ds2_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 = 2
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 2, con valores
atípicos.
Valores del documento csv.

         X0     X1     X2     X3     X4      X5     X6      X7     X8      X9  \
0     9.744  6.154  1.538  3.077  0.513   2.051  1.538   5.641  2.051   4.103
1    14.068  2.662  1.141  5.323  0.000   4.183  2.281   9.125  0.000   5.703
2    10.795  6.250  1.420  4.830  0.000   6.534  1.420   7.386  0.568   3.977
3    10.825  4.639  2.577  3.093  1.546   2.577  0.000   6.701  3.608   9.278
4     5.128  6.410  3.846  7.692  0.641  11.538  0.641   3.846  0.641   6.410
..      ...    ...    ...    ...    ...     ...    ...     ...    ...     ...
995   6.040  3.356  4.362  3.020  0.671   3.691  2.349   4.362  1.007  14.094
996  10.256  4.103  1.538  1.538  1.538   1.538  2.051  10.769  4.103   7.692
997  16.114  7.109  0.474  1.896  0.474   7.583  0.948   8.057  2.370   0.474
998   2.510  2.510  5.021  6.276  0.418   5.858  2.092   6.695  3.347  13.389
999  16.327  4.898  2.041  0.816  0.000   4.490  1.633   8.980  0.408   2.449
```

```
     …     X11    X12    X13    X14    X15    X16    X17    X18     X19  \
0    …   1.026  2.051  6.154  8.718  7.692  4.103  1.538  5.128  12.308
1    …   1.141  1.901  5.323  3.802  3.422  4.943  1.521  3.802  14.068
2    …   0.000  0.852  4.261  4.545  7.670  6.818  1.420  2.841  14.489
3    …   1.031  4.639  9.278  3.093  7.216  2.577  0.000  7.216   5.670
4    …   8.974  4.487  1.282  2.564  7.692  3.205  0.641  3.205   7.051
..   …     …    …    …    …    …    …    …    …     …
995  …   6.376  1.342  9.732  2.685  6.711  1.678  1.007  7.047   7.718
996  …   0.000  1.538  4.103  5.641  5.641  8.205  1.538  1.538   8.718
997  …   0.474  0.948  4.739  7.583  6.161  2.844  2.844  3.791  11.374
998  …  11.297  0.837  7.113  2.929  4.603  4.184  0.837  4.603   4.603
999  …   0.408  0.816  3.265  5.306  4.082  3.265  1.633  4.082  16.735

            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 2, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      9.388283     5.991061     2.736202     5.876961     0.681756
std       4.539442     2.715770     2.189533     2.720898     0.954970
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       5.830500     4.173750     1.139750     3.670000     0.000000
50%       8.925500     5.714000     2.244000     5.462000     0.341500
75%      12.277750     7.692000     3.925750     7.465750     0.981000
max      24.843000    16.867000    14.815000    16.149000     6.557000

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      7.090872     2.296036     7.523201     1.801976     5.793005
std       3.922801     1.597126     2.845401     1.205634     3.367724
```

min	0.452000	0.000000	0.000000	0.000000	0.000000
25%	3.960750	1.248000	5.371250	0.820750	3.204750
50%	7.038000	2.015500	7.402500	1.710500	5.183500
75%	9.602750	3.052750	9.382250	2.610250	7.775500
max	49.091000	11.194000	17.453000	10.000000	18.841000

	X10	X11	X12	X13	X14 \
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	11.346235	4.455205	2.083967	3.968835	4.220368
std	3.333506	4.328307	1.275947	2.017317	1.826923
min	0.943000	0.000000	0.000000	0.000000	0.000000
25%	8.943000	1.052250	1.129750	2.612750	2.985000
50%	11.384000	2.446000	1.818000	3.857000	4.098500
75%	13.688000	7.514000	2.667000	5.012750	5.310750
max	21.739000	20.690000	9.375000	17.105000	14.286000

	X15	X16	X17	X18	X19
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	5.945905	5.475645	1.275196	3.371392	8.677816
std	2.252498	2.075446	1.048246	1.670689	3.495097
min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	4.465250	4.124000	0.547500	2.261500	5.970000
50%	5.740000	5.392500	1.149000	3.205000	8.273500
75%	7.243000	6.603000	1.778250	4.217000	11.239500
max	17.347000	15.054000	6.989000	11.111000	18.947000

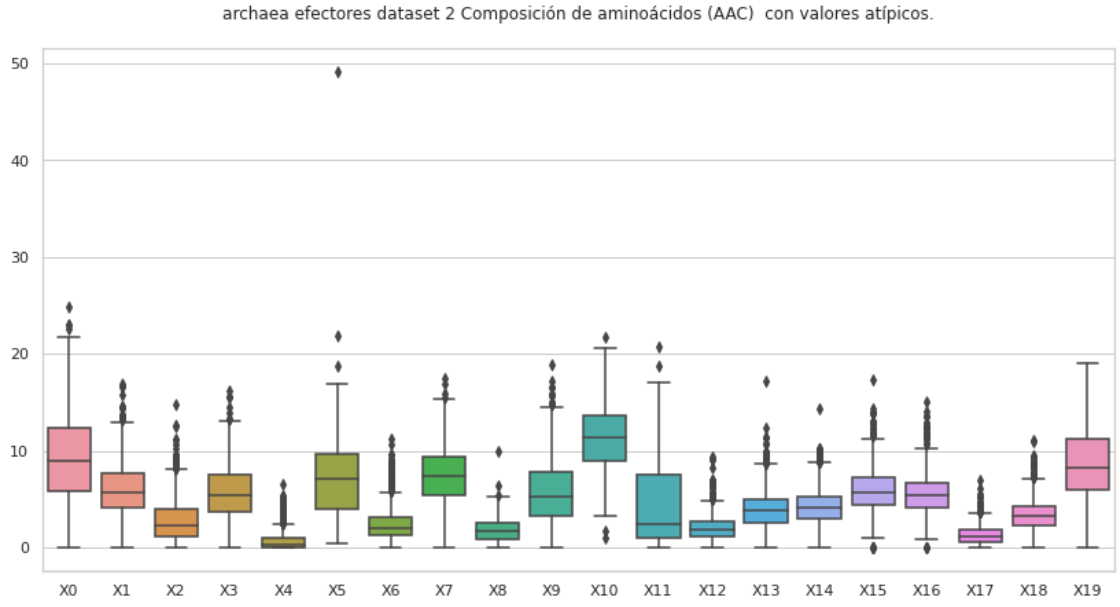
no_efectores
Composición de aminoácidos (AAC) no_efectores archaea dataset 2, con valores atípicos.
Valores del documento csv.

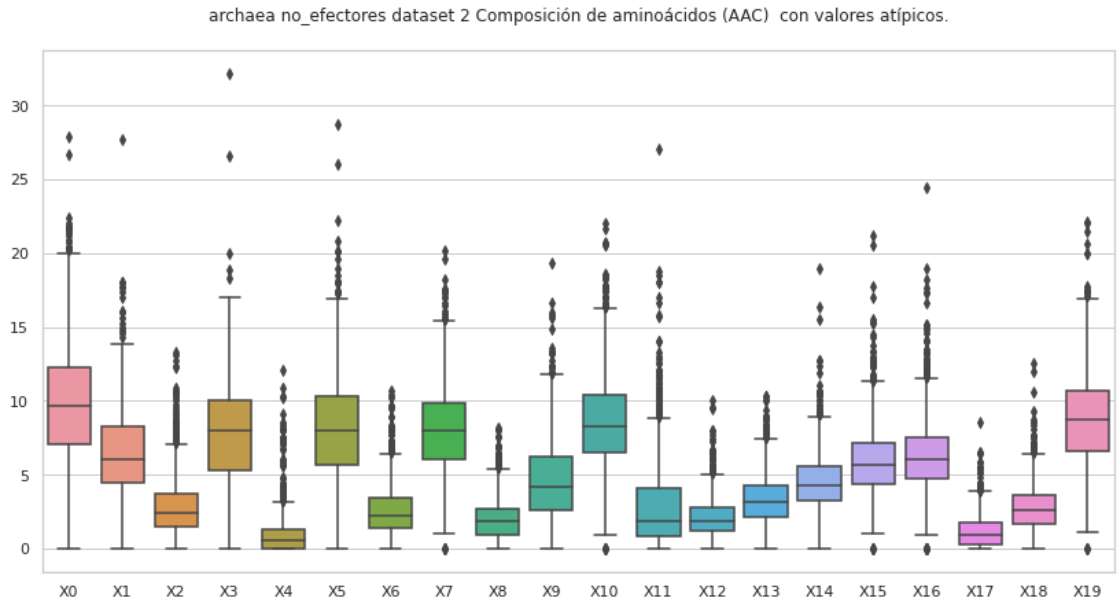
	X0	X1	X2	X3	X4	X5	X6	X7	X8 \
0	6.612	4.132	3.636	4.463	0.992	5.620	1.653	8.595	1.983
1	14.286	12.500	0.000	8.929	1.786	5.357	0.000	8.929	3.571
2	15.618	3.905	2.386	3.471	0.868	4.555	3.037	8.026	1.735
3	9.180	0.328	8.197	11.148	0.000	5.246	2.623	9.836	0.328
4	9.341	5.495	2.198	3.846	0.549	3.297	2.198	7.692	0.549
..	…	…	…	…	…	…	…	…	…
995	5.310	2.655	4.425	3.540	0.000	14.159	2.655	7.965	6.195
996	13.147	6.773	2.390	7.371	0.199	12.550	1.594	8.566	1.394
997	4.494	11.236	0.000	11.236	0.000	8.989	5.618	8.989	1.124
998	1.600	4.800	9.600	6.400	0.800	4.800	3.200	1.600	0.800
999	17.143	1.429	0.714	5.714	0.000	8.571	1.429	13.571	0.714

	X9	…	X11	X12	X13	X14	X15	X16	X17	X18 \
0	8.760	…	2.975	4.628	5.785	3.306	9.587	4.959	0.661	1.488

```
1     0.000  …   1.786  1.786  1.786  0.000  14.286   7.143  0.000  0.000
2     7.375  …   1.518  0.868  4.989  3.471   5.206   5.640  2.386  2.386
3     5.246  …   0.656  0.000  4.262  1.639  13.770  11.803  0.000  1.311
4     7.692  …   2.198  2.198  2.747  4.945   7.143   4.945  2.198  4.396
..      …  …     …    …    …    …      …     …    …    …
995   4.425  …   0.000  2.655  4.425  9.735   8.850   6.195  0.885  1.770
996   4.183  …   2.390  2.191  1.793  5.179   5.976   6.574  0.199  1.195
997   4.494  …   2.247  3.371  2.247  4.494   7.865   5.618  0.000  1.124
998  10.400  …  12.800  1.600  7.200  4.800   4.800  10.400  0.000  1.600
999   5.000  …   1.429  3.571  6.429  5.000   6.429   5.714  0.000  0.000

        X19            X20
0     9.421   no_efectores
1    10.714   no_efectores
2     9.328   no_efectores
3     9.508   no_efectores
4    11.538   no_efectores
..      …            …
995  10.619   no_efectores
996  10.956   no_efectores
997   6.742   no_efectores
998   4.800   no_efectores
999  10.000   no_efectores

[1000 rows x 21 columns]


Composición de aminoácidos (AAC) no_efectores archaea dataset 2, con valores
atípicos.
Estadísticas.

                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      9.821385     6.427476     2.808316     7.809643     1.009754
std       4.146631     3.012274     2.049276     3.454087     1.502877
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       7.035500     4.450000     1.459000     5.315000     0.000000
50%       9.709500     6.081000     2.382500     7.956500     0.601500
75%      12.259500     8.246750     3.690500    10.000000     1.283500
max      27.907000    27.660000    13.333000    32.143000    12.069000

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      8.119545     2.533834     8.044250     1.957467     4.674055
std       3.730139     1.745558     2.981488     1.422887     2.908685
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       5.706000     1.370000     6.008250     0.934250     2.564000
50%       8.030500     2.252000     7.991000     1.814000     4.167000
```

```
75%      10.291500     3.426250     9.806250     2.719250     6.256500
max      28.696000    10.695000    20.139000     8.163000    19.298000

                X10          X11          X12          X13          X14  \
count   1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean       8.690738     2.989865     2.194341     3.366549     4.477421
std        3.403554     3.175181     1.364341     1.890369     2.051766
min        0.000000     0.000000     0.000000     0.000000     0.000000
25%        6.522000     0.853000     1.244750     2.131750     3.295750
50%        8.260000     1.895000     1.887000     3.175000     4.310000
75%       10.435000     4.085250     2.766500     4.286000     5.556000
max       22.034000    27.027000    10.000000    10.345000    18.966000

                X15          X16          X17          X18          X19
count   1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean       5.962261     6.344573     1.144604     2.788839     8.835055
std        2.476274     2.597943     1.136318     1.624518     3.276260
min        0.000000     0.000000     0.000000     0.000000     0.000000
25%        4.348000     4.754250     0.267000     1.695000     6.557000
50%        5.678500     6.073500     0.889000     2.641000     8.696000
75%        7.143000     7.487000     1.724000     3.593750    10.704250
max       21.154000    24.444000     8.511000    12.500000    22.093000
```





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 2, sin valores
atípicos.
Valores del documento csv.

         X0     X1     X2     X3     X4      X5     X6      X7     X8      X9  \
0     9.744  6.154  1.538  3.077  0.513   2.051  1.538   5.641  2.051   4.103
1    14.068  2.662  1.141  5.323  0.000   4.183  2.281   9.125  0.000   5.703
2    10.795  6.250  1.420  4.830  0.000   6.534  1.420   7.386  0.568   3.977
3    10.825  4.639  2.577  3.093  1.546   2.577  0.000   6.701  3.608   9.278
4     5.128  6.410  3.846  7.692  0.641  11.538  0.641   3.846  0.641   6.410
..      …      …      …      …      …       …      …       …      …       …
995   6.040  3.356  4.362  3.020  0.671   3.691  2.349   4.362  1.007  14.094
996  10.256  4.103  1.538  1.538  1.538   1.538  2.051  10.769  4.103   7.692
997  16.114  7.109  0.474  1.896  0.474   7.583  0.948   8.057  2.370   0.474
998   2.510  2.510  5.021  6.276  0.418   5.858  2.092   6.695  3.347  13.389
999  16.327  4.898  2.041  0.816  0.000   4.490  1.633   8.980  0.408   2.449

     …     X11    X12    X13    X14    X15    X16    X17    X18     X19  \
0    …   1.026  2.051  6.154  8.718  7.692  4.103  1.538  5.128  12.308
1    …   1.141  1.901  5.323  3.802  3.422  4.943  1.521  3.802  14.068
2    …   0.000  0.852  4.261  4.545  7.670  6.818  1.420  2.841  14.489
3    …   1.031  4.639  9.278  3.093  7.216  2.577  0.000  7.216   5.670
4    …   8.974  4.487  1.282  2.564  7.692  3.205  0.641  3.205   7.051
..   …      …      …      …      …      …      …      …      …       …
995  …   6.376  1.342  9.732  2.685  6.711  1.678  1.007  7.047   7.718
```

```
996  …   0.000  1.538  4.103  5.641  5.641  8.205  1.538  1.538   8.718
997  …   0.474  0.948  4.739  7.583  6.161  2.844  2.844  3.791  11.374
998  …  11.297  0.837  7.113  2.929  4.603  4.184  0.837  4.603   4.603
999  …   0.408  0.816  3.265  5.306  4.082  3.265  1.633  4.082  16.735

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

[868 rows x 21 columns]
```

Composición de aminoácidos (AAC) efectores archaea dataset 2, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  868.000000  868.000000  868.000000  868.000000  868.000000  868.000000
mean     9.745713    6.077215    2.594006    5.894933    0.560365    7.046325
std      4.478980    2.544731    1.968343    2.670755    0.725551    3.626940
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.452000
25%      6.214000    4.261750    1.124000    3.674500    0.000000    3.828000
50%      9.472000    5.849500    2.145000    5.477500    0.299000    7.133500
75%     12.873250    7.803500    3.780250    7.500000    0.862000    9.548750
max     22.977000   13.830000    9.244000   13.846000    3.448000   18.785000

                X6          X7          X8          X9         X10         X11  \
count  868.000000  868.000000  868.000000  868.000000  868.000000  868.000000
mean     2.175639    7.659039    1.819866    5.647392   11.504399    4.262575
std      1.307240    2.813850    1.153977    3.197497    3.219809    4.239073
min      0.000000    0.654000    0.000000    0.000000    3.401000    0.000000
25%      1.250000    5.525000    0.847000    3.196250    9.091000    0.976000
50%      1.990000    7.692000    1.734000    5.058000   11.615000    2.279500
75%      2.977500    9.524000    2.632000    7.595000   13.745500    7.248500
max      6.977000   15.942000    5.263000   14.734000   20.619000   17.021000

               X12         X13         X14         X15         X16         X17  \
count  868.000000  868.000000  868.000000  868.000000  868.000000  868.000000
mean     1.978179    3.997961    4.201523    5.845031    5.435785    1.277044
```

std	1.119395	1.805593	1.734136	2.002912	1.847657	0.974751
min	0.000000	0.000000	0.000000	0.000000	1.010000	0.000000
25%	1.093000	2.854500	3.003000	4.478000	4.181750	0.584000
50%	1.778000	3.922000	4.098500	5.736000	5.426000	1.171000
75%	2.574000	5.012750	5.285750	7.143000	6.579000	1.783750
max	5.660000	9.831000	9.605000	12.389000	11.562000	4.255000

	X18	X19
count	868.000000	868.000000
mean	3.312780	8.964134
std	1.514553	3.489050
min	0.000000	0.000000
25%	2.271750	6.279750
50%	3.208000	8.718000
75%	4.167000	11.505000
max	8.380000	18.947000

no_efectores
Composición de aminoácidos (AAC) no_efectores archaea dataset 2, sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	X7	X8	\
0	6.612	4.132	3.636	4.463	0.992	5.620	1.653	8.595	1.983	
2	15.618	3.905	2.386	3.471	0.868	4.555	3.037	8.026	1.735	
4	9.341	5.495	2.198	3.846	0.549	3.297	2.198	7.692	0.549	
5	3.672	3.240	3.672	6.263	0.216	6.695	1.944	8.855	2.376	
6	5.730	3.312	5.819	5.730	1.074	3.402	2.596	9.669	1.791	
..	…	…	…	…	…	…	…	…	…	
994	9.632	4.533	1.983	7.082	2.550	10.765	3.116	11.048	3.116	
995	5.310	2.655	4.425	3.540	0.000	14.159	2.655	7.965	6.195	
996	13.147	6.773	2.390	7.371	0.199	12.550	1.594	8.566	1.394	
997	4.494	11.236	0.000	11.236	0.000	8.989	5.618	8.989	1.124	
999	17.143	1.429	0.714	5.714	0.000	8.571	1.429	13.571	0.714	

	X9	…	X11	X12	X13	X14	X15	X16	X17	X18	\
0	8.760	…	2.975	4.628	5.785	3.306	9.587	4.959	0.661	1.488	
2	7.375	…	1.518	0.868	4.989	3.471	5.206	5.640	2.386	2.386	
4	7.692	…	2.198	2.198	2.747	4.945	7.143	4.945	2.198	4.396	
5	11.231	…	7.991	3.672	3.888	5.400	6.263	5.400	0.648	5.400	
6	5.819	…	3.044	1.253	4.834	3.671	10.474	11.638	2.417	4.745	
..	…	…	…	…	…	…	…	…	…	…	
994	5.382	…	1.133	1.983	2.833	5.099	4.816	4.816	1.416	1.416	
995	4.425	…	0.000	2.655	4.425	9.735	8.850	6.195	0.885	1.770	
996	4.183	…	2.390	2.191	1.793	5.179	5.976	6.574	0.199	1.195	
997	4.494	…	2.247	3.371	2.247	4.494	7.865	5.618	0.000	1.124	

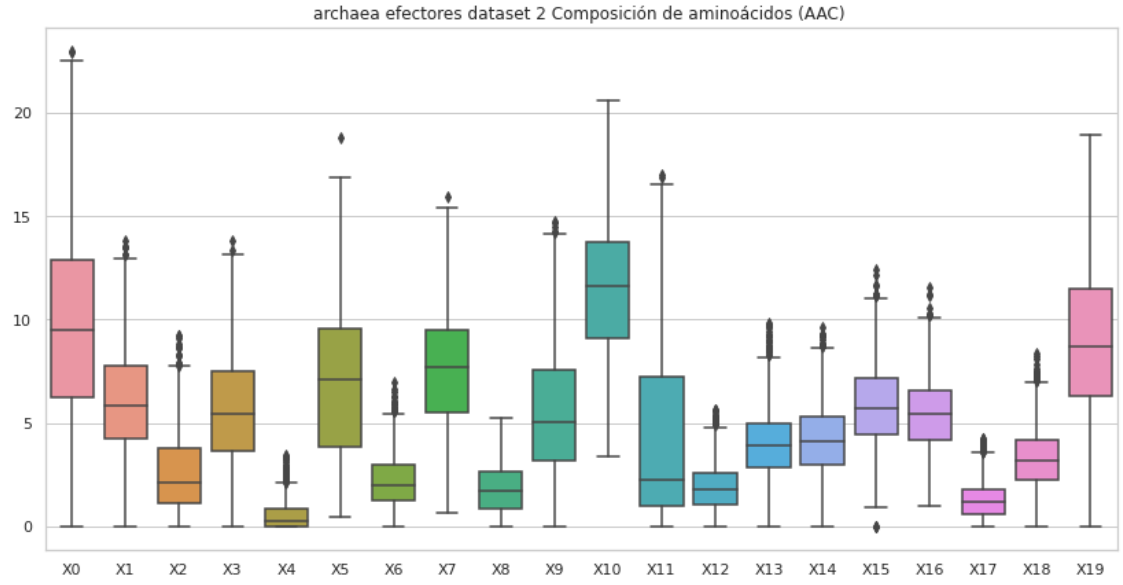
```
999   5.000  …  1.429  3.571  6.429  5.000   6.429   5.714  0.000  0.000

        X19           X20
0     9.421  no_efectores
2     9.328  no_efectores
4    11.538  no_efectores
5     6.695  no_efectores
6     6.893  no_efectores
..      …           …
994   9.348  no_efectores
995  10.619  no_efectores
996  10.956  no_efectores
997   6.742  no_efectores
999  10.000  no_efectores

