ds5_nematoda_limpieza_de_datos

December 14, 2020

Limpieza de datos

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

1 Declaración de variables

[2]:
```
organismo ="nematoda"
dataset = 5
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 5, con valores
atípicos.
Valores del documento csv.

        X0      X1      X2     X3     X4      X5     X6     X7     X8      X9  \
0    5.907   2.954   6.118  8.861  0.211  13.502  3.797  4.008  3.586   6.329
1    6.098   7.317   1.220  3.659  0.000   2.439  7.317  4.878  1.220  13.415
2    6.522   8.696   4.710  4.348  1.449   9.783  3.261  3.261  4.348   3.623
3    8.224   5.421   3.925  4.299  1.869   6.916  1.869  7.664  1.869   5.794
4    7.812  12.500   3.125  7.031  3.125  10.156  3.906  3.125  3.125   2.344
..     …       …       …      …      …       …      …      …      …       …
495  4.422   5.782   3.061  7.143  2.381   6.803  2.041  5.782  3.741   8.503
496  6.000   4.857   7.714  6.000  1.714   5.143  2.857  4.571  2.857   5.714
497  8.475   8.475  10.169  3.390  0.000   5.085  5.085  1.695  1.695   6.780
498  4.724  11.811   2.362  5.512  2.362   3.937  3.937  7.087  0.787   4.724
```

```
499  8.152   4.891   4.891  6.522  2.717   8.696  2.717  3.804  7.065   5.435

     …     X11    X12    X13    X14     X15    X16    X17    X18    X19  \
0    …   6.329  4.430  2.743  3.376   7.173  5.696  0.422  1.477  6.118
1    …  17.073  0.000  6.098  2.439   8.537  4.878  0.000  1.220  4.878
2    …   4.348  1.812  3.261  7.971   7.971  5.435  1.449  2.174  8.333
3    …   4.673  3.925  4.860  6.355   5.421  4.673  0.935  3.551  7.664
4    …   7.812  5.469  3.906  5.469   2.344  1.562  4.688  4.688  1.562
..   …     …    …    …    …     …     …    …    …    …
495  …   8.844  2.041  5.782  2.721   6.803  3.741  1.701  4.082  4.082
496  …   4.857  2.286  4.857  3.714  10.857  9.429  0.286  2.286  5.429
497  …   5.085  3.390  8.475  1.695   3.390  5.085  0.000  5.085  8.475
498  …   5.512  3.937  4.724  6.299   3.150  6.299  2.362  5.512  3.937
499  …   6.522  3.261  4.348  2.717   3.261  5.978  1.630  2.717  5.978

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

[500 rows x 21 columns]
```

Composición de aminoácidos (AAC) efectores nematoda dataset 5, con valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     6.904866    6.282428    4.282582    5.093660    2.243218    6.501042
std      2.664158    2.654732    1.988779    2.113812    1.715614    3.175540
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      5.260000    4.423000    2.963000    3.770250    1.207250    4.317000
50%      6.702500    5.946000    4.065000    5.042000    1.969000    6.275500
75%      8.364250    7.732250    5.408500    6.222750    2.913000    8.046000
max     17.021000   16.867000   11.017000   16.279000   11.818000   20.958000

                X6          X7          X8          X9         X10         X11  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     3.637606    5.681830    2.529814    5.650834    9.025518    5.987834
```

```
std      1.659719    3.046573    1.495270    2.176342    2.909455    2.893297
min      0.000000    0.000000    0.000000    0.000000    1.266000    0.000000
25%      2.536500    3.808500    1.569750    4.258000    7.241750    3.981000
50%      3.520000    5.110000    2.299000    5.608500    9.059500    5.599500
75%      4.590500    6.972250    3.262000    6.744500   10.811000    7.482750
max     11.656000   28.736000   10.227000   15.385000   20.779000   17.241000

               X12         X13         X14         X15         X16         X17  \
count  500.000000  500.00000  500.000000  500.000000  500.000000  500.000000
mean     2.792214    4.41418    4.868696    7.770532    5.536588    1.148546
std      1.533969    2.00706    2.710884    3.147775    2.346949    0.984989
min      0.000000    0.00000    0.000000    1.087000    0.000000    0.000000
25%      1.906500    3.09500    3.154750    5.773500    4.162000    0.458500
50%      2.500000    4.25500    4.391000    7.399500    5.197500    0.952000
75%      3.340000    5.69300    6.173000    9.400500    6.667000    1.676500
max     14.286000   12.12100   19.874000   23.256000   20.379000    8.197000

               X18         X19
count  500.000000  500.000000
mean     3.043842    6.604252
std      1.517253    2.485096
min      0.000000    0.000000
25%      1.988500    5.019500
50%      2.931500    6.510500
75%      3.928000    8.096000
max      8.571000   15.054000
```

no_efectores

Composición de aminoácidos (AAC) no_efectores nematoda dataset 5, con valores atípicos.

Valores del documento csv.

```
        X0     X1     X2     X3     X4      X5     X6      X7     X8     X9  \
0    5.972  4.555  4.757  2.733  1.619   6.883  4.049   5.668  1.518  8.806
1    6.098  2.439  4.878  1.220  1.220   7.317  8.537   2.439  2.439  4.878
2    7.753  6.439  4.993  6.176  1.051   6.702  2.891   5.650  2.365  5.125
3    4.545  3.030  4.545  4.545  0.000   1.515  3.030   7.576  3.030  4.545
4    7.312  6.324  3.557  4.941  0.988   4.941  3.162   5.929  1.581  7.510
..     …      …      …      …      …       …      …       …      …      …
495  4.813  2.674  9.091  4.813  1.070  12.834  6.417   4.813  2.139  3.209
496  4.545  3.977  5.114  7.955  0.000   6.250  2.841   8.523  3.977  6.818
497  7.160  5.986  3.873  3.638  0.704   6.338  5.516   5.869  2.230  1.878
498  7.506  9.051  3.974  4.857  2.870   9.051  4.194   2.870  3.753  3.974
499  5.187  4.704  6.031  6.634  0.844   5.066  3.619  20.507  0.965  5.549

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

```
0    …  5.972  1.721  6.579   4.251  8.502  6.174  0.709  2.429   6.478
1    …  7.317  2.439  6.098   0.000  1.220  0.000  2.439  4.878  15.854
2    …  6.307  1.183  5.519   3.942  5.125  4.993  1.314  4.336   7.227
3    …  6.061  6.061  7.576   3.030  3.030  3.030  1.515  7.576  13.636
4    …  4.743  1.976  4.941   4.941  7.708  5.336  1.186  4.150   7.312
..   …    …    …    …     …    …    …    …    …     …
495  …  9.626  3.209  2.139   0.535  4.813  4.813  0.535  2.674   5.348
496  …  7.955  2.841  7.955   7.386  3.977  5.682  1.136  4.545   3.977
497  …  5.282  3.873  3.286  19.366  8.803  4.460  0.000  1.761   3.756
498  …  3.753  2.428  3.974   6.402  7.947  4.857  0.883  1.325   6.843
499  …  3.860  3.136  5.187   4.704  5.549  2.895  1.568  2.051   4.946

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

[500 rows x 21 columns]
```

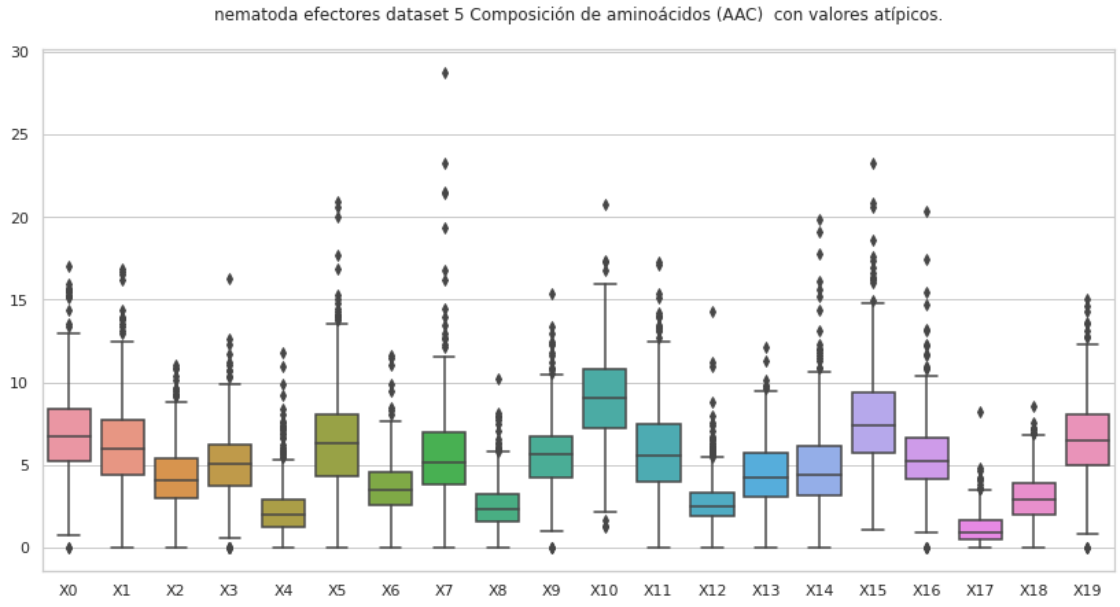
Composición de aminoácidos (AAC) no_efectores nematoda dataset 5, con valores atípicos.
Estadísticas.

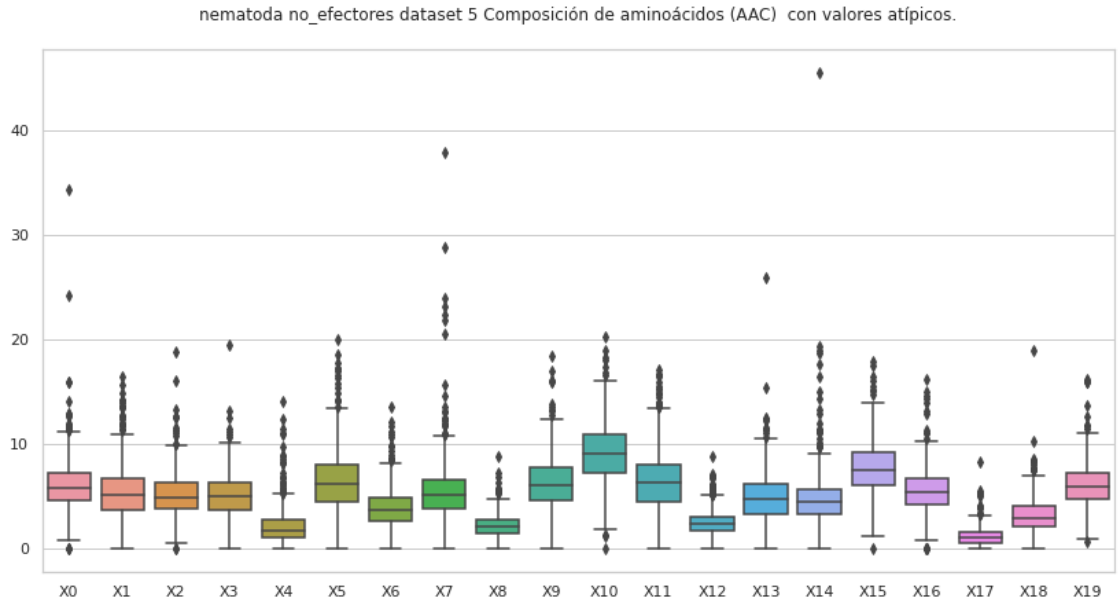
```
                X0          X1          X2          X3          X4          X5  \
count   500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean      6.060952    5.470108    5.162626    5.039412    2.168206    6.496074
std       2.809496    2.630561    2.248411    2.247570    1.851724    3.099753
min       0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%       4.545000    3.701000    3.843250    3.621250    1.083750    4.417250
50%       5.825000    5.138500    4.923500    4.960500    1.737000    6.154000
75%       7.218000    6.687750    6.299500    6.316500    2.742500    8.017000
max      34.340000   16.471000   18.831000   19.481000   14.103000   20.000000

                X6          X7          X8          X9          X10         X11  \
count   500.000000  500.000000  500.000000  500.00000  500.000000  500.000000
mean      3.895962    5.597716    2.174830    6.28540    9.138406    6.592418
std       1.935476    3.430192    1.132651    2.64429    2.931715    2.903530
min       0.000000    0.000000    0.000000    0.00000    0.000000    0.000000
25%       2.597000    3.872250    1.433500    4.58725    7.279500    4.422500
```

50%	3.673000	5.106500	2.091500	6.07200	9.001500	6.262500
75%	4.897750	6.613500	2.761250	7.71550	10.927500	8.051000
max	13.531000	37.805000	8.750000	18.46200	20.238000	17.080000

	X12	X13	X14	X15	X16	X17	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	2.511918	4.878656	4.846922	7.727094	5.515996	1.204710	
std	1.235287	2.398246	3.159222	2.611917	2.120987	1.007192	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	1.676500	3.270000	3.271000	6.028500	4.164250	0.563500	
50%	2.333500	4.675000	4.441000	7.455000	5.338000	1.016500	
75%	3.061000	6.154000	5.714000	9.244250	6.667000	1.630250	
max	8.861000	25.888000	45.455000	17.808000	16.098000	8.219000	

	X18	X19
count	500.000000	500.000000
mean	3.150798	6.081802
std	1.808428	2.249503
min	0.000000	0.719000
25%	2.098000	4.688000
50%	2.894500	5.956000
75%	4.082000	7.261750
max	18.919000	16.162000





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 5, sin valores
atípicos.
Valores del documento csv.

        X0      X1      X2     X3     X4      X5     X6     X7     X8     X9  \
0    5.907   2.954   6.118  8.861  0.211  13.502  3.797  4.008  3.586  6.329
2    6.522   8.696   4.710  4.348  1.449   9.783  3.261  3.261  4.348  3.623
3    8.224   5.421   3.925  4.299  1.869   6.916  1.869  7.664  1.869  5.794
5    7.246   4.831   3.382  4.831  2.899  11.111  3.865  4.831  2.899  3.865
6    4.298  10.602   3.152  3.438  3.725   4.585  3.438  7.736  3.438  4.871
..     …       …       …      …      …       …      …      …      …      …
494  9.524   5.442   2.041  6.122  2.041   3.401  4.082  4.762  3.401  8.163
495  4.422   5.782   3.061  7.143  2.381   6.803  2.041  5.782  3.741  8.503
496  6.000   4.857   7.714  6.000  1.714   5.143  2.857  4.571  2.857  5.714
497  8.475   8.475  10.169  3.390  0.000   5.085  5.085  1.695  1.695  6.780
498  4.724  11.811   2.362  5.512  2.362   3.937  3.937  7.087  0.787  4.724

     …    X11    X12    X13    X14     X15    X16    X17    X18    X19  \
0    …  6.329  4.430  2.743  3.376   7.173  5.696  0.422  1.477  6.118
2    …  4.348  1.812  3.261  7.971   7.971  5.435  1.449  2.174  8.333
3    …  4.673  3.925  4.860  6.355   5.421  4.673  0.935  3.551  7.664
5    …  6.763  5.797  5.314  3.382   6.763  5.314  0.966  1.932  5.797
6    …  4.585  2.006  5.731  4.298   7.163  6.017  2.579  3.152  5.158
..   …    …      …      …      …       …      …      …      …      …
```

```
494  …  5.442  2.041  3.401  4.082   7.483  4.762  1.361  4.762  8.844
495  …  8.844  2.041  5.782  2.721   6.803  3.741  1.701  4.082  4.082
496  …  4.857  2.286  4.857  3.714  10.857  9.429  0.286  2.286  5.429
497  …  5.085  3.390  8.475  1.695   3.390  5.085  0.000  5.085  8.475
498  …  5.512  3.937  4.724  6.299   3.150  6.299  2.362  5.512  3.937

           X20
0     efectores
2     efectores
3     efectores
5     efectores
6     efectores
..          …
494   efectores
495   efectores
496   efectores
497   efectores
498   efectores

[409 rows x 21 columns]
```

Composición de aminoácidos (AAC) efectores nematoda dataset 5, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  409.000000  409.000000  409.000000  409.000000  409.000000  409.000000
mean     6.877325    6.259078    4.292257    5.112342    2.196183    6.507472
std      2.283204    2.433393    1.808035    1.866631    1.434194    2.700361
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      5.319000    4.494000    3.077000    3.968000    1.262000    4.667000
50%      6.771000    5.970000    4.087000    5.056000    2.000000    6.379000
75%      8.313000    7.733000    5.405000    6.122000    2.899000    8.000000
max     14.400000   13.953000   10.169000   11.215000    7.317000   15.000000

                X6          X7          X8          X9         X10         X11  \
count  409.000000  409.000000  409.000000  409.000000  409.000000  409.000000
mean     3.559149    5.399773    2.517159    5.721293    9.399939    6.042680
std      1.413562    2.267623    1.288224    1.955260    2.696920    2.584299
min      0.000000    0.000000    0.000000    0.980000    2.143000    0.000000
25%      2.588000    3.846000    1.639000    4.444000    7.692000    4.167000
50%      3.497000    5.063000    2.326000    5.650000    9.467000    5.675000
75%      4.502000    6.761000    3.214000    6.752000   11.055000    7.460000
max      8.475000   13.986000    6.593000   11.765000   17.391000   14.198000

               X12         X13         X14         X15         X16         X17  \
count  409.000000  409.000000  409.000000  409.000000  409.000000  409.000000
```

mean	2.701924	4.420616	4.739765	7.729902	5.536822	1.139753
std	1.219336	1.886544	2.093527	2.810689	1.932662	0.859476
min	0.000000	0.000000	0.000000	1.087000	0.000000	0.000000
25%	1.929000	3.125000	3.297000	5.882000	4.348000	0.529000
50%	2.475000	4.315000	4.403000	7.407000	5.269000	0.987000
75%	3.261000	5.691000	5.926000	9.391000	6.667000	1.669000
max	7.042000	10.101000	12.048000	16.949000	11.712000	4.098000

	X18	X19
count	409.000000	409.000000
mean	3.068359	6.778308
std	1.427646	2.205811
min	0.000000	0.901000
25%	2.091000	5.333000
50%	3.022000	6.667000
75%	3.943000	8.197000
max	7.254000	13.158000

no_efectores
Composición de aminoácidos (AAC) no_efectores nematoda dataset 5, sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9	\
0	5.972	4.555	4.757	2.733	1.619	6.883	4.049	5.668	1.518	8.806	
2	7.753	6.439	4.993	6.176	1.051	6.702	2.891	5.650	2.365	5.125	
4	7.312	6.324	3.557	4.941	0.988	4.941	3.162	5.929	1.581	7.510	
5	3.855	7.029	5.215	7.710	1.814	4.989	2.948	5.669	3.855	6.576	
6	4.673	3.271	1.869	4.206	0.935	4.206	4.673	3.738	0.935	10.748	
..	…	…	…	…	…	…	…	…	…	…	
492	5.405	6.950	5.212	7.239	1.351	6.081	3.089	6.564	2.992	7.432	
493	4.626	7.048	4.626	8.370	2.203	9.471	2.203	4.405	2.423	5.286	
495	4.813	2.674	9.091	4.813	1.070	12.834	6.417	4.813	2.139	3.209	
496	4.545	3.977	5.114	7.955	0.000	6.250	2.841	8.523	3.977	6.818	
498	7.506	9.051	3.974	4.857	2.870	9.051	4.194	2.870	3.753	3.974	

	…	X11	X12	X13	X14	X15	X16	X17	X18	X19	\
0	…	5.972	1.721	6.579	4.251	8.502	6.174	0.709	2.429	6.478	
2	…	6.307	1.183	5.519	3.942	5.125	4.993	1.314	4.336	7.227	
4	…	4.743	1.976	4.941	4.941	7.708	5.336	1.186	4.150	7.312	
5	…	7.483	2.041	5.669	4.989	6.803	4.535	0.680	4.082	3.401	
6	…	5.607	4.206	5.140	4.206	9.346	7.944	2.336	4.673	5.140	
..	…	…	…	…	…	…	…	…	…	…	
492	…	5.309	3.089	4.247	4.247	6.371	3.958	0.772	2.896	7.625	
493	…	5.947	2.423	7.930	4.185	6.608	6.167	2.423	3.084	3.965	
495	…	9.626	3.209	2.139	0.535	4.813	4.813	0.535	2.674	5.348	

```
496  …  7.955  2.841  7.955  7.386  3.977  5.682  1.136  4.545  3.977
498  …  3.753  2.428  3.974  6.402  7.947  4.857  0.883  1.325  6.843

              X20
0     no_efectores
2     no_efectores
4     no_efectores
5     no_efectores
6     no_efectores
..             …
492   no_efectores
493   no_efectores
495   no_efectores
496   no_efectores
498   no_efectores

[397 rows x 21 columns]
```

Composición de aminoácidos (AAC) no_efectores nematoda dataset 5, sin valores atípicos.
Estadísticas.

