ds2_fusarium_oxysporum_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 ="fusarium_oxysporum"
dataset = 2
nombre = ("ds" + str(dataset) + "_" + str(organismo))
nombre2 = (str(organismo)+ " dataset " + str(dataset))
r2 = ("Datos/resultados/"+ str(organismo) + "/" + str(nombre) + "/
 ↪transformaciones/sin_filtrar")
r3 = ("Datos/resultados/"+ str(organismo) + "/" + str(nombre) + "/
 ↪transformaciones/sin_atipicos")

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

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

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

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

2 Composición de aminoácidos (AAC)

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

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

    if etiq == "efectores":
        df=AAC_efec

    if etiq == "no_efectores":
        df=AAC_no_efec

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

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

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

         X0      X1     X2     X3     X4     X5     X6      X7     X8     X9  \
0     8.029   8.029  2.798  6.204  0.852  7.178  4.136   5.961  3.285  5.596
1     2.985  14.925  1.493  4.478  4.478  2.985  2.985  10.448  1.493  7.463
2     7.990   7.216  2.835  6.443  0.773  7.216  1.804   7.732  2.062  4.381
3     9.395   2.389  3.185  3.025  2.707  3.822  2.548   9.395  1.274  7.803
4    10.940   2.879  3.071  3.071  2.111  4.415  3.263  10.557  1.344  7.678
..      …       …      …      …      …      …      …       …      …      …
995   8.850   5.752  5.752  6.195  1.770  7.522  3.097   8.407  2.212  5.752
996   7.616   4.057  3.203  5.267  1.281  5.267  3.274   7.402  2.420  5.694
997  12.704   5.064  2.575  4.807  1.545  6.266  3.605   7.468  1.717  5.579
998   8.867  10.099  4.433  4.680  2.709  7.143  3.448   8.374  3.202  4.187
999   6.797   4.758  4.418  7.222  0.680  3.398  4.843   5.183  1.274  5.438
```

```
     …    X11    X12    X13    X14    X15    X16    X17    X18     X19  \
0    …  4.501  1.703  4.380  5.109  7.421  4.745  1.825  2.190   7.178
1    …  4.478  2.985  0.000  4.478  5.970  1.493  1.493  1.493  17.910
2    …  2.835  2.835  2.320  6.701  6.959  5.412  2.577  3.608   7.990
3    …  3.822  3.025  4.140  4.459  7.962  6.847  2.070  3.981   8.280
4    …  4.607  3.263  6.718  2.303  5.950  5.950  2.687  2.687   7.869
..   …    …    …    …    …    …    …    …    …      …
995  …  3.097  3.982  5.310  8.850  4.425  2.655  2.212  2.212   4.867
996  …  5.267  2.420  5.338  4.057  8.185  6.690  1.779  4.128   7.687
997  …  4.206  1.717  2.661  4.034  9.871  5.837  1.202  1.974   7.811
998  …  4.680  1.724  1.478  5.665  9.113  5.172  1.724  1.478   4.433
999  …  4.588  2.124  3.483  8.496  8.156  8.666  1.444  3.398   6.882

            X20
0     efectores
1     efectores
2     efectores
3     efectores
4     efectores
..          …
995   efectores
996   efectores
997   efectores
998   efectores
999   efectores

[1000 rows x 21 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      8.084039     5.968340     3.867431     5.764845     1.495861
std       2.257301     2.265874     1.463264     1.929056     1.260573
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       6.802500     4.545000     3.003750     4.663000     0.734750
50%       8.019000     5.841500     3.742500     5.783500     1.254500
75%       9.355250     7.170250     4.616500     6.861500     1.961000
max      17.778000    19.749000    11.765000    16.471000    11.594000

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      6.115676     4.022522     6.816109     2.407696     5.145848
std       2.317334     1.749427     2.302148     1.162049     1.742931
```

min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	4.774750	2.899000	5.247750	1.617500	4.005000
50%	5.882000	3.831000	6.650000	2.349000	5.061500
75%	7.255750	4.806500	8.163000	3.045500	6.170250
max	20.833000	19.943000	19.512000	9.524000	12.698000

	X10	X11	X12	X13	X14 \
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	8.751835	5.150916	2.316540	3.807906	5.883382
std	2.418649	2.294641	1.009545	1.503359	2.405214
min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	7.253000	3.597750	1.639000	2.857000	4.402750
50%	8.851500	4.888000	2.204500	3.765000	5.646000
75%	10.233250	6.303250	2.895000	4.697500	7.047500
max	17.164000	21.260000	6.931000	9.259000	29.208000

	X15	X16	X17	X18	X19
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	7.968727	6.002635	1.562825	2.858308	6.008516
std	2.448946	2.134896	0.939087	1.254151	1.898723
min	1.220000	0.000000	0.000000	0.000000	0.000000
25%	6.328000	4.948000	0.884000	2.066750	4.846000
50%	7.692000	5.787000	1.464000	2.746500	5.935000
75%	9.286250	6.779500	2.143000	3.608000	7.071000
max	20.513000	26.923000	6.349000	10.000000	17.910000

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

	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9 \
0	9.804	9.804	4.902	6.863	1.961	3.922	0.980	5.882	4.902	6.863
1	7.281	5.201	2.972	8.172	1.189	5.944	2.229	9.212	3.269	4.309
2	8.932	5.243	6.408	5.437	0.194	6.990	3.883	7.379	1.553	6.214
3	9.231	7.385	3.385	6.923	2.000	5.385	6.000	4.923	2.923	5.846
4	7.483	4.592	3.912	6.463	2.041	4.082	5.782	4.932	2.041	5.442
..	…	…	…	…	…	…	…	…	…	
995	6.114	5.523	3.748	5.128	0.394	5.325	4.142	5.523	2.959	6.312
996	5.316	5.980	5.648	6.977	1.329	8.306	2.658	5.316	2.326	5.316
997	6.806	7.135	2.415	7.245	2.964	5.598	3.513	8.342	2.086	4.061
998	6.475	6.763	2.446	6.331	1.295	7.194	3.453	4.029	1.583	5.755
999	3.817	9.160	2.290	6.107	0.000	6.107	3.817	3.053	5.344	11.450

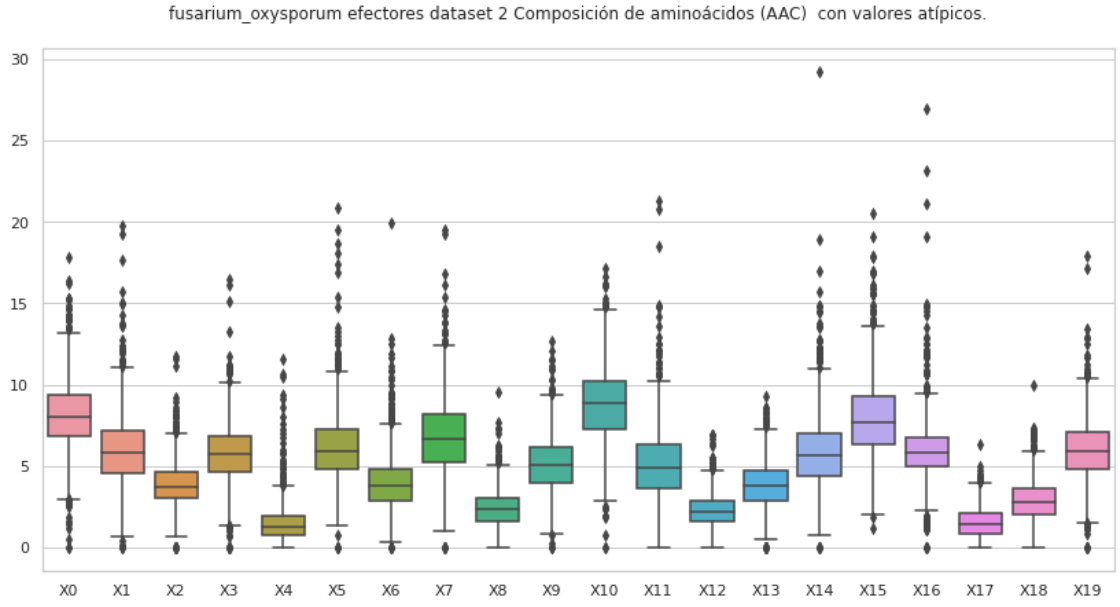
	…	X11	X12	X13	X14	X15	X16	X17	X18	X19 \
0	…	0.980	2.941	2.941	5.882	8.824	4.902	0.980	1.961	8.824

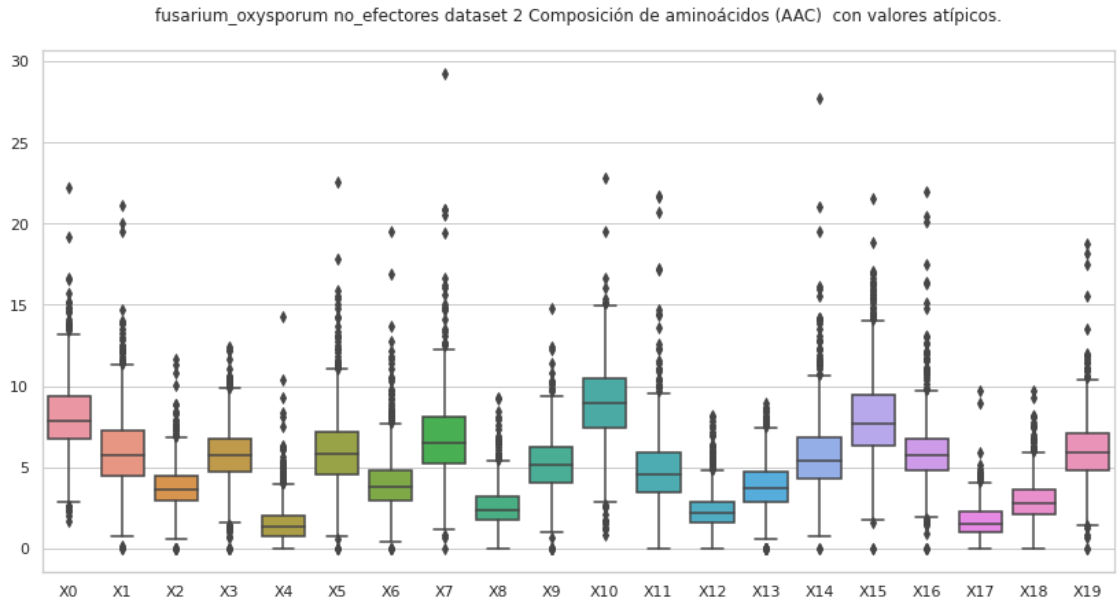
```
1    …   7.132  2.080  3.418   5.795   9.212  4.755  1.040  2.823  5.052
2    …   6.602  3.301  3.883   3.883   5.631  4.854  1.359  2.524  5.825
3    …   2.923  2.462  4.000   6.308   7.231  5.385  1.846  2.615  4.308
4    …   2.551  3.061  2.721   7.993  10.714  5.782  1.531  2.211  7.143
..   …     …    …    …     …     …    …    …    …    …
995  …   5.325  3.550  5.128   7.692   5.720  5.720  1.578  3.550  6.312
996  …  12.292  0.664  1.993   6.977   8.306  6.312  1.661  0.997  4.651
997  …   5.269  1.756  4.610   5.928   7.245  5.708  1.537  3.952  5.049
998  …   3.453  2.734  2.446  10.647  13.237  7.338  0.863  2.014  5.180
999  …   4.580  5.344  2.290   9.160   6.870  1.527  0.000  5.344  3.817

              X20
0     no_efectores
1     no_efectores
2     no_efectores
3     no_efectores
4     no_efectores
..             …
995   no_efectores
996   no_efectores
997   no_efectores
998   no_efectores
999   no_efectores

[1000 rows x 21 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      8.081971     5.982579     3.774005     5.723120     1.533014
std       2.257999     2.332568     1.393609     1.799321     1.265017
min       1.714000     0.000000     0.000000     0.000000     0.000000
25%       6.719000     4.483750     2.923000     4.694750     0.733750
50%       7.881000     5.764500     3.636000     5.748500     1.311500
75%       9.370500     7.241250     4.506500     6.785250     2.034250
max      22.222000    21.088000    11.667000    12.444000    14.286000

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      6.013300     4.054809     6.794741     2.552682     5.186182
std       2.321126     1.808385     2.479296     1.255400     1.829931
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       4.589000     2.930000     5.238250     1.736500     4.049250
50%       5.795000     3.824500     6.523000     2.406000     5.137500
```

```
75%       7.175250     4.839250     8.081000     3.206250     6.269750
max      22.581000    19.481000    29.231000     9.312000    14.801000

                X10          X11          X12          X13          X14  \
count   1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean       8.903964     4.876088     2.351781     3.822632     5.762764
std        2.497026     2.303928     1.099078     1.530475     2.367110
min        0.847000     0.000000     0.000000     0.000000     0.000000
25%        7.407000     3.446750     1.638250     2.832750     4.290000
50%        8.982000     4.605000     2.226000     3.737000     5.376000
75%       10.465750     5.894500     2.902000     4.703000     6.833500
max       22.782000    21.739000     8.163000     8.982000    27.679000

                X15          X16          X17          X18          X19
count   1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean       8.022859     5.945313     1.670316     2.923617     6.024345
std        2.641068     2.063759     1.013056     1.330154     1.963598
min        0.000000     0.000000     0.000000     0.000000     0.000000
25%        6.302750     4.818250     1.017000     2.083000     4.835500
50%        7.718000     5.752500     1.515000     2.827000     5.925500
75%        9.435500     6.783500     2.258500     3.625000     7.093250
max       21.547000    21.963000     9.677000     9.701000    18.727000
```





2.1 Composición de aminoácidos (AAC), sin valores atípicos

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

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


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

    if etiq == "efectores":
        df=AAC_efec

    if etiq == "no_efectores":
        df=AAC_no_efec

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

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

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



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

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

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

         X0      X1     X2     X3     X4     X5     X6      X7     X8     X9  \
0     8.029   8.029  2.798  6.204  0.852  7.178  4.136   5.961  3.285  5.596
2     7.990   7.216  2.835  6.443  0.773  7.216  1.804   7.732  2.062  4.381
3     9.395   2.389  3.185  3.025  2.707  3.822  2.548   9.395  1.274  7.803
4    10.940   2.879  3.071  3.071  2.111  4.415  3.263  10.557  1.344  7.678
5     8.735   2.711  3.614  2.410  2.410  4.819  5.120   7.831  1.205  7.229
..      …       …      …      …      …      …      …       …      …      …
995   8.850   5.752  5.752  6.195  1.770  7.522  3.097   8.407  2.212  5.752
996   7.616   4.057  3.203  5.267  1.281  5.267  3.274   7.402  2.420  5.694
997  12.704   5.064  2.575  4.807  1.545  6.266  3.605   7.468  1.717  5.579
998   8.867  10.099  4.433  4.680  2.709  7.143  3.448   8.374  3.202  4.187
999   6.797   4.758  4.418  7.222  0.680  3.398  4.843   5.183  1.274  5.438

     …    X11    X12    X13    X14    X15    X16    X17    X18    X19  \
0    …  4.501  1.703  4.380  5.109  7.421  4.745  1.825  2.190  7.178
2    …  2.835  2.835  2.320  6.701  6.959  5.412  2.577  3.608  7.990
3    …  3.822  3.025  4.140  4.459  7.962  6.847  2.070  3.981  8.280
4    …  4.607  3.263  6.718  2.303  5.950  5.950  2.687  2.687  7.869
5    …  5.422  2.410  6.928  3.916  8.735  5.120  2.108  3.916  6.928
..   …    …      …      …      …      …      …      …      …      …
995  …  3.097  3.982  5.310  8.850  4.425  2.655  2.212  2.212  4.867
```

```
996  …  5.267  2.420  5.338  4.057  8.185  6.690  1.779  4.128  7.687
997  …  4.206  1.717  2.661  4.034  9.871  5.837  1.202  1.974  7.811
998  …  4.680  1.724  1.478  5.665  9.113  5.172  1.724  1.478  4.433
999  …  4.588  2.124  3.483  8.496  8.156  8.666  1.444  3.398  6.882

          X20
0    efectores
2    efectores
3    efectores
4    efectores
5    efectores
..         …
995  efectores
996  efectores
997  efectores
998  efectores
999  efectores

[864 rows x 21 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	864.000000	864.000000	864.000000	864.000000	864.000000	864.000000	
mean	8.106328	5.924609	3.863396	5.823760	1.397188	6.085594	
std	1.992971	1.875144	1.265167	1.655558	0.876975	1.922951	
min	2.695000	1.159000	0.662000	0.000000	0.000000	0.746000	
25%	6.926000	4.627500	3.052250	4.852750	0.758500	4.891500	
50%	8.031000	5.882000	3.742500	5.840000	1.254500	5.972500	
75%	9.325250	7.111000	4.563750	6.847500	1.868250	7.224500	
max	14.844000	12.097000	8.209000	11.194000	5.263000	12.763000	

	X6	X7	X8	X9	X10	X11	\
count	864.000000	864.000000	864.000000	864.000000	864.000000	864.000000	
mean	3.912611	6.798045	2.403172	5.286729	8.990286	5.076084	
std	1.410652	1.985078	0.997205	1.535960	2.073134	1.874329	
min	0.000000	0.000000	0.000000	0.877000	3.060000	0.877000	
25%	2.905500	5.338000	1.667000	4.218500	7.585750	3.713750	
50%	3.821500	6.679000	2.363000	5.169000	9.015500	4.911500	
75%	4.729500	8.122750	3.003250	6.206000	10.346500	6.255250	
max	9.231000	13.158000	5.660000	10.317000	15.033000	11.935000	

	X12	X13	X14	X15	X16	X17	\
count	864.000000	864.000000	864.000000	864.000000	864.000000	864.000000	
mean	2.277344	3.890914	5.815551	7.822388	5.937105	1.605748	

std	0.907029	1.342388	1.951308	2.134756	1.498577	0.862203
min	0.000000	0.000000	0.000000	1.835000	0.000000	0.000000
25%	1.670500	3.019000	4.461250	6.333000	5.069000	0.951250
50%	2.201000	3.826000	5.646000	7.624500	5.805000	1.498500
75%	2.821000	4.679250	6.936750	9.120750	6.716000	2.164250
max	5.252000	8.244000	12.712000	14.873000	12.018000	4.333000

	X18	X19
count	864.000000	864.000000
mean	2.879271	6.103829
std	1.091351	1.609175
min	0.000000	1.064000
25%	2.128000	4.993500
50%	2.784500	6.042500
75%	3.601250	7.074000
max	6.593000	11.667000

no_efectores

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

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9	\
0	9.804	9.804	4.902	6.863	1.961	3.922	0.980	5.882	4.902	6.863	
1	7.281	5.201	2.972	8.172	1.189	5.944	2.229	9.212	3.269	4.309	
2	8.932	5.243	6.408	5.437	0.194	6.990	3.883	7.379	1.553	6.214	
3	9.231	7.385	3.385	6.923	2.000	5.385	6.000	4.923	2.923	5.846	
4	7.483	4.592	3.912	6.463	2.041	4.082	5.782	4.932	2.041	5.442	
..	…	…	…	…	…	…	…	…	…		
991	6.761	5.352	4.225	3.380	2.254	3.662	1.690	6.197	1.127	10.141	
994	8.707	4.741	3.534	4.569	1.466	3.103	4.655	5.259	1.897	7.155	
995	6.114	5.523	3.748	5.128	0.394	5.325	4.142	5.523	2.959	6.312	
997	6.806	7.135	2.415	7.245	2.964	5.598	3.513	8.342	2.086	4.061	
998	6.475	6.763	2.446	6.331	1.295	7.194	3.453	4.029	1.583	5.755	

	…	X11	X12	X13	X14	X15	X16	X17	X18	X19	\
0	…	0.980	2.941	2.941	5.882	8.824	4.902	0.980	1.961	8.824	
1	…	7.132	2.080	3.418	5.795	9.212	4.755	1.040	2.823	5.052	
2	…	6.602	3.301	3.883	3.883	5.631	4.854	1.359	2.524	5.825	
3	…	2.923	2.462	4.000	6.308	7.231	5.385	1.846	2.615	4.308	
4	…	2.551	3.061	2.721	7.993	10.714	5.782	1.531	2.211	7.143	
..	…	…	…	…	…	…	…	…	…		
991	…	3.662	3.944	5.634	3.380	8.169	7.606	2.254	3.380	7.606	
994	…	2.759	2.069	4.310	5.259	10.603	6.207	1.121	2.328	6.638	
995	…	5.325	3.550	5.128	7.692	5.720	5.720	1.578	3.550	6.312	
997	…	5.269	1.756	4.610	5.928	7.245	5.708	1.537	3.952	5.049	

```
998  …  3.453  2.734  2.446  10.647  13.237  7.338  0.863  2.014  5.180

              X20
0     no_efectores
1     no_efectores
2     no_efectores
3     no_efectores
4     no_efectores
..             …
991   no_efectores
994   no_efectores
995   no_efectores
997   no_efectores
998   no_efectores