[840 rows x 21 columns]
```

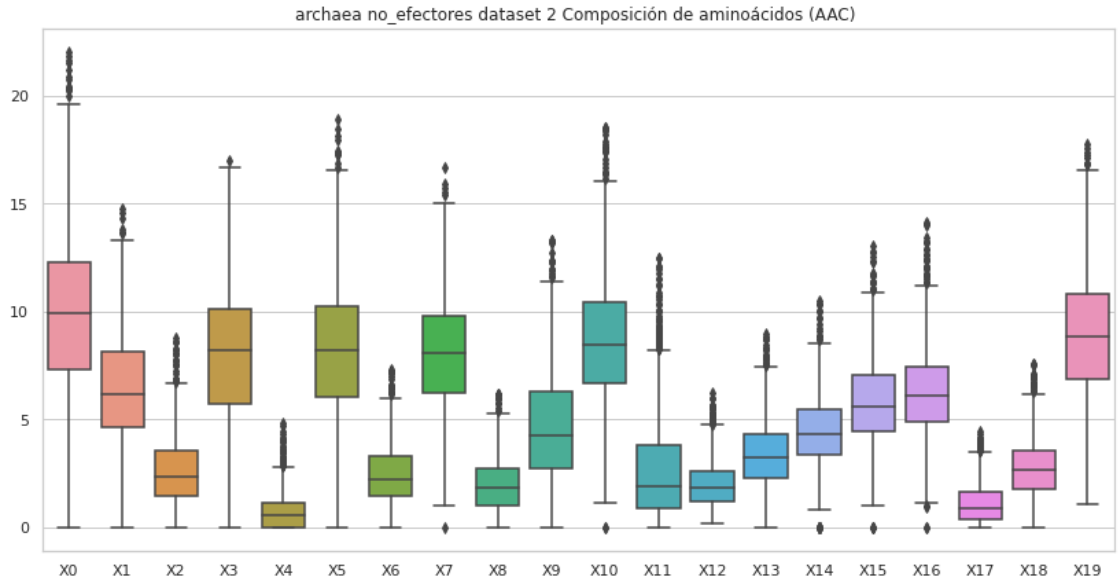
Composición de aminoácidos (AAC) no_efectores archaea dataset 2, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  840.000000  840.000000  840.000000  840.000000  840.000000  840.000000
mean    10.009806    6.436877    2.668132    7.949104    0.792120    8.193194
std      3.807951    2.611548    1.669650    3.181972    0.886392    3.407703
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      7.293500    4.630750    1.463750    5.710250    0.000000    6.047000
50%      9.932500    6.176500    2.387500    8.197000    0.575000    8.185500
75%     12.264000    8.163000    3.571000   10.106500    1.136750   10.268000
max     22.026000   14.773000    8.777000   16.981000    4.808000   18.935000

                X6          X7          X8          X9         X10         X11  \
count  840.000000  840.000000  840.000000  840.000000  840.000000  840.000000
mean     2.479088    8.112352    1.958492    4.729979    8.829605    2.803425
std      1.468840    2.638414    1.269088    2.677931    3.118602    2.669269
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      1.457250    6.244000    1.015000    2.759750    6.667000    0.880000
50%      2.260500    8.057500    1.839500    4.268000    8.459500    1.912000
75%      3.327750    9.797000    2.746750    6.295250   10.448000    3.820500
max      7.339000   16.667000    6.197000   13.333000   18.548000   12.500000

               X12         X13         X14         X15         X16         X17  \
count  840.000000  840.000000  840.000000  840.000000  840.000000  840.000000
mean     2.056410    3.423199    4.479454    5.858969    6.317520    1.100961
std      1.164631    1.743041    1.710352    2.125146    2.156858    0.955390
min      0.199000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      1.209250    2.273000    3.399250    4.428750    4.886250    0.388750
```

50%	1.826500	3.226000	4.348000	5.616500	6.135000	0.901000
75%	2.632000	4.358750	5.455000	7.050000	7.429500	1.657000
max	6.250000	8.978000	10.484000	13.043000	14.124000	4.444000

	X18	X19
count	840.000000	840.000000
mean	2.788706	9.012543
std	1.413735	2.966334
min	0.000000	1.099000
25%	1.758500	6.896000
50%	2.703000	8.871500
75%	3.571000	10.849000
max	7.584000	17.757000





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
2, con valores atípicos.
Valores del documento csv.
```

	X0	X1	X2	X3	X4	X5	X6	\
0	0.029208	0.001537	0.009224	0.006149	0.018447	0.016910	0.006149	
1	0.019266	0.000000	0.007290	0.005728	0.007290	0.012497	0.000000	
2	0.022130	0.000000	0.009900	0.013395	0.008736	0.015142	0.001165	
3	0.029458	0.004208	0.008417	0.007014	0.025250	0.018236	0.009819	
4	0.037163	0.004645	0.055745	0.083617	0.009291	0.027873	0.004645	
..	…	…	…	…	…	…	…	
995	0.016932	0.001881	0.008466	0.010347	0.027279	0.012229	0.002822	
996	0.017296	0.002594	0.002594	0.002594	0.006919	0.018161	0.006919	
997	0.041126	0.001210	0.004838	0.019354	0.012096	0.020563	0.006048	
998	0.025290	0.004215	0.063224	0.059009	0.071654	0.067439	0.033720	
999	0.023083	0.000000	0.001154	0.006348	0.004617	0.012695	0.000577	

	X7	X8	X9	…	X74	X75	X76	\
0	0.012298	0.003075	0.044581	…	0.014407	0.000241	-0.003610	
1	0.007811	0.001562	0.021349	…	0.019046	0.002514	0.013814	
2	0.008153	0.000000	0.028537	…	0.021313	0.003590	0.018538	
3	0.025250	0.002806	0.039277	…	0.024276	0.004543	-0.003502	
4	0.046454	0.065036	0.102199	…	0.006080	0.023937	-0.021472	
..	…	…	…	…	…	…	…	
995	0.039508	0.017873	0.035745	…	0.000823	-0.003767	0.001644	
996	0.012972	0.000000	0.030269	…	0.011273	0.002692	0.017193	
997	0.001210	0.001210	0.035078	…	0.002746	-0.001241	0.015066	
998	0.134879	0.113804	0.109589	…	-0.022002	-0.086485	-0.011014	
999	0.003462	0.000577	0.025968	…	0.020497	0.001538	0.012265	

	X77	X78	X79	X80	X81	X82	X83
0	0.014488	0.004276	-0.010369	0.015479	0.004256	0.014991	efectores
1	0.027512	0.008055	0.010009	0.020183	-0.002058	0.003396	efectores
2	0.010589	-0.002200	0.023844	0.017372	0.003816	0.017261	efectores
3	0.016830	-0.002598	0.016586	0.005648	-0.000403	0.007535	efectores
4	-0.057620	0.006883	-0.006780	-0.035410	-0.001048	0.010215	efectores
..	…	…	…	…	…	…	
995	0.017368	0.012238	0.000062	0.021156	-0.001376	-0.005957	efectores
996	0.019029	0.004450	0.014508	0.012719	0.004687	0.016268	efectores

```
997  0.022813  0.006532  0.035620  0.012692  0.002235  0.024560  efectores
998  0.029531  0.032603  0.037754 -0.031086 -0.017250 -0.037757  efectores
999  0.022021  0.005453  0.005946  0.009990 -0.001091  0.011137  efectores

[1000 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro_mass efectores archaea dataset 2, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.033468     0.003694     0.027151     0.038140     0.018806
std       0.021982     0.006501     0.025776     0.065834     0.054156
min       0.000000     0.000000     0.000000     0.000694     0.000000
25%       0.021662     0.000000     0.009421     0.008585     0.007475
50%       0.029708     0.000885     0.021014     0.027228     0.012567
75%       0.041132     0.004606     0.038445     0.053530     0.020813
max       0.448072     0.056794     0.314333     1.257331     1.571664

                X5           X6           X7           X8           X9  …  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
mean      0.030140     0.008419     0.031913     0.029841     0.051044  …
std       0.052374     0.010529     0.073730     0.074696     0.080296  …
min       0.000000     0.000000     0.000000     0.000000     0.001727  …
25%       0.017234     0.002313     0.007759     0.002288     0.024994  …
50%       0.025016     0.006094     0.017310     0.009163     0.038573  …
75%       0.035344     0.010775     0.038395     0.040522     0.058260  …
max       1.571664     0.149357     1.885997     1.257331     1.572090  …

               X73          X74          X75          X76          X77  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.012086     0.002989     0.006899     0.012576     0.005116
std       0.050773     0.048234     0.045949     0.024458     0.032509
min      -1.397925    -0.520899    -1.041572    -0.285842    -0.329305
25%       0.004184    -0.007185    -0.001597     0.003155    -0.004215
50%       0.014908     0.007671     0.004545     0.015848     0.010338
75%       0.024706     0.016184     0.016896     0.024239     0.019066
max       0.126454     0.764896     0.208495     0.105641     0.159605

               X78          X79          X80          X81          X82
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.008302     0.013554     0.001933     0.007389     0.012684
std       0.041865     0.029736     0.050355     0.064749     0.036768
min      -0.601208    -0.362373    -0.761900    -0.854787    -0.943676
25%      -0.000330     0.005014    -0.005052    -0.000770     0.004339
50%       0.006206     0.015415     0.010836     0.005455     0.014995
```

```
75%       0.016662     0.025266     0.018899     0.017018     0.023780
max       0.719197     0.471552     0.163691     1.352774     0.104955

[8 rows x 83 columns]
```

```
no_efectores
Composición de pseudo aminoácidos (PseAAC)  hidro_mass no_efectores archaea
dataset 2, con valores atípicos.
Valores del documento csv.
```

	X0	X1	X2	X3	X4	X5	X6	\
0	0.015828	0.002374	0.010684	0.013454	0.013849	0.020576	0.004748	
1	0.032574	0.004072	0.020359	0.012215	0.004072	0.020359	0.008143	
2	0.033896	0.001883	0.007532	0.009886	0.010828	0.017419	0.003766	
3	0.018434	0.000000	0.022384	0.010534	0.008559	0.019751	0.000658	
4	0.022749	0.001338	0.009367	0.008029	0.006691	0.018735	0.001338	
..	…	…	…	…	…	…	…	
995	0.027029	0.000000	0.018019	0.072077	0.022524	0.040543	0.031534	
996	0.031871	0.000483	0.017867	0.030422	0.004346	0.020764	0.003380	
997	0.014001	0.000000	0.035003	0.028003	0.007001	0.028003	0.003500	
998	0.016249	0.008124	0.064995	0.048746	0.073119	0.016249	0.008124	
999	0.022716	0.000000	0.007572	0.011358	0.008518	0.017983	0.000946	

	X7	X8	X9	…	X74	X75	X76	\
0	0.020972	0.007123	0.025720	…	0.007815	0.005175	0.014828	
1	0.000000	0.004072	0.016287	…	-0.045436	-0.006526	0.018385	
2	0.016006	0.003295	0.028717	…	0.004424	-0.001240	0.020444	
3	0.010534	0.001317	0.009875	…	-0.003717	-0.007761	0.040307	
4	0.018735	0.005353	0.036131	…	0.014780	0.001271	0.009879	
..	…	…	…	…	…	…	…	
995	0.022524	0.000000	0.018019	…	0.009806	0.013011	-0.019928	
996	0.010141	0.005795	0.013038	…	0.007039	0.017690	0.032664	
997	0.014001	0.007001	0.031503	…	0.004765	0.053387	-0.019897	
998	0.105617	0.129990	0.081244	…	-0.031942	-0.074732	-0.015621	
999	0.006625	0.001893	0.009465	…	0.013790	-0.001104	0.030989	

	X77	X78	X79	X80	X81	X82	X83
0	0.013975	0.007454	0.010588	0.015772	0.011753	0.017825	no_efectores
1	-0.041437	-0.009876	0.033724	0.007202	0.010319	0.014286	no_efectores
2	0.017636	0.004162	0.022557	0.018871	0.004344	0.024093	no_efectores
3	0.006427	0.002561	0.035678	0.006098	0.001040	0.038502	no_efectores
4	0.012973	0.005436	0.015523	0.008732	0.000544	0.012553	no_efectores
..	…	…	…	…	…	…	
995	-0.013979	-0.000842	0.025760	-0.000517	0.011657	0.034203	no_efectores
996	-0.000402	0.010093	0.023449	-0.001733	0.009544	0.021929	no_efectores
997	0.006887	0.035050	0.030864	-0.005206	0.020606	-0.002171	no_efectores

```
998  0.010838  0.016893  0.018228 -0.053875 -0.123349  0.038521  no_efectores
999  0.006746 -0.006158  0.031477  0.002493 -0.000101  0.025315  no_efectores

[1000 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores archaea
dataset 2, con valores atípicos.
Estadísticas.

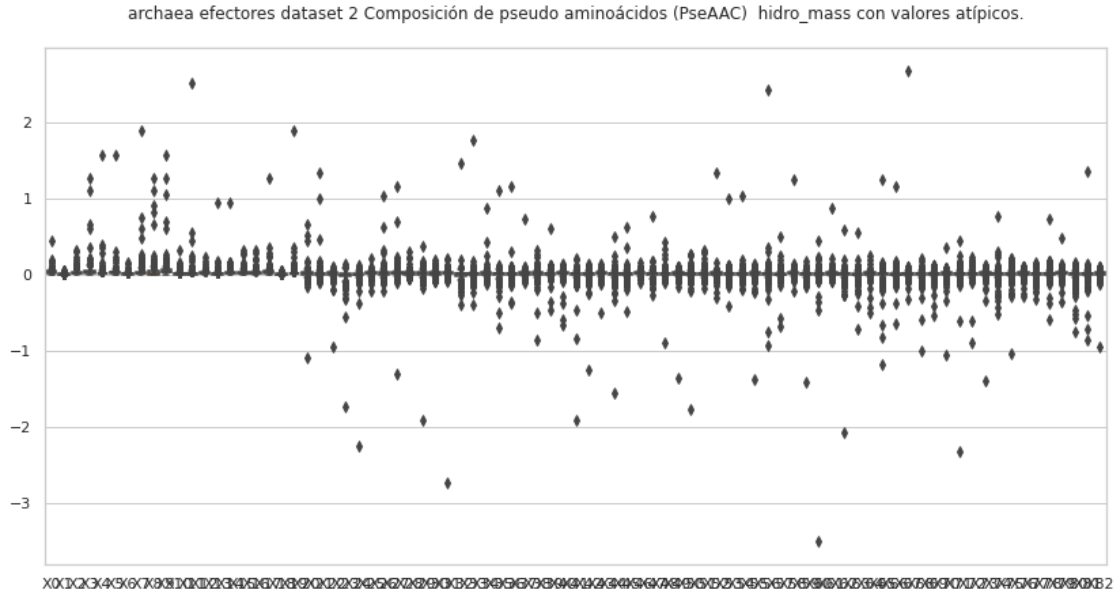
	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.029744	0.005585	0.029932	0.028334	0.008305	
std	0.208887	0.029192	0.125046	0.207081	0.204271	
min	-6.412790	0.000000	-3.206395	-6.412790	-6.412790	
25%	0.023155	0.000000	0.017972	0.016514	0.006008	
50%	0.031866	0.001953	0.028554	0.029919	0.010956	
75%	0.041982	0.005103	0.041385	0.046940	0.017700	
max	1.326461	0.884308	2.210769	0.884308	0.442154	

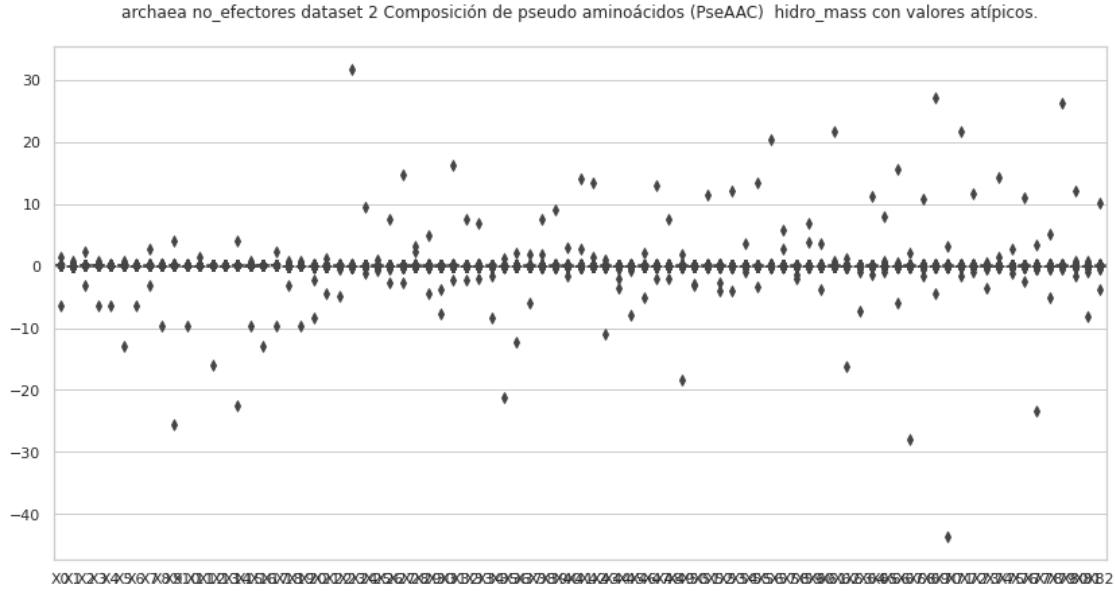
	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.017536	0.002630	0.020354	0.005824	0.012393	…	
std	0.407837	0.203760	0.134250	0.306011	0.822328	…	
min	-12.825580	-6.412790	-3.206395	-9.619185	-25.651161	…	
25%	0.019246	0.002609	0.007342	0.002234	0.019835	…	
50%	0.026769	0.006056	0.013684	0.006537	0.027480	…	
75%	0.035319	0.011558	0.025284	0.016647	0.040824	…	
max	0.884308	0.442154	2.652923	0.442154	3.979384	…	

	X73	X74	X75	X76	X77	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.012706	0.017256	0.008831	0.024374	-0.019199	
std	0.120841	0.452182	0.095443	0.355243	0.748549	
min	-3.675908	-0.294966	-1.077459	-2.494230	-23.395754	
25%	0.007502	-0.007258	-0.001082	0.008336	-0.007290	
50%	0.018630	0.003374	0.006970	0.018123	0.004064	
75%	0.027723	0.013547	0.017569	0.028819	0.014019	
max	0.472783	14.188518	2.681273	10.928191	3.357436	

	X78	X79	X80	X81	X82
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.007738	0.041743	0.012735	-0.001013	0.021738
std	0.230116	0.830283	0.385975	0.259536	0.342382
min	-5.010775	-0.434996	-1.708978	-8.063588	-3.824489
25%	-0.001637	0.007186	-0.005886	-0.001947	0.007256
50%	0.005855	0.018225	0.003663	0.005473	0.018289
75%	0.018260	0.027137	0.013346	0.018504	0.026763

```
max        5.147594     26.252397     12.030190      0.761076     10.090176

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.029208	0.001537	0.009224	0.006149	0.018447	0.016910	0.006149	
1	0.019266	0.000000	0.007290	0.005728	0.007290	0.012497	0.000000	
2	0.022130	0.000000	0.009900	0.013395	0.008736	0.015142	0.001165	
3	0.029458	0.004208	0.008417	0.007014	0.025250	0.018236	0.009819	
4	0.037163	0.004645	0.055745	0.083617	0.009291	0.027873	0.004645	
..	…	…	…	…	…	…	…	
995	0.016932	0.001881	0.008466	0.010347	0.027279	0.012229	0.002822	
996	0.017296	0.002594	0.002594	0.002594	0.006919	0.018161	0.006919	
997	0.041126	0.001210	0.004838	0.019354	0.012096	0.020563	0.006048	
998	0.025290	0.004215	0.063224	0.059009	0.071654	0.067439	0.033720	
999	0.023083	0.000000	0.001154	0.006348	0.004617	0.012695	0.000577	

	X7	X8	X9	…	X74	X75	X76	\
0	0.012298	0.003075	0.044581	…	0.014407	0.000241	-0.003610	
1	0.007811	0.001562	0.021349	…	0.019046	0.002514	0.013814	
2	0.008153	0.000000	0.028537	…	0.021313	0.003590	0.018538	
3	0.025250	0.002806	0.039277	…	0.024276	0.004543	-0.003502	
4	0.046454	0.065036	0.102199	…	0.006080	0.023937	-0.021472	
..	…	…	…	…	…	…	…	
995	0.039508	0.017873	0.035745	…	0.000823	-0.003767	0.001644	
996	0.012972	0.000000	0.030269	…	0.011273	0.002692	0.017193	
997	0.001210	0.001210	0.035078	…	0.002746	-0.001241	0.015066	
998	0.134879	0.113804	0.109589	…	-0.022002	-0.086485	-0.011014	
999	0.003462	0.000577	0.025968	…	0.020497	0.001538	0.012265	

	X77	X78	X79	X80	X81	X82	X83
0	0.014488	0.004276	-0.010369	0.015479	0.004256	0.014991	efectores
1	0.027512	0.008055	0.010009	0.020183	-0.002058	0.003396	efectores
2	0.010589	-0.002200	0.023844	0.017372	0.003816	0.017261	efectores
3	0.016830	-0.002598	0.016586	0.005648	-0.000403	0.007535	efectores
4	-0.057620	0.006883	-0.006780	-0.035410	-0.001048	0.010215	efectores
..	…	…	…	…	…	…	
995	0.017368	0.012238	0.000062	0.021156	-0.001376	-0.005957	efectores
996	0.019029	0.004450	0.014508	0.012719	0.004687	0.016268	efectores
997	0.022813	0.006532	0.035620	0.012692	0.002235	0.024560	efectores
998	0.029531	0.032603	0.037754	-0.031086	-0.017250	-0.037757	efectores
999	0.022021	0.005453	0.005946	0.009990	-0.001091	0.011137	efectores

[907 rows x 84 columns]

Composición de pseudo aminoácidos (PseAAC) hidro_mass efectores archaea dataset 2, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	907.000000	907.000000	907.000000	907.000000	907.000000	907.000000	
mean	0.031057	0.002760	0.023317	0.029918	0.014676	0.025757	
std	0.013442	0.004293	0.017279	0.025432	0.011264	0.011582	
min	0.000000	0.000000	0.000000	0.000694	0.000000	0.000000	
25%	0.021542	0.000000	0.008974	0.007539	0.007280	0.016751	
50%	0.028730	0.000623	0.018951	0.022746	0.011648	0.023623	
75%	0.038826	0.003910	0.035129	0.047813	0.019421	0.032754	
max	0.086600	0.023089	0.080638	0.143510	0.099850	0.093194	

	X6	X7	X8	X9	…	X73	\
count	907.000000	907.000000	907.000000	907.000000	…	907.000000	
mean	0.007074	0.023880	0.020808	0.041237	…	0.014837	
std	0.006448	0.023860	0.027637	0.023410	…	0.016992	
min	0.000000	0.000000	0.000000	0.001727	…	-0.070359	
25%	0.002140	0.007219	0.001960	0.024164	…	0.005433	
50%	0.005817	0.014813	0.007775	0.036272	…	0.015235	
75%	0.009755	0.032700	0.032559	0.052346	…	0.024609	
max	0.038236	0.138042	0.193044	0.179339	…	0.120876	

	X74	X75	X76	X77	X78	X79	\
count	907.000000	907.000000	907.000000	907.000000	907.000000	907.000000	
mean	0.004843	0.008384	0.014341	0.006898	0.008862	0.014806	
std	0.021447	0.019075	0.015827	0.021264	0.019097	0.016552	
min	-0.123862	-0.086485	-0.060211	-0.081437	-0.082187	-0.066948	
25%	-0.005360	-0.001232	0.005549	-0.002301	0.000227	0.006023	
50%	0.008179	0.004358	0.016215	0.010772	0.006207	0.015963	
75%	0.015950	0.015378	0.024204	0.018671	0.015726	0.025034	
max	0.111119	0.116878	0.077323	0.093847	0.097163	0.100530	

	X80	X81	X82
count	907.000000	907.000000	907.000000
mean	0.007237	0.009157	0.014846
std	0.022699	0.021021	0.016330
min	-0.102177	-0.104345	-0.078956
25%	-0.003107	-0.000341	0.005938
50%	0.011360	0.005291	0.015599
75%	0.018823	0.016139	0.023682
max	0.134910	0.135581	0.080634

[8 rows x 83 columns]

