ds3_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 = 3
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 3, con
valores atípicos.
Valores del documento csv.

         X0      X1     X2     X3     X4      X5     X6      X7     X8     X9  \
0    10.204   5.102  2.721  5.782  1.361   6.463  2.381  10.544  3.061  4.082
1     8.304   4.152  4.498  6.228  0.865   8.304  3.633   5.363  2.941  3.806
2     9.179   7.729  1.932  8.213  0.483   7.246  3.865   5.314  2.899  2.899
3    10.526   3.204  5.034  4.119  1.373   3.204  4.348   7.323  2.975  4.805
4     5.545   8.688  5.176  5.915  1.664   7.948  4.806   5.176  2.588  4.991
..      …       …      …      …      …       …      …       …      …      …
995  10.127   3.797  1.266  7.595  1.582   6.962  4.114   7.595  3.165  6.962
996   6.535   5.847  3.955  5.503  0.860   7.395  5.589   5.159  2.150  6.019
997   8.937  10.894  3.327  7.502  0.587  11.220  4.044   5.023  3.653  2.609
998   6.977   6.566  3.967  5.198  1.778   6.566  4.651   5.198  2.462  4.514
999   9.846   3.089  3.282  3.089  1.544   3.282  1.544   7.915  1.737  7.722
```

```
     …     X11    X12    X13    X14    X15    X16    X17    X18    X19  \
0    …   5.102  2.381  3.741  3.061  7.483  5.782  2.041  1.361  9.524
1    …  10.035  2.422  4.498  5.190  5.536  4.325  1.211  3.287  6.920
2    …   5.314  4.348  2.415  2.899  6.763  6.280  0.966  2.899  7.729
3    …   6.865  1.144  5.263  3.890  6.865  7.323  3.204  5.034  6.407
4    …   6.839  2.033  3.142  5.176  8.503  6.470  1.664  1.479  3.512
..   …     …    …    …    …    …    …    …    …    …
995  …   5.063  1.899  1.899  5.063  6.013  6.013  2.215  1.899  7.278
996  …   6.793  2.923  3.869  5.073  6.793  5.331  0.860  3.611  5.159
997  …   8.806  1.044  1.044  6.001  9.132  3.979  0.652  1.566  3.327
998  …   3.283  1.915  4.378  5.198  9.986  7.387  2.052  2.462  4.651
999  …   2.124  3.089  7.143  3.861  6.564  8.880  2.703  3.475  7.722

            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 3, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      8.155106     5.825163     3.873596     5.663506     1.504349
std       2.384638     2.216708     1.507853     1.795504     1.193571
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       6.703000     4.446500     2.964500     4.585250     0.755750
50%       7.993500     5.653000     3.722500     5.714000     1.289500
75%       9.357000     7.042000     4.587750     6.667000     1.954250
max      23.158000    17.143000    16.981000    12.791000    10.526000

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      6.201953     4.068910     6.616051     2.484965     5.225028
std       2.330782     1.830632     2.293229     1.190881     1.713239
```

```
min       0.000000     0.000000     0.000000     0.000000     0.833000
25%       4.823750     2.982000     5.085000     1.741750     4.224500
50%       6.077000     3.857500     6.446000     2.392000     5.097000
75%       7.401000     4.771250     7.874000     3.137250     6.189250
max      25.676000    21.667000    24.363000     8.654000    13.793000

               X10          X11          X12          X13          X14  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      8.871221     5.050389     2.362169     3.802543     5.788848
std       2.489582     2.202072     1.142870     1.596464     2.254008
min       0.000000     0.000000     0.238000     0.000000     0.000000
25%       7.343000     3.533000     1.652250     2.781000     4.385000
50%       8.840000     4.796500     2.180500     3.682500     5.526000
75%      10.344500     6.263250     2.894500     4.673750     6.835750
max      19.921000    16.438000    16.129000    12.000000    16.667000

               X15          X16          X17          X18          X19
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      8.000149     6.066231     1.610260     2.876146     5.953405
std       2.516416     2.097649     0.970564     1.344996     1.856549
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       6.393750     4.911500     0.955000     2.049250     4.842250
50%       7.746000     5.913000     1.518000     2.753500     5.843000
75%       9.337500     6.878750     2.181250     3.641750     7.026500
max      24.719000    21.839000     9.091000    13.514000    12.931000
```

no_efectores

Composición de aminoácidos (AAC) no_efectores fusarium_oxysporum dataset 3, con valores atípicos.

Valores del documento csv.

```
         X0      X1     X2     X3     X4      X5     X6     X7     X8     X9  \
0     9.315   7.671  3.014  4.932  0.000  10.959  3.836  3.836  1.644  8.767
1    10.963   3.654  3.987  3.987  0.332   5.648  2.326  7.973  0.997  5.648
2     5.417   2.917  1.667  5.417  1.667   7.083  5.417  5.417  2.500  7.917
3     7.427   3.714  4.775  2.387  1.061   2.122  2.653  8.223  1.061  9.019
4     2.721  10.204  4.082  6.122  1.361   3.401  2.041  6.122  3.401  3.401
..      …      …    …    …      …       …    …      …      …    …
995  10.919   7.279  5.199  7.452  1.386   6.066  2.946  4.853  3.640  5.893
996   9.958   5.085  4.661  4.873  1.271   6.780  2.754  6.568  2.331  6.144
997   8.668   5.708  4.228  4.651  1.903   6.554  3.805  7.188  2.537  6.554
998   8.976   7.854  2.665  5.750  2.945   5.750  5.610  6.592  2.104  3.787
999   6.119   5.402  2.103  4.159  3.107   8.748  2.390  4.637  5.354  7.553

     …    X11    X12    X13    X14     X15     X16    X17    X18    X19  \
0    …  9.315  2.466  1.918  6.849   5.479   3.562  0.274  3.014  4.932
```

```
1    …  8.638  1.993  3.654  2.326   6.977   8.306  0.997  3.322  6.977
2    …  4.583  4.167  2.917  3.333   9.583   4.167  2.083  5.000  6.250
3    …  4.775  1.857  9.284  5.040   5.305   5.570  2.387  4.509  7.162
4    …  3.401  0.680  4.762  9.524  13.605  10.204  1.361  1.361  5.442
..   …    …    …    …    …     …      …     …    …    …
995  …  4.506  1.560  2.600  3.293   9.705   5.546  1.213  1.906  3.466
996  …  4.661  1.907  4.873  5.085   7.627   6.144  1.271  2.966  8.475
997  …  5.285  3.171  3.171  4.863   7.400   5.074  0.846  2.326  6.131
998  …  2.525  2.525  4.067  5.330   7.714   5.610  1.823  2.384  6.031
999  …  6.931  1.530  3.920  1.195   6.788   7.505  1.912  3.537  6.501

                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]
```

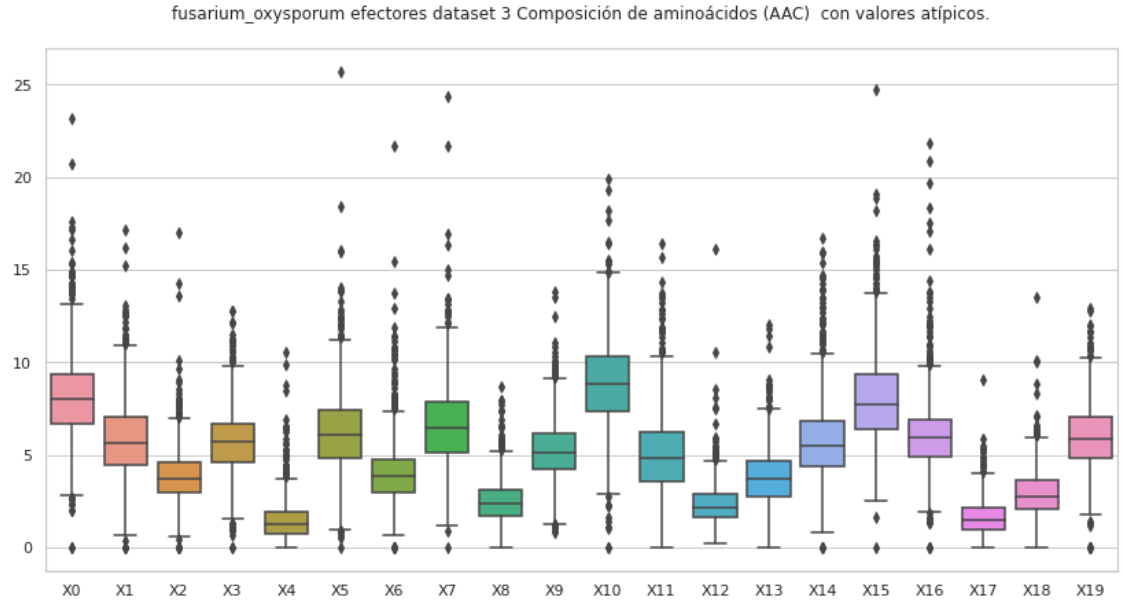
Composición de aminoácidos (AAC) no_efectores fusarium_oxysporum dataset 3, con valores atípicos.
Estadísticas.

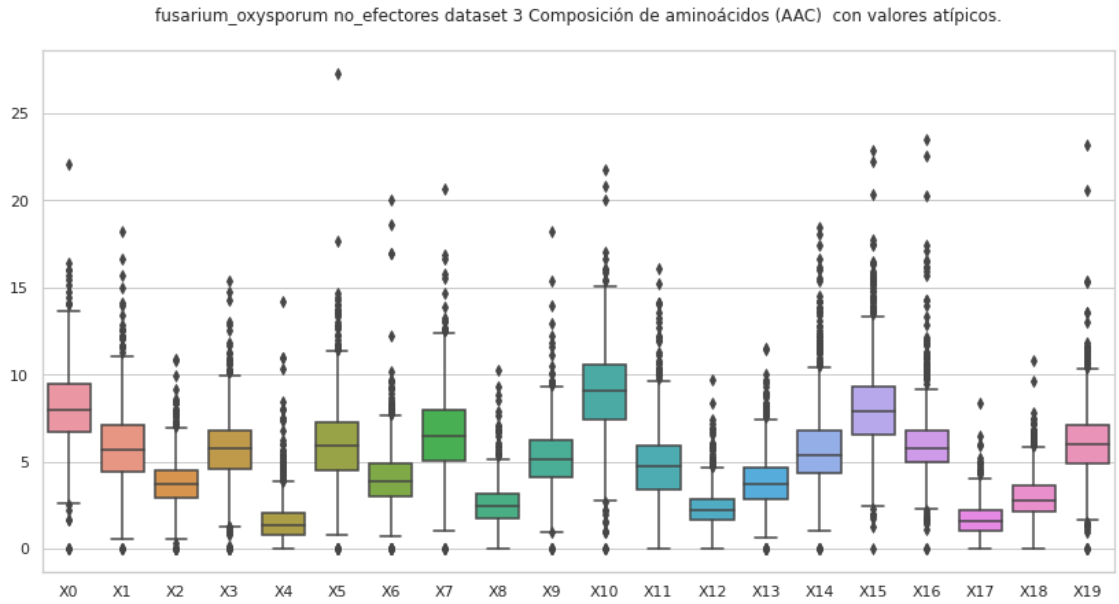
```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      8.138903     5.839727     3.801352     5.674745     1.584666
std       2.270966     2.265445     1.464776     2.011903     1.355950
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       6.716000     4.396000     2.897750     4.580750     0.785750
50%       7.993000     5.701000     3.704000     5.748500     1.365000
75%       9.484500     7.092250     4.525750     6.774000     2.014750
max      22.078000    18.182000    10.914000    15.363000    14.156000

                X5          X6           X7           X8           X9  \
count  1000.000000  1000.00000  1000.000000  1000.000000  1000.000000
mean      5.975062     4.02935     6.640659     2.498192     5.223348
std       2.525931     1.79665     2.328634     1.235208     1.866947
min       0.000000     0.00000     0.000000     0.000000     0.000000
25%       4.492250     2.97775     5.058000     1.702500     4.093750
50%       5.882000     3.86650     6.488500     2.422000     5.135000
```

75%	7.238250	4.85025	8.000000	3.125000	6.189500
max	27.273000	20.05500	20.690000	10.280000	18.182000

	X10	X11	X12	X13	X14 \
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	8.951983	4.868723	2.358161	3.861031	5.754984
std	2.492257	2.180254	1.108679	1.594823	2.472857
min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	7.413000	3.396000	1.625000	2.861750	4.304750
50%	9.042000	4.723000	2.222000	3.741500	5.337500
75%	10.545500	5.906000	2.865250	4.682750	6.759500
max	21.739000	16.058000	9.677000	11.538000	18.447000

	X15	X16	X17	X18	X19
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	8.133277	6.006643	1.675747	2.903690	6.079772
std	2.623650	2.163883	0.989689	1.332576	2.078128
min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	6.550000	4.935500	1.026250	2.112250	4.906000
50%	7.910000	5.784500	1.597500	2.769500	5.953000
75%	9.324250	6.742250	2.238250	3.626750	7.101000
max	22.857000	23.529000	8.333000	10.811000	23.171000





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 3, sin
valores atípicos.
Valores del documento csv.

         X0      X1     X2     X3     X4      X5     X6      X7     X8     X9  \
0    10.204   5.102  2.721  5.782  1.361   6.463  2.381  10.544  3.061  4.082
1     8.304   4.152  4.498  6.228  0.865   8.304  3.633   5.363  2.941  3.806
2     9.179   7.729  1.932  8.213  0.483   7.246  3.865   5.314  2.899  2.899
3    10.526   3.204  5.034  4.119  1.373   3.204  4.348   7.323  2.975  4.805
4     5.545   8.688  5.176  5.915  1.664   7.948  4.806   5.176  2.588  4.991
..      …       …      …      …      …       …      …       …      …      …
995  10.127   3.797  1.266  7.595  1.582   6.962  4.114   7.595  3.165  6.962
996   6.535   5.847  3.955  5.503  0.860   7.395  5.589   5.159  2.150  6.019
997   8.937  10.894  3.327  7.502  0.587  11.220  4.044   5.023  3.653  2.609
998   6.977   6.566  3.967  5.198  1.778   6.566  4.651   5.198  2.462  4.514
999   9.846   3.089  3.282  3.089  1.544   3.282  1.544   7.915  1.737  7.722

     …     X11    X12    X13    X14    X15    X16    X17    X18    X19  \
0    …   5.102  2.381  3.741  3.061  7.483  5.782  2.041  1.361  9.524
1    …  10.035  2.422  4.498  5.190  5.536  4.325  1.211  3.287  6.920
2    …   5.314  4.348  2.415  2.899  6.763  6.280  0.966  2.899  7.729
3    …   6.865  1.144  5.263  3.890  6.865  7.323  3.204  5.034  6.407
4    …   6.839  2.033  3.142  5.176  8.503  6.470  1.664  1.479  3.512
..   …      …      …      …      …      …      …      …      …
995  …   5.063  1.899  1.899  5.063  6.013  6.013  2.215  1.899  7.278
```

```
996  …   6.793  2.923  3.869  5.073  6.793  5.331  0.860  3.611  5.159
997  …   8.806  1.044  1.044  6.001  9.132  3.979  0.652  1.566  3.327
998  …   3.283  1.915  4.378  5.198  9.986  7.387  2.052  2.462  4.651
999  …   2.124  3.089  7.143  3.861  6.564  8.880  2.703  3.475  7.722

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

[850 rows x 21 columns]
```

Composición de aminoácidos (AAC) efectores fusarium_oxysporum dataset 3, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count   850.000000  850.000000  850.000000  850.000000  850.000000  850.000000
mean      8.093745    5.827448    3.850478    5.703171    1.457832    6.206824
std       2.005597    1.888448    1.222740    1.570214    0.904769    1.936622
min       2.000000    0.000000    0.000000    0.641000    0.000000    0.806000
25%       6.812750    4.540500    3.062750    4.753000    0.827250    4.970000
50%       8.000000    5.734000    3.747500    5.769000    1.312000    6.160000
75%       9.244750    6.974500    4.558500    6.656500    1.907250    7.360000
max      14.842000   11.864000    7.944000   10.932000    5.054000   12.346000

                X6          X7          X8          X9         X10         X11  \
count   850.000000  850.000000  850.000000  850.000000  850.000000  850.000000
mean      3.919240    6.640548    2.480524    5.330852    9.140553    4.988766
std       1.346946    2.027162    1.003076    1.529310    2.108156    1.823871
min       0.000000    1.460000    0.000000    1.031000    3.248000    0.000000
25%       3.012000    5.161500    1.817000    4.388750    7.761000    3.618000
50%       3.847000    6.477500    2.414500    5.193000    9.083000    4.806500
75%       4.675250    7.910500    3.127250    6.245000   10.493000    6.164750
max       9.440000   13.462000    5.769000   10.354000   15.528000   11.616000

               X12         X13         X14         X15         X16         X17  \
count   850.000000  850.000000  850.000000  850.000000  850.000000  850.000000
mean      2.325808    3.873460    5.688402    7.926239    5.957600    1.640894
```

std	0.914029	1.394529	1.837036	2.153476	1.568919	0.851868
min	0.238000	0.000000	0.000000	2.667000	0.000000	0.000000
25%	1.695000	3.009250	4.525500	6.472000	4.993250	1.011250
50%	2.193500	3.766000	5.518000	7.734500	5.914000	1.561000
75%	2.888750	4.700500	6.688750	9.163250	6.823750	2.217250
max	5.512000	8.387000	12.392000	15.468000	12.274000	4.518000

	X18	X19
count	850.000000	850.000000
mean	2.885991	6.061618
std	1.131027	1.648222
min	0.000000	1.205000
25%	2.140750	4.986250
50%	2.775500	5.915000
75%	3.643250	7.051000
max	6.614000	11.079000

no_efectores
Composición de aminoácidos (AAC) no_efectores fusarium_oxysporum dataset 3, sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9	\
0	9.315	7.671	3.014	4.932	0.000	10.959	3.836	3.836	1.644	8.767	
1	10.963	3.654	3.987	3.987	0.332	5.648	2.326	7.973	0.997	5.648	
2	5.417	2.917	1.667	5.417	1.667	7.083	5.417	5.417	2.500	7.917	
4	2.721	10.204	4.082	6.122	1.361	3.401	2.041	6.122	3.401	3.401	
5	7.303	3.933	3.933	5.056	1.966	5.337	3.371	8.146	2.809	6.180	
..	…	…	…	…	…	…	…	…	…		
995	10.919	7.279	5.199	7.452	1.386	6.066	2.946	4.853	3.640	5.893	
996	9.958	5.085	4.661	4.873	1.271	6.780	2.754	6.568	2.331	6.144	
997	8.668	5.708	4.228	4.651	1.903	6.554	3.805	7.188	2.537	6.554	
998	8.976	7.854	2.665	5.750	2.945	5.750	5.610	6.592	2.104	3.787	
999	6.119	5.402	2.103	4.159	3.107	8.748	2.390	4.637	5.354	7.553	

	…	X11	X12	X13	X14	X15	X16	X17	X18	X19	\
0	…	9.315	2.466	1.918	6.849	5.479	3.562	0.274	3.014	4.932	
1	…	8.638	1.993	3.654	2.326	6.977	8.306	0.997	3.322	6.977	
2	…	4.583	4.167	2.917	3.333	9.583	4.167	2.083	5.000	6.250	
4	…	3.401	0.680	4.762	9.524	13.605	10.204	1.361	1.361	5.442	
5	…	5.337	1.966	4.213	6.461	7.584	6.180	1.685	1.685	6.180	
..	…	…	…	…	…	…	…	…	…		
995	…	4.506	1.560	2.600	3.293	9.705	5.546	1.213	1.906	3.466	
996	…	4.661	1.907	4.873	5.085	7.627	6.144	1.271	2.966	8.475	
997	…	5.285	3.171	3.171	4.863	7.400	5.074	0.846	2.326	6.131	
998	…	2.525	2.525	4.067	5.330	7.714	5.610	1.823	2.384	6.031	