[844 rows x 21 columns]
```

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

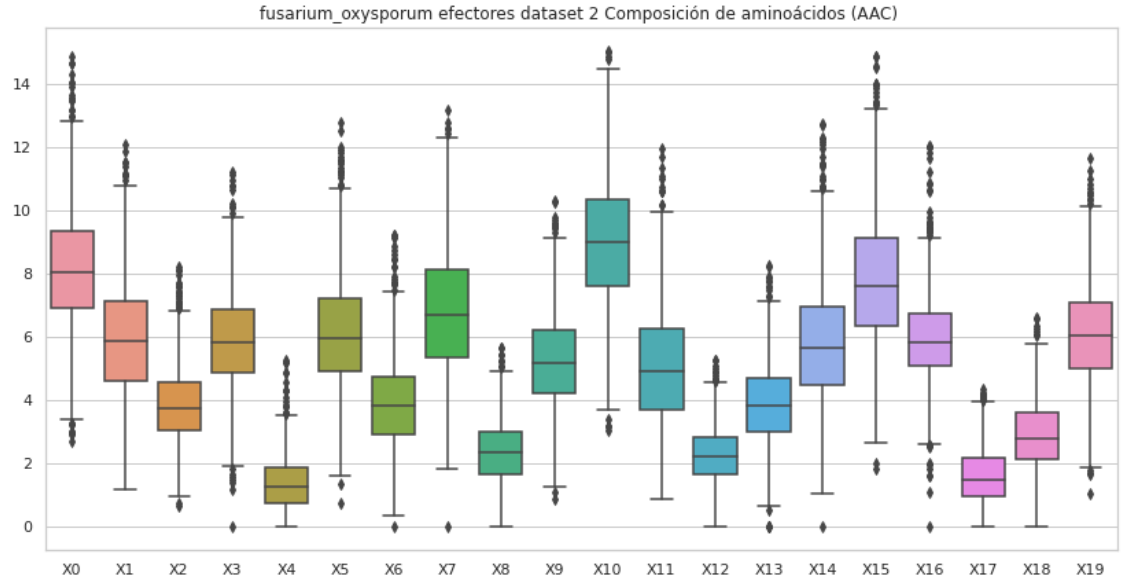
	X0	X1	X2	X3	X4	X5
count	844.000000	844.000000	844.000000	844.00000	844.000000	844.000000
mean	8.042955	5.934887	3.756224	5.82386	1.454709	5.980408
std	2.005369	1.886203	1.184487	1.59332	0.922627	1.877577
min	1.714000	1.176000	0.000000	1.03600	0.000000	0.881000
25%	6.792750	4.654000	2.974250	4.87450	0.784000	4.738000
50%	7.920500	5.785000	3.650000	5.85150	1.333000	5.869500
75%	9.275750	7.148000	4.450000	6.79650	2.000000	7.085250
max	14.815000	12.874000	7.674000	10.65100	5.155000	12.766000

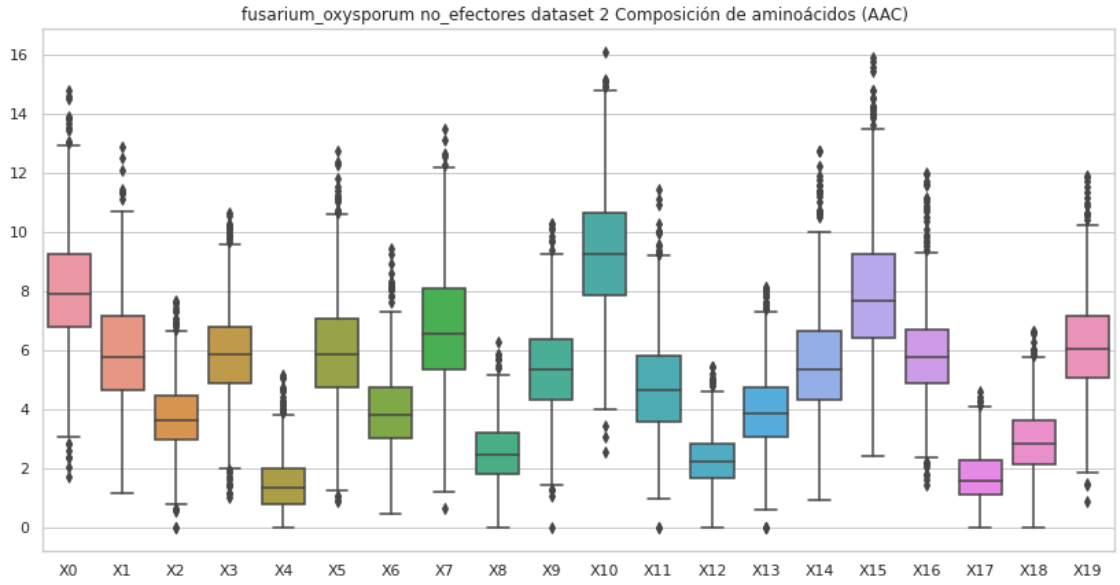
	X6	X7	X8	X9	X10	X11
count	844.000000	844.000000	844.000000	844.000000	844.000000	844.000000
mean	3.946046	6.704528	2.532570	5.375812	9.260762	4.797647
std	1.376129	1.993224	1.038517	1.606691	2.113081	1.749186
min	0.459000	0.641000	0.000000	0.000000	2.564000	0.000000
25%	3.006000	5.337000	1.807000	4.344250	7.874500	3.561500
50%	3.824500	6.548500	2.459000	5.338500	9.268500	4.640500
75%	4.762000	8.084750	3.197000	6.350250	10.657000	5.826250
max	9.444000	13.497000	6.269000	10.280000	16.092000	11.429000

	X12	X13	X14	X15	X16	X17
count	844.000000	844.000000	844.00000	844.000000	844.000000	844.000000
mean	2.297929	3.934895	5.58616	7.934218	5.862477	1.721083
std	0.883050	1.353332	1.83518	2.214282	1.550701	0.882219
min	0.000000	0.000000	0.93000	2.439000	1.429000	0.000000
25%	1.667000	3.051250	4.33125	6.433000	4.900750	1.099000

50%	2.234500	3.844500	5.35150	7.680000	5.749500	1.598000
75%	2.857000	4.764000	6.66700	9.277250	6.683000	2.299000
max	5.426000	8.120000	12.73500	15.929000	12.000000	4.615000

	X18	X19
count	844.000000	844.000000
mean	2.918155	6.134758
std	1.123699	1.716813
min	0.000000	0.885000
25%	2.136250	5.050500
50%	2.838500	6.061000
75%	3.606000	7.179500
max	6.645000	11.898000





3 Composición de pseudo aminoácidos (PseAAC) hidro_mass

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

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

    if etiq == "efectores":
        df=PseAAC_hidro_mass_efec

    if etiq == "no_efectores":
        df=PseAAC_hidro_mass_no_efec

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

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

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

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.049017  0.005199  0.037877  0.043818  0.026736  0.036391  0.020052
1    0.018666  0.028000  0.028000  0.018666  0.000000  0.065332  0.009333
2    0.057162  0.005532  0.046098  0.051630  0.016595  0.055318  0.014751
3    0.023804  0.006859  0.007666  0.009683  0.010490  0.023804  0.003228
4    0.031263  0.006033  0.008776  0.012615  0.019197  0.030166  0.003839
..        …         …         …         …         …         …         …
995  0.061750  0.012350  0.043225  0.052487  0.037050  0.058662  0.015437
996  0.027739  0.004666  0.019184  0.019184  0.019443  0.026961  0.008814
997  0.043582  0.005301  0.016490  0.021497  0.009129  0.025619  0.005889
998  0.030239  0.009240  0.015959  0.024359  0.005040  0.028559  0.010920
999  0.037467  0.003747  0.039808  0.018733  0.019202  0.028568  0.007025

           X7        X8        X9  …       X74       X75       X76  \
0    0.034163  0.027479  0.054216  … -0.012173 -0.006541  0.018043
1    0.046666  0.028000  0.037333  … -0.084218 -0.017884 -0.024786
2    0.031347  0.020283  0.073757  …  0.007499  0.005852  0.018213
3    0.019770  0.009683  0.025015  …  0.012570  0.005405  0.019372
4    0.021939  0.013163  0.024681  …  0.006902 -0.001628  0.007626
..        …         …         …  …  …       …         …         …
995  0.040137  0.021612  0.049400  …  0.017878  0.014336  0.028260
996  0.020739  0.019184  0.032664  …  0.004471  0.012304  0.009238
997  0.019141  0.014429  0.032098  …  0.002075 -0.000994  0.030465
998  0.014279  0.015959  0.025199  … -0.013267  0.013750  0.003449
999  0.029973  0.025290  0.048238  … -0.012037 -0.000977  0.026166

          X77       X78       X79       X80       X81       X82         X83
0    0.002286  0.014598  0.006827  0.009042  0.008263  0.007744  efectores
1   -0.049621  0.037887  0.055870 -0.014832  0.024674 -0.007702  efectores
2   -0.014128 -0.012039  0.019731 -0.036760 -0.010441  0.037169  efectores
3    0.006167 -0.002268  0.018752  0.009141  0.004660  0.016709  efectores
4    0.000073 -0.006669  0.026447  0.013518  0.000829  0.000559  efectores
..        …         …         …         …         …         …         …
995 -0.029360 -0.017615  0.019567  0.002802  0.023193  0.003541  efectores
996  0.007979  0.008603  0.015977  0.006635  0.002501  0.013982  efectores
```

```
997 -0.005345 -0.003452  0.036736  0.010436  0.013584  0.029576  efectores
998 -0.009595  0.014570  0.005855  0.014910  0.018726  0.009851  efectores
999  0.008652 -0.000428  0.027660  0.017809  0.009630  0.027741  efectores

[1000 rows x 84 columns]
```

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

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.037693	0.007699	0.028485	0.030115	0.019337	
std	0.018183	0.010734	0.018401	0.023419	0.015385	
min	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.026880	0.002860	0.017033	0.017817	0.011056	
50%	0.035156	0.005596	0.025590	0.026523	0.016622	
75%	0.044432	0.009377	0.035105	0.036432	0.024187	
max	0.210136	0.245916	0.184546	0.369359	0.246062	

	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.031728	0.012246	0.025492	0.024508	0.043797	…	
std	0.015880	0.009416	0.016540	0.017335	0.028794	…	
min	0.000000	0.000000	0.000000	0.000000	0.000000	…	
25%	0.021300	0.006063	0.015765	0.014424	0.027170	…	
50%	0.029441	0.010410	0.022749	0.022169	0.039346	…	
75%	0.038993	0.015881	0.031429	0.030819	0.054323	…	
max	0.152826	0.081972	0.182354	0.305652	0.401169	…	

	X73	X74	X75	X76	X77	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.014569	0.000727	0.005836	0.013921	0.001982	
std	0.022094	0.029628	0.023212	0.027924	0.029642	
min	-0.231098	-0.306590	-0.192012	-0.578071	-0.412480	
25%	0.007135	-0.007756	-0.001804	0.006278	-0.006049	
50%	0.014475	0.002728	0.006298	0.015395	0.003433	
75%	0.023685	0.010846	0.015699	0.023869	0.011114	
max	0.310077	0.492553	0.222378	0.188335	0.310771	

	X78	X79	X80	X81	X82
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.005793	0.014913	0.000894	0.006965	0.014120
std	0.023357	0.022699	0.046945	0.024768	0.025443
min	-0.317526	-0.167647	-1.174354	-0.325715	-0.333062
25%	-0.002304	0.006816	-0.005646	-0.001649	0.006076
50%	0.006526	0.015371	0.003679	0.006689	0.015125

```
75%       0.014776     0.024761     0.012359     0.017626     0.024601
max       0.117136     0.298763     0.527219     0.246257     0.324711

[8 rows x 83 columns]



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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.046961  0.009392  0.032873  0.018784  0.014088  0.028177  0.023481
1    0.026893  0.004391  0.030186  0.021953  0.012623  0.034027  0.012074
2    0.043292  0.000941  0.026352  0.033881  0.018823  0.035763  0.007529
3    0.049580  0.010742  0.037185  0.028922  0.021485  0.026443  0.015700
4    0.027967  0.007628  0.024154  0.015255  0.010170  0.018433  0.007628
..        …         …         …         …         …         …         …
995  0.051697  0.003335  0.043359  0.045026  0.043359  0.046694  0.025015
996  0.013501  0.003375  0.017719  0.021095  0.005063  0.013501  0.005906
997  0.025871  0.011267  0.027541  0.021281  0.017526  0.031713  0.007928
998  0.025325  0.005065  0.024763  0.028139  0.009567  0.015758  0.006191
999  0.026112  0.000000  0.041779  0.041779  0.015667  0.020889  0.036556

           X7        X8        X9  …       X74       X75       X76  \
0    0.032873  0.004696  0.028177  … -0.065594 -0.051234 -0.014171
1    0.015916  0.026344  0.032930  …  0.004169  0.005404  0.018049
2    0.030116  0.031999  0.047998  …  0.018400 -0.002067  0.017143
3    0.031401  0.015700  0.047927  …  0.001833  0.011921 -0.006465
4    0.020340  0.009534  0.035595  … -0.001226  0.004856  0.023020
..        …         …         …  …         …         …         …
995  0.053364  0.045026  0.086717  …  0.009087  0.015268  0.013564
996  0.013501  0.031220  0.017719  …  0.009313  0.017000  0.000313
997  0.015439  0.020029  0.033382  … -0.014896  0.003610  0.017249
998  0.022511  0.013507  0.026451  … -0.006449 -0.004169  0.020665
999  0.078335  0.031334  0.067890  …  0.012646  0.012020 -0.015515

           X77       X78       X79       X80       X81       X82           X83
0   -0.027507 -0.011433 -0.002124 -0.076538 -0.050540  0.028355  no_efectores
1   -0.000861  0.001844  0.022413 -0.004350  0.007434  0.001116  no_efectores
2   -0.013085  0.007401  0.025457 -0.006637  0.004961  0.015868  no_efectores
3    0.018711  0.001084  0.001750 -0.008244 -0.000153  0.002286  no_efectores
4    0.000562 -0.003168  0.013907 -0.013586 -0.012348  0.015879  no_efectores
..        …         …         …         …         …         …           …
995 -0.007659  0.009246  0.013876  0.009466  0.019952  0.004590  no_efectores
996  0.001716  0.014667  0.001910  0.010191  0.009810  0.005489  no_efectores
997 -0.003797 -0.009057  0.011939 -0.008373 -0.003074  0.018235  no_efectores
```

```
998  0.011834  0.017942  0.013959  0.001401  0.013446  0.034327  no_efectores
999  0.000974  0.015326 -0.003926  0.038609  0.032120 -0.013148  no_efectores

[1000 rows x 84 columns]
```

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

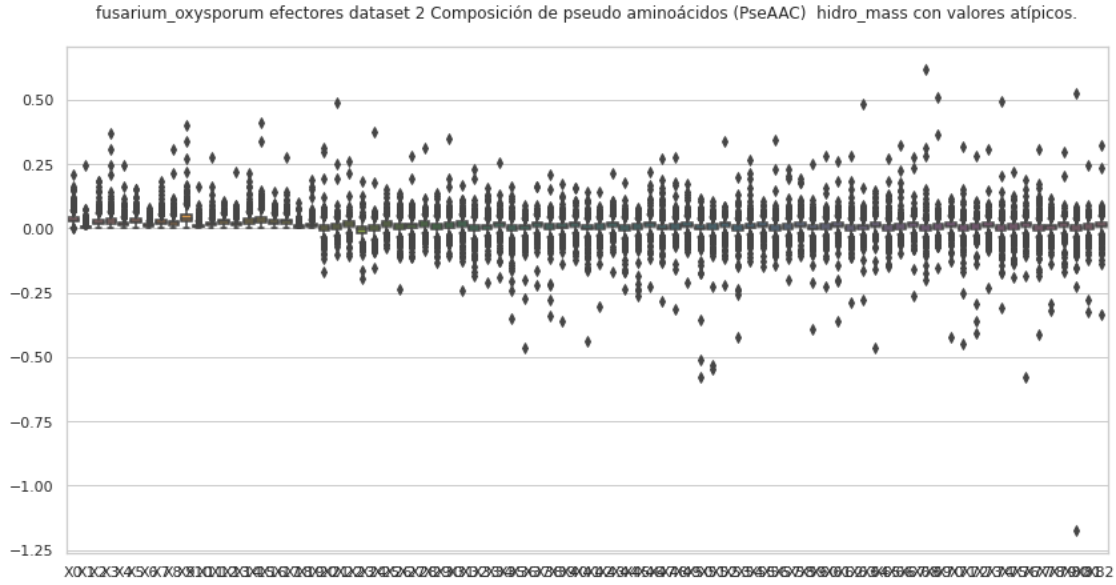
	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.038820	0.008090	0.028547	0.029796	0.019818	
std	0.024271	0.012842	0.016242	0.018675	0.016555	
min	0.003184	0.000000	0.000000	0.000000	0.000000	
25%	0.027636	0.003108	0.017435	0.017912	0.011217	
50%	0.036849	0.005708	0.026521	0.027463	0.017480	
75%	0.046480	0.010370	0.037020	0.037161	0.024863	
max	0.494502	0.329668	0.164834	0.259237	0.329668	

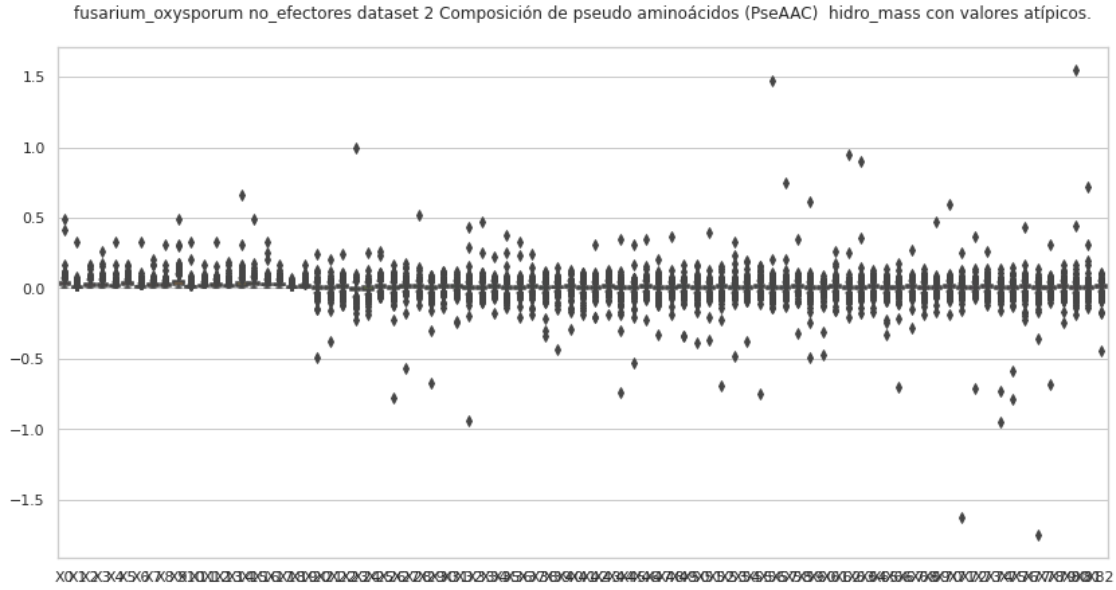
	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.032237	0.013593	0.026022	0.023777	0.045706	…	
std	0.015901	0.014470	0.015898	0.016727	0.030540	…	
min	0.000000	0.000000	0.000000	0.000000	0.002289	…	
25%	0.021954	0.006373	0.016347	0.013665	0.027539	…	
50%	0.030021	0.011272	0.023948	0.021757	0.041725	…	
75%	0.039838	0.017002	0.032953	0.030194	0.056180	…	
max	0.164834	0.329668	0.207389	0.311084	0.494502	…	

	X73	X74	X75	X76	X77	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.014727	0.001010	0.005333	0.014136	-0.000335	
std	0.020403	0.043177	0.036385	0.026613	0.060652	
min	-0.116163	-0.952912	-0.783490	-0.220252	-1.748575	
25%	0.005829	-0.006117	-0.001390	0.004264	-0.006420	
50%	0.014648	0.003348	0.006637	0.014951	0.002731	
75%	0.023597	0.012174	0.015554	0.024340	0.011839	
max	0.263218	0.140820	0.143045	0.429741	0.140338	

	X78	X79	X80	X81	X82
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.006019	0.014253	0.004412	0.006839	0.012743
std	0.030090	0.022632	0.055389	0.031568	0.024631
min	-0.685965	-0.244569	-0.181389	-0.140044	-0.446698
25%	-0.002176	0.005570	-0.006413	-0.002196	0.005304
50%	0.006886	0.014892	0.003279	0.005237	0.013781
75%	0.015009	0.024274	0.011687	0.014433	0.023126

```
max       0.310217      0.168812      1.547452      0.717223      0.111275

[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
fusarium_oxysporum dataset 2, sin valores atípicos.
Valores del documento csv.
```

	X0	X1	X2	X3	X4	X5	X6	\
0	0.049017	0.005199	0.037877	0.043818	0.026736	0.036391	0.020052	
2	0.057162	0.005532	0.046098	0.051630	0.016595	0.055318	0.014751	
3	0.023804	0.006859	0.007666	0.009683	0.010490	0.023804	0.003228	
4	0.031263	0.006033	0.008776	0.012615	0.019197	0.030166	0.003839	
5	0.030246	0.008344	0.008344	0.016687	0.023988	0.027117	0.004172	
..	…	…	…	…	…	…	…	
995	0.061750	0.012350	0.043225	0.052487	0.037050	0.058662	0.015437	
996	0.027739	0.004666	0.019184	0.019184	0.019443	0.026961	0.008814	
997	0.043582	0.005301	0.016490	0.021497	0.009129	0.025619	0.005889	
998	0.030239	0.009240	0.015959	0.024359	0.005040	0.028559	0.010920	
999	0.037467	0.003747	0.039808	0.018733	0.019202	0.028568	0.007025	

	X7	X8	X9	…	X74	X75	X76	\
0	0.034163	0.027479	0.054216	…	-0.012173	-0.006541	0.018043	
2	0.031347	0.020283	0.073757	…	0.007499	0.005852	0.018213	
3	0.019770	0.009683	0.025015	…	0.012570	0.005405	0.019372	
4	0.021939	0.013163	0.024681	…	0.006902	-0.001628	0.007626	
5	0.025031	0.018773	0.029203	…	0.012372	0.008271	0.030624	
..	…	…	…	…	…	…	…	
995	0.040137	0.021612	0.049400	…	0.017878	0.014336	0.028260	
996	0.020739	0.019184	0.032664	…	0.004471	0.012304	0.009238	
997	0.019141	0.014429	0.032098	…	0.002075	-0.000994	0.030465	
998	0.014279	0.015959	0.025199	…	-0.013267	0.013750	0.003449	
999	0.029973	0.025290	0.048238	…	-0.012037	-0.000977	0.026166	

	X77	X78	X79	X80	X81	X82	X83
0	0.002286	0.014598	0.006827	0.009042	0.008263	0.007744	efectores
2	-0.014128	-0.012039	0.019731	-0.036760	-0.010441	0.037169	efectores
3	0.006167	-0.002268	0.018752	0.009141	0.004660	0.016709	efectores
4	0.000073	-0.006669	0.026447	0.013518	0.000829	0.000559	efectores
5	0.012668	0.005181	0.018296	0.009943	0.001094	0.003595	efectores
..	…	…	…	…	…	…	
995	-0.029360	-0.017615	0.019567	0.002802	0.023193	0.003541	efectores
996	0.007979	0.008603	0.015977	0.006635	0.002501	0.013982	efectores
997	-0.005345	-0.003452	0.036736	0.010436	0.013584	0.029576	efectores
998	-0.009595	0.014570	0.005855	0.014910	0.018726	0.009851	efectores
999	0.008652	-0.000428	0.027660	0.017809	0.009630	0.027741	efectores

[881 rows x 84 columns]

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

	X0	X1	X2	X3	X4	X5	\
count	881.000000	881.000000	881.000000	881.000000	881.000000	881.000000	
mean	0.034771	0.006359	0.025459	0.026591	0.016897	0.029220	
std	0.012156	0.005116	0.012132	0.012834	0.008939	0.011435	
min	0.001052	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.026316	0.002781	0.016410	0.017008	0.010548	0.020636	
50%	0.033752	0.005305	0.024582	0.025061	0.015802	0.028490	
75%	0.042178	0.008360	0.033089	0.034218	0.022294	0.036785	
max	0.082983	0.034137	0.071141	0.088041	0.058524	0.071001	

	X6	X7	X8	X9	…	X73	\
count	881.000000	881.000000	881.000000	881.000000	…	881.000000	
mean	0.010787	0.022691	0.022153	0.039315	…	0.015772	
std	0.006721	0.010863	0.010661	0.019189	…	0.012469	
min	0.000000	0.000000	0.001514	0.000000	…	-0.025242	
25%	0.005780	0.015113	0.014055	0.025822	…	0.008045	
50%	0.009943	0.021462	0.020980	0.036724	…	0.015059	
75%	0.014475	0.029253	0.028718	0.049989	…	0.023361	
max	0.037458	0.062481	0.060397	0.122586	…	0.070502	

