ds1_archaea_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 ="archaea"
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 archaea dataset 1, con valores
atípicos.
Valores del documento csv.

         X0     X1     X2     X3     X4      X5     X6      X7     X8      X9  \
0     6.566  4.040  5.051  6.566  0.000   3.535  1.010   9.596  2.020  15.152
1    11.650  4.854  3.236  5.502  1.618   6.472  1.942  12.945  0.647   3.236
2    12.857  8.571  1.429  8.571  2.857   8.571  2.857   5.714  0.000   4.286
3     2.353  7.059  4.706  1.176  0.000   2.353  5.882  11.765  2.353   7.059
4     8.434  1.205  2.410  6.024  1.205  10.843  1.205   6.024  1.205   6.024
..      …      …      …      …      …       …      …       …      …       …
995   7.263  8.380  3.352  7.263  1.490   8.194  1.676   6.890  2.048   7.263
996   9.266  5.019  0.772  2.703  0.000   2.317  2.317   9.266  0.772   4.633
997  21.316  7.105  0.789  4.737  0.000   2.105  0.263   9.737  1.579   0.789
998   5.578  5.578  5.179  5.578  0.398   8.367  2.789   5.179  1.992   7.570
999   7.263  1.676  3.352  1.676  0.559   0.559  1.117  12.849  5.028   8.380
```

```
     …     X11    X12    X13    X14    X15    X16    X17    X18     X19  \
0    …  10.606  2.020  3.535  5.556  4.040  5.051  0.000  6.061   5.051
1    …   0.971  2.265  3.883  5.178  7.120  9.061  1.294  3.236   7.767
2    …   4.286  4.286  2.857  4.286  1.429  2.857  0.000  2.857   8.571
3    …  14.118  2.353  3.529  4.706  3.529  5.882  1.176  2.353  14.118
4    …  14.458  2.410  0.000  7.229  6.024  3.614  1.205  3.614   9.639
..   …      …      …      …      …      …      …      …      …
995  …   4.842  3.538  4.469  5.400  5.587  4.842  2.048  3.538   3.724
996  …   1.158  2.317  4.633  3.089  7.336  5.792  0.772  4.247  17.375
997  …   0.000  0.526  3.684  6.579  3.421  6.842  1.842  1.842  13.421
998  …   8.566  1.793  5.578  5.976  3.984  4.582  1.195  4.183   5.777
999  …   2.235  0.559  7.263  5.028  5.587  8.380  1.676  2.793   6.145

            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 archaea dataset 1, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      9.118060     6.047344     2.862079     5.918869     0.657181
std       4.608496     2.711928     2.206975     2.735271     0.880402
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       5.549000     4.167000     1.195000     3.797000     0.000000
50%       8.625000     5.909000     2.326000     5.528000     0.385000
75%      12.138000     7.787500     4.167000     7.597500     0.943000
max      24.791000    19.588000    12.689000    16.770000     4.969000

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      7.091078     2.420070     7.520467     1.732860     5.782507
std       3.634599     1.657847     2.849248     1.247369     3.321188
```

min	0.000000	0.000000	0.524000	0.000000	0.000000
25%	4.221500	1.282000	5.508500	0.772000	3.104250
50%	7.038500	2.156500	7.520500	1.584500	5.189500
75%	9.543250	3.170250	9.280000	2.500000	7.937750
max	18.478000	10.067000	19.301000	16.230000	18.750000

	X10	X11	X12	X13	X14 \
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	11.401743	4.673757	2.050011	3.900800	4.086679
std	3.457218	4.359361	1.264837	1.973956	1.746493
min	1.478000	0.000000	0.000000	0.000000	0.000000
25%	8.952500	1.114750	1.122250	2.532000	2.946250
50%	11.314500	2.837000	1.724000	3.786000	3.964000
75%	13.740250	7.799500	2.647500	5.030000	5.150000
max	27.083000	20.513000	10.112000	15.328000	13.043000

	X15	X16	X17	X18	X19
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	5.944855	5.491203	1.224950	3.367709	8.707728
std	2.236009	2.068974	1.024791	1.686504	3.584697
min	0.000000	0.000000	0.000000	0.000000	0.806000
25%	4.344500	4.075000	0.547500	2.277500	5.913250
50%	5.859000	5.352500	1.062000	3.175000	8.163500
75%	7.229000	6.785250	1.697500	4.242750	11.315000
max	20.197000	13.265000	7.487000	13.889000	19.880000

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

	X0	X1	X2	X3	X4	X5	X6	X7	X8 \
0	10.784	5.882	1.961	6.863	1.961	10.784	2.941	5.882	2.941
1	9.068	4.282	3.778	7.557	0.252	7.809	3.526	7.557	1.511
2	8.911	5.941	0.990	1.980	0.000	0.990	2.970	7.921	0.000
3	4.858	2.429	6.478	7.287	4.049	5.668	4.453	3.239	2.024
4	6.452	8.602	3.226	4.301	0.000	4.301	0.000	9.677	1.075
..	…	…	…	…	…	…	…	…	…
995	10.559	8.075	0.621	16.149	0.621	3.106	0.000	8.075	1.242
996	4.000	4.667	6.000	4.667	2.667	7.333	6.000	3.000	2.333
997	7.285	7.781	3.808	6.291	1.159	8.940	2.318	8.609	2.980
998	12.500	5.921	1.974	8.333	0.439	7.675	2.412	8.114	2.632
999	7.143	5.714	2.857	7.143	4.286	5.714	0.000	11.429	1.429

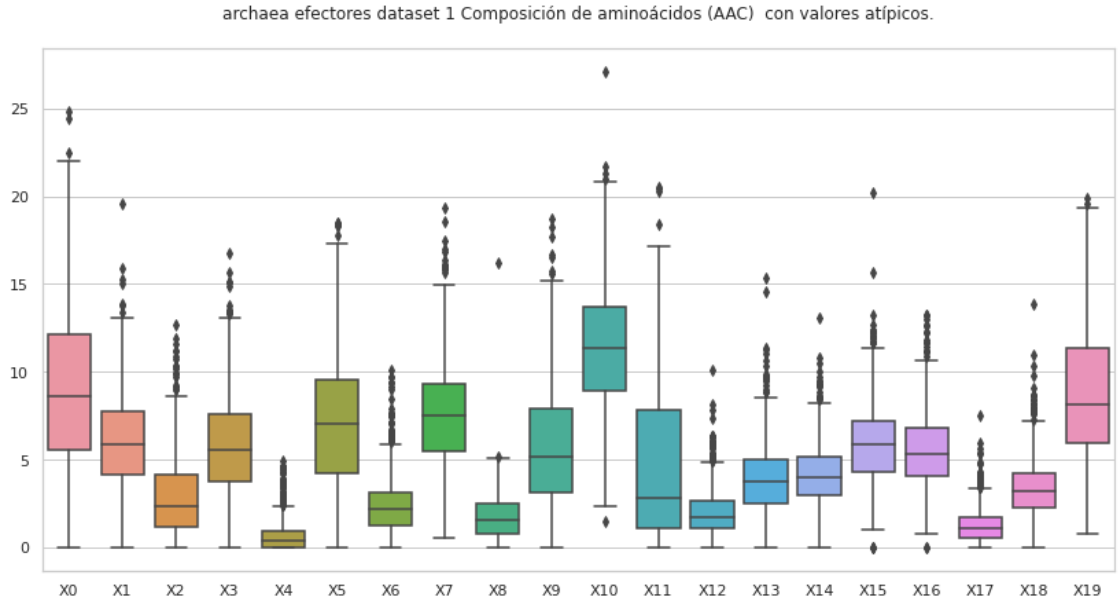
	X9	…	X11	X12	X13	X14	X15	X16	X17	X18 \
0	7.843	…	2.941	0.980	3.922	6.863	3.922	4.902	0.980	3.922

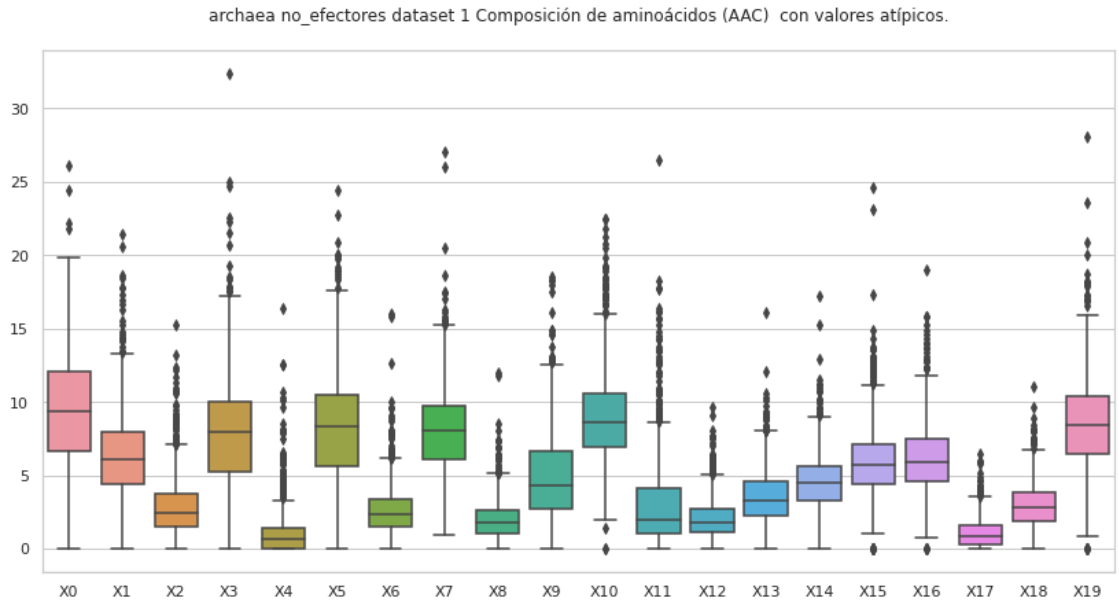
```
1     5.038  …   3.275  1.008  4.786  6.801   9.320   6.801  1.259  3.526
2     5.941  …   0.000  2.970  3.960  2.970   8.911   8.911  0.000  3.960
3     7.692  …  13.765  3.239  2.834  2.834   7.692   4.049  0.810  5.263
4     2.151  …   1.075  2.151  6.452  8.602   8.602   4.301  2.151  3.226
..      …  …     …    …    …    …      …     …    …    …
995   1.863  …   0.621  0.621  1.863  5.590  10.559  11.180  1.242  3.106
996  11.333  …   6.000  2.000  5.667  4.333   9.000   4.667  1.667  4.333
997   5.629  …   4.801  3.642  3.311  4.801   5.464   5.132  0.000  2.483
998   4.167  …   3.728  1.535  3.289  4.605   5.044   6.360  1.316  5.263
999   4.286  …  15.714  1.429  2.857  1.429   7.143   7.143  1.429  1.429

        X19            X20
0     5.882   no_efectores
1     6.045   no_efectores
2    12.871   no_efectores
3     2.834   no_efectores
4     9.677   no_efectores
..      …            …
995   9.317   no_efectores
996   2.667   no_efectores
997   8.113   no_efectores
998   8.114   no_efectores
999   5.714   no_efectores

[1000 rows x 21 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      9.532387     6.314992     2.879678     7.770255     1.084167
std       3.907225     2.959476     2.126248     3.630185     1.586237
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       6.622250     4.389750     1.464500     5.217000     0.000000
50%       9.380500     6.097000     2.442000     7.937000     0.668000
75%      12.091500     7.937000     3.725750    10.032500     1.370000
max      26.066000    21.429000    15.278000    32.364000    16.364000

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      8.200776     2.575877     8.000639     1.905816     4.929730
std       3.753682     1.770116     2.931036     1.364817     3.099709
min       0.000000     0.000000     0.935000     0.000000     0.000000
25%       5.631750     1.515000     6.046500     1.001500     2.675500
50%       8.292000     2.308000     8.022000     1.735500     4.286000
```

```
75%      10.484750     3.391750     9.723500     2.648250     6.607500
max      24.444000    16.000000    27.000000    12.000000    18.519000

                X10          X11          X12          X13          X14  \
count   1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean       8.980274     3.086971     2.092398     3.513206     4.503635
std        3.269611     3.165057     1.306708     1.916823     1.969996
min        0.000000     0.000000     0.000000     0.000000     0.000000
25%        6.893750     1.020500     1.153500     2.222000     3.268000
50%        8.565000     2.005000     1.798500     3.245500     4.461500
75%       10.526000     4.086000     2.718000     4.563500     5.594500
max       22.449000    26.471000     9.615000    16.129000    17.241000

                X15          X16          X17          X18          X19
count   1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean       5.975091     6.120461     1.093959     2.912363     8.527344
std        2.533543     2.414516     1.031221     1.545722     3.165516
min        0.000000     0.000000     0.000000     0.000000     0.000000
25%        4.438750     4.579250     0.310750     1.878250     6.466750
50%        5.682000     5.937000     0.885500     2.776500     8.393500
75%        7.145250     7.497000     1.601750     3.846000    10.353750
max       24.581000    19.000000     6.410000    11.024000    28.070000
```





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

         X0     X1     X2     X3     X4      X5     X6      X7     X8      X9  \
0     6.566  4.040  5.051  6.566  0.000   3.535  1.010   9.596  2.020  15.152
1    11.650  4.854  3.236  5.502  1.618   6.472  1.942  12.945  0.647   3.236
2    12.857  8.571  1.429  8.571  2.857   8.571  2.857   5.714  0.000   4.286
3     2.353  7.059  4.706  1.176  0.000   2.353  5.882  11.765  2.353   7.059
4     8.434  1.205  2.410  6.024  1.205  10.843  1.205   6.024  1.205   6.024
..      …      …      …      …      …       …      …       …      …       …
995   7.263  8.380  3.352  7.263  1.490   8.194  1.676   6.890  2.048   7.263
996   9.266  5.019  0.772  2.703  0.000   2.317  2.317   9.266  0.772   4.633
997  21.316  7.105  0.789  4.737  0.000   2.105  0.263   9.737  1.579   0.789
998   5.578  5.578  5.179  5.578  0.398   8.367  2.789   5.179  1.992   7.570
999   7.263  1.676  3.352  1.676  0.559   0.559  1.117  12.849  5.028   8.380

     …     X11    X12    X13    X14    X15    X16    X17    X18     X19  \
0    …  10.606  2.020  3.535  5.556  4.040  5.051  0.000  6.061   5.051
1    …   0.971  2.265  3.883  5.178  7.120  9.061  1.294  3.236   7.767
2    …   4.286  4.286  2.857  4.286  1.429  2.857  0.000  2.857   8.571
3    …  14.118  2.353  3.529  4.706  3.529  5.882  1.176  2.353  14.118
4    …  14.458  2.410  0.000  7.229  6.024  3.614  1.205  3.614   9.639
..   …      …      …      …      …      …      …      …      …       …
995  …   4.842  3.538  4.469  5.400  5.587  4.842  2.048  3.538   3.724
```

```
996  …   1.158  2.317  4.633  3.089  7.336  5.792  0.772  4.247  17.375
997  …   0.000  0.526  3.684  6.579  3.421  6.842  1.842  1.842  13.421
998  …   8.566  1.793  5.578  5.976  3.984  4.582  1.195  4.183   5.777
999  …   2.235  0.559  7.263  5.028  5.587  8.380  1.676  2.793   6.145

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

[873 rows x 21 columns]
```

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

```
                X0          X1          X2          X3          X4          X5  \
count  873.000000  873.000000  873.000000  873.000000  873.000000  873.000000
mean     9.421088    6.115538    2.704102    5.909593    0.574471    7.009608
std      4.569246    2.589550    2.019834    2.633874    0.707456    3.554589
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      5.814000    4.277000    1.136000    3.846000    0.000000    4.096000
50%      8.929000    5.993000    2.210000    5.528000    0.380000    7.000000
75%     12.434000    7.843000    3.968000    7.539000    0.855000    9.483000
max     22.436000   13.889000    9.231000   13.750000    2.994000   17.742000

                X6          X7          X8          X9         X10         X11  \
count  873.000000  873.000000  873.000000  873.000000  873.000000  873.000000
mean     2.345667    7.621140    1.751636    5.675415   11.565714    4.462553
std      1.461981    2.719463    1.164989    3.211088    3.349164    4.177898
min      0.000000    0.667000    0.000000    0.000000    2.381000    0.000000
25%      1.325000    5.634000    0.778000    3.086000    9.056000    1.058000
50%      2.155000    7.658000    1.613000    5.147000   11.504000    2.488000
75%      3.140000    9.392000    2.534000    7.821000   13.855000    7.692000
max      7.143000   16.049000    5.172000   15.714000   21.711000   17.172000

               X12         X13         X14         X15         X16         X17  \
count  873.000000  873.000000  873.000000  873.000000  873.000000  873.000000
mean     1.961329    3.937433    4.093777    5.880874    5.484307    1.219159
```

std	1.119701	1.824546	1.655331	2.088377	1.908418	0.947344
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	1.105000	2.646000	3.053000	4.360000	4.230000	0.565000
50%	1.705000	3.906000	3.968000	5.842000	5.376000	1.091000
75%	2.532000	5.145000	5.150000	7.172000	6.765000	1.709000
max	5.842000	9.794000	9.290000	12.366000	11.650000	4.255000

	X18	X19
count	873.000000	873.000000
mean	3.335706	8.930852
std	1.545700	3.531441
min	0.000000	0.806000
25%	2.299000	6.087000
50%	3.170000	8.527000
75%	4.183000	11.538000
max	8.387000	19.318000

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

	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9	\
0	10.784	5.882	1.961	6.863	1.961	10.784	2.941	5.882	2.941	7.843	
1	9.068	4.282	3.778	7.557	0.252	7.809	3.526	7.557	1.511	5.038	
4	6.452	8.602	3.226	4.301	0.000	4.301	0.000	9.677	1.075	2.151	
5	11.146	4.954	4.644	7.121	1.548	5.882	2.786	6.192	1.238	7.121	
7	11.765	9.050	3.620	7.240	1.357	9.050	2.262	6.335	2.262	0.905	
..	…	…	…	…	…	…	…	…	…	…	
994	5.556	5.556	0.000	9.259	1.852	14.815	3.704	9.259	0.000	3.704	
995	10.559	8.075	0.621	16.149	0.621	3.106	0.000	8.075	1.242	1.863	
996	4.000	4.667	6.000	4.667	2.667	7.333	6.000	3.000	2.333	11.333	
997	7.285	7.781	3.808	6.291	1.159	8.940	2.318	8.609	2.980	5.629	
998	12.500	5.921	1.974	8.333	0.439	7.675	2.412	8.114	2.632	4.167	

	…	X11	X12	X13	X14	X15	X16	X17	X18	X19	\
0	…	2.941	0.980	3.922	6.863	3.922	4.902	0.980	3.922	5.882	
1	…	3.275	1.008	4.786	6.801	9.320	6.801	1.259	3.526	6.045	
4	…	1.075	2.151	6.452	8.602	8.602	4.301	2.151	3.226	9.677	
5	…	4.025	2.786	3.406	3.715	12.074	3.715	0.619	1.548	8.050	
7	…	0.452	2.262	4.072	7.240	4.525	5.430	0.905	3.620	8.145	
..	…	…	…	…	…	…	…	…	…		
994	…	3.704	1.852	3.704	5.556	1.852	5.556	1.852	7.407	5.556	
995	…	0.621	0.621	1.863	5.590	10.559	11.180	1.242	3.106	9.317	
996	…	6.000	2.000	5.667	4.333	9.000	4.667	1.667	4.333	2.667	
997	…	4.801	3.642	3.311	4.801	5.464	5.132	0.000	2.483	8.113	

```
998  …  3.728  1.535  3.289  4.605   5.044   6.360  1.316  5.263  8.114

              X20
0     no_efectores
1     no_efectores
4     no_efectores
5     no_efectores
7     no_efectores
..             …
994   no_efectores
995   no_efectores
996   no_efectores
997   no_efectores
998   no_efectores

[816 rows x 21 columns]
```

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

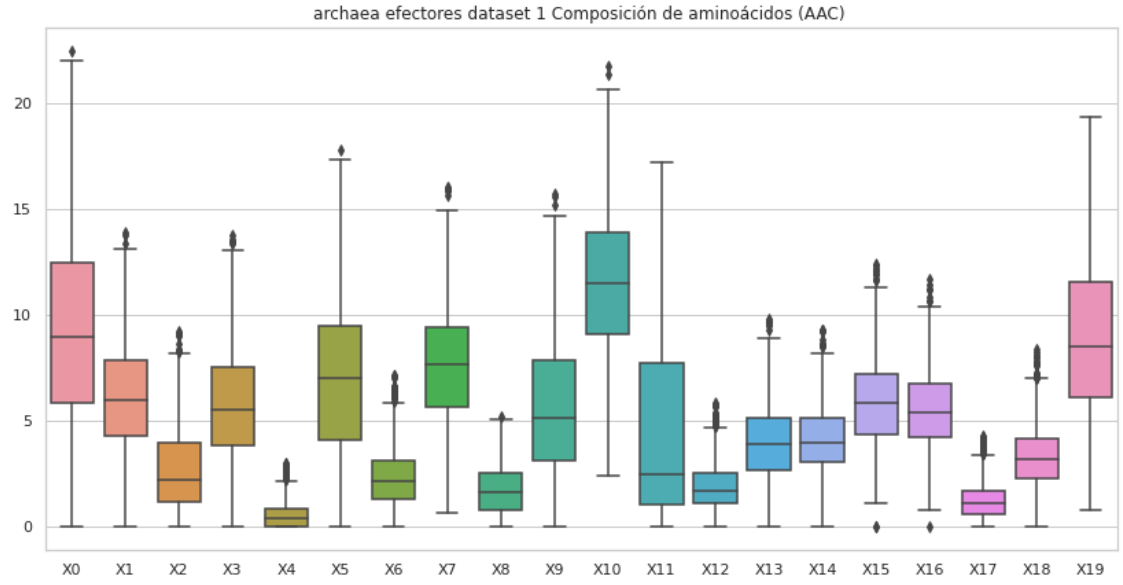
	X0	X1	X2	X3	X4	X5	\
count	816.000000	816.000000	816.000000	816.000000	816.000000	816.000000	
mean	9.868194	6.378243	2.710842	7.935316	0.893347	8.291961	
std	3.582069	2.448549	1.726575	3.205652	0.956791	3.381524	
min	1.987000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	7.286500	4.626250	1.470000	5.817250	0.154500	6.099500	
50%	9.767500	6.266500	2.434000	8.193500	0.670500	8.382000	
75%	12.272750	7.904250	3.448000	10.134000	1.283750	10.484000	
max	19.872000	14.444000	9.205000	18.519000	5.797000	19.207000	

