106     0.020027
min     -0.069122    -0.072606    -0.072714
25%      0.001827     0.002910     0.001431
50%      0.014071     0.014948     0.014721
75%      0.025701     0.026025     0.026372
max      0.084036     0.086265     0.093259

[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 4, con valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.024303	0.000000	0.004861	0.038885	0.024303	0.058327	0.019442	
1	0.011485	0.011485	0.021693	0.038282	0.010208	0.012761	0.007656	
2	0.054088	0.044254	0.039336	0.068839	0.059005	0.059005	0.034419	
3	0.083388	0.037061	0.101919	0.083388	0.083388	0.037061	0.064857	
4	0.025209	0.004583	0.025209	0.018334	0.011459	0.013750	0.016042	
..	…	…	…	…	…	…	…	
995	0.018877	0.018877	0.042473	0.047192	0.018877	0.018877	0.023596	
996	0.091965	0.110358	0.147144	0.147144	0.091965	0.073572	0.055179	
997	0.009511	0.006341	0.023250	0.030648	0.008455	0.011625	0.004227	
998	0.084230	0.037758	0.056638	0.058090	0.052281	0.071160	0.036306	
999	0.028399	0.017580	0.039218	0.018933	0.028399	0.028399	0.010819	

	X7	X8	X9	…	X53	X54	X55	\
0	0.029163	0.029163	0.024303	…	0.060656	0.028392	0.016374	
1	0.008932	0.037006	0.020417	…	0.027957	-0.003450	0.013792	
2	0.088507	0.054088	0.118009	…	0.084570	-0.042173	-0.036058	
3	0.064857	0.083388	0.166776	…	-0.046430	-0.029511	-0.076904	
4	0.022917	0.032084	0.038959	…	-0.013659	-0.005171	0.000568	
..	…	…	…	…	…	…	…	
995	0.014158	0.042473	0.033034	…	0.018032	0.059823	0.084079	
996	0.165537	0.073572	0.165537	…	0.047943	0.074204	0.043863	
997	0.023250	0.026421	0.023250	…	0.025755	0.022731	0.017291	
998	0.060994	0.063899	0.121989	…	0.009721	-0.012301	-0.001233	
999	0.033809	0.022990	0.054094	…	-0.003816	-0.003205	-0.005452	

	X56	X57	X58	X59	X60	X61	X62
0	0.004145	0.011839	-0.011834	0.008247	-0.004571	-0.001074	efectores
1	-0.015124	0.006430	0.019208	0.033855	0.015037	0.027065	efectores
2	-0.038895	-0.029058	0.049971	-0.013994	0.059967	0.037089	efectores
3	0.062550	0.006159	0.082912	0.057655	0.021127	-0.020474	efectores
4	0.030276	0.019133	0.005940	0.003514	0.042491	0.033347	efectores
..	…	…	…	…	…	…	…
995	0.048561	0.050759	0.042645	0.056350	0.010908	0.018808	efectores
996	-0.279101	-0.207425	-0.174610	-0.002180	0.063647	0.034659	efectores
997	0.001909	0.009143	0.008660	0.021010	-0.003324	0.027810	efectores

```
998  0.022202  0.014739 -0.002738 -0.029263  0.052492  0.011617  efectores
999  0.013599  0.025965 -0.004127  0.001943  0.011564  0.028902  efectores

[1000 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro efectores nematoda dataset 4, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.044251     0.013679     0.032868     0.042260     0.029812
std       0.220929     0.094339     0.157789     0.100210     0.081695
min      -6.534626    -2.178209    -3.820143    -2.178209    -1.910071
25%       0.028669     0.005851     0.021025     0.026118     0.016114
50%       0.046486     0.013001     0.035708     0.042819     0.028147
75%       0.066421     0.023557     0.052867     0.058640     0.041835
max       0.739123     0.171720     0.923903     0.739123     0.237408

                X5           X6           X7           X8           X9  …  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
mean      0.037708     0.016175     0.034752     0.038583     0.051663  …
std       0.188179     0.079673     0.170741     0.082313     0.416152  …
min      -5.730214    -1.910071    -3.820143    -1.910071    -9.550357  …
25%       0.021670     0.007543     0.022132     0.023643     0.037833  …
50%       0.035333     0.014620     0.038089     0.037409     0.060219  …
75%       0.056522     0.025083     0.056464     0.053869     0.089600  …
max       0.554342     0.369561     0.259468     0.283033     1.293465  …

               X52          X53          X54          X55          X56  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.003642     0.014114     0.006109     0.006729     0.004437
std       0.321892     0.140193     0.433378     0.197044     0.225582
min      -5.390572    -0.268170    -4.440178    -3.271405    -1.902555
25%      -0.015837    -0.004327    -0.017000    -0.007627    -0.016167
50%       0.004876     0.011908     0.002918     0.009752     0.004288
75%       0.020015     0.025771     0.018862     0.023591     0.020451
max       8.490951     4.224455    12.847298     5.141992     6.706872

               X57          X58          X59          X60          X61
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.014261     0.016039     0.018858     0.008829     0.003202
std       0.477822     0.448131     0.329678     0.217080     0.416362
min      -4.801111    -1.435807    -0.437214    -1.235083   -11.338064
25%      -0.007194    -0.012554    -0.005758    -0.012371    -0.005631
50%       0.009282     0.006600     0.011065     0.005735     0.011954
75%       0.024871     0.022316     0.025874     0.024081     0.027093
```

```
max      14.149628    13.971111    10.311849     6.456712     6.525238

[8 rows x 62 columns]
```

no_efectores
Composición de pseudo aminoácidos (PseAAC) hidro no_efectores nematoda dataset 4, con valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.020634	0.000000	0.036109	0.015475	0.002579	0.012896	0.007738	
1	0.014654	0.014654	0.073272	0.065945	0.043963	0.087926	0.065945	
2	0.021266	0.005316	0.027911	0.058481	0.027911	0.014620	0.013291	
3	0.014480	0.008688	0.011584	0.014480	0.039096	0.005792	0.002896	
4	0.039613	0.017865	0.039613	0.066799	0.050487	0.032623	0.017865	
..	…	…	…	…	…	…	…	
995	0.056979	0.000000	0.068375	0.056979	0.022792	0.022792	0.034187	
996	0.027940	0.012895	0.027940	0.023642	0.030089	0.124656	0.008597	
997	0.053264	0.028199	0.053264	0.054831	0.046998	0.070497	0.034465	
998	0.024023	0.027026	0.105101	0.066064	0.075072	0.033032	0.024023	
999	0.045862	0.000000	0.061150	0.137587	0.045862	0.030575	0.015287	

	X7	X8	X9	…	X53	X54	X55	\
0	0.005158	0.007738	0.007738	…	0.033734	0.009663	0.027564	
1	0.131890	0.014654	0.065945	…	-0.052488	0.045608	0.003510	
2	0.031899	0.035886	0.066456	…	0.026830	0.024593	0.041873	
3	0.026064	0.013032	0.041992	…	-0.000351	0.008830	0.003177	
4	0.064469	0.055148	0.095538	…	0.013614	-0.006009	0.017341	
..	…	…	…	…	…	…	…	
995	0.068375	0.056979	0.091166	…	-0.004153	-0.123187	0.032273	
996	0.030089	0.012895	0.025791	…	0.005970	-0.007269	0.000814	
997	0.062664	0.059531	0.098695	…	-0.005267	0.007244	0.014143	
998	0.081078	0.060058	0.090087	…	0.006234	-0.010402	-0.031553	
999	0.107012	0.076437	0.122300	…	0.037446	-0.012612	-0.036930	

