ds1_Meloidogyne_limpieza_de_datos

January 19, 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 ="Meloidogyne"
dataset = 1
nombre = ("ds" + str(dataset) + "_" + str(organismo))
nombre2 = (str(organismo)+ " dataset " + str(dataset))
r2 = ("Datos/resultados/"+ str(organismo) + "/" + str(nombre) + "/
 ↪transformaciones/sin_filtrar")
r3 = ("Datos/resultados/"+ str(organismo) + "/" + str(nombre) + "/
 ↪transformaciones/sin_atipicos")

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

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

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

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

2 Composición de aminoácidos (AAC)

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

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

    if etiq == "efectores":
        df=AAC_efec

    if etiq == "no_efectores":
        df=AAC_no_efec

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

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

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

        X0      X1      X2     X3     X4      X5     X6      X7     X8  \
0    7.083   2.083   5.417  6.250  0.833  10.833  0.833   3.333  1.667
1    2.214   7.011   7.011  5.904  3.321   9.963  3.321   8.487  2.583
2    3.846   4.327   9.615  9.135  5.288   2.885  4.808  11.538  3.365
3    4.762   0.595   7.738  5.952  2.976   4.167  4.167   7.143  1.786
4    3.185   1.911   3.822  1.911  3.185   3.185  7.643   2.548  1.274
..     …       …       …      …      …       …      …       …      …
95   4.188   3.665  10.471  9.948  2.094   3.141  4.712   1.571  0.524
96   4.152   1.730   6.574  2.768  3.114   3.114  2.768   6.228  2.422
97   8.054   2.685   6.040  6.040  2.013   2.013  1.342   9.396  0.671
98  12.800  13.600   4.000  4.000  0.000   2.400  1.600  12.000  0.800
99   7.429   1.714   7.429  1.714  2.857   7.429  3.429   8.571  1.143
```

```
        X9  …     X11    X12     X13     X14    X15     X16    X17    X18  \
0    5.000  …  17.500  2.083   2.917   6.250  5.833   3.750  1.250  4.583
1    7.380  …   7.380  1.476   5.535   3.690  5.904   2.583  1.476  2.583
2    6.250  …   3.846  0.962   6.250   2.404  5.769   4.808  0.962  2.885
3    6.548  …  11.905  3.571   2.976   5.357  7.738   4.762  0.000  2.381
4   12.739  …   5.732  2.548  10.191   3.185  7.006   5.732  1.274  3.822
..     …  …     …    …      …       …    …       …    …      …
95   8.377  …   9.948  2.618   3.141   3.141  5.236   3.665  1.047  3.141
96   4.498  …   6.920  2.076   1.730  16.609  9.689   7.266  1.730  4.498
97   6.040  …   7.383  5.369   4.027   2.685  8.725  12.081  1.342  2.013
98   8.000  …   2.400  4.000   4.000   3.200  6.400   4.000  2.400  0.800
99   8.571  …   9.714  2.286   4.571   2.857  5.714  10.286  0.000  1.143

      X19         X20
0   4.583   efectores
1   5.535   efectores
2   6.731   efectores
3   9.524   efectores
4   5.732   efectores
..    …         …
95  6.806   efectores
96  5.190   efectores
97  5.369   efectores
98  5.600   efectores
99  5.714   efectores

[100 rows x 21 columns]
```

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

```
                X0          X1          X2          X3          X4          X5  \
count   100.000000  100.000000  100.000000  100.000000  100.00000  100.00000
mean      5.567740    2.934040    6.294810    4.758190    2.95682    6.53192
std       2.939291    2.413183    3.110864    2.788327    3.40403    5.92157
min       0.000000    0.000000    0.000000    0.000000    0.00000    0.00000
25%       3.303750    1.200250    4.017250    2.775500    0.63825    2.77800
50%       5.204000    2.213500    6.222500    4.456000    2.15100    4.44450
75%       7.333750    4.074500    8.107000    6.009750    3.51250    8.38950
max      14.000000   13.600000   14.667000   13.253000   18.34900   34.94000

                X6          X7          X8          X9          X10         X11  \
count   100.000000  100.000000  100.000000  100.000000  100.000000  100.000000
mean      3.482180    9.018310    1.534920    6.713010    8.200210    8.908050
std       2.717604    6.948871    1.257462    2.933336    3.559774    4.174185
```

min	0.000000	1.031000	0.000000	1.149000	1.961000	1.961000
25%	1.740750	4.895250	0.634500	4.571250	5.634000	6.103500
50%	3.288500	7.060000	1.315500	6.726500	7.407500	8.929000
75%	4.586750	10.632250	2.290750	7.844000	10.069250	10.918750
max	20.803000	40.196000	6.593000	17.021000	18.557000	21.858000

	X12	X13	X14	X15	X16	X17 \
count	100.000000	100.000000	100.00000	100.000000	100.000000	100.000000
mean	2.509160	4.994650	4.59916	7.081730	5.019710	0.984940
std	1.573719	2.893151	3.49829	2.904576	2.977821	1.432658
min	0.000000	0.365000	0.82600	2.128000	0.000000	0.000000
25%	1.266000	3.061000	2.51975	5.184500	3.007750	0.000000
50%	2.260000	4.494500	3.90600	6.742500	4.348000	0.458500
75%	3.571000	6.250000	5.33900	8.578750	6.667000	1.317250
max	9.302000	13.514000	22.78900	16.456000	15.238000	6.494000

	X18	X19
count	100.000000	100.000000
mean	2.867530	5.042900
std	2.022923	2.228547
min	0.000000	0.000000
25%	1.724000	3.790750
50%	2.521500	5.120500
75%	4.040000	6.377250
max	12.644000	10.811000

no_efectores
Composición de aminoácidos (AAC) no_efectores Meloidogyne dataset 1, con valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9 \
0	5.362	3.485	6.434	5.362	2.413	6.971	5.094	3.485	2.145	8.311
1	4.800	4.000	5.600	2.400	2.400	8.800	7.200	5.600	3.200	7.200
2	5.618	6.742	1.124	5.618	5.618	0.000	5.618	6.742	1.124	5.618
3	5.852	4.811	8.322	3.121	1.951	6.762	7.022	4.941	1.951	5.982
4	2.597	3.896	8.442	3.247	5.195	11.688	1.948	5.195	2.597	5.844
..	…	…	…	…	…	…	…	…	…	
495	9.189	4.865	2.703	3.784	2.162	5.946	2.703	5.405	3.243	3.784
496	4.627	5.656	6.170	5.656	0.514	10.026	10.797	4.884	2.314	7.969
497	4.077	3.837	6.795	5.596	5.196	7.114	6.235	7.354	2.078	7.194
498	1.724	6.034	6.897	4.310	0.862	7.759	3.448	2.586	2.586	6.034
499	7.417	3.836	5.882	5.882	0.767	8.184	4.604	4.092	1.023	4.859

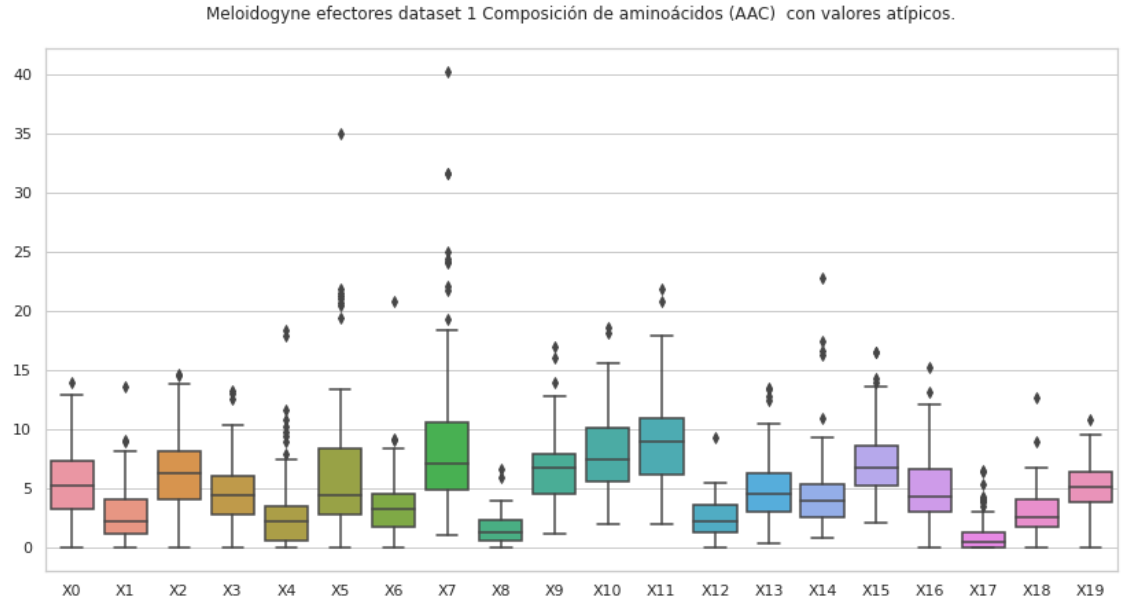
	…	X11	X12	X13	X14	X15	X16	X17	X18	X19 \
0	…	8.847	1.877	2.949	1.609	8.311	6.166	1.877	1.609	4.290

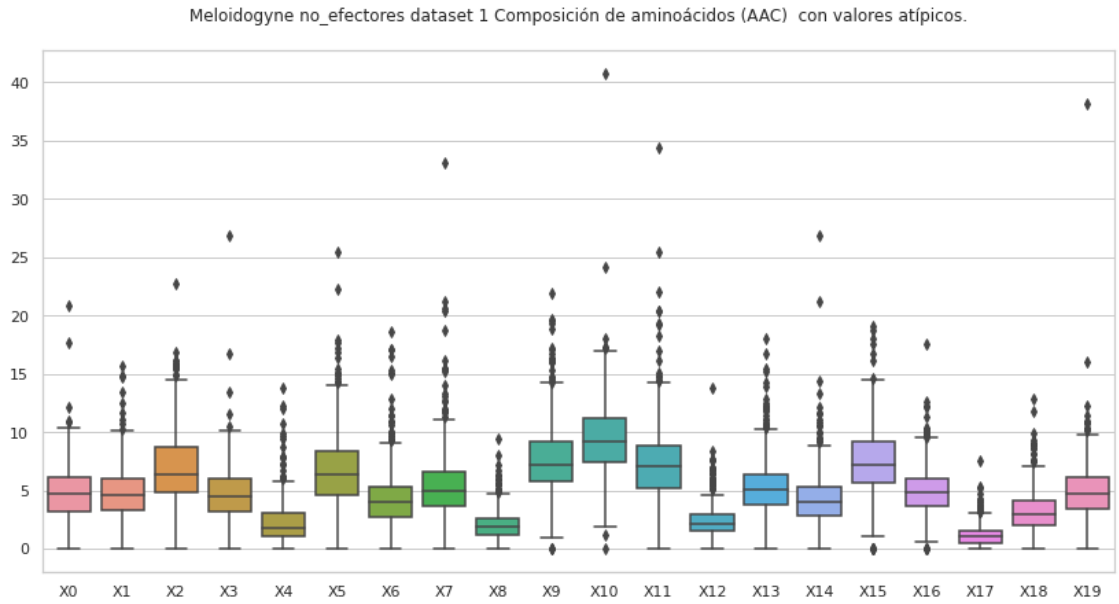
```
1    …  11.200  3.200  4.000  4.800  4.000  7.200  0.800  0.800  6.400
2    …   7.865  2.247  8.989  5.618  7.865  7.865  1.124  0.000  3.371
3    …   4.291  1.170  5.982  7.282  9.103  5.722  0.390  2.081  3.251
4    …   7.792  1.299  5.195  3.247  4.545  5.844  2.597  4.545  5.844
..   …     …    …    …    …    …    …    …    …    …
495  …   6.486  2.703  8.108  3.243  4.324  8.108  3.243  4.324  4.324
496  …   6.170  2.571  1.542  4.370  5.398  7.198  0.000  1.285  4.370
497  …   5.915  0.639  4.956  4.716  9.193  4.077  1.119  3.197  3.917
498  …   6.897  3.448  7.759  5.172  6.897  6.034  5.172  3.448  3.448
499  …  13.299  1.790  4.348  4.092  4.859  9.719  1.535  2.558  5.115