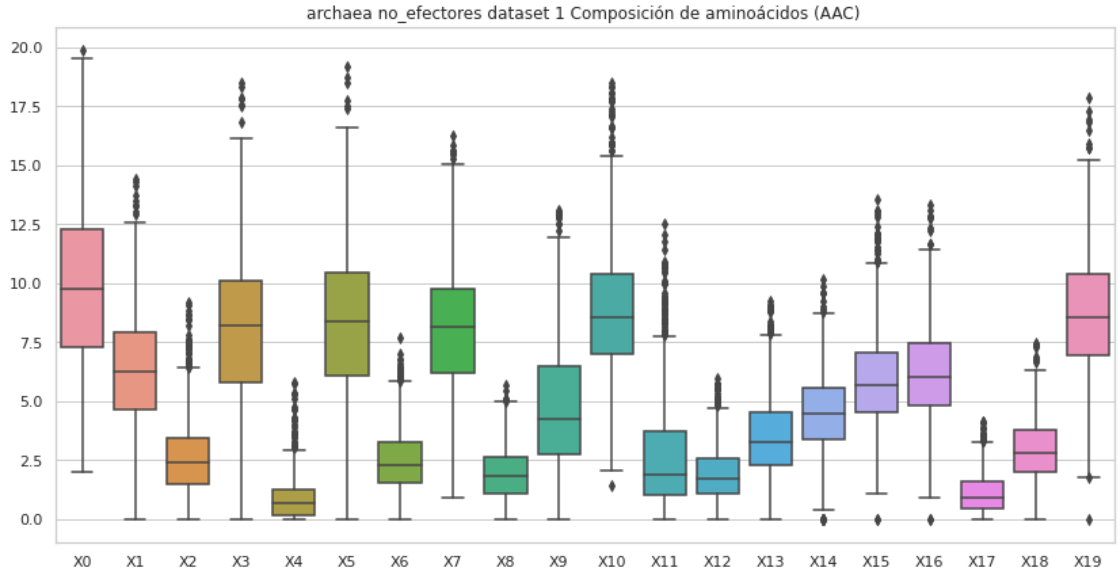
	X6	X7	X8	X9	X10	X11	\
count	816.000000	816.000000	816.000000	816.000000	816.00000	816.000000	
mean	2.472730	8.106701	1.906054	4.845086	8.99698	2.747523	
std	1.403086	2.574482	1.104546	2.767248	2.94342	2.452998	
min	0.000000	0.935000	0.000000	0.000000	1.42900	0.000000	
25%	1.538000	6.207750	1.110000	2.778000	7.01050	1.030750	
50%	2.306000	8.163500	1.847000	4.278000	8.58200	1.923000	
75%	3.265000	9.751000	2.664000	6.484750	10.40950	3.732500	
max	7.692000	16.250000	5.714000	13.125000	18.51900	12.500000	

	X12	X13	X14	X15	X16	X17	\
count	816.000000	816.000000	816.000000	816.000000	816.000000	816.000000	
mean	1.965023	3.528882	4.530237	5.929722	6.156359	1.091831	
std	1.118373	1.751069	1.655060	2.121792	2.049438	0.891156	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	1.116500	2.307250	3.408000	4.530750	4.809250	0.432000	

50%	1.740500	3.282000	4.497500	5.665500	6.036500	0.917000
75%	2.594000	4.521000	5.556000	7.067250	7.463000	1.581000
max	6.000000	9.237000	10.145000	13.568000	13.333000	4.124000

	X18	X19
count	816.000000	816.000000
mean	2.916382	8.728580
std	1.351443	2.666602
min	0.000000	0.000000
25%	2.000000	6.942000
50%	2.826500	8.537500
75%	3.794750	10.398250
max	7.447000	17.857000





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

          X0        X1        X2        X3        X4        X5        X6  \
0    0.032772  0.000000  0.032772  0.017647  0.017647  0.047898  0.010084
1    0.021614  0.003002  0.010207  0.012008  0.007205  0.024016  0.001201
2    0.104789  0.023286  0.069859  0.069859  0.023286  0.046573  0.000000
3    0.020152  0.000000  0.010076  0.020152  0.030228  0.100760  0.020152
4    0.027380  0.003911  0.019557  0.035203  0.000000  0.019557  0.003911
..        …         …         …         …         …         …         …
995  0.055761  0.011438  0.055761  0.062910  0.034314  0.052901  0.015727
996  0.015041  0.000000  0.004387  0.003760  0.007520  0.015041  0.001253
997  0.034099  0.000000  0.007577  0.003368  0.005894  0.015576  0.002526
998  0.052356  0.003740  0.052356  0.078534  0.052356  0.048616  0.018699
999  0.012599  0.000969  0.002908  0.000969  0.012599  0.022291  0.008723

          X7        X8        X9  …       X74       X75       X76  \
0    0.075628  0.052940  0.022688  …  0.011747  0.021181  0.008723
1    0.006004  0.001801  0.013209  …  0.009105  0.004780  0.038248
2    0.034930  0.034930  0.104789  …  0.099243 -0.007861 -0.032309
3    0.060456  0.120913  0.030228  …  0.011412  0.005113  0.129013
4    0.019557  0.046938  0.023469  … -0.017982 -0.013372  0.009306
..        …         …         …  …         …         …         …
995  0.055761  0.037174  0.062910  …  0.000391  0.030852  0.000971
996  0.007520  0.001880  0.026321  …  0.019045 -0.000162  0.024976
997  0.001263  0.000000  0.021469  …  0.011208  0.003835  0.023135
998  0.071054  0.080404  0.095363  … -0.013045  0.007527  0.014387
999  0.014538  0.003877  0.031014  …  0.014927  0.003325  0.021765

          X77       X78       X79       X80       X81       X82         X83
0    0.020828  0.030253  0.014110 -0.002039 -0.010414  0.002373  efectores
1    0.002823  0.004991  0.021308  0.009210  0.007710  0.031077  efectores
2   -0.078690 -0.004337  0.006774 -0.097897 -0.092551 -0.040824  efectores
3   -0.099279  0.003247 -0.022474  0.110382 -0.023527  0.014274  efectores
4    0.007419  0.007882  0.020850 -0.011498  0.013418  0.029676  efectores
..        …         …         …         …         …         …         …
995  0.038766  0.044851  0.007040 -0.008284  0.002510  0.007614  efectores
996  0.030490  0.005799  0.016470  0.012366 -0.002724  0.008503  efectores
997  0.002639 -0.004194  0.030881  0.009257  0.002160  0.024737  efectores
```

```
998  0.040514  0.041490  0.003398  0.018081  0.008794 -0.014112  efectores
999  0.010076  0.004183  0.006278  0.007527  0.000245  0.006605  efectores

[1000 rows x 84 columns]
```

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

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.032081	0.003576	0.026987	0.035025	0.016646	
std	0.016578	0.006117	0.021864	0.029977	0.014457	
min	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.021196	0.000000	0.009453	0.008894	0.007244	
50%	0.028653	0.000987	0.020834	0.027645	0.013142	
75%	0.038889	0.004350	0.038820	0.054871	0.021585	
max	0.175742	0.044344	0.152967	0.233103	0.160472	

	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.028136	0.007881	0.029135	0.027235	0.047290	…	
std	0.013898	0.008680	0.030000	0.033815	0.031005	…	
min	0.002038	0.000000	0.000000	0.000000	0.005002	…	
25%	0.017526	0.002066	0.007358	0.002379	0.025067	…	
50%	0.025041	0.005829	0.018469	0.010205	0.040368	…	
75%	0.035514	0.010592	0.041780	0.043511	0.059397	…	
max	0.108002	0.130968	0.247795	0.277342	0.227817	…	

	X73	X74	X75	X76	X77	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.012234	0.005367	0.009941	0.013493	0.006109	
std	0.021752	0.033867	0.027435	0.019667	0.032692	
min	-0.194333	-0.409445	-0.137772	-0.145071	-0.208076	
25%	0.002836	-0.006761	-0.001250	0.003982	-0.004605	
50%	0.013911	0.008810	0.004958	0.015985	0.009189	
75%	0.023368	0.017225	0.017036	0.024345	0.018812	
max	0.151521	0.187654	0.233050	0.152266	0.326451	

	X78	X79	X80	X81	X82
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.008409	0.014223	0.005553	0.008953	0.013108
std	0.027768	0.020819	0.032121	0.027840	0.019816
min	-0.145799	-0.113129	-0.229497	-0.176735	-0.113692
25%	-0.001411	0.004430	-0.004045	-0.001162	0.004551
50%	0.005474	0.014310	0.009708	0.005103	0.014378
75%	0.016578	0.025149	0.018489	0.017263	0.023418

```
max       0.256908     0.172295     0.192947     0.207187     0.114204

[8 rows x 83 columns]
```

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

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.028104  0.005110  0.017884  0.028104  0.010220  0.015330  0.007665
1    0.038942  0.001082  0.032452  0.033534  0.020553  0.032452  0.006490
2    0.017683  0.000000  0.003929  0.001965  0.007859  0.015718  0.000000
3    0.033227  0.027689  0.049841  0.038765  0.019383  0.022151  0.013845
4    0.013200  0.000000  0.008800  0.008800  0.013200  0.019801  0.002200
..        …         …         …         …         …         …         …
995  0.013880  0.000816  0.021227  0.004082  0.002449  0.010614  0.001633
996  0.032401  0.021601  0.037801  0.059402  0.045902  0.024301  0.018901
997  0.033968  0.005404  0.029336  0.041688  0.015440  0.040144  0.013896
998  0.043975  0.001543  0.029317  0.027002  0.011572  0.028545  0.009258
999  0.013378  0.008027  0.013378  0.010702  0.005351  0.021405  0.002676

           X7        X8        X9  …        X74       X75       X76  \
0    0.020439  0.007665  0.020439  …   0.026663  0.045387  0.009157
1    0.021635  0.014062  0.029207  …   0.003907 -0.005249  0.026993
2    0.011789  0.000000  0.039295  … -0.020667 -0.008760  0.014884
3    0.052610  0.094144  0.058148  … -0.021665 -0.015068 -0.014152
4    0.004400  0.002200  0.028601  …   0.015620  0.009919  0.001192
..        …         …         …  …         …         …         …
995  0.002449  0.000816  0.007348  …   0.013299  0.015171  0.013149
996  0.091803  0.048602  0.062102  …   0.019774 -0.000843 -0.023995
997  0.026248  0.022388  0.034740  … -0.007832  0.002170  0.022641
998  0.014658  0.013115  0.023145  …   0.001358 -0.000682  0.017330
999  0.008027  0.029431  0.010702  …   0.019979  0.016015  0.010323

           X77       X78       X79       X80       X81       X82           X83
0    0.016468  0.002887  0.018217  0.000025 -0.005388  0.003341  no_efectores
1    0.007216  0.007550  0.016494  0.014991  0.014013  0.032731  no_efectores
2    0.014141 -0.000981  0.023174  0.020626  0.011050  0.022555  no_efectores
3    0.021765  0.035514 -0.011465  0.012155  0.029990  0.019618  no_efectores
4    0.009641  0.005837  0.001300 -0.006469 -0.005668 -0.001778  no_efectores
..        …         …         …         …         …         …           …
995  0.007810  0.015912  0.010521  0.004932  0.016789  0.013834  no_efectores
996 -0.024125 -0.009501  0.009909  0.028394  0.032805 -0.023157  no_efectores
997 -0.011710  0.014256  0.013532  0.011836  0.021045  0.005348  no_efectores
998 -0.007416 -0.010333  0.020495  0.003752  0.001714  0.015302  no_efectores
```

```
999  0.015498  0.011627  0.027849 -0.008178  0.007241  0.008590  no_efectores

[1000 rows x 84 columns]
```

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

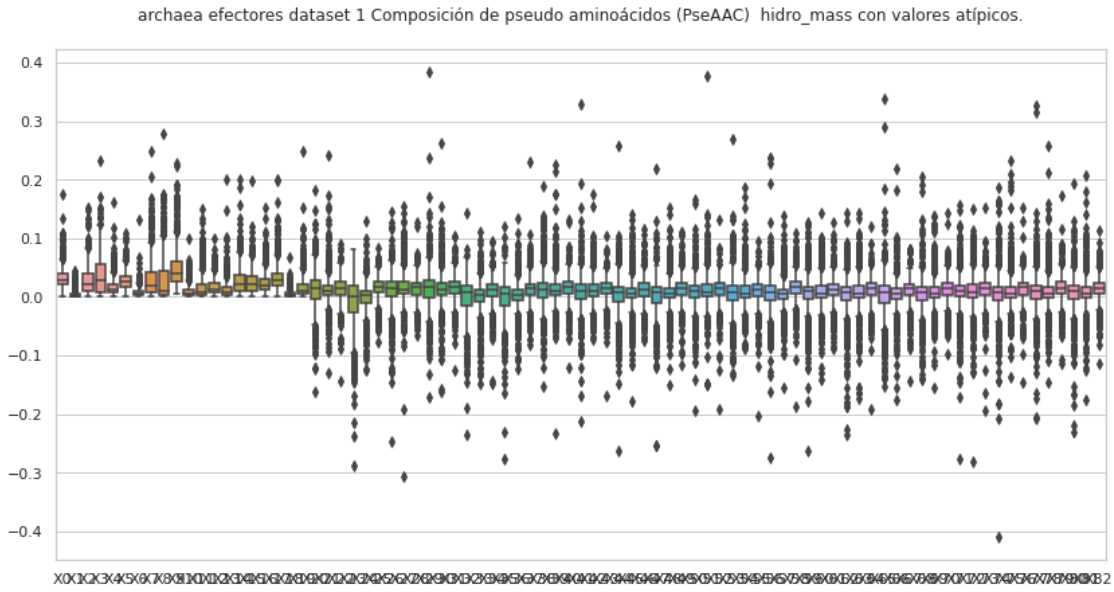
	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.033668	0.004833	0.031188	0.033452	0.014152	
std	0.020150	0.007975	0.020968	0.025496	0.015753	
min	-0.361457	0.000000	-0.120486	-0.240971	-0.240971	
25%	0.022895	0.000000	0.017022	0.017317	0.006659	
50%	0.032070	0.002231	0.029730	0.031063	0.010990	
75%	0.042461	0.005513	0.041262	0.046021	0.017482	
max	0.117033	0.070824	0.180714	0.266236	0.137200	

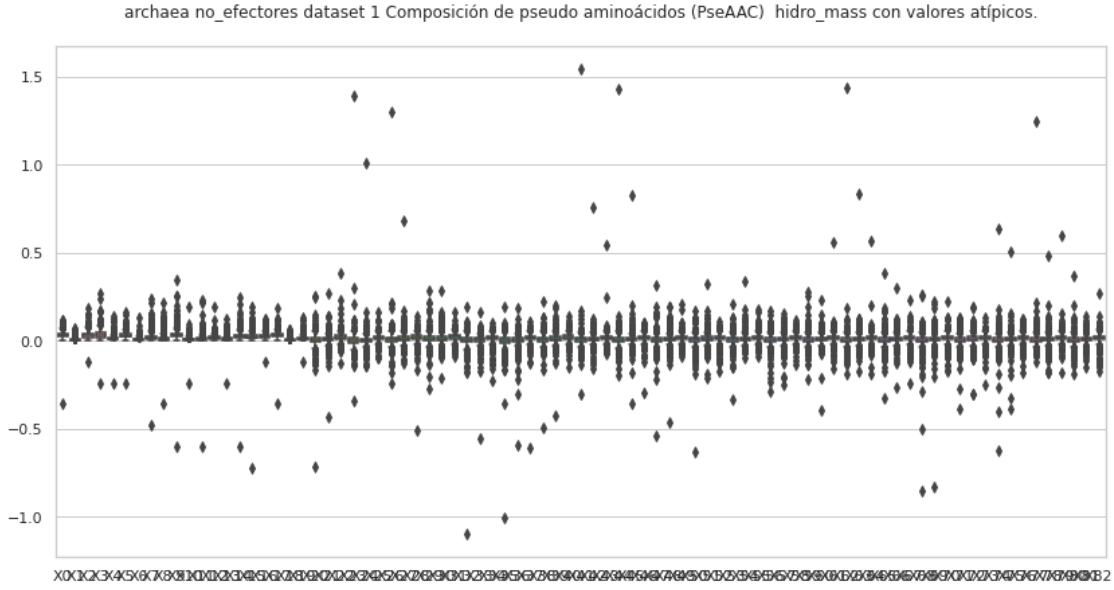
	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.028978	0.008322	0.020874	0.014632	0.034723	…	
std	0.016980	0.009060	0.027916	0.024700	0.032910	…	
min	-0.240971	0.000000	-0.481943	-0.361457	-0.602428	…	
25%	0.020133	0.002765	0.007924	0.002878	0.021074	…	
50%	0.027035	0.006232	0.014391	0.006922	0.029117	…	
75%	0.035580	0.011182	0.026115	0.016906	0.041291	…	
max	0.158467	0.133118	0.240100	0.216857	0.342304	…	

	X73	X74	X75	X76	X77	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.015398	0.001818	0.007691	0.015573	0.003002	
std	0.022767	0.039935	0.032044	0.021908	0.046112	
min	-0.247624	-0.624507	-0.391533	-0.179370	-0.120514	
25%	0.007463	-0.006799	-0.001034	0.008020	-0.008149	
50%	0.017912	0.003173	0.005899	0.017113	0.002696	
75%	0.026949	0.013127	0.017407	0.026198	0.012552	
max	0.130988	0.631698	0.503921	0.119424	1.246451	

	X78	X79	X80	X81	X82
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.008208	0.016371	0.002571	0.007694	0.015420
std	0.028951	0.027716	0.027386	0.022200	0.024841
min	-0.183963	-0.181473	-0.187262	-0.148181	-0.174774
25%	-0.001403	0.007205	-0.007545	-0.001768	0.006451
50%	0.006721	0.017298	0.003559	0.006125	0.018157
75%	0.016927	0.027191	0.013383	0.017628	0.027417
max	0.479002	0.596600	0.370523	0.135010	0.270708

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.032772	0.000000	0.032772	0.017647	0.017647	0.047898	0.010084	
1	0.021614	0.003002	0.010207	0.012008	0.007205	0.024016	0.001201	
4	0.027380	0.003911	0.019557	0.035203	0.000000	0.019557	0.003911	
7	0.038315	0.004789	0.028736	0.047894	0.038315	0.031131	0.004789	
8	0.028822	0.000848	0.003391	0.000000	0.005934	0.017802	0.004239	
..	…	…	…	…	…	…	…	
994	0.031288	0.001738	0.017382	0.008691	0.015644	0.029550	0.008691	
995	0.055761	0.011438	0.055761	0.062910	0.034314	0.052901	0.015727	
996	0.015041	0.000000	0.004387	0.003760	0.007520	0.015041	0.001253	
997	0.034099	0.000000	0.007577	0.003368	0.005894	0.015576	0.002526	
999	0.012599	0.000969	0.002908	0.000969	0.012599	0.022291	0.008723	

	X7	X8	X9	…	X74	X75	X76	\
0	0.075628	0.052940	0.022688	…	0.011747	0.021181	0.008723	
1	0.006004	0.001801	0.013209	…	0.009105	0.004780	0.038248	
4	0.019557	0.046938	0.023469	…	-0.017982	-0.013372	0.009306	
7	0.083814	0.071841	0.052683	…	0.028777	0.039234	0.033994	
8	0.003391	0.002543	0.020345	…	0.011262	0.002845	0.031031	
..	…	…	…	…	…	…	…	
994	0.012167	0.003476	0.041717	…	0.010363	-0.012625	0.011439	
995	0.055761	0.037174	0.062910	…	0.000391	0.030852	0.000971	
996	0.007520	0.001880	0.026321	…	0.019045	-0.000162	0.024976	
997	0.001263	0.000000	0.021469	…	0.011208	0.003835	0.023135	
999	0.014538	0.003877	0.031014	…	0.014927	0.003325	0.021765	

	X77	X78	X79	X80	X81	X82	X83
0	0.020828	0.030253	0.014110	-0.002039	-0.010414	0.002373	efectores
1	0.002823	0.004991	0.021308	0.009210	0.007710	0.031077	efectores
4	0.007419	0.007882	0.020850	-0.011498	0.013418	0.029676	efectores
7	0.009309	0.014724	0.007524	-0.002395	-0.021141	-0.003321	efectores
8	0.009997	-0.000537	0.016324	0.013940	0.002711	0.015238	efectores
..	…	…	…	…	…	…	
994	0.016550	-0.007303	0.011034	0.027649	-0.003812	0.019843	efectores
995	0.038766	0.044851	0.007040	-0.008284	0.002510	0.007614	efectores
996	0.030490	0.005799	0.016470	0.012366	-0.002724	0.008503	efectores
997	0.002639	-0.004194	0.030881	0.009257	0.002160	0.024737	efectores
999	0.010076	0.004183	0.006278	0.007527	0.000245	0.006605	efectores

[795 rows x 84 columns]

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

	X0	X1	X2	X3	X4	X5	\
count	795.000000	795.000000	795.000000	795.000000	795.000000	795.000000	
mean	0.028926	0.002241	0.020979	0.025350	0.013561	0.024625	
std	0.011718	0.003587	0.015572	0.021036	0.009601	0.010070	
min	0.002648	0.000000	0.000000	0.000000	0.000000	0.002038	
25%	0.020509	0.000000	0.008296	0.007014	0.006458	0.016586	
50%	0.027185	0.000600	0.016407	0.018846	0.011107	0.022918	
75%	0.035279	0.003007	0.030993	0.041646	0.018521	0.030658	
max	0.073632	0.021405	0.071799	0.088428	0.056213	0.058826	

	X6	X7	X8	X9	…	X73	\
count	795.000000	795.000000	795.000000	795.000000	…	795.000000	
mean	0.006166	0.020365	0.017341	0.037566	…	0.014798	
std	0.005342	0.018982	0.022353	0.018388	…	0.014922	
min	0.000000	0.000000	0.000000	0.005683	…	-0.044821	
25%	0.001837	0.006327	0.001849	0.022873	…	0.005933	
50%	0.004992	0.012179	0.005866	0.034165	…	0.015750	
75%	0.008984	0.029980	0.027774	0.049435	…	0.024011	
max	0.031750	0.118240	0.099842	0.120809	…	0.066453	

	X74	X75	X76	X77	X78	X79	\
count	795.000000	795.000000	795.000000	795.000000	795.000000	795.000000	
mean	0.006463	0.007753	0.015766	0.007425	0.007758	0.015340	
std	0.017019	0.016120	0.014249	0.018867	0.015988	0.014208	
min	-0.093677	-0.050322	-0.037411	-0.084382	-0.056642	-0.046576	
25%	-0.003761	-0.000901	0.007321	-0.000805	-0.000006	0.006398	
50%	0.009331	0.004574	0.017587	0.009990	0.005360	0.015079	
75%	0.016286	0.013432	0.024484	0.018303	0.013719	0.024309	
max	0.084842	0.082329	0.059603	0.079192	0.083745	0.074947	