	X56	X57	X58	X59	X60	X61	X62
0	0.006526	0.051213	-0.009632	0.012527	0.012969	0.039473	no_efectores
1	-0.015658	-0.026905	-0.046913	-0.025083	-0.007805	-0.067153	no_efectores
2	-0.001751	-0.006358	-0.003600	0.034460	-0.002407	0.017921	no_efectores
3	0.022805	0.011853	0.003106	-0.012950	0.015807	0.010594	no_efectores
4	0.014351	0.028148	0.020899	0.022317	0.005402	0.008979	no_efectores
..	…	…	…	…	…	…	…
995	0.066337	0.049961	-0.053522	0.061591	0.042430	0.052401	no_efectores
996	0.021049	0.011813	0.008788	0.009757	0.007435	-0.005388	no_efectores
997	0.009622	0.013554	0.003451	-0.005808	-0.005854	-0.002937	no_efectores
998	0.007572	-0.001934	0.054239	0.077645	-0.039412	-0.039070	no_efectores

```
999 -0.077966 -0.003527 -0.024474 -0.032376 -0.079350 -0.036140  no_efectores

[1000 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro no_efectores nematoda dataset 4, con valores atípicos.
Estadísticas.

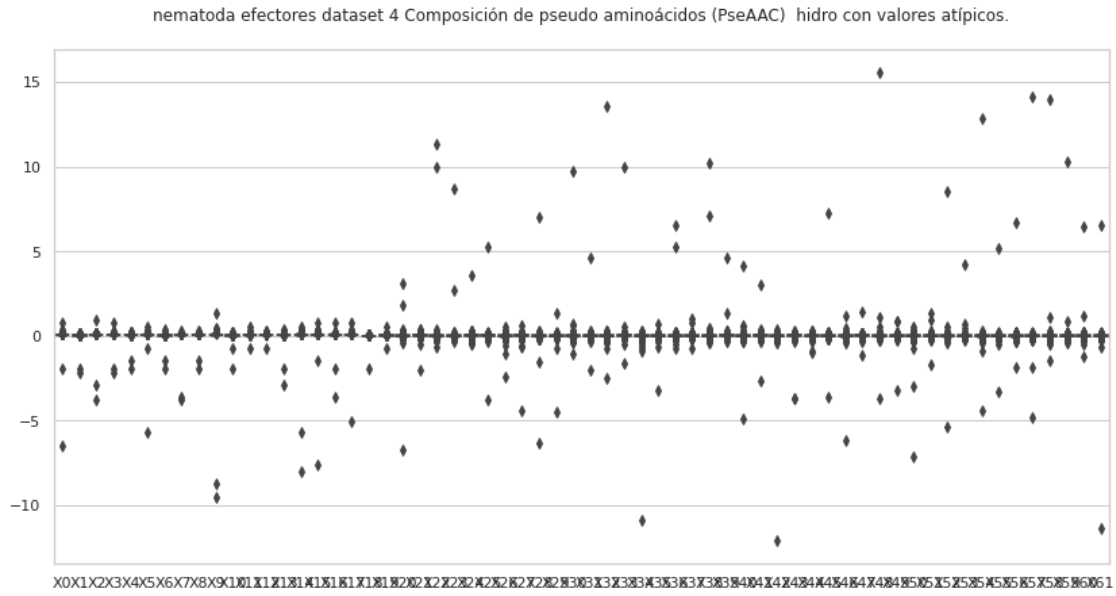
	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.040034	0.015757	0.035471	0.043913	0.037242	
std	0.028462	0.016373	0.023849	0.029943	0.030866	
min	-0.158073	-0.008195	-0.105382	-0.105382	-0.263454	
25%	0.022406	0.006058	0.018731	0.025030	0.019260	
50%	0.034894	0.011808	0.032417	0.041581	0.031529	
75%	0.051655	0.020459	0.047475	0.056872	0.047755	
max	0.310841	0.207227	0.238195	0.518068	0.285834	

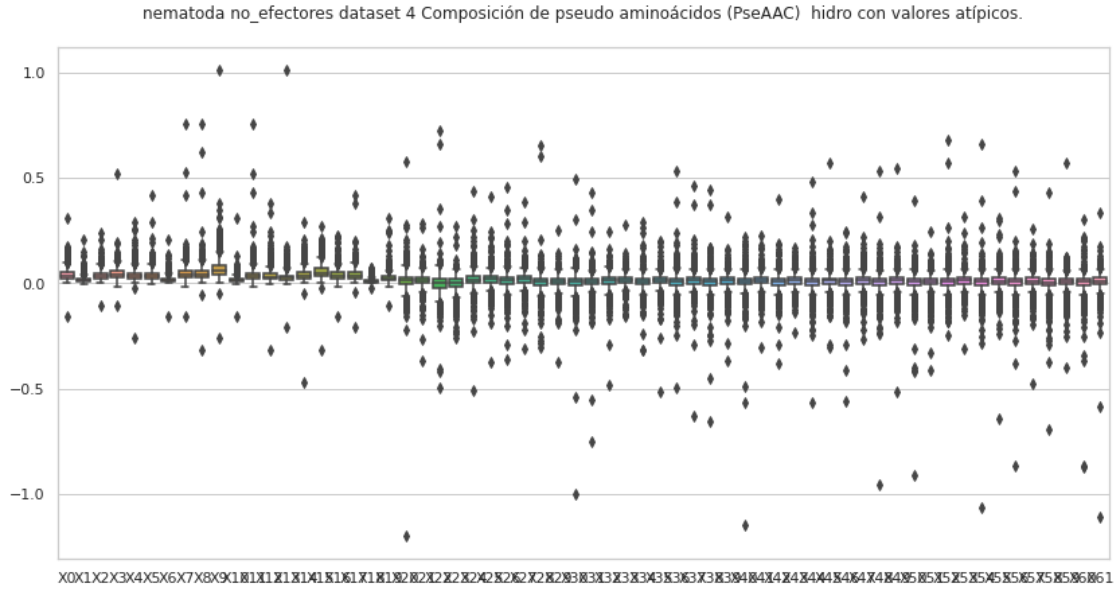
	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.038271	0.017099	0.046320	0.045908	0.066668	…	
std	0.029839	0.017270	0.042395	0.043974	0.053789	…	
min	-0.008195	-0.158073	-0.158073	-0.316145	-0.263454	…	
25%	0.019247	0.007343	0.024193	0.024379	0.038269	…	
50%	0.031783	0.013714	0.039213	0.040922	0.058036	…	
75%	0.049157	0.022349	0.058100	0.058829	0.083244	…	
max	0.414455	0.207227	0.757140	0.757140	1.009520	…	

	X52	X53	X54	X55	X56	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.004500	0.010478	0.000220	0.008161	0.001865	
std	0.047000	0.033492	0.056857	0.041987	0.054180	
min	-0.246398	-0.311690	-1.066054	-0.641001	-0.863911	
25%	-0.010601	-0.003200	-0.012400	-0.003715	-0.012019	
50%	0.006349	0.011686	0.005421	0.010713	0.005087	
75%	0.018988	0.025414	0.018143	0.024320	0.019339	
max	0.679434	0.312398	0.661113	0.306970	0.530677	

	X57	X58	X59	X60	X61
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.008119	0.000786	0.006873	0.000240	0.006222
std	0.037298	0.048628	0.039080	0.057861	0.053009
min	-0.476691	-0.697423	-0.401770	-0.870778	-1.106541
25%	-0.003502	-0.011252	-0.003906	-0.011919	-0.005577
50%	0.009592	0.003980	0.009776	0.004618	0.009562
75%	0.023854	0.018253	0.021781	0.019289	0.024479
max	0.255729	0.429037	0.569158	0.303261	0.336058

[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 4, sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.024303	0.000000	0.004861	0.038885	0.024303	0.058327	0.019442	
1	0.011485	0.011485	0.021693	0.038282	0.010208	0.012761	0.007656	
2	0.054088	0.044254	0.039336	0.068839	0.059005	0.059005	0.034419	
3	0.083388	0.037061	0.101919	0.083388	0.083388	0.037061	0.064857	
4	0.025209	0.004583	0.025209	0.018334	0.011459	0.013750	0.016042	
..	…	…	…	…	…	…	…	
995	0.018877	0.018877	0.042473	0.047192	0.018877	0.018877	0.023596	
996	0.091965	0.110358	0.147144	0.147144	0.091965	0.073572	0.055179	
997	0.009511	0.006341	0.023250	0.030648	0.008455	0.011625	0.004227	
998	0.084230	0.037758	0.056638	0.058090	0.052281	0.071160	0.036306	
999	0.028399	0.017580	0.039218	0.018933	0.028399	0.028399	0.010819	

	X7	X8	X9	…	X53	X54	X55	\
0	0.029163	0.029163	0.024303	…	0.060656	0.028392	0.016374	
1	0.008932	0.037006	0.020417	…	0.027957	-0.003450	0.013792	
2	0.088507	0.054088	0.118009	…	0.084570	-0.042173	-0.036058	
3	0.064857	0.083388	0.166776	…	-0.046430	-0.029511	-0.076904	
4	0.022917	0.032084	0.038959	…	-0.013659	-0.005171	0.000568	
..	…	…	…	…	…	…	…	
995	0.014158	0.042473	0.033034	…	0.018032	0.059823	0.084079	
996	0.165537	0.073572	0.165537	…	0.047943	0.074204	0.043863	
997	0.023250	0.026421	0.023250	…	0.025755	0.022731	0.017291	
998	0.060994	0.063899	0.121989	…	0.009721	-0.012301	-0.001233	
999	0.033809	0.022990	0.054094	…	-0.003816	-0.003205	-0.005452	

	X56	X57	X58	X59	X60	X61	X62
0	0.004145	0.011839	-0.011834	0.008247	-0.004571	-0.001074	efectores
1	-0.015124	0.006430	0.019208	0.033855	0.015037	0.027065	efectores
2	-0.038895	-0.029058	0.049971	-0.013994	0.059967	0.037089	efectores
3	0.062550	0.006159	0.082912	0.057655	0.021127	-0.020474	efectores
4	0.030276	0.019133	0.005940	0.003514	0.042491	0.033347	efectores
..	…	…	…	…	…	…	…
995	0.048561	0.050759	0.042645	0.056350	0.010908	0.018808	efectores
996	-0.279101	-0.207425	-0.174610	-0.002180	0.063647	0.034659	efectores
997	0.001909	0.009143	0.008660	0.021010	-0.003324	0.027810	efectores
998	0.022202	0.014739	-0.002738	-0.029263	0.052492	0.011617	efectores
999	0.013599	0.025965	-0.004127	0.001943	0.011564	0.028902	efectores

[972 rows x 63 columns]

Composición de pseudo aminoácidos (PseAAC) efectores nematoda dataset 4, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	972.000000	972.000000	972.000000	972.000000	972.000000	972.000000	
mean	0.050488	0.017060	0.038281	0.044554	0.031886	0.041851	
std	0.031078	0.017258	0.023147	0.025404	0.023069	0.029870	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.028455	0.005854	0.021083	0.025999	0.015979	0.021524	
50%	0.045949	0.012793	0.035535	0.042463	0.027813	0.035053	
75%	0.064962	0.022489	0.051889	0.058119	0.041239	0.055578	
max	0.356112	0.136235	0.158391	0.202610	0.237408	0.205908	

	X6	X7	X8	X9	…	X52	\
count	972.000000	972.000000	972.000000	972.000000	…	972.000000	
mean	0.018442	0.041186	0.040974	0.065973	…	0.000946	
std	0.015958	0.025282	0.024398	0.039228	…	0.041335	
min	0.000000	0.000000	0.000000	0.000000	…	-0.279868	
25%	0.007543	0.022110	0.023565	0.037770	…	-0.014604	
50%	0.014534	0.038002	0.037037	0.059115	…	0.005103	
75%	0.024643	0.055904	0.053281	0.088039	…	0.019851	
max	0.127155	0.169286	0.192280	0.348460	…	0.280519	

	X53	X54	X55	X56	X57	X58	\
count	972.000000	972.000000	972.000000	972.000000	972.000000	972.000000	
mean	0.009415	-0.000108	0.007339	0.001106	0.007957	0.003465	
std	0.031729	0.039098	0.032525	0.040842	0.033683	0.045154	
min	-0.194787	-0.247987	-0.177784	-0.279101	-0.215162	-0.401919	
25%	-0.003932	-0.015891	-0.006785	-0.015609	-0.006448	-0.012157	
50%	0.012002	0.003020	0.010405	0.004405	0.009763	0.006645	
75%	0.025652	0.018749	0.023858	0.020451	0.024876	0.022179	
max	0.170478	0.148401	0.201248	0.154188	0.156722	0.312093	

	X59	X60	X61
count	972.000000	972.000000	972.000000
mean	0.009177	0.003244	0.009872
std	0.036395	0.044495	0.034399
min	-0.188848	-0.528452	-0.261153
25%	-0.005299	-0.012075	-0.005080
50%	0.011399	0.005777	0.012202
75%	0.025851	0.023917	0.026802
max	0.169614	0.186936	0.134981

[8 rows x 62 columns]

```
no_efectores
Composición de pseudo aminoácidos (PseAAC)  no_efectores nematoda dataset 4, sin
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.020634  0.000000  0.036109  0.015475  0.002579  0.012896  0.007738
1    0.014654  0.014654  0.073272  0.065945  0.043963  0.087926  0.065945
2    0.021266  0.005316  0.027911  0.058481  0.027911  0.014620  0.013291
3    0.014480  0.008688  0.011584  0.014480  0.039096  0.005792  0.002896
4    0.039613  0.017865  0.039613  0.066799  0.050487  0.032623  0.017865
..        …         …         …         …         …         …
991  0.125259  0.038541  0.019271  0.077083  0.009635  0.048177  0.009635
993  0.036759  0.010503  0.028357  0.053563  0.031508  0.018905  0.016804
994  0.018096  0.004524  0.031668  0.027144  0.006032  0.016588  0.013572
997  0.053264  0.028199  0.053264  0.054831  0.046998  0.070497  0.034465
998  0.024023  0.027026  0.105101  0.066064  0.075072  0.033032  0.024023

           X7        X8        X9  …        X53       X54       X55  \
0    0.005158  0.007738  0.007738  …   0.033734  0.009663  0.027564
1    0.131890  0.014654  0.065945  … -0.052488  0.045608  0.003510
2    0.031899  0.035886  0.066456  …   0.026830  0.024593  0.041873
3    0.026064  0.013032  0.041992  … -0.000351  0.008830  0.003177
4    0.064469  0.055148  0.095538  …   0.013614 -0.006009  0.017341
..        …         …         …  …         …         …         …
991  0.067447  0.057812  0.115624  … -0.012819 -0.027096  0.038037
993  0.054613  0.056714  0.075618  …   0.026042  0.013491  0.017347
994  0.007540  0.010556  0.033176  …   0.006119  0.029957  0.035947
997  0.062664  0.059531  0.098695  … -0.005267  0.007244  0.014143
998  0.081078  0.060058  0.090087  …   0.006234 -0.010402 -0.031553

          X56       X57       X58       X59       X60       X61           X62
0    0.006526  0.051213 -0.009632  0.012527  0.012969  0.039473  no_efectores
1   -0.015658 -0.026905 -0.046913 -0.025083 -0.007805 -0.067153  no_efectores
2   -0.001751 -0.006358 -0.003600  0.034460 -0.002407  0.017921  no_efectores
3    0.022805  0.011853  0.003106 -0.012950  0.015807  0.010594  no_efectores
4    0.014351  0.028148  0.020899  0.022317  0.005402  0.008979  no_efectores
..        …         …         …         …         …         …
991 -0.032443 -0.007979  0.002036 -0.047143  0.018200 -0.016032  no_efectores
993  0.009686  0.029500 -0.031741  0.007278 -0.003108  0.000652  no_efectores
994 -0.009714  0.014319  0.001866  0.021094  0.028312  0.016303  no_efectores
997  0.009622  0.013554  0.003451 -0.005808 -0.005854 -0.002937  no_efectores
998  0.007572 -0.001934  0.054239  0.077645 -0.039412 -0.039070  no_efectores

[855 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC) no_efectores nematoda dataset 4, sin valores atípicos.
Estadísticas.

