ds1_Heterodera_limpieza_de_datos

February 1, 2021

Limpieza de datos

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

1 Declaración de variables

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

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

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

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

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

2 Composición de aminoácidos (AAC)

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

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

    if etiq == "efectores":
        df=AAC_efec

    if etiq == "no_efectores":
        df=AAC_no_efec

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

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

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

        X0      X1     X2     X3     X4      X5     X6      X7     X8      X9  \
0   10.662   2.941  6.250  4.412  2.941   6.985  5.882   5.147  0.735   4.044
1    8.571  14.286  1.714  0.571  4.571   3.429  2.286   5.143  1.714   4.571
2   12.069   6.034  4.310  5.172  0.000  14.655  7.759   1.724  0.862   2.586
3    3.030   2.273  9.091  6.061  3.788   3.788  4.545   3.788  2.273   4.545
4    3.815   4.632  9.537  5.177  3.270   4.905  3.815  10.899  2.180  12.534
..     …       …      …      …      …       …      …       …      …       …
76  12.230   5.755  5.036  3.597  0.000   3.597  2.158  10.072  2.878   5.036
77   5.056   5.618  3.933  6.180  2.247   9.551  2.809   5.056  0.562   5.618
78   6.531   4.082  4.082  3.673  2.449   2.449  2.857   6.939  0.408   8.980
79   9.172   5.325  4.438  3.846  2.959   7.396  4.142   6.509  1.775   5.325
80   7.660   4.255  7.234  3.830  1.702   4.681  2.979   7.234  1.702   5.106
```

```
    …     X11    X12    X13    X14     X15    X16    X17    X18    X19  \
0   …  10.662  3.676  5.515  2.206   5.515  5.515  0.735  1.103  6.618
1   …   6.857  4.000  4.000  2.286  12.571  6.286  2.286  0.571  6.286
2   …   6.897  3.448  4.310  0.862   9.483  4.310  0.000  1.724  6.897
3   …   8.333  2.273  4.545  5.303   5.303  8.333  0.758  4.545  8.333
4   …   4.632  2.725  1.635  2.452  11.444  3.270  0.000  1.635  4.905
..  …     …    …    …    …     …    …    …    …    …
76  …   7.194  2.878  2.878  7.914   8.633  6.475  0.000  0.000  4.317
77  …  12.921  1.685  2.247  9.551   7.303  8.989  0.000  1.124  3.933
78  …   6.122  2.449  5.306  4.082   4.898  6.531  2.857  6.531  6.939
79  …   7.988  3.254  4.734  6.509   5.325  4.734  0.592  1.479  4.734
80  …   3.404  2.979  5.532  5.106   7.660  5.532  0.426  1.277  8.511

           X20
0    efectores
1    efectores
2    efectores
3    efectores
4    efectores
..         …
76   efectores
77   efectores
78   efectores
79   efectores
80   efectores

[81 rows x 21 columns]


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

              X0         X1         X2         X3         X4         X5  \
count  81.000000  81.000000  81.000000  81.000000  81.000000  81.000000
mean    7.749864   5.444457   5.213407   5.193407   2.258358   6.962333
std     3.342051   3.643674   2.165037   2.294265   1.836720   3.148565
min     1.562000   0.599000   0.704000   0.000000   0.000000   0.000000
25%     5.056000   2.941000   3.750000   3.846000   0.803000   4.681000
50%     7.489000   5.291000   4.528000   5.221000   1.970000   7.171000
75%     9.172000   6.707000   6.522000   6.218000   2.959000   9.346000
max    17.105000  25.385000  10.638000  12.895000   9.211000  14.655000

              X6         X7         X8         X9        X10        X11  \
count  81.000000  81.000000  81.000000  81.000000  81.000000  81.000000
mean    4.065617   7.092062   1.776691   4.834914   7.424444   8.386012
std     1.808155   3.542836   1.385771   2.283387   2.783518   4.046048
```

```
min     0.769000   1.724000   0.000000   0.000000   1.992000   0.427000
25%     2.857000   5.147000   0.617000   3.252000   5.525000   5.469000
50%     3.846000   6.075000   1.714000   5.012000   6.897000   7.460000
75%     5.046000   8.082000   2.488000   5.882000   9.312000  12.351000
max    10.359000  24.219000   8.099000  12.534000  15.464000  21.127000

             X12        X13        X14        X15        X16        X17  \
count  81.000000  81.000000  81.000000  81.000000  81.000000  81.000000
mean    2.635790   4.196889   5.124519   7.450037   5.828160   0.854321
std     1.085574   2.467956   2.709287   3.313887   2.570675   0.980050
min     0.775000   0.000000   0.000000   1.829000   1.064000   0.000000
25%     1.685000   2.381000   3.125000   5.515000   4.464000   0.000000
50%     2.591000   3.750000   4.478000   6.630000   5.515000   0.592000
75%     3.257000   5.419000   6.805000   7.792000   7.389000   1.493000
max     5.488000  13.830000  12.150000  18.085000  16.154000   3.670000

             X18        X19
count  81.000000  81.000000
mean    2.052062   5.456691
std     1.674045   2.243983
min     0.000000   0.704000
25%     1.093000   4.167000
50%     1.434000   5.093000
75%     3.163000   6.739000
max     8.485000  12.308000
```

no_efectores

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

Valores del documento csv.

```
        X0     X1     X2     X3     X4      X5     X6      X7     X8      X9  \
0    6.452  4.301  3.943  9.319  1.075  12.545  2.509   5.018  1.792   5.376
1    0.901  0.000  3.604  0.901  0.000   3.604  0.000   1.802  0.901   4.505
2    6.916  2.948  4.762  6.009  1.361   7.256  3.175   4.762  2.381   4.308
3   12.048  3.614  4.819  3.614  1.205   1.205  4.819   6.024  2.410   0.000
4    5.298  0.662  1.987  2.649  1.987   0.662  1.325  11.258  3.311  10.596
..     …      …      …      …      …       …      …       …      …       …
76   9.719  1.512  6.479  3.024  2.160   2.592  4.752   9.935  2.160   5.616
77   2.174  2.174  1.812  2.174  0.725   3.261  1.087   1.812  0.725   6.522
78   6.723  0.840  1.681  3.361  0.840   0.000  0.840  10.084  2.521  11.765
79   8.286  4.857  2.857  5.143  1.714   7.714  3.429   6.571  2.286   6.857
80   9.562  4.781  4.382  7.968  0.797  10.757  5.976   5.179  0.797   3.586

    …     X11    X12     X13    X14     X15    X16    X17    X18     X19  \
0   …  12.903  1.434   4.659  2.509   5.376  4.659  1.075  2.867   6.093
```

```
1   …   2.703  4.505  43.243  0.000   6.306  2.703  2.703  2.703   1.802
2   …   8.163  2.721   5.896  3.628   7.370  5.556  1.814  3.288   6.349
3   …   6.024  4.819   7.229  7.229   4.819  6.024  0.000  2.410   7.229
4   …   2.649  3.311  11.258  3.311   5.298  3.974  2.649  5.298   9.934
..  …     …    …      …      …       …      …      …      …       …
76  …   4.968  2.160   2.160  7.343  10.367  4.752  2.808  4.104   6.911
77  …   2.174  1.812  35.145  1.449  10.145  1.087  0.362  3.986   3.623
78  …   2.521  4.202  10.084  1.681   9.244  3.361  2.521  5.042  10.084
79  …   5.429  4.857   3.714  5.143   6.571  5.714  1.143  4.000   6.286
80  …   8.367  2.789   1.992  1.992   6.773  3.586  0.797  4.781   5.179

              X20
0    no_efectores
1    no_efectores
2    no_efectores
3    no_efectores
4    no_efectores
..            …
76   no_efectores
77   no_efectores
78   no_efectores
79   no_efectores
80   no_efectores

[81 rows x 21 columns]
```

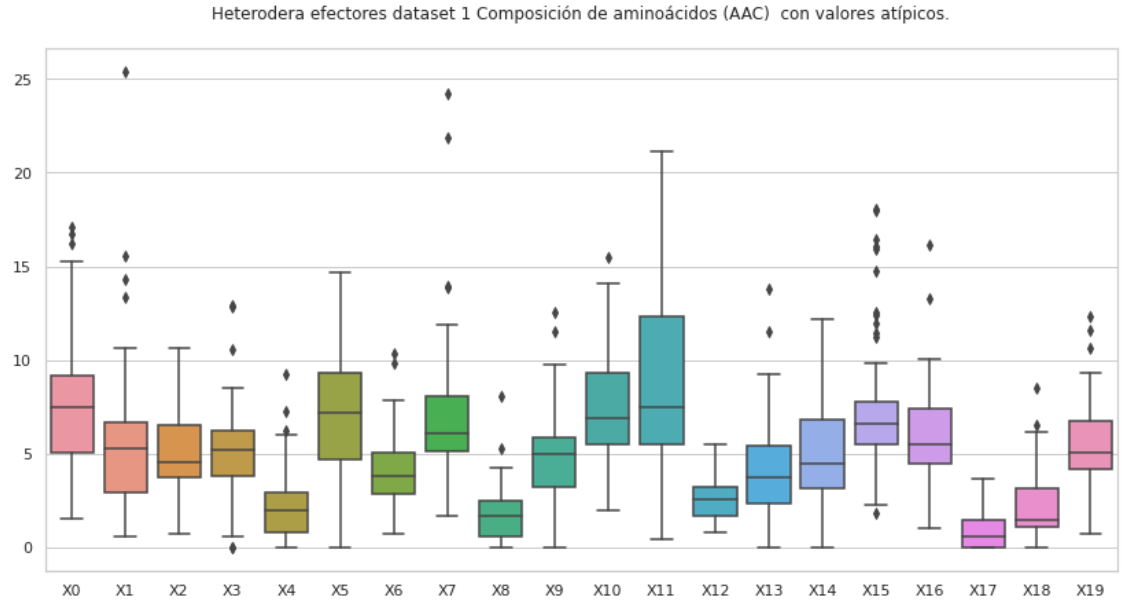
Composición de aminoácidos (AAC) no_efectores Heterodera dataset 1, con valores atípicos.
Estadísticas.

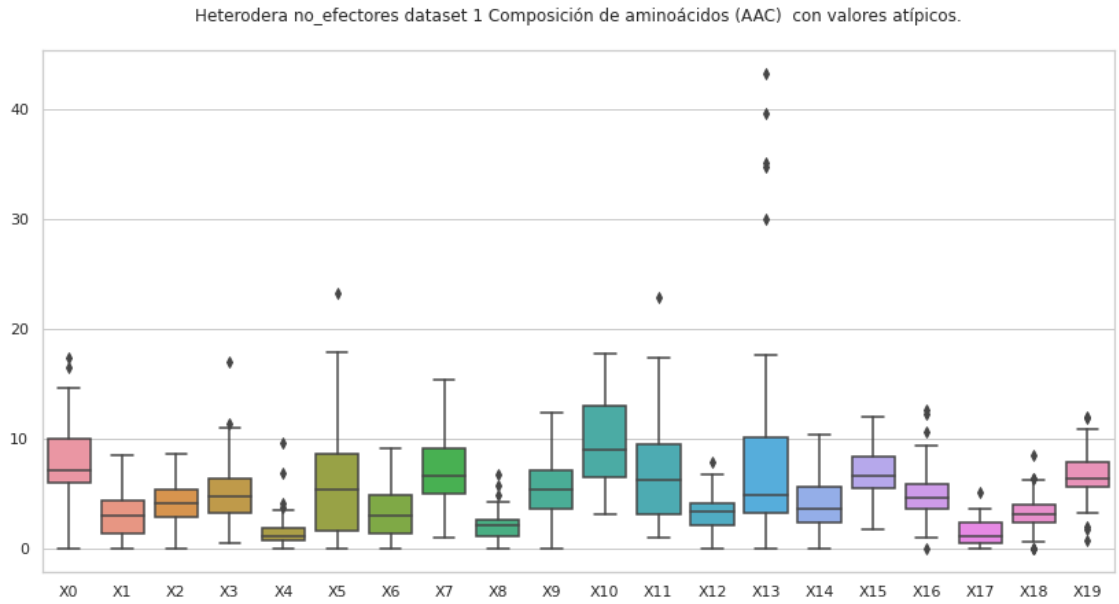
```
              X0         X1         X2         X3         X4         X5  \
count  81.000000  81.000000  81.000000  81.000000  81.000000  81.000000
mean    7.851198   3.126667   4.242605   5.045815   1.470346   5.859309
std     3.641319   2.034914   1.919191   2.651774   1.483617   4.730678
min     0.000000   0.000000   0.000000   0.529000   0.000000   0.000000
25%     5.952000   1.325000   2.899000   3.213000   0.690000   1.681000
50%     7.080000   2.959000   4.072000   4.762000   1.075000   5.430000
75%     9.963000   4.377000   5.422000   6.349000   1.905000   8.661000
max    17.436000   8.451000   8.571000  16.949000   9.677000  23.239000

              X6         X7         X8         X9        X10        X11  \
count  81.000000  81.000000  81.000000  81.000000  81.000000  81.000000
mean    3.233494   6.858938   2.100963   5.633086   9.642593   7.043531
std     2.088124   3.153170   1.214513   3.214618   3.763242   4.478361
min     0.000000   1.056000   0.000000   0.000000   3.150000   0.971000
25%     1.333000   5.000000   1.110000   3.676000   6.479000   3.109000
50%     2.959000   6.571000   2.147000   5.325000   9.048000   6.213000
```

75%	4.819000	9.155000	2.594000	7.143000	12.977000	9.524000
max	9.091000	15.323000	6.780000	12.319000	17.754000	22.835000

	X12	X13	X14	X15	X16	X17	\
count	81.000000	81.000000	81.000000	81.000000	81.000000	81.000000	
mean	3.203531	7.755815	4.105852	6.823617	4.780667	1.391889	
std	1.462540	8.424486	2.224698	2.082376	2.235368	1.071602	
min	0.000000	0.000000	0.000000	1.695000	0.000000	0.000000	
25%	2.083000	3.209000	2.410000	5.523000	3.571000	0.555000	
50%	3.321000	4.902000	3.628000	6.571000	4.603000	1.099000	
75%	4.184000	10.084000	5.622000	8.411000	5.903000	2.381000	
max	7.888000	43.243000	10.317000	11.966000	12.613000	5.128000	

	X18	X19
count	81.000000	81.000000
mean	3.122988	6.706963
std	1.557597	2.135855
min	0.000000	0.787000
25%	2.381000	5.575000
50%	3.170000	6.417000
75%	3.947000	7.850000
max	8.462000	11.972000





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 Heterodera dataset 1, sin valores
atípicos.
Valores del documento csv.

        X0      X1     X2     X3     X4      X5     X6      X7     X8     X9  \
0   10.662   2.941  6.250  4.412  2.941   6.985  5.882   5.147  0.735  4.044
1    8.571  14.286  1.714  0.571  4.571   3.429  2.286   5.143  1.714  4.571
2   12.069   6.034  4.310  5.172  0.000  14.655  7.759   1.724  0.862  2.586
3    3.030   2.273  9.091  6.061  3.788   3.788  4.545   3.788  2.273  4.545
5    4.972   5.525  4.420  6.077  2.210   9.392  2.762   5.525  0.552  5.525
..     …       …      …      …      …       …      …       …      …      …
76  12.230   5.755  5.036  3.597  0.000   3.597  2.158  10.072  2.878  5.036
77   5.056   5.618  3.933  6.180  2.247   9.551  2.809   5.056  0.562  5.618
78   6.531   4.082  4.082  3.673  2.449   2.449  2.857   6.939  0.408  8.980
79   9.172   5.325  4.438  3.846  2.959   7.396  4.142   6.509  1.775  5.325
80   7.660   4.255  7.234  3.830  1.702   4.681  2.979   7.234  1.702  5.106

    …     X11    X12    X13    X14     X15    X16    X17    X18    X19  \
0   …  10.662  3.676  5.515  2.206   5.515  5.515  0.735  1.103  6.618
1   …   6.857  4.000  4.000  2.286  12.571  6.286  2.286  0.571  6.286
2   …   6.897  3.448  4.310  0.862   9.483  4.310  0.000  1.724  6.897
3   …   8.333  2.273  4.545  5.303   5.303  8.333  0.758  4.545  8.333
5   …  13.260  1.657  2.210  9.945   6.630  8.287  0.000  1.105  4.420
..  …      …      …      …      …       …      …      …      …      …
76  …   7.194  2.878  2.878  7.914   8.633  6.475  0.000  0.000  4.317
```

```
77  …  12.921  1.685  2.247  9.551   7.303  8.989  0.000  1.124  3.933
78  …   6.122  2.449  5.306  4.082   4.898  6.531  2.857  6.531  6.939
79  …   7.988  3.254  4.734  6.509   5.325  4.734  0.592  1.479  4.734
80  …   3.404  2.979  5.532  5.106   7.660  5.532  0.426  1.277  8.511

          X20
0   efectores
1   efectores
2   efectores
3   efectores
5   efectores
..        …
76  efectores
77  efectores
78  efectores
79  efectores
80  efectores

[68 rows x 21 columns]
```

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

```
              X0         X1         X2         X3         X4         X5  \
count  68.000000  68.000000  68.000000  68.000000  68.000000  68.000000
mean    7.776132   4.988059   5.065412   5.227574   2.214603   7.062147
std     3.081262   2.452396   1.875103   1.639595   1.596462   2.915333
min     3.030000   0.599000   1.538000   0.571000   0.000000   0.000000
25%     5.149750   3.166000   3.844500   3.864750   0.867000   4.759500
50%     7.518000   5.249500   4.451000   5.375500   2.099500   7.269500
75%     9.314500   6.184250   6.064000   6.226000   2.959000   9.280750
max    16.766000  14.286000  10.185000   8.527000   6.250000  14.655000

              X6         X7         X8         X9        X10        X11  \
count  68.000000  68.000000  68.000000  68.000000  68.000000  68.000000
mean    4.024574   6.544824   1.670147   5.134338   7.762412   8.461015
std     1.599305   2.301885   1.221847   1.872575   2.760068   3.693802
min     0.769000   1.724000   0.000000   1.796000   2.679000   2.308000
25%     2.948500   5.146000   0.595000   3.756500   5.699750   5.714500
50%     3.928000   5.971500   1.624000   5.120500   7.559000   7.865000
75%     5.074250   7.701750   2.468500   6.018500   9.530500  11.645500
max     7.834000  13.843000   5.309000  11.538000  15.464000  17.365000

             X12        X13        X14        X15        X16        X17  \
count  68.000000  68.000000  68.000000  68.000000  68.000000  68.000000
mean    2.695162   4.215735   5.340441   7.211809   5.984456   0.883676
```

std	1.094315	2.084984	2.686762	2.674838	1.984465	0.984041
min	0.775000	0.599000	0.000000	1.829000	2.139000	0.000000
25%	1.760750	2.504250	3.224500	5.515000	4.734000	0.000000
50%	2.619500	3.990000	4.990000	6.688000	5.773000	0.592000
75%	3.434000	5.334250	6.884750	7.668000	7.536750	1.526000
max	5.488000	11.538000	12.150000	16.043000	10.063000	3.670000

	X18	X19
count	68.000000	68.000000
mean	1.971750	5.765735
std	1.460519	1.839689
min	0.000000	2.020000
25%	1.094750	4.381500
50%	1.375000	5.524500
75%	3.127000	6.868750
max	6.531000	11.613000

no_efectores

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

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9	\
0	6.452	4.301	3.943	9.319	1.075	12.545	2.509	5.018	1.792	5.376	
2	6.916	2.948	4.762	6.009	1.361	7.256	3.175	4.762	2.381	4.308	
3	12.048	3.614	4.819	3.614	1.205	1.205	4.819	6.024	2.410	0.000	
4	5.298	0.662	1.987	2.649	1.987	0.662	1.325	11.258	3.311	10.596	
5	7.002	5.819	5.621	4.832	1.775	5.819	3.945	7.002	2.170	5.325	
..	…	…	…	…	…	…	…	…	…	…	
75	5.952	5.952	8.333	7.143	1.190	8.333	2.381	1.190	3.571	4.762	
76	9.719	1.512	6.479	3.024	2.160	2.592	4.752	9.935	2.160	5.616	
78	6.723	0.840	1.681	3.361	0.840	0.000	0.840	10.084	2.521	11.765	
79	8.286	4.857	2.857	5.143	1.714	7.714	3.429	6.571	2.286	6.857	
80	9.562	4.781	4.382	7.968	0.797	10.757	5.976	5.179	0.797	3.586	

	…	X11	X12	X13	X14	X15	X16	X17	X18	X19	\
0	…	12.903	1.434	4.659	2.509	5.376	4.659	1.075	2.867	6.093	
2	…	8.163	2.721	5.896	3.628	7.370	5.556	1.814	3.288	6.349	
3	…	6.024	4.819	7.229	7.229	4.819	6.024	0.000	2.410	7.229	
4	…	2.649	3.311	11.258	3.311	5.298	3.974	2.649	5.298	9.934	
5	…	6.213	2.071	4.832	6.016	5.523	7.002	0.592	3.550	7.396	
..	…	…	…	…	…	…	…	…	…	…	
75	…	9.524	1.190	8.333	4.762	5.952	2.381	2.381	2.381	5.952	
76	…	4.968	2.160	2.160	7.343	10.367	4.752	2.808	4.104	6.911	
78	…	2.521	4.202	10.084	1.681	9.244	3.361	2.521	5.042	10.084	
79	…	5.429	4.857	3.714	5.143	6.571	5.714	1.143	4.000	6.286	

```
80  …   8.367  2.789   1.992  1.992   6.773  3.586  0.797  4.781   5.179

             X20
0   no_efectores
2   no_efectores
3   no_efectores
4   no_efectores
5   no_efectores
..           …
75  no_efectores
76  no_efectores
78  no_efectores
79  no_efectores
80  no_efectores

[68 rows x 21 columns]
```

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

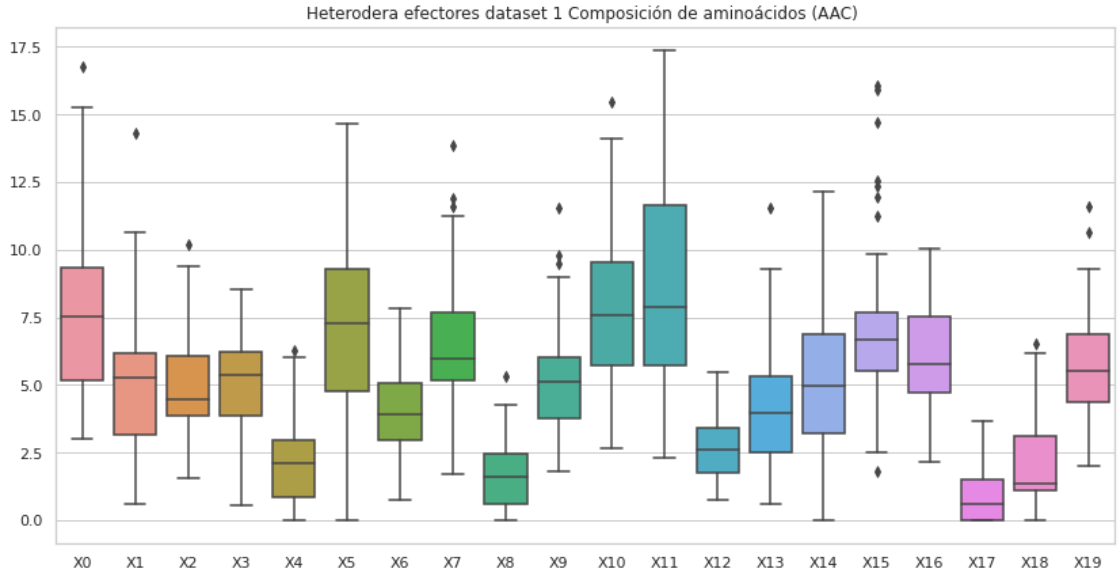
	X0	X1	X2	X3	X4	X5	\
count	68.000000	68.000000	68.000000	68.000000	68.000000	68.000000	
mean	8.098779	3.159353	4.374250	4.960309	1.419294	5.502765	
std	3.188064	1.954242	1.902396	2.194776	1.003003	4.367380	
min	0.000000	0.513000	0.820000	0.840000	0.000000	0.000000	
25%	6.024000	1.313000	3.052500	3.354000	0.725000	1.205000	
50%	7.163000	3.015500	4.273000	4.638000	1.190000	5.426000	
75%	9.916500	4.410500	5.686250	6.389750	1.981750	8.358250	
max	17.436000	8.407000	8.571000	11.409000	4.167000	17.874000	

