ds5_archaea_limpieza_de_datos

February 1, 2021

Limpieza de datos

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

1 Declaración de variables

[2]:
```
organismo ="archaea"
dataset = 5
nombre = ("ds" + str(dataset) + "_" + str(organismo))
nombre2 = (str(organismo)+ " dataset " + str(dataset))
r2 = ("Datos/resultados/"+ str(organismo) + "/" + str(nombre) + "/
 ↪transformaciones/sin_filtrar")
r3 = ("Datos/resultados/"+ str(organismo) + "/" + str(nombre) + "/
 ↪transformaciones/sin_atipicos")

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

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

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

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

2 Composición de aminoácidos (AAC)

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

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

    if etiq == "efectores":
        df=AAC_efec

    if etiq == "no_efectores":
        df=AAC_no_efec

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

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

```
efectores
Composición de aminoácidos (AAC) efectores archaea dataset 5, con valores
atípicos.
Valores del documento csv.

         X0      X1      X2      X3     X4      X5     X6      X7     X8  \
0     7.317   6.098   3.659   9.756  1.220   9.756  2.439   6.098  3.659
1     7.292   4.688   6.250   7.292  0.521   6.771  2.604   5.208  1.562
2     9.244  10.084   3.361  11.765  0.000  15.126  4.202   3.361  1.681
3     6.711   0.671   6.040   4.027  0.671  11.409  1.342   0.671  2.013
4     4.054   6.081   2.027   9.459  0.000  16.892  2.703   7.432  0.676
..      …       …       …       …      …       …      …       …      …
995  10.370   8.889   0.741  10.741  0.000   7.037  1.481  10.741  0.741
996  17.105   3.947   0.987   2.961  0.000   2.632  0.658  10.197  2.632
997   7.977   9.402   4.274   7.123  0.000   9.972  5.413   3.989  1.709
998   6.825   2.671  10.979   2.671  0.593   2.374  2.374  13.947  1.484
999   9.360   8.374   3.941   5.419  0.493   7.882  2.956   6.897  0.493
```

```
         X9  …     X11    X12    X13    X14    X15     X16    X17    X18  \
0     4.878  …   3.659  2.439  0.000  3.659  3.659  12.195  1.220  1.220
1     7.292  …   8.854  2.604  4.688  1.562  6.250   7.292  0.521  3.646
2     5.042  …   3.361  1.681  2.521  0.840  4.202   5.882  1.681  3.361
3    12.752  …  16.107  3.356  3.356  2.685  4.027   4.698  0.671  5.369
4     3.378  …   3.378  2.703  4.730  7.432  8.784   4.054  1.351  1.351
..      …  …     …    …    …    …    …      …    …    …
995   2.593  …   1.111  0.741  1.481  4.074  5.556   8.148  0.370  2.222
996   2.632  …   0.329  0.658  5.592  5.263  3.947   7.566  2.303  2.303
997   6.268  …   5.698  3.419  4.274  1.994  4.558   6.268  1.425  2.849
998   7.122  …   3.264  1.484  4.154  3.858  9.199   8.902  0.297  4.451
999   4.926  …   1.478  1.478  1.478  3.941  8.374  10.345  0.985  1.478

        X19         X20
0    10.976   efectores
1     3.125   efectores
2     3.361   efectores
3     6.040   efectores
4     4.730   efectores
..      …         …
995  16.296   efectores
996  13.158   efectores
997   4.843   efectores
998   7.418   efectores
999   9.360   efectores

[1000 rows x 21 columns]
```

Composición de aminoácidos (AAC) efectores archaea dataset 5, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      9.312007     6.041979     2.736560     5.853095     0.668347
std       4.416351     2.679709     2.112026     2.634872     0.929032
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       5.960000     4.216500     1.176000     3.814250     0.000000
50%       8.861000     5.797000     2.258000     5.435000     0.317500
75%      12.191250     7.696500     3.864000     7.468000     1.031000
max      27.536000    16.667000    11.786000    16.170000     6.504000

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      7.246533     2.408505     7.470858     1.779710     5.736234
std       3.698923     1.717751     2.893738     1.200427     3.361682
```

min	0.000000	0.000000	0.671000	0.000000	0.000000
25%	4.141500	1.205000	5.221750	0.824500	3.122500
50%	7.317000	2.130000	7.393000	1.639000	5.078500
75%	9.911750	3.164500	9.325500	2.469000	7.741000
max	18.898000	14.576000	20.849000	7.143000	21.154000

	X10	X11	X12	X13	X14 \
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	11.489897	4.527528	2.001903	3.909896	4.107703
std	3.256529	4.364804	1.239011	1.975062	1.745631
min	2.941000	0.000000	0.000000	0.000000	0.000000
25%	9.104500	1.064000	1.089000	2.532000	2.985000
50%	11.450500	2.540000	1.697000	3.779000	3.931500
75%	13.783250	7.867250	2.607500	5.002250	5.110750
max	23.077000	20.833000	10.145000	15.217000	10.345000

	X15	X16	X17	X18	X19
count	1000.000000	1000.000000	1000.000000	1000.00000	1000.000000
mean	5.866871	5.652249	1.227529	3.34744	8.615070
std	2.049265	2.176315	1.002127	1.62138	3.529489
min	0.000000	0.000000	0.000000	0.00000	0.714000
25%	4.483000	4.136500	0.574500	2.24275	5.704250
50%	5.651000	5.549000	1.071000	3.18000	8.169500
75%	7.143000	6.936000	1.685500	4.23350	11.224750
max	12.791000	14.365000	6.486000	10.34500	19.288000

no_efectores
Composición de aminoácidos (AAC) no_efectores archaea dataset 5, con valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	X7	X8 \
0	8.621	8.621	6.897	6.897	1.724	12.069	0.000	8.621	6.897
1	9.859	4.225	0.000	15.493	0.000	14.085	4.225	8.451	0.000
2	13.383	3.717	1.115	6.320	0.000	8.178	2.602	8.550	1.487
3	17.361	3.125	2.083	2.431	0.694	3.125	0.694	12.847	3.472
4	3.846	0.000	3.846	1.923	0.000	9.615	0.000	0.000	0.000
..	…	…	…	…	…	…	…	…	…
995	10.680	4.369	2.427	5.583	1.699	8.252	3.641	10.922	0.971
996	12.088	17.582	3.297	4.396	5.495	10.989	3.297	3.297	3.297
997	6.322	2.874	1.724	4.598	0.575	3.448	1.149	5.747	1.149
998	18.121	14.094	0.000	10.738	0.671	5.369	0.671	13.423	0.671
999	10.191	8.280	1.911	7.006	0.637	5.732	2.548	8.280	1.911

	X9	…	X11	X12	X13	X14	X15	X16	X17	X18 \
0	5.172	…	1.724	3.448	3.448	3.448	6.897	0.000	5.172	1.724

```
1     1.408  …   1.408  1.408  2.817  5.634  12.676   5.634  0.000  0.000
2     3.346  …   0.743  1.487  1.115  8.178   7.807  11.524  0.743  1.859
3     5.208  …   1.042  2.778  2.778  3.125   7.639   5.556  1.389  1.736
4     1.923  …  30.769  3.846  1.923  1.923  25.000   0.000  0.000  0.000
..      …  …     …    …    …    …      …      …    …      …
995   8.495  …   4.369  2.427  2.913  1.699   5.825   6.553  0.243  2.670
996   2.198  …   3.297  1.099  2.198  3.297   4.396   2.198  0.000  3.297
997  12.644  …   4.598  5.172  8.046  4.023   6.897   3.448  0.575  4.023
998   2.013  …   1.342  1.342  3.356  3.356   4.698   4.698  0.671  0.000
999   5.732  …   7.643  1.274  4.459  2.548   7.006   5.732  0.000  0.637

        X19            X20
0     3.448   no_efectores
1     8.451   no_efectores
2     8.178   no_efectores
3    10.764   no_efectores
4     5.769   no_efectores
..      …            …
995   8.010   no_efectores
996   5.495   no_efectores
997   9.195   no_efectores
998   7.383   no_efectores
999   8.280   no_efectores

[1000 rows x 21 columns]
```

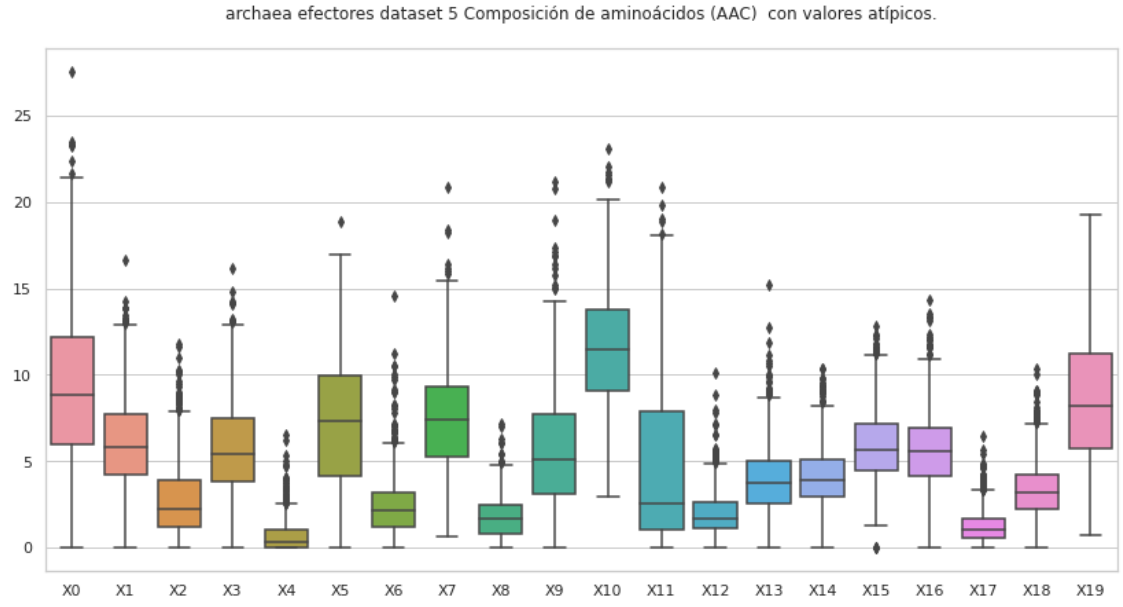
Composición de aminoácidos (AAC) no_efectores archaea dataset 5, con valores atípicos.
Estadísticas.

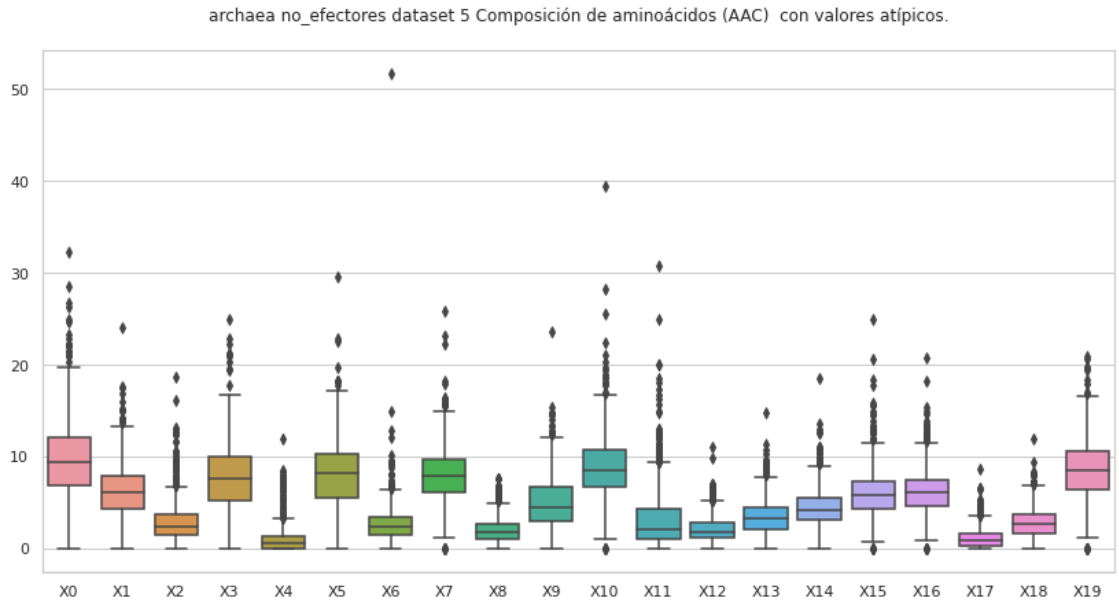
```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      9.750585     6.162593     2.888580     7.769388     1.002727
std       4.199130     2.842553     2.202007     3.587238     1.394213
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       6.892250     4.269250     1.523750     5.263000     0.000000
50%       9.363000     6.099000     2.439000     7.692000     0.629500
75%      12.163500     7.955000     3.660500    10.027500     1.317000
max      32.258000    24.074000    18.667000    25.000000    12.000000

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      8.046567     2.641187     7.962188     1.927219     4.988629
std       3.662728     2.303964     2.979715     1.315367     2.916511
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       5.556000     1.482500     6.061000     0.997500     2.934500
50%       8.170500     2.343000     7.880500     1.778000     4.464000
```

75%	10.306000	3.473500	9.732000	2.633250	6.667000
max	29.630000	51.667000	25.899000	7.692000	23.636000

	X10	X11	X12	X13	X14 \
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	8.939702	3.224681	2.138326	3.398581	4.376465
std	3.465950	3.321525	1.315798	1.898920	2.008141
min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	6.722000	1.013750	1.156000	2.125750	3.194750
50%	8.499500	2.156500	1.852000	3.215500	4.212500
75%	10.714000	4.367500	2.778750	4.478000	5.498750
max	39.437000	30.769000	11.111000	14.815000	18.497000

	X15	X16	X17	X18	X19
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	6.033865	6.260608	1.116002	2.818352	8.553728
std	2.560443	2.419679	1.089181	1.622280	3.118689
min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	4.389750	4.700750	0.285000	1.695000	6.448500
50%	5.846000	6.066000	0.915000	2.698500	8.444500
75%	7.289750	7.449250	1.617000	3.785000	10.542500
max	25.000000	20.765000	8.696000	12.000000	20.896000





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 5, sin valores
atípicos.
Valores del documento csv.

         X0      X1     X2      X3     X4      X5     X6      X7     X8  \
1     7.292   4.688  6.250   7.292  0.521   6.771  2.604   5.208  1.562
2     9.244  10.084  3.361  11.765  0.000  15.126  4.202   3.361  1.681
3     6.711   0.671  6.040   4.027  0.671  11.409  1.342   0.671  2.013
4     4.054   6.081  2.027   9.459  0.000  16.892  2.703   7.432  0.676
5    18.953   2.494  1.496   3.990  0.000   0.499  0.998  12.718  1.247
..      …     …    …     …     …     …    …     …    …
994   5.517   4.483  5.517   3.793  1.379   7.241  3.448   3.103  1.379
995  10.370   8.889  0.741  10.741  0.000   7.037  1.481  10.741  0.741
996  17.105   3.947  0.987   2.961  0.000   2.632  0.658  10.197  2.632
997   7.977   9.402  4.274   7.123  0.000   9.972  5.413   3.989  1.709
999   9.360   8.374  3.941   5.419  0.493   7.882  2.956   6.897  0.493

         X9  …     X11    X12    X13    X14    X15     X16    X17    X18  \
1     7.292  …   8.854  2.604  4.688  1.562  6.250   7.292  0.521  3.646
2     5.042  …   3.361  1.681  2.521  0.840  4.202   5.882  1.681  3.361
3    12.752  …  16.107  3.356  3.356  2.685  4.027   4.698  0.671  5.369
4     3.378  …   3.378  2.703  4.730  7.432  8.784   4.054  1.351  1.351
5     2.494  …   0.998  0.998  3.242  3.491  4.738   7.481  3.242  2.244
..      …  …     …    …    …    …    …     …    …    …
994   8.966  …  10.345  2.069  7.931  6.897  4.483   3.103  0.000  5.172
```

```
995   2.593  …   1.111  0.741  1.481  4.074  5.556   8.148  0.370  2.222
996   2.632  …   0.329  0.658  5.592  5.263  3.947   7.566  2.303  2.303
997   6.268  …   5.698  3.419  4.274  1.994  4.558   6.268  1.425  2.849
999   4.926  …   1.478  1.478  1.478  3.941  8.374  10.345  0.985  1.478

        X19        X20
1     3.125  efectores
2     3.361  efectores
3     6.040  efectores
4     4.730  efectores
5    13.466  efectores
..      …        …
994   3.448  efectores
995  16.296  efectores
996  13.158  efectores
997   4.843  efectores
999   9.360  efectores

[870 rows x 21 columns]
```

Composición de aminoácidos (AAC) efectores archaea dataset 5, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  870.000000  870.000000  870.000000  870.000000  870.000000  870.000000
mean     9.542966    6.196080    2.643178    5.959401    0.582132    7.286009
std      4.281846    2.618873    1.974030    2.583928    0.763354    3.597374
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      6.166250    4.318000    1.156500    3.937500    0.000000    4.230250
50%      9.295000    5.973500    2.199000    5.595000    0.299500    7.350500
75%     12.425250    7.910750    3.797000    7.639000    0.919500    9.873500
max     22.404000   13.889000    8.966000   13.194000    3.306000   16.923000

                X6          X7          X8          X9         X10         X11  \
count  870.000000  870.000000  870.000000  870.000000  870.000000  870.000000
mean     2.351668    7.546300    1.761094    5.558889   11.521801    4.278991
std      1.475652    2.777616    1.150979    3.118791    3.087108    4.171317
min      0.000000    0.671000    0.000000    0.000000    2.941000    0.000000
25%      1.268000    5.406750    0.820000    3.065500    9.167750    1.019250
50%      2.137000    7.509500    1.620500    5.000000   11.558500    2.412500
75%      3.134500    9.386250    2.446000    7.600250   13.768000    7.483000
max      7.143000   15.972000    5.046000   15.756000   21.171000   16.923000

               X12         X13         X14         X15         X16         X17  \
count  870.000000  870.000000  870.000000  870.000000  870.000000  870.000000
mean     1.920090    3.883044    4.079962    5.890909    5.695925    1.200649
```

std	1.104324	1.768507	1.669232	1.962721	2.031542	0.926863
min	0.000000	0.000000	0.000000	1.250000	1.000000	0.000000
25%	1.071000	2.689500	3.030000	4.525750	4.242500	0.573500
50%	1.658000	3.831000	3.910500	5.698500	5.599000	1.079000
75%	2.530250	5.000000	5.059000	7.143000	6.936000	1.686500
max	5.696000	9.655000	9.302000	11.765000	12.048000	4.225000

	X18	X19
count	870.000000	870.000000
mean	3.323430	8.777380
std	1.490156	3.535062
min	0.000000	1.418000
25%	2.274500	5.844500
50%	3.183500	8.275500
75%	4.205000	11.416750
max	7.976000	18.992000

no_efectores
Composición de aminoácidos (AAC) no_efectores archaea dataset 5, sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	X7	X8	\
1	9.859	4.225	0.000	15.493	0.000	14.085	4.225	8.451	0.000	
2	13.383	3.717	1.115	6.320	0.000	8.178	2.602	8.550	1.487	
3	17.361	3.125	2.083	2.431	0.694	3.125	0.694	12.847	3.472	
5	8.225	7.792	3.896	9.091	0.866	12.554	5.195	3.896	2.165	
6	4.933	1.794	8.520	7.175	0.897	1.794	3.363	6.278	1.345	
..	…	…	…	…	…	…	…	…	…	
994	5.914	0.538	8.602	5.914	0.000	1.613	3.763	5.914	2.151	
995	10.680	4.369	2.427	5.583	1.699	8.252	3.641	10.922	0.971	
997	6.322	2.874	1.724	4.598	0.575	3.448	1.149	5.747	1.149	
998	18.121	14.094	0.000	10.738	0.671	5.369	0.671	13.423	0.671	
999	10.191	8.280	1.911	7.006	0.637	5.732	2.548	8.280	1.911	

	X9	…	X11	X12	X13	X14	X15	X16	X17	X18	\
1	1.408	…	1.408	1.408	2.817	5.634	12.676	5.634	0.000	0.000	
2	3.346	…	0.743	1.487	1.115	8.178	7.807	11.524	0.743	1.859	
3	5.208	…	1.042	2.778	2.778	3.125	7.639	5.556	1.389	1.736	
5	6.926	…	3.463	0.433	5.195	1.732	6.494	3.896	1.299	3.463	
6	9.417	…	5.157	1.121	6.054	6.054	10.314	7.623	1.345	4.933	
..	…	…	…	…	…	…	…	…	…	…	
994	6.452	…	8.602	2.151	3.763	3.226	4.301	11.290	0.538	6.452	
995	8.495	…	4.369	2.427	2.913	1.699	5.825	6.553	0.243	2.670	
997	12.644	…	4.598	5.172	8.046	4.023	6.897	3.448	0.575	4.023	
998	2.013	…	1.342	1.342	3.356	3.356	4.698	4.698	0.671	0.000	

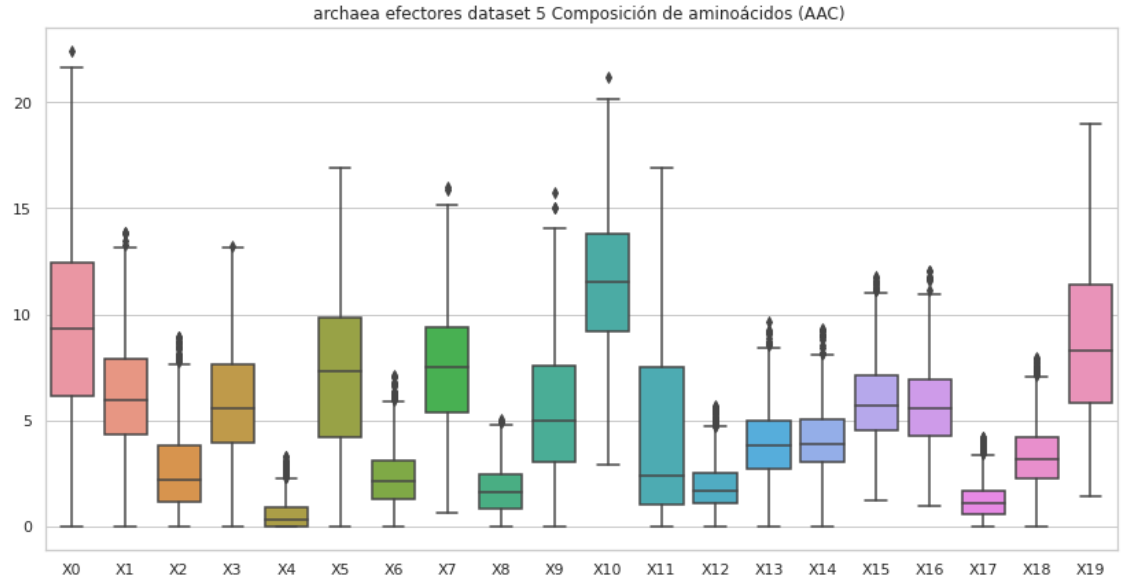
```
999   5.732  …  7.643  1.274  4.459  2.548   7.006   5.732  0.000  0.637

        X19           X20
1     8.451  no_efectores
2     8.178  no_efectores
3    10.764  no_efectores
5     6.494  no_efectores
6     5.157  no_efectores
..      …           …
994  11.828  no_efectores
995   8.010  no_efectores
997   9.195  no_efectores
998   7.383  no_efectores
999   8.280  no_efectores

[841 rows x 21 columns]
```

