ds1_fusarium_oxysporum_limpieza_de_datos

February 1, 2021

Limpieza de datos

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

1 Declaración de variables

[2]:
```
organismo ="fusarium_oxysporum"
dataset = 1
nombre = ("ds" + str(dataset) + "_" + str(organismo))
nombre2 = (str(organismo)+ " dataset " + str(dataset))
r2 = ("Datos/resultados/"+ str(organismo) + "/" + str(nombre) + "/
 ↪transformaciones/sin_filtrar")
r3 = ("Datos/resultados/"+ str(organismo) + "/" + str(nombre) + "/
 ↪transformaciones/sin_atipicos")

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

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

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

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

2 Composición de aminoácidos (AAC)

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

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

    if etiq == "efectores":
        df=AAC_efec

    if etiq == "no_efectores":
        df=AAC_no_efec

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

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

```
efectores
Composición de aminoácidos (AAC) efectores fusarium_oxysporum dataset 1, con
valores atípicos.
Valores del documento csv.

         X0     X1     X2     X3     X4     X5     X6     X7     X8     X9  \
0     9.695  4.668  3.591  6.463  2.154  6.284  3.950  5.745  3.232  4.668
1     9.761  6.773  3.785  5.378  2.590  5.578  2.988  4.980  2.390  6.175
2     8.048  5.231  2.414  4.225  1.207  3.823  4.024  5.231  2.414  3.823
3     6.966  7.191  3.446  6.667  1.423  7.491  3.596  6.067  3.221  4.419
4     8.333  7.540  1.190  7.143  0.794  8.333  2.381  7.143  2.381  1.984
..      …      …      …      …      …      …      …      …      …      …
995   4.762  6.667  3.492  8.254  0.952  5.397  1.905  6.032  3.492  5.079
996  13.971  2.206  4.412  3.676  1.471  3.676  3.676  5.882  4.412  5.147
997   9.714  5.429  3.000  4.000  1.143  3.714  3.571  8.000  1.857  8.571
998   8.255  3.066  6.604  5.660  0.943  3.774  4.481  7.783  2.123  4.009
999   3.356  7.383  2.685  6.040  1.342  6.711  6.711  6.040  4.027  6.040
```

```
     …    X11    X12    X13     X14     X15    X16    X17    X18    X19  \
0    …  4.309  3.591  4.309   6.643   8.079  7.181  1.795  2.154  3.950
1    …  2.789  2.988  4.582   5.777   8.167  7.371  1.992  3.386  5.179
2    …  3.219  1.610  4.829  13.481  11.469  8.048  1.207  1.811  6.439
3    …  3.745  1.798  5.019   6.517   8.315  5.843  1.273  3.071  7.116
4    …  2.381  2.778  1.984   9.921   9.127  4.365  1.587  0.397  8.333
..   …    …    …    …     …       …      …    …      …      …
995  …  4.444  2.222  3.492   7.619   5.714  6.032  3.175  2.857  7.937
996  …  6.618  2.941  2.941   7.353   6.618  6.618  0.000  2.941  5.147
997  …  3.000  3.143  5.286   4.571   7.429  5.714  2.000  4.714  5.857
998  …  4.953  1.887  4.245   3.538   6.840  7.783  4.481  4.009  6.840
999  …  3.356  2.013  2.685   7.383  12.752  4.027  1.342  2.685  4.698

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

[1000 rows x 21 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      8.418299     5.775609     3.753980     5.701968     1.551843
std       2.352574     2.239349     1.481839     1.996339     1.348520
min       1.829000     0.000000     0.000000     0.000000     0.000000
25%       6.921750     4.290500     2.855000     4.510500     0.690000
50%       8.142000     5.826000     3.598500     5.684000     1.302000
75%       9.735500     7.047000     4.545000     6.807750     2.015250
max      20.455000    19.266000    14.172000    17.647000    13.462000

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      5.994206     4.028894     6.714977     2.487024     5.164738
std       2.330624     1.860946     2.448200     1.224416     1.817357
```

min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	4.351000	2.915250	5.155750	1.709000	4.031250
50%	5.953000	3.799500	6.485000	2.418500	5.128000
75%	7.274750	4.786250	7.925000	3.150250	6.251750
max	17.976000	18.103000	31.737000	11.111000	12.676000

	X10	X11	X12	X13	X14 \
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	8.750275	4.996038	2.313000	3.774536	5.852927
std	2.539614	2.186132	1.116646	1.567017	2.554912
min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	7.183750	3.550000	1.612250	2.828000	4.323250
50%	8.715500	4.791000	2.194500	3.661000	5.534500
75%	10.340500	6.164750	2.832500	4.670000	6.803750
max	21.495000	16.788000	10.938000	11.000000	30.242000

	X15	X16	X17	X18	X19
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	8.089448	6.038957	1.601067	2.857967	6.134286
std	2.895915	2.218685	0.984200	1.338220	1.988856
min	0.877000	0.000000	0.000000	0.000000	0.000000
25%	6.272750	4.847250	0.903250	2.040250	4.869000
50%	7.692000	5.792000	1.508500	2.757500	5.943500
75%	9.285250	6.765000	2.168500	3.561250	7.252500
max	29.677000	22.222000	7.692000	12.821000	23.393000

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

	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9 \
0	6.273	5.351	2.768	4.428	1.661	6.089	4.244	6.642	2.583	6.458
1	9.172	5.478	4.076	4.841	1.529	5.987	3.694	7.516	2.038	4.713
2	7.705	6.616	3.183	6.198	1.675	6.449	4.523	5.779	2.345	5.025
3	8.908	6.226	4.262	7.615	1.964	6.418	2.826	5.987	3.161	5.029
4	9.651	8.311	2.547	4.558	1.609	5.362	7.373	6.434	2.011	5.228
..	…	…	…	…	…	…	…	…	…	…
995	6.462	5.654	5.170	7.593	1.292	6.139	4.200	6.947	3.554	5.977
996	8.237	6.590	3.789	5.766	2.471	6.096	4.448	4.448	1.977	5.931
997	8.203	3.125	3.516	5.078	1.562	1.953	3.516	10.156	1.953	5.469
998	5.866	7.821	2.235	6.704	1.955	5.587	3.911	5.028	3.073	3.631
999	7.164	5.848	2.632	7.018	3.216	6.579	5.702	6.871	3.070	5.848

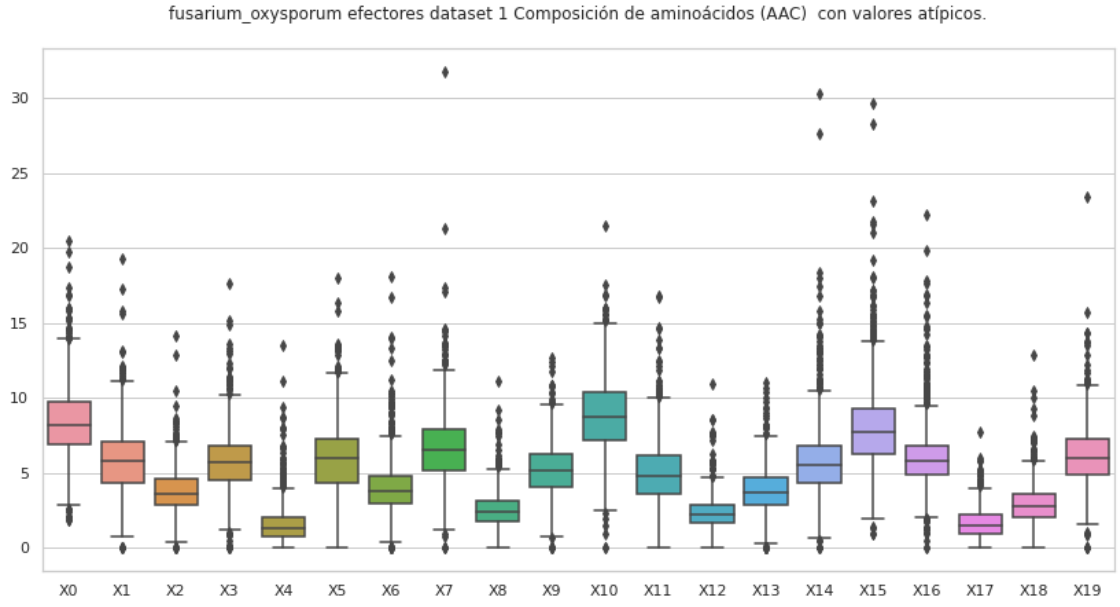
	…	X11	X12	X13	X14	X15	X16	X17	X18	X19 \
0	…	5.535	1.845	5.166	6.827	5.904	7.011	2.030	2.952	4.428

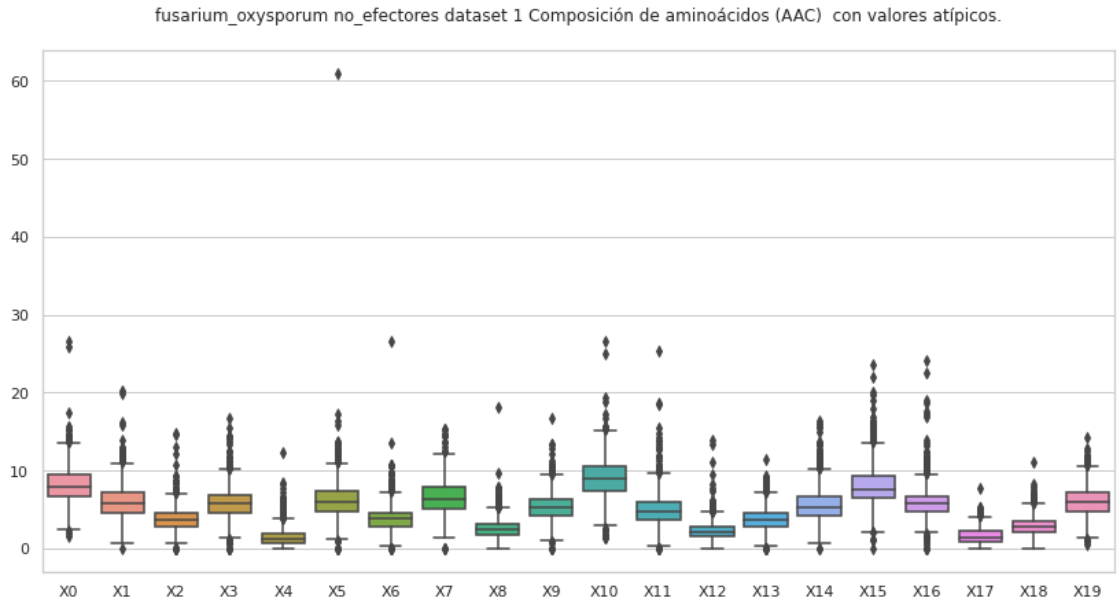
```
1    …  4.076  1.911  4.076  5.605   7.261  5.987  1.019  1.911   8.153
2    …  3.769  1.926  3.518  5.611   9.883  7.705  1.005  2.010   5.695
3    …  4.598  1.724  2.921  5.029   6.034  4.646  1.437  3.257   5.603
4    …  1.072  2.011  3.083  5.496   9.651  5.630  1.072  2.681   5.898
..   …    …    …    …    …     …     …    …     …     …
995  …  5.331  2.423  2.746  5.977   4.685  4.200  2.100  3.393   7.916
996  …  3.954  3.130  3.624  5.107   8.896  5.601  1.812  2.471   4.942
997  …  3.516  2.344  7.422  2.734   9.766  6.641  3.125  1.953  10.156
998  …  5.028  2.235  3.911  6.983  10.615  6.983  1.117  2.235   6.145
999  …  5.117  1.608  3.509  4.532   8.918  2.632  1.023  3.070   4.678

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

[1000 rows x 21 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      8.135655     5.965433     3.795652     5.786065     1.501478
std       2.334120     2.206399     1.551986     2.089206     1.205164
min       1.667000     0.000000     0.000000     0.000000     0.000000
25%       6.624750     4.540250     2.861500     4.643750     0.732250
50%       7.904000     5.882000     3.704000     5.818000     1.299000
75%       9.443000     7.149750     4.549750     6.855250     2.000000
max      26.531000    20.175000    14.754000    16.667000    12.281000

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      6.196397     3.911379     6.615316     2.557886     5.274027
std       2.844936     1.690500     2.170946     1.343486     1.832582
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       4.827250     2.889250     5.097500     1.774000     4.166250
50%       6.024000     3.804500     6.392000     2.439000     5.270500
```

```
75%       7.319750     4.663750     8.000000     3.156500     6.302250
max      60.920000    26.593000    15.385000    18.062000    16.667000

                X10          X11          X12          X13          X14  \
count   1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean       8.946782     4.994128     2.336929     3.809883     5.616647
std        2.598105     2.260369     1.208988     1.554498     2.160142
min        1.322000     0.000000     0.000000     0.000000     0.000000
25%        7.398750     3.653250     1.608750     2.794750     4.261750
50%        8.982000     4.747000     2.175000     3.759500     5.353500
75%       10.485500     6.061000     2.837750     4.594000     6.631250
max       26.667000    25.446000    13.934000    11.409000    16.410000

                X15          X16          X17          X18          X19
count   1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean       8.077016     5.929471     1.607529     2.873559     6.068837
std        2.657465     2.107250     0.980726     1.293227     1.890959
min        0.000000     0.000000     0.000000     0.000000     0.441000
25%        6.482500     4.805250     0.945250     2.061250     4.833750
50%        7.624000     5.781000     1.484000     2.803500     5.950500
75%        9.355250     6.744750     2.216000     3.584500     7.143000
max       23.636000    24.165000     7.692000    11.013000    14.286000
```





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

         X0     X1     X2     X3     X4     X5     X6     X7     X8     X9  \
0     9.695  4.668  3.591  6.463  2.154  6.284  3.950  5.745  3.232  4.668
1     9.761  6.773  3.785  5.378  2.590  5.578  2.988  4.980  2.390  6.175
2     8.048  5.231  2.414  4.225  1.207  3.823  4.024  5.231  2.414  3.823
3     6.966  7.191  3.446  6.667  1.423  7.491  3.596  6.067  3.221  4.419
4     8.333  7.540  1.190  7.143  0.794  8.333  2.381  7.143  2.381  1.984
..      …      …      …      …      …      …      …      …      …      …
995   4.762  6.667  3.492  8.254  0.952  5.397  1.905  6.032  3.492  5.079
996  13.971  2.206  4.412  3.676  1.471  3.676  3.676  5.882  4.412  5.147
997   9.714  5.429  3.000  4.000  1.143  3.714  3.571  8.000  1.857  8.571
998   8.255  3.066  6.604  5.660  0.943  3.774  4.481  7.783  2.123  4.009
999   3.356  7.383  2.685  6.040  1.342  6.711  6.711  6.040  4.027  6.040

     …    X11    X12    X13     X14     X15    X16    X17    X18    X19  \
0    …  4.309  3.591  4.309   6.643   8.079  7.181  1.795  2.154  3.950
1    …  2.789  2.988  4.582   5.777   8.167  7.371  1.992  3.386  5.179
2    …  3.219  1.610  4.829  13.481  11.469  8.048  1.207  1.811  6.439
3    …  3.745  1.798  5.019   6.517   8.315  5.843  1.273  3.071  7.116
4    …  2.381  2.778  1.984   9.921   9.127  4.365  1.587  0.397  8.333
..   …    …      …      …       …       …      …      …      …      …
995  …  4.444  2.222  3.492   7.619   5.714  6.032  3.175  2.857  7.937
```

```
996  …  6.618  2.941  2.941   7.353   6.618  6.618  0.000  2.941  5.147
997  …  3.000  3.143  5.286   4.571   7.429  5.714  2.000  4.714  5.857
998  …  4.953  1.887  4.245   3.538   6.840  7.783  4.481  4.009  6.840
999  …  3.356  2.013  2.685   7.383  12.752  4.027  1.342  2.685  4.698

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

[834 rows x 21 columns]
```

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

```
                X0          X1          X2          X3         X4          X5  \
count  834.000000  834.000000  834.000000  834.000000  834.00000  834.000000
mean     8.346590    5.809586    3.777675    5.752608    1.43174    6.074753
std      2.047942    1.947347    1.248814    1.652728    0.96829    2.042940
min      1.829000    0.000000    0.000000    0.000000    0.00000    0.000000
25%      6.998500    4.429000    2.976250    4.728250    0.73925    4.558500
50%      8.089500    5.900000    3.656500    5.777000    1.29450    6.055500
75%      9.619000    7.012500    4.531750    6.811250    1.95850    7.317000
max     15.278000   11.950000    8.092000   11.377000    5.55600   12.821000

                X6          X7          X8          X9         X10         X11  \
count  834.000000  834.000000  834.000000  834.000000  834.000000  834.000000
mean     3.948362    6.759604    2.477499    5.342940    9.001679    4.938555
std      1.426883    2.061771    1.003926    1.569959    2.229176    1.852721
min      0.000000    0.000000    0.000000    0.669000    1.942000    0.000000
25%      2.988000    5.290000    1.799750    4.313250    7.583250    3.652500
50%      3.795000    6.604000    2.455000    5.233000    8.950000    4.821500
75%      4.687500    7.918750    3.125000    6.341750   10.415250    6.050500
max      9.578000   13.559000    6.000000   10.377000   16.000000   11.142000

               X12         X13         X14         X15         X16         X17  \
count  834.000000  834.000000  834.000000  834.000000  834.000000  834.000000
mean     2.249841    3.868940    5.667859    7.968163    5.904912    1.604532
```

std	0.874448	1.339687	1.947794	2.387680	1.543800	0.831894
min	0.000000	0.000000	0.000000	0.877000	1.258000	0.000000
25%	1.669000	2.986000	4.352250	6.345250	4.986250	1.007000
50%	2.209000	3.782000	5.491000	7.692000	5.792500	1.534500
75%	2.778000	4.687000	6.646000	9.135250	6.685250	2.165750
max	5.618000	8.173000	13.481000	16.667000	11.765000	4.481000

	X18	X19
count	834.000000	834.000000
mean	2.851891	6.222386
std	1.127340	1.630364
min	0.000000	0.833000
25%	2.127250	5.106750
50%	2.778000	6.102000
75%	3.535500	7.281250
max	6.614000	11.856000

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

	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9	\
0	6.273	5.351	2.768	4.428	1.661	6.089	4.244	6.642	2.583	6.458	
1	9.172	5.478	4.076	4.841	1.529	5.987	3.694	7.516	2.038	4.713	
2	7.705	6.616	3.183	6.198	1.675	6.449	4.523	5.779	2.345	5.025	
3	8.908	6.226	4.262	7.615	1.964	6.418	2.826	5.987	3.161	5.029	
4	9.651	8.311	2.547	4.558	1.609	5.362	7.373	6.434	2.011	5.228	
..	…	…	…	…	…	…	…	…	…	…	
995	6.462	5.654	5.170	7.593	1.292	6.139	4.200	6.947	3.554	5.977	
996	8.237	6.590	3.789	5.766	2.471	6.096	4.448	4.448	1.977	5.931	
997	8.203	3.125	3.516	5.078	1.562	1.953	3.516	10.156	1.953	5.469	
998	5.866	7.821	2.235	6.704	1.955	5.587	3.911	5.028	3.073	3.631	
999	7.164	5.848	2.632	7.018	3.216	6.579	5.702	6.871	3.070	5.848	

	…	X11	X12	X13	X14	X15	X16	X17	X18	X19	\
0	…	5.535	1.845	5.166	6.827	5.904	7.011	2.030	2.952	4.428	
1	…	4.076	1.911	4.076	5.605	7.261	5.987	1.019	1.911	8.153	
2	…	3.769	1.926	3.518	5.611	9.883	7.705	1.005	2.010	5.695	
3	…	4.598	1.724	2.921	5.029	6.034	4.646	1.437	3.257	5.603	
4	…	1.072	2.011	3.083	5.496	9.651	5.630	1.072	2.681	5.898	
..	…	…	…	…	…	…	…	…	…	…	
995	…	5.331	2.423	2.746	5.977	4.685	4.200	2.100	3.393	7.916	
996	…	3.954	3.130	3.624	5.107	8.896	5.601	1.812	2.471	4.942	
997	…	3.516	2.344	7.422	2.734	9.766	6.641	3.125	1.953	10.156	
998	…	5.028	2.235	3.911	6.983	10.615	6.983	1.117	2.235	6.145	

```
999  …  5.117  1.608  3.509  4.532   8.918  2.632  1.023  3.070   4.678

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

[864 rows x 21 columns]
```

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

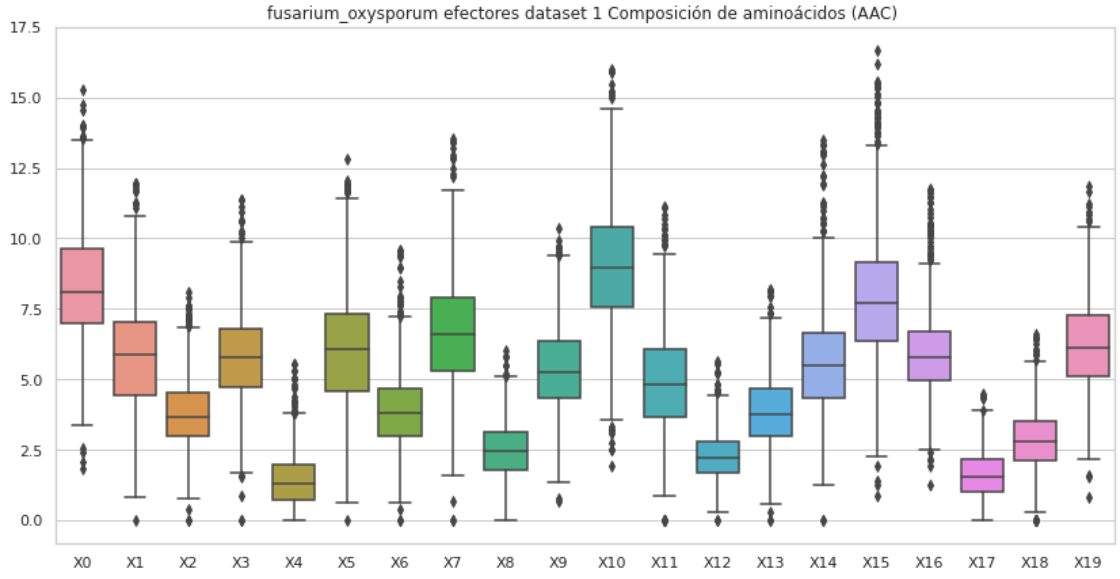
	X0	X1	X2	X3	X4	X5	\
count	864.000000	864.000000	864.000000	864.000000	864.000000	864.000000	
mean	8.097862	5.938706	3.744139	5.895795	1.402397	6.181601	
std	2.045032	1.894779	1.255016	1.732129	0.894509	1.957562	
min	2.410000	1.374000	0.000000	0.000000	0.000000	0.000000	
25%	6.667000	4.661000	2.914750	4.910000	0.757000	4.971250	
50%	7.904000	5.886000	3.724500	5.929000	1.282000	6.075500	
75%	9.375000	7.108750	4.500500	6.877250	1.923500	7.310750	
max	14.701000	12.290000	8.370000	11.828000	4.959000	13.492000	