                X20
0     no_efectores
1     no_efectores
2     no_efectores
3     no_efectores
4     no_efectores
..             …
495   no_efectores
496   no_efectores
497   no_efectores
498   no_efectores
499   no_efectores

[500 rows x 21 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     4.780436    4.693216    6.925346    4.668378    2.236658    6.729016
std      2.335963    2.353719    3.270207    2.284615    1.873544    3.288728
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      3.218250    3.257250    4.815500    3.180500    1.050250    4.545000
50%      4.758000    4.634500    6.336000    4.476500    1.790500    6.406500
75%      6.152750    6.011250    8.738000    5.977500    3.030000    8.395750
max     20.833000   15.686000   22.785000   26.804000   13.793000   25.490000

                X6          X7         X8          X9         X10         X11  \
count  500.000000  500.000000  500.00000  500.000000  500.000000  500.000000
mean     4.280964    5.451268    1.99866    7.634752    9.381432    7.396768
std      2.614778    3.159025    1.25667    3.037188    3.274357    3.538920
min      0.000000    0.000000    0.00000    0.000000    0.000000    0.000000
25%      2.718500    3.618500    1.16175    5.749500    7.401750    5.230250
50%      3.966500    4.922000    1.91150    7.234500    9.210500    7.078000
```

75%	5.343250	6.652000	2.58650	9.160750	11.230000	8.875500
max	18.634000	33.113000	9.41200	21.887000	40.741000	34.356000

	X12	X13	X14	X15	X16	X17 \
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.00000
mean	2.365038	5.312168	4.283774	7.551558	4.987046	1.14997
std	1.386561	2.540780	2.496549	2.983131	2.132160	0.96718
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000
25%	1.485000	3.729500	2.817250	5.639250	3.629000	0.49275
50%	2.131000	5.026000	4.036500	7.229000	4.858500	1.02300
75%	2.907000	6.330750	5.301000	9.193500	6.002500	1.54325
max	13.750000	18.033000	26.797000	19.118000	17.526000	7.50000

	X18	X19
count	500.000000	500.000000
mean	3.228444	4.945082
std	1.833946	2.576994
min	0.000000	0.000000
25%	2.035250	3.448000
50%	3.000000	4.720000
75%	4.115500	6.076250
max	12.857000	38.144000





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

      X0     X1      X2     X3     X4      X5     X6      X7     X8      X9  \
0   7.083  2.083   5.417  6.250  0.833  10.833  0.833   3.333  1.667   5.000
1   2.214  7.011   7.011  5.904  3.321   9.963  3.321   8.487  2.583   7.380
2   3.846  4.327   9.615  9.135  5.288   2.885  4.808  11.538  3.365   6.250
3   4.762  0.595   7.738  5.952  2.976   4.167  4.167   7.143  1.786   6.548
4   3.185  1.911   3.822  1.911  3.185   3.185  7.643   2.548  1.274  12.739
..    …      …       …      …      …       …      …       …      …       …
92  5.357  3.214   7.143  5.000  3.571   2.857  1.786   8.571  1.786   6.786
94  1.899  2.532  14.557  3.797  3.797  11.392  2.532   3.165  0.633   7.595
95  4.188  3.665  10.471  9.948  2.094   3.141  4.712   1.571  0.524   8.377
97  8.054  2.685   6.040  6.040  2.013   2.013  1.342   9.396  0.671   6.040
99  7.429  1.714   7.429  1.714  2.857   7.429  3.429   8.571  1.143   8.571

    …     X11    X12     X13    X14    X15     X16    X17    X18    X19  \
0   …  17.500  2.083   2.917  6.250  5.833   3.750  1.250  4.583  4.583
1   …   7.380  1.476   5.535  3.690  5.904   2.583  1.476  2.583  5.535
2   …   3.846  0.962   6.250  2.404  5.769   4.808  0.962  2.885  6.731
3   …  11.905  3.571   2.976  5.357  7.738   4.762  0.000  2.381  9.524
4   …   5.732  2.548  10.191  3.185  7.006   5.732  1.274  3.822  5.732
..  …      …      …       …      …      …       …      …      …      …
92  …   7.857  1.071   6.071  3.929  8.571   8.571  1.429  2.857  6.786
```

```
94  …   6.962  3.797   9.494  2.532  2.532   6.329  0.633  1.266  5.696
95  …   9.948  2.618   3.141  3.141  5.236   3.665  1.047  3.141  6.806
97  …   7.383  5.369   4.027  2.685  8.725  12.081  1.342  2.013  5.369
99  …   9.714  2.286   4.571  2.857  5.714  10.286  0.000  1.143  5.714

          X20
0   efectores
1   efectores
2   efectores
3   efectores
4   efectores
..        …
92  efectores
94  efectores
95  efectores
97  efectores
99  efectores

[78 rows x 21 columns]
```

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

```
              X0         X1         X2         X3         X4         X5  \
count  78.000000  78.000000  78.000000  78.000000  78.000000  78.000000
mean    5.854167   2.754821   6.823974   5.031615   2.783462   6.867513
std     2.739625   1.987811   2.863793   2.551107   2.647225   5.065803
min     0.000000   0.000000   0.000000   0.000000   0.000000   0.000000
25%     4.201500   1.186000   5.362750   3.501000   0.695750   3.146250
50%     5.505500   2.075000   6.999000   5.000000   2.701000   5.648500
75%     7.397250   4.057000   8.174500   6.040000   3.551500   8.628750
max    14.000000   8.145000  14.667000  13.016000  11.594000  21.515000

              X6         X7         X8         X9        X10        X11  \
count  78.000000  78.000000  78.000000  78.000000  78.000000  78.000000
mean    3.431167   7.424885   1.642115   6.849231   8.257487   9.468756
std     2.046629   4.527084   1.017198   2.583696   3.504864   3.811352
min     0.000000   1.031000   0.000000   1.905000   2.190000   2.174000
25%     1.786000   4.728250   0.893000   4.808750   5.773500   6.740750
50%     3.419000   6.667000   1.635000   7.032500   7.546000   9.569000
75%     4.545000   9.372500   2.319250   8.377000  10.119750  11.111000
max     9.044000  25.000000   3.933000  13.978000  18.557000  20.803000

             X12        X13        X14        X15        X16        X17  \
count  78.000000  78.000000  78.000000  78.000000  78.000000  78.000000
mean    2.434192   5.050910   4.113769   6.774718   5.435000   0.838769
```

std	1.393912	2.908362	1.871240	2.219453	2.757104	1.131489
min	0.000000	0.365000	0.826000	2.532000	0.985000	0.000000
25%	1.261000	3.081000	2.708250	5.236000	3.275250	0.000000
50%	2.292500	4.130500	3.909500	6.620000	4.785000	0.458500
75%	3.563250	6.227000	5.257250	7.768000	6.861750	1.268000
max	5.369000	13.514000	9.278000	13.636000	13.139000	4.082000