	X74	X75	X76	X77	X78	X79	\
count	881.000000	881.000000	881.000000	881.000000	881.000000	881.000000	
mean	0.002127	0.007307	0.015468	0.002277	0.006583	0.016055	
std	0.013837	0.013501	0.013522	0.015124	0.013365	0.013119	
min	-0.054824	-0.060081	-0.045880	-0.064427	-0.047444	-0.029783	
25%	-0.005721	-0.000761	0.007393	-0.005257	-0.001075	0.007710	
50%	0.003134	0.006855	0.015582	0.003440	0.006358	0.015717	
75%	0.010632	0.015070	0.023602	0.010093	0.014147	0.024213	
max	0.047183	0.057272	0.072742	0.080243	0.071258	0.079364	

	X80	X81	X82
count	881.000000	881.000000	881.000000
mean	0.003680	0.008640	0.015264
std	0.015175	0.014612	0.013490
min	-0.065545	-0.044710	-0.050588
25%	-0.004474	-0.000128	0.007243
50%	0.004317	0.006981	0.015363
75%	0.012035	0.017241	0.023894
max	0.077721	0.065709	0.067005

[8 rows x 83 columns]

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.046961  0.009392  0.032873  0.018784  0.014088  0.028177  0.023481
1    0.026893  0.004391  0.030186  0.021953  0.012623  0.034027  0.012074
2    0.043292  0.000941  0.026352  0.033881  0.018823  0.035763  0.007529
3    0.049580  0.010742  0.037185  0.028922  0.021485  0.026443  0.015700
4    0.027967  0.007628  0.024154  0.015255  0.010170  0.018433  0.007628
..        …         …         …         …         …         …         …
994  0.029141  0.004905  0.015292  0.010387  0.014426  0.017600  0.006348
995  0.051697  0.003335  0.043359  0.045026  0.043359  0.046694  0.025015
996  0.013501  0.003375  0.017719  0.021095  0.005063  0.013501  0.005906
997  0.025871  0.011267  0.027541  0.021281  0.017526  0.031713  0.007928
998  0.025325  0.005065  0.024763  0.028139  0.009567  0.015758  0.006191

           X7        X8        X9  …       X74       X75       X76  \
0    0.032873  0.004696  0.028177  … -0.065594 -0.051234 -0.014171
1    0.015916  0.026344  0.032930  …  0.004169  0.005404  0.018049
2    0.030116  0.031999  0.047998  …  0.018400 -0.002067  0.017143
3    0.031401  0.015700  0.047927  …  0.001833  0.011921 -0.006465
4    0.020340  0.009534  0.035595  … -0.001226  0.004856  0.023020
..        …         …         …  …         …         …         …
994  0.023948  0.009233  0.045588  … -0.003859 -0.001870  0.019812
995  0.053364  0.045026  0.086717  …  0.009087  0.015268  0.013564
996  0.013501  0.031220  0.017719  …  0.009313  0.017000  0.000313
997  0.015439  0.020029  0.033382  … -0.014896  0.003610  0.017249
998  0.022511  0.013507  0.026451  … -0.006449 -0.004169  0.020665

           X77       X78       X79       X80       X81       X82           X83
0   -0.027507 -0.011433 -0.002124 -0.076538 -0.050540  0.028355  no_efectores
1   -0.000861  0.001844  0.022413 -0.004350  0.007434  0.001116  no_efectores
2   -0.013085  0.007401  0.025457 -0.006637  0.004961  0.015868  no_efectores
3    0.018711  0.001084  0.001750 -0.008244 -0.000153  0.002286  no_efectores
4    0.000562 -0.003168  0.013907 -0.013586 -0.012348  0.015879  no_efectores
..        …         …         …         …         …         …           …
994  0.016560  0.002528  0.013700  0.008737  0.001876  0.017625  no_efectores
995 -0.007659  0.009246  0.013876  0.009466  0.019952  0.004590  no_efectores
996  0.001716  0.014667  0.001910  0.010191  0.009810  0.005489  no_efectores
997 -0.003797 -0.009057  0.011939 -0.008373 -0.003074  0.018235  no_efectores
998  0.011834  0.017942  0.013959  0.001401  0.013446  0.034327  no_efectores

[892 rows x 84 columns]
```

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

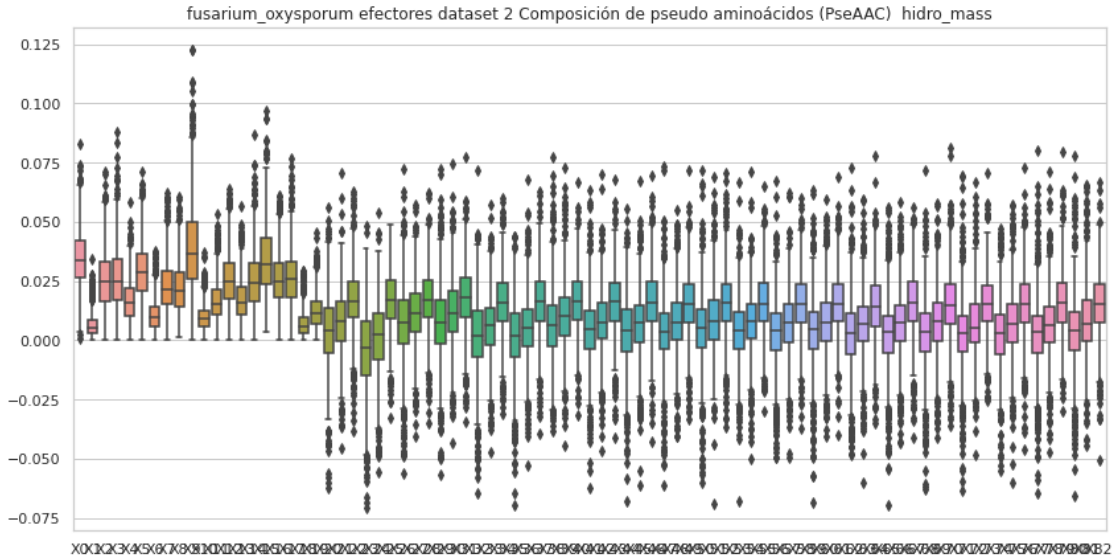
	X0	X1	X2	X3	X4	X5	\
count	892.000000	892.000000	892.000000	892.000000	892.000000	892.000000	
mean	0.036038	0.006857	0.026572	0.027247	0.017785	0.029977	
std	0.012698	0.005457	0.012656	0.013546	0.009854	0.011908	
min	0.003184	0.000000	0.000000	0.000000	0.000000	0.002064	
25%	0.026876	0.003050	0.016930	0.017348	0.010823	0.021527	
50%	0.035563	0.005517	0.025676	0.026021	0.016738	0.028962	
75%	0.044449	0.009532	0.035089	0.034699	0.023598	0.037585	
max	0.083038	0.037891	0.075815	0.080338	0.067746	0.073642	

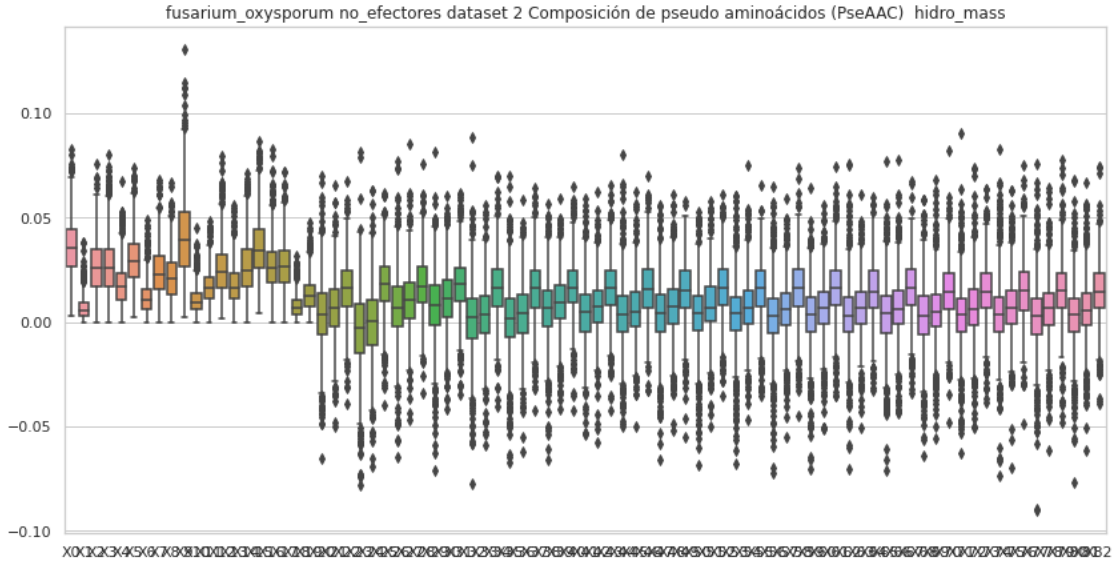
	X6	X7	X8	X9	…	X73	\
count	892.000000	892.000000	892.000000	892.000000	…	892.000000	
mean	0.011793	0.024207	0.021723	0.041471	…	0.014751	
std	0.007540	0.011885	0.010920	0.020085	…	0.013581	
min	0.000000	0.000000	0.000000	0.002289	…	-0.038546	
25%	0.006075	0.016016	0.013370	0.026474	…	0.006742	
50%	0.010706	0.023007	0.020837	0.039503	…	0.014714	
75%	0.015667	0.031456	0.028397	0.052881	…	0.023209	
max	0.048906	0.067885	0.066647	0.130616	…	0.063500	

	X74	X75	X76	X77	X78	X79	\
count	892.000000	892.000000	892.000000	892.000000	892.000000	892.000000	
mean	0.003854	0.007342	0.015127	0.002239	0.006532	0.015147	
std	0.016253	0.013848	0.014605	0.016474	0.013734	0.014718	
min	-0.073883	-0.069527	-0.048406	-0.090213	-0.061624	-0.048386	
25%	-0.004834	-0.000707	0.005747	-0.005687	-0.001299	0.006695	
50%	0.003873	0.006755	0.015369	0.003016	0.006831	0.014961	
75%	0.012039	0.015057	0.024044	0.011031	0.014153	0.023608	
max	0.083045	0.062882	0.075064	0.075903	0.064407	0.077792	

	X80	X81	X82
count	892.000000	892.000000	892.000000
mean	0.003122	0.005909	0.014945
std	0.014913	0.014109	0.013836
min	-0.076538	-0.061300	-0.039295
25%	-0.004909	-0.001492	0.006697
50%	0.003493	0.005354	0.014363
75%	0.011235	0.013709	0.023126
max	0.068295	0.066655	0.074725

[8 rows x 83 columns]





4 Composición de pseudo aminoácidos (PseAAC) mass

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

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

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

    if etiq == "efectores":
        df=PseAAC_mass_efec

    if etiq == "no_efectores":
        df=PseAAC_mass_no_efec

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

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

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.055800  0.005918  0.043118  0.049882  0.030436  0.041427  0.022827
1    0.016610  0.024915  0.024915  0.016610  0.000000  0.058135  0.008305
2    0.051787  0.005012  0.041764  0.046775  0.015035  0.050116  0.013364
3    0.037114  0.010694  0.011952  0.015097  0.016355  0.037114  0.005032
4    0.048177  0.009297  0.013523  0.019440  0.029583  0.046487  0.005917
..        …         …         …         …         …         …         …
995  0.058077  0.011615  0.040654  0.049365  0.034846  0.055173  0.014519
996  0.042597  0.007166  0.029460  0.029460  0.029858  0.041403  0.013536
997  0.044000  0.005351  0.016649  0.021703  0.009216  0.025865  0.005946
998  0.049110  0.015006  0.025919  0.039561  0.008185  0.046382  0.017734
999  0.035961  0.003596  0.038208  0.017980  0.018430  0.027420  0.006743

           X7        X8        X9  …       X32       X33       X34  \
0    0.038891  0.031282  0.061718  …  0.012827  0.010037  0.008688
```

```
1    0.041525  0.024915  0.033220  …  0.036455  0.077152  0.035713
2    0.028399  0.018376  0.066822  …  0.016393 -0.006829  0.014406
3    0.030824  0.015097  0.039001  …  0.030064  0.031143  0.018436
4    0.033809  0.020285  0.038035  …  0.026956  0.041875  0.012748
..        …         …        …  …        …         …
995  0.037750  0.020327  0.046461  …  0.036415  0.010441  0.020211
996  0.031849  0.029460  0.050161  …  0.006739  0.018921  0.026180
997  0.019324  0.014568  0.032406  …  0.032279  0.029603  0.028931
998  0.023191  0.025919  0.040925  …  0.055339  0.012147  0.022158
999  0.028769  0.024274  0.046300  …  0.013386  0.018552  0.024289

          X35       X36       X37       X38       X39       X40         X41
0    0.035410  0.014253  0.032839  0.020540  0.007772  0.008816  efectores
1    0.010686 -0.004543  0.023761 -0.022056  0.049715 -0.006853  efectores
2    0.025277  0.010249  0.034113  0.016500  0.017876  0.033674  efectores
3    0.028463  0.019956  0.029847  0.030203  0.029237  0.026051  efectores
4    0.041358  0.021563  0.016232  0.011751  0.040756  0.000862  efectores
..        …         …         …         …         …         …         …
995 -0.010306  0.035484  0.010641  0.026579  0.018404  0.003330  efectores
996  0.021144  0.022942  0.024138  0.014187  0.024536  0.021472  efectores
997  0.029715  0.031116  0.030245  0.030758  0.037089  0.029860  efectores
998  0.029539  0.003832  0.025266  0.005601  0.009509  0.015999  efectores
999  0.019885  0.022887  0.018162  0.025114  0.026548  0.026626  efectores

[1000 rows x 42 columns]
```

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

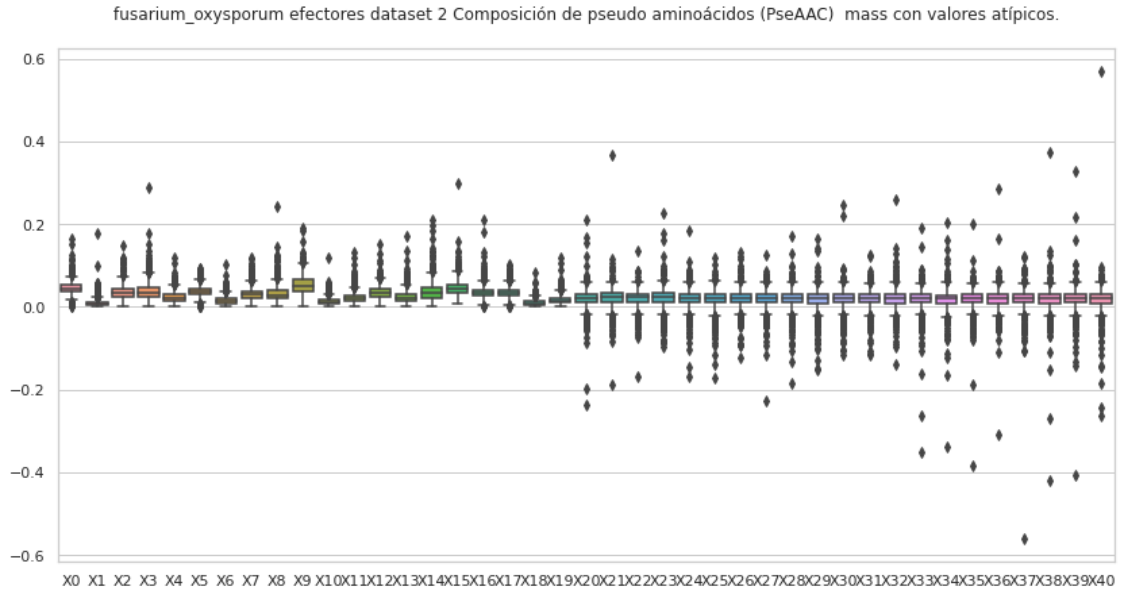
```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.046726     0.009167     0.035217     0.037916     0.023365
std       0.014955     0.009752     0.017255     0.021765     0.012926
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       0.038350     0.003926     0.023160     0.023924     0.014743
50%       0.045039     0.007321     0.033559     0.034322     0.021024
75%       0.053400     0.011612     0.044101     0.047546     0.030270
max       0.163512     0.178569     0.147100     0.287636     0.119046

                X5           X6           X7           X8           X9  …  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
mean      0.038932     0.015041     0.031119     0.031460     0.053041  …
std       0.012928     0.009817     0.014722     0.018967     0.022525  …
min       0.000000     0.000000     0.000000     0.000000     0.000000  …
25%       0.031004     0.008206     0.021377     0.019868     0.038030  …
50%       0.037694     0.013674     0.029432     0.028083     0.050288  …
```

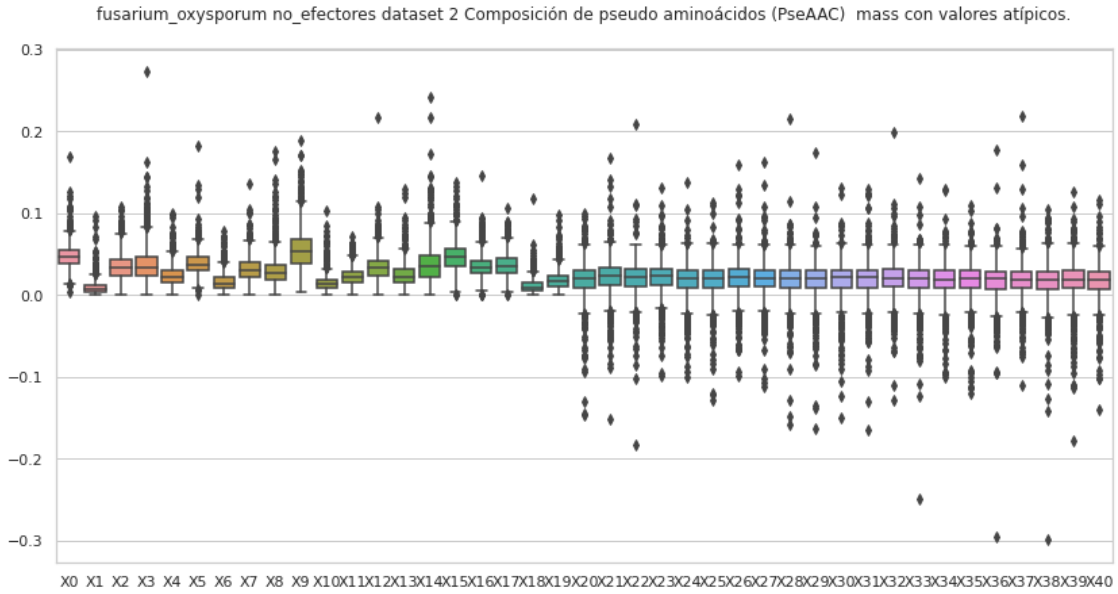
```
75%       0.044915     0.019521     0.038740     0.039338     0.065637  …
max       0.093006     0.101281     0.119046     0.243074     0.189489  …

               X31          X32          X33          X34          X35  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.018343     0.018986     0.018304     0.017352     0.018775
std       0.020442     0.022337     0.025814     0.024990     0.024849
min      -0.115348    -0.138978    -0.349605    -0.337435    -0.383779
25%       0.009843     0.009078     0.010239     0.007902     0.010867
50%       0.020509     0.020525     0.020495     0.019572     0.021186
75%       0.029729     0.029756     0.029612     0.029029     0.030446
max       0.124855     0.258357     0.190243     0.203940     0.201432

               X36          X37          X38          X39          X40
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.018065     0.017915     0.017932     0.018553     0.017571
std       0.024213     0.027653     0.029193     0.027656     0.031052
min      -0.308091    -0.559718    -0.419759    -0.406040    -0.261734
25%       0.008479     0.010117     0.009125     0.009796     0.009102
50%       0.020071     0.020187     0.020526     0.020533     0.020722
75%       0.029412     0.029608     0.029462     0.030046     0.029803
max       0.284890     0.122159     0.374266     0.327521     0.568964

[8 rows x 41 columns]
```

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

          X0        X1        X2        X3        X4        X5        X6  \
0    0.056173  0.011235  0.039321  0.022469  0.016852  0.033704  0.028087
1    0.037672  0.006151  0.042285  0.030753  0.017683  0.047667  0.016914
2    0.048131  0.001046  0.029297  0.037668  0.020927  0.039761  0.008371
3    0.062638  0.013572  0.046978  0.036539  0.027143  0.033407  0.019835
4    0.035228  0.009608  0.030424  0.019215  0.012810  0.023218  0.009608
..        …         …         …         …         …         …         …
995  0.049106  0.003168  0.041186  0.042770  0.041186  0.044354  0.023761
996  0.032909  0.008227  0.043193  0.051420  0.012341  0.032909  0.014398
997  0.031136  0.013559  0.033145  0.025612  0.021092  0.038167  0.009542
998  0.033208  0.006642  0.032470  0.036897  0.012545  0.020663  0.008117
999  0.045228  0.000000  0.072365  0.072365  0.027137  0.036182  0.063319

          X7        X8        X9  …       X32       X33       X34  \
0    0.039321  0.005617  0.033704  …  0.054415  0.056914 -0.008586
1    0.022296  0.036903  0.046129  …  0.010029  0.029965  0.011921
```

```
2    0.033483  0.035575  0.053363  …  0.031604  0.024683  0.028818
3    0.039671  0.019835  0.060550  …  0.040137  0.007439  0.009184
4    0.025620  0.012009  0.044835  …  0.005744  0.022508  0.026992
..        …         …        …  …        …         …
995  0.050691  0.042770  0.082372  …  0.019947  0.015007 -0.004795
996  0.032909  0.076102  0.043193  …  0.014022  0.025641  0.037984
997  0.018581  0.024105  0.040176  …  0.023845  0.029640  0.017231
998  0.029518  0.017711  0.034684  …  0.010390  0.031122  0.022353
999  0.135684  0.054274  0.117593  … -0.044127 -0.006965  0.055002

          X35       X36       X37       X38       X39       X40           X41
0    0.036300 -0.030908  0.027585 -0.016951 -0.002540  0.033918  no_efectores
1    0.018893  0.022658  0.016451  0.025283  0.031397  0.001563  no_efectores
2    0.026257  0.027912  0.021154  0.019059  0.028303  0.017642  no_efectores
3    0.009263  0.014699  0.016295 -0.008168  0.002211  0.002888  no_efectores
4    0.038072  0.024837  0.024668  0.028996  0.017517  0.020001  no_efectores
..        …         …         …         …         …         …           …
995 -0.003064 -0.006706 -0.003426  0.012885  0.013181  0.004360  no_efectores
996  0.031532  0.019380  0.010270  0.000763  0.004655  0.013380  no_efectores
997  0.027356  0.033582  0.012304  0.020759  0.014369  0.021946  no_efectores
998  0.029013  0.013048  0.027847  0.027096  0.018304  0.045011  no_efectores
999 -0.001349  0.014268 -0.023800 -0.026873 -0.006800 -0.022773  no_efectores

[1000 rows x 42 columns]


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

                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.047529     0.009585     0.035272     0.038104     0.023854
std       0.015152     0.009308     0.016209     0.022595     0.013343
min       0.003845     0.000000     0.000000     0.000000     0.000000
25%       0.038417     0.003971     0.023866     0.023389     0.014861
50%       0.046304     0.007512     0.033481     0.034237     0.022111
75%       0.054527     0.012618     0.044254     0.047506     0.030292
max       0.168606     0.096413     0.108236     0.272993     0.100307

