# ds5_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 = 5
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 5, con
valores atípicos.
Valores del documento csv.

         X0      X1     X2     X3     X4      X5     X6      X7     X8     X9  \
0     6.306   4.204  4.505  9.910  0.000   8.709  3.604   8.408  1.201  3.904
1    11.386  15.347  4.455  2.475  0.990   6.931  3.960   4.455  3.960  4.950
2     6.250   5.625  3.750  8.438  0.625   7.500  2.188   5.938  1.875  7.187
3     9.174   5.505  4.587  5.505  1.835   4.587  0.917   2.752  2.752  3.670
4    10.490   5.245  5.245  7.692  0.350   4.196  3.147   8.042  1.399  4.196
..      …       …      …      …      …       …      …       …      …      …
995   7.732   5.155  3.093  5.155  1.031   8.247  4.124  10.825  3.608  3.608
996   9.173   6.466  2.556  6.316  2.105   5.714  4.211   5.113  4.511  4.211
997   5.760   5.760  1.613  7.604  2.304  11.751  4.839  12.442  1.843  3.917
998   7.078   7.623  4.719  5.989  1.452   5.808  6.171   4.174  2.722  3.448
999  10.417   5.952  4.762  5.655  0.595   7.738  3.869  10.714  1.488  4.167
```

```
     …    X11    X12    X13     X14     X15    X16    X17    X18    X19  \
0    …  6.607  1.201  5.706   6.607   5.706  4.805  1.802  4.805  4.204
1    …  9.901  2.475  0.990   1.980   5.941  2.475  1.485  1.980  3.960
2    …  5.000  2.500  3.750   4.688   8.438  5.938  2.500  3.750  5.312
3    …  5.505  0.917  6.422   1.835  14.679  4.587  0.917  5.505  7.339
4    …  5.594  1.399  2.797   9.790   8.741  9.441  0.350  2.448  4.196
..   …    …    …    …     …       …     …    …      …      …
995  …  5.670  2.062  4.124   5.155   6.186  3.608  2.577  2.577  6.701
996  …  2.857  2.707  4.060   4.511   8.271  5.263  1.805  2.707  5.714
997  …  8.065  2.765  3.226   5.991   3.687  3.687  0.461  2.765  5.530
998  …  3.267  2.541  5.082   4.174   8.530  5.626  1.633  1.452  6.715
999  …  2.679  1.488  3.571  10.714   5.952  2.976  1.488  2.381  5.357

            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 5, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      8.217572     5.804001     3.792625     5.711564     1.542625
std       2.368095     2.309476     1.415353     1.965850     1.359820
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       6.762750     4.255000     2.893750     4.460000     0.744000
50%       8.069500     5.756000     3.672500     5.779000     1.270000
75%       9.594250     7.045500     4.503500     6.854000     1.980500
max      23.750000    21.951000    12.500000    14.851000    12.698000

                X5           X6           X7           X8          X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.00000
mean      6.067156     4.001818     6.699165     2.448744     5.13884
std       2.412092     1.723836     2.103368     1.335918     1.79066
```

| | | | | | |
|---|---|---|---|---|---|
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.00000 |
| 25% | 4.668750 | 2.947000 | 5.260750 | 1.613000 | 4.00675 |
| 50% | 5.816500 | 3.767500 | 6.626000 | 2.276500 | 5.01250 |
| 75% | 7.190750 | 4.793500 | 8.000000 | 3.073250 | 6.21225 |
| max | 20.339000 | 16.944000 | 13.826000 | 14.545000 | 17.18800 |

| | X10 | X11 | X12 | X13 | X14 \ |
|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 8.756428 | 5.121770 | 2.382582 | 3.855107 | 5.742724 |
| std | 2.661199 | 2.269347 | 1.228740 | 1.687148 | 2.391768 |
| min | 1.282000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 25% | 7.143000 | 3.592250 | 1.642750 | 2.732250 | 4.253000 |
| 50% | 8.897500 | 4.930000 | 2.219500 | 3.728500 | 5.381000 |
| 75% | 10.387000 | 6.224000 | 2.857000 | 4.789250 | 6.798500 |
| max | 23.438000 | 18.644000 | 12.903000 | 13.953000 | 19.298000 |

| | X15 | X16 | X17 | X18 | X19 |
|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 8.086215 | 6.083547 | 1.584883 | 2.852784 | 6.109900 |
| std | 2.593189 | 2.051541 | 1.004145 | 1.301314 | 1.934901 |
| min | 2.500000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 25% | 6.328250 | 4.964000 | 0.870000 | 1.999250 | 4.870500 |
| 50% | 7.683000 | 5.897500 | 1.500500 | 2.746000 | 6.033000 |
| 75% | 9.392000 | 6.898750 | 2.145250 | 3.645750 | 7.279750 |
| max | 24.837000 | 18.953000 | 6.000000 | 10.938000 | 18.605000 |

no_efectores
Composición de aminoácidos (AAC) no_efectores fusarium_oxysporum dataset 5, con valores atípicos.
Valores del documento csv.

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 \ |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 5.765 | 5.322 | 4.656 | 6.208 | 1.774 | 3.326 | 4.435 | 7.761 | 2.217 | 8.204 |
| 1 | 5.856 | 5.405 | 6.306 | 6.081 | 2.703 | 4.955 | 4.054 | 7.207 | 4.505 | 2.703 |
| 2 | 4.895 | 4.429 | 4.895 | 6.993 | 2.098 | 6.993 | 4.196 | 5.361 | 3.263 | 6.294 |
| 3 | 7.466 | 6.835 | 4.627 | 6.099 | 1.052 | 6.519 | 5.152 | 6.519 | 2.313 | 3.891 |
| 4 | 10.087 | 3.304 | 5.043 | 2.087 | 1.739 | 3.478 | 2.609 | 11.478 | 1.391 | 9.391 |
| .. | … | … | … | … | … | … | … | … | … | … |
| 995 | 7.483 | 4.422 | 5.782 | 5.442 | 1.361 | 5.782 | 4.082 | 8.844 | 3.741 | 7.483 |
| 996 | 10.909 | 10.909 | 3.636 | 7.273 | 0.000 | 10.909 | 5.455 | 5.455 | 3.636 | 5.455 |
| 997 | 6.273 | 5.904 | 4.059 | 7.749 | 2.214 | 5.904 | 3.690 | 5.535 | 1.845 | 5.166 |
| 998 | 8.780 | 7.073 | 5.610 | 6.098 | 1.463 | 5.854 | 5.610 | 4.390 | 1.463 | 7.073 |
| 999 | 7.534 | 5.479 | 2.055 | 7.534 | 2.055 | 10.274 | 3.425 | 5.479 | 1.370 | 3.425 |

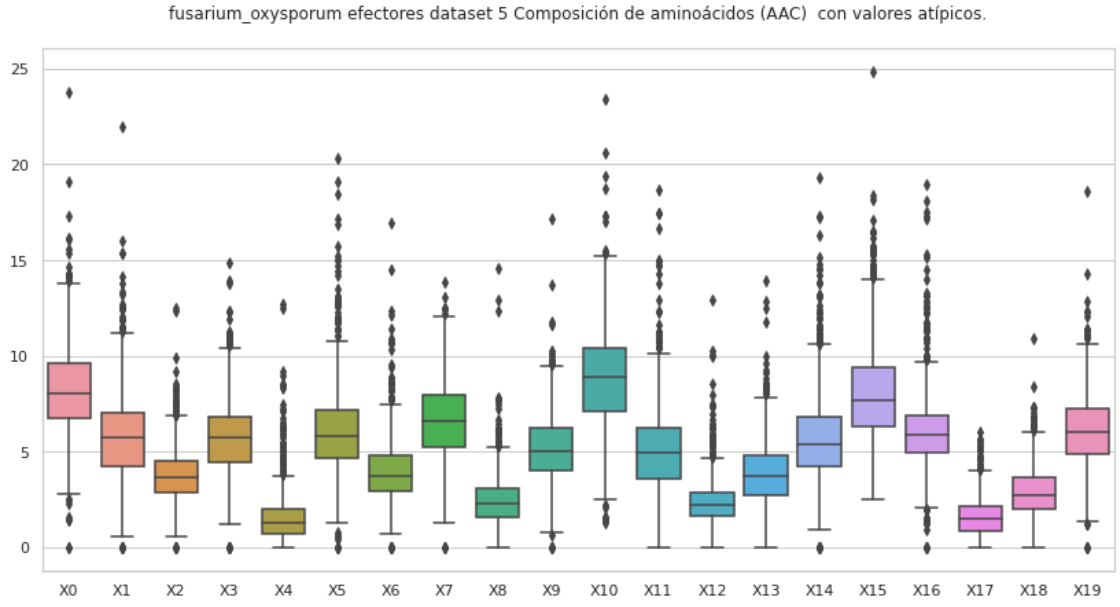
| | … | X11 | X12 | X13 | X14 | X15 | X16 | X17 | X18 | X19 \ |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | … | 4.213 | 0.665 | 3.548 | 4.656 | 9.091 | 6.652 | 0.887 | 3.104 | 8.426 |

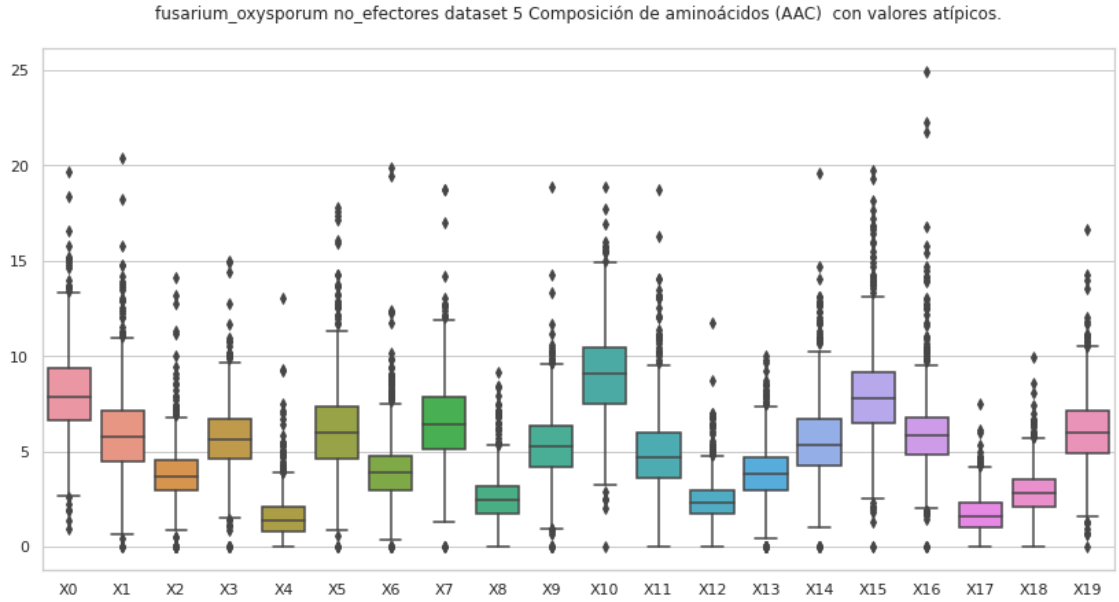
```
1    …  4.279  2.477  2.477  8.784  11.486  5.856  1.351  2.027  3.829
2    …  7.226  2.564  4.429  3.263   6.527  4.662  1.166  2.797  6.061
3    …  3.575  3.260  3.575  7.045   8.728  5.678  1.893  2.629  4.942
4    …  2.957  2.609  6.261  4.522   5.391  4.348  2.087  4.000  7.130
..   …    …    …    …    …     …     …    …    …    …
995  …  7.143  3.061  3.061  4.422   7.143  5.782  0.340  3.061  4.762
996  …  3.636  1.818  5.455  3.636   1.818  1.818  0.000  5.455  5.455
997  …  4.059  0.738  4.428  8.487   7.380  5.904  3.321  3.321  5.904
998  …  5.122  2.683  4.390  2.195   7.561  6.341  0.732  4.390  5.610
999  …  3.425  4.795  6.164  4.795   2.055  6.164  2.740  5.479  4.795

              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 5, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      8.017179     5.904579     3.803822     5.645502     1.570370
std       2.231760     2.257123     1.501280     1.834587     1.212219
min       0.926000     0.000000     0.000000     0.000000     0.000000
25%       6.663750     4.479500     2.949250     4.595250     0.825750
50%       7.863500     5.778000     3.704000     5.644000     1.370000
75%       9.333500     7.096750     4.520250     6.704250     2.063000
max      19.685000    20.424000    14.085000    14.963000    13.043000

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      6.090490     4.040003     6.605010     2.540427     5.329070
std       2.364177     1.817553     2.177473     1.261240     1.898677
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       4.642500     2.955750     5.097500     1.727000     4.201500
50%       5.958000     3.911500     6.385500     2.459000     5.247000
```

```
75%       7.314750     4.786250     7.875750     3.175750     6.367500
max      17.791000    19.876000    18.750000     9.130000    18.868000

                X10          X11          X12          X13          X14  \
count   1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean       9.011176     4.922288     2.418390     3.885183     5.628259
std        2.375725     2.164022     1.101053     1.546438     2.106382
min        0.000000     0.000000     0.000000     0.000000     0.000000
25%        7.469000     3.574000     1.738750     2.919250     4.282500
50%        9.082500     4.706000     2.297500     3.810000     5.352500
75%       10.449000     6.002500     2.936500     4.710000     6.667000
max       18.868000    18.750000    11.765000    10.037000    19.621000

                X15          X16          X17          X18          X19
count   1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean       7.979213     5.967215     1.678137     2.878164     6.085633
std        2.439519     2.153065     1.014487     1.261912     1.836184
min        0.000000     0.000000     0.000000     0.000000     0.000000
25%        6.452000     4.837500     0.989500     2.089250     4.896750
50%        7.792000     5.800500     1.603500     2.787500     5.980000
75%        9.176750     6.740500     2.274000     3.544750     7.143000
max       19.714000    24.922000     7.500000     9.917000    16.667000
```






## 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 5, sin
valores atípicos.
Valores del documento csv.

         X0     X1     X2     X3     X4      X5     X6      X7     X8     X9  \
0     6.306  4.204  4.505  9.910  0.000   8.709  3.604   8.408  1.201  3.904
2     6.250  5.625  3.750  8.438  0.625   7.500  2.188   5.938  1.875  7.187
3     9.174  5.505  4.587  5.505  1.835   4.587  0.917   2.752  2.752  3.670
4    10.490  5.245  5.245  7.692  0.350   4.196  3.147   8.042  1.399  4.196
5     7.604  3.443  3.730  4.017  1.578   7.461  2.582   8.034  2.869  5.739
..      …      …      …      …      …       …      …       …      …      …
995   7.732  5.155  3.093  5.155  1.031   8.247  4.124  10.825  3.608  3.608
996   9.173  6.466  2.556  6.316  2.105   5.714  4.211   5.113  4.511  4.211
997   5.760  5.760  1.613  7.604  2.304  11.751  4.839  12.442  1.843  3.917
998   7.078  7.623  4.719  5.989  1.452   5.808  6.171   4.174  2.722  3.448
999  10.417  5.952  4.762  5.655  0.595   7.738  3.869  10.714  1.488  4.167

     …    X11    X12    X13     X14     X15    X16    X17    X18    X19  \
0    …  6.607  1.201  5.706   6.607   5.706  4.805  1.802  4.805  4.204
2    …  5.000  2.500  3.750   4.688   8.438  5.938  2.500  3.750  5.312
3    …  5.505  0.917  6.422   1.835  14.679  4.587  0.917  5.505  7.339
4    …  5.594  1.399  2.797   9.790   8.741  9.441  0.350  2.448  4.196
5    …  7.461  2.582  3.013   5.452   6.169  7.317  1.435  3.013  7.174
..   …    …      …      …       …       …      …      …      …      …
995  …  5.670  2.062  4.124   5.155   6.186  3.608  2.577  2.577  6.701
```

```
996  …  2.857  2.707  4.060   4.511   8.271  5.263  1.805  2.707  5.714
997  …  8.065  2.765  3.226   5.991   3.687  3.687  0.461  2.765  5.530
998  …  3.267  2.541  5.082   4.174   8.530  5.626  1.633  1.452  6.715
999  …  2.679  1.488  3.571  10.714   5.952  2.976  1.488  2.381  5.357

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

[852 rows x 21 columns]
```

Composición de aminoácidos (AAC) efectores fusarium_oxysporum dataset 5, sin valores atípicos.
Estadísticas.

| | X0 | X1 | X2 | X3 | X4 | X5 | \ |
|---|---|---|---|---|---|---|---|
| count | 852.000000 | 852.000000 | 852.000000 | 852.000000 | 852.000000 | 852.000000 | |
| mean | 8.222141 | 5.791381 | 3.809457 | 5.772310 | 1.431263 | 5.992190 | |
| std | 2.018161 | 2.019648 | 1.217779 | 1.727579 | 0.933866 | 1.932743 | |
| min | 2.273000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 1.282000 | |
| 25% | 6.854000 | 4.348000 | 2.970000 | 4.689250 | 0.799250 | 4.858000 | |
| 50% | 8.154000 | 5.755000 | 3.715500 | 5.830500 | 1.279500 | 5.821000 | |
| 75% | 9.480500 | 6.973000 | 4.500250 | 6.831000 | 1.890750 | 7.076500 | |
| max | 14.170000 | 12.713000 | 8.000000 | 11.231000 | 5.479000 | 12.963000 | |

| | X6 | X7 | X8 | X9 | X10 | X11 | \ |
|---|---|---|---|---|---|---|---|
| count | 852.000000 | 852.000000 | 852.000000 | 852.000000 | 852.000000 | 852.000000 | |
| mean | 3.930930 | 6.756046 | 2.420330 | 5.258925 | 8.988934 | 5.048285 | |
| std | 1.399348 | 1.942071 | 1.018361 | 1.604905 | 2.248799 | 1.817244 | |
| min | 0.000000 | 1.370000 | 0.000000 | 1.250000 | 2.692000 | 0.000000 | |
| 25% | 3.012000 | 5.368750 | 1.687000 | 4.150750 | 7.611250 | 3.729750 | |
| 50% | 3.773500 | 6.713500 | 2.316000 | 5.115500 | 9.087000 | 4.932500 | |
| 75% | 4.746000 | 8.004250 | 3.055750 | 6.284750 | 10.523250 | 6.140750 | |
| max | 8.929000 | 12.442000 | 6.188000 | 10.256000 | 15.487000 | 11.604000 | |

| | X12 | X13 | X14 | X15 | X16 | X17 | \ |
|---|---|---|---|---|---|---|---|
| count | 852.000000 | 852.00000 | 852.000000 | 852.000000 | 852.000000 | 852.000000 | |
| mean | 2.298306 | 3.92415 | 5.655360 | 8.004553 | 6.037234 | 1.609677 | |

| | | | | | | |
|---|---|---|---|---|---|---|
| std | 0.918201 | 1.46725 | 1.948716 | 2.360510 | 1.507679 | 0.889366 |
| min | 0.177000 | 0.00000 | 0.000000 | 2.500000 | 1.220000 | 0.000000 |
| 25% | 1.676000 | 2.97350 | 4.321750 | 6.421250 | 5.076750 | 0.930750 |
| 50% | 2.222000 | 3.81900 | 5.381000 | 7.616500 | 5.929500 | 1.534500 |
| 75% | 2.768000 | 4.81550 | 6.667000 | 9.259750 | 6.829000 | 2.160000 |
| max | 5.952000 | 8.80500 | 12.621000 | 15.716000 | 12.222000 | 4.332000 |

| | X18 | X19 |
|---|---|---|
| count | 852.000000 | 852.000000 |
| mean | 2.876988 | 6.171602 |
| std | 1.162784 | 1.682824 |
| min | 0.000000 | 1.205000 |
| 25% | 2.072500 | 5.000000 |
| 50% | 2.778000 | 6.088500 |
| 75% | 3.638000 | 7.279750 |
| max | 6.599000 | 11.351000 |

no_efectores
Composición de aminoácidos (AAC) no_efectores fusarium_oxysporum dataset 5, sin valores atípicos.
Valores del documento csv.

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 | \ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 5.765 | 5.322 | 4.656 | 6.208 | 1.774 | 3.326 | 4.435 | 7.761 | 2.217 | 8.204 | |
| 1 | 5.856 | 5.405 | 6.306 | 6.081 | 2.703 | 4.955 | 4.054 | 7.207 | 4.505 | 2.703 | |
| 2 | 4.895 | 4.429 | 4.895 | 6.993 | 2.098 | 6.993 | 4.196 | 5.361 | 3.263 | 6.294 | |
| 3 | 7.466 | 6.835 | 4.627 | 6.099 | 1.052 | 6.519 | 5.152 | 6.519 | 2.313 | 3.891 | |
| 4 | 10.087 | 3.304 | 5.043 | 2.087 | 1.739 | 3.478 | 2.609 | 11.478 | 1.391 | 9.391 | |
| .. | … | … | … | … | … | … | … | … | … | … | |
| 995 | 7.483 | 4.422 | 5.782 | 5.442 | 1.361 | 5.782 | 4.082 | 8.844 | 3.741 | 7.483 | |
| 996 | 10.909 | 10.909 | 3.636 | 7.273 | 0.000 | 10.909 | 5.455 | 5.455 | 3.636 | 5.455 | |
| 997 | 6.273 | 5.904 | 4.059 | 7.749 | 2.214 | 5.904 | 3.690 | 5.535 | 1.845 | 5.166 | |
| 998 | 8.780 | 7.073 | 5.610 | 6.098 | 1.463 | 5.854 | 5.610 | 4.390 | 1.463 | 7.073 | |
| 999 | 7.534 | 5.479 | 2.055 | 7.534 | 2.055 | 10.274 | 3.425 | 5.479 | 1.370 | 3.425 | |

| | … | X11 | X12 | X13 | X14 | X15 | X16 | X17 | X18 | X19 | \ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | … | 4.213 | 0.665 | 3.548 | 4.656 | 9.091 | 6.652 | 0.887 | 3.104 | 8.426 | |
| 1 | … | 4.279 | 2.477 | 2.477 | 8.784 | 11.486 | 5.856 | 1.351 | 2.027 | 3.829 | |
| 2 | … | 7.226 | 2.564 | 4.429 | 3.263 | 6.527 | 4.662 | 1.166 | 2.797 | 6.061 | |
| 3 | … | 3.575 | 3.260 | 3.575 | 7.045 | 8.728 | 5.678 | 1.893 | 2.629 | 4.942 | |
| 4 | … | 2.957 | 2.609 | 6.261 | 4.522 | 5.391 | 4.348 | 2.087 | 4.000 | 7.130 | |
| .. | … | … | … | … | … | … | … | … | … | | |
| 995 | … | 7.143 | 3.061 | 3.061 | 4.422 | 7.143 | 5.782 | 0.340 | 3.061 | 4.762 | |
| 996 | … | 3.636 | 1.818 | 5.455 | 3.636 | 1.818 | 1.818 | 0.000 | 5.455 | 5.455 | |
| 997 | … | 4.059 | 0.738 | 4.428 | 8.487 | 7.380 | 5.904 | 3.321 | 3.321 | 5.904 | |
| 998 | … | 5.122 | 2.683 | 4.390 | 2.195 | 7.561 | 6.341 | 0.732 | 4.390 | 5.610 | |

```
999  …  3.425  4.795  6.164  4.795   2.055  6.164  2.740  5.479  4.795

              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

[850 rows x 21 columns]
```

Composición de aminoácidos (AAC) no_efectores fusarium_oxysporum dataset 5, sin valores atípicos.
Estadísticas.