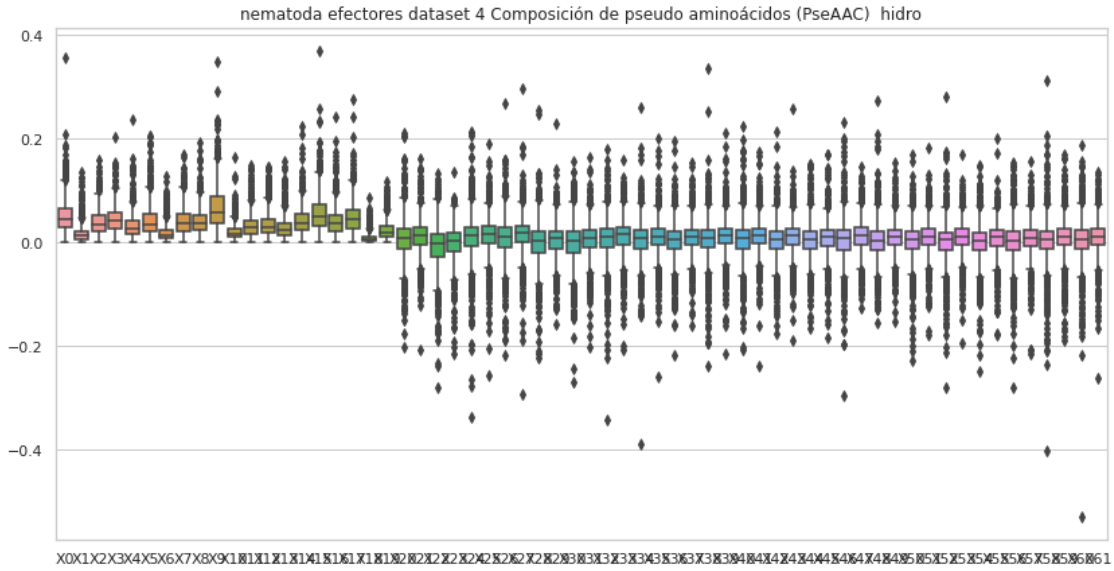
	X0	X1	X2	X3	X4	X5	\
count	855.000000	855.000000	855.000000	855.000000	855.000000	855.000000	
mean	0.036192	0.013968	0.031921	0.040154	0.031986	0.034336	
std	0.020815	0.011348	0.018388	0.021191	0.019169	0.021462	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.021170	0.006015	0.017254	0.024276	0.018205	0.018438	
50%	0.033163	0.011140	0.029558	0.040157	0.029536	0.029841	
75%	0.048074	0.019020	0.044345	0.053046	0.043070	0.045276	
max	0.125259	0.064469	0.105101	0.120302	0.112180	0.113562	

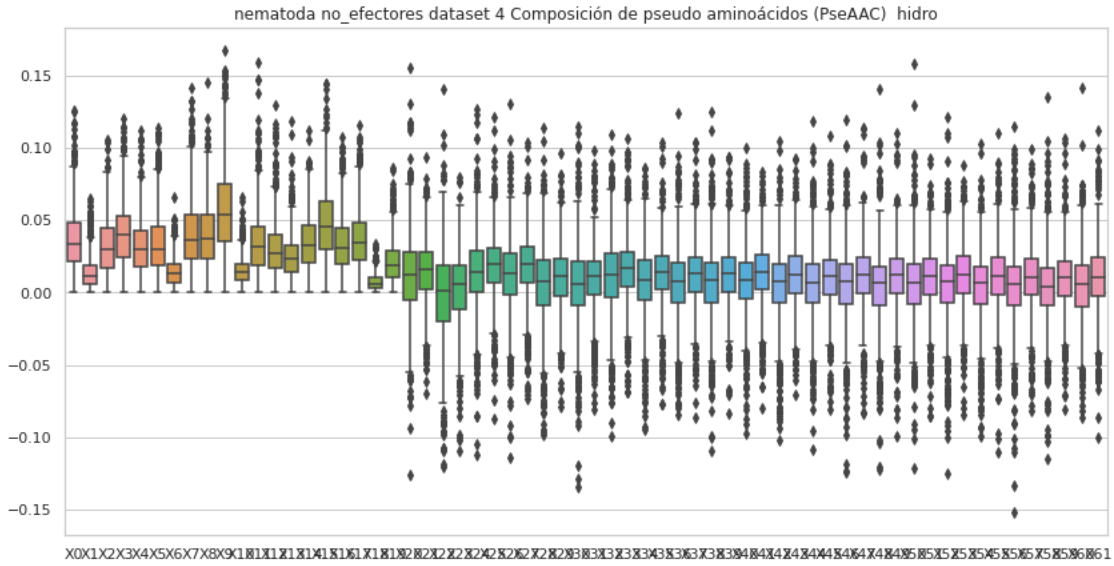
	X6	X7	X8	X9	…	X52	\
count	855.000000	855.000000	855.000000	855.000000	…	855.000000	
mean	0.014357	0.039995	0.039882	0.057339	…	0.005022	
std	0.009996	0.023300	0.021821	0.029442	…	0.025945	
min	0.000000	0.000000	0.000000	0.000000	…	-0.124713	
25%	0.007003	0.022866	0.023251	0.035278	…	-0.007557	
50%	0.012761	0.036379	0.037528	0.054051	…	0.007217	
75%	0.020049	0.054100	0.054000	0.075012	…	0.018376	
max	0.065945	0.141724	0.144721	0.167086	…	0.120883	

	X53	X54	X55	X56	X57	X58	\
count	855.000000	855.000000	855.000000	855.000000	855.000000	855.000000	
mean	0.011329	0.004083	0.010333	0.003732	0.010267	0.003956	
std	0.022010	0.024984	0.022949	0.027137	0.022539	0.025744	
min	-0.084805	-0.099472	-0.110427	-0.151861	-0.081931	-0.115165	
25%	-0.000273	-0.008170	-0.001175	-0.009383	-0.001773	-0.008695	
50%	0.012619	0.006377	0.011759	0.005535	0.010180	0.004407	
75%	0.025172	0.017867	0.023968	0.017838	0.023034	0.017274	
max	0.088017	0.102157	0.109554	0.114812	0.098998	0.134454	

	X59	X60	X61
count	855.000000	855.000000	855.000000
mean	0.009468	0.004695	0.010300
std	0.020886	0.025536	0.024441
min	-0.080471	-0.086084	-0.099988
25%	-0.002186	-0.009515	-0.002526
50%	0.010278	0.006058	0.010447
75%	0.021040	0.018934	0.024227
max	0.106585	0.141148	0.111943

[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 4,
con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.218244  0.165539  0.056992  0.004227  0.105088 -0.095810  0.028425
1   -0.004767  0.000896  0.027689 -0.034625 -0.071647  0.048405  0.116707
2    0.048562  0.017845  0.013823 -0.053341 -0.035738 -0.070553  0.024781
3   -0.027969  0.022099 -0.068296 -0.010149 -0.030623 -0.070653  0.019133
4    0.013399  0.025362  0.053115  0.044309  0.042439  0.042985  0.008109
..        …         …         …         …         …         …         …
995 -0.004752  0.037677 -0.081200 -0.065022  0.032421 -0.024451  0.098704
996  0.120346 -0.015528 -0.066367 -0.113780 -0.087586 -0.150389  0.024614
997 -0.077321 -0.006474  0.027995  0.008948  0.022489 -0.028513 -0.026591
998  0.014293 -0.022930 -0.001020  0.037801  0.009239 -0.034653 -0.008831
999  0.025367  0.018190  0.067727  0.059868 -0.007035 -0.036952 -0.008875

           X7        X8        X9       X10       X11       X12        X13
0   -0.133577  0.008114 -0.039043  0.079485  0.202914 -0.003074  efectores
```

```
1   -0.099995 -0.016323  0.000019 -0.058355 -0.110902 -0.056524  efectores
2   -0.051986  0.045369  0.055354 -0.012706  0.033834  0.020331  efectores
3    0.049939 -0.054186  0.014173 -0.107624  0.020429  0.009875  efectores
4    0.038602  0.003697  0.059413 -0.004282  0.009442  0.074815  efectores
..        …       …       …       …       …       …        …
995 -0.015361  0.020740 -0.054453 -0.100104 -0.021016 -0.021210  efectores
996  0.097103  0.090843 -0.016548 -0.074788 -0.059871 -0.041684  efectores
997  0.005472  0.056446 -0.009037  0.025306  0.047533 -0.047140  efectores
998  0.018961 -0.016942  0.044234 -0.013668  0.004967 -0.006670  efectores
999  0.022677  0.028692 -0.071891 -0.075295  0.052553  0.015800  efectores

[1000 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro_mass efectores nematoda dataset 4, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.019717     0.010988     0.012530     0.006967     0.002797
std       0.070778     0.071387     0.082123     0.074978     0.070583
min      -0.397525    -0.280178    -0.421457    -0.325324    -0.380042
25%      -0.019637    -0.027366    -0.033221    -0.031188    -0.034945
50%       0.020963     0.011466     0.009165     0.006303     0.005283
75%       0.058358     0.048684     0.052210     0.043394     0.041829
max       0.393351     0.405397     0.624939     0.738565     0.315124

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.010352     0.008773     0.006263     0.006164     0.005261
std       0.077535     0.071488     0.073118     0.077218     0.070304
min      -0.269688    -0.395533    -0.334495    -0.370642    -0.281060
25%      -0.031376    -0.028510    -0.032034    -0.032471    -0.033774
50%       0.011264     0.008941     0.007122     0.004097     0.006086
75%       0.044203     0.049040     0.043567     0.043403     0.043942
max       0.723713     0.268890     0.528696     0.701850     0.303544

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.004195     0.005732     0.003884
std       0.072098     0.076366     0.070976
min      -0.355985    -0.291456    -0.427231
25%      -0.033910    -0.032866    -0.034704
50%       0.004146     0.004703     0.003367
75%       0.043379     0.040880     0.041065
max       0.371363     0.602932     0.530164
```

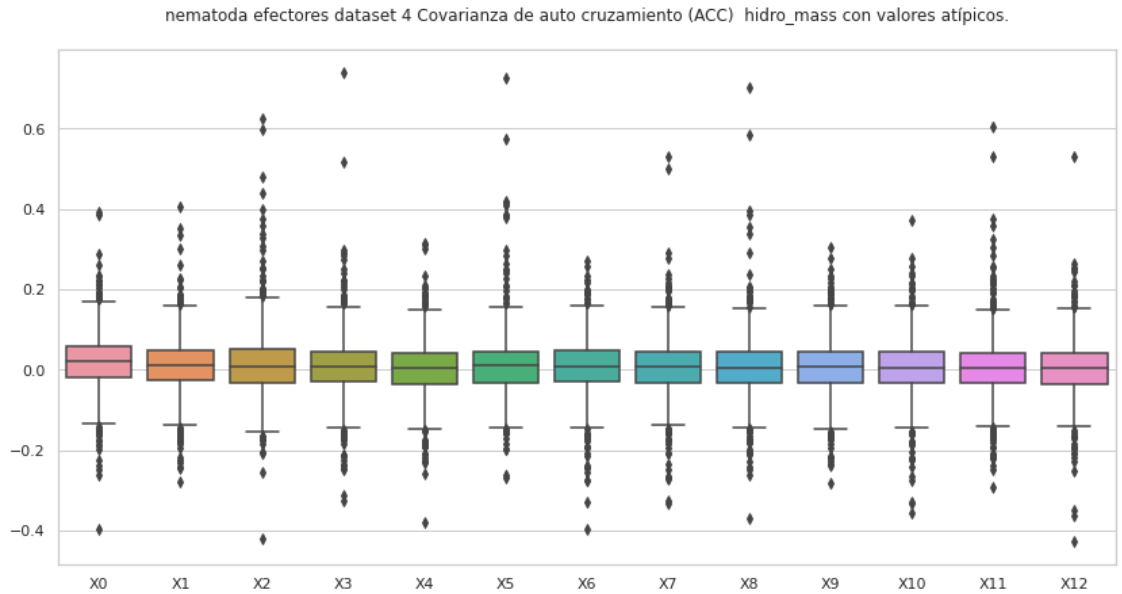
no_efectores
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores nematoda dataset 4, con valores atípicos.
Valores del documento csv.

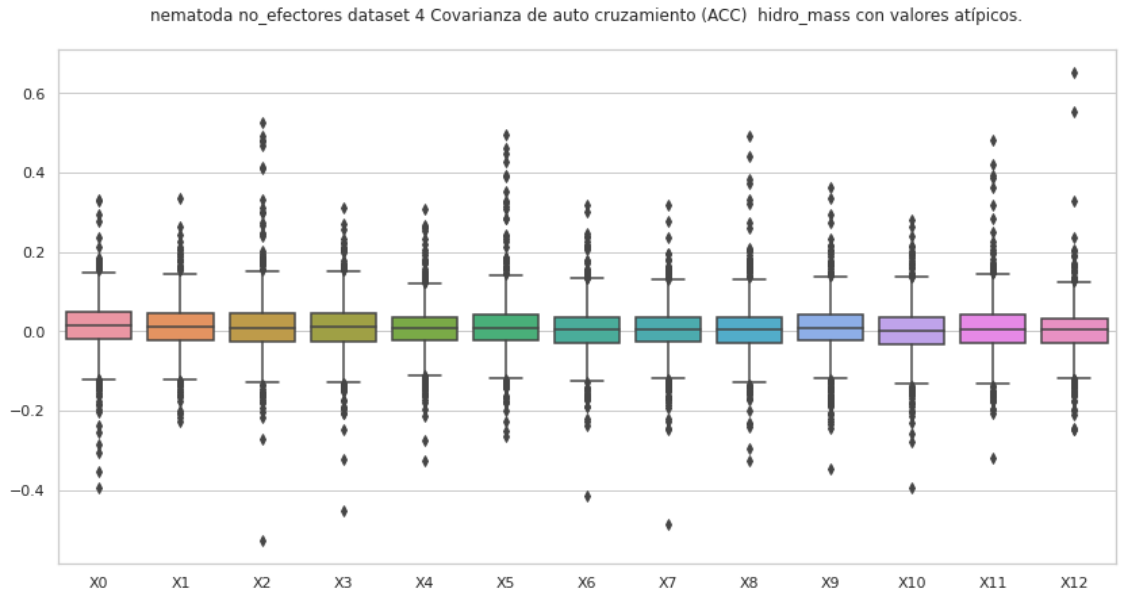
```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.072182  0.007824 -0.003928 -0.060508 -0.062722  0.179127  0.041200
1   -0.092526  0.103721 -0.023417  0.033337 -0.065660  0.121807 -0.000318
2   -0.004360  0.041435  0.048941  0.013346 -0.018648  0.026165 -0.009966
3    0.030595  0.027905 -0.046264 -0.001627  0.038008  0.090834  0.038055
4    0.014832  0.020576  0.031638  0.016525 -0.011857  0.029440 -0.002449
..        …         …         …         …         …         …
995 -0.044889 -0.008165 -0.106210  0.038989 -0.021179 -0.039688  0.023541
996  0.010167  0.090368  0.524797  0.077094  0.136784  0.446394  0.035848
997  0.021535 -0.019859  0.037854 -0.012890  0.041507  0.007881 -0.007748
998  0.006935  0.025471  0.016383  0.013073  0.102687 -0.015004  0.037112
999  0.072536 -0.013713 -0.133992 -0.193287 -0.131249  0.041754  0.056345