```
999  …  6.931  1.530  3.920  1.195   6.788   7.505  1.912  3.537  6.501

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

[838 rows x 21 columns]
```

Composición de aminoácidos (AAC) no_efectores fusarium_oxysporum dataset 3, sin valores atípicos.
Estadísticas.

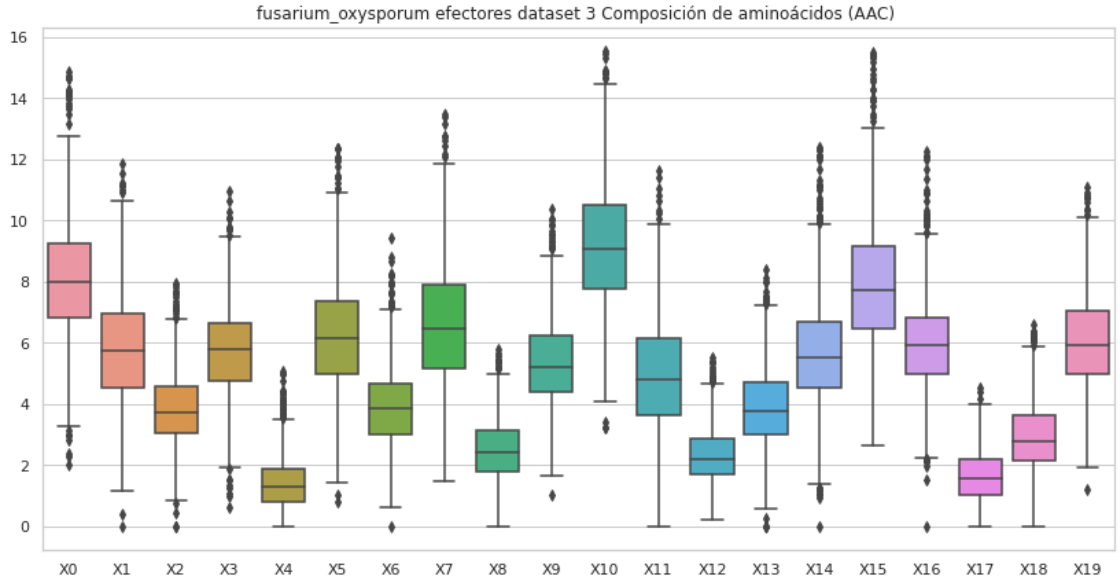
	X0	X1	X2	X3	X4	X5	\
count	838.000000	838.000000	838.000000	838.000000	838.000000	838.000000	
mean	8.172169	5.819342	3.806365	5.762327	1.491835	5.962805	
std	1.936518	1.923949	1.245379	1.705135	0.965456	2.087258	
min	2.247000	1.222000	0.000000	0.000000	0.000000	0.000000	
25%	6.864750	4.507000	2.989000	4.827000	0.847250	4.695000	
50%	8.031000	5.701500	3.715500	5.811500	1.366000	5.981500	
75%	9.427250	6.939000	4.471750	6.772500	1.906750	7.133500	
max	14.463000	12.500000	8.065000	11.243000	5.584000	13.462000	

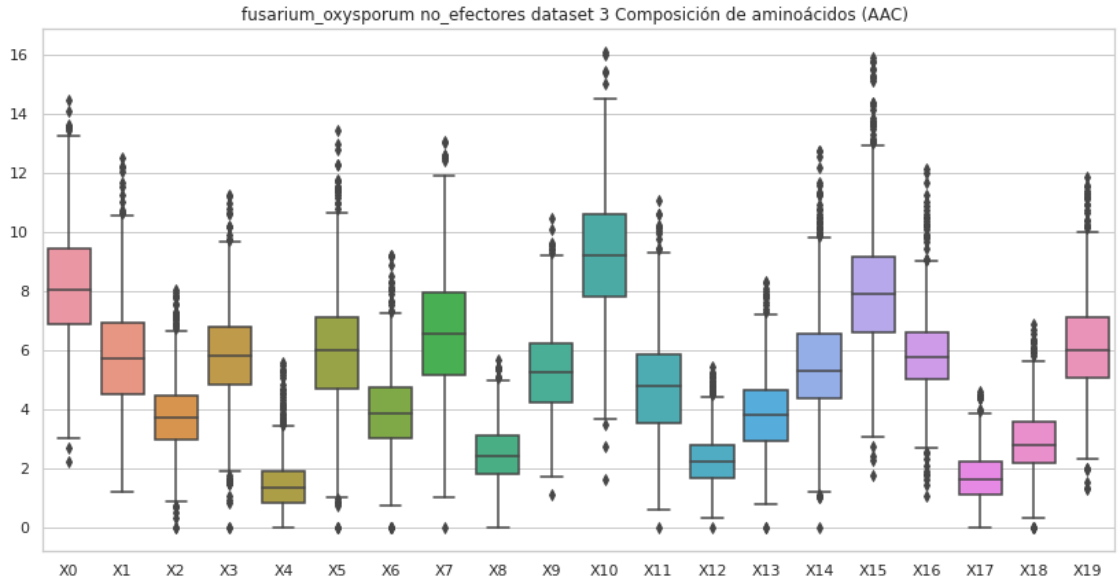
	X6	X7	X8	X9	X10	X11	\
count	838.000000	838.000000	838.000000	838.000000	838.000000	838.000000	
mean	3.961779	6.667852	2.497222	5.332249	9.219408	4.867294	
std	1.398324	2.044811	0.980129	1.541302	2.037519	1.772639	
min	0.000000	0.000000	0.000000	1.111000	1.613000	0.000000	
25%	3.042250	5.162500	1.818000	4.255000	7.832250	3.537750	
50%	3.850000	6.541000	2.439000	5.249500	9.219500	4.799500	
75%	4.751500	7.965250	3.120500	6.250000	10.606000	5.856250	
max	9.233000	13.084000	5.660000	10.458000	16.092000	11.087000	

	X12	X13	X14	X15	X16	X17	\
count	838.000000	838.000000	838.000000	838.000000	838.000000	838.000000	
mean	2.275977	3.891364	5.598016	8.011778	5.867004	1.708142	
std	0.901143	1.364920	1.916232	2.238143	1.472851	0.844929	
min	0.000000	0.000000	0.000000	1.754000	1.075000	0.000000	
25%	1.669500	2.950000	4.386750	6.595750	5.010500	1.128500	

50%	2.211000	3.802000	5.326000	7.900500	5.787500	1.644000
75%	2.778000	4.662500	6.569000	9.150500	6.610000	2.239000
max	5.444000	8.333000	12.766000	15.909000	12.150000	4.587000

	X18	X19
count	838.000000	838.000000
mean	2.924128	6.162936
std	1.124133	1.716983
min	0.000000	1.307000
25%	2.199500	5.080500
50%	2.794000	6.011000
75%	3.597500	7.106250
max	6.877000	11.852000





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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.042668	0.005689	0.024179	0.027023	0.015645	0.044090	0.012800	
1	0.034472	0.003591	0.025854	0.034472	0.018672	0.022263	0.012209	
2	0.040510	0.002132	0.036246	0.031982	0.010661	0.023453	0.012793	
3	0.076976	0.010040	0.030121	0.023428	0.038488	0.053549	0.021754	
4	0.021000	0.006300	0.022400	0.030100	0.011900	0.019600	0.009800	
..	…	…	…	…	…	…	…	
995	0.044521	0.006956	0.033390	0.030608	0.008348	0.033390	0.013913	
996	0.048412	0.006370	0.040768	0.054782	0.028665	0.038220	0.015925	
997	0.010560	0.000694	0.008864	0.013258	0.001233	0.005935	0.004316	
998	0.036227	0.009234	0.026992	0.034096	0.022730	0.026992	0.012786	
999	0.022141	0.003473	0.006946	0.007380	0.016063	0.017800	0.003907	

	X7	X8	X9	…	X74	X75	X76	\
0	0.017067	0.021334	0.032712	…	0.006321	-0.004214	0.029241	
1	0.015800	0.041654	0.035190	…	0.019440	0.024490	0.028272	
2	0.012793	0.023453	0.046906	…	0.014456	0.011120	0.030628	
3	0.035141	0.050202	0.051875	…	-0.024408	-0.042011	0.012833	
4	0.018900	0.025900	0.032900	…	-0.010903	0.016396	0.007286	
..	…	…	…	…	…	…	…	
995	0.030608	0.022260	0.041738	…	-0.011039	0.016321	0.017154	
996	0.044590	0.050323	0.078352	…	-0.026484	-0.014254	0.006328	
997	0.003083	0.010406	0.007862	…	0.014114	0.022680	0.005533	
998	0.023441	0.017048	0.056116	…	-0.000121	0.000916	0.015536	
999	0.017366	0.004776	0.025615	…	0.014248	0.005994	0.013519	

	X77	X78	X79	X80	X81	X82	X83
0	-0.017139	-0.008381	0.040627	0.013711	0.008792	0.016236	efectores
1	-0.005404	0.015658	0.015887	0.014641	0.014646	0.007376	efectores
2	0.019014	0.025494	-0.011759	-0.002178	0.008474	0.000648	efectores
3	-0.004395	0.003368	0.023455	-0.005885	0.006130	0.032685	efectores
4	0.006686	0.010985	0.015335	0.020213	0.030161	-0.003858	efectores
..	…	…	…	…	…	…	
995	0.008615	0.019753	0.025955	-0.008624	-0.009699	0.022561	efectores
996	0.015040	0.026196	0.018350	0.015661	0.025335	0.012971	efectores

```
997  0.013151  0.021507  0.003354  0.015794  0.024870  0.006370  efectores
998 -0.011307 -0.008477  0.006141  0.016402  0.012582  0.013580  efectores
999  0.018311  0.001907  0.008756  0.011769  0.004915  0.013155  efectores

[1000 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro_mass efectores
fusarium_oxysporum dataset 3, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.031097	0.010250	0.028397	0.029763	0.024093	
std	0.517731	0.095592	0.243464	0.219694	0.352315	
min	-14.212789	0.000000	-4.737596	-4.737596	-4.737596	
25%	0.027018	0.003001	0.017009	0.018290	0.010760	
50%	0.035158	0.005665	0.025085	0.026818	0.016814	
75%	0.044790	0.010046	0.034652	0.038607	0.023372	
max	8.062005	3.023252	6.046504	5.038753	10.077506	

	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.022279	0.009947	0.023766	0.015818	0.060673	…	
std	0.302990	0.163423	0.394412	0.489314	0.541441	…	
min	-9.475192	-4.737596	-9.475192	-14.212789	-0.489813	…	
25%	0.021004	0.006306	0.016020	0.014328	0.027389	…	
50%	0.028074	0.010497	0.023035	0.021361	0.039675	…	
75%	0.038183	0.016266	0.031836	0.031092	0.054182	…	
max	1.007751	2.015501	8.062005	6.046504	17.131761	…	

	X73	X74	X75	X76	X77	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.040197	-0.011830	-0.007888	0.011637	0.013494	
std	0.750137	0.370846	0.357610	0.131190	0.630895	
min	-0.212762	-11.408293	-8.999686	-3.434205	-7.128048	
25%	0.005778	-0.006545	-0.001140	0.005871	-0.006124	
50%	0.015485	0.002626	0.006557	0.014491	0.003722	
75%	0.024053	0.011400	0.015843	0.023307	0.012353	
max	23.675281	0.589240	0.828264	1.945707	18.606521	

	X78	X79	X80	X81	X82
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.023330	0.022400	-0.040461	-0.012841	0.043507
std	0.614457	0.372451	1.023720	0.451745	0.947068
min	-2.187845	-2.469827	-30.507632	-12.875697	-0.200344
25%	-0.002099	0.005826	-0.006450	-0.001736	0.004912
50%	0.006325	0.014193	0.003200	0.006228	0.014623

```
75%       0.015725     0.023974     0.011499     0.014834     0.023043
max      19.298124    11.491556     0.120112     0.590223    29.956152

[8 rows x 83 columns]
```

```
no_efectores
Composición de pseudo aminoácidos (PseAAC)  hidro_mass no_efectores
fusarium_oxysporum dataset 3, con valores atípicos.
Valores del documento csv.
```

	X0	X1	X2	X3	X4	X5	X6	\
0	0.020136	0.000000	0.010660	0.023690	0.004146	0.008291	0.003553	
1	0.042116	0.001276	0.015315	0.021696	0.014039	0.030630	0.003829	
2	0.036852	0.011339	0.036852	0.048191	0.019844	0.036852	0.017009	
3	0.026166	0.003738	0.008410	0.007476	0.032707	0.028969	0.003738	
4	0.012712	0.006356	0.028603	0.015890	0.022247	0.028603	0.015890	
..	…	…	…	…	…	…	…	
995	0.061986	0.007871	0.042308	0.034436	0.014758	0.027549	0.020662	
996	0.049551	0.006326	0.024248	0.033737	0.024248	0.032683	0.011597	
997	0.035654	0.007826	0.019131	0.026958	0.013044	0.029567	0.010435	
998	0.042233	0.013858	0.027055	0.027055	0.019137	0.031015	0.009898	
999	0.052667	0.026745	0.035797	0.075298	0.033740	0.039912	0.046084	

	X7	X8	X9	…	X74	X75	X76	\
0	0.018952	0.020136	0.017767	…	0.007650	0.020405	0.006394	
1	0.021696	0.033183	0.043393	…	-0.010003	-0.006996	0.031526	
2	0.053861	0.031183	0.085044	…	0.016044	0.030460	-0.021458	
3	0.031772	0.016821	0.041117	…	0.012285	-0.001369	0.006609	
4	0.015890	0.015890	0.031781	…	0.032366	0.009329	0.019218	
..	…	…	…	…	…	…	…	
995	0.033453	0.025581	0.060018	…	0.010690	0.006681	0.041610	
996	0.030574	0.023194	0.032683	…	-0.001407	-0.000236	0.019100	
997	0.026958	0.021740	0.040871	…	0.004784	0.006692	0.024454	
998	0.017817	0.011878	0.046852	…	0.016407	0.002626	0.022285	
999	0.065011	0.059662	0.091345	…	-0.012293	-0.017788	0.011832	

	X77	X78	X79	X80	X81	X82	X83
0	0.005960	0.021479	0.009386	0.015071	0.027855	0.007733	no_efectores
1	0.001775	0.009281	0.020974	0.002019	0.000368	0.032048	no_efectores
2	0.004257	0.000903	-0.019083	0.079297	0.028508	0.031583	no_efectores
3	0.022137	0.009373	0.011527	0.006938	-0.006661	-0.001153	no_efectores
4	0.022124	0.033193	0.007522	-0.007215	0.018399	0.002788	no_efectores
..	…	…	…	…	…	…	…
995	0.011503	0.009492	0.005292	-0.004569	0.025898	-0.009402	no_efectores
996	-0.009406	0.013742	0.030772	-0.008408	-0.007550	0.022130	no_efectores
997	0.004889	0.002807	0.018981	-0.000699	-0.002860	0.041516	no_efectores

```
998  0.005025  0.013991  0.010827 -0.013707 -0.007703  0.004503  no_efectores
999  0.015601  0.011593  0.017071  0.003722  0.022451  0.011907  no_efectores

[1000 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores
fusarium_oxysporum dataset 3, con valores atípicos.
Estadísticas.

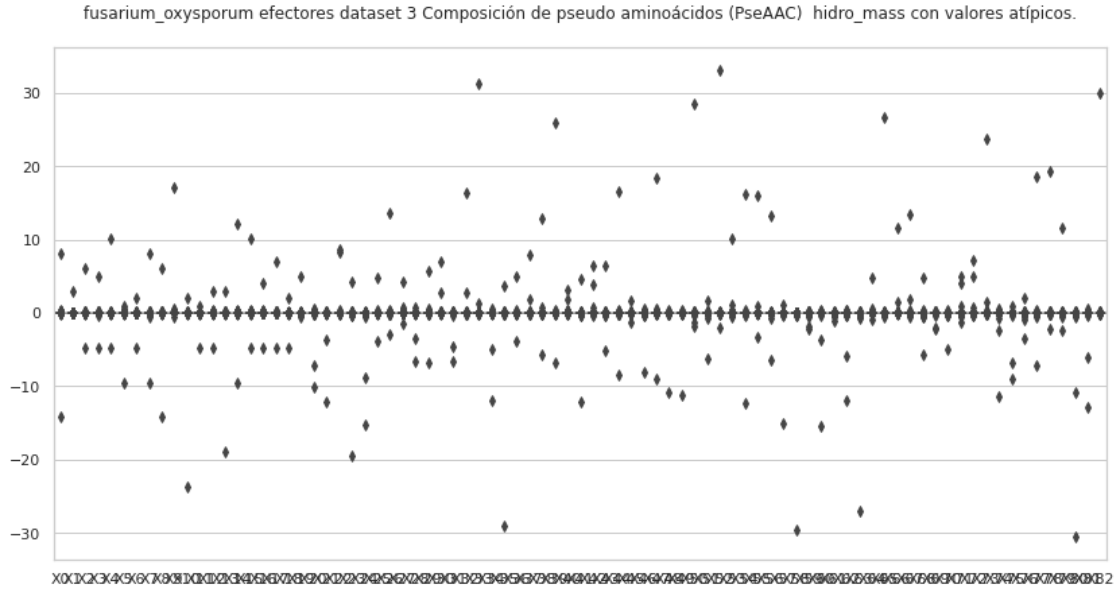
	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.038007	0.007921	0.028252	0.028668	0.018626	
std	0.039917	0.009938	0.024178	0.054704	0.052354	
min	-0.983953	-0.157425	-0.472276	-1.475929	-1.475929	
25%	0.027927	0.003052	0.017515	0.017152	0.011457	
50%	0.036942	0.006213	0.026425	0.027431	0.017481	
75%	0.046492	0.010202	0.037224	0.038166	0.025060	
max	0.320751	0.084139	0.220725	0.577493	0.384995	

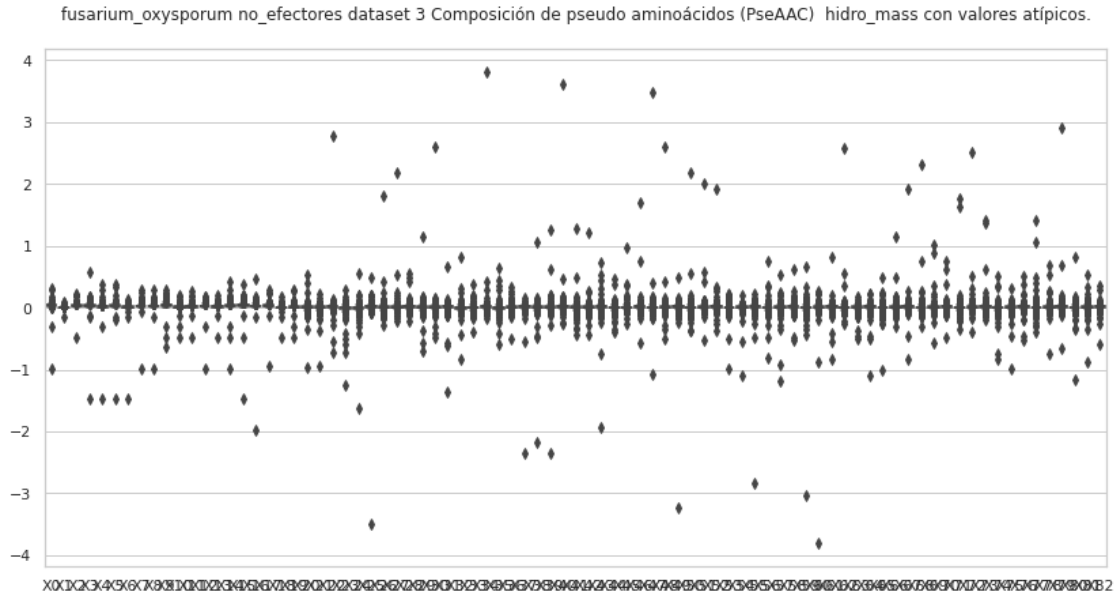
	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.030318	0.011653	0.025680	0.023589	0.044858	…	
std	0.053068	0.048869	0.037537	0.037372	0.041142	…	
min	-1.475929	-1.475929	-0.983953	-0.983953	-0.629701	…	
25%	0.021458	0.006206	0.016386	0.013802	0.028536	…	
50%	0.029027	0.011243	0.023702	0.021896	0.040943	…	
75%	0.039909	0.016651	0.032214	0.031212	0.057321	…	
max	0.384995	0.128301	0.288746	0.288746	0.278886	…	

	X73	X74	X75	X76	X77	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.017467	-0.000096	0.003840	0.015150	0.004339	
std	0.068787	0.048982	0.042680	0.037857	0.067618	
min	-0.202653	-0.826020	-0.978984	-0.536330	-0.383785	
25%	0.005531	-0.006335	-0.002092	0.006306	-0.006853	
50%	0.014426	0.002564	0.006040	0.015833	0.002733	
75%	0.024012	0.010661	0.014299	0.024418	0.011903	
max	1.406025	0.517978	0.306396	0.513924	1.419353	

	X78	X79	X80	X81	X82
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.006682	0.017721	0.001502	0.005846	0.014720
std	0.043761	0.098984	0.053540	0.043607	0.032781
min	-0.735519	-0.652277	-1.159645	-0.870393	-0.586594
25%	-0.002880	0.006119	-0.007208	-0.002638	0.006496
50%	0.005728	0.015326	0.003010	0.005817	0.015490
75%	0.015073	0.025248	0.012079	0.015703	0.025899

```
max        0.675102       2.914833       0.806128       0.538359       0.364653

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.042668	0.005689	0.024179	0.027023	0.015645	0.044090	0.012800	
1	0.034472	0.003591	0.025854	0.034472	0.018672	0.022263	0.012209	
2	0.040510	0.002132	0.036246	0.031982	0.010661	0.023453	0.012793	
3	0.076976	0.010040	0.030121	0.023428	0.038488	0.053549	0.021754	
4	0.021000	0.006300	0.022400	0.030100	0.011900	0.019600	0.009800	
..	…	…	…	…	…	…	…	
995	0.044521	0.006956	0.033390	0.030608	0.008348	0.033390	0.013913	
996	0.048412	0.006370	0.040768	0.054782	0.028665	0.038220	0.015925	
997	0.010560	0.000694	0.008864	0.013258	0.001233	0.005935	0.004316	
998	0.036227	0.009234	0.026992	0.034096	0.022730	0.026992	0.012786	
999	0.022141	0.003473	0.006946	0.007380	0.016063	0.017800	0.003907	

	X7	X8	X9	…	X74	X75	X76	\
0	0.017067	0.021334	0.032712	…	0.006321	-0.004214	0.029241	
1	0.015800	0.041654	0.035190	…	0.019440	0.024490	0.028272	
2	0.012793	0.023453	0.046906	…	0.014456	0.011120	0.030628	
3	0.035141	0.050202	0.051875	…	-0.024408	-0.042011	0.012833	
4	0.018900	0.025900	0.032900	…	-0.010903	0.016396	0.007286	
..	…	…	…	…	…	…	…	
995	0.030608	0.022260	0.041738	…	-0.011039	0.016321	0.017154	
996	0.044590	0.050323	0.078352	…	-0.026484	-0.014254	0.006328	
997	0.003083	0.010406	0.007862	…	0.014114	0.022680	0.005533	
998	0.023441	0.017048	0.056116	…	-0.000121	0.000916	0.015536	
999	0.017366	0.004776	0.025615	…	0.014248	0.005994	0.013519	

	X77	X78	X79	X80	X81	X82	X83
0	-0.017139	-0.008381	0.040627	0.013711	0.008792	0.016236	efectores
1	-0.005404	0.015658	0.015887	0.014641	0.014646	0.007376	efectores
2	0.019014	0.025494	-0.011759	-0.002178	0.008474	0.000648	efectores
3	-0.004395	0.003368	0.023455	-0.005885	0.006130	0.032685	efectores
4	0.006686	0.010985	0.015335	0.020213	0.030161	-0.003858	efectores
..	…	…	…	…	…	…	
995	0.008615	0.019753	0.025955	-0.008624	-0.009699	0.022561	efectores
996	0.015040	0.026196	0.018350	0.015661	0.025335	0.012971	efectores
997	0.013151	0.021507	0.003354	0.015794	0.024870	0.006370	efectores
998	-0.011307	-0.008477	0.006141	0.016402	0.012582	0.013580	efectores
999	0.018311	0.001907	0.008756	0.011769	0.004915	0.013155	efectores

[995 rows x 84 columns]

Composición de pseudo aminoácidos (PseAAC) hidro_mass efectores
fusarium_oxysporum dataset 3, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	995.000000	995.000000	995.000000	995.000000	995.000000	995.000000	
mean	0.037125	0.007199	0.027288	0.029645	0.018882	0.030551	
std	0.015628	0.006189	0.015173	0.016412	0.014637	0.014571	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.027048	0.003049	0.017091	0.018328	0.010786	0.021029	
50%	0.035144	0.005682	0.025056	0.026796	0.016837	0.028073	
75%	0.044756	0.010043	0.034612	0.038420	0.023340	0.038140	
max	0.164447	0.046305	0.163129	0.106851	0.212988	0.133118	

	X6	X7	X8	X9	…	X73	\
count	995.000000	995.000000	995.000000	995.000000	…	995.000000	
mean	0.012606	0.025528	0.024031	0.043515	…	0.014726	
std	0.009942	0.016179	0.014754	0.025172	…	0.019821	
min	0.000000	0.001487	0.000000	0.000000	…	-0.171751	
25%	0.006313	0.016081	0.014354	0.027461	…	0.005748	
50%	0.010510	0.023035	0.021334	0.039658	…	0.015466	
75%	0.016214	0.031719	0.030994	0.054127	…	0.023866	
max	0.114381	0.208858	0.156644	0.277783	…	0.275189	

	X74	X75	X76	X77	X78	X79	\
count	995.000000	995.000000	995.000000	995.000000	995.000000	995.000000	
mean	0.002306	0.007224	0.014724	0.001956	0.006167	0.014310	
std	0.020327	0.018635	0.019023	0.026359	0.020672	0.019391	
min	-0.204623	-0.113813	-0.123846	-0.462898	-0.255896	-0.148374	
25%	-0.006396	-0.001077	0.006007	-0.006030	-0.002088	0.006026	
50%	0.002644	0.006560	0.014682	0.003747	0.006316	0.014385	
75%	0.011427	0.015839	0.023323	0.012289	0.015592	0.023985	
max	0.104147	0.126048	0.155204	0.213801	0.110839	0.137153	