	X18	X19
count	78.000000	78.000000
mean	2.801936	5.361513
std	1.583119	2.187300
min	0.000000	1.075000
25%	1.786000	3.953500
50%	2.573500	5.477000
75%	3.972250	6.772250
max	6.731000	10.811000

no_efectores

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

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9	\
0	5.362	3.485	6.434	5.362	2.413	6.971	5.094	3.485	2.145	8.311	
1	4.800	4.000	5.600	2.400	2.400	8.800	7.200	5.600	3.200	7.200	
2	5.618	6.742	1.124	5.618	5.618	0.000	5.618	6.742	1.124	5.618	
3	5.852	4.811	8.322	3.121	1.951	6.762	7.022	4.941	1.951	5.982	
4	2.597	3.896	8.442	3.247	5.195	11.688	1.948	5.195	2.597	5.844	
..	…	…	…	…	…	…	…	…	…		
494	5.337	3.652	5.056	4.775	1.685	8.708	5.056	5.899	2.528	3.933	
495	9.189	4.865	2.703	3.784	2.162	5.946	2.703	5.405	3.243	3.784	
496	4.627	5.656	6.170	5.656	0.514	10.026	10.797	4.884	2.314	7.969	
497	4.077	3.837	6.795	5.596	5.196	7.114	6.235	7.354	2.078	7.194	
499	7.417	3.836	5.882	5.882	0.767	8.184	4.604	4.092	1.023	4.859	

	…	X11	X12	X13	X14	X15	X16	X17	X18	X19	\
0	…	8.847	1.877	2.949	1.609	8.311	6.166	1.877	1.609	4.290	
1	…	11.200	3.200	4.000	4.800	4.000	7.200	0.800	0.800	6.400	
2	…	7.865	2.247	8.989	5.618	7.865	7.865	1.124	0.000	3.371	
3	…	4.291	1.170	5.982	7.282	9.103	5.722	0.390	2.081	3.251	
4	…	7.792	1.299	5.195	3.247	4.545	5.844	2.597	4.545	5.844	
..	…	…	…	…	…	…	…	…	…		
494	…	7.865	1.404	5.056	5.899	7.865	6.461	0.000	3.652	4.775	
495	…	6.486	2.703	8.108	3.243	4.324	8.108	3.243	4.324	4.324	
496	…	6.170	2.571	1.542	4.370	5.398	7.198	0.000	1.285	4.370	
497	…	5.915	0.639	4.956	4.716	9.193	4.077	1.119	3.197	3.917	

```
499  …  13.299  1.790  4.348  4.092  4.859  9.719  1.535  2.558  5.115

              X20
0     no_efectores
1     no_efectores
2     no_efectores
3     no_efectores
4     no_efectores
..             …
494   no_efectores
495   no_efectores
496   no_efectores
497   no_efectores
499   no_efectores

[409 rows x 21 columns]
```

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

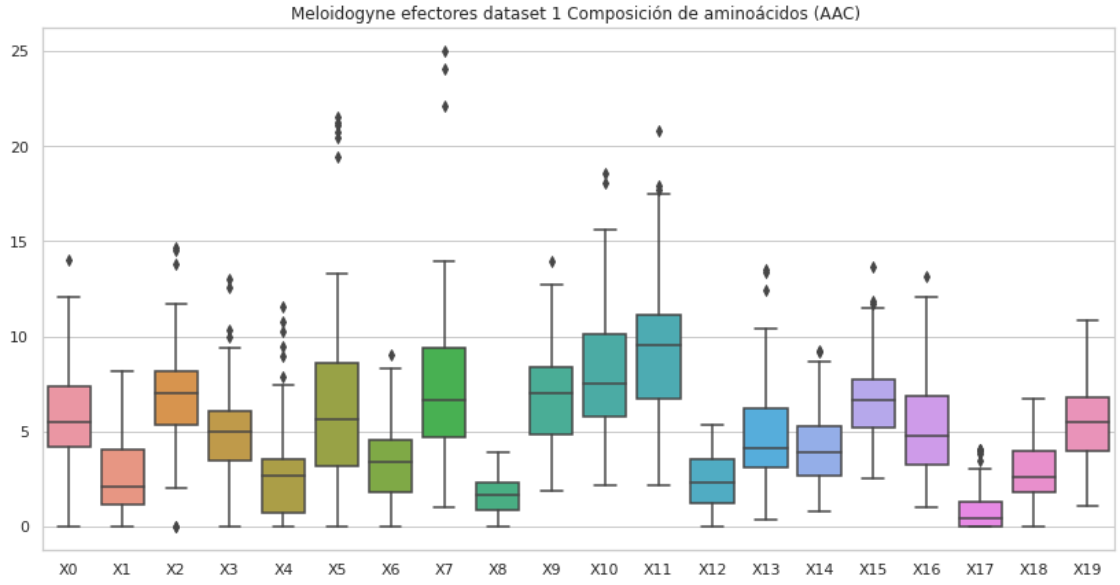
	X0	X1	X2	X3	X4	X5	\
count	409.000000	409.000000	409.000000	409.000000	409.000000	409.000000	
mean	4.844166	4.735579	6.997677	4.745998	2.207863	6.875892	
std	2.028092	1.955667	2.981323	1.828262	1.435209	2.752463	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	3.394000	3.497000	5.028000	3.492000	1.202000	5.075000	
50%	4.848000	4.713000	6.395000	4.633000	1.890000	6.711000	
75%	6.122000	6.000000	8.738000	6.019000	3.051000	8.391000	
max	11.009000	10.738000	16.129000	10.448000	7.750000	15.385000	

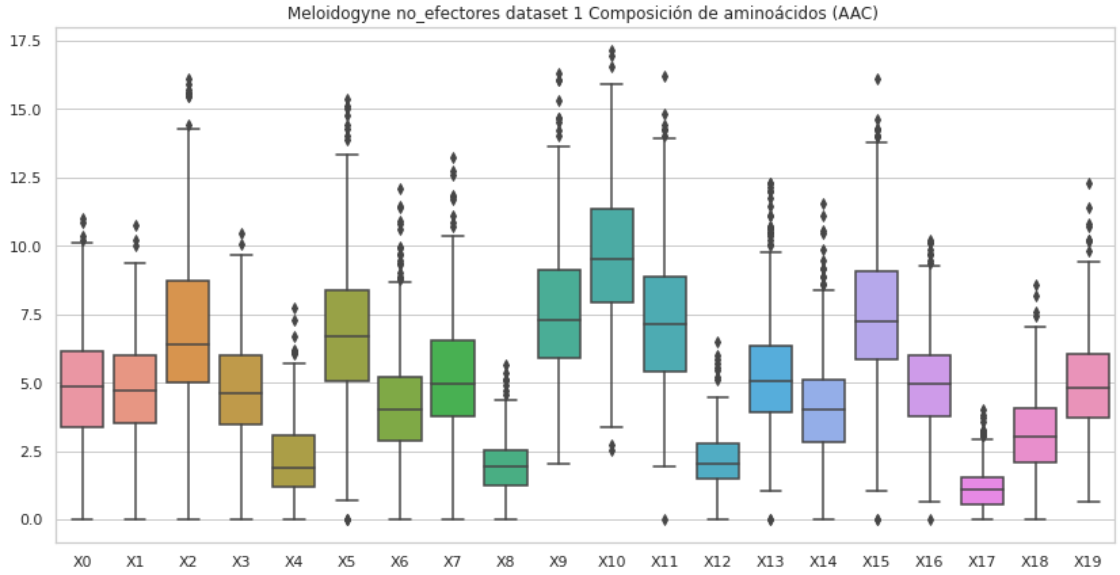
	X6	X7	X8	X9	X10	X11	\
count	409.000000	409.000000	409.000000	409.000000	409.000000	409.000000	
mean	4.237237	5.225892	1.965399	7.632924	9.602071	7.315660	
std	2.057767	2.222668	1.003343	2.459172	2.629077	2.682171	
min	0.000000	0.000000	0.000000	2.041000	2.532000	0.000000	
25%	2.857000	3.774000	1.250000	5.904000	7.910000	5.421000	
50%	4.023000	4.958000	1.916000	7.296000	9.501000	7.143000	
75%	5.187000	6.542000	2.528000	9.123000	11.333000	8.861000	
max	12.069000	13.253000	5.625000	16.327000	17.143000	16.187000	

	X12	X13	X14	X15	X16	X17	\
count	409.000000	409.000000	409.000000	409.000000	409.000000	409.000000	
mean	2.150523	5.374281	4.167851	7.552176	5.045315	1.146941	
std	1.005446	2.149369	1.894023	2.551748	1.835050	0.820744	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	1.471000	3.937000	2.844000	5.851000	3.791000	0.568000	