	X80	X81	X82
count	795.000000	795.000000	795.000000
mean	0.008612	0.008033	0.015234
std	0.018541	0.016717	0.013914
min	-0.089944	-0.057029	-0.044129
25%	-0.000605	-0.000474	0.006857
50%	0.010443	0.004641	0.015650
75%	0.017917	0.013844	0.023498
max	0.066696	0.087614	0.068094

[8 rows x 83 columns]

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.028104  0.005110  0.017884  0.028104  0.010220  0.015330  0.007665
1    0.038942  0.001082  0.032452  0.033534  0.020553  0.032452  0.006490
2    0.017683  0.000000  0.003929  0.001965  0.007859  0.015718  0.000000
4    0.013200  0.000000  0.008800  0.008800  0.013200  0.019801  0.002200
5    0.037661  0.005231  0.024061  0.019877  0.011508  0.020923  0.004185
..        …         …         …         …         …         …         …
995  0.013880  0.000816  0.021227  0.004082  0.002449  0.010614  0.001633
996  0.032401  0.021601  0.037801  0.059402  0.045902  0.024301  0.018901
997  0.033968  0.005404  0.029336  0.041688  0.015440  0.040144  0.013896
998  0.043975  0.001543  0.029317  0.027002  0.011572  0.028545  0.009258
999  0.013378  0.008027  0.013378  0.010702  0.005351  0.021405  0.002676

           X7        X8        X9  …       X74       X75       X76  \
0    0.020439  0.007665  0.020439  …  0.026663  0.045387  0.009157
1    0.021635  0.014062  0.029207  …  0.003907 -0.005249  0.026993
2    0.011789  0.000000  0.039295  … -0.020667 -0.008760  0.014884
4    0.004400  0.002200  0.028601  …  0.015620  0.009919  0.001192
5    0.024061  0.013600  0.025107  … -0.002956  0.003679  0.018322
..        …         …         …  …         …         …         …
995  0.002449  0.000816  0.007348  …  0.013299  0.015171  0.013149
996  0.091803  0.048602  0.062102  …  0.019774 -0.000843 -0.023995
997  0.026248  0.022388  0.034740  … -0.007832  0.002170  0.022641
998  0.014658  0.013115  0.023145  …  0.001358 -0.000682  0.017330
999  0.008027  0.029431  0.010702  …  0.019979  0.016015  0.010323

           X77       X78       X79       X80       X81       X82           X83
0    0.016468  0.002887  0.018217  0.000025 -0.005388  0.003341  no_efectores
1    0.007216  0.007550  0.016494  0.014991  0.014013  0.032731  no_efectores
2    0.014141 -0.000981  0.023174  0.020626  0.011050  0.022555  no_efectores
4    0.009641  0.005837  0.001300 -0.006469 -0.005668 -0.001778  no_efectores
5    0.003446  0.002223  0.028377  0.007388  0.008279  0.020602  no_efectores
..        …         …         …         …         …         …           …
995  0.007810  0.015912  0.010521  0.004932  0.016789  0.013834  no_efectores
996 -0.024125 -0.009501  0.009909  0.028394  0.032805 -0.023157  no_efectores
997 -0.011710  0.014256  0.013532  0.011836  0.021045  0.005348  no_efectores
998 -0.007416 -0.010333  0.020495  0.003752  0.001714  0.015302  no_efectores
999  0.015498  0.011627  0.027849 -0.008178  0.007241  0.008590  no_efectores

[874 rows x 84 columns]
```

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

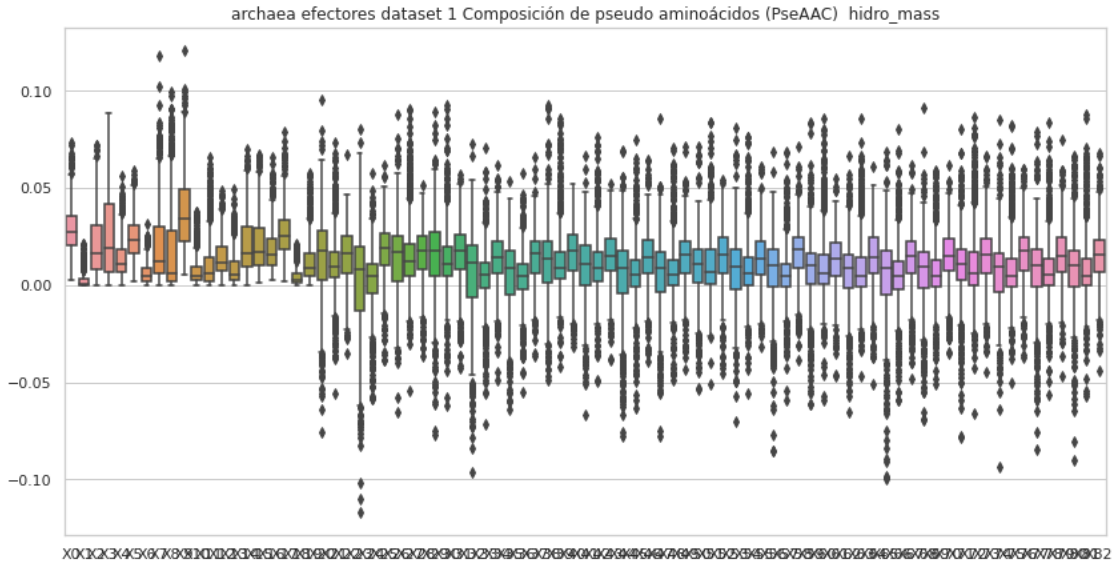
	X0	X1	X2	X3	X4	X5	\
count	874.000000	874.000000	874.000000	874.000000	874.000000	874.000000	
mean	0.032555	0.003542	0.028172	0.029895	0.012119	0.027009	
std	0.013196	0.004584	0.016093	0.017903	0.008362	0.010292	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.002753	
25%	0.022702	0.000000	0.015671	0.016278	0.006436	0.019725	
50%	0.031350	0.002060	0.028733	0.029148	0.010329	0.025977	
75%	0.041203	0.004707	0.037971	0.041880	0.015450	0.033218	
max	0.082489	0.028306	0.089974	0.092099	0.059116	0.067560	

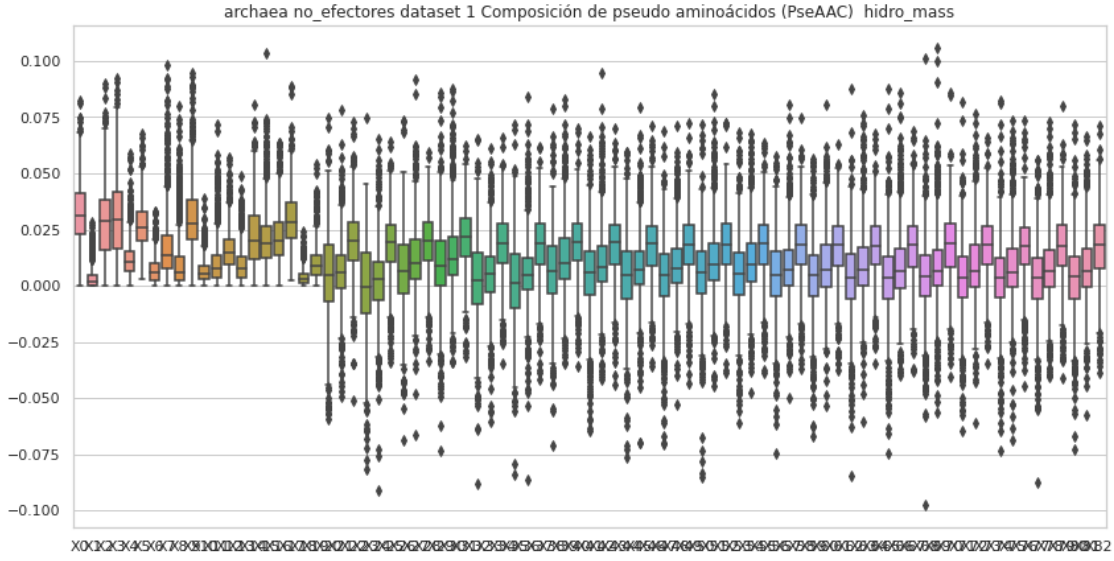
	X6	X7	X8	X9	…	X73	\
count	874.000000	874.000000	874.000000	874.000000	…	874.000000	
mean	0.006999	0.017652	0.010642	0.030596	…	0.017742	
std	0.005811	0.015115	0.012925	0.014833	…	0.013809	
min	0.000000	0.000000	0.000000	0.000000	…	-0.034874	
25%	0.002635	0.007629	0.002568	0.020337	…	0.009384	
50%	0.005754	0.013314	0.005876	0.027518	…	0.018243	
75%	0.010042	0.022279	0.012876	0.038050	…	0.026741	
max	0.033148	0.098329	0.079959	0.094588	…	0.066046	

	X74	X75	X76	X77	X78	X79	\
count	874.000000	874.000000	874.000000	874.000000	874.000000	874.000000	
mean	0.003084	0.008327	0.016999	0.002687	0.008028	0.017145	
std	0.016677	0.015761	0.014470	0.016152	0.015240	0.014981	
min	-0.073423	-0.068686	-0.044575	-0.087355	-0.040269	-0.041247	
25%	-0.005418	-0.000439	0.009456	-0.005994	-0.000376	0.008737	
50%	0.003479	0.005751	0.017552	0.003355	0.006580	0.017851	
75%	0.012563	0.016204	0.026092	0.012252	0.016068	0.027080	
max	0.082510	0.073688	0.073266	0.056065	0.072738	0.079797	

	X80	X81	X82
count	874.000000	874.000000	874.000000
mean	0.003366	0.008046	0.017105
std	0.016228	0.015622	0.014899
min	-0.072645	-0.057506	-0.039066
25%	-0.005662	-0.000794	0.007934
50%	0.004005	0.006236	0.018339
75%	0.012871	0.016769	0.027139
max	0.071899	0.066238	0.071171

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.036977  0.000000  0.036977  0.019911  0.019911  0.054043  0.011378
1    0.030081  0.004178  0.014205  0.016712  0.010027  0.033423  0.001671
2    0.093247  0.020722  0.062165  0.062165  0.020722  0.041443  0.000000
3    0.015230  0.000000  0.007615  0.015230  0.022844  0.076148  0.015230
4    0.036850  0.005264  0.026322  0.047379  0.000000  0.026322  0.005264
..        …         …         …         …         …         …         …
995  0.066126  0.013564  0.066126  0.074603  0.040693  0.062734  0.018651
996  0.035063  0.000000  0.010227  0.008766  0.017532  0.035063  0.002922
997  0.049022  0.000000  0.010894  0.004842  0.008473  0.022393  0.003631
998  0.054294  0.003878  0.054294  0.081441  0.054294  0.050416  0.019391
999  0.026573  0.002044  0.006132  0.002044  0.026573  0.047015  0.018397

           X7        X8        X9  …       X32       X33       X34  \
0    0.085332  0.059732  0.025600  …  0.014813  0.030359  0.009721
```

```
1    0.008356  0.002507  0.018383  …  0.029747  0.020855  0.021246
2    0.031082  0.031082  0.093247  … -0.028490  0.084934  0.027700
3    0.045689  0.091377  0.022844  … -0.032380 -0.010731  0.069062
4    0.026322  0.063172  0.031586  …  0.034800  0.029857  0.014262
..        …         …        …  …        …         …
995  0.066126  0.044084  0.074603  …  0.025599  0.022066 -0.060542
996  0.017532  0.004383  0.061360  …  0.047989  0.009113  0.012772
997  0.001816  0.000000  0.030866  …  0.037166  0.032048  0.037463
998  0.073685  0.083380  0.098893  …  0.019113  0.005518 -0.005255
999  0.030662  0.008176  0.065412  …  0.026800  0.021075  0.036953

          X35       X36       X37       X38       X39       X40        X41
0    0.033186  0.007943  0.030143  0.009842  0.015920  0.002678  efectores
1    0.029853  0.033673  0.044445  0.053231  0.029655  0.043250  efectores
2   -0.026470 -0.048724 -0.030951 -0.028750  0.006028 -0.036328  efectores
3    0.022570 -0.002633  0.017448  0.097499 -0.016985  0.010787  efectores
4   -0.010295  0.037285  0.029761  0.012525  0.028061  0.039940  efectores
..        …         …         …         …         …         …
995 -0.006866 -0.016807 -0.014938  0.001151  0.008349  0.009029  efectores
996  0.025555  0.040035  0.035306  0.058224  0.038397  0.019823  efectores
997  0.034100  0.032210  0.048631  0.033260  0.044397  0.035564  efectores
998 -0.015655  0.011717  0.029576  0.014920  0.003524 -0.014634  efectores
999  0.016145  0.045011  0.024879  0.045905  0.013242  0.013930  efectores

[1000 rows x 42 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.045564     0.004620     0.036287     0.048095     0.023774
std       0.017866     0.007372     0.024733     0.038959     0.017642
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       0.034436     0.000000     0.017155     0.016140     0.011572
50%       0.042981     0.001497     0.028980     0.037677     0.019511
75%       0.054171     0.006464     0.050094     0.072576     0.031241
max       0.217726     0.053715     0.144814     0.242099     0.132900

                X5           X6           X7           X8           X9  …  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
mean      0.039571     0.010908     0.039455     0.036083     0.066972  …
std       0.013397     0.010534     0.033335     0.041164     0.032389  …
min       0.004607     0.000000     0.000000     0.000000     0.006038  …
25%       0.030031     0.003366     0.011992     0.003928     0.043115  …
50%       0.038120     0.008773     0.028492     0.016053     0.062407  …
```

```
75%       0.047455     0.015434     0.059615     0.062986     0.087198  …
max       0.108667     0.142828     0.172928     0.268686     0.178648  …

               X31          X32          X33          X34          X35  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.014778     0.021833     0.015973     0.017560     0.016535
std       0.026138     0.026819     0.025574     0.027789     0.027322
min      -0.211336    -0.194582    -0.113186    -0.175856    -0.150051
25%       0.002218     0.010196     0.003520     0.005125     0.004162
50%       0.019367     0.025911     0.021079     0.022522     0.021968
75%       0.032441     0.038499     0.032821     0.035296     0.034532
max       0.103732     0.143660     0.127980     0.137430     0.160511

               X36          X37          X38          X39          X40
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.019078     0.017743     0.019938     0.020748     0.019175
std       0.025982     0.028186     0.025362     0.026081     0.026601
min      -0.103919    -0.288681    -0.150669    -0.102329    -0.144459
25%       0.007088     0.004436     0.006249     0.007150     0.007420
50%       0.023426     0.021483     0.025033     0.022953     0.022614
75%       0.035241     0.034762     0.036648     0.036172     0.035083
max       0.145503     0.166333     0.097499     0.126084     0.104247

[8 rows x 41 columns]
```

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

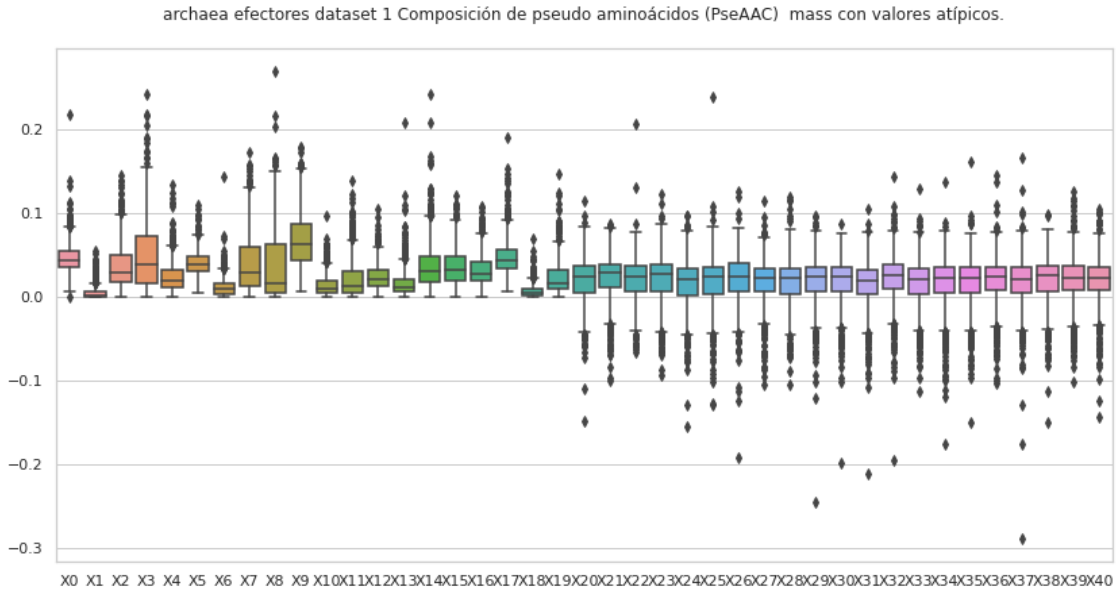
```
          X0        X1        X2        X3        X4        X5        X6  \
0    0.066507  0.012092  0.042323  0.066507  0.024185  0.036277  0.018138
1    0.041499  0.001153  0.034583  0.035736  0.021902  0.034583  0.006917
2    0.030107  0.000000  0.006690  0.003345  0.013381  0.026761  0.000000
3    0.045627  0.038023  0.068441  0.053232  0.026616  0.030418  0.019011
4    0.035210  0.000000  0.023473  0.023473  0.035210  0.052815  0.005868
..        …         …         …         …         …         …         …
995  0.032833  0.001931  0.050215  0.009657  0.005794  0.025108  0.003863
996  0.044099  0.029399  0.051449  0.080848  0.062474  0.033074  0.025725
997  0.045362  0.007217  0.039176  0.055672  0.020619  0.053610  0.018557
998  0.050445  0.001770  0.033630  0.030975  0.013275  0.032745  0.010620
999  0.022204  0.013323  0.022204  0.017764  0.008882  0.035527  0.004441

          X7        X8        X9  …       X32       X33       X34  \
0    0.048369  0.018138  0.048369  …  0.019247  0.051078 -0.002224
1    0.023055  0.014986  0.031125  …  0.002026  0.036495  0.024051
```

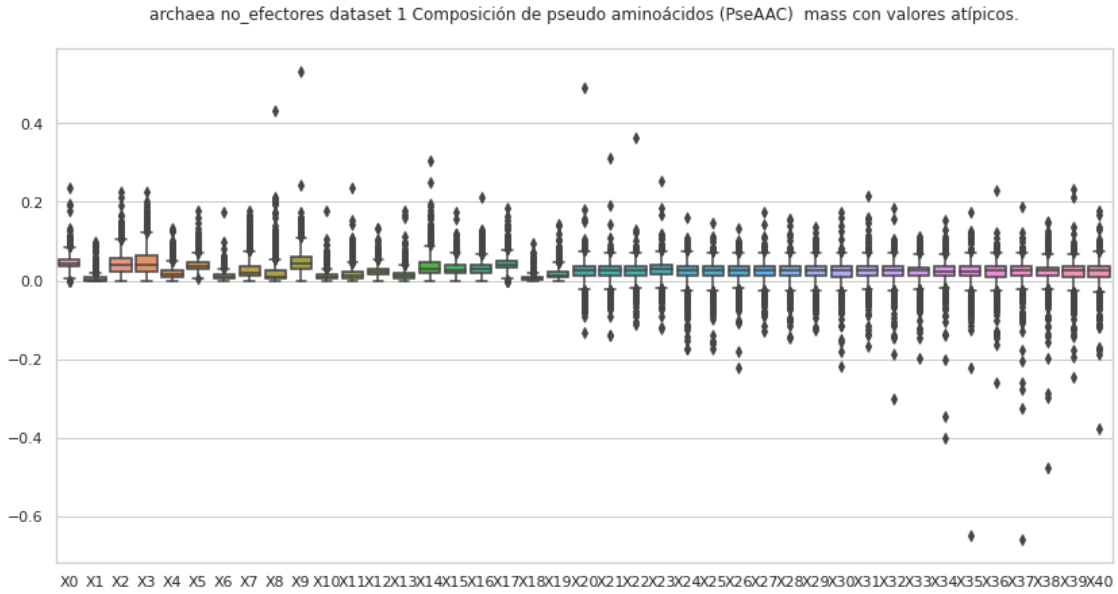
```
2    0.020071  0.000000  0.066904  …  0.030613  0.038563  0.053622
3    0.072243  0.129277  0.079848  …  0.005714  0.016883  0.019003
4    0.011737  0.005868  0.076288  …  0.013848  0.037579  0.044759
..        …         …        …  …        …         …         …
995  0.005794  0.001931  0.017382  …  0.034599  0.006217  0.041593
996  0.124948  0.066149  0.084523  … -0.028949  0.015053 -0.047511
997  0.035052  0.029898  0.046393  …  0.018014  0.017184  0.008578
998  0.016815  0.015045  0.026550  …  0.023483  0.032529  0.030655
999  0.013323  0.048850  0.017764  …  0.015907  0.014433  0.013038

          X35       X36       X37       X38       X39       X40           X41
0   -0.018876  0.063533  0.044704  0.021669  0.043110  0.007907  no_efectores
1    0.034519  0.016114  0.012812  0.028765  0.017577  0.034880  no_efectores
2    0.011022  0.029972  0.015314  0.025341  0.039456  0.038403  no_efectores
3   -0.020638 -0.038044  0.031977 -0.019433 -0.015744  0.026939  no_efectores
4   -0.008732  0.006148  0.030283  0.003180  0.003468 -0.004742  no_efectores
..        …         …         …         …         …         …           …
995  0.036060  0.019967  0.042635  0.031104  0.024889  0.032725  no_efectores
996  0.000072 -0.010082 -0.015898 -0.032658  0.013486 -0.031518  no_efectores
997  0.026880  0.023928  0.024378  0.030235  0.018071  0.007142  no_efectores
998  0.023145  0.025311  0.035413  0.019880  0.023510  0.017553  no_efectores
999  0.001042  0.049954  0.029234  0.017134  0.046223  0.014257  no_efectores

[1000 rows x 42 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.046629     0.006451     0.042826     0.047417     0.019873
std       0.019981     0.011244     0.028129     0.034425     0.016267
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       0.034615     0.000000     0.023551     0.023381     0.009165
50%       0.043692     0.002945     0.038098     0.040182     0.015482
75%       0.054839     0.007588     0.055965     0.063465     0.025933
max       0.236464     0.099780     0.225615     0.226168     0.132891

                X5           X6           X7           X8           X9  …  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
mean      0.039793     0.011083     0.028932     0.021186     0.049117  …
std       0.016275     0.011666     0.026645     0.032025     0.031773  …
min       0.004074     0.000000     0.000000     0.000000     0.000000  …
25%       0.030075     0.004170     0.011470     0.004052     0.029335  …
50%       0.037407     0.008480     0.020589     0.009470     0.042076  …
75%       0.046161     0.014841     0.037293     0.024420     0.061975  …
```

```
max       0.177348     0.175114     0.177348     0.433645     0.532044  …

                X31          X32          X33          X34          X35  \
count   1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean       0.021908     0.020806     0.020443     0.020164     0.019949
std        0.027680     0.029448     0.026766     0.031293     0.035386
min       -0.167189    -0.301907    -0.199750    -0.402289    -0.649431
25%        0.011498     0.010750     0.010914     0.012839     0.010424
50%        0.026328     0.025105     0.025185     0.023601     0.023898
75%        0.035261     0.034850     0.034126     0.034562     0.035014
max        0.216195     0.184873     0.112797     0.152465     0.175391

                X36          X37          X38          X39          X40
count   1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean       0.019897     0.019276     0.019578     0.019889     0.019907
std        0.029998     0.038614     0.035204     0.030892     0.032696
min       -0.259804    -0.659791    -0.477928    -0.245659    -0.376494
25%        0.010143     0.011413     0.011452     0.010197     0.009744
50%        0.024578     0.025411     0.024122     0.024336     0.025372
75%        0.034500     0.035319     0.033816     0.034751     0.035317
max        0.230759     0.188726     0.148257     0.233598     0.177767

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.036977  0.000000  0.036977  0.019911  0.019911  0.054043  0.011378
1    0.030081  0.004178  0.014205  0.016712  0.010027  0.033423  0.001671
2    0.093247  0.020722  0.062165  0.062165  0.020722  0.041443  0.000000
4    0.036850  0.005264  0.026322  0.047379  0.000000  0.026322  0.005264
6    0.077303  0.006442  0.045094  0.103071  0.032210  0.032210  0.019326
..        …         …         …         …         …         …         …
995  0.066126  0.013564  0.066126  0.074603  0.040693  0.062734  0.018651
996  0.035063  0.000000  0.010227  0.008766  0.017532  0.035063  0.002922
997  0.049022  0.000000  0.010894  0.004842  0.008473  0.022393  0.003631
998  0.054294  0.003878  0.054294  0.081441  0.054294  0.050416  0.019391
999  0.026573  0.002044  0.006132  0.002044  0.026573  0.047015  0.018397

           X7        X8        X9  …       X32       X33       X34  \
0    0.085332  0.059732  0.025600  …  0.014813  0.030359  0.009721
1    0.008356  0.002507  0.018383  …  0.029747  0.020855  0.021246
2    0.031082  0.031082  0.093247  … -0.028490  0.084934  0.027700
4    0.026322  0.063172  0.031586  …  0.034800  0.029857  0.014262
6    0.083745  0.103071  0.083745  …  0.011386  0.021590 -0.034999
..        …         …         …  …         …         …         …
995  0.066126  0.044084  0.074603  …  0.025599  0.022066 -0.060542
996  0.017532  0.004383  0.061360  …  0.047989  0.009113  0.012772
997  0.001816  0.000000  0.030866  …  0.037166  0.032048  0.037463
998  0.073685  0.083380  0.098893  …  0.019113  0.005518 -0.005255
999  0.030662  0.008176  0.065412  …  0.026800  0.021075  0.036953
```

```
          X35       X36       X37       X38       X39       X40        X41
0    0.033186  0.007943  0.030143  0.009842  0.015920  0.002678  efectores
1    0.029853  0.033673  0.044445  0.053231  0.029655  0.043250  efectores
2   -0.026470 -0.048724 -0.030951 -0.028750  0.006028 -0.036328  efectores
4   -0.010295  0.037285  0.029761  0.012525  0.028061  0.039940  efectores
6    0.021973 -0.015529 -0.048461  0.011432  0.057769  0.005882  efectores
..        …         …         …         …         …         …
995 -0.006866 -0.016807 -0.014938  0.001151  0.008349  0.009029  efectores
996  0.025555  0.040035  0.035306  0.058224  0.038397  0.019823  efectores
997  0.034100  0.032210  0.048631  0.033260  0.044397  0.035564  efectores
998 -0.015655  0.011717  0.029576  0.014920  0.003524 -0.014634  efectores
999  0.016145  0.045011  0.024879  0.045905  0.013242  0.013930  efectores

[813 rows x 42 columns]
```