	X80	X81	X82
count	995.000000	995.000000	995.000000
mean	0.001376	0.006038	0.013806
std	0.023323	0.019606	0.017525
min	-0.281737	-0.231772	-0.101374
25%	-0.006335	-0.001658	0.004970
50%	0.003241	0.006294	0.014631
75%	0.011500	0.014869	0.022994
max	0.120112	0.107505	0.121166

[8 rows x 83 columns]

```
no_efectores
Composición de pseudo aminoácidos (PseAAC)  hidro_mass no_efectores
fusarium_oxysporum dataset 3, sin valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.020136  0.000000  0.010660  0.023690  0.004146  0.008291  0.003553
1    0.042116  0.001276  0.015315  0.021696  0.014039  0.030630  0.003829
2    0.036852  0.011339  0.036852  0.048191  0.019844  0.036852  0.017009
3    0.026166  0.003738  0.008410  0.007476  0.032707  0.028969  0.003738
4    0.012712  0.006356  0.028603  0.015890  0.022247  0.028603  0.015890
..        …         …         …         …         …         …         …
995  0.061986  0.007871  0.042308  0.034436  0.014758  0.027549  0.020662
996  0.049551  0.006326  0.024248  0.033737  0.024248  0.032683  0.011597
997  0.035654  0.007826  0.019131  0.026958  0.013044  0.029567  0.010435
998  0.042233  0.013858  0.027055  0.027055  0.019137  0.031015  0.009898
999  0.052667  0.026745  0.035797  0.075298  0.033740  0.039912  0.046084

           X7        X8        X9  …       X74       X75       X76  \
0    0.018952  0.020136  0.017767  …  0.007650  0.020405  0.006394
1    0.021696  0.033183  0.043393  … -0.010003 -0.006996  0.031526
2    0.053861  0.031183  0.085044  …  0.016044  0.030460 -0.021458
3    0.031772  0.016821  0.041117  …  0.012285 -0.001369  0.006609
4    0.015890  0.015890  0.031781  …  0.032366  0.009329  0.019218
..        …         …         …  …         …         …         …
995  0.033453  0.025581  0.060018  …  0.010690  0.006681  0.041610
996  0.030574  0.023194  0.032683  … -0.001407 -0.000236  0.019100
997  0.026958  0.021740  0.040871  …  0.004784  0.006692  0.024454
998  0.017817  0.011878  0.046852  …  0.016407  0.002626  0.022285
999  0.065011  0.059662  0.091345  … -0.012293 -0.017788  0.011832

           X77       X78       X79       X80       X81       X82           X83
0    0.005960  0.021479  0.009386  0.015071  0.027855  0.007733  no_efectores
1    0.001775  0.009281  0.020974  0.002019  0.000368  0.032048  no_efectores
2    0.004257  0.000903 -0.019083  0.079297  0.028508  0.031583  no_efectores
3    0.022137  0.009373  0.011527  0.006938 -0.006661 -0.001153  no_efectores
4    0.022124  0.033193  0.007522 -0.007215  0.018399  0.002788  no_efectores
..        …         …         …         …         …         …           …
995  0.011503  0.009492  0.005292 -0.004569  0.025898 -0.009402  no_efectores
996 -0.009406  0.013742  0.030772 -0.008408 -0.007550  0.022130  no_efectores
997  0.004889  0.002807  0.018981 -0.000699 -0.002860  0.041516  no_efectores
998  0.005025  0.013991  0.010827 -0.013707 -0.007703  0.004503  no_efectores
999  0.015601  0.011593  0.017071  0.003722  0.022451  0.011907  no_efectores

[956 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores
fusarium_oxysporum dataset 3, sin valores atípicos.
Estadísticas.

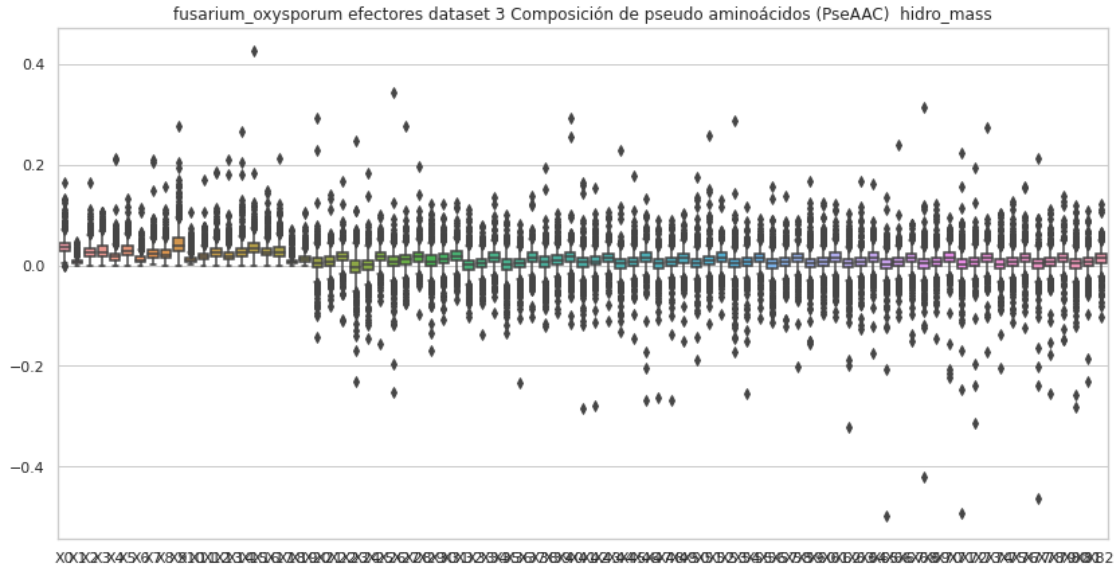
	X0	X1	X2	X3	X4	X5	\
count	956.000000	956.000000	956.000000	956.000000	956.000000	956.000000	
mean	0.037510	0.007306	0.027417	0.028514	0.018703	0.030477	
std	0.014038	0.005880	0.014167	0.015545	0.010939	0.012839	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.027753	0.003078	0.017279	0.016977	0.011150	0.021385	
50%	0.036303	0.006064	0.026056	0.026985	0.016924	0.028615	
75%	0.045044	0.009921	0.036312	0.036457	0.024291	0.038444	
max	0.101297	0.032802	0.088211	0.110506	0.066594	0.091738	

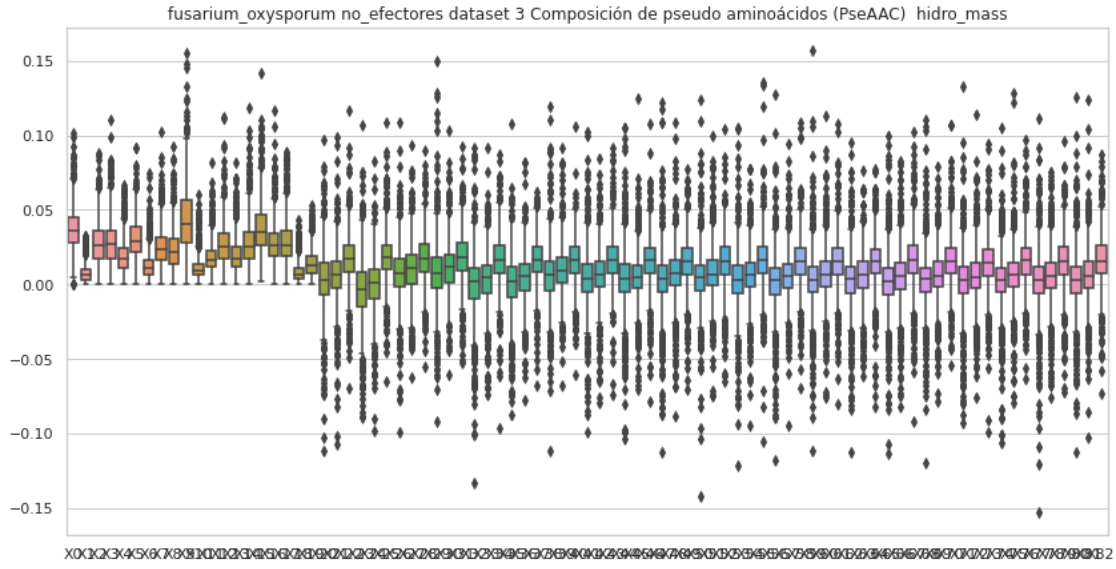
	X6	X7	X8	X9	…	X73	\
count	956.000000	956.000000	956.000000	956.000000	…	956.000000	
mean	0.012307	0.025016	0.023121	0.043585	…	0.014538	
std	0.008713	0.012844	0.013003	0.022233	…	0.017355	
min	0.000000	0.000000	0.000000	0.000000	…	-0.098796	
25%	0.006088	0.016373	0.013734	0.028355	…	0.005934	
50%	0.010962	0.023484	0.021580	0.040082	…	0.014461	
75%	0.015991	0.031584	0.030409	0.056377	…	0.023775	
max	0.074866	0.102197	0.092460	0.155183	…	0.094545	

	X74	X75	X76	X77	X78	X79	\
count	956.000000	956.000000	956.000000	956.000000	956.000000	956.000000	
mean	0.001454	0.006166	0.015468	0.001923	0.005233	0.014897	
std	0.018839	0.017665	0.016702	0.020379	0.017573	0.016831	
min	-0.106709	-0.079061	-0.063667	-0.152850	-0.076117	-0.072511	
25%	-0.005474	-0.001432	0.006583	-0.006472	-0.002764	0.006822	
50%	0.002794	0.006122	0.015894	0.002753	0.005563	0.015411	
75%	0.010567	0.014147	0.024028	0.011609	0.014347	0.024969	
max	0.105263	0.128433	0.096830	0.110992	0.075020	0.095715	

	X80	X81	X82
count	956.000000	956.000000	956.000000
mean	0.002570	0.006055	0.015684
std	0.020099	0.017476	0.017043
min	-0.112971	-0.102445	-0.073034
25%	-0.006478	-0.002494	0.006991
50%	0.002988	0.005773	0.015557
75%	0.011633	0.015161	0.025795
max	0.125263	0.123818	0.087077

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.043079  0.005744  0.024412  0.027284  0.015796  0.044515  0.012924
1    0.058191  0.006062  0.043643  0.058191  0.031520  0.037582  0.020609
2    0.050991  0.002684  0.045624  0.040256  0.013419  0.029521  0.016102
3    0.065546  0.008549  0.025648  0.019949  0.032773  0.045597  0.018524
4    0.045360  0.013608  0.048384  0.065016  0.025704  0.042336  0.021168
..        …         …         …         …         …         …         …
995  0.046003  0.007188  0.034502  0.031627  0.008626  0.034502  0.014376
996  0.053294  0.007012  0.044879  0.060307  0.031556  0.042074  0.017531
997  0.049883  0.003277  0.041872  0.062626  0.005826  0.028036  0.020390
998  0.045830  0.011682  0.034148  0.043134  0.028756  0.034148  0.016175
999  0.048838  0.007661  0.015322  0.016279  0.035431  0.039262  0.008618

           X7        X8        X9  …       X32       X33       X34  \
```

```
0    0.017232  0.021540  0.033027  …  0.030732  0.018096  0.046392
1    0.026671  0.070314  0.059404  … -0.012280  0.019624  0.018991
2    0.016102  0.029521  0.059042  …  0.018999  0.014254  0.038128
3    0.029923  0.042747  0.044172  … -0.002484  0.027664  0.015354
4    0.040824  0.055944  0.071064  …  0.008183  0.003207  0.035150
..        …         …         …  …         …         …
995  0.031627  0.023002  0.043128  …  0.021518  0.025650  0.029339
996  0.049087  0.055398  0.086252  …  0.015070  0.027758  0.016148
997  0.014564  0.049154  0.037139  …  0.012293  0.020327  0.018595
998  0.029655  0.021567  0.070991  …  0.020497  0.016435 -0.005107
999  0.038304  0.010534  0.056498  …  0.017858  0.014313  0.017853

          X35       X36       X37       X38       X39       X40        X41
0    0.042328  0.028314  0.011836  0.029523  0.041018  0.016393  efectores
1    0.018646  0.009855  0.038390  0.047725  0.026817  0.012452  efectores
2    0.019618  0.017378  0.016695  0.038553 -0.014801  0.000816  efectores
3   -0.006010  0.043995  0.032249  0.010928  0.019972  0.027832  efectores
4   -0.001300 -0.003257  0.008593  0.015737  0.033125 -0.008334  efectores
..        …         …         …         …         …         …
995  0.034137  0.015995  0.025230  0.017725  0.026820  0.023312  efectores
996  0.006054  0.003740 -0.000587  0.006967  0.020200  0.014279  efectores
997  0.025743  0.014560  0.020689  0.026136  0.015843  0.030092  efectores
998  0.018624  0.023877  0.008596  0.019654  0.007769  0.017179  efectores
999  0.007510  0.035539  0.023932  0.029819  0.019314  0.029016  efectores

[1000 rows x 42 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass efectores fusarium_oxysporum dataset 3, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.047342     0.009071     0.034945     0.039139     0.023584
std       0.015214     0.007517     0.017246     0.022577     0.013596
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       0.037945     0.004078     0.023394     0.023445     0.014362
50%       0.046314     0.007586     0.032794     0.035457     0.021067
75%       0.054433     0.012487     0.043350     0.049602     0.029956
max       0.213377     0.067324     0.149040     0.257330     0.101011

                X5           X6           X7           X8           X9  …  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
mean      0.038268     0.015760     0.032299     0.031655     0.054871  …
std       0.013366     0.010477     0.015985     0.020382     0.024619  …
min       0.000000     0.000000     0.001574     0.000000     0.000000  …
25%       0.029624     0.008900     0.021623     0.018979     0.038235  …
```

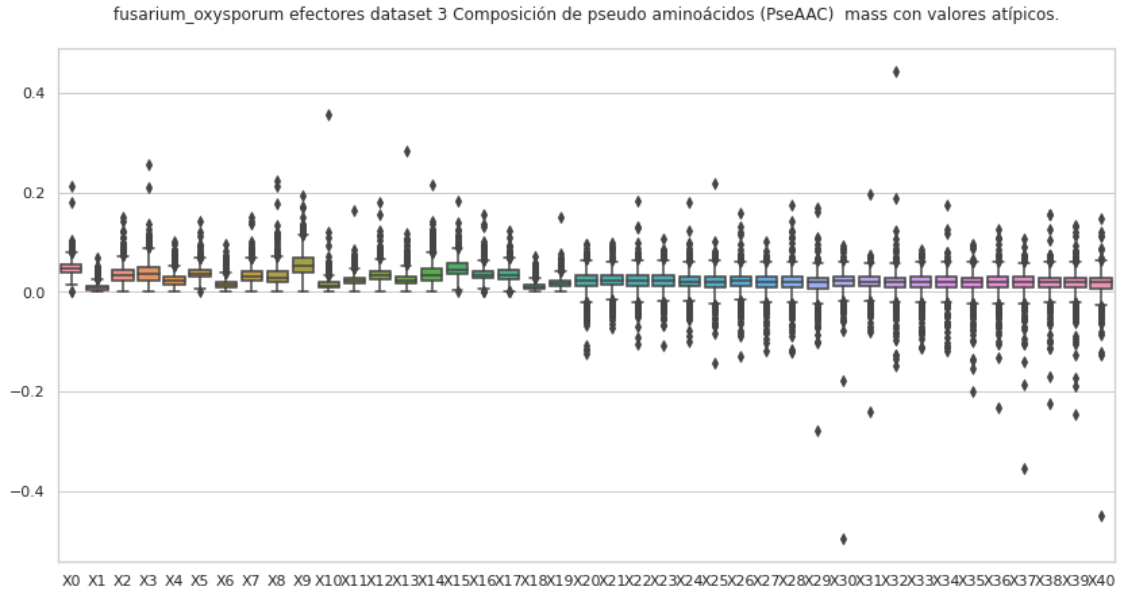
```
50%       0.036919     0.013875     0.030106     0.027796     0.052729  …
75%       0.044911     0.020409     0.040123     0.040115     0.068841  …
max       0.142251     0.096502     0.148772     0.222196     0.192881  …

               X31          X32          X33          X34          X35  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.018421     0.017606     0.017578     0.017968     0.016094
std       0.020682     0.026364     0.021567     0.022257     0.023175
min      -0.239927    -0.148242    -0.112234    -0.117624    -0.200263
25%       0.010385     0.008634     0.009510     0.009570     0.008115
50%       0.020873     0.019837     0.020238     0.020351     0.019660
75%       0.029618     0.028558     0.029761     0.029964     0.028731
max       0.196105     0.443239     0.085598     0.174734     0.096568

               X36          X37          X38          X39          X40
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.017597     0.017403     0.017715     0.017637     0.016323
std       0.023033     0.024926     0.023868     0.024691     0.025732
min      -0.234045    -0.355437    -0.224041    -0.246503    -0.449732
25%       0.008207     0.008386     0.008478     0.008685     0.006558
50%       0.019875     0.020474     0.020171     0.019710     0.018558
75%       0.029412     0.029513     0.029107     0.029060     0.029047
max       0.126566     0.107946     0.155219     0.133540     0.147699

[8 rows x 41 columns]



no_efectores
Composición de pseudo aminoácidos (PseAAC)  mass no_efectores fusarium_oxysporum
dataset 3, con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.057061  0.000000  0.030209  0.067131  0.011748  0.023496  0.010070
1    0.049288  0.001494  0.017923  0.025391  0.016429  0.035846  0.004481
2    0.041434  0.012749  0.041434  0.054183  0.022311  0.041434  0.019123
3    0.048512  0.006930  0.015593  0.013861  0.060640  0.053710  0.006930
4    0.016138  0.008069  0.036310  0.020172  0.028241  0.036310  0.020172
..        …         …         …         …         …         …         …
995  0.070773  0.008987  0.048305  0.039318  0.016851  0.031455  0.023591
996  0.052103  0.006652  0.025497  0.035475  0.025497  0.034366  0.012194
997  0.042375  0.009302  0.022738  0.032040  0.015503  0.035140  0.012403
998  0.053748  0.017636  0.034432  0.034432  0.024355  0.039471  0.012597
999  0.048348  0.024552  0.032862  0.069123  0.030973  0.036639  0.042305

           X7        X8        X9  …       X32       X33       X34  \
0    0.053704  0.057061  0.050348  …  0.018930  0.012169 -0.011679
```

