ds3_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 = 3
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 3, con valores
atípicos.
Valores del documento csv.

         X0     X1     X2     X3     X4      X5     X6     X7     X8      X9  \
0     4.237  7.627  2.542  5.932  0.000  15.254  3.390  7.627  1.695   5.085
1    19.792  5.208  2.083  5.208  2.083   0.000  2.083  9.375  1.042   7.292
2     6.486  4.324  5.946  3.243  2.162   5.405  2.162  9.730  0.541   7.027
3     8.108  5.405  4.054  4.054  4.054  13.514  1.351  5.405  1.351   9.459
4     2.874  4.598  6.322  5.172  1.149   7.471  1.149  6.322  3.448  12.069
..      …      …      …      …      …       …      …      …      …       …
995  10.569  6.504  4.065  8.943  0.813   8.943  0.813  6.504  0.000  10.569
996  11.610  6.367  1.873  9.738  0.000   8.614  1.498  9.738  1.498   2.247
997  15.672  5.970  1.493  7.463  0.746   3.731  1.493  8.209  3.731   1.493
998   8.397  6.107  4.580  8.397  1.527   9.924  8.397  5.344  2.290   2.290
999   1.515  6.061  3.030  7.576  0.000  12.121  1.515  3.030  0.000  12.121
```

```
     …     X11    X12    X13    X14     X15    X16    X17    X18     X19  \
0    …   2.542  3.390  3.390  2.542  11.017  3.390  1.695  1.695   5.085
1    …   2.083  1.042  5.208  4.167  13.542  5.208  2.083  0.000   6.250
2    …   7.027  2.162  7.568  4.324   6.486  5.946  1.622  6.486   4.324
3    …   9.459  4.054  0.000  5.405   4.054  6.757  1.351  0.000   9.459
4    …  13.793  0.575  6.897  3.448   2.874  1.724  0.575  8.046   6.897
..   …     …    …    …    …     …    …    …    …     …
995  …   4.878  2.439  3.252  0.000   4.878  5.691  3.252  3.252   7.317
996  …   1.498  1.124  1.124  3.745   6.367  6.742  0.375  2.622  15.730
997  …   1.493  0.746  5.970  2.985   4.478  1.493  0.746  0.746  16.418
998  …   3.817  3.817  3.817  1.527   3.817  3.817  0.763  5.344   6.870
999  …  16.667  1.515  1.515  0.000   7.576  1.515  1.515  4.545   7.576

            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 3, con valores atípicos.
Estadísticas.

```
                X0          X1           X2           X3           X4  \
count  1000.000000  1000.00000  1000.000000  1000.000000  1000.000000
mean      9.531321     5.97132     2.717362     5.835994     0.632572
std       4.593026     2.64957     2.101232     2.707522     0.870401
min       0.000000     0.00000     0.000000     0.000000     0.000000
25%       5.819250     4.10250     1.198000     3.685000     0.000000
50%       9.387000     5.85700     2.165500     5.376000     0.273500
75%      12.583500     7.61400     3.892250     7.523750     0.982000
max      27.174000    17.18800    12.481000    15.942000     4.651000

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      6.887740     2.397531     7.756977     1.725395     5.749409
std       3.766571     1.444022     2.947423     1.143056     3.391303
```

min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	3.846000	1.350000	5.586250	0.754500	3.151750
50%	6.722000	2.265000	7.705500	1.613000	5.066000
75%	9.449000	3.215500	9.583000	2.521000	7.830000
max	20.661000	8.629000	21.627000	5.128000	20.000000

	X10	X11	X12	X13	X14 \
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	11.388093	4.280730	2.056576	4.026021	4.162701
std	3.271647	4.372627	1.209158	2.010717	1.780785
min	2.529000	0.000000	0.000000	0.000000	0.000000
25%	9.014500	1.017250	1.129500	2.724000	2.978500
50%	11.429000	2.279500	1.754000	3.867000	4.091000
75%	13.679500	6.953750	2.694000	5.025750	5.204500
max	25.543000	20.388000	7.634000	12.887000	11.538000

	X15	X16	X17	X18	X19
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	5.922135	5.566603	1.303707	3.297797	8.789890
std	2.144198	2.159371	1.122894	1.561664	3.560367
min	0.000000	0.000000	0.000000	0.000000	1.087000
25%	4.478000	4.054000	0.580250	2.222000	6.029500
50%	5.733500	5.480500	1.088500	3.146500	8.565500
75%	7.183000	6.757000	1.807500	4.111000	11.284000
max	15.232000	23.930000	10.326000	11.029000	19.881000

no_efectores
Composición de aminoácidos (AAC) no_efectores archaea dataset 3, con valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	X7	X8 \
0	6.250	4.688	6.250	6.250	2.344	8.594	2.344	3.906	3.125
1	9.883	7.457	1.527	7.367	0.809	9.434	3.504	6.559	2.606
2	11.735	6.122	1.020	12.245	2.551	6.633	1.531	13.265	1.531
3	0.971	2.913	11.650	2.913	0.971	4.854	2.913	2.913	0.971
4	11.321	7.075	3.774	7.547	0.943	11.792	1.887	7.075	2.830
..	…	…	…	…	…	…	…	…	…
995	12.575	5.090	0.599	3.892	0.299	4.192	1.198	11.377	0.299
996	10.219	5.109	0.000	2.920	0.000	8.029	0.730	13.139	0.730
997	9.655	8.966	4.138	7.586	0.690	8.966	1.379	7.586	0.000
998	7.143	5.143	4.857	5.143	0.571	7.714	1.714	9.714	2.286
999	10.417	6.845	1.190	9.226	0.000	10.119	2.083	8.631	1.190

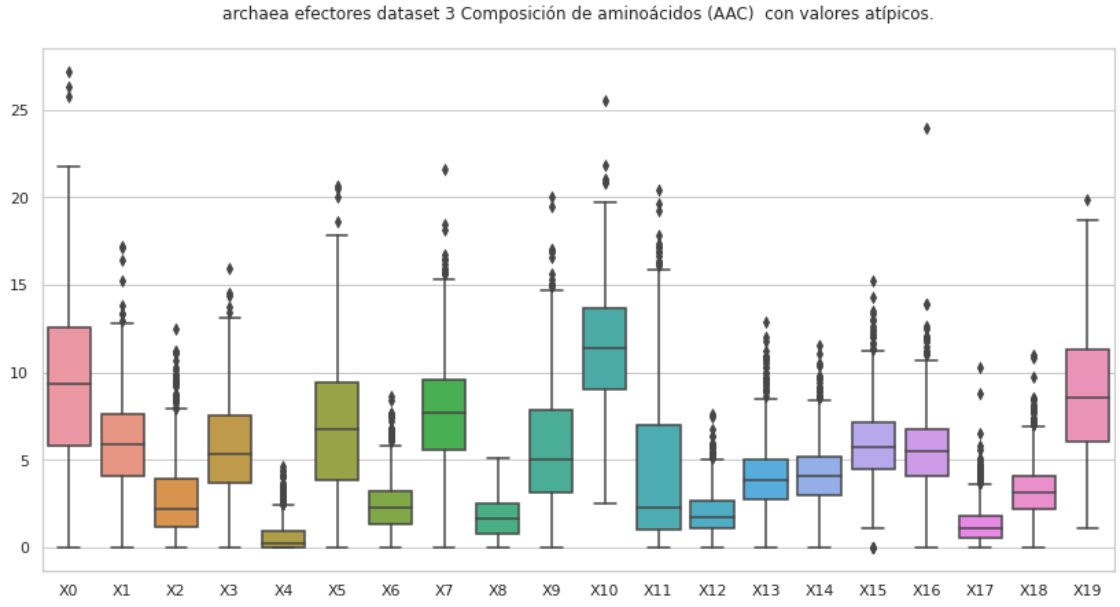
	X9	…	X11	X12	X13	X14	X15	X16	X17	X18 \
0	7.031	…	10.156	3.906	5.469	2.344	4.688	3.125	0.000	4.688

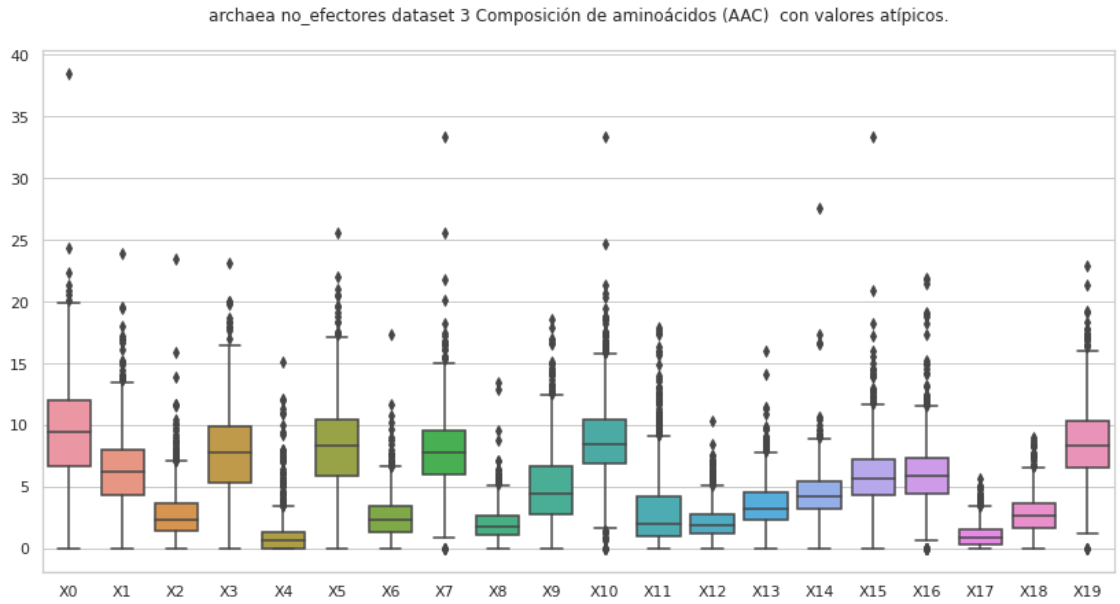
```
1     3.953  …   0.898  0.898  2.875  6.379   6.649   8.086  0.449  3.324
2     3.061  …   0.000  2.551  4.082  6.122   5.612   4.082  0.000  2.041
3    11.650  …   5.825  1.942  7.767  1.942  14.563  10.680  0.971  3.883
4     2.358  …   0.472  0.943  4.717  4.245   3.774   7.075  1.415  3.774
..      …  …     …    …    …    …      …     …    …    …
995   7.784  …   0.000  2.096  5.689  3.593   5.689   6.587  1.198  0.898
996   5.839  …   1.460  1.460  8.029  0.000   8.029   3.650  1.460  5.109
997   4.138  …   5.517  2.759  2.069  4.138   6.207   2.759  0.690  4.138
998   7.143  …   4.857  2.857  4.286  5.429   5.143   4.571  2.286  5.714
999   6.250  …   1.488  2.679  3.571  5.060   5.060   7.440  1.190  2.679

        X19            X20
0     8.594   no_efectores
1     7.098   no_efectores
2     9.694   no_efectores
3     4.854   no_efectores
4     7.547   no_efectores
..      …            …
995  12.575   no_efectores
996   6.569   no_efectores
997  12.414   no_efectores
998   6.857   no_efectores
999   6.548   no_efectores

[1000 rows x 21 columns]
```

Composición de aminoácidos (AAC) no_efectores archaea dataset 3, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      9.551993     6.417151     2.807094     7.659095     1.066207
std       3.961634     2.971403     2.128019     3.361097     1.595708
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       6.658500     4.348000     1.458250     5.374250     0.000000
50%       9.403500     6.193500     2.357500     7.827000     0.649000
75%      11.968500     8.050750     3.707750     9.841250     1.379750
max      38.462000    23.864000    23.438000    23.077000    15.094000

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      8.254448     2.544129     7.933455     1.982636     5.022300
std       3.681905     1.688343     3.044880     1.419415     3.111424
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       5.876000     1.342000     5.997750     1.070750     2.828750
50%       8.333000     2.281000     7.796500     1.818000     4.413500
```

```
75%      10.407250     3.452750     9.595500     2.718000     6.726750
max      25.610000    17.308000    33.333000    13.462000    18.519000

                X10          X11          X12          X13          X14  \
count   1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean       8.915180     3.130147     2.157898     3.543922     4.412601
std        3.341559     3.157412     1.349238     1.902238     2.050783
min        0.000000     0.000000     0.000000     0.000000     0.000000
25%        6.929000     1.000000     1.186500     2.306750     3.205000
50%        8.472000     2.010000     1.902000     3.274500     4.270500
75%       10.479500     4.259750     2.757500     4.511500     5.503500
max       33.333000    17.949000    10.345000    16.000000    27.536000

                X15          X16          X17          X18          X19
count   1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean       6.015665     6.089317     1.083434     2.805353     8.607933
std        2.678443     2.637689     0.986097     1.550702     3.090493
min        0.000000     0.000000     0.000000     0.000000     0.000000
25%        4.302250     4.492250     0.318750     1.722250     6.563750
50%        5.695500     5.882000     0.941000     2.683000     8.392000
75%        7.277750     7.317000     1.583250     3.675250    10.345000
max       33.333000    21.944000     5.660000     8.989000    22.917000
```





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 3, sin valores
atípicos.
Valores del documento csv.

         X0     X1     X2      X3     X4      X5     X6     X7     X8      X9  \
0     4.237  7.627  2.542   5.932  0.000  15.254  3.390  7.627  1.695   5.085
2     6.486  4.324  5.946   3.243  2.162   5.405  2.162  9.730  0.541   7.027
5     9.677  9.677  3.226  10.753  0.000   7.527  1.075  8.602  3.226   5.376
6     3.378  3.378  4.730   4.730  0.676   8.784  4.054  3.378  1.351   5.405
7    12.613  5.405  1.351   6.306  0.000   3.604  4.054  9.910  1.802   6.306
..      …      …      …       …      …       …      …      …      …       …
993  14.627  3.881  2.090  12.239  0.299   7.463  2.687  5.672  1.194   3.881
995  10.569  6.504  4.065   8.943  0.813   8.943  0.813  6.504  0.000  10.569
996  11.610  6.367  1.873   9.738  0.000   8.614  1.498  9.738  1.498   2.247
997  15.672  5.970  1.493   7.463  0.746   3.731  1.493  8.209  3.731   1.493
999   1.515  6.061  3.030   7.576  0.000  12.121  1.515  3.030  0.000  12.121

     …     X11    X12    X13    X14     X15    X16    X17    X18     X19  \
0    …   2.542  3.390  3.390  2.542  11.017  3.390  1.695  1.695   5.085
2    …   7.027  2.162  7.568  4.324   6.486  5.946  1.622  6.486   4.324
5    …   2.151  1.075  2.151  1.075   6.452  7.527  0.000  2.151   4.301
6    …  13.514  4.054  2.703  2.703   8.784  8.784  0.676  4.054   5.405
7    …   0.901  1.802  4.054  5.405   5.405  5.856  2.252  2.252  11.261
..   …      …      …      …      …       …      …      …      …       …
993  …   0.299  1.791  1.791  4.179   8.657  8.060  0.299  4.478   9.552
```

```
995  …   4.878  2.439  3.252  0.000   4.878  5.691  3.252  3.252   7.317
996  …   1.498  1.124  1.124  3.745   6.367  6.742  0.375  2.622  15.730
997  …   1.493  0.746  5.970  2.985   4.478  1.493  0.746  0.746  16.418
999  …  16.667  1.515  1.515  0.000   7.576  1.515  1.515  4.545   7.576

           X20
0     efectores
2     efectores
5     efectores
6     efectores
7     efectores
..          …
993   efectores
995   efectores
996   efectores
997   efectores
999   efectores

[875 rows x 21 columns]
```

Composición de aminoácidos (AAC) efectores archaea dataset 3, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  875.000000  875.000000  875.000000  875.000000  875.000000  875.000000
mean     9.879151    6.080456    2.534314    5.915321    0.541743    6.839521
std      4.444430    2.484564    1.842263    2.673662    0.744061    3.668082
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      6.278000    4.307500    1.160500    3.760000    0.000000    3.853000
50%      9.848000    5.926000    2.041000    5.455000    0.227000    6.678000
75%     12.799500    7.670000    3.734500    7.595000    0.810000    9.309500
max     21.767000   13.805000    8.731000   13.750000    3.226000   17.829000

                X6          X7          X8          X9         X10         X11  \
count  875.000000  875.000000  875.000000  875.000000  875.000000  875.000000
mean     2.353632    7.939050    1.755142    5.492768   11.411459    3.918955
std      1.328344    2.832442    1.149150    3.163705    3.166984    4.067375
min      0.000000    0.000000    0.000000    0.000000    2.721000    0.000000
25%      1.371500    5.882000    0.761500    3.076500    9.080500    0.941500
50%      2.258000    7.944000    1.657000    4.918000   11.538000    1.946000
75%      3.166500    9.744000    2.564000    7.418000   13.650500    6.332500
max      6.716000   16.516000    5.128000   15.652000   20.904000   17.365000

               X12         X13         X14         X15         X16         X17  \
count  875.000000  875.000000  875.000000  875.000000  875.000000  875.000000
mean     1.980705    4.013565    4.201483    5.894499    5.649955    1.310955
```

std	1.094718	1.833531	1.684031	1.991122	1.933550	1.010736
min	0.266000	0.000000	0.000000	0.000000	0.901000	0.000000
25%	1.124000	2.844000	3.077000	4.578500	4.204000	0.637000
50%	1.739000	3.913000	4.167000	5.742000	5.596000	1.136000
75%	2.573500	5.021000	5.215000	7.169500	6.815500	1.831000
max	5.628000	10.000000	9.392000	12.338000	12.000000	4.615000

	X18	X19
count	875.000000	875.000000
mean	3.237230	9.049977
std	1.380477	3.496130
min	0.000000	1.399000
25%	2.261500	6.314000
50%	3.150000	8.943000
75%	4.054000	11.533500
max	7.864000	18.657000

no_efectores

Composición de aminoácidos (AAC) no_efectores archaea dataset 3, sin valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9	\
0	6.250	4.688	6.250	6.250	2.344	8.594	2.344	3.906	3.125	7.031	
1	9.883	7.457	1.527	7.367	0.809	9.434	3.504	6.559	2.606	3.953	
2	11.735	6.122	1.020	12.245	2.551	6.633	1.531	13.265	1.531	3.061	
4	11.321	7.075	3.774	7.547	0.943	11.792	1.887	7.075	2.830	2.358	
6	6.226	4.669	5.058	8.755	1.167	5.642	2.335	6.615	1.751	7.588	
..	…	…	…	…	…	…	…	…	…	…	
995	12.575	5.090	0.599	3.892	0.299	4.192	1.198	11.377	0.299	7.784	
996	10.219	5.109	0.000	2.920	0.000	8.029	0.730	13.139	0.730	5.839	
997	9.655	8.966	4.138	7.586	0.690	8.966	1.379	7.586	0.000	4.138	
998	7.143	5.143	4.857	5.143	0.571	7.714	1.714	9.714	2.286	7.143	
999	10.417	6.845	1.190	9.226	0.000	10.119	2.083	8.631	1.190	6.250	

	…	X11	X12	X13	X14	X15	X16	X17	X18	X19	\
0	…	10.156	3.906	5.469	2.344	4.688	3.125	0.000	4.688	8.594	
1	…	0.898	0.898	2.875	6.379	6.649	8.086	0.449	3.324	7.098	
2	…	0.000	2.551	4.082	6.122	5.612	4.082	0.000	2.041	9.694	
4	…	0.472	0.943	4.717	4.245	3.774	7.075	1.415	3.774	7.547	
6	…	6.615	2.724	3.891	2.724	9.533	5.447	0.584	3.891	7.977	
..	…	…	…	…	…	…	…	…	…		
995	…	0.000	2.096	5.689	3.593	5.689	6.587	1.198	0.898	12.575	
996	…	1.460	1.460	8.029	0.000	8.029	3.650	1.460	5.109	6.569	
997	…	5.517	2.759	2.069	4.138	6.207	2.759	0.690	4.138	12.414	
998	…	4.857	2.857	4.286	5.429	5.143	4.571	2.286	5.714	6.857	