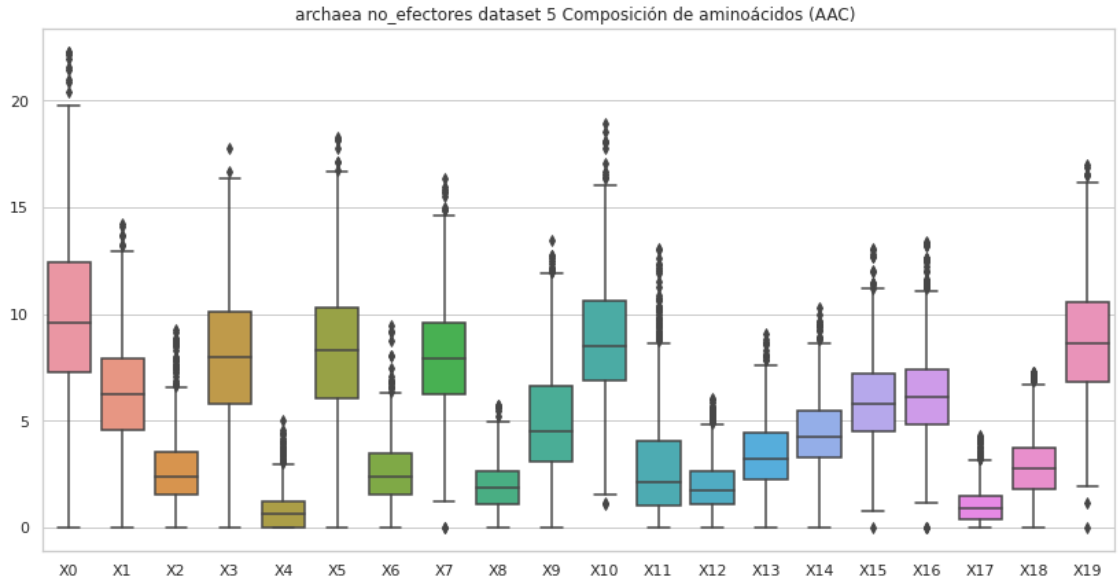
Composición de aminoácidos (AAC) no_efectores archaea dataset 5, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  841.000000  841.000000  841.000000  841.000000  841.000000  841.000000
mean     9.956361    6.266677    2.715163    7.969535    0.825401    8.212207
std      3.842093    2.469402    1.723922    3.160031    0.899128    3.324728
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      7.256000    4.563000    1.533000    5.800000    0.000000    6.024000
50%      9.598000    6.218000    2.410000    8.000000    0.619000    8.333000
75%     12.424000    7.907000    3.553000   10.101000    1.205000   10.309000
max     22.271000   14.208000    9.272000   17.757000    5.000000   18.269000

                X6          X7          X8          X9         X10         X11  \
count  841.000000  841.000000  841.000000  841.000000  841.000000  841.000000
mean     2.601707    7.953653    1.921612    4.986669    8.951590    3.031756
std      1.507680    2.602888    1.145896    2.737824    2.946046    2.768601
min      0.000000    0.000000    0.000000    0.000000    1.064000    0.000000
25%      1.575000    6.250000    1.111000    3.061000    6.857000    1.031000
50%      2.370000    7.937000    1.843000    4.483000    8.516000    2.137000
75%      3.473000    9.622000    2.640000    6.626000   10.625000    4.065000
max      9.489000   16.327000    5.714000   13.483000   18.957000   13.043000

               X12         X13         X14         X15         X16         X17  \
count  841.000000  841.000000  841.000000  841.000000  841.000000  841.000000
mean     2.013615    3.365057    4.384957    5.901845    6.276256    1.040014
std      1.159582    1.681526    1.730277    2.112556    2.146033    0.886429
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      1.111000    2.239000    3.255000    4.500000    4.825000    0.382000
```

50%	1.770000	3.192000	4.268000	5.817000	6.107000	0.909000
75%	2.613000	4.459000	5.441000	7.180000	7.379000	1.508000
max	6.024000	9.091000	10.294000	13.043000	13.415000	4.286000

	X18	X19
count	841.000000	841.000000
mean	2.834776	8.791106
std	1.452296	2.758686
min	0.000000	0.000000
25%	1.786000	6.813000
50%	2.743000	8.613000
75%	3.759000	10.584000
max	7.303000	16.972000





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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.027290	0.004548	0.036387	0.036387	0.000000	0.022742	0.013645	
1	0.056916	0.004065	0.056916	0.052851	0.036589	0.040655	0.012196	
2	0.025763	0.000000	0.032790	0.042158	0.007026	0.009369	0.004684	
3	0.059355	0.005935	0.035613	0.100903	0.029677	0.005935	0.017806	
4	0.013629	0.000000	0.031800	0.056786	0.015900	0.024986	0.002271	
..	…	…	…	…	…	…	…	
995	0.021109	0.000000	0.021863	0.014324	0.003016	0.021863	0.001508	
996	0.027533	0.000000	0.004765	0.004236	0.009001	0.016414	0.004236	
997	0.046357	0.000000	0.041390	0.057946	0.024834	0.023179	0.009934	
998	0.019553	0.001700	0.007651	0.006801	0.011902	0.039957	0.004251	
999	0.024890	0.001310	0.014410	0.020960	0.003930	0.018340	0.001310	

	X7	X8	X9	…	X74	X75	X76	\
0	0.018193	0.013645	0.022742	…	-0.014943	-0.026250	0.017608	
1	0.056916	0.069113	0.093506	…	0.027050	-0.030629	0.000718	
2	0.014053	0.009369	0.025763	…	-0.007124	0.016577	0.005957	
3	0.112774	0.142451	0.065290	…	0.029951	-0.060374	-0.010987	
4	0.011357	0.011357	0.029529	…	0.031165	0.020700	0.012999	
..	…	…	…	…	…	…	…	
995	0.005277	0.002262	0.013570	…	0.013898	0.007999	0.020556	
996	0.004236	0.000529	0.024356	…	0.019168	0.002848	0.023317	
997	0.036423	0.033112	0.049668	…	-0.034220	0.009621	0.014670	
998	0.020403	0.009352	0.017003	…	0.014004	0.007787	0.033230	
999	0.013100	0.003930	0.027510	…	-0.006126	-0.007352	0.021377	

	X77	X78	X79	X80	X81	X82	X83
0	-0.034386	0.009535	0.049062	-0.020638	0.015523	0.016076	efectores
1	-0.026662	0.040715	-0.021942	0.020715	0.034541	0.016432	efectores
2	0.035004	0.046248	-0.003680	-0.004405	0.015845	0.008546	efectores
3	-0.003037	0.037748	-0.014979	-0.045686	0.029542	-0.000725	efectores
4	0.037878	0.053133	-0.009540	-0.024387	0.002085	0.009429	efectores
..	…	…	…	…	…	…	
995	0.012179	0.010901	0.033801	0.009143	0.009126	0.022936	efectores
996	0.020602	0.005587	0.026168	0.022481	0.006246	0.018236	efectores

```
997  0.017065  0.039756 -0.014862  0.009615  0.020754  0.011234  efectores
998 -0.000165  0.000140  0.031012  0.000030 -0.007075  0.027714  efectores
999  0.004485  0.005598  0.018704 -0.010352  0.000533  0.011256  efectores

[1000 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro_mass efectores archaea dataset 5, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.033800	0.003945	0.026767	0.036716	0.016431	
std	0.032718	0.011388	0.027180	0.041036	0.016888	
min	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.021014	0.000000	0.009327	0.008444	0.007233	
50%	0.028944	0.000784	0.021312	0.028958	0.012471	
75%	0.041221	0.005001	0.038400	0.056092	0.020425	
max	0.918120	0.306040	0.612080	0.918120	0.306040	

	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.027569	0.007998	0.028202	0.026057	0.047277	…	
std	0.015903	0.007764	0.033286	0.033886	0.034615	…	
min	0.004412	0.000000	0.000000	0.000000	0.008018	…	
25%	0.017114	0.002413	0.007507	0.002113	0.024602	…	
50%	0.024194	0.006001	0.017428	0.009481	0.039495	…	
75%	0.034386	0.010932	0.039047	0.040546	0.059098	…	
max	0.306040	0.058794	0.612080	0.306040	0.612080	…	

	X73	X74	X75	X76	X77	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.013821	0.003479	0.006397	0.012009	0.007444	
std	0.021875	0.037685	0.091397	0.025689	0.047385	
min	-0.089895	-0.783721	-2.762704	-0.514404	-0.140770	
25%	0.003127	-0.007348	-0.000840	0.005467	-0.004835	
50%	0.014908	0.008360	0.005320	0.015222	0.009929	
75%	0.024318	0.017066	0.017867	0.023324	0.018676	
max	0.281181	0.168203	0.139084	0.086507	1.250066	

	X78	X79	X80	X81	X82
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.009984	0.014149	0.001886	0.008646	0.012179
std	0.027474	0.021809	0.074972	0.053985	0.065337
min	-0.127898	-0.156858	-2.181570	-1.507890	-1.946656
25%	-0.000466	0.004852	-0.005959	-0.000604	0.005274
50%	0.005701	0.014997	0.009746	0.005379	0.014533

```
75%       0.018173     0.025575     0.017522     0.017778     0.023822
max       0.486408     0.148511     0.167197     0.161151     0.120391

[8 rows x 83 columns]
```

```
no_efectores
Composición de pseudo aminoácidos (PseAAC)  hidro_mass no_efectores archaea
dataset 5, con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.030508  0.006102  0.024406  0.042711  0.012203  0.030508  0.024406
1    0.018225  0.000000  0.028639  0.026036  0.005207  0.015621  0.000000
2    0.026853  0.000000  0.012680  0.016410  0.002238  0.017156  0.002984
3    0.025023  0.001001  0.003503  0.004504  0.004004  0.018517  0.005005
4    0.006878  0.000000  0.003439  0.017196  0.003439  0.000000  0.000000
..        …         …         …         …         …         …
995  0.035731  0.005685  0.018678  0.027611  0.009745  0.036543  0.003248
996  0.034788  0.015813  0.012650  0.031626  0.006325  0.009488  0.009488
997  0.012521  0.001138  0.009106  0.006830  0.015936  0.011383  0.002277
998  0.030860  0.001143  0.018287  0.009144  0.005715  0.022859  0.001143
999  0.040113  0.002507  0.027578  0.022564  0.017550  0.032592  0.007521

           X7        X8        X9  …       X74       X75       X76  \
0    0.018305  0.006102  0.018305  …  0.013823 -0.018567  0.040794
1    0.002604  0.002604  0.007811  …  0.013881  0.048936  0.012713
2    0.006713  0.001492  0.019394  … -0.001339 -0.003624  0.021730
3    0.007507  0.001501  0.017516  …  0.004436 -0.002780  0.020438
4    0.003439  0.055027  0.017196  … -0.028785  0.002300  0.007787
..        …         …         …  …         …         …         …
995  0.028423  0.014617  0.027611  …  0.002385 -0.008907  0.017248
996  0.006325  0.009488  0.025301  …  0.078370  0.072420  0.032827
997  0.025043  0.009106  0.027319  …  0.038347  0.012825  0.009043
998  0.003429  0.002286  0.012572  … -0.004638  0.004289  0.027101
999  0.022564  0.030085  0.040113  …  0.006961  0.038654  0.000896

           X77       X78       X79       X80       X81       X82           X83
0   -0.039452 -0.001849 -0.019358  0.014971  0.019288 -0.071250  no_efectores
1   -0.017937  0.023962  0.016995 -0.005157  0.024025  0.012616  no_efectores
2   -0.001595  0.003841  0.029239  0.007966  0.003715  0.039946  no_efectores
3    0.006639 -0.000779  0.029686  0.006912 -0.004055  0.018277  no_efectores
4    0.042816  0.051754  0.007866 -0.000817  0.018860 -0.001380  no_efectores
..        …         …         …         …         …         …
995 -0.006695 -0.007834  0.032909  0.015269  0.009804  0.036385  no_efectores
996  0.004738 -0.009441  0.021647 -0.018861  0.000238 -0.003807  no_efectores
997  0.006448 -0.003581  0.008323  0.028585  0.004932  0.005196  no_efectores
```

```
998 -0.005966  0.004920  0.024332  0.008558  0.014933  0.018345  no_efectores
999 -0.001554  0.024415  0.022771 -0.015484 -0.012346  0.027167  no_efectores

[1000 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores archaea
dataset 5, con valores atípicos.
Estadísticas.

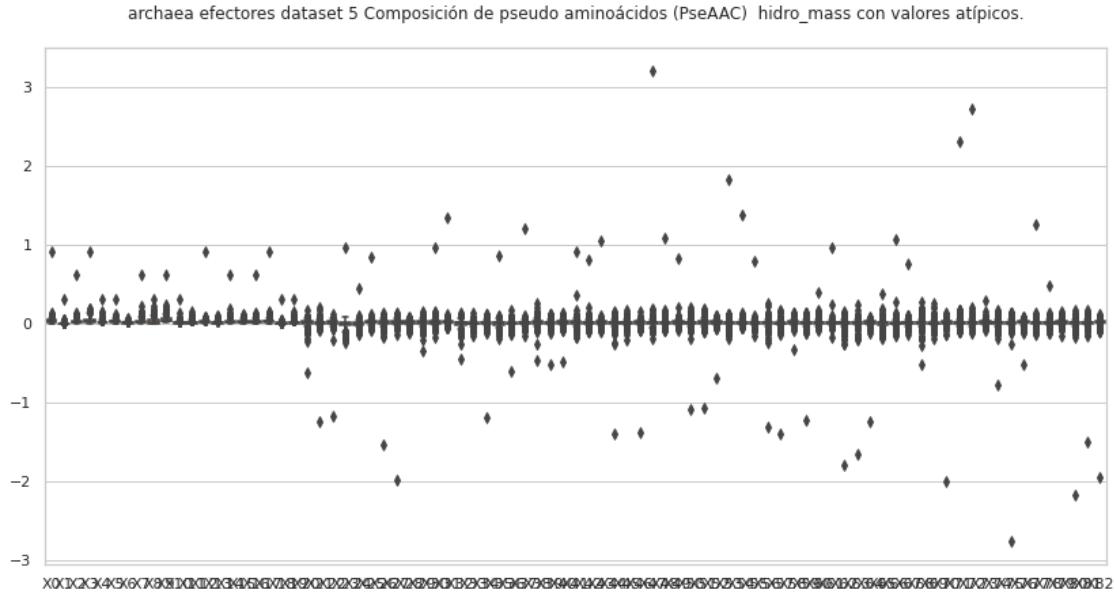
	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.034811	0.004888	0.027220	0.015505	0.014979	
std	0.019846	0.012519	0.145472	0.571031	0.020515	
min	0.000000	0.000000	-4.502181	-18.008723	0.000000	
25%	0.023790	0.000000	0.018032	0.016883	0.006323	
50%	0.032043	0.002100	0.029352	0.029893	0.010777	
75%	0.042274	0.005128	0.041532	0.045087	0.017535	
max	0.355711	0.268097	0.355711	0.188858	0.283287	

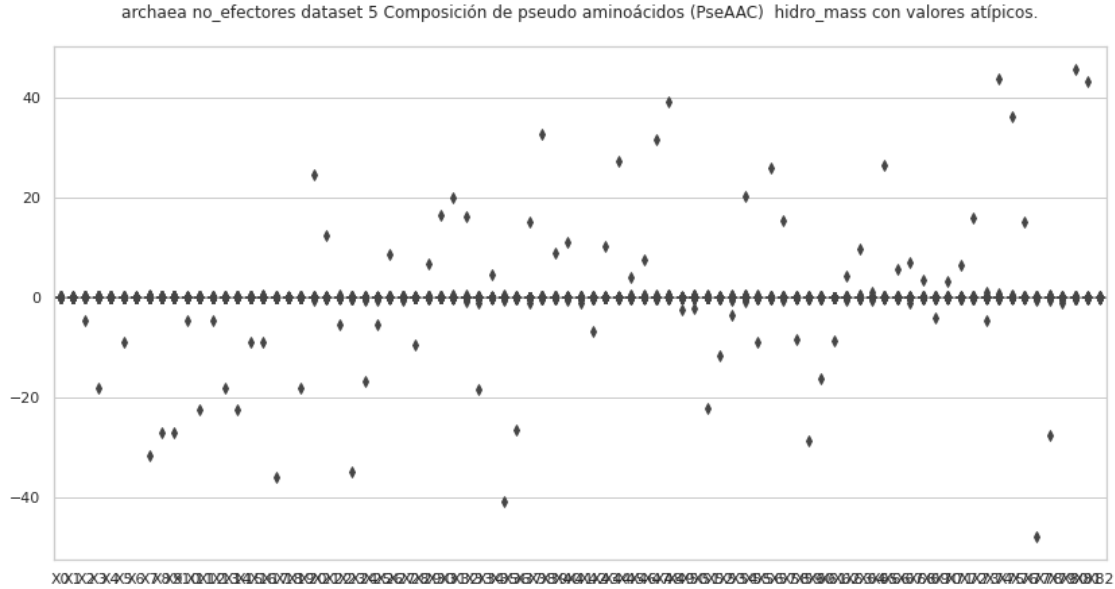
	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.020472	0.009038	-0.009026	-0.010399	0.008582	…	
std	0.286235	0.015605	0.997772	0.855273	0.855841	…	
min	-9.004362	0.000000	-31.515266	-27.013085	-27.013085	…	
25%	0.019768	0.002629	0.008044	0.002791	0.020188	…	
50%	0.027065	0.006417	0.014761	0.007366	0.029619	…	
75%	0.035143	0.011215	0.027126	0.018618	0.041844	…	
max	0.268097	0.355711	0.434307	0.355711	0.349002	…	

	X73	X74	X75	X76	X77	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.013344	0.046568	0.044303	0.031172	-0.045427	
std	0.148642	1.385950	1.144046	0.476116	1.513079	
min	-4.510307	-0.403077	-0.188510	-0.324099	-47.834593	
25%	0.007341	-0.007765	-0.001498	0.007559	-0.006663	
50%	0.018754	0.003912	0.006079	0.018358	0.004320	
75%	0.027903	0.012984	0.017384	0.026489	0.013924	
max	0.993381	43.813342	36.175058	15.052938	0.222934	

	X78	X79	X80	X81	X82
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	-0.018820	0.013783	0.047598	0.050939	0.015915
std	0.873303	0.044493	1.442180	1.361579	0.024841
min	-27.584578	-0.980599	-0.224493	-0.267336	-0.208297
25%	-0.002021	0.007397	-0.006539	-0.001424	0.006963
50%	0.006577	0.017974	0.003899	0.006154	0.018002
75%	0.017856	0.026577	0.014042	0.018739	0.027274

```
max        0.461703      0.173255      45.597625      43.055876      0.231838

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.027290	0.004548	0.036387	0.036387	0.000000	0.022742	0.013645	
2	0.025763	0.000000	0.032790	0.042158	0.007026	0.009369	0.004684	
4	0.013629	0.000000	0.031800	0.056786	0.015900	0.024986	0.002271	
5	0.022110	0.000000	0.004655	0.000582	0.003782	0.014837	0.001455	
6	0.025098	0.012549	0.012549	0.041830	0.004183	0.020915	0.004183	
..	…	…	…	…	…	…	…	
995	0.021109	0.000000	0.021863	0.014324	0.003016	0.021863	0.001508	
996	0.027533	0.000000	0.004765	0.004236	0.009001	0.016414	0.004236	
997	0.046357	0.000000	0.041390	0.057946	0.024834	0.023179	0.009934	
998	0.019553	0.001700	0.007651	0.006801	0.011902	0.039957	0.004251	
999	0.024890	0.001310	0.014410	0.020960	0.003930	0.018340	0.001310	

	X7	X8	X9	…	X74	X75	X76	\
0	0.018193	0.013645	0.022742	…	-0.014943	-0.026250	0.017608	
2	0.014053	0.009369	0.025763	…	-0.007124	0.016577	0.005957	
4	0.011357	0.011357	0.029529	…	0.031165	0.020700	0.012999	
5	0.002909	0.001164	0.017746	…	0.010242	-0.000496	0.019631	
6	0.025098	0.046013	0.020915	…	0.008640	0.039496	0.024144	
..	…	…	…	…	…	…	…	
995	0.005277	0.002262	0.013570	…	0.013898	0.007999	0.020556	
996	0.004236	0.000529	0.024356	…	0.019168	0.002848	0.023317	
997	0.036423	0.033112	0.049668	…	-0.034220	0.009621	0.014670	
998	0.020403	0.009352	0.017003	…	0.014004	0.007787	0.033230	
999	0.013100	0.003930	0.027510	…	-0.006126	-0.007352	0.021377	

	X77	X78	X79	X80	X81	X82	X83
0	-0.034386	0.009535	0.049062	-0.020638	0.015523	0.016076	efectores
2	0.035004	0.046248	-0.003680	-0.004405	0.015845	0.008546	efectores
4	0.037878	0.053133	-0.009540	-0.024387	0.002085	0.009429	efectores
5	0.018570	0.002685	0.020774	0.014451	-0.000988	0.021790	efectores
6	0.009486	0.035739	0.004931	-0.028437	0.016443	0.006054	efectores
..	…	…	…	…	…	…	
995	0.012179	0.010901	0.033801	0.009143	0.009126	0.022936	efectores
996	0.020602	0.005587	0.026168	0.022481	0.006246	0.018236	efectores
997	0.017065	0.039756	-0.014862	0.009615	0.020754	0.011234	efectores
998	-0.000165	0.000140	0.031012	0.000030	-0.007075	0.027714	efectores
999	0.004485	0.005598	0.018704	-0.010352	0.000533	0.011256	efectores

[877 rows x 84 columns]

Composición de pseudo aminoácidos (PseAAC) hidro_mass efectores archaea dataset 5, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	877.000000	877.000000	877.000000	877.000000	877.000000	877.000000	
mean	0.030691	0.002655	0.022705	0.029608	0.014032	0.025207	
std	0.013820	0.004348	0.016548	0.024557	0.010241	0.010814	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.004412	
25%	0.020760	0.000000	0.008469	0.007377	0.006909	0.016589	
50%	0.027405	0.000537	0.018221	0.022103	0.011413	0.022825	
75%	0.038170	0.003739	0.034741	0.050197	0.018478	0.032252	
max	0.089359	0.031396	0.075086	0.119363	0.066857	0.072398	

	X6	X7	X8	X9	…	X73	\
count	877.000000	877.000000	877.000000	877.000000	…	877.000000	
mean	0.006977	0.022280	0.019635	0.040305	…	0.013668	
std	0.006196	0.021076	0.025272	0.021167	…	0.016249	
min	0.000000	0.000000	0.000000	0.008018	…	-0.045499	
25%	0.002296	0.006941	0.001797	0.023111	…	0.003766	
50%	0.005535	0.013988	0.007160	0.036232	…	0.014980	
75%	0.009613	0.030913	0.030029	0.052013	…	0.023776	
max	0.031270	0.108333	0.126752	0.131400	…	0.068823	

	X74	X75	X76	X77	X78	X79	\
count	877.000000	877.000000	877.000000	877.000000	877.000000	877.000000	
mean	0.005681	0.009103	0.014449	0.007689	0.009200	0.015298	
std	0.020687	0.018838	0.014648	0.020093	0.018059	0.015537	
min	-0.107128	-0.064835	-0.043380	-0.090460	-0.061222	-0.040342	
25%	-0.005563	-0.000496	0.007188	-0.002482	-0.000142	0.006454	
50%	0.009141	0.005092	0.015974	0.010433	0.005649	0.015541	
75%	0.016714	0.015138	0.023346	0.018472	0.015769	0.025229	
max	0.109995	0.133624	0.069742	0.086129	0.073845	0.065361	