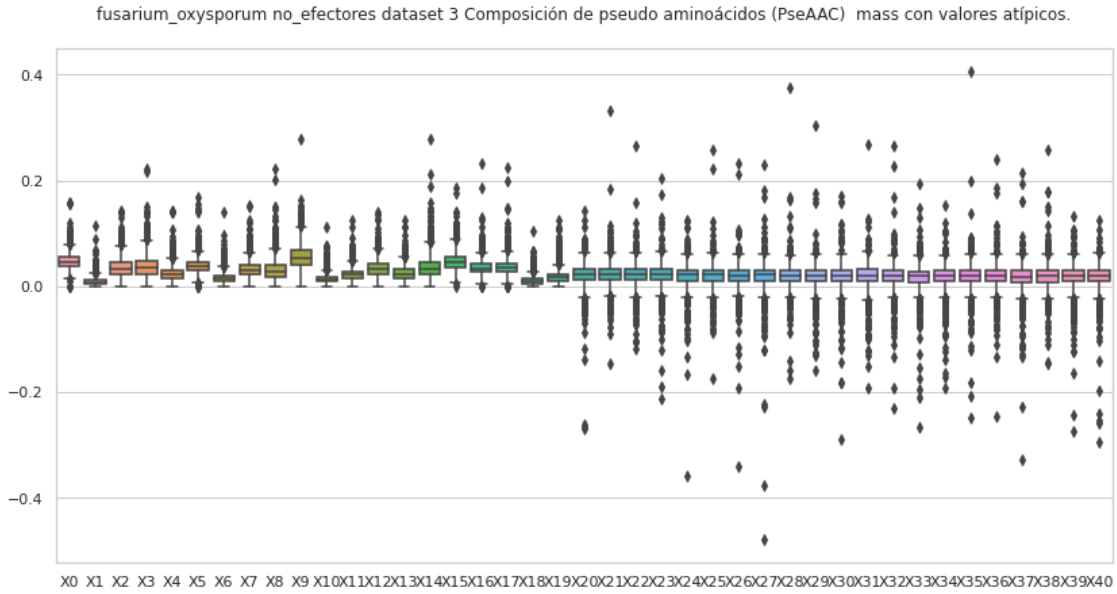
```
1    0.025391  0.038833  0.050782  …  0.002951  0.024086  0.030914
2    0.060557  0.035059  0.095617  … -0.026331  0.019610  0.036739
3    0.058907  0.031186  0.076233  …  0.022984  0.006493  0.047290
4    0.020172  0.020172  0.040344  …  0.008998  0.011761 -0.002476
..        …         …        …  …        …         …
995  0.038195  0.029208  0.068526  … -0.003346  0.005002  0.013303
996  0.032149  0.024389  0.034366  …  0.017948  0.040826  0.026261
997  0.032040  0.025839  0.048577  …  0.034593  0.026675  0.036477
998  0.022675  0.015117  0.059627  …  0.033852  0.016624 -0.003917
999  0.059680  0.054769  0.083854  …  0.002248  0.012326  0.017148

          X35       X36       X37       X38       X39       X40           X41
0    0.020818  0.011369  0.007916  0.018120  0.026597  0.021912  no_efectores
1    0.055448  0.013663  0.025123  0.036895  0.024546  0.037505  no_efectores
2    0.008252  0.044765  0.022613 -0.024125 -0.021455  0.035510  no_efectores
3   -0.009578  0.011922  0.024690  0.012253  0.021372 -0.002138  no_efectores
4    0.046877  0.045182  0.016788  0.024396  0.009548  0.003539  no_efectores
..        …         …         …         …         …         …           …
995  0.049223  0.026745  0.015152  0.047508  0.006042 -0.010735  no_efectores
996  0.039064  0.031233  0.008573  0.020084  0.032357  0.023270  no_efectores
997 -0.004198  0.049528  0.012141  0.029064  0.022560  0.049343  no_efectores
998  0.017654  0.014547  0.010209  0.028361  0.013779  0.005731  no_efectores
999  0.014039  0.006326  0.002425  0.010861  0.015671  0.010931  no_efectores

[1000 rows x 42 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass no_efectores fusarium_oxysporum dataset 3, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.047451     0.009680     0.035096     0.037910     0.024250
std       0.015557     0.009308     0.018354     0.023267     0.015179
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       0.038506     0.004252     0.022260     0.022954     0.015159
50%       0.045874     0.007758     0.033535     0.034210     0.021772
75%       0.054669     0.012421     0.044468     0.048838     0.030479
max       0.158133     0.115567     0.143055     0.220766     0.143771

                X5           X6           X7           X8           X9  …  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
mean      0.038544     0.016063     0.031882     0.030808     0.055549  …
std       0.015222     0.011910     0.016151     0.020512     0.025552  …
min       0.000000     0.000000     0.000000     0.000000     0.000000  …
25%       0.030279     0.008638     0.022139     0.017369     0.039229  …
50%       0.037191     0.014056     0.030231     0.027049     0.052635  …
```

```
75%       0.045025     0.020162     0.039115     0.039119     0.068908  …
max       0.167625     0.139253     0.151614     0.220766     0.279375  …

                X31          X32          X33          X34          X35  \
count   1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean       0.017478     0.017560     0.016668     0.016829     0.018803
std        0.026836     0.026787     0.027229     0.025319     0.029149
min       -0.193700    -0.230970    -0.266710    -0.193238    -0.248260
25%        0.008782     0.009263     0.008257     0.009145     0.009563
50%        0.020231     0.019665     0.019404     0.019881     0.020947
75%        0.031356     0.030009     0.028378     0.029361     0.030889
max        0.267294     0.264413     0.193806     0.153511     0.405958

                X36          X37          X38          X39          X40
count   1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean       0.018618     0.016058     0.018606     0.016831     0.016808
std        0.025035     0.029076     0.025550     0.026982     0.029036
min       -0.247011    -0.328156    -0.146033    -0.275013    -0.294251
25%        0.009555     0.007919     0.008295     0.008861     0.008908
50%        0.019775     0.018194     0.020382     0.019657     0.020483
75%        0.030460     0.028967     0.029591     0.030036     0.030421
max        0.240185     0.213431     0.257798     0.133660     0.125956

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.043079  0.005744  0.024412  0.027284  0.015796  0.044515  0.012924
1    0.058191  0.006062  0.043643  0.058191  0.031520  0.037582  0.020609
2    0.050991  0.002684  0.045624  0.040256  0.013419  0.029521  0.016102
3    0.065546  0.008549  0.025648  0.019949  0.032773  0.045597  0.018524
4    0.045360  0.013608  0.048384  0.065016  0.025704  0.042336  0.021168
..        …         …         …         …         …         …         …
995  0.046003  0.007188  0.034502  0.031627  0.008626  0.034502  0.014376
996  0.053294  0.007012  0.044879  0.060307  0.031556  0.042074  0.017531
997  0.049883  0.003277  0.041872  0.062626  0.005826  0.028036  0.020390
998  0.045830  0.011682  0.034148  0.043134  0.028756  0.034148  0.016175
999  0.048838  0.007661  0.015322  0.016279  0.035431  0.039262  0.008618

           X7        X8        X9  …       X32       X33       X34  \
0    0.017232  0.021540  0.033027  …  0.030732  0.018096  0.046392
1    0.026671  0.070314  0.059404  … -0.012280  0.019624  0.018991
2    0.016102  0.029521  0.059042  …  0.018999  0.014254  0.038128
3    0.029923  0.042747  0.044172  … -0.002484  0.027664  0.015354
4    0.040824  0.055944  0.071064  …  0.008183  0.003207  0.035150
..        …         …         …  …       …         …         …
995  0.031627  0.023002  0.043128  …  0.021518  0.025650  0.029339
996  0.049087  0.055398  0.086252  …  0.015070  0.027758  0.016148
997  0.014564  0.049154  0.037139  …  0.012293  0.020327  0.018595
998  0.029655  0.021567  0.070991  …  0.020497  0.016435 -0.005107
999  0.038304  0.010534  0.056498  …  0.017858  0.014313  0.017853
```

```
          X35       X36       X37       X38       X39       X40        X41
0    0.042328  0.028314  0.011836  0.029523  0.041018  0.016393  efectores
1    0.018646  0.009855  0.038390  0.047725  0.026817  0.012452  efectores
2    0.019618  0.017378  0.016695  0.038553 -0.014801  0.000816  efectores
3   -0.006010  0.043995  0.032249  0.010928  0.019972  0.027832  efectores
4   -0.001300 -0.003257  0.008593  0.015737  0.033125 -0.008334  efectores
..        …         …         …         …         …         …
995  0.034137  0.015995  0.025230  0.017725  0.026820  0.023312  efectores
996  0.006054  0.003740 -0.000587  0.006967  0.020200  0.014279  efectores
997  0.025743  0.014560  0.020689  0.026136  0.015843  0.030092  efectores
998  0.018624  0.023877  0.008596  0.019654  0.007769  0.017179  efectores
999  0.007510  0.035539  0.023932  0.029819  0.019314  0.029016  efectores

[865 rows x 42 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass efectores fusarium_oxysporum dataset 3, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  865.000000  865.000000  865.000000  865.000000  865.000000  865.000000
mean     0.045854    0.008323    0.032587    0.035462    0.021983    0.036518
std      0.011854    0.005837    0.013452    0.016831    0.010997    0.010559
min      0.014482    0.000000    0.000000    0.001847    0.000000    0.006912
25%      0.037724    0.004102    0.023127    0.022873    0.014149    0.029317
50%      0.045670    0.007179    0.031427    0.033483    0.020612    0.036011
75%      0.053332    0.011468    0.041338    0.046084    0.028130    0.042792
max      0.092500    0.029763    0.081470    0.102121    0.061833    0.074831

                X6          X7          X8          X9  …         X31  \
count  865.000000  865.000000  865.000000  865.000000  …  865.000000
mean     0.014269    0.029919    0.028629    0.051417  …    0.020488
std      0.007740    0.012538    0.014489    0.020703  …    0.014627
min      0.000000    0.001574    0.000000    0.003611  …   -0.038632
25%      0.008563    0.020961    0.017822    0.037139  …    0.011812
50%      0.013336    0.029028    0.026323    0.049910  …    0.021931
75%      0.019239    0.038060    0.037795    0.064984  …    0.029853
max      0.044584    0.074440    0.087432    0.125058  …    0.068113

               X32         X33         X34         X35         X36         X37  \
count  865.000000  865.000000  865.000000  865.000000  865.000000  865.000000
mean     0.019027    0.020436    0.020297    0.019248    0.020152    0.019782
std      0.015149    0.015826    0.015032    0.015296    0.015508    0.015753
min     -0.057649   -0.042489   -0.029830   -0.050692   -0.041412   -0.038121
25%      0.010683    0.011791    0.011388    0.010540    0.010700    0.011505
50%      0.021027    0.021583    0.021573    0.020898    0.020788    0.021251
75%      0.028532    0.030133    0.030073    0.029115    0.029476    0.029637
```

```
max      0.070754    0.077625    0.078221    0.066512    0.085669    0.075290

              X38         X39         X40
count  865.000000  865.000000  865.000000
mean     0.019556    0.019708    0.018911
std      0.015292    0.015446    0.015708
min     -0.040202   -0.041508   -0.046440
25%      0.010800    0.011324    0.010199
50%      0.020644    0.020489    0.020022
75%      0.029001    0.029136    0.029326
max      0.068168    0.070971    0.067894

[8 rows x 41 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass no_efectores fusarium_oxysporum
dataset 3, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.057061  0.000000  0.030209  0.067131  0.011748  0.023496  0.010070
1    0.049288  0.001494  0.017923  0.025391  0.016429  0.035846  0.004481
2    0.041434  0.012749  0.041434  0.054183  0.022311  0.041434  0.019123
3    0.048512  0.006930  0.015593  0.013861  0.060640  0.053710  0.006930
4    0.016138  0.008069  0.036310  0.020172  0.028241  0.036310  0.020172
..        …         …         …         …         …         …
995  0.070773  0.008987  0.048305  0.039318  0.016851  0.031455  0.023591
996  0.052103  0.006652  0.025497  0.035475  0.025497  0.034366  0.012194
997  0.042375  0.009302  0.022738  0.032040  0.015503  0.035140  0.012403
998  0.053748  0.017636  0.034432  0.034432  0.024355  0.039471  0.012597
999  0.048348  0.024552  0.032862  0.069123  0.030973  0.036639  0.042305

           X7        X8        X9  …       X32       X33       X34  \
0    0.053704  0.057061  0.050348  …  0.018930  0.012169 -0.011679
1    0.025391  0.038833  0.050782  …  0.002951  0.024086  0.030914
2    0.060557  0.035059  0.095617  … -0.026331  0.019610  0.036739
3    0.058907  0.031186  0.076233  …  0.022984  0.006493  0.047290
4    0.020172  0.020172  0.040344  …  0.008998  0.011761 -0.002476
..        …         …         …  …         …         …         …
995  0.038195  0.029208  0.068526  … -0.003346  0.005002  0.013303
996  0.032149  0.024389  0.034366  …  0.017948  0.040826  0.026261
997  0.032040  0.025839  0.048577  …  0.034593  0.026675  0.036477
998  0.022675  0.015117  0.059627  …  0.033852  0.016624 -0.003917
999  0.059680  0.054769  0.083854  …  0.002248  0.012326  0.017148

           X35       X36       X37       X38       X39       X40           X41
0    0.020818  0.011369  0.007916  0.018120  0.026597  0.021912  no_efectores
```

```
1    0.055448  0.013663  0.025123  0.036895  0.024546  0.037505  no_efectores
2    0.008252  0.044765  0.022613 -0.024125 -0.021455  0.035510  no_efectores
3   -0.009578  0.011922  0.024690  0.012253  0.021372 -0.002138  no_efectores
4    0.046877  0.045182  0.016788  0.024396  0.009548  0.003539  no_efectores
..        …         …         …         …         …         …
995  0.049223  0.026745  0.015152  0.047508  0.006042 -0.010735  no_efectores
996  0.039064  0.031233  0.008573  0.020084  0.032357  0.023270  no_efectores
997 -0.004198  0.049528  0.012141  0.029064  0.022560  0.049343  no_efectores
998  0.017654  0.014547  0.010209  0.028361  0.013779  0.005731  no_efectores
999  0.014039  0.006326  0.002425  0.010861  0.015671  0.010931  no_efectores

[874 rows x 42 columns]
```

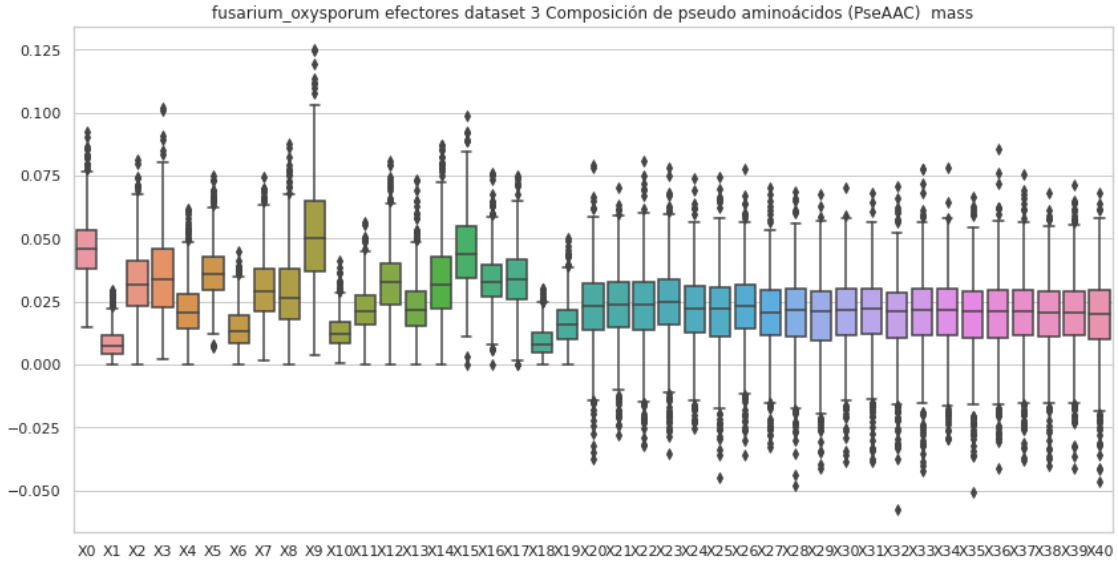
Composición de pseudo aminoácidos (PseAAC) mass no_efectores fusarium_oxysporum dataset 3, sin valores atípicos.
Estadísticas.

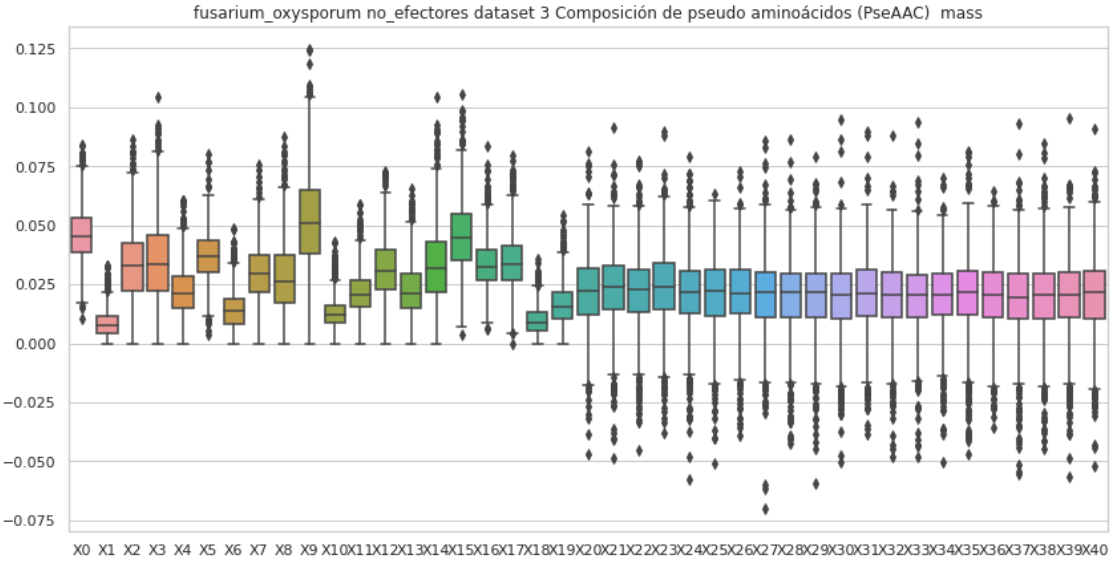
	X0	X1	X2	X3	X4	X5	\
count	874.000000	874.000000	874.000000	874.000000	874.000000	874.000000	
mean	0.045910	0.008439	0.033397	0.035346	0.022168	0.037082	
std	0.011929	0.006004	0.014715	0.018060	0.010716	0.010592	
min	0.010532	0.000000	0.000000	0.000000	0.000000	0.003490	
25%	0.038364	0.004225	0.022271	0.022486	0.014854	0.030299	
50%	0.045052	0.007331	0.032920	0.033595	0.021191	0.036658	
75%	0.053119	0.011395	0.042635	0.046103	0.028504	0.043699	
max	0.084325	0.032952	0.086535	0.104289	0.060640	0.080344	

	X6	X7	X8	X9	…	X31	\
count	874.000000	874.000000	874.000000	874.000000	…	874.000000	
mean	0.014376	0.029924	0.028141	0.051983	…	0.020385	
std	0.008061	0.012085	0.014512	0.020358	…	0.016222	
min	0.000000	0.000000	0.000000	0.000000	…	-0.038270	
25%	0.008268	0.021684	0.017298	0.038201	…	0.011393	
50%	0.013574	0.029265	0.025998	0.050691	…	0.020979	
75%	0.018718	0.037640	0.037153	0.064734	…	0.031364	
max	0.048845	0.075795	0.087536	0.124530	…	0.089861	

	X32	X33	X34	X35	X36	X37	\
count	874.000000	874.000000	874.000000	874.000000	874.000000	874.000000	
mean	0.019749	0.019605	0.019308	0.020665	0.019915	0.018939	
std	0.015497	0.015925	0.015483	0.017351	0.015084	0.016964	
min	-0.048092	-0.048138	-0.050307	-0.046775	-0.035535	-0.055221	
25%	0.011186	0.010883	0.011872	0.011848	0.010760	0.010379	
50%	0.020648	0.020271	0.020609	0.021498	0.020295	0.019190	
75%	0.030084	0.028883	0.029306	0.030862	0.030300	0.029352	
max	0.088287	0.093753	0.070174	0.081126	0.064549	0.092973	

```
              X38         X39         X40
count  874.000000  874.000000  874.000000
mean     0.019317    0.019420    0.020265
std      0.016744    0.016853    0.017154
min     -0.044715   -0.056478   -0.052147
25%      0.010231    0.011190    0.010561
50%      0.020734    0.020706    0.021422
75%      0.029383    0.030228    0.030748
max      0.084499    0.095106    0.090581

[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 3, con valores atípicos.
Valores del documento csv.
```

	X0	X1	X2	X3	X4	X5	X6	\
0	0.099764	0.013302	0.056533	0.063184	0.036580	0.103089	0.029929	
1	0.041901	0.004365	0.031426	0.041901	0.022696	0.027061	0.014840	
2	0.062622	0.003296	0.056031	0.049439	0.016480	0.036255	0.019775	
3	0.138226	0.018030	0.054089	0.042069	0.069113	0.096157	0.039064	
4	0.022932	0.006880	0.024461	0.032869	0.012995	0.021403	0.010702	
..	…	…	…	…	…	…	…	
995	0.094350	0.014742	0.070762	0.064865	0.017691	0.070762	0.029484	
996	0.058157	0.007652	0.048975	0.065810	0.034435	0.045914	0.019131	
997	0.011650	0.000765	0.009779	0.014626	0.001361	0.006548	0.004762	
998	0.049708	0.012671	0.037037	0.046784	0.031189	0.037037	0.017544	
999	0.028698	0.004502	0.009003	0.009566	0.020820	0.023071	0.005064	

	X7	X8	X9	…	X53	X54	X55	\
0	0.039905	0.049882	0.076486	…	0.051846	0.002622	-0.003148	
1	0.019205	0.050630	0.042774	…	0.034389	0.000025	0.015360	
2	0.019775	0.036255	0.072510	…	0.023206	-0.039416	-0.035913	
3	0.063103	0.090148	0.093153	…	0.022751	-0.046300	-0.046001	
4	0.020639	0.028283	0.035927	…	0.020464	0.003115	0.010322	
..	…	…	…	…	…	…	…	
995	0.064865	0.047175	0.088453	…	0.000547	-0.034817	-0.014870	
996	0.053566	0.060453	0.094123	…	0.002931	-0.001052	0.026165	
997	0.003401	0.011479	0.008673	…	0.023830	0.016208	0.025722	
998	0.032164	0.023392	0.076999	…	0.004861	-0.000053	0.001569	
999	0.022509	0.006190	0.033200	…	0.008486	0.008629	0.000074	

	X56	X57	X58	X59	X60	X61	X62
0	0.014780	-0.009853	-0.040073	-0.019595	0.032058	0.020557	efectores
1	0.023629	0.029768	-0.006569	0.019032	0.017797	0.017802	efectores
2	0.022347	0.017189	0.029393	0.039410	-0.003367	0.013100	efectores
3	-0.043830	-0.075439	-0.007893	0.006049	-0.010568	0.011008	efectores
4	-0.011907	0.017905	0.007301	0.011996	0.022072	0.032936	efectores
..	…	…	…	…	…	…	
995	-0.023394	0.034588	0.018256	0.041861	-0.018276	-0.020555	efectores
996	-0.031815	-0.017123	0.018068	0.031470	0.018813	0.030435	efectores
997	0.015570	0.025020	0.014508	0.023726	0.017424	0.027437	efectores

```
998 -0.000166  0.001256 -0.015514 -0.011632  0.022506  0.017264  efectores
999  0.018467  0.007769  0.023734  0.002472  0.015254  0.006371  efectores

[1000 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro efectores fusarium_oxysporum dataset 3, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.060274	0.011302	0.041480	0.043838	0.029382	
std	0.047125	0.011807	0.026029	0.028026	0.036036	
min	-0.146701	0.000000	-0.293402	-0.440104	-0.146701	
25%	0.038673	0.004371	0.026402	0.029220	0.016895	
50%	0.054545	0.008900	0.040063	0.043108	0.025843	
75%	0.075412	0.015053	0.053153	0.056396	0.035749	
max	1.099788	0.176543	0.353086	0.294239	0.684714	

	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.050288	0.018974	0.039317	0.037101	0.067111	…	
std	0.056404	0.015986	0.043495	0.034606	0.070179	…	
min	-0.293402	0.000000	-0.733506	-0.293402	-0.733506	…	
25%	0.029307	0.009949	0.024155	0.022023	0.041243	…	
50%	0.043700	0.016627	0.036625	0.033424	0.063212	…	
75%	0.064207	0.024387	0.048776	0.047901	0.082843	…	
max	1.539704	0.272929	0.659873	0.659873	1.539704	…	

	X52	X53	X54	X55	X56	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.000674	0.010128	-0.000846	0.007743	-0.000627	
std	0.078762	0.041821	0.144064	0.067465	0.088848	
min	-1.571570	-0.471446	-3.899028	-1.241590	-2.499870	
25%	-0.009411	-0.001949	-0.009694	-0.002698	-0.010692	
50%	0.005561	0.010441	0.004164	0.010063	0.004435	
75%	0.016855	0.022551	0.016692	0.021418	0.016726	
max	0.818813	0.845139	1.415994	1.024610	0.445511	

	X57	X58	X59	X60	X61
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.008814	0.001290	0.007342	-0.001211	0.006926
std	0.066284	0.055835	0.039806	0.076001	0.059076
min	-1.338604	-1.209798	-0.668793	-1.791713	-1.120997
25%	-0.001579	-0.010367	-0.003310	-0.010477	-0.002881
50%	0.010621	0.005416	0.009610	0.004928	0.009602
75%	0.022354	0.018370	0.022873	0.016727	0.022100

```
max       1.240346     0.424960     0.266416     0.384922     0.883873

[8 rows x 62 columns]
```

no_efectores
Composición de pseudo aminoácidos (PseAAC) hidro no_efectores
fusarium_oxysporum dataset 3, con valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.023325	0.000000	0.012349	0.027442	0.004802	0.009605	0.004116	
1	0.079516	0.002410	0.028915	0.040963	0.026505	0.057829	0.007229	
2	0.046594	0.014337	0.046594	0.060930	0.025089	0.046594	0.021505	
3	0.032186	0.004598	0.010346	0.009196	0.040233	0.035635	0.004598	
4	0.018710	0.009355	0.042098	0.023388	0.032743	0.042098	0.023388	
..	…	…	…	…	…	…	…	
995	0.089591	0.011377	0.061150	0.049773	0.021331	0.039818	0.029864	
996	0.090652	0.011573	0.044362	0.061721	0.044362	0.059792	0.021216	
997	0.062557	0.013732	0.033567	0.047299	0.022887	0.051876	0.018309	
998	0.061677	0.020238	0.039512	0.039512	0.027947	0.045294	0.014455	
999	0.068271	0.034669	0.046403	0.097606	0.043736	0.051737	0.059737	

	X7	X8	X9	…	X53	X54	X55	\
0	0.021953	0.023325	0.020581	…	0.015884	0.016625	0.023904	
1	0.040963	0.062649	0.081925	…	-0.011192	0.026689	0.031706	
2	0.068099	0.039426	0.107524	…	-0.029331	0.014373	-0.006477	
3	0.039083	0.020691	0.050579	…	-0.002694	-0.012271	-0.020946	
4	0.023388	0.023388	0.046776	…	0.010095	-0.053768	-0.035460	
..	…	…	…	…	…	…	…	
995	0.048351	0.036974	0.086747	…	0.040779	-0.005070	0.032694	
996	0.055934	0.042433	0.059792	…	0.013433	-0.006018	-0.013481	
997	0.047299	0.038144	0.071712	…	0.021597	0.020719	-0.003089	
998	0.026020	0.017347	0.068423	…	0.000304	0.003626	0.003737	
999	0.084272	0.077338	0.118407	…	0.033346	0.018875	0.022934	

	X56	X57	X58	X59	X60	X61	X62
0	0.008862	0.023636	0.006904	0.024881	0.017457	0.032266	no_efectores
1	-0.018886	-0.013208	0.003351	0.017522	0.003811	0.000695	no_efectores
2	0.020286	0.038512	0.005383	0.001142	0.100258	0.036044	no_efectores
3	0.015112	-0.001684	0.027231	0.011529	0.008535	-0.008193	no_efectores
4	0.047636	0.013731	0.032562	0.048854	-0.010619	0.027081	no_efectores
..	…	…	…	…	…	…	…
995	0.015451	0.009657	0.016625	0.013719	-0.006604	0.037432	no_efectores
996	-0.002574	-0.000432	-0.017208	0.025141	-0.015383	-0.013812	no_efectores
997	0.008393	0.011741	0.008578	0.004925	-0.001226	-0.005017	no_efectores
998	0.023961	0.003836	0.007339	0.020432	-0.020017	-0.011250	no_efectores

```
999 -0.015935 -0.023058  0.020224  0.015028  0.004825  0.029103  no_efectores

[1000 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro no_efectores
fusarium_oxysporum dataset 3, con valores atípicos.
Estadísticas.