```
999  …   1.488  2.679  3.571  5.060  5.060  7.440  1.190  2.679   6.548

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

[843 rows x 21 columns]
```

Composición de aminoácidos (AAC) no_efectores archaea dataset 3, sin valores atípicos.
Estadísticas.

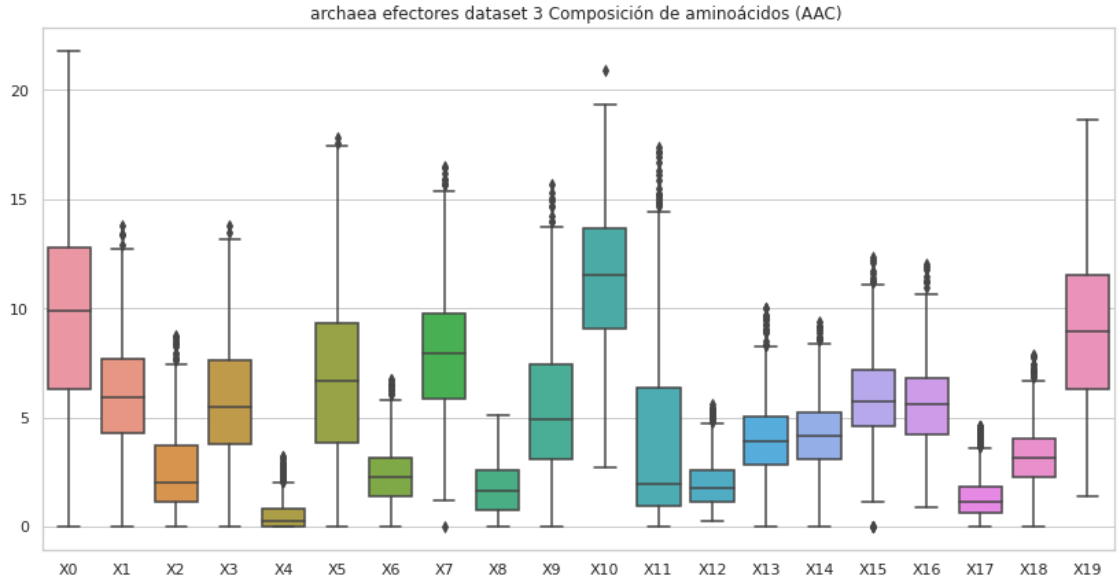
	X0	X1	X2	X3	X4	X5	\
count	843.000000	843.00000	843.000000	843.000000	843.000000	843.000000	
mean	9.815931	6.47702	2.719580	7.930383	0.884881	8.436094	
std	3.649878	2.56515	1.784412	2.950018	0.971053	3.272339	
min	1.639000	0.00000	0.000000	0.000000	0.000000	0.565000	
25%	7.222500	4.68050	1.481500	5.970000	0.000000	6.363500	
50%	9.655000	6.35800	2.370000	8.180000	0.651000	8.556000	
75%	12.107500	8.01700	3.675500	9.929000	1.259500	10.444000	
max	21.311000	14.94300	8.989000	17.708000	5.833000	19.118000	

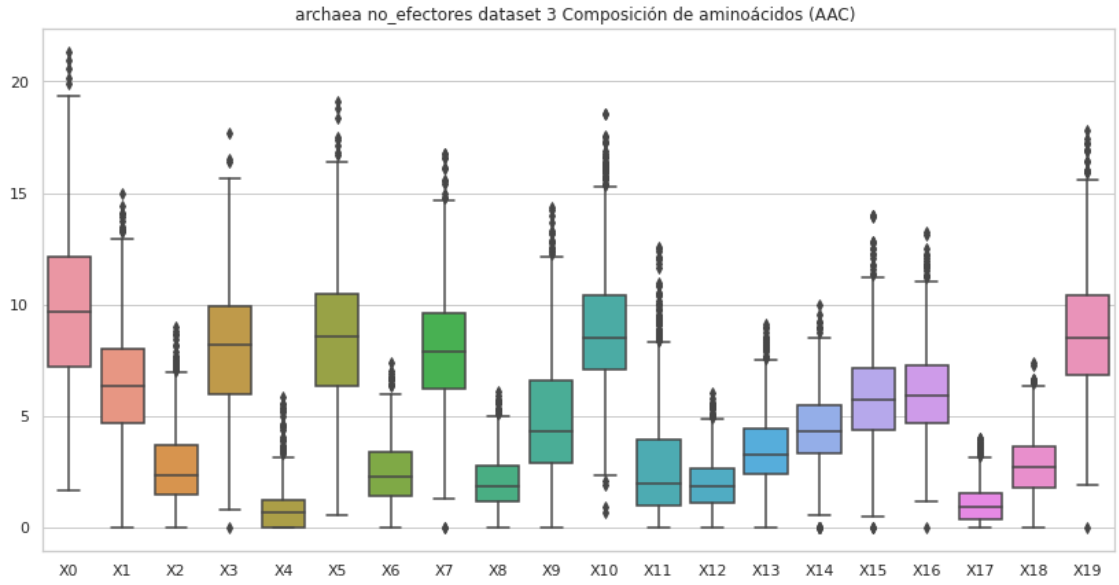
	X6	X7	X8	X9	X10	X11	\
count	843.000000	843.000000	843.000000	843.000000	843.000000	843.000000	
mean	2.487572	8.002045	2.017785	4.884292	8.927107	2.875173	
std	1.399774	2.605759	1.190246	2.783134	2.850783	2.679813	
min	0.000000	0.000000	0.000000	0.000000	0.667000	0.000000	
25%	1.408000	6.222000	1.173000	2.866500	7.088500	1.004000	
50%	2.299000	7.877000	1.877000	4.310000	8.511000	1.954000	
75%	3.372000	9.602000	2.742000	6.591500	10.390000	3.942000	
max	7.407000	16.779000	6.081000	14.348000	18.568000	12.583000	

	X12	X13	X14	X15	X16	X17	\
count	843.000000	843.000000	843.000000	843.000000	843.000000	843.000000	
mean	2.001287	3.506849	4.409372	5.920024	6.074122	1.058070	
std	1.126563	1.675972	1.641859	2.208131	2.057585	0.885773	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	1.128000	2.372500	3.357000	4.367000	4.674000	0.392500	

50%	1.835000	3.250000	4.308000	5.702000	5.926000	0.939000
75%	2.635500	4.430000	5.467500	7.143000	7.259500	1.511000
max	6.024000	9.091000	10.000000	14.013000	13.260000	4.032000

	X18	X19
count	843.000000	843.000000
mean	2.808624	8.763713
std	1.371449	2.783782
min	0.000000	0.000000
25%	1.789500	6.833500
50%	2.713000	8.511000
75%	3.638000	10.437500
max	7.418000	17.778000





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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.021143	0.000000	0.029600	0.076113	0.016914	0.038057	0.008457	
1	0.029124	0.003066	0.007664	0.000000	0.007664	0.013795	0.001533	
2	0.027553	0.009184	0.013777	0.022961	0.032146	0.041330	0.002296	
3	0.029836	0.014918	0.014918	0.049726	0.000000	0.019891	0.004973	
4	0.025475	0.010190	0.045854	0.066234	0.061139	0.056044	0.030570	
..	…	…	…	…	…	…	…	
995	0.066206	0.005093	0.056021	0.056021	0.020371	0.040742	0.000000	
996	0.028940	0.000000	0.024272	0.021472	0.002801	0.024272	0.003734	
997	0.025420	0.001210	0.012105	0.006052	0.009684	0.013315	0.006052	
998	0.086366	0.015703	0.086366	0.102069	0.039257	0.054960	0.023554	
999	0.008464	0.000000	0.042320	0.067712	0.008464	0.016928	0.000000	

	X7	X8	X9	…	X74	X75	X76	\
0	0.025371	0.012686	0.059199	…	0.039986	0.022281	0.000034	
1	0.010730	0.003066	0.009197	…	0.007554	0.000805	0.025320	
2	0.029849	0.029849	0.029849	…	0.020479	0.030724	0.008055	
3	0.034808	0.034808	0.009945	…	0.013676	0.040495	0.000304	
4	0.106993	0.122278	0.040759	…	0.017630	0.057975	0.084942	
..	…	…	…	…	…	…	…	
995	0.066206	0.030557	0.045835	…	-0.058890	-0.046709	0.006606	
996	0.005601	0.003734	0.018671	…	0.001167	0.008084	0.022887	
997	0.002421	0.002421	0.024209	…	0.014643	0.001746	0.008609	
998	0.023554	0.039257	0.094217	…	0.128269	0.081980	-0.009784	
999	0.067712	0.093104	0.059248	…	-0.079957	0.049127	0.021547	

	X77	X78	X79	X80	X81	X82	X83
0	0.044181	0.090421	-0.045323	-0.014501	0.039582	-0.015357	efectores
1	0.014901	0.006389	0.049982	0.013817	0.008960	0.026874	efectores
2	-0.011958	-0.009848	-0.009777	0.005443	0.007274	0.000077	efectores
3	0.020920	0.028420	0.039843	-0.004420	-0.002039	0.003083	efectores
4	0.022943	0.079490	-0.006831	0.018112	-0.014340	-0.023306	efectores
..	…	…	…	…	…	…	
995	-0.007157	-0.026082	0.065560	-0.001754	-0.004374	0.021369	efectores
996	0.002309	0.007698	0.030904	-0.000724	-0.003317	0.020575	efectores

```
997  0.019717  0.007057  0.012735  0.018009  0.007001  0.010140  efectores
998  0.057004  0.027305  0.061577 -0.109090  0.001263  0.060370  efectores
999 -0.060103 -0.015581 -0.005535  0.071475  0.080733 -0.014792  efectores

[1000 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro_mass efectores archaea dataset 3, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.031464	0.003240	0.024139	0.032510	0.015768	
std	0.025935	0.007644	0.034940	0.029792	0.016466	
min	-0.593736	-0.148434	-0.890604	-0.148434	-0.296868	
25%	0.020559	0.000000	0.009026	0.007481	0.007185	
50%	0.028255	0.000513	0.019777	0.023623	0.012552	
75%	0.039442	0.004260	0.036804	0.050668	0.020833	
max	0.131626	0.044733	0.115064	0.209405	0.090706	

	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.026893	0.006977	0.026155	0.023383	0.043501	…	
std	0.024000	0.016052	0.031127	0.033343	0.037183	…	
min	-0.593736	-0.445302	-0.445302	-0.296868	-0.742170	…	
25%	0.016908	0.001772	0.007312	0.001952	0.023825	…	
50%	0.024669	0.005554	0.015547	0.007520	0.038200	…	
75%	0.034677	0.010375	0.037103	0.037029	0.055851	…	
max	0.101613	0.044085	0.158440	0.196824	0.203227	…	

	X73	X74	X75	X76	X77	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.014529	0.006844	0.009533	0.015096	0.007175	
std	0.023291	0.028436	0.024971	0.035528	0.046540	
min	-0.438281	-0.165282	-0.204425	-0.079783	-0.264966	
25%	0.005643	-0.004527	-0.001655	0.004383	-0.003335	
50%	0.015814	0.009290	0.004176	0.016638	0.009827	
75%	0.025167	0.016967	0.016328	0.024398	0.018386	
max	0.114159	0.323291	0.171060	0.970213	1.180889	

	X78	X79	X80	X81	X82
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.009684	0.014487	0.004519	0.008506	0.014839
std	0.033874	0.020018	0.033620	0.024835	0.020221
min	-0.171278	-0.166701	-0.466707	-0.298167	-0.113640
25%	-0.001065	0.005114	-0.002725	-0.000433	0.005173
50%	0.005698	0.016043	0.009985	0.005489	0.016228

```
75%       0.016371     0.025509     0.017965     0.016621     0.024492
max       0.745374     0.104846     0.156463     0.100974     0.129773

[8 rows x 83 columns]
```

```
no_efectores
Composición de pseudo aminoácidos (PseAAC)  hidro_mass no_efectores archaea
dataset 3, con valores atípicos.
Valores del documento csv.
```

	X0	X1	X2	X3	X4	X5	X6	\
0	0.077249	0.028968	0.077249	0.106217	0.067593	0.048280	0.038624	
1	0.043714	0.003577	0.032587	0.041727	0.012717	0.029010	0.011525	
2	0.024588	0.005345	0.025657	0.013898	0.008552	0.027795	0.003207	
3	0.011272	0.011272	0.033816	0.056360	0.090176	0.033816	0.011272	
4	0.056443	0.004704	0.037629	0.058795	0.023518	0.035277	0.014111	
..	…	…	…	…	…	…	…	
995	0.019752	0.000470	0.006114	0.006584	0.008935	0.017871	0.000470	
996	0.025867	0.000000	0.007390	0.020324	0.020324	0.033257	0.001848	
997	0.031748	0.002268	0.024945	0.029480	0.006803	0.024945	0.000000	
998	0.049752	0.003980	0.035822	0.053732	0.029851	0.067663	0.015921	
999	0.047890	0.000000	0.042417	0.046521	0.016419	0.039680	0.005473	

	X7	X8	X9	…	X74	X75	X76	\
0	0.086905	0.125529	0.077249	…	0.003931	0.043633	-0.077059	
1	0.017486	0.003974	0.045304	…	0.010041	0.012500	0.024822	
2	0.006414	0.000000	0.012828	…	0.002036	0.001431	0.033102	
3	0.135264	0.067632	0.056360	…	0.043646	0.022115	-0.024630	
4	0.011759	0.002352	0.047036	…	-0.022326	-0.039026	0.054593	
..	…	…	…	…	…	…	…	
995	0.012227	0.000000	0.022574	…	0.023694	0.004252	0.015186	
996	0.014781	0.003695	0.044343	…	-0.009516	-0.026839	0.012733	
997	0.013606	0.018142	0.020410	…	0.026705	0.004052	0.031245	
998	0.049752	0.033831	0.045772	…	0.019862	0.022107	0.015538	
999	0.028734	0.006841	0.038312	…	0.000051	-0.012262	0.018278	

	X77	X78	X79	X80	X81	X82	X83
0	0.080748	-0.020324	0.032843	-0.070573	-0.015658	-0.018350	no_efectores
1	-0.010128	-0.000257	0.015319	0.000560	0.016461	0.018130	no_efectores
2	0.011121	0.011176	0.020173	-0.006555	-0.010935	0.021229	no_efectores
3	-0.012136	-0.045092	-0.071794	-0.012185	-0.070846	-0.001679	no_efectores
4	0.012921	0.021868	0.038927	0.059164	0.065285	0.027377	no_efectores
..	…	…	…	…	…	…	…
995	0.019337	0.002129	0.014356	0.013559	-0.002406	0.027305	no_efectores
996	0.041477	0.007294	0.033681	-0.001188	-0.002544	0.014669	no_efectores
997	-0.025161	0.007295	0.020720	-0.004579	0.014499	0.024313	no_efectores

```
998  0.017480  0.014678  0.029604  0.003603 -0.000285  0.010842  no_efectores
999  0.005245  0.018761  0.034058  0.008550 -0.006606  0.021858  no_efectores

[1000 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores archaea
dataset 3, con valores atípicos.
Estadísticas.

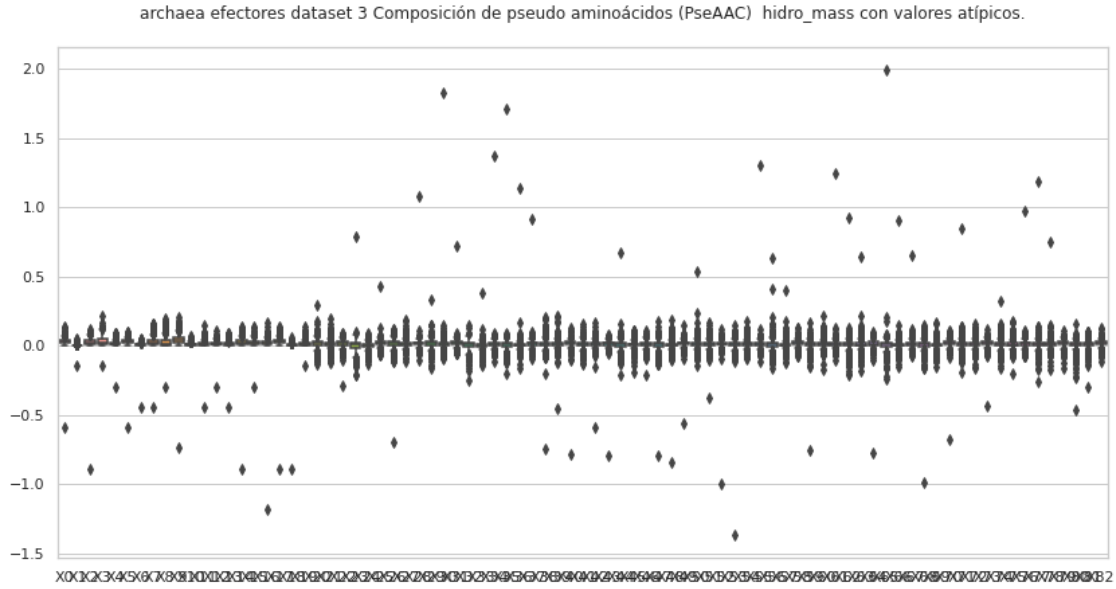
	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.034546	0.005200	0.031271	0.034905	0.014971	
std	0.024497	0.013245	0.019511	0.032380	0.016911	
min	-0.503049	-0.143728	-0.071864	-0.466181	-0.215592	
25%	0.024614	0.000000	0.018752	0.018304	0.007078	
50%	0.033409	0.002166	0.029981	0.031813	0.011527	
75%	0.043421	0.005755	0.041401	0.047176	0.019045	
max	0.164772	0.167482	0.186678	0.277356	0.188221	

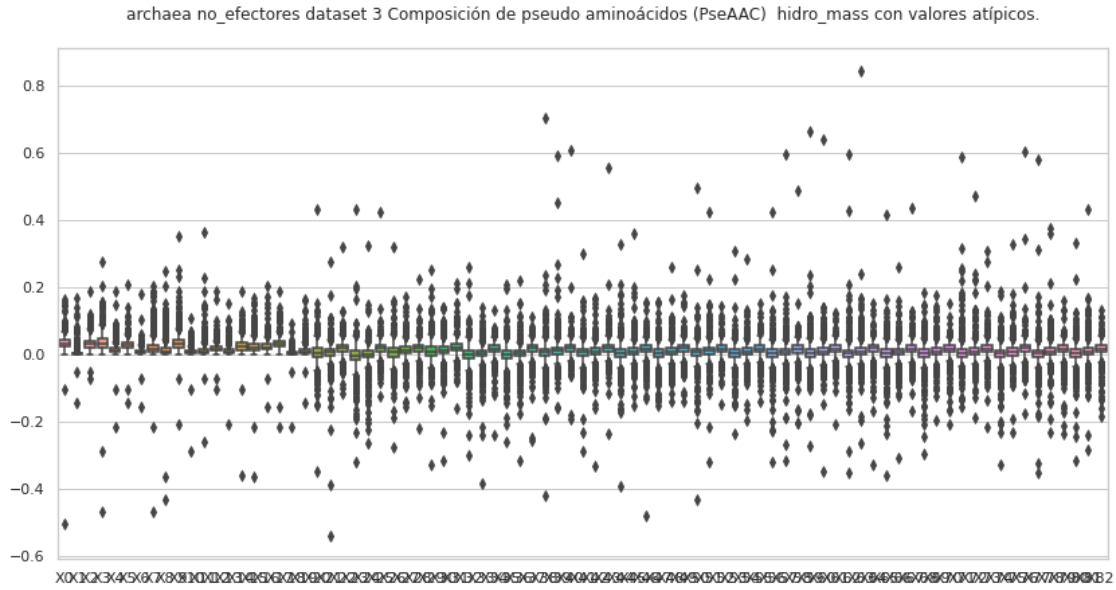
	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.029626	0.008779	0.021817	0.015184	0.035889	…	
std	0.016860	0.011953	0.029690	0.030386	0.027597	…	
min	-0.143728	-0.155394	-0.466181	-0.431184	-0.207192	…	
25%	0.020195	0.002882	0.008545	0.002777	0.021543	…	
50%	0.027788	0.006380	0.015588	0.007393	0.030324	…	
75%	0.036168	0.012407	0.028248	0.018832	0.043829	…	
max	0.206129	0.180363	0.205476	0.249546	0.353083	…	

	X73	X74	X75	X76	X77	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.016771	0.000808	0.008065	0.016963	0.002214	
std	0.025828	0.030487	0.027215	0.030669	0.036928	
min	-0.216062	-0.326420	-0.231937	-0.177423	-0.352072	
25%	0.007467	-0.007904	-0.002250	0.007052	-0.007665	
50%	0.017546	0.002787	0.006012	0.018391	0.003634	
75%	0.026771	0.013729	0.017950	0.027248	0.013571	
max	0.309073	0.188280	0.328901	0.604321	0.578233	

	X78	X79	X80	X81	X82
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.009685	0.015181	0.002100	0.009031	0.015464
std	0.030454	0.024951	0.031741	0.031268	0.023550
min	-0.233136	-0.233941	-0.314878	-0.285652	-0.184222
25%	-0.000660	0.006679	-0.007209	-0.001455	0.006984
50%	0.006911	0.017635	0.003598	0.006639	0.017699
75%	0.018636	0.026811	0.014682	0.018288	0.027495

```
max        0.374743      0.160413      0.331276      0.430900      0.131726

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.021143	0.000000	0.029600	0.076113	0.016914	0.038057	0.008457	
1	0.029124	0.003066	0.007664	0.000000	0.007664	0.013795	0.001533	
2	0.027553	0.009184	0.013777	0.022961	0.032146	0.041330	0.002296	
3	0.029836	0.014918	0.014918	0.049726	0.000000	0.019891	0.004973	
5	0.044469	0.000000	0.049409	0.034587	0.009882	0.039528	0.014823	
..	…	…	…	…	…	…	…	
993	0.040832	0.000833	0.034165	0.020832	0.005000	0.015833	0.003333	
994	0.031923	0.000000	0.012492	0.006940	0.020820	0.023596	0.011104	
995	0.066206	0.005093	0.056021	0.056021	0.020371	0.040742	0.000000	
996	0.028940	0.000000	0.024272	0.021472	0.002801	0.024272	0.003734	
997	0.025420	0.001210	0.012105	0.006052	0.009684	0.013315	0.006052	

	X7	X8	X9	…	X74	X75	X76	\
0	0.025371	0.012686	0.059199	…	0.039986	0.022281	0.000034	
1	0.010730	0.003066	0.009197	…	0.007554	0.000805	0.025320	
2	0.029849	0.029849	0.029849	…	0.020479	0.030724	0.008055	
3	0.034808	0.034808	0.009945	…	0.013676	0.040495	0.000304	
5	0.024705	0.009882	0.064232	…	0.031235	0.017096	0.025425	
..	…	…	…	…	…	…	…	
993	0.010833	0.000833	0.019166	…	-0.006268	-0.007896	0.021928	
994	0.011104	0.005552	0.030535	…	0.017712	0.014379	0.011906	
995	0.066206	0.030557	0.045835	…	-0.058890	-0.046709	0.006606	
996	0.005601	0.003734	0.018671	…	0.001167	0.008084	0.022887	
997	0.002421	0.002421	0.024209	…	0.014643	0.001746	0.008609	

	X77	X78	X79	X80	X81	X82	X83
0	0.044181	0.090421	-0.045323	-0.014501	0.039582	-0.015357	efectores
1	0.014901	0.006389	0.049982	0.013817	0.008960	0.026874	efectores
2	-0.011958	-0.009848	-0.009777	0.005443	0.007274	0.000077	efectores
3	0.020920	0.028420	0.039843	-0.004420	-0.002039	0.003083	efectores
5	-0.015634	-0.003217	0.022705	-0.028812	-0.006157	-0.015728	efectores
..	…	…	…	…	…	…	
993	0.007090	0.003285	0.032863	0.001966	0.009161	0.034633	efectores
994	0.024292	0.007826	-0.008089	0.010002	-0.001873	-0.002669	efectores
995	-0.007157	-0.026082	0.065560	-0.001754	-0.004374	0.021369	efectores
996	0.002309	0.007698	0.030904	-0.000724	-0.003317	0.020575	efectores
997	0.019717	0.007057	0.012735	0.018009	0.007001	0.010140	efectores

[852 rows x 84 columns]

Composición de pseudo aminoácidos (PseAAC) hidro_mass efectores archaea dataset 3, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	852.000000	852.000000	852.000000	852.000000	852.000000	852.000000	
mean	0.030004	0.002236	0.020923	0.025467	0.013691	0.025231	
std	0.013636	0.003736	0.015636	0.022386	0.009846	0.011128	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.002551	
25%	0.020247	0.000000	0.008022	0.006470	0.006606	0.016206	
50%	0.027382	0.000366	0.015957	0.017550	0.011238	0.023446	
75%	0.036571	0.003027	0.031541	0.040678	0.018360	0.031879	
max	0.083642	0.020853	0.087325	0.104952	0.062567	0.077434	

	X6	X7	X8	X9	…	X73	\
count	852.000000	852.000000	852.000000	852.000000	…	852.000000	
mean	0.006263	0.020564	0.016304	0.037564	…	0.015792	
std	0.005671	0.019168	0.022343	0.019758	…	0.014722	
min	0.000000	0.000000	0.000000	0.004803	…	-0.042634	
25%	0.001699	0.006579	0.001742	0.022020	…	0.007190	
50%	0.005116	0.012295	0.004952	0.034023	…	0.016284	
75%	0.008988	0.030006	0.021810	0.048086	…	0.025034	
max	0.037466	0.102790	0.112486	0.131845	…	0.077144	

	X74	X75	X76	X77	X78	X79	\
count	852.000000	852.000000	852.000000	852.000000	852.000000	852.000000	
mean	0.006499	0.007233	0.015247	0.008067	0.008591	0.016334	
std	0.017078	0.016147	0.013987	0.018127	0.017515	0.014854	
min	-0.076600	-0.065223	-0.064904	-0.065493	-0.079976	-0.045323	
25%	-0.003010	-0.001477	0.007231	-0.001424	-0.000451	0.007192	
50%	0.009499	0.003691	0.017108	0.010335	0.005561	0.016871	
75%	0.015917	0.013140	0.024178	0.018052	0.014580	0.025509	
max	0.086157	0.072199	0.054754	0.077863	0.090421	0.073944	