50%	2.023000	5.065000	4.021000	7.243000	4.961000	1.068000
75%	2.753000	6.336000	5.120000	9.091000	5.990000	1.533000
max	6.475000	12.295000	11.538000	16.107000	10.204000	4.027000

	X18	X19
count	409.000000	409.000000
mean	3.174535	5.001919
std	1.558608	1.913924
min	0.000000	0.625000
25%	2.081000	3.738000
50%	3.000000	4.814000
75%	4.082000	6.051000
max	8.571000	12.281000





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

          X0        X1        X2        X3        X4        X5        X6  \
0   0.018089  0.002128  0.015961  0.027665  0.007448  0.008512  0.004256
1   0.007533  0.011300  0.020088  0.033899  0.018833  0.028877  0.008789
2   0.018464  0.025389  0.043853  0.013848  0.030005  0.055393  0.016156
3   0.021094  0.013184  0.026367  0.018457  0.013184  0.031640  0.007910
4   0.011986  0.011986  0.007191  0.011986  0.038354  0.009588  0.004794
..       …         …         …         …         …         …         …
95  0.028301  0.014151  0.067215  0.021226  0.021226  0.010613  0.003538
96  0.015365  0.011524  0.010243  0.011524  0.006402  0.023047  0.008963
97  0.026163  0.006541  0.019622  0.006541  0.013082  0.030524  0.002180
98  0.036651  0.000000  0.011454  0.006872  0.011454  0.034361  0.002291
99  0.027096  0.010422  0.006253  0.027096  0.016674  0.031265  0.004169

          X7        X8        X9  …       X74       X75       X76       X77  \
0   0.012769  0.044690  0.020217  … -0.009316  0.025479 -0.007277  0.006793
1   0.025110  0.025110  0.022599  …  0.000366  0.014015  0.003277 -0.014408
2   0.030005  0.018464  0.020773  … -0.002843 -0.011345  0.030853 -0.009755
3   0.029004  0.052734  0.026367  …  0.023597  0.039282  0.028849  0.004633
4   0.047943  0.021574  0.050340  … -0.017744 -0.021278  0.004536  0.025964
..       …         …         …  …         …         …         …         …
95  0.056602  0.067215  0.084903  …  0.032844  0.013685 -0.003060 -0.024645
96  0.016645  0.025608  0.025608  …  0.001883 -0.001411  0.026700 -0.008885
97  0.019622  0.023983  0.021803  … -0.008106  0.012377  0.021119 -0.020834
98  0.022907  0.006872  0.022907  …  0.010971  0.017764  0.011819  0.012884
99  0.031265  0.035433  0.027096  …  0.015019  0.018419  0.024560  0.012424

          X78       X79       X80       X81       X82        X83
0   0.029438 -0.010309  0.013829  0.031464  0.004764  efectores
1   0.006823  0.007498  0.007286  0.005308 -0.000583  efectores
2  -0.006535  0.031973 -0.009334  0.003945  0.030918  efectores
3   0.016806  0.023887  0.022568  0.026698  0.030483  efectores
4   0.007621 -0.005523  0.010482  0.018174 -0.025315  efectores
..       …         …         …         …         …          …
95 -0.019282  0.008118 -0.028192  0.002257 -0.005306  efectores
96 -0.005600  0.032131 -0.003610 -0.001811  0.033788  efectores
97 -0.010543  0.010942 -0.021709 -0.013243  0.021919  efectores
```

```
98  0.009370  0.010808 -0.005212 -0.012546  0.025260  efectores
99 -0.006581  0.035818 -0.028596 -0.012169  0.021654  efectores

[100 rows x 84 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	
mean	0.019270	0.010919	0.018609	0.021691	0.020699	0.028094	
std	0.017840	0.012906	0.016664	0.023106	0.022861	0.022491	
min	-0.082991	-0.020748	-0.020748	-0.062243	-0.062243	-0.082991	
25%	0.008584	0.001145	0.007208	0.008526	0.008044	0.014674	
50%	0.018681	0.007059	0.012735	0.016512	0.016469	0.028012	
75%	0.027134	0.017122	0.025707	0.027891	0.024652	0.036329	
max	0.083706	0.053948	0.077366	0.125559	0.132534	0.104632	

	X6	X7	X8	X9	…	X73	\
count	100.000000	100.000000	100.000000	100.000000	…	100.000000	
mean	0.006629	0.025442	0.031468	0.034053	…	0.016251	
std	0.008574	0.027478	0.026012	0.034033	…	0.021489	
min	-0.020748	-0.165982	-0.103739	-0.124487	…	-0.086561	
25%	0.001319	0.012437	0.017470	0.014744	…	0.002076	
50%	0.004159	0.025066	0.028315	0.025405	…	0.017137	
75%	0.009667	0.036513	0.044278	0.048428	…	0.029813	
max	0.051111	0.087619	0.132534	0.125559	…	0.084253	

	X74	X75	X76	X77	X78	X79	\
count	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	
mean	0.001514	0.006323	0.012776	0.001227	0.003721	0.015296	
std	0.033968	0.035260	0.026379	0.044536	0.044969	0.023408	
min	-0.110165	-0.061263	-0.075722	-0.082236	-0.122913	-0.115578	
25%	-0.007435	-0.007343	0.002913	-0.010093	-0.008052	0.002701	
50%	0.003147	0.008289	0.010937	0.000074	0.001027	0.015312	
75%	0.010923	0.020239	0.026189	0.007687	0.015752	0.028889	
max	0.226319	0.239278	0.181337	0.388708	0.388600	0.065022	

	X80	X81	X82
count	100.000000	100.000000	100.000000
mean	0.000834	0.005563	0.017972
std	0.035749	0.032044	0.040482
min	-0.066864	-0.062812	-0.040528
25%	-0.019623	-0.007945	0.001303
50%	0.002580	0.003034	0.015245
75%	0.013729	0.018546	0.028876

```
max      0.254767    0.203697    0.358751

[8 rows x 83 columns]
```

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.039065	0.017579	0.039065	0.050785	0.021486	0.025392	0.015626	
1	0.028372	0.014186	0.014186	0.052016	0.023644	0.033101	0.018915	
2	0.030464	0.030464	0.030464	0.000000	0.048743	0.036557	0.006093	
3	0.025232	0.008411	0.013457	0.029157	0.025793	0.021307	0.008411	
4	0.032107	0.064215	0.040134	0.144483	0.064215	0.064215	0.032107	
..	…	…	…	…	…	…	…	
495	0.032941	0.007751	0.013564	0.021315	0.029065	0.019377	0.011626	
496	0.018443	0.002049	0.022542	0.039961	0.006148	0.019468	0.009222	
497	0.022861	0.029136	0.031377	0.039894	0.027791	0.041239	0.011654	
498	0.010648	0.005324	0.026621	0.047917	0.047917	0.015973	0.015973	
499	0.025833	0.002672	0.020488	0.028506	0.015144	0.014253	0.003563	

	X7	X8	X9	…	X74	X75	X76	\
0	0.060551	0.064458	0.097663	…	0.052215	0.030363	0.020696	
1	0.042559	0.066202	0.037830	…	0.040588	-0.015780	0.025957	
2	0.030464	0.042650	0.060929	…	0.026437	0.003766	-0.035026	
3	0.025793	0.018504	0.043175	…	-0.008101	-0.004952	0.009563	
4	0.072242	0.096322	0.104349	…	0.012182	-0.025465	-0.018115	
..	…	…	…	…	…	…	…	
495	0.013564	0.023252	0.040692	…	0.015004	0.009584	0.014769	
496	0.031763	0.024591	0.033813	…	0.017690	0.014198	0.007955	
497	0.040342	0.033170	0.038101	…	-0.007846	-0.003829	0.031113	
498	0.037269	0.042593	0.058566	…	0.040889	-0.017555	0.037387	
499	0.016925	0.046322	0.021379	…	-0.016852	0.012893	0.006053	

	X77	X78	X79	X80	X81	X82	X83
0	-0.006135	0.029009	0.022644	-0.039227	-0.004298	0.018084	no_efectores
1	0.018748	0.054556	0.019468	0.006973	0.030021	0.009896	no_efectores
2	0.002829	-0.003823	0.038196	-0.015821	0.015996	0.054366	no_efectores
3	0.003408	0.004060	0.021043	0.010314	0.003666	0.016709	no_efectores
4	0.021946	0.089238	0.009704	-0.011997	-0.002453	0.020929	no_efectores
..	…	…	…	…	…	…	…
495	0.016505	0.015909	-0.008704	0.006972	0.013036	-0.006538	no_efectores
496	-0.003684	0.012564	-0.004337	0.000596	0.004875	0.004450	no_efectores
497	0.006318	0.006953	0.010578	0.007409	0.010737	0.010231	no_efectores
498	-0.004032	-0.009744	-0.006898	0.025509	0.015437	0.001452	no_efectores

```
499 -0.010505  0.009681  0.002419 -0.009739  0.013010  0.007026  no_efectores

[500 rows x 84 columns]
```

