ds2_archaea_limpieza_de_datos

January 19, 2021

Limpieza de datos

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

1 Declaración de variables

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

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

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

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

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

2 Composición de aminoácidos (AAC)

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

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

    if etiq == "efectores":
        df=AAC_efec

    if etiq == "no_efectores":
        df=AAC_no_efec

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

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

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

         X0      X1     X2      X3     X4      X5     X6      X7     X8  \
0     3.896   3.896  2.597   4.545  0.649  14.286  2.597   5.195  0.000
1    12.444   8.889  1.778   8.444  0.000   3.556  1.333   9.333  3.111
2    12.871   7.426  0.990   2.475  0.990   3.960  0.495   9.406  1.980
3     2.479  10.744  2.479   4.959  3.306   9.917  2.479   8.264  0.826
4     7.092   4.965  6.028   9.220  0.709   7.447  2.837   3.901  1.773
..      …      …     …       …      …       …      …       …      …
495  16.140   4.912  0.702   2.456  0.000   2.807  1.404  10.877  0.351
496   0.000   1.905  3.810   5.714  0.000   9.524  3.810   3.810  2.857
497   8.048   7.746  2.113  10.614  0.252   6.439  2.264  12.274  2.414
498   3.226   5.806  2.581   9.032  0.645   9.032  3.226   4.516  0.645
499   3.518   4.020  3.015   5.025  0.000  10.050  3.015   7.538  4.020
```

```
         X9  …     X11    X12    X13    X14    X15    X16    X17    X18  \
0     3.896  …  12.338  3.896  3.247  5.195  2.597  5.844  0.000  1.948
1     1.778  …   0.000  0.889  4.889  4.444  4.000  4.000  2.222  2.667
2     0.990  …   0.990  0.495  6.931  7.921  5.446  4.455  2.475  2.970
3    10.744  …   5.785  0.826  0.826  4.132  4.959  2.479  1.653  4.132
4     6.383  …  10.993  3.191  5.319  0.709  7.092  3.191  1.064  5.319
..      …  …     …    …    …    …    …    …    …    …
495   2.807  …   0.351  1.053  4.561  3.860  4.561  7.368  1.404  3.860
496   8.571  …  15.238  2.857  3.810  0.952  6.667  3.810  0.000  5.714
497   2.716  …   0.855  1.006  2.918  4.829  5.835  9.054  1.459  4.879
498   7.742  …  10.968  5.161  2.581  1.935  5.806  6.452  0.000  3.226
499   7.538  …  10.050  1.508  3.518  5.528  4.020  4.020  0.000  4.020

        X19         X20
0     7.792   efectores
1    13.333   efectores
2    11.386   efectores
3     8.264   efectores
4     4.610   efectores
..      …         …
495  15.088   efectores
496   6.667   efectores
497   6.942   efectores
498   5.806   efectores
499  11.055   efectores

[500 rows x 21 columns]
```

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

```
                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     9.293928    5.879204    2.827724    5.714486    0.654948    7.017696
std      4.577502    2.596324    2.225907    2.692226    0.873107    3.846546
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      5.705000    4.190250    1.186500    3.532750    0.000000    3.691000
50%      8.765000    5.538500    2.288000    5.197500    0.381000    6.902000
75%     12.082000    7.391250    4.039250    7.503750    0.975750    9.542000
max     21.569000   15.000000   11.585000   14.433000    5.310000   21.084000

                X6          X7          X8          X9         X10         X11  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     2.310540    7.444660    1.813850    5.846092   11.387872    4.629216
std      1.618139    2.900599    1.271786    3.307801    3.293133    4.451381
```

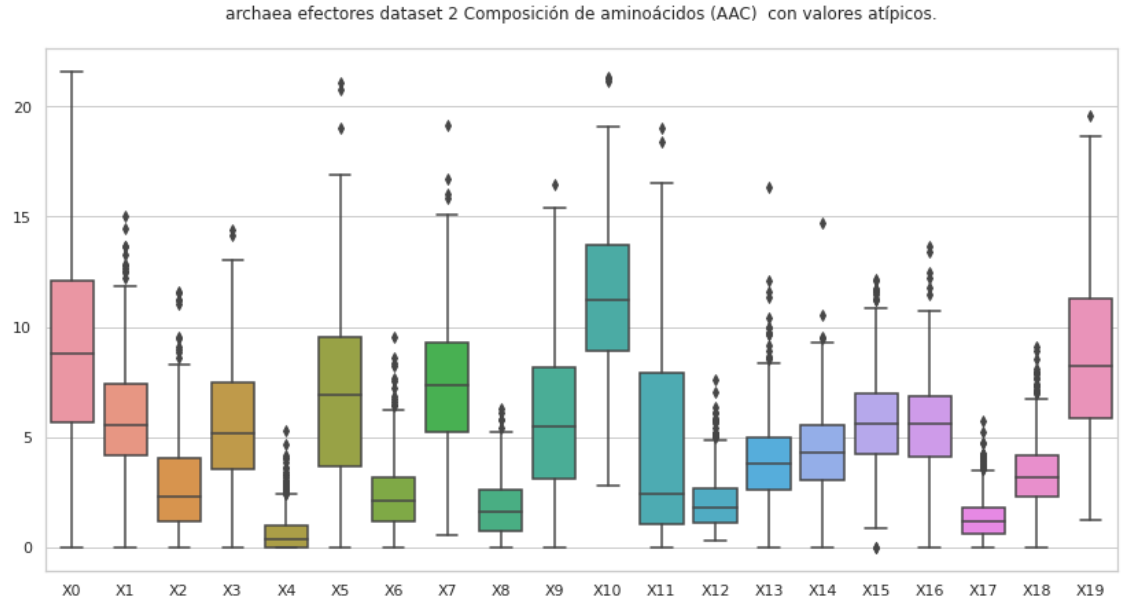
```
min      0.000000    0.562000    0.000000    0.000000    2.778000    0.000000
25%      1.154250    5.207250    0.778000    3.109000    8.942500    1.058000
50%      2.105500    7.365000    1.649000    5.487500   11.223000    2.451000
75%      3.200000    9.291750    2.608750    8.152000   13.732000    7.933000
max      9.559000   19.171000    6.294000   16.471000   21.306000   19.048000

              X12         X13         X14         X15         X16         X17  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     2.091384    3.955648    4.348774    5.736496    5.593642    1.330784
std      1.257811    2.048707    1.925857    2.192748    2.123232    1.028223
min      0.309000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      1.149000    2.608250    3.038250    4.245750    4.097750    0.631000
50%      1.789000    3.776500    4.316500    5.620000    5.600000    1.205000
75%      2.670000    4.982250    5.556000    6.965000    6.873000    1.790000
max      7.634000   16.327000   14.695000   12.174000   13.661000    5.714000

              X18         X19
count  500.000000  500.000000
mean     3.339754    8.783204
std      1.552078    3.590057
min      0.000000    1.250000
25%      2.296250    5.863000
50%      3.191000    8.206000
75%      4.173250   11.301250
max      9.091000   19.585000



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

         X0      X1     X2      X3     X4      X5     X6      X7     X8  \
0     8.171   6.420  3.307   9.339  0.584   9.339  1.556   9.144  2.140
1     0.000  18.750  2.083  12.500  2.083   4.167  4.167   4.167  0.000
2     6.870   4.580  0.763  10.687  1.527  12.214  4.580   7.634  3.053
3     9.392   6.906  1.657   7.735  0.552   9.945  1.934  11.326  1.105
4     7.372   4.167  6.410   4.487  0.321   8.013  0.962   5.769  2.885
..      …       …      …       …      …       …      …       …      …
495   4.478  10.448  0.746   9.701  0.000  16.418  2.985   9.701  2.985
496  11.765   8.333  2.451  11.275  0.980   7.353  1.471   9.314  2.941
497   7.266   9.689  1.730   7.266  1.384   9.343  2.768   7.958  1.384
498   6.623   9.272  1.987  10.596  1.325   6.623  0.662   4.636  6.623
499   6.092   4.606  2.526  12.333  0.149  11.887  2.972   8.915  1.783

        X9  …    X11    X12    X13    X14    X15     X16    X17    X18  \
0    2.335  …  1.556  1.556  3.891  6.420  3.891   5.058  2.140  2.918
```

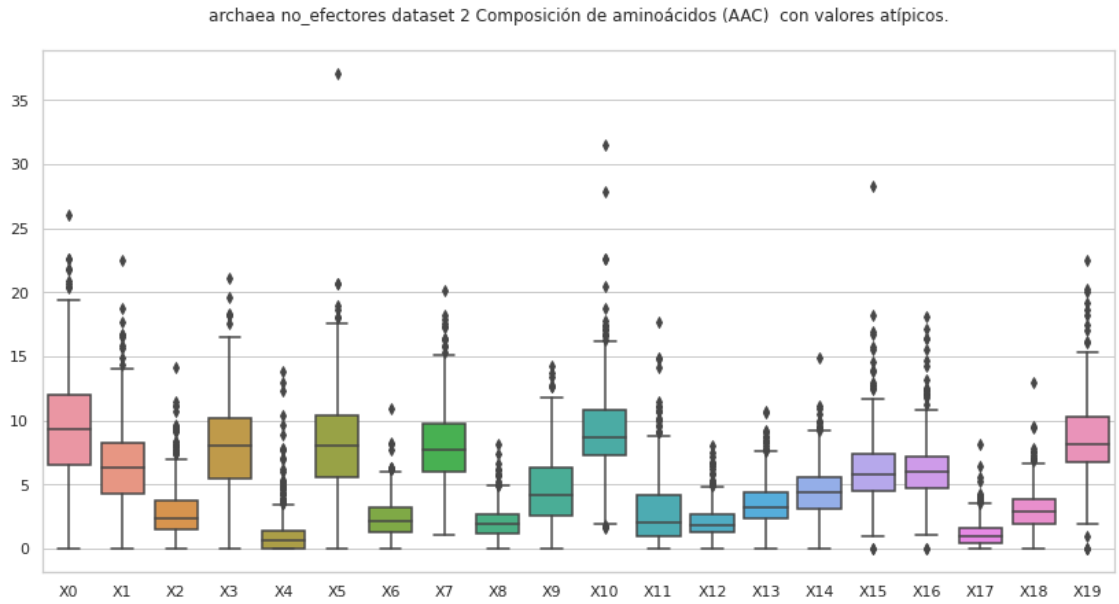
```
1    4.167  …  2.083  6.250  8.333  2.083  2.083  12.500  2.083  0.000
2    1.527  …  1.527  3.053  4.580  5.344  2.290   9.160  2.290  3.053
3    4.696  …  1.657  3.039  3.039  6.077  3.867   7.459  0.552  1.934
4    8.654  …  6.410  3.526  1.923  3.205  6.090   7.051  0.000  2.564
..     …  …    …    …    …    …    …     …    …    …
495  0.746  …  1.493  2.239  1.493  2.239  6.716   5.970  1.493  2.985
496  0.000  …  1.471  1.471  1.961  2.451  8.824   7.843  0.490  3.431
497  7.612  …  5.536  2.076  2.076  3.114  5.190   4.844  0.346  1.730
498  1.325  …  1.325  1.987  3.311  7.285  7.947   7.285  2.649  2.649
499  6.241  …  2.972  3.120  2.229  2.972  5.795   6.241  0.000  0.743

        X19            X20
0    10.117   no_efectores
1     4.167   no_efectores
2     6.107   no_efectores
3     9.945   no_efectores
4    10.897   no_efectores
..      …            …
495   9.701   no_efectores
496   9.804   no_efectores
497   6.920   no_efectores
498   7.947   no_efectores
499  10.550   no_efectores

[500 rows x 21 columns]
```

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

```
                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     9.491770    6.455726    2.808270    7.858204    1.147068    8.139134
std      4.097394    3.175598    2.088261    3.483814    1.708688    3.968479
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      6.547250    4.293750    1.477500    5.479000    0.000000    5.592500
50%      9.268500    6.357500    2.402500    8.055000    0.669000    8.061000
75%     11.949500    8.237000    3.706000   10.194000    1.395250   10.409750
max     26.000000   22.449000   14.101000   21.127000   13.846000   37.037000

                X6          X7          X8          X9         X10         X11  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     2.348506    7.981860    1.972046    4.680960    9.199484    2.930212
std      1.568329    2.883158    1.259091    2.861257    3.430249    2.756998
min      0.000000    1.020000    0.000000    0.000000    1.613000    0.000000
25%      1.275750    6.033250    1.174500    2.599750    7.242250    0.955750
50%      2.140000    7.692000    1.887000    4.175500    8.696000    2.052000
```

```
75%      3.199750    9.719750    2.656750    6.341000   10.850000    4.137750
max     10.929000   20.093000    8.108000   14.211000   31.507000   17.647000

              X12         X13         X14         X15         X16         X17  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     2.104110    3.530652    4.453420    6.088602    6.152298    1.174124
std      1.254439    1.800986    1.973356    2.698442    2.439634    1.083980
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      1.253750    2.324000    3.144000    4.456750    4.669250    0.383000
50%      1.858500    3.198000    4.359000    5.817000    6.006500    0.962000
75%      2.691000    4.432500    5.561250    7.366500    7.211500    1.655250
max      8.000000   10.714000   14.894000   28.247000   18.056000    8.163000

              X18         X19
count  500.000000  500.000000
mean     2.955378    8.528208
std      1.615106    3.106625
min      0.000000    0.000000
25%      1.896000    6.717500
50%      2.862500    8.148000
75%      3.846000   10.237500
max     12.963000   22.449000
```





2.1 Composición de aminoácidos (AAC), sin valores atípicos

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

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


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

    if etiq == "efectores":
        df=AAC_efec

    if etiq == "no_efectores":
        df=AAC_no_efec

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

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

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



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

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

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

         X0     X1     X2      X3     X4      X5     X6      X7     X8     X9  \
0     3.896  3.896  2.597   4.545  0.649  14.286  2.597   5.195  0.000  3.896
1    12.444  8.889  1.778   8.444  0.000   3.556  1.333   9.333  3.111  1.778
2    12.871  7.426  0.990   2.475  0.990   3.960  0.495   9.406  1.980  0.990
4     7.092  4.965  6.028   9.220  0.709   7.447  2.837   3.901  1.773  6.383
5    14.953  8.411  0.467   4.673  0.467   3.271  1.402   8.879  0.467  4.206
..      …      …      …       …      …       …      …       …      …      …
495  16.140  4.912  0.702   2.456  0.000   2.807  1.404  10.877  0.351  2.807
496   0.000  1.905  3.810   5.714  0.000   9.524  3.810   3.810  2.857  8.571
497   8.048  7.746  2.113  10.614  0.252   6.439  2.264  12.274  2.414  2.716
498   3.226  5.806  2.581   9.032  0.645   9.032  3.226   4.516  0.645  7.742
499   3.518  4.020  3.015   5.025  0.000  10.050  3.015   7.538  4.020  7.538

     …     X11    X12    X13    X14    X15    X16    X17    X18     X19  \
0    …  12.338  3.896  3.247  5.195  2.597  5.844  0.000  1.948   7.792
1    …   0.000  0.889  4.889  4.444  4.000  4.000  2.222  2.667  13.333
2    …   0.990  0.495  6.931  7.921  5.446  4.455  2.475  2.970  11.386
4    …  10.993  3.191  5.319  0.709  7.092  3.191  1.064  5.319   4.610
5    …   0.467  0.467  1.869  4.206  6.075  6.542  1.402  4.206  16.822
..   …      …      …      …      …      …      …      …      …
495  …   0.351  1.053  4.561  3.860  4.561  7.368  1.404  3.860  15.088
```

```
496  …  15.238  2.857  3.810  0.952  6.667  3.810  0.000  5.714   6.667
497  …   0.855  1.006  2.918  4.829  5.835  9.054  1.459  4.879   6.942
498  …  10.968  5.161  2.581  1.935  5.806  6.452  0.000  3.226   5.806
499  …  10.050  1.508  3.518  5.528  4.020  4.020  0.000  4.020  11.055

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

[435 rows x 21 columns]
```

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

```
                X0          X1          X2          X3          X4          X5  \
count   435.000000  435.000000  435.000000  435.000000  435.000000  435.000000
mean      9.722956    5.932566    2.693538    5.702720    0.540706    6.865543
std       4.567164    2.429107    1.996712    2.663769    0.684005    3.703827
min       0.000000    0.792000    0.000000    0.000000    0.000000    0.000000
25%       6.145500    4.275000    1.155000    3.531500    0.000000    3.554500
50%       9.402000    5.648000    2.264000    5.161000    0.307000    6.829000
75%      12.436500    7.451500    4.000000    7.515500    0.806000    9.384500
max      21.569000   13.636000    9.462000   13.043000    3.000000   16.883000

                X6          X7          X8          X9         X10         X11  \
count   435.000000  435.000000  435.000000  435.000000  435.000000  435.000000
mean      2.216251    7.615577    1.814568    5.633471   11.532814    4.372660
std       1.474527    2.743821    1.238896    3.191816    3.187079    4.338234
min       0.000000    0.752000    0.000000    0.000000    2.778000    0.000000
25%       1.137000    5.479000    0.770500    3.056500    9.146000    1.006500
50%       2.020000    7.547000    1.661000    5.147000   11.315000    2.006000
75%       3.156000    9.423000    2.639500    7.815500   13.809500    7.788500
max       6.849000   16.049000    5.435000   14.388000   21.127000   16.539000

               X12         X13         X14         X15         X16         X17  \
count   435.000000  435.000000  435.000000  435.000000  435.000000  435.000000
mean      2.024103    4.009457    4.351786    5.754457    5.599471    1.337593
```

std	1.156144	1.798559	1.785201	2.111423	1.999165	0.971973
min	0.309000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	1.143000	2.823500	3.133000	4.309500	4.146000	0.665500
50%	1.774000	3.874000	4.390000	5.642000	5.666000	1.227000
75%	2.632000	5.004500	5.542000	6.897000	6.863500	1.794000
max	5.851000	10.000000	9.494000	12.174000	11.795000	4.255000

	X18	X19
count	435.000000	435.000000
mean	3.297715	8.981922
std	1.422830	3.549856
min	0.000000	2.564000
25%	2.326000	6.253500
50%	3.200000	8.527000
75%	4.105000	11.649000
max	7.895000	18.638000

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

	X0	X1	X2	X3	X4	X5	X6	X7	X8	\
0	8.171	6.420	3.307	9.339	0.584	9.339	1.556	9.144	2.140	
2	6.870	4.580	0.763	10.687	1.527	12.214	4.580	7.634	3.053	
3	9.392	6.906	1.657	7.735	0.552	9.945	1.934	11.326	1.105	
4	7.372	4.167	6.410	4.487	0.321	8.013	0.962	5.769	2.885	
5	12.150	4.984	1.869	10.280	0.312	9.346	2.181	4.984	1.869	
..	…	…	…	…	…	…	…	…	…	
494	12.760	7.812	1.302	10.156	1.042	9.375	2.083	9.115	1.823	
495	4.478	10.448	0.746	9.701	0.000	16.418	2.985	9.701	2.985	
496	11.765	8.333	2.451	11.275	0.980	7.353	1.471	9.314	2.941	
497	7.266	9.689	1.730	7.266	1.384	9.343	2.768	7.958	1.384	
499	6.092	4.606	2.526	12.333	0.149	11.887	2.972	8.915	1.783	

	X9	…	X11	X12	X13	X14	X15	X16	X17	X18	\
0	2.335	…	1.556	1.556	3.891	6.420	3.891	5.058	2.140	2.918	
2	1.527	…	1.527	3.053	4.580	5.344	2.290	9.160	2.290	3.053	
3	4.696	…	1.657	3.039	3.039	6.077	3.867	7.459	0.552	1.934	
4	8.654	…	6.410	3.526	1.923	3.205	6.090	7.051	0.000	2.564	
5	4.673	…	1.869	1.246	2.804	4.673	5.296	6.231	0.935	3.427	
..	…	…	…	…	…	…	…	…	…	…	
494	2.344	…	1.042	2.344	2.865	7.031	3.125	4.948	0.260	1.302	
495	0.746	…	1.493	2.239	1.493	2.239	6.716	5.970	1.493	2.985	
496	0.000	…	1.471	1.471	1.961	2.451	8.824	7.843	0.490	3.431	
497	7.612	…	5.536	2.076	2.076	3.114	5.190	4.844	0.346	1.730	

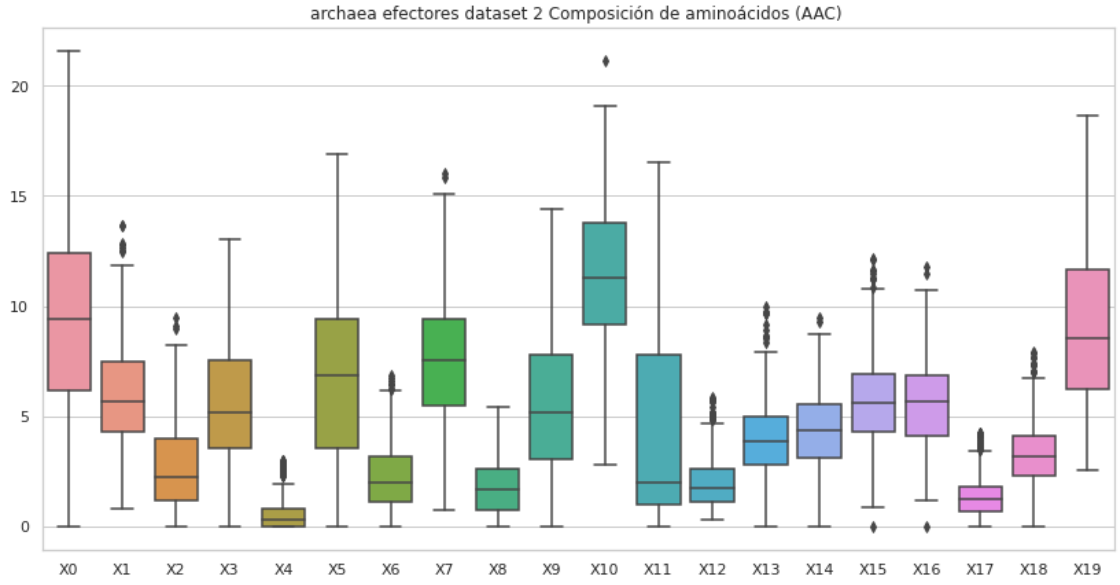
```
499  6.241  …  2.972  3.120  2.229  2.972  5.795  6.241  0.000  0.743

        X19           X20
0    10.117  no_efectores
2     6.107  no_efectores
3     9.945  no_efectores
4    10.897  no_efectores
5    12.773  no_efectores
..      …           …
494  10.677  no_efectores
495   9.701  no_efectores
496   9.804  no_efectores
497   6.920  no_efectores
499  10.550  no_efectores

[412 rows x 21 columns]
```

