ds1_archaea_limpieza_de_datos

January 19, 2021

Limpieza de datos

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

1 Declaración de variables

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

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

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

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

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

2 Composición de aminoácidos (AAC)

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

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

    if etiq == "efectores":
        df=AAC_efec

    if etiq == "no_efectores":
        df=AAC_no_efec

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

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

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

         X0     X1     X2     X3     X4      X5     X6      X7     X8     X9  \
0    11.081  7.027  3.243  4.865  0.270   8.919  6.216   9.459  2.162  6.216
1    14.614  5.819  7.037  7.037  0.271   6.225  2.436  19.215  0.677  1.759
2    15.709  4.981  1.149  1.149  0.000   3.831  1.149   8.812  1.149  4.981
3     8.922  5.948  1.859  5.948  1.115   7.807  3.717   9.665  0.743  5.948
4    10.405  7.514  2.312  2.312  0.578  15.029  1.734   8.092  1.734  6.936
..      …      …      …      …      …       …      …       …      …      …
495   7.216  7.216  4.124  4.124  0.000   5.155  3.093   5.155  2.062  8.247
496   6.637  6.195  5.310  5.752  0.885   7.522  0.442   3.982  0.885  8.850
497  17.598  3.073  1.117  4.749  0.000   1.955  1.676  12.011  0.838  3.631
498   6.047  6.047  2.791  6.047  0.000   3.721  1.860   9.302  4.186  4.186
499   9.810  7.595  2.215  6.013  0.000   6.646  7.911   6.013  0.949  6.013
```

```
     …     X11    X12    X13    X14    X15    X16    X17    X18     X19  \
0    …   3.784  2.162  2.162  9.459  3.243  4.595  0.811  3.514   4.865
1    …   1.083  0.541  2.977  2.571  6.495  8.525  0.677  1.083   7.307
2    …   0.383  1.533  3.831  5.364  4.981  8.046  1.149  4.598  14.176
3    …   4.089  2.974  2.974  2.602  5.948  7.063  0.743  3.346  10.781
4    …   6.358  4.046  3.468  1.156  4.624  2.890  0.578  4.624   3.468
..   …     …    …    …    …    …    …    …    …     …
495  …  10.309  3.093  5.155  3.093  9.278  6.186  3.093  2.062   5.155
496  …  10.177  1.770  2.655  5.310  5.752  4.425  0.885  6.637   3.982
497  …   0.838  1.397  4.469  3.073  5.028  6.704  1.676  1.676  12.849
498  …   1.395  1.860  5.116  3.721  3.721  6.512  2.326  4.651  13.488
499  …   1.899  0.949  2.848  5.063  9.177  6.646  2.215  2.215   4.747

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

[500 rows x 21 columns]
```

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

```
                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     9.299878    6.213176    2.685162    5.794782    0.680588    7.285994
std      4.586333    3.027055    2.022717    2.727923    0.935679    4.088747
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      5.948500    4.098000    1.218000    3.755250    0.000000    3.880750
50%      8.786000    6.074000    2.268500    5.414000    0.391500    7.308000
75%     12.060000    8.111500    3.797000    7.502000    0.977000    9.891250
max     23.699000   20.732000   11.504000   15.385000    6.857000   21.831000

                X6          X7          X8          X9         X10         X11  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     2.382412    7.408346    1.789684    5.938128   11.476852    4.561968
std      1.671821    2.875955    1.185776    3.633332    3.557328    4.399490
```

min	0.000000	0.000000	0.000000	0.000000	0.957000	0.000000
25%	1.282000	5.333000	0.911250	3.200000	8.943000	1.106750
50%	2.013500	7.208000	1.626000	5.178500	11.394500	2.459000
75%	3.146250	9.231000	2.578500	7.976500	13.836250	7.433750
max	10.606000	19.215000	6.422000	20.000000	23.529000	18.341000

	X12	X13	X14	X15	X16	X17 \
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	1.943614	3.994054	4.216818	5.796374	5.463444	1.184856
std	1.111649	1.944646	2.049692	2.108975	2.188451	0.973119
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	1.068500	2.680500	2.982750	4.478000	3.954500	0.556250
50%	1.675500	3.901500	3.948000	5.719000	5.456500	1.023500
75%	2.567250	5.136750	5.237500	7.117000	6.652750	1.642500
max	7.865000	11.837000	21.311000	15.652000	13.988000	4.918000

	X18	X19
count	500.000000	500.000000
mean	3.349146	8.534586
std	1.663598	3.626598
min	0.000000	1.316000
25%	2.220250	5.771000
50%	3.155000	8.059500
75%	4.134750	11.034500
max	10.050000	18.657000

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

	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9 \
0	8.994	6.860	2.896	8.232	0.000	7.317	3.811	8.079	1.067	5.793
1	14.067	4.281	0.917	1.835	2.446	3.976	0.612	9.786	0.917	6.116
2	8.000	6.667	0.000	6.667	0.000	6.667	1.333	12.000	1.333	0.000
3	9.091	5.455	1.818	7.273	9.091	7.273	1.818	10.909	0.000	1.818
4	10.227	9.773	0.909	9.091	1.818	8.864	1.818	8.182	1.591	4.091
..	…	…	…	…	…	…	…	…	…	…
495	9.402	4.274	4.274	7.692	0.855	11.966	2.564	6.838	0.000	2.564
496	9.223	3.883	2.427	14.078	1.942	10.194	2.427	10.194	2.427	4.369
497	7.182	5.525	1.381	4.420	0.000	3.039	3.039	8.564	0.829	7.735
498	6.623	9.272	4.636	9.272	0.662	7.285	2.649	7.285	2.649	5.960
499	9.091	9.091	0.000	5.195	0.000	10.390	2.597	5.195	2.597	3.896

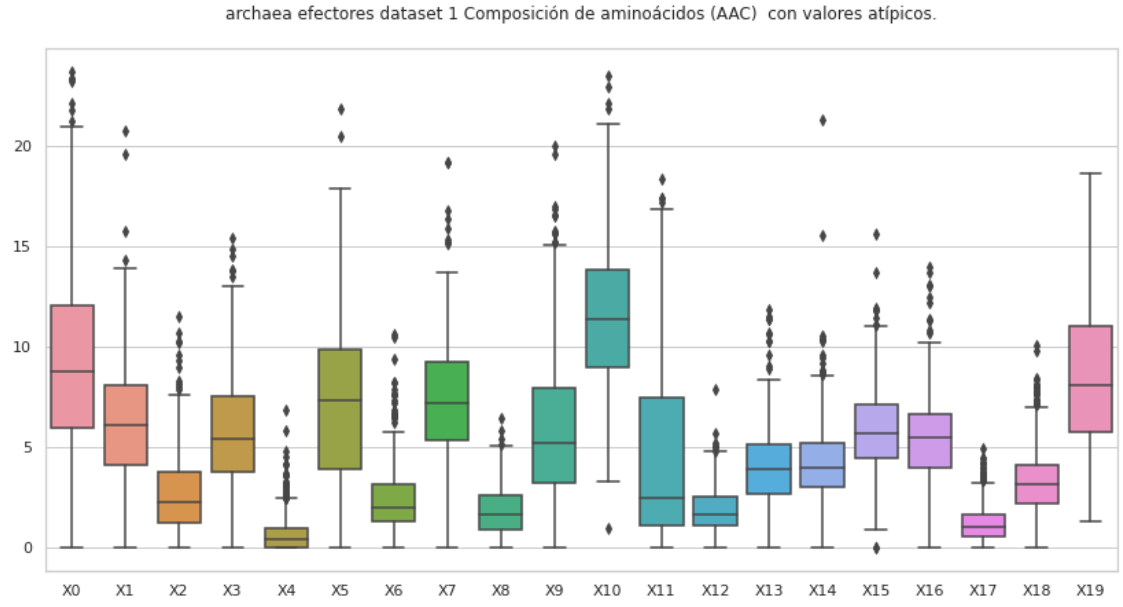
	…	X11	X12	X13	X14	X15	X16	X17	X18	X19 \
0	…	1.524	2.439	5.488	4.268	3.811	6.555	1.372	2.744	8.994

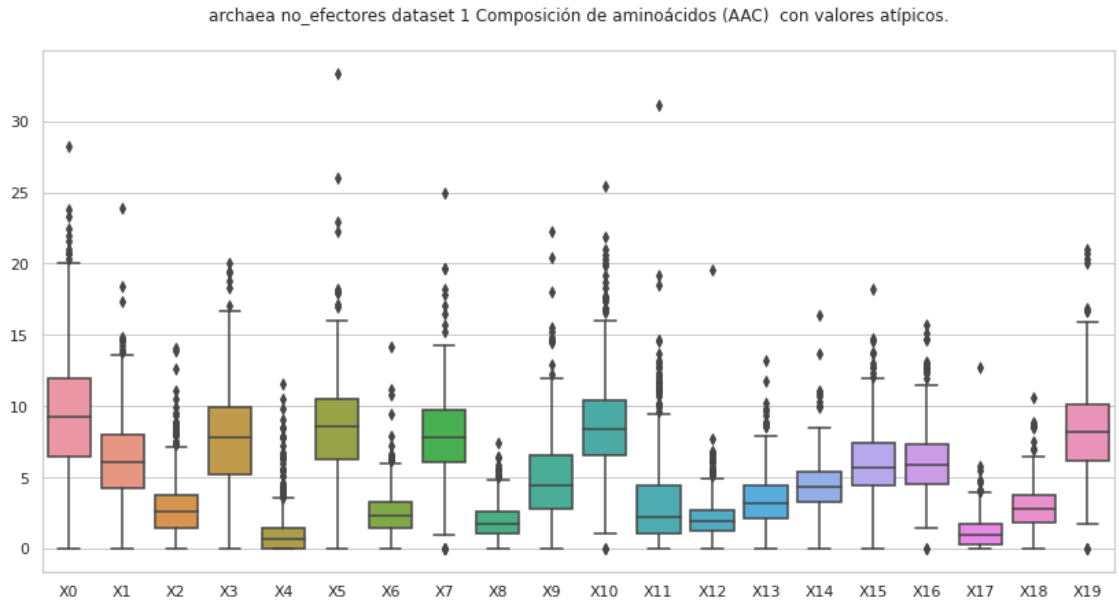
```
1    …   1.835  1.223  2.752  6.422   6.116  5.505  0.917  3.058  10.398
2    …   1.333  1.333  1.333  6.667   5.333  8.000  0.000  4.000  20.000
3    …  14.545  5.455  5.455  1.818   5.455  3.636  1.818  3.636   1.818
4    …   0.682  1.136  2.500  6.591   5.909  5.227  0.682  2.045   7.500
..   …     …    …    …    …     …    …    …    …     …
495  …   5.128  0.855  1.709  4.274   8.547  9.402  2.564  2.564   7.692
496  …   0.000  1.942  3.398  3.883   4.369  9.223  1.456  2.913   6.311
497  …   1.105  1.934  7.459  5.249   3.591  5.249  2.210  4.420  13.260
498  …   3.974  1.325  3.974  4.636   5.960  5.298  0.662  3.974   7.285
499  …   3.896  1.299  2.597  2.597  18.182  2.597  2.597  5.195   3.896

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

[500 rows x 21 columns]
```

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

```
                X0          X1          X2          X3          X4          X5  \
count   500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean      9.525232    6.314100    2.901802    7.685832    1.182536    8.433174
std       4.298959    3.102931    2.137834    3.572755    1.754967    3.799854
min       0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%       6.467000    4.279250    1.493000    5.184500    0.000000    6.250000
50%       9.256500    6.110500    2.569500    7.788000    0.643000    8.534000
75%      11.984250    8.007250    3.775000    9.924000    1.425750   10.530750
max      28.261000   23.864000   14.085000   20.000000   11.538000   33.333000

                X6          X7          X8          X9         X10         X11  \
count   500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean      2.492216    7.895222    1.867148    4.994648    8.851166    3.338816
std       1.687633    2.969510    1.274913    3.077497    3.468636    3.456091
min       0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%       1.423750    6.076750    1.042000    2.797000    6.562750    1.065500
50%       2.300500    7.759000    1.691500    4.444000    8.395500    2.186000
```

```
75%      3.264250    9.696000    2.598250    6.510750   10.417000    4.457250
max     14.151000   25.000000    7.447000   22.222000   25.424000   31.111000

              X12         X13         X14         X15         X16         X17  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     2.211642    3.409468    4.398112    6.049296    6.104368    1.149534
std      1.557007    1.933631    1.872950    2.470342    2.476377    1.148660
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      1.226000    2.105500    3.245500    4.404250    4.487750    0.295250
50%      1.912500    3.166500    4.294500    5.675500    5.907000    0.916500
75%      2.741750    4.468500    5.438500    7.409500    7.371000    1.773500
max     19.565000   13.208000   16.364000   18.182000   15.686000   12.766000

              X18         X19
count  500.000000  500.000000
mean     2.929218    8.266620
std      1.555266    3.174805
min      0.000000    0.000000
25%      1.863250    6.200750
50%      2.784500    8.167000
75%      3.783000   10.156250
max     10.638000   21.053000
```





2.1 Composición de aminoácidos (AAC), sin valores atípicos

[4]:
```
transf = "Composición de aminoácidos (AAC) "
estado = "sin valores atípicos.\n"
transf2="AAC"

out = (str(r3) + '/ds' + str(dataset) + '_' + str(transf2) + '_' +␣
 ↪str(organismo) + '.csv')
os.makedirs(str(r3), exist_ok=True)
df=""
df_out = pd.DataFrame()


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

    if etiq == "efectores":
        df=AAC_efec

    if etiq == "no_efectores":
        df=AAC_no_efec

    del df['X20']
    #Se eliminan todas las filas que tengan valores atípicos en al menos una de␣
 ↪sus columnas.
    df = (df[(np.abs(stats.zscore(df)) < 3).all(axis=1)])
```

```
    df['X20'] = etiq
    df_out = pd.concat([df_out,df])

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



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

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

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

         X0     X1     X2     X3     X4      X5     X6      X7     X8     X9  \
0    11.081  7.027  3.243  4.865  0.270   8.919  6.216   9.459  2.162  6.216
2    15.709  4.981  1.149  1.149  0.000   3.831  1.149   8.812  1.149  4.981
3     8.922  5.948  1.859  5.948  1.115   7.807  3.717   9.665  0.743  5.948
4    10.405  7.514  2.312  2.312  0.578  15.029  1.734   8.092  1.734  6.936
5    18.465  2.878  1.199  2.638  0.000   1.918  1.439  12.710  0.719  4.317
..      …      …      …      …      …       …      …       …      …      …
492  11.083  7.053  2.519  5.793  0.000   8.816  3.023   8.060  0.504  5.793
495   7.216  7.216  4.124  4.124  0.000   5.155  3.093   5.155  2.062  8.247
496   6.637  6.195  5.310  5.752  0.885   7.522  0.442   3.982  0.885  8.850
497  17.598  3.073  1.117  4.749  0.000   1.955  1.676  12.011  0.838  3.631
498   6.047  6.047  2.791  6.047  0.000   3.721  1.860   9.302  4.186  4.186

     …     X11    X12    X13    X14    X15    X16    X17    X18     X19  \
0    …   3.784  2.162  2.162  9.459  3.243  4.595  0.811  3.514   4.865
2    …   0.383  1.533  3.831  5.364  4.981  8.046  1.149  4.598  14.176
3    …   4.089  2.974  2.974  2.602  5.948  7.063  0.743  3.346  10.781
4    …   6.358  4.046  3.468  1.156  4.624  2.890  0.578  4.624   3.468
5    …   1.199  1.679  3.357  3.597  4.556  5.516  2.878  1.918  14.149
..   …      …      …      …      …      …      …      …      …      …
492  …   1.259  1.008  5.793  3.275  5.793  5.793  1.008  2.267   8.564
```

```
495  …  10.309  3.093  5.155  3.093  9.278  6.186  3.093  2.062   5.155
496  …  10.177  1.770  2.655  5.310  5.752  4.425  0.885  6.637   3.982
497  …   0.838  1.397  4.469  3.073  5.028  6.704  1.676  1.676  12.849
498  …   1.395  1.860  5.116  3.721  3.721  6.512  2.326  4.651  13.488

          X20
0    efectores
2    efectores
3    efectores
4    efectores
5    efectores
..         …
492  efectores
495  efectores
496  efectores
497  efectores
498  efectores

[420 rows x 21 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	420.000000	420.000000	420.000000	420.000000	420.000000	420.000000	
mean	9.612490	6.430798	2.519131	5.914648	0.583629	7.335162	
std	4.419274	2.809640	1.805836	2.645622	0.725762	3.927140	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	6.492500	4.535000	1.169500	3.843000	0.000000	3.980500	
50%	9.202000	6.297000	2.210000	5.528000	0.381000	7.511000	
75%	12.529500	8.238000	3.597500	7.592000	0.866250	9.891250	
max	22.118000	14.286000	8.293000	13.821000	3.306000	17.895000	

	X6	X7	X8	X9	X10	X11	\
count	420.000000	420.000000	420.000000	420.000000	420.00000	420.000000	
mean	2.256183	7.447610	1.817329	5.688764	11.54770	4.294083	
std	1.379372	2.733534	1.137544	3.408184	3.35898	4.209880	
min	0.000000	0.000000	0.000000	0.372000	0.95700	0.000000	
25%	1.354750	5.512000	0.989500	3.092000	9.07225	1.052500	
50%	1.988000	7.339500	1.707000	5.018000	11.66900	2.326000	
75%	3.077000	9.231000	2.587500	7.528000	13.89275	7.041500	
max	7.216000	15.883000	5.128000	16.556000	22.13100	17.188000	

	X12	X13	X14	X15	X16	X17	\
count	420.000000	420.000000	420.000000	420.000000	420.000000	420.000000	
mean	1.894067	3.985750	4.232824	5.780476	5.463238	1.188581	

std	1.041399	1.705442	1.766086	1.879754	2.001282	0.914157
min	0.000000	0.000000	0.000000	0.000000	0.905000	0.000000
25%	1.068500	2.775250	3.050250	4.615500	4.014250	0.593000
50%	1.663000	3.950000	4.018000	5.729000	5.469000	1.069500
75%	2.500000	5.155000	5.285500	7.044000	6.667000	1.677250
max	5.202000	9.605000	10.345000	11.828000	11.345000	4.000000

	X18	X19
count	420.000000	420.000000
mean	3.273031	8.734369
std	1.477846	3.599459
min	0.000000	1.316000
25%	2.271500	5.927750
50%	3.162500	8.333000
75%	4.022750	11.309250
max	8.122000	18.657000

no_efectores

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

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	X7	X8	\
0	8.994	6.860	2.896	8.232	0.000	7.317	3.811	8.079	1.067	
1	14.067	4.281	0.917	1.835	2.446	3.976	0.612	9.786	0.917	
4	10.227	9.773	0.909	9.091	1.818	8.864	1.818	8.182	1.591	
5	7.774	7.067	3.180	4.947	0.707	6.360	2.473	6.714	4.594	
6	13.458	9.938	1.656	8.696	1.035	7.039	1.863	8.282	3.106	
..	…	…	…	…	…	…	…	…	…	
494	15.698	5.814	0.581	9.884	1.163	8.721	2.326	8.140	1.163	
495	9.402	4.274	4.274	7.692	0.855	11.966	2.564	6.838	0.000	
496	9.223	3.883	2.427	14.078	1.942	10.194	2.427	10.194	2.427	
497	7.182	5.525	1.381	4.420	0.000	3.039	3.039	8.564	0.829	
498	6.623	9.272	4.636	9.272	0.662	7.285	2.649	7.285	2.649	

	X9	…	X11	X12	X13	X14	X15	X16	X17	X18	\
0	5.793	…	1.524	2.439	5.488	4.268	3.811	6.555	1.372	2.744	
1	6.116	…	1.835	1.223	2.752	6.422	6.116	5.505	0.917	3.058	
4	4.091	…	0.682	1.136	2.500	6.591	5.909	5.227	0.682	2.045	
5	10.954	…	5.654	4.947	2.473	1.060	9.187	1.767	0.707	6.007	
6	2.484	…	0.828	1.035	2.277	5.176	5.176	4.141	2.277	3.520	
..	…	…	…	…	…	…	…	…	…	…	
494	2.326	…	1.163	1.163	3.488	5.814	6.395	5.233	1.163	3.488	
495	2.564	…	5.128	0.855	1.709	4.274	8.547	9.402	2.564	2.564	
496	4.369	…	0.000	1.942	3.398	3.883	4.369	9.223	1.456	2.913	
497	7.735	…	1.105	1.934	7.459	5.249	3.591	5.249	2.210	4.420	

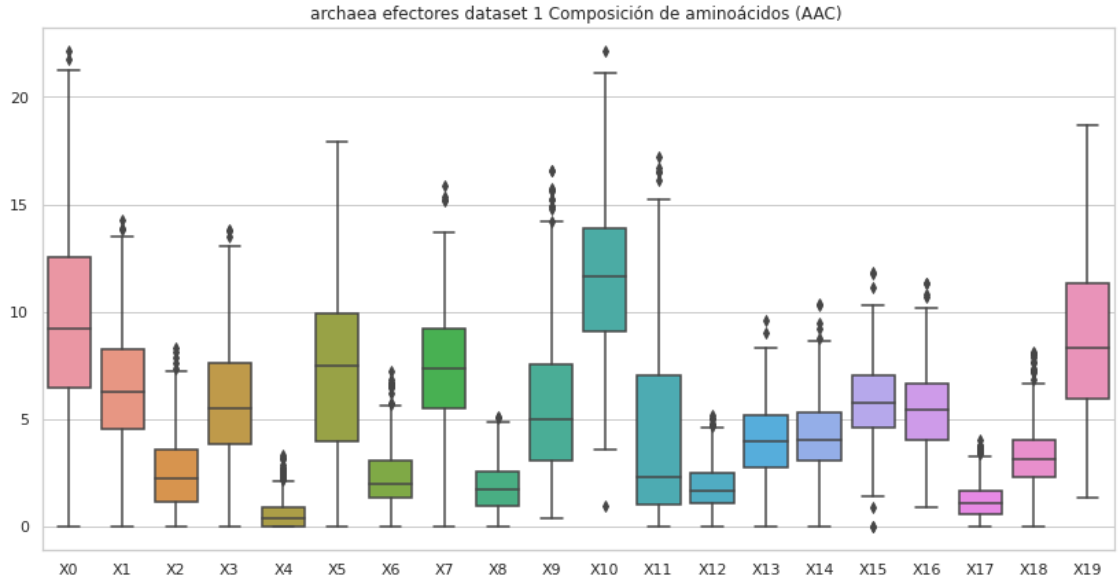
```
498   5.960  …  3.974  1.325  3.974  4.636  5.960  5.298  0.662  3.974

        X19           X20
0     8.994  no_efectores
1    10.398  no_efectores
4     7.500  no_efectores
5     3.887  no_efectores
6     9.524  no_efectores
..      …           …
494   7.558  no_efectores
495   7.692  no_efectores
496   6.311  no_efectores
497  13.260  no_efectores
498   7.285  no_efectores

[416 rows x 21 columns]
```

