

# ds1\_Globodera\_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 = "Globodera"
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)+"\n
↪ "+str(transf)+" " +str(estado))
```

efectores

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

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9	\
0	6.494	1.948	7.143	6.494	7.143	2.597	2.597	7.792	0.649	5.195	
1	8.097	3.239	7.692	4.453	4.049	5.263	3.644	12.146	1.215	7.692	
2	3.750	1.250	0.000	22.500	0.000	16.250	5.000	0.000	0.000	3.750	
3	8.097	3.644	7.692	4.453	4.049	5.263	3.644	12.146	1.215	7.692	
4	2.069	1.034	2.759	1.724	0.000	1.724	1.034	5.862	2.069	5.862	
..	...	...	...	...	...	...	...	...	...	...	
119	7.595	4.219	7.173	3.797	3.376	3.797	4.219	13.924	0.422	8.017	
120	7.287	3.644	8.502	3.644	4.453	5.263	4.049	13.360	0.000	6.883	
121	7.287	3.644	8.502	3.644	4.453	5.263	4.049	13.360	0.000	6.883	
122	8.502	3.239	7.692	4.453	4.049	5.263	3.644	12.146	1.215	7.692	
123	8.097	3.239	7.692	4.453	4.049	4.858	3.644	12.551	1.215	7.692	

	...	X11	X12	X13	X14	X15	X16	X17	X18	X19	\
0	...	9.091	1.948	3.247	6.494	7.792	7.143	0.649	2.597	4.545	
1	...	7.692	0.810	4.858	1.619	6.073	6.073	0.810	1.619	8.097	
2	...	10.000	0.000	0.000	15.000	5.000	2.500	5.000	0.000	6.250	
3	...	7.287	0.810	5.263	1.619	5.263	6.073	0.810	1.619	8.097	
4	...	3.793	2.069	22.414	2.069	11.379	0.690	2.414	4.483	10.000	
..	...	...	...	...	...	...	...	...	...	...	
119	...	8.861	0.000	4.641	2.110	7.173	6.329	0.844	1.688	6.329	
120	...	7.287	0.000	5.263	1.215	4.453	6.883	0.810	2.429	8.097	
121	...	7.287	0.000	5.263	1.215	4.049	7.287	0.810	2.429	8.097	
122	...	7.692	0.810	4.858	1.619	5.263	5.668	0.810	2.024	8.097	
123	...	7.692	0.810	5.263	1.619	5.263	6.073	0.810	1.619	8.097	

	X20
0	efectores
1	efectores
2	efectores
3	efectores
4	efectores
..	...
119	efectores
120	efectores
121	efectores
122	efectores
123	efectores

[124 rows x 21 columns]

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

	X0	X1	X2	X3	X4	X5	\
count	124.000000	124.000000	124.000000	124.000000	124.000000	124.000000	
mean	6.559984	2.919540	6.897274	4.341621	4.110097	3.852718	
std	2.101041	1.779504	2.266423	2.297426	2.508419	2.081722	
min	2.062000	0.851000	0.000000	0.000000	0.000000	0.000000	
25%	6.494000	1.948000	7.012500	3.644000	2.807500	2.597000	
50%	7.105000	3.239000	7.692000	4.175500	4.049000	3.797000	
75%	7.595000	3.644000	7.850250	5.844000	7.051000	5.263000	
max	12.911000	17.241000	13.223000	22.500000	10.959000	16.250000	

	X6	X7	X8	X9	X10	X11	\
count	124.000000	124.000000	124.000000	124.000000	124.000000	124.000000	
mean	3.180371	9.564589	0.968831	6.210081	8.560145	7.463734	
std	1.318249	3.900435	0.811943	1.491626	3.883355	1.893831	

min	0.000000	0.000000	0.000000	1.370000	3.750000	0.000000
25%	2.597000	6.484250	0.422000	5.195000	5.485000	7.183000
50%	3.488500	8.387500	0.649000	5.921000	7.860000	7.692000
75%	4.049000	13.360000	1.384500	7.287000	8.442000	9.091000
max	7.792000	25.490000	3.846000	10.400000	18.213000	10.256000

	X12	X13	X14	X15	X16	X17 \
count	124.000000	124.000000	124.000000	124.000000	124.000000	124.000000
mean	1.400081	6.959669	3.677387	7.102669	5.783935	1.112250
std	1.108688	6.261115	2.558102	2.269398	2.391423	0.810707
min	0.000000	0.000000	0.810000	2.597000	0.690000	0.000000
25%	0.000000	3.247000	1.619000	5.263000	5.882000	0.649000
50%	1.948000	5.046500	2.110000	7.173000	6.444000	0.810000
75%	2.063750	5.263000	6.494000	7.792000	7.143000	0.852250
max	4.190000	22.414000	15.000000	12.340000	15.068000	5.000000

	X18	X19
count	124.000000	124.000000
mean	2.628677	6.705960
std	1.277402	1.958951
min	0.000000	3.247000
25%	1.688000	4.567500
50%	2.429000	6.429000
75%	2.597000	8.097000
max	8.000000	10.690000

#### no\_efectores

Composición de aminoácidos (AAC) no\_efectores Globodera dataset 1, con valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9 \
0	3.883	1.942	3.883	2.913	1.942	6.796	0.971	11.650	2.913	6.796
1	6.557	2.459	4.918	4.508	3.279	9.016	1.639	8.197	2.459	4.918
2	7.197	3.030	4.545	4.167	2.652	8.712	1.515	7.955	2.273	5.303
3	7.170	3.019	4.528	4.528	2.264	7.925	1.132	7.925	2.264	5.283
4	6.415	3.396	4.906	4.151	3.019	7.925	1.132	8.679	1.509	5.283
..	...	...	...	...	...	...	...	...	...	...
119	1.923	0.385	2.308	1.923	1.154	1.923	1.154	7.692	2.692	5.000
120	13.811	1.535	6.394	3.325	1.790	2.302	3.069	6.650	2.813	4.604
121	7.725	1.288	10.300	6.009	4.292	1.288	2.575	12.017	0.429	6.438
122	11.921	5.298	5.960	2.649	0.662	4.636	3.974	8.609	3.311	2.649
123	7.170	2.642	4.528	4.528	2.264	7.925	1.509	7.925	2.264	5.283
...	X11	X12	X13	X14	X15	X16	X17	X18	X19 \	
0	...	9.709	0.971	4.854	3.883	8.738	3.883	1.942	6.796	8.738

1	...	7.377	1.639	4.098	5.738	6.967	6.148	1.639	3.689	6.967
2	...	6.439	1.894	5.303	6.061	6.818	5.682	1.515	3.409	6.818
3	...	7.547	1.132	5.283	5.660	7.547	5.660	1.887	3.396	7.547
4	...	6.792	1.509	4.906	5.283	7.170	6.415	1.509	3.774	7.170
..	...	...	...	...	...	...	...	...	...	...
119	...	1.538	1.923	21.154	1.923	13.462	1.923	2.692	3.846	8.077
120	...	8.440	1.023	3.581	6.394	8.184	5.627	2.558	3.581	6.905
121	...	6.009	1.288	3.433	6.009	9.442	6.438	1.288	3.004	6.009
122	...	7.947	3.974	3.974	7.285	7.285	2.649	0.662	1.987	5.960
123	...	7.170	1.887	4.906	6.415	7.170	5.283	1.887	3.396	6.792

```

                X20
0      no_efectores
1      no_efectores
2      no_efectores
3      no_efectores
4      no_efectores
..
119    no_efectores
120    no_efectores
121    no_efectores
122    no_efectores
123    no_efectores

```

[124 rows x 21 columns]

Composición de aminoácidos (AAC) no\_efectores Globodera dataset 1, con valores atípicos.

Estadísticas.

	X0	X1	X2	X3	X4	X5 \
count	124.000000	124.000000	124.000000	124.000000	124.000000	124.000000
mean	7.385750	2.877815	5.534460	4.319347	2.261371	5.186685
std	2.621619	1.721118	2.197215	1.699767	1.573211	3.261205
min	0.826000	0.000000	1.347000	1.333000	0.000000	0.000000
25%	5.876000	1.282000	4.514000	3.176500	0.789500	1.688000
50%	7.170000	2.642000	5.284000	4.321000	2.264000	5.505000
75%	9.119500	4.147000	6.426750	5.213250	3.252000	7.932500
max	13.990000	7.432000	10.970000	9.091000	7.000000	13.214000

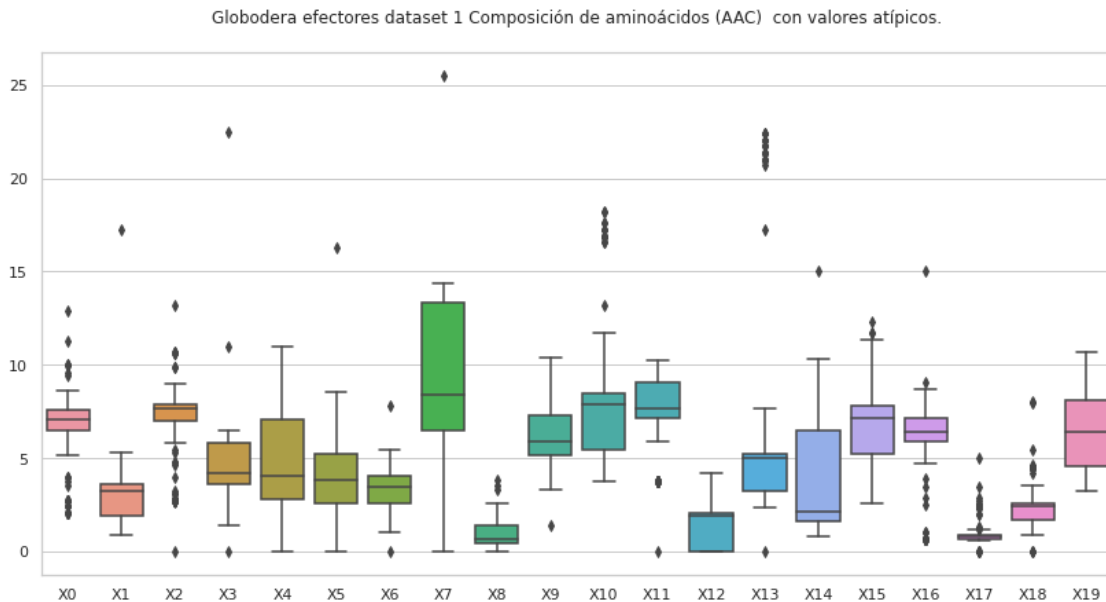
  

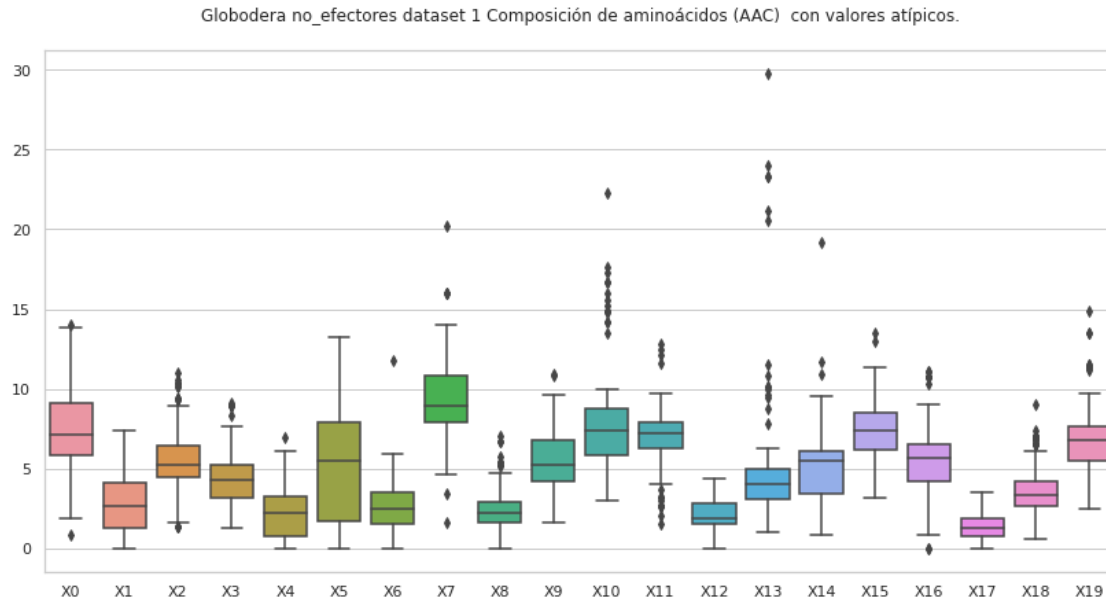
	X6	X7	X8	X9	X10	X11 \
count	124.000000	124.000000	124.000000	124.000000	124.000000	124.000000
mean	2.654879	9.319129	2.336895	5.409298	8.033089	6.940435
std	1.654124	2.420224	1.375713	2.005484	3.283738	1.931791
min	0.000000	1.667000	0.000000	1.667000	3.000000	1.538000
25%	1.507750	7.925000	1.614250	4.174500	5.859000	6.309250
50%	2.500000	8.945000	2.264000	5.289500	7.412000	7.238500

75%	3.526000	10.832250	2.913000	6.796000	8.780250	7.939500
max	11.785000	20.202000	7.031000	10.897000	22.314000	12.857000

	X12	X13	X14	X15	X16	X17 \
count	124.000000	124.000000	124.000000	124.000000	124.000000	124.000000
mean	2.173556	5.276589	5.306677	7.525702	5.548581	1.395694
std	0.988863	4.615922	2.533225	1.689506	2.357858	0.768836
min	0.000000	1.010000	0.826000	3.175000	0.000000	0.000000
25%	1.509000	3.125000	3.448000	6.230750	4.180000	0.781000
50%	1.887000	4.020000	5.469000	7.432000	5.660000	1.266000
75%	2.857000	5.003500	6.132500	8.477750	6.564500	1.887000
max	4.348000	29.752000	19.192000	13.462000	11.111000	3.529000

	X18	X19
count	124.000000	124.000000
mean	3.577306	6.936710
std	1.733181	2.042954
min	0.602000	2.479000
25%	2.675500	5.462250
50%	3.396000	6.797500
75%	4.196000	7.692000
max	9.000000	14.865000





## 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)+"\n
↪"+str(transf))

```

efectores

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

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9	\
0	6.494	1.948	7.143	6.494	7.143	2.597	2.597	7.792	0.649	5.195	
1	8.097	3.239	7.692	4.453	4.049	5.263	3.644	12.146	1.215	7.692	
3	8.097	3.644	7.692	4.453	4.049	5.263	3.644	12.146	1.215	7.692	
4	2.069	1.034	2.759	1.724	0.000	1.724	1.034	5.862	2.069	5.862	
5	6.883	3.644	8.097	3.644	4.453	5.263	4.453	12.955	0.405	6.883	
..	...	...	...	...	...	...	...	...	...	...	
119	7.595	4.219	7.173	3.797	3.376	3.797	4.219	13.924	0.422	8.017	
120	7.287	3.644	8.502	3.644	4.453	5.263	4.049	13.360	0.000	6.883	
121	7.287	3.644	8.502	3.644	4.453	5.263	4.049	13.360	0.000	6.883	
122	8.502	3.239	7.692	4.453	4.049	5.263	3.644	12.146	1.215	7.692	
123	8.097	3.239	7.692	4.453	4.049	4.858	3.644	12.551	1.215	7.692	
...	...	...	...	...	...	...	...	...	...	...	
0	...	9.091	1.948	3.247	6.494	7.792	7.143	0.649	2.597	4.545	\
1	...	7.692	0.810	4.858	1.619	6.073	6.073	0.810	1.619	8.097	
3	...	7.287	0.810	5.263	1.619	5.263	6.073	0.810	1.619	8.097	
4	...	3.793	2.069	22.414	2.069	11.379	0.690	2.414	4.483	10.000	
5	...	6.883	0.000	5.263	1.215	5.263	6.883	0.810	2.429	8.097	
..	...	...	...	...	...	...	...	...	...	...	
119	...	8.861	0.000	4.641	2.110	7.173	6.329	0.844	1.688	6.329	

120	...	7.287	0.000	5.263	1.215	4.453	6.883	0.810	2.429	8.097
121	...	7.287	0.000	5.263	1.215	4.049	7.287	0.810	2.429	8.097
122	...	7.692	0.810	4.858	1.619	5.263	5.668	0.810	2.024	8.097
123	...	7.692	0.810	5.263	1.619	5.263	6.073	0.810	1.619	8.097

```

      X20
0    efectores
1    efectores
3    efectores
4    efectores
5    efectores
..    ""
119  efectores
120  efectores
121  efectores
122  efectores
123  efectores

```

[114 rows x 21 columns]

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

Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	
mean	6.587772	2.817246	7.062833	4.286912	4.122754	3.785526	
std	2.018559	1.215937	2.151151	1.535492	2.449557	1.657648	
min	2.062000	0.851000	2.759000	1.375000	0.000000	1.379000	
25%	6.494000	1.948000	7.143000	3.644000	2.893500	2.597000	
50%	7.143000	3.239000	7.692000	4.219000	4.049000	3.797000	
75%	7.595000	3.644000	8.040750	5.844000	7.120000	5.263000	
max	11.258000	5.294000	13.223000	10.952000	7.792000	8.571000	

	X6	X7	X8	X9	X10	X11	\
count	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	
mean	3.054649	9.650263	0.903877	6.231325	8.643044	7.473079	
std	1.232868	3.367446	0.702676	1.278501	3.939829	1.798101	
min	0.000000	3.306000	0.000000	3.306000	4.706000	3.780000	
25%	2.597000	6.699750	0.422000	5.195000	6.018250	7.287000	
50%	3.244000	8.442000	0.649000	6.073000	7.975000	7.692000	
75%	4.049000	13.360000	1.215000	7.287000	8.442000	9.091000	
max	5.300000	14.346000	2.549000	8.571000	18.213000	10.256000	

	X12	X13	X14	X15	X16	X17	\
count	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	
mean	1.312658	7.113193	3.547465	7.165088	5.688333	1.091825	

std	1.047350	6.365623	2.327032	2.263282	2.258676	0.684889
min	0.000000	2.649000	0.810000	4.049000	0.690000	0.000000
25%	0.000000	3.247000	1.619000	5.263000	5.905500	0.649000
50%	1.935500	5.095500	2.110000	7.173000	6.410000	0.810000
75%	1.957750	5.263000	6.494000	7.792000	7.143000	0.844000
max	4.190000	22.414000	7.843000	12.340000	8.723000	2.857000

	X18	X19
count	114.000000	114.000000
mean	2.612789	6.848921
std	0.970090	1.960241
min	1.000000	3.247000
25%	1.695500	4.552500
50%	2.429000	6.751000
75%	2.597000	8.097000
max	4.636000	10.690000

no\_efectores

Composición de aminoácidos (AAC) no\_efectores Globodera dataset 1, sin valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9	\
0	3.883	1.942	3.883	2.913	1.942	6.796	0.971	11.650	2.913	6.796	
1	6.557	2.459	4.918	4.508	3.279	9.016	1.639	8.197	2.459	4.918	
2	7.197	3.030	4.545	4.167	2.652	8.712	1.515	7.955	2.273	5.303	
3	7.170	3.019	4.528	4.528	2.264	7.925	1.132	7.925	2.264	5.283	
4	6.415	3.396	4.906	4.151	3.019	7.925	1.132	8.679	1.509	5.283	
..	...	...	...	...	...	...	...	...	...	...	
118	3.883	1.942	4.854	2.913	1.942	7.767	1.942	11.650	2.913	6.796	
120	13.811	1.535	6.394	3.325	1.790	2.302	3.069	6.650	2.813	4.604	
121	7.725	1.288	10.300	6.009	4.292	1.288	2.575	12.017	0.429	6.438	
122	11.921	5.298	5.960	2.649	0.662	4.636	3.974	8.609	3.311	2.649	
123	7.170	2.642	4.528	4.528	2.264	7.925	1.509	7.925	2.264	5.283	

	X11	X12	X13	X14	X15	X16	X17	X18	X19	\
0	9.709	0.971	4.854	3.883	8.738	3.883	1.942	6.796	8.738	
1	7.377	1.639	4.098	5.738	6.967	6.148	1.639	3.689	6.967	
2	6.439	1.894	5.303	6.061	6.818	5.682	1.515	3.409	6.818	
3	7.547	1.132	5.283	5.660	7.547	5.660	1.887	3.396	7.547	
4	6.792	1.509	4.906	5.283	7.170	6.415	1.509	3.774	7.170	
..	...	...	...	...	...	...	...	...	...	
118	6.796	0.971	5.825	3.883	8.738	3.883	0.971	6.796	8.738	
120	8.440	1.023	3.581	6.394	8.184	5.627	2.558	3.581	6.905	
121	6.009	1.288	3.433	6.009	9.442	6.438	1.288	3.004	6.009	
122	7.947	3.974	3.974	7.285	7.285	2.649	0.662	1.987	5.960	

123 ... 7.170 1.887 4.906 6.415 7.170 5.283 1.887 3.396 6.792

```

                X20
0   no_efectores
1   no_efectores
2   no_efectores
3   no_efectores
4   no_efectores
..
118 no_efectores
120 no_efectores
121 no_efectores
122 no_efectores
123 no_efectores

```

[107 rows x 21 columns]

Composición de aminoácidos (AAC) no\_efectores Globodera dataset 1, sin valores atípicos.

Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	107.000000	107.000000	107.000000	107.000000	107.000000	107.000000	
mean	7.619112	3.001748	5.877009	4.473318	2.424991	5.394636	
std	2.391028	1.646898	1.968365	1.539877	1.499963	3.016064	
min	3.846000	0.676000	1.653000	1.533000	0.000000	0.000000	
25%	6.081000	1.534000	4.545000	3.413000	1.071000	1.923000	
50%	7.170000	2.652000	5.442000	4.472000	2.564000	5.759000	
75%	9.163000	4.162000	6.495000	5.469000	3.265500	7.940000	
max	13.990000	7.432000	10.970000	9.091000	6.087000	13.214000	

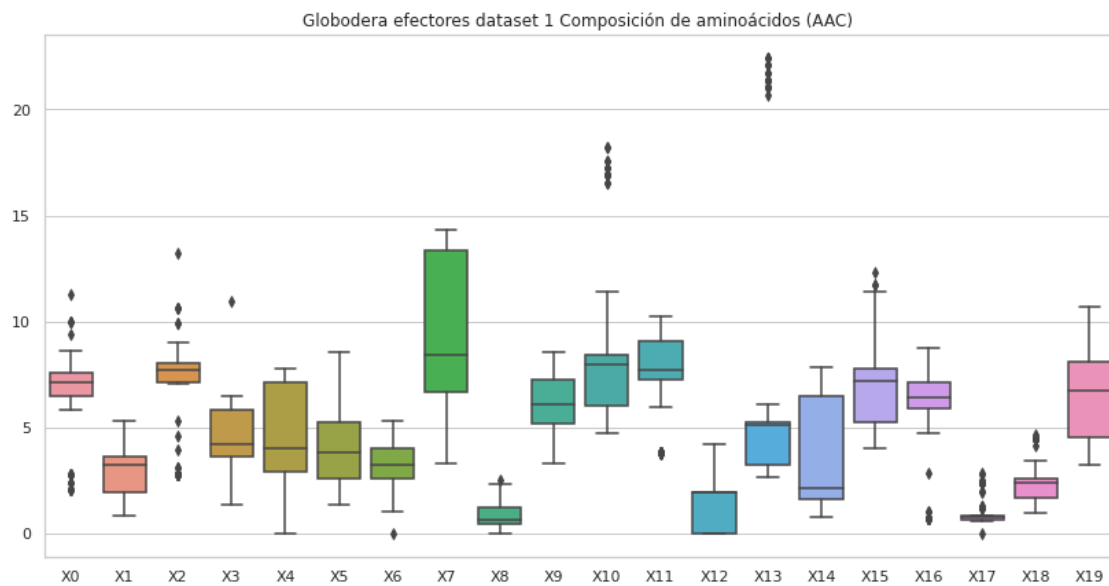
	X6	X7	X8	X9	X10	X11	\
count	107.000000	107.000000	107.000000	107.000000	107.000000	107.000000	
mean	2.666636	9.419860	2.203065	5.536953	7.479280	7.222355	
std	1.406244	2.076386	1.216756	1.967478	2.220564	1.574169	
min	0.800000	3.478000	0.000000	1.852000	4.348000	2.027000	
25%	1.509000	7.955000	1.544000	4.536500	5.803000	6.427000	
50%	2.532000	9.091000	2.264000	5.303000	7.229000	7.317000	
75%	3.497000	11.053000	2.776500	6.833500	8.679000	7.973500	
max	5.944000	16.026000	5.759000	10.897000	15.541000	12.500000	

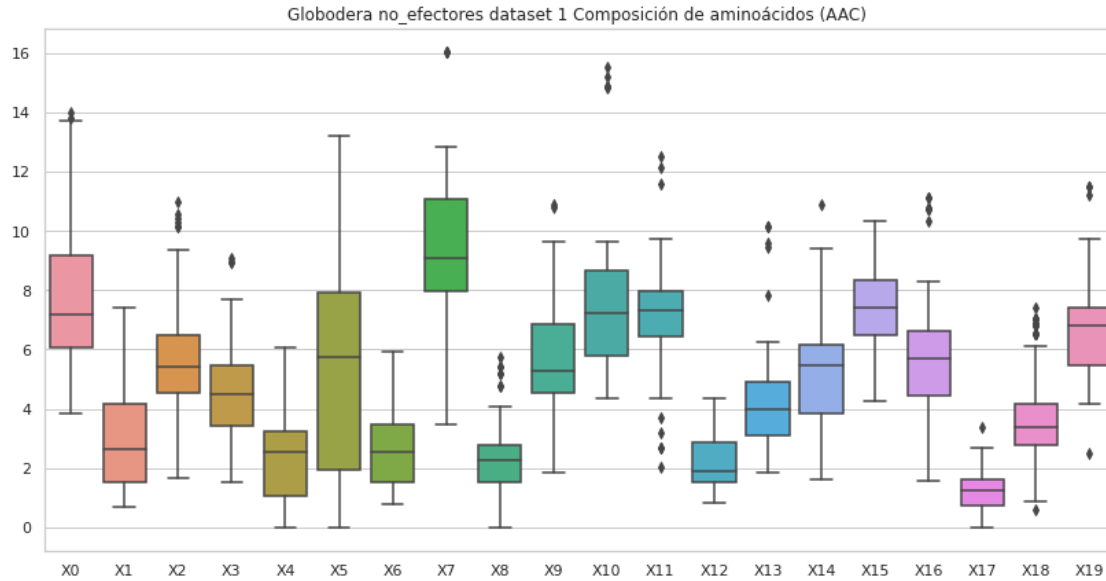
  

	X12	X13	X14	X15	X16	X17	\
count	107.000000	107.000000	107.000000	107.000000	107.000000	107.000000	
mean	2.183290	4.229561	5.377299	7.415430	5.899701	1.296953	
std	0.971846	1.550741	1.969346	1.275494	2.088022	0.645093	
min	0.855000	1.863000	1.600000	4.274000	1.571000	0.000000	
25%	1.509000	3.125000	3.857000	6.484000	4.452500	0.733000	

50%	1.887000	4.000000	5.469000	7.432000	5.682000	1.250000
75%	2.857000	4.906000	6.143000	8.333000	6.627500	1.632500
max	4.348000	10.135000	10.881000	10.345000	11.111000	3.390000

	X18	X19
count	107.000000	107.000000
mean	3.603458	6.675327
std	1.688177	1.609468
min	0.602000	2.479000
25%	2.795500	5.455500
50%	3.396000	6.792000
75%	4.188500	7.395500
max	7.407000	11.486000





### 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) +",\n"
    ↪" + 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 Globodera dataset 1, con valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	0.026925	0.029617	0.026925	0.010770	0.013462	0.032310	0.002692
1	0.028851	0.014426	0.015868	0.018753	0.017311	0.043277	0.004328
2	0.005843	0.000000	0.035060	0.025321	0.000000	0.000000	0.000000
3	0.031386	0.015693	0.017263	0.020401	0.020401	0.047080	0.004708
4	0.003170	0.000000	0.002642	0.002642	0.034343	0.008982	0.003170
..	...	...	...	...	...	...	...
119	0.025978	0.011546	0.012989	0.012989	0.015875	0.047626	0.001443
120	0.028841	0.017625	0.014421	0.020830	0.020830	0.052876	0.000000
121	0.029063	0.017761	0.014532	0.020990	0.020990	0.053282	0.000000
122	0.032341	0.015401	0.016941	0.020021	0.018481	0.046202	0.004620
123	0.029946	0.014973	0.016470	0.017967	0.019465	0.046416	0.004492

	X7	X8	X9 ...	X74	X75	X76 \
0	0.021540	0.037695	0.035002 ...	0.017599	-0.007318	0.015308
1	0.027409	0.027409	0.017311 ...	0.002610	0.009824	0.015526
2	0.005843	0.015582	0.005843 ...	0.008463	0.061221	0.018124
3	0.029817	0.028248	0.020401 ...	0.000542	0.010097	0.007572
4	0.008982	0.005812	0.025361 ...	0.023094	0.013264	-0.001554
..	...	...	...	...	...	...
119	0.027421	0.030307	0.018762 ...	0.008415	0.015502	0.024947
120	0.027239	0.028841	0.025637 ...	-0.000400	0.005884	0.009712
121	0.027448	0.029063	0.025834 ...	-0.000191	0.006114	0.008668
122	0.029261	0.029261	0.020021 ...	0.002205	0.009884	0.009482
123	0.028448	0.028448	0.019465 ...	0.004354	0.011411	0.009131

	X77	X78	X79	X80	X81	X82	X83
0	-0.011429	0.003381	0.011562	-0.006313	0.000640	0.017155	efectores
1	0.009401	0.011447	0.030113	-0.010029	-0.012344	0.021630	efectores
2	-0.002174	0.032520	-0.003085	-0.002872	0.027661	0.003855	efectores
3	0.008261	0.011305	0.031108	-0.008798	-0.015521	0.022464	efectores
4	0.028477	0.015071	-0.002723	0.024517	0.012591	-0.002753	efectores
..	...	...	...	...	...	...	...
119	-0.011390	-0.006509	0.040574	-0.006887	-0.005581	0.030166	efectores
120	-0.001153	-0.000958	0.029995	-0.017784	-0.019358	0.024858	efectores

121	-0.001415	-0.001593	0.030288	-0.017836	-0.019357	0.024709	efectores
122	0.008070	0.011162	0.031665	-0.007919	-0.015178	0.021534	efectores
123	0.008016	0.009652	0.029893	-0.010240	-0.016497	0.020286	efectores

[124 rows x 84 columns]

Composición de pseudo aminoácidos (PseAAC) hidro\_mass efectores Globodera dataset 1, con valores atípicos.

Estadísticas.

	X0	X1	X2	X3	X4	X5 \
count	124.000000	124.000000	124.000000	124.000000	124.000000	124.000000
mean	0.024832	0.016232	0.016755	0.015156	0.018689	0.035373
std	0.013705	0.011266	0.008985	0.010371	0.016998	0.016339
min	-0.058065	-0.029032	-0.000000	-0.000000	-0.145162	-0.000000
25%	0.025353	0.011041	0.012876	0.010464	0.013646	0.027403
50%	0.027678	0.016889	0.015875	0.012713	0.018663	0.034181
75%	0.029172	0.028602	0.024209	0.020405	0.021568	0.047531
max	0.061832	0.036766	0.044173	0.057415	0.042835	0.076232

	X6	X7	X8	X9 ...	X73 \
count	124.000000	124.000000	124.000000	124.000000 ...	124.000000
mean	0.003314	0.022325	0.028834	0.027943 ...	0.023893
std	0.003315	0.009843	0.012376	0.015193 ...	0.018113
min	0.000000	-0.029032	-0.000000	-0.087097 ...	-0.098235
25%	0.001470	0.019770	0.027199	0.020108 ...	0.010584
50%	0.002727	0.023263	0.029289	0.026203 ...	0.028259
75%	0.003322	0.027670	0.037394	0.034918 ...	0.034710
max	0.019138	0.055056	0.053168	0.085386 ...	0.083800

	X74	X75	X76	X77	X78	X79 \
count	124.000000	124.000000	124.000000	124.000000	124.000000	124.000000
mean	0.013560	0.007515	0.012555	0.001764	0.005207	0.019774
std	0.044120	0.030656	0.012394	0.018582	0.015942	0.015538
min	-0.089173	-0.036601	-0.071758	-0.047701	-0.020328	-0.005384
25%	0.002116	-0.008457	0.008078	-0.010795	-0.001596	0.010250
50%	0.010112	0.006780	0.013952	-0.002556	0.003063	0.018994
75%	0.019160	0.013100	0.018199	0.008917	0.011315	0.031260
max	0.475724	0.312216	0.036771	0.077622	0.144809	0.071835

	X80	X81	X82
count	124.000000	124.000000	124.000000
mean	-0.003535	-0.002042	0.016953
std	0.023976	0.019199	0.018594
min	-0.033313	-0.031491	-0.042313
25%	-0.014002	-0.015308	0.009275
50%	-0.008025	-0.003475	0.019417



75%	-0.002739	0.004015	0.024864
max	0.204323	0.153054	0.157695

[8 rows x 83 columns]

no\_efectores

Composición de pseudo aminoácidos (PseAAC) hidro\_mass no\_efectores Globodera dataset 1, con valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	0.030936	0.015468	0.023202	0.054138	0.038670	0.092808	0.023202
1	0.036878	0.018439	0.025354	0.050707	0.023049	0.046097	0.013829
2	0.041471	0.015279	0.024009	0.050201	0.030557	0.045836	0.013096
3	0.039088	0.012344	0.024687	0.043203	0.028802	0.043203	0.012344
4	0.035137	0.016535	0.022736	0.043405	0.026870	0.047539	0.008268
..	...	...	...	...	...	...	
119	0.002596	0.001557	0.002596	0.002596	0.028551	0.010382	0.003634
120	0.038323	0.004968	0.009226	0.006387	0.009935	0.018452	0.007806
121	0.020457	0.011365	0.015911	0.003409	0.009092	0.031822	0.001136
122	0.044014	0.002445	0.009781	0.017117	0.014671	0.031788	0.012226
123	0.040549	0.012805	0.025610	0.044817	0.027744	0.044817	0.012805

	X7	X8	X9 ...	X74	X75	X76 \
0	0.054138	0.077340	0.054138 ...	-0.049355	-0.090928	0.031378
1	0.027658	0.041488	0.043793 ...	-0.015609	-0.010030	0.033433
2	0.030557	0.037105	0.050201 ...	-0.021556	-0.012656	0.035198
3	0.028802	0.041146	0.045260 ...	-0.012176	-0.010928	0.029415
4	0.028937	0.037204	0.049606 ...	-0.001751	-0.009842	0.036493
..	...	...	...	...	...	
119	0.006748	0.002076	0.023360 ...	0.022254	0.012280	-0.002257
120	0.012774	0.023419	0.020581 ...	0.002316	-0.001468	0.017586
121	0.017047	0.015911	0.012501 ...	0.007564	-0.007357	0.022607
122	0.009781	0.029343	0.031788 ...	0.011176	0.006779	0.022918
123	0.029878	0.040549	0.051220 ...	-0.017518	-0.020837	0.030097

	X77	X78	X79	X80	X81	X82	X83
0	0.039313	0.008129	0.062444	0.022041	-0.044566	0.082349	no_efectores
1	-0.021945	-0.023573	0.044632	0.011590	0.017624	0.049628	no_efectores
2	-0.016841	-0.010432	0.048263	-0.002101	-0.011366	0.038724	no_efectores
3	-0.020585	-0.020110	0.041026	-0.004566	0.011643	0.040987	no_efectores
4	-0.032614	-0.025132	0.055779	0.000176	0.012538	0.039278	no_efectores
..	...	...	...	...	...	...	
119	0.019938	0.010880	0.005265	0.025778	0.011887	-0.000158	no_efectores
120	0.008212	0.002324	0.036228	-0.008566	-0.001221	0.023525	no_efectores
121	-0.005184	-0.002788	0.025976	-0.005877	0.002111	0.025323	no_efectores

```

122 -0.004622 -0.000306  0.007000 -0.007235 -0.003353 -0.004973  no_efectores
123 -0.016623 -0.016227  0.049913  0.000086  0.006216  0.047137  no_efectores

```

[124 rows x 84 columns]

Composición de pseudo aminoácidos (PseAAC) hidro\_mass no\_efectores Globodera dataset 1, con valores atípicos.  
Estadísticas.

	X0	X1	X2	X3	X4	X5 \
count	124.000000	124.000000	124.000000	124.000000	124.000000	124.000000
mean	0.029561	0.009918	0.018590	0.024289	0.019509	0.037780
std	0.014609	0.007945	0.013278	0.021217	0.014048	0.019500
min	0.000785	0.000000	0.001569	0.000000	0.001191	0.007512
25%	0.019503	0.001919	0.009641	0.004679	0.008666	0.024954
50%	0.029276	0.011449	0.016040	0.020730	0.016393	0.037469
75%	0.038634	0.016376	0.024558	0.043254	0.027810	0.045226
max	0.106171	0.042812	0.106171	0.148639	0.127405	0.140110

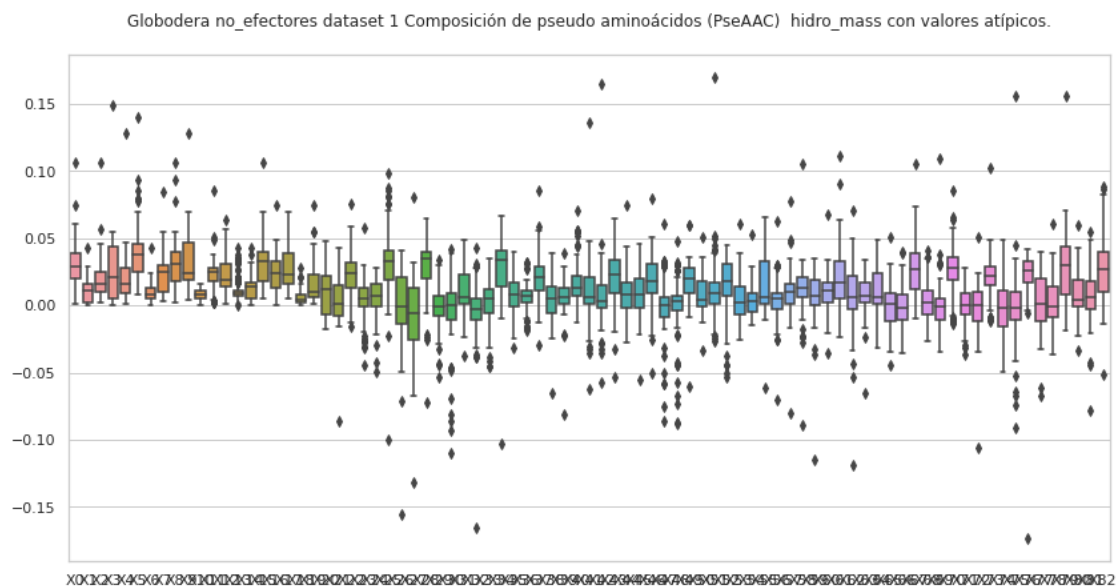
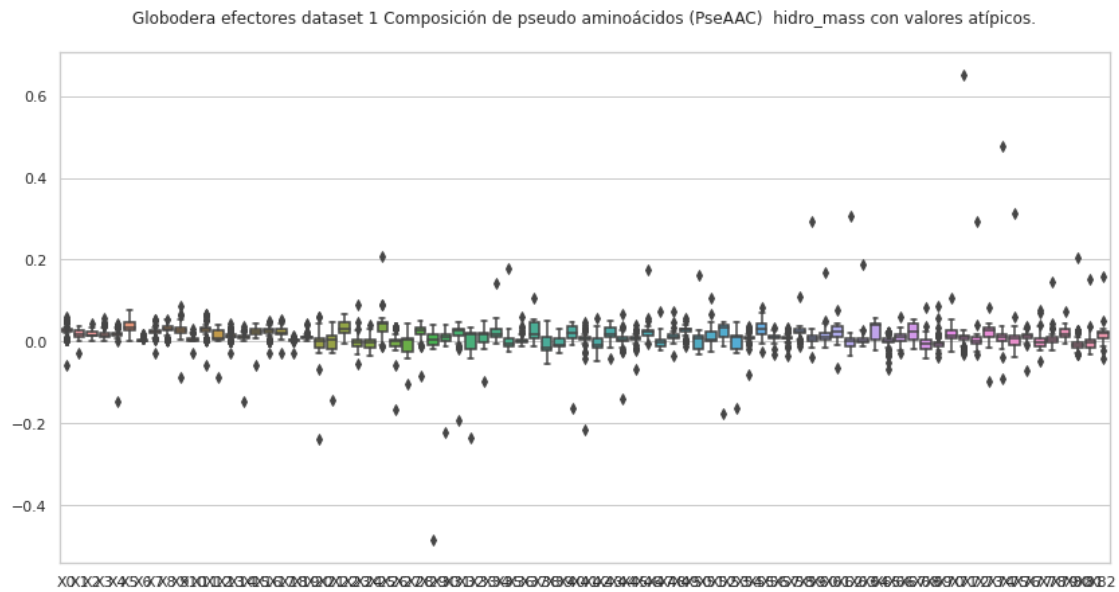
	X6	X7	X8	X9 ...	X73 \
count	124.000000	124.000000	124.000000	124.000000	124.000000
mean	0.009485	0.022354	0.029539	0.031541	0.021167
std	0.006576	0.013193	0.016225	0.017175	0.014649
min	0.000000	0.003177	0.002076	0.003574	-0.011709
25%	0.004762	0.010019	0.018010	0.019341	0.014954
50%	0.008222	0.024388	0.031201	0.023496	0.021879
75%	0.013176	0.030093	0.039375	0.046578	0.029103
max	0.042468	0.084066	0.106171	0.127405	0.102255

	X74	X75	X76	X77	X78	X79 \
count	124.000000	124.000000	124.000000	124.000000	124.000000	124.000000
mean	-0.001141	-0.001214	0.021382	0.001543	0.001429	0.027812
std	0.017981	0.025299	0.021449	0.020729	0.017998	0.023331
min	-0.049355	-0.090928	-0.173343	-0.066859	-0.036841	-0.019120
25%	-0.015704	-0.010956	0.017152	-0.012212	-0.009120	0.008441
50%	-0.001771	-0.001476	0.025988	0.001403	-0.000552	0.029837
75%	0.011431	0.010122	0.032297	0.020316	0.013825	0.043162
max	0.048528	0.155333	0.041440	0.039313	0.060165	0.155614

	X80	X81	X82
count	124.000000	124.000000	124.000000
mean	0.008459	0.006253	0.025493
std	0.015849	0.016914	0.022340
min	-0.033995	-0.077836	-0.051272
25%	-0.001331	-0.002737	0.010341
50%	0.003557	0.006278	0.027110
75%	0.018835	0.017804	0.039455

max 0.059376 0.054626 0.087943

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

	X0	X1	X2	X3	X4	X5	X6 \
0	0.026925	0.029617	0.026925	0.010770	0.013462	0.032310	0.002692
1	0.028851	0.014426	0.015868	0.018753	0.017311	0.043277	0.004328
3	0.031386	0.015693	0.017263	0.020401	0.020401	0.047080	0.004708
4	0.003170	0.000000	0.002642	0.002642	0.034343	0.008982	0.003170
5	0.028195	0.018244	0.014927	0.021561	0.021561	0.053073	0.001659
..	...	...	...	...	...	...	...
119	0.025978	0.011546	0.012989	0.012989	0.015875	0.047626	0.001443
120	0.028841	0.017625	0.014421	0.020830	0.020830	0.052876	0.000000
121	0.029063	0.017761	0.014532	0.020990	0.020990	0.053282	0.000000
122	0.032341	0.015401	0.016941	0.020021	0.018481	0.046202	0.004620
123	0.029946	0.014973	0.016470	0.017967	0.019465	0.046416	0.004492

	X7	X8	X9	...	X74	X75	X76 \
0	0.021540	0.037695	0.035002	...	0.017599	-0.007318	0.015308
1	0.027409	0.027409	0.017311	...	0.002610	0.009824	0.015526
3	0.029817	0.028248	0.020401	...	0.000542	0.010097	0.007572
4	0.008982	0.005812	0.025361	...	0.023094	0.013264	-0.001554
5	0.028195	0.028195	0.026536	...	0.002654	0.005925	0.010787
..	...	...	...	...	...	...	...
119	0.027421	0.030307	0.018762	...	0.008415	0.015502	0.024947
120	0.027239	0.028841	0.025637	...	-0.000400	0.005884	0.009712
121	0.027448	0.029063	0.025834	...	-0.000191	0.006114	0.008668
122	0.029261	0.029261	0.020021	...	0.002205	0.009884	0.009482
123	0.028448	0.028448	0.019465	...	0.004354	0.011411	0.009131

	X77	X78	X79	X80	X81	X82	X83
0	-0.011429	0.003381	0.011562	-0.006313	0.000640	0.017155	efectores
1	0.009401	0.011447	0.030113	-0.010029	-0.012344	0.021630	efectores
3	0.008261	0.011305	0.031108	-0.008798	-0.015521	0.022464	efectores
4	0.028477	0.015071	-0.002723	0.024517	0.012591	-0.002753	efectores
5	-0.002003	-0.001668	0.029714	-0.018219	-0.018168	0.023475	efectores
..	...	...	...	...	...	...	...
119	-0.011390	-0.006509	0.040574	-0.006887	-0.005581	0.030166	efectores
120	-0.001153	-0.000958	0.029995	-0.017784	-0.019358	0.024858	efectores
121	-0.001415	-0.001593	0.030288	-0.017836	-0.019357	0.024709	efectores
122	0.008070	0.011162	0.031665	-0.007919	-0.015178	0.021534	efectores
123	0.008016	0.009652	0.029893	-0.010240	-0.016497	0.020286	efectores

[110 rows x 84 columns]

Composición de pseudo aminoácidos (PseAAC) hidro\_mass efectores Globodera  
dataset 1, sin valores atípicos.  
Estadísticas.