	X80	X81	X82
count	852.000000	852.000000	852.000000
mean	0.008030	0.008358	0.015758
std	0.018485	0.015435	0.014485
min	-0.092932	-0.042359	-0.041574
25%	-0.000672	0.000187	0.007450
50%	0.010664	0.005211	0.016954
75%	0.017565	0.014534	0.023975
max	0.075528	0.079287	0.068372

[8 rows x 83 columns]

```
no_efectores
Composición de pseudo aminoácidos (PseAAC)  hidro_mass no_efectores archaea
dataset 3, sin valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
1    0.043714  0.003577  0.032587  0.041727  0.012717  0.029010  0.011525
2    0.024588  0.005345  0.025657  0.013898  0.008552  0.027795  0.003207
4    0.056443  0.004704  0.037629  0.058795  0.023518  0.035277  0.014111
5    0.003410  0.000000  0.017050  0.012504  0.001137  0.010230  0.001137
6    0.034892  0.006542  0.049066  0.031620  0.021807  0.037072  0.009813
..        …         …         …         …         …         …         …
995  0.019752  0.000470  0.006114  0.006584  0.008935  0.017871  0.000470
996  0.025867  0.000000  0.007390  0.020324  0.020324  0.033257  0.001848
997  0.031748  0.002268  0.024945  0.029480  0.006803  0.024945  0.000000
998  0.049752  0.003980  0.035822  0.053732  0.029851  0.067663  0.015921
999  0.047890  0.000000  0.042417  0.046521  0.016419  0.039680  0.005473

           X7        X8        X9  …       X74       X75       X76  \
1    0.017486  0.003974  0.045304  …  0.010041  0.012500  0.024822
2    0.006414  0.000000  0.012828  …  0.002036  0.001431  0.033102
4    0.011759  0.002352  0.047036  … -0.022326 -0.039026  0.054593
5    0.002273  0.000000  0.007957  …  0.002322  0.018635 -0.009994
6    0.042524  0.037072  0.038163  … -0.012536  0.005448  0.020527
..        …         …         …  …         …         …         …
995  0.012227  0.000000  0.022574  …  0.023694  0.004252  0.015186
996  0.014781  0.003695  0.044343  … -0.009516 -0.026839  0.012733
997  0.013606  0.018142  0.020410  …  0.026705  0.004052  0.031245
998  0.049752  0.033831  0.045772  …  0.019862  0.022107  0.015538
999  0.028734  0.006841  0.038312  …  0.000051 -0.012262  0.018278

           X77       X78       X79       X80       X81       X82           X83
1   -0.010128 -0.000257  0.015319  0.000560  0.016461  0.018130  no_efectores
2    0.011121  0.011176  0.020173 -0.006555 -0.010935  0.021229  no_efectores
4    0.012921  0.021868  0.038927  0.059164  0.065285  0.027377  no_efectores
5   -0.005447  0.028846 -0.002306  0.018035  0.036564 -0.003168  no_efectores
6    0.009626 -0.004030  0.017815 -0.002036  0.020218  0.017059  no_efectores
..        …         …         …         …         …         …           …
995  0.019337  0.002129  0.014356  0.013559 -0.002406  0.027305  no_efectores
996  0.041477  0.007294  0.033681 -0.001188 -0.002544  0.014669  no_efectores
997 -0.025161  0.007295  0.020720 -0.004579  0.014499  0.024313  no_efectores
998  0.017480  0.014678  0.029604  0.003603 -0.000285  0.010842  no_efectores
999  0.005245  0.018761  0.034058  0.008550 -0.006606  0.021858  no_efectores

[872 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores archaea
dataset 3, sin valores atípicos.
Estadísticas.

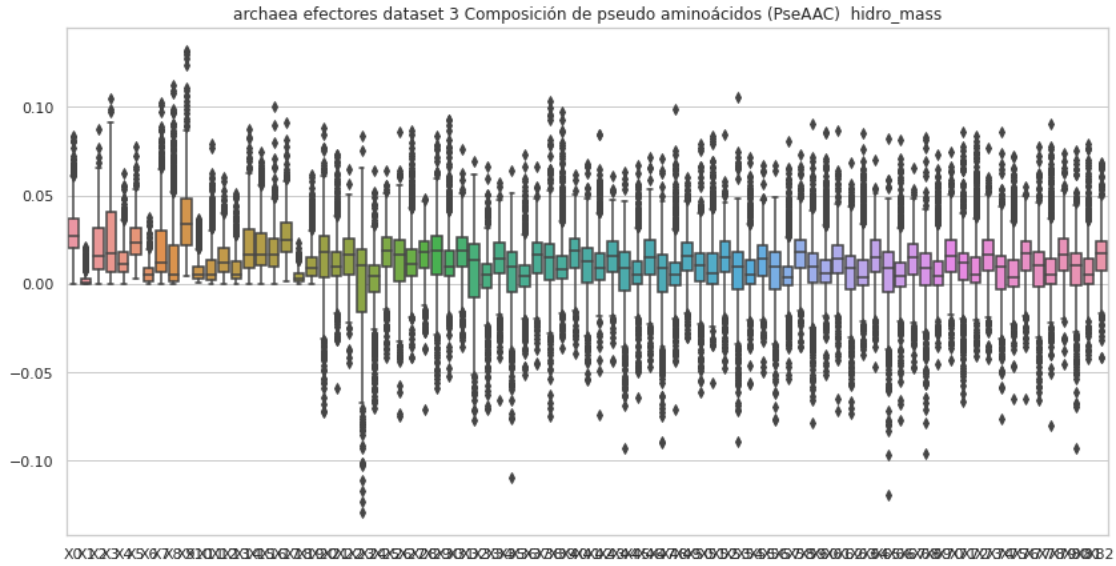
	X0	X1	X2	X3	X4	X5	\
count	872.000000	872.000000	872.000000	872.000000	872.000000	872.000000	
mean	0.033250	0.003559	0.028969	0.030825	0.012437	0.027719	
std	0.012935	0.004919	0.015822	0.018670	0.008233	0.011221	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.024282	0.000000	0.017483	0.016474	0.006842	0.019926	
50%	0.032530	0.002038	0.029051	0.029549	0.010826	0.026879	
75%	0.041323	0.004908	0.039284	0.042306	0.016405	0.034069	
max	0.090926	0.043129	0.078973	0.116376	0.050675	0.067666	

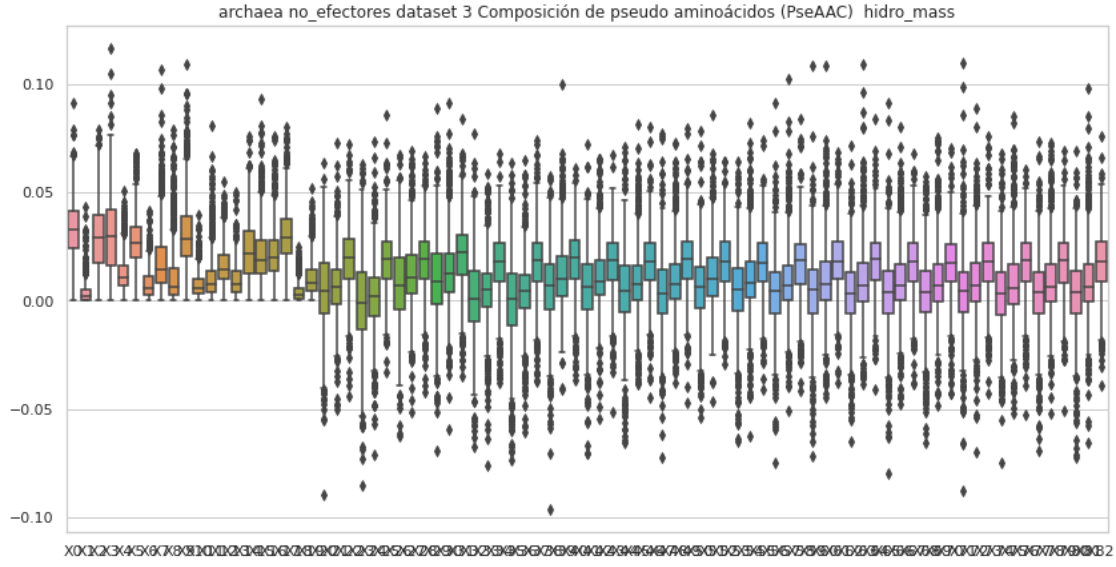
	X6	X7	X8	X9	…	X73	\
count	872.000000	872.000000	872.000000	872.000000	…	872.000000	
mean	0.007601	0.018375	0.011327	0.031527	…	0.017003	
std	0.006510	0.014768	0.013520	0.015716	…	0.014753	
min	0.000000	0.000000	0.000000	0.000000	…	-0.043065	
25%	0.002882	0.008214	0.002579	0.020695	…	0.008966	
50%	0.005790	0.014166	0.006355	0.028457	…	0.017954	
75%	0.011026	0.024542	0.014718	0.039130	…	0.026473	
max	0.041696	0.106611	0.078999	0.108836	…	0.076009	

	X74	X75	X76	X77	X78	X79	\
count	872.000000	872.000000	872.000000	872.000000	872.000000	872.000000	
mean	0.002863	0.008101	0.017016	0.003096	0.008726	0.017123	
std	0.017101	0.017187	0.014577	0.017251	0.015908	0.014557	
min	-0.075061	-0.052011	-0.052863	-0.069618	-0.051907	-0.051155	
25%	-0.006349	-0.001361	0.008915	-0.006092	-0.000248	0.008419	
50%	0.003247	0.005998	0.018700	0.003761	0.006603	0.018370	
75%	0.013380	0.016748	0.026707	0.013085	0.017008	0.026695	
max	0.067803	0.085228	0.060338	0.073472	0.072732	0.069022	

	X80	X81	X82
count	872.000000	872.000000	872.000000
mean	0.003323	0.008580	0.017478
std	0.017238	0.017106	0.014550
min	-0.072393	-0.065650	-0.039432
25%	-0.005884	-0.000353	0.008986
50%	0.003905	0.006322	0.018324
75%	0.013547	0.016631	0.027222
max	0.068804	0.097735	0.075775

[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 3,
con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.039202  0.000000  0.054883  0.141128  0.031362  0.070564  0.015681
1    0.038610  0.004064  0.010161  0.000000  0.010161  0.018289  0.002032
2    0.047420  0.015807  0.023710  0.039516  0.055323  0.071130  0.003952
3    0.037934  0.018967  0.018967  0.063223  0.000000  0.025289  0.006322
4    0.029512  0.011805  0.053121  0.076731  0.070828  0.064926  0.035414
..        …         …         …         …         …         …         …
995  0.065732  0.005056  0.055619  0.055619  0.020225  0.040450  0.000000
996  0.033479  0.000000  0.028079  0.024839  0.003240  0.028079  0.004320
997  0.042643  0.002031  0.020306  0.010153  0.016245  0.022337  0.010153
998  0.083283  0.015142  0.083283  0.098425  0.037856  0.052998  0.022714
999  0.018537  0.000000  0.092685  0.148296  0.018537  0.037074  0.000000

           X7        X8        X9  …       X32       X33       X34  \
```

```
0    0.047043  0.023521  0.109766  …  0.063445 -0.000213 -0.017543
1    0.014225  0.004064  0.012193  …  0.043037  0.015588  0.053593
2    0.051371  0.051371  0.051371  …  0.025875 -0.001301  0.039705
3    0.044256  0.044256  0.012645  …  0.042402  0.029940  0.014568
4    0.123949  0.141657  0.047219  …  0.039332  0.030888  0.036978
..        …         …        …  …        …         …
995  0.065732  0.030338  0.045507  …  0.031025  0.052053  0.039585
996  0.006480  0.004320  0.021599  …  0.031959  0.027247  0.028214
997  0.004061  0.004061  0.040612  …  0.015933  0.029891  0.033552
998  0.022714  0.037856  0.090854  …  0.019460 -0.020343  0.030955
999  0.148296  0.203907  0.129759  … -0.125084  0.050623 -0.018469

          X35       X36       X37       X38       X39       X40        X41
0   -0.061894  0.011373  0.058144  0.000064 -0.084036 -0.028475  efectores
1    0.045948  0.049631  0.022915  0.033567  0.066264  0.035628  efectores
2   -0.004510  0.004042  0.044163  0.013862 -0.016827  0.000133  efectores
3   -0.013220  0.033971  0.036514  0.000386  0.050658  0.003920  efectores
4    0.012807 -0.031585 -0.055161  0.098403 -0.007913 -0.026999  efectores
..        …         …         …         …         …         …
995  0.032993 -0.012878  0.048671  0.006559  0.065090  0.021216  efectores
996  0.035348  0.030037  0.034610  0.026477  0.035752  0.023803  efectores
997  0.022827  0.029165  0.003338  0.014442  0.021363  0.017010  efectores
998 -0.032229 -0.015040  0.008885 -0.009435  0.059379  0.058215  efectores
999 -0.044477  0.126342 -0.068051  0.047190 -0.012121 -0.032397  efectores

[1000 rows x 42 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass efectores archaea dataset 3, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.045610     0.004303     0.034281     0.044876     0.023444
std       0.016849     0.006907     0.023501     0.038434     0.017594
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       0.034746     0.000000     0.016552     0.014634     0.011658
50%       0.042843     0.000886     0.028696     0.032465     0.018858
75%       0.053777     0.005800     0.048532     0.070073     0.030829
max       0.151824     0.052358     0.194617     0.219592     0.126104

                X5           X6           X7           X8           X9  …  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
mean      0.038842     0.010289     0.037510     0.032336     0.064256  …
std       0.012842     0.008949     0.033606     0.041181     0.032581  …
min       0.000000     0.000000     0.000000     0.000000     0.004942  …
25%       0.029396     0.003039     0.011941     0.003315     0.040859  …
```

```
50%       0.037397     0.008559     0.024539     0.010826     0.058338  …
75%       0.046297     0.015026     0.054478     0.050485     0.085097  …
max       0.112076     0.063563     0.189156     0.203907     0.263378  …

               X31          X32          X33          X34          X35  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.016871     0.022653     0.018613     0.019738     0.015629
std       0.024043     0.025110     0.031622     0.024347     0.026159
min      -0.185143    -0.125084    -0.177707    -0.117730    -0.109977
25%       0.004044     0.011048     0.006180     0.006940     0.003030
50%       0.020670     0.026060     0.022254     0.023857     0.021108
75%       0.033340     0.037800     0.034112     0.035337     0.033303
max       0.127547     0.114304     0.652221     0.111463     0.083173

               X36          X37          X38          X39          X40
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.020783     0.021377     0.020425     0.021228     0.022034
std       0.026258     0.025298     0.025836     0.025466     0.032697
min      -0.300764    -0.254841    -0.138490    -0.120892    -0.093768
25%       0.007768     0.009364     0.006893     0.008223     0.007367
50%       0.024571     0.025431     0.025632     0.025210     0.025712
75%       0.035514     0.036100     0.036953     0.037224     0.036598
max       0.132745     0.149029     0.115604     0.113606     0.656568

[8 rows x 41 columns]
```

no_efectores
Composición de pseudo aminoácidos (PseAAC) mass no_efectores archaea dataset 3, con valores atípicos.
Valores del documento csv.

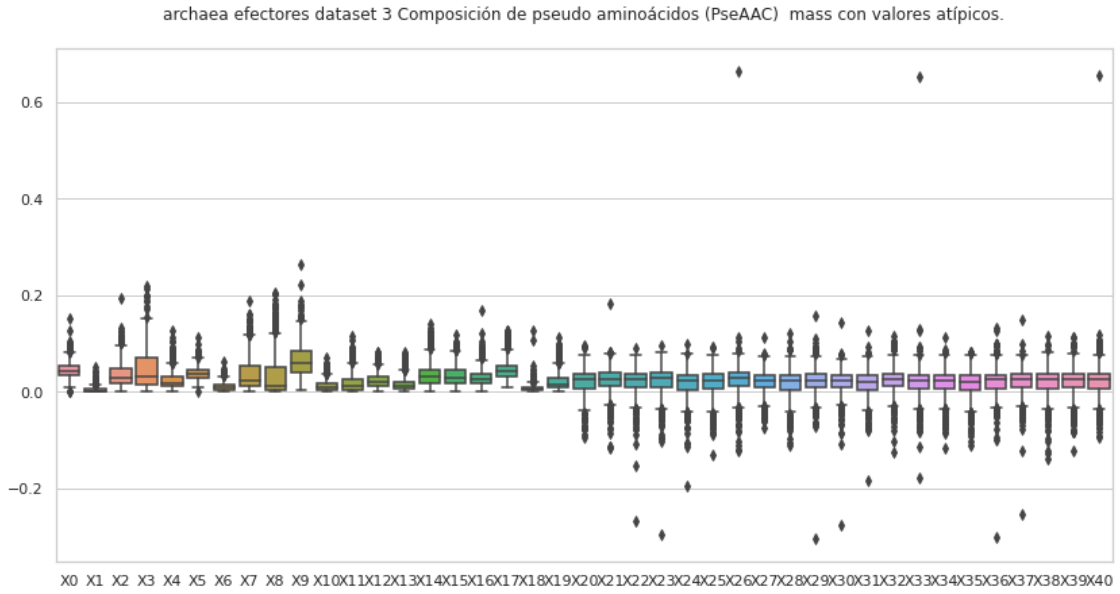
```
          X0        X1        X2        X3        X4        X5        X6  \
0    0.076430  0.028661  0.076430  0.105092  0.066877  0.047769  0.038215
1    0.050377  0.004122  0.037554  0.048088  0.014655  0.033432  0.013281
2    0.028778  0.006256  0.030030  0.016266  0.010010  0.032532  0.003754
3    0.008027  0.008027  0.024080  0.040133  0.064214  0.024080  0.008027
4    0.071772  0.005981  0.047848  0.074762  0.029905  0.044857  0.017943
..        …         …         …         …         …         …         …
995  0.033822  0.000805  0.010469  0.011274  0.015300  0.030601  0.000805
996  0.039999  0.000000  0.011428  0.031428  0.031428  0.051428  0.002857
997  0.048245  0.003446  0.037907  0.044799  0.010338  0.037907  0.000000
998  0.046999  0.003760  0.033839  0.050759  0.028199  0.063919  0.015040
999  0.054661  0.000000  0.048414  0.053100  0.018741  0.045291  0.006247

          X7        X8        X9  …       X32       X33       X34  \
0    0.085984  0.124200  0.076430  … -0.043919  0.042286 -0.106005
```