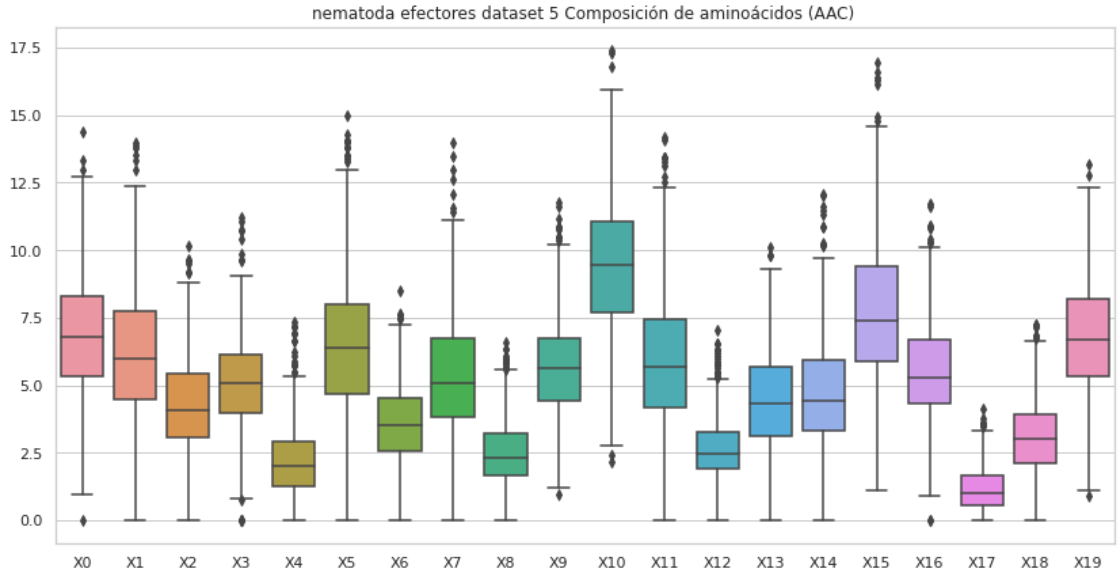
	X0	X1	X2	X3	X4	X5	\
count	397.000000	397.000000	397.000000	397.000000	397.000000	397.000000	
mean	6.044040	5.398461	5.089668	5.187806	2.031950	6.501224	
std	2.051463	2.147994	1.887707	1.962174	1.279322	2.599402	
min	0.000000	0.909000	0.000000	0.775000	0.000000	0.735000	
25%	4.762000	3.891000	3.930000	3.922000	1.149000	4.706000	
50%	5.936000	5.224000	4.944000	5.143000	1.754000	6.250000	
75%	7.110000	6.631000	6.269000	6.349000	2.589000	7.925000	
max	14.062000	12.648000	11.248000	11.441000	6.818000	15.405000	

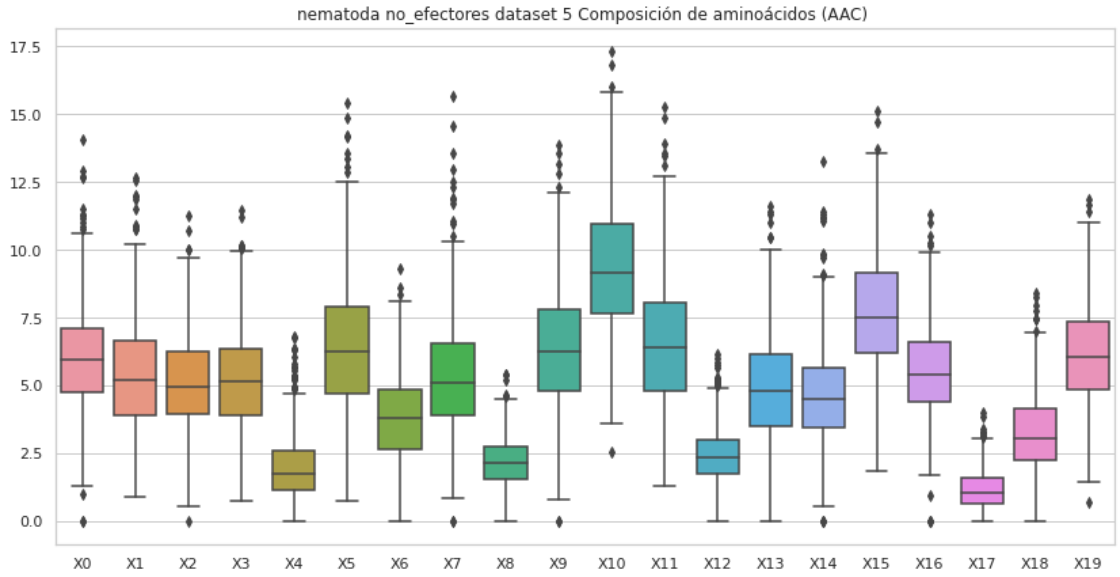
	X6	X7	X8	X9	X10	X11	\
count	397.000000	397.000000	397.000000	397.000000	397.000000	397.000000	
mean	3.847469	5.361418	2.191096	6.428262	9.366927	6.614035	
std	1.634337	2.267474	0.936184	2.328244	2.551594	2.437420	
min	0.000000	0.000000	0.000000	0.000000	2.542000	1.322000	
25%	2.658000	3.927000	1.556000	4.815000	7.634000	4.795000	
50%	3.778000	5.120000	2.174000	6.250000	9.170000	6.406000	
75%	4.854000	6.534000	2.761000	7.805000	10.959000	8.054000	
max	9.326000	15.646000	5.426000	13.861000	17.308000	15.267000	

	X12	X13	X14	X15	X16	X17	\
count	397.000000	397.000000	397.000000	397.000000	397.000000	397.000000	
mean	2.488373	4.916987	4.698917	7.714368	5.516647	1.174476	
std	1.055583	2.010476	2.002406	2.255969	1.715872	0.778561	
min	0.000000	0.000000	0.000000	1.840000	0.000000	0.000000	

25%	1.746000	3.490000	3.448000	6.226000	4.380000	0.643000
50%	2.353000	4.817000	4.500000	7.519000	5.398000	1.031000
75%	3.020000	6.154000	5.678000	9.174000	6.617000	1.621000
max	6.140000	11.616000	13.235000	15.126000	11.304000	4.000000

	X18	X19
count	397.000000	397.000000
mean	3.239343	6.188511
std	1.526464	1.948502
min	0.000000	0.719000
25%	2.273000	4.854000
50%	3.030000	6.068000
75%	4.155000	7.332000
max	8.393000	11.842000





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 5, con valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.023918	0.000854	0.035877	0.054670	0.011105	0.016230	0.014522	
1	0.020419	0.000000	0.012252	0.008168	0.020419	0.016336	0.004084	
2	0.035114	0.007803	0.023410	0.052671	0.017557	0.017557	0.023410	
3	0.031792	0.007225	0.016618	0.026734	0.018786	0.029624	0.007225	
4	0.020044	0.008018	0.018040	0.026058	0.010022	0.008018	0.008018	
..	…	…	…	…	…	…	…	
495	0.032721	0.017619	0.052858	0.050340	0.042789	0.042789	0.027687	
496	0.027135	0.007753	0.027135	0.023259	0.021967	0.020675	0.012922	
497	0.142291	0.000000	0.056916	0.085375	0.142291	0.028458	0.028458	
498	0.029687	0.014844	0.034635	0.024739	0.029687	0.044531	0.004948	
499	0.092390	0.030797	0.073912	0.098550	0.049275	0.043116	0.080072	

	X7	X8	X9	…	X74	X75	X76	\
0	0.025626	0.025626	0.028189	…	0.000005	0.034888	0.011323	
1	0.044923	0.057174	0.024503	…	-0.004035	-0.022083	0.008982	
2	0.019508	0.023410	0.039016	…	0.044711	0.025192	-0.014887	
3	0.022399	0.018063	0.039017	…	0.015523	0.013134	0.017595	
4	0.006013	0.020044	0.016036	…	0.036611	0.024920	0.010128	
..	…	…	…	…	…	…	…	
495	0.062926	0.065443	0.078028	…	-0.044896	-0.030442	0.004509	
496	0.025843	0.021967	0.038765	…	0.010535	0.006034	0.012726	
497	0.113833	0.085375	0.142291	…	-0.211466	-0.178494	0.038460	
498	0.029687	0.034635	0.069270	…	-0.042067	0.007479	-0.011347	
499	0.061594	0.073912	0.098550	…	0.033506	0.053402	0.086009	

	X77	X78	X79	X80	X81	X82	X83
0	-0.004262	-0.007366	0.000712	-0.004239	0.021984	0.013252	efectores
1	0.023777	0.041813	0.005362	0.004562	0.005433	0.045855	efectores
2	0.018115	0.034526	-0.028871	-0.010482	-0.018786	0.008064	efectores
3	-0.006813	-0.002715	0.015139	-0.014424	-0.005677	0.007176	efectores
4	0.014575	0.026664	-0.013920	0.010408	0.015858	0.005704	efectores
..	…	…	…	…	…	…	
495	-0.002839	0.017366	0.020774	-0.002288	-0.003907	0.013377	efectores
496	0.013285	0.012651	0.011451	-0.030791	-0.030241	0.031476	efectores

```
497  0.416952  0.170616  0.082279 -0.228294 -0.235874 -0.121083  efectores
498  0.057192 -0.013961  0.033309 -0.009164 -0.006034  0.002869  efectores
499 -0.032592  0.033422 -0.018537 -0.069504 -0.021492 -0.007670  efectores

[500 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro_mass efectores nematoda
dataset 5, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.037236	0.012977	0.029510	0.036789	0.027612	0.030480	
std	0.033150	0.014139	0.030606	0.035984	0.042916	0.027331	
min	-0.267885	0.000000	-0.000000	-0.089295	-0.267885	-0.089295	
25%	0.022339	0.004722	0.014386	0.017751	0.011830	0.016236	
50%	0.033089	0.009485	0.024428	0.031138	0.020362	0.025300	
75%	0.045491	0.017004	0.037229	0.046842	0.032800	0.038313	
max	0.398522	0.159409	0.482764	0.482764	0.675869	0.386211	

	X6	X7	X8	X9	…	X73	\
count	500.000000	500.000000	500.000000	500.000000	…	500.000000	
mean	0.015377	0.033576	0.035278	0.053361	…	0.009348	
std	0.015405	0.031597	0.040743	0.050812	…	0.056909	
min	0.000000	-0.089295	-0.089295	-0.178590	…	-0.774609	
25%	0.006129	0.017179	0.017707	0.027280	…	0.000296	
50%	0.011337	0.027137	0.027950	0.045333	…	0.010361	
75%	0.019404	0.040225	0.043245	0.066771	…	0.023557	
max	0.136692	0.386211	0.579316	0.729023	…	0.814431	

	X74	X75	X76	X77	X78	X79	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.003093	0.008095	0.010476	0.002251	0.006120	0.008115	
std	0.063981	0.043711	0.043975	0.065047	0.070141	0.038119	
min	-0.492493	-0.334372	-0.450185	-0.676155	-1.287588	-0.510732	
25%	-0.011032	-0.003732	0.000430	-0.011691	-0.004314	-0.000853	
50%	0.003331	0.007731	0.011482	0.001610	0.008030	0.010426	
75%	0.016755	0.019883	0.023571	0.016670	0.020921	0.023245	
max	0.924382	0.583629	0.490736	0.636003	0.321247	0.103342	

	X80	X81	X82
count	500.000000	500.000000	500.000000
mean	-0.004627	0.003183	0.008684
std	0.086959	0.061561	0.045250
min	-1.300089	-0.799507	-0.624921
25%	-0.014491	-0.005530	0.000994
50%	0.002392	0.005442	0.011711

```
75%      0.014630    0.020810    0.021651
max      0.588748    0.646817    0.105304

[8 rows x 83 columns]
```

```
no_efectores
Composición de pseudo aminoácidos (PseAAC)  hidro_mass no_efectores nematoda
dataset 5, con valores atípicos.
Valores del documento csv.
```

	X0	X1	X2	X3	X4	X5	X6	\
0	0.018784	0.005094	0.008596	0.021650	0.020695	0.017829	0.004776	
1	0.028057	0.005611	0.005611	0.033669	0.028057	0.011223	0.011223	
2	0.060218	0.008165	0.047970	0.052052	0.042867	0.043887	0.018371	
3	0.017683	0.000000	0.017683	0.005894	0.029471	0.029471	0.011789	
4	0.027522	0.003719	0.018596	0.018596	0.018596	0.022315	0.005951	
..	…	…	…	…	…	…	…	
495	0.021890	0.004864	0.021890	0.058374	0.009729	0.021890	0.009729	
496	0.034378	0.000000	0.060161	0.047269	0.060161	0.064458	0.030080	
497	0.018974	0.001866	0.009642	0.016796	0.008709	0.015552	0.005910	
498	0.037441	0.014316	0.024226	0.045149	0.019822	0.014316	0.018720	
499	0.009896	0.001611	0.012657	0.009666	0.009896	0.039123	0.001841	

	X7	X8	X9	…	X74	X75	X76	\
0	0.027699	0.018784	0.033430	…	0.009010	0.005298	0.000551	
1	0.022446	0.033669	0.084172	…	0.032274	0.009244	-0.024029	
2	0.039805	0.048991	0.084713	…	0.002676	0.000580	0.016003	
3	0.017683	0.023577	0.047154	…	0.064639	0.031630	0.022804	
4	0.028266	0.017852	0.043143	…	0.004938	0.000253	0.005661	
..	…	…	…	…	…	…	…	
495	0.014593	0.043780	0.065670	…	0.002606	0.011127	0.024259	
496	0.051566	0.060161	0.034378	…	0.002495	-0.000955	-0.024963	
497	0.004977	0.013997	0.016485	…	0.011121	0.021027	0.016287	
498	0.019822	0.018720	0.047351	…	0.031125	0.018403	0.013785	
499	0.010586	0.007364	0.013348	…	0.004062	0.005189	0.037068	

	X77	X78	X79	X80	X81	X82	X83
0	0.014897	0.013477	0.001271	0.008197	0.010806	0.012658	no_efectores
1	0.050158	-0.007394	0.009774	0.066965	0.020483	0.041491	no_efectores
2	-0.001979	-0.023601	0.017991	-0.005541	0.003276	0.037310	no_efectores
3	0.098700	0.042004	0.037277	-0.035123	-0.000730	0.014809	no_efectores
4	0.018498	0.005001	0.007361	0.003080	-0.007595	0.009091	no_efectores
..	…	…	…	…	…	…	…
495	-0.040387	0.005595	0.006697	-0.004823	0.034408	0.002120	no_efectores
496	0.005550	0.023354	0.000570	0.024018	0.025436	0.001869	no_efectores
497	0.000680	0.011418	0.019288	0.000287	0.008859	0.020177	no_efectores

```
498  0.003673  0.009379 -0.002360 -0.010707  0.018860 -0.000566  no_efectores
499 -0.001014  0.002373  0.038917 -0.004043 -0.001652  0.039802  no_efectores

[500 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores nematoda
dataset 5, con valores atípicos.
Estadísticas.

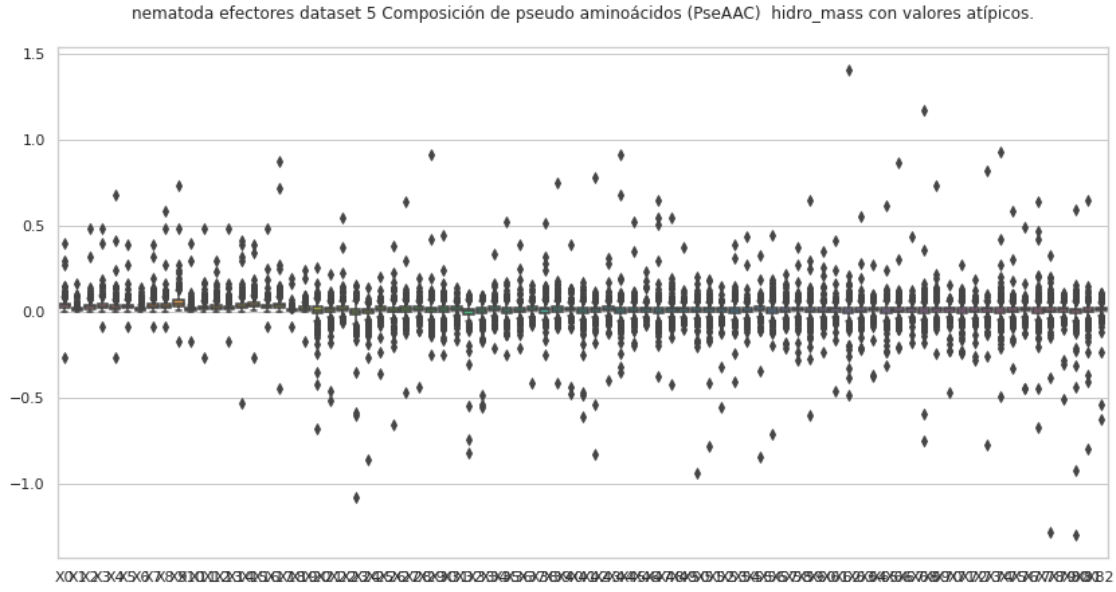
	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.030435	0.012024	0.027035	0.033695	0.027038	0.027847	
std	0.018244	0.013332	0.019285	0.022123	0.021180	0.019577	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.018363	0.004129	0.013548	0.017989	0.012679	0.014846	
50%	0.027190	0.008291	0.023904	0.030076	0.021369	0.023968	
75%	0.038808	0.015345	0.036579	0.043687	0.035380	0.035805	
max	0.158176	0.121745	0.187174	0.145053	0.165183	0.234993	

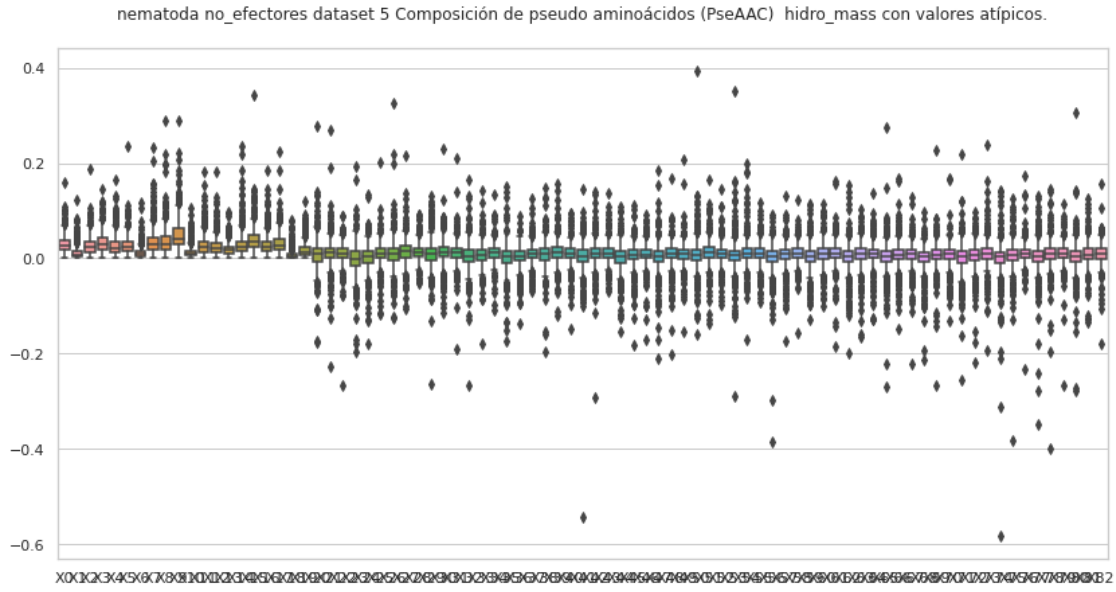
	X6	X7	X8	X9	…	X73	\
count	500.000000	500.000000	500.000000	500.000000	…	500.000000	
mean	0.012260	0.034510	0.035875	0.050180	…	0.010303	
std	0.011808	0.026867	0.029286	0.034016	…	0.028193	
min	0.000000	0.000000	0.000000	0.000000	…	-0.193668	
25%	0.005363	0.017927	0.018142	0.028327	…	-0.000047	
50%	0.009455	0.029091	0.029655	0.042061	…	0.009648	
75%	0.015058	0.043055	0.045277	0.062383	…	0.020184	
max	0.117198	0.233346	0.289990	0.289990	…	0.237159	

	X74	X75	X76	X77	X78	X79	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	-0.001518	0.004994	0.010431	0.002188	0.007705	0.010294	
std	0.042534	0.030228	0.023146	0.036096	0.032352	0.023179	
min	-0.582296	-0.382231	-0.232510	-0.348947	-0.400364	-0.266532	
25%	-0.008700	-0.004758	0.001519	-0.007003	-0.001261	0.001565	
50%	0.002996	0.006353	0.009366	0.003961	0.009129	0.009833	
75%	0.012668	0.017856	0.019362	0.015540	0.020148	0.020137	
max	0.142534	0.124374	0.172015	0.131641	0.144147	0.115163	

	X80	X81	X82
count	500.000000	500.000000	500.000000
mean	0.002746	0.008665	0.009242
std	0.035344	0.025835	0.022653
min	-0.277371	-0.134813	-0.177726
25%	-0.007710	-0.001685	-0.000036
50%	0.003887	0.008356	0.008488
75%	0.015351	0.021145	0.020193

```
max       0.305613     0.126365     0.156463

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

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.023918  0.000854  0.035877  0.054670  0.011105  0.016230  0.014522
1    0.020419  0.000000  0.012252  0.008168  0.020419  0.016336  0.004084
2    0.035114  0.007803  0.023410  0.052671  0.017557  0.017557  0.023410
3    0.031792  0.007225  0.016618  0.026734  0.018786  0.029624  0.007225
4    0.020044  0.008018  0.018040  0.026058  0.010022  0.008018  0.008018
..        …         …         …         …         …         …         …
493  0.053244  0.015973  0.042595  0.042595  0.042595  0.021298  0.015973
494  0.064945  0.013917  0.041750  0.023195  0.023195  0.032472  0.023195
495  0.032721  0.017619  0.052858  0.050340  0.042789  0.042789  0.027687
496  0.027135  0.007753  0.027135  0.023259  0.021967  0.020675  0.012922
498  0.029687  0.014844  0.034635  0.024739  0.029687  0.044531  0.004948

           X7        X8        X9  …       X74       X75       X76  \
0    0.025626  0.025626  0.028189  …  0.000005  0.034888  0.011323
1    0.044923  0.057174  0.024503  … -0.004035 -0.022083  0.008982
2    0.019508  0.023410  0.039016  …  0.044711  0.025192 -0.014887
3    0.022399  0.018063  0.039017  …  0.015523  0.013134  0.017595
4    0.006013  0.020044  0.016036  …  0.036611  0.024920  0.010128
..        …         …         …  …         …         …         …
493  0.037271  0.047920  0.085190  …  0.035214  0.027878  0.009278
494  0.055667  0.037111  0.060306  …  0.010296  0.004815  0.001940
495  0.062926  0.065443  0.078028  … -0.044896 -0.030442  0.004509
496  0.025843  0.021967  0.038765  …  0.010535  0.006034  0.012726
498  0.029687  0.034635  0.069270  … -0.042067  0.007479 -0.011347

           X77       X78       X79       X80       X81       X82         X83
0   -0.004262 -0.007366  0.000712 -0.004239  0.021984  0.013252  efectores
1    0.023777  0.041813  0.005362  0.004562  0.005433  0.045855  efectores
2    0.018115  0.034526 -0.028871 -0.010482 -0.018786  0.008064  efectores
3   -0.006813 -0.002715  0.015139 -0.014424 -0.005677  0.007176  efectores
4    0.014575  0.026664 -0.013920  0.010408  0.015858  0.005704  efectores
..        …         …         …         …         …         …         …
493 -0.043340 -0.010690  0.030939  0.013232 -0.000149  0.025164  efectores
494 -0.060487 -0.048822  0.068155  0.013160  0.009173 -0.025413  efectores
495 -0.002839  0.017366  0.020774 -0.002288 -0.003907  0.013377  efectores
496  0.013285  0.012651  0.011451 -0.030791 -0.030241  0.031476  efectores
498  0.057192 -0.013961  0.033309 -0.009164 -0.006034  0.002869  efectores
```

[453 rows x 84 columns]

Composición de pseudo aminoácidos (PseAAC) hidro_mass efectores nematoda
dataset 5, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	453.000000	453.000000	453.000000	453.000000	453.000000	453.000000	
mean	0.033577	0.011261	0.025809	0.032359	0.022423	0.027273	
std	0.016426	0.009154	0.015160	0.019229	0.015626	0.014834	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.022114	0.004632	0.013871	0.017443	0.011323	0.016108	
50%	0.031802	0.008806	0.023152	0.029632	0.018780	0.024192	
75%	0.041990	0.016015	0.034608	0.042534	0.029685	0.035992	
max	0.119448	0.054098	0.089421	0.119448	0.109722	0.084624	

	X6	X7	X8	X9	…	X73	\
count	453.000000	453.000000	453.000000	453.000000	…	453.000000	
mean	0.012727	0.028680	0.030110	0.046286	…	0.011237	
std	0.009761	0.017652	0.019131	0.026465	…	0.017750	
min	0.000000	0.000000	0.000000	0.001339	…	-0.072567	
25%	0.005665	0.016736	0.017038	0.026131	…	0.001162	
50%	0.010297	0.025231	0.026628	0.041608	…	0.010441	
75%	0.017751	0.036913	0.039897	0.060510	…	0.022761	
max	0.058174	0.115505	0.108237	0.165390	…	0.079079	