	X0	X1	X2	X3	X4	X5 \
count	110.000000	110.000000	110.000000	110.000000	110.000000	110.000000
mean	0.024719	0.016873	0.015869	0.013795	0.019831	0.035514
std	0.010646	0.010367	0.007885	0.008461	0.007562	0.015117
min	0.003138	0.000000	0.002125	0.002059	0.007917	0.008438
25%	0.025885	0.011504	0.012934	0.010403	0.013794	0.030207
50%	0.027627	0.017116	0.014937	0.011700	0.018891	0.034047
75%	0.029053	0.028669	0.023840	0.020035	0.021366	0.047594
max	0.054031	0.036766	0.034120	0.045386	0.036394	0.056438

	X6	X7	X8	X9 ...	X73 \
count	110.000000	110.000000	110.000000	110.000000	110.000000
mean	0.002668	0.021893	0.028045	0.027509	0.026019
std	0.001904	0.007003	0.011632	0.008066	0.011663
min	0.000000	0.007463	0.005662	0.010556	-0.009754
25%	0.001462	0.020169	0.027388	0.020060	0.021729
50%	0.002693	0.022372	0.029273	0.026091	0.029210
75%	0.003214	0.027448	0.037005	0.034493	0.034957
max	0.009133	0.030950	0.052317	0.052317	0.042178

	X74	X75	X76	X77	X78	X79 \
count	110.000000	110.000000	110.000000	110.000000	110.000000	110.000000
mean	0.011120	0.004137	0.012619	0.000910	0.003892	0.020034
std	0.009538	0.011899	0.008971	0.014865	0.007610	0.014751
min	-0.018296	-0.036601	-0.004580	-0.022202	-0.016065	-0.004141
25%	0.002159	-0.008581	0.008128	-0.010417	-0.001560	0.010920
50%	0.011297	0.006451	0.013194	-0.002556	0.003020	0.022444
75%	0.019108	0.012498	0.016407	0.008551	0.011056	0.031459
max	0.033645	0.035709	0.031247	0.032278	0.017303	0.043591

	X80	X81	X82
count	110.000000	110.000000	110.000000
mean	-0.004839	-0.003636	0.017234
std	0.014611	0.011706	0.011828
min	-0.033313	-0.024791	-0.010096
25%	-0.012471	-0.015314	0.012622
50%	-0.008052	-0.002531	0.020297
75%	-0.004851	0.003281	0.024876
max	0.026063	0.022402	0.049515

[8 rows x 83 columns]

no\_efectores

Composición de pseudo aminoácidos (PseAAC) hidro\_mass no\_efectores Globodera dataset 1, sin valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
1	0.036878	0.018439	0.025354	0.050707	0.023049	0.046097	0.013829
2	0.041471	0.015279	0.024009	0.050201	0.030557	0.045836	0.013096
3	0.039088	0.012344	0.024687	0.043203	0.028802	0.043203	0.012344
4	0.035137	0.016535	0.022736	0.043405	0.026870	0.047539	0.008268
5	0.010826	0.001804	0.015036	0.022253	0.007819	0.007819	0.003007
..	...	...	...	...	...	...	
119	0.002596	0.001557	0.002596	0.002596	0.028551	0.010382	0.003634
120	0.038323	0.004968	0.009226	0.006387	0.009935	0.018452	0.007806
121	0.020457	0.011365	0.015911	0.003409	0.009092	0.031822	0.001136
122	0.044014	0.002445	0.009781	0.017117	0.014671	0.031788	0.012226
123	0.040549	0.012805	0.025610	0.044817	0.027744	0.044817	0.012805

	X7	X8	X9	...	X74	X75	X76 \
1	0.027658	0.041488	0.043793	...	-0.015609	-0.010030	0.033433
2	0.030557	0.037105	0.050201	...	-0.021556	-0.012656	0.035198
3	0.028802	0.041146	0.045260	...	-0.012176	-0.010928	0.029415
4	0.028937	0.037204	0.049606	...	-0.001751	-0.009842	0.036493
5	0.010224	0.020448	0.010224	...	0.005200	0.033396	0.000713
..	...	...	...	...	...	...	
119	0.006748	0.002076	0.023360	...	0.022254	0.012280	-0.002257
120	0.012774	0.023419	0.020581	...	0.002316	-0.001468	0.017586
121	0.017047	0.015911	0.012501	...	0.007564	-0.007357	0.022607
122	0.009781	0.029343	0.031788	...	0.011176	0.006779	0.022918
123	0.029878	0.040549	0.051220	...	-0.017518	-0.020837	0.030097

	X77	X78	X79	X80	X81	X82	X83
1	-0.021945	-0.023573	0.044632	0.011590	0.017624	0.049628	no_efectores
2	-0.016841	-0.010432	0.048263	-0.002101	-0.011366	0.038724	no_efectores
3	-0.020585	-0.020110	0.041026	-0.004566	0.011643	0.040987	no_efectores
4	-0.032614	-0.025132	0.055779	0.000176	0.012538	0.039278	no_efectores
5	0.013633	0.036150	0.008497	0.002626	0.023729	-0.002583	no_efectores
..	...	...	...	...	...	...	
119	0.019938	0.010880	0.005265	0.025778	0.011887	-0.000158	no_efectores
120	0.008212	0.002324	0.036228	-0.008566	-0.001221	0.023525	no_efectores
121	-0.005184	-0.002788	0.025976	-0.005877	0.002111	0.025323	no_efectores
122	-0.004622	-0.000306	0.007000	-0.007235	-0.003353	-0.004973	no_efectores
123	-0.016623	-0.016227	0.049913	0.000086	0.006216	0.047137	no_efectores

[111 rows x 84 columns]

Composición de pseudo aminoácidos (PseAAC) hidro\_mass no\_efectores Globodera  
dataset 1, sin valores atípicos.  
Estadísticas.

	X0	X1	X2	X3	X4	X5 \
count	111.000000	111.000000	111.000000	111.000000	111.000000	111.000000
mean	0.027952	0.009082	0.016759	0.021760	0.017372	0.034689
std	0.012106	0.007004	0.010076	0.017855	0.009243	0.014196
min	0.000785	0.000000	0.001569	0.000000	0.001191	0.007512
25%	0.019296	0.001881	0.009503	0.004605	0.008588	0.024905
50%	0.028872	0.011355	0.014564	0.018934	0.015750	0.035089
75%	0.038103	0.015744	0.024058	0.042938	0.027083	0.044695
max	0.055092	0.022337	0.045716	0.053930	0.037554	0.069804

	X6	X7	X8	X9 ...	X73 \
count	111.000000	111.000000	111.000000	111.000000 ...	111.000000
mean	0.008596	0.020516	0.026837	0.029265 ...	0.021027
std	0.005244	0.010714	0.012330	0.013827 ...	0.012109
min	0.000785	0.003177	0.002076	0.003574 ...	-0.010816
25%	0.004637	0.009870	0.017468	0.019266 ...	0.015684
50%	0.007806	0.022660	0.029071	0.023200 ...	0.022292
75%	0.012921	0.029614	0.037524	0.044275 ...	0.029071
max	0.021780	0.046387	0.045437	0.054525 ...	0.048707

	X74	X75	X76	X77	X78	X79 \
count	111.000000	111.000000	111.000000	111.000000	111.000000	111.000000
mean	-0.000178	0.000213	0.023298	0.000572	0.000761	0.026026
std	0.016096	0.015621	0.012051	0.018213	0.017061	0.018659
min	-0.031594	-0.052543	-0.003804	-0.032614	-0.032141	-0.007303
25%	-0.015131	-0.009096	0.020112	-0.013122	-0.009126	0.008385
50%	0.000081	-0.001206	0.026131	-0.001599	-0.001816	0.029200
75%	0.012278	0.010193	0.032556	0.019046	0.013027	0.040927
max	0.034004	0.034372	0.041440	0.029796	0.036741	0.063680

	X80	X81	X82
count	111.000000	111.000000	111.000000
mean	0.008412	0.008102	0.024174
std	0.014019	0.013442	0.018001
min	-0.020898	-0.017304	-0.005113
25%	-0.001131	-0.000295	0.010617
50%	0.003556	0.007210	0.026862
75%	0.016729	0.017984	0.038374
max	0.042196	0.054626	0.080779

[8 rows x 83 columns]





```

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

for etiq in "efectores", "no_efectores":
    titulo = (str(transf)+" "+ str(comp)+" "+ str(etiq) + " "+ str(nombre2) +",\n
↳" + 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)+"\n
↳"+str(transf)+" "+str(comp)+" "+str(estado))

```

efectores

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

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	0.025602	0.028162	0.025602	0.010241	0.012801	0.030722	0.002560
1	0.030226	0.015113	0.016625	0.019647	0.018136	0.045340	0.004534
2	0.038552	0.000000	0.231314	0.167060	0.000000	0.000000	0.000000
3	0.032353	0.016176	0.017794	0.021029	0.021029	0.048529	0.004853
4	0.021066	0.000000	0.017555	0.017555	0.228216	0.059687	0.021066
..	...	...	...	...	...	...	...
119	0.026440	0.011751	0.013220	0.013220	0.016158	0.048473	0.001469
120	0.028797	0.017598	0.014398	0.020798	0.020798	0.052794	0.000000
121	0.029032	0.017741	0.014516	0.020967	0.020967	0.053224	0.000000
122	0.033361	0.015886	0.017475	0.020652	0.019063	0.047658	0.004766
123	0.030997	0.015499	0.017049	0.018598	0.020148	0.048046	0.004650
	X7	X8	X9 ...	X32	X33	X34 \	

0	0.020481	0.035843	0.033282	...	0.029464	0.022259	0.040654
1	0.028715	0.028715	0.018136	...	0.021587	0.041172	0.044513
2	0.038552	0.102806	0.038552	...	-0.073516	0.021814	-0.003083
3	0.030735	0.029118	0.021029	...	0.021777	0.038706	0.047669
4	0.059687	0.038621	0.168529	...	0.024751	0.028437	0.025311
..	...	...	...	...	...	...	...
119	0.027908	0.030846	0.019095	...	0.024063	0.036598	0.042487
120	0.027197	0.028797	0.025597	...	0.022573	0.037104	0.046131
121	0.027419	0.029032	0.025806	...	0.022416	0.036960	0.046422
122	0.030184	0.030184	0.020652	...	0.023830	0.039523	0.048549
123	0.029447	0.029447	0.020148	...	0.022883	0.040420	0.049976

	X35	X36	X37	X38	X39	X40	X41
0	0.046910	0.026478	0.032980	0.014556	0.010994	0.016312	efectores
1	0.031961	0.012039	0.032534	0.016266	0.031549	0.022661	efectores
2	-0.017459	0.045811	-0.019417	0.119574	-0.020353	0.025431	efectores
3	0.030367	0.007702	0.026764	0.007805	0.032066	0.023156	efectores
4	-0.046043	-0.021633	0.067594	-0.010327	-0.018092	-0.018294	efectores
..	...	...	...	...	...	...	...
119	0.021816	0.014746	0.037394	0.025391	0.041296	0.030702	efectores
120	0.022795	0.006296	0.028179	0.009697	0.029949	0.024819	efectores
121	0.023041	0.005892	0.027926	0.008659	0.030255	0.024682	efectores
122	0.028790	0.009288	0.028321	0.009781	0.032663	0.022213	efectores
123	0.028380	0.009280	0.030842	0.009451	0.030942	0.020998	efectores

[124 rows x 42 columns]

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

Estadísticas.

	X0	X1	X2	X3	X4	X5 \
count	124.000000	124.000000	124.000000	124.000000	124.000000	124.000000
mean	0.030551	0.017163	0.021550	0.019636	0.049554	0.043627
std	0.011356	0.010850	0.021991	0.018352	0.073102	0.013901
min	0.008875	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.025600	0.011617	0.014486	0.010419	0.012901	0.031173
50%	0.027590	0.017159	0.017497	0.016831	0.019647	0.047965
75%	0.030232	0.027495	0.023067	0.020668	0.024091	0.053166
max	0.097829	0.048915	0.231314	0.167060	0.244573	0.073287

	X6	X7	X8	X9 ...	X31 \
count	124.000000	124.000000	124.000000	124.000000	124.000000
mean	0.006183	0.031567	0.035470	0.052619	0.026091
std	0.007385	0.015049	0.012002	0.052661	0.027705
min	0.000000	0.005836	0.000000	0.011833	-0.102855
25%	0.001483	0.020901	0.029213	0.021928	0.017691

50%	0.002578	0.027690	0.034898	0.032011	...	0.030566
75%	0.007925	0.030447	0.037373	0.049492	...	0.043255
max	0.028992	0.088502	0.102806	0.182440	...	0.080065

	X32	X33	X34	X35	X36	X37 \
count	124.000000	124.000000	124.000000	124.000000	124.000000	124.000000
mean	0.022306	0.030990	0.034009	0.015863	0.010391	0.033357
std	0.022365	0.013718	0.016676	0.032494	0.025561	0.022536
min	-0.183907	-0.006893	-0.022010	-0.069360	-0.181470	-0.025888
25%	0.022255	0.022374	0.027411	0.006342	0.006069	0.027776
50%	0.024446	0.035252	0.040654	0.023991	0.014070	0.032482
75%	0.029463	0.037856	0.044590	0.041355	0.026413	0.036956
max	0.056563	0.110526	0.051342	0.050771	0.054942	0.165509

	X38	X39	X40
count	124.000000	124.000000	124.000000
mean	0.013809	0.017991	0.011021
std	0.019210	0.020153	0.032010
min	-0.031017	-0.027340	-0.265689
25%	0.008378	0.009887	0.010781
50%	0.013796	0.023257	0.018490
75%	0.021711	0.032314	0.025558
max	0.120901	0.065224	0.055506

[8 rows x 41 columns]

no\_efectores

Composición de pseudo aminoácidos (PseAAC) mass no\_efectores Globodera dataset 1, con valores atípicos.  
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	0.025063	0.012532	0.018797	0.043860	0.031329	0.075189	0.018797
1	0.036412	0.018206	0.025033	0.050066	0.022757	0.045515	0.013654
2	0.039242	0.014457	0.022719	0.047503	0.028915	0.043372	0.012392
3	0.038320	0.012101	0.024202	0.042354	0.028236	0.042354	0.012101
4	0.033740	0.015878	0.021832	0.041679	0.025802	0.045649	0.007939
..	...	...	...	...	...	...	...
119	0.012439	0.007464	0.012439	0.012439	0.136833	0.049757	0.017415
120	0.044915	0.005822	0.010813	0.007486	0.011645	0.021626	0.009149
121	0.020789	0.011549	0.016169	0.003465	0.009240	0.032338	0.001155
122	0.048597	0.002700	0.010799	0.018899	0.016199	0.035098	0.013499
123	0.037975	0.011992	0.023984	0.041973	0.025983	0.041973	0.011992

	X7	X8	X9	...	X32	X33	X34 \
0	0.043860	0.062658	0.043860	...	0.016781	-0.012049	-0.020015

1	0.027309	0.040963	0.043239	...	0.008276	0.003463	0.001788
2	0.028915	0.035111	0.047503	...	0.005656	0.006798	0.001990
3	0.028236	0.040337	0.044371	...	0.015435	0.008046	0.005603
4	0.027786	0.035725	0.047634	...	0.009820	0.005075	0.005393
..	...	...	...	...	...	...	...
119	0.032342	0.009951	0.111954	...	0.026497	0.010677	0.011767
120	0.014972	0.027448	0.024121	...	0.034786	0.035007	0.032910
121	0.017324	0.016169	0.012704	...	0.028035	0.043964	0.031105
122	0.010799	0.032398	0.035098	...	0.017505	0.027606	0.026073
123	0.027982	0.037975	0.047969	...	0.010406	0.005771	0.003920

	X35	X36	X37	X38	X39	X40	X41
0	0.051003	0.069377	0.022337	0.025421	0.050589	0.066715	no_efectores
1	0.027817	0.023638	0.021538	0.033010	0.044068	0.049001	no_efectores
2	0.039487	0.028643	0.027513	0.033306	0.045669	0.036642	no_efectores
3	0.034595	0.030897	0.026624	0.028838	0.040220	0.040181	no_efectores
4	0.028874	0.038371	0.022587	0.035042	0.053561	0.037716	no_efectores
..	...	...	...	...	...	...	...
119	0.034488	0.036122	0.016255	-0.010817	0.025234	-0.000758	no_efectores
120	0.025988	0.026182	0.021499	0.020611	0.042460	0.027572	no_efectores
121	0.031822	0.045734	0.032167	0.022973	0.026398	0.025734	no_efectores
122	0.033829	0.025487	0.019580	0.025304	0.007729	-0.005491	no_efectores
123	0.032158	0.028049	0.029318	0.028187	0.046745	0.044145	no_efectores

[124 rows x 42 columns]

Composición de pseudo aminoácidos (PseAAC) mass no\_efectores Globodera dataset 1, con valores atípicos.  
Estadísticas.

	X0	X1	X2	X3	X4	X5 \
count	124.000000	124.000000	124.000000	124.000000	124.000000	124.000000
mean	0.035125	0.010773	0.022163	0.027921	0.033103	0.045120
std	0.013119	0.007884	0.014201	0.022900	0.049865	0.013747
min	0.008658	0.000000	0.002115	0.000000	0.001269	0.021417
25%	0.023843	0.003696	0.014243	0.005190	0.011816	0.034453
50%	0.036377	0.011467	0.018013	0.024261	0.021995	0.042752
75%	0.041235	0.015934	0.025217	0.044122	0.029140	0.049474
max	0.109240	0.054996	0.109240	0.152936	0.311672	0.086576

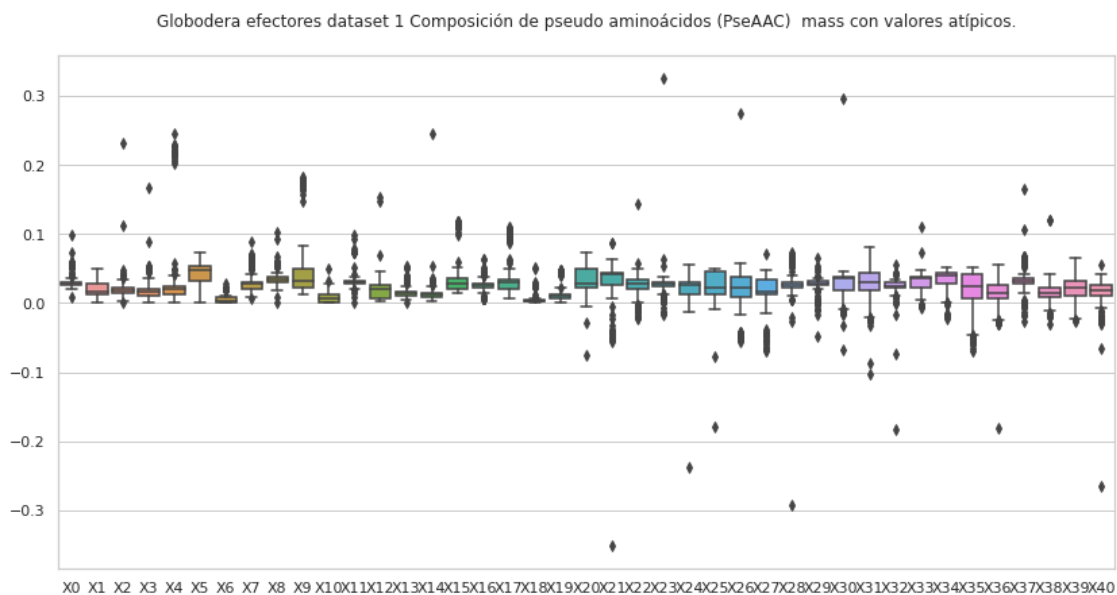
	X6	X7	X8	X9 ...	X31 \
count	124.000000	124.000000	124.000000	124.000000	124.000000
mean	0.012296	0.027568	0.035034	0.044564	0.017788
std	0.007749	0.013686	0.015695	0.036091	0.020272
min	0.000000	0.003384	0.003807	0.003807	-0.063653
25%	0.006929	0.015658	0.027382	0.024076	0.001415
50%	0.012052	0.028337	0.035500	0.039806	0.015295

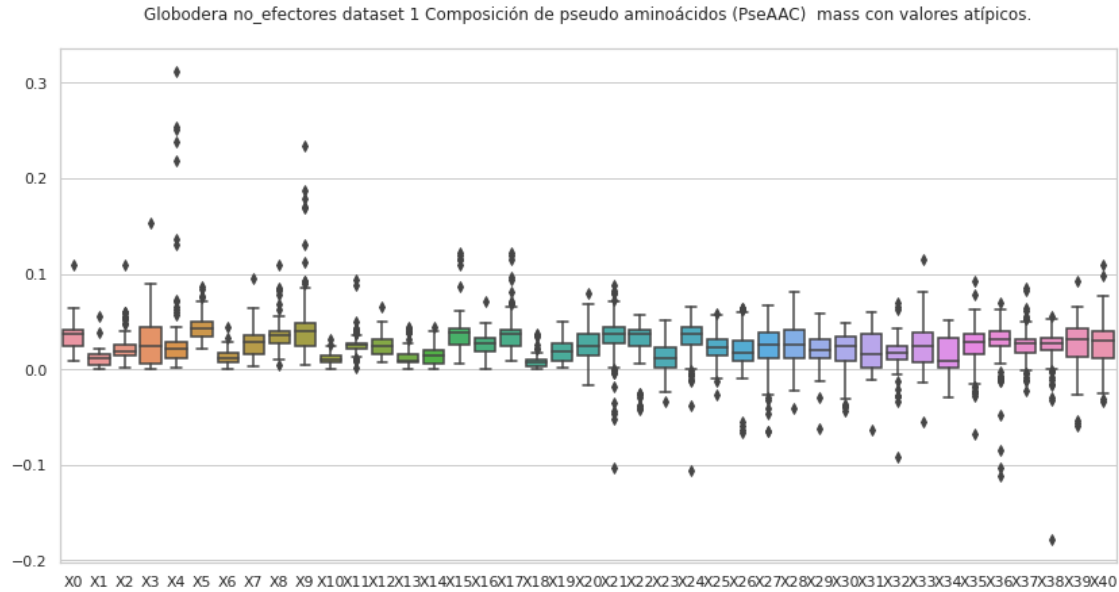
75%	0.016748	0.035715	0.040463	0.048717	...	0.037369
max	0.043696	0.095233	0.109240	0.233754	...	0.059075