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

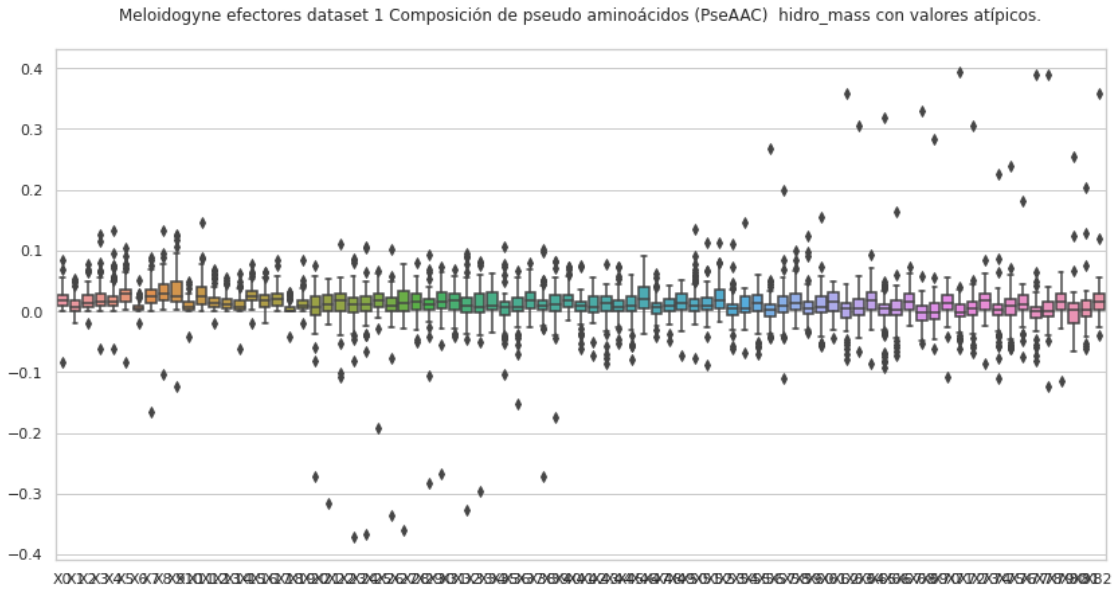
	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.026963	0.013859	0.027932	0.038157	0.033688	0.030091	
std	0.020760	0.017020	0.025602	0.026700	0.036047	0.024755	
min	-0.000000	-0.000000	-0.051522	-0.051522	-0.051522	-0.103043	
25%	0.013710	0.004324	0.013470	0.021273	0.015185	0.016271	
50%	0.023792	0.008727	0.023098	0.034290	0.026275	0.024822	
75%	0.034485	0.018767	0.036097	0.048900	0.040688	0.038433	
max	0.217528	0.197369	0.351501	0.174978	0.479320	0.270046	

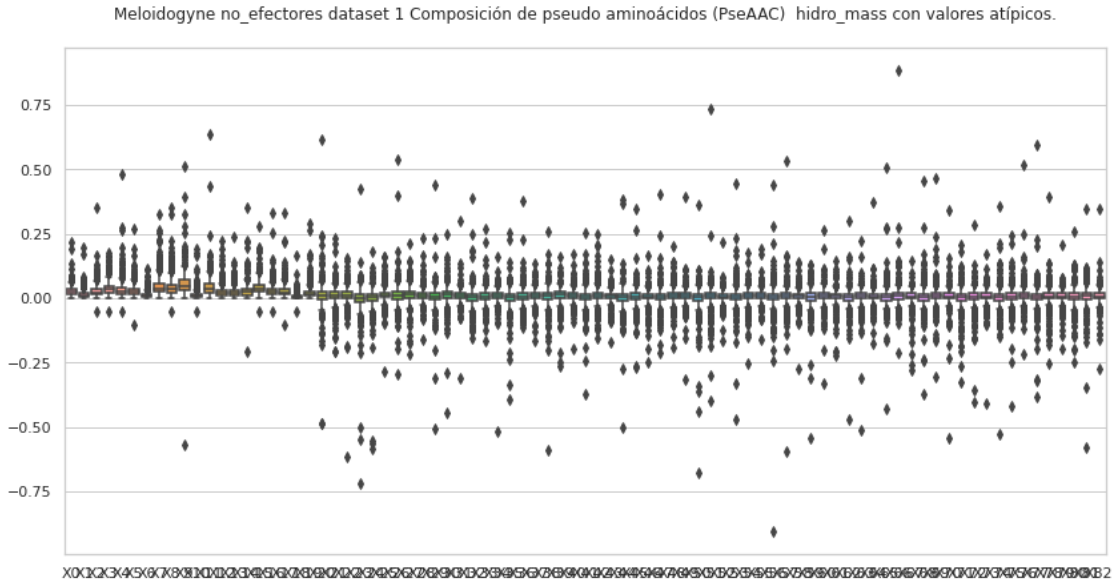
	X6	X7	X8	X9	…	X73	\
count	500.000000	500.000000	500.000000	500.000000	…	500.000000	
mean	0.012302	0.047278	0.043529	0.055393	…	0.006876	
std	0.012975	0.038881	0.036376	0.053370	…	0.029729	
min	-0.000000	-0.000000	-0.051522	-0.566738	…	-0.409195	
25%	0.005282	0.024310	0.022694	0.031736	…	-0.001808	
50%	0.009636	0.037560	0.034800	0.046941	…	0.007278	
75%	0.015124	0.057764	0.052988	0.070443	…	0.018971	
max	0.115630	0.326293	0.351501	0.511275	…	0.172452	

	X74	X75	X76	X77	X78	X79	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.002434	0.006203	0.010205	0.002256	0.006648	0.006742	
std	0.043510	0.037661	0.036217	0.048740	0.037238	0.025010	
min	-0.527856	-0.418718	-0.218851	-0.381996	-0.254432	-0.195166	
25%	-0.008514	-0.003346	-0.000640	-0.006939	-0.001146	-0.001265	
50%	0.003723	0.007031	0.007977	0.004022	0.007513	0.006813	
75%	0.014883	0.018360	0.019319	0.015747	0.018398	0.018108	
max	0.358378	0.244474	0.517042	0.593238	0.392842	0.208875	

	X80	X81	X82
count	500.000000	500.000000	500.000000
mean	0.003028	0.006828	0.007513
std	0.037249	0.045592	0.031418
min	-0.246064	-0.578207	-0.276155
25%	-0.007639	-0.003703	-0.002538
50%	0.004830	0.008010	0.007360
75%	0.018588	0.020703	0.018601
max	0.257154	0.346554	0.345800

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.018089	0.002128	0.015961	0.027665	0.007448	0.008512	0.004256	
1	0.007533	0.011300	0.020088	0.033899	0.018833	0.028877	0.008789	
2	0.018464	0.025389	0.043853	0.013848	0.030005	0.055393	0.016156	
3	0.021094	0.013184	0.026367	0.018457	0.013184	0.031640	0.007910	
4	0.011986	0.011986	0.007191	0.011986	0.038354	0.009588	0.004794	
..	…	…	…	…	…	…	…	
93	0.008825	0.002942	0.008825	0.005884	0.017651	0.032359	0.017651	
95	0.028301	0.014151	0.067215	0.021226	0.021226	0.010613	0.003538	
97	0.026163	0.006541	0.019622	0.006541	0.013082	0.030524	0.002180	
98	0.036651	0.000000	0.011454	0.006872	0.011454	0.034361	0.002291	
99	0.027096	0.010422	0.006253	0.027096	0.016674	0.031265	0.004169	

	X7	X8	X9	…	X74	X75	X76	X77	\
0	0.012769	0.044690	0.020217	…	-0.009316	0.025479	-0.007277	0.006793	
1	0.025110	0.025110	0.022599	…	0.000366	0.014015	0.003277	-0.014408	
2	0.030005	0.018464	0.020773	…	-0.002843	-0.011345	0.030853	-0.009755	
3	0.029004	0.052734	0.026367	…	0.023597	0.039282	0.028849	0.004633	
4	0.047943	0.021574	0.050340	…	-0.017744	-0.021278	0.004536	0.025964	
..	…	…	…	…	…	…	…	…	
93	0.017651	0.008825	0.026476	…	0.009135	0.004868	0.029275	-0.001566	
95	0.056602	0.067215	0.084903	…	0.032844	0.013685	-0.003060	-0.024645	
97	0.019622	0.023983	0.021803	…	-0.008106	0.012377	0.021119	-0.020834	
98	0.022907	0.006872	0.022907	…	0.010971	0.017764	0.011819	0.012884	
99	0.031265	0.035433	0.027096	…	0.015019	0.018419	0.024560	0.012424	

	X78	X79	X80	X81	X82	X83
0	0.029438	-0.010309	0.013829	0.031464	0.004764	efectores
1	0.006823	0.007498	0.007286	0.005308	-0.000583	efectores
2	-0.006535	0.031973	-0.009334	0.003945	0.030918	efectores
3	0.016806	0.023887	0.022568	0.026698	0.030483	efectores
4	0.007621	-0.005523	0.010482	0.018174	-0.025315	efectores
..	…	…	…	…	…	
93	0.005680	0.026699	0.013656	0.009100	0.021343	efectores
95	-0.019282	0.008118	-0.028192	0.002257	-0.005306	efectores
97	-0.010543	0.010942	-0.021709	-0.013243	0.021919	efectores
98	0.009370	0.010808	-0.005212	-0.012546	0.025260	efectores
99	-0.006581	0.035818	-0.028596	-0.012169	0.021654	efectores

[76 rows x 84 columns]