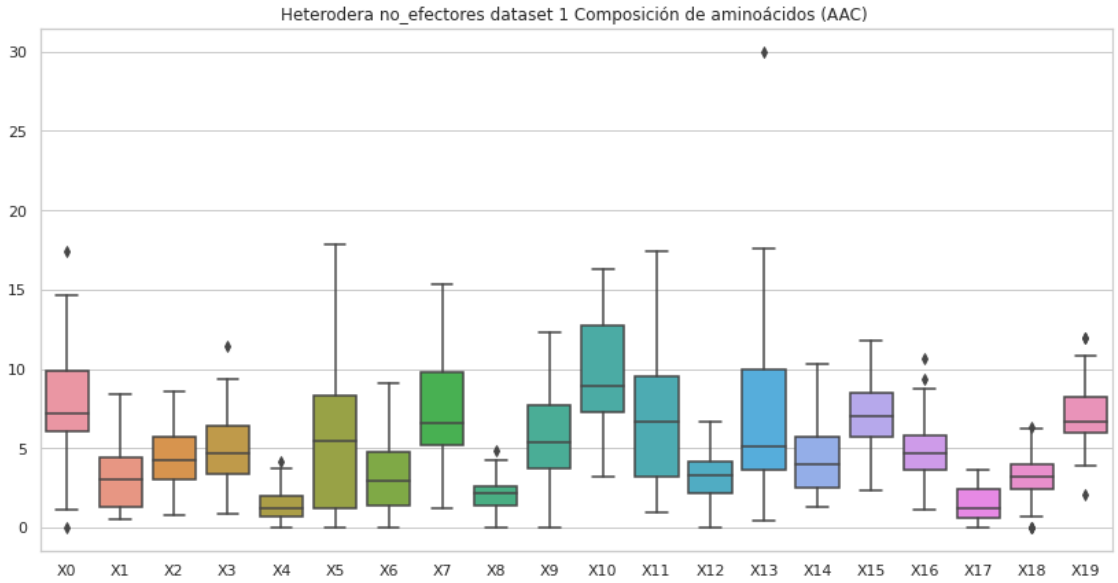
	X6	X7	X8	X9	X10	X11	\
count	68.000000	68.000000	68.000000	68.000000	68.000000	68.000000	
mean	3.196235	7.263029	2.126397	5.767176	9.527324	7.092397	
std	2.057563	3.025008	0.966329	3.269858	3.363181	4.262607	
min	0.000000	1.190000	0.000000	0.000000	3.175000	0.971000	
25%	1.351000	5.171000	1.423750	3.686500	7.271250	3.183000	
50%	2.908000	6.571000	2.194000	5.347000	8.901000	6.642500	
75%	4.768750	9.798000	2.603500	7.692000	12.742500	9.541500	
max	9.091000	15.323000	4.854000	12.319000	16.246000	17.436000	

	X12	X13	X14	X15	X16	X17	\
count	68.000000	68.000000	68.000000	68.000000	68.000000	68.000000	
mean	3.174206	6.630853	4.371882	6.989250	4.792382	1.419824	
std	1.367815	4.853545	2.149902	1.902721	1.702498	1.005138	
min	0.000000	0.426000	1.342000	2.381000	1.120000	0.000000	
25%	2.194500	3.623500	2.536750	5.726000	3.653500	0.582750	

50%	3.316000	5.070000	3.963500	7.020000	4.651000	1.245000
75%	4.123250	9.964000	5.734750	8.492250	5.798750	2.412250
max	6.711000	29.972000	10.317000	11.806000	10.628000	3.676000

	X18	X19
count	68.000000	68.000000
mean	3.106941	7.027088
std	1.446172	1.930361
min	0.000000	2.045000
25%	2.410000	5.952000
50%	3.192000	6.638000
75%	3.960250	8.268500
max	6.338000	11.972000





3 Composición de pseudo aminoácidos (PseAAC) hidro_mass

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

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

    if etiq == "efectores":
        df=PseAAC_hidro_mass_efec

    if etiq == "no_efectores":
        df=PseAAC_hidro_mass_no_efec

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

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

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

          X0        X1        X2        X3        X4        X5        X6  \
0   0.058009  0.016002  0.024004  0.038006  0.030005  0.028004  0.004001
1   0.038327  0.020441  0.002555  0.015331  0.017886  0.022996  0.007665
2   0.040111  0.000000  0.017190  0.048706  0.014325  0.005730  0.002865
3   0.014892  0.018614  0.029783  0.018614  0.022337  0.018614  0.011169
4   0.013831  0.011855  0.018770  0.017782  0.005927  0.039516  0.007903
..       …         …         …         …         …         …         …
76  0.023367  0.000000  0.006873  0.006873  0.005498  0.019244  0.005498
77  0.007189  0.003195  0.008787  0.013580  0.003195  0.007189  0.000799
78  0.024521  0.009195  0.013793  0.009195  0.019924  0.026054  0.001533
79  0.039337  0.012689  0.016496  0.031723  0.020303  0.027916  0.007614
80  0.029235  0.006497  0.014617  0.017866  0.021114  0.027611  0.006497

          X7        X8        X9  …       X74       X75       X76       X77  \
0   0.022003  0.058009  0.046007  …  0.000902  0.014766  0.032452 -0.025553
1   0.020441  0.030661  0.035772  …  0.011216  0.006042  0.003969 -0.001780
2   0.008595  0.022920  0.022920  …  0.024963  0.057482  0.000219  0.001392
3   0.022337  0.040952  0.044675  …  0.009159  0.001242  0.028082 -0.024448
4   0.045444  0.016795  0.023710  … -0.020561 -0.009907  0.042428 -0.012528
..       …         …         …  …         …         …         …         …
76  0.009622  0.013745  0.017869  …  0.000288  0.009898  0.038290 -0.010012
77  0.007988  0.018373  0.007988  …  0.015498  0.023211 -0.000205  0.009766
78  0.033717  0.022989  0.044445  …  0.005821  0.005943  0.032243 -0.004808
79  0.022841  0.034261  0.041874  … -0.007310  0.006377  0.027387 -0.000846
80  0.019490  0.012993  0.050349  …  0.007172  0.000685  0.015535  0.014967

         X78       X79       X80       X81       X82        X83
0  -0.023318  0.011503  0.002070 -0.015834  0.015188  efectores
1   0.014038 -0.023848  0.027777  0.024300 -0.000282  efectores
2   0.036096 -0.010760 -0.004943  0.005944  0.014890  efectores
3  -0.006427 -0.003922  0.026091  0.041316 -0.004104  efectores
4   0.001211  0.012631 -0.006430 -0.004450  0.006922  efectores
..       …         …         …         …         …          …
76  0.003715  0.024033 -0.002319 -0.001532  0.035198  efectores
77  0.020071  0.002793  0.006442  0.015011  0.004970  efectores
```

```
78 -0.004617  0.024709  0.004247 -0.001978  0.014143  efectores
79  0.014595  0.022144 -0.004928  0.004507  0.001693  efectores
80  0.011526  0.020328  0.016956 -0.002194  0.031754  efectores

[81 rows x 84 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	81.000000	81.000000	81.000000	81.000000	81.000000	81.000000	
mean	0.025795	0.007916	0.017285	0.022261	0.015774	0.022634	
std	0.016220	0.007955	0.011971	0.013569	0.013566	0.014074	
min	0.002680	0.000000	0.000000	0.000000	0.000000	0.005730	
25%	0.013089	0.002138	0.008773	0.013244	0.004848	0.010074	
50%	0.023367	0.004185	0.016496	0.018641	0.013513	0.019454	
75%	0.035570	0.012848	0.023284	0.031948	0.022337	0.028114	
max	0.075554	0.039983	0.056035	0.060578	0.063226	0.069665	

	X6	X7	X8	X9	…	X73	X74	\
count	81.000000	81.000000	81.000000	81.000000	…	81.000000	81.000000	
mean	0.006141	0.016649	0.024916	0.027269	…	0.012599	0.005310	
std	0.005543	0.011669	0.013145	0.020870	…	0.013428	0.015653	
min	0.000000	0.000000	0.000802	0.002178	…	-0.009016	-0.027552	
25%	0.001621	0.007988	0.017029	0.009285	…	0.002363	-0.006769	
50%	0.005172	0.013899	0.020236	0.020956	…	0.008636	0.005888	
75%	0.009763	0.022337	0.032667	0.038962	…	0.020896	0.014303	
max	0.023039	0.048142	0.060578	0.091246	…	0.056454	0.053632	

	X75	X76	X77	X78	X79	X80	\
count	81.000000	81.000000	81.000000	81.000000	81.000000	81.000000	
mean	0.012757	0.012754	0.001109	0.008985	0.011293	0.003605	
std	0.013165	0.015716	0.017256	0.015563	0.017031	0.014623	
min	-0.018557	-0.027522	-0.093528	-0.044489	-0.023848	-0.049908	
25%	0.005424	-0.000205	-0.004808	-0.001456	0.001274	-0.004928	
50%	0.011652	0.009905	0.004485	0.011705	0.008000	0.005932	
75%	0.022896	0.026845	0.010209	0.019962	0.022058	0.010400	
max	0.057482	0.047977	0.034882	0.050229	0.059342	0.046868	

	X81	X82
count	81.000000	81.000000
mean	0.009759	0.010388
std	0.014702	0.014596
min	-0.036146	-0.041397
25%	0.000702	0.001443
50%	0.011782	0.007822

```
75%     0.018737   0.018553
max     0.046311   0.044552

[8 rows x 83 columns]



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

          X0        X1        X2        X3        X4        X5        X6  \
0   0.010374  0.001729  0.014984  0.020171  0.007492  0.008068  0.002882
1   0.000726  0.000000  0.000726  0.002904  0.034850  0.001452  0.000726
2   0.058305  0.011470  0.050658  0.061172  0.049702  0.040144  0.020072
3   0.057638  0.005764  0.017291  0.005764  0.034583  0.028819  0.011528
4   0.008522  0.003196  0.004261  0.001065  0.018110  0.018110  0.005326
..       …         …         …         …         …         …         …
76  0.028750  0.006389  0.008945  0.007667  0.006389  0.029389  0.006389
77  0.002605  0.000868  0.002605  0.003908  0.042115  0.002171  0.000868
78  0.009510  0.001189  0.004755  0.000000  0.014265  0.014265  0.003566
79  0.055835  0.011552  0.034656  0.051984  0.025030  0.044283  0.015403
80  0.039592  0.003299  0.032993  0.044541  0.008248  0.021446  0.003299

          X7        X8        X9  …       X74       X75       X76       X77  \
0   0.008645  0.020747  0.009797  …  0.005841  0.032475 -0.000798  0.015290
1   0.003630  0.002178  0.013795  …  0.017737  0.013717  0.002840  0.017866
2   0.036321  0.068819  0.095581  … -0.006697  0.011839  0.006543 -0.002585
3   0.000000  0.028819  0.069166  …  0.010370 -0.010239  0.038574 -0.005314
4   0.017044  0.004261  0.020240  …  0.023730  0.013605 -0.003006  0.026576
..       …         …         …  …         …         …         …         …
76  0.016611  0.014695  0.019167  …  0.009027  0.003026  0.027732 -0.002514
77  0.007815  0.002605  0.021274  …  0.018493  0.015463  0.001514  0.019453
78  0.016643  0.003566  0.017832  …  0.025718  0.015980 -0.000378  0.028227
79  0.046208  0.036582  0.050059  …  0.040920  0.034248  0.023612  0.012300
80  0.014847  0.034643  0.041241  …  0.022571  0.021508 -0.004184  0.006463

         X78       X79       X80       X81       X82           X83
0   0.035387  0.009171  0.005185  0.024512 -0.002140  no_efectores
1   0.019567  0.001987  0.018406  0.015473  0.002266  no_efectores
2  -0.003377  0.015921 -0.003715 -0.019895  0.007387  no_efectores
3   0.001297  0.001729  0.051285  0.017272  0.023879  no_efectores
4   0.009444 -0.008516  0.026204  0.011481 -0.002107  no_efectores
..       …         …         …         …         …           …
76  0.000945  0.026303 -0.002695 -0.006661  0.030137  no_efectores
77  0.014430 -0.001668  0.020182  0.015340  0.006048  no_efectores
78  0.006754 -0.004904  0.023007  0.008573  0.002599  no_efectores
```

```
79  0.007343  0.026013  0.013127 -0.010185  0.010573  no_efectores
80 -0.010494  0.004765 -0.011080  0.028235  0.010211  no_efectores

[81 rows x 84 columns]
```

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

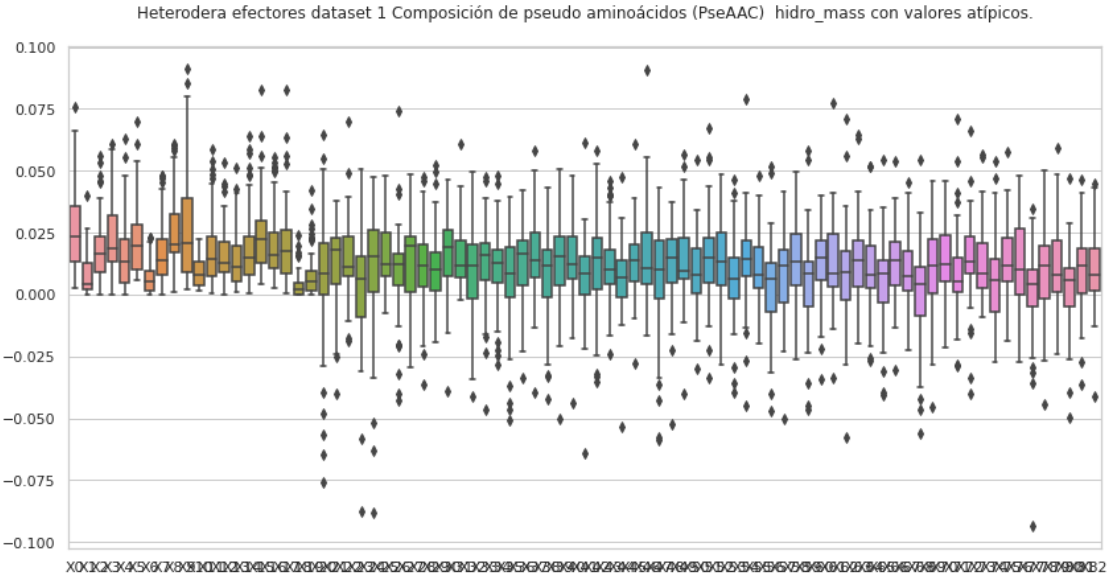
	X0	X1	X2	X3	X4	X5	\
count	81.000000	81.000000	81.000000	81.000000	81.000000	81.000000	
mean	0.028795	0.006222	0.019382	0.022986	0.022719	0.021960	
std	0.020412	0.008210	0.016528	0.022623	0.019321	0.012838	
min	0.000000	0.000000	0.000428	0.000000	0.000000	0.000856	
25%	0.010597	0.000996	0.006549	0.003908	0.008458	0.013546	
50%	0.025327	0.003196	0.015079	0.015457	0.017467	0.017741	
75%	0.042231	0.009896	0.025972	0.035544	0.031304	0.030070	
max	0.077283	0.048469	0.078576	0.091672	0.091672	0.054375	

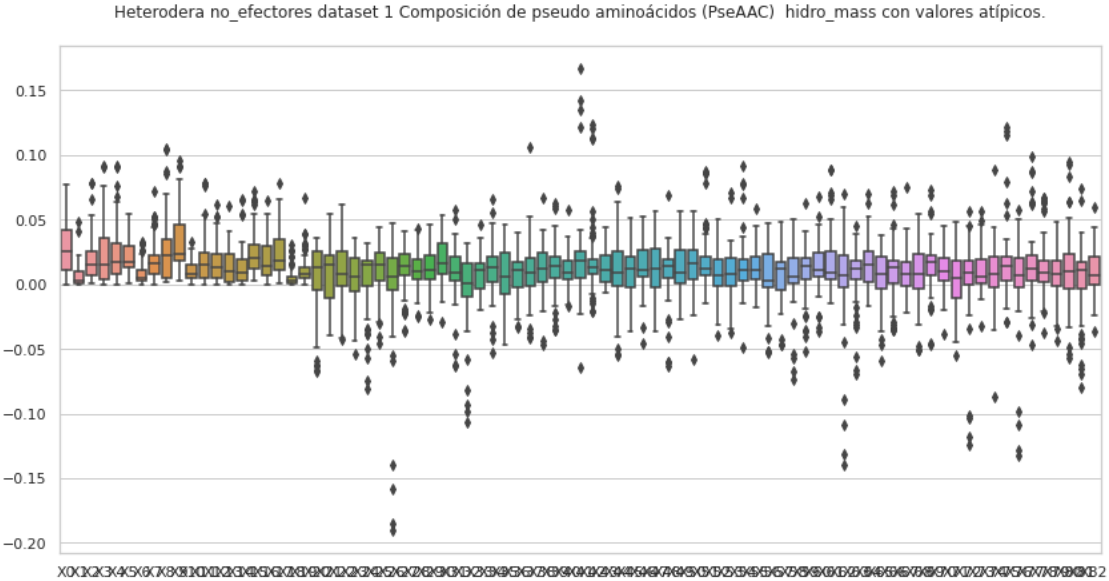
	X6	X7	X8	X9	…	X73	X74	\
count	81.000000	81.000000	81.000000	81.000000	…	81.000000	81.000000	
mean	0.007950	0.018614	0.027052	0.033384	…	0.009821	0.008224	
std	0.007689	0.014436	0.023776	0.023030	…	0.014989	0.020876	
min	0.000000	0.000000	0.001637	0.002676	…	-0.032028	-0.087075	
25%	0.003016	0.008380	0.004847	0.017832	…	0.000298	-0.001884	
50%	0.004852	0.016643	0.021994	0.022988	…	0.005744	0.008445	
75%	0.011528	0.022702	0.034619	0.046157	…	0.019530	0.021177	
max	0.032736	0.072500	0.104768	0.095581	…	0.056149	0.088525	

	X75	X76	X77	X78	X79	X80	\
count	81.000000	81.000000	81.000000	81.000000	81.000000	81.000000	
mean	0.016209	0.004554	0.014622	0.013083	0.007987	0.012408	
std	0.026763	0.032595	0.024418	0.020427	0.016076	0.027374	
min	-0.029655	-0.132490	-0.046807	-0.036959	-0.044076	-0.057267	
25%	0.002645	-0.002024	0.002361	0.002158	-0.001763	-0.003834	
50%	0.013937	0.006543	0.012331	0.009307	0.008351	0.009659	
75%	0.021176	0.020059	0.022474	0.017839	0.018676	0.023113	
max	0.121383	0.057922	0.098826	0.067114	0.047758	0.094905	

	X81	X82
count	81.000000	81.000000
mean	0.004547	0.010891
std	0.025993	0.014957
min	-0.080155	-0.036534
25%	-0.003752	0.000142
50%	0.010669	0.007387
75%	0.017028	0.021772

```
max      0.073946   0.059266

[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 Heterodera
dataset 1, sin valores atípicos.
Valores del documento csv.
```

	X0	X1	X2	X3	X4	X5	X6	\
0	0.058009	0.016002	0.024004	0.038006	0.030005	0.028004	0.004001	
5	0.007185	0.003193	0.008782	0.013572	0.003193	0.007983	0.000798	
7	0.012829	0.000000	0.024055	0.015235	0.004811	0.020046	0.006415	
8	0.014355	0.004101	0.020507	0.022557	0.017431	0.018456	0.008203	
10	0.039830	0.012848	0.016703	0.032121	0.020557	0.028266	0.007709	
..	…	…	…	…	…	…	…	
76	0.023367	0.000000	0.006873	0.006873	0.005498	0.019244	0.005498	
77	0.007189	0.003195	0.008787	0.013580	0.003195	0.007189	0.000799	
78	0.024521	0.009195	0.013793	0.009195	0.019924	0.026054	0.001533	
79	0.039337	0.012689	0.016496	0.031723	0.020303	0.027916	0.007614	
80	0.029235	0.006497	0.014617	0.017866	0.021114	0.027611	0.006497	

	X7	X8	X9	…	X74	X75	X76	X77	\
0	0.022003	0.058009	0.046007	…	0.000902	0.014766	0.032452	-0.025553	
5	0.007983	0.019160	0.007983	…	0.014303	0.022593	-0.001398	0.008791	
7	0.006415	0.000802	0.007217	…	0.001181	0.015622	0.016373	0.003906	
8	0.012304	0.021532	0.038962	…	-0.008065	0.007515	-0.009622	0.000128	
10	0.023127	0.034691	0.042400	…	-0.008279	0.005493	0.027654	-0.001591	
..	…	…	…	…	…	…	…	…	
76	0.009622	0.013745	0.017869	…	0.000288	0.009898	0.038290	-0.010012	
77	0.007988	0.018373	0.007988	…	0.015498	0.023211	-0.000205	0.009766	
78	0.033717	0.022989	0.044445	…	0.005821	0.005943	0.032243	-0.004808	
79	0.022841	0.034261	0.041874	…	-0.007310	0.006377	0.027387	-0.000846	
80	0.019490	0.012993	0.050349	…	0.007172	0.000685	0.015535	0.014967	

	X78	X79	X80	X81	X82	X83
0	-0.023318	0.011503	0.002070	-0.015834	0.015188	efectores
5	0.021689	0.000906	0.008579	0.016866	0.002050	efectores
7	0.017139	0.015391	0.001113	0.023935	0.019692	efectores
8	0.010950	0.004020	0.007813	0.010486	0.011634	efectores
10	0.013496	0.022345	-0.006743	0.003278	0.001262	efectores
..	…	…	…	…	…	…
76	0.003715	0.024033	-0.002319	-0.001532	0.035198	efectores
77	0.020071	0.002793	0.006442	0.015011	0.004970	efectores
78	-0.004617	0.024709	0.004247	-0.001978	0.014143	efectores
79	0.014595	0.022144	-0.004928	0.004507	0.001693	efectores
80	0.011526	0.020328	0.016956	-0.002194	0.031754	efectores

[62 rows x 84 columns]

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

	X0	X1	X2	X3	X4	X5	\
count	62.000000	62.000000	62.000000	62.000000	62.000000	62.000000	
mean	0.021806	0.006835	0.014381	0.019006	0.012261	0.019397	
std	0.013881	0.006686	0.008141	0.010326	0.009657	0.010876	
min	0.002680	0.000000	0.000000	0.000000	0.000000	0.006293	
25%	0.008370	0.002151	0.008249	0.012618	0.003451	0.009494	
50%	0.021602	0.003366	0.013431	0.017293	0.009226	0.016606	
75%	0.030100	0.012502	0.020439	0.025658	0.019389	0.027259	
max	0.062856	0.026487	0.037305	0.039755	0.038120	0.049239	