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

```
                X0          X1          X2          X3          X4          X5  \
count  813.000000  813.000000  813.000000  813.000000  813.000000  813.000000
mean     0.044301    0.003310    0.031178    0.037963    0.020822    0.037581
std      0.014053    0.004881    0.019811    0.029862    0.014429    0.011150
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.012294
25%      0.035034    0.000000    0.016018    0.013180    0.010426    0.029285
50%      0.042746    0.001072    0.025505    0.028552    0.017420    0.036635
75%      0.052056    0.004822    0.043495    0.056564    0.027693    0.044728
max      0.097417    0.026422    0.109944    0.144772    0.068470    0.073647

                X6          X7          X8          X9  …         X31  \
count  813.000000  813.000000  813.000000  813.000000  …  813.000000
mean     0.009282    0.032325    0.026806    0.059519  …    0.019308
std      0.007647    0.028041    0.032559    0.027225  …    0.020003
min      0.000000    0.000000    0.000000    0.006465  …   -0.060739
25%      0.003082    0.010233    0.003138    0.040027  …    0.007104
50%      0.007668    0.021509    0.009123    0.054996  …    0.022236
75%      0.013547    0.050412    0.042351    0.078957  …    0.033759
max      0.042351    0.137382    0.149871    0.151584  …    0.077201

               X32         X33         X34         X35         X36         X37  \
count  813.000000  813.000000  813.000000  813.000000  813.000000  813.000000
mean     0.024814    0.018926    0.022358    0.021035    0.022326    0.021724
std      0.020491    0.020902    0.020447    0.021650    0.019562    0.021086
min     -0.045458   -0.059348   -0.060595   -0.061086   -0.053477   -0.054208
25%      0.014197    0.007536    0.010761    0.009646    0.012029    0.009433
50%      0.027506    0.023046    0.025783    0.024276    0.025939    0.024712
75%      0.038647    0.033185    0.036414    0.035859    0.035349    0.036198
```

```
max      0.089348    0.092529    0.088753    0.092302    0.082978    0.084253

              X38         X39         X40
count  813.000000  813.000000  813.000000
mean     0.023618    0.023438    0.022859
std      0.020468    0.019455    0.020348
min     -0.051347   -0.055107   -0.056550
25%      0.011423    0.011491    0.011229
50%      0.027947    0.025192    0.024593
75%      0.037057    0.035915    0.035850
max      0.073990    0.085692    0.097817

[8 rows x 41 columns]
```

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

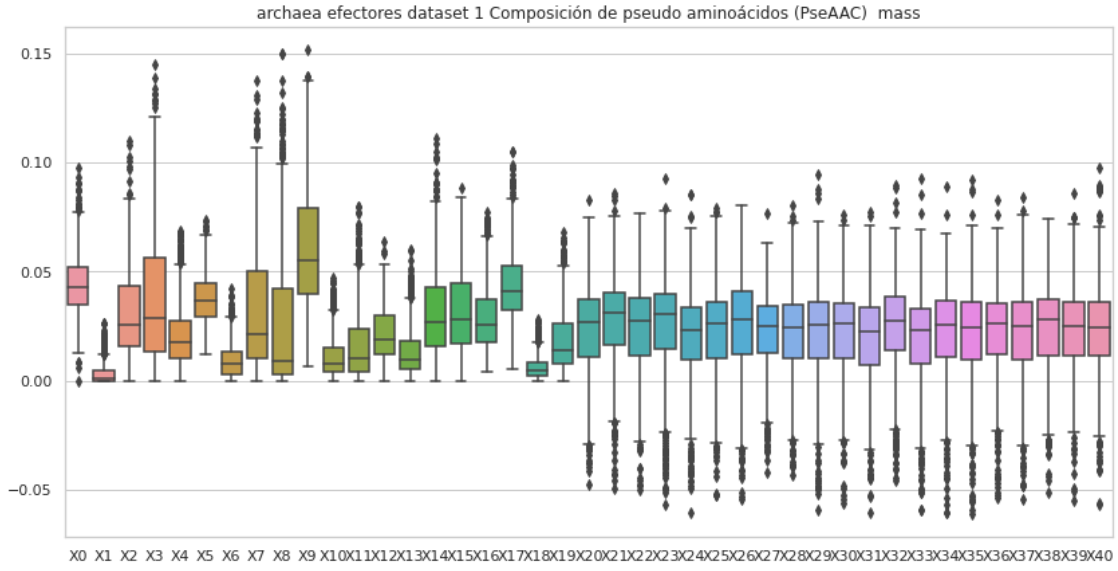
```
           X0        X1        X2        X3        X4        X5        X6  \
1    0.041499  0.001153  0.034583  0.035736  0.021902  0.034583  0.006917
2    0.030107  0.000000  0.006690  0.003345  0.013381  0.026761  0.000000
4    0.035210  0.000000  0.023473  0.023473  0.035210  0.052815  0.005868
5    0.040581  0.005636  0.025927  0.021418  0.012400  0.022545  0.004509
7    0.068507  0.007905  0.042158  0.052697  0.023714  0.036888  0.013174
..        …         …         …         …         …         …         …
993  0.050499  0.005941  0.036637  0.026735  0.018813  0.043568  0.011882
995  0.032833  0.001931  0.050215  0.009657  0.005794  0.025108  0.003863
997  0.045362  0.007217  0.039176  0.055672  0.020619  0.053610  0.018557
998  0.050445  0.001770  0.033630  0.030975  0.013275  0.032745  0.010620
999  0.022204  0.013323  0.022204  0.017764  0.008882  0.035527  0.004441

           X7        X8        X9  …       X32       X33       X34  \
1    0.023055  0.014986  0.031125  …  0.002026  0.036495  0.024051
2    0.020071  0.000000  0.066904  …  0.030613  0.038563  0.053622
4    0.011737  0.005868  0.076288  …  0.013848  0.037579  0.044759
5    0.025927  0.014654  0.027054  …  0.038088  0.023022  0.037095
7    0.005270  0.002635  0.055332  …  0.010093 -0.024052  0.034590
..        …         …         …  …  …       …         …         …
993  0.003961  0.004951  0.025745  …  0.030263  0.040489  0.031073
995  0.005794  0.001931  0.017382  …  0.034599  0.006217  0.041593
997  0.035052  0.029898  0.046393  …  0.018014  0.017184  0.008578
998  0.016815  0.015045  0.026550  …  0.023483  0.032529  0.030655
999  0.013323  0.048850  0.017764  …  0.015907  0.014433  0.013038

           X35       X36       X37       X38       X39       X40           X41
1    0.034519  0.016114  0.012812  0.028765  0.017577  0.034880  no_efectores
```

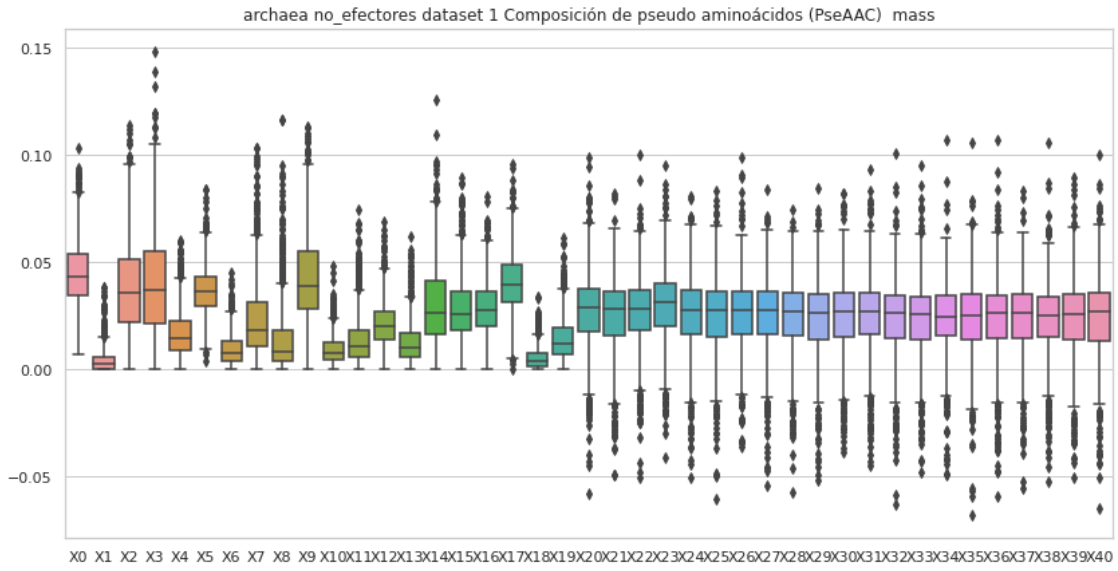
```
2    0.011022  0.029972  0.015314  0.025341  0.039456  0.038403  no_efectores
4   -0.008732  0.006148  0.030283  0.003180  0.003468 -0.004742  no_efectores
5    0.022759  0.038655  0.025685  0.019743  0.030577  0.022200  no_efectores
7   -0.015983  0.049759  0.006384 -0.007558 -0.023048  0.003765  no_efectores
..        …         …         …         …         …         …
993  0.018209  0.016657  0.013436  0.030121  0.025389  0.028754  no_efectores
995  0.036060  0.019967  0.042635  0.031104  0.024889  0.032725  no_efectores
997  0.026880  0.023928  0.024378  0.030235  0.018071  0.007142  no_efectores
998  0.023145  0.025311  0.035413  0.019880  0.023510  0.017553  no_efectores
999  0.001042  0.049954  0.029234  0.017134  0.046223  0.014257  no_efectores

[842 rows x 42 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass no_efectores archaea 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     0.044803    0.004620    0.038030    0.040296    0.016758    0.037240
std      0.015195    0.005962    0.021592    0.024968    0.010899    0.011019
min      0.006710    0.000000    0.000000    0.000000    0.000000    0.004074
25%      0.034338    0.000000    0.022114    0.021443    0.008873    0.029767
50%      0.043252    0.002713    0.035517    0.036720    0.014415    0.036430
75%      0.053724    0.006067    0.051649    0.055202    0.022604    0.043563
max      0.103552    0.038307    0.113707    0.148199    0.060039    0.083953

                X6          X7          X8          X9  …         X31  \
count  842.000000  842.000000  842.000000  842.000000  …  842.000000
mean     0.009284    0.023880    0.014313    0.042774  …    0.024756
std      0.007303    0.018689    0.017130    0.020035  …    0.017555
min      0.000000    0.000000    0.000000    0.000000  …   -0.045044
25%      0.004075    0.010725    0.003646    0.028368  …    0.016082
50%      0.007764    0.018397    0.007986    0.038663  …    0.027319
75%      0.013317    0.031694    0.018294    0.055256  …    0.035477
max      0.045001    0.103474    0.116232    0.113187  …    0.093093

               X32         X33         X34         X35         X36         X37  \
count  842.000000  842.000000  842.000000  842.000000  842.000000  842.000000
mean     0.024022    0.023213    0.023921    0.023906    0.023825    0.024481
std      0.018163    0.018154    0.016735    0.018304    0.018976    0.016989
min     -0.062888   -0.048212   -0.049256   -0.067983   -0.059241   -0.055392
25%      0.014603    0.014129    0.015853    0.013698    0.014460    0.014799
50%      0.026161    0.025963    0.024818    0.025108    0.026117    0.026564
75%      0.034806    0.034089    0.034796    0.035291    0.034524    0.035264
max      0.100605    0.094884    0.106761    0.105710    0.107069    0.082980
```

```
              X38          X39          X40
count  842.000000   842.000000   842.000000
mean     0.023751     0.023800     0.023879
std      0.017544     0.018379     0.018883
min     -0.052314    -0.050606    -0.065076
25%      0.014888     0.013908     0.013528
50%      0.025350     0.025855     0.026727
75%      0.033823     0.035002     0.035533
max      0.105807     0.089815     0.100134

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.053474	0.000000	0.053474	0.028794	0.028794	0.078154	0.016454	
1	0.046284	0.006428	0.021856	0.025713	0.015428	0.051427	0.002571	
2	0.151593	0.033687	0.101062	0.101062	0.033687	0.067375	0.000000	
3	0.037792	0.000000	0.018896	0.037792	0.056687	0.188958	0.037792	
4	0.047074	0.006725	0.033625	0.060524	0.000000	0.033625	0.006725	
..	…	…	…	…	…	…	…	
995	0.060313	0.012372	0.060313	0.068045	0.037116	0.057220	0.017011	
996	0.020509	0.000000	0.005982	0.005127	0.010255	0.020509	0.001709	
997	0.073425	0.000000	0.016317	0.007252	0.012691	0.033540	0.005439	
998	0.053733	0.003838	0.053733	0.080600	0.053733	0.049895	0.019190	
999	0.018016	0.001386	0.004157	0.001386	0.018016	0.031874	0.012472	

	X7	X8	X9	…	X53	X54	X55	\
0	0.123402	0.086381	0.037020	…	-0.029064	0.004072	0.024123	
1	0.012857	0.003857	0.028285	…	0.004390	0.025004	0.019205	
2	0.050531	0.050531	0.151593	…	-0.025008	0.011633	-0.033519	
3	0.113375	0.226750	0.056687	…	-0.014631	-0.180273	0.010996	
4	0.033625	0.080699	0.040349	…	-0.011718	-0.011598	0.059110	
..	…	…	…	…	…	…	…	
995	0.060313	0.040209	0.068045	…	0.029212	0.008373	-0.011983	
996	0.010255	0.002564	0.035891	…	0.006208	0.015130	0.000295	
997	0.002719	0.000000	0.046231	…	-0.000278	0.009131	-0.002734	
998	0.072924	0.082519	0.097871	…	0.003123	-0.013140	-0.002847	
999	0.020787	0.005543	0.044346	…	0.008929	0.008110	-0.001062	

	X56	X57	X58	X59	X60	X61	X62
0	0.019168	0.034561	0.033985	0.049363	-0.003326	-0.016992	efectores
1	0.019498	0.010236	0.006045	0.010689	0.019721	0.016511	efectores
2	0.143570	-0.011372	-0.113837	-0.006274	-0.141623	-0.133889	efectores
3	0.021402	0.009588	-0.186179	0.006089	0.207002	-0.044121	efectores
4	-0.030917	-0.022991	0.012755	0.013551	-0.019769	0.023069	efectores
..	…	…	…	…	…	…	…
995	0.000423	0.033370	0.041930	0.048512	-0.008960	0.002715	efectores
996	0.025970	-0.000221	0.041576	0.007907	0.016862	-0.003714	efectores
997	0.024135	0.008258	0.005682	-0.009032	0.019933	0.004652	efectores

```
998 -0.013388  0.007725  0.041580  0.042581  0.018556  0.009025  efectores
999  0.021344  0.004755  0.014408  0.005982  0.010763  0.000351  efectores

[1000 rows x 63 columns]
```

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

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.047597	0.004613	0.036851	0.045329	0.021753	
std	0.027119	0.007435	0.026774	0.032834	0.015319	
min	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.029190	0.000000	0.014189	0.013608	0.011247	
50%	0.041338	0.001729	0.029951	0.044870	0.018482	
75%	0.060597	0.005851	0.055427	0.067512	0.027800	
max	0.282517	0.054774	0.262818	0.262818	0.150297	

	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.040937	0.010481	0.036666	0.033027	0.062955	…	
std	0.022099	0.010131	0.032637	0.037558	0.033057	…	
min	0.002113	0.000000	0.000000	0.000000	0.009808	…	
25%	0.025809	0.003663	0.012272	0.003963	0.040188	…	
50%	0.035577	0.008739	0.025523	0.015182	0.054810	…	
75%	0.051299	0.014303	0.051968	0.057117	0.077381	…	
max	0.188958	0.147161	0.273871	0.310925	0.295670	…	

	X52	X53	X54	X55	X56	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.006953	0.010033	0.011797	0.014400	0.007360	
std	0.035893	0.027947	0.037069	0.032450	0.039983	
min	-0.203945	-0.188973	-0.309786	-0.203182	-0.356057	
25%	-0.007355	-0.002016	0.000127	-0.000527	-0.009698	
50%	0.011951	0.007241	0.016577	0.010723	0.013255	
75%	0.025862	0.021323	0.028898	0.030358	0.025676	
max	0.175221	0.151268	0.169327	0.135170	0.187872	

	X57	X58	X59	X60	X61
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.011466	0.008646	0.011052	0.009206	0.011551
std	0.030722	0.039511	0.033092	0.038543	0.033005
min	-0.154985	-0.299614	-0.250501	-0.209982	-0.180516
25%	-0.002087	-0.007489	-0.002090	-0.006942	-0.001626
50%	0.007692	0.013028	0.008407	0.015745	0.007588
75%	0.024277	0.028110	0.024436	0.028435	0.025344

```
max       0.218122     0.366813     0.236961     0.207002     0.202238

[8 rows x 62 columns]
```

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.033536	0.006097	0.021341	0.033536	0.012195	0.018292	0.009146	
1	0.079302	0.002203	0.066085	0.068288	0.041854	0.066085	0.013217	
2	0.028936	0.000000	0.006430	0.003215	0.012860	0.025721	0.000000	
3	0.034768	0.028973	0.052151	0.040562	0.020281	0.023178	0.014486	
4	0.015910	0.000000	0.010607	0.010607	0.015910	0.023865	0.002652	
..	…	…	…	…	…	…	…	
995	0.019584	0.001152	0.029952	0.005760	0.003456	0.014976	0.002304	
996	0.030132	0.020088	0.035154	0.055243	0.042687	0.022599	0.017577	
997	0.047345	0.007532	0.040888	0.058105	0.021520	0.055953	0.019368	
998	0.091606	0.003214	0.061070	0.056249	0.024107	0.059463	0.019285	
999	0.022876	0.013726	0.022876	0.018301	0.009150	0.036601	0.004575	

	X7	X8	X9	…	X53	X54	X55	\
0	0.024390	0.009146	0.024390	…	-0.003755	0.013955	0.027985	
1	0.044057	0.028637	0.059476	…	0.050623	-0.015956	-0.018020	
2	0.019291	0.000000	0.064302	…	0.003963	-0.001490	0.001006	
3	0.055049	0.098508	0.060843	…	0.071928	-0.032621	-0.059613	
4	0.005303	0.002652	0.034471	…	0.028270	0.007563	0.023846	
..	…	…	…	…	…	…	…	
995	0.003456	0.001152	0.010368	…	0.015536	0.011555	0.026069	
996	0.085375	0.045198	0.057754	…	0.001736	0.022495	0.008983	
997	0.036584	0.031204	0.048421	…	0.005204	-0.011904	0.025668	
998	0.030535	0.027321	0.048213	…	0.029904	-0.020479	-0.018231	
999	0.013726	0.050327	0.018301	…	-0.019263	-0.025978	0.028593	

	X56	X57	X58	X59	X60	X61	X62
0	0.031816	0.054159	0.019650	0.003445	0.000030	-0.006429	no_efectores
1	0.007955	-0.010689	0.014695	0.015375	0.030527	0.028536	no_efectores
2	-0.033820	-0.014335	0.023140	-0.001606	0.033752	0.018082	no_efectores
3	-0.022670	-0.015766	0.022774	0.037161	0.012718	0.031380	no_efectores
4	0.018827	0.011955	0.011620	0.007035	-0.007797	-0.006832	no_efectores
..	…	…	…	…	…	…	…
995	0.018765	0.021406	0.011020	0.022452	0.006959	0.023689	no_efectores
996	0.018390	-0.000784	-0.022436	-0.008836	0.026406	0.030508	no_efectores
997	-0.010916	0.003025	-0.016321	0.019869	0.016497	0.029332	no_efectores
998	0.002828	-0.001421	-0.015449	-0.021525	0.007815	0.003570	no_efectores