	X80	X81	X82
count	877.000000	877.000000	877.000000
mean	0.005534	0.008597	0.015254
std	0.022501	0.018753	0.016631
min	-0.114858	-0.067038	-0.075795
25%	-0.003791	-0.000471	0.006448
50%	0.010095	0.004781	0.015353
75%	0.017024	0.015845	0.023721
max	0.113800	0.097922	0.100564

[8 rows x 83 columns]

```
no_efectores
Composición de pseudo aminoácidos (PseAAC)  hidro_mass no_efectores archaea
dataset 5, sin valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
1    0.018225  0.000000  0.028639  0.026036  0.005207  0.015621  0.000000
2    0.026853  0.000000  0.012680  0.016410  0.002238  0.017156  0.002984
3    0.025023  0.001001  0.003503  0.004504  0.004004  0.018517  0.005005
4    0.006878  0.000000  0.003439  0.017196  0.003439  0.000000  0.000000
5    0.049858  0.005248  0.055106  0.076099  0.031489  0.023617  0.013120
..        …         …         …         …         …         …         …
995  0.035731  0.005685  0.018678  0.027611  0.009745  0.036543  0.003248
996  0.034788  0.015813  0.012650  0.031626  0.006325  0.009488  0.009488
997  0.012521  0.001138  0.009106  0.006830  0.015936  0.011383  0.002277
998  0.030860  0.001143  0.018287  0.009144  0.005715  0.022859  0.001143
999  0.040113  0.002507  0.027578  0.022564  0.017550  0.032592  0.007521

           X7        X8        X9  …       X74       X75       X76  \
1    0.002604  0.002604  0.007811  …  0.013881  0.048936  0.012713
2    0.006713  0.001492  0.019394  … -0.001339 -0.003624  0.021730
3    0.007507  0.001501  0.017516  …  0.004436 -0.002780  0.020438
4    0.003439  0.055027  0.017196  … -0.028785  0.002300  0.007787
5    0.041986  0.020993  0.041986  …  0.050327  0.043812  0.015552
..        …         …         …  …         …         …         …
995  0.028423  0.014617  0.027611  …  0.002385 -0.008907  0.017248
996  0.006325  0.009488  0.025301  …  0.078370  0.072420  0.032827
997  0.025043  0.009106  0.027319  …  0.038347  0.012825  0.009043
998  0.003429  0.002286  0.012572  … -0.004638  0.004289  0.027101
999  0.022564  0.030085  0.040113  …  0.006961  0.038654  0.000896

           X77       X78       X79       X80       X81       X82           X83
1   -0.017937  0.023962  0.016995 -0.005157  0.024025  0.012616  no_efectores
2   -0.001595  0.003841  0.029239  0.007966  0.003715  0.039946  no_efectores
3    0.006639 -0.000779  0.029686  0.006912 -0.004055  0.018277  no_efectores
4    0.042816  0.051754  0.007866 -0.000817  0.018860 -0.001380  no_efectores
5   -0.026072  0.006386  0.001070 -0.046352 -0.025709  0.024871  no_efectores
..        …         …         …         …         …         …           …
995 -0.006695 -0.007834  0.032909  0.015269  0.009804  0.036385  no_efectores
996  0.004738 -0.009441  0.021647 -0.018861  0.000238 -0.003807  no_efectores
997  0.006448 -0.003581  0.008323  0.028585  0.004932  0.005196  no_efectores
998 -0.005966  0.004920  0.024332  0.008558  0.014933  0.018345  no_efectores
999 -0.001554  0.024415  0.022771 -0.015484 -0.012346  0.027167  no_efectores

[963 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores archaea
dataset 5, sin valores atípicos.
Estadísticas.

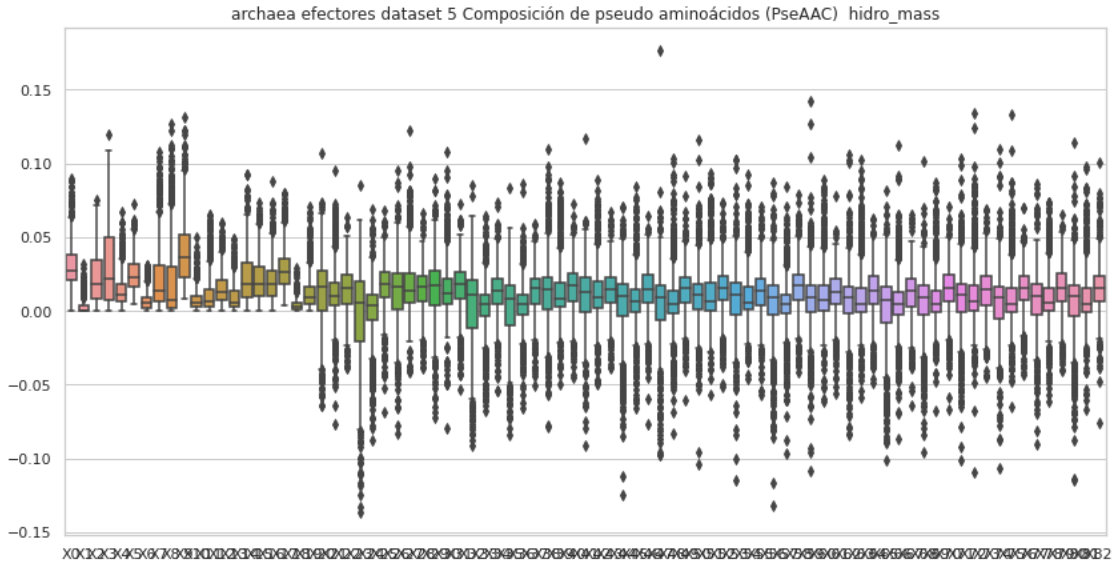
	X0	X1	X2	X3	X4	X5	\
count	963.000000	963.000000	963.000000	963.000000	963.000000	963.000000	
mean	0.033233	0.003928	0.029789	0.031693	0.012962	0.027981	
std	0.013397	0.005699	0.017291	0.020222	0.010081	0.012703	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.023689	0.000000	0.017859	0.016466	0.006301	0.019591	
50%	0.031909	0.002076	0.028794	0.029269	0.010572	0.026715	
75%	0.041497	0.004927	0.040720	0.044144	0.017181	0.034130	
max	0.081934	0.041687	0.103589	0.112928	0.071683	0.111765	

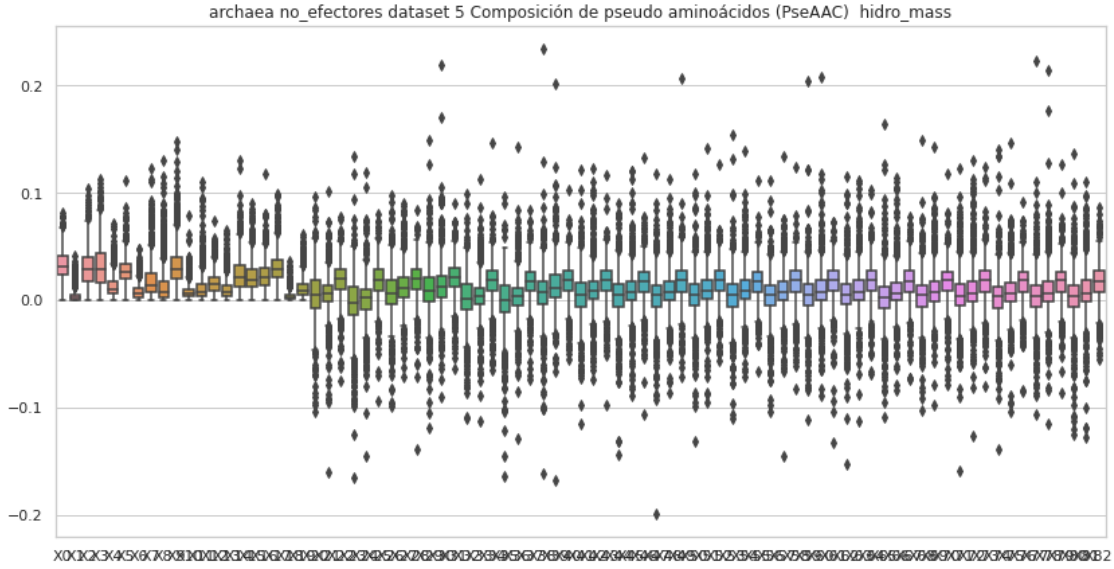
	X6	X7	X8	X9	…	X73	\
count	963.000000	963.000000	963.000000	963.000000	…	963.000000	
mean	0.007846	0.019593	0.013837	0.032836	…	0.017385	
std	0.007132	0.017164	0.017956	0.019265	…	0.018216	
min	0.000000	0.000000	0.000000	0.000000	…	-0.079593	
25%	0.002645	0.007894	0.002730	0.020078	…	0.007960	
50%	0.006293	0.014231	0.007129	0.028925	…	0.018774	
75%	0.010773	0.025548	0.017472	0.040232	…	0.027528	
max	0.047700	0.122397	0.130301	0.147675	…	0.124332	

	X74	X75	X76	X77	X78	X79	\
count	963.000000	963.000000	963.000000	963.000000	963.000000	963.000000	
mean	0.002352	0.008519	0.016518	0.003526	0.008709	0.016831	
std	0.021523	0.019876	0.016842	0.022236	0.021917	0.017415	
min	-0.138740	-0.103301	-0.084946	-0.105997	-0.115577	-0.085884	
25%	-0.007567	-0.001112	0.007851	-0.006096	-0.001701	0.008006	
50%	0.003827	0.006056	0.018376	0.004435	0.006554	0.018302	
75%	0.012815	0.016784	0.026229	0.013677	0.017099	0.026602	
max	0.140749	0.146240	0.091028	0.222934	0.213983	0.126401	

	X80	X81	X82
count	963.000000	963.000000	963.000000
mean	0.003251	0.008776	0.016858
std	0.022438	0.021638	0.017071
min	-0.124998	-0.127872	-0.054609
25%	-0.005918	-0.000974	0.007466
50%	0.004239	0.006269	0.018102
75%	0.013895	0.018602	0.027050
max	0.136391	0.110028	0.086540

[8 rows x 83 columns]





4 Composición de pseudo aminoácidos (PseAAC) mass

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

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

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

    if etiq == "efectores":
        df=PseAAC_mass_efec

    if etiq == "no_efectores":
        df=PseAAC_mass_no_efec

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

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

```
efectores
Composición de pseudo aminoácidos (PseAAC)  mass efectores archaea dataset 5,
con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.033111  0.005518  0.044148  0.044148  0.000000  0.027592  0.016555
1    0.056843  0.004060  0.056843  0.052783  0.036542  0.040602  0.012181
2    0.078136  0.000000  0.099446  0.127859  0.021310  0.028413  0.014207
3    0.073065  0.007306  0.043839  0.124210  0.036532  0.007306  0.021919
4    0.033241  0.000000  0.077563  0.138505  0.038781  0.060942  0.005540
..        …         …         …         …         …         …         …
995  0.027764  0.000000  0.028756  0.018840  0.003966  0.028756  0.001983
996  0.045736  0.000000  0.007916  0.007036  0.014952  0.027266  0.007036
997  0.070692  0.000000  0.063118  0.088366  0.037871  0.035346  0.015148
998  0.019257  0.001675  0.007535  0.006698  0.011722  0.039351  0.004186
999  0.038139  0.002007  0.022081  0.032117  0.006022  0.028103  0.002007

           X7        X8        X9  …       X32       X33       X34  \
```

```
0    0.022074  0.016555  0.027592  …  0.005305  0.008414 -0.006864
1    0.056843  0.069024  0.093386  … -0.046684  0.025830  0.004897
2    0.042620  0.028413  0.078136  …  0.021528  0.031021 -0.065949
3    0.138823  0.175356  0.080371  … -0.008420 -0.041214  0.000475
4    0.027701  0.027701  0.072022  …  0.054131 -0.032513 -0.018198
..        …         …        …  …         …         …
995  0.006941  0.002975  0.017848  …  0.031028  0.037475  0.042217
996  0.007036  0.000880  0.040459  …  0.034150  0.041883  0.025356
997  0.055544  0.050495  0.075742  … -0.008467 -0.023258  0.009328
998  0.020094  0.009210  0.016745  …  0.028033  0.030606  0.051957
999  0.020073  0.006022  0.042154  …  0.046583  0.016768  0.008048

          X35       X36       X37       X38       X39       X40        X41
0    0.050975  0.001327  0.034761  0.021363  0.059526  0.019504  efectores
1    0.011699  0.055134 -0.037160  0.000717 -0.021914  0.016411  efectores
2   -0.010133 -0.038199 -0.017153  0.018066 -0.011160  0.025919  efectores
3    0.007822 -0.021218  0.024503 -0.013525 -0.018439 -0.000892  efectores
4   -0.007510  0.028681  0.039624  0.031706 -0.023268  0.022998  efectores
..        …         …         …         …         …         …
995  0.041387  0.027386  0.029623  0.027037  0.044458  0.030167  efectores
996  0.035280  0.031634  0.047551  0.038734  0.043469  0.030293  efectores
997  0.031926 -0.007977 -0.030076  0.022371 -0.022663  0.017131  efectores
998  0.025676  0.040041  0.030529  0.032726  0.030542  0.027294  efectores
999  0.013077  0.009811  0.036206  0.032757  0.028660  0.017248  efectores

[1000 rows x 42 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass efectores archaea dataset 5, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.046897     0.004758     0.035846     0.049496     0.023321
std       0.017457     0.007686     0.023781     0.039738     0.017499
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       0.036169     0.000000     0.016984     0.015355     0.011678
50%       0.044312     0.001225     0.030284     0.038803     0.018589
75%       0.054528     0.006755     0.049850     0.074275     0.030853
max       0.152395     0.051076     0.121507     0.204206     0.162214

                X5           X6           X7           X8           X9  …  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
mean      0.038632     0.011024     0.038613     0.034892     0.066925  …
std       0.012173     0.009687     0.033516     0.041037     0.032757  …
min       0.007306     0.000000     0.000000     0.000000     0.008333  …
25%       0.029649     0.003814     0.012282     0.003688     0.042528  …
```

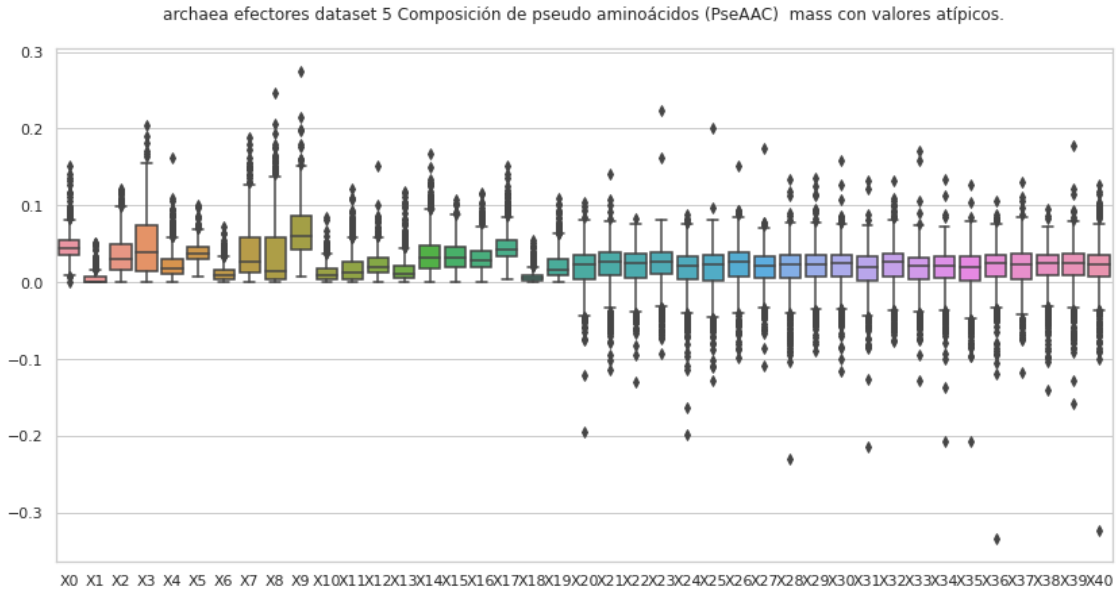
```
50%       0.037060     0.008877     0.026655     0.014759     0.059394  …
75%       0.045801     0.015619     0.058331     0.057612     0.086862  …
max       0.099911     0.072904     0.188560     0.246943     0.274417  …

               X31          X32          X33          X34          X35  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.015886     0.021976     0.016726     0.016587     0.015325
std       0.026601     0.026171     0.024966     0.026435     0.027879
min      -0.214630    -0.077707    -0.128447    -0.206910    -0.206144
25%       0.002313     0.008257     0.004401     0.005376     0.001627
50%       0.020371     0.026590     0.021412     0.021943     0.020385
75%       0.032977     0.037842     0.032685     0.033458     0.034000
max       0.132445     0.131495     0.170523     0.134169     0.126668

               X36          X37          X38          X39          X40
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.019572     0.019772     0.019137     0.020372     0.020236
std       0.027919     0.026688     0.025615     0.027731     0.029004
min      -0.333442    -0.117388    -0.140138    -0.158295    -0.323118
25%       0.007761     0.004705     0.008361     0.008343     0.007048
50%       0.025653     0.023098     0.024741     0.025100     0.023342
75%       0.035271     0.036467     0.034870     0.037284     0.036052
max       0.106143     0.130123     0.095723     0.178081     0.127481

[8 rows x 41 columns]



no_efectores
Composición de pseudo aminoácidos (PseAAC)  mass no_efectores archaea dataset 5,
con valores atípicos.
Valores del documento csv.

          X0        X1        X2        X3        X4        X5        X6  \
0    0.039085  0.007817  0.031268  0.054719  0.015634  0.039085  0.031268
1    0.032324  0.000000  0.050794  0.046177  0.009235  0.027706  0.000000
2    0.031257  0.000000  0.014760  0.019101  0.002605  0.019969  0.003473
3    0.034333  0.001373  0.004807  0.006180  0.005493  0.025407  0.006867
4    0.022808  0.000000  0.011404  0.057019  0.011404  0.000000  0.000000
..        …         …         …         …         …         …         …
995  0.037682  0.005995  0.019697  0.029118  0.010277  0.038538  0.003426
996  0.125367  0.056985  0.045588  0.113970  0.022794  0.034191  0.034191
997  0.041956  0.003814  0.030514  0.022885  0.053399  0.038142  0.007628
998  0.046537  0.001724  0.027577  0.013789  0.008618  0.034472  0.001724
999  0.047060  0.002941  0.032354  0.026471  0.020589  0.038236  0.008824

          X7        X8        X9  …       X32       X33       X34  \
0    0.023451  0.007817  0.023451  …  0.032735  0.071063  0.050321
```

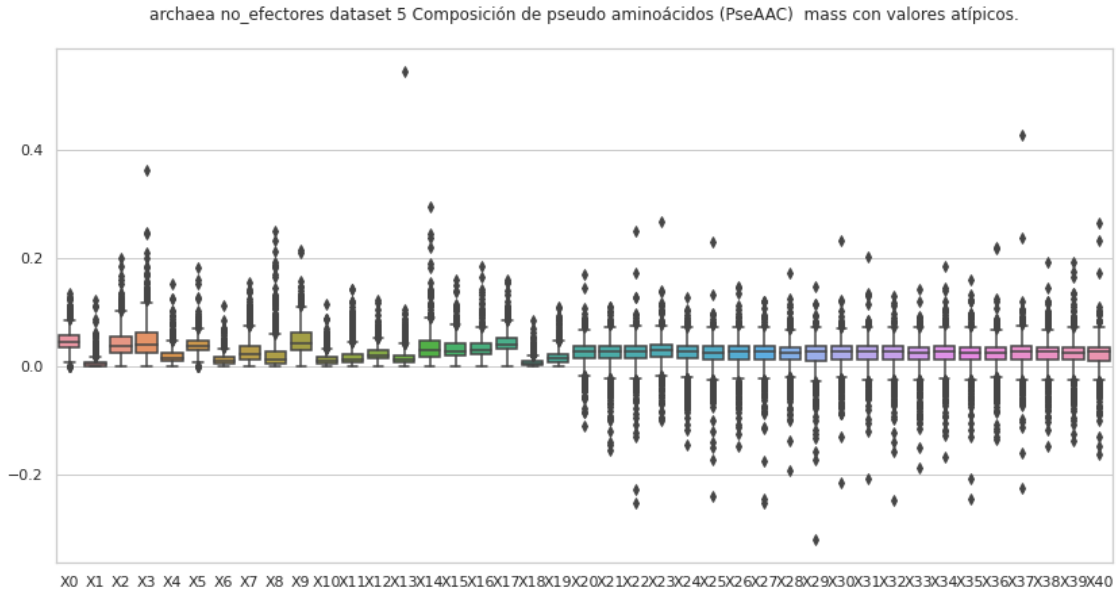
```
1    0.004618  0.004618  0.013853  …  0.017414  0.017759 -0.001672
2    0.007814  0.001736  0.022574  …  0.027111  0.037020  0.031356
3    0.010300  0.002060  0.024033  …  0.026727  0.036949  0.035326
4    0.011404  0.182460  0.057019  …  0.022673  0.009365  0.021343
..        …         …         …  …         …         …
995  0.029974  0.015415  0.029118  …  0.039061  0.036113  0.030888
996  0.022794  0.034191  0.091176  … -0.086933 -0.042084  0.027788
997  0.083912  0.030514  0.091541  …  0.047980 -0.023888 -0.036363
998  0.005171  0.003447  0.018960  …  0.044125  0.032664  0.033627
999  0.026471  0.035295  0.047060  …  0.025535 -0.025079  0.054310

          X35       X36       X37       X38       X39       X40           X41
0    0.008296 -0.050336  0.043490  0.052264 -0.024801 -0.091284  no_efectores
1    0.019966  0.014530  0.021395  0.022547  0.030141  0.022375  no_efectores
2    0.039974  0.034842  0.043115  0.025294  0.034034  0.046497  no_efectores
3    0.035064  0.042930  0.044613  0.028043  0.040732  0.025078  no_efectores
4    0.041213  0.037264  0.027300  0.025820  0.026082 -0.004575  no_efectores
..        …         …         …         …         …         …           …
995  0.011749  0.020645  0.044808  0.018189  0.034705  0.038372  no_efectores
996  0.029275 -0.059184 -0.020392  0.118298  0.078011 -0.013719  no_efectores
997  0.044841  0.035246  0.003709  0.030301  0.027887  0.017411  no_efectores
998  0.027955  0.024247  0.035396  0.040869  0.036693  0.027664  no_efectores
999  0.025265  0.029965  0.066819  0.001051  0.026714  0.031872  no_efectores

[1000 rows x 42 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass no_efectores archaea dataset 5, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.046168     0.006025     0.041737     0.045939     0.018982
std       0.018067     0.010986     0.026097     0.034464     0.016087
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       0.034453     0.000000     0.023459     0.022858     0.008758
50%       0.044182     0.002755     0.037629     0.038199     0.015054
75%       0.055427     0.006865     0.054851     0.061628     0.024131
max       0.134359     0.121952     0.198567     0.361115     0.152428