	X6	X7	X8	X9	…	X73	X74	\
count	62.000000	62.000000	62.000000	62.000000	…	62.000000	62.000000	
mean	0.004994	0.013896	0.022919	0.021097	…	0.011145	0.003447	
std	0.004293	0.009102	0.011607	0.014197	…	0.011162	0.012362	
min	0.000000	0.000000	0.000802	0.002178	…	-0.007199	-0.027552	
25%	0.000851	0.006999	0.017216	0.008206	…	0.002296	-0.006623	
50%	0.003395	0.011323	0.019675	0.017799	…	0.008045	0.005241	
75%	0.008175	0.019644	0.029410	0.031354	…	0.019251	0.011961	
max	0.017288	0.042885	0.060149	0.057593	…	0.036798	0.031054	

	X75	X76	X77	X78	X79	X80	\
count	62.000000	62.000000	62.000000	62.000000	62.000000	62.000000	
mean	0.012711	0.011698	0.004435	0.011019	0.012672	0.004083	
std	0.011107	0.015003	0.009855	0.011757	0.014165	0.010777	
min	-0.018557	-0.027522	-0.025553	-0.023318	-0.021028	-0.039844	
25%	0.005804	-0.000515	-0.001246	0.000555	0.002168	-0.002222	
50%	0.012514	0.009289	0.006253	0.014247	0.008758	0.006646	
75%	0.022839	0.025135	0.010098	0.020044	0.022122	0.009946	
max	0.030674	0.045287	0.030979	0.030419	0.044824	0.028609	

	X81	X82
count	62.000000	62.000000
mean	0.010241	0.012350
std	0.012233	0.013666
min	-0.027444	-0.011241
25%	0.002435	0.002526
50%	0.013042	0.009118
75%	0.018074	0.021935
max	0.031270	0.044552

[8 rows x 83 columns]

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

          X0        X1        X2        X3        X4        X5        X6  \
0   0.010374  0.001729  0.014984  0.020171  0.007492  0.008068  0.002882
1   0.000726  0.000000  0.000726  0.002904  0.034850  0.001452  0.000726
2   0.058305  0.011470  0.050658  0.061172  0.049702  0.040144  0.020072
3   0.057638  0.005764  0.017291  0.005764  0.034583  0.028819  0.011528
4   0.008522  0.003196  0.004261  0.001065  0.018110  0.018110  0.005326
..       …         …         …         …         …         …
76  0.028750  0.006389  0.008945  0.007667  0.006389  0.029389  0.006389
77  0.002605  0.000868  0.002605  0.003908  0.042115  0.002171  0.000868
78  0.009510  0.001189  0.004755  0.000000  0.014265  0.014265  0.003566
79  0.055835  0.011552  0.034656  0.051984  0.025030  0.044283  0.015403
80  0.039592  0.003299  0.032993  0.044541  0.008248  0.021446  0.003299

          X7        X8        X9  …       X74       X75       X76       X77  \
0   0.008645  0.020747  0.009797  …  0.005841  0.032475 -0.000798  0.015290
1   0.003630  0.002178  0.013795  …  0.017737  0.013717  0.002840  0.017866
2   0.036321  0.068819  0.095581  … -0.006697  0.011839  0.006543 -0.002585
3   0.000000  0.028819  0.069166  …  0.010370 -0.010239  0.038574 -0.005314
4   0.017044  0.004261  0.020240  …  0.023730  0.013605 -0.003006  0.026576
..       …         …         …  …         …         …         …         …
76  0.016611  0.014695  0.019167  …  0.009027  0.003026  0.027732 -0.002514
77  0.007815  0.002605  0.021274  …  0.018493  0.015463  0.001514  0.019453
78  0.016643  0.003566  0.017832  …  0.025718  0.015980 -0.000378  0.028227
79  0.046208  0.036582  0.050059  …  0.040920  0.034248  0.023612  0.012300
80  0.014847  0.034643  0.041241  …  0.022571  0.021508 -0.004184  0.006463

         X78       X79       X80       X81       X82           X83
0   0.035387  0.009171  0.005185  0.024512 -0.002140  no_efectores
1   0.019567  0.001987  0.018406  0.015473  0.002266  no_efectores
2  -0.003377  0.015921 -0.003715 -0.019895  0.007387  no_efectores
3   0.001297  0.001729  0.051285  0.017272  0.023879  no_efectores
4   0.009444 -0.008516  0.026204  0.011481 -0.002107  no_efectores
..       …         …         …         …         …           …
76  0.000945  0.026303 -0.002695 -0.006661  0.030137  no_efectores
77  0.014430 -0.001668  0.020182  0.015340  0.006048  no_efectores
78  0.006754 -0.004904  0.023007  0.008573  0.002599  no_efectores
79  0.007343  0.026013  0.013127 -0.010185  0.010573  no_efectores
80 -0.010494  0.004765 -0.011080  0.028235  0.010211  no_efectores

[65 rows x 84 columns]
```

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

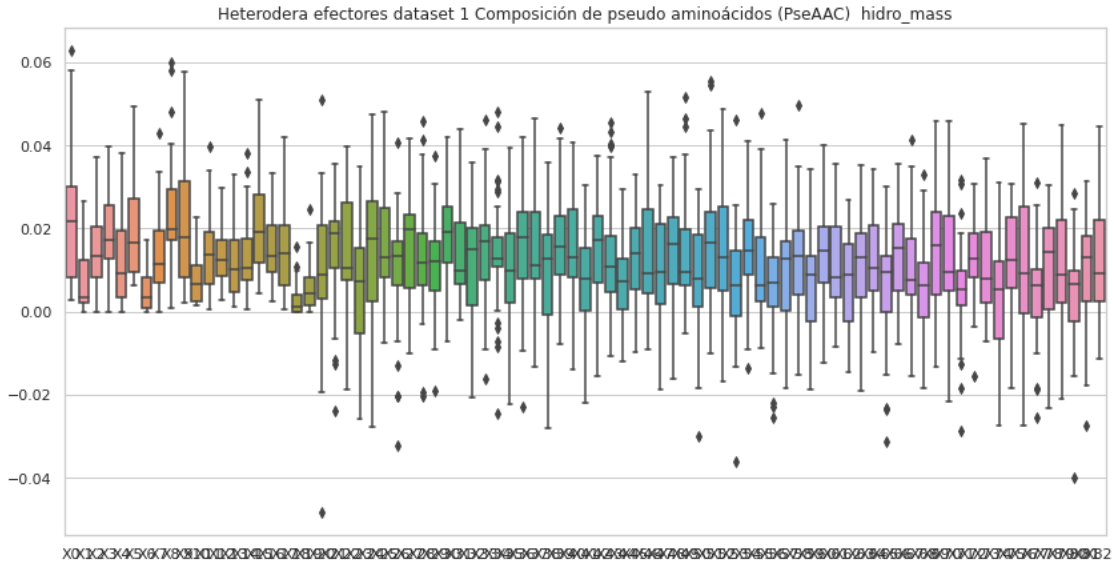
	X0	X1	X2	X3	X4	X5	\
count	65.000000	65.000000	65.000000	65.000000	65.000000	65.000000	
mean	0.024259	0.004852	0.014717	0.016621	0.017744	0.020230	
std	0.017276	0.005404	0.011376	0.016298	0.012133	0.011347	
min	0.000000	0.000000	0.000428	0.000000	0.000000	0.000856	
25%	0.010358	0.000903	0.004847	0.002141	0.008243	0.013645	
50%	0.021994	0.002330	0.011644	0.011063	0.016503	0.016926	
75%	0.035065	0.007497	0.020496	0.027442	0.025030	0.026087	
max	0.061296	0.022227	0.050658	0.061172	0.049702	0.053344	

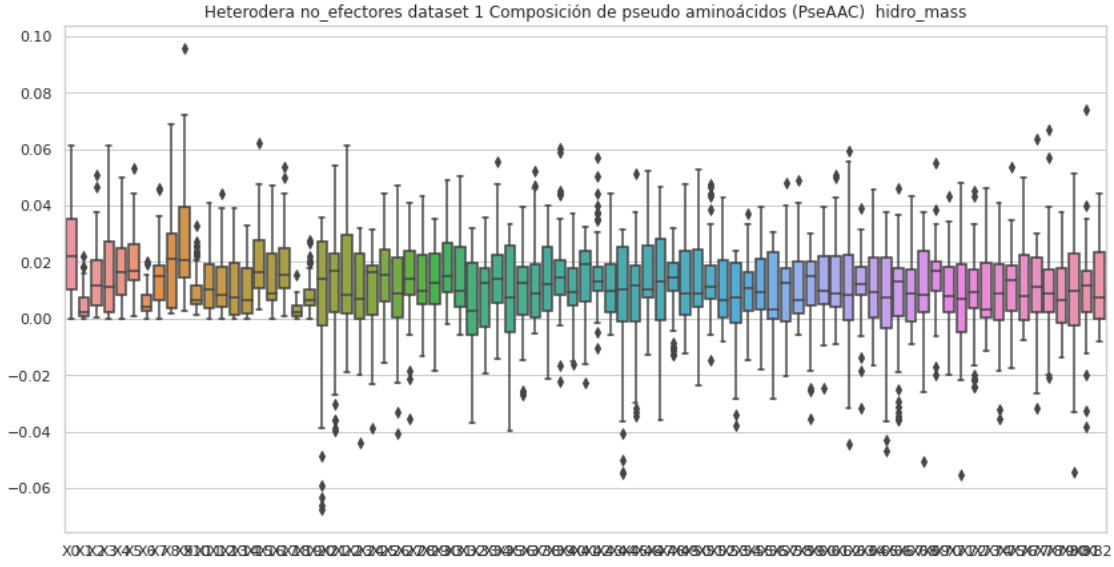
	X6	X7	X8	X9	…	X73	X74	\
count	65.000000	65.000000	65.000000	65.000000	…	65.000000	65.000000	
mean	0.006044	0.014763	0.020791	0.027880	…	0.008861	0.008575	
std	0.004706	0.010429	0.016532	0.019468	…	0.013055	0.014210	
min	0.000000	0.000000	0.001637	0.002676	…	-0.011462	-0.035680	
25%	0.002882	0.006652	0.003884	0.014735	…	0.000217	-0.000758	
50%	0.004128	0.014935	0.021110	0.020780	…	0.003057	0.009027	
75%	0.008315	0.018938	0.030254	0.039419	…	0.019492	0.019195	
max	0.020072	0.046208	0.068819	0.095581	…	0.046179	0.040920	

	X75	X76	X77	X78	X79	X80	\
count	65.000000	65.000000	65.000000	65.000000	65.000000	65.000000	
mean	0.012036	0.011719	0.010414	0.010894	0.008612	0.010750	
std	0.013548	0.014827	0.015656	0.015445	0.012392	0.018874	
min	-0.017444	-0.007769	-0.031445	-0.021007	-0.013912	-0.054241	
25%	0.002645	-0.000502	0.002361	0.002158	-0.001289	-0.002576	
50%	0.013717	0.007944	0.011239	0.009005	0.006395	0.009659	
75%	0.018954	0.022739	0.021455	0.017304	0.017822	0.023007	
max	0.053490	0.050070	0.063654	0.066759	0.037670	0.051285	

	X81	X82
count	65.000000	65.000000
mean	0.010433	0.011367
std	0.016538	0.012704
min	-0.038245	-0.007928
25%	0.002504	0.000142
50%	0.011517	0.007249
75%	0.017118	0.023345
max	0.073946	0.044253

[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))
```

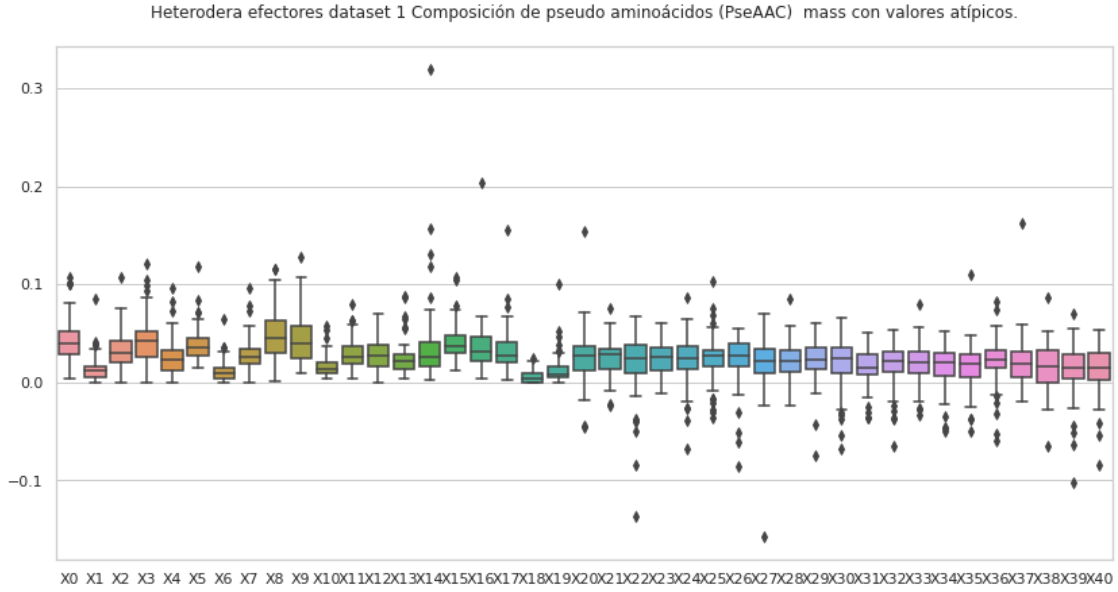
```
efectores
Composición de pseudo aminoácidos (PseAAC)  mass efectores Heterodera dataset 1,
con valores atípicos.
Valores del documento csv.

          X0        X1        X2        X3        X4        X5        X6  \
0   0.060335  0.016644  0.024966  0.039530  0.031208  0.029127  0.004161
1   0.071047  0.037892  0.004736  0.028419  0.033155  0.042628  0.014209
2   0.099546  0.000000  0.042663  0.120877  0.035552  0.014221  0.007110
3   0.019767  0.024708  0.039533  0.024708  0.029650  0.024708  0.014825
4   0.014423  0.012363  0.019574  0.018544  0.006181  0.041209  0.008242
..       …         …         …         …         …         …         …
76  0.029099  0.000000  0.008559  0.008559  0.006847  0.023964  0.006847
77  0.025844  0.011486  0.031587  0.048817  0.011486  0.025844  0.002872
78  0.038419  0.014407  0.021611  0.014407  0.031215  0.040820  0.002401
79  0.051086  0.016479  0.021423  0.041198  0.026367  0.036255  0.009888
80  0.033304  0.007401  0.016652  0.020353  0.024053  0.031454  0.007401

          X7        X8        X9  …       X32       X33       X34       X35  \
```

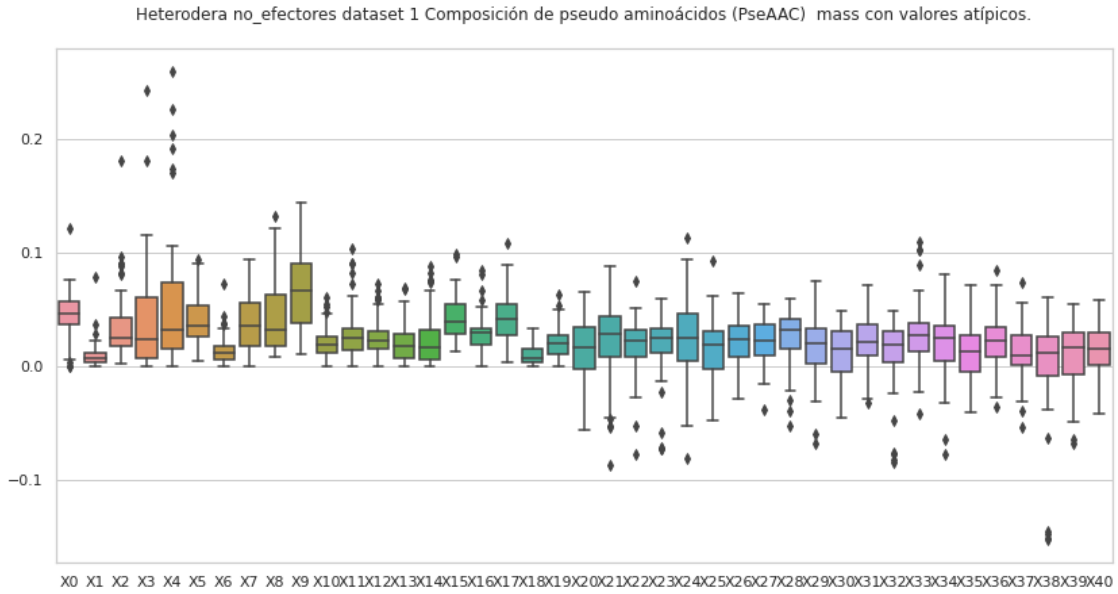
```
0   0.022886  0.060335  0.047852  …  0.029866  0.025941  0.012656  0.001198
1   0.037892  0.056837  0.066310  …  0.001540 -0.026201 -0.047765  0.019060
2   0.021331  0.056883  0.056883  … -0.065569  0.020140  0.049278 -0.009868
3   0.029650  0.054358  0.059300  …  0.040292  0.026592  0.021186 -0.002321
4   0.047390  0.017514  0.024725  …  0.011847  0.032178  0.029237  0.027870
..       …         …        …  …        …         …         …
76  0.011982  0.017117  0.022253  …  0.036068  0.032763  0.036653  0.025407
77  0.028716  0.066046  0.028716  …  0.020057  0.017693  0.022440  0.015977
78  0.052826  0.036018  0.069634  …  0.028346  0.038424  0.028308 -0.019825
79  0.029663  0.044494  0.054382  …  0.024933  0.009746  0.021250  0.017982
80  0.022203  0.014802  0.057357  …  0.032463  0.037197  0.021045  0.039357

         X36       X37       X38       X39       X40        X41
0   0.017419  0.022492  0.033754  0.011964  0.015797  efectores
1   0.056832  0.037431  0.007357 -0.044208 -0.000523  efectores
2  -0.003927  0.026060  0.000543 -0.026703  0.036954  efectores
3   0.008925  0.018859  0.037276 -0.005206 -0.005448  efectores
4   0.037862  0.058871  0.044245  0.013172  0.007218  efectores
..       …         …         …         …         …         …
76  0.037558  0.036819  0.047683  0.029929  0.043832  efectores
77  0.020431  0.005524 -0.000739  0.010041  0.017865  efectores
78  0.007948  0.028842  0.050517  0.038713  0.022158  efectores
79  0.032635  0.003297  0.035567  0.028758  0.002198  efectores
80  0.019453  0.030680  0.017698  0.023157  0.036175  efectores

[81 rows x 42 columns]


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

              X0         X1         X2         X3         X4         X5  \
count  81.000000  81.000000  81.000000  81.000000  81.000000  81.000000
mean    0.043926   0.012921   0.030976   0.041519   0.025073   0.037981
std     0.021357   0.012601   0.017724   0.024977   0.018009   0.016416
min     0.003824   0.000000   0.000000   0.000000   0.000000   0.014221
25%     0.028544   0.004990   0.019574   0.024978   0.011355   0.027451
50%     0.039998   0.011229   0.029281   0.041542   0.022254   0.035967
75%     0.052226   0.016644   0.042522   0.051403   0.031966   0.044432
max     0.106388   0.085194   0.106520   0.120877   0.096400   0.117424

              X6         X7         X8         X9  …        X31        X32  \
count  81.000000  81.000000  81.000000  81.000000  …  81.000000  81.000000
mean    0.010700   0.028117   0.047892   0.044386  …   0.016203   0.017574
std     0.010319   0.016452   0.025034   0.024768  …   0.018341   0.021431
min     0.000000   0.000000   0.001508   0.009441  …  -0.036475  -0.065569
25%     0.003217   0.019174   0.029519   0.024725  …   0.007296   0.010445
```

```
50%     0.009587   0.025566   0.045086   0.039922  …   0.014675   0.021476
75%     0.014825   0.034521   0.063307   0.058030  …   0.028356   0.030809
max     0.064828   0.096400   0.115262   0.126960  …   0.049949   0.053756

             X33        X34        X35        X36        X37        X38  \
count  81.000000  81.000000  81.000000  81.000000  81.000000  81.000000
mean    0.018174   0.015966   0.016047   0.021471   0.020159   0.016568
std     0.019571   0.022635   0.022624   0.022117   0.023163   0.022769
min    -0.033408  -0.050801  -0.049865  -0.059165  -0.019302  -0.065808
25%     0.009468   0.006828   0.004639   0.014198   0.005524  -0.000532
50%     0.019881   0.020691   0.018664   0.023430   0.019119   0.016452
75%     0.030994   0.029237   0.028356   0.032911   0.031107   0.033009
max     0.078943   0.052320   0.109660   0.082318   0.161898   0.086865

             X39        X40
count  81.000000  81.000000
mean    0.013397   0.014300
std     0.026177   0.023365
min    -0.102526  -0.085331
25%     0.003221   0.001632
50%     0.015174   0.015284
75%     0.028950   0.030046
max     0.069319   0.052817

[8 rows x 41 columns]



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

          X0        X1        X2        X3        X4        X5        X6  \
0   0.043343  0.007224  0.062607  0.084278  0.031303  0.033711  0.012040
1   0.003607  0.000000  0.003607  0.014427  0.173123  0.007213  0.003607
2   0.057264  0.011265  0.049754  0.060081  0.048816  0.039428  0.019714
3   0.061688  0.006169  0.018507  0.006169  0.037013  0.030844  0.012338
4   0.044316  0.016618  0.022158  0.005539  0.094171  0.094171  0.027697
..       …         …         …         …         …         …         …
76  0.029377  0.006528  0.009139  0.007834  0.006528  0.030030  0.006528
77  0.013951  0.004650  0.013951  0.020926  0.225536  0.011626  0.004650
78  0.043550  0.005444  0.021775  0.000000  0.065325  0.065325  0.016331
79  0.052498  0.010862  0.032585  0.048877  0.023533  0.041636  0.014482
80  0.072080  0.006007  0.060066  0.081090  0.015017  0.039043  0.006007

          X7        X8        X9  …       X32       X33       X34       X35  \
0   0.036119  0.086686  0.040935  …  0.008227  0.013665  0.038255  0.044099
```

```
1   0.018034  0.010820  0.068528  …  0.024799  0.018705  0.034902  0.009449
2   0.035673  0.067591  0.093876  …  0.015342  0.003965  0.019154  0.016062
3   0.000000  0.030844  0.074026  … -0.003724  0.023506  0.022226  0.016978
4   0.088631  0.022158  0.105250  …  0.013815  0.040842 -0.020656 -0.025782
..       …         …        …  …        …         …         …         …
76  0.016973  0.015015  0.019584  …  0.034645  0.033330  0.032472  0.035423
77  0.041852  0.013951  0.113931  …  0.033529  0.011782  0.007488  0.015336
78  0.076213  0.016331  0.081656  … -0.004066  0.042660 -0.023316 -0.015435
79  0.043446  0.034395  0.047067  …  0.016308 -0.004267  0.021610  0.026354
80  0.027030  0.063070  0.075083  …  0.033385  0.007927 -0.004275 -0.014733

          X36       X37       X38       X39       X40           X41
0   0.010736  0.006827 -0.003336  0.038316 -0.008943  no_efectores
1   0.004614  0.028536  0.014109  0.009872  0.011257  no_efectores
2   0.007116 -0.004856  0.006426  0.015637  0.007255  no_efectores
3  -0.014163  0.001684  0.041285  0.001851  0.025556  no_efectores
4   0.034119 -0.000149 -0.015633 -0.044286 -0.010955  no_efectores
..       …         …         …         …         …           …
76  0.025017  0.026294  0.028337  0.026877  0.030794  no_efectores
77  0.023511  0.023403  0.008106 -0.008933  0.032391  no_efectores
78  0.031936  0.028329 -0.001729 -0.022458  0.011900  no_efectores
79  0.012610  0.007993  0.022201  0.024458  0.009941  no_efectores
80  0.036654 -0.005993 -0.007617  0.008675  0.018590  no_efectores

[81 rows x 42 columns]


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

                X0          X1          X2          X3          X4          X5  \
count   81.000000   81.000000   81.000000   81.000000   81.000000   81.000000
mean     0.045653    0.008924    0.033565    0.039937    0.051402    0.040500
std      0.019417    0.010146    0.027404    0.042712    0.053877    0.020685
min      0.000000    0.000000    0.002259    0.000000    0.000000    0.004518
25%      0.036440    0.003637    0.018026    0.006774    0.015077    0.026297
50%      0.045721    0.006412    0.024355    0.023752    0.032096    0.035897
75%      0.056974    0.011966    0.042169    0.060081    0.073709    0.052770
max      0.121140    0.077771    0.180493    0.242281    0.259210    0.094171

                X6          X7          X8          X9  …         X31         X32  \
count   81.000000   81.000000   81.000000   81.000000  …   81.000000   81.000000
mean     0.014085    0.037276    0.043252    0.062706  …    0.021438    0.011955
std      0.011487    0.025434    0.030912    0.032861  …    0.020276    0.027341
min      0.000000    0.000000    0.007649    0.010838  …   -0.032919   -0.085014
25%      0.005983    0.017948    0.017651    0.038030  …    0.009320    0.003509
50%      0.012061    0.035673    0.032252    0.066345  …    0.021396    0.018717
```

```
75%     0.016958   0.055544   0.062576   0.090460  …   0.036621   0.030822
max     0.072197   0.094136   0.131786   0.143935  …   0.070530   0.047976

              X33        X34        X35        X36        X37        X38  \
count   81.000000  81.000000  81.000000  81.000000  81.000000  81.000000
mean     0.026735   0.018984   0.011329   0.021492   0.012055   0.003854
std      0.025064   0.027683   0.023399   0.023439   0.021799   0.041805
min     -0.041853  -0.077572  -0.040858  -0.036532  -0.053549  -0.152931
25%      0.012421   0.004279  -0.005376   0.007366   0.001029  -0.008251
50%      0.026658   0.024749   0.013109   0.022038   0.008759   0.011607
75%      0.037607   0.035025   0.026625   0.034595   0.027361   0.026285
max      0.109302   0.080240   0.071167   0.083641   0.073630   0.060580

              X39        X40
count   81.000000  81.000000
mean     0.009237   0.014568
std      0.027255   0.020266
min     -0.067889  -0.041633
25%     -0.007220   0.000467
50%      0.016117   0.015483
75%      0.028948   0.029300
max      0.054425   0.057666

[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 Heterodera dataset 1, sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.060335	0.016644	0.024966	0.039530	0.031208	0.029127	0.004161	
3	0.019767	0.024708	0.039533	0.024708	0.029650	0.024708	0.014825	
4	0.014423	0.012363	0.019574	0.018544	0.006181	0.041209	0.008242	
5	0.025549	0.011355	0.031227	0.048260	0.011355	0.028388	0.002839	
7	0.024131	0.000000	0.045245	0.028655	0.009049	0.037704	0.012065	
..	…	…	…	…	…	…	…	
76	0.029099	0.000000	0.008559	0.008559	0.006847	0.023964	0.006847	
77	0.025844	0.011486	0.031587	0.048817	0.011486	0.025844	0.002872	
78	0.038419	0.014407	0.021611	0.014407	0.031215	0.040820	0.002401	
79	0.051086	0.016479	0.021423	0.041198	0.026367	0.036255	0.009888	
80	0.033304	0.007401	0.016652	0.020353	0.024053	0.031454	0.007401	

	X7	X8	X9	…	X32	X33	X34	X35	\
0	0.022886	0.060335	0.047852	…	0.029866	0.025941	0.012656	0.001198	
3	0.029650	0.054358	0.059300	…	0.040292	0.026592	0.021186	-0.002321	
4	0.047390	0.017514	0.024725	…	0.011847	0.032178	0.029237	0.027870	
5	0.028388	0.068132	0.028388	…	0.009621	0.011914	0.015134	0.025036	
7	0.012065	0.001508	0.013574	…	0.027625	0.026850	0.036457	0.041178	
..	…	…	…	…	…	…	…	…	
76	0.011982	0.017117	0.022253	…	0.036068	0.032763	0.036653	0.025407	
77	0.028716	0.066046	0.028716	…	0.020057	0.017693	0.022440	0.015977	
78	0.052826	0.036018	0.069634	…	0.028346	0.038424	0.028308	-0.019825	
79	0.029663	0.044494	0.054382	…	0.024933	0.009746	0.021250	0.017982	
80	0.022203	0.014802	0.057357	…	0.032463	0.037197	0.021045	0.039357	

	X36	X37	X38	X39	X40	X41
0	0.017419	0.022492	0.033754	0.011964	0.015797	efectores
3	0.008925	0.018859	0.037276	-0.005206	-0.005448	efectores
4	0.037862	0.058871	0.044245	0.013172	0.007218	efectores
5	0.028493	0.020724	-0.004970	0.003221	0.007290	efectores
7	0.027383	0.034595	0.030795	0.028950	0.037040	efectores
..	…	…	…	…	…	…
76	0.037558	0.036819	0.047683	0.029929	0.043832	efectores
77	0.020431	0.005524	-0.000739	0.010041	0.017865	efectores
78	0.007948	0.028842	0.050517	0.038713	0.022158	efectores
79	0.032635	0.003297	0.035567	0.028758	0.002198	efectores
80	0.019453	0.030680	0.017698	0.023157	0.036175	efectores

[63 rows x 42 columns]

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

	X0	X1	X2	X3	X4	X5	\
count	63.000000	63.000000	63.000000	63.000000	63.000000	63.000000	
mean	0.037761	0.010847	0.028569	0.037317	0.020610	0.034524	
std	0.017106	0.007435	0.014761	0.020436	0.012971	0.011100	
min	0.003824	0.000000	0.000000	0.002915	0.000000	0.014574	
25%	0.026028	0.005096	0.019299	0.019647	0.010456	0.025936	
50%	0.036388	0.010376	0.028298	0.041198	0.019210	0.031515	
75%	0.048365	0.015098	0.037120	0.048682	0.026925	0.041041	
max	0.101755	0.032495	0.074941	0.099023	0.056662	0.064565	

	X6	X7	X8	X9	…	X31	X32	\
count	63.000000	63.000000	63.000000	63.000000	…	63.000000	63.000000	
mean	0.008074	0.025707	0.046060	0.038584	…	0.018754	0.022955	
std	0.007279	0.012429	0.023209	0.020378	…	0.014066	0.014031	
min	0.000000	0.000000	0.001508	0.009441	…	-0.009128	-0.024159	
25%	0.002994	0.017424	0.029333	0.022841	…	0.009490	0.015935	
50%	0.005707	0.025566	0.045086	0.032167	…	0.014842	0.024923	
75%	0.011319	0.031369	0.063264	0.054761	…	0.028024	0.032636	
max	0.035691	0.056662	0.115262	0.090882	…	0.049949	0.053756	

	X33	X34	X35	X36	X37	X38	\
count	63.000000	63.000000	63.000000	63.000000	63.000000	63.000000	
mean	0.019710	0.022062	0.017978	0.024509	0.019428	0.019406	
std	0.014306	0.015455	0.015891	0.014288	0.015369	0.019415	
min	-0.033408	-0.019663	-0.025735	-0.014859	-0.019302	-0.017646	
25%	0.011063	0.013433	0.010311	0.018153	0.007071	-0.000128	
50%	0.020886	0.021186	0.018857	0.024119	0.021180	0.022784	
75%	0.030179	0.030945	0.027327	0.032532	0.031012	0.034262	