                X5           X6           X7           X8           X9  …  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
mean      0.039261     0.016188     0.031727     0.030346     0.055456  …
std       0.013881     0.010762     0.014958     0.019198     0.025367  …
min       0.000000     0.000000     0.000000     0.000000     0.003789  …
25%       0.030929     0.009037     0.021494     0.018254     0.038387  …
50%       0.037639     0.014094     0.030510     0.027065     0.053382  …
75%       0.046059     0.021322     0.039740     0.037191     0.068595  …
```

```
max       0.182978     0.077998     0.135684     0.175474     0.188484  …

                X31          X32          X33          X34          X35  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.018905     0.019016     0.018271     0.017362     0.017663
std       0.022409     0.021775     0.023481     0.020960     0.022372
min      -0.163949    -0.128616    -0.249108    -0.100457    -0.119334
25%       0.009259     0.010030     0.009349     0.007977     0.009206
50%       0.021435     0.020706     0.020790     0.019182     0.020545
75%       0.030327     0.031071     0.030272     0.029894     0.030216
max       0.129153     0.198179     0.142676     0.128740     0.108840

                X36          X37          X38          X39          X40
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.016746     0.018070     0.016747     0.017057     0.016231
std       0.023604     0.021575     0.024462     0.023626     0.021907
min      -0.294574    -0.109877    -0.298567    -0.177590    -0.140426
25%       0.007161     0.009025     0.006382     0.008062     0.007300
50%       0.019629     0.019224     0.019433     0.018952     0.018457
75%       0.029320     0.028917     0.029243     0.030126     0.028507
max       0.177738     0.218809     0.104620     0.126405     0.115734

[8 rows x 41 columns]
```





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

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

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

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

    if etiq == "efectores":
        df=PseAAC_mass_efec

    if etiq == "no_efectores":
        df=PseAAC_mass_no_efec

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

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

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

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

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

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.055800  0.005918  0.043118  0.049882  0.030436  0.041427  0.022827
2    0.051787  0.005012  0.041764  0.046775  0.015035  0.050116  0.013364
3    0.037114  0.010694  0.011952  0.015097  0.016355  0.037114  0.005032
4    0.048177  0.009297  0.013523  0.019440  0.029583  0.046487  0.005917
5    0.049455  0.013643  0.013643  0.027286  0.039223  0.044339  0.006821
..        …         …         …         …         …         …         …
995  0.058077  0.011615  0.040654  0.049365  0.034846  0.055173  0.014519
996  0.042597  0.007166  0.029460  0.029460  0.029858  0.041403  0.013536
997  0.044000  0.005351  0.016649  0.021703  0.009216  0.025865  0.005946
998  0.049110  0.015006  0.025919  0.039561  0.008185  0.046382  0.017734
999  0.035961  0.003596  0.038208  0.017980  0.018430  0.027420  0.006743

           X7        X8        X9  …       X32       X33       X34  \
0    0.038891  0.031282  0.061718  …  0.012827  0.010037  0.008688
2    0.028399  0.018376  0.066822  …  0.016393 -0.006829  0.014406
3    0.030824  0.015097  0.039001  …  0.030064  0.031143  0.018436
4    0.033809  0.020285  0.038035  …  0.026956  0.041875  0.012748
5    0.040928  0.030696  0.047750  …  0.003580  0.030150 -0.012678
..        …         …         …  …         …         …         …
995  0.037750  0.020327  0.046461  …  0.036415  0.010441  0.020211
996  0.031849  0.029460  0.050161  …  0.006739  0.018921  0.026180
997  0.019324  0.014568  0.032406  …  0.032279  0.029603  0.028931
998  0.023191  0.025919  0.040925  …  0.055339  0.012147  0.022158
999  0.028769  0.024274  0.046300  …  0.013386  0.018552  0.024289
```

```
          X35       X36       X37       X38       X39       X40        X41
0    0.035410  0.014253  0.032839  0.020540  0.007772  0.008816  efectores
2    0.025277  0.010249  0.034113  0.016500  0.017876  0.033674  efectores
3    0.028463  0.019956  0.029847  0.030203  0.029237  0.026051  efectores
4    0.041358  0.021563  0.016232  0.011751  0.040756  0.000862  efectores
5    0.043403  0.022114  0.046966  0.050074  0.029916  0.005879  efectores
..        …         …         …         …         …         …
995 -0.010306  0.035484  0.010641  0.026579  0.018404  0.003330  efectores
996  0.021144  0.022942  0.024138  0.014187  0.024536  0.021472  efectores
997  0.029715  0.031116  0.030245  0.030758  0.037089  0.029860  efectores
998  0.029539  0.003832  0.025266  0.005601  0.009509  0.015999  efectores
999  0.019885  0.022887  0.018162  0.025114  0.026548  0.026626  efectores

[868 rows x 42 columns]
```

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

```
                X0          X1          X2          X3          X4          X5  \
count  868.000000  868.000000  868.000000  868.000000  868.000000  868.000000
mean     0.045277    0.007984    0.032946    0.034699    0.021963    0.037427
std      0.011502    0.005752    0.013914    0.016286    0.010125    0.010038
min      0.006599    0.000000    0.000000    0.000000    0.000000    0.005427
25%      0.038350    0.003863    0.022864    0.023143    0.014712    0.030811
50%      0.044460    0.006937    0.032177    0.033086    0.020631    0.037092
75%      0.052077    0.010609    0.041662    0.044546    0.028781    0.043347
max      0.091328    0.037973    0.086584    0.102219    0.060032    0.074553

                X6          X7          X8          X9  …         X31  \
count  868.000000  868.000000  868.000000  868.000000  …  868.000000
mean     0.013741    0.029575    0.028628    0.050866  …    0.020491
std      0.007564    0.011939    0.013788    0.019699  …    0.015073
min      0.000000    0.000775    0.002101    0.003545  …   -0.034740
25%      0.007855    0.021291    0.018917    0.037448  …    0.012172
50%      0.012950    0.028664    0.026889    0.049194  …    0.021494
75%      0.018225    0.036778    0.036230    0.063042  …    0.030137
max      0.044197    0.067510    0.087355    0.116192  …    0.074718

               X32         X33         X34         X35         X36         X37  \
count  868.000000  868.000000  868.000000  868.000000  868.000000  868.000000
mean     0.019857    0.020200    0.018981    0.020559    0.019210    0.019904
std      0.014550    0.014849    0.015156    0.015385    0.016709    0.014990
min     -0.032793   -0.054221   -0.033769   -0.040729   -0.039509   -0.050760
25%      0.011261    0.012209    0.010010    0.012404    0.010674    0.011568
50%      0.021171    0.020953    0.020185    0.021608    0.020564    0.020597
75%      0.029234    0.029251    0.029014    0.030576    0.029217    0.029506
```

```
max      0.078884    0.068542    0.079452    0.068553    0.086411    0.094252

              X38         X39         X40
count  868.000000  868.000000  868.000000
mean     0.019894    0.020305    0.019654
std      0.015322    0.015090    0.015935
min     -0.041257   -0.047882   -0.068570
25%      0.011256    0.011270    0.011194
50%      0.021208    0.021143    0.021065
75%      0.029167    0.030046    0.029564
max      0.068933    0.082645    0.078653

[8 rows x 41 columns]
```

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

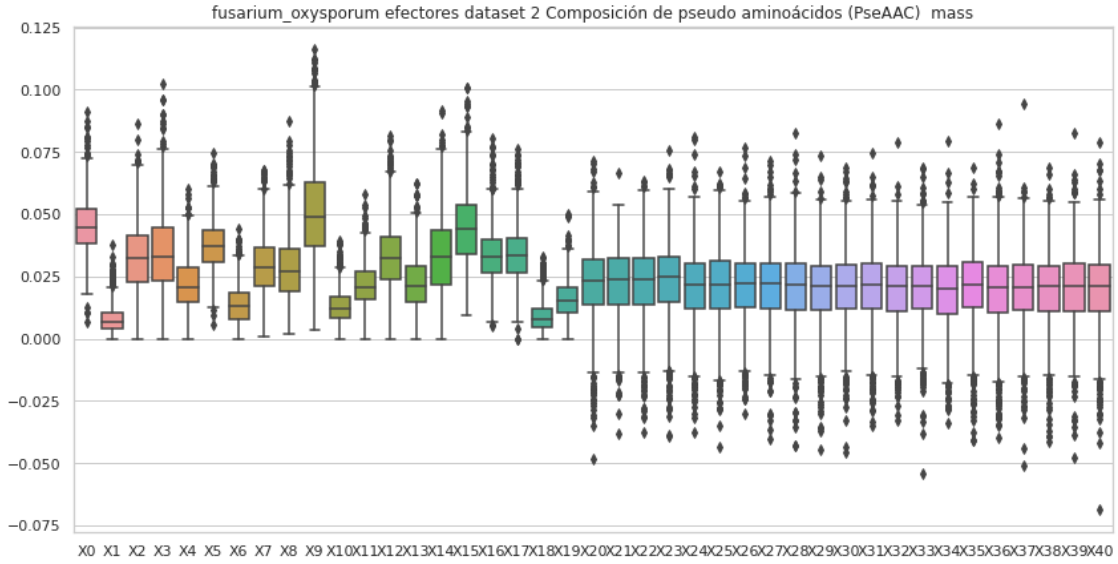
```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.056173  0.011235  0.039321  0.022469  0.016852  0.033704  0.028087
1    0.037672  0.006151  0.042285  0.030753  0.017683  0.047667  0.016914
2    0.048131  0.001046  0.029297  0.037668  0.020927  0.039761  0.008371
3    0.062638  0.013572  0.046978  0.036539  0.027143  0.033407  0.019835
4    0.035228  0.009608  0.030424  0.019215  0.012810  0.023218  0.009608
..        …         …         …         …         …         …         …
994  0.040647  0.006842  0.021330  0.014488  0.020122  0.024549  0.008854
995  0.049106  0.003168  0.041186  0.042770  0.041186  0.044354  0.023761
996  0.032909  0.008227  0.043193  0.051420  0.012341  0.032909  0.014398
997  0.031136  0.013559  0.033145  0.025612  0.021092  0.038167  0.009542
998  0.033208  0.006642  0.032470  0.036897  0.012545  0.020663  0.008117

           X7        X8        X9  …       X32       X33       X34  \
0    0.039321  0.005617  0.033704  …  0.054415  0.056914 -0.008586
1    0.022296  0.036903  0.046129  …  0.010029  0.029965  0.011921
2    0.033483  0.035575  0.053363  …  0.031604  0.024683  0.028818
3    0.039671  0.019835  0.060550  …  0.040137  0.007439  0.009184
4    0.025620  0.012009  0.044835  …  0.005744  0.022508  0.026992
..        …         …         …  …  …       …         …         …
994  0.033403  0.012878  0.063586  …  0.026737  0.024495  0.034182
995  0.050691  0.042770  0.082372  …  0.019947  0.015007 -0.004795
996  0.032909  0.076102  0.043193  …  0.014022  0.025641  0.037984
997  0.018581  0.024105  0.040176  …  0.023845  0.029640  0.017231
998  0.029518  0.017711  0.034684  …  0.010390  0.031122  0.022353

           X35       X36       X37       X38       X39       X40           X41
0    0.036300 -0.030908  0.027585 -0.016951 -0.002540  0.033918  no_efectores
```

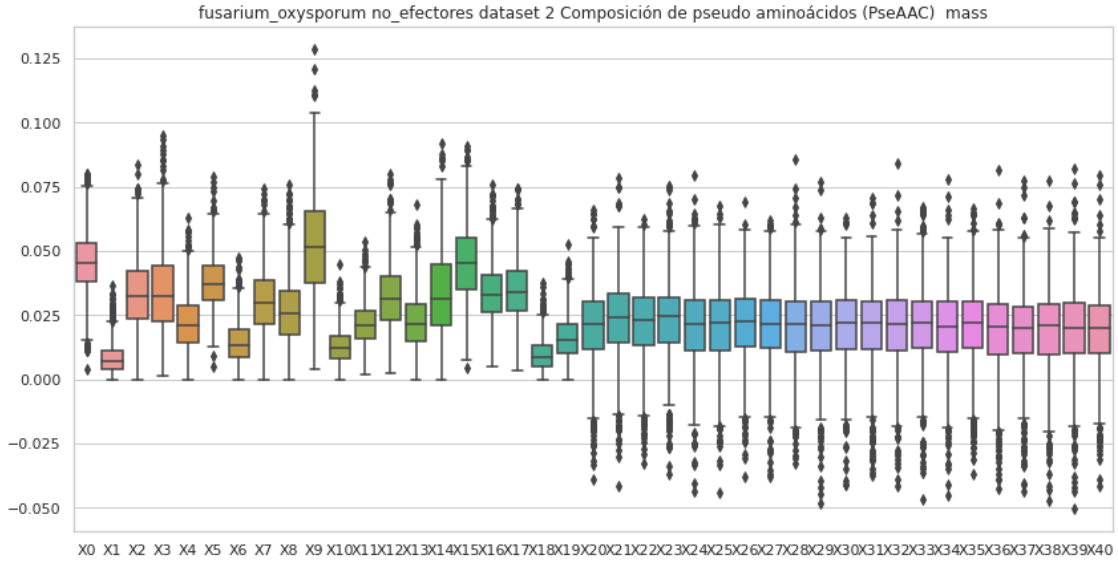
```
1    0.018893  0.022658  0.016451  0.025283  0.031397  0.001563  no_efectores
2    0.026257  0.027912  0.021154  0.019059  0.028303  0.017642  no_efectores
3    0.009263  0.014699  0.016295 -0.008168  0.002211  0.002888  no_efectores
4    0.038072  0.024837  0.024668  0.028996  0.017517  0.020001  no_efectores
..        …         …         …         …         …         …
994  0.028103  0.021625  0.019231  0.027634  0.019109  0.024584  no_efectores
995 -0.003064 -0.006706 -0.003426  0.012885  0.013181  0.004360  no_efectores
996  0.031532  0.019380  0.010270  0.000763  0.004655  0.013380  no_efectores
997  0.027356  0.033582  0.012304  0.020759  0.014369  0.021946  no_efectores
998  0.029013  0.013048  0.027847  0.027096  0.018304  0.045011  no_efectores

[860 rows x 42 columns]
```

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

```
                X0          X1          X2          X3          X4          X5  \
count  860.000000  860.000000  860.000000  860.000000  860.000000  860.000000
mean     0.045614    0.008350    0.033391    0.034242    0.022279    0.037626
std      0.012139    0.006074    0.013611    0.016040    0.010745    0.010366
min      0.003845    0.000000    0.000000    0.001529    0.000000    0.004884
25%      0.037878    0.003954    0.023422    0.022468    0.014395    0.030678
50%      0.045342    0.007180    0.032499    0.032176    0.021072    0.037146
75%      0.053094    0.011454    0.042507    0.044185    0.028725    0.044548
max      0.080078    0.036385    0.083552    0.095033    0.062953    0.078861

                X6          X7          X8          X9  …         X31  \
count  860.000000  860.000000  860.000000  860.000000  …  860.000000
mean     0.014570    0.030479    0.027241    0.052135  …    0.020278
std      0.007951    0.012611    0.013217    0.020677  …    0.015554
min      0.000000    0.000000    0.000000    0.003789  …   -0.037281
25%      0.008854    0.021335    0.017658    0.037727  …    0.011908
50%      0.013365    0.029933    0.025709    0.051557  …    0.022241
75%      0.019635    0.038689    0.034683    0.065428  …    0.030192
max      0.047314    0.074561    0.076102    0.128615  …    0.070970

               X32         X33         X34         X35         X36         X37  \
count  860.000000  860.000000  860.000000  860.000000  860.000000  860.000000
mean     0.020594    0.020808    0.019607    0.020545    0.018767    0.018526
std      0.015564    0.015670    0.015882    0.015199    0.016382    0.015832
min     -0.041392   -0.046983   -0.045341   -0.036794   -0.042484   -0.043544
25%      0.011349    0.012194    0.010606    0.012085    0.009508    0.010261
50%      0.021719    0.021829    0.020685    0.021989    0.020442    0.019854
75%      0.031124    0.030246    0.030410    0.030394    0.029307    0.028485
max      0.083995    0.067244    0.078030    0.066713    0.081748    0.077457
```

```
              X38          X39          X40
count  860.000000   860.000000   860.000000
mean     0.018854     0.019444     0.018953
std      0.016634     0.016012     0.015386
min     -0.047456    -0.050350    -0.041445
25%      0.009444     0.010304     0.010236
50%      0.020804     0.019815     0.019998
75%      0.029397     0.029819     0.028849
max      0.077275     0.081956     0.079619

[8 rows x 41 columns]
```





5 Composición de pseudo aminoácidos (PseAAC) hidro

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

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

    if etiq == "efectores":
        df=PseAAC_hidro_efec

    if etiq == "no_efectores":
        df=PseAAC_hidro_no_efec

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

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

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.066959	0.007102	0.051741	0.059857	0.036523	0.049712	0.027392	
1	0.037219	0.055829	0.055829	0.037219	0.000000	0.130268	0.018610	
2	0.093455	0.009044	0.075367	0.084411	0.027132	0.090440	0.024117	
3	0.038896	0.011207	0.012526	0.015822	0.017141	0.038896	0.005274	
4	0.049091	0.009474	0.013780	0.019809	0.030143	0.047368	0.006029	
..	…	…	…	…	…	…	…	
995	0.097315	0.019463	0.068121	0.082718	0.058389	0.092450	0.024329	
996	0.038901	0.006544	0.026904	0.026904	0.027267	0.037811	0.012361	
997	0.123614	0.015034	0.046773	0.060972	0.025892	0.072665	0.016705	
998	0.041692	0.012739	0.022004	0.033585	0.006949	0.039376	0.015056	
999	0.073557	0.007356	0.078154	0.036778	0.037698	0.056087	0.013792	

	X7	X8	X9	…	X53	X54	X55	\
0	0.046668	0.037537	0.074060	…	0.026748	0.013631	0.016151	
1	0.093049	0.055829	0.074439	…	-0.118948	0.207031	0.052478	
2	0.051249	0.033161	0.120587	…	-0.003937	0.105541	0.048624	
3	0.032304	0.015822	0.040874	…	0.002855	0.010028	0.008385	
4	0.034450	0.020670	0.038756	…	-0.005763	0.011584	0.007032	
..	…	…	…	…	…	…	…	
995	0.063255	0.034060	0.077852	…	-0.019186	0.109598	0.066215	
996	0.029085	0.026904	0.045809	…	0.013151	-0.007011	0.004067	
997	0.054290	0.040926	0.091040	…	0.032793	0.016775	0.004401	
998	0.019688	0.022004	0.034744	…	0.004772	0.016076	0.033575	
999	0.058845	0.049651	0.094704	…	0.005544	-0.007626	-0.000371	

	X56	X57	X58	X59	X60	X61	X62
0	-0.016628	-0.008935	0.003123	0.019942	0.012351	0.011288	efectores
1	-0.167925	-0.035659	-0.098940	0.075543	-0.029573	0.049199	efectores
2	0.012261	0.009567	-0.023098	-0.019683	-0.060099	-0.017069	efectores
3	0.020540	0.008831	0.010077	-0.003706	0.014937	0.007614	efectores
4	0.010838	-0.002557	0.000115	-0.010472	0.021227	0.001303	efectores
..	…	…	…	…	…	…	…
995	0.028175	0.022593	-0.046271	-0.027760	0.004416	0.036552	efectores
996	0.006271	0.017256	0.011189	0.012065	0.009305	0.003508	efectores
997	0.005886	-0.002819	-0.015159	-0.009790	0.029600	0.038530	efectores

```
998 -0.018293  0.018959 -0.013230  0.020089  0.020557  0.025819  efectores
999 -0.023631 -0.001919  0.016986 -0.000841  0.034964  0.018905  efectores

[1000 rows x 63 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.059219     0.011624     0.042641     0.044501     0.028882
std       0.029852     0.012835     0.023704     0.029100     0.021358
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       0.038586     0.004573     0.026237     0.028908     0.016775
50%       0.055337     0.008472     0.041573     0.041734     0.027064
75%       0.074445     0.014483     0.054825     0.056029     0.036978
max       0.307544     0.148994     0.284539     0.474232     0.379385

                X5           X6           X7           X8           X9  …  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
mean      0.050877     0.018295     0.038530     0.036846     0.066256  …
std       0.034495     0.013192     0.021714     0.022146     0.037787  …
min       0.000000     0.000000     0.000000     0.000000     0.000000  …
25%       0.031218     0.009134     0.024671     0.022951     0.042293  …
50%       0.045322     0.016379     0.036098     0.033409     0.062448  …
75%       0.066857     0.024302     0.049189     0.047240     0.084551  …
max       0.744969     0.148994     0.187701     0.331962     0.521655  …

               X52          X53          X54          X55          X56  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.003336     0.010058     0.001450     0.007730    -0.000283
std       0.038296     0.027752     0.044924     0.032747     0.042990
min      -0.284862    -0.211992    -0.695490    -0.452256    -0.772919
25%      -0.010544    -0.002545    -0.008979    -0.002779    -0.012241
50%       0.005106     0.011901     0.004258     0.008866     0.004292
75%       0.017424     0.024137     0.016785     0.022445     0.016252
max       0.322711     0.244436     0.213825     0.189909     0.330838

               X57          X58          X59          X60          X61
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.007988     0.002819     0.007189     0.000996     0.009108
std       0.030638     0.038794     0.030957     0.052970     0.039820
min      -0.208407    -0.424576    -0.377021    -0.914640    -0.745199
25%      -0.003017    -0.010100    -0.003709    -0.009304    -0.002731
50%       0.010218     0.005113     0.009532     0.005549     0.011224
75%       0.022411     0.016692     0.021667     0.017554     0.025097
```

```
max       0.223415     0.296995     0.109691     0.442809     0.206830

[8 rows x 62 columns]
```

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.073034	0.014607	0.051124	0.029214	0.021910	0.043820	0.036517	
1	0.041026	0.006698	0.046050	0.033491	0.019257	0.051910	0.018420	
2	0.073975	0.001608	0.045028	0.057894	0.032163	0.061110	0.012865	
3	0.066498	0.014408	0.049874	0.038791	0.028816	0.035466	0.021058	
4	0.048233	0.013154	0.041656	0.026309	0.017539	0.031790	0.013154	
..	…	…	…	…	…	…	…	
995	0.065212	0.004207	0.054694	0.056798	0.054694	0.058902	0.031554	
996	0.016001	0.004000	0.021001	0.025001	0.006000	0.016001	0.007000	
997	0.047105	0.020514	0.050144	0.038748	0.031910	0.057742	0.014435	
998	0.040295	0.008059	0.039399	0.044772	0.015222	0.025072	0.009850	
999	0.023592	0.000000	0.037747	0.037747	0.014155	0.018874	0.033029	