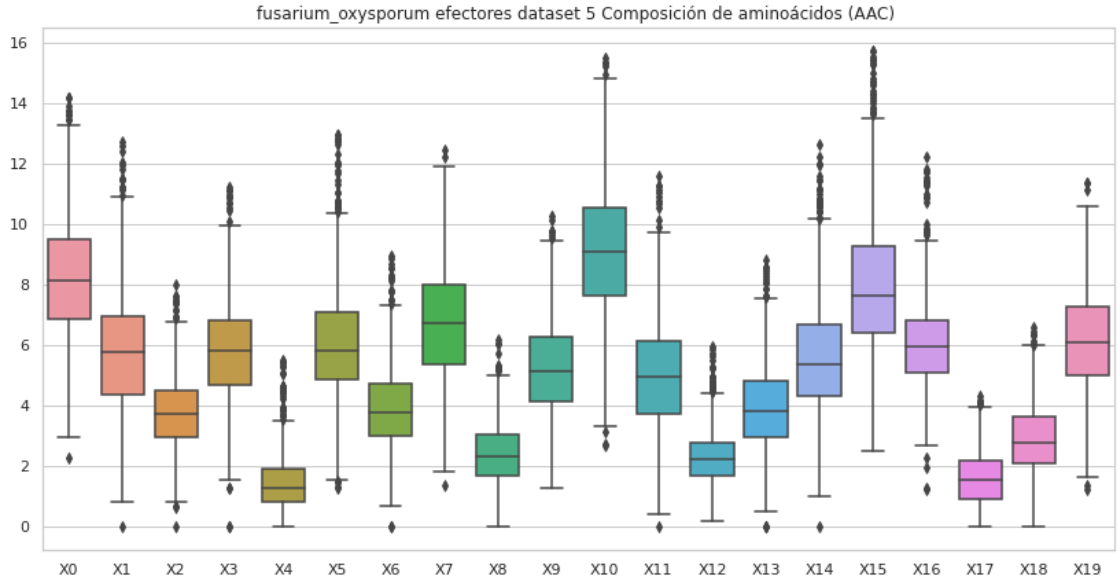
| | X0 | X1 | X2 | X3 | X4 | X5 | \ |
|---|---|---|---|---|---|---|---|
| count | 850.000000 | 850.000000 | 850.000000 | 850.000000 | 850.000000 | 850.000000 | |
| mean | 7.943946 | 5.897034 | 3.793712 | 5.772744 | 1.489152 | 6.102142 | |
| std | 1.962344 | 1.883114 | 1.221307 | 1.601264 | 0.899109 | 1.960872 | |
| min | 1.351000 | 0.000000 | 0.000000 | 0.877000 | 0.000000 | 1.117000 | |
| 25% | 6.682250 | 4.683250 | 3.030000 | 4.829250 | 0.873750 | 4.766000 | |
| 50% | 7.826000 | 5.812500 | 3.726500 | 5.794500 | 1.362000 | 6.027500 | |
| 75% | 9.210250 | 7.012500 | 4.506500 | 6.737500 | 1.983750 | 7.258750 | |
| max | 14.634000 | 12.360000 | 8.228000 | 10.938000 | 4.902000 | 12.903000 | |

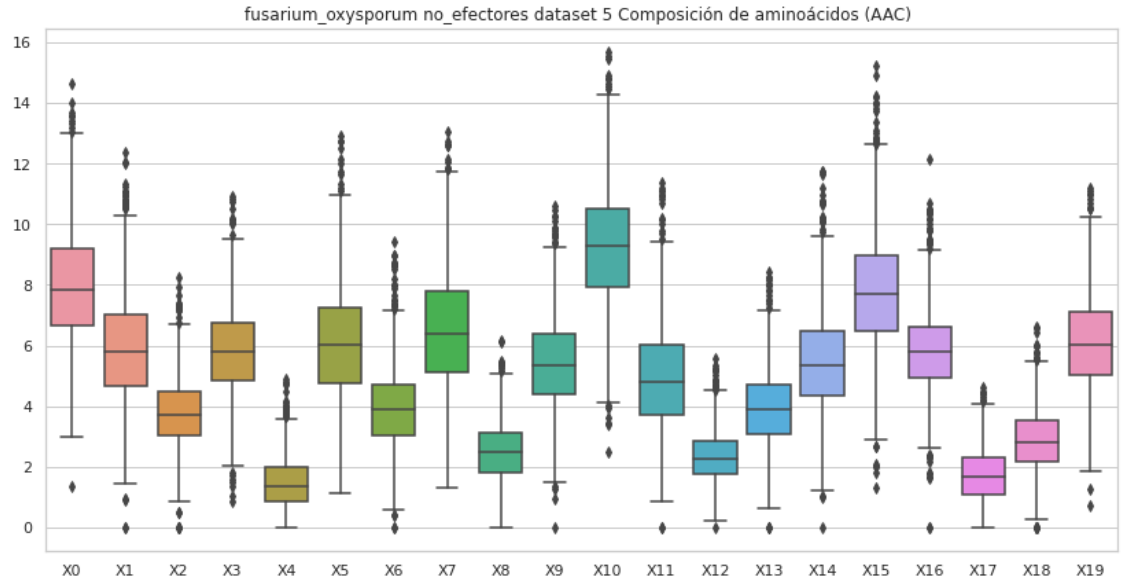
| | X6 | X7 | X8 | X9 | X10 | X11 | \ |
|---|---|---|---|---|---|---|---|
| count | 850.000000 | 850.000000 | 850.000000 | 850.000000 | 850.000000 | 850.000000 | |
| mean | 3.939511 | 6.565465 | 2.526679 | 5.473461 | 9.236561 | 4.929661 | |
| std | 1.436134 | 1.969926 | 1.013235 | 1.627829 | 2.067078 | 1.773320 | |
| min | 0.000000 | 1.333000 | 0.000000 | 0.000000 | 2.500000 | 0.000000 | |
| 25% | 3.015500 | 5.122750 | 1.808500 | 4.403750 | 7.926000 | 3.704000 | |
| 50% | 3.909000 | 6.389500 | 2.497500 | 5.365000 | 9.299500 | 4.787500 | |
| 75% | 4.689500 | 7.785500 | 3.144750 | 6.389000 | 10.506750 | 6.007500 | |
| max | 9.412000 | 13.043000 | 6.154000 | 10.588000 | 15.672000 | 11.392000 | |

| | X12 | X13 | X14 | X15 | X16 | X17 | \ |
|---|---|---|---|---|---|---|---|
| count | 850.000000 | 850.000000 | 850.000000 | 850.000000 | 850.000000 | 850.000000 | |
| mean | 2.345504 | 3.997058 | 5.549867 | 7.820841 | 5.830851 | 1.732480 | |
| std | 0.878393 | 1.353965 | 1.708768 | 2.046942 | 1.501040 | 0.900993 | |
| min | 0.000000 | 0.000000 | 0.000000 | 1.333000 | 0.000000 | 0.000000 | |
| 25% | 1.761250 | 3.062750 | 4.348000 | 6.491000 | 4.950750 | 1.109000 | |

| | | | | | | |
|---|---|---|---|---|---|---|
| 50% | 2.273000 | 3.897500 | 5.352500 | 7.692000 | 5.792500 | 1.654500 |
| 75% | 2.871500 | 4.714500 | 6.488750 | 8.952750 | 6.632500 | 2.292500 |
| max | 5.583000 | 8.447000 | 11.741000 | 15.236000 | 12.162000 | 4.632000 |

| | X18 | X19 |
|---|---|---|
| count | 850.000000 | 850.000000 |
| mean | 2.902035 | 6.151405 |
| std | 1.104100 | 1.604106 |
| min | 0.000000 | 0.746000 |
| 25% | 2.195000 | 5.028500 |
| 50% | 2.815500 | 6.037000 |
| 75% | 3.515500 | 7.130750 |
| max | 6.633000 | 11.194000 |






## 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 5, con valores atípicos.
Valores del documento csv.
```

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.028218  0.000000  0.044343  0.038968  0.025531  0.037624  0.005375
1    0.028988  0.002521  0.006302  0.017645  0.002521  0.011343  0.010083
2    0.034020  0.003402  0.045927  0.040824  0.020412  0.032319  0.010206
3    0.098652  0.019730  0.059191  0.049326  0.069057  0.029596  0.029596
4    0.024772  0.000826  0.018166  0.009909  0.006606  0.018992  0.003303
..        …         …         …         …         …         …         …
995  0.039929  0.005324  0.026619  0.042591  0.021295  0.055900  0.018633
996  0.046776  0.010736  0.032207  0.029139  0.020704  0.026072  0.023005
997  0.016844  0.006738  0.022234  0.034362  0.009433  0.036384  0.005390
998  0.040158  0.008238  0.033980  0.032950  0.028832  0.023683  0.015445
999  0.027178  0.001553  0.014754  0.020189  0.009318  0.027954  0.003883

           X7        X8        X9  …       X74       X75       X76  \
0    0.017468  0.029562  0.034937  …  0.008863  0.026526 -0.001237
1    0.012603  0.025207  0.025207  …  0.013782  0.018259  0.006638
2    0.039123  0.027216  0.047628  …  0.003095  0.012462  0.018223
3    0.039461  0.059191  0.118383  …  0.054917  0.047651  0.036197
4    0.009909  0.013212  0.012386  … -0.007055 -0.001778  0.027723
..        …         …         …  …         …         …         …
995  0.018633  0.029281  0.045252  … -0.009077 -0.007014  0.053442
996  0.021471  0.014570  0.059812  … -0.010009  0.006660 -0.005951
997  0.011454  0.023582  0.017518  …  0.000988  0.008599  0.006408
998  0.019564  0.018535  0.066930  … -0.012009 -0.013285  0.008417
999  0.010871  0.006989  0.020966  …  0.002478  0.002962  0.005059

          X77       X78       X79       X80       X81       X82         X83
0   -0.007800  0.025439  0.017349  0.002164  0.011506  0.004750  efectores
1    0.003131  0.004558 -0.000723  0.005863  0.009455 -0.001407  efectores
2    0.008445  0.021100  0.016816 -0.013625  0.018107  0.016092  efectores
3   -0.025716 -0.077376  0.071743 -0.004742 -0.031730  0.081886  efectores
4    0.009351  0.016926  0.023330 -0.010420  0.004789  0.018271  efectores
..        …         …         …         …         …         …         …
995 -0.001944 -0.023660  0.003497  0.026164  0.022077  0.033162  efectores
996  0.010138 -0.002473  0.011798  0.000414 -0.002855  0.020094  efectores
```

```
997  0.000840  0.019822  0.014940  0.003491  0.017171  0.008405  efectores
998 -0.025850 -0.000185  0.012897  0.012110  0.006035  0.027774  efectores
999  0.009366  0.010623  0.025053  0.000411  0.002185  0.029521  efectores

[1000 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC)  hidro_mass efectores
fusarium_oxysporum dataset 5, con valores atípicos.
Estadísticas.

| | X0 | X1 | X2 | X3 | X4 | \ |
|---|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | |
| mean | 0.294147 | 0.007581 | 0.283241 | 0.030540 | 0.275840 | |
| std | 8.042428 | 0.008673 | 8.042472 | 0.065522 | 8.043006 | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | |
| 25% | 0.026467 | 0.002798 | 0.015853 | 0.017083 | 0.010273 | |
| 50% | 0.034819 | 0.005480 | 0.024797 | 0.025264 | 0.016575 | |
| 75% | 0.044134 | 0.009592 | 0.033750 | 0.035416 | 0.023702 | |
| max | 254.346056 | 0.099090 | 254.346056 | 1.989174 | 254.346056 | |

| | X5 | X6 | X7 | X8 | X9 | … | \ |
|---|---|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | … | |
| mean | 0.540127 | 1.029402 | 0.536200 | 0.534995 | 0.299304 | … | |
| std | 16.085300 | 32.172135 | 16.085669 | 16.085703 | 8.042286 | … | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.002416 | … | |
| 25% | 0.020656 | 0.005876 | 0.015235 | 0.013970 | 0.026208 | … | |
| 50% | 0.028242 | 0.010236 | 0.022033 | 0.020949 | 0.037658 | … | |
| 75% | 0.037121 | 0.015495 | 0.029875 | 0.030476 | 0.053007 | … | |
| max | 508.692111 | 1017.384223 | 508.692111 | 508.692111 | 254.346056 | … | |

| | X73 | X74 | X75 | X76 | X77 | \ |
|---|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | |
| mean | 0.799598 | -0.273765 | 0.080621 | -0.538845 | -0.803320 | |
| std | 24.955801 | 8.664804 | 2.567472 | 17.319667 | 25.573344 | |
| min | -4.282193 | -273.994288 | -6.193744 | -547.660047 | -808.686347 | |
| 25% | 0.006049 | -0.006282 | -0.001726 | 0.005623 | -0.005716 | |
| 50% | 0.014528 | 0.003110 | 0.005741 | 0.014955 | 0.002986 | |
| 75% | 0.023222 | 0.011478 | 0.014792 | 0.024021 | 0.010784 | |
| max | 789.170268 | 0.191857 | 80.949871 | 0.171865 | 3.611544 | |

| | X78 | X79 | X80 | X81 | X82 |
|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | -0.292374 | 0.673207 | -3.255454 | -0.788488 | -1.091791 |
| std | 9.374344 | 21.045574 | 103.052346 | 25.314816 | 34.979947 |
| min | -296.429923 | -5.038079 | -3258.797392 | -800.496982 | -1106.148453 |
| 25% | -0.001322 | 0.006023 | -0.006359 | -0.002402 | 0.006255 |
| 50% | 0.005443 | 0.014597 | 0.002820 | 0.006348 | 0.014950 |

```
75%       0.014152     0.023473     0.011578     0.014706     0.024561
max       0.297252   665.506860     1.741853     4.938123     0.210337

[8 rows x 83 columns]
```

```
no_efectores
Composición de pseudo aminoácidos (PseAAC)  hidro_mass no_efectores
fusarium_oxysporum dataset 5, con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.023574  0.007254  0.025388  0.013601  0.014507  0.031735  0.009067
1    0.020943  0.009666  0.021748  0.017721  0.008860  0.025776  0.016110
2    0.032416  0.013893  0.046308  0.046308  0.029329  0.035503  0.021611
3    0.031589  0.004449  0.025805  0.027585  0.015127  0.027585  0.009788
4    0.022574  0.003892  0.004671  0.007784  0.014012  0.025688  0.003114
..        …         …         …         …         …         …
995  0.028110  0.005111  0.020443  0.021721  0.011499  0.033221  0.014055
996  0.058618  0.000000  0.039079  0.058618  0.029309  0.029309  0.019539
997  0.039035  0.013777  0.048220  0.036739  0.027554  0.034443  0.011481
998  0.058822  0.009804  0.040849  0.039215  0.029411  0.029411  0.009804
999  0.048293  0.013171  0.048293  0.065853  0.039512  0.035122  0.008780

           X7        X8        X9  …       X74       X75       X76  \
0    0.033548  0.017227  0.037175  …  0.008567 -0.001573  0.026257
1    0.009666  0.015304  0.027387  …  0.019754  0.014942  0.014008
2    0.041678  0.047852  0.078724  … -0.028284 -0.012220 -0.001624
3    0.016462  0.015127  0.034703  …  0.004474  0.007909  0.004962
4    0.021018  0.006617  0.022574  …  0.004276  0.000083  0.011065
..        …         …         …  …         …         …
995  0.028110  0.026832  0.025554  …  0.002063  0.009665  0.017234
996  0.029309  0.019539  0.039079  …  0.062372  0.049144 -0.004896
997  0.032147  0.025258  0.050516  …  0.011631  0.021996  0.026262
998  0.047385  0.034313  0.050653  …  0.021134  0.036162  0.029702
999  0.021951  0.021951  0.070244  …  0.023533  0.021476  0.022495

           X77       X78       X79       X80       X81       X82           X83
0   -0.001512 -0.005252  0.015090  0.016980  0.012720  0.029910  no_efectores
1    0.006458  0.005747  0.019061  0.006176  0.011242  0.021324  no_efectores
2    0.003332 -0.007746  0.025909 -0.016658  0.001132  0.022278  no_efectores
3    0.011769  0.008106  0.018690 -0.003272  0.000990  0.004227  no_efectores
4    0.016997  0.005028  0.011293  0.012944  0.002557  0.018930  no_efectores
..        …         …         …         …         …         …
995  0.004028  0.001888  0.020196 -0.006619 -0.002941  0.026909  no_efectores
996  0.128040  0.092426 -0.053730  0.062618  0.069950 -0.042057  no_efectores
997 -0.031061 -0.012445  0.032959  0.037269  0.046257  0.008607  no_efectores
```

```
998  0.019038  0.014539  0.009497  0.020259  0.025991  0.008334  no_efectores
999 -0.022016 -0.038903  0.013138  0.027384  0.046128 -0.041660  no_efectores

[1000 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC)  hidro_mass no_efectores
fusarium_oxysporum dataset 5, con valores atípicos.
Estadísticas.

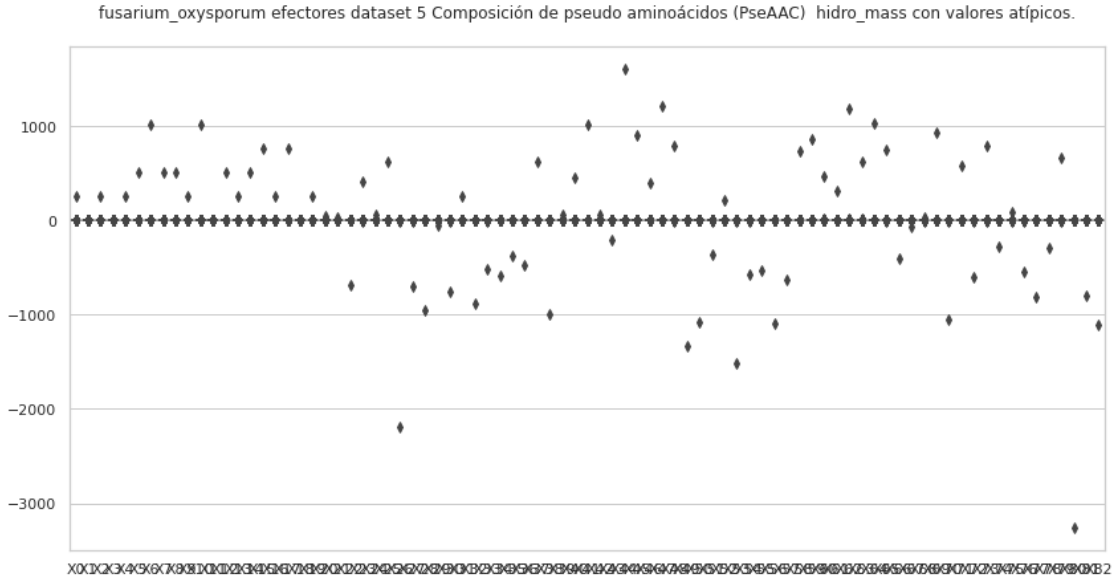
| | X0 | X1 | X2 | X3 | X4 | \ |
|---|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | |
| mean | 0.038716 | 0.008173 | 0.028443 | 0.030652 | 0.020298 | |
| std | 0.019347 | 0.008438 | 0.016140 | 0.019005 | 0.015331 | |
| min | 0.003567 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | |
| 25% | 0.028037 | 0.003180 | 0.017763 | 0.018296 | 0.011777 | |
| 50% | 0.035841 | 0.006042 | 0.026218 | 0.028089 | 0.017879 | |
| 75% | 0.045858 | 0.010673 | 0.036417 | 0.038723 | 0.025318 | |
| max | 0.239137 | 0.116829 | 0.176095 | 0.229417 | 0.220119 | |

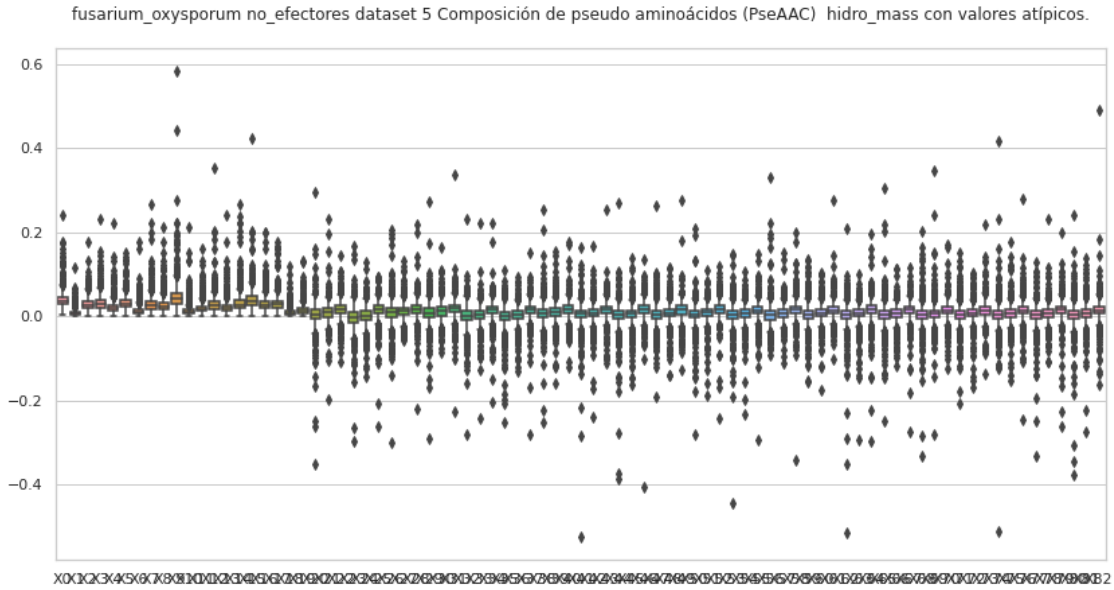
| | X5 | X6 | X7 | X8 | X9 | … | \ |
|---|---|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | … | |
| mean | 0.031606 | 0.013524 | 0.027501 | 0.024793 | 0.046755 | … | |
| std | 0.015568 | 0.012351 | 0.020120 | 0.017459 | 0.034498 | … | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | … | |
| 25% | 0.021718 | 0.006746 | 0.016814 | 0.014724 | 0.029467 | … | |
| 50% | 0.029101 | 0.011237 | 0.024224 | 0.022233 | 0.041495 | … | |
| 75% | 0.038226 | 0.016541 | 0.034025 | 0.030663 | 0.055903 | … | |
| max | 0.150003 | 0.176095 | 0.264696 | 0.211757 | 0.582332 | … | |

| | X73 | X74 | X75 | X76 | X77 | \ |
|---|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | |
| mean | 0.013787 | 0.001751 | 0.005899 | 0.013871 | 0.001305 | |
| std | 0.021190 | 0.031610 | 0.020867 | 0.023143 | 0.025912 | |
| min | -0.136308 | -0.512769 | -0.152647 | -0.247911 | -0.331963 | |
| 25% | 0.003749 | -0.007387 | -0.002629 | 0.005193 | -0.006612 | |
| 50% | 0.013144 | 0.003082 | 0.006067 | 0.014079 | 0.003621 | |
| 75% | 0.023632 | 0.011392 | 0.014950 | 0.022854 | 0.012628 | |
| max | 0.219087 | 0.417028 | 0.176290 | 0.277208 | 0.128040 | |

| | X78 | X79 | X80 | X81 | X82 |
|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 0.005585 | 0.014335 | 0.001619 | 0.006179 | 0.013952 |
| std | 0.020077 | 0.023897 | 0.030688 | 0.023557 | 0.026197 |
| min | -0.108998 | -0.261822 | -0.377578 | -0.274461 | -0.161573 |
| 25% | -0.002380 | 0.005170 | -0.007242 | -0.002455 | 0.005053 |
| 50% | 0.005188 | 0.015046 | 0.002564 | 0.006083 | 0.013905 |
| 75% | 0.014742 | 0.023769 | 0.012645 | 0.016080 | 0.023527 |

```
max       0.231535      0.199788      0.239729      0.139138      0.490056

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

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 | \ |
|---|---|---|---|---|---|---|---|---|
| 0 | 0.028218 | 0.000000 | 0.044343 | 0.038968 | 0.025531 | 0.037624 | 0.005375 | |
| 1 | 0.028988 | 0.002521 | 0.006302 | 0.017645 | 0.002521 | 0.011343 | 0.010083 | |
| 2 | 0.034020 | 0.003402 | 0.045927 | 0.040824 | 0.020412 | 0.032319 | 0.010206 | |
| 4 | 0.024772 | 0.000826 | 0.018166 | 0.009909 | 0.006606 | 0.018992 | 0.003303 | |
| 5 | 0.035064 | 0.007277 | 0.018525 | 0.034403 | 0.013893 | 0.037049 | 0.013232 | |
| .. | … | … | … | … | … | … | … | |
| 995 | 0.039929 | 0.005324 | 0.026619 | 0.042591 | 0.021295 | 0.055900 | 0.018633 | |
| 996 | 0.046776 | 0.010736 | 0.032207 | 0.029139 | 0.020704 | 0.026072 | 0.023005 | |
| 997 | 0.016844 | 0.006738 | 0.022234 | 0.034362 | 0.009433 | 0.036384 | 0.005390 | |
| 998 | 0.040158 | 0.008238 | 0.033980 | 0.032950 | 0.028832 | 0.023683 | 0.015445 | |
| 999 | 0.027178 | 0.001553 | 0.014754 | 0.020189 | 0.009318 | 0.027954 | 0.003883 | |

| | X7 | X8 | X9 | … | X74 | X75 | X76 | \ |
|---|---|---|---|---|---|---|---|---|
| 0 | 0.017468 | 0.029562 | 0.034937 | … | 0.008863 | 0.026526 | -0.001237 | |
| 1 | 0.012603 | 0.025207 | 0.025207 | … | 0.013782 | 0.018259 | 0.006638 | |
| 2 | 0.039123 | 0.027216 | 0.047628 | … | 0.003095 | 0.012462 | 0.018223 | |
| 4 | 0.009909 | 0.013212 | 0.012386 | … | -0.007055 | -0.001778 | 0.027723 | |
| 5 | 0.026464 | 0.034403 | 0.043003 | … | -0.000172 | -0.004592 | 0.040908 | |
| .. | … | … | … | … | … | … | … | |
| 995 | 0.018633 | 0.029281 | 0.045252 | … | -0.009077 | -0.007014 | 0.053442 | |
| 996 | 0.021471 | 0.014570 | 0.059812 | … | -0.010009 | 0.006660 | -0.005951 | |
| 997 | 0.011454 | 0.023582 | 0.017518 | … | 0.000988 | 0.008599 | 0.006408 | |
| 998 | 0.019564 | 0.018535 | 0.066930 | … | -0.012009 | -0.013285 | 0.008417 | |
| 999 | 0.010871 | 0.006989 | 0.020966 | … | 0.002478 | 0.002962 | 0.005059 | |

| | X77 | X78 | X79 | X80 | X81 | X82 | X83 |
|---|---|---|---|---|---|---|---|
| 0 | -0.007800 | 0.025439 | 0.017349 | 0.002164 | 0.011506 | 0.004750 | efectores |
| 1 | 0.003131 | 0.004558 | -0.000723 | 0.005863 | 0.009455 | -0.001407 | efectores |
| 2 | 0.008445 | 0.021100 | 0.016816 | -0.013625 | 0.018107 | 0.016092 | efectores |
| 4 | 0.009351 | 0.016926 | 0.023330 | -0.010420 | 0.004789 | 0.018271 | efectores |
| 5 | 0.002229 | 0.003397 | 0.015281 | -0.005167 | -0.000990 | 0.018135 | efectores |
| .. | … | … | … | … | … | … | |
| 995 | -0.001944 | -0.023660 | 0.003497 | 0.026164 | 0.022077 | 0.033162 | efectores |
| 996 | 0.010138 | -0.002473 | 0.011798 | 0.000414 | -0.002855 | 0.020094 | efectores |
| 997 | 0.000840 | 0.019822 | 0.014940 | 0.003491 | 0.017171 | 0.008405 | efectores |
| 998 | -0.025850 | -0.000185 | 0.012897 | 0.012110 | 0.006035 | 0.027774 | efectores |
| 999 | 0.009366 | 0.010623 | 0.025053 | 0.000411 | 0.002185 | 0.029521 | efectores |

[967 rows x 84 columns]

Composición de pseudo aminoácidos (PseAAC)  hidro_mass efectores
fusarium_oxysporum dataset 5, sin valores atípicos.
Estadísticas.

| | X0 | X1 | X2 | X3 | X4 | X5 | \ |
|---|---|---|---|---|---|---|---|
| count | 967.000000 | 967.000000 | 967.000000 | 967.000000 | 967.000000 | 967.000000 | |
| mean | 0.035461 | 0.006734 | 0.025946 | 0.027359 | 0.017563 | 0.029559 | |
| std | 0.013235 | 0.005806 | 0.013884 | 0.015744 | 0.010652 | 0.013456 | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | |
| 25% | 0.026301 | 0.002761 | 0.015685 | 0.016981 | 0.010168 | 0.020445 | |
| 50% | 0.034219 | 0.005324 | 0.024313 | 0.024910 | 0.016240 | 0.027770 | |
| 75% | 0.043531 | 0.009080 | 0.032984 | 0.035023 | 0.023038 | 0.036664 | |
| max | 0.100928 | 0.033304 | 0.129036 | 0.157141 | 0.125713 | 0.161295 | |

| | X6 | X7 | X8 | X9 | … | X73 | \ |
|---|---|---|---|---|---|---|---|
| count | 967.000000 | 967.000000 | 967.000000 | 967.000000 | … | 967.000000 | |
| mean | 0.011288 | 0.023357 | 0.022600 | 0.040330 | … | 0.015117 | |
| std | 0.007802 | 0.012864 | 0.012174 | 0.021450 | … | 0.015782 | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.002416 | … | -0.090632 | |
| 25% | 0.005813 | 0.014839 | 0.013744 | 0.025923 | … | 0.006754 | |
| 50% | 0.010052 | 0.021683 | 0.020731 | 0.037114 | … | 0.014770 | |
| 75% | 0.014988 | 0.029249 | 0.029552 | 0.051498 | … | 0.023230 | |
| max | 0.062856 | 0.141300 | 0.081774 | 0.177424 | … | 0.090579 | |

| | X74 | X75 | X76 | X77 | X78 | X79 | \ |
|---|---|---|---|---|---|---|---|
| count | 967.000000 | 967.000000 | 967.000000 | 967.000000 | 967.000000 | 967.000000 | |
| mean | 0.002428 | 0.006019 | 0.014502 | 0.002107 | 0.006333 | 0.014556 | |
| std | 0.020755 | 0.017818 | 0.016343 | 0.018102 | 0.016255 | 0.017464 | |
| min | -0.200647 | -0.204369 | -0.071978 | -0.148131 | -0.098999 | -0.099414 | |
| 25% | -0.005976 | -0.001540 | 0.005884 | -0.005351 | -0.001075 | 0.006393 | |
| 50% | 0.003281 | 0.005876 | 0.014938 | 0.002972 | 0.005535 | 0.014637 | |
| 75% | 0.011372 | 0.014799 | 0.023824 | 0.010529 | 0.014035 | 0.023347 | |
| max | 0.148339 | 0.147146 | 0.133640 | 0.084969 | 0.105900 | 0.134540 | |

| | X80 | X81 | X82 |
|---|---|---|---|
| count | 967.000000 | 967.000000 | 967.000000 |
| mean | 0.001806 | 0.006557 | 0.014917 |
| std | 0.021981 | 0.023009 | 0.018863 |
| min | -0.230912 | -0.165962 | -0.089733 |
| 25% | -0.006062 | -0.002111 | 0.006659 |
| 50% | 0.002996 | 0.006374 | 0.015008 |
| 75% | 0.011286 | 0.014581 | 0.024569 |
| max | 0.166513 | 0.369602 | 0.210337 |

[8 rows x 83 columns]