	X74	X75	X76	X77	X78	X79	\
count	453.000000	453.000000	453.000000	453.000000	453.000000	453.000000	
mean	0.002091	0.007585	0.011742	0.002818	0.008400	0.010583	
std	0.027754	0.022394	0.019169	0.025328	0.023817	0.019477	
min	-0.150371	-0.083427	-0.056367	-0.115102	-0.117424	-0.070711	
25%	-0.009791	-0.002665	0.001475	-0.009792	-0.002958	0.000203	
50%	0.003796	0.007722	0.011574	0.002082	0.008084	0.010589	
75%	0.015523	0.019056	0.022179	0.015490	0.020398	0.023044	
max	0.168902	0.098163	0.092976	0.108974	0.126870	0.079690	

	X80	X81	X82
count	453.000000	453.000000	453.000000
mean	0.002686	0.007486	0.011511
std	0.026563	0.023033	0.018253
min	-0.129962	-0.103208	-0.068227
25%	-0.011331	-0.003370	0.001537
50%	0.003022	0.005727	0.011756
75%	0.014620	0.020481	0.021159
max	0.119252	0.103817	0.075882

[8 rows x 83 columns]

no_efectores
Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores nematoda
dataset 5, sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.018784	0.005094	0.008596	0.021650	0.020695	0.017829	0.004776	
1	0.028057	0.005611	0.005611	0.033669	0.028057	0.011223	0.011223	
2	0.060218	0.008165	0.047970	0.052052	0.042867	0.043887	0.018371	
3	0.017683	0.000000	0.017683	0.005894	0.029471	0.029471	0.011789	
4	0.027522	0.003719	0.018596	0.018596	0.018596	0.022315	0.005951	
..	…	…	…	…	…	…	…	
495	0.021890	0.004864	0.021890	0.058374	0.009729	0.021890	0.009729	
496	0.034378	0.000000	0.060161	0.047269	0.060161	0.064458	0.030080	
497	0.018974	0.001866	0.009642	0.016796	0.008709	0.015552	0.005910	
498	0.037441	0.014316	0.024226	0.045149	0.019822	0.014316	0.018720	
499	0.009896	0.001611	0.012657	0.009666	0.009896	0.039123	0.001841	

	X7	X8	X9	…	X74	X75	X76	\
0	0.027699	0.018784	0.033430	…	0.009010	0.005298	0.000551	
1	0.022446	0.033669	0.084172	…	0.032274	0.009244	-0.024029	
2	0.039805	0.048991	0.084713	…	0.002676	0.000580	0.016003	
3	0.017683	0.023577	0.047154	…	0.064639	0.031630	0.022804	
4	0.028266	0.017852	0.043143	…	0.004938	0.000253	0.005661	
..	…	…	…	…	…	…	…	
495	0.014593	0.043780	0.065670	…	0.002606	0.011127	0.024259	
496	0.051566	0.060161	0.034378	…	0.002495	-0.000955	-0.024963	
497	0.004977	0.013997	0.016485	…	0.011121	0.021027	0.016287	
498	0.019822	0.018720	0.047351	…	0.031125	0.018403	0.013785	
499	0.010586	0.007364	0.013348	…	0.004062	0.005189	0.037068	

	X77	X78	X79	X80	X81	X82	X83
0	0.014897	0.013477	0.001271	0.008197	0.010806	0.012658	no_efectores
1	0.050158	-0.007394	0.009774	0.066965	0.020483	0.041491	no_efectores
2	-0.001979	-0.023601	0.017991	-0.005541	0.003276	0.037310	no_efectores
3	0.098700	0.042004	0.037277	-0.035123	-0.000730	0.014809	no_efectores
4	0.018498	0.005001	0.007361	0.003080	-0.007595	0.009091	no_efectores
..	…	…	…	…	…	…	…
495	-0.040387	0.005595	0.006697	-0.004823	0.034408	0.002120	no_efectores
496	0.005550	0.023354	0.000570	0.024018	0.025436	0.001869	no_efectores
497	0.000680	0.011418	0.019288	0.000287	0.008859	0.020177	no_efectores
498	0.003673	0.009379	-0.002360	-0.010707	0.018860	-0.000566	no_efectores
499	-0.001014	0.002373	0.038917	-0.004043	-0.001652	0.039802	no_efectores

[401 rows x 84 columns]

Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores nematoda
dataset 5, sin valores atípicos.
Estadísticas.

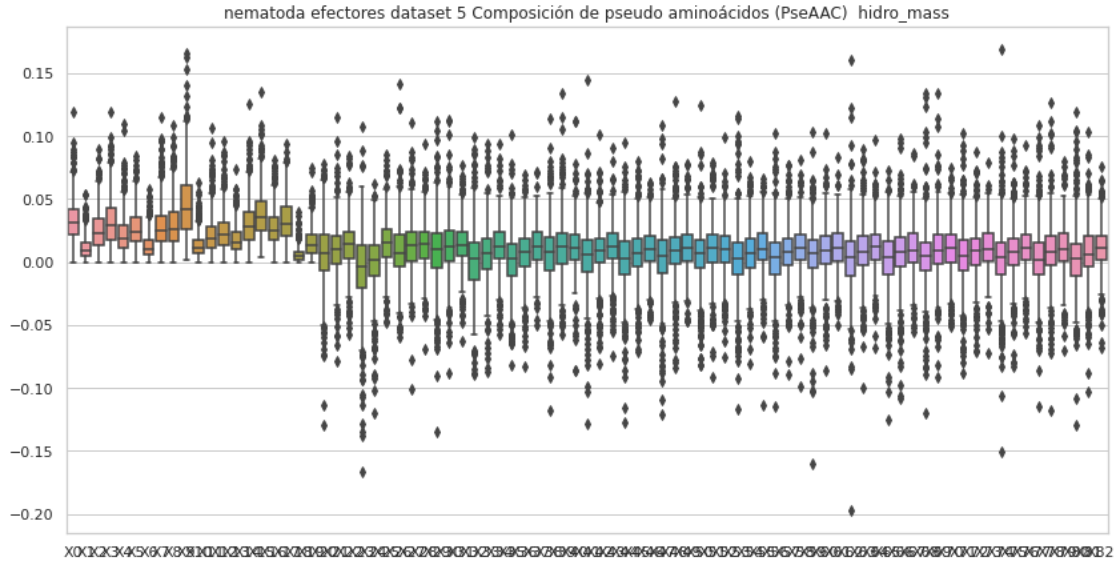
	X0	X1	X2	X3	X4	X5	\
count	401.000000	401.000000	401.000000	401.000000	401.000000	401.000000	
mean	0.026640	0.009637	0.023121	0.028473	0.021872	0.024317	
std	0.013261	0.008456	0.013748	0.015711	0.013482	0.013459	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.017657	0.004041	0.012291	0.016844	0.011548	0.014065	
50%	0.024955	0.007568	0.020596	0.026223	0.019798	0.021899	
75%	0.034132	0.012787	0.032111	0.038956	0.030047	0.032305	
max	0.082083	0.049049	0.069041	0.087217	0.071359	0.073396	

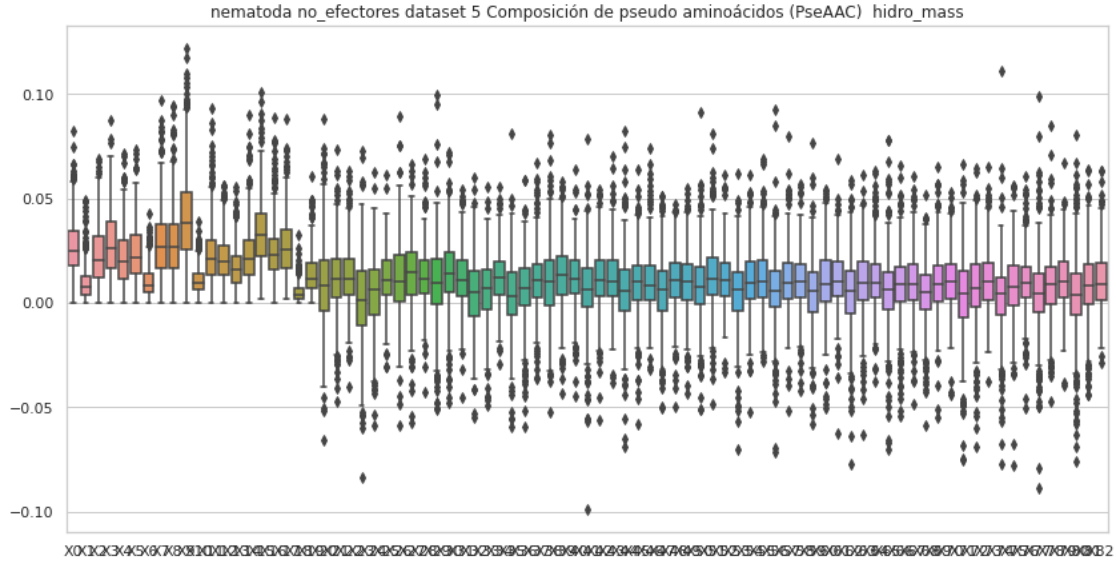
	X6	X7	X8	X9	…	X73	\
count	401.000000	401.000000	401.000000	401.000000	…	401.000000	
mean	0.010058	0.028491	0.029011	0.041358	…	0.010636	
std	0.006869	0.016351	0.016822	0.022518	…	0.014106	
min	0.000000	0.000000	0.000000	0.000000	…	-0.037837	
25%	0.004926	0.016396	0.016640	0.025753	…	0.001452	
50%	0.008536	0.026915	0.026602	0.038053	…	0.009903	
75%	0.013821	0.037454	0.037511	0.052942	…	0.019221	
max	0.042431	0.097004	0.094053	0.121731	…	0.065154	

	X74	X75	X76	X77	X78	X79	\
count	401.000000	401.000000	401.000000	401.000000	401.000000	401.000000	
mean	0.002566	0.007508	0.010063	0.003991	0.008310	0.011213	
std	0.017305	0.015712	0.013806	0.019412	0.016277	0.013472	
min	-0.077069	-0.077755	-0.050583	-0.088725	-0.047546	-0.039565	
25%	-0.005922	-0.001511	0.002661	-0.004382	-0.000725	0.002596	
50%	0.004163	0.007401	0.009257	0.004327	0.008955	0.010207	
75%	0.011936	0.017703	0.017052	0.014179	0.017256	0.019719	
max	0.110985	0.056001	0.060621	0.098700	0.084942	0.067225	

	X80	X81	X82
count	401.000000	401.000000	401.000000
mean	0.003439	0.008753	0.010797
std	0.019785	0.016816	0.014387
min	-0.076056	-0.044321	-0.028913
25%	-0.005491	-0.000382	0.001472
50%	0.003917	0.008376	0.009054
75%	0.013989	0.018597	0.019297
max	0.080586	0.063863	0.063385

[8 rows x 83 columns]


nematoda efectores dataset 5 Composición de pseudo aminoácidos (PseAAC) hidro_mass
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
−0.20


nematoda no_efectores dataset 5 Composición de pseudo aminoácidos (PseAAC) hidro_mass
0.10
0.05
0.00
−0.05
−0.10

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 5,
con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.042854  0.001531  0.064281  0.097953  0.019897  0.029080  0.026019
1    0.043424  0.000000  0.026054  0.017370  0.043424  0.034739  0.008685
2    0.057470  0.012771  0.038314  0.086206  0.028735  0.028735  0.038314
3    0.041266  0.009379  0.021571  0.034701  0.024384  0.038452  0.009379
4    0.058240  0.023296  0.052416  0.075713  0.029120  0.023296  0.023296
..        …         …         …         …         …         …         …
495  0.039465  0.021250  0.063751  0.060715  0.051608  0.051608  0.033393
```

```
496  0.026362  0.007532  0.026362  0.022596  0.021341  0.020085  0.012553
497  0.082347  0.000000  0.032939  0.049408  0.082347  0.016469  0.016469
498  0.044417  0.022208  0.051820  0.037014  0.044417  0.066625  0.007403
499  0.082275  0.027425  0.065820  0.087760  0.043880  0.038395  0.071305

           X7        X8        X9  …       X32       X33       X34  \
0    0.045915  0.045915  0.050507  …  0.004751  0.011412  0.006996
1    0.095533  0.121588  0.052109  … -0.000698 -0.005374 -0.021652
2    0.031928  0.038314  0.063856  …  0.023926  0.039063  0.016762
3    0.029074  0.023446  0.050644  …  0.023312  0.018987  0.015387
4    0.017472  0.058240  0.046592  …  0.017682  0.030241  0.003202
..        …         …         …  …         …         …         …
495  0.075894  0.078929  0.094108  …  0.008918  0.001590 -0.002503
496  0.025107  0.021341  0.037660  …  0.030363  0.028814  0.025690
497  0.065877  0.049408  0.082347  …  0.058979 -0.001097  0.003965
498  0.044417  0.051820  0.103639  … -0.054963  0.030405 -0.044193
499  0.054850  0.065820  0.087760  …  0.030283 -0.001447 -0.001394

           X35       X36       X37       X38       X39       X40         X41
0    0.009581  0.019231  0.018746  0.020288  0.001276  0.023744  efectores
1   -0.005021 -0.005564  0.054367  0.019100  0.011403  0.097515  efectores
2    0.012905 -0.014809 -0.020793 -0.024365 -0.047252  0.013197  efectores
3    0.021540  0.019789  0.013127  0.022838  0.019651  0.009315  efectores
4   -0.035824  0.046260  0.025987  0.029428 -0.040446  0.016574  efectores
..        …         …         …         …         …         …         …
495  0.022555  0.019919  0.046888  0.005438  0.025055  0.016133  efectores
496  0.026932  0.020172  0.023719  0.012363  0.011125  0.030579  efectores
497  0.037963 -0.110788  0.058848  0.022258  0.047617 -0.070073  efectores
498  0.011029 -0.042257  0.009850 -0.016977  0.049837  0.004292  efectores
499  0.026221  0.004726 -0.074604  0.076592 -0.016507 -0.006830  efectores

[500 rows x 42 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass efectores nematoda dataset 5, con valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.046907    0.015687    0.036635    0.048243    0.032709    0.037117
std      0.021108    0.012921    0.021240    0.034617    0.021439    0.017110
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.034335    0.006958    0.023025    0.026320    0.018631    0.025352
50%      0.044494    0.012682    0.033003    0.040538    0.027948    0.035225
75%      0.056281    0.021520    0.046305    0.061307    0.043400    0.045122
max      0.205922    0.088816    0.147087    0.294174    0.160423    0.176504
```

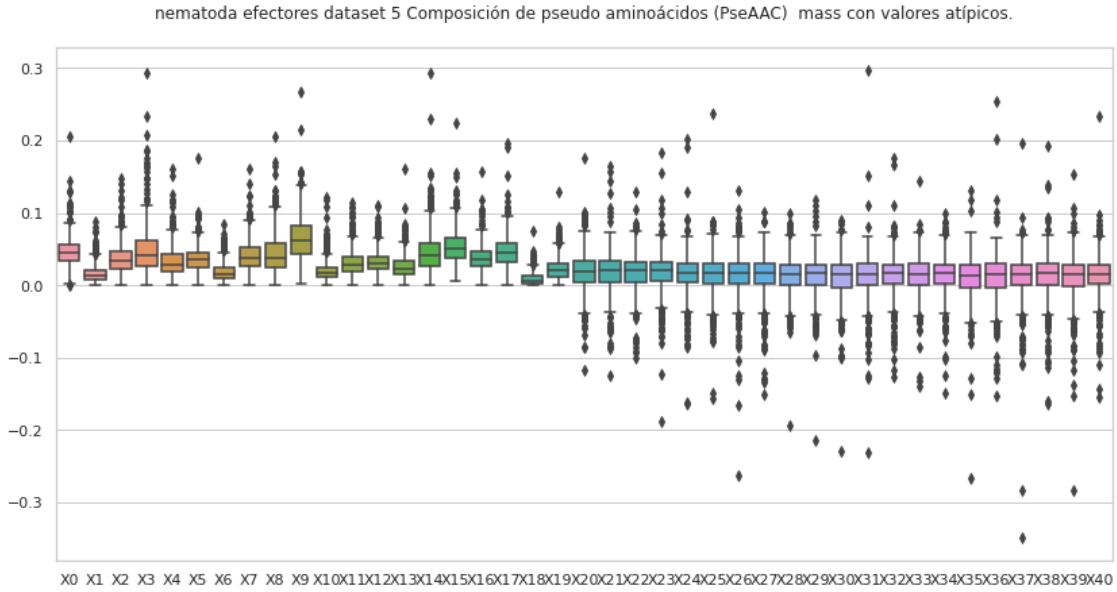
```
              X6          X7          X8          X9  …         X31  \
count  500.000000  500.000000  500.000000  500.000000  …  500.000000
mean     0.018320    0.040455    0.044461    0.064681  …    0.012641
std      0.013077    0.021141    0.028915    0.030812  …    0.033254
min      0.000000    0.000000    0.000000    0.002898  …   -0.232289
25%      0.008735    0.026525    0.023726    0.043298  …    0.000885
50%      0.015605    0.037199    0.038381    0.062255  …    0.015304
75%      0.023729    0.052238    0.058320    0.081770  …    0.030110
max      0.083816    0.160423    0.205922    0.267589  …    0.296931

              X32         X33         X34         X35         X36         X37  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.013635    0.013505    0.013313    0.010819    0.011770    0.011466
std      0.030252    0.027088    0.027481    0.030715    0.033307    0.035074
min     -0.127360   -0.139015   -0.149718   -0.266178   -0.152465   -0.349047
25%      0.001687    0.000676    0.001601   -0.003416   -0.002907    0.000188
50%      0.016913    0.015910    0.016167    0.012623    0.015799    0.015575
75%      0.028625    0.030141    0.029024    0.029039    0.029507    0.028337
max      0.176254    0.144017    0.099803    0.131397    0.253750    0.196120

              X38         X39         X40
count  500.000000  500.000000  500.000000
mean     0.013477    0.010904    0.013394
std      0.031990    0.032614    0.029961
min     -0.164778   -0.283641   -0.155616
25%      0.000894   -0.002178    0.001798
50%      0.017644    0.015090    0.015530
75%      0.029873    0.028181    0.028352
max      0.192939    0.154013    0.233535

[8 rows x 41 columns]
```

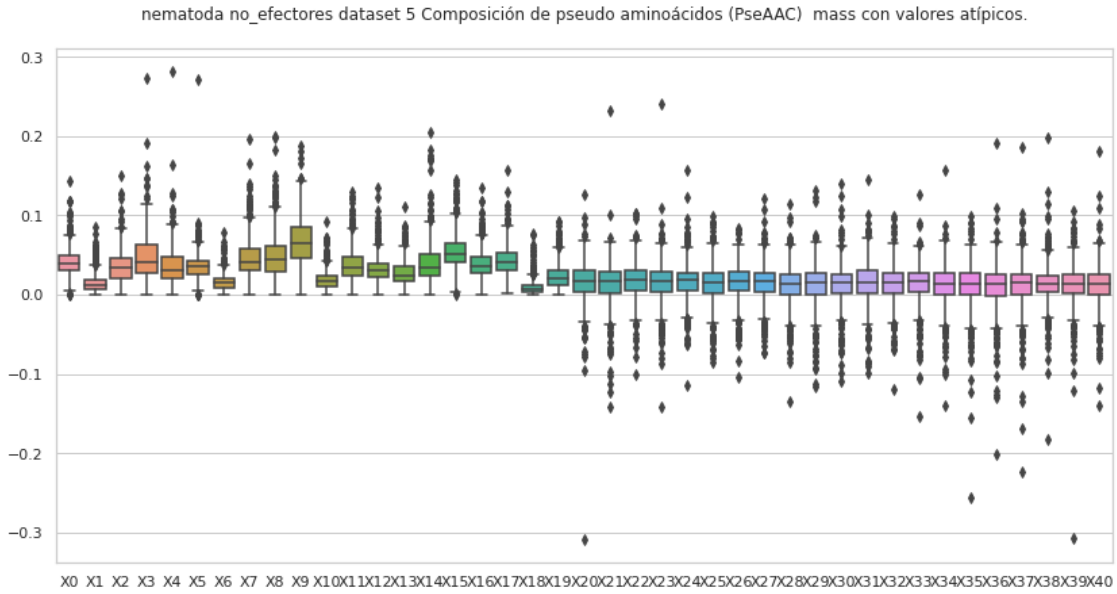
no_efectores
Composición de pseudo aminoácidos (PseAAC) mass no_efectores nematoda dataset 5, con valores atípicos.
Valores del documento csv.

```
          X0        X1        X2        X3        X4        X5        X6  \
0    0.034087  0.009244  0.015599  0.039286  0.037553  0.032354  0.008666
1    0.060294  0.012059  0.012059  0.072353  0.060294  0.024118  0.024118
2    0.058833  0.007977  0.046867  0.050855  0.041881  0.042878  0.017949
3    0.037562  0.000000  0.037562  0.012521  0.062604  0.062604  0.025042
4    0.044568  0.006023  0.030114  0.030114  0.030114  0.036137  0.009636
..        …         …         …         …         …         …         …
495  0.038666  0.008592  0.038666  0.103110  0.017185  0.038666  0.017185
496  0.046368  0.000000  0.081145  0.063756  0.081145  0.086941  0.040572
```

```
497  0.030347  0.002985  0.015422  0.026865  0.013930  0.024875  0.009452
498  0.060428  0.023105  0.039101  0.072869  0.031991  0.023105  0.030214
499  0.010284  0.001674  0.013154  0.010045  0.010284  0.040657  0.001913

           X7        X8        X9  …       X32       X33       X34  \
0    0.050264  0.034087  0.060663  …  0.033242  0.027892  0.014610
1    0.048235  0.072353  0.180882  … -0.035125 -0.076708  0.064014
2    0.038889  0.047864  0.082765  …  0.009177  0.016438  0.010084
3    0.037562  0.050083  0.100166  … -0.061685 -0.021140  0.065605
4    0.045773  0.028909  0.069864  … -0.005735  0.036424  0.030783
..        …         …         …  …        …         …         …
495  0.025777  0.077332  0.115998  … -0.037615  0.020269  0.002863
496  0.069552  0.081145  0.046368  … -0.030577  0.026215 -0.030155
497  0.007960  0.022387  0.026367  …  0.037805  0.018961  0.027774
498  0.031991  0.030214  0.076424  …  0.025087  0.001655 -0.003580
499  0.011001  0.007653  0.013871  …  0.035569  0.033290  0.036844

          X35       X36       X37       X38       X39       X40           X41
0    0.022830  0.021876  0.020100  0.001001  0.002307  0.022969  no_efectores
1    0.029987 -0.008448 -0.036996 -0.051636  0.021005  0.089163  no_efectores
2   -0.001645 -0.001954  0.015648  0.015635  0.017578  0.036452  no_efectores
3    0.037293  0.041721 -0.021586  0.048440  0.079184  0.031457  no_efectores
4   -0.007124  0.033642  0.025515  0.009168  0.011920  0.014721  no_efectores
..        …         …         …         …         …         …           …
495  0.019793  0.067702 -0.002240  0.042851  0.011830  0.003745  no_efectores
496  0.025804 -0.026795  0.052847 -0.033670  0.000769  0.002521  no_efectores
497  0.029559  0.020512  0.016894  0.026050  0.030850  0.032273  no_efectores
498  0.022986  0.022074  0.001724  0.022249 -0.003810 -0.000913  no_efectores
499  0.038016  0.026392  0.036350  0.038521  0.040443  0.041363  no_efectores

[500 rows x 42 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.041154    0.015364    0.036641    0.047784    0.036745    0.036487
std      0.017933    0.013252    0.020887    0.029493    0.024805    0.018015
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.030618    0.006953    0.021395    0.026534    0.019929    0.026273
50%      0.039750    0.012091    0.034676    0.041525    0.031231    0.035120
75%      0.048930    0.019455    0.046589    0.063637    0.048249    0.043079
max      0.142520    0.084750    0.149391    0.272695    0.281403    0.271105

                X6          X7          X8          X9  …         X31  \
```

```
count  500.000000  500.000000  500.000000  500.000000  …  500.000000
mean     0.015877    0.046253    0.049085    0.066851  …    0.014394
std      0.010198    0.025957    0.029834    0.030270  …    0.026006
min      0.000000    0.000000    0.000000    0.000000  …   -0.098799
25%      0.008950    0.030542    0.029374    0.045660  …    0.002389
50%      0.014815    0.041599    0.043704    0.064295  …    0.015476
75%      0.020716    0.057546    0.062157    0.085122  …    0.030053
max      0.077822    0.196814    0.199976    0.186914  …    0.144943

              X32         X33         X34         X35         X36         X37  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.012767    0.013443    0.012138    0.010491    0.009647    0.012079
std      0.024644    0.025934    0.027080    0.029684    0.029364    0.030906
min     -0.120132   -0.153946   -0.140373   -0.255771   -0.201352   -0.223430
25%      0.001871    0.003561    0.000704   -0.000680   -0.001945   -0.000104
50%      0.014976    0.016450    0.013689    0.012966    0.014209    0.014680
75%      0.027613    0.027426    0.027904    0.027314    0.025350    0.026047
max      0.099104    0.126780    0.156327    0.099590    0.190941    0.185872

              X38         X39         X40
count  500.000000  500.000000  500.000000
mean     0.013624    0.013039    0.012215
std      0.026165    0.028071    0.026797
min     -0.182229   -0.307491   -0.139292
25%      0.002989    0.002354   -0.000074
50%      0.014473    0.014460    0.013394
75%      0.024576    0.026213    0.026315
max      0.198449    0.105903    0.180508

[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 5,
sin valores atípicos.
Valores del documento csv.