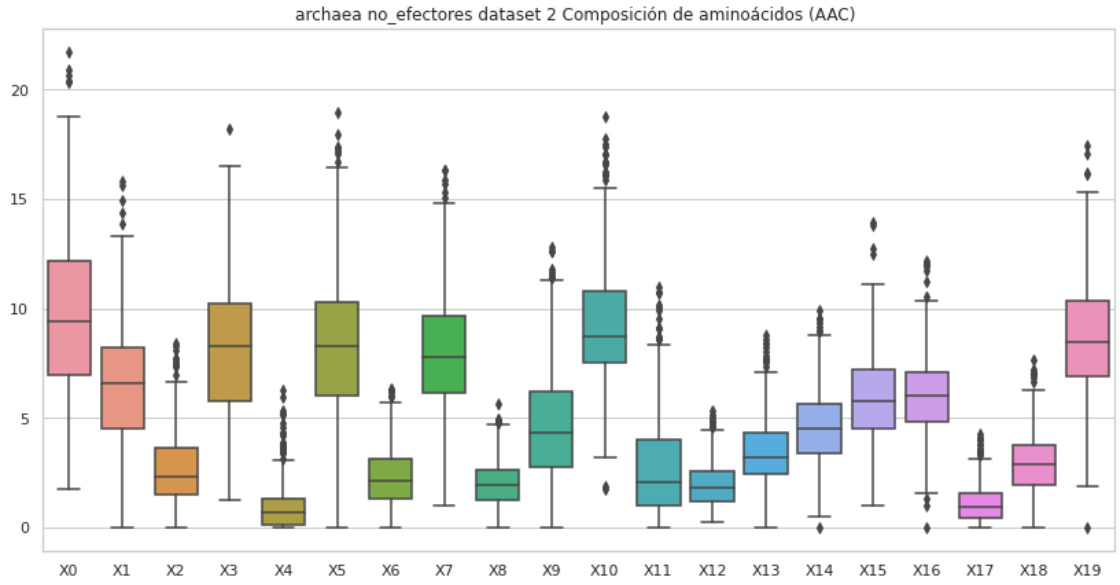
Composición de aminoácidos (AAC) no_efectores archaea dataset 2, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  412.000000  412.000000  412.000000  412.000000  412.000000  412.000000
mean     9.715522    6.532187    2.680687    8.077660    0.957631    8.211665
std      3.801601    2.677667    1.733083    3.065256    1.071693    3.389549
min      1.786000    0.000000    0.000000    1.250000    0.000000    0.000000
25%      6.991500    4.505750    1.483250    5.802750    0.148000    6.002750
50%      9.442500    6.572500    2.322500    8.286500    0.672000    8.300500
75%     12.197500    8.192500    3.617750   10.244750    1.321250   10.261250
max     21.707000   15.810000    8.402000   18.182000    6.250000   18.935000

                X6          X7          X8          X9         X10         X11  \
count  412.000000  412.000000  412.000000  412.000000  412.000000  412.000000
mean     2.317170    7.978854    1.996182    4.679539    9.217524    2.838684
std      1.409812    2.543895    1.075901    2.674265    2.884047    2.463372
min      0.000000    1.020000    0.000000    0.000000    1.786000    0.000000
25%      1.296750    6.156250    1.266000    2.779250    7.501500    1.010000
50%      2.150000    7.762000    1.921000    4.308500    8.729500    2.049000
75%      3.120750    9.636750    2.665750    6.205500   10.787250    4.012000
max      6.329000   16.340000    5.674000   12.780000   18.750000   10.976000

               X12         X13         X14         X15         X16         X17  \
count  412.000000  412.000000  412.000000  412.000000  412.000000  412.000000
mean     1.974243    3.525500    4.552578    5.942714    6.031058    1.112447
std      1.050911    1.622558    1.688466    2.125467    1.983677    0.923071
min      0.248000    0.000000    0.000000    1.000000    0.000000    0.000000
25%      1.210750    2.472000    3.407750    4.497500    4.804750    0.450750
```

50%	1.829000	3.190000	4.530000	5.803000	6.018500	0.952000
75%	2.553500	4.358500	5.623750	7.231000	7.091750	1.543750
max	5.303000	8.791000	9.934000	13.918000	12.150000	4.245000

	X18	X19
count	412.00000	412.000000
mean	2.93057	8.727604
std	1.37608	2.718518
min	0.00000	0.000000
25%	1.93275	6.892250
50%	2.85700	8.486000
75%	3.74175	10.381250
max	7.63400	17.419000





3 Composición de pseudo aminoácidos (PseAAC) hidro_mass

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

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

    if etiq == "efectores":
        df=PseAAC_hidro_mass_efec

    if etiq == "no_efectores":
        df=PseAAC_hidro_mass_no_efec

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

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

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.018802	0.003134	0.021936	0.068941	0.015668	0.025069	0.000000	
1	0.050137	0.000000	0.034021	0.014325	0.019697	0.037603	0.012534	
2	0.030174	0.002321	0.005803	0.009284	0.016248	0.022050	0.004642	
3	0.020946	0.027927	0.041891	0.083782	0.006982	0.069819	0.006982	
4	0.046757	0.004676	0.060784	0.049095	0.035068	0.025716	0.011689	
..	…	…	…	…	…	…	…	
495	0.022150	0.000000	0.003371	0.003852	0.006260	0.014927	0.000482	
496	0.000000	0.000000	0.048603	0.081006	0.032402	0.032402	0.024302	
497	0.027844	0.000870	0.036719	0.022275	0.010094	0.042462	0.008353	
498	0.022526	0.004505	0.063073	0.063073	0.018021	0.031536	0.004505	
499	0.020263	0.000000	0.028947	0.057893	0.020263	0.043420	0.023157	

	X7	X8	X9	…	X74	X75	X76	\
0	0.018802	0.059540	0.075208	…	0.009747	0.040938	0.012554	
1	0.007162	0.000000	0.051927	…	-0.004186	0.001079	-0.005614	
2	0.002321	0.002321	0.035977	…	0.009896	-0.005774	0.017936	
3	0.090764	0.048873	0.090764	…	-0.107809	-0.072461	0.033369	
4	0.042081	0.072474	0.053771	…	-0.024961	0.008235	-0.015420	
..	…	…	…	…	…	…	…	
495	0.003852	0.000482	0.021187	…	0.016570	0.004242	0.023243	
496	0.072905	0.129609	0.121509	…	0.125894	0.163324	0.023751	
497	0.009397	0.002958	0.025408	…	-0.002531	0.009924	0.028367	
498	0.054063	0.076589	0.081094	…	-0.052563	-0.007216	-0.003197	
499	0.043420	0.057893	0.049209	…	-0.010479	-0.021759	0.022948	

	X77	X78	X79	X80	X81	X82	X83
0	0.005021	0.030354	0.000408	0.001063	0.011334	0.011976	efectores
1	-0.002604	-0.036967	-0.000371	-0.013827	-0.006040	0.012629	efectores
2	0.004323	-0.005912	0.020768	-0.003046	-0.009587	0.025053	efectores
3	-0.071457	0.029871	0.030098	0.017191	-0.013744	-0.021230	efectores
4	0.030538	0.041747	-0.002484	-0.002353	-0.000388	0.014609	efectores
..	…	…	…	…	…	…	
495	0.024519	0.007799	0.012092	0.007423	0.001260	0.016617	efectores
496	0.027115	0.083765	-0.016294	-0.087844	-0.089283	-0.030714	efectores

```
497  0.001484  0.006192  0.019865 -0.003918  0.011222  0.024197  efectores
498 -0.033848 -0.004829  0.004526 -0.013021 -0.035399  0.023955  efectores
499  0.002436  0.031168 -0.013981 -0.010159  0.015840 -0.015495  efectores

[500 rows x 84 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.032590	0.003535	0.026133	0.035603	0.016455	0.027470	
std	0.023527	0.006052	0.021328	0.038635	0.014610	0.015387	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.001544	
25%	0.021783	0.000000	0.008644	0.007742	0.007379	0.017060	
50%	0.028925	0.001129	0.020677	0.025605	0.013027	0.024508	
75%	0.039848	0.004335	0.039907	0.055794	0.021217	0.033963	
max	0.448072	0.048019	0.199143	0.597429	0.149357	0.199143	

	X6	X7	X8	X9	…	X73	\
count	500.000000	500.000000	500.000000	500.000000	…	500.000000	
mean	0.008050	0.030560	0.027593	0.046846	…	0.012469	
std	0.009854	0.043803	0.043402	0.038141	…	0.019159	
min	0.000000	0.000000	0.000000	0.007573	…	-0.110300	
25%	0.002435	0.006982	0.002344	0.025731	…	0.003433	
50%	0.005997	0.019079	0.008860	0.038122	…	0.013679	
75%	0.010680	0.042243	0.046025	0.057416	…	0.023185	
max	0.149357	0.746787	0.647215	0.597429	…	0.102122	

	X74	X75	X76	X77	X78	X79	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.003005	0.006183	0.013740	0.003991	0.008452	0.013964	
std	0.033247	0.028784	0.021129	0.027884	0.022454	0.019624	
min	-0.286148	-0.290081	-0.178442	-0.169982	-0.106416	-0.091158	
25%	-0.007939	-0.002722	0.005082	-0.004884	-0.001376	0.003295	
50%	0.007784	0.003636	0.016221	0.008667	0.005215	0.015614	
75%	0.016460	0.013873	0.025354	0.017202	0.016361	0.024683	
max	0.159929	0.163324	0.129013	0.115635	0.142272	0.137885	

	X80	X81	X82
count	500.000000	500.000000	500.000000
mean	0.004453	0.009090	0.013698
std	0.041281	0.041616	0.020696
min	-0.572918	-0.721344	-0.114587
25%	-0.004505	-0.000699	0.004836
50%	0.009271	0.004885	0.015663

```
75%      0.018717    0.017434    0.023395
max      0.127265    0.182906    0.095891

[8 rows x 83 columns]
```

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.033981	0.002427	0.038836	0.038836	0.016182	0.038027	0.008900	
1	0.000000	0.007657	0.045944	0.015315	0.030630	0.015315	0.000000	
2	0.051882	0.011529	0.080706	0.092235	0.034588	0.057647	0.023059	
3	0.033394	0.001964	0.027501	0.035358	0.010804	0.040269	0.003929	
4	0.033079	0.001438	0.020135	0.035955	0.008629	0.025888	0.012944	
..	…	…	…	…	…	…	…	
495	0.013740	0.000000	0.029770	0.050379	0.004580	0.029770	0.009160	
496	0.034300	0.002858	0.032871	0.021438	0.005717	0.027154	0.008575	
497	0.041318	0.007870	0.041318	0.053123	0.011805	0.045253	0.007870	
498	0.038142	0.007628	0.061027	0.038142	0.019071	0.026699	0.038142	
499	0.016291	0.000397	0.032980	0.031787	0.005960	0.023841	0.004768	

	X7	X8	X9	…	X74	X75	X76	\
0	0.009709	0.006473	0.042072	…	-0.014284	-0.002084	0.034870	
1	0.015315	0.007657	0.030630	…	-0.001446	0.046274	-0.002606	
2	0.011529	0.011529	0.069176	…	0.001572	0.044636	-0.016734	
3	0.016697	0.005893	0.025537	…	-0.002892	0.001339	0.031850	
4	0.038832	0.028764	0.041708	…	-0.014898	-0.010583	0.025917	
..	…	…	…	…	…	…	…	
495	0.002290	0.004580	0.022900	…	0.044920	0.069844	-0.000645	
496	0.000000	0.004287	0.018579	…	0.011209	0.009484	0.025043	
497	0.043285	0.031480	0.066895	…	0.011328	0.026399	-0.003360	
498	0.007628	0.007628	0.045770	…	0.006161	0.026246	-0.008749	
499	0.016688	0.007947	0.021059	…	0.004720	0.017737	0.019724	

	X77	X78	X79	X80	X81	X82	X83
0	-0.006859	-0.004676	0.033110	0.003668	0.022699	0.024586	no_efectores
1	0.067924	0.021270	0.034588	-0.108856	-0.058032	0.003674	no_efectores
2	0.055115	0.013154	-0.032694	-0.041508	-0.079093	0.000345	no_efectores
3	-0.008283	0.008977	0.017159	-0.005442	-0.009005	0.014553	no_efectores
4	-0.022162	-0.010335	0.031039	0.004038	-0.004065	0.016320	no_efectores
..	…	…	…	…	…	…	
495	-0.010017	0.027299	0.005935	0.022390	0.021400	-0.006952	no_efectores
496	-0.003547	0.015188	0.028236	-0.000795	0.017267	0.004124	no_efectores
497	-0.037905	0.005517	-0.002842	-0.012005	-0.000488	-0.014719	no_efectores

```
498 -0.004114  0.006240 -0.016454  0.043103  0.066246  0.000460  no_efectores
499  0.002241  0.009167  0.019579 -0.010180  0.014170  0.015102  no_efectores

[500 rows x 84 columns]
```

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

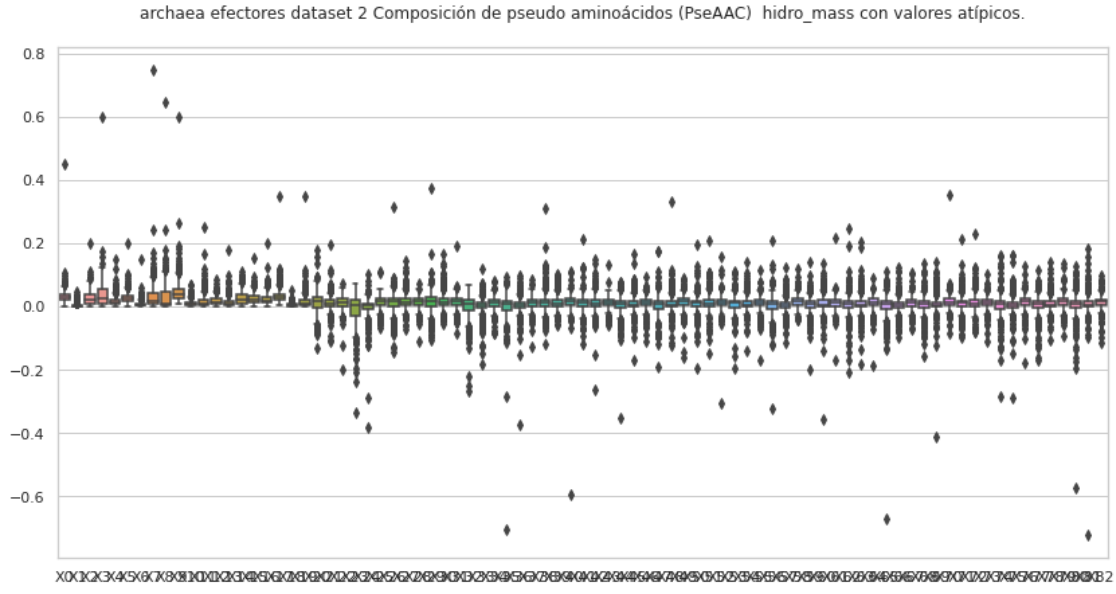
	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.034716	0.005131	0.032274	0.034456	0.014975	0.029775	
std	0.018408	0.008705	0.019576	0.023268	0.014338	0.013412	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.002789	
25%	0.023705	0.000000	0.017457	0.017244	0.007016	0.020616	
50%	0.032785	0.002368	0.030955	0.032145	0.011603	0.028084	
75%	0.041891	0.006231	0.043586	0.048033	0.018012	0.036942	
max	0.190911	0.090179	0.099241	0.160745	0.190911	0.119320	

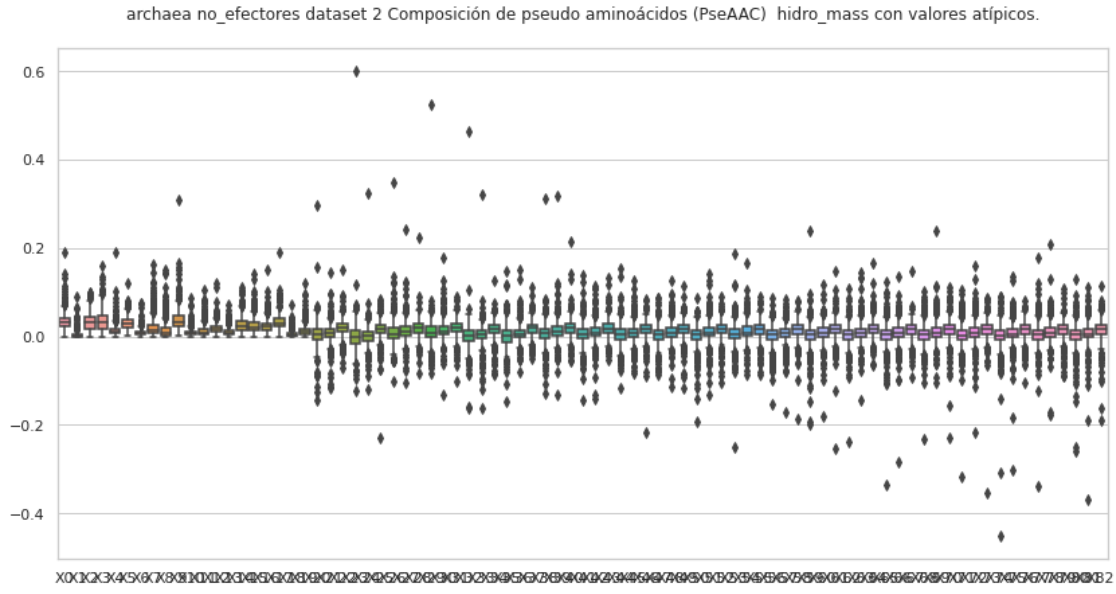
	X6	X7	X8	X9	…	X73	\
count	500.000000	500.000000	500.000000	500.000000	…	500.000000	
mean	0.008857	0.020679	0.014702	0.036909	…	0.014898	
std	0.008866	0.021114	0.020214	0.025185	…	0.024503	
min	0.000000	0.000000	0.000000	0.002791	…	-0.352940	
25%	0.003338	0.007541	0.002841	0.021949	…	0.005551	
50%	0.006743	0.014345	0.007669	0.031899	…	0.016701	
75%	0.011948	0.026216	0.018550	0.045971	…	0.026043	
max	0.075645	0.162704	0.151483	0.310231	…	0.086535	

	X74	X75	X76	X77	X78	X79	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	-0.000117	0.006993	0.014648	0.003551	0.009218	0.013765	
std	0.033634	0.026915	0.019466	0.028895	0.024495	0.020918	
min	-0.451472	-0.301669	-0.105628	-0.339675	-0.176253	-0.090161	
25%	-0.008562	-0.001532	0.005933	-0.006887	-0.001123	0.006035	
50%	0.002495	0.006183	0.017019	0.003910	0.006468	0.015912	
75%	0.013670	0.016175	0.026682	0.014718	0.018945	0.025436	
max	0.089870	0.136508	0.079800	0.178430	0.209086	0.108910	

	X80	X81	X82
count	500.000000	500.000000	500.000000
mean	0.001519	0.007133	0.014551
std	0.029673	0.028806	0.024260
min	-0.259147	-0.367877	-0.189447
25%	-0.007976	-0.002183	0.004567
50%	0.003700	0.005768	0.016986
75%	0.014043	0.018291	0.026910

```
max       0.130324     0.110042     0.113773

[8 rows x 83 columns]
```





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

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

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


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

    if etiq == "efectores":
        df=PseAAC_hidro_mass_efec

    if etiq == "no_efectores":
        df=PseAAC_hidro_mass_no_efec

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

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


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

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

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

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.018802	0.003134	0.021936	0.068941	0.015668	0.025069	0.000000	
1	0.050137	0.000000	0.034021	0.014325	0.019697	0.037603	0.012534	
2	0.030174	0.002321	0.005803	0.009284	0.016248	0.022050	0.004642	
4	0.046757	0.004676	0.060784	0.049095	0.035068	0.025716	0.011689	
5	0.031347	0.000980	0.009796	0.006857	0.003918	0.018612	0.000980	
..	…	…	…	…	…	…	…	
493	0.019288	0.009644	0.024110	0.038576	0.000000	0.014466	0.000000	
494	0.022838	0.000000	0.003915	0.005873	0.008483	0.014356	0.001305	
495	0.022150	0.000000	0.003371	0.003852	0.006260	0.014927	0.000482	
497	0.027844	0.000870	0.036719	0.022275	0.010094	0.042462	0.008353	
499	0.020263	0.000000	0.028947	0.057893	0.020263	0.043420	0.023157	

	X7	X8	X9	…	X74	X75	X76	\
0	0.018802	0.059540	0.075208	…	0.009747	0.040938	0.012554	
1	0.007162	0.000000	0.051927	…	-0.004186	0.001079	-0.005614	
2	0.002321	0.002321	0.035977	…	0.009896	-0.005774	0.017936	
4	0.042081	0.072474	0.053771	…	-0.024961	0.008235	-0.015420	
5	0.008816	0.000980	0.022530	…	0.005643	-0.003338	0.026670	
..	…	…	…	…	…	…	…	
493	0.014466	0.019288	0.043398	…	0.024828	0.026473	-0.005013	
494	0.007178	0.001305	0.026754	…	0.018516	0.006207	0.021082	
495	0.003852	0.000482	0.021187	…	0.016570	0.004242	0.023243	
497	0.009397	0.002958	0.025408	…	-0.002531	0.009924	0.028367	
499	0.043420	0.057893	0.049209	…	-0.010479	-0.021759	0.022948	

	X77	X78	X79	X80	X81	X82	X83
0	0.005021	0.030354	0.000408	0.001063	0.011334	0.011976	efectores
1	-0.002604	-0.036967	-0.000371	-0.013827	-0.006040	0.012629	efectores
2	0.004323	-0.005912	0.020768	-0.003046	-0.009587	0.025053	efectores
4	0.030538	0.041747	-0.002484	-0.002353	-0.000388	0.014609	efectores
5	0.008946	-0.009068	0.020809	0.004133	-0.007339	0.009954	efectores
..	…	…	…	…	…	…	
493	-0.061090	-0.011468	-0.006630	0.034681	0.062673	0.022237	efectores
494	0.015660	0.004085	0.019807	0.018843	-0.001708	0.015734	efectores
495	0.024519	0.007799	0.012092	0.007423	0.001260	0.016617	efectores
497	0.001484	0.006192	0.019865	-0.003918	0.011222	0.024197	efectores
499	0.002436	0.031168	-0.013981	-0.010159	0.015840	-0.015495	efectores

[398 rows x 84 columns]

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

	X0	X1	X2	X3	X4	X5	\
count	398.000000	398.000000	398.000000	398.000000	398.000000	398.000000	
mean	0.028955	0.002051	0.020523	0.025333	0.013995	0.024109	
std	0.010961	0.003139	0.015887	0.022380	0.009897	0.009840	
min	0.001368	0.000000	0.000000	0.000000	0.000000	0.001544	
25%	0.021285	0.000000	0.007420	0.006377	0.006682	0.016530	
50%	0.027573	0.000782	0.014666	0.017305	0.011414	0.022493	
75%	0.036054	0.002929	0.033217	0.043525	0.018665	0.030473	
max	0.065041	0.015592	0.068624	0.105912	0.052921	0.059282	

	X6	X7	X8	X9	…	X73	\
count	398.000000	398.000000	398.000000	398.000000	…	398.000000	
mean	0.006205	0.021070	0.017790	0.036968	…	0.014253	
std	0.005218	0.021044	0.023649	0.017676	…	0.013873	
min	0.000000	0.000000	0.000000	0.007573	…	-0.040629	
25%	0.002040	0.006014	0.001796	0.023345	…	0.006757	
50%	0.004926	0.012752	0.005601	0.033609	…	0.014971	
75%	0.009078	0.030520	0.029767	0.045955	…	0.023325	
max	0.028073	0.114871	0.120225	0.108775	…	0.060608	

	X74	X75	X76	X77	X78	X79	\
count	398.000000	398.000000	398.000000	398.000000	398.000000	398.000000	
mean	0.005173	0.006303	0.015716	0.006870	0.007272	0.015191	
std	0.018055	0.015574	0.012684	0.016803	0.014909	0.013861	
min	-0.091093	-0.035415	-0.029328	-0.061090	-0.036967	-0.030841	
25%	-0.004241	-0.002047	0.008312	-0.001579	-0.000580	0.006662	
50%	0.008588	0.003399	0.016754	0.009383	0.004496	0.017191	
75%	0.015643	0.011069	0.024791	0.017135	0.013090	0.024616	
max	0.084842	0.085131	0.056731	0.060540	0.057563	0.048946	