                X5           X6           X7           X8           X9  …  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
mean      0.038657     0.011290     0.028331     0.021224     0.047414  …
std       0.016212     0.011234     0.023726     0.029609     0.027082  …
min       0.000000     0.000000     0.000000     0.000000     0.000000  …
25%       0.029275     0.003782     0.011893     0.003811     0.027945  …
50%       0.036025     0.008459     0.021559     0.010722     0.040839  …
```

```
75%       0.045356      0.015260      0.036718      0.026236      0.061555  …
max       0.180558      0.110650      0.154770      0.248962      0.213271  …

                X31           X32           X33           X34           X35  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.021250     0.020310     0.020275     0.022214     0.021424
std       0.026216     0.027445     0.027379     0.027224     0.027817
min      -0.208162    -0.249203    -0.189526    -0.169750    -0.245611
25%       0.010344     0.011449     0.010992     0.011493     0.010677
50%       0.025266     0.025484     0.025079     0.025447     0.024818
75%       0.035424     0.035667     0.035179     0.036021     0.034923
max       0.202388     0.128293     0.140781     0.184250     0.159613

                X36           X37           X38           X39           X40
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.021234     0.022410     0.021575     0.021162     0.021006
std       0.027052     0.030869     0.026811     0.028389     0.028793
min      -0.135821    -0.225828    -0.148766    -0.137581    -0.163219
25%       0.012147     0.011366     0.011093     0.010372     0.010034
50%       0.025020     0.025489     0.025597     0.024689     0.025358
75%       0.034325     0.036422     0.035167     0.035148     0.035195
max       0.219654     0.425374     0.192616     0.190770     0.262707

[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 5,
sin valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.033111  0.005518  0.044148  0.044148  0.000000  0.027592  0.016555
1    0.056843  0.004060  0.056843  0.052783  0.036542  0.040602  0.012181
4    0.033241  0.000000  0.077563  0.138505  0.038781  0.060942  0.005540
5    0.036885  0.000000  0.007765  0.000971  0.006309  0.024752  0.002427
6    0.040558  0.020279  0.020279  0.067596  0.006760  0.033798  0.006760
..        …         …         …         …         …         …         …
995  0.027764  0.000000  0.028756  0.018840  0.003966  0.028756  0.001983
996  0.045736  0.000000  0.007916  0.007036  0.014952  0.027266  0.007036
997  0.070692  0.000000  0.063118  0.088366  0.037871  0.035346  0.015148
998  0.019257  0.001675  0.007535  0.006698  0.011722  0.039351  0.004186
999  0.038139  0.002007  0.022081  0.032117  0.006022  0.028103  0.002007

           X7        X8        X9  …       X32       X33       X34  \
0    0.022074  0.016555  0.027592  …  0.005305  0.008414 -0.006864
1    0.056843  0.069024  0.093386  … -0.046684  0.025830  0.004897
4    0.027701  0.027701  0.072022  …  0.054131 -0.032513 -0.018198
5    0.004853  0.001941  0.029605  …  0.033400  0.039271  0.033958
6    0.040558  0.074356  0.033798  …  0.031435  0.041259  0.010695
..        …         …         …  …         …         …         …
995  0.006941  0.002975  0.017848  …  0.031028  0.037475  0.042217
996  0.007036  0.000880  0.040459  …  0.034150  0.041883  0.025356
997  0.055544  0.050495  0.075742  … -0.008467 -0.023258  0.009328
998  0.020094  0.009210  0.016745  …  0.028033  0.030606  0.051957
999  0.020073  0.006022  0.042154  …  0.046583  0.016768  0.008048
```

	X35	X36	X37	X38	X39	X40	X41
0	0.050975	0.001327	0.034761	0.021363	0.059526	0.019504	efectores
1	0.011699	0.055134	-0.037160	0.000717	-0.021914	0.016411	efectores
4	-0.007510	0.028681	0.039624	0.031706	-0.023268	0.022998	efectores
5	0.039646	0.043598	0.045507	0.032750	0.034656	0.036351	efectores
6	-0.001480	0.048993	-0.000488	0.039017	0.007968	0.009783	efectores
..	…	…	…	…	…	…	
995	0.041387	0.027386	0.029623	0.027037	0.044458	0.030167	efectores
996	0.035280	0.031634	0.047551	0.038734	0.043469	0.030293	efectores
997	0.031926	-0.007977	-0.030076	0.022371	-0.022663	0.017131	efectores
998	0.025676	0.040041	0.030529	0.032726	0.030542	0.027294	efectores
999	0.013077	0.009811	0.036206	0.032757	0.028660	0.017248	efectores

[806 rows x 42 columns]

Composición de pseudo aminoácidos (PseAAC) mass efectores archaea dataset 5, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	806.000000	806.000000	806.000000	806.000000	806.000000	806.000000	
mean	0.045079	0.003143	0.030804	0.039076	0.020707	0.036943	
std	0.013529	0.004866	0.019951	0.031689	0.013909	0.010401	
min	0.007790	0.000000	0.003246	0.000000	0.000000	0.010074	
25%	0.036344	0.000000	0.015433	0.013436	0.011139	0.028767	
50%	0.043773	0.000773	0.024850	0.028136	0.017106	0.035664	
75%	0.052405	0.004846	0.043314	0.058974	0.026852	0.043372	
max	0.090831	0.027417	0.103644	0.150794	0.074508	0.071403	

	X6	X7	X8	X9	…	X31	\
count	806.000000	806.000000	806.000000	806.000000	…	806.000000	
mean	0.009286	0.031432	0.026687	0.058949	…	0.020794	
std	0.007356	0.027440	0.034328	0.026799	…	0.019798	
min	0.000000	0.000000	0.000000	0.008333	…	-0.058672	
25%	0.003457	0.010707	0.002888	0.040172	…	0.008559	
50%	0.007800	0.020961	0.009199	0.053906	…	0.024059	
75%	0.013421	0.045727	0.040229	0.076920	…	0.034505	
max	0.037124	0.131190	0.156803	0.146182	…	0.082061	

	X32	X33	X34	X35	X36	X37	\
count	806.000000	806.000000	806.000000	806.000000	806.000000	806.000000	
mean	0.024838	0.019690	0.020612	0.020426	0.023010	0.022272	
std	0.021292	0.019673	0.020568	0.020990	0.019665	0.022142	
min	-0.053973	-0.056548	-0.062229	-0.055014	-0.057866	-0.055843	
25%	0.013929	0.009410	0.009586	0.009300	0.013146	0.009097	
50%	0.028557	0.022905	0.024741	0.022993	0.027114	0.025492	
75%	0.038657	0.033418	0.034429	0.035132	0.035482	0.036735	

```
max      0.098149    0.089736    0.087991    0.079309    0.081816    0.084075

              X38         X39         X40
count  806.000000  806.000000  806.000000
mean     0.023416    0.024019    0.023390
std      0.019430    0.020294    0.020717
min     -0.052951   -0.050992   -0.052359
25%      0.013418    0.012537    0.011308
50%      0.026981    0.027326    0.025187
75%      0.035981    0.037316    0.036308
max      0.082999    0.085364    0.102513

[8 rows x 41 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass no_efectores archaea dataset 5, sin valores atípicos.
Valores del documento csv.

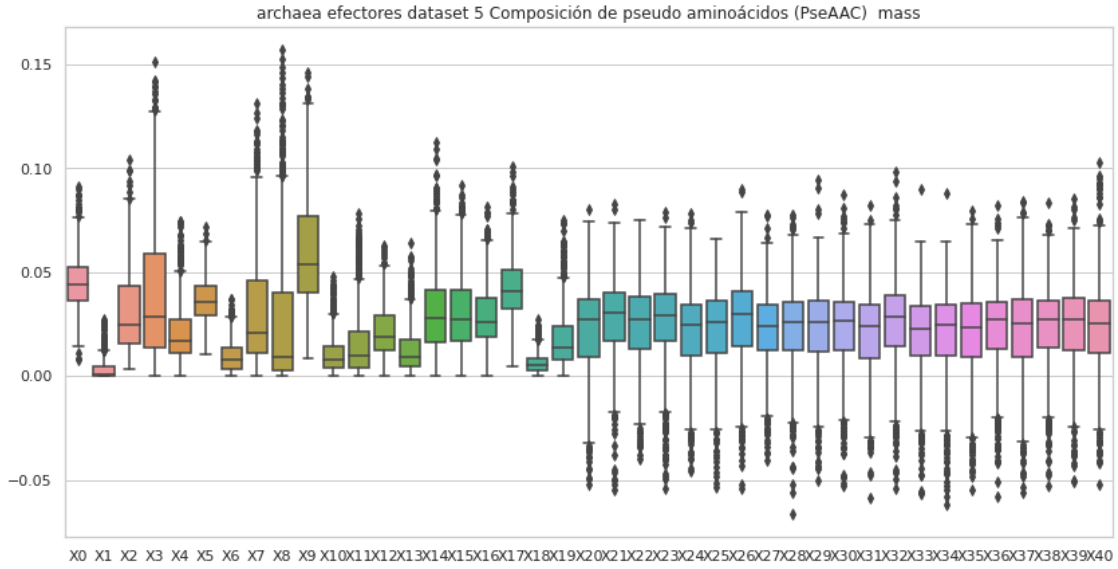
```
           X0        X1        X2        X3        X4        X5        X6  \
1    0.032324  0.000000  0.050794  0.046177  0.009235  0.027706  0.000000
2    0.031257  0.000000  0.014760  0.019101  0.002605  0.019969  0.003473
3    0.034333  0.001373  0.004807  0.006180  0.005493  0.025407  0.006867
5    0.075739  0.007973  0.083712  0.115602  0.047835  0.035877  0.019931
6    0.027199  0.004945  0.039562  0.009891  0.033381  0.034617  0.007418
..        …         …         …         …         …         …         …
994  0.031454  0.000000  0.031454  0.008578  0.020016  0.031454  0.011438
995  0.037682  0.005995  0.019697  0.029118  0.010277  0.038538  0.003426
997  0.041956  0.003814  0.030514  0.022885  0.053399  0.038142  0.007628
998  0.046537  0.001724  0.027577  0.013789  0.008618  0.034472  0.001724
999  0.047060  0.002941  0.032354  0.026471  0.020589  0.038236  0.008824

           X7        X8        X9  …       X32       X33       X34  \
1    0.004618  0.004618  0.013853  …  0.017414  0.017759 -0.001672
2    0.007814  0.001736  0.022574  …  0.027111  0.037020  0.031356
3    0.010300  0.002060  0.024033  …  0.026727  0.036949  0.035326
5    0.063780  0.031890  0.063780  … -0.010269 -0.003826  0.006248
6    0.051926  0.028436  0.037090  …  0.015169  0.014505  0.034424
..        …         …         …  …         …         …         …
994  0.034314  0.045751  0.037173  …  0.030718  0.020667  0.035202
995  0.029974  0.015415  0.029118  …  0.039061  0.036113  0.030888
997  0.083912  0.030514  0.091541  …  0.047980 -0.023888 -0.036363
998  0.005171  0.003447  0.018960  …  0.044125  0.032664  0.033627
999  0.026471  0.035295  0.047060  …  0.025535 -0.025079  0.054310

           X35       X36       X37       X38       X39       X40           X41
1    0.019966  0.014530  0.021395  0.022547  0.030141  0.022375  no_efectores
```

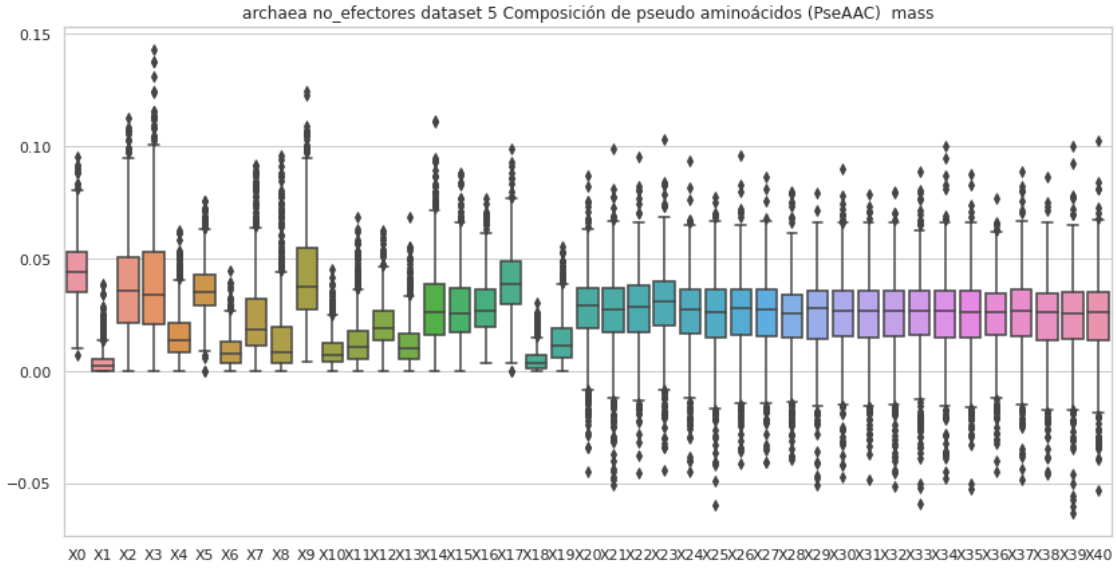
```
2    0.039974  0.034842  0.043115  0.025294  0.034034  0.046497  no_efectores
3    0.035064  0.042930  0.044613  0.028043  0.040732  0.025078  no_efectores
5   -0.000291  0.009013 -0.024823  0.023626  0.001625  0.037782  no_efectores
6    0.017499  0.005160  0.015371  0.038908  0.037551  0.046907  no_efectores
..        …         …         …         …         …         …
994  0.020200  0.032914  0.027053  0.014855  0.001388  0.027830  no_efectores
995  0.011749  0.020645  0.044808  0.018189  0.034705  0.038372  no_efectores
997  0.044841  0.035246  0.003709  0.030301  0.027887  0.017411  no_efectores
998  0.027955  0.024247  0.035396  0.040869  0.036693  0.027664  no_efectores
999  0.025265  0.029965  0.066819  0.001051  0.026714  0.031872  no_efectores

[837 rows x 42 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass no_efectores archaea dataset 5, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  837.000000  837.000000  837.000000  837.000000  837.000000  837.000000
mean     0.044776    0.004271    0.037808    0.039368    0.016193    0.036380
std      0.014143    0.005555    0.021609    0.025522    0.011147    0.011378
min      0.007440    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.034929    0.000000    0.021658    0.020792    0.008316    0.028932
50%      0.043933    0.002665    0.035620    0.034105    0.013737    0.035059
75%      0.053210    0.005575    0.050873    0.053204    0.021273    0.042798
max      0.095018    0.038557    0.112311    0.142900    0.062355    0.075611

                X6          X7          X8          X9  …         X31  \
count  837.000000  837.000000  837.000000  837.000000  …  837.000000
mean     0.009120    0.024178    0.015010    0.042421  …    0.024657
std      0.007282    0.018334    0.017701    0.021251  …    0.017487
min      0.000000    0.000000    0.000000    0.004332  …   -0.048418
25%      0.003495    0.011182    0.003369    0.027205  …    0.015176
50%      0.007631    0.018733    0.008389    0.037829  …    0.026621
75%      0.013049    0.032343    0.019887    0.054662  …    0.035717
max      0.044830    0.091867    0.096070    0.124777  …    0.078913

               X32         X33         X34         X35         X36         X37  \
count  837.000000  837.000000  837.000000  837.000000  837.000000  837.000000
mean     0.024400    0.024633    0.024866    0.024369    0.024417    0.025202
std      0.017902    0.018337    0.017987    0.017353    0.016388    0.017432
min     -0.051212   -0.058692   -0.047391   -0.052246   -0.044736   -0.048212
25%      0.015428    0.016112    0.014832    0.014932    0.015871    0.015382
50%      0.026731    0.026633    0.026659    0.026342    0.026446    0.026619
75%      0.035788    0.035699    0.035802    0.035792    0.034321    0.036078
max      0.079785    0.088797    0.100280    0.087387    0.076872    0.088685
```

```
              X38          X39          X40
count  837.000000   837.000000   837.000000
mean     0.023835     0.023781     0.023949
std      0.017653     0.018259     0.018472
min     -0.045621    -0.063062    -0.052790
25%      0.013689     0.014044     0.013701
50%      0.026366     0.025893     0.026487
75%      0.034880     0.035032     0.035296
max      0.086465     0.100298     0.102550

[8 rows x 41 columns]
```





5 Composición de pseudo aminoácidos (PseAAC) hidro

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

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

    if etiq == "efectores":
        df=PseAAC_hidro_efec

    if etiq == "no_efectores":
        df=PseAAC_hidro_no_efec

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

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

efectores
Composición de pseudo aminoácidos (PseAAC) hidro efectores archaea dataset 5, con valores atípicos.
Valores del documento csv.

```
          X0        X1        X2        X3        X4        X5        X6  \
0    0.049731  0.008288  0.066308  0.066308  0.000000  0.041442  0.024865
1    0.073037  0.005217  0.073037  0.067820  0.046952  0.052169  0.015651
2    0.027148  0.000000  0.034552  0.044425  0.007404  0.009872  0.004936
3    0.055367  0.005537  0.033220  0.094123  0.027683  0.005537  0.016610
4    0.014715  0.000000  0.034335  0.061312  0.017167  0.026977  0.002452
..        …         …         …         …         …         …         …
995  0.047622  0.000000  0.049323  0.032315  0.006803  0.049323  0.003402
996  0.049257  0.000000  0.008525  0.007578  0.016103  0.029365  0.007578
997  0.050096  0.000000  0.044729  0.062620  0.026837  0.025048  0.010735
998  0.072123  0.006272  0.028222  0.025086  0.043901  0.147383  0.015679
999  0.040581  0.002136  0.023494  0.034174  0.006408  0.029902  0.002136

          X7        X8        X9  …       X53       X54       X55  \
0    0.033154  0.024865  0.041442  …  0.002981  0.005368  0.076908
1    0.073037  0.088688  0.119989  … -0.030219 -0.014926  0.024617
2    0.014808  0.009872  0.027148  …  0.049099  0.014473  0.034066
3    0.105197  0.132880  0.060903  … -0.024699  0.028479  0.070147
4    0.012262  0.012262  0.031882  … -0.001729  0.024447  0.057116
..        …         …         …  …       …         …         …
995  0.011905  0.005102  0.030614  … -0.014078  0.010623  0.022484
996  0.007578  0.000947  0.043573  …  0.001496  0.030400  0.002788
997  0.039361  0.035783  0.053675  …  0.004007  0.003448  0.036412
998  0.075259  0.034494  0.062716  … -0.029739 -0.041286 -0.008103
999  0.021359  0.006408  0.044853  …  0.014094 -0.006701 -0.000822

          X56       X57       X58       X59       X60       X61         X62
0   -0.027231 -0.047835 -0.062661  0.017376 -0.037609  0.028287  efectores
1    0.034712 -0.039304 -0.034214  0.052247  0.026582  0.044323  efectores
2   -0.007506  0.017468  0.036885  0.048735 -0.004641  0.016697  efectores
3    0.027938 -0.056317 -0.002833  0.035212 -0.042616  0.027557  efectores
4    0.033649  0.022349  0.040897  0.057368 -0.026331  0.002251  efectores
..        …         …         …         …         …         …         …
995  0.031353  0.018045  0.027475  0.024592  0.020628  0.020588  efectores
996  0.034291  0.005096  0.036858  0.009996  0.040219  0.011173  efectores
997 -0.036980  0.010397  0.018441  0.042963  0.010391  0.022427  efectores
```

```
998  0.051656  0.028723 -0.000608  0.000515  0.000109 -0.026096  efectores
999 -0.009988 -0.011987  0.007313  0.009128 -0.016879  0.000869  efectores

[1000 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro efectores archaea dataset 5, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.048744	0.004803	0.036031	0.046610	0.021485	
std	0.027551	0.008074	0.025214	0.033867	0.015347	
min	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.029049	0.000000	0.013628	0.012390	0.010943	
50%	0.043420	0.001385	0.031014	0.046925	0.018012	
75%	0.060785	0.006488	0.054422	0.070048	0.027448	
max	0.238438	0.079479	0.158959	0.238438	0.114948	

	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.040171	0.010811	0.035138	0.031290	0.062843	…	
std	0.022341	0.009492	0.030405	0.036060	0.031610	…	
min	0.004232	0.000000	0.000000	0.000000	0.014001	…	
25%	0.025048	0.003651	0.012506	0.003460	0.039317	…	
50%	0.035230	0.009302	0.024057	0.014367	0.055587	…	
75%	0.050000	0.014987	0.049365	0.054172	0.078001	…	
max	0.170525	0.070141	0.216319	0.196013	0.224015	…	

	X52	X53	X54	X55	X56	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.005228	0.009880	0.013338	0.017267	0.005769	
std	0.043391	0.030760	0.039258	0.038592	0.037272	
min	-0.511184	-0.271572	-0.151675	-0.123419	-0.290783	
25%	-0.009600	-0.002016	-0.002735	0.000085	-0.010432	
50%	0.013515	0.007195	0.017174	0.010464	0.012141	
75%	0.025443	0.020357	0.028513	0.029077	0.026210	
max	0.237579	0.229576	0.597546	0.708873	0.170223	

	X57	X58	X59	X60	X61
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.010173	0.009803	0.012853	0.006423	0.012570
std	0.038295	0.035252	0.028332	0.040442	0.032573
min	-0.717482	-0.170582	-0.121535	-0.566560	-0.391603
25%	-0.001525	-0.006956	-0.000772	-0.008335	-0.000949
50%	0.008002	0.015432	0.008675	0.014329	0.008306
75%	0.024532	0.027977	0.026018	0.026873	0.023830