```
max     0.050248   0.052320   0.047527   0.074096   0.058871   0.051977

             X39        X40
count  63.000000  63.000000
mean    0.020596   0.020344
std     0.015337   0.015775
min    -0.005206  -0.014812
25%     0.008396   0.009971
50%     0.019039   0.017205
75%     0.030985   0.031212
max     0.055256   0.052817

[8 rows x 41 columns]
```

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

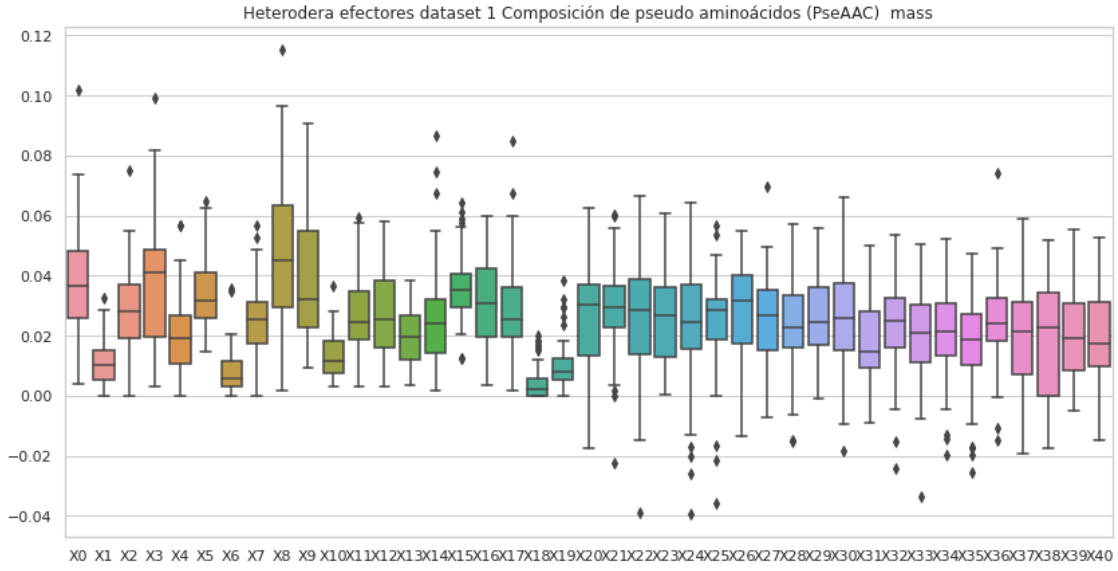
```
          X0        X1        X2        X3        X4        X5        X6  \
0   0.043343  0.007224  0.062607  0.084278  0.031303  0.033711  0.012040
1   0.003607  0.000000  0.003607  0.014427  0.173123  0.007213  0.003607
2   0.057264  0.011265  0.049754  0.060081  0.048816  0.039428  0.019714
3   0.061688  0.006169  0.018507  0.006169  0.037013  0.030844  0.012338
4   0.044316  0.016618  0.022158  0.005539  0.094171  0.094171  0.027697
..       …         …         …         …         …         …         …
74  0.050041  0.001472  0.020605  0.020605  0.001472  0.022077  0.007359
76  0.029377  0.006528  0.009139  0.007834  0.006528  0.030030  0.006528
78  0.043550  0.005444  0.021775  0.000000  0.065325  0.065325  0.016331
79  0.052498  0.010862  0.032585  0.048877  0.023533  0.041636  0.014482
80  0.072080  0.006007  0.060066  0.081090  0.015017  0.039043  0.006007

          X7        X8        X9  …       X32       X33       X34       X35  \
0   0.036119  0.086686  0.040935  …  0.008227  0.013665  0.038255  0.044099
1   0.018034  0.010820  0.068528  …  0.024799  0.018705  0.034902  0.009449
2   0.035673  0.067591  0.093876  …  0.015342  0.003965  0.019154  0.016062
3   0.000000  0.030844  0.074026  … -0.003724  0.023506  0.022226  0.016978
4   0.088631  0.022158  0.105250  …  0.013815  0.040842 -0.020656 -0.025782
..       …         …         …  …       …         …         …         …
74  0.007359  0.052985  0.016190  …  0.030822  0.041842  0.035025  0.013109
76  0.016973  0.015015  0.019584  …  0.034645  0.033330  0.032472  0.035423
78  0.076213  0.016331  0.081656  … -0.004066  0.042660 -0.023316 -0.015435
79  0.043446  0.034395  0.047067  …  0.016308 -0.004267  0.021610  0.026354
80  0.027030  0.063070  0.075083  …  0.033385  0.007927 -0.004275 -0.014733

         X36       X37       X38       X39       X40           X41
0   0.010736  0.006827 -0.003336  0.038316 -0.008943  no_efectores
```

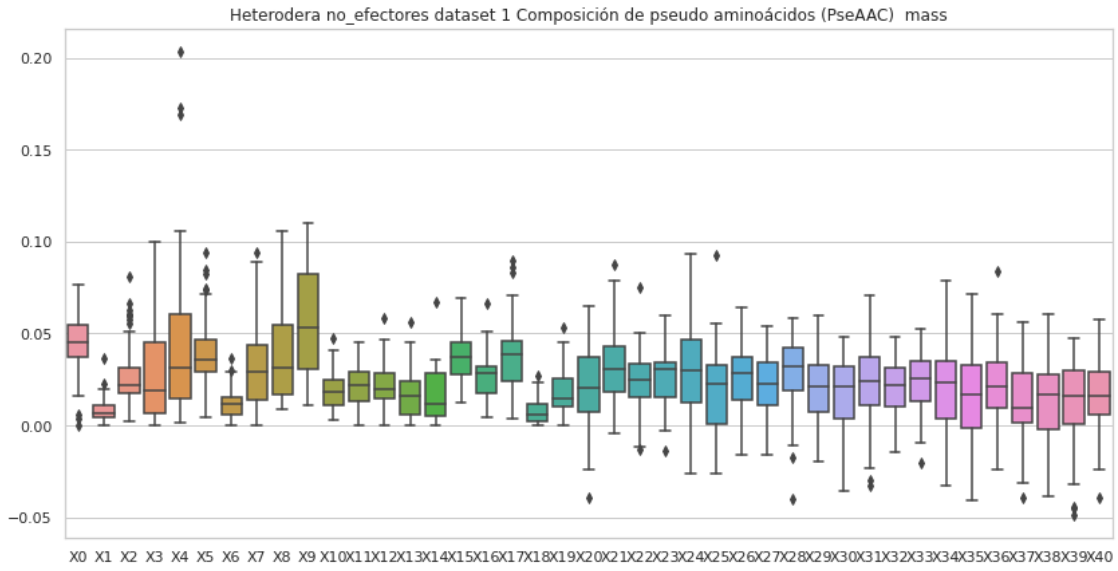
```
1   0.004614  0.028536  0.014109  0.009872  0.011257  no_efectores
2   0.007116 -0.004856  0.006426  0.015637  0.007255  no_efectores
3  -0.014163  0.001684  0.041285  0.001851  0.025556  no_efectores
4   0.034119 -0.000149 -0.015633 -0.044286 -0.010955  no_efectores
..       …       …       …       …       …           …
74  0.029444  0.022977  0.025664  0.042672  0.035429  no_efectores
76  0.025017  0.026294  0.028337  0.026877  0.030794  no_efectores
78  0.031936  0.028329 -0.001729 -0.022458  0.011900  no_efectores
79  0.012610  0.007993  0.022201  0.024458  0.009941  no_efectores
80  0.036654 -0.005993 -0.007617  0.008675  0.018590  no_efectores

[64 rows x 42 columns]
```

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

```
              X0         X1         X2         X3         X4         X5  \
count  64.000000  64.000000  64.000000  64.000000  64.000000  64.000000
mean    0.044537   0.008119   0.026954   0.030858   0.041441   0.040370
std     0.015279   0.006226   0.016614   0.029307   0.041612   0.019099
min     0.000000   0.000000   0.002259   0.000000   0.001455   0.004518
25%     0.037495   0.004427   0.017881   0.006290   0.014428   0.028885
50%     0.045035   0.006371   0.022141   0.019186   0.031246   0.035598
75%     0.054895   0.011188   0.031209   0.045611   0.060790   0.046756
max     0.076369   0.036237   0.080813   0.099828   0.203319   0.094171

              X6         X7         X8         X9  …        X31        X32  \
count  64.000000  64.000000  64.000000  64.000000  …  64.000000  64.000000
mean    0.011786   0.033356   0.038510   0.055695  …   0.022656   0.020322
std     0.007625   0.024744   0.025426   0.029050  …   0.020502   0.014019
min     0.000000   0.000000   0.009036   0.010838  …  -0.032919  -0.014257
25%     0.005880   0.013894   0.016501   0.030640  …   0.011026   0.010037
50%     0.011786   0.029244   0.031661   0.053315  …   0.024324   0.021580
75%     0.015468   0.043902   0.054458   0.082417  …   0.037315   0.031272
max     0.036578   0.094136   0.105462   0.109945  …   0.070530   0.047976

             X33        X34        X35        X36        X37        X38  \
count  64.000000  64.000000  64.000000  64.000000  64.000000  64.000000
mean    0.023957   0.018208   0.014766   0.022813   0.013297   0.014044
std     0.015846   0.022443   0.023803   0.019997   0.020023   0.020664
min    -0.020585  -0.032361  -0.040858  -0.024101  -0.039565  -0.038619
25%     0.013362   0.003416  -0.001039   0.009600   0.001497  -0.001829
50%     0.025768   0.023044   0.016922   0.021389   0.009691   0.016745
75%     0.035026   0.034933   0.032929   0.034238   0.028381   0.027905
max     0.052849   0.079067   0.071167   0.083641   0.055872   0.060580
```

```
              X39         X40
count   64.000000   64.000000
mean     0.012312    0.015879
std      0.022680    0.019443
min     -0.048738   -0.039412
25%      0.000824    0.005663
50%      0.016307    0.015913
75%      0.029969    0.029102
max      0.047442    0.057666

[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 Heterodera dataset
1, con valores atípicos.
Valores del documento csv.

          X0        X1        X2        X3        X4        X5        X6  \
0   0.099562  0.027465  0.041198  0.065230  0.051498  0.048064  0.006866
1   0.042225  0.022520  0.002815  0.016890  0.019705  0.025335  0.008445
2   0.043157  0.000000  0.018496  0.052405  0.015413  0.006165  0.003083
3   0.020177  0.025221  0.040354  0.025221  0.030265  0.025221  0.015133
4   0.034266  0.029371  0.046504  0.044057  0.014686  0.097904  0.019581
..       …         …         …         …         …         …         …
76  0.060217  0.000000  0.017711  0.017711  0.014169  0.049590  0.014169
77  0.008321  0.003698  0.010170  0.015717  0.003698  0.008321  0.000925
78  0.033262  0.012473  0.018710  0.012473  0.027025  0.035341  0.002079
79  0.059702  0.019259  0.025036  0.048147  0.030814  0.042369  0.011555
80  0.058021  0.012893  0.029010  0.035457  0.041904  0.054797  0.012893

          X7        X8        X9  …       X53       X54       X55       X56  \
0   0.037765  0.099562  0.078963  …  0.023174 -0.019445  0.044478  0.001549
1   0.022520  0.033780  0.039410  … -0.001912  0.045022  0.025549  0.012357
2   0.009248  0.024661  0.024661  … -0.013674  0.040076  0.071050  0.026859
3   0.030265  0.055486  0.060530  … -0.015506  0.028424  0.050183  0.012410
4   0.112589  0.041609  0.058742  … -0.034235  0.069589  0.076685 -0.050941
..       …         …         …  …         …         …         …         …
76  0.024795  0.035422  0.046048  …  0.010888  0.018649  0.011181  0.000742
77  0.009245  0.021264  0.009245  …  0.030659  0.007585  0.016404  0.017937
78  0.045735  0.031183  0.060287  … -0.009143  0.041900  0.020874  0.007895
79  0.034665  0.051998  0.063553  …  0.003999  0.027475  0.041144 -0.011095
80  0.038680  0.025787  0.099924  …  0.001890  0.002630  0.001003  0.014234

         X57       X58       X59       X60       X61        X62
0   0.025343 -0.043857 -0.040021  0.003554 -0.027176  efectores
1   0.006657 -0.001961  0.015465  0.030602  0.026771  efectores
2   0.061847  0.001498  0.038837 -0.005319  0.006396  efectores
3   0.001683 -0.033125 -0.008708  0.035351  0.055980  efectores
4  -0.024546 -0.031038  0.003000 -0.015930 -0.011026  efectores
..       …         …         …         …         …         …
76  0.025507 -0.025800  0.009574 -0.005976 -0.003948  efectores
77  0.026865  0.011303  0.023230  0.007456  0.017373  efectores
78  0.008061 -0.006522 -0.006263  0.005760 -0.002683  efectores
```

```
79  0.009678 -0.001283  0.022151 -0.007480  0.006841  efectores
80  0.001360  0.029704  0.022874  0.033652 -0.004354  efectores

[81 rows x 63 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	81.000000	81.000000	81.000000	81.000000	81.000000	81.000000	
mean	0.040912	0.012748	0.025517	0.031886	0.023668	0.036814	
std	0.032246	0.014349	0.018947	0.019721	0.020400	0.028887	
min	0.005694	0.000000	0.000000	0.000000	0.000000	0.005639	
25%	0.015754	0.002564	0.010624	0.016971	0.006180	0.013485	
50%	0.033262	0.005694	0.021691	0.028590	0.017974	0.025714	
75%	0.058021	0.019436	0.035401	0.042839	0.032176	0.050573	
max	0.143215	0.067994	0.085929	0.100250	0.092222	0.142302	

	X6	X7	X8	X9	…	X52	X53	\
count	81.000000	81.000000	81.000000	81.000000	…	81.000000	81.000000	
mean	0.009589	0.025744	0.037136	0.041629	…	0.000849	0.013079	
std	0.009296	0.021038	0.023236	0.033982	…	0.028078	0.022272	
min	0.000000	0.000000	0.001222	0.002224	…	-0.092880	-0.061806	
25%	0.001923	0.009545	0.021264	0.015907	…	-0.011402	0.000539	
50%	0.006866	0.021109	0.029825	0.034353	…	0.005576	0.016838	
75%	0.014907	0.035090	0.048448	0.060064	…	0.014324	0.028532	
max	0.042964	0.112589	0.100030	0.167050	…	0.084030	0.066320	

	X54	X55	X56	X57	X58	X59	\
count	81.000000	81.000000	81.000000	81.000000	81.000000	81.000000	
mean	0.008954	0.018759	0.005691	0.016058	0.000242	0.010022	
std	0.024111	0.024172	0.024195	0.017741	0.026374	0.023219	
min	-0.103201	-0.090165	-0.050941	-0.036962	-0.126876	-0.084330	
25%	0.001330	0.011181	-0.010086	0.008061	-0.006522	-0.002465	
50%	0.006379	0.018388	0.010109	0.017159	0.007506	0.015465	
75%	0.018265	0.029517	0.016433	0.027292	0.013873	0.024966	
max	0.075841	0.087025	0.101660	0.061847	0.054306	0.070928	

	X60	X61
count	81.000000	81.000000
mean	0.004204	0.011296
std	0.022497	0.021015
min	-0.079364	-0.054665
25%	-0.007077	0.001577
50%	0.007456	0.014842
75%	0.015198	0.025061

```
max     0.072967   0.062824

[8 rows x 62 columns]
```

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.011258	0.001876	0.016262	0.021891	0.008131	0.008756	0.003127	
1	0.000826	0.000000	0.000826	0.003303	0.039634	0.001651	0.000826	
2	0.070684	0.013905	0.061414	0.074160	0.060255	0.048668	0.024334	
3	0.105942	0.010594	0.031783	0.010594	0.063565	0.052971	0.021188	
4	0.008799	0.003300	0.004399	0.001100	0.018698	0.018698	0.005499	
..	…	…	…	…	…	…	…	
76	0.090657	0.020146	0.028204	0.024175	0.020146	0.092672	0.020146	
77	0.002792	0.000931	0.002792	0.004188	0.045134	0.002326	0.000931	
78	0.010303	0.001288	0.005151	0.000000	0.015454	0.015454	0.003863	
79	0.091488	0.018929	0.056786	0.085179	0.041012	0.072560	0.025238	
80	0.045782	0.003815	0.038152	0.051505	0.009538	0.024799	0.003815	