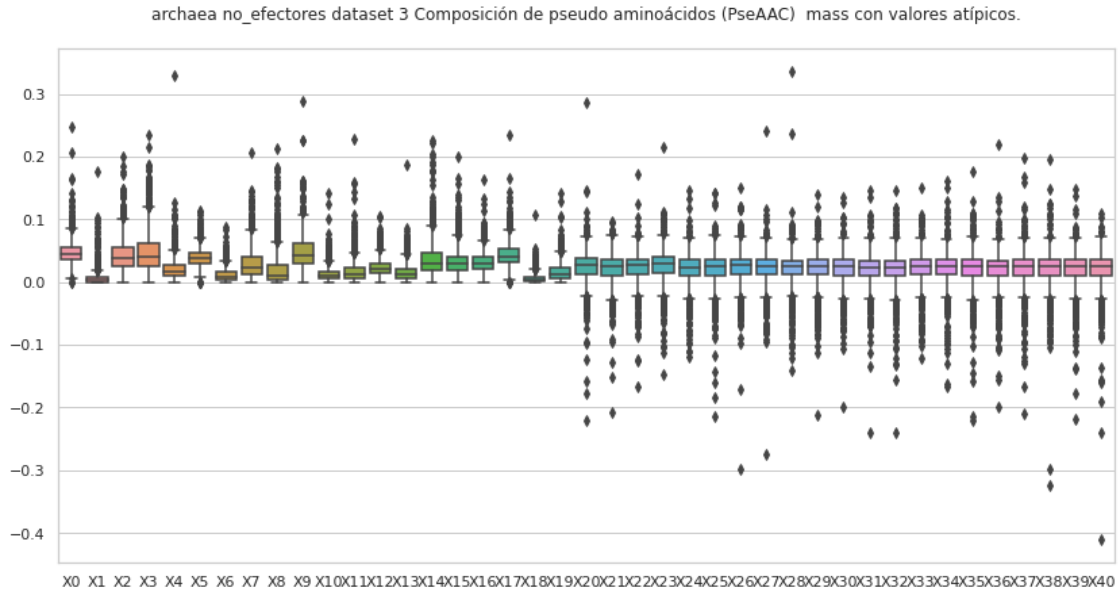
```
1    0.020151  0.004580  0.052209  …  0.028144  0.026647  0.018365
2    0.007507  0.000000  0.015015  …  0.039072  0.030599  0.025032
3    0.096320  0.048160  0.040133  … -0.014153  0.013710 -0.010422
4    0.014952  0.002990  0.059810  …  0.004428 -0.031184  0.012023
..        …         …        …  …         …         …
995  0.020937  0.000000  0.038654  …  0.023089  0.039683  0.029551
996  0.022857  0.005714  0.068570  … -0.002543  0.010729  0.030897
997  0.020676  0.027569  0.031015  …  0.024025  0.050984  0.011036
998  0.046999  0.031959  0.043239  …  0.019144  0.029257  0.029687
999  0.032797  0.007809  0.043729  …  0.058992  0.038667  0.022688

          X35       X36       X37       X38       X39       X40           X41
0   -0.038840  0.037279  0.018796 -0.076243  0.032495 -0.018155  no_efectores
1    0.022509  0.026050  0.033183  0.028605  0.017654  0.020893  no_efectores
2    0.042364  0.027731  0.030548  0.038743  0.023611  0.024847  no_efectores
3   -0.001730  0.037545  0.006532 -0.017539 -0.051124 -0.001196  no_efectores
4    0.022350 -0.003696  0.006762  0.069420  0.049499  0.034812  no_efectores
..        …         …        …         …         …         …           …
995  0.031719  0.027753  0.041208  0.026003  0.024582  0.046756  no_efectores
996 -0.000139  0.022687  0.034985  0.019690  0.052084  0.022684  no_efectores
997  0.043923 -0.000644  0.016423  0.047480  0.031487  0.036946  no_efectores
998  0.038991  0.011746  0.004394  0.014678  0.027966  0.010242  no_efectores
999  0.001171  0.011162  0.011237  0.020863  0.038874  0.024949  no_efectores

[1000 rows x 42 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass no_efectores archaea dataset 3, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.047384     0.006607     0.042588     0.048212     0.020631
std       0.020419     0.012836     0.026938     0.034410     0.018831
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       0.035961     0.000000     0.024686     0.024240     0.009641
50%       0.045027     0.002876     0.037937     0.041253     0.016433
75%       0.055653     0.007495     0.055407     0.062457     0.026624
max       0.247914     0.176345     0.200523     0.234607     0.329119

                X5           X6           X7           X8           X9  …  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
mean      0.039514     0.011629     0.029706     0.021074     0.048487  …
std       0.014580     0.011105     0.025613     0.028189     0.027684  …
min       0.000000     0.000000     0.000000     0.000000     0.000000  …
25%       0.030280     0.004212     0.011626     0.003957     0.029642  …
50%       0.037244     0.008824     0.022299     0.010073     0.043232  …
```

```
75%       0.046433     0.015885     0.040079     0.027890     0.062207  …
max       0.114293     0.088033     0.205700     0.212498     0.287980  …

                X31          X32          X33          X34          X35  \
count   1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean       0.020170     0.019916     0.021944     0.021400     0.020493
std        0.026107     0.028367     0.025435     0.026889     0.028789
min       -0.240395    -0.239933    -0.122118    -0.165749    -0.221437
25%        0.009667     0.008969     0.011665     0.011256     0.009556
50%        0.023971     0.024035     0.025365     0.024886     0.024202
75%        0.034241     0.033831     0.035952     0.035431     0.034847
max        0.145178     0.145001     0.151069     0.160719     0.175153

                X36          X37          X38          X39          X40
count   1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean       0.019827     0.020992     0.020238     0.020322     0.019353
std        0.027939     0.028897     0.031192     0.028172     0.031616
min       -0.200054    -0.210360    -0.323963    -0.218611    -0.410435
25%        0.010661     0.010622     0.010947     0.010469     0.010266
50%        0.024361     0.024329     0.024916     0.024499     0.024537
75%        0.034425     0.035003     0.035390     0.035285     0.035500
max        0.218489     0.197380     0.194557     0.147366     0.109814

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

	X0	X1	X2	X3	X4	X5	X6	\
1	0.038610	0.004064	0.010161	0.000000	0.010161	0.018289	0.002032	
2	0.047420	0.015807	0.023710	0.039516	0.055323	0.071130	0.003952	
3	0.037934	0.018967	0.018967	0.063223	0.000000	0.025289	0.006322	
5	0.051240	0.000000	0.056934	0.039853	0.011387	0.045547	0.017080	
7	0.046558	0.000000	0.023279	0.013302	0.014965	0.036581	0.006651	
..	…	…	…	…	…	…	…	
991	0.038229	0.002249	0.015741	0.013493	0.013493	0.037479	0.005247	
993	0.048441	0.000989	0.040532	0.024715	0.005932	0.018783	0.003954	
995	0.065732	0.005056	0.055619	0.055619	0.020225	0.040450	0.000000	
996	0.033479	0.000000	0.028079	0.024839	0.003240	0.028079	0.004320	
997	0.042643	0.002031	0.020306	0.010153	0.016245	0.022337	0.010153	

	X7	X8	X9	…	X32	X33	X34	\
1	0.014225	0.004064	0.012193	…	0.043037	0.015588	0.053593	
2	0.051371	0.051371	0.051371	…	0.025875	-0.001301	0.039705	
3	0.044256	0.044256	0.012645	…	0.042402	0.029940	0.014568	
5	0.028467	0.011387	0.074014	…	0.001809	0.026554	0.033912	
7	0.023279	0.003326	0.034919	…	0.025836	0.010078	0.050401	
..	…	…	…	…	…	…	…	
991	0.012743	0.003748	0.053220	…	0.043334	0.021157	0.018080	
993	0.012852	0.000989	0.022738	…	0.026207	0.022918	0.027686	
995	0.065732	0.030338	0.045507	…	0.031025	0.052053	0.039585	
996	0.006480	0.004320	0.021599	…	0.031959	0.027247	0.028214	
997	0.004061	0.004061	0.040612	…	0.015933	0.029891	0.033552	

	X35	X36	X37	X38	X39	X40	X41
1	0.045948	0.049631	0.022915	0.033567	0.066264	0.035628	efectores
2	-0.004510	0.004042	0.044163	0.013862	-0.016827	0.000133	efectores
3	-0.013220	0.033971	0.036514	0.000386	0.050658	0.003920	efectores
5	0.025705	0.042615	0.023868	0.029297	0.026163	-0.018123	efectores
7	0.036372	0.025002	0.043469	0.018901	0.014815	0.019861	efectores
..	…	…	…	…	…	…	
991	0.030414	0.030103	0.020735	0.026502	0.032870	0.025662	efectores
993	0.039079	0.036805	0.021058	0.026015	0.038988	0.041087	efectores
995	0.032993	-0.012878	0.048671	0.006559	0.065090	0.021216	efectores
996	0.035348	0.030037	0.034610	0.026477	0.035752	0.023803	efectores
997	0.022827	0.029165	0.003338	0.014442	0.021363	0.017010	efectores

[805 rows x 42 columns]

Composición de pseudo aminoácidos (PseAAC) mass efectores archaea dataset 3, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	805.000000	805.000000	805.000000	805.000000	805.000000	805.000000	
mean	0.044426	0.002957	0.028917	0.034573	0.020213	0.037165	
std	0.013897	0.004663	0.018658	0.029164	0.013847	0.010793	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.010009	
25%	0.035175	0.000000	0.014911	0.012666	0.010636	0.029015	
50%	0.042613	0.000646	0.024050	0.024576	0.016825	0.035805	
75%	0.052097	0.004409	0.039727	0.049900	0.026224	0.044831	
max	0.092361	0.024506	0.091175	0.149800	0.069303	0.072638	

	X6	X7	X8	X9	…	X31	\
count	805.000000	805.000000	805.000000	805.000000	…	805.000000	
mean	0.008818	0.029660	0.022104	0.055282	…	0.021112	
std	0.007224	0.026682	0.030360	0.025330	…	0.018408	
min	0.000000	0.000000	0.000000	0.004942	…	-0.037330	
25%	0.002756	0.010266	0.002795	0.037182	…	0.010378	
50%	0.007208	0.019955	0.006986	0.050691	…	0.023592	
75%	0.013218	0.041535	0.030708	0.071091	…	0.033924	
max	0.036815	0.136334	0.149592	0.137271	…	0.083108	

	X32	X33	X34	X35	X36	X37	\
count	805.000000	805.000000	805.000000	805.000000	805.000000	805.000000	
mean	0.026335	0.020526	0.023643	0.021341	0.023840	0.024796	
std	0.018910	0.020130	0.019476	0.020493	0.019184	0.018879	
min	-0.035407	-0.074493	-0.043243	-0.062174	-0.044463	-0.048247	
25%	0.016177	0.009901	0.012868	0.010834	0.013425	0.013595	
50%	0.027862	0.023656	0.026617	0.024607	0.026288	0.027547	
75%	0.038236	0.034367	0.036116	0.035023	0.035892	0.036603	

```
max      0.094280    0.085095    0.081917    0.083173    0.098350    0.075389

              X38         X39         X40
count  805.000000  805.000000  805.000000
mean     0.024346    0.025367    0.024925
std      0.019454    0.018957    0.019907
min     -0.055970   -0.040908   -0.053006
25%      0.013724    0.014042    0.012796
50%      0.027987    0.027715    0.027678
75%      0.037635    0.037940    0.036921
max      0.080617    0.096592    0.106842

[8 rows x 41 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass no_efectores archaea dataset 3, sin valores atípicos.
Valores del documento csv.

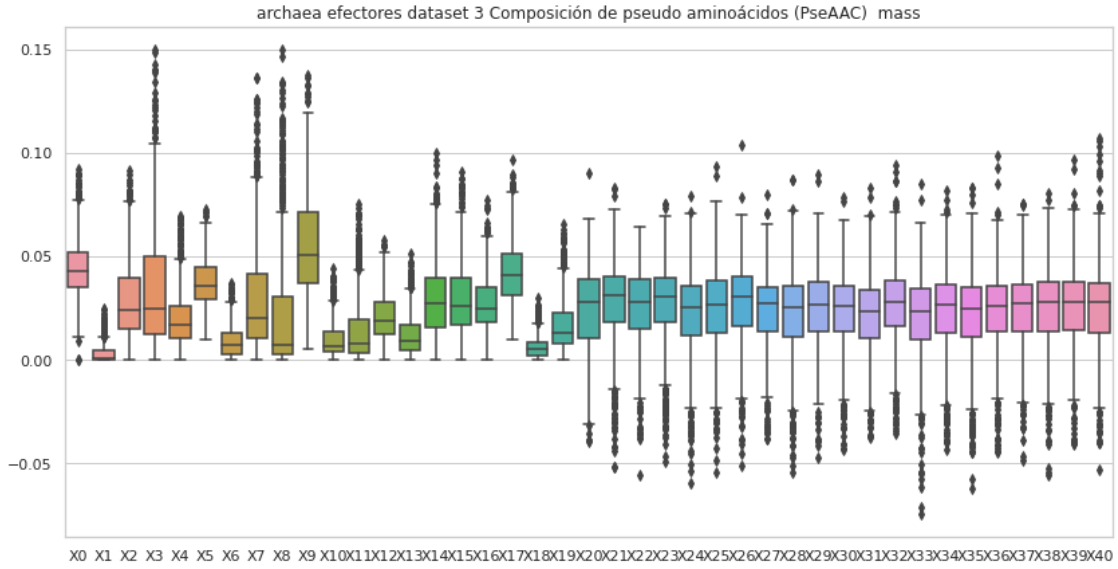
```
           X0        X1        X2        X3        X4        X5        X6  \
1    0.050377  0.004122  0.037554  0.048088  0.014655  0.033432  0.013281
2    0.028778  0.006256  0.030030  0.016266  0.010010  0.032532  0.003754
4    0.071772  0.005981  0.047848  0.074762  0.029905  0.044857  0.017943
6    0.037110  0.006958  0.052186  0.033631  0.023194  0.039430  0.010437
7    0.023686  0.016918  0.060906  0.054138  0.033836  0.040604  0.003384
..        …         …         …         …         …         …         …
995  0.033822  0.000805  0.010469  0.011274  0.015300  0.030601  0.000805
996  0.039999  0.000000  0.011428  0.031428  0.031428  0.051428  0.002857
997  0.048245  0.003446  0.037907  0.044799  0.010338  0.037907  0.000000
998  0.046999  0.003760  0.033839  0.050759  0.028199  0.063919  0.015040
999  0.054661  0.000000  0.048414  0.053100  0.018741  0.045291  0.006247

           X7        X8        X9  …       X32       X33       X34  \
1    0.020151  0.004580  0.052209  …  0.028144  0.026647  0.018365
2    0.007507  0.000000  0.015015  …  0.039072  0.030599  0.025032
4    0.014952  0.002990  0.059810  …  0.004428 -0.031184  0.012023
6    0.045228  0.039430  0.040589  …  0.036664  0.037114  0.013445
7    0.084591  0.071057  0.050755  …  0.019173  0.029348 -0.015419
..        …         …         …  …  …       …         …         …
995  0.020937  0.000000  0.038654  …  0.023089  0.039683  0.029551
996  0.022857  0.005714  0.068570  … -0.002543  0.010729  0.030897
997  0.020676  0.027569  0.031015  …  0.024025  0.050984  0.011036
998  0.046999  0.031959  0.043239  …  0.019144  0.029257  0.029687
999  0.032797  0.007809  0.043729  …  0.058992  0.038667  0.022688

           X35       X36       X37       X38       X39       X40           X41
1    0.022509  0.026050  0.033183  0.028605  0.017654  0.020893  no_efectores
```

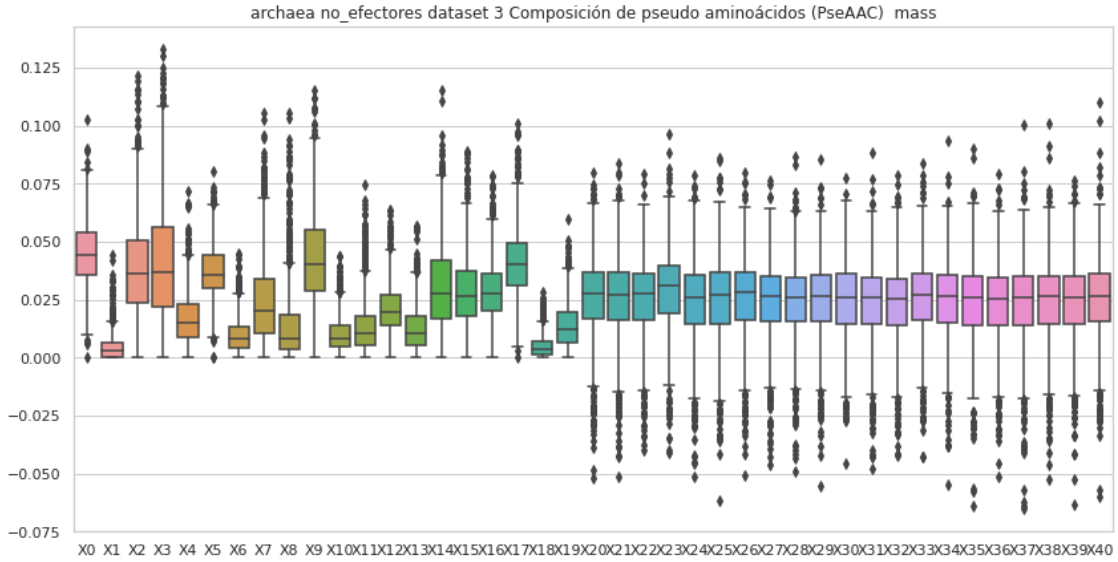
```
2    0.042364  0.027731  0.030548  0.038743  0.023611  0.024847  no_efectores
4    0.022350 -0.003696  0.006762  0.069420  0.049499  0.034812  no_efectores
6   -0.008056  0.024924  0.037307  0.021833  0.018948  0.018143  no_efectores
7    0.013936  0.038565  0.014179 -0.021200  0.039721  0.005253  no_efectores
..        …         …         …         …         …         …           …
995  0.031719  0.027753  0.041208  0.026003  0.024582  0.046756  no_efectores
996 -0.000139  0.022687  0.034985  0.019690  0.052084  0.022684  no_efectores
997  0.043923 -0.000644  0.016423  0.047480  0.031487  0.036946  no_efectores
998  0.038991  0.011746  0.004394  0.014678  0.027966  0.010242  no_efectores
999  0.001171  0.011162  0.011237  0.020863  0.038874  0.024949  no_efectores

[841 rows x 42 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass no_efectores archaea dataset 3, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  841.000000  841.000000  841.000000  841.000000  841.000000  841.000000
mean     0.044782    0.004704    0.038253    0.040886    0.017128    0.037353
std      0.014099    0.006110    0.021377    0.025316    0.011324    0.011142
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.035709    0.000000    0.023410    0.021683    0.008955    0.029981
50%      0.044118    0.002741    0.036125    0.036487    0.014874    0.035880
75%      0.053804    0.006256    0.050386    0.056498    0.023054    0.044306
max      0.102696    0.043945    0.121229    0.132893    0.071576    0.080516

                X6          X7          X8          X9  …         X31  \
count  841.000000  841.000000  841.000000  841.000000  …  841.000000
mean     0.009731    0.025142    0.015001    0.043316  …    0.023288
std      0.007746    0.019406    0.018234    0.020152  …    0.017606
min      0.000000    0.000000    0.000000    0.000000  …   -0.047805
25%      0.004130    0.010708    0.003560    0.028570  …    0.014181
50%      0.007936    0.020303    0.008086    0.040096  …    0.025902
75%      0.013534    0.034073    0.018383    0.055091  …    0.034729
max      0.044691    0.105591    0.105591    0.115331  …    0.088418

               X32         X33         X34         X35         X36         X37  \
count  841.000000  841.000000  841.000000  841.000000  841.000000  841.000000
mean     0.023212    0.025007    0.024565    0.023781    0.023343    0.023493
std      0.018052    0.016979    0.017203    0.017751    0.017449    0.018964
min     -0.042263   -0.042593   -0.054855   -0.064100   -0.051598   -0.065136
25%      0.013592    0.016055    0.015005    0.013936    0.013777    0.013763
50%      0.025467    0.026753    0.026196    0.025910    0.025483    0.025626
75%      0.034137    0.035909    0.035634    0.035056    0.034647    0.035108
max      0.078527    0.083949    0.093242    0.090145    0.079142    0.100392
```

```
              X38          X39          X40
count  841.000000   841.000000   841.000000
mean     0.023951     0.023842     0.025071
std      0.018227     0.018271     0.018426
min     -0.052652    -0.063367    -0.059963
25%      0.014638     0.014403     0.015808
50%      0.026181     0.026065     0.026610
75%      0.035267     0.035327     0.036163
max      0.100828     0.076119     0.109814

[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 3,
con valores atípicos.
Valores del documento csv.
```

	X0	X1	X2	X3	X4	X5	X6	\
0	0.022032	0.000000	0.030844	0.079314	0.017625	0.039657	0.008813	
1	0.074133	0.007803	0.019509	0.000000	0.019509	0.035116	0.003902	
2	0.032657	0.010886	0.016328	0.027214	0.038099	0.048985	0.002721	
3	0.051313	0.025656	0.025656	0.085521	0.000000	0.034208	0.008552	
4	0.024894	0.009958	0.044810	0.064725	0.059746	0.054767	0.029873	
..	…	…	…	…	…	…	…	
995	0.106923	0.008225	0.090473	0.090473	0.032899	0.065799	0.000000	
996	0.075200	0.000000	0.063071	0.055794	0.007277	0.063071	0.009703	
997	0.044904	0.002138	0.021383	0.010691	0.017106	0.023521	0.010691	
998	0.087105	0.015837	0.087105	0.102942	0.039593	0.055430	0.023756	
999	0.007680	0.000000	0.038402	0.061443	0.007680	0.015361	0.000000	

	X7	X8	X9	…	X53	X54	X55	\
0	0.026438	0.013219	0.061689	…	0.049275	0.025448	0.015546	
1	0.027312	0.007803	0.023410	…	0.003483	0.013368	0.002522	
2	0.035378	0.035378	0.035378	…	0.013525	-0.013953	-0.012468	
3	0.059865	0.059865	0.017104	…	0.023037	0.050526	0.078501	
4	0.104556	0.119492	0.039831	…	-0.050560	0.046312	0.008567	
..	…	…	…	…	…	…	…	
995	0.106923	0.049349	0.074024	…	0.060191	-0.034534	-0.002169	
996	0.014555	0.009703	0.048516	…	0.003387	-0.000593	0.021093	
997	0.004277	0.004277	0.042765	…	-0.007402	0.016431	-0.005522	
998	0.023756	0.039593	0.095023	…	-0.017127	-0.038352	-0.030675	
999	0.061443	0.084484	0.053763	…	0.012330	-0.007706	-0.022483	

	X56	X57	X58	X59	X60	X61	X62
0	0.041668	0.023218	0.046038	0.094223	-0.015111	0.041247	efectores
1	0.019228	0.002048	0.037930	0.016264	0.035171	0.022808	efectores
2	0.024272	0.036414	-0.014173	-0.011673	0.006451	0.008621	efectores
3	0.023521	0.069644	0.035978	0.048877	-0.007601	-0.003507	efectores
4	0.017228	0.056654	0.022420	0.077679	0.017700	-0.014014	efectores
..	…	…	…	…	…	…	…
995	-0.095107	-0.075436	-0.011558	-0.042123	-0.002832	-0.007064	efectores
996	0.003034	0.021005	0.006000	0.020004	-0.001881	-0.008618	efectores
997	0.025867	0.003083	0.034830	0.012466	0.031812	0.012368	efectores

```
998  0.129366  0.082681  0.057492  0.027539 -0.110023  0.001274  efectores
999 -0.072554  0.044579 -0.054539 -0.014138  0.064858  0.073259  efectores

[1000 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro efectores archaea dataset 3, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.049792     0.004620     0.035927     0.043834     0.022029
std       0.032017     0.007983     0.028283     0.033914     0.015814
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       0.029232     0.000000     0.012914     0.010919     0.011082
50%       0.042326     0.000853     0.029999     0.041492     0.018283
75%       0.062953     0.006031     0.054017     0.066739     0.028832
max       0.343287     0.071095     0.426568     0.238539     0.142189

                X5           X6           X7           X8           X9  …  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
mean      0.042050     0.010630     0.034936     0.029376     0.061562  …
std       0.025780     0.011327     0.032237     0.035952     0.033028  …
min       0.000000     0.000000     0.000000     0.000000     0.006179  …
25%       0.025896     0.003143     0.012891     0.003469     0.038164  …
50%       0.035663     0.008667     0.022305     0.011822     0.054886  …
75%       0.051829     0.014436     0.049583     0.049132     0.076405  …
max       0.284379     0.213284     0.301069     0.270186     0.355474  …

               X52          X53          X54          X55          X56  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.006873     0.008983     0.012977     0.013961     0.008755
std       0.040923     0.029232     0.039753     0.031935     0.035424
min      -0.281194    -0.242006    -0.500177    -0.207095    -0.257188
25%      -0.007080    -0.002299    -0.000229    -0.000365    -0.006905
50%       0.012932     0.007105     0.017886     0.009500     0.014553
75%       0.025736     0.020991     0.029252     0.027099     0.025809
max       0.474164     0.136506     0.164409     0.180802     0.201735

               X57          X58          X59          X60          X61
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.011684     0.008003     0.011277     0.008257     0.011843
std       0.029561     0.042495     0.032872     0.041832     0.030138
min      -0.140791    -0.565605    -0.357008    -0.456517    -0.176620
25%      -0.002549    -0.005476    -0.001701    -0.003599    -0.000638
50%       0.006667     0.014869     0.008970     0.015966     0.007923
75%       0.023344     0.027327     0.026041     0.028291     0.024965
```

```
max       0.178743     0.156926     0.180998     0.223536     0.158380

[8 rows x 62 columns]
```

no_efectores
Composición de pseudo aminoácidos (PseAAC) hidro no_efectores archaea dataset 3, con valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.063046  0.023642  0.063046  0.086688  0.055165  0.039404  0.031523
1    0.076080  0.006225  0.056714  0.072622  0.022132  0.050489  0.020057
2    0.069233  0.015051  0.072244  0.039132  0.024081  0.078264  0.009030
3    0.014896  0.014896  0.044689  0.074481  0.119170  0.044689  0.014896
4    0.079257  0.006605  0.052838  0.082559  0.033024  0.049536  0.019814
..        …         …         …         …         …         …         …
995  0.034467  0.000821  0.010668  0.011489  0.015592  0.031184  0.000821
996  0.042653  0.000000  0.012186  0.033513  0.033513  0.054839  0.003047
997  0.047332  0.003381  0.037189  0.043951  0.010143  0.037189  0.000000
998  0.077987  0.006239  0.056151  0.084226  0.046792  0.106063  0.024956
999  0.082055  0.000000  0.072678  0.079711  0.028133  0.067989  0.009378

           X7        X8        X9  …       X53       X54       X55  \
0    0.070927  0.102450  0.063046  … -0.053807  0.002130  0.050512
1    0.030432  0.006916  0.078847  …  0.025412  0.013028  0.024685
2    0.018061  0.000000  0.036122  …  0.017847  0.002550 -0.028005
3    0.178755  0.089377  0.074481  … -0.000435 -0.079847 -0.049663
4    0.016512  0.003302  0.066048  …  0.018718  0.052352  0.058916
..        …         …         …  …         …         …         …
995  0.021337  0.000000  0.039391  … -0.010384  0.028287 -0.002807
996  0.024373  0.006093  0.073119  … -0.020541  0.028139 -0.003711
997  0.020285  0.027047  0.030428  …  0.029178  0.033148  0.059592
998  0.077987  0.053031  0.071748  … -0.019231 -0.005476 -0.019882
999  0.049233  0.011722  0.065644  …  0.026244 -0.022875  0.002706

           X56       X57       X58       X59       X60       X61           X62
0    0.003208  0.035611  0.065902 -0.016587 -0.057598 -0.012779  no_efectores
1    0.017475  0.021754 -0.017627 -0.000448  0.000975  0.028649  no_efectores
2    0.005733  0.004030  0.031313  0.031468 -0.018456 -0.030791  no_efectores
3    0.057679  0.029225 -0.016038 -0.059590 -0.016103 -0.093625  no_efectores
4   -0.031350 -0.054799  0.018144  0.030707  0.083077  0.091672  no_efectores
..        …         …         …         …         …         …           …
995  0.041346  0.007420  0.033743  0.003716  0.023659 -0.004198  no_efectores
996 -0.015691 -0.044257  0.068393  0.012027 -0.001959 -0.004195  no_efectores
997  0.039813  0.006041 -0.037511  0.010876 -0.006827  0.021615  no_efectores
998  0.031134  0.034652  0.027400  0.023007  0.005648 -0.000447  no_efectores
```

```
999  0.000087 -0.021010  0.008987  0.032145  0.014649 -0.011319  no_efectores

[1000 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro no_efectores archaea dataset 3, con valores atípicos.
Estadísticas.