          X0        X1        X2        X3        X4        X5        X6  \
0   0.042854  0.001531  0.064281  0.097953  0.019897  0.029080  0.026019
1   0.043424  0.000000  0.026054  0.017370  0.043424  0.034739  0.008685
2   0.057470  0.012771  0.038314  0.086206  0.028735  0.028735  0.038314
3   0.041266  0.009379  0.021571  0.034701  0.024384  0.038452  0.009379
```

```
4    0.058240  0.023296  0.052416  0.075713  0.029120  0.023296  0.023296
..        …         …         …         …         …         …
492  0.039765  0.012557  0.031393  0.048136  0.023022  0.039765  0.008372
493  0.058540  0.017562  0.046832  0.046832  0.046832  0.023416  0.017562
494  0.054556  0.011691  0.035072  0.019484  0.019484  0.027278  0.019484
495  0.039465  0.021250  0.063751  0.060715  0.051608  0.051608  0.033393
496  0.026362  0.007532  0.026362  0.022596  0.021341  0.020085  0.012553

           X7        X8        X9  …       X32       X33       X34  \
0    0.045915  0.045915  0.050507  …  0.004751  0.011412  0.006996
1    0.095533  0.121588  0.052109  … -0.000698 -0.005374 -0.021652
2    0.031928  0.038314  0.063856  …  0.023926  0.039063  0.016762
3    0.029074  0.023446  0.050644  …  0.023312  0.018987  0.015387
4    0.017472  0.058240  0.046592  …  0.017682  0.030241  0.003202
..        …         …         …  …         …         …         …
492  0.046043  0.025115  0.062786  …  0.028680  0.000273 -0.009613
493  0.040978  0.052686  0.093664  … -0.008848  0.025749 -0.020098
494  0.046762  0.031175  0.050659  … -0.020809  0.033530  0.021915
495  0.075894  0.078929  0.094108  …  0.008918  0.001590 -0.002503
496  0.025107  0.021341  0.037660  …  0.030363  0.028814  0.025690

           X35       X36       X37       X38       X39       X40         X41
0    0.009581  0.019231  0.018746  0.020288  0.001276  0.023744  efectores
1   -0.005021 -0.005564  0.054367  0.019100  0.011403  0.097515  efectores
2    0.012905 -0.014809 -0.020793 -0.024365 -0.047252  0.013197  efectores
3    0.021540  0.019789  0.013127  0.022838  0.019651  0.009315  efectores
4   -0.035824  0.046260  0.025987  0.029428 -0.040446  0.016574  efectores
..        …         …         …         …         …         …
492  0.033000  0.023453  0.024549  0.023391  0.013852 -0.000651  efectores
493  0.046683  0.008257 -0.008762  0.010201  0.034016  0.027667  efectores
494  0.007575  0.019307  0.021070  0.001630  0.057253 -0.021348  efectores
495  0.022555  0.019919  0.046888  0.005438  0.025055  0.016133  efectores
496  0.026932  0.020172  0.023719  0.012363  0.011125  0.030579  efectores

[407 rows x 42 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass efectores nematoda dataset 5, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  407.000000  407.000000  407.000000  407.000000  407.000000  407.000000
mean     0.044249    0.014004    0.033267    0.042004    0.028530    0.034610
std      0.015328    0.009656    0.016365    0.023190    0.016125    0.013054
min      0.002898    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.034226    0.006787    0.021892    0.025396    0.017187    0.024489
50%      0.043424    0.011963    0.031393    0.038384    0.025537    0.034035
```

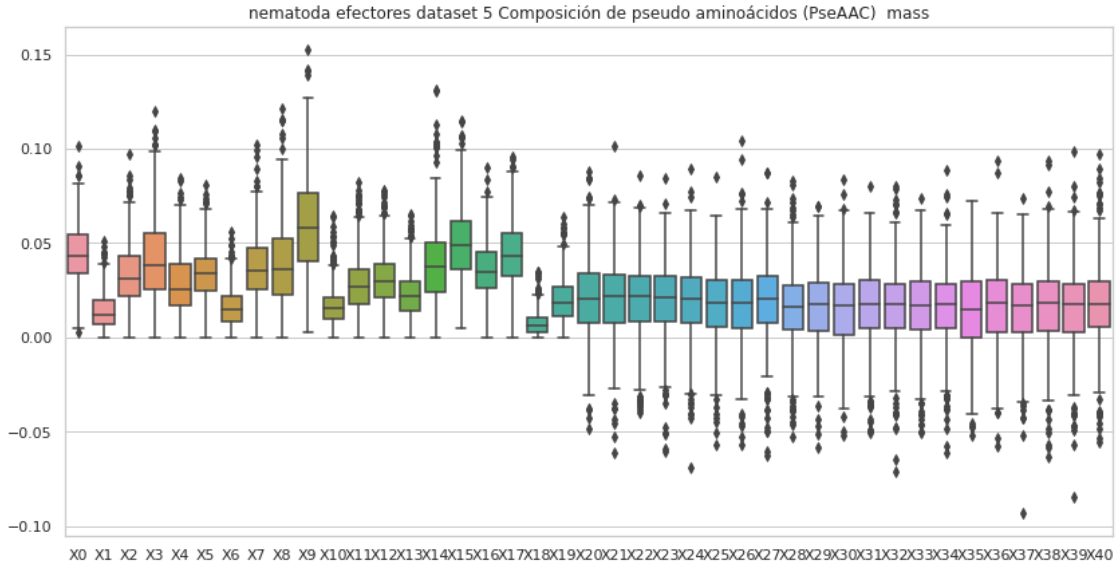
```
75%      0.054550    0.019815    0.043134    0.055010    0.039142    0.042045
max      0.101608    0.050958    0.097120    0.119626    0.084685    0.080499

                X6          X7          X8          X9  …         X31  \
count   407.000000  407.000000  407.000000  407.000000  …  407.000000
mean      0.016306    0.037167    0.039037    0.059950  …    0.016578
std       0.010163    0.017458    0.022685    0.026722  …    0.020899
min       0.000000    0.000000    0.000000    0.002898  …   -0.050879
25%       0.008426    0.025613    0.022559    0.040230  …    0.004938
50%       0.014604    0.035387    0.036293    0.058408  …    0.017820
75%       0.021708    0.047315    0.052095    0.076813  …    0.030458
max       0.055684    0.101861    0.121588    0.152554  …    0.080070

                X32         X33         X34         X35         X36         X37  \
count   407.000000  407.000000  407.000000  407.000000  407.000000  407.000000
mean      0.016058    0.015936    0.015765    0.014114    0.015746    0.015240
std       0.021595    0.020383    0.020795    0.021937    0.021716    0.020440
min      -0.071252   -0.050818   -0.061439   -0.051924   -0.057837   -0.093223
25%       0.004789    0.003927    0.004880   -0.000215    0.002473    0.002969
50%       0.017879    0.016738    0.017504    0.015110    0.018002    0.016966
75%       0.028540    0.029872    0.028439    0.029639    0.030148    0.028076
max       0.080230    0.073843    0.088652    0.072109    0.093664    0.073774

                X38         X39         X40
count   407.000000  407.000000  407.000000
mean      0.016039    0.014810    0.017234
std       0.022533    0.021970    0.022247
min      -0.063043   -0.084467   -0.055403
25%       0.003683    0.002645    0.005285
50%       0.018456    0.017331    0.017430
75%       0.029706    0.028319    0.029778
max       0.093433    0.098311    0.097515

[8 rows x 41 columns]
```

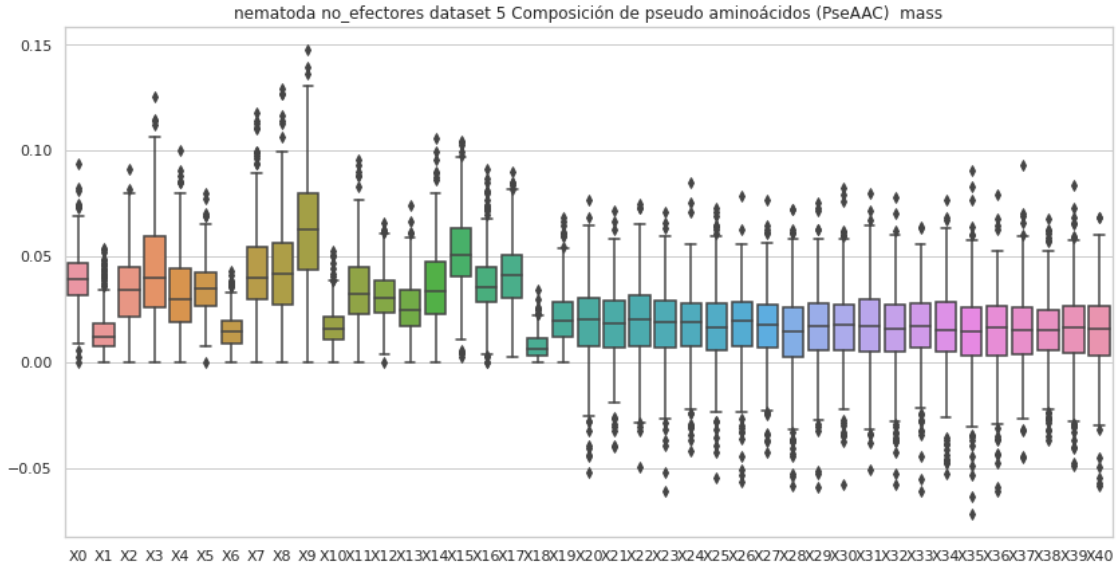
Composición de pseudo aminoácidos (PseAAC) mass no_efectores nematoda dataset 5, sin valores atípicos.
Valores del documento csv.

```
          X0        X1        X2        X3        X4        X5        X6  \
0    0.034087  0.009244  0.015599  0.039286  0.037553  0.032354  0.008666
2    0.058833  0.007977  0.046867  0.050855  0.041881  0.042878  0.017949
4    0.044568  0.006023  0.030114  0.030114  0.030114  0.036137  0.009636
5    0.035801  0.016847  0.071602  0.046330  0.052648  0.052648  0.035801
6    0.042414  0.008483  0.038173  0.038173  0.046656  0.033931  0.008483
..        …         …         …         …         …         …         …
```

```
493  0.042724  0.020345  0.077310  0.087482  0.073241  0.040689  0.022379
494  0.082099  0.000000  0.000902  0.013533  0.003609  0.007218  0.000000
495  0.038666  0.008592  0.038666  0.103110  0.017185  0.038666  0.017185
498  0.060428  0.023105  0.039101  0.072869  0.031991  0.023105  0.030214
499  0.010284  0.001674  0.013154  0.010045  0.010284  0.040657  0.001913

           X7        X8        X9  …       X32       X33       X34  \
0    0.050264  0.034087  0.060663  …  0.033242  0.027892  0.014610
2    0.038889  0.047864  0.082765  …  0.009177  0.016438  0.010084
4    0.045773  0.028909  0.069864  … -0.005735  0.036424  0.030783
5    0.061072  0.069496  0.098979  …  0.009226  0.036449  0.000600
6    0.097553  0.050897  0.110277  …  0.004301 -0.019992 -0.022299
..        …         …         …  …         …         …
493  0.048827  0.054930  0.061034  …  0.030995 -0.014684  0.010474
494  0.009924  0.004511  0.013533  …  0.036128  0.033099  0.041533
495  0.025777  0.077332  0.115998  … -0.037615  0.020269  0.002863
498  0.031991  0.030214  0.076424  …  0.025087  0.001655 -0.003580
499  0.011001  0.007653  0.013871  …  0.035569  0.033290  0.036844

           X35       X36       X37       X38       X39       X40           X41
0    0.022830  0.021876  0.020100  0.001001  0.002307  0.022969  no_efectores
2   -0.001645 -0.001954  0.015648  0.015635  0.017578  0.036452  no_efectores
4   -0.007124  0.033642  0.025515  0.009168  0.011920  0.014721  no_efectores
5   -0.003312  0.023705  0.003114  0.008023 -0.011753 -0.003835  no_efectores
6    0.021234  0.030192  0.009077  0.010811  0.007493 -0.004241  no_efectores
..        …         …         …         …         …         …
493  0.021831 -0.017970 -0.022973  0.017713 -0.017367  0.008927  no_efectores
494  0.039139  0.021038  0.065833  0.020707  0.048901  0.014625  no_efectores
495  0.019793  0.067702 -0.002240  0.042851  0.011830  0.003745  no_efectores
498  0.022986  0.022074  0.001724  0.022249 -0.003810 -0.000913  no_efectores
499  0.038016  0.026392  0.036350  0.038521  0.040443  0.041363  no_efectores

[399 rows x 42 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  399.000000  399.000000  399.000000  399.000000  399.000000  399.000000
mean     0.039502    0.013866    0.034530    0.043822    0.032791    0.035208
std      0.013458    0.010161    0.016865    0.023852    0.018477    0.012093
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.031321    0.007214    0.021618    0.025627    0.018853    0.026725
50%      0.039055    0.011803    0.033806    0.039586    0.029750    0.034916
75%      0.046715    0.018019    0.044878    0.059131    0.044401    0.042190
max      0.093532    0.053456    0.091390    0.125348    0.099808    0.079974
```

```
              X6          X7          X8          X9  …         X31  \
count  399.000000  399.000000  399.000000  399.000000  …  399.000000
mean     0.014845    0.042599    0.043952    0.062547  …    0.016604
std      0.007948    0.020994    0.023253    0.026792  …    0.018582
min      0.000000    0.000000    0.000000    0.000000  …   -0.051039
25%      0.008890    0.029509    0.026921    0.043660  …    0.004989
50%      0.014548    0.039497    0.041541    0.062390  …    0.016715
75%      0.019286    0.054178    0.056341    0.079773  …    0.029675
max      0.042873    0.117806    0.128890    0.147480  …    0.079981

              X32         X33         X34         X35         X36         X37  \
count  399.000000  399.000000  399.000000  399.000000  399.000000  399.000000
mean     0.014996    0.016255    0.015365    0.013791    0.013950    0.015152
std      0.018901    0.017736    0.020237    0.020282    0.018807    0.018703
min     -0.057830   -0.061174   -0.052749   -0.071849   -0.060825   -0.045474
25%      0.004949    0.006840    0.004593    0.002920    0.003281    0.003592
50%      0.015795    0.017139    0.014779    0.014183    0.016522    0.015144
75%      0.027101    0.027449    0.028162    0.025583    0.026114    0.025966
max      0.077534    0.063746    0.076349    0.090463    0.078778    0.092977

              X38         X39         X40
count  399.000000  399.000000  399.000000
mean     0.014246    0.015482    0.014702
std      0.017049    0.018808    0.019148
min     -0.036732   -0.048787   -0.058297
25%      0.005230    0.004491    0.003203
50%      0.014851    0.016192    0.015537
75%      0.024216    0.026237    0.026317
max      0.067513    0.083266    0.068459

[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 5,
con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.028246  0.001009  0.042369  0.064562  0.013114  0.019167  0.017149
1    0.023616  0.000000  0.014170  0.009446  0.023616  0.018893  0.004723
2    0.037862  0.008414  0.025242  0.056793  0.018931  0.018931  0.025242
3    0.051598  0.011727  0.026972  0.043389  0.030490  0.048080  0.011727
4    0.021969  0.008788  0.019772  0.028560  0.010984  0.008788  0.008788
..        …         …         …         …         …         …         …
495  0.035923  0.019343  0.058029  0.055266  0.046976  0.046976  0.030396
496  0.064162  0.018332  0.064162  0.054996  0.051940  0.048885  0.030553
497  0.149609  0.000000  0.059844  0.089766  0.149609  0.029922  0.029922
498  0.030924  0.015462  0.036078  0.025770  0.030924  0.046386  0.005154
499  0.091442  0.030481  0.073154  0.097538  0.048769  0.042673  0.079250

           X7        X8        X9  …       X53       X54       X55  \
0    0.030264  0.030264  0.033290  …  0.019214  0.003962  0.039135
1    0.051955  0.066125  0.028339  …  0.089448  0.017463  0.020136
```

```
2    0.021035  0.025242  0.042069  …  0.017439 -0.001500  0.033658
3    0.036353  0.029317  0.063324  … -0.007326  0.014628  0.021260
4    0.006591  0.021969  0.017575  …  0.033898  0.053083  0.036214
..        …         …         …  …         …         …
495  0.069083  0.071846  0.085662  …  0.027614  0.002679 -0.016956
496  0.061106  0.051940  0.091659  …  0.012061  0.049604  0.035023
497  0.119687  0.089766  0.149609  …  0.008743 -0.136389 -0.017790
498  0.030924  0.036078  0.072156  …  0.006274  0.024128 -0.009363
499  0.060961  0.073154  0.097538  … -0.002631 -0.029568  0.012890

          X56       X57       X58       X59       X60       X61         X62
0    0.000006  0.041201 -0.005034 -0.008699 -0.005006  0.025962  efectores
1   -0.004667 -0.025540  0.027500  0.048359  0.005277  0.006284  efectores
2    0.048211  0.027163  0.019533  0.037228 -0.011303 -0.020256  efectores
3    0.025194  0.021316 -0.011057 -0.004406 -0.023410 -0.009214  efectores
4    0.040126  0.027313  0.015974  0.029224  0.011408  0.017381  efectores
..        …         …         …         …         …         …         …
495 -0.049289 -0.033420 -0.003117  0.019066 -0.002512 -0.004289  efectores
496  0.024909  0.014268  0.031412  0.029914 -0.072805 -0.071504  efectores
497 -0.222342 -0.187674  0.438396  0.179391 -0.240035 -0.248005  efectores
498 -0.043820  0.007791  0.059575 -0.014543 -0.009545 -0.006285  efectores
499  0.033162  0.052854 -0.032257  0.033079 -0.068790 -0.021271  efectores

[500 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro efectores nematoda dataset 5, con valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.033622    0.015578    0.037769    0.040021    0.010696    0.033002
std      0.411313    0.042218    0.023588    0.140493    0.425634    0.160977
min     -9.115046   -0.847887   -0.000000   -3.038349   -9.115046   -3.038349
25%      0.028239    0.006370    0.019553    0.025917    0.016637    0.020203
50%      0.045142    0.012964    0.035615    0.043744    0.028133    0.035017
75%      0.067208    0.022923    0.049786    0.060969    0.044336    0.055928
max      0.259363    0.154576    0.162461    0.181855    0.204211    0.264706

                X6          X7          X8          X9  …         X52  \
count  500.000000  500.000000  500.000000  500.000000  …  500.000000
mean     0.018382    0.034484    0.036309    0.050296  …    0.081621
std      0.042639    0.161301    0.146905    0.318249  …    1.782675
min     -0.847887   -3.038349   -3.038349   -6.076697  …   -0.757439
25%      0.008187    0.023646    0.023079    0.038868  …   -0.012453
50%      0.015353    0.037381    0.039321    0.061728  …    0.007032
75%      0.027729    0.054861    0.054965    0.087593  …    0.022304
```

```
max      0.109113    0.278236    0.382972    0.406153  …   39.841258

              X53         X54         X55         X56         X57         X58  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.061697   -0.007981   -0.008576    0.006991    0.017342    0.028396
std      1.110787    0.367952    0.441393    0.481228    0.271427    1.034503
min     -0.198843   -8.014580   -9.737843   -6.406940   -1.576987   -8.043441
25%     -0.003875   -0.012139   -0.007104   -0.015708   -0.005495   -0.015821
50%      0.012477    0.007307    0.011826    0.005436    0.011222    0.002150
75%      0.028703    0.022834    0.024833    0.022195    0.025287    0.021218
max     24.835583    1.321644    1.236402    8.543364    5.800421   21.640638

              X59         X60         X61
count  500.000000  500.000000  500.000000
mean     0.023112    0.009563    0.015456
std      0.515579    0.867227    0.334371
min     -3.513393  -10.817509   -2.265943
25%     -0.005828   -0.019068   -0.007338
50%      0.010767    0.003423    0.008718
75%      0.028284    0.019681    0.027002
max     10.930754   16.032078    7.048629