```
999  0.034163  0.027386  0.026502  0.019882 -0.013985  0.012381  no_efectores

[1000 rows x 63 columns]
```

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

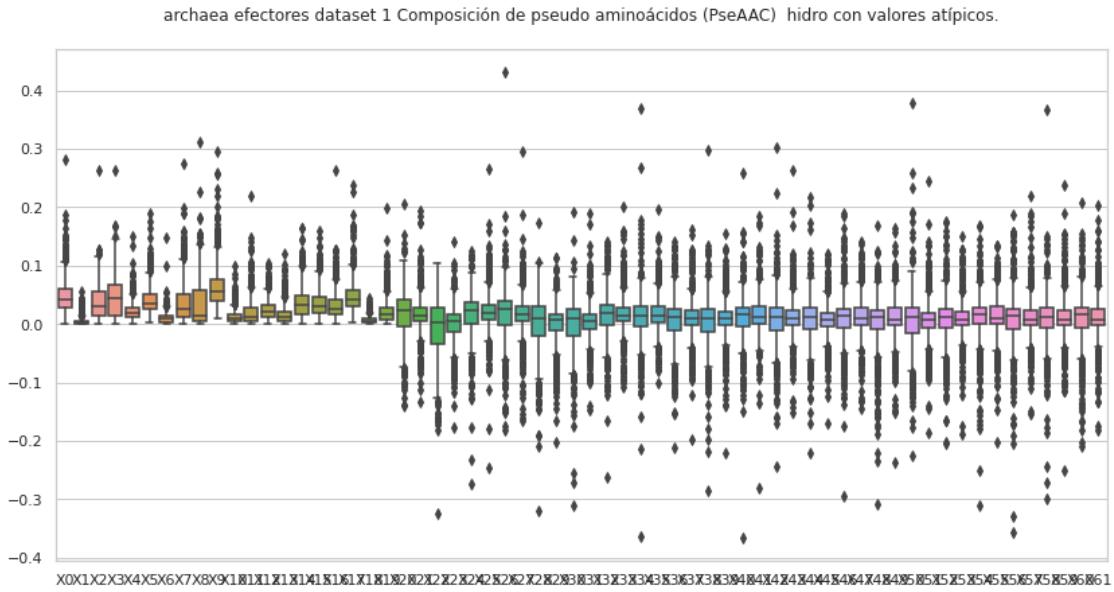
	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.059144	0.008491	0.051088	0.052270	0.022475	
std	0.048069	0.033712	0.054655	0.034156	0.026607	
min	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.034226	0.000000	0.029912	0.033141	0.011682	
50%	0.053454	0.003871	0.052540	0.053633	0.018282	
75%	0.076149	0.009042	0.065729	0.067580	0.027513	
max	0.999946	0.999946	1.499919	0.499973	0.499973	

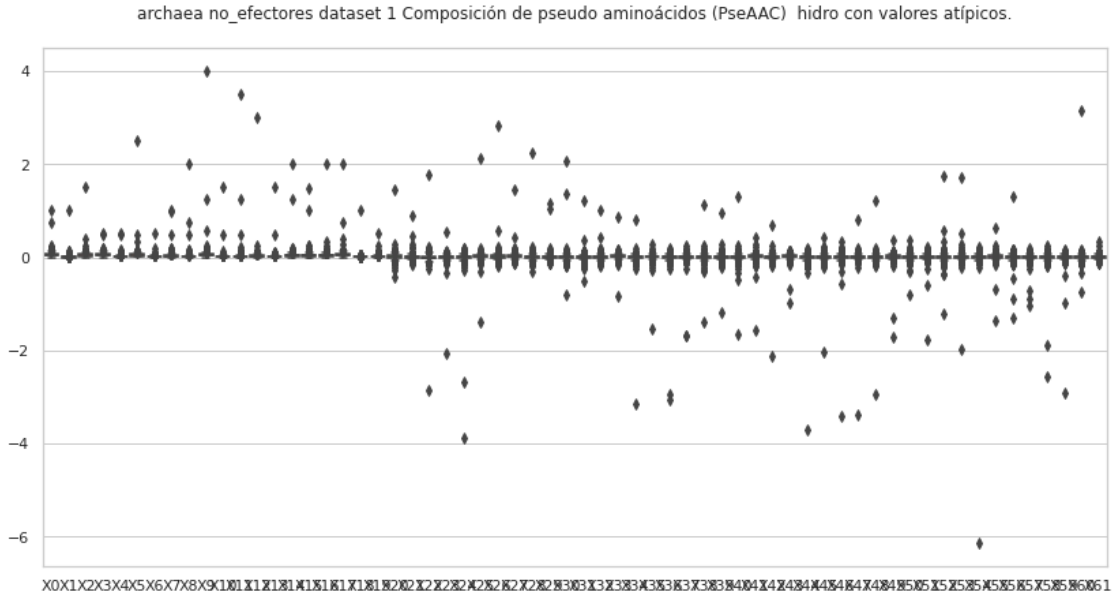
	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.051567	0.013267	0.033938	0.023488	0.059225	…	
std	0.082965	0.019805	0.052902	0.072336	0.133903	…	
min	0.003168	0.000000	0.000000	0.000000	0.000000	…	
25%	0.029895	0.004449	0.014247	0.005083	0.035936	…	
50%	0.045112	0.010566	0.023530	0.011146	0.049724	…	
75%	0.062760	0.018251	0.040415	0.025868	0.066850	…	
max	2.499865	0.499973	0.999946	1.999892	3.999784	…	

	X52	X53	X54	X55	X56	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.004762	0.011975	-0.001337	0.013479	0.002326	
std	0.080634	0.090848	0.197379	0.061605	0.074408	
min	-1.219700	-1.991296	-6.140820	-1.369110	-1.296666	
25%	-0.009870	-0.001353	-0.010688	-0.000887	-0.011328	
50%	0.007153	0.010870	0.005829	0.012794	0.004863	
75%	0.022973	0.026922	0.021425	0.030696	0.021930	
max	1.747595	1.709468	0.213636	0.623502	1.316850	

	X57	X58	X59	X60	X61
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.008897	-0.001614	0.007078	0.005959	0.011718
std	0.057746	0.106430	0.102581	0.108213	0.031812
min	-1.034381	-2.558549	-2.902057	-0.760560	-0.128755
25%	-0.002032	-0.012765	-0.002576	-0.013090	-0.003011
50%	0.010300	0.004815	0.011560	0.005793	0.011219
75%	0.026514	0.019845	0.026695	0.021831	0.027633
max	0.183302	0.263780	0.174300	3.157648	0.351683

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.053474	0.000000	0.053474	0.028794	0.028794	0.078154	0.016454	
1	0.046284	0.006428	0.021856	0.025713	0.015428	0.051427	0.002571	
4	0.047074	0.006725	0.033625	0.060524	0.000000	0.033625	0.006725	
6	0.046391	0.003866	0.027061	0.061854	0.019329	0.019329	0.011598	
7	0.043476	0.005435	0.032607	0.054345	0.043476	0.035325	0.005435	
..	…	…	…	…	…	…	…	
995	0.060313	0.012372	0.060313	0.068045	0.037116	0.057220	0.017011	
996	0.020509	0.000000	0.005982	0.005127	0.010255	0.020509	0.001709	
997	0.073425	0.000000	0.016317	0.007252	0.012691	0.033540	0.005439	
998	0.053733	0.003838	0.053733	0.080600	0.053733	0.049895	0.019190	
999	0.018016	0.001386	0.004157	0.001386	0.018016	0.031874	0.012472	

	X7	X8	X9	…	X53	X54	X55	\
0	0.123402	0.086381	0.037020	…	-0.029064	0.004072	0.024123	
1	0.012857	0.003857	0.028285	…	0.004390	0.025004	0.019205	
4	0.033625	0.080699	0.040349	…	-0.011718	-0.011598	0.059110	
6	0.050257	0.061854	0.050257	…	0.041251	-0.003204	0.020913	
7	0.095104	0.081518	0.059780	…	0.052252	-0.035830	-0.001738	
..	…	…	…	…	…	…	…	
995	0.060313	0.040209	0.068045	…	0.029212	0.008373	-0.011983	
996	0.010255	0.002564	0.035891	…	0.006208	0.015130	0.000295	
997	0.002719	0.000000	0.046231	…	-0.000278	0.009131	-0.002734	
998	0.072924	0.082519	0.097871	…	0.003123	-0.013140	-0.002847	
999	0.020787	0.005543	0.044346	…	0.008929	0.008110	-0.001062	

	X56	X57	X58	X59	X60	X61	X62
0	0.019168	0.034561	0.033985	0.049363	-0.003326	-0.016992	efectores
1	0.019498	0.010236	0.006045	0.010689	0.019721	0.016511	efectores
4	-0.030917	-0.022991	0.012755	0.013551	-0.019769	0.023069	efectores
6	0.013345	0.004699	0.096988	0.080730	0.057616	0.066965	efectores
7	0.032654	0.044519	0.010563	0.016707	-0.002717	-0.023988	efectores
..	…	…	…	…	…	…	…
995	0.000423	0.033370	0.041930	0.048512	-0.008960	0.002715	efectores
996	0.025970	-0.000221	0.041576	0.007907	0.016862	-0.003714	efectores
997	0.024135	0.008258	0.005682	-0.009032	0.019933	0.004652	efectores
998	-0.013388	0.007725	0.041580	0.042581	0.018556	0.009025	efectores
999	0.021344	0.004755	0.014408	0.005982	0.010763	0.000351	efectores

[796 rows x 63 columns]

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

	X0	X1	X2	X3	X4	X5	\
count	796.000000	796.000000	796.000000	796.000000	796.000000	796.000000	
mean	0.044160	0.002992	0.031005	0.036289	0.019072	0.036275	
std	0.022317	0.004548	0.022194	0.026654	0.011152	0.016467	
min	0.002723	0.000000	0.000000	0.000000	0.000000	0.002113	
25%	0.028557	0.000000	0.012098	0.010309	0.010571	0.024787	
50%	0.038934	0.001123	0.024059	0.032342	0.017122	0.033115	
75%	0.057214	0.004124	0.047529	0.057611	0.025750	0.044890	
max	0.127929	0.025246	0.100867	0.120084	0.059095	0.106700	

	X6	X7	X8	X9	…	X52	\
count	796.000000	796.000000	796.000000	796.000000	…	796.000000	
mean	0.008962	0.028714	0.023595	0.053926	…	0.009448	
std	0.007292	0.024293	0.028100	0.021092	…	0.024048	
min	0.000000	0.000000	0.000000	0.010288	…	-0.090086	
25%	0.002996	0.010494	0.002917	0.037402	…	-0.002256	
50%	0.007839	0.019909	0.009440	0.049554	…	0.013918	
75%	0.012940	0.040455	0.039193	0.067147	…	0.025462	
max	0.038883	0.133005	0.118097	0.152874	…	0.110236	

	X53	X54	X55	X56	X57	X58	\
count	796.000000	796.000000	796.000000	796.000000	796.000000	796.000000	
mean	0.009227	0.015713	0.015115	0.010759	0.011581	0.011209	
std	0.018609	0.023187	0.023307	0.023556	0.020467	0.025330	
min	-0.062952	-0.097921	-0.068313	-0.076472	-0.061661	-0.102652	
25%	-0.000844	0.004421	0.000440	-0.004438	-0.000770	-0.001265	
50%	0.006756	0.017492	0.010156	0.015195	0.007882	0.014671	
75%	0.018481	0.028120	0.028413	0.025281	0.020881	0.027635	
max	0.088100	0.122700	0.098897	0.125035	0.091921	0.096988	

	X59	X60	X61
count	796.000000	796.000000	796.000000
mean	0.010762	0.012586	0.010902
std	0.021203	0.024856	0.021767
min	-0.061531	-0.097748	-0.078312
25%	-0.000206	-0.001288	-0.000689
50%	0.008342	0.016708	0.007274
75%	0.020712	0.027330	0.020820
max	0.087422	0.112532	0.103848

[8 rows x 62 columns]

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.033536	0.006097	0.021341	0.033536	0.012195	0.018292	0.009146	
1	0.079302	0.002203	0.066085	0.068288	0.041854	0.066085	0.013217	
2	0.028936	0.000000	0.006430	0.003215	0.012860	0.025721	0.000000	
3	0.034768	0.028973	0.052151	0.040562	0.020281	0.023178	0.014486	
4	0.015910	0.000000	0.010607	0.010607	0.015910	0.023865	0.002652	
..	…	…	…	…	…	…	…	
995	0.019584	0.001152	0.029952	0.005760	0.003456	0.014976	0.002304	
996	0.030132	0.020088	0.035154	0.055243	0.042687	0.022599	0.017577	
997	0.047345	0.007532	0.040888	0.058105	0.021520	0.055953	0.019368	
998	0.091606	0.003214	0.061070	0.056249	0.024107	0.059463	0.019285	
999	0.022876	0.013726	0.022876	0.018301	0.009150	0.036601	0.004575	

	X7	X8	X9	…	X53	X54	X55	\
0	0.024390	0.009146	0.024390	…	-0.003755	0.013955	0.027985	
1	0.044057	0.028637	0.059476	…	0.050623	-0.015956	-0.018020	
2	0.019291	0.000000	0.064302	…	0.003963	-0.001490	0.001006	
3	0.055049	0.098508	0.060843	…	0.071928	-0.032621	-0.059613	
4	0.005303	0.002652	0.034471	…	0.028270	0.007563	0.023846	
..	…	…	…	…	…	…	…	
995	0.003456	0.001152	0.010368	…	0.015536	0.011555	0.026069	
996	0.085375	0.045198	0.057754	…	0.001736	0.022495	0.008983	
997	0.036584	0.031204	0.048421	…	0.005204	-0.011904	0.025668	
998	0.030535	0.027321	0.048213	…	0.029904	-0.020479	-0.018231	
999	0.013726	0.050327	0.018301	…	-0.019263	-0.025978	0.028593	

	X56	X57	X58	X59	X60	X61	X62
0	0.031816	0.054159	0.019650	0.003445	0.000030	-0.006429	no_efectores
1	0.007955	-0.010689	0.014695	0.015375	0.030527	0.028536	no_efectores
2	-0.033820	-0.014335	0.023140	-0.001606	0.033752	0.018082	no_efectores
3	-0.022670	-0.015766	0.022774	0.037161	0.012718	0.031380	no_efectores
4	0.018827	0.011955	0.011620	0.007035	-0.007797	-0.006832	no_efectores
..	…	…	…	…	…	…	…
995	0.018765	0.021406	0.011020	0.022452	0.006959	0.023689	no_efectores
996	0.018390	-0.000784	-0.022436	-0.008836	0.026406	0.030508	no_efectores
997	-0.010916	0.003025	-0.016321	0.019869	0.016497	0.029332	no_efectores
998	0.002828	-0.001421	-0.015449	-0.021525	0.007815	0.003570	no_efectores
999	0.034163	0.027386	0.026502	0.019882	-0.013985	0.012381	no_efectores

[956 rows x 63 columns]

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

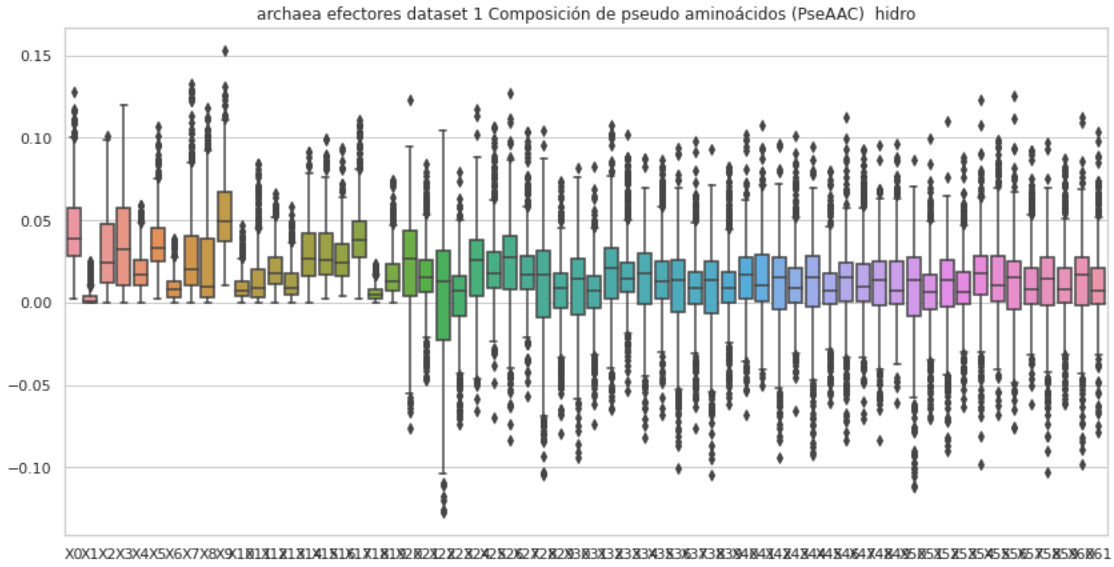
	X0	X1	X2	X3	X4	X5	\
count	956.000000	956.000000	956.000000	956.000000	956.000000	956.000000	
mean	0.056483	0.006674	0.047961	0.050373	0.020384	0.047318	
std	0.029513	0.009527	0.025375	0.026283	0.012830	0.023748	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.003168	
25%	0.033841	0.000000	0.028893	0.031984	0.011596	0.029446	
50%	0.052441	0.003793	0.052053	0.053359	0.017939	0.044515	
75%	0.075048	0.008652	0.065197	0.066859	0.026554	0.061469	
max	0.187526	0.080976	0.140823	0.145114	0.079047	0.145083	

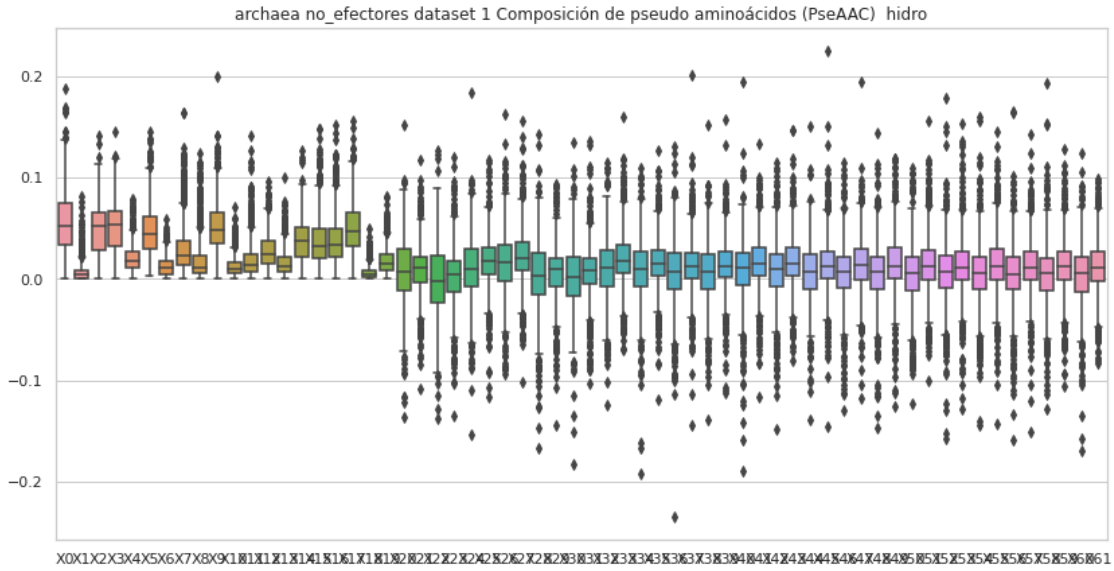
	X6	X7	X8	X9	…	X52	\
count	956.000000	956.000000	956.000000	956.000000	…	956.000000	
mean	0.012057	0.029892	0.018634	0.051893	…	0.004867	
std	0.009559	0.024474	0.021187	0.023459	…	0.030684	
min	0.000000	0.000000	0.000000	0.000000	…	-0.157120	
25%	0.004467	0.014054	0.005010	0.035478	…	-0.008621	
50%	0.010405	0.022378	0.010817	0.048815	…	0.007376	
75%	0.017730	0.038095	0.023162	0.065152	…	0.022361	
max	0.058315	0.164295	0.124419	0.199021	…	0.178832	

	X53	X54	X55	X56	X57	X58	\
count	956.000000	956.000000	956.000000	956.000000	956.000000	956.000000	
mean	0.012176	0.004304	0.014313	0.004560	0.012585	0.003613	
std	0.028185	0.029883	0.028153	0.029168	0.025952	0.030391	
min	-0.128164	-0.144032	-0.142571	-0.159288	-0.151143	-0.128603	
25%	-0.000801	-0.010106	-0.000710	-0.010254	-0.000953	-0.011811	
50%	0.010870	0.005829	0.012303	0.004967	0.011077	0.005341	
75%	0.026648	0.021121	0.029876	0.021490	0.026423	0.019943	
max	0.152645	0.159748	0.144903	0.165296	0.142696	0.192682	

	X59	X60	X61
count	956.000000	956.000000	956.000000
mean	0.011742	0.004236	0.011551
std	0.026153	0.028723	0.025672
min	-0.106235	-0.169436	-0.082949
25%	-0.001595	-0.012275	-0.002656
50%	0.011748	0.005998	0.011330
75%	0.026606	0.021369	0.027186
max	0.127789	0.123707	0.098258

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

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.004467  0.025270  0.029109 -0.014898 -0.034573  0.021768  0.088447
1    0.174532  0.037062  0.073368 -0.046045 -0.037508  0.003810  0.021214
2    0.121882 -0.048610 -0.126531  0.176028  0.156178  0.037505 -0.226181
3   -0.022762 -0.076583  0.020413  0.081808 -0.131319 -0.157695  0.113266
4   -0.028106 -0.020061  0.024813  0.014477 -0.080353  0.033493  0.161468
..        …         …         …         …         …         …         …
995 -0.028978  0.004694  0.017077 -0.068358 -0.072025  0.057904  0.049565
996 -0.010575 -0.015659  0.005466 -0.100160 -0.073163  0.030508  0.076399
997 -0.015739 -0.043570 -0.063631  0.010209  0.076073 -0.077238 -0.029054
998  0.003045 -0.006731  0.010004 -0.085439  0.034774  0.002450 -0.095348
999  0.015063 -0.002766  0.062789  0.050233  0.074395 -0.047095 -0.049058

           X7        X8        X9       X10       X11       X12        X13
0    0.116823  0.085791  0.026360 -0.002731 -0.056457  0.001458  efectores
```

```
1    0.082825  0.033241  0.030027  0.030046  0.029351 -0.014977  efectores
2    0.001765  0.116190  0.122720 -0.059632 -0.186723 -0.137583  efectores
3    0.129175 -0.064325 -0.019066 -0.005863  0.130961 -0.145167  efectores
4   -0.048152  0.011490  0.159687  0.087638  0.126576  0.053434  efectores
..        …         …         …         …         …         …         …
995  0.049814 -0.018185 -0.004198  0.108470 -0.057275  0.038206  efectores
996 -0.098265 -0.056314  0.016383 -0.081312  0.039864  0.090599  efectores
997 -0.058863 -0.048790  0.008715  0.046763 -0.004816 -0.000429  efectores
998 -0.044134 -0.010362  0.002989 -0.064141  0.085057  0.030626  efectores
999 -0.046605 -0.066869 -0.039698 -0.032557  0.050323 -0.023956  efectores

[1000 rows x 14 columns]
```