	X32	X33	X34	X35	X36	X37 \
count	124.000000	124.000000	124.000000	124.000000	124.000000	124.000000
mean	0.015240	0.023834	0.014165	0.025577	0.027532	0.025491
std	0.018189	0.021864	0.019325	0.021733	0.026314	0.016518
min	-0.092032	-0.054688	-0.029504	-0.067155	-0.111964	-0.022503
25%	0.009848	0.007406	0.001574	0.016077	0.023523	0.017543
50%	0.016667	0.024277	0.008112	0.028744	0.030922	0.027172
75%	0.023824	0.038585	0.032452	0.037577	0.040506	0.031000
max	0.069366	0.114718	0.051675	0.092593	0.069377	0.084928

	X38	X39	X40
count	124.000000	124.000000	124.000000
mean	0.022491	0.027160	0.026358
std	0.023833	0.024153	0.024505
min	-0.178355	-0.059206	-0.034279
25%	0.020063	0.013500	0.012148
50%	0.026638	0.030972	0.029658
75%	0.033414	0.043317	0.040306
max	0.055798	0.092805	0.109106

[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) +",\n"
             ' ' + 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)+"\n
↳"+str(transf)+" "+str(comp))

```

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

	X0	X1	X2	X3	X4	X5	X6 \
0	0.025602	0.028162	0.025602	0.010241	0.012801	0.030722	0.002560
1	0.030226	0.015113	0.016625	0.019647	0.018136	0.045340	0.004534
3	0.032353	0.016176	0.017794	0.021029	0.021029	0.048529	0.004853
4	0.021066	0.000000	0.017555	0.017555	0.228216	0.059687	0.021066
5	0.027883	0.018042	0.014762	0.021322	0.021322	0.052486	0.001640
..	...	...	...	...	...	...	...
119	0.026440	0.011751	0.013220	0.013220	0.016158	0.048473	0.001469
120	0.028797	0.017598	0.014398	0.020798	0.020798	0.052794	0.000000
121	0.029032	0.017741	0.014516	0.020967	0.020967	0.053224	0.000000
122	0.033361	0.015886	0.017475	0.020652	0.019063	0.047658	0.004766
123	0.030997	0.015499	0.017049	0.018598	0.020148	0.048046	0.004650

	X7	X8	X9 ...	X32	X33	X34 \
0	0.020481	0.035843	0.033282 ...	0.029464	0.022259	0.040654
1	0.028715	0.028715	0.018136 ...	0.021587	0.041172	0.044513
3	0.030735	0.029118	0.021029 ...	0.021777	0.038706	0.047669
4	0.059687	0.038621	0.168529 ...	0.024751	0.028437	0.025311
5	0.027883	0.027883	0.026243 ...	0.021397	0.036446	0.046934
..	...	...	... ..	...	...	...
119	0.027908	0.030846	0.019095 ...	0.024063	0.036598	0.042487
120	0.027197	0.028797	0.025597 ...	0.022573	0.037104	0.046131
121	0.027419	0.029032	0.025806 ...	0.022416	0.036960	0.046422
122	0.030184	0.030184	0.020652 ...	0.023830	0.039523	0.048549
123	0.029447	0.029447	0.020148 ...	0.022883	0.040420	0.049976



	X35	X36	X37	X38	X39	X40	X41
0	0.046910	0.026478	0.032980	0.014556	0.010994	0.016312	efectores
1	0.031961	0.012039	0.032534	0.016266	0.031549	0.022661	efectores
3	0.030367	0.007702	0.026764	0.007805	0.032066	0.023156	efectores
4	-0.046043	-0.021633	0.067594	-0.010327	-0.018092	-0.018294	efectores
5	0.025124	0.004490	0.030066	0.010667	0.029385	0.023215	efectores
..	...	...	...	...	...	...	
119	0.021816	0.014746	0.037394	0.025391	0.041296	0.030702	efectores
120	0.022795	0.006296	0.028179	0.009697	0.029949	0.024819	efectores
121	0.023041	0.005892	0.027926	0.008659	0.030255	0.024682	efectores
122	0.028790	0.009288	0.028321	0.009781	0.032663	0.022213	efectores
123	0.028380	0.009280	0.030842	0.009451	0.030942	0.020998	efectores

[113 rows x 42 columns]

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

	X0	X1	X2	X3	X4	X5 \
count	113.000000	113.000000	113.000000	113.000000	113.000000	113.000000
mean	0.028388	0.016370	0.018632	0.016543	0.049369	0.044965
std	0.007080	0.009809	0.006165	0.008215	0.074005	0.012024
min	0.008875	0.000000	0.002958	0.002958	0.008712	0.020389
25%	0.025589	0.011604	0.014495	0.010415	0.012901	0.031393
50%	0.027428	0.016800	0.017440	0.014617	0.018975	0.048046
75%	0.029190	0.027250	0.022736	0.020499	0.021258	0.053147
max	0.058995	0.033830	0.048674	0.054351	0.230215	0.073287

	X6	X7	X8	X9 ...	X31 \
count	113.000000	113.000000	113.000000	113.000000	113.000000
mean	0.005999	0.030835	0.033446	0.051780	0.028335
std	0.007327	0.014036	0.006491	0.054150	0.023377
min	0.000000	0.008875	0.008875	0.011833	-0.032391
25%	0.001484	0.020829	0.029141	0.020803	0.026846
50%	0.002573	0.027606	0.033274	0.031381	0.032236
75%	0.006135	0.029859	0.036642	0.034682	0.043457
max	0.021925	0.071945	0.059941	0.182440	0.080065

	X32	X33	X34	X35	X36	X37 \
count	113.000000	113.000000	113.000000	113.000000	113.000000	113.000000
mean	0.025600	0.031297	0.037609	0.017378	0.011082	0.034693
std	0.006931	0.009733	0.011762	0.032865	0.018265	0.014522
min	0.000004	-0.006893	-0.015115	-0.069360	-0.029499	-0.013607
25%	0.022623	0.022906	0.034857	0.019934	0.005993	0.028179
50%	0.024751	0.035796	0.041221	0.024642	0.013271	0.032701
75%	0.029464	0.037873	0.045198	0.043055	0.026345	0.037059

max	0.056563	0.047195	0.051342	0.050771	0.050214	0.067864
-----	----------	----------	----------	----------	----------	----------

	X38	X39	X40
count	113.000000	113.000000	113.000000
mean	0.010982	0.018397	0.014409
std	0.013221	0.020430	0.018872
min	-0.031017	-0.027340	-0.030164
25%	0.008112	0.010531	0.013755
50%	0.013350	0.027087	0.020998
75%	0.016936	0.032730	0.026058
max	0.041190	0.065224	0.055506

[8 rows x 41 columns]

Composición de pseudo aminoácidos (PseAAC) mass no\_efectores Globodera dataset 1, sin valores atípicos.  
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
1	0.036412	0.018206	0.025033	0.050066	0.022757	0.045515	0.013654
2	0.039242	0.014457	0.022719	0.047503	0.028915	0.043372	0.012392
3	0.038320	0.012101	0.024202	0.042354	0.028236	0.042354	0.012101
4	0.033740	0.015878	0.021832	0.041679	0.025802	0.045649	0.007939
5	0.043210	0.007202	0.060014	0.088821	0.031207	0.031207	0.012003
..	...	...	...	...	...	...	...
119	0.012439	0.007464	0.012439	0.012439	0.136833	0.049757	0.017415
120	0.044915	0.005822	0.010813	0.007486	0.011645	0.021626	0.009149
121	0.020789	0.011549	0.016169	0.003465	0.009240	0.032338	0.001155
122	0.048597	0.002700	0.010799	0.018899	0.016199	0.035098	0.013499
123	0.037975	0.011992	0.023984	0.041973	0.025983	0.041973	0.011992

	X7	X8	X9	...	X32	X33	X34 \
1	0.027309	0.040963	0.043239	...	0.008276	0.003463	0.001788
2	0.028915	0.035111	0.047503	...	0.005656	0.006798	0.001990
3	0.028236	0.040337	0.044371	...	0.015435	0.008046	0.005603
4	0.027786	0.035725	0.047634	...	0.009820	0.005075	0.005393
5	0.040810	0.081619	0.040810	...	0.011314	0.009374	0.035823
..	...	...	...	...	...	...	...
119	0.032342	0.009951	0.111954	...	0.026497	0.010677	0.011767
120	0.014972	0.027448	0.024121	...	0.034786	0.035007	0.032910
121	0.017324	0.016169	0.012704	...	0.028035	0.043964	0.031105
122	0.010799	0.032398	0.035098	...	0.017505	0.027606	0.026073
123	0.027982	0.037975	0.047969	...	0.010406	0.005771	0.003920

	X35	X36	X37	X38	X39	X40	X41
1	0.027817	0.023638	0.021538	0.033010	0.044068	0.049001	no_efectores

2	0.039487	0.028643	0.027513	0.033306	0.045669	0.036642	no_efectores
3	0.034595	0.030897	0.026624	0.028838	0.040220	0.040181	no_efectores
4	0.028874	0.038371	0.022587	0.035042	0.053561	0.037716	no_efectores
5	0.035253	0.015665	0.012247	0.002846	0.033915	-0.010309	no_efectores
..	...	...	...	...	...	...	
119	0.034488	0.036122	0.016255	-0.010817	0.025234	-0.000758	no_efectores
120	0.025988	0.026182	0.021499	0.020611	0.042460	0.027572	no_efectores
121	0.031822	0.045734	0.032167	0.022973	0.026398	0.025734	no_efectores
122	0.033829	0.025487	0.019580	0.025304	0.007729	-0.005491	no_efectores
123	0.032158	0.028049	0.029318	0.028187	0.046745	0.044145	no_efectores

[106 rows x 42 columns]

Composición de pseudo aminoácidos (PseAAC) mass no\_efectores Globodera dataset 1, sin valores atípicos.

Estadísticas.

	X0	X1	X2	X3	X4	X5 \
count	106.000000	106.000000	106.000000	106.000000	106.000000	106.000000
mean	0.034333	0.010509	0.019801	0.025407	0.022225	0.042785
std	0.010701	0.006328	0.009752	0.018432	0.017284	0.010763
min	0.012439	0.000000	0.002115	0.000000	0.001269	0.021417
25%	0.023803	0.004075	0.013953	0.005135	0.011593	0.034177
50%	0.036377	0.011536	0.017265	0.024014	0.018733	0.042405
75%	0.040225	0.015984	0.024648	0.042760	0.027865	0.048028
max	0.063422	0.021057	0.060014	0.088821	0.136833	0.069680

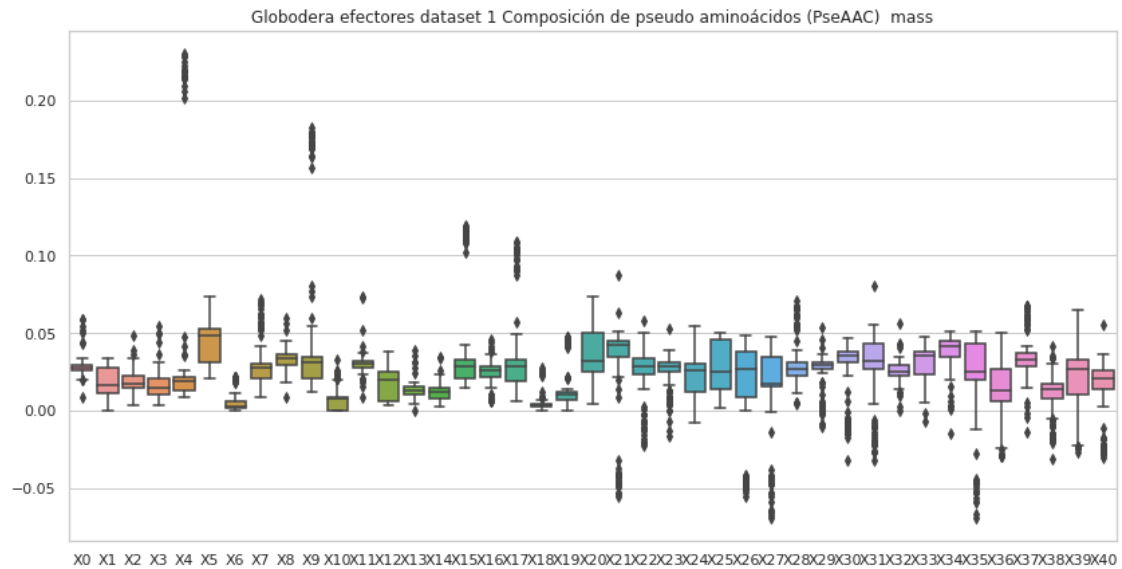
	X6	X7	X8	X9 ...	X31 \
count	106.000000	106.000000	106.000000	106.000000	106.000000
mean	0.010962	0.025530	0.031679	0.037648	0.018489
std	0.006569	0.011503	0.011614	0.019692	0.018785
min	0.000000	0.003384	0.003807	0.003807	-0.010081
25%	0.005748	0.014863	0.023826	0.023537	0.002381
50%	0.011857	0.027829	0.034332	0.034530	0.015295
75%	0.014062	0.032299	0.037848	0.047921	0.038139
max	0.029165	0.063934	0.081619	0.111954	0.054241

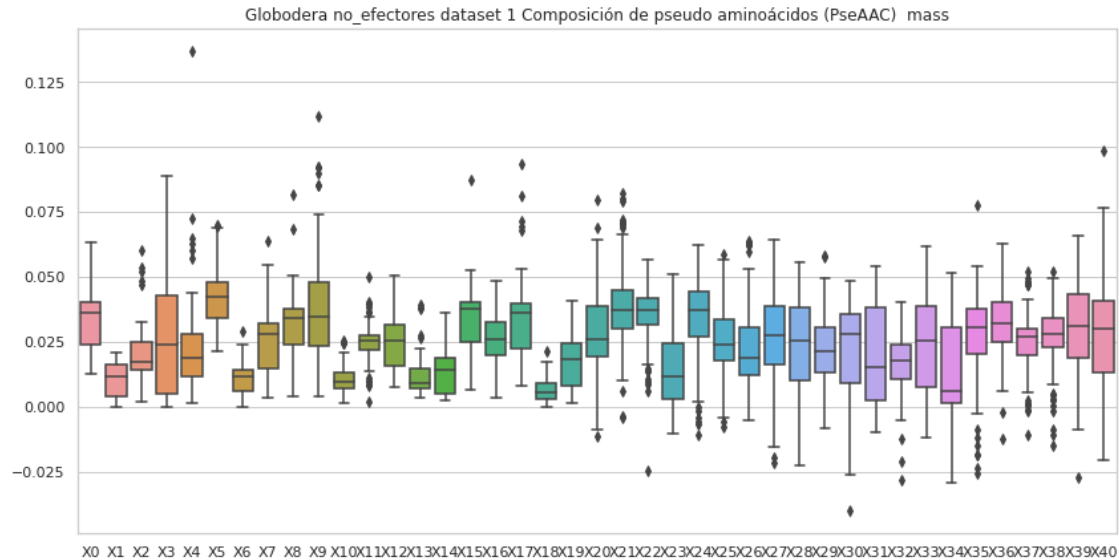
  

	X32	X33	X34	X35	X36	X37 \
count	106.000000	106.000000	106.000000	106.000000	106.000000	106.000000
mean	0.015984	0.024309	0.012586	0.027231	0.032584	0.024891
std	0.011236	0.018246	0.017778	0.016990	0.012215	0.011895
min	-0.028065	-0.012158	-0.029504	-0.025698	-0.012646	-0.010946
25%	0.010421	0.007739	0.001468	0.020254	0.024717	0.019542
50%	0.017523	0.025229	0.006067	0.030465	0.031899	0.027172
75%	0.023650	0.038737	0.030286	0.037512	0.040457	0.029986
max	0.040280	0.061877	0.051675	0.077378	0.062748	0.051880

	X38	X39	X40
count	106.000000	106.000000	106.000000
mean	0.026976	0.030118	0.028912
std	0.012164	0.017999	0.019938
min	-0.015089	-0.027041	-0.020499
25%	0.023074	0.018705	0.013384
50%	0.028011	0.031117	0.030251
75%	0.033896	0.043336	0.040616
max	0.052224	0.065841	0.098389

[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) +",\n"
↳      + 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)+"\n
↪ "+str(transf)+" "+str(comp)+" "+str(estado))
```

efectores

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

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	0.074179	0.081597	0.074179	0.029672	0.037089	0.089015	0.007418
1	0.071803	0.035901	0.039491	0.046672	0.043082	0.107704	0.010770
2	0.005819	0.000000	0.034912	0.025214	0.000000	0.000000	0.000000
3	0.075178	0.037589	0.041348	0.048866	0.048866	0.112767	0.011277
4	0.003162	0.000000	0.002635	0.002635	0.034250	0.008958	0.003162
..	...	...	...	...	...	...	
119	0.072258	0.032115	0.036129	0.036129	0.044158	0.132473	0.004014
120	0.073159	0.044708	0.036580	0.052837	0.052837	0.134125	0.000000
121	0.073073	0.044656	0.036537	0.052775	0.052775	0.133968	0.000000
122	0.078696	0.037474	0.041222	0.048717	0.044969	0.112423	0.011242
123	0.074168	0.037084	0.040792	0.044501	0.048209	0.114961	0.011125

	X7	X8	X9 ...	X53	X54	X55 \
0	0.059343	0.103851	0.096433	... -0.035576	0.030856	0.003311
1	0.068212	0.068212	0.043082	... -0.007605	0.011018	-0.013150
2	0.005819	0.015516	0.005819	... 0.055549	0.000753	0.014484
3	0.071419	0.067660	0.048866	... -0.008260	0.015149	-0.011273
4	0.008958	0.005796	0.025292	... 0.013946	0.022026	0.014363
..	...	...	...	...	...	
119	0.076272	0.084301	0.052186	... -0.029341	0.039545	0.031311
120	0.069095	0.073159	0.065030	... -0.016028	0.010618	-0.033337
121	0.069014	0.073073	0.064954	... -0.015114	0.010526	-0.033136
122	0.071202	0.071202	0.048717	... -0.008322	0.016447	-0.011281
123	0.070460	0.070460	0.048209	... -0.005664	0.021722	-0.009810

	X56	X57	X58	X59	X60	X61	X62
0	0.048487	-0.020161	-0.031488	0.009314	-0.017392	0.001762	efectores
1	0.006495	0.024449	0.023395	0.028489	-0.024960	-0.030721	efectores
2	0.008427	0.060962	-0.002165	0.032382	-0.002860	0.027544	efectores
3	0.001298	0.024185	0.019787	0.027078	-0.021073	-0.037177	efectores
4	0.023031	0.013228	0.028399	0.015030	0.024450	0.012557	efectores
..	...	...	...	...	...	...	
119	0.023406	0.043121	-0.031681	-0.018106	-0.019157	-0.015525	efectores
120	-0.001016	0.014926	-0.002925	-0.002430	-0.045112	-0.049105	efectores
121	-0.000481	0.015372	-0.003557	-0.004004	-0.044846	-0.048669	efectores

```

122  0.005365  0.024051  0.019637  0.027160 -0.019270 -0.036934  efectores
123  0.010783  0.028263  0.019852  0.023906 -0.025363 -0.040860  efectores

```

[124 rows x 63 columns]

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

	X0	X1	X2	X3	X4	X5 \
count	124.000000	124.000000	124.000000	124.000000	124.000000	124.000000
mean	0.059487	0.041346	0.039413	0.033561	0.038673	0.086989
std	0.029879	0.029934	0.021884	0.017769	0.023233	0.047277
min	-0.077255	-0.038628	-0.000000	-0.000000	-0.193138	-0.000000
25%	0.058138	0.023169	0.032919	0.026897	0.034144	0.044758
50%	0.073078	0.037342	0.037453	0.033242	0.039092	0.093480
75%	0.076549	0.074794	0.060183	0.048754	0.047858	0.131490
max	0.101470	0.094737	0.074179	0.082679	0.062009	0.155996

	X6	X7	X8	X9 ...	X52 \
count	124.000000	124.000000	124.000000	124.000000	124.000000
mean	0.006567	0.052872	0.068834	0.061917	-0.015224
std	0.005374	0.025993	0.033740	0.031042	0.032617
min	0.000000	-0.038628	-0.000000	-0.115883	-0.077264
25%	0.003196	0.032488	0.060841	0.044501	-0.048306
50%	0.006724	0.061766	0.072867	0.063349	-0.010667
75%	0.008029	0.069970	0.097908	0.090915	0.009028
max	0.027560	0.112664	0.112582	0.122959	0.110009

	X53	X54	X55	X56	X57	X58 \
count	124.000000	124.000000	124.000000	124.000000	124.000000	124.000000
mean	-0.013136	0.024034	0.002122	0.025031	0.010086	-0.004201
std	0.023947	0.079741	0.042572	0.060697	0.046073	0.030457
min	-0.068623	-0.092025	-0.070208	-0.118296	-0.086199	-0.072616
25%	-0.031204	0.009889	-0.011613	0.004906	-0.020927	-0.028202
50%	-0.015780	0.022158	0.001795	0.022244	0.013144	-0.006486
75%	-0.004049	0.032008	0.013925	0.043439	0.024202	0.021762
max	0.115722	0.866156	0.392351	0.632954	0.415405	0.103276

	X59	X60	X61
count	124.000000	124.000000	124.000000
mean	0.007167	-0.014956	-0.009376
std	0.024901	0.034722	0.032686
min	-0.064318	-0.058018	-0.075406
25%	-0.004023	-0.033947	-0.034765
50%	0.008156	-0.022269	-0.009301
75%	0.015366	-0.006512	0.008859

max            0.192669      0.271852      0.203639

[8 rows x 62 columns]

no\_efectores

Composición de pseudo aminoácidos (PseAAC) hidro no\_efectores Globodera dataset  
1, con valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	0.055020	0.027510	0.041265	0.096284	0.068774	0.165059	0.041265
1	0.067101	0.033550	0.046132	0.092264	0.041938	0.083876	0.025163
2	0.079840	0.029415	0.046223	0.096648	0.058829	0.088244	0.025213
3	0.074435	0.023506	0.047011	0.082270	0.054847	0.082270	0.023506
4	0.069397	0.032657	0.044904	0.085725	0.053068	0.093890	0.016329
..	...	...	...	...	...	...	...
119	0.002802	0.001681	0.002802	0.002802	0.030822	0.011208	0.003923
120	0.090328	0.011709	0.021746	0.015055	0.023418	0.043491	0.018400
121	0.072859	0.040477	0.056668	0.012143	0.032382	0.113337	0.004048
122	0.094954	0.005275	0.021101	0.036927	0.031651	0.068578	0.026376
123	0.081460	0.025724	0.051448	0.090035	0.055736	0.090035	0.025724

	X7	X8	X9 ...	X53	X54	X55 \
0	0.096284	0.137549	0.096284 ...	0.017457	-0.004688	0.040397
1	0.050326	0.075489	0.079682 ...	-0.039573	-0.006015	-0.020445
2	0.058829	0.071436	0.096648 ...	-0.065221	0.001544	0.000680
3	0.054847	0.078352	0.086187 ...	-0.035917	-0.012753	-0.031695
4	0.057150	0.073479	0.097972 ...	-0.021833	-0.039997	-0.047481
..	...	...	...	...	...	...
119	0.007285	0.002242	0.025218 ...	0.014407	0.021772	0.011655
120	0.030109	0.055201	0.048510 ...	0.010504	-0.001185	0.004722
121	0.060716	0.056668	0.044525 ...	-0.008121	0.027810	0.039956
122	0.021101	0.063303	0.068578 ...	-0.046655	0.027644	0.003512
123	0.060023	0.081460	0.102897 ...	-0.050393	-0.013461	-0.025680

	X56	X57	X58	X59	X60	X61	X62
0	-0.087778	-0.161714	0.069918	0.014458	0.039200	-0.079260	no_efectores
1	-0.028402	-0.018250	-0.039929	-0.042892	0.021088	0.032068	no_efectores
2	-0.041500	-0.024366	-0.032422	-0.020083	-0.004045	-0.021882	no_efectores
3	-0.023187	-0.020810	-0.039199	-0.038296	-0.008694	0.022172	no_efectores
4	-0.003459	-0.019437	-0.064412	-0.049635	0.000348	0.024762	no_efectores
..	...	...	...	...	...	...	...
119	0.024025	0.013256	0.021524	0.011746	0.027828	0.012833	no_efectores
120	0.005458	-0.003461	0.019356	0.005477	-0.020190	-0.002878	no_efectores
121	0.026940	-0.026202	-0.018463	-0.009930	-0.020930	0.007520	no_efectores
122	0.024111	0.014625	-0.009970	-0.000660	-0.015609	-0.007234	no_efectores



123 -0.035193 -0.041860 -0.033395 -0.032600 0.000172 0.012486 no\_efectores

[124 rows x 63 columns]

Composición de pseudo aminoácidos (PseAAC) hidro no\_efectores Globodera dataset  
1, con valores atípicos.  
Estadísticas.