```
no_efectores
Composición de pseudo aminoácidos (PseAAC)  hidro_mass no_efectores archaea
dataset 2, sin valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.015828  0.002374  0.010684  0.013454  0.013849  0.020576  0.004748
1    0.032574  0.004072  0.020359  0.012215  0.004072  0.020359  0.008143
2    0.033896  0.001883  0.007532  0.009886  0.010828  0.017419  0.003766
3    0.018434  0.000000  0.022384  0.010534  0.008559  0.019751  0.000658
4    0.022749  0.001338  0.009367  0.008029  0.006691  0.018735  0.001338
..        …         …         …         …         …         …         …
995  0.027029  0.000000  0.018019  0.072077  0.022524  0.040543  0.031534
996  0.031871  0.000483  0.017867  0.030422  0.004346  0.020764  0.003380
997  0.014001  0.000000  0.035003  0.028003  0.007001  0.028003  0.003500
998  0.016249  0.008124  0.064995  0.048746  0.073119  0.016249  0.008124
999  0.022716  0.000000  0.007572  0.011358  0.008518  0.017983  0.000946

           X7        X8        X9  …       X74       X75       X76  \
0    0.020972  0.007123  0.025720  …  0.007815  0.005175  0.014828
1    0.000000  0.004072  0.016287  … -0.045436 -0.006526  0.018385
2    0.016006  0.003295  0.028717  …  0.004424 -0.001240  0.020444
3    0.010534  0.001317  0.009875  … -0.003717 -0.007761  0.040307
4    0.018735  0.005353  0.036131  …  0.014780  0.001271  0.009879
..        …         …         …  …         …         …         …
995  0.022524  0.000000  0.018019  …  0.009806  0.013011 -0.019928
996  0.010141  0.005795  0.013038  …  0.007039  0.017690  0.032664
997  0.014001  0.007001  0.031503  …  0.004765  0.053387 -0.019897
998  0.105617  0.129990  0.081244  … -0.031942 -0.074732 -0.015621
999  0.006625  0.001893  0.009465  …  0.013790 -0.001104  0.030989

           X77       X78       X79       X80       X81       X82           X83
0    0.013975  0.007454  0.010588  0.015772  0.011753  0.017825  no_efectores
1   -0.041437 -0.009876  0.033724  0.007202  0.010319  0.014286  no_efectores
2    0.017636  0.004162  0.022557  0.018871  0.004344  0.024093  no_efectores
3    0.006427  0.002561  0.035678  0.006098  0.001040  0.038502  no_efectores
4    0.012973  0.005436  0.015523  0.008732  0.000544  0.012553  no_efectores
..        …         …         …         …         …         …           …
995 -0.013979 -0.000842  0.025760 -0.000517  0.011657  0.034203  no_efectores
996 -0.000402  0.010093  0.023449 -0.001733  0.009544  0.021929  no_efectores
997  0.006887  0.035050  0.030864 -0.005206  0.020606 -0.002171  no_efectores
998  0.010838  0.016893  0.018228 -0.053875 -0.123349  0.038521  no_efectores
999  0.006746 -0.006158  0.031477  0.002493 -0.000101  0.025315  no_efectores

[972 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores archaea
dataset 2, sin valores atípicos.
Estadísticas.

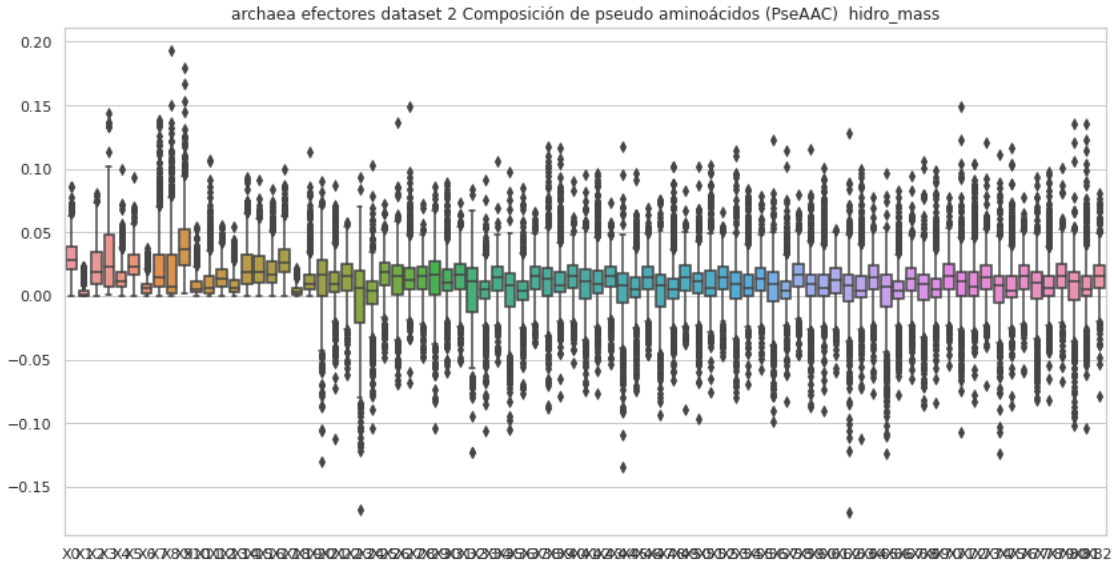
	X0	X1	X2	X3	X4	X5	\
count	972.000000	972.000000	972.000000	972.000000	972.000000	972.000000	
mean	0.033898	0.004309	0.030219	0.032494	0.013421	0.028507	
std	0.016735	0.007932	0.018463	0.021497	0.012394	0.014603	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.023035	0.000000	0.017753	0.016119	0.005976	0.019214	
50%	0.031525	0.001893	0.028249	0.029545	0.010743	0.026439	
75%	0.041281	0.004953	0.040853	0.046055	0.016967	0.034771	
max	0.193582	0.089421	0.132165	0.135056	0.158215	0.184584	

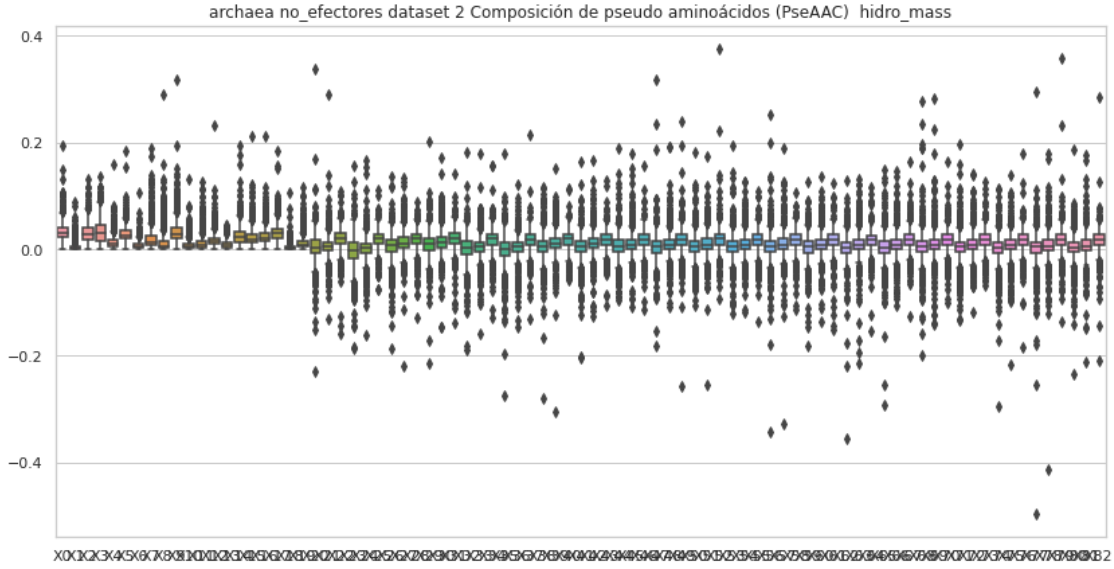
	X6	X7	X8	X9	…	X73	\
count	972.000000	972.000000	972.000000	972.000000	…	972.000000	
mean	0.008300	0.019442	0.013795	0.032500	…	0.017373	
std	0.008810	0.020703	0.021195	0.022394	…	0.018364	
min	0.000000	0.000000	0.000000	0.000000	…	-0.067803	
25%	0.002620	0.007235	0.002214	0.019731	…	0.007944	
50%	0.005977	0.013333	0.006249	0.027128	…	0.018793	
75%	0.011291	0.024147	0.015943	0.039603	…	0.027707	
max	0.105477	0.189751	0.290061	0.316430	…	0.127182	

	X74	X75	X76	X77	X78	X79	\
count	972.000000	972.000000	972.000000	972.000000	972.000000	972.000000	
mean	0.001683	0.007719	0.017137	0.002134	0.008186	0.017124	
std	0.024810	0.022871	0.020584	0.031215	0.026443	0.023060	
min	-0.294966	-0.216083	-0.182969	-0.496140	-0.412830	-0.135429	
25%	-0.006961	-0.000972	0.008707	-0.006582	-0.001309	0.007609	
50%	0.003442	0.006881	0.018314	0.004166	0.005848	0.018335	
75%	0.013246	0.017221	0.028661	0.013944	0.017926	0.027137	
max	0.113218	0.139236	0.177873	0.294676	0.178035	0.357822	

	X80	X81	X82
count	972.000000	972.000000	972.000000
mean	0.003008	0.008026	0.016352
std	0.023864	0.024202	0.022482
min	-0.235161	-0.212503	-0.209154
25%	-0.005065	-0.001613	0.007651
50%	0.003753	0.005624	0.018328
75%	0.013346	0.018468	0.026666
max	0.186655	0.177058	0.285695

[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 2,
con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.053806  0.002832  0.016992  0.011328  0.033983  0.031151  0.011328
1    0.048228  0.000000  0.018248  0.014338  0.018248  0.031283  0.000000
2    0.038068  0.000000  0.017031  0.023041  0.015027  0.026047  0.002004
3    0.074927  0.010704  0.021408  0.017840  0.064223  0.046383  0.024976
4    0.051063  0.006383  0.076594  0.114891  0.012766  0.038297  0.006383
..        …         …         …         …         …         …         …
995  0.055000  0.006111  0.027500  0.033611  0.088611  0.039722  0.009167
996  0.032304  0.004846  0.004846  0.004846  0.012922  0.033919  0.012922
997  0.064006  0.001883  0.007530  0.030121  0.018825  0.032003  0.009413
998  0.024703  0.004117  0.061757  0.057639  0.069991  0.065874  0.032937
999  0.049527  0.000000  0.002476  0.013620  0.009905  0.027240  0.001238

           X7        X8        X9  …       X32       X33       X34  \
```

```
0    0.022655  0.005664  0.082126  …  0.041668 -0.007840  0.037496
1    0.019552  0.003910  0.053442  …  0.047362  0.016906  0.032676
2    0.014025  0.000000  0.049088  …  0.042427  0.018044  0.011255
3    0.064223  0.007136  0.099903  … -0.005678  0.075388  0.036467
4    0.063829  0.089360  0.140423  …  0.017537  0.012573  0.007613
..        …       …       …  …       …        …       …
995  0.128333  0.058055  0.116111  …  0.027130 -0.018276 -0.007566
996  0.024228  0.000000  0.056532  …  0.031068  0.040972  0.042051
997  0.001883  0.001883  0.054593  …  0.039961  0.057951  0.052912
998  0.131747  0.111162  0.107045  … -0.036560  0.041286 -0.021767
999  0.007429  0.001238  0.055718  …  0.041764  0.015277  0.023478

          X35       X36       X37       X38       X39       X40        X41
0    0.028432  0.061973  0.026573 -0.006651 -0.019102  0.027615  efectores
1    0.051805  0.030775  0.037002  0.034580  0.025054  0.008501  efectores
2    0.043799  0.020190  0.033655  0.031889  0.041016  0.029693  efectores
3   -0.039168  0.022705 -0.008809 -0.008906  0.042187  0.019165  efectores
4    0.022232 -0.003969 -0.042083 -0.029503 -0.009316  0.014036  efectores
..        …       …       …       …       …       …        …
995  0.011272 -0.026872  0.003643  0.005342  0.000201 -0.019350  efectores
996  0.023932  0.021710  0.014794  0.032112  0.027097  0.030383  efectores
997  0.004396  0.023807  0.036906  0.023447  0.055436  0.038224  efectores
998  0.050611 -0.003075 -0.026697 -0.010758  0.036877 -0.036880  efectores
999  0.025885  0.033375  0.051036  0.026315  0.012757  0.023896  efectores

[1000 rows x 42 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass efectores archaea dataset 2, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.046600     0.004739     0.035383     0.047163     0.023614
std       0.016861     0.007912     0.023468     0.038614     0.018081
min       0.000000     0.000000     0.000000     0.001003     0.000000
25%       0.035731     0.000000     0.017112     0.015910     0.012351
50%       0.044983     0.001365     0.030054     0.035422     0.018661
75%       0.055516     0.006372     0.048864     0.068450     0.029946
max       0.131162     0.083553     0.126138     0.258959     0.187265

                X5           X6           X7           X8           X9  …  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
mean      0.039163     0.011145     0.038907     0.034080     0.066123  …
std       0.013728     0.009837     0.034175     0.041022     0.033550  …
min       0.000000     0.000000     0.000000     0.000000     0.001733  …
25%       0.030509     0.003452     0.013009     0.003687     0.041181  …
```

```
50%       0.037577     0.009169     0.026561     0.013502     0.059642  …
75%       0.046525     0.015950     0.057771     0.056978     0.086879  …
max       0.172640     0.071670     0.245316     0.208882     0.256970  …

               X31          X32          X33          X34          X35  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.016389     0.021987     0.016488     0.017907     0.016024
std       0.025735     0.025505     0.025751     0.028371     0.029887
min      -0.119161    -0.147415    -0.104445    -0.328025    -0.146748
25%       0.002767     0.009381     0.001321     0.005817     0.003194
50%       0.020823     0.026243     0.020525     0.023212     0.020595
75%       0.033290     0.038288     0.033719     0.035862     0.032963
max       0.095408     0.112660     0.203285     0.135531     0.401512

               X36          X37          X38          X39          X40
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.018977     0.020604     0.019616     0.019542     0.019387
std       0.026341     0.025456     0.026121     0.027915     0.026990
min      -0.126154    -0.100638    -0.127705    -0.153497    -0.137463
25%       0.005993     0.006893     0.005639     0.007868     0.006382
50%       0.024642     0.023918     0.024932     0.024387     0.023360
75%       0.035626     0.036221     0.036617     0.035816     0.035378
max       0.127703     0.111402     0.140316     0.129672     0.104192

[8 rows x 41 columns]
```

no_efectores
Composición de pseudo aminoácidos (PseAAC) mass no_efectores archaea dataset 2, con valores atípicos.
Valores del documento csv.

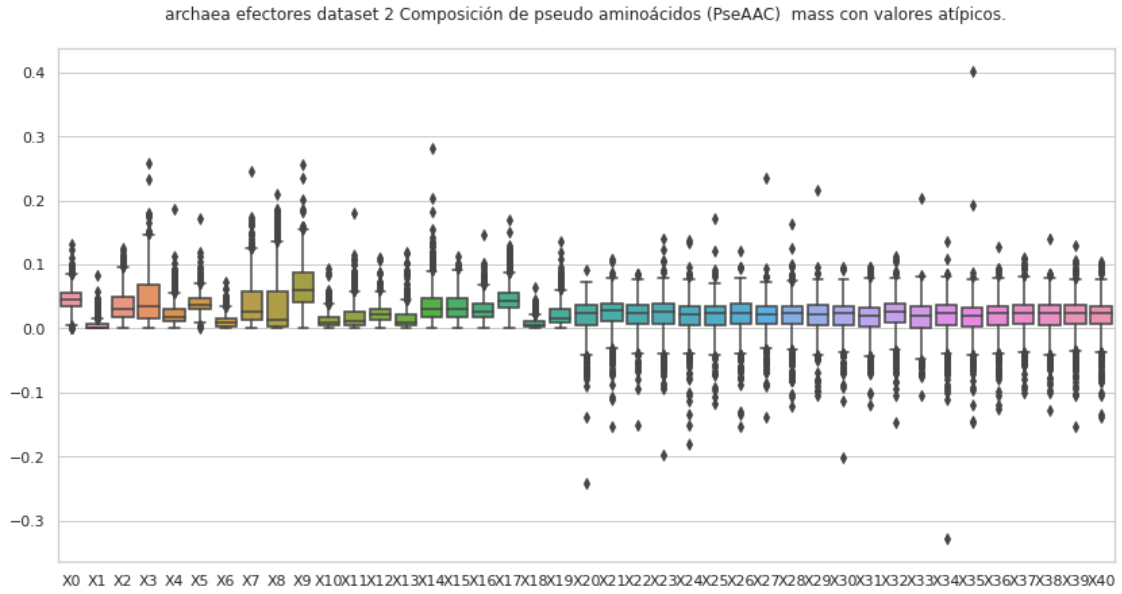
```
          X0        X1        X2        X3        X4        X5        X6  \
0    0.029017  0.004353  0.019586  0.024664  0.025390  0.037722  0.008705
1    0.040620  0.005077  0.025387  0.015232  0.005077  0.025387  0.010155
2    0.055274  0.003071  0.012283  0.016121  0.017657  0.028405  0.006142
3    0.020043  0.000000  0.024338  0.011453  0.009306  0.021475  0.000716
4    0.048365  0.002845  0.019915  0.017070  0.014225  0.039830  0.002845
..        …         …         …         …         …         …         …
995  0.026554  0.000000  0.017703  0.070811  0.022129  0.039831  0.030980
996  0.043142  0.000654  0.024185  0.041181  0.005883  0.028107  0.004576
997  0.023076  0.000000  0.057691  0.046153  0.011538  0.046153  0.005769
998  0.017439  0.008720  0.069758  0.052318  0.078477  0.017439  0.008720
999  0.031609  0.000000  0.010536  0.015805  0.011854  0.025024  0.001317

          X7        X8        X9  …       X32       X33       X34  \
0    0.038448  0.013058  0.047153  …  0.021834  0.025389  0.024375
```

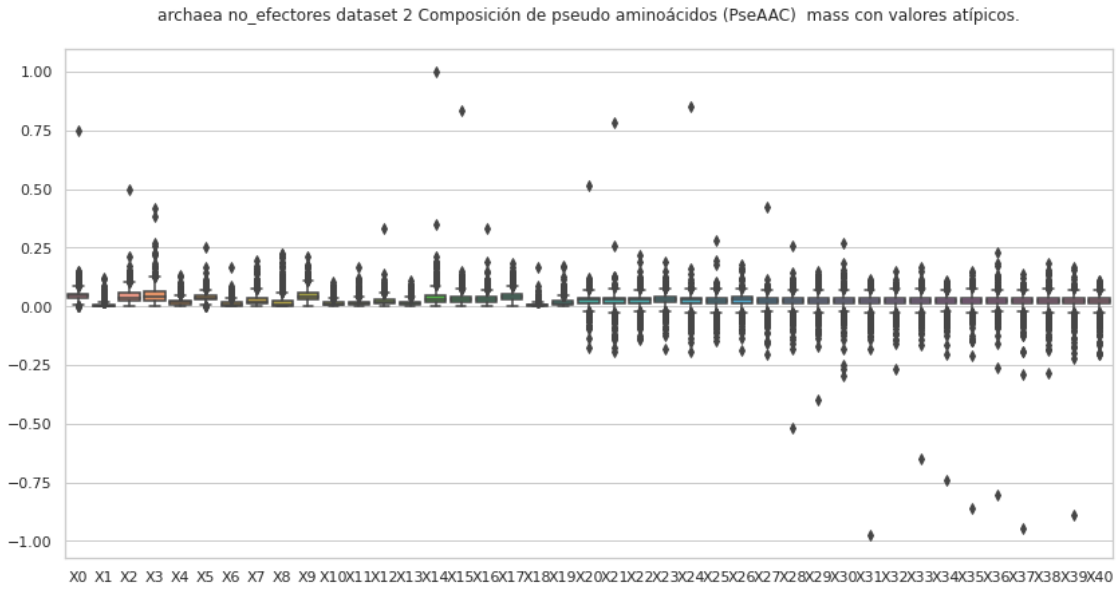
```
1    0.000000  0.005077  0.020310  …  0.036005  0.023165  0.049923
2    0.026101  0.005374  0.046829  …  0.021512  0.030014  0.039739
3    0.011453  0.001432  0.010737  …  0.035808  0.034232  0.039172
4    0.039830  0.011380  0.076814  …  0.039897  0.025598  0.045934
..        …         …        …  …        …         …         …
995  0.022129  0.000000  0.017703  …  0.052941  0.025372  0.061507
996  0.013727  0.007844  0.017649  …  0.030186  0.035472  0.028380
997  0.023076  0.011538  0.051922  … -0.003480  0.051813 -0.011992
998  0.113356  0.139515  0.087197  … -0.022445  0.021906 -0.013157
999  0.009219  0.002634  0.013171  …  0.040183  0.037632  0.040533

          X35       X36       X37       X38       X39       X40           X41
0    0.028316  0.024842  0.032183  0.027183  0.019410  0.032679  no_efectores
1    0.059489  0.047402  0.056362  0.022926  0.042055  0.017815  no_efectores
2    0.029588  0.040788  0.019735  0.033339  0.036784  0.039288  no_efectores
3    0.036282  0.036500  0.041565  0.043825  0.038792  0.041862  no_efectores
4    0.035120  0.062036  0.011961  0.021002  0.033001  0.026687  no_efectores
..        …         …         …         …         …         …           …
995  0.020147  0.062379  0.030667 -0.019578  0.025308  0.033602  no_efectores
996  0.035997  0.020489  0.033352  0.044215  0.031741  0.029684  no_efectores
997  0.041883  0.028448  0.054959 -0.032793  0.050869 -0.003578  no_efectores
998  0.026253 -0.020961  0.006297 -0.016766  0.019563  0.041344  no_efectores
999  0.039496  0.039487  0.045870  0.043121  0.043801  0.035226  no_efectores

[1000 rows x 42 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass no_efectores archaea dataset 2, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.047241     0.006010     0.042936     0.047477     0.018866
std       0.029420     0.011206     0.031290     0.038342     0.015531
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       0.034502     0.000000     0.023296     0.021758     0.008599
50%       0.044034     0.002611     0.037833     0.038969     0.015316
75%       0.055220     0.006619     0.055376     0.064265     0.024934
max       0.749927     0.119744     0.499951     0.416626     0.132770