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

	X0	X1	X2	X3	X4
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.006550	0.018382	0.011528	0.015090	-0.008103
std	0.068541	0.069876	0.067231	0.069609	0.069673
min	-0.207577	-0.213337	-0.194444	-0.209807	-0.231342
25%	-0.033842	-0.023677	-0.030544	-0.030347	-0.051492
50%	0.006308	0.019963	0.007771	0.017977	-0.005772
75%	0.049617	0.060892	0.051595	0.061488	0.036433
max	0.256098	0.236765	0.253958	0.249920	0.227940

	X5	X6	X7	X8	X9
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	-0.000176	0.017620	0.001203	-0.003690	0.007148
std	0.069848	0.071538	0.066100	0.072418	0.074518
min	-0.296697	-0.226181	-0.276376	-0.259188	-0.254148
25%	-0.043196	-0.026889	-0.037444	-0.051666	-0.036318
50%	0.003229	0.015729	0.000448	-0.001389	0.005180
75%	0.040572	0.063151	0.043360	0.042279	0.048959
max	0.295946	0.274666	0.259286	0.267947	0.237354

	X10	X11	X12
count	1000.000000	1000.000000	1000.000000
mean	0.002407	-0.008446	0.016022
std	0.067974	0.067191	0.071811
min	-0.244414	-0.305499	-0.247439
25%	-0.037571	-0.048502	-0.028383
50%	0.004953	-0.006316	0.012143
75%	0.043788	0.031939	0.065285
max	0.240354	0.234573	0.276764

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

	X0	X1	X2	X3	X4	X5	X6	\
0	-0.023608	0.016581	0.050176	0.088799	0.088735	0.006631	-0.242662	
1	0.076231	-0.038437	0.020548	-0.007432	0.048997	-0.021988	0.096001	
2	0.012784	0.082061	-0.004551	0.063337	-0.021487	-0.013913	0.000099	
3	-0.020406	0.014784	0.053661	0.047365	-0.082953	0.046269	-0.023208	
4	0.196390	-0.146285	0.070285	0.044591	0.073889	0.088384	-0.021129	
..	…	…	…	…	…	…	…	
995	0.044649	0.100367	0.084165	-0.089450	0.103239	0.061377	0.111764	
996	-0.053369	-0.023012	0.021413	0.015762	0.025735	-0.003694	0.063903	
997	-0.056326	-0.022288	-0.024945	-0.001347	-0.069617	-0.021421	0.030651	
998	0.153364	0.000734	0.114013	0.153007	0.034945	0.056845	0.078581	
999	0.074229	0.104071	0.258920	0.162793	0.071392	0.327072	0.035270	

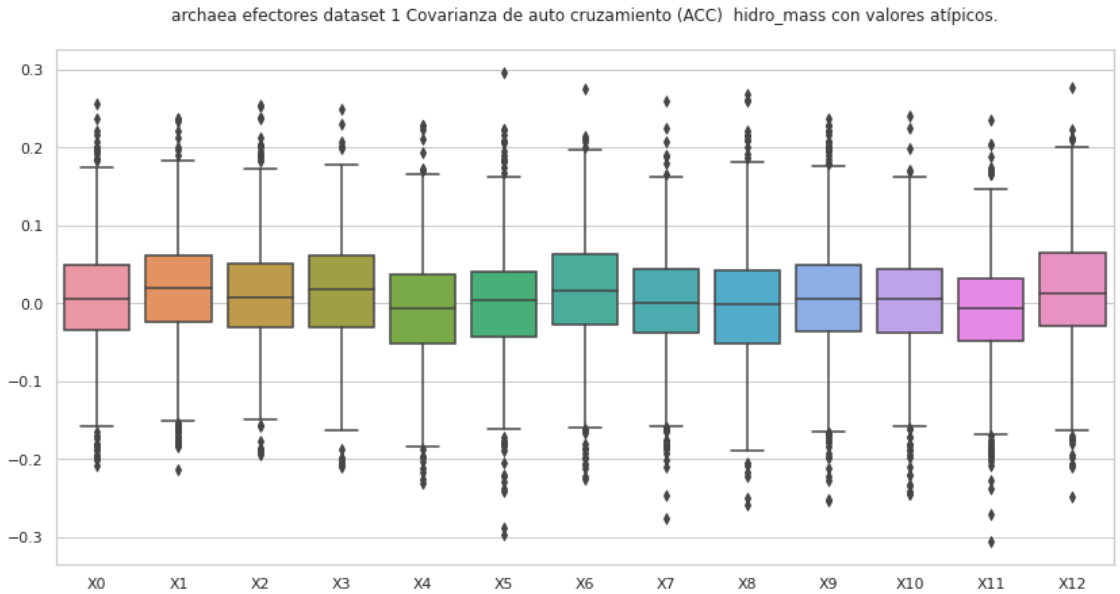
	X7	X8	X9	X10	X11	X12	X13
0	0.153345	0.010670	-0.042241	-0.119489	-0.031693	0.017364	no_efectores
1	0.012017	0.025248	-0.024683	-0.032070	0.028766	-0.095301	no_efectores
2	0.079815	-0.029214	0.043530	-0.042490	0.027532	0.034218	no_efectores
3	-0.014847	-0.061738	-0.000289	0.055549	-0.041674	0.001385	no_efectores
4	-0.078481	-0.114637	-0.032571	0.084423	0.059866	-0.010267	no_efectores
..	…	…	…	…	…	…	…
995	-0.044043	0.030475	0.130789	-0.071428	0.015855	0.030420	no_efectores
996	-0.002448	0.047333	0.014310	-0.033347	-0.067822	-0.057850	no_efectores
997	0.006465	-0.011653	0.026763	-0.029144	0.003890	-0.008506	no_efectores
998	0.027297	0.041865	-0.011640	0.106487	-0.037063	0.029315	no_efectores
999	0.207343	0.084416	0.071199	0.099584	0.314936	0.006451	no_efectores

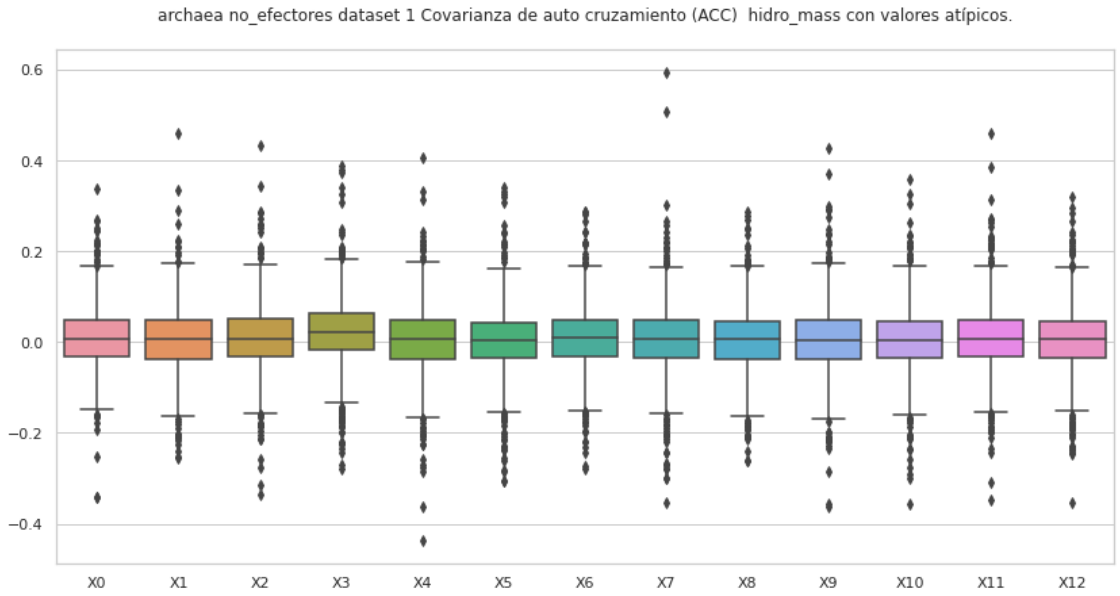
[1000 rows x 14 columns]

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

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.010427	0.004245	0.007274	0.022168	0.006151	
std	0.072422	0.072746	0.074314	0.076124	0.077493	
min	-0.342930	-0.256097	-0.337309	-0.279149	-0.437261	
25%	-0.033221	-0.036449	-0.033247	-0.018196	-0.037899	
50%	0.006023	0.005524	0.008017	0.021246	0.005987	
75%	0.047456	0.048669	0.050668	0.063215	0.048866	

```
max       0.336806     0.460173     0.432147     0.388635     0.405202

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.003571     0.009008     0.004716     0.004057     0.004543
std       0.075507     0.071602     0.078870     0.071770     0.076824
min      -0.307510    -0.279469    -0.355057    -0.261391    -0.364012
25%      -0.036230    -0.031848    -0.034761    -0.038369    -0.039254
50%       0.004286     0.008992     0.005920     0.006099     0.003025
75%       0.043902     0.047894     0.047889     0.044159     0.047102
max       0.339799     0.286500     0.593796     0.287446     0.427335

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.003888     0.007338     0.005832
std       0.075108     0.074641     0.075705
min      -0.355750    -0.348601    -0.353689
25%      -0.036161    -0.032750    -0.034065
50%       0.004516     0.005763     0.006619
75%       0.046291     0.048410     0.045337
max       0.359647     0.458860     0.320888
```





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

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.004467  0.025270  0.029109 -0.014898 -0.034573  0.021768  0.088447
1    0.174532  0.037062  0.073368 -0.046045 -0.037508  0.003810  0.021214
3   -0.022762 -0.076583  0.020413  0.081808 -0.131319 -0.157695  0.113266
4   -0.028106 -0.020061  0.024813  0.014477 -0.080353  0.033493  0.161468
5   -0.011271  0.009650  0.106747 -0.003978  0.006774  0.025265  0.037600
..        …         …         …         …         …         …         …
995 -0.028978  0.004694  0.017077 -0.068358 -0.072025  0.057904  0.049565
996 -0.010575 -0.015659  0.005466 -0.100160 -0.073163  0.030508  0.076399
997 -0.015739 -0.043570 -0.063631  0.010209  0.076073 -0.077238 -0.029054
998  0.003045 -0.006731  0.010004 -0.085439  0.034774  0.002450 -0.095348
999  0.015063 -0.002766  0.062789  0.050233  0.074395 -0.047095 -0.049058

           X7        X8        X9       X10       X11       X12        X13
0    0.116823  0.085791  0.026360 -0.002731 -0.056457  0.001458  efectores
1    0.082825  0.033241  0.030027  0.030046  0.029351 -0.014977  efectores
3    0.129175 -0.064325 -0.019066 -0.005863  0.130961 -0.145167  efectores
4   -0.048152  0.011490  0.159687  0.087638  0.126576  0.053434  efectores
5   -0.149196 -0.010759 -0.012205 -0.111435 -0.084777 -0.038512  efectores
..        …         …         …         …         …         …         …
995  0.049814 -0.018185 -0.004198  0.108470 -0.057275  0.038206  efectores
```

```
996 -0.098265 -0.056314  0.016383 -0.081312  0.039864  0.090599  efectores
997 -0.058863 -0.048790  0.008715  0.046763 -0.004816 -0.000429  efectores
998 -0.044134 -0.010362  0.002989 -0.064141  0.085057  0.030626  efectores
999 -0.046605 -0.066869 -0.039698 -0.032557  0.050323 -0.023956  efectores

[931 rows x 14 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	931.000000	931.000000	931.000000	931.000000	931.000000	931.000000	
mean	0.005469	0.019997	0.012066	0.014359	-0.008953	-0.000224	
std	0.064928	0.066305	0.063796	0.065592	0.066209	0.064208	
min	-0.198197	-0.182578	-0.187647	-0.163111	-0.212280	-0.203828	
25%	-0.033873	-0.020081	-0.028541	-0.030853	-0.049527	-0.041111	
50%	0.005396	0.021193	0.007829	0.017817	-0.005937	0.002999	
75%	0.047765	0.060834	0.051467	0.060951	0.034695	0.038571	
max	0.199898	0.220448	0.212835	0.199477	0.192909	0.207717	

	X6	X7	X8	X9	X10	X11	\
count	931.000000	931.000000	931.000000	931.000000	931.000000	931.000000	
mean	0.019218	0.001347	-0.004142	0.008005	0.003893	-0.007065	
std	0.067682	0.061908	0.067497	0.070779	0.062941	0.062438	
min	-0.192238	-0.191659	-0.217562	-0.212382	-0.198532	-0.207266	
25%	-0.023323	-0.035936	-0.050165	-0.034415	-0.035892	-0.045642	
50%	0.016686	0.000477	-0.001470	0.005145	0.004913	-0.005928	
75%	0.063152	0.043257	0.040797	0.048476	0.043102	0.030880	
max	0.214690	0.187509	0.210004	0.228184	0.198513	0.188858	

	X12
count	931.000000
mean	0.016596
std	0.067596
min	-0.197727
25%	-0.026915
50%	0.012222
75%	0.063244
max	0.222822

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

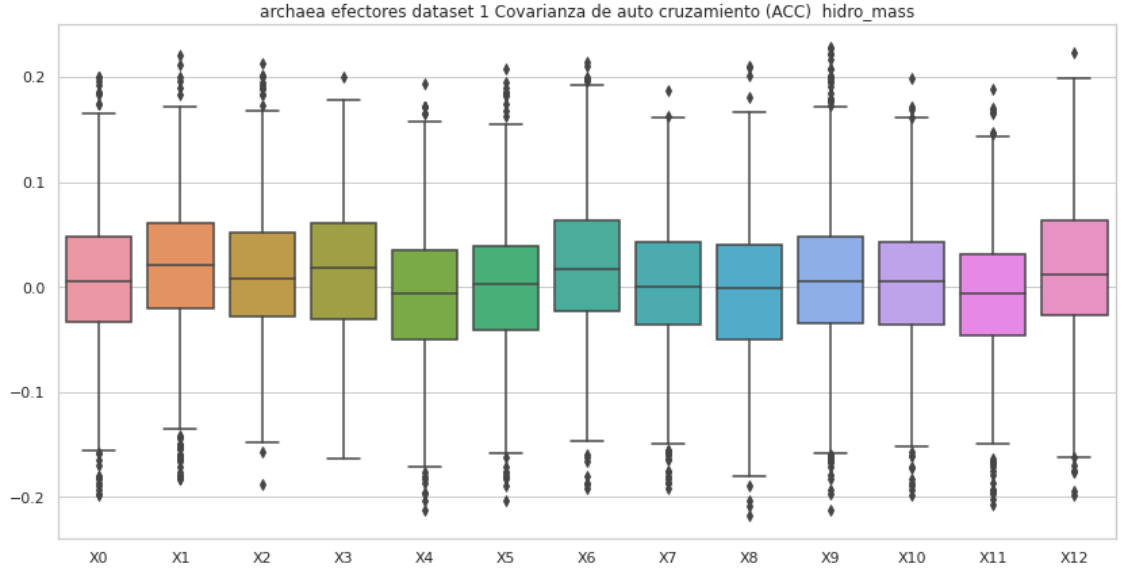
```
           X0        X1        X2        X3        X4        X5        X6  \
1    0.076231 -0.038437  0.020548 -0.007432  0.048997 -0.021988  0.096001
2    0.012784  0.082061 -0.004551  0.063337 -0.021487 -0.013913  0.000099
3   -0.020406  0.014784  0.053661  0.047365 -0.082953  0.046269 -0.023208
4    0.196390 -0.146285  0.070285  0.044591  0.073889  0.088384 -0.021129
5    0.010448  0.070843  0.010297  0.007297 -0.033660  0.007049 -0.005346
..        …         …         …         …         …         …
994 -0.010648  0.015710 -0.206250  0.065003 -0.081886 -0.030950  0.020072
995  0.044649  0.100367  0.084165 -0.089450  0.103239  0.061377  0.111764
996 -0.053369 -0.023012  0.021413  0.015762  0.025735 -0.003694  0.063903
997 -0.056326 -0.022288 -0.024945 -0.001347 -0.069617 -0.021421  0.030651
998  0.153364  0.000734  0.114013  0.153007  0.034945  0.056845  0.078581

           X7        X8        X9       X10       X11       X12           X13
1    0.012017  0.025248 -0.024683 -0.032070  0.028766 -0.095301  no_efectores
2    0.079815 -0.029214  0.043530 -0.042490  0.027532  0.034218  no_efectores
3   -0.014847 -0.061738 -0.000289  0.055549 -0.041674  0.001385  no_efectores
4   -0.078481 -0.114637 -0.032571  0.084423  0.059866 -0.010267  no_efectores
5   -0.057112  0.046953  0.092937  0.024314  0.020300  0.056526  no_efectores
..        …         …         …         …         …         …
994 -0.085142  0.179154 -0.098943 -0.172711  0.080495  0.165621  no_efectores
995 -0.044043  0.030475  0.130789 -0.071428  0.015855  0.030420  no_efectores
996 -0.002448  0.047333  0.014310 -0.033347 -0.067822 -0.057850  no_efectores
997  0.006465 -0.011653  0.026763 -0.029144  0.003890 -0.008506  no_efectores
998  0.027297  0.041865 -0.011640  0.106487 -0.037063  0.029315  no_efectores

[910 rows x 14 columns]
```

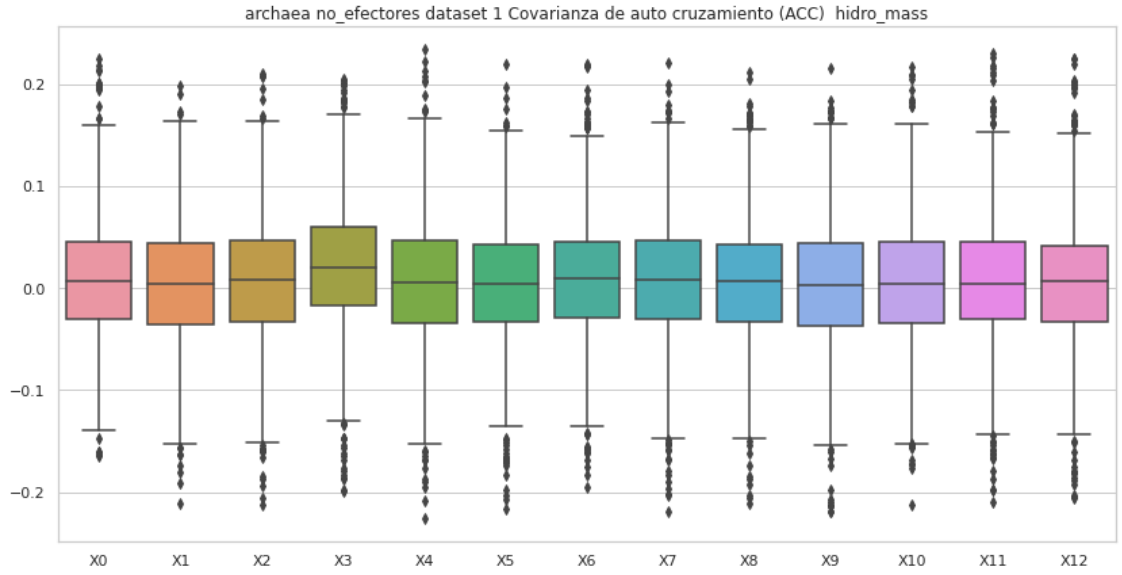
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores archaea dataset 1, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  910.000000  910.000000  910.000000  910.000000  910.000000  910.000000
mean     0.010030    0.002272    0.006070    0.021176    0.007253    0.003889
std      0.064285    0.062664    0.062795    0.065553    0.066994    0.062923
min     -0.164696   -0.210676   -0.212569   -0.199312   -0.225923   -0.216027
25%     -0.030763   -0.035361   -0.032507   -0.016837   -0.034353   -0.033349
50%      0.006276    0.004050    0.007954    0.020753    0.006196    0.004209
75%      0.045951    0.044591    0.046719    0.060046    0.046688    0.042258
max      0.225209    0.198224    0.209912    0.205645    0.234284    0.220198