	X7	X8	X9	…	X53	X54	X55	\
0	0.051124	0.007303	0.043820	…	-0.046139	-0.040507	0.008700	
1	0.024281	0.040189	0.050236	…	0.006752	-0.002218	0.015289	
2	0.051461	0.054677	0.082016	…	0.031317	-0.020067	0.020600	
3	0.042116	0.021058	0.064282	…	0.012885	-0.024473	0.001891	
4	0.035079	0.016443	0.061387	…	-0.006064	0.007183	0.013594	
..	…	…	…	…	…	…	…	
995	0.067316	0.056798	0.109389	…	0.010488	-0.012842	-0.032628	
996	0.016001	0.037002	0.021001	…	0.032394	0.007338	0.018016	
997	0.028111	0.036469	0.060781	…	0.007006	-0.014809	-0.005960	
998	0.035818	0.021491	0.042086	…	0.014800	-0.006388	0.015223	
999	0.070776	0.028311	0.061340	…	-0.042653	0.043571	0.002282	

	X56	X57	X58	X59	X60	X61	X62
0	-0.102012	-0.079679	-0.042779	-0.017781	-0.119032	-0.078600	no_efectores
1	0.006360	0.008245	-0.001314	0.002813	-0.006637	0.011341	no_efectores
2	0.031441	-0.003532	-0.022360	0.012646	-0.011341	0.008478	no_efectores
3	0.002458	0.015988	0.025096	0.001454	-0.011058	-0.000205	no_efectores
4	-0.002114	0.008374	0.000968	-0.005463	-0.023431	-0.021295	no_efectores
..	…	…	…	…	…	…	…
995	0.011462	0.019260	-0.009662	0.011664	0.011941	0.025168	no_efectores
996	0.011038	0.020148	0.002034	0.017383	0.012078	0.011627	no_efectores
997	-0.027122	0.006573	-0.006913	-0.016491	-0.015246	-0.005597	no_efectores
998	-0.010261	-0.006633	0.018829	0.028548	0.002229	0.021394	no_efectores

```
999  0.011425  0.010860  0.000880  0.013847  0.034884  0.029021  no_efectores

[1000 rows x 63 columns]
```

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

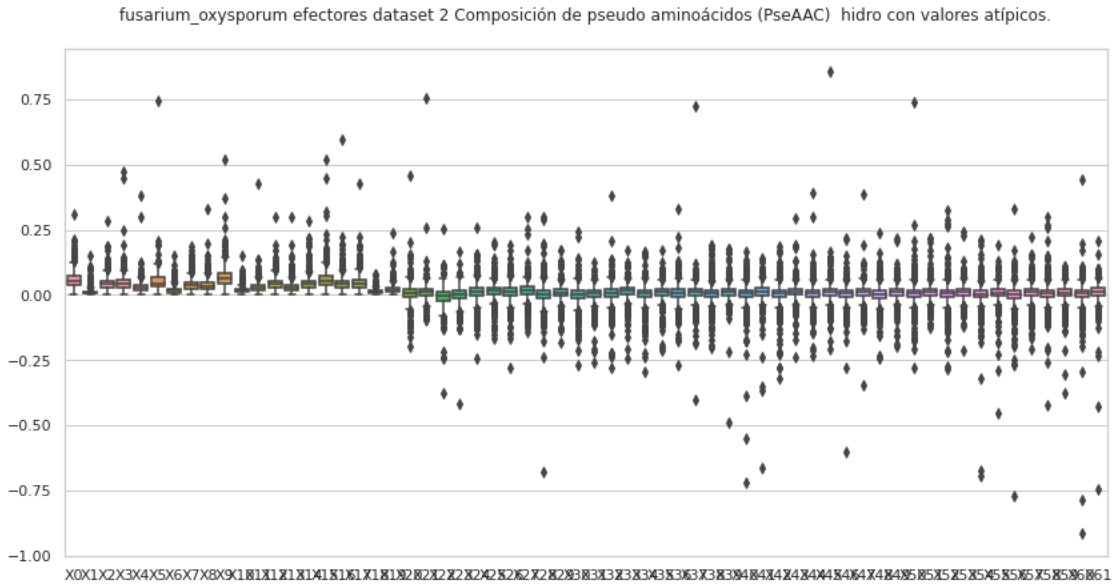
	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.060729	0.012061	0.043017	0.043785	0.029450	
std	0.031369	0.013331	0.020767	0.021223	0.017907	
min	0.005889	0.000000	0.000000	0.000000	0.000000	
25%	0.040943	0.004577	0.027095	0.028970	0.017969	
50%	0.056970	0.009014	0.042872	0.042233	0.027611	
75%	0.075308	0.015999	0.055866	0.056362	0.038611	
max	0.323104	0.177997	0.125607	0.152903	0.177997	

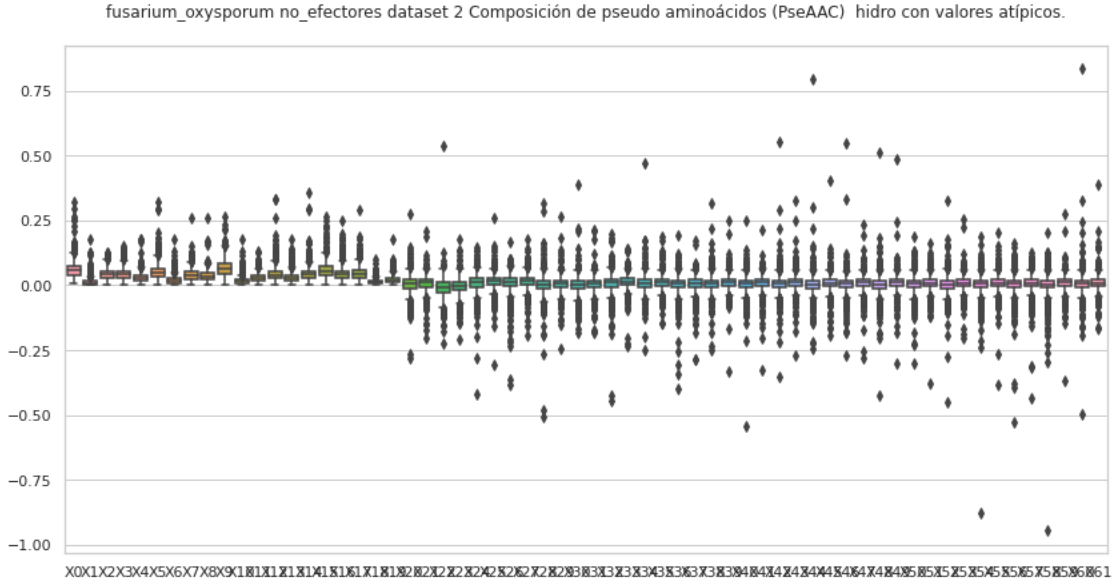
	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.051941	0.019924	0.039631	0.035584	0.068126	…	
std	0.030976	0.014777	0.021761	0.020212	0.033805	…	
min	0.000000	0.000000	0.000000	0.000000	0.002697	…	
25%	0.031673	0.010052	0.025112	0.021005	0.044769	…	
50%	0.046798	0.017673	0.037422	0.034338	0.065678	…	
75%	0.065573	0.025946	0.051414	0.046316	0.087178	…	
max	0.323104	0.177997	0.258483	0.258483	0.266996	…	

	X52	X53	X54	X55	X56	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.002835	0.007886	0.001891	0.008315	0.002666	
std	0.035699	0.028397	0.041816	0.030630	0.038835	
min	-0.450488	-0.203904	-0.879295	-0.384908	-0.527538	
25%	-0.011500	-0.003752	-0.009204	-0.003328	-0.009113	
50%	0.004723	0.008459	0.005163	0.009542	0.005222	
75%	0.017847	0.021307	0.017215	0.022589	0.017825	
max	0.326888	0.255148	0.185347	0.204679	0.161594	

	X57	X58	X59	X60	X61
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.008175	0.000480	0.007945	0.004430	0.008158
std	0.032636	0.046411	0.030529	0.045051	0.030610
min	-0.433746	-0.944107	-0.370373	-0.498777	-0.167910
25%	-0.002329	-0.010117	-0.003560	-0.010006	-0.003492
50%	0.009899	0.004430	0.010429	0.005296	0.007868
75%	0.023529	0.017411	0.022384	0.018218	0.022035
max	0.126119	0.202238	0.273908	0.835515	0.387250

[8 rows x 62 columns]





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

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

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


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

    if etiq == "efectores":
        df=PseAAC_hidro_efec

    if etiq == "no_efectores":
        df=PseAAC_hidro_no_efec

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

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


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

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

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

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.066959	0.007102	0.051741	0.059857	0.036523	0.049712	0.027392	
2	0.093455	0.009044	0.075367	0.084411	0.027132	0.090440	0.024117	
3	0.038896	0.011207	0.012526	0.015822	0.017141	0.038896	0.005274	
4	0.049091	0.009474	0.013780	0.019809	0.030143	0.047368	0.006029	
5	0.041168	0.011357	0.011357	0.022714	0.032651	0.036910	0.005678	
..	…	…	…	…	…	…	…	
995	0.097315	0.019463	0.068121	0.082718	0.058389	0.092450	0.024329	
996	0.038901	0.006544	0.026904	0.026904	0.027267	0.037811	0.012361	
997	0.123614	0.015034	0.046773	0.060972	0.025892	0.072665	0.016705	
998	0.041692	0.012739	0.022004	0.033585	0.006949	0.039376	0.015056	
999	0.073557	0.007356	0.078154	0.036778	0.037698	0.056087	0.013792	

	X7	X8	X9	…	X53	X54	X55	\
0	0.046668	0.037537	0.074060	…	0.026748	0.013631	0.016151	
2	0.051249	0.033161	0.120587	…	-0.003937	0.105541	0.048624	
3	0.032304	0.015822	0.040874	…	0.002855	0.010028	0.008385	
4	0.034450	0.020670	0.038756	…	-0.005763	0.011584	0.007032	
5	0.034070	0.025553	0.039749	…	0.004273	0.010096	0.002154	
..	…	…	…	…	…	…	…	
995	0.063255	0.034060	0.077852	…	-0.019186	0.109598	0.066215	
996	0.029085	0.026904	0.045809	…	0.013151	-0.007011	0.004067	
997	0.054290	0.040926	0.091040	…	0.032793	0.016775	0.004401	
998	0.019688	0.022004	0.034744	…	0.004772	0.016076	0.033575	
999	0.058845	0.049651	0.094704	…	0.005544	-0.007626	-0.000371	

	X56	X57	X58	X59	X60	X61	X62
0	-0.016628	-0.008935	0.003123	0.019942	0.012351	0.011288	efectores
2	0.012261	0.009567	-0.023098	-0.019683	-0.060099	-0.017069	efectores
3	0.020540	0.008831	0.010077	-0.003706	0.014937	0.007614	efectores
4	0.010838	-0.002557	0.000115	-0.010472	0.021227	0.001303	efectores
5	0.016840	0.011257	0.017243	0.007052	0.013534	0.001489	efectores
..	…	…	…	…	…	…	…
995	0.028175	0.022593	-0.046271	-0.027760	0.004416	0.036552	efectores
996	0.006271	0.017256	0.011189	0.012065	0.009305	0.003508	efectores
997	0.005886	-0.002819	-0.015159	-0.009790	0.029600	0.038530	efectores
998	-0.018293	0.018959	-0.013230	0.020089	0.020557	0.025819	efectores
999	-0.023631	-0.001919	0.016986	-0.000841	0.034964	0.018905	efectores

```
[882 rows x 63 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	882.000000	882.000000	882.000000	882.000000	882.000000	882.000000	
mean	0.055126	0.009803	0.039666	0.041332	0.026501	0.046855	
std	0.023606	0.007720	0.018103	0.018433	0.013606	0.022758	
min	0.001305	0.000000	0.000000	0.003110	0.000000	0.000000	
25%	0.037688	0.004435	0.025385	0.028141	0.016240	0.029835	
50%	0.052396	0.007879	0.040187	0.040441	0.025940	0.042751	
75%	0.070699	0.013183	0.051998	0.054095	0.035168	0.062421	
max	0.132933	0.044248	0.101548	0.115081	0.085677	0.114217	

	X6	X7	X8	X9	…	X52	\
count	882.000000	882.000000	882.000000	882.000000	…	882.000000	
mean	0.016781	0.035523	0.034349	0.061451	…	0.004162	
std	0.009874	0.016805	0.016314	0.028932	…	0.022654	
min	0.000000	0.000000	0.002531	0.000000	…	-0.088563	
25%	0.008758	0.023523	0.022110	0.040459	…	-0.008542	
50%	0.015708	0.034299	0.032383	0.059815	…	0.005483	
75%	0.022974	0.046840	0.045204	0.079193	…	0.016837	
max	0.052329	0.095063	0.101038	0.168845	…	0.082364	

	X53	X54	X55	X56	X57	X58	\
count	882.000000	882.000000	882.000000	882.000000	882.000000	882.000000	
mean	0.010821	0.003768	0.009655	0.003111	0.010138	0.003638	
std	0.020870	0.022275	0.019880	0.021355	0.019205	0.023142	
min	-0.070732	-0.089764	-0.064726	-0.078907	-0.070967	-0.093436	
25%	-0.000536	-0.007625	-0.001685	-0.009692	-0.001029	-0.007871	
50%	0.012164	0.004535	0.009185	0.005247	0.010718	0.005291	
75%	0.023741	0.016070	0.022097	0.015834	0.022297	0.015884	
max	0.083153	0.109598	0.098379	0.094625	0.080500	0.105404	

	X59	X60	X61
count	882.000000	882.000000	882.000000
mean	0.009976	0.004942	0.011804
std	0.019427	0.023543	0.021243
min	-0.082865	-0.099816	-0.070071
25%	-0.001070	-0.006261	-0.000568
50%	0.010028	0.006755	0.011771
75%	0.021321	0.017379	0.024697
max	0.096790	0.090179	0.097131

[8 rows x 62 columns]

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

           X0        X1        X2        X3        X4        X5        X6  \
1    0.041026  0.006698  0.046050  0.033491  0.019257  0.051910  0.018420
2    0.073975  0.001608  0.045028  0.057894  0.032163  0.061110  0.012865
3    0.066498  0.014408  0.049874  0.038791  0.028816  0.035466  0.021058
4    0.048233  0.013154  0.041656  0.026309  0.017539  0.031790  0.013154
6    0.025002  0.010870  0.021741  0.015218  0.020654  0.028263  0.014131
..        …         …         …         …         …         …
994  0.047173  0.007940  0.024754  0.016814  0.023353  0.028491  0.010275
995  0.065212  0.004207  0.054694  0.056798  0.054694  0.058902  0.031554
996  0.016001  0.004000  0.021001  0.025001  0.006000  0.016001  0.007000
997  0.047105  0.020514  0.050144  0.038748  0.031910  0.057742  0.014435
998  0.040295  0.008059  0.039399  0.044772  0.015222  0.025072  0.009850

           X7        X8        X9  …       X53       X54       X55  \
1    0.024281  0.040189  0.050236  …  0.006752 -0.002218  0.015289
2    0.051461  0.054677  0.082016  …  0.031317 -0.020067  0.020600
3    0.042116  0.021058  0.064282  …  0.012885 -0.024473  0.001891
4    0.035079  0.016443  0.061387  … -0.006064  0.007183  0.013594
6    0.029350  0.018480  0.052178  …  0.005498  0.012279  0.000441
..        …         …         …  …         …         …
994  0.038766  0.014946  0.073795  …  0.001048  0.038184  0.016124
995  0.067316  0.056798  0.109389  …  0.010488 -0.012842 -0.032628
996  0.016001  0.037002  0.021001  …  0.032394  0.007338  0.018016
997  0.028111  0.036469  0.060781  …  0.007006 -0.014809 -0.005960
998  0.035818  0.021491  0.042086  …  0.014800 -0.006388  0.015223

          X56       X57       X58       X59       X60       X61           X62
1    0.006360  0.008245 -0.001314  0.002813 -0.006637  0.011341  no_efectores
2    0.031441 -0.003532 -0.022360  0.012646 -0.011341  0.008478  no_efectores
3    0.002458  0.015988  0.025096  0.001454 -0.011058 -0.000205  no_efectores
4   -0.002114  0.008374  0.000968 -0.005463 -0.023431 -0.021295  no_efectores
6   -0.011508 -0.009892  0.012505  0.009486  0.022670 -0.006690  no_efectores
..        …         …         …         …         …         …
994 -0.006246 -0.003027  0.026807  0.004092  0.014143  0.003036  no_efectores
995  0.011462  0.019260 -0.009662  0.011664  0.011941  0.025168  no_efectores
996  0.011038  0.020148  0.002034  0.017383  0.012078  0.011627  no_efectores
997 -0.027122  0.006573 -0.006913 -0.016491 -0.015246 -0.005597  no_efectores
998 -0.010261 -0.006633  0.018829  0.028548  0.002229  0.021394  no_efectores

[867 rows x 63 columns]
```

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

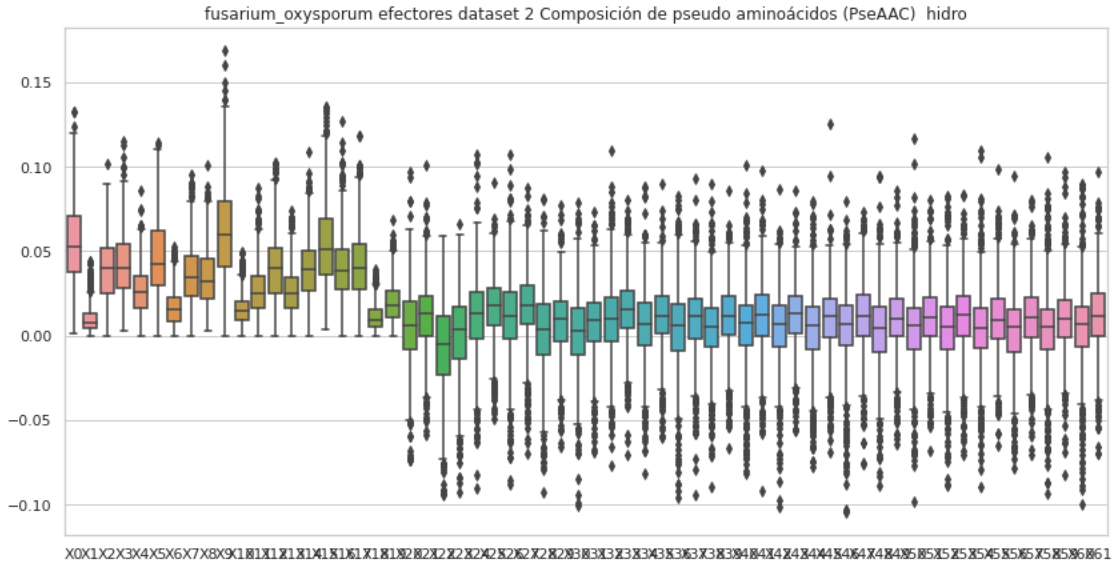
	X0	X1	X2	X3	X4	X5	\
count	867.000000	867.000000	867.000000	867.000000	867.000000	867.000000	
mean	0.056461	0.010104	0.040718	0.041407	0.027411	0.046970	
std	0.024383	0.007533	0.018531	0.018167	0.014075	0.022729	
min	0.005889	0.000000	0.000000	0.000000	0.000000	0.002152	
25%	0.038611	0.004500	0.026488	0.028034	0.017394	0.029724	
50%	0.054361	0.008462	0.041019	0.040784	0.026383	0.044730	
75%	0.071579	0.014304	0.053528	0.054006	0.037026	0.060616	
max	0.154373	0.051403	0.097708	0.103083	0.083022	0.144733	

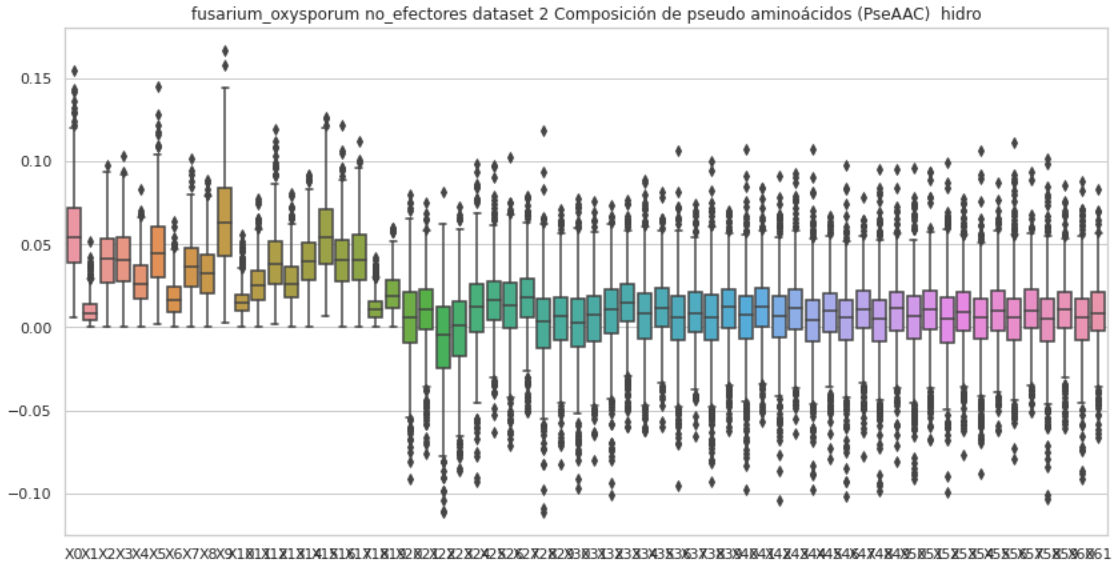
	X6	X7	X8	X9	…	X52	\
count	867.000000	867.000000	867.000000	867.000000	…	867.000000	
mean	0.017879	0.037014	0.033599	0.063669	…	0.004561	
std	0.010398	0.017595	0.016278	0.028730	…	0.023148	
min	0.000000	0.000000	0.000000	0.002697	…	-0.098807	
25%	0.009576	0.024806	0.020253	0.042859	…	-0.009475	
50%	0.016787	0.036203	0.032827	0.062806	…	0.005430	
75%	0.024435	0.048004	0.043792	0.083648	…	0.017871	
max	0.063545	0.101445	0.088416	0.166483	…	0.099223	

	X53	X54	X55	X56	X57	X58	\
count	867.000000	867.000000	867.000000	867.000000	867.000000	867.000000	
mean	0.009211	0.004780	0.009717	0.005564	0.010353	0.003830	
std	0.020309	0.022066	0.020428	0.022486	0.018823	0.023506	
min	-0.076748	-0.088185	-0.062607	-0.079106	-0.051050	-0.103213	
25%	-0.001907	-0.006482	-0.002394	-0.007716	-0.000812	-0.008451	
50%	0.009263	0.006181	0.010135	0.006006	0.010181	0.005112	
75%	0.021042	0.016891	0.022133	0.017572	0.022523	0.016960	
max	0.077200	0.106155	0.086992	0.110881	0.093203	0.101709	

	X59	X60	X61
count	867.000000	867.000000	867.000000
mean	0.009679	0.004614	0.008568
std	0.019209	0.022341	0.019358
min	-0.060830	-0.091232	-0.066642
25%	-0.000665	-0.007813	-0.001983
50%	0.010686	0.005634	0.008346
75%	0.021389	0.017312	0.021507
max	0.085094	0.088030	0.083309