                X5           X6           X7           X8           X9  …  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
mean      0.039007     0.011445     0.026884     0.020359     0.046252  …
std       0.017007     0.012320     0.024507     0.029848     0.027373  …
min       0.000000     0.000000     0.000000     0.000000     0.000000  …
25%       0.028947     0.003565     0.010500     0.003076     0.027239  …
50%       0.036806     0.008327     0.019581     0.009107     0.040341  …
```

```
75%       0.046236     0.015948     0.035628     0.025771     0.059946  …
max       0.249976     0.166650     0.192773     0.226006     0.212012  …

                X31          X32          X33          X34          X35  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.018779     0.020372     0.021246     0.020048     0.021302
std       0.042541     0.028864     0.034831     0.035976     0.039737
min      -0.975554    -0.265537    -0.652781    -0.742471    -0.859449
25%       0.011059     0.010202     0.010453     0.012612     0.010739
50%       0.024800     0.025936     0.026202     0.025382     0.026024
75%       0.035926     0.035678     0.036310     0.035639     0.036358
max       0.115855     0.151064     0.165024     0.112004     0.148596

                X36          X37          X38          X39          X40
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.021447     0.019827     0.021669     0.019642     0.019817
std       0.039420     0.042939     0.031982     0.041526     0.031520
min      -0.805386    -0.948314    -0.283893    -0.888564    -0.204625
25%       0.013114     0.011362     0.011986     0.009888     0.010470
50%       0.024971     0.025907     0.025520     0.025628     0.025500
75%       0.035975     0.035559     0.036586     0.035389     0.034515
max       0.228035     0.139985     0.182641     0.166370     0.111071

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.053806	0.002832	0.016992	0.011328	0.033983	0.031151	0.011328	
1	0.048228	0.000000	0.018248	0.014338	0.018248	0.031283	0.000000	
2	0.038068	0.000000	0.017031	0.023041	0.015027	0.026047	0.002004	
3	0.074927	0.010704	0.021408	0.017840	0.064223	0.046383	0.024976	
4	0.051063	0.006383	0.076594	0.114891	0.012766	0.038297	0.006383	
..	…	…	…	…	…	…	…	
993	0.000000	0.025596	0.025596	0.093850	0.000000	0.051191	0.000000	
996	0.032304	0.004846	0.004846	0.004846	0.012922	0.033919	0.012922	
997	0.064006	0.001883	0.007530	0.030121	0.018825	0.032003	0.009413	
998	0.024703	0.004117	0.061757	0.057639	0.069991	0.065874	0.032937	
999	0.049527	0.000000	0.002476	0.013620	0.009905	0.027240	0.001238	

	X7	X8	X9	…	X32	X33	X34	\
0	0.022655	0.005664	0.082126	…	0.041668	-0.007840	0.037496	
1	0.019552	0.003910	0.053442	…	0.047362	0.016906	0.032676	
2	0.014025	0.000000	0.049088	…	0.042427	0.018044	0.011255	
3	0.064223	0.007136	0.099903	…	-0.005678	0.075388	0.036467	
4	0.063829	0.089360	0.140423	…	0.017537	0.012573	0.007613	
..	…	…	…	…	…	…	…	
993	0.051191	0.059723	0.034127	…	0.005596	0.004561	0.027934	
996	0.024228	0.000000	0.056532	…	0.031068	0.040972	0.042051	
997	0.001883	0.001883	0.054593	…	0.039961	0.057951	0.052912	
998	0.131747	0.111162	0.107045	…	-0.036560	0.041286	-0.021767	
999	0.007429	0.001238	0.055718	…	0.041764	0.015277	0.023478	

```
          X35       X36       X37       X38       X39       X40        X41
0    0.028432  0.061973  0.026573 -0.006651 -0.019102  0.027615  efectores
1    0.051805  0.030775  0.037002  0.034580  0.025054  0.008501  efectores
2    0.043799  0.020190  0.033655  0.031889  0.041016  0.029693  efectores
3   -0.039168  0.022705 -0.008809 -0.008906  0.042187  0.019165  efectores
4    0.022232 -0.003969 -0.042083 -0.029503 -0.009316  0.014036  efectores
..        …         …         …         …         …         …
993  0.057667  0.019377  0.022517  0.014656  0.001606 -0.020838  efectores
996  0.023932  0.021710  0.014794  0.032112  0.027097  0.030383  efectores
997  0.004396  0.023807  0.036906  0.023447  0.055436  0.038224  efectores
998  0.050611 -0.003075 -0.026697 -0.010758  0.036877 -0.036880  efectores
999  0.025885  0.033375  0.051036  0.026315  0.012757  0.023896  efectores

[817 rows x 42 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass efectores archaea dataset 2, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  817.000000  817.000000  817.000000  817.000000  817.000000  817.000000
mean     0.045676    0.003223    0.030695    0.037860    0.020614    0.036979
std      0.014261    0.004781    0.019843    0.030683    0.013980    0.010701
min      0.000000    0.000000    0.000000    0.001003    0.000000    0.005212
25%      0.036154    0.000000    0.015888    0.013114    0.011301    0.029723
50%      0.044562    0.000984    0.024737    0.028446    0.017208    0.036179
75%      0.053996    0.005146    0.044135    0.056678    0.026561    0.043311
max      0.092556    0.027588    0.104004    0.146118    0.076982    0.080027

                X6          X7          X8          X9  …         X31  \
count  817.000000  817.000000  817.000000  817.000000  …  817.000000
mean     0.009482    0.031981    0.025407    0.058284  …    0.019766
std      0.007592    0.028729    0.032672    0.027203  …    0.020316
min      0.000000    0.000000    0.000000    0.001733  …   -0.053807
25%      0.003247    0.010230    0.002921    0.038845  …    0.008403
50%      0.007825    0.021172    0.008944    0.053569  …    0.023587
75%      0.013961    0.044022    0.039570    0.075531  …    0.033803
max      0.039841    0.131747    0.152003    0.158796  …    0.080460

               X32         X33         X34         X35         X36         X37  \
count  817.000000  817.000000  817.000000  817.000000  817.000000  817.000000
mean     0.024851    0.018901    0.022678    0.020564    0.023282    0.023379
std      0.019904    0.020527    0.020056    0.020301    0.019914    0.021503
min     -0.044736   -0.057753   -0.058085   -0.059463   -0.047208   -0.051511
25%      0.013615    0.007317    0.011562    0.010623    0.011599    0.011384
50%      0.028049    0.022639    0.025497    0.023666    0.026953    0.025958
75%      0.038911    0.033764    0.036883    0.033891    0.036173    0.036711
```

```
max      0.087230    0.075388    0.086233    0.077211    0.092905    0.095945

              X38         X39         X40
count  817.000000  817.000000  817.000000
mean     0.023643    0.023709    0.023779
std      0.020785    0.019932    0.019869
min     -0.055051   -0.060127   -0.041965
25%      0.012656    0.012916    0.012591
50%      0.027448    0.026425    0.025586
75%      0.037491    0.036344    0.036253
max      0.084531    0.091134    0.100009

[8 rows x 41 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass no_efectores archaea dataset 2, sin valores atípicos.
Valores del documento csv.

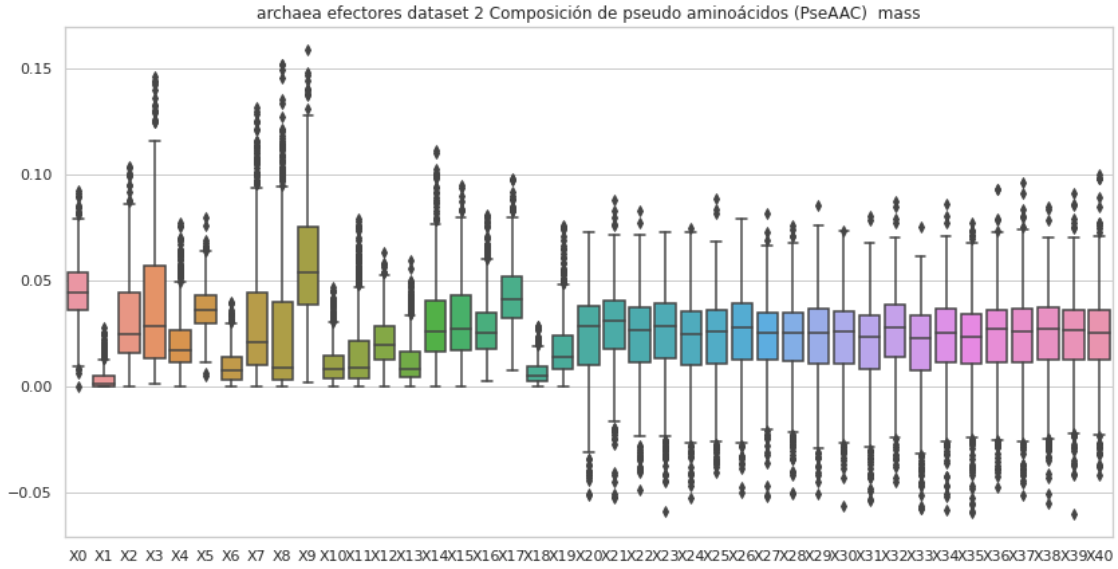
```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.029017  0.004353  0.019586  0.024664  0.025390  0.037722  0.008705
1    0.040620  0.005077  0.025387  0.015232  0.005077  0.025387  0.010155
2    0.055274  0.003071  0.012283  0.016121  0.017657  0.028405  0.006142
3    0.020043  0.000000  0.024338  0.011453  0.009306  0.021475  0.000716
4    0.048365  0.002845  0.019915  0.017070  0.014225  0.039830  0.002845
..        …         …         …         …         …         …         …
994  0.034909  0.009241  0.025668  0.039016  0.010267  0.040043  0.011294
995  0.026554  0.000000  0.017703  0.070811  0.022129  0.039831  0.030980
996  0.043142  0.000654  0.024185  0.041181  0.005883  0.028107  0.004576
997  0.023076  0.000000  0.057691  0.046153  0.011538  0.046153  0.005769
999  0.031609  0.000000  0.010536  0.015805  0.011854  0.025024  0.001317

           X7        X8        X9  …       X32       X33       X34  \
0    0.038448  0.013058  0.047153  …  0.021834  0.025389  0.024375
1    0.000000  0.005077  0.020310  …  0.036005  0.023165  0.049923
2    0.026101  0.005374  0.046829  …  0.021512  0.030014  0.039739
3    0.011453  0.001432  0.010737  …  0.035808  0.034232  0.039172
4    0.039830  0.011380  0.076814  …  0.039897  0.025598  0.045934
..        …         …         …  …  …       …         …         …
994  0.019508  0.004107  0.028749  …  0.034241  0.033074  0.028740
995  0.022129  0.000000  0.017703  …  0.052941  0.025372  0.061507
996  0.013727  0.007844  0.017649  …  0.030186  0.035472  0.028380
997  0.023076  0.011538  0.051922  … -0.003480  0.051813 -0.011992
999  0.009219  0.002634  0.013171  …  0.040183  0.037632  0.040533

           X35       X36       X37       X38       X39       X40           X41
0    0.028316  0.024842  0.032183  0.027183  0.019410  0.032679  no_efectores
```

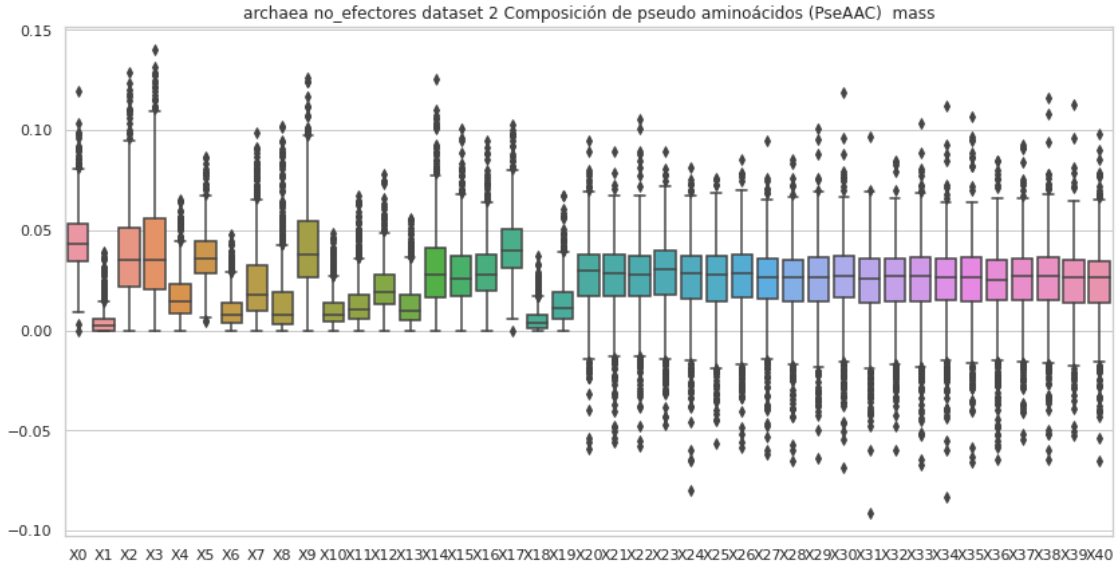
```
1    0.059489  0.047402  0.056362  0.022926  0.042055  0.017815  no_efectores
2    0.029588  0.040788  0.019735  0.033339  0.036784  0.039288  no_efectores
3    0.036282  0.036500  0.041565  0.043825  0.038792  0.041862  no_efectores
4    0.035120  0.062036  0.011961  0.021002  0.033001  0.026687  no_efectores
..        …         …         …         …         …         …           …
994  0.026342  0.028564  0.031875  0.035255  0.035939  0.030560  no_efectores
995  0.020147  0.062379  0.030667 -0.019578  0.025308  0.033602  no_efectores
996  0.035997  0.020489  0.033352  0.044215  0.031741  0.029684  no_efectores
997  0.041883  0.028448  0.054959 -0.032793  0.050869 -0.003578  no_efectores
999  0.039496  0.039487  0.045870  0.043121  0.043801  0.035226  no_efectores

[871 rows x 42 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass no_efectores archaea dataset 2, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  871.000000  871.000000  871.000000  871.000000  871.000000  871.000000
mean     0.044901    0.004327    0.038100    0.040536    0.016815    0.037176
std      0.015164    0.005955    0.022371    0.026886    0.012099    0.012157
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.004414
25%      0.034489    0.000000    0.022061    0.020205    0.008274    0.028740
50%      0.043466    0.002523    0.035291    0.035203    0.014215    0.036130
75%      0.052965    0.005861    0.050986    0.055870    0.022784    0.044424
max      0.119140    0.039223    0.128916    0.140123    0.064999    0.086378

                X6          X7          X8          X9  …         X31  \
count  871.000000  871.000000  871.000000  871.000000  …  871.000000
mean     0.009872    0.023523    0.014638    0.042035  …    0.022856
std      0.008500    0.018777    0.018365    0.021650  …    0.019600
min      0.000000    0.000000    0.000000    0.000000  …   -0.091332
25%      0.003588    0.009860    0.002891    0.026225  …    0.013563
50%      0.007899    0.018029    0.007626    0.038104  …    0.025771
75%      0.013809    0.032217    0.019000    0.054668  …    0.035880
max      0.048158    0.098932    0.102041    0.126292  …    0.096767

               X32         X33         X34         X35         X36         X37  \
count  871.000000  871.000000  871.000000  871.000000  871.000000  871.000000
mean     0.023572    0.023971    0.023540    0.024543    0.023319    0.024225
std      0.018897    0.020144    0.019127    0.019450    0.020068    0.018719
min     -0.059967   -0.067380   -0.083547   -0.066138   -0.064461   -0.054860
25%      0.014684    0.014425    0.015335    0.014653    0.015093    0.015028
50%      0.027123    0.027092    0.026320    0.026291    0.025234    0.027035
75%      0.035797    0.036238    0.035616    0.036189    0.035374    0.035522
max      0.084948    0.103211    0.112004    0.106812    0.084763    0.092504
```

```
              X38          X39          X40
count  871.000000   871.000000   871.000000
mean     0.024892     0.024201     0.024106
std      0.019505     0.018208     0.018982
min     -0.064367    -0.052190    -0.065553
25%      0.014972     0.013842     0.013741
50%      0.026804     0.026620     0.026687
75%      0.036366     0.035320     0.034467
max      0.116044     0.113038     0.097918

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.038588	0.002031	0.012186	0.008124	0.024371	0.022340	0.008124	
1	0.026125	0.000000	0.009885	0.007767	0.009885	0.016946	0.000000	
2	0.035485	0.000000	0.015875	0.021477	0.014007	0.024279	0.001868	
3	0.033514	0.004788	0.009575	0.007979	0.028726	0.020747	0.011171	
4	0.037279	0.004660	0.055919	0.083878	0.009320	0.027959	0.004660	
..	…	…	…	…	…	…	…	
995	0.017411	0.001935	0.008706	0.010640	0.028052	0.012575	0.002902	
996	0.027315	0.004097	0.004097	0.004097	0.010926	0.028681	0.010926	
997	0.067124	0.001974	0.007897	0.031588	0.019742	0.033562	0.009871	
998	0.025711	0.004285	0.064278	0.059993	0.072849	0.068563	0.034282	
999	0.034180	0.000000	0.001709	0.009400	0.006836	0.018799	0.000855	

	X7	X8	X9	…	X53	X54	X55	\
0	0.016248	0.004062	0.058897	…	0.006720	0.051142	0.021240	
1	0.010591	0.002118	0.028949	…	0.002233	0.027569	0.007380	
2	0.013073	0.000000	0.045756	…	0.011762	0.017205	-0.010124	
3	0.028726	0.003192	0.044685	…	-0.004065	0.011228	-0.002934	
4	0.046599	0.065239	0.102518	…	0.019187	0.016953	0.050324	
..	…	…	…	…	…	…	…	
995	0.040626	0.018379	0.036757	…	0.014086	-0.002247	-0.002430	
996	0.020486	0.000000	0.047802	…	0.007738	0.015759	-0.000306	
997	0.001974	0.001974	0.057253	…	0.002613	0.030150	0.007426	
998	0.137127	0.115701	0.111415	…	0.090452	0.030592	0.010011	
999	0.005127	0.000855	0.038453	…	0.005454	0.019605	-0.004212	

	X56	X57	X58	X59	X60	X61	X62
0	0.019033	0.000318	0.019141	0.005649	0.020450	0.005622	efectores
1	0.025826	0.003409	0.037306	0.010922	0.027368	-0.002790	efectores
2	0.034175	0.005757	0.016978	-0.003528	0.027854	0.006119	efectores
3	0.027618	0.005168	0.019147	-0.002956	0.006425	-0.000458	efectores
4	0.006099	0.024012	-0.057799	0.006905	-0.035521	-0.001051	efectores
..	…	…	…	…	…	…	…
995	0.000846	-0.003873	0.017860	0.012585	0.021756	-0.001415	efectores
996	0.017802	0.004251	0.030052	0.007027	0.020087	0.007401	efectores
997	0.004481	-0.002026	0.037235	0.010661	0.020716	0.003648	efectores

```
998 -0.022369 -0.087927  0.030023  0.033147 -0.031604 -0.017538  efectores
999  0.030352  0.002278  0.032609  0.008075  0.014793 -0.001615  efectores

[1000 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro efectores archaea dataset 2, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.049245	0.004848	0.036653	0.046776	0.022721	
std	0.028280	0.008047	0.027180	0.040564	0.022717	
min	0.000000	0.000000	0.000000	0.000853	0.000000	
25%	0.030238	0.000000	0.013593	0.011966	0.011059	
50%	0.043394	0.001437	0.031148	0.045358	0.018583	
75%	0.063808	0.006390	0.054896	0.069270	0.028352	
max	0.298393	0.081380	0.323508	0.438985	0.384896	

	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.041819	0.011102	0.037542	0.033242	0.064386	…	
std	0.026846	0.011415	0.042767	0.052622	0.047489	…	
min	0.000000	0.000000	0.000000	0.000000	0.008088	…	
25%	0.025369	0.003639	0.013252	0.003852	0.039173	…	
50%	0.035589	0.009304	0.023977	0.013047	0.055192	…	
75%	0.052433	0.014761	0.050341	0.051950	0.077544	…	
max	0.384896	0.149957	0.647016	0.970524	0.739447	…	

	X52	X53	X54	X55	X56	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.005802	0.008740	0.013548	0.015429	0.004562	
std	0.041166	0.034838	0.043320	0.039191	0.044316	
min	-0.630490	-0.573346	-0.651246	-0.653968	-0.464862	
25%	-0.006620	-0.002562	-0.000413	-0.000153	-0.010592	
50%	0.014270	0.007215	0.017148	0.011223	0.011633	
75%	0.026003	0.020957	0.030323	0.029604	0.025081	
max	0.188454	0.173541	0.187650	0.209592	0.302645	

	X57	X58	X59	X60	X61
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.009694	0.007785	0.010875	0.006928	0.010545
std	0.034742	0.039896	0.035597	0.046068	0.040694
min	-0.271387	-0.362277	-0.319247	-0.557559	-0.433868
25%	-0.002414	-0.006313	-0.000499	-0.007050	-0.001060
50%	0.007194	0.015648	0.009311	0.015540	0.008305
75%	0.023288	0.028360	0.024125	0.028989	0.024695