	X80	X81	X82
count	398.000000	398.000000	398.000000
mean	0.007755	0.008150	0.014811
std	0.018321	0.016241	0.013532
min	-0.067451	-0.035780	-0.034700
25%	-0.000736	-0.000182	0.006518
50%	0.010592	0.003954	0.016122
75%	0.017758	0.013653	0.022807
max	0.063431	0.087386	0.056056

[8 rows x 83 columns]

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.033981  0.002427  0.038836  0.038836  0.016182  0.038027  0.008900
3    0.033394  0.001964  0.027501  0.035358  0.010804  0.040269  0.003929
4    0.033079  0.001438  0.020135  0.035955  0.008629  0.025888  0.012944
5    0.045751  0.001173  0.038713  0.035193  0.010558  0.018770  0.007039
6    0.027495  0.002500  0.014997  0.022496  0.009998  0.034994  0.009998
..        …         …         …         …         …         …         …
494  0.036875  0.003010  0.029349  0.027092  0.008278  0.026339  0.005268
495  0.013740  0.000000  0.029770  0.050379  0.004580  0.029770  0.009160
496  0.034300  0.002858  0.032871  0.021438  0.005717  0.027154  0.008575
497  0.041318  0.007870  0.041318  0.053123  0.011805  0.045253  0.007870
499  0.016291  0.000397  0.032980  0.031787  0.005960  0.023841  0.004768

           X7        X8        X9  …       X74       X75       X76  \
0    0.009709  0.006473  0.042072  … -0.014284 -0.002084  0.034870
3    0.016697  0.005893  0.025537  … -0.002892  0.001339  0.031850
4    0.038832  0.028764  0.041708  … -0.014898 -0.010583  0.025917
5    0.017597  0.007039  0.030501  …  0.016117  0.014584  0.037015
6    0.039993  0.029995  0.029995  … -0.021405  0.011555  0.035044
..        …         …         …  …       …         …         …
494  0.006773  0.003010  0.024834  …  0.011656  0.008850  0.020437
495  0.002290  0.004580  0.022900  …  0.044920  0.069844 -0.000645
496  0.000000  0.004287  0.018579  …  0.011209  0.009484  0.025043
497  0.043285  0.031480  0.066895  …  0.011328  0.026399 -0.003360
499  0.016688  0.007947  0.021059  …  0.004720  0.017737  0.019724

           X77       X78       X79       X80       X81       X82           X83
0   -0.006859 -0.004676  0.033110  0.003668  0.022699  0.024586  no_efectores
3   -0.008283  0.008977  0.017159 -0.005442 -0.009005  0.014553  no_efectores
4   -0.022162 -0.010335  0.031039  0.004038 -0.004065  0.016320  no_efectores
5    0.021982  0.006267  0.015745 -0.004374  0.003362  0.034771  no_efectores
6    0.015660  0.000268  0.022798 -0.024125 -0.020604 -0.000740  no_efectores
..        …         …         …         …         …         …           …
494 -0.003380  0.012738  0.022867 -0.001282  0.006253  0.023653  no_efectores
495 -0.010017  0.027299  0.005935  0.022390  0.021400 -0.006952  no_efectores
496 -0.003547  0.015188  0.028236 -0.000795  0.017267  0.004124  no_efectores
497 -0.037905  0.005517 -0.002842 -0.012005 -0.000488 -0.014719  no_efectores
499  0.002241  0.009167  0.019579 -0.010180  0.014170  0.015102  no_efectores

[416 rows x 84 columns]
```

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

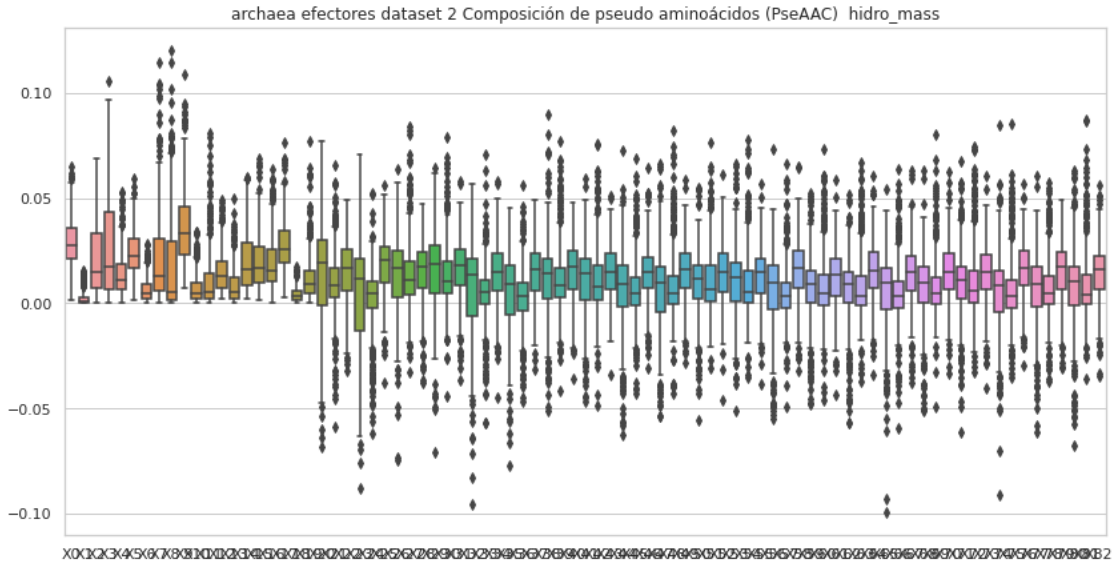
	X0	X1	X2	X3	X4	X5	\
count	416.000000	416.000000	416.000000	416.000000	416.000000	416.000000	
mean	0.032692	0.003716	0.029406	0.030005	0.012216	0.027382	
std	0.012997	0.004888	0.016831	0.018395	0.007865	0.010543	
min	0.004037	0.000000	0.000000	0.000000	0.000000	0.002789	
25%	0.023730	0.000000	0.016211	0.015318	0.006617	0.019787	
50%	0.032183	0.002070	0.029342	0.028651	0.010492	0.026919	
75%	0.040694	0.004822	0.040398	0.043891	0.016261	0.034538	
max	0.074163	0.027712	0.087686	0.087686	0.043626	0.065530	

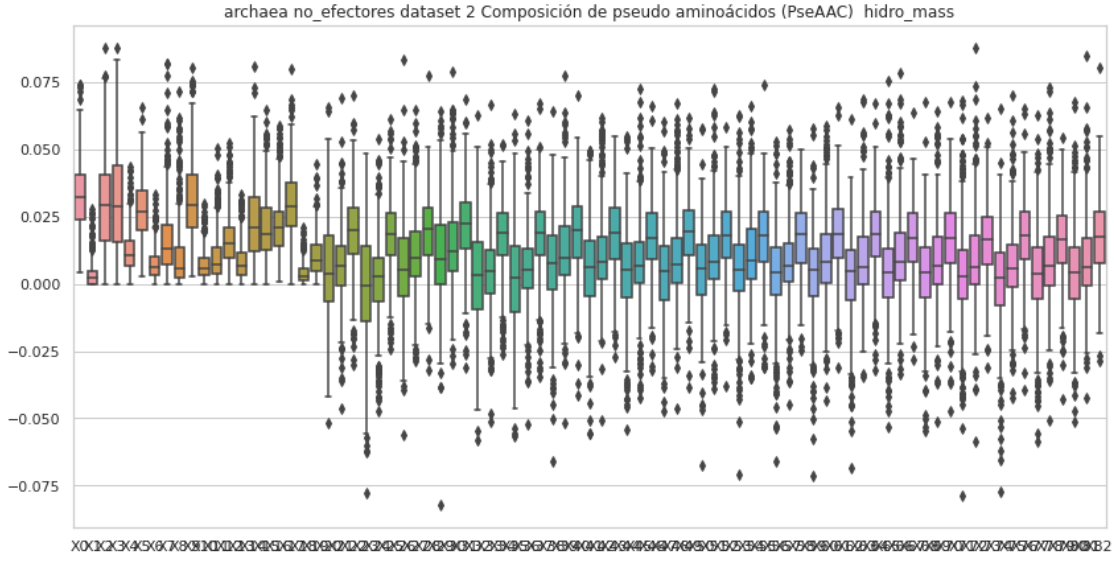
	X6	X7	X8	X9	…	X73	\
count	416.000000	416.000000	416.000000	416.000000	…	416.000000	
mean	0.007350	0.016521	0.010318	0.031572	…	0.015735	
std	0.005872	0.013314	0.012220	0.015005	…	0.013486	
min	0.000000	0.000000	0.000000	0.002791	…	-0.031035	
25%	0.003168	0.007120	0.002250	0.020712	…	0.007048	
50%	0.005971	0.012979	0.005913	0.029130	…	0.016564	
75%	0.010293	0.022011	0.013537	0.040408	…	0.025077	
max	0.033384	0.081942	0.071545	0.080416	…	0.061704	

	X74	X75	X76	X77	X78	X79	\
count	416.000000	416.000000	416.000000	416.000000	416.000000	416.000000	
mean	0.001385	0.007289	0.017338	0.003561	0.009112	0.015799	
std	0.017040	0.014751	0.014162	0.016844	0.016225	0.014036	
min	-0.077306	-0.042581	-0.039560	-0.058710	-0.050631	-0.042846	
25%	-0.007829	-0.001178	0.009309	-0.005639	-0.000740	0.007767	
50%	0.002167	0.005854	0.017921	0.003967	0.006451	0.016549	
75%	0.011616	0.014739	0.026936	0.013512	0.017671	0.025228	
max	0.064944	0.069844	0.064917	0.062543	0.073066	0.054379	

	X80	X81	X82
count	416.000000	416.000000	416.000000
mean	0.004145	0.008756	0.017091
std	0.016685	0.015310	0.014474
min	-0.051187	-0.042364	-0.028646
25%	-0.005566	-0.000662	0.007727
50%	0.004094	0.006188	0.017705
75%	0.013425	0.016793	0.026816
max	0.067431	0.084888	0.080488

[8 rows x 83 columns]





4 Composición de pseudo aminoácidos (PseAAC) mass

[7]:
```
#mass
transf = "Composición de pseudo aminoácidos (PseAAC) "
transf2 = "PseAAC"
estado = "con valores atípicos.\n"
```

```
comp = "mass"
df=""

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

    if etiq == "efectores":
        df=PseAAC_mass_efec

    if etiq == "no_efectores":
        df=PseAAC_mass_no_efec

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

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

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.033567  0.005595  0.039162  0.123079  0.027973  0.044756  0.000000
1    0.062001  0.000000  0.042072  0.017714  0.024357  0.046500  0.015500
2    0.048558  0.003735  0.009338  0.014941  0.026147  0.035485  0.007470
3    0.022074  0.029432  0.044148  0.088296  0.007358  0.073580  0.007358
4    0.072085  0.007208  0.093710  0.075689  0.054063  0.039647  0.018021
..        …         …         …         …         …         …         …
495  0.044307  0.000000  0.006742  0.007706  0.012522  0.029859  0.000963
496  0.000000  0.000000  0.062424  0.104039  0.041616  0.041616  0.031212
497  0.035336  0.001104  0.046599  0.028269  0.012809  0.053887  0.010601
498  0.029992  0.005998  0.083977  0.083977  0.023993  0.041988  0.005998
499  0.025028  0.000000  0.035754  0.071508  0.025028  0.053631  0.028603

           X7        X8        X9  …       X32       X33       X34  \
0    0.033567  0.106296  0.134268  …  0.015960  0.036551  0.007272
```

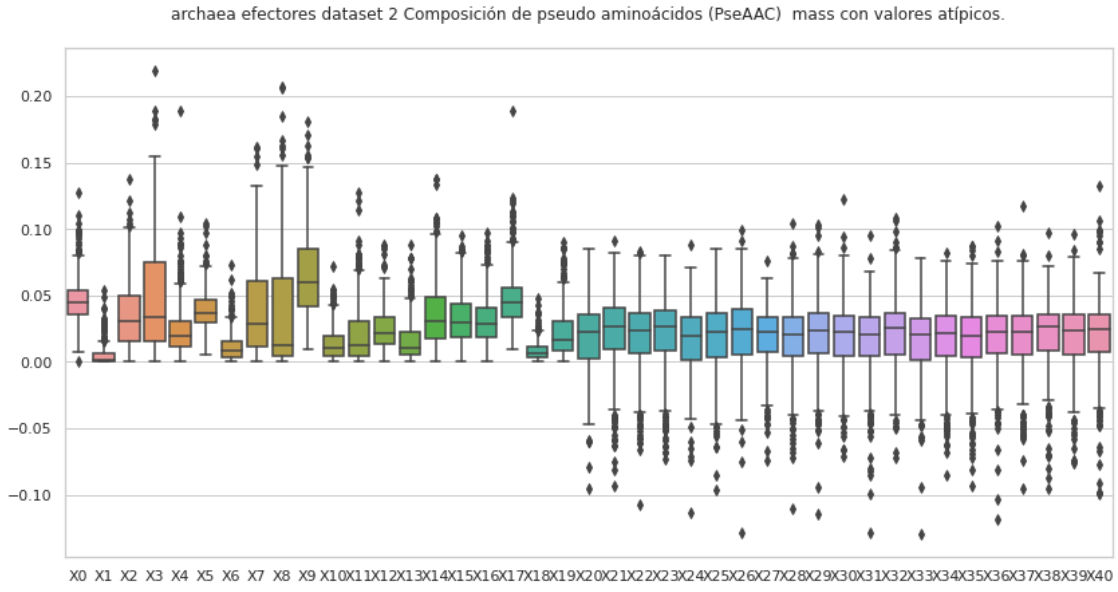
```
1    0.008857  0.000000  0.064215  …  0.028535  0.012222  0.038497
2    0.003735  0.003735  0.057896  …  0.013564  0.020904  0.057116
3    0.095654  0.051506  0.095654  …  0.007141  0.003659 -0.046022
4    0.064876  0.111731  0.082897  … -0.033181 -0.044334 -0.010108
..        …         …        …  …        …         …
495  0.007706  0.000963  0.042381  …  0.043242  0.019611  0.040990
496  0.093635  0.166463  0.156059  …  0.009817  0.051003 -0.010002
497  0.011926  0.003754  0.032244  …  0.021587  0.029156  0.008385
498  0.071980  0.101972  0.107970  … -0.000981  0.024005 -0.000330
499  0.053631  0.071508  0.060782  …  0.039030  0.007334  0.042411

          X35       X36       X37       X38       X39       X40        X41
0    0.001752  0.008733 -0.021295  0.022412  0.000729  0.021380  efectores
1    0.012900  0.002876  0.042637 -0.006943 -0.000459  0.015618  efectores
2    0.012502  0.035508  0.047416  0.028864  0.033422  0.040317  efectores
3   -0.008772  0.063553 -0.028586  0.035166  0.031720 -0.022374  efectores
4    0.025516  0.014726 -0.021089 -0.023773 -0.003829  0.022523  efectores
..        …         …        …         …         …         …
495  0.037941  0.024883  0.027286  0.046493  0.024188  0.033240  efectores
496 -0.031234 -0.003457  0.005451  0.030505 -0.020927 -0.039447  efectores
497  0.021572  0.039295  0.028155  0.036000  0.025209  0.030707  efectores
498  0.002142  0.014197 -0.011585 -0.004256  0.006025  0.031894  efectores
499  0.073502  0.027598  0.004626  0.028345 -0.017269 -0.019139  efectores

[500 rows x 42 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.046145    0.004606    0.035456    0.048536    0.023927    0.038900
std      0.016351    0.007322    0.023949    0.040318    0.019004    0.013607
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.005107
25%      0.035392    0.000000    0.015547    0.015264    0.011500    0.029303
50%      0.044570    0.001734    0.030301    0.034019    0.019299    0.036961
75%      0.053837    0.006238    0.049820    0.075066    0.030884    0.046540
max      0.127280    0.053715    0.137481    0.219051    0.189000    0.103770

                X6          X7          X8          X9  …         X31  \
count  500.000000  500.000000  500.000000  500.000000  …  500.000000
mean     0.011089    0.040488    0.036273    0.066086  …    0.015608
std      0.009946    0.034340    0.042618    0.031489  …    0.025677
min      0.000000    0.000000    0.000000    0.008929  …   -0.129039
25%      0.003821    0.011296    0.004013    0.041707  …    0.004348
50%      0.008731    0.028959    0.012763    0.060001  …    0.020371
```

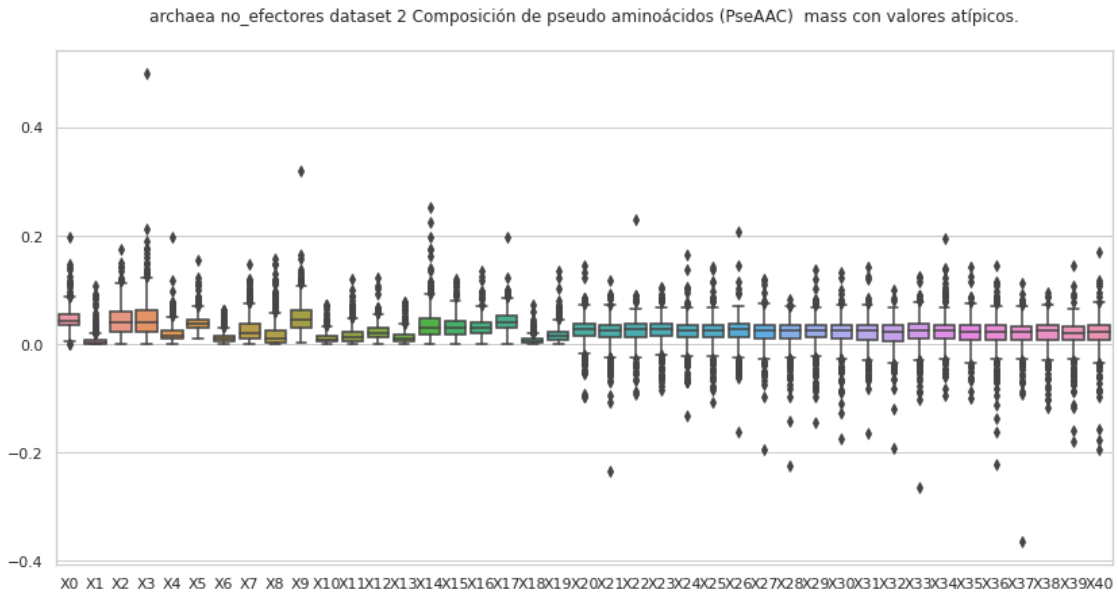
```
75%      0.015804    0.060954    0.063258    0.085473  …    0.033295
max      0.072904    0.161828    0.206769    0.181125  …    0.095408

              X32         X33         X34         X35         X36         X37  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.020778    0.016038    0.018077    0.015674    0.019824    0.018577
std      0.025446    0.023725    0.024982    0.026645    0.024818    0.025499
min     -0.072668   -0.129720   -0.084900   -0.093848   -0.118639   -0.095232
25%      0.005543    0.001451    0.004699    0.003168    0.006441    0.005560
50%      0.025456    0.020829    0.021936    0.019946    0.022871    0.022372
75%      0.036943    0.032711    0.035098    0.033553    0.034958    0.034885
max      0.107726    0.077495    0.081764    0.087396    0.102090    0.116863

              X38         X39         X40
count  500.000000  500.000000  500.000000
mean     0.020912    0.019980    0.020216
std      0.025025    0.024749    0.026897
min     -0.095124   -0.076380   -0.099425
25%      0.008859    0.005693    0.007034
50%      0.026173    0.023102    0.024146
75%      0.036015    0.035544    0.035448
max      0.097499    0.096376    0.132712

[8 rows x 41 columns]
```

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

          X0        X1        X2        X3        X4        X5        X6  \
0    0.041107  0.002936  0.046980  0.046980  0.019575  0.046001  0.010766
1    0.000000  0.019446  0.116674  0.038891  0.077783  0.038891  0.000000
2    0.051538  0.011453  0.080170  0.091623  0.034359  0.057264  0.022906
3    0.037331  0.002196  0.030743  0.039527  0.012078  0.045017  0.004392
4    0.036842  0.001602  0.022426  0.040046  0.009611  0.028833  0.014417
..        …         …         …         …         …         …         …
495  0.033694  0.000000  0.073004  0.123545  0.011231  0.073004  0.022463
496  0.048465  0.004039  0.046446  0.030291  0.008078  0.038368  0.012116
497  0.055997  0.010666  0.055997  0.071996  0.015999  0.061330  0.010666
498  0.062106  0.012421  0.099370  0.062106  0.031053  0.043474  0.062106
499  0.023156  0.000565  0.046877  0.045182  0.008472  0.033887  0.006777

          X7        X8        X9  …       X32       X33       X34  \
0    0.011745  0.007830  0.050895  …  0.010514  0.019674  0.035625
1    0.038891  0.019446  0.077783  … -0.014173  0.001358  0.104741
```

```
2    0.011453  0.011453  0.068717  … -0.023199  0.005766 -0.001684
3    0.018666  0.006588  0.028547  …  0.015612  0.017812  0.015623
4    0.043250  0.032037  0.046453  …  0.007543  0.028971  0.024560
..        …         …        …  …         …         …
495  0.005616  0.011231  0.056157  …  0.022677  0.066040  0.028653
496  0.000000  0.006058  0.026252  …  0.043057  0.051626  0.033398
497  0.058663  0.042664  0.090661  … -0.012869 -0.007040  0.007914
498  0.012421  0.012421  0.074527  …  0.030073 -0.010201 -0.035539
499  0.023721  0.011296  0.029933  …  0.028528  0.026283  0.035247

          X35       X36       X37       X38       X39       X40           X41
0    0.030712  0.026785  0.024541  0.042183  0.040053  0.029742  no_efectores
1    0.116053 -0.020178 -0.031594 -0.006619  0.087836  0.009329  no_efectores
2    0.067138  0.090403  0.059768 -0.016623 -0.032477  0.000343  no_efectores
3    0.022569  0.024846  0.053246  0.035605  0.019182  0.016268  no_efectores
4    0.019153  0.027802  0.025043  0.028866  0.034570  0.018177  no_efectores
..        …         …         …         …         …         …           …
495  0.039913  0.020139  0.037969 -0.001581  0.014555 -0.017049  no_efectores
496  0.007223  0.018723  0.030980  0.035385  0.039897  0.005827  no_efectores
497  0.013161 -0.008993 -0.002334 -0.004554 -0.003852 -0.019949  no_efectores
498 -0.008322 -0.031924  0.033593 -0.014246 -0.026791  0.000748  no_efectores
499  0.032341  0.022393  0.024727  0.028036  0.027829  0.021466  no_efectores

[500 rows x 42 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.046936    0.007070    0.044214    0.048736    0.020331    0.040249
std      0.020646    0.012553    0.027626    0.040070    0.017094    0.015960
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.010255
25%      0.034997    0.000000    0.023832    0.023110    0.009675    0.030337
50%      0.043670    0.003022    0.040051    0.040594    0.016481    0.038554
75%      0.055840    0.008184    0.059868    0.063615    0.026339    0.046460
max      0.197030    0.106745    0.174901    0.499718    0.197030    0.155546

                X6          X7          X8          X9  …         X31  \
count  500.000000  500.000000  500.000000  500.000000  …  500.000000
mean     0.011585    0.027399    0.019505    0.050186  …    0.019511
std      0.010153    0.023765    0.024602    0.028522  …    0.027160
min      0.000000    0.000000    0.000000    0.003620  …   -0.165594
25%      0.004723    0.010482    0.003745    0.031504  …    0.008384
50%      0.009389    0.020558    0.010819    0.044408  …    0.024137
75%      0.015627    0.037071    0.026056    0.063878  …    0.034684
```

```
max      0.062904    0.147772    0.158121    0.320174  …    0.142028

              X32         X33         X34         X35         X36         X37  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.018768    0.021866    0.023108    0.019394    0.019435    0.019229
std      0.025515    0.029737    0.026890    0.025180    0.031330    0.029544
min     -0.192217   -0.263442   -0.095697   -0.100635   -0.221714   -0.364251
25%      0.006514    0.011005    0.009815    0.006986    0.008683    0.008343
50%      0.023347    0.026252    0.024626    0.023094    0.023011    0.022630
75%      0.034626    0.037870    0.035547    0.035232    0.035005    0.033558
max      0.099368    0.125525    0.196266    0.141759    0.144864    0.113855

              X38         X39         X40
count  500.000000  500.000000  500.000000
mean     0.019362    0.017364    0.019371
std      0.026143    0.028951    0.031223
min     -0.116695   -0.179256   -0.195519
25%      0.008013    0.008497    0.007566
50%      0.024474    0.021034    0.023557
75%      0.034847    0.033731    0.035607
max      0.095307    0.146104    0.170700

[8 rows x 41 columns]
```