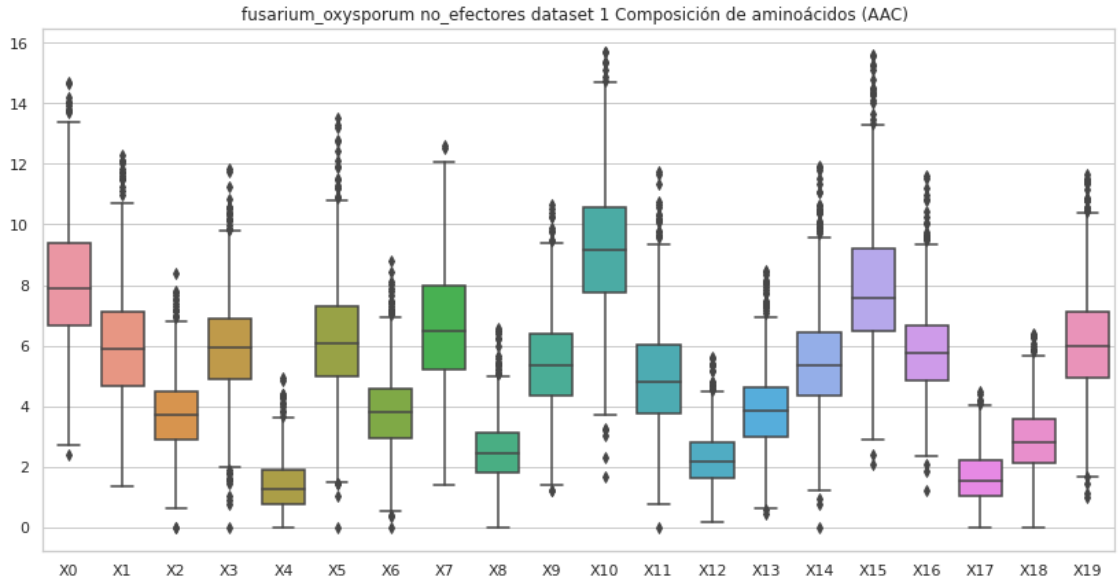
	X6	X7	X8	X9	X10	X11	\
count	864.000000	864.000000	864.000000	864.000000	864.000000	864.000000	
mean	3.841056	6.637844	2.527962	5.406657	9.170638	4.987685	
std	1.295858	1.959372	1.057923	1.544206	2.185428	1.787734	
min	0.000000	1.408000	0.000000	1.242000	1.695000	0.000000	
25%	2.941000	5.227500	1.835000	4.367750	7.754250	3.755250	
50%	3.817000	6.473000	2.459500	5.371500	9.149500	4.825000	
75%	4.582750	7.982250	3.126750	6.374000	10.550750	6.027750	
max	8.802000	12.590000	6.557000	10.638000	15.686000	11.728000	

	X12	X13	X14	X15	X16	X17	\
count	864.000000	864.000000	864.000000	864.000000	864.000000	864.000000	
mean	2.252462	3.898669	5.542385	7.944096	5.828351	1.659498	
std	0.910859	1.365044	1.827683	2.231921	1.483898	0.874172	
min	0.174000	0.437000	0.000000	2.105000	1.242000	0.000000	
25%	1.645000	3.003000	4.332750	6.497500	4.874250	1.036000	

50%	2.161000	3.846000	5.357500	7.574500	5.779500	1.533000
75%	2.791500	4.607750	6.455000	9.215000	6.667000	2.233000
max	5.634000	8.462000	11.945000	15.612000	11.628000	4.481000

	X18	X19
count	864.000000	864.000000
mean	2.899034	6.143234
std	1.127282	1.748396
min	0.000000	1.000000
25%	2.150250	4.947000
50%	2.817000	5.985000
75%	3.587750	7.130250
max	6.383000	11.667000





3 Composición de pseudo aminoácidos (PseAAC) hidro_mass

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

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

    if etiq == "efectores":
        df=PseAAC_hidro_mass_efec

    if etiq == "no_efectores":
        df=PseAAC_hidro_mass_no_efec

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

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

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

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.045680  0.010151  0.030454  0.029608  0.020302  0.027070  0.015227
1    0.040306  0.010693  0.022210  0.023032  0.018919  0.020564  0.009871
2    0.031153  0.004673  0.016355  0.014798  0.018692  0.020249  0.009346
3    0.036734  0.007505  0.035154  0.039499  0.026464  0.031994  0.016985
4    0.029328  0.002793  0.025138  0.029328  0.006983  0.025138  0.008379
..        …         …         …         …         …         …         …
995  0.044570  0.008914  0.077255  0.050513  0.032685  0.056455  0.032685
996  0.054663  0.005754  0.014385  0.014385  0.011508  0.023016  0.017262
997  0.026366  0.003102  0.010856  0.010081  0.014346  0.021713  0.005041
998  0.060157  0.006875  0.041250  0.027500  0.030938  0.056719  0.015469
999  0.020817  0.008327  0.037471  0.041634  0.016654  0.037471  0.024981

           X7        X8        X9  …       X74       X75       X76  \
0    0.021994  0.020302  0.035529  …  0.003367  0.001473  0.022592
1    0.025500  0.011516  0.030435  …  0.007212  0.008720  0.000661
2    0.014798  0.012461  0.028816  …  0.006883 -0.002083  0.019227
3    0.023305  0.019750  0.035944  …  0.004565  0.014878  0.003703
4    0.006983  0.008379  0.041897  …  0.005206  0.011373  0.013645
..        …         …         …  …  …       …         …         …
995  0.047541  0.041599  0.098054  …  0.098419  0.085387 -0.036418
996  0.020139  0.025893  0.040278  … -0.015971 -0.003453  0.041540
997  0.023264  0.008142  0.025203  …  0.006420  0.004885  0.003445
998  0.029219  0.036094  0.063594  … -0.005605 -0.000779  0.021124
999  0.037471  0.020817  0.054124  … -0.045219 -0.010261 -0.029538

          X77       X78       X79       X80       X81       X82         X83
0    0.005668  0.006364  0.017767 -0.000279  0.009226  0.004546  efectores
1    0.004182 -0.002863  0.030669  0.007158  0.003549  0.013244  efectores
2    0.001168  0.002922  0.022808 -0.012169 -0.003829  0.022709  efectores
3   -0.013315  0.007181  0.012377 -0.002451  0.003787  0.019735  efectores
4   -0.002361 -0.004343  0.001012 -0.000720  0.003318  0.014286  efectores
..        …         …         …         …         …         …         …
995  0.007151 -0.011258  0.027360  0.002222  0.002339 -0.002689  efectores
996 -0.022406 -0.020361  0.039246 -0.007811 -0.006770  0.036252  efectores
```

```
997  0.003242  0.005587  0.016147  0.009287  0.004615  0.016079  efectores
998  0.026352  0.011704  0.045607  0.030011  0.021884  0.017506  efectores
999  0.039752 -0.000358  0.012661 -0.028649  0.003396 -0.015214  efectores

[1000 rows x 84 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.038308     0.008025     0.027566     0.028867     0.018564
std       0.018159     0.017179     0.017103     0.018266     0.012508
min       0.002176     0.000000     0.000000     0.000000     0.000000
25%       0.027433     0.002687     0.016339     0.016983     0.010160
50%       0.035112     0.005702     0.025445     0.025747     0.015790
75%       0.045711     0.009766     0.034902     0.037803     0.024352
max       0.263998     0.461996     0.194325     0.168288     0.094943

                X5           X6           X7           X8           X9  …  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
mean      0.030591     0.012886     0.025117     0.023997     0.042667  …
std       0.016179     0.014291     0.016129     0.017423     0.026076  …
min       0.000000     0.000000     0.000000     0.000000     0.000000  …
25%       0.020994     0.006017     0.015497     0.013645     0.026676  …
50%       0.028669     0.010310     0.022410     0.020928     0.037696  …
75%       0.037441     0.016117     0.031723     0.030371     0.053461  …
max       0.210498     0.263998     0.197998     0.228457     0.315779  …

               X73          X74          X75          X76          X77  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.014655     0.001437     0.005549     0.013822     0.002265
std       0.019802     0.021951     0.019086     0.026700     0.027766
min      -0.189473    -0.170832    -0.172615    -0.352890    -0.411469
25%       0.006664    -0.005868    -0.002093     0.006488    -0.005783
50%       0.014969     0.002975     0.006043     0.015343     0.003701
75%       0.024052     0.011174     0.014599     0.024749     0.012410
max       0.189895     0.125549     0.085387     0.106662     0.341279

               X78          X79          X80          X81          X82
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.006323     0.014524     0.001513     0.006590     0.014790
std       0.023084     0.025895     0.024699     0.023066     0.021676
min      -0.185480    -0.325844    -0.232766    -0.218881    -0.249953
25%      -0.001540     0.006886    -0.005863    -0.001966     0.006190
50%       0.005977     0.015728     0.002790     0.006502     0.015213
```

```
75%       0.015215     0.024758     0.011541     0.015542     0.024773
max       0.257550     0.112438     0.245569     0.276251     0.241422

[8 rows x 83 columns]



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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.049743  0.013167  0.035112  0.048280  0.040965  0.052669  0.020482
1    0.038723  0.006454  0.020437  0.025277  0.017210  0.031731  0.008605
2    0.038162  0.008296  0.030696  0.031940  0.017422  0.028622  0.011615
3    0.059251  0.013061  0.050650  0.042686  0.019432  0.039819  0.021025
4    0.043250  0.007208  0.020423  0.024028  0.013816  0.028833  0.009010
..        …         …         …         …         …         …         …
995  0.041326  0.008265  0.048558  0.039260  0.017564  0.044425  0.022729
996  0.041364  0.012409  0.028955  0.030609  0.018200  0.022337  0.009927
997  0.021306  0.004058  0.013189  0.005073  0.019277  0.026378  0.005073
998  0.024119  0.008039  0.027564  0.022970  0.016079  0.020673  0.012634
999  0.039810  0.017874  0.038998  0.036561  0.019499  0.038186  0.017062

           X7        X8        X9  …       X74       X75       X76  \
0    0.051206  0.043891  0.093633  …  0.001043 -0.010814 -0.005533
1    0.019899  0.017210  0.046252  …  0.002264 -0.000205  0.011518
2    0.024888  0.018666  0.046458  … -0.003677  0.004750  0.017194
3    0.033448  0.030581  0.082187  … -0.013545 -0.004012  0.020692
4    0.023427  0.004806  0.046253  …  0.001464  0.008069  0.022335
..        …         …         …  …         …         …         …
995  0.038226  0.034094  0.052691  …  0.024979  0.033473  0.016869
996  0.029782  0.019855  0.053773  … -0.000463  0.010406  0.013923
997  0.014204  0.009131  0.020291  …  0.019675  0.004217  0.022476
998  0.014931  0.020673  0.036752  …  0.011113  0.025100  0.004690
999  0.032498  0.028436  0.060934  … -0.008223 -0.011439  0.009160

           X77       X78       X79       X80       X81       X82           X83
0    0.013487  0.004899  0.016598  0.004444  0.012179 -0.018620  no_efectores
1    0.007245  0.003308  0.033182 -0.004497  0.002319  0.023617  no_efectores
2   -0.010594 -0.003249  0.022968 -0.007016  0.006238  0.006325  no_efectores
3   -0.013916 -0.009397  0.012042  0.012702  0.015477  0.015615  no_efectores
4    0.009234  0.007074  0.019444 -0.015352 -0.012608  0.011500  no_efectores
..        …         …         …         …         …         …           …
995  0.005947  0.015799  0.005160 -0.020404 -0.001514  0.018368  no_efectores
996 -0.002606  0.014136  0.022630  0.031651  0.010800  0.018021  no_efectores
997  0.006145  0.001802  0.008242 -0.001239 -0.002545  0.000868  no_efectores
```

```
998  0.038792  0.035912  0.016810 -0.008366  0.009581  0.020606  no_efectores
999 -0.009087  0.004675  0.028934  0.000475  0.021671  0.025996  no_efectores

[1000 rows x 84 columns]
```

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

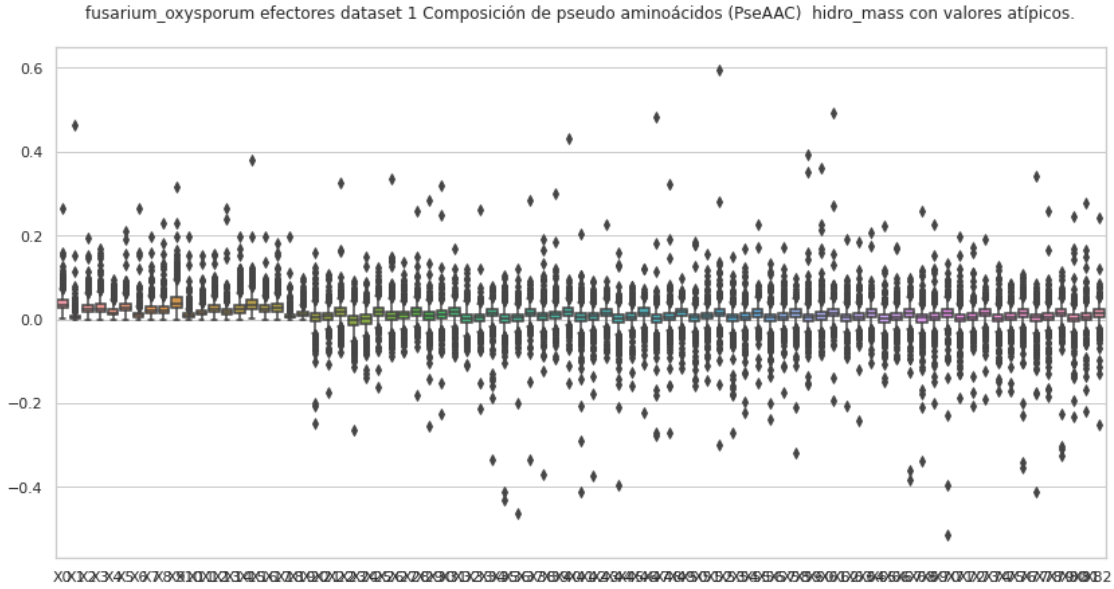
	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.038452	0.007687	0.028454	0.030064	0.019419	
std	0.017998	0.007482	0.016396	0.016942	0.013985	
min	0.002071	0.000000	0.000000	0.000000	0.000000	
25%	0.027605	0.002894	0.017048	0.018423	0.010950	
50%	0.036270	0.006003	0.026424	0.027491	0.017228	
75%	0.045785	0.010106	0.037033	0.038131	0.024000	
max	0.264176	0.069726	0.173661	0.148599	0.156888	

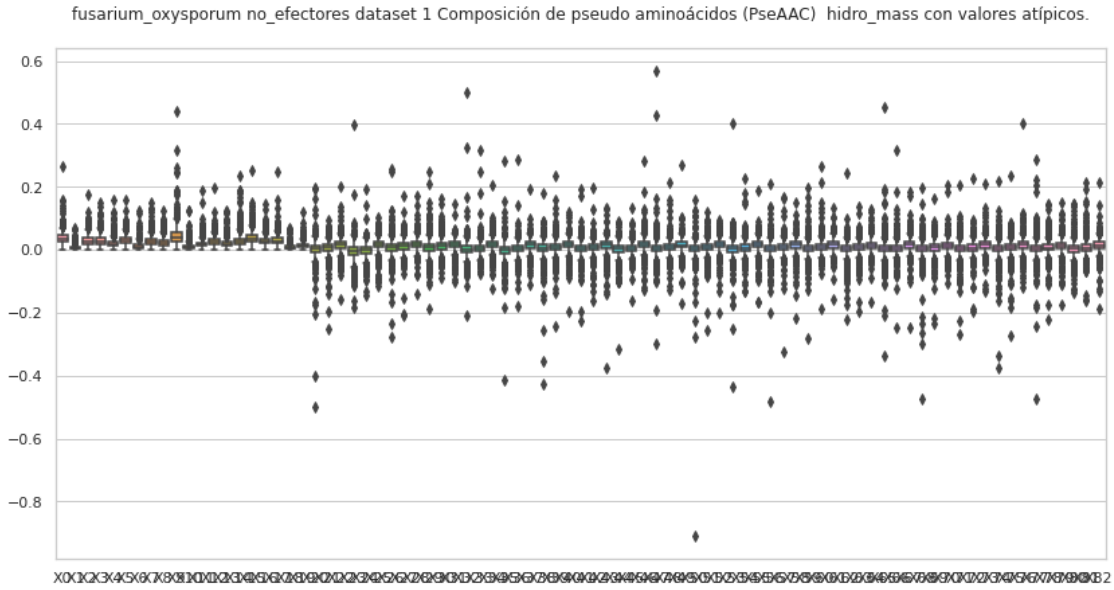
	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.031303	0.013222	0.026334	0.024212	0.045793	…	
std	0.014868	0.010861	0.015549	0.014260	0.031043	…	
min	0.000000	0.000000	0.000000	0.000000	0.001481	…	
25%	0.021487	0.006506	0.016599	0.014329	0.027864	…	
50%	0.029735	0.010861	0.024280	0.022175	0.040676	…	
75%	0.038472	0.017184	0.033117	0.030779	0.057146	…	
max	0.157484	0.124475	0.151444	0.124475	0.439288	…	

	X73	X74	X75	X76	X77	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.015143	0.001555	0.006592	0.015084	0.002334	
std	0.019378	0.027945	0.023503	0.023113	0.031292	
min	-0.132859	-0.377524	-0.274943	-0.119079	-0.473299	
25%	0.006000	-0.006048	-0.001149	0.005513	-0.005911	
50%	0.015246	0.002863	0.006720	0.015193	0.003520	
75%	0.024006	0.011456	0.015664	0.024238	0.012190	
max	0.207902	0.219502	0.236600	0.402963	0.285704	

	X78	X79	X80	X81	X82
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.006429	0.013274	0.003027	0.007562	0.014811
std	0.022863	0.021268	0.023026	0.023500	0.020953
min	-0.222300	-0.176016	-0.136986	-0.161153	-0.189120
25%	-0.000983	0.004626	-0.006467	-0.002475	0.006338
50%	0.006225	0.014577	0.003489	0.006068	0.014877
75%	0.015242	0.023591	0.012530	0.015480	0.024609

```
max        0.145073       0.149056       0.145939       0.215890       0.214830

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.045680	0.010151	0.030454	0.029608	0.020302	0.027070	0.015227	
1	0.040306	0.010693	0.022210	0.023032	0.018919	0.020564	0.009871	
2	0.031153	0.004673	0.016355	0.014798	0.018692	0.020249	0.009346	
3	0.036734	0.007505	0.035154	0.039499	0.026464	0.031994	0.016985	
4	0.029328	0.002793	0.025138	0.029328	0.006983	0.025138	0.008379	
..	…	…	…	…	…	…	…	
994	0.018332	0.005238	0.031426	0.007857	0.000000	0.023570	0.010475	
996	0.054663	0.005754	0.014385	0.014385	0.011508	0.023016	0.017262	
997	0.026366	0.003102	0.010856	0.010081	0.014346	0.021713	0.005041	
998	0.060157	0.006875	0.041250	0.027500	0.030938	0.056719	0.015469	
999	0.020817	0.008327	0.037471	0.041634	0.016654	0.037471	0.024981	

	X7	X8	X9	…	X74	X75	X76	\
0	0.021994	0.020302	0.035529	…	0.003367	0.001473	0.022592	
1	0.025500	0.011516	0.030435	…	0.007212	0.008720	0.000661	
2	0.014798	0.012461	0.028816	…	0.006883	-0.002083	0.019227	
3	0.023305	0.019750	0.035944	…	0.004565	0.014878	0.003703	
4	0.006983	0.008379	0.041897	…	0.005206	0.011373	0.013645	
..	…	…	…	…	…	…	…	
994	0.005238	0.005238	0.060234	…	0.004526	-0.003257	-0.000685	
996	0.020139	0.025893	0.040278	…	-0.015971	-0.003453	0.041540	
997	0.023264	0.008142	0.025203	…	0.006420	0.004885	0.003445	
998	0.029219	0.036094	0.063594	…	-0.005605	-0.000779	0.021124	
999	0.037471	0.020817	0.054124	…	-0.045219	-0.010261	-0.029538	

	X77	X78	X79	X80	X81	X82	X83
0	0.005668	0.006364	0.017767	-0.000279	0.009226	0.004546	efectores
1	0.004182	-0.002863	0.030669	0.007158	0.003549	0.013244	efectores
2	0.001168	0.002922	0.022808	-0.012169	-0.003829	0.022709	efectores
3	-0.013315	0.007181	0.012377	-0.002451	0.003787	0.019735	efectores
4	-0.002361	-0.004343	0.001012	-0.000720	0.003318	0.014286	efectores
..	…	…	…	…	…	…	
994	0.032256	0.033613	0.006242	0.012924	0.021913	0.014784	efectores
996	-0.022406	-0.020361	0.039246	-0.007811	-0.006770	0.036252	efectores
997	0.003242	0.005587	0.016147	0.009287	0.004615	0.016079	efectores
998	0.026352	0.011704	0.045607	0.030011	0.021884	0.017506	efectores
999	0.039752	-0.000358	0.012661	-0.028649	0.003396	-0.015214	efectores

[881 rows x 84 columns]

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

	X0	X1	X2	X3	X4	X5	\
count	881.000000	881.000000	881.000000	881.000000	881.000000	881.000000	
mean	0.035421	0.006506	0.024764	0.025767	0.016489	0.028363	
std	0.012183	0.005349	0.012374	0.012785	0.009003	0.011062	
min	0.002176	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.027017	0.002740	0.015530	0.016209	0.009755	0.020479	
50%	0.033356	0.005440	0.023339	0.024086	0.015030	0.027296	
75%	0.043287	0.008951	0.032839	0.034483	0.022439	0.035810	
max	0.081029	0.050653	0.070353	0.066552	0.052487	0.064440	

	X6	X7	X8	X9	…	X73	\
count	881.000000	881.000000	881.000000	881.000000	…	881.000000	
mean	0.011005	0.022676	0.021283	0.038310	…	0.016232	
std	0.006885	0.011088	0.010951	0.018390	…	0.013658	
min	0.000000	0.000000	0.000000	0.000000	…	-0.043648	
25%	0.005863	0.014914	0.013140	0.025033	…	0.008019	
50%	0.009876	0.021622	0.019853	0.036022	…	0.015527	
75%	0.015020	0.029886	0.028082	0.049196	…	0.024050	
max	0.043414	0.073262	0.060150	0.107874	…	0.073094	