```
max       0.208613     0.232076     0.221137     0.223754     0.331290

[8 rows x 62 columns]
```

no_efectores
Composición de pseudo aminoácidos (PseAAC) hidro no_efectores archaea dataset 2, con valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.022809  0.003421  0.015396  0.019388  0.019958  0.029652  0.006843
1    0.076446  0.009556  0.047779  0.028667  0.009556  0.047779  0.019111
2    0.056139  0.003119  0.012475  0.016374  0.017933  0.028849  0.006238
3    0.065584  0.000000  0.079638  0.037477  0.030450  0.070269  0.002342
4    0.029423  0.001731  0.012115  0.010385  0.008654  0.024231  0.001731
..        …         …         …         …         …         …         …
995  0.055030  0.000000  0.036686  0.146745  0.045858  0.082544  0.064201
996  0.063278  0.000959  0.035474  0.060402  0.008629  0.041227  0.006711
997  0.019866  0.000000  0.049665  0.039732  0.009933  0.039732  0.004966
998  0.014992  0.007496  0.059968  0.044976  0.067465  0.014992  0.007496
999  0.054887  0.000000  0.018296  0.027443  0.020583  0.043452  0.002287

           X7        X8        X9  …       X53       X54       X55  \
0    0.030222  0.010264  0.037065  …  0.000504  0.013761  0.004942
1    0.000000  0.009556  0.038223  …  0.035367  0.091853  0.094252
2    0.026510  0.005458  0.047562  …  0.004910  0.016643  0.003827
3    0.037477  0.004685  0.035134  … -0.001135  0.022070  0.030078
4    0.024231  0.006923  0.046731  … -0.000183  0.014359 -0.000814
..        …         …         …  …         …         …         …
995  0.045858  0.000000  0.036686  …  0.105316  0.003187 -0.008910
996  0.020134  0.011505  0.025886  …  0.022165  0.000013  0.022291
997  0.019866  0.009933  0.044698  …  0.089731 -0.031397 -0.008924
998  0.097449  0.119937  0.074961  …  0.075291  0.000335 -0.016432
999  0.016009  0.004574  0.022870  …  0.007404  0.028911  0.006190

          X56       X57       X58       X59       X60       X61           X62
0    0.011262  0.007458  0.020139  0.010742  0.022729  0.016937  no_efectores
1   -0.106631 -0.015317 -0.097248 -0.023178  0.016903  0.024217  no_efectores
2    0.007328 -0.002053  0.029209  0.006893  0.031255  0.007195  no_efectores
3   -0.013225 -0.027611  0.022866  0.009110  0.021694  0.003699  no_efectores
4    0.019116  0.001644  0.016778  0.007031  0.011293  0.000703  no_efectores
..        …         …         …         …         …         …           …
995  0.019964  0.026489 -0.028461 -0.001715 -0.001052  0.023732  no_efectores
996  0.013976  0.035123 -0.000798  0.020040 -0.003441  0.018949  no_efectores
997  0.006760  0.075749  0.009771  0.049732 -0.007386  0.029237  no_efectores
998 -0.029472 -0.068952  0.010000  0.015587 -0.049708 -0.113810  no_efectores
```

```
999  0.033320 -0.002667  0.016299 -0.014878  0.006024 -0.000245  no_efectores

[1000 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro no_efectores archaea dataset 2, con valores atípicos.
Estadísticas.

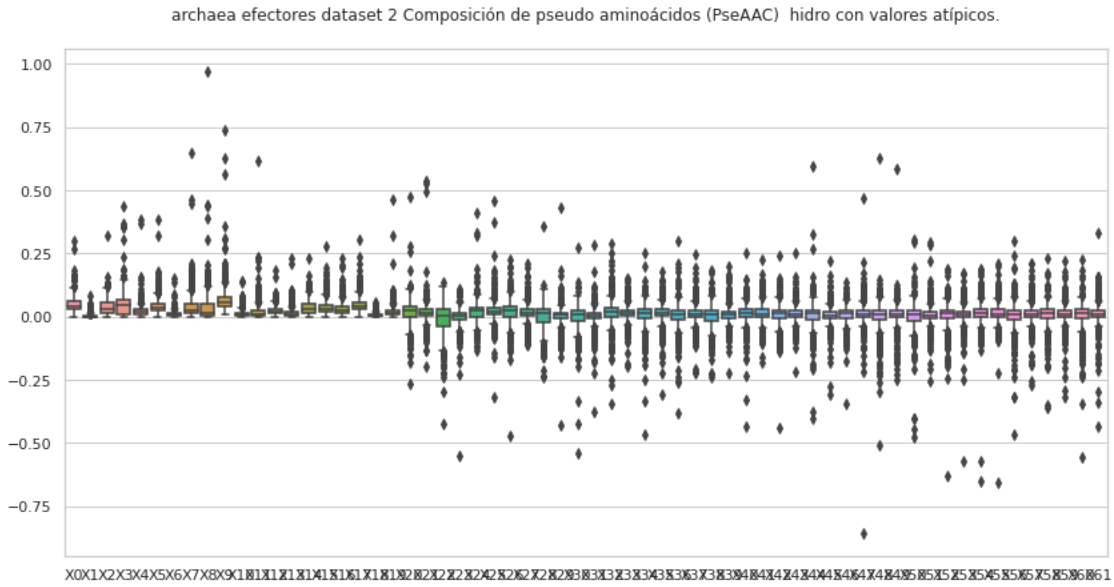
	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.059754	0.007553	0.049292	0.051539	0.021186	
std	0.048231	0.019157	0.032875	0.036499	0.024692	
min	-0.559770	-0.111954	-0.335862	-0.335862	-0.334742	
25%	0.035669	0.000000	0.029563	0.031001	0.010839	
50%	0.054688	0.003247	0.050748	0.051820	0.018483	
75%	0.078476	0.008331	0.066674	0.068677	0.026804	
max	0.658539	0.439026	0.316231	0.329270	0.289816	

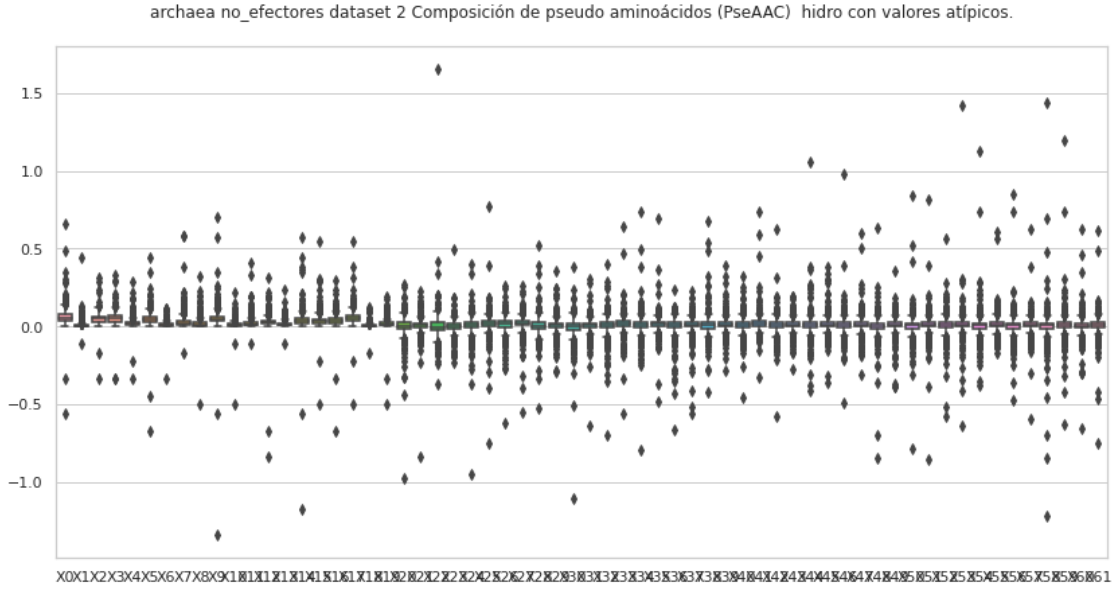
	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.049372	0.013197	0.032403	0.020350	0.051701	…	
std	0.042567	0.017691	0.040185	0.032902	0.062459	…	
min	-0.669484	-0.334742	-0.167371	-0.502113	-1.338967	…	
25%	0.030135	0.004217	0.012751	0.004183	0.034236	…	
50%	0.044884	0.010324	0.023062	0.011218	0.046851	…	
75%	0.064890	0.018487	0.040286	0.025310	0.065202	…	
max	0.439026	0.117222	0.585588	0.322361	0.702706	…	

	X52	X53	X54	X55	X56	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.005079	0.014180	0.005516	0.014568	0.004175	
std	0.050877	0.063486	0.060969	0.042338	0.054997	
min	-0.580254	-0.635620	-0.377510	-0.174571	-0.470106	
25%	-0.010549	-0.000177	-0.011441	-0.002022	-0.011970	
50%	0.008821	0.012396	0.006646	0.011502	0.006705	
75%	0.024135	0.028536	0.022095	0.029624	0.022469	
max	0.561738	1.419386	1.128825	0.608423	0.852935	

	X57	X58	X59	X60	X61
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.011334	0.002906	0.012662	0.003975	0.010223
std	0.045133	0.085733	0.063088	0.051987	0.051688
min	-0.592960	-1.221237	-0.633613	-0.657921	-0.752988
25%	-0.002057	-0.011630	-0.002759	-0.010060	-0.003239
50%	0.011358	0.007672	0.010680	0.006176	0.009684
75%	0.027840	0.023272	0.027970	0.021874	0.029072
max	0.624834	1.434660	1.193755	0.627965	0.614482

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.038588	0.002031	0.012186	0.008124	0.024371	0.022340	0.008124	
1	0.026125	0.000000	0.009885	0.007767	0.009885	0.016946	0.000000	
2	0.035485	0.000000	0.015875	0.021477	0.014007	0.024279	0.001868	
3	0.033514	0.004788	0.009575	0.007979	0.028726	0.020747	0.011171	
4	0.037279	0.004660	0.055919	0.083878	0.009320	0.027959	0.004660	
..	…	…	…	…	…	…	…	
995	0.017411	0.001935	0.008706	0.010640	0.028052	0.012575	0.002902	
996	0.027315	0.004097	0.004097	0.004097	0.010926	0.028681	0.010926	
997	0.067124	0.001974	0.007897	0.031588	0.019742	0.033562	0.009871	
998	0.025711	0.004285	0.064278	0.059993	0.072849	0.068563	0.034282	
999	0.034180	0.000000	0.001709	0.009400	0.006836	0.018799	0.000855	

	X7	X8	X9	…	X53	X54	X55	\
0	0.016248	0.004062	0.058897	…	0.006720	0.051142	0.021240	
1	0.010591	0.002118	0.028949	…	0.002233	0.027569	0.007380	
2	0.013073	0.000000	0.045756	…	0.011762	0.017205	-0.010124	
3	0.028726	0.003192	0.044685	…	-0.004065	0.011228	-0.002934	
4	0.046599	0.065239	0.102518	…	0.019187	0.016953	0.050324	
..	…	…	…	…	…	…	…	
995	0.040626	0.018379	0.036757	…	0.014086	-0.002247	-0.002430	
996	0.020486	0.000000	0.047802	…	0.007738	0.015759	-0.000306	
997	0.001974	0.001974	0.057253	…	0.002613	0.030150	0.007426	
998	0.137127	0.115701	0.111415	…	0.090452	0.030592	0.010011	
999	0.005127	0.000855	0.038453	…	0.005454	0.019605	-0.004212	

	X56	X57	X58	X59	X60	X61	X62
0	0.019033	0.000318	0.019141	0.005649	0.020450	0.005622	efectores
1	0.025826	0.003409	0.037306	0.010922	0.027368	-0.002790	efectores
2	0.034175	0.005757	0.016978	-0.003528	0.027854	0.006119	efectores
3	0.027618	0.005168	0.019147	-0.002956	0.006425	-0.000458	efectores
4	0.006099	0.024012	-0.057799	0.006905	-0.035521	-0.001051	efectores
..	…	…	…	…	…	…	…
995	0.000846	-0.003873	0.017860	0.012585	0.021756	-0.001415	efectores
996	0.017802	0.004251	0.030052	0.007027	0.020087	0.007401	efectores
997	0.004481	-0.002026	0.037235	0.010661	0.020716	0.003648	efectores
998	-0.022369	-0.087927	0.030023	0.033147	-0.031604	-0.017538	efectores
999	0.030352	0.002278	0.032609	0.008075	0.014793	-0.001615	efectores

```
[850 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC) efectores archaea dataset 2, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	850.000000	850.000000	850.000000	850.000000	850.000000	850.000000	
mean	0.046006	0.003558	0.031638	0.038387	0.019886	0.037579	
std	0.022835	0.005454	0.021977	0.028549	0.012046	0.019015	
min	0.000000	0.000000	0.000000	0.000853	0.000000	0.000000	
25%	0.029488	0.000000	0.012289	0.010237	0.010749	0.024005	
50%	0.041675	0.000957	0.026427	0.035315	0.017824	0.032774	
75%	0.060333	0.004839	0.049307	0.060794	0.026863	0.047847	
max	0.121836	0.027086	0.097498	0.131837	0.081061	0.119432	

	X6	X7	X8	X9	…	X52	\
count	850.000000	850.000000	850.000000	850.000000	…	850.000000	
mean	0.009455	0.029204	0.023870	0.055342	…	0.009330	
std	0.007532	0.024151	0.029361	0.023533	…	0.025439	
min	0.000000	0.000000	0.000000	0.008173	…	-0.097850	
25%	0.003421	0.011735	0.002841	0.037576	…	-0.003092	
50%	0.008660	0.020541	0.010396	0.050883	…	0.015090	
75%	0.013210	0.040368	0.037895	0.069062	…	0.025740	
max	0.042371	0.137127	0.156719	0.179379	…	0.085825	

	X53	X54	X55	X56	X57	X58	\
count	850.000000	850.000000	850.000000	850.000000	850.000000	850.000000	
mean	0.010196	0.014944	0.014733	0.008750	0.010748	0.011159	
std	0.020552	0.025376	0.024467	0.025101	0.022381	0.026340	
min	-0.068984	-0.108125	-0.084191	-0.104740	-0.087927	-0.087458	
25%	-0.001295	0.002182	0.000514	-0.007047	-0.001721	-0.001928	
50%	0.007366	0.017420	0.010129	0.013355	0.006985	0.016740	
75%	0.019611	0.029084	0.026704	0.024883	0.021342	0.027854	
max	0.108601	0.120770	0.116923	0.119278	0.107730	0.118778	

	X59	X60	X61
count	850.000000	850.000000	850.000000
mean	0.012213	0.012224	0.011993
std	0.022617	0.027250	0.022467
min	-0.066722	-0.119135	-0.065315
25%	0.000692	-0.002429	-0.000159
50%	0.009146	0.017161	0.008334
75%	0.022504	0.028799	0.022147
max	0.099019	0.118770	0.114465

[8 rows x 62 columns]

no_efectores
Composición de pseudo aminoácidos (PseAAC) no_efectores archaea dataset 2, sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.022809	0.003421	0.015396	0.019388	0.019958	0.029652	0.006843	
2	0.056139	0.003119	0.012475	0.016374	0.017933	0.028849	0.006238	
3	0.065584	0.000000	0.079638	0.037477	0.030450	0.070269	0.002342	
4	0.029423	0.001731	0.012115	0.010385	0.008654	0.024231	0.001731	
5	0.036252	0.002132	0.061842	0.066107	0.038385	0.087432	0.023457	
..	…	…	…	…	…	…	…	
993	0.036408	0.005689	0.039822	0.067128	0.019342	0.039822	0.011378	
994	0.074517	0.019725	0.054792	0.083284	0.021917	0.085476	0.024109	
996	0.063278	0.000959	0.035474	0.060402	0.008629	0.041227	0.006711	
997	0.019866	0.000000	0.049665	0.039732	0.009933	0.039732	0.004966	
999	0.054887	0.000000	0.018296	0.027443	0.020583	0.043452	0.002287	

	X7	X8	X9	…	X53	X54	X55	\
0	0.030222	0.010264	0.037065	…	0.000504	0.013761	0.004942	
2	0.026510	0.005458	0.047562	…	0.004910	0.016643	0.003827	
3	0.037477	0.004685	0.035134	…	-0.001135	0.022070	0.030078	
4	0.024231	0.006923	0.046731	…	-0.000183	0.014359	-0.000814	
5	0.110889	0.078902	0.063975	…	-0.009638	0.018587	0.011404	
..	…	…	…	…	…	…	…	
993	0.029582	0.031857	0.028444	…	0.035474	0.009716	0.010177	
994	0.041642	0.008767	0.061367	…	0.049424	-0.010857	-0.023122	
996	0.020134	0.011505	0.025886	…	0.022165	0.000013	0.022291	
997	0.019866	0.009933	0.044698	…	0.089731	-0.031397	-0.008924	
999	0.016009	0.004574	0.022870	…	0.007404	0.028911	0.006190	

	X56	X57	X58	X59	X60	X61	X62
0	0.011262	0.007458	0.020139	0.010742	0.022729	0.016937	no_efectores
2	0.007328	-0.002053	0.029209	0.006893	0.031255	0.007195	no_efectores
3	-0.013225	-0.027611	0.022866	0.009110	0.021694	0.003699	no_efectores
4	0.019116	0.001644	0.016778	0.007031	0.011293	0.000703	no_efectores
5	0.054422	0.010501	-0.009001	-0.016472	0.040332	0.025057	no_efectores
..	…	…	…	…	…	…	…
993	-0.000457	0.016345	-0.014150	0.002870	0.017011	0.030436	no_efectores
994	0.019852	0.002420	0.002523	-0.015853	0.018818	-0.004528	no_efectores
996	0.013976	0.035123	-0.000798	0.020040	-0.003441	0.018949	no_efectores
997	0.006760	0.075749	0.009771	0.049732	-0.007386	0.029237	no_efectores
999	0.033320	-0.002667	0.016299	-0.014878	0.006024	-0.000245	no_efectores

[898 rows x 63 columns]

Composición de pseudo aminoácidos (PseAAC) no_efectores archaea dataset 2, sin valores atípicos.
Estadísticas.

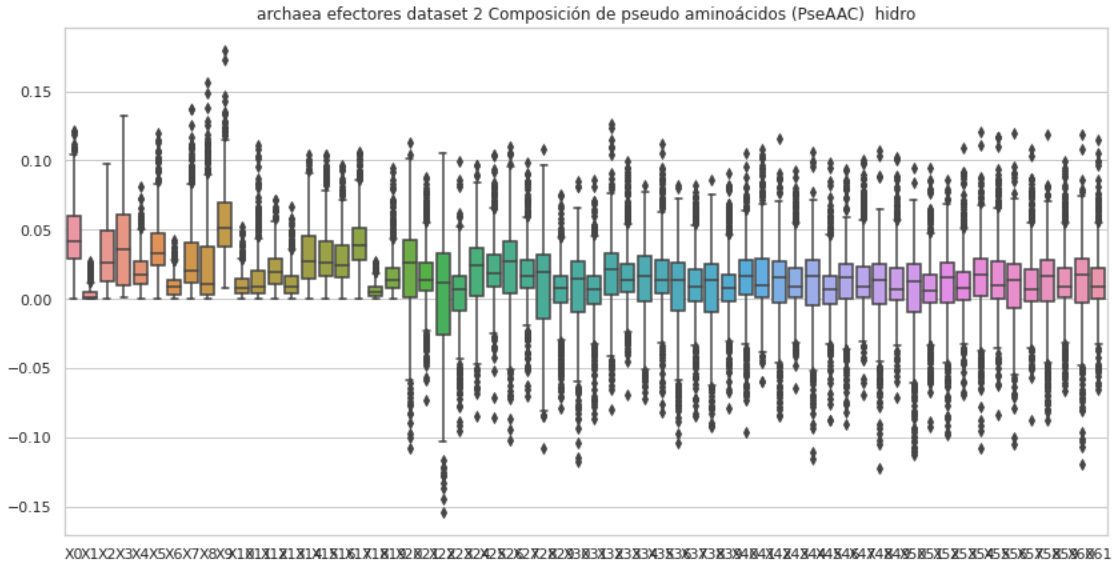
	X0	X1	X2	X3	X4	X5	\
count	898.000000	898.000000	898.000000	898.000000	898.000000	898.000000	
mean	0.056790	0.005797	0.047087	0.048229	0.019351	0.046842	
std	0.028679	0.008229	0.024931	0.025341	0.012652	0.023234	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.035602	0.000000	0.028217	0.029393	0.010670	0.029452	
50%	0.053338	0.003198	0.049504	0.050132	0.017945	0.043526	
75%	0.075814	0.007679	0.064761	0.066088	0.025082	0.061862	
max	0.167076	0.057145	0.128872	0.130474	0.080763	0.130812	

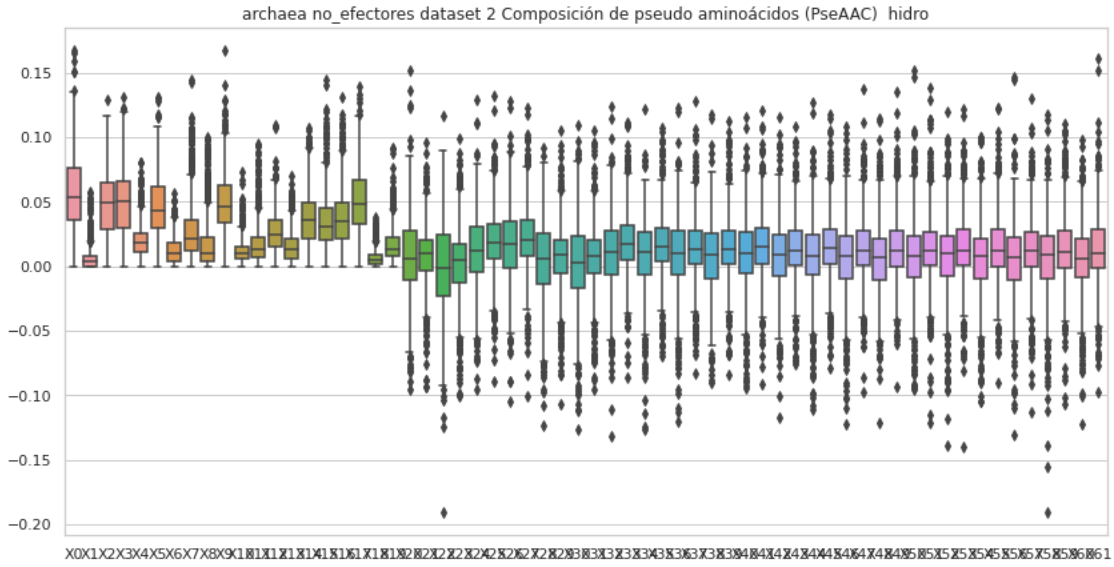
	X6	X7	X8	X9	…	X52	\
count	898.000000	898.000000	898.000000	898.000000	…	898.000000	
mean	0.011969	0.027256	0.016989	0.048729	…	0.007209	
std	0.010028	0.022111	0.019315	0.022153	…	0.028338	
min	0.000000	0.000000	0.000000	0.000000	…	-0.139094	
25%	0.004212	0.011845	0.003982	0.033548	…	-0.007507	
50%	0.009659	0.021643	0.010310	0.045720	…	0.009416	
75%	0.017741	0.035737	0.022263	0.062312	…	0.023597	
max	0.056510	0.144556	0.100077	0.166795	…	0.120076	

	X53	X54	X55	X56	X57	X58	\
count	898.000000	898.000000	898.000000	898.000000	898.000000	898.000000	
mean	0.014163	0.005720	0.014126	0.005308	0.013083	0.005735	
std	0.023944	0.028062	0.025316	0.027696	0.024926	0.028994	
min	-0.139691	-0.105204	-0.090744	-0.131221	-0.112731	-0.190768	
25%	0.001021	-0.009430	-0.000710	-0.010281	-0.000236	-0.009453	
50%	0.012396	0.007655	0.011603	0.007188	0.011589	0.008502	
75%	0.028027	0.021665	0.028397	0.022051	0.026931	0.022893	
max	0.121158	0.099628	0.122736	0.146439	0.130067	0.117875	

	X59	X60	X61
count	898.000000	898.000000	898.000000
mean	0.012609	0.005669	0.012624
std	0.025976	0.026652	0.025572
min	-0.106820	-0.122356	-0.097443
25%	-0.001614	-0.008323	-0.001961
50%	0.010524	0.006246	0.009990
75%	0.027050	0.021273	0.028360
max	0.110694	0.097967	0.160653

[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 2,
con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.005443 -0.018635 -0.084997 -0.007834  0.002934 -0.015078  0.046394
1   -0.027923 -0.077917  0.055378 -0.034381 -0.085988  0.008007  0.077016
2    0.023057  0.030394  0.147859  0.089264 -0.045016  0.050764  0.016967
3    0.007678 -0.063872  0.030566 -0.079442 -0.095993 -0.145327  0.064974
4    0.039778 -0.020321 -0.075481 -0.036650 -0.078937 -0.097488 -0.092399
..        …        …        …        …        …        …        …
995 -0.013519 -0.003511  0.034309  0.041849  0.003351  0.099324 -0.014219
996 -0.001671 -0.001633 -0.062107 -0.074465  0.104763 -0.001856  0.164325
997 -0.145829  0.071592 -0.035244 -0.073373 -0.002276 -0.028309 -0.012430
998 -0.097416 -0.001915 -0.090449  0.010720 -0.042461  0.079594 -0.040465
999  0.030004  0.041750  0.013307  0.039298 -0.071875  0.085857 -0.022942