	X7	X8	X9	…	X53	X54	X55	X56	\
0	0.009382	0.022516	0.010633	…	0.033195	0.005289	0.020423	0.006339	
1	0.004129	0.002477	0.015688	…	0.014153	0.019415	0.014147	0.020173	
2	0.044033	0.083430	0.115876	…	-0.003780	-0.016699	-0.011888	-0.008118	
3	0.000000	0.052971	0.127130	…	0.027416	-0.026198	-0.039038	0.019060	
4	0.017598	0.004399	0.020897	…	0.023096	0.025177	0.010038	0.024500	
..	…	…	…	…	…	…	…	…	
76	0.052380	0.046336	0.060438	…	-0.007620	-0.026901	-0.014154	0.028465	
77	0.008375	0.002792	0.022800	…	0.013802	0.016115	0.015060	0.019819	
78	0.018030	0.003863	0.019317	…	0.020441	0.023634	0.009454	0.027861	
79	0.075715	0.059941	0.082024	…	-0.032707	-0.028000	-0.007531	0.067049	
80	0.017168	0.040059	0.047690	…	0.022586	-0.010103	0.001830	0.026100	

	X57	X58	X59	X60	X61	X62
0	0.035243	0.016593	0.038404	0.005627	0.026601	no_efectores
1	0.015600	0.020318	0.022253	0.020932	0.017597	no_efectores
2	0.014353	-0.003133	-0.004095	-0.004503	-0.024119	no_efectores
3	-0.018820	-0.009767	0.002385	0.094264	0.031746	no_efectores
4	0.014047	0.027439	0.009751	0.027055	0.011854	no_efectores
..	…	…	…	…	…	…
76	0.009541	-0.007928	0.002981	-0.008498	-0.021002	no_efectores
77	0.016571	0.020848	0.015464	0.021629	0.016440	no_efectores
78	0.017311	0.030579	0.007317	0.024923	0.009287	no_efectores
79	0.056117	0.020154	0.012033	0.021509	-0.016688	no_efectores

```
80  0.024870  0.007473 -0.012134 -0.012812  0.032649  no_efectores

[81 rows x 63 columns]
```

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

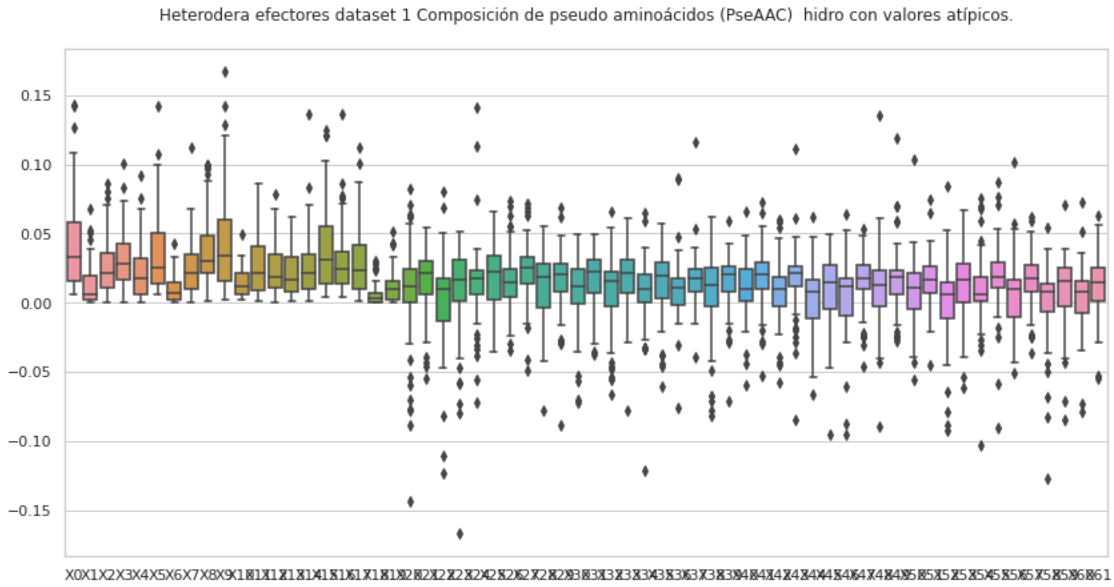
	X0	X1	X2	X3	X4	X5	\
count	81.000000	81.000000	81.000000	81.000000	81.000000	81.000000	
mean	0.046064	0.010168	0.027490	0.031192	0.030449	0.035943	
std	0.038354	0.014082	0.019742	0.026418	0.021618	0.029409	
min	0.000000	0.000000	0.000480	0.000000	0.000000	0.000961	
25%	0.011378	0.001172	0.008242	0.008169	0.015454	0.015436	
50%	0.039714	0.003815	0.026429	0.026861	0.023774	0.022118	
75%	0.074499	0.015016	0.040876	0.049218	0.041379	0.050507	
max	0.166074	0.088573	0.069751	0.091831	0.075159	0.138451	

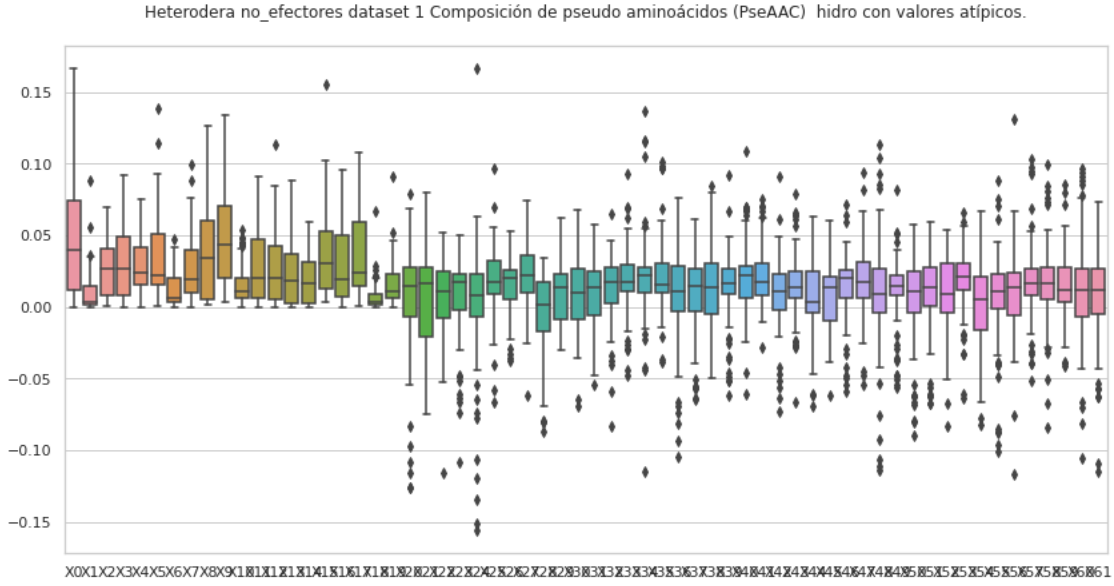
	X6	X7	X8	X9	…	X52	X53	\
count	81.000000	81.000000	81.000000	81.000000	…	81.000000	81.000000	
mean	0.011419	0.026380	0.039687	0.048582	…	0.008877	0.017059	
std	0.010419	0.021059	0.031654	0.035063	…	0.026833	0.022406	
min	0.000000	0.000000	0.001715	0.002971	…	-0.083070	-0.063682	
25%	0.003815	0.010139	0.005162	0.020087	…	-0.004469	0.011350	
50%	0.006453	0.019514	0.034284	0.043721	…	0.009079	0.021456	
75%	0.020146	0.039624	0.059941	0.070105	…	0.030784	0.030105	
max	0.047311	0.099332	0.126385	0.133657	…	0.053328	0.066015	

	X54	X55	X56	X57	X58	X59	\
count	81.000000	81.000000	81.000000	81.000000	81.000000	81.000000	
mean	0.001775	0.006030	0.009017	0.016992	0.014500	0.014911	
std	0.027818	0.035347	0.028963	0.028288	0.029339	0.023938	
min	-0.082285	-0.101012	-0.116418	-0.051109	-0.084074	-0.041143	
25%	-0.016037	-0.000835	-0.005518	0.008085	0.004899	0.003863	
50%	0.005327	0.010923	0.013959	0.016072	0.016593	0.011577	
75%	0.020997	0.022912	0.024153	0.026347	0.027439	0.027527	
max	0.067109	0.088475	0.131161	0.103395	0.099040	0.085785	

	X60	X61
count	81.000000	81.000000
mean	0.012874	0.004551
std	0.036636	0.033152
min	-0.105471	-0.114731
25%	-0.006604	-0.005204
50%	0.012257	0.011573
75%	0.027055	0.026398
max	0.096465	0.073088

[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 Heterodera dataset 1, sin
valores atípicos.
Valores del documento csv.
```

	X0	X1	X2	X3	X4	X5	X6	\
0	0.099562	0.027465	0.041198	0.065230	0.051498	0.048064	0.006866	
1	0.042225	0.022520	0.002815	0.016890	0.019705	0.025335	0.008445	
2	0.043157	0.000000	0.018496	0.052405	0.015413	0.006165	0.003083	
5	0.008323	0.003699	0.010173	0.015721	0.003699	0.009248	0.000925	
7	0.019558	0.000000	0.036671	0.023225	0.007334	0.030559	0.009779	
..	…	…	…	…	…	…	…	
76	0.060217	0.000000	0.017711	0.017711	0.014169	0.049590	0.014169	
77	0.008321	0.003698	0.010170	0.015717	0.003698	0.008321	0.000925	
78	0.033262	0.012473	0.018710	0.012473	0.027025	0.035341	0.002079	
79	0.059702	0.019259	0.025036	0.048147	0.030814	0.042369	0.011555	
80	0.058021	0.012893	0.029010	0.035457	0.041904	0.054797	0.012893	

	X7	X8	X9	…	X53	X54	X55	X56	\
0	0.037765	0.099562	0.078963	…	0.023174	-0.019445	0.044478	0.001549	
1	0.022520	0.033780	0.039410	…	-0.001912	0.045022	0.025549	0.012357	
2	0.009248	0.024661	0.024661	…	-0.013674	0.040076	0.071050	0.026859	
5	0.009248	0.022195	0.009248	…	0.028203	0.007396	0.016914	0.016569	
7	0.009779	0.001222	0.011001	…	0.027589	0.001496	0.020426	0.001800	
..	…	…	…	…	…	…	…	…	
76	0.024795	0.035422	0.046048	…	0.010888	0.018649	0.011181	0.000742	
77	0.009245	0.021264	0.009245	…	0.030659	0.007585	0.016404	0.017937	
78	0.045735	0.031183	0.060287	…	-0.009143	0.041900	0.020874	0.007895	
79	0.034665	0.051998	0.063553	…	0.003999	0.027475	0.041144	-0.011095	
80	0.038680	0.025787	0.099924	…	0.001890	0.002630	0.001003	0.014234	

	X57	X58	X59	X60	X61	X62
0	0.025343	-0.043857	-0.040021	0.003554	-0.027176	efectores
1	0.006657	-0.001961	0.015465	0.030602	0.026771	efectores
2	0.061847	0.001498	0.038837	-0.005319	0.006396	efectores
5	0.026172	0.010183	0.025124	0.009938	0.019537	efectores
7	0.023815	0.005954	0.026128	0.001697	0.036487	efectores
..	…	…	…	…	…	…
76	0.025507	-0.025800	0.009574	-0.005976	-0.003948	efectores
77	0.026865	0.011303	0.023230	0.007456	0.017373	efectores
78	0.008061	-0.006522	-0.006263	0.005760	-0.002683	efectores
79	0.009678	-0.001283	0.022151	-0.007480	0.006841	efectores
80	0.001360	0.029704	0.022874	0.033652	-0.004354	efectores

```
[65 rows x 63 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	65.000000	65.000000	65.000000	65.000000	65.000000	65.000000	
mean	0.032420	0.009408	0.020616	0.028012	0.018265	0.030050	
std	0.024055	0.010298	0.013911	0.016528	0.015241	0.024716	
min	0.005694	0.000000	0.000000	0.000000	0.000000	0.005639	
25%	0.010110	0.002336	0.010170	0.016347	0.004772	0.010719	
50%	0.025006	0.004052	0.018496	0.025287	0.014169	0.020137	
75%	0.051580	0.014737	0.028477	0.036309	0.029582	0.042760	
max	0.099562	0.045678	0.068160	0.073686	0.060100	0.107098	

	X6	X7	X8	X9	…	X52	X53	\
count	65.000000	65.000000	65.000000	65.000000	…	65.000000	65.000000	
mean	0.008000	0.020570	0.032698	0.031664	…	0.006414	0.018128	
std	0.007514	0.016500	0.020706	0.023803	…	0.018189	0.014867	
min	0.000000	0.000000	0.001222	0.002224	…	-0.045382	-0.013904	
25%	0.001772	0.009245	0.020763	0.010131	…	-0.002162	0.006979	
50%	0.005850	0.014039	0.024661	0.024797	…	0.008835	0.019941	
75%	0.011697	0.031455	0.036454	0.047579	…	0.015717	0.028667	
max	0.029068	0.068198	0.099562	0.099924	…	0.052369	0.057092	

	X54	X55	X56	X57	X58	X59	\
count	65.000000	65.000000	65.000000	65.000000	65.000000	65.000000	
mean	0.010712	0.019224	0.004301	0.017383	0.004684	0.013646	
std	0.016118	0.014562	0.016924	0.013787	0.014577	0.015384	
min	-0.030312	-0.024473	-0.042974	-0.016157	-0.043857	-0.040021	
25%	0.002630	0.013045	-0.003719	0.009525	-0.002880	0.002745	
50%	0.007396	0.018313	0.010109	0.017817	0.008475	0.020225	
75%	0.018265	0.026019	0.015533	0.026865	0.012330	0.024966	
max	0.068146	0.071050	0.033161	0.061847	0.031981	0.038837	

	X60	X61
count	65.000000	65.000000
mean	0.005521	0.013783
std	0.013205	0.015313
min	-0.023755	-0.029161
25%	-0.005100	0.006441
50%	0.007456	0.016662
75%	0.011947	0.025061
max	0.035626	0.047109

[8 rows x 62 columns]

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

```
          X0        X1        X2        X3        X4        X5        X6  \
0   0.011258  0.001876  0.016262  0.021891  0.008131  0.008756  0.003127
1   0.000826  0.000000  0.000826  0.003303  0.039634  0.001651  0.000826
2   0.070684  0.013905  0.061414  0.074160  0.060255  0.048668  0.024334
3   0.105942  0.010594  0.031783  0.010594  0.063565  0.052971  0.021188
4   0.008799  0.003300  0.004399  0.001100  0.018698  0.018698  0.005499
..       …         …         …         …         …         …         …
74  0.051290  0.001509  0.021119  0.021119  0.001509  0.022628  0.007543
76  0.090657  0.020146  0.028204  0.024175  0.020146  0.092672  0.020146
77  0.002792  0.000931  0.002792  0.004188  0.045134  0.002326  0.000931
78  0.010303  0.001288  0.005151  0.000000  0.015454  0.015454  0.003863
80  0.045782  0.003815  0.038152  0.051505  0.009538  0.024799  0.003815

          X7        X8        X9  …        X53       X54       X55       X56  \
0   0.009382  0.022516  0.010633  …   0.033195  0.005289  0.020423  0.006339
1   0.004129  0.002477  0.015688  …   0.014153  0.019415  0.014147  0.020173
2   0.044033  0.083430  0.115876  … -0.003780 -0.016699 -0.011888 -0.008118
3   0.000000  0.052971  0.127130  …   0.027416 -0.026198 -0.039038  0.019060
4   0.017598  0.004399  0.020897  …   0.023096  0.025177  0.010038  0.024500
..       …         …         …  …         …         …         …         …
74  0.007543  0.054307  0.016594  …   0.022182 -0.001022  0.018050  0.016862
76  0.052380  0.046336  0.060438  … -0.007620 -0.026901 -0.014154  0.028465
77  0.008375  0.002792  0.022800  …   0.013802  0.016115  0.015060  0.019819
78  0.018030  0.003863  0.019317  …   0.020441  0.023634  0.009454  0.027861
80  0.017168  0.040059  0.047690  …   0.022586 -0.010103  0.001830  0.026100

         X57       X58       X59       X60       X61           X62
0   0.035243  0.016593  0.038404  0.005627  0.026601  no_efectores
1   0.015600  0.020318  0.022253  0.020932  0.017597  no_efectores
2   0.014353 -0.003133 -0.004095 -0.004503 -0.024119  no_efectores
3  -0.018820 -0.009767  0.002385  0.094264  0.031746  no_efectores
4   0.014047  0.027439  0.009751  0.027055  0.011854  no_efectores
..       …         …         …         …         …           …
74  0.031662  0.016679  0.028582 -0.013683  0.019239  no_efectores
76  0.009541 -0.007928  0.002981 -0.008498 -0.021002  no_efectores
77  0.016571  0.020848  0.015464  0.021629  0.016440  no_efectores
78  0.017311  0.030579  0.007317  0.024923  0.009287  no_efectores
80  0.024870  0.007473 -0.012134 -0.012812  0.032649  no_efectores

[61 rows x 63 columns]
```

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

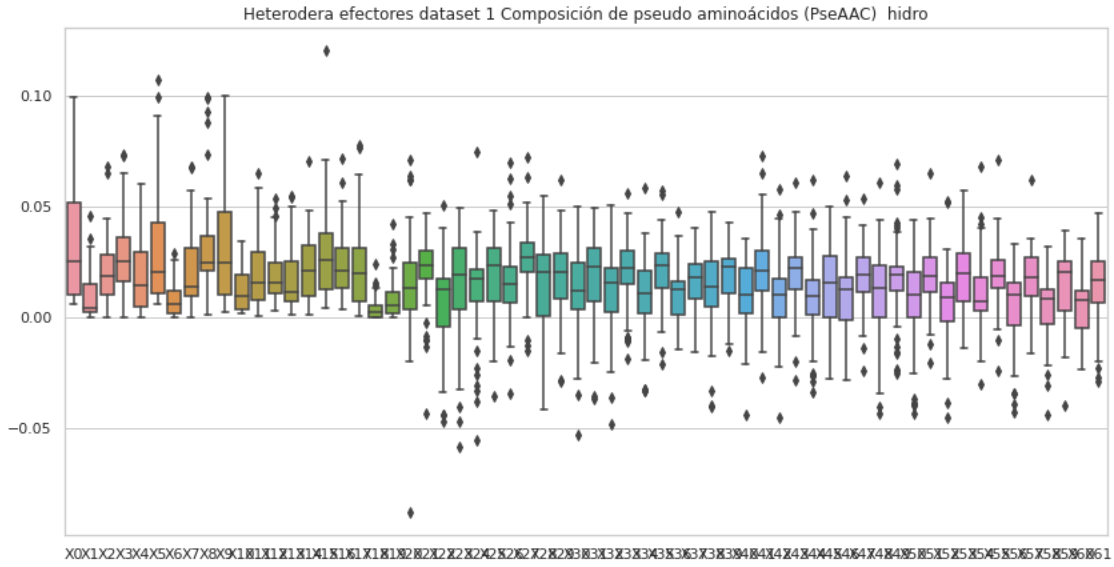
	X0	X1	X2	X3	X4	X5	\
count	61.000000	61.000000	61.000000	61.000000	61.000000	61.000000	
mean	0.036223	0.007360	0.021164	0.023711	0.024087	0.029726	
std	0.034285	0.009980	0.017082	0.022761	0.018380	0.025236	
min	0.000000	0.000000	0.000480	0.000000	0.000000	0.000961	
25%	0.011041	0.001125	0.005151	0.002098	0.009538	0.015242	
50%	0.021489	0.001968	0.017177	0.018878	0.018404	0.018698	
75%	0.053202	0.010951	0.032262	0.036349	0.036569	0.047548	
max	0.122196	0.036004	0.061414	0.078684	0.069602	0.114013	

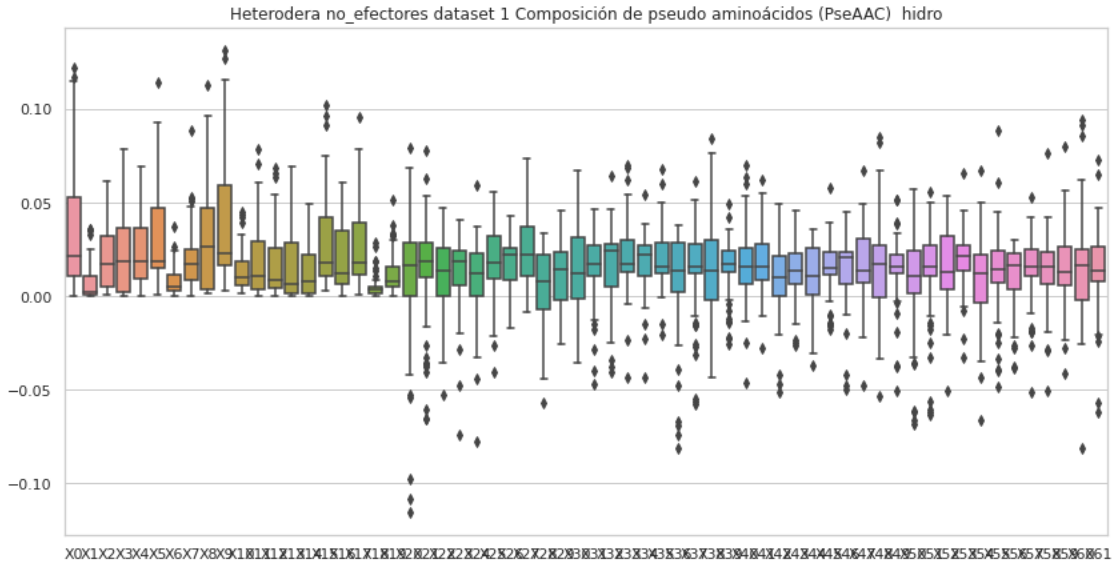
	X6	X7	X8	X9	…	X52	X53	\
count	61.000000	61.000000	61.000000	61.000000	…	61.000000	61.000000	
mean	0.008475	0.020499	0.031256	0.039139	…	0.014684	0.019593	
std	0.008172	0.016803	0.028850	0.032177	…	0.019546	0.015539	
min	0.000000	0.000000	0.001715	0.002971	…	-0.050488	-0.032840	
25%	0.003281	0.008850	0.003863	0.016594	…	0.003827	0.013862	
50%	0.004992	0.017208	0.026249	0.022734	…	0.012981	0.021456	
75%	0.011797	0.024916	0.046864	0.059113	…	0.032286	0.027416	
max	0.037271	0.088670	0.113103	0.131415	…	0.053326	0.066015	

	X54	X55	X56	X57	X58	X59	\
count	61.000000	61.000000	61.000000	61.000000	61.000000	61.000000	
mean	0.007350	0.012347	0.010022	0.014244	0.012965	0.014742	
std	0.023045	0.023272	0.017386	0.018173	0.019100	0.018942	
min	-0.066245	-0.048591	-0.038347	-0.051109	-0.050323	-0.041143	
25%	-0.003510	0.007323	0.003662	0.011048	0.006442	0.006081	
50%	0.012464	0.014147	0.016862	0.016072	0.016024	0.012862	
75%	0.021919	0.024417	0.023105	0.025063	0.023794	0.026695	
max	0.067109	0.088475	0.030280	0.052869	0.076469	0.080200	

	X60	X61
count	61.000000	61.000000
mean	0.014478	0.013443
std	0.026900	0.022602
min	-0.081129	-0.062348
25%	-0.002215	0.007610
50%	0.016315	0.013519
75%	0.024923	0.026601
max	0.094264	0.073088

[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 Heterodera dataset 1,
con valores atípicos.
Valores del documento csv.