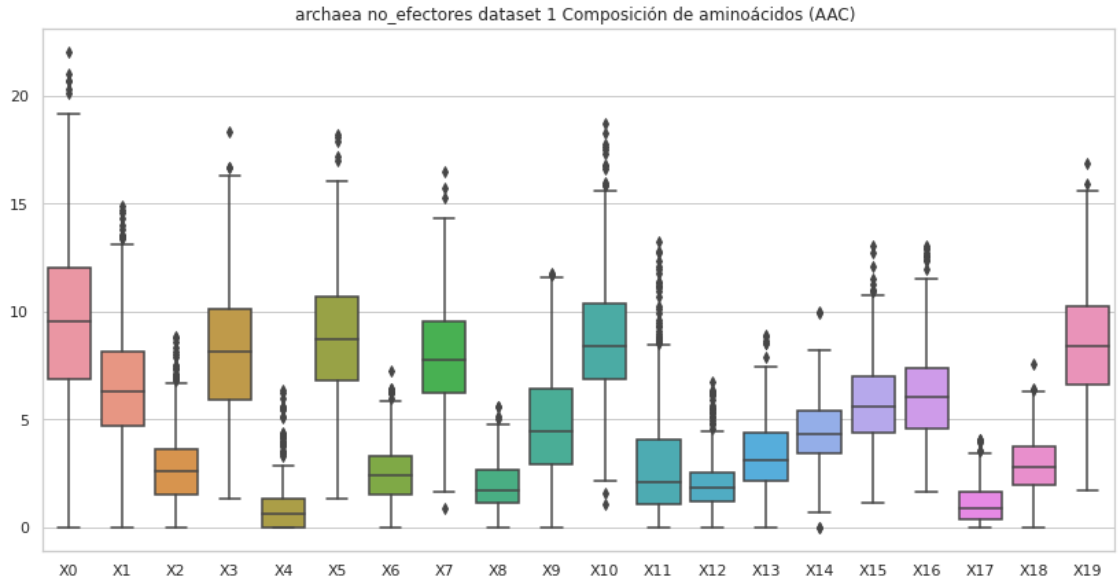
Composición de aminoácidos (AAC) no_efectores archaea dataset 1, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  416.000000  416.000000  416.000000  416.000000  416.000000  416.000000
mean     9.731683    6.460959    2.762623    8.078591    0.990928    8.717368
std      3.890718    2.814210    1.713027    3.096431    1.204548    3.177279
min      0.000000    0.000000    0.000000    1.333000    0.000000    1.347000
25%      6.870750    4.714250    1.495250    5.911250    0.000000    6.829500
50%      9.571000    6.272000    2.580500    8.157500    0.664500    8.734000
75%     12.037000    8.145250    3.619250   10.096500    1.327750   10.699000
max     22.000000   14.876000    8.841000   18.333000    6.383000   18.182000

                X6          X7          X8          X9         X10         X11  \
count  416.000000  416.000000  416.000000  416.000000  416.000000  416.000000
mean     2.499168    7.863923    1.903512    4.826315    8.820873    3.052356
std      1.414958    2.410501    1.145828    2.549152    2.937600    2.853596
min      0.000000    0.917000    0.000000    0.000000    1.099000    0.000000
25%      1.505750    6.208500    1.118000    2.919500    6.846250    1.059000
50%      2.387500    7.769500    1.744000    4.431000    8.408500    2.090500
75%      3.293000    9.567250    2.660250    6.398250   10.345000    4.040000
max      7.258000   16.484000    5.590000   11.785000   18.667000   13.235000

               X12         X13         X14         X15         X16         X17  \
count  416.000000  416.000000  416.000000  416.000000  416.000000  416.000000
mean     2.040534    3.354430    4.400969    5.860947    6.179221    1.072851
std      1.228627    1.675103    1.530300    2.123103    2.259453    0.890808
min      0.000000    0.000000    0.000000    1.136000    1.653000    0.000000
25%      1.186500    2.171500    3.404250    4.418000    4.590000    0.374000
```

50%	1.823500	3.125000	4.315500	5.621000	6.021000	0.913500
75%	2.518750	4.387250	5.376000	7.013500	7.407000	1.633250
max	6.723000	8.889000	10.000000	13.043000	13.043000	4.089000

	X18	X19
count	416.000000	416.000000
mean	2.917702	8.465135
std	1.324364	2.733891
min	0.000000	1.724000
25%	1.977500	6.587000
50%	2.800500	8.394500
75%	3.737750	10.243250
max	7.547000	16.883000





3 Composición de pseudo aminoácidos (PseAAC) hidro_mass

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

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

    if etiq == "efectores":
        df=PseAAC_hidro_mass_efec

    if etiq == "no_efectores":
        df=PseAAC_hidro_mass_no_efec

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

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

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.037370  0.000911  0.016407  0.030079  0.007292  0.031902  0.007292
1    0.021043  0.000390  0.010132  0.008963  0.004287  0.027668  0.000974
2    0.026860  0.000000  0.001965  0.006551  0.006551  0.015068  0.001965
3    0.036558  0.004570  0.024372  0.031988  0.012186  0.039605  0.003047
4    0.052606  0.002923  0.011690  0.075987  0.017535  0.040916  0.008768
..        …         …         …         …         …         …         …
495  0.043471  0.000000  0.024841  0.031051  0.031051  0.031051  0.012420
496  0.088867  0.011849  0.077018  0.100716  0.035547  0.053320  0.011849
497  0.022576  0.000000  0.006092  0.002508  0.005734  0.015409  0.001075
498  0.019667  0.000000  0.019667  0.012103  0.016642  0.030258  0.013616
499  0.052236  0.000000  0.032016  0.035386  0.015165  0.032016  0.005055

           X7        X8        X9  …       X74       X75       X76  \
0    0.020964  0.012761  0.020052  … -0.006863  0.000870  0.022776
1    0.002533  0.001559  0.005261  …  0.000797  0.003792  0.031478
2    0.008517  0.000655  0.022274  …  0.019840  0.003997  0.022842
3    0.024372  0.016756  0.031988  …  0.022899  0.011032  0.023047
4    0.035071  0.032148  0.061374  …  0.001918  0.018828  0.015106
..        …         …         …  …  …         …         …         …
495  0.049681  0.062102  0.037261  … -0.002592  0.006220  0.037846
496  0.118489  0.136263  0.159961  …  0.054849  0.128895 -0.052399
497  0.004659  0.001075  0.020067  …  0.008377 -0.002316  0.020141
498  0.013616  0.004539  0.042361  …  0.024252  0.013945  0.016787
499  0.032016  0.010110  0.058976  … -0.012250  0.006472  0.000367

          X77       X78       X79       X80       X81       X82         X83
0    0.016784  0.020726  0.012031  0.011605  0.013009  0.044115  efectores
1   -0.002872  0.005131  0.033373 -0.002586  0.005266  0.032751  efectores
2    0.010008  0.002019  0.012836  0.011530 -0.000942  0.010788  efectores
3   -0.002615 -0.003113  0.029436 -0.014764 -0.008663  0.036250  efectores
4    0.016812  0.014502  0.010085 -0.007225 -0.023159  0.008385  efectores
..        …         …         …         …         …         …         …
495 -0.031896  0.017054 -0.016840  0.017985  0.002401  0.019280  efectores
496 -0.001804  0.007947  0.047916 -0.102334 -0.082801 -0.028390  efectores
```

```
497  0.022416  0.002182  0.021289  0.011264 -0.003545  0.026104  efectores
498  0.013018 -0.004670  0.006798  0.023654  0.015291  0.011044  efectores
499  0.003651  0.021715  0.031545  0.037090  0.035365  0.023023  efectores

[500 rows x 84 columns]
```

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

```
                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.033434    0.003677    0.026929    0.038489    0.017825    0.028638
std      0.019484    0.006095    0.026032    0.058890    0.025636    0.020982
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.021527    0.000000    0.009242    0.008461    0.007721    0.016748
50%      0.030065    0.001100    0.020665    0.030223    0.013087    0.024905
75%      0.042011    0.004711    0.036459    0.054630    0.021512    0.034401
max      0.221514    0.038423    0.303203    1.100463    0.393022    0.303203

                X6          X7          X8          X9  …         X73  \
count  500.000000  500.000000  500.000000  500.000000  …  500.000000
mean     0.008247    0.030610    0.029468    0.051392  …    0.012583
std      0.009917    0.044003    0.070664    0.080475  …    0.024601
min      0.000000    0.000000    0.000000    0.001579  …   -0.337279
25%      0.002371    0.007765    0.002410    0.025489  …    0.003656
50%      0.006179    0.019265    0.009463    0.040219  …    0.014642
75%      0.010514    0.039930    0.039929    0.058503  …    0.023538
max      0.129944    0.606406    1.100463    1.572090  …    0.126454

               X74         X75         X76         X77         X78         X79  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.004319    0.009530    0.011863    0.005256    0.008796    0.013858
std      0.040572    0.032889    0.022264    0.031805    0.039104    0.026211
min     -0.520899   -0.277176   -0.112599   -0.329305   -0.601208   -0.362373
25%     -0.006229   -0.002255    0.003309   -0.003340    0.000418    0.004828
50%      0.008241    0.004983    0.014715    0.011037    0.007329    0.014238
75%      0.017089    0.018225    0.023443    0.019417    0.017114    0.025583
max      0.222752    0.208878    0.088845    0.094350    0.126313    0.111532

               X80         X81         X82
count  500.000000  500.000000  500.000000
mean     0.002759    0.006579    0.013348
std      0.043177    0.047836    0.020007
min     -0.522564   -0.854787   -0.101552
25%     -0.007152   -0.000705    0.003471
50%      0.009271    0.005692    0.014881
```

```
75%      0.018902    0.018181    0.023484
max      0.167197    0.109199    0.098262

[8 rows x 83 columns]
```

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.032334	0.000000	0.029594	0.026306	0.019729	0.029046	0.003836	
1	0.021260	0.003697	0.002773	0.006008	0.004160	0.014790	0.001387	
2	0.020224	0.000000	0.016853	0.016853	0.003371	0.030335	0.003371	
3	0.040858	0.040858	0.032687	0.032687	0.024515	0.049030	0.000000	
4	0.035753	0.006356	0.031781	0.030986	0.008740	0.028603	0.005562	
..	…	…	…	…	…	…	…	
495	0.039365	0.003579	0.032208	0.050101	0.007157	0.028629	0.000000	
496	0.027575	0.005805	0.042088	0.030478	0.010159	0.030478	0.007257	
497	0.014604	0.000000	0.008987	0.006179	0.015166	0.017413	0.001685	
498	0.041643	0.004164	0.058301	0.045808	0.024986	0.045808	0.016657	
499	0.056514	0.000000	0.032294	0.064588	0.016147	0.032294	0.016147	

	X7	X8	X9	…	X74	X75	X76	\
0	0.020825	0.005480	0.035074	…	0.002470	0.002987	0.013903	
1	0.009243	0.002773	0.025419	…	0.008849	-0.000076	0.018825	
2	0.000000	0.003371	0.023594	…	0.016447	-0.001304	0.029993	
3	0.008172	0.065373	0.008172	…	0.006510	0.004205	0.002253	
4	0.014301	0.002384	0.039726	…	-0.008287	-0.001023	0.024735	
..	…	…	…	…	…	…	…	
495	0.010736	0.021472	0.028629	…	-0.010261	0.000371	0.023527	
496	0.013062	0.000000	0.015964	…	0.002159	0.005535	0.023199	
497	0.015728	0.002247	0.028085	…	0.016472	0.003284	0.008240	
498	0.037479	0.024986	0.041643	…	0.017741	0.058662	-0.014252	
499	0.024220	0.024220	0.056514	…	-0.033860	0.009638	0.076921	

	X77	X78	X79	X80	X81	X82	X83
0	0.016073	0.008758	0.010862	0.019353	0.007702	0.022360	no_efectores
1	0.013539	0.002870	0.015011	0.013225	0.004058	0.024319	no_efectores
2	0.016652	-0.009555	0.027220	-0.022299	-0.008778	0.031638	no_efectores
3	-0.031908	0.014173	-0.080820	0.021244	0.069961	-0.009395	no_efectores
4	0.006211	0.013207	0.012339	0.007392	0.016944	0.029974	no_efectores
..	…	…	…	…	…	…	…
495	0.007493	0.021542	0.044026	0.019444	0.036784	0.001639	no_efectores
496	-0.003760	0.014160	0.006077	0.012924	-0.000150	0.013805	no_efectores
497	0.009326	-0.002255	0.002029	0.019077	-0.003406	0.008811	no_efectores

```
498 -0.043106 -0.015297  0.045240 -0.045976 -0.014207 -0.045529  no_efectores
499  0.027196  0.077907  0.032003  0.062497  0.078027  0.036635  no_efectores

[500 rows x 84 columns]
```

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

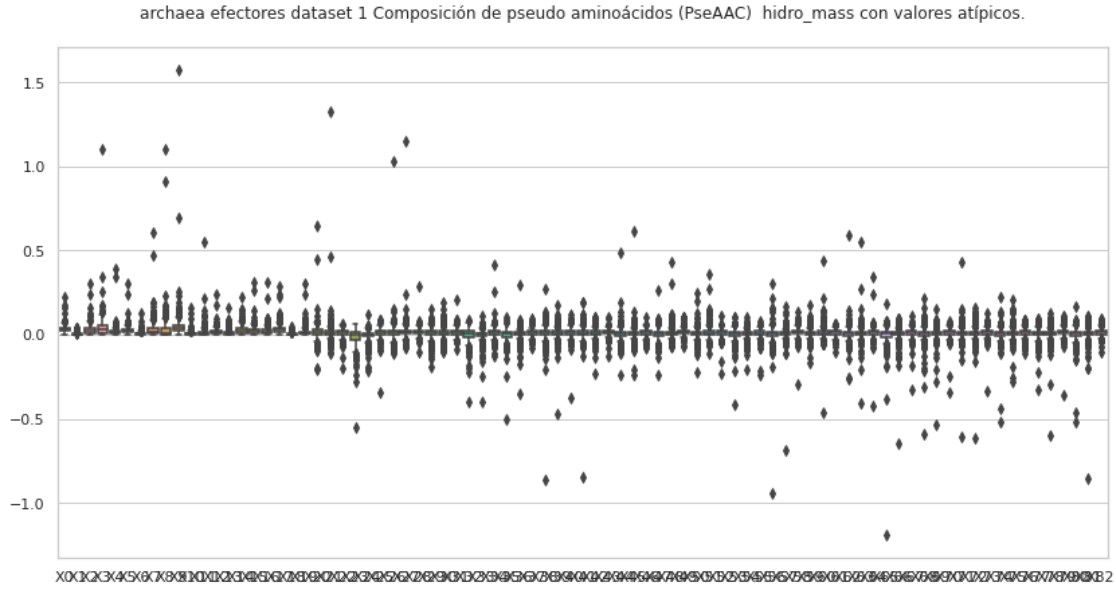
	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.031666	0.003400	0.028858	0.032464	0.013921	0.028480	
std	0.052419	0.049263	0.048473	0.063335	0.018327	0.030494	
min	-0.790137	-1.053516	-0.790137	-1.053516	-0.263379	-0.526758	
25%	0.024507	0.000000	0.018531	0.018762	0.006488	0.019655	
50%	0.033121	0.002460	0.029834	0.033073	0.011056	0.028574	
75%	0.043893	0.005817	0.043368	0.048752	0.018302	0.036828	
max	0.152390	0.087080	0.140288	0.151735	0.130620	0.108850	

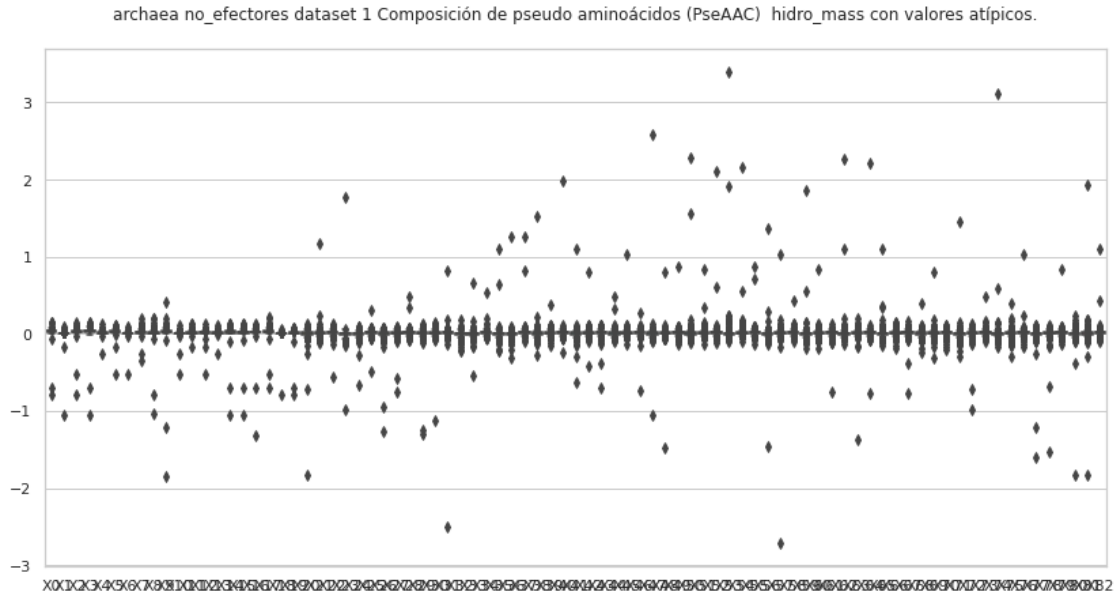
	X6	X7	X8	X9	…	X73	\
count	500.000000	500.000000	500.000000	500.000000	…	500.000000	
mean	0.007448	0.021342	0.012723	0.029537	…	0.017965	
std	0.025724	0.031149	0.064136	0.104793	…	0.029962	
min	-0.526758	-0.347995	-1.043985	-1.843654	…	-0.095402	
25%	0.002759	0.008275	0.002925	0.020648	…	0.007624	
50%	0.006313	0.015801	0.007817	0.030142	…	0.018270	
75%	0.012042	0.029448	0.018273	0.043883	…	0.027037	
max	0.101157	0.202313	0.206471	0.413629	…	0.488330	

	X74	X75	X76	X77	X78	X79	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.008256	0.006799	0.017427	-0.003551	0.004010	0.017583	
std	0.143886	0.036304	0.051616	0.094108	0.079777	0.042225	
min	-0.170810	-0.289119	-0.169881	-1.609585	-1.538625	-0.130588	
25%	-0.007960	-0.001930	0.006934	-0.007261	-0.001295	0.007317	
50%	0.002969	0.006091	0.017492	0.004088	0.006926	0.017521	
75%	0.012328	0.017381	0.026511	0.014449	0.019890	0.027303	
max	3.110132	0.396329	1.026028	0.108034	0.109236	0.827453	

	X80	X81	X82
count	500.000000	500.000000	500.000000
mean	-0.001136	0.008511	0.018905
std	0.088503	0.122662	0.055666
min	-1.824734	-1.828495	-0.102108
25%	-0.009080	-0.003112	0.006225
50%	0.002498	0.006302	0.017914
75%	0.012577	0.017339	0.026889

```
max       0.243448     1.928067     1.096466

[8 rows x 83 columns]
```





3.1 Composición de pseudo aminoácidos (PseAAC) hidro_mass, sin valores atípicos

[6]:
```
#hidro_mass
transf = "Composición de pseudo aminoácidos (PseAAC) "
transf2 = "PseAAC"
estado = "sin valores atípicos.\n"
comp = "hidro_mass"
df=""

out = (str(r3) + '/ds' + str(dataset) + '_' + str(transf2) + '_' + str(comp) +␣
↪'_' + str(organismo) + '.csv')
os.makedirs(str(r3), exist_ok=True)
df_out = pd.DataFrame()


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

    if etiq == "efectores":
        df=PseAAC_hidro_mass_efec

    if etiq == "no_efectores":
        df=PseAAC_hidro_mass_no_efec

    del df['X83']
    #Se eliminan todas las filas que tengan valores atípicos en al menos una de␣
↪sus columnas.
    df = (df[(np.abs(stats.zscore(df)) < 3).all(axis=1)])
    df['X83'] = etiq
    df_out = pd.concat([df_out,df])

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


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

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

```
    ax = sns.boxplot(data=df)
    ax.set_title(organismo +' '+str(etiq)+" dataset "+str(dataset)+"␣
↪"+str(transf)+" "+str(comp))
```

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.037370	0.000911	0.016407	0.030079	0.007292	0.031902	0.007292	
1	0.021043	0.000390	0.010132	0.008963	0.004287	0.027668	0.000974	
2	0.026860	0.000000	0.001965	0.006551	0.006551	0.015068	0.001965	
3	0.036558	0.004570	0.024372	0.031988	0.012186	0.039605	0.003047	
4	0.052606	0.002923	0.011690	0.075987	0.017535	0.040916	0.008768	
..	…	…	…	…	…	…	…	
493	0.006949	0.011582	0.016214	0.046327	0.002316	0.020847	0.009265	
495	0.043471	0.000000	0.024841	0.031051	0.031051	0.031051	0.012420	
497	0.022576	0.000000	0.006092	0.002508	0.005734	0.015409	0.001075	
498	0.019667	0.000000	0.019667	0.012103	0.016642	0.030258	0.013616	
499	0.052236	0.000000	0.032016	0.035386	0.015165	0.032016	0.005055	

	X7	X8	X9	…	X74	X75	X76	\
0	0.020964	0.012761	0.020052	…	-0.006863	0.000870	0.022776	
1	0.002533	0.001559	0.005261	…	0.000797	0.003792	0.031478	
2	0.008517	0.000655	0.022274	…	0.019840	0.003997	0.022842	
3	0.024372	0.016756	0.031988	…	0.022899	0.011032	0.023047	
4	0.035071	0.032148	0.061374	…	0.001918	0.018828	0.015106	
..	…	…	…	…	…	…	…	
493	0.018531	0.032429	0.037061	…	0.034979	0.026567	-0.001717	
495	0.049681	0.062102	0.037261	…	-0.002592	0.006220	0.037846	
497	0.004659	0.001075	0.020067	…	0.008377	-0.002316	0.020141	
498	0.013616	0.004539	0.042361	…	0.024252	0.013945	0.016787	
499	0.032016	0.010110	0.058976	…	-0.012250	0.006472	0.000367	

	X77	X78	X79	X80	X81	X82	X83
0	0.016784	0.020726	0.012031	0.011605	0.013009	0.044115	efectores
1	-0.002872	0.005131	0.033373	-0.002586	0.005266	0.032751	efectores
2	0.010008	0.002019	0.012836	0.011530	-0.000942	0.010788	efectores
3	-0.002615	-0.003113	0.029436	-0.014764	-0.008663	0.036250	efectores
4	0.016812	0.014502	0.010085	-0.007225	-0.023159	0.008385	efectores
..	…	…	…	…	…	…	
493	-0.015855	0.013986	-0.008565	-0.017221	0.015536	0.015143	efectores
495	-0.031896	0.017054	-0.016840	0.017985	0.002401	0.019280	efectores
497	0.022416	0.002182	0.021289	0.011264	-0.003545	0.026104	efectores
498	0.013018	-0.004670	0.006798	0.023654	0.015291	0.011044	efectores
499	0.003651	0.021715	0.031545	0.037090	0.035365	0.023023	efectores

[434 rows x 84 columns]

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

	X0	X1	X2	X3	X4	X5	\
count	434.000000	434.000000	434.000000	434.000000	434.000000	434.000000	
mean	0.031355	0.002544	0.022413	0.029265	0.014705	0.025456	
std	0.013478	0.003951	0.016643	0.024374	0.010535	0.011632	
min	0.002020	0.000000	0.001439	0.000000	0.000000	0.000000	
25%	0.021494	0.000000	0.008870	0.007194	0.007297	0.016242	
50%	0.029509	0.000776	0.018240	0.022828	0.012351	0.023713	
75%	0.039975	0.003696	0.033245	0.048763	0.019277	0.032015	
max	0.078318	0.021343	0.088272	0.105933	0.076235	0.074306	

	X6	X7	X8	X9	…	X73	\
count	434.000000	434.000000	434.000000	434.000000	…	434.000000	
mean	0.007097	0.022422	0.018595	0.040537	…	0.013887	
std	0.006263	0.020805	0.024566	0.022549	…	0.015508	
min	0.000000	0.000000	0.000000	0.001579	…	-0.040159	
25%	0.002367	0.007112	0.002060	0.023412	…	0.005093	
50%	0.005785	0.014811	0.006897	0.036721	…	0.015061	
75%	0.009713	0.032741	0.026143	0.051207	…	0.023361	
max	0.035487	0.122947	0.174950	0.141947	…	0.084314	

	X74	X75	X76	X77	X78	X79	\
count	434.000000	434.000000	434.000000	434.000000	434.000000	434.000000	
mean	0.006666	0.008746	0.014116	0.007462	0.009965	0.014314	
std	0.018754	0.018584	0.014813	0.019953	0.018109	0.016214	
min	-0.068891	-0.058078	-0.045492	-0.081072	-0.061913	-0.042273	
25%	-0.002938	-0.001231	0.005692	-0.001386	0.000740	0.005478	
50%	0.008782	0.004656	0.015167	0.011257	0.007058	0.014205	
75%	0.016484	0.014882	0.023516	0.018507	0.015261	0.023998	
max	0.073558	0.081301	0.056083	0.094350	0.101168	0.080029	

	X80	X81	X82
count	434.000000	434.000000	434.000000
mean	0.007992	0.009641	0.014005
std	0.021254	0.018461	0.015704
min	-0.074710	-0.052709	-0.041275
25%	-0.003880	-0.000182	0.006215
50%	0.010585	0.005653	0.015227
75%	0.018742	0.015886	0.023164
max	0.087159	0.100137	0.066773

[8 rows x 83 columns]

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.032334  0.000000  0.029594  0.026306  0.019729  0.029046  0.003836
1    0.021260  0.003697  0.002773  0.006008  0.004160  0.014790  0.001387
2    0.020224  0.000000  0.016853  0.016853  0.003371  0.030335  0.003371
4    0.035753  0.006356  0.031781  0.030986  0.008740  0.028603  0.005562
5    0.063816  0.005801  0.040610  0.052213  0.020305  0.055113  0.037709
..        …         …         …         …         …         …         …
495  0.039365  0.003579  0.032208  0.050101  0.007157  0.028629  0.000000
496  0.027575  0.005805  0.042088  0.030478  0.010159  0.030478  0.007257
497  0.014604  0.000000  0.008987  0.006179  0.015166  0.017413  0.001685
498  0.041643  0.004164  0.058301  0.045808  0.024986  0.045808  0.016657
499  0.056514  0.000000  0.032294  0.064588  0.016147  0.032294  0.016147

           X7        X8        X9  …       X74       X75       X76  \
0    0.020825  0.005480  0.035074  …  0.002470  0.002987  0.013903
1    0.009243  0.002773  0.025419  …  0.008849 -0.000076  0.018825
2    0.000000  0.003371  0.023594  …  0.016447 -0.001304  0.029993
4    0.014301  0.002384  0.039726  … -0.008287 -0.001023  0.024735
5    0.089922  0.046411  0.078319  …  0.007217  0.009728  0.015528
..        …         …         …  …         …         …         …
495  0.010736  0.021472  0.028629  … -0.010261  0.000371  0.023527
496  0.013062  0.000000  0.015964  …  0.002159  0.005535  0.023199
497  0.015728  0.002247  0.028085  …  0.016472  0.003284  0.008240
498  0.037479  0.024986  0.041643  …  0.017741  0.058662 -0.014252
499  0.024220  0.024220  0.056514  … -0.033860  0.009638  0.076921

           X77       X78       X79       X80       X81       X82           X83
0    0.016073  0.008758  0.010862  0.019353  0.007702  0.022360  no_efectores
1    0.013539  0.002870  0.015011  0.013225  0.004058  0.024319  no_efectores
2    0.016652 -0.009555  0.027220 -0.022299 -0.008778  0.031638  no_efectores
4    0.006211  0.013207  0.012339  0.007392  0.016944  0.029974  no_efectores
5   -0.023263 -0.011911  0.015322  0.018695  0.030943  0.014155  no_efectores
..        …         …         …         …         …         …           …
495  0.007493  0.021542  0.044026  0.019444  0.036784  0.001639  no_efectores
496 -0.003760  0.014160  0.006077  0.012924 -0.000150  0.013805  no_efectores
497  0.009326 -0.002255  0.002029  0.019077 -0.003406  0.008811  no_efectores
498 -0.043106 -0.015297  0.045240 -0.045976 -0.014207 -0.045529  no_efectores
499  0.027196  0.077907  0.032003  0.062497  0.078027  0.036635  no_efectores

[462 rows x 84 columns]
```