                X6          X7          X8          X9         X10         X11  \
count  910.000000  910.000000  910.000000  910.000000  910.000000  910.000000
mean     0.009888    0.006593    0.005724    0.001910    0.004614    0.005445
std      0.062369    0.063112    0.063236    0.064770    0.063679    0.063222
```

min	-0.194938	-0.219342	-0.211014	-0.219785	-0.212512	-0.210290
25%	-0.028676	-0.030966	-0.033032	-0.036838	-0.034113	-0.030567
50%	0.009369	0.007665	0.006822	0.002690	0.004611	0.004790
75%	0.044942	0.046614	0.043309	0.043790	0.045290	0.045103
max	0.219105	0.220559	0.211950	0.215494	0.217332	0.230282

	X12
count	910.000000
mean	0.005975
std	0.065821
min	-0.205466
25%	-0.032394
50%	0.006619
75%	0.041925
max	0.225277





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

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.004467  0.025270  0.029109 -0.014898 -0.034573  0.021768  0.088447
1    0.174532  0.037062  0.073368 -0.046045 -0.037508  0.003810  0.021214
2    0.121882 -0.048610 -0.126531  0.176028  0.156178  0.037505 -0.226181
3   -0.022762 -0.076583  0.020413  0.081808 -0.131319 -0.157695  0.113266
4   -0.028106 -0.020061  0.024813  0.014477 -0.080353  0.033493  0.161468
..        …         …         …         …         …         …         …
995 -0.028978  0.004694  0.017077 -0.068358 -0.072025  0.057904  0.049565
996 -0.010575 -0.015659  0.005466 -0.100160 -0.073163  0.030508  0.076399
997 -0.015739 -0.043570 -0.063631  0.010209  0.076073 -0.077238 -0.029054
998  0.003045 -0.006731  0.010004 -0.085439  0.034774  0.002450 -0.095348
999  0.015063 -0.002766  0.062789  0.050233  0.074395 -0.047095 -0.049058

           X7        X8        X9       X10       X11       X12         X13
0    0.116823  0.085791  0.026360 -0.002731 -0.056457  0.001458  efectores
1    0.082825  0.033241  0.030027  0.030046  0.029351 -0.014977  efectores
2    0.001765  0.116190  0.122720 -0.059632 -0.186723 -0.137583  efectores
3    0.129175 -0.064325 -0.019066 -0.005863  0.130961 -0.145167  efectores
4   -0.048152  0.011490  0.159687  0.087638  0.126576  0.053434  efectores
..        …         …         …         …         …         …         …
995  0.049814 -0.018185 -0.004198  0.108470 -0.057275  0.038206  efectores
996 -0.098265 -0.056314  0.016383 -0.081312  0.039864  0.090599  efectores
997 -0.058863 -0.048790  0.008715  0.046763 -0.004816 -0.000429  efectores
998 -0.044134 -0.010362  0.002989 -0.064141  0.085057  0.030626  efectores
999 -0.046605 -0.066869 -0.039698 -0.032557  0.050323 -0.023956  efectores

[1000 rows x 14 columns]


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

                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.006550     0.018382     0.011528     0.015090    -0.008103
```

```
std       0.068541     0.069876     0.067231     0.069609     0.069673
min      -0.207577    -0.213337    -0.194444    -0.209807    -0.231342
25%      -0.033842    -0.023677    -0.030544    -0.030347    -0.051492
50%       0.006308     0.019963     0.007771     0.017977    -0.005772
75%       0.049617     0.060892     0.051595     0.061488     0.036433
max       0.256098     0.236765     0.253958     0.249920     0.227940

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean     -0.000176     0.017620     0.001203    -0.003690     0.007148
std       0.069848     0.071538     0.066100     0.072418     0.074518
min      -0.296697    -0.226181    -0.276376    -0.259188    -0.254148
25%      -0.043196    -0.026889    -0.037444    -0.051666    -0.036318
50%       0.003229     0.015729     0.000448    -0.001389     0.005180
75%       0.040572     0.063151     0.043360     0.042279     0.048959
max       0.295946     0.274666     0.259286     0.267947     0.237354

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.002407    -0.008446     0.016022
std       0.067974     0.067191     0.071811
min      -0.244414    -0.305499    -0.247439
25%      -0.037571    -0.048502    -0.028383
50%       0.004953    -0.006316     0.012143
75%       0.043788     0.031939     0.065285
max       0.240354     0.234573     0.276764
```

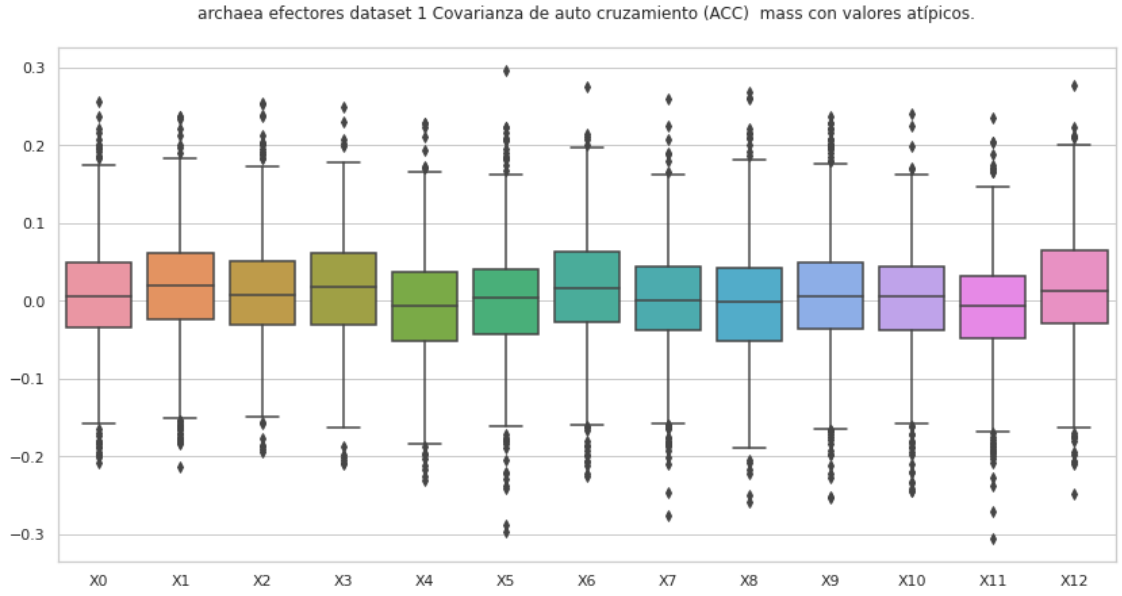
no_efectores
Covarianza de auto cruzamiento (ACC) mass no_efectores archaea dataset 1, con valores atípicos.
Valores del documento csv.

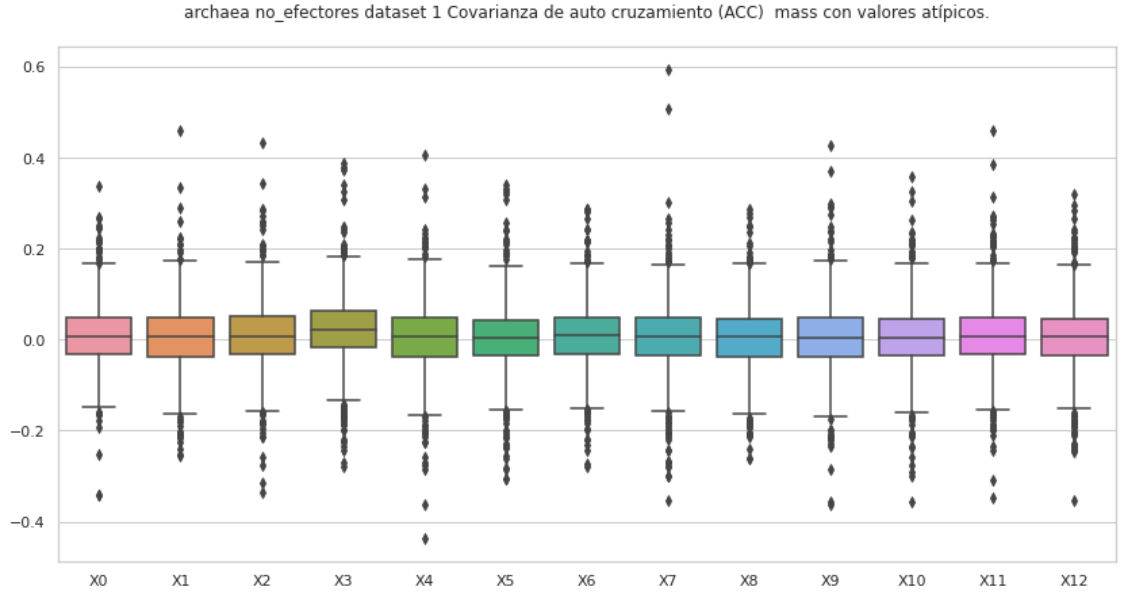
```
           X0        X1        X2        X3        X4        X5        X6  \
0   -0.023608  0.016581  0.050176  0.088799  0.088735  0.006631 -0.242662
1    0.076231 -0.038437  0.020548 -0.007432  0.048997 -0.021988  0.096001
2    0.012784  0.082061 -0.004551  0.063337 -0.021487 -0.013913  0.000099
3   -0.020406  0.014784  0.053661  0.047365 -0.082953  0.046269 -0.023208
4    0.196390 -0.146285  0.070285  0.044591  0.073889  0.088384 -0.021129
..        …         …         …         …         …         …         …
995  0.044649  0.100367  0.084165 -0.089450  0.103239  0.061377  0.111764
996 -0.053369 -0.023012  0.021413  0.015762  0.025735 -0.003694  0.063903
997 -0.056326 -0.022288 -0.024945 -0.001347 -0.069617 -0.021421  0.030651
998  0.153364  0.000734  0.114013  0.153007  0.034945  0.056845  0.078581
999  0.074229  0.104071  0.258920  0.162793  0.071392  0.327072  0.035270

           X7        X8        X9       X10       X11       X12           X13
```

```
0    0.153345  0.010670 -0.042241 -0.119489 -0.031693  0.017364  no_efectores
1    0.012017  0.025248 -0.024683 -0.032070  0.028766 -0.095301  no_efectores
2    0.079815 -0.029214  0.043530 -0.042490  0.027532  0.034218  no_efectores
3   -0.014847 -0.061738 -0.000289  0.055549 -0.041674  0.001385  no_efectores
4   -0.078481 -0.114637 -0.032571  0.084423  0.059866 -0.010267  no_efectores
..        …       …       …       …       …       …           …
995 -0.044043  0.030475  0.130789 -0.071428  0.015855  0.030420  no_efectores
996 -0.002448  0.047333  0.014310 -0.033347 -0.067822 -0.057850  no_efectores
997  0.006465 -0.011653  0.026763 -0.029144  0.003890 -0.008506  no_efectores
998  0.027297  0.041865 -0.011640  0.106487 -0.037063  0.029315  no_efectores
999  0.207343  0.084416  0.071199  0.099584  0.314936  0.006451  no_efectores

[1000 rows x 14 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.010427     0.004245     0.007274     0.022168     0.006151
std       0.072422     0.072746     0.074314     0.076124     0.077493
min      -0.342930    -0.256097    -0.337309    -0.279149    -0.437261
25%      -0.033221    -0.036449    -0.033247    -0.018196    -0.037899
50%       0.006023     0.005524     0.008017     0.021246     0.005987
75%       0.047456     0.048669     0.050668     0.063215     0.048866
max       0.336806     0.460173     0.432147     0.388635     0.405202

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.003571     0.009008     0.004716     0.004057     0.004543
std       0.075507     0.071602     0.078870     0.071770     0.076824
min      -0.307510    -0.279469    -0.355057    -0.261391    -0.364012
25%      -0.036230    -0.031848    -0.034761    -0.038369    -0.039254
50%       0.004286     0.008992     0.005920     0.006099     0.003025
75%       0.043902     0.047894     0.047889     0.044159     0.047102
max       0.339799     0.286500     0.593796     0.287446     0.427335

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.003888     0.007338     0.005832
std       0.075108     0.074641     0.075705
min      -0.355750    -0.348601    -0.353689
25%      -0.036161    -0.032750    -0.034065
50%       0.004516     0.005763     0.006619
75%       0.046291     0.048410     0.045337
max       0.359647     0.458860     0.320888
```


archaea efectores dataset 1 Covarianza de auto cruzamiento (ACC) mass con valores atípicos.
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12


archaea no_efectores dataset 1 Covarianza de auto cruzamiento (ACC) mass con valores atípicos.
0.6
0.4
0.2
0.0
−0.2
−0.4
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12

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

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

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

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

    if etiq == "efectores":
        df=ACC_mass_efec

    if etiq == "no_efectores":
        df=ACC_mass_no_efec

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

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

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

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

```
    ax.set_title(organismo +' '+str(etiq)+" dataset "+str(dataset)+"␣
↪"+str(transf)+" "+str(comp))
```

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

	X0	X1	X2	X3	X4	X5	X6	\
0	-0.004467	0.025270	0.029109	-0.014898	-0.034573	0.021768	0.088447	
1	0.174532	0.037062	0.073368	-0.046045	-0.037508	0.003810	0.021214	
3	-0.022762	-0.076583	0.020413	0.081808	-0.131319	-0.157695	0.113266	
4	-0.028106	-0.020061	0.024813	0.014477	-0.080353	0.033493	0.161468	
5	-0.011271	0.009650	0.106747	-0.003978	0.006774	0.025265	0.037600	
..	…	…	…	…	…	…	…	
995	-0.028978	0.004694	0.017077	-0.068358	-0.072025	0.057904	0.049565	
996	-0.010575	-0.015659	0.005466	-0.100160	-0.073163	0.030508	0.076399	
997	-0.015739	-0.043570	-0.063631	0.010209	0.076073	-0.077238	-0.029054	
998	0.003045	-0.006731	0.010004	-0.085439	0.034774	0.002450	-0.095348	
999	0.015063	-0.002766	0.062789	0.050233	0.074395	-0.047095	-0.049058	

	X7	X8	X9	X10	X11	X12	X13
0	0.116823	0.085791	0.026360	-0.002731	-0.056457	0.001458	efectores
1	0.082825	0.033241	0.030027	0.030046	0.029351	-0.014977	efectores
3	0.129175	-0.064325	-0.019066	-0.005863	0.130961	-0.145167	efectores
4	-0.048152	0.011490	0.159687	0.087638	0.126576	0.053434	efectores
5	-0.149196	-0.010759	-0.012205	-0.111435	-0.084777	-0.038512	efectores
..	…	…	…	…	…	…	…
995	0.049814	-0.018185	-0.004198	0.108470	-0.057275	0.038206	efectores
996	-0.098265	-0.056314	0.016383	-0.081312	0.039864	0.090599	efectores
997	-0.058863	-0.048790	0.008715	0.046763	-0.004816	-0.000429	efectores
998	-0.044134	-0.010362	0.002989	-0.064141	0.085057	0.030626	efectores
999	-0.046605	-0.066869	-0.039698	-0.032557	0.050323	-0.023956	efectores

```
[931 rows x 14 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	931.000000	931.000000	931.000000	931.000000	931.000000	931.000000	
mean	0.005469	0.019997	0.012066	0.014359	-0.008953	-0.000224	
std	0.064928	0.066305	0.063796	0.065592	0.066209	0.064208	
min	-0.198197	-0.182578	-0.187647	-0.163111	-0.212280	-0.203828	
25%	-0.033873	-0.020081	-0.028541	-0.030853	-0.049527	-0.041111	
50%	0.005396	0.021193	0.007829	0.017817	-0.005937	0.002999	
75%	0.047765	0.060834	0.051467	0.060951	0.034695	0.038571	

```
max      0.199898    0.220448    0.212835    0.199477    0.192909    0.207717

               X6          X7          X8          X9         X10         X11  \
count  931.000000  931.000000  931.000000  931.000000  931.000000  931.000000
mean     0.019218    0.001347   -0.004142    0.008005    0.003893   -0.007065
std      0.067682    0.061908    0.067497    0.070779    0.062941    0.062438
min     -0.192238   -0.191659   -0.217562   -0.212382   -0.198532   -0.207266
25%     -0.023323   -0.035936   -0.050165   -0.034415   -0.035892   -0.045642
50%      0.016686    0.000477   -0.001470    0.005145    0.004913   -0.005928
75%      0.063152    0.043257    0.040797    0.048476    0.043102    0.030880
max      0.214690    0.187509    0.210004    0.228184    0.198513    0.188858

              X12
count  931.000000
mean     0.016596
std      0.067596
min     -0.197727
25%     -0.026915
50%      0.012222
75%      0.063244
max      0.222822
```

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

```
           X0        X1        X2        X3        X4        X5        X6  \
1    0.076231 -0.038437  0.020548 -0.007432  0.048997 -0.021988  0.096001
2    0.012784  0.082061 -0.004551  0.063337 -0.021487 -0.013913  0.000099
3   -0.020406  0.014784  0.053661  0.047365 -0.082953  0.046269 -0.023208
4    0.196390 -0.146285  0.070285  0.044591  0.073889  0.088384 -0.021129
5    0.010448  0.070843  0.010297  0.007297 -0.033660  0.007049 -0.005346
..        …         …         …         …         …         …
994 -0.010648  0.015710 -0.206250  0.065003 -0.081886 -0.030950  0.020072
995  0.044649  0.100367  0.084165 -0.089450  0.103239  0.061377  0.111764
996 -0.053369 -0.023012  0.021413  0.015762  0.025735 -0.003694  0.063903
997 -0.056326 -0.022288 -0.024945 -0.001347 -0.069617 -0.021421  0.030651
998  0.153364  0.000734  0.114013  0.153007  0.034945  0.056845  0.078581

           X7        X8        X9       X10       X11       X12           X13
1    0.012017  0.025248 -0.024683 -0.032070  0.028766 -0.095301  no_efectores
2    0.079815 -0.029214  0.043530 -0.042490  0.027532  0.034218  no_efectores
3   -0.014847 -0.061738 -0.000289  0.055549 -0.041674  0.001385  no_efectores
4   -0.078481 -0.114637 -0.032571  0.084423  0.059866 -0.010267  no_efectores
5   -0.057112  0.046953  0.092937  0.024314  0.020300  0.056526  no_efectores
..        …         …         …         …         …         …           …
```

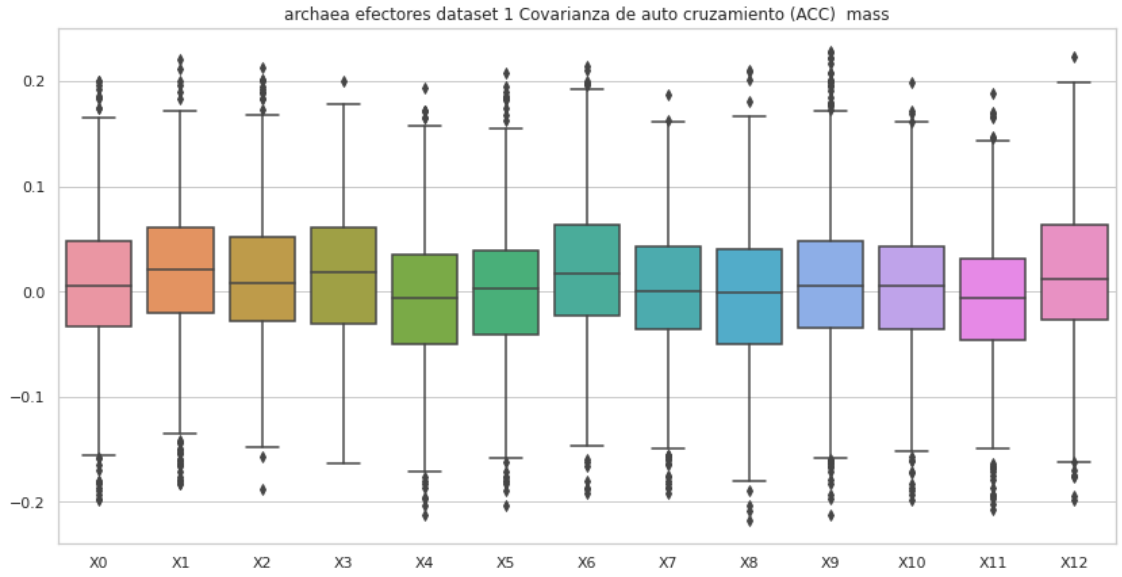
```
994 -0.085142  0.179154 -0.098943 -0.172711  0.080495  0.165621  no_efectores
995 -0.044043  0.030475  0.130789 -0.071428  0.015855  0.030420  no_efectores
996 -0.002448  0.047333  0.014310 -0.033347 -0.067822 -0.057850  no_efectores
997  0.006465 -0.011653  0.026763 -0.029144  0.003890 -0.008506  no_efectores
998  0.027297  0.041865 -0.011640  0.106487 -0.037063  0.029315  no_efectores

[910 rows x 14 columns]
```

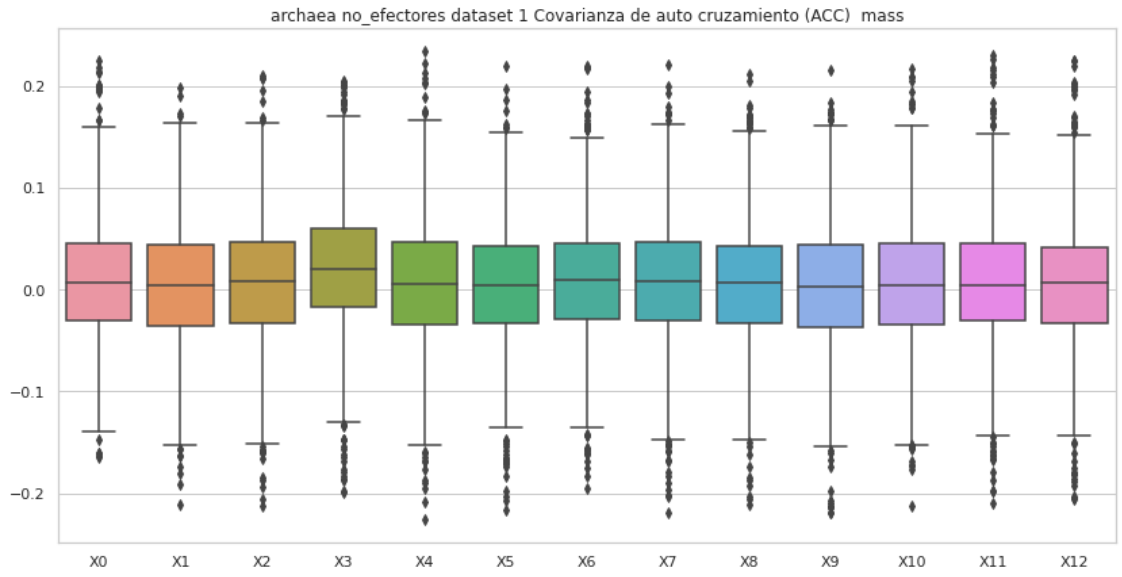
Covarianza de auto cruzamiento (ACC) mass no_efectores archaea dataset 1, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5
count	910.000000	910.000000	910.000000	910.000000	910.000000	910.000000
mean	0.010030	0.002272	0.006070	0.021176	0.007253	0.003889
std	0.064285	0.062664	0.062795	0.065553	0.066994	0.062923
min	-0.164696	-0.210676	-0.212569	-0.199312	-0.225923	-0.216027
25%	-0.030763	-0.035361	-0.032507	-0.016837	-0.034353	-0.033349
50%	0.006276	0.004050	0.007954	0.020753	0.006196	0.004209
75%	0.045951	0.044591	0.046719	0.060046	0.046688	0.042258
max	0.225209	0.198224	0.209912	0.205645	0.234284	0.220198

	X6	X7	X8	X9	X10	X11
count	910.000000	910.000000	910.000000	910.000000	910.000000	910.000000
mean	0.009888	0.006593	0.005724	0.001910	0.004614	0.005445
std	0.062369	0.063112	0.063236	0.064770	0.063679	0.063222
min	-0.194938	-0.219342	-0.211014	-0.219785	-0.212512	-0.210290
25%	-0.028676	-0.030966	-0.033032	-0.036838	-0.034113	-0.030567
50%	0.009369	0.007665	0.006822	0.002690	0.004611	0.004790
75%	0.044942	0.046614	0.043309	0.043790	0.045290	0.045103
max	0.219105	0.220559	0.211950	0.215494	0.217332	0.230282

	X12
count	910.000000
mean	0.005975
std	0.065821
min	-0.205466
25%	-0.032394
50%	0.006619
75%	0.041925
max	0.225277





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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.039762 -0.121160  0.046815  0.038100 -0.014330  0.020466  0.002374
1    0.089078  0.085529  0.164658  0.058391  0.093552  0.069839  0.027884
2   -0.034468 -0.076228 -0.046261 -0.092294 -0.142525 -0.153825  0.036229
3    0.039921  0.015301 -0.168705  0.088319  0.080902 -0.017531 -0.130694
4   -0.059767 -0.087097  0.175296  0.232745  0.076060 -0.078998 -0.031112
..        …         …         …         …         …         …         …
995 -0.034737 -0.176673 -0.017216  0.026185  0.070100 -0.136641 -0.021024
996  0.110527 -0.027128  0.102080  0.085090  0.053182  0.009982  0.085851
997  0.118052  0.085712  0.015242  0.030000  0.086867  0.070415  0.077697
998  0.034211 -0.163536 -0.013067  0.001532 -0.039201  0.043288  0.050365
999 -0.001770  0.007376  0.038181  0.030045  0.039600 -0.006506 -0.001123

           X7        X8        X9       X10       X11       X12        X13
0   -0.069621 -0.018454  0.101224  0.080683 -0.023190 -0.148151  efectores
1   -0.010880  0.061180  0.018912  0.004052  0.108526  0.096870  efectores
2    0.218504  0.133326  0.048071  0.036323 -0.138014 -0.080164  efectores
```

```
3   -0.047128  0.119209  0.093434 -0.383855 -0.219232 -0.082221  efectores
4    0.085351 -0.005391 -0.102273 -0.140827  0.050743  0.102831  efectores
..        …       …        …        …        …        …        …
995  0.083219 -0.029190 -0.125920 -0.020569  0.050607 -0.089524  efectores
996  0.093861  0.033035 -0.003754  0.011807  0.049906 -0.030605  efectores
997  0.077952 -0.023559 -0.003262  0.010419  0.011495 -0.028471  efectores
998  0.057132 -0.018085 -0.155163 -0.079608  0.026331  0.029440  efectores
999  0.103807 -0.023006  0.020405  0.024045  0.044606  0.034815  efectores

[1000 rows x 14 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.024496    -0.026452     0.044671     0.033658    -0.019627
std       0.090375     0.098422     0.085002     0.092073     0.099840
min      -0.280702    -0.372374    -0.318515    -0.312696    -0.460252
25%      -0.029113    -0.094839    -0.010792    -0.013229    -0.083312
50%       0.027323    -0.015959     0.037843     0.034455    -0.011833
75%       0.084134     0.048278     0.097665     0.082247     0.048916
max       0.325564     0.251843     0.318952     0.650649     0.262869

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean     -0.016955     0.029618     0.022989    -0.000515    -0.000282
std       0.085896     0.087750     0.082907     0.087409     0.084891
min      -0.387770    -0.283130    -0.271607    -0.309934    -0.343687
25%      -0.072590    -0.024278    -0.027014    -0.048426    -0.044065
50%      -0.015419     0.022961     0.021307     0.004410     0.000961
75%       0.036427     0.077671     0.071184     0.047235     0.045196
max       0.227382     0.375477     0.430776     0.260837     0.291494

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.017168     0.008356    -0.007673
std       0.081345     0.085066     0.080281
min      -0.383855    -0.302499    -0.361826
25%      -0.029755    -0.042774    -0.052808
50%       0.007996     0.002995    -0.006106
75%       0.067841     0.056779     0.038063
max       0.308909     0.424556     0.272967
```