           X7        X8        X9       X10       X11       X12        X13
0    0.060522  0.186961 -0.000218 -0.088922 -0.139434  0.063320  efectores
```

```
1   -0.032107  0.047364  0.050714 -0.045058 -0.059875  0.083870  efectores
2   -0.046013  0.039564  0.024939 -0.073187 -0.004072  0.081296  efectores
3    0.045862 -0.113507  0.044521  0.121701  0.063497 -0.073121  efectores
4   -0.002665  0.021018 -0.008155  0.070357  0.105667  0.017713  efectores
..        …        …        …        …        …        …        …
995  0.079241 -0.041507 -0.047778 -0.001032  0.002203  0.050089  efectores
996  0.037454 -0.024362  0.045248 -0.047640 -0.025376 -0.010043  efectores
997  0.051975 -0.013730 -0.057041  0.023549 -0.031107  0.068816  efectores
998  0.088117 -0.038272  0.061181 -0.078103  0.081104 -0.083219  efectores
999  0.036628 -0.117445  0.053187  0.001043  0.044585  0.053890  efectores

[1000 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro_mass efectores archaea dataset 2, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.005704     0.019238     0.009245     0.011975    -0.002747
std       0.071498     0.069503     0.067815     0.072728     0.074425
min      -0.288239    -0.263362    -0.221731    -0.328625    -0.279168
25%      -0.037723    -0.021341    -0.031696    -0.033717    -0.046670
50%       0.007683     0.020929     0.008020     0.014001    -0.001834
75%       0.052304     0.060322     0.051280     0.060371     0.042571
max       0.240172     0.280200     0.279326     0.267981     0.246180

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.002615     0.011469     0.000510     0.000965     0.005450
std       0.071049     0.069775     0.067489     0.072901     0.074874
min      -0.268516    -0.272288    -0.275302    -0.254238    -0.240528
25%      -0.038933    -0.031756    -0.040692    -0.045458    -0.040829
50%       0.004592     0.012231     0.001178     0.003485     0.006540
75%       0.048291     0.055993     0.044875     0.044636     0.050723
max       0.255310     0.230651     0.279565     0.296241     0.257050

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.002491    -0.004675     0.015066
std       0.072246     0.068194     0.069542
min      -0.343092    -0.266614    -0.251885
25%      -0.040014    -0.046998    -0.031421
50%       0.001269    -0.004096     0.011460
75%       0.048002     0.037843     0.062672
max       0.208882     0.384496     0.247189
```

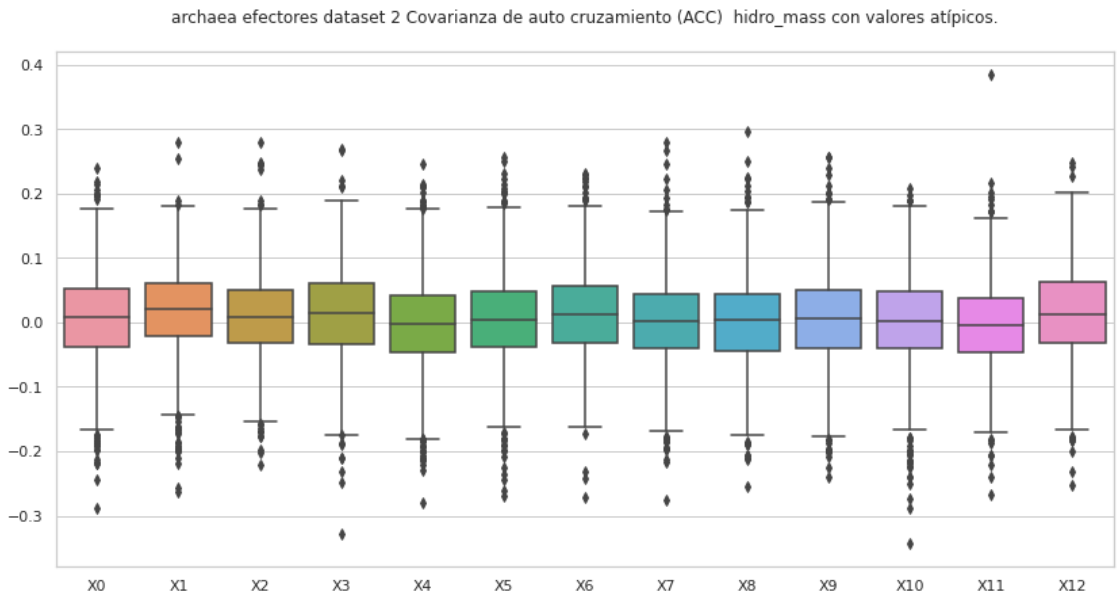
no_efectores
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores archaea dataset 2,
con valores atípicos.
Valores del documento csv.

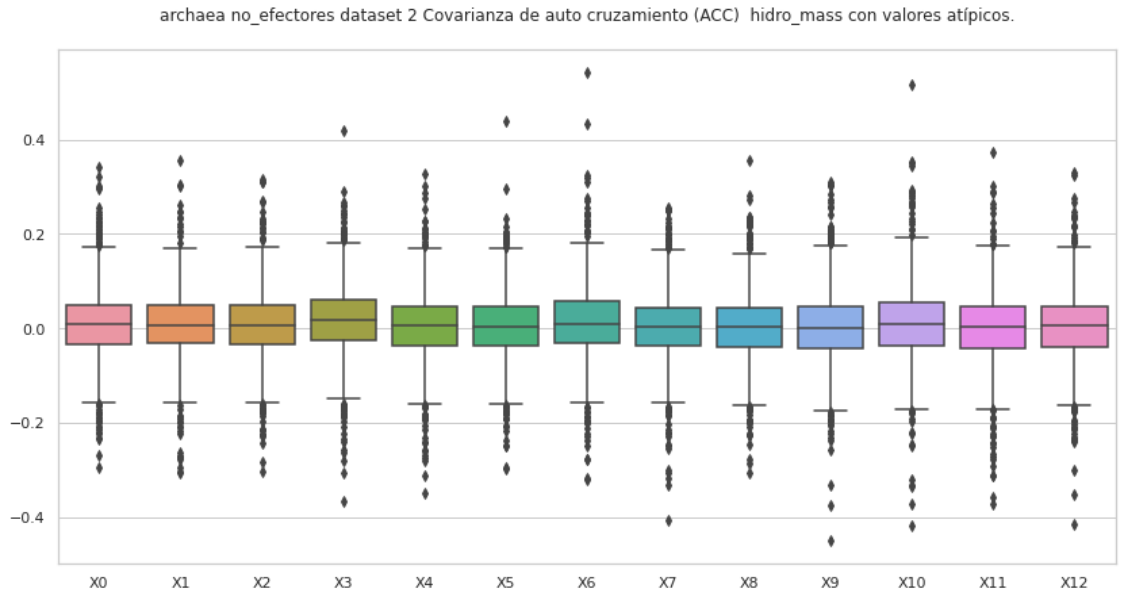
```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.035777 -0.000696 -0.015934 -0.019013 -0.022376 -0.046919 -0.026761
1   -0.095522  0.020359 -0.197128  0.266886 -0.156052  0.053897 -0.195423
2   -0.044814 -0.042353  0.053133 -0.002771  0.020951  0.056514 -0.006538
3    0.031429  0.020576 -0.008731 -0.002777  0.000790 -0.030101 -0.034517
4    0.032967 -0.002815  0.092848 -0.000931 -0.087647 -0.094754 -0.092877
..        …       …       …       …       …       …       …
995 -0.039702 -0.040522  0.053232 -0.119806 -0.001014 -0.029432 -0.130018
996  0.022590  0.029463 -0.039808  0.073215 -0.076629 -0.002488 -0.034594
997 -0.056812  0.129133 -0.025008  0.066403 -0.015866  0.192747 -0.023293
998  0.012404  0.029974 -0.018173 -0.017288  0.021667 -0.028013 -0.044589
999 -0.012844 -0.033228 -0.087665  0.036276  0.004147  0.082489 -0.084467

           X7        X8        X9       X10       X11       X12           X13
0    0.066634  0.024145  0.020167 -0.002941  0.019606 -0.022116  no_efectores
1    0.108045  0.094838 -0.034535 -0.090834  0.072988  0.024905  no_efectores
2    0.016612 -0.000786 -0.075658  0.044028 -0.054618 -0.045816  no_efectores
3    0.014429 -0.095531  0.002909  0.020919 -0.018494 -0.003396  no_efectores
4   -0.253573 -0.154219 -0.035814  0.002671  0.030237  0.049257  no_efectores
..        …       …       …       …       …       …           …
995  0.035035  0.004656 -0.027292 -0.077032 -0.018723  0.114534  no_efectores
996  0.009819  0.074126  0.075080 -0.096116 -0.003635 -0.008089  no_efectores
997  0.130843 -0.226724  0.185644 -0.062958  0.123454 -0.082783  no_efectores
998 -0.017176 -0.055874 -0.031499 -0.085771 -0.051345 -0.046825  no_efectores
999  0.078884 -0.070880  0.100605 -0.117218  0.019784  0.034981  no_efectores

[1000 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores archaea dataset 2,
con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.009371     0.007519     0.006046     0.017315     0.004939
std       0.077181     0.074947     0.074478     0.078219     0.079044
min      -0.293860    -0.304968    -0.303053    -0.367565    -0.350528
25%      -0.033233    -0.032132    -0.034023    -0.024565    -0.035928
50%       0.008294     0.006587     0.005147     0.018026     0.006328
75%       0.049987     0.050546     0.048637     0.059594     0.047324
```

```
max       0.340503     0.356153     0.315695     0.420034     0.325809

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.002858     0.012934     0.002281     0.002076     0.003786
std       0.074990     0.081209     0.077353     0.076281     0.081353
min      -0.298457    -0.319387    -0.405955    -0.307032    -0.449415
25%      -0.037419    -0.030827    -0.037889    -0.040296    -0.041708
50%       0.002629     0.009409     0.002361     0.004266     0.001120
75%       0.045782     0.056712     0.044381     0.043278     0.047040
max       0.440056     0.541764     0.255430     0.355540     0.309617

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.011002     0.000722     0.004145
std       0.085426     0.080000     0.077686
min      -0.416512    -0.372297    -0.415588
25%      -0.035977    -0.041123    -0.038337
50%       0.007642     0.003644     0.006109
75%       0.055579     0.045960     0.047405
max       0.515834     0.373322     0.329582
```





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 2,
sin valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.005443 -0.018635 -0.084997 -0.007834  0.002934 -0.015078  0.046394
1   -0.027923 -0.077917  0.055378 -0.034381 -0.085988  0.008007  0.077016
2    0.023057  0.030394  0.147859  0.089264 -0.045016  0.050764  0.016967
3    0.007678 -0.063872  0.030566 -0.079442 -0.095993 -0.145327  0.064974
4    0.039778 -0.020321 -0.075481 -0.036650 -0.078937 -0.097488 -0.092399
..        …         …         …         …         …         …         …
995 -0.013519 -0.003511  0.034309  0.041849  0.003351  0.099324 -0.014219
996 -0.001671 -0.001633 -0.062107 -0.074465  0.104763 -0.001856  0.164325
997 -0.145829  0.071592 -0.035244 -0.073373 -0.002276 -0.028309 -0.012430
998 -0.097416 -0.001915 -0.090449  0.010720 -0.042461  0.079594 -0.040465
999  0.030004  0.041750  0.013307  0.039298 -0.071875  0.085857 -0.022942

           X7        X8        X9       X10       X11       X12        X13
0    0.060522  0.186961 -0.000218 -0.088922 -0.139434  0.063320  efectores
1   -0.032107  0.047364  0.050714 -0.045058 -0.059875  0.083870  efectores
2   -0.046013  0.039564  0.024939 -0.073187 -0.004072  0.081296  efectores
3    0.045862 -0.113507  0.044521  0.121701  0.063497 -0.073121  efectores
4   -0.002665  0.021018 -0.008155  0.070357  0.105667  0.017713  efectores
..        …         …         …         …         …         …         …
995  0.079241 -0.041507 -0.047778 -0.001032  0.002203  0.050089  efectores
```

```
996  0.037454 -0.024362  0.045248 -0.047640 -0.025376 -0.010043  efectores
997  0.051975 -0.013730 -0.057041  0.023549 -0.031107  0.068816  efectores
998  0.088117 -0.038272  0.061181 -0.078103  0.081104 -0.083219  efectores
999  0.036628 -0.117445  0.053187  0.001043  0.044585  0.053890  efectores

[926 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro_mass efectores archaea dataset 2,
sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	926.000000	926.000000	926.000000	926.000000	926.000000	926.000000	
mean	0.007371	0.021685	0.010584	0.011301	-0.003145	0.001796	
std	0.066184	0.062369	0.062037	0.067484	0.070588	0.065604	
min	-0.197650	-0.186867	-0.156994	-0.190481	-0.221905	-0.208499	
25%	-0.035197	-0.017638	-0.029985	-0.032357	-0.046193	-0.037374	
50%	0.008979	0.021272	0.010331	0.012985	-0.001834	0.003712	
75%	0.051020	0.060177	0.051251	0.057915	0.041146	0.045863	
max	0.213739	0.188358	0.188834	0.221510	0.212889	0.201266	

	X6	X7	X8	X9	X10	X11	\
count	926.000000	926.000000	926.000000	926.000000	926.000000	926.000000	
mean	0.012145	-0.000545	0.000170	0.005622	0.004908	-0.004472	
std	0.064862	0.062369	0.069048	0.071403	0.066499	0.063919	
min	-0.172609	-0.197845	-0.212038	-0.207831	-0.214098	-0.206485	
25%	-0.029117	-0.039812	-0.044056	-0.039015	-0.038491	-0.046357	
50%	0.012976	0.000426	0.001510	0.006441	0.001411	-0.004096	
75%	0.055124	0.042051	0.042592	0.050515	0.048045	0.037379	
max	0.218983	0.192632	0.202796	0.212975	0.197265	0.195926	

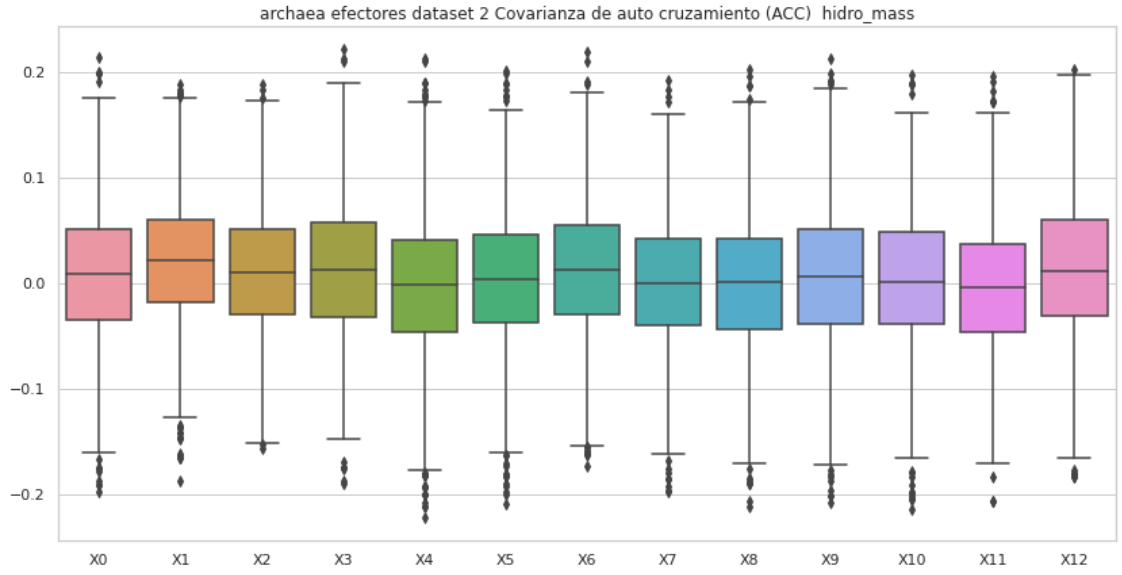
	X12
count	926.000000
mean	0.014438
std	0.066416
min	-0.183306
25%	-0.031005
50%	0.010783
75%	0.060475
max	0.201889

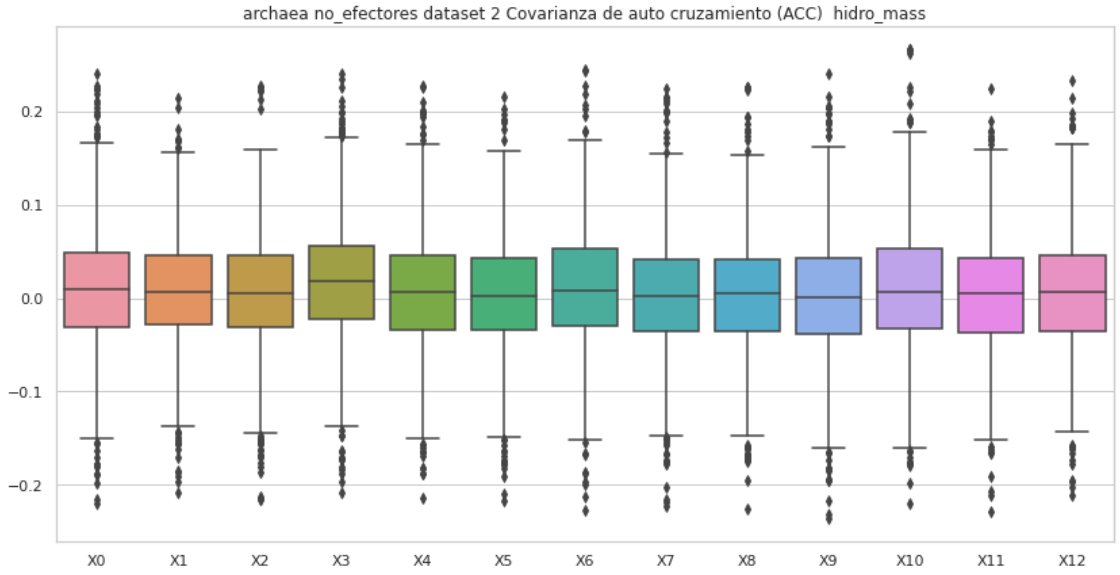
no_efectores
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores archaea dataset 2,
sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.035777 -0.000696 -0.015934 -0.019013 -0.022376 -0.046919 -0.026761
2   -0.044814 -0.042353  0.053133 -0.002771  0.020951  0.056514 -0.006538
3    0.031429  0.020576 -0.008731 -0.002777  0.000790 -0.030101 -0.034517
5    0.049197  0.005522  0.043128 -0.015035 -0.005444 -0.010401  0.071239
6    0.053937  0.027326 -0.049574 -0.033971  0.001463  0.025072  0.034759
..        …         …         …         …         …         …         …
994 -0.025449 -0.056890  0.046894  0.025409  0.012662 -0.022677 -0.035210
995 -0.039702 -0.040522  0.053232 -0.119806 -0.001014 -0.029432 -0.130018
996  0.022590  0.029463 -0.039808  0.073215 -0.076629 -0.002488 -0.034594
998  0.012404  0.029974 -0.018173 -0.017288  0.021667 -0.028013 -0.044589
999 -0.012844 -0.033228 -0.087665  0.036276  0.004147  0.082489 -0.084467

           X7        X8        X9       X10       X11       X12           X13
0    0.066634  0.024145  0.020167 -0.002941  0.019606 -0.022116  no_efectores
2    0.016612 -0.000786 -0.075658  0.044028 -0.054618 -0.045816  no_efectores
3    0.014429 -0.095531  0.002909  0.020919 -0.018494 -0.003396  no_efectores
5   -0.028248 -0.044053 -0.064007 -0.025271 -0.072197 -0.012357  no_efectores
6   -0.014751 -0.001890 -0.004821 -0.018820  0.038043  0.014484  no_efectores
..        …         …         …         …         …         …           …
994 -0.016565  0.041977  0.007653  0.035522  0.052019  0.033237  no_efectores
995  0.035035  0.004656 -0.027292 -0.077032 -0.018723  0.114534  no_efectores
996  0.009819  0.074126  0.075080 -0.096116 -0.003635 -0.008089  no_efectores
998 -0.017176 -0.055874 -0.031499 -0.085771 -0.051345 -0.046825  no_efectores
999  0.078884 -0.070880  0.100605 -0.117218  0.019784  0.034981  no_efectores

[903 rows x 14 columns]


Covarianza de auto cruzamiento (ACC)  hidro_mass no_efectores archaea dataset 2,
sin valores atípicos.
Estadísticas.

                X0          X1          X2          X3          X4          X5  \
count  903.000000  903.000000  903.000000  903.000000  903.000000  903.000000
mean     0.009535    0.007531    0.005321    0.017312    0.006992    0.002773
std      0.067551    0.062796    0.065197    0.067127    0.066670    0.064903
min     -0.220305   -0.208385   -0.215590   -0.208090   -0.213871   -0.217250
25%     -0.031123   -0.028954   -0.031178   -0.022527   -0.033847   -0.033954
50%      0.009067    0.006795    0.005018    0.018015    0.007251    0.002566
75%      0.048879    0.046118    0.045702    0.055418    0.045599    0.043538
max      0.240638    0.214026    0.226836    0.240338    0.228089    0.216097

                X6          X7          X8          X9          X10         X11  \
count  903.000000  903.000000  903.000000  903.000000  903.000000  903.000000
mean     0.011072    0.002799    0.002850    0.002273    0.009535    0.001669
std      0.067272    0.066171    0.064306    0.068176    0.069460    0.065298
```

min	-0.228096	-0.222412	-0.226646	-0.236099	-0.219484	-0.228819
25%	-0.029573	-0.034949	-0.034849	-0.038830	-0.032715	-0.037046
50%	0.008573	0.001542	0.004656	0.000484	0.007203	0.004652
75%	0.052625	0.041109	0.041284	0.043123	0.052870	0.043457
max	0.245329	0.225093	0.226086	0.240719	0.266594	0.224872

	X12
count	903.000000
mean	0.006205
std	0.064586
min	-0.211995
25%	-0.035155
50%	0.007079
75%	0.045914
max	0.233949