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

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

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

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

    if etiq == "efectores":
        df=PseAAC_mass_efec

    if etiq == "no_efectores":
        df=PseAAC_mass_no_efec

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

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

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

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

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

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.033567  0.005595  0.039162  0.123079  0.027973  0.044756  0.000000
1    0.062001  0.000000  0.042072  0.017714  0.024357  0.046500  0.015500
2    0.048558  0.003735  0.009338  0.014941  0.026147  0.035485  0.007470
4    0.072085  0.007208  0.093710  0.075689  0.054063  0.039647  0.018021
5    0.046515  0.001454  0.014536  0.010175  0.005814  0.027618  0.001454
..        …         …         …         …         …         …         …
494  0.049399  0.000000  0.008468  0.012703  0.018348  0.031051  0.002823
495  0.044307  0.000000  0.006742  0.007706  0.012522  0.029859  0.000963
497  0.035336  0.001104  0.046599  0.028269  0.012809  0.053887  0.010601
498  0.029992  0.005998  0.083977  0.083977  0.023993  0.041988  0.005998
499  0.025028  0.000000  0.035754  0.071508  0.025028  0.053631  0.028603

           X7        X8        X9  …       X32       X33       X34  \
0    0.033567  0.106296  0.134268  …  0.015960  0.036551  0.007272
1    0.008857  0.000000  0.064215  …  0.028535  0.012222  0.038497
2    0.003735  0.003735  0.057896  …  0.013564  0.020904  0.057116
4    0.064876  0.111731  0.082897  … -0.033181 -0.044334 -0.010108
5    0.013082  0.001454  0.033433  …  0.049673  0.029494  0.012672
..        …         …         …  …       …         …         …
494  0.015525  0.002823  0.057868  …  0.047845  0.020897  0.001520
495  0.007706  0.000963  0.042381  …  0.043242  0.019611  0.040990
497  0.011926  0.003754  0.032244  …  0.021587  0.029156  0.008385
498  0.071980  0.101972  0.107970  … -0.000981  0.024005 -0.000330
499  0.053631  0.071508  0.060782  …  0.039030  0.007334  0.042411
```

	X35	X36	X37	X38	X39	X40	X41
0	0.001752	0.008733	-0.021295	0.022412	0.000729	0.021380	efectores
1	0.012900	0.002876	0.042637	-0.006943	-0.000459	0.015618	efectores
2	0.012502	0.035508	0.047416	0.028864	0.033422	0.040317	efectores
4	0.025516	0.014726	-0.021089	-0.023773	-0.003829	0.022523	efectores
5	0.037607	0.024914	0.031262	0.039575	0.030878	0.014771	efectores
..	…	…	…	…	…	…	
494	0.030045	0.028872	0.035087	0.045600	0.042841	0.034033	efectores
495	0.037941	0.024883	0.027286	0.046493	0.024188	0.033240	efectores
497	0.021572	0.039295	0.028155	0.036000	0.025209	0.030707	efectores
498	0.002142	0.014197	-0.011585	-0.004256	0.006025	0.031894	efectores
499	0.073502	0.027598	0.004626	0.028345	-0.017269	-0.019139	efectores

[396 rows x 42 columns]

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

	X0	X1	X2	X3	X4	X5	\
count	396.000000	396.000000	396.000000	396.000000	396.000000	396.000000	
mean	0.044456	0.003025	0.030462	0.037482	0.020600	0.036432	
std	0.012774	0.004450	0.020655	0.032042	0.014160	0.010809	
min	0.010939	0.000000	0.000000	0.000000	0.000000	0.006218	
25%	0.035797	0.000000	0.013677	0.012663	0.010707	0.028510	
50%	0.043934	0.001158	0.024471	0.025681	0.017054	0.035258	
75%	0.051507	0.004764	0.044572	0.056998	0.026531	0.043677	
max	0.093140	0.026388	0.102473	0.147121	0.076227	0.069615	

	X6	X7	X8	X9	…	X31	\
count	396.000000	396.000000	396.000000	396.000000	…	396.000000	
mean	0.009260	0.031754	0.026149	0.058235	…	0.020703	
std	0.007541	0.028601	0.034058	0.026255	…	0.019293	
min	0.000000	0.000000	0.000000	0.008929	…	-0.046336	
25%	0.003189	0.009295	0.002941	0.039483	…	0.008592	
50%	0.007268	0.020656	0.008054	0.053352	…	0.023803	
75%	0.013736	0.048959	0.040205	0.077208	…	0.034387	
max	0.038868	0.125195	0.147784	0.139351	…	0.078463	

	X32	X33	X34	X35	X36	X37	\
count	396.000000	396.000000	396.000000	396.000000	396.000000	396.000000	
mean	0.023835	0.018952	0.022372	0.021388	0.023164	0.022315	
std	0.021004	0.020162	0.020072	0.021003	0.018963	0.020202	
min	-0.049616	-0.048127	-0.051912	-0.061663	-0.050292	-0.055843	
25%	0.013414	0.006573	0.010930	0.011132	0.013002	0.010748	
50%	0.027608	0.023743	0.026049	0.023443	0.025574	0.025494	
75%	0.037836	0.033935	0.036498	0.035253	0.035312	0.035659	

```
max      0.086070    0.073596    0.075898    0.087396    0.072298    0.075312

              X38         X39         X40
count  396.000000  396.000000  396.000000
mean     0.024028    0.023916    0.024279
std      0.019881    0.019302    0.020112
min     -0.051347   -0.037536   -0.047703
25%      0.013089    0.013305    0.012729
50%      0.027960    0.027191    0.026933
75%      0.036273    0.036268    0.036033
max      0.069923    0.074342    0.098956

[8 rows x 41 columns]
```

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

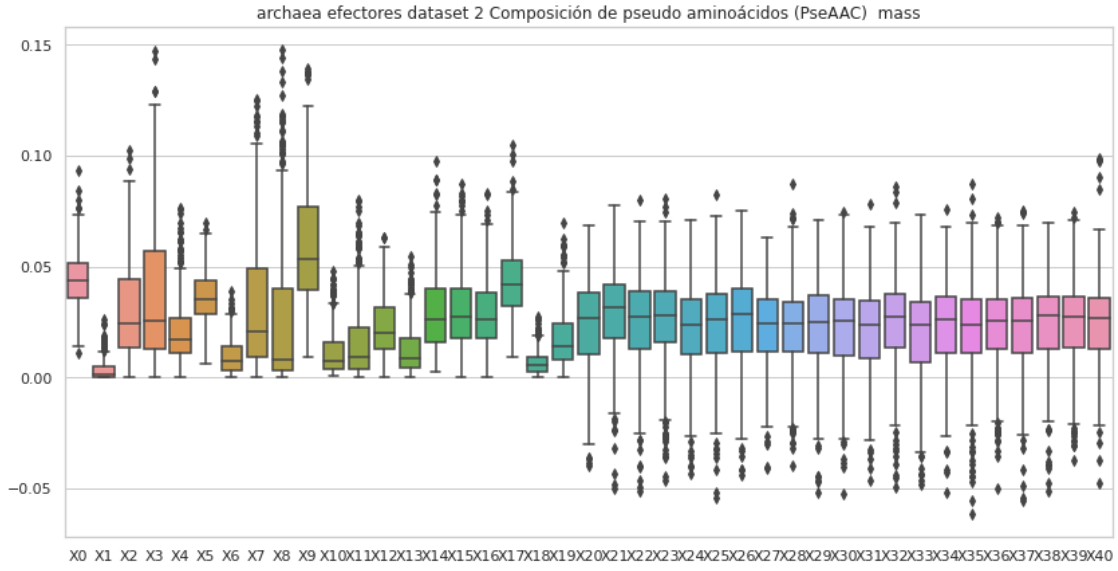
```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.041107  0.002936  0.046980  0.046980  0.019575  0.046001  0.010766
3    0.037331  0.002196  0.030743  0.039527  0.012078  0.045017  0.004392
4    0.036842  0.001602  0.022426  0.040046  0.009611  0.028833  0.014417
5    0.055311  0.001418  0.046801  0.042547  0.012764  0.022692  0.008509
6    0.031711  0.002883  0.017297  0.025945  0.011531  0.040359  0.011531
..        …         …         …         …         …         …         …
494  0.048009  0.003919  0.038211  0.035272  0.010777  0.034292  0.006858
495  0.033694  0.000000  0.073004  0.123545  0.011231  0.073004  0.022463
496  0.048465  0.004039  0.046446  0.030291  0.008078  0.038368  0.012116
497  0.055997  0.010666  0.055997  0.071996  0.015999  0.061330  0.010666
499  0.023156  0.000565  0.046877  0.045182  0.008472  0.033887  0.006777

           X7        X8        X9  …       X32       X33       X34  \
0    0.011745  0.007830  0.050895  …  0.010514  0.019674  0.035625
3    0.018666  0.006588  0.028547  …  0.015612  0.017812  0.015623
4    0.043250  0.032037  0.046453  …  0.007543  0.028971  0.024560
5    0.021273  0.008509  0.036874  …  0.005143  0.046373  0.032758
6    0.046125  0.034593  0.034593  …  0.034504  0.019607  0.027901
..        …         …         …  …  …       …         …         …
494  0.008818  0.003919  0.032332  …  0.026433  0.018099  0.037137
495  0.005616  0.011231  0.056157  …  0.022677  0.066040  0.028653
496  0.000000  0.006058  0.026252  …  0.043057  0.051626  0.033398
497  0.058663  0.042664  0.090661  … -0.012869 -0.007040  0.007914
499  0.023721  0.011296  0.029933  …  0.028528  0.026283  0.035247

          X35       X36       X37       X38       X39       X40           X41
0    0.030712  0.026785  0.024541  0.042183  0.040053  0.029742  no_efectores
```

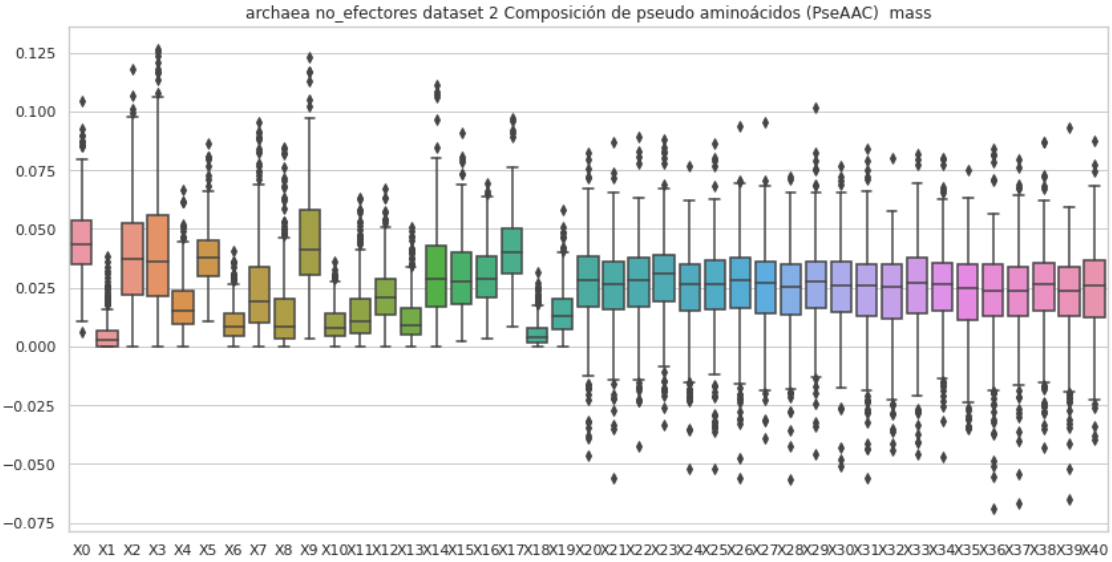
```
3    0.022569  0.024846  0.053246  0.035605  0.019182  0.016268  no_efectores
4    0.019153  0.027802  0.025043  0.028866  0.034570  0.018177  no_efectores
5    0.022660  0.008230  0.024989  0.044749  0.019035  0.042037  no_efectores
6    0.012925  0.046623  0.015072  0.040416  0.026293 -0.000854  no_efectores
..        …         …         …         …         …         …
494  0.043787  0.024961  0.023656  0.026608  0.029772  0.030795  no_efectores
495  0.039913  0.020139  0.037969 -0.001581  0.014555 -0.017049  no_efectores
496  0.007223  0.018723  0.030980  0.035385  0.039897  0.005827  no_efectores
497  0.013161 -0.008993 -0.002334 -0.004554 -0.003852 -0.019949  no_efectores
499  0.032341  0.022393  0.024727  0.028036  0.027829  0.021466  no_efectores

[420 rows x 42 columns]
```

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

```
                X0          X1          X2          X3          X4          X5  \
count   420.000000  420.000000  420.000000  420.000000  420.000000  420.000000
mean      0.044636    0.005039    0.039198    0.040430    0.017495    0.038077
std       0.015328    0.006521    0.022195    0.026274    0.011264    0.012040
min       0.006126    0.000000    0.000000    0.000000    0.000000    0.010566
25%       0.034997    0.000000    0.021975    0.021346    0.009322    0.030072
50%       0.043447    0.002711    0.037283    0.036034    0.014968    0.037767
75%       0.053753    0.006853    0.052643    0.055877    0.023769    0.045021
max       0.104236    0.038467    0.117932    0.126393    0.066569    0.086378

                X6          X7          X8          X9  …         X31  \
count   420.000000  420.000000  420.000000  420.000000  …  420.000000
mean      0.010062    0.023877    0.014822    0.044749  …    0.023076
std       0.007530    0.018717    0.016553    0.020764  …    0.019387
min       0.000000    0.000000    0.000000    0.003620  …   -0.055918
25%       0.004498    0.010065    0.003281    0.030675  …    0.012763
50%       0.008506    0.019280    0.008438    0.040975  …    0.025693
75%       0.013846    0.033748    0.020500    0.057867  …    0.034859
max       0.040439    0.095272    0.084654    0.123148  …    0.084183

               X32         X33         X34         X35         X36         X37  \
count   420.000000  420.000000  420.000000  420.000000  420.000000  420.000000
mean      0.022006    0.024514    0.024269    0.021586    0.022048    0.022159
std       0.018255    0.019497    0.017384    0.018399    0.020214    0.018738
min      -0.043843   -0.045550   -0.046907   -0.035057   -0.068942   -0.066521
25%       0.011562    0.014009    0.015450    0.011426    0.013204    0.013145
50%       0.025190    0.027099    0.026229    0.024922    0.023714    0.023552
75%       0.034854    0.037647    0.035316    0.035232    0.034758    0.033577
max       0.080138    0.082084    0.079874    0.075159    0.084124    0.079490
```

```
              X38         X39         X40
count  420.000000  420.000000  420.000000
mean     0.023984    0.021653    0.023464
std      0.018875    0.018569    0.019370
min     -0.043109   -0.064772   -0.039560
25%      0.015119    0.012685    0.012332
50%      0.026595    0.023616    0.025908
75%      0.035644    0.033926    0.036391
max      0.087152    0.093348    0.087410

[8 rows x 41 columns]
```





5 Composición de pseudo aminoácidos (PseAAC) hidro

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

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

    if etiq == "efectores":
        df=PseAAC_hidro_efec

    if etiq == "no_efectores":
        df=PseAAC_hidro_no_efec

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

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

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

```
          X0        X1        X2        X3        X4        X5        X6  \
0    0.020383  0.003397  0.023780  0.074737  0.016986  0.027177  0.000000
1    0.084372  0.000000  0.057252  0.024106  0.033146  0.063279  0.021093
2    0.049221  0.003786  0.009466  0.015145  0.026504  0.035969  0.007572
3    0.023379  0.031172  0.046757  0.093515  0.007793  0.077929  0.007793
4    0.046265  0.004627  0.060145  0.048578  0.034699  0.025446  0.011566
..        …         …         …         …         …         …         …
495  0.034755  0.000000  0.005289  0.006044  0.009822  0.023422  0.000756
496  0.000000  0.000000  0.045341  0.075568  0.030227  0.030227  0.022671
497  0.049902  0.001559  0.065808  0.039921  0.018089  0.076100  0.014971
498  0.023781  0.004756  0.066587  0.066587  0.019025  0.033293  0.004756
499  0.026438  0.000000  0.037768  0.075536  0.026438  0.056652  0.030214

          X7        X8        X9  …       X53       X54       X55  \
0    0.020383  0.064545  0.081531  …  0.005332 -0.010823  0.013162
1    0.012053  0.000000  0.087385  … -0.002618 -0.002240 -0.054572
2    0.003786  0.003786  0.058687  … -0.003614  0.011870 -0.001094
3    0.101308  0.054550  0.101308  …  0.072739 -0.019486 -0.014384
4    0.041639  0.071711  0.053205  … -0.005175  0.042315  0.063325
..        …         …         …  …         …         …         …
495  0.006044  0.000756  0.033244  …  0.011481  0.020088  0.000843
496  0.068012  0.120909  0.113353  … -0.014502  0.047058  0.051091
497  0.016842  0.005302  0.045535  …  0.023626  0.003830  0.026025
498  0.057074  0.080855  0.085612  …  0.025787 -0.006576  0.067820
499  0.056652  0.075536  0.064206  …  0.005677  0.014381  0.037374

          X56       X57       X58       X59       X60       X61         X62
0    0.010566  0.044380  0.005443  0.032906  0.001152  0.012287  efectores
1   -0.007044  0.001816 -0.004381 -0.062210 -0.023269 -0.010164  efectores
2    0.016143 -0.009418  0.007053 -0.009644 -0.004968 -0.015638  efectores
3   -0.120332 -0.080879 -0.079758  0.033341  0.019187 -0.015340  efectores
4   -0.024698  0.008148  0.030216  0.041308 -0.002328 -0.000384  efectores
..        …         …         …         …         …         …         …
495  0.026000  0.006656  0.038472  0.012238  0.011648  0.001977  efectores
496  0.117443  0.152361  0.025295  0.078143 -0.081947 -0.083290  efectores
497 -0.004536  0.017786  0.002659  0.011097 -0.007021  0.020112  efectores
```

```
498 -0.055492 -0.007618 -0.035734 -0.005098 -0.013746 -0.037371  efectores
499 -0.013673 -0.028390  0.003178  0.040667 -0.013255  0.020667  efectores

[500 rows x 63 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.047424	0.004592	0.035330	0.044650	0.021332	0.039433	
std	0.026583	0.007551	0.025635	0.034557	0.014010	0.020869	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.001587	
25%	0.029426	0.000000	0.012584	0.011025	0.011446	0.025229	
50%	0.043812	0.001831	0.029148	0.044134	0.018170	0.034811	
75%	0.059009	0.005679	0.053635	0.068750	0.027421	0.048221	
max	0.269503	0.047228	0.130825	0.359337	0.098965	0.188958	

	X6	X7	X8	X9	…	X52	\
count	500.000000	500.000000	500.000000	500.000000	…	500.000000	
mean	0.010536	0.037298	0.032437	0.061842	…	0.007505	
std	0.009551	0.037407	0.040158	0.033050	…	0.035095	
min	0.000000	0.000000	0.000000	0.014266	…	-0.204845	
25%	0.003762	0.012178	0.003711	0.040538	…	-0.006228	
50%	0.008975	0.025155	0.012853	0.053626	…	0.013784	
75%	0.014569	0.052932	0.056482	0.076095	…	0.026023	
max	0.089834	0.449171	0.389282	0.359337	…	0.194952	

	X53	X54	X55	X56	X57	X58	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.010191	0.013807	0.016288	0.004477	0.007202	0.006583	
std	0.029368	0.031699	0.030489	0.039318	0.031449	0.034761	
min	-0.247961	-0.180273	-0.085035	-0.290783	-0.228315	-0.201803	
25%	-0.002016	0.000080	0.000053	-0.011489	-0.004256	-0.007504	
50%	0.007223	0.017295	0.011293	0.012416	0.004981	0.012805	
75%	0.020809	0.028712	0.030043	0.025045	0.018979	0.025713	
max	0.183132	0.175510	0.161154	0.187872	0.167339	0.135419	

	X59	X60	X61
count	500.000000	500.000000	500.000000
mean	0.010640	0.008588	0.012740
std	0.026620	0.043701	0.035980
min	-0.104986	-0.344594	-0.433868
25%	-0.002252	-0.006072	-0.001166
50%	0.008261	0.013140	0.007601
75%	0.021730	0.029562	0.024396

```
max      0.147493    0.207002    0.174452

[8 rows x 62 columns]



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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.057672  0.004119  0.065911  0.065911  0.027463  0.064538  0.015105
1    0.000000  0.007864  0.047182  0.015727  0.031455  0.015727  0.000000
2    0.069315  0.015403  0.107824  0.123228  0.046210  0.077017  0.030807
3    0.072437  0.004261  0.059654  0.076698  0.023435  0.087350  0.008522
4    0.060048  0.002611  0.036551  0.065270  0.015665  0.046994  0.023497
..        …         …         …         …         …         …         …
495  0.015282  0.000000  0.033111  0.056035  0.005094  0.033111  0.010188
496  0.058751  0.004896  0.056303  0.036719  0.009792  0.046511  0.014688
497  0.049735  0.009473  0.049735  0.063945  0.014210  0.054472  0.009473
498  0.039657  0.007931  0.063451  0.039657  0.019828  0.027760  0.039657
499  0.028891  0.000705  0.058487  0.056373  0.010570  0.042280  0.008456

           X7        X8        X9  …       X53       X54       X55  \
0    0.016478  0.010985  0.071404  …  0.019691  0.022198  0.032943
1    0.015727  0.007864  0.031455  …  0.047157  0.029591  0.047993
2    0.015403  0.015403  0.092421  …  0.026090  0.044519 -0.020427
3    0.036218  0.012783  0.055393  …  0.044956  0.014060 -0.000637
4    0.070492  0.052216  0.075713  … -0.000424 -0.017321  0.024486
..        …         …         …  …       …         …         …
495  0.002547  0.005094  0.025470  …  0.035161 -0.017763  0.032380
496  0.000000  0.007344  0.031823  …  0.008957 -0.008991 -0.005183
497  0.052103  0.037893  0.080523  …  0.008395  0.027797  0.042098
498  0.007931  0.007931  0.047588  …  0.015408  0.002418  0.006407
499  0.029596  0.014093  0.037347  …  0.017603 -0.003492  0.022454

           X56       X57       X58       X59       X60       X61           X62
0   -0.024243 -0.003538 -0.011641 -0.007937  0.006225  0.038524  no_efectores
1   -0.001485  0.047520  0.069754  0.021843 -0.111788 -0.059595  no_efectores
2    0.002100  0.059635  0.073635  0.017574 -0.055455 -0.105669  no_efectores
3   -0.006273  0.002905 -0.017968  0.019472 -0.011804 -0.019533  no_efectores
4   -0.027044 -0.019211 -0.040230 -0.018762  0.007330 -0.007379  no_efectores
..        …         …         …         …         …         …           …
495  0.049963  0.077684 -0.011141  0.030363  0.024903  0.023802  no_efectores
496  0.019199  0.016245 -0.006075  0.026015 -0.001361  0.029576  no_efectores
497  0.013636  0.031777 -0.045627  0.006641 -0.014451 -0.000587  no_efectores
498  0.006406  0.027289 -0.004278  0.006488  0.044815  0.068877  no_efectores
```

```
499  0.008370  0.031455  0.003974  0.016257 -0.018054  0.025129  no_efectores

[500 rows x 63 columns]
```