```
max       0.142703     0.324646     0.147493     0.144998     0.228597

[8 rows x 62 columns]
```

no_efectores
Composición de pseudo aminoácidos (PseAAC) hidro no_efectores archaea dataset 5, con valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.053209  0.010642  0.042568  0.074493  0.021284  0.053209  0.042568
1    0.029347  0.000000  0.046116  0.041924  0.008385  0.025154  0.000000
2    0.078622  0.000000  0.037127  0.048047  0.006552  0.050231  0.008736
3    0.060250  0.002410  0.008435  0.010845  0.009640  0.044585  0.012050
4    0.007841  0.000000  0.003920  0.019602  0.003920  0.000000  0.000000
..        …         …         …         …         …         …         …
995  0.092486  0.014714  0.048345  0.071467  0.025224  0.094588  0.008408
996  0.034434  0.015652  0.012521  0.031303  0.006261  0.009391  0.009391
997  0.013918  0.001265  0.010122  0.007592  0.017714  0.012653  0.002531
998  0.060846  0.002254  0.036057  0.018028  0.011268  0.045071  0.002254
999  0.074116  0.004632  0.050955  0.041690  0.032426  0.060219  0.013897

           X7        X8        X9  …       X53       X54       X55  \
0    0.031926  0.010642  0.031926  …  0.036895 -0.024394  0.008231
1    0.004192  0.004192  0.012577  … -0.008183 -0.013317  0.032007
2    0.019655  0.004368  0.056783  …  0.023138 -0.000736  0.012637
3    0.018075  0.003615  0.042175  … -0.002648  0.013914 -0.008229
4    0.003920  0.062727  0.019602  …  0.049552  0.017996  0.039179
..        …         …         …  …       …         …         …
995  0.073569  0.037835  0.071467  …  0.037339  0.037878  0.004828
996  0.006261  0.009391  0.025043  …  0.004733  0.014408  0.043521
997  0.027837  0.010122  0.030367  …  0.014318  0.027616  0.007042
998  0.006761  0.004507  0.024789  … -0.012778  0.003132  0.026539
999  0.041690  0.055587  0.074116  … -0.006736  0.025820  0.023586

          X56       X57       X58       X59       X60       X61           X62
0    0.024110 -0.032383 -0.068811 -0.003224  0.026111  0.033642  no_efectores
1    0.022351  0.078799 -0.028883  0.038584 -0.008304  0.038685  no_efectores
2   -0.003922 -0.010612 -0.004671  0.011247  0.023325  0.010878  no_efectores
3    0.010681 -0.006694  0.015986 -0.001875  0.016643 -0.009764  no_efectores
4   -0.032812  0.002621  0.048807  0.058996 -0.000932  0.021499  no_efectores
..        …         …         …         …         …         …           …
995  0.006175 -0.023056 -0.017330 -0.020277  0.039523  0.025378  no_efectores
996  0.077571  0.071681  0.004690 -0.009345 -0.018669  0.000235  no_efectores
997  0.042626  0.014256  0.007167 -0.003981  0.031775  0.005482  no_efectores
998 -0.009144  0.008456 -0.011763  0.009702  0.016874  0.029444  no_efectores
```

```
999  0.012861  0.071419 -0.002872  0.045111 -0.028609 -0.022812  no_efectores

[1000 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro no_efectores archaea dataset 5, con valores atípicos.
Estadísticas.

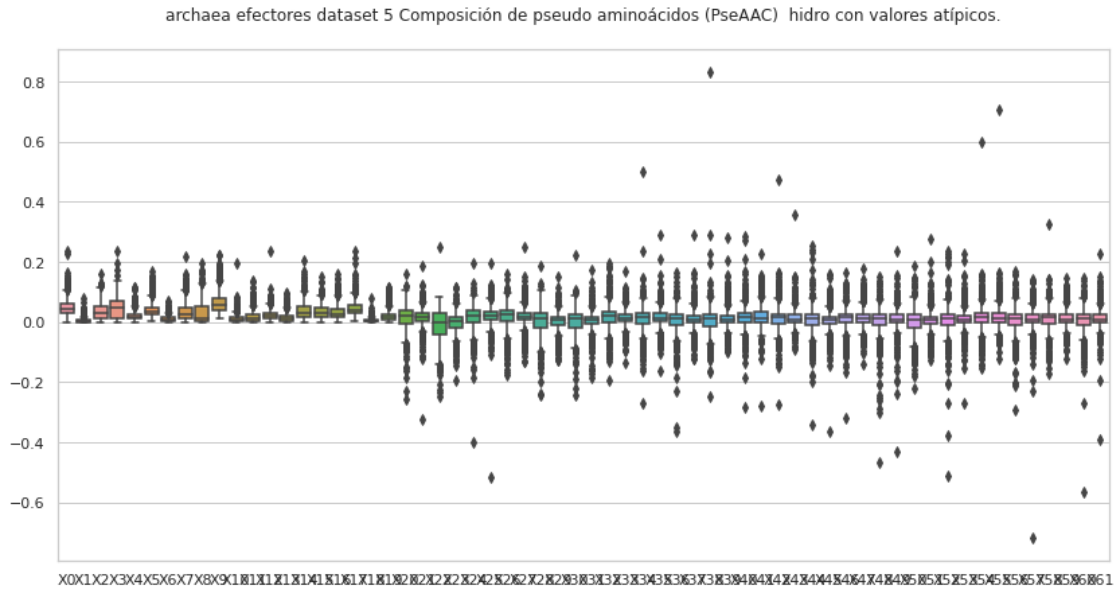
	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.058628	0.007156	0.049641	0.051853	0.021617	
std	0.030850	0.013013	0.027467	0.033937	0.018207	
min	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.036478	0.000000	0.030727	0.033251	0.011325	
50%	0.053989	0.003730	0.051621	0.052121	0.017954	
75%	0.075684	0.008639	0.066129	0.068431	0.027451	
max	0.230542	0.174781	0.254792	0.650052	0.196574	

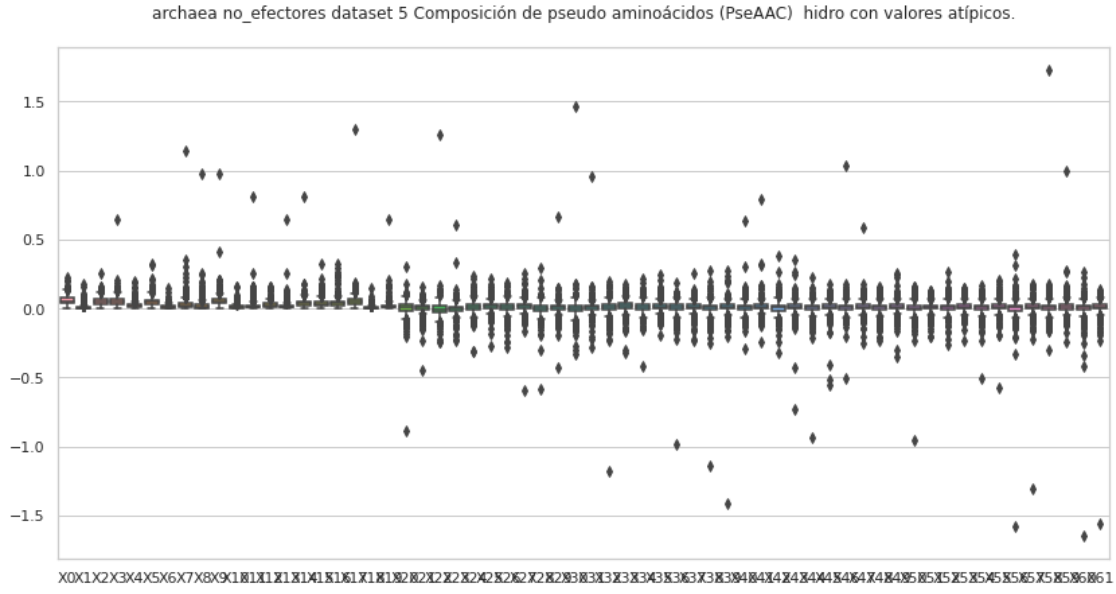
	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.049680	0.013090	0.033884	0.023197	0.055376	…	
std	0.029063	0.012448	0.046034	0.041233	0.042072	…	
min	0.000000	0.000000	0.000000	0.000000	0.000000	…	
25%	0.029940	0.004632	0.014406	0.004953	0.035901	…	
50%	0.045970	0.010725	0.025800	0.012312	0.050035	…	
75%	0.064039	0.018327	0.041775	0.027706	0.066724	…	
max	0.325026	0.148903	1.137591	0.975078	0.975078	…	

	X52	X53	X54	X55	X56	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.005560	0.013710	0.005720	0.013590	0.001810	
std	0.035606	0.031636	0.038601	0.038140	0.063815	
min	-0.257856	-0.227685	-0.501728	-0.576279	-1.581509	
25%	-0.011193	-0.001983	-0.010312	-0.001527	-0.012940	
50%	0.008294	0.012289	0.006948	0.011520	0.006805	
75%	0.023147	0.029421	0.023759	0.028506	0.021693	
max	0.260826	0.152584	0.163768	0.206574	0.389290	

	X57	X58	X59	X60	X61
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.010459	0.006101	0.013877	0.001627	0.009803
std	0.053513	0.066140	0.046961	0.066428	0.060219
min	-1.305793	-0.302242	-0.242564	-1.645915	-1.554167
25%	-0.002705	-0.011673	-0.003689	-0.010629	-0.002635
50%	0.010177	0.007143	0.011550	0.006698	0.010885
75%	0.026992	0.023024	0.028711	0.023354	0.029467
max	0.178639	1.726662	0.995707	0.261266	0.134267

[8 rows x 62 columns]





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

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

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


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

    if etiq == "efectores":
        df=PseAAC_hidro_efec

    if etiq == "no_efectores":
        df=PseAAC_hidro_no_efec

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

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


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

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

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

```
efectores
Composición de pseudo aminoácidos (PseAAC)  efectores archaea dataset 5, sin
valores atípicos.
Valores del documento csv.
```

	X0	X1	X2	X3	X4	X5	X6	\
0	0.049731	0.008288	0.066308	0.066308	0.000000	0.041442	0.024865	
2	0.027148	0.000000	0.034552	0.044425	0.007404	0.009872	0.004936	
4	0.014715	0.000000	0.034335	0.061312	0.017167	0.026977	0.002452	
5	0.042747	0.000000	0.008999	0.001125	0.007312	0.028685	0.002812	
6	0.036117	0.018059	0.018059	0.060195	0.006020	0.030098	0.006020	
..	…	…	…	…	…	…	…	
993	0.060086	0.018488	0.030043	0.039287	0.027732	0.067019	0.018488	
995	0.047622	0.000000	0.049323	0.032315	0.006803	0.049323	0.003402	
996	0.049257	0.000000	0.008525	0.007578	0.016103	0.029365	0.007578	
997	0.050096	0.000000	0.044729	0.062620	0.026837	0.025048	0.010735	
999	0.040581	0.002136	0.023494	0.034174	0.006408	0.029902	0.002136	

	X7	X8	X9	…	X53	X54	X55	\
0	0.033154	0.024865	0.041442	…	0.002981	0.005368	0.076908	
2	0.014808	0.009872	0.027148	…	0.049099	0.014473	0.034066	
4	0.012262	0.012262	0.031882	…	-0.001729	0.024447	0.057116	
5	0.005625	0.002250	0.034310	…	0.000724	0.029469	0.005243	
6	0.036117	0.066215	0.030098	…	0.016269	0.024014	0.049027	
..	…	…	…	…	…	…	…	
993	0.039287	0.055464	0.087818	…	0.004904	0.021860	0.019797	
995	0.011905	0.005102	0.030614	…	-0.014078	0.010623	0.022484	
996	0.007578	0.000947	0.043573	…	0.001496	0.030400	0.002788	
997	0.039361	0.035783	0.053675	…	0.004007	0.003448	0.036412	
999	0.021359	0.006408	0.044853	…	0.014094	-0.006701	-0.000822	

	X56	X57	X58	X59	X60	X61	X62
0	-0.027231	-0.047835	-0.062661	0.017376	-0.037609	0.028287	efectores
2	-0.007506	0.017468	0.036885	0.048735	-0.004641	0.016697	efectores
4	0.033649	0.022349	0.040897	0.057368	-0.026331	0.002251	efectores
5	0.019801	-0.000959	0.035902	0.005191	0.027939	-0.001910	efectores
6	0.012434	0.056836	0.013651	0.051430	-0.040923	0.023663	efectores
..	…	…	…	…	…	…	…
993	-0.019468	0.020989	0.018925	0.022466	0.000523	0.008492	efectores
995	0.031353	0.018045	0.027475	0.024592	0.020628	0.020588	efectores
996	0.034291	0.005096	0.036858	0.009996	0.040219	0.011173	efectores
997	-0.036980	0.010397	0.018441	0.042963	0.010391	0.022427	efectores
999	-0.009988	-0.011987	0.007313	0.009128	-0.016879	0.000869	efectores

```
[805 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC) efectores archaea dataset 5, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	805.000000	805.000000	805.000000	805.000000	805.000000	805.000000	
mean	0.044762	0.003243	0.030880	0.037258	0.018842	0.035823	
std	0.022596	0.005152	0.022054	0.027797	0.011004	0.016767	
min	0.004006	0.000000	0.002139	0.000000	0.000000	0.004232	
25%	0.028462	0.000000	0.011728	0.010142	0.010505	0.023701	
50%	0.040742	0.000809	0.024035	0.035692	0.016433	0.032262	
75%	0.056259	0.004284	0.048054	0.060754	0.024522	0.045433	
max	0.125291	0.028829	0.108642	0.110103	0.058830	0.105986	

	X6	X7	X8	X9	…	X52	\
count	805.000000	805.000000	805.000000	805.000000	…	805.000000	
mean	0.009298	0.027863	0.022958	0.054516	…	0.009527	
std	0.007510	0.023294	0.027635	0.022188	…	0.024083	
min	0.000000	0.000000	0.000000	0.014001	…	-0.084176	
25%	0.003143	0.011504	0.002824	0.037363	…	-0.001781	
50%	0.008461	0.019398	0.010066	0.050158	…	0.016319	
75%	0.012645	0.038617	0.037795	0.068494	…	0.025560	
max	0.038299	0.116014	0.126537	0.149386	…	0.095565	

	X53	X54	X55	X56	X57	X58	\
count	805.000000	805.000000	805.000000	805.000000	805.000000	805.000000	
mean	0.010504	0.013783	0.014302	0.010290	0.012013	0.012738	
std	0.019403	0.024520	0.023503	0.024453	0.021159	0.024590	
min	-0.080768	-0.089225	-0.070008	-0.081599	-0.057833	-0.071399	
25%	-0.000215	0.000557	0.000743	-0.004897	-0.000137	-0.000511	
50%	0.007375	0.017488	0.009613	0.015093	0.007783	0.016903	
75%	0.018689	0.027236	0.024343	0.026237	0.020842	0.027845	
max	0.098912	0.085311	0.114046	0.097906	0.101993	0.101671	

	X59	X60	X61
count	805.000000	805.000000	805.000000
mean	0.012345	0.010820	0.011581
std	0.021107	0.026195	0.021582
min	-0.064468	-0.105065	-0.067586
25%	-0.000103	-0.002494	-0.000384
50%	0.008441	0.016271	0.007796
75%	0.022454	0.026855	0.020831
max	0.090123	0.106651	0.102144

[8 rows x 62 columns]

```
no_efectores
Composición de pseudo aminoácidos (PseAAC)  no_efectores archaea dataset 5, sin
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.053209  0.010642  0.042568  0.074493  0.021284  0.053209  0.042568
1    0.029347  0.000000  0.046116  0.041924  0.008385  0.025154  0.000000
2    0.078622  0.000000  0.037127  0.048047  0.006552  0.050231  0.008736
3    0.060250  0.002410  0.008435  0.010845  0.009640  0.044585  0.012050
4    0.007841  0.000000  0.003920  0.019602  0.003920  0.000000  0.000000
..        …         …         …         …         …         …
995  0.092486  0.014714  0.048345  0.071467  0.025224  0.094588  0.008408
996  0.034434  0.015652  0.012521  0.031303  0.006261  0.009391  0.009391
997  0.013918  0.001265  0.010122  0.007592  0.017714  0.012653  0.002531
998  0.060846  0.002254  0.036057  0.018028  0.011268  0.045071  0.002254
999  0.074116  0.004632  0.050955  0.041690  0.032426  0.060219  0.013897

           X7        X8        X9  …       X53       X54       X55  \
0    0.031926  0.010642  0.031926  …  0.036895 -0.024394  0.008231
1    0.004192  0.004192  0.012577  … -0.008183 -0.013317  0.032007
2    0.019655  0.004368  0.056783  …  0.023138 -0.000736  0.012637
3    0.018075  0.003615  0.042175  … -0.002648  0.013914 -0.008229
4    0.003920  0.062727  0.019602  …  0.049552  0.017996  0.039179
..        …         …         …  …         …         …
995  0.073569  0.037835  0.071467  …  0.037339  0.037878  0.004828
996  0.006261  0.009391  0.025043  …  0.004733  0.014408  0.043521
997  0.027837  0.010122  0.030367  …  0.014318  0.027616  0.007042
998  0.006761  0.004507  0.024789  … -0.012778  0.003132  0.026539
999  0.041690  0.055587  0.074116  … -0.006736  0.025820  0.023586

          X56       X57       X58       X59       X60       X61           X62
0    0.024110 -0.032383 -0.068811 -0.003224  0.026111  0.033642  no_efectores
1    0.022351  0.078799 -0.028883  0.038584 -0.008304  0.038685  no_efectores
2   -0.003922 -0.010612 -0.004671  0.011247  0.023325  0.010878  no_efectores
3    0.010681 -0.006694  0.015986 -0.001875  0.016643 -0.009764  no_efectores
4   -0.032812  0.002621  0.048807  0.058996 -0.000932  0.021499  no_efectores
..        …         …         …         …         …         …
995  0.006175 -0.023056 -0.017330 -0.020277  0.039523  0.025378  no_efectores
996  0.077571  0.071681  0.004690 -0.009345 -0.018669  0.000235  no_efectores
997  0.042626  0.014256  0.007167 -0.003981  0.031775  0.005482  no_efectores
998 -0.009144  0.008456 -0.011763  0.009702  0.016874  0.029444  no_efectores
999  0.012861  0.071419 -0.002872  0.045111 -0.028609 -0.022812  no_efectores

[880 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC) no_efectores archaea dataset 5, sin valores atípicos.
Estadísticas.

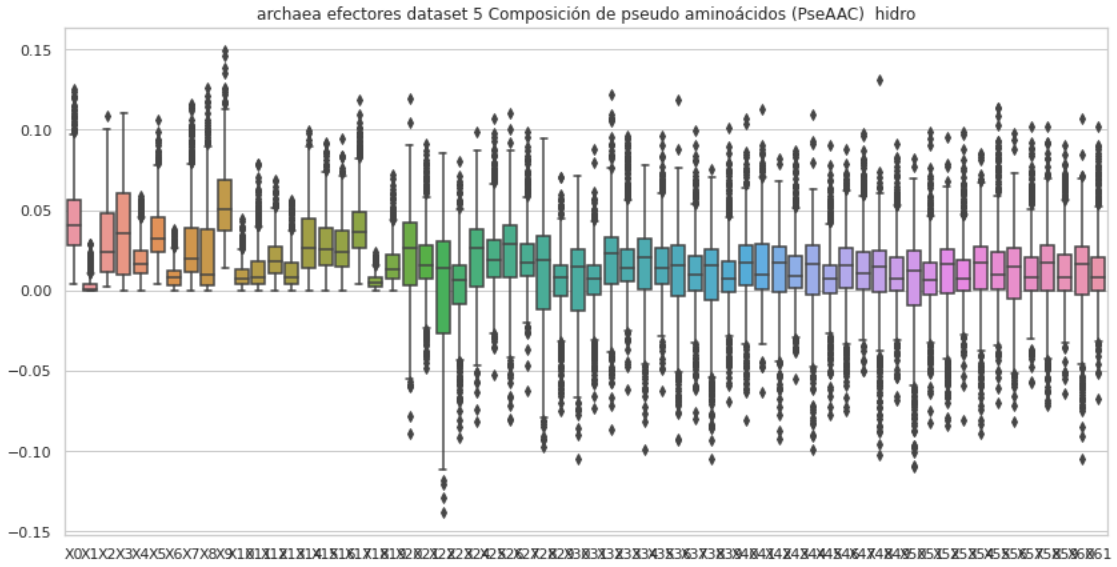
	X0	X1	X2	X3	X4	X5	\
count	880.000000	880.000000	880.000000	880.000000	880.000000	880.000000	
mean	0.056934	0.005325	0.047550	0.048366	0.019290	0.046281	
std	0.027682	0.006629	0.024934	0.025265	0.011847	0.022238	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.036174	0.000000	0.028989	0.030578	0.011148	0.029220	
50%	0.053157	0.003568	0.051087	0.051097	0.017313	0.043827	
75%	0.074452	0.007554	0.064538	0.066191	0.025653	0.061091	
max	0.149882	0.042563	0.114139	0.126784	0.066574	0.122552	

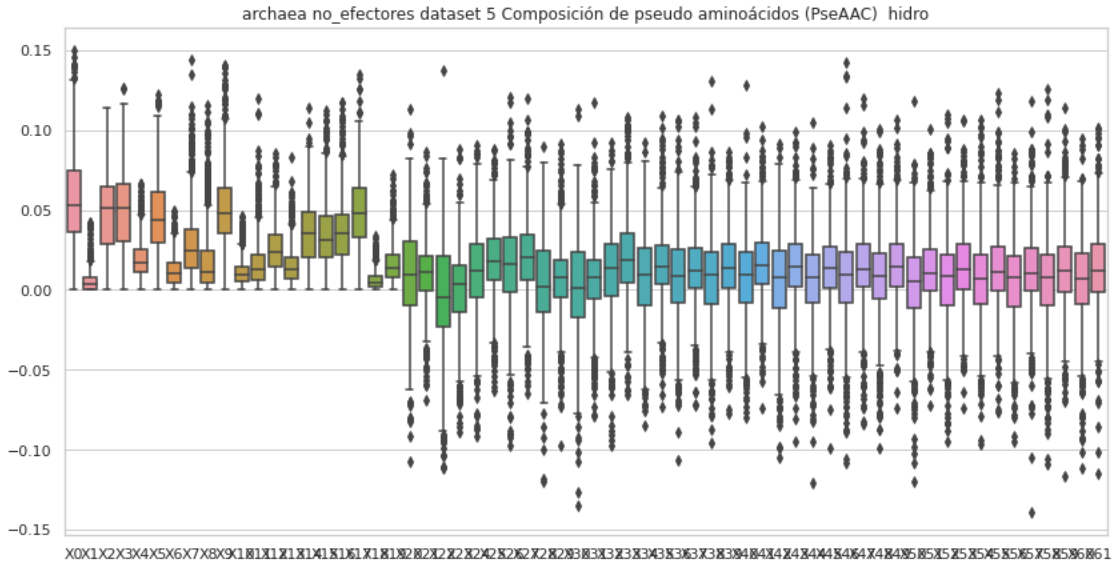
	X6	X7	X8	X9	…	X52	\
count	880.000000	880.000000	880.000000	880.000000	…	880.000000	
mean	0.011600	0.029090	0.018466	0.050801	…	0.005962	
std	0.008876	0.021601	0.020623	0.022895	…	0.026697	
min	0.000000	0.000000	0.000000	0.000000	…	-0.094844	
25%	0.004531	0.013894	0.004624	0.035217	…	-0.009573	
50%	0.010263	0.024361	0.011071	0.048243	…	0.008457	
75%	0.017093	0.037848	0.024307	0.063926	…	0.021851	
max	0.050054	0.143782	0.116087	0.141059	…	0.109922	

	X53	X54	X55	X56	X57	X58	\
count	880.000000	880.000000	880.000000	880.000000	880.000000	880.000000	
mean	0.014254	0.006591	0.014125	0.004364	0.012623	0.005688	
std	0.023462	0.026020	0.024083	0.025067	0.024180	0.027783	
min	-0.073660	-0.096369	-0.076621	-0.094587	-0.139133	-0.108802	
25%	0.000097	-0.008408	-0.000627	-0.010616	-0.000844	-0.009262	
50%	0.012854	0.007210	0.011576	0.007521	0.010400	0.007610	
75%	0.028413	0.022508	0.027196	0.020946	0.026264	0.022151	
max	0.106335	0.106195	0.122948	0.085734	0.118281	0.126131	

	X59	X60	X61
count	880.000000	880.000000	880.000000
mean	0.012869	0.005786	0.013456
std	0.025046	0.026508	0.025033
min	-0.116646	-0.111652	-0.114967
25%	-0.001681	-0.008396	-0.000879
50%	0.011769	0.007455	0.011778
75%	0.027240	0.022693	0.029011
max	0.114380	0.094774	0.101460

[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 5,
con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.015888  0.060600  0.021128  0.175120 -0.042455 -0.025333  0.072468
1   -0.112544  0.093547 -0.008473  0.003300  0.110713 -0.091982  0.096355
2    0.068264  0.039092 -0.007323  0.004073 -0.012690 -0.004975  0.034127
3    0.049046 -0.045433 -0.029467 -0.085142 -0.118806 -0.017147  0.039711
4   -0.021547 -0.096534  0.041258  0.115039 -0.078061  0.006559 -0.044745
..        …         …         …         …         …         …         …
995  0.064766 -0.017841 -0.070164  0.147441  0.093923  0.095305  0.027736
996 -0.078849  0.073196 -0.113261  0.059961 -0.005648  0.015498  0.041653
997  0.023128  0.075865  0.003858  0.013955  0.027401  0.015069 -0.017081
998 -0.053129  0.059062 -0.060425 -0.098835  0.054804  0.027718  0.089254
999  0.160396  0.075309 -0.005707 -0.042872  0.024529  0.078772  0.086375

           X7        X8        X9       X10       X11       X12        X13
0    0.085899 -0.077145 -0.002226  0.089885 -0.024457 -0.081080  efectores
```

```
1   -0.016797  0.131995 -0.033148 -0.055136  0.053453 -0.145517  efectores
2   -0.039125  0.068679  0.140358  0.174350  0.001696  0.050186  efectores
3    0.051694 -0.007213  0.040958  0.003637  0.054353 -0.017675  efectores
4    0.018083  0.074150 -0.106667 -0.074817 -0.044245  0.106460  efectores
..        …       …       …       …       …       …        …
995  0.107874  0.106333  0.065946  0.001640  0.000423  0.027021  efectores
996  0.106638 -0.143074 -0.051620 -0.059269  0.058394  0.005048  efectores
997 -0.016367  0.000739  0.055297  0.048400  0.013840 -0.022282  efectores
998  0.134793  0.008506  0.047533 -0.046301  0.046333 -0.028229  efectores
999 -0.086757 -0.008986 -0.017921 -0.000169  0.043513  0.092020  efectores

[1000 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro_mass efectores archaea dataset 5, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.003804     0.017101     0.009915     0.015535    -0.005843
std       0.071714     0.067656     0.067399     0.068485     0.071362
min      -0.297766    -0.274422    -0.348021    -0.228345    -0.296786
25%      -0.040183    -0.025912    -0.030464    -0.028208    -0.046471
50%       0.004498     0.020300     0.007656     0.017660    -0.002226
75%       0.047194     0.058008     0.049777     0.058943     0.037974
max       0.224149     0.248910     0.263949     0.273984     0.226649

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean     -0.000485     0.018395    -0.001129    -0.000185     0.004682
std       0.066256     0.070279     0.064823     0.077263     0.073817
min      -0.277328    -0.211212    -0.249870    -0.324798    -0.317040
25%      -0.038771    -0.025785    -0.039004    -0.044226    -0.037979
50%       0.001685     0.014979    -0.001207     0.002766     0.004945
75%       0.038973     0.063059     0.038027     0.047555     0.048852
max       0.256319     0.483294     0.331801     0.315391     0.237331

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.007547    -0.004682     0.016271
std       0.071549     0.068713     0.071882
min      -0.225828    -0.302639    -0.218270
25%      -0.034762    -0.045027    -0.031879
50%       0.006970    -0.004149     0.011628
75%       0.046609     0.034862     0.063822
max       0.373070     0.239947     0.276165
```

