ds4_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 = 4
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]:
```python
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 4, con valores
atípicos.
Valores del documento csv.

         X0      X1     X2     X3     X4      X5     X6      X7     X8  \
0    12.130   4.438  1.183  2.959  0.296   2.367  2.367  12.426  0.296
1     4.925   4.711  5.567  4.069  0.642   5.139  1.927   5.353  1.713
2    12.598   4.724  1.575  5.512  0.000  13.386  0.787   7.087  0.787
3     0.000  12.857  1.429  4.286  2.857  11.429  8.571   7.143  0.000
4     4.292   5.579  4.721  5.365  0.429   4.721  1.931   4.936  1.073
..      …       …     …     …     …      …      …     …      …
495   2.247   3.371  7.865  6.742  0.000   4.494  2.247  11.236  1.124
496   3.828   4.306  5.263  4.785  0.478   5.742  0.957   3.828  1.435
497  12.698   5.291  0.529  5.291  0.529   1.587  1.587   8.466  1.587
498  12.857   2.381  2.619  3.571  0.238   2.619  2.857  11.190  0.714
499  10.550   5.046  2.064  3.899  0.229   5.963  2.982   9.404  0.000
```

	X9	…	X11	X12	X13	X14	X15	X16	X17	X18	\
0	6.805	…	0.592	1.183	4.142	5.325	6.805	6.509	1.775	3.254	
1	12.206	…	8.351	2.570	7.281	4.925	4.497	4.497	1.071	4.069	
2	6.299	…	7.874	1.575	3.150	2.362	3.937	2.362	0.787	6.299	
3	5.714	…	1.429	2.857	1.429	4.286	8.571	5.714	1.429	0.000	
4	13.305	…	7.940	3.433	7.296	4.721	4.936	4.077	0.858	5.150	
..	…	…	…	…	…	…	…	…	…	…	
495	8.989	…	7.865	2.247	3.371	7.865	5.618	6.742	2.247	2.247	
496	11.005	…	15.311	1.435	6.220	5.263	3.349	2.871	1.914	7.656	
497	1.058	…	1.587	0.529	8.466	6.349	6.349	5.291	3.704	3.704	
498	7.143	…	1.190	4.762	5.476	3.810	5.238	4.524	1.905	2.381	
499	5.734	…	4.128	2.294	5.046	4.587	5.734	4.358	1.376	3.211	

	X19	X20
0	10.947	efectores
1	7.709	efectores
2	8.661	efectores
3	5.714	efectores
4	6.652	efectores
..	…	…
495	6.742	efectores
496	6.220	efectores
497	10.582	efectores
498	12.143	efectores
499	10.321	efectores

[500 rows x 21 columns]

Composición de aminoácidos (AAC) efectores archaea dataset 4, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	9.206016	6.217028	2.763046	6.276422	0.724464	7.200110	
std	4.445086	2.979604	2.155206	2.656069	0.912346	3.520843	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	5.927000	4.151000	1.178000	4.323250	0.000000	4.494000	
50%	8.523500	5.961000	2.230500	5.952000	0.461000	7.190000	
75%	12.191250	7.957500	4.082000	8.153500	1.094500	9.677000	
max	24.294000	29.545000	12.124000	14.231000	5.941000	17.073000	

	X6	X7	X8	X9	X10	X11	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	2.429538	7.543938	1.839340	5.702760	11.148056	4.453934	
std	1.553636	2.793033	1.290184	3.423166	3.451995	4.319660	

min	0.000000	0.535000	0.000000	0.000000	3.295000	0.000000
25%	1.394750	5.396000	0.859250	2.995500	8.720250	1.134250
50%	2.292000	7.303500	1.726000	5.000000	11.147500	2.469000
75%	3.164500	9.347500	2.598750	7.621000	13.291000	7.853000
max	11.111000	17.765000	16.043000	15.842000	23.256000	17.857000

	X12	X13	X14	X15	X16	X17 \
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	1.975588	3.817078	4.138012	5.866694	5.641626	1.199948
std	1.201279	1.864853	1.742502	2.207923	2.047838	0.976930
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	1.068500	2.484750	2.985000	4.279750	4.349500	0.561250
50%	1.700000	3.733000	4.016000	5.643500	5.645500	1.050000
75%	2.565500	4.893500	5.057750	7.087000	6.842250	1.703750
max	8.861000	10.277000	13.043000	15.248000	12.684000	5.208000

	X18	X19
count	500.000000	500.000000
mean	3.335584	8.520892
std	1.758867	3.471172
min	0.000000	0.000000
25%	2.190500	5.610000
50%	3.230500	8.192000
75%	4.130000	11.034000
max	13.889000	18.657000

no_efectores
Composición de aminoácidos (AAC) no_efectores archaea dataset 4, con valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9 \
0	2.865	5.469	7.292	6.510	1.562	5.469	2.865	5.469	2.083	8.854
1	8.592	7.637	2.864	6.205	2.148	5.489	3.580	7.399	2.387	3.580
2	10.526	9.023	2.632	9.023	1.128	6.767	1.880	7.143	1.128	5.263
3	5.714	6.667	4.762	5.714	0.000	13.333	4.762	8.571	0.952	6.667
4	13.580	9.877	0.823	9.465	1.235	8.642	0.000	11.111	2.881	1.235
..	…	…	…	…	…	…	…	…	…	…
495	6.392	7.629	3.402	8.763	1.031	10.515	3.711	7.526	2.268	6.392
496	14.932	5.656	1.584	5.656	1.131	2.941	0.226	12.670	0.905	2.941
497	5.357	8.929	2.679	3.571	0.000	10.714	3.571	3.571	1.786	8.036
498	5.797	5.797	1.449	17.391	0.000	8.696	7.246	5.797	2.899	7.246
499	10.784	6.863	1.961	9.804	0.000	15.686	2.941	2.941	4.902	1.961

	…	X11	X12	X13	X14	X15	X16	X17	X18	X19 \
0	…	8.073	2.604	5.208	2.344	10.417	3.385	0.781	3.906	3.906

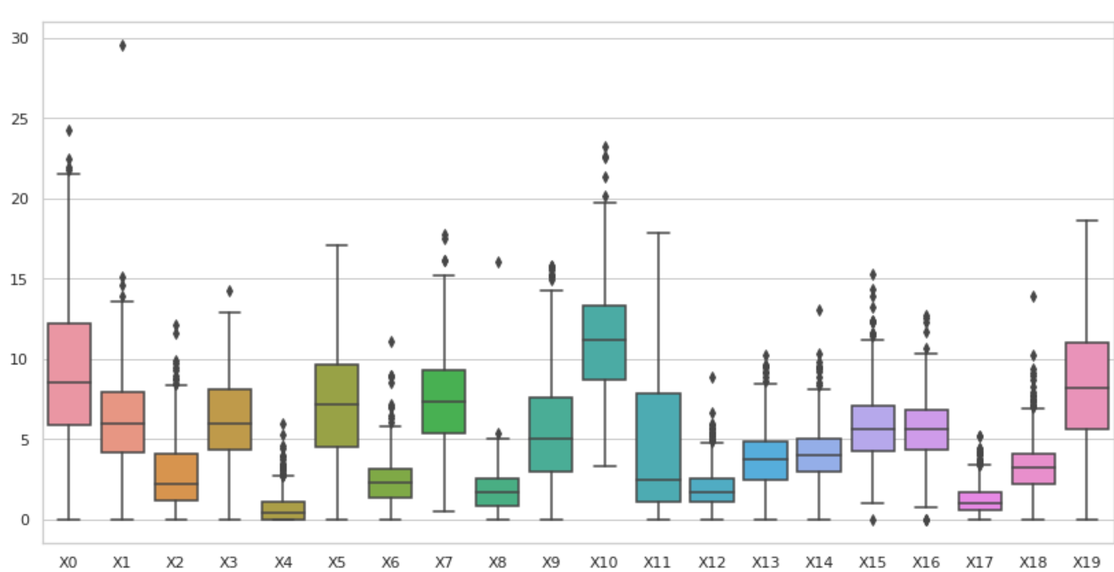
```
1    …  1.432  5.251  4.296  7.637   7.160  5.012  0.716  2.148   7.399
2    …  1.504  0.376  3.759  4.887   6.391  7.519  1.128  1.880   8.271
3    …  9.524  1.905  2.857  0.952   4.762  6.667  0.952  4.762   4.762
4    …  0.412  0.412  3.704  4.527   7.819  4.938  1.235  2.469   8.642
..   …    …      …     …     …       …      …     …      …      …
495  …  2.784  2.784  3.196  5.052   5.876  6.082  0.619  2.062   6.495
496  …  0.226  0.452  2.489  6.109   5.882  8.145  1.584  2.941  11.312
497  …  5.357  4.464  0.893  1.786   8.036  9.821  0.893  5.357   3.571
498  …  4.348  1.449  0.000  0.000   2.899  4.348  0.000  2.899   5.797
499  …  0.000  2.941  3.922  5.882   2.941  3.922  0.980  2.941   8.824

             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]
```

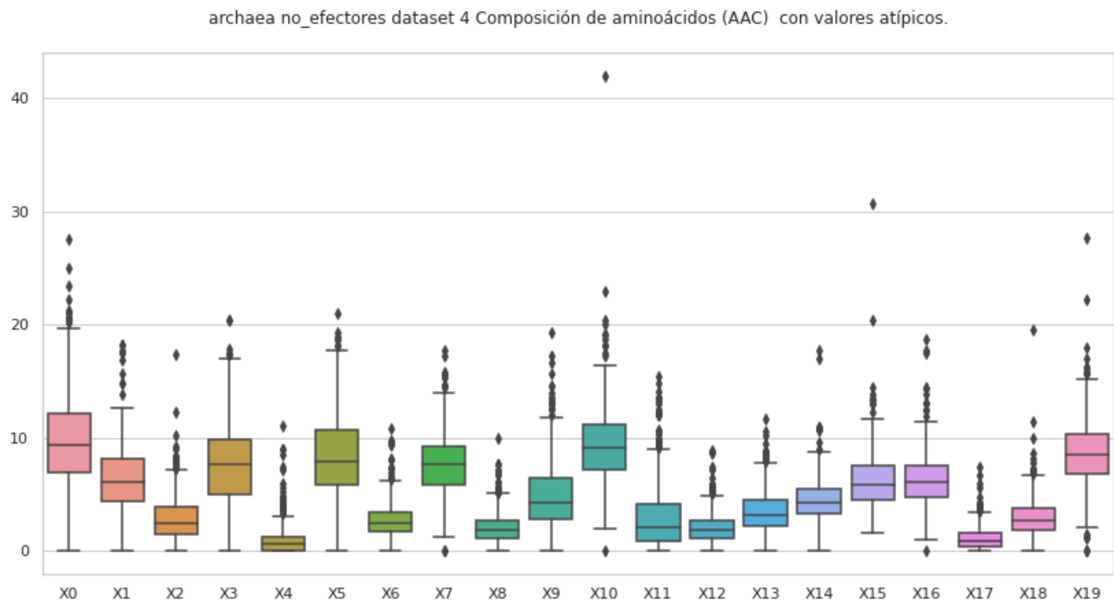
Composición de aminoácidos (AAC) no_efectores archaea dataset 4, con valores atípicos.
Estadísticas.

```
              X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     9.647648    6.264514    2.871438    7.528664    0.952476    8.162410
std      4.266084    2.926411    2.025049    3.491542    1.303477    3.625781
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      6.897000    4.373250    1.512000    4.952000    0.000000    5.765500
50%      9.298000    6.027000    2.463000    7.599000    0.626000    7.901000
75%     12.085750    8.069750    3.823500    9.776250    1.260000   10.611000
max     27.459000   18.182000   17.308000   20.339000   11.039000   20.930000

              X6          X7          X8          X9         X10         X11  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     2.621092    7.744046    2.024946    4.873540    9.332784    3.029026
std      1.693496    2.722094    1.383763    3.037662    3.519490    3.013873
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      1.638000    5.844250    1.106500    2.797500    7.118000    0.872250
50%      2.408000    7.692000    1.857000    4.221000    9.091000    2.073000
```

```
75%       3.448000    9.241000    2.718000    6.423000   11.096750    4.148500
max      10.840000   17.647000   10.000000   19.298000   41.935000   15.385000

              X12         X13         X14         X15         X16        X17  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.00000
mean     2.083822    3.474000    4.416004    6.201966    6.198486    1.08670
std      1.326604    1.951042    2.042957    2.622536    2.392961    1.04436
min      0.000000    0.000000    0.000000    1.562000    0.000000    0.00000
25%      1.124000    2.187250    3.214250    4.460250    4.715750    0.32275
50%      1.782000    3.211500    4.192500    5.821500    6.010000    0.88300
75%      2.639500    4.444750    5.483500    7.495500    7.482750    1.57425
max      8.889000   11.589000   17.742000   30.645000   18.699000    7.40700

              X18         X19
count  500.000000  500.000000
mean     2.917274    8.569198
std      1.743566    3.060521
min      0.000000    0.000000
25%      1.835000    6.772500
50%      2.703000    8.465500
75%      3.774000   10.277000
max     19.512000   27.660000
```

archaea efectores dataset 4 Composición de aminoácidos (AAC) con valores atípicos.





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 4, sin valores
atípicos.
Valores del documento csv.

         X0     X1     X2     X3     X4      X5     X6      X7     X8      X9  \
0    12.130  4.438  1.183  2.959  0.296   2.367  2.367  12.426  0.296   6.805
1     4.925  4.711  5.567  4.069  0.642   5.139  1.927   5.353  1.713  12.206
2    12.598  4.724  1.575  5.512  0.000  13.386  0.787   7.087  0.787   6.299
4     4.292  5.579  4.721  5.365  0.429   4.721  1.931   4.936  1.073  13.305
5     9.375  7.143  1.786  6.696  0.000   4.911  2.679  11.161  2.679   3.125
..      …      …     …      …     …       …     …       …      …       …
495   2.247  3.371  7.865  6.742  0.000   4.494  2.247  11.236  1.124   8.989
496   3.828  4.306  5.263  4.785  0.478   5.742  0.957   3.828  1.435  11.005
497  12.698  5.291  0.529  5.291  0.529   1.587  1.587   8.466  1.587   1.058
498  12.857  2.381  2.619  3.571  0.238   2.619  2.857  11.190  0.714   7.143
499  10.550  5.046  2.064  3.899  0.229   5.963  2.982   9.404  0.000   5.734

     …    X11    X12    X13    X14    X15    X16    X17    X18     X19  \
0    …  0.592  1.183  4.142  5.325  6.805  6.509  1.775  3.254  10.947
1    …  8.351  2.570  7.281  4.925  4.497  4.497  1.071  4.069   7.709
2    …  7.874  1.575  3.150  2.362  3.937  2.362  0.787  6.299   8.661
4    …  7.940  3.433  7.296  4.721  4.936  4.077  0.858  5.150   6.652
5    …  0.000  1.339  4.464  6.250  6.250  4.018  1.786  1.786  12.500
..   …    …      …     …      …     …      …     …      …       …
495  …  7.865  2.247  3.371  7.865  5.618  6.742  2.247  2.247   6.742
```

```
496  …  15.311  1.435  6.220  5.263  3.349  2.871  1.914  7.656   6.220
497  …   1.587  0.529  8.466  6.349  6.349  5.291  3.704  3.704  10.582
498  …   1.190  4.762  5.476  3.810  5.238  4.524  1.905  2.381  12.143
499  …   4.128  2.294  5.046  4.587  5.734  4.358  1.376  3.211  10.321

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

[437 rows x 21 columns]
```

```
Composición de aminoácidos (AAC) efectores archaea dataset 4, sin valores
atípicos.
Estadísticas.

               X0          X1          X2          X3          X4          X5  \
count  437.000000  437.000000  437.000000  437.000000  437.000000  437.000000
mean     9.565954    6.342469    2.549664    6.410863    0.642851    7.299838
std      4.341430    2.743115    1.893682    2.652882    0.757349    3.493775
min      0.000000    0.000000    0.000000    1.024000    0.000000    0.000000
25%      6.400000    4.420000    1.141000    4.478000    0.000000    4.630000
50%      8.883000    6.237000    2.105000    6.107000    0.420000    7.339000
75%     12.587000    8.209000    3.834000    8.293000    1.036000    9.653000
max     22.436000   15.079000    9.009000   14.231000    3.433000   17.073000

               X6          X7          X8          X9         X10         X11  \
count  437.000000  437.000000  437.000000  437.000000  437.000000  437.000000
mean     2.369346    7.664156    1.831568    5.439471   11.187069    4.174126
std      1.392059    2.667465    1.107682    3.212299    3.276477    4.127676
min      0.000000    1.961000    0.000000    0.000000    3.409000    0.000000
25%      1.415000    5.600000    0.885000    2.927000    8.876000    1.049000
50%      2.283000    7.722000    1.765000    4.851000   11.220000    2.381000
75%      3.150000    9.463000    2.604000    7.143000   13.291000    7.194000
max      7.042000   15.198000    5.405000   15.714000   21.311000   17.143000

              X12         X13         X14         X15         X16         X17  \
count  437.000000  437.000000  437.000000  437.000000  437.000000  437.000000
mean     1.904304    3.811240    4.103316    5.819082    5.727574    1.177423
```

std	1.072736	1.754891	1.590671	2.039205	1.856422	0.917600
min	0.000000	0.000000	0.000000	1.053000	0.962000	0.000000
25%	1.053000	2.586000	3.053000	4.348000	4.500000	0.559000
50%	1.667000	3.779000	4.035000	5.618000	5.714000	1.044000
75%	2.532000	4.878000	5.000000	7.035000	6.875000	1.703000
max	5.435000	9.259000	9.231000	12.366000	11.728000	4.082000

	X18	X19
count	437.000000	437.000000
mean	3.233224	8.746551
std	1.468882	3.461245
min	0.000000	0.000000
25%	2.229000	5.882000
50%	3.211000	8.302000
75%	4.032000	11.227000
max	8.295000	18.657000

no_efectores
Composición de aminoácidos (AAC) no_efectores archaea dataset 4, sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9	\
0	2.865	5.469	7.292	6.510	1.562	5.469	2.865	5.469	2.083	8.854	
1	8.592	7.637	2.864	6.205	2.148	5.489	3.580	7.399	2.387	3.580	
2	10.526	9.023	2.632	9.023	1.128	6.767	1.880	7.143	1.128	5.263	
3	5.714	6.667	4.762	5.714	0.000	13.333	4.762	8.571	0.952	6.667	
4	13.580	9.877	0.823	9.465	1.235	8.642	0.000	11.111	2.881	1.235	
..	…	…	…	…	…	…	…	…	…	…	
495	6.392	7.629	3.402	8.763	1.031	10.515	3.711	7.526	2.268	6.392	
496	14.932	5.656	1.584	5.656	1.131	2.941	0.226	12.670	0.905	2.941	
497	5.357	8.929	2.679	3.571	0.000	10.714	3.571	3.571	1.786	8.036	
498	5.797	5.797	1.449	17.391	0.000	8.696	7.246	5.797	2.899	7.246	
499	10.784	6.863	1.961	9.804	0.000	15.686	2.941	2.941	4.902	1.961	

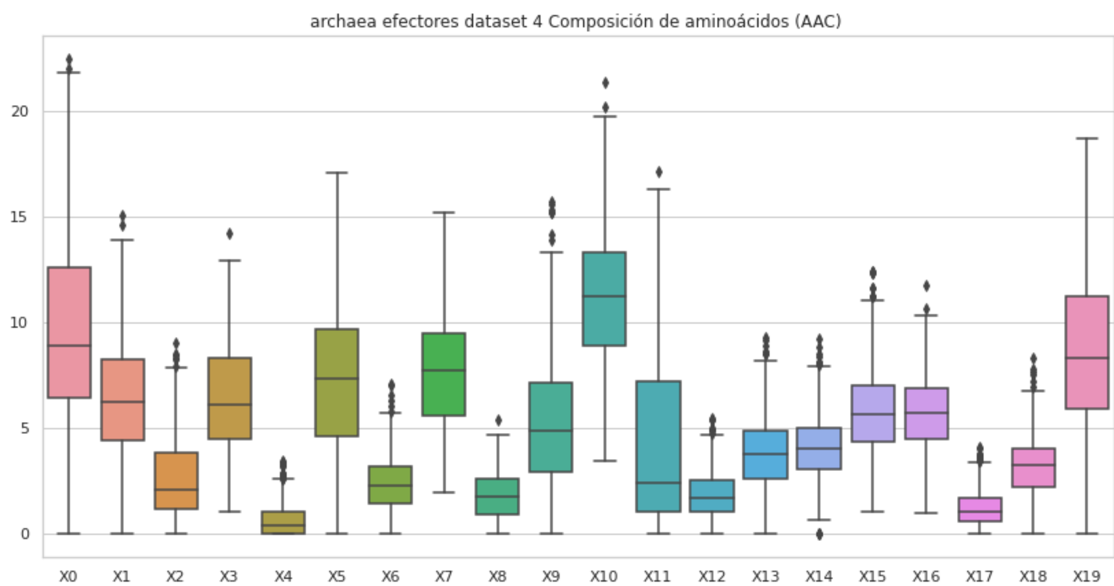
	…	X11	X12	X13	X14	X15	X16	X17	X18	X19	\
0	…	8.073	2.604	5.208	2.344	10.417	3.385	0.781	3.906	3.906	
1	…	1.432	5.251	4.296	7.637	7.160	5.012	0.716	2.148	7.399	
2	…	1.504	0.376	3.759	4.887	6.391	7.519	1.128	1.880	8.271	
3	…	9.524	1.905	2.857	0.952	4.762	6.667	0.952	4.762	4.762	
4	…	0.412	0.412	3.704	4.527	7.819	4.938	1.235	2.469	8.642	
..	…	…	…	…	…	…	…	…	…	…	
495	…	2.784	2.784	3.196	5.052	5.876	6.082	0.619	2.062	6.495	
496	…	0.226	0.452	2.489	6.109	5.882	8.145	1.584	2.941	11.312	
497	…	5.357	4.464	0.893	1.786	8.036	9.821	0.893	5.357	3.571	
498	…	4.348	1.449	0.000	0.000	2.899	4.348	0.000	2.899	5.797	

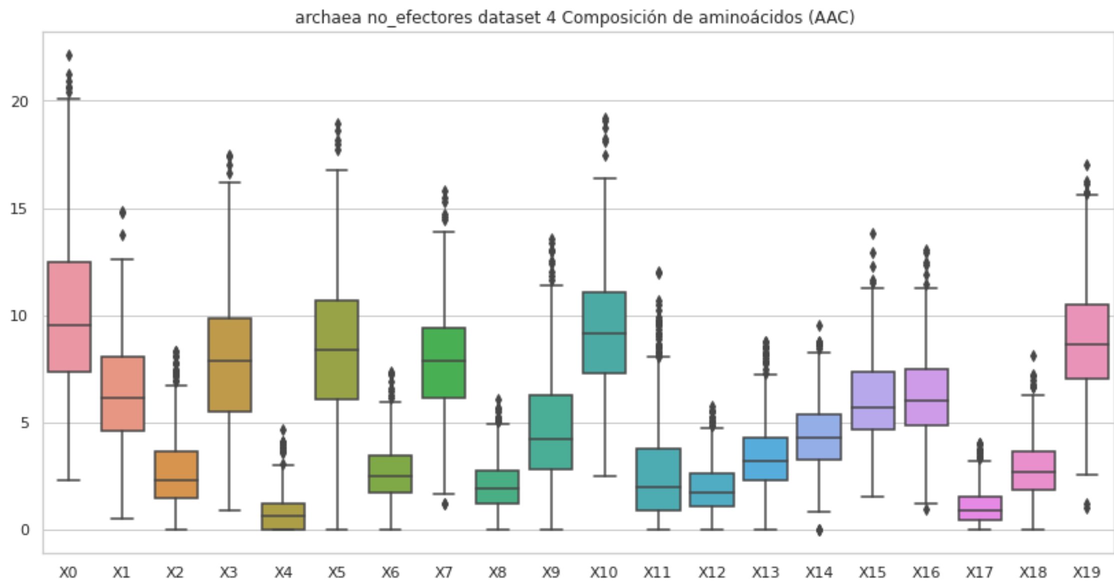
```
499  …  0.000  2.941  3.922  5.882   2.941  3.922  0.980  2.941   8.824

             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

[418 rows x 21 columns]
```

```
Composición de aminoácidos (AAC) no_efectores archaea dataset 4, sin valores
atípicos.
Estadísticas.