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))
```

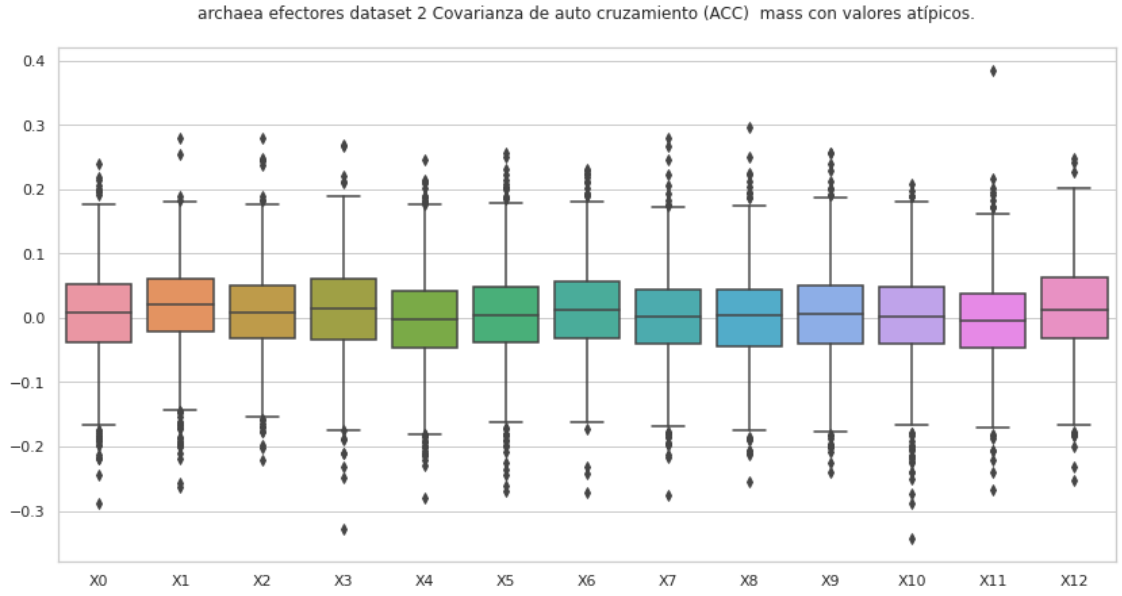
```
efectores
Covarianza de auto cruzamiento (ACC)  mass efectores archaea dataset 2, con
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.005443 -0.018635 -0.084997 -0.007834  0.002934 -0.015078  0.046394
1   -0.027923 -0.077917  0.055378 -0.034381 -0.085988  0.008007  0.077016
2    0.023057  0.030394  0.147859  0.089264 -0.045016  0.050764  0.016967
3    0.007678 -0.063872  0.030566 -0.079442 -0.095993 -0.145327  0.064974
4    0.039778 -0.020321 -0.075481 -0.036650 -0.078937 -0.097488 -0.092399
..        …        …        …        …        …        …        …
995 -0.013519 -0.003511  0.034309  0.041849  0.003351  0.099324 -0.014219
996 -0.001671 -0.001633 -0.062107 -0.074465  0.104763 -0.001856  0.164325
997 -0.145829  0.071592 -0.035244 -0.073373 -0.002276 -0.028309 -0.012430
998 -0.097416 -0.001915 -0.090449  0.010720 -0.042461  0.079594 -0.040465
999  0.030004  0.041750  0.013307  0.039298 -0.071875  0.085857 -0.022942

           X7        X8        X9       X10       X11       X12        X13
0    0.060522  0.186961 -0.000218 -0.088922 -0.139434  0.063320  efectores
1   -0.032107  0.047364  0.050714 -0.045058 -0.059875  0.083870  efectores
2   -0.046013  0.039564  0.024939 -0.073187 -0.004072  0.081296  efectores
3    0.045862 -0.113507  0.044521  0.121701  0.063497 -0.073121  efectores
4   -0.002665  0.021018 -0.008155  0.070357  0.105667  0.017713  efectores
..        …        …        …        …        …        …        …
995  0.079241 -0.041507 -0.047778 -0.001032  0.002203  0.050089  efectores
996  0.037454 -0.024362  0.045248 -0.047640 -0.025376 -0.010043  efectores
997  0.051975 -0.013730 -0.057041  0.023549 -0.031107  0.068816  efectores
998  0.088117 -0.038272  0.061181 -0.078103  0.081104 -0.083219  efectores
999  0.036628 -0.117445  0.053187  0.001043  0.044585  0.053890  efectores

[1000 rows x 14 columns]


Covarianza de auto cruzamiento (ACC)  mass efectores archaea dataset 2, con
valores atípicos.
Estadísticas.

                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.005704     0.019238     0.009245     0.011975    -0.002747
```

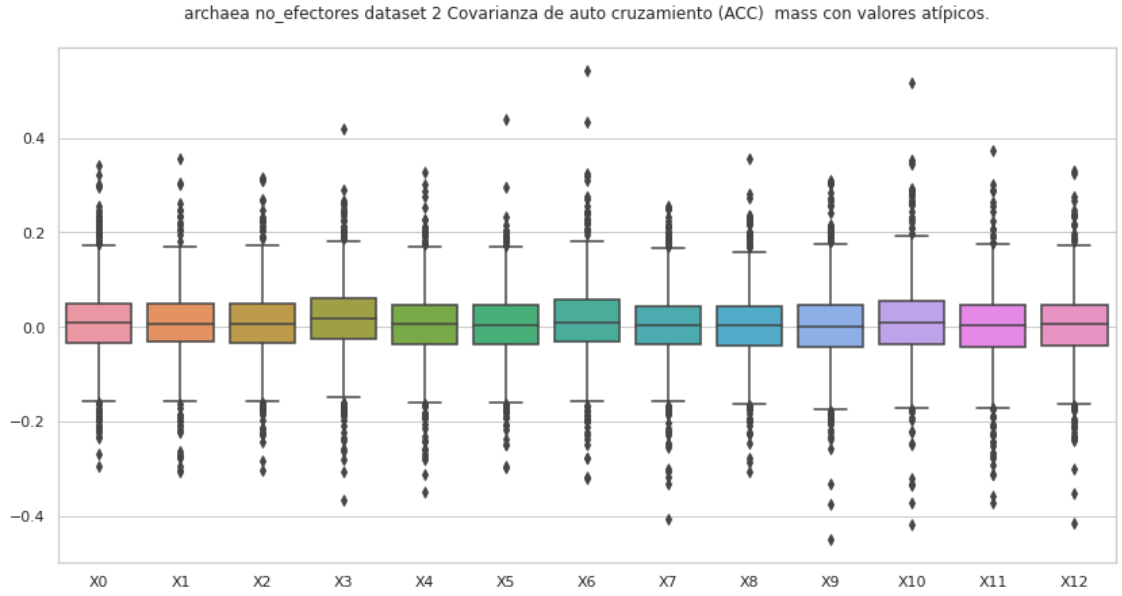
```
std       0.071498     0.069503     0.067815     0.072728     0.074425
min      -0.288239    -0.263362    -0.221731    -0.328625    -0.279168
25%      -0.037723    -0.021341    -0.031696    -0.033717    -0.046670
50%       0.007683     0.020929     0.008020     0.014001    -0.001834
75%       0.052304     0.060322     0.051280     0.060371     0.042571
max       0.240172     0.280200     0.279326     0.267981     0.246180

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.002615     0.011469     0.000510     0.000965     0.005450
std       0.071049     0.069775     0.067489     0.072901     0.074874
min      -0.268516    -0.272288    -0.275302    -0.254238    -0.240528
25%      -0.038933    -0.031756    -0.040692    -0.045458    -0.040829
50%       0.004592     0.012231     0.001178     0.003485     0.006540
75%       0.048291     0.055993     0.044875     0.044636     0.050723
max       0.255310     0.230651     0.279565     0.296241     0.257050

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.002491    -0.004675     0.015066
std       0.072246     0.068194     0.069542
min      -0.343092    -0.266614    -0.251885
25%      -0.040014    -0.046998    -0.031421
50%       0.001269    -0.004096     0.011460
75%       0.048002     0.037843     0.062672
max       0.208882     0.384496     0.247189



no_efectores
Covarianza de auto cruzamiento (ACC)  mass no_efectores archaea dataset 2, con
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.035777 -0.000696 -0.015934 -0.019013 -0.022376 -0.046919 -0.026761
1   -0.095522  0.020359 -0.197128  0.266886 -0.156052  0.053897 -0.195423
2   -0.044814 -0.042353  0.053133 -0.002771  0.020951  0.056514 -0.006538
3    0.031429  0.020576 -0.008731 -0.002777  0.000790 -0.030101 -0.034517
4    0.032967 -0.002815  0.092848 -0.000931 -0.087647 -0.094754 -0.092877
..        …       …       …       …       …       …       …
995 -0.039702 -0.040522  0.053232 -0.119806 -0.001014 -0.029432 -0.130018
996  0.022590  0.029463 -0.039808  0.073215 -0.076629 -0.002488 -0.034594
997 -0.056812  0.129133 -0.025008  0.066403 -0.015866  0.192747 -0.023293
998  0.012404  0.029974 -0.018173 -0.017288  0.021667 -0.028013 -0.044589
999 -0.012844 -0.033228 -0.087665  0.036276  0.004147  0.082489 -0.084467

           X7        X8        X9       X10       X11       X12           X13
```

```
0    0.066634  0.024145  0.020167 -0.002941  0.019606 -0.022116  no_efectores
1    0.108045  0.094838 -0.034535 -0.090834  0.072988  0.024905  no_efectores
2    0.016612 -0.000786 -0.075658  0.044028 -0.054618 -0.045816  no_efectores
3    0.014429 -0.095531  0.002909  0.020919 -0.018494 -0.003396  no_efectores
4   -0.253573 -0.154219 -0.035814  0.002671  0.030237  0.049257  no_efectores
..        …       …       …       …       …       …           …
995  0.035035  0.004656 -0.027292 -0.077032 -0.018723  0.114534  no_efectores
996  0.009819  0.074126  0.075080 -0.096116 -0.003635 -0.008089  no_efectores
997  0.130843 -0.226724  0.185644 -0.062958  0.123454 -0.082783  no_efectores
998 -0.017176 -0.055874 -0.031499 -0.085771 -0.051345 -0.046825  no_efectores
999  0.078884 -0.070880  0.100605 -0.117218  0.019784  0.034981  no_efectores

[1000 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) mass no_efectores archaea dataset 2, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.009371     0.007519     0.006046     0.017315     0.004939
std       0.077181     0.074947     0.074478     0.078219     0.079044
min      -0.293860    -0.304968    -0.303053    -0.367565    -0.350528
25%      -0.033233    -0.032132    -0.034023    -0.024565    -0.035928
50%       0.008294     0.006587     0.005147     0.018026     0.006328
75%       0.049987     0.050546     0.048637     0.059594     0.047324
max       0.340503     0.356153     0.315695     0.420034     0.325809

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.002858     0.012934     0.002281     0.002076     0.003786
std       0.074990     0.081209     0.077353     0.076281     0.081353
min      -0.298457    -0.319387    -0.405955    -0.307032    -0.449415
25%      -0.037419    -0.030827    -0.037889    -0.040296    -0.041708
50%       0.002629     0.009409     0.002361     0.004266     0.001120
75%       0.045782     0.056712     0.044381     0.043278     0.047040
max       0.440056     0.541764     0.255430     0.355540     0.309617

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.011002     0.000722     0.004145
std       0.085426     0.080000     0.077686
min      -0.416512    -0.372297    -0.415588
25%      -0.035977    -0.041123    -0.038337
50%       0.007642     0.003644     0.006109
75%       0.055579     0.045960     0.047405
max       0.515834     0.373322     0.329582
```


archaea efectores dataset 2 Covarianza de auto cruzamiento (ACC) mass con valores atípicos.
0.4
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 2 Covarianza de auto cruzamiento (ACC) mass con valores atípicos.
0.4
0.2
0.0
−0.2
−0.4
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12

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

	X0	X1	X2	X3	X4	X5	X6	\
0	-0.005443	-0.018635	-0.084997	-0.007834	0.002934	-0.015078	0.046394	
1	-0.027923	-0.077917	0.055378	-0.034381	-0.085988	0.008007	0.077016	
2	0.023057	0.030394	0.147859	0.089264	-0.045016	0.050764	0.016967	
3	0.007678	-0.063872	0.030566	-0.079442	-0.095993	-0.145327	0.064974	
4	0.039778	-0.020321	-0.075481	-0.036650	-0.078937	-0.097488	-0.092399	
..	…	…	…	…	…	…	…	
995	-0.013519	-0.003511	0.034309	0.041849	0.003351	0.099324	-0.014219	
996	-0.001671	-0.001633	-0.062107	-0.074465	0.104763	-0.001856	0.164325	
997	-0.145829	0.071592	-0.035244	-0.073373	-0.002276	-0.028309	-0.012430	
998	-0.097416	-0.001915	-0.090449	0.010720	-0.042461	0.079594	-0.040465	
999	0.030004	0.041750	0.013307	0.039298	-0.071875	0.085857	-0.022942	

	X7	X8	X9	X10	X11	X12	X13
0	0.060522	0.186961	-0.000218	-0.088922	-0.139434	0.063320	efectores
1	-0.032107	0.047364	0.050714	-0.045058	-0.059875	0.083870	efectores
2	-0.046013	0.039564	0.024939	-0.073187	-0.004072	0.081296	efectores
3	0.045862	-0.113507	0.044521	0.121701	0.063497	-0.073121	efectores
4	-0.002665	0.021018	-0.008155	0.070357	0.105667	0.017713	efectores
..	…	…	…	…	…	…	…
995	0.079241	-0.041507	-0.047778	-0.001032	0.002203	0.050089	efectores
996	0.037454	-0.024362	0.045248	-0.047640	-0.025376	-0.010043	efectores
997	0.051975	-0.013730	-0.057041	0.023549	-0.031107	0.068816	efectores
998	0.088117	-0.038272	0.061181	-0.078103	0.081104	-0.083219	efectores
999	0.036628	-0.117445	0.053187	0.001043	0.044585	0.053890	efectores

```
[926 rows x 14 columns]
```

```
Covarianza de auto cruzamiento (ACC)  mass efectores archaea dataset 2, sin
valores atípicos.
Estadísticas.
```

	X0	X1	X2	X3	X4	X5	\
count	926.000000	926.000000	926.000000	926.000000	926.000000	926.000000	
mean	0.007371	0.021685	0.010584	0.011301	-0.003145	0.001796	
std	0.066184	0.062369	0.062037	0.067484	0.070588	0.065604	
min	-0.197650	-0.186867	-0.156994	-0.190481	-0.221905	-0.208499	
25%	-0.035197	-0.017638	-0.029985	-0.032357	-0.046193	-0.037374	
50%	0.008979	0.021272	0.010331	0.012985	-0.001834	0.003712	
75%	0.051020	0.060177	0.051251	0.057915	0.041146	0.045863	

```
max      0.213739    0.188358    0.188834    0.221510    0.212889    0.201266

               X6          X7          X8          X9         X10         X11  \
count  926.000000  926.000000  926.000000  926.000000  926.000000  926.000000
mean     0.012145   -0.000545    0.000170    0.005622    0.004908   -0.004472
std      0.064862    0.062369    0.069048    0.071403    0.066499    0.063919
min     -0.172609   -0.197845   -0.212038   -0.207831   -0.214098   -0.206485
25%     -0.029117   -0.039812   -0.044056   -0.039015   -0.038491   -0.046357
50%      0.012976    0.000426    0.001510    0.006441    0.001411   -0.004096
75%      0.055124    0.042051    0.042592    0.050515    0.048045    0.037379
max      0.218983    0.192632    0.202796    0.212975    0.197265    0.195926

              X12
count  926.000000
mean     0.014438
std      0.066416
min     -0.183306
25%     -0.031005
50%      0.010783
75%      0.060475
max      0.201889
```

Covarianza de auto cruzamiento (ACC) mass no_efectores archaea dataset 2, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.035777 -0.000696 -0.015934 -0.019013 -0.022376 -0.046919 -0.026761
2   -0.044814 -0.042353  0.053133 -0.002771  0.020951  0.056514 -0.006538
3    0.031429  0.020576 -0.008731 -0.002777  0.000790 -0.030101 -0.034517
5    0.049197  0.005522  0.043128 -0.015035 -0.005444 -0.010401  0.071239
6    0.053937  0.027326 -0.049574 -0.033971  0.001463  0.025072  0.034759
..        …         …         …         …         …         …         …
994 -0.025449 -0.056890  0.046894  0.025409  0.012662 -0.022677 -0.035210
995 -0.039702 -0.040522  0.053232 -0.119806 -0.001014 -0.029432 -0.130018
996  0.022590  0.029463 -0.039808  0.073215 -0.076629 -0.002488 -0.034594
998  0.012404  0.029974 -0.018173 -0.017288  0.021667 -0.028013 -0.044589
999 -0.012844 -0.033228 -0.087665  0.036276  0.004147  0.082489 -0.084467

           X7        X8        X9       X10       X11       X12           X13
0    0.066634  0.024145  0.020167 -0.002941  0.019606 -0.022116  no_efectores
2    0.016612 -0.000786 -0.075658  0.044028 -0.054618 -0.045816  no_efectores
3    0.014429 -0.095531  0.002909  0.020919 -0.018494 -0.003396  no_efectores
5   -0.028248 -0.044053 -0.064007 -0.025271 -0.072197 -0.012357  no_efectores
6   -0.014751 -0.001890 -0.004821 -0.018820  0.038043  0.014484  no_efectores
..        …         …         …         …         …         …           …
```

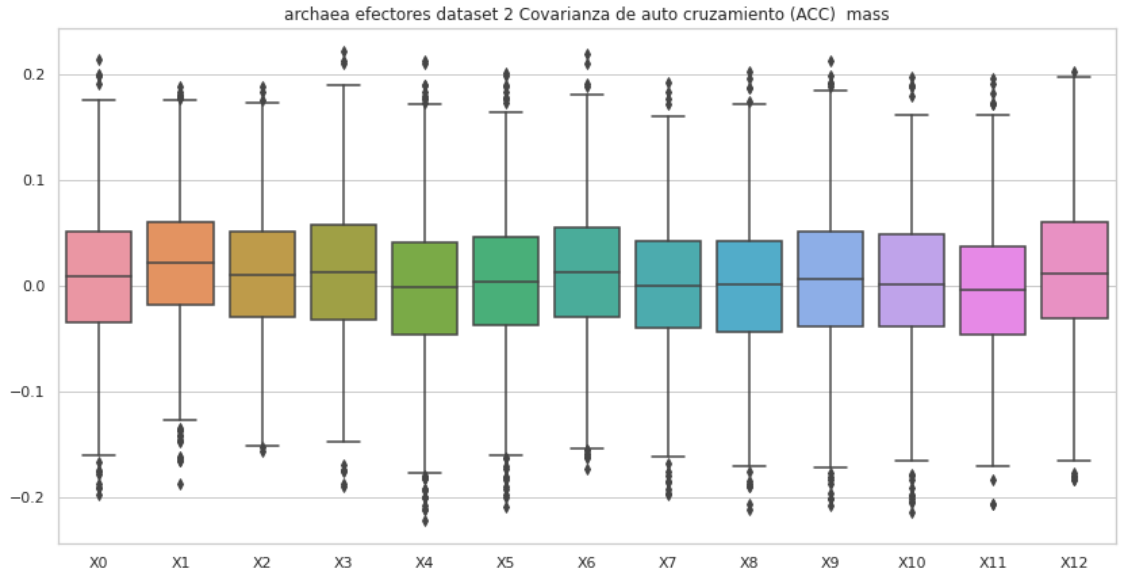
```
994 -0.016565  0.041977  0.007653  0.035522  0.052019  0.033237  no_efectores
995  0.035035  0.004656 -0.027292 -0.077032 -0.018723  0.114534  no_efectores
996  0.009819  0.074126  0.075080 -0.096116 -0.003635 -0.008089  no_efectores
998 -0.017176 -0.055874 -0.031499 -0.085771 -0.051345 -0.046825  no_efectores
999  0.078884 -0.070880  0.100605 -0.117218  0.019784  0.034981  no_efectores

[903 rows x 14 columns]
```

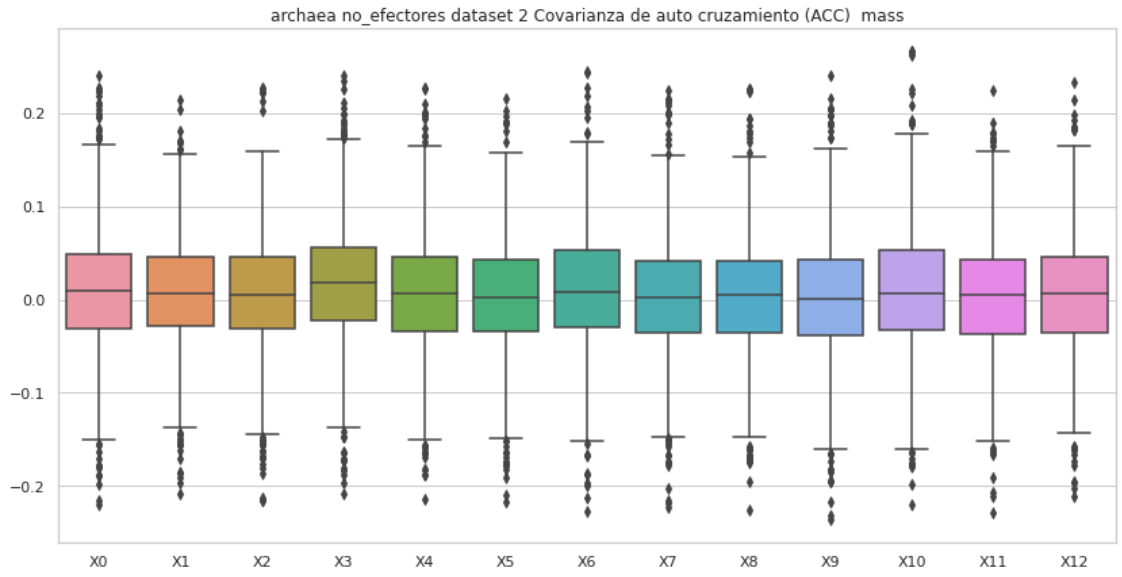
Covarianza de auto cruzamiento (ACC) mass no_efectores archaea dataset 2, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5
count	903.000000	903.000000	903.000000	903.000000	903.000000	903.000000
mean	0.009535	0.007531	0.005321	0.017312	0.006992	0.002773
std	0.067551	0.062796	0.065197	0.067127	0.066670	0.064903
min	-0.220305	-0.208385	-0.215590	-0.208090	-0.213871	-0.217250
25%	-0.031123	-0.028954	-0.031178	-0.022527	-0.033847	-0.033954
50%	0.009067	0.006795	0.005018	0.018015	0.007251	0.002566
75%	0.048879	0.046118	0.045702	0.055418	0.045599	0.043538
max	0.240638	0.214026	0.226836	0.240338	0.228089	0.216097

	X6	X7	X8	X9	X10	X11
count	903.000000	903.000000	903.000000	903.000000	903.000000	903.000000
mean	0.011072	0.002799	0.002850	0.002273	0.009535	0.001669
std	0.067272	0.066171	0.064306	0.068176	0.069460	0.065298
min	-0.228096	-0.222412	-0.226646	-0.236099	-0.219484	-0.228819
25%	-0.029573	-0.034949	-0.034849	-0.038830	-0.032715	-0.037046
50%	0.008573	0.001542	0.004656	0.000484	0.007203	0.004652
75%	0.052625	0.041109	0.041284	0.043123	0.052870	0.043457
max	0.245329	0.225093	0.226086	0.240719	0.266594	0.224872

	X12
count	903.000000
mean	0.006205
std	0.064586
min	-0.211995
25%	-0.035155
50%	0.007079
75%	0.045914
max	0.233949