           X7        X8        X9       X10       X11       X12           X13
0    0.099911  0.075786  0.164674  0.013027  0.150372  0.200904  no_efectores
1    0.004995  0.002893  0.168939 -0.127742  0.035131 -0.046130  no_efectores
2    0.060863 -0.023079  0.045247  0.015929  0.024000 -0.008897  no_efectores
3    0.019530 -0.004538  0.010877  0.027380  0.043401  0.012660  no_efectores
4   -0.022869  0.009028  0.013956  0.024796 -0.006975 -0.014553  no_efectores
..        …         …         …         …         …         …
995 -0.221079  0.121084  0.060641  0.240350 -0.159562 -0.047046  no_efectores
996  0.059553  0.442438  0.024478  0.119017  0.421888  0.024818  no_efectores
997  0.006646  0.026086  0.009867  0.017550 -0.014288 -0.080522  no_efectores
998 -0.064020  0.023967  0.024910 -0.040851 -0.002838 -0.068431  no_efectores
999  0.158260  0.105881 -0.003633 -0.093601  0.021425 -0.049456  no_efectores

[1000 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores nematoda dataset 4, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.013567     0.011857     0.011553     0.009637     0.006395
std       0.067458     0.060482     0.077625     0.065511     0.061744
min      -0.392752    -0.228828    -0.527224    -0.453371    -0.325301
25%      -0.018816    -0.021640    -0.026414    -0.024553    -0.023035
50%       0.013390     0.012462     0.008799     0.010214     0.008150
75%       0.049917     0.046092     0.044744     0.046104     0.035377
```

```
max       0.332311     0.334293     0.524797     0.312002     0.306611

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.013389     0.004605     0.004544     0.005043     0.006411
std       0.073173     0.063019     0.061576     0.068713     0.066409
min      -0.265967    -0.416051    -0.487146    -0.324624    -0.345353
25%      -0.021884    -0.028903    -0.025771    -0.030581    -0.023924
50%       0.007879     0.004957     0.003934     0.003488     0.007230
75%       0.043670     0.035884     0.036647     0.035095     0.041014
max       0.496452     0.317175     0.319196     0.492474     0.362965

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.001607     0.009169     0.002914
std       0.064042     0.068506     0.064668
min      -0.394386    -0.320108    -0.248956
25%      -0.032280    -0.028327    -0.028367
50%       0.002638     0.005995     0.003398
75%       0.035807     0.041669     0.032966
max       0.279266     0.480739     0.651881
```





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 4,
sin valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.218244  0.165539  0.056992  0.004227  0.105088 -0.095810  0.028425
1   -0.004767  0.000896  0.027689 -0.034625 -0.071647  0.048405  0.116707
2    0.048562  0.017845  0.013823 -0.053341 -0.035738 -0.070553  0.024781
3   -0.027969  0.022099 -0.068296 -0.010149 -0.030623 -0.070653  0.019133
4    0.013399  0.025362  0.053115  0.044309  0.042439  0.042985  0.008109
..        …         …         …         …         …         …         …
995 -0.004752  0.037677 -0.081200 -0.065022  0.032421 -0.024451  0.098704
996  0.120346 -0.015528 -0.066367 -0.113780 -0.087586 -0.150389  0.024614
997 -0.077321 -0.006474  0.027995  0.008948  0.022489 -0.028513 -0.026591
998  0.014293 -0.022930 -0.001020  0.037801  0.009239 -0.034653 -0.008831
999  0.025367  0.018190  0.067727  0.059868 -0.007035 -0.036952 -0.008875

           X7        X8        X9       X10       X11       X12        X13
0   -0.133577  0.008114 -0.039043  0.079485  0.202914 -0.003074  efectores
1   -0.099995 -0.016323  0.000019 -0.058355 -0.110902 -0.056524  efectores
2   -0.051986  0.045369  0.055354 -0.012706  0.033834  0.020331  efectores
3    0.049939 -0.054186  0.014173 -0.107624  0.020429  0.009875  efectores
4    0.038602  0.003697  0.059413 -0.004282  0.009442  0.074815  efectores
..        …         …         …         …         …         …         …
995 -0.015361  0.020740 -0.054453 -0.100104 -0.021016 -0.021210  efectores
```

```
996  0.097103  0.090843 -0.016548 -0.074788 -0.059871 -0.041684  efectores
997  0.005472  0.056446 -0.009037  0.025306  0.047533 -0.047140  efectores
998  0.018961 -0.016942  0.044234 -0.013668  0.004967 -0.006670  efectores
999  0.022677  0.028692 -0.071891 -0.075295  0.052553  0.015800  efectores

[914 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro_mass efectores nematoda dataset 4, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  914.000000  914.000000  914.000000  914.000000  914.000000  914.000000
mean     0.020557    0.010243    0.008202    0.006983    0.003407    0.006000
std      0.060387    0.061507    0.063626    0.059389    0.061646    0.058201
min     -0.178705   -0.194327   -0.208736   -0.212915   -0.206556   -0.185879
25%     -0.016678   -0.026009   -0.029754   -0.028188   -0.031619   -0.030297
50%      0.021635    0.011466    0.008936    0.006303    0.005735    0.010587
75%      0.056221    0.045204    0.046694    0.041875    0.038400    0.040323
max      0.218244    0.222371    0.251378    0.224372    0.209225    0.205185

                X6          X7          X8          X9         X10         X11  \
count  914.000000  914.000000  914.000000  914.000000  914.000000  914.000000
mean     0.012047    0.005451    0.004442    0.005716    0.003285    0.002281
std      0.060053    0.058533    0.060928    0.061428    0.059777    0.061645
min     -0.195528   -0.209799   -0.206054   -0.194374   -0.206254   -0.220879
25%     -0.023808   -0.029806   -0.031887   -0.030553   -0.032428   -0.032556
50%      0.009991    0.006041    0.003721    0.006122    0.003318    0.003692
75%      0.048389    0.040810    0.039722    0.042136    0.039858    0.038093
max      0.197648    0.217350    0.235273    0.205696    0.195782    0.214701

               X12
count  914.000000
mean     0.004617
std      0.059467
min     -0.208437
25%     -0.031767
50%      0.004075
75%      0.040188
max      0.188592
```

no_efectores
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores nematoda dataset 4, sin valores atípicos.
Valores del documento csv.

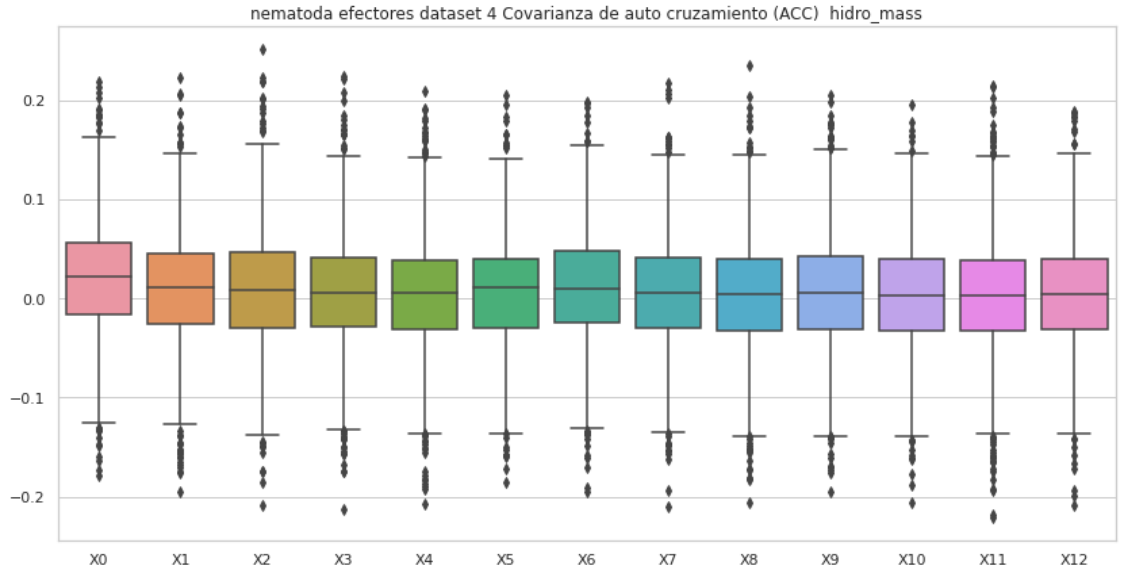
```
           X0        X1        X2        X3        X4        X5        X6  \
1   -0.092526  0.103721 -0.023417  0.033337 -0.065660  0.121807 -0.000318
2   -0.004360  0.041435  0.048941  0.013346 -0.018648  0.026165 -0.009966
3    0.030595  0.027905 -0.046264 -0.001627  0.038008  0.090834  0.038055
4    0.014832  0.020576  0.031638  0.016525 -0.011857  0.029440 -0.002449
5    0.045381 -0.014438 -0.010325 -0.006002  0.035778  0.017757 -0.015057
..        …         …         …         …         …         …
992  0.064955 -0.011127 -0.050326 -0.063582  0.018350  0.075964 -0.008018
993 -0.016269  0.036703  0.044865  0.002749  0.011786 -0.006508  0.015848
994  0.083983 -0.003926  0.022056  0.061683 -0.006963 -0.017464  0.012343
997  0.021535 -0.019859  0.037854 -0.012890  0.041507  0.007881 -0.007748
998  0.006935  0.025471  0.016383  0.013073  0.102687 -0.015004  0.037112

           X7        X8        X9       X10       X11       X12           X13
1    0.004995  0.002893  0.168939 -0.127742  0.035131 -0.046130  no_efectores
2    0.060863 -0.023079  0.045247  0.015929  0.024000 -0.008897  no_efectores
3    0.019530 -0.004538  0.010877  0.027380  0.043401  0.012660  no_efectores
4   -0.022869  0.009028  0.013956  0.024796 -0.006975 -0.014553  no_efectores
5    0.032237 -0.025727 -0.016514  0.022232 -0.003794 -0.003417  no_efectores
..        …         …         …         …         …         …
992  0.041599  0.117900  0.008253  0.085993  0.124121 -0.156959  no_efectores
993  0.001173  0.022116 -0.009640 -0.022176  0.001486 -0.043365  no_efectores
994  0.066238  0.014387  0.090311  0.041663  0.010109  0.032082  no_efectores
997  0.006646  0.026086  0.009867  0.017550 -0.014288 -0.080522  no_efectores
998 -0.064020  0.023967  0.024910 -0.040851 -0.002838 -0.068431  no_efectores

[906 rows x 14 columns]
```