              X0          X1          X2          X3          X4          X5  \
count  418.000000  418.000000  418.000000  418.000000  418.000000  418.000000
mean     9.975524    6.355730    2.691055    7.756323    0.828072    8.368433
std      3.888779    2.564144    1.718013    3.187574    0.849850    3.454789
min      2.304000    0.526000    0.000000    0.901000    0.000000    0.000000
25%      7.330750    4.589750    1.474500    5.512250    0.000000    6.082250
50%      9.544000    6.151500    2.328000    7.889500    0.636000    8.366000
75%     12.448000    8.086250    3.625750    9.826250    1.219000   10.710750
max     22.137000   14.815000    8.333000   17.442000    4.688000   18.966000

              X6          X7          X8          X9         X10         X11  \
count  418.000000  418.000000  418.000000  418.000000  418.000000  418.000000
mean     2.607981    7.896402    2.022986    4.731809    9.308093    2.789455
std      1.421913    2.613660    1.204962    2.629776    2.993198    2.557335
min      0.000000    1.220000    0.000000    0.000000    2.469000    0.000000
25%      1.714500    6.133250    1.184750    2.833500    7.312500    0.908250
50%      2.497000    7.879000    1.897500    4.203000    9.126000    2.014500
75%      3.428000    9.392750    2.728000    6.279250   11.076250    3.771750
max      7.339000   15.789000    6.061000   13.542000   19.178000   12.012000

             X12         X13         X14         X15         X16         X17  \
count  418.000000  418.000000  418.000000  418.000000  418.000000  418.000000
mean     1.980541    3.427653    4.386706    6.052555    6.156215    1.053675
std      1.123606    1.748372    1.681030    2.124182    2.042782    0.872493
min      0.000000    0.000000    0.000000    1.562000    0.943000    0.000000
25%      1.105500    2.326500    3.281500    4.653250    4.877500    0.444000
```

50%	1.741500	3.211500	4.270500	5.721000	6.042000	0.897500
75%	2.603250	4.312250	5.348750	7.333500	7.467000	1.562000
max	5.747000	8.772000	9.524000	13.814000	13.043000	4.000000

	X18	X19
count	418.000000	418.000000
mean	2.816060	8.794756
std	1.382235	2.676954
min	0.000000	1.042000
25%	1.863000	7.031500
50%	2.679500	8.670500
75%	3.674500	10.470500
max	8.097000	17.012000





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
4, con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.018092  0.000441  0.004413  0.003530  0.006178  0.018533  0.000441
1    0.024391  0.003181  0.020149  0.025452  0.036056  0.026512  0.008484
2    0.051462  0.000000  0.022515  0.054679  0.012866  0.028948  0.003216
3    0.000000  0.020274  0.030410  0.081094  0.010137  0.050684  0.000000
4    0.019840  0.001984  0.024799  0.021824  0.033727  0.022816  0.004960
..        …         …         …         …         …         …         …
495  0.012481  0.000000  0.037444  0.024963  0.018722  0.062407  0.006241
496  0.037726  0.004716  0.047157  0.056589  0.061305  0.037726  0.014147
497  0.035952  0.001498  0.014980  0.004494  0.023968  0.023968  0.004494
498  0.020465  0.000379  0.005685  0.004169  0.008717  0.017812  0.001137
499  0.027067  0.000588  0.010003  0.015299  0.012945  0.024125  0.000000

           X7        X8        X9  …       X74       X75       X76  \
0    0.010149  0.000883  0.021181  …  0.023480  0.008249  0.025712
1    0.060448  0.041359  0.043480  … -0.008698 -0.012267  0.009733
2    0.025731  0.032164  0.041813  …  0.023974  0.037101  0.017928
3    0.040547  0.010137  0.101368  … -0.033329 -0.018181 -0.034435
4    0.061503  0.036703  0.039679  …  0.019156  0.001858  0.009389
..        …         …         …  …       …         …         …
495  0.049926  0.043685  0.037444  … -0.032674 -0.018169  0.032768
496  0.108462  0.150904  0.080168  …  0.041887  0.037180  0.030294
497  0.002996  0.004494  0.041943  …  0.018907  0.005675  0.027654
498  0.011370  0.001895  0.019707  …  0.016208  0.002077  0.025630
499  0.014710  0.010592  0.033540  …  0.013057  0.006644  0.004113

          X77       X78       X79       X80       X81       X82        X83
0    0.009754 -0.001881  0.020667  0.016523  0.000978  0.024025  efectores
1   -0.002395  0.005594  0.002704  0.019168  0.006180  0.000970  efectores
2   -0.000082  0.028069  0.037572 -0.025853 -0.029535  0.044225  efectores
3    0.085894  0.065426  0.060645  0.052053  0.075805  0.042504  efectores
4    0.004815  0.006009  0.001047  0.019681  0.011235 -0.007660  efectores
..        …         …         …         …         …         …          …
495 -0.013364  0.030969 -0.001823  0.007404  0.030916 -0.012257  efectores
496  0.042264  0.049625 -0.023143 -0.041804 -0.009255  0.035541  efectores
```

```
497  0.005918 -0.007734  0.025497  0.026112  0.002262  0.019850  efectores
498  0.015674 -0.000714  0.009969  0.016575  0.001281  0.021888  efectores
499  0.008085  0.011291  0.009055 -0.001652 -0.001615  0.012305  efectores

[500 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro_mass efectores archaea dataset 4, con valores atípicos.
Estadísticas.

```
               X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.032004    0.004020    0.027855    0.033882    0.015793    0.028018
std      0.014872    0.007046    0.020219    0.026721    0.013136    0.013679
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.004342
25%      0.021055    0.000000    0.010158    0.009959    0.007359    0.017244
50%      0.028669    0.001283    0.024396    0.028013    0.012312    0.025652
75%      0.040426    0.005483    0.040397    0.053133    0.020453    0.035029
max      0.097984    0.082527    0.115064    0.134933    0.096890    0.112948

               X6          X7          X8          X9  …         X73  \
count  500.000000  500.000000  500.000000  500.000000  …  500.000000
mean     0.008397    0.027851    0.025041    0.044970  …    0.013444
std      0.009424    0.029502    0.031668    0.027484  …    0.020529
min      0.000000    0.000000    0.000000    0.005594  …   -0.164656
25%      0.002786    0.007736    0.002266    0.024722  …    0.003475
50%      0.006313    0.016723    0.009531    0.038140  …    0.015603
75%      0.010890    0.039377    0.039362    0.055896  …    0.024566
max      0.130252    0.220072    0.192563    0.154209  …    0.067906

              X74         X75         X76         X77         X78         X79  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.004895    0.009665    0.012950    0.007243    0.011275    0.014636
std      0.027674    0.027410    0.018127    0.028868    0.026052    0.019624
min     -0.138602   -0.166557   -0.100955   -0.084728   -0.080395   -0.095537
25%     -0.007541   -0.001202    0.003433   -0.005251   -0.000274    0.004931
50%      0.007160    0.004480    0.014941    0.009139    0.006718    0.015754
75%      0.016410    0.016268    0.023310    0.018950    0.020911    0.026377
max      0.163579    0.233050    0.102497    0.318613    0.207266    0.083651

              X80         X81         X82
count  500.000000  500.000000  500.000000
mean     0.003688    0.010156    0.015016
std      0.029292    0.021882    0.017115
min     -0.137781   -0.088231   -0.058597
25%     -0.004569   -0.000136    0.005241
50%      0.008352    0.006876    0.016104
```

```
75%      0.017903    0.019912    0.024441
max      0.130458    0.114410    0.074680

[8 rows x 83 columns]
```

no_efectores
Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores archaea
dataset 4, con valores atípicos.
Valores del documento csv.

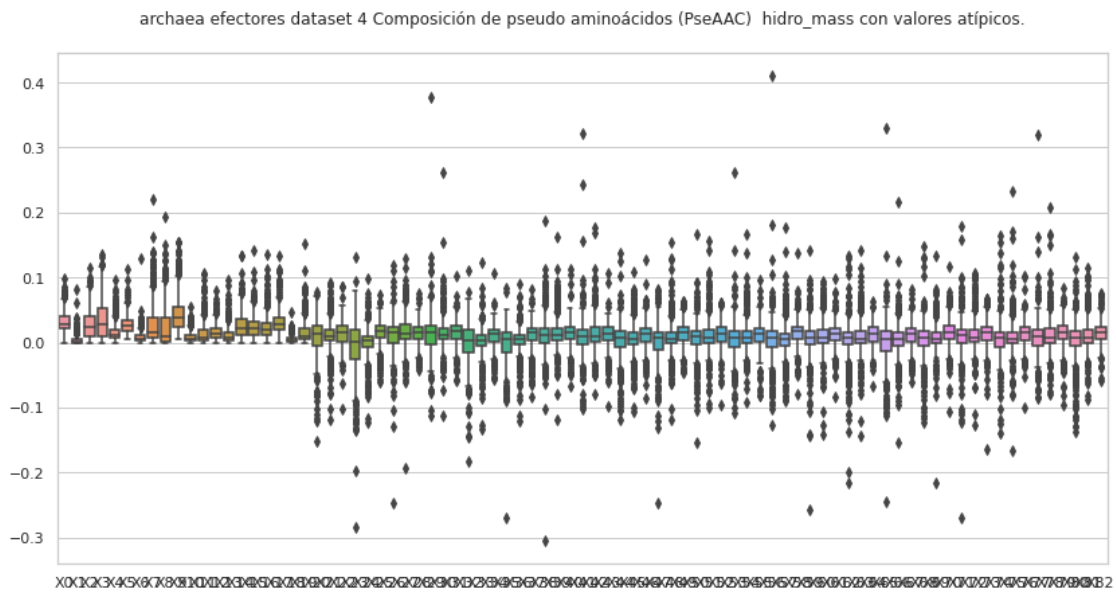
```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.023975  0.013077  0.054488  0.045770  0.043591  0.045770  0.017436
1    0.044634  0.011158  0.032236  0.028516  0.022317  0.038435  0.012398
2    0.051715  0.005541  0.044327  0.033245  0.018470  0.035092  0.005541
3    0.021846  0.000000  0.021846  0.050975  0.010923  0.032770  0.003641
4    0.043426  0.003948  0.030267  0.027635  0.011844  0.035531  0.009212
..        …         …         …         …         …         …         …
495  0.026329  0.004247  0.036096  0.043315  0.013164  0.031000  0.009343
496  0.023558  0.001785  0.008923  0.004640  0.003926  0.019988  0.001428
497  0.041462  0.000000  0.027641  0.082924  0.006910  0.027641  0.013821
498  0.029495  0.000000  0.088485  0.044243  0.000000  0.029495  0.014748
499  0.062614  0.000000  0.056922  0.091075  0.022769  0.017077  0.028461

           X7        X8        X9  …       X74       X75       X76  \
0    0.074104  0.067565  0.091540  … -0.015817 -0.038610  0.026719
1    0.018597  0.007439  0.047113  … -0.008080  0.001165  0.039050
2    0.025857  0.007388  0.048021  … -0.004345 -0.004513  0.032675
3    0.025488  0.036411  0.021846  …  0.025972  0.027509 -0.008526
4    0.003948  0.001316  0.022371  …  0.006782  0.013409  0.026907
..        …         …         … …       …         …         …
495  0.026329  0.011466  0.030576  …  0.004144  0.012662  0.013546
496  0.004640  0.000357  0.019274  …  0.002159  0.003162  0.025794
497  0.062193  0.041462  0.089835  … -0.009796 -0.041058  0.026794
498  0.036869  0.022121  0.081112  …  0.037598  0.095299 -0.015672
499  0.011384  0.000000  0.056922  …  0.008787 -0.029239  0.029508

          X77       X78       X79       X80       X81       X82          X83
0   -0.014233 -0.000812 -0.003252  0.000911  0.038180 -0.001601  no_efectores
1    0.002059  0.008439  0.025268  0.001672  0.001550  0.013470  no_efectores
2   -0.045728 -0.016406  0.031993 -0.008243  0.012032  0.019934  no_efectores
3   -0.022567 -0.024531 -0.017774  0.002891  0.002016  0.014504  no_efectores
4   -0.012865  0.011435 -0.000119 -0.011685  0.006652  0.008253  no_efectores
..        …         …         …         …         …         …            …
495  0.004265  0.018724  0.013721  0.009201  0.022157  0.006117  no_efectores
496  0.005380 -0.000686  0.023851  0.005245  0.001480  0.026280  no_efectores
497  0.047769  0.053367  0.082871 -0.017099 -0.045872  0.010241  no_efectores
```

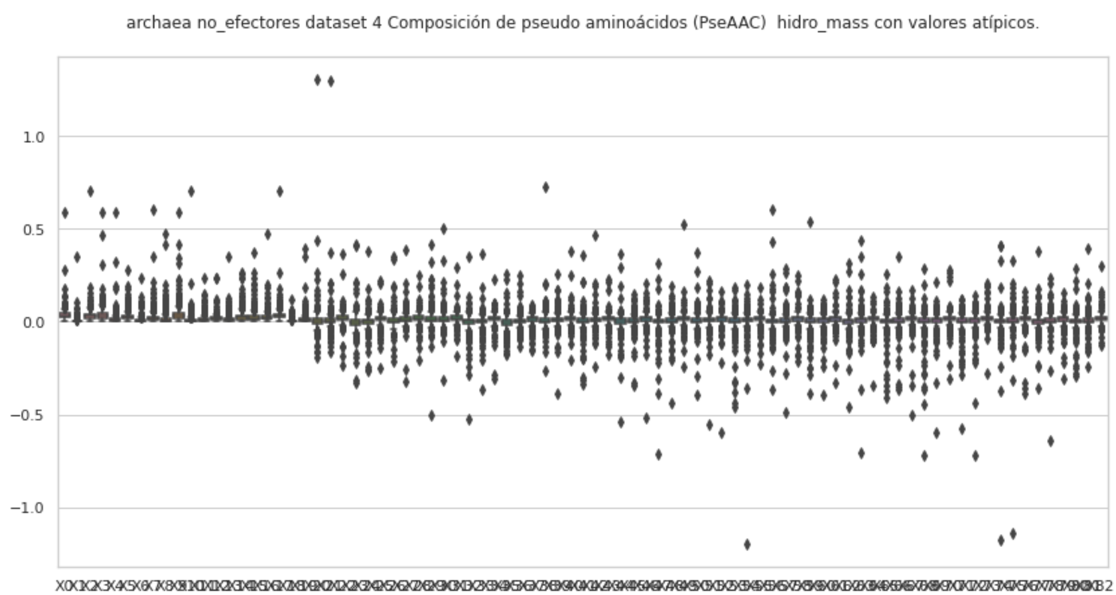
```
498 -0.059928 -0.035440 -0.031154  0.026201  0.071867  0.002258  no_efectores
499 -0.047207  0.010240  0.009531 -0.039341  0.019780 -0.030533  no_efectores

[500 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores archaea
dataset 4, con valores atípicos.
Estadísticas.

```
               X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.036809    0.005379    0.032681    0.037202    0.017063    0.030643
std      0.032588    0.018046    0.036898    0.041064    0.033709    0.022423
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.023513    0.000000    0.016898    0.018418    0.006779    0.019971
50%      0.032655    0.002137    0.029765    0.032047    0.011092    0.026870
75%      0.043635    0.005291    0.040479    0.046998    0.018276    0.034622
max      0.589697    0.353818    0.707636    0.589697    0.589697    0.278277

               X6          X7          X8          X9  …         X73  \
count  500.000000  500.000000  500.000000  500.000000  …  500.000000
mean     0.010474    0.023936    0.017949    0.041242  …    0.018625
std      0.017098    0.038551    0.037078    0.044482  …    0.027431
min      0.000000    0.000000    0.000000    0.000000  …   -0.276885
25%      0.002990    0.008307    0.002307    0.021367  …    0.009427
50%      0.006556    0.015566    0.007056    0.032469  …    0.018897
75%      0.011908    0.026835    0.018281    0.047125  …    0.027206
max      0.235879    0.602933    0.471757    0.589697  …    0.204050

              X74         X75         X76         X77         X78         X79  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.000408    0.006166    0.013730    0.001325    0.007139    0.014461
std      0.068664    0.061841    0.037420    0.038745    0.041056    0.030477
min     -1.177143   -1.139009   -0.388831   -0.358621   -0.637340   -0.307292
25%     -0.006705   -0.002183    0.007968   -0.007401   -0.002669    0.006662
50%      0.003616    0.005624    0.017275    0.003467    0.005730    0.017374
75%      0.012988    0.016801    0.027309    0.013085    0.018796    0.026488
max      0.410794    0.330486    0.161076    0.379155    0.234022    0.154776

              X80         X81         X82
count  500.000000  500.000000  500.000000
mean     0.000821    0.008070    0.017788
std      0.038600    0.034451    0.026022
min     -0.293733   -0.245058   -0.127497
25%     -0.006677   -0.000929    0.007232
50%      0.003841    0.006400    0.018075
75%      0.013013    0.016162    0.027611
```

```
max       0.283173    0.390861    0.297886

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.018092  0.000441  0.004413  0.003530  0.006178  0.018533  0.000441
1    0.024391  0.003181  0.020149  0.025452  0.036056  0.026512  0.008484
2    0.051462  0.000000  0.022515  0.054679  0.012866  0.028948  0.003216
4    0.019840  0.001984  0.024799  0.021824  0.033727  0.022816  0.004960
5    0.026105  0.000000  0.018646  0.013674  0.012431  0.031077  0.007459
..        …         …         …         …         …         …         …
493  0.022816  0.000000  0.007081  0.002360  0.008654  0.015735  0.004721
494  0.042966  0.000880  0.023244  0.017433  0.003346  0.024476  0.003874
497  0.035952  0.001498  0.014980  0.004494  0.023968  0.023968  0.004494
498  0.020465  0.000379  0.005685  0.004169  0.008717  0.017812  0.001137
499  0.027067  0.000588  0.010003  0.015299  0.012945  0.024125  0.000000

           X7        X8        X9  …       X74       X75       X76  \
0    0.010149  0.000883  0.021181  …  0.023480  0.008249  0.025712
1    0.060448  0.041359  0.043480  … -0.008698 -0.012267  0.009733
2    0.025731  0.032164  0.041813  …  0.023974  0.037101  0.017928
4    0.061503  0.036703  0.039679  …  0.019156  0.001858  0.009389
5    0.008702  0.000000  0.033564  … -0.001075  0.001842  0.003403
..        …         …         … …       …         …         …
493  0.007868  0.003147  0.016522  …  0.010913 -0.004112  0.025653
494  0.002465  0.001057  0.020426  …  0.008204  0.013950  0.028931
497  0.002996  0.004494  0.041943  …  0.018907  0.005675  0.027654
498  0.011370  0.001895  0.019707  …  0.016208  0.002077  0.025630
499  0.014710  0.010592  0.033540  …  0.013057  0.006644  0.004113

          X77       X78       X79       X80       X81       X82        X83
0    0.009754 -0.001881  0.020667  0.016523  0.000978  0.024025  efectores
1   -0.002395  0.005594  0.002704  0.019168  0.006180  0.000970  efectores
2   -0.000082  0.028069  0.037572 -0.025853 -0.029535  0.044225  efectores
4    0.004815  0.006009  0.001047  0.019681  0.011235 -0.007660  efectores
5    0.008444 -0.003178  0.013596  0.004460 -0.000470  0.018938  efectores
..        …         …         …         …         …         …          …
493  0.014039  0.006929  0.027267  0.012704  0.000350  0.032565  efectores
494  0.003608  0.010511  0.027293 -0.001718  0.009431  0.029524  efectores
497  0.005918 -0.007734  0.025497  0.026112  0.002262  0.019850  efectores
498  0.015674 -0.000714  0.009969  0.016575  0.001281  0.021888  efectores
499  0.008085  0.011291  0.009055 -0.001652 -0.001615  0.012305  efectores
```

```
[390 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro_mass efectores archaea dataset 4, sin valores atípicos.
Estadísticas.