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

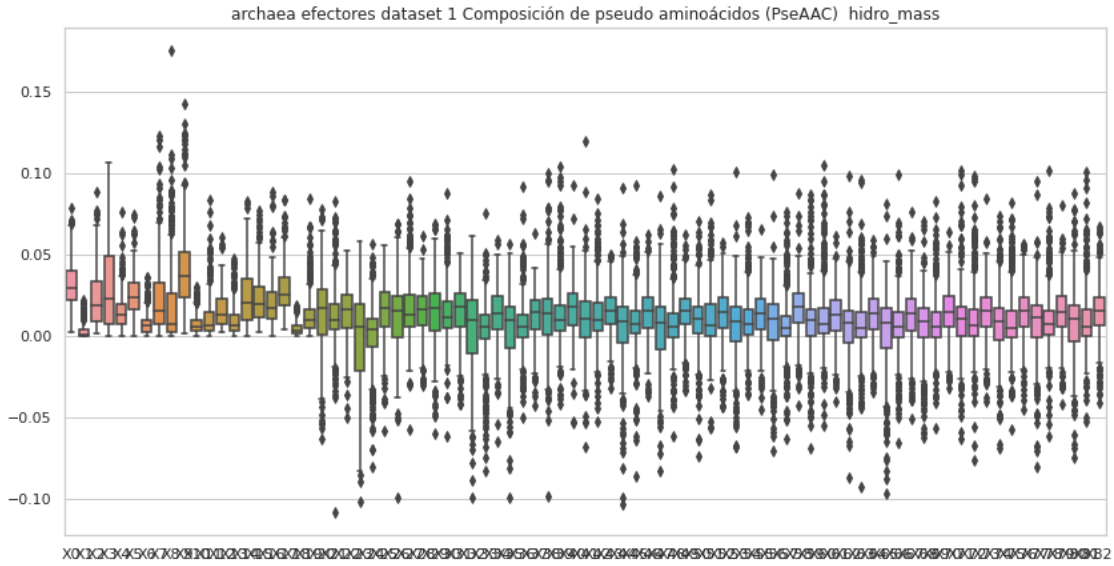
	X0	X1	X2	X3	X4	X5	\
count	462.000000	462.000000	462.000000	462.000000	462.000000	462.000000	
mean	0.033627	0.004663	0.030141	0.033836	0.013383	0.028818	
std	0.014378	0.007630	0.017546	0.020691	0.010391	0.012708	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.024483	0.000000	0.018143	0.018332	0.006497	0.019638	
50%	0.032474	0.002342	0.029172	0.031816	0.010894	0.028323	
75%	0.043239	0.005536	0.041100	0.046207	0.017095	0.035255	
max	0.077226	0.065039	0.103085	0.101335	0.064208	0.088814	

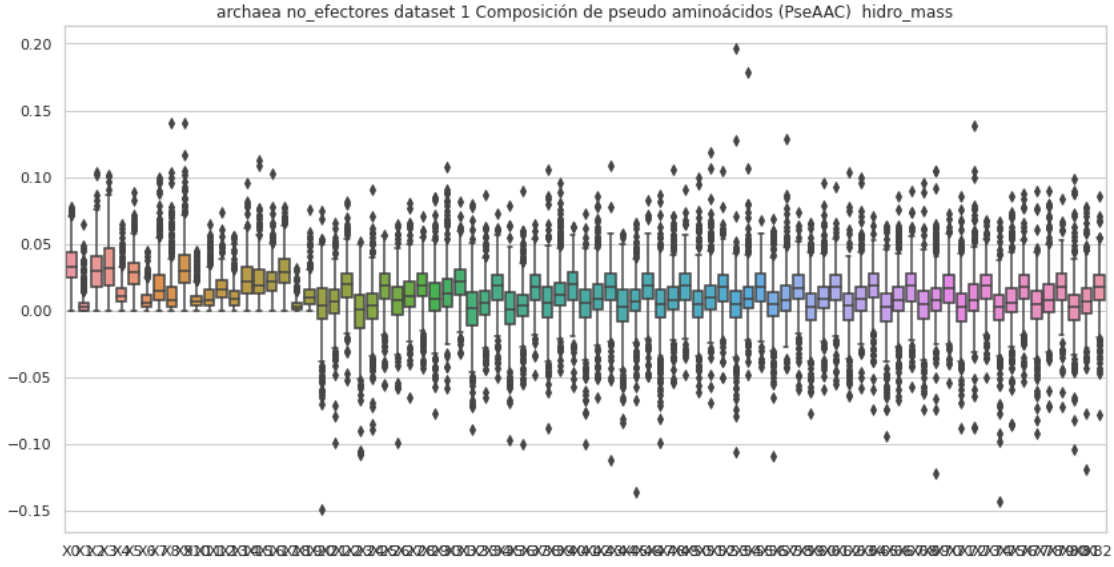
	X6	X7	X8	X9	…	X73	\
count	462.000000	462.000000	462.000000	462.000000	…	462.000000	
mean	0.007831	0.020024	0.014085	0.033059	…	0.017002	
std	0.007083	0.017168	0.018214	0.018565	…	0.015733	
min	0.000000	0.000000	0.000000	0.000000	…	-0.050202	
25%	0.002802	0.007944	0.002833	0.020312	…	0.008515	
50%	0.006167	0.015017	0.007415	0.029416	…	0.018227	
75%	0.011433	0.026651	0.017419	0.041615	…	0.026513	
max	0.044407	0.099089	0.140917	0.140917	…	0.067112	

	X74	X75	X76	X77	X78	X79	\
count	462.000000	462.000000	462.000000	462.000000	462.000000	462.000000	
mean	0.001430	0.007194	0.016832	0.002836	0.009229	0.016994	
std	0.020344	0.020708	0.017609	0.020398	0.019788	0.017374	
min	-0.142862	-0.085728	-0.053320	-0.092441	-0.072164	-0.072690	
25%	-0.007015	-0.001582	0.008439	-0.006032	-0.000933	0.008074	
50%	0.003043	0.006041	0.017745	0.004337	0.007178	0.017840	
75%	0.012002	0.017146	0.026465	0.014358	0.019551	0.027195	
max	0.066546	0.085237	0.087563	0.089554	0.089837	0.083505	

	X80	X81	X82
count	462.000000	462.000000	462.000000
mean	0.002170	0.008425	0.016465
std	0.020815	0.018851	0.017758
min	-0.104565	-0.119267	-0.078402
25%	-0.007001	-0.001944	0.007351
50%	0.002787	0.006336	0.018134
75%	0.012032	0.017047	0.026224
max	0.098510	0.078027	0.085451

[8 rows x 83 columns]





4 Composición de pseudo aminoácidos (PseAAC) mass

[7]:
```
#mass
transf = "Composición de pseudo aminoácidos (PseAAC) "
transf2 = "PseAAC"
```

```
estado = "con valores atípicos.\n"
comp = "mass"
df=""

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

    if etiq == "efectores":
        df=PseAAC_mass_efec

    if etiq == "no_efectores":
        df=PseAAC_mass_no_efec

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

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

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.046768  0.001141  0.020532  0.037642  0.009125  0.039924  0.009125
1    0.023286  0.000431  0.011212  0.009918  0.004743  0.030617  0.001078
2    0.048880  0.000000  0.003577  0.011922  0.011922  0.027420  0.003577
3    0.039226  0.004903  0.026151  0.034323  0.013075  0.042495  0.003269
4    0.077794  0.004322  0.017288  0.112369  0.025931  0.060506  0.012966
..        …         …         …         …         …         …         …
495  0.046413  0.000000  0.026521  0.033152  0.033152  0.033152  0.013261
496  0.071706  0.009561  0.062145  0.081267  0.028682  0.043024  0.009561
497  0.036567  0.000000  0.009867  0.004063  0.009287  0.024959  0.001741
498  0.039109  0.000000  0.039109  0.024067  0.033092  0.060168  0.027075
499  0.059125  0.000000  0.036238  0.040053  0.017165  0.036238  0.005722

           X7        X8        X9  …       X32       X33       X34  \
```

```
0    0.026236  0.015969  0.025095  …  0.019315  0.031418  0.015877
1    0.002803  0.001725  0.005821  …  0.035203  0.043975  0.060366
2    0.015498  0.001192  0.040534  …  0.042319  0.035981  0.020843
3    0.026151  0.017979  0.034323  …  0.037315  0.026689  0.018541
4    0.051863  0.047541  0.090760  … -0.034259  0.012252  0.009244
..        …         …       …  …        …         …
495  0.053043  0.066304  0.039782  … -0.014481  0.050139  0.016750
496  0.095608  0.109949  0.129071  … -0.052468  0.018574 -0.057432
497  0.007546  0.001741  0.032504  …  0.036172  0.037452  0.027378
498  0.027075  0.009025  0.084235  …  0.032862  0.016737 -0.003796
499  0.036238  0.011444  0.066754  …  0.021834  0.004474  0.028106

          X35       X36       X37       X38       X39       X40        X41
0    0.031871  0.039507  0.028670  0.028504  0.015056  0.055208  efectores
1    0.044914  0.036715  0.034508  0.034833  0.036930  0.036242  efectores
2    0.024213  0.031380  0.038341  0.041568  0.023359  0.019632  efectores
3    0.020484  0.041331  0.043901  0.024729  0.031584  0.038895  efectores
4    0.047734 -0.000446  0.013668  0.022338  0.014913  0.012400  efectores
..        …         …         …         …         …         …
495  0.012664 -0.002364 -0.011041  0.040406 -0.017979  0.020584  efectores
496  0.043152 -0.044161  0.019277 -0.042280  0.038663 -0.022907  efectores
497  0.037853  0.037317  0.037708  0.032624  0.034483  0.042283  efectores
498  0.055575  0.006406  0.023172  0.033382  0.013518  0.021960  efectores
499  0.009529  0.038892  0.026477  0.000415  0.035705  0.026059  efectores

[500 rows x 42 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.047523    0.004901    0.035949    0.050475    0.024288    0.039856
std      0.017822    0.008271    0.024591    0.041634    0.016706    0.014988
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.036147    0.000000    0.017108    0.015557    0.013021    0.030029
50%      0.045103    0.001857    0.031321    0.038614    0.020078    0.038374
75%      0.058112    0.006451    0.048833    0.080408    0.032532    0.047492
max      0.128670    0.076047    0.163777    0.198841    0.104523    0.171106

                X6          X7          X8          X9  …         X31  \
count  500.000000  500.000000  500.000000  500.000000  …  500.000000
mean     0.011224    0.040830    0.035791    0.069019  …    0.015289
std      0.009821    0.035353    0.043000    0.035134  …    0.024045
min      0.000000    0.000000    0.000000    0.002487  …   -0.085281
25%      0.003689    0.012736    0.004156    0.043418  …    0.003087
```

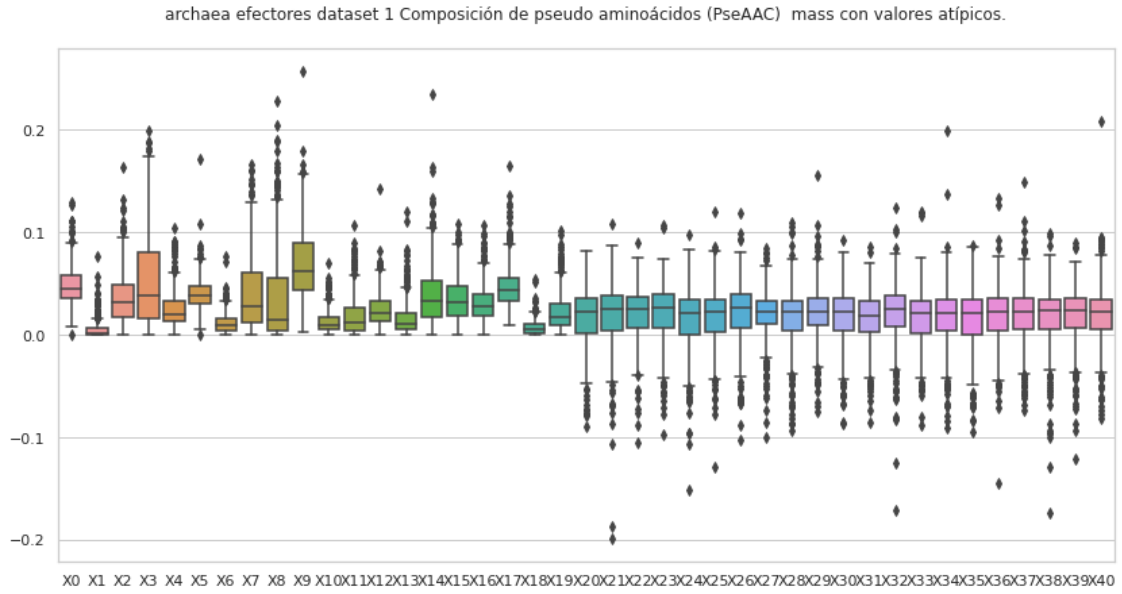
```
50%      0.009373    0.027336    0.014938    0.062039  …    0.018805
75%      0.015742    0.060387    0.055740    0.089211  …    0.032920
max      0.076047    0.166188    0.228141    0.256970  …    0.086350

              X32         X33         X34         X35         X36         X37  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.021429    0.015771    0.017180    0.014667    0.018571    0.019503
std      0.028826    0.025671    0.027564    0.028063    0.027001    0.025872
min     -0.170793   -0.088129   -0.090750   -0.094774   -0.144801   -0.073780
25%      0.008811    0.001549    0.003608   -0.000171    0.003672    0.005852
50%      0.025397    0.020831    0.021658    0.020974    0.022814    0.022829
75%      0.038437    0.032855    0.034704    0.034446    0.036026    0.035252
max      0.123996    0.120200    0.199733    0.086917    0.132745    0.149029

              X38         X39         X40
count  500.000000  500.000000  500.000000
mean     0.017745    0.019812    0.019467
std      0.028535    0.026535    0.027833
min     -0.173879   -0.120892   -0.081802
25%      0.005503    0.007459    0.005795
50%      0.023640    0.024513    0.022253
75%      0.034984    0.036467    0.035111
max      0.098403    0.089438    0.208707

[8 rows x 41 columns]



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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.052094  0.000000  0.047679  0.042381  0.031786  0.046796  0.006181
1    0.035063  0.006098  0.004573  0.009909  0.006860  0.024391  0.002287
2    0.023799  0.000000  0.019832  0.019832  0.003966  0.035698  0.003966
3    0.103990  0.103990  0.083192  0.083192  0.062394  0.124788  0.000000
4    0.046059  0.008188  0.040941  0.039918  0.011259  0.036847  0.007165
..        …         …         …         …         …         …         …
495  0.049636  0.004512  0.040612  0.063174  0.009025  0.036099  0.000000
496  0.033123  0.006973  0.050556  0.036609  0.012203  0.036609  0.008717
497  0.044731  0.000000  0.027526  0.018924  0.046451  0.053333  0.005161
498  0.061898  0.006190  0.086658  0.068088  0.037139  0.068088  0.024759
499  0.072204  0.000000  0.041259  0.082518  0.020630  0.041259  0.020630

           X7        X8        X9  …       X32       X33       X34  \
0    0.033552  0.008829  0.056509  …  0.017368  0.010598  0.026206
```

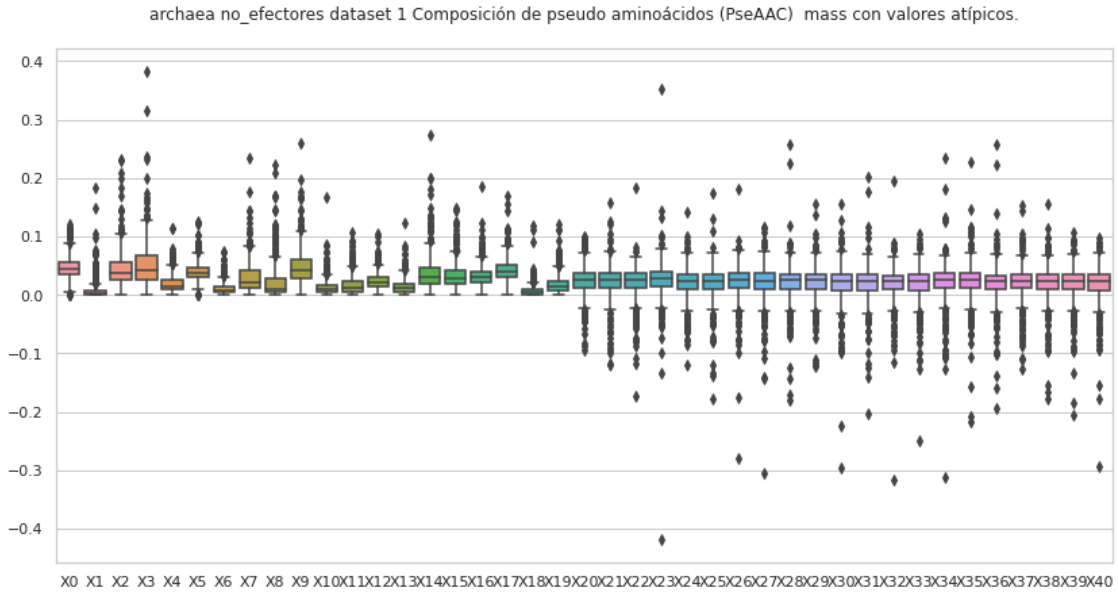
```
1    0.015245  0.004573  0.041923  …  0.030496  0.032778  0.032671
2    0.000000  0.003966  0.027765  …  0.016058  0.033871  0.054316
3    0.020798  0.166384  0.020798  …  0.079687 -0.249668 -0.018834
4    0.018424  0.003071  0.051177  … -0.000582  0.025685  0.023064
..        …         …         …  …         …         …
495  0.013537  0.027074  0.036099  …  0.011303  0.018066  0.049685
496  0.015690  0.000000  0.019176  …  0.046431  0.030586  0.024070
497  0.048171  0.006882  0.086020  …  0.007606  0.035483 -0.009811
498  0.055709  0.037139  0.061898  … -0.003781  0.083121 -0.076076
499  0.030944  0.030944  0.072204  … -0.007586 -0.046507  0.037910

          X35       X36       X37       X38       X39       X40           X41
0    0.013712  0.000681  0.025850  0.022399  0.017500  0.036025  no_efectores
1    0.034955  0.033864  0.042843  0.031047  0.024756  0.040108  no_efectores
2    0.012008  0.049439  0.029778  0.035296  0.032032  0.037231  no_efectores
3    0.228040 -0.193739 -0.006939  0.005735 -0.205699 -0.023912  no_efectores
4    0.029641  0.024518  0.026154  0.031865  0.015896  0.038614  no_efectores
..        …         …         …         …         …         …           …
495  0.013694  0.024024  0.020520  0.029666  0.055514  0.002067  no_efectores
496  0.034594  0.041787  0.034197  0.027866  0.007299  0.016583  no_efectores
497 -0.002252  0.014663  0.019280  0.025237  0.006215  0.026987  no_efectores
498  0.062356  0.002124  0.028225 -0.021185  0.067245 -0.067673  no_efectores
499  0.114713  0.016360  0.002892  0.098275  0.040887  0.046805  no_efectores

[500 rows x 42 columns]
```