[8 rows x 62 columns]
```

no_efectores
Composición de pseudo aminoácidos (PseAAC) hidro no_efectores nematoda dataset 5, con valores atípicos.
Valores del documento csv.

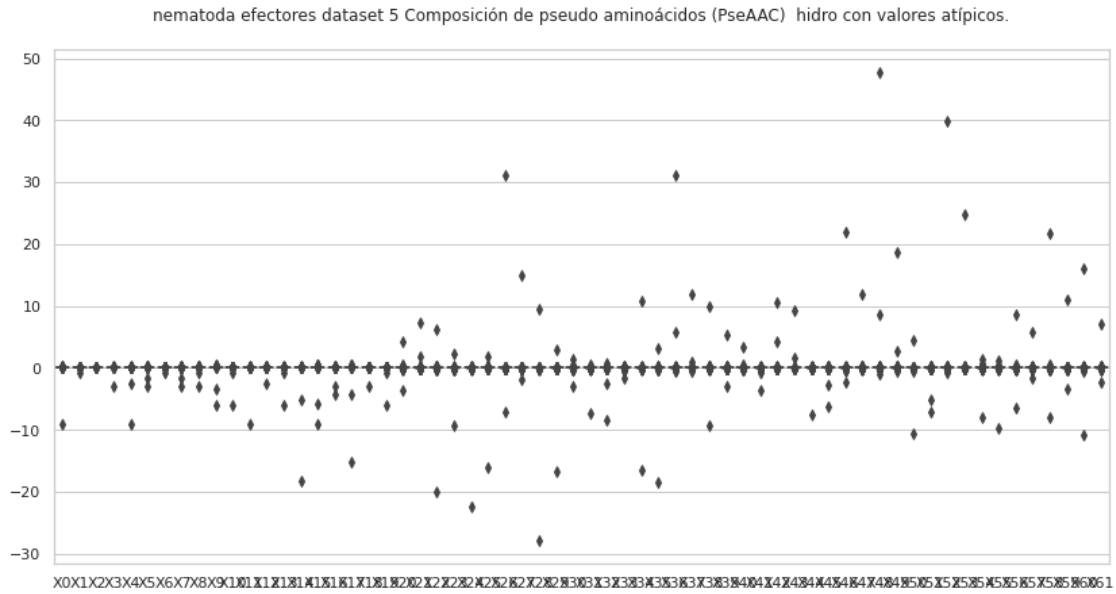
```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.024603  0.006672  0.011259  0.028357  0.027106  0.023352  0.006255
1    0.028204  0.005641  0.005641  0.033845  0.028204  0.011282  0.011282
2    0.079953  0.010841  0.063691  0.069112  0.056915  0.058271  0.024392
3    0.019257  0.000000  0.019257  0.006419  0.032095  0.032095  0.012838
4    0.036268  0.004901  0.024505  0.024505  0.024505  0.029407  0.007842
..        …         …         …         …         …         …         …
495  0.024632  0.005474  0.024632  0.065684  0.010947  0.024632  0.010947
496  0.033873  0.000000  0.059277  0.046575  0.059277  0.063511  0.029639
497  0.029656  0.002917  0.015071  0.026253  0.013612  0.024308  0.009237
498  0.042582  0.016281  0.027553  0.051349  0.022544  0.016281  0.021291
499  0.043304  0.007050  0.055389  0.042297  0.043304  0.171204  0.008057

           X7        X8        X9  …       X53       X54       X55  \
0    0.036280  0.024603  0.043786  …  0.008544 -0.000818  0.005595
1    0.022563  0.033845  0.084612  … -0.020179 -0.049361 -0.060210
2    0.052850  0.065046  0.112476  …  0.019586  0.031022  0.029725
```

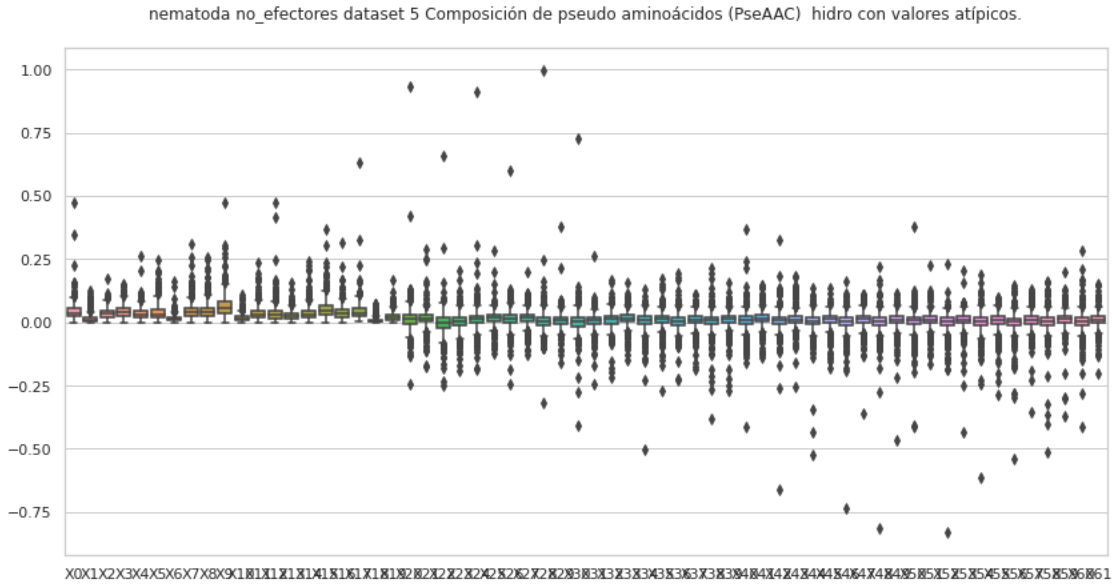
```
3    0.019257  0.025676  0.051352  … -0.002579  0.048420  0.004767
4    0.037248  0.023525  0.056853  … -0.024668  0.005258 -0.006460
..        …         …         …  …         …         …
495  0.016421  0.049263  0.073895  …  0.026625  0.021004  0.041110
496  0.050809  0.059277  0.033873  …  0.031043  0.014763  0.048113
497  0.007779  0.021877  0.025767  …  0.020487  0.012915  0.026244
498  0.022544  0.021291  0.053854  …  0.017219  0.035451  0.053887
499  0.046326  0.032227  0.058411  …  0.029181  0.009234 -0.011729

          X56       X57       X58       X59       X60       X61           X62
0    0.011801  0.006939  0.019512  0.017652  0.010736  0.014153  no_efectores
1    0.032443  0.009293  0.050420 -0.007433  0.067315  0.020590  no_efectores
2    0.003553  0.000769 -0.002627 -0.031336 -0.007357  0.004350  no_efectores
3    0.070392  0.034446  0.107486  0.045743 -0.038249 -0.000795  no_efectores
4    0.006507  0.000334  0.024377  0.006590  0.004058 -0.010009  no_efectores
..        …         …         …         …         …         …           …
495  0.002932  0.012520 -0.045445  0.006295 -0.005427  0.038717  no_efectores
496  0.002459 -0.000941  0.005468  0.023011  0.023666  0.025062  no_efectores
497  0.017382  0.032865  0.001063  0.017846  0.000449  0.013847  no_efectores
498  0.035399  0.020930  0.004177  0.010667 -0.012178  0.021450  no_efectores
499  0.017776  0.022708 -0.004438  0.010386 -0.017693 -0.007228  no_efectores

[500 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro no_efectores nematoda dataset 5, con valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.042859    0.016118    0.034986    0.043149    0.034416    0.040072
std      0.035701    0.017524    0.022077    0.025025    0.025319    0.032208
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.022790    0.005448    0.018082    0.024547    0.017961    0.018179
50%      0.035767    0.010918    0.032912    0.040247    0.030544    0.031461
75%      0.054159    0.020612    0.048329    0.056186    0.045012    0.053209
max      0.472310    0.125819    0.174058    0.150956    0.262395    0.247339

                X6          X7          X8          X9  …         X52  \
count  500.000000  500.000000  500.000000  500.000000  …  500.000000
mean     0.016003    0.044685    0.045960    0.065828  …   -0.000659
std      0.015046    0.032577    0.033025    0.045391  …    0.050104
min      0.000000    0.000000    0.000000    0.000000  …   -0.829902
25%      0.007003    0.024060    0.023737    0.037145  …   -0.009657
50%      0.012002    0.039263    0.039237    0.057019  …    0.005428
75%      0.020437    0.054628    0.058718    0.083400  …    0.017481
max      0.162930    0.310232    0.259186    0.472310  …    0.231255
```

```
              X53         X54         X55         X56         X57         X58  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.007936   -0.000209    0.005458   -0.002667    0.005770    0.001835
std      0.037805    0.047783    0.036540    0.047996    0.036092    0.050758
min     -0.433840   -0.616013   -0.285351   -0.541490   -0.355445   -0.513623
25%     -0.002529   -0.014344   -0.005306   -0.012682   -0.005531   -0.010244
50%      0.010957    0.005445    0.007509    0.004115    0.008597    0.005376
75%      0.023829    0.019573    0.023985    0.015987    0.022651    0.019504
max      0.203715    0.190650    0.100735    0.148227    0.130695    0.162008

              X59         X60         X61
count  500.000000  500.000000  500.000000
mean     0.009197    0.002755    0.010275
std      0.039829    0.044785    0.032268
min     -0.372307   -0.412916   -0.204550
25%     -0.001984   -0.010321   -0.002777
50%      0.012309    0.005621    0.011172
75%      0.024486    0.019009    0.024773
max      0.200101    0.284196    0.151621

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.028246  0.001009  0.042369  0.064562  0.013114  0.019167  0.017149
1    0.023616  0.000000  0.014170  0.009446  0.023616  0.018893  0.004723
2    0.037862  0.008414  0.025242  0.056793  0.018931  0.018931  0.025242
3    0.051598  0.011727  0.026972  0.043389  0.030490  0.048080  0.011727
4    0.021969  0.008788  0.019772  0.028560  0.010984  0.008788  0.008788
..        …         …         …         …         …         …         …
495  0.035923  0.019343  0.058029  0.055266  0.046976  0.046976  0.030396
496  0.064162  0.018332  0.064162  0.054996  0.051940  0.048885  0.030553
497  0.149609  0.000000  0.059844  0.089766  0.149609  0.029922  0.029922
498  0.030924  0.015462  0.036078  0.025770  0.030924  0.046386  0.005154
499  0.091442  0.030481  0.073154  0.097538  0.048769  0.042673  0.079250

           X7        X8        X9  …       X53       X54       X55  \
0    0.030264  0.030264  0.033290  …  0.019214  0.003962  0.039135
1    0.051955  0.066125  0.028339  …  0.089448  0.017463  0.020136
2    0.021035  0.025242  0.042069  …  0.017439 -0.001500  0.033658
3    0.036353  0.029317  0.063324  … -0.007326  0.014628  0.021260
4    0.006591  0.021969  0.017575  …  0.033898  0.053083  0.036214
```

```
..        …       …       …  …       …       …       …
495  0.069083  0.071846  0.085662  …  0.027614  0.002679 -0.016956
496  0.061106  0.051940  0.091659  …  0.012061  0.049604  0.035023
497  0.119687  0.089766  0.149609  …  0.008743 -0.136389 -0.017790
498  0.030924  0.036078  0.072156  …  0.006274  0.024128 -0.009363
499  0.060961  0.073154  0.097538  … -0.002631 -0.029568  0.012890

          X56       X57       X58       X59       X60       X61        X62
0    0.000006  0.041201 -0.005034 -0.008699 -0.005006  0.025962  efectores
1   -0.004667 -0.025540  0.027500  0.048359  0.005277  0.006284  efectores
2    0.048211  0.027163  0.019533  0.037228 -0.011303 -0.020256  efectores
3    0.025194  0.021316 -0.011057 -0.004406 -0.023410 -0.009214  efectores
4    0.040126  0.027313  0.015974  0.029224  0.011408  0.017381  efectores
..        …       …       …       …       …       …       …
495 -0.049289 -0.033420 -0.003117  0.019066 -0.002512 -0.004289  efectores
496  0.024909  0.014268  0.031412  0.029914 -0.072805 -0.071504  efectores
497 -0.222342 -0.187674  0.438396  0.179391 -0.240035 -0.248005  efectores
498 -0.043820  0.007791  0.059575 -0.014543 -0.009545 -0.006285  efectores
499  0.033162  0.052854 -0.032257  0.033079 -0.068790 -0.021271  efectores

[488 rows x 63 columns]


Composición de pseudo aminoácidos (PseAAC)  efectores nematoda dataset 5, sin
valores atípicos.
Estadísticas.

               X0          X1          X2          X3          X4          X5  \
count  488.000000  488.000000  488.000000  488.000000  488.000000  488.000000
mean     0.050400    0.016435    0.036386    0.045263    0.032974    0.042135
std      0.030638    0.014924    0.020302    0.024483    0.023006    0.030223
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.028202    0.006291    0.019553    0.026159    0.016637    0.020203
50%      0.044328    0.012776    0.035450    0.043539    0.028082    0.034598
75%      0.065814    0.022509    0.049275    0.060431    0.043656    0.055508
max      0.259363    0.138533    0.106564    0.144116    0.166193    0.264706

               X6          X7          X8          X9  …         X52  \
count  488.000000  488.000000  488.000000  488.000000  …  488.000000
mean     0.019362    0.041960    0.042017    0.067260  …    0.004783
std      0.016507    0.026670    0.024723    0.039179  …    0.041795
min      0.000000    0.000000    0.000000    0.001628  …   -0.206935
25%      0.008175    0.023287    0.023029    0.038697  …   -0.011554
50%      0.015148    0.036834    0.039018    0.061302  …    0.007197
75%      0.026809    0.054050    0.054064    0.086232  …    0.022244
max      0.108697    0.165332    0.153702    0.295453  …    0.201040

              X53         X54         X55         X56         X57         X58  \
```

count	488.000000	488.000000	488.000000	488.000000	488.000000	488.000000
mean	0.012546	0.005662	0.008634	0.002753	0.009842	0.004087
std	0.035436	0.039599	0.031406	0.042507	0.034068	0.046103
min	-0.131931	-0.229994	-0.199047	-0.231021	-0.187674	-0.180896
25%	-0.003354	-0.011129	-0.006275	-0.014783	-0.004578	-0.014415
50%	0.012533	0.007572	0.012163	0.005570	0.011422	0.002955
75%	0.028290	0.022755	0.024833	0.022026	0.025262	0.021267
max	0.210072	0.199245	0.118061	0.174190	0.163663	0.438396

	X59	X60	X61
count	488.000000	488.000000	488.000000
mean	0.010549	0.000702	0.007010
std	0.035279	0.043570	0.038722
min	-0.129876	-0.240035	-0.248005
25%	-0.005012	-0.018431	-0.007094
50%	0.010897	0.003573	0.008855
75%	0.028048	0.019681	0.026933
max	0.179391	0.184391	0.120989

[8 rows x 62 columns]

no_efectores
Composición de pseudo aminoácidos (PseAAC) no_efectores nematoda dataset 5, sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.024603	0.006672	0.011259	0.028357	0.027106	0.023352	0.006255	
1	0.028204	0.005641	0.005641	0.033845	0.028204	0.011282	0.011282	
2	0.079953	0.010841	0.063691	0.069112	0.056915	0.058271	0.024392	
3	0.019257	0.000000	0.019257	0.006419	0.032095	0.032095	0.012838	
4	0.036268	0.004901	0.024505	0.024505	0.024505	0.029407	0.007842	
..	…	…	…	…	…	…	…	
493	0.018627	0.008870	0.033707	0.038142	0.031932	0.017740	0.009757	
495	0.024632	0.005474	0.024632	0.065684	0.010947	0.024632	0.010947	
496	0.033873	0.000000	0.059277	0.046575	0.059277	0.063511	0.029639	
497	0.029656	0.002917	0.015071	0.026253	0.013612	0.024308	0.009237	
498	0.042582	0.016281	0.027553	0.051349	0.022544	0.016281	0.021291	

	X7	X8	X9	…	X53	X54	X55	\
0	0.036280	0.024603	0.043786	…	0.008544	-0.000818	0.005595	
1	0.022563	0.033845	0.084612	…	-0.020179	-0.049361	-0.060210	
2	0.052850	0.065046	0.112476	…	0.019586	0.031022	0.029725	
3	0.019257	0.025676	0.051352	…	-0.002579	0.048420	0.004767	
4	0.037248	0.023525	0.056853	…	-0.024668	0.005258	-0.006460	
..	…	…	…	…	…	…	…	

```
493  0.021288  0.023949  0.026610  …  0.010851 -0.008834  0.013584
495  0.016421  0.049263  0.073895  …  0.026625  0.021004  0.041110
496  0.050809  0.059277  0.033873  …  0.031043  0.014763  0.048113
497  0.007779  0.021877  0.025767  …  0.020487  0.012915  0.026244
498  0.022544  0.021291  0.053854  …  0.017219  0.035451  0.053887

          X56       X57       X58       X59       X60       X61           X62
0    0.011801  0.006939  0.019512  0.017652  0.010736  0.014153  no_efectores
1    0.032443  0.009293  0.050420 -0.007433  0.067315  0.020590  no_efectores
2    0.003553  0.000769 -0.002627 -0.031336 -0.007357  0.004350  no_efectores
3    0.070392  0.034446  0.107486  0.045743 -0.038249 -0.000795  no_efectores
4    0.006507  0.000334  0.024377  0.006590  0.004058 -0.010009  no_efectores
..        …         …         …         …         …         …           …
493  0.015037  0.016600 -0.003436  0.017792  0.017964  0.034791  no_efectores
495  0.002932  0.012520 -0.045445  0.006295 -0.005427  0.038717  no_efectores
496  0.002459 -0.000941  0.005468  0.023011  0.023666  0.025062  no_efectores
497  0.017382  0.032865  0.001063  0.017846  0.000449  0.013847  no_efectores
498  0.035399  0.020930  0.004177  0.010667 -0.012178  0.021450  no_efectores

[408 rows x 63 columns]
```

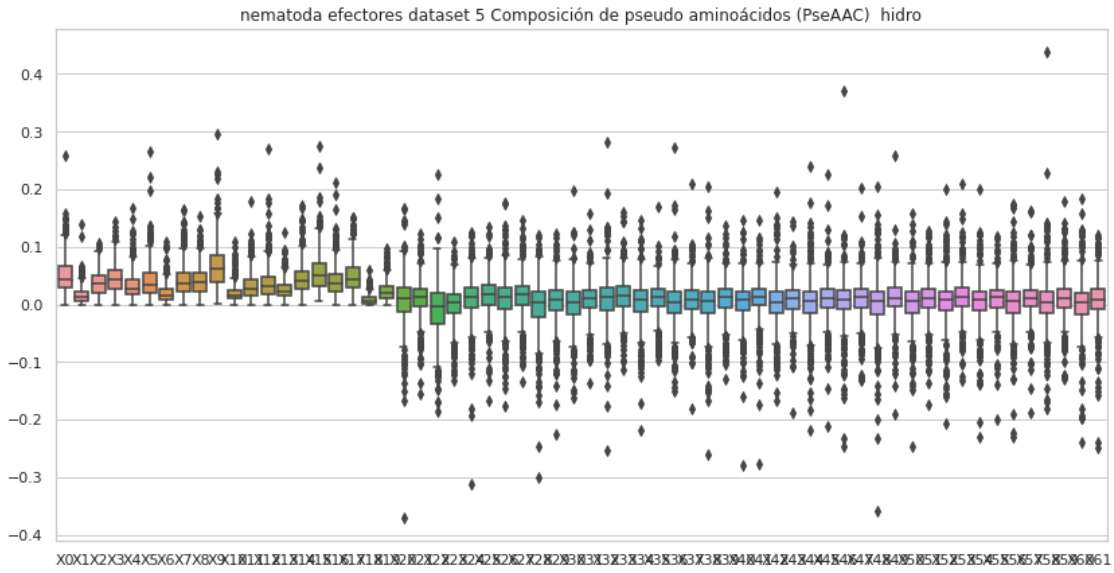
Composición de pseudo aminoácidos (PseAAC) no_efectores nematoda dataset 5, sin valores atípicos.
Estadísticas.

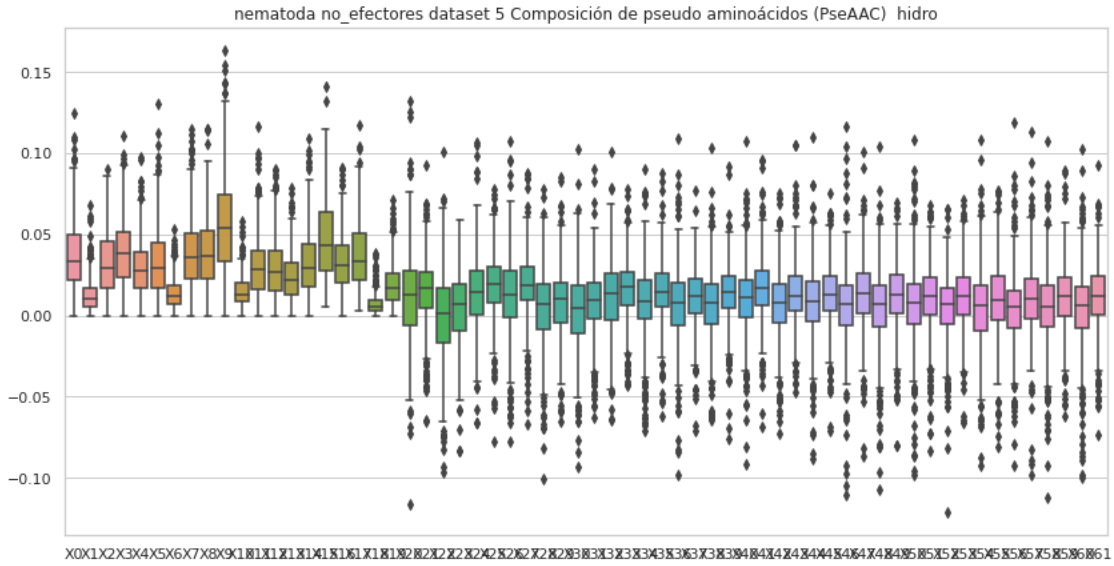
```
                X0          X1          X2          X3          X4          X5  \
count  408.000000  408.000000  408.000000  408.000000  408.000000  408.000000
mean     0.037111    0.012638    0.031650    0.038696    0.029653    0.033496
std      0.020915    0.010502    0.018784    0.020113    0.017449    0.021958
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.021898    0.005289    0.016553    0.023529    0.017214    0.016830
50%      0.033014    0.010084    0.029229    0.038084    0.027255    0.029052
75%      0.049597    0.017277    0.045373    0.051501    0.039186    0.045152
max      0.124410    0.067657    0.090214    0.111033    0.097154    0.130499

                X6          X7          X8          X9  …         X52  \
count  408.000000  408.000000  408.000000  408.000000  …  408.000000
mean     0.013890    0.038009    0.039274    0.056639  …    0.004180
std      0.009446    0.021390    0.021551    0.030282  …    0.021235
min      0.000000    0.000000    0.000000    0.000000  …   -0.121324
25%      0.006794    0.022493    0.022469    0.033388  …   -0.005203
50%      0.011540    0.035992    0.036496    0.053822  …    0.006898
75%      0.018719    0.050275    0.052048    0.074729  …    0.017031
max      0.052974    0.114937    0.114436    0.163242  …    0.065703