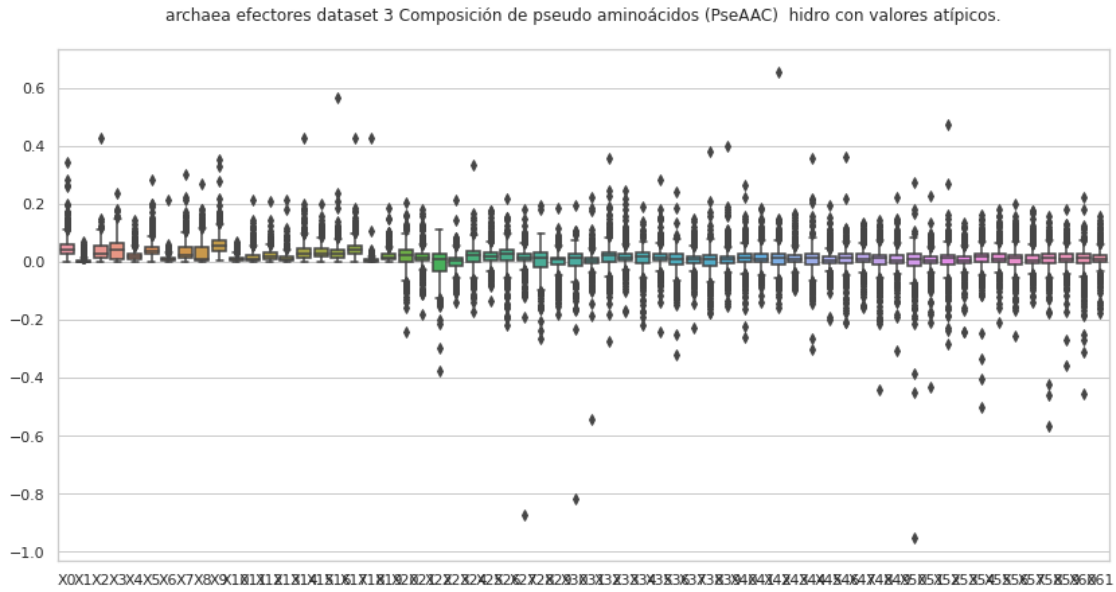
	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.060494	0.008265	0.049753	0.054178	0.022892	
std	0.048682	0.018094	0.028536	0.038699	0.019737	
min	-0.109358	-0.054679	-0.054679	-0.492109	-0.109358	
25%	0.035423	0.000000	0.033000	0.034317	0.012541	
50%	0.055301	0.003758	0.051347	0.054585	0.019165	
75%	0.076085	0.009510	0.066216	0.071649	0.028664	
max	1.007426	0.217910	0.377785	0.377785	0.276517	

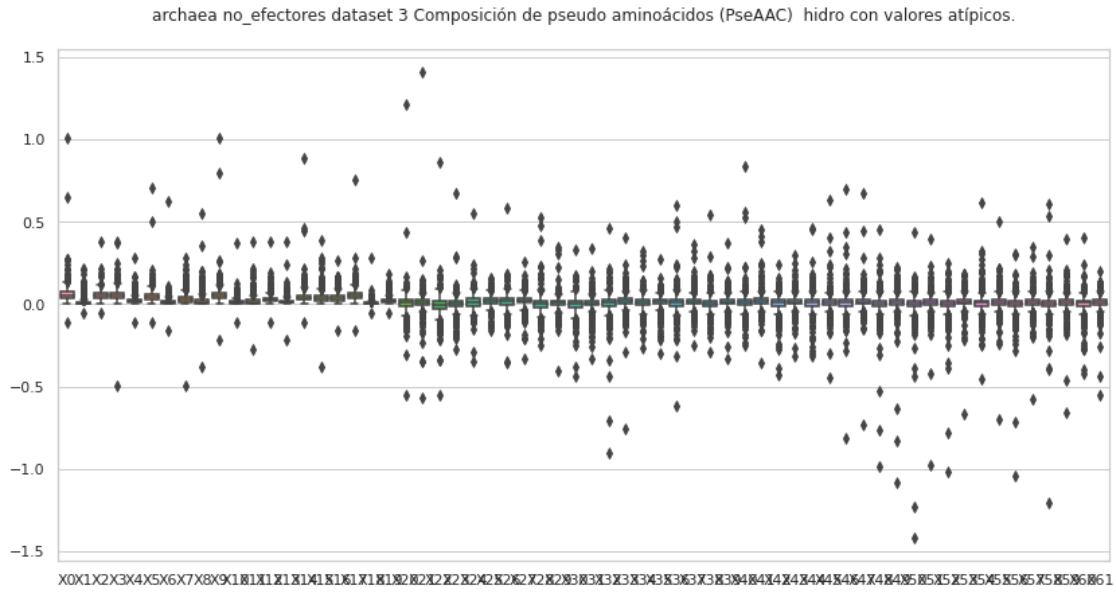
	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.050649	0.014074	0.032847	0.022204	0.056502	…	
std	0.037292	0.023525	0.034469	0.035863	0.048514	…	
min	-0.109358	-0.164036	-0.492109	-0.382751	-0.218715	…	
25%	0.030014	0.004673	0.014852	0.004965	0.036588	…	
50%	0.045427	0.011250	0.025181	0.012055	0.049988	…	
75%	0.063777	0.019115	0.042418	0.027061	0.068396	…	
max	0.708303	0.619765	0.276517	0.553034	1.007426	…	

	X52	X53	X54	X55	X56	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.002269	0.013651	0.006702	0.012756	0.001233	
std	0.058167	0.039994	0.048336	0.044810	0.056978	
min	-1.021464	-0.663639	-0.457349	-0.698421	-1.044419	
25%	-0.011710	-0.001339	-0.010876	-0.001969	-0.012560	
50%	0.006274	0.012471	0.006862	0.012612	0.005410	
75%	0.023114	0.030417	0.022646	0.028370	0.022167	
max	0.246313	0.186270	0.618717	0.498909	0.302494	

	X57	X58	X59	X60	X61
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.010382	0.002735	0.012532	0.002519	0.011147
std	0.040408	0.062117	0.043733	0.045405	0.041329
min	-0.578311	-1.209794	-0.659605	-0.424893	-0.552669
25%	-0.003808	-0.012706	-0.001383	-0.011619	-0.002473
50%	0.010176	0.005851	0.011588	0.006913	0.011306
75%	0.027166	0.021429	0.027881	0.022615	0.029018
max	0.347193	0.610393	0.395585	0.402803	0.201285

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.022032	0.000000	0.030844	0.079314	0.017625	0.039657	0.008813	
1	0.074133	0.007803	0.019509	0.000000	0.019509	0.035116	0.003902	
2	0.032657	0.010886	0.016328	0.027214	0.038099	0.048985	0.002721	
3	0.051313	0.025656	0.025656	0.085521	0.000000	0.034208	0.008552	
6	0.018577	0.003715	0.026008	0.048301	0.014862	0.018577	0.007431	
..	…	…	…	…	…	…	…	
993	0.093598	0.001910	0.078317	0.047754	0.011461	0.036293	0.007641	
994	0.035550	0.000000	0.013911	0.007728	0.023185	0.026276	0.012365	
995	0.106923	0.008225	0.090473	0.090473	0.032899	0.065799	0.000000	
996	0.075200	0.000000	0.063071	0.055794	0.007277	0.063071	0.009703	
997	0.044904	0.002138	0.021383	0.010691	0.017106	0.023521	0.010691	

	X7	X8	X9	…	X53	X54	X55	\
0	0.026438	0.013219	0.061689	…	0.049275	0.025448	0.015546	
1	0.027312	0.007803	0.023410	…	0.003483	0.013368	0.002522	
2	0.035378	0.035378	0.035378	…	0.013525	-0.013953	-0.012468	
3	0.059865	0.059865	0.017104	…	0.023037	0.050526	0.078501	
6	0.029724	0.074309	0.052016	…	-0.013642	0.008462	0.040918	
..	…	…	…	…	…	…	…	
993	0.024832	0.001910	0.043934	…	-0.001571	0.011277	0.033943	
994	0.012365	0.006183	0.034004	…	0.004481	0.019398	0.007825	
995	0.106923	0.049349	0.074024	…	0.060191	-0.034534	-0.002169	
996	0.014555	0.009703	0.048516	…	0.003387	-0.000593	0.021093	
997	0.004277	0.004277	0.042765	…	-0.007402	0.016431	-0.005522	

	X56	X57	X58	X59	X60	X61	X62
0	0.041668	0.023218	0.046038	0.094223	-0.015111	0.041247	efectores
1	0.019228	0.002048	0.037930	0.016264	0.035171	0.022808	efectores
2	0.024272	0.036414	-0.014173	-0.011673	0.006451	0.008621	efectores
3	0.023521	0.069644	0.035978	0.048877	-0.007601	-0.003507	efectores
6	-0.014048	0.054961	-0.029378	-0.001450	-0.052288	-0.013968	efectores
..	…	…	…	…	…	…	…
993	-0.014369	-0.018099	0.016251	0.007530	0.004506	0.021000	efectores
994	0.019724	0.016012	0.027052	0.008715	0.011139	-0.002086	efectores
995	-0.095107	-0.075436	-0.011558	-0.042123	-0.002832	-0.007064	efectores
996	0.003034	0.021005	0.006000	0.020004	-0.001881	-0.008618	efectores
997	0.025867	0.003083	0.034830	0.012466	0.031812	0.012368	efectores

```
[824 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC) efectores archaea dataset 3, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	824.000000	824.000000	824.000000	824.000000	824.000000	824.000000	
mean	0.046135	0.003113	0.030478	0.035501	0.019059	0.037702	
std	0.023748	0.005107	0.022047	0.027593	0.011241	0.018496	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.002523	
25%	0.028781	0.000000	0.011324	0.009482	0.010579	0.024896	
50%	0.040873	0.000510	0.023874	0.030826	0.017063	0.033529	
75%	0.060627	0.004138	0.046696	0.056922	0.025450	0.048400	
max	0.138279	0.027661	0.103607	0.113937	0.068709	0.112068	

	X6	X7	X8	X9	…	X52	\
count	824.000000	824.000000	824.000000	824.000000	…	824.000000	
mean	0.008739	0.027224	0.020942	0.053464	…	0.010360	
std	0.007274	0.022077	0.026283	0.022150	…	0.023615	
min	0.000000	0.000000	0.000000	0.006179	…	-0.099212	
25%	0.002582	0.011729	0.002738	0.036382	…	-0.000794	
50%	0.007751	0.018931	0.007819	0.048861	…	0.014810	
75%	0.012505	0.037214	0.030841	0.066666	…	0.025856	
max	0.044024	0.126224	0.124374	0.156005	…	0.106248	

	X53	X54	X55	X56	X57	X58	\
count	824.000000	824.000000	824.000000	824.000000	824.000000	824.000000	
mean	0.010135	0.015454	0.013683	0.010811	0.010714	0.012643	
std	0.018276	0.023281	0.022692	0.023047	0.021379	0.023173	
min	-0.044817	-0.074814	-0.076229	-0.095107	-0.075436	-0.087527	
25%	-0.000513	0.004527	0.001264	-0.003542	-0.001699	0.000189	
50%	0.007120	0.018344	0.009378	0.015563	0.006402	0.016723	
75%	0.018873	0.027630	0.023468	0.025567	0.020064	0.027322	
max	0.081987	0.104970	0.095770	0.111512	0.089250	0.102539	

	X59	X60	X61
count	824.000000	824.000000	824.000000
mean	0.012548	0.012048	0.011470
std	0.022476	0.025469	0.019949
min	-0.057562	-0.110266	-0.066770
25%	-0.000399	-0.001233	0.000290
50%	0.009005	0.016824	0.007556
75%	0.023379	0.028153	0.020975
max	0.100634	0.081620	0.091302

[8 rows x 62 columns]

no_efectores
Composición de pseudo aminoácidos (PseAAC) no_efectores archaea dataset 3, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.063046  0.023642  0.063046  0.086688  0.055165  0.039404  0.031523
1    0.076080  0.006225  0.056714  0.072622  0.022132  0.050489  0.020057
2    0.069233  0.015051  0.072244  0.039132  0.024081  0.078264  0.009030
4    0.079257  0.006605  0.052838  0.082559  0.033024  0.049536  0.019814
5    0.003477  0.000000  0.017385  0.012749  0.001159  0.010431  0.001159
..        …         …         …         …         …         …
995  0.034467  0.000821  0.010668  0.011489  0.015592  0.031184  0.000821
996  0.042653  0.000000  0.012186  0.033513  0.033513  0.054839  0.003047
997  0.047332  0.003381  0.037189  0.043951  0.010143  0.037189  0.000000
998  0.077987  0.006239  0.056151  0.084226  0.046792  0.106063  0.024956
999  0.082055  0.000000  0.072678  0.079711  0.028133  0.067989  0.009378

           X7        X8        X9  …       X53       X54       X55  \
0    0.070927  0.102450  0.063046  … -0.053807  0.002130  0.050512
1    0.030432  0.006916  0.078847  …  0.025412  0.013028  0.024685
2    0.018061  0.000000  0.036122  …  0.017847  0.002550 -0.028005
4    0.016512  0.003302  0.066048  …  0.018718  0.052352  0.058916
5    0.002318  0.000000  0.008113  …  0.033244  0.011246  0.034615
..        …         …         …  …       …         …         …
995  0.021337  0.000000  0.039391  … -0.010384  0.028287 -0.002807
996  0.024373  0.006093  0.073119  … -0.020541  0.028139 -0.003711
997  0.020285  0.027047  0.030428  …  0.029178  0.033148  0.059592
998  0.077987  0.053031  0.071748  … -0.019231 -0.005476 -0.019882
999  0.049233  0.011722  0.065644  …  0.026244 -0.022875  0.002706

          X56       X57       X58       X59       X60       X61           X62
0    0.003208  0.035611  0.065902 -0.016587 -0.057598 -0.012779  no_efectores
1    0.017475  0.021754 -0.017627 -0.000448  0.000975  0.028649  no_efectores
2    0.005733  0.004030  0.031313  0.031468 -0.018456 -0.030791  no_efectores
4   -0.031350 -0.054799  0.018144  0.030707  0.083077  0.091672  no_efectores
5    0.002368  0.019001 -0.005554  0.029412  0.018389  0.037282  no_efectores
..        …         …         …         …         …         …
995  0.041346  0.007420  0.033743  0.003716  0.023659 -0.004198  no_efectores
996 -0.015691 -0.044257  0.068393  0.012027 -0.001959 -0.004195  no_efectores
997  0.039813  0.006041 -0.037511  0.010876 -0.006827  0.021615  no_efectores
998  0.031134  0.034652  0.027400  0.023007  0.005648 -0.000447  no_efectores
999  0.000087 -0.021010  0.008987  0.032145  0.014649 -0.011319  no_efectores

[889 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC) no_efectores archaea dataset 3, sin valores atípicos.
Estadísticas.

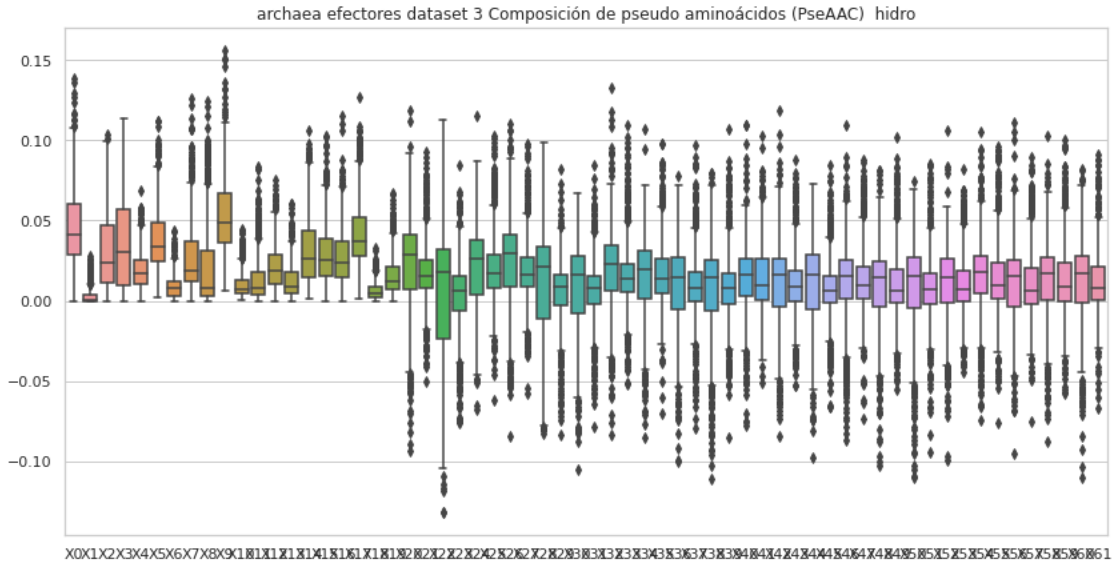
	X0	X1	X2	X3	X4	X5	\
count	889.000000	889.000000	889.000000	889.000000	889.000000	889.000000	
mean	0.057242	0.005741	0.047892	0.050345	0.020245	0.046871	
std	0.028801	0.007456	0.024125	0.026393	0.012317	0.022573	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.034921	0.000000	0.031742	0.032955	0.012258	0.029651	
50%	0.054579	0.003428	0.050844	0.052602	0.018381	0.044131	
75%	0.075255	0.008119	0.064744	0.068385	0.026027	0.060182	
max	0.185780	0.052232	0.105304	0.162819	0.077158	0.135657	

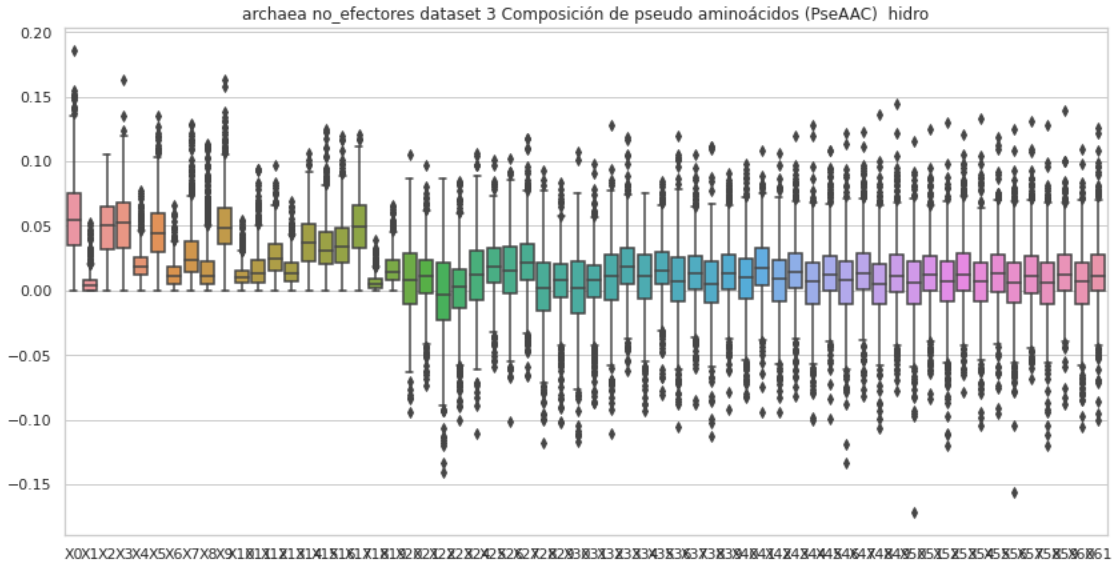
	X6	X7	X8	X9	…	X52	\
count	889.000000	889.000000	889.000000	889.000000	…	889.000000	
mean	0.012687	0.028872	0.017943	0.051327	…	0.005185	
std	0.010293	0.021502	0.020327	0.023260	…	0.027413	
min	0.000000	0.000000	0.000000	0.000000	…	-0.119864	
25%	0.004731	0.014269	0.004676	0.035634	…	-0.008863	
50%	0.010727	0.023775	0.011048	0.048253	…	0.006713	
75%	0.017968	0.038132	0.023038	0.063566	…	0.022798	
max	0.066246	0.128994	0.113272	0.163388	…	0.129600	

	X53	X54	X55	X56	X57	X58	\
count	889.000000	889.000000	889.000000	889.000000	889.000000	889.000000	
mean	0.014468	0.006102	0.013976	0.005048	0.012574	0.005139	
std	0.025241	0.026547	0.024710	0.027387	0.025218	0.026727	
min	-0.079193	-0.104860	-0.094798	-0.156530	-0.068191	-0.119716	
25%	-0.000201	-0.008607	-0.000834	-0.009822	-0.001858	-0.010516	
50%	0.012420	0.007244	0.013177	0.005909	0.010792	0.006408	
75%	0.029011	0.021563	0.027899	0.021877	0.027023	0.021201	
max	0.120666	0.133017	0.118427	0.124629	0.131420	0.127717	

	X59	X60	X61
count	889.000000	889.000000	889.000000
mean	0.013777	0.005228	0.013114
std	0.024072	0.026680	0.025807
min	-0.088207	-0.105726	-0.100032
25%	0.000050	-0.010217	-0.000446
50%	0.011944	0.007185	0.011451
75%	0.027190	0.022013	0.028154
max	0.139199	0.109165	0.125758

[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 3,
con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.114843 -0.196889 -0.044435  0.098150  0.006701 -0.062600 -0.081355
1   -0.186083 -0.092446  0.045695  0.029906  0.234190 -0.396699  0.088197
2    0.022518 -0.049605  0.094243 -0.005566 -0.089440 -0.011582  0.100343
3   -0.039257  0.005039  0.116611  0.039954  0.034853  0.073094  0.005434
4   -0.073199 -0.016393 -0.076531  0.093333 -0.091301  0.066326 -0.020607
..        …         …         …         …         …         …         …
995  0.074901  0.107089 -0.062123  0.114744 -0.103667  0.004536  0.102645
996  0.009365  0.081069 -0.005977  0.048423  0.062020  0.081839  0.102280
997  0.180901  0.146636  0.216451  0.151302  0.227469  0.220173  0.163990
998  0.050236 -0.080113  0.009945  0.010906 -0.001176 -0.102179 -0.013181
999 -0.128161 -0.017222  0.145927 -0.078703 -0.033723 -0.051159 -0.013460

           X7        X8        X9       X10       X11       X12        X13
0    0.029848  0.231728 -0.007268 -0.088224 -0.061602  0.108636  efectores
```

```
1    0.028338 -0.019930  0.127342 -0.082932  0.241337  0.091652  efectores
2   -0.067346  0.021595 -0.009946 -0.062915 -0.005177  0.027914  efectores
3    0.062402 -0.082918  0.063462 -0.038027  0.125903  0.088533  efectores
4   -0.020088 -0.018200  0.039638 -0.112889 -0.010651  0.077114  efectores
..        …        …        …        …        …        …
995 -0.070860 -0.151738 -0.133297 -0.031298 -0.047527  0.054482  efectores
996  0.011582  0.063357 -0.003469  0.012590  0.054926  0.016833  efectores
997  0.141502  0.037710  0.124105  0.085191 -0.013688 -0.057822  efectores
998  0.033659 -0.200928 -0.007268  0.051083 -0.017091  0.031338  efectores
999  0.074499 -0.022332 -0.108072  0.083329 -0.089329 -0.203880  efectores

[1000 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro_mass efectores archaea dataset 3, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.008453     0.018032     0.008607     0.015441    -0.007213
std       0.069059     0.070693     0.069076     0.069254     0.071354
min      -0.268809    -0.303772    -0.524088    -0.579639    -0.272140
25%      -0.035686    -0.022661    -0.033850    -0.025328    -0.048729
50%       0.006395     0.019816     0.008259     0.019759    -0.003279
75%       0.053152     0.059877     0.051540     0.057825     0.040378
max       0.245221     0.285015     0.226452     0.222028     0.234190

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.003959     0.020215     0.000804    -0.003513     0.009182
std       0.071357     0.077691     0.064536     0.074582     0.073594
min      -0.396699    -0.229729    -0.266405    -0.349572    -0.592968
25%      -0.035785    -0.023653    -0.038767    -0.048842    -0.032120
50%       0.007250     0.020073     0.003916    -0.004453     0.007127
75%       0.049705     0.060625     0.041062     0.043789     0.053259
max       0.221366     1.283083     0.238644     0.231728     0.256232

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.002246    -0.006680     0.014876
std       0.069340     0.064274     0.071045
min      -0.538224    -0.270956    -0.221888
25%      -0.036628    -0.046745    -0.032681
50%       0.003891    -0.004788     0.009454
75%       0.045174     0.032971     0.061090
max       0.228148     0.241337     0.266121
```