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

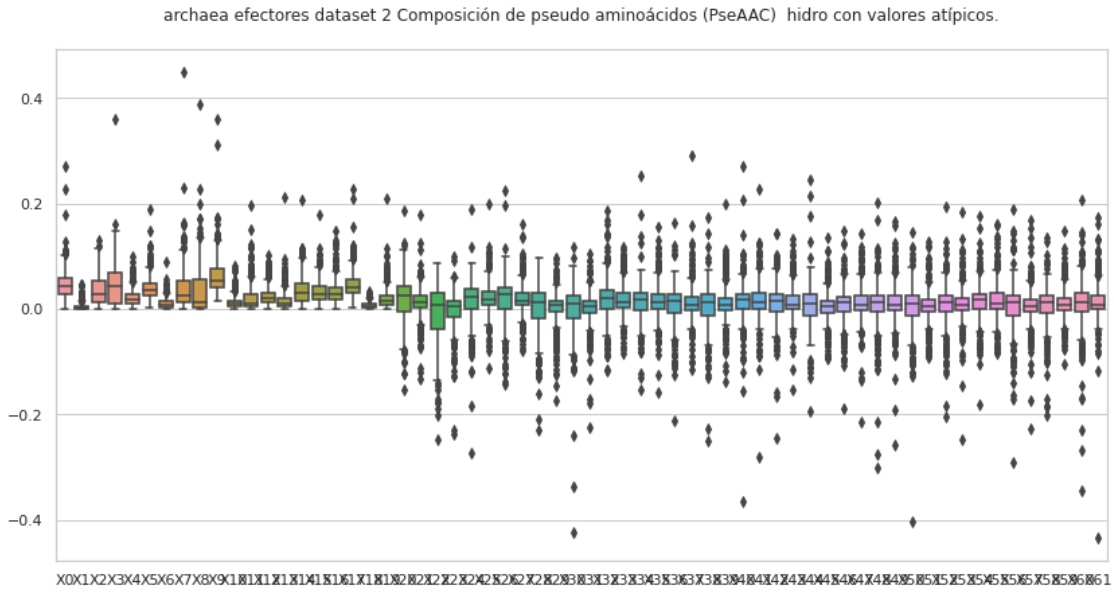
	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.057464	0.007626	0.050085	0.051416	0.022089	0.049342	
std	0.032787	0.012347	0.027300	0.028031	0.015354	0.028462	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.002785	
25%	0.033594	0.000000	0.030673	0.032182	0.012554	0.031834	
50%	0.053282	0.004258	0.053341	0.052850	0.018616	0.045226	
75%	0.077662	0.009306	0.068154	0.069829	0.027768	0.062061	
max	0.306024	0.116836	0.233673	0.161883	0.118033	0.350509	

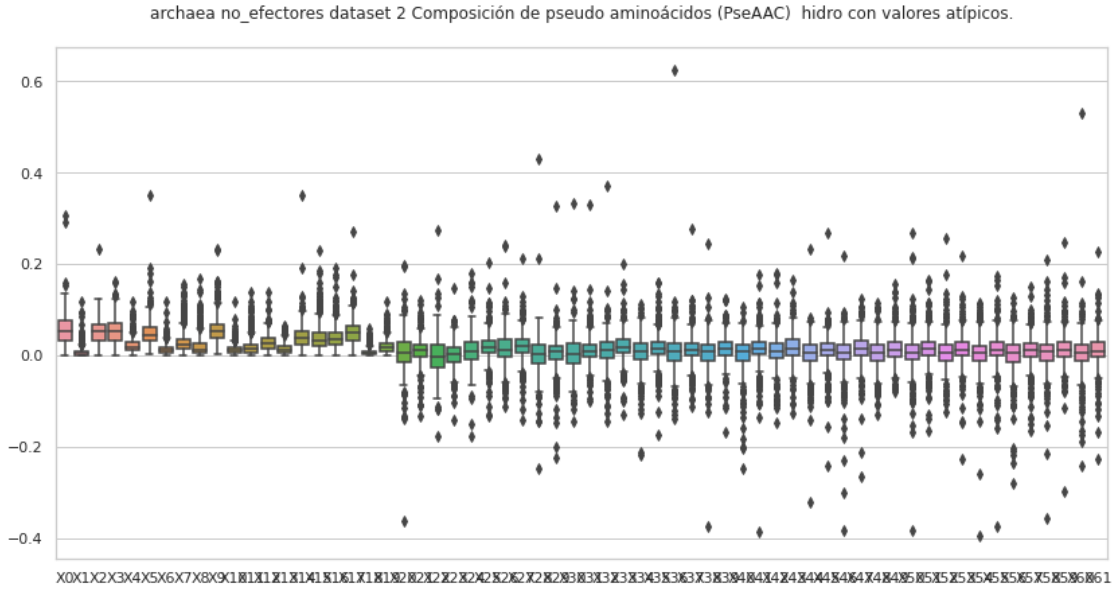
	X6	X7	X8	X9	…	X52	\
count	500.000000	500.000000	500.000000	500.000000	…	500.000000	
mean	0.013408	0.030412	0.020516	0.056378	…	0.006684	
std	0.012017	0.025959	0.024216	0.029555	…	0.033340	
min	0.000000	0.000000	0.000000	0.002948	…	-0.122991	
25%	0.005378	0.013412	0.005037	0.037824	…	-0.011324	
50%	0.011227	0.024012	0.011698	0.052345	…	0.006615	
75%	0.018146	0.036225	0.026436	0.068913	…	0.022265	
max	0.116836	0.156297	0.169162	0.233673	…	0.255430	

	X53	X54	X55	X56	X57	X58	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.013233	0.003066	0.013215	0.000263	0.009536	0.005659	
std	0.032414	0.039853	0.037448	0.039186	0.031901	0.039395	
min	-0.226432	-0.395381	-0.374098	-0.278955	-0.165827	-0.358208	
25%	-0.001415	-0.013461	-0.001646	-0.014038	-0.002632	-0.011529	
50%	0.010833	0.004629	0.011910	0.004311	0.010187	0.007568	
75%	0.029535	0.021741	0.029378	0.022231	0.026008	0.023791	
max	0.217890	0.114002	0.172642	0.127496	0.151500	0.208907	

	X59	X60	X61
count	500.000000	500.000000	500.000000
mean	0.012988	0.003998	0.011151
std	0.033653	0.044044	0.034251
min	-0.297081	-0.240775	-0.225953
25%	-0.001880	-0.013467	-0.003485
50%	0.010882	0.006190	0.009543
75%	0.028345	0.022591	0.028424
max	0.248022	0.530306	0.226597

[8 rows x 62 columns]





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

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

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


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

    if etiq == "efectores":
        df=PseAAC_hidro_efec

    if etiq == "no_efectores":
        df=PseAAC_hidro_no_efec

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

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


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

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

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

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.020383	0.003397	0.023780	0.074737	0.016986	0.027177	0.000000	
1	0.084372	0.000000	0.057252	0.024106	0.033146	0.063279	0.021093	
2	0.049221	0.003786	0.009466	0.015145	0.026504	0.035969	0.007572	
4	0.046265	0.004627	0.060145	0.048578	0.034699	0.025446	0.011566	
5	0.058512	0.001829	0.018285	0.012800	0.007314	0.034742	0.001829	
..	…	…	…	…	…	…	…	
494	0.032140	0.000000	0.005510	0.008265	0.011938	0.020202	0.001837	
495	0.034755	0.000000	0.005289	0.006044	0.009822	0.023422	0.000756	
497	0.049902	0.001559	0.065808	0.039921	0.018089	0.076100	0.014971	
498	0.023781	0.004756	0.066587	0.066587	0.019025	0.033293	0.004756	
499	0.026438	0.000000	0.037768	0.075536	0.026438	0.056652	0.030214	

	X7	X8	X9	…	X53	X54	X55	\
0	0.020383	0.064545	0.081531	…	0.005332	-0.010823	0.013162	
1	0.012053	0.000000	0.087385	…	-0.002618	-0.002240	-0.054572	
2	0.003786	0.003786	0.058687	…	-0.003614	0.011870	-0.001094	
4	0.041639	0.071711	0.053205	…	-0.005175	0.042315	0.063325	
5	0.016457	0.001829	0.042056	…	0.005475	0.017267	-0.001939	
..	…	…	…	…	…	…	…	
494	0.010101	0.001837	0.037650	…	0.006845	0.013281	-0.001109	
495	0.006044	0.000756	0.033244	…	0.011481	0.020088	0.000843	
497	0.016842	0.005302	0.045535	…	0.023626	0.003830	0.026025	
498	0.057074	0.080855	0.085612	…	0.025787	-0.006576	0.067820	
499	0.056652	0.075536	0.064206	…	0.005677	0.014381	0.037374	

	X56	X57	X58	X59	X60	X61	X62
0	0.010566	0.044380	0.005443	0.032906	0.001152	0.012287	efectores
1	-0.007044	0.001816	-0.004381	-0.062210	-0.023269	-0.010164	efectores
2	0.016143	-0.009418	0.007053	-0.009644	-0.004968	-0.015638	efectores
4	-0.024698	0.008148	0.030216	0.041308	-0.002328	-0.000384	efectores
5	0.010534	-0.006232	0.016698	-0.016926	0.007715	-0.013699	efectores
..	…	…	…	…	…	…	…
494	0.026057	0.008735	0.022038	0.005749	0.026518	-0.002403	efectores
495	0.026000	0.006656	0.038472	0.012238	0.011648	0.001977	efectores
497	-0.004536	0.017786	0.002659	0.011097	-0.007021	0.020112	efectores
498	-0.055492	-0.007618	-0.035734	-0.005098	-0.013746	-0.037371	efectores
499	-0.013673	-0.028390	0.003178	0.040667	-0.013255	0.020667	efectores

```
[404 rows x 63 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	404.000000	404.000000	404.000000	404.000000	404.000000	404.000000	
mean	0.044627	0.002912	0.029536	0.035525	0.019676	0.035694	
std	0.021498	0.004459	0.022111	0.027220	0.011044	0.015518	
min	0.001332	0.000000	0.000000	0.000000	0.000000	0.001587	
25%	0.028953	0.000000	0.010714	0.009620	0.011388	0.024146	
50%	0.041306	0.001273	0.022530	0.031360	0.017315	0.033540	
75%	0.055902	0.003711	0.044387	0.058670	0.026361	0.043894	
max	0.111933	0.023912	0.100758	0.102136	0.059056	0.089474	

	X6	X7	X8	X9	…	X52	\
count	404.000000	404.000000	404.000000	404.000000	…	404.000000	
mean	0.008823	0.028516	0.023724	0.053694	…	0.010831	
std	0.006580	0.023948	0.028831	0.020825	…	0.022673	
min	0.000000	0.000000	0.000000	0.014266	…	-0.083700	
25%	0.003387	0.010491	0.003053	0.038678	…	0.000316	
50%	0.008315	0.019589	0.008870	0.048936	…	0.015068	
75%	0.012663	0.040286	0.041741	0.067094	…	0.025780	
max	0.036826	0.130071	0.115939	0.138798	…	0.087877	

	X53	X54	X55	X56	X57	X58	\
count	404.000000	404.000000	404.000000	404.000000	404.000000	404.000000	
mean	0.010054	0.015662	0.014111	0.009455	0.009362	0.011088	
std	0.018374	0.022926	0.022296	0.024689	0.019799	0.022944	
min	-0.054916	-0.080936	-0.054572	-0.076472	-0.043109	-0.063724	
25%	-0.000363	0.003600	0.000706	-0.005515	-0.002737	-0.001238	
50%	0.007024	0.017867	0.009716	0.014712	0.005835	0.014825	
75%	0.017340	0.028601	0.026093	0.025168	0.018279	0.026118	
max	0.088100	0.108079	0.103861	0.111512	0.082299	0.091836	

	X59	X60	X61
count	404.000000	404.000000	404.000000
mean	0.010766	0.011864	0.010686
std	0.021051	0.026188	0.020477
min	-0.062210	-0.085481	-0.038558
25%	-0.001091	-0.002136	-0.000717
50%	0.007822	0.015213	0.006648
75%	0.019413	0.028344	0.019977
max	0.087422	0.091582	0.106513

[8 rows x 62 columns]

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.057672	0.004119	0.065911	0.065911	0.027463	0.064538	0.015105	
3	0.072437	0.004261	0.059654	0.076698	0.023435	0.087350	0.008522	
4	0.060048	0.002611	0.036551	0.065270	0.015665	0.046994	0.023497	
5	0.083275	0.002135	0.070463	0.064058	0.019217	0.034164	0.012812	
6	0.051842	0.004713	0.028277	0.042416	0.018852	0.065980	0.018852	
..	…	…	…	…	…	…	…	
495	0.015282	0.000000	0.033111	0.056035	0.005094	0.033111	0.010188	
496	0.058751	0.004896	0.056303	0.036719	0.009792	0.046511	0.014688	
497	0.049735	0.009473	0.049735	0.063945	0.014210	0.054472	0.009473	
498	0.039657	0.007931	0.063451	0.039657	0.019828	0.027760	0.039657	
499	0.028891	0.000705	0.058487	0.056373	0.010570	0.042280	0.008456	

	X7	X8	X9	…	X53	X54	X55	\
0	0.016478	0.010985	0.071404	…	0.019691	0.022198	0.032943	
3	0.036218	0.012783	0.055393	…	0.044956	0.014060	-0.000637	
4	0.070492	0.052216	0.075713	…	-0.000424	-0.017321	0.024486	
5	0.032029	0.012812	0.055517	…	0.008540	-0.005070	0.001552	
6	0.075406	0.056555	0.056555	…	-0.026920	0.038575	0.032721	
..	…	…	…	…	…	…	…	
495	0.002547	0.005094	0.025470	…	0.035161	-0.017763	0.032380	
496	0.000000	0.007344	0.031823	…	0.008957	-0.008991	-0.005183	
497	0.052103	0.037893	0.080523	…	0.008395	0.027797	0.042098	
498	0.007931	0.007931	0.047588	…	0.015408	0.002418	0.006407	
499	0.029596	0.014093	0.037347	…	0.017603	-0.003492	0.022454	

	X56	X57	X58	X59	X60	X61	X62
0	-0.024243	-0.003538	-0.011641	-0.007937	0.006225	0.038524	no_efectores
3	-0.006273	0.002905	-0.017968	0.019472	-0.011804	-0.019533	no_efectores
4	-0.027044	-0.019211	-0.040230	-0.018762	0.007330	-0.007379	no_efectores
5	0.029336	0.026545	0.040011	0.011407	-0.007961	0.006119	no_efectores
6	-0.040359	0.021787	0.029527	0.000506	-0.045486	-0.038849	no_efectores
..	…	…	…	…	…	…	…
495	0.049963	0.077684	-0.011141	0.030363	0.024903	0.023802	no_efectores
496	0.019199	0.016245	-0.006075	0.026015	-0.001361	0.029576	no_efectores
497	0.013636	0.031777	-0.045627	0.006641	-0.014451	-0.000587	no_efectores
498	0.006406	0.027289	-0.004278	0.006488	0.044815	0.068877	no_efectores
499	0.008370	0.031455	0.003974	0.016257	-0.018054	0.025129	no_efectores

[407 rows x 63 columns]

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

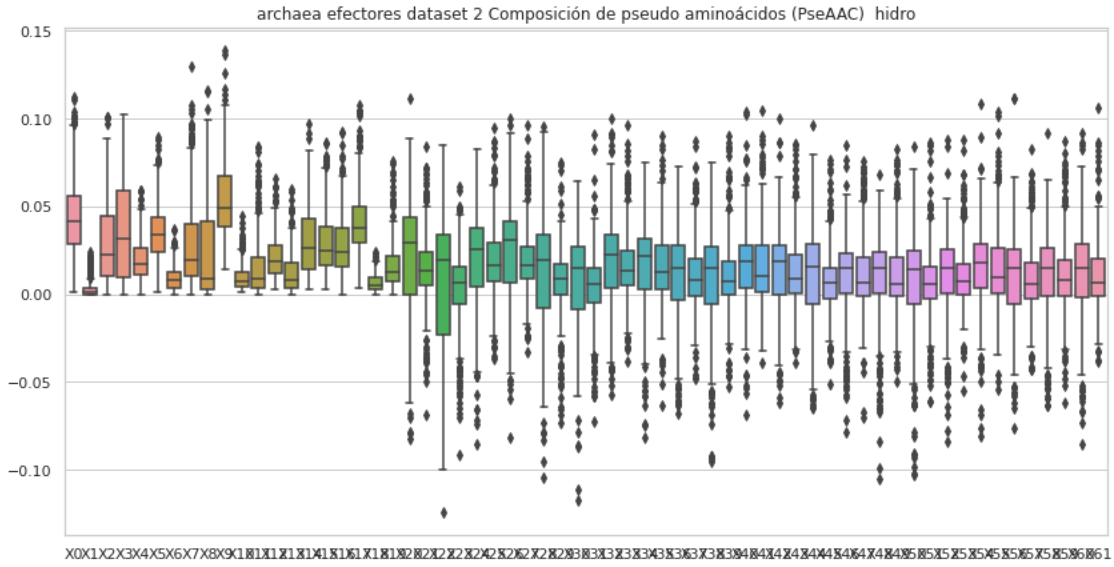
	X0	X1	X2	X3	X4	X5	\
count	407.000000	407.000000	407.000000	407.000000	407.000000	407.000000	
mean	0.054552	0.005802	0.046587	0.047693	0.019165	0.044506	
std	0.027665	0.006933	0.024643	0.025549	0.010357	0.020334	
min	0.004561	0.000000	0.000000	0.000000	0.000000	0.002785	
25%	0.032988	0.000000	0.024480	0.028990	0.012435	0.029158	
50%	0.049695	0.003854	0.049856	0.050886	0.017422	0.042053	
75%	0.073819	0.008214	0.064348	0.065787	0.023749	0.058600	
max	0.152492	0.044395	0.101662	0.135147	0.063666	0.105046	

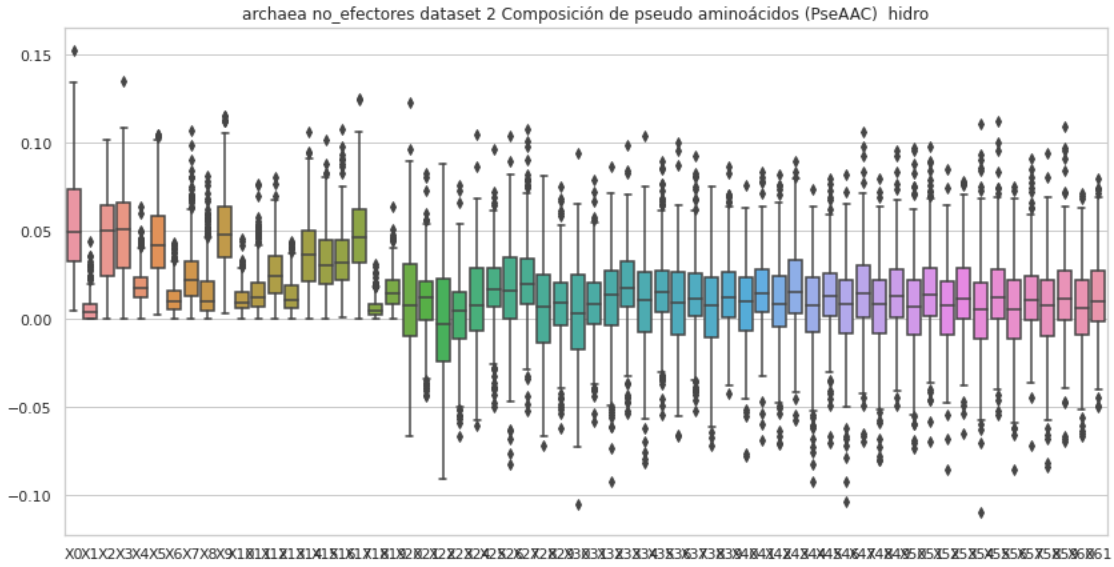
	X6	X7	X8	X9	…	X52	\
count	407.000000	407.000000	407.000000	407.000000	…	407.000000	
mean	0.011680	0.025449	0.015416	0.050113	…	0.005994	
std	0.008718	0.017950	0.016046	0.021138	…	0.022975	
min	0.000000	0.000000	0.000000	0.002948	…	-0.085558	
25%	0.005143	0.013049	0.004416	0.035327	…	-0.009049	
50%	0.009978	0.022135	0.009793	0.047676	…	0.007649	
75%	0.016156	0.032868	0.021185	0.063738	…	0.021191	
max	0.043326	0.107178	0.080978	0.115711	…	0.084831	

	X53	X54	X55	X56	X57	X58	\
count	407.000000	407.000000	407.000000	407.000000	407.000000	407.000000	
mean	0.013855	0.004461	0.014914	0.004456	0.011809	0.005937	
std	0.022252	0.027550	0.023720	0.024472	0.021124	0.025339	
min	-0.065429	-0.109794	-0.054316	-0.085213	-0.071781	-0.083773	
25%	0.000219	-0.011416	0.000203	-0.011145	-0.000350	-0.010068	
50%	0.011266	0.005379	0.011969	0.005566	0.010428	0.007717	
75%	0.029007	0.020831	0.027884	0.021858	0.024591	0.022391	
max	0.077882	0.110990	0.112727	0.075019	0.090910	0.094404	

	X59	X60	X61
count	407.000000	407.000000	407.000000
mean	0.013562	0.006109	0.012729
std	0.023555	0.024389	0.021404
min	-0.069259	-0.066527	-0.049765
25%	-0.000915	-0.008607	-0.001322
50%	0.011277	0.006504	0.010377
75%	0.027183	0.021945	0.027319
max	0.109571	0.071208	0.079602

[8 rows x 62 columns]





6 Covarianza de auto cruzamiento (ACC) hidro_mass

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

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

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

    if etiq == "efectores":
        df=ACC_hidro_mass_efec

    if etiq == "no_efectores":
        df=ACC_hidro_mass_no_efec

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

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

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

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.092583 -0.045864  0.013643 -0.018513  0.050855 -0.017522 -0.009819
1    0.080302  0.058291 -0.030669  0.000233 -0.023781  0.050628 -0.050247
2   -0.069382  0.051525  0.070289 -0.022288  0.102035 -0.112882  0.100776
3   -0.060528 -0.108142  0.108169 -0.073748 -0.163644  0.014162  0.005269
4   -0.077586  0.000537  0.013399 -0.076046  0.091719 -0.087441  0.031330
..        …         …         …         …         …         …         …
495 -0.054183 -0.036246  0.127586  0.058238 -0.124509 -0.030947  0.047807
496  0.055665 -0.082717 -0.038970  0.001607 -0.009556 -0.025194  0.051599
497  0.009076  0.049655  0.030958 -0.070982 -0.025746 -0.011156 -0.036681
498  0.036769 -0.114714 -0.117652 -0.041165  0.036207  0.010075  0.055455
499  0.026212 -0.038008 -0.000551 -0.002722 -0.055878 -0.080016  0.048680

           X7        X8        X9       X10       X11       X12        X13
0   -0.003504 -0.019897  0.059127 -0.093668 -0.016244  0.012880  efectores
```

```
1   -0.041517  0.048843  0.079395 -0.013989  0.054372  0.018889  efectores
2   -0.027805 -0.171157  0.072685 -0.014110 -0.003368 -0.083080  efectores
3   -0.001512  0.045317  0.045186  0.113467  0.022608 -0.004064  efectores
4    0.018726  0.057302  0.045060  0.040170 -0.010870 -0.070976  efectores
..        …         …         …         …         …         …
495 -0.008366 -0.021239 -0.021116 -0.006238 -0.015193  0.055699  efectores
496 -0.044144 -0.084803 -0.039022  0.032714  0.027727  0.017817  efectores
497  0.037251  0.015116  0.033526  0.043610  0.042661 -0.012675  efectores
498  0.013608 -0.059207 -0.030496  0.040891  0.059416 -0.012138  efectores
499  0.000501  0.047413 -0.097374 -0.003773  0.028219  0.076941  efectores

[500 rows x 14 columns]
```