              X53         X54         X55         X56         X57         X58  \
count  408.000000  408.000000  408.000000  408.000000  408.000000  408.000000
```

mean	0.011439	0.003850	0.009725	0.003782	0.009436	0.004435
std	0.019824	0.025907	0.023448	0.022836	0.022951	0.025374
min	-0.066207	-0.092403	-0.080999	-0.092464	-0.098132	-0.111940
25%	0.000101	-0.009758	-0.002662	-0.007917	-0.002353	-0.006777
50%	0.011808	0.006122	0.009082	0.005647	0.010495	0.005698
75%	0.023735	0.018756	0.023985	0.015466	0.022919	0.018439
max	0.070931	0.108582	0.076548	0.119203	0.112821	0.107486

	X59	X60	X61
count	408.000000	408.000000	408.000000
mean	0.010807	0.004342	0.011601
std	0.020800	0.025760	0.021500
min	-0.061789	-0.100091	-0.073480
25%	-0.000167	-0.007437	0.000193
50%	0.012348	0.006487	0.012011
75%	0.023382	0.017913	0.024199
max	0.088957	0.102042	0.092792

[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 5,
con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.004622  0.021521 -0.007020 -0.002234  0.014137 -0.006020 -0.063299
1    0.065861  0.099888  0.115225  0.071350 -0.014651 -0.175079  0.083323
2   -0.011966  0.054191  0.019246 -0.022138  0.053399 -0.059781  0.072857
3    0.082915  0.026638  0.073253  0.028404  0.004208  0.046544  0.045879
4   -0.000618 -0.056232 -0.076153  0.077672  0.068254 -0.020627 -0.083698
..        …       …       …       …       …       …       …
495 -0.021948  0.032248 -0.010003  0.022332  0.007417  0.007493 -0.037968
496  0.057319  0.087196  0.100188  0.022304  0.052471  0.064041  0.059096
497 -0.043424 -0.023200 -0.094913 -0.050880  0.111304  0.026114  0.155105
498 -0.181910  0.133994 -0.092320 -0.022772  0.033055  0.088942  0.006578
499 -0.017375 -0.016556 -0.021813  0.070237  0.035177  0.009242 -0.039807

           X7        X8        X9       X10       X11       X12        X13
0    0.025683  0.001941 -0.013940  0.034448 -0.012567 -0.025834  efectores
1    0.015955 -0.038148  0.015743  0.053098 -0.063528 -0.007009  efectores
2   -0.046511 -0.054151 -0.033940  0.015368  0.030925  0.034573  efectores
3   -0.003023  0.079197  0.033078  0.054154  0.027731  0.023326  efectores
4    0.052851  0.150041 -0.078806 -0.070721  0.066841  0.014354  efectores
..        …       …       …       …       …       …        …
495 -0.018370 -0.073003 -0.025523 -0.041398 -0.008318  0.012562  efectores
496  0.082631 -0.002345  0.065305  0.094917  0.071823  0.068536  efectores
497  0.021170 -0.085518 -0.055961  0.050482 -0.048699  0.121860  efectores
498  0.103463  0.023689  0.109609 -0.109729  0.141867 -0.117117  efectores
499 -0.047536 -0.067483 -0.050028 -0.040587 -0.053042  0.048112  efectores

[500 rows x 14 columns]


Covarianza de auto cruzamiento (ACC)  hidro_mass efectores nematoda dataset 5,
con valores atípicos.
Estadísticas.

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

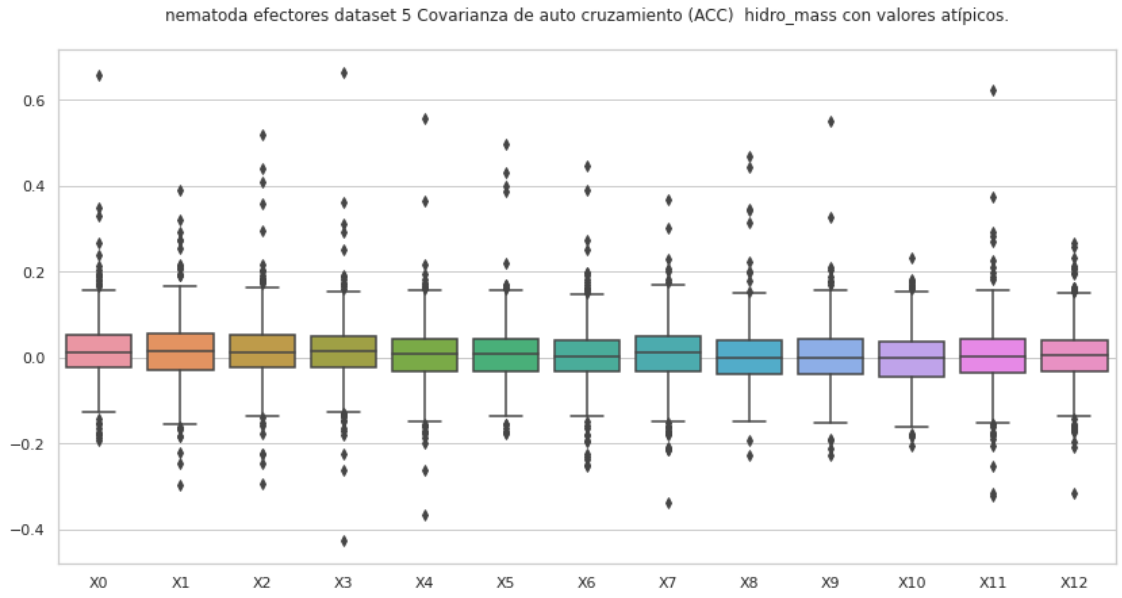
```
mean     0.017079    0.015079    0.014660    0.014244    0.007627    0.007986
std      0.077709    0.076643    0.076686    0.076867    0.073361    0.072181
min     -0.193093   -0.297706   -0.294531   -0.425990   -0.366646   -0.175681
25%     -0.022697   -0.029703   -0.024420   -0.021688   -0.033560   -0.033309
50%      0.012837    0.014637    0.010353    0.013941    0.008556    0.007416
75%      0.052196    0.054433    0.051740    0.049302    0.043024    0.042872
max      0.657781    0.388844    0.517513    0.663383    0.556382    0.495386

                X6          X7          X8          X9         X10         X11  \
count   500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean      0.006570    0.007250    0.003480    0.004524   -0.003810    0.005204
std       0.074427    0.073578    0.072207    0.071697    0.067188    0.080205
min      -0.253302   -0.337294   -0.227673   -0.228443   -0.206742   -0.322807
25%      -0.032416   -0.032452   -0.037876   -0.038436   -0.043634   -0.034667
50%       0.001712    0.010392    0.000625    0.000478   -0.000495    0.003796
75%       0.039729    0.047985    0.038866    0.042653    0.036636    0.044210
max       0.447357    0.367702    0.468705    0.550472    0.233571    0.623797

               X12
count   500.000000
mean      0.004559
std       0.070788
min      -0.314624
25%      -0.032194
50%       0.005032
75%       0.041439
max       0.265259
```

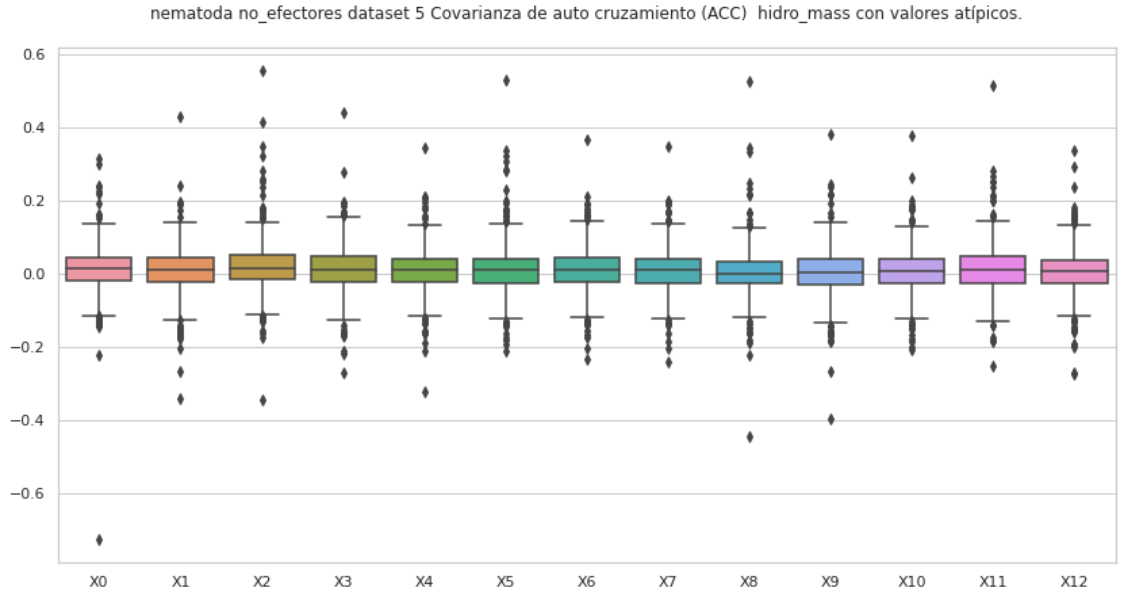
no_efectores
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores nematoda dataset 5, con valores atípicos.
Valores del documento csv.

```
            X0        X1        X2        X3        X4        X5        X6  \
0    0.041656 -0.008989  0.054909  0.067907  0.002478  0.062140  0.017208
1   -0.018267  0.010369 -0.010773  0.033851  0.024248  0.063632 -0.024304
2   -0.004069  0.022637 -0.013418 -0.007399  0.064431 -0.002769  0.027306
3    0.079650  0.028390 -0.049272  0.153428  0.158789 -0.034182 -0.120187
4    0.012218 -0.016649  0.009331  0.026244  0.019812  0.045782  0.056438
..        …         …         …         …         …         …         …
495  0.019198  0.021302  0.027451 -0.067075  0.020520  0.012482  0.014099
496 -0.069403 -0.069587 -0.128997  0.027374  0.058314 -0.079835 -0.066858
497  0.043453 -0.007622 -0.008797 -0.001983  0.020591  0.035658 -0.004292
498 -0.042217  0.022555 -0.029549  0.008094  0.024395  0.045304 -0.058010
499  0.297554  0.194138  0.169708  0.163544  0.133555  0.153411  0.208631
```

```
           X7        X8        X9       X10       X11       X12           X13
0    0.026875  0.016499  0.042661 -0.015428 -0.025516  0.062789  no_efectores
1   -0.033676  0.011571 -0.011409 -0.053864  0.036514 -0.071381  no_efectores
2   -0.080980 -0.025799 -0.025649  0.036599  0.031816 -0.000779  no_efectores
3   -0.186278 -0.064429  0.030008 -0.056427 -0.115960 -0.160156  no_efectores
4   -0.017654 -0.009760  0.021604  0.089228 -0.014257 -0.065658  no_efectores
..        …         …         …         …         …         …           …
495 -0.012399 -0.015832 -0.050836  0.033870 -0.028486 -0.109200  no_efectores
496  0.017824  0.032227 -0.028218 -0.009270  0.109764 -0.076734  no_efectores
497  0.046409  0.016984 -0.004860  0.058795  0.056238  0.069334  no_efectores
498 -0.044435 -0.072076  0.025428  0.020339  0.062836  0.037676  no_efectores
499  0.194794  0.164737  0.189940  0.140611  0.213508  0.143059  no_efectores

[500 rows x 14 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.011645    0.007081    0.017769    0.010986    0.008891    0.008855
std      0.069862    0.065532    0.071549    0.065249    0.062796    0.071274
min     -0.727407   -0.341683   -0.344162   -0.271959   -0.323499   -0.212572
25%     -0.020316   -0.025469   -0.017270   -0.025442   -0.023439   -0.028337
50%      0.012047    0.008022    0.014048    0.009803    0.010194    0.008330
75%      0.044302    0.042340    0.048669    0.046496    0.039512    0.037601
max      0.315182    0.426943    0.553871    0.439271    0.342961    0.529563

                X6          X7          X8          X9         X10         X11  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.008983    0.008416    0.003308    0.001250    0.004995    0.012056
std      0.060803    0.061062    0.070075    0.067276    0.061139    0.067959
min     -0.236000   -0.242599   -0.445889   -0.397801   -0.209637   -0.253624
25%     -0.023567   -0.026978   -0.028092   -0.032824   -0.028222   -0.026443
50%      0.010434    0.010014   -0.001190    0.002936    0.005315    0.011178
75%      0.043302    0.040243    0.033708    0.037621    0.037603    0.045666
max      0.364131    0.348325    0.524533    0.381874    0.377474    0.512788

               X12
count  500.000000
mean     0.005205
std      0.064719
min     -0.275196
25%     -0.026637
50%      0.006774
75%      0.036426
```

```
max       0.333801
```


nematoda efectores dataset 5 Covarianza de auto cruzamiento (ACC) hidro_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 5 Covarianza de auto cruzamiento (ACC) hidro_mass con valores atípicos.
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12

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 5,
sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0   -0.004622  0.021521 -0.007020 -0.002234  0.014137 -0.006020 -0.063299
1    0.065861  0.099888  0.115225  0.071350 -0.014651 -0.175079  0.083323
2   -0.011966  0.054191  0.019246 -0.022138  0.053399 -0.059781  0.072857
3    0.082915  0.026638  0.073253  0.028404  0.004208  0.046544  0.045879
4   -0.000618 -0.056232 -0.076153  0.077672  0.068254 -0.020627 -0.083698
..        …         …         …         …         …         …         …
495 -0.021948  0.032248 -0.010003  0.022332  0.007417  0.007493 -0.037968
496  0.057319  0.087196  0.100188  0.022304  0.052471  0.064041  0.059096
497 -0.043424 -0.023200 -0.094913 -0.050880  0.111304  0.026114  0.155105
498 -0.181910  0.133994 -0.092320 -0.022772  0.033055  0.088942  0.006578
499 -0.017375 -0.016556 -0.021813  0.070237  0.035177  0.009242 -0.039807

           X7        X8        X9       X10       X11       X12        X13
0    0.025683  0.001941 -0.013940  0.034448 -0.012567 -0.025834  efectores
1    0.015955 -0.038148  0.015743  0.053098 -0.063528 -0.007009  efectores
2   -0.046511 -0.054151 -0.033940  0.015368  0.030925  0.034573  efectores
3   -0.003023  0.079197  0.033078  0.054154  0.027731  0.023326  efectores
4    0.052851  0.150041 -0.078806 -0.070721  0.066841  0.014354  efectores
..        …         …         …         …         …         …         …
495 -0.018370 -0.073003 -0.025523 -0.041398 -0.008318  0.012562  efectores
496  0.082631 -0.002345  0.065305  0.094917  0.071823  0.068536  efectores
497  0.021170 -0.085518 -0.055961  0.050482 -0.048699  0.121860  efectores
498  0.103463  0.023689  0.109609 -0.109729  0.141867 -0.117117  efectores
499 -0.047536 -0.067483 -0.050028 -0.040587 -0.053042  0.048112  efectores

[459 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro_mass efectores nematoda dataset 5,
sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  459.000000  459.000000  459.000000  459.000000  459.000000  459.000000
mean     0.012315    0.013307    0.012429    0.011435    0.005430    0.003153
std      0.064727    0.063060    0.058910    0.057543    0.060976    0.058601
min     -0.193093   -0.184825   -0.176556   -0.180496   -0.198338   -0.175681
25%     -0.022147   -0.027660   -0.021392   -0.020639   -0.033054   -0.034726
50%      0.011266    0.014682    0.010119    0.013374    0.007417    0.005883
75%      0.050045    0.051026    0.047902    0.044775    0.041222    0.038981
max      0.204075    0.206665    0.200645    0.191792    0.193364    0.220812
```

```
                X6          X7          X8          X9          X10          X11  \
count  459.000000  459.000000  459.000000  459.000000  459.000000  459.000000
mean     0.005646    0.006792   -0.001285    0.003449   -0.003239    0.003803
std      0.060566    0.065008    0.057033    0.059240    0.061978    0.062513
min     -0.195519   -0.212102   -0.147297   -0.148398   -0.182494   -0.204743
25%     -0.030096   -0.029158   -0.037516   -0.036612   -0.041468   -0.031642
50%      0.001939    0.010373    0.000493   -0.000167   -0.000182    0.004714
75%      0.037183    0.044561    0.034061    0.039143    0.034653    0.041987
max      0.196682    0.206014    0.198671    0.203326    0.176985    0.209607

               X12
count  459.000000
mean     0.001355
std      0.061903
min     -0.207221
25%     -0.031853
50%      0.003537
75%      0.037333
max      0.193586
```

no_efectores
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores nematoda dataset 5, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.041656 -0.008989  0.054909  0.067907  0.002478  0.062140  0.017208
1   -0.018267  0.010369 -0.010773  0.033851  0.024248  0.063632 -0.024304
2   -0.004069  0.022637 -0.013418 -0.007399  0.064431 -0.002769  0.027306
4    0.012218 -0.016649  0.009331  0.026244  0.019812  0.045782  0.056438
5    0.010133  0.017403 -0.001420 -0.002480 -0.035501 -0.056899 -0.118508
..        …         …         …         …         …         …
493  0.010421  0.074682  0.009525  0.007835  0.006894  0.019970 -0.031968
495  0.019198  0.021302  0.027451 -0.067075  0.020520  0.012482  0.014099
496 -0.069403 -0.069587 -0.128997  0.027374  0.058314 -0.079835 -0.066858
497  0.043453 -0.007622 -0.008797 -0.001983  0.020591  0.035658 -0.004292
498 -0.042217  0.022555 -0.029549  0.008094  0.024395  0.045304 -0.058010

           X7        X8        X9       X10       X11       X12           X13
0    0.026875  0.016499  0.042661 -0.015428 -0.025516  0.062789  no_efectores
1   -0.033676  0.011571 -0.011409 -0.053864  0.036514 -0.071381  no_efectores
2   -0.080980 -0.025799 -0.025649  0.036599  0.031816 -0.000779  no_efectores
4   -0.017654 -0.009760  0.021604  0.089228 -0.014257 -0.065658  no_efectores
5    0.043755 -0.004402  0.002490  0.020497  0.031926  0.010595  no_efectores
..        …         …         …         …         …         …
493  0.018805 -0.060558  0.037767  0.018477 -0.003399  0.065179  no_efectores
```

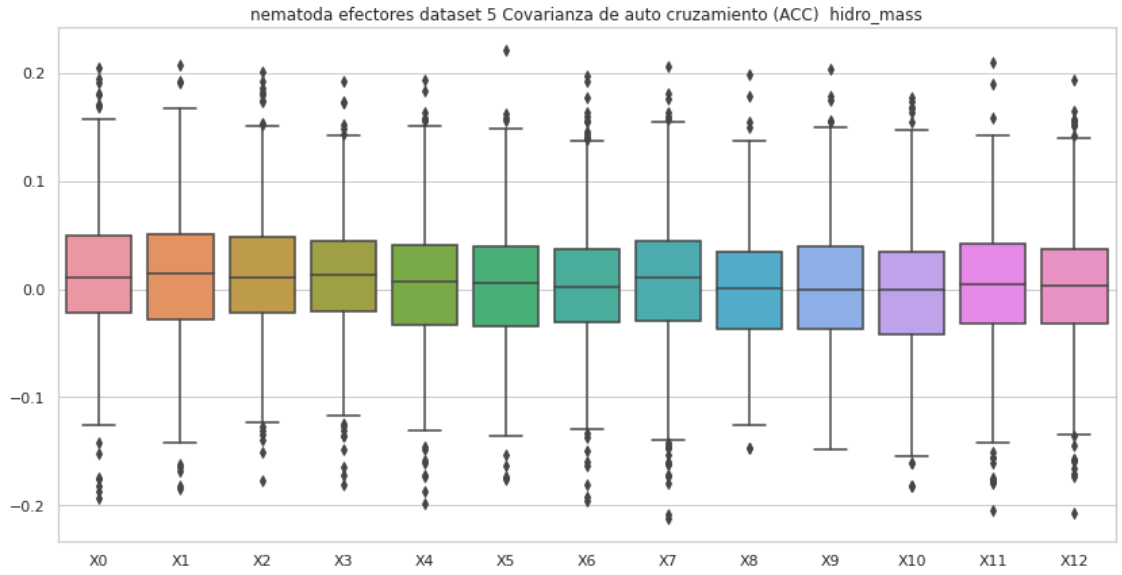
```
495 -0.012399 -0.015832 -0.050836  0.033870 -0.028486 -0.109200  no_efectores
496  0.017824  0.032227 -0.028218 -0.009270  0.109764 -0.076734  no_efectores
497  0.046409  0.016984 -0.004860  0.058795  0.056238  0.069334  no_efectores
498 -0.044435 -0.072076  0.025428  0.020339  0.062836  0.037676  no_efectores