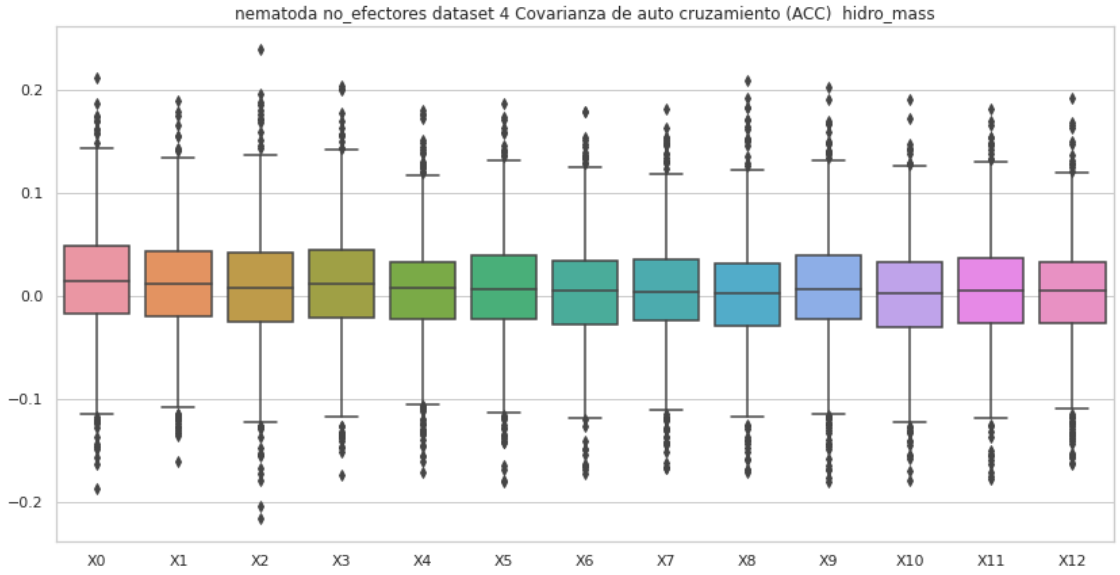
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores nematoda dataset 4, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  906.000000  906.000000  906.000000  906.000000  906.000000  906.000000
mean     0.014818    0.010904    0.007224    0.010368    0.005247    0.007343
std      0.054484    0.050912    0.055913    0.054917    0.050957    0.052557
min     -0.186583   -0.160266   -0.215710   -0.172995   -0.171576   -0.180749
25%     -0.017035   -0.019342   -0.024631   -0.020930   -0.022218   -0.021791
50%      0.014051    0.012045    0.008242    0.011282    0.008068    0.006470
75%      0.048073    0.043541    0.041498    0.044515    0.033363    0.039971
max      0.212087    0.189456    0.239181    0.204113    0.180429    0.186539

                X6          X7          X8          X9         X10         X11  \
count  906.000000  906.000000  906.000000  906.000000  906.000000  906.000000
mean     0.003804    0.004809    0.002192    0.006571    0.000751    0.005864
std      0.051998    0.052030    0.052602    0.054861    0.051509    0.052322
```

min	-0.171710	-0.167505	-0.171232	-0.180407	-0.178510	-0.177511
25%	-0.027548	-0.023358	-0.028372	-0.022210	-0.030231	-0.026597
50%	0.004677	0.003934	0.002851	0.007151	0.002502	0.004963
75%	0.033486	0.035446	0.032130	0.039431	0.033080	0.036360
max	0.178758	0.181231	0.209391	0.202536	0.190981	0.180825

	X12
count	906.000000
mean	0.003057
std	0.051594
min	-0.163065
25%	-0.025860
50%	0.004612
75%	0.032503
max	0.192531





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 4, con
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.218244  0.165539  0.056992  0.004227  0.105088 -0.095810  0.028425
1   -0.004767  0.000896  0.027689 -0.034625 -0.071647  0.048405  0.116707
2    0.048562  0.017845  0.013823 -0.053341 -0.035738 -0.070553  0.024781
3   -0.027969  0.022099 -0.068296 -0.010149 -0.030623 -0.070653  0.019133
4    0.013399  0.025362  0.053115  0.044309  0.042439  0.042985  0.008109
..        …         …         …         …         …         …         …
995 -0.004752  0.037677 -0.081200 -0.065022  0.032421 -0.024451  0.098704
996  0.120346 -0.015528 -0.066367 -0.113780 -0.087586 -0.150389  0.024614
997 -0.077321 -0.006474  0.027995  0.008948  0.022489 -0.028513 -0.026591
998  0.014293 -0.022930 -0.001020  0.037801  0.009239 -0.034653 -0.008831
999  0.025367  0.018190  0.067727  0.059868 -0.007035 -0.036952 -0.008875

           X7        X8        X9       X10       X11       X12         X13
0   -0.133577  0.008114 -0.039043  0.079485  0.202914 -0.003074  efectores
1   -0.099995 -0.016323  0.000019 -0.058355 -0.110902 -0.056524  efectores
2   -0.051986  0.045369  0.055354 -0.012706  0.033834  0.020331  efectores
3    0.049939 -0.054186  0.014173 -0.107624  0.020429  0.009875  efectores
4    0.038602  0.003697  0.059413 -0.004282  0.009442  0.074815  efectores
..        …         …         …         …         …         …         …
995 -0.015361  0.020740 -0.054453 -0.100104 -0.021016 -0.021210  efectores
996  0.097103  0.090843 -0.016548 -0.074788 -0.059871 -0.041684  efectores
997  0.005472  0.056446 -0.009037  0.025306  0.047533 -0.047140  efectores
998  0.018961 -0.016942  0.044234 -0.013668  0.004967 -0.006670  efectores
999  0.022677  0.028692 -0.071891 -0.075295  0.052553  0.015800  efectores

[1000 rows x 14 columns]


Covarianza de auto cruzamiento (ACC)  mass efectores nematoda dataset 4, con
valores atípicos.
Estadísticas.

                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.019717     0.010988     0.012530     0.006967     0.002797
```

```
std       0.070778     0.071387     0.082123     0.074978     0.070583
min      -0.397525    -0.280178    -0.421457    -0.325324    -0.380042
25%      -0.019637    -0.027366    -0.033221    -0.031188    -0.034945
50%       0.020963     0.011466     0.009165     0.006303     0.005283
75%       0.058358     0.048684     0.052210     0.043394     0.041829
max       0.393351     0.405397     0.624939     0.738565     0.315124

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.010352     0.008773     0.006263     0.006164     0.005261
std       0.077535     0.071488     0.073118     0.077218     0.070304
min      -0.269688    -0.395533    -0.334495    -0.370642    -0.281060
25%      -0.031376    -0.028510    -0.032034    -0.032471    -0.033774
50%       0.011264     0.008941     0.007122     0.004097     0.006086
75%       0.044203     0.049040     0.043567     0.043403     0.043942
max       0.723713     0.268890     0.528696     0.701850     0.303544

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.004195     0.005732     0.003884
std       0.072098     0.076366     0.070976
min      -0.355985    -0.291456    -0.427231
25%      -0.033910    -0.032866    -0.034704
50%       0.004146     0.004703     0.003367
75%       0.043379     0.040880     0.041065
max       0.371363     0.602932     0.530164
```

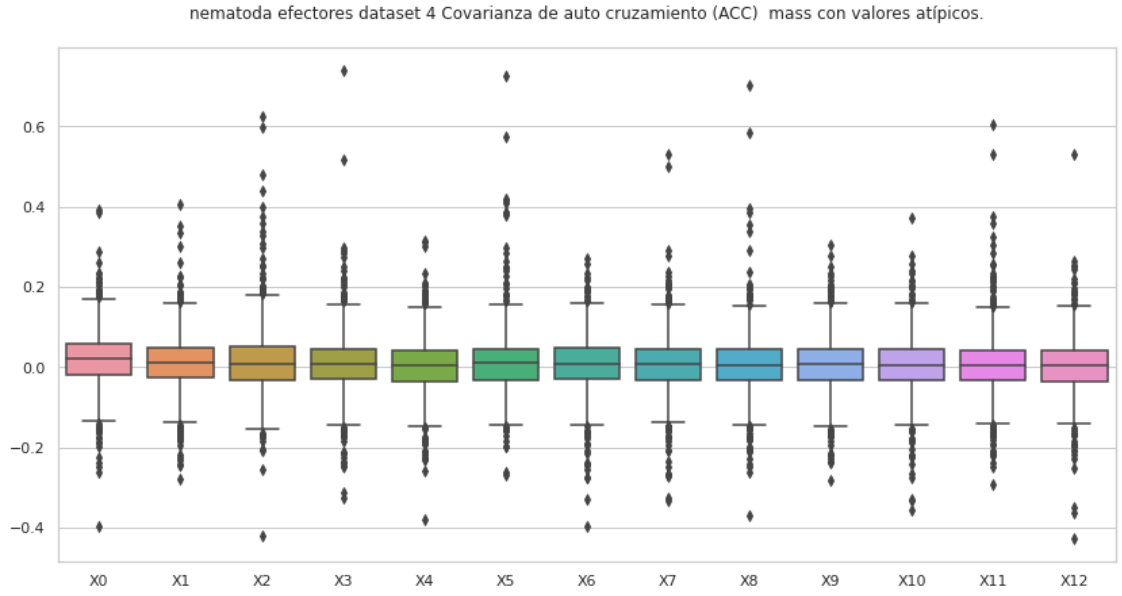
no_efectores
Covarianza de auto cruzamiento (ACC) mass no_efectores nematoda dataset 4, con valores atípicos.
Valores del documento csv.

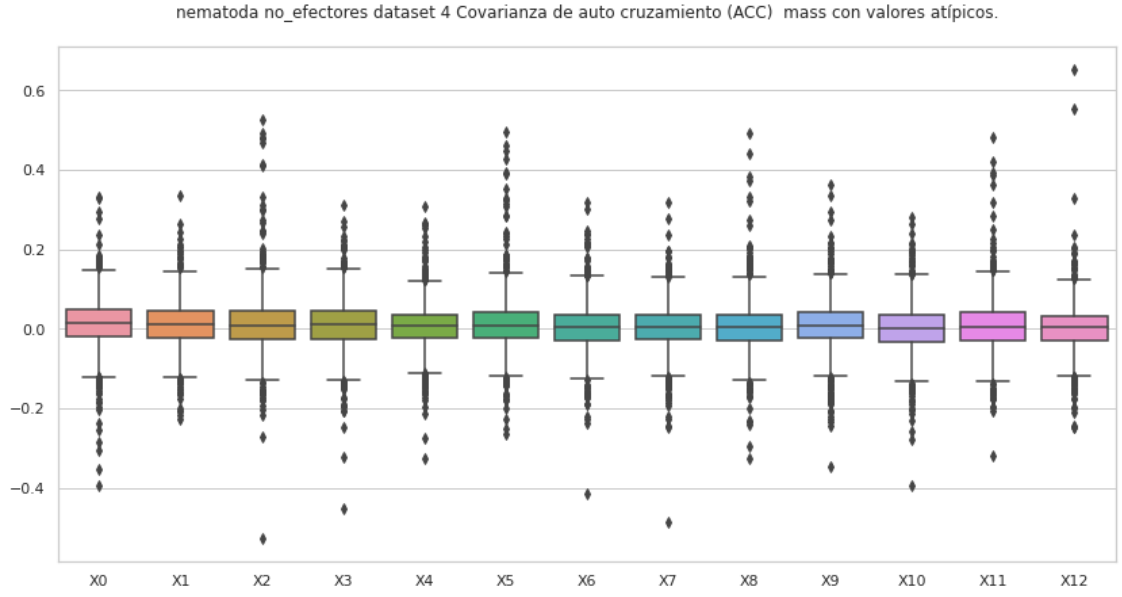
```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.072182  0.007824 -0.003928 -0.060508 -0.062722  0.179127  0.041200
1   -0.092526  0.103721 -0.023417  0.033337 -0.065660  0.121807 -0.000318
2   -0.004360  0.041435  0.048941  0.013346 -0.018648  0.026165 -0.009966
3    0.030595  0.027905 -0.046264 -0.001627  0.038008  0.090834  0.038055
4    0.014832  0.020576  0.031638  0.016525 -0.011857  0.029440 -0.002449
..        …         …         …         …         …         …         …
995 -0.044889 -0.008165 -0.106210  0.038989 -0.021179 -0.039688  0.023541
996  0.010167  0.090368  0.524797  0.077094  0.136784  0.446394  0.035848
997  0.021535 -0.019859  0.037854 -0.012890  0.041507  0.007881 -0.007748
998  0.006935  0.025471  0.016383  0.013073  0.102687 -0.015004  0.037112
999  0.072536 -0.013713 -0.133992 -0.193287 -0.131249  0.041754  0.056345

           X7        X8        X9       X10       X11       X12           X13
```

```
0    0.099911  0.075786  0.164674  0.013027  0.150372  0.200904  no_efectores
1    0.004995  0.002893  0.168939 -0.127742  0.035131 -0.046130  no_efectores
2    0.060863 -0.023079  0.045247  0.015929  0.024000 -0.008897  no_efectores
3    0.019530 -0.004538  0.010877  0.027380  0.043401  0.012660  no_efectores
4   -0.022869  0.009028  0.013956  0.024796 -0.006975 -0.014553  no_efectores
..        …         …         …         …         …         …           …
995 -0.221079  0.121084  0.060641  0.240350 -0.159562 -0.047046  no_efectores
996  0.059553  0.442438  0.024478  0.119017  0.421888  0.024818  no_efectores
997  0.006646  0.026086  0.009867  0.017550 -0.014288 -0.080522  no_efectores
998 -0.064020  0.023967  0.024910 -0.040851 -0.002838 -0.068431  no_efectores
999  0.158260  0.105881 -0.003633 -0.093601  0.021425 -0.049456  no_efectores

[1000 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) mass no_efectores nematoda dataset 4, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.013567     0.011857     0.011553     0.009637     0.006395
std       0.067458     0.060482     0.077625     0.065511     0.061744
min      -0.392752    -0.228828    -0.527224    -0.453371    -0.325301
25%      -0.018816    -0.021640    -0.026414    -0.024553    -0.023035
50%       0.013390     0.012462     0.008799     0.010214     0.008150
75%       0.049917     0.046092     0.044744     0.046104     0.035377
max       0.332311     0.334293     0.524797     0.312002     0.306611

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.013389     0.004605     0.004544     0.005043     0.006411
std       0.073173     0.063019     0.061576     0.068713     0.066409
min      -0.265967    -0.416051    -0.487146    -0.324624    -0.345353
25%      -0.021884    -0.028903    -0.025771    -0.030581    -0.023924
50%       0.007879     0.004957     0.003934     0.003488     0.007230
75%       0.043670     0.035884     0.036647     0.035095     0.041014
max       0.496452     0.317175     0.319196     0.492474     0.362965

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.001607     0.009169     0.002914
std       0.064042     0.068506     0.064668
min      -0.394386    -0.320108    -0.248956
25%      -0.032280    -0.028327    -0.028367
50%       0.002638     0.005995     0.003398
75%       0.035807     0.041669     0.032966
max       0.279266     0.480739     0.651881
```