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

	X0	X1	X2	X3	X4	X5	\
count	76.000000	76.000000	76.000000	76.000000	76.000000	76.000000	
mean	0.018196	0.008867	0.015771	0.019557	0.015636	0.024764	
std	0.010589	0.010456	0.012656	0.014241	0.012331	0.013804	
min	0.000000	0.000000	0.000000	0.000000	0.000305	0.001831	
25%	0.009089	0.000720	0.006913	0.008630	0.007390	0.014596	
50%	0.018277	0.004914	0.011466	0.016512	0.013412	0.027521	
75%	0.026008	0.011975	0.023411	0.026181	0.020807	0.031820	
max	0.040409	0.049391	0.067215	0.066990	0.055115	0.061023	

	X6	X7	X8	X9	…	X73	X74	\
count	76.000000	76.000000	76.000000	76.000000	…	76.000000	76.000000	
mean	0.005607	0.022641	0.029015	0.026132	…	0.015700	0.004511	
std	0.005772	0.014918	0.017063	0.021144	…	0.015548	0.014043	
min	0.000000	0.001006	0.001350	0.001350	…	-0.019232	-0.041699	
25%	0.001601	0.007195	0.017395	0.009560	…	0.002076	-0.003661	
50%	0.003779	0.022975	0.026986	0.022864	…	0.017137	0.005410	
75%	0.007987	0.031624	0.037811	0.030163	…	0.028342	0.011850	
max	0.029389	0.058777	0.069798	0.088044	…	0.061959	0.037492	

	X75	X76	X77	X78	X79	X80	\
count	76.000000	76.000000	76.000000	76.000000	76.000000	76.000000	
mean	0.010758	0.014211	0.000156	0.005316	0.016244	-0.000627	
std	0.017369	0.014629	0.015269	0.016766	0.015575	0.016622	
min	-0.037523	-0.023311	-0.060083	-0.037520	-0.018925	-0.049454	
25%	0.001343	0.004254	-0.006971	-0.003574	0.004905	-0.008412	
50%	0.011061	0.011196	0.002377	0.004318	0.016788	0.003498	
75%	0.022366	0.025379	0.008769	0.017274	0.026305	0.012764	
max	0.055671	0.045388	0.031005	0.039354	0.062402	0.029720	

	X81	X82
count	76.000000	76.000000
mean	0.005447	0.015548
std	0.018638	0.015784
min	-0.052154	-0.025315
25%	-0.004721	0.002823
50%	0.004089	0.021160
75%	0.021073	0.027772
max	0.034333	0.048968

[8 rows x 83 columns]

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.039065  0.017579  0.039065  0.050785  0.021486  0.025392  0.015626
1    0.028372  0.014186  0.014186  0.052016  0.023644  0.033101  0.018915
2    0.030464  0.030464  0.030464  0.000000  0.048743  0.036557  0.006093
3    0.025232  0.008411  0.013457  0.029157  0.025793  0.021307  0.008411
5    0.023709  0.011290  0.030484  0.031613  0.024838  0.029354  0.006774
..        …         …         …         …         …         …         …
495  0.032941  0.007751  0.013564  0.021315  0.029065  0.019377  0.011626
496  0.018443  0.002049  0.022542  0.039961  0.006148  0.019468  0.009222
497  0.022861  0.029136  0.031377  0.039894  0.027791  0.041239  0.011654
498  0.010648  0.005324  0.026621  0.047917  0.047917  0.015973  0.015973
499  0.025833  0.002672  0.020488  0.028506  0.015144  0.014253  0.003563

           X7        X8        X9  …       X74       X75       X76  \
0    0.060551  0.064458  0.097663  …  0.052215  0.030363  0.020696
1    0.042559  0.066202  0.037830  …  0.040588 -0.015780  0.025957
2    0.030464  0.042650  0.060929  …  0.026437  0.003766 -0.035026
3    0.025793  0.018504  0.043175  … -0.008101 -0.004952  0.009563
5    0.038387  0.035000  0.053064  …  0.009847  0.017747  0.005199
..        …         …         …  …         …         …         …
495  0.013564  0.023252  0.040692  …  0.015004  0.009584  0.014769
496  0.031763  0.024591  0.033813  …  0.017690  0.014198  0.007955
497  0.040342  0.033170  0.038101  … -0.007846 -0.003829  0.031113
498  0.037269  0.042593  0.058566  …  0.040889 -0.017555  0.037387
499  0.016925  0.046322  0.021379  … -0.016852  0.012893  0.006053

           X77       X78       X79       X80       X81       X82           X83
0   -0.006135  0.029009  0.022644 -0.039227 -0.004298  0.018084  no_efectores
1    0.018748  0.054556  0.019468  0.006973  0.030021  0.009896  no_efectores
2    0.002829 -0.003823  0.038196 -0.015821  0.015996  0.054366  no_efectores
3    0.003408  0.004060  0.021043  0.010314  0.003666  0.016709  no_efectores
5   -0.002757  0.011311  0.027000  0.011323  0.015667  0.015941  no_efectores
..        …         …         …         …         …         …           …
495  0.016505  0.015909 -0.008704  0.006972  0.013036 -0.006538  no_efectores
496 -0.003684  0.012564 -0.004337  0.000596  0.004875  0.004450  no_efectores
497  0.006318  0.006953  0.010578  0.007409  0.010737  0.010231  no_efectores
498 -0.004032 -0.009744 -0.006898  0.025509  0.015437  0.001452  no_efectores
499 -0.010505  0.009681  0.002419 -0.009739  0.013010  0.007026  no_efectores

[438 rows x 84 columns]
```

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

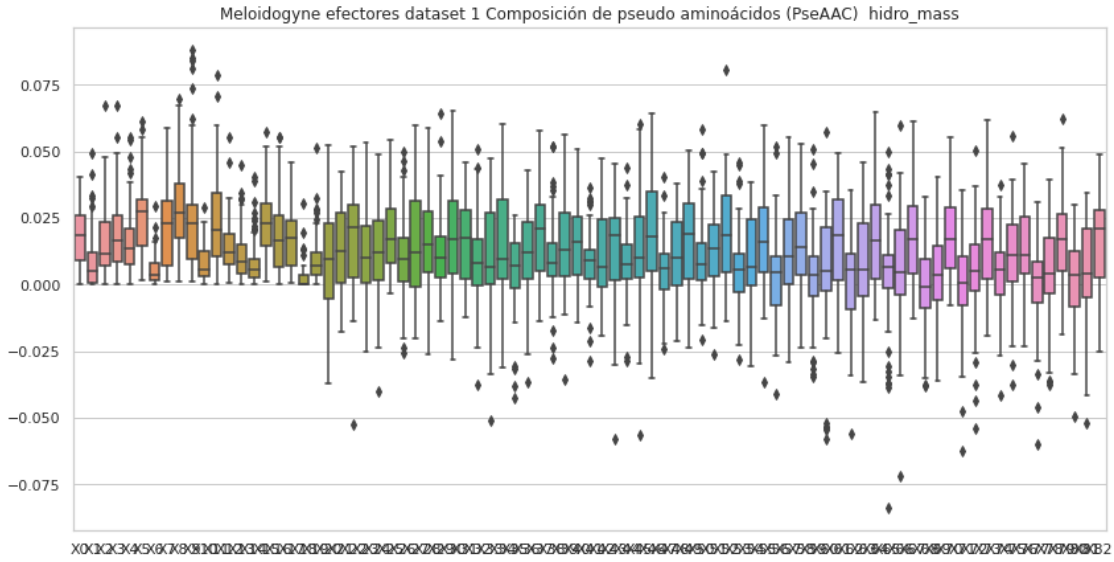
	X0	X1	X2	X3	X4	X5	\
count	438.000000	438.000000	438.000000	438.000000	438.000000	438.000000	
mean	0.024078	0.011587	0.024138	0.033629	0.026660	0.025960	
std	0.013889	0.010469	0.014473	0.018453	0.016798	0.014373	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.013126	0.004123	0.013254	0.020456	0.014717	0.015938	
50%	0.022994	0.008160	0.021631	0.032123	0.023772	0.023878	
75%	0.031690	0.016132	0.033672	0.045717	0.034454	0.034510	
max	0.077664	0.057438	0.072762	0.109777	0.112471	0.081070	

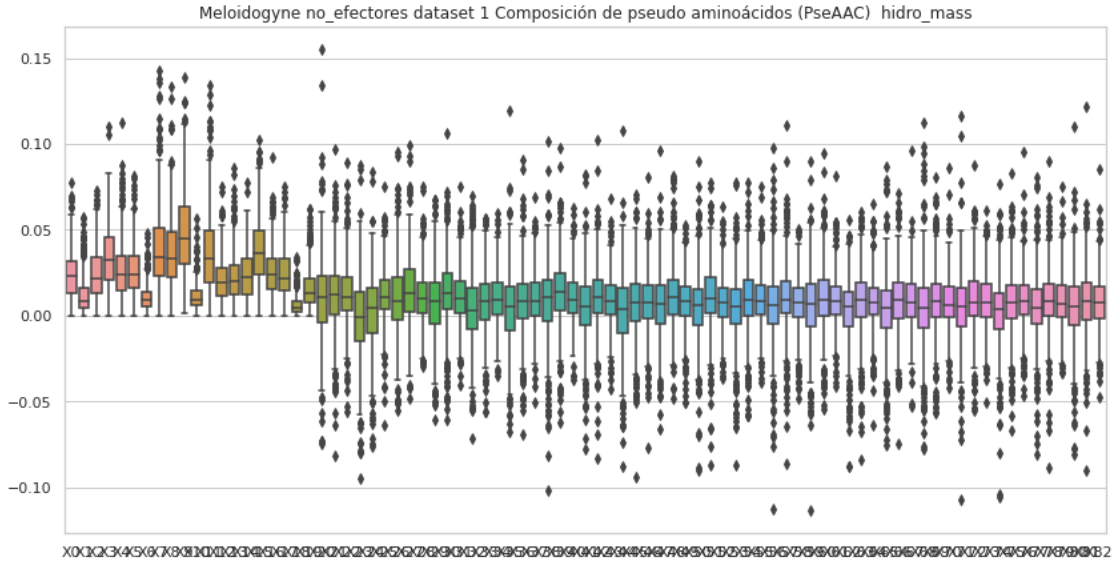
	X6	X7	X8	X9	…	X73	\
count	438.000000	438.000000	438.000000	438.000000	…	438.000000	
mean	0.010331	0.038879	0.036415	0.048169	…	0.008667	
std	0.007482	0.023891	0.020748	0.025653	…	0.015516	
min	0.000000	0.000000	0.000000	0.001394	…	-0.050973	
25%	0.005270	0.022883	0.022259	0.029834	…	-0.000962	
50%	0.009069	0.033958	0.033061	0.044523	…	0.007517	
75%	0.013954	0.050725	0.048771	0.063195	…	0.018173	
max	0.047972	0.142640	0.133437	0.138604	…	0.061202	

	X74	X75	X76	X77	X78	X79	\
count	438.000000	438.000000	438.000000	438.000000	438.000000	438.000000	
mean	0.002140	0.007046	0.009435	0.004485	0.009006	0.008961	
std	0.020892	0.018680	0.016243	0.020028	0.018803	0.014967	
min	-0.105662	-0.066868	-0.046951	-0.081143	-0.088923	-0.033767	
25%	-0.007802	-0.001722	0.000857	-0.004468	-0.000011	-0.000641	
50%	0.003613	0.007471	0.008306	0.004521	0.008481	0.007078	
75%	0.013045	0.017847	0.018048	0.015012	0.018325	0.018025	
max	0.077110	0.083095	0.095480	0.081012	0.090026	0.075025	

	X80	X81	X82
count	438.000000	438.000000	438.000000
mean	0.004331	0.008598	0.008217
std	0.021681	0.019934	0.016028
min	-0.081257	-0.090200	-0.047893
25%	-0.005956	-0.002330	-0.001351
50%	0.005136	0.008190	0.007260
75%	0.016774	0.019161	0.016726
max	0.110162	0.121556	0.085334

[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))
```