```
no_efectores
Composición de pseudo aminoácidos (PseAAC)  hidro_mass no_efectores
fusarium_oxysporum dataset 5, sin valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.023574  0.007254  0.025388  0.013601  0.014507  0.031735  0.009067
1    0.020943  0.009666  0.021748  0.017721  0.008860  0.025776  0.016110
2    0.032416  0.013893  0.046308  0.046308  0.029329  0.035503  0.021611
3    0.031589  0.004449  0.025805  0.027585  0.015127  0.027585  0.009788
4    0.022574  0.003892  0.004671  0.007784  0.014012  0.025688  0.003114
..        …         …         …         …         …         …         …
993  0.034690  0.007434  0.034690  0.024779  0.013215  0.019823  0.014867
994  0.044021  0.022010  0.039236  0.028709  0.021053  0.023924  0.022967
995  0.028110  0.005111  0.020443  0.021721  0.011499  0.033221  0.014055
997  0.039035  0.013777  0.048220  0.036739  0.027554  0.034443  0.011481
998  0.058822  0.009804  0.040849  0.039215  0.029411  0.029411  0.009804

           X7        X8        X9  …       X74       X75       X76  \
0    0.033548  0.017227  0.037175  …  0.008567 -0.001573  0.026257
1    0.009666  0.015304  0.027387  …  0.019754  0.014942  0.014008
2    0.041678  0.047852  0.078724  … -0.028284 -0.012220 -0.001624
3    0.016462  0.015127  0.034703  …  0.004474  0.007909  0.004962
4    0.021018  0.006617  0.022574  …  0.004276  0.000083  0.011065
..        …         …         …  …         …         …         …
993  0.023127  0.009911  0.047906  … -0.019503 -0.019072  0.012023
994  0.038279  0.020096  0.064117  …  0.016095  0.008541  0.009976
995  0.028110  0.026832  0.025554  …  0.002063  0.009665  0.017234
997  0.032147  0.025258  0.050516  …  0.011631  0.021996  0.026262
998  0.047385  0.034313  0.050653  …  0.021134  0.036162  0.029702

           X77       X78       X79       X80       X81       X82           X83
0   -0.001512 -0.005252  0.015090  0.016980  0.012720  0.029910  no_efectores
1    0.006458  0.005747  0.019061  0.006176  0.011242  0.021324  no_efectores
2    0.003332 -0.007746  0.025909 -0.016658  0.001132  0.022278  no_efectores
3    0.011769  0.008106  0.018690 -0.003272  0.000990  0.004227  no_efectores
4    0.016997  0.005028  0.011293  0.012944  0.002557  0.018930  no_efectores
..        …         …         …         …         …         …           …
993 -0.006504  0.005195 -0.001364  0.022459  0.026409  0.016164  no_efectores
994 -0.004349 -0.007741  0.020880  0.029995  0.005911  0.018161  no_efectores
995  0.004028  0.001888  0.020196 -0.006619 -0.002941  0.026909  no_efectores
997 -0.031061 -0.012445  0.032959  0.037269  0.046257  0.008607  no_efectores
998  0.019038  0.014539  0.009497  0.020259  0.025991  0.008334  no_efectores

[878 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC)  hidro_mass no_efectores
fusarium_oxysporum dataset 5, sin valores atípicos.
Estadísticas.

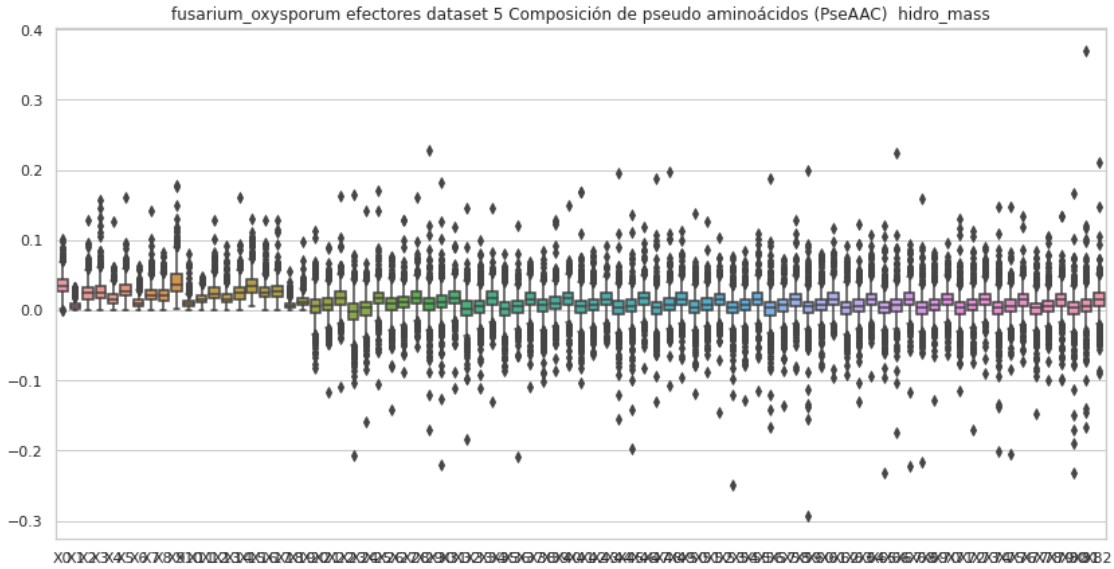
| | X0 | X1 | X2 | X3 | X4 | X5 | \ |
|---|---|---|---|---|---|---|---|
| count | 878.000000 | 878.000000 | 878.000000 | 878.000000 | 878.000000 | 878.000000 | |
| mean | 0.035523 | 0.006976 | 0.026213 | 0.027383 | 0.017876 | 0.029117 | |
| std | 0.012208 | 0.005286 | 0.012497 | 0.013081 | 0.009373 | 0.010955 | |
| min | 0.003567 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.003757 | |
| 25% | 0.027525 | 0.003172 | 0.016948 | 0.017709 | 0.011209 | 0.021105 | |
| 50% | 0.034647 | 0.005651 | 0.024878 | 0.026379 | 0.016617 | 0.027683 | |
| 75% | 0.043019 | 0.009671 | 0.033684 | 0.035657 | 0.023577 | 0.036020 | |
| max | 0.087392 | 0.033293 | 0.069024 | 0.073823 | 0.052875 | 0.072729 | |

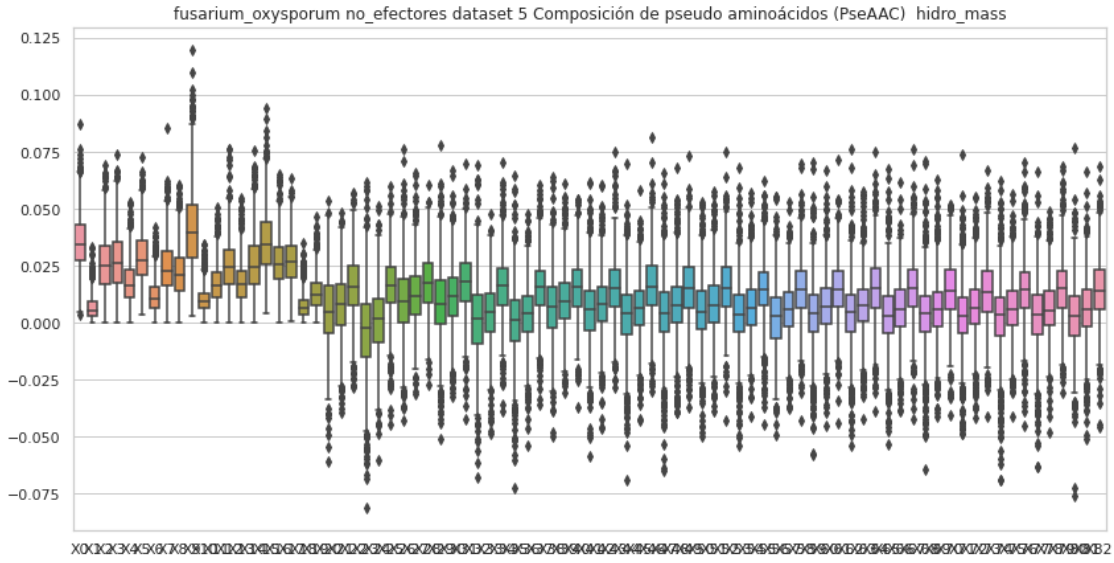
| | X6 | X7 | X8 | X9 | … | X73 | \ |
|---|---|---|---|---|---|---|---|
| count | 878.000000 | 878.000000 | 878.000000 | 878.000000 | … | 878.000000 | |
| mean | 0.011847 | 0.024175 | 0.022016 | 0.041230 | … | 0.014067 | |
| std | 0.007301 | 0.011430 | 0.010871 | 0.018627 | … | 0.014157 | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.003229 | … | -0.041254 | |
| 25% | 0.006524 | 0.016430 | 0.014357 | 0.028377 | … | 0.004908 | |
| 50% | 0.010500 | 0.022996 | 0.021277 | 0.039417 | … | 0.013384 | |
| 75% | 0.015587 | 0.031558 | 0.028739 | 0.052009 | … | 0.023017 | |
| max | 0.042066 | 0.085610 | 0.060442 | 0.119734 | … | 0.066655 | |

| | X74 | X75 | X76 | X77 | X78 | X79 | \ |
|---|---|---|---|---|---|---|---|
| count | 878.000000 | 878.000000 | 878.000000 | 878.000000 | 878.000000 | 878.000000 | |
| mean | 0.002072 | 0.006388 | 0.014510 | 0.002975 | 0.006114 | 0.015152 | |
| std | 0.015775 | 0.013020 | 0.014081 | 0.014830 | 0.012759 | 0.014097 | |
| min | -0.068829 | -0.040806 | -0.045403 | -0.063318 | -0.048187 | -0.030642 | |
| 25% | -0.005435 | -0.001285 | 0.006503 | -0.005317 | -0.001198 | 0.006425 | |
| 50% | 0.003402 | 0.006158 | 0.014406 | 0.003891 | 0.005293 | 0.015358 | |
| 75% | 0.011008 | 0.014311 | 0.022305 | 0.012126 | 0.013808 | 0.022993 | |
| max | 0.062205 | 0.056354 | 0.070302 | 0.066573 | 0.048921 | 0.068571 | |

| | X80 | X81 | X82 |
|---|---|---|---|
| count | 878.000000 | 878.000000 | 878.000000 |
| mean | 0.002832 | 0.006376 | 0.014668 |
| std | 0.014471 | 0.013939 | 0.013874 |
| min | -0.075673 | -0.051155 | -0.046012 |
| 25% | -0.005792 | -0.001393 | 0.006030 |
| 50% | 0.002808 | 0.005988 | 0.014158 |
| 75% | 0.011699 | 0.014376 | 0.023364 |
| max | 0.076549 | 0.066339 | 0.068842 |

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.047824  0.000000  0.075153  0.066043  0.043270  0.063766  0.009109
1    0.126693  0.011017  0.027542  0.077118  0.011017  0.049576  0.044067
2    0.039145  0.003915  0.052846  0.046974  0.023487  0.037188  0.011744
3    0.067644  0.013529  0.040586  0.033822  0.047351  0.020293  0.020293
4    0.032689  0.001090  0.023972  0.013076  0.008717  0.025062  0.004359
..        …         …         …         …         …         …         …
995  0.046348  0.006180  0.030898  0.049437  0.024719  0.064887  0.021629
996  0.057129  0.013112  0.039335  0.035589  0.025287  0.031842  0.028096
997  0.031153  0.012461  0.041122  0.063553  0.017446  0.067291  0.009969
998  0.050965  0.010454  0.043124  0.041817  0.036590  0.030056  0.019602
999  0.037094  0.002120  0.020137  0.027555  0.012718  0.038154  0.005299

           X7        X8        X9  …       X32       X33       X34  \
```

```
0    0.029606  0.050102  0.059211  …  0.024397  0.008879  0.007614
1    0.055084  0.110168  0.110168  … -0.092046  0.115780 -0.011006
2    0.045017  0.031316  0.054803  …  0.019117 -0.005513 -0.000634
3    0.027058  0.040586  0.081173  …  0.013663  0.037414  0.050672
4    0.013076  0.017434  0.016345  …  0.031308  0.028759  0.020888
..        …         …         …  …         …         …
995  0.021629  0.033988  0.052527  …  0.027060  0.014514 -0.001899
996  0.026223  0.017794  0.073050  …  0.017405  0.007450  0.020319
997  0.021184  0.043615  0.032399  …  0.001375  0.017517 -0.003714
998  0.024829  0.023522  0.084942  … -0.002617  0.008185  0.017410
999  0.014838  0.009538  0.028615  …  0.024057  0.016143  0.038002

          X35       X36       X37       X38       X39       X40        X41
0    0.014649  0.012023  0.023114 -0.002097  0.029404  0.008050  efectores
1    0.054277 -0.059872  0.060205  0.029013 -0.003162 -0.006151  efectores
2    0.021262  0.039099  0.030045  0.020969  0.019349  0.018516  efectores
3    0.009227  0.005586 -0.003581  0.024820  0.049192  0.056148  efectores
4    0.034316  0.036983  0.040901  0.036582  0.030786  0.024110  efectores
..        …         …         …         …         …         …
995  0.031250  0.011240  0.032726  0.062033  0.004059  0.038493  efectores
996  0.006777  0.025981  0.008763 -0.007268  0.014409  0.024541  efectores
997  0.014274  0.031575  0.030912  0.011851  0.027632  0.015546  efectores
998  0.012943  0.016702 -0.004107  0.010682  0.016368  0.035247  efectores
999  0.017378  0.024554  0.038601  0.006904  0.034194  0.040292  efectores

[1000 rows x 42 columns]
```

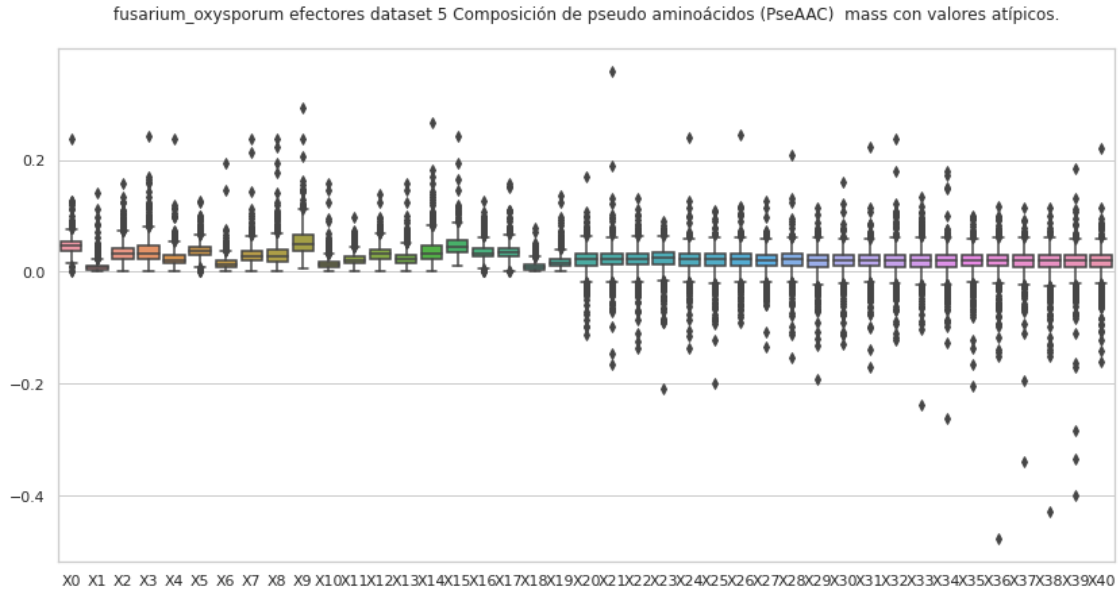
Composición de pseudo aminoácidos (PseAAC)  mass efectores fusarium_oxysporum dataset 5, con valores atípicos.
Estadísticas.

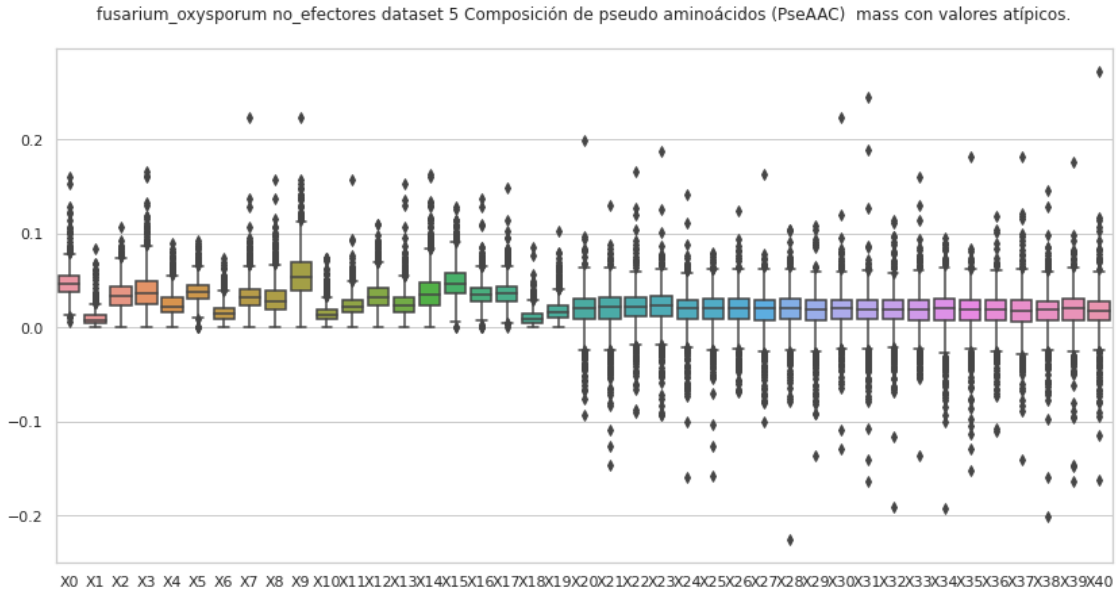
```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.046953     0.009360     0.034708     0.037396     0.023657
std       0.015661     0.010386     0.018282     0.022776     0.015577
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       0.037469     0.004070     0.022479     0.023349     0.014512
50%       0.045830     0.007247     0.031841     0.032881     0.021363
75%       0.053083     0.011722     0.042526     0.045949     0.030204
max       0.236525     0.140873     0.157683     0.242462     0.236525

                X5           X6           X7           X8           X9  …  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
mean      0.038261     0.015327     0.031062     0.031365     0.053195  …
std       0.013388     0.012219     0.017097     0.021137     0.025804  …
min       0.000000     0.000000     0.000000     0.000000     0.005809  …
25%       0.029816     0.008167     0.020797     0.018710     0.036641  …
```

```
50%       0.037142     0.013516     0.028742     0.026631     0.050257  …
75%       0.044460     0.019587     0.038408     0.038826     0.067036  …
max       0.126016     0.194435     0.236525     0.236525     0.292262  …

               X31          X32          X33          X34          X35  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.018857     0.019315     0.019450     0.018101     0.017939
std       0.022678     0.023571     0.022962     0.024441     0.023486
min      -0.169248    -0.122843    -0.237029    -0.263173    -0.203253
25%       0.010253     0.009365     0.009244     0.009273     0.010082
50%       0.020904     0.021156     0.021340     0.020384     0.020578
75%       0.029820     0.029847     0.030691     0.030287     0.029947
max       0.221946     0.236975     0.133999     0.179297     0.099181

               X36          X37          X38          X39          X40
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.018139     0.017601     0.016107     0.016586     0.017188
std       0.027681     0.023838     0.027521     0.030110     0.024725
min      -0.476979    -0.339452    -0.427518    -0.399372    -0.159543
25%       0.009974     0.008649     0.007957     0.008693     0.009101
50%       0.020389     0.019619     0.019316     0.019491     0.019872
75%       0.030257     0.029511     0.029699     0.029125     0.029683
max       0.117921     0.113824     0.115040     0.185237     0.220346

[8 rows x 41 columns]



no_efectores
Composición de pseudo aminoácidos (PseAAC)  mass no_efectores fusarium_oxysporum
dataset 5, con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.025502  0.007847  0.027463  0.014712  0.015693  0.034329  0.009808
1    0.028876  0.013327  0.029986  0.024433  0.012217  0.035540  0.022212
2    0.034811  0.014919  0.049730  0.049730  0.031495  0.038126  0.023207
3    0.045512  0.006410  0.037179  0.039743  0.021794  0.039743  0.014102
4    0.041049  0.007077  0.008493  0.014155  0.025478  0.046710  0.005662
..        …         …         …         …         …         …         …
995  0.033816  0.006148  0.024593  0.026130  0.013834  0.039964  0.016908
996  0.160583  0.000000  0.107055  0.160583  0.080291  0.080291  0.053528
997  0.042371  0.014954  0.052340  0.039878  0.029909  0.037386  0.012462
998  0.075908  0.012651  0.052714  0.050605  0.037954  0.037954  0.012651
999  0.077576  0.021157  0.077576  0.105786  0.063471  0.056419  0.014105

           X7        X8        X9  …       X32       X33       X34  \
0    0.036291  0.018636  0.040214  …  0.022223  0.019151  0.020539
```

```
1    0.013327  0.021102  0.037761  …  0.003625  0.026763  0.014123
2    0.044757  0.051387  0.084540  …  0.017726  0.003023  0.014543
3    0.023717  0.021794  0.049999  …  0.012311  0.009714  0.035312
4    0.038218  0.012031  0.041049  …  0.034564  0.027022  0.010493
..        …         …         …  …         …         …
995  0.033816  0.032279  0.030742  …  0.041645  0.036866 -0.012739
996  0.080291  0.053528  0.107055  … -0.007760 -0.135540  0.081884
997  0.034894  0.027416  0.054833  …  0.020402  0.012035  0.010202
998  0.061148  0.044279  0.065365  … -0.013419  0.022461 -0.025748
999  0.035262  0.035262  0.112838  …  0.006429 -0.029169  0.082147

          X35       X36       X37       X38       X39       X40           X41
0    0.027269  0.015145  0.030224  0.028404  0.016324  0.032355  no_efectores
1    0.027244  0.015085  0.020630  0.019314  0.026281  0.029401  no_efectores
2    0.021999  0.032867 -0.004361 -0.001743  0.027824  0.023923  no_efectores
3    0.005130  0.014375  0.028381  0.007149  0.026928  0.006090  no_efectores
4    0.041439  0.028461  0.022267  0.020121  0.020535  0.034421  no_efectores
..        …         …         …         …         …         …           …
995  0.026578  0.023057  0.036258  0.020733  0.024295  0.032372  no_efectores
996 -0.076177 -0.049227 -0.011384 -0.013414 -0.147193 -0.115213  no_efectores
997  0.006997  0.000893  0.022500  0.028506  0.035775  0.009342  no_efectores
998  0.000382 -0.000968  0.004134  0.038329  0.012255  0.010754  no_efectores
999  0.006227 -0.070293  0.008328  0.036135  0.021105 -0.066922  no_efectores

[1000 rows x 42 columns]


Composición de pseudo aminoácidos (PseAAC)  mass no_efectores fusarium_oxysporum
dataset 5, con valores atípicos.
Estadísticas.