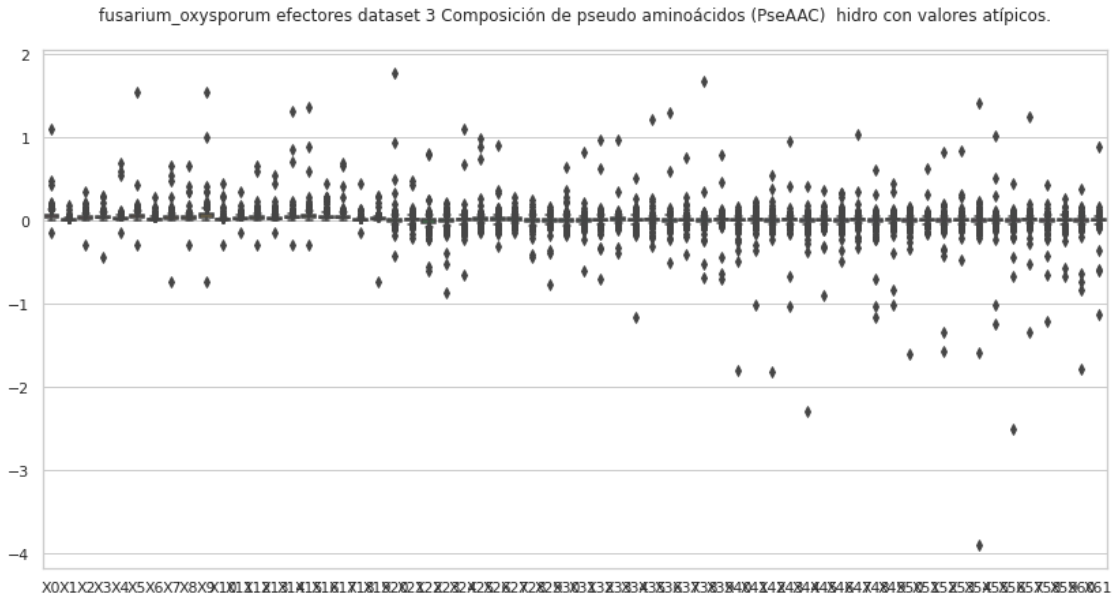
	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.062582	0.012735	0.044062	0.044161	0.030697	
std	0.031304	0.014115	0.027154	0.029634	0.022424	
min	0.000000	0.000000	0.000000	-0.492612	0.000000	
25%	0.042375	0.004670	0.028431	0.028723	0.017871	
50%	0.058926	0.009844	0.043282	0.042712	0.028321	
75%	0.077940	0.015809	0.055838	0.057985	0.038040	
max	0.379123	0.166095	0.498286	0.287800	0.332191	

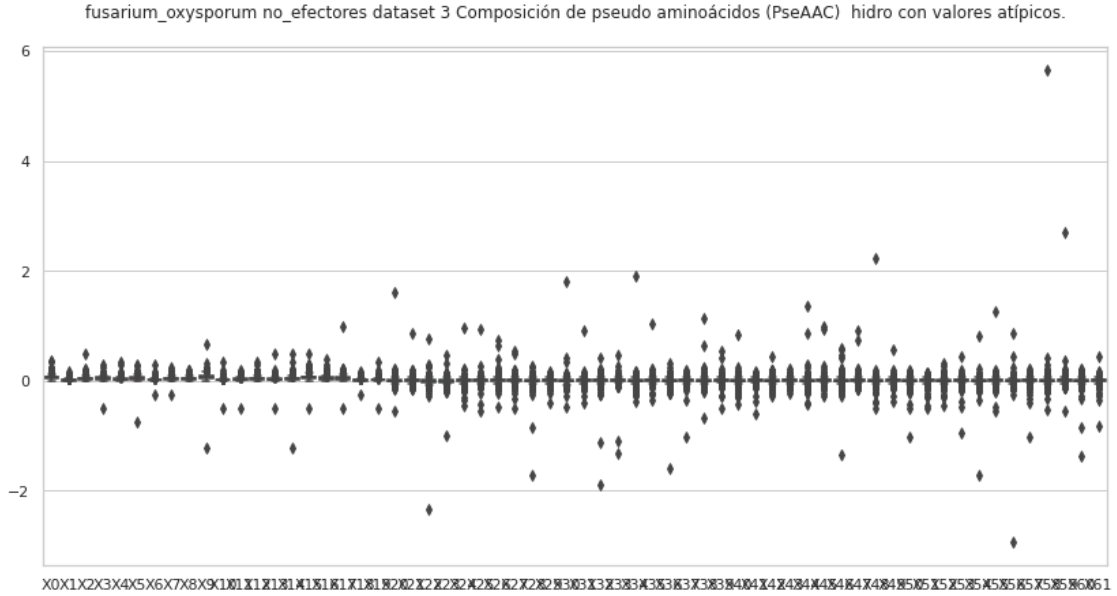
	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.050901	0.019899	0.040446	0.037110	0.069219	…	
std	0.038958	0.018834	0.024934	0.022167	0.057564	…	
min	-0.738919	-0.246306	-0.246306	0.000000	-1.231531	…	
25%	0.031466	0.010177	0.025580	0.021609	0.045221	…	
50%	0.046392	0.017839	0.037981	0.034184	0.067656	…	
75%	0.067079	0.025873	0.051746	0.048517	0.090238	…	
max	0.287800	0.287800	0.246892	0.202199	0.664381	…	

	X52	X53	X54	X55	X56	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.002552	0.006580	0.000550	0.007608	-0.000515	
std	0.040633	0.047904	0.070632	0.055009	0.103510	
min	-0.450685	-0.951955	-1.715161	-0.541923	-2.933474	
25%	-0.008922	-0.003684	-0.010907	-0.005069	-0.010086	
50%	0.005279	0.009120	0.004719	0.008256	0.004358	
75%	0.017229	0.022543	0.017517	0.022187	0.016637	
max	0.321961	0.430624	0.823125	1.250127	0.871512	

	X57	X58	X59	X60	X61
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.005232	0.008262	0.010259	0.001318	0.007071
std	0.047926	0.184283	0.092736	0.063373	0.045354
min	-1.015845	-0.516837	-0.560237	-1.363018	-0.814904
25%	-0.003802	-0.010342	-0.004814	-0.010982	-0.004695
50%	0.009439	0.004481	0.009139	0.004976	0.009130
75%	0.021157	0.018112	0.021939	0.018490	0.022458
max	0.216620	5.650475	2.687595	0.226126	0.445875

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.099764	0.013302	0.056533	0.063184	0.036580	0.103089	0.029929	
1	0.041901	0.004365	0.031426	0.041901	0.022696	0.027061	0.014840	
2	0.062622	0.003296	0.056031	0.049439	0.016480	0.036255	0.019775	
3	0.138226	0.018030	0.054089	0.042069	0.069113	0.096157	0.039064	
4	0.022932	0.006880	0.024461	0.032869	0.012995	0.021403	0.010702	
..	…	…	…	…	…	…	…	
995	0.094350	0.014742	0.070762	0.064865	0.017691	0.070762	0.029484	
996	0.058157	0.007652	0.048975	0.065810	0.034435	0.045914	0.019131	
997	0.011650	0.000765	0.009779	0.014626	0.001361	0.006548	0.004762	
998	0.049708	0.012671	0.037037	0.046784	0.031189	0.037037	0.017544	
999	0.028698	0.004502	0.009003	0.009566	0.020820	0.023071	0.005064	

	X7	X8	X9	…	X53	X54	X55	\
0	0.039905	0.049882	0.076486	…	0.051846	0.002622	-0.003148	
1	0.019205	0.050630	0.042774	…	0.034389	0.000025	0.015360	
2	0.019775	0.036255	0.072510	…	0.023206	-0.039416	-0.035913	
3	0.063103	0.090148	0.093153	…	0.022751	-0.046300	-0.046001	
4	0.020639	0.028283	0.035927	…	0.020464	0.003115	0.010322	
..	…	…	…	…	…	…	…	
995	0.064865	0.047175	0.088453	…	0.000547	-0.034817	-0.014870	
996	0.053566	0.060453	0.094123	…	0.002931	-0.001052	0.026165	
997	0.003401	0.011479	0.008673	…	0.023830	0.016208	0.025722	
998	0.032164	0.023392	0.076999	…	0.004861	-0.000053	0.001569	
999	0.022509	0.006190	0.033200	…	0.008486	0.008629	0.000074	

	X56	X57	X58	X59	X60	X61	X62
0	0.014780	-0.009853	-0.040073	-0.019595	0.032058	0.020557	efectores
1	0.023629	0.029768	-0.006569	0.019032	0.017797	0.017802	efectores
2	0.022347	0.017189	0.029393	0.039410	-0.003367	0.013100	efectores
3	-0.043830	-0.075439	-0.007893	0.006049	-0.010568	0.011008	efectores
4	-0.011907	0.017905	0.007301	0.011996	0.022072	0.032936	efectores
..	…	…	…	…	…	…	
995	-0.023394	0.034588	0.018256	0.041861	-0.018276	-0.020555	efectores
996	-0.031815	-0.017123	0.018068	0.031470	0.018813	0.030435	efectores
997	0.015570	0.025020	0.014508	0.023726	0.017424	0.027437	efectores
998	-0.000166	0.001256	-0.015514	-0.011632	0.022506	0.017264	efectores
999	0.018467	0.007769	0.023734	0.002472	0.015254	0.006371	efectores

```
[938 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC) efectores fusarium_oxysporum dataset 3, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	938.000000	938.000000	938.000000	938.000000	938.000000	938.000000	
mean	0.056738	0.010177	0.039942	0.042645	0.026934	0.046901	
std	0.025549	0.007739	0.018405	0.018775	0.014658	0.024146	
min	0.003820	0.000000	0.000000	0.002022	0.000000	0.002888	
25%	0.037492	0.004374	0.025907	0.029130	0.016740	0.028829	
50%	0.053041	0.008605	0.039382	0.042364	0.025640	0.042067	
75%	0.073362	0.014442	0.051770	0.055228	0.034963	0.061959	
max	0.196516	0.044096	0.100533	0.110165	0.108911	0.134587	

	X6	X7	X8	X9	…	X52	\
count	938.000000	938.000000	938.000000	938.000000	…	938.000000	
mean	0.017673	0.036650	0.034629	0.062933	…	0.004112	
std	0.010592	0.017737	0.017474	0.028518	…	0.024140	
min	0.000000	0.001807	0.000000	0.000000	…	-0.116018	
25%	0.009775	0.023634	0.021473	0.040855	…	-0.007872	
50%	0.016124	0.035630	0.032826	0.062730	…	0.005826	
75%	0.023528	0.047342	0.045853	0.080750	…	0.016625	
max	0.064252	0.110165	0.109147	0.164327	…	0.110733	

	X53	X54	X55	X56	X57	X58	\
count	938.000000	938.000000	938.000000	938.000000	938.000000	938.000000	
mean	0.009528	0.003359	0.008883	0.003280	0.009756	0.004640	
std	0.021780	0.025285	0.021155	0.024051	0.021802	0.022691	
min	-0.089836	-0.177212	-0.104845	-0.099683	-0.094774	-0.083933	
25%	-0.001840	-0.008769	-0.001471	-0.009073	-0.000794	-0.008798	
50%	0.010534	0.004282	0.010171	0.004745	0.010706	0.005797	
75%	0.021852	0.016194	0.020853	0.016525	0.021797	0.018243	
max	0.108204	0.143768	0.096122	0.126143	0.095434	0.084039	

	X59	X60	X61
count	938.000000	938.000000	938.000000
mean	0.009406	0.003365	0.009162
std	0.021317	0.024823	0.022379
min	-0.101665	-0.127330	-0.095484
25%	-0.002265	-0.008893	-0.002312
50%	0.009988	0.005144	0.009824
75%	0.022855	0.016542	0.021927
max	0.108168	0.172565	0.141471

[8 rows x 62 columns]

no_efectores
Composición de pseudo aminoácidos (PseAAC) no_efectores fusarium_oxysporum
dataset 3, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.023325  0.000000  0.012349  0.027442  0.004802  0.009605  0.004116
1    0.079516  0.002410  0.028915  0.040963  0.026505  0.057829  0.007229
2    0.046594  0.014337  0.046594  0.060930  0.025089  0.046594  0.021505
3    0.032186  0.004598  0.010346  0.009196  0.040233  0.035635  0.004598
4    0.018710  0.009355  0.042098  0.023388  0.032743  0.042098  0.023388
..        …         …         …         …         …         …
995  0.089591  0.011377  0.061150  0.049773  0.021331  0.039818  0.029864
996  0.090652  0.011573  0.044362  0.061721  0.044362  0.059792  0.021216
997  0.062557  0.013732  0.033567  0.047299  0.022887  0.051876  0.018309
998  0.061677  0.020238  0.039512  0.039512  0.027947  0.045294  0.014455
999  0.068271  0.034669  0.046403  0.097606  0.043736  0.051737  0.059737

           X7        X8        X9  …       X53       X54       X55  \
0    0.021953  0.023325  0.020581  …  0.015884  0.016625  0.023904
1    0.040963  0.062649  0.081925  … -0.011192  0.026689  0.031706
2    0.068099  0.039426  0.107524  … -0.029331  0.014373 -0.006477
3    0.039083  0.020691  0.050579  … -0.002694 -0.012271 -0.020946
4    0.023388  0.023388  0.046776  …  0.010095 -0.053768 -0.035460
..        …         …         …  …       …         …         …
995  0.048351  0.036974  0.086747  …  0.040779 -0.005070  0.032694
996  0.055934  0.042433  0.059792  …  0.013433 -0.006018 -0.013481
997  0.047299  0.038144  0.071712  …  0.021597  0.020719 -0.003089
998  0.026020  0.017347  0.068423  …  0.000304  0.003626  0.003737
999  0.084272  0.077338  0.118407  …  0.033346  0.018875  0.022934

          X56       X57       X58       X59       X60       X61           X62
0    0.008862  0.023636  0.006904  0.024881  0.017457  0.032266  no_efectores
1   -0.018886 -0.013208  0.003351  0.017522  0.003811  0.000695  no_efectores
2    0.020286  0.038512  0.005383  0.001142  0.100258  0.036044  no_efectores
3    0.015112 -0.001684  0.027231  0.011529  0.008535 -0.008193  no_efectores
4    0.047636  0.013731  0.032562  0.048854 -0.010619  0.027081  no_efectores
..        …         …         …         …         …         …
995  0.015451  0.009657  0.016625  0.013719 -0.006604  0.037432  no_efectores
996 -0.002574 -0.000432 -0.017208  0.025141 -0.015383 -0.013812  no_efectores
997  0.008393  0.011741  0.008578  0.004925 -0.001226 -0.005017  no_efectores
998  0.023961  0.003836  0.007339  0.020432 -0.020017 -0.011250  no_efectores
999 -0.015935 -0.023058  0.020224  0.015028  0.004825  0.029103  no_efectores

[896 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC) no_efectores fusarium_oxysporum
dataset 3, sin valores atípicos.
Estadísticas.

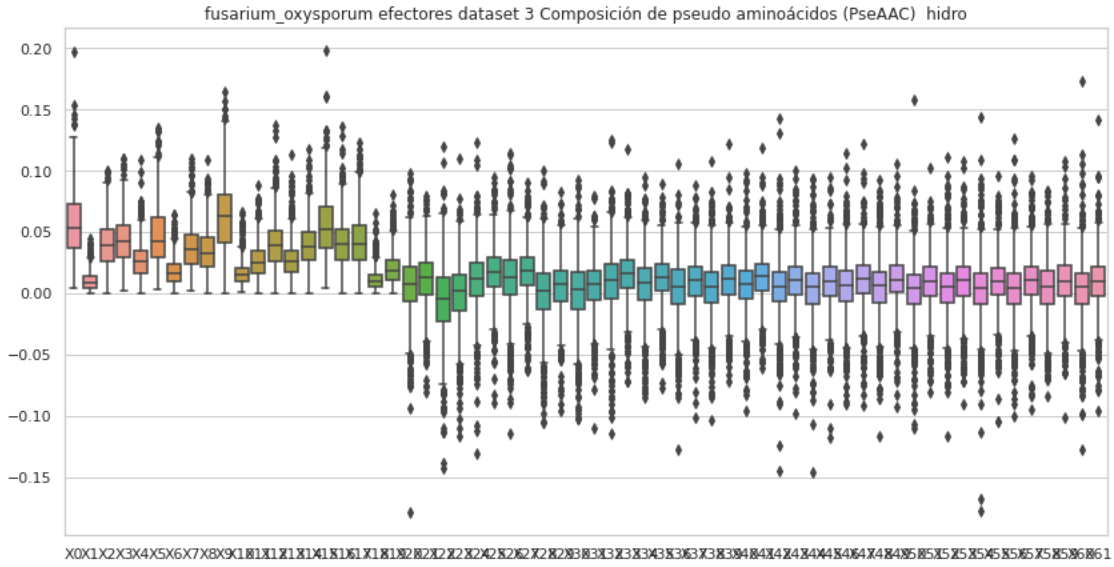
	X0	X1	X2	X3	X4	X5	\
count	896.000000	896.000000	896.000000	896.000000	896.000000	896.000000	
mean	0.058595	0.010838	0.041779	0.042789	0.027753	0.048162	
std	0.025095	0.008446	0.019248	0.019166	0.013929	0.023954	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.040967	0.004694	0.028015	0.029033	0.017564	0.030375	
50%	0.056264	0.009467	0.042622	0.042366	0.027434	0.043768	
75%	0.073125	0.014613	0.054066	0.056423	0.036540	0.063542	
max	0.151408	0.054545	0.108888	0.112806	0.085018	0.143327	

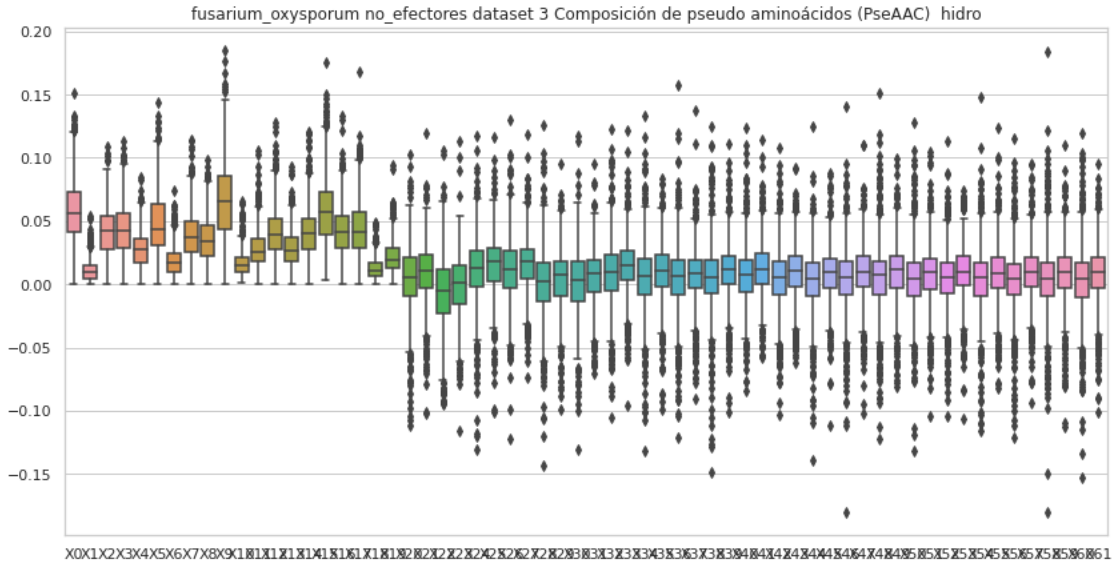
	X6	X7	X8	X9	…	X52	\
count	896.000000	896.000000	896.000000	896.000000	…	896.000000	
mean	0.018451	0.038302	0.034762	0.065657	…	0.004288	
std	0.010938	0.018451	0.016971	0.029325	…	0.022686	
min	0.000000	0.000000	0.000000	0.000000	…	-0.104689	
25%	0.010032	0.025313	0.021901	0.043810	…	-0.007047	
50%	0.017277	0.036620	0.033570	0.065278	…	0.005909	
75%	0.024547	0.049632	0.046266	0.085682	…	0.016712	
max	0.074514	0.114332	0.098591	0.184839	…	0.112720	

	X53	X54	X55	X56	X57	X58	\
count	896.000000	896.000000	896.000000	896.000000	896.000000	896.000000	
mean	0.009240	0.003553	0.008556	0.002831	0.009055	0.003689	
std	0.021253	0.025737	0.022503	0.024913	0.021353	0.027129	
min	-0.106843	-0.115717	-0.081651	-0.121184	-0.093838	-0.180303	
25%	-0.001291	-0.009216	-0.003008	-0.007783	-0.001596	-0.009441	
50%	0.009749	0.005460	0.008830	0.004658	0.010006	0.004580	
75%	0.022143	0.017289	0.021178	0.016281	0.021157	0.017583	
max	0.073663	0.147773	0.124195	0.114890	0.094745	0.183406	

	X59	X60	X61
count	896.000000	896.000000	896.000000
mean	0.008249	0.003406	0.008660
std	0.023846	0.026981	0.022538
min	-0.113131	-0.153209	-0.101082
25%	-0.003433	-0.009926	-0.003210
50%	0.009749	0.004917	0.009417
75%	0.021247	0.017291	0.021812
max	0.110328	0.119417	0.094742

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

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.013952 -0.055806 -0.016258  0.085893 -0.058251 -0.090093  0.057338
1    0.079270  0.056174  0.036470  0.066935 -0.052461 -0.073548 -0.091442
2   -0.005653  0.073230 -0.024098  0.124723 -0.015845  0.041716  0.048514
3    0.063594 -0.023570 -0.003357  0.041148 -0.058540  0.052066 -0.002636
4   -0.026582  0.001799  0.007927  0.023703  0.005890 -0.078885  0.019054
..        …         …         …         …         …         …         …
995 -0.029162 -0.047127  0.042632  0.042425  0.014037 -0.034396  0.039998
996 -0.009585 -0.026947 -0.006991  0.025855  0.008368 -0.056399  0.021435
997  0.071460  0.074796  0.039679  0.051660  0.050697  0.031004  0.001941
998  0.026544 -0.015703  0.018220 -0.025364  0.060470 -0.026365  0.045653
999 -0.053670  0.016357 -0.041410  0.032333 -0.041042  0.109892 -0.025769

           X7        X8        X9       X10       X11       X12        X13
0   -0.057351 -0.120778  0.033428  0.016161 -0.044863  0.004956  efectores
```

```
1    0.009447 -0.062696  0.024572 -0.024681 -0.022205 -0.066736  efectores
2    0.112347  0.078609 -0.014261  0.155434 -0.034611  0.024162  efectores
3   -0.054396  0.007558  0.011947 -0.001724 -0.070259 -0.063530  efectores
4    0.031834 -0.020183 -0.075950 -0.005706  0.073329 -0.002064  efectores
..        …        …        …        …        …        …        …
995  0.038504 -0.015822  0.018644  0.209351 -0.088961 -0.005836  efectores
996 -0.019324 -0.028057 -0.029680 -0.020381  0.011931  0.011891  efectores
997  0.039878  0.036302  0.070155  0.061502  0.015388  0.013143  efectores
998  0.044963  0.015439 -0.037238 -0.011701  0.022741  0.017734  efectores
999  0.048549 -0.059975  0.016429  0.027688  0.057160 -0.023499  efectores

[1000 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro_mass efectores fusarium_oxysporum
dataset 3, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.010348	0.014272	0.012095	0.015219	0.006169	
std	0.056958	0.052961	0.058724	0.055861	0.054444	
min	-0.263027	-0.183567	-0.229413	-0.254152	-0.259674	
25%	-0.024015	-0.015746	-0.019766	-0.016778	-0.027608	
50%	0.011114	0.012355	0.013946	0.015729	0.006843	
75%	0.042358	0.045974	0.043814	0.045682	0.036343	
max	0.285876	0.260422	0.343249	0.209486	0.279217	

	X5	X6	X7	X8	X9	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.003705	0.011595	0.003101	0.005200	0.001172	
std	0.058424	0.057415	0.053266	0.062578	0.060317	
min	-0.294020	-0.254184	-0.261201	-0.282011	-0.450844	
25%	-0.026543	-0.019069	-0.025323	-0.026031	-0.033725	
50%	0.002735	0.009159	0.003093	0.004801	0.002620	
75%	0.038076	0.040139	0.033747	0.037029	0.034925	
max	0.346640	0.422461	0.264442	0.326305	0.457589	

	X10	X11	X12
count	1000.000000	1000.000000	1000.000000
mean	0.005889	0.005761	0.002333
std	0.053645	0.052559	0.059495
min	-0.437327	-0.218862	-0.437060
25%	-0.021162	-0.024041	-0.027850
50%	0.007207	0.006723	0.003775
75%	0.035127	0.036622	0.033518
max	0.210045	0.425213	0.438232

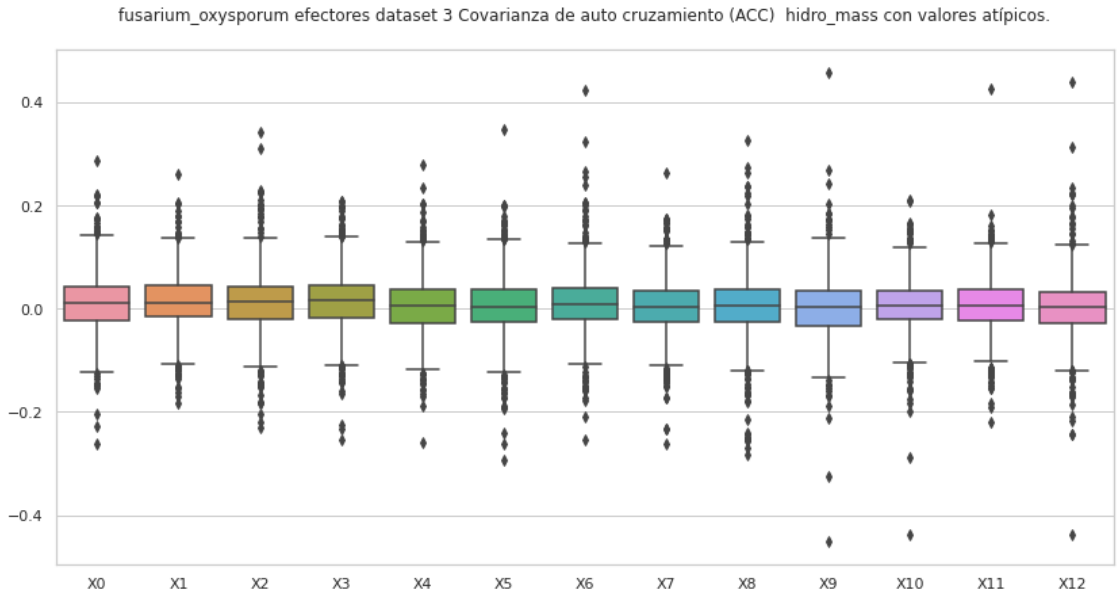
no_efectores
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores fusarium_oxysporum
dataset 3, con valores atípicos.
Valores del documento csv.

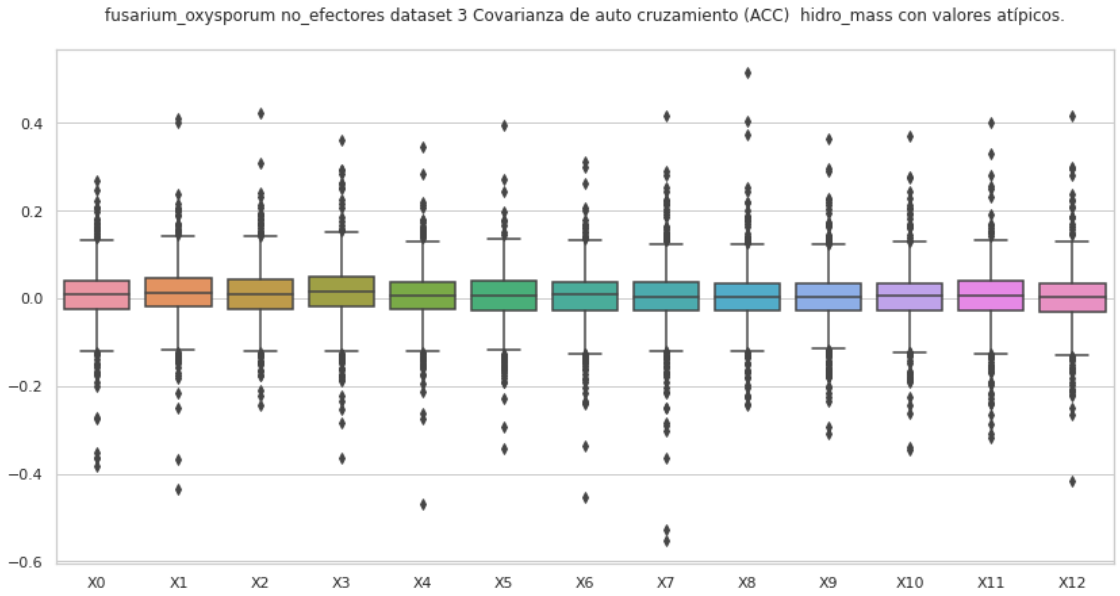
```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.053348  0.141385  0.043101  0.117595  0.024031  0.001686  0.092818
1    0.007214 -0.013362  0.001787 -0.098859  0.071497 -0.060806  0.001910
2    0.027217 -0.038720 -0.045296 -0.008883  0.043747  0.001794 -0.008728
3   -0.039541  0.048749  0.063758 -0.019483  0.117205 -0.110431  0.022551
4   -0.107434 -0.033984  0.010014  0.039722 -0.049609  0.033628  0.010495
..        …         …         …         …         …         …
995 -0.019027 -0.050533  0.030565 -0.028120 -0.057040 -0.029417  0.022302
996  0.019079 -0.007947 -0.118229 -0.042741  0.025989 -0.040300 -0.018607
997 -0.059939 -0.003591  0.024773  0.052936 -0.007596  0.032044  0.030041
998  0.045852 -0.015087 -0.006784  0.013931  0.011105  0.044499 -0.024902
999 -0.008411  0.009243 -0.020681  0.032634  0.024320 -0.021888  0.036567