	X74	X75	X76	X77	X78	X79	\
count	881.000000	881.000000	881.000000	881.000000	881.000000	881.000000	
mean	0.002896	0.006288	0.015766	0.003239	0.006669	0.016341	
std	0.013966	0.013141	0.013358	0.014855	0.013198	0.013212	
min	-0.060114	-0.042106	-0.039969	-0.065823	-0.035817	-0.036881	
25%	-0.004317	-0.001293	0.007574	-0.004927	-0.000897	0.008171	
50%	0.003367	0.006249	0.015537	0.003659	0.005914	0.015965	
75%	0.010769	0.014347	0.024215	0.011560	0.014312	0.024402	
max	0.050610	0.052454	0.065248	0.065980	0.066242	0.069267	

	X80	X81	X82
count	881.000000	881.000000	881.000000
mean	0.002458	0.006558	0.015788
std	0.014131	0.012912	0.014307
min	-0.054668	-0.041994	-0.047430
25%	-0.005097	-0.001503	0.007131
50%	0.002958	0.006209	0.015563
75%	0.010416	0.014427	0.024585
max	0.054738	0.073853	0.066331

[8 rows x 83 columns]

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.049743  0.013167  0.035112  0.048280  0.040965  0.052669  0.020482
1    0.038723  0.006454  0.020437  0.025277  0.017210  0.031731  0.008605
2    0.038162  0.008296  0.030696  0.031940  0.017422  0.028622  0.011615
3    0.059251  0.013061  0.050650  0.042686  0.019432  0.039819  0.021025
4    0.043250  0.007208  0.020423  0.024028  0.013816  0.028833  0.009010
..        …         …         …         …         …         …         …
995  0.041326  0.008265  0.048558  0.039260  0.017564  0.044425  0.022729
996  0.041364  0.012409  0.028955  0.030609  0.018200  0.022337  0.009927
997  0.021306  0.004058  0.013189  0.005073  0.019277  0.026378  0.005073
998  0.024119  0.008039  0.027564  0.022970  0.016079  0.020673  0.012634
999  0.039810  0.017874  0.038998  0.036561  0.019499  0.038186  0.017062

           X7        X8        X9  …       X74       X75       X76  \
0    0.051206  0.043891  0.093633  …  0.001043 -0.010814 -0.005533
1    0.019899  0.017210  0.046252  …  0.002264 -0.000205  0.011518
2    0.024888  0.018666  0.046458  … -0.003677  0.004750  0.017194
3    0.033448  0.030581  0.082187  … -0.013545 -0.004012  0.020692
4    0.023427  0.004806  0.046253  …  0.001464  0.008069  0.022335
..        …         …         …  …         …         …         …
995  0.038226  0.034094  0.052691  …  0.024979  0.033473  0.016869
996  0.029782  0.019855  0.053773  … -0.000463  0.010406  0.013923
997  0.014204  0.009131  0.020291  …  0.019675  0.004217  0.022476
998  0.014931  0.020673  0.036752  …  0.011113  0.025100  0.004690
999  0.032498  0.028436  0.060934  … -0.008223 -0.011439  0.009160

           X77       X78       X79       X80       X81       X82           X83
0    0.013487  0.004899  0.016598  0.004444  0.012179 -0.018620  no_efectores
1    0.007245  0.003308  0.033182 -0.004497  0.002319  0.023617  no_efectores
2   -0.010594 -0.003249  0.022968 -0.007016  0.006238  0.006325  no_efectores
3   -0.013916 -0.009397  0.012042  0.012702  0.015477  0.015615  no_efectores
4    0.009234  0.007074  0.019444 -0.015352 -0.012608  0.011500  no_efectores
..        …         …         …         …         …         …           …
995  0.005947  0.015799  0.005160 -0.020404 -0.001514  0.018368  no_efectores
996 -0.002606  0.014136  0.022630  0.031651  0.010800  0.018021  no_efectores
997  0.006145  0.001802  0.008242 -0.001239 -0.002545  0.000868  no_efectores
998  0.038792  0.035912  0.016810 -0.008366  0.009581  0.020606  no_efectores
999 -0.009087  0.004675  0.028934  0.000475  0.021671  0.025996  no_efectores

[858 rows x 84 columns]
```

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

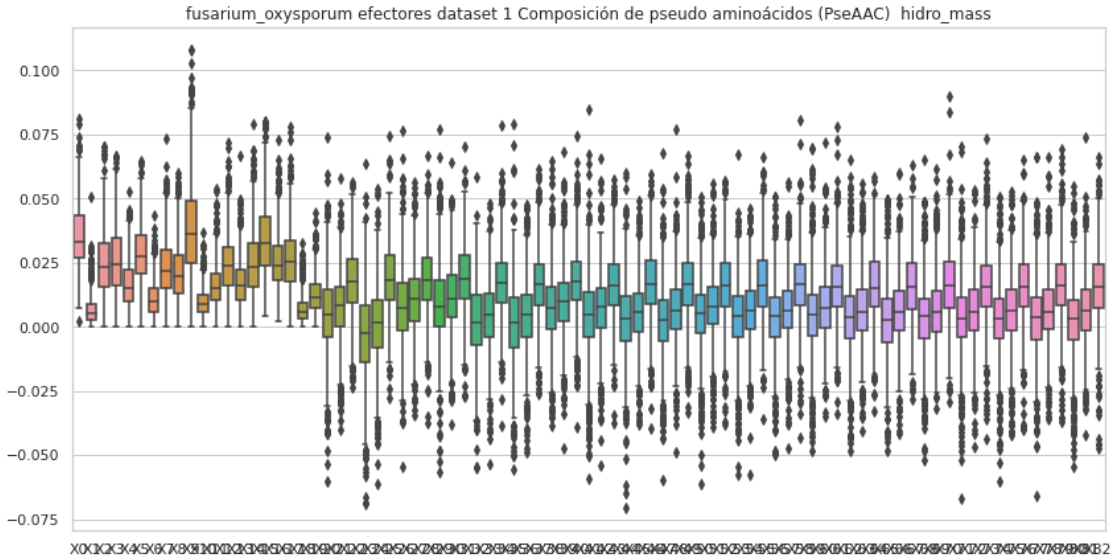
	X0	X1	X2	X3	X4	X5	\
count	858.000000	858.000000	858.000000	858.000000	858.000000	858.000000	
mean	0.035013	0.006526	0.026087	0.027246	0.017113	0.029117	
std	0.011806	0.005136	0.012574	0.013022	0.008939	0.011569	
min	0.002071	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.026459	0.002765	0.016647	0.017693	0.010524	0.020907	
50%	0.034589	0.005577	0.025335	0.026055	0.016452	0.028171	
75%	0.042566	0.008877	0.034364	0.034759	0.022256	0.036038	
max	0.082392	0.029366	0.072334	0.073069	0.053504	0.071937	

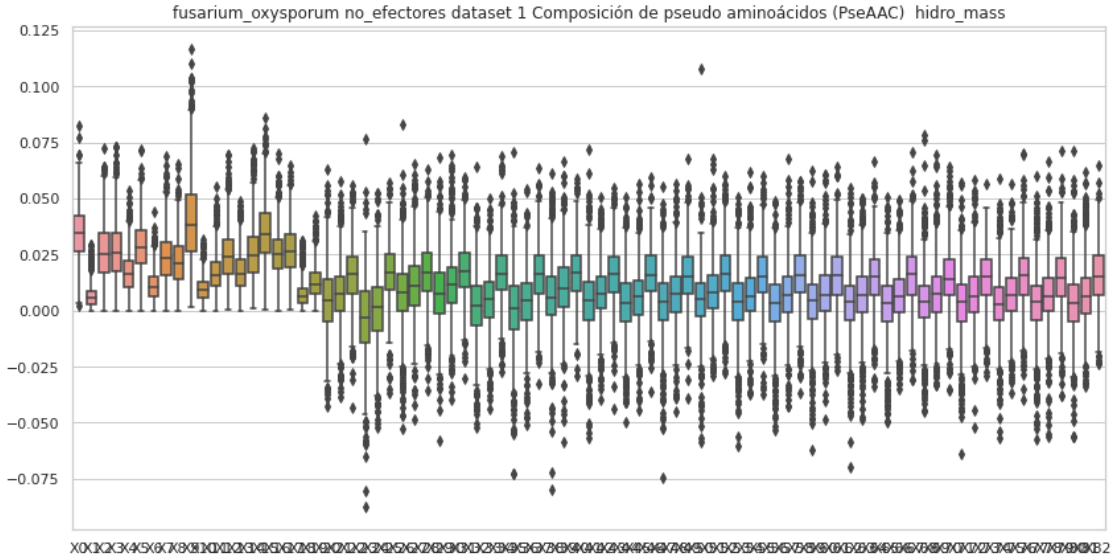
	X6	X7	X8	X9	…	X73	\
count	858.000000	858.000000	858.000000	858.000000	…	858.000000	
mean	0.011320	0.023587	0.021856	0.040212	…	0.014988	
std	0.006871	0.011060	0.010654	0.019395	…	0.012673	
min	0.000000	0.000000	0.000000	0.001481	…	-0.029324	
25%	0.006166	0.015520	0.013826	0.026188	…	0.006889	
50%	0.010250	0.023337	0.021168	0.038160	…	0.015314	
75%	0.015297	0.030563	0.028718	0.051780	…	0.023124	
max	0.044347	0.068803	0.065414	0.116588	…	0.056472	

	X74	X75	X76	X77	X78	X79	\
count	858.000000	858.000000	858.000000	858.000000	858.000000	858.000000	
mean	0.002824	0.007134	0.015139	0.003261	0.007061	0.014648	
std	0.013562	0.012597	0.013381	0.014401	0.013810	0.013198	
min	-0.056525	-0.036897	-0.039590	-0.057217	-0.056011	-0.037939	
25%	-0.004324	-0.000422	0.006710	-0.004105	-0.000109	0.006048	
50%	0.002958	0.006691	0.015542	0.003875	0.006555	0.014837	
75%	0.010669	0.014786	0.023387	0.010953	0.014326	0.023213	
max	0.058870	0.056355	0.070073	0.059887	0.066499	0.071151	

	X80	X81	X82
count	858.000000	858.000000	858.000000
mean	0.003467	0.007201	0.015566
std	0.015010	0.014128	0.013032
min	-0.056370	-0.034034	-0.023825
25%	-0.004858	-0.001272	0.007129
50%	0.003592	0.006130	0.015111
75%	0.011569	0.014313	0.024321
max	0.071499	0.061820	0.064701

[8 rows x 83 columns]





4 Composición de pseudo aminoácidos (PseAAC) mass

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

```
comp = "mass"
df=""

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

    if etiq == "efectores":
        df=PseAAC_mass_efec

    if etiq == "no_efectores":
        df=PseAAC_mass_no_efec

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

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

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.052569  0.011682  0.035046  0.034073  0.023364  0.031152  0.017523
1    0.059660  0.015828  0.032874  0.034092  0.028004  0.030439  0.014611
2    0.031538  0.004731  0.016557  0.014980  0.018923  0.020499  0.009461
3    0.044399  0.009071  0.042489  0.047741  0.031986  0.038670  0.020529
4    0.040410  0.003849  0.034637  0.040410  0.009621  0.034637  0.011546
..        …         …         …         …         …         …         …
995  0.039253  0.007851  0.068039  0.044487  0.028786  0.049721  0.028786
996  0.051662  0.005438  0.013595  0.013595  0.010876  0.021752  0.016314
997  0.053111  0.006248  0.021869  0.020307  0.028899  0.043739  0.010154
998  0.056753  0.006486  0.038916  0.025944  0.029187  0.053510  0.014594
999  0.028715  0.011486  0.051686  0.057429  0.022972  0.051686  0.034457

           X7        X8        X9  …       X32       X33       X34  \
0    0.025311  0.023364  0.040887  …  0.020324  0.027602  0.025310
```

```
1    0.037744  0.017046  0.045050  …  0.026989  0.007544  0.022017
2    0.014980  0.012615  0.029172  …  0.017832  0.022694  0.035640
3    0.028167  0.023870  0.043444  …  0.036983  0.040307  0.011801
4    0.009621  0.011546  0.057729  …  0.032855  0.021110  0.037606
..        …         …         …  …         …         …         …
995  0.041870  0.036636  0.086357  …  0.023692 -0.022364  0.003516
996  0.019033  0.024471  0.038067  …  0.015245  0.049615  0.036711
997  0.046863  0.016402  0.050768  …  0.014353  0.029749  0.013100
998  0.027566  0.034052  0.059996  …  0.004621 -0.000249  0.015086
999  0.051686  0.028715  0.074658  …  0.021772  0.026301  0.028472

          X35       X36       X37       X38       X39       X40         X41
0    0.034894  0.003172  0.010621  0.025999  0.020447  0.005232  efectores
1    0.005560  0.004994  0.034774  0.000978  0.045395  0.019603  efectores
2    0.028491  0.021740  0.023812  0.019464  0.023090  0.022989  efectores
3    0.005194  0.023540  0.009772  0.004475  0.014959  0.023853  efectores
4    0.013116  0.034813  0.011904  0.018801  0.001395  0.019685  efectores
..        …         …         …         …         …         …         …
995  0.025302  0.019493  0.029452 -0.032074  0.024096 -0.002368  efectores
996  0.013372  0.018592  0.014342  0.039260  0.037091  0.034262  efectores
997  0.027060  0.033844  0.030305  0.006939  0.032527  0.032390  efectores
998  0.011470  0.033757  0.016663  0.019929  0.043027  0.016516  efectores
999 -0.019900 -0.033260 -0.047196 -0.040744  0.017464 -0.020985  efectores

[1000 rows x 42 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.047756     0.009414     0.034572     0.036687     0.022954
std       0.015920     0.012897     0.018958     0.021030     0.013958
min       0.009304     0.000000     0.000000     0.000000     0.000000
25%       0.038073     0.003675     0.021548     0.021309     0.013783
50%       0.045662     0.007292     0.032605     0.033525     0.020988
75%       0.055106     0.012199     0.044052     0.047392     0.029228
max       0.161476     0.268033     0.182532     0.156543     0.182532

                X5           X6           X7           X8           X9  …  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
mean      0.037650     0.015418     0.030805     0.030484     0.052518  …
std       0.014878     0.011177     0.015322     0.019390     0.023731  …
min       0.000000     0.000000     0.000000     0.000000     0.000000  …
25%       0.029989     0.008523     0.020590     0.018371     0.036974  …
50%       0.037065     0.013736     0.029762     0.027070     0.050096  …
```

```
75%       0.043551     0.019863     0.038196     0.037715     0.066047  …
max       0.243376     0.153162     0.114871     0.243376     0.174688  …

               X31          X32          X33          X34          X35  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.019350     0.018236     0.019373     0.019099     0.017075
std       0.020615     0.023011     0.022967     0.021082     0.022437
min      -0.083951    -0.185589    -0.137841    -0.161985    -0.221637
25%       0.009630     0.009977     0.010121     0.009680     0.007178
50%       0.021607     0.020866     0.020705     0.020671     0.019619
75%       0.030542     0.030888     0.029692     0.030746     0.030141
max       0.281355     0.161422     0.285912     0.180386     0.090743

               X36          X37          X38          X39          X40
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.017502     0.017718     0.017579     0.018443     0.018084
std       0.029981     0.025103     0.030802     0.028214     0.023295
min      -0.423026    -0.415171    -0.375054    -0.496226    -0.185642
25%       0.008436     0.009654     0.008865     0.009691     0.008642
50%       0.020429     0.019862     0.020757     0.021257     0.020363
75%       0.030732     0.029503     0.029752     0.030634     0.030485
max       0.133225     0.155242     0.465619     0.215501     0.139308

[8 rows x 41 columns]
```

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

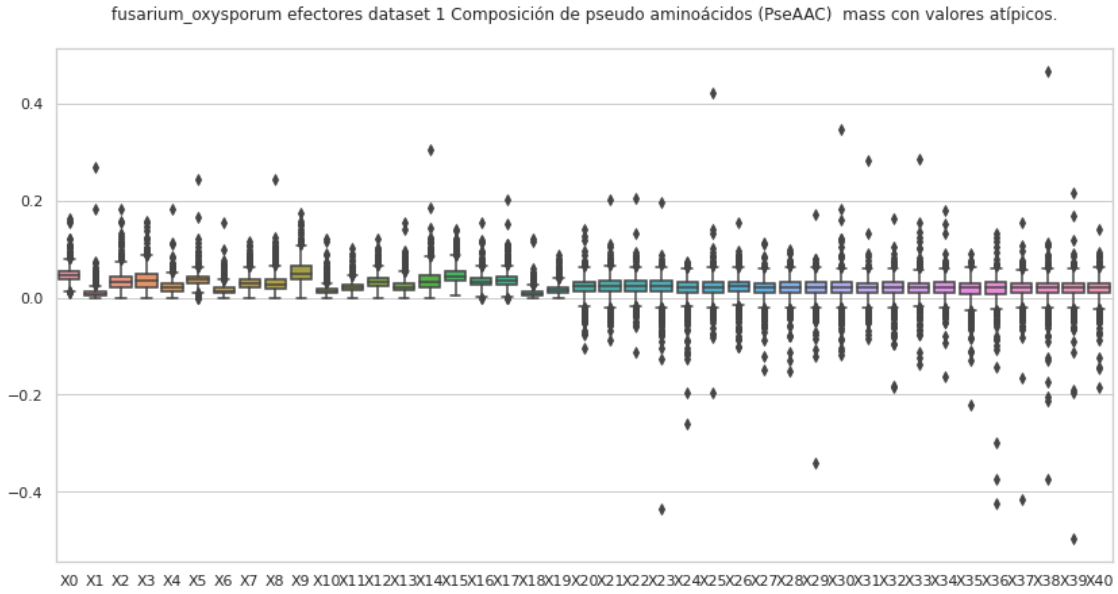
```
          X0        X1        X2        X3        X4        X5        X6  \
0    0.048270  0.012777  0.034073  0.046850  0.039752  0.051110  0.019876
1    0.040342  0.006724  0.021292  0.026334  0.017930  0.033058  0.008965
2    0.041988  0.009128  0.033773  0.035142  0.019168  0.031491  0.012779
3    0.056687  0.012495  0.048458  0.040839  0.018591  0.038096  0.020115
4    0.051002  0.008500  0.024084  0.028334  0.016292  0.034001  0.010625
..        …         …         …         …         …         …         …
995  0.047256  0.009451  0.055525  0.044893  0.020084  0.050800  0.025991
996  0.050567  0.015170  0.035397  0.037419  0.022249  0.027306  0.012136
997  0.031073  0.005919  0.019236  0.007398  0.028114  0.038471  0.007398
998  0.032697  0.010899  0.037368  0.031140  0.021798  0.028026  0.017127
999  0.044810  0.020119  0.043895  0.041152  0.021948  0.042981  0.019204

          X7        X8        X9  …       X32       X33       X34  \
0    0.049690  0.042591  0.090861  …  0.012316  0.015010 -0.003004
1    0.020731  0.017930  0.048186  …  0.022652  0.037650  0.032439
```

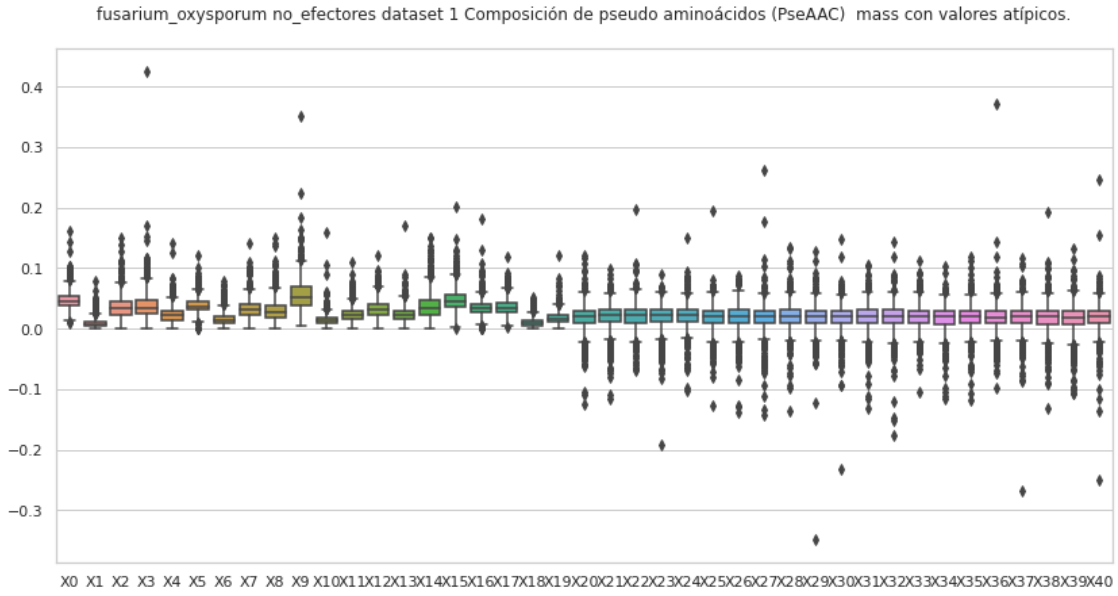
```
2    0.027383  0.020537  0.051115  …  0.022069  0.022185  0.020048
3    0.032001  0.029258  0.078630  …  0.016160  0.014310  0.011372
4    0.027626  0.005667  0.054544  …  0.019859  0.012643  0.026760
..        …         …        …  …        …         …
995  0.043711  0.038986  0.060251  …  0.007430 -0.006872 -0.012104
996  0.036408  0.024272  0.065737  …  0.010548  0.021692  0.009384
997  0.020715  0.013317  0.029593  …  0.037351  0.037692  0.023374
998  0.020241  0.028026  0.049824  …  0.010641  0.020914  0.013303
999  0.036579  0.032007  0.068586  …  0.012760  0.028378  0.025470

          X35       X36       X37       X38       X39       X40           X41
0    0.030079  0.013662  0.025428 -0.005369  0.016107 -0.018069  no_efectores
1    0.027158  0.023240  0.024295  0.011999  0.034569  0.024605  no_efectores
2    0.033996  0.013176  0.006771  0.018918  0.025270  0.006959  no_efectores
3    0.013533  0.027176  0.026701  0.019796  0.011521  0.014939  no_efectores
4    0.021274  0.028430  0.008132  0.026338  0.022929  0.013561  no_efectores
..        …         …         …         …         …         …           …
995  0.011054  0.004931  0.016019  0.019289  0.005901  0.021003  no_efectores
996  0.019689  0.030153  0.013776  0.017020  0.027664  0.022030  no_efectores
997  0.031879  0.038205  0.042779  0.032780  0.012020  0.001266  no_efectores
998 -0.010140 -0.000603  0.044190  0.006358  0.022789  0.027936  no_efectores
999  0.023151  0.030378  0.030347  0.010310  0.032568  0.029261  no_efectores

[1000 rows x 42 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.047390     0.009294     0.035784     0.038858     0.023690
std       0.014467     0.008200     0.018461     0.024478     0.013613
min       0.008559     0.000000     0.000000     0.000000     0.000000
25%       0.038626     0.003825     0.023345     0.023989     0.014472
50%       0.045223     0.007724     0.033906     0.034546     0.021798
75%       0.054750     0.012397     0.044806     0.048288     0.030248
max       0.162220     0.078977     0.149855     0.425052     0.141942