[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 fusarium_oxysporum
dataset 2, con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.029254 -0.027115  0.054639 -0.040025 -0.023540  0.057690  0.014828
1   -0.017197 -0.044290  0.093608  0.101439  0.131142  0.004135 -0.176308
2   -0.003907 -0.048486 -0.023992 -0.055206 -0.009835 -0.004075  0.006016
3   -0.045057  0.038780  0.031591  0.077752  0.010091 -0.001557  0.033936
4   -0.045676  0.057307 -0.012145  0.088963 -0.027765 -0.050090  0.021919
..        …         …         …         …         …         …         …
995  0.001308  0.020169  0.014361 -0.030617  0.108970 -0.063608 -0.066512
996  0.025264  0.026069  0.012862  0.026300  0.009456  0.025527  0.000527
997 -0.003600  0.034699  0.029189 -0.008363 -0.013482  0.037532 -0.032246
998  0.034409 -0.010677  0.009393 -0.027888 -0.001469 -0.137049 -0.044194
999  0.043652  0.025815  0.023371  0.001224  0.033454  0.035920  0.012011

           X7        X8        X9       X10       X11       X12        X13
0    0.016836 -0.004217 -0.029513 -0.012428 -0.046910 -0.000580  efectores
```

```
1    0.020039 -0.019548  0.067384 -0.134807 -0.012458  0.035451  efectores
2   -0.008027 -0.009609 -0.050187  0.012267  0.027931 -0.012786  efectores
3   -0.013203  0.038242 -0.059766 -0.004613 -0.024407  0.007478  efectores
4   -0.062960  0.053577 -0.029200  0.075409  0.039584  0.001800  efectores
..        …       …       …       …       …       …        …
995  0.041824 -0.159779 -0.076974 -0.022098  0.096538  0.053413  efectores
996 -0.061908  0.026446  0.024620  0.054589 -0.020487 -0.041351  efectores
997 -0.004137  0.023620  0.026100 -0.002333 -0.010776  0.013066  efectores
998  0.055840  0.021000  0.029757  0.098726 -0.031333  0.121496  efectores
999  0.002595  0.013030 -0.036075 -0.010247 -0.000937 -0.023715  efectores

[1000 rows x 14 columns]
```

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

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.007898	0.010420	0.009771	0.012434	0.004012	
std	0.061605	0.057858	0.055317	0.061637	0.056894	
min	-0.437624	-0.300528	-0.317518	-0.268226	-0.275322	
25%	-0.024044	-0.022511	-0.019393	-0.019029	-0.025805	
50%	0.007296	0.011517	0.011124	0.014836	0.003930	
75%	0.042218	0.043510	0.039212	0.046760	0.035127	
max	0.363650	0.503993	0.263736	0.308047	0.247435	

	X5	X6	X7	X8	X9	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.004624	0.006079	0.005542	0.004797	0.002341	
std	0.058329	0.057497	0.058375	0.057872	0.064432	
min	-0.254366	-0.356856	-0.373323	-0.290646	-0.333390	
25%	-0.025717	-0.023414	-0.025742	-0.024469	-0.029303	
50%	0.007558	0.006667	0.006720	0.003416	0.003081	
75%	0.039471	0.036022	0.038368	0.034903	0.035504	
max	0.303819	0.346334	0.264890	0.263900	0.601063	

	X10	X11	X12
count	1000.000000	1000.000000	1000.000000
mean	0.005533	0.004067	0.005411
std	0.059977	0.054971	0.058445
min	-0.329572	-0.200135	-0.242843
25%	-0.025719	-0.026114	-0.027755
50%	0.004631	0.004587	0.002939
75%	0.036612	0.034365	0.034876
max	0.335026	0.256460	0.421009

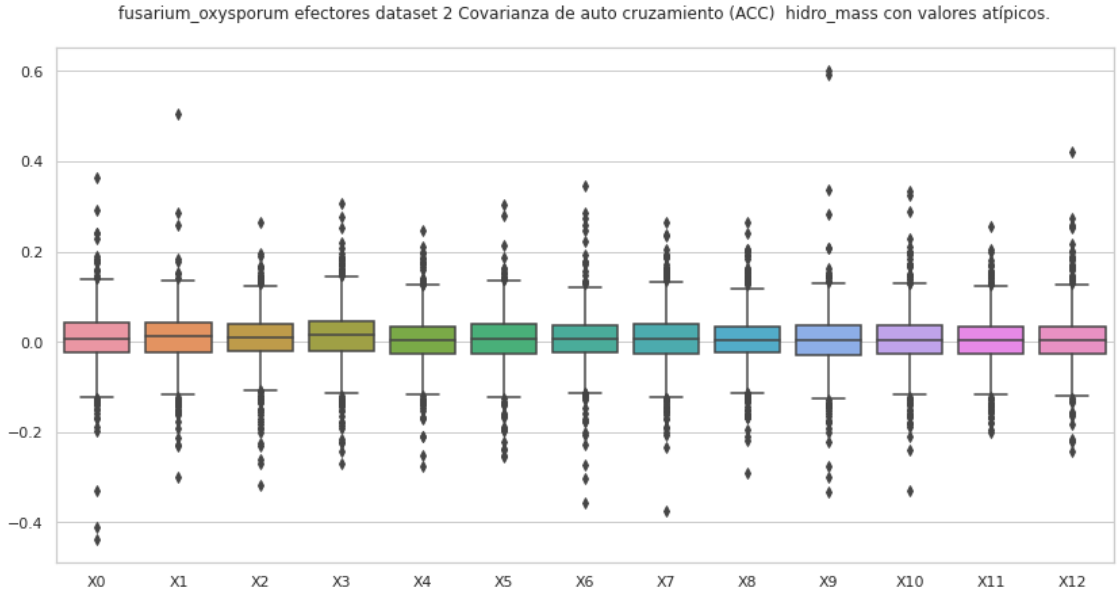
no_efectores
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores fusarium_oxysporum
dataset 2, con valores atípicos.
Valores del documento csv.

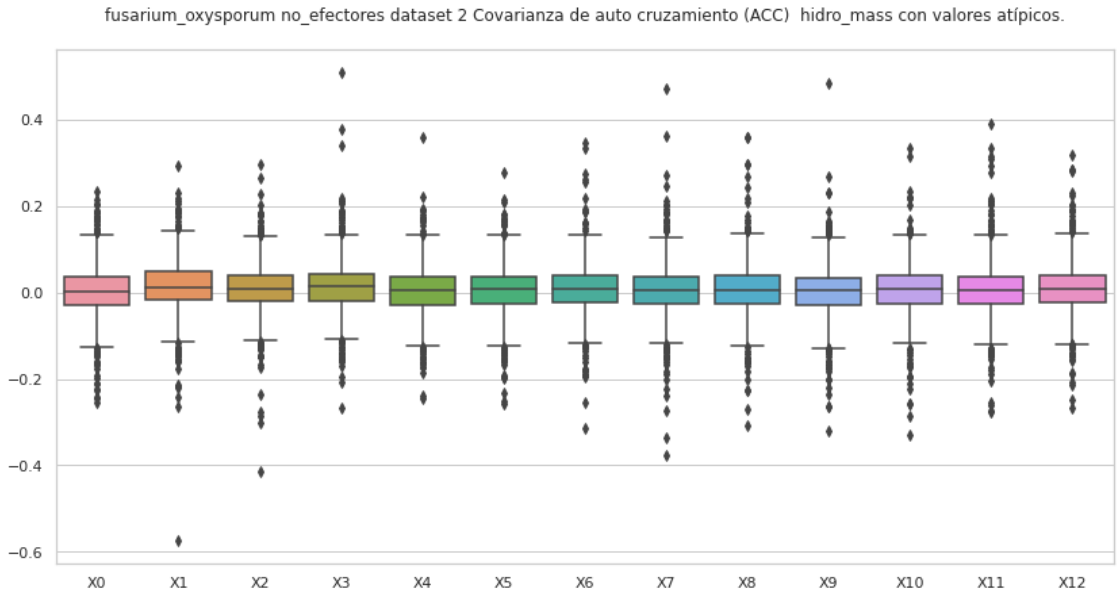
```
           X0        X1        X2        X3        X4        X5        X6  \
0   -0.002669  0.120923  0.026152  0.015756 -0.017041 -0.078143 -0.090430
1   -0.010897 -0.004304  0.037379 -0.009984 -0.037635  0.035549  0.021690
2   -0.001690  0.024345  0.042145  0.024123  0.022003  0.024760 -0.044088
3    0.059653 -0.007338  0.023733  0.024210  0.015698  0.050625 -0.020775
4    0.004753  0.059749 -0.058299 -0.023815  0.017752  0.036672  0.067263
..        …         …         …         …         …         …
995  0.041971  0.020205  0.019927  0.020961  0.035720  0.023011  0.068797
996  0.086793 -0.003113  0.060965  0.020451 -0.054065 -0.038042  0.037406
997  0.040111  0.097349  0.093317  0.083541  0.125803  0.052805  0.082016
998 -0.008591  0.042661  0.011164 -0.027529 -0.042756  0.041233  0.010834
999  0.111388  0.008244 -0.036812 -0.004888 -0.108528 -0.103451 -0.001071

           X7        X8        X9       X10       X11       X12           X13
0   -0.106577 -0.004407  0.028100  0.019005 -0.028356  0.084659  no_efectores
1   -0.031093 -0.040008 -0.025540  0.038243  0.069147 -0.057762  no_efectores
2   -0.073360  0.043471  0.029918  0.014036  0.027699  0.053503  no_efectores
3    0.021126 -0.013553 -0.011375  0.006641  0.049111  0.078575  no_efectores
4    0.029570  0.072363 -0.000631  0.021275 -0.032045 -0.072984  no_efectores
..        …         …         …         …         …         …
995 -0.069726 -0.056327 -0.006280 -0.034501 -0.020839  0.023166  no_efectores
996  0.064998 -0.036694 -0.012359  0.020862  0.074632  0.005162  no_efectores
997  0.081828  0.084251  0.071214  0.021394  0.064115  0.057364  no_efectores
998 -0.008263  0.023523 -0.015279 -0.057878 -0.006442 -0.047461  no_efectores
999 -0.030808 -0.060940 -0.015558  0.012852  0.007294 -0.075209  no_efectores

[1000 rows x 14 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.002724     0.013014     0.008418     0.014119     0.003951
std       0.060283     0.062846     0.058339     0.060392     0.057975
min      -0.253277    -0.574021    -0.412833    -0.265775    -0.246493
25%      -0.029156    -0.017738    -0.021195    -0.019159    -0.027948
50%       0.002822     0.012207     0.008516     0.013352     0.005519
75%       0.036728     0.047567     0.039261     0.042466     0.036431
```

```
max       0.232821     0.294194     0.295874     0.508912     0.358408

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.005140     0.008539     0.006066     0.006655     0.001777
std       0.057223     0.057802     0.063724     0.060098     0.060436
min      -0.258857    -0.313191    -0.377632    -0.308958    -0.320190
25%      -0.027152    -0.023922    -0.025621    -0.026114    -0.030749
50%       0.007696     0.007837     0.005965     0.004990     0.004640
75%       0.037601     0.039838     0.036601     0.039107     0.034756
max       0.275991     0.347595     0.470728     0.360053     0.485141

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.005309     0.007732     0.008145
std       0.060015     0.062182     0.060838
min      -0.328373    -0.276018    -0.268542
25%      -0.026172    -0.025883    -0.023262
50%       0.007935     0.006179     0.006792
75%       0.038190     0.037545     0.041088
max       0.335035     0.389110     0.319251
```





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

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

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

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

    if etiq == "efectores":
        df=ACC_hidro_mass_efec

    if etiq == "no_efectores":
        df=ACC_hidro_mass_no_efec

    del df['X13']
```

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

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

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

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

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.029254 -0.027115  0.054639 -0.040025 -0.023540  0.057690  0.014828
2   -0.003907 -0.048486 -0.023992 -0.055206 -0.009835 -0.004075  0.006016
3   -0.045057  0.038780  0.031591  0.077752  0.010091 -0.001557  0.033936
4   -0.045676  0.057307 -0.012145  0.088963 -0.027765 -0.050090  0.021919
5   -0.032851  0.040783 -0.077271  0.001231  0.015273 -0.023072 -0.144986
..        …         …         …         …         …         …         …
995  0.001308  0.020169  0.014361 -0.030617  0.108970 -0.063608 -0.066512
996  0.025264  0.026069  0.012862  0.026300  0.009456  0.025527  0.000527
997 -0.003600  0.034699  0.029189 -0.008363 -0.013482  0.037532 -0.032246
998  0.034409 -0.010677  0.009393 -0.027888 -0.001469 -0.137049 -0.044194
999  0.043652  0.025815  0.023371  0.001224  0.033454  0.035920  0.012011

           X7        X8        X9       X10       X11       X12        X13
0    0.016836 -0.004217 -0.029513 -0.012428 -0.046910 -0.000580  efectores
2   -0.008027 -0.009609 -0.050187  0.012267  0.027931 -0.012786  efectores
3   -0.013203  0.038242 -0.059766 -0.004613 -0.024407  0.007478  efectores
4   -0.062960  0.053577 -0.029200  0.075409  0.039584  0.001800  efectores
5    0.075941 -0.046110 -0.018824 -0.024182  0.081187 -0.062469  efectores
..        …         …         …         …         …         …         …
995  0.041824 -0.159779 -0.076974 -0.022098  0.096538  0.053413  efectores
```

```
996 -0.061908  0.026446  0.024620  0.054589 -0.020487 -0.041351  efectores
997 -0.004137  0.023620  0.026100 -0.002333 -0.010776  0.013066  efectores
998  0.055840  0.021000  0.029757  0.098726 -0.031333  0.121496  efectores
999  0.002595  0.013030 -0.036075 -0.010247 -0.000937 -0.023715  efectores

[918 rows x 14 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	918.000000	918.000000	918.000000	918.000000	918.000000	918.000000	
mean	0.009196	0.010459	0.011565	0.014444	0.005355	0.005726	
std	0.051065	0.048738	0.045995	0.051473	0.048163	0.050203	
min	-0.170327	-0.162061	-0.149797	-0.165322	-0.153209	-0.167414	
25%	-0.021739	-0.020671	-0.015821	-0.015726	-0.023053	-0.022843	
50%	0.007491	0.011569	0.011124	0.015406	0.005193	0.008110	
75%	0.041199	0.041550	0.037231	0.045986	0.034696	0.037714	
max	0.188969	0.142762	0.168408	0.194584	0.165795	0.163543	

	X6	X7	X8	X9	X10	X11	\
count	918.000000	918.000000	918.000000	918.000000	918.000000	918.000000	
mean	0.006154	0.005698	0.003796	0.002168	0.003783	0.004656	
std	0.045290	0.049183	0.049130	0.049242	0.049343	0.047211	
min	-0.158329	-0.161532	-0.168410	-0.190278	-0.166643	-0.153698	
25%	-0.021225	-0.024128	-0.022958	-0.026054	-0.024879	-0.022873	
50%	0.006651	0.006902	0.002858	0.003105	0.003761	0.005470	
75%	0.034157	0.037232	0.032565	0.034356	0.033405	0.032999	
max	0.177463	0.171115	0.159040	0.152625	0.184213	0.155163	

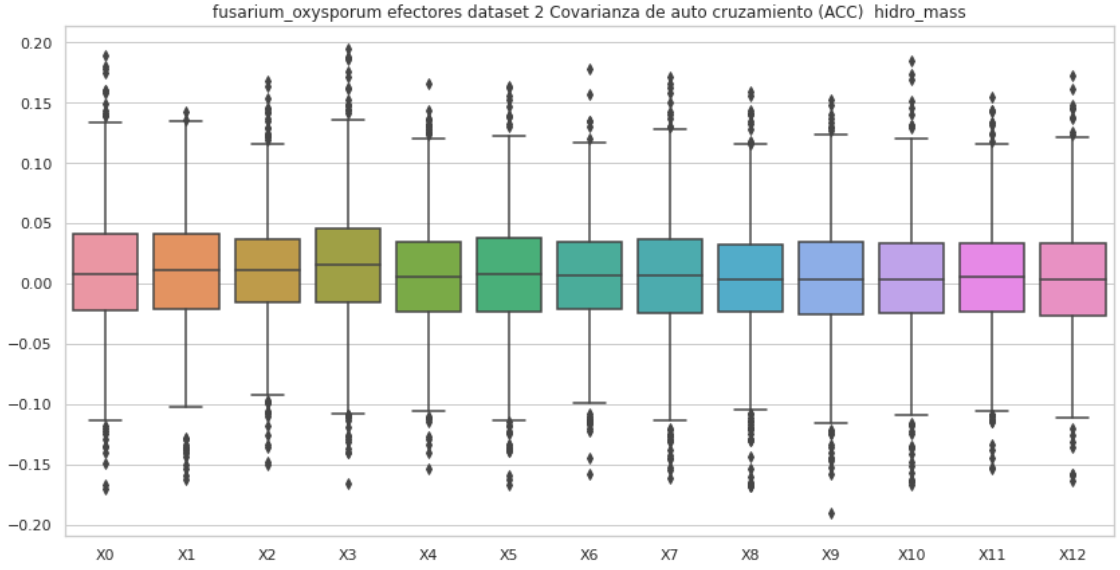
	X12
count	918.000000
mean	0.003417
std	0.047298
min	-0.163182
25%	-0.026636
50%	0.002939
75%	0.033034
max	0.172038

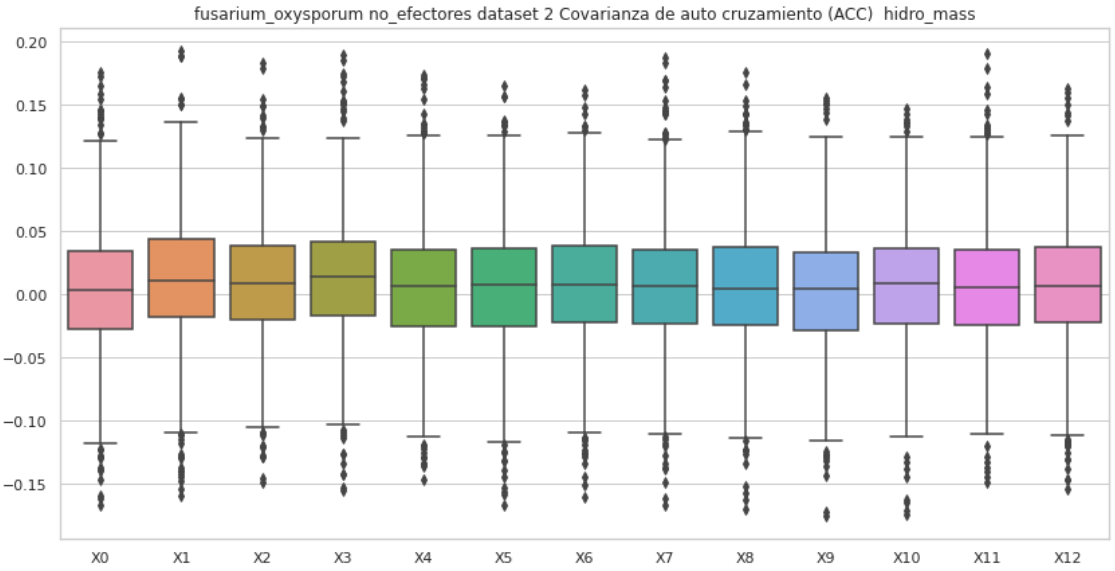
no_efectores
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores fusarium_oxysporum
dataset 2, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0   -0.002669  0.120923  0.026152  0.015756 -0.017041 -0.078143 -0.090430
1   -0.010897 -0.004304  0.037379 -0.009984 -0.037635  0.035549  0.021690
2   -0.001690  0.024345  0.042145  0.024123  0.022003  0.024760 -0.044088
3    0.059653 -0.007338  0.023733  0.024210  0.015698  0.050625 -0.020775
4    0.004753  0.059749 -0.058299 -0.023815  0.017752  0.036672  0.067263
..        …         …         …         …         …         …         …
995  0.041971  0.020205  0.019927  0.020961  0.035720  0.023011  0.068797
996  0.086793 -0.003113  0.060965  0.020451 -0.054065 -0.038042  0.037406
997  0.040111  0.097349  0.093317  0.083541  0.125803  0.052805  0.082016
998 -0.008591  0.042661  0.011164 -0.027529 -0.042756  0.041233  0.010834
999  0.111388  0.008244 -0.036812 -0.004888 -0.108528 -0.103451 -0.001071

           X7        X8        X9       X10       X11       X12           X13
0   -0.106577 -0.004407  0.028100  0.019005 -0.028356  0.084659  no_efectores
1   -0.031093 -0.040008 -0.025540  0.038243  0.069147 -0.057762  no_efectores
2   -0.073360  0.043471  0.029918  0.014036  0.027699  0.053503  no_efectores
3    0.021126 -0.013553 -0.011375  0.006641  0.049111  0.078575  no_efectores
4    0.029570  0.072363 -0.000631  0.021275 -0.032045 -0.072984  no_efectores
..        …         …         …         …         …         …         …
995 -0.069726 -0.056327 -0.006280 -0.034501 -0.020839  0.023166  no_efectores
996  0.064998 -0.036694 -0.012359  0.020862  0.074632  0.005162  no_efectores
997  0.081828  0.084251  0.071214  0.021394  0.064115  0.057364  no_efectores
998 -0.008263  0.023523 -0.015279 -0.057878 -0.006442 -0.047461  no_efectores
999 -0.030808 -0.060940 -0.015558  0.012852  0.007294 -0.075209  no_efectores

[920 rows x 14 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  920.000000  920.000000  920.000000  920.000000  920.000000  920.000000
mean     0.003803    0.012385    0.009758    0.014199    0.005380    0.005151
std      0.050565    0.051703    0.048123    0.048751    0.049802    0.049440
min     -0.166689   -0.159097   -0.148321   -0.155372   -0.146916   -0.166310
25%     -0.027230   -0.017260   -0.019414   -0.016208   -0.024948   -0.025172
50%      0.003184    0.010837    0.009069    0.014200    0.006258    0.007696
75%      0.034218    0.044269    0.038814    0.041778    0.035872    0.036283
max      0.176477    0.192930    0.182905    0.189500    0.173735    0.164886