                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.047572     0.009848     0.035217     0.038817     0.024578
std       0.014762     0.008792     0.015874     0.021278     0.013170
min       0.006425     0.000000     0.000000     0.000000     0.000000
25%       0.038287     0.004502     0.024124     0.024371     0.016223
50%       0.046472     0.007812     0.034043     0.035887     0.022561
75%       0.054629     0.012953     0.044284     0.049843     0.031906
max       0.160583     0.083455     0.107055     0.165823     0.089519

                X5           X6           X7           X8           X9  …  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
mean      0.038750     0.016256     0.033449     0.030979     0.056733  …
std       0.012207     0.010399     0.017140     0.017522     0.024709  …
min       0.000000     0.000000     0.000000     0.000000     0.000000  …
25%       0.030959     0.009344     0.023407     0.019809     0.040013  …
50%       0.037896     0.014521     0.031691     0.028223     0.054292  …
```

```
75%       0.044924     0.021196     0.040847     0.038841     0.069713  …
max       0.092718     0.074364     0.223797     0.156658     0.223797  …

               X31          X32          X33          X34          X35  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.017093     0.017649     0.018091     0.017497     0.017175
std       0.023905     0.021505     0.020514     0.022968     0.022411
min      -0.163639    -0.190125    -0.135540    -0.191489    -0.152003
25%       0.008474     0.008907     0.008038     0.007507     0.007840
50%       0.019519     0.019474     0.019205     0.020166     0.019316
75%       0.029465     0.028763     0.029220     0.030352     0.029419
max       0.244452     0.113625     0.160439     0.095701     0.182245

               X36          X37          X38          X39          X40
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.017152     0.016500     0.016733     0.017966     0.016798
std       0.020603     0.022513     0.023327     0.023150     0.024163
min      -0.109782    -0.140855    -0.201005    -0.163561    -0.161901
25%       0.007340     0.005945     0.007082     0.007605     0.007332
50%       0.018686     0.017673     0.019136     0.020181     0.018502
75%       0.029061     0.028991     0.028364     0.030222     0.028555
max       0.117747     0.182089     0.146257     0.175924     0.272240

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.047824  0.000000  0.075153  0.066043  0.043270  0.063766  0.009109
2    0.039145  0.003915  0.052846  0.046974  0.023487  0.037188  0.011744
4    0.032689  0.001090  0.023972  0.013076  0.008717  0.025062  0.004359
5    0.037672  0.007819  0.019902  0.036961  0.014927  0.039804  0.014216
6    0.063162  0.011484  0.022968  0.017226  0.037323  0.034452  0.007177
..        …         …         …         …         …         …         …
995  0.046348  0.006180  0.030898  0.049437  0.024719  0.064887  0.021629
996  0.057129  0.013112  0.039335  0.035589  0.025287  0.031842  0.028096
997  0.031153  0.012461  0.041122  0.063553  0.017446  0.067291  0.009969
998  0.050965  0.010454  0.043124  0.041817  0.036590  0.030056  0.019602
999  0.037094  0.002120  0.020137  0.027555  0.012718  0.038154  0.005299

           X7        X8        X9  …       X32       X33       X34  \
0    0.029606  0.050102  0.059211  …  0.024397  0.008879  0.007614
2    0.045017  0.031316  0.054803  …  0.019117 -0.005513 -0.000634
4    0.013076  0.017434  0.016345  …  0.031308  0.028759  0.020888
5    0.028431  0.036961  0.046201  …  0.016271  0.025813  0.026421
6    0.047371  0.018661  0.051678  …  0.004930  0.015197  0.016233
..        …         …         …  …       …         …         …
995  0.021629  0.033988  0.052527  …  0.027060  0.014514 -0.001899
996  0.026223  0.017794  0.073050  …  0.017405  0.007450  0.020319
997  0.021184  0.043615  0.032399  …  0.001375  0.017517 -0.003714
998  0.024829  0.023522  0.084942  … -0.002617  0.008185  0.017410
999  0.014838  0.009538  0.028615  …  0.024057  0.016143  0.038002
```

| | X35 | X36 | X37 | X38 | X39 | X40 | X41 |
|---|---|---|---|---|---|---|---|
| 0 | 0.014649 | 0.012023 | 0.023114 | -0.002097 | 0.029404 | 0.008050 | efectores |
| 2 | 0.021262 | 0.039099 | 0.030045 | 0.020969 | 0.019349 | 0.018516 | efectores |
| 4 | 0.034316 | 0.036983 | 0.040901 | 0.036582 | 0.030786 | 0.024110 | efectores |
| 5 | 0.041221 | 0.032309 | 0.018584 | 0.043950 | 0.016417 | 0.019483 | efectores |
| 6 | 0.046319 | 0.003175 | 0.007833 | 0.014400 | 0.008448 | 0.029972 | efectores |
| .. | … | … | … | … | … | … | |
| 995 | 0.031250 | 0.011240 | 0.032726 | 0.062033 | 0.004059 | 0.038493 | efectores |
| 996 | 0.006777 | 0.025981 | 0.008763 | -0.007268 | 0.014409 | 0.024541 | efectores |
| 997 | 0.014274 | 0.031575 | 0.030912 | 0.011851 | 0.027632 | 0.015546 | efectores |
| 998 | 0.012943 | 0.016702 | -0.004107 | 0.010682 | 0.016368 | 0.035247 | efectores |
| 999 | 0.017378 | 0.024554 | 0.038601 | 0.006904 | 0.034194 | 0.040292 | efectores |

[869 rows x 42 columns]

Composición de pseudo aminoácidos (PseAAC)  mass efectores fusarium_oxysporum
dataset 5, sin valores atípicos.
Estadísticas.

| | X0 | X1 | X2 | X3 | X4 | X5 | \ |
|---|---|---|---|---|---|---|---|
| count | 869.000000 | 869.000000 | 869.000000 | 869.000000 | 869.000000 | 869.000000 | |
| mean | 0.045430 | 0.008169 | 0.032099 | 0.033634 | 0.021732 | 0.036841 | |
| std | 0.011651 | 0.006018 | 0.013916 | 0.015962 | 0.010566 | 0.010696 | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.005427 | |
| 25% | 0.037335 | 0.004190 | 0.021957 | 0.022396 | 0.014271 | 0.029287 | |
| 50% | 0.045393 | 0.007030 | 0.031002 | 0.031240 | 0.020652 | 0.036534 | |
| 75% | 0.052248 | 0.011055 | 0.040869 | 0.042500 | 0.028204 | 0.043498 | |
| max | 0.090751 | 0.037335 | 0.088297 | 0.098307 | 0.060741 | 0.077088 | |

| | X6 | X7 | X8 | X9 | … | X31 | \ |
|---|---|---|---|---|---|---|---|
| count | 869.000000 | 869.000000 | 869.000000 | 869.000000 | … | 869.000000 | |
| mean | 0.013569 | 0.028853 | 0.027887 | 0.049932 | … | 0.020780 | |
| std | 0.007600 | 0.011848 | 0.012994 | 0.020182 | … | 0.015841 | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.005809 | … | -0.044895 | |
| 25% | 0.007891 | 0.020558 | 0.018298 | 0.035679 | … | 0.012118 | |
| 50% | 0.012778 | 0.027542 | 0.025532 | 0.048796 | … | 0.021599 | |
| 75% | 0.018099 | 0.036033 | 0.035895 | 0.063274 | … | 0.029791 | |
| max | 0.050241 | 0.072596 | 0.081388 | 0.121087 | … | 0.085326 | |

| | X32 | X33 | X34 | X35 | X36 | X37 | \ |
|---|---|---|---|---|---|---|---|
| count | 869.000000 | 869.000000 | 869.000000 | 869.000000 | 869.000000 | 869.000000 | |
| mean | 0.020327 | 0.020982 | 0.019976 | 0.020498 | 0.020055 | 0.020073 | |
| std | 0.015147 | 0.015538 | 0.015980 | 0.015480 | 0.015168 | 0.014804 | |
| min | -0.042311 | -0.034072 | -0.040379 | -0.042805 | -0.048383 | -0.043338 | |
| 25% | 0.011083 | 0.012177 | 0.010801 | 0.011885 | 0.011240 | 0.011180 | |
| 50% | 0.021523 | 0.022174 | 0.021255 | 0.021494 | 0.020755 | 0.020511 | |
| 75% | 0.029473 | 0.030687 | 0.030274 | 0.029974 | 0.029588 | 0.029512 | |

```
max      0.088623    0.084835    0.075165    0.086137    0.077954    0.071927

              X38         X39         X40
count  869.000000  869.000000  869.000000
mean     0.019358    0.019500    0.020014
std      0.015588    0.014878    0.016046
min     -0.052928   -0.046126   -0.056549
25%      0.010829    0.010684    0.011788
50%      0.020583    0.020218    0.020915
75%      0.030098    0.029268    0.030020
max      0.072081    0.068797    0.078216

[8 rows x 41 columns]
```

Composición de pseudo aminoácidos (PseAAC)  mass no_efectores fusarium_oxysporum
dataset 5, sin valores atípicos.
Valores del documento csv.

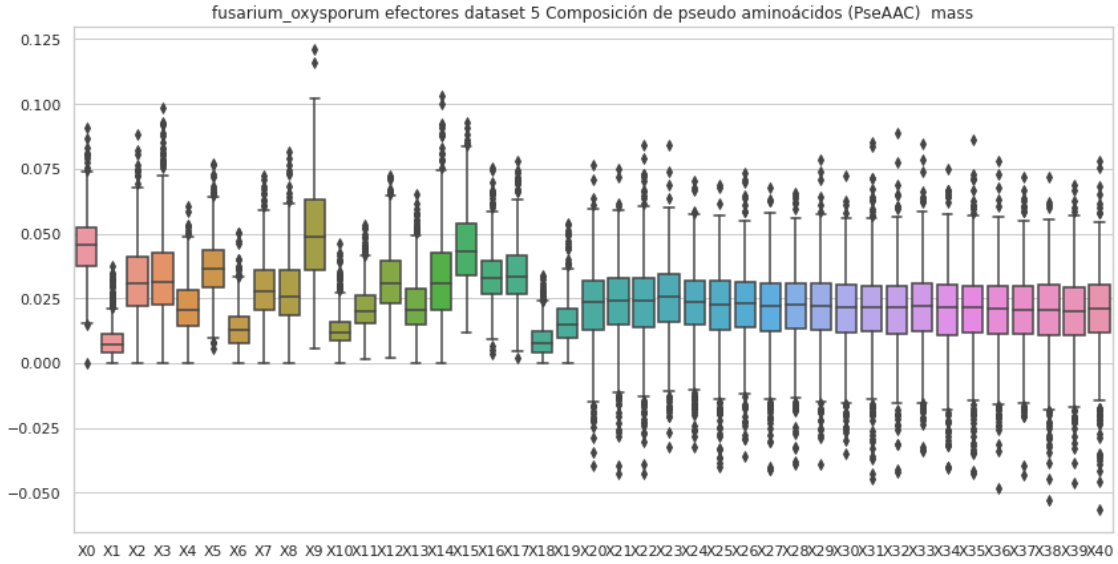
```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.025502  0.007847  0.027463  0.014712  0.015693  0.034329  0.009808
1    0.028876  0.013327  0.029986  0.024433  0.012217  0.035540  0.022212
2    0.034811  0.014919  0.049730  0.049730  0.031495  0.038126  0.023207
3    0.045512  0.006410  0.037179  0.039743  0.021794  0.039743  0.014102
4    0.041049  0.007077  0.008493  0.014155  0.025478  0.046710  0.005662
..        …         …         …         …         …         …         …
993  0.040682  0.008718  0.040682  0.029058  0.015498  0.023247  0.017435
994  0.046495  0.023248  0.041441  0.030323  0.022237  0.025269  0.024258
995  0.033816  0.006148  0.024593  0.026130  0.013834  0.039964  0.016908
997  0.042371  0.014954  0.052340  0.039878  0.029909  0.037386  0.012462
998  0.075908  0.012651  0.052714  0.050605  0.037954  0.037954  0.012651

           X7        X8        X9  …       X32       X33       X34  \
0    0.036291  0.018636  0.040214  …  0.022223  0.019151  0.020539
1    0.013327  0.021102  0.037761  …  0.003625  0.026763  0.014123
2    0.044757  0.051387  0.084540  …  0.017726  0.003023  0.014543
3    0.023717  0.021794  0.049999  …  0.012311  0.009714  0.035312
4    0.038218  0.012031  0.041049  …  0.034564  0.027022  0.010493
..        …         …         …  …  …       …         …         …
993  0.027121  0.011623  0.056179  …  0.026902  0.022555  0.033332
994  0.040431  0.021226  0.067721  …  0.013029  0.013195  0.017877
995  0.033816  0.032279  0.030742  …  0.041645  0.036866 -0.012739
997  0.034894  0.027416  0.054833  …  0.020402  0.012035  0.010202
998  0.061148  0.044279  0.065365  … -0.013419  0.022461 -0.025748

           X35       X36       X37       X38       X39       X40           X41
0    0.027269  0.015145  0.030224  0.028404  0.016324  0.032355  no_efectores
```

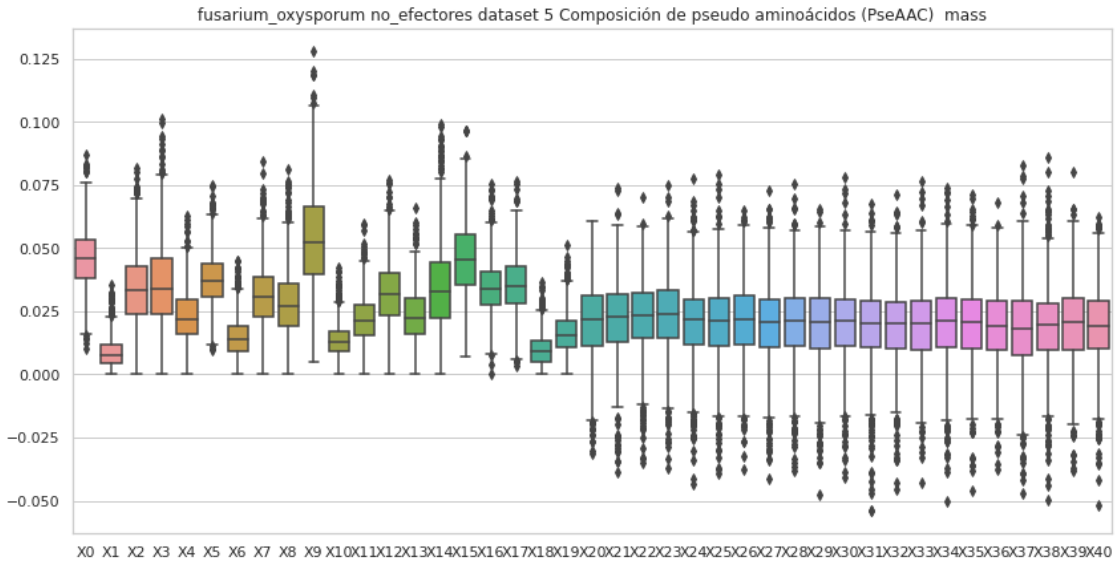
```
1    0.027244  0.015085  0.020630  0.019314  0.026281  0.029401  no_efectores
2    0.021999  0.032867 -0.004361 -0.001743  0.027824  0.023923  no_efectores
3    0.005130  0.014375  0.028381  0.007149  0.026928  0.006090  no_efectores
4    0.041439  0.028461  0.022267  0.020121  0.020535  0.034421  no_efectores
..        …         …         …         …         …         …
993  0.023924  0.026576  0.018196  0.014099 -0.001599  0.018955  no_efectores
994  0.035120 -0.003490  0.003087  0.010537  0.022054  0.019182  no_efectores
995  0.026578  0.023057  0.036258  0.020733  0.024295  0.032372  no_efectores
997  0.006997  0.000893  0.022500  0.028506  0.035775  0.009342  no_efectores
998  0.000382 -0.000968  0.004134  0.038329  0.012255  0.010754  no_efectores

[852 rows x 42 columns]
```

Composición de pseudo aminoácidos (PseAAC)  mass no_efectores fusarium_oxysporum
dataset 5, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  852.000000  852.000000  852.000000  852.000000  852.000000  852.000000
mean     0.045952    0.008847    0.034010    0.035833    0.022981    0.037543
std      0.011555    0.006225    0.014396    0.016863    0.010743    0.010292
min      0.010340    0.000000    0.000000    0.000000    0.000000    0.009778
25%      0.038190    0.004499    0.023636    0.023734    0.015901    0.030805
50%      0.045878    0.007560    0.033317    0.033735    0.021624    0.037270
75%      0.053405    0.011954    0.042706    0.045911    0.029707    0.043824
max      0.087142    0.035458    0.081721    0.101038    0.062970    0.074697

                X6          X7          X8          X9  …         X31  \
count  852.000000  852.000000  852.000000  852.000000  …  852.000000
mean     0.014891    0.031088    0.028379    0.053231  …    0.019009
std      0.008047    0.012819    0.013582    0.020261  …    0.016119
min      0.000000    0.000000    0.000000    0.005040  …   -0.053866
25%      0.009145    0.022839    0.019326    0.039463  …    0.010980
50%      0.014062    0.030639    0.027174    0.052257  …    0.020410
75%      0.019128    0.038520    0.036113    0.066337  …    0.029380
max      0.045038    0.084427    0.081187    0.127854  …    0.067424

               X32         X33         X34         X35         X36         X37  \
count  852.000000  852.000000  852.000000  852.000000  852.000000  852.000000
mean     0.019117    0.019437    0.019824    0.019854    0.018841    0.018095
std      0.015234    0.015056    0.016152    0.015785    0.014839    0.016882
min     -0.045408   -0.043021   -0.050033   -0.046078   -0.037578   -0.047248
25%      0.010204    0.009900    0.010587    0.010359    0.009682    0.007796
50%      0.020217    0.020130    0.021418    0.020715    0.019413    0.018278
75%      0.028639    0.029299    0.030376    0.029842    0.028987    0.029108
max      0.071147    0.076434    0.073793    0.071357    0.068000    0.082974
```

```
              X38         X39         X40
count  852.000000  852.000000  852.000000
mean     0.018790    0.019302    0.018797
std      0.016257    0.016384    0.015367
min     -0.049620   -0.037901   -0.051989
25%      0.009959    0.009437    0.010079
50%      0.019647    0.020686    0.019292
75%      0.027942    0.030044    0.028885
max      0.085869    0.079930    0.062243

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

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.032910  0.000000  0.051716  0.045448  0.029776  0.043881  0.006269
1    0.028259  0.002457  0.006143  0.017201  0.002457  0.011058  0.009829
2    0.050382  0.005038  0.068016  0.060459  0.030229  0.047863  0.015115
3    0.159915  0.031983  0.095949  0.079958  0.111941  0.047975  0.047975
4    0.051788  0.001726  0.037978  0.020715  0.013810  0.039704  0.006905
..        …         …         …         …         …         …         …
995  0.060968  0.008129  0.040646  0.065033  0.032516  0.085356  0.028452
996  0.067678  0.015533  0.046598  0.042160  0.029956  0.037722  0.033284
997  0.022407  0.008963  0.029578  0.045711  0.012548  0.048400  0.007170
998  0.051524  0.010569  0.043597  0.042276  0.036992  0.030386  0.019817
999  0.051450  0.002940  0.027930  0.038220  0.017640  0.052920  0.007350

           X7        X8        X9  …       X53       X54       X55  \
0    0.020373  0.034478  0.040746  …  0.014745  0.023600  0.056038
1    0.012287  0.024573  0.024573  …  0.030830  0.046088  0.041822
2    0.057940  0.040306  0.070535  …  0.019543 -0.018370  0.014500
3    0.063966  0.095949  0.191898  …  0.067746  0.051143  0.044064
4    0.020715  0.027620  0.025894  …  0.025087  0.003461  0.014569
..        …         …         …  …  …         …         …         …
995  0.028452  0.044710  0.069098  … -0.000022  0.014805  0.028685
996  0.031065  0.021080  0.086539  …  0.012343 -0.003785  0.006773
997  0.015237  0.031370  0.023304  …  0.014978  0.015601  0.041312
998  0.025102  0.023780  0.085874  …  0.010059 -0.023035 -0.004836
999  0.020580  0.013230  0.039690  …  0.007728  0.019065  0.015743

          X56       X57       X58       X59       X60       X61         X62
0    0.010337  0.030937 -0.009097  0.029670  0.002524  0.013419  efectores
1    0.013436  0.017800  0.003052  0.004444  0.005715  0.009217  efectores
2    0.004584  0.018455  0.012506  0.031248 -0.020177  0.026816  efectores
3    0.089020  0.077241 -0.041685 -0.125427 -0.007686 -0.051435  efectores
4   -0.014749 -0.003716  0.019548  0.035385 -0.021782  0.010012  efectores
..        …         …         …         …         …         …         …
995 -0.013860 -0.010710 -0.002969 -0.036127  0.039951  0.033711  efectores
996 -0.014481  0.009636  0.014668 -0.003578  0.000598 -0.004131  efectores
997  0.001314  0.011439  0.001118  0.026369  0.004644  0.022843  efectores
```

```
998 -0.015408 -0.017045 -0.033166 -0.000237  0.015537  0.007743  efectores
999  0.004691  0.005607  0.017730  0.020110  0.000779  0.004136  efectores

[1000 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC)  hidro efectores fusarium_oxysporum dataset 5, con valores atípicos.
Estadísticas.

| | X0 | X1 | X2 | X3 | X4 | \ |
|---|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | |
| mean | 0.061783 | 0.012836 | 0.041339 | 0.043823 | 0.028947 | |
| std | 0.078303 | 0.032183 | 0.022943 | 0.037848 | 0.023684 | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | |
| 25% | 0.037634 | 0.004225 | 0.024720 | 0.027355 | 0.016551 | |
| 50% | 0.054781 | 0.008514 | 0.040636 | 0.041038 | 0.025900 | |
| 75% | 0.074746 | 0.015238 | 0.053846 | 0.054662 | 0.036463 | |
| max | 2.331874 | 0.932750 | 0.267031 | 0.932750 | 0.466375 | |

| | X5 | X6 | X7 | X8 | X9 | … | \ |
|---|---|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | … | |
| mean | 0.050546 | 0.019147 | 0.039228 | 0.037395 | 0.066361 | … | |
| std | 0.042706 | 0.035737 | 0.039872 | 0.038045 | 0.067601 | … | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.002381 | … | |
| 25% | 0.029504 | 0.009046 | 0.023076 | 0.021706 | 0.040646 | … | |
| 50% | 0.043468 | 0.016375 | 0.035121 | 0.033234 | 0.058937 | … | |
| 75% | 0.064581 | 0.024038 | 0.048560 | 0.047273 | 0.083075 | … | |
| max | 0.932750 | 1.068124 | 0.932750 | 0.932750 | 1.865500 | … | |

| | X52 | X53 | X54 | X55 | X56 | \ |
|---|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | |
| mean | 0.001189 | 0.012250 | 0.005744 | 0.009451 | 0.003518 | |
| std | 0.070626 | 0.040365 | 0.062811 | 0.042310 | 0.040837 | |
| min | -1.901326 | -0.237073 | -0.314784 | -0.629762 | -0.300525 | |
| 25% | -0.008847 | -0.000757 | -0.008754 | -0.001721 | -0.009928 | |
| 50% | 0.004777 | 0.010779 | 0.005584 | 0.010371 | 0.005107 | |
| 75% | 0.017253 | 0.024865 | 0.017046 | 0.022658 | 0.016648 | |
| max | 0.310876 | 0.972336 | 1.600242 | 0.423249 | 0.634255 | |

| | X57 | X58 | X59 | X60 | X61 |
|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 0.007444 | 0.002688 | 0.008081 | -0.002816 | 0.007538 |
| std | 0.031057 | 0.058733 | 0.032787 | 0.118773 | 0.049323 |
| min | -0.339894 | -0.849018 | -0.311214 | -3.421324 | -0.840420 |
| 25% | -0.002891 | -0.009101 | -0.002011 | -0.009918 | -0.004047 |
| 50% | 0.009559 | 0.004603 | 0.008267 | 0.004604 | 0.010046 |
| 75% | 0.021438 | 0.015702 | 0.021287 | 0.017162 | 0.022850 |

```
max       0.233095     1.255683     0.470215     0.204206     0.470193

[8 rows x 62 columns]
```

no_efectores
Composición de pseudo aminoácidos (PseAAC)  hidro no_efectores
fusarium_oxysporum dataset 5, con valores atípicos.
Valores del documento csv.

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 | \ |
|---|---|---|---|---|---|---|---|---|
| 0 | 0.048657 | 0.014971 | 0.052400 | 0.028071 | 0.029943 | 0.065500 | 0.018714 | |
| 1 | 0.033118 | 0.015285 | 0.034392 | 0.028023 | 0.014012 | 0.040761 | 0.025475 | |
| 2 | 0.044344 | 0.019005 | 0.063349 | 0.063349 | 0.040121 | 0.048567 | 0.029563 | |
| 3 | 0.043330 | 0.006103 | 0.035396 | 0.037837 | 0.020749 | 0.037837 | 0.013426 | |
| 4 | 0.033500 | 0.005776 | 0.006931 | 0.011552 | 0.020793 | 0.038121 | 0.004621 | |
| .. | … | … | … | … | … | … | … | |
| 995 | 0.051635 | 0.009388 | 0.037553 | 0.039900 | 0.021124 | 0.061023 | 0.025818 | |
| 996 | 0.050003 | 0.000000 | 0.033335 | 0.050003 | 0.025001 | 0.025001 | 0.016668 | |
| 997 | 0.055686 | 0.019654 | 0.068789 | 0.052411 | 0.039308 | 0.049135 | 0.016378 | |
| 998 | 0.065723 | 0.010954 | 0.045641 | 0.043816 | 0.032862 | 0.032862 | 0.010954 | |
| 999 | 0.047417 | 0.012932 | 0.047417 | 0.064660 | 0.038796 | 0.034485 | 0.008621 | |

| | X7 | X8 | X9 | … | X53 | X54 | X55 | \ |
|---|---|---|---|---|---|---|---|---|
| 0 | 0.069243 | 0.035557 | 0.076729 | … | 0.006923 | -0.000764 | -0.007822 | |
| 1 | 0.015285 | 0.024202 | 0.043308 | … | 0.012229 | 0.016405 | 0.023096 | |
| 2 | 0.057014 | 0.065461 | 0.107693 | … | 0.075279 | 0.039275 | 0.043004 | |
| 3 | 0.022580 | 0.020749 | 0.047602 | … | 0.012302 | 0.009089 | 0.004747 | |
| 4 | 0.031190 | 0.009819 | 0.033500 | … | -0.003395 | 0.013944 | 0.007978 | |
| .. | … | … | … | … | … | … | … | |
| 995 | 0.051635 | 0.049288 | 0.046941 | … | 0.011587 | -0.009759 | -0.004235 | |
| 996 | 0.025001 | 0.016668 | 0.033335 | … | 0.022133 | 0.056631 | 0.067737 | |
| 997 | 0.045859 | 0.036032 | 0.072064 | … | -0.056095 | 0.012119 | 0.028869 | |
| 998 | 0.052944 | 0.038339 | 0.056595 | … | 0.019977 | 0.018537 | 0.006629 | |
| 999 | 0.021553 | 0.021553 | 0.068971 | … | -0.050677 | 0.036788 | 0.052857 | |

| | X56 | X57 | X58 | X59 | X60 | X61 | X62 |
|---|---|---|---|---|---|---|---|
| 0 | 0.017682 | -0.003248 | -0.003121 | -0.010840 | 0.035047 | 0.026254 | no_efectores |
| 1 | 0.031238 | 0.023629 | 0.010212 | 0.009087 | 0.009766 | 0.017777 | no_efectores |
| 2 | -0.038693 | -0.016716 | 0.004559 | -0.010596 | -0.022788 | 0.001549 | no_efectores |
| 3 | 0.006137 | 0.010849 | 0.016144 | 0.011119 | -0.004488 | 0.001358 | no_efectores |
| 4 | 0.006346 | 0.000124 | 0.025223 | 0.007461 | 0.019208 | 0.003794 | no_efectores |
| .. | … | … | … | … | … | … | … |
| 995 | 0.003790 | 0.017754 | 0.007398 | 0.003468 | -0.012158 | -0.005403 | no_efectores |
| 996 | 0.053205 | 0.041921 | 0.109221 | 0.078842 | 0.053415 | 0.059669 | no_efectores |
| 997 | 0.016592 | 0.031379 | -0.044310 | -0.017754 | 0.053166 | 0.065989 | no_efectores |
| 998 | 0.023613 | 0.040404 | 0.021271 | 0.016245 | 0.022636 | 0.029040 | no_efectores |

```
999  0.023107  0.021086 -0.021617 -0.038198  0.026888  0.045292  no_efectores

[1000 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC)  hidro no_efectores
fusarium_oxysporum dataset 5, con valores atípicos.
Estadísticas.