                X5           X6           X7           X8           X9  …  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
mean      0.038287     0.016206     0.032168     0.030904     0.055540  …
std       0.012644     0.010678     0.015053     0.017734     0.026717  …
min       0.000000     0.000000     0.000000     0.000000     0.005796  …
25%       0.030733     0.009061     0.022719     0.019117     0.038455  …
50%       0.037183     0.014268     0.030947     0.028207     0.052181  …
75%       0.044327     0.021132     0.040281     0.039027     0.068843  …
```

```
max       0.121665     0.079210     0.140574     0.150142     0.351027  …

                X31          X32          X33          X34          X35  \
count   1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean       0.018521     0.018383     0.019122     0.017161     0.018371
std        0.021788     0.022580     0.019646     0.020198     0.020522
min       -0.132210    -0.175654    -0.104437    -0.115876    -0.118294
25%        0.009315     0.008929     0.009164     0.007857     0.009136
50%        0.020572     0.020947     0.019752     0.019764     0.020528
75%        0.030821     0.030734     0.030425     0.028889     0.029637
max        0.105862     0.142908     0.111570     0.104360     0.117803

                X36          X37          X38          X39          X40
count   1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean       0.017862     0.017769     0.017893     0.016492     0.018061
std        0.023089     0.021511     0.022415     0.022380     0.023515
min       -0.098926    -0.269008    -0.131605    -0.107128    -0.249410
25%        0.008497     0.009268     0.008063     0.006726     0.009368
50%        0.018548     0.019766     0.019703     0.019047     0.019560
75%        0.028400     0.029702     0.029406     0.029057     0.029967
max        0.371098     0.115585     0.191778     0.132627     0.247044

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.052569  0.011682  0.035046  0.034073  0.023364  0.031152  0.017523
1    0.059660  0.015828  0.032874  0.034092  0.028004  0.030439  0.014611
2    0.031538  0.004731  0.016557  0.014980  0.018923  0.020499  0.009461
3    0.044399  0.009071  0.042489  0.047741  0.031986  0.038670  0.020529
4    0.040410  0.003849  0.034637  0.040410  0.009621  0.034637  0.011546
..        …         …         …         …         …         …         …
994  0.029603  0.008458  0.050747  0.012687  0.000000  0.038060  0.016916
995  0.039253  0.007851  0.068039  0.044487  0.028786  0.049721  0.028786
996  0.051662  0.005438  0.013595  0.013595  0.010876  0.021752  0.016314
997  0.053111  0.006248  0.021869  0.020307  0.028899  0.043739  0.010154
998  0.056753  0.006486  0.038916  0.025944  0.029187  0.053510  0.014594

           X7        X8        X9  …       X32       X33       X34  \
0    0.025311  0.023364  0.040887  …  0.020324  0.027602  0.025310
1    0.037744  0.017046  0.045050  …  0.026989  0.007544  0.022017
2    0.014980  0.012615  0.029172  …  0.017832  0.022694  0.035640
3    0.028167  0.023870  0.043444  …  0.036983  0.040307  0.011801
4    0.009621  0.011546  0.057729  …  0.032855  0.021110  0.037606
..        …         …         …  …  …       …         …         …
994  0.008458  0.008458  0.097266  …  0.028128  0.020299  0.038040
995  0.041870  0.036636  0.086357  …  0.023692 -0.022364  0.003516
996  0.019033  0.024471  0.038067  …  0.015245  0.049615  0.036711
997  0.046863  0.016402  0.050768  …  0.014353  0.029749  0.013100
998  0.027566  0.034052  0.059996  …  0.004621 -0.000249  0.015086
```

	X35	X36	X37	X38	X39	X40	X41
0	0.034894	0.003172	0.010621	0.025999	0.020447	0.005232	efectores
1	0.005560	0.004994	0.034774	0.000978	0.045395	0.019603	efectores
2	0.028491	0.021740	0.023812	0.019464	0.023090	0.022989	efectores
3	0.005194	0.023540	0.009772	0.004475	0.014959	0.023853	efectores
4	0.013116	0.034813	0.011904	0.018801	0.001395	0.019685	efectores
..	…	…	…	…	…	…	
994	0.018503	0.021614	0.020373	-0.001107	0.010079	0.023873	efectores
995	0.025302	0.019493	0.029452	-0.032074	0.024096	-0.002368	efectores
996	0.013372	0.018592	0.014342	0.039260	0.037091	0.034262	efectores
997	0.027060	0.033844	0.030305	0.006939	0.032527	0.032390	efectores
998	0.011470	0.033757	0.016663	0.019929	0.043027	0.016516	efectores

[879 rows x 42 columns]

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

	X0	X1	X2	X3	X4	X5	\
count	879.000000	879.000000	879.000000	879.000000	879.000000	879.000000	
mean	0.045977	0.008239	0.032031	0.034001	0.021440	0.036377	
std	0.012477	0.006338	0.014471	0.017127	0.010803	0.010769	
min	0.009304	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.037921	0.003760	0.021168	0.020609	0.013541	0.029807	
50%	0.044839	0.007097	0.031556	0.032199	0.020243	0.036498	
75%	0.053787	0.011646	0.041507	0.044578	0.028112	0.042769	
max	0.091951	0.046318	0.088158	0.097456	0.061821	0.071642	

	X6	X7	X8	X9	…	X31	\
count	879.000000	879.000000	879.000000	879.000000	…	879.000000	
mean	0.013972	0.029072	0.027630	0.049802	…	0.020957	
std	0.007897	0.012298	0.014361	0.020655	…	0.015199	
min	0.000000	0.000000	0.000000	0.003024	…	-0.036666	
25%	0.008186	0.020581	0.017712	0.035831	…	0.011966	
50%	0.012981	0.029115	0.025566	0.048393	…	0.022584	
75%	0.018781	0.036925	0.034752	0.063495	…	0.031112	
max	0.047761	0.073795	0.083801	0.121364	…	0.076412	

	X32	X33	X34	X35	X36	X37	\
count	879.000000	879.000000	879.000000	879.000000	879.000000	879.000000	
mean	0.020467	0.019833	0.019921	0.019862	0.020446	0.020239	
std	0.015948	0.014741	0.015513	0.016333	0.015884	0.015709	
min	-0.041047	-0.040798	-0.041454	-0.037885	-0.045241	-0.046450	
25%	0.011809	0.011286	0.011024	0.010505	0.010388	0.011822	
50%	0.021873	0.020722	0.021131	0.020694	0.021614	0.020796	
75%	0.031162	0.029016	0.030458	0.030698	0.031000	0.029660	

```
max      0.080890    0.074841    0.068059    0.075782    0.103859    0.070065

              X38         X39         X40
count  879.000000  879.000000  879.000000
mean     0.019818    0.020654    0.020165
std      0.015847    0.015424    0.016381
min     -0.050860   -0.065221   -0.046225
25%      0.010930    0.011552    0.010139
50%      0.021157    0.021899    0.021600
75%      0.029760    0.030460    0.030453
max      0.075301    0.075075    0.080877

[8 rows x 41 columns]
```

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

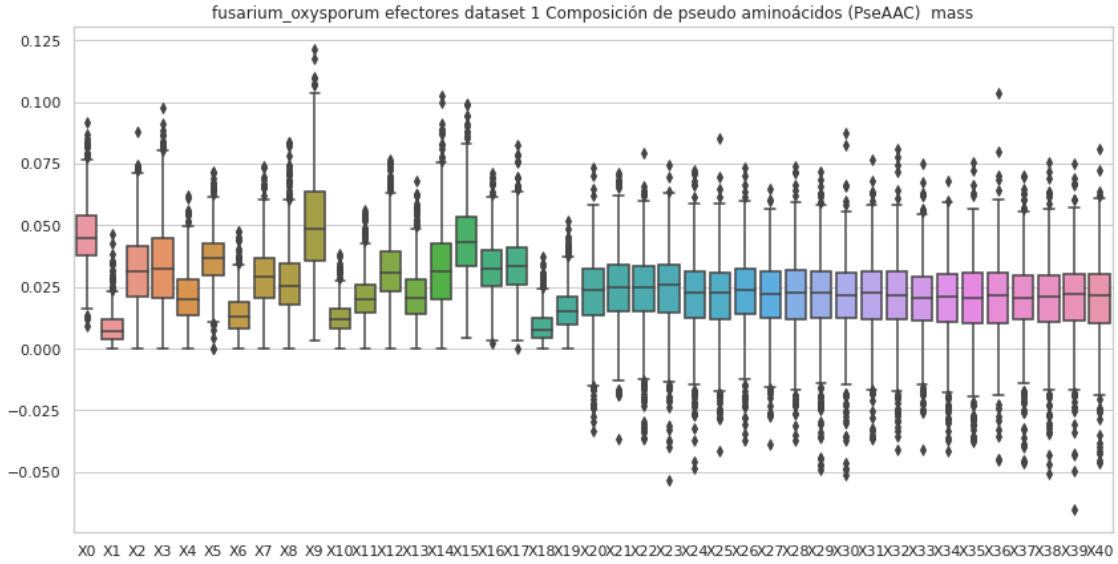
```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.048270  0.012777  0.034073  0.046850  0.039752  0.051110  0.019876
1    0.040342  0.006724  0.021292  0.026334  0.017930  0.033058  0.008965
2    0.041988  0.009128  0.033773  0.035142  0.019168  0.031491  0.012779
3    0.056687  0.012495  0.048458  0.040839  0.018591  0.038096  0.020115
4    0.051002  0.008500  0.024084  0.028334  0.016292  0.034001  0.010625
..        …         …         …         …         …         …         …
995  0.047256  0.009451  0.055525  0.044893  0.020084  0.050800  0.025991
996  0.050567  0.015170  0.035397  0.037419  0.022249  0.027306  0.012136
997  0.031073  0.005919  0.019236  0.007398  0.028114  0.038471  0.007398
998  0.032697  0.010899  0.037368  0.031140  0.021798  0.028026  0.017127
999  0.044810  0.020119  0.043895  0.041152  0.021948  0.042981  0.019204

           X7        X8        X9  …       X32       X33       X34  \
0    0.049690  0.042591  0.090861  …  0.012316  0.015010 -0.003004
1    0.020731  0.017930  0.048186  …  0.022652  0.037650  0.032439
2    0.027383  0.020537  0.051115  …  0.022069  0.022185  0.020048
3    0.032001  0.029258  0.078630  …  0.016160  0.014310  0.011372
4    0.027626  0.005667  0.054544  …  0.019859  0.012643  0.026760
..        …         …         …  …  …       …         …         …
995  0.043711  0.038986  0.060251  …  0.007430 -0.006872 -0.012104
996  0.036408  0.024272  0.065737  …  0.010548  0.021692  0.009384
997  0.020715  0.013317  0.029593  …  0.037351  0.037692  0.023374
998  0.020241  0.028026  0.049824  …  0.010641  0.020914  0.013303
999  0.036579  0.032007  0.068586  …  0.012760  0.028378  0.025470

          X35       X36       X37       X38       X39       X40           X41
0    0.030079  0.013662  0.025428 -0.005369  0.016107 -0.018069  no_efectores
```

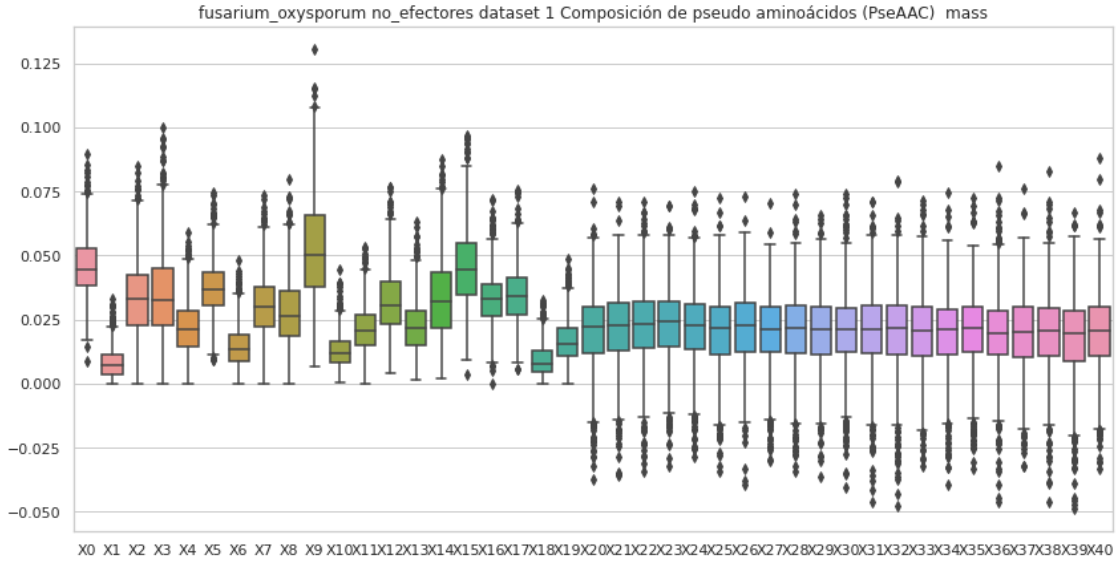
```
1    0.027158  0.023240  0.024295  0.011999  0.034569  0.024605  no_efectores
2    0.033996  0.013176  0.006771  0.018918  0.025270  0.006959  no_efectores
3    0.013533  0.027176  0.026701  0.019796  0.011521  0.014939  no_efectores
4    0.021274  0.028430  0.008132  0.026338  0.022929  0.013561  no_efectores
..        …         …         …         …         …         …
995  0.011054  0.004931  0.016019  0.019289  0.005901  0.021003  no_efectores
996  0.019689  0.030153  0.013776  0.017020  0.027664  0.022030  no_efectores
997  0.031879  0.038205  0.042779  0.032780  0.012020  0.001266  no_efectores
998 -0.010140 -0.000603  0.044190  0.006358  0.022789  0.027936  no_efectores
999  0.023151  0.030378  0.030347  0.010310  0.032568  0.029261  no_efectores

[842 rows x 42 columns]
```

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

```
                X0          X1          X2          X3          X4          X5  \
count  842.000000  842.000000  842.000000  842.000000  842.000000  842.000000
mean     0.045643    0.008186    0.033608    0.035036    0.022080    0.037080
std      0.011403    0.005910    0.014668    0.016409    0.010395    0.010403
min      0.008559    0.000000    0.000000    0.000000    0.000000    0.009257
25%      0.038024    0.003747    0.022860    0.022942    0.014384    0.030476
50%      0.044563    0.007323    0.032931    0.032592    0.021060    0.036588
75%      0.052528    0.011290    0.042243    0.044784    0.028250    0.043354
max      0.089768    0.033313    0.085089    0.100105    0.058797    0.074724

                X6          X7          X8          X9  …         X31  \
count  842.000000  842.000000  842.000000  842.000000  …  842.000000
mean     0.014544    0.030462    0.028011    0.052039  …    0.020512
std      0.008030    0.012098    0.013331    0.020907  …    0.015650
min      0.000000    0.000000    0.000000    0.006722  …   -0.045901
25%      0.008661    0.022307    0.018528    0.037754  …    0.012045
50%      0.013572    0.029787    0.026531    0.050072  …    0.021181
75%      0.019340    0.037925    0.036094    0.065661  …    0.030729
max      0.048175    0.073641    0.079549    0.130424  …    0.071067

               X32         X33         X34         X35         X36         X37  \
count  842.000000  842.000000  842.000000  842.000000  842.000000  842.000000
mean     0.020481    0.020108    0.019569    0.020448    0.019395    0.019628
std      0.015900    0.015516    0.015270    0.014793    0.015415    0.014544
min     -0.047626   -0.032125   -0.039643   -0.033318   -0.046199   -0.032218
25%      0.011564    0.010869    0.011254    0.012130    0.011145    0.010535
50%      0.021710    0.020758    0.021007    0.021725    0.019726    0.020370
75%      0.030708    0.030128    0.029190    0.029949    0.028371    0.029809
max      0.078982    0.071363    0.074507    0.072632    0.084687    0.076272
```

```
              X38          X39          X40
count  842.000000   842.000000   842.000000
mean     0.019741     0.018284     0.020018
std      0.015588     0.015833     0.015650
min     -0.046325    -0.048727    -0.033405
25%      0.011007     0.009000     0.010865
50%      0.020563     0.019731     0.020554
75%      0.029282     0.028684     0.030059
max      0.082753     0.066999     0.087977

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.075852  0.016856  0.050568  0.049164  0.033712  0.044950  0.025284
1    0.054665  0.014503  0.030121  0.031237  0.025659  0.027890  0.013387
2    0.078023  0.011703  0.040962  0.037061  0.046814  0.050715  0.023407
3    0.052481  0.010722  0.050224  0.056432  0.037809  0.045710  0.024266
4    0.046836  0.004461  0.040145  0.046836  0.011151  0.040145  0.013382
..        …         …         …         …         …         …         …
995  0.055676  0.011135  0.096506  0.063100  0.040829  0.070523  0.040829
996  0.164062  0.017270  0.043174  0.043174  0.034539  0.069079  0.051809
997  0.034021  0.004002  0.014009  0.013008  0.018511  0.028017  0.006504
998  0.089950  0.010280  0.061680  0.041120  0.046260  0.084810  0.023130
999  0.023249  0.009300  0.041849  0.046499  0.018600  0.041849  0.027899

           X7        X8        X9  …       X53       X54       X55  \
0    0.036521  0.033712  0.058996  … -0.019293  0.025812  0.022054
1    0.034584  0.015618  0.041277  …  0.009412  0.013915 -0.001121
2    0.037061  0.031209  0.072171  … -0.021812 -0.016243 -0.021886
3    0.033295  0.028216  0.051353  …  0.021283 -0.007372  0.009610
4    0.011151  0.013382  0.066909  …  0.032419 -0.025073 -0.006417
..        …         …         …  …         …         …         …
995  0.059388  0.051965  0.122488  … -0.040610  0.017069  0.016827
996  0.060444  0.077713  0.120888  …  0.011522  0.066620  0.056040
997  0.030018  0.010506  0.032520  …  0.009935  0.007182  0.008915
998  0.043690  0.053970  0.095090  … -0.025113  0.007944 -0.018633
999  0.041849  0.023249  0.060448  … -0.013462 -0.050072 -0.000442

          X56       X57       X58       X59       X60       X61         X62
0    0.005591  0.002447  0.009412  0.010567 -0.000463  0.015320  efectores
1    0.009782  0.011826  0.005671 -0.003883  0.009708  0.004813  efectores
2    0.017239 -0.005217  0.002925  0.007318 -0.030477 -0.009589  efectores
3    0.006522  0.021256 -0.019023  0.010259 -0.003502  0.005410  efectores
4    0.008314  0.018162 -0.003770 -0.006936 -0.001150  0.005299  efectores
..        …         …         …         …         …         …         …
995  0.122943  0.106664  0.008933 -0.014063  0.002776  0.002921  efectores
996 -0.047934 -0.010362 -0.067247 -0.061112 -0.023443 -0.020321  efectores
997  0.008284  0.006303  0.004183  0.007210  0.011983  0.005954  efectores
```

```
998 -0.008381 -0.001165  0.039403  0.017501  0.044875  0.032722  efectores
999 -0.050502 -0.011460  0.044397 -0.000400 -0.031996  0.003793  efectores

[1000 rows x 63 columns]
```

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

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.061030	0.011966	0.041716	0.043043	0.027983	
std	0.028286	0.014116	0.021142	0.022373	0.015770	
min	0.002187	0.000000	0.000000	0.000000	0.000000	
25%	0.040044	0.004343	0.027089	0.028136	0.016826	
50%	0.057547	0.008883	0.040951	0.041267	0.026560	
75%	0.076571	0.015113	0.054184	0.055032	0.036897	
max	0.204275	0.170584	0.170289	0.239372	0.116720	

	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.049741	0.018952	0.038260	0.035927	0.064630	…	
std	0.028194	0.014843	0.020308	0.020765	0.032230	…	
min	0.000000	0.000000	0.000000	0.000000	0.000000	…	
25%	0.029974	0.009313	0.023805	0.022327	0.041241	…	
50%	0.043803	0.016234	0.036394	0.032318	0.060757	…	
75%	0.064912	0.024838	0.050364	0.047259	0.083389	…	
max	0.251546	0.188050	0.188660	0.255434	0.255434	…	

	X52	X53	X54	X55	X56	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.002521	0.007773	0.003217	0.008082	0.002500	
std	0.032826	0.027507	0.030679	0.026526	0.029814	
min	-0.377326	-0.167604	-0.154839	-0.194583	-0.169332	
25%	-0.010678	-0.005394	-0.009845	-0.003787	-0.009599	
50%	0.005517	0.008839	0.005337	0.010207	0.004424	
75%	0.017256	0.022104	0.017550	0.022291	0.016559	
max	0.144436	0.270612	0.184003	0.149201	0.122943	

	X57	X58	X59	X60	X61
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.007771	0.003671	0.008911	0.001978	0.008833
std	0.025243	0.035948	0.027872	0.031184	0.027933
min	-0.144387	-0.488991	-0.220425	-0.220693	-0.244727
25%	-0.003438	-0.009159	-0.002660	-0.009681	-0.003331
50%	0.009416	0.005998	0.009807	0.004459	0.010800
75%	0.021703	0.019096	0.022986	0.017319	0.022694

```
max       0.106664     0.407830     0.199001     0.230054     0.162985

[8 rows x 62 columns]
```

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.065241	0.017270	0.046052	0.063322	0.053728	0.069078	0.026864	
1	0.083758	0.013960	0.044205	0.054675	0.037226	0.068635	0.018613	
2	0.065080	0.014148	0.052347	0.054469	0.029710	0.048810	0.019807	
3	0.095581	0.021069	0.081707	0.068860	0.031347	0.064235	0.033916	
4	0.072069	0.012011	0.034033	0.040038	0.023022	0.048046	0.015014	
..	…	…	…	…	…	…	…	
995	0.054021	0.010804	0.063474	0.051320	0.022959	0.058072	0.029711	
996	0.060461	0.018138	0.042322	0.044741	0.026603	0.032649	0.014511	
997	0.037114	0.007069	0.022975	0.008837	0.033579	0.045950	0.008837	
998	0.035809	0.011936	0.040925	0.034104	0.023873	0.030693	0.018757	
999	0.059661	0.026786	0.058443	0.054790	0.029221	0.057225	0.025569	

	X7	X8	X9	…	X53	X54	X55	\
0	0.067159	0.057565	0.122806	…	0.005470	-0.029379	-0.009484	
1	0.043042	0.037226	0.100044	…	0.002034	0.032746	0.010783	
2	0.042443	0.031832	0.079227	…	0.019747	-0.009049	0.023406	
3	0.053957	0.049332	0.132580	…	0.004405	0.004523	0.017629	
4	0.039037	0.008008	0.077074	…	0.013968	-0.015066	-0.007561	
..	…	…	…	…	…	…	…	
995	0.049969	0.044567	0.068876	…	-0.018678	0.008178	0.004712	
996	0.043532	0.029021	0.078599	…	0.005306	0.007960	-0.006870	
997	0.024742	0.015906	0.035346	…	-0.009276	0.030911	0.002500	
998	0.022167	0.030693	0.054566	…	0.012521	0.004630	0.006987	
999	0.048702	0.042615	0.091317	…	0.021317	0.005377	0.023039	

	X56	X57	X58	X59	X60	X61	X62
0	0.001368	-0.014184	0.017689	0.006425	0.005829	0.015974	no_efectores
1	0.004897	-0.000444	0.015671	0.007155	-0.009728	0.005015	no_efectores
2	-0.006270	0.008101	-0.018067	-0.005541	-0.011964	0.010638	no_efectores
3	-0.021850	-0.006471	-0.022448	-0.015159	0.020490	0.024967	no_efectores
4	0.002440	0.013446	0.015388	0.011788	-0.025581	-0.021010	no_efectores
..	…	…	…	…	…	…	…
995	0.032652	0.043755	0.007773	0.020652	-0.026672	-0.001979	no_efectores
996	-0.000677	0.015210	-0.003808	0.020662	0.046263	0.015786	no_efectores
997	0.034273	0.007346	0.010705	0.003139	-0.002158	-0.004434	no_efectores
998	0.016500	0.037265	0.057595	0.053319	-0.012421	0.014224	no_efectores