[457 rows x 14 columns]
```

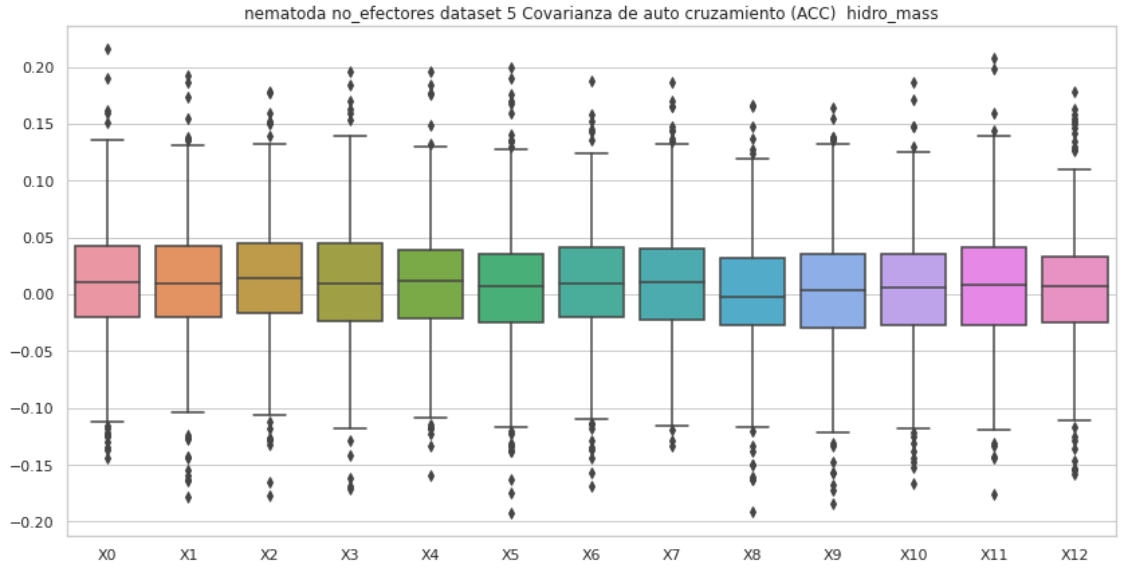
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores nematoda dataset 5, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	457.000000	457.000000	457.000000	457.000000	457.000000	457.000000	
mean	0.011774	0.009557	0.014369	0.010045	0.010324	0.005075	
std	0.054355	0.055003	0.052528	0.055275	0.051655	0.056106	
min	-0.144296	-0.178399	-0.177080	-0.171631	-0.158774	-0.192411	
25%	-0.019560	-0.019777	-0.016210	-0.023927	-0.021495	-0.025124	
50%	0.011246	0.009017	0.013863	0.009298	0.011387	0.007433	
75%	0.042658	0.042202	0.045319	0.045100	0.039293	0.036097	
max	0.216118	0.192835	0.178332	0.195678	0.196532	0.199116	

	X6	X7	X8	X9	X10	X11	\
count	457.000000	457.000000	457.000000	457.000000	457.000000	457.000000	
mean	0.008557	0.010224	0.000511	0.001162	0.004142	0.007551	
std	0.051922	0.052240	0.053528	0.054351	0.052354	0.053680	
min	-0.169009	-0.133751	-0.191673	-0.183586	-0.166792	-0.175713	
25%	-0.020340	-0.022815	-0.027352	-0.029453	-0.026629	-0.026642	
50%	0.009889	0.010852	-0.002250	0.003381	0.005674	0.008683	
75%	0.040848	0.040297	0.032494	0.035393	0.035481	0.041109	
max	0.187858	0.186616	0.166087	0.164175	0.186664	0.208343	

	X12
count	457.000000
mean	0.004918
std	0.053275
min	-0.158223
25%	-0.024471
50%	0.006712
75%	0.033583
max	0.178479





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 5, con
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.004622  0.021521 -0.007020 -0.002234  0.014137 -0.006020 -0.063299
1    0.065861  0.099888  0.115225  0.071350 -0.014651 -0.175079  0.083323
2   -0.011966  0.054191  0.019246 -0.022138  0.053399 -0.059781  0.072857
3    0.082915  0.026638  0.073253  0.028404  0.004208  0.046544  0.045879
4   -0.000618 -0.056232 -0.076153  0.077672  0.068254 -0.020627 -0.083698
..        …         …         …         …         …         …         …
495 -0.021948  0.032248 -0.010003  0.022332  0.007417  0.007493 -0.037968
496  0.057319  0.087196  0.100188  0.022304  0.052471  0.064041  0.059096
497 -0.043424 -0.023200 -0.094913 -0.050880  0.111304  0.026114  0.155105
498 -0.181910  0.133994 -0.092320 -0.022772  0.033055  0.088942  0.006578
499 -0.017375 -0.016556 -0.021813  0.070237  0.035177  0.009242 -0.039807

           X7        X8        X9       X10       X11       X12        X13
0    0.025683  0.001941 -0.013940  0.034448 -0.012567 -0.025834  efectores
1    0.015955 -0.038148  0.015743  0.053098 -0.063528 -0.007009  efectores
```

```
2   -0.046511 -0.054151 -0.033940  0.015368  0.030925  0.034573  efectores
3   -0.003023  0.079197  0.033078  0.054154  0.027731  0.023326  efectores
4    0.052851  0.150041 -0.078806 -0.070721  0.066841  0.014354  efectores
..        …         …         …         …         …         …
495 -0.018370 -0.073003 -0.025523 -0.041398 -0.008318  0.012562  efectores
496  0.082631 -0.002345  0.065305  0.094917  0.071823  0.068536  efectores
497  0.021170 -0.085518 -0.055961  0.050482 -0.048699  0.121860  efectores
498  0.103463  0.023689  0.109609 -0.109729  0.141867 -0.117117  efectores
499 -0.047536 -0.067483 -0.050028 -0.040587 -0.053042  0.048112  efectores

[500 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) mass efectores nematoda dataset 5, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.017079	0.015079	0.014660	0.014244	0.007627	0.007986	
std	0.077709	0.076643	0.076686	0.076867	0.073361	0.072181	
min	-0.193093	-0.297706	-0.294531	-0.425990	-0.366646	-0.175681	
25%	-0.022697	-0.029703	-0.024420	-0.021688	-0.033560	-0.033309	
50%	0.012837	0.014637	0.010353	0.013941	0.008556	0.007416	
75%	0.052196	0.054433	0.051740	0.049302	0.043024	0.042872	
max	0.657781	0.388844	0.517513	0.663383	0.556382	0.495386	

	X6	X7	X8	X9	X10	X11	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.006570	0.007250	0.003480	0.004524	-0.003810	0.005204	
std	0.074427	0.073578	0.072207	0.071697	0.067188	0.080205	
min	-0.253302	-0.337294	-0.227673	-0.228443	-0.206742	-0.322807	
25%	-0.032416	-0.032452	-0.037876	-0.038436	-0.043634	-0.034667	
50%	0.001712	0.010392	0.000625	0.000478	-0.000495	0.003796	
75%	0.039729	0.047985	0.038866	0.042653	0.036636	0.044210	
max	0.447357	0.367702	0.468705	0.550472	0.233571	0.623797	

	X12
count	500.000000
mean	0.004559
std	0.070788
min	-0.314624
25%	-0.032194
50%	0.005032
75%	0.041439
max	0.265259

no_efectores
Covarianza de auto cruzamiento (ACC) mass no_efectores nematoda dataset 5, con valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.041656	-0.008989	0.054909	0.067907	0.002478	0.062140	0.017208	
1	-0.018267	0.010369	-0.010773	0.033851	0.024248	0.063632	-0.024304	
2	-0.004069	0.022637	-0.013418	-0.007399	0.064431	-0.002769	0.027306	
3	0.079650	0.028390	-0.049272	0.153428	0.158789	-0.034182	-0.120187	
4	0.012218	-0.016649	0.009331	0.026244	0.019812	0.045782	0.056438	
..	…	…	…	…	…	…	…	
495	0.019198	0.021302	0.027451	-0.067075	0.020520	0.012482	0.014099	
496	-0.069403	-0.069587	-0.128997	0.027374	0.058314	-0.079835	-0.066858	
497	0.043453	-0.007622	-0.008797	-0.001983	0.020591	0.035658	-0.004292	
498	-0.042217	0.022555	-0.029549	0.008094	0.024395	0.045304	-0.058010	
499	0.297554	0.194138	0.169708	0.163544	0.133555	0.153411	0.208631	

	X7	X8	X9	X10	X11	X12	X13
0	0.026875	0.016499	0.042661	-0.015428	-0.025516	0.062789	no_efectores
1	-0.033676	0.011571	-0.011409	-0.053864	0.036514	-0.071381	no_efectores
2	-0.080980	-0.025799	-0.025649	0.036599	0.031816	-0.000779	no_efectores
3	-0.186278	-0.064429	0.030008	-0.056427	-0.115960	-0.160156	no_efectores
4	-0.017654	-0.009760	0.021604	0.089228	-0.014257	-0.065658	no_efectores
..	…	…	…	…	…	…	…
495	-0.012399	-0.015832	-0.050836	0.033870	-0.028486	-0.109200	no_efectores
496	0.017824	0.032227	-0.028218	-0.009270	0.109764	-0.076734	no_efectores
497	0.046409	0.016984	-0.004860	0.058795	0.056238	0.069334	no_efectores
498	-0.044435	-0.072076	0.025428	0.020339	0.062836	0.037676	no_efectores
499	0.194794	0.164737	0.189940	0.140611	0.213508	0.143059	no_efectores

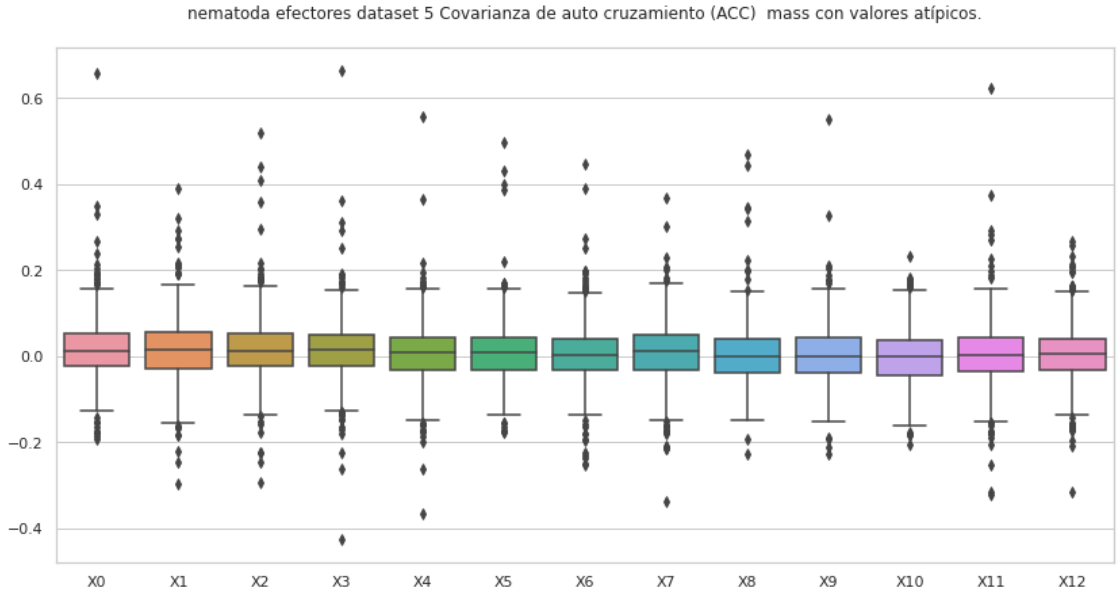
[500 rows x 14 columns]

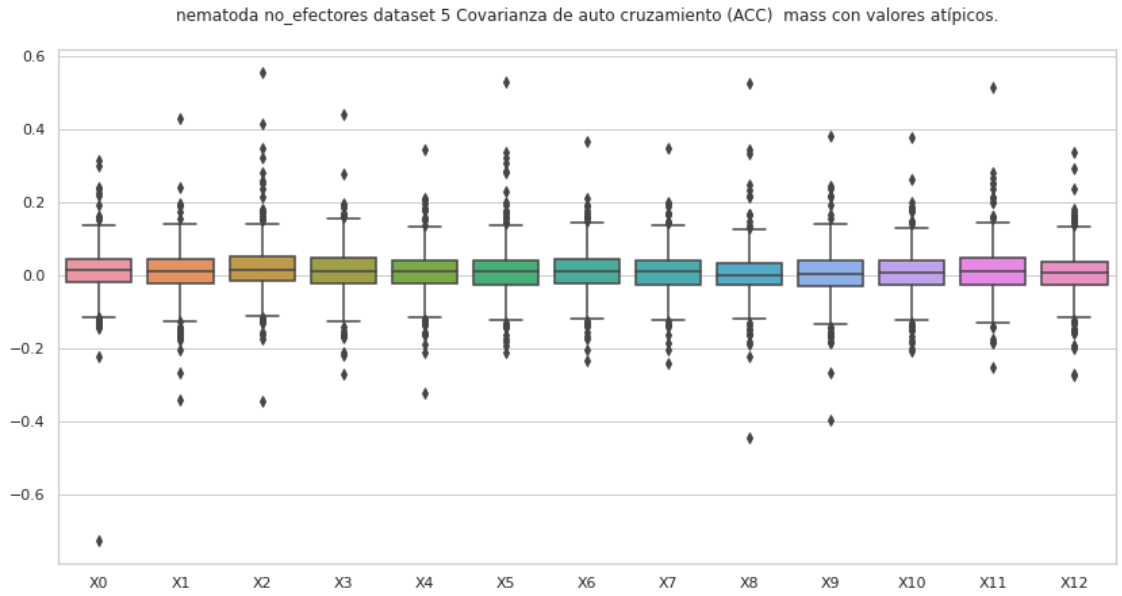
Covarianza de auto cruzamiento (ACC) mass no_efectores nematoda dataset 5, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.011645	0.007081	0.017769	0.010986	0.008891	0.008855	
std	0.069862	0.065532	0.071549	0.065249	0.062796	0.071274	
min	-0.727407	-0.341683	-0.344162	-0.271959	-0.323499	-0.212572	
25%	-0.020316	-0.025469	-0.017270	-0.025442	-0.023439	-0.028337	
50%	0.012047	0.008022	0.014048	0.009803	0.010194	0.008330	
75%	0.044302	0.042340	0.048669	0.046496	0.039512	0.037601	
max	0.315182	0.426943	0.553871	0.439271	0.342961	0.529563	

	X6	X7	X8	X9	X10	X11	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.008983	0.008416	0.003308	0.001250	0.004995	0.012056	
std	0.060803	0.061062	0.070075	0.067276	0.061139	0.067959	
min	-0.236000	-0.242599	-0.445889	-0.397801	-0.209637	-0.253624	
25%	-0.023567	-0.026978	-0.028092	-0.032824	-0.028222	-0.026443	
50%	0.010434	0.010014	-0.001190	0.002936	0.005315	0.011178	
75%	0.043302	0.040243	0.033708	0.037621	0.037603	0.045666	
max	0.364131	0.348325	0.524533	0.381874	0.377474	0.512788	

	X12
count	500.000000
mean	0.005205
std	0.064719
min	-0.275196
25%	-0.026637
50%	0.006774
75%	0.036426
max	0.333801





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 5, sin
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.004622  0.021521 -0.007020 -0.002234  0.014137 -0.006020 -0.063299
1    0.065861  0.099888  0.115225  0.071350 -0.014651 -0.175079  0.083323
2   -0.011966  0.054191  0.019246 -0.022138  0.053399 -0.059781  0.072857
3    0.082915  0.026638  0.073253  0.028404  0.004208  0.046544  0.045879
4   -0.000618 -0.056232 -0.076153  0.077672  0.068254 -0.020627 -0.083698
..        …         …         …         …         …         …         …
495 -0.021948  0.032248 -0.010003  0.022332  0.007417  0.007493 -0.037968
496  0.057319  0.087196  0.100188  0.022304  0.052471  0.064041  0.059096
497 -0.043424 -0.023200 -0.094913 -0.050880  0.111304  0.026114  0.155105
498 -0.181910  0.133994 -0.092320 -0.022772  0.033055  0.088942  0.006578
499 -0.017375 -0.016556 -0.021813  0.070237  0.035177  0.009242 -0.039807

           X7        X8        X9       X10       X11       X12        X13
0    0.025683  0.001941 -0.013940  0.034448 -0.012567 -0.025834  efectores
1    0.015955 -0.038148  0.015743  0.053098 -0.063528 -0.007009  efectores
2   -0.046511 -0.054151 -0.033940  0.015368  0.030925  0.034573  efectores
3   -0.003023  0.079197  0.033078  0.054154  0.027731  0.023326  efectores
4    0.052851  0.150041 -0.078806 -0.070721  0.066841  0.014354  efectores
..        …         …         …         …         …         …         …
```

```
495 -0.018370 -0.073003 -0.025523 -0.041398 -0.008318  0.012562  efectores
496  0.082631 -0.002345  0.065305  0.094917  0.071823  0.068536  efectores
497  0.021170 -0.085518 -0.055961  0.050482 -0.048699  0.121860  efectores
498  0.103463  0.023689  0.109609 -0.109729  0.141867 -0.117117  efectores
499 -0.047536 -0.067483 -0.050028 -0.040587 -0.053042  0.048112  efectores

[459 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) mass efectores nematoda dataset 5, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	459.000000	459.000000	459.000000	459.000000	459.000000	459.000000	
mean	0.012315	0.013307	0.012429	0.011435	0.005430	0.003153	
std	0.064727	0.063060	0.058910	0.057543	0.060976	0.058601	
min	-0.193093	-0.184825	-0.176556	-0.180496	-0.198338	-0.175681	
25%	-0.022147	-0.027660	-0.021392	-0.020639	-0.033054	-0.034726	
50%	0.011266	0.014682	0.010119	0.013374	0.007417	0.005883	
75%	0.050045	0.051026	0.047902	0.044775	0.041222	0.038981	
max	0.204075	0.206665	0.200645	0.191792	0.193364	0.220812	

	X6	X7	X8	X9	X10	X11	\
count	459.000000	459.000000	459.000000	459.000000	459.000000	459.000000	
mean	0.005646	0.006792	-0.001285	0.003449	-0.003239	0.003803	
std	0.060566	0.065008	0.057033	0.059240	0.061978	0.062513	
min	-0.195519	-0.212102	-0.147297	-0.148398	-0.182494	-0.204743	
25%	-0.030096	-0.029158	-0.037516	-0.036612	-0.041468	-0.031642	
50%	0.001939	0.010373	0.000493	-0.000167	-0.000182	0.004714	
75%	0.037183	0.044561	0.034061	0.039143	0.034653	0.041987	
max	0.196682	0.206014	0.198671	0.203326	0.176985	0.209607	

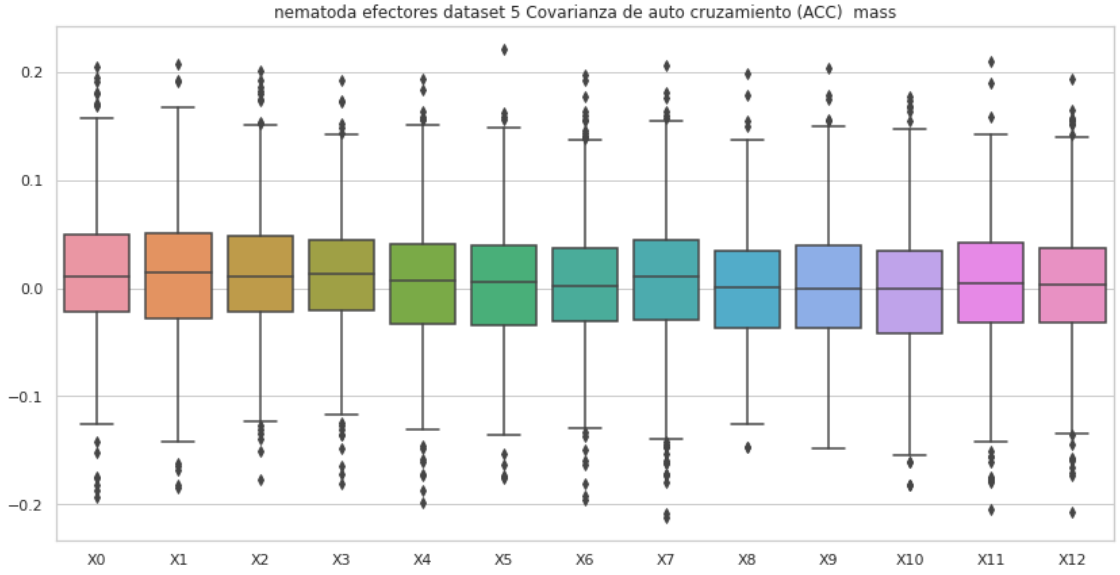
	X12
count	459.000000
mean	0.001355
std	0.061903
min	-0.207221
25%	-0.031853
50%	0.003537
75%	0.037333
max	0.193586

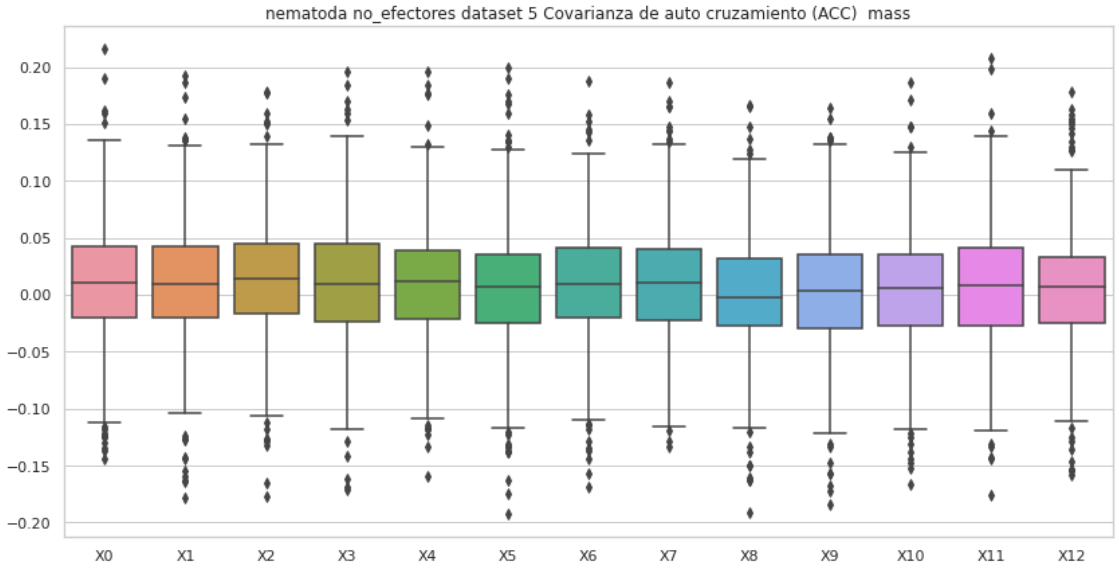
Covarianza de auto cruzamiento (ACC) mass no_efectores nematoda dataset 5, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.041656 -0.008989  0.054909  0.067907  0.002478  0.062140  0.017208
1   -0.018267  0.010369 -0.010773  0.033851  0.024248  0.063632 -0.024304
2   -0.004069  0.022637 -0.013418 -0.007399  0.064431 -0.002769  0.027306
4    0.012218 -0.016649  0.009331  0.026244  0.019812  0.045782  0.056438
5    0.010133  0.017403 -0.001420 -0.002480 -0.035501 -0.056899 -0.118508
..        …         …         …         …         …         …
493  0.010421  0.074682  0.009525  0.007835  0.006894  0.019970 -0.031968
495  0.019198  0.021302  0.027451 -0.067075  0.020520  0.012482  0.014099
496 -0.069403 -0.069587 -0.128997  0.027374  0.058314 -0.079835 -0.066858
497  0.043453 -0.007622 -0.008797 -0.001983  0.020591  0.035658 -0.004292
498 -0.042217  0.022555 -0.029549  0.008094  0.024395  0.045304 -0.058010

           X7        X8        X9       X10       X11       X12           X13
0    0.026875  0.016499  0.042661 -0.015428 -0.025516  0.062789  no_efectores
1   -0.033676  0.011571 -0.011409 -0.053864  0.036514 -0.071381  no_efectores
2   -0.080980 -0.025799 -0.025649  0.036599  0.031816 -0.000779  no_efectores
4   -0.017654 -0.009760  0.021604  0.089228 -0.014257 -0.065658  no_efectores
5    0.043755 -0.004402  0.002490  0.020497  0.031926  0.010595  no_efectores
..        …         …         …         …         …         …
493  0.018805 -0.060558  0.037767  0.018477 -0.003399  0.065179  no_efectores
495 -0.012399 -0.015832 -0.050836  0.033870 -0.028486 -0.109200  no_efectores
496  0.017824  0.032227 -0.028218 -0.009270  0.109764 -0.076734  no_efectores
497  0.046409  0.016984 -0.004860  0.058795  0.056238  0.069334  no_efectores
498 -0.044435 -0.072076  0.025428  0.020339  0.062836  0.037676  no_efectores

[457 rows x 14 columns]


Covarianza de auto cruzamiento (ACC)  mass no_efectores nematoda dataset 5, sin
valores atípicos.
Estadísticas.

                X0          X1          X2          X3          X4          X5  \
count  457.000000  457.000000  457.000000  457.000000  457.000000  457.000000
mean     0.011774    0.009557    0.014369    0.010045    0.010324    0.005075
std      0.054355    0.055003    0.052528    0.055275    0.051655    0.056106
min     -0.144296   -0.178399   -0.177080   -0.171631   -0.158774   -0.192411
25%     -0.019560   -0.019777   -0.016210   -0.023927   -0.021495   -0.025124
50%      0.011246    0.009017    0.013863    0.009298    0.011387    0.007433
75%      0.042658    0.042202    0.045319    0.045100    0.039293    0.036097
max      0.216118    0.192835    0.178332    0.195678    0.196532    0.199116

                X6          X7          X8          X9         X10         X11  \
count  457.000000  457.000000  457.000000  457.000000  457.000000  457.000000
mean     0.008557    0.010224    0.000511    0.001162    0.004142    0.007551
std      0.051922    0.052240    0.053528    0.054351    0.052354    0.053680
```

min	-0.169009	-0.133751	-0.191673	-0.183586	-0.166792	-0.175713
25%	-0.020340	-0.022815	-0.027352	-0.029453	-0.026629	-0.026642
50%	0.009889	0.010852	-0.002250	0.003381	0.005674	0.008683
75%	0.040848	0.040297	0.032494	0.035393	0.035481	0.041109
max	0.187858	0.186616	0.166087	0.164175	0.186664	0.208343

	X12
count	457.000000
mean	0.004918
std	0.053275
min	-0.158223
25%	-0.024471
50%	0.006712
75%	0.033583
max	0.178479





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 5, con
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.065449 -0.181147 -0.076902  0.054192 -0.016721  0.012068  0.047374
1    0.062272 -0.296558  0.080816  0.219622  0.074763 -0.104301  0.003782
2    0.001888 -0.031313 -0.060426 -0.000681 -0.017064 -0.110276 -0.060509
3    0.046185 -0.028828  0.024607  0.016371 -0.034182  0.029293  0.102199
4    0.168342  0.066908  0.061463 -0.125254 -0.026610 -0.215291  0.143828
..        …       …       …       …       …       …       …
495 -0.038636  0.018320 -0.016168 -0.045454  0.035099  0.042829  0.006529
496 -0.018126 -0.001221 -0.094503 -0.026380  0.004365  0.072932 -0.067343
497 -0.176960  0.060679  0.102384  0.098178  0.071239 -0.047339  0.102444
498  0.020969  0.077289 -0.003677 -0.020676  0.025965  0.009992  0.047705
499  0.001033 -0.177276  0.098064  0.079071 -0.062161 -0.039300 -0.032269

           X7        X8        X9       X10       X11       X12        X13
0    0.039483 -0.039406 -0.028236  0.081010  0.007719  0.002401  efectores
1    0.185030  0.195701 -0.174684 -0.055190  0.139330  0.125565  efectores
2    0.038951 -0.094353  0.043500  0.043780 -0.012163 -0.018647  efectores
3    0.086985  0.031084  0.034815  0.074027 -0.018158  0.032862  efectores
4   -0.192572 -0.091838  0.149808 -0.007533  0.207057  0.051471  efectores
..        …       …       …       …       …       …        …
495 -0.111524 -0.029119  0.135768 -0.027309  0.067161 -0.159224  efectores
496  0.057168 -0.021027  0.022426  0.075119  0.002902  0.002995  efectores
497 -0.070403  0.108740  0.099587 -0.019684 -0.212874  0.257808  efectores
498 -0.029399  0.108737  0.006403  0.040387  0.163797 -0.042315  efectores
499 -0.058634  0.044869 -0.200197 -0.033778  0.045400 -0.093588  efectores

[500 rows x 14 columns]


Covarianza de auto cruzamiento (ACC)  hidro efectores nematoda dataset 5, con
valores atípicos.
Estadísticas.