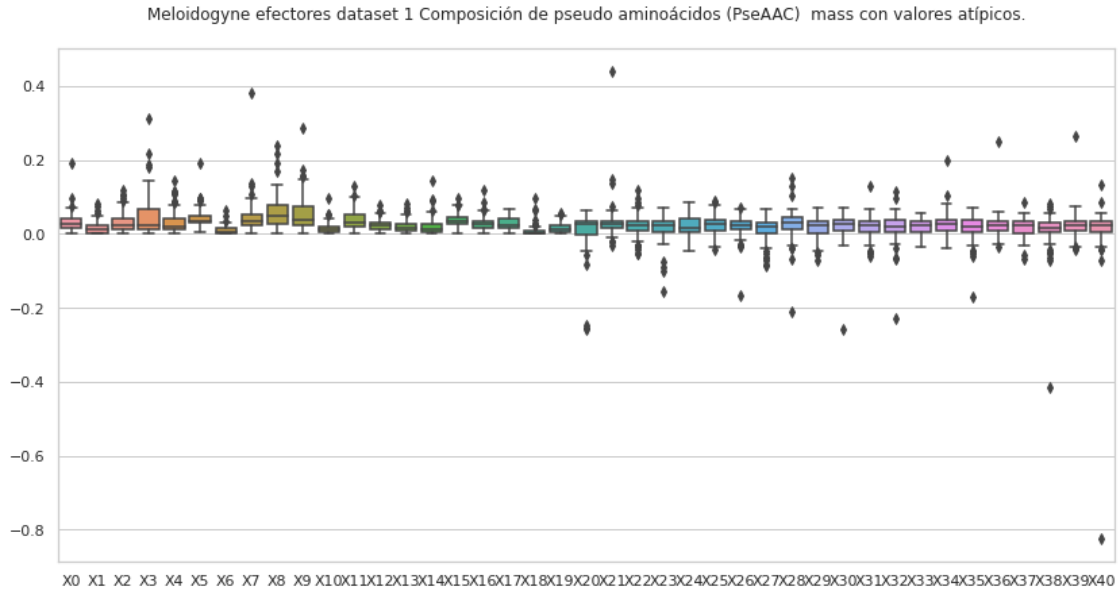
```
efectores
Composición de pseudo aminoácidos (PseAAC)  mass efectores Meloidogyne dataset
1, con valores atípicos.
Valores del documento csv.

          X0        X1        X2        X3        X4        X5        X6  \
0   0.068085  0.008010  0.060075  0.104130  0.028035  0.032040  0.016020
1   0.018035  0.027052  0.048093  0.081157  0.045087  0.069134  0.021041
2   0.017770  0.024433  0.042203  0.013327  0.028876  0.053309  0.015548
3   0.021619  0.013512  0.027024  0.018917  0.013512  0.032429  0.008107
4   0.034475  0.034475  0.020685  0.034475  0.110320  0.027580  0.013790
..       …         …         …         …         …         …         …
95  0.040175  0.020088  0.095416  0.030132  0.030132  0.015066  0.005022
96  0.015383  0.011538  0.010256  0.011538  0.006410  0.023075  0.008974
97  0.030023  0.007506  0.022517  0.007506  0.015012  0.035027  0.002502
98  0.051194  0.000000  0.015998  0.009599  0.015998  0.047994  0.003200
99  0.028961  0.011139  0.006683  0.028961  0.017822  0.033417  0.004456

          X7        X8        X9  …       X32       X33       X34       X35  \
0   0.048060  0.168211  0.076095  …  0.023088  0.040040  0.008774 -0.045788
```