```
999 -0.012323 -0.017142 -0.013618  0.007006  0.000712  0.032477  no_efectores

[1000 rows x 63 columns]
```

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

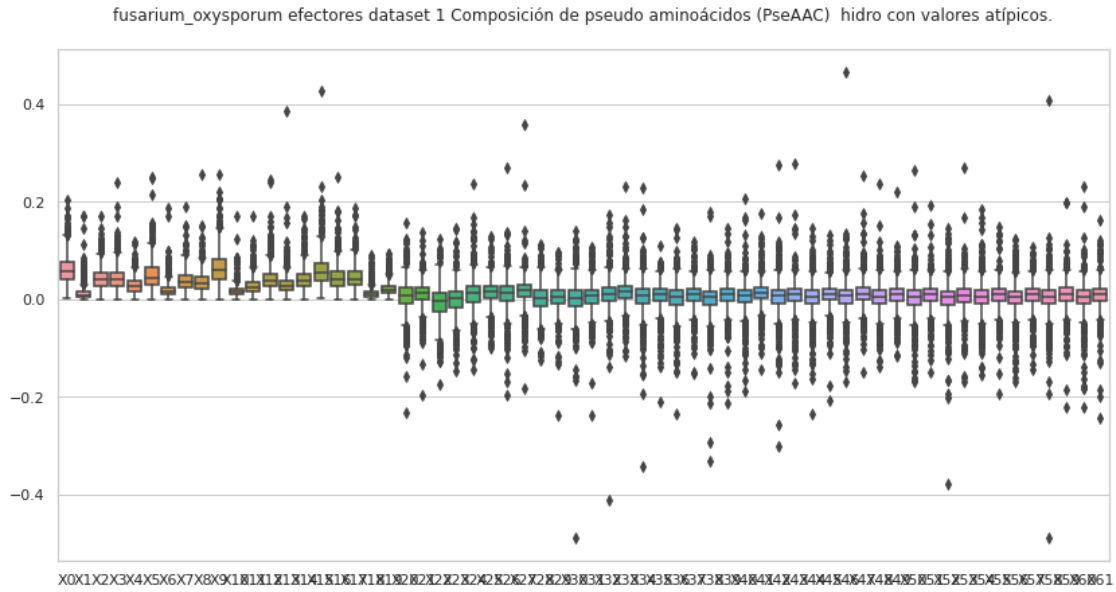
	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.060554	0.011600	0.042552	0.044565	0.028882	
std	0.031340	0.010971	0.021245	0.021892	0.017789	
min	0.002112	0.000000	0.000000	0.000000	0.000000	
25%	0.040717	0.004470	0.027225	0.030346	0.017550	
50%	0.056155	0.008970	0.041777	0.043197	0.027078	
75%	0.074359	0.015601	0.055696	0.056922	0.036873	
max	0.409518	0.102380	0.152638	0.205209	0.204759	

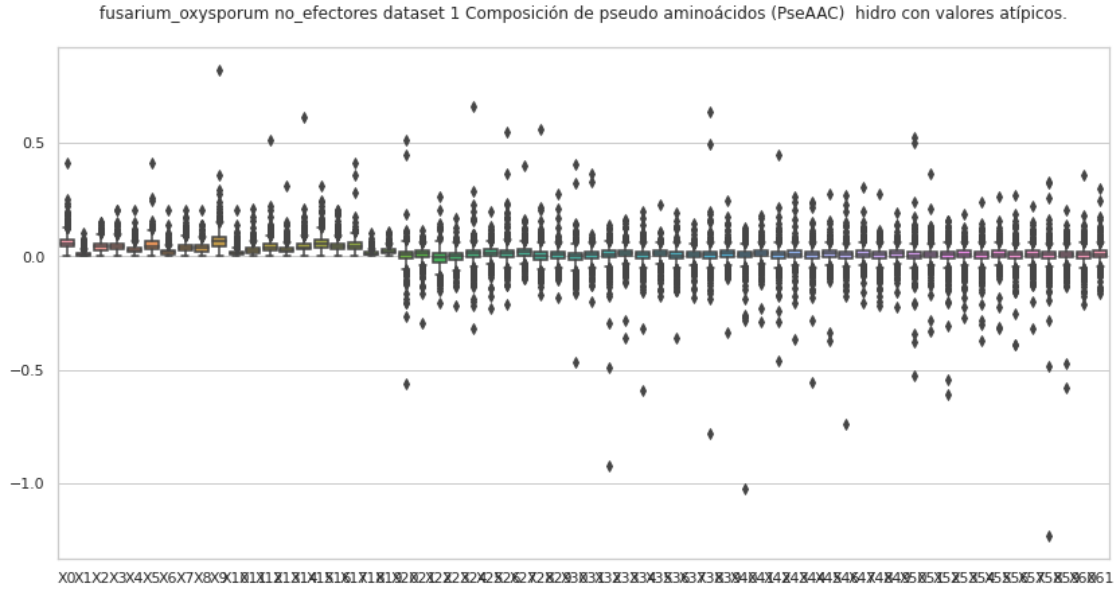
	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.049925	0.019501	0.040058	0.036329	0.068589	…	
std	0.028920	0.014702	0.021984	0.020118	0.042755	…	
min	0.000000	0.000000	0.000000	0.000000	0.001730	…	
25%	0.030681	0.010126	0.026070	0.022303	0.043715	…	
50%	0.045058	0.017405	0.037893	0.034362	0.064881	…	
75%	0.065753	0.025755	0.051933	0.047466	0.086812	…	
max	0.409518	0.204759	0.204759	0.203517	0.819036	…	

	X52	X53	X54	X55	X56	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.001834	0.008265	0.002823	0.008403	0.002734	
std	0.042616	0.030281	0.035282	0.031303	0.036830	
min	-0.610131	-0.270125	-0.371154	-0.315504	-0.388333	
25%	-0.008989	-0.003466	-0.009672	-0.003318	-0.009182	
50%	0.005936	0.010396	0.005709	0.009568	0.004778	
75%	0.018121	0.022682	0.018012	0.023047	0.017324	
max	0.157360	0.182751	0.236838	0.259922	0.269625	

	X57	X58	X59	X60	X61
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.009223	0.002376	0.007923	0.003866	0.009618
std	0.030331	0.055701	0.036897	0.034665	0.031454
min	-0.316955	-1.230758	-0.578063	-0.212262	-0.166845
25%	-0.001677	-0.008784	-0.001761	-0.010310	-0.004178
50%	0.010140	0.005348	0.009305	0.005351	0.008858
75%	0.023029	0.017694	0.022109	0.018264	0.022562
max	0.218531	0.327531	0.203912	0.355035	0.295738

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.075852	0.016856	0.050568	0.049164	0.033712	0.044950	0.025284	
1	0.054665	0.014503	0.030121	0.031237	0.025659	0.027890	0.013387	
3	0.052481	0.010722	0.050224	0.056432	0.037809	0.045710	0.024266	
4	0.046836	0.004461	0.040145	0.046836	0.011151	0.040145	0.013382	
6	0.065016	0.007314	0.019505	0.021943	0.029257	0.052826	0.008127	
..	…	…	…	…	…	…	…	
991	0.074003	0.007162	0.042969	0.057292	0.014323	0.045357	0.019097	
992	0.093144	0.012419	0.058991	0.062096	0.018629	0.055886	0.037258	
994	0.027736	0.007925	0.047547	0.011887	0.000000	0.035660	0.015849	
997	0.034021	0.004002	0.014009	0.013008	0.018511	0.028017	0.006504	
999	0.023249	0.009300	0.041849	0.046499	0.018600	0.041849	0.027899	

	X7	X8	X9	…	X53	X54	X55	\
0	0.036521	0.033712	0.058996	…	-0.019293	0.025812	0.022054	
1	0.034584	0.015618	0.041277	…	0.009412	0.013915	-0.001121	
3	0.033295	0.028216	0.051353	…	0.021283	-0.007372	0.009610	
4	0.011151	0.013382	0.066909	…	0.032419	-0.025073	-0.006417	
6	0.029257	0.027632	0.073143	…	-0.001851	0.008701	0.002494	
..	…	…	…	…	…	…	…	
991	0.062067	0.059680	0.078777	…	-0.003899	0.028519	0.015864	
992	0.065201	0.052782	0.099354	…	0.014432	-0.033662	-0.008605	
994	0.007925	0.007925	0.091132	…	0.029382	0.014349	0.001230	
997	0.030018	0.010506	0.032520	…	0.009935	0.007182	0.008915	
999	0.041849	0.023249	0.060448	…	-0.013462	-0.050072	-0.000442	

	X56	X57	X58	X59	X60	X61	X62
0	0.005591	0.002447	0.009412	0.010567	-0.000463	0.015320	efectores
1	0.009782	0.011826	0.005671	-0.003883	0.009708	0.004813	efectores
3	0.006522	0.021256	-0.019023	0.010259	-0.003502	0.005410	efectores
4	0.008314	0.018162	-0.003770	-0.006936	-0.001150	0.005299	efectores
6	0.008304	-0.003163	-0.005287	-0.002314	0.013013	0.004397	efectores
..	…	…	…	…	…	…	…
991	0.012579	0.032194	-0.044608	0.007519	-0.009091	-0.009226	efectores
992	-0.013909	-0.016991	-0.030249	-0.035827	0.026672	0.035568	efectores
994	0.006848	-0.004927	0.048803	0.050856	0.019553	0.033154	efectores
997	0.008284	0.006303	0.004183	0.007210	0.011983	0.005954	efectores
999	-0.050502	-0.011460	0.044397	-0.000400	-0.031996	0.003793	efectores

[833 rows x 63 columns]

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

	X0	X1	X2	X3	X4	X5	\
count	833.000000	833.000000	833.000000	833.000000	833.000000	833.000000	
mean	0.056075	0.010000	0.038741	0.039907	0.025491	0.045159	
std	0.023614	0.008064	0.017757	0.017575	0.013001	0.022188	
min	0.002187	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.037694	0.004298	0.024974	0.027056	0.016036	0.028854	
50%	0.053219	0.008355	0.038829	0.039789	0.024682	0.040774	
75%	0.070756	0.013504	0.051001	0.052553	0.033975	0.060022	
max	0.136898	0.047281	0.098197	0.095351	0.069343	0.121735	

	X6	X7	X8	X9	…	X52	\
count	833.000000	833.000000	833.000000	833.000000	…	833.000000	
mean	0.017030	0.035410	0.033325	0.059937	…	0.004486	
std	0.009975	0.016896	0.016241	0.026879	…	0.021295	
min	0.000000	0.000000	0.000000	0.000000	…	-0.072121	
25%	0.009056	0.022932	0.021388	0.039499	…	-0.007058	
50%	0.015444	0.034249	0.031480	0.058286	…	0.006036	
75%	0.023397	0.046583	0.043964	0.078257	…	0.016341	
max	0.061659	0.089456	0.086511	0.149692	…	0.076958	

	X53	X54	X55	X56	X57	X58	\
count	833.000000	833.000000	833.000000	833.000000	833.000000	833.000000	
mean	0.009390	0.004761	0.009598	0.004872	0.009940	0.005171	
std	0.019240	0.021105	0.019197	0.020864	0.018322	0.022867	
min	-0.072397	-0.077937	-0.064951	-0.081712	-0.065391	-0.099860	
25%	-0.002569	-0.006308	-0.001166	-0.006208	-0.000723	-0.007579	
50%	0.009632	0.005879	0.010956	0.005461	0.010746	0.006047	
75%	0.021759	0.017238	0.022216	0.016063	0.021411	0.018368	
max	0.073833	0.086520	0.065567	0.083003	0.075891	0.103677	

	X59	X60	X61
count	833.000000	833.000000	833.000000
mean	0.009934	0.004004	0.009962
std	0.019243	0.021004	0.019202
min	-0.062545	-0.079547	-0.068487
25%	-0.000875	-0.007359	-0.001179
50%	0.010259	0.005196	0.011091
75%	0.021860	0.017099	0.022412
max	0.080515	0.070894	0.088940

[8 rows x 62 columns]

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

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.065241  0.017270  0.046052  0.063322  0.053728  0.069078  0.026864
1    0.083758  0.013960  0.044205  0.054675  0.037226  0.068635  0.018613
2    0.065080  0.014148  0.052347  0.054469  0.029710  0.048810  0.019807
3    0.095581  0.021069  0.081707  0.068860  0.031347  0.064235  0.033916
4    0.072069  0.012011  0.034033  0.040038  0.023022  0.048046  0.015014
..        …         …         …         …         …         …         …
995  0.054021  0.010804  0.063474  0.051320  0.022959  0.058072  0.029711
996  0.060461  0.018138  0.042322  0.044741  0.026603  0.032649  0.014511
997  0.037114  0.007069  0.022975  0.008837  0.033579  0.045950  0.008837
998  0.035809  0.011936  0.040925  0.034104  0.023873  0.030693  0.018757
999  0.059661  0.026786  0.058443  0.054790  0.029221  0.057225  0.025569

           X7        X8        X9  …       X53       X54       X55  \
0    0.067159  0.057565  0.122806  …  0.005470 -0.029379 -0.009484
1    0.043042  0.037226  0.100044  …  0.002034  0.032746  0.010783
2    0.042443  0.031832  0.079227  …  0.019747 -0.009049  0.023406
3    0.053957  0.049332  0.132580  …  0.004405  0.004523  0.017629
4    0.039037  0.008008  0.077074  …  0.013968 -0.015066 -0.007561
..        …         …         …  …       …         …         …
995  0.049969  0.044567  0.068876  … -0.018678  0.008178  0.004712
996  0.043532  0.029021  0.078599  …  0.005306  0.007960 -0.006870
997  0.024742  0.015906  0.035346  … -0.009276  0.030911  0.002500
998  0.022167  0.030693  0.054566  …  0.012521  0.004630  0.006987
999  0.048702  0.042615  0.091317  …  0.021317  0.005377  0.023039

          X56       X57       X58       X59       X60       X61           X62
0    0.001368 -0.014184  0.017689  0.006425  0.005829  0.015974  no_efectores
1    0.004897 -0.000444  0.015671  0.007155 -0.009728  0.005015  no_efectores
2   -0.006270  0.008101 -0.018067 -0.005541 -0.011964  0.010638  no_efectores
3   -0.021850 -0.006471 -0.022448 -0.015159  0.020490  0.024967  no_efectores
4    0.002440  0.013446  0.015388  0.011788 -0.025581 -0.021010  no_efectores
..        …         …         …         …         …         …           …
995  0.032652  0.043755  0.007773  0.020652 -0.026672 -0.001979  no_efectores
996 -0.000677  0.015210 -0.003808  0.020662  0.046263  0.015786  no_efectores
997  0.034273  0.007346  0.010705  0.003139 -0.002158 -0.004434  no_efectores
998  0.016500  0.037265  0.057595  0.053319 -0.012421  0.014224  no_efectores
999 -0.012323 -0.017142 -0.013618  0.007006  0.000712  0.032477  no_efectores

[860 rows x 63 columns]
```

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

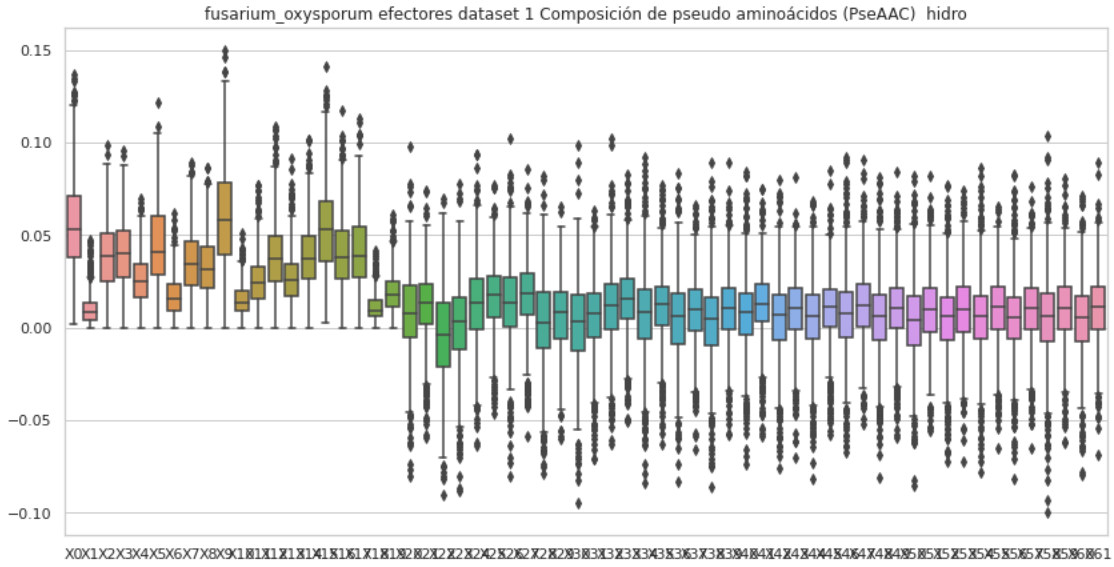
	X0	X1	X2	X3	X4	X5	\
count	860.000000	860.000000	860.000000	860.000000	860.000000	860.000000	
mean	0.055110	0.009896	0.040595	0.042091	0.026561	0.046143	
std	0.022888	0.007480	0.018453	0.018059	0.013271	0.022775	
min	0.002112	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.038968	0.004439	0.026810	0.028886	0.017342	0.029429	
50%	0.053467	0.008379	0.041305	0.041779	0.025848	0.043354	
75%	0.068371	0.014021	0.053633	0.054296	0.035024	0.061016	
max	0.140449	0.043699	0.101654	0.108069	0.077696	0.135486	

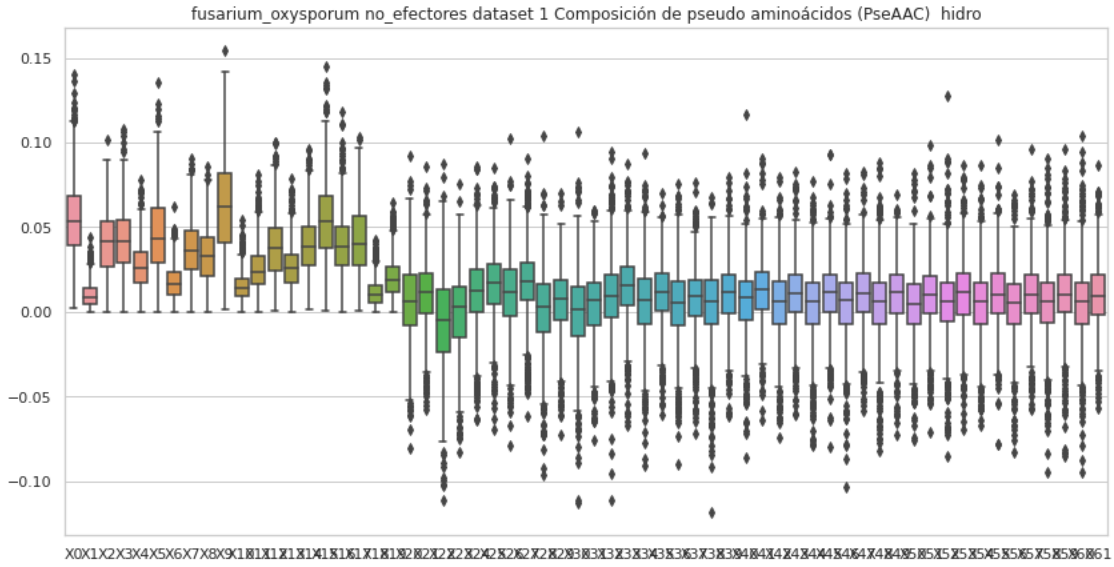
	X6	X7	X8	X9	…	X52	\
count	860.000000	860.000000	860.000000	860.000000	…	860.000000	
mean	0.017526	0.036652	0.033771	0.062549	…	0.005278	
std	0.009919	0.017075	0.016077	0.028341	…	0.022293	
min	0.000000	0.000000	0.000000	0.001730	…	-0.084915	
25%	0.009814	0.025024	0.021526	0.041174	…	-0.005542	
50%	0.016430	0.036033	0.033040	0.062491	…	0.006416	
75%	0.023613	0.047836	0.044048	0.082169	…	0.017594	
max	0.062205	0.090915	0.085722	0.154288	…	0.127822	

	X53	X54	X55	X56	X57	X58	\
count	860.000000	860.000000	860.000000	860.000000	860.000000	860.000000	
mean	0.010645	0.005061	0.010417	0.004078	0.010462	0.005436	
std	0.020501	0.020980	0.019796	0.020554	0.019381	0.022377	
min	-0.069244	-0.067129	-0.078432	-0.082901	-0.057937	-0.094361	
25%	-0.001428	-0.007066	-0.001073	-0.007150	-0.000554	-0.006021	
50%	0.011859	0.006540	0.010317	0.005088	0.010270	0.006295	
75%	0.022556	0.017603	0.022928	0.016504	0.021836	0.017115	
max	0.086618	0.086839	0.101852	0.069114	0.095953	0.090225	

	X59	X60	X61
count	860.000000	860.000000	860.000000
mean	0.010129	0.005505	0.010358
std	0.021259	0.022628	0.020155
min	-0.085236	-0.094740	-0.056424
25%	-0.000101	-0.007222	-0.001734
50%	0.010109	0.006100	0.009406
75%	0.021709	0.017346	0.022142
max	0.096241	0.104233	0.086582