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.142877	0.050328	0.043193	0.074674	0.157371	-0.017938	0.056920	
1	-0.049283	-0.150823	-0.014428	0.017522	-0.025395	-0.031961	0.075864	
2	-0.055686	-0.041942	0.052189	0.151169	-0.053262	-0.072608	0.068049	
3	0.020190	-0.033354	0.059725	0.027032	-0.151143	-0.047166	-0.011113	
4	0.327120	0.026398	0.239750	0.271274	0.114581	0.144157	0.217013	
..	…	…	…	…	…	…	…	
995	0.199178	0.338843	0.245793	0.273023	0.277500	0.107143	0.369219	
996	-0.019117	-0.059150	0.065125	-0.091222	0.114946	0.082187	-0.001551	
997	0.018037	-0.025500	-0.058443	-0.008068	0.042766	-0.027799	-0.043659	
998	-0.069709	-0.062702	-0.031853	0.074894	-0.069951	0.003910	-0.049446	
999	-0.029378	-0.119047	0.241290	-0.012778	0.084675	-0.004010	0.091379	

	X7	X8	X9	X10	X11	X12	X13
0	0.151663	-0.011647	0.133410	0.232357	0.095623	0.017287	no_efectores
1	0.024849	-0.072310	0.023962	-0.014032	0.006031	-0.053492	no_efectores
2	0.055371	0.018979	-0.079546	-0.000960	-0.045562	0.026857	no_efectores
3	0.005142	-0.072586	-0.035991	-0.048064	-0.053641	-0.011660	no_efectores
4	0.085482	0.026151	0.075958	0.267726	0.132929	0.091095	no_efectores
..	…	…	…	…	…	…	…
995	0.113766	0.325806	0.152451	0.204731	0.149492	0.194239	no_efectores
996	-0.024768	0.049299	0.005893	0.054478	0.022979	-0.031294	no_efectores
997	-0.012401	0.023211	0.056428	-0.016406	-0.033675	0.050311	no_efectores
998	-0.089603	0.028885	0.102730	0.014952	0.013084	0.080266	no_efectores
999	-0.011442	0.028730	0.200693	0.112019	0.014180	0.238867	no_efectores

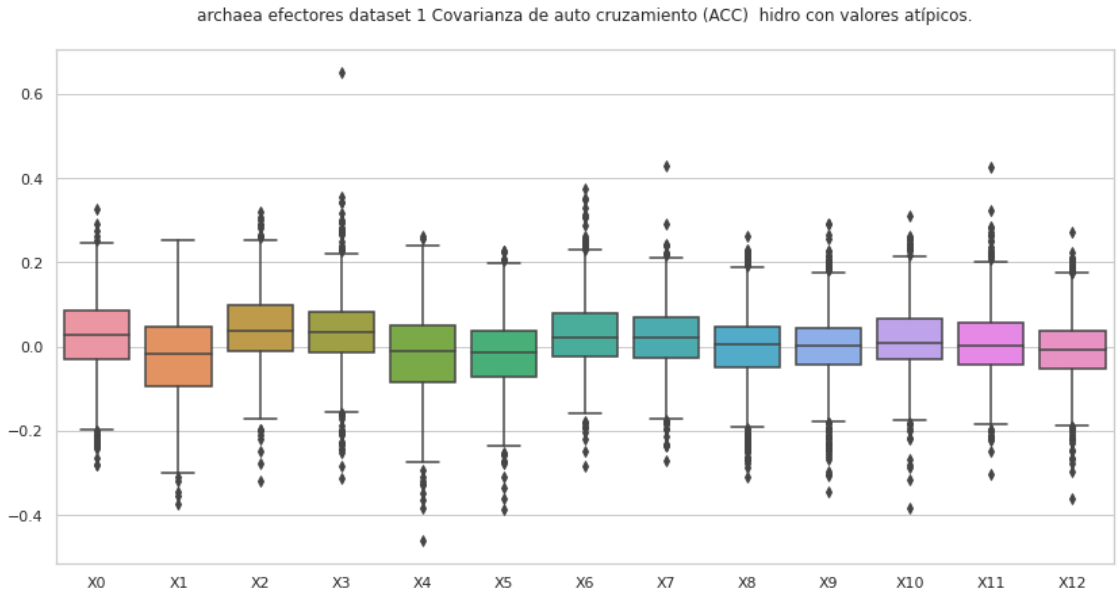
[1000 rows x 14 columns]

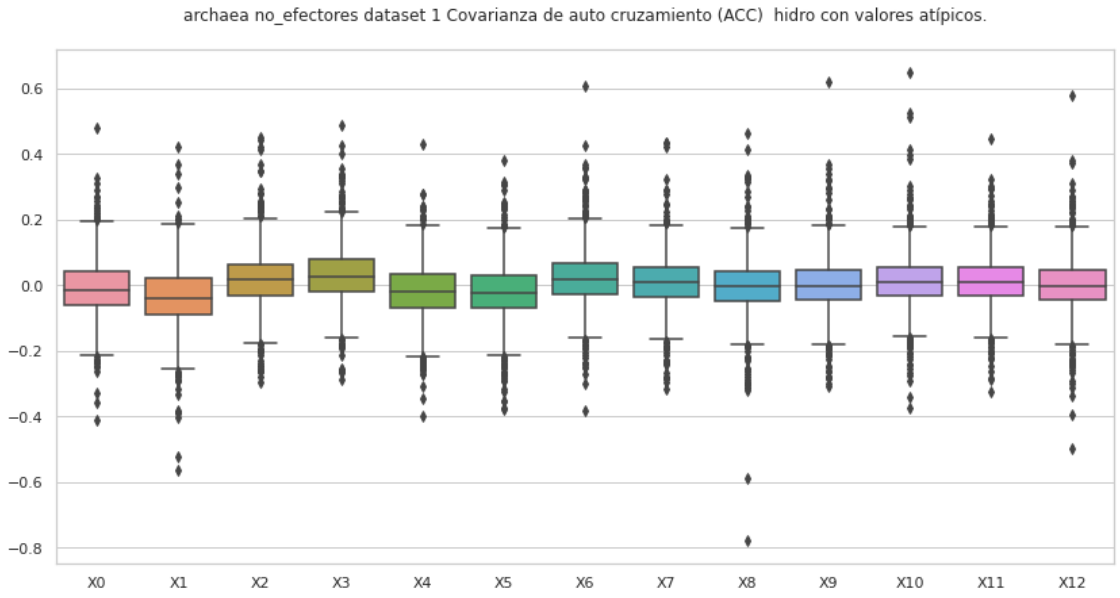
Covarianza de auto cruzamiento (ACC) hidro no_efectores archaea dataset 1, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	-0.010196	-0.036551	0.018051	0.031023	-0.019124	
std	0.088347	0.097688	0.088090	0.090421	0.087193	
min	-0.411084	-0.563163	-0.295444	-0.288859	-0.400287	
25%	-0.062201	-0.087708	-0.031021	-0.018566	-0.067828	
50%	-0.013639	-0.039531	0.015954	0.027705	-0.020655	
75%	0.040925	0.023284	0.065078	0.078400	0.032874	
max	0.479647	0.422766	0.451431	0.488480	0.429852	

	X5	X6	X7	X8	X9	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	-0.020102	0.020681	0.007562	-0.001087	0.002902	
std	0.089951	0.090125	0.082100	0.091381	0.086793	
min	-0.379694	-0.383885	-0.315735	-0.778888	-0.309147	
25%	-0.068536	-0.027407	-0.035304	-0.046713	-0.043070	
50%	-0.021526	0.017179	0.010145	-0.001239	-0.002021	
75%	0.029308	0.065501	0.055105	0.042980	0.047752	
max	0.382667	0.609672	0.436725	0.462729	0.621872	

	X10	X11	X12
count	1000.000000	1000.000000	1000.000000
mean	0.014929	0.011182	-0.000589
std	0.088945	0.081449	0.088311
min	-0.372467	-0.323966	-0.498754
25%	-0.029742	-0.032305	-0.045552
50%	0.011426	0.008911	-0.003157
75%	0.055933	0.053570	0.045610
max	0.649306	0.445628	0.578011





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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.039762 -0.121160  0.046815  0.038100 -0.014330  0.020466  0.002374
1    0.089078  0.085529  0.164658  0.058391  0.093552  0.069839  0.027884
2   -0.034468 -0.076228 -0.046261 -0.092294 -0.142525 -0.153825  0.036229
4   -0.059767 -0.087097  0.175296  0.232745  0.076060 -0.078998 -0.031112
7   -0.028080  0.072486  0.037975 -0.109441 -0.025264 -0.039056 -0.006822
..        …         …         …         …         …         …         …
995 -0.034737 -0.176673 -0.017216  0.026185  0.070100 -0.136641 -0.021024
996  0.110527 -0.027128  0.102080  0.085090  0.053182  0.009982  0.085851
997  0.118052  0.085712  0.015242  0.030000  0.086867  0.070415  0.077697
998  0.034211 -0.163536 -0.013067  0.001532 -0.039201  0.043288  0.050365
999 -0.001770  0.007376  0.038181  0.030045  0.039600 -0.006506 -0.001123

           X7        X8        X9       X10       X11       X12        X13
0   -0.069621 -0.018454  0.101224  0.080683 -0.023190 -0.148151  efectores
1   -0.010880  0.061180  0.018912  0.004052  0.108526  0.096870  efectores
2    0.218504  0.133326  0.048071  0.036323 -0.138014 -0.080164  efectores
4    0.085351 -0.005391 -0.102273 -0.140827  0.050743  0.102831  efectores
7   -0.008190 -0.018578  0.022364  0.076125  0.003091 -0.010763  efectores
..        …         …         …         …         …         …         …
995  0.083219 -0.029190 -0.125920 -0.020569  0.050607 -0.089524  efectores
```

```
996  0.093861  0.033035 -0.003754  0.011807  0.049906 -0.030605  efectores
997  0.077952 -0.023559 -0.003262  0.010419  0.011495 -0.028471  efectores
998  0.057132 -0.018085 -0.155163 -0.079608  0.026331  0.029440  efectores
999  0.103807 -0.023006  0.020405  0.024045  0.044606  0.034815  efectores

[932 rows x 14 columns]
```

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

	X0	X1	X2	X3	X4	X5
count	932.000000	932.000000	932.000000	932.000000	932.000000	932.000000
mean	0.026318	-0.024544	0.043398	0.032218	-0.016273	-0.015758
std	0.083919	0.093940	0.078590	0.084156	0.092244	0.079788
min	-0.242745	-0.318946	-0.198823	-0.241831	-0.318275	-0.258627
25%	-0.026109	-0.093815	-0.009751	-0.011057	-0.075999	-0.069795
50%	0.028812	-0.013712	0.035843	0.034348	-0.009261	-0.014799
75%	0.082346	0.047602	0.094139	0.080233	0.048716	0.034399
max	0.291803	0.251843	0.299548	0.301632	0.262869	0.227382

	X6	X7	X8	X9	X10	X11
count	932.000000	932.000000	932.000000	932.000000	932.000000	932.000000
mean	0.026473	0.022553	0.000355	-0.000367	0.015654	0.004779
std	0.080059	0.077437	0.081728	0.074975	0.074410	0.076861
min	-0.220771	-0.213470	-0.262118	-0.254814	-0.218614	-0.221274
25%	-0.022566	-0.026256	-0.046588	-0.042480	-0.029425	-0.042436
50%	0.022219	0.021283	0.004410	0.000961	0.006252	0.002087
75%	0.072996	0.067567	0.045498	0.042381	0.061244	0.051109
max	0.286704	0.243770	0.230265	0.227843	0.251445	0.236755

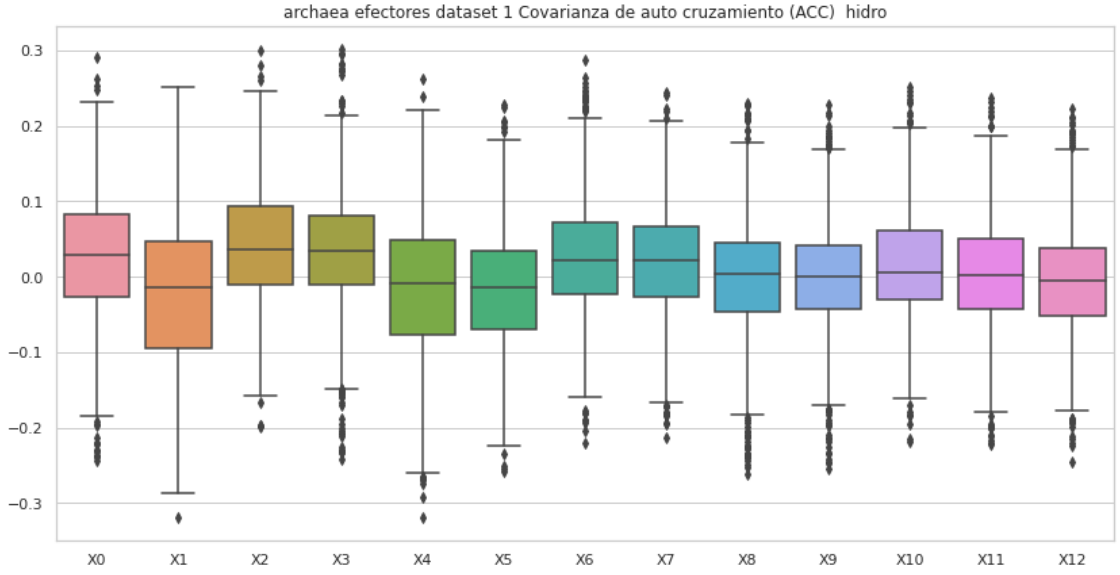
	X12
count	932.000000
mean	-0.006139
std	0.074437
min	-0.244797
25%	-0.050872
50%	-0.004519
75%	0.037506
max	0.222688

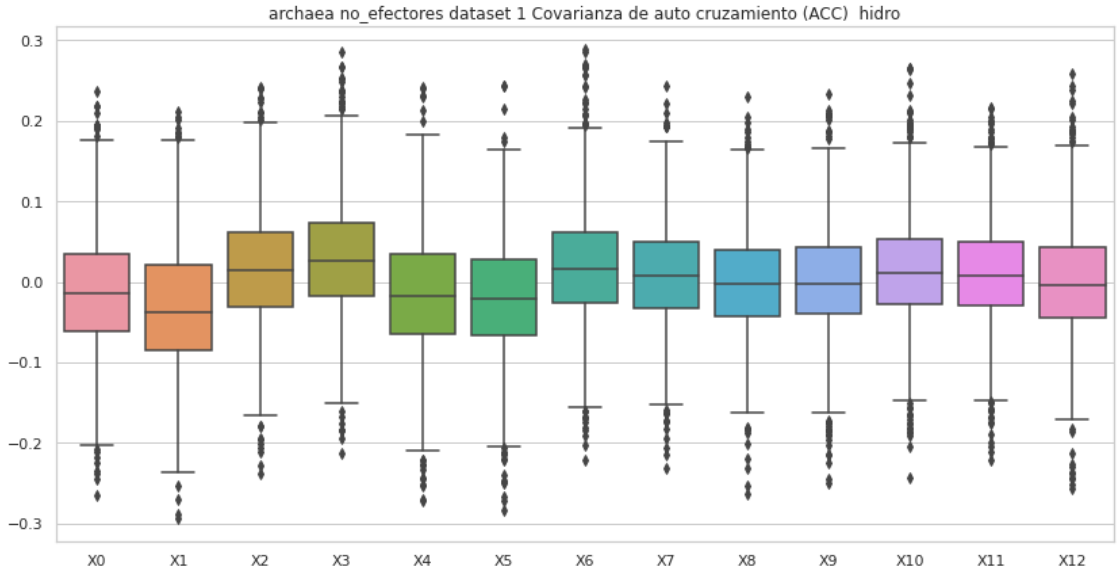
no_efectores
Covarianza de auto cruzamiento (ACC) no_efectores archaea dataset 1, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.142877  0.050328  0.043193  0.074674  0.157371 -0.017938  0.056920
1   -0.049283 -0.150823 -0.014428  0.017522 -0.025395 -0.031961  0.075864
2   -0.055686 -0.041942  0.052189  0.151169 -0.053262 -0.072608  0.068049
3    0.020190 -0.033354  0.059725  0.027032 -0.151143 -0.047166 -0.011113
5    0.025374  0.026980 -0.009436 -0.006427  0.001437 -0.027924 -0.032427
..        …         …         …         …         …         …
994 -0.005614 -0.047872  0.065495 -0.105835 -0.121761 -0.068208 -0.076136
996 -0.019117 -0.059150  0.065125 -0.091222  0.114946  0.082187 -0.001551
997  0.018037 -0.025500 -0.058443 -0.008068  0.042766 -0.027799 -0.043659
998 -0.069709 -0.062702 -0.031853  0.074894 -0.069951  0.003910 -0.049446
999 -0.029378 -0.119047  0.241290 -0.012778  0.084675 -0.004010  0.091379

           X7        X8        X9       X10       X11       X12           X13
0    0.151663 -0.011647  0.133410  0.232357  0.095623  0.017287  no_efectores
1    0.024849 -0.072310  0.023962 -0.014032  0.006031 -0.053492  no_efectores
2    0.055371  0.018979 -0.079546 -0.000960 -0.045562  0.026857  no_efectores
3    0.005142 -0.072586 -0.035991 -0.048064 -0.053641 -0.011660  no_efectores
5   -0.035429 -0.052702 -0.087322 -0.084816  0.012965  0.015266  no_efectores
..        …         …         …         …         …         …
994 -0.069623  0.061718  0.135764  0.019396  0.067492 -0.130917  no_efectores
996 -0.024768  0.049299  0.005893  0.054478  0.022979 -0.031294  no_efectores
997 -0.012401  0.023211  0.056428 -0.016406 -0.033675  0.050311  no_efectores
998 -0.089603  0.028885  0.102730  0.014952  0.013084  0.080266  no_efectores
999 -0.011442  0.028730  0.200693  0.112019  0.014180  0.238867  no_efectores

[905 rows x 14 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  905.000000  905.000000  905.000000  905.000000  905.000000  905.000000
mean    -0.011466   -0.033495    0.014439    0.027861   -0.014919   -0.020482
std      0.077019    0.082399    0.073106    0.076233    0.078993    0.074208
min     -0.264381   -0.293606   -0.237688   -0.212300   -0.272238   -0.284064
25%     -0.060901   -0.084873   -0.030391   -0.016720   -0.063828   -0.065998
50%     -0.013594   -0.038492    0.014304    0.026830   -0.017227   -0.020890
75%      0.035337    0.020652    0.061249    0.073120    0.035399    0.027168
max      0.236639    0.211776    0.241290    0.285844    0.241633    0.242922

                X6          X7          X8          X9         X10         X11  \
count  905.000000  905.000000  905.000000  905.000000  905.000000  905.000000
mean     0.017443    0.007302   -0.000781    0.001067    0.012741    0.008805
std      0.073803    0.068245    0.069483    0.072375    0.070666    0.068690
```

min	-0.220503	-0.230680	-0.263864	-0.249831	-0.243430	-0.221501
25%	-0.026260	-0.032483	-0.042731	-0.040002	-0.027987	-0.029733
50%	0.015382	0.008283	-0.001486	-0.002385	0.010918	0.008420
75%	0.061088	0.050702	0.040138	0.042867	0.052901	0.049460
max	0.288858	0.243390	0.229944	0.233495	0.265642	0.216969

	X12
count	905.000000
mean	-0.001415
std	0.072531
min	-0.256747
25%	-0.044134
50%	-0.003420
75%	0.042620
max	0.258298




archaea no_efectores dataset 1 Covarianza de auto cruzamiento (ACC) hidro
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12