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

```
                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.046891    0.008059    0.043970    0.050859    0.019673    0.040134
std      0.019333    0.016788    0.031056    0.039464    0.015987    0.016799
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.034911    0.000000    0.024872    0.025071    0.009005    0.029740
50%      0.045551    0.003156    0.038582    0.042814    0.014988    0.037655
75%      0.056174    0.008006    0.057039    0.067469    0.026702    0.047410
max      0.121961    0.182942    0.232700    0.382213    0.114907    0.124788

                X6          X7          X8          X9  …         X31  \
count  500.000000  500.000000  500.000000  500.000000  …  500.000000
mean     0.011179    0.029863    0.023019    0.048971  …    0.020521
std      0.010831    0.026358    0.030998    0.029542  …    0.030186
min      0.000000    0.000000    0.000000    0.000000  …   -0.203088
25%      0.004182    0.011902    0.004275    0.028740  …    0.007913
50%      0.008595    0.021924    0.010819    0.043223  …    0.024264
```

```
75%      0.015014    0.041019    0.029051    0.061694  …    0.035624
max      0.074075    0.233874    0.222377    0.259262  …    0.202753

              X32         X33         X34         X35         X36         X37  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.017949    0.018530    0.021114    0.022215    0.019525    0.020516
std      0.030803    0.030011    0.034200    0.032299    0.032360    0.028375
min     -0.316325   -0.249668   -0.311693   -0.216921   -0.193739   -0.127485
25%      0.009250    0.008514    0.012336    0.011243    0.008790    0.011168
50%      0.022885    0.023281    0.025214    0.025633    0.022980    0.024762
75%      0.033908    0.034161    0.036599    0.037035    0.034056    0.034520
max      0.195487    0.105810    0.233477    0.228040    0.256806    0.153080

              X38         X39         X40
count  500.000000  500.000000  500.000000
mean     0.019600    0.019848    0.018195
std      0.030904    0.029428    0.032013
min     -0.176999   -0.205699   -0.293281
25%      0.010100    0.009946    0.008437
50%      0.023715    0.023919    0.024046
75%      0.034679    0.035157    0.034189
max      0.155958    0.107379    0.097216

[8 rows x 41 columns]
```





4.1 Composición de pseudo aminoácidos (PseAAC) mass, sin valores atípicos

[8]:
```
#mass
transf = "Composición de pseudo aminoácidos (PseAAC) "
transf2 = "PseAAC"
estado = "sin valores atípicos.\n"
comp = "mass"
df=""

out = (str(r3) + '/ds' + str(dataset) + '_' + str(transf2) + '_' + str(comp) +␣
 ↪'_' + str(organismo) + '.csv')
os.makedirs(str(r3), exist_ok=True)
df_out = pd.DataFrame()

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

    if etiq == "efectores":
        df=PseAAC_mass_efec

    if etiq == "no_efectores":
        df=PseAAC_mass_no_efec

    del df['X41']
    df = (df[(np.abs(stats.zscore(df)) < 3).all(axis=1)])
    df['X41'] = etiq
```

```
    df_out = pd.concat([df_out,df])

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

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

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

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.046768  0.001141  0.020532  0.037642  0.009125  0.039924  0.009125
1    0.023286  0.000431  0.011212  0.009918  0.004743  0.030617  0.001078
2    0.048880  0.000000  0.003577  0.011922  0.011922  0.027420  0.003577
3    0.039226  0.004903  0.026151  0.034323  0.013075  0.042495  0.003269
4    0.077794  0.004322  0.017288  0.112369  0.025931  0.060506  0.012966
..        …         …         …         …         …         …         …
495  0.046413  0.000000  0.026521  0.033152  0.033152  0.033152  0.013261
496  0.071706  0.009561  0.062145  0.081267  0.028682  0.043024  0.009561
497  0.036567  0.000000  0.009867  0.004063  0.009287  0.024959  0.001741
498  0.039109  0.000000  0.039109  0.024067  0.033092  0.060168  0.027075
499  0.059125  0.000000  0.036238  0.040053  0.017165  0.036238  0.005722

           X7        X8        X9  …       X32       X33       X34  \
0    0.026236  0.015969  0.025095  …  0.019315  0.031418  0.015877
1    0.002803  0.001725  0.005821  …  0.035203  0.043975  0.060366
2    0.015498  0.001192  0.040534  …  0.042319  0.035981  0.020843
3    0.026151  0.017979  0.034323  …  0.037315  0.026689  0.018541
4    0.051863  0.047541  0.090760  … -0.034259  0.012252  0.009244
..        …         …         …  …  …       …         …         …
495  0.053043  0.066304  0.039782  … -0.014481  0.050139  0.016750
496  0.095608  0.109949  0.129071  … -0.052468  0.018574 -0.057432
497  0.007546  0.001741  0.032504  …  0.036172  0.037452  0.027378
498  0.027075  0.009025  0.084235  …  0.032862  0.016737 -0.003796
499  0.036238  0.011444  0.066754  …  0.021834  0.004474  0.028106
```

```
          X35       X36       X37       X38       X39       X40        X41
0    0.031871  0.039507  0.028670  0.028504  0.015056  0.055208  efectores
1    0.044914  0.036715  0.034508  0.034833  0.036930  0.036242  efectores
2    0.024213  0.031380  0.038341  0.041568  0.023359  0.019632  efectores
3    0.020484  0.041331  0.043901  0.024729  0.031584  0.038895  efectores
4    0.047734 -0.000446  0.013668  0.022338  0.014913  0.012400  efectores
..        …         …         …         …         …         …
495  0.012664 -0.002364 -0.011041  0.040406 -0.017979  0.020584  efectores
496  0.043152 -0.044161  0.019277 -0.042280  0.038663 -0.022907  efectores
497  0.037853  0.037317  0.037708  0.032624  0.034483  0.042283  efectores
498  0.055575  0.006406  0.023172  0.033382  0.013518  0.021960  efectores
499  0.009529  0.038892  0.026477  0.000415  0.035705  0.026059  efectores

[395 rows x 42 columns]
```

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

```
                X0          X1          X2          X3          X4          X5  \
count  395.000000  395.000000  395.000000  395.000000  395.000000  395.000000
mean     0.046444    0.003422    0.031075    0.039071    0.021554    0.038252
std      0.014068    0.004909    0.019946    0.033033    0.013618    0.012057
min      0.008219    0.000000    0.002692    0.000000    0.000000    0.012564
25%      0.037293    0.000000    0.016144    0.013268    0.011552    0.029286
50%      0.044458    0.001442    0.025531    0.028948    0.018769    0.036624
75%      0.055577    0.005504    0.042777    0.058550    0.029177    0.045511
max      0.092440    0.029316    0.104747    0.163668    0.066235    0.082794

                X6          X7          X8          X9  …         X31  \
count  395.000000  395.000000  395.000000  395.000000  …  395.000000
mean     0.009541    0.032458    0.024492    0.059798  …    0.020092
std      0.007636    0.028966    0.031572    0.028173  …    0.019492
min      0.000000    0.000000    0.000000    0.002487  …   -0.049465
25%      0.003346    0.011240    0.002892    0.039561  …    0.008788
50%      0.008151    0.022074    0.008679    0.053910  …    0.022823
75%      0.014041    0.045887    0.037268    0.080125  …    0.034144
max      0.039597    0.139331    0.163489    0.156480  …    0.080460

               X32         X33         X34         X35         X36         X37  \
count  395.000000  395.000000  395.000000  395.000000  395.000000  395.000000
mean     0.025583    0.018582    0.021416    0.021181    0.021230    0.022519
std      0.021512    0.020757    0.021271    0.021856    0.021899    0.020678
min     -0.055270   -0.055095   -0.057432   -0.061086   -0.051289   -0.048461
25%      0.015137    0.007539    0.009215    0.009871    0.008491    0.012612
50%      0.027354    0.022533    0.024629    0.024213    0.025820    0.023953
75%      0.038632    0.034117    0.036234    0.036708    0.036049    0.035388
```

```
max      0.103634    0.058736    0.080708    0.086181    0.092905    0.091768

              X38         X39         X40
count  395.000000  395.000000  395.000000
mean     0.022193    0.023824    0.022128
std      0.019998    0.020894    0.021496
min     -0.057145   -0.058423   -0.060518
25%      0.010546    0.012528    0.010748
50%      0.025829    0.027131    0.025355
75%      0.035296    0.036834    0.035558
max      0.078497    0.086239    0.095045

[8 rows x 41 columns]
```

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

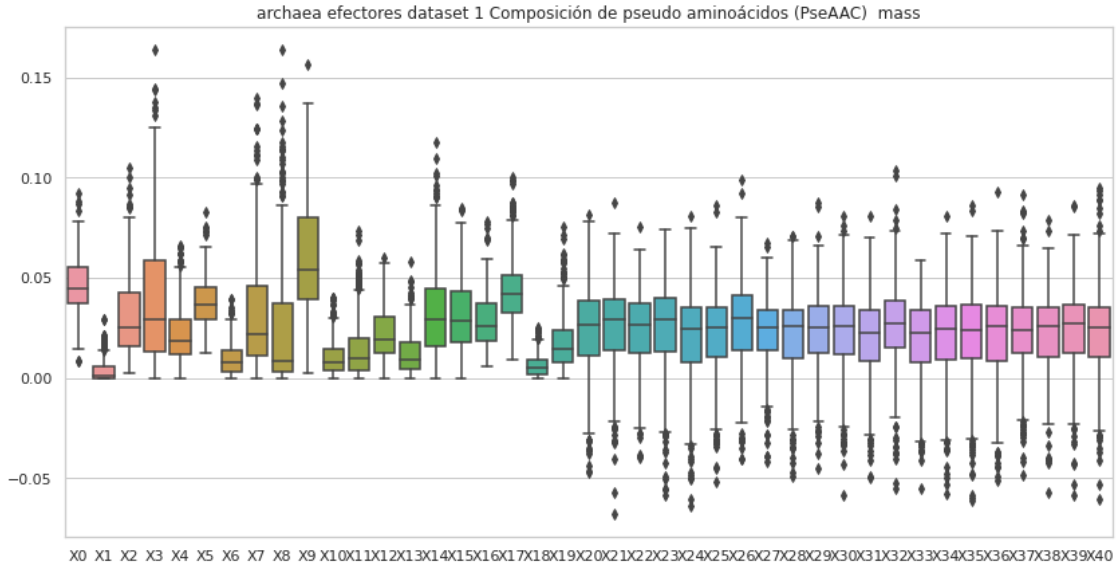
```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.052094  0.000000  0.047679  0.042381  0.031786  0.046796  0.006181
1    0.035063  0.006098  0.004573  0.009909  0.006860  0.024391  0.002287
2    0.023799  0.000000  0.019832  0.019832  0.003966  0.035698  0.003966
4    0.046059  0.008188  0.040941  0.039918  0.011259  0.036847  0.007165
5    0.060500  0.005500  0.038500  0.049500  0.019250  0.052250  0.035750
..        …         …         …         …         …         …         …
494  0.060212  0.004460  0.037911  0.033451  0.013380  0.031221  0.004460
495  0.049636  0.004512  0.040612  0.063174  0.009025  0.036099  0.000000
496  0.033123  0.006973  0.050556  0.036609  0.012203  0.036609  0.008717
497  0.044731  0.000000  0.027526  0.018924  0.046451  0.053333  0.005161
498  0.061898  0.006190  0.086658  0.068088  0.037139  0.068088  0.024759

           X7        X8        X9  …       X32       X33       X34  \
0    0.033552  0.008829  0.056509  …  0.017368  0.010598  0.026206
1    0.015245  0.004573  0.041923  …  0.030496  0.032778  0.032671
2    0.000000  0.003966  0.027765  …  0.016058  0.033871  0.054316
4    0.018424  0.003071  0.051177  … -0.000582  0.025685  0.023064
5    0.085250  0.044000  0.074250  …  0.019717  0.023049  0.019405
..        …         …         …  …  …       …         …         …
494  0.008920  0.004460  0.033451  …  0.027258  0.011205  0.000848
495  0.013537  0.027074  0.036099  …  0.011303  0.018066  0.049685
496  0.015690  0.000000  0.019176  …  0.046431  0.030586  0.024070
497  0.048171  0.006882  0.086020  …  0.007606  0.035483 -0.009811
498  0.055709  0.037139  0.061898  … -0.003781  0.083121 -0.076076

           X35       X36       X37       X38       X39       X40           X41
0    0.013712  0.000681  0.025850  0.022399  0.017500  0.036025  no_efectores
```

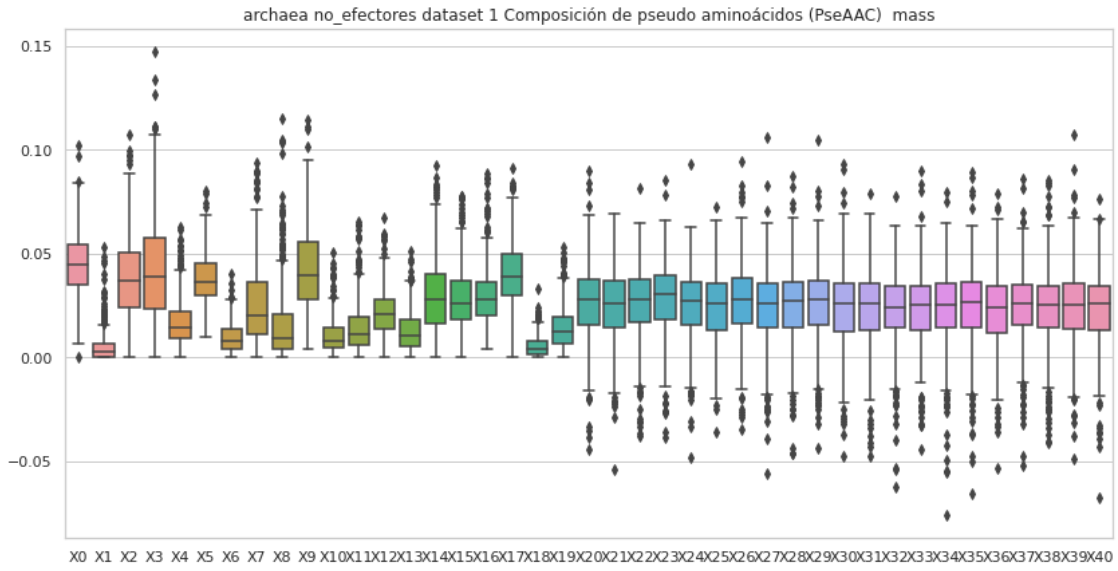
```
1    0.034955  0.033864  0.042843  0.031047  0.024756  0.040108  no_efectores
2    0.012008  0.049439  0.029778  0.035296  0.032032  0.037231  no_efectores
4    0.029641  0.024518  0.026154  0.031865  0.015896  0.038614  no_efectores
5   -0.011808  0.016239  0.005274  0.014721  0.014526  0.013419  no_efectores
..        …         …         …         …         …         …
494  0.012349  0.022980  0.009252  0.008150  0.007099  0.017398  no_efectores
495  0.013694  0.024024  0.020520  0.029666  0.055514  0.002067  no_efectores
496  0.034594  0.041787  0.034197  0.027866  0.007299  0.016583  no_efectores
497 -0.002252  0.014663  0.019280  0.025237  0.006215  0.026987  no_efectores
498  0.062356  0.002124  0.028225 -0.021185  0.067245 -0.067673  no_efectores

[416 rows x 42 columns]
```

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

```
                X0          X1          X2          X3          X4          X5  \
count  416.000000  416.000000  416.000000  416.000000  416.000000  416.000000
mean     0.044974    0.005186    0.038000    0.041785    0.017084    0.037590
std      0.015020    0.007619    0.020899    0.025552    0.011630    0.011725
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.009761
25%      0.034751    0.000000    0.023766    0.023363    0.008902    0.029590
50%      0.044408    0.002893    0.036807    0.038711    0.014108    0.036019
75%      0.054630    0.006376    0.050314    0.057539    0.022280    0.045377
max      0.102168    0.053009    0.107277    0.147215    0.062478    0.080325

                X6          X7          X8          X9  …         X31  \
count  416.000000  416.000000  416.000000  416.000000  …  416.000000
mean     0.009323    0.025231    0.016010    0.042938  …    0.023754
std      0.007116    0.019008    0.019087    0.020262  …    0.018619
min      0.000000    0.000000    0.000000    0.003729  …   -0.047706
25%      0.004045    0.010850    0.003724    0.027930  …    0.012875
50%      0.007835    0.019921    0.008979    0.039215  …    0.025650
75%      0.013722    0.035940    0.020957    0.055475  …    0.035892
max      0.040162    0.093679    0.115004    0.114431  …    0.079022

               X32         X33         X34         X35         X36         X37  \
count  416.000000  416.000000  416.000000  416.000000  416.000000  416.000000
mean     0.022488    0.023228    0.023659    0.024293    0.022655    0.024538
std      0.017742    0.017845    0.019098    0.019193    0.018298    0.017350
min     -0.062647   -0.044167   -0.076076   -0.066028   -0.053305   -0.052401
25%      0.014387    0.013184    0.014602    0.014595    0.011872    0.015507
50%      0.024166    0.025171    0.025538    0.026373    0.024256    0.026140
75%      0.034486    0.034161    0.035456    0.036405    0.034237    0.034747
max      0.077808    0.089553    0.079695    0.088935    0.079060    0.085812
```

```
              X38          X39          X40
count  416.000000   416.000000   416.000000
mean     0.023573     0.024672     0.023254
std      0.019105     0.018753     0.018633
min     -0.041083    -0.049052    -0.067673
25%      0.014092     0.013755     0.012844
50%      0.024968     0.025535     0.025904
75%      0.034293     0.035771     0.034263
max      0.085278     0.107379     0.076332

[8 rows x 41 columns]
```





5 Composición de pseudo aminoácidos (PseAAC) hidro

[9]:
```
#hidro
transf = "Composición de pseudo aminoácidos (PseAAC) "
transf2 = "PseAAC"
estado = "con valores atípicos.\n"
comp = "hidro"
df=""

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

    if etiq == "efectores":
        df=PseAAC_hidro_efec

    if etiq == "no_efectores":
        df=PseAAC_hidro_no_efec

    #del df['X62']
    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")
```

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

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.069438	0.001694	0.030485	0.055889	0.013549	0.059277	0.013549	
1	0.087569	0.001622	0.042163	0.037298	0.017838	0.115137	0.004054	
2	0.043220	0.000000	0.003162	0.010541	0.010541	0.024245	0.003162	
3	0.076519	0.009565	0.051013	0.066954	0.025506	0.082896	0.006377	
4	0.063428	0.003524	0.014095	0.091619	0.021143	0.049333	0.010571	
..	…	…	…	…	…	…	…	
495	0.065296	0.000000	0.037312	0.046640	0.046640	0.046640	0.018656	
496	0.080817	0.010776	0.070042	0.091593	0.032327	0.048490	0.010776	
497	0.044187	0.000000	0.011924	0.004910	0.011222	0.030160	0.002104	
498	0.023915	0.000000	0.023915	0.014717	0.020236	0.036793	0.016557	
499	0.080488	0.000000	0.049332	0.054524	0.023368	0.049332	0.007789	

	X7	X8	X9	…	X53	X54	X55	\
0	0.038953	0.023711	0.037260	…	0.012969	-0.005957	-0.004929	
1	0.010541	0.006487	0.021892	…	0.020820	-0.003128	0.014765	
2	0.013704	0.001054	0.035841	…	-0.002629	0.013416	-0.003523	
3	0.051013	0.035071	0.066954	…	-0.004249	0.017967	0.016112	
4	0.042286	0.038762	0.074000	…	0.016835	0.024520	0.015157	
..	…	…	…	…	…	…	…	
495	0.074624	0.093279	0.055968	…	-0.030378	0.148285	0.098025	
496	0.107756	0.123920	0.145471	…	0.056831	0.043637	0.023741	
497	0.009118	0.002104	0.039278	…	-0.000954	0.030366	-0.002602	
498	0.016557	0.005519	0.051510	…	-0.003241	0.010163	0.000394	
499	0.049332	0.015578	0.090874	…	0.040134	0.000530	-0.011398	

	X56	X57	X58	X59	X60	X61	X62
0	-0.012751	0.001617	0.031186	0.038512	0.021563	0.024171	efectores
1	0.003316	0.015782	-0.011952	0.021350	-0.010761	0.021915	efectores
2	0.031924	0.006432	0.016103	0.003248	0.018553	-0.001516	efectores
3	0.047930	0.023091	-0.005473	-0.006516	-0.030902	-0.018131	efectores
4	0.002312	0.022701	0.020270	0.017485	-0.008712	-0.027923	efectores
..	…	…	…	…	…	…	
495	-0.003893	0.009343	-0.047910	0.025616	0.027014	0.003607	efectores
496	0.049880	0.117219	-0.001641	0.007228	-0.093064	-0.075301	efectores
497	0.016396	-0.004534	0.043874	0.004271	0.022047	-0.006938	efectores

```
498  0.029490  0.016957  0.015829 -0.005679  0.028763  0.018594  efectores
499 -0.018875  0.009972  0.005625  0.033459  0.057150  0.054492  efectores

[500 rows x 63 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.048892	0.004747	0.036183	0.047386	0.023045	0.041251	
std	0.027310	0.007424	0.028180	0.040838	0.022616	0.027524	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.029168	0.000000	0.013543	0.012240	0.011403	0.025187	
50%	0.043022	0.001722	0.030205	0.046906	0.018278	0.034712	
75%	0.063441	0.006397	0.053268	0.067960	0.028401	0.050794	
max	0.184862	0.040665	0.323508	0.438985	0.369724	0.323508	

	X6	X7	X8	X9	…	X52	\
count	500.000000	500.000000	500.000000	500.000000	…	500.000000	
mean	0.010893	0.037671	0.033665	0.065089	…	0.006174	
std	0.010891	0.043222	0.059472	0.051996	…	0.047742	
min	0.000000	0.000000	0.000000	0.002977	…	-0.630490	
25%	0.003824	0.012931	0.004187	0.039273	…	-0.005510	
50%	0.009313	0.025516	0.013586	0.054798	…	0.011953	
75%	0.013998	0.049533	0.049423	0.076167	…	0.024795	
max	0.138646	0.647016	0.970524	0.739447	…	0.230898	

	X53	X54	X55	X56	X57	X58	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.008647	0.012912	0.012843	0.006844	0.011549	0.008997	
std	0.040278	0.044652	0.043641	0.040766	0.035213	0.036432	
min	-0.573346	-0.651246	-0.653968	-0.464862	-0.271387	-0.351358	
25%	-0.001643	-0.001762	-0.000750	-0.009000	-0.003276	-0.004969	
50%	0.008165	0.015944	0.010279	0.012379	0.007781	0.016011	
75%	0.022022	0.026971	0.028323	0.025675	0.024665	0.029310	
max	0.213985	0.213945	0.184566	0.196478	0.279800	0.159180	

	X59	X60	X61
count	500.000000	500.000000	500.000000
mean	0.012703	0.005561	0.009955
std	0.034146	0.047379	0.038586
min	-0.319247	-0.557559	-0.340982
25%	0.000522	-0.009728	-0.001127
50%	0.010290	0.014081	0.008491
75%	0.025373	0.029442	0.025769

```
max      0.111210    0.138515    0.127106

[8 rows x 62 columns]
```

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.043764	0.000000	0.040055	0.035605	0.026704	0.039314	0.005192	
1	0.039024	0.006787	0.005090	0.011029	0.007635	0.027147	0.002545	
2	0.050179	0.000000	0.041816	0.041816	0.008363	0.075269	0.008363	
3	0.038672	0.038672	0.030938	0.030938	0.023203	0.046406	0.000000	
4	0.062359	0.011086	0.055431	0.054045	0.015243	0.049888	0.009700	
..	…	…	…	…	…	…	…	
495	0.062920	0.005720	0.051480	0.080080	0.011440	0.045760	0.000000	
496	0.059115	0.012445	0.090229	0.065338	0.021779	0.065338	0.015557	
497	0.016656	0.000000	0.010250	0.007047	0.017296	0.019859	0.001922	
498	0.043558	0.004356	0.060981	0.047914	0.026135	0.047914	0.017423	
499	0.067363	0.000000	0.038493	0.076986	0.019247	0.038493	0.019247	

	X7	X8	X9	…	X53	X54	X55	\
0	0.028187	0.007418	0.047473	…	0.004143	-0.004235	-0.003340	
1	0.016967	0.005090	0.046659	…	0.004071	0.035353	0.010508	
2	0.000000	0.008363	0.058542	…	0.033335	0.008621	-0.003631	
3	0.007734	0.061875	0.007734	…	-0.025246	0.007121	-0.040979	
4	0.024944	0.004157	0.069288	…	0.022638	-0.014881	0.012370	
..	…	…	…	…	…	…	…	
495	0.017160	0.034320	0.045760	…	0.021875	-0.045565	-0.015963	
496	0.028002	0.000000	0.034225	…	-0.023184	-0.015434	-0.020793	
497	0.017937	0.002562	0.032030	…	-0.000145	0.017170	-0.002314	
498	0.039202	0.026135	0.043558	…	-0.003410	0.000247	-0.036928	
499	0.028870	0.028870	0.067363	…	-0.060167	-0.026713	0.039999	

	X56	X57	X58	X59	X60	X61	X62
0	0.003343	0.004043	0.021755	0.011854	0.026195	0.010425	no_efectores
1	0.016243	-0.000140	0.024851	0.005268	0.024276	0.007448	no_efectores
2	0.040809	-0.003236	0.041316	-0.023708	-0.055328	-0.021780	no_efectores
3	0.006162	0.003980	-0.030200	0.013414	0.020107	0.066217	no_efectores
4	-0.014454	-0.001784	0.010834	0.023035	0.012893	0.029553	no_efectores
..	…	…	…	…	…	…	…
495	-0.016400	0.000592	0.011977	0.034432	0.031079	0.058795	no_efectores
496	0.004629	0.011866	-0.008061	0.030356	0.027707	-0.000321	no_efectores
497	0.018786	0.003745	0.010635	-0.002572	0.021756	-0.003884	no_efectores
498	0.018556	0.061359	-0.045088	-0.016000	-0.048090	-0.014860	no_efectores