no_efectores
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores archaea dataset 5,
con valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.289354	0.163037	0.143272	0.482456	0.211624	0.074577	-0.073506	
1	0.080026	-0.032431	0.103822	0.149245	-0.060713	0.059636	0.194522	
2	0.060699	0.011629	0.081504	0.029490	0.062128	0.055391	0.011149	
3	-0.087508	0.040694	0.056822	0.065032	0.100864	0.065769	-0.016954	
4	0.037192	-0.019966	-0.050050	-0.042743	-0.061527	0.031641	0.000363	
..	…	…	…	…	…	…	…	
995	0.023676	0.058927	-0.040920	0.017641	0.032638	-0.001646	0.010442	
996	-0.012135	0.004585	-0.003821	-0.092199	0.016003	-0.126202	0.108575	
997	0.020419	-0.194169	-0.006013	-0.018656	-0.010752	0.095136	0.079212	
998	-0.130746	0.105529	-0.007510	0.123763	0.085111	0.190545	-0.040033	
999	-0.029587	-0.001096	-0.034820	-0.131356	0.005997	0.053616	0.000204	

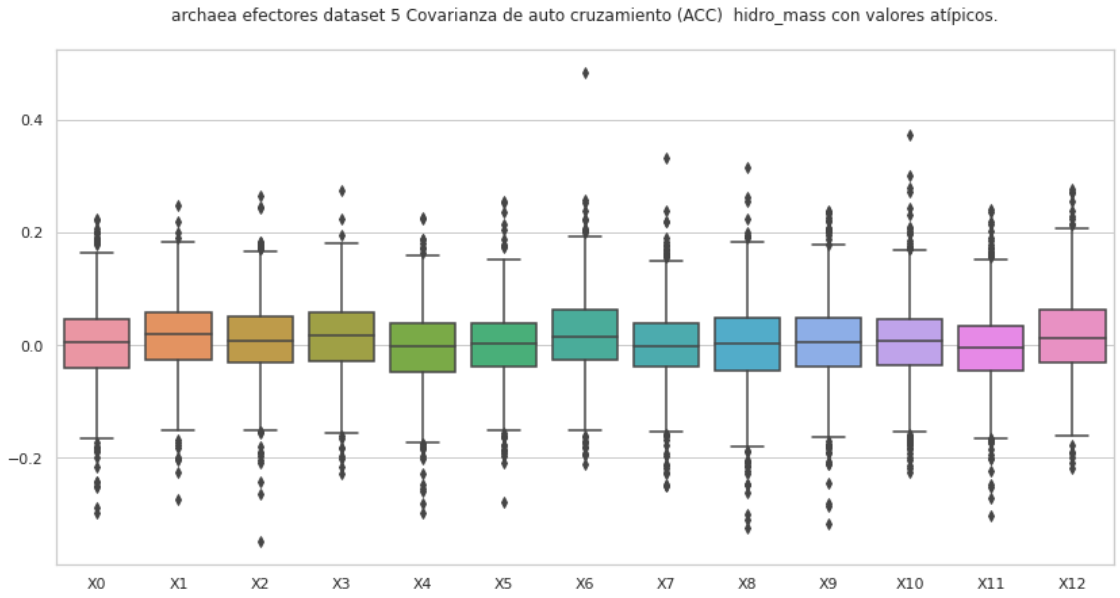
	X7	X8	X9	X10	X11	X12	X13
0	0.140005	0.184886	0.164242	0.156792	-0.042461	0.148923	no_efectores
1	0.063453	0.030572	0.050420	0.039967	0.130524	-0.043992	no_efectores
2	-0.005342	-0.010882	0.014255	-0.024583	0.009951	-0.056087	no_efectores
3	0.060539	-0.016741	-0.057371	0.088469	0.032080	-0.098137	no_efectores
4	0.100639	-0.010084	-0.021324	-0.012185	-0.011518	0.016712	no_efectores
..	…	…	…	…	…	…	…
995	0.003137	-0.037532	0.044270	-0.045214	0.037449	0.058529	no_efectores
996	-0.122039	0.014260	0.150582	-0.065986	-0.074404	-0.170801	no_efectores
997	-0.154020	-0.040941	0.116862	0.003368	0.087313	0.095176	no_efectores
998	0.144520	0.002247	0.034367	0.049143	-0.052735	0.104162	no_efectores
999	0.044638	0.129188	-0.009300	0.062938	-0.082059	0.008649	no_efectores

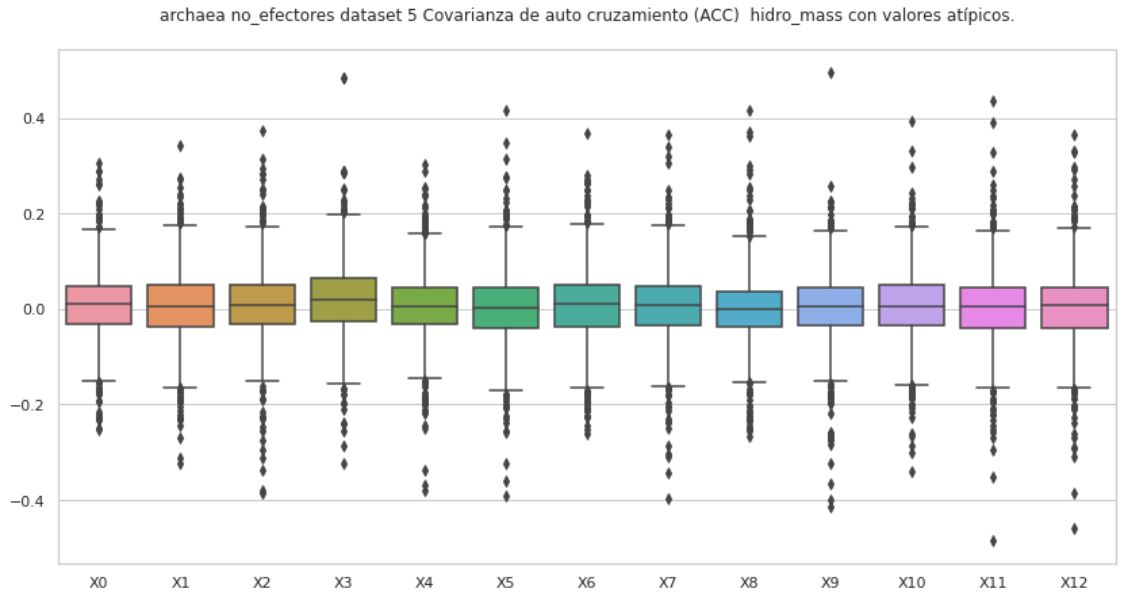
[1000 rows x 14 columns]

Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores archaea dataset 5,
con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.009303	0.006101	0.009699	0.020745	0.005347	
std	0.072171	0.076712	0.076524	0.079435	0.075353	
min	-0.252802	-0.323394	-0.384568	-0.323729	-0.380453	
25%	-0.031226	-0.036459	-0.032151	-0.025723	-0.032646	
50%	0.011270	0.004396	0.007019	0.020308	0.005784	
75%	0.048339	0.049260	0.050482	0.065140	0.043463	

```
max       0.304973     0.340848     0.373340     0.484579     0.303703

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.001541     0.006987     0.006023     0.000675     0.003679
std       0.076957     0.076728     0.075971     0.074469     0.075889
min      -0.390313    -0.261885    -0.395905    -0.267932    -0.412624
25%      -0.041482    -0.037293    -0.035318    -0.038542    -0.034946
50%       0.002523     0.009678     0.007278    -0.000649     0.005875
75%       0.044635     0.049779     0.048846     0.037540     0.045462
max       0.417063     0.368375     0.364827     0.416948     0.494751

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.006632     0.004022     0.003559
std       0.074729     0.078896     0.078142
min      -0.339390    -0.484584    -0.458479
25%      -0.033890    -0.039184    -0.039058
50%       0.005357     0.006268     0.007598
75%       0.049245     0.044334     0.045568
max       0.394096     0.436612     0.364752
```





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

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

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

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

    if etiq == "efectores":
        df=ACC_hidro_mass_efec

    if etiq == "no_efectores":
        df=ACC_hidro_mass_no_efec

    del df['X13']
```

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

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

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

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

```
efectores
Covarianza de auto cruzamiento (ACC)  hidro_mass efectores archaea dataset 5,
sin valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.015888  0.060600  0.021128  0.175120 -0.042455 -0.025333  0.072468
1   -0.112544  0.093547 -0.008473  0.003300  0.110713 -0.091982  0.096355
2    0.068264  0.039092 -0.007323  0.004073 -0.012690 -0.004975  0.034127
3    0.049046 -0.045433 -0.029467 -0.085142 -0.118806 -0.017147  0.039711
4   -0.021547 -0.096534  0.041258  0.115039 -0.078061  0.006559 -0.044745
..        …         …         …         …         …         …         …
995  0.064766 -0.017841 -0.070164  0.147441  0.093923  0.095305  0.027736
996 -0.078849  0.073196 -0.113261  0.059961 -0.005648  0.015498  0.041653
997  0.023128  0.075865  0.003858  0.013955  0.027401  0.015069 -0.017081
998 -0.053129  0.059062 -0.060425 -0.098835  0.054804  0.027718  0.089254
999  0.160396  0.075309 -0.005707 -0.042872  0.024529  0.078772  0.086375

           X7        X8        X9       X10       X11       X12        X13
0    0.085899 -0.077145 -0.002226  0.089885 -0.024457 -0.081080  efectores
1   -0.016797  0.131995 -0.033148 -0.055136  0.053453 -0.145517  efectores
2   -0.039125  0.068679  0.140358  0.174350  0.001696  0.050186  efectores
3    0.051694 -0.007213  0.040958  0.003637  0.054353 -0.017675  efectores
4    0.018083  0.074150 -0.106667 -0.074817 -0.044245  0.106460  efectores
..        …         …         …         …         …         …         …
995  0.107874  0.106333  0.065946  0.001640  0.000423  0.027021  efectores
```

```
996  0.106638 -0.143074 -0.051620 -0.059269  0.058394  0.005048  efectores
997 -0.016367  0.000739  0.055297  0.048400  0.013840 -0.022282  efectores
998  0.134793  0.008506  0.047533 -0.046301  0.046333 -0.028229  efectores
999 -0.086757 -0.008986 -0.017921 -0.000169  0.043513  0.092020  efectores

[918 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro_mass efectores archaea dataset 5, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	918.000000	918.000000	918.000000	918.000000	918.000000	918.000000	
mean	0.005498	0.017343	0.011070	0.014501	-0.004291	-0.000599	
std	0.066231	0.061997	0.061403	0.064838	0.065694	0.062467	
min	-0.197952	-0.181738	-0.180701	-0.181800	-0.202583	-0.196772	
25%	-0.037614	-0.023255	-0.028703	-0.027927	-0.043555	-0.038290	
50%	0.004636	0.020036	0.007788	0.016909	-0.001490	0.001622	
75%	0.045186	0.056296	0.048231	0.056168	0.037128	0.038928	
max	0.206642	0.182271	0.183038	0.181346	0.187027	0.176608	

	X6	X7	X8	X9	X10	X11	\
count	918.000000	918.000000	918.000000	918.000000	918.000000	918.000000	
mean	0.018055	0.000417	0.000743	0.005323	0.006467	-0.003534	
std	0.063627	0.057526	0.070234	0.066766	0.064587	0.062622	
min	-0.180613	-0.191788	-0.226893	-0.210567	-0.204867	-0.200621	
25%	-0.023299	-0.036860	-0.042648	-0.034973	-0.033882	-0.042741	
50%	0.014979	-0.000459	0.003199	0.004945	0.006280	-0.003589	
75%	0.061229	0.037868	0.045555	0.048250	0.044779	0.034665	
max	0.220763	0.190465	0.223667	0.207986	0.209027	0.190036	

	X12
count	918.000000
mean	0.015370
std	0.067056
min	-0.189123
25%	-0.030571
50%	0.010610
75%	0.060261
max	0.213417

no_efectores
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores archaea dataset 5, sin valores atípicos.
Valores del documento csv.

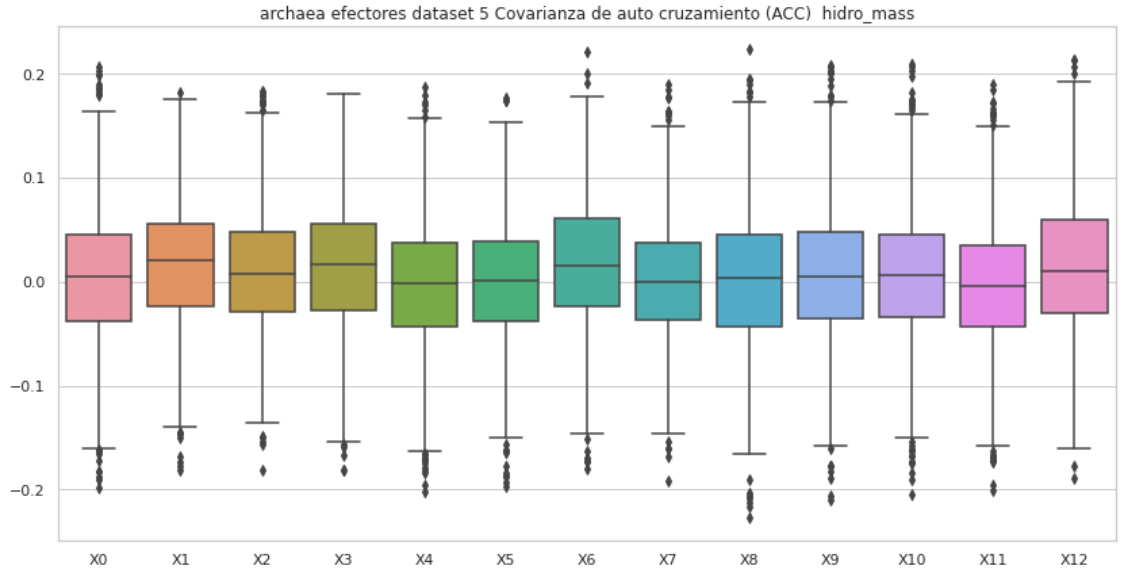
```
           X0        X1        X2        X3        X4        X5        X6  \
1    0.080026 -0.032431  0.103822  0.149245 -0.060713  0.059636  0.194522
2    0.060699  0.011629  0.081504  0.029490  0.062128  0.055391  0.011149
3   -0.087508  0.040694  0.056822  0.065032  0.100864  0.065769 -0.016954
4    0.037192 -0.019966 -0.050050 -0.042743 -0.061527  0.031641  0.000363
5   -0.011557 -0.053341  0.008578  0.007840 -0.030196  0.064348  0.068808
..        …         …         …         …         …         …
995  0.023676  0.058927 -0.040920  0.017641  0.032638 -0.001646  0.010442
996 -0.012135  0.004585 -0.003821 -0.092199  0.016003 -0.126202  0.108575
997  0.020419 -0.194169 -0.006013 -0.018656 -0.010752  0.095136  0.079212
998 -0.130746  0.105529 -0.007510  0.123763  0.085111  0.190545 -0.040033
999 -0.029587 -0.001096 -0.034820 -0.131356  0.005997  0.053616  0.000204

           X7        X8        X9       X10       X11       X12           X13
1    0.063453  0.030572  0.050420  0.039967  0.130524 -0.043992  no_efectores
2   -0.005342 -0.010882  0.014255 -0.024583  0.009951 -0.056087  no_efectores
3    0.060539 -0.016741 -0.057371  0.088469  0.032080 -0.098137  no_efectores
4    0.100639 -0.010084 -0.021324 -0.012185 -0.011518  0.016712  no_efectores
5   -0.057018  0.045682 -0.013058  0.043817 -0.087080 -0.029315  no_efectores
..        …         …         …         …         …         …           …
995  0.003137 -0.037532  0.044270 -0.045214  0.037449  0.058529  no_efectores
996 -0.122039  0.014260  0.150582 -0.065986 -0.074404 -0.170801  no_efectores
997 -0.154020 -0.040941  0.116862  0.003368  0.087313  0.095176  no_efectores
998  0.144520  0.002247  0.034367  0.049143 -0.052735  0.104162  no_efectores
999  0.044638  0.129188 -0.009300  0.062938 -0.082059  0.008649  no_efectores

[899 rows x 14 columns]
```

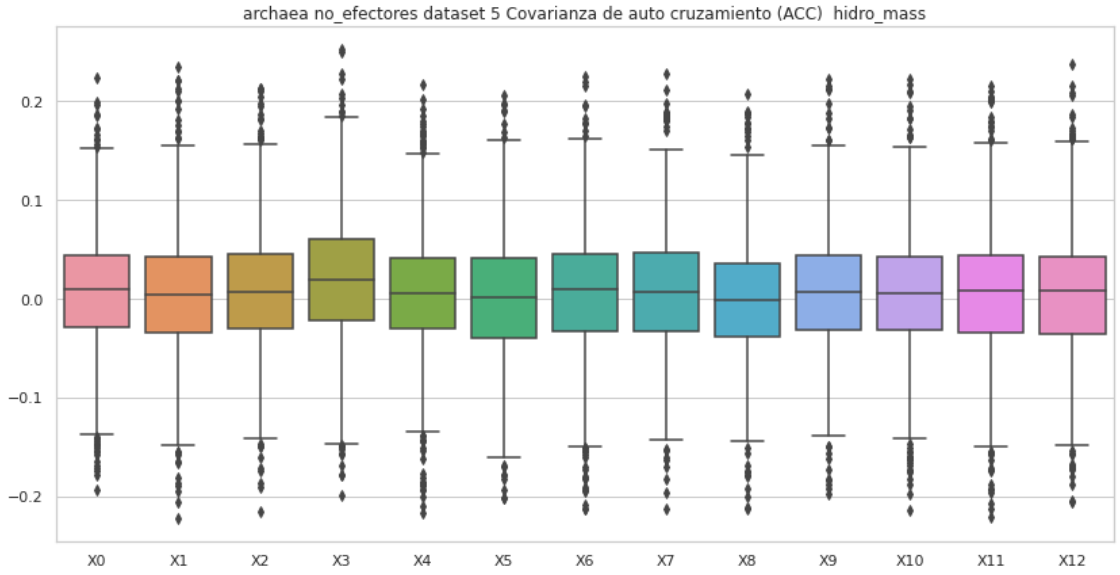
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores archaea dataset 5, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  899.000000  899.000000  899.000000  899.000000  899.000000  899.000000
mean     0.007688    0.005476    0.009131    0.019578    0.005748    0.000622
std      0.061818    0.065794    0.062168    0.066644    0.064067    0.063044
min     -0.192789   -0.222121   -0.215665   -0.198480   -0.216087   -0.201870
25%     -0.029357   -0.034697   -0.030655   -0.022107   -0.030624   -0.039753
50%      0.009837    0.003701    0.006898    0.019938    0.005997    0.000920
75%      0.043792    0.042504    0.045933    0.060601    0.040652    0.041172
max      0.223433    0.234617    0.213011    0.252445    0.216769    0.206343

                X6          X7          X8          X9         X10         X11  \
count  899.000000  899.000000  899.000000  899.000000  899.000000  899.000000
mean     0.006968    0.007028   -0.000512    0.006317    0.006101    0.005842
std      0.064395    0.062872    0.061558    0.061861    0.063552    0.066386
```

min	-0.212791	-0.212695	-0.212024	-0.197885	-0.214471	-0.220524
25%	-0.032507	-0.032335	-0.037943	-0.031534	-0.031913	-0.034181
50%	0.009868	0.007402	-0.000628	0.006507	0.005356	0.007913
75%	0.045570	0.046544	0.036283	0.043886	0.043329	0.043720
max	0.224926	0.227464	0.206776	0.222488	0.222676	0.216188

	X12
count	899.000000
mean	0.004298
std	0.063408
min	-0.205368
25%	-0.035561
50%	0.007908
75%	0.042954
max	0.236953





7 Covarianza de auto cruzamiento (ACC) mass

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

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

    if etiq == "efectores":
        df=ACC_mass_efec

    if etiq == "no_efectores":
        df=ACC_mass_no_efec

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

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