nematoda efectores dataset 4 Covarianza de auto cruzamiento (ACC) mass con valores atípicos.
0.6
0.4
0.2
0.0
−0.2
−0.4
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12


nematoda no_efectores dataset 4 Covarianza de auto cruzamiento (ACC) mass con valores atípicos.
0.6
0.4
0.2
0.0
−0.2
−0.4
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12

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 4, sin
valores atípicos.
Valores del documento csv.
```

	X0	X1	X2	X3	X4	X5	X6	\
0	0.218244	0.165539	0.056992	0.004227	0.105088	-0.095810	0.028425	
1	-0.004767	0.000896	0.027689	-0.034625	-0.071647	0.048405	0.116707	
2	0.048562	0.017845	0.013823	-0.053341	-0.035738	-0.070553	0.024781	
3	-0.027969	0.022099	-0.068296	-0.010149	-0.030623	-0.070653	0.019133	
4	0.013399	0.025362	0.053115	0.044309	0.042439	0.042985	0.008109	
..	…	…	…	…	…	…	…	
995	-0.004752	0.037677	-0.081200	-0.065022	0.032421	-0.024451	0.098704	
996	0.120346	-0.015528	-0.066367	-0.113780	-0.087586	-0.150389	0.024614	
997	-0.077321	-0.006474	0.027995	0.008948	0.022489	-0.028513	-0.026591	
998	0.014293	-0.022930	-0.001020	0.037801	0.009239	-0.034653	-0.008831	
999	0.025367	0.018190	0.067727	0.059868	-0.007035	-0.036952	-0.008875	

	X7	X8	X9	X10	X11	X12	X13
0	-0.133577	0.008114	-0.039043	0.079485	0.202914	-0.003074	efectores
1	-0.099995	-0.016323	0.000019	-0.058355	-0.110902	-0.056524	efectores
2	-0.051986	0.045369	0.055354	-0.012706	0.033834	0.020331	efectores
3	0.049939	-0.054186	0.014173	-0.107624	0.020429	0.009875	efectores
4	0.038602	0.003697	0.059413	-0.004282	0.009442	0.074815	efectores
..	…	…	…	…	…	…	…
995	-0.015361	0.020740	-0.054453	-0.100104	-0.021016	-0.021210	efectores
996	0.097103	0.090843	-0.016548	-0.074788	-0.059871	-0.041684	efectores
997	0.005472	0.056446	-0.009037	0.025306	0.047533	-0.047140	efectores
998	0.018961	-0.016942	0.044234	-0.013668	0.004967	-0.006670	efectores
999	0.022677	0.028692	-0.071891	-0.075295	0.052553	0.015800	efectores

```
[914 rows x 14 columns]
```

```
Covarianza de auto cruzamiento (ACC)  mass efectores nematoda dataset 4, sin
valores atípicos.
Estadísticas.
```

	X0	X1	X2	X3	X4	X5	\
count	914.000000	914.000000	914.000000	914.000000	914.000000	914.000000	
mean	0.020557	0.010243	0.008202	0.006983	0.003407	0.006000	
std	0.060387	0.061507	0.063626	0.059389	0.061646	0.058201	
min	-0.178705	-0.194327	-0.208736	-0.212915	-0.206556	-0.185879	
25%	-0.016678	-0.026009	-0.029754	-0.028188	-0.031619	-0.030297	
50%	0.021635	0.011466	0.008936	0.006303	0.005735	0.010587	
75%	0.056221	0.045204	0.046694	0.041875	0.038400	0.040323	

```
max      0.218244    0.222371    0.251378    0.224372    0.209225    0.205185

               X6          X7          X8          X9         X10         X11  \
count  914.000000  914.000000  914.000000  914.000000  914.000000  914.000000
mean     0.012047    0.005451    0.004442    0.005716    0.003285    0.002281
std      0.060053    0.058533    0.060928    0.061428    0.059777    0.061645
min     -0.195528   -0.209799   -0.206054   -0.194374   -0.206254   -0.220879
25%     -0.023808   -0.029806   -0.031887   -0.030553   -0.032428   -0.032556
50%      0.009991    0.006041    0.003721    0.006122    0.003318    0.003692
75%      0.048389    0.040810    0.039722    0.042136    0.039858    0.038093
max      0.197648    0.217350    0.235273    0.205696    0.195782    0.214701

              X12
count  914.000000
mean     0.004617
std      0.059467
min     -0.208437
25%     -0.031767
50%      0.004075
75%      0.040188
max      0.188592
```

Covarianza de auto cruzamiento (ACC) mass no_efectores nematoda dataset 4, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
1   -0.092526  0.103721 -0.023417  0.033337 -0.065660  0.121807 -0.000318
2   -0.004360  0.041435  0.048941  0.013346 -0.018648  0.026165 -0.009966
3    0.030595  0.027905 -0.046264 -0.001627  0.038008  0.090834  0.038055
4    0.014832  0.020576  0.031638  0.016525 -0.011857  0.029440 -0.002449
5    0.045381 -0.014438 -0.010325 -0.006002  0.035778  0.017757 -0.015057
..        …         …         …         …         …         …
992  0.064955 -0.011127 -0.050326 -0.063582  0.018350  0.075964 -0.008018
993 -0.016269  0.036703  0.044865  0.002749  0.011786 -0.006508  0.015848
994  0.083983 -0.003926  0.022056  0.061683 -0.006963 -0.017464  0.012343
997  0.021535 -0.019859  0.037854 -0.012890  0.041507  0.007881 -0.007748
998  0.006935  0.025471  0.016383  0.013073  0.102687 -0.015004  0.037112

           X7        X8        X9       X10       X11       X12           X13
1    0.004995  0.002893  0.168939 -0.127742  0.035131 -0.046130  no_efectores
2    0.060863 -0.023079  0.045247  0.015929  0.024000 -0.008897  no_efectores
3    0.019530 -0.004538  0.010877  0.027380  0.043401  0.012660  no_efectores
4   -0.022869  0.009028  0.013956  0.024796 -0.006975 -0.014553  no_efectores
5    0.032237 -0.025727 -0.016514  0.022232 -0.003794 -0.003417  no_efectores
..        …         …         …         …         …         …           …
```

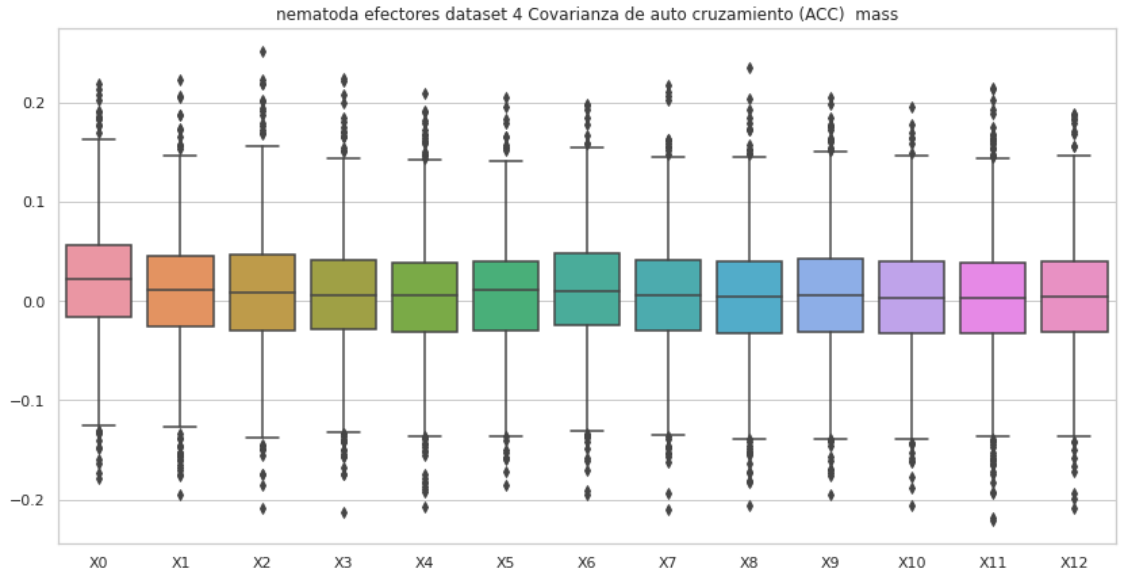
```
992  0.041599  0.117900  0.008253  0.085993  0.124121 -0.156959  no_efectores
993  0.001173  0.022116 -0.009640 -0.022176  0.001486 -0.043365  no_efectores
994  0.066238  0.014387  0.090311  0.041663  0.010109  0.032082  no_efectores
997  0.006646  0.026086  0.009867  0.017550 -0.014288 -0.080522  no_efectores
998 -0.064020  0.023967  0.024910 -0.040851 -0.002838 -0.068431  no_efectores

[906 rows x 14 columns]
```

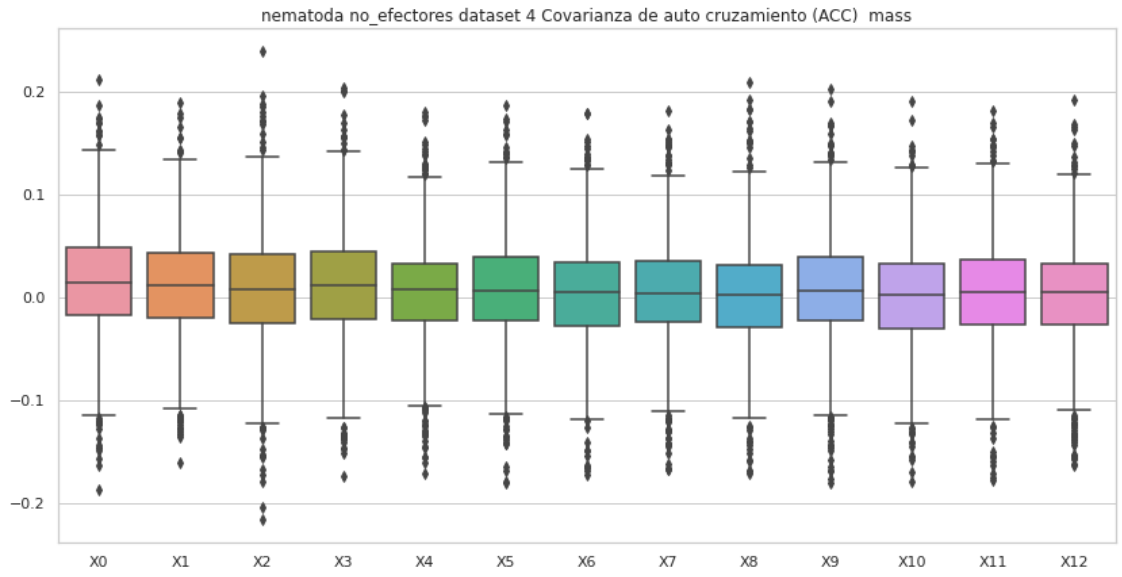
Covarianza de auto cruzamiento (ACC) mass no_efectores nematoda dataset 4, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	906.000000	906.000000	906.000000	906.000000	906.000000	906.000000	
mean	0.014818	0.010904	0.007224	0.010368	0.005247	0.007343	
std	0.054484	0.050912	0.055913	0.054917	0.050957	0.052557	
min	-0.186583	-0.160266	-0.215710	-0.172995	-0.171576	-0.180749	
25%	-0.017035	-0.019342	-0.024631	-0.020930	-0.022218	-0.021791	
50%	0.014051	0.012045	0.008242	0.011282	0.008068	0.006470	
75%	0.048073	0.043541	0.041498	0.044515	0.033363	0.039971	
max	0.212087	0.189456	0.239181	0.204113	0.180429	0.186539	