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

```
                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.002893    0.017388    0.006960    0.013300   -0.005358    0.001517
std      0.067670    0.069455    0.069489    0.066450    0.067815    0.065990
min     -0.241955   -0.200475   -0.249135   -0.182319   -0.281667   -0.234601
25%     -0.037988   -0.025749   -0.034783   -0.029048   -0.045266   -0.037648
50%      0.004126    0.022675    0.005499    0.013018   -0.003661   -0.000158
75%      0.046165    0.059162    0.051979    0.057178    0.035733    0.045273
max      0.215634    0.233958    0.237077    0.221048    0.227624    0.194723

                X6          X7          X8          X9         X10         X11  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.014135    0.001845   -0.003304    0.006215    0.002972   -0.006638
std      0.070873    0.062416    0.071892    0.075461    0.072331    0.067548
min     -0.232398   -0.222139   -0.291102   -0.254148   -0.244905   -0.238109
25%     -0.028637   -0.034618   -0.049731   -0.042506   -0.037955   -0.044268
50%      0.011708    0.001014   -0.005502    0.006058    0.001057   -0.003046
75%      0.059816    0.042851    0.038734    0.050424    0.046318    0.031723
max      0.274666    0.225515    0.267947    0.236190    0.256147    0.210196

               X12
count  500.000000
mean     0.011736
std      0.070501
min     -0.209102
25%     -0.030270
50%      0.010773
75%      0.056306
max      0.241867
```

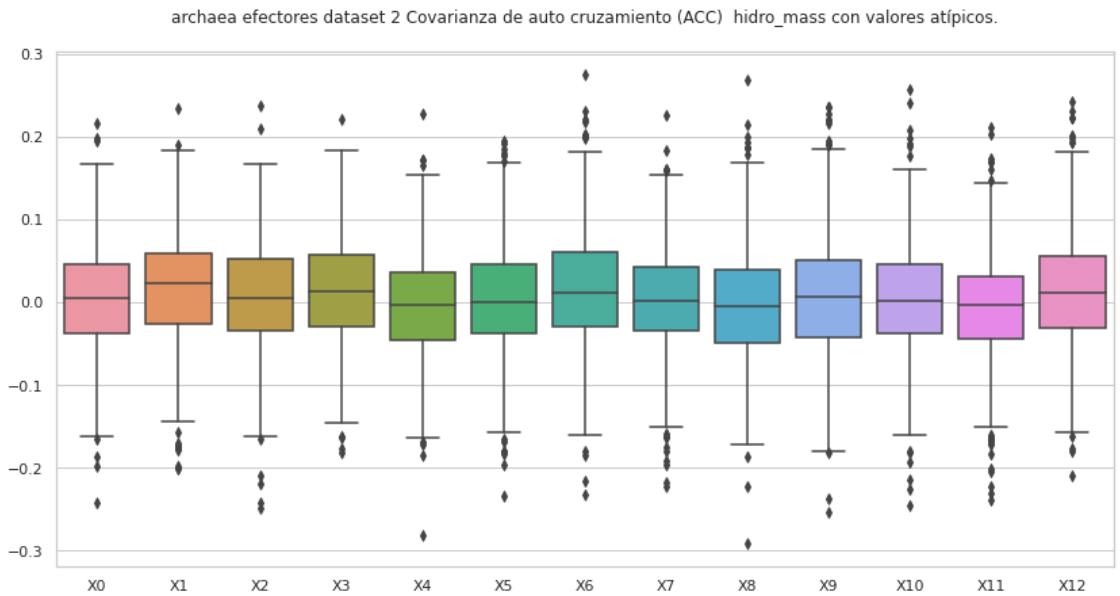
no_efectores
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores archaea dataset 2,
con valores atípicos.
Valores del documento csv.

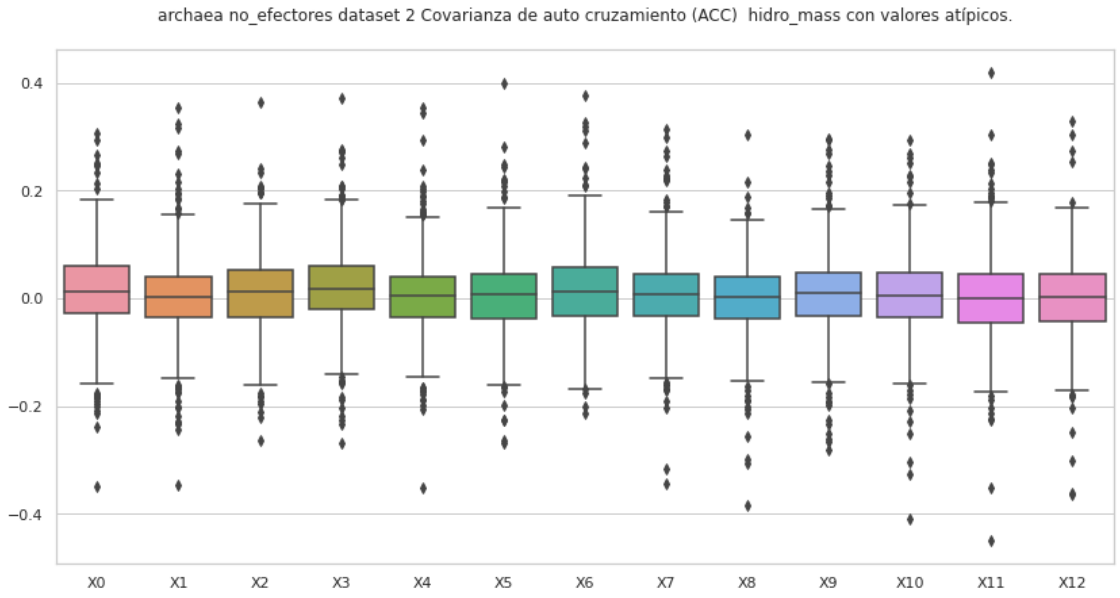
```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.101866 -0.047975  0.035152  0.070423  0.030389  0.062847 -0.023510
1    0.265035 -0.026883 -0.051708 -0.047006  0.024846 -0.159642 -0.157277
2    0.043084 -0.056615 -0.070286 -0.156307 -0.197525 -0.008673  0.115076
3   -0.027815 -0.006666 -0.014372  0.030202 -0.057483  0.072936  0.018772
4    0.001314  0.066281  0.014012 -0.031977  0.020172  0.010746 -0.002735
..        …         …         …         …         …         …
495 -0.153306 -0.010063 -0.025664  0.190976 -0.166325 -0.017941 -0.032504
496 -0.035679  0.089293  0.053457 -0.036842  0.007443  0.010630 -0.100716
497  0.052284  0.037414  0.077570 -0.017418  0.057906  0.019559 -0.048490
498  0.107626 -0.008890  0.014193  0.071589 -0.008673 -0.018127 -0.104323
499  0.006378  0.026192  0.051863  0.072080  0.125144  0.031814  0.050589

           X7        X8        X9       X10       X11       X12           X13
0   -0.064861 -0.009969 -0.063614 -0.054692 -0.033496 -0.033716  no_efectores
1   -0.148260 -0.206942 -0.006760 -0.047053 -0.126270 -0.047328  no_efectores
2    0.087422  0.187135  0.130844 -0.157570 -0.026815 -0.097310  no_efectores
3    0.010015 -0.091793  0.006307  0.060368  0.057858 -0.073815  no_efectores
4   -0.023600 -0.028569  0.017826 -0.013722  0.050688 -0.056874  no_efectores
..        …         …         …         …         …         …
495 -0.065739  0.035219  0.076857 -0.018461 -0.034240  0.034164  no_efectores
496 -0.035132  0.038993 -0.176972 -0.047886 -0.095113  0.057535  no_efectores
497 -0.041311 -0.015996 -0.050123 -0.033326 -0.025628 -0.086298  no_efectores
498 -0.134655 -0.004973  0.031016  0.108940 -0.000281  0.046934  no_efectores
499  0.053487  0.033238  0.039336  0.012097  0.023240  0.030554  no_efectores

[500 rows x 14 columns]
```

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

```
                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.015865    0.004846    0.008005    0.020436    0.004421    0.004284
std      0.080057    0.079675    0.073608    0.076359    0.073473    0.076465
min     -0.350086   -0.346003   -0.264761   -0.268109   -0.351585   -0.267414
25%     -0.027628   -0.035992   -0.035332   -0.019979   -0.035352   -0.037316
50%      0.013542    0.003066    0.012755    0.018466    0.005249    0.007436
75%      0.059170    0.040665    0.052004    0.061190    0.040692    0.045773
```

```
max      0.306850    0.353480    0.364640    0.371369    0.353323    0.399541

                X6          X7          X8          X9         X10         X11  \
count   500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean      0.013015    0.008017   -0.000138    0.009789    0.006319    0.000538
std       0.079145    0.076007    0.071090    0.077633    0.078128    0.082125
min      -0.214376   -0.344527   -0.384472   -0.280105   -0.408079   -0.449299
25%      -0.032160   -0.031900   -0.038177   -0.033898   -0.034951   -0.045564
50%       0.013612    0.006581    0.002943    0.010498    0.005098   -0.001198
75%       0.058358    0.045888    0.039329    0.047040    0.048255    0.044253
max       0.375245    0.312634    0.303440    0.296388    0.292727    0.418923

               X12
count   500.000000
mean      0.000482
std       0.077168
min      -0.364102
25%      -0.042685
50%       0.002123
75%       0.044558
max       0.327833
```





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

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

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

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

    if etiq == "efectores":
        df=ACC_hidro_mass_efec

    if etiq == "no_efectores":
        df=ACC_hidro_mass_no_efec

    del df['X13']
```

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

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

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

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

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

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.092583 -0.045864  0.013643 -0.018513  0.050855 -0.017522 -0.009819
1    0.080302  0.058291 -0.030669  0.000233 -0.023781  0.050628 -0.050247
2   -0.069382  0.051525  0.070289 -0.022288  0.102035 -0.112882  0.100776
3   -0.060528 -0.108142  0.108169 -0.073748 -0.163644  0.014162  0.005269
4   -0.077586  0.000537  0.013399 -0.076046  0.091719 -0.087441  0.031330
..        …         …         …         …         …         …         …
495 -0.054183 -0.036246  0.127586  0.058238 -0.124509 -0.030947  0.047807
496  0.055665 -0.082717 -0.038970  0.001607 -0.009556 -0.025194  0.051599
497  0.009076  0.049655  0.030958 -0.070982 -0.025746 -0.011156 -0.036681
498  0.036769 -0.114714 -0.117652 -0.041165  0.036207  0.010075  0.055455
499  0.026212 -0.038008 -0.000551 -0.002722 -0.055878 -0.080016  0.048680

           X7        X8        X9       X10       X11       X12        X13
0   -0.003504 -0.019897  0.059127 -0.093668 -0.016244  0.012880  efectores
1   -0.041517  0.048843  0.079395 -0.013989  0.054372  0.018889  efectores
2   -0.027805 -0.171157  0.072685 -0.014110 -0.003368 -0.083080  efectores
3   -0.001512  0.045317  0.045186  0.113467  0.022608 -0.004064  efectores
4    0.018726  0.057302  0.045060  0.040170 -0.010870 -0.070976  efectores
..        …         …         …         …         …         …         …
495 -0.008366 -0.021239 -0.021116 -0.006238 -0.015193  0.055699  efectores
```

```
496 -0.044144 -0.084803 -0.039022  0.032714  0.027727  0.017817  efectores
497  0.037251  0.015116  0.033526  0.043610  0.042661 -0.012675  efectores
498  0.013608 -0.059207 -0.030496  0.040891  0.059416 -0.012138  efectores
499  0.000501  0.047413 -0.097374 -0.003773  0.028219  0.076941  efectores

[462 rows x 14 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	462.000000	462.000000	462.000000	462.000000	462.000000	462.000000	
mean	0.003315	0.018545	0.007948	0.013853	-0.006849	0.001891	
std	0.064038	0.064894	0.064325	0.064477	0.065436	0.062843	
min	-0.197650	-0.178612	-0.165367	-0.182319	-0.183976	-0.195876	
25%	-0.035307	-0.023784	-0.031984	-0.027306	-0.045290	-0.035099	
50%	0.004126	0.022456	0.005499	0.013056	-0.004148	-0.000353	
75%	0.045939	0.058177	0.050368	0.056942	0.034909	0.043832	
max	0.198471	0.189312	0.208928	0.182671	0.171347	0.194723	

	X6	X7	X8	X9	X10	X11	\
count	462.000000	462.000000	462.000000	462.000000	462.000000	462.000000	
mean	0.016185	0.004352	-0.002496	0.007190	0.001088	-0.005012	
std	0.064891	0.058521	0.066918	0.071009	0.064691	0.062352	
min	-0.158551	-0.179666	-0.171157	-0.182286	-0.193585	-0.203745	
25%	-0.021316	-0.031540	-0.047966	-0.038471	-0.037598	-0.040952	
50%	0.013624	0.004261	-0.004885	0.006355	0.000391	-0.001769	
75%	0.059665	0.043215	0.037358	0.050317	0.042000	0.030719	
max	0.220704	0.182616	0.199436	0.227596	0.188031	0.173023	

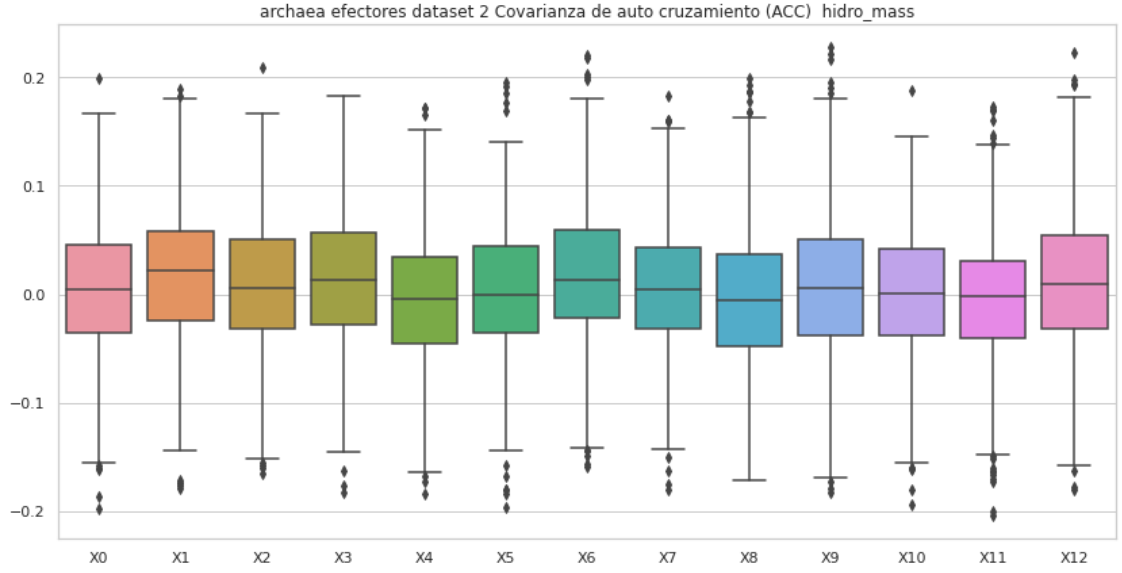
	X12
count	462.000000
mean	0.008989
std	0.067564
min	-0.179826
25%	-0.031605
50%	0.009635
75%	0.054465
max	0.222822

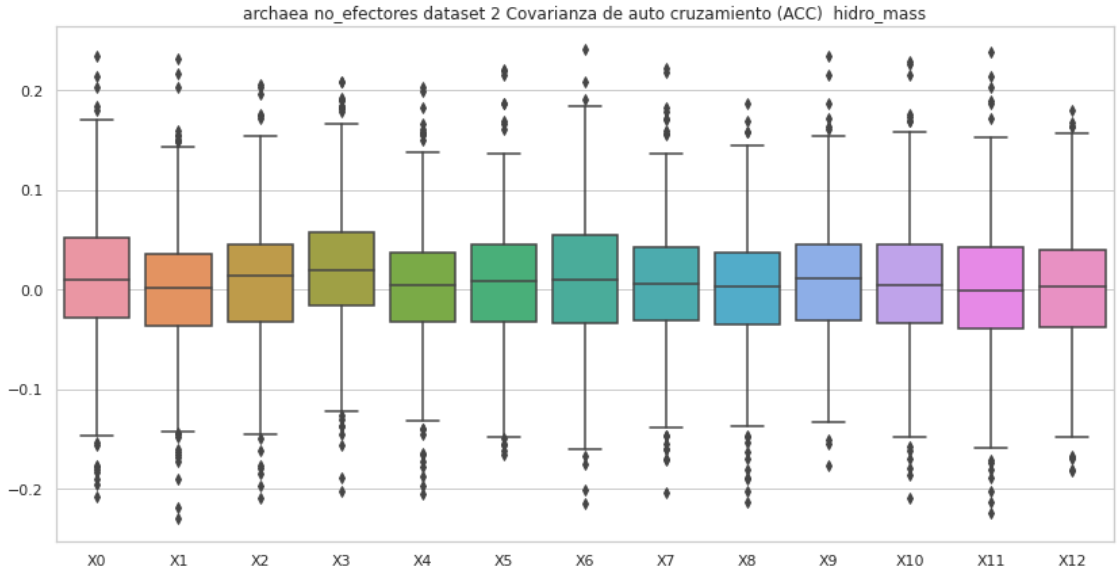
no_efectores
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores archaea dataset 2, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.101866 -0.047975  0.035152  0.070423  0.030389  0.062847 -0.023510
2    0.043084 -0.056615 -0.070286 -0.156307 -0.197525 -0.008673  0.115076
3   -0.027815 -0.006666 -0.014372  0.030202 -0.057483  0.072936  0.018772
4    0.001314  0.066281  0.014012 -0.031977  0.020172  0.010746 -0.002735
5   -0.015492 -0.002249  0.035824 -0.031419  0.030763  0.025858 -0.021329
..        …         …         …         …         …         …
495 -0.153306 -0.010063 -0.025664  0.190976 -0.166325 -0.017941 -0.032504
496 -0.035679  0.089293  0.053457 -0.036842  0.007443  0.010630 -0.100716
497  0.052284  0.037414  0.077570 -0.017418  0.057906  0.019559 -0.048490
498  0.107626 -0.008890  0.014193  0.071589 -0.008673 -0.018127 -0.104323
499  0.006378  0.026192  0.051863  0.072080  0.125144  0.031814  0.050589

           X7        X8        X9       X10       X11       X12           X13
0   -0.064861 -0.009969 -0.063614 -0.054692 -0.033496 -0.033716  no_efectores
2    0.087422  0.187135  0.130844 -0.157570 -0.026815 -0.097310  no_efectores
3    0.010015 -0.091793  0.006307  0.060368  0.057858 -0.073815  no_efectores
4   -0.023600 -0.028569  0.017826 -0.013722  0.050688 -0.056874  no_efectores
5   -0.003770 -0.012487  0.011748  0.004003  0.015876 -0.081726  no_efectores
..        …         …         …         …         …         …           …
495 -0.065739  0.035219  0.076857 -0.018461 -0.034240  0.034164  no_efectores
496 -0.035132  0.038993 -0.176972 -0.047886 -0.095113  0.057535  no_efectores
497 -0.041311 -0.015996 -0.050123 -0.033326 -0.025628 -0.086298  no_efectores
498 -0.134655 -0.004973  0.031016  0.108940 -0.000281  0.046934  no_efectores
499  0.053487  0.033238  0.039336  0.012097  0.023240  0.030554  no_efectores

[450 rows x 14 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  450.000000  450.000000  450.000000  450.000000  450.000000  450.000000
mean     0.012651    0.001689    0.007237    0.021298    0.002275    0.005282
std      0.068826    0.065121    0.064388    0.062919    0.062615    0.063240
min     -0.208194   -0.229651   -0.209805   -0.202427   -0.205836   -0.165508
25%     -0.027753   -0.036144   -0.031993   -0.015807   -0.032717   -0.032665
50%      0.009170    0.001409    0.013156    0.018788    0.004972    0.008898
75%      0.052563    0.035808    0.044748    0.057325    0.036826    0.044488
max      0.234161    0.231207    0.205205    0.208645    0.202563    0.220497

                X6          X7          X8          X9         X10         X11  \
count  450.000000  450.000000  450.000000  450.000000  450.000000  450.000000
mean     0.007711    0.005795    0.000987    0.010432    0.005092    0.000537
std      0.069797    0.062376    0.061417    0.061218    0.065584    0.068886
```

min	-0.214376	-0.203744	-0.213037	-0.176972	-0.208565	-0.224725
25%	-0.033934	-0.031695	-0.034490	-0.030785	-0.033318	-0.039030
50%	0.010364	0.006201	0.002943	0.010732	0.004372	-0.001078
75%	0.053980	0.042451	0.037455	0.044869	0.045218	0.042853
max	0.240883	0.222474	0.187135	0.234860	0.228195	0.238739

	X12
count	450.000000
mean	0.001225
std	0.063291
min	-0.182555
25%	-0.037541
50%	0.002325
75%	0.040154
max	0.179339