```
efectores
Covarianza de auto cruzamiento (ACC)  mass efectores archaea dataset 5, con
valores atípicos.
Valores del documento csv.
```

	X0	X1	X2	X3	X4	X5	X6	\
0	-0.015888	0.060600	0.021128	0.175120	-0.042455	-0.025333	0.072468	
1	-0.112544	0.093547	-0.008473	0.003300	0.110713	-0.091982	0.096355	
2	0.068264	0.039092	-0.007323	0.004073	-0.012690	-0.004975	0.034127	
3	0.049046	-0.045433	-0.029467	-0.085142	-0.118806	-0.017147	0.039711	
4	-0.021547	-0.096534	0.041258	0.115039	-0.078061	0.006559	-0.044745	
..	…	…	…	…	…	…	…	
995	0.064766	-0.017841	-0.070164	0.147441	0.093923	0.095305	0.027736	
996	-0.078849	0.073196	-0.113261	0.059961	-0.005648	0.015498	0.041653	
997	0.023128	0.075865	0.003858	0.013955	0.027401	0.015069	-0.017081	
998	-0.053129	0.059062	-0.060425	-0.098835	0.054804	0.027718	0.089254	
999	0.160396	0.075309	-0.005707	-0.042872	0.024529	0.078772	0.086375	

	X7	X8	X9	X10	X11	X12	X13
0	0.085899	-0.077145	-0.002226	0.089885	-0.024457	-0.081080	efectores
1	-0.016797	0.131995	-0.033148	-0.055136	0.053453	-0.145517	efectores
2	-0.039125	0.068679	0.140358	0.174350	0.001696	0.050186	efectores
3	0.051694	-0.007213	0.040958	0.003637	0.054353	-0.017675	efectores
4	0.018083	0.074150	-0.106667	-0.074817	-0.044245	0.106460	efectores
..	…	…	…	…	…	…	
995	0.107874	0.106333	0.065946	0.001640	0.000423	0.027021	efectores
996	0.106638	-0.143074	-0.051620	-0.059269	0.058394	0.005048	efectores
997	-0.016367	0.000739	0.055297	0.048400	0.013840	-0.022282	efectores
998	0.134793	0.008506	0.047533	-0.046301	0.046333	-0.028229	efectores
999	-0.086757	-0.008986	-0.017921	-0.000169	0.043513	0.092020	efectores

```
[1000 rows x 14 columns]
```

```
Covarianza de auto cruzamiento (ACC)  mass efectores archaea dataset 5, con
valores atípicos.
Estadísticas.
```

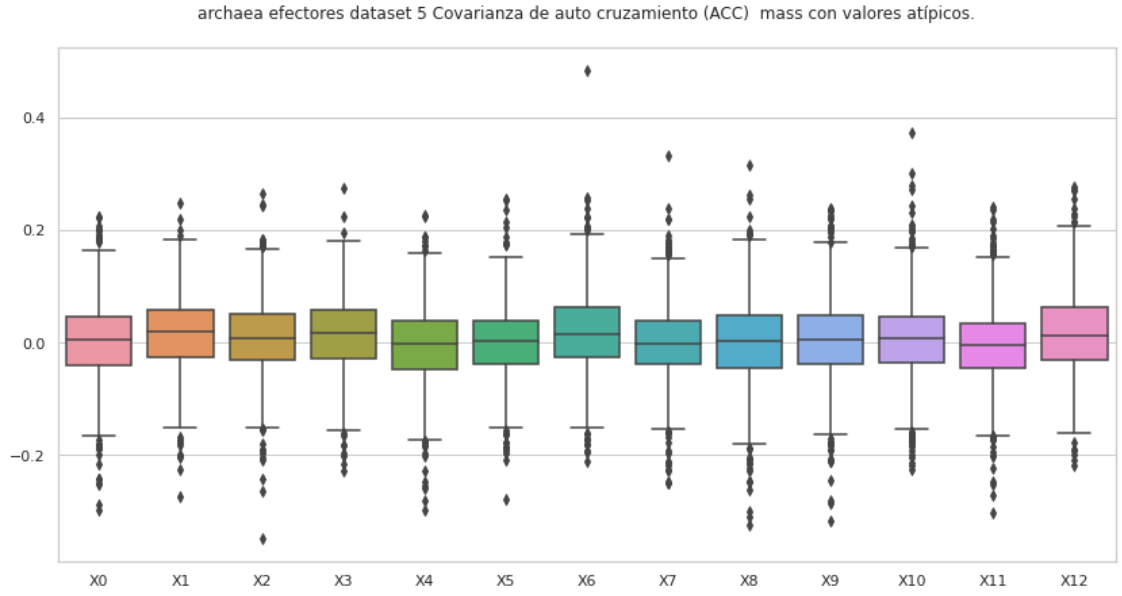
	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.003804	0.017101	0.009915	0.015535	-0.005843	

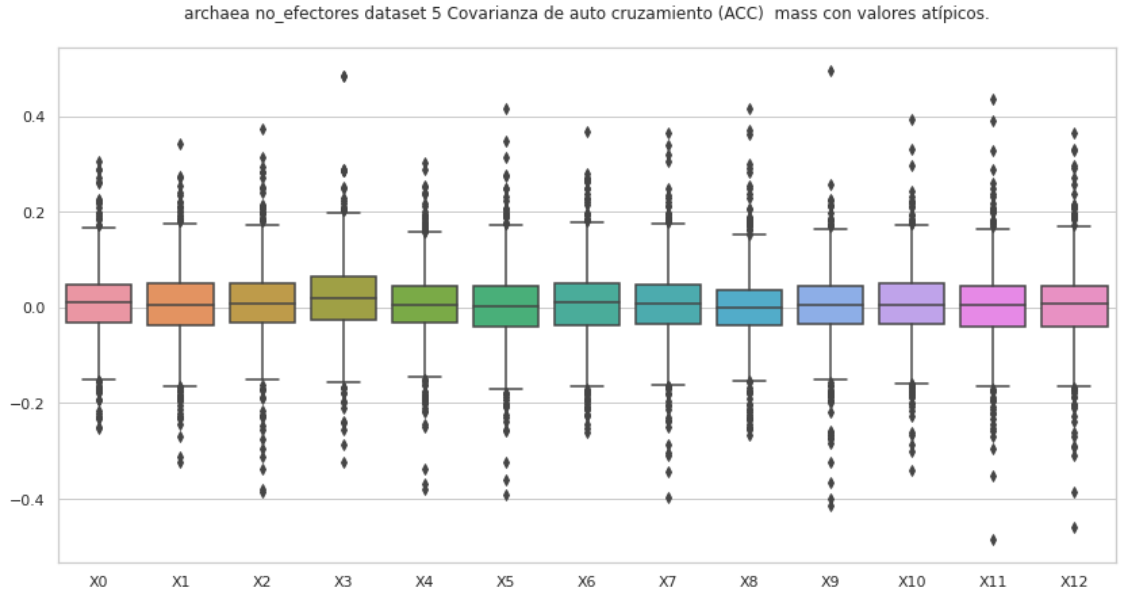
```
std       0.071714     0.067656     0.067399     0.068485     0.071362
min      -0.297766    -0.274422    -0.348021    -0.228345    -0.296786
25%      -0.040183    -0.025912    -0.030464    -0.028208    -0.046471
50%       0.004498     0.020300     0.007656     0.017660    -0.002226
75%       0.047194     0.058008     0.049777     0.058943     0.037974
max       0.224149     0.248910     0.263949     0.273984     0.226649

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean     -0.000485     0.018395    -0.001129    -0.000185     0.004682
std       0.066256     0.070279     0.064823     0.077263     0.073817
min      -0.277328    -0.211212    -0.249870    -0.324798    -0.317040
25%      -0.038771    -0.025785    -0.039004    -0.044226    -0.037979
50%       0.001685     0.014979    -0.001207     0.002766     0.004945
75%       0.038973     0.063059     0.038027     0.047555     0.048852
max       0.256319     0.483294     0.331801     0.315391     0.237331

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.007547    -0.004682     0.016271
std       0.071549     0.068713     0.071882
min      -0.225828    -0.302639    -0.218270
25%      -0.034762    -0.045027    -0.031879
50%       0.006970    -0.004149     0.011628
75%       0.046609     0.034862     0.063822
max       0.373070     0.239947     0.276165



no_efectores
Covarianza de auto cruzamiento (ACC)  mass no_efectores archaea dataset 5, con
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.289354  0.163037  0.143272  0.482456  0.211624  0.074577 -0.073506
1    0.080026 -0.032431  0.103822  0.149245 -0.060713  0.059636  0.194522
2    0.060699  0.011629  0.081504  0.029490  0.062128  0.055391  0.011149
3   -0.087508  0.040694  0.056822  0.065032  0.100864  0.065769 -0.016954
4    0.037192 -0.019966 -0.050050 -0.042743 -0.061527  0.031641  0.000363
..        …       …       …       …       …       …       …
995  0.023676  0.058927 -0.040920  0.017641  0.032638 -0.001646  0.010442
996 -0.012135  0.004585 -0.003821 -0.092199  0.016003 -0.126202  0.108575
997  0.020419 -0.194169 -0.006013 -0.018656 -0.010752  0.095136  0.079212
998 -0.130746  0.105529 -0.007510  0.123763  0.085111  0.190545 -0.040033
999 -0.029587 -0.001096 -0.034820 -0.131356  0.005997  0.053616  0.000204

           X7        X8        X9       X10       X11       X12           X13
```

```
0    0.140005  0.184886  0.164242  0.156792 -0.042461  0.148923  no_efectores
1    0.063453  0.030572  0.050420  0.039967  0.130524 -0.043992  no_efectores
2   -0.005342 -0.010882  0.014255 -0.024583  0.009951 -0.056087  no_efectores
3    0.060539 -0.016741 -0.057371  0.088469  0.032080 -0.098137  no_efectores
4    0.100639 -0.010084 -0.021324 -0.012185 -0.011518  0.016712  no_efectores
..        …       …       …       …       …       …           …
995  0.003137 -0.037532  0.044270 -0.045214  0.037449  0.058529  no_efectores
996 -0.122039  0.014260  0.150582 -0.065986 -0.074404 -0.170801  no_efectores
997 -0.154020 -0.040941  0.116862  0.003368  0.087313  0.095176  no_efectores
998  0.144520  0.002247  0.034367  0.049143 -0.052735  0.104162  no_efectores
999  0.044638  0.129188 -0.009300  0.062938 -0.082059  0.008649  no_efectores

[1000 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) mass no_efectores archaea dataset 5, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.009303     0.006101     0.009699     0.020745     0.005347
std       0.072171     0.076712     0.076524     0.079435     0.075353
min      -0.252802    -0.323394    -0.384568    -0.323729    -0.380453
25%      -0.031226    -0.036459    -0.032151    -0.025723    -0.032646
50%       0.011270     0.004396     0.007019     0.020308     0.005784
75%       0.048339     0.049260     0.050482     0.065140     0.043463
max       0.304973     0.340848     0.373340     0.484579     0.303703

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.001541     0.006987     0.006023     0.000675     0.003679
std       0.076957     0.076728     0.075971     0.074469     0.075889
min      -0.390313    -0.261885    -0.395905    -0.267932    -0.412624
25%      -0.041482    -0.037293    -0.035318    -0.038542    -0.034946
50%       0.002523     0.009678     0.007278    -0.000649     0.005875
75%       0.044635     0.049779     0.048846     0.037540     0.045462
max       0.417063     0.368375     0.364827     0.416948     0.494751

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.006632     0.004022     0.003559
std       0.074729     0.078896     0.078142
min      -0.339390    -0.484584    -0.458479
25%      -0.033890    -0.039184    -0.039058
50%       0.005357     0.006268     0.007598
75%       0.049245     0.044334     0.045568
max       0.394096     0.436612     0.364752
```


archaea efectores dataset 5 Covarianza de auto cruzamiento (ACC) mass con valores atípicos.
0.4
0.2
0.0
−0.2
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12


archaea no_efectores dataset 5 Covarianza de auto cruzamiento (ACC) mass con valores atípicos.
0.4
0.2
0.0
−0.2
−0.4
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12

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

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

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

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

    if etiq == "efectores":
        df=ACC_mass_efec

    if etiq == "no_efectores":
        df=ACC_mass_no_efec

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

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

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

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

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

```
Covarianza de auto cruzamiento (ACC)  mass efectores archaea dataset 5, sin
valores atípicos.
Valores del documento csv.
```

```
           X0        X1        X2        X3        X4        X5        X6  \
0   -0.015888  0.060600  0.021128  0.175120 -0.042455 -0.025333  0.072468
1   -0.112544  0.093547 -0.008473  0.003300  0.110713 -0.091982  0.096355
2    0.068264  0.039092 -0.007323  0.004073 -0.012690 -0.004975  0.034127
3    0.049046 -0.045433 -0.029467 -0.085142 -0.118806 -0.017147  0.039711
4   -0.021547 -0.096534  0.041258  0.115039 -0.078061  0.006559 -0.044745
..        …         …         …         …         …         …
995  0.064766 -0.017841 -0.070164  0.147441  0.093923  0.095305  0.027736
996 -0.078849  0.073196 -0.113261  0.059961 -0.005648  0.015498  0.041653
997  0.023128  0.075865  0.003858  0.013955  0.027401  0.015069 -0.017081
998 -0.053129  0.059062 -0.060425 -0.098835  0.054804  0.027718  0.089254
999  0.160396  0.075309 -0.005707 -0.042872  0.024529  0.078772  0.086375

           X7        X8        X9       X10       X11       X12        X13
0    0.085899 -0.077145 -0.002226  0.089885 -0.024457 -0.081080  efectores
1   -0.016797  0.131995 -0.033148 -0.055136  0.053453 -0.145517  efectores
2   -0.039125  0.068679  0.140358  0.174350  0.001696  0.050186  efectores
3    0.051694 -0.007213  0.040958  0.003637  0.054353 -0.017675  efectores
4    0.018083  0.074150 -0.106667 -0.074817 -0.044245  0.106460  efectores
..        …         …         …         …         …         …
995  0.107874  0.106333  0.065946  0.001640  0.000423  0.027021  efectores
996  0.106638 -0.143074 -0.051620 -0.059269  0.058394  0.005048  efectores
997 -0.016367  0.000739  0.055297  0.048400  0.013840 -0.022282  efectores
998  0.134793  0.008506  0.047533 -0.046301  0.046333 -0.028229  efectores
999 -0.086757 -0.008986 -0.017921 -0.000169  0.043513  0.092020  efectores

[918 rows x 14 columns]
```

```
Covarianza de auto cruzamiento (ACC)  mass efectores archaea dataset 5, sin
valores atípicos.
Estadísticas.
```

```
                X0          X1          X2          X3          X4          X5  \
count  918.000000  918.000000  918.000000  918.000000  918.000000  918.000000
mean     0.005498    0.017343    0.011070    0.014501   -0.004291   -0.000599
std      0.066231    0.061997    0.061403    0.064838    0.065694    0.062467
min     -0.197952   -0.181738   -0.180701   -0.181800   -0.202583   -0.196772
25%     -0.037614   -0.023255   -0.028703   -0.027927   -0.043555   -0.038290
50%      0.004636    0.020036    0.007788    0.016909   -0.001490    0.001622
75%      0.045186    0.056296    0.048231    0.056168    0.037128    0.038928
```

```
max      0.206642    0.182271    0.183038    0.181346    0.187027    0.176608

               X6          X7          X8          X9         X10         X11  \
count  918.000000  918.000000  918.000000  918.000000  918.000000  918.000000
mean     0.018055    0.000417    0.000743    0.005323    0.006467   -0.003534
std      0.063627    0.057526    0.070234    0.066766    0.064587    0.062622
min     -0.180613   -0.191788   -0.226893   -0.210567   -0.204867   -0.200621
25%     -0.023299   -0.036860   -0.042648   -0.034973   -0.033882   -0.042741
50%      0.014979   -0.000459    0.003199    0.004945    0.006280   -0.003589
75%      0.061229    0.037868    0.045555    0.048250    0.044779    0.034665
max      0.220763    0.190465    0.223667    0.207986    0.209027    0.190036

              X12
count  918.000000
mean     0.015370
std      0.067056
min     -0.189123
25%     -0.030571
50%      0.010610
75%      0.060261
max      0.213417
```

Covarianza de auto cruzamiento (ACC) mass no_efectores archaea dataset 5, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
1    0.080026 -0.032431  0.103822  0.149245 -0.060713  0.059636  0.194522
2    0.060699  0.011629  0.081504  0.029490  0.062128  0.055391  0.011149
3   -0.087508  0.040694  0.056822  0.065032  0.100864  0.065769 -0.016954
4    0.037192 -0.019966 -0.050050 -0.042743 -0.061527  0.031641  0.000363
5   -0.011557 -0.053341  0.008578  0.007840 -0.030196  0.064348  0.068808
..        …         …         …         …         …         …
995  0.023676  0.058927 -0.040920  0.017641  0.032638 -0.001646  0.010442
996 -0.012135  0.004585 -0.003821 -0.092199  0.016003 -0.126202  0.108575
997  0.020419 -0.194169 -0.006013 -0.018656 -0.010752  0.095136  0.079212
998 -0.130746  0.105529 -0.007510  0.123763  0.085111  0.190545 -0.040033
999 -0.029587 -0.001096 -0.034820 -0.131356  0.005997  0.053616  0.000204

           X7        X8        X9       X10       X11       X12           X13
1    0.063453  0.030572  0.050420  0.039967  0.130524 -0.043992  no_efectores
2   -0.005342 -0.010882  0.014255 -0.024583  0.009951 -0.056087  no_efectores
3    0.060539 -0.016741 -0.057371  0.088469  0.032080 -0.098137  no_efectores
4    0.100639 -0.010084 -0.021324 -0.012185 -0.011518  0.016712  no_efectores
5   -0.057018  0.045682 -0.013058  0.043817 -0.087080 -0.029315  no_efectores
..        …         …         …         …         …         …           …
```

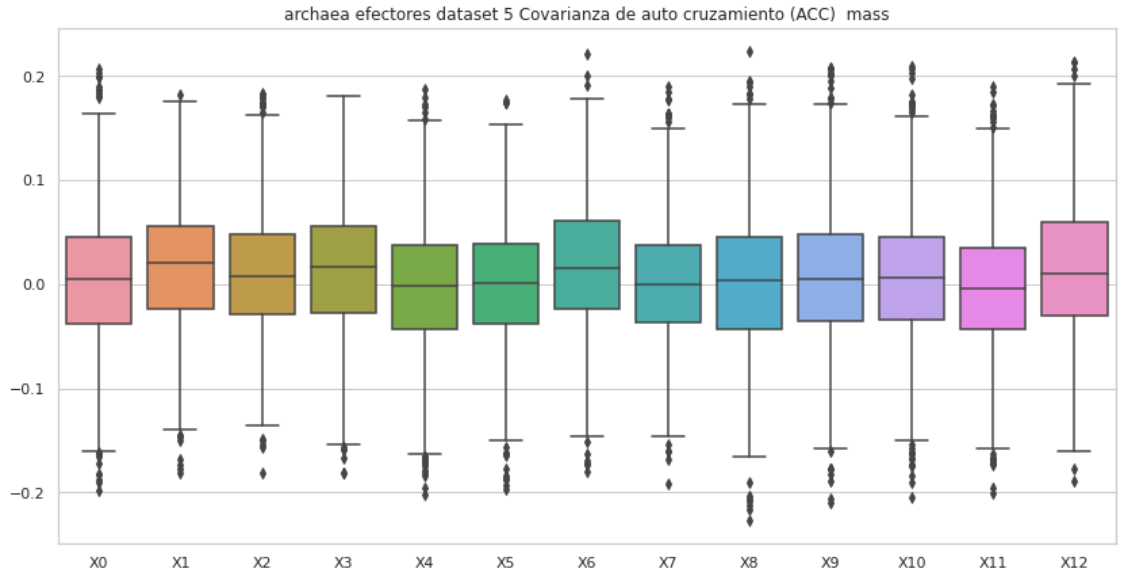
```
995  0.003137 -0.037532  0.044270 -0.045214  0.037449  0.058529  no_efectores
996 -0.122039  0.014260  0.150582 -0.065986 -0.074404 -0.170801  no_efectores
997 -0.154020 -0.040941  0.116862  0.003368  0.087313  0.095176  no_efectores
998  0.144520  0.002247  0.034367  0.049143 -0.052735  0.104162  no_efectores
999  0.044638  0.129188 -0.009300  0.062938 -0.082059  0.008649  no_efectores

[899 rows x 14 columns]
```

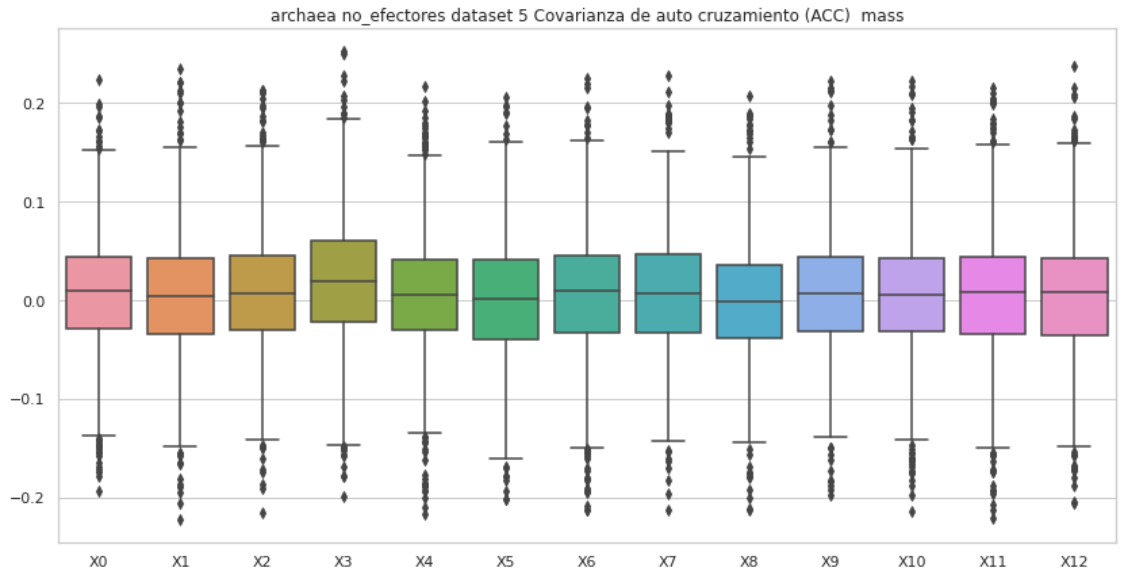
Covarianza de auto cruzamiento (ACC) mass no_efectores archaea dataset 5, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5
count	899.000000	899.000000	899.000000	899.000000	899.000000	899.000000
mean	0.007688	0.005476	0.009131	0.019578	0.005748	0.000622
std	0.061818	0.065794	0.062168	0.066644	0.064067	0.063044
min	-0.192789	-0.222121	-0.215665	-0.198480	-0.216087	-0.201870
25%	-0.029357	-0.034697	-0.030655	-0.022107	-0.030624	-0.039753
50%	0.009837	0.003701	0.006898	0.019938	0.005997	0.000920
75%	0.043792	0.042504	0.045933	0.060601	0.040652	0.041172
max	0.223433	0.234617	0.213011	0.252445	0.216769	0.206343

	X6	X7	X8	X9	X10	X11
count	899.000000	899.000000	899.000000	899.000000	899.000000	899.000000
mean	0.006968	0.007028	-0.000512	0.006317	0.006101	0.005842
std	0.064395	0.062872	0.061558	0.061861	0.063552	0.066386
min	-0.212791	-0.212695	-0.212024	-0.197885	-0.214471	-0.220524
25%	-0.032507	-0.032335	-0.037943	-0.031534	-0.031913	-0.034181
50%	0.009868	0.007402	-0.000628	0.006507	0.005356	0.007913
75%	0.045570	0.046544	0.036283	0.043886	0.043329	0.043720
max	0.224926	0.227464	0.206776	0.222488	0.222676	0.216188

	X12
count	899.000000
mean	0.004298
std	0.063408
min	-0.205368
25%	-0.035561
50%	0.007908
75%	0.042954
max	0.236953