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.059361  0.058924  0.043385  0.019266 -0.050025 -0.012213 -0.049867
1   -0.038358 -0.022014  0.084707  0.028736 -0.011468  0.028057  0.035392
2    0.084211  0.036179  0.016931  0.032327 -0.004568  0.087915  0.025747
3    0.012996  0.062573  0.044595 -0.022979 -0.019777 -0.000854 -0.058178
4    0.028805  0.067778 -0.075469  0.054776 -0.039264  0.018130 -0.043181
..        …         …         …         …         …         …         …
995 -0.021303  0.004344 -0.047501 -0.030772 -0.021117  0.034647 -0.019821
996  0.020442 -0.036786  0.004765 -0.182379 -0.007104  0.016541  0.223794
997  0.034182  0.050604 -0.007599  0.058100 -0.004085 -0.027304  0.000403
998 -0.043523 -0.045795  0.017393 -0.098307  0.048070 -0.054248  0.017152
999  0.083563  0.017567 -0.024205 -0.123925  0.089546  0.159116 -0.015107

           X7        X8        X9       X10       X11       X12        X13
0   -0.057264  0.007241  0.079565  0.013705  0.058798 -0.000752  efectores
```

```
1   -0.000954 -0.035263  0.034839 -0.019392  0.021979  0.038419  efectores
2    0.032613  0.074832  0.042960 -0.012147  0.046505 -0.035839  efectores
3   -0.007489 -0.011409  0.003605  0.034205  0.039099  0.072545  efectores
4   -0.020093  0.040531 -0.011302 -0.029099 -0.025713  0.027003  efectores
..        …         …         …         …         …         …         …
995  0.080332  0.106620 -0.095693 -0.011724  0.011147  0.037502  efectores
996  0.095954  0.012273 -0.018609  0.006520 -0.006303 -0.055766  efectores
997  0.004985  0.003930 -0.047405 -0.097623  0.034636 -0.019929  efectores
998 -0.036561 -0.057524  0.106336 -0.024249 -0.009742 -0.036122  efectores
999  0.112711 -0.004530 -0.045630 -0.035828 -0.145422  0.023255  efectores

[1000 rows x 14 columns]
```

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

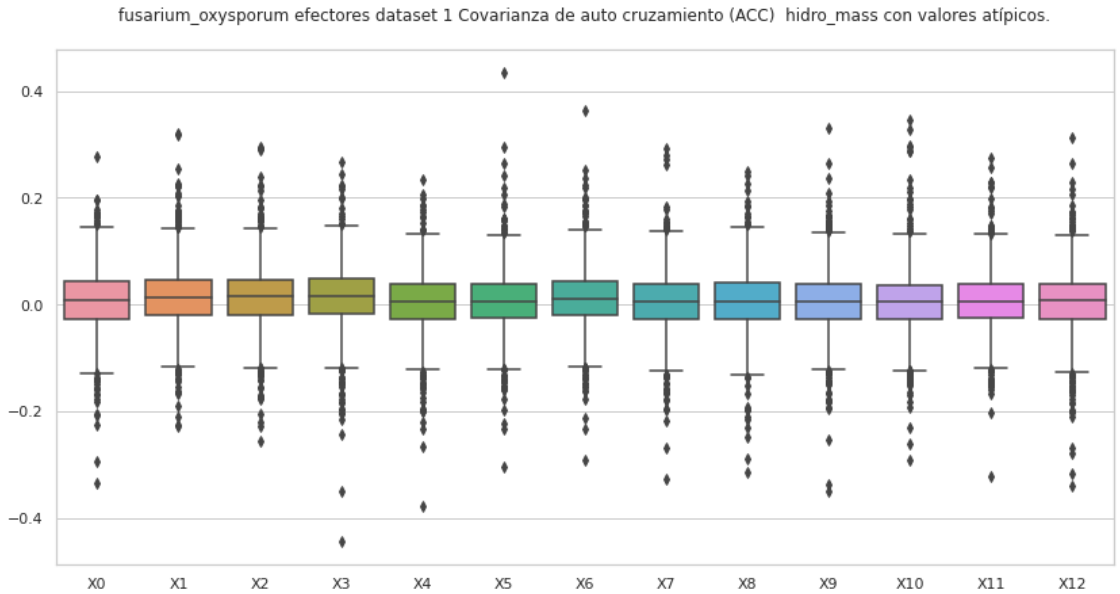
	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.007914	0.015334	0.013748	0.013974	0.003726	
std	0.059289	0.060870	0.060545	0.062419	0.060000	
min	-0.333863	-0.227512	-0.255981	-0.444166	-0.378492	
25%	-0.026228	-0.019780	-0.019805	-0.017625	-0.027220	
50%	0.009159	0.013294	0.015488	0.016071	0.005539	
75%	0.042773	0.046534	0.046278	0.049796	0.037462	
max	0.277788	0.320850	0.296214	0.268561	0.234797	

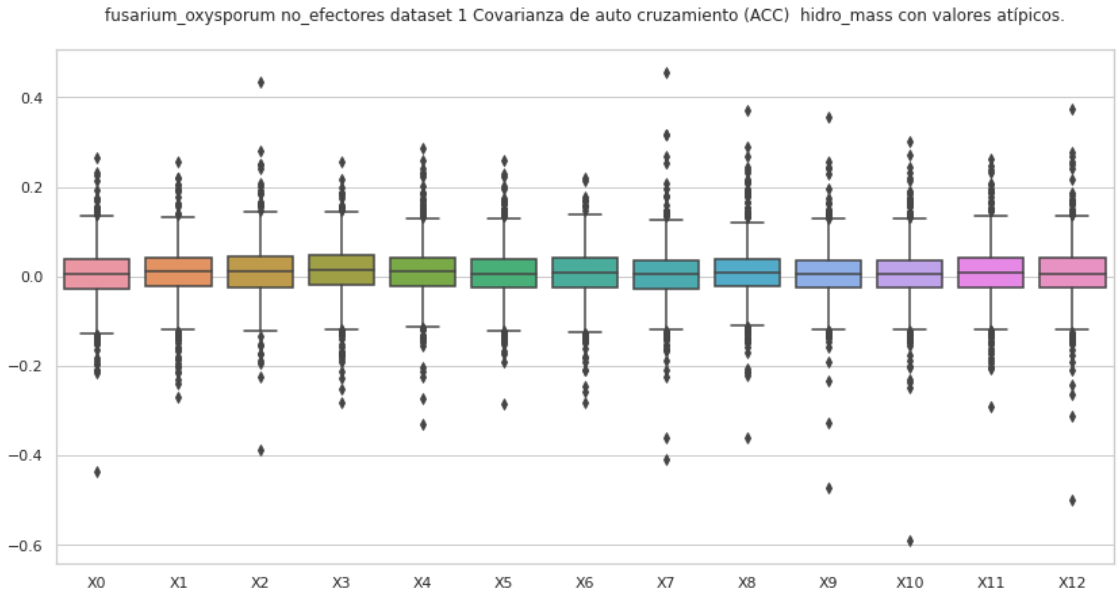
	X5	X6	X7	X8	X9	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.005929	0.012089	0.004652	0.006341	0.005663	
std	0.060220	0.058654	0.057759	0.059293	0.059467	
min	-0.303600	-0.291733	-0.327682	-0.314630	-0.350388	
25%	-0.025608	-0.020385	-0.027651	-0.028212	-0.026803	
50%	0.005079	0.011898	0.004618	0.006235	0.006710	
75%	0.037652	0.043938	0.038454	0.041485	0.038299	
max	0.434801	0.363294	0.291784	0.249002	0.330457	

	X10	X11	X12
count	1000.000000	1000.000000	1000.000000
mean	0.006423	0.006220	0.003878
std	0.061849	0.057005	0.062113
min	-0.291373	-0.321068	-0.341191
25%	-0.027370	-0.025269	-0.027726
50%	0.004683	0.006542	0.007091
75%	0.036773	0.037836	0.037599
max	0.346803	0.274804	0.314213

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

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.015650 -0.011331  0.028389  0.032013 -0.022398 -0.017179  0.025394
1    0.041161  0.030963 -0.000365 -0.018599  0.029021 -0.003269 -0.039617
2    0.010331  0.040050  0.005810  0.001890 -0.051507  0.032012  0.008658
3    0.022248 -0.030544  0.002908 -0.015868  0.014437 -0.027792  0.010491
4   -0.000055  0.029249  0.028753  0.032387  0.027074  0.003731  0.041759
..        …         …         …         …         …         …
995  0.052177 -0.006901 -0.005753  0.029276  0.000347  0.010830 -0.020796
996 -0.046018  0.013061  0.037168 -0.009167 -0.024973  0.060783  0.090649
997  0.000815 -0.022999  0.010311  0.076356  0.015550  0.082516  0.063399
998  0.034308  0.051967  0.081666  0.091435  0.023750  0.056765  0.038937
999  0.013296 -0.006418 -0.001484 -0.006932  0.029627 -0.018526  0.006531

           X7        X8        X9       X10       X11       X12           X13
0   -0.019560 -0.045433  0.019501  0.006152 -0.033507 -0.001731  no_efectores
1   -0.010741  0.024122  0.012430  0.015638  0.024028 -0.009259  no_efectores
2   -0.000690  0.047390  0.001453 -0.001497  0.019164  0.003748  no_efectores
3   -0.005095  0.054626  0.004013 -0.018259 -0.012590 -0.003643  no_efectores
4   -0.001602 -0.084648  0.014938 -0.022521  0.053480 -0.006299  no_efectores
..        …         …         …         …         …         …
995  0.009640  0.078996 -0.057196  0.029834 -0.010140 -0.017023  no_efectores
996  0.011661  0.039312 -0.019038  0.008169 -0.008315 -0.011261  no_efectores
997 -0.052386 -0.018233 -0.036478  0.093462  0.151383  0.067000  no_efectores
998  0.058807  0.027715  0.010362  0.031771  0.007733 -0.008922  no_efectores
999 -0.034849 -0.008139 -0.092508 -0.007543 -0.055925 -0.015299  no_efectores

[1000 rows x 14 columns]


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

                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.003693     0.007011     0.010673     0.012145     0.010608
std       0.059905     0.058273     0.059600     0.057878     0.057683
min      -0.434742    -0.270548    -0.387666    -0.281609    -0.329958
25%      -0.029524    -0.023009    -0.024225    -0.019131    -0.021426
50%       0.005548     0.009745     0.010408     0.015202     0.010058
75%       0.036910     0.040853     0.043179     0.046941     0.039947
```

```
max       0.266218     0.254689     0.433576     0.257133     0.284839

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.006093     0.006912     0.004509     0.007801     0.006284
std       0.056198     0.058348     0.059442     0.059892     0.057307
min      -0.284003    -0.282700    -0.409931    -0.361897    -0.471840
25%      -0.026162    -0.026109    -0.026675    -0.021942    -0.026013
50%       0.005675     0.008179     0.005655     0.007067     0.005440
75%       0.036955     0.040659     0.034822     0.037854     0.036375
max       0.260137     0.220528     0.455226     0.371280     0.355382

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.005545     0.007547     0.007818
std       0.062004     0.059849     0.062538
min      -0.590356    -0.291501    -0.500593
25%      -0.026603    -0.024145    -0.024871
50%       0.005320     0.006666     0.006398
75%       0.036109     0.040272     0.040003
max       0.302310     0.261207     0.373581
```





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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.059361  0.058924  0.043385  0.019266 -0.050025 -0.012213 -0.049867
1   -0.038358 -0.022014  0.084707  0.028736 -0.011468  0.028057  0.035392
2    0.084211  0.036179  0.016931  0.032327 -0.004568  0.087915  0.025747
3    0.012996  0.062573  0.044595 -0.022979 -0.019777 -0.000854 -0.058178
4    0.028805  0.067778 -0.075469  0.054776 -0.039264  0.018130 -0.043181
..        …         …         …         …         …         …         …
994  0.065591 -0.003433  0.036059  0.047216 -0.088258  0.090745  0.102642
995 -0.021303  0.004344 -0.047501 -0.030772 -0.021117  0.034647 -0.019821
997  0.034182  0.050604 -0.007599  0.058100 -0.004085 -0.027304  0.000403
998 -0.043523 -0.045795  0.017393 -0.098307  0.048070 -0.054248  0.017152
999  0.083563  0.017567 -0.024205 -0.123925  0.089546  0.159116 -0.015107

           X7        X8        X9       X10       X11       X12        X13
0   -0.057264  0.007241  0.079565  0.013705  0.058798 -0.000752  efectores
1   -0.000954 -0.035263  0.034839 -0.019392  0.021979  0.038419  efectores
2    0.032613  0.074832  0.042960 -0.012147  0.046505 -0.035839  efectores
3   -0.007489 -0.011409  0.003605  0.034205  0.039099  0.072545  efectores
4   -0.020093  0.040531 -0.011302 -0.029099 -0.025713  0.027003  efectores
..        …         …         …         …         …         …         …
994  0.009412  0.012955 -0.084085 -0.034807 -0.016416  0.018030  efectores
```

```
995  0.080332  0.106620 -0.095693 -0.011724  0.011147  0.037502  efectores
997  0.004985  0.003930 -0.047405 -0.097623  0.034636 -0.019929  efectores
998 -0.036561 -0.057524  0.106336 -0.024249 -0.009742 -0.036122  efectores
999  0.112711 -0.004530 -0.045630 -0.035828 -0.145422  0.023255  efectores

[907 rows x 14 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	907.000000	907.000000	907.000000	907.000000	907.000000	907.000000	
mean	0.008199	0.015381	0.012643	0.015676	0.005422	0.005276	
std	0.051217	0.051843	0.051333	0.050192	0.050216	0.050256	
min	-0.168255	-0.167004	-0.155818	-0.148647	-0.157364	-0.159943	
25%	-0.024131	-0.016029	-0.018935	-0.015706	-0.024804	-0.024283	
50%	0.009054	0.013919	0.015397	0.016054	0.005682	0.004853	
75%	0.040908	0.045568	0.045082	0.047749	0.036236	0.035708	
max	0.172623	0.186005	0.195200	0.171822	0.177678	0.160696	

	X6	X7	X8	X9	X10	X11	\
count	907.000000	907.000000	907.000000	907.000000	907.000000	907.000000	
mean	0.010138	0.005935	0.006512	0.004624	0.004376	0.005506	
std	0.049667	0.048062	0.050058	0.048604	0.049494	0.048370	
min	-0.161446	-0.163397	-0.151329	-0.164149	-0.150826	-0.159206	
25%	-0.020019	-0.024372	-0.027215	-0.026432	-0.027260	-0.023968	
50%	0.010926	0.005940	0.005768	0.005695	0.004295	0.006766	
75%	0.040845	0.037494	0.040144	0.036367	0.034272	0.036065	
max	0.185500	0.154017	0.165896	0.161853	0.187291	0.146926	

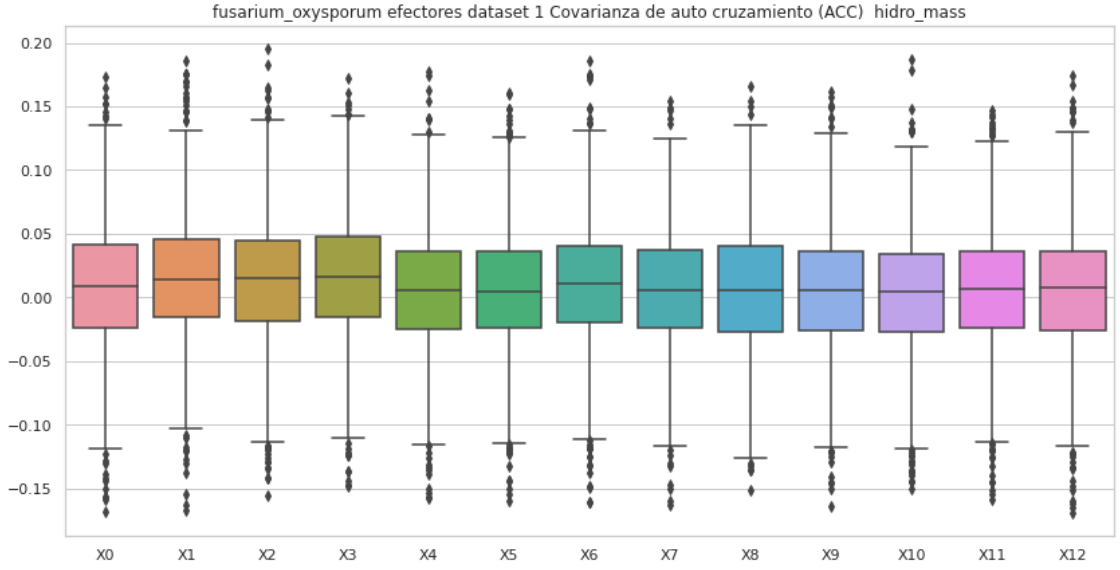
	X12
count	907.000000
mean	0.004883
std	0.050951
min	-0.169414
25%	-0.025849
50%	0.007257
75%	0.036636
max	0.173802

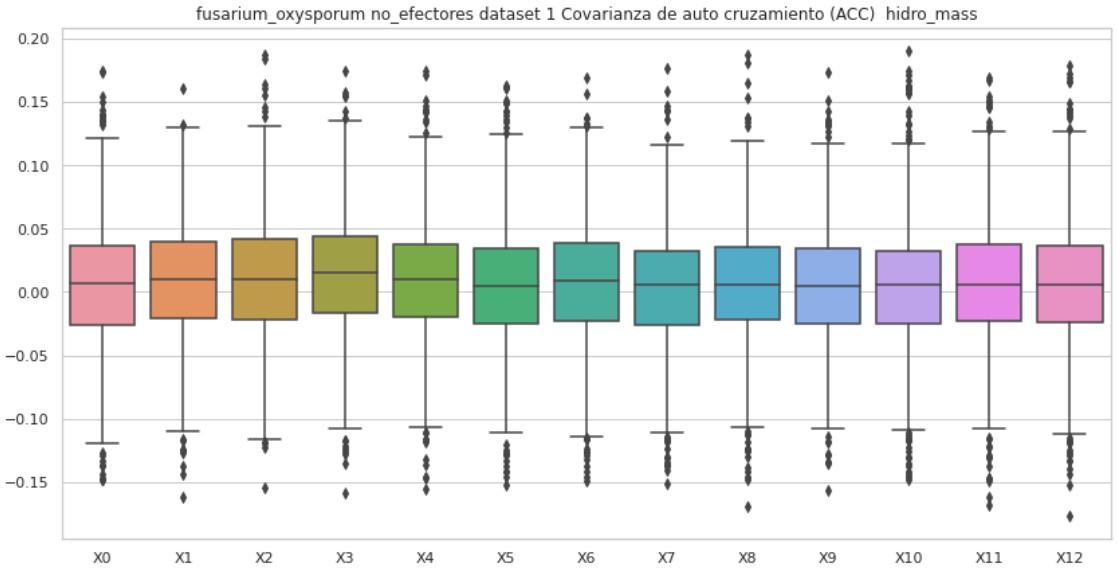
no_efectores
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores fusarium_oxysporum
dataset 1, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0   -0.015650 -0.011331  0.028389  0.032013 -0.022398 -0.017179  0.025394
1    0.041161  0.030963 -0.000365 -0.018599  0.029021 -0.003269 -0.039617
2    0.010331  0.040050  0.005810  0.001890 -0.051507  0.032012  0.008658
3    0.022248 -0.030544  0.002908 -0.015868  0.014437 -0.027792  0.010491
4   -0.000055  0.029249  0.028753  0.032387  0.027074  0.003731  0.041759
..        …         …         …         …         …         …
995  0.052177 -0.006901 -0.005753  0.029276  0.000347  0.010830 -0.020796
996 -0.046018  0.013061  0.037168 -0.009167 -0.024973  0.060783  0.090649
997  0.000815 -0.022999  0.010311  0.076356  0.015550  0.082516  0.063399
998  0.034308  0.051967  0.081666  0.091435  0.023750  0.056765  0.038937
999  0.013296 -0.006418 -0.001484 -0.006932  0.029627 -0.018526  0.006531

           X7        X8        X9       X10       X11       X12           X13
0   -0.019560 -0.045433  0.019501  0.006152 -0.033507 -0.001731  no_efectores
1   -0.010741  0.024122  0.012430  0.015638  0.024028 -0.009259  no_efectores
2   -0.000690  0.047390  0.001453 -0.001497  0.019164  0.003748  no_efectores
3   -0.005095  0.054626  0.004013 -0.018259 -0.012590 -0.003643  no_efectores
4   -0.001602 -0.084648  0.014938 -0.022521  0.053480 -0.006299  no_efectores
..        …         …         …         …         …         …           …
995  0.009640  0.078996 -0.057196  0.029834 -0.010140 -0.017023  no_efectores
996  0.011661  0.039312 -0.019038  0.008169 -0.008315 -0.011261  no_efectores
997 -0.052386 -0.018233 -0.036478  0.093462  0.151383  0.067000  no_efectores
998  0.058807  0.027715  0.010362  0.031771  0.007733 -0.008922  no_efectores
999 -0.034849 -0.008139 -0.092508 -0.007543 -0.055925 -0.015299  no_efectores

[912 rows x 14 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  912.000000  912.000000  912.000000  912.000000  912.000000  912.000000
mean     0.005156    0.008186    0.010298    0.013103    0.008633    0.005146
std      0.050231    0.046643    0.050200    0.047904    0.047643    0.049874
min     -0.148503   -0.161618   -0.154155   -0.158553   -0.155914   -0.152119
25%     -0.026234   -0.021064   -0.021867   -0.016843   -0.020121   -0.025339
50%      0.006467    0.010292    0.010356    0.014870    0.009776    0.005126
75%      0.036307    0.039914    0.041395    0.044352    0.037329    0.034613
max      0.173941    0.161038    0.186806    0.174103    0.173955    0.162811