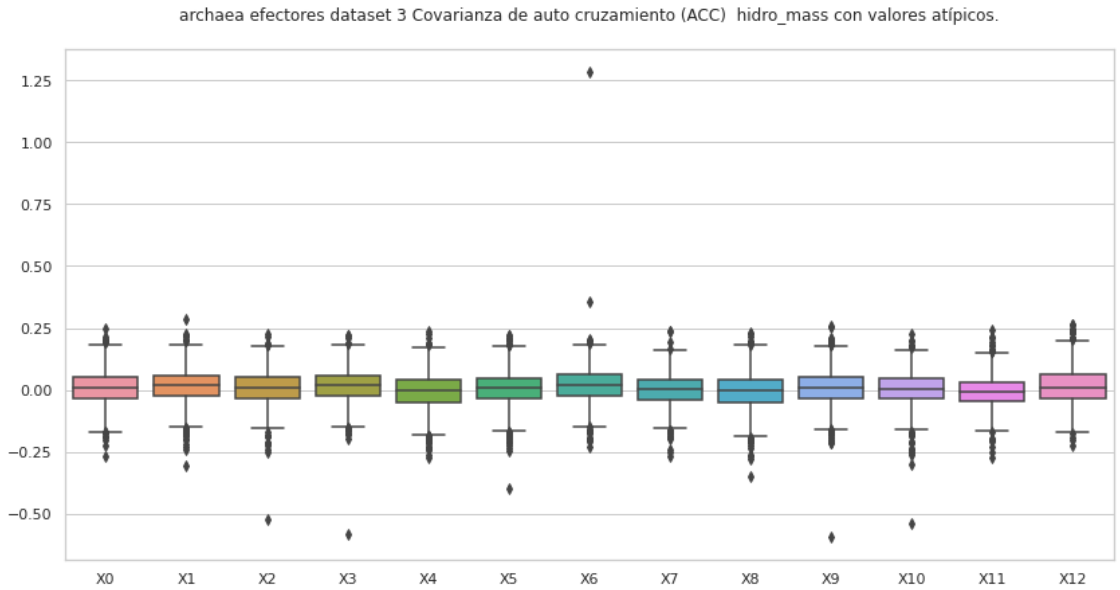
no_efectores
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores archaea dataset 3,
con valores atípicos.
Valores del documento csv.

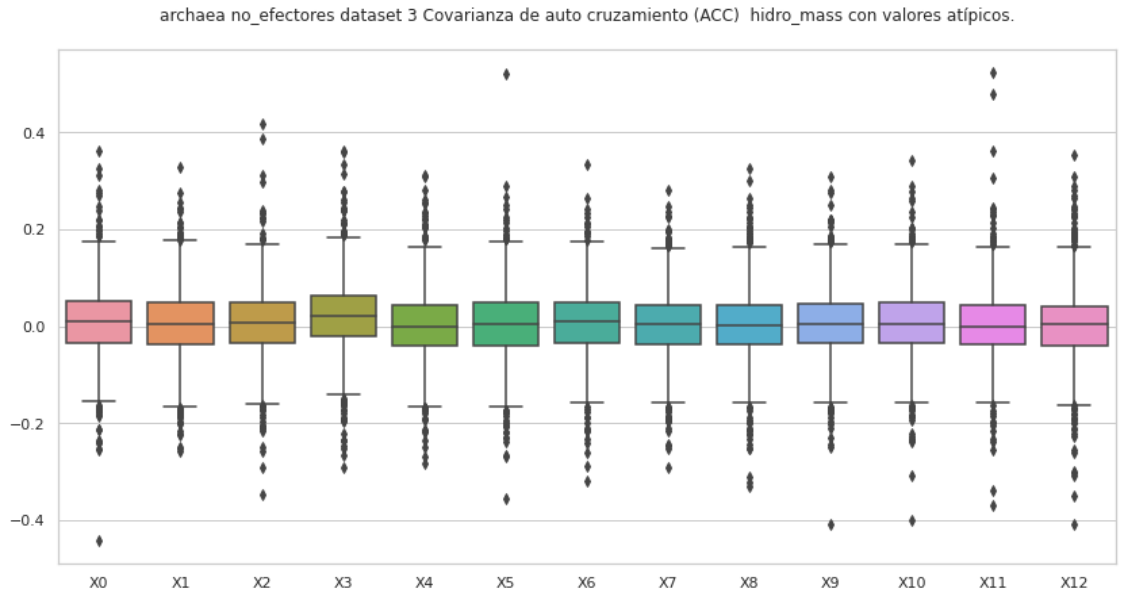
```
           X0        X1        X2        X3        X4        X5        X6  \
0   -0.093575 -0.054025  0.111097 -0.029496 -0.002633  0.071206 -0.099662
1    0.053833  0.003480  0.034158 -0.010562  0.016269 -0.013977 -0.025022
2    0.034271 -0.089845  0.179669  0.164192 -0.028484  0.053993  0.097806
3    0.065695 -0.054165  0.052648  0.171342  0.012820  0.050907  0.050346
4    0.019685 -0.172411 -0.007750  0.069830 -0.130615  0.063996  0.108494
..        …       …       …       …       …       …       …
995  0.072839 -0.033914  0.023246  0.114870 -0.019730  0.026667  0.045837
996 -0.000271 -0.071894  0.145556 -0.021958  0.089256  0.081605 -0.048531
997 -0.061372 -0.059628  0.092630 -0.027111 -0.049593  0.052719  0.145747
998 -0.056212 -0.055255  0.004168  0.013530  0.070374  0.104577 -0.007666
999 -0.092469 -0.015937 -0.035896 -0.026836 -0.033983 -0.049760  0.169589

           X7        X8        X9       X10       X11       X12           X13
0   -0.016025  0.024374 -0.042007 -0.073815 -0.006026 -0.077929  no_efectores
1   -0.043154 -0.002770 -0.062808 -0.010051  0.013118  0.019045  no_efectores
2    0.011149  0.057423  0.096364  0.103880  0.071016  0.064433  no_efectores
3   -0.000056 -0.090361  0.011262  0.081510 -0.097293 -0.048842  no_efectores
4   -0.068257 -0.054625  0.010700  0.015132  0.067584 -0.044183  no_efectores
..        …       …       …       …       …       …           …
995  0.022863 -0.015005  0.047525 -0.028106 -0.050720 -0.087528  no_efectores
996  0.024978  0.190676  0.038631  0.186926  0.141108 -0.114943  no_efectores
997 -0.031989  0.061052 -0.025347 -0.027390  0.035343  0.035416  no_efectores
998 -0.034841  0.005925 -0.006677 -0.000882 -0.084647  0.016582  no_efectores
999 -0.093887 -0.044394 -0.010495  0.022586 -0.018815  0.126722  no_efectores

[1000 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores archaea dataset 3,
con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.009299     0.003970     0.005921     0.019354     0.000940
std       0.076820     0.074941     0.073639     0.078057     0.073723
min      -0.442782    -0.258500    -0.348365    -0.293199    -0.283444
25%      -0.033877    -0.037957    -0.034076    -0.021005    -0.041411
50%       0.010987     0.003696     0.006147     0.020079    -0.002081
75%       0.050632     0.048405     0.050186     0.063080     0.042955
```

```
max       0.361488     0.328654     0.418595     0.360585     0.311632

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.003018     0.009038     0.001611     0.002875     0.005176
std       0.076944     0.073203     0.070696     0.074537     0.071654
min      -0.357161    -0.318640    -0.292956    -0.330434    -0.408843
25%      -0.039469    -0.034663    -0.037301    -0.038293    -0.034544
50%       0.004567     0.010806     0.003335     0.000871     0.004631
75%       0.047903     0.050226     0.043947     0.044489     0.047440
max       0.522532     0.333901     0.280582     0.325811     0.307657

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.004356     0.002544     0.002591
std       0.073132     0.074399     0.079116
min      -0.399831    -0.371200    -0.409532
25%      -0.035700    -0.037726    -0.040449
50%       0.003365    -0.001121     0.003871
75%       0.047850     0.043578     0.041648
max       0.343566     0.523727     0.353039
```





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 3,
sin valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
2    0.022518 -0.049605  0.094243 -0.005566 -0.089440 -0.011582  0.100343
3   -0.039257  0.005039  0.116611  0.039954  0.034853  0.073094  0.005434
4   -0.073199 -0.016393 -0.076531  0.093333 -0.091301  0.066326 -0.020607
5   -0.099857 -0.088624  0.111897 -0.007385 -0.135371  0.058074  0.157932
6   -0.006381 -0.080682 -0.027686 -0.023641 -0.098453 -0.010610  0.104092
..        …         …         …         …         …         …         …
992  0.014630 -0.003432 -0.060542  0.125027 -0.011092  0.020385 -0.019435
993  0.063303 -0.020508  0.042068 -0.038679 -0.014255  0.002170 -0.017359
995  0.074901  0.107089 -0.062123  0.114744 -0.103667  0.004536  0.102645
996  0.009365  0.081069 -0.005977  0.048423  0.062020  0.081839  0.102280
998  0.050236 -0.080113  0.009945  0.010906 -0.001176 -0.102179 -0.013181

           X7        X8        X9       X10       X11       X12        X13
2   -0.067346  0.021595 -0.009946 -0.062915 -0.005177  0.027914  efectores
3    0.062402 -0.082918  0.063462 -0.038027  0.125903  0.088533  efectores
4   -0.020088 -0.018200  0.039638 -0.112889 -0.010651  0.077114  efectores
5    0.014303 -0.093978 -0.016411  0.068438 -0.045543 -0.080704  efectores
6   -0.015475  0.008513  0.008183  0.038664 -0.101622 -0.008161  efectores
..        …         …         …         …         …         …         …
992  0.083743 -0.018752  0.085996 -0.078040 -0.082516 -0.089319  efectores
```

```
993  0.064970 -0.026873 -0.015315  0.022068  0.032185 -0.025146  efectores
995 -0.070860 -0.151738 -0.133297 -0.031298 -0.047527  0.054482  efectores
996  0.011582  0.063357 -0.003469  0.012590  0.054926  0.016833  efectores
998  0.033659 -0.200928 -0.007268  0.051083 -0.017091  0.031338  efectores

[927 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro_mass efectores archaea dataset 3,
sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	927.000000	927.000000	927.000000	927.000000	927.000000	927.000000	
mean	0.007808	0.018783	0.012679	0.015882	-0.006933	0.005811	
std	0.064957	0.065800	0.062999	0.065255	0.066397	0.066018	
min	-0.189932	-0.188604	-0.187324	-0.182319	-0.210908	-0.205896	
25%	-0.034552	-0.022251	-0.029180	-0.025341	-0.047098	-0.034240	
50%	0.005725	0.019355	0.011147	0.019943	-0.003276	0.007674	
75%	0.049967	0.058766	0.054442	0.057136	0.038367	0.049003	
max	0.204771	0.227640	0.187346	0.188868	0.183584	0.207431	

	X6	X7	X8	X9	X10	X11	\
count	927.000000	927.000000	927.000000	927.000000	927.000000	927.000000	
mean	0.021174	0.000614	-0.003400	0.009615	0.004647	-0.005248	
std	0.062847	0.060053	0.068890	0.066837	0.061707	0.059573	
min	-0.195450	-0.191421	-0.221686	-0.189537	-0.181470	-0.198226	
25%	-0.017716	-0.036479	-0.047110	-0.030890	-0.035008	-0.043857	
50%	0.021329	0.003635	-0.004479	0.007046	0.003931	-0.004235	
75%	0.061059	0.040014	0.042025	0.052199	0.043915	0.032886	
max	0.198039	0.159603	0.215797	0.210233	0.200565	0.182102	

	X12
count	927.000000
mean	0.015298
std	0.066225
min	-0.176796
25%	-0.030819
50%	0.010482
75%	0.060232
max	0.211342

no_efectores
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores archaea dataset 3,
sin valores atípicos.
Valores del documento csv.

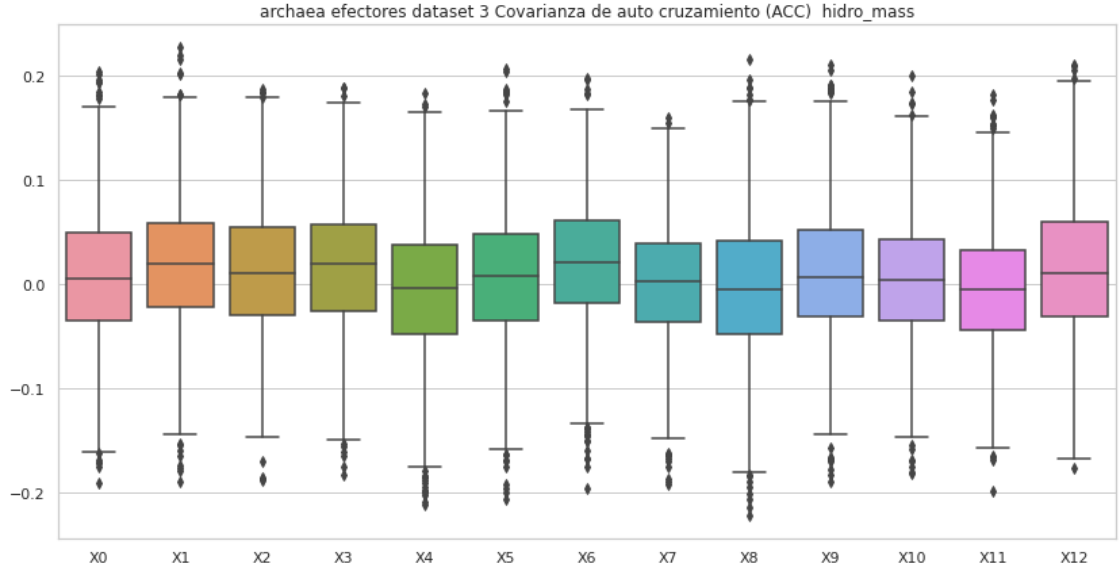
```
           X0        X1        X2        X3        X4        X5        X6  \
0   -0.093575 -0.054025  0.111097 -0.029496 -0.002633  0.071206 -0.099662
1    0.053833  0.003480  0.034158 -0.010562  0.016269 -0.013977 -0.025022
2    0.034271 -0.089845  0.179669  0.164192 -0.028484  0.053993  0.097806
3    0.065695 -0.054165  0.052648  0.171342  0.012820  0.050907  0.050346
4    0.019685 -0.172411 -0.007750  0.069830 -0.130615  0.063996  0.108494
..        …        …        …        …        …        …
995  0.072839 -0.033914  0.023246  0.114870 -0.019730  0.026667  0.045837
996 -0.000271 -0.071894  0.145556 -0.021958  0.089256  0.081605 -0.048531
997 -0.061372 -0.059628  0.092630 -0.027111 -0.049593  0.052719  0.145747
998 -0.056212 -0.055255  0.004168  0.013530  0.070374  0.104577 -0.007666
999 -0.092469 -0.015937 -0.035896 -0.026836 -0.033983 -0.049760  0.169589

           X7        X8        X9       X10       X11       X12           X13
0   -0.016025  0.024374 -0.042007 -0.073815 -0.006026 -0.077929  no_efectores
1   -0.043154 -0.002770 -0.062808 -0.010051  0.013118  0.019045  no_efectores
2    0.011149  0.057423  0.096364  0.103880  0.071016  0.064433  no_efectores
3   -0.000056 -0.090361  0.011262  0.081510 -0.097293 -0.048842  no_efectores
4   -0.068257 -0.054625  0.010700  0.015132  0.067584 -0.044183  no_efectores
..        …        …        …        …        …        …
995  0.022863 -0.015005  0.047525 -0.028106 -0.050720 -0.087528  no_efectores
996  0.024978  0.190676  0.038631  0.186926  0.141108 -0.114943  no_efectores
997 -0.031989  0.061052 -0.025347 -0.027390  0.035343  0.035416  no_efectores
998 -0.034841  0.005925 -0.006677 -0.000882 -0.084647  0.016582  no_efectores
999 -0.093887 -0.044394 -0.010495  0.022586 -0.018815  0.126722  no_efectores

[907 rows x 14 columns]
```

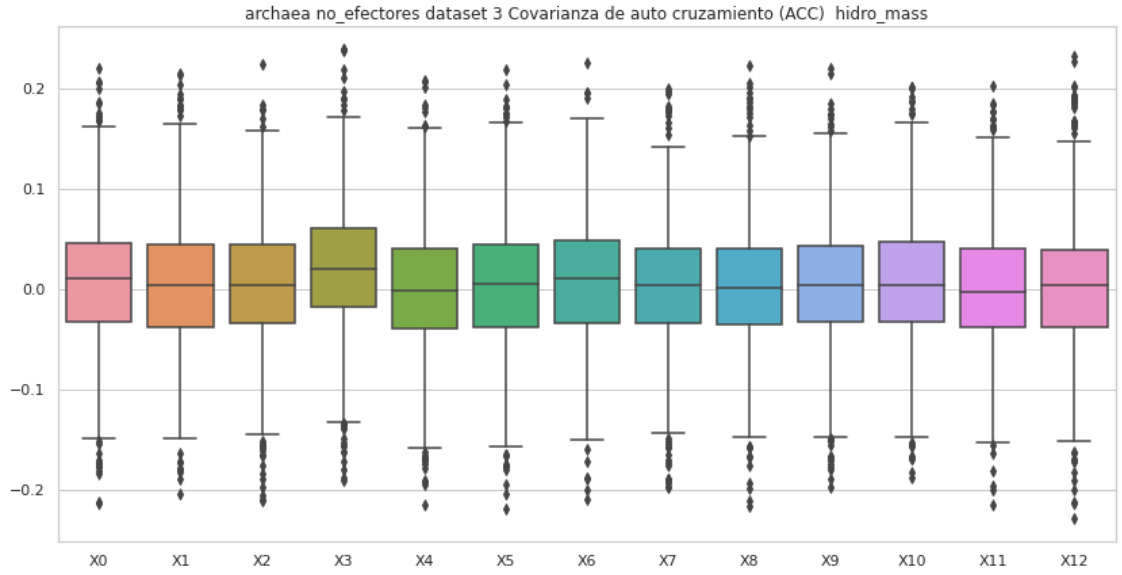
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores archaea dataset 3, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  907.000000  907.000000  907.000000  907.000000  907.000000  907.000000
mean     0.007799    0.002767    0.002884    0.019544   -0.000139    0.003631
std      0.066489    0.066292    0.063185    0.066207    0.063729    0.066191
min     -0.213859   -0.203471   -0.211184   -0.190966   -0.214963   -0.218877
25%     -0.032213   -0.037597   -0.033607   -0.017288   -0.039246   -0.037739
50%      0.010438    0.003480    0.003519    0.020084   -0.001783    0.005182
75%      0.045655    0.044297    0.044821    0.060476    0.040867    0.044233
max      0.220124    0.214445    0.223815    0.239075    0.207505    0.219458

                X6          X7          X8          X9         X10         X11  \
count  907.000000  907.000000  907.000000  907.000000  907.000000  907.000000
mean     0.008055    0.002137    0.002801    0.003802    0.006018    0.000555
std      0.063151    0.062246    0.062832    0.062414    0.062458    0.061323
```

min	-0.209539	-0.198022	-0.216098	-0.197069	-0.188202	-0.215448
25%	-0.033493	-0.034351	-0.034945	-0.032858	-0.032465	-0.037379
50%	0.010541	0.003803	0.000934	0.003550	0.004620	-0.002648
75%	0.048237	0.039994	0.040217	0.043654	0.047095	0.040406
max	0.225135	0.199311	0.222862	0.219679	0.201344	0.202166

	X12
count	907.000000
mean	0.003244
std	0.065310
min	-0.228479
25%	-0.038003
50%	0.004148
75%	0.039063
max	0.232621





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))
```