7 Covarianza de auto cruzamiento (ACC) mass

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

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

    if etiq == "efectores":
        df=ACC_mass_efec

    if etiq == "no_efectores":
        df=ACC_mass_no_efec

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

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

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

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.092583 -0.045864  0.013643 -0.018513  0.050855 -0.017522 -0.009819
1    0.080302  0.058291 -0.030669  0.000233 -0.023781  0.050628 -0.050247
2   -0.069382  0.051525  0.070289 -0.022288  0.102035 -0.112882  0.100776
3   -0.060528 -0.108142  0.108169 -0.073748 -0.163644  0.014162  0.005269
4   -0.077586  0.000537  0.013399 -0.076046  0.091719 -0.087441  0.031330
..        …         …         …         …         …         …         …
495 -0.054183 -0.036246  0.127586  0.058238 -0.124509 -0.030947  0.047807
496  0.055665 -0.082717 -0.038970  0.001607 -0.009556 -0.025194  0.051599
497  0.009076  0.049655  0.030958 -0.070982 -0.025746 -0.011156 -0.036681
498  0.036769 -0.114714 -0.117652 -0.041165  0.036207  0.010075  0.055455
499  0.026212 -0.038008 -0.000551 -0.002722 -0.055878 -0.080016  0.048680

           X7        X8        X9       X10       X11       X12         X13
0   -0.003504 -0.019897  0.059127 -0.093668 -0.016244  0.012880  efectores
1   -0.041517  0.048843  0.079395 -0.013989  0.054372  0.018889  efectores
2   -0.027805 -0.171157  0.072685 -0.014110 -0.003368 -0.083080  efectores
3   -0.001512  0.045317  0.045186  0.113467  0.022608 -0.004064  efectores
4    0.018726  0.057302  0.045060  0.040170 -0.010870 -0.070976  efectores
..        …         …         …         …         …         …         …
495 -0.008366 -0.021239 -0.021116 -0.006238 -0.015193  0.055699  efectores
496 -0.044144 -0.084803 -0.039022  0.032714  0.027727  0.017817  efectores
497  0.037251  0.015116  0.033526  0.043610  0.042661 -0.012675  efectores
498  0.013608 -0.059207 -0.030496  0.040891  0.059416 -0.012138  efectores
499  0.000501  0.047413 -0.097374 -0.003773  0.028219  0.076941  efectores

[500 rows x 14 columns]


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

                X0          X1          X2          X3          X4          X5  \
count   500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean      0.002893    0.017388    0.006960    0.013300   -0.005358    0.001517
```

```
std      0.067670    0.069455    0.069489    0.066450    0.067815    0.065990
min     -0.241955   -0.200475   -0.249135   -0.182319   -0.281667   -0.234601
25%     -0.037988   -0.025749   -0.034783   -0.029048   -0.045266   -0.037648
50%      0.004126    0.022675    0.005499    0.013018   -0.003661   -0.000158
75%      0.046165    0.059162    0.051979    0.057178    0.035733    0.045273
max      0.215634    0.233958    0.237077    0.221048    0.227624    0.194723

                X6          X7          X8          X9          X10         X11  \
count   500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean      0.014135    0.001845   -0.003304    0.006215    0.002972   -0.006638
std       0.070873    0.062416    0.071892    0.075461    0.072331    0.067548
min      -0.232398   -0.222139   -0.291102   -0.254148   -0.244905   -0.238109
25%      -0.028637   -0.034618   -0.049731   -0.042506   -0.037955   -0.044268
50%       0.011708    0.001014   -0.005502    0.006058    0.001057   -0.003046
75%       0.059816    0.042851    0.038734    0.050424    0.046318    0.031723
max       0.274666    0.225515    0.267947    0.236190    0.256147    0.210196

               X12
count   500.000000
mean      0.011736
std       0.070501
min      -0.209102
25%      -0.030270
50%       0.010773
75%       0.056306
max       0.241867



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

            X0        X1        X2        X3        X4        X5        X6  \
0     0.101866 -0.047975  0.035152  0.070423  0.030389  0.062847 -0.023510
1     0.265035 -0.026883 -0.051708 -0.047006  0.024846 -0.159642 -0.157277
2     0.043084 -0.056615 -0.070286 -0.156307 -0.197525 -0.008673  0.115076
3    -0.027815 -0.006666 -0.014372  0.030202 -0.057483  0.072936  0.018772
4     0.001314  0.066281  0.014012 -0.031977  0.020172  0.010746 -0.002735
..         …         …         …         …         …         …         …
495 -0.153306 -0.010063 -0.025664  0.190976 -0.166325 -0.017941 -0.032504
496 -0.035679  0.089293  0.053457 -0.036842  0.007443  0.010630 -0.100716
497  0.052284  0.037414  0.077570 -0.017418  0.057906  0.019559 -0.048490
498  0.107626 -0.008890  0.014193  0.071589 -0.008673 -0.018127 -0.104323
499  0.006378  0.026192  0.051863  0.072080  0.125144  0.031814  0.050589

            X7        X8        X9        X10       X11       X12           X13
```

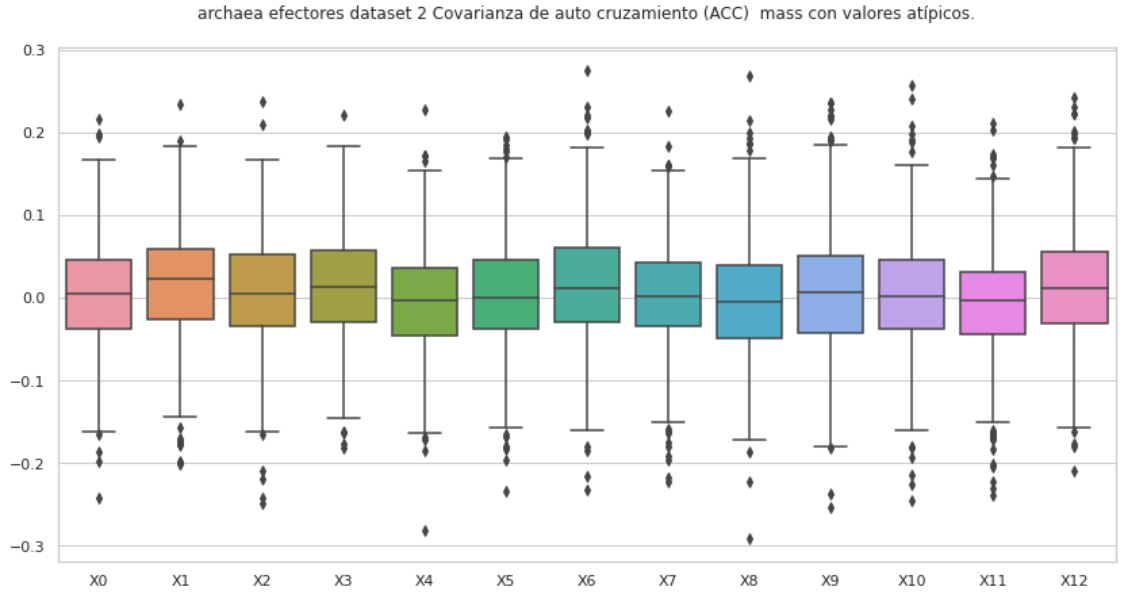
```
0   -0.064861 -0.009969 -0.063614 -0.054692 -0.033496 -0.033716  no_efectores
1   -0.148260 -0.206942 -0.006760 -0.047053 -0.126270 -0.047328  no_efectores
2    0.087422  0.187135  0.130844 -0.157570 -0.026815 -0.097310  no_efectores
3    0.010015 -0.091793  0.006307  0.060368  0.057858 -0.073815  no_efectores
4   -0.023600 -0.028569  0.017826 -0.013722  0.050688 -0.056874  no_efectores
..        …       …       …       …       …       …           …
495 -0.065739  0.035219  0.076857 -0.018461 -0.034240  0.034164  no_efectores
496 -0.035132  0.038993 -0.176972 -0.047886 -0.095113  0.057535  no_efectores
497 -0.041311 -0.015996 -0.050123 -0.033326 -0.025628 -0.086298  no_efectores
498 -0.134655 -0.004973  0.031016  0.108940 -0.000281  0.046934  no_efectores
499  0.053487  0.033238  0.039336  0.012097  0.023240  0.030554  no_efectores

[500 rows x 14 columns]
```

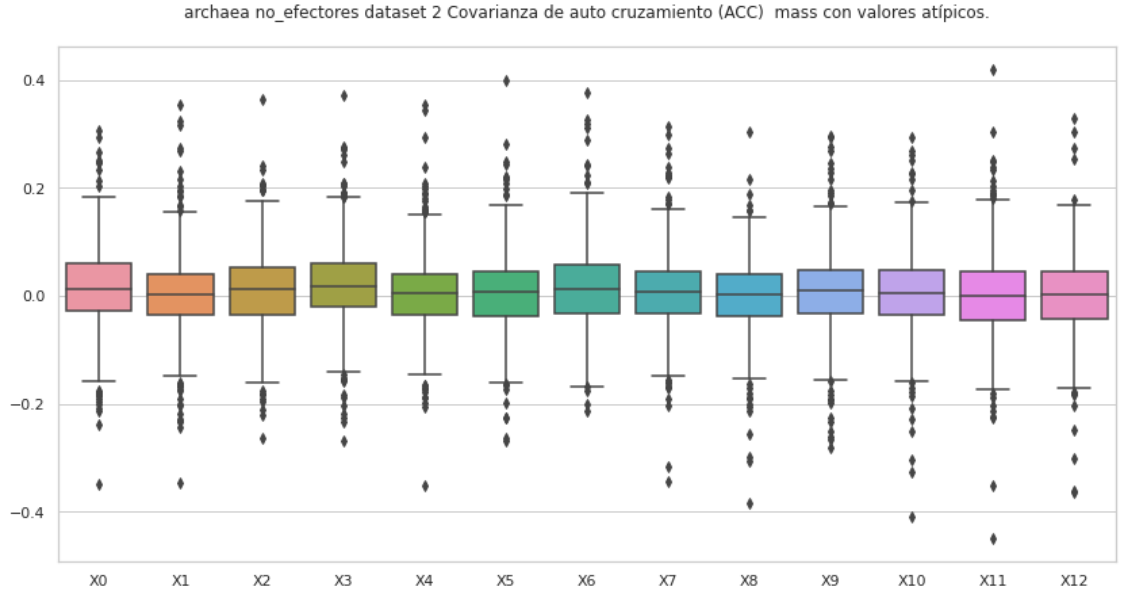
Covarianza de auto cruzamiento (ACC) mass no_efectores archaea dataset 2, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.015865	0.004846	0.008005	0.020436	0.004421	0.004284	
std	0.080057	0.079675	0.073608	0.076359	0.073473	0.076465	
min	-0.350086	-0.346003	-0.264761	-0.268109	-0.351585	-0.267414	
25%	-0.027628	-0.035992	-0.035332	-0.019979	-0.035352	-0.037316	
50%	0.013542	0.003066	0.012755	0.018466	0.005249	0.007436	
75%	0.059170	0.040665	0.052004	0.061190	0.040692	0.045773	
max	0.306850	0.353480	0.364640	0.371369	0.353323	0.399541	

	X6	X7	X8	X9	X10	X11	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.013015	0.008017	-0.000138	0.009789	0.006319	0.000538	
std	0.079145	0.076007	0.071090	0.077633	0.078128	0.082125	
min	-0.214376	-0.344527	-0.384472	-0.280105	-0.408079	-0.449299	
25%	-0.032160	-0.031900	-0.038177	-0.033898	-0.034951	-0.045564	
50%	0.013612	0.006581	0.002943	0.010498	0.005098	-0.001198	
75%	0.058358	0.045888	0.039329	0.047040	0.048255	0.044253	
max	0.375245	0.312634	0.303440	0.296388	0.292727	0.418923	

	X12
count	500.000000
mean	0.000482
std	0.077168
min	-0.364102
25%	-0.042685
50%	0.002123
75%	0.044558
max	0.327833


archaea efectores dataset 2 Covarianza de auto cruzamiento (ACC) mass con valores atípicos.
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12


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

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

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

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

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

    if etiq == "efectores":
        df=ACC_mass_efec

    if etiq == "no_efectores":
        df=ACC_mass_no_efec

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

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

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

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

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

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

	X0	X1	X2	X3	X4	X5	X6	\
0	-0.092583	-0.045864	0.013643	-0.018513	0.050855	-0.017522	-0.009819	
1	0.080302	0.058291	-0.030669	0.000233	-0.023781	0.050628	-0.050247	
2	-0.069382	0.051525	0.070289	-0.022288	0.102035	-0.112882	0.100776	
3	-0.060528	-0.108142	0.108169	-0.073748	-0.163644	0.014162	0.005269	
4	-0.077586	0.000537	0.013399	-0.076046	0.091719	-0.087441	0.031330	
..	…	…	…	…	…	…	…	
495	-0.054183	-0.036246	0.127586	0.058238	-0.124509	-0.030947	0.047807	
496	0.055665	-0.082717	-0.038970	0.001607	-0.009556	-0.025194	0.051599	
497	0.009076	0.049655	0.030958	-0.070982	-0.025746	-0.011156	-0.036681	
498	0.036769	-0.114714	-0.117652	-0.041165	0.036207	0.010075	0.055455	
499	0.026212	-0.038008	-0.000551	-0.002722	-0.055878	-0.080016	0.048680	

	X7	X8	X9	X10	X11	X12	X13
0	-0.003504	-0.019897	0.059127	-0.093668	-0.016244	0.012880	efectores
1	-0.041517	0.048843	0.079395	-0.013989	0.054372	0.018889	efectores
2	-0.027805	-0.171157	0.072685	-0.014110	-0.003368	-0.083080	efectores
3	-0.001512	0.045317	0.045186	0.113467	0.022608	-0.004064	efectores
4	0.018726	0.057302	0.045060	0.040170	-0.010870	-0.070976	efectores
..	…	…	…	…	…	…	…
495	-0.008366	-0.021239	-0.021116	-0.006238	-0.015193	0.055699	efectores
496	-0.044144	-0.084803	-0.039022	0.032714	0.027727	0.017817	efectores
497	0.037251	0.015116	0.033526	0.043610	0.042661	-0.012675	efectores
498	0.013608	-0.059207	-0.030496	0.040891	0.059416	-0.012138	efectores
499	0.000501	0.047413	-0.097374	-0.003773	0.028219	0.076941	efectores

```
[462 rows x 14 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	462.000000	462.000000	462.000000	462.000000	462.000000	462.000000	
mean	0.003315	0.018545	0.007948	0.013853	-0.006849	0.001891	
std	0.064038	0.064894	0.064325	0.064477	0.065436	0.062843	
min	-0.197650	-0.178612	-0.165367	-0.182319	-0.183976	-0.195876	
25%	-0.035307	-0.023784	-0.031984	-0.027306	-0.045290	-0.035099	
50%	0.004126	0.022456	0.005499	0.013056	-0.004148	-0.000353	
75%	0.045939	0.058177	0.050368	0.056942	0.034909	0.043832	

```
max      0.198471    0.189312    0.208928    0.182671    0.171347    0.194723

               X6          X7          X8          X9         X10         X11  \
count  462.000000  462.000000  462.000000  462.000000  462.000000  462.000000
mean     0.016185    0.004352   -0.002496    0.007190    0.001088   -0.005012
std      0.064891    0.058521    0.066918    0.071009    0.064691    0.062352
min     -0.158551   -0.179666   -0.171157   -0.182286   -0.193585   -0.203745
25%     -0.021316   -0.031540   -0.047966   -0.038471   -0.037598   -0.040952
50%      0.013624    0.004261   -0.004885    0.006355    0.000391   -0.001769
75%      0.059665    0.043215    0.037358    0.050317    0.042000    0.030719
max      0.220704    0.182616    0.199436    0.227596    0.188031    0.173023

              X12
count  462.000000
mean     0.008989
std      0.067564
min     -0.179826
25%     -0.031605
50%      0.009635
75%      0.054465
max      0.222822
```

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

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.101866 -0.047975  0.035152  0.070423  0.030389  0.062847 -0.023510
2    0.043084 -0.056615 -0.070286 -0.156307 -0.197525 -0.008673  0.115076
3   -0.027815 -0.006666 -0.014372  0.030202 -0.057483  0.072936  0.018772
4    0.001314  0.066281  0.014012 -0.031977  0.020172  0.010746 -0.002735
5   -0.015492 -0.002249  0.035824 -0.031419  0.030763  0.025858 -0.021329
..        …         …         …         …         …         …
495 -0.153306 -0.010063 -0.025664  0.190976 -0.166325 -0.017941 -0.032504
496 -0.035679  0.089293  0.053457 -0.036842  0.007443  0.010630 -0.100716
497  0.052284  0.037414  0.077570 -0.017418  0.057906  0.019559 -0.048490
498  0.107626 -0.008890  0.014193  0.071589 -0.008673 -0.018127 -0.104323
499  0.006378  0.026192  0.051863  0.072080  0.125144  0.031814  0.050589

           X7        X8        X9       X10       X11       X12           X13
0   -0.064861 -0.009969 -0.063614 -0.054692 -0.033496 -0.033716  no_efectores
2    0.087422  0.187135  0.130844 -0.157570 -0.026815 -0.097310  no_efectores
3    0.010015 -0.091793  0.006307  0.060368  0.057858 -0.073815  no_efectores
4   -0.023600 -0.028569  0.017826 -0.013722  0.050688 -0.056874  no_efectores
5   -0.003770 -0.012487  0.011748  0.004003  0.015876 -0.081726  no_efectores
..        …         …         …         …         …         …           …
```

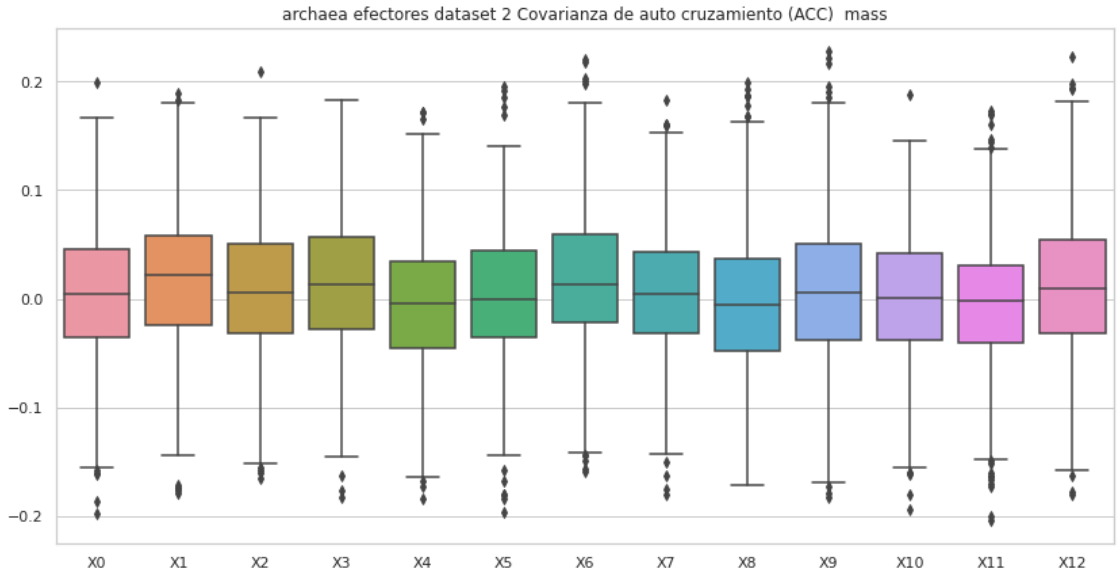
```
495 -0.065739  0.035219  0.076857 -0.018461 -0.034240  0.034164  no_efectores
496 -0.035132  0.038993 -0.176972 -0.047886 -0.095113  0.057535  no_efectores
497 -0.041311 -0.015996 -0.050123 -0.033326 -0.025628 -0.086298  no_efectores
498 -0.134655 -0.004973  0.031016  0.108940 -0.000281  0.046934  no_efectores
499  0.053487  0.033238  0.039336  0.012097  0.023240  0.030554  no_efectores