                X6          X7          X8          X9         X10         X11  \
count  920.000000  920.000000  920.000000  920.000000  920.000000  920.000000
mean     0.008728    0.006994    0.005982    0.003030    0.005927    0.006248
std      0.047587    0.050592    0.049865    0.048510    0.049215    0.048295
```

min	-0.160831	-0.166836	-0.170293	-0.175300	-0.174057	-0.148228
25%	-0.021678	-0.023290	-0.023875	-0.028300	-0.023458	-0.024530
50%	0.008161	0.006605	0.004652	0.004899	0.008553	0.006055
75%	0.038632	0.035197	0.037528	0.032960	0.036648	0.035435
max	0.162178	0.187922	0.176218	0.155930	0.147028	0.190402

	X12
count	920.000000
mean	0.006774
std	0.050284
min	-0.153649
25%	-0.022113
50%	0.006717
75%	0.037650
max	0.163302





7 Covarianza de auto cruzamiento (ACC) mass

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

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

    if etiq == "efectores":
        df=ACC_mass_efec

    if etiq == "no_efectores":
        df=ACC_mass_no_efec

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

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

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.029254 -0.027115  0.054639 -0.040025 -0.023540  0.057690  0.014828
1   -0.017197 -0.044290  0.093608  0.101439  0.131142  0.004135 -0.176308
2   -0.003907 -0.048486 -0.023992 -0.055206 -0.009835 -0.004075  0.006016
3   -0.045057  0.038780  0.031591  0.077752  0.010091 -0.001557  0.033936
4   -0.045676  0.057307 -0.012145  0.088963 -0.027765 -0.050090  0.021919
..        …         …         …         …         …         …         …
995  0.001308  0.020169  0.014361 -0.030617  0.108970 -0.063608 -0.066512
996  0.025264  0.026069  0.012862  0.026300  0.009456  0.025527  0.000527
997 -0.003600  0.034699  0.029189 -0.008363 -0.013482  0.037532 -0.032246
998  0.034409 -0.010677  0.009393 -0.027888 -0.001469 -0.137049 -0.044194
999  0.043652  0.025815  0.023371  0.001224  0.033454  0.035920  0.012011

           X7        X8        X9       X10       X11       X12        X13
0    0.016836 -0.004217 -0.029513 -0.012428 -0.046910 -0.000580  efectores
1    0.020039 -0.019548  0.067384 -0.134807 -0.012458  0.035451  efectores
2   -0.008027 -0.009609 -0.050187  0.012267  0.027931 -0.012786  efectores
3   -0.013203  0.038242 -0.059766 -0.004613 -0.024407  0.007478  efectores
4   -0.062960  0.053577 -0.029200  0.075409  0.039584  0.001800  efectores
..        …         …         …         …         …         …         …
995  0.041824 -0.159779 -0.076974 -0.022098  0.096538  0.053413  efectores
996 -0.061908  0.026446  0.024620  0.054589 -0.020487 -0.041351  efectores
997 -0.004137  0.023620  0.026100 -0.002333 -0.010776  0.013066  efectores
998  0.055840  0.021000  0.029757  0.098726 -0.031333  0.121496  efectores
999  0.002595  0.013030 -0.036075 -0.010247 -0.000937 -0.023715  efectores

[1000 rows x 14 columns]


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

                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.007898     0.010420     0.009771     0.012434     0.004012
```

```
std       0.061605     0.057858     0.055317     0.061637     0.056894
min      -0.437624    -0.300528    -0.317518    -0.268226    -0.275322
25%      -0.024044    -0.022511    -0.019393    -0.019029    -0.025805
50%       0.007296     0.011517     0.011124     0.014836     0.003930
75%       0.042218     0.043510     0.039212     0.046760     0.035127
max       0.363650     0.503993     0.263736     0.308047     0.247435

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.004624     0.006079     0.005542     0.004797     0.002341
std       0.058329     0.057497     0.058375     0.057872     0.064432
min      -0.254366    -0.356856    -0.373323    -0.290646    -0.333390
25%      -0.025717    -0.023414    -0.025742    -0.024469    -0.029303
50%       0.007558     0.006667     0.006720     0.003416     0.003081
75%       0.039471     0.036022     0.038368     0.034903     0.035504
max       0.303819     0.346334     0.264890     0.263900     0.601063

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.005533     0.004067     0.005411
std       0.059977     0.054971     0.058445
min      -0.329572    -0.200135    -0.242843
25%      -0.025719    -0.026114    -0.027755
50%       0.004631     0.004587     0.002939
75%       0.036612     0.034365     0.034876
max       0.335026     0.256460     0.421009
```

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

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.002669  0.120923  0.026152  0.015756 -0.017041 -0.078143 -0.090430
1   -0.010897 -0.004304  0.037379 -0.009984 -0.037635  0.035549  0.021690
2   -0.001690  0.024345  0.042145  0.024123  0.022003  0.024760 -0.044088
3    0.059653 -0.007338  0.023733  0.024210  0.015698  0.050625 -0.020775
4    0.004753  0.059749 -0.058299 -0.023815  0.017752  0.036672  0.067263
..        …         …         …         …         …         …         …
995  0.041971  0.020205  0.019927  0.020961  0.035720  0.023011  0.068797
996  0.086793 -0.003113  0.060965  0.020451 -0.054065 -0.038042  0.037406
997  0.040111  0.097349  0.093317  0.083541  0.125803  0.052805  0.082016
998 -0.008591  0.042661  0.011164 -0.027529 -0.042756  0.041233  0.010834
999  0.111388  0.008244 -0.036812 -0.004888 -0.108528 -0.103451 -0.001071

           X7        X8        X9       X10       X11       X12           X13
```

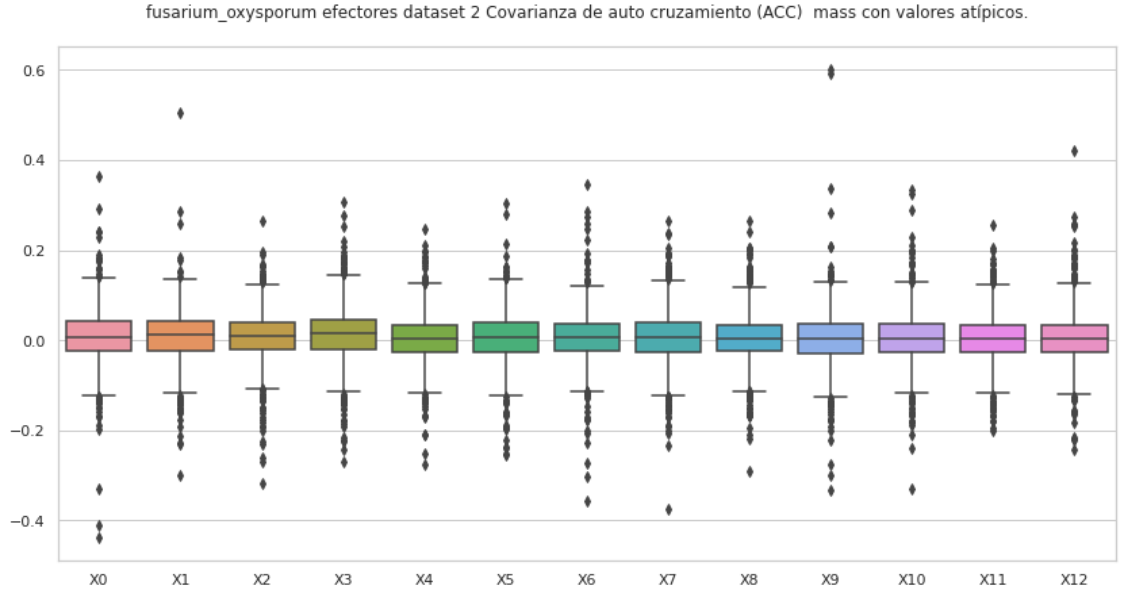
```
0   -0.106577 -0.004407  0.028100  0.019005 -0.028356  0.084659  no_efectores
1   -0.031093 -0.040008 -0.025540  0.038243  0.069147 -0.057762  no_efectores
2   -0.073360  0.043471  0.029918  0.014036  0.027699  0.053503  no_efectores
3    0.021126 -0.013553 -0.011375  0.006641  0.049111  0.078575  no_efectores
4    0.029570  0.072363 -0.000631  0.021275 -0.032045 -0.072984  no_efectores
..        …       …       …       …       …       …           …
995 -0.069726 -0.056327 -0.006280 -0.034501 -0.020839  0.023166  no_efectores
996  0.064998 -0.036694 -0.012359  0.020862  0.074632  0.005162  no_efectores
997  0.081828  0.084251  0.071214  0.021394  0.064115  0.057364  no_efectores
998 -0.008263  0.023523 -0.015279 -0.057878 -0.006442 -0.047461  no_efectores
999 -0.030808 -0.060940 -0.015558  0.012852  0.007294 -0.075209  no_efectores

[1000 rows x 14 columns]
```

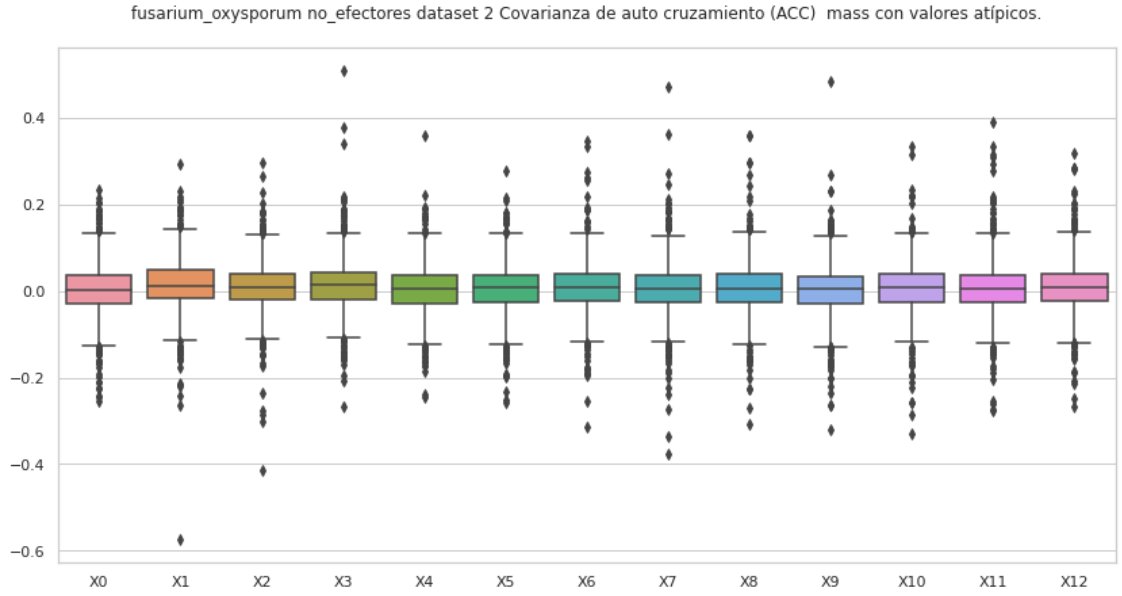
Covarianza de auto cruzamiento (ACC) mass no_efectores fusarium_oxysporum
dataset 2, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.002724	0.013014	0.008418	0.014119	0.003951	
std	0.060283	0.062846	0.058339	0.060392	0.057975	
min	-0.253277	-0.574021	-0.412833	-0.265775	-0.246493	
25%	-0.029156	-0.017738	-0.021195	-0.019159	-0.027948	
50%	0.002822	0.012207	0.008516	0.013352	0.005519	
75%	0.036728	0.047567	0.039261	0.042466	0.036431	
max	0.232821	0.294194	0.295874	0.508912	0.358408	

	X5	X6	X7	X8	X9	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.005140	0.008539	0.006066	0.006655	0.001777	
std	0.057223	0.057802	0.063724	0.060098	0.060436	
min	-0.258857	-0.313191	-0.377632	-0.308958	-0.320190	
25%	-0.027152	-0.023922	-0.025621	-0.026114	-0.030749	
50%	0.007696	0.007837	0.005965	0.004990	0.004640	
75%	0.037601	0.039838	0.036601	0.039107	0.034756	
max	0.275991	0.347595	0.470728	0.360053	0.485141	

	X10	X11	X12
count	1000.000000	1000.000000	1000.000000
mean	0.005309	0.007732	0.008145
std	0.060015	0.062182	0.060838
min	-0.328373	-0.276018	-0.268542
25%	-0.026172	-0.025883	-0.023262
50%	0.007935	0.006179	0.006792
75%	0.038190	0.037545	0.041088
max	0.335035	0.389110	0.319251


fusarium_oxysporum efectores dataset 2 Covarianza de auto cruzamiento (ACC) mass con valores atípicos.
0.6
0.4
0.2
0.0
−0.2
−0.4
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12


fusarium_oxysporum no_efectores dataset 2 Covarianza de auto cruzamiento (ACC) mass con valores atípicos.
0.4
0.2
0.0
−0.2
−0.4
−0.6
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12

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

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

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

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

    if etiq == "efectores":
        df=ACC_mass_efec

    if etiq == "no_efectores":
        df=ACC_mass_no_efec

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

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

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

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

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

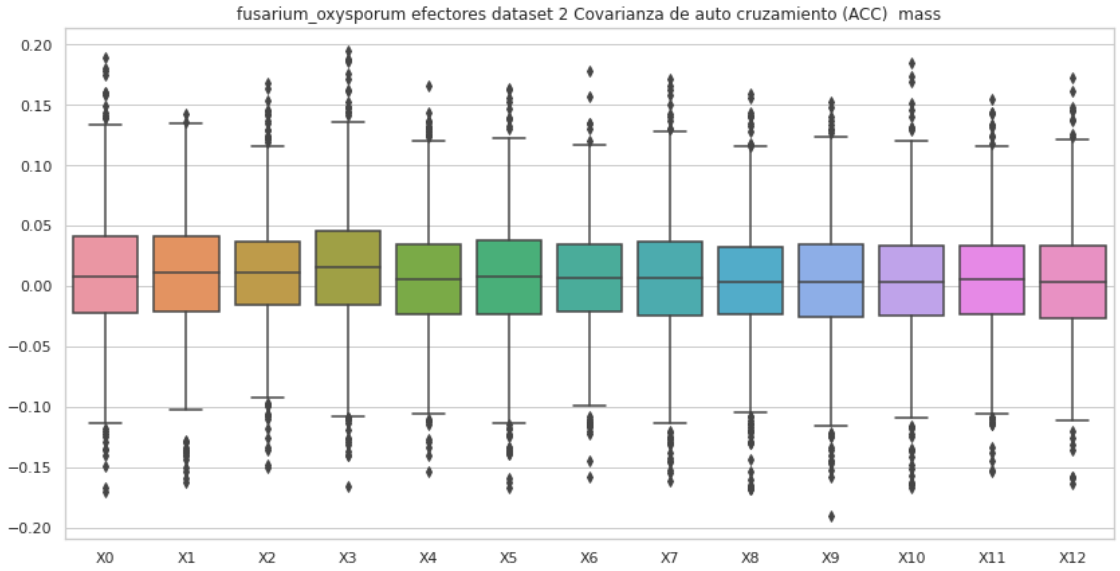
```
Covarianza de auto cruzamiento (ACC)  mass efectores fusarium_oxysporum dataset
2, sin valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.029254 -0.027115  0.054639 -0.040025 -0.023540  0.057690  0.014828
2   -0.003907 -0.048486 -0.023992 -0.055206 -0.009835 -0.004075  0.006016
3   -0.045057  0.038780  0.031591  0.077752  0.010091 -0.001557  0.033936
4   -0.045676  0.057307 -0.012145  0.088963 -0.027765 -0.050090  0.021919
5   -0.032851  0.040783 -0.077271  0.001231  0.015273 -0.023072 -0.144986
..        …         …         …         …         …         …         …
995  0.001308  0.020169  0.014361 -0.030617  0.108970 -0.063608 -0.066512
996  0.025264  0.026069  0.012862  0.026300  0.009456  0.025527  0.000527
997 -0.003600  0.034699  0.029189 -0.008363 -0.013482  0.037532 -0.032246
998  0.034409 -0.010677  0.009393 -0.027888 -0.001469 -0.137049 -0.044194
999  0.043652  0.025815  0.023371  0.001224  0.033454  0.035920  0.012011

           X7        X8        X9       X10       X11       X12         X13
0    0.016836 -0.004217 -0.029513 -0.012428 -0.046910 -0.000580  efectores
2   -0.008027 -0.009609 -0.050187  0.012267  0.027931 -0.012786  efectores
3   -0.013203  0.038242 -0.059766 -0.004613 -0.024407  0.007478  efectores
4   -0.062960  0.053577 -0.029200  0.075409  0.039584  0.001800  efectores
5    0.075941 -0.046110 -0.018824 -0.024182  0.081187 -0.062469  efectores
..        …         …         …         …         …         …         …
995  0.041824 -0.159779 -0.076974 -0.022098  0.096538  0.053413  efectores
996 -0.061908  0.026446  0.024620  0.054589 -0.020487 -0.041351  efectores
997 -0.004137  0.023620  0.026100 -0.002333 -0.010776  0.013066  efectores
998  0.055840  0.021000  0.029757  0.098726 -0.031333  0.121496  efectores
999  0.002595  0.013030 -0.036075 -0.010247 -0.000937 -0.023715  efectores

[918 rows x 14 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  918.000000  918.000000  918.000000  918.000000  918.000000  918.000000
mean     0.009196    0.010459    0.011565    0.014444    0.005355    0.005726
std      0.051065    0.048738    0.045995    0.051473    0.048163    0.050203
min     -0.170327   -0.162061   -0.149797   -0.165322   -0.153209   -0.167414
25%     -0.021739   -0.020671   -0.015821   -0.015726   -0.023053   -0.022843
50%      0.007491    0.011569    0.011124    0.015406    0.005193    0.008110
75%      0.041199    0.041550    0.037231    0.045986    0.034696    0.037714
```

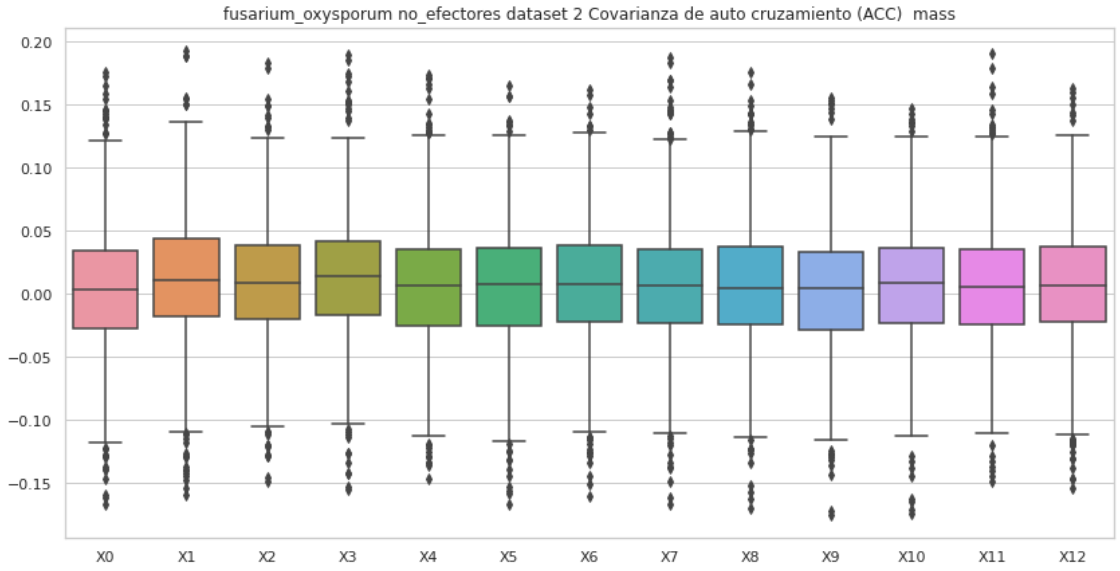
```
max      0.188969    0.142762    0.168408    0.194584    0.165795    0.163543

               X6          X7          X8          X9         X10         X11  \
count  918.000000  918.000000  918.000000  918.000000  918.000000  918.000000
mean     0.006154    0.005698    0.003796    0.002168    0.003783    0.004656
std      0.045290    0.049183    0.049130    0.049242    0.049343    0.047211
min     -0.158329   -0.161532   -0.168410   -0.190278   -0.166643   -0.153698
25%     -0.021225   -0.024128   -0.022958   -0.026054   -0.024879   -0.022873
50%      0.006651    0.006902    0.002858    0.003105    0.003761    0.005470
75%      0.034157    0.037232    0.032565    0.034356    0.033405    0.032999
max      0.177463    0.171115    0.159040    0.152625    0.184213    0.155163

              X12
count  918.000000
mean     0.003417
std      0.047298
min     -0.163182
25%     -0.026636
50%      0.002939
75%      0.033034
max      0.172038
```

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

```
           X0        X1        X2        X3        X4        X5        X6  \
0   -0.002669  0.120923  0.026152  0.015756 -0.017041 -0.078143 -0.090430
1   -0.010897 -0.004304  0.037379 -0.009984 -0.037635  0.035549  0.021690
2   -0.001690  0.024345  0.042145  0.024123  0.022003  0.024760 -0.044088
3    0.059653 -0.007338  0.023733  0.024210  0.015698  0.050625 -0.020775
4    0.004753  0.059749 -0.058299 -0.023815  0.017752  0.036672  0.067263
..        …         …         …         …         …         …         …
995  0.041971  0.020205  0.019927  0.020961  0.035720  0.023011  0.068797
996  0.086793 -0.003113  0.060965  0.020451 -0.054065 -0.038042  0.037406
997  0.040111  0.097349  0.093317  0.083541  0.125803  0.052805  0.082016
998 -0.008591  0.042661  0.011164 -0.027529 -0.042756  0.041233  0.010834
999  0.111388  0.008244 -0.036812 -0.004888 -0.108528 -0.103451 -0.001071

           X7        X8        X9       X10       X11       X12           X13
0   -0.106577 -0.004407  0.028100  0.019005 -0.028356  0.084659  no_efectores
1   -0.031093 -0.040008 -0.025540  0.038243  0.069147 -0.057762  no_efectores
2   -0.073360  0.043471  0.029918  0.014036  0.027699  0.053503  no_efectores
3    0.021126 -0.013553 -0.011375  0.006641  0.049111  0.078575  no_efectores
4    0.029570  0.072363 -0.000631  0.021275 -0.032045 -0.072984  no_efectores
..        …         …         …         …         …         …           …
```

```
995 -0.069726 -0.056327 -0.006280 -0.034501 -0.020839  0.023166  no_efectores
996  0.064998 -0.036694 -0.012359  0.020862  0.074632  0.005162  no_efectores
997  0.081828  0.084251  0.071214  0.021394  0.064115  0.057364  no_efectores
998 -0.008263  0.023523 -0.015279 -0.057878 -0.006442 -0.047461  no_efectores
999 -0.030808 -0.060940 -0.015558  0.012852  0.007294 -0.075209  no_efectores

[920 rows x 14 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  920.000000  920.000000  920.000000  920.000000  920.000000  920.000000
mean     0.003803    0.012385    0.009758    0.014199    0.005380    0.005151
std      0.050565    0.051703    0.048123    0.048751    0.049802    0.049440
min     -0.166689   -0.159097   -0.148321   -0.155372   -0.146916   -0.166310
25%     -0.027230   -0.017260   -0.019414   -0.016208   -0.024948   -0.025172
50%      0.003184    0.010837    0.009069    0.014200    0.006258    0.007696
75%      0.034218    0.044269    0.038814    0.041778    0.035872    0.036283
max      0.176477    0.192930    0.182905    0.189500    0.173735    0.164886

                X6          X7          X8          X9         X10         X11  \
count  920.000000  920.000000  920.000000  920.000000  920.000000  920.000000
mean     0.008728    0.006994    0.005982    0.003030    0.005927    0.006248
std      0.047587    0.050592    0.049865    0.048510    0.049215    0.048295
min     -0.160831   -0.166836   -0.170293   -0.175300   -0.174057   -0.148228
25%     -0.021678   -0.023290   -0.023875   -0.028300   -0.023458   -0.024530
50%      0.008161    0.006605    0.004652    0.004899    0.008553    0.006055
75%      0.038632    0.035197    0.037528    0.032960    0.036648    0.035435
max      0.162178    0.187922    0.176218    0.155930    0.147028    0.190402