          X0        X1        X2        X3        X4        X5        X6  \
0  -0.038769  0.031259  0.002415  0.011131 -0.020154  0.030932  0.038987
1   0.077828  0.012054 -0.121769  0.056352 -0.023107  0.101445  0.000900
2  -0.137930 -0.007231  0.020661  0.056659 -0.034381 -0.009534  0.011462
3  -0.006994  0.011514  0.120517 -0.003738 -0.002641  0.036674 -0.073480
4   0.119770 -0.044673 -0.024220 -0.031648 -0.063063 -0.025467  0.105083
..       …         …         …         …         …         …         …
76  0.154887  0.142653 -0.035713  0.035094 -0.038459 -0.034522 -0.059078
77  0.052752  0.045389  0.076326  0.037815  0.114526  0.059162  0.095152
78  0.129281 -0.014146  0.040115  0.079577  0.038408  0.131465 -0.063595
79  0.064474  0.017109 -0.073589 -0.033216 -0.010125  0.043477  0.105925
80  0.028575  0.070338 -0.061186 -0.001816 -0.002058  0.015923  0.051552

          X7        X8        X9       X10       X11       X12        X13
0  -0.010492  0.053747  0.054128  0.002833 -0.022410  0.040367  efectores
```

```
1   0.057265  0.009527 -0.048216  0.047199 -0.113203 -0.021783  efectores
2  -0.006708 -0.019858  0.032800  0.058270  0.040313 -0.185185  efectores
3  -0.040932  0.139373  0.035370  0.019275  0.040369  0.096267  efectores
4   0.045049 -0.045470  0.020705  0.044133 -0.004578 -0.085961  efectores
..       …       …       …       …       …       …       …
76  0.146802  0.060316  0.148114  0.168644  0.039016  0.057945  efectores
77  0.076969  0.015624  0.010191  0.072782 -0.014074  0.018624  efectores
78 -0.008268 -0.020587  0.029253 -0.060016  0.013854  0.067851  efectores
79  0.002035  0.052976 -0.038353  0.056429 -0.051149  0.020793  efectores
80 -0.049432  0.027267 -0.058554 -0.093642  0.002049  0.027824  efectores

[81 rows x 14 columns]
```

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

```
              X0         X1         X2         X3         X4         X5  \
count  81.000000  81.000000  81.000000  81.000000  81.000000  81.000000
mean    0.021458   0.032031   0.009564   0.029125   0.011675   0.023817
std     0.078131   0.068911   0.077599   0.061383   0.072703   0.071746
min    -0.216972  -0.096133  -0.222997  -0.104221  -0.184244  -0.124998
25%    -0.020616  -0.007729  -0.048777  -0.015219  -0.037567  -0.019418
50%     0.028575   0.031259   0.020661   0.029907   0.001549   0.030932
75%     0.060420   0.057391   0.071666   0.056352   0.062580   0.057397
max     0.239309   0.334429   0.148569   0.260304   0.175573   0.209816

              X6         X7         X8         X9        X10        X11  \
count  81.000000  81.000000  81.000000  81.000000  81.000000  81.000000
mean    0.022296   0.015529   0.020861   0.020626   0.023931   0.000656
std     0.073495   0.068944   0.072314   0.063286   0.089909   0.056084
min    -0.244908  -0.253155  -0.169734  -0.141630  -0.121615  -0.175983
25%    -0.013138  -0.017228  -0.019858  -0.026273  -0.021926  -0.032602
50%     0.028720   0.010465   0.010040   0.014681   0.020088  -0.006727
75%     0.081158   0.057265   0.052010   0.058106   0.058802   0.040313
max     0.164990   0.181859   0.283057   0.166436   0.411132   0.106875

             X12
count  81.000000
mean    0.009283
std     0.067297
min    -0.185185
25%    -0.029733
50%     0.018624
75%     0.040367
max     0.252684
```

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.016356	-0.005967	0.067647	0.029868	-0.027867	-0.066628	0.035680	
1	0.087264	0.024981	0.124344	0.122770	0.031860	0.061186	0.215549	
2	-0.044589	0.004445	0.040506	-0.033726	-0.025209	-0.000943	-0.014504	
3	0.043150	0.004835	0.007311	0.016026	0.067616	0.083122	-0.170190	
4	-0.032342	0.135388	-0.064402	0.027851	0.125018	-0.114852	0.088371	
..	…	…	…	…	…	…	…	
76	0.099114	0.053502	0.092183	0.047710	0.112952	0.057618	0.074242	
77	-0.019287	0.039536	-0.041690	0.100384	0.031247	0.032895	0.038710	
78	-0.100984	0.075671	-0.005136	-0.052009	0.157874	-0.104895	0.047333	
79	0.072141	-0.008676	0.075559	0.045462	-0.093315	0.000801	-0.008817	
80	0.123031	-0.043752	0.005875	0.051749	-0.093804	-0.016990	0.117190	

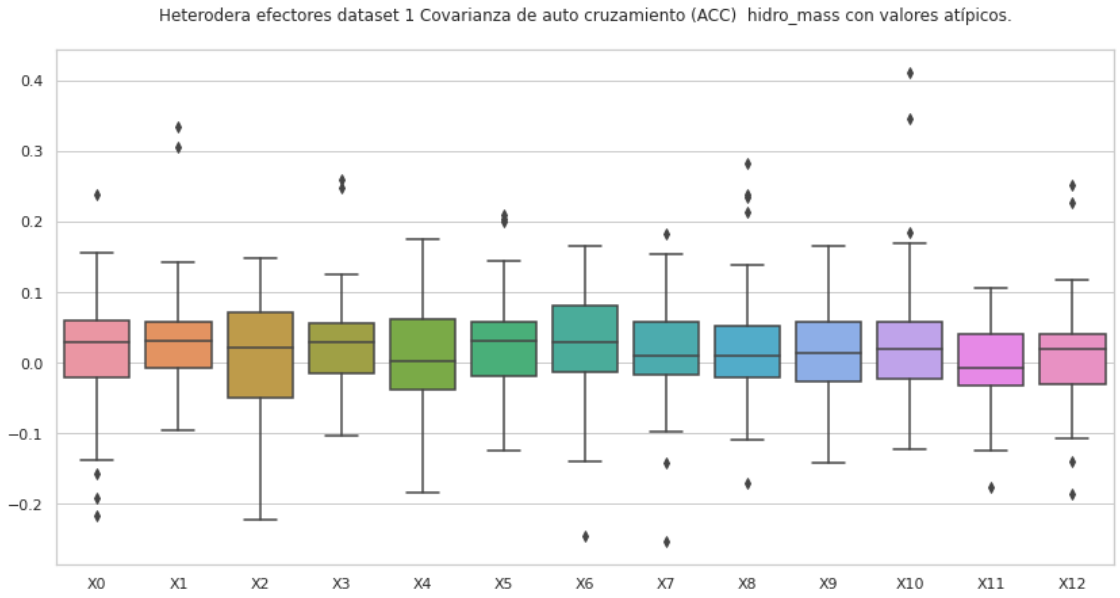
	X7	X8	X9	X10	X11	X12	X13
0	0.101642	0.073049	0.003192	0.067882	0.075707	0.010843	no_efectores
1	0.106735	0.098941	0.010120	0.071415	-0.017568	0.042418	no_efectores
2	0.001867	-0.010332	-0.028280	-0.012661	-0.019196	0.008247	no_efectores
3	0.019508	0.067043	0.056666	0.042142	-0.108111	-0.119458	no_efectores
4	-0.092463	0.037825	-0.102604	-0.140277	-0.053252	-0.026306	no_efectores
..	…	…	…	…	…	…	…
76	0.081382	0.047450	0.026265	-0.037723	0.017300	0.052555	no_efectores
77	0.122265	0.016981	0.004828	-0.051503	0.027153	0.075163	no_efectores
78	-0.095514	0.099682	-0.055245	-0.118194	0.138249	-0.086957	no_efectores
79	0.023318	-0.020391	0.042288	-0.066117	0.036480	0.002699	no_efectores
80	-0.017556	0.009057	0.079370	0.009253	-0.102467	0.064053	no_efectores

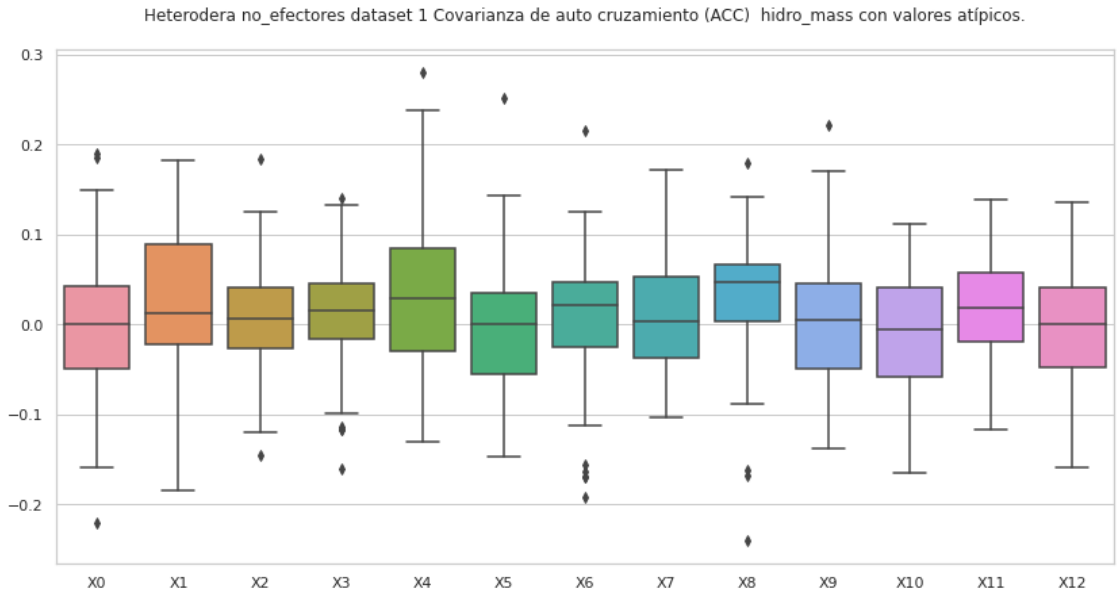
[81 rows x 14 columns]

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

	X0	X1	X2	X3	X4	X5	\
count	81.000000	81.000000	81.000000	81.000000	81.000000	81.000000	
mean	-0.003567	0.029406	0.006690	0.009861	0.032399	-0.004135	
std	0.079208	0.074722	0.060888	0.059054	0.084697	0.070672	
min	-0.220949	-0.184909	-0.145183	-0.160370	-0.131271	-0.147052	
25%	-0.049103	-0.021636	-0.026312	-0.016637	-0.029140	-0.055896	
50%	-0.000053	0.011896	0.005875	0.015923	0.029341	0.000801	
75%	0.042866	0.088728	0.040530	0.045455	0.084655	0.035477	

```
max     0.189693   0.181984   0.184017   0.139800   0.279843   0.250784

               X6         X7         X8         X9        X10        X11  \
count  81.000000  81.000000  81.000000  81.000000  81.000000  81.000000
mean    0.010225   0.007086   0.028191   0.002657  -0.011629   0.013621
std     0.070256   0.062814   0.067427   0.067657   0.065964   0.062432
min    -0.191915  -0.103552  -0.240303  -0.137878  -0.165737  -0.117730
25%    -0.025987  -0.037019   0.003223  -0.049266  -0.057752  -0.019196
50%     0.021109   0.003728   0.046992   0.004693  -0.005526   0.018842
75%     0.046526   0.052792   0.067043   0.046093   0.041124   0.057696
max     0.215549   0.171175   0.179643   0.220772   0.110875   0.138249

              X12
count  81.000000
mean   -0.008622
std     0.063997
min    -0.159479
25%    -0.048151
50%    -0.000443
75%     0.041096
max     0.134893
```





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 Heterodera dataset 1,
sin valores atípicos.
Valores del documento csv.

          X0        X1        X2        X3        X4        X5        X6  \
0  -0.038769  0.031259  0.002415  0.011131 -0.020154  0.030932  0.038987
1   0.077828  0.012054 -0.121769  0.056352 -0.023107  0.101445  0.000900
2  -0.137930 -0.007231  0.020661  0.056659 -0.034381 -0.009534  0.011462
3  -0.006994  0.011514  0.120517 -0.003738 -0.002641  0.036674 -0.073480
4   0.119770 -0.044673 -0.024220 -0.031648 -0.063063 -0.025467  0.105083
..       …         …         …         …         …         …         …
76  0.154887  0.142653 -0.035713  0.035094 -0.038459 -0.034522 -0.059078
77  0.052752  0.045389  0.076326  0.037815  0.114526  0.059162  0.095152
78  0.129281 -0.014146  0.040115  0.079577  0.038408  0.131465 -0.063595
79  0.064474  0.017109 -0.073589 -0.033216 -0.010125  0.043477  0.105925
80  0.028575  0.070338 -0.061186 -0.001816 -0.002058  0.015923  0.051552

          X7        X8        X9       X10       X11       X12        X13
0  -0.010492  0.053747  0.054128  0.002833 -0.022410  0.040367  efectores
1   0.057265  0.009527 -0.048216  0.047199 -0.113203 -0.021783  efectores
2  -0.006708 -0.019858  0.032800  0.058270  0.040313 -0.185185  efectores
3  -0.040932  0.139373  0.035370  0.019275  0.040369  0.096267  efectores
4   0.045049 -0.045470  0.020705  0.044133 -0.004578 -0.085961  efectores
..       …         …         …         …         …         …         …
76  0.146802  0.060316  0.148114  0.168644  0.039016  0.057945  efectores
```

```
77  0.076969  0.015624  0.010191  0.072782 -0.014074  0.018624  efectores
78 -0.008268 -0.020587  0.029253 -0.060016  0.013854  0.067851  efectores
79  0.002035  0.052976 -0.038353  0.056429 -0.051149  0.020793  efectores
80 -0.049432  0.027267 -0.058554 -0.093642  0.002049  0.027824  efectores

[75 rows x 14 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	75.000000	75.000000	75.000000	75.000000	75.000000	75.000000	
mean	0.025847	0.026595	0.011904	0.024140	0.012769	0.015707	
std	0.067170	0.052025	0.075139	0.050184	0.073701	0.064387	
min	-0.157326	-0.096133	-0.151052	-0.104221	-0.184244	-0.124998	
25%	-0.005321	-0.007480	-0.046922	-0.012032	-0.038013	-0.022443	
50%	0.037918	0.032832	0.020661	0.029907	-0.001823	0.029526	
75%	0.060803	0.057224	0.075500	0.053786	0.064067	0.050151	
max	0.155608	0.142653	0.148569	0.125409	0.175573	0.199367	

	X6	X7	X8	X9	X10	X11	\
count	75.000000	75.000000	75.000000	75.000000	75.000000	75.000000	
mean	0.025134	0.020712	0.012689	0.025678	0.019436	0.000383	
std	0.065052	0.055256	0.057481	0.060544	0.069134	0.052186	
min	-0.111139	-0.092492	-0.169734	-0.141630	-0.121615	-0.124822	
25%	-0.012618	-0.013364	-0.018845	-0.017306	-0.017645	-0.030585	
50%	0.028720	0.011256	0.009527	0.019438	0.022147	-0.007540	
75%	0.076108	0.056628	0.049307	0.068474	0.058536	0.039040	
max	0.164990	0.181859	0.213635	0.166436	0.185237	0.106875	

	X12
count	75.000000
mean	0.001032
std	0.056615
min	-0.185185
25%	-0.033048
50%	0.017155
75%	0.033988
max	0.117940

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

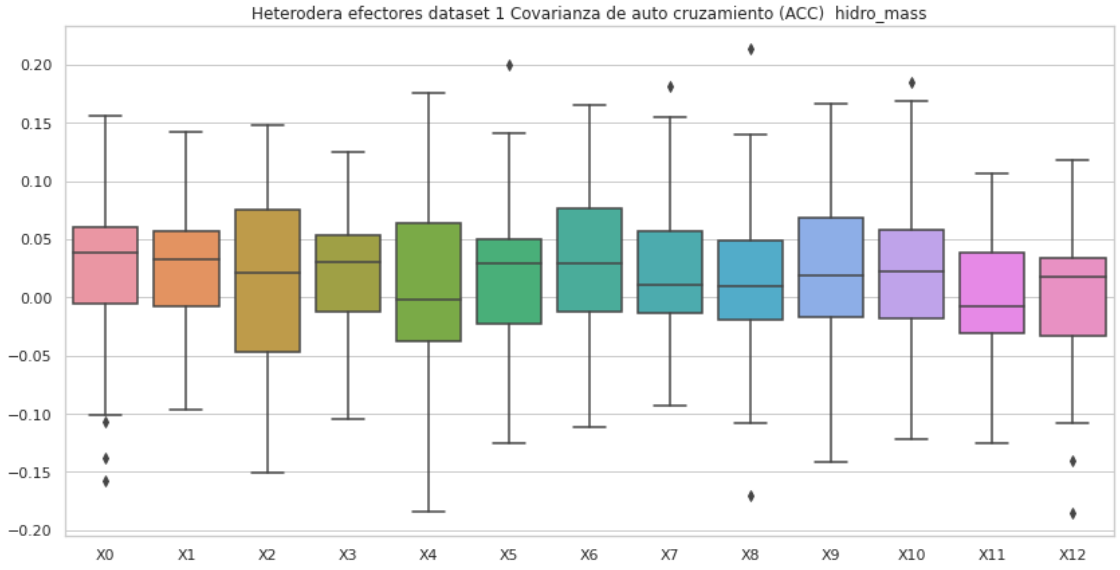
```
          X0        X1        X2        X3        X4        X5        X6  \
0   0.016356 -0.005967  0.067647  0.029868 -0.027867 -0.066628  0.035680
1   0.087264  0.024981  0.124344  0.122770  0.031860  0.061186  0.215549
2  -0.044589  0.004445  0.040506 -0.033726 -0.025209 -0.000943 -0.014504
3   0.043150  0.004835  0.007311  0.016026  0.067616  0.083122 -0.170190
4  -0.032342  0.135388 -0.064402  0.027851  0.125018 -0.114852  0.088371
..       …         …         …         …         …         …         …
76  0.099114  0.053502  0.092183  0.047710  0.112952  0.057618  0.074242
77 -0.019287  0.039536 -0.041690  0.100384  0.031247  0.032895  0.038710
78 -0.100984  0.075671 -0.005136 -0.052009  0.157874 -0.104895  0.047333
79  0.072141 -0.008676  0.075559  0.045462 -0.093315  0.000801 -0.008817
80  0.123031 -0.043752  0.005875  0.051749 -0.093804 -0.016990  0.117190

          X7        X8        X9       X10       X11       X12           X13
0   0.101642  0.073049  0.003192  0.067882  0.075707  0.010843  no_efectores
1   0.106735  0.098941  0.010120  0.071415 -0.017568  0.042418  no_efectores
2   0.001867 -0.010332 -0.028280 -0.012661 -0.019196  0.008247  no_efectores
3   0.019508  0.067043  0.056666  0.042142 -0.108111 -0.119458  no_efectores
4  -0.092463  0.037825 -0.102604 -0.140277 -0.053252 -0.026306  no_efectores
..       …         …         …         …         …         …           …
76  0.081382  0.047450  0.026265 -0.037723  0.017300  0.052555  no_efectores
77  0.122265  0.016981  0.004828 -0.051503  0.027153  0.075163  no_efectores
78 -0.095514  0.099682 -0.055245 -0.118194  0.138249 -0.086957  no_efectores
79  0.023318 -0.020391  0.042288 -0.066117  0.036480  0.002699  no_efectores
80 -0.017556  0.009057  0.079370  0.009253 -0.102467  0.064053  no_efectores

[79 rows x 14 columns]
```

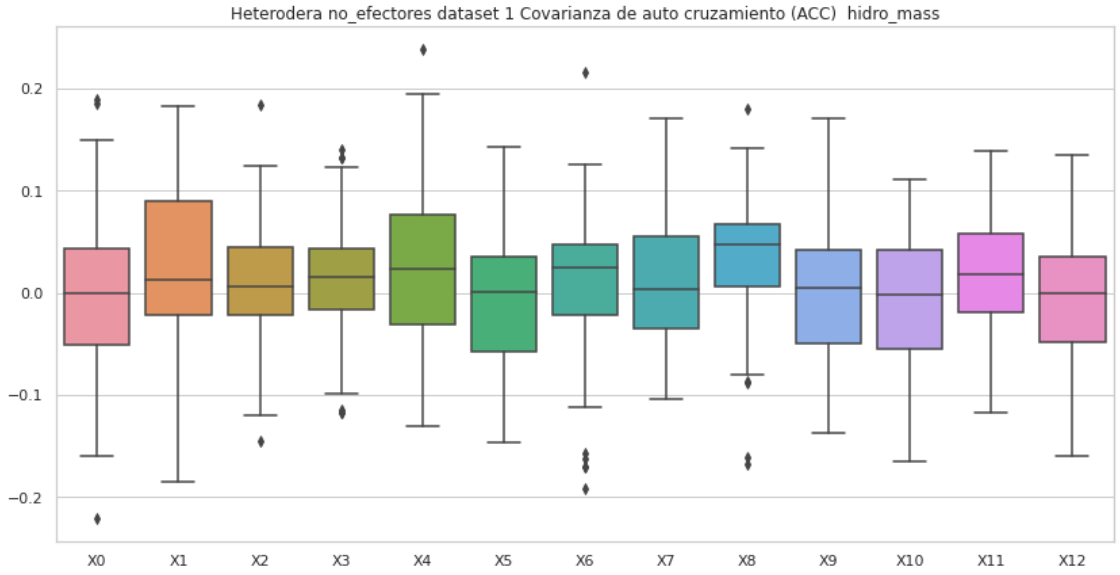
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores Heterodera dataset 1, sin valores atípicos.
Estadísticas.