	X0	X1	X2	X3	X4	X5 \
count	124.000000	124.000000	124.000000	124.000000	124.000000	124.000000
mean	0.059613	0.021346	0.035968	0.043958	0.034941	0.078324
std	0.029604	0.017993	0.020938	0.033299	0.018613	0.046644
min	0.000781	0.000000	0.001562	0.000000	0.006356	0.007812
25%	0.039362	0.002707	0.019076	0.015880	0.021896	0.052265
50%	0.062577	0.023446	0.042772	0.033984	0.031157	0.076814
75%	0.077104	0.032931	0.049440	0.082672	0.049667	0.094045
max	0.179594	0.085967	0.134695	0.115180	0.112246	0.336738

	X6	X7	X8	X9 ...	X52 \
count	124.000000	124.000000	124.000000	124.000000	124.000000
mean	0.017110	0.045436	0.057955	0.059266	0.003482
std	0.010639	0.029687	0.030096	0.028040	0.023325
min	0.000000	0.005220	0.002242	0.010698	-0.059271
25%	0.007649	0.016203	0.040007	0.038764	-0.014473
50%	0.016963	0.053261	0.061690	0.054402	0.003522
75%	0.024009	0.060778	0.072330	0.085601	0.020398
max	0.044898	0.202043	0.224492	0.134695	0.071555

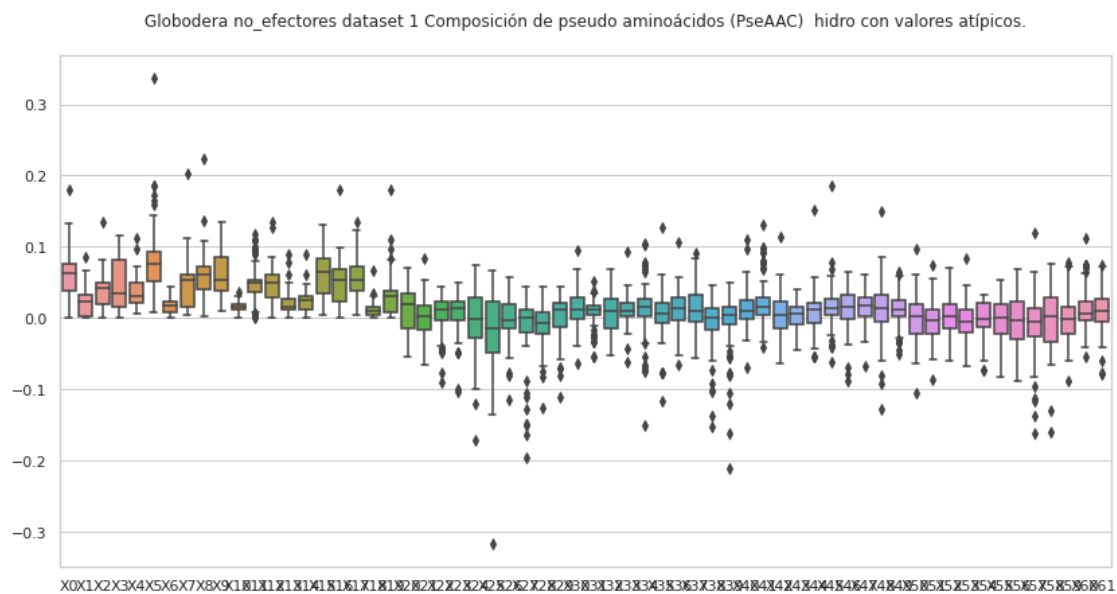
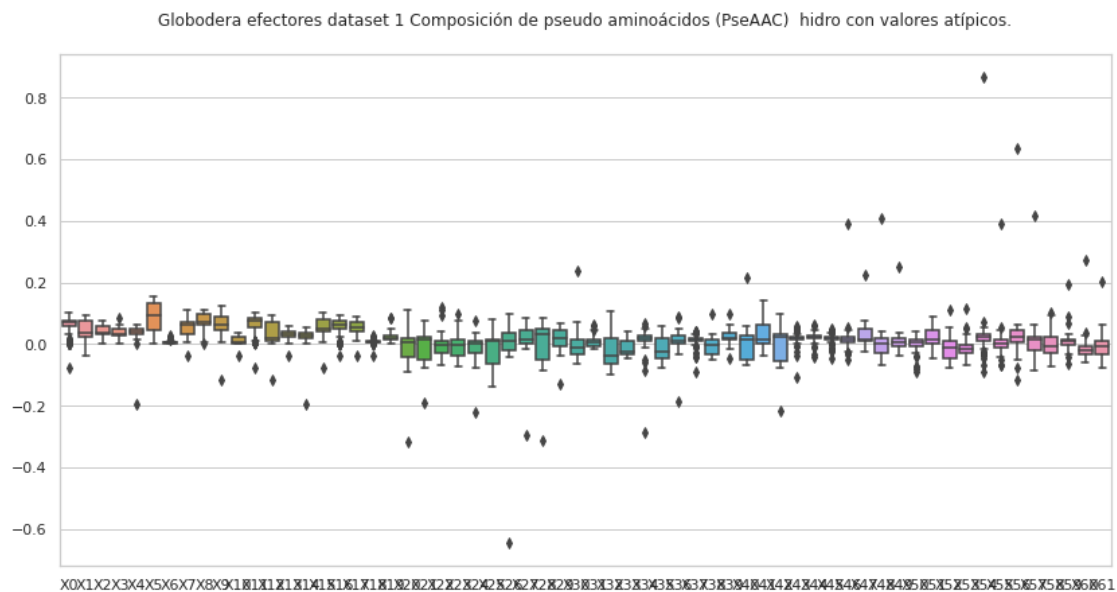
  

	X53	X54	X55	X56	X57	X58 \
count	124.000000	124.000000	124.000000	124.000000	124.000000	124.000000
mean	-0.007265	0.001213	-0.001362	-0.003269	-0.007171	-0.002048
std	0.024893	0.023696	0.028290	0.031273	0.038651	0.039723
min	-0.067107	-0.072738	-0.082686	-0.087778	-0.161714	-0.160687
25%	-0.020453	-0.012613	-0.022539	-0.028578	-0.025908	-0.033396
50%	-0.004643	-0.000252	0.000559	-0.003228	-0.004132	0.002492
75%	0.011136	0.022107	0.020148	0.023150	0.014550	0.028527
max	0.084010	0.033716	0.053776	0.068371	0.119290	0.075608

	X59	X60	X61
count	124.000000	124.000000	124.000000
mean	-0.000742	0.012579	0.009773
std	0.031761	0.026817	0.025960
min	-0.088543	-0.054587	-0.079260
25%	-0.021441	-0.003224	-0.004769
50%	-0.001362	0.006447	0.009401
75%	0.016179	0.023682	0.026489
max	0.078534	0.112963	0.073758

[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)+"\n
↳"+str(transf)+" "+str(comp))
```

efectores

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

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	0.074179	0.081597	0.074179	0.029672	0.037089	0.089015	0.007418
1	0.071803	0.035901	0.039491	0.046672	0.043082	0.107704	0.010770
3	0.075178	0.037589	0.041348	0.048866	0.048866	0.112767	0.011277
4	0.003162	0.000000	0.002635	0.002635	0.034250	0.008958	0.003162
5	0.070758	0.045785	0.037460	0.054109	0.054109	0.133192	0.004162
..	...	...	...	...	...	...	...
119	0.072258	0.032115	0.036129	0.036129	0.044158	0.132473	0.004014
120	0.073159	0.044708	0.036580	0.052837	0.052837	0.134125	0.000000
121	0.073073	0.044656	0.036537	0.052775	0.052775	0.133968	0.000000
122	0.078696	0.037474	0.041222	0.048717	0.044969	0.112423	0.011242
123	0.074168	0.037084	0.040792	0.044501	0.048209	0.114961	0.011125

	X7	X8	X9	...	X53	X54	X55 \
0	0.059343	0.103851	0.096433	...	-0.035576	0.030856	0.003311
1	0.068212	0.068212	0.043082	...	-0.007605	0.011018	-0.013150
3	0.071419	0.067660	0.048866	...	-0.008260	0.015149	-0.011273
4	0.008958	0.005796	0.025292	...	0.013946	0.022026	0.014363
5	0.070758	0.070758	0.066596	...	-0.014472	0.006744	-0.037797
..	...	...	...	...	...	...	...
119	0.076272	0.084301	0.052186	...	-0.029341	0.039545	0.031311
120	0.069095	0.073159	0.065030	...	-0.016028	0.010618	-0.033337
121	0.069014	0.073073	0.064954	...	-0.015114	0.010526	-0.033136
122	0.071202	0.071202	0.048717	...	-0.008322	0.016447	-0.011281
123	0.070460	0.070460	0.048209	...	-0.005664	0.021722	-0.009810

	X56	X57	X58	X59	X60	X61	X62
0	0.048487	-0.020161	-0.031488	0.009314	-0.017392	0.001762	efectores
1	0.006495	0.024449	0.023395	0.028489	-0.024960	-0.030721	efectores
3	0.001298	0.024185	0.019787	0.027078	-0.021073	-0.037177	efectores
4	0.023031	0.013228	0.028399	0.015030	0.024450	0.012557	efectores
5	0.006660	0.014870	-0.005027	-0.004187	-0.045723	-0.045596	efectores
..	...	...	...	...	...	...	...
119	0.023406	0.043121	-0.031681	-0.018106	-0.019157	-0.015525	efectores
120	-0.001016	0.014926	-0.002925	-0.002430	-0.045112	-0.049105	efectores
121	-0.000481	0.015372	-0.003557	-0.004004	-0.044846	-0.048669	efectores
122	0.005365	0.024051	0.019637	0.027160	-0.019270	-0.036934	efectores
123	0.010783	0.028263	0.019852	0.023906	-0.025363	-0.040860	efectores

[113 rows x 63 columns]

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

	X0	X1	X2	X3	X4	X5 \
count	113.000000	113.000000	113.000000	113.000000	113.000000	113.000000
mean	0.060966	0.042807	0.039991	0.033744	0.041338	0.087882
std	0.027145	0.029005	0.021735	0.016995	0.008699	0.046077
min	0.003153	0.000000	0.002131	0.002083	0.013930	0.008427
25%	0.066109	0.024073	0.034166	0.028009	0.034953	0.047131
50%	0.073118	0.037357	0.037460	0.034450	0.039951	0.094113
75%	0.076512	0.076896	0.060686	0.048717	0.048209	0.131162
max	0.087167	0.094737	0.074179	0.063456	0.054520	0.141560

	X6	X7	X8	X9 ...	X52 \
count	113.000000	113.000000	113.000000	113.000000	113.000000
mean	0.006114	0.053279	0.069602	0.063120	-0.015840
std	0.004656	0.023576	0.033456	0.025896	0.029801
min	0.000000	0.007459	0.005728	0.021949	-0.063506
25%	0.003196	0.032499	0.064998	0.045168	-0.047956
50%	0.005068	0.061782	0.073083	0.063404	-0.010953
75%	0.007791	0.069854	0.098031	0.091028	0.008817
max	0.022150	0.085414	0.112582	0.111962	0.034207

	X53	X54	X55	X56	X57	X58 \
count	113.000000	113.000000	113.000000	113.000000	113.000000	113.000000
mean	-0.014804	0.019752	-0.000057	0.023419	0.007432	-0.006210
std	0.017936	0.018067	0.021414	0.020710	0.026273	0.025455
min	-0.054740	-0.034627	-0.050239	-0.049907	-0.034798	-0.062870
25%	-0.030962	0.010595	-0.011273	0.006328	-0.021181	-0.028419
50%	-0.015794	0.022374	0.001916	0.022971	0.013152	-0.006950
75%	-0.005664	0.031998	0.013839	0.044043	0.024095	0.020733
max	0.027837	0.071216	0.045700	0.062528	0.054235	0.046931

	X59	X60	X61
count	113.000000	113.000000	113.000000
mean	0.005836	-0.017449	-0.010961
std	0.013422	0.022829	0.025373
min	-0.030037	-0.058018	-0.056673
25%	-0.003857	-0.030991	-0.036888
50%	0.008442	-0.022387	-0.007304
75%	0.015321	-0.013425	0.008200
max	0.032239	0.034881	0.061067

[8 rows x 62 columns]

no\_efectores

Composición de pseudo aminoácidos (PseAAC) no\_efectores Globodera dataset 1,  
sin valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
1	0.067101	0.033550	0.046132	0.092264	0.041938	0.083876	0.025163
2	0.079840	0.029415	0.046223	0.096648	0.058829	0.088244	0.025213
3	0.074435	0.023506	0.047011	0.082270	0.054847	0.082270	0.023506
4	0.069397	0.032657	0.044904	0.085725	0.053068	0.093890	0.016329
5	0.011794	0.001966	0.016381	0.024244	0.008518	0.008518	0.003276
..	...	...	...	...	...	...	
119	0.002802	0.001681	0.002802	0.002802	0.030822	0.011208	0.003923
120	0.090328	0.011709	0.021746	0.015055	0.023418	0.043491	0.018400
121	0.072859	0.040477	0.056668	0.012143	0.032382	0.113337	0.004048
122	0.094954	0.005275	0.021101	0.036927	0.031651	0.068578	0.026376
123	0.081460	0.025724	0.051448	0.090035	0.055736	0.090035	0.025724
	X7	X8	X9	...	X53	X54	X55 \
1	0.050326	0.075489	0.079682	...	-0.039573	-0.006015	-0.020445
2	0.058829	0.071436	0.096648	...	-0.065221	0.001544	0.000680
3	0.054847	0.078352	0.086187	...	-0.035917	-0.012753	-0.031695
4	0.057150	0.073479	0.097972	...	-0.021833	-0.039997	-0.047481
5	0.011139	0.022278	0.011139	...	0.032146	0.008606	0.022057
..	...	...	...	...	...	...	
119	0.007285	0.002242	0.025218	...	0.014407	0.021772	0.011655
120	0.030109	0.055201	0.048510	...	0.010504	-0.001185	0.004722
121	0.060716	0.056668	0.044525	...	-0.008121	0.027810	0.039956
122	0.021101	0.063303	0.068578	...	-0.046655	0.027644	0.003512
123	0.060023	0.081460	0.102897	...	-0.050393	-0.013461	-0.025680
	X56	X57	X58	X59	X60	X61	X62
1	-0.028402	-0.018250	-0.039929	-0.042892	0.021088	0.032068	no_efectores
2	-0.041500	-0.024366	-0.032422	-0.020083	-0.004045	-0.021882	no_efectores
3	-0.023187	-0.020810	-0.039199	-0.038296	-0.008694	0.022172	no_efectores
4	-0.003459	-0.019437	-0.064412	-0.049635	0.000348	0.024762	no_efectores
5	0.005665	0.036385	0.014852	0.039385	0.002861	0.025852	no_efectores
..	...	...	...	...	...	...	
119	0.024025	0.013256	0.021524	0.011746	0.027828	0.012833	no_efectores
120	0.005458	-0.003461	0.019356	0.005477	-0.020190	-0.002878	no_efectores
121	0.026940	-0.026202	-0.018463	-0.009930	-0.020930	0.007520	no_efectores
122	0.024111	0.014625	-0.009970	-0.000660	-0.015609	-0.007234	no_efectores
123	-0.035193	-0.041860	-0.033395	-0.032600	0.000172	0.012486	no_efectores

[110 rows x 63 columns]

Composición de pseudo aminoácidos (PseAAC) no\_efectores Globodera dataset 1,  
sin valores atípicos.

Estadísticas.

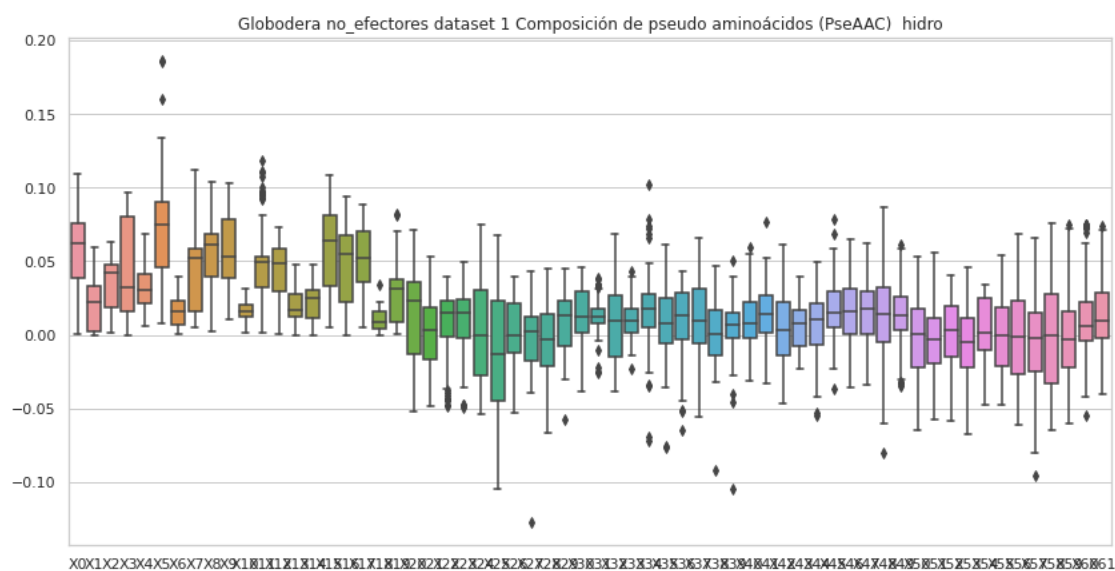
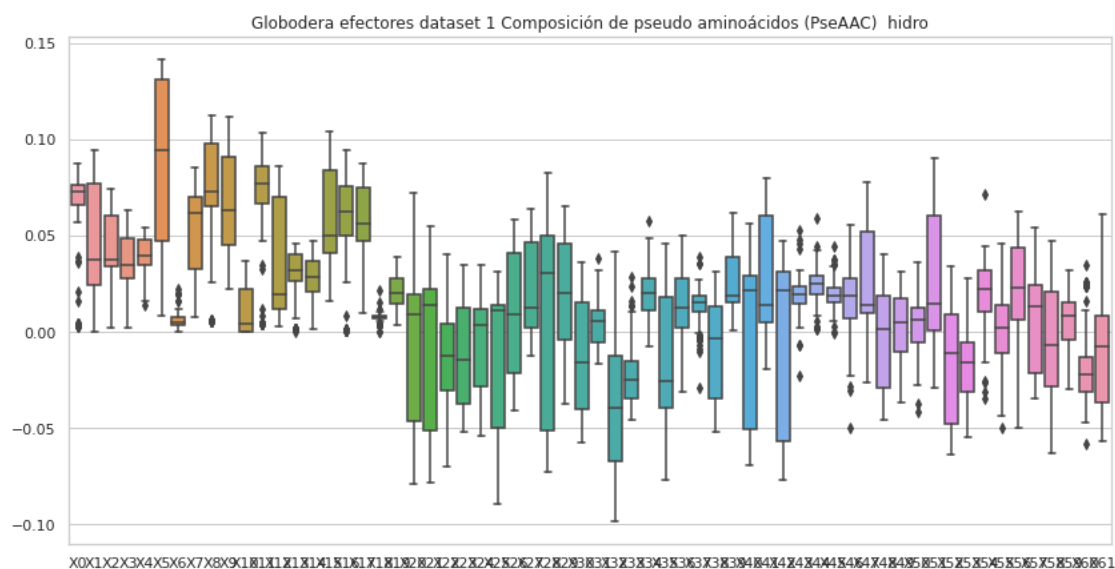
	X0	X1	X2	X3	X4	X5 \
count	110.000000	110.000000	110.000000	110.000000	110.000000	110.000000
mean	0.056711	0.020384	0.033689	0.040455	0.032444	0.071285
std	0.026872	0.016935	0.018559	0.031986	0.015264	0.037578
min	0.000781	0.000000	0.001562	0.000000	0.006356	0.007812
25%	0.038571	0.002648	0.018540	0.015529	0.021133	0.046027
50%	0.061731	0.022387	0.042459	0.032230	0.030712	0.074487
75%	0.075605	0.033004	0.047592	0.079916	0.041781	0.090388
max	0.108969	0.059600	0.062957	0.096648	0.068957	0.186250

	X6	X7	X8	X9 ...	X52 \
count	110.000000	110.000000	110.000000	110.000000 ...	110.000000
mean	0.015669	0.042136	0.053809	0.056567 ...	0.002657
std	0.009159	0.024658	0.023835	0.025732 ...	0.020068
min	0.000781	0.005220	0.002242	0.010698 ...	-0.058630
25%	0.006703	0.015830	0.039755	0.038211 ...	-0.014427
50%	0.015550	0.052622	0.061055	0.053254 ...	0.003023
75%	0.023476	0.058790	0.068777	0.078384 ...	0.019739
max	0.039129	0.111750	0.104300	0.102897 ...	0.040565

	X53	X54	X55	X56	X57	X58 \
count	110.000000	110.000000	110.000000	110.000000	110.000000	110.000000
mean	-0.008173	0.004021	-0.000417	-0.001153	-0.002713	-0.002440
std	0.024221	0.020827	0.026319	0.028495	0.028549	0.033655
min	-0.067107	-0.047210	-0.047481	-0.061153	-0.095807	-0.064412
25%	-0.022408	-0.010578	-0.021572	-0.027033	-0.024423	-0.033392
50%	-0.004978	0.001445	-0.000428	-0.001357	-0.002588	-0.000146
75%	0.011367	0.024625	0.019039	0.023091	0.014608	0.027662
max	0.045566	0.033716	0.053776	0.068371	0.065604	0.075608

	X59	X60	X61
count	110.000000	110.000000	110.000000
mean	-0.001288	0.011017	0.011851
std	0.030809	0.025045	0.022870
min	-0.060357	-0.054587	-0.040618
25%	-0.022027	-0.003804	-0.001718
50%	-0.003380	0.005716	0.010054
75%	0.016364	0.022755	0.028332
max	0.075242	0.075051	0.073758

[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 Globodera dataset 1,  
con valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	-0.010549	0.079950	-0.021663	-0.003517	0.010475	0.104994	-0.084835
1	0.087442	0.074751	-0.020555	0.011417	-0.089592	-0.080496	-0.010581
2	-0.064158	0.042774	-0.023861	0.035360	0.055761	-0.158534	0.105877
3	0.093568	0.091161	-0.014655	0.008886	-0.107150	-0.090436	0.001904
4	0.016486	-0.080949	-0.004929	0.000070	0.065487	0.070790	-0.087497
..	...	...	...	...	...	...	...
119	0.062843	0.007681	0.012457	-0.019704	-0.083203	-0.066870	0.069773
120	0.108285	0.051721	0.008466	-0.012157	-0.102047	-0.108167	0.034955
121	0.106425	0.052584	0.009448	-0.015887	-0.097062	-0.107316	0.036854
122	0.093445	0.069846	-0.010262	0.001120	-0.098703	-0.084250	0.000165
123	0.088304	0.061832	0.016245	-0.003739	-0.105054	-0.086628	0.014074
	X7	X8	X9	X10	X11	X12	X13
0	0.042210	0.001973	0.016051	0.049825	0.005194	0.015993	efectores