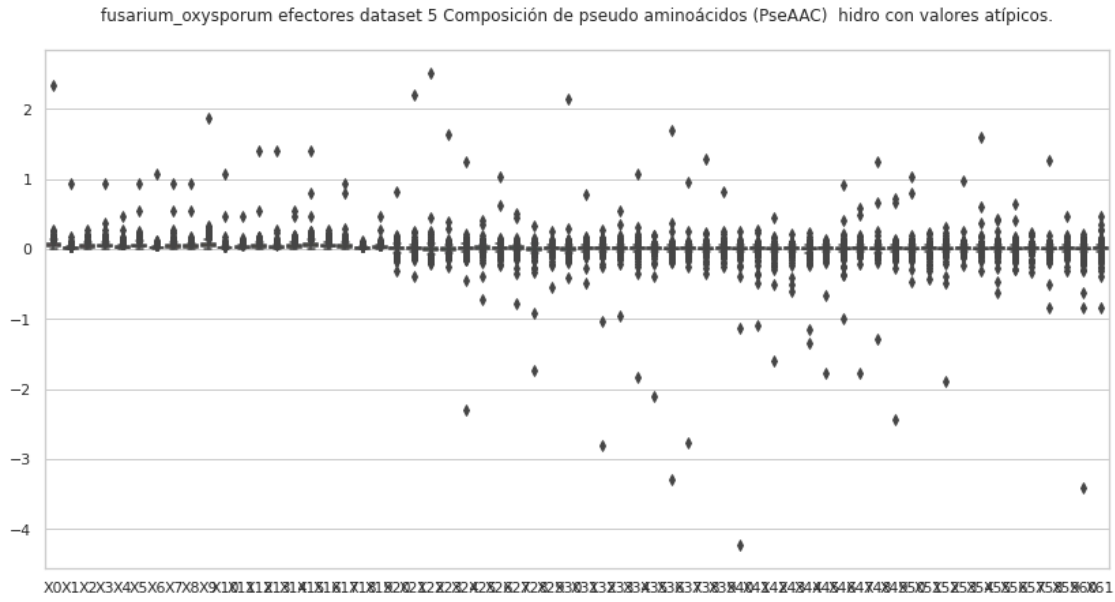
| | X0 | X1 | X2 | X3 | X4 | \ |
|---|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | |
| mean | 0.059963 | 0.012088 | 0.042389 | 0.044218 | 0.029653 | |
| std | 0.034471 | 0.011491 | 0.024373 | 0.022350 | 0.021595 | |
| min | 0.003942 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | |
| 25% | 0.039073 | 0.004620 | 0.026809 | 0.028200 | 0.018030 | |
| 50% | 0.054329 | 0.009347 | 0.040153 | 0.042879 | 0.027054 | |
| 75% | 0.075020 | 0.016740 | 0.054840 | 0.056603 | 0.038010 | |
| max | 0.628691 | 0.117699 | 0.419128 | 0.181226 | 0.419128 | |

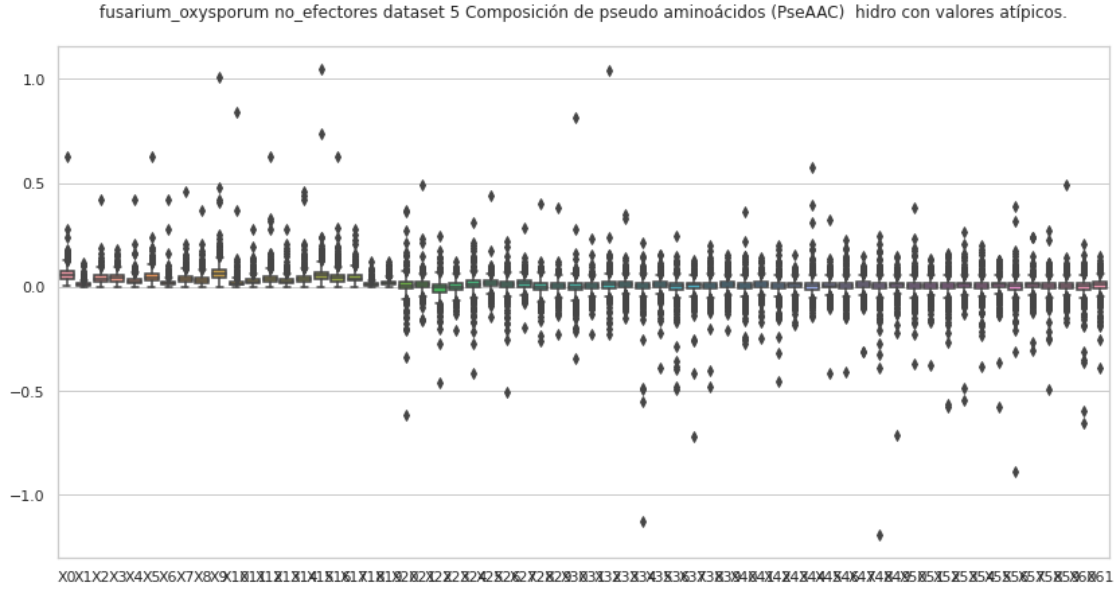
| | X5 | X6 | X7 | X8 | X9 | … | \ |
|---|---|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | … | |
| mean | 0.049698 | 0.019853 | 0.040237 | 0.036361 | 0.068756 | … | |
| std | 0.031920 | 0.020314 | 0.025942 | 0.023410 | 0.048208 | … | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | … | |
| 25% | 0.030565 | 0.010008 | 0.025067 | 0.021160 | 0.044069 | … | |
| 50% | 0.044265 | 0.017196 | 0.036932 | 0.033694 | 0.063760 | … | |
| 75% | 0.063990 | 0.025101 | 0.050461 | 0.046371 | 0.084496 | … | |
| max | 0.628691 | 0.419128 | 0.459964 | 0.367971 | 1.011921 | … | |

| | X52 | X53 | X54 | X55 | X56 | \ |
|---|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | |
| mean | 0.002400 | 0.007837 | 0.003496 | 0.008048 | 0.001732 | |
| std | 0.042026 | 0.036617 | 0.035687 | 0.034189 | 0.046482 | |
| min | -0.579804 | -0.542645 | -0.383110 | -0.577651 | -0.891041 | |
| 25% | -0.009125 | -0.003889 | -0.008281 | -0.001520 | -0.010881 | |
| 50% | 0.005608 | 0.009599 | 0.004175 | 0.009430 | 0.004540 | |
| 75% | 0.018148 | 0.020857 | 0.018098 | 0.021866 | 0.016840 | |
| max | 0.160282 | 0.262853 | 0.201452 | 0.129928 | 0.387803 | |

| | X57 | X58 | X59 | X60 | X61 |
|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 0.007612 | 0.002245 | 0.007524 | 0.001310 | 0.007753 |
| std | 0.030607 | 0.036169 | 0.030155 | 0.045534 | 0.032901 |
| min | -0.304846 | -0.492963 | -0.168028 | -0.656478 | -0.389233 |
| 25% | -0.003902 | -0.010243 | -0.004312 | -0.010915 | -0.004029 |
| 50% | 0.009248 | 0.005565 | 0.008020 | 0.004265 | 0.009373 |
| 75% | 0.020526 | 0.017698 | 0.021165 | 0.018121 | 0.022888 |
| max | 0.241750 | 0.268220 | 0.493229 | 0.204875 | 0.149545 |

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

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 | \ |
|---|---|---|---|---|---|---|---|---|
| 0 | 0.032910 | 0.000000 | 0.051716 | 0.045448 | 0.029776 | 0.043881 | 0.006269 | |
| 1 | 0.028259 | 0.002457 | 0.006143 | 0.017201 | 0.002457 | 0.011058 | 0.009829 | |
| 2 | 0.050382 | 0.005038 | 0.068016 | 0.060459 | 0.030229 | 0.047863 | 0.015115 | |
| 4 | 0.051788 | 0.001726 | 0.037978 | 0.020715 | 0.013810 | 0.039704 | 0.006905 | |
| 5 | 0.066116 | 0.013722 | 0.034929 | 0.064869 | 0.026197 | 0.069859 | 0.024950 | |
| .. | … | … | … | … | … | … | … | |
| 995 | 0.060968 | 0.008129 | 0.040646 | 0.065033 | 0.032516 | 0.085356 | 0.028452 | |
| 996 | 0.067678 | 0.015533 | 0.046598 | 0.042160 | 0.029956 | 0.037722 | 0.033284 | |
| 997 | 0.022407 | 0.008963 | 0.029578 | 0.045711 | 0.012548 | 0.048400 | 0.007170 | |
| 998 | 0.051524 | 0.010569 | 0.043597 | 0.042276 | 0.036992 | 0.030386 | 0.019817 | |
| 999 | 0.051450 | 0.002940 | 0.027930 | 0.038220 | 0.017640 | 0.052920 | 0.007350 | |

| | X7 | X8 | X9 | … | X53 | X54 | X55 | \ |
|---|---|---|---|---|---|---|---|---|
| 0 | 0.020373 | 0.034478 | 0.040746 | … | 0.014745 | 0.023600 | 0.056038 | |
| 1 | 0.012287 | 0.024573 | 0.024573 | … | 0.030830 | 0.046088 | 0.041822 | |
| 2 | 0.057940 | 0.040306 | 0.070535 | … | 0.019543 | -0.018370 | 0.014500 | |
| 4 | 0.020715 | 0.027620 | 0.025894 | … | 0.025087 | 0.003461 | 0.014569 | |
| 5 | 0.049899 | 0.064869 | 0.081086 | … | 0.019029 | 0.019864 | 0.009258 | |
| .. | … | … | … | … | … | … | … | |
| 995 | 0.028452 | 0.044710 | 0.069098 | … | -0.000022 | 0.014805 | 0.028685 | |
| 996 | 0.031065 | 0.021080 | 0.086539 | … | 0.012343 | -0.003785 | 0.006773 | |
| 997 | 0.015237 | 0.031370 | 0.023304 | … | 0.014978 | 0.015601 | 0.041312 | |
| 998 | 0.025102 | 0.023780 | 0.085874 | … | 0.010059 | -0.023035 | -0.004836 | |
| 999 | 0.020580 | 0.013230 | 0.039690 | … | 0.007728 | 0.019065 | 0.015743 | |

| | X56 | X57 | X58 | X59 | X60 | X61 | X62 |
|---|---|---|---|---|---|---|---|
| 0 | 0.010337 | 0.030937 | -0.009097 | 0.029670 | 0.002524 | 0.013419 | efectores |
| 1 | 0.013436 | 0.017800 | 0.003052 | 0.004444 | 0.005715 | 0.009217 | efectores |
| 2 | 0.004584 | 0.018455 | 0.012506 | 0.031248 | -0.020177 | 0.026816 | efectores |
| 4 | -0.014749 | -0.003716 | 0.019548 | 0.035385 | -0.021782 | 0.010012 | efectores |
| 5 | -0.000324 | -0.008659 | 0.004204 | 0.006405 | -0.009742 | -0.001867 | efectores |
| .. | … | … | … | … | … | … | … |
| 995 | -0.013860 | -0.010710 | -0.002969 | -0.036127 | 0.039951 | 0.033711 | efectores |
| 996 | -0.014481 | 0.009636 | 0.014668 | -0.003578 | 0.000598 | -0.004131 | efectores |
| 997 | 0.001314 | 0.011439 | 0.001118 | 0.026369 | 0.004644 | 0.022843 | efectores |
| 998 | -0.015408 | -0.017045 | -0.033166 | -0.000237 | 0.015537 | 0.007743 | efectores |
| 999 | 0.004691 | 0.005607 | 0.017730 | 0.020110 | 0.000779 | 0.004136 | efectores |

```
[943 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC)  efectores fusarium_oxysporum dataset 5, sin valores atípicos.
Estadísticas.

| | X0 | X1 | X2 | X3 | X4 | X5 | \ |
|---|---|---|---|---|---|---|---|
| count | 943.000000 | 943.000000 | 943.000000 | 943.000000 | 943.000000 | 943.000000 | |
| mean | 0.056963 | 0.010698 | 0.039866 | 0.040779 | 0.026916 | 0.046798 | |
| std | 0.026333 | 0.010249 | 0.018997 | 0.018616 | 0.014702 | 0.024140 | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | |
| 25% | 0.037142 | 0.004216 | 0.024221 | 0.027045 | 0.016292 | 0.029058 | |
| 50% | 0.053343 | 0.008199 | 0.040456 | 0.040208 | 0.025231 | 0.042137 | |
| 75% | 0.071863 | 0.014694 | 0.052868 | 0.053572 | 0.035479 | 0.061970 | |
| max | 0.180653 | 0.103577 | 0.104204 | 0.138641 | 0.091156 | 0.146799 | |

| | X6 | X7 | X8 | X9 | … | X52 | \ |
|---|---|---|---|---|---|---|---|
| count | 943.000000 | 943.000000 | 943.000000 | 943.000000 | … | 943.000000 | |
| mean | 0.017152 | 0.035587 | 0.034866 | 0.061357 | … | 0.003880 | |
| std | 0.010927 | 0.017590 | 0.018062 | 0.029596 | … | 0.023653 | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.002381 | … | -0.105826 | |
| 25% | 0.008884 | 0.022813 | 0.021356 | 0.040211 | … | -0.007843 | |
| 50% | 0.016039 | 0.034127 | 0.032409 | 0.057942 | … | 0.004780 | |
| 75% | 0.023283 | 0.046534 | 0.045619 | 0.080737 | … | 0.016583 | |
| max | 0.099478 | 0.106110 | 0.104342 | 0.193094 | … | 0.124707 | |

| | X53 | X54 | X55 | X56 | X57 | X58 | \ |
|---|---|---|---|---|---|---|---|
| count | 943.000000 | 943.000000 | 943.000000 | 943.000000 | 943.000000 | 943.000000 | |
| mean | 0.011650 | 0.004077 | 0.010535 | 0.003560 | 0.008408 | 0.003365 | |
| std | 0.020610 | 0.023497 | 0.020628 | 0.024713 | 0.020861 | 0.024190 | |
| min | -0.067709 | -0.099990 | -0.110345 | -0.106100 | -0.067368 | -0.115870 | |
| 25% | -0.000127 | -0.007899 | -0.000363 | -0.009053 | -0.001955 | -0.007846 | |
| 50% | 0.011079 | 0.005864 | 0.010570 | 0.005161 | 0.010021 | 0.004662 | |
| 75% | 0.024118 | 0.016707 | 0.022438 | 0.016483 | 0.021339 | 0.015370 | |
| max | 0.115794 | 0.094648 | 0.107306 | 0.097833 | 0.066124 | 0.099536 | |

| | X59 | X60 | X61 |
|---|---|---|---|
| count | 943.000000 | 943.000000 | 943.000000 |
| mean | 0.008617 | 0.004349 | 0.009679 |
| std | 0.020324 | 0.024649 | 0.022153 |
| min | -0.071628 | -0.146227 | -0.096940 |
| 25% | -0.001297 | -0.008721 | -0.002505 |
| 50% | 0.008552 | 0.004838 | 0.010151 |
| 75% | 0.020841 | 0.017044 | 0.022691 |
| max | 0.098757 | 0.204206 | 0.149051 |

[8 rows x 62 columns]

no_efectores
Composición de pseudo aminoácidos (PseAAC)  no_efectores fusarium_oxysporum
dataset 5, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.048657  0.014971  0.052400  0.028071  0.029943  0.065500  0.018714
1    0.033118  0.015285  0.034392  0.028023  0.014012  0.040761  0.025475
2    0.044344  0.019005  0.063349  0.063349  0.040121  0.048567  0.029563
3    0.043330  0.006103  0.035396  0.037837  0.020749  0.037837  0.013426
4    0.033500  0.005776  0.006931  0.011552  0.020793  0.038121  0.004621
..        …         …         …         …         …         …         …
994  0.065656  0.032828  0.058520  0.042819  0.031401  0.035683  0.034256
995  0.051635  0.009388  0.037553  0.039900  0.021124  0.061023  0.025818
997  0.055686  0.019654  0.068789  0.052411  0.039308  0.049135  0.016378
998  0.065723  0.010954  0.045641  0.043816  0.032862  0.032862  0.010954
999  0.047417  0.012932  0.047417  0.064660  0.038796  0.034485  0.008621

           X7        X8        X9  …       X53       X54       X55  \
0    0.069243  0.035557  0.076729  …  0.006923 -0.000764 -0.007822
1    0.015285  0.024202  0.043308  …  0.012229  0.016405  0.023096
2    0.057014  0.065461  0.107693  …  0.075279  0.039275  0.043004
3    0.022580  0.020749  0.047602  …  0.012302  0.009089  0.004747
4    0.031190  0.009819  0.033500  … -0.003395  0.013944  0.007978
..        …         …         …  …         …         …         …
994  0.057093  0.029974  0.095630  … -0.000436 -0.001242  0.000160
995  0.051635  0.049288  0.046941  …  0.011587 -0.009759 -0.004235
997  0.045859  0.036032  0.072064  … -0.056095  0.012119  0.028869
998  0.052944  0.038339  0.056595  …  0.019977  0.018537  0.006629
999  0.021553  0.021553  0.068971  … -0.050677  0.036788  0.052857

          X56       X57       X58       X59       X60       X61           X62
0    0.017682 -0.003248 -0.003121 -0.010840  0.035047  0.026254  no_efectores
1    0.031238  0.023629  0.010212  0.009087  0.009766  0.017777  no_efectores
2   -0.038693 -0.016716  0.004559 -0.010596 -0.022788  0.001549  no_efectores
3    0.006137  0.010849  0.016144  0.011119 -0.004488  0.001358  no_efectores
4    0.006346  0.000124  0.025223  0.007461  0.019208  0.003794  no_efectores
..        …         …         …         …         …         …           …
994  0.024006  0.012738 -0.006487 -0.011545  0.044737  0.008816  no_efectores
995  0.003790  0.017754  0.007398  0.003468 -0.012158 -0.005403  no_efectores
997  0.016592  0.031379 -0.044310 -0.017754  0.053166  0.065989  no_efectores
998  0.023613  0.040404  0.021271  0.016245  0.022636  0.029040  no_efectores
999  0.023107  0.021086 -0.021617 -0.038198  0.026888  0.045292  no_efectores

[874 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC)  no_efectores fusarium_oxysporum
dataset 5, sin valores atípicos.
Estadísticas.

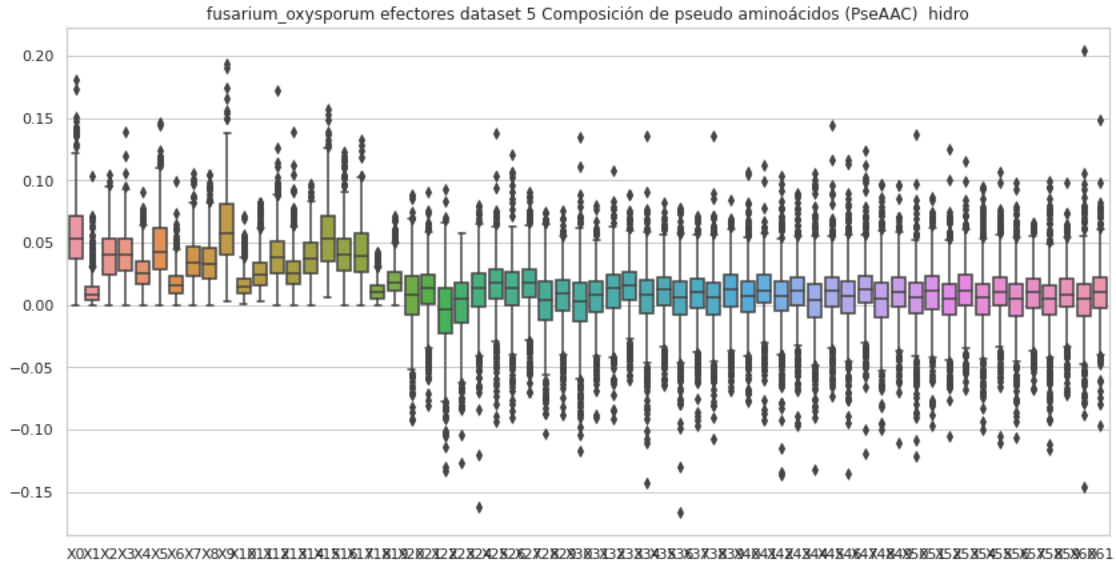
| | X0 | X1 | X2 | X3 | X4 | X5 | \ |
|---|---|---|---|---|---|---|---|
| count | 874.000000 | 874.000000 | 874.000000 | 874.000000 | 874.000000 | 874.000000 | |
| mean | 0.054923 | 0.010469 | 0.039857 | 0.041170 | 0.027156 | 0.045488 | |
| std | 0.023763 | 0.007692 | 0.018364 | 0.017815 | 0.013616 | 0.021417 | |
| min | 0.003942 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.004181 | |
| 25% | 0.037739 | 0.004605 | 0.025856 | 0.027175 | 0.017697 | 0.029446 | |
| 50% | 0.051958 | 0.008818 | 0.038568 | 0.041159 | 0.026092 | 0.041307 | |
| 75% | 0.068816 | 0.015230 | 0.052714 | 0.054159 | 0.035912 | 0.059463 | |
| max | 0.160373 | 0.044495 | 0.107390 | 0.092981 | 0.092083 | 0.129081 | |

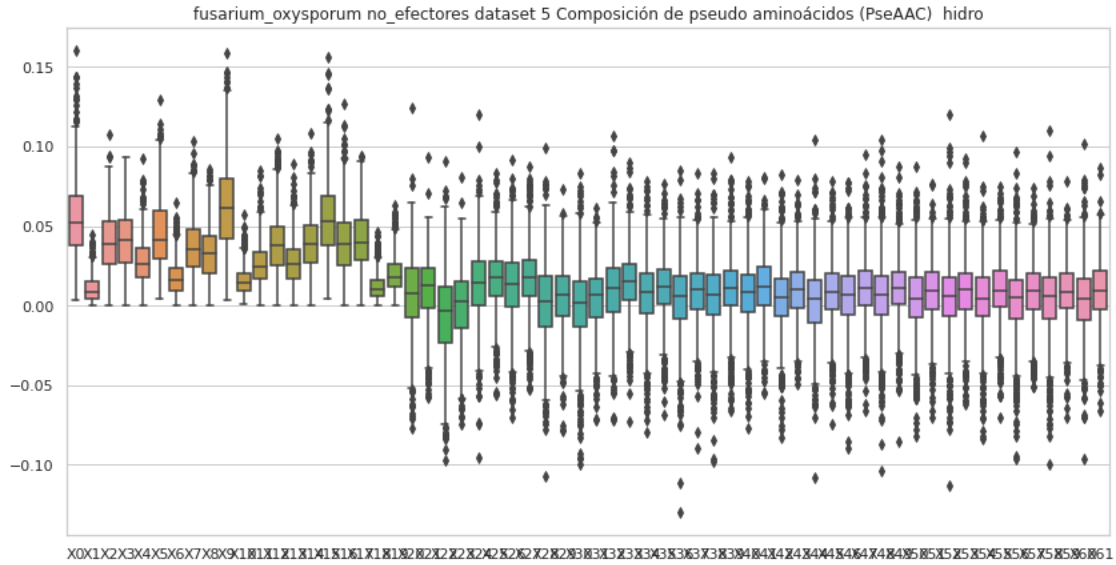
| | X6 | X7 | X8 | X9 | … | X52 | \ |
|---|---|---|---|---|---|---|---|
| count | 874.000000 | 874.000000 | 874.000000 | 874.000000 | … | 874.000000 | |
| mean | 0.017875 | 0.037039 | 0.033500 | 0.062746 | … | 0.005287 | |
| std | 0.010401 | 0.017237 | 0.016022 | 0.027203 | … | 0.022711 | |
| min | 0.000000 | 0.000000 | 0.000000 | 0.003387 | … | -0.112714 | |
| 25% | 0.009839 | 0.024579 | 0.020759 | 0.042395 | … | -0.006386 | |
| 50% | 0.016524 | 0.035702 | 0.032626 | 0.061276 | … | 0.006253 | |
| 75% | 0.023823 | 0.048019 | 0.043873 | 0.079688 | … | 0.018040 | |
| max | 0.065150 | 0.103693 | 0.085826 | 0.159123 | … | 0.120066 | |

| | X53 | X54 | X55 | X56 | X57 | X58 | \ |
|---|---|---|---|---|---|---|---|
| count | 874.000000 | 874.000000 | 874.000000 | 874.000000 | 874.000000 | 874.000000 | |
| mean | 0.009373 | 0.004705 | 0.009906 | 0.002921 | 0.008707 | 0.004305 | |
| std | 0.019863 | 0.022484 | 0.019125 | 0.022900 | 0.018715 | 0.022471 | |
| min | -0.062031 | -0.083480 | -0.070374 | -0.096452 | -0.051053 | -0.099241 | |
| 25% | -0.002197 | -0.006458 | -0.000262 | -0.008300 | -0.001975 | -0.007754 | |
| 50% | 0.010305 | 0.004790 | 0.009529 | 0.005275 | 0.009633 | 0.005941 | |
| 75% | 0.020269 | 0.017632 | 0.021664 | 0.016221 | 0.020157 | 0.017467 | |
| max | 0.092689 | 0.107006 | 0.074963 | 0.097102 | 0.073975 | 0.110322 | |

| | X59 | X60 | X61 |
|---|---|---|---|
| count | 874.000000 | 874.000000 | 874.000000 |
| mean | 0.008435 | 0.004482 | 0.009459 |
| std | 0.018415 | 0.021422 | 0.019348 |
| min | -0.070111 | -0.096177 | -0.066017 |
| 25% | -0.001828 | -0.008715 | -0.001973 |
| 50% | 0.008447 | 0.004754 | 0.009660 |
| 75% | 0.020768 | 0.017233 | 0.021641 |
| max | 0.078264 | 0.101594 | 0.086456 |

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.009715  0.000520 -0.076166 -0.044523  0.031587 -0.110099  0.053749
1   -0.047070  0.066631 -0.134354  0.069049 -0.102367  0.059562 -0.128934
2   -0.024975  0.019980  0.029167  0.084420 -0.024428  0.041260 -0.014064
3   -0.020038  0.047527  0.015308 -0.067246 -0.020417 -0.015069 -0.123033
4    0.066602 -0.003272 -0.007416  0.064446 -0.010361  0.054844  0.091152
..        …         …         …         …         …         …         …
995 -0.146291 -0.063328  0.071222  0.079739 -0.026965  0.066375 -0.052653
996  0.003814  0.053309 -0.015354  0.034205  0.072789  0.040366 -0.002819
997  0.012738  0.116541 -0.010380  0.022918 -0.040516  0.029206  0.062672
998  0.018288  0.058056 -0.024623  0.047567 -0.073932  0.046799  0.035913
999  0.046895  0.076009  0.101244 -0.041055  0.058772  0.057384 -0.015844

           X7        X8        X9       X10       X11       X12        X13
0    0.002173 -0.044896 -0.026252  0.025089  0.059151  0.031293  efectores
```