           X7        X8        X9       X10       X11       X12           X13
0   -0.036484  0.079850 -0.023570  0.044934 -0.057383  0.030276  no_efectores
1   -0.015156  0.012392  0.013532  0.034523 -0.037565 -0.099802  no_efectores
2   -0.028171  0.046146  0.030212  0.039604 -0.070270 -0.096562  no_efectores
3   -0.050111  0.022263  0.088514 -0.045660 -0.039715  0.021147  no_efectores
4    0.143631  0.030380  0.074750 -0.005950 -0.054161 -0.033390  no_efectores
..        …         …         …         …         …         …
995 -0.085616 -0.055882  0.000161  0.013521  0.037701 -0.078678  no_efectores
996 -0.015602  0.008786  0.016317 -0.012117  0.025125 -0.020400  no_efectores
997 -0.022220 -0.019662  0.040208 -0.007746 -0.053893  0.054517  no_efectores
998  0.009030  0.009868 -0.003145  0.022436  0.060962  0.054522  no_efectores
999  0.012826 -0.016231  0.000218  0.025107  0.040599 -0.009592  no_efectores

[1000 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores fusarium_oxysporum
dataset 3, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.006811     0.012647     0.009877     0.014037     0.006733
std       0.063225     0.062775     0.060695     0.065203     0.061202
min      -0.384333    -0.436285    -0.243138    -0.364879    -0.470186
25%      -0.024660    -0.020142    -0.023839    -0.019200    -0.025756
50%       0.007953     0.012419     0.009699     0.014093     0.007102
75%       0.039295     0.045033     0.042206     0.049153     0.038405
```

```
max       0.269161     0.410313     0.423256     0.362413     0.344166

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.005687     0.004595     0.003664     0.004195     0.002411
std       0.060546     0.061549     0.069284     0.062430     0.062212
min      -0.344119    -0.453110    -0.552803    -0.244072    -0.310392
25%      -0.026830    -0.027451    -0.026448    -0.027931    -0.027012
50%       0.006142     0.008176     0.002275     0.003817     0.002038
75%       0.040125     0.038250     0.036749     0.033653     0.033570
max       0.395409     0.310543     0.416296     0.515183     0.363748

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.003310     0.004739     0.002840
std       0.062501     0.064264     0.063630
min      -0.347046    -0.317403    -0.417323
25%      -0.029012    -0.027769    -0.031195
50%       0.004657     0.004899     0.003202
75%       0.034664     0.038851     0.034901
max       0.370957     0.402715     0.418022
```





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 3, sin valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.013952 -0.055806 -0.016258  0.085893 -0.058251 -0.090093  0.057338
1    0.079270  0.056174  0.036470  0.066935 -0.052461 -0.073548 -0.091442
2   -0.005653  0.073230 -0.024098  0.124723 -0.015845  0.041716  0.048514
3    0.063594 -0.023570 -0.003357  0.041148 -0.058540  0.052066 -0.002636
4   -0.026582  0.001799  0.007927  0.023703  0.005890 -0.078885  0.019054
..        …         …         …         …         …         …         …
994  0.045926  0.032206 -0.015503  0.026210 -0.012588 -0.045361 -0.034476
996 -0.009585 -0.026947 -0.006991  0.025855  0.008368 -0.056399  0.021435
997  0.071460  0.074796  0.039679  0.051660  0.050697  0.031004  0.001941
998  0.026544 -0.015703  0.018220 -0.025364  0.060470 -0.026365  0.045653
999 -0.053670  0.016357 -0.041410  0.032333 -0.041042  0.109892 -0.025769

           X7        X8        X9       X10       X11       X12        X13
0   -0.057351 -0.120778  0.033428  0.016161 -0.044863  0.004956  efectores
1    0.009447 -0.062696  0.024572 -0.024681 -0.022205 -0.066736  efectores
2    0.112347  0.078609 -0.014261  0.155434 -0.034611  0.024162  efectores
3   -0.054396  0.007558  0.011947 -0.001724 -0.070259 -0.063530  efectores
4    0.031834 -0.020183 -0.075950 -0.005706  0.073329 -0.002064  efectores
..        …         …         …         …         …         …         …
994 -0.020546  0.028726  0.031524 -0.052547  0.007874  0.005590  efectores
```

```
996 -0.019324 -0.028057 -0.029680 -0.020381  0.011931  0.011891  efectores
997  0.039878  0.036302  0.070155  0.061502  0.015388  0.013143  efectores
998  0.044963  0.015439 -0.037238 -0.011701  0.022741  0.017734  efectores
999  0.048549 -0.059975  0.016429  0.027688  0.057160 -0.023499  efectores

[911 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro_mass efectores fusarium_oxysporum
dataset 3, sin valores atípicos.
Estadísticas.

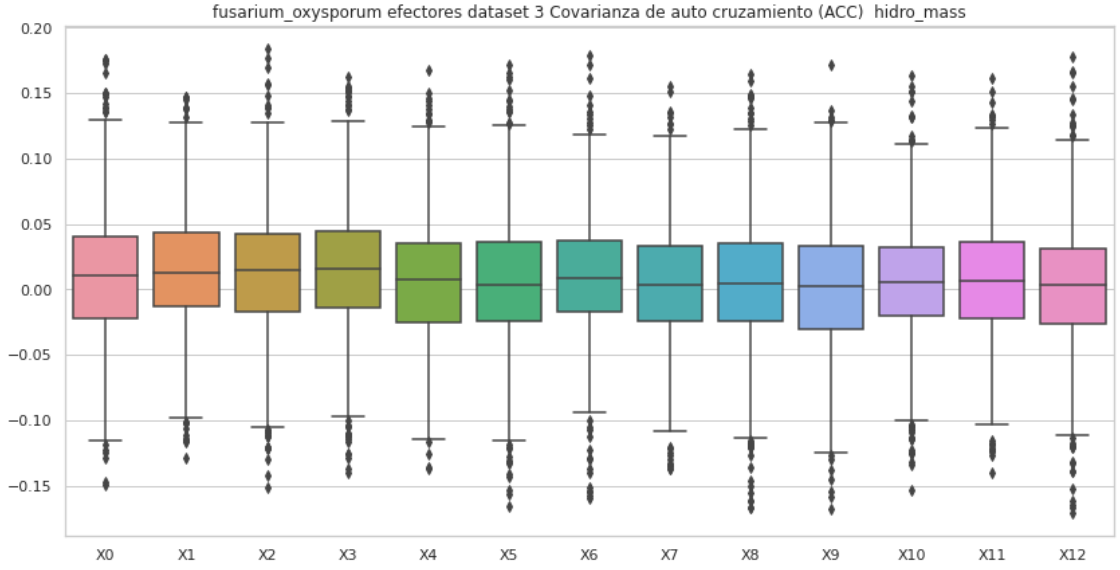
```
                X0          X1          X2          X3          X4          X5  \
count  911.000000  911.000000  911.000000  911.000000  911.000000  911.000000
mean     0.009664    0.014501    0.012312    0.015324    0.006509    0.004131
std      0.049283    0.045882    0.049519    0.048343    0.047094    0.050251
min     -0.149004   -0.129051   -0.151751   -0.140028   -0.137606   -0.165358
25%     -0.021987   -0.013380   -0.017133   -0.013642   -0.025408   -0.024592
50%      0.010504    0.012302    0.014403    0.015731    0.007152    0.002977
75%      0.040654    0.043832    0.042533    0.044159    0.034815    0.035783
max      0.176091    0.146824    0.183725    0.161899    0.167813    0.171128

                X6          X7          X8          X9         X10         X11  \
count  911.000000  911.000000  911.000000  911.000000  911.000000  911.000000
mean     0.009223    0.003790    0.004963    0.000639    0.005666    0.006012
std      0.046863    0.045871    0.050070    0.049753    0.045447    0.045931
min     -0.160145   -0.137238   -0.167186   -0.167845   -0.153076   -0.140681
25%     -0.017379   -0.024102   -0.024675   -0.030140   -0.020718   -0.022597
50%      0.008211    0.003363    0.004849    0.002536    0.005605    0.006750
75%      0.036943    0.032709    0.034784    0.032881    0.032284    0.036068
max      0.178498    0.155451    0.164220    0.171383    0.162904    0.161640

               X12
count  911.000000
mean     0.001927
std      0.048756
min     -0.170955
25%     -0.026385
50%      0.003709
75%      0.031007
max      0.177903
```

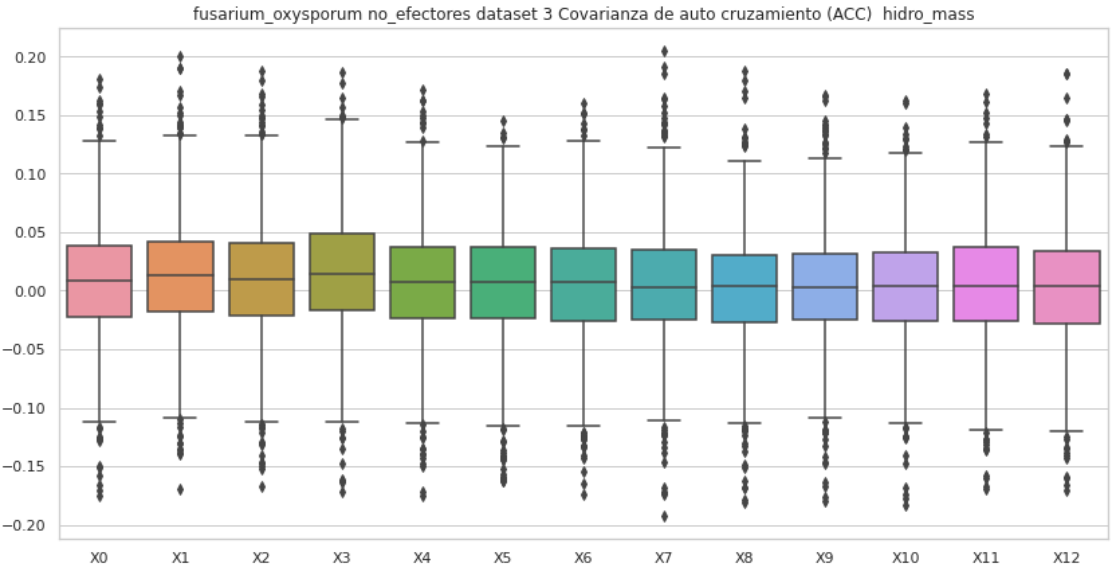
no_efectores
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores fusarium_oxysporum
dataset 3, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.053348  0.141385  0.043101  0.117595  0.024031  0.001686  0.092818
1    0.007214 -0.013362  0.001787 -0.098859  0.071497 -0.060806  0.001910
2    0.027217 -0.038720 -0.045296 -0.008883  0.043747  0.001794 -0.008728
3   -0.039541  0.048749  0.063758 -0.019483  0.117205 -0.110431  0.022551
4   -0.107434 -0.033984  0.010014  0.039722 -0.049609  0.033628  0.010495
..        …         …         …         …         …         …         …
995 -0.019027 -0.050533  0.030565 -0.028120 -0.057040 -0.029417  0.022302
996  0.019079 -0.007947 -0.118229 -0.042741  0.025989 -0.040300 -0.018607
997 -0.059939 -0.003591  0.024773  0.052936 -0.007596  0.032044  0.030041
998  0.045852 -0.015087 -0.006784  0.013931  0.011105  0.044499 -0.024902
999 -0.008411  0.009243 -0.020681  0.032634  0.024320 -0.021888  0.036567

           X7        X8        X9       X10       X11       X12           X13
0   -0.036484  0.079850 -0.023570  0.044934 -0.057383  0.030276  no_efectores
1   -0.015156  0.012392  0.013532  0.034523 -0.037565 -0.099802  no_efectores
2   -0.028171  0.046146  0.030212  0.039604 -0.070270 -0.096562  no_efectores
3   -0.050111  0.022263  0.088514 -0.045660 -0.039715  0.021147  no_efectores
4    0.143631  0.030380  0.074750 -0.005950 -0.054161 -0.033390  no_efectores
..        …         …         …         …         …         …           …
995 -0.085616 -0.055882  0.000161  0.013521  0.037701 -0.078678  no_efectores
996 -0.015602  0.008786  0.016317 -0.012117  0.025125 -0.020400  no_efectores
997 -0.022220 -0.019662  0.040208 -0.007746 -0.053893  0.054517  no_efectores
998  0.009030  0.009868 -0.003145  0.022436  0.060962  0.054522  no_efectores
999  0.012826 -0.016231  0.000218  0.025107  0.040599 -0.009592  no_efectores

[912 rows x 14 columns]


Covarianza de auto cruzamiento (ACC)  hidro_mass no_efectores fusarium_oxysporum
dataset 3, sin valores atípicos.
Estadísticas.

              X0          X1          X2          X3          X4          X5  \
count  912.000000  912.000000  912.000000  912.000000  912.000000  912.000000
mean     0.008236    0.012668    0.009868    0.014778    0.007509    0.006516
std      0.050804    0.050289    0.051210    0.051618    0.050040    0.050994
min     -0.174670   -0.169784   -0.166874   -0.171734   -0.175190   -0.163100
25%     -0.021872   -0.018266   -0.020871   -0.016937   -0.023529   -0.023659
50%      0.008439    0.013036    0.009581    0.014257    0.007610    0.006944
75%      0.038043    0.042392    0.040803    0.048265    0.037132    0.037489
max      0.181215    0.199946    0.187243    0.186208    0.171221    0.145730

              X6          X7          X8          X9         X10         X11  \
count  912.000000  912.000000  912.000000  912.000000  912.000000  912.000000
mean     0.005297    0.005078    0.003005    0.003239    0.003279    0.005764
std      0.048651    0.051464    0.048637    0.049451    0.047680    0.050624
```

min	-0.174214	-0.192474	-0.181119	-0.179458	-0.183367	-0.169892
25%	-0.025596	-0.025154	-0.026896	-0.024270	-0.026337	-0.025656
50%	0.007947	0.002365	0.003861	0.002688	0.004480	0.004389
75%	0.036002	0.035322	0.030727	0.031889	0.032182	0.037104
max	0.159992	0.204944	0.187392	0.167019	0.162611	0.168252

	X12
count	912.000000
mean	0.003435
std	0.049084
min	-0.170560
25%	-0.028005
50%	0.003723
75%	0.033334
max	0.185607