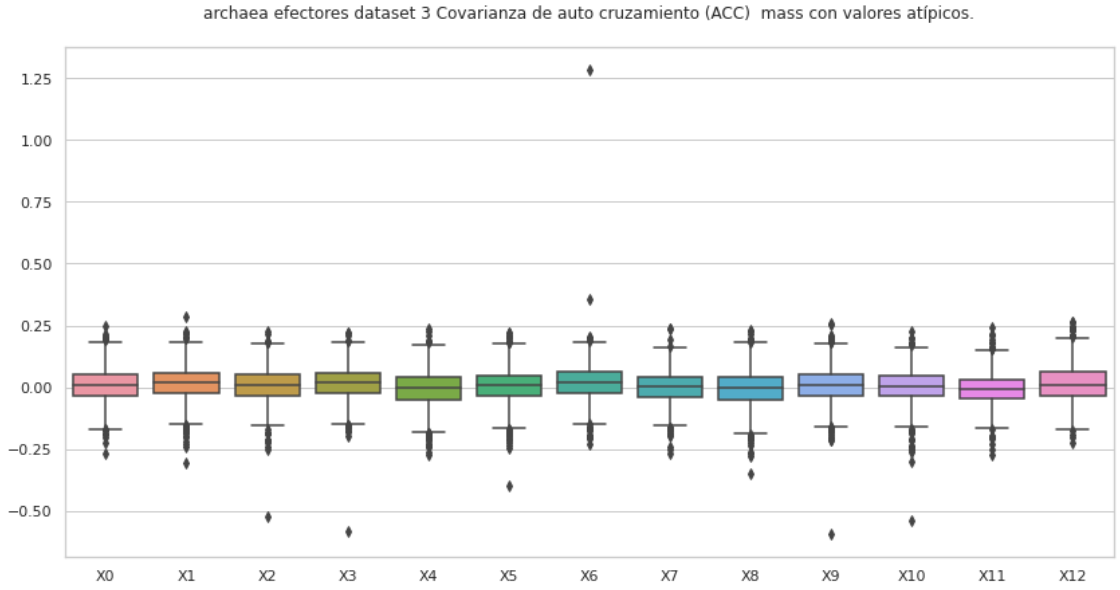
```
efectores
Covarianza de auto cruzamiento (ACC)  mass efectores archaea dataset 3, con
valores atípicos.
Valores del documento csv.

          X0        X1        X2        X3        X4        X5        X6  \
0    0.114843 -0.196889 -0.044435  0.098150  0.006701 -0.062600 -0.081355
1   -0.186083 -0.092446  0.045695  0.029906  0.234190 -0.396699  0.088197
2    0.022518 -0.049605  0.094243 -0.005566 -0.089440 -0.011582  0.100343
3   -0.039257  0.005039  0.116611  0.039954  0.034853  0.073094  0.005434
4   -0.073199 -0.016393 -0.076531  0.093333 -0.091301  0.066326 -0.020607
..        …         …         …         …         …         …         …
995  0.074901  0.107089 -0.062123  0.114744 -0.103667  0.004536  0.102645
996  0.009365  0.081069 -0.005977  0.048423  0.062020  0.081839  0.102280
997  0.180901  0.146636  0.216451  0.151302  0.227469  0.220173  0.163990
998  0.050236 -0.080113  0.009945  0.010906 -0.001176 -0.102179 -0.013181
999 -0.128161 -0.017222  0.145927 -0.078703 -0.033723 -0.051159 -0.013460

          X7        X8        X9       X10       X11       X12         X13
0    0.029848  0.231728 -0.007268 -0.088224 -0.061602  0.108636  efectores
1    0.028338 -0.019930  0.127342 -0.082932  0.241337  0.091652  efectores
2   -0.067346  0.021595 -0.009946 -0.062915 -0.005177  0.027914  efectores
3    0.062402 -0.082918  0.063462 -0.038027  0.125903  0.088533  efectores
4   -0.020088 -0.018200  0.039638 -0.112889 -0.010651  0.077114  efectores
..        …         …         …         …         …         …         …
995 -0.070860 -0.151738 -0.133297 -0.031298 -0.047527  0.054482  efectores
996  0.011582  0.063357 -0.003469  0.012590  0.054926  0.016833  efectores
997  0.141502  0.037710  0.124105  0.085191 -0.013688 -0.057822  efectores
998  0.033659 -0.200928 -0.007268  0.051083 -0.017091  0.031338  efectores
999  0.074499 -0.022332 -0.108072  0.083329 -0.089329 -0.203880  efectores

[1000 rows x 14 columns]


Covarianza de auto cruzamiento (ACC)  mass efectores archaea dataset 3, con
valores atípicos.
Estadísticas.

                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.008453     0.018032     0.008607     0.015441    -0.007213
```

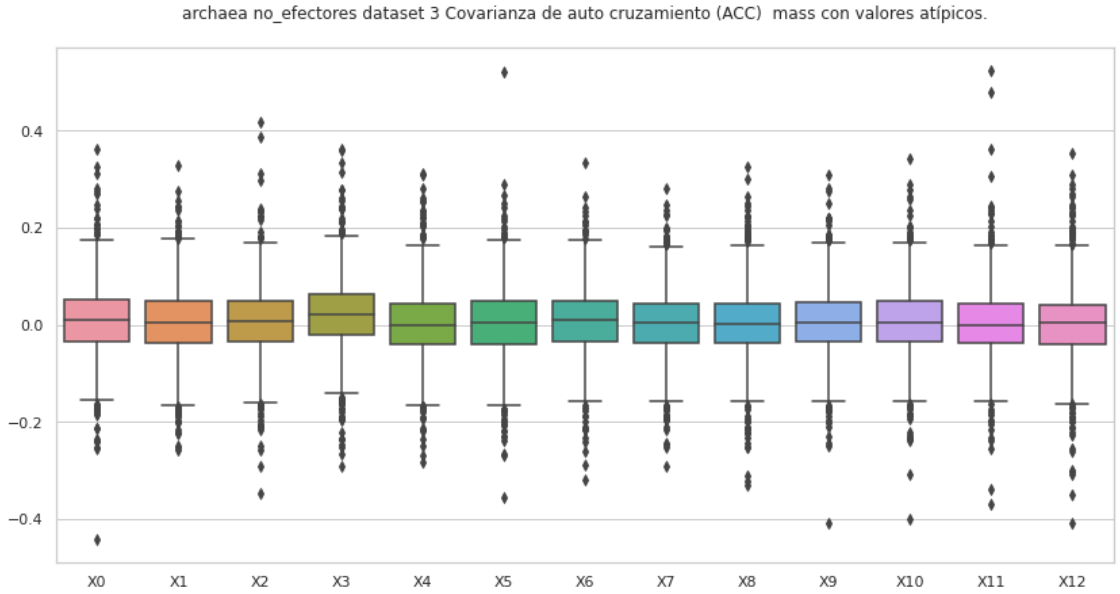
```
std       0.069059     0.070693     0.069076     0.069254     0.071354
min      -0.268809    -0.303772    -0.524088    -0.579639    -0.272140
25%      -0.035686    -0.022661    -0.033850    -0.025328    -0.048729
50%       0.006395     0.019816     0.008259     0.019759    -0.003279
75%       0.053152     0.059877     0.051540     0.057825     0.040378
max       0.245221     0.285015     0.226452     0.222028     0.234190

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.003959     0.020215     0.000804    -0.003513     0.009182
std       0.071357     0.077691     0.064536     0.074582     0.073594
min      -0.396699    -0.229729    -0.266405    -0.349572    -0.592968
25%      -0.035785    -0.023653    -0.038767    -0.048842    -0.032120
50%       0.007250     0.020073     0.003916    -0.004453     0.007127
75%       0.049705     0.060625     0.041062     0.043789     0.053259
max       0.221366     1.283083     0.238644     0.231728     0.256232

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.002246    -0.006680     0.014876
std       0.069340     0.064274     0.071045
min      -0.538224    -0.270956    -0.221888
25%      -0.036628    -0.046745    -0.032681
50%       0.003891    -0.004788     0.009454
75%       0.045174     0.032971     0.061090
max       0.228148     0.241337     0.266121



no_efectores
Covarianza de auto cruzamiento (ACC)  mass no_efectores archaea dataset 3, con
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.093575 -0.054025  0.111097 -0.029496 -0.002633  0.071206 -0.099662
1    0.053833  0.003480  0.034158 -0.010562  0.016269 -0.013977 -0.025022
2    0.034271 -0.089845  0.179669  0.164192 -0.028484  0.053993  0.097806
3    0.065695 -0.054165  0.052648  0.171342  0.012820  0.050907  0.050346
4    0.019685 -0.172411 -0.007750  0.069830 -0.130615  0.063996  0.108494
..        …        …         …         …         …         …         …
995  0.072839 -0.033914  0.023246  0.114870 -0.019730  0.026667  0.045837
996 -0.000271 -0.071894  0.145556 -0.021958  0.089256  0.081605 -0.048531
997 -0.061372 -0.059628  0.092630 -0.027111 -0.049593  0.052719  0.145747
998 -0.056212 -0.055255  0.004168  0.013530  0.070374  0.104577 -0.007666
999 -0.092469 -0.015937 -0.035896 -0.026836 -0.033983 -0.049760  0.169589

           X7        X8        X9       X10       X11       X12           X13
```

```
0   -0.016025  0.024374 -0.042007 -0.073815 -0.006026 -0.077929  no_efectores
1   -0.043154 -0.002770 -0.062808 -0.010051  0.013118  0.019045  no_efectores
2    0.011149  0.057423  0.096364  0.103880  0.071016  0.064433  no_efectores
3   -0.000056 -0.090361  0.011262  0.081510 -0.097293 -0.048842  no_efectores
4   -0.068257 -0.054625  0.010700  0.015132  0.067584 -0.044183  no_efectores
..        …       …       …       …       …       …           …
995  0.022863 -0.015005  0.047525 -0.028106 -0.050720 -0.087528  no_efectores
996  0.024978  0.190676  0.038631  0.186926  0.141108 -0.114943  no_efectores
997 -0.031989  0.061052 -0.025347 -0.027390  0.035343  0.035416  no_efectores
998 -0.034841  0.005925 -0.006677 -0.000882 -0.084647  0.016582  no_efectores
999 -0.093887 -0.044394 -0.010495  0.022586 -0.018815  0.126722  no_efectores

[1000 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) mass no_efectores archaea dataset 3, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.009299     0.003970     0.005921     0.019354     0.000940
std       0.076820     0.074941     0.073639     0.078057     0.073723
min      -0.442782    -0.258500    -0.348365    -0.293199    -0.283444
25%      -0.033877    -0.037957    -0.034076    -0.021005    -0.041411
50%       0.010987     0.003696     0.006147     0.020079    -0.002081
75%       0.050632     0.048405     0.050186     0.063080     0.042955
max       0.361488     0.328654     0.418595     0.360585     0.311632

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.003018     0.009038     0.001611     0.002875     0.005176
std       0.076944     0.073203     0.070696     0.074537     0.071654
min      -0.357161    -0.318640    -0.292956    -0.330434    -0.408843
25%      -0.039469    -0.034663    -0.037301    -0.038293    -0.034544
50%       0.004567     0.010806     0.003335     0.000871     0.004631
75%       0.047903     0.050226     0.043947     0.044489     0.047440
max       0.522532     0.333901     0.280582     0.325811     0.307657

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.004356     0.002544     0.002591
std       0.073132     0.074399     0.079116
min      -0.399831    -0.371200    -0.409532
25%      -0.035700    -0.037726    -0.040449
50%       0.003365    -0.001121     0.003871
75%       0.047850     0.043578     0.041648
max       0.343566     0.523727     0.353039
```


archaea efectores dataset 3 Covarianza de auto cruzamiento (ACC) mass con valores atípicos.
1.25
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12


archaea no_efectores dataset 3 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 3, sin
valores atípicos.
Valores del documento csv.
```

	X0	X1	X2	X3	X4	X5	X6	\
2	0.022518	-0.049605	0.094243	-0.005566	-0.089440	-0.011582	0.100343	
3	-0.039257	0.005039	0.116611	0.039954	0.034853	0.073094	0.005434	
4	-0.073199	-0.016393	-0.076531	0.093333	-0.091301	0.066326	-0.020607	
5	-0.099857	-0.088624	0.111897	-0.007385	-0.135371	0.058074	0.157932	
6	-0.006381	-0.080682	-0.027686	-0.023641	-0.098453	-0.010610	0.104092	
..	…	…	…	…	…	…	…	
992	0.014630	-0.003432	-0.060542	0.125027	-0.011092	0.020385	-0.019435	
993	0.063303	-0.020508	0.042068	-0.038679	-0.014255	0.002170	-0.017359	
995	0.074901	0.107089	-0.062123	0.114744	-0.103667	0.004536	0.102645	
996	0.009365	0.081069	-0.005977	0.048423	0.062020	0.081839	0.102280	
998	0.050236	-0.080113	0.009945	0.010906	-0.001176	-0.102179	-0.013181	

	X7	X8	X9	X10	X11	X12	X13
2	-0.067346	0.021595	-0.009946	-0.062915	-0.005177	0.027914	efectores
3	0.062402	-0.082918	0.063462	-0.038027	0.125903	0.088533	efectores
4	-0.020088	-0.018200	0.039638	-0.112889	-0.010651	0.077114	efectores
5	0.014303	-0.093978	-0.016411	0.068438	-0.045543	-0.080704	efectores
6	-0.015475	0.008513	0.008183	0.038664	-0.101622	-0.008161	efectores
..	…	…	…	…	…	…	…
992	0.083743	-0.018752	0.085996	-0.078040	-0.082516	-0.089319	efectores
993	0.064970	-0.026873	-0.015315	0.022068	0.032185	-0.025146	efectores
995	-0.070860	-0.151738	-0.133297	-0.031298	-0.047527	0.054482	efectores
996	0.011582	0.063357	-0.003469	0.012590	0.054926	0.016833	efectores
998	0.033659	-0.200928	-0.007268	0.051083	-0.017091	0.031338	efectores

```
[927 rows x 14 columns]
```

```
Covarianza de auto cruzamiento (ACC)  mass efectores archaea dataset 3, sin
valores atípicos.
Estadísticas.
```

	X0	X1	X2	X3	X4	X5	\
count	927.000000	927.000000	927.000000	927.000000	927.000000	927.000000	
mean	0.007808	0.018783	0.012679	0.015882	-0.006933	0.005811	
std	0.064957	0.065800	0.062999	0.065255	0.066397	0.066018	
min	-0.189932	-0.188604	-0.187324	-0.182319	-0.210908	-0.205896	
25%	-0.034552	-0.022251	-0.029180	-0.025341	-0.047098	-0.034240	
50%	0.005725	0.019355	0.011147	0.019943	-0.003276	0.007674	
75%	0.049967	0.058766	0.054442	0.057136	0.038367	0.049003	

```
max      0.204771    0.227640    0.187346    0.188868    0.183584    0.207431

               X6          X7          X8          X9         X10         X11  \
count  927.000000  927.000000  927.000000  927.000000  927.000000  927.000000
mean     0.021174    0.000614   -0.003400    0.009615    0.004647   -0.005248
std      0.062847    0.060053    0.068890    0.066837    0.061707    0.059573
min     -0.195450   -0.191421   -0.221686   -0.189537   -0.181470   -0.198226
25%     -0.017716   -0.036479   -0.047110   -0.030890   -0.035008   -0.043857
50%      0.021329    0.003635   -0.004479    0.007046    0.003931   -0.004235
75%      0.061059    0.040014    0.042025    0.052199    0.043915    0.032886
max      0.198039    0.159603    0.215797    0.210233    0.200565    0.182102

              X12
count  927.000000
mean     0.015298
std      0.066225
min     -0.176796
25%     -0.030819
50%      0.010482
75%      0.060232
max      0.211342
```

Covarianza de auto cruzamiento (ACC) mass no_efectores archaea dataset 3, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0   -0.093575 -0.054025  0.111097 -0.029496 -0.002633  0.071206 -0.099662
1    0.053833  0.003480  0.034158 -0.010562  0.016269 -0.013977 -0.025022
2    0.034271 -0.089845  0.179669  0.164192 -0.028484  0.053993  0.097806
3    0.065695 -0.054165  0.052648  0.171342  0.012820  0.050907  0.050346
4    0.019685 -0.172411 -0.007750  0.069830 -0.130615  0.063996  0.108494
..        …        …        …        …        …        …        …
995  0.072839 -0.033914  0.023246  0.114870 -0.019730  0.026667  0.045837
996 -0.000271 -0.071894  0.145556 -0.021958  0.089256  0.081605 -0.048531
997 -0.061372 -0.059628  0.092630 -0.027111 -0.049593  0.052719  0.145747
998 -0.056212 -0.055255  0.004168  0.013530  0.070374  0.104577 -0.007666
999 -0.092469 -0.015937 -0.035896 -0.026836 -0.033983 -0.049760  0.169589

           X7        X8        X9       X10       X11       X12           X13
0   -0.016025  0.024374 -0.042007 -0.073815 -0.006026 -0.077929  no_efectores
1   -0.043154 -0.002770 -0.062808 -0.010051  0.013118  0.019045  no_efectores
2    0.011149  0.057423  0.096364  0.103880  0.071016  0.064433  no_efectores
3   -0.000056 -0.090361  0.011262  0.081510 -0.097293 -0.048842  no_efectores
4   -0.068257 -0.054625  0.010700  0.015132  0.067584 -0.044183  no_efectores
..        …        …        …        …        …        …            …
```

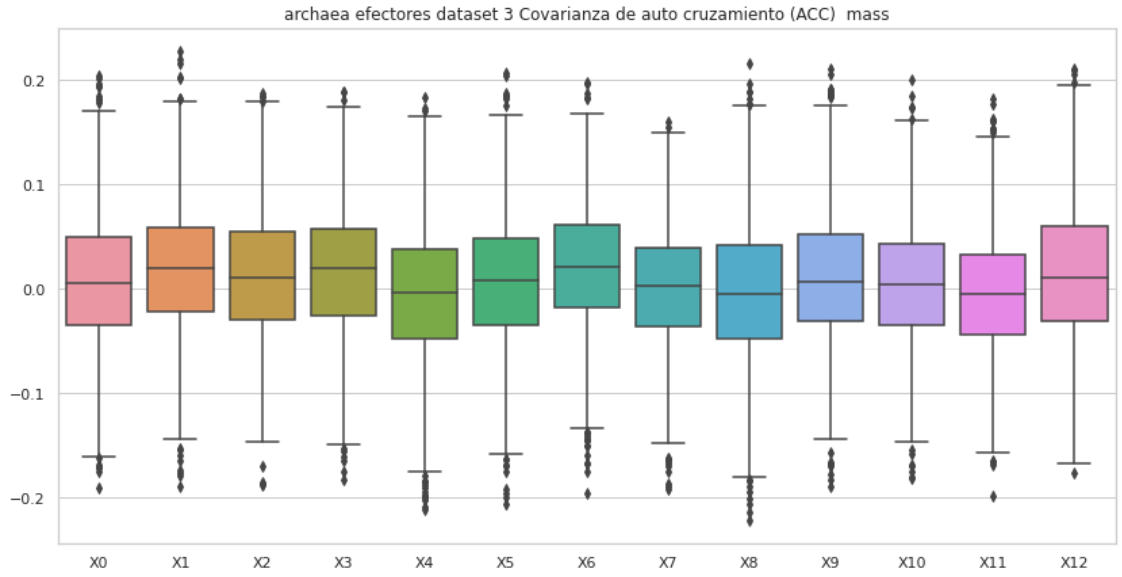
```
995  0.022863 -0.015005  0.047525 -0.028106 -0.050720 -0.087528  no_efectores
996  0.024978  0.190676  0.038631  0.186926  0.141108 -0.114943  no_efectores
997 -0.031989  0.061052 -0.025347 -0.027390  0.035343  0.035416  no_efectores
998 -0.034841  0.005925 -0.006677 -0.000882 -0.084647  0.016582  no_efectores
999 -0.093887 -0.044394 -0.010495  0.022586 -0.018815  0.126722  no_efectores

[907 rows x 14 columns]
```

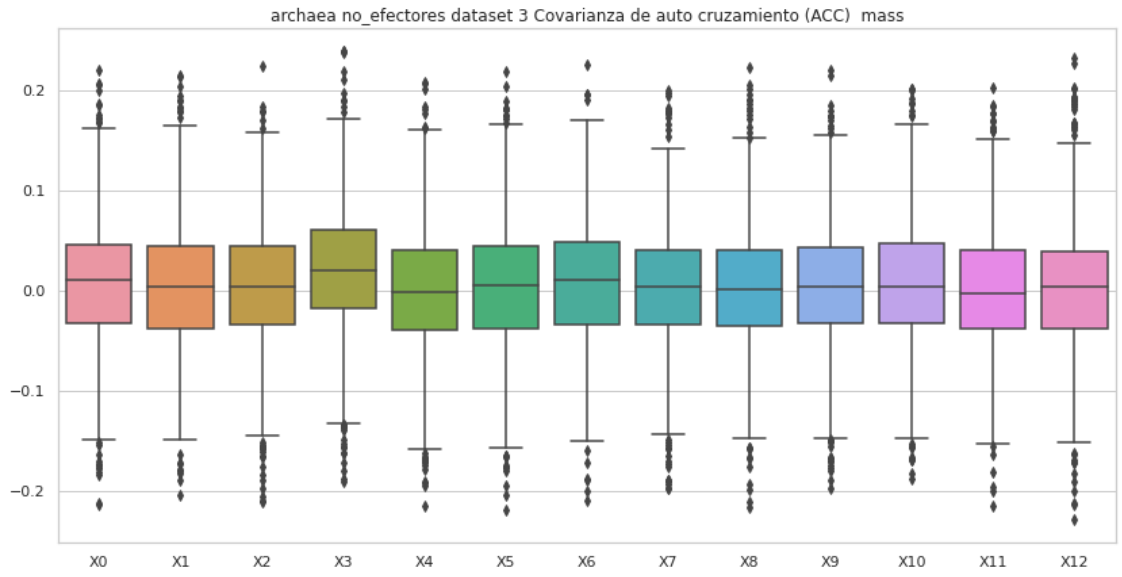
Covarianza de auto cruzamiento (ACC) mass no_efectores archaea dataset 3, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	907.000000	907.000000	907.000000	907.000000	907.000000	907.000000	
mean	0.007799	0.002767	0.002884	0.019544	-0.000139	0.003631	
std	0.066489	0.066292	0.063185	0.066207	0.063729	0.066191	
min	-0.213859	-0.203471	-0.211184	-0.190966	-0.214963	-0.218877	
25%	-0.032213	-0.037597	-0.033607	-0.017288	-0.039246	-0.037739	
50%	0.010438	0.003480	0.003519	0.020084	-0.001783	0.005182	
75%	0.045655	0.044297	0.044821	0.060476	0.040867	0.044233	
max	0.220124	0.214445	0.223815	0.239075	0.207505	0.219458	

	X6	X7	X8	X9	X10	X11	\
count	907.000000	907.000000	907.000000	907.000000	907.000000	907.000000	
mean	0.008055	0.002137	0.002801	0.003802	0.006018	0.000555	
std	0.063151	0.062246	0.062832	0.062414	0.062458	0.061323	
min	-0.209539	-0.198022	-0.216098	-0.197069	-0.188202	-0.215448	
25%	-0.033493	-0.034351	-0.034945	-0.032858	-0.032465	-0.037379	
50%	0.010541	0.003803	0.000934	0.003550	0.004620	-0.002648	
75%	0.048237	0.039994	0.040217	0.043654	0.047095	0.040406	
max	0.225135	0.199311	0.222862	0.219679	0.201344	0.202166	

	X12
count	907.000000
mean	0.003244
std	0.065310
min	-0.228479
25%	-0.038003
50%	0.004148
75%	0.039063
max	0.232621