```

1  -0.063674 -0.000746  0.001506  0.029297  0.051266 -0.047259  efectores
2  -0.112671 -0.059166 -0.037901 -0.138312  0.062583 -0.146305  efectores
3  -0.071321 -0.020210  0.001595  0.017997  0.057693 -0.041386  efectores
4  -0.088773  0.105527 -0.017113 -0.028421 -0.052765  0.048569  efectores
..      ...      ...      ...      ...      ...
119 -0.101180 -0.101671  0.008942  0.020309  0.053943 -0.046297  efectores
120 -0.067454 -0.046498 -0.011636  0.050587  0.087886 -0.041194  efectores
121 -0.070281 -0.044722 -0.015348  0.046718  0.085861 -0.041362  efectores
122 -0.072415 -0.027387  0.002758  0.017747  0.056809 -0.033485  efectores
123 -0.080820 -0.025647  0.007020  0.028570  0.053436 -0.041788  efectores

```

[124 rows x 14 columns]

Covarianza de auto cruzamiento (ACC) hidro\_mass efectores Globodera dataset 1,  
con valores atípicos.

Estadísticas.

	X0	X1	X2	X3	X4	X5 \
count	124.000000	124.000000	124.000000	124.000000	124.000000	124.000000
mean	0.041515	0.031174	-0.004925	-0.000276	-0.020619	-0.002568
std	0.066352	0.087484	0.042344	0.061259	0.079160	0.090881
min	-0.117886	-0.503538	-0.090270	-0.187075	-0.343665	-0.256220
25%	-0.008009	0.004066	-0.023784	-0.015472	-0.094270	-0.087966
50%	0.029163	0.055562	-0.010589	-0.003000	-0.003425	0.006121
75%	0.091795	0.080227	0.008164	0.007965	0.023904	0.092380
max	0.464974	0.316647	0.193164	0.440767	0.157839	0.131299

	X6	X7	X8	X9	X10	X11 \
count	124.000000	124.000000	124.000000	124.000000	124.000000	124.000000
mean	-0.014944	-0.036240	-0.003822	0.008564	0.022343	0.029879
std	0.075500	0.069502	0.077602	0.031524	0.056190	0.061086
min	-0.110454	-0.167294	-0.415399	-0.072703	-0.138312	-0.187073
25%	-0.084883	-0.088542	-0.043824	-0.008427	-0.006318	-0.007271
50%	-0.005363	-0.067235	-0.004181	0.007182	0.027068	0.020111
75%	0.034960	0.037404	0.023542	0.016824	0.049699	0.062992
max	0.387758	0.161485	0.161947	0.197792	0.416201	0.304052

	X12
count	124.000000
mean	-0.006512
std	0.050917
min	-0.247599
25%	-0.041512
50%	-0.012433
75%	0.022073
max	0.165858

no\_efectores

Covarianza de auto cruzamiento (ACC) hidro\_mass no\_efectores Globodera dataset  
1, con valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	-0.007934	0.149593	0.018527	0.008338	0.103889	-0.083495	-0.089614
1	-0.019927	0.051389	0.085881	-0.049360	0.046373	0.024052	-0.061740
2	0.011830	0.047089	0.086867	-0.089937	0.067371	0.008432	-0.067699
3	-0.002923	0.060430	0.069633	-0.105039	0.069781	-0.004421	-0.033805
4	-0.000843	0.093577	0.057585	-0.069250	0.025734	-0.001145	-0.040913
..	...	...	...	...	...	...	
119	0.013204	0.104309	0.094787	0.055603	0.038557	0.015047	0.077456
120	0.039099	0.043180	0.119983	-0.017972	0.139362	0.055141	0.046884
121	0.118540	0.007486	0.032157	-0.056051	0.107429	0.159474	-0.070768
122	0.149094	0.038679	0.049007	-0.040010	-0.015238	0.071576	-0.019588
123	0.001224	0.067488	0.064445	-0.104767	0.071490	-0.016061	-0.049272

	X7	X8	X9	X10	X11	X12	X13
0	-0.073022	-0.177967	-0.046539	-0.176807	-0.093743	-0.019910	no_efectores
1	-0.019324	-0.051417	-0.033550	-0.055260	-0.065876	-0.047493	no_efectores
2	0.019694	-0.043341	-0.006750	-0.049723	-0.102480	-0.057332	no_efectores
3	0.026946	-0.051479	0.002180	-0.047125	-0.077480	-0.021516	no_efectores
4	-0.000748	-0.052139	-0.028762	-0.062232	-0.069021	-0.047349	no_efectores
..	...	...	...	...	...	...	
119	0.057797	0.030293	-0.030493	0.010036	0.001627	0.066129	no_efectores
120	0.088372	0.098059	0.052733	0.014243	0.023640	0.062680	no_efectores
121	0.040778	-0.012200	-0.037553	0.045189	0.110732	-0.026798	no_efectores
122	0.204239	0.156980	0.066204	0.062484	-0.087189	-0.022955	no_efectores
123	0.057831	-0.030673	-0.011330	-0.036026	-0.079360	-0.040350	no_efectores

[124 rows x 14 columns]

Covarianza de auto cruzamiento (ACC) hidro\_mass no\_efectores Globodera dataset  
1, con valores atípicos.

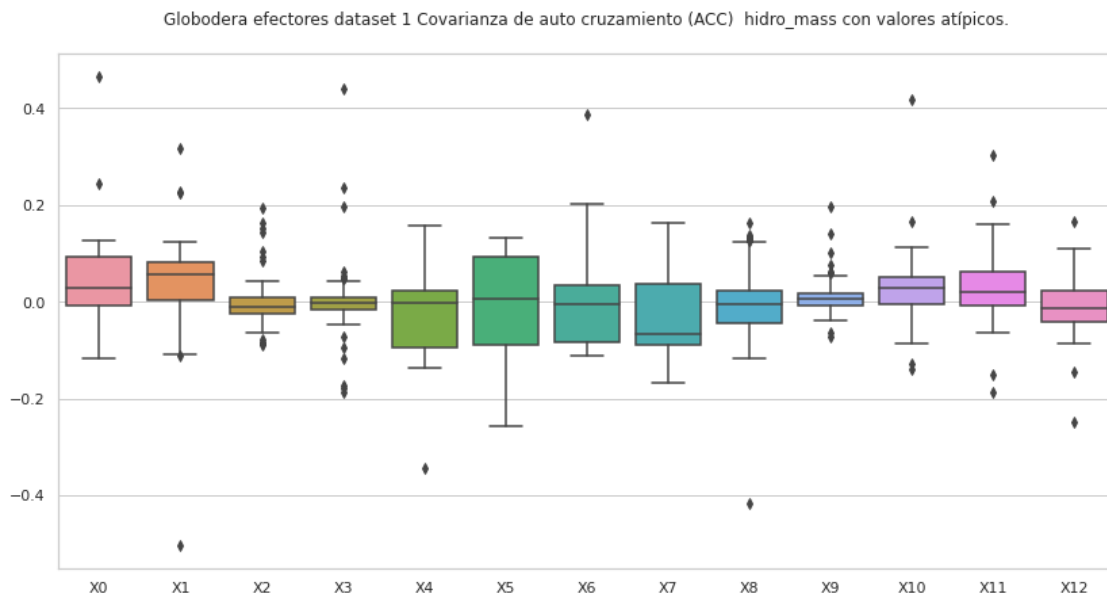
Estadísticas.

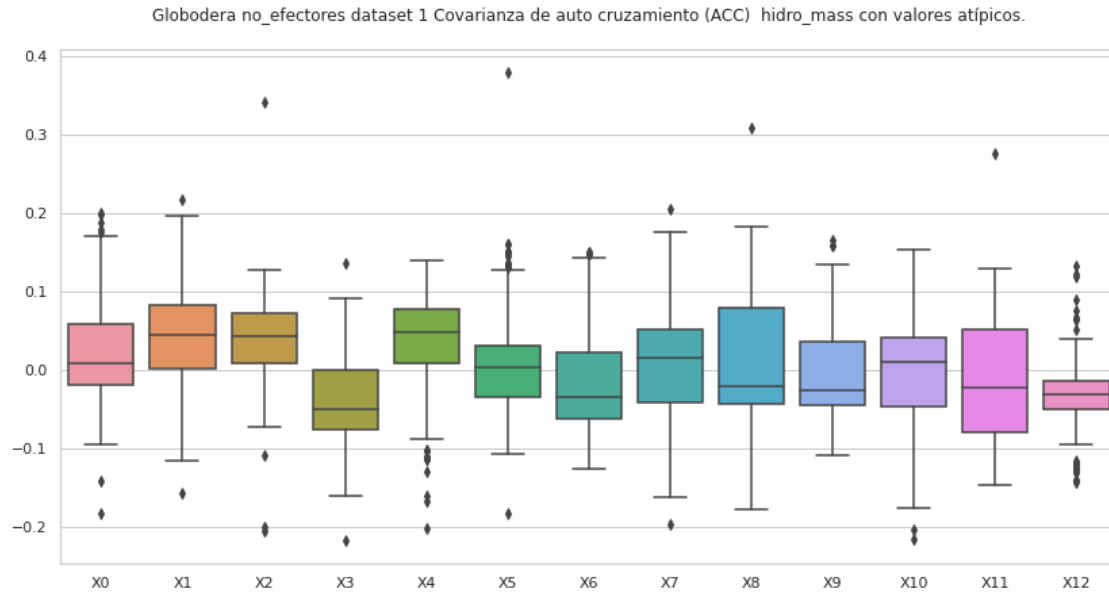
	X0	X1	X2	X3	X4	X5 \
count	124.000000	124.000000	124.000000	124.000000	124.000000	124.000000
mean	0.023332	0.046149	0.033501	-0.044288	0.030525	0.009802
std	0.072712	0.069315	0.062403	0.063106	0.070670	0.072309
min	-0.184051	-0.157437	-0.205958	-0.218116	-0.203192	-0.184146
25%	-0.020036	0.000962	0.007560	-0.077390	0.008463	-0.035525
50%	0.007349	0.044149	0.042522	-0.050286	0.047602	0.002650
75%	0.057589	0.081860	0.071791	-0.000359	0.076731	0.030136

max	0.199982	0.216889	0.340880	0.135031	0.139362	0.378927
-----	----------	----------	----------	----------	----------	----------

	X6	X7	X8	X9	X10	X11 \
count	124.000000	124.000000	124.000000	124.000000	124.000000	124.000000
mean	-0.018571	0.005676	0.013756	-0.004901	-0.005273	-0.014209
std	0.066485	0.079465	0.084746	0.058742	0.068531	0.074776
min	-0.126615	-0.197375	-0.177967	-0.110068	-0.216828	-0.147431
25%	-0.061940	-0.042288	-0.043956	-0.045601	-0.047512	-0.079930
50%	-0.036001	0.014592	-0.021770	-0.026120	0.010224	-0.023741
75%	0.021615	0.051857	0.078129	0.034818	0.040976	0.051555
max	0.150148	0.204239	0.308672	0.165127	0.152627	0.276172

	X12
count	124.000000
mean	-0.030696
std	0.051789
min	-0.143409
25%	-0.051404
50%	-0.032277
75%	-0.013778
max	0.132850





## 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) +",\n"
      '↪' + 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)+"\n
→"+str(transf)+" "+str(comp))

```

efectores

Covarianza de auto cruzamiento (ACC) hidro\_mass efectores Globodera dataset 1,  
sin valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	-0.010549	0.079950	-0.021663	-0.003517	0.010475	0.104994	-0.084835
1	0.087442	0.074751	-0.020555	0.011417	-0.089592	-0.080496	-0.010581
2	-0.064158	0.042774	-0.023861	0.035360	0.055761	-0.158534	0.105877
3	0.093568	0.091161	-0.014655	0.008886	-0.107150	-0.090436	0.001904
4	0.016486	-0.080949	-0.004929	0.000070	0.065487	0.070790	-0.087497
..	...	...	...	...	...	...	
119	0.062843	0.007681	0.012457	-0.019704	-0.083203	-0.066870	0.069773
120	0.108285	0.051721	0.008466	-0.012157	-0.102047	-0.108167	0.034955
121	0.106425	0.052584	0.009448	-0.015887	-0.097062	-0.107316	0.036854
122	0.093445	0.069846	-0.010262	0.001120	-0.098703	-0.084250	0.000165
123	0.088304	0.061832	0.016245	-0.003739	-0.105054	-0.086628	0.014074
	X7	X8	X9	X10	X11	X12	X13
0	0.042210	0.001973	0.016051	0.049825	0.005194	0.015993	efectores
1	-0.063674	-0.000746	0.001506	0.029297	0.051266	-0.047259	efectores
2	-0.112671	-0.059166	-0.037901	-0.138312	0.062583	-0.146305	efectores
3	-0.071321	-0.020210	0.001595	0.017997	0.057693	-0.041386	efectores
4	-0.088773	0.105527	-0.017113	-0.028421	-0.052765	0.048569	efectores
..	...	...	...	...	...	...	
119	-0.101180	-0.101671	0.008942	0.020309	0.053943	-0.046297	efectores

```

120 -0.067454 -0.046498 -0.011636 0.050587 0.087886 -0.041194 efectores
121 -0.070281 -0.044722 -0.015348 0.046718 0.085861 -0.041362 efectores
122 -0.072415 -0.027387 0.002758 0.017747 0.056809 -0.033485 efectores
123 -0.080820 -0.025647 0.007020 0.028570 0.053436 -0.041788 efectores

```

[116 rows x 14 columns]

Covarianza de auto cruzamiento (ACC) hidro\_mass efectores Globodera dataset 1,  
sin valores atípicos.  
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	116.000000	116.000000	116.000000	116.000000	116.000000	116.000000	
mean	0.039879	0.031257	-0.010420	-0.005554	-0.021974	-0.001244	
std	0.048747	0.067097	0.029619	0.032398	0.072381	0.088629	
min	-0.064158	-0.110488	-0.090270	-0.178040	-0.137897	-0.158534	
25%	-0.007525	0.004066	-0.023912	-0.014410	-0.094489	-0.087966	
50%	0.031747	0.055562	-0.012522	-0.003080	-0.005607	0.006121	
75%	0.091795	0.080067	0.007764	0.006391	0.019932	0.092380	
max	0.125288	0.226622	0.104776	0.062859	0.122434	0.121527	

	X6	X7	X8	X9	X10	X11	\
count	116.000000	116.000000	116.000000	116.000000	116.000000	116.000000	
mean	-0.020457	-0.040766	-0.002921	0.004392	0.017964	0.030494	
std	0.064297	0.064642	0.067581	0.020139	0.039666	0.048172	
min	-0.110454	-0.155074	-0.117263	-0.072703	-0.138312	-0.063735	
25%	-0.085258	-0.088773	-0.043607	-0.008756	-0.006051	-0.006533	
50%	-0.005363	-0.068393	-0.005285	0.005140	0.027068	0.031493	
75%	0.034615	0.033811	0.018951	0.015887	0.049440	0.062992	
max	0.105877	0.098034	0.136839	0.061478	0.076060	0.158915	

	X12
count	116.000000
mean	-0.006972
std	0.041708
min	-0.146305
25%	-0.041395
50%	-0.010771
75%	0.021568
max	0.072321

no\_efectores

Covarianza de auto cruzamiento (ACC) hidro\_mass no\_efectores Globodera dataset  
1, sin valores atípicos.  
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	-0.007934	0.149593	0.018527	0.008338	0.103889	-0.083495	-0.089614
1	-0.019927	0.051389	0.085881	-0.049360	0.046373	0.024052	-0.061740
2	0.011830	0.047089	0.086867	-0.089937	0.067371	0.008432	-0.067699
3	-0.002923	0.060430	0.069633	-0.105039	0.069781	-0.004421	-0.033805
4	-0.000843	0.093577	0.057585	-0.069250	0.025734	-0.001145	-0.040913
..	...	...	...	...	...	...	
119	0.013204	0.104309	0.094787	0.055603	0.038557	0.015047	0.077456
120	0.039099	0.043180	0.119983	-0.017972	0.139362	0.055141	0.046884
121	0.118540	0.007486	0.032157	-0.056051	0.107429	0.159474	-0.070768
122	0.149094	0.038679	0.049007	-0.040010	-0.015238	0.071576	-0.019588
123	0.001224	0.067488	0.064445	-0.104767	0.071490	-0.016061	-0.049272

	X7	X8	X9	X10	X11	X12	X13
0	-0.073022	-0.177967	-0.046539	-0.176807	-0.093743	-0.019910	no_efectores
1	-0.019324	-0.051417	-0.033550	-0.055260	-0.065876	-0.047493	no_efectores
2	0.019694	-0.043341	-0.006750	-0.049723	-0.102480	-0.057332	no_efectores
3	0.026946	-0.051479	0.002180	-0.047125	-0.077480	-0.021516	no_efectores
4	-0.000748	-0.052139	-0.028762	-0.062232	-0.069021	-0.047349	no_efectores
..	...	...	...	...	...	...	
119	0.057797	0.030293	-0.030493	0.010036	0.001627	0.066129	no_efectores
120	0.088372	0.098059	0.052733	0.014243	0.023640	0.062680	no_efectores
121	0.040778	-0.012200	-0.037553	0.045189	0.110732	-0.026798	no_efectores
122	0.204239	0.156980	0.066204	0.062484	-0.087189	-0.022955	no_efectores
123	0.057831	-0.030673	-0.011330	-0.036026	-0.079360	-0.040350	no_efectores

[118 rows x 14 columns]

Covarianza de auto cruzamiento (ACC) hidro\_mass no\_efectores Globodera dataset 1, sin valores atípicos.  
Estadísticas.