```
               X0          X1          X2          X3          X4          X5  \
count  390.000000  390.000000  390.000000  390.000000  390.000000  390.000000
mean     0.029659    0.002253    0.021694    0.025309    0.013077    0.024415
std      0.011672    0.003454    0.015200    0.020242    0.008958    0.010177
min      0.002178    0.000000    0.000000    0.000000    0.000000    0.004714
25%      0.020561    0.000000    0.008921    0.006527    0.006608    0.016242
50%      0.027237    0.000584    0.019011    0.020521    0.010682    0.022977
75%      0.037178    0.003035    0.032151    0.039594    0.017577    0.029601
max      0.070311    0.017653    0.072870    0.084240    0.047535    0.059060

               X6          X7          X8          X9  …         X73  \
count  390.000000  390.000000  390.000000  390.000000  …  390.000000
mean     0.006370    0.019676    0.015712    0.035763  …    0.015758
std      0.005490    0.018937    0.021237    0.017577  …    0.014736
min      0.000000    0.000000    0.000000    0.005594  …   -0.031414
25%      0.002117    0.006484    0.001773    0.022325  …    0.007817
50%      0.005462    0.011714    0.005666    0.033073  …    0.017115
75%      0.008811    0.026566    0.021810    0.044674  …    0.024413
max      0.029273    0.101424    0.108882    0.101713  …    0.060128

              X74         X75         X76         X77         X78         X79  \
count  390.000000  390.000000  390.000000  390.000000  390.000000  390.000000
mean     0.005087    0.007667    0.015182    0.007755    0.009078    0.015980
std      0.016809    0.014898    0.013628    0.016313    0.016136    0.013530
min     -0.072242   -0.036866   -0.029684   -0.075873   -0.051336   -0.035014
25%     -0.004675   -0.001023    0.006929   -0.000836   -0.000058    0.007618
50%      0.007699    0.003644    0.016834    0.009771    0.006000    0.016167
75%      0.016112    0.012946    0.023884    0.018204    0.017247    0.025418
max      0.066785    0.061648    0.059603    0.052059    0.075790    0.060589

              X80         X81         X82
count  390.000000  390.000000  390.000000
mean     0.007465    0.008888    0.016141
std      0.016165    0.013395    0.013009
min     -0.072273   -0.030434   -0.021326
25%     -0.001443    0.000299    0.007962
50%      0.009588    0.006357    0.016999
75%      0.016825    0.016877    0.024003
max      0.077700    0.064452    0.060500

[8 rows x 83 columns]
```

no_efectores
Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores archaea
dataset 4, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.023975  0.013077  0.054488  0.045770  0.043591  0.045770  0.017436
1    0.044634  0.011158  0.032236  0.028516  0.022317  0.038435  0.012398
2    0.051715  0.005541  0.044327  0.033245  0.018470  0.035092  0.005541
3    0.021846  0.000000  0.021846  0.050975  0.010923  0.032770  0.003641
4    0.043426  0.003948  0.030267  0.027635  0.011844  0.035531  0.009212
..        …         …         …         …         …         …         …
495  0.026329  0.004247  0.036096  0.043315  0.013164  0.031000  0.009343
496  0.023558  0.001785  0.008923  0.004640  0.003926  0.019988  0.001428
497  0.041462  0.000000  0.027641  0.082924  0.006910  0.027641  0.013821
498  0.029495  0.000000  0.088485  0.044243  0.000000  0.029495  0.014748
499  0.062614  0.000000  0.056922  0.091075  0.022769  0.017077  0.028461

           X7        X8        X9  …       X74       X75       X76  \
0    0.074104  0.067565  0.091540  … -0.015817 -0.038610  0.026719
1    0.018597  0.007439  0.047113  … -0.008080  0.001165  0.039050
2    0.025857  0.007388  0.048021  … -0.004345 -0.004513  0.032675
3    0.025488  0.036411  0.021846  …  0.025972  0.027509 -0.008526
4    0.003948  0.001316  0.022371  …  0.006782  0.013409  0.026907
..        …         …         …  …       …         …         …
495  0.026329  0.011466  0.030576  …  0.004144  0.012662  0.013546
496  0.004640  0.000357  0.019274  …  0.002159  0.003162  0.025794
497  0.062193  0.041462  0.089835  … -0.009796 -0.041058  0.026794
498  0.036869  0.022121  0.081112  …  0.037598  0.095299 -0.015672
499  0.011384  0.000000  0.056922  …  0.008787 -0.029239  0.029508

          X77       X78       X79       X80       X81       X82           X83
0   -0.014233 -0.000812 -0.003252  0.000911  0.038180 -0.001601  no_efectores
1    0.002059  0.008439  0.025268  0.001672  0.001550  0.013470  no_efectores
2   -0.045728 -0.016406  0.031993 -0.008243  0.012032  0.019934  no_efectores
3   -0.022567 -0.024531 -0.017774  0.002891  0.002016  0.014504  no_efectores
4   -0.012865  0.011435 -0.000119 -0.011685  0.006652  0.008253  no_efectores
..        …         …         …         …         …         …           …
495  0.004265  0.018724  0.013721  0.009201  0.022157  0.006117  no_efectores
496  0.005380 -0.000686  0.023851  0.005245  0.001480  0.026280  no_efectores
497  0.047769  0.053367  0.082871 -0.017099 -0.045872  0.010241  no_efectores
498 -0.059928 -0.035440 -0.031154  0.026201  0.071867  0.002258  no_efectores
499 -0.047207  0.010240  0.009531 -0.039341  0.019780 -0.030533  no_efectores

[462 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores archaea dataset 4, sin valores atípicos.
Estadísticas.

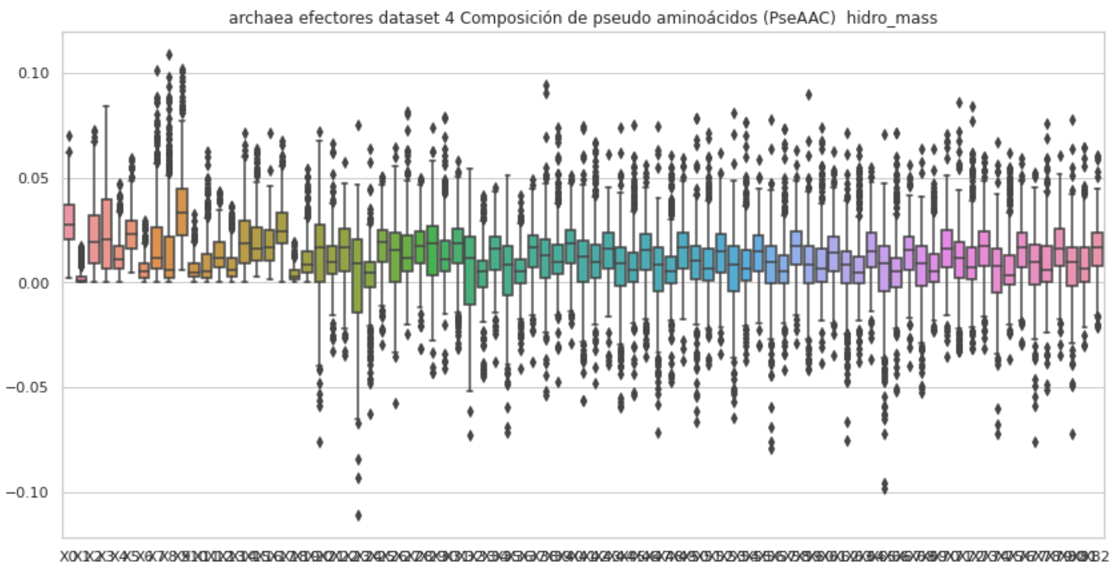
	X0	X1	X2	X3	X4	X5	\
count	462.000000	462.000000	462.000000	462.000000	462.000000	462.000000	
mean	0.033564	0.003664	0.029181	0.032426	0.013111	0.027274	
std	0.013665	0.005309	0.016793	0.020145	0.010510	0.011179	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.002271	
25%	0.023445	0.000000	0.016617	0.018168	0.006439	0.019585	
50%	0.032297	0.002054	0.029174	0.030650	0.010874	0.026155	
75%	0.042385	0.004672	0.039386	0.045158	0.016766	0.032648	
max	0.092818	0.043340	0.089471	0.139701	0.078658	0.080220	

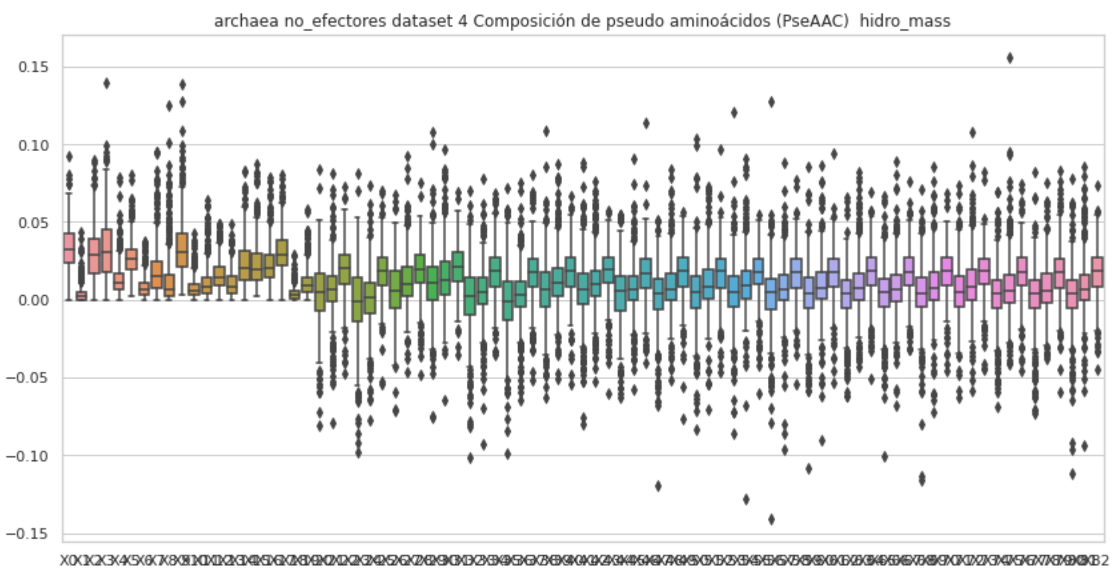
	X6	X7	X8	X9	…	X73	\
count	462.000000	462.000000	462.000000	462.000000	…	462.000000	
mean	0.008097	0.018674	0.012669	0.034152	…	0.017572	
std	0.007259	0.016019	0.016800	0.018768	…	0.014783	
min	0.000000	0.000000	0.000000	0.003190	…	-0.044677	
25%	0.002861	0.007838	0.002287	0.020976	…	0.010083	
50%	0.006391	0.014849	0.006434	0.030765	…	0.018897	
75%	0.010639	0.024707	0.015794	0.042718	…	0.026699	
max	0.037199	0.095396	0.125424	0.138587	…	0.081999	

	X74	X75	X76	X77	X78	X79	\
count	462.000000	462.000000	462.000000	462.000000	462.000000	462.000000	
mean	0.002914	0.008439	0.016921	0.002559	0.007658	0.017467	
std	0.018273	0.018600	0.016059	0.019189	0.016708	0.016876	
min	-0.068527	-0.041058	-0.045956	-0.073471	-0.038213	-0.059378	
25%	-0.005479	-0.001593	0.008820	-0.005586	-0.001718	0.007592	
50%	0.003815	0.005857	0.017978	0.003927	0.005764	0.018215	
75%	0.012854	0.015863	0.027272	0.012783	0.017066	0.026596	
max	0.069614	0.155902	0.074050	0.081959	0.075437	0.082871	

	X80	X81	X82
count	462.000000	462.000000	462.000000
mean	0.002709	0.008228	0.017255
std	0.019319	0.017663	0.015230
min	-0.111373	-0.093667	-0.045026
25%	-0.005391	-0.000220	0.008183
50%	0.003852	0.006439	0.018347
75%	0.012556	0.015889	0.027248
max	0.077948	0.085391	0.073599

[8 rows x 83 columns]





4 Composición de pseudo aminoácidos (PseAAC) mass

[7]:
```python
#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 4,
con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.030765  0.000750  0.007504  0.006003  0.010505  0.031515  0.000750
1    0.045758  0.005968  0.037800  0.047747  0.067642  0.049737  0.015916
2    0.061291  0.000000  0.026815  0.065122  0.015323  0.034476  0.003831
3    0.000000  0.030365  0.045547  0.121459  0.015182  0.075912  0.000000
4    0.040318  0.004032  0.050398  0.044350  0.068541  0.046366  0.010080
..        …         …         …         …         …         …         …
495  0.011428  0.000000  0.034283  0.022855  0.017141  0.057138  0.005714
496  0.038905  0.004863  0.048631  0.058358  0.063221  0.038905  0.014589
497  0.061577  0.002566  0.025657  0.007697  0.041051  0.041051  0.007697
498  0.039401  0.000730  0.010945  0.008026  0.016782  0.034294  0.002189
499  0.045292  0.000985  0.016738  0.025600  0.021661  0.040369  0.000000

           X7        X8        X9  …       X32       X33       X34  \
```

0	0.017258	0.001501	0.036017	…	0.011503	0.027663	0.039157
1	0.113400	0.077590	0.081568	…	-0.029061	-0.007623	0.029471
2	0.030646	0.038307	0.049799	…	0.003156	0.014485	0.037015
3	0.060729	0.015182	0.151824	…	-0.067590	-0.015335	0.002669
4	0.124987	0.074589	0.080637	…	-0.027392	0.037251	0.029525
..	…	…	…	…	…	…	…
495	0.045710	0.039996	0.034283	…	0.061268	-0.024893	0.024920
496	0.111852	0.155621	0.082673	…	0.018379	-0.006112	-0.007217
497	0.005131	0.007697	0.071839	…	0.011226	0.005262	0.046463
498	0.021890	0.003648	0.037942	…	0.028004	0.034414	0.041931
499	0.024615	0.017723	0.056122	…	0.008645	0.031114	0.022794

	X35	X36	X37	X38	X39	X40	X41
0	0.043664	0.052347	0.037519	0.043723	0.035143	0.040855	efectores
1	-0.024477	0.002469	-0.011768	0.018259	0.005073	0.001819	efectores
2	0.041446	0.023355	0.046412	0.021352	0.044749	0.052672	efectores
3	-0.012044	-0.036851	0.056991	-0.051574	0.090831	0.063660	efectores
4	-0.000687	-0.002120	-0.006956	0.019081	0.002127	-0.015566	efectores
..	…	…	…	…	…	…	…
495	-0.008265	0.053850	0.029233	0.030001	-0.001669	-0.011222	efectores
496	0.007483	-0.003890	-0.007609	0.031241	-0.023867	0.036652	efectores
497	0.004415	0.033338	0.009254	0.047364	0.043671	0.033998	efectores
498	0.018878	0.038444	0.035524	0.049345	0.019194	0.042141	efectores
499	0.033967	0.043901	0.025492	0.006882	0.015152	0.020590	efectores

[500 rows x 42 columns]

Composición de pseudo aminoácidos (PseAAC) mass efectores archaea dataset 4,
con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.045717	0.005029	0.038069	0.047211	0.022677	0.039235	
std	0.016288	0.007461	0.024335	0.036461	0.016391	0.012217	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.004694	
25%	0.034032	0.000000	0.019224	0.017326	0.011539	0.030317	
50%	0.044327	0.002111	0.034409	0.038204	0.018786	0.037740	
75%	0.054245	0.007475	0.051980	0.068263	0.029537	0.046538	
max	0.113294	0.063629	0.117787	0.177945	0.090919	0.099911	

	X6	X7	X8	X9	…	X31	\
count	500.000000	500.000000	500.000000	500.000000	…	500.000000	
mean	0.011384	0.038293	0.033397	0.064789	…	0.015678	
std	0.010728	0.034252	0.039114	0.032804	…	0.025430	
min	0.000000	0.000000	0.000000	0.005998	…	-0.103055	
25%	0.004203	0.011999	0.003674	0.040942	…	0.000566	

```
50%       0.008943    0.026122    0.015374    0.058005  …    0.020798
75%       0.015883    0.053648    0.057631    0.085569  …    0.033050
max       0.140808    0.169678    0.162933    0.258459  …    0.103468

                X32         X33         X34         X35         X36         X37  \
count   500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean      0.022432    0.018169    0.017818    0.015976    0.021054    0.019961
std       0.024422    0.024791    0.024753    0.025919    0.024207    0.026554
min      -0.067590   -0.183487   -0.092796   -0.083938   -0.082459   -0.128548
25%       0.008619    0.005199    0.002664    0.002582    0.008002    0.005207
50%       0.025944    0.022097    0.022720    0.021452    0.026350    0.024484
75%       0.039247    0.034306    0.034753    0.033974    0.035294    0.035727
max       0.116950    0.128552    0.149861    0.097732    0.109619    0.101308

                X38         X39         X40
count   500.000000  500.000000  500.000000
mean      0.018950    0.021153    0.022209
std       0.024202    0.027010    0.025905
min      -0.077984   -0.172500   -0.082459
25%       0.005177    0.007634    0.006997
50%       0.021948    0.024462    0.024695
75%       0.034357    0.037075    0.036430
max       0.091410    0.102306    0.232347

[8 rows x 41 columns]
```

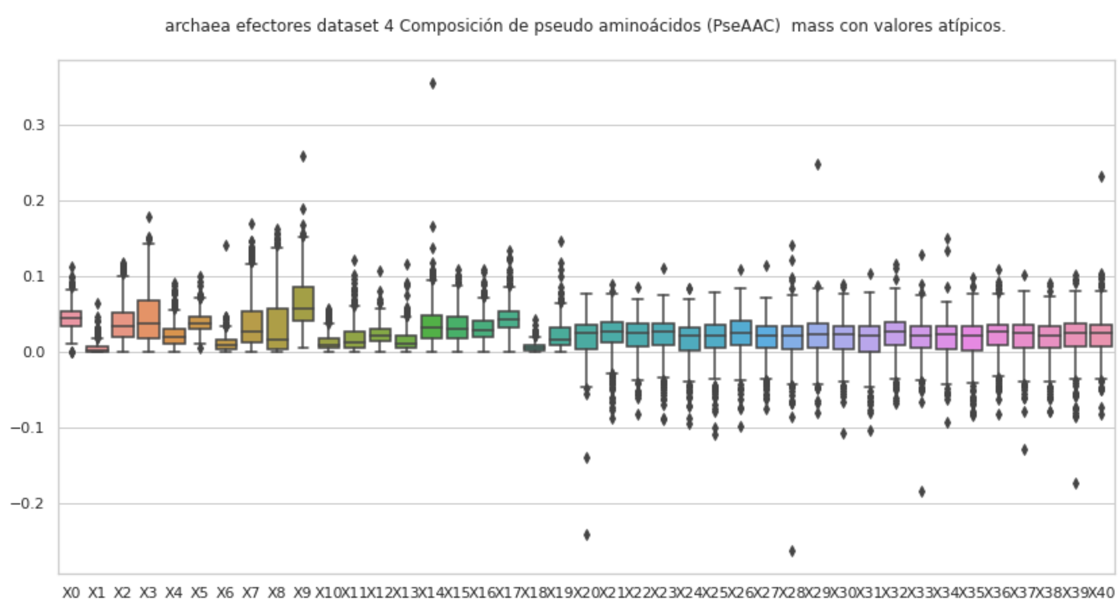
```
no_efectores
Composición de pseudo aminoácidos (PseAAC)  mass no_efectores archaea dataset 4,
con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.025236  0.013765  0.057355  0.048178  0.045884  0.048178  0.018353
1    0.047453  0.011863  0.034271  0.030317  0.023726  0.040862  0.013181
2    0.051978  0.005569  0.044553  0.033415  0.018564  0.035271  0.005569
3    0.055673  0.000000  0.055673  0.129903  0.027836  0.083509  0.009279
4    0.052731  0.004794  0.036752  0.033556  0.014381  0.043143  0.011185
..        …         …         …         …         …         …         …
495  0.041996  0.006774  0.057575  0.069090  0.020998  0.049447  0.014902
496  0.032855  0.002489  0.012445  0.006472  0.005476  0.027877  0.001991
497  0.056248  0.000000  0.037499  0.112496  0.009375  0.037499  0.018749
498  0.047621  0.000000  0.142864  0.071432  0.000000  0.047621  0.023811
499  0.104898  0.000000  0.095362  0.152579  0.038145  0.028609  0.047681

           X7        X8        X9  …       X32       X33       X34  \
0    0.078002  0.071120  0.096356  …  0.012580 -0.012922  0.017558
```

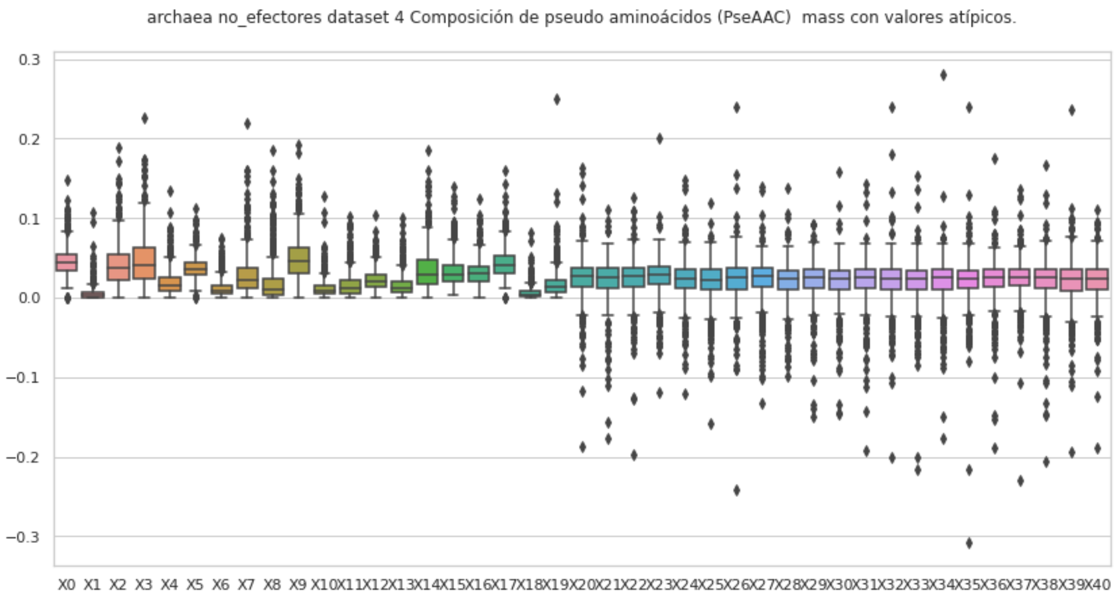
```
1    0.019772  0.007909  0.050089  …  0.019754  0.023904  0.029609
2    0.025989  0.007425  0.048265  …  0.020377  0.034918  0.034402
3    0.064951  0.092788  0.055673  …  0.062327 -0.017043  0.008362
4    0.004794  0.001598  0.027164  …  0.015919  0.011825  0.012354
..        …         …         …  …        …         …         …
495  0.041996  0.018289  0.048770  …  0.017338  0.012909  0.017057
496  0.006472  0.000498  0.026882  …  0.042410  0.047049  0.044951
497  0.084372  0.056248  0.121870  …  0.048221  0.005522 -0.017672
498  0.059527  0.035716  0.130959  …  0.006407 -0.006345  0.004161
499  0.019072  0.000000  0.095362  … -0.035298  0.045041  0.012696

          X35       X36       X37       X38       X39       X40           X41
0   -0.011112  0.005002  0.003936  0.028125 -0.003424 -0.001685  no_efectores
1    0.038108  0.013595  0.019989  0.041516  0.026864  0.014320  no_efectores
2    0.056726  0.024018  0.028560  0.032841  0.032156  0.020035  no_efectores
3   -0.020931 -0.012694 -0.023465 -0.021727 -0.045294  0.036962  no_efectores
4    0.036586  0.032773  0.057040  0.032673 -0.000144  0.010021  no_efectores
..        …         …         …        …         …         …            …
495  0.012927  0.017955  0.017286  0.021606  0.021885  0.009757  no_efectores
496  0.033611  0.032879  0.034045  0.035975  0.033264  0.036651  no_efectores
497 -0.015858 -0.006469 -0.026732  0.036349  0.112423  0.013893  no_efectores
498  0.088690  0.051576  0.015221 -0.025303 -0.050299  0.003646  no_efectores
499  0.093878  0.049566 -0.030028  0.049435  0.015968 -0.051152  no_efectores

[500 rows x 42 columns]
```

```
Composición de pseudo aminoácidos (PseAAC)  mass no_efectores archaea dataset 4,
con valores atípicos.
Estadísticas.
```

```
               X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.046211    0.005739    0.041211    0.046819    0.020029    0.038286
std      0.018517    0.010114    0.026857    0.032343    0.017187    0.014016
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.034774    0.000000    0.022351    0.023909    0.008975    0.029620
50%      0.043394    0.002872    0.036864    0.041035    0.015260    0.036136
75%      0.054368    0.006776    0.053697    0.062629    0.026194    0.044600
max      0.148466    0.107905    0.188395    0.225657    0.134970    0.112295