                X6          X7          X8          X9         X10         X11  \
count  912.000000  912.000000  912.000000  912.000000  912.000000  912.000000
mean     0.007674    0.004185    0.004676    0.004412    0.006042    0.006669
std      0.049601    0.046616    0.047350    0.046362    0.050484    0.050029
```

min	-0.149266	-0.151005	-0.168995	-0.156808	-0.148221	-0.167786
25%	-0.022754	-0.025708	-0.021942	-0.024941	-0.024969	-0.022624
50%	0.008422	0.005419	0.005989	0.004771	0.005399	0.005403
75%	0.038406	0.032717	0.035327	0.033976	0.032088	0.037242
max	0.169436	0.177030	0.187374	0.173685	0.190253	0.169126

	X12
count	912.000000
mean	0.006107
std	0.049449
min	-0.176716
25%	-0.023954
50%	0.005963
75%	0.036415
max	0.178993





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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.059361  0.058924  0.043385  0.019266 -0.050025 -0.012213 -0.049867
1   -0.038358 -0.022014  0.084707  0.028736 -0.011468  0.028057  0.035392
2    0.084211  0.036179  0.016931  0.032327 -0.004568  0.087915  0.025747
3    0.012996  0.062573  0.044595 -0.022979 -0.019777 -0.000854 -0.058178
4    0.028805  0.067778 -0.075469  0.054776 -0.039264  0.018130 -0.043181
..        …         …         …         …         …         …         …
995 -0.021303  0.004344 -0.047501 -0.030772 -0.021117  0.034647 -0.019821
996  0.020442 -0.036786  0.004765 -0.182379 -0.007104  0.016541  0.223794
997  0.034182  0.050604 -0.007599  0.058100 -0.004085 -0.027304  0.000403
998 -0.043523 -0.045795  0.017393 -0.098307  0.048070 -0.054248  0.017152
999  0.083563  0.017567 -0.024205 -0.123925  0.089546  0.159116 -0.015107

           X7        X8        X9       X10       X11       X12        X13
0   -0.057264  0.007241  0.079565  0.013705  0.058798 -0.000752  efectores
1   -0.000954 -0.035263  0.034839 -0.019392  0.021979  0.038419  efectores
2    0.032613  0.074832  0.042960 -0.012147  0.046505 -0.035839  efectores
3   -0.007489 -0.011409  0.003605  0.034205  0.039099  0.072545  efectores
4   -0.020093  0.040531 -0.011302 -0.029099 -0.025713  0.027003  efectores
..        …         …         …         …         …         …         …
995  0.080332  0.106620 -0.095693 -0.011724  0.011147  0.037502  efectores
996  0.095954  0.012273 -0.018609  0.006520 -0.006303 -0.055766  efectores
997  0.004985  0.003930 -0.047405 -0.097623  0.034636 -0.019929  efectores
998 -0.036561 -0.057524  0.106336 -0.024249 -0.009742 -0.036122  efectores
999  0.112711 -0.004530 -0.045630 -0.035828 -0.145422  0.023255  efectores

[1000 rows x 14 columns]


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

                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.007914     0.015334     0.013748     0.013974     0.003726
```

```
std       0.059289     0.060870     0.060545     0.062419     0.060000
min      -0.333863    -0.227512    -0.255981    -0.444166    -0.378492
25%      -0.026228    -0.019780    -0.019805    -0.017625    -0.027220
50%       0.009159     0.013294     0.015488     0.016071     0.005539
75%       0.042773     0.046534     0.046278     0.049796     0.037462
max       0.277788     0.320850     0.296214     0.268561     0.234797

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.005929     0.012089     0.004652     0.006341     0.005663
std       0.060220     0.058654     0.057759     0.059293     0.059467
min      -0.303600    -0.291733    -0.327682    -0.314630    -0.350388
25%      -0.025608    -0.020385    -0.027651    -0.028212    -0.026803
50%       0.005079     0.011898     0.004618     0.006235     0.006710
75%       0.037652     0.043938     0.038454     0.041485     0.038299
max       0.434801     0.363294     0.291784     0.249002     0.330457

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.006423     0.006220     0.003878
std       0.061849     0.057005     0.062113
min      -0.291373    -0.321068    -0.341191
25%      -0.027370    -0.025269    -0.027726
50%       0.004683     0.006542     0.007091
75%       0.036773     0.037836     0.037599
max       0.346803     0.274804     0.314213



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

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.015650 -0.011331  0.028389  0.032013 -0.022398 -0.017179  0.025394
1    0.041161  0.030963 -0.000365 -0.018599  0.029021 -0.003269 -0.039617
2    0.010331  0.040050  0.005810  0.001890 -0.051507  0.032012  0.008658
3    0.022248 -0.030544  0.002908 -0.015868  0.014437 -0.027792  0.010491
4   -0.000055  0.029249  0.028753  0.032387  0.027074  0.003731  0.041759
..        …         …         …         …         …         …         …
995  0.052177 -0.006901 -0.005753  0.029276  0.000347  0.010830 -0.020796
996 -0.046018  0.013061  0.037168 -0.009167 -0.024973  0.060783  0.090649
997  0.000815 -0.022999  0.010311  0.076356  0.015550  0.082516  0.063399
998  0.034308  0.051967  0.081666  0.091435  0.023750  0.056765  0.038937
999  0.013296 -0.006418 -0.001484 -0.006932  0.029627 -0.018526  0.006531

           X7        X8        X9       X10       X11       X12           X13
```

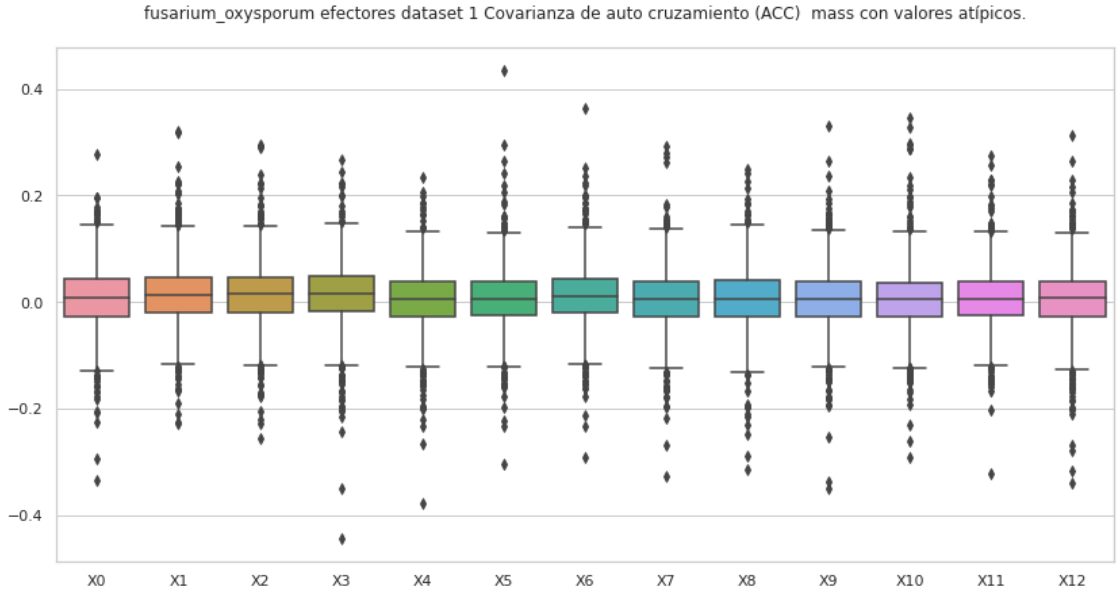
```
0   -0.019560 -0.045433  0.019501  0.006152 -0.033507 -0.001731  no_efectores
1   -0.010741  0.024122  0.012430  0.015638  0.024028 -0.009259  no_efectores
2   -0.000690  0.047390  0.001453 -0.001497  0.019164  0.003748  no_efectores
3   -0.005095  0.054626  0.004013 -0.018259 -0.012590 -0.003643  no_efectores
4   -0.001602 -0.084648  0.014938 -0.022521  0.053480 -0.006299  no_efectores
..        …       …       …       …       …       …           …
995  0.009640  0.078996 -0.057196  0.029834 -0.010140 -0.017023  no_efectores
996  0.011661  0.039312 -0.019038  0.008169 -0.008315 -0.011261  no_efectores
997 -0.052386 -0.018233 -0.036478  0.093462  0.151383  0.067000  no_efectores
998  0.058807  0.027715  0.010362  0.031771  0.007733 -0.008922  no_efectores
999 -0.034849 -0.008139 -0.092508 -0.007543 -0.055925 -0.015299  no_efectores

[1000 rows x 14 columns]
```

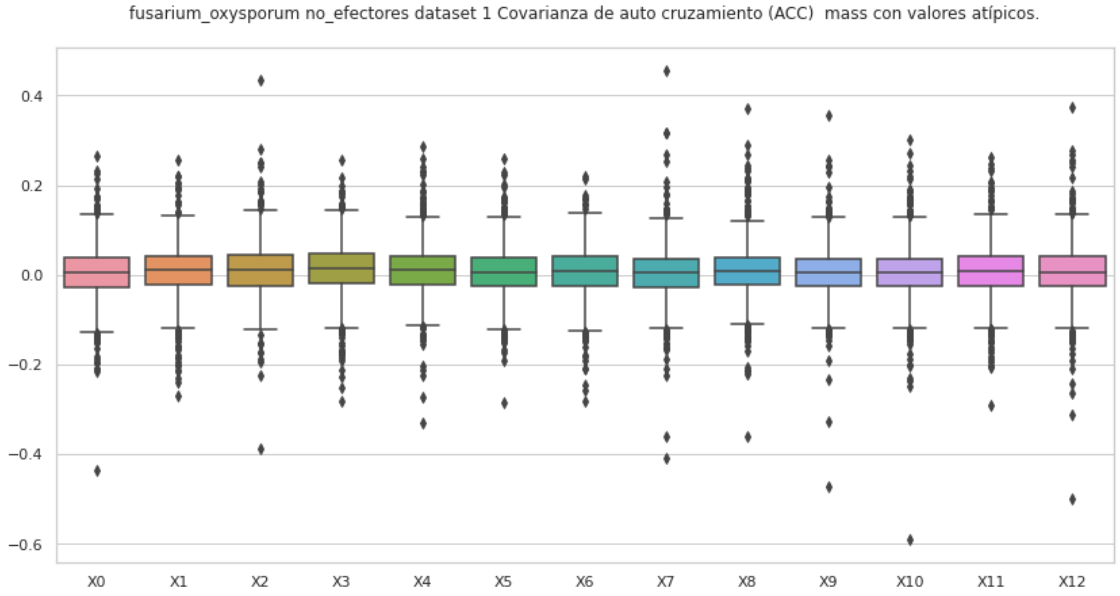
Covarianza de auto cruzamiento (ACC) mass no_efectores fusarium_oxysporum
dataset 1, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.003693	0.007011	0.010673	0.012145	0.010608	
std	0.059905	0.058273	0.059600	0.057878	0.057683	
min	-0.434742	-0.270548	-0.387666	-0.281609	-0.329958	
25%	-0.029524	-0.023009	-0.024225	-0.019131	-0.021426	
50%	0.005548	0.009745	0.010408	0.015202	0.010058	
75%	0.036910	0.040853	0.043179	0.046941	0.039947	
max	0.266218	0.254689	0.433576	0.257133	0.284839	

	X5	X6	X7	X8	X9	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.006093	0.006912	0.004509	0.007801	0.006284	
std	0.056198	0.058348	0.059442	0.059892	0.057307	
min	-0.284003	-0.282700	-0.409931	-0.361897	-0.471840	
25%	-0.026162	-0.026109	-0.026675	-0.021942	-0.026013	
50%	0.005675	0.008179	0.005655	0.007067	0.005440	
75%	0.036955	0.040659	0.034822	0.037854	0.036375	
max	0.260137	0.220528	0.455226	0.371280	0.355382	

	X10	X11	X12
count	1000.000000	1000.000000	1000.000000
mean	0.005545	0.007547	0.007818
std	0.062004	0.059849	0.062538
min	-0.590356	-0.291501	-0.500593
25%	-0.026603	-0.024145	-0.024871
50%	0.005320	0.006666	0.006398
75%	0.036109	0.040272	0.040003
max	0.302310	0.261207	0.373581


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


fusarium_oxysporum no_efectores dataset 1 Covarianza de auto cruzamiento (ACC) mass con valores atípicos.
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

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))
```

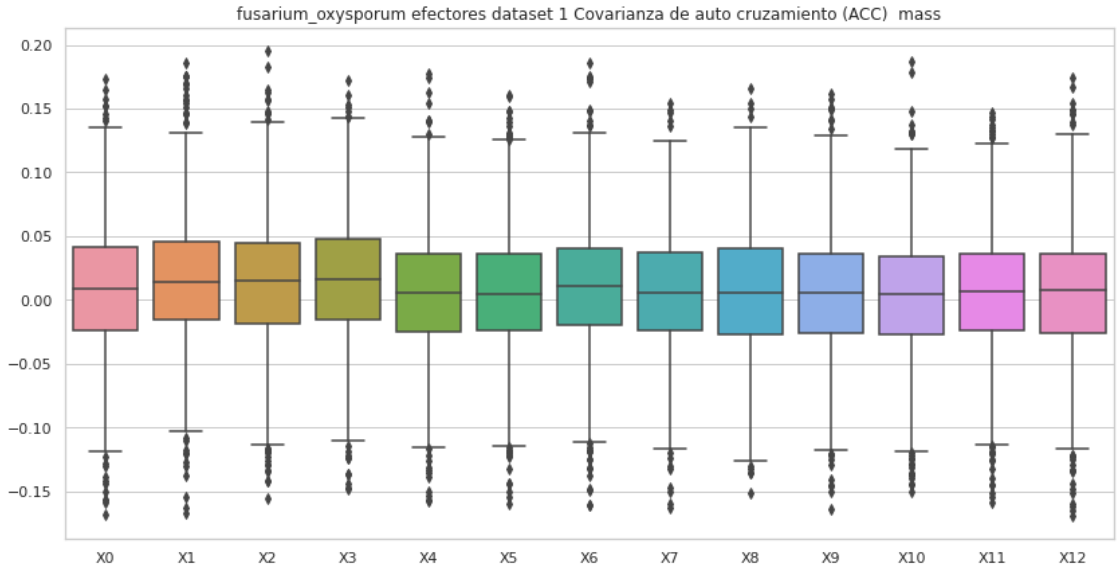
```
Covarianza de auto cruzamiento (ACC)  mass efectores fusarium_oxysporum dataset
1, sin valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.059361  0.058924  0.043385  0.019266 -0.050025 -0.012213 -0.049867
1   -0.038358 -0.022014  0.084707  0.028736 -0.011468  0.028057  0.035392
2    0.084211  0.036179  0.016931  0.032327 -0.004568  0.087915  0.025747
3    0.012996  0.062573  0.044595 -0.022979 -0.019777 -0.000854 -0.058178
4    0.028805  0.067778 -0.075469  0.054776 -0.039264  0.018130 -0.043181
..        …         …         …         …         …         …         …
994  0.065591 -0.003433  0.036059  0.047216 -0.088258  0.090745  0.102642
995 -0.021303  0.004344 -0.047501 -0.030772 -0.021117  0.034647 -0.019821
997  0.034182  0.050604 -0.007599  0.058100 -0.004085 -0.027304  0.000403
998 -0.043523 -0.045795  0.017393 -0.098307  0.048070 -0.054248  0.017152
999  0.083563  0.017567 -0.024205 -0.123925  0.089546  0.159116 -0.015107

           X7        X8        X9       X10       X11       X12         X13
0   -0.057264  0.007241  0.079565  0.013705  0.058798 -0.000752  efectores
1   -0.000954 -0.035263  0.034839 -0.019392  0.021979  0.038419  efectores
2    0.032613  0.074832  0.042960 -0.012147  0.046505 -0.035839  efectores
3   -0.007489 -0.011409  0.003605  0.034205  0.039099  0.072545  efectores
4   -0.020093  0.040531 -0.011302 -0.029099 -0.025713  0.027003  efectores
..        …         …         …         …         …         …         …
994  0.009412  0.012955 -0.084085 -0.034807 -0.016416  0.018030  efectores
995  0.080332  0.106620 -0.095693 -0.011724  0.011147  0.037502  efectores
997  0.004985  0.003930 -0.047405 -0.097623  0.034636 -0.019929  efectores
998 -0.036561 -0.057524  0.106336 -0.024249 -0.009742 -0.036122  efectores
999  0.112711 -0.004530 -0.045630 -0.035828 -0.145422  0.023255  efectores

[907 rows x 14 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  907.000000  907.000000  907.000000  907.000000  907.000000  907.000000
mean     0.008199    0.015381    0.012643    0.015676    0.005422    0.005276
std      0.051217    0.051843    0.051333    0.050192    0.050216    0.050256
min     -0.168255   -0.167004   -0.155818   -0.148647   -0.157364   -0.159943
25%     -0.024131   -0.016029   -0.018935   -0.015706   -0.024804   -0.024283
50%      0.009054    0.013919    0.015397    0.016054    0.005682    0.004853
75%      0.040908    0.045568    0.045082    0.047749    0.036236    0.035708
```

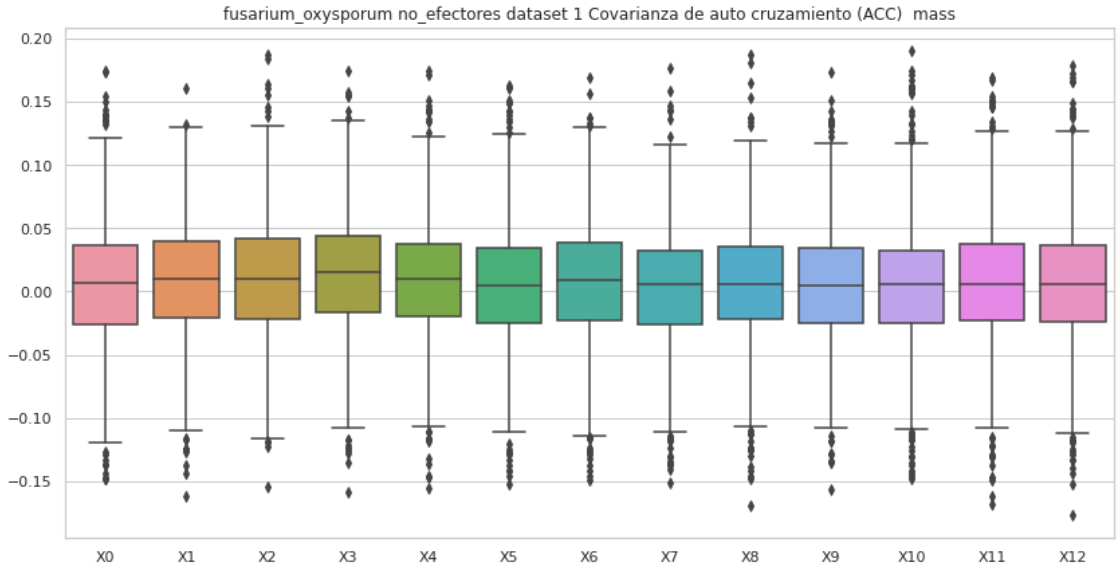
```
max      0.172623    0.186005    0.195200    0.171822    0.177678    0.160696

               X6          X7          X8          X9         X10         X11  \
count  907.000000  907.000000  907.000000  907.000000  907.000000  907.000000
mean     0.010138    0.005935    0.006512    0.004624    0.004376    0.005506
std      0.049667    0.048062    0.050058    0.048604    0.049494    0.048370
min     -0.161446   -0.163397   -0.151329   -0.164149   -0.150826   -0.159206
25%     -0.020019   -0.024372   -0.027215   -0.026432   -0.027260   -0.023968
50%      0.010926    0.005940    0.005768    0.005695    0.004295    0.006766
75%      0.040845    0.037494    0.040144    0.036367    0.034272    0.036065
max      0.185500    0.154017    0.165896    0.161853    0.187291    0.146926

              X12
count  907.000000
mean     0.004883
std      0.050951
min     -0.169414
25%     -0.025849
50%      0.007257
75%      0.036636
max      0.173802
```

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

```
           X0        X1        X2        X3        X4        X5        X6  \
0   -0.015650 -0.011331  0.028389  0.032013 -0.022398 -0.017179  0.025394
1    0.041161  0.030963 -0.000365 -0.018599  0.029021 -0.003269 -0.039617
2    0.010331  0.040050  0.005810  0.001890 -0.051507  0.032012  0.008658
3    0.022248 -0.030544  0.002908 -0.015868  0.014437 -0.027792  0.010491
4   -0.000055  0.029249  0.028753  0.032387  0.027074  0.003731  0.041759
..        …         …         …         …         …         …
995  0.052177 -0.006901 -0.005753  0.029276  0.000347  0.010830 -0.020796
996 -0.046018  0.013061  0.037168 -0.009167 -0.024973  0.060783  0.090649
997  0.000815 -0.022999  0.010311  0.076356  0.015550  0.082516  0.063399
998  0.034308  0.051967  0.081666  0.091435  0.023750  0.056765  0.038937
999  0.013296 -0.006418 -0.001484 -0.006932  0.029627 -0.018526  0.006531

           X7        X8        X9       X10       X11       X12           X13
0   -0.019560 -0.045433  0.019501  0.006152 -0.033507 -0.001731  no_efectores
1   -0.010741  0.024122  0.012430  0.015638  0.024028 -0.009259  no_efectores
2   -0.000690  0.047390  0.001453 -0.001497  0.019164  0.003748  no_efectores
3   -0.005095  0.054626  0.004013 -0.018259 -0.012590 -0.003643  no_efectores
4   -0.001602 -0.084648  0.014938 -0.022521  0.053480 -0.006299  no_efectores
..        …         …         …         …         …         …           …
```

```
995  0.009640  0.078996 -0.057196  0.029834 -0.010140 -0.017023  no_efectores
996  0.011661  0.039312 -0.019038  0.008169 -0.008315 -0.011261  no_efectores
997 -0.052386 -0.018233 -0.036478  0.093462  0.151383  0.067000  no_efectores
998  0.058807  0.027715  0.010362  0.031771  0.007733 -0.008922  no_efectores
999 -0.034849 -0.008139 -0.092508 -0.007543 -0.055925 -0.015299  no_efectores

[912 rows x 14 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  912.000000  912.000000  912.000000  912.000000  912.000000  912.000000
mean     0.005156    0.008186    0.010298    0.013103    0.008633    0.005146
std      0.050231    0.046643    0.050200    0.047904    0.047643    0.049874
min     -0.148503   -0.161618   -0.154155   -0.158553   -0.155914   -0.152119
25%     -0.026234   -0.021064   -0.021867   -0.016843   -0.020121   -0.025339
50%      0.006467    0.010292    0.010356    0.014870    0.009776    0.005126
75%      0.036307    0.039914    0.041395    0.044352    0.037329    0.034613
max      0.173941    0.161038    0.186806    0.174103    0.173955    0.162811

                X6          X7          X8          X9         X10         X11  \
count  912.000000  912.000000  912.000000  912.000000  912.000000  912.000000
mean     0.007674    0.004185    0.004676    0.004412    0.006042    0.006669
std      0.049601    0.046616    0.047350    0.046362    0.050484    0.050029
min     -0.149266   -0.151005   -0.168995   -0.156808   -0.148221   -0.167786
25%     -0.022754   -0.025708   -0.021942   -0.024941   -0.024969   -0.022624
50%      0.008422    0.005419    0.005989    0.004771    0.005399    0.005403
75%      0.038406    0.032717    0.035327    0.033976    0.032088    0.037242
max      0.169436    0.177030    0.187374    0.173685    0.190253    0.169126