	X0	X1	X2	X3	X4	X5 \
count	118.000000	118.000000	118.000000	118.000000	118.000000	118.000000
mean	0.021903	0.048527	0.036266	-0.044799	0.032817	0.006770
std	0.070668	0.067908	0.047091	0.063470	0.068177	0.064397
min	-0.184051	-0.157437	-0.109834	-0.218116	-0.167325	-0.184146
25%	-0.020254	0.004352	0.012023	-0.077821	0.014636	-0.035232
50%	0.005493	0.045490	0.043566	-0.051149	0.047937	0.000814
75%	0.053532	0.082203	0.072057	-0.001483	0.076146	0.027663
max	0.199982	0.216889	0.127200	0.135031	0.139362	0.159686

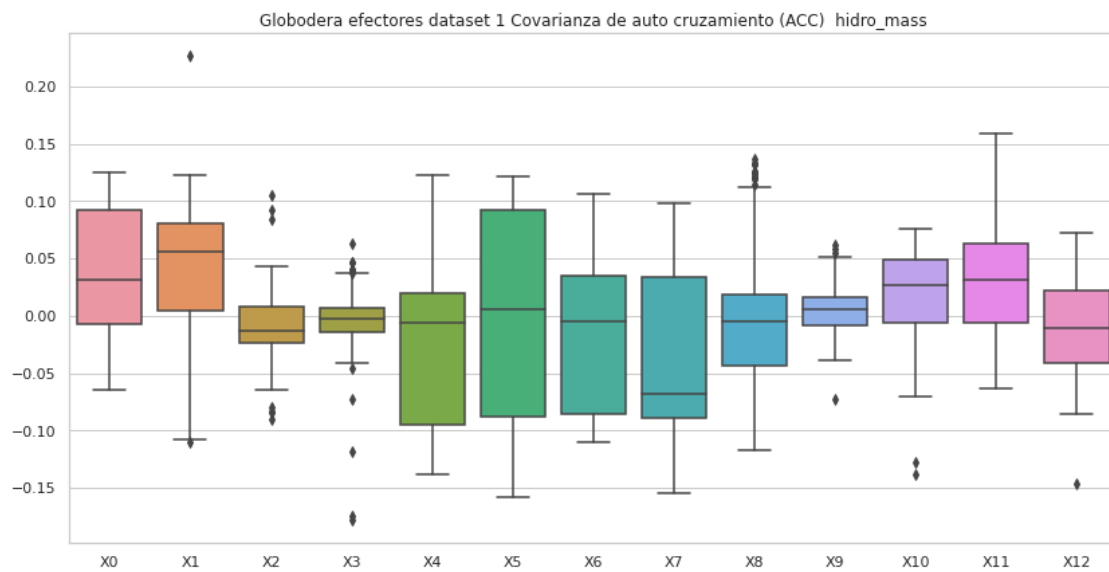
  

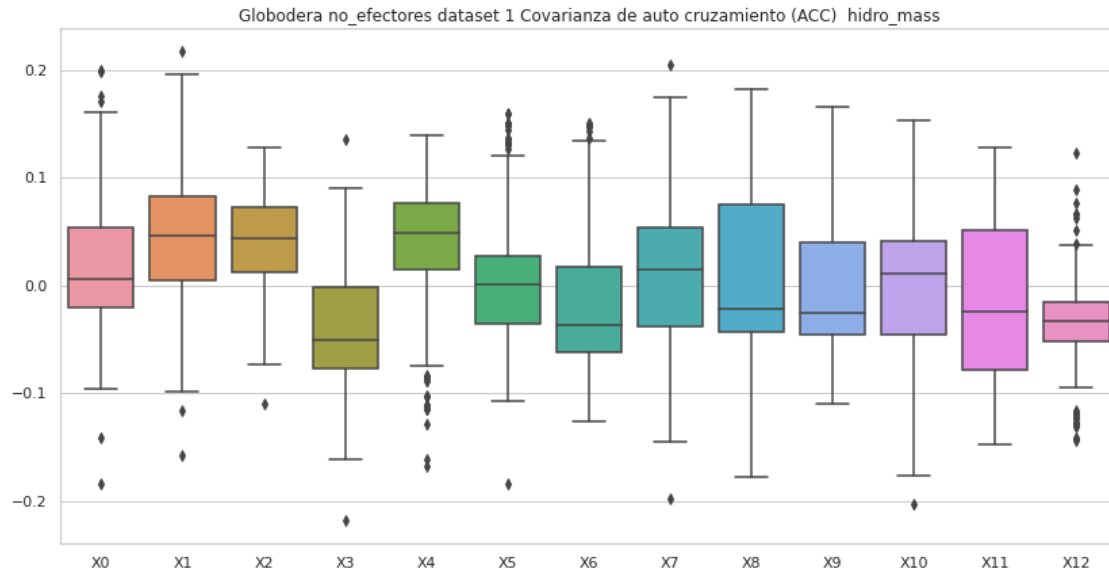
	X6	X7	X8	X9	X10	X11 \
count	118.000000	118.000000	118.000000	118.000000	118.000000	118.000000
mean	-0.020202	0.007475	0.012308	-0.004026	-0.001255	-0.016680
std	0.066103	0.078937	0.079113	0.059193	0.063999	0.069513



min	-0.126615	-0.197375	-0.177967	-0.110068	-0.203588	-0.147431
25%	-0.062341	-0.038507	-0.042840	-0.045285	-0.045968	-0.079127
50%	-0.037590	0.014969	-0.021770	-0.026120	0.010723	-0.024840
75%	0.016676	0.052932	0.075384	0.039197	0.041539	0.050611
max	0.150148	0.204239	0.182028	0.165127	0.152627	0.128297

	X12
count	118.000000
mean	-0.034052
std	0.048349
min	-0.143409
25%	-0.051679
50%	-0.032792
75%	-0.015896
max	0.122133





## 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) +",\n"
    ↪ " + 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)+"\n
↪"+str(transf)+" "+str(comp)+" "+str(estado))

```

efectores

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

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	-0.010549	0.079950	-0.021663	-0.003517	0.010475	0.104994	-0.084835
1	0.087442	0.074751	-0.020555	0.011417	-0.089592	-0.080496	-0.010581
2	-0.064158	0.042774	-0.023861	0.035360	0.055761	-0.158534	0.105877
3	0.093568	0.091161	-0.014655	0.008886	-0.107150	-0.090436	0.001904
4	0.016486	-0.080949	-0.004929	0.000070	0.065487	0.070790	-0.087497
..	...	...	...	...	...	...	
119	0.062843	0.007681	0.012457	-0.019704	-0.083203	-0.066870	0.069773
120	0.108285	0.051721	0.008466	-0.012157	-0.102047	-0.108167	0.034955
121	0.106425	0.052584	0.009448	-0.015887	-0.097062	-0.107316	0.036854
122	0.093445	0.069846	-0.010262	0.001120	-0.098703	-0.084250	0.000165
123	0.088304	0.061832	0.016245	-0.003739	-0.105054	-0.086628	0.014074
	X7	X8	X9	X10	X11	X12	X13
0	0.042210	0.001973	0.016051	0.049825	0.005194	0.015993	efectores
1	-0.063674	-0.000746	0.001506	0.029297	0.051266	-0.047259	efectores
2	-0.112671	-0.059166	-0.037901	-0.138312	0.062583	-0.146305	efectores
3	-0.071321	-0.020210	0.001595	0.017997	0.057693	-0.041386	efectores
4	-0.088773	0.105527	-0.017113	-0.028421	-0.052765	0.048569	efectores
..	...	...	...	...	...	...	
119	-0.101180	-0.101671	0.008942	0.020309	0.053943	-0.046297	efectores
120	-0.067454	-0.046498	-0.011636	0.050587	0.087886	-0.041194	efectores
121	-0.070281	-0.044722	-0.015348	0.046718	0.085861	-0.041362	efectores
122	-0.072415	-0.027387	0.002758	0.017747	0.056809	-0.033485	efectores
123	-0.080820	-0.025647	0.007020	0.028570	0.053436	-0.041788	efectores

[124 rows x 14 columns]

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

Estadísticas.

	X0	X1	X2	X3	X4	X5 \
count	124.000000	124.000000	124.000000	124.000000	124.000000	124.000000
mean	0.041515	0.031174	-0.004925	-0.000276	-0.020619	-0.002568

std	0.066352	0.087484	0.042344	0.061259	0.079160	0.090881
min	-0.117886	-0.503538	-0.090270	-0.187075	-0.343665	-0.256220
25%	-0.008009	0.004066	-0.023784	-0.015472	-0.094270	-0.087966
50%	0.029163	0.055562	-0.010589	-0.003000	-0.003425	0.006121
75%	0.091795	0.080227	0.008164	0.007965	0.023904	0.092380
max	0.464974	0.316647	0.193164	0.440767	0.157839	0.131299

	X6	X7	X8	X9	X10	X11 \
count	124.000000	124.000000	124.000000	124.000000	124.000000	124.000000
mean	-0.014944	-0.036240	-0.003822	0.008564	0.022343	0.029879
std	0.075500	0.069502	0.077602	0.031524	0.056190	0.061086
min	-0.110454	-0.167294	-0.415399	-0.072703	-0.138312	-0.187073
25%	-0.084883	-0.088542	-0.043824	-0.008427	-0.006318	-0.007271
50%	-0.005363	-0.067235	-0.004181	0.007182	0.027068	0.020111
75%	0.034960	0.037404	0.023542	0.016824	0.049699	0.062992
max	0.387758	0.161485	0.161947	0.197792	0.416201	0.304052

	X12
count	124.000000
mean	-0.006512
std	0.050917
min	-0.247599
25%	-0.041512
50%	-0.012433
75%	0.022073
max	0.165858

no\_efectores

Covarianza de auto cruzamiento (ACC) mass no\_efectores Globodera dataset 1, con valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	-0.007934	0.149593	0.018527	0.008338	0.103889	-0.083495	-0.089614
1	-0.019927	0.051389	0.085881	-0.049360	0.046373	0.024052	-0.061740
2	0.011830	0.047089	0.086867	-0.089937	0.067371	0.008432	-0.067699
3	-0.002923	0.060430	0.069633	-0.105039	0.069781	-0.004421	-0.033805
4	-0.000843	0.093577	0.057585	-0.069250	0.025734	-0.001145	-0.040913
..	...	...	...	...	...	...	...
119	0.013204	0.104309	0.094787	0.055603	0.038557	0.015047	0.077456
120	0.039099	0.043180	0.119983	-0.017972	0.139362	0.055141	0.046884
121	0.118540	0.007486	0.032157	-0.056051	0.107429	0.159474	-0.070768
122	0.149094	0.038679	0.049007	-0.040010	-0.015238	0.071576	-0.019588
123	0.001224	0.067488	0.064445	-0.104767	0.071490	-0.016061	-0.049272
	X7	X8	X9	X10	X11	X12	X13

```

0  -0.073022 -0.177967 -0.046539 -0.176807 -0.093743 -0.019910 no_efectores
1  -0.019324 -0.051417 -0.033550 -0.055260 -0.065876 -0.047493 no_efectores
2   0.019694 -0.043341 -0.006750 -0.049723 -0.102480 -0.057332 no_efectores
3   0.026946 -0.051479  0.002180 -0.047125 -0.077480 -0.021516 no_efectores
4  -0.000748 -0.052139 -0.028762 -0.062232 -0.069021 -0.047349 no_efectores
..      ...      ...      ...      ...      ...      ...
119 0.057797  0.030293 -0.030493  0.010036  0.001627  0.066129 no_efectores
120 0.088372  0.098059  0.052733  0.014243  0.023640  0.062680 no_efectores
121 0.040778 -0.012200 -0.037553  0.045189  0.110732 -0.026798 no_efectores
122 0.204239  0.156980  0.066204  0.062484 -0.087189 -0.022955 no_efectores
123 0.057831 -0.030673 -0.011330 -0.036026 -0.079360 -0.040350 no_efectores

```

[124 rows x 14 columns]

Covarianza de auto cruzamiento (ACC) mass no\_efectores Globodera dataset 1, con valores atípicos.

Estadísticas.

	X0	X1	X2	X3	X4	X5 \
count	124.000000	124.000000	124.000000	124.000000	124.000000	124.000000
mean	0.023332	0.046149	0.033501	-0.044288	0.030525	0.009802
std	0.072712	0.069315	0.062403	0.063106	0.070670	0.072309
min	-0.184051	-0.157437	-0.205958	-0.218116	-0.203192	-0.184146
25%	-0.020036	0.000962	0.007560	-0.077390	0.008463	-0.035525
50%	0.007349	0.044149	0.042522	-0.050286	0.047602	0.002650
75%	0.057589	0.081860	0.071791	-0.000359	0.076731	0.030136
max	0.199982	0.216889	0.340880	0.135031	0.139362	0.378927

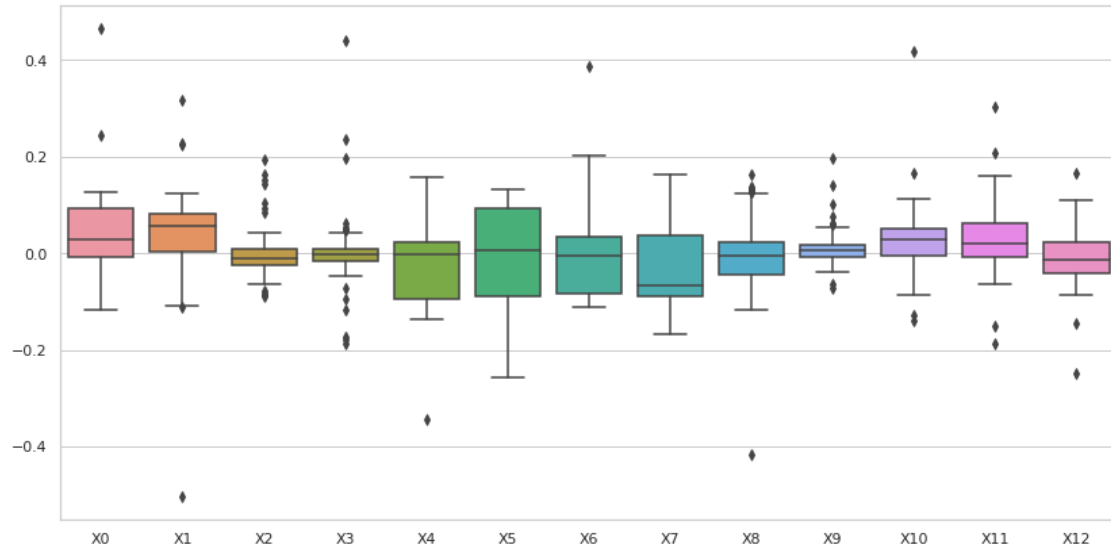
  

	X6	X7	X8	X9	X10	X11 \
count	124.000000	124.000000	124.000000	124.000000	124.000000	124.000000
mean	-0.018571	0.005676	0.013756	-0.004901	-0.005273	-0.014209
std	0.066485	0.079465	0.084746	0.058742	0.068531	0.074776
min	-0.126615	-0.197375	-0.177967	-0.110068	-0.216828	-0.147431
25%	-0.061940	-0.042288	-0.043956	-0.045601	-0.047512	-0.079930
50%	-0.036001	0.014592	-0.021770	-0.026120	0.010224	-0.023741
75%	0.021615	0.051857	0.078129	0.034818	0.040976	0.051555
max	0.150148	0.204239	0.308672	0.165127	0.152627	0.276172

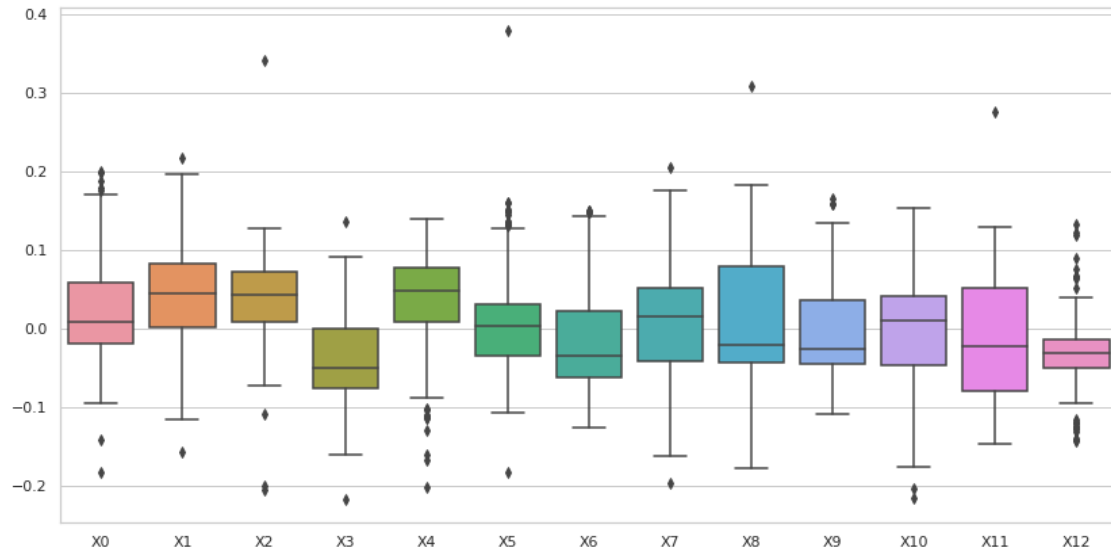
  

	X12
count	124.000000
mean	-0.030696
std	0.051789
min	-0.143409
25%	-0.051404
50%	-0.032277
75%	-0.013778
max	0.132850

Globodera efectores dataset 1 Covarianza de auto cruzamiento (ACC) mass con valores atípicos.



Globodera no\_efectores dataset 1 Covarianza de auto cruzamiento (ACC) mass con valores atípicos.



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

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

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

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

    if etiq == "efectores":
        df=ACC_mass_efec

    if etiq == "no_efectores":
        df=ACC_mass_no_efec

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

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

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

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

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

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

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	-0.010549	0.079950	-0.021663	-0.003517	0.010475	0.104994	-0.084835
1	0.087442	0.074751	-0.020555	0.011417	-0.089592	-0.080496	-0.010581
2	-0.064158	0.042774	-0.023861	0.035360	0.055761	-0.158534	0.105877
3	0.093568	0.091161	-0.014655	0.008886	-0.107150	-0.090436	0.001904
4	0.016486	-0.080949	-0.004929	0.000070	0.065487	0.070790	-0.087497
..	...	...	...	...	...	...	
119	0.062843	0.007681	0.012457	-0.019704	-0.083203	-0.066870	0.069773
120	0.108285	0.051721	0.008466	-0.012157	-0.102047	-0.108167	0.034955
121	0.106425	0.052584	0.009448	-0.015887	-0.097062	-0.107316	0.036854
122	0.093445	0.069846	-0.010262	0.001120	-0.098703	-0.084250	0.000165
123	0.088304	0.061832	0.016245	-0.003739	-0.105054	-0.086628	0.014074

	X7	X8	X9	X10	X11	X12	X13
0	0.042210	0.001973	0.016051	0.049825	0.005194	0.015993	efectores
1	-0.063674	-0.000746	0.001506	0.029297	0.051266	-0.047259	efectores
2	-0.112671	-0.059166	-0.037901	-0.138312	0.062583	-0.146305	efectores
3	-0.071321	-0.020210	0.001595	0.017997	0.057693	-0.041386	efectores
4	-0.088773	0.105527	-0.017113	-0.028421	-0.052765	0.048569	efectores
..	...	...	...	...	...	...	
119	-0.101180	-0.101671	0.008942	0.020309	0.053943	-0.046297	efectores
120	-0.067454	-0.046498	-0.011636	0.050587	0.087886	-0.041194	efectores
121	-0.070281	-0.044722	-0.015348	0.046718	0.085861	-0.041362	efectores
122	-0.072415	-0.027387	0.002758	0.017747	0.056809	-0.033485	efectores
123	-0.080820	-0.025647	0.007020	0.028570	0.053436	-0.041788	efectores

[116 rows x 14 columns]

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

Estadísticas.

	X0	X1	X2	X3	X4	X5 \
count	116.000000	116.000000	116.000000	116.000000	116.000000	116.000000
mean	0.039879	0.031257	-0.010420	-0.005554	-0.021974	-0.001244
std	0.048747	0.067097	0.029619	0.032398	0.072381	0.088629
min	-0.064158	-0.110488	-0.090270	-0.178040	-0.137897	-0.158534
25%	-0.007525	0.004066	-0.023912	-0.014410	-0.094489	-0.087966
50%	0.031747	0.055562	-0.012522	-0.003080	-0.005607	0.006121
75%	0.091795	0.080067	0.007764	0.006391	0.019932	0.092380



max	0.125288	0.226622	0.104776	0.062859	0.122434	0.121527
-----	----------	----------	----------	----------	----------	----------

	X6	X7	X8	X9	X10	X11 \
count	116.000000	116.000000	116.000000	116.000000	116.000000	116.000000
mean	-0.020457	-0.040766	-0.002921	0.004392	0.017964	0.030494
std	0.064297	0.064642	0.067581	0.020139	0.039666	0.048172
min	-0.110454	-0.155074	-0.117263	-0.072703	-0.138312	-0.063735
25%	-0.085258	-0.088773	-0.043607	-0.008756	-0.006051	-0.006533
50%	-0.005363	-0.068393	-0.005285	0.005140	0.027068	0.031493
75%	0.034615	0.033811	0.018951	0.015887	0.049440	0.062992
max	0.105877	0.098034	0.136839	0.061478	0.076060	0.158915

	X12
count	116.000000
mean	-0.006972
std	0.041708
min	-0.146305
25%	-0.041395
50%	-0.010771
75%	0.021568
max	0.072321

Covarianza de auto cruzamiento (ACC) mass no\_efectores Globodera dataset 1, sin valores atípicos.  
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	-0.007934	0.149593	0.018527	0.008338	0.103889	-0.083495	-0.089614
1	-0.019927	0.051389	0.085881	-0.049360	0.046373	0.024052	-0.061740
2	0.011830	0.047089	0.086867	-0.089937	0.067371	0.008432	-0.067699
3	-0.002923	0.060430	0.069633	-0.105039	0.069781	-0.004421	-0.033805
4	-0.000843	0.093577	0.057585	-0.069250	0.025734	-0.001145	-0.040913
..	...	...	...	...	...	...	...
119	0.013204	0.104309	0.094787	0.055603	0.038557	0.015047	0.077456
120	0.039099	0.043180	0.119983	-0.017972	0.139362	0.055141	0.046884
121	0.118540	0.007486	0.032157	-0.056051	0.107429	0.159474	-0.070768
122	0.149094	0.038679	0.049007	-0.040010	-0.015238	0.071576	-0.019588
123	0.001224	0.067488	0.064445	-0.104767	0.071490	-0.016061	-0.049272

	X7	X8	X9	X10	X11	X12	X13
0	-0.073022	-0.177967	-0.046539	-0.176807	-0.093743	-0.019910	no_efectores
1	-0.019324	-0.051417	-0.033550	-0.055260	-0.065876	-0.047493	no_efectores
2	0.019694	-0.043341	-0.006750	-0.049723	-0.102480	-0.057332	no_efectores
3	0.026946	-0.051479	0.002180	-0.047125	-0.077480	-0.021516	no_efectores
4	-0.000748	-0.052139	-0.028762	-0.062232	-0.069021	-0.047349	no_efectores
..	...	...	...	...	...	...	...