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 2, con
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.082525  0.046163  0.024685  0.021231 -0.042095 -0.003890 -0.100917
1    0.131098 -0.028597  0.154758  0.169085  0.117067 -0.021076  0.101543
2    0.048496  0.025507  0.064423  0.102697  0.002263 -0.015332  0.070060
3    0.082444  0.074506  0.060104  0.117511  0.026384 -0.051381 -0.055270
4    0.143698 -0.160690  0.019492  0.012487 -0.136476 -0.156022  0.014325
..        …         …         …         …         …         …         …
995  0.083275  0.046875  0.038874  0.075819  0.068174  0.100438 -0.014336
996  0.042503  0.000670  0.003545 -0.058447  0.075533  0.008006  0.018239
997  0.069936  0.135565  0.145020  0.161899 -0.024660  0.000620 -0.040247
998 -0.095351 -0.158566  0.032330  0.003219 -0.159586  0.083778  0.121088
999  0.078567  0.066573  0.103448  0.041503  0.078296  0.084101  0.034307

           X7        X8        X9       X10       X11       X12        X13
0    0.028626 -0.081452  0.026419 -0.010840 -0.047991 -0.099049  efectores
1    0.125102  0.025473  0.007859  0.072365  0.078935  0.030298  efectores
2    0.103424  0.105309  0.009555 -0.032279  0.012647  0.012509  efectores
```

```
3    0.002732  0.026700  0.023177 -0.012665 -0.004397  0.001176  efectores
4    0.104235 -0.048447 -0.086605 -0.005546 -0.060541  0.086582  efectores
..        …       …        …        …        …         …
995  0.122905  0.025363  0.094937 -0.064848  0.032317  0.075783  efectores
996  0.027054 -0.001145  0.041802 -0.004403 -0.029137 -0.019570  efectores
997  0.008024 -0.084908 -0.021080 -0.001695  0.005089  0.016472  efectores
998 -0.051206  0.035123 -0.123006 -0.003184  0.059149  0.004523  efectores
999  0.036815 -0.018402  0.000946 -0.021201  0.033945 -0.012502  efectores

[1000 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro efectores archaea dataset 2, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.022416    -0.027177     0.041236     0.031805    -0.016834
std       0.090854     0.098804     0.080719     0.089812     0.091954
min      -0.482349    -0.451793    -0.395963    -0.393037    -0.340485
25%      -0.030823    -0.098203    -0.010366    -0.018414    -0.078013
50%       0.026057    -0.019805     0.036417     0.033950    -0.011464
75%       0.080566     0.046358     0.091070     0.082797     0.048698
max       0.318436     0.247248     0.276040     0.333588     0.230570

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean     -0.013992     0.023364     0.022759    -0.002003    -0.004559
std       0.084642     0.089712     0.085434     0.088914     0.084329
min      -0.332032    -0.335820    -0.262545    -0.330730    -0.316194
25%      -0.064130    -0.028175    -0.027099    -0.052243    -0.055360
50%      -0.011181     0.023130     0.019810     0.004627    -0.006684
75%       0.040906     0.070802     0.072625     0.049280     0.045162
max       0.245337     0.344254     0.349116     0.278329     0.265349

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.016247     0.010315    -0.012921
std       0.082501     0.077042     0.082136
min      -0.344083    -0.298290    -0.339636
25%      -0.037159    -0.039722    -0.059192
50%       0.010075     0.004698    -0.010123
75%       0.066666     0.052025     0.034792
max       0.332047     0.364189     0.452714
```

no_efectores
Covarianza de auto cruzamiento (ACC) hidro no_efectores archaea dataset 2, con valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.032461	0.022009	0.177854	0.077261	0.034890	0.075708	0.116846	
1	0.010786	-0.322060	-0.027323	0.160021	0.092085	-0.155797	-0.069362	
2	0.066106	0.050517	0.030738	0.062816	0.027098	0.005305	-0.035442	
3	-0.059433	0.010710	-0.027360	-0.033702	0.031347	0.008599	-0.059248	
4	0.153921	0.045357	0.104230	0.172594	0.064788	0.017686	0.009631	
..	…	…	…	…	…	…	…	
995	0.092793	0.068323	-0.073672	-0.169312	-0.161854	-0.010420	0.032801	
996	0.028910	-0.013692	0.001550	0.082903	0.032692	-0.032891	-0.007863	
997	-0.260024	0.070355	0.004770	-0.073108	-0.094716	0.057829	0.003603	
998	-0.145443	0.044709	-0.072042	0.040519	0.101831	-0.255668	0.098346	
999	0.124393	0.077687	0.042291	0.000034	0.071956	0.100183	0.055470	

	X7	X8	X9	X10	X11	X12	X13
0	-0.001791	0.072028	0.095803	0.049221	0.031206	0.037376	no_efectores
1	0.008366	0.089335	-0.021564	-0.146201	-0.004791	0.133546	no_efectores
2	-0.039632	-0.063186	-0.046210	0.006160	-0.003085	0.012846	no_efectores
3	0.115201	-0.029472	-0.003064	-0.064850	-0.016222	0.003234	no_efectores
4	0.016312	0.125561	0.088528	-0.005317	0.026291	-0.005096	no_efectores
..	…	…	…	…	…	…	…
995	0.055359	-0.044522	-0.087217	-0.038311	-0.179028	-0.105573	no_efectores
996	-0.012971	0.052861	0.010897	-0.037331	0.004585	0.007479	no_efectores
997	-0.067317	-0.019576	-0.009490	-0.007989	-0.066225	0.092291	no_efectores
998	0.091650	-0.038942	0.084660	-0.265487	0.175137	0.155495	no_efectores
999	-0.044868	0.040739	-0.013146	-0.008906	0.044020	-0.051066	no_efectores

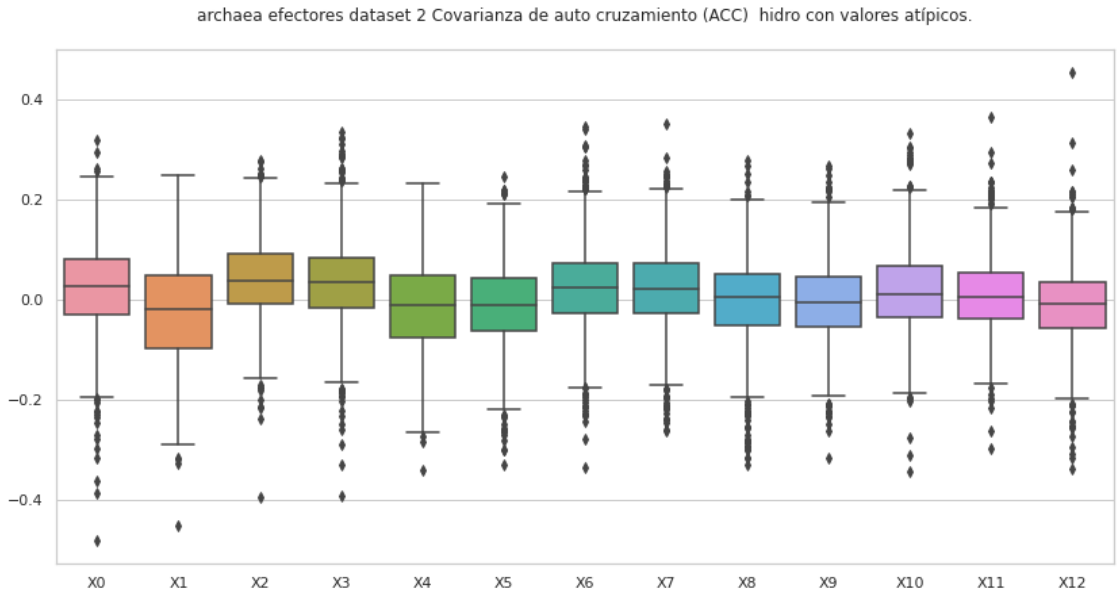
[1000 rows x 14 columns]

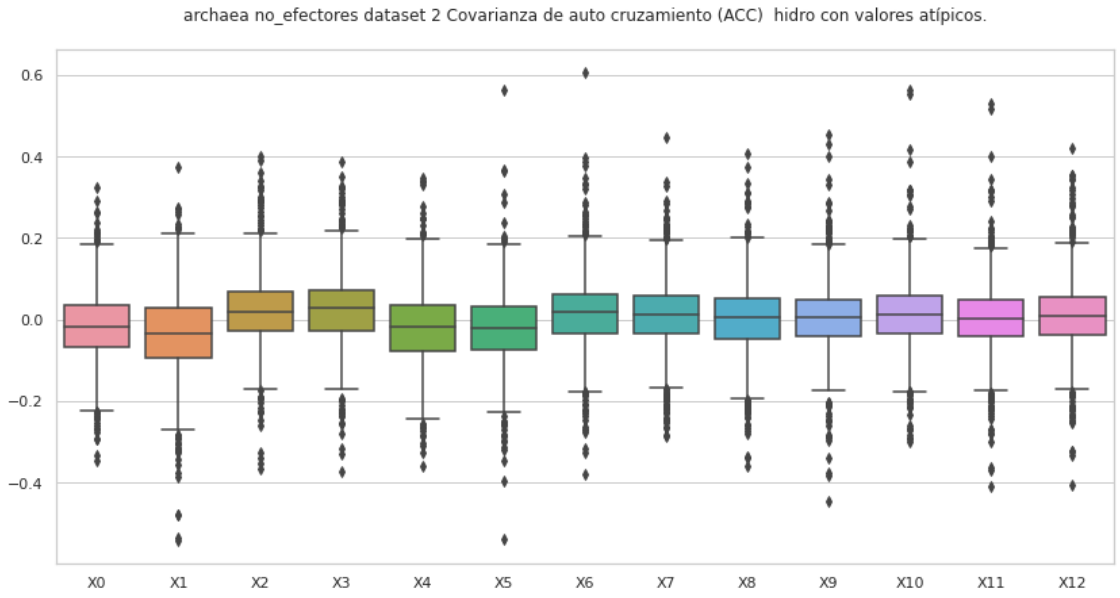
Covarianza de auto cruzamiento (ACC) hidro no_efectores archaea dataset 2, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	-0.016908	-0.036039	0.020721	0.025418	-0.019013	
std	0.088021	0.101356	0.089305	0.089276	0.090524	
min	-0.347686	-0.542357	-0.367266	-0.372261	-0.358051	
25%	-0.068555	-0.093811	-0.028358	-0.026007	-0.077334	
50%	-0.017519	-0.035323	0.018747	0.027486	-0.018387	
75%	0.035313	0.028479	0.068184	0.073294	0.036039	
max	0.322970	0.374734	0.399567	0.386499	0.347650	

	X5	X6	X7	X8	X9	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	-0.022604	0.016404	0.011842	0.003378	0.005516	
std	0.091468	0.090587	0.085565	0.090218	0.088171	
min	-0.538433	-0.380581	-0.286211	-0.361171	-0.447172	
25%	-0.073849	-0.032636	-0.033162	-0.046627	-0.039706	
50%	-0.020459	0.017785	0.012622	0.005343	0.004895	
75%	0.032708	0.063286	0.059246	0.052583	0.050083	
max	0.564228	0.606488	0.446902	0.408606	0.452843	

	X10	X11	X12
count	1000.000000	1000.000000	1000.000000
mean	0.012774	0.003471	0.009351
std	0.089567	0.087545	0.088827
min	-0.298566	-0.410456	-0.405330
25%	-0.035651	-0.040659	-0.037075
50%	0.011271	0.002727	0.007353
75%	0.058848	0.047641	0.054093
max	0.564747	0.531087	0.420779





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 2, sin valores
atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.082525  0.046163  0.024685  0.021231 -0.042095 -0.003890 -0.100917
1    0.131098 -0.028597  0.154758  0.169085  0.117067 -0.021076  0.101543
2    0.048496  0.025507  0.064423  0.102697  0.002263 -0.015332  0.070060
3    0.082444  0.074506  0.060104  0.117511  0.026384 -0.051381 -0.055270
4    0.143698 -0.160690  0.019492  0.012487 -0.136476 -0.156022  0.014325
..        …       …       …       …       …       …       …
995  0.083275  0.046875  0.038874  0.075819  0.068174  0.100438 -0.014336
996  0.042503  0.000670  0.003545 -0.058447  0.075533  0.008006  0.018239
997  0.069936  0.135565  0.145020  0.161899 -0.024660  0.000620 -0.040247
998 -0.095351 -0.158566  0.032330  0.003219 -0.159586  0.083778  0.121088
999  0.078567  0.066573  0.103448  0.041503  0.078296  0.084101  0.034307

           X7        X8        X9       X10       X11       X12        X13
0    0.028626 -0.081452  0.026419 -0.010840 -0.047991 -0.099049  efectores
1    0.125102  0.025473  0.007859  0.072365  0.078935  0.030298  efectores
2    0.103424  0.105309  0.009555 -0.032279  0.012647  0.012509  efectores
3    0.002732  0.026700  0.023177 -0.012665 -0.004397  0.001176  efectores
4    0.104235 -0.048447 -0.086605 -0.005546 -0.060541  0.086582  efectores
..        …       …       …       …       …       …       …
995  0.122905  0.025363  0.094937 -0.064848  0.032317  0.075783  efectores
```

```
996  0.027054 -0.001145  0.041802 -0.004403 -0.029137 -0.019570  efectores
997  0.008024 -0.084908 -0.021080 -0.001695  0.005089  0.016472  efectores
998 -0.051206  0.035123 -0.123006 -0.003184  0.059149  0.004523  efectores
999  0.036815 -0.018402  0.000946 -0.021201  0.033945 -0.012502  efectores

[923 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) efectores archaea dataset 2, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	923.000000	923.000000	923.000000	923.000000	923.000000	923.000000	
mean	0.024440	-0.025040	0.041724	0.033677	-0.014338	-0.011952	
std	0.081677	0.094429	0.075555	0.082945	0.087898	0.078802	
min	-0.248119	-0.319638	-0.200124	-0.222470	-0.284337	-0.266503	
25%	-0.027633	-0.093335	-0.008513	-0.014839	-0.074750	-0.061750	
50%	0.026601	-0.018713	0.035678	0.035428	-0.007384	-0.009935	
75%	0.077641	0.046269	0.088384	0.081938	0.049632	0.040050	
max	0.293358	0.223364	0.276040	0.295772	0.230570	0.218831	

	X6	X7	X8	X9	X10	X11	\
count	923.000000	923.000000	923.000000	923.000000	923.000000	923.000000	
mean	0.022185	0.021336	0.001532	-0.004868	0.011687	0.009302	
std	0.081656	0.078342	0.079755	0.078353	0.073094	0.070344	
min	-0.243809	-0.229219	-0.257562	-0.236754	-0.202894	-0.204386	
25%	-0.025866	-0.026496	-0.048670	-0.053088	-0.037868	-0.038950	
50%	0.022082	0.018699	0.005748	-0.006666	0.006511	0.004429	
75%	0.065754	0.069128	0.047667	0.041942	0.059005	0.048244	
max	0.276649	0.254334	0.250899	0.235065	0.226964	0.234151	

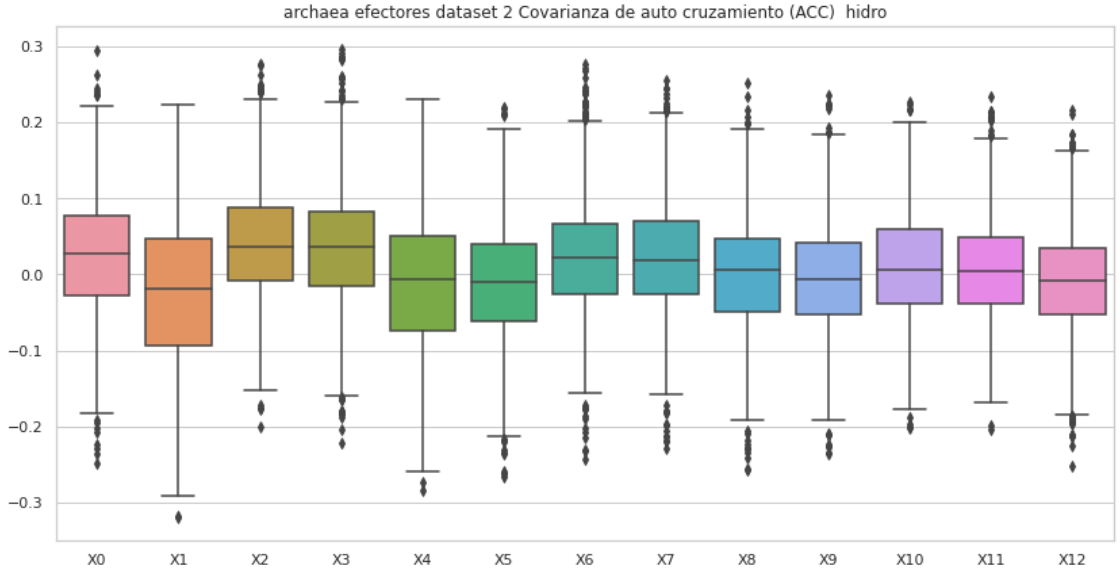
	X12
count	923.000000
mean	-0.010041
std	0.071509
min	-0.252704
25%	-0.053520
50%	-0.008648
75%	0.034399
max	0.216091

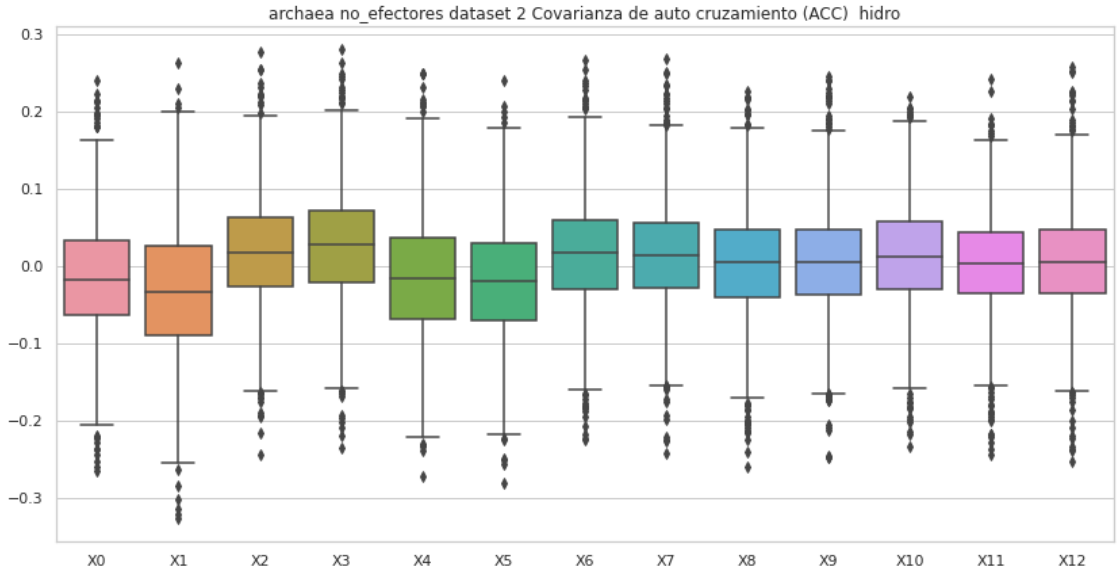
no_efectores
Covarianza de auto cruzamiento (ACC) no_efectores archaea dataset 2, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.032461  0.022009  0.177854  0.077261  0.034890  0.075708  0.116846
1    0.010786 -0.322060 -0.027323  0.160021  0.092085 -0.155797 -0.069362
2    0.066106  0.050517  0.030738  0.062816  0.027098  0.005305 -0.035442
3   -0.059433  0.010710 -0.027360 -0.033702  0.031347  0.008599 -0.059248
4    0.153921  0.045357  0.104230  0.172594  0.064788  0.017686  0.009631
..        …         …         …         …         …         …
994 -0.133073 -0.014576  0.097859 -0.092274  0.032873  0.014241  0.018271
995  0.092793  0.068323 -0.073672 -0.169312 -0.161854 -0.010420  0.032801
996  0.028910 -0.013692  0.001550  0.082903  0.032692 -0.032891 -0.007863
997 -0.260024  0.070355  0.004770 -0.073108 -0.094716  0.057829  0.003603
999  0.124393  0.077687  0.042291  0.000034  0.071956  0.100183  0.055470

           X7        X8        X9       X10       X11       X12           X13
0   -0.001791  0.072028  0.095803  0.049221  0.031206  0.037376  no_efectores
1    0.008366  0.089335 -0.021564 -0.146201 -0.004791  0.133546  no_efectores
2   -0.039632 -0.063186 -0.046210  0.006160 -0.003085  0.012846  no_efectores
3    0.115201 -0.029472 -0.003064 -0.064850 -0.016222  0.003234  no_efectores
4    0.016312  0.125561  0.088528 -0.005317  0.026291 -0.005096  no_efectores
..        …         …         …         …         …         …
994  0.042299  0.023294  0.009120  0.024626 -0.053569 -0.017893  no_efectores
995  0.055359 -0.044522 -0.087217 -0.038311 -0.179028 -0.105573  no_efectores
996 -0.012971  0.052861  0.010897 -0.037331  0.004585  0.007479  no_efectores
997 -0.067317 -0.019576 -0.009490 -0.007989 -0.066225  0.092291  no_efectores
999 -0.044868  0.040739 -0.013146 -0.008906  0.044020 -0.051066  no_efectores

[897 rows x 14 columns]


Covarianza de auto cruzamiento (ACC)  no_efectores archaea dataset 2, sin
valores atípicos.
Estadísticas.

                X0          X1          X2          X3          X4          X5  \
count  897.000000  897.000000  897.000000  897.000000  897.000000  897.000000
mean    -0.014938   -0.031844    0.017458    0.025904   -0.015535   -0.021183
std      0.079249    0.085270    0.074446    0.075831    0.080206    0.077346
min     -0.265977   -0.326886   -0.245248   -0.235512   -0.272844   -0.281807
25%     -0.063284   -0.089209   -0.027224   -0.021176   -0.068286   -0.071033
50%     -0.017084   -0.033350    0.017686    0.027787   -0.016087   -0.020243
75%      0.032461    0.026253    0.062505    0.071081    0.035984    0.029229
max      0.239243    0.263720    0.276127    0.280338    0.249650    0.239404

                X6          X7          X8          X9         X10         X11  \
count  897.000000  897.000000  897.000000  897.000000  897.000000  897.000000
mean     0.015303    0.013758    0.002744    0.005826    0.011205    0.002453
std      0.074498    0.071695    0.074771    0.072968    0.072053    0.069137
```

min	-0.224875	-0.241782	-0.260773	-0.248753	-0.234702	-0.244770
25%	-0.029784	-0.028674	-0.041061	-0.037985	-0.031086	-0.036145
50%	0.017839	0.012877	0.004998	0.004901	0.011228	0.002631
75%	0.060172	0.055631	0.047786	0.047142	0.056682	0.044116
max	0.266254	0.267804	0.226734	0.246069	0.218741	0.242076

	X12
count	897.000000
mean	0.006736
std	0.071889
min	-0.252695
25%	-0.036026
50%	0.005721
75%	0.047596
max	0.258051




archaea no_efectores dataset 2 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