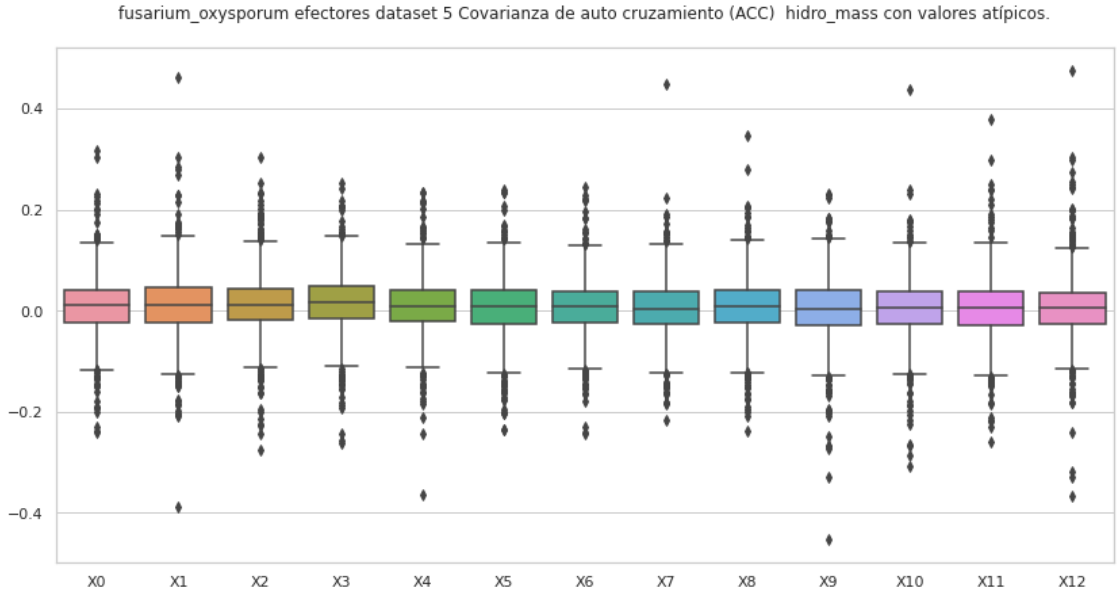
```
1    0.097181 -0.208400  0.071286 -0.068116 -0.094719 -0.169502  efectores
2    0.059918  0.114560  0.046208 -0.017126  0.001618  0.031668  efectores
3    0.000316 -0.079621 -0.209254 -0.042398  0.060060 -0.004191  efectores
4    0.001117 -0.083297  0.064339  0.066756  0.078989  0.010679  efectores
..        …        …        …        …        …        …
995 -0.026309 -0.047577 -0.080583  0.080542 -0.037148  0.021763  efectores
996 -0.007433  0.040178  0.009606  0.023677  0.029601  0.012132  efectores
997  0.027105  0.015319  0.045230 -0.002773 -0.021141 -0.064073  efectores
998  0.009694  0.020795  0.018241 -0.011883 -0.017278 -0.034690  efectores
999  0.006180  0.082226  0.068744 -0.010000  0.106379 -0.008417  efectores

[1000 rows x 14 columns]
```

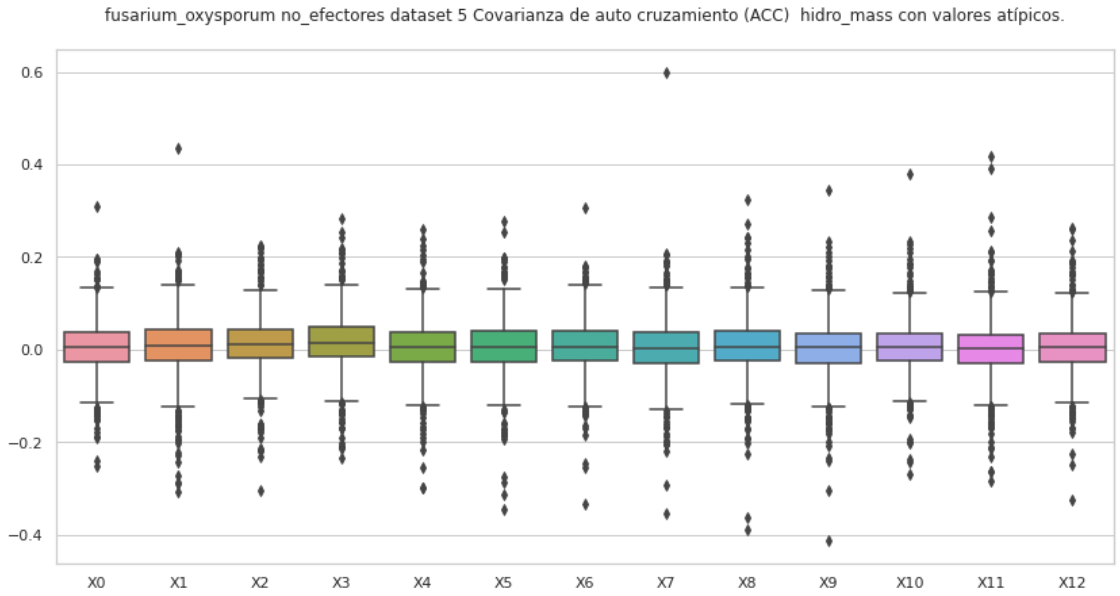
Covarianza de auto cruzamiento (ACC)  hidro_mass efectores fusarium_oxysporum
dataset 5, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.008786     0.011810     0.012648     0.016101     0.008549
std       0.058600     0.063333     0.060473     0.058645     0.056455
min      -0.240910    -0.389017    -0.274770    -0.261354    -0.365513
25%      -0.022646    -0.022851    -0.019310    -0.015646    -0.021172
50%       0.010425     0.011706     0.011678     0.017790     0.009448
75%       0.040294     0.046707     0.044214     0.049634     0.041007
max       0.316845     0.462082     0.303677     0.252158     0.233359

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.006040     0.007563     0.003365     0.007956     0.002759
std       0.058601     0.056442     0.055493     0.056929     0.061549
min      -0.236181    -0.244787    -0.215459    -0.238937    -0.452724
25%      -0.025871    -0.023329    -0.027734    -0.024444    -0.027881
50%       0.007602     0.009069     0.004298     0.007872     0.003140
75%       0.039672     0.038322     0.036672     0.041136     0.040474
max       0.238519     0.243211     0.448019     0.346665     0.231211

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.003280     0.005550     0.006060
std       0.059659     0.059907     0.063594
min      -0.308986    -0.258435    -0.366551
25%      -0.027145    -0.028473    -0.025938
50%       0.005249     0.005063     0.004484
75%       0.037965     0.038892     0.034656
max       0.437510     0.378083     0.474760
```

```
no_efectores
Covarianza de auto cruzamiento (ACC)  hidro_mass no_efectores fusarium_oxysporum
dataset 5, con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.062453  0.023433  0.023638  0.115169  0.068954 -0.006815  0.009393
1    0.058356  0.027170  0.038701  0.011180  0.103074 -0.044269  0.023315
2    0.030536 -0.061034 -0.012768  0.007648  0.000491 -0.006463  0.021106
3   -0.013352  0.017059  0.068466  0.000129  0.040656  0.010577 -0.021191
4    0.017873 -0.018288 -0.027092  0.022807 -0.007655 -0.008843 -0.049344
..        …         …         …         …         …         …
995 -0.006176  0.032783  0.034584 -0.007126 -0.011054  0.021773  0.062234
996  0.067214  0.004278  0.224288  0.171269 -0.017389 -0.168740  0.048729
997 -0.069707  0.035365  0.033013 -0.060010 -0.088631  0.031651  0.053052
998  0.043945  0.014359  0.002858  0.000321 -0.061820 -0.064469 -0.023915
999 -0.078321  0.010490 -0.046935 -0.027604  0.049669 -0.068051  0.041167

           X7        X8        X9       X10       X11       X12           X13
0    0.061936  0.012058  0.011067  0.040215 -0.028771  0.006251  no_efectores
1   -0.040153  0.023114 -0.037008  0.051082  0.027151 -0.069039  no_efectores
2    0.108493  0.040252 -0.037912  0.000665  0.052704  0.026046  no_efectores
3    0.068411  0.042965 -0.048515 -0.030199  0.042293 -0.011806  no_efectores
4    0.007356 -0.046218  0.025928  0.101260  0.148905  0.043595  no_efectores
..        …         …         …         …         …         …
995  0.055535  0.083335  0.056208  0.027111  0.000747  0.093770  no_efectores
996 -0.016098 -0.021882 -0.181469 -0.070267 -0.027160 -0.010133  no_efectores
997 -0.065087  0.101203  0.093995 -0.042853  0.108375  0.030308  no_efectores
998  0.002423 -0.064082 -0.024096  0.032817  0.021130 -0.058060  no_efectores
999  0.127845 -0.070236  0.009437  0.057222 -0.082412  0.010568  no_efectores

[1000 rows x 14 columns]


Covarianza de auto cruzamiento (ACC)  hidro_mass no_efectores fusarium_oxysporum
dataset 5, con valores atípicos.
Estadísticas.

                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.005491     0.007144     0.012143     0.015477     0.004840
std       0.056383     0.062574     0.056868     0.058908     0.057728
min      -0.253343    -0.306694    -0.304586    -0.234911    -0.297840
25%      -0.025528    -0.023227    -0.017709    -0.016160    -0.026657
50%       0.005661     0.009510     0.012650     0.014666     0.005950
75%       0.038723     0.042960     0.042204     0.048194     0.036971
```

```
max       0.309722     0.436423     0.224288     0.284081     0.259404

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.004549     0.006706     0.003558     0.006259     0.002476
std       0.059323     0.054875     0.061360     0.060388     0.060325
min      -0.346435    -0.332872    -0.355538    -0.388539    -0.412724
25%      -0.027320    -0.024537    -0.029300    -0.023618    -0.029839
50%       0.005111     0.006665     0.003538     0.005006     0.004148
75%       0.039696     0.041860     0.036321     0.040628     0.034564
max       0.278656     0.307536     0.598643     0.323006     0.344545

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.007025     0.003103     0.004046
std       0.056716     0.061614     0.057415
min      -0.270820    -0.283420    -0.325193
25%      -0.023316    -0.029313    -0.026939
50%       0.005861     0.002800     0.004081
75%       0.035136     0.032440     0.033568
max       0.380560     0.417414     0.261919
```






### 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 5, sin valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.009715  0.000520 -0.076166 -0.044523  0.031587 -0.110099  0.053749
2   -0.024975  0.019980  0.029167  0.084420 -0.024428  0.041260 -0.014064
4    0.066602 -0.003272 -0.007416  0.064446 -0.010361  0.054844  0.091152
5    0.023988 -0.001712 -0.042395 -0.012470  0.028812 -0.033189 -0.029020
6   -0.015862  0.062257 -0.061294  0.029002 -0.002349 -0.043555  0.023316
..        …         …         …         …         …         …         …
995 -0.146291 -0.063328  0.071222  0.079739 -0.026965  0.066375 -0.052653
996  0.003814  0.053309 -0.015354  0.034205  0.072789  0.040366 -0.002819
997  0.012738  0.116541 -0.010380  0.022918 -0.040516  0.029206  0.062672
998  0.018288  0.058056 -0.024623  0.047567 -0.073932  0.046799  0.035913
999  0.046895  0.076009  0.101244 -0.041055  0.058772  0.057384 -0.015844

           X7        X8        X9       X10       X11       X12        X13
0    0.002173 -0.044896 -0.026252  0.025089  0.059151  0.031293  efectores
2    0.059918  0.114560  0.046208 -0.017126  0.001618  0.031668  efectores
4    0.001117 -0.083297  0.064339  0.066756  0.078989  0.010679  efectores
5    0.051467  0.000718  0.005321  0.068016 -0.031789 -0.020347  efectores
6    0.061350 -0.008490 -0.008107  0.044822  0.004994 -0.039802  efectores
..        …         …         …         …         …         …         …
995 -0.026309 -0.047577 -0.080583  0.080542 -0.037148  0.021763  efectores
```

```
996 -0.007433  0.040178  0.009606  0.023677  0.029601  0.012132  efectores
997  0.027105  0.015319  0.045230 -0.002773 -0.021141 -0.064073  efectores
998  0.009694  0.020795  0.018241 -0.011883 -0.017278 -0.034690  efectores
999  0.006180  0.082226  0.068744 -0.010000  0.106379 -0.008417  efectores

[905 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC)  hidro_mass efectores fusarium_oxysporum
dataset 5, sin valores atípicos.
Estadísticas.

| | X0 | X1 | X2 | X3 | X4 | X5 | \ |
|---|---|---|---|---|---|---|---|
| count | 905.000000 | 905.000000 | 905.000000 | 905.000000 | 905.000000 | 905.000000 | |
| mean | 0.007999 | 0.011353 | 0.011919 | 0.018099 | 0.009102 | 0.006464 | |
| std | 0.049243 | 0.050360 | 0.049966 | 0.049722 | 0.047939 | 0.050247 | |
| min | -0.159396 | -0.175444 | -0.163969 | -0.154296 | -0.136397 | -0.161343 | |
| 25% | -0.021782 | -0.020717 | -0.017582 | -0.012182 | -0.018872 | -0.024228 | |
| 50% | 0.010322 | 0.012113 | 0.011744 | 0.018542 | 0.009526 | 0.007352 | |
| 75% | 0.039068 | 0.044528 | 0.041146 | 0.048406 | 0.039845 | 0.037749 | |
| max | 0.151643 | 0.189455 | 0.190073 | 0.177800 | 0.163184 | 0.169394 | |

| | X6 | X7 | X8 | X9 | X10 | X11 | \ |
|---|---|---|---|---|---|---|---|
| count | 905.000000 | 905.000000 | 905.000000 | 905.000000 | 905.000000 | 905.000000 | |
| mean | 0.006758 | 0.003844 | 0.008579 | 0.004801 | 0.005582 | 0.004695 | |
| std | 0.047982 | 0.047730 | 0.047199 | 0.049326 | 0.048716 | 0.050076 | |
| min | -0.155970 | -0.153572 | -0.148189 | -0.163006 | -0.162979 | -0.165447 | |
| 25% | -0.022525 | -0.025664 | -0.022482 | -0.025575 | -0.024303 | -0.026570 | |
| 50% | 0.008568 | 0.005410 | 0.008763 | 0.002344 | 0.005718 | 0.004225 | |
| 75% | 0.036258 | 0.035335 | 0.039938 | 0.038084 | 0.035702 | 0.036319 | |
| max | 0.153019 | 0.155931 | 0.163273 | 0.182699 | 0.167215 | 0.180631 | |

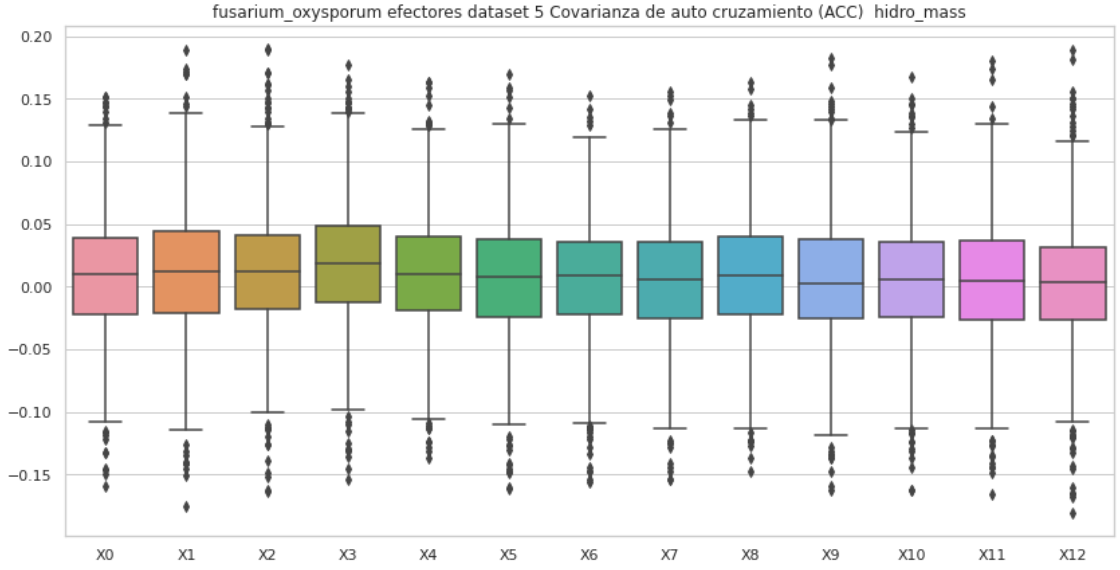
| | X12 |
|---|---|
| count | 905.000000 |
| mean | 0.003227 |
| std | 0.049886 |
| min | -0.180906 |
| 25% | -0.026065 |
| 50% | 0.003776 |
| 75% | 0.031668 |
| max | 0.189114 |

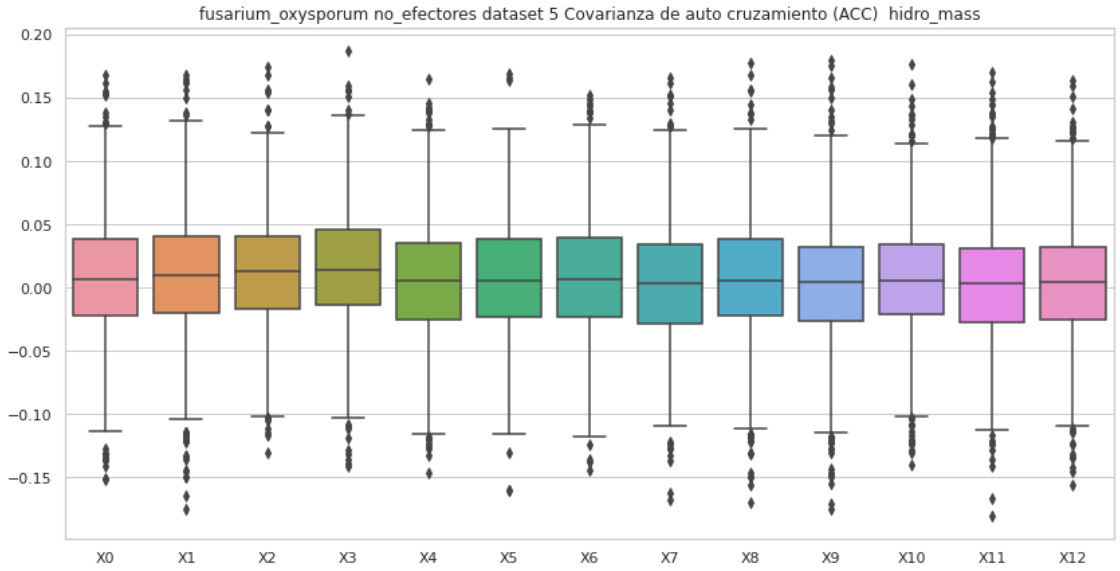
no_efectores
Covarianza de auto cruzamiento (ACC)  hidro_mass no_efectores fusarium_oxysporum
dataset 5, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.062453  0.023433  0.023638  0.115169  0.068954 -0.006815  0.009393
1    0.058356  0.027170  0.038701  0.011180  0.103074 -0.044269  0.023315
2    0.030536 -0.061034 -0.012768  0.007648  0.000491 -0.006463  0.021106
3   -0.013352  0.017059  0.068466  0.000129  0.040656  0.010577 -0.021191
4    0.017873 -0.018288 -0.027092  0.022807 -0.007655 -0.008843 -0.049344
..        …         …         …         …         …         …
994  0.052028 -0.014813  0.019281  0.007604  0.026593  0.017490  0.008255
995 -0.006176  0.032783  0.034584 -0.007126 -0.011054  0.021773  0.062234
997 -0.069707  0.035365  0.033013 -0.060010 -0.088631  0.031651  0.053052
998  0.043945  0.014359  0.002858  0.000321 -0.061820 -0.064469 -0.023915
999 -0.078321  0.010490 -0.046935 -0.027604  0.049669 -0.068051  0.041167

           X7        X8        X9       X10       X11       X12           X13
0    0.061936  0.012058  0.011067  0.040215 -0.028771  0.006251  no_efectores
1   -0.040153  0.023114 -0.037008  0.051082  0.027151 -0.069039  no_efectores
2    0.108493  0.040252 -0.037912  0.000665  0.052704  0.026046  no_efectores
3    0.068411  0.042965 -0.048515 -0.030199  0.042293 -0.011806  no_efectores
4    0.007356 -0.046218  0.025928  0.101260  0.148905  0.043595  no_efectores
..        …         …         …         …         …         …
994 -0.018331  0.041290 -0.026590 -0.001760  0.041811 -0.002940  no_efectores
995  0.055535  0.083335  0.056208  0.027111  0.000747  0.093770  no_efectores
997 -0.065087  0.101203  0.093995 -0.042853  0.108375  0.030308  no_efectores
998  0.002423 -0.064082 -0.024096  0.032817  0.021130 -0.058060  no_efectores
999  0.127845 -0.070236  0.009437  0.057222 -0.082412  0.010568  no_efectores

[898 rows x 14 columns]


Covarianza de auto cruzamiento (ACC)  hidro_mass no_efectores fusarium_oxysporum
dataset 5, sin valores atípicos.
Estadísticas.

                X0          X1          X2          X3          X4          X5  \
count  898.000000  898.000000  898.000000  898.000000  898.000000  898.000000
mean     0.006400    0.009180    0.013256    0.016169    0.005343    0.006213
std      0.049720    0.050124    0.045601    0.048278    0.048520    0.046843
min     -0.151222   -0.175161   -0.130777   -0.140661   -0.145973   -0.160551
25%     -0.022519   -0.019994   -0.016311   -0.014027   -0.025175   -0.023387
50%      0.006101    0.010129    0.013024    0.014580    0.005774    0.005390
75%      0.038330    0.040826    0.041021    0.046204    0.035397    0.038169
max      0.168299    0.167579    0.174290    0.187116    0.164975    0.169425

                X6          X7          X8          X9         X10         X11  \
count  898.000000  898.000000  898.000000  898.000000  898.000000  898.000000
mean     0.006639    0.003493    0.005824    0.003282    0.007246    0.003772
std      0.046518    0.048860    0.049218    0.049725    0.046706    0.048928
```

| | | | | | | |
|---|---|---|---|---|---|---|
| min | -0.143855 | -0.167450 | -0.170211 | -0.175136 | -0.140277 | -0.180326 |
| 25% | -0.023505 | -0.028109 | -0.022500 | -0.026576 | -0.020803 | -0.026938 |
| 50% | 0.006281 | 0.003959 | 0.005057 | 0.004325 | 0.006044 | 0.003083 |
| 75% | 0.039056 | 0.034118 | 0.038880 | 0.032428 | 0.033896 | 0.031183 |
| max | 0.152334 | 0.165912 | 0.178114 | 0.180147 | 0.176477 | 0.170716 |

| | X12 |
|---|---|
| count | 898.000000 |
| mean | 0.003402 |
| std | 0.047875 |
| min | -0.155535 |
| 25% | -0.025332 |
| 50% | 0.004328 |
| 75% | 0.031756 |
| max | 0.164246 |