               X12
count  912.000000
mean     0.006107
std      0.049449
min     -0.176716
25%     -0.023954
50%      0.005963
75%      0.036415
max      0.178993
```





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

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.000461 -0.004334  0.023134  0.048356 -0.073241 -0.032438 -0.000373
1    0.063419  0.088255  0.068386  0.045264  0.079740  0.068848  0.019216
2    0.061592  0.001960  0.022593  0.015100  0.066094  0.018108  0.025488
3   -0.008914 -0.033415  0.037748 -0.051408 -0.047108 -0.014367  0.012157
4    0.054133  0.045064  0.071708 -0.000610 -0.017166  0.029958 -0.006605
..        …         …         …         …         …         …         …
995 -0.165410 -0.174248  0.070022 -0.077380  0.069009  0.004755 -0.018122
996 -0.047391 -0.042880 -0.029777  0.052456 -0.036528 -0.115430  0.137704
997  0.072776  0.050291  0.087403  0.135648  0.062473  0.121869  0.062115
998 -0.041551 -0.098190  0.020192  0.007920  0.034997 -0.041395  0.004266
999  0.027965  0.052821  0.101492 -0.072336 -0.092150 -0.021299  0.133091

           X7        X8        X9       X10       X11       X12        X13
0    0.023138  0.001188 -0.072609  0.020884  0.001565  0.004342  efectores
1    0.061047  0.117663  0.009719  0.028682  0.003266  0.003784  efectores
2   -0.014804 -0.019261 -0.023445 -0.017723 -0.037947 -0.023578  efectores
```

```
3   -0.015727  0.013308  0.022664  0.003331  0.030104  0.014595  efectores
4    0.022124  0.012886  0.008419  0.039666 -0.002734  0.092423  efectores
..        …         …         …         …         …         …         …
995  0.026898 -0.050885  0.016014  0.003488 -0.174671  0.088351  efectores
996 -0.091268 -0.044708 -0.003878  0.024228 -0.034586 -0.043934  efectores
997  0.108811  0.028741  0.049397  0.082839  0.104023  0.069161  efectores
998  0.007055  0.019687 -0.055182  0.002361 -0.048522  0.016139  efectores
999 -0.037161  0.067400  0.139877  0.046892 -0.018371 -0.006224  efectores

[1000 rows x 14 columns]
```

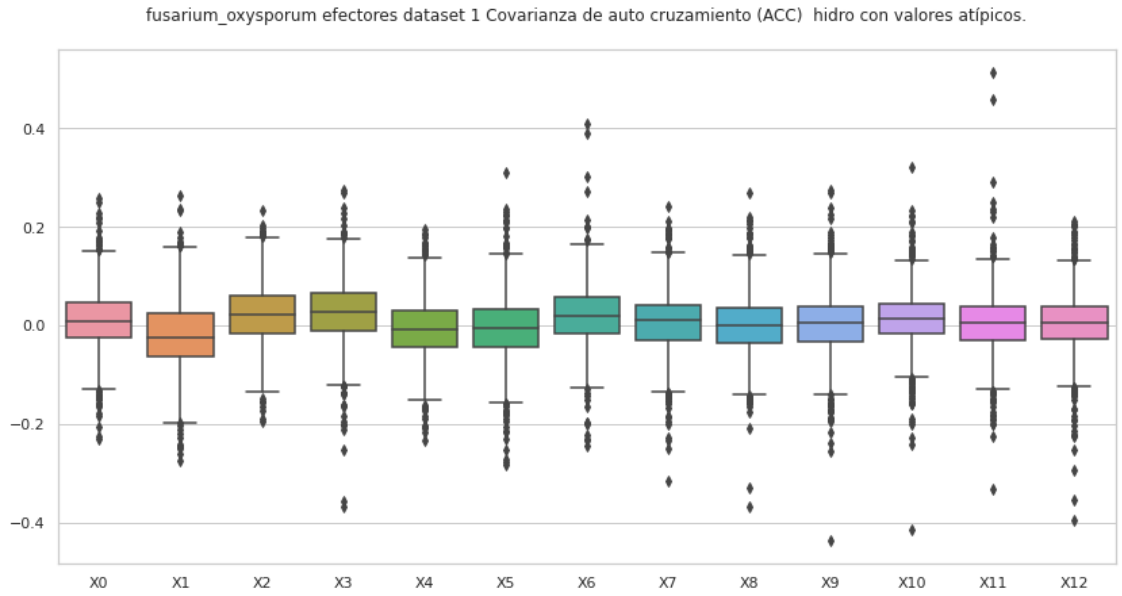
Covarianza de auto cruzamiento (ACC) hidro efectores fusarium_oxysporum dataset 1, con valores atípicos.
Estadísticas.

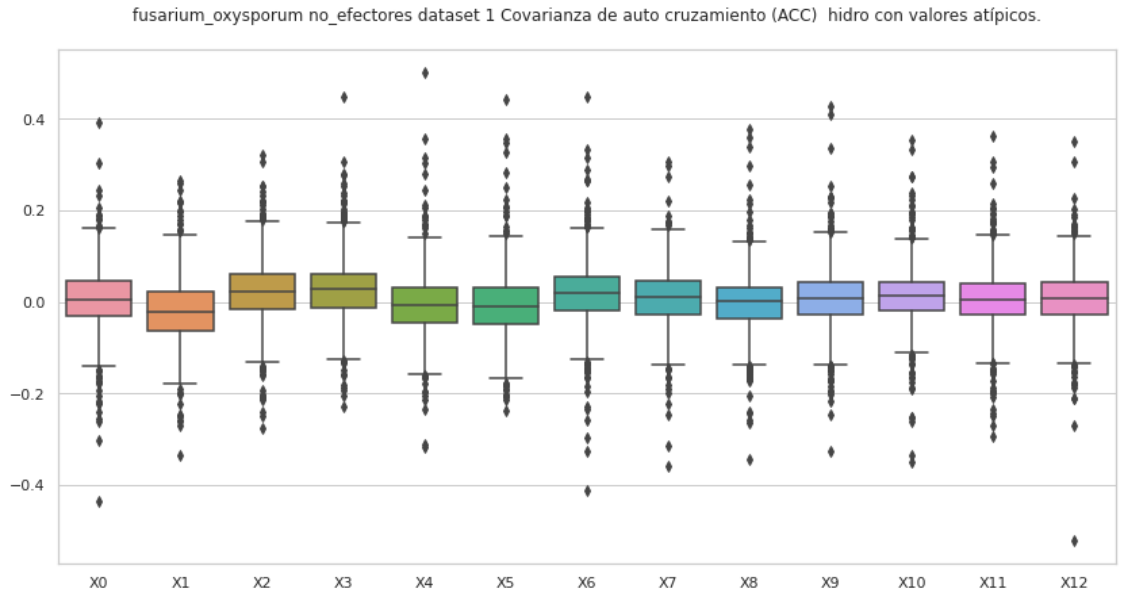
```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.010792    -0.018590     0.022191     0.025416    -0.005637
std       0.062895     0.070961     0.061910     0.064630     0.061156
min      -0.229807    -0.275394    -0.193897    -0.368315    -0.231928
25%      -0.024283    -0.063404    -0.016921    -0.010206    -0.043257
50%       0.009595    -0.023584     0.023253     0.027676    -0.007909
75%       0.046247     0.026048     0.061699     0.065315     0.029769
max       0.256589     0.262287     0.233177     0.273111     0.196188

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean     -0.004413     0.020944     0.006837     0.001172     0.004198
std       0.067638     0.063727     0.061649     0.061727     0.064736
min      -0.282717    -0.243354    -0.315024    -0.366816    -0.435824
25%      -0.043056    -0.016748    -0.028769    -0.034807    -0.032736
50%      -0.005332     0.020097     0.010372     0.000017     0.005007
75%       0.032497     0.057200     0.042175     0.036654     0.039003
max       0.310096     0.409180     0.242445     0.268177     0.273667

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.012149     0.004684     0.002431
std       0.059816     0.063213     0.062206
min      -0.413048    -0.331740    -0.395427
25%      -0.016591    -0.029200    -0.027136
50%       0.014300     0.004419     0.005136
75%       0.043034     0.037518     0.037365
max       0.322329     0.512739     0.211069
```

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

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.032370 -0.134751 -0.006392  0.081700 -0.007065 -0.039368  0.030710
1   -0.019036  0.001600 -0.015253 -0.036691 -0.030385  0.000430 -0.011471
2   -0.024921 -0.053537 -0.020730 -0.002957 -0.018104 -0.005434  0.010209
3   -0.006289 -0.060792  0.026345  0.037772 -0.039991 -0.053299 -0.016168
4    0.001523  0.000259  0.030574  0.031580  0.026940  0.002753  0.010906
..        …         …         …         …         …         …
995  0.032384 -0.013887 -0.009400  0.008690 -0.020652 -0.079187 -0.027817
996 -0.041964  0.032709 -0.018875  0.051993 -0.018307  0.047253  0.066275
997  0.139547  0.046858 -0.071239  0.090464  0.064284  0.132748  0.037334
998 -0.008488 -0.057757 -0.022654  0.095286  0.067037 -0.094220 -0.030364
999 -0.029405 -0.100193  0.000512  0.030789 -0.008638 -0.020228  0.079492

           X7        X8        X9       X10       X11       X12           X13
0    0.043421  0.007133 -0.001880  0.016711 -0.022774  0.004936  no_efectores
1    0.044044 -0.003215 -0.019064 -0.034927  0.000178 -0.037233  no_efectores
2    0.007483 -0.029241  0.019649  0.034152  0.016691 -0.002297  no_efectores
3   -0.011084 -0.049524 -0.015965 -0.025337  0.009799  0.022310  no_efectores
4   -0.004704  0.015278  0.040377  0.049544  0.036454 -0.000713  no_efectores
..        …         …         …         …         …         …           …
995  0.047186 -0.054502  0.019569  0.025412  0.007582  0.008733  no_efectores
996 -0.042687 -0.019275 -0.013873  0.028355  0.018417 -0.002534  no_efectores
997 -0.017399 -0.021123 -0.019476 -0.036941  0.003477  0.058594  no_efectores
998  0.031358 -0.030374  0.028633 -0.084257  0.055437 -0.069819  no_efectores
999 -0.027202 -0.038001  0.028929  0.003877 -0.043984 -0.001023  no_efectores

[1000 rows x 14 columns]


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

                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.006244    -0.020596     0.024611     0.027053    -0.004989
std       0.068535     0.069269     0.067456     0.068561     0.068488
min      -0.437135    -0.335778    -0.276763    -0.228541    -0.318298
25%      -0.030658    -0.062282    -0.015236    -0.014128    -0.045251
50%       0.003752    -0.021504     0.023254     0.027229    -0.005682
75%       0.046152     0.022096     0.061790     0.061522     0.031155
max       0.392726     0.265034     0.322876     0.449752     0.501671
```

	X5	X6	X7	X8	X9	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	-0.007791	0.019309	0.007966	-0.000863	0.007721	
std	0.069720	0.069223	0.062575	0.065849	0.066224	
min	-0.238084	-0.412022	-0.358973	-0.344285	-0.327322	
25%	-0.047857	-0.017289	-0.028684	-0.035655	-0.028884	
50%	-0.008504	0.020201	0.010385	0.000707	0.006673	
75%	0.030961	0.055221	0.046796	0.032070	0.043345	
max	0.444050	0.447395	0.307536	0.378532	0.426511	

	X10	X11	X12
count	1000.000000	1000.000000	1000.000000
mean	0.012256	0.006213	0.006945
std	0.062331	0.065545	0.062624
min	-0.349086	-0.295332	-0.521648
25%	-0.018496	-0.028737	-0.028322
50%	0.012774	0.005818	0.007748
75%	0.044501	0.041221	0.042479
max	0.355163	0.362966	0.350121





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

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.000461 -0.004334  0.023134  0.048356 -0.073241 -0.032438 -0.000373
1    0.063419  0.088255  0.068386  0.045264  0.079740  0.068848  0.019216
2    0.061592  0.001960  0.022593  0.015100  0.066094  0.018108  0.025488
3   -0.008914 -0.033415  0.037748 -0.051408 -0.047108 -0.014367  0.012157
4    0.054133  0.045064  0.071708 -0.000610 -0.017166  0.029958 -0.006605
..        …         …         …         …         …         …         …
995 -0.165410 -0.174248  0.070022 -0.077380  0.069009  0.004755 -0.018122
996 -0.047391 -0.042880 -0.029777  0.052456 -0.036528 -0.115430  0.137704
997  0.072776  0.050291  0.087403  0.135648  0.062473  0.121869  0.062115
998 -0.041551 -0.098190  0.020192  0.007920  0.034997 -0.041395  0.004266
999  0.027965  0.052821  0.101492 -0.072336 -0.092150 -0.021299  0.133091

           X7        X8        X9       X10       X11       X12        X13
0    0.023138  0.001188 -0.072609  0.020884  0.001565  0.004342  efectores
1    0.061047  0.117663  0.009719  0.028682  0.003266  0.003784  efectores
2   -0.014804 -0.019261 -0.023445 -0.017723 -0.037947 -0.023578  efectores
3   -0.015727  0.013308  0.022664  0.003331  0.030104  0.014595  efectores
4    0.022124  0.012886  0.008419  0.039666 -0.002734  0.092423  efectores
..        …         …         …         …         …         …         …
995  0.026898 -0.050885  0.016014  0.003488 -0.174671  0.088351  efectores
```

```
996 -0.091268 -0.044708 -0.003878  0.024228 -0.034586 -0.043934  efectores
997  0.108811  0.028741  0.049397  0.082839  0.104023  0.069161  efectores
998  0.007055  0.019687 -0.055182  0.002361 -0.048522  0.016139  efectores
999 -0.037161  0.067400  0.139877  0.046892 -0.018371 -0.006224  efectores

[917 rows x 14 columns]
```

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

	X0	X1	X2	X3	X4	X5
count	917.000000	917.000000	917.000000	917.000000	917.000000	917.000000
mean	0.010437	-0.019125	0.023276	0.027290	-0.005727	-0.004637
std	0.056860	0.064680	0.056958	0.055089	0.055000	0.057392
min	-0.165410	-0.218361	-0.155910	-0.163846	-0.171304	-0.193761
25%	-0.023286	-0.063015	-0.015011	-0.007599	-0.041526	-0.041545
50%	0.009065	-0.023993	0.023845	0.028818	-0.008308	-0.006043
75%	0.043864	0.024179	0.061578	0.063574	0.027396	0.029794
max	0.191563	0.188209	0.184377	0.217546	0.163277	0.197026

	X6	X7	X8	X9	X10	X11
count	917.000000	917.000000	917.000000	917.000000	917.000000	917.000000
mean	0.021161	0.007839	0.000371	0.004356	0.012497	0.004464
std	0.053738	0.054147	0.052284	0.053447	0.050262	0.050438
min	-0.149570	-0.166676	-0.174674	-0.189888	-0.158480	-0.183561
25%	-0.013907	-0.026747	-0.032506	-0.030918	-0.015198	-0.027002
50%	0.020442	0.010480	-0.000154	0.005095	0.014152	0.004669
75%	0.055703	0.041472	0.034113	0.036641	0.041645	0.036549
max	0.200916	0.185434	0.185462	0.166304	0.188271	0.177178

	X12
count	917.000000
mean	0.004655
std	0.051287
min	-0.182470
25%	-0.024040
50%	0.006461
75%	0.035966
max	0.188444

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

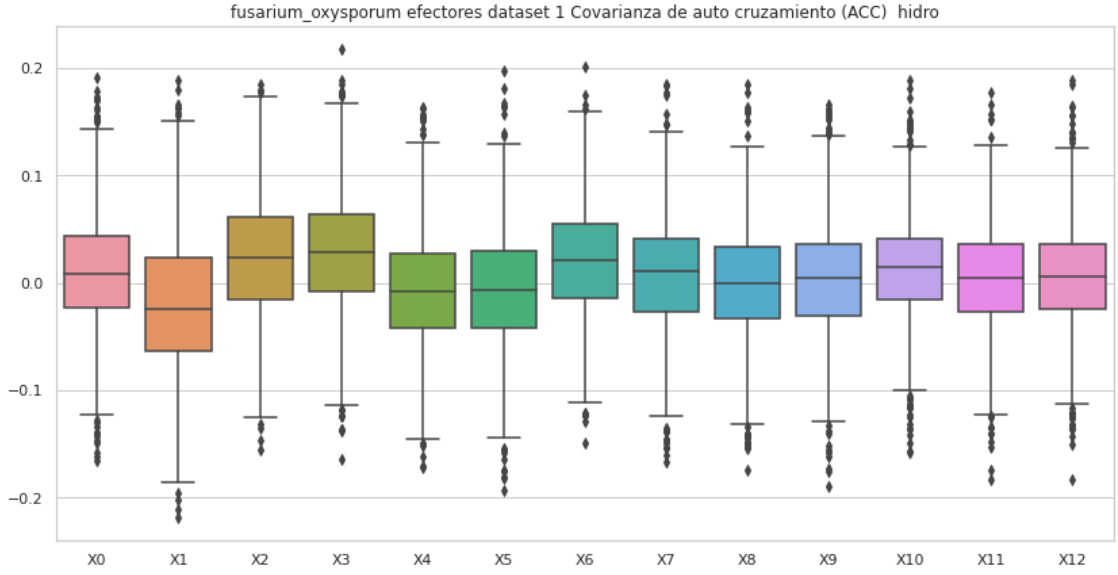
```
           X0        X1        X2        X3        X4        X5        X6  \
0   -0.032370 -0.134751 -0.006392  0.081700 -0.007065 -0.039368  0.030710
1   -0.019036  0.001600 -0.015253 -0.036691 -0.030385  0.000430 -0.011471
2   -0.024921 -0.053537 -0.020730 -0.002957 -0.018104 -0.005434  0.010209
3   -0.006289 -0.060792  0.026345  0.037772 -0.039991 -0.053299 -0.016168
4    0.001523  0.000259  0.030574  0.031580  0.026940  0.002753  0.010906
..        …         …         …         …         …         …
995  0.032384 -0.013887 -0.009400  0.008690 -0.020652 -0.079187 -0.027817
996 -0.041964  0.032709 -0.018875  0.051993 -0.018307  0.047253  0.066275
997  0.139547  0.046858 -0.071239  0.090464  0.064284  0.132748  0.037334
998 -0.008488 -0.057757 -0.022654  0.095286  0.067037 -0.094220 -0.030364
999 -0.029405 -0.100193  0.000512  0.030789 -0.008638 -0.020228  0.079492

           X7        X8        X9       X10       X11       X12           X13
0    0.043421  0.007133 -0.001880  0.016711 -0.022774  0.004936  no_efectores
1    0.044044 -0.003215 -0.019064 -0.034927  0.000178 -0.037233  no_efectores
2    0.007483 -0.029241  0.019649  0.034152  0.016691 -0.002297  no_efectores
3   -0.011084 -0.049524 -0.015965 -0.025337  0.009799  0.022310  no_efectores
4   -0.004704  0.015278  0.040377  0.049544  0.036454 -0.000713  no_efectores
..        …         …         …         …         …         …
995  0.047186 -0.054502  0.019569  0.025412  0.007582  0.008733  no_efectores
996 -0.042687 -0.019275 -0.013873  0.028355  0.018417 -0.002534  no_efectores
997 -0.017399 -0.021123 -0.019476 -0.036941  0.003477  0.058594  no_efectores
998  0.031358 -0.030374  0.028633 -0.084257  0.055437 -0.069819  no_efectores
999 -0.027202 -0.038001  0.028929  0.003877 -0.043984 -0.001023  no_efectores

[920 rows x 14 columns]
```

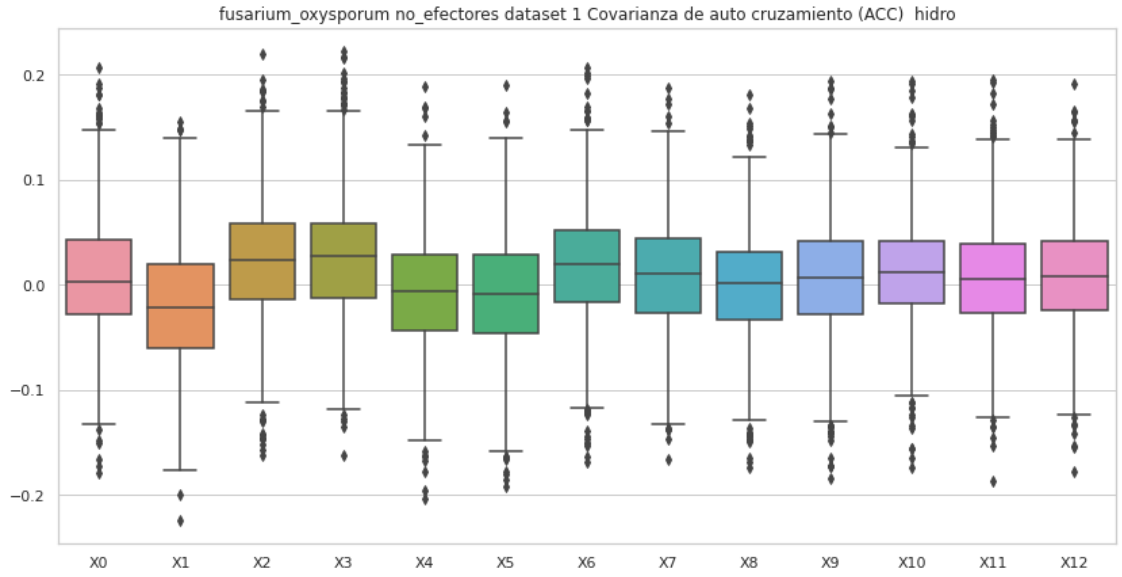
Covarianza de auto cruzamiento (ACC) no_efectores fusarium_oxysporum dataset 1, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  920.000000  920.000000  920.000000  920.000000  920.000000  920.000000
mean     0.008326   -0.021647    0.023896    0.025679   -0.005948   -0.009214
std      0.058056    0.059695    0.057562    0.058804    0.056900    0.057807
min     -0.178954   -0.224199   -0.161899   -0.161901   -0.203907   -0.192297
25%     -0.028108   -0.060913   -0.013636   -0.012065   -0.043354   -0.046388
50%      0.003514   -0.021504    0.023431    0.027229   -0.005682   -0.008504
75%      0.042938    0.020091    0.058312    0.059013    0.028076    0.028481
max      0.206654    0.155261    0.219544    0.222164    0.188421    0.189456

                X6          X7          X8          X9         X10         X11  \
count  920.000000  920.000000  920.000000  920.000000  920.000000  920.000000
mean     0.018772    0.008497   -0.000170    0.006536    0.011615    0.006410
std      0.055767    0.054179    0.054210    0.055137    0.050246    0.054223
```

min	-0.168109	-0.166661	-0.173315	-0.184082	-0.174135	-0.187300
25%	-0.015916	-0.026696	-0.033532	-0.027817	-0.017481	-0.027136
50%	0.019962	0.010621	0.002097	0.007138	0.012469	0.005848
75%	0.051936	0.044143	0.031147	0.041029	0.042185	0.038912
max	0.206925	0.187617	0.180566	0.193746	0.194296	0.194907

	X12
count	920.000000
mean	0.008514
std	0.052368
min	-0.178303
25%	-0.024707
50%	0.008600
75%	0.041915
max	0.191896




fusarium_oxysporum no_efectores dataset 1 Covarianza de auto cruzamiento (ACC) hidro
0.2
0.1
0.0
−0.1
−0.2
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12