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
3, con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.013952 -0.055806 -0.016258  0.085893 -0.058251 -0.090093  0.057338
1    0.079270  0.056174  0.036470  0.066935 -0.052461 -0.073548 -0.091442
2   -0.005653  0.073230 -0.024098  0.124723 -0.015845  0.041716  0.048514
3    0.063594 -0.023570 -0.003357  0.041148 -0.058540  0.052066 -0.002636
4   -0.026582  0.001799  0.007927  0.023703  0.005890 -0.078885  0.019054
..        …         …         …         …         …         …         …
995 -0.029162 -0.047127  0.042632  0.042425  0.014037 -0.034396  0.039998
996 -0.009585 -0.026947 -0.006991  0.025855  0.008368 -0.056399  0.021435
997  0.071460  0.074796  0.039679  0.051660  0.050697  0.031004  0.001941
998  0.026544 -0.015703  0.018220 -0.025364  0.060470 -0.026365  0.045653
999 -0.053670  0.016357 -0.041410  0.032333 -0.041042  0.109892 -0.025769

           X7        X8        X9       X10       X11       X12         X13
0   -0.057351 -0.120778  0.033428  0.016161 -0.044863  0.004956  efectores
1    0.009447 -0.062696  0.024572 -0.024681 -0.022205 -0.066736  efectores
2    0.112347  0.078609 -0.014261  0.155434 -0.034611  0.024162  efectores
3   -0.054396  0.007558  0.011947 -0.001724 -0.070259 -0.063530  efectores
4    0.031834 -0.020183 -0.075950 -0.005706  0.073329 -0.002064  efectores
..        …         …         …         …         …         …         …
995  0.038504 -0.015822  0.018644  0.209351 -0.088961 -0.005836  efectores
996 -0.019324 -0.028057 -0.029680 -0.020381  0.011931  0.011891  efectores
997  0.039878  0.036302  0.070155  0.061502  0.015388  0.013143  efectores
998  0.044963  0.015439 -0.037238 -0.011701  0.022741  0.017734  efectores
999  0.048549 -0.059975  0.016429  0.027688  0.057160 -0.023499  efectores

[1000 rows x 14 columns]


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

                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.010348     0.014272     0.012095     0.015219     0.006169
```

```
std       0.056958     0.052961     0.058724     0.055861     0.054444
min      -0.263027    -0.183567    -0.229413    -0.254152    -0.259674
25%      -0.024015    -0.015746    -0.019766    -0.016778    -0.027608
50%       0.011114     0.012355     0.013946     0.015729     0.006843
75%       0.042358     0.045974     0.043814     0.045682     0.036343
max       0.285876     0.260422     0.343249     0.209486     0.279217

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.003705     0.011595     0.003101     0.005200     0.001172
std       0.058424     0.057415     0.053266     0.062578     0.060317
min      -0.294020    -0.254184    -0.261201    -0.282011    -0.450844
25%      -0.026543    -0.019069    -0.025323    -0.026031    -0.033725
50%       0.002735     0.009159     0.003093     0.004801     0.002620
75%       0.038076     0.040139     0.033747     0.037029     0.034925
max       0.346640     0.422461     0.264442     0.326305     0.457589

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.005889     0.005761     0.002333
std       0.053645     0.052559     0.059495
min      -0.437327    -0.218862    -0.437060
25%      -0.021162    -0.024041    -0.027850
50%       0.007207     0.006723     0.003775
75%       0.035127     0.036622     0.033518
max       0.210045     0.425213     0.438232



no_efectores
Covarianza de auto cruzamiento (ACC)  mass no_efectores fusarium_oxysporum
dataset 3, con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.053348  0.141385  0.043101  0.117595  0.024031  0.001686  0.092818
1    0.007214 -0.013362  0.001787 -0.098859  0.071497 -0.060806  0.001910
2    0.027217 -0.038720 -0.045296 -0.008883  0.043747  0.001794 -0.008728
3   -0.039541  0.048749  0.063758 -0.019483  0.117205 -0.110431  0.022551
4   -0.107434 -0.033984  0.010014  0.039722 -0.049609  0.033628  0.010495
..        …       …         …         …         …         …         …
995 -0.019027 -0.050533  0.030565 -0.028120 -0.057040 -0.029417  0.022302
996  0.019079 -0.007947 -0.118229 -0.042741  0.025989 -0.040300 -0.018607
997 -0.059939 -0.003591  0.024773  0.052936 -0.007596  0.032044  0.030041
998  0.045852 -0.015087 -0.006784  0.013931  0.011105  0.044499 -0.024902
999 -0.008411  0.009243 -0.020681  0.032634  0.024320 -0.021888  0.036567

           X7        X8        X9       X10       X11       X12           X13
```

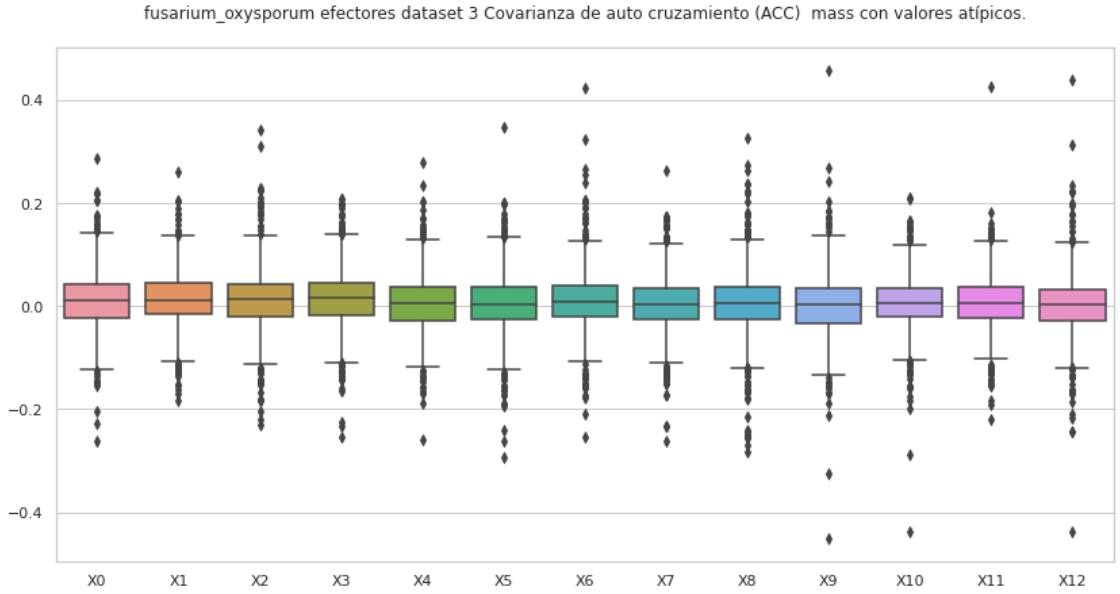
```
0   -0.036484  0.079850 -0.023570  0.044934 -0.057383  0.030276  no_efectores
1   -0.015156  0.012392  0.013532  0.034523 -0.037565 -0.099802  no_efectores
2   -0.028171  0.046146  0.030212  0.039604 -0.070270 -0.096562  no_efectores
3   -0.050111  0.022263  0.088514 -0.045660 -0.039715  0.021147  no_efectores
4    0.143631  0.030380  0.074750 -0.005950 -0.054161 -0.033390  no_efectores
..        …       …       …       …       …       …           …
995 -0.085616 -0.055882  0.000161  0.013521  0.037701 -0.078678  no_efectores
996 -0.015602  0.008786  0.016317 -0.012117  0.025125 -0.020400  no_efectores
997 -0.022220 -0.019662  0.040208 -0.007746 -0.053893  0.054517  no_efectores
998  0.009030  0.009868 -0.003145  0.022436  0.060962  0.054522  no_efectores
999  0.012826 -0.016231  0.000218  0.025107  0.040599 -0.009592  no_efectores

[1000 rows x 14 columns]
```

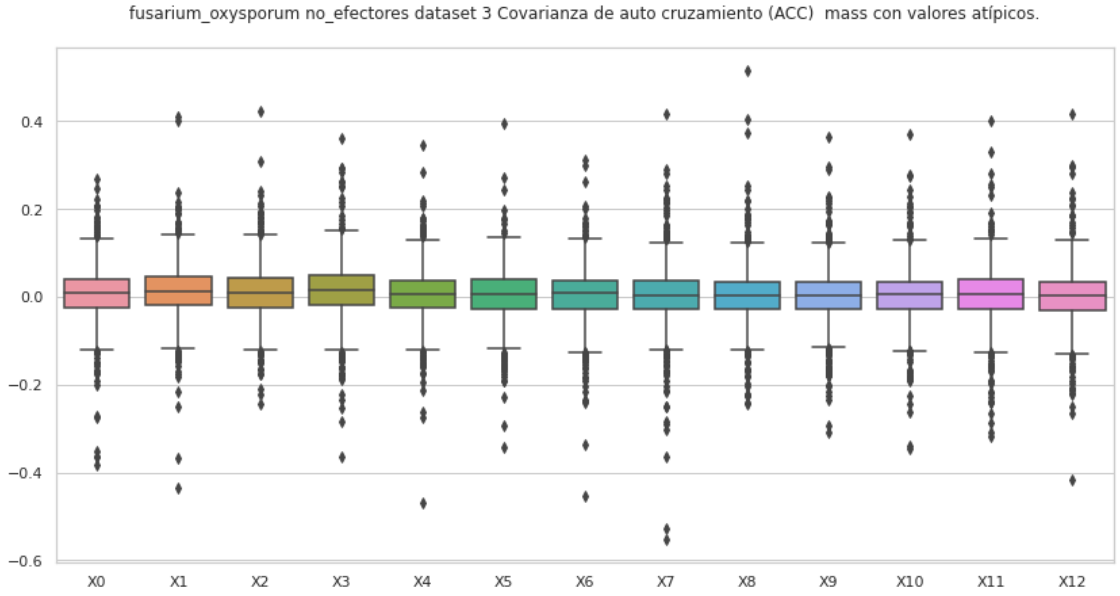
Covarianza de auto cruzamiento (ACC) mass no_efectores fusarium_oxysporum
dataset 3, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.006811	0.012647	0.009877	0.014037	0.006733	
std	0.063225	0.062775	0.060695	0.065203	0.061202	
min	-0.384333	-0.436285	-0.243138	-0.364879	-0.470186	
25%	-0.024660	-0.020142	-0.023839	-0.019200	-0.025756	
50%	0.007953	0.012419	0.009699	0.014093	0.007102	
75%	0.039295	0.045033	0.042206	0.049153	0.038405	
max	0.269161	0.410313	0.423256	0.362413	0.344166	

	X5	X6	X7	X8	X9	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.005687	0.004595	0.003664	0.004195	0.002411	
std	0.060546	0.061549	0.069284	0.062430	0.062212	
min	-0.344119	-0.453110	-0.552803	-0.244072	-0.310392	
25%	-0.026830	-0.027451	-0.026448	-0.027931	-0.027012	
50%	0.006142	0.008176	0.002275	0.003817	0.002038	
75%	0.040125	0.038250	0.036749	0.033653	0.033570	
max	0.395409	0.310543	0.416296	0.515183	0.363748	

	X10	X11	X12
count	1000.000000	1000.000000	1000.000000
mean	0.003310	0.004739	0.002840
std	0.062501	0.064264	0.063630
min	-0.347046	-0.317403	-0.417323
25%	-0.029012	-0.027769	-0.031195
50%	0.004657	0.004899	0.003202
75%	0.034664	0.038851	0.034901
max	0.370957	0.402715	0.418022


fusarium_oxysporum efectores dataset 3 Covarianza de auto cruzamiento (ACC) mass con valores atípicos.
0.4
0.2
0.0
−0.2
−0.4
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12


fusarium_oxysporum no_efectores dataset 3 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))
```

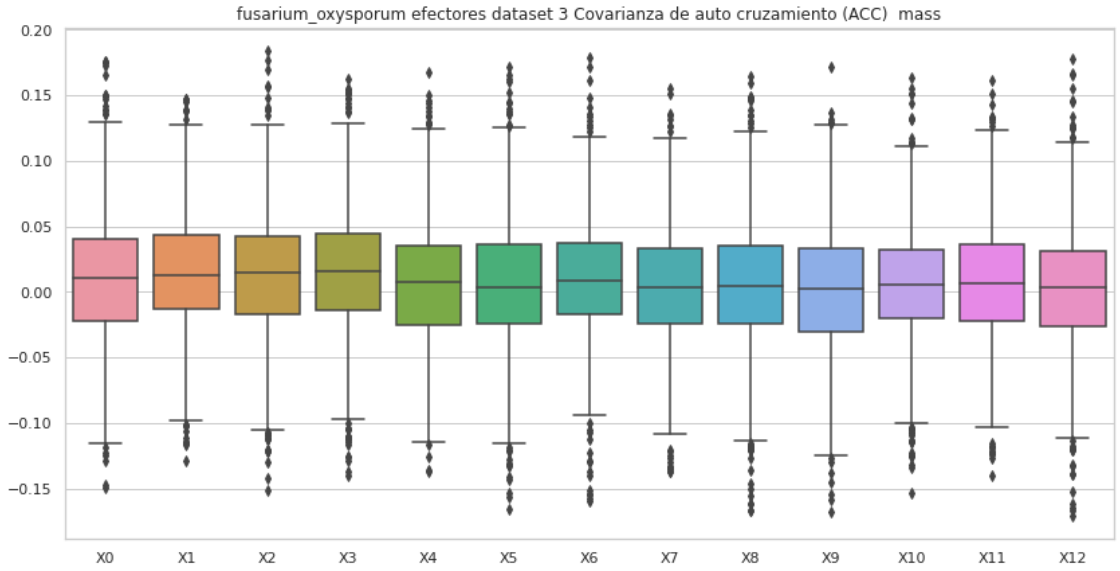
```
Covarianza de auto cruzamiento (ACC)  mass efectores fusarium_oxysporum dataset
3, sin valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.013952 -0.055806 -0.016258  0.085893 -0.058251 -0.090093  0.057338
1    0.079270  0.056174  0.036470  0.066935 -0.052461 -0.073548 -0.091442
2   -0.005653  0.073230 -0.024098  0.124723 -0.015845  0.041716  0.048514
3    0.063594 -0.023570 -0.003357  0.041148 -0.058540  0.052066 -0.002636
4   -0.026582  0.001799  0.007927  0.023703  0.005890 -0.078885  0.019054
..        …         …         …         …         …         …         …
994  0.045926  0.032206 -0.015503  0.026210 -0.012588 -0.045361 -0.034476
996 -0.009585 -0.026947 -0.006991  0.025855  0.008368 -0.056399  0.021435
997  0.071460  0.074796  0.039679  0.051660  0.050697  0.031004  0.001941
998  0.026544 -0.015703  0.018220 -0.025364  0.060470 -0.026365  0.045653
999 -0.053670  0.016357 -0.041410  0.032333 -0.041042  0.109892 -0.025769

           X7        X8        X9       X10       X11       X12         X13
0   -0.057351 -0.120778  0.033428  0.016161 -0.044863  0.004956  efectores
1    0.009447 -0.062696  0.024572 -0.024681 -0.022205 -0.066736  efectores
2    0.112347  0.078609 -0.014261  0.155434 -0.034611  0.024162  efectores
3   -0.054396  0.007558  0.011947 -0.001724 -0.070259 -0.063530  efectores
4    0.031834 -0.020183 -0.075950 -0.005706  0.073329 -0.002064  efectores
..        …         …         …         …         …         …         …
994 -0.020546  0.028726  0.031524 -0.052547  0.007874  0.005590  efectores
996 -0.019324 -0.028057 -0.029680 -0.020381  0.011931  0.011891  efectores
997  0.039878  0.036302  0.070155  0.061502  0.015388  0.013143  efectores
998  0.044963  0.015439 -0.037238 -0.011701  0.022741  0.017734  efectores
999  0.048549 -0.059975  0.016429  0.027688  0.057160 -0.023499  efectores

[911 rows x 14 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  911.000000  911.000000  911.000000  911.000000  911.000000  911.000000
mean     0.009664    0.014501    0.012312    0.015324    0.006509    0.004131
std      0.049283    0.045882    0.049519    0.048343    0.047094    0.050251
min     -0.149004   -0.129051   -0.151751   -0.140028   -0.137606   -0.165358
25%     -0.021987   -0.013380   -0.017133   -0.013642   -0.025408   -0.024592
50%      0.010504    0.012302    0.014403    0.015731    0.007152    0.002977
75%      0.040654    0.043832    0.042533    0.044159    0.034815    0.035783
```

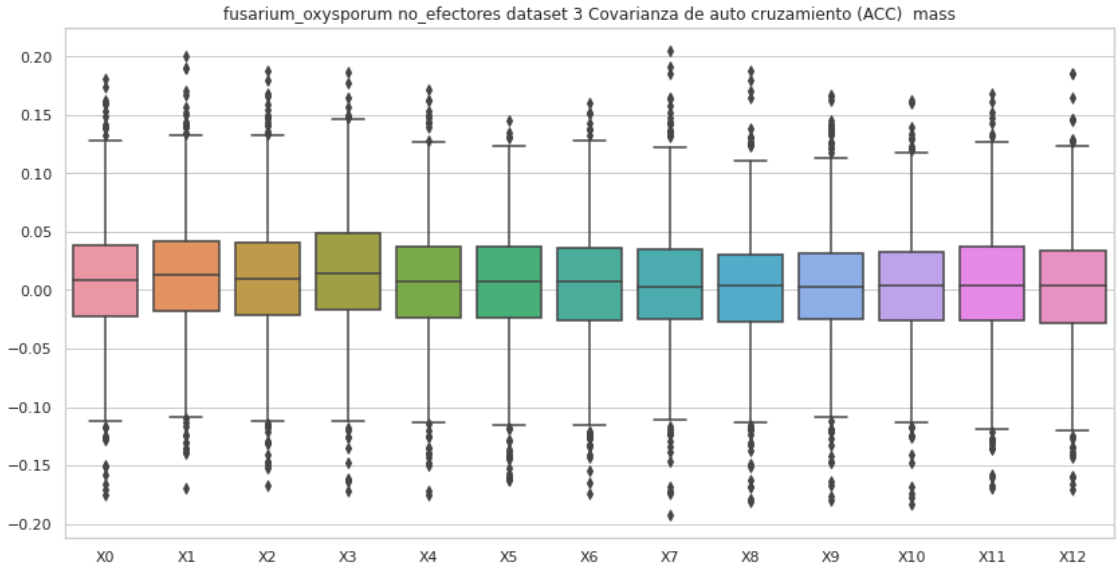
```
max      0.176091    0.146824    0.183725    0.161899    0.167813    0.171128

               X6          X7          X8          X9         X10         X11  \
count  911.000000  911.000000  911.000000  911.000000  911.000000  911.000000
mean     0.009223    0.003790    0.004963    0.000639    0.005666    0.006012
std      0.046863    0.045871    0.050070    0.049753    0.045447    0.045931
min     -0.160145   -0.137238   -0.167186   -0.167845   -0.153076   -0.140681
25%     -0.017379   -0.024102   -0.024675   -0.030140   -0.020718   -0.022597
50%      0.008211    0.003363    0.004849    0.002536    0.005605    0.006750
75%      0.036943    0.032709    0.034784    0.032881    0.032284    0.036068
max      0.178498    0.155451    0.164220    0.171383    0.162904    0.161640

              X12
count  911.000000
mean     0.001927
std      0.048756
min     -0.170955
25%     -0.026385
50%      0.003709
75%      0.031007
max      0.177903
```

Covarianza de auto cruzamiento (ACC) mass no_efectores fusarium_oxysporum
dataset 3, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.053348  0.141385  0.043101  0.117595  0.024031  0.001686  0.092818
1    0.007214 -0.013362  0.001787 -0.098859  0.071497 -0.060806  0.001910
2    0.027217 -0.038720 -0.045296 -0.008883  0.043747  0.001794 -0.008728
3   -0.039541  0.048749  0.063758 -0.019483  0.117205 -0.110431  0.022551
4   -0.107434 -0.033984  0.010014  0.039722 -0.049609  0.033628  0.010495
..        …         …         …         …         …         …         …
995 -0.019027 -0.050533  0.030565 -0.028120 -0.057040 -0.029417  0.022302
996  0.019079 -0.007947 -0.118229 -0.042741  0.025989 -0.040300 -0.018607
997 -0.059939 -0.003591  0.024773  0.052936 -0.007596  0.032044  0.030041
998  0.045852 -0.015087 -0.006784  0.013931  0.011105  0.044499 -0.024902
999 -0.008411  0.009243 -0.020681  0.032634  0.024320 -0.021888  0.036567

           X7        X8        X9       X10       X11       X12           X13
0   -0.036484  0.079850 -0.023570  0.044934 -0.057383  0.030276  no_efectores
1   -0.015156  0.012392  0.013532  0.034523 -0.037565 -0.099802  no_efectores
2   -0.028171  0.046146  0.030212  0.039604 -0.070270 -0.096562  no_efectores
3   -0.050111  0.022263  0.088514 -0.045660 -0.039715  0.021147  no_efectores
4    0.143631  0.030380  0.074750 -0.005950 -0.054161 -0.033390  no_efectores
..        …         …         …         …         …         …           …
```

```
995 -0.085616 -0.055882  0.000161  0.013521  0.037701 -0.078678  no_efectores
996 -0.015602  0.008786  0.016317 -0.012117  0.025125 -0.020400  no_efectores
997 -0.022220 -0.019662  0.040208 -0.007746 -0.053893  0.054517  no_efectores
998  0.009030  0.009868 -0.003145  0.022436  0.060962  0.054522  no_efectores
999  0.012826 -0.016231  0.000218  0.025107  0.040599 -0.009592  no_efectores

[912 rows x 14 columns]


Covarianza de auto cruzamiento (ACC)  mass no_efectores fusarium_oxysporum
dataset 3, sin valores atípicos.
Estadísticas.