```
              X0         X1         X2         X3         X4         X5  \
count  79.000000  79.000000  79.000000  79.000000  79.000000  79.000000
mean   -0.004008   0.031462   0.008979   0.010584   0.028243  -0.006751
std     0.080154   0.074012   0.059724   0.055143   0.080443   0.065212
min    -0.220949  -0.184909  -0.145183  -0.117650  -0.131271  -0.147052
25%    -0.051357  -0.021344  -0.022392  -0.016615  -0.030846  -0.057273
50%    -0.000939   0.012157   0.006314   0.015923   0.023114   0.000801
75%     0.043008   0.089735   0.044418   0.042741   0.076227   0.035392
max     0.189693   0.181984   0.184017   0.139800   0.238180   0.143211

              X6         X7         X8         X9        X10        X11  \
count  79.000000  79.000000  79.000000  79.000000  79.000000  79.000000
mean    0.010988   0.008286   0.032378  -0.000988  -0.010432   0.012189
std     0.070822   0.062533   0.060591   0.063316   0.066335   0.062180
```

min	-0.191915	-0.103552	-0.167665	-0.137878	-0.165737	-0.117730
25%	-0.022466	-0.034988	0.005834	-0.049340	-0.054628	-0.019267
50%	0.024083	0.003728	0.047450	0.004625	-0.001401	0.017300
75%	0.046735	0.054878	0.067280	0.042402	0.041633	0.057277
max	0.215549	0.171175	0.179643	0.170430	0.110875	0.138249

	X12
count	79.000000
mean	-0.010779
std	0.063214
min	-0.159479
25%	-0.048561
50%	-0.001127
75%	0.035629
max	0.134893





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))
```

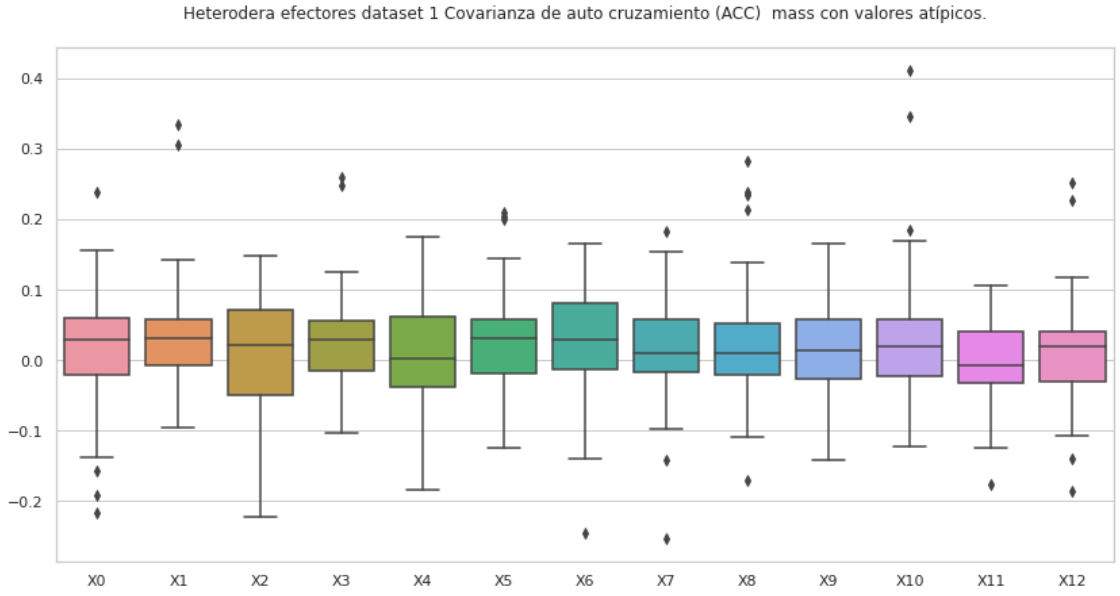
```
efectores
Covarianza de auto cruzamiento (ACC)  mass efectores Heterodera dataset 1, con
valores atípicos.
Valores del documento csv.

          X0        X1        X2        X3        X4        X5        X6  \
0  -0.038769  0.031259  0.002415  0.011131 -0.020154  0.030932  0.038987
1   0.077828  0.012054 -0.121769  0.056352 -0.023107  0.101445  0.000900
2  -0.137930 -0.007231  0.020661  0.056659 -0.034381 -0.009534  0.011462
3  -0.006994  0.011514  0.120517 -0.003738 -0.002641  0.036674 -0.073480
4   0.119770 -0.044673 -0.024220 -0.031648 -0.063063 -0.025467  0.105083
..       …         …         …         …         …         …         …
76  0.154887  0.142653 -0.035713  0.035094 -0.038459 -0.034522 -0.059078
77  0.052752  0.045389  0.076326  0.037815  0.114526  0.059162  0.095152
78  0.129281 -0.014146  0.040115  0.079577  0.038408  0.131465 -0.063595
79  0.064474  0.017109 -0.073589 -0.033216 -0.010125  0.043477  0.105925
80  0.028575  0.070338 -0.061186 -0.001816 -0.002058  0.015923  0.051552

          X7        X8        X9       X10       X11       X12         X13
0  -0.010492  0.053747  0.054128  0.002833 -0.022410  0.040367  efectores
1   0.057265  0.009527 -0.048216  0.047199 -0.113203 -0.021783  efectores
2  -0.006708 -0.019858  0.032800  0.058270  0.040313 -0.185185  efectores
3  -0.040932  0.139373  0.035370  0.019275  0.040369  0.096267  efectores
4   0.045049 -0.045470  0.020705  0.044133 -0.004578 -0.085961  efectores
..       …         …         …         …         …         …         …
76  0.146802  0.060316  0.148114  0.168644  0.039016  0.057945  efectores
77  0.076969  0.015624  0.010191  0.072782 -0.014074  0.018624  efectores
78 -0.008268 -0.020587  0.029253 -0.060016  0.013854  0.067851  efectores
79  0.002035  0.052976 -0.038353  0.056429 -0.051149  0.020793  efectores
80 -0.049432  0.027267 -0.058554 -0.093642  0.002049  0.027824  efectores

[81 rows x 14 columns]


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

              X0         X1         X2         X3         X4         X5  \
count  81.000000  81.000000  81.000000  81.000000  81.000000  81.000000
mean    0.021458   0.032031   0.009564   0.029125   0.011675   0.023817
```

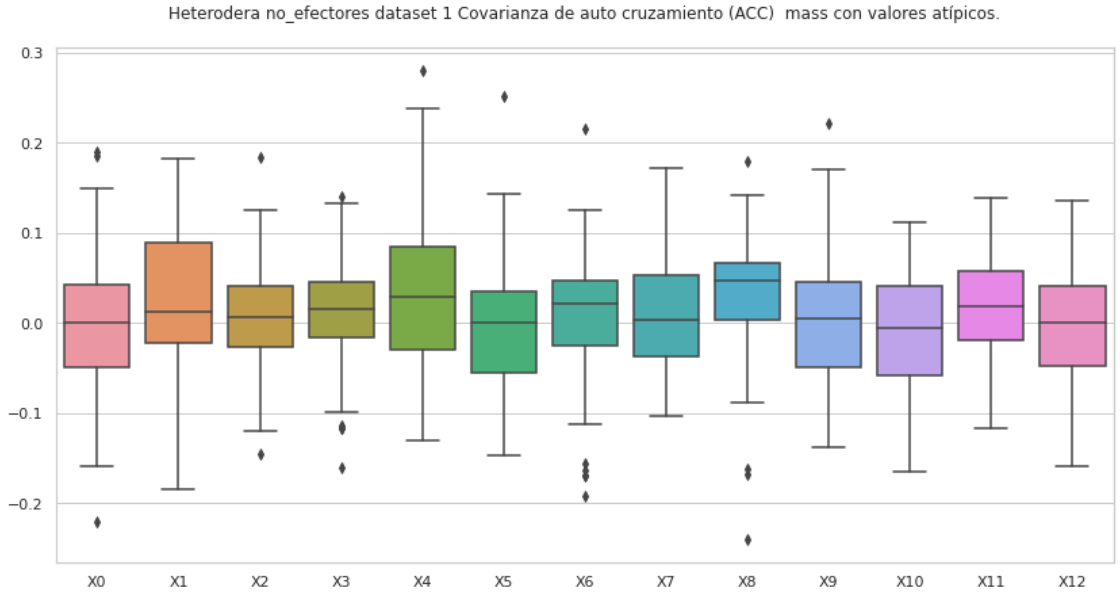
```
std     0.078131   0.068911   0.077599   0.061383   0.072703   0.071746
min    -0.216972  -0.096133  -0.222997  -0.104221  -0.184244  -0.124998
25%    -0.020616  -0.007729  -0.048777  -0.015219  -0.037567  -0.019418
50%     0.028575   0.031259   0.020661   0.029907   0.001549   0.030932
75%     0.060420   0.057391   0.071666   0.056352   0.062580   0.057397
max     0.239309   0.334429   0.148569   0.260304   0.175573   0.209816

              X6         X7         X8         X9        X10        X11  \
count  81.000000  81.000000  81.000000  81.000000  81.000000  81.000000
mean    0.022296   0.015529   0.020861   0.020626   0.023931   0.000656
std     0.073495   0.068944   0.072314   0.063286   0.089909   0.056084
min    -0.244908  -0.253155  -0.169734  -0.141630  -0.121615  -0.175983
25%    -0.013138  -0.017228  -0.019858  -0.026273  -0.021926  -0.032602
50%     0.028720   0.010465   0.010040   0.014681   0.020088  -0.006727
75%     0.081158   0.057265   0.052010   0.058106   0.058802   0.040313
max     0.164990   0.181859   0.283057   0.166436   0.411132   0.106875

             X12
count  81.000000
mean    0.009283
std     0.067297
min    -0.185185
25%    -0.029733
50%     0.018624
75%     0.040367
max     0.252684



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

          X0        X1        X2        X3        X4        X5        X6  \
0   0.016356 -0.005967  0.067647  0.029868 -0.027867 -0.066628  0.035680
1   0.087264  0.024981  0.124344  0.122770  0.031860  0.061186  0.215549
2  -0.044589  0.004445  0.040506 -0.033726 -0.025209 -0.000943 -0.014504
3   0.043150  0.004835  0.007311  0.016026  0.067616  0.083122 -0.170190
4  -0.032342  0.135388 -0.064402  0.027851  0.125018 -0.114852  0.088371
..       …         …         …         …         …         …         …
76  0.099114  0.053502  0.092183  0.047710  0.112952  0.057618  0.074242
77 -0.019287  0.039536 -0.041690  0.100384  0.031247  0.032895  0.038710
78 -0.100984  0.075671 -0.005136 -0.052009  0.157874 -0.104895  0.047333
79  0.072141 -0.008676  0.075559  0.045462 -0.093315  0.000801 -0.008817
80  0.123031 -0.043752  0.005875  0.051749 -0.093804 -0.016990  0.117190

          X7        X8        X9       X10       X11       X12           X13
```

```
0   0.101642  0.073049  0.003192  0.067882  0.075707  0.010843  no_efectores
1   0.106735  0.098941  0.010120  0.071415 -0.017568  0.042418  no_efectores
2   0.001867 -0.010332 -0.028280 -0.012661 -0.019196  0.008247  no_efectores
3   0.019508  0.067043  0.056666  0.042142 -0.108111 -0.119458  no_efectores
4  -0.092463  0.037825 -0.102604 -0.140277 -0.053252 -0.026306  no_efectores
..       …       …       …       …       …       …           …
76  0.081382  0.047450  0.026265 -0.037723  0.017300  0.052555  no_efectores
77  0.122265  0.016981  0.004828 -0.051503  0.027153  0.075163  no_efectores
78 -0.095514  0.099682 -0.055245 -0.118194  0.138249 -0.086957  no_efectores
79  0.023318 -0.020391  0.042288 -0.066117  0.036480  0.002699  no_efectores
80 -0.017556  0.009057  0.079370  0.009253 -0.102467  0.064053  no_efectores

[81 rows x 14 columns]
```

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

```
              X0         X1         X2         X3         X4         X5  \
count  81.000000  81.000000  81.000000  81.000000  81.000000  81.000000
mean   -0.003567   0.029406   0.006690   0.009861   0.032399  -0.004135
std     0.079208   0.074722   0.060888   0.059054   0.084697   0.070672
min    -0.220949  -0.184909  -0.145183  -0.160370  -0.131271  -0.147052
25%    -0.049103  -0.021636  -0.026312  -0.016637  -0.029140  -0.055896
50%    -0.000053   0.011896   0.005875   0.015923   0.029341   0.000801
75%     0.042866   0.088728   0.040530   0.045455   0.084655   0.035477
max     0.189693   0.181984   0.184017   0.139800   0.279843   0.250784

              X6         X7         X8         X9        X10        X11  \
count  81.000000  81.000000  81.000000  81.000000  81.000000  81.000000
mean    0.010225   0.007086   0.028191   0.002657  -0.011629   0.013621
std     0.070256   0.062814   0.067427   0.067657   0.065964   0.062432
min    -0.191915  -0.103552  -0.240303  -0.137878  -0.165737  -0.117730
25%    -0.025987  -0.037019   0.003223  -0.049266  -0.057752  -0.019196
50%     0.021109   0.003728   0.046992   0.004693  -0.005526   0.018842
75%     0.046526   0.052792   0.067043   0.046093   0.041124   0.057696
max     0.215549   0.171175   0.179643   0.220772   0.110875   0.138249

             X12
count  81.000000
mean   -0.008622
std     0.063997
min    -0.159479
25%    -0.048151
50%    -0.000443
75%     0.041096
max     0.134893
```


Heterodera efectores dataset 1 Covarianza de auto cruzamiento (ACC) mass con valores atípicos.
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12


Heterodera no_efectores dataset 1 Covarianza de auto cruzamiento (ACC) mass con valores atípicos.
0.3
0.2
0.1
0.0
−0.1
−0.2
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 Heterodera dataset 1, sin
valores atípicos.
Valores del documento csv.
```

	X0	X1	X2	X3	X4	X5	X6	\
0	-0.038769	0.031259	0.002415	0.011131	-0.020154	0.030932	0.038987	
1	0.077828	0.012054	-0.121769	0.056352	-0.023107	0.101445	0.000900	
2	-0.137930	-0.007231	0.020661	0.056659	-0.034381	-0.009534	0.011462	
3	-0.006994	0.011514	0.120517	-0.003738	-0.002641	0.036674	-0.073480	
4	0.119770	-0.044673	-0.024220	-0.031648	-0.063063	-0.025467	0.105083	
..	…	…	…	…	…	…	…	
76	0.154887	0.142653	-0.035713	0.035094	-0.038459	-0.034522	-0.059078	
77	0.052752	0.045389	0.076326	0.037815	0.114526	0.059162	0.095152	
78	0.129281	-0.014146	0.040115	0.079577	0.038408	0.131465	-0.063595	
79	0.064474	0.017109	-0.073589	-0.033216	-0.010125	0.043477	0.105925	
80	0.028575	0.070338	-0.061186	-0.001816	-0.002058	0.015923	0.051552	

	X7	X8	X9	X10	X11	X12	X13
0	-0.010492	0.053747	0.054128	0.002833	-0.022410	0.040367	efectores
1	0.057265	0.009527	-0.048216	0.047199	-0.113203	-0.021783	efectores
2	-0.006708	-0.019858	0.032800	0.058270	0.040313	-0.185185	efectores
3	-0.040932	0.139373	0.035370	0.019275	0.040369	0.096267	efectores
4	0.045049	-0.045470	0.020705	0.044133	-0.004578	-0.085961	efectores
..	…	…	…	…	…	…	…
76	0.146802	0.060316	0.148114	0.168644	0.039016	0.057945	efectores
77	0.076969	0.015624	0.010191	0.072782	-0.014074	0.018624	efectores
78	-0.008268	-0.020587	0.029253	-0.060016	0.013854	0.067851	efectores
79	0.002035	0.052976	-0.038353	0.056429	-0.051149	0.020793	efectores
80	-0.049432	0.027267	-0.058554	-0.093642	0.002049	0.027824	efectores

```
[75 rows x 14 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	75.000000	75.000000	75.000000	75.000000	75.000000	75.000000	
mean	0.025847	0.026595	0.011904	0.024140	0.012769	0.015707	
std	0.067170	0.052025	0.075139	0.050184	0.073701	0.064387	
min	-0.157326	-0.096133	-0.151052	-0.104221	-0.184244	-0.124998	
25%	-0.005321	-0.007480	-0.046922	-0.012032	-0.038013	-0.022443	
50%	0.037918	0.032832	0.020661	0.029907	-0.001823	0.029526	
75%	0.060803	0.057224	0.075500	0.053786	0.064067	0.050151	

```
max     0.155608   0.142653   0.148569   0.125409   0.175573   0.199367

              X6         X7         X8         X9        X10        X11  \
count  75.000000  75.000000  75.000000  75.000000  75.000000  75.000000
mean    0.025134   0.020712   0.012689   0.025678   0.019436   0.000383
std     0.065052   0.055256   0.057481   0.060544   0.069134   0.052186
min    -0.111139  -0.092492  -0.169734  -0.141630  -0.121615  -0.124822
25%    -0.012618  -0.013364  -0.018845  -0.017306  -0.017645  -0.030585
50%     0.028720   0.011256   0.009527   0.019438   0.022147  -0.007540
75%     0.076108   0.056628   0.049307   0.068474   0.058536   0.039040
max     0.164990   0.181859   0.213635   0.166436   0.185237   0.106875

             X12
count  75.000000
mean    0.001032
std     0.056615
min    -0.185185
25%    -0.033048
50%     0.017155
75%     0.033988
max     0.117940
```

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

```
          X0        X1        X2        X3        X4        X5        X6  \
0   0.016356 -0.005967  0.067647  0.029868 -0.027867 -0.066628  0.035680
1   0.087264  0.024981  0.124344  0.122770  0.031860  0.061186  0.215549
2  -0.044589  0.004445  0.040506 -0.033726 -0.025209 -0.000943 -0.014504
3   0.043150  0.004835  0.007311  0.016026  0.067616  0.083122 -0.170190
4  -0.032342  0.135388 -0.064402  0.027851  0.125018 -0.114852  0.088371
..       …         …         …         …         …         …         …
76  0.099114  0.053502  0.092183  0.047710  0.112952  0.057618  0.074242
77 -0.019287  0.039536 -0.041690  0.100384  0.031247  0.032895  0.038710
78 -0.100984  0.075671 -0.005136 -0.052009  0.157874 -0.104895  0.047333
79  0.072141 -0.008676  0.075559  0.045462 -0.093315  0.000801 -0.008817
80  0.123031 -0.043752  0.005875  0.051749 -0.093804 -0.016990  0.117190

          X7        X8        X9       X10       X11       X12           X13
0   0.101642  0.073049  0.003192  0.067882  0.075707  0.010843  no_efectores
1   0.106735  0.098941  0.010120  0.071415 -0.017568  0.042418  no_efectores
2   0.001867 -0.010332 -0.028280 -0.012661 -0.019196  0.008247  no_efectores
3   0.019508  0.067043  0.056666  0.042142 -0.108111 -0.119458  no_efectores
4  -0.092463  0.037825 -0.102604 -0.140277 -0.053252 -0.026306  no_efectores
..       …         …         …         …         …         …           …
```

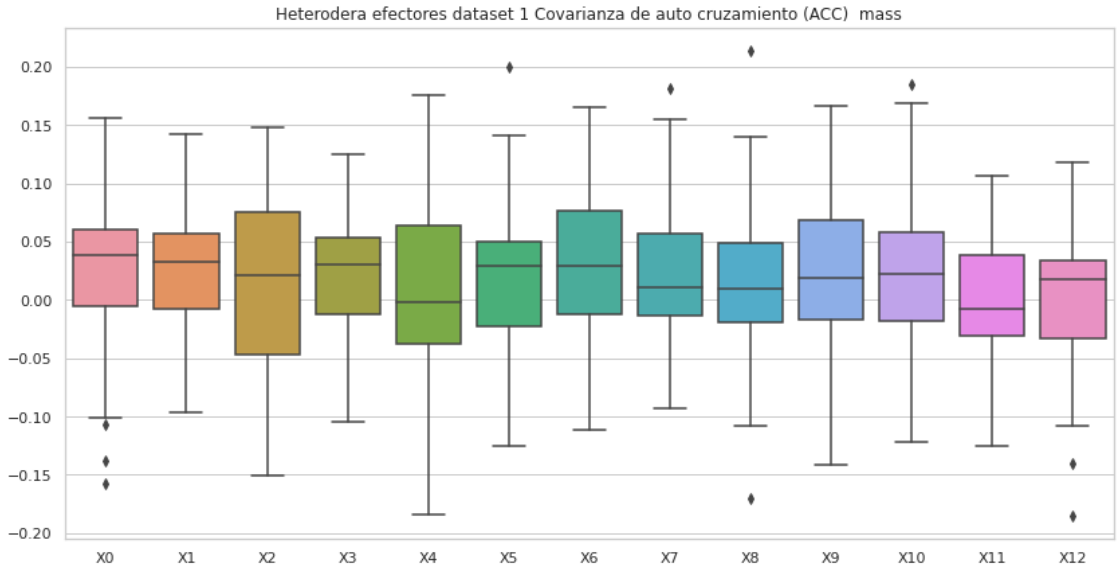
```
76  0.081382  0.047450  0.026265 -0.037723  0.017300  0.052555  no_efectores
77  0.122265  0.016981  0.004828 -0.051503  0.027153  0.075163  no_efectores
78 -0.095514  0.099682 -0.055245 -0.118194  0.138249 -0.086957  no_efectores
79  0.023318 -0.020391  0.042288 -0.066117  0.036480  0.002699  no_efectores
80 -0.017556  0.009057  0.079370  0.009253 -0.102467  0.064053  no_efectores

[79 rows x 14 columns]
```

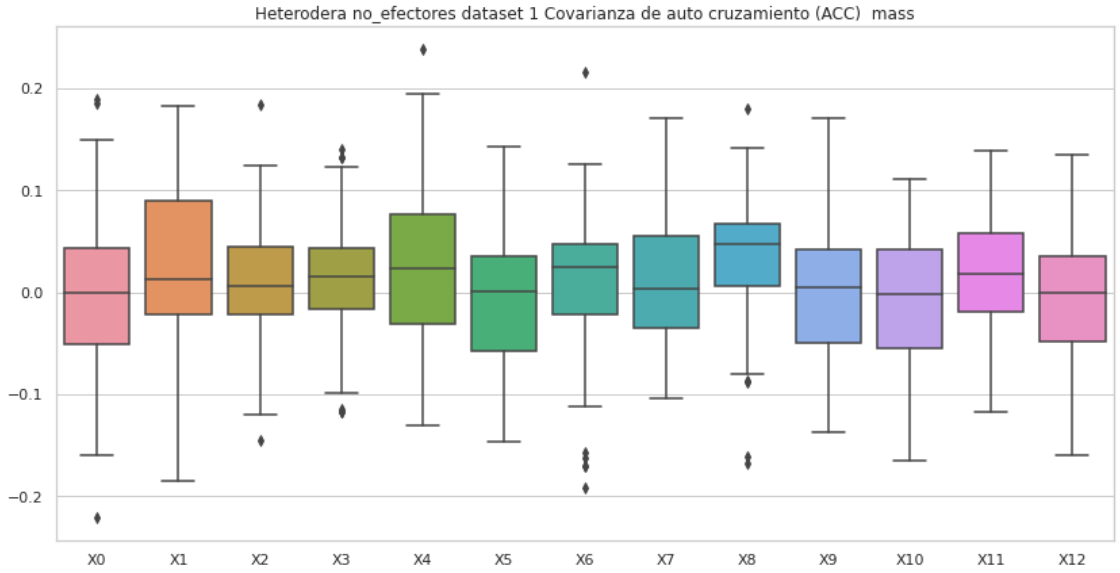
Covarianza de auto cruzamiento (ACC) mass no_efectores Heterodera dataset 1, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	79.000000	79.000000	79.000000	79.000000	79.000000	79.000000	
mean	-0.004008	0.031462	0.008979	0.010584	0.028243	-0.006751	
std	0.080154	0.074012	0.059724	0.055143	0.080443	0.065212	
min	-0.220949	-0.184909	-0.145183	-0.117650	-0.131271	-0.147052	
25%	-0.051357	-0.021344	-0.022392	-0.016615	-0.030846	-0.057273	
50%	-0.000939	0.012157	0.006314	0.015923	0.023114	0.000801	
75%	0.043008	0.089735	0.044418	0.042741	0.076227	0.035392	
max	0.189693	0.181984	0.184017	0.139800	0.238180	0.143211	

	X6	X7	X8	X9	X10	X11	\
count	79.000000	79.000000	79.000000	79.000000	79.000000	79.000000	
mean	0.010988	0.008286	0.032378	-0.000988	-0.010432	0.012189	
std	0.070822	0.062533	0.060591	0.063316	0.066335	0.062180	
min	-0.191915	-0.103552	-0.167665	-0.137878	-0.165737	-0.117730	
25%	-0.022466	-0.034988	0.005834	-0.049340	-0.054628	-0.019267	
50%	0.024083	0.003728	0.047450	0.004625	-0.001401	0.017300	
75%	0.046735	0.054878	0.067280	0.042402	0.041633	0.057277	
max	0.215549	0.171175	0.179643	0.170430	0.110875	0.138249	

	X12
count	79.000000
mean	-0.010779
std	0.063214
min	-0.159479
25%	-0.048561
50%	-0.001127
75%	0.035629
max	0.134893





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 Heterodera dataset 1, con
valores atípicos.
Valores del documento csv.