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

```
mean     0.012705   -0.023858    0.025432    0.022038   -0.007304    0.000770
std      0.084163    0.083272    0.084723    0.087464    0.089921    0.085139
min     -0.227791   -0.340708   -0.246477   -0.278880   -0.343077   -0.268681
25%     -0.035116   -0.073592   -0.028144   -0.027209   -0.058964   -0.051367
50%      0.008007   -0.024524    0.030883    0.018734   -0.007493    0.004443
75%      0.054331    0.020867    0.071869    0.072291    0.047065    0.048705
max      0.351128    0.281419    0.340031    0.423707    0.375234    0.336974

               X6          X7          X8          X9         X10         X11  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.022493    0.006716    0.000038    0.004455    0.011748    0.001434
std      0.089203    0.082449    0.083317    0.085247    0.081058    0.083695
min     -0.333538   -0.323031   -0.353739   -0.418536   -0.317442   -0.265034
25%     -0.028764   -0.038989   -0.050311   -0.042662   -0.028448   -0.043521
50%      0.020605    0.006003   -0.000803    0.009416    0.015447   -0.000637
75%      0.070275    0.055490    0.045639    0.053349    0.055933    0.041758
max      0.335412    0.333394    0.327024    0.304677    0.292291    0.688250

              X12
count  500.000000
mean     0.005972
std      0.080808
min     -0.469906
25%     -0.042657
50%      0.009558
75%      0.053267
max      0.257808
```

no_efectores
Covarianza de auto cruzamiento (ACC) hidro no_efectores nematoda dataset 5, con valores atípicos.
Valores del documento csv.

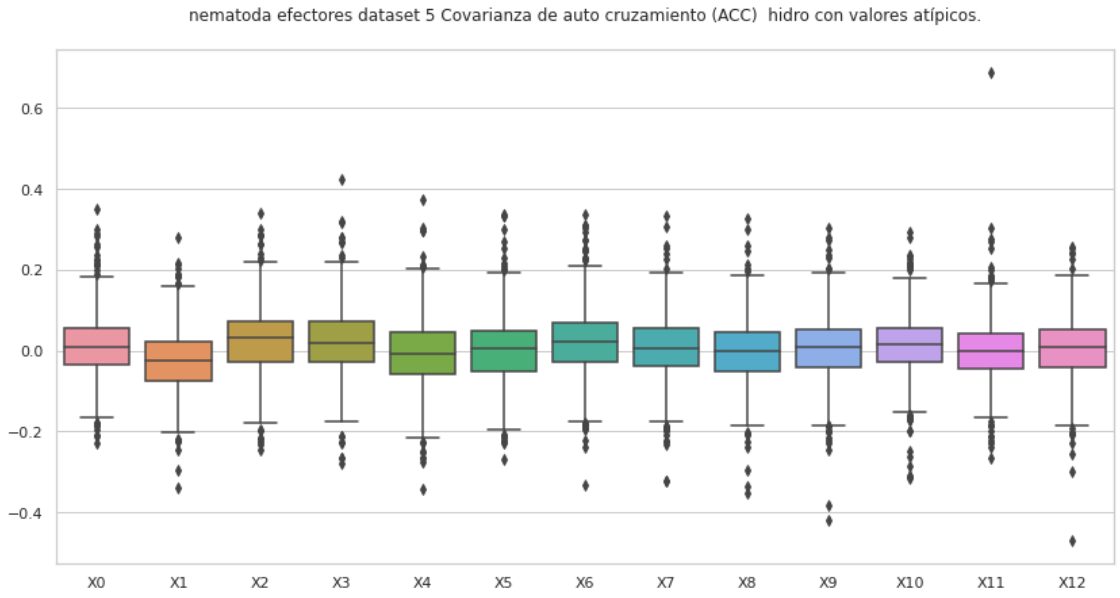
```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.096572  0.006321  0.048899  0.076821  0.043546  0.102528  0.097434
1   -0.095629  0.175052 -0.256324  0.211855 -0.124108  0.050768  0.130348
2   -0.026905 -0.061926 -0.008408  0.018808 -0.066431 -0.031572 -0.055957
3    0.013305  0.034633  0.011647 -0.153853 -0.123691 -0.123910  0.228294
4    0.071869  0.069960  0.065305  0.041792  0.164730  0.070459  0.026324
..        …         …         …         …         …         …         …
495  0.056560 -0.019064  0.005954  0.054416 -0.040938 -0.072531  0.172065
496  0.107099  0.041686  0.131490 -0.067317 -0.101211 -0.074666 -0.011550
497  0.058273  0.026851  0.019402  0.053807  0.063772  0.014892  0.041415
498  0.001318 -0.129490  0.084089  0.061811 -0.003121 -0.014188  0.082529
499  0.018014 -0.046953  0.008701  0.048947 -0.037287 -0.026646  0.011671
```

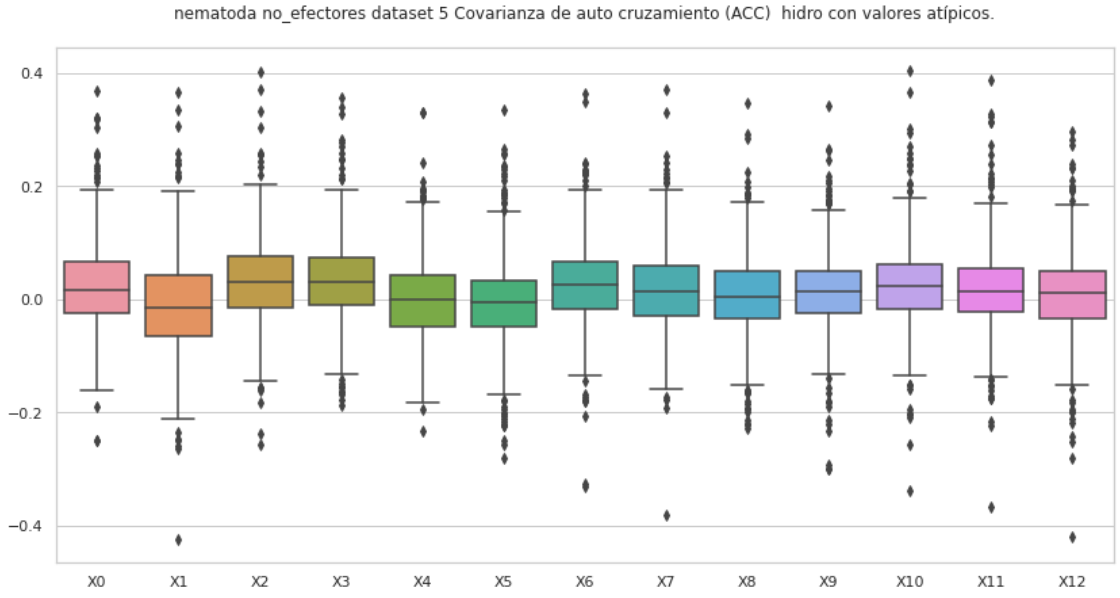
```
           X7        X8        X9       X10       X11       X12           X13
0    0.065845  0.055813  0.070269  0.077024  0.041476  0.049044  no_efectores
1    0.124776  0.025900 -0.138697 -0.004180 -0.050374 -0.029132  no_efectores
2   -0.025621 -0.011381 -0.032625 -0.036936  0.075421  0.023216  no_efectores
3    0.011172 -0.026984  0.062815 -0.043817 -0.004879  0.084365  no_efectores
4    0.075946  0.004776  0.045297 -0.097330  0.004188  0.008553  no_efectores
..        …         …         …         …         …         …
495  0.106629  0.009019  0.026181  0.107273  0.006954 -0.127720  no_efectores
496 -0.086165 -0.151217 -0.032526  0.047597 -0.063823 -0.029458  no_efectores
497  0.036206  0.058985  0.017451  0.026728  0.055553  0.034127  no_efectores
498 -0.027057  0.035768 -0.056479 -0.006759  0.086855 -0.019499  no_efectores
499 -0.025053 -0.024304  0.031028 -0.053772 -0.012348  0.023708  no_efectores

[500 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro no_efectores nematoda dataset 5, con valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count   500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean      0.024760   -0.010285    0.031810    0.035786    0.002756   -0.004414
std       0.082429    0.091283    0.078616    0.077790    0.076121    0.080903
min      -0.250831   -0.424922   -0.256324   -0.186971   -0.234072   -0.281497
25%      -0.025315   -0.064675   -0.014664   -0.010480   -0.048222   -0.047955
50%       0.016846   -0.014092    0.029751    0.029249   -0.001011   -0.005321
75%       0.066538    0.041441    0.076146    0.074497    0.041357    0.033894
max       0.367452    0.365085    0.400670    0.356711    0.330576    0.333860

                X6          X7          X8          X9         X10         X11  \
count   500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean      0.027140    0.015540    0.006345    0.013109    0.025020    0.016879
std       0.076792    0.079067    0.075092    0.075235    0.078807    0.078319
min      -0.330825   -0.381499   -0.227550   -0.301328   -0.338875   -0.367765
25%      -0.017344   -0.030193   -0.033665   -0.025347   -0.017879   -0.023109
50%       0.025763    0.012959    0.003329    0.014428    0.023610    0.012431
75%       0.067410    0.059490    0.050408    0.049152    0.062137    0.054515
max       0.364355    0.370954    0.347108    0.341267    0.404052    0.388510

               X12
count   500.000000
mean      0.008030
std       0.078912
min      -0.419919
25%      -0.033210
50%       0.010289
75%       0.049430
```

```
max       0.296722
```


nematoda efectores dataset 5 Covarianza de auto cruzamiento (ACC) hidro 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 5 Covarianza de auto cruzamiento (ACC) hidro con valores atípicos.
0.4
0.2
0.0
−0.2
−0.4
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12

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 5, sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.065449	-0.181147	-0.076902	0.054192	-0.016721	0.012068	0.047374	
2	0.001888	-0.031313	-0.060426	-0.000681	-0.017064	-0.110276	-0.060509	
3	0.046185	-0.028828	0.024607	0.016371	-0.034182	0.029293	0.102199	
4	0.168342	0.066908	0.061463	-0.125254	-0.026610	-0.215291	0.143828	
5	-0.076580	-0.068564	0.186762	-0.000599	-0.074997	0.094595	-0.098404	
..	…	…	…	…	…	…	…	
494	-0.150077	-0.135383	0.110230	-0.010398	-0.023000	-0.056699	0.046604	
495	-0.038636	0.018320	-0.016168	-0.045454	0.035099	0.042829	0.006529	
496	-0.018126	-0.001221	-0.094503	-0.026380	0.004365	0.072932	-0.067343	
498	0.020969	0.077289	-0.003677	-0.020676	0.025965	0.009992	0.047705	
499	0.001033	-0.177276	0.098064	0.079071	-0.062161	-0.039300	-0.032269	

	X7	X8	X9	X10	X11	X12	X13
0	0.039483	-0.039406	-0.028236	0.081010	0.007719	0.002401	efectores
2	0.038951	-0.094353	0.043500	0.043780	-0.012163	-0.018647	efectores
3	0.086985	0.031084	0.034815	0.074027	-0.018158	0.032862	efectores
4	-0.192572	-0.091838	0.149808	-0.007533	0.207057	0.051471	efectores
5	-0.190110	0.051057	0.095726	-0.019219	0.062449	-0.019484	efectores
..	…	…	…	…	…	…	…
494	0.054735	-0.003911	-0.056434	0.039323	-0.027268	-0.134853	efectores
495	-0.111524	-0.029119	0.135768	-0.027309	0.067161	-0.159224	efectores
496	0.057168	-0.021027	0.022426	0.075119	0.002902	0.002995	efectores
498	-0.029399	0.108737	0.006403	0.040387	0.163797	-0.042315	efectores
499	-0.058634	0.044869	-0.200197	-0.033778	0.045400	-0.093588	efectores

[458 rows x 14 columns]

Covarianza de auto cruzamiento (ACC) efectores nematoda dataset 5, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	458.000000	458.000000	458.000000	458.000000	458.000000	458.000000	
mean	0.011125	-0.024350	0.023658	0.021035	-0.009761	-0.000207	
std	0.075926	0.075314	0.072741	0.075589	0.080446	0.074706	
min	-0.210230	-0.244431	-0.216851	-0.230159	-0.274186	-0.225596	
25%	-0.034101	-0.072686	-0.025987	-0.025845	-0.057301	-0.048515	
50%	0.007762	-0.023848	0.029922	0.018734	-0.010500	0.004443	
75%	0.051729	0.018345	0.067776	0.069386	0.043030	0.047515	
max	0.263665	0.186105	0.241461	0.279705	0.211669	0.228822	

```
                X6           X7           X8           X9          X10          X11  \
count   458.000000   458.000000   458.000000   458.000000   458.000000   458.000000
mean      0.020313     0.007434    -0.001341     0.002664     0.012558    -0.001760
std       0.078022     0.072752     0.071958     0.076254     0.067932     0.069259
min      -0.221004    -0.230583    -0.239126    -0.245327    -0.199480    -0.240322
25%      -0.028457    -0.034011    -0.046727    -0.043317    -0.026954    -0.042831
50%       0.021009     0.006337    -0.002450     0.007260     0.015678    -0.001178
75%       0.066182     0.052554     0.041243     0.052076     0.054121     0.038438
max       0.272392     0.253729     0.246804     0.250552     0.232710     0.207057

               X12
count   458.000000
mean      0.004978
std       0.071657
min      -0.228485
25%      -0.042542
50%       0.008326
75%       0.050998
max       0.241607
```

no_efectores
Covarianza de auto cruzamiento (ACC) no_efectores nematoda dataset 5, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.096572  0.006321  0.048899  0.076821  0.043546  0.102528  0.097434
2   -0.026905 -0.061926 -0.008408  0.018808 -0.066431 -0.031572 -0.055957
3    0.013305  0.034633  0.011647 -0.153853 -0.123691 -0.123910  0.228294
4    0.071869  0.069960  0.065305  0.041792  0.164730  0.070459  0.026324
5    0.054957 -0.117148 -0.015336 -0.029880 -0.053025  0.024331 -0.036885
..        …         …         …         …         …         …         …
495  0.056560 -0.019064  0.005954  0.054416 -0.040938 -0.072531  0.172065
496  0.107099  0.041686  0.131490 -0.067317 -0.101211 -0.074666 -0.011550
497  0.058273  0.026851  0.019402  0.053807  0.063772  0.014892  0.041415
498  0.001318 -0.129490  0.084089  0.061811 -0.003121 -0.014188  0.082529
499  0.018014 -0.046953  0.008701  0.048947 -0.037287 -0.026646  0.011671

           X7        X8        X9       X10       X11       X12           X13
0    0.065845  0.055813  0.070269  0.077024  0.041476  0.049044  no_efectores
2   -0.025621 -0.011381 -0.032625 -0.036936  0.075421  0.023216  no_efectores
3    0.011172 -0.026984  0.062815 -0.043817 -0.004879  0.084365  no_efectores
4    0.075946  0.004776  0.045297 -0.097330  0.004188  0.008553  no_efectores
5   -0.004116 -0.030017 -0.027407  0.027739  0.045095 -0.001934  no_efectores
..        …         …         …         …         …         …           …
495  0.106629  0.009019  0.026181  0.107273  0.006954 -0.127720  no_efectores
```

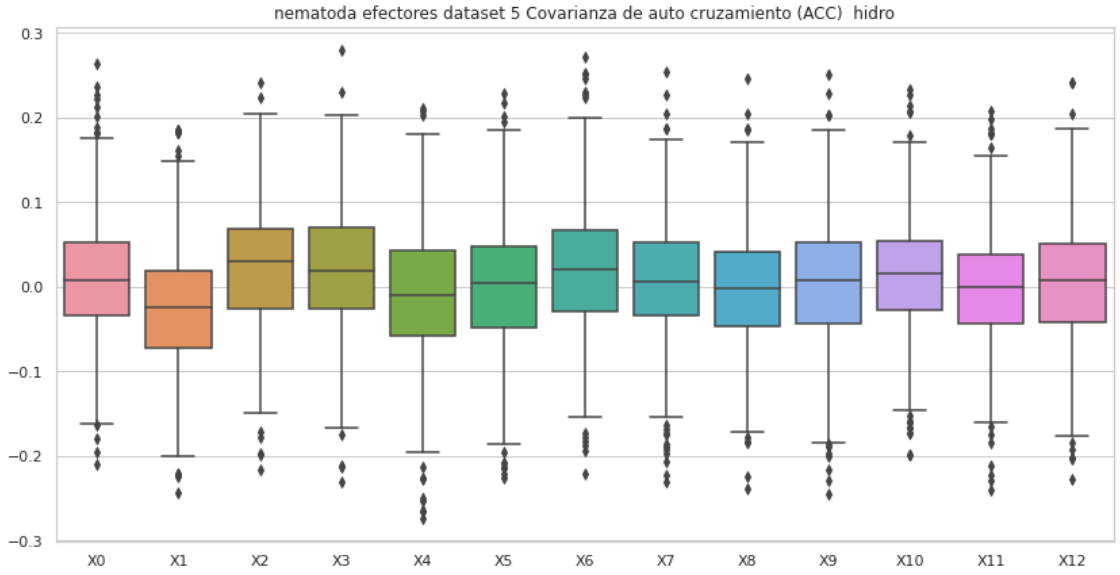
```
496 -0.086165 -0.151217 -0.032526  0.047597 -0.063823 -0.029458  no_efectores
497  0.036206  0.058985  0.017451  0.026728  0.055553  0.034127  no_efectores
498 -0.027057  0.035768 -0.056479 -0.006759  0.086855 -0.019499  no_efectores
499 -0.025053 -0.024304  0.031028 -0.053772 -0.012348  0.023708  no_efectores

[454 rows x 14 columns]
```

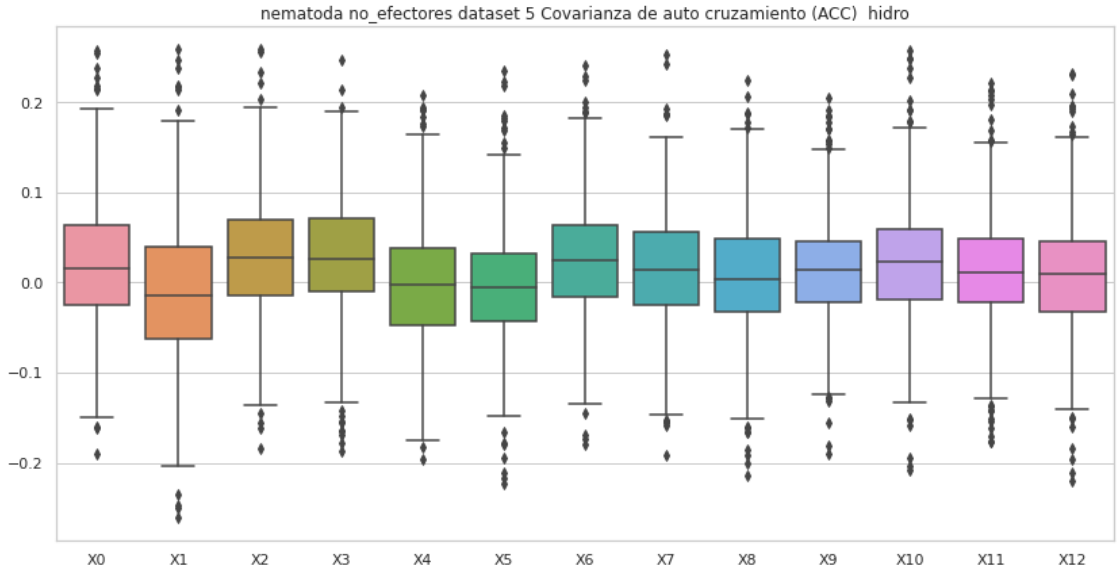
Covarianza de auto cruzamiento (ACC) no_efectores nematoda dataset 5, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	454.000000	454.000000	454.000000	454.000000	454.000000	454.000000	
mean	0.021837	-0.011317	0.028141	0.029054	0.000515	-0.004075	
std	0.071580	0.080916	0.067713	0.067329	0.070424	0.066712	
min	-0.190666	-0.260724	-0.183718	-0.186971	-0.195423	-0.223015	
25%	-0.024286	-0.062355	-0.015112	-0.010326	-0.047778	-0.042386	
50%	0.015464	-0.014092	0.027511	0.026457	-0.001754	-0.004862	
75%	0.063337	0.039037	0.069868	0.071226	0.037831	0.032567	
max	0.257509	0.258714	0.258503	0.246070	0.208147	0.234841	

	X6	X7	X8	X9	X10	X11	\
count	454.000000	454.000000	454.000000	454.000000	454.000000	454.000000	
mean	0.024981	0.015476	0.006151	0.013022	0.022320	0.012411	
std	0.064988	0.067162	0.066838	0.061005	0.068030	0.064582	
min	-0.179880	-0.191906	-0.214094	-0.190265	-0.208468	-0.176758	
25%	-0.016106	-0.025692	-0.032421	-0.022510	-0.018595	-0.022676	
50%	0.024558	0.013882	0.004131	0.013707	0.023305	0.011478	
75%	0.063373	0.055510	0.049114	0.046047	0.059727	0.049117	
max	0.241558	0.252353	0.224356	0.205229	0.257240	0.221497	

	X12
count	454.000000
mean	0.007306
std	0.066727
min	-0.219946
25%	-0.031908
50%	0.009169
75%	0.046216
max	0.231627


nematoda efectores dataset 5 Covarianza de auto cruzamiento (ACC) hidro
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12


nematoda no_efectores dataset 5 Covarianza de auto cruzamiento (ACC) hidro
0.2
0.1
0.0
−0.1
−0.2
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12