8 Covarianza de auto cruzamiento (ACC) hidro

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

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

    if etiq == "efectores":
        df=ACC_hidro_efec

    if etiq == "no_efectores":
        df=ACC_hidro_no_efec

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

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

```
efectores
Covarianza de auto cruzamiento (ACC)  hidro efectores archaea dataset 5, con
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.086808 -0.060071  0.106535  0.005706 -0.143475  0.044461  0.039605
1   -0.029296 -0.091842  0.029109  0.032544  0.151887 -0.117773  0.067374
2    0.023214 -0.097114  0.098091 -0.133866 -0.106998 -0.121658 -0.134333
3   -0.019141 -0.015334 -0.003481  0.238151 -0.029291  0.051587 -0.054556
4    0.087904  0.100339 -0.054993  0.171182 -0.178162  0.125066 -0.078625
..        …         …         …         …         …         …         …
995 -0.079801 -0.046038 -0.007581  0.092936 -0.123283 -0.027306 -0.009879
996  0.072190  0.001485  0.018378  0.027570 -0.015850 -0.008889 -0.004375
997  0.020305 -0.018772 -0.036150 -0.028519 -0.054142 -0.030809  0.085235
998  0.004783 -0.029647 -0.092031 -0.016957 -0.068959  0.006580  0.062737
999  0.007136  0.085062  0.164691  0.043048  0.065152  0.045562  0.130569

           X7        X8        X9       X10       X11       X12        X13
0   -0.093423  0.009724 -0.044214  0.191778  0.035058 -0.035203  efectores
1   -0.065668 -0.140462  0.057068  0.067026 -0.079233 -0.200423  efectores
2    0.091233  0.021591  0.144286  0.041865  0.061708  0.032499  efectores
```

```
3    0.029505  0.081400  0.028634 -0.043219 -0.136256 -0.202603  efectores
4    0.161906 -0.156916  0.129145 -0.167105  0.149917 -0.263818  efectores
..        …       …        …        …        …        …        …
995  0.069261  0.035572 -0.020354  0.130227  0.002537 -0.046422  efectores
996 -0.051715 -0.051145 -0.050267 -0.013925 -0.037330 -0.032537  efectores
997  0.058028 -0.207407  0.047137 -0.018132 -0.038182 -0.062455  efectores
998  0.034322 -0.004179 -0.048017 -0.024147 -0.061812  0.044579  efectores
999  0.013573  0.064429  0.000135  0.107766  0.087521  0.038045  efectores

[1000 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro efectores archaea dataset 5, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.025817    -0.032317     0.046391     0.034698    -0.020385
std       0.088105     0.104395     0.083944     0.088245     0.096667
min      -0.330013    -0.413852    -0.452865    -0.312581    -0.403740
25%      -0.024092    -0.100775    -0.005657    -0.017144    -0.085183
50%       0.030315    -0.026638     0.038823     0.034563    -0.015722
75%       0.080276     0.043908     0.097856     0.084537     0.044155
max       0.263371     0.217667     0.342199     0.328033     0.273635

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean     -0.017234     0.026742     0.021825     0.003627    -0.000574
std       0.085045     0.092815     0.082757     0.090408     0.092816
min      -0.424210    -0.363646    -0.304760    -0.439255    -0.372742
25%      -0.068361    -0.029013    -0.026764    -0.045223    -0.047805
50%      -0.017404     0.020806     0.018234     0.007588    -0.000634
75%       0.035465     0.078084     0.072185     0.055114     0.048120
max       0.286502     0.341324     0.297549     0.396656     0.398532

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.020993     0.011501    -0.014767
std       0.083524     0.082918     0.088050
min      -0.328031    -0.281709    -0.394579
25%      -0.035890    -0.039533    -0.060769
50%       0.011580     0.003740    -0.007330
75%       0.074275     0.058935     0.038469
max       0.391802     0.346977     0.255782
```

no_efectores
Covarianza de auto cruzamiento (ACC) hidro no_efectores archaea dataset 5, con valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	-0.124282	-0.148604	-0.032309	0.045352	0.212934	0.055651	-0.197682	
1	-0.026290	-0.227239	0.100394	0.140682	-0.017932	0.201812	-0.057182	
2	0.076883	-0.027711	0.022071	0.043944	-0.008155	-0.033450	-0.014130	
3	0.018345	0.058591	0.067316	0.033776	-0.025988	0.052549	0.018251	
4	-0.000674	-0.000258	0.006347	-0.079995	-0.118047	0.139997	0.188834	
..	…	…	…	…	…	…	…	
995	-0.017965	-0.054763	-0.045027	-0.001475	0.027988	-0.130973	0.045384	
996	-0.067791	-0.255067	0.092690	0.015179	-0.124649	-0.091443	0.063882	
997	0.141293	0.054463	0.016128	0.118317	0.051663	-0.002910	0.119256	
998	0.041070	0.014635	0.207352	0.125073	0.072561	0.079200	0.088183	
999	-0.043664	-0.055854	-0.071759	-0.147123	-0.100819	0.037867	-0.134606	

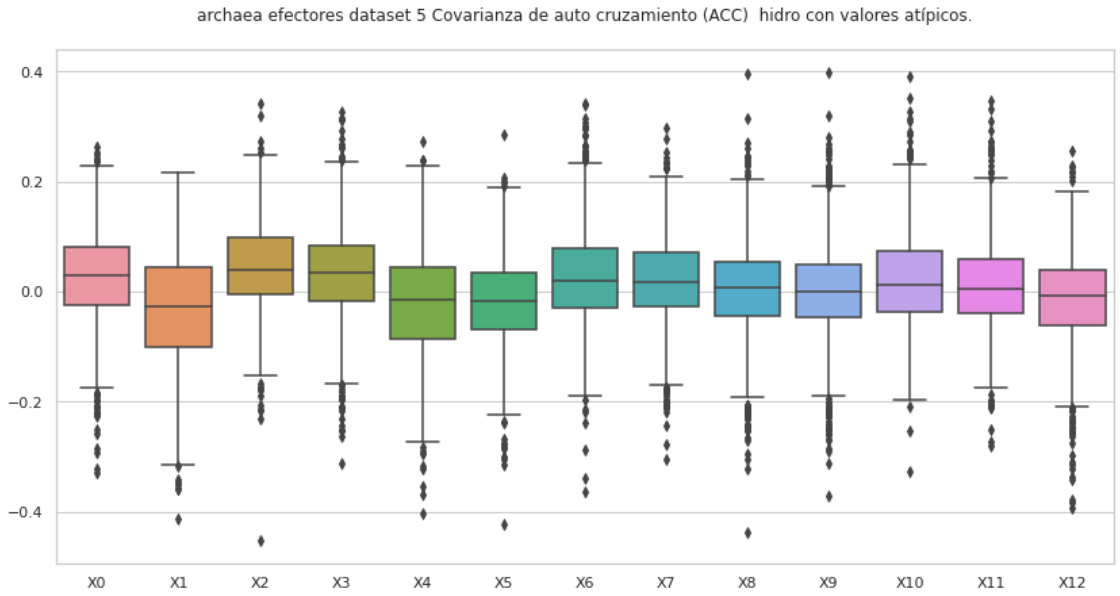
	X7	X8	X9	X10	X11	X12	X13
0	-0.033873	0.047319	0.021451	0.172673	0.120702	-0.096455	no_efectores
1	-0.071138	0.183800	0.038175	-0.179799	0.144034	0.229968	no_efectores
2	0.072723	0.070793	0.066749	0.034584	0.072273	0.007036	no_efectores
3	0.023037	0.001533	0.012719	-0.061052	-0.054452	0.028209	no_efectores
4	0.035025	0.012678	0.230195	0.113589	-0.119303	-0.158376	no_efectores
..	…	…	…	…	…	…	…
995	-0.071084	-0.004232	-0.017549	-0.005038	0.035719	-0.037355	no_efectores
996	0.174461	-0.211167	-0.099619	0.324525	0.145757	-0.358140	no_efectores
997	-0.057512	0.011019	0.068570	-0.020280	-0.028512	0.012837	no_efectores
998	0.009316	0.134542	-0.015618	-0.006093	0.099336	0.006704	no_efectores
999	0.113443	0.104022	-0.024204	-0.113040	-0.108053	-0.045550	no_efectores

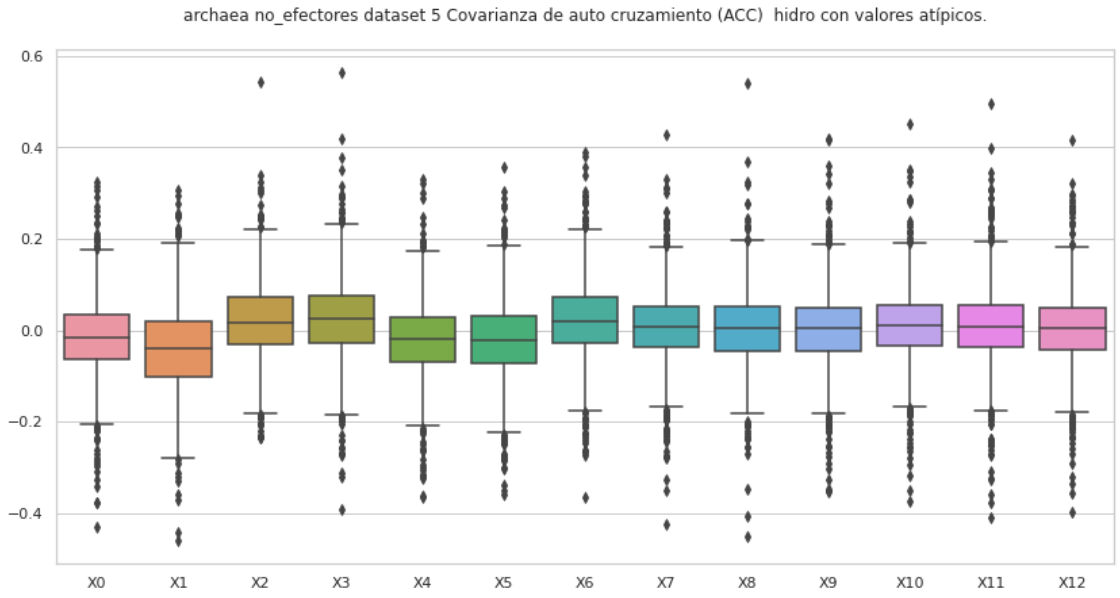
[1000 rows x 14 columns]

Covarianza de auto cruzamiento (ACC) hidro no_efectores archaea dataset 5, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	-0.016350	-0.040424	0.020434	0.024315	-0.020512	
std	0.088319	0.096060	0.084137	0.092703	0.086524	
min	-0.431319	-0.460339	-0.235007	-0.392575	-0.366497	
25%	-0.062955	-0.101566	-0.029942	-0.027771	-0.068151	
50%	-0.017466	-0.039037	0.016713	0.026338	-0.019418	
75%	0.033253	0.018171	0.072574	0.076932	0.029038	
max	0.323914	0.306280	0.542639	0.563864	0.328833	

```
                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean     -0.022604     0.022794     0.004864     0.002511     0.001323
std       0.088583     0.091958     0.086458     0.084996     0.087823
min      -0.358318    -0.365890    -0.425066    -0.451580    -0.353629
25%      -0.072507    -0.027481    -0.037123    -0.046471    -0.044536
50%      -0.022082     0.020008     0.006952     0.003725     0.003899
75%       0.030141     0.073359     0.050984     0.050672     0.048766
max       0.356613     0.390115     0.426923     0.539498     0.420466

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.010204     0.008629     0.001522
std       0.084284     0.090167     0.086401
min      -0.375057    -0.409761    -0.399361
25%      -0.035223    -0.037129    -0.042680
50%       0.010518     0.007484     0.003289
75%       0.055365     0.055015     0.048364
max       0.452740     0.495208     0.416628
```





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 5, sin valores
atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.086808 -0.060071  0.106535  0.005706 -0.143475  0.044461  0.039605
1   -0.029296 -0.091842  0.029109  0.032544  0.151887 -0.117773  0.067374
2    0.023214 -0.097114  0.098091 -0.133866 -0.106998 -0.121658 -0.134333
3   -0.019141 -0.015334 -0.003481  0.238151 -0.029291  0.051587 -0.054556
4    0.087904  0.100339 -0.054993  0.171182 -0.178162  0.125066 -0.078625
..        …         …         …         …         …         …         …
995 -0.079801 -0.046038 -0.007581  0.092936 -0.123283 -0.027306 -0.009879
996  0.072190  0.001485  0.018378  0.027570 -0.015850 -0.008889 -0.004375
997  0.020305 -0.018772 -0.036150 -0.028519 -0.054142 -0.030809  0.085235
998  0.004783 -0.029647 -0.092031 -0.016957 -0.068959  0.006580  0.062737
999  0.007136  0.085062  0.164691  0.043048  0.065152  0.045562  0.130569

           X7        X8        X9       X10       X11       X12        X13
0   -0.093423  0.009724 -0.044214  0.191778  0.035058 -0.035203  efectores
1   -0.065668 -0.140462  0.057068  0.067026 -0.079233 -0.200423  efectores
2    0.091233  0.021591  0.144286  0.041865  0.061708  0.032499  efectores
3    0.029505  0.081400  0.028634 -0.043219 -0.136256 -0.202603  efectores
4    0.161906 -0.156916  0.129145 -0.167105  0.149917 -0.263818  efectores
..        …         …         …         …         …         …         …
995  0.069261  0.035572 -0.020354  0.130227  0.002537 -0.046422  efectores
```

```
996 -0.051715 -0.051145 -0.050267 -0.013925 -0.037330 -0.032537  efectores
997  0.058028 -0.207407  0.047137 -0.018132 -0.038182 -0.062455  efectores
998  0.034322 -0.004179 -0.048017 -0.024147 -0.061812  0.044579  efectores
999  0.013573  0.064429  0.000135  0.107766  0.087521  0.038045  efectores

[926 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) efectores archaea dataset 5, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  926.000000  926.000000  926.000000  926.000000  926.000000  926.000000
mean     0.027855   -0.028402    0.045295    0.035079   -0.017269   -0.015568
std      0.082015    0.097212    0.075919    0.081070    0.090026    0.077395
min     -0.225486   -0.343170   -0.190016   -0.216598   -0.294305   -0.239883
25%     -0.021210   -0.096472   -0.005129   -0.014432   -0.079716   -0.063812
50%      0.031094   -0.023847    0.037423    0.034335   -0.013712   -0.017247
75%      0.078385    0.044064    0.095097    0.084425    0.043972    0.033809
max      0.263371    0.217667    0.261090    0.293489    0.239365    0.191628

                X6          X7          X8          X9         X10         X11  \
count  926.000000  926.000000  926.000000  926.000000  926.000000  926.000000
mean     0.025121    0.021621    0.002640   -0.002617    0.015230    0.008548
std      0.085498    0.076467    0.082150    0.084210    0.073416    0.074071
min     -0.218075   -0.218877   -0.264884   -0.271892   -0.209474   -0.211880
25%     -0.026619   -0.025629   -0.044339   -0.046862   -0.036471   -0.038630
50%      0.019353    0.017996    0.006526   -0.003564    0.008526    0.002709
75%      0.072905    0.071375    0.050999    0.044466    0.067007    0.052100
max      0.301099    0.243061    0.261464    0.269257    0.259561    0.255089

               X12
count  926.000000
mean    -0.012862
std      0.079673
min     -0.274901
25%     -0.059480
50%     -0.006272
75%      0.038270
max      0.229786
```

no_efectores
Covarianza de auto cruzamiento (ACC) no_efectores archaea dataset 5, sin valores atípicos.
Valores del documento csv.

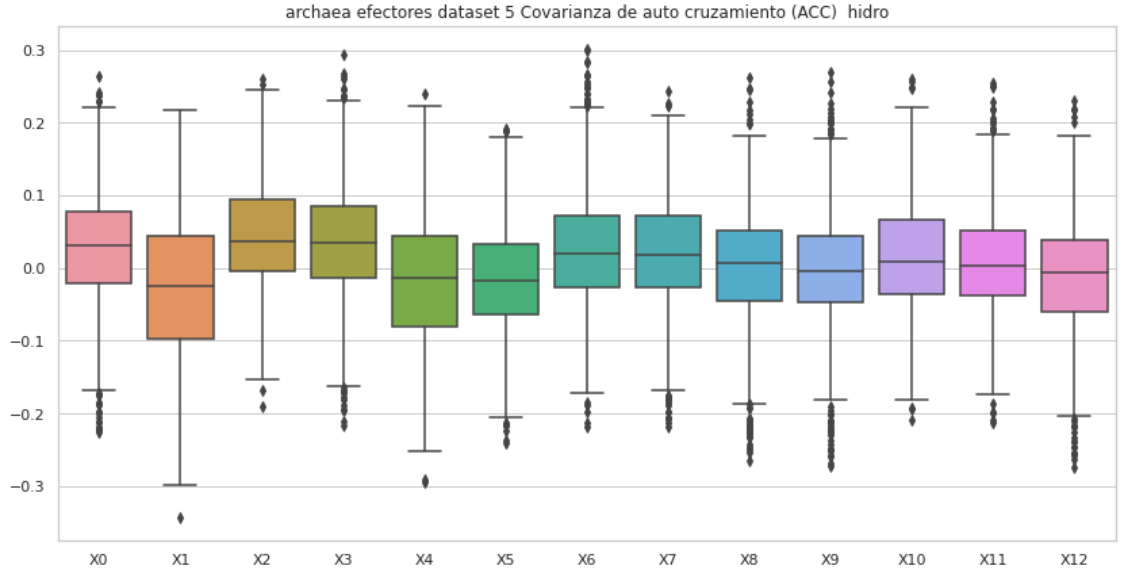
```
           X0        X1        X2        X3        X4        X5        X6  \
0   -0.124282 -0.148604 -0.032309  0.045352  0.212934  0.055651 -0.197682
1   -0.026290 -0.227239  0.100394  0.140682 -0.017932  0.201812 -0.057182
2    0.076883 -0.027711  0.022071  0.043944 -0.008155 -0.033450 -0.014130
3    0.018345  0.058591  0.067316  0.033776 -0.025988  0.052549  0.018251
4   -0.000674 -0.000258  0.006347 -0.079995 -0.118047  0.139997  0.188834
..        …         …         …         …         …         …
994 -0.003860  0.110954  0.087545 -0.045162  0.088609 -0.039937 -0.032410
995 -0.017965 -0.054763 -0.045027 -0.001475  0.027988 -0.130973  0.045384
997  0.141293  0.054463  0.016128  0.118317  0.051663 -0.002910  0.119256
998  0.041070  0.014635  0.207352  0.125073  0.072561  0.079200  0.088183
999 -0.043664 -0.055854 -0.071759 -0.147123 -0.100819  0.037867 -0.134606

           X7        X8        X9       X10       X11       X12           X13
0   -0.033873  0.047319  0.021451  0.172673  0.120702 -0.096455  no_efectores
1   -0.071138  0.183800  0.038175 -0.179799  0.144034  0.229968  no_efectores
2    0.072723  0.070793  0.066749  0.034584  0.072273  0.007036  no_efectores
3    0.023037  0.001533  0.012719 -0.061052 -0.054452  0.028209  no_efectores
4    0.035025  0.012678  0.230195  0.113589 -0.119303 -0.158376  no_efectores
..        …         …         …         …         …         …
994  0.011201  0.021202  0.086909  0.052477 -0.070837  0.044939  no_efectores
995 -0.071084 -0.004232 -0.017549 -0.005038  0.035719 -0.037355  no_efectores
997 -0.057512  0.011019  0.068570 -0.020280 -0.028512  0.012837  no_efectores
998  0.009316  0.134542 -0.015618 -0.006093  0.099336  0.006704  no_efectores
999  0.113443  0.104022 -0.024204 -0.113040 -0.108053 -0.045550  no_efectores

[898 rows x 14 columns]
```

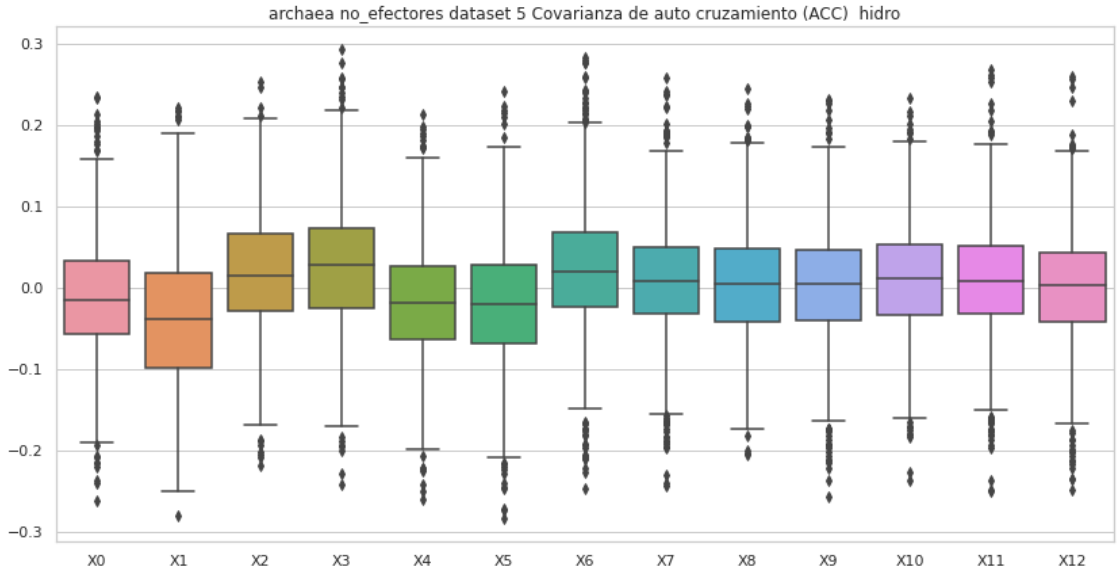
Covarianza de auto cruzamiento (ACC) no_efectores archaea dataset 5, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  898.000000  898.000000  898.000000  898.000000  898.000000  898.000000
mean    -0.012056   -0.039502    0.016811    0.024573   -0.018883   -0.021500
std      0.074197    0.083926    0.072640    0.079706    0.073886    0.076365
min     -0.262188   -0.280125   -0.219793   -0.242508   -0.261271   -0.284042
25%     -0.057154   -0.099019   -0.028686   -0.025239   -0.064015   -0.068602
50%     -0.015300   -0.039356    0.015251    0.027286   -0.018980   -0.020891
75%      0.032539    0.017177    0.065799    0.073241    0.026474    0.027787
max      0.234463    0.221730    0.252725    0.293019    0.212934    0.240889

                X6          X7          X8          X9         X10         X11  \
count  898.000000  898.000000  898.000000  898.000000  898.000000  898.000000
mean     0.022223    0.006885    0.003874    0.001529    0.010799    0.008242
std      0.080009    0.072433    0.069918    0.072751    0.068630    0.071237
```

min	-0.247044	-0.243497	-0.205486	-0.256590	-0.237447	-0.250560
25%	-0.023583	-0.032889	-0.041458	-0.040870	-0.033327	-0.032624
50%	0.019525	0.008230	0.004175	0.005166	0.010785	0.007731
75%	0.067135	0.050244	0.047460	0.045675	0.052344	0.052023
max	0.282363	0.258495	0.244082	0.231676	0.232871	0.267394

	X12
count	898.000000
mean	-0.000077
std	0.071722
min	-0.248372
25%	-0.041896
50%	0.002998
75%	0.043377
max	0.260087




archaea no_efectores dataset 5 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