          X0        X1        X2        X3        X4        X5        X6  \
0   0.055143 -0.134738 -0.071098  0.121843  0.029562  0.014554 -0.055971
1  -0.066495 -0.072227 -0.084825  0.075908  0.057859 -0.182268  0.016332
2   0.023471 -0.226590  0.023738  0.138031 -0.129894 -0.195536  0.103991
3  -0.071158  0.060737  0.193862 -0.021740  0.041220  0.044213 -0.088489
4  -0.081184  0.085448 -0.071963 -0.004516  0.048894 -0.084502  0.001859
..       …         …         …         …         …         …         …
76  0.075705  0.003740  0.094708  0.178378  0.145760 -0.024424  0.102125
77  0.071354  0.195788  0.113852  0.084953  0.039603  0.113079  0.055913
78  0.066221  0.042181  0.033988  0.155107  0.109672  0.001978  0.110386
79  0.062521 -0.025851  0.075452  0.057947  0.065510  0.109329 -0.058690
80  0.042315 -0.106566  0.002724  0.017448  0.049227 -0.048924  0.042819

          X7        X8        X9       X10       X11       X12        X13
0   0.053867 -0.014916 -0.000657 -0.047876 -0.023351 -0.078027  efectores
1   0.039664  0.037957  0.089430 -0.093018  0.041407 -0.090391  efectores
2   0.185823 -0.059569  0.064699  0.271513  0.075486 -0.218760  efectores
```

```
3   0.097589  0.107394 -0.044710 -0.045041  0.067313  0.111252  efectores
4   0.007269  0.028226 -0.029112  0.014805  0.038857  0.062023  efectores
..       …       …       …       …       …       …       …
76  0.146282  0.023882 -0.023658  0.019774  0.016036  0.077603  efectores
77  0.128872  0.069856  0.107117  0.108638  0.104435 -0.028012  efectores
78 -0.099087 -0.001191 -0.031057  0.069074  0.015368 -0.115891  efectores
79 -0.049629 -0.024610  0.003134 -0.022958  0.039837 -0.044647  efectores
80  0.008796  0.004373 -0.025518 -0.041815  0.033521  0.093776  efectores

[81 rows x 14 columns]
```

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

```
              X0         X1         X2         X3         X4         X5  \
count  81.000000  81.000000  81.000000  81.000000  81.000000  81.000000
mean    0.043890   0.029808   0.046450   0.060003   0.035369   0.041017
std     0.088581   0.116288   0.091827   0.077345   0.085305   0.082418
min    -0.159007  -0.311203  -0.275033  -0.182478  -0.174814  -0.195536
25%    -0.023825  -0.025946  -0.008406   0.015979  -0.010418  -0.011126
50%     0.055143   0.037996   0.053994   0.056905   0.039117   0.044687
75%     0.081577   0.092642   0.112164   0.111162   0.066223   0.096495
max     0.392558   0.323245   0.358437   0.230075   0.323501   0.248145

              X6         X7         X8         X9        X10        X11  \
count  81.000000  81.000000  81.000000  81.000000  81.000000  81.000000
mean    0.042912   0.041648   0.031067   0.035448   0.049783   0.036645
std     0.075726   0.083715   0.063023   0.071590   0.078080   0.068721
min    -0.155636  -0.120701  -0.124676  -0.151967  -0.105052  -0.161582
25%    -0.009336  -0.018955  -0.015064  -0.023658   0.010799  -0.002124
50%     0.048435   0.039664   0.032243   0.028480   0.040779   0.039837
75%     0.102125   0.097589   0.070650   0.090046   0.081962   0.086218
max     0.250655   0.341001   0.187669   0.200368   0.271513   0.220542

             X12
count  81.000000
mean    0.002152
std     0.078526
min    -0.218760
25%    -0.044647
50%     0.001145
75%     0.055734
max     0.178424
```

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.226658	0.084194	0.159737	0.254054	0.197648	0.179695	0.168897	
1	0.103072	0.101293	-0.028286	0.119011	0.095672	0.044875	0.142234	
2	-0.062914	-0.067220	0.051837	0.012794	-0.058074	-0.040781	0.022180	
3	-0.152720	-0.000601	0.055350	-0.011331	-0.043694	-0.007263	0.020588	
4	0.066283	0.048306	-0.039091	-0.039650	0.005094	-0.032097	0.030204	
..	…	…	…	…	…	…	…	
76	-0.070730	-0.012107	0.037525	0.013477	-0.018011	-0.004021	-0.009671	
77	0.143190	0.192399	0.105409	0.123443	0.136470	0.047700	0.078469	
78	0.056198	0.101842	-0.025396	0.007001	0.043202	0.030709	0.048461	
79	0.082728	-0.017515	-0.117860	0.032443	-0.031661	-0.087266	-0.019403	
80	0.018831	-0.104662	-0.035841	0.098069	-0.095252	-0.010006	0.200566	

	X7	X8	X9	X10	X11	X12	X13
0	0.252916	0.194454	0.183461	0.199373	0.173710	0.084469	no_efectores
1	0.081903	0.034741	0.014519	-0.006677	0.109224	0.061454	no_efectores
2	-0.004209	0.022613	-0.035968	-0.001451	-0.012600	0.029774	no_efectores
3	0.054088	-0.128845	0.074619	0.022814	-0.062791	-0.012528	no_efectores
4	0.007612	0.028350	0.024790	-0.084218	-0.032379	-0.007732	no_efectores
..	…	…	…	…	…	…	…
76	0.021189	-0.058846	-0.030556	0.036514	-0.006113	0.024907	no_efectores
77	0.035614	0.017912	0.020998	0.004834	-0.020782	0.049741	no_efectores
78	0.009002	0.027210	0.039686	-0.016310	-0.001046	0.016948	no_efectores
79	-0.025955	-0.028583	-0.002819	0.038869	0.098875	-0.070721	no_efectores
80	-0.026115	-0.104649	0.047420	0.123574	-0.030162	0.015551	no_efectores

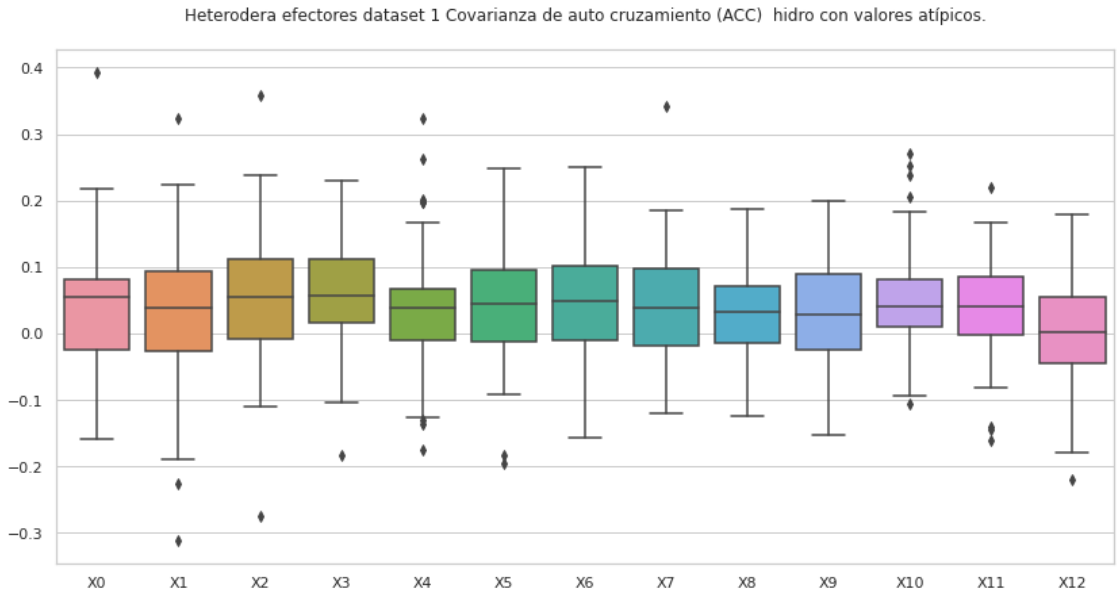
[81 rows x 14 columns]

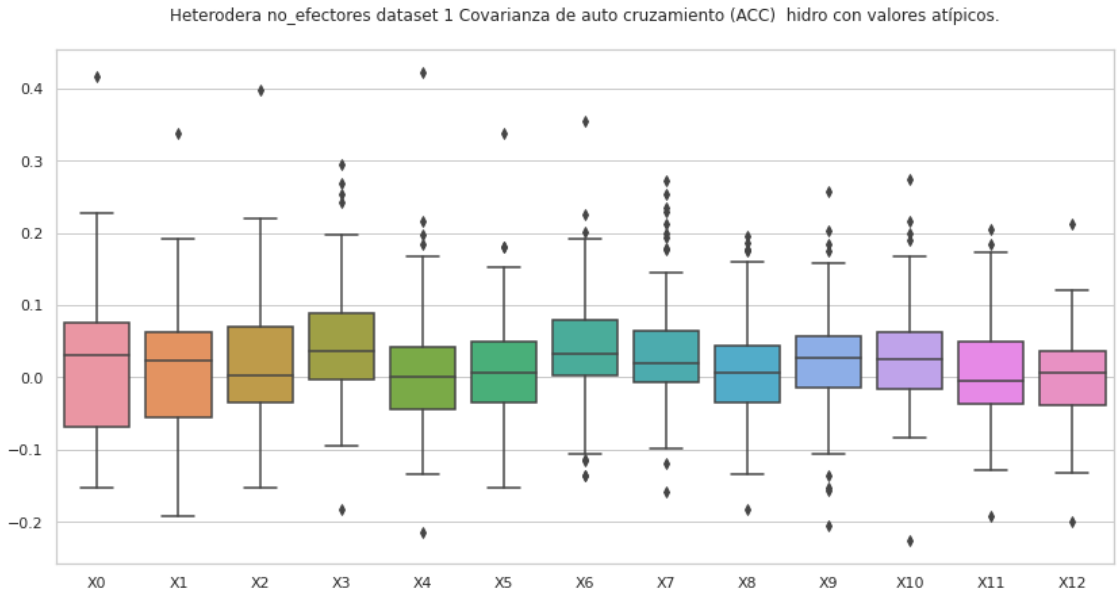
Covarianza de auto cruzamiento (ACC) hidro no_efectores Heterodera dataset 1,
con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	81.000000	81.000000	81.000000	81.000000	81.000000	81.000000	
mean	0.017174	0.007570	0.025859	0.047383	0.010331	0.011274	
std	0.105785	0.093922	0.089349	0.085472	0.087933	0.080068	
min	-0.152720	-0.192691	-0.153089	-0.182412	-0.214892	-0.152545	
25%	-0.069311	-0.054823	-0.035408	-0.002383	-0.043694	-0.034810	
50%	0.031110	0.022752	0.002489	0.035607	-0.000041	0.007027	
75%	0.075384	0.062225	0.069200	0.089278	0.042094	0.050088	
max	0.416583	0.337290	0.398210	0.294685	0.421750	0.338430	

	X6	X7	X8	X9	X10	X11	\
count	81.000000	81.000000	81.000000	81.000000	81.000000	81.000000	
mean	0.040679	0.038011	0.008788	0.020877	0.032054	0.011316	
std	0.083063	0.082442	0.082265	0.079533	0.076636	0.073053	
min	-0.136636	-0.157929	-0.181900	-0.205376	-0.225804	-0.192239	
25%	0.002147	-0.006251	-0.034284	-0.015143	-0.015240	-0.037361	
50%	0.032024	0.020075	0.007158	0.026417	0.024891	-0.005520	
75%	0.078469	0.063539	0.043548	0.056106	0.062882	0.049788	
max	0.355250	0.272776	0.194454	0.257945	0.274108	0.204236	

	X12
count	81.000000
mean	0.000392
std	0.067840
min	-0.199589
25%	-0.038973
50%	0.006634
75%	0.036937
max	0.211968





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 Heterodera dataset 1, sin
valores atípicos.
Valores del documento csv.

          X0        X1        X2        X3        X4        X5        X6  \
0   0.055143 -0.134738 -0.071098  0.121843  0.029562  0.014554 -0.055971
1  -0.066495 -0.072227 -0.084825  0.075908  0.057859 -0.182268  0.016332
2   0.023471 -0.226590  0.023738  0.138031 -0.129894 -0.195536  0.103991
3  -0.071158  0.060737  0.193862 -0.021740  0.041220  0.044213 -0.088489
4  -0.081184  0.085448 -0.071963 -0.004516  0.048894 -0.084502  0.001859
..       …         …         …         …         …         …         …
76  0.075705  0.003740  0.094708  0.178378  0.145760 -0.024424  0.102125
77  0.071354  0.195788  0.113852  0.084953  0.039603  0.113079  0.055913
78  0.066221  0.042181  0.033988  0.155107  0.109672  0.001978  0.110386
79  0.062521 -0.025851  0.075452  0.057947  0.065510  0.109329 -0.058690
80  0.042315 -0.106566  0.002724  0.017448  0.049227 -0.048924  0.042819

          X7        X8        X9       X10       X11       X12        X13
0   0.053867 -0.014916 -0.000657 -0.047876 -0.023351 -0.078027  efectores
1   0.039664  0.037957  0.089430 -0.093018  0.041407 -0.090391  efectores
2   0.185823 -0.059569  0.064699  0.271513  0.075486 -0.218760  efectores
3   0.097589  0.107394 -0.044710 -0.045041  0.067313  0.111252  efectores
4   0.007269  0.028226 -0.029112  0.014805  0.038857  0.062023  efectores
..       …         …         …         …         …         …         …
76  0.146282  0.023882 -0.023658  0.019774  0.016036  0.077603  efectores
```

```
77  0.128872  0.069856  0.107117  0.108638  0.104435 -0.028012  efectores
78 -0.099087 -0.001191 -0.031057  0.069074  0.015368 -0.115891  efectores
79 -0.049629 -0.024610  0.003134 -0.022958  0.039837 -0.044647  efectores
80  0.008796  0.004373 -0.025518 -0.041815  0.033521  0.093776  efectores

[78 rows x 14 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	78.000000	78.000000	78.000000	78.000000	78.000000	78.000000	
mean	0.035399	0.027908	0.045494	0.059946	0.031304	0.039297	
std	0.076559	0.112814	0.077858	0.070706	0.080365	0.080370	
min	-0.159007	-0.311203	-0.109746	-0.103702	-0.174814	-0.195536	
25%	-0.024732	-0.025941	-0.007634	0.016346	-0.011124	-0.012869	
50%	0.049218	0.040088	0.053884	0.056698	0.035305	0.046570	
75%	0.079869	0.091384	0.106200	0.109007	0.066045	0.094921	
max	0.218298	0.223532	0.238406	0.222476	0.263275	0.219520	

	X6	X7	X8	X9	X10	X11	\
count	78.000000	78.000000	78.000000	78.000000	78.000000	78.000000	
mean	0.040815	0.037678	0.030006	0.034305	0.049314	0.036406	
std	0.076287	0.075582	0.058634	0.069247	0.076615	0.068221	
min	-0.155636	-0.120701	-0.100562	-0.151967	-0.105052	-0.161582	
25%	-0.016068	-0.014736	-0.015027	-0.023437	0.011801	-0.000709	
50%	0.044536	0.039336	0.030235	0.027855	0.040174	0.039347	
75%	0.093484	0.091637	0.070187	0.087905	0.074992	0.085672	
max	0.250655	0.185823	0.187669	0.200368	0.271513	0.220542	

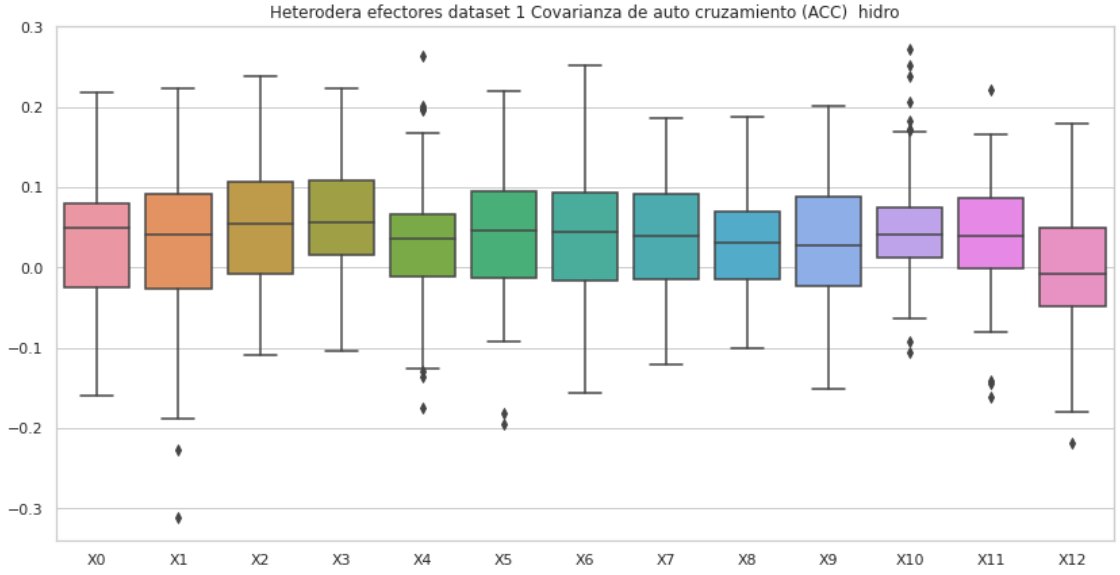
	X12
count	78.000000
mean	-0.002478
std	0.076247
min	-0.218760
25%	-0.048943
50%	-0.008471
75%	0.049373
max	0.178424

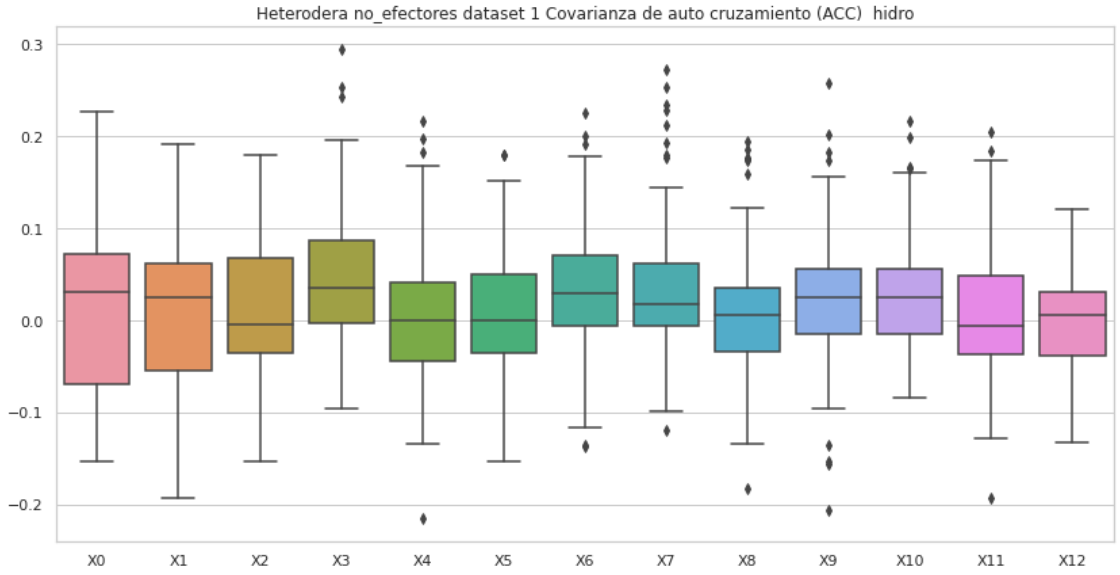
no_efectores
Covarianza de auto cruzamiento (ACC) no_efectores Heterodera dataset 1, sin valores atípicos.
Valores del documento csv.

```
          X0        X1        X2        X3        X4        X5        X6  \
0   0.226658  0.084194  0.159737  0.254054  0.197648  0.179695  0.168897
1   0.103072  0.101293 -0.028286  0.119011  0.095672  0.044875  0.142234
2  -0.062914 -0.067220  0.051837  0.012794 -0.058074 -0.040781  0.022180
3  -0.152720 -0.000601  0.055350 -0.011331 -0.043694 -0.007263  0.020588
4   0.066283  0.048306 -0.039091 -0.039650  0.005094 -0.032097  0.030204
..       …         …         …         …         …         …         …
76 -0.070730 -0.012107  0.037525  0.013477 -0.018011 -0.004021 -0.009671
77  0.143190  0.192399  0.105409  0.123443  0.136470  0.047700  0.078469
78  0.056198  0.101842 -0.025396  0.007001  0.043202  0.030709  0.048461
79  0.082728 -0.017515 -0.117860  0.032443 -0.031661 -0.087266 -0.019403
80  0.018831 -0.104662 -0.035841  0.098069 -0.095252 -0.010006  0.200566

          X7        X8        X9       X10       X11       X12           X13
0   0.252916  0.194454  0.183461  0.199373  0.173710  0.084469  no_efectores
1   0.081903  0.034741  0.014519 -0.006677  0.109224  0.061454  no_efectores
2  -0.004209  0.022613 -0.035968 -0.001451 -0.012600  0.029774  no_efectores
3   0.054088 -0.128845  0.074619  0.022814 -0.062791 -0.012528  no_efectores
4   0.007612  0.028350  0.024790 -0.084218 -0.032379 -0.007732  no_efectores
..       …         …         …         …         …         …           …
76  0.021189 -0.058846 -0.030556  0.036514 -0.006113  0.024907  no_efectores
77  0.035614  0.017912  0.020998  0.004834 -0.020782  0.049741  no_efectores
78  0.009002  0.027210  0.039686 -0.016310 -0.001046  0.016948  no_efectores
79 -0.025955 -0.028583 -0.002819  0.038869  0.098875 -0.070721  no_efectores
80 -0.026115 -0.104649  0.047420  0.123574 -0.030162  0.015551  no_efectores

[77 rows x 14 columns]


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

              X0         X1         X2         X3         X4         X5  \
count  77.000000  77.000000  77.000000  77.000000  77.000000  77.000000
mean    0.012007   0.006734   0.018758   0.046089   0.006342   0.008286
std     0.095381   0.085607   0.077604   0.077749   0.075584   0.071232
min    -0.152720  -0.192691  -0.153089  -0.095142  -0.214892  -0.152545
25%    -0.069311  -0.054823  -0.035608  -0.002383  -0.043694  -0.034810
50%     0.031110   0.025134  -0.004698   0.035607  -0.000041   0.000626
75%     0.072323   0.062225   0.067807   0.087283   0.042094   0.050088
max     0.227613   0.192399   0.179288   0.294685   0.216472   0.179924

              X6         X7         X8         X9        X10        X11  \
count  77.000000  77.000000  77.000000  77.000000  77.000000  77.000000
mean    0.033759   0.036116   0.006939   0.020015   0.030060   0.010048
std     0.074901   0.077988   0.082234   0.078580   0.064833   0.071766
```

min	-0.136636	-0.119388	-0.181900	-0.205376	-0.084218	-0.192239
25%	-0.005244	-0.006251	-0.034284	-0.015143	-0.015240	-0.037361
50%	0.030020	0.017170	0.006466	0.024790	0.024715	-0.006113
75%	0.071557	0.061840	0.035789	0.055685	0.055894	0.049342
max	0.225841	0.272776	0.194454	0.257945	0.216601	0.204236

	X12
count	77.000000
mean	0.000483
std	0.058400
min	-0.132973
25%	-0.038339
50%	0.006634
75%	0.031089
max	0.120429




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