               X6          X7          X8          X9  …         X31  \
count  500.000000  500.000000  500.000000  500.000000  …  500.000000
mean     0.011953    0.028749    0.020419    0.050809  …    0.021079
std      0.011318    0.026761    0.027671    0.028555  …    0.028118
min      0.000000    0.000000    0.000000    0.000000  …   -0.191843
25%      0.004496    0.011249    0.003529    0.030805  …    0.012369
50%      0.009240    0.021915    0.009775    0.045470  …    0.025808
```

```
75%       0.015914    0.036831    0.023098    0.062370  …     0.036118
max       0.074863    0.219326    0.184980    0.191634  …     0.142033

                X32         X33         X34         X35         X36         X37  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.020972    0.019874    0.021229    0.020666    0.021397    0.023288
std      0.029424    0.028436    0.030084    0.031314    0.027700    0.025714
min     -0.200717   -0.215589   -0.176908   -0.307744   -0.188735   -0.229695
25%      0.010877    0.010403    0.010689    0.011985    0.013754    0.014463
50%      0.024487    0.024278    0.025515    0.024621    0.024666    0.025521
75%      0.034939    0.034169    0.035689    0.034789    0.035221    0.035416
max      0.239701    0.152643    0.280381    0.239798    0.174783    0.136065

                X38         X39         X40
count  500.000000  500.000000  500.000000
mean     0.019897    0.020585    0.021130
std      0.030130    0.029832    0.026074
min     -0.206400   -0.194146   -0.188739
25%      0.011417    0.008531    0.011010
50%      0.025132    0.024177    0.023851
75%      0.035359    0.035682    0.035414
max      0.167209    0.236291    0.110648

[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 4,
sin valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.030765  0.000750  0.007504  0.006003  0.010505  0.031515  0.000750
1    0.045758  0.005968  0.037800  0.047747  0.067642  0.049737  0.015916
2    0.061291  0.000000  0.026815  0.065122  0.015323  0.034476  0.003831
4    0.040318  0.004032  0.050398  0.044350  0.068541  0.046366  0.010080
5    0.036985  0.000000  0.026418  0.019373  0.017612  0.044030  0.010567
..        …         …         …         …         …         …         …
494  0.054458  0.001116  0.029461  0.022096  0.004241  0.031023  0.004910
495  0.011428  0.000000  0.034283  0.022855  0.017141  0.057138  0.005714
497  0.061577  0.002566  0.025657  0.007697  0.041051  0.041051  0.007697
498  0.039401  0.000730  0.010945  0.008026  0.016782  0.034294  0.002189
499  0.045292  0.000985  0.016738  0.025600  0.021661  0.040369  0.000000

           X7        X8        X9  …       X32       X33       X34  \
0    0.017258  0.001501  0.036017  …  0.011503  0.027663  0.039157
1    0.113400  0.077590  0.081568  … -0.029061 -0.007623  0.029471
2    0.030646  0.038307  0.049799  …  0.003156  0.014485  0.037015
4    0.124987  0.074589  0.080637  … -0.027392  0.037251  0.029525
5    0.012328  0.000000  0.047552  …  0.026388  0.022486  0.041689
..        …         …         …  …   …        …         …
494  0.003125  0.001339  0.025890  …  0.024783  0.032502  0.030626
495  0.045710  0.039996  0.034283  …  0.061268 -0.024893  0.024920
497  0.005131  0.007697  0.071839  …  0.011226  0.005262  0.046463
498  0.021890  0.003648  0.037942  …  0.028004  0.034414  0.041931
499  0.024615  0.017723  0.056122  …  0.008645  0.031114  0.022794
```

	X35	X36	X37	X38	X39	X40	X41
0	0.043664	0.052347	0.037519	0.043723	0.035143	0.040855	efectores
1	-0.024477	0.002469	-0.011768	0.018259	0.005073	0.001819	efectores
2	0.041446	0.023355	0.046412	0.021352	0.044749	0.052672	efectores
4	-0.000687	-0.002120	-0.006956	0.019081	0.002127	-0.015566	efectores
5	0.025261	0.013184	0.033865	0.004821	0.019263	0.026830	efectores
..	…	…	…	…	…	…	…
494	0.040051	0.031013	0.028552	0.036669	0.034594	0.037421	efectores
495	-0.008265	0.053850	0.029233	0.030001	-0.001669	-0.011222	efectores
497	0.004415	0.033338	0.009254	0.047364	0.043671	0.033998	efectores
498	0.018878	0.038444	0.035524	0.049345	0.019194	0.042141	efectores
499	0.033967	0.043901	0.025492	0.006882	0.015152	0.020590	efectores

[409 rows x 42 columns]

Composición de pseudo aminoácidos (PseAAC) mass efectores archaea dataset 4, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5 \
count	409.000000	409.000000	409.000000	409.000000	409.000000	409.000000
mean	0.044919	0.003707	0.032525	0.038364	0.020114	0.037321
std	0.012928	0.005097	0.019446	0.029436	0.013422	0.010116
min	0.011428	0.000000	0.002659	0.000000	0.000000	0.010495
25%	0.035320	0.000000	0.017132	0.013552	0.010931	0.029774
50%	0.043749	0.001597	0.029821	0.029411	0.016661	0.036176
75%	0.052748	0.005549	0.043749	0.056195	0.026237	0.043721
max	0.092556	0.026752	0.104632	0.148229	0.068541	0.075810

	X6	X7	X8	X9	…	X31 \
count	409.000000	409.000000	409.000000	409.000000	…	409.000000
mean	0.009658	0.031703	0.026478	0.056670	…	0.020568
std	0.007418	0.028807	0.033250	0.025945	…	0.019792
min	0.000000	0.000000	0.000000	0.005998	…	-0.058672
25%	0.003797	0.009953	0.002953	0.038815	…	0.010453
50%	0.007800	0.021793	0.009729	0.051809	…	0.024277
75%	0.014487	0.044239	0.043005	0.075732	…	0.033667
max	0.037124	0.137382	0.149871	0.143598	…	0.077201

	X32	X33	X34	X35	X36	X37 \
count	409.000000	409.000000	409.000000	409.000000	409.000000	409.000000
mean	0.025114	0.020464	0.021450	0.021167	0.024225	0.022376
std	0.019344	0.019887	0.019674	0.019956	0.019653	0.022087
min	-0.046684	-0.047962	-0.053480	-0.049643	-0.051369	-0.046286
25%	0.013836	0.007962	0.008822	0.009451	0.012236	0.010294
50%	0.027303	0.023641	0.024885	0.023873	0.027955	0.025489
75%	0.039537	0.034663	0.035370	0.034537	0.036157	0.035986

```
max      0.078227    0.062597    0.086233    0.085756    0.092840    0.080164

              X38         X39         X40
count  409.000000  409.000000  409.000000
mean     0.023110    0.023936    0.024068
std      0.018766    0.019068    0.019250
min     -0.029528   -0.041189   -0.036364
25%      0.012215    0.013268    0.012588
50%      0.025369    0.026374    0.026559
75%      0.036071    0.036712    0.036596
max      0.073990    0.077198    0.098788

[8 rows x 41 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass no_efectores archaea dataset 4, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.025236  0.013765  0.057355  0.048178  0.045884  0.048178  0.018353
1    0.047453  0.011863  0.034271  0.030317  0.023726  0.040862  0.013181
2    0.051978  0.005569  0.044553  0.033415  0.018564  0.035271  0.005569
4    0.052731  0.004794  0.036752  0.033556  0.014381  0.043143  0.011185
5    0.058640  0.002606  0.097733  0.108158  0.035184  0.043003  0.019547
..        …        …        …        …        …        …        …
492  0.036491  0.005614  0.046315  0.033684  0.018245  0.050526  0.011228
493  0.048182  0.000000  0.021414  0.026768  0.045505  0.053535  0.021414
494  0.026320  0.003760  0.031333  0.037600  0.008773  0.035093  0.003760
495  0.041996  0.006774  0.057575  0.069090  0.020998  0.049447  0.014902
496  0.032855  0.002489  0.012445  0.006472  0.005476  0.027877  0.001991

           X7        X8        X9  …       X32       X33       X34  \
0    0.078002  0.071120  0.096356  …  0.012580 -0.012922  0.017558
1    0.019772  0.007909  0.050089  …  0.019754  0.023904  0.029609
2    0.025989  0.007425  0.048265  …  0.020377  0.034918  0.034402
4    0.004794  0.001598  0.027164  …  0.015919  0.011825  0.012354
5    0.043003  0.046912  0.069065  …  0.002205  0.004924  0.018221
..        …        …        … …       …        …        …
492  0.035087  0.007017  0.033684  …  0.021289  0.024050  0.014293
493  0.032121  0.013384  0.112424  …  0.052824  0.049380  0.020804
494  0.011280  0.002507  0.010027  …  0.027695  0.031930  0.033107
495  0.041996  0.018289  0.048770  …  0.017338  0.012909  0.017057
496  0.006472  0.000498  0.026882  …  0.042410  0.047049  0.044951

          X35       X36       X37       X38       X39       X40           X41
0   -0.011112  0.005002  0.003936  0.028125 -0.003424 -0.001685  no_efectores
```

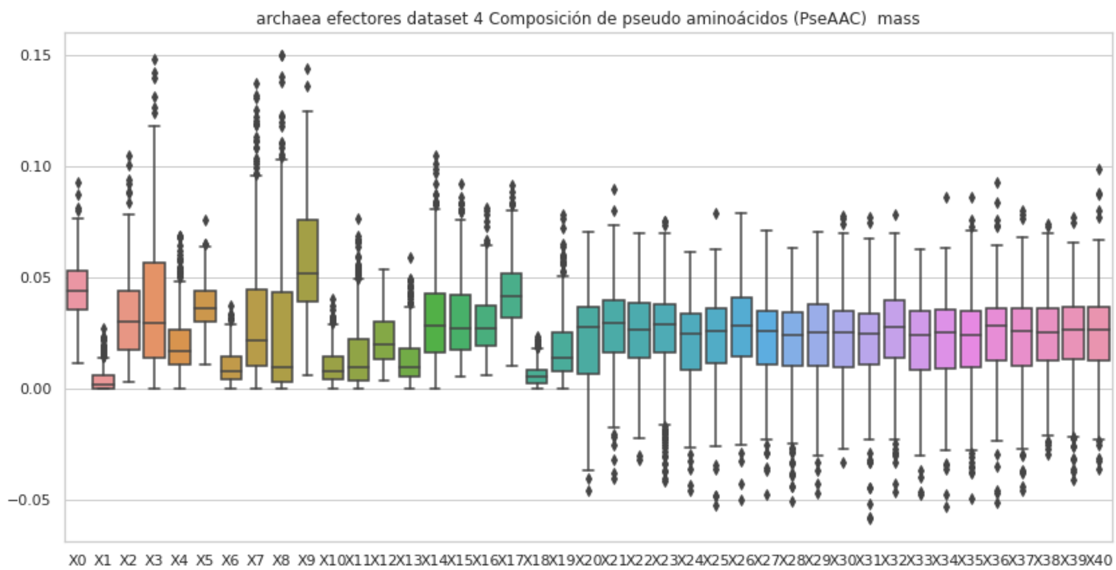
```
1    0.038108  0.013595  0.019989  0.041516  0.026864  0.014320  no_efectores
2    0.056726  0.024018  0.028560  0.032841  0.032156  0.020035  no_efectores
4    0.036586  0.032773  0.057040  0.032673 -0.000144  0.010021  no_efectores
5    0.014914  0.001369  0.002667  0.011093 -0.007897  0.001457  no_efectores
..        …         …         …         …         …         …            …
492  0.011885  0.012290  0.031049  0.017046  0.018635  0.023577  no_efectores
493 -0.022538 -0.002836  0.057734  0.025357 -0.030641 -0.042891  no_efectores
494  0.014722  0.022151  0.028549  0.041703  0.043946  0.035218  no_efectores
495  0.012927  0.017955  0.017286  0.021606  0.021885  0.009757  no_efectores
496  0.033611  0.032879  0.034045  0.035975  0.033264  0.036651  no_efectores

[413 rows x 42 columns]
```

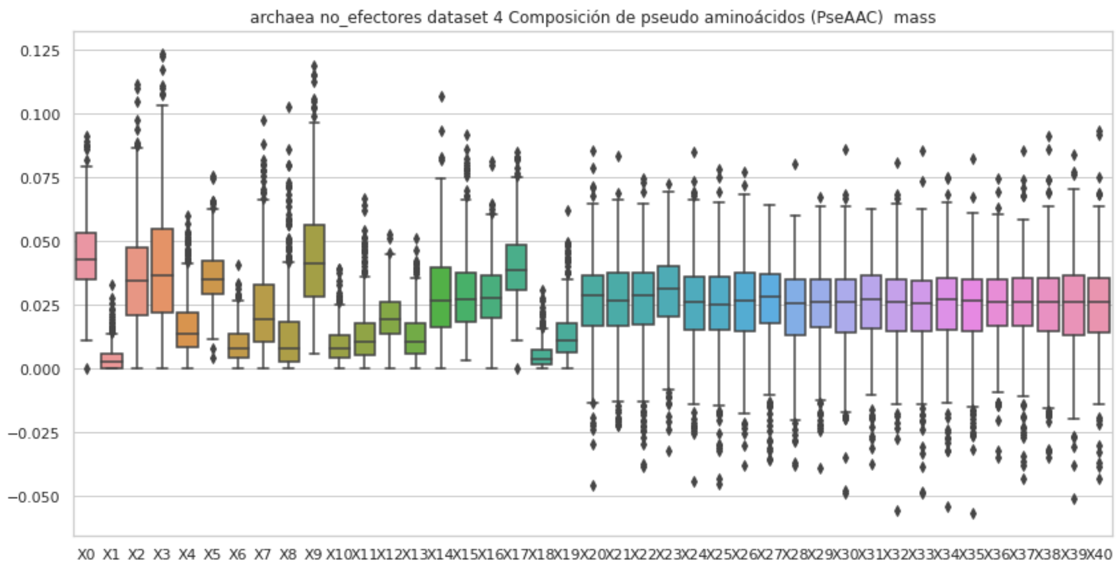
Composición de pseudo aminoácidos (PseAAC) mass no_efectores archaea dataset 4,
sin valores atípicos.
Estadísticas.

```
               X0          X1          X2          X3          X4          X5  \
count  413.000000  413.000000  413.000000  413.000000  413.000000  413.000000
mean     0.044564    0.004129    0.036560    0.040106    0.016732    0.036058
std      0.014383    0.004934    0.020430    0.024501    0.011644    0.010610
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.004081
25%      0.034815    0.000000    0.021064    0.021838    0.008715    0.029067
50%      0.042910    0.002714    0.034352    0.036591    0.013866    0.035093
75%      0.053311    0.005616    0.047644    0.054761    0.021977    0.042561
max      0.091160    0.033161    0.111492    0.123549    0.059881    0.075660

               X6          X7          X8          X9  …         X31  \
count  413.000000  413.000000  413.000000  413.000000  …  413.000000
mean     0.009645    0.023489    0.014102    0.044081  …    0.025055
std      0.007382    0.017745    0.016757    0.021178  …    0.016192
min      0.000000    0.000000    0.000000    0.005692  …   -0.037296
25%      0.004100    0.010596    0.002889    0.028454  …    0.015786
50%      0.007946    0.019426    0.007947    0.041097  …    0.027478
75%      0.013549    0.032918    0.018521    0.056197  …    0.036373
max      0.040550    0.097311    0.102718    0.118651  …    0.062696

              X32         X33         X34         X35         X36         X37  \
count  413.000000  413.000000  413.000000  413.000000  413.000000  413.000000
mean     0.024310    0.023576    0.024753    0.023655    0.025174    0.025174
std      0.016585    0.017089    0.017876    0.017329    0.015136    0.017039
min     -0.055606   -0.048648   -0.053830   -0.056646   -0.034997   -0.043068
25%      0.014976    0.014796    0.015012    0.014638    0.016564    0.016913
50%      0.026036    0.025578    0.027396    0.026540    0.026045    0.026154
75%      0.035234    0.034449    0.035674    0.034838    0.035140    0.035329
max      0.081103    0.085316    0.074920    0.082255    0.074439    0.085745
```

	X38	X39	X40
count	413.000000	413.000000	413.000000
mean	0.024346	0.025030	0.024752
std	0.017686	0.017918	0.017700
min	-0.034917	-0.050778	-0.042891
25%	0.014870	0.013319	0.014320
50%	0.026052	0.026146	0.026145
75%	0.035574	0.036645	0.035731
max	0.091495	0.084237	0.093142

[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")
```

```python
#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 4,
con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.032245  0.000786  0.007865  0.006292  0.011010  0.033031  0.000786
1    0.025349  0.003306  0.020941  0.026451  0.037473  0.027554  0.008817
2    0.090468  0.000000  0.039580  0.096122  0.022617  0.050888  0.005654
3    0.000000  0.019458  0.029187  0.077833  0.009729  0.048645  0.000000
4    0.020449  0.002045  0.025561  0.022494  0.034764  0.023517  0.005112
..        …         …         …         …         …         …         …
495  0.026946  0.000000  0.080837  0.053891  0.040419  0.134729  0.013473
496  0.037135  0.004642  0.046419  0.055703  0.060345  0.037135  0.013926
497  0.051413  0.002142  0.021422  0.006427  0.034275  0.034275  0.006427
498  0.031989  0.000592  0.008886  0.006516  0.013625  0.027842  0.001777
499  0.041078  0.000893  0.015181  0.023218  0.019646  0.036613  0.000000

           X7        X8        X9  …       X53       X54       X55  \
0    0.018089  0.001573  0.037750  …  0.001427  0.034071  0.009318
1    0.062822  0.042984  0.045188  …  0.001869  0.022418  0.012157
2    0.045234  0.056542  0.073505  … -0.003685 -0.002361 -0.032694
3    0.038916  0.009729  0.097291  …  0.045297 -0.027964  0.014022
4    0.063392  0.037831  0.040898  …  0.015611  0.019880  0.007899
..        …         …         …  …       …         …         …
495  0.107783  0.094310  0.080837  …  0.078048  0.071365  0.003346
496  0.106764  0.148542  0.078913  …  0.012031  0.033909  0.004183
497  0.004284  0.006427  0.059982  …  0.005768  0.020190 -0.008807
498  0.017771  0.002962  0.030804  …  0.001127  0.031139  0.005412
499  0.022325  0.016074  0.050901  …  0.019356 -0.007326 -0.010416

          X56       X57       X58       X59       X60       X61        X62
0    0.041847  0.014703  0.017385 -0.003353  0.029449  0.001743  efectores
1   -0.009040 -0.012749 -0.002489  0.005813  0.019921  0.006423  efectores
2    0.042145  0.065221 -0.000143  0.049343 -0.045448 -0.051921  efectores
3   -0.031988 -0.017450  0.082439  0.062794  0.049960  0.072756  efectores
4    0.019744  0.001915  0.004963  0.006194  0.020285  0.011580  efectores
..        …         …         …         …         …         …        …
495 -0.070540 -0.039225 -0.028852  0.066859  0.015984  0.066744  efectores
496  0.041232  0.036598  0.041603  0.048848 -0.041150 -0.009110  efectores
497  0.027038  0.008116  0.008464 -0.011060  0.037342  0.003235  efectores
```

```
498  0.025333  0.003246  0.024499 -0.001116  0.025908  0.002002  efectores
499  0.019816  0.010082  0.012270  0.017135 -0.002506 -0.002450  efectores

[500 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro efectores archaea dataset 4, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.047953	0.005237	0.038579	0.045099	0.021202	0.041265	
std	0.025577	0.008244	0.025094	0.030295	0.014357	0.021814	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.004582	
25%	0.029357	0.000000	0.015485	0.014486	0.011132	0.025747	
50%	0.043500	0.001993	0.035397	0.046196	0.017912	0.036303	
75%	0.061928	0.007794	0.056877	0.065857	0.027726	0.052041	
max	0.204202	0.075560	0.131534	0.145859	0.099138	0.134729	

	X6	X7	X8	X9	…	X52	\
count	500.000000	500.000000	500.000000	500.000000	…	500.000000	
mean	0.011344	0.035712	0.031043	0.060969	…	0.006860	
std	0.011046	0.032882	0.036237	0.031802	…	0.037329	
min	0.000000	0.000000	0.000000	0.014010	…	-0.304999	
25%	0.004374	0.012836	0.003711	0.040171	…	-0.006720	
50%	0.009399	0.022612	0.014260	0.054712	…	0.012636	
75%	0.015134	0.049619	0.053085	0.074344	…	0.025649	
max	0.146866	0.201493	0.179365	0.379233	…	0.237579	

	X53	X54	X55	X56	X57	X58	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.011139	0.013627	0.016652	0.005928	0.011230	0.010207	
std	0.026585	0.040809	0.031362	0.035424	0.031414	0.035752	
min	-0.197317	-0.304223	-0.143787	-0.240028	-0.152496	-0.114083	
25%	-0.000238	-0.001999	0.001812	-0.010106	-0.001968	-0.007640	
50%	0.007687	0.016481	0.011959	0.009753	0.006411	0.013551	
75%	0.023060	0.030693	0.029233	0.025061	0.024673	0.027326	
max	0.208379	0.487055	0.222833	0.169375	0.218122	0.359255	

	X59	X60	X61
count	500.000000	500.000000	500.000000
mean	0.015037	0.006817	0.013769
std	0.031732	0.038464	0.027185
min	-0.121535	-0.268224	-0.096049
25%	-0.000354	-0.005723	-0.000186
50%	0.009895	0.011746	0.010564
75%	0.030496	0.026548	0.027523