# 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))
```

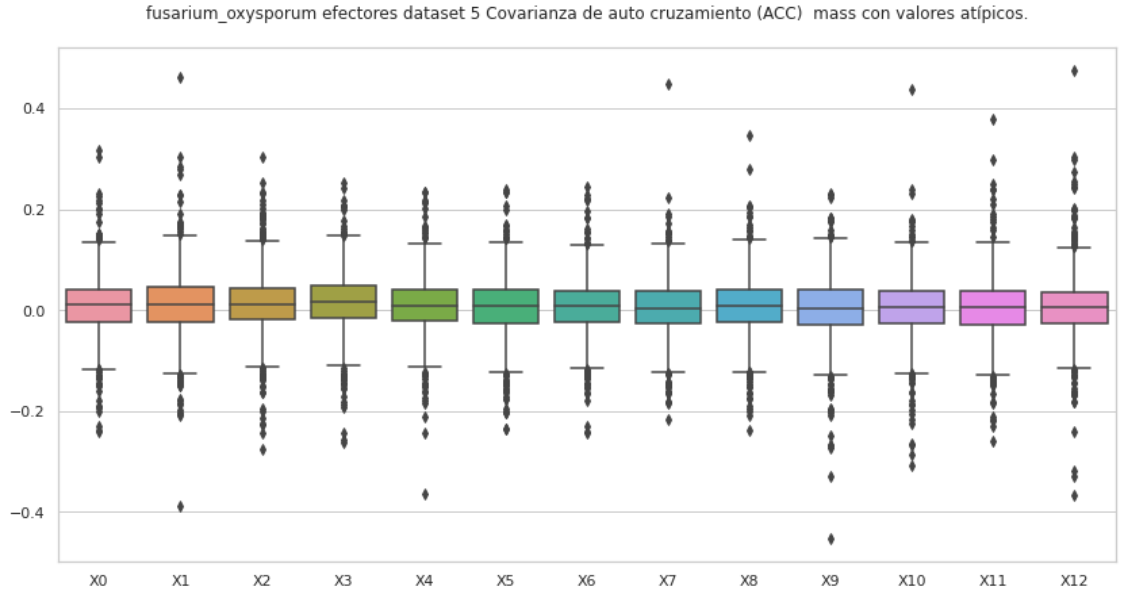
```
efectores
Covarianza de auto cruzamiento (ACC)  mass efectores fusarium_oxysporum dataset
5, con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.009715  0.000520 -0.076166 -0.044523  0.031587 -0.110099  0.053749
1   -0.047070  0.066631 -0.134354  0.069049 -0.102367  0.059562 -0.128934
2   -0.024975  0.019980  0.029167  0.084420 -0.024428  0.041260 -0.014064
3   -0.020038  0.047527  0.015308 -0.067246 -0.020417 -0.015069 -0.123033
4    0.066602 -0.003272 -0.007416  0.064446 -0.010361  0.054844  0.091152
..        …       …       …       …       …       …       …
995 -0.146291 -0.063328  0.071222  0.079739 -0.026965  0.066375 -0.052653
996  0.003814  0.053309 -0.015354  0.034205  0.072789  0.040366 -0.002819
997  0.012738  0.116541 -0.010380  0.022918 -0.040516  0.029206  0.062672
998  0.018288  0.058056 -0.024623  0.047567 -0.073932  0.046799  0.035913
999  0.046895  0.076009  0.101244 -0.041055  0.058772  0.057384 -0.015844

           X7        X8        X9       X10       X11       X12        X13
0    0.002173 -0.044896 -0.026252  0.025089  0.059151  0.031293  efectores
1    0.097181 -0.208400  0.071286 -0.068116 -0.094719 -0.169502  efectores
2    0.059918  0.114560  0.046208 -0.017126  0.001618  0.031668  efectores
3    0.000316 -0.079621 -0.209254 -0.042398  0.060060 -0.004191  efectores
4    0.001117 -0.083297  0.064339  0.066756  0.078989  0.010679  efectores
..        …       …       …       …       …       …        …
995 -0.026309 -0.047577 -0.080583  0.080542 -0.037148  0.021763  efectores
996 -0.007433  0.040178  0.009606  0.023677  0.029601  0.012132  efectores
997  0.027105  0.015319  0.045230 -0.002773 -0.021141 -0.064073  efectores
998  0.009694  0.020795  0.018241 -0.011883 -0.017278 -0.034690  efectores
999  0.006180  0.082226  0.068744 -0.010000  0.106379 -0.008417  efectores

[1000 rows x 14 columns]


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

                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.008786     0.011810     0.012648     0.016101     0.008549
```

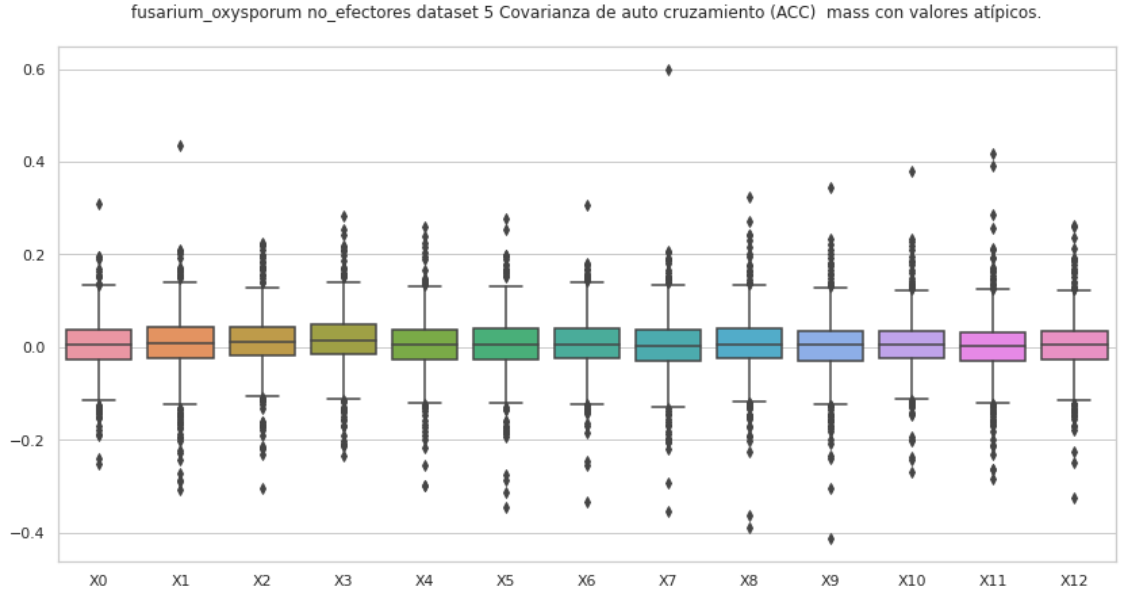
```
std       0.058600     0.063333     0.060473     0.058645     0.056455
min      -0.240910    -0.389017    -0.274770    -0.261354    -0.365513
25%      -0.022646    -0.022851    -0.019310    -0.015646    -0.021172
50%       0.010425     0.011706     0.011678     0.017790     0.009448
75%       0.040294     0.046707     0.044214     0.049634     0.041007
max       0.316845     0.462082     0.303677     0.252158     0.233359

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.006040     0.007563     0.003365     0.007956     0.002759
std       0.058601     0.056442     0.055493     0.056929     0.061549
min      -0.236181    -0.244787    -0.215459    -0.238937    -0.452724
25%      -0.025871    -0.023329    -0.027734    -0.024444    -0.027881
50%       0.007602     0.009069     0.004298     0.007872     0.003140
75%       0.039672     0.038322     0.036672     0.041136     0.040474
max       0.238519     0.243211     0.448019     0.346665     0.231211

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.003280     0.005550     0.006060
std       0.059659     0.059907     0.063594
min      -0.308986    -0.258435    -0.366551
25%      -0.027145    -0.028473    -0.025938
50%       0.005249     0.005063     0.004484
75%       0.037965     0.038892     0.034656
max       0.437510     0.378083     0.474760



no_efectores
Covarianza de auto cruzamiento (ACC)  mass no_efectores fusarium_oxysporum
dataset 5, con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.062453  0.023433  0.023638  0.115169  0.068954 -0.006815  0.009393
1    0.058356  0.027170  0.038701  0.011180  0.103074 -0.044269  0.023315
2    0.030536 -0.061034 -0.012768  0.007648  0.000491 -0.006463  0.021106
3   -0.013352  0.017059  0.068466  0.000129  0.040656  0.010577 -0.021191
4    0.017873 -0.018288 -0.027092  0.022807 -0.007655 -0.008843 -0.049344
..        …       …       …       …       …       …       …
995 -0.006176  0.032783  0.034584 -0.007126 -0.011054  0.021773  0.062234
996  0.067214  0.004278  0.224288  0.171269 -0.017389 -0.168740  0.048729
997 -0.069707  0.035365  0.033013 -0.060010 -0.088631  0.031651  0.053052
998  0.043945  0.014359  0.002858  0.000321 -0.061820 -0.064469 -0.023915
999 -0.078321  0.010490 -0.046935 -0.027604  0.049669 -0.068051  0.041167

           X7        X8        X9       X10       X11       X12           X13
```

```
0    0.061936  0.012058  0.011067  0.040215 -0.028771  0.006251  no_efectores
1   -0.040153  0.023114 -0.037008  0.051082  0.027151 -0.069039  no_efectores
2    0.108493  0.040252 -0.037912  0.000665  0.052704  0.026046  no_efectores
3    0.068411  0.042965 -0.048515 -0.030199  0.042293 -0.011806  no_efectores
4    0.007356 -0.046218  0.025928  0.101260  0.148905  0.043595  no_efectores
..        …         …         …         …         …         …           …
995  0.055535  0.083335  0.056208  0.027111  0.000747  0.093770  no_efectores
996 -0.016098 -0.021882 -0.181469 -0.070267 -0.027160 -0.010133  no_efectores
997 -0.065087  0.101203  0.093995 -0.042853  0.108375  0.030308  no_efectores
998  0.002423 -0.064082 -0.024096  0.032817  0.021130 -0.058060  no_efectores
999  0.127845 -0.070236  0.009437  0.057222 -0.082412  0.010568  no_efectores

[1000 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC)  mass no_efectores fusarium_oxysporum
dataset 5, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.005491     0.007144     0.012143     0.015477     0.004840
std       0.056383     0.062574     0.056868     0.058908     0.057728
min      -0.253343    -0.306694    -0.304586    -0.234911    -0.297840
25%      -0.025528    -0.023227    -0.017709    -0.016160    -0.026657
50%       0.005661     0.009510     0.012650     0.014666     0.005950
75%       0.038723     0.042960     0.042204     0.048194     0.036971
max       0.309722     0.436423     0.224288     0.284081     0.259404

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.004549     0.006706     0.003558     0.006259     0.002476
std       0.059323     0.054875     0.061360     0.060388     0.060325
min      -0.346435    -0.332872    -0.355538    -0.388539    -0.412724
25%      -0.027320    -0.024537    -0.029300    -0.023618    -0.029839
50%       0.005111     0.006665     0.003538     0.005006     0.004148
75%       0.039696     0.041860     0.036321     0.040628     0.034564
max       0.278656     0.307536     0.598643     0.323006     0.344545

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.007025     0.003103     0.004046
std       0.056716     0.061614     0.057415
min      -0.270820    -0.283420    -0.325193
25%      -0.023316    -0.029313    -0.026939
50%       0.005861     0.002800     0.004081
75%       0.035136     0.032440     0.033568
max       0.380560     0.417414     0.261919
```


fusarium_oxysporum efectores dataset 5 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 5 Covarianza de auto cruzamiento (ACC)  mass con valores atípicos.
0.6
0.4
0.2
0.0
−0.2
−0.4
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12

### 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
5, sin valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.009715  0.000520 -0.076166 -0.044523  0.031587 -0.110099  0.053749
2   -0.024975  0.019980  0.029167  0.084420 -0.024428  0.041260 -0.014064
4    0.066602 -0.003272 -0.007416  0.064446 -0.010361  0.054844  0.091152
5    0.023988 -0.001712 -0.042395 -0.012470  0.028812 -0.033189 -0.029020
6   -0.015862  0.062257 -0.061294  0.029002 -0.002349 -0.043555  0.023316
..        …         …         …         …         …         …         …
995 -0.146291 -0.063328  0.071222  0.079739 -0.026965  0.066375 -0.052653
996  0.003814  0.053309 -0.015354  0.034205  0.072789  0.040366 -0.002819
997  0.012738  0.116541 -0.010380  0.022918 -0.040516  0.029206  0.062672
998  0.018288  0.058056 -0.024623  0.047567 -0.073932  0.046799  0.035913
999  0.046895  0.076009  0.101244 -0.041055  0.058772  0.057384 -0.015844

           X7        X8        X9       X10       X11       X12         X13
0    0.002173 -0.044896 -0.026252  0.025089  0.059151  0.031293  efectores
2    0.059918  0.114560  0.046208 -0.017126  0.001618  0.031668  efectores
4    0.001117 -0.083297  0.064339  0.066756  0.078989  0.010679  efectores
5    0.051467  0.000718  0.005321  0.068016 -0.031789 -0.020347  efectores
6    0.061350 -0.008490 -0.008107  0.044822  0.004994 -0.039802  efectores
..        …         …         …         …         …         …         …
995 -0.026309 -0.047577 -0.080583  0.080542 -0.037148  0.021763  efectores
996 -0.007433  0.040178  0.009606  0.023677  0.029601  0.012132  efectores
997  0.027105  0.015319  0.045230 -0.002773 -0.021141 -0.064073  efectores
998  0.009694  0.020795  0.018241 -0.011883 -0.017278 -0.034690  efectores
999  0.006180  0.082226  0.068744 -0.010000  0.106379 -0.008417  efectores

[905 rows x 14 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  905.000000  905.000000  905.000000  905.000000  905.000000  905.000000
mean     0.007999    0.011353    0.011919    0.018099    0.009102    0.006464
std      0.049243    0.050360    0.049966    0.049722    0.047939    0.050247
min     -0.159396   -0.175444   -0.163969   -0.154296   -0.136397   -0.161343
25%     -0.021782   -0.020717   -0.017582   -0.012182   -0.018872   -0.024228
50%      0.010322    0.012113    0.011744    0.018542    0.009526    0.007352
75%      0.039068    0.044528    0.041146    0.048406    0.039845    0.037749
```

```
max      0.151643    0.189455    0.190073    0.177800    0.163184    0.169394

               X6          X7          X8          X9         X10         X11  \
count  905.000000  905.000000  905.000000  905.000000  905.000000  905.000000
mean     0.006758    0.003844    0.008579    0.004801    0.005582    0.004695
std      0.047982    0.047730    0.047199    0.049326    0.048716    0.050076
min     -0.155970   -0.153572   -0.148189   -0.163006   -0.162979   -0.165447
25%     -0.022525   -0.025664   -0.022482   -0.025575   -0.024303   -0.026570
50%      0.008568    0.005410    0.008763    0.002344    0.005718    0.004225
75%      0.036258    0.035335    0.039938    0.038084    0.035702    0.036319
max      0.153019    0.155931    0.163273    0.182699    0.167215    0.180631

              X12
count  905.000000
mean     0.003227
std      0.049886
min     -0.180906
25%     -0.026065
50%      0.003776
75%      0.031668
max      0.189114
```

Covarianza de auto cruzamiento (ACC)  mass no_efectores fusarium_oxysporum
dataset 5, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.062453  0.023433  0.023638  0.115169  0.068954 -0.006815  0.009393
1    0.058356  0.027170  0.038701  0.011180  0.103074 -0.044269  0.023315
2    0.030536 -0.061034 -0.012768  0.007648  0.000491 -0.006463  0.021106
3   -0.013352  0.017059  0.068466  0.000129  0.040656  0.010577 -0.021191
4    0.017873 -0.018288 -0.027092  0.022807 -0.007655 -0.008843 -0.049344
..        …         …         …         …         …         …
994  0.052028 -0.014813  0.019281  0.007604  0.026593  0.017490  0.008255
995 -0.006176  0.032783  0.034584 -0.007126 -0.011054  0.021773  0.062234
997 -0.069707  0.035365  0.033013 -0.060010 -0.088631  0.031651  0.053052
998  0.043945  0.014359  0.002858  0.000321 -0.061820 -0.064469 -0.023915
999 -0.078321  0.010490 -0.046935 -0.027604  0.049669 -0.068051  0.041167

           X7        X8        X9       X10       X11       X12           X13
0    0.061936  0.012058  0.011067  0.040215 -0.028771  0.006251  no_efectores
1   -0.040153  0.023114 -0.037008  0.051082  0.027151 -0.069039  no_efectores
2    0.108493  0.040252 -0.037912  0.000665  0.052704  0.026046  no_efectores
3    0.068411  0.042965 -0.048515 -0.030199  0.042293 -0.011806  no_efectores
4    0.007356 -0.046218  0.025928  0.101260  0.148905  0.043595  no_efectores
..        …         …         …         …         …         …           …
```

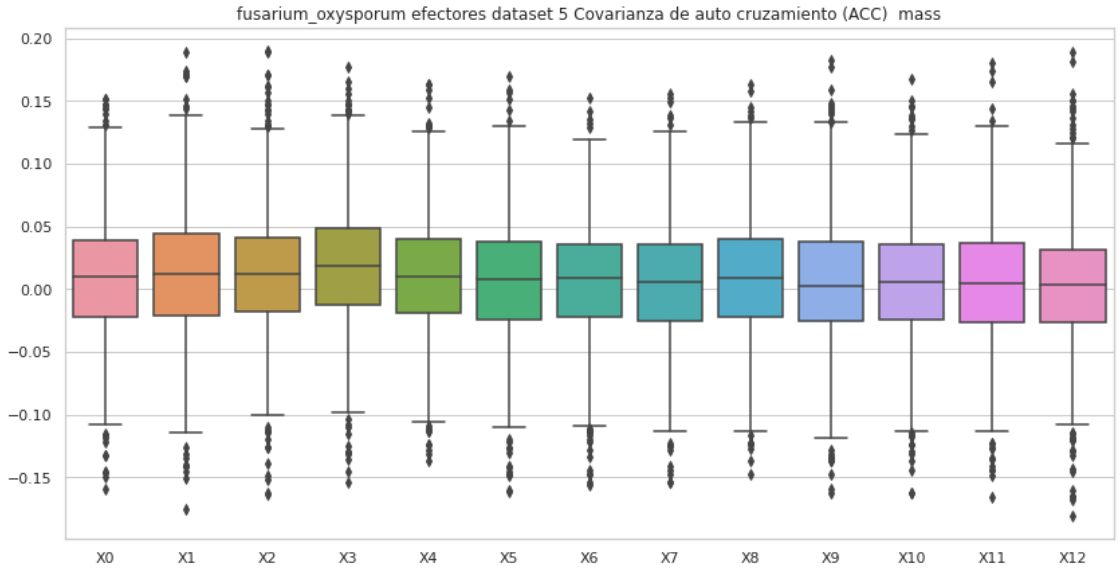
```
994 -0.018331  0.041290 -0.026590 -0.001760  0.041811 -0.002940  no_efectores
995  0.055535  0.083335  0.056208  0.027111  0.000747  0.093770  no_efectores
997 -0.065087  0.101203  0.093995 -0.042853  0.108375  0.030308  no_efectores
998  0.002423 -0.064082 -0.024096  0.032817  0.021130 -0.058060  no_efectores
999  0.127845 -0.070236  0.009437  0.057222 -0.082412  0.010568  no_efectores