```
499 -0.040360  0.011488  0.032417  0.092863  0.074494  0.093006  no_efectores

[500 rows x 63 columns]
```

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

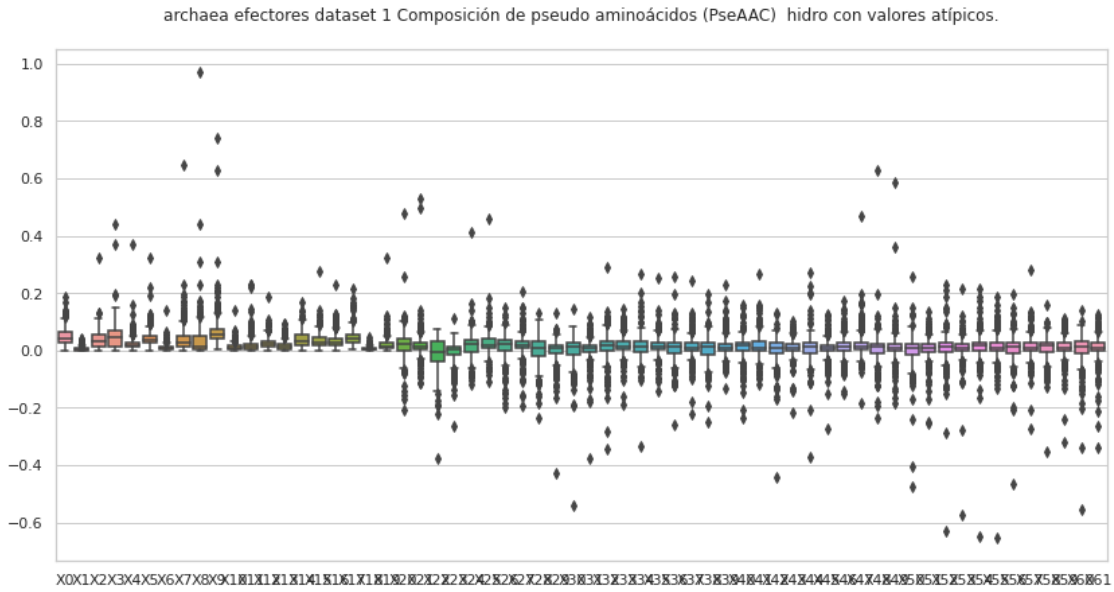
	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.059747	0.009072	0.050364	0.055593	0.021826	0.050979	
std	0.035602	0.016863	0.030248	0.032407	0.017039	0.033563	
min	-0.062334	0.000000	-0.041556	-0.020778	-0.041556	-0.020778	
25%	0.037093	0.000000	0.031642	0.035316	0.011409	0.031664	
50%	0.054737	0.004131	0.051694	0.056133	0.018239	0.046665	
75%	0.078393	0.009652	0.066393	0.071671	0.027256	0.063268	
max	0.339785	0.165266	0.254839	0.339785	0.158213	0.445555	

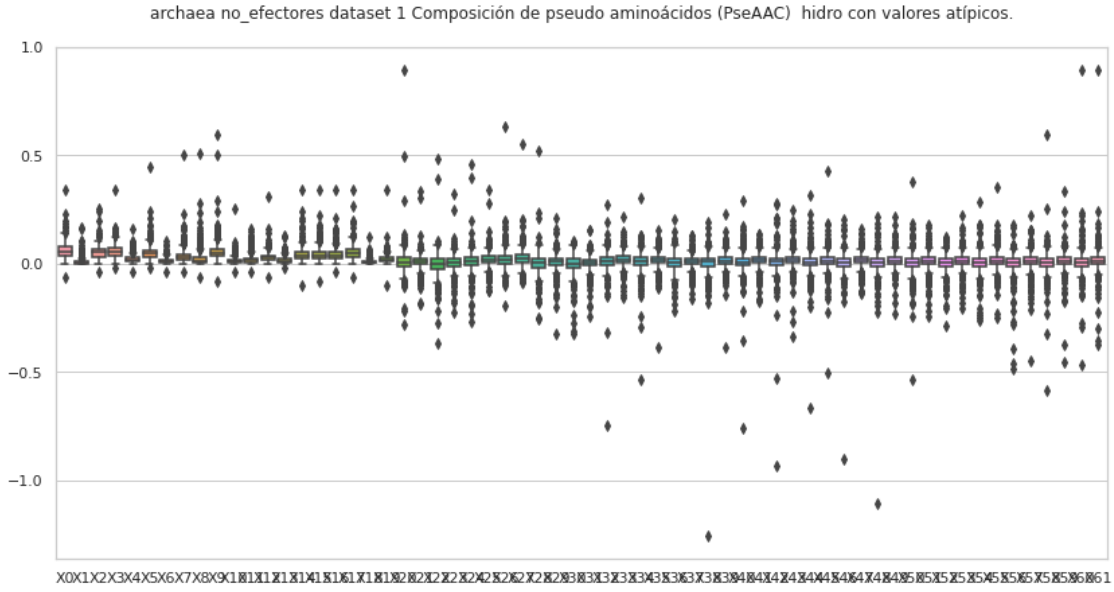
	X6	X7	X8	X9	…	X52	\
count	500.000000	500.000000	500.000000	500.000000	…	500.000000	
mean	0.013249	0.034983	0.024726	0.057641	…	0.002854	
std	0.012843	0.036534	0.038602	0.045278	…	0.038699	
min	-0.041556	-0.041556	-0.062334	-0.083112	…	-0.287555	
25%	0.004751	0.014149	0.005174	0.036507	…	-0.012016	
50%	0.010446	0.026062	0.012752	0.049915	…	0.006979	
75%	0.018976	0.043675	0.028831	0.067559	…	0.022934	
max	0.103566	0.501250	0.509677	0.594624	…	0.124981	

	X53	X54	X55	X56	X57	X58	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.012801	0.000585	0.014206	-0.001586	0.008633	0.003539	
std	0.037173	0.044374	0.041738	0.052615	0.044471	0.054601	
min	-0.207764	-0.264968	-0.252713	-0.487888	-0.447104	-0.586933	
25%	-0.002426	-0.015794	-0.002317	-0.013141	-0.003064	-0.012348	
50%	0.012047	0.003447	0.012320	0.004628	0.010997	0.006921	
75%	0.028612	0.020557	0.031496	0.020476	0.028129	0.024080	
max	0.220281	0.285661	0.353078	0.179547	0.222972	0.592497	

	X59	X60	X61
count	500.000000	500.000000	500.000000
mean	0.011492	0.003225	0.011325
std	0.046030	0.062349	0.060448
min	-0.457156	-0.465559	-0.372111
25%	-0.002104	-0.014089	-0.004666
50%	0.011363	0.004256	0.010748
75%	0.029360	0.019812	0.027307
max	0.333814	0.890842	0.892678

[8 rows x 62 columns]





5.1 Composición de pseudo aminoácidos (PseAAC) hidro, sin valores atípicos

[10]:
```
#hidro
transf = "Composición de pseudo aminoácidos (PseAAC) "
transf2 = "PseAAC"
estado = "sin valores atípicos.\n"
comp = "hidro"
df=""

out = (str(r3) + '/ds' + str(dataset) + '_' + str(transf2) + '_' + str(comp) +␣
↪'_' + str(organismo) + '.csv')
os.makedirs(str(r3), exist_ok=True)
df_out = pd.DataFrame()


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

    if etiq == "efectores":
        df=PseAAC_hidro_efec

    if etiq == "no_efectores":
        df=PseAAC_hidro_no_efec

    del df['X62']
    #Se eliminan todas las filas que tengan valores atípicos en al menos una de␣
↪sus columnas.
    df = (df[(np.abs(stats.zscore(df)) < 3).all(axis=1)])
    df['X62'] = etiq
    df_out = pd.concat([df_out,df])

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


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

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

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

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.069438	0.001694	0.030485	0.055889	0.013549	0.059277	0.013549	
1	0.087569	0.001622	0.042163	0.037298	0.017838	0.115137	0.004054	
2	0.043220	0.000000	0.003162	0.010541	0.010541	0.024245	0.003162	
3	0.076519	0.009565	0.051013	0.066954	0.025506	0.082896	0.006377	
4	0.063428	0.003524	0.014095	0.091619	0.021143	0.049333	0.010571	
..	…	…	…	…	…	…	…	
491	0.024908	0.000000	0.007866	0.005244	0.009177	0.022286	0.009177	
492	0.021760	0.000000	0.011375	0.017309	0.011375	0.015826	0.000989	
493	0.007446	0.012410	0.017374	0.049640	0.002482	0.022338	0.009928	
497	0.044187	0.000000	0.011924	0.004910	0.011222	0.030160	0.002104	
498	0.023915	0.000000	0.023915	0.014717	0.020236	0.036793	0.016557	

	X7	X8	X9	…	X53	X54	X55	\
0	0.038953	0.023711	0.037260	…	0.012969	-0.005957	-0.004929	
1	0.010541	0.006487	0.021892	…	0.020820	-0.003128	0.014765	
2	0.013704	0.001054	0.035841	…	-0.002629	0.013416	-0.003523	
3	0.051013	0.035071	0.066954	…	-0.004249	0.017967	0.016112	
4	0.042286	0.038762	0.074000	…	0.016835	0.024520	0.015157	
..	…	…	…	…	…	…	…	
491	0.007866	0.006555	0.043261	…	0.000897	0.020053	0.007943	
492	0.011375	0.002473	0.024728	…	0.018145	0.016254	0.016734	
493	0.019856	0.034748	0.039712	…	0.035442	-0.001923	0.002557	
497	0.009118	0.002104	0.039278	…	-0.000954	0.030366	-0.002602	
498	0.016557	0.005519	0.051510	…	-0.003241	0.010163	0.000394	

	X56	X57	X58	X59	X60	X61	X62
0	-0.012751	0.001617	0.031186	0.038512	0.021563	0.024171	efectores
1	0.003316	0.015782	-0.011952	0.021350	-0.010761	0.021915	efectores
2	0.031924	0.006432	0.016103	0.003248	0.018553	-0.001516	efectores
3	0.047930	0.023091	-0.005473	-0.006516	-0.030902	-0.018131	efectores
4	0.002312	0.022701	0.020270	0.017485	-0.008712	-0.027923	efectores
..	…	…	…	…	…	…	…
491	0.019287	0.004001	0.033325	0.014066	0.029523	0.007832	efectores
492	0.012784	0.015852	0.016907	0.012660	0.016414	0.014043	efectores
493	0.037481	0.028468	-0.016989	0.014986	-0.018453	0.016648	efectores
497	0.016396	-0.004534	0.043874	0.004271	0.022047	-0.006938	efectores
498	0.029490	0.016957	0.015829	-0.005679	0.028763	0.018594	efectores

```
[428 rows x 63 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	428.000000	428.000000	428.000000	428.000000	428.000000	428.000000	
mean	0.046524	0.003356	0.031595	0.039279	0.020109	0.037954	
std	0.024579	0.005118	0.022080	0.028799	0.012223	0.020795	
min	0.002144	0.000000	0.002144	0.000000	0.000000	0.000000	
25%	0.028403	0.000000	0.012289	0.010288	0.011104	0.024583	
50%	0.041728	0.001158	0.025639	0.037627	0.017404	0.032456	
75%	0.060242	0.004862	0.048027	0.062816	0.025676	0.048175	
max	0.128860	0.026914	0.100168	0.121639	0.074520	0.123372	

	X6	X7	X8	X9	…	X52	\
count	428.000000	428.000000	428.000000	428.000000	…	428.000000	
mean	0.009542	0.029688	0.024007	0.055517	…	0.009156	
std	0.007699	0.024219	0.028874	0.023906	…	0.024429	
min	0.000000	0.000000	0.000000	0.002977	…	-0.121708	
25%	0.003705	0.011981	0.003276	0.037652	…	-0.001833	
50%	0.008700	0.021847	0.010487	0.049865	…	0.013826	
75%	0.013041	0.041405	0.036227	0.070183	…	0.024352	
max	0.041159	0.139103	0.169948	0.164637	…	0.094751	

	X53	X54	X55	X56	X57	X58	\
count	428.000000	428.000000	428.000000	428.000000	428.000000	428.000000	
mean	0.009946	0.014484	0.014314	0.009235	0.010741	0.011182	
std	0.020239	0.023789	0.024085	0.024796	0.022686	0.026420	
min	-0.072709	-0.072995	-0.088642	-0.079071	-0.060849	-0.099795	
25%	-0.000755	0.000347	0.000144	-0.004881	-0.002221	-0.001763	
50%	0.007717	0.016218	0.010106	0.012941	0.006509	0.017441	
75%	0.019279	0.026636	0.023966	0.024735	0.020580	0.028802	
max	0.103240	0.109616	0.119773	0.100169	0.111690	0.096988	

	X59	X60	X61
count	428.000000	428.000000	428.000000
mean	0.013945	0.011798	0.012453
std	0.023409	0.027830	0.023681
min	-0.065751	-0.129204	-0.099220
25%	0.001234	-0.005997	-0.000410
50%	0.010110	0.016461	0.008491
75%	0.023875	0.029586	0.022848
max	0.111210	0.081847	0.115300

[8 rows x 62 columns]

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

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.043764  0.000000  0.040055  0.035605  0.026704  0.039314  0.005192
1    0.039024  0.006787  0.005090  0.011029  0.007635  0.027147  0.002545
2    0.050179  0.000000  0.041816  0.041816  0.008363  0.075269  0.008363
3    0.038672  0.038672  0.030938  0.030938  0.023203  0.046406  0.000000
4    0.062359  0.011086  0.055431  0.054045  0.015243  0.049888  0.009700
..        …         …         …         …         …         …
495  0.062920  0.005720  0.051480  0.080080  0.011440  0.045760  0.000000
496  0.059115  0.012445  0.090229  0.065338  0.021779  0.065338  0.015557
497  0.016656  0.000000  0.010250  0.007047  0.017296  0.019859  0.001922
498  0.043558  0.004356  0.060981  0.047914  0.026135  0.047914  0.017423
499  0.067363  0.000000  0.038493  0.076986  0.019247  0.038493  0.019247

           X7        X8        X9  …       X53       X54       X55  \
0    0.028187  0.007418  0.047473  …  0.004143 -0.004235 -0.003340
1    0.016967  0.005090  0.046659  …  0.004071  0.035353  0.010508
2    0.000000  0.008363  0.058542  …  0.033335  0.008621 -0.003631
3    0.007734  0.061875  0.007734  … -0.025246  0.007121 -0.040979
4    0.024944  0.004157  0.069288  …  0.022638 -0.014881  0.012370
..        …         …         …  …         …         …
495  0.017160  0.034320  0.045760  …  0.021875 -0.045565 -0.015963
496  0.028002  0.000000  0.034225  … -0.023184 -0.015434 -0.020793
497  0.017937  0.002562  0.032030  … -0.000145  0.017170 -0.002314
498  0.039202  0.026135  0.043558  … -0.003410  0.000247 -0.036928
499  0.028870  0.028870  0.067363  … -0.060167 -0.026713  0.039999

          X56       X57       X58       X59       X60       X61           X62
0    0.003343  0.004043  0.021755  0.011854  0.026195  0.010425  no_efectores
1    0.016243 -0.000140  0.024851  0.005268  0.024276  0.007448  no_efectores
2    0.040809 -0.003236  0.041316 -0.023708 -0.055328 -0.021780  no_efectores
3    0.006162  0.003980 -0.030200  0.013414  0.020107  0.066217  no_efectores
4   -0.014454 -0.001784  0.010834  0.023035  0.012893  0.029553  no_efectores
..        …         …         …         …         …         …
495 -0.016400  0.000592  0.011977  0.034432  0.031079  0.058795  no_efectores
496  0.004629  0.011866 -0.008061  0.030356  0.027707 -0.000321  no_efectores
497  0.018786  0.003745  0.010635 -0.002572  0.021756 -0.003884  no_efectores
498  0.018556  0.061359 -0.045088 -0.016000 -0.048090 -0.014860  no_efectores
499 -0.040360  0.011488  0.032417  0.092863  0.074494  0.093006  no_efectores

[435 rows x 63 columns]
```

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

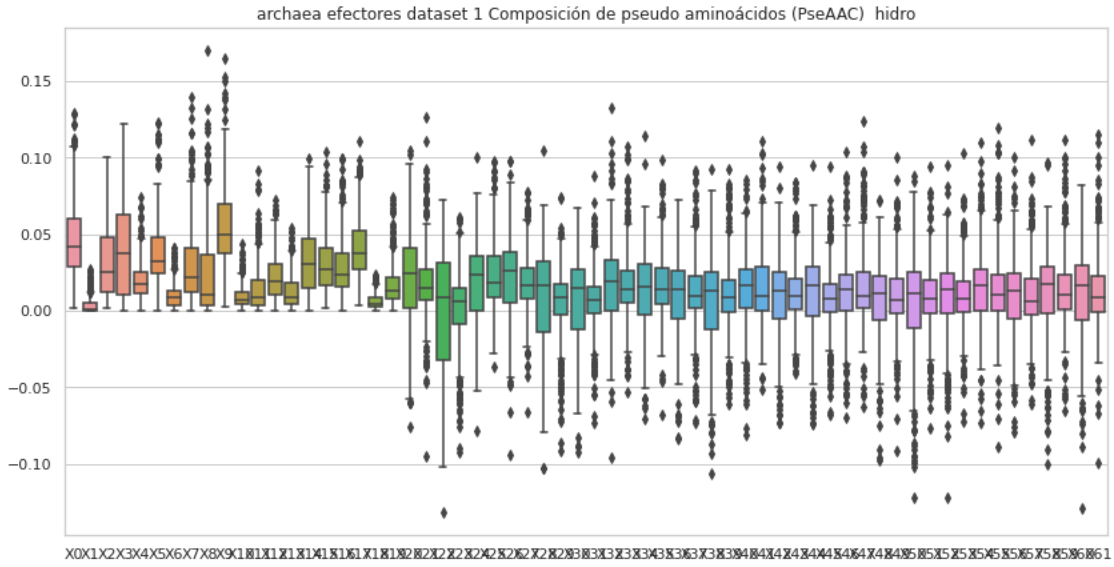
	X0	X1	X2	X3	X4	X5	\
count	435.000000	435.000000	435.000000	435.000000	435.000000	435.000000	
mean	0.056354	0.006286	0.047078	0.050605	0.019635	0.046653	
std	0.028133	0.008906	0.024587	0.025768	0.011960	0.021965	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.035997	0.000000	0.029344	0.032138	0.011370	0.030674	
50%	0.053150	0.003570	0.050574	0.053794	0.017846	0.044633	
75%	0.074700	0.008094	0.063935	0.067425	0.025865	0.059932	
max	0.141205	0.052767	0.100860	0.122862	0.072629	0.116009	

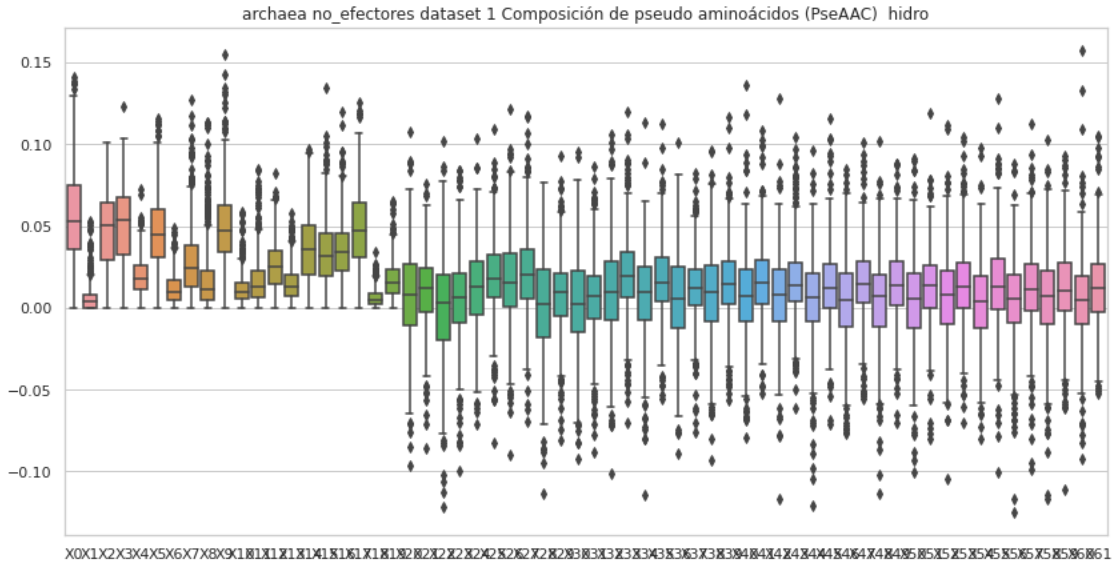
	X6	X7	X8	X9	…	X52	\
count	435.000000	435.000000	435.000000	435.000000	…	435.000000	
mean	0.011810	0.029032	0.018809	0.050523	…	0.006440	
std	0.009633	0.021955	0.021413	0.023662	…	0.026605	
min	0.000000	0.000000	0.000000	0.000000	…	-0.104690	
25%	0.004626	0.013184	0.004879	0.034452	…	-0.009679	
50%	0.009636	0.024238	0.011289	0.047432	…	0.007971	
75%	0.017262	0.038345	0.023111	0.062992	…	0.022754	
max	0.049222	0.127173	0.113185	0.154688	…	0.112030	

	X53	X54	X55	X56	X57	X58	\
count	435.000000	435.000000	435.000000	435.000000	435.000000	435.000000	
mean	0.014837	0.003396	0.014337	0.004336	0.012049	0.004750	
std	0.024686	0.025493	0.027087	0.026471	0.026593	0.028413	
min	-0.070563	-0.080283	-0.100908	-0.124997	-0.099163	-0.116808	
25%	-0.000195	-0.012177	-0.000941	-0.009271	-0.001829	-0.010162	
50%	0.012649	0.004231	0.012589	0.005718	0.011009	0.006918	
75%	0.027635	0.019707	0.029975	0.020334	0.026986	0.023011	
max	0.104521	0.098168	0.128365	0.087592	0.112833	0.102658	

	X59	X60	X61
count	435.000000	435.000000	435.000000
mean	0.012505	0.004002	0.013143
std	0.025452	0.027673	0.024053
min	-0.111161	-0.092313	-0.052723
25%	-0.001664	-0.010152	-0.002564
50%	0.010838	0.005024	0.011965
75%	0.027821	0.019396	0.026434
max	0.092863	0.157283	0.104944

[8 rows x 62 columns]