```
max      0.233705    0.160016    0.140331

[8 rows x 62 columns]
```

no_efectores
Composición de pseudo aminoácidos (PseAAC) hidro no_efectores archaea dataset 4, con valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.027032  0.014745  0.061436  0.051606  0.049149  0.051606  0.019659
1    0.077103  0.019276  0.055685  0.049260  0.038551  0.066394  0.021417
2    0.104188  0.011163  0.089304  0.066978  0.037210  0.070699  0.011163
3    0.022069  0.000000  0.022069  0.051495  0.011035  0.033104  0.003678
4    0.087513  0.007956  0.060994  0.055690  0.023867  0.071602  0.018563
..        …         …         …         …         …         …         …
495  0.033541  0.005410  0.045983  0.055180  0.016770  0.039492  0.011902
496  0.053448  0.004049  0.020246  0.010528  0.008908  0.045350  0.003239
497  0.039990  0.000000  0.026660  0.079979  0.006665  0.026660  0.013330
498  0.033162  0.000000  0.099486  0.049743  0.000000  0.033162  0.016581
499  0.063652  0.000000  0.057865  0.092584  0.023146  0.017360  0.028933

           X7        X8        X9  …       X53       X54       X55  \
0    0.083553  0.076180  0.103212  …  0.024466  0.002753  0.021283
1    0.032126  0.012850  0.081386  …  0.016545  0.027295  0.009730
2    0.052094  0.014884  0.096746  …  0.057701  0.043236  0.019775
3    0.025748  0.036782  0.022069  …  0.012868 -0.011489  0.009696
4    0.007956  0.002652  0.045083  …  0.030521 -0.018048  0.011363
..        …         …         …  …       …         …         …
495  0.033541  0.014606  0.038951  …  0.034519  0.004193  0.017682
496  0.010528  0.000810  0.043731  … -0.000293  0.015588 -0.000555
497  0.059984  0.039990  0.086644  …  0.055584 -0.018040 -0.011006
498  0.041453  0.024872  0.091196  …  0.017345 -0.000787  0.006423
499  0.011573  0.000000  0.057865  …  0.030766  0.022694  0.080753

          X56       X57       X58       X59       X60       X61          X62
0   -0.017834 -0.043533 -0.016048 -0.000916  0.001027  0.043048  no_efectores
1   -0.013957  0.002013  0.003557  0.014578  0.002888  0.002677  no_efectores
2   -0.008754 -0.009093 -0.092127 -0.033052 -0.016606  0.024242  no_efectores
3    0.026237  0.027790 -0.022797 -0.024782  0.002921  0.002037  no_efectores
4    0.013668  0.027022 -0.025926  0.023044 -0.023549  0.013406  no_efectores
..        …         …         …         …         …         …           …
495  0.005279  0.016131  0.005433  0.023853  0.011721  0.028226  no_efectores
496  0.004898  0.007174  0.012207 -0.001555  0.011900  0.003357  no_efectores
497 -0.009448 -0.039599  0.046072  0.051471 -0.016491 -0.044242  no_efectores
498  0.042272  0.107146 -0.067379 -0.039846  0.029458  0.080801  no_efectores
```

```
499  0.008932 -0.029724 -0.047989  0.010410 -0.039993  0.020108  no_efectores

[500 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro no_efectores archaea dataset 4, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	0.061114	0.007161	0.049932	0.054876	0.023755	0.050454
std	0.043775	0.012662	0.033318	0.039767	0.027159	0.035178
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.035155	0.000000	0.030479	0.036704	0.011968	0.031052
50%	0.055559	0.003678	0.049790	0.053701	0.019008	0.044386
75%	0.077323	0.008527	0.066520	0.070475	0.028804	0.060902
max	0.506867	0.142713	0.351477	0.546501	0.341563	0.410056

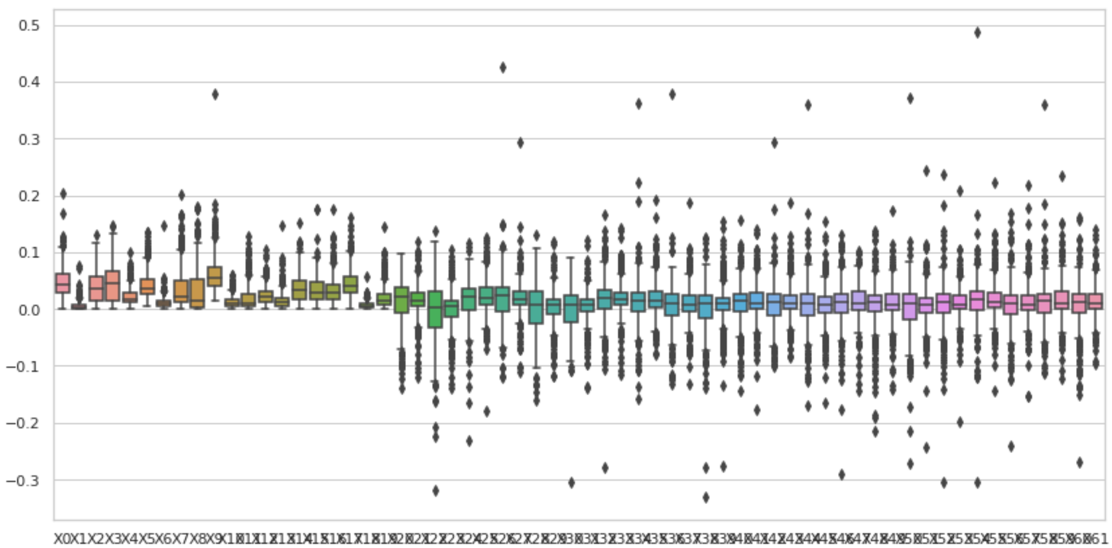
	X6	X7	X8	X9	…	X52
count	500.000000	500.000000	500.000000	500.000000	…	500.000000
mean	0.015291	0.034027	0.023101	0.060850	…	-0.000509
std	0.021205	0.036856	0.032424	0.043824	…	0.068450
min	0.000000	0.000000	0.000000	0.000000	…	-0.792957
25%	0.004900	0.014084	0.004153	0.038305	…	-0.009696
50%	0.011005	0.025157	0.011713	0.052307	…	0.006824
75%	0.019639	0.040520	0.026442	0.072482	…	0.022466
max	0.292897	0.461703	0.319640	0.614814	…	0.379448

	X53	X54	X55	X56	X57	X58
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	0.010316	0.003323	0.011996	0.002179	0.011362	0.000960
std	0.047407	0.052788	0.043394	0.062978	0.052111	0.054732
min	-0.594255	-0.612472	-0.307088	-0.474801	-0.459420	-0.724280
25%	-0.001532	-0.009895	-0.000642	-0.011382	-0.003760	-0.012272
50%	0.012157	0.008327	0.013864	0.006109	0.009869	0.006148
75%	0.028064	0.025437	0.029981	0.020991	0.026142	0.021182
max	0.291743	0.194348	0.132075	0.615420	0.444346	0.220783

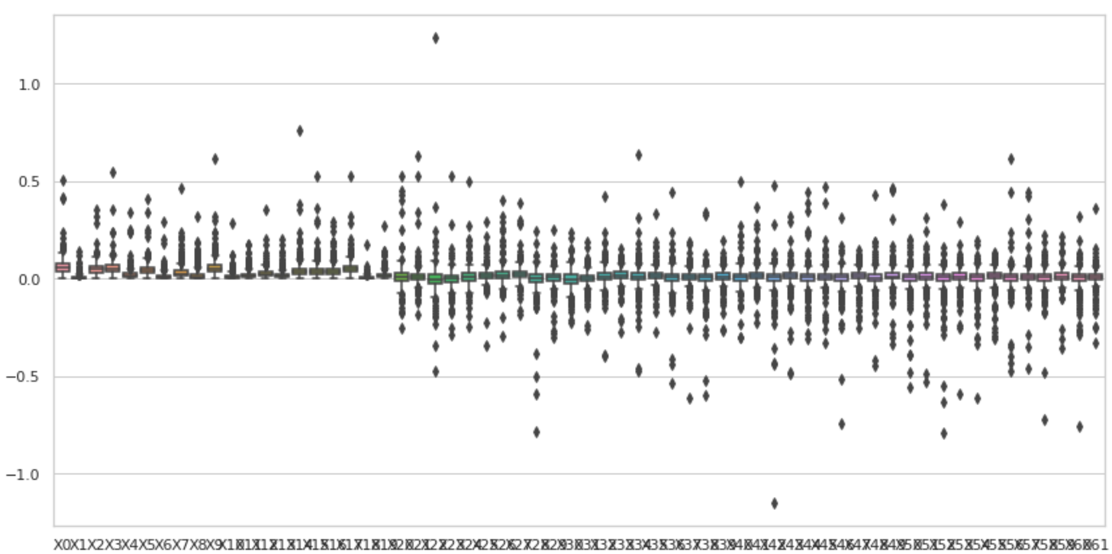
	X59	X60	X61
count	500.000000	500.000000	500.000000
mean	0.010597	0.000232	0.010297
std	0.039234	0.057217	0.042521
min	-0.356158	-0.758084	-0.329487
25%	-0.004120	-0.010824	-0.001519
50%	0.010079	0.005949	0.010307
75%	0.028663	0.020627	0.025564
max	0.218412	0.322468	0.363180

```
[8 rows x 62 columns]
```

archaea efectores dataset 4 Composición de pseudo aminoácidos (PseAAC) hidro con valores atípicos.



archaea no_efectores dataset 4 Composición de pseudo aminoácidos (PseAAC) hidro con valores atípicos.



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)
```

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

```
efectores
Composición de pseudo aminoácidos (PseAAC)  efectores archaea dataset 4, sin
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.032245  0.000786  0.007865  0.006292  0.011010  0.033031  0.000786
1    0.025349  0.003306  0.020941  0.026451  0.037473  0.027554  0.008817
2    0.090468  0.000000  0.039580  0.096122  0.022617  0.050888  0.005654
4    0.020449  0.002045  0.025561  0.022494  0.034764  0.023517  0.005112
5    0.045588  0.000000  0.032563  0.023880  0.021709  0.054272  0.013025
..        …         …         …         …         …         …         …
493  0.049755  0.000000  0.015441  0.005147  0.018872  0.034313  0.010294
494  0.098728  0.002023  0.053410  0.040058  0.007688  0.056243  0.008902
497  0.051413  0.002142  0.021422  0.006427  0.034275  0.034275  0.006427
498  0.031989  0.000592  0.008886  0.006516  0.013625  0.027842  0.001777
499  0.041078  0.000893  0.015181  0.023218  0.019646  0.036613  0.000000

           X7        X8        X9  …       X53       X54       X55  \
0    0.018089  0.001573  0.037750  …  0.001427  0.034071  0.009318
1    0.062822  0.042984  0.045188  …  0.001869  0.022418  0.012157
2    0.045234  0.056542  0.073505  … -0.003685 -0.002361 -0.032694
4    0.063392  0.037831  0.040898  …  0.015611  0.019880  0.007899
5    0.015196  0.000000  0.058613  …  0.004944 -0.014218 -0.042345
..        …         …         …  …       …         …         …
493  0.017157  0.006863  0.036029  … -0.005329  0.019325 -0.001876
494  0.005665  0.002428  0.046936  …  0.033383 -0.015116  0.013931
497  0.004284  0.006427  0.059982  …  0.005768  0.020190 -0.008807
498  0.017771  0.002962  0.030804  …  0.001127  0.031139  0.005412
499  0.022325  0.016074  0.050901  …  0.019356 -0.007326 -0.010416

          X56       X57       X58       X59       X60       X61        X62
0    0.041847  0.014703  0.017385 -0.003353  0.029449  0.001743  efectores
1   -0.009040 -0.012749 -0.002489  0.005813  0.019921  0.006423  efectores
2    0.042145  0.065221 -0.000143  0.049343 -0.045448 -0.051921  efectores
4    0.019744  0.001915  0.004963  0.006194  0.020285  0.011580  efectores
5   -0.001878  0.003217  0.014747 -0.005551  0.007788 -0.000822  efectores
..        …         …         …         …         …         …          …
493  0.023798 -0.008968  0.030615  0.015110  0.027704  0.000764  efectores
494  0.018852  0.032054  0.008290  0.024154 -0.003948  0.021671  efectores
497  0.027038  0.008116  0.008464 -0.011060  0.037342  0.003235  efectores
498  0.025333  0.003246  0.024499 -0.001116  0.025908  0.002002  efectores
499  0.019816  0.010082  0.012270  0.017135 -0.002506 -0.002450  efectores

[393 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC) efectores archaea dataset 4, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5
count	393.000000	393.000000	393.000000	393.000000	393.000000	393.000000
mean	0.045671	0.003173	0.032718	0.037640	0.018772	0.037110
std	0.022150	0.004752	0.022119	0.026969	0.010817	0.018395
min	0.002254	0.000000	0.000000	0.000000	0.000000	0.004582
25%	0.028745	0.000000	0.012972	0.010352	0.010584	0.024588
50%	0.042371	0.000931	0.026655	0.035950	0.016318	0.032250
75%	0.058622	0.004291	0.051447	0.061110	0.024779	0.046277
max	0.121241	0.025315	0.097498	0.121241	0.058928	0.106314

	X6	X7	X8	X9	…	X52
count	393.000000	393.000000	393.000000	393.000000	…	393.000000
mean	0.009381	0.026746	0.020922	0.053053	…	0.009803
std	0.007364	0.023051	0.026349	0.022182	…	0.023314
min	0.000000	0.000000	0.000000	0.014010	…	-0.070294
25%	0.003678	0.010810	0.002862	0.037401	…	-0.002122
50%	0.008610	0.018606	0.009016	0.049606	…	0.014661
75%	0.013135	0.035212	0.029662	0.063467	…	0.025145
max	0.036170	0.128104	0.135618	0.149386	…	0.078429

	X53	X54	X55	X56	X57	X58
count	393.000000	393.000000	393.000000	393.000000	393.000000	393.000000
mean	0.010828	0.015435	0.014621	0.009351	0.011409	0.011338
std	0.017640	0.022369	0.020828	0.023418	0.020322	0.021957
min	-0.047680	-0.055519	-0.051938	-0.090162	-0.050707	-0.063635
25%	0.000264	0.002345	0.002367	-0.006161	-0.000817	-0.001142
50%	0.007401	0.017112	0.011154	0.013702	0.006572	0.015348
75%	0.020459	0.029359	0.026391	0.025333	0.021976	0.026115
max	0.085794	0.104970	0.086806	0.097906	0.089250	0.082295

	X59	X60	X61
count	393.000000	393.000000	393.000000
mean	0.012829	0.010211	0.013479
std	0.022127	0.024941	0.019953
min	-0.062735	-0.105065	-0.051921
25%	-0.000085	-0.003004	0.000656
50%	0.009378	0.014733	0.010614
75%	0.027525	0.026433	0.023846
max	0.101141	0.071224	0.073345

[8 rows x 62 columns]

no_efectores
Composición de pseudo aminoácidos (PseAAC) no_efectores archaea dataset 4, sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.027032	0.014745	0.061436	0.051606	0.049149	0.051606	0.019659	
1	0.077103	0.019276	0.055685	0.049260	0.038551	0.066394	0.021417	
2	0.104188	0.011163	0.089304	0.066978	0.037210	0.070699	0.011163	
3	0.022069	0.000000	0.022069	0.051495	0.011035	0.033104	0.003678	
4	0.087513	0.007956	0.060994	0.055690	0.023867	0.071602	0.018563	
..	…	…	…	…	…	…	…	
495	0.033541	0.005410	0.045983	0.055180	0.016770	0.039492	0.011902	
496	0.053448	0.004049	0.020246	0.010528	0.008908	0.045350	0.003239	
497	0.039990	0.000000	0.026660	0.079979	0.006665	0.026660	0.013330	
498	0.033162	0.000000	0.099486	0.049743	0.000000	0.033162	0.016581	
499	0.063652	0.000000	0.057865	0.092584	0.023146	0.017360	0.028933	

	X7	X8	X9	…	X53	X54	X55	\
0	0.083553	0.076180	0.103212	…	0.024466	0.002753	0.021283	
1	0.032126	0.012850	0.081386	…	0.016545	0.027295	0.009730	
2	0.052094	0.014884	0.096746	…	0.057701	0.043236	0.019775	
3	0.025748	0.036782	0.022069	…	0.012868	-0.011489	0.009696	
4	0.007956	0.002652	0.045083	…	0.030521	-0.018048	0.011363	
..	…	…	…	…	…	…	…	
495	0.033541	0.014606	0.038951	…	0.034519	0.004193	0.017682	
496	0.010528	0.000810	0.043731	…	-0.000293	0.015588	-0.000555	
497	0.059984	0.039990	0.086644	…	0.055584	-0.018040	-0.011006	
498	0.041453	0.024872	0.091196	…	0.017345	-0.000787	0.006423	
499	0.011573	0.000000	0.057865	…	0.030766	0.022694	0.080753	

	X56	X57	X58	X59	X60	X61	X62
0	-0.017834	-0.043533	-0.016048	-0.000916	0.001027	0.043048	no_efectores
1	-0.013957	0.002013	0.003557	0.014578	0.002888	0.002677	no_efectores
2	-0.008754	-0.009093	-0.092127	-0.033052	-0.016606	0.024242	no_efectores
3	0.026237	0.027790	-0.022797	-0.024782	0.002921	0.002037	no_efectores
4	0.013668	0.027022	-0.025926	0.023044	-0.023549	0.013406	no_efectores
..	…	…	…	…	…	…	…
495	0.005279	0.016131	0.005433	0.023853	0.011721	0.028226	no_efectores
496	0.004898	0.007174	0.012207	-0.001555	0.011900	0.003357	no_efectores
497	-0.009448	-0.039599	0.046072	0.051471	-0.016491	-0.044242	no_efectores
498	0.042272	0.107146	-0.067379	-0.039846	0.029458	0.080801	no_efectores
499	0.008932	-0.029724	-0.047989	0.010410	-0.039993	0.020108	no_efectores

[450 rows x 63 columns]

Composición de pseudo aminoácidos (PseAAC) no_efectores archaea dataset 4, sin valores atípicos.
Estadísticas.

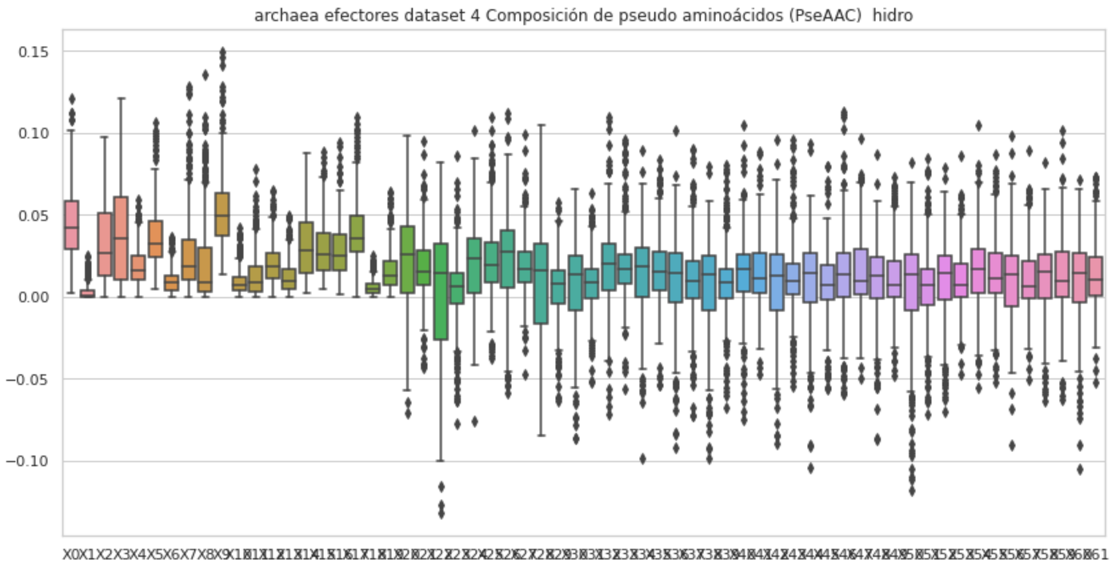
	X0	X1	X2	X3	X4	X5	\
count	450.000000	450.000000	450.000000	450.000000	450.000000	450.000000	
mean	0.057166	0.005413	0.047016	0.050757	0.020109	0.045933	
std	0.028793	0.006592	0.024173	0.026433	0.012599	0.022356	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.034528	0.000000	0.029950	0.031319	0.011788	0.030213	
50%	0.054374	0.003410	0.049242	0.051964	0.018241	0.043431	
75%	0.075441	0.007550	0.064696	0.068105	0.025839	0.058941	
max	0.176347	0.034990	0.104889	0.124039	0.085486	0.122801	

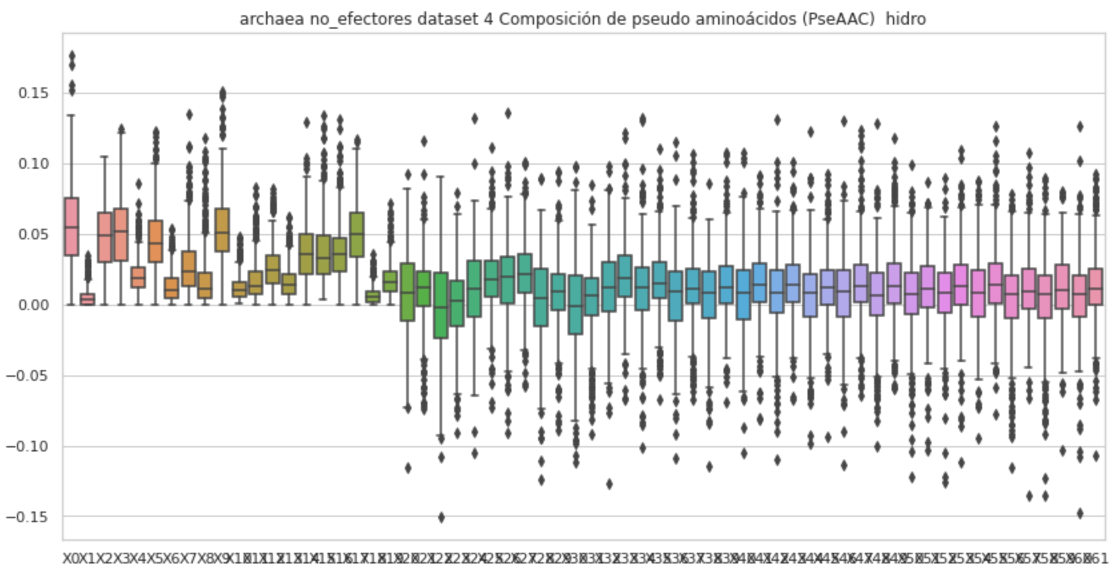
	X6	X7	X8	X9	…	X52	\
count	450.000000	450.000000	450.000000	450.000000	…	450.000000	
mean	0.012696	0.028482	0.018137	0.053957	…	0.006809	
std	0.010203	0.021056	0.021031	0.025283	…	0.025766	
min	0.000000	0.000000	0.000000	0.000000	…	-0.125824	
25%	0.004681	0.013030	0.004000	0.037374	…	-0.005785	
50%	0.010310	0.023748	0.010810	0.050847	…	0.007992	
75%	0.018512	0.037277	0.022769	0.067555	…	0.022654	
max	0.053129	0.134406	0.118222	0.150617	…	0.089560	

	X53	X54	X55	X56	X57	X58	\
count	450.000000	450.000000	450.000000	450.000000	450.000000	450.000000	
mean	0.014226	0.007893	0.016005	0.003438	0.011068	0.004359	
std	0.024928	0.026160	0.026179	0.028055	0.025268	0.028307	
min	-0.111496	-0.094467	-0.077573	-0.115489	-0.135004	-0.135628	
25%	0.000044	-0.008402	0.000323	-0.009941	-0.003206	-0.009916	
50%	0.012810	0.008690	0.013864	0.006817	0.009496	0.007416	
75%	0.027621	0.023934	0.028632	0.020337	0.024979	0.020570	
max	0.109722	0.087255	0.126591	0.078127	0.107146	0.089467	

	X59	X60	X61
count	450.000000	450.000000	450.000000
mean	0.011362	0.005225	0.013013
std	0.024545	0.026763	0.024528
min	-0.103157	-0.147609	-0.106998
25%	-0.003021	-0.008629	-0.000208
50%	0.010312	0.006970	0.011483
75%	0.027986	0.020612	0.025493
max	0.080470	0.126656	0.092510

[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 4,
con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.093449  0.103198  0.024476  0.004098 -0.034269  0.011867 -0.004485
1   -0.026317 -0.029884  0.133822  0.019048 -0.013586 -0.002166  0.008559
2   -0.125665  0.091094  0.044624  0.148923  0.045763  0.090525 -0.071119
3    0.023092  0.056525  0.034590 -0.110905 -0.096042 -0.016991 -0.153657
4    0.046318 -0.017352  0.048512 -0.002018  0.009757 -0.032974 -0.049354
..        …         …         …         …         …         …         …
495  0.077209  0.130857 -0.001088  0.020975  0.128191  0.060350 -0.030527
496  0.058449  0.047566  0.053842  0.021811  0.019159 -0.053528 -0.026796
497  0.038507  0.050909  0.064274  0.069369  0.035609 -0.066948  0.044645
498 -0.010068  0.033854  0.006513  0.027894 -0.012882  0.013830  0.052302
499  0.033132  0.020666  0.110568 -0.025619  0.012764  0.044836  0.070477

           X7        X8        X9       X10       X11       X12        X13
0   -0.113482 -0.001697  0.065546 -0.054329 -0.043361 -0.174597  efectores
```

```
1   -0.063992  0.006311 -0.011478 -0.019367 -0.006123 -0.078499  efectores
2   -0.124123  0.010813  0.087544 -0.069030  0.086131 -0.055990  efectores
3    0.105717 -0.116607 -0.071618 -0.059377  0.083385 -0.129894  efectores
4   -0.042160 -0.006835 -0.005385  0.053977 -0.036500 -0.066198  efectores
..        …        …        …        …        …        …         …
495 -0.162259  0.035005 -0.039880  0.019705  0.064380  0.163298  efectores
496 -0.070468 -0.052426 -0.092558  0.003850 -0.030561  0.036381  efectores
497 -0.129053 -0.099917  0.055308 -0.081601 -0.122854 -0.052929  efectores
498  0.006947 -0.090983 -0.098329  0.032305 -0.028716 -0.033192  efectores
499  0.006391  0.006202 -0.040613 -0.045563 -0.011576 -0.085682  efectores

[500 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro_mass efectores archaea dataset 4, con valores atípicos.
Estadísticas.