               X12
count  920.000000
mean     0.006774
std      0.050284
min     -0.153649
25%     -0.022113
50%      0.006717
75%      0.037650
max      0.163302
```





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 fusarium_oxysporum dataset
2, con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.094137 -0.065201 -0.019587 -0.040861 -0.011276  0.033596  0.021987
1    0.125398  0.097842  0.168729 -0.049869  0.029677  0.078909  0.150173
2   -0.032418 -0.128094  0.042796  0.012409 -0.024097 -0.061371  0.031789
3    0.044149 -0.014993  0.067450  0.070612  0.082249 -0.010255  0.035898
4    0.051764  0.114512  0.033180  0.059411  0.041010  0.084905  0.025737
..        …         …         …         …         …         …         …
995  0.033738 -0.100029  0.078302  0.032144 -0.115762 -0.072146 -0.020075
996  0.079593  0.027876  0.083336  0.049230 -0.011094  0.074444  0.087196
997 -0.021456 -0.033678  0.000266 -0.007823 -0.062855 -0.027728  0.015990
998  0.025784 -0.048446  0.067398  0.035877 -0.009344 -0.036641 -0.019036
999 -0.076676 -0.037880 -0.034982 -0.007734  0.002982 -0.045348  0.001137

           X7        X8        X9       X10       X11       X12        X13
0    0.044953 -0.042449  0.011405 -0.001525  0.060797  0.027889  efectores
1   -0.056984 -0.143666 -0.041678 -0.077792 -0.246211  0.099972  efectores
2    0.025403 -0.030426 -0.059121 -0.045512 -0.020293 -0.103129  efectores
```

```
3    0.104541  0.029003 -0.009017 -0.000111  0.060676  0.057337  efectores
4    0.053961  0.008187 -0.025103 -0.039466  0.024717  0.022372  efectores
..        …       …        …        …        …         …        …
995  0.040706  0.011009  0.017060 -0.016672  0.003936 -0.069432  efectores
996  0.025398  0.061455  0.080925 -0.009013  0.039406  0.044597  efectores
997  0.000198 -0.031937 -0.007625 -0.012907 -0.046566  0.005687  efectores
998 -0.017320 -0.016800 -0.005480 -0.017440  0.071638  0.032435  efectores
999 -0.037100  0.010640 -0.005271  0.000514  0.019842 -0.020818  efectores

[1000 rows x 14 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.008834    -0.016862     0.025170     0.028317    -0.000782
std       0.063251     0.070750     0.064982     0.066502     0.065305
min      -0.367020    -0.277917    -0.312373    -0.242449    -0.229515
25%      -0.029497    -0.059763    -0.013318    -0.011631    -0.036925
50%       0.008240    -0.018092     0.025004     0.026908    -0.004782
75%       0.046208     0.022612     0.062849     0.068512     0.037155
max       0.322639     0.421303     0.435771     0.452772     0.371286

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean     -0.003241     0.019975     0.008677     0.002491     0.008955
std       0.065555     0.065795     0.064907     0.067672     0.062515
min      -0.354816    -0.425843    -0.335312    -0.311257    -0.268394
25%      -0.040921    -0.018218    -0.025880    -0.030956    -0.029674
50%      -0.003985     0.020569     0.007868     0.007162     0.010982
75%       0.035012     0.058283     0.044411     0.040437     0.041408
max       0.403999     0.356524     0.371954     0.382188     0.403345

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.010684     0.009871     0.004772
std       0.068501     0.066216     0.067665
min      -0.570916    -0.394336    -0.321398
25%      -0.023923    -0.023729    -0.034008
50%       0.011860     0.010489     0.005663
75%       0.048368     0.045492     0.040410
max       0.384463     0.364217     0.408269
```

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.127843	0.247060	0.105906	0.188204	0.044875	0.078057	-0.115462	
1	0.038565	0.053407	0.140912	0.060944	0.045415	-0.035036	0.015974	
2	-0.008475	-0.043868	-0.080271	0.069621	-0.043782	0.016885	-0.036962	
3	0.118964	-0.062463	-0.024118	0.032381	0.050950	-0.013069	-0.011878	
4	0.033462	-0.003860	0.007301	0.088887	0.011326	-0.004329	0.006842	
..	…	…	…	…	…	…	…	
995	-0.005968	-0.087170	0.065661	0.016165	-0.039013	0.053640	-0.018549	
996	0.077637	-0.008517	0.043984	-0.031139	-0.017858	-0.005752	0.061961	
997	0.038523	-0.058547	0.001796	0.009286	-0.024215	-0.006867	-0.039591	
998	0.036451	-0.071031	-0.003442	0.018739	0.010509	-0.032165	0.064101	
999	0.156846	0.039771	-0.027168	0.078630	-0.086827	0.138716	0.160085	

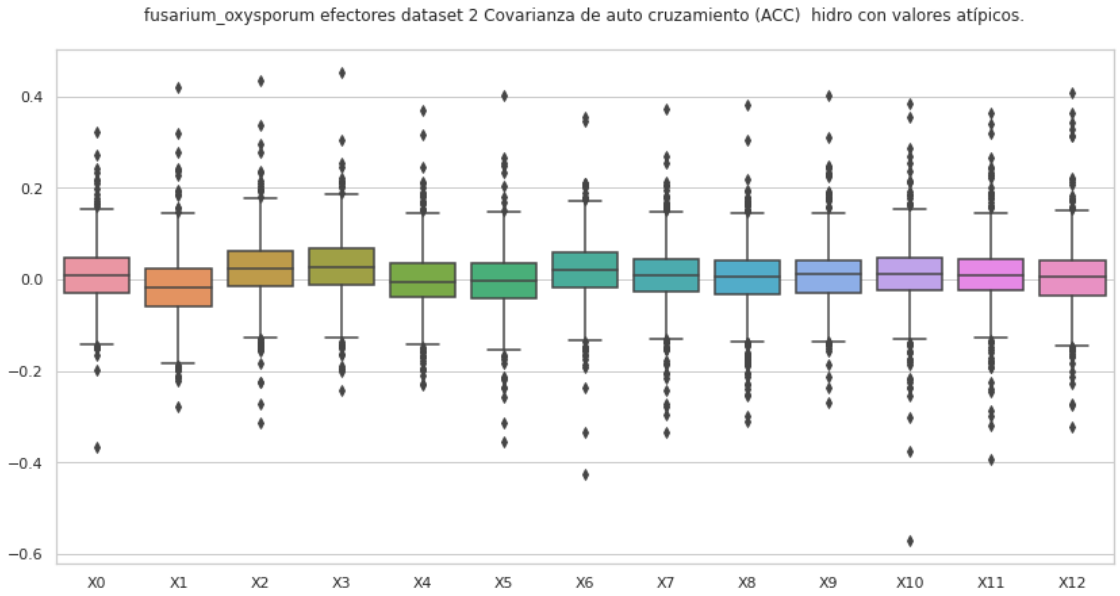
	X7	X8	X9	X10	X11	X12	X13
0	0.100173	0.032001	0.122347	0.007029	0.083424	0.083605	no_efectores
1	0.033185	-0.062923	0.006817	-0.000866	0.050034	0.003383	no_efectores
2	0.007995	0.039370	0.058137	0.011844	-0.021754	-0.005142	no_efectores
3	-0.030740	0.004477	0.090616	0.053325	0.037941	-0.079041	no_efectores
4	0.024606	0.019965	0.050126	0.015663	0.049120	-0.037189	no_efectores
..	…	…	…	…	…	…	…
995	-0.001135	0.061599	-0.066113	-0.021461	-0.004163	0.001742	no_efectores
996	0.089118	-0.088676	-0.064307	-0.081366	-0.171581	0.004402	no_efectores
997	0.064838	0.044713	0.012469	0.008170	0.007526	0.045121	no_efectores
998	0.007310	-0.041264	0.003133	0.020924	0.030442	-0.008109	no_efectores
999	0.086589	-0.133838	-0.003748	-0.095754	-0.020778	-0.049564	no_efectores

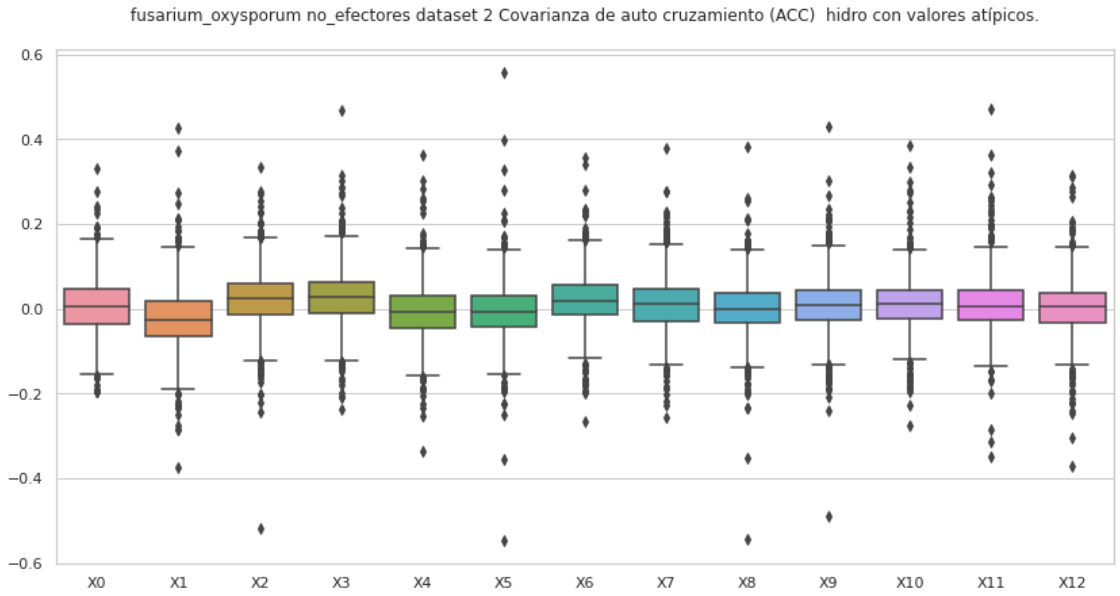
[1000 rows x 14 columns]

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

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.006446	-0.022379	0.023373	0.026660	-0.004950	
std	0.066167	0.075470	0.065865	0.067822	0.067483	
min	-0.196466	-0.375960	-0.519908	-0.238664	-0.337415	
25%	-0.035777	-0.066553	-0.013761	-0.011949	-0.044640	
50%	0.004201	-0.025541	0.022725	0.026320	-0.008447	
75%	0.045156	0.019250	0.058672	0.062932	0.031125	
max	0.330603	0.425550	0.333845	0.468662	0.361432	

```
                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean     -0.004348     0.020614     0.011247     0.000789     0.009578
std       0.070975     0.063383     0.063105     0.064718     0.066820
min      -0.547246    -0.265452    -0.255951    -0.544081    -0.490735
25%      -0.043922    -0.015455    -0.028738    -0.033741    -0.027083
50%      -0.006579     0.018214     0.010020    -0.001583     0.009555
75%       0.031187     0.055174     0.045436     0.035998     0.043376
max       0.557074     0.355594     0.377985     0.382216     0.428982

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.011132     0.010262     0.001755
std       0.063363     0.065728     0.065452
min      -0.276040    -0.349088    -0.372649
25%      -0.022369    -0.027925    -0.033368
50%       0.010560     0.006123     0.004279
75%       0.042591     0.043805     0.037567
max       0.385716     0.471762     0.314567
```





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

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

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

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

    if etiq == "efectores":
        df=ACC_hidro_efec

    if etiq == "no_efectores":
        df=ACC_hidro_no_efec

    del df['X13']
```

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

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

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

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

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

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.094137 -0.065201 -0.019587 -0.040861 -0.011276  0.033596  0.021987
2   -0.032418 -0.128094  0.042796  0.012409 -0.024097 -0.061371  0.031789
3    0.044149 -0.014993  0.067450  0.070612  0.082249 -0.010255  0.035898
4    0.051764  0.114512  0.033180  0.059411  0.041010  0.084905  0.025737
5    0.103107  0.043424  0.030739  0.068554  0.105321  0.033849  0.091880
..        …         …         …         …         …         …         …
995  0.033738 -0.100029  0.078302  0.032144 -0.115762 -0.072146 -0.020075
996  0.079593  0.027876  0.083336  0.049230 -0.011094  0.074444  0.087196
997 -0.021456 -0.033678  0.000266 -0.007823 -0.062855 -0.027728  0.015990
998  0.025784 -0.048446  0.067398  0.035877 -0.009344 -0.036641 -0.019036
999 -0.076676 -0.037880 -0.034982 -0.007734  0.002982 -0.045348  0.001137

           X7        X8        X9       X10       X11       X12        X13
0    0.044953 -0.042449  0.011405 -0.001525  0.060797  0.027889  efectores
2    0.025403 -0.030426 -0.059121 -0.045512 -0.020293 -0.103129  efectores
3    0.104541  0.029003 -0.009017 -0.000111  0.060676  0.057337  efectores
4    0.053961  0.008187 -0.025103 -0.039466  0.024717  0.022372  efectores
5    0.045078  0.087138  0.003291 -0.053373  0.058190  0.049220  efectores
..        …         …         …         …         …         …         …
995  0.040706  0.011009  0.017060 -0.016672  0.003936 -0.069432  efectores
```

```
996  0.025398  0.061455  0.080925 -0.009013  0.039406  0.044597  efectores
997  0.000198 -0.031937 -0.007625 -0.012907 -0.046566  0.005687  efectores
998 -0.017320 -0.016800 -0.005480 -0.017440  0.071638  0.032435  efectores
999 -0.037100  0.010640 -0.005271  0.000514  0.019842 -0.020818  efectores

[917 rows x 14 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	917.000000	917.000000	917.000000	917.000000	917.000000	917.000000	
mean	0.007997	-0.019597	0.024262	0.027189	-0.002006	-0.002454	
std	0.055621	0.062679	0.053999	0.058772	0.055562	0.055486	
min	-0.165512	-0.220391	-0.154115	-0.146279	-0.195525	-0.184154	
25%	-0.027350	-0.059376	-0.011408	-0.010328	-0.034928	-0.039650	
50%	0.008165	-0.019116	0.024481	0.026374	-0.005295	-0.003818	
75%	0.044149	0.019137	0.059972	0.064094	0.035182	0.033588	
max	0.187270	0.193867	0.199601	0.214458	0.182264	0.169732	

	X6	X7	X8	X9	X10	X11	\
count	917.000000	917.000000	917.000000	917.000000	917.000000	917.000000	
mean	0.020070	0.010009	0.004982	0.006574	0.010615	0.008753	
std	0.056197	0.052901	0.056057	0.051547	0.052533	0.050885	
min	-0.174143	-0.183317	-0.182743	-0.155265	-0.179524	-0.181843	
25%	-0.016624	-0.023609	-0.027697	-0.028548	-0.022347	-0.022282	
50%	0.019716	0.008690	0.007357	0.009972	0.011800	0.010227	
75%	0.056128	0.042246	0.039148	0.039568	0.046056	0.041233	
max	0.209767	0.176539	0.193857	0.185928	0.185330	0.189575	

	X12
count	917.000000
mean	0.003217
std	0.054110
min	-0.183688
25%	-0.031885
50%	0.005574
75%	0.038069
max	0.176506

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

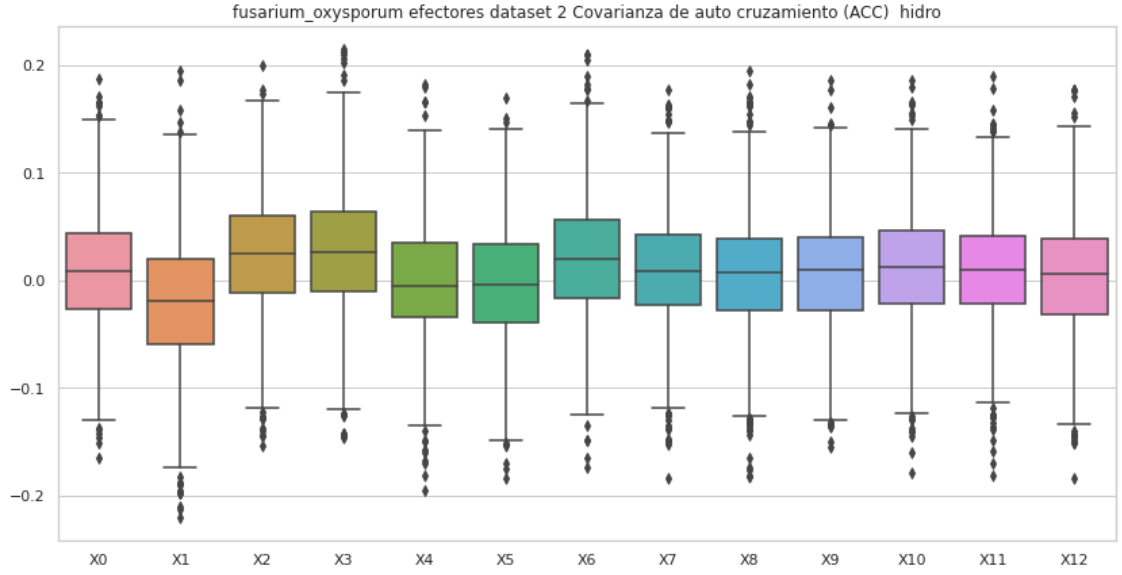
```
           X0        X1        X2        X3        X4        X5        X6  \
1    0.038565  0.053407  0.140912  0.060944  0.045415 -0.035036  0.015974
2   -0.008475 -0.043868 -0.080271  0.069621 -0.043782  0.016885 -0.036962
3    0.118964 -0.062463 -0.024118  0.032381  0.050950 -0.013069 -0.011878
4    0.033462 -0.003860  0.007301  0.088887  0.011326 -0.004329  0.006842
5    0.104719 -0.020709  0.016395 -0.052619  0.007183 -0.037348  0.042107
..        …        …        …        …        …        …        …
995 -0.005968 -0.087170  0.065661  0.016165 -0.039013  0.053640 -0.018549
996  0.077637 -0.008517  0.043984 -0.031139 -0.017858 -0.005752  0.061961
997  0.038523 -0.058547  0.001796  0.009286 -0.024215 -0.006867 -0.039591
998  0.036451 -0.071031 -0.003442  0.018739  0.010509 -0.032165  0.064101
999  0.156846  0.039771 -0.027168  0.078630 -0.086827  0.138716  0.160085

           X7        X8        X9       X10       X11       X12           X13
1    0.033185 -0.062923  0.006817 -0.000866  0.050034  0.003383  no_efectores
2    0.007995  0.039370  0.058137  0.011844 -0.021754 -0.005142  no_efectores
3   -0.030740  0.004477  0.090616  0.053325  0.037941 -0.079041  no_efectores
4    0.024606  0.019965  0.050126  0.015663  0.049120 -0.037189  no_efectores
5    0.040583 -0.100798 -0.053438  0.054604 -0.090010 -0.076618  no_efectores
..        …        …        …        …        …        …        …
995 -0.001135  0.061599 -0.066113 -0.021461 -0.004163  0.001742  no_efectores
996  0.089118 -0.088676 -0.064307 -0.081366 -0.171581  0.004402  no_efectores
997  0.064838  0.044713  0.012469  0.008170  0.007526  0.045121  no_efectores
998  0.007310 -0.041264  0.003133  0.020924  0.030442 -0.008109  no_efectores
999  0.086589 -0.133838 -0.003748 -0.095754 -0.020778 -0.049564  no_efectores

[909 rows x 14 columns]
```

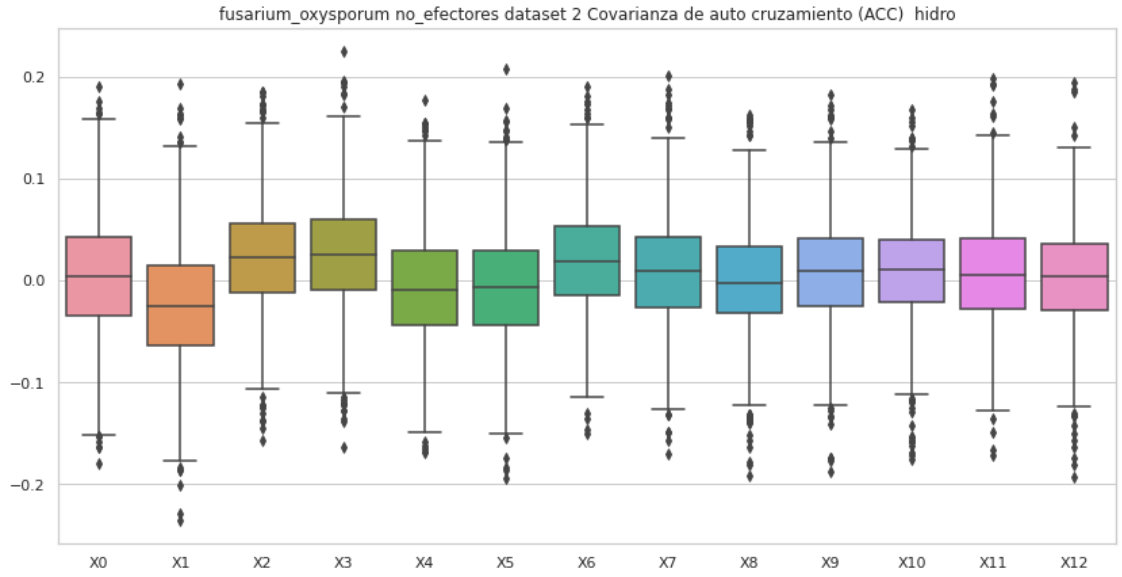
Covarianza de auto cruzamiento (ACC) no_efectores fusarium_oxysporum dataset 2, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  909.000000  909.000000  909.000000  909.000000  909.000000  909.000000
mean     0.005293   -0.022999    0.022993    0.024617   -0.006590   -0.005236
std      0.059648    0.063080    0.053931    0.056334    0.056938    0.058297
min     -0.179253   -0.235282   -0.156680   -0.163454   -0.169239   -0.193615
25%     -0.034891   -0.063526   -0.011952   -0.009926   -0.043890   -0.043225
50%      0.003671   -0.025370    0.022745    0.025761   -0.009768   -0.006868
75%      0.043031    0.014981    0.055623    0.059583    0.029024    0.029111
max      0.190123    0.192611    0.184735    0.224589    0.177137    0.207870

                X6          X7          X8          X9         X10         X11  \
count  909.000000  909.000000  909.000000  909.000000  909.000000  909.000000
mean     0.020281    0.009341    0.000449    0.008575    0.009026    0.006759
std      0.052621    0.053707    0.052925    0.053949    0.052081    0.052309
```

min	-0.149947	-0.169937	-0.190771	-0.187055	-0.175824	-0.171581
25%	-0.014155	-0.026684	-0.032195	-0.024714	-0.021461	-0.027640
50%	0.018421	0.009727	-0.002067	0.009046	0.010280	0.005594
75%	0.053689	0.041994	0.033261	0.040660	0.039374	0.040944
max	0.190506	0.200257	0.161629	0.181614	0.167739	0.198444

	X12
count	909.000000
mean	0.002929
std	0.052848
min	-0.192960
25%	-0.028852
50%	0.004622
75%	0.036294
max	0.194197




fusarium_oxysporum no_efectores dataset 2 Covarianza de auto cruzamiento (ACC) hidro
0.2
0.1
0.0
−0.1
−0.2
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12