	X6	X7	X8	X9	X10	X11	\
count	906.000000	906.000000	906.000000	906.000000	906.000000	906.000000	
mean	0.003804	0.004809	0.002192	0.006571	0.000751	0.005864	
std	0.051998	0.052030	0.052602	0.054861	0.051509	0.052322	
min	-0.171710	-0.167505	-0.171232	-0.180407	-0.178510	-0.177511	
25%	-0.027548	-0.023358	-0.028372	-0.022210	-0.030231	-0.026597	
50%	0.004677	0.003934	0.002851	0.007151	0.002502	0.004963	
75%	0.033486	0.035446	0.032130	0.039431	0.033080	0.036360	
max	0.178758	0.181231	0.209391	0.202536	0.190981	0.180825	

	X12
count	906.000000
mean	0.003057
std	0.051594
min	-0.163065
25%	-0.025860
50%	0.004612
75%	0.032503
max	0.192531





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 4, con
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.050496 -0.221279  0.004202  0.120614 -0.006600 -0.013983 -0.037731
1   -0.046629 -0.093572 -0.032753  0.030833  0.049962  0.017556  0.128570
2   -0.118273 -0.131529  0.056141 -0.016719 -0.055235 -0.073870 -0.017822
3    0.057330 -0.231126 -0.072878 -0.043912  0.006700 -0.090982  0.001974
4   -0.015430  0.030615 -0.011865 -0.102593  0.080050 -0.050338  0.035169
..        …       …       …       …       …       …       …
995  0.002941  0.027240 -0.154167 -0.064263 -0.046394 -0.152431 -0.092003
996  0.027931 -0.128791 -0.101648  0.071394  0.043592 -0.112924  0.056260
997 -0.082439  0.053807  0.034348 -0.049024 -0.099415  0.034551  0.102971
998 -0.033174 -0.034189 -0.046915  0.007485 -0.044830 -0.060160 -0.033140
999  0.022854  0.048701  0.035738  0.025890  0.048908  0.028899  0.016300

           X7        X8        X9       X10       X11       X12        X13
0    0.176974  0.110349  0.056140 -0.153101  0.045121  0.045728  efectores
1   -0.071268  0.003278  0.043296 -0.000128 -0.076226  0.073026  efectores
2    0.055977  0.021319 -0.058667 -0.056035  0.033226 -0.108184  efectores
```

```
3    0.077189 -0.100286 -0.050791  0.041818 -0.122318  0.059289  efectores
4    0.062997 -0.042103  0.053920 -0.111131 -0.051739 -0.012883  efectores
..        …        …        …        …        …        …
995  0.133659 -0.055528 -0.024118 -0.075211  0.012242  0.050104  efectores
996 -0.067366 -0.144678 -0.036203 -0.003249  0.058139  0.016567  efectores
997 -0.089463 -0.020014  0.097472  0.075302  0.003500  0.052859  efectores
998 -0.026476 -0.001730 -0.000936  0.007447 -0.057440  0.006331  efectores
999  0.015180  0.030246  0.027375 -0.031563  0.061028  0.045875  efectores

[1000 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro efectores nematoda dataset 4, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.018521    -0.018271     0.020443     0.023831    -0.004553
std       0.083595     0.090765     0.087978     0.089083     0.083847
min      -0.386905    -0.368721    -0.403834    -0.394315    -0.325208
25%      -0.031388    -0.068979    -0.029402    -0.026912    -0.056870
50%       0.013688    -0.021599     0.021042     0.024144    -0.006752
75%       0.066466     0.035448     0.069051     0.076001     0.043736
max       0.423737     0.473558     0.498484     0.437276     0.360679

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean     -0.003965     0.017933     0.006969     0.005906     0.010584
std       0.086152     0.089242     0.080411     0.080988     0.081520
min      -0.402784    -0.413980    -0.310769    -0.273575    -0.359847
25%      -0.053379    -0.033354    -0.039061    -0.041199    -0.032138
50%      -0.002964     0.019288     0.009124     0.004106     0.010918
75%       0.048993     0.067561     0.054404     0.049117     0.055126
max       0.294223     0.390205     0.352952     0.452377     0.358453

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.014989     0.008824     0.003842
std       0.077446     0.081664     0.081471
min      -0.302821    -0.433188    -0.274646
25%      -0.025940    -0.035453    -0.042543
50%       0.013645     0.007614     0.005510
75%       0.057739     0.051528     0.047920
max       0.520309     0.498759     0.314528
```

no_efectores
Covarianza de auto cruzamiento (ACC) hidro no_efectores nematoda dataset 4, con valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	-0.030916	-0.037230	-0.007332	0.013296	0.005730	-0.173982	-0.020127	
1	0.116989	0.025699	0.076570	-0.029711	0.047103	0.192560	0.012237	
2	-0.052938	-0.020380	-0.014068	0.033740	-0.028991	-0.073260	0.192562	
3	0.092846	0.052778	0.124112	0.070985	0.022072	-0.065795	0.039673	
4	0.033529	-0.057213	0.040960	0.016331	-0.028824	-0.040351	0.004154	
..	…	…	…	…	…	…	…	
995	-0.109965	0.090432	0.141821	-0.143231	0.060419	-0.161167	0.027182	
996	-0.047438	0.086121	0.052284	0.010117	-0.025143	0.058306	-0.078490	
997	-0.002066	-0.115666	-0.021651	-0.002449	0.029193	-0.039616	-0.047444	
998	-0.029726	-0.008290	-0.040065	-0.025307	-0.030000	-0.093725	-0.013219	
999	0.049208	-0.169176	0.138807	0.026913	-0.112341	-0.198534	-0.301353	

	X7	X8	X9	X10	X11	X12	X13
0	0.074306	-0.027165	0.015167	0.249750	0.022024	0.099860	no_efectores
1	0.048536	0.000709	0.046154	0.073352	0.022357	-0.161272	no_efectores
2	0.080135	0.000650	-0.133064	0.085080	-0.065503	-0.022113	no_efectores
3	0.110099	0.007205	-0.015793	-0.028560	0.119257	0.034138	no_efectores
4	0.048623	-0.062414	0.064929	-0.021611	-0.023211	0.030860	no_efectores
..	…	…	…	…	…	…	…
995	-0.033276	-0.018583	0.022373	-0.128993	0.183292	-0.157170	no_efectores
996	0.049341	-0.014154	0.018777	0.028977	0.057772	-0.054217	no_efectores
997	0.005094	0.076808	-0.009374	-0.070071	-0.026506	-0.003653	no_efectores
998	-0.041868	0.019970	-0.075581	0.042456	-0.004976	0.061892	no_efectores
999	0.106023	0.236710	-0.117490	-0.060822	0.146309	0.261099	no_efectores

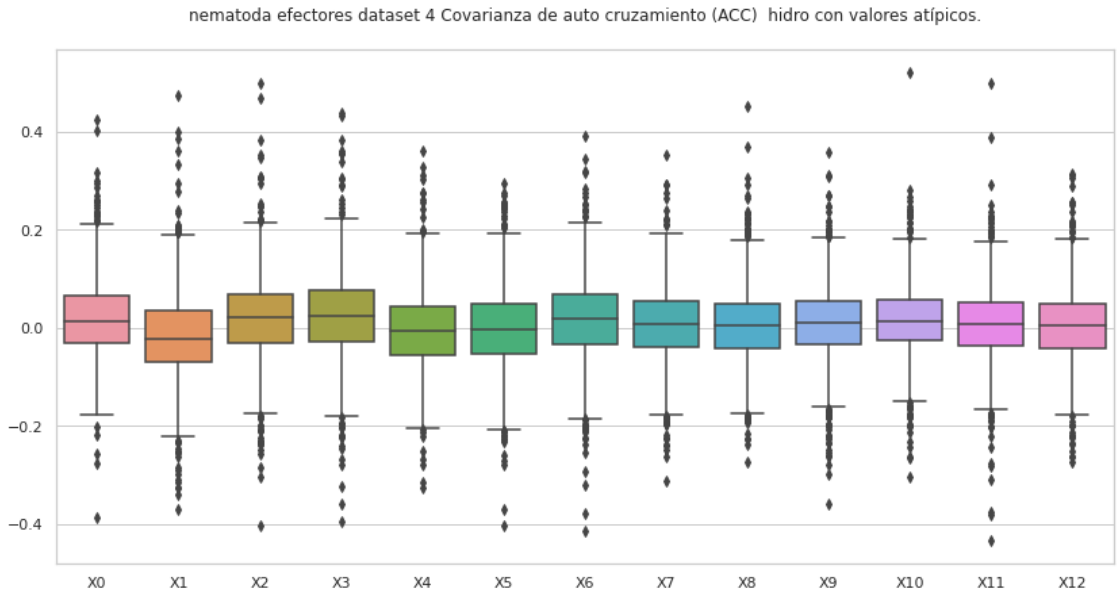
[1000 rows x 14 columns]

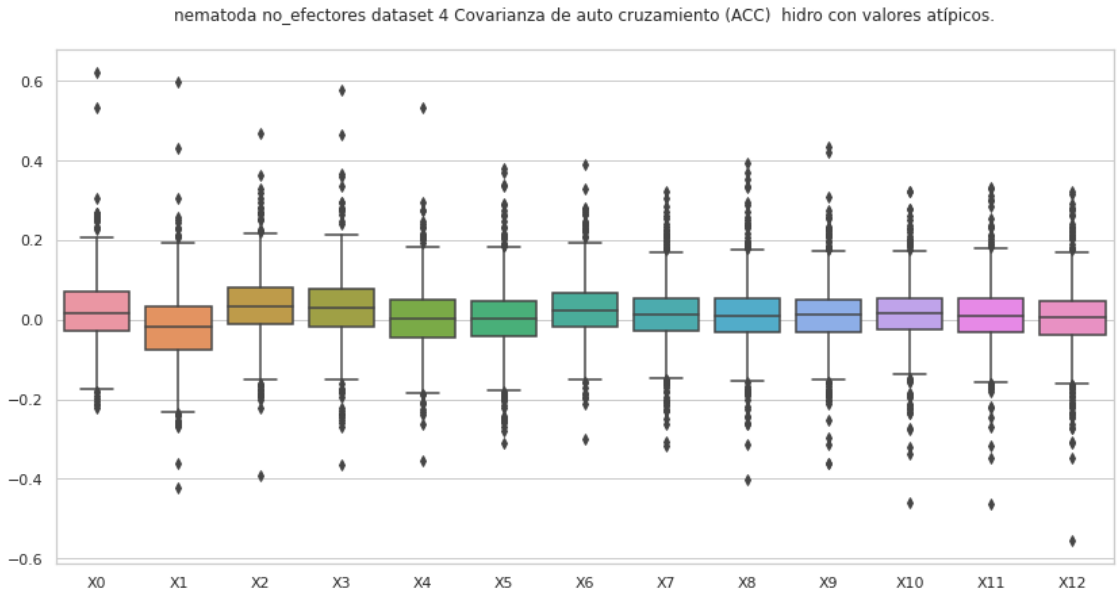
Covarianza de auto cruzamiento (ACC) hidro no_efectores nematoda dataset 4, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.022521	-0.017897	0.034390	0.029885	0.003052	
std	0.080843	0.088072	0.080178	0.083464	0.078510	
min	-0.221288	-0.421394	-0.390157	-0.365325	-0.353147	
25%	-0.029460	-0.074024	-0.012195	-0.018242	-0.044754	
50%	0.017349	-0.016947	0.032473	0.030515	0.001820	
75%	0.069477	0.032685	0.080839	0.077077	0.049561	
max	0.621014	0.595930	0.466973	0.578669	0.534469	

	X5	X6	X7	X8	X9	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.004313	0.025909	0.011681	0.011762	0.010947	
std	0.082234	0.074617	0.074902	0.080349	0.078258	
min	-0.310935	-0.301353	-0.317803	-0.400913	-0.361086	
25%	-0.042787	-0.019433	-0.027658	-0.030073	-0.032176	
50%	0.003696	0.023406	0.011981	0.009127	0.013054	
75%	0.047703	0.068104	0.052052	0.053453	0.049876	
max	0.380211	0.390687	0.322978	0.393013	0.435383	

	X10	X11	X12
count	1000.000000	1000.000000	1000.000000
mean	0.014115	0.011256	0.004013
std	0.076496	0.077997	0.083138
min	-0.460166	-0.462837	-0.554841
25%	-0.025250	-0.031138	-0.037069
50%	0.015665	0.009677	0.006716
75%	0.054757	0.054102	0.047204
max	0.321802	0.333590	0.320650





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 4, sin valores
atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.050496 -0.221279  0.004202  0.120614 -0.006600 -0.013983 -0.037731
1   -0.046629 -0.093572 -0.032753  0.030833  0.049962  0.017556  0.128570
2   -0.118273 -0.131529  0.056141 -0.016719 -0.055235 -0.073870 -0.017822
3    0.057330 -0.231126 -0.072878 -0.043912  0.006700 -0.090982  0.001974
4   -0.015430  0.030615 -0.011865 -0.102593  0.080050 -0.050338  0.035169
..        …         …         …         …         …         …         …
995  0.002941  0.027240 -0.154167 -0.064263 -0.046394 -0.152431 -0.092003
996  0.027931 -0.128791 -0.101648  0.071394  0.043592 -0.112924  0.056260
997 -0.082439  0.053807  0.034348 -0.049024 -0.099415  0.034551  0.102971
998 -0.033174 -0.034189 -0.046915  0.007485 -0.044830 -0.060160 -0.033140
999  0.022854  0.048701  0.035738  0.025890  0.048908  0.028899  0.016300

           X7        X8        X9       X10       X11       X12        X13
0    0.176974  0.110349  0.056140 -0.153101  0.045121  0.045728  efectores
1   -0.071268  0.003278  0.043296 -0.000128 -0.076226  0.073026  efectores
2    0.055977  0.021319 -0.058667 -0.056035  0.033226 -0.108184  efectores
3    0.077189 -0.100286 -0.050791  0.041818 -0.122318  0.059289  efectores
4    0.062997 -0.042103  0.053920 -0.111131 -0.051739 -0.012883  efectores
..        …         …         …         …         …         …         …
995  0.133659 -0.055528 -0.024118 -0.075211  0.012242  0.050104  efectores
```

```
996 -0.067366 -0.144678 -0.036203 -0.003249  0.058139  0.016567  efectores
997 -0.089463 -0.020014  0.097472  0.075302  0.003500  0.052859  efectores
998 -0.026476 -0.001730 -0.000936  0.007447 -0.057440  0.006331  efectores
999  0.015180  0.030246  0.027375 -0.031563  0.061028  0.045875  efectores

[912 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) efectores nematoda dataset 4, sin valores atípicos.
Estadísticas.

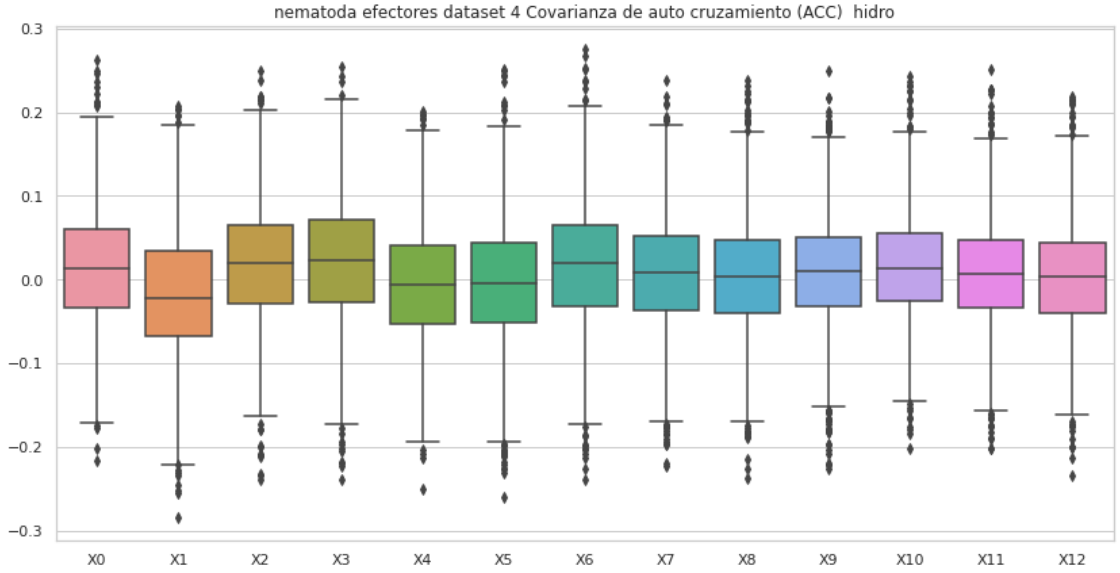
```
                X0          X1          X2          X3          X4          X5  \
count  912.000000  912.000000  912.000000  912.000000  912.000000  912.000000
mean     0.013966   -0.019427    0.018561    0.019749   -0.006044   -0.005456
std      0.073713    0.077228    0.074776    0.075800    0.074782    0.077835
min     -0.216807   -0.284239   -0.238464   -0.239140   -0.250416   -0.260024
25%     -0.033043   -0.068161   -0.028014   -0.026912   -0.053525   -0.051364
50%      0.012529   -0.021540    0.020323    0.022415   -0.006655   -0.004006
75%      0.059496    0.033919    0.065147    0.071260    0.041302    0.044157
max      0.261502    0.207971    0.248692    0.254066    0.200674    0.250519