```
               X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.005620    0.017056    0.011178    0.014521   -0.009450    0.000990
std      0.071455    0.068047    0.067464    0.067014    0.069897    0.066528
min     -0.411431   -0.218952   -0.258376   -0.209807   -0.258579   -0.204442
25%     -0.039204   -0.020888   -0.031667   -0.024978   -0.046481   -0.038551
50%      0.006825    0.017552    0.006734    0.016236   -0.002745    0.003903
75%      0.051702    0.058392    0.051930    0.057831    0.033261    0.041053
max      0.222100    0.241034    0.271296    0.250239    0.171421    0.201266

               X6          X7          X8          X9         X10         X11  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.021188   -0.000286   -0.004550    0.007669    0.002708   -0.004569
std      0.069940    0.066317    0.079660    0.077076    0.069242    0.067342
min     -0.194555   -0.198585   -0.451489   -0.213733   -0.261583   -0.222320
25%     -0.026817   -0.037502   -0.050406   -0.040899   -0.041721   -0.044494
50%      0.018307    0.003269   -0.001683    0.006546    0.000539   -0.002359
75%      0.066886    0.043293    0.041267    0.056753    0.045885    0.035030
max      0.207433    0.259834    0.291169    0.400687    0.193456    0.236683

              X12
count  500.000000
mean     0.018125
std      0.070402
min     -0.223231
25%     -0.029602
50%      0.015106
75%      0.070658
max      0.215409
```

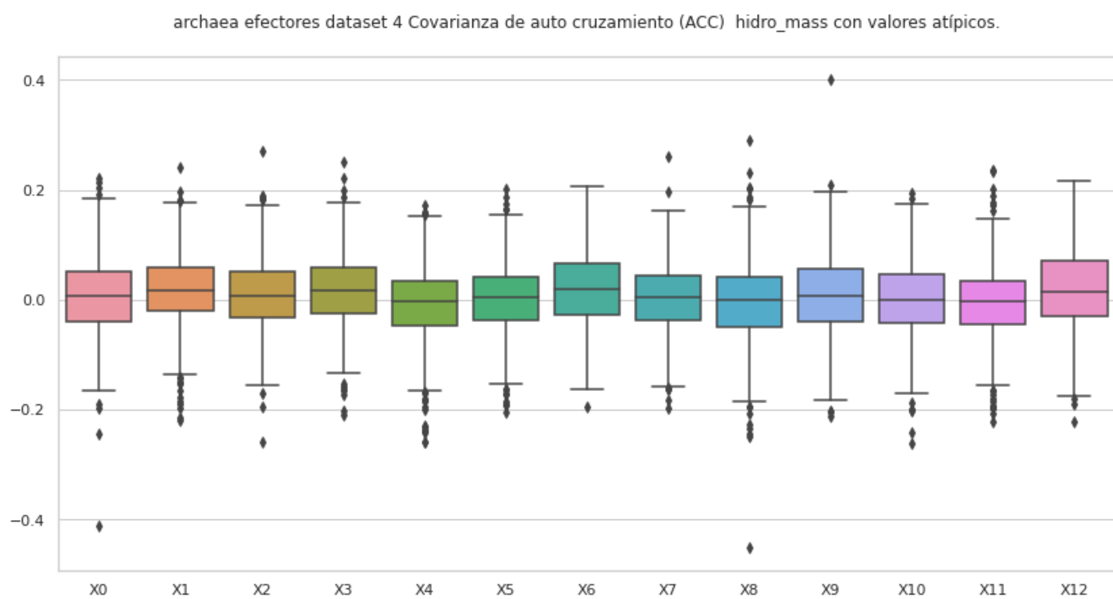
no_efectores
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores archaea dataset 4, con valores atípicos.
Valores del documento csv.

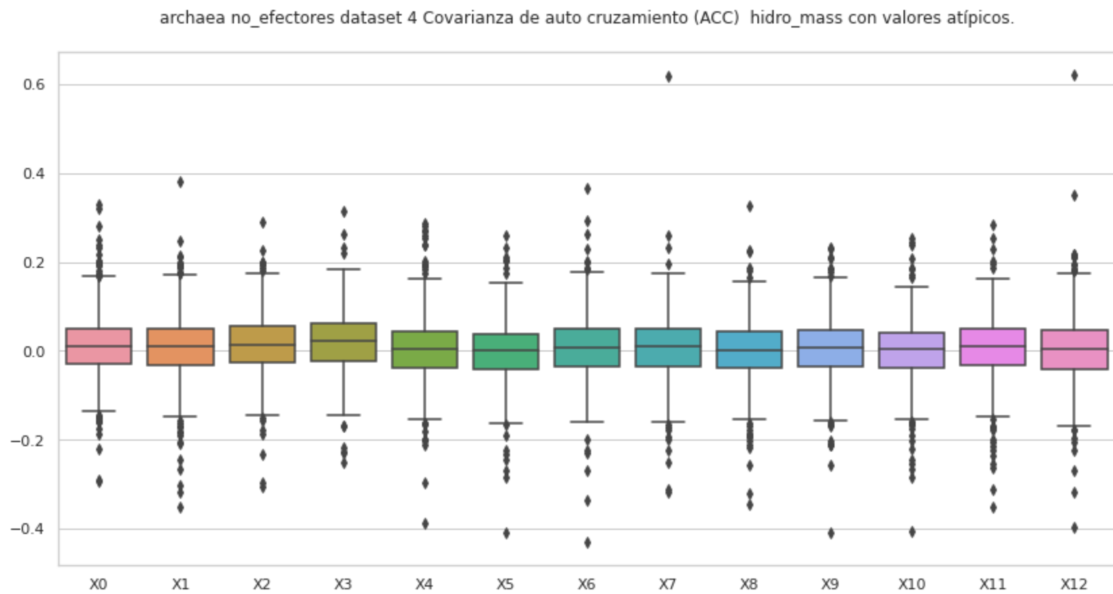
```
           X0        X1        X2        X3        X4        X5        X6  \
0   -0.010315 -0.061416 -0.039641  0.021242  0.054646 -0.006399 -0.036350
1   -0.059650  0.010867 -0.034683  0.041115 -0.058475 -0.014403  0.010750
2    0.130371  0.044571  0.018364  0.013825 -0.098359 -0.021772 -0.117048
3   -0.113533  0.010284  0.059626  0.039931 -0.022758 -0.022429  0.004175
4   -0.030411  0.147539 -0.131060  0.085225 -0.105771  0.120964 -0.092066
..        …         …         …         …         …         …         …
495 -0.004383 -0.053191  0.040575  0.032951 -0.021183 -0.005124 -0.008713
496 -0.024239  0.027761  0.077498 -0.020696 -0.083397 -0.040206  0.008057
497 -0.007380  0.006793 -0.129347 -0.097535  0.011029 -0.071428 -0.043595
498  0.038281 -0.029552 -0.076342 -0.030407 -0.081197  0.022590  0.093086
499  0.043463  0.009990 -0.002767 -0.031961 -0.054017 -0.061926  0.011879

           X7        X8        X9       X10       X11       X12           X13
0    0.054648  0.034022 -0.005984 -0.011492 -0.015608  0.011187  no_efectores
1   -0.032587 -0.062418  0.031327  0.002862 -0.078312 -0.013321  no_efectores
2   -0.095068 -0.109064 -0.057665 -0.057392  0.038816 -0.011748  no_efectores
3    0.039729  0.006060  0.010642 -0.101951  0.002585  0.110592  no_efectores
4    0.066765 -0.145637  0.111991 -0.060156  0.124649 -0.075251  no_efectores
..        …         …         …         …         …         …             …
495 -0.016122 -0.008854 -0.006918 -0.002632  0.008072  0.000604  no_efectores
496 -0.072841  0.047593  0.018943 -0.032165  0.038450  0.050388  no_efectores
497  0.064920 -0.014805 -0.202582  0.027678  0.012787  0.087755  no_efectores
498 -0.018013  0.032592  0.020547 -0.070968 -0.049678 -0.012778  no_efectores
499 -0.155346  0.093062 -0.067696 -0.085322 -0.142310 -0.087601  no_efectores

[500 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores archaea dataset 4, con valores atípicos.
Estadísticas.

```
               X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.013518    0.007634    0.013228    0.020872    0.005881   -0.002230
std      0.076011    0.078510    0.072496    0.070562    0.080854    0.073203
min     -0.294151   -0.350489   -0.306199   -0.251341   -0.386883   -0.407816
25%     -0.027910   -0.033052   -0.026266   -0.022624   -0.038234   -0.041890
50%      0.010733    0.010045    0.014540    0.022068    0.003195   -0.000284
75%      0.050005    0.048476    0.055494    0.061848    0.043716    0.038721
```

```
max      0.328910    0.379892    0.290847    0.313523    0.286807    0.260690

                 X6          X7          X8          X9         X10         X11  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.007958    0.006695   -0.000371    0.003822   -0.000322    0.006538
std      0.075588    0.077428    0.073616    0.073755    0.072718    0.074902
min     -0.430380   -0.316491   -0.345979   -0.410057   -0.407038   -0.350356
25%     -0.036819   -0.034138   -0.039089   -0.035857   -0.037306   -0.031710
50%      0.006116    0.010372    0.000665    0.008191    0.002724    0.009060
75%      0.050457    0.050811    0.042596    0.045926    0.041213    0.049143
max      0.364983    0.617846    0.326724    0.231231    0.252051    0.284585

                X12
count  500.000000
mean     0.004524
std      0.080955
min     -0.396610
25%     -0.041707
50%      0.003800
75%      0.045941
max      0.620833
```





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']
```

```python
    #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 4,
sin valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.093449  0.103198  0.024476  0.004098 -0.034269  0.011867 -0.004485
1   -0.026317 -0.029884  0.133822  0.019048 -0.013586 -0.002166  0.008559
2   -0.125665  0.091094  0.044624  0.148923  0.045763  0.090525 -0.071119
3    0.023092  0.056525  0.034590 -0.110905 -0.096042 -0.016991 -0.153657
4    0.046318 -0.017352  0.048512 -0.002018  0.009757 -0.032974 -0.049354
..        …         …         …         …         …         …         …
495  0.077209  0.130857 -0.001088  0.020975  0.128191  0.060350 -0.030527
496  0.058449  0.047566  0.053842  0.021811  0.019159 -0.053528 -0.026796
497  0.038507  0.050909  0.064274  0.069369  0.035609 -0.066948  0.044645
498 -0.010068  0.033854  0.006513  0.027894 -0.012882  0.013830  0.052302
499  0.033132  0.020666  0.110568 -0.025619  0.012764  0.044836  0.070477

           X7        X8        X9       X10       X11       X12        X13
0   -0.113482 -0.001697  0.065546 -0.054329 -0.043361 -0.174597  efectores
1   -0.063992  0.006311 -0.011478 -0.019367 -0.006123 -0.078499  efectores
2   -0.124123  0.010813  0.087544 -0.069030  0.086131 -0.055990  efectores
3    0.105717 -0.116607 -0.071618 -0.059377  0.083385 -0.129894  efectores
4   -0.042160 -0.006835 -0.005385  0.053977 -0.036500 -0.066198  efectores
..        …         …         …         …         …         …          …
495 -0.162259  0.035005 -0.039880  0.019705  0.064380  0.163298  efectores
```

```
496 -0.070468 -0.052426 -0.092558  0.003850 -0.030561  0.036381  efectores
497 -0.129053 -0.099917  0.055308 -0.081601 -0.122854 -0.052929  efectores
498  0.006947 -0.090983 -0.098329  0.032305 -0.028716 -0.033192  efectores
499  0.006391  0.006202 -0.040613 -0.045563 -0.011576 -0.085682  efectores

[467 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro_mass efectores archaea dataset 4,
sin valores atípicos.
Estadísticas.

```
               X0          X1          X2          X3          X4          X5  \
count  467.000000  467.000000  467.000000  467.000000  467.000000  467.000000
mean     0.004531    0.018075    0.013627    0.014127   -0.006280    0.001970
std      0.066341    0.063874    0.063090    0.062753    0.062113    0.064172
min     -0.197952   -0.184630   -0.148304   -0.174287   -0.199892   -0.192566
25%     -0.039696   -0.019881   -0.028171   -0.023284   -0.042076   -0.036415
50%      0.005060    0.017673    0.010465    0.016147   -0.001657    0.005072
75%      0.047230    0.058430    0.052042    0.054141    0.033050    0.041022
max      0.213739    0.196858    0.188834    0.199065    0.171421    0.185627

               X6          X7          X8          X9         X10         X11  \
count  467.000000  467.000000  467.000000  467.000000  467.000000  467.000000
mean     0.022595   -0.000820   -0.003949    0.007687    0.003860   -0.003108
std      0.067921    0.064364    0.073109    0.072441    0.064624    0.062283
min     -0.162351   -0.198585   -0.235007   -0.205981   -0.202402   -0.196794
25%     -0.022513   -0.036396   -0.049484   -0.038786   -0.040519   -0.042126
50%      0.020586    0.002986   -0.001670    0.006659    0.000731   -0.001662
75%      0.066419    0.043126    0.038357    0.056474    0.044855    0.035007
max      0.204287    0.196071    0.203619    0.209361    0.193456    0.188858

              X12
count  467.000000
mean     0.016381
std      0.068119
min     -0.189123
25%     -0.029670
50%      0.012093
75%      0.066112
max      0.213417
```

no_efectores
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores archaea dataset 4,
sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6
0	-0.010315	-0.061416	-0.039641	0.021242	0.054646	-0.006399	-0.036350
1	-0.059650	0.010867	-0.034683	0.041115	-0.058475	-0.014403	0.010750
2	0.130371	0.044571	0.018364	0.013825	-0.098359	-0.021772	-0.117048
3	-0.113533	0.010284	0.059626	0.039931	-0.022758	-0.022429	0.004175
4	-0.030411	0.147539	-0.131060	0.085225	-0.105771	0.120964	-0.092066
..	…	…	…	…	…	…	…
495	-0.004383	-0.053191	0.040575	0.032951	-0.021183	-0.005124	-0.008713
496	-0.024239	0.027761	0.077498	-0.020696	-0.083397	-0.040206	0.008057
497	-0.007380	0.006793	-0.129347	-0.097535	0.011029	-0.071428	-0.043595
498	0.038281	-0.029552	-0.076342	-0.030407	-0.081197	0.022590	0.093086
499	0.043463	0.009990	-0.002767	-0.031961	-0.054017	-0.061926	0.011879

	X7	X8	X9	X10	X11	X12	X13
0	0.054648	0.034022	-0.005984	-0.011492	-0.015608	0.011187	no_efectores
1	-0.032587	-0.062418	0.031327	0.002862	-0.078312	-0.013321	no_efectores
2	-0.095068	-0.109064	-0.057665	-0.057392	0.038816	-0.011748	no_efectores
3	0.039729	0.006060	0.010642	-0.101951	0.002585	0.110592	no_efectores
4	0.066765	-0.145637	0.111991	-0.060156	0.124649	-0.075251	no_efectores
..	…	…	…	…	…	…	…
495	-0.016122	-0.008854	-0.006918	-0.002632	0.008072	0.000604	no_efectores
496	-0.072841	0.047593	0.018943	-0.032165	0.038450	0.050388	no_efectores
497	0.064920	-0.014805	-0.202582	0.027678	0.012787	0.087755	no_efectores
498	-0.018013	0.032592	0.020547	-0.070968	-0.049678	-0.012778	no_efectores
499	-0.155346	0.093062	-0.067696	-0.085322	-0.142310	-0.087601	no_efectores

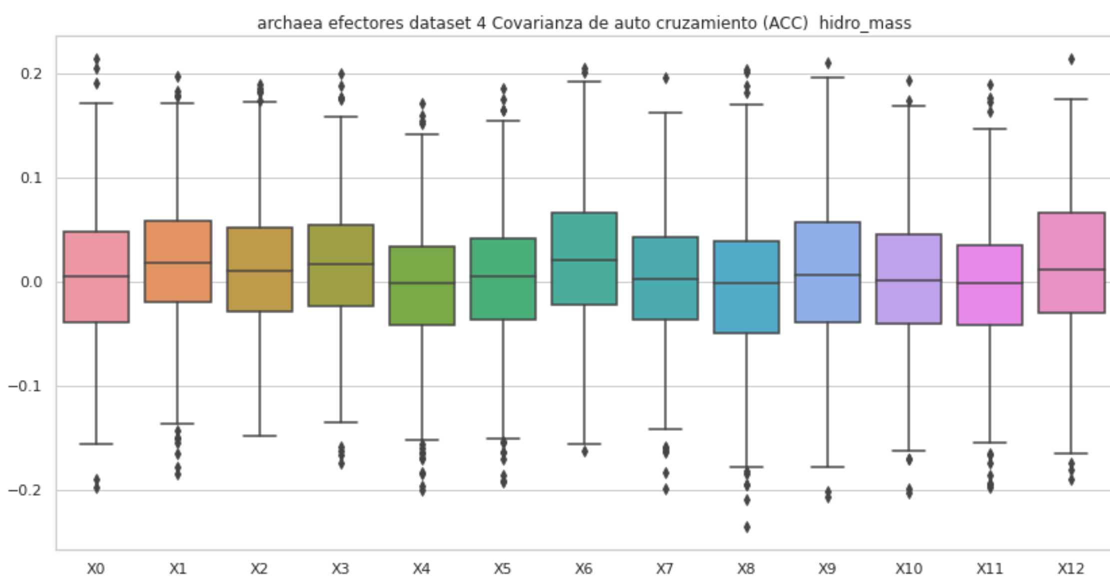
[453 rows x 14 columns]

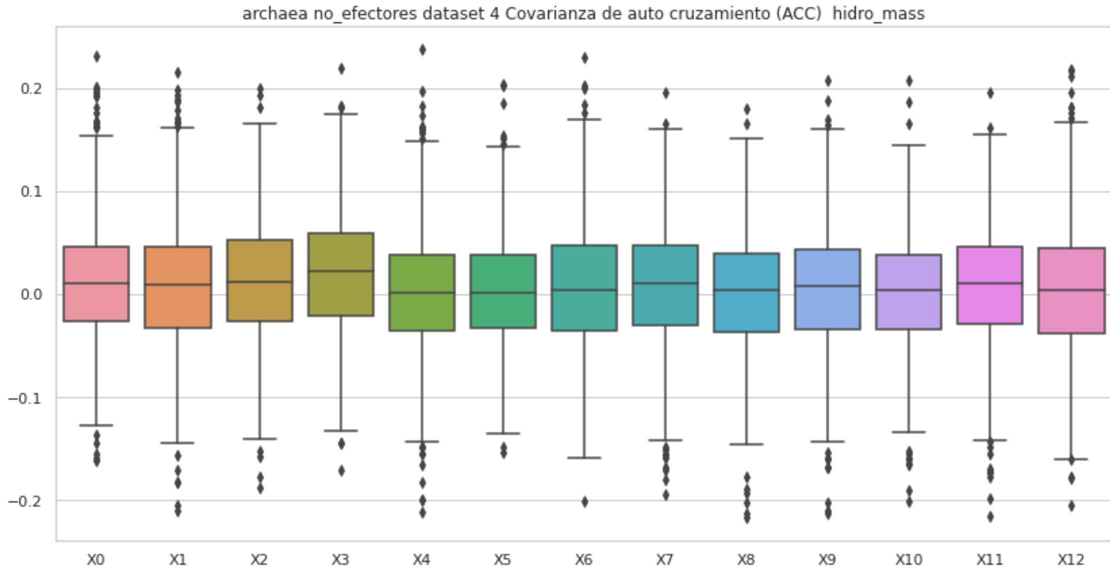
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores archaea dataset 4, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5
count	453.000000	453.000000	453.000000	453.000000	453.000000	453.000000
mean	0.012988	0.008212	0.012341	0.021632	0.002707	0.001990
std	0.064634	0.064788	0.063306	0.062315	0.067376	0.059296
min	-0.160703	-0.209916	-0.187512	-0.170319	-0.211070	-0.153614
25%	-0.026177	-0.032362	-0.025842	-0.021238	-0.035635	-0.033166
50%	0.010835	0.009121	0.012226	0.022548	0.001202	0.001757
75%	0.046554	0.045349	0.052186	0.059397	0.038211	0.038138
max	0.231248	0.214626	0.199214	0.218613	0.237465	0.203765

	X6	X7	X8	X9	X10	X11
count	453.000000	453.000000	453.000000	453.000000	453.000000	453.000000
mean	0.007705	0.006625	0.000621	0.004163	0.001579	0.007958
std	0.063261	0.062792	0.062970	0.065032	0.059568	0.063147

```
min     -0.200511  -0.194039  -0.216476  -0.212710  -0.201184  -0.215448
25%     -0.035778  -0.030109  -0.037116  -0.034651  -0.034419  -0.029323
50%      0.003863   0.011194   0.003831   0.008350   0.003903   0.009916
75%      0.047560   0.046871   0.039015   0.043977   0.038400   0.046354
max      0.229503   0.195577   0.179989   0.207666   0.206784   0.196071

               X12
count   453.000000
mean      0.005416
std       0.065648
min      -0.204388
25%      -0.037351
50%       0.003848
75%       0.044632
max       0.217461
```





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")
```

```python
    #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 4, con
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.093449  0.103198  0.024476  0.004098 -0.034269  0.011867 -0.004485
1   -0.026317 -0.029884  0.133822  0.019048 -0.013586 -0.002166  0.008559
2   -0.125665  0.091094  0.044624  0.148923  0.045763  0.090525 -0.071119
3    0.023092  0.056525  0.034590 -0.110905 -0.096042 -0.016991 -0.153657
4    0.046318 -0.017352  0.048512 -0.002018  0.009757 -0.032974 -0.049354
..        …        …        …        …        …        …        …
495  0.077209  0.130857 -0.001088  0.020975  0.128191  0.060350 -0.030527
496  0.058449  0.047566  0.053842  0.021811  0.019159 -0.053528 -0.026796
497  0.038507  0.050909  0.064274  0.069369  0.035609 -0.066948  0.044645
498 -0.010068  0.033854  0.006513  0.027894 -0.012882  0.013830  0.052302
499  0.033132  0.020666  0.110568 -0.025619  0.012764  0.044836  0.070477

           X7        X8        X9       X10       X11       X12        X13
0   -0.113482 -0.001697  0.065546 -0.054329 -0.043361 -0.174597  efectores
1   -0.063992  0.006311 -0.011478 -0.019367 -0.006123 -0.078499  efectores
2   -0.124123  0.010813  0.087544 -0.069030  0.086131 -0.055990  efectores
3    0.105717 -0.116607 -0.071618 -0.059377  0.083385 -0.129894  efectores
4   -0.042160 -0.006835 -0.005385  0.053977 -0.036500 -0.066198  efectores
..        …        …        …        …        …        …          …
495 -0.162259  0.035005 -0.039880  0.019705  0.064380  0.163298  efectores
496 -0.070468 -0.052426 -0.092558  0.003850 -0.030561  0.036381  efectores
497 -0.129053 -0.099917  0.055308 -0.081601 -0.122854 -0.052929  efectores
498  0.006947 -0.090983 -0.098329  0.032305 -0.028716 -0.033192  efectores
499  0.006391  0.006202 -0.040613 -0.045563 -0.011576 -0.085682  efectores

[500 rows x 14 columns]


Covarianza de auto cruzamiento (ACC)  mass efectores archaea dataset 4, con
valores atípicos.
Estadísticas.