6 Covarianza de auto cruzamiento (ACC) hidro_mass

[11]:
```
#hidro_mass
transf = "Covarianza de auto cruzamiento (ACC) "
transf2 = "ACC"
estado = "con valores atípicos.\n"
```

```
comp = "hidro_mass"
df=""

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

    if etiq == "efectores":
        df=ACC_hidro_mass_efec

    if etiq == "no_efectores":
        df=ACC_hidro_mass_no_efec

    #del df['X13']
    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")

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

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

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.014333  0.058410 -0.015800  0.050499  0.033702  0.011739  0.033106
1    0.117089  0.022280 -0.010059 -0.055715  0.019496  0.030200  0.194636
2   -0.022653 -0.067049  0.038619 -0.016568 -0.025107  0.036969  0.018184
3   -0.014102 -0.186867 -0.058138  0.028357  0.017154 -0.014069 -0.006509
4   -0.042041 -0.037866 -0.132133  0.067086 -0.011917  0.054880 -0.013365
..        …         …         …         …         …         …         …
495  0.068230  0.040814  0.013470 -0.024109 -0.020386  0.079698  0.291283
496 -0.098494  0.090414 -0.046123  0.031132 -0.029106 -0.015877 -0.005387
497  0.035241  0.074552 -0.051782  0.104516  0.053124  0.030135  0.009295
498 -0.008424 -0.098538  0.183161 -0.018165 -0.110064  0.127099  0.025780
499 -0.089477 -0.068173  0.059083 -0.068624 -0.016302  0.084140 -0.104055

           X7        X8        X9       X10       X11       X12        X13
0    0.066307  0.070958  0.002458 -0.054490  0.074434 -0.009852  efectores
```

```
1    0.181927  0.061440  0.052418 -0.095182  0.017137 -0.008054  efectores
2   -0.011953 -0.012152  0.022458  0.069529 -0.032195  0.065258  efectores
3    0.071194 -0.012853 -0.061057  0.058445  0.040567  0.055157  efectores
4    0.052872  0.046645  0.133642  0.040733 -0.058345 -0.119692  efectores
..        …         …         …         …         …         …
495  0.096959  0.082475 -0.075844 -0.001409  0.104725 -0.054715  efectores
496 -0.014981 -0.070751  0.022246 -0.034248  0.045241 -0.110333  efectores
497  0.018779  0.029106 -0.015120  0.000557 -0.008222 -0.000135  efectores
498 -0.096359  0.058341 -0.034121 -0.117291 -0.031005  0.054966  efectores
499  0.073335  0.010683 -0.070616  0.008012  0.042056  0.006054  efectores

[500 rows x 14 columns]
```

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

```
                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.002939    0.008203    0.007979    0.014099   -0.004549   -0.000841
std      0.070068    0.075840    0.066766    0.070813    0.076525    0.068372
min     -0.222323   -0.339124   -0.222691   -0.187396   -0.310859   -0.237262
25%     -0.040690   -0.034304   -0.041085   -0.033020   -0.046808   -0.042516
50%      0.000953    0.013606    0.007162    0.020809   -0.004059   -0.000808
75%      0.048414    0.056372    0.052288    0.062198    0.043345    0.038743
max      0.280399    0.253401    0.247227    0.212324    0.204496    0.249863

                X6          X7          X8          X9         X10         X11  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.018172    0.003362   -0.002521    0.007883    0.000809   -0.007018
std      0.070236    0.063886    0.073789    0.070119    0.070408    0.061428
min     -0.205625   -0.248853   -0.299462   -0.206259   -0.273953   -0.207756
25%     -0.023447   -0.034456   -0.051159   -0.036106   -0.041130   -0.041249
50%      0.017333    0.005366   -0.002557    0.005596    0.001569   -0.006423
75%      0.058472    0.044924    0.043056    0.048046    0.043449    0.027856
max      0.291283    0.200966    0.252725    0.253159    0.299910    0.216093

               X12
count  500.000000
mean     0.013017
std      0.074332
min     -0.296465
25%     -0.032943
50%      0.011452
75%      0.063365
max      0.244813
```

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

	X0	X1	X2	X3	X4	X5	X6	\
0	-0.007453	-0.004358	0.022996	0.041770	0.031562	-0.001634	-0.016052	
1	0.077597	-0.028794	0.007985	0.041786	0.075589	0.003971	0.009067	
2	-0.169125	0.099419	-0.018895	-0.163102	-0.007650	-0.082962	-0.021851	
3	-0.035957	-0.261776	0.281942	-0.005707	-0.154203	0.256104	0.092575	
4	-0.002941	0.056854	-0.050010	-0.038256	0.026012	-0.029620	0.105664	
..	…	…	…	…	…	…	…	
495	-0.025754	-0.119795	0.025343	-0.108952	0.072790	0.069994	0.075011	
496	0.128380	0.026249	0.073039	0.035285	0.003243	0.071292	0.093109	
497	-0.013349	0.000164	0.040691	-0.050973	0.099268	-0.010383	-0.007907	
498	-0.127940	0.081952	-0.098889	0.008579	-0.022997	0.046067	-0.027528	
499	-0.012926	-0.022720	0.002323	0.037684	-0.114223	-0.004831	-0.146306	

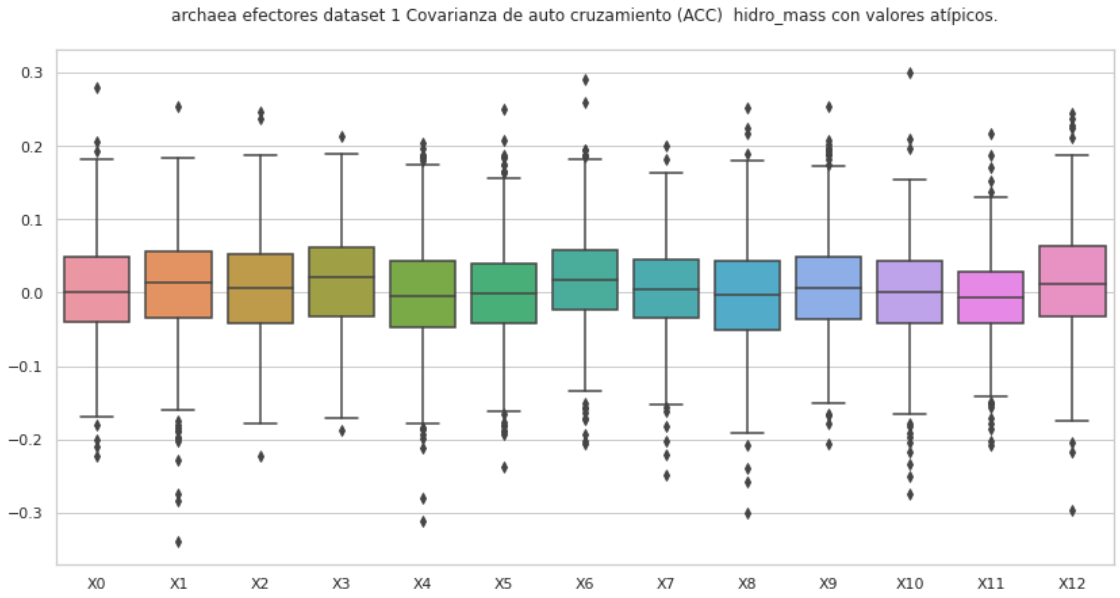
	X7	X8	X9	X10	X11	X12	X13
0	0.011972	0.041875	0.025961	-0.023854	-0.043768	-0.004662	no_efectores
1	-0.033402	-0.011621	0.053565	0.057959	0.039933	-0.026798	no_efectores
2	0.079553	0.016118	0.058210	-0.106449	0.174852	-0.098204	no_efectores
3	-0.294665	0.358922	0.079103	-0.416854	0.036443	0.113940	no_efectores
4	-0.013605	0.021535	0.024658	0.023564	0.033117	-0.112959	no_efectores
..	…	…	…	…	…	…	…
495	0.046176	-0.077435	-0.126524	-0.039933	0.055277	-0.038658	no_efectores
496	-0.009276	0.020455	0.032325	0.070916	-0.005286	0.125299	no_efectores
497	0.030171	0.015544	0.010503	-0.033135	-0.017511	-0.037388	no_efectores
498	-0.148681	-0.075304	-0.079266	0.073382	0.000108	-0.038254	no_efectores
499	-0.189282	0.027878	-0.124328	0.020047	-0.035824	-0.048783	no_efectores

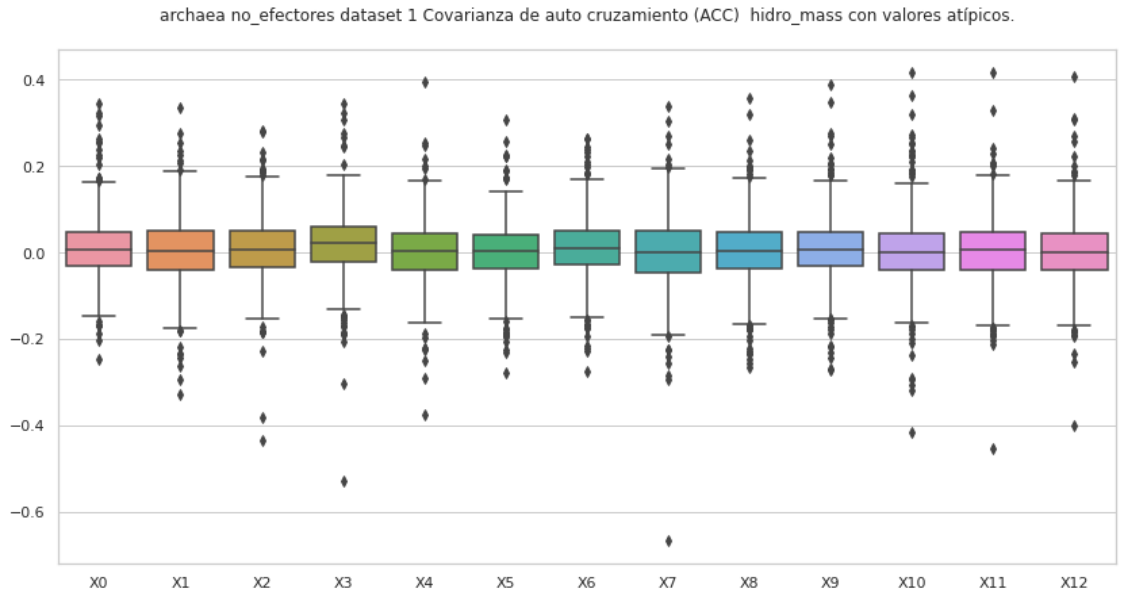
[500 rows x 14 columns]

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

	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.012023	0.005795	0.005676	0.021019	0.001277	-0.000015	
std	0.074929	0.079772	0.076852	0.080765	0.075763	0.071589	
min	-0.248673	-0.327442	-0.434693	-0.530479	-0.375185	-0.277203	
25%	-0.029441	-0.040566	-0.034124	-0.020548	-0.039769	-0.036475	
50%	0.006062	0.004824	0.006565	0.021511	0.002389	0.002121	
75%	0.048774	0.051565	0.050485	0.060987	0.043946	0.042136	

```
max      0.344385    0.334743    0.281942    0.343865    0.395156    0.308510

                X6          X7          X8          X9         X10         X11  \
count   500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean      0.011894    0.001503    0.004246    0.007693    0.003425    0.004518
std       0.072772    0.086517    0.079204    0.078800    0.086653    0.080911
min      -0.276890   -0.667267   -0.267040   -0.273796   -0.416854   -0.453421
25%      -0.028582   -0.046016   -0.038396   -0.031605   -0.039098   -0.041306
50%       0.009486    0.001190    0.003672    0.007697    0.000781    0.005677
75%       0.051764    0.051047    0.047262    0.048624    0.043493    0.047736
max       0.264049    0.337730    0.358922    0.388279    0.416983    0.415995

               X12
count   500.000000
mean      0.000399
std       0.076327
min      -0.400128
25%      -0.041488
50%       0.000645
75%       0.043943
max       0.407235
```





6.1 Covarianza de auto cruzamiento (ACC) hidro_mass, sin valores atípicos

[12]:
```
#hidro_mass
transf = "Covarianza de auto cruzamiento (ACC) "
transf2 = "ACC"
estado = "sin valores atípicos.\n"
comp = "hidro_mass"
df=""

out = (str(r3) + '/ds' + str(dataset) + '_' + str(transf2) + '_' + str(comp) +␣
↪'_' + str(organismo) + '.csv')
os.makedirs(str(r3), exist_ok=True)
df_out = pd.DataFrame()

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

    if etiq == "efectores":
        df=ACC_hidro_mass_efec

    if etiq == "no_efectores":
        df=ACC_hidro_mass_no_efec

    del df['X13']
```

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

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

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

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

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

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.014333  0.058410 -0.015800  0.050499  0.033702  0.011739  0.033106
1    0.117089  0.022280 -0.010059 -0.055715  0.019496  0.030200  0.194636
2   -0.022653 -0.067049  0.038619 -0.016568 -0.025107  0.036969  0.018184
3   -0.014102 -0.186867 -0.058138  0.028357  0.017154 -0.014069 -0.006509
4   -0.042041 -0.037866 -0.132133  0.067086 -0.011917  0.054880 -0.013365
..        …       …       …       …       …       …       …
494 -0.078251  0.022220  0.009384 -0.016942 -0.108257 -0.109375 -0.008845
496 -0.098494  0.090414 -0.046123  0.031132 -0.029106 -0.015877 -0.005387
497  0.035241  0.074552 -0.051782  0.104516  0.053124  0.030135  0.009295
498 -0.008424 -0.098538  0.183161 -0.018165 -0.110064  0.127099  0.025780
499 -0.089477 -0.068173  0.059083 -0.068624 -0.016302  0.084140 -0.104055

           X7        X8        X9       X10       X11       X12        X13
0    0.066307  0.070958  0.002458 -0.054490  0.074434 -0.009852  efectores
1    0.181927  0.061440  0.052418 -0.095182  0.017137 -0.008054  efectores
2   -0.011953 -0.012152  0.022458  0.069529 -0.032195  0.065258  efectores
3    0.071194 -0.012853 -0.061057  0.058445  0.040567  0.055157  efectores
4    0.052872  0.046645  0.133642  0.040733 -0.058345 -0.119692  efectores
..        …       …       …       …       …       …        …
494  0.065117  0.044981  0.032289  0.036142  0.048335 -0.118593  efectores
```

```
496 -0.014981 -0.070751  0.022246 -0.034248  0.045241 -0.110333  efectores
497  0.018779  0.029106 -0.015120  0.000557 -0.008222 -0.000135  efectores
498 -0.096359  0.058341 -0.034121 -0.117291 -0.031005  0.054966  efectores
499  0.073335  0.010683 -0.070616  0.008012  0.042056  0.006054  efectores

[463 rows x 14 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	463.000000	463.000000	463.000000	463.000000	463.000000	463.000000	
mean	0.003403	0.011309	0.008601	0.012042	-0.003815	-0.001904	
std	0.064829	0.068062	0.063867	0.068499	0.072637	0.063039	
min	-0.180330	-0.202726	-0.177774	-0.187396	-0.212280	-0.193211	
25%	-0.037879	-0.030429	-0.036038	-0.034355	-0.045543	-0.040754	
50%	0.001878	0.014133	0.006504	0.015855	-0.003225	-0.002568	
75%	0.047399	0.055251	0.051397	0.059899	0.040696	0.035035	
max	0.206320	0.182761	0.187346	0.212324	0.204496	0.184461	

	X6	X7	X8	X9	X10	X11	\
count	463.000000	463.000000	463.000000	463.000000	463.000000	463.000000	
mean	0.019625	0.004096	-0.001484	0.007492	0.000611	-0.007612	
std	0.062989	0.058850	0.068576	0.067941	0.066034	0.057330	
min	-0.172609	-0.160876	-0.207629	-0.179212	-0.203896	-0.186031	
25%	-0.020101	-0.032215	-0.049952	-0.034980	-0.041943	-0.040109	
50%	0.017638	0.005405	-0.001833	0.005646	0.000557	-0.006214	
75%	0.057948	0.043441	0.042487	0.047208	0.042882	0.026146	
max	0.194636	0.181927	0.217009	0.207579	0.209027	0.170779	

	X12
count	463.000000
mean	0.013592
std	0.069768
min	-0.203880
25%	-0.032722
50%	0.011391
75%	0.062490
max	0.228887

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

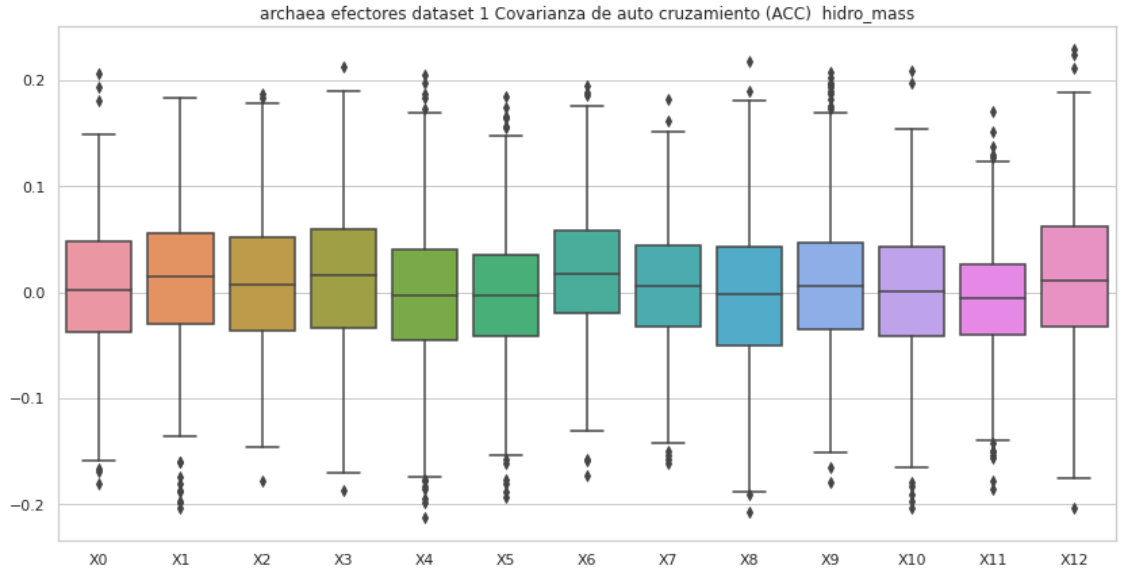
```
           X0        X1        X2        X3        X4        X5        X6  \
0   -0.007453 -0.004358  0.022996  0.041770  0.031562 -0.001634 -0.016052
1    0.077597 -0.028794  0.007985  0.041786  0.075589  0.003971  0.009067
2   -0.169125  0.099419 -0.018895 -0.163102 -0.007650 -0.082962 -0.021851
4   -0.002941  0.056854 -0.050010 -0.038256  0.026012 -0.029620  0.105664
5    0.040365 -0.031385 -0.018865 -0.054045 -0.033356 -0.000632  0.098324
..        …         …         …         …         …         …
495 -0.025754 -0.119795  0.025343 -0.108952  0.072790  0.069994  0.075011
496  0.128380  0.026249  0.073039  0.035285  0.003243  0.071292  0.093109
497 -0.013349  0.000164  0.040691 -0.050973  0.099268 -0.010383 -0.007907
498 -0.127940  0.081952 -0.098889  0.008579 -0.022997  0.046067 -0.027528
499 -0.012926 -0.022720  0.002323  0.037684 -0.114223 -0.004831 -0.146306

           X7        X8        X9       X10       X11       X12           X13
0    0.011972  0.041875  0.025961 -0.023854 -0.043768 -0.004662  no_efectores
1   -0.033402 -0.011621  0.053565  0.057959  0.039933 -0.026798  no_efectores
2    0.079553  0.016118  0.058210 -0.106449  0.174852 -0.098204  no_efectores
4   -0.013605  0.021535  0.024658  0.023564  0.033117 -0.112959  no_efectores
5   -0.031198  0.036242  0.047570  0.000683 -0.030940  0.029588  no_efectores
..        …         …         …         …         …         …
495  0.046176 -0.077435 -0.126524 -0.039933  0.055277 -0.038658  no_efectores
496 -0.009276  0.020455  0.032325  0.070916 -0.005286  0.125299  no_efectores
497  0.030171  0.015544  0.010503 -0.033135 -0.017511 -0.037388  no_efectores
498 -0.148681 -0.075304 -0.079266  0.073382  0.000108 -0.038254  no_efectores
499 -0.189282  0.027878 -0.124328  0.020047 -0.035824 -0.048783  no_efectores

[446 rows x 14 columns]
```

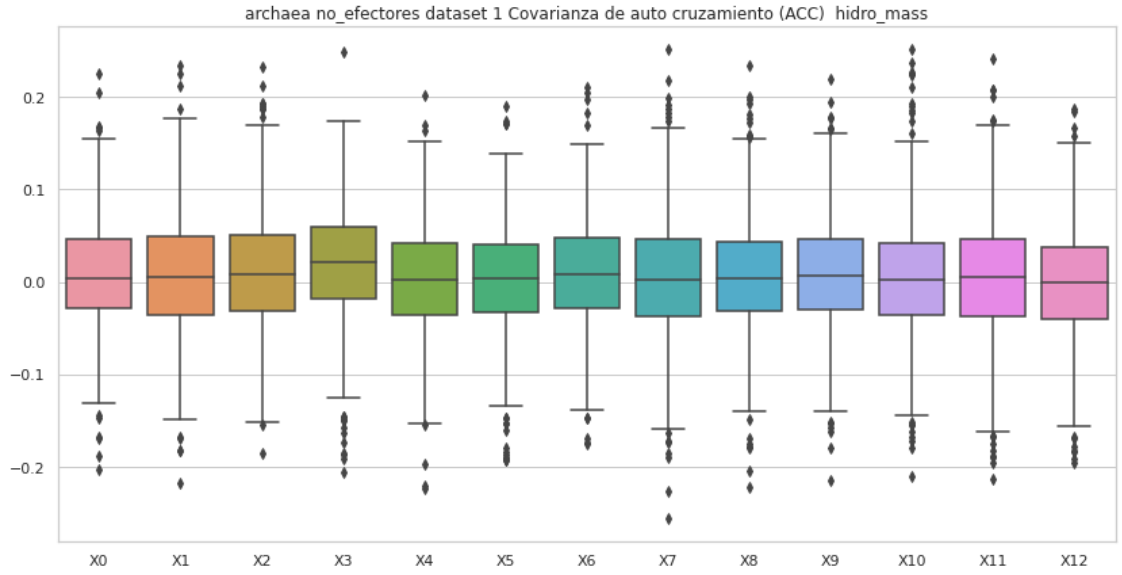
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores archaea dataset 1, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  446.000000  446.000000  446.000000  446.000000  446.000000  446.000000
mean     0.007665    0.007125    0.009014    0.019831    0.002145    0.000505
std      0.062327    0.067033    0.064606    0.066620    0.063105    0.061649
min     -0.202478   -0.217515   -0.184838   -0.206192   -0.223705   -0.193069
25%     -0.028503   -0.036152   -0.030675   -0.018003   -0.036202   -0.033072
50%      0.004339    0.004824    0.007632    0.021236    0.002802    0.003883
75%      0.046747    0.049267    0.050256    0.059178    0.042114    0.041206
max      0.225241    0.234617    0.232387    0.249355    0.201380    0.190655

                X6          X7          X8          X9         X10         X11  \
count  446.000000  446.000000  446.000000  446.000000  446.000000  446.000000
mean     0.009846    0.003437    0.006247    0.006376    0.004872    0.004491
std      0.060037    0.070824    0.065546    0.064467    0.068593    0.069030
```

min	-0.174996	-0.255489	-0.221664	-0.214500	-0.210503	-0.213039
25%	-0.028234	-0.037194	-0.030506	-0.029954	-0.034957	-0.036797
50%	0.008086	0.001933	0.004224	0.007260	0.002456	0.005677
75%	0.047306	0.046775	0.044062	0.046528	0.042001	0.046915
max	0.211027	0.251549	0.234209	0.219879	0.251512	0.241907

	X12
count	446.000000
mean	-0.001569
std	0.062912
min	-0.195438
25%	-0.040161
50%	0.000154
75%	0.037655
max	0.187897





7 Covarianza de auto cruzamiento (ACC) mass

[13]:
```
#mass
transf = "Covarianza de auto cruzamiento (ACC) "
transf2 = "ACC"
estado = "con valores atípicos.\n"
comp = "mass"
df=""