```

119  0.057797  0.030293 -0.030493  0.010036  0.001627  0.066129  no_efectores
120  0.088372  0.098059  0.052733  0.014243  0.023640  0.062680  no_efectores
121  0.040778 -0.012200 -0.037553  0.045189  0.110732 -0.026798  no_efectores
122  0.204239  0.156980  0.066204  0.062484 -0.087189 -0.022955  no_efectores
123  0.057831 -0.030673 -0.011330 -0.036026 -0.079360 -0.040350  no_efectores

```

[118 rows x 14 columns]

Covarianza de auto cruzamiento (ACC) mass no\_efectores Globodera dataset 1, sin valores atípicos.

Estadísticas.

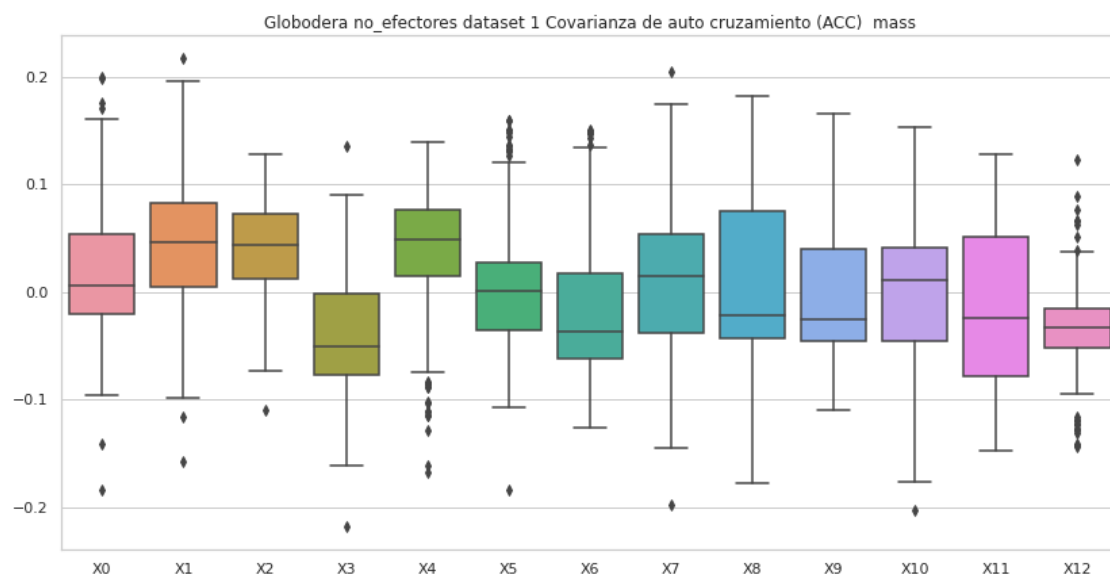
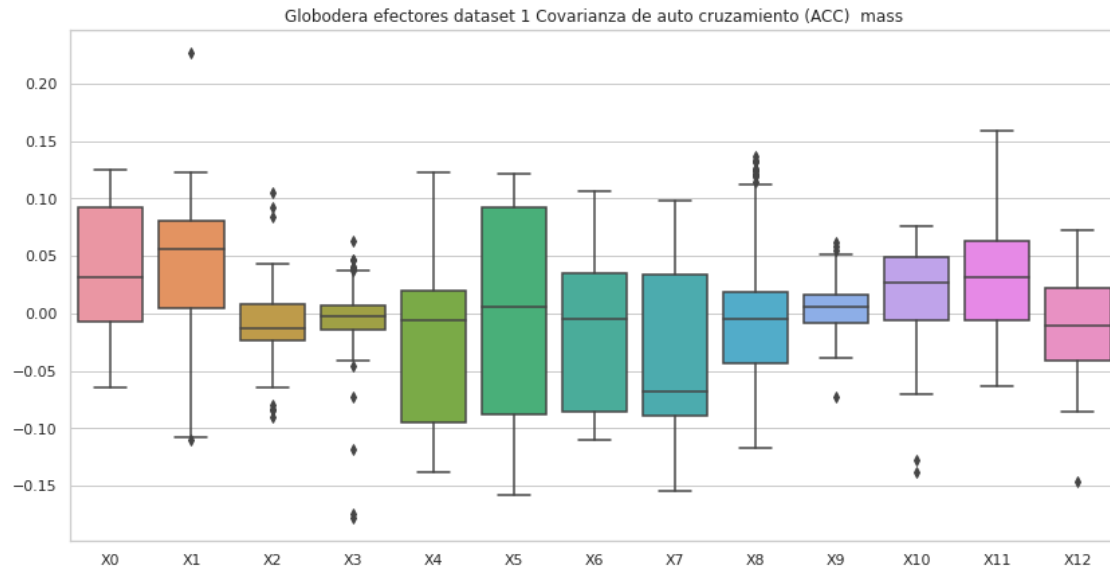
	X0	X1	X2	X3	X4	X5 \
count	118.000000	118.000000	118.000000	118.000000	118.000000	118.000000
mean	0.021903	0.048527	0.036266	-0.044799	0.032817	0.006770
std	0.070668	0.067908	0.047091	0.063470	0.068177	0.064397
min	-0.184051	-0.157437	-0.109834	-0.218116	-0.167325	-0.184146
25%	-0.020254	0.004352	0.012023	-0.077821	0.014636	-0.035232
50%	0.005493	0.045490	0.043566	-0.051149	0.047937	0.000814
75%	0.053532	0.082203	0.072057	-0.001483	0.076146	0.027663
max	0.199982	0.216889	0.127200	0.135031	0.139362	0.159686

	X6	X7	X8	X9	X10	X11 \
count	118.000000	118.000000	118.000000	118.000000	118.000000	118.000000
mean	-0.020202	0.007475	0.012308	-0.004026	-0.001255	-0.016680
std	0.066103	0.078937	0.079113	0.059193	0.063999	0.069513
min	-0.126615	-0.197375	-0.177967	-0.110068	-0.203588	-0.147431
25%	-0.062341	-0.038507	-0.042840	-0.045285	-0.045968	-0.079127
50%	-0.037590	0.014969	-0.021770	-0.026120	0.010723	-0.024840
75%	0.016676	0.052932	0.075384	0.039197	0.041539	0.050611
max	0.150148	0.204239	0.182028	0.165127	0.152627	0.128297

	X12
count	118.000000
mean	-0.034052
std	0.048349
min	-0.143409
25%	-0.051679
50%	-0.032792
75%	-0.015896
max	0.122133



## 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 Globodera dataset 1, con valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	-0.123281	0.009368	-0.160451	0.073226	-0.023615	0.000064	-0.079555
1	0.017515	-0.096585	0.037832	-0.009383	0.079429	-0.030363	-0.057120
2	-0.114669	-0.182504	0.116009	-0.261545	0.168657	0.295811	-0.089210
3	0.022154	-0.097727	0.020161	-0.008216	0.086245	-0.024061	-0.067988
4	0.094713	-0.037312	-0.028884	-0.029602	0.062516	0.016304	-0.065760
..	...	...	...	...	...	...	...
119	0.055140	-0.101706	-0.002728	-0.024328	0.081574	-0.009781	-0.071001
120	0.013938	-0.078822	0.006466	0.002621	0.081785	-0.030197	-0.075177
121	0.015690	-0.080110	0.003802	0.003756	0.083421	-0.028963	-0.077856
122	0.022832	-0.098430	0.019951	-0.008344	0.086340	-0.023962	-0.068226
123	0.020656	-0.083613	0.020081	-0.012561	0.097713	-0.017765	-0.067609

	X7	X8	X9	X10	X11	X12	X13
0	-0.122671	0.019076	0.071766	0.123552	0.005047	0.023861	efectores
1	-0.055091	0.021631	0.006855	-0.022101	0.031099	0.014453	efectores
2	0.264589	-0.133720	-0.219929	0.373555	-0.036655	-0.011333	efectores

3	-0.055420	0.026429	0.010124	-0.039458	0.025677	0.033227	efectores
4	0.045986	0.003977	0.036264	0.014666	-0.008302	0.002893	efectores
..	...	...	...	...	...	...	
119	-0.093315	-0.018797	0.005223	0.008975	0.031150	0.025466	efectores
120	-0.060301	0.020236	0.031042	0.002461	0.038020	0.019527	efectores
121	-0.060508	0.019191	0.031338	0.003658	0.039884	0.022230	efectores
122	-0.053997	0.026904	0.011079	-0.040950	0.025216	0.033716	efectores
123	-0.059699	0.022137	0.009476	-0.032627	0.030359	0.027717	efectores

[124 rows x 14 columns]

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

	X0	X1	X2	X3	X4	X5	\
count	124.000000	124.000000	124.000000	124.000000	124.000000	124.000000	
mean	0.003829	-0.042148	-0.029196	0.022023	0.042213	0.002385	
std	0.090697	0.068024	0.080874	0.075512	0.059318	0.046004	
min	-0.139245	-0.182504	-0.160451	-0.261545	-0.067710	-0.046481	
25%	-0.108240	-0.087824	-0.114939	-0.018547	-0.020551	-0.023321	
50%	0.019339	-0.054773	-0.006981	0.002650	0.063246	-0.006485	
75%	0.054696	0.009846	0.015488	0.075346	0.081734	0.013996	
max	0.329217	0.346279	0.289122	0.531124	0.244329	0.295811	

	X6	X7	X8	X9	X10	X11	\
count	124.000000	124.000000	124.000000	124.000000	124.000000	124.000000	
mean	-0.056088	-0.037530	0.011182	0.037288	0.030408	0.025384	
std	0.045311	0.076511	0.039779	0.043398	0.070925	0.045805	
min	-0.133182	-0.143350	-0.133720	-0.219929	-0.166205	-0.110496	
25%	-0.076030	-0.097808	-0.004369	0.012161	-0.010200	0.011056	
50%	-0.070937	-0.060405	0.017539	0.032056	0.010392	0.022919	
75%	-0.060973	0.038938	0.024476	0.062094	0.093638	0.033954	
max	0.127873	0.264589	0.209367	0.260661	0.373555	0.288820	

	X12
count	124.000000
mean	0.021268
std	0.029477
min	-0.098664
25%	0.009038
50%	0.022755
75%	0.033227
max	0.175938

no\_efectores

Covarianza de auto cruzamiento (ACC) hidro no\_efectores Globodera dataset 1,  
con valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	0.024849	0.004041	-0.170572	-0.277845	-0.098585	-0.040764	0.038626
1	-0.058060	0.027591	-0.111051	-0.093267	-0.073351	0.008339	-0.017610
2	-0.014666	0.009757	-0.080331	-0.050469	-0.029395	0.050877	-0.029392
3	-0.003139	0.035502	-0.111566	-0.073729	-0.020763	0.009867	-0.027307
4	-0.025714	0.066144	-0.111332	-0.039149	-0.042260	0.018342	0.005264
..	...	...	...	...	...	...	
119	0.005124	0.043223	0.025664	-0.004589	0.034867	0.023739	0.015257
120	0.040518	0.016848	0.070768	0.024538	0.058577	0.011243	0.036186
121	-0.066898	-0.010606	-0.048494	-0.016799	0.034596	0.008789	-0.007599
122	0.107626	0.101017	0.147748	0.028942	0.038990	-0.084023	0.050449
123	-0.015782	0.022656	-0.081025	-0.052122	-0.038009	0.015996	-0.021433

	X7	X8	X9	X10	X11	X12	X13
0	0.000381	0.112278	-0.170570	0.110093	-0.020824	0.035078	no_efectores
1	-0.034259	0.074712	-0.009761	0.057281	-0.018082	0.037854	no_efectores
2	0.003458	0.033986	-0.027885	0.021803	0.025791	0.041966	no_efectores
3	0.005966	0.057769	-0.000677	0.049796	-0.054679	0.050704	no_efectores
4	-0.039784	0.070192	0.015539	0.029634	0.000706	-0.020405	no_efectores
..	...	...	...	...	...	...	
119	0.003572	-0.023706	0.042339	-0.000863	0.021983	0.026262	no_efectores
120	0.052351	-0.010204	0.051144	0.029473	0.008974	0.043424	no_efectores
121	-0.015466	0.007714	0.051931	0.042349	0.050734	-0.012105	no_efectores
122	0.115957	-0.068779	-0.049548	-0.021974	-0.063990	-0.114843	no_efectores
123	-0.008299	0.047306	-0.002409	0.029608	-0.025762	0.038068	no_efectores

[124 rows x 14 columns]

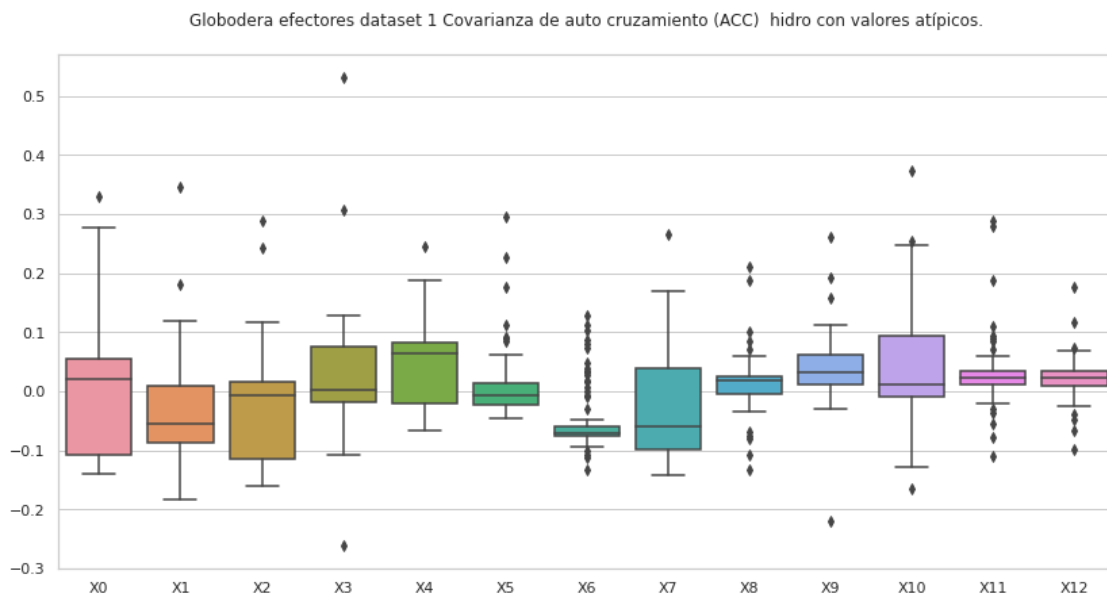
Covarianza de auto cruzamiento (ACC) hidro no\_efectores Globodera dataset 1,  
con valores atípicos.

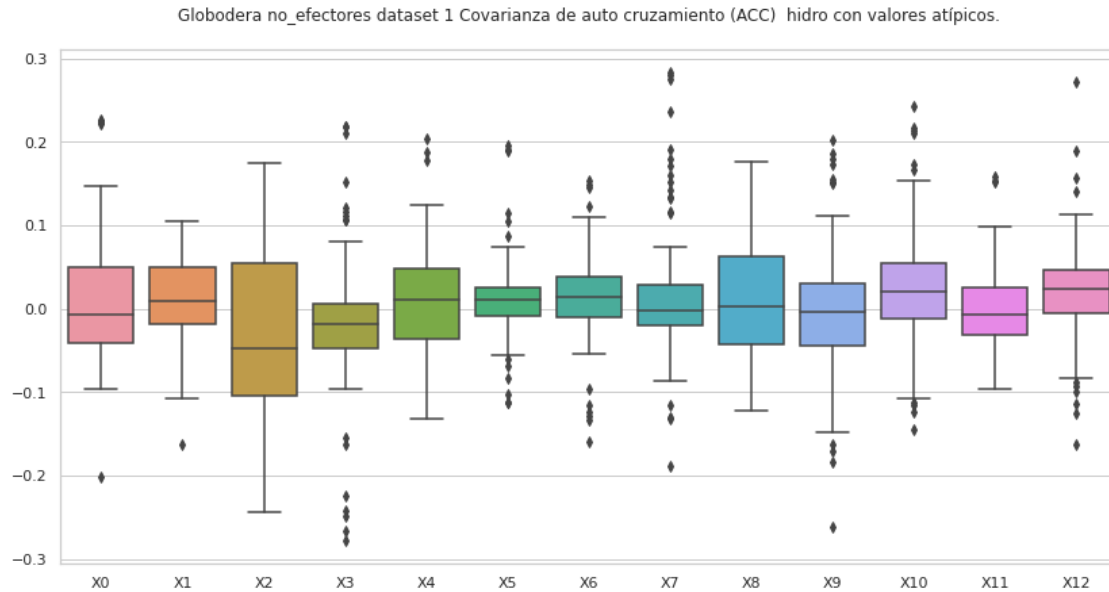
Estadísticas.

	X0	X1	X2	X3	X4	X5 \
count	124.000000	124.000000	124.000000	124.000000	124.000000	124.000000
mean	0.005076	0.009031	-0.030610	-0.019041	0.005910	0.009825
std	0.073021	0.052938	0.093910	0.077742	0.059651	0.046850
min	-0.201731	-0.163171	-0.244158	-0.277845	-0.131673	-0.112454
25%	-0.040900	-0.018543	-0.104497	-0.046939	-0.037099	-0.009051
50%	-0.007042	0.008684	-0.047901	-0.018474	0.010360	0.009880
75%	0.049048	0.049390	0.054628	0.005195	0.048765	0.024623
max	0.225746	0.105021	0.174312	0.218968	0.202990	0.194872

	X6	X7	X8	X9	X10	X11	\
count	124.000000	124.000000	124.000000	124.000000	124.000000	124.000000	
mean	0.013494	0.010788	0.011492	-0.008242	0.022974	-0.004234	
std	0.050119	0.076741	0.062673	0.073547	0.070911	0.048407	
min	-0.159453	-0.189139	-0.121477	-0.261701	-0.144267	-0.096230	
25%	-0.010886	-0.020107	-0.043397	-0.044464	-0.011649	-0.031778	
50%	0.013410	-0.002334	0.002900	-0.003144	0.020803	-0.007048	
75%	0.038186	0.028418	0.062720	0.030790	0.054487	0.025912	
max	0.153710	0.282871	0.175914	0.201787	0.242291	0.157579	

	X12
count	124.000000
mean	0.021003
std	0.057742
min	-0.162254
25%	-0.005742
50%	0.024280
75%	0.046275
max	0.271373





## 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)+"\n
→"+str(transf)+" "+str(comp))

```

efectores

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

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
0	-0.123281	0.009368	-0.160451	0.073226	-0.023615	0.000064	-0.079555
1	0.017515	-0.096585	0.037832	-0.009383	0.079429	-0.030363	-0.057120
3	0.022154	-0.097727	0.020161	-0.008216	0.086245	-0.024061	-0.067988
4	0.094713	-0.037312	-0.028884	-0.029602	0.062516	0.016304	-0.065760
5	0.015061	-0.085555	-0.009455	0.002679	0.085459	-0.035578	-0.072768
..	...	...	...	...	...	...	...
119	0.055140	-0.101706	-0.002728	-0.024328	0.081574	-0.009781	-0.071001
120	0.013938	-0.078822	0.006466	0.002621	0.081785	-0.030197	-0.075177
121	0.015690	-0.080110	0.003802	0.003756	0.083421	-0.028963	-0.077856
122	0.022832	-0.098430	0.019951	-0.008344	0.086340	-0.023962	-0.068226
123	0.020656	-0.083613	0.020081	-0.012561	0.097713	-0.017765	-0.067609
	X7	X8	X9	X10	X11	X12	X13
0	-0.122671	0.019076	0.071766	0.123552	0.005047	0.023861	efectores
1	-0.055091	0.021631	0.006855	-0.022101	0.031099	0.014453	efectores
3	-0.055420	0.026429	0.010124	-0.039458	0.025677	0.033227	efectores
4	0.045986	0.003977	0.036264	0.014666	-0.008302	0.002893	efectores
5	-0.062521	0.018227	0.041546	-0.009713	0.029754	0.021906	efectores
..	...	...	...	...	...	...	...
119	-0.093315	-0.018797	0.005223	0.008975	0.031150	0.025466	efectores

120	-0.060301	0.020236	0.031042	0.002461	0.038020	0.019527	efectores
121	-0.060508	0.019191	0.031338	0.003658	0.039884	0.022230	efectores
122	-0.053997	0.026904	0.011079	-0.040950	0.025216	0.033716	efectores
123	-0.059699	0.022137	0.009476	-0.032627	0.030359	0.027717	efectores

[114 rows x 14 columns]

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

Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	
mean	-0.003475	-0.050090	-0.039905	0.017227	0.037408	-0.006111	
std	0.081249	0.050450	0.068854	0.045778	0.055022	0.023530	
min	-0.139245	-0.153055	-0.160451	-0.047489	-0.067710	-0.046481	
25%	-0.109149	-0.089166	-0.126881	-0.018505	-0.028270	-0.024061	
50%	0.018120	-0.055550	-0.011476	0.002170	0.063246	-0.007072	
75%	0.051634	0.003776	0.010174	0.073031	0.081504	0.008658	
max	0.177135	0.095079	0.081285	0.101264	0.129713	0.088366	

	X6	X7	X8	X9	X10	X11	\
count	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	
mean	-0.064143	-0.049559	0.008219	0.035056	0.025852	0.019435	
std	0.028147	0.062309	0.024483	0.025465	0.055636	0.023321	
min	-0.112340	-0.143350	-0.081222	-0.030508	-0.127769	-0.077135	
25%	-0.076147	-0.100408	-0.002682	0.011440	-0.011174	0.009919	
50%	-0.071656	-0.061606	0.016172	0.032028	0.008120	0.019926	
75%	-0.065464	-0.036132	0.023350	0.055904	0.093386	0.031900	
max	0.048289	0.149442	0.060183	0.112050	0.127153	0.108576	

	X12
count	114.000000
mean	0.020787
std	0.019686
min	-0.048527
25%	0.014141
50%	0.022724
75%	0.031312
max	0.072453

no\_efectores

Covarianza de auto cruzamiento (ACC) no\_efectores Globodera dataset 1, sin valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6 \
1	-0.058060	0.027591	-0.111051	-0.093267	-0.073351	0.008339	-0.017610
2	-0.014666	0.009757	-0.080331	-0.050469	-0.029395	0.050877	-0.029392
3	-0.003139	0.035502	-0.111566	-0.073729	-0.020763	0.009867	-0.027307
4	-0.025714	0.066144	-0.111332	-0.039149	-0.042260	0.018342	0.005264
6	-0.036660	-0.007274	-0.044676	-0.016570	0.052801	0.021777	0.024537
..	...	...	...	...	...	...	
119	0.005124	0.043223	0.025664	-0.004589	0.034867	0.023739	0.015257
120	0.040518	0.016848	0.070768	0.024538	0.058577	0.011243	0.036186
121	-0.066898	-0.010606	-0.048494	-0.016799	0.034596	0.008789	-0.007599
122	0.107626	0.101017	0.147748	0.028942	0.038990	-0.084023	0.050449
123	-0.015782	0.022656	-0.081025	-0.052122	-0.038009	0.015996	-0.021433

	X7	X8	X9	X10	X11	X12	X13
1	-0.034259	0.074712	-0.009761	0.057281	-0.018082	0.037854	no_efectores
2	0.003458	0.033986	-0.027885	0.021803	0.025791	0.041966	no_efectores
3	0.005966	0.057769	-0.000677	0.049796	-0.054679	0.050704	no_efectores
4	-0.039784	0.070192	0.015539	0.029634	0.000706	-0.020405	no_efectores
6	-0.062978	-0.058349	0.005260	-0.015476	0.042238	0.026507	no_efectores
..	...	...	...	...	...	...	
119	0.003572	-0.023706	0.042339	-0.000863	0.021983	0.026262	no_efectores
120	0.052351	-0.010204	0.051144	0.029473	0.008974	0.043424	no_efectores
121	-0.015466	0.007714	0.051931	0.042349	0.050734	-0.012105	no_efectores
122	0.115957	-0.068779	-0.049548	-0.021974	-0.063990	-0.114843	no_efectores
123	-0.008299	0.047306	-0.002409	0.029608	-0.025762	0.038068	no_efectores

[115 rows x 14 columns]

Covarianza de auto cruzamiento (ACC) no\_efectores Globodera dataset 1, sin valores atípicos.  
Estadísticas.

	X0	X1	X2	X3	X4	X5 \
count	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000
mean	-0.000319	0.010196	-0.032867	-0.020266	0.004163	0.005775
std	0.060359	0.048533	0.088796	0.057782	0.050664	0.037026
min	-0.096871	-0.095922	-0.244158	-0.248601	-0.120591	-0.112454
25%	-0.041305	-0.017964	-0.102406	-0.045008	-0.036027	-0.008341
50%	-0.007790	0.007611	-0.048210	-0.018654	0.012153	0.009893
75%	0.043883	0.049168	0.020094	0.000097	0.048607	0.023975
max	0.146223	0.105021	0.174312	0.151479	0.124906	0.114102

	X6	X7	X8	X9	X10	X11 \
count	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000
mean	0.009486	0.004397	0.004619	-0.009225	0.013769	-0.008645
std	0.042837	0.059878	0.053432	0.059528	0.060306	0.041788

min	-0.133625	-0.132185	-0.121477	-0.183482	-0.144267	-0.096230
25%	-0.012299	-0.019679	-0.044435	-0.039698	-0.015476	-0.031952
50%	0.010716	-0.002586	0.000476	-0.003879	0.018334	-0.007106
75%	0.033208	0.025699	0.058010	0.023886	0.046972	0.024200
max	0.123028	0.190181	0.115282	0.173446	0.172651	0.098850

	X12
count	115.000000
mean	0.018367
std	0.050912
min	-0.125585
25%	-0.006688
50%	0.022731
75%	0.044815
max	0.189688