               X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.005620    0.017056    0.011178    0.014521   -0.009450    0.000990
```

std	0.071455	0.068047	0.067464	0.067014	0.069897	0.066528
min	-0.411431	-0.218952	-0.258376	-0.209807	-0.258579	-0.204442
25%	-0.039204	-0.020888	-0.031667	-0.024978	-0.046481	-0.038551
50%	0.006825	0.017552	0.006734	0.016236	-0.002745	0.003903
75%	0.051702	0.058392	0.051930	0.057831	0.033261	0.041053
max	0.222100	0.241034	0.271296	0.250239	0.171421	0.201266

	X6	X7	X8	X9	X10	X11 \
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	0.021188	-0.000286	-0.004550	0.007669	0.002708	-0.004569
std	0.069940	0.066317	0.079660	0.077076	0.069242	0.067342
min	-0.194555	-0.198585	-0.451489	-0.213733	-0.261583	-0.222320
25%	-0.026817	-0.037502	-0.050406	-0.040899	-0.041721	-0.044494
50%	0.018307	0.003269	-0.001683	0.006546	0.000539	-0.002359
75%	0.066886	0.043293	0.041267	0.056753	0.045885	0.035030
max	0.207433	0.259834	0.291169	0.400687	0.193456	0.236683

	X12
count	500.000000
mean	0.018125
std	0.070402
min	-0.223231
25%	-0.029602
50%	0.015106
75%	0.070658
max	0.215409

no_efectores
Covarianza de auto cruzamiento (ACC) mass no_efectores archaea dataset 4, con valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	-0.010315	-0.061416	-0.039641	0.021242	0.054646	-0.006399	-0.036350
1	-0.059650	0.010867	-0.034683	0.041115	-0.058475	-0.014403	0.010750
2	0.130371	0.044571	0.018364	0.013825	-0.098359	-0.021772	-0.117048
3	-0.113533	0.010284	0.059626	0.039931	-0.022758	-0.022429	0.004175
4	-0.030411	0.147539	-0.131060	0.085225	-0.105771	0.120964	-0.092066
..	…	…	…	…	…	…	…
495	-0.004383	-0.053191	0.040575	0.032951	-0.021183	-0.005124	-0.008713
496	-0.024239	0.027761	0.077498	-0.020696	-0.083397	-0.040206	0.008057
497	-0.007380	0.006793	-0.129347	-0.097535	0.011029	-0.071428	-0.043595
498	0.038281	-0.029552	-0.076342	-0.030407	-0.081197	0.022590	0.093086
499	0.043463	0.009990	-0.002767	-0.031961	-0.054017	-0.061926	0.011879

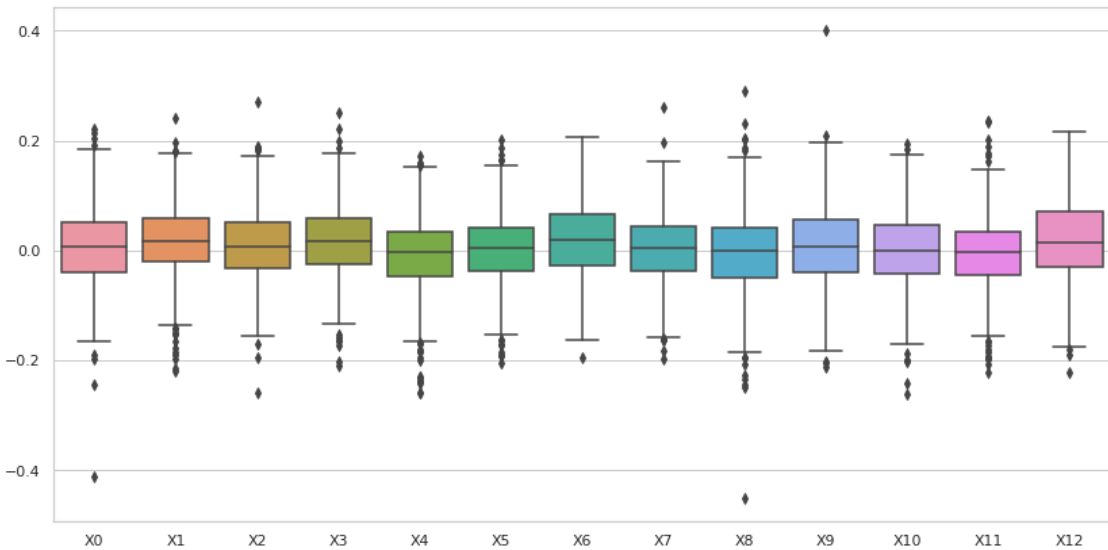
	X7	X8	X9	X10	X11	X12	X13

```
0    0.054648  0.034022 -0.005984 -0.011492 -0.015608  0.011187  no_efectores
1   -0.032587 -0.062418  0.031327  0.002862 -0.078312 -0.013321  no_efectores
2   -0.095068 -0.109064 -0.057665 -0.057392  0.038816 -0.011748  no_efectores
3    0.039729  0.006060  0.010642 -0.101951  0.002585  0.110592  no_efectores
4    0.066765 -0.145637  0.111991 -0.060156  0.124649 -0.075251  no_efectores
..        …         …         …         …         …         …             …
495 -0.016122 -0.008854 -0.006918 -0.002632  0.008072  0.000604  no_efectores
496 -0.072841  0.047593  0.018943 -0.032165  0.038450  0.050388  no_efectores
497  0.064920 -0.014805 -0.202582  0.027678  0.012787  0.087755  no_efectores
498 -0.018013  0.032592  0.020547 -0.070968 -0.049678 -0.012778  no_efectores
499 -0.155346  0.093062 -0.067696 -0.085322 -0.142310 -0.087601  no_efectores

[500 rows x 14 columns]
```

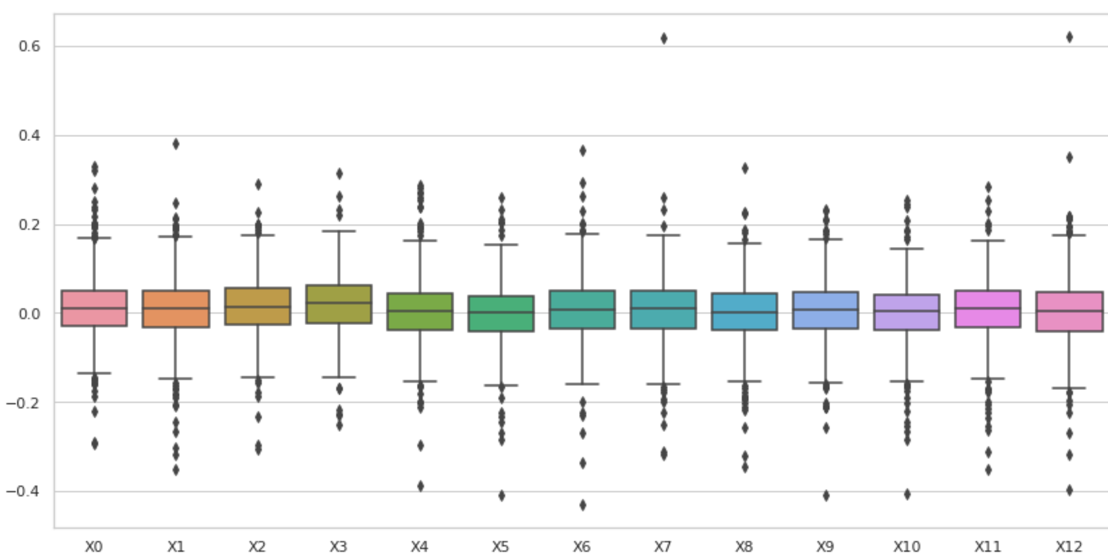
Covarianza de auto cruzamiento (ACC) mass no_efectores archaea dataset 4, con valores atípicos.
Estadísticas.

```
               X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.013518    0.007634    0.013228    0.020872    0.005881   -0.002230
std      0.076011    0.078510    0.072496    0.070562    0.080854    0.073203
min     -0.294151   -0.350489   -0.306199   -0.251341   -0.386883   -0.407816
25%     -0.027910   -0.033052   -0.026266   -0.022624   -0.038234   -0.041890
50%      0.010733    0.010045    0.014540    0.022068    0.003195   -0.000284
75%      0.050005    0.048476    0.055494    0.061848    0.043716    0.038721
max      0.328910    0.379892    0.290847    0.313523    0.286807    0.260690

               X6          X7          X8          X9         X10         X11  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.007958    0.006695   -0.000371    0.003822   -0.000322    0.006538
std      0.075588    0.077428    0.073616    0.073755    0.072718    0.074902
min     -0.430380   -0.316491   -0.345979   -0.410057   -0.407038   -0.350356
25%     -0.036819   -0.034138   -0.039089   -0.035857   -0.037306   -0.031710
50%      0.006116    0.010372    0.000665    0.008191    0.002724    0.009060
75%      0.050457    0.050811    0.042596    0.045926    0.041213    0.049143
max      0.364983    0.617846    0.326724    0.231231    0.252051    0.284585

              X12
count  500.000000
mean     0.004524
std      0.080955
min     -0.396610
25%     -0.041707
50%      0.003800
75%      0.045941
max      0.620833
```

archaea efectores dataset 4 Covarianza de auto cruzamiento (ACC) mass con valores atípicos.



archaea no_efectores dataset 4 Covarianza de auto cruzamiento (ACC) mass con valores atípicos.



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 4, sin
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.093449  0.103198  0.024476  0.004098 -0.034269  0.011867 -0.004485
1   -0.026317 -0.029884  0.133822  0.019048 -0.013586 -0.002166  0.008559
2   -0.125665  0.091094  0.044624  0.148923  0.045763  0.090525 -0.071119
3    0.023092  0.056525  0.034590 -0.110905 -0.096042 -0.016991 -0.153657
4    0.046318 -0.017352  0.048512 -0.002018  0.009757 -0.032974 -0.049354
..        …         …         …         …         …         …         …
495  0.077209  0.130857 -0.001088  0.020975  0.128191  0.060350 -0.030527
496  0.058449  0.047566  0.053842  0.021811  0.019159 -0.053528 -0.026796
497  0.038507  0.050909  0.064274  0.069369  0.035609 -0.066948  0.044645
498 -0.010068  0.033854  0.006513  0.027894 -0.012882  0.013830  0.052302
499  0.033132  0.020666  0.110568 -0.025619  0.012764  0.044836  0.070477

           X7        X8        X9       X10       X11       X12        X13
0   -0.113482 -0.001697  0.065546 -0.054329 -0.043361 -0.174597  efectores
1   -0.063992  0.006311 -0.011478 -0.019367 -0.006123 -0.078499  efectores
2   -0.124123  0.010813  0.087544 -0.069030  0.086131 -0.055990  efectores
3    0.105717 -0.116607 -0.071618 -0.059377  0.083385 -0.129894  efectores
4   -0.042160 -0.006835 -0.005385  0.053977 -0.036500 -0.066198  efectores
..        …         …         …         …         …         …          …
495 -0.162259  0.035005 -0.039880  0.019705  0.064380  0.163298  efectores
496 -0.070468 -0.052426 -0.092558  0.003850 -0.030561  0.036381  efectores
497 -0.129053 -0.099917  0.055308 -0.081601 -0.122854 -0.052929  efectores
498  0.006947 -0.090983 -0.098329  0.032305 -0.028716 -0.033192  efectores
499  0.006391  0.006202 -0.040613 -0.045563 -0.011576 -0.085682  efectores

[467 rows x 14 columns]


Covarianza de auto cruzamiento (ACC)  mass efectores archaea dataset 4, sin
valores atípicos.
Estadísticas.

               X0          X1          X2          X3          X4          X5  \
count  467.000000  467.000000  467.000000  467.000000  467.000000  467.000000
mean     0.004531    0.018075    0.013627    0.014127   -0.006280    0.001970
std      0.066341    0.063874    0.063090    0.062753    0.062113    0.064172
min     -0.197952   -0.184630   -0.148304   -0.174287   -0.199892   -0.192566
25%     -0.039696   -0.019881   -0.028171   -0.023284   -0.042076   -0.036415
50%      0.005060    0.017673    0.010465    0.016147   -0.001657    0.005072
75%      0.047230    0.058430    0.052042    0.054141    0.033050    0.041022
```

max	0.213739	0.196858	0.188834	0.199065	0.171421	0.185627

	X6	X7	X8	X9	X10	X11 \
count	467.000000	467.000000	467.000000	467.000000	467.000000	467.000000
mean	0.022595	-0.000820	-0.003949	0.007687	0.003860	-0.003108
std	0.067921	0.064364	0.073109	0.072441	0.064624	0.062283
min	-0.162351	-0.198585	-0.235007	-0.205981	-0.202402	-0.196794
25%	-0.022513	-0.036396	-0.049484	-0.038786	-0.040519	-0.042126
50%	0.020586	0.002986	-0.001670	0.006659	0.000731	-0.001662
75%	0.066419	0.043126	0.038357	0.056474	0.044855	0.035007
max	0.204287	0.196071	0.203619	0.209361	0.193456	0.188858

	X12
count	467.000000
mean	0.016381
std	0.068119
min	-0.189123
25%	-0.029670
50%	0.012093
75%	0.066112
max	0.213417

Covarianza de auto cruzamiento (ACC) mass no_efectores archaea dataset 4, sin valores atípicos.
Valores del documento csv.

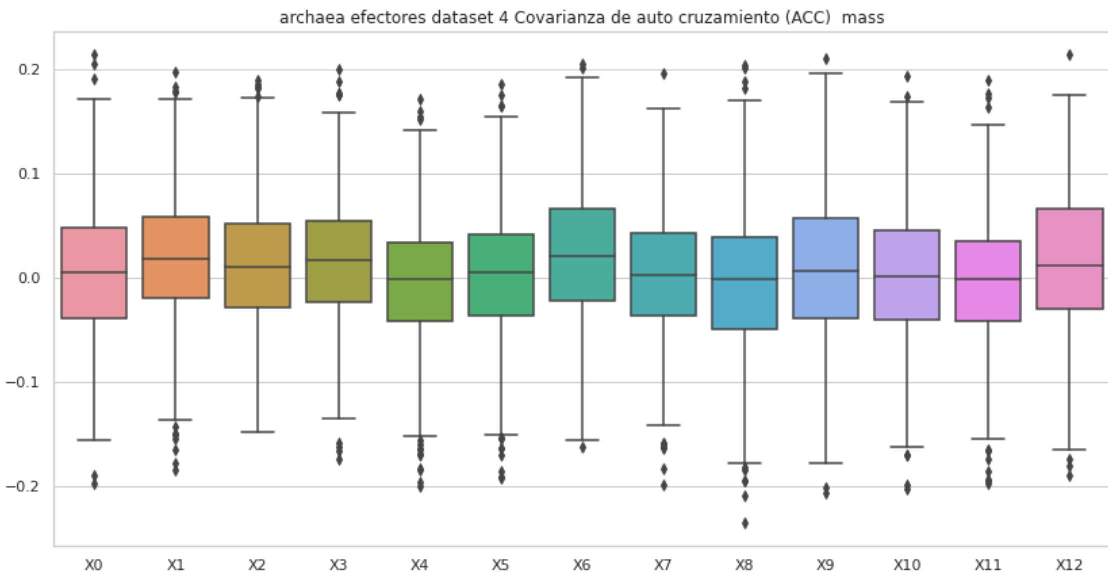
	X0	X1	X2	X3	X4	X5	X6 \
0	-0.010315	-0.061416	-0.039641	0.021242	0.054646	-0.006399	-0.036350
1	-0.059650	0.010867	-0.034683	0.041115	-0.058475	-0.014403	0.010750
2	0.130371	0.044571	0.018364	0.013825	-0.098359	-0.021772	-0.117048
3	-0.113533	0.010284	0.059626	0.039931	-0.022758	-0.022429	0.004175
4	-0.030411	0.147539	-0.131060	0.085225	-0.105771	0.120964	-0.092066
..	…	…	…	…	…	…	…
495	-0.004383	-0.053191	0.040575	0.032951	-0.021183	-0.005124	-0.008713
496	-0.024239	0.027761	0.077498	-0.020696	-0.083397	-0.040206	0.008057
497	-0.007380	0.006793	-0.129347	-0.097535	0.011029	-0.071428	-0.043595
498	0.038281	-0.029552	-0.076342	-0.030407	-0.081197	0.022590	0.093086
499	0.043463	0.009990	-0.002767	-0.031961	-0.054017	-0.061926	0.011879

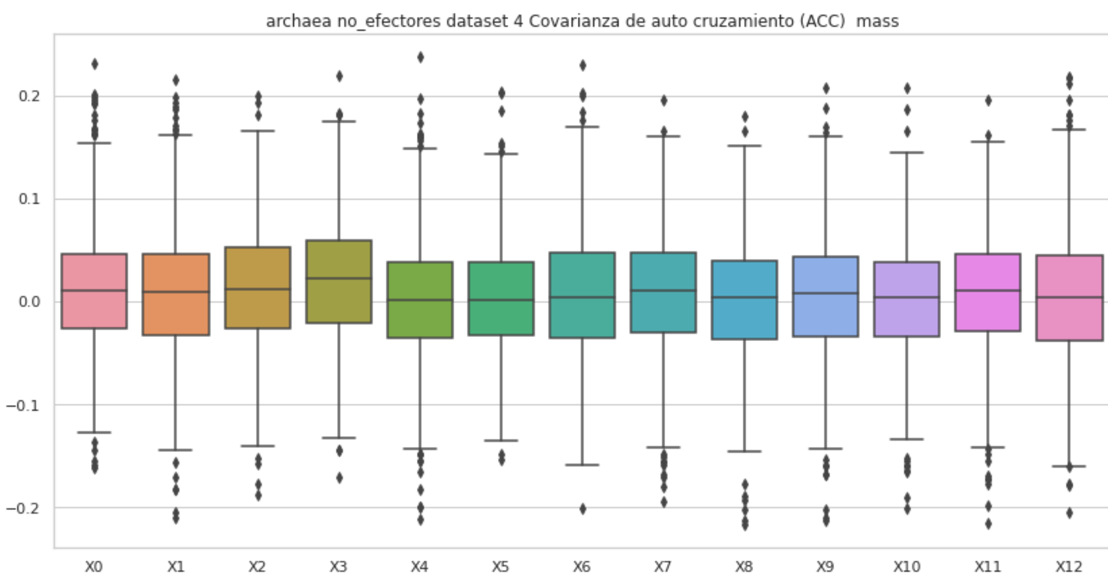
	X7	X8	X9	X10	X11	X12	X13
0	0.054648	0.034022	-0.005984	-0.011492	-0.015608	0.011187	no_efectores
1	-0.032587	-0.062418	0.031327	0.002862	-0.078312	-0.013321	no_efectores
2	-0.095068	-0.109064	-0.057665	-0.057392	0.038816	-0.011748	no_efectores
3	0.039729	0.006060	0.010642	-0.101951	0.002585	0.110592	no_efectores
4	0.066765	-0.145637	0.111991	-0.060156	0.124649	-0.075251	no_efectores
..	…	…	…	…	…	…	…

```
495 -0.016122 -0.008854 -0.006918 -0.002632  0.008072  0.000604  no_efectores
496 -0.072841  0.047593  0.018943 -0.032165  0.038450  0.050388  no_efectores
497  0.064920 -0.014805 -0.202582  0.027678  0.012787  0.087755  no_efectores
498 -0.018013  0.032592  0.020547 -0.070968 -0.049678 -0.012778  no_efectores
499 -0.155346  0.093062 -0.067696 -0.085322 -0.142310 -0.087601  no_efectores

[453 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) mass no_efectores archaea dataset 4, sin valores atípicos.
Estadísticas.