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

    if etiq == "efectores":
        df=ACC_mass_efec

    if etiq == "no_efectores":
        df=ACC_mass_no_efec

    #del df['X13']
    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")
```

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

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

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.014333  0.058410 -0.015800  0.050499  0.033702  0.011739  0.033106
1    0.117089  0.022280 -0.010059 -0.055715  0.019496  0.030200  0.194636
2   -0.022653 -0.067049  0.038619 -0.016568 -0.025107  0.036969  0.018184
3   -0.014102 -0.186867 -0.058138  0.028357  0.017154 -0.014069 -0.006509
4   -0.042041 -0.037866 -0.132133  0.067086 -0.011917  0.054880 -0.013365
..        …         …         …         …         …         …         …
495  0.068230  0.040814  0.013470 -0.024109 -0.020386  0.079698  0.291283
496 -0.098494  0.090414 -0.046123  0.031132 -0.029106 -0.015877 -0.005387
497  0.035241  0.074552 -0.051782  0.104516  0.053124  0.030135  0.009295
498 -0.008424 -0.098538  0.183161 -0.018165 -0.110064  0.127099  0.025780
499 -0.089477 -0.068173  0.059083 -0.068624 -0.016302  0.084140 -0.104055

           X7        X8        X9       X10       X11       X12        X13
0    0.066307  0.070958  0.002458 -0.054490  0.074434 -0.009852  efectores
1    0.181927  0.061440  0.052418 -0.095182  0.017137 -0.008054  efectores
2   -0.011953 -0.012152  0.022458  0.069529 -0.032195  0.065258  efectores
3    0.071194 -0.012853 -0.061057  0.058445  0.040567  0.055157  efectores
4    0.052872  0.046645  0.133642  0.040733 -0.058345 -0.119692  efectores
..        …         …         …         …         …         …         …
495  0.096959  0.082475 -0.075844 -0.001409  0.104725 -0.054715  efectores
496 -0.014981 -0.070751  0.022246 -0.034248  0.045241 -0.110333  efectores
497  0.018779  0.029106 -0.015120  0.000557 -0.008222 -0.000135  efectores
498 -0.096359  0.058341 -0.034121 -0.117291 -0.031005  0.054966  efectores
499  0.073335  0.010683 -0.070616  0.008012  0.042056  0.006054  efectores

[500 rows x 14 columns]


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

                X0          X1          X2          X3          X4          X5  \
count   500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean      0.002939    0.008203    0.007979    0.014099   -0.004549   -0.000841
```

```
std      0.070068    0.075840    0.066766    0.070813    0.076525    0.068372
min     -0.222323   -0.339124   -0.222691   -0.187396   -0.310859   -0.237262
25%     -0.040690   -0.034304   -0.041085   -0.033020   -0.046808   -0.042516
50%      0.000953    0.013606    0.007162    0.020809   -0.004059   -0.000808
75%      0.048414    0.056372    0.052288    0.062198    0.043345    0.038743
max      0.280399    0.253401    0.247227    0.212324    0.204496    0.249863

                X6          X7          X8          X9          X10         X11  \
count   500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean      0.018172    0.003362   -0.002521    0.007883    0.000809   -0.007018
std       0.070236    0.063886    0.073789    0.070119    0.070408    0.061428
min      -0.205625   -0.248853   -0.299462   -0.206259   -0.273953   -0.207756
25%      -0.023447   -0.034456   -0.051159   -0.036106   -0.041130   -0.041249
50%       0.017333    0.005366   -0.002557    0.005596    0.001569   -0.006423
75%       0.058472    0.044924    0.043056    0.048046    0.043449    0.027856
max       0.291283    0.200966    0.252725    0.253159    0.299910    0.216093

               X12
count   500.000000
mean      0.013017
std       0.074332
min      -0.296465
25%      -0.032943
50%       0.011452
75%       0.063365
max       0.244813
```

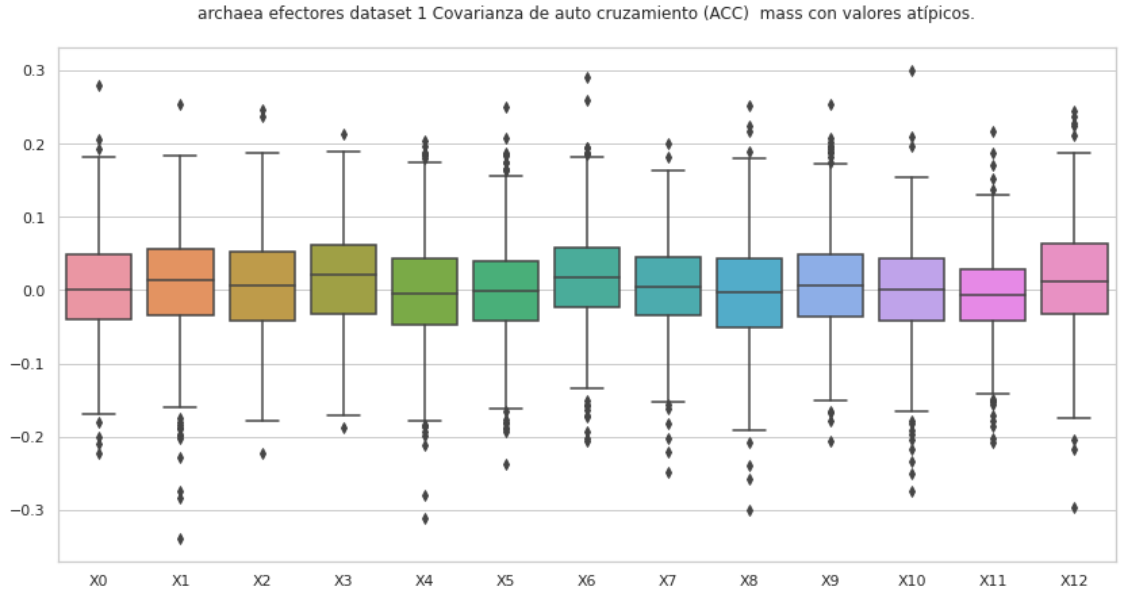
no_efectores
Covarianza de auto cruzamiento (ACC) mass no_efectores archaea dataset 1, con valores atípicos.
Valores del documento csv.

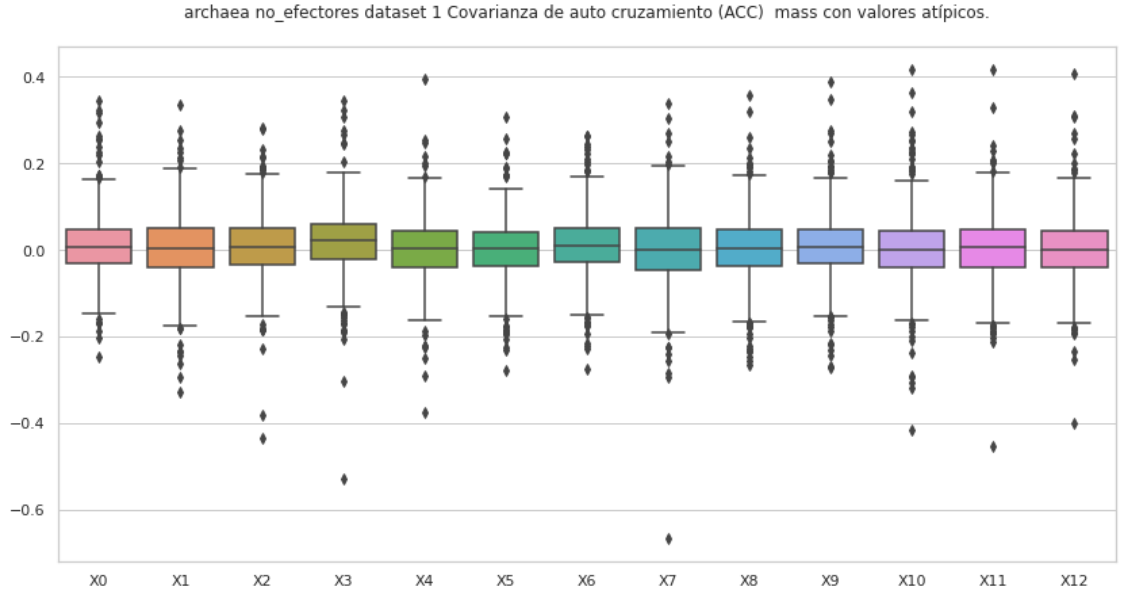
```
           X0        X1        X2        X3        X4        X5        X6  \
0   -0.007453 -0.004358  0.022996  0.041770  0.031562 -0.001634 -0.016052
1    0.077597 -0.028794  0.007985  0.041786  0.075589  0.003971  0.009067
2   -0.169125  0.099419 -0.018895 -0.163102 -0.007650 -0.082962 -0.021851
3   -0.035957 -0.261776  0.281942 -0.005707 -0.154203  0.256104  0.092575
4   -0.002941  0.056854 -0.050010 -0.038256  0.026012 -0.029620  0.105664
..        …         …         …         …         …         …         …
495 -0.025754 -0.119795  0.025343 -0.108952  0.072790  0.069994  0.075011
496  0.128380  0.026249  0.073039  0.035285  0.003243  0.071292  0.093109
497 -0.013349  0.000164  0.040691 -0.050973  0.099268 -0.010383 -0.007907
498 -0.127940  0.081952 -0.098889  0.008579 -0.022997  0.046067 -0.027528
499 -0.012926 -0.022720  0.002323  0.037684 -0.114223 -0.004831 -0.146306

           X7        X8        X9       X10       X11       X12           X13
```

```
0    0.011972  0.041875  0.025961 -0.023854 -0.043768 -0.004662  no_efectores
1   -0.033402 -0.011621  0.053565  0.057959  0.039933 -0.026798  no_efectores
2    0.079553  0.016118  0.058210 -0.106449  0.174852 -0.098204  no_efectores
3   -0.294665  0.358922  0.079103 -0.416854  0.036443  0.113940  no_efectores
4   -0.013605  0.021535  0.024658  0.023564  0.033117 -0.112959  no_efectores
..        …         …         …         …         …         …           …
495  0.046176 -0.077435 -0.126524 -0.039933  0.055277 -0.038658  no_efectores
496 -0.009276  0.020455  0.032325  0.070916 -0.005286  0.125299  no_efectores
497  0.030171  0.015544  0.010503 -0.033135 -0.017511 -0.037388  no_efectores
498 -0.148681 -0.075304 -0.079266  0.073382  0.000108 -0.038254  no_efectores
499 -0.189282  0.027878 -0.124328  0.020047 -0.035824 -0.048783  no_efectores

[500 rows x 14 columns]
```

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

```
                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.012023    0.005795    0.005676    0.021019    0.001277   -0.000015
std      0.074929    0.079772    0.076852    0.080765    0.075763    0.071589
min     -0.248673   -0.327442   -0.434693   -0.530479   -0.375185   -0.277203
25%     -0.029441   -0.040566   -0.034124   -0.020548   -0.039769   -0.036475
50%      0.006062    0.004824    0.006565    0.021511    0.002389    0.002121
75%      0.048774    0.051565    0.050485    0.060987    0.043946    0.042136
max      0.344385    0.334743    0.281942    0.343865    0.395156    0.308510

                X6          X7          X8          X9         X10         X11  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.011894    0.001503    0.004246    0.007693    0.003425    0.004518
std      0.072772    0.086517    0.079204    0.078800    0.086653    0.080911
min     -0.276890   -0.667267   -0.267040   -0.273796   -0.416854   -0.453421
25%     -0.028582   -0.046016   -0.038396   -0.031605   -0.039098   -0.041306
50%      0.009486    0.001190    0.003672    0.007697    0.000781    0.005677
75%      0.051764    0.051047    0.047262    0.048624    0.043493    0.047736
max      0.264049    0.337730    0.358922    0.388279    0.416983    0.415995

               X12
count  500.000000
mean     0.000399
std      0.076327
min     -0.400128
25%     -0.041488
50%      0.000645
75%      0.043943
max      0.407235
```


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


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

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

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

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

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

    if etiq == "efectores":
        df=ACC_mass_efec

    if etiq == "no_efectores":
        df=ACC_mass_no_efec

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

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

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

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

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

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

	X0	X1	X2	X3	X4	X5	X6	\
0	-0.014333	0.058410	-0.015800	0.050499	0.033702	0.011739	0.033106	
1	0.117089	0.022280	-0.010059	-0.055715	0.019496	0.030200	0.194636	
2	-0.022653	-0.067049	0.038619	-0.016568	-0.025107	0.036969	0.018184	
3	-0.014102	-0.186867	-0.058138	0.028357	0.017154	-0.014069	-0.006509	
4	-0.042041	-0.037866	-0.132133	0.067086	-0.011917	0.054880	-0.013365	
..	…	…	…	…	…	…	…	
494	-0.078251	0.022220	0.009384	-0.016942	-0.108257	-0.109375	-0.008845	
496	-0.098494	0.090414	-0.046123	0.031132	-0.029106	-0.015877	-0.005387	
497	0.035241	0.074552	-0.051782	0.104516	0.053124	0.030135	0.009295	
498	-0.008424	-0.098538	0.183161	-0.018165	-0.110064	0.127099	0.025780	
499	-0.089477	-0.068173	0.059083	-0.068624	-0.016302	0.084140	-0.104055	

	X7	X8	X9	X10	X11	X12	X13
0	0.066307	0.070958	0.002458	-0.054490	0.074434	-0.009852	efectores
1	0.181927	0.061440	0.052418	-0.095182	0.017137	-0.008054	efectores
2	-0.011953	-0.012152	0.022458	0.069529	-0.032195	0.065258	efectores
3	0.071194	-0.012853	-0.061057	0.058445	0.040567	0.055157	efectores
4	0.052872	0.046645	0.133642	0.040733	-0.058345	-0.119692	efectores
..	…	…	…	…	…	…	
494	0.065117	0.044981	0.032289	0.036142	0.048335	-0.118593	efectores
496	-0.014981	-0.070751	0.022246	-0.034248	0.045241	-0.110333	efectores
497	0.018779	0.029106	-0.015120	0.000557	-0.008222	-0.000135	efectores
498	-0.096359	0.058341	-0.034121	-0.117291	-0.031005	0.054966	efectores
499	0.073335	0.010683	-0.070616	0.008012	0.042056	0.006054	efectores

```
[463 rows x 14 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	463.000000	463.000000	463.000000	463.000000	463.000000	463.000000	
mean	0.003403	0.011309	0.008601	0.012042	-0.003815	-0.001904	
std	0.064829	0.068062	0.063867	0.068499	0.072637	0.063039	
min	-0.180330	-0.202726	-0.177774	-0.187396	-0.212280	-0.193211	
25%	-0.037879	-0.030429	-0.036038	-0.034355	-0.045543	-0.040754	
50%	0.001878	0.014133	0.006504	0.015855	-0.003225	-0.002568	
75%	0.047399	0.055251	0.051397	0.059899	0.040696	0.035035	

```
max      0.206320    0.182761    0.187346    0.212324    0.204496    0.184461

               X6          X7          X8          X9         X10         X11  \
count  463.000000  463.000000  463.000000  463.000000  463.000000  463.000000
mean     0.019625    0.004096   -0.001484    0.007492    0.000611   -0.007612
std      0.062989    0.058850    0.068576    0.067941    0.066034    0.057330
min     -0.172609   -0.160876   -0.207629   -0.179212   -0.203896   -0.186031
25%     -0.020101   -0.032215   -0.049952   -0.034980   -0.041943   -0.040109
50%      0.017638    0.005405   -0.001833    0.005646    0.000557   -0.006214
75%      0.057948    0.043441    0.042487    0.047208    0.042882    0.026146
max      0.194636    0.181927    0.217009    0.207579    0.209027    0.170779

              X12
count  463.000000
mean     0.013592
std      0.069768
min     -0.203880
25%     -0.032722
50%      0.011391
75%      0.062490
max      0.228887
```

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

```
           X0        X1        X2        X3        X4        X5        X6  \
0   -0.007453 -0.004358  0.022996  0.041770  0.031562 -0.001634 -0.016052
1    0.077597 -0.028794  0.007985  0.041786  0.075589  0.003971  0.009067
2   -0.169125  0.099419 -0.018895 -0.163102 -0.007650 -0.082962 -0.021851
4   -0.002941  0.056854 -0.050010 -0.038256  0.026012 -0.029620  0.105664
5    0.040365 -0.031385 -0.018865 -0.054045 -0.033356 -0.000632  0.098324
..        …         …         …         …         …         …
495 -0.025754 -0.119795  0.025343 -0.108952  0.072790  0.069994  0.075011
496  0.128380  0.026249  0.073039  0.035285  0.003243  0.071292  0.093109
497 -0.013349  0.000164  0.040691 -0.050973  0.099268 -0.010383 -0.007907
498 -0.127940  0.081952 -0.098889  0.008579 -0.022997  0.046067 -0.027528
499 -0.012926 -0.022720  0.002323  0.037684 -0.114223 -0.004831 -0.146306

           X7        X8        X9       X10       X11       X12           X13
0    0.011972  0.041875  0.025961 -0.023854 -0.043768 -0.004662  no_efectores
1   -0.033402 -0.011621  0.053565  0.057959  0.039933 -0.026798  no_efectores
2    0.079553  0.016118  0.058210 -0.106449  0.174852 -0.098204  no_efectores
4   -0.013605  0.021535  0.024658  0.023564  0.033117 -0.112959  no_efectores
5   -0.031198  0.036242  0.047570  0.000683 -0.030940  0.029588  no_efectores
..        …         …         …         …         …         …           …
```

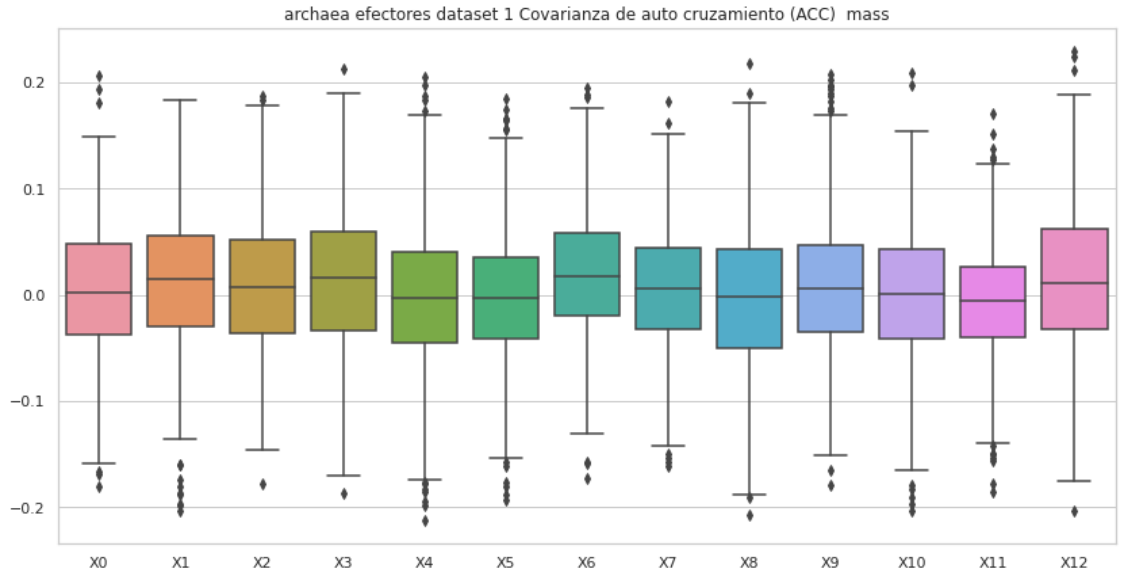
```
495  0.046176 -0.077435 -0.126524 -0.039933  0.055277 -0.038658  no_efectores
496 -0.009276  0.020455  0.032325  0.070916 -0.005286  0.125299  no_efectores
497  0.030171  0.015544  0.010503 -0.033135 -0.017511 -0.037388  no_efectores
498 -0.148681 -0.075304 -0.079266  0.073382  0.000108 -0.038254  no_efectores
499 -0.189282  0.027878 -0.124328  0.020047 -0.035824 -0.048783  no_efectores

[446 rows x 14 columns]
```

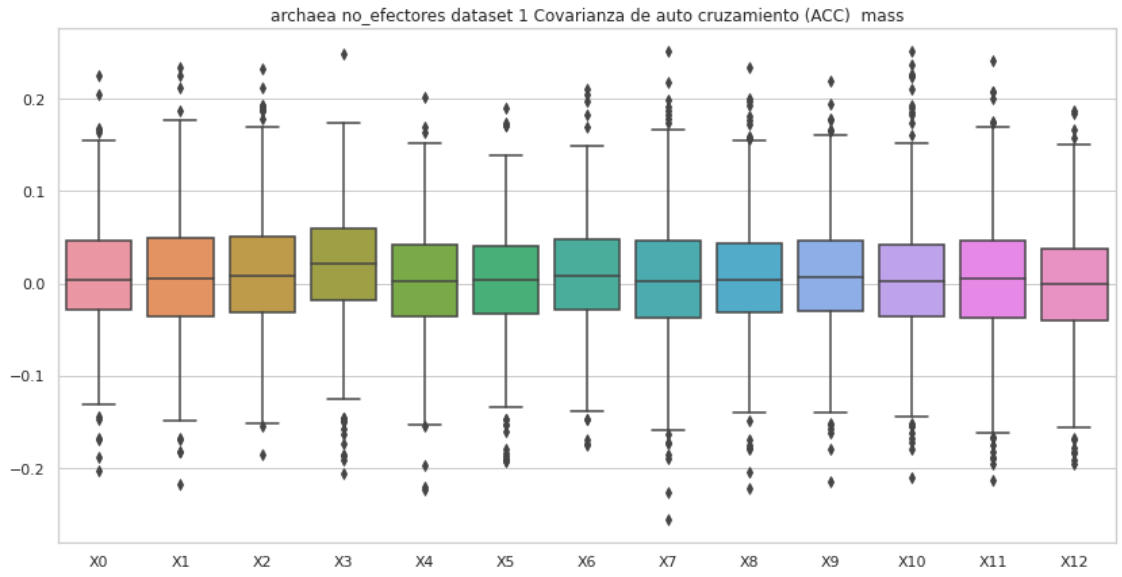
Covarianza de auto cruzamiento (ACC) mass no_efectores archaea dataset 1, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5
count	446.000000	446.000000	446.000000	446.000000	446.000000	446.000000
mean	0.007665	0.007125	0.009014	0.019831	0.002145	0.000505
std	0.062327	0.067033	0.064606	0.066620	0.063105	0.061649
min	-0.202478	-0.217515	-0.184838	-0.206192	-0.223705	-0.193069
25%	-0.028503	-0.036152	-0.030675	-0.018003	-0.036202	-0.033072
50%	0.004339	0.004824	0.007632	0.021236	0.002802	0.003883
75%	0.046747	0.049267	0.050256	0.059178	0.042114	0.041206
max	0.225241	0.234617	0.232387	0.249355	0.201380	0.190655

	X6	X7	X8	X9	X10	X11
count	446.000000	446.000000	446.000000	446.000000	446.000000	446.000000
mean	0.009846	0.003437	0.006247	0.006376	0.004872	0.004491
std	0.060037	0.070824	0.065546	0.064467	0.068593	0.069030
min	-0.174996	-0.255489	-0.221664	-0.214500	-0.210503	-0.213039
25%	-0.028234	-0.037194	-0.030506	-0.029954	-0.034957	-0.036797
50%	0.008086	0.001933	0.004224	0.007260	0.002456	0.005677
75%	0.047306	0.046775	0.044062	0.046528	0.042001	0.046915
max	0.211027	0.251549	0.234209	0.219879	0.251512	0.241907

	X12
count	446.000000
mean	-0.001569
std	0.062912
min	-0.195438
25%	-0.040161
50%	0.000154
75%	0.037655
max	0.187897





8 Covarianza de auto cruzamiento (ACC) hidro

[15]:
```
#hidro
transf = "Covarianza de auto cruzamiento (ACC) "
transf2 = "ACC"
estado = "con valores atípicos.\n"
comp = "hidro"
df=""
```

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

    if etiq == "efectores":
        df=ACC_hidro_efec

    if etiq == "no_efectores":
        df=ACC_hidro_no_efec

    #del df['X13']
    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")

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

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.010813 -0.043708 -0.015619  0.110187  0.034273 -0.052725 -0.034041
1    0.018466  0.035859  0.045762 -0.006163  0.020743 -0.012747  0.002079
2    0.062457  0.023670  0.065404  0.121756 -0.002160  0.021997  0.087227
3    0.005361 -0.143134 -0.009402 -0.083602 -0.128791 -0.025536  0.073148
4    0.047844 -0.131639  0.153584  0.062186 -0.076578  0.063472  0.071524
..        …       …         …         …         …         …         …
495 -0.089987 -0.114351 -0.136477  0.253919 -0.181981  0.033551 -0.120484
496  0.110535 -0.173183  0.012241 -0.052766 -0.056464  0.039465  0.069483
497  0.035339  0.065686  0.007001  0.034806  0.110848  0.026836  0.105358
498  0.050122 -0.002242  0.087628  0.037271 -0.016424  0.006568  0.027550
499 -0.045795 -0.044247  0.034310  0.006797 -0.015120 -0.028852  0.016193

           X7        X8        X9       X10       X11       X12        X13
0    0.014762  0.053089 -0.007727  0.028033 -0.059612  0.019024  efectores
1    0.020930 -0.006431  0.037323 -0.019954  0.015630  0.020927  efectores
2    0.080268 -0.018311 -0.008097  0.017690  0.042552 -0.046469  efectores
```

```
3   -0.047546  0.096980  0.017039 -0.074668 -0.092614  0.024981  efectores
4   -0.116800  0.037860  0.014940 -0.122099  0.037634  0.155464  efectores
..        …       …        …        …        …        …        …
495 -0.020502 -0.098145  0.263118 -0.079966  0.020860 -0.068351  efectores
496 -0.092323 -0.124977 -0.048655  0.059218  0.018686 -0.029556  efectores
497 -0.052290  0.008072 -0.024165 -0.022710 -0.027212 -0.007112  efectores
498 -0.026495  0.067510 -0.001381 -0.039646  0.044862  0.017056  efectores
499 -0.074017  0.040207 -0.042146  0.034993 -0.001887 -0.051076  efectores

[500 rows x 14 columns]
```