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 3, con
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.125123 -0.014344 -0.121566 -0.063276 -0.015017  0.027330 -0.023358
1    0.061213  0.133562 -0.029225 -0.025185  0.051441 -0.001130 -0.020975
2    0.096108  0.063699 -0.048484  0.111859  0.030988  0.017427 -0.025002
3   -0.047206 -0.001800 -0.115818 -0.046125  0.010053  0.030834 -0.017295
4   -0.140211 -0.243482 -0.058864  0.085605  0.070412 -0.187966  0.133234
..        …         …         …         …         …         …         …
995 -0.011090 -0.090030 -0.094758 -0.034044 -0.094216  0.121375  0.127612
996 -0.051072 -0.077653 -0.016299 -0.009708 -0.042362 -0.030508 -0.028339
997 -0.027315  0.028542  0.094072 -0.011935  0.110988 -0.000842  0.083411
998  0.011200 -0.221969  0.005880  0.197378 -0.130622 -0.098152  0.032837
999  0.264934  0.028277 -0.064362 -0.077181 -0.269540 -0.113227 -0.070963

           X7        X8        X9       X10       X11       X12        X13
0   -0.008412 -0.099988 -0.062623 -0.092080 -0.041259 -0.163131  efectores
1   -0.116062 -0.059122 -0.043897 -0.070713  0.031399  0.026063  efectores
2    0.125162  0.131961  0.205128  0.116189  0.142579 -0.028937  efectores
```

```
3   -0.136027 -0.145415 -0.011180  0.029156  0.167767 -0.044300  efectores
4    0.195612 -0.170576 -0.137727 -0.028606  0.154358  0.104064  efectores
..        …       …        …        …        …        …        …
995  0.066348  0.095534 -0.138501 -0.022457 -0.044774  0.124190  efectores
996  0.039369  0.021842 -0.011637  0.048818 -0.008509 -0.099623  efectores
997  0.042841  0.022715  0.001825 -0.004557  0.041978  0.040319  efectores
998 -0.073056 -0.031895 -0.035525  0.123537 -0.225638 -0.023945  efectores
999 -0.050480 -0.141097 -0.052976 -0.011062 -0.102427 -0.044999  efectores

[1000 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro efectores archaea dataset 3, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.027591    -0.025862     0.043195     0.033941    -0.014586
std       0.089117     0.098904     0.086646     0.090532     0.094107
min      -0.292003    -0.385520    -0.366295    -0.364802    -0.389944
25%      -0.021437    -0.089368    -0.014477    -0.016688    -0.073393
50%       0.028017    -0.015825     0.035656     0.036439    -0.008332
75%       0.079871     0.045176     0.096277     0.083828     0.047261
max       0.337942     0.354982     0.471753     0.334931     0.331652

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean     -0.010010     0.030786     0.024948    -0.002084    -0.005599
std       0.084166     0.092023     0.082298     0.091566     0.088290
min      -0.471736    -0.328167    -0.334885    -0.376935    -0.299527
25%      -0.063080    -0.026372    -0.023485    -0.046396    -0.053294
50%      -0.006516     0.023989     0.022730     0.004606    -0.002810
75%       0.046192     0.072043     0.072693     0.043483     0.041804
max       0.294412     0.653119     0.438358     0.319849     0.298862

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.016917     0.004473    -0.008250
std       0.081909     0.080864     0.081551
min      -0.257953    -0.306746    -0.369458
25%      -0.034866    -0.046035    -0.054062
50%       0.005817     0.000226    -0.007043
75%       0.065395     0.049445     0.035456
max       0.362145     0.347520     0.303443
```

no_efectores
Covarianza de auto cruzamiento (ACC) hidro no_efectores archaea dataset 3, con valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.062844	-0.157762	-0.022051	0.023256	0.017547	0.003971	0.109516	
1	-0.098285	-0.125094	0.028901	0.059199	-0.025950	-0.046054	0.026281	
2	0.110084	-0.018068	0.074028	0.007895	-0.011371	-0.062626	-0.073483	
3	0.006178	-0.131851	-0.041097	0.153451	0.029662	-0.116856	-0.083441	
4	-0.085484	-0.196705	-0.012594	0.006111	-0.068265	-0.033082	0.168117	
..	…	…	…	…	…	…	…	
995	0.041009	-0.058650	0.079014	0.091924	-0.046468	0.003196	0.076182	
996	-0.005843	0.024562	0.139436	0.099389	0.008920	0.041891	-0.033428	
997	0.009603	-0.105520	-0.097722	0.049098	-0.174892	0.025683	-0.117053	
998	0.003387	-0.054266	0.063457	0.028288	-0.111363	-0.018410	-0.013991	
999	-0.040659	-0.137542	0.107670	0.091908	-0.042494	0.004412	0.048292	

	X7	X8	X9	X10	X11	X12	X13
0	0.026007	-0.139760	0.002216	0.134379	-0.045137	-0.090073	no_efectores
1	0.037233	0.005104	-0.044409	-0.025771	-0.011200	0.033514	no_efectores
2	-0.015542	0.007644	-0.110538	-0.030666	-0.026515	0.077507	no_efectores
3	0.074349	0.076380	-0.018147	0.005013	0.049484	0.001052	no_efectores
4	0.032599	-0.052029	-0.032534	0.019610	-0.086835	-0.012327	no_efectores
..	…	…	…	…	…	…	…
995	-0.039405	-0.041522	-0.015524	-0.033569	-0.027686	-0.012793	no_efectores
996	0.017183	-0.052104	-0.112436	-0.055857	0.027027	-0.045342	no_efectores
997	0.190664	-0.003628	-0.015883	-0.050053	-0.075875	-0.248767	no_efectores
998	0.024274	-0.025863	-0.047845	0.063590	-0.008495	0.050257	no_efectores
999	-0.034037	-0.087483	0.003445	-0.049921	-0.034424	0.036358	no_efectores

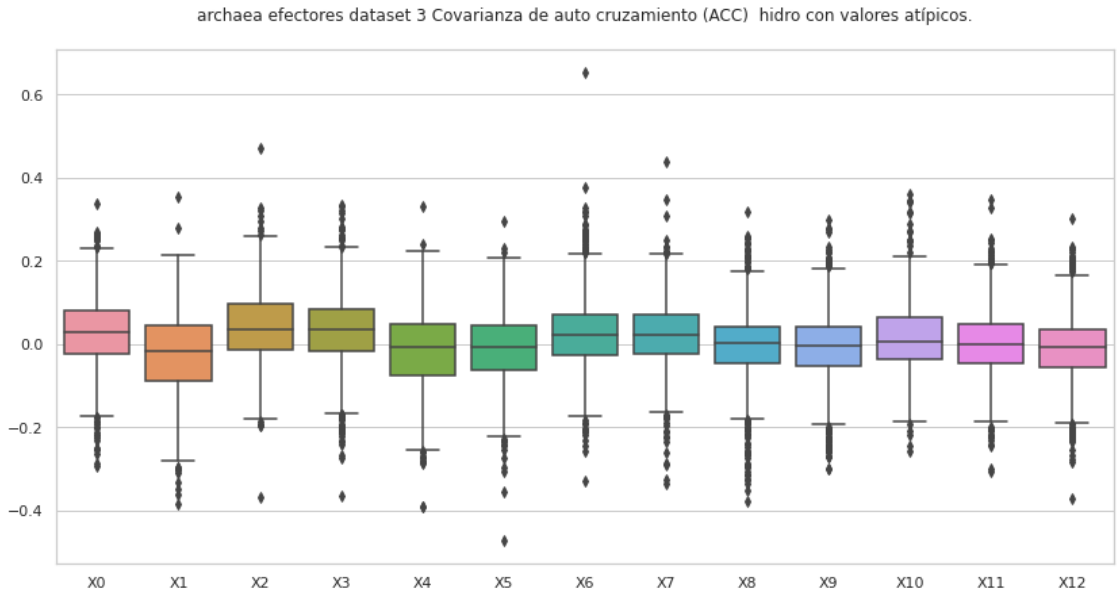
[1000 rows x 14 columns]

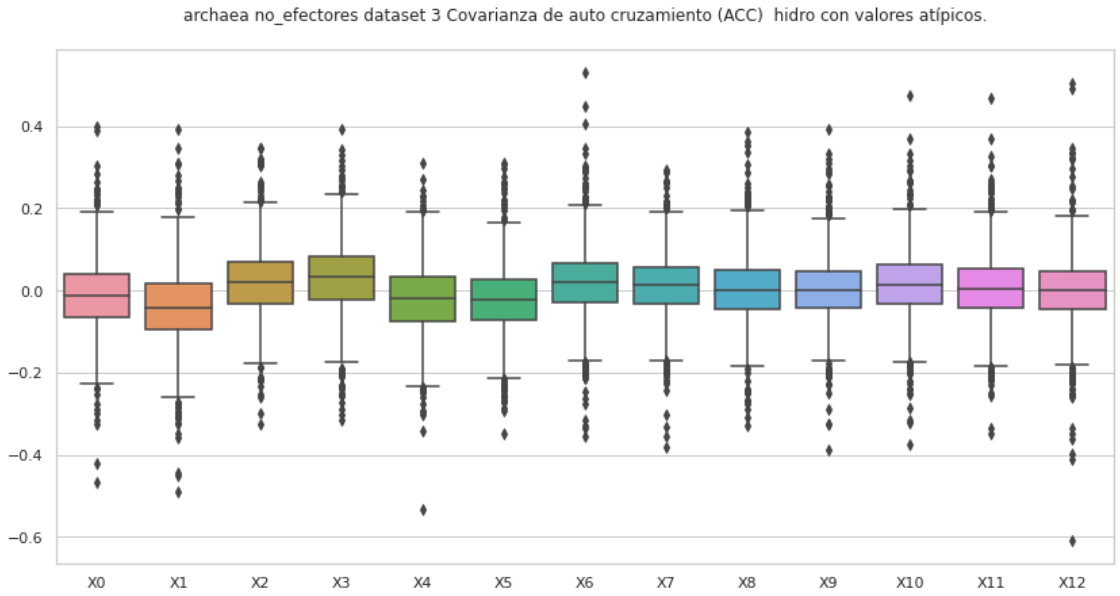
Covarianza de auto cruzamiento (ACC) hidro no_efectores archaea dataset 3, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	-0.011323	-0.037732	0.020053	0.030082	-0.020292	
std	0.090541	0.098817	0.085457	0.091266	0.088905	
min	-0.465565	-0.490861	-0.326700	-0.314169	-0.531769	
25%	-0.066824	-0.094383	-0.030994	-0.021684	-0.074439	
50%	-0.013898	-0.042991	0.018604	0.031896	-0.019737	
75%	0.041380	0.017947	0.068216	0.081182	0.032231	
max	0.397809	0.392866	0.347472	0.393305	0.310119	

	X5	X6	X7	X8	X9	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	-0.025217	0.019254	0.010675	0.004585	0.003421	
std	0.088571	0.090585	0.079636	0.086341	0.082818	
min	-0.349588	-0.353940	-0.382031	-0.328472	-0.388295	
25%	-0.070367	-0.029467	-0.034044	-0.045973	-0.040719	
50%	-0.023381	0.018976	0.014673	0.002036	0.001138	
75%	0.025455	0.065782	0.057515	0.050171	0.046649	
max	0.309605	0.530451	0.292214	0.387234	0.392169	

	X10	X11	X12
count	1000.000000	1000.000000	1000.000000
mean	0.015183	0.006004	-0.000308
std	0.084114	0.085563	0.090741
min	-0.375367	-0.348591	-0.608824
25%	-0.032979	-0.042159	-0.045268
50%	0.013135	0.004974	0.000640
75%	0.061630	0.053029	0.047509
max	0.475845	0.467440	0.504201





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 3, sin valores
atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.125123 -0.014344 -0.121566 -0.063276 -0.015017  0.027330 -0.023358
1    0.061213  0.133562 -0.029225 -0.025185  0.051441 -0.001130 -0.020975
2    0.096108  0.063699 -0.048484  0.111859  0.030988  0.017427 -0.025002
3   -0.047206 -0.001800 -0.115818 -0.046125  0.010053  0.030834 -0.017295
4   -0.140211 -0.243482 -0.058864  0.085605  0.070412 -0.187966  0.133234
..        …         …         …         …         …         …         …
995 -0.011090 -0.090030 -0.094758 -0.034044 -0.094216  0.121375  0.127612
996 -0.051072 -0.077653 -0.016299 -0.009708 -0.042362 -0.030508 -0.028339
997 -0.027315  0.028542  0.094072 -0.011935  0.110988 -0.000842  0.083411
998  0.011200 -0.221969  0.005880  0.197378 -0.130622 -0.098152  0.032837
999  0.264934  0.028277 -0.064362 -0.077181 -0.269540 -0.113227 -0.070963

           X7        X8        X9       X10       X11       X12        X13
0   -0.008412 -0.099988 -0.062623 -0.092080 -0.041259 -0.163131  efectores
1   -0.116062 -0.059122 -0.043897 -0.070713  0.031399  0.026063  efectores
2    0.125162  0.131961  0.205128  0.116189  0.142579 -0.028937  efectores
3   -0.136027 -0.145415 -0.011180  0.029156  0.167767 -0.044300  efectores
4    0.195612 -0.170576 -0.137727 -0.028606  0.154358  0.104064  efectores
..        …         …         …         …         …         …         …
995  0.066348  0.095534 -0.138501 -0.022457 -0.044774  0.124190  efectores
```

```
996  0.039369  0.021842 -0.011637  0.048818 -0.008509 -0.099623  efectores
997  0.042841  0.022715  0.001825 -0.004557  0.041978  0.040319  efectores
998 -0.073056 -0.031895 -0.035525  0.123537 -0.225638 -0.023945  efectores
999 -0.050480 -0.141097 -0.052976 -0.011062 -0.102427 -0.044999  efectores

[928 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) efectores archaea dataset 3, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	928.000000	928.000000	928.000000	928.000000	928.000000	928.000000	
mean	0.029260	-0.020453	0.040399	0.036298	-0.008773	-0.005701	
std	0.083817	0.092158	0.081119	0.083659	0.087384	0.076681	
min	-0.231661	-0.297944	-0.195874	-0.233316	-0.282070	-0.253118	
25%	-0.020642	-0.082765	-0.014690	-0.013542	-0.067852	-0.056577	
50%	0.027845	-0.011627	0.033840	0.038286	-0.004034	-0.004038	
75%	0.077743	0.047039	0.094092	0.083245	0.048471	0.046807	
max	0.269399	0.215275	0.274428	0.303100	0.240608	0.230132	

	X6	X7	X8	X9	X10	X11	\
count	928.000000	928.000000	928.000000	928.000000	928.000000	928.000000	
mean	0.027106	0.025263	0.001986	-0.004082	0.012966	0.005696	
std	0.083197	0.073341	0.078898	0.080045	0.073781	0.073384	
min	-0.230031	-0.211821	-0.273021	-0.258525	-0.219673	-0.230645	
25%	-0.026623	-0.022426	-0.043574	-0.050238	-0.034866	-0.042778	
50%	0.021704	0.022389	0.005979	-0.001573	0.002997	0.001167	
75%	0.067360	0.067739	0.040973	0.039635	0.057192	0.048632	
max	0.290735	0.232656	0.259995	0.227740	0.251335	0.232163	

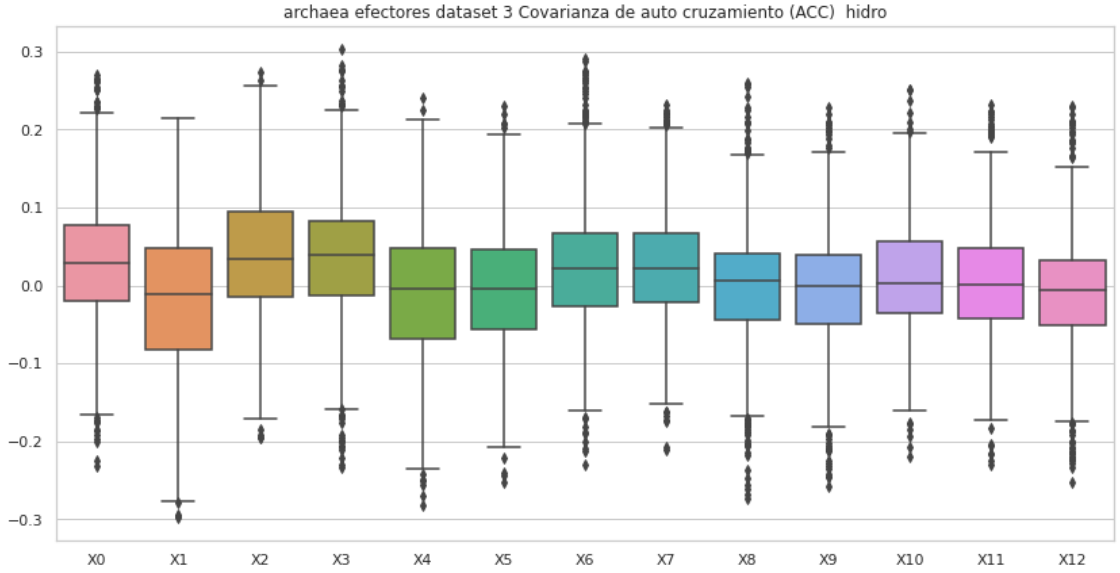
	X12
count	928.000000
mean	-0.006943
std	0.075010
min	-0.252675
25%	-0.050638
50%	-0.006374
75%	0.032785
max	0.229559

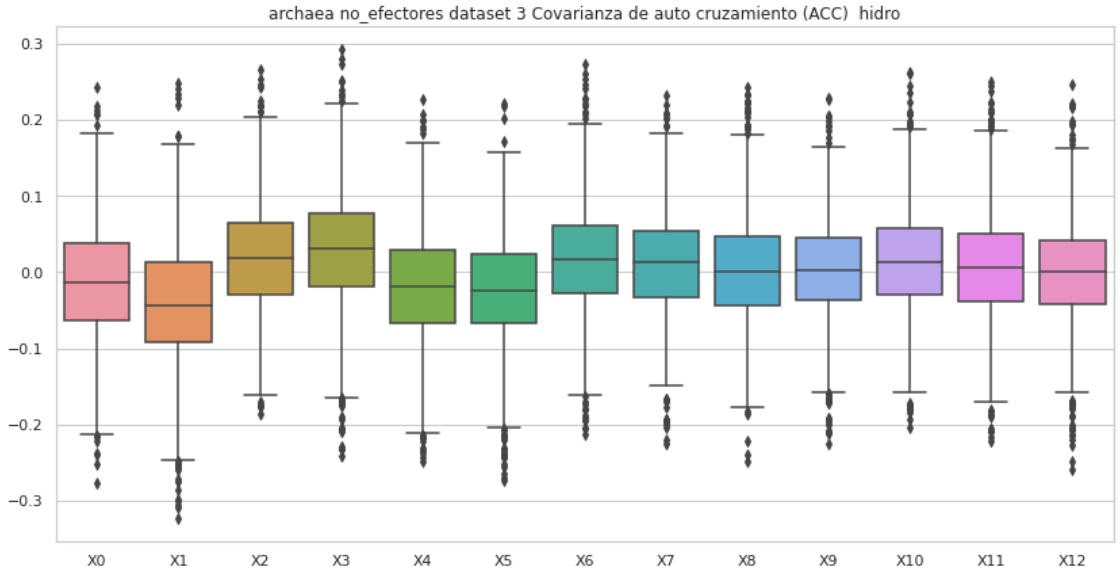
no_efectores
Covarianza de auto cruzamiento (ACC) no_efectores archaea dataset 3, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.062844 -0.157762 -0.022051  0.023256  0.017547  0.003971  0.109516
1   -0.098285 -0.125094  0.028901  0.059199 -0.025950 -0.046054  0.026281
2    0.110084 -0.018068  0.074028  0.007895 -0.011371 -0.062626 -0.073483
3    0.006178 -0.131851 -0.041097  0.153451  0.029662 -0.116856 -0.083441
4   -0.085484 -0.196705 -0.012594  0.006111 -0.068265 -0.033082  0.168117
..        …         …         …         …         …         …
995  0.041009 -0.058650  0.079014  0.091924 -0.046468  0.003196  0.076182
996 -0.005843  0.024562  0.139436  0.099389  0.008920  0.041891 -0.033428
997  0.009603 -0.105520 -0.097722  0.049098 -0.174892  0.025683 -0.117053
998  0.003387 -0.054266  0.063457  0.028288 -0.111363 -0.018410 -0.013991
999 -0.040659 -0.137542  0.107670  0.091908 -0.042494  0.004412  0.048292

           X7        X8        X9       X10       X11       X12           X13
0    0.026007 -0.139760  0.002216  0.134379 -0.045137 -0.090073  no_efectores
1    0.037233  0.005104 -0.044409 -0.025771 -0.011200  0.033514  no_efectores
2   -0.015542  0.007644 -0.110538 -0.030666 -0.026515  0.077507  no_efectores
3    0.074349  0.076380 -0.018147  0.005013  0.049484  0.001052  no_efectores
4    0.032599 -0.052029 -0.032534  0.019610 -0.086835 -0.012327  no_efectores
..        …         …         …         …         …         …           …
995 -0.039405 -0.041522 -0.015524 -0.033569 -0.027686 -0.012793  no_efectores
996  0.017183 -0.052104 -0.112436 -0.055857  0.027027 -0.045342  no_efectores
997  0.190664 -0.003628 -0.015883 -0.050053 -0.075875 -0.248767  no_efectores
998  0.024274 -0.025863 -0.047845  0.063590 -0.008495  0.050257  no_efectores
999 -0.034037 -0.087483  0.003445 -0.049921 -0.034424  0.036358  no_efectores

[905 rows x 14 columns]


Covarianza de auto cruzamiento (ACC)  no_efectores archaea dataset 3, sin
valores atípicos.
Estadísticas.

                X0          X1          X2          X3          X4          X5  \
count  905.000000  905.000000  905.000000  905.000000  905.000000  905.000000
mean    -0.012127   -0.039064    0.019284    0.028284   -0.020189   -0.027249
std      0.077999    0.084297    0.073042    0.080922    0.077196    0.077694
min     -0.277358   -0.323248   -0.186275   -0.241482   -0.248205   -0.274198
25%     -0.063143   -0.091352   -0.028784   -0.019727   -0.067563   -0.067574
50%     -0.014392   -0.043938    0.018593    0.031409   -0.018687   -0.023591
75%      0.037371    0.012654    0.065248    0.077138    0.028614    0.023837
max      0.242789    0.248598    0.265455    0.292326    0.226280    0.221487

                X6          X7          X8          X9         X10         X11  \
count  905.000000  905.000000  905.000000  905.000000  905.000000  905.000000
mean     0.016367    0.009943    0.003550    0.002783    0.015722    0.006220
std      0.074744    0.070083    0.074009    0.068846    0.069013    0.075119
```

min	-0.213029	-0.225688	-0.248700	-0.225644	-0.203149	-0.221126
25%	-0.027798	-0.032624	-0.043890	-0.036938	-0.029820	-0.039248
50%	0.017438	0.013167	0.001225	0.001770	0.013303	0.005107
75%	0.061206	0.054480	0.046309	0.044515	0.058035	0.050524
max	0.272655	0.232386	0.242020	0.229117	0.262986	0.248971

	X12
count	905.000000
mean	-0.001254
std	0.070963
min	-0.259683
25%	-0.042512
50%	-0.000103
75%	0.041513
max	0.246922




archaea no_efectores dataset 3 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