                X6          X7          X8          X9         X10         X11  \
count  912.000000  912.000000  912.000000  912.000000  912.000000  912.000000
mean     0.018889    0.006149    0.003354    0.008691    0.014271    0.007872
std      0.079263    0.071674    0.073205    0.070222    0.068340    0.069695
min     -0.238768   -0.223047   -0.236565   -0.226621   -0.201199   -0.202103
25%     -0.031185   -0.037237   -0.040418   -0.031417   -0.025597   -0.033813
50%      0.019474    0.008569    0.003633    0.009963    0.013774    0.007359
75%      0.065156    0.051607    0.046635    0.050718    0.055543    0.047749
max      0.275109    0.238256    0.237284    0.248605    0.243421    0.251616

               X12
count  912.000000
mean     0.002429
std      0.072534
min     -0.234493
25%     -0.039843
50%      0.004389
75%      0.044723
max      0.218340
```

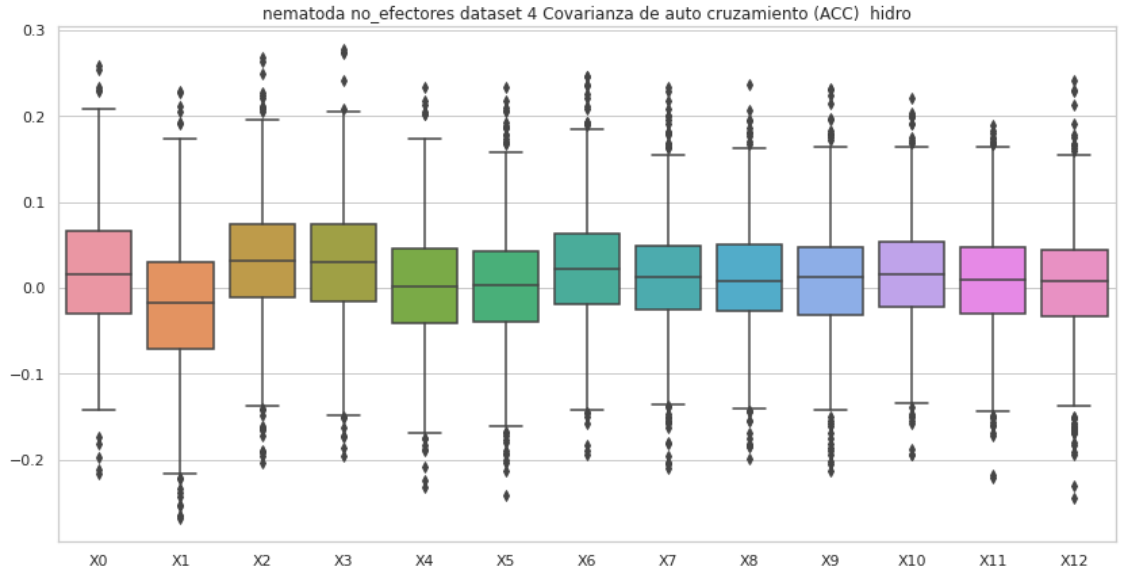
no_efectores
Covarianza de auto cruzamiento (ACC) no_efectores nematoda dataset 4, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
1    0.116989  0.025699  0.076570 -0.029711  0.047103  0.192560  0.012237
2   -0.052938 -0.020380 -0.014068  0.033740 -0.028991 -0.073260  0.192562
3    0.092846  0.052778  0.124112  0.070985  0.022072 -0.065795  0.039673
4    0.033529 -0.057213  0.040960  0.016331 -0.028824 -0.040351  0.004154
5    0.026444 -0.035866  0.084273  0.031094  0.007756 -0.028383  0.042939
..        …         …         …         …         …         …         …
994 -0.000380 -0.144278  0.018254  0.005954  0.034856  0.049998 -0.006647
995 -0.109965  0.090432  0.141821 -0.143231  0.060419 -0.161167  0.027182
996 -0.047438  0.086121  0.052284  0.010117 -0.025143  0.058306 -0.078490
997 -0.002066 -0.115666 -0.021651 -0.002449  0.029193 -0.039616 -0.047444
998 -0.029726 -0.008290 -0.040065 -0.025307 -0.030000 -0.093725 -0.013219

           X7        X8        X9       X10       X11       X12           X13
1    0.048536  0.000709  0.046154  0.073352  0.022357 -0.161272  no_efectores
2    0.080135  0.000650 -0.133064  0.085080 -0.065503 -0.022113  no_efectores
3    0.110099  0.007205 -0.015793 -0.028560  0.119257  0.034138  no_efectores
4    0.048623 -0.062414  0.064929 -0.021611 -0.023211  0.030860  no_efectores
5    0.026644 -0.036952  0.000472  0.043014 -0.005246  0.035248  no_efectores
..        …         …         …         …         …         …           …
994  0.005348  0.080717  0.051253 -0.127525  0.058099  0.016789  no_efectores
995 -0.033276 -0.018583  0.022373 -0.128993  0.183292 -0.157170  no_efectores
996  0.049341 -0.014154  0.018777  0.028977  0.057772 -0.054217  no_efectores
997  0.005094  0.076808 -0.009374 -0.070071 -0.026506 -0.003653  no_efectores
998 -0.041868  0.019970 -0.075581  0.042456 -0.004976  0.061892  no_efectores

[903 rows x 14 columns]


Covarianza de auto cruzamiento (ACC)  no_efectores nematoda dataset 4, sin
valores atípicos.
Estadísticas.

                X0          X1          X2          X3          X4          X5  \
count  903.000000  903.000000  903.000000  903.000000  903.000000  903.000000
mean     0.019316   -0.019328    0.031165    0.029083    0.002082    0.002529
std      0.071259    0.076659    0.069287    0.069830    0.068334    0.068041
min     -0.216141   -0.268399   -0.202930   -0.195017   -0.232096   -0.241307
25%     -0.029421   -0.070614   -0.011581   -0.015517   -0.040464   -0.039451
50%      0.016162   -0.017479    0.031687    0.030372    0.001994    0.003367
75%      0.066389    0.029425    0.073755    0.073246    0.045759    0.042857
max      0.258714    0.228920    0.268164    0.277476    0.233623    0.233772

                X6          X7          X8          X9         X10         X11  \
count  903.000000  903.000000  903.000000  903.000000  903.000000  903.000000
mean     0.022299    0.011281    0.009933    0.010006    0.015116    0.008509
std      0.065455    0.063905    0.064869    0.065563    0.062091    0.064610
```

min	-0.193918	-0.209235	-0.199732	-0.212395	-0.194619	-0.221079
25%	-0.018920	-0.025524	-0.026657	-0.031044	-0.021684	-0.029968
50%	0.022686	0.012085	0.008040	0.013033	0.015706	0.008743
75%	0.063618	0.048580	0.050034	0.047307	0.053235	0.047940
max	0.245953	0.232944	0.236086	0.232248	0.221222	0.188690

	X12
count	903.000000
mean	0.004789
std	0.066067
min	-0.244691
25%	-0.033352
50%	0.007538
75%	0.043389
max	0.241340




nematoda no_efectores dataset 4 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