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

```
                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.022998   -0.036776    0.045352    0.026961   -0.022786   -0.011669
std      0.092663    0.101237    0.082437    0.089213    0.097333    0.082806
min     -0.279380   -0.338283   -0.252217   -0.343051   -0.390413   -0.301237
25%     -0.028693   -0.108293   -0.012531   -0.022172   -0.086950   -0.062430
50%      0.024269   -0.031226    0.043545    0.032782   -0.015305   -0.014121
75%      0.081199    0.040741    0.090570    0.079184    0.044174    0.042444
max      0.293358    0.217252    0.380130    0.341833    0.301426    0.227149

                X6          X7          X8          X9         X10         X11  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.022262    0.018177    0.002517    0.001366    0.010901    0.001032
std      0.094539    0.081980    0.088084    0.093011    0.085835    0.075242
min     -0.287856   -0.290159   -0.364356   -0.298832   -0.238090   -0.251363
25%     -0.033826   -0.031808   -0.041450   -0.050465   -0.041402   -0.042787
50%      0.019869    0.018351    0.004804   -0.000626   -0.000562   -0.002187
75%      0.075570    0.066434    0.044515    0.044613    0.060493    0.038592
max      0.389377    0.267646    0.246266    0.329916    0.340721    0.273557

               X12
count  500.000000
mean    -0.011007
std      0.083748
min     -0.339210
25%     -0.053374
50%     -0.008093
75%      0.035670
max      0.259563
```

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.025839	-0.013484	0.064962	0.042764	0.061521	0.012659	0.098122	
1	0.051400	-0.019094	0.095371	0.137372	0.055921	-0.011560	0.007717	
2	0.073366	-0.230503	0.054053	0.062910	-0.140882	-0.058623	0.056409	
3	0.000101	-0.232138	-0.010728	-0.213806	-0.019699	-0.148524	0.196655	
4	-0.055621	-0.059121	-0.028236	-0.006432	0.010757	0.047635	-0.017354	
..	…	…	…	…	…	…	…	
495	-0.077947	0.071496	-0.000664	-0.117423	0.002488	-0.028192	-0.014107	
496	-0.014575	0.033931	0.085516	-0.052635	0.069765	0.069930	-0.023560	
497	0.112248	-0.000675	0.119442	0.181694	0.059348	0.051371	0.033442	
498	-0.105151	0.070458	-0.033886	0.037594	-0.092236	-0.027610	0.042822	
499	0.105092	-0.072957	-0.104645	-0.239808	-0.046419	0.045789	0.065272	

	X7	X8	X9	X10	X11	X12	X13
0	0.094641	0.069041	0.030236	0.049520	0.090141	0.037043	no_efectores
1	0.026657	0.062821	-0.042997	-0.050652	0.019933	0.055072	no_efectores
2	-0.047640	0.129296	0.064202	-0.087319	-0.037917	0.000677	no_efectores
3	0.070149	0.002569	0.078883	-0.080996	0.286289	0.197663	no_efectores
4	-0.011155	-0.015526	-0.014901	0.025979	0.035159	0.022108	no_efectores
..	…	…	…	…	…	…	…
495	-0.090727	0.085249	0.006734	-0.059456	0.165084	-0.036906	no_efectores
496	0.121599	-0.111122	0.048290	0.022574	-0.052737	0.029373	no_efectores
497	0.083488	0.047107	-0.025636	0.024511	0.078436	-0.002252	no_efectores
498	-0.163513	0.083801	0.052230	0.030366	-0.001355	0.068289	no_efectores
499	0.014544	-0.168106	-0.027516	0.226528	0.065615	0.233856	no_efectores

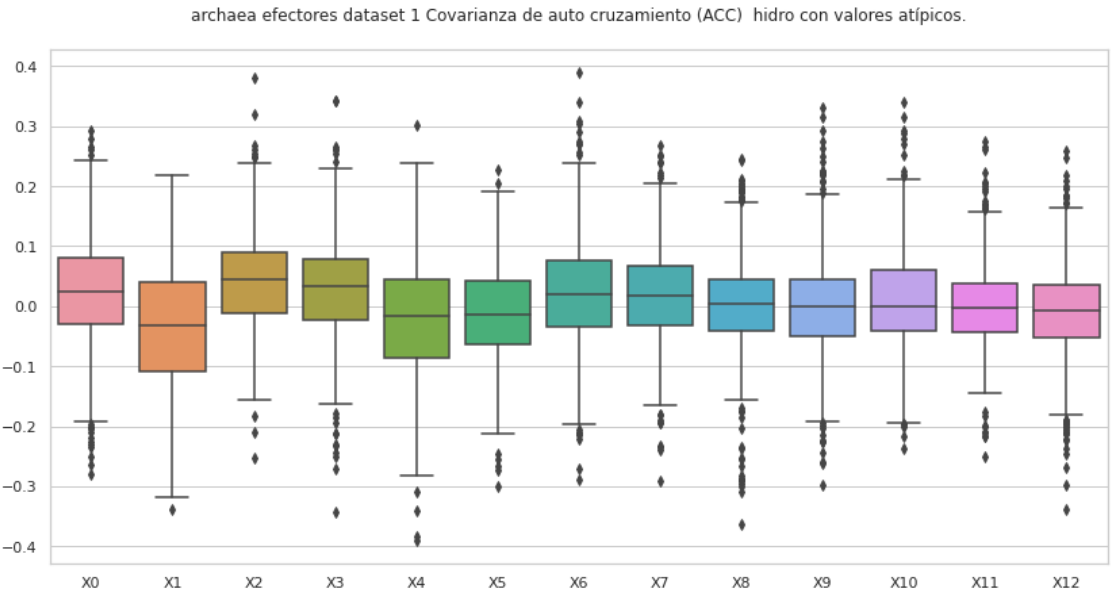
[500 rows x 14 columns]

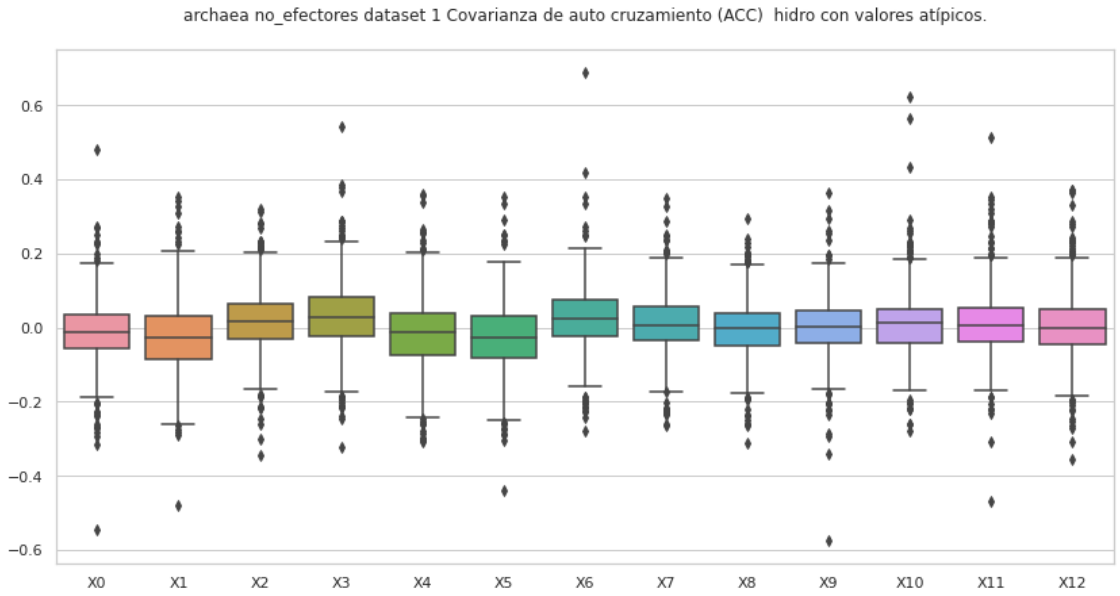
Covarianza de auto cruzamiento (ACC) hidro no_efectores archaea dataset 1, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	-0.013778	-0.027063	0.018051	0.031348	-0.014320	-0.026011	
std	0.091836	0.102373	0.087081	0.097127	0.096177	0.093608	
min	-0.546580	-0.480687	-0.343605	-0.323134	-0.308300	-0.438688	
25%	-0.057468	-0.085981	-0.029380	-0.022659	-0.073599	-0.080043	
50%	-0.013840	-0.027858	0.015841	0.027736	-0.012666	-0.026516	
75%	0.036754	0.031529	0.064940	0.080974	0.039260	0.032925	
max	0.480790	0.354606	0.319176	0.540806	0.359060	0.353786	

	X6	X7	X8	X9	X10	X11	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.024874	0.011648	-0.004019	0.000361	0.014283	0.013111	
std	0.092538	0.085031	0.082716	0.085493	0.092302	0.093376	
min	-0.279858	-0.264266	-0.310896	-0.574275	-0.279131	-0.468643	
25%	-0.023344	-0.034015	-0.050166	-0.041433	-0.040238	-0.038033	
50%	0.022698	0.007861	-0.002363	0.001784	0.012115	0.007329	
75%	0.075543	0.058513	0.040208	0.045608	0.050202	0.052178	
max	0.688296	0.350964	0.293381	0.365474	0.621669	0.512927	

	X12
count	500.000000
mean	0.004756
std	0.092725
min	-0.356503
25%	-0.045572
50%	-0.002120
75%	0.051142
max	0.370548





8.1 Covarianza de auto cruzamiento (ACC) hidro, sin valores atípicos

[16]:
```
#hidro
transf = "Covarianza de auto cruzamiento (ACC) "
transf2 = "ACC"
estado = "sin valores atípicos.\n"
comp = "hidro"
df=""

out = (str(r3) + '/ds' + str(dataset) + '_' + str(transf2) + '_' + str(comp) +␣
 ↪'_' + str(organismo) + '.csv')
os.makedirs(str(r3), exist_ok=True)
df_out = pd.DataFrame()

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

    if etiq == "efectores":
        df=ACC_hidro_efec

    if etiq == "no_efectores":
        df=ACC_hidro_no_efec

    del df['X13']
```

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

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

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

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

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.010813 -0.043708 -0.015619  0.110187  0.034273 -0.052725 -0.034041
1    0.018466  0.035859  0.045762 -0.006163  0.020743 -0.012747  0.002079
2    0.062457  0.023670  0.065404  0.121756 -0.002160  0.021997  0.087227
3    0.005361 -0.143134 -0.009402 -0.083602 -0.128791 -0.025536  0.073148
4    0.047844 -0.131639  0.153584  0.062186 -0.076578  0.063472  0.071524
..        …       …         …         …         …         …         …
495 -0.089987 -0.114351 -0.136477  0.253919 -0.181981  0.033551 -0.120484
496  0.110535 -0.173183  0.012241 -0.052766 -0.056464  0.039465  0.069483
497  0.035339  0.065686  0.007001  0.034806  0.110848  0.026836  0.105358
498  0.050122 -0.002242  0.087628  0.037271 -0.016424  0.006568  0.027550
499 -0.045795 -0.044247  0.034310  0.006797 -0.015120 -0.028852  0.016193

           X7        X8        X9       X10       X11       X12        X13
0    0.014762  0.053089 -0.007727  0.028033 -0.059612  0.019024  efectores
1    0.020930 -0.006431  0.037323 -0.019954  0.015630  0.020927  efectores
2    0.080268 -0.018311 -0.008097  0.017690  0.042552 -0.046469  efectores
3   -0.047546  0.096980  0.017039 -0.074668 -0.092614  0.024981  efectores
4   -0.116800  0.037860  0.014940 -0.122099  0.037634  0.155464  efectores
..        …       …         …         …         …         …         …
495 -0.020502 -0.098145  0.263118 -0.079966  0.020860 -0.068351  efectores
```

```
496 -0.092323 -0.124977 -0.048655  0.059218  0.018686 -0.029556  efectores
497 -0.052290  0.008072 -0.024165 -0.022710 -0.027212 -0.007112  efectores
498 -0.026495  0.067510 -0.001381 -0.039646  0.044862  0.017056  efectores
499 -0.074017  0.040207 -0.042146  0.034993 -0.001887 -0.051076  efectores

[458 rows x 14 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	458.000000	458.000000	458.000000	458.000000	458.000000	458.000000	
mean	0.026306	-0.032496	0.042045	0.029169	-0.017839	-0.009336	
std	0.087230	0.095670	0.075472	0.081251	0.086501	0.076129	
min	-0.250127	-0.338283	-0.154960	-0.232786	-0.269540	-0.254186	
25%	-0.022178	-0.098437	-0.013552	-0.018051	-0.075537	-0.059626	
50%	0.025977	-0.029019	0.039563	0.034991	-0.011699	-0.012126	
75%	0.080970	0.043342	0.086772	0.078403	0.046068	0.041194	
max	0.293358	0.181418	0.266799	0.266048	0.197467	0.204343	

	X6	X7	X8	X9	X10	X11	\
count	458.000000	458.000000	458.000000	458.000000	458.000000	458.000000	
mean	0.019586	0.019570	0.006972	-0.001182	0.004905	0.001170	
std	0.082603	0.074209	0.074005	0.080481	0.076987	0.068318	
min	-0.220771	-0.193252	-0.255159	-0.225613	-0.238090	-0.217089	
25%	-0.031442	-0.030009	-0.033819	-0.050173	-0.042760	-0.040592	
50%	0.019682	0.018587	0.006893	-0.001381	-0.004373	-0.002119	
75%	0.070592	0.064836	0.044437	0.040961	0.052919	0.037695	
max	0.304071	0.251496	0.242204	0.263118	0.251445	0.223450	

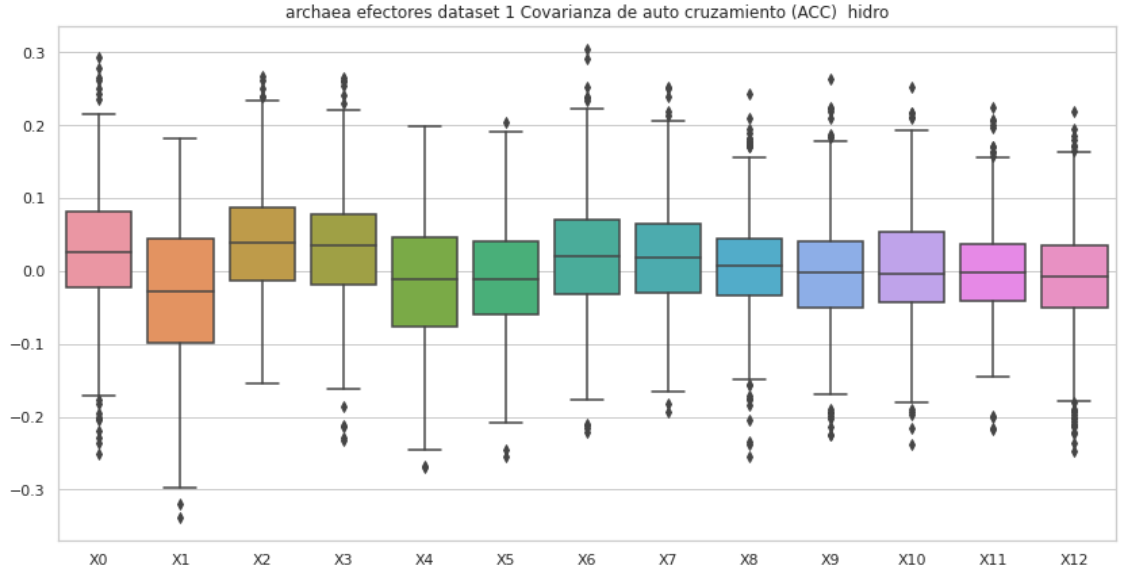
	X12
count	458.000000
mean	-0.010059
std	0.075846
min	-0.246765
25%	-0.051054
50%	-0.007924
75%	0.034915
max	0.218638

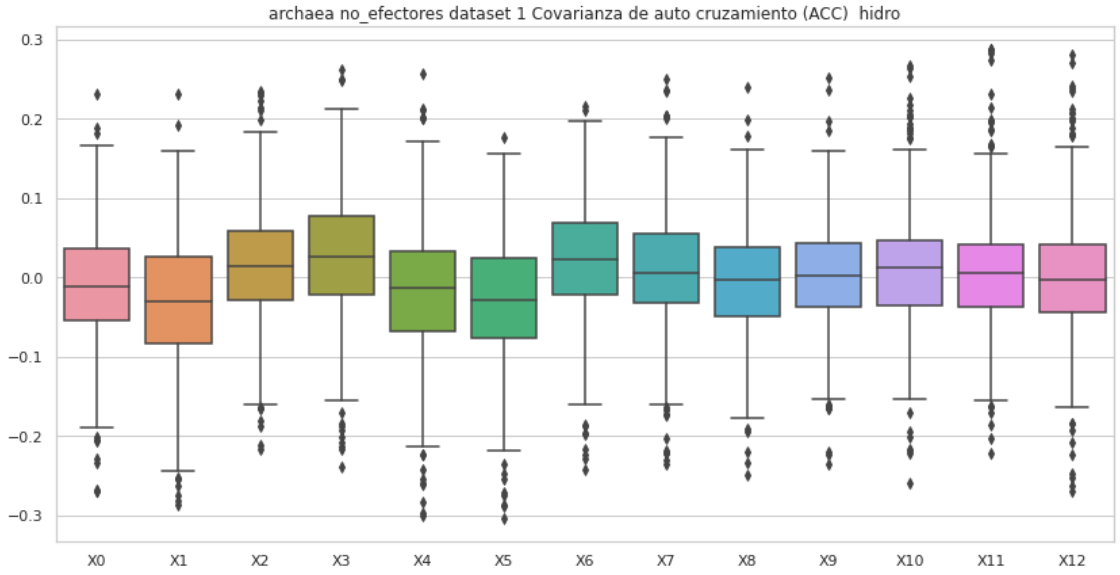
no_efectores
Covarianza de auto cruzamiento (ACC) no_efectores archaea dataset 1, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.025839 -0.013484  0.064962  0.042764  0.061521  0.012659  0.098122
1    0.051400 -0.019094  0.095371  0.137372  0.055921 -0.011560  0.007717
2    0.073366 -0.230503  0.054053  0.062910 -0.140882 -0.058623  0.056409
3    0.000101 -0.232138 -0.010728 -0.213806 -0.019699 -0.148524  0.196655
4   -0.055621 -0.059121 -0.028236 -0.006432  0.010757  0.047635 -0.017354
..        …         …         …         …         …         …
495 -0.077947  0.071496 -0.000664 -0.117423  0.002488 -0.028192 -0.014107
496 -0.014575  0.033931  0.085516 -0.052635  0.069765  0.069930 -0.023560
497  0.112248 -0.000675  0.119442  0.181694  0.059348  0.051371  0.033442
498 -0.105151  0.070458 -0.033886  0.037594 -0.092236 -0.027610  0.042822
499  0.105092 -0.072957 -0.104645 -0.239808 -0.046419  0.045789  0.065272

           X7        X8        X9       X10       X11       X12           X13
0    0.094641  0.069041  0.030236  0.049520  0.090141  0.037043  no_efectores
1    0.026657  0.062821 -0.042997 -0.050652  0.019933  0.055072  no_efectores
2   -0.047640  0.129296  0.064202 -0.087319 -0.037917  0.000677  no_efectores
3    0.070149  0.002569  0.078883 -0.080996  0.286289  0.197663  no_efectores
4   -0.011155 -0.015526 -0.014901  0.025979  0.035159  0.022108  no_efectores
..        …         …         …         …         …         …
495 -0.090727  0.085249  0.006734 -0.059456  0.165084 -0.036906  no_efectores
496  0.121599 -0.111122  0.048290  0.022574 -0.052737  0.029373  no_efectores
497  0.083488  0.047107 -0.025636  0.024511  0.078436 -0.002252  no_efectores
498 -0.163513  0.083801  0.052230  0.030366 -0.001355  0.068289  no_efectores
499  0.014544 -0.168106 -0.027516  0.226528  0.065615  0.233856  no_efectores

[451 rows x 14 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  451.000000  451.000000  451.000000  451.000000  451.000000  451.000000
mean    -0.013085   -0.032445    0.014071    0.024955   -0.017301   -0.030933
std      0.075767    0.086568    0.074053    0.080766    0.081820    0.079204
min     -0.269838   -0.287777   -0.216903   -0.239808   -0.300777   -0.304096
25%     -0.055122   -0.083306   -0.029174   -0.022338   -0.068885   -0.076258
50%     -0.012410   -0.030529    0.013693    0.025586   -0.012667   -0.027980
75%      0.035555    0.025501    0.058924    0.076602    0.033289    0.024508
max      0.231116    0.230565    0.233917    0.261685    0.256267    0.176124

                X6          X7          X8          X9         X10         X11  \
count  451.000000  451.000000  451.000000  451.000000  451.000000  451.000000
mean     0.020531    0.009359   -0.006053    0.001197    0.009562    0.007719
std      0.077562    0.074555    0.070191    0.067905    0.074872    0.075011
```

min	-0.241992	-0.236329	-0.248776	-0.235410	-0.259553	-0.221494
25%	-0.022790	-0.032312	-0.048881	-0.037912	-0.036072	-0.038065
50%	0.021931	0.006188	-0.002467	0.002296	0.011615	0.004965
75%	0.068808	0.054673	0.037238	0.043646	0.046054	0.041699
max	0.215516	0.250357	0.239115	0.252464	0.267700	0.287388

	X12
count	451.000000
mean	0.001567
std	0.078947
min	-0.269549
25%	-0.043988
50%	-0.003946
75%	0.040990
max	0.281382




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