                X0          X1          X2          X3          X4          X5  \
count  912.000000  912.000000  912.000000  912.000000  912.000000  912.000000
mean     0.008236    0.012668    0.009868    0.014778    0.007509    0.006516
std      0.050804    0.050289    0.051210    0.051618    0.050040    0.050994
min     -0.174670   -0.169784   -0.166874   -0.171734   -0.175190   -0.163100
25%     -0.021872   -0.018266   -0.020871   -0.016937   -0.023529   -0.023659
50%      0.008439    0.013036    0.009581    0.014257    0.007610    0.006944
75%      0.038043    0.042392    0.040803    0.048265    0.037132    0.037489
max      0.181215    0.199946    0.187243    0.186208    0.171221    0.145730

                X6          X7          X8          X9         X10         X11  \
count  912.000000  912.000000  912.000000  912.000000  912.000000  912.000000
mean     0.005297    0.005078    0.003005    0.003239    0.003279    0.005764
std      0.048651    0.051464    0.048637    0.049451    0.047680    0.050624
min     -0.174214   -0.192474   -0.181119   -0.179458   -0.183367   -0.169892
25%     -0.025596   -0.025154   -0.026896   -0.024270   -0.026337   -0.025656
50%      0.007947    0.002365    0.003861    0.002688    0.004480    0.004389
75%      0.036002    0.035322    0.030727    0.031889    0.032182    0.037104
max      0.159992    0.204944    0.187392    0.167019    0.162611    0.168252

               X12
count  912.000000
mean     0.003435
std      0.049084
min     -0.170560
25%     -0.028005
50%      0.003723
75%      0.033334
max      0.185607
```





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
3, con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.068279 -0.064832 -0.129218 -0.031865 -0.037906  0.045728 -0.019197
1    0.082729 -0.010446  0.088468  0.090486  0.058365  0.033042 -0.062850
2    0.018646 -0.150613  0.077810  0.060207  0.005621  0.044534 -0.020327
3    0.011703 -0.091254  0.049492 -0.068715 -0.033710 -0.027945 -0.009435
4   -0.005527 -0.062334  0.030022  0.045908  0.018507 -0.041892  0.125287
..        …         …         …         …         …         …         …
995 -0.112615  0.031008  0.058008  0.066753  0.050346 -0.103660  0.050554
996  0.001451 -0.045537 -0.000740 -0.009778 -0.059500  0.011729 -0.000996
997  0.089102  0.104744  0.116950  0.114540  0.128826  0.091404  0.097111
998  0.026800  0.014432  0.063959  0.038476 -0.005332  0.020018  0.038805
999  0.089752  0.092060  0.055072  0.067811  0.007192  0.040600  0.113621

           X7        X8        X9       X10       X11       X12        X13
0    0.022370 -0.019135  0.087817  0.008303  0.052908 -0.016795  efectores
1    0.016824  0.041118 -0.045535 -0.010931 -0.029313  0.057552  efectores
2   -0.089690 -0.024951  0.045351 -0.020420 -0.043547  0.063616  efectores
```

```
3   -0.069476 -0.003804 -0.001651 -0.012783  0.013338 -0.023804  efectores
4   -0.062702 -0.004490  0.040809  0.020880  0.085647 -0.045636  efectores
..        …       …        …        …        …        …        …
995  0.056688 -0.024686 -0.148477 -0.000850  0.010237 -0.075101  efectores
996  0.005557 -0.031282  0.009170 -0.055231 -0.022038 -0.022566  efectores
997  0.080449  0.139343  0.122175  0.112651  0.121416  0.115933  efectores
998  0.034999 -0.023856  0.007450 -0.016533  0.045339 -0.020211  efectores
999  0.046654  0.063684  0.024944 -0.013439 -0.043936  0.027612  efectores

[1000 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro efectores fusarium_oxysporum dataset 3, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.011870    -0.021154     0.025988     0.029286    -0.005311
std       0.064655     0.073052     0.066270     0.064381     0.064153
min      -0.267801    -0.324418    -0.363712    -0.282006    -0.464266
25%      -0.030622    -0.061908    -0.012446    -0.007947    -0.041971
50%       0.009006    -0.021919     0.026824     0.030436    -0.005328
75%       0.051785     0.019612     0.065348     0.066169     0.031126
max       0.276397     0.317205     0.284427     0.377371     0.333651

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean     -0.002564     0.022454     0.012774     0.006049     0.010286
std       0.063987     0.065231     0.063198     0.064021     0.060731
min      -0.246508    -0.376221    -0.449786    -0.287035    -0.281771
25%      -0.039272    -0.012225    -0.019818    -0.029003    -0.026673
50%      -0.002635     0.022479     0.013242     0.003382     0.009357
75%       0.035419     0.058960     0.047925     0.041961     0.042973
max       0.249927     0.314419     0.304805     0.378747     0.307820

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.015286     0.004847     0.003702
std       0.057810     0.062605     0.061344
min      -0.341000    -0.318298    -0.349122
25%      -0.017582    -0.028705    -0.028382
50%       0.015212     0.005973     0.003300
75%       0.046886     0.040292     0.040048
max       0.378315     0.358841     0.252956
```

no_efectores
Covarianza de auto cruzamiento (ACC) hidro no_efectores fusarium_oxysporum
dataset 3, con valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.052759	0.042304	0.115265	0.174268	0.077804	0.127750	0.266015	
1	0.037065	-0.037607	0.017315	0.048315	0.012567	-0.003120	0.117895	
2	0.036156	-0.125834	-0.121387	0.007845	0.066295	0.073209	-0.045575	
3	0.040532	0.007921	0.087314	0.048320	0.060518	0.050469	0.027279	
4	-0.125378	-0.104711	0.023996	0.021510	0.015422	0.005466	0.048688	
..	…	…	…	…	…	…	…	
995	-0.050237	-0.095571	0.153159	0.109246	-0.158970	0.036458	0.071922	
996	0.008825	-0.066833	0.034382	0.040247	0.039644	-0.046200	-0.009022	
997	0.031435	-0.090082	0.071759	0.042927	-0.044422	0.042274	0.015140	
998	0.069085	0.012179	-0.035606	0.085599	0.008826	-0.009335	-0.015649	
999	-0.024780	-0.122178	0.033751	0.040638	-0.078313	-0.033271	0.038098	

	X7	X8	X9	X10	X11	X12	X13
0	0.062369	0.000393	0.171993	0.163327	-0.085301	0.035767	no_efectores
1	0.097446	-0.089686	0.003335	0.080299	0.042653	0.038499	no_efectores
2	-0.057456	-0.073176	0.011971	0.047158	-0.099967	-0.069332	no_efectores
3	0.028038	-0.010534	0.005516	0.008964	-0.017945	0.046218	no_efectores
4	0.002201	-0.027232	0.058083	-0.029549	-0.006432	-0.027933	no_efectores
..	…	…	…	…	…	…	…
995	-0.078091	0.013339	-0.034847	0.030664	-0.176916	-0.033166	no_efectores
996	0.054069	-0.054607	-0.096885	0.020414	0.051737	0.025533	no_efectores
997	-0.073715	0.046070	0.039889	0.041622	0.020308	0.020436	no_efectores
998	0.009804	-0.006456	0.051195	0.035110	0.001577	-0.009140	no_efectores
999	-0.000710	-0.051294	0.047889	-0.053591	-0.033193	-0.018163	no_efectores

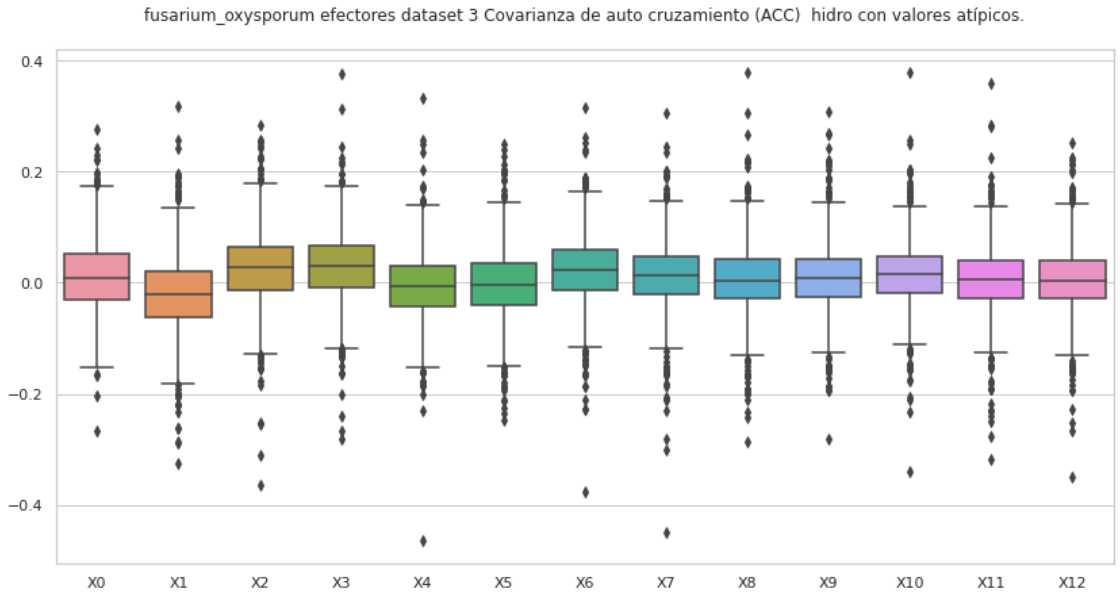
[1000 rows x 14 columns]

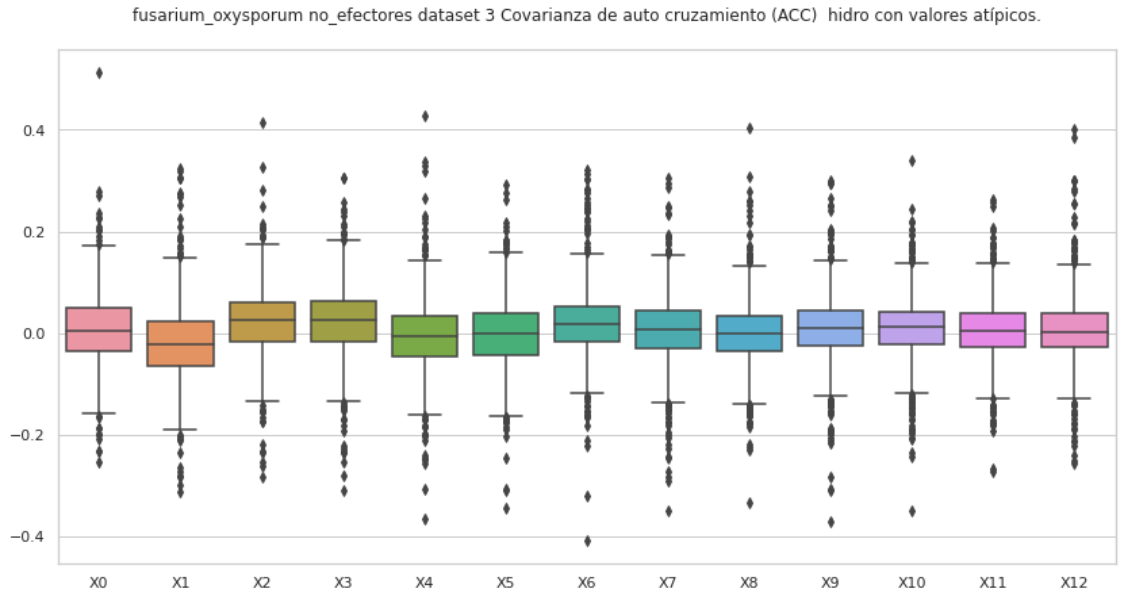
Covarianza de auto cruzamiento (ACC) hidro no_efectores fusarium_oxysporum
dataset 3, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.006895	-0.019209	0.020949	0.022681	-0.004858	
std	0.071683	0.075302	0.066279	0.067924	0.071413	
min	-0.254274	-0.311874	-0.282845	-0.310809	-0.366483	
25%	-0.035116	-0.063866	-0.017754	-0.016984	-0.045406	
50%	0.003715	-0.021349	0.024330	0.025874	-0.005796	
75%	0.048892	0.021827	0.060540	0.062795	0.033471	
max	0.512950	0.323348	0.413321	0.306368	0.427504	

	X5	X6	X7	X8	X9	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	-0.002720	0.019391	0.006752	-0.001414	0.007650	
std	0.066885	0.067311	0.068202	0.066107	0.067261	
min	-0.345160	-0.408776	-0.350757	-0.335065	-0.372198	
25%	-0.042465	-0.017356	-0.029926	-0.035938	-0.025498	
50%	-0.000613	0.017135	0.006188	0.000053	0.010881	
75%	0.038465	0.053052	0.043203	0.033480	0.043068	
max	0.292450	0.320523	0.306300	0.403996	0.300159	

	X10	X11	X12
count	1000.000000	1000.000000	1000.000000
mean	0.010541	0.004339	0.004563
std	0.063667	0.061588	0.066077
min	-0.350509	-0.273977	-0.256156
25%	-0.023547	-0.028578	-0.028685
50%	0.011455	0.004192	0.001966
75%	0.041668	0.038379	0.037605
max	0.340243	0.262514	0.400153





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 3,
sin valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.068279 -0.064832 -0.129218 -0.031865 -0.037906  0.045728 -0.019197
1    0.082729 -0.010446  0.088468  0.090486  0.058365  0.033042 -0.062850
2    0.018646 -0.150613  0.077810  0.060207  0.005621  0.044534 -0.020327
3    0.011703 -0.091254  0.049492 -0.068715 -0.033710 -0.027945 -0.009435
4   -0.005527 -0.062334  0.030022  0.045908  0.018507 -0.041892  0.125287
..        …       …       …       …       …       …       …
995 -0.112615  0.031008  0.058008  0.066753  0.050346 -0.103660  0.050554
996  0.001451 -0.045537 -0.000740 -0.009778 -0.059500  0.011729 -0.000996
997  0.089102  0.104744  0.116950  0.114540  0.128826  0.091404  0.097111
998  0.026800  0.014432  0.063959  0.038476 -0.005332  0.020018  0.038805
999  0.089752  0.092060  0.055072  0.067811  0.007192  0.040600  0.113621

           X7        X8        X9       X10       X11       X12        X13
0    0.022370 -0.019135  0.087817  0.008303  0.052908 -0.016795  efectores
1    0.016824  0.041118 -0.045535 -0.010931 -0.029313  0.057552  efectores
2   -0.089690 -0.024951  0.045351 -0.020420 -0.043547  0.063616  efectores
3   -0.069476 -0.003804 -0.001651 -0.012783  0.013338 -0.023804  efectores
4   -0.062702 -0.004490  0.040809  0.020880  0.085647 -0.045636  efectores
..        …       …       …       …       …       …        …
995  0.056688 -0.024686 -0.148477 -0.000850  0.010237 -0.075101  efectores
```

```
996  0.005557 -0.031282  0.009170 -0.055231 -0.022038 -0.022566  efectores
997  0.080449  0.139343  0.122175  0.112651  0.121416  0.115933  efectores
998  0.034999 -0.023856  0.007450 -0.016533  0.045339 -0.020211  efectores
999  0.046654  0.063684  0.024944 -0.013439 -0.043936  0.027612  efectores

[922 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) efectores fusarium_oxysporum dataset 3, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5
count	922.000000	922.000000	922.000000	922.000000	922.000000	922.000000
mean	0.010691	-0.020735	0.025686	0.027757	-0.003672	-0.002053
std	0.059965	0.063861	0.056585	0.057226	0.056601	0.056842
min	-0.167053	-0.206008	-0.155620	-0.163420	-0.186708	-0.194353
25%	-0.029252	-0.060059	-0.010133	-0.007781	-0.038613	-0.036030
50%	0.007957	-0.021693	0.026824	0.029545	-0.004383	-0.002243
75%	0.049144	0.018099	0.063929	0.064007	0.030935	0.033877
max	0.197987	0.195369	0.222087	0.196153	0.173382	0.187299

	X6	X7	X8	X9	X10	X11
count	922.000000	922.000000	922.000000	922.000000	922.000000	922.000000
mean	0.023865	0.014475	0.005167	0.009632	0.014297	0.006786
std	0.056771	0.051685	0.053644	0.051274	0.048311	0.051950
min	-0.166997	-0.166507	-0.169102	-0.158087	-0.156905	-0.176374
25%	-0.009899	-0.016606	-0.028091	-0.024820	-0.016807	-0.027603
50%	0.023169	0.013616	0.003138	0.009287	0.014452	0.006679
75%	0.058233	0.046218	0.039908	0.042077	0.045121	0.039524
max	0.188921	0.194179	0.169827	0.183113	0.175925	0.190653

	X12
count	922.000000
mean	0.003958
std	0.051572
min	-0.163573
25%	-0.027072
50%	0.002795
75%	0.038333
max	0.171768

no_efectores
Covarianza de auto cruzamiento (ACC) no_efectores fusarium_oxysporum dataset 3, sin valores atípicos.
Valores del documento csv.

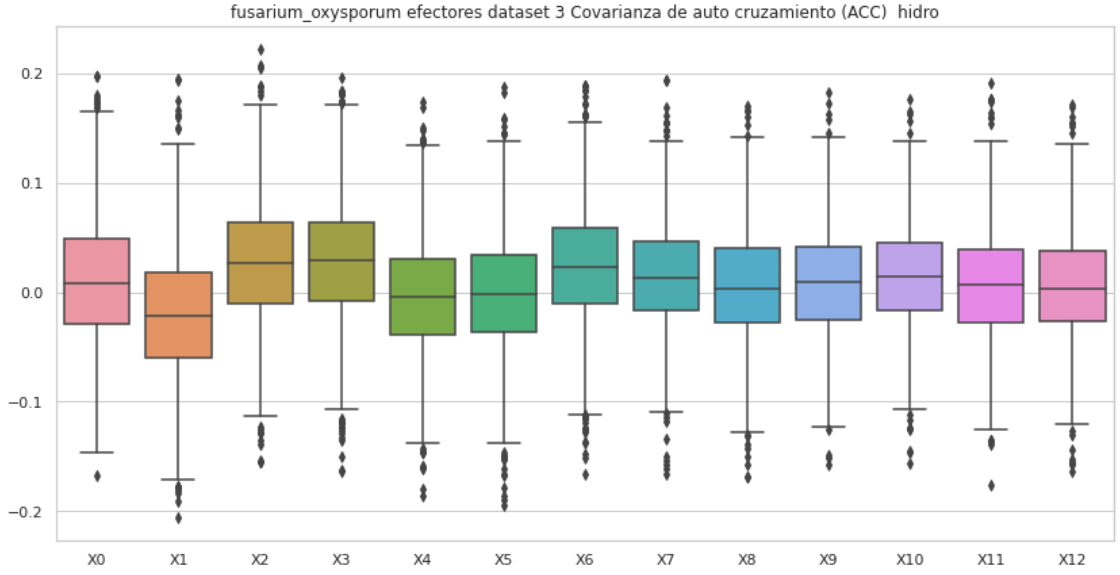
```
           X0        X1        X2        X3        X4        X5        X6  \
1    0.037065 -0.037607  0.017315  0.048315  0.012567 -0.003120  0.117895
2    0.036156 -0.125834 -0.121387  0.007845  0.066295  0.073209 -0.045575
3    0.040532  0.007921  0.087314  0.048320  0.060518  0.050469  0.027279
4   -0.125378 -0.104711  0.023996  0.021510  0.015422  0.005466  0.048688
5    0.004383 -0.016233  0.054917 -0.020247 -0.022097 -0.083243 -0.007400
..        …         …         …         …         …         …         …
995 -0.050237 -0.095571  0.153159  0.109246 -0.158970  0.036458  0.071922
996  0.008825 -0.066833  0.034382  0.040247  0.039644 -0.046200 -0.009022
997  0.031435 -0.090082  0.071759  0.042927 -0.044422  0.042274  0.015140
998  0.069085  0.012179 -0.035606  0.085599  0.008826 -0.009335 -0.015649
999 -0.024780 -0.122178  0.033751  0.040638 -0.078313 -0.033271  0.038098

           X7        X8        X9       X10       X11       X12           X13
1    0.097446 -0.089686  0.003335  0.080299  0.042653  0.038499  no_efectores
2   -0.057456 -0.073176  0.011971  0.047158 -0.099967 -0.069332  no_efectores
3    0.028038 -0.010534  0.005516  0.008964 -0.017945  0.046218  no_efectores
4    0.002201 -0.027232  0.058083 -0.029549 -0.006432 -0.027933  no_efectores
5    0.063429  0.027096 -0.034839 -0.044568  0.041946 -0.074074  no_efectores
..        …         …         …         …         …         …           …
995 -0.078091  0.013339 -0.034847  0.030664 -0.176916 -0.033166  no_efectores
996  0.054069 -0.054607 -0.096885  0.020414  0.051737  0.025533  no_efectores
997 -0.073715  0.046070  0.039889  0.041622  0.020308  0.020436  no_efectores
998  0.009804 -0.006456  0.051195  0.035110  0.001577 -0.009140  no_efectores
999 -0.000710 -0.051294  0.047889 -0.053591 -0.033193 -0.018163  no_efectores

[906 rows x 14 columns]
```

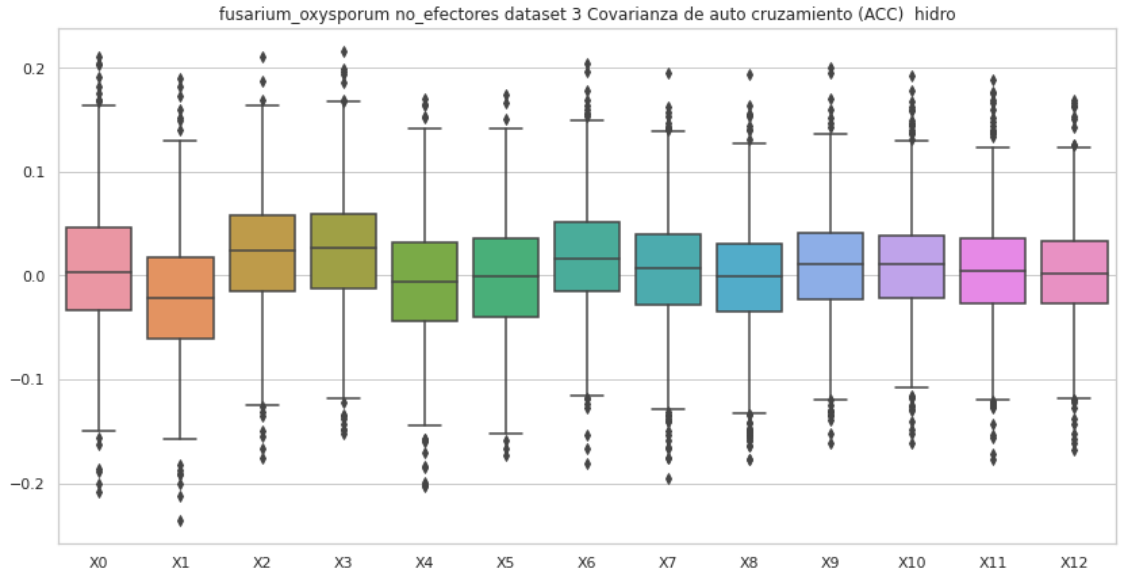
Covarianza de auto cruzamiento (ACC) no_efectores fusarium_oxysporum dataset 3, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  906.000000  906.000000  906.000000  906.000000  906.000000  906.000000
mean     0.006452   -0.021029    0.020924    0.024083   -0.006192   -0.002792
std      0.063096    0.061252    0.055512    0.055911    0.058321    0.055185
min     -0.207723   -0.235758   -0.175308   -0.152037   -0.202721   -0.173040
25%     -0.032843   -0.060971   -0.015211   -0.012620   -0.043465   -0.040020
50%      0.003715   -0.021808    0.024330    0.026442   -0.006232   -0.001060
75%      0.046547    0.017345    0.058203    0.059446    0.031294    0.035500
max      0.210172    0.189875    0.210594    0.215772    0.170079    0.173704

                X6          X7          X8          X9         X10         X11  \
count  906.000000  906.000000  906.000000  906.000000  906.000000  906.000000
mean     0.017641    0.006216   -0.003223    0.008403    0.010835    0.004450
std      0.052280    0.055840    0.053970    0.052503    0.050720    0.052863
```

min	-0.181352	-0.194971	-0.177009	-0.161815	-0.161053	-0.176916
25%	-0.015511	-0.028349	-0.034878	-0.023089	-0.022072	-0.026367
50%	0.016751	0.006815	-0.000969	0.011085	0.011321	0.004192
75%	0.050983	0.039042	0.030425	0.041196	0.038990	0.036370
max	0.203865	0.194580	0.193713	0.200109	0.192424	0.188647

	X12
count	906.000000
mean	0.002279
std	0.050561
min	-0.168456
25%	-0.027309
50%	0.001345
75%	0.033696
max	0.168960




fusarium_oxysporum no_efectores dataset 3 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