[898 rows x 14 columns]
```

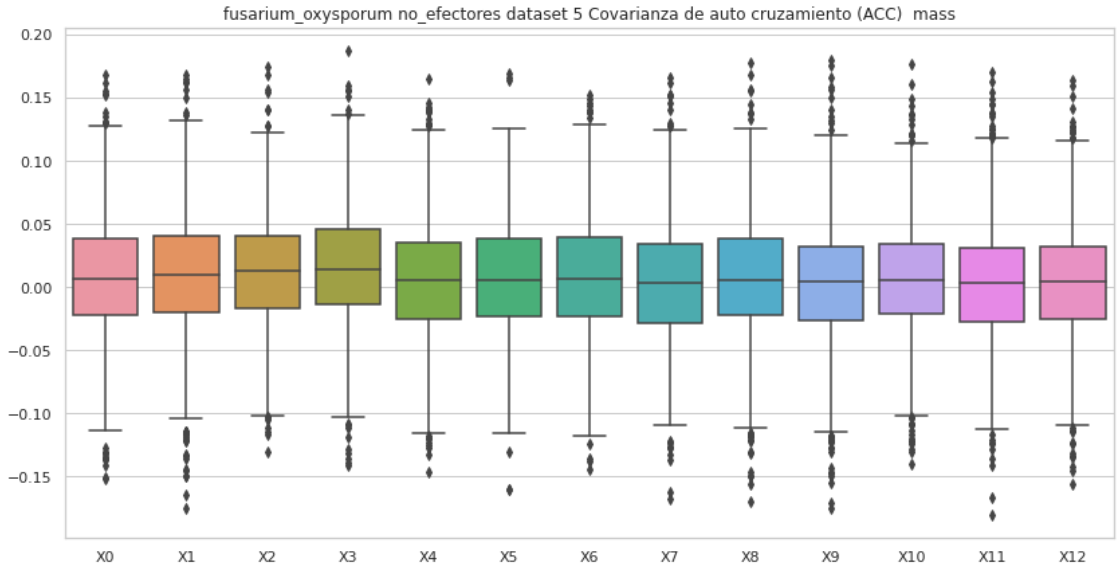
Covarianza de auto cruzamiento (ACC)  mass no_efectores fusarium_oxysporum
dataset 5, sin valores atípicos.
Estadísticas.

| | X0 | X1 | X2 | X3 | X4 | X5 | \ |
|---|---|---|---|---|---|---|---|
| count | 898.000000 | 898.000000 | 898.000000 | 898.000000 | 898.000000 | 898.000000 | |
| mean | 0.006400 | 0.009180 | 0.013256 | 0.016169 | 0.005343 | 0.006213 | |
| std | 0.049720 | 0.050124 | 0.045601 | 0.048278 | 0.048520 | 0.046843 | |
| min | -0.151222 | -0.175161 | -0.130777 | -0.140661 | -0.145973 | -0.160551 | |
| 25% | -0.022519 | -0.019994 | -0.016311 | -0.014027 | -0.025175 | -0.023387 | |
| 50% | 0.006101 | 0.010129 | 0.013024 | 0.014580 | 0.005774 | 0.005390 | |
| 75% | 0.038330 | 0.040826 | 0.041021 | 0.046204 | 0.035397 | 0.038169 | |
| max | 0.168299 | 0.167579 | 0.174290 | 0.187116 | 0.164975 | 0.169425 | |

| | X6 | X7 | X8 | X9 | X10 | X11 | \ |
|---|---|---|---|---|---|---|---|
| count | 898.000000 | 898.000000 | 898.000000 | 898.000000 | 898.000000 | 898.000000 | |
| mean | 0.006639 | 0.003493 | 0.005824 | 0.003282 | 0.007246 | 0.003772 | |
| std | 0.046518 | 0.048860 | 0.049218 | 0.049725 | 0.046706 | 0.048928 | |
| min | -0.143855 | -0.167450 | -0.170211 | -0.175136 | -0.140277 | -0.180326 | |
| 25% | -0.023505 | -0.028109 | -0.022500 | -0.026576 | -0.020803 | -0.026938 | |
| 50% | 0.006281 | 0.003959 | 0.005057 | 0.004325 | 0.006044 | 0.003083 | |
| 75% | 0.039056 | 0.034118 | 0.038880 | 0.032428 | 0.033896 | 0.031183 | |
| max | 0.152334 | 0.165912 | 0.178114 | 0.180147 | 0.176477 | 0.170716 | |

| | X12 |
|---|---|
| count | 898.000000 |
| mean | 0.003402 |
| std | 0.047875 |
| min | -0.155535 |
| 25% | -0.025332 |
| 50% | 0.004328 |
| 75% | 0.031756 |
| max | 0.164246 |




# 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
5, con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.030466  0.066635 -0.084677  0.152869  0.019768  0.108211 -0.063237
1    0.091653 -0.039449 -0.128817  0.002492  0.084600  0.120613 -0.015479
2    0.018127 -0.032502  0.063867  0.029231 -0.042173  0.018989 -0.026705
3    0.045250  0.004833  0.047489 -0.053428  0.049318  0.032860 -0.128509
4    0.039206  0.012438  0.073714 -0.035819  0.078560 -0.012359  0.013217
..        …         …         …         …         …         …         …
995 -0.048753 -0.074106 -0.081441  0.106468  0.058785 -0.160434  0.070117
996  0.098980  0.016100  0.044292  0.066376  0.061996 -0.035566  0.003999
997 -0.044521 -0.044794  0.038691  0.066006  0.018253 -0.007308 -0.038864
998  0.087498  0.053818  0.048244  0.090715  0.048482 -0.029928  0.053961
999  0.067333 -0.008878  0.057444  0.117065 -0.002817 -0.008101  0.028859

           X7        X8        X9       X10       X11       X12        X13
0    0.083579 -0.071536  0.096742 -0.013742  0.113625 -0.036505  efectores
1    0.026540  0.007017  0.123924  0.088916  0.115739 -0.023704  efectores
2   -0.055033 -0.014497 -0.082385 -0.041091 -0.037252 -0.120458  efectores
```

```
3    0.042659 -0.044972 -0.100530  0.048772 -0.186932  0.012390  efectores
4   -0.007215 -0.031749  0.010486  0.062208  0.003792  0.061033  efectores
..        …       …        …        …        …        …        …
995 -0.064896  0.093722  0.006175 -0.051491  0.151961 -0.043145  efectores
996 -0.008703  0.048872  0.036560  0.006479 -0.001680 -0.021708  efectores
997  0.097790  0.006767 -0.018749 -0.058366  0.062868 -0.078554  efectores
998 -0.022878 -0.050505 -0.072081 -0.007671  0.040607  0.021837  efectores
999  0.102513 -0.011511 -0.035940 -0.038717  0.031053  0.048352  efectores

[1000 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro efectores fusarium_oxysporum dataset 5, con valores atípicos.
Estadísticas.

| | X0 | X1 | X2 | X3 | X4 | \ |
|---|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | |
| mean | 0.012138 | -0.013580 | 0.025576 | 0.027155 | -0.000708 | |
| std | 0.070555 | 0.072480 | 0.069118 | 0.070601 | 0.068752 | |
| min | -0.549889 | -0.309226 | -0.468359 | -0.428663 | -0.327760 | |
| 25% | -0.027622 | -0.058409 | -0.012117 | -0.010000 | -0.039329 | |
| 50% | 0.010323 | -0.017036 | 0.027246 | 0.028633 | -0.002962 | |
| 75% | 0.052695 | 0.031593 | 0.065485 | 0.065767 | 0.037676 | |
| max | 0.436814 | 0.385805 | 0.419653 | 0.337553 | 0.410259 | |

| | X5 | X6 | X7 | X8 | X9 | \ |
|---|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | |
| mean | -0.002393 | 0.022695 | 0.007899 | 0.001260 | 0.008414 | |
| std | 0.064503 | 0.069417 | 0.065762 | 0.065546 | 0.066261 | |
| min | -0.393938 | -0.279802 | -0.359318 | -0.370035 | -0.354770 | |
| 25% | -0.040216 | -0.012885 | -0.028595 | -0.030578 | -0.028290 | |
| 50% | -0.004786 | 0.023626 | 0.008344 | 0.005233 | 0.010654 | |
| 75% | 0.039594 | 0.058506 | 0.044236 | 0.038296 | 0.045876 | |
| max | 0.289354 | 0.657287 | 0.312563 | 0.274757 | 0.367148 | |

| | X10 | X11 | X12 |
|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 0.011373 | 0.005797 | 0.002654 |
| std | 0.064891 | 0.064080 | 0.063877 |
| min | -0.399716 | -0.494752 | -0.267063 |
| 25% | -0.020161 | -0.026529 | -0.030124 |
| 50% | 0.010315 | 0.007983 | 0.006612 |
| 75% | 0.047439 | 0.039024 | 0.039769 |
| max | 0.417052 | 0.297204 | 0.335859 |

no_efectores
Covarianza de auto cruzamiento (ACC)  hidro no_efectores fusarium_oxysporum
dataset 5, con valores atípicos.
Valores del documento csv.

| | X0 | X1 | X2 | X3 | X4 | X5 | X6 | \ |
|---|---|---|---|---|---|---|---|---|
| 0 | -0.076936 | -0.036690 | -0.063968 | 0.036578 | -0.084878 | 0.021293 | 0.020290 | |
| 1 | -0.000545 | -0.090221 | 0.027127 | 0.011076 | -0.016988 | 0.007823 | 0.037878 | |
| 2 | -0.047990 | -0.086111 | 0.002849 | 0.013158 | -0.044674 | -0.016941 | -0.036610 | |
| 3 | 0.011812 | 0.058422 | 0.076768 | 0.052506 | 0.007100 | -0.006390 | 0.050045 | |
| 4 | 0.070388 | 0.064234 | 0.054695 | 0.049310 | 0.059041 | 0.050989 | 0.039800 | |
| .. | … | … | … | … | … | … | … | |
| 995 | 0.072084 | -0.018198 | 0.038298 | 0.014235 | 0.090472 | -0.076515 | 0.040143 | |
| 996 | -0.102482 | 0.200376 | 0.217070 | 0.026696 | 0.045315 | -0.096640 | -0.077230 | |
| 997 | -0.045761 | 0.032276 | 0.138367 | -0.013866 | -0.155094 | 0.058246 | -0.087526 | |
| 998 | 0.060058 | -0.113457 | -0.062267 | 0.048635 | -0.008727 | -0.025041 | -0.059134 | |
| 999 | -0.193222 | 0.149842 | 0.112696 | -0.071310 | 0.124874 | -0.054916 | 0.075881 | |

| | X7 | X8 | X9 | X10 | X11 | X12 | X13 |
|---|---|---|---|---|---|---|---|
| 0 | 0.035737 | -0.059553 | -0.011206 | 0.074774 | -0.002129 | 0.069316 | no_efectores |
| 1 | -0.027252 | -0.015683 | 0.040721 | -0.011745 | 0.012077 | 0.086725 | no_efectores |
| 2 | 0.030499 | 0.052392 | 0.035941 | 0.032270 | 0.060163 | -0.033407 | no_efectores |
| 3 | 0.038693 | -0.003246 | 0.025016 | 0.013103 | 0.006640 | 0.031557 | no_efectores |
| 4 | 0.040449 | 0.001809 | 0.075293 | 0.068337 | 0.067141 | 0.022243 | no_efectores |
| .. | … | … | … | … | … | … | … |
| 995 | 0.037056 | 0.013771 | 0.033268 | -0.019950 | 0.030531 | 0.006708 | no_efectores |
| 996 | -0.120100 | -0.147624 | -0.216370 | -0.390637 | 0.068964 | -0.222604 | no_efectores |
| 997 | -0.054297 | -0.028249 | 0.062236 | -0.076921 | -0.007384 | 0.108458 | no_efectores |
| 998 | -0.024193 | 0.104209 | 0.027430 | -0.104913 | -0.085135 | 0.049880 | no_efectores |
| 999 | -0.066228 | 0.255341 | -0.024125 | 0.111053 | 0.032281 | -0.121670 | no_efectores |

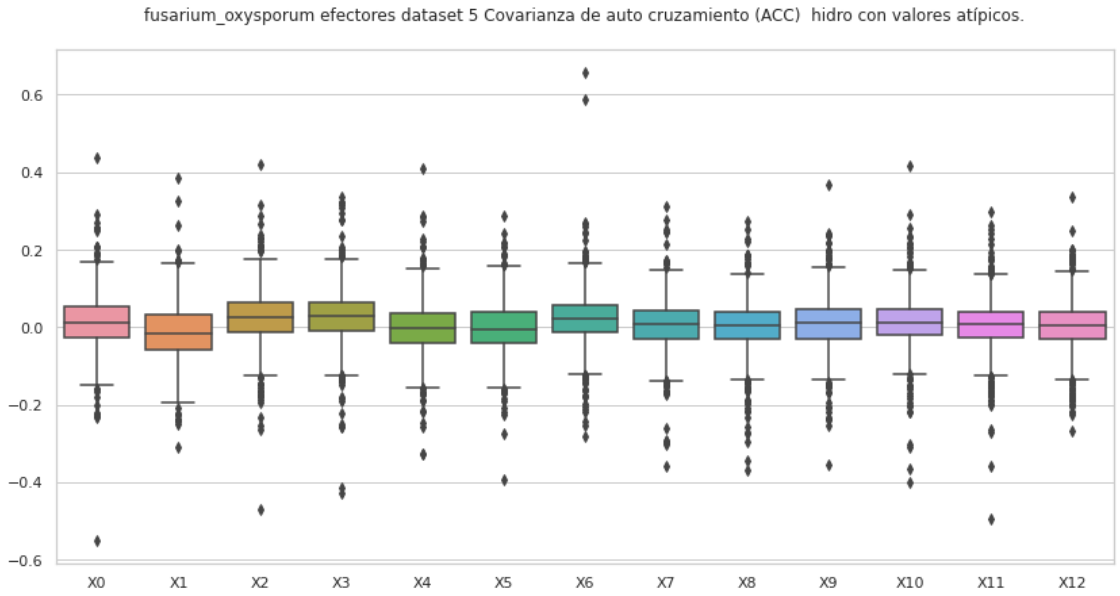
[1000 rows x 14 columns]

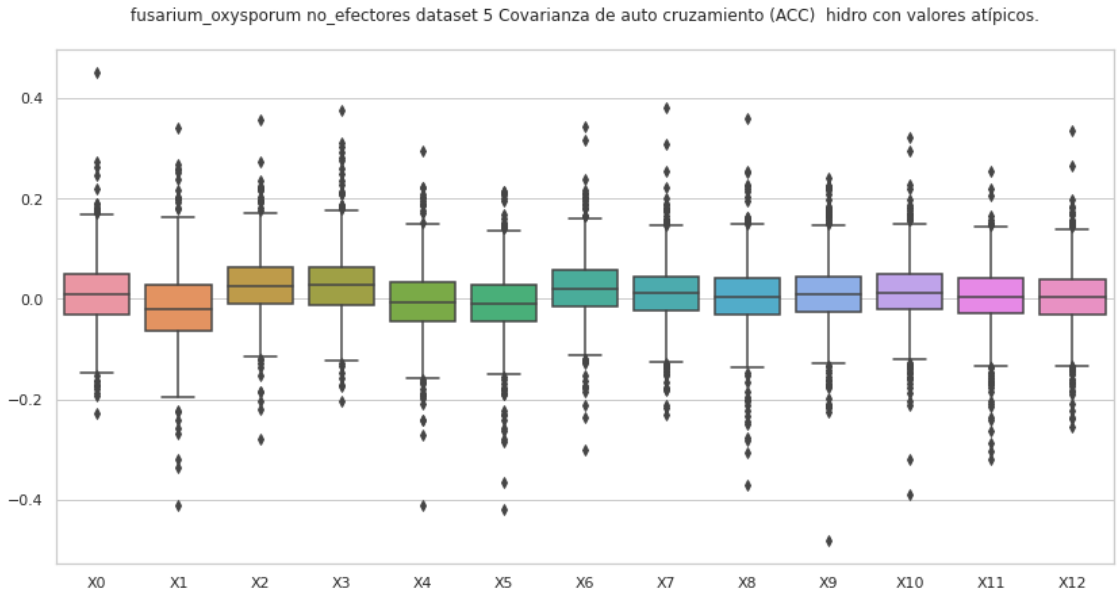
Covarianza de auto cruzamiento (ACC)  hidro no_efectores fusarium_oxysporum
dataset 5, con valores atípicos.
Estadísticas.

| | X0 | X1 | X2 | X3 | X4 | \ |
|---|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | |
| mean | 0.010550 | -0.015321 | 0.027153 | 0.028856 | -0.005333 | |
| std | 0.066992 | 0.075074 | 0.062267 | 0.066199 | 0.067099 | |
| min | -0.227147 | -0.412287 | -0.278431 | -0.204947 | -0.410568 | |
| 25% | -0.030149 | -0.062219 | -0.008911 | -0.013162 | -0.044078 | |
| 50% | 0.009813 | -0.019827 | 0.025953 | 0.028215 | -0.006308 | |
| 75% | 0.049910 | 0.028076 | 0.064329 | 0.063185 | 0.033306 | |
| max | 0.451147 | 0.341640 | 0.357349 | 0.375367 | 0.294016 | |

| | X5 | X6 | X7 | X8 | X9 | \ |
|---|---|---|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | |
| mean | -0.007317 | 0.021264 | 0.011572 | 0.003296 | 0.009958 | |
| std | 0.066543 | 0.063452 | 0.062714 | 0.066942 | 0.064627 | |
| min | -0.419963 | -0.300355 | -0.231496 | -0.371968 | -0.481312 | |
| 25% | -0.043569 | -0.014100 | -0.023046 | -0.032180 | -0.024973 | |
| 50% | -0.008260 | 0.019963 | 0.010810 | 0.004964 | 0.009028 | |
| 75% | 0.028738 | 0.057010 | 0.045306 | 0.040618 | 0.044746 | |
| max | 0.214099 | 0.344804 | 0.382157 | 0.360915 | 0.240024 | |

| | X10 | X11 | X12 |
|---|---|---|---|
| count | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 0.013660 | 0.003816 | 0.002703 |
| std | 0.062198 | 0.060898 | 0.061033 |
| min | -0.390637 | -0.320810 | -0.253766 |
| 25% | -0.021432 | -0.028993 | -0.029993 |
| 50% | 0.012275 | 0.003735 | 0.004895 |
| 75% | 0.048552 | 0.041279 | 0.038787 |
| max | 0.321838 | 0.255648 | 0.335822 |






## 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 5,
sin valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.030466  0.066635 -0.084677  0.152869  0.019768  0.108211 -0.063237
1    0.091653 -0.039449 -0.128817  0.002492  0.084600  0.120613 -0.015479
2    0.018127 -0.032502  0.063867  0.029231 -0.042173  0.018989 -0.026705
4    0.039206  0.012438  0.073714 -0.035819  0.078560 -0.012359  0.013217
5    0.006328 -0.074122  0.010892 -0.029626  0.025528 -0.001818  0.056133
..        …         …         …         …         …         …         …
995 -0.048753 -0.074106 -0.081441  0.106468  0.058785 -0.160434  0.070117
996  0.098980  0.016100  0.044292  0.066376  0.061996 -0.035566  0.003999
997 -0.044521 -0.044794  0.038691  0.066006  0.018253 -0.007308 -0.038864
998  0.087498  0.053818  0.048244  0.090715  0.048482 -0.029928  0.053961
999  0.067333 -0.008878  0.057444  0.117065 -0.002817 -0.008101  0.028859

           X7        X8        X9       X10       X11       X12        X13
0    0.083579 -0.071536  0.096742 -0.013742  0.113625 -0.036505  efectores
1    0.026540  0.007017  0.123924  0.088916  0.115739 -0.023704  efectores
2   -0.055033 -0.014497 -0.082385 -0.041091 -0.037252 -0.120458  efectores
4   -0.007215 -0.031749  0.010486  0.062208  0.003792  0.061033  efectores
5    0.048519 -0.019963 -0.011734 -0.024644 -0.007478 -0.094260  efectores
..        …         …         …         …         …         …         …
995 -0.064896  0.093722  0.006175 -0.051491  0.151961 -0.043145  efectores
```

```
996 -0.008703  0.048872  0.036560  0.006479 -0.001680 -0.021708  efectores
997  0.097790  0.006767 -0.018749 -0.058366  0.062868 -0.078554  efectores
998 -0.022878 -0.050505 -0.072081 -0.007671  0.040607  0.021837  efectores
999  0.102513 -0.011511 -0.035940 -0.038717  0.031053  0.048352  efectores

[918 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) efectores fusarium_oxysporum dataset 5, sin valores atípicos.
Estadísticas.

| | X0 | X1 | X2 | X3 | X4 | X5 | \ |
|---|---|---|---|---|---|---|---|
| count | 918.000000 | 918.000000 | 918.000000 | 918.000000 | 918.000000 | 918.000000 | |
| mean | 0.012435 | -0.015232 | 0.026100 | 0.026215 | -0.001558 | -0.003064 | |
| std | 0.058767 | 0.063857 | 0.057403 | 0.057097 | 0.056608 | 0.055110 | |
| min | -0.179104 | -0.225719 | -0.180539 | -0.184048 | -0.187068 | -0.169786 | |
| 25% | -0.025910 | -0.057622 | -0.009375 | -0.008127 | -0.037564 | -0.038667 | |
| 50% | 0.010376 | -0.017855 | 0.027246 | 0.027435 | -0.003270 | -0.004372 | |
| 75% | 0.050286 | 0.027907 | 0.063838 | 0.064492 | 0.035642 | 0.035695 | |
| max | 0.207796 | 0.198131 | 0.229028 | 0.185540 | 0.162465 | 0.185306 | |

| | X6 | X7 | X8 | X9 | X10 | X11 | \ |
|---|---|---|---|---|---|---|---|
| count | 918.000000 | 918.000000 | 918.000000 | 918.000000 | 918.000000 | 918.000000 | |
| mean | 0.022278 | 0.007612 | 0.003695 | 0.009358 | 0.012633 | 0.006922 | |
| std | 0.054126 | 0.054085 | 0.054374 | 0.056408 | 0.052901 | 0.050582 | |
| min | -0.176484 | -0.172915 | -0.186546 | -0.168070 | -0.182562 | -0.177121 | |
| 25% | -0.009073 | -0.026599 | -0.028147 | -0.024976 | -0.018710 | -0.024650 | |
| 50% | 0.023626 | 0.007749 | 0.006418 | 0.012618 | 0.010034 | 0.007820 | |
| 75% | 0.055671 | 0.040895 | 0.037378 | 0.045181 | 0.044440 | 0.036743 | |
| max | 0.180583 | 0.171714 | 0.185892 | 0.187513 | 0.205607 | 0.175301 | |

| | X12 |
|---|---|
| count | 918.000000 |
| mean | 0.005054 |
| std | 0.055255 |
| min | -0.176024 |
| 25% | -0.027251 |
| 50% | 0.007459 |
| 75% | 0.039230 |
| max | 0.185728 |

no_efectores
Covarianza de auto cruzamiento (ACC) no_efectores fusarium_oxysporum dataset 5, sin valores atípicos.
Valores del documento csv.

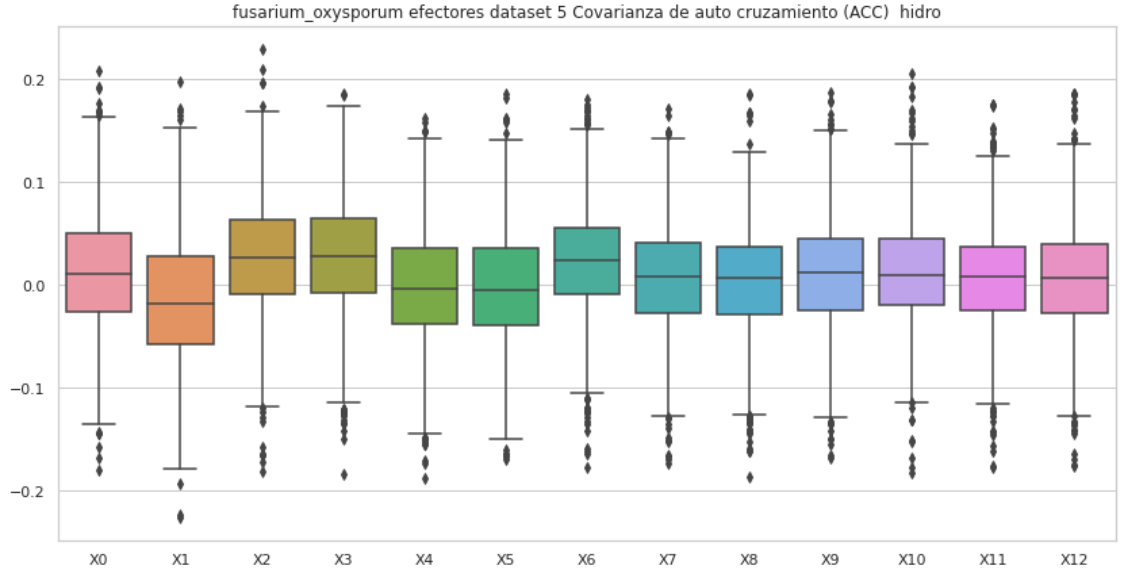
```
           X0        X1        X2        X3        X4        X5        X6  \
0   -0.076936 -0.036690 -0.063968  0.036578 -0.084878  0.021293  0.020290
1   -0.000545 -0.090221  0.027127  0.011076 -0.016988  0.007823  0.037878
2   -0.047990 -0.086111  0.002849  0.013158 -0.044674 -0.016941 -0.036610
3    0.011812  0.058422  0.076768  0.052506  0.007100 -0.006390  0.050045
4    0.070388  0.064234  0.054695  0.049310  0.059041  0.050989  0.039800
..        …         …         …         …         …         …
993 -0.023360 -0.075536  0.028805  0.102100  0.056703 -0.048275  0.028600
994 -0.011229  0.025706 -0.023829  0.061298 -0.079929  0.020271  0.016247
995  0.072084 -0.018198  0.038298  0.014235  0.090472 -0.076515  0.040143
997 -0.045761  0.032276  0.138367 -0.013866 -0.155094  0.058246 -0.087526
998  0.060058 -0.113457 -0.062267  0.048635 -0.008727 -0.025041 -0.059134

           X7        X8        X9       X10       X11       X12           X13
0    0.035737 -0.059553 -0.011206  0.074774 -0.002129  0.069316  no_efectores
1   -0.027252 -0.015683  0.040721 -0.011745  0.012077  0.086725  no_efectores
2    0.030499  0.052392  0.035941  0.032270  0.060163 -0.033407  no_efectores
3    0.038693 -0.003246  0.025016  0.013103  0.006640  0.031557  no_efectores
4    0.040449  0.001809  0.075293  0.068337  0.067141  0.022243  no_efectores
..        …         …         …         …         …         …
993  0.055427 -0.014571  0.045788  0.000327 -0.066240  0.005313  no_efectores
994  0.061747 -0.007272 -0.039469 -0.034733 -0.080902  0.006439  no_efectores
995  0.037056  0.013771  0.033268 -0.019950  0.030531  0.006708  no_efectores
997 -0.054297 -0.028249  0.062236 -0.076921 -0.007384  0.108458  no_efectores
998 -0.024193  0.104209  0.027430 -0.104913 -0.085135  0.049880  no_efectores

[912 rows x 14 columns]
```

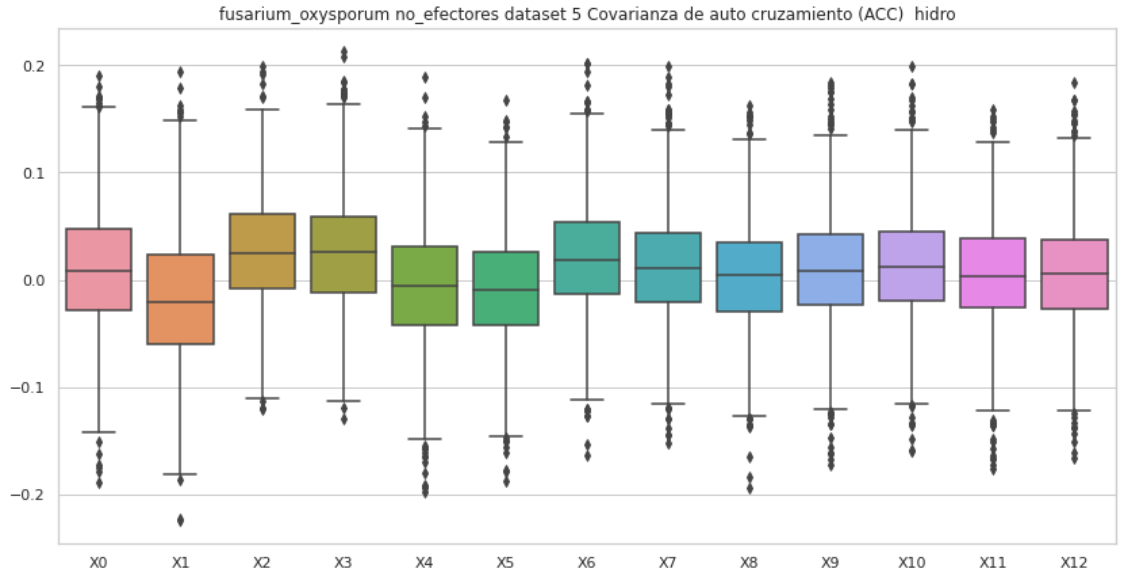
Covarianza de auto cruzamiento (ACC) no_efectores fusarium_oxysporum dataset 5, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  912.000000  912.000000  912.000000  912.000000  912.000000  912.000000
mean     0.008838   -0.017038    0.025467    0.026297   -0.005554   -0.006888
std      0.060447    0.063598    0.053934    0.056147    0.057765    0.055146
min     -0.188474   -0.224296   -0.120896   -0.129015   -0.197747   -0.187775
25%     -0.028789   -0.059509   -0.007631   -0.011435   -0.042109   -0.042820
50%      0.008507   -0.020811    0.025109    0.026500   -0.006160   -0.009216
75%      0.047149    0.023887    0.061590    0.059252    0.031540    0.025841
max      0.190512    0.193888    0.199254    0.212936    0.189090    0.167831

                X6          X7          X8          X9         X10         X11  \
count  912.000000  912.000000  912.000000  912.000000  912.000000  912.000000
mean     0.020322    0.011358    0.002440    0.008993    0.013631    0.004675
std      0.054570    0.053252    0.053232    0.053729    0.051812    0.050954
```

| | | | | | | |
|---|---|---|---|---|---|---|
| min | -0.163199 | -0.151793 | -0.194490 | -0.172182 | -0.160334 | -0.176603 |
| 25% | -0.013628 | -0.020963 | -0.030226 | -0.023436 | -0.019194 | -0.026362 |
| 50% | 0.018789 | 0.010810 | 0.004400 | 0.008755 | 0.011462 | 0.003735 |
| 75% | 0.053836 | 0.043426 | 0.035352 | 0.042078 | 0.045161 | 0.038084 |
| max | 0.201799 | 0.199429 | 0.162276 | 0.183473 | 0.198862 | 0.158848 |

| | X12 |
|---|---|
| count | 912.000000 |
| mean | 0.003900 |
| std | 0.051969 |
| min | -0.165805 |
| 25% | -0.027492 |
| 50% | 0.005474 |
| 75% | 0.036721 |
| max | 0.184225 |




fusarium_oxysporum no_efectores dataset 5 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