[450 rows x 14 columns]
```

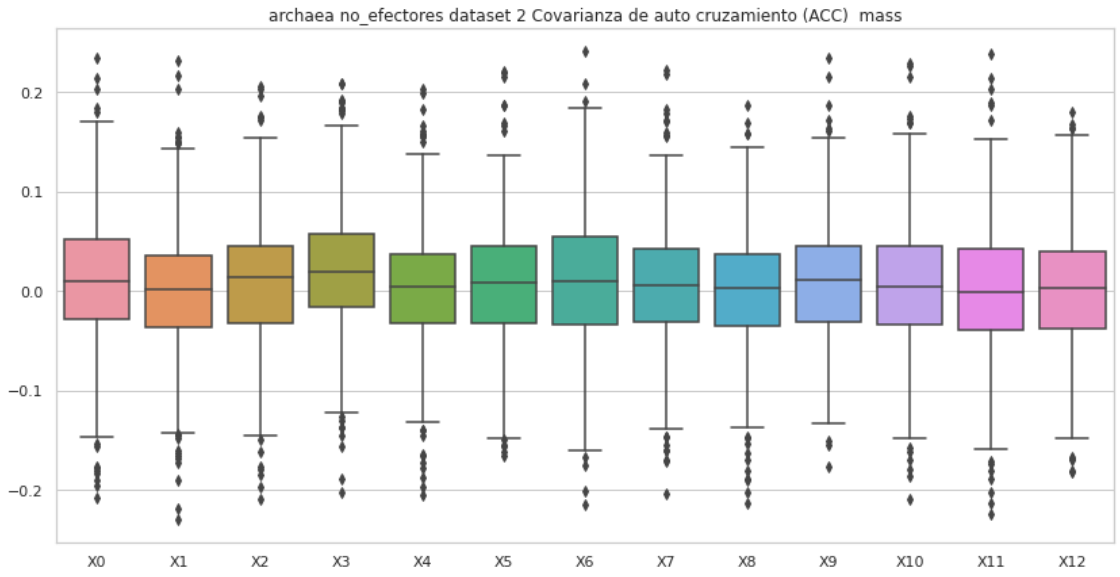
Covarianza de auto cruzamiento (ACC) mass no_efectores archaea dataset 2, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5
count	450.000000	450.000000	450.000000	450.000000	450.000000	450.000000
mean	0.012651	0.001689	0.007237	0.021298	0.002275	0.005282
std	0.068826	0.065121	0.064388	0.062919	0.062615	0.063240
min	-0.208194	-0.229651	-0.209805	-0.202427	-0.205836	-0.165508
25%	-0.027753	-0.036144	-0.031993	-0.015807	-0.032717	-0.032665
50%	0.009170	0.001409	0.013156	0.018788	0.004972	0.008898
75%	0.052563	0.035808	0.044748	0.057325	0.036826	0.044488
max	0.234161	0.231207	0.205205	0.208645	0.202563	0.220497

	X6	X7	X8	X9	X10	X11
count	450.000000	450.000000	450.000000	450.000000	450.000000	450.000000
mean	0.007711	0.005795	0.000987	0.010432	0.005092	0.000537
std	0.069797	0.062376	0.061417	0.061218	0.065584	0.068886
min	-0.214376	-0.203744	-0.213037	-0.176972	-0.208565	-0.224725
25%	-0.033934	-0.031695	-0.034490	-0.030785	-0.033318	-0.039030
50%	0.010364	0.006201	0.002943	0.010732	0.004372	-0.001078
75%	0.053980	0.042451	0.037455	0.044869	0.045218	0.042853
max	0.240883	0.222474	0.187135	0.234860	0.228195	0.238739

	X12
count	450.000000
mean	0.001225
std	0.063291
min	-0.182555
25%	-0.037541
50%	0.002325
75%	0.040154
max	0.179339





8 Covarianza de auto cruzamiento (ACC) hidro

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

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

    if etiq == "efectores":
        df=ACC_hidro_efec

    if etiq == "no_efectores":
        df=ACC_hidro_no_efec

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

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

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

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.148376 -0.238817  0.236870  0.120488 -0.070904 -0.075873  0.062061
1    0.082591  0.105927 -0.061563  0.133595  0.101711 -0.048840  0.065995
2    0.117382  0.096093  0.136032  0.123158  0.083386 -0.060564 -0.076272
3    0.040503 -0.254903  0.079351 -0.142513 -0.118800  0.001029 -0.093562
4   -0.174099 -0.166784  0.014806  0.045433  0.144467 -0.031648  0.053160
..        …       …         …         …         …         …         …
495  0.024180 -0.035491  0.135559  0.111512  0.021861  0.042879  0.106252
496  0.044231 -0.118907  0.002048  0.202256 -0.229334 -0.230857  0.019316
497 -0.064393  0.060393 -0.005137  0.005766 -0.046718 -0.000246 -0.007293
498  0.045470 -0.192066  0.094580 -0.034074 -0.113922 -0.102054 -0.090961
499 -0.006717  0.053547  0.100184  0.046393 -0.137337 -0.119125 -0.008248

           X7        X8        X9       X10       X11       X12        X13
0    0.080234  0.085404 -0.019421  0.043211  0.099347  0.143774  efectores
1    0.039520  0.084104  0.037462 -0.005013 -0.053932 -0.090092  efectores
2   -0.013336 -0.002601 -0.098587 -0.003031  0.102595 -0.058341  efectores
```

```
3    0.140563 -0.048618 -0.027596  0.049749 -0.004146  0.000667  efectores
4   -0.015053 -0.117518  0.054280 -0.010014  0.060151 -0.061002  efectores
..        …       …        …        …        …        …
495  0.052976  0.065099  0.038343  0.048026  0.017318  0.000780  efectores
496 -0.104254  0.080629  0.092747 -0.015615 -0.040945  0.036360  efectores
497  0.034340  0.051147  0.059194 -0.002056 -0.013918  0.006600  efectores
498 -0.068746 -0.071140  0.144904  0.025489  0.055052  0.115384  efectores
499 -0.213015 -0.021861  0.011153  0.041967  0.043315  0.085829  efectores

[500 rows x 14 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.015974	-0.031572	0.038555	0.035222	-0.019182	-0.019862	
std	0.087542	0.102333	0.081884	0.078588	0.092634	0.081264	
min	-0.388992	-0.393791	-0.215363	-0.224891	-0.365089	-0.268584	
25%	-0.027490	-0.094633	-0.014268	-0.011002	-0.074831	-0.076021	
50%	0.020945	-0.021167	0.035179	0.038312	-0.014565	-0.016355	
75%	0.073142	0.043467	0.082109	0.080834	0.042790	0.031346	
max	0.259757	0.247248	0.342199	0.283277	0.221750	0.246616	

	X6	X7	X8	X9	X10	X11	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.025424	0.018975	0.001522	-0.000939	0.016659	0.009584	
std	0.087923	0.080841	0.084446	0.087953	0.082803	0.077457	
min	-0.363646	-0.213015	-0.325120	-0.313101	-0.383855	-0.219232	
25%	-0.025229	-0.031922	-0.048690	-0.040688	-0.028837	-0.040954	
50%	0.023023	0.017826	0.002502	-0.001830	0.009506	0.002687	
75%	0.072446	0.061482	0.046627	0.045974	0.060827	0.054576	
max	0.391443	0.343253	0.396656	0.412087	0.289762	0.346977	

	X12
count	500.000000
mean	-0.017535
std	0.081903
min	-0.377806
25%	-0.061787
50%	-0.015191
75%	0.029114
max	0.232792

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

	X0	X1	X2	X3	X4	X5	X6	\
0	-0.037977	-0.057671	0.071618	-0.063457	-0.043556	-0.016981	-0.016343	
1	-0.180581	0.289155	0.075671	-0.145514	0.158724	-0.235105	-0.412641	
2	-0.004894	0.051031	0.068440	0.024537	-0.309559	-0.046374	-0.033137	
3	-0.044422	-0.075573	-0.003574	0.010354	-0.014309	-0.029749	-0.024903	
4	0.075881	-0.031384	0.063547	0.032540	0.018128	-0.046458	-0.143017	
..	…	…	…	…	…	…	…	
495	0.005881	-0.190253	-0.072287	0.092624	-0.113838	-0.242464	0.104073	
496	-0.029016	0.072848	0.086011	-0.079648	0.122966	-0.111799	-0.027096	
497	-0.033505	-0.060407	-0.114517	0.077742	0.063421	-0.183964	-0.082957	
498	-0.050303	-0.039962	-0.001686	0.093187	0.085198	-0.107833	0.049092	
499	-0.121185	-0.052492	0.059996	0.062652	-0.017400	-0.061112	0.082213	

	X7	X8	X9	X10	X11	X12	X13
0	0.017766	0.002461	0.014854	-0.010810	0.007139	-0.010153	no_efectores
1	0.109784	-0.408795	0.131708	-0.222651	0.045091	0.176635	no_efectores
2	-0.008640	0.050340	0.066804	-0.079045	-0.057135	0.016974	no_efectores
3	0.031089	0.009579	-0.032215	0.057303	0.009157	-0.034880	no_efectores
4	-0.007946	0.045881	0.043740	-0.004880	0.083924	0.001590	no_efectores
..	…	…	…	…	…	…	…
495	0.205932	0.011709	-0.260251	0.070887	0.101423	-0.046988	no_efectores
496	0.071684	0.062803	-0.055477	0.069317	0.067106	0.115015	no_efectores
497	0.036369	0.063960	0.055351	-0.059058	-0.044264	-0.058041	no_efectores
498	0.062396	0.068218	-0.052184	-0.129140	-0.010564	0.059297	no_efectores
499	-0.011256	-0.011757	0.006303	0.013290	0.007240	-0.073189	no_efectores

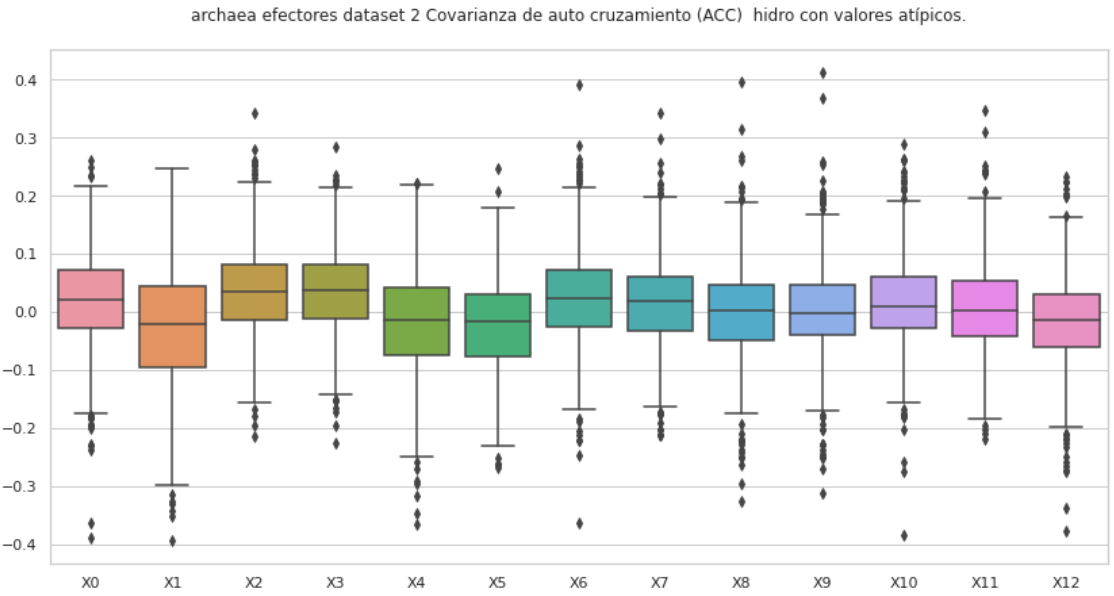
[500 rows x 14 columns]

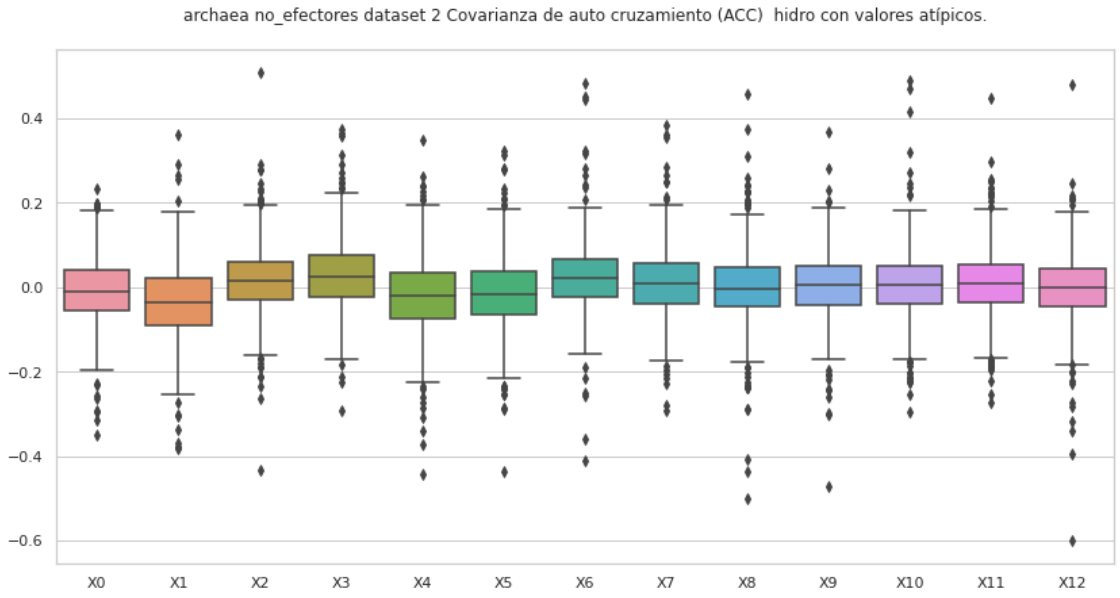
Covarianza de auto cruzamiento (ACC) hidro no_efectores archaea dataset 2, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	-0.010172	-0.037403	0.016240	0.027949	-0.021878	-0.012824	
std	0.082535	0.094372	0.084612	0.086152	0.092426	0.089509	
min	-0.351411	-0.383729	-0.431934	-0.293262	-0.441822	-0.436337	
25%	-0.055980	-0.091584	-0.029214	-0.023408	-0.073844	-0.066586	
50%	-0.010425	-0.037876	0.014964	0.025236	-0.020111	-0.017057	
75%	0.039732	0.019848	0.060858	0.077197	0.034059	0.037649	
max	0.233294	0.360928	0.507910	0.373433	0.349133	0.321783	

	X6	X7	X8	X9	X10	X11	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.020900	0.010089	-0.003474	0.002883	0.005483	0.011562	
std	0.088002	0.085729	0.092913	0.082405	0.087791	0.084197	
min	-0.412641	-0.292722	-0.499334	-0.471643	-0.296999	-0.273690	
25%	-0.024896	-0.039548	-0.047237	-0.041441	-0.040089	-0.035793	
50%	0.021932	0.009309	-0.004099	0.005631	0.006343	0.008991	
75%	0.065457	0.056172	0.045680	0.050421	0.051588	0.053649	
max	0.481138	0.383059	0.457338	0.366300	0.488181	0.447558	

	X12
count	500.000000
mean	-0.005109
std	0.086994
min	-0.600080
25%	-0.046743
50%	0.000436
75%	0.045071
max	0.479034





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

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

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

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

    if etiq == "efectores":
        df=ACC_hidro_efec

    if etiq == "no_efectores":
        df=ACC_hidro_no_efec

    del df['X13']
```

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

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

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

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

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

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.148376 -0.238817  0.236870  0.120488 -0.070904 -0.075873  0.062061
1    0.082591  0.105927 -0.061563  0.133595  0.101711 -0.048840  0.065995
2    0.117382  0.096093  0.136032  0.123158  0.083386 -0.060564 -0.076272
3    0.040503 -0.254903  0.079351 -0.142513 -0.118800  0.001029 -0.093562
4   -0.174099 -0.166784  0.014806  0.045433  0.144467 -0.031648  0.053160
..        …       …       …       …       …       …       …
495  0.024180 -0.035491  0.135559  0.111512  0.021861  0.042879  0.106252
496  0.044231 -0.118907  0.002048  0.202256 -0.229334 -0.230857  0.019316
497 -0.064393  0.060393 -0.005137  0.005766 -0.046718 -0.000246 -0.007293
498  0.045470 -0.192066  0.094580 -0.034074 -0.113922 -0.102054 -0.090961
499 -0.006717  0.053547  0.100184  0.046393 -0.137337 -0.119125 -0.008248

           X7        X8        X9       X10       X11       X12        X13
0    0.080234  0.085404 -0.019421  0.043211  0.099347  0.143774  efectores
1    0.039520  0.084104  0.037462 -0.005013 -0.053932 -0.090092  efectores
2   -0.013336 -0.002601 -0.098587 -0.003031  0.102595 -0.058341  efectores
3    0.140563 -0.048618 -0.027596  0.049749 -0.004146  0.000667  efectores
4   -0.015053 -0.117518  0.054280 -0.010014  0.060151 -0.061002  efectores
..        …       …       …       …       …       …        …
495  0.052976  0.065099  0.038343  0.048026  0.017318  0.000780  efectores
```

```
496 -0.104254  0.080629  0.092747 -0.015615 -0.040945  0.036360  efectores
497  0.034340  0.051147  0.059194 -0.002056 -0.013918  0.006600  efectores
498 -0.068746 -0.071140  0.144904  0.025489  0.055052  0.115384  efectores
499 -0.213015 -0.021861  0.011153  0.041967  0.043315  0.085829  efectores

[463 rows x 14 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	463.000000	463.000000	463.000000	463.000000	463.000000	463.000000	
mean	0.019381	-0.025452	0.038955	0.036738	-0.014222	-0.020633	
std	0.081876	0.095313	0.077482	0.074221	0.087229	0.077492	
min	-0.238035	-0.331278	-0.196158	-0.172986	-0.295009	-0.262518	
25%	-0.023268	-0.084803	-0.012066	-0.008231	-0.068353	-0.075532	
50%	0.024180	-0.016846	0.036101	0.038666	-0.010382	-0.017264	
75%	0.073382	0.044710	0.081314	0.079970	0.044853	0.030517	
max	0.259757	0.213884	0.279991	0.236152	0.221750	0.207437	

	X6	X7	X8	X9	X10	X11	\
count	463.000000	463.000000	463.000000	463.000000	463.000000	463.000000	
mean	0.026999	0.017919	0.000887	-0.001912	0.015090	0.008992	
std	0.080150	0.073591	0.073247	0.075103	0.074001	0.069333	
min	-0.220689	-0.213015	-0.249241	-0.236754	-0.202894	-0.210099	
25%	-0.019953	-0.030720	-0.046711	-0.038690	-0.028091	-0.038001	
50%	0.024469	0.017744	0.002642	-0.002521	0.005970	0.003579	
75%	0.071141	0.058938	0.045632	0.042236	0.055884	0.052978	
max	0.286704	0.240401	0.216923	0.258115	0.263213	0.236755	

	X12
count	463.000000
mean	-0.013034
std	0.071919
min	-0.249062
25%	-0.057971
50%	-0.013830
75%	0.028853
max	0.223611

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

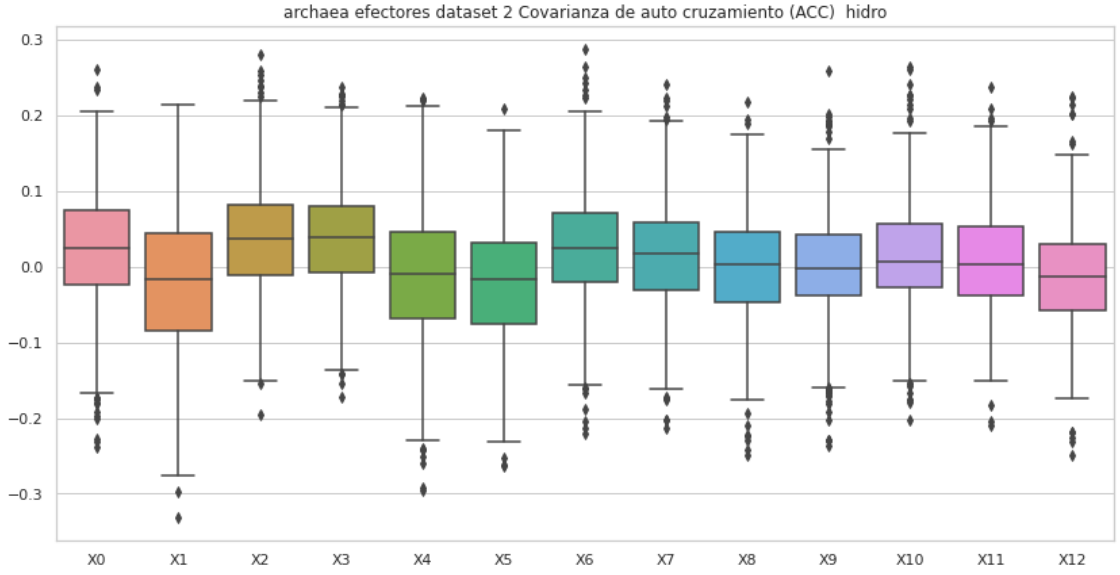
```
           X0        X1        X2        X3        X4        X5        X6  \
0   -0.037977 -0.057671  0.071618 -0.063457 -0.043556 -0.016981 -0.016343
3   -0.044422 -0.075573 -0.003574  0.010354 -0.014309 -0.029749 -0.024903
4    0.075881 -0.031384  0.063547  0.032540  0.018128 -0.046458 -0.143017
5    0.026013 -0.058595 -0.041030 -0.004315 -0.002690 -0.029059  0.027754
6    0.037357 -0.044045 -0.017960  0.129858 -0.070056 -0.005917  0.033265
..        …        …        …        …        …        …        …
494 -0.029526 -0.096795 -0.023290  0.045928 -0.057312 -0.079805 -0.006048
496 -0.029016  0.072848  0.086011 -0.079648  0.122966 -0.111799 -0.027096
497 -0.033505 -0.060407 -0.114517  0.077742  0.063421 -0.183964 -0.082957
498 -0.050303 -0.039962 -0.001686  0.093187  0.085198 -0.107833  0.049092
499 -0.121185 -0.052492  0.059996  0.062652 -0.017400 -0.061112  0.082213

           X7        X8        X9       X10       X11       X12           X13
0    0.017766  0.002461  0.014854 -0.010810  0.007139 -0.010153  no_efectores
3    0.031089  0.009579 -0.032215  0.057303  0.009157 -0.034880  no_efectores
4   -0.007946  0.045881  0.043740 -0.004880  0.083924  0.001590  no_efectores
5   -0.095910 -0.022275 -0.033992 -0.091332 -0.013276  0.076638  no_efectores
6    0.004486 -0.006467  0.001813  0.042593  0.135349 -0.023188  no_efectores
..        …        …        …        …        …        …        …
494 -0.061052  0.075498  0.039663  0.033189  0.004344  0.131894  no_efectores
496  0.071684  0.062803 -0.055477  0.069317  0.067106  0.115015  no_efectores
497  0.036369  0.063960  0.055351 -0.059058 -0.044264 -0.058041  no_efectores
498  0.062396  0.068218 -0.052184 -0.129140 -0.010564  0.059297  no_efectores
499 -0.011256 -0.011757  0.006303  0.013290  0.007240 -0.073189  no_efectores

[459 rows x 14 columns]
```

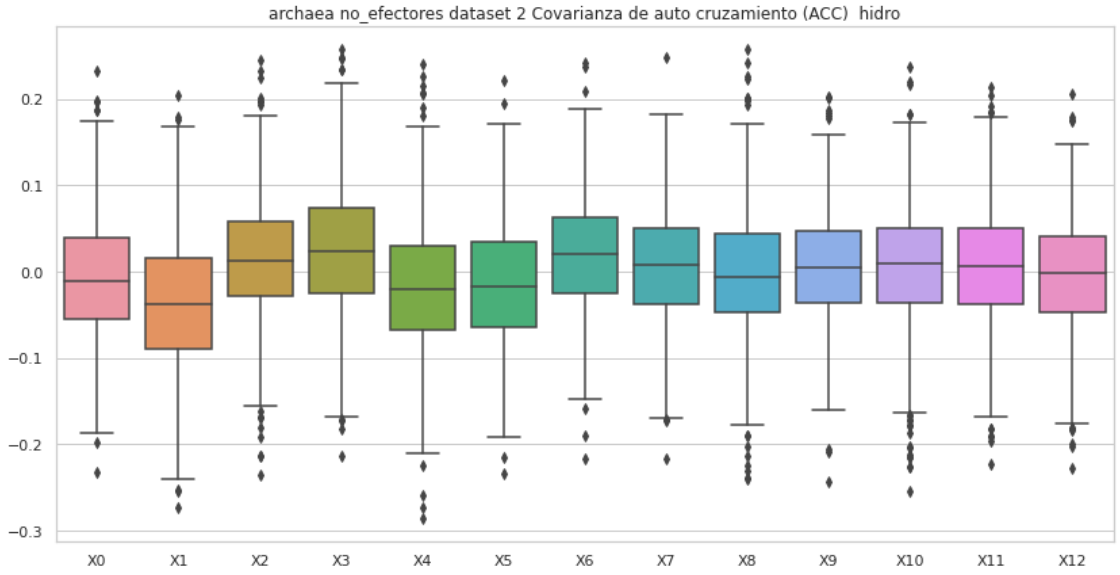
Covarianza de auto cruzamiento (ACC) no_efectores archaea dataset 2, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  459.000000  459.000000  459.000000  459.000000  459.000000  459.000000
mean    -0.006546   -0.035977    0.014773    0.024437   -0.019147   -0.013651
std      0.074578    0.080669    0.072106    0.076704    0.079681    0.073158
min     -0.232879   -0.273248   -0.235804   -0.213935   -0.285661   -0.234431
25%     -0.054713   -0.089537   -0.027500   -0.024155   -0.067540   -0.064752
50%     -0.010248   -0.038057    0.013466    0.024017   -0.019487   -0.017312
75%      0.039107    0.015708    0.058022    0.073247    0.030297    0.034425
max      0.233294    0.204767    0.245199    0.257675    0.240222    0.221600

                X6          X7          X8          X9         X10         X11  \
count  459.000000  459.000000  459.000000  459.000000  459.000000  459.000000
mean     0.018922    0.007217   -0.003554    0.004054    0.003624    0.006102
std      0.071299    0.072102    0.075270    0.067666    0.075175    0.073022
```

min	-0.216219	-0.215936	-0.239553	-0.243357	-0.254305	-0.222359
25%	-0.024400	-0.037430	-0.046295	-0.036263	-0.035600	-0.037569
50%	0.020195	0.008050	-0.006137	0.005657	0.009139	0.006643
75%	0.062699	0.050696	0.043552	0.046847	0.050169	0.049821
max	0.242055	0.248405	0.257502	0.202380	0.236908	0.214539

	X12
count	459.000000
mean	-0.005240
std	0.068184
min	-0.227801
25%	-0.046445
50%	-0.000925
75%	0.041635
max	0.205521




archaea no_efectores dataset 2 Covarianza de auto cruzamiento (ACC) hidro
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