```
               X0          X1          X2          X3          X4          X5  \
count  453.000000  453.000000  453.000000  453.000000  453.000000  453.000000
mean     0.012988    0.008212    0.012341    0.021632    0.002707    0.001990
std      0.064634    0.064788    0.063306    0.062315    0.067376    0.059296
min     -0.160703   -0.209916   -0.187512   -0.170319   -0.211070   -0.153614
25%     -0.026177   -0.032362   -0.025842   -0.021238   -0.035635   -0.033166
50%      0.010835    0.009121    0.012226    0.022548    0.001202    0.001757
75%      0.046554    0.045349    0.052186    0.059397    0.038211    0.038138
max      0.231248    0.214626    0.199214    0.218613    0.237465    0.203765

               X6          X7          X8          X9         X10         X11  \
count  453.000000  453.000000  453.000000  453.000000  453.000000  453.000000
mean     0.007705    0.006625    0.000621    0.004163    0.001579    0.007958
std      0.063261    0.062792    0.062970    0.065032    0.059568    0.063147
min     -0.200511   -0.194039   -0.216476   -0.212710   -0.201184   -0.215448
25%     -0.035778   -0.030109   -0.037116   -0.034651   -0.034419   -0.029323
50%      0.003863    0.011194    0.003831    0.008350    0.003903    0.009916
75%      0.047560    0.046871    0.039015    0.043977    0.038400    0.046354
max      0.229503    0.195577    0.179989    0.207666    0.206784    0.196071

              X12
count  453.000000
mean     0.005416
std      0.065648
min     -0.204388
25%     -0.037351
50%      0.003848
75%      0.044632
max      0.217461
```





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=""
```

```python
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 4, con
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.023599 -0.119102  0.045873  0.089409  0.000330  0.009812 -0.007358
1    0.027539 -0.002890  0.072907  0.101504 -0.054247  0.045809  0.080831
2    0.084010 -0.052690  0.192480  0.011721 -0.044626  0.086301 -0.027815
3   -0.003458 -0.023364 -0.112025 -0.261674  0.001403 -0.042670 -0.084924
4    0.053982  0.044888  0.092037  0.051263 -0.068760  0.046713  0.074821
..        …         …         …         …         …         …         …
495 -0.025609 -0.155105  0.015078 -0.039138  0.059628  0.069175 -0.031114
496  0.122544  0.003166 -0.131903 -0.078440 -0.120873 -0.148583 -0.002221
497  0.035108 -0.010148 -0.044492  0.000693 -0.017713 -0.111841  0.030730
498  0.032727  0.007099  0.082129  0.038846  0.013278  0.072186  0.008628
499 -0.003781  0.005260  0.192254  0.092038  0.040065  0.131077  0.016567

           X7        X8        X9       X10       X11       X12        X13
0   -0.072046  0.018149 -0.009583 -0.031504  0.048961  0.046599  efectores
1    0.071706  0.068315  0.062549  0.030347  0.081079  0.043004  efectores
2   -0.175399  0.141757 -0.100644 -0.129573 -0.079594  0.088971  efectores
```

```
3    0.033704  0.092932 -0.013958  0.118043 -0.017958 -0.256219  efectores
4    0.075436  0.052882  0.060129 -0.002297  0.087722  0.049985  efectores
..        …         …         …         …         …         …         …
495 -0.122825 -0.074691 -0.006329  0.194587 -0.138758 -0.297530  efectores
496  0.110210  0.034469 -0.014412 -0.077239  0.095034  0.041380  efectores
497  0.087745 -0.029940 -0.067257 -0.091665 -0.091948 -0.045276  efectores
498  0.022808 -0.007419 -0.002994 -0.003159 -0.031936 -0.013246  efectores
499  0.098559  0.110076 -0.035266  0.113409  0.050167  0.062576  efectores

[500 rows x 14 columns]
```

```
Covarianza de auto cruzamiento (ACC)  hidro efectores archaea dataset 4, con
valores atípicos.
Estadísticas.

               X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.021162   -0.027917    0.042736    0.032232   -0.017824   -0.013873
std      0.088161    0.097659    0.088446    0.090297    0.090339    0.080312
min     -0.482349   -0.358469   -0.326423   -0.261674   -0.284337   -0.361520
25%     -0.029445   -0.102429   -0.009579   -0.015678   -0.082939   -0.060165
50%      0.025171   -0.018476    0.034511    0.032119   -0.020574   -0.013316
75%      0.074386    0.045052    0.101058    0.080648    0.048649    0.035274
max      0.286193    0.223364    0.274428    0.642387    0.227553    0.192785

               X6          X7          X8          X9         X10         X11  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.031859    0.025249   -0.002385   -0.005141    0.013323    0.008923
std      0.086202    0.086440    0.091671    0.089359    0.082168    0.084661
min     -0.234784   -0.311959   -0.318012   -0.372742   -0.344083   -0.211401
25%     -0.019444   -0.026432   -0.050567   -0.049129   -0.034622   -0.042985
50%      0.026860    0.022126    0.001736   -0.000946    0.003058   -0.002579
75%      0.077630    0.079402    0.048331    0.042455    0.067036    0.057637
max      0.372214    0.417418    0.321769    0.273102    0.341757    0.410753

              X12
count  500.000000
mean    -0.011918
std      0.088920
min     -0.339636
25%     -0.059701
50%     -0.008337
75%      0.043100
max      0.229559
```

no_efectores
Covarianza de auto cruzamiento (ACC) hidro no_efectores archaea dataset 4, con valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.060676	-0.086753	-0.052375	0.013684	0.016610	0.002621	-0.020795	
1	0.006303	-0.113237	0.027335	0.067276	0.008722	0.006787	0.011732	
2	-0.055415	-0.044293	0.042772	-0.005444	0.031207	0.000153	-0.171071	
3	0.067628	-0.146925	0.024326	0.086630	-0.217580	-0.032058	0.084243	
4	-0.025899	0.070184	0.025430	0.051474	-0.082660	-0.087294	0.052721	
..	…	…	…	…	…	…	…	
495	-0.044775	-0.051500	0.002761	0.073964	-0.005189	-0.068522	0.016585	
496	0.101361	0.016394	0.011841	0.085937	0.101521	-0.010971	0.047655	
497	-0.153980	0.001925	0.038271	0.083735	-0.278127	0.095579	0.194309	
498	-0.249557	-0.156518	0.115002	0.046949	-0.017756	-0.216465	0.061443	
499	-0.096826	-0.016107	0.181018	0.053107	-0.170969	-0.099005	0.056286	

	X7	X8	X9	X10	X11	X12	X13
0	-0.081234	-0.126053	0.027890	0.073236	0.011599	-0.061858	no_efectores
1	0.004658	-0.074066	0.028770	-0.043993	-0.017353	0.049799	no_efectores
2	0.045547	-0.126074	-0.024066	0.019529	-0.060886	0.015539	no_efectores
3	0.279619	0.015159	-0.043973	-0.140974	-0.039288	-0.029561	no_efectores
4	0.041423	-0.033315	-0.030556	-0.024614	-0.015525	-0.097494	no_efectores
..	…	…	…	…	…	…	…
495	0.018141	-0.037611	0.009500	0.057016	0.063540	0.001711	no_efectores
496	0.056301	0.081062	0.040469	-0.040485	0.020755	-0.033451	no_efectores
497	-0.163889	0.106296	0.100483	-0.042605	-0.198932	0.205641	no_efectores
498	0.069270	0.037233	-0.077652	-0.031279	0.215341	-0.166172	no_efectores
499	0.156439	-0.065703	0.049009	0.049246	-0.029018	-0.155976	no_efectores

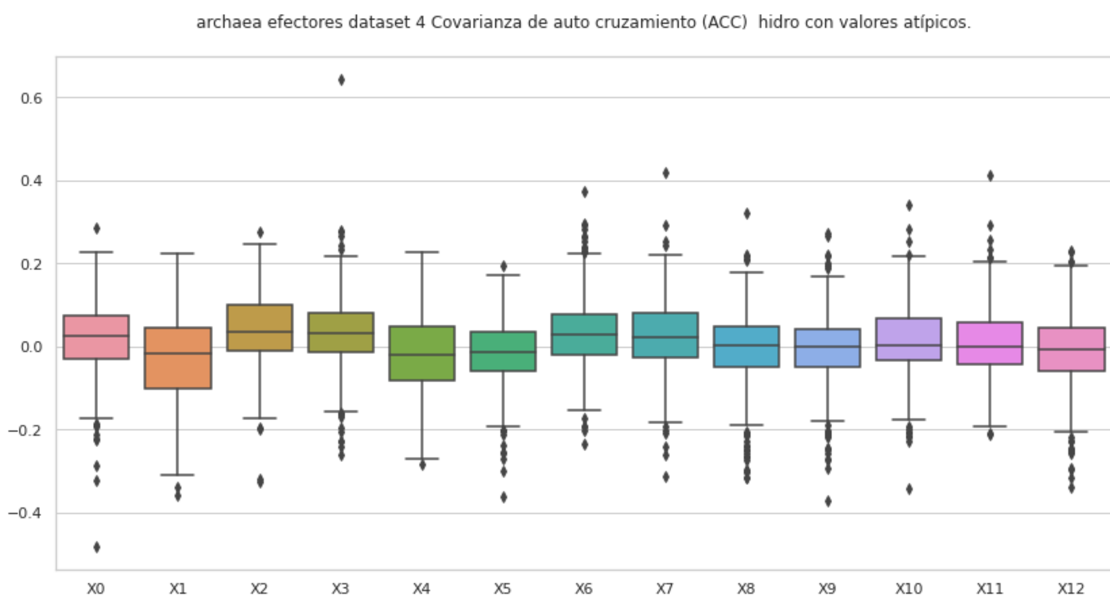
[500 rows x 14 columns]

Covarianza de auto cruzamiento (ACC) hidro no_efectores archaea dataset 4, con valores atípicos.
Estadísticas.

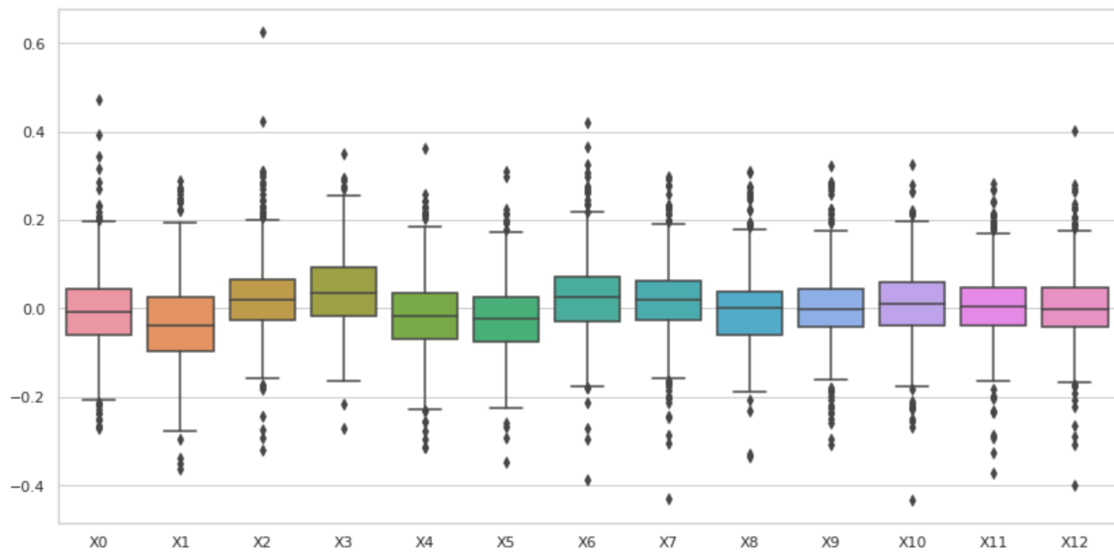
	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	-0.005634	-0.034183	0.024313	0.038716	-0.016422	-0.023055	
std	0.095721	0.099967	0.093607	0.087359	0.090857	0.087926	
min	-0.270496	-0.362807	-0.320673	-0.270569	-0.315408	-0.347567	
25%	-0.059334	-0.097135	-0.027800	-0.018684	-0.069878	-0.075540	
50%	-0.007248	-0.038937	0.018398	0.034722	-0.016265	-0.023361	
75%	0.043707	0.025996	0.064018	0.091411	0.036115	0.025130	
max	0.472152	0.289668	0.624902	0.349323	0.361651	0.309771	

	X6	X7	X8	X9	X10	X11	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.026002	0.015772	-0.002973	0.001949	0.012125	0.005898	
std	0.091365	0.085984	0.085391	0.085078	0.086253	0.083874	
min	-0.387151	-0.430523	-0.334609	-0.307918	-0.433256	-0.372052	
25%	-0.029173	-0.027255	-0.058723	-0.043227	-0.038332	-0.038356	
50%	0.024164	0.017991	0.000924	-0.001463	0.010462	0.005628	
75%	0.069954	0.061154	0.037330	0.045014	0.057760	0.045857	
max	0.421137	0.297983	0.310669	0.321897	0.325266	0.282550	

	X12
count	500.000000
mean	0.001911
std	0.082677
min	-0.399022
25%	-0.041390
50%	-0.002275
75%	0.045521
max	0.401415



archaea no_efectores dataset 4 Covarianza de auto cruzamiento (ACC) hidro con valores atípicos.



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 4, sin valores
atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.023599 -0.119102  0.045873  0.089409  0.000330  0.009812 -0.007358
1    0.027539 -0.002890  0.072907  0.101504 -0.054247  0.045809  0.080831
2    0.084010 -0.052690  0.192480  0.011721 -0.044626  0.086301 -0.027815
4    0.053982  0.044888  0.092037  0.051263 -0.068760  0.046713  0.074821
5    0.031278  0.146464 -0.001636  0.050394  0.073368  0.021528  0.079785
..        …         …         …         …         …         …         …
494 -0.029347 -0.109702  0.084385  0.007307 -0.077029 -0.077176  0.048612
496  0.122544  0.003166 -0.131903 -0.078440 -0.120873 -0.148583 -0.002221
497  0.035108 -0.010148 -0.044492  0.000693 -0.017713 -0.111841  0.030730
498  0.032727  0.007099  0.082129  0.038846  0.013278  0.072186  0.008628
499 -0.003781  0.005260  0.192254  0.092038  0.040065  0.131077  0.016567

           X7        X8        X9       X10       X11       X12        X13
0   -0.072046  0.018149 -0.009583 -0.031504  0.048961  0.046599  efectores
1    0.071706  0.068315  0.062549  0.030347  0.081079  0.043004  efectores
2   -0.175399  0.141757 -0.100644 -0.129573 -0.079594  0.088971  efectores
4    0.075436  0.052882  0.060129 -0.002297  0.087722  0.049985  efectores
5    0.008911 -0.011505  0.020736 -0.029281  0.034120  0.051431  efectores
..        …         …         …         …         …         …          …
494  0.007593  0.021645  0.007873 -0.005033 -0.031050  0.010838  efectores
```

```
496  0.110210  0.034469 -0.014412 -0.077239  0.095034  0.041380  efectores
497  0.087745 -0.029940 -0.067257 -0.091665 -0.091948 -0.045276  efectores
498  0.022808 -0.007419 -0.002994 -0.003159 -0.031936 -0.013246  efectores
499  0.098559  0.110076 -0.035266  0.113409  0.050167  0.062576  efectores

[469 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) efectores archaea dataset 4, sin valores atípicos.
Estadísticas.

```
               X0          X1          X2          X3          X4          X5  \
count  469.000000  469.000000  469.000000  469.000000  469.000000  469.000000
mean     0.023572   -0.023267    0.043810    0.032079   -0.015065   -0.007974
std      0.081763    0.093512    0.084209    0.082544    0.090299    0.072640
min     -0.224249   -0.310964   -0.200124   -0.229024   -0.284337   -0.237068
25%     -0.028365   -0.099240   -0.008476   -0.014723   -0.078559   -0.053834
50%      0.025218   -0.014219    0.034547    0.032430   -0.017635   -0.009352
75%      0.073751    0.047700    0.099425    0.080380    0.051617    0.037470
max      0.227979    0.223364    0.274428    0.263904    0.227553    0.192785

               X6          X7          X8          X9         X10         X11  \
count  469.000000  469.000000  469.000000  469.000000  469.000000  469.000000
mean     0.030660    0.026744    0.003079   -0.002796    0.011825    0.009198
std      0.079757    0.079545    0.081997    0.082124    0.076858    0.080449
min     -0.202080   -0.201381   -0.273021   -0.270800   -0.228159   -0.211401
25%     -0.016770   -0.025285   -0.045181   -0.046613   -0.033393   -0.041259
50%      0.026195    0.022450    0.003571    0.001231    0.001352   -0.001038
75%      0.074208    0.079089    0.049479    0.041912    0.062176    0.057055
max      0.282254    0.252490    0.220815    0.219420    0.251161    0.255024

              X12
count  469.000000
mean    -0.006750
std      0.081731
min     -0.247170
25%     -0.055256
50%     -0.006633
75%      0.043484
max      0.229559
```

no_efectores
Covarianza de auto cruzamiento (ACC) no_efectores archaea dataset 4, sin valores atípicos.
Valores del documento csv.

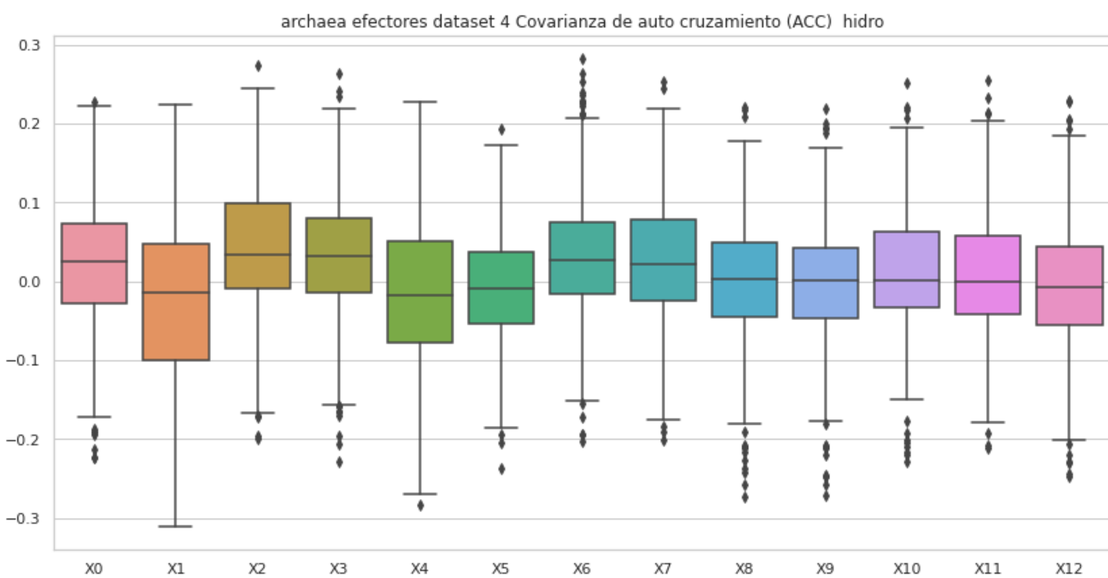
```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.060676 -0.086753 -0.052375  0.013684  0.016610  0.002621 -0.020795
1    0.006303 -0.113237  0.027335  0.067276  0.008722  0.006787  0.011732
2   -0.055415 -0.044293  0.042772 -0.005444  0.031207  0.000153 -0.171071
4   -0.025899  0.070184  0.025430  0.051474 -0.082660 -0.087294  0.052721
5   -0.020909 -0.121002  0.044084  0.129387 -0.042171 -0.076037 -0.015632
..        …         …         …         …         …         …         …
495 -0.044775 -0.051500  0.002761  0.073964 -0.005189 -0.068522  0.016585
496  0.101361  0.016394  0.011841  0.085937  0.101521 -0.010971  0.047655
497 -0.153980  0.001925  0.038271  0.083735 -0.278127  0.095579  0.194309
498 -0.249557 -0.156518  0.115002  0.046949 -0.017756 -0.216465  0.061443
499 -0.096826 -0.016107  0.181018  0.053107 -0.170969 -0.099005  0.056286

           X7        X8        X9       X10       X11       X12           X13
0   -0.081234 -0.126053  0.027890  0.073236  0.011599 -0.061858  no_efectores
1    0.004658 -0.074066  0.028770 -0.043993 -0.017353  0.049799  no_efectores
2    0.045547 -0.126074 -0.024066  0.019529 -0.060886  0.015539  no_efectores
4    0.041423 -0.033315 -0.030556 -0.024614 -0.015525 -0.097494  no_efectores
5    0.039420 -0.123933 -0.024359  0.012455 -0.021397  0.001160  no_efectores
..        …         …         …         …         …         …             …
495  0.018141 -0.037611  0.009500  0.057016  0.063540  0.001711  no_efectores
496  0.056301  0.081062  0.040469 -0.040485  0.020755 -0.033451  no_efectores
497 -0.163889  0.106296  0.100483 -0.042605 -0.198932  0.205641  no_efectores
498  0.069270  0.037233 -0.077652 -0.031279  0.215341 -0.166172  no_efectores
499  0.156439 -0.065703  0.049009  0.049246 -0.029018 -0.155976  no_efectores

[452 rows x 14 columns]
```

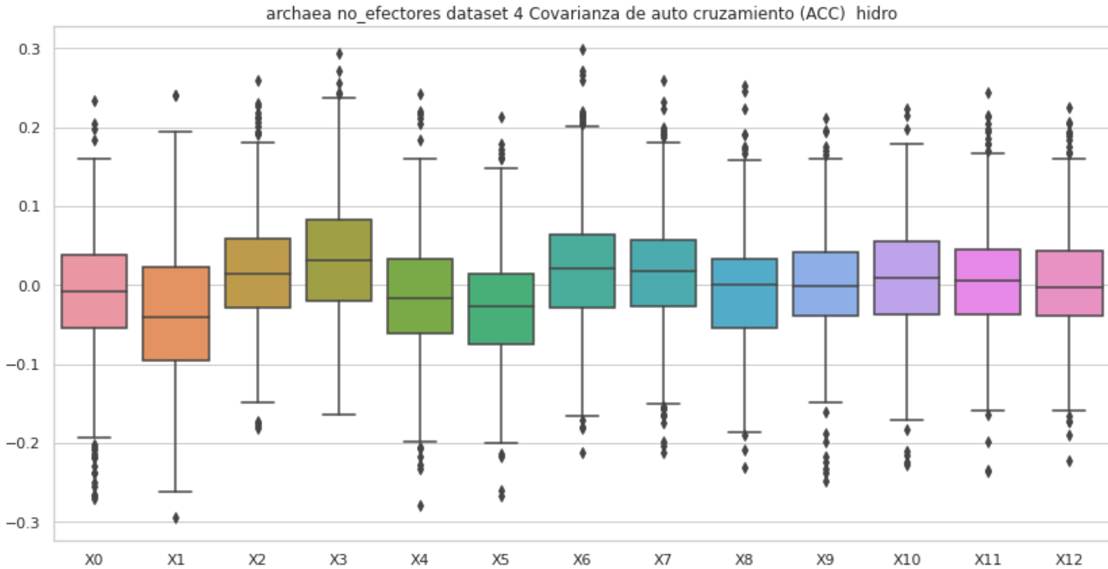
Covarianza de auto cruzamiento (ACC) no_efectores archaea dataset 4, sin valores atípicos.
Estadísticas.

```
               X0          X1          X2          X3          X4          X5  \
count  452.000000  452.000000  452.000000  452.000000  452.000000  452.000000
mean    -0.010171   -0.037580    0.017185    0.032256   -0.016053   -0.028695
std      0.081306    0.085601    0.073742    0.079008    0.078814    0.076148
min     -0.270496   -0.294157   -0.181663   -0.164174   -0.278127   -0.266808
25%     -0.054585   -0.095526   -0.028588   -0.019973   -0.061951   -0.075290
50%     -0.007248   -0.039798    0.014390    0.030910   -0.016127   -0.026716
75%      0.038290    0.022797    0.058421    0.082902    0.032736    0.014813
max      0.233945    0.241080    0.259369    0.293254    0.241633    0.212506

               X6          X7          X8          X9         X10         X11  \
count  452.000000  452.000000  452.000000  452.000000  452.000000  452.000000
mean     0.021837    0.012904   -0.005657   -0.000869    0.010478    0.005184
std      0.076826    0.071802    0.073018    0.069597    0.072967    0.070157
```

min	-0.212074	-0.212141	-0.230372	-0.248753	-0.227266	-0.235345
25%	-0.027819	-0.027255	-0.054155	-0.038221	-0.036930	-0.037805
50%	0.021370	0.017239	0.000800	-0.001655	0.009707	0.005027
75%	0.064211	0.056211	0.033569	0.042124	0.055953	0.044301
max	0.298376	0.259158	0.252280	0.211970	0.222779	0.243098

	X12
count	452.000000
mean	0.000888
std	0.066931
min	-0.222524
25%	-0.039239
50%	-0.002362
75%	0.042790
max	0.224108




archaea no_efectores dataset 4 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