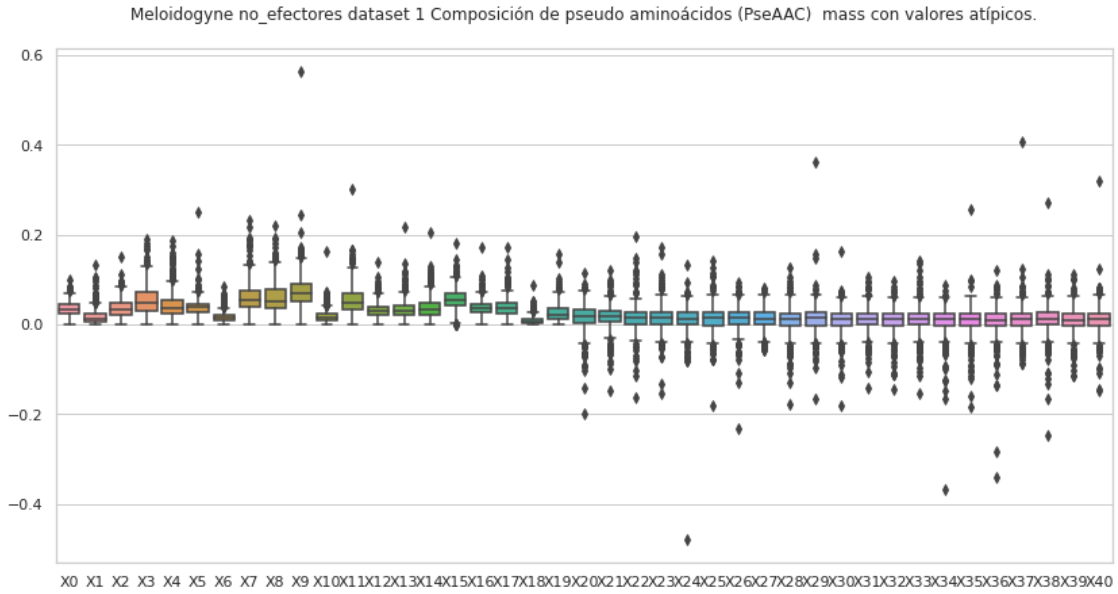
```
1   0.060116  0.060116  0.054105  …  0.020824  0.024052 -0.017308 -0.011172
2   0.028876  0.017770  0.019991  …  0.001470  0.035289  0.040876  0.052135
3   0.029726  0.054048  0.027024  …  0.013181  0.026930  0.020476  0.022721
4   0.137900  0.062055  0.144795  …  0.045283  0.037764  0.002932  0.020837
..       …         …        …  …        …         …         …         …
95  0.080351  0.095416  0.120526  … -0.001153 -0.005857 -0.018664  0.040327
96  0.016665  0.025639  0.025639  …  0.021306  0.018779  0.043689  0.031970
97  0.022517  0.027521  0.025019  …  0.035698  0.034458  0.029946  0.009651
98  0.031996  0.009599  0.031996  …  0.030078  0.012167 -0.009185  0.011758
99  0.033417  0.037872  0.028961  …  0.019347  0.023954  0.032441  0.040017

          X36       X37       X38       X39       X40        X41
0   0.031213  0.043926 -0.027391 -0.038804  0.017931  efectores
1   0.007770  0.032816  0.007845  0.017950 -0.001396  efectores
2   0.043505  0.036778  0.029692  0.030770  0.029755  efectores
3   0.038526  0.029827  0.029568  0.024482  0.031242  efectores
4   0.006395  0.019518  0.013047 -0.015887 -0.072814  efectores
..       …         …         …         …         …         …
95 -0.011422  0.045603 -0.004344  0.011523 -0.007532  efectores
96  0.038839  0.041756  0.026733  0.032170  0.033829  efectores
97  0.035498  0.022932  0.024234  0.012556  0.025153  efectores
98  0.025093 -0.001299  0.016508  0.015096  0.035283  efectores
99  0.021038  0.021369  0.026250  0.038284  0.023144  efectores

[100 rows x 42 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  100.000000  100.000000  100.000000  100.000000  100.000000  100.000000
mean     0.031555    0.015067    0.030606    0.045072    0.031409    0.041753
std      0.024048    0.016509    0.025130    0.052296    0.028689    0.024460
min      0.000000    0.000000    0.000000    0.000000    0.001479    0.006094
25%      0.017935    0.003238    0.012059    0.011530    0.012050    0.029452
50%      0.027386    0.011022    0.025213    0.024859    0.021133    0.034592
75%      0.040714    0.022588    0.042885    0.065671    0.040463    0.049833
max      0.190589    0.081043    0.118478    0.312351    0.142942    0.190589

                X6          X7          X8          X9  …         X31  \
count  100.000000  100.000000  100.000000  100.000000  …  100.000000
mean     0.009989    0.043057    0.057227    0.054430  …    0.017774
std      0.010744    0.043124    0.043666    0.046608  …    0.028584
min      0.000000    0.002973    0.001692    0.001692  …   -0.059173
25%      0.002495    0.023280    0.028824    0.024630  …    0.004371
50%      0.006693    0.034297    0.048172    0.036550  …    0.024305
```

```
75%      0.015612    0.053866    0.078996    0.075571  …    0.033487
max      0.063459    0.381179    0.238237    0.285884  …    0.129813

              X32         X33         X34         X35         X36         X37  \
count  100.000000  100.000000  100.000000  100.000000  100.000000  100.000000
mean     0.018779    0.018844    0.025265    0.017648    0.021945    0.019712
std      0.036920    0.021673    0.029617    0.030834    0.029798    0.023517
min     -0.227387   -0.036745   -0.038888   -0.168867   -0.036384   -0.069442
25%      0.006812    0.005404    0.007166    0.006287    0.007535    0.003048
50%      0.020323    0.023697    0.026654    0.020814    0.022450    0.021919
75%      0.037180    0.034441    0.039665    0.037744    0.034632    0.035021
max      0.114928    0.056602    0.198028    0.070558    0.248838    0.086040

              X38         X39         X40
count  100.000000  100.000000  100.000000
mean     0.011288    0.021758    0.009772
std      0.050643    0.034324    0.088392
min     -0.416441   -0.042773   -0.823872
25%      0.005212    0.007634    0.003552
50%      0.016750    0.023982    0.025135
75%      0.029309    0.035184    0.033205
max      0.080786    0.265426    0.133728

[8 rows x 41 columns]



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

          X0        X1        X2        X3        X4        X5        X6  \
0    0.040282  0.018127  0.040282  0.052366  0.022155  0.026183  0.016113
1    0.037920  0.018960  0.018960  0.069519  0.031600  0.044240  0.025280
2    0.027369  0.027369  0.027369  0.000000  0.043790  0.032843  0.005474
3    0.031086  0.010362  0.016579  0.035922  0.031777  0.026251  0.010362
4    0.026074  0.052149  0.032593  0.117335  0.052149  0.052149  0.026074
..        …         …         …         …         …         …         …
495  0.100832  0.023725  0.041519  0.065244  0.088970  0.059313  0.035588
496  0.034479  0.003831  0.042142  0.074705  0.011493  0.036395  0.017240
497  0.026948  0.034345  0.036987  0.047026  0.032760  0.048612  0.013738
498  0.012520  0.006260  0.031300  0.056341  0.056341  0.018780  0.018780
499  0.055038  0.005694  0.043651  0.060732  0.032264  0.030366  0.007591

          X7        X8        X9  …       X32       X33       X34  \
0    0.062437  0.066465  0.100705  …  0.016491  0.008064 -0.000392
1    0.056879  0.088479  0.050560  … -0.006368 -0.046941  0.015923
```

```
2    0.027369  0.038317  0.054738  …  0.043934  0.029153  0.008909
3    0.031777  0.022797  0.053192  …  0.022062  0.027379  0.017751
4    0.058667  0.078223  0.084742  …  0.009360  0.002659 -0.016512
..        …         …         …  …         …         …
495  0.041519  0.071176  0.124557  …  0.000393 -0.020493  0.007540
496  0.059381  0.045973  0.063212  …  0.006448 -0.008978 -0.005994
497  0.047555  0.039101  0.044913  …  0.016559  0.014200  0.004912
498  0.043821  0.050081  0.068861  …  0.029037 -0.004820  0.023251
499  0.036060  0.098689  0.045549  …  0.011292  0.017977  0.021060

          X35       X36       X37       X38       X39       X40           X41
0    0.021723  0.009806  0.023432  0.021341  0.023350  0.018648  no_efectores
1    0.022599 -0.019049 -0.001007  0.034691  0.026019  0.013226  no_efectores
2    0.049489  0.013194 -0.006174 -0.031467  0.034315  0.048842  no_efectores
3    0.028778  0.014893  0.010081  0.011782  0.025925  0.020586  no_efectores
4    0.016366 -0.002329  0.020651 -0.014711  0.007880  0.016996  no_efectores
..        …         …         …         …         …         …           …
495 -0.005534 -0.015425  0.004925  0.045209 -0.026643 -0.020014  no_efectores
496  0.008998  0.014549  0.021219  0.014871 -0.008108  0.008320  no_efectores
497  0.018309  0.005712  0.021756  0.036676  0.012469  0.012060  no_efectores
498  0.030577  0.057698  0.071960  0.043959 -0.008110  0.001708  no_efectores
499  0.013486  0.008780  0.011943  0.012896  0.005154  0.014969  no_efectores

[500 rows x 42 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.034502    0.017358    0.035806    0.053235    0.043336    0.038838
std      0.015269    0.016191    0.019200    0.031950    0.028852    0.019865
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.024167    0.007222    0.021809    0.030267    0.024066    0.027339
50%      0.034515    0.013322    0.033584    0.048738    0.035348    0.037602
75%      0.043920    0.023497    0.047052    0.070782    0.054113    0.045568
max      0.100832    0.131126    0.149505    0.189740    0.187437    0.249250

                X6          X7          X8          X9  …         X31  \
count  500.000000  500.000000  500.000000  500.000000  …  500.000000
mean     0.015590    0.060952    0.058836    0.073912  …    0.011424
std      0.011401    0.034181    0.034036    0.039605  …    0.026271
min      0.000000    0.000000    0.000000    0.000000  …   -0.140764
25%      0.008447    0.037851    0.034628    0.051065  …   -0.000837
50%      0.013694    0.054111    0.051174    0.070149  …    0.013429
75%      0.019913    0.076080    0.077792    0.090600  …    0.025217
```

```
max      0.083492    0.232845    0.219093    0.563348  …    0.104742

              X32         X33         X34         X35         X36         X37  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.008325    0.009678    0.007648    0.008294    0.006996    0.010632
std      0.027029    0.029087    0.032075    0.030727    0.033140    0.031379
min     -0.144368   -0.153287   -0.369273   -0.184470   -0.340423   -0.088924
25%     -0.003151   -0.001353   -0.002204   -0.003571   -0.002773   -0.002988
50%      0.012126    0.013179    0.010793    0.011481    0.010456    0.010873
75%      0.022777    0.024128    0.023259    0.023307    0.023438    0.024388
max      0.095172    0.140471    0.086245    0.255579    0.119569    0.406748

              X38         X39         X40
count  500.000000  500.000000  500.000000
mean     0.010771    0.008720    0.010251
std      0.032526    0.026049    0.030993
min     -0.247100   -0.115868   -0.147952
25%     -0.000952   -0.002961   -0.003591
50%      0.012868    0.010198    0.011153
75%      0.026003    0.024086    0.025373
max      0.270151    0.110837    0.319170

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

          X0        X1        X2        X3        X4        X5        X6  \
0   0.068085  0.008010  0.060075  0.104130  0.028035  0.032040  0.016020
1   0.018035  0.027052  0.048093  0.081157  0.045087  0.069134  0.021041
2   0.017770  0.024433  0.042203  0.013327  0.028876  0.053309  0.015548
3   0.021619  0.013512  0.027024  0.018917  0.013512  0.032429  0.008107
5   0.042880  0.000000  0.028586  0.021440  0.038115  0.019058  0.019058
..       …         …         …         …         …         …         …
95  0.040175  0.020088  0.095416  0.030132  0.030132  0.015066  0.005022
96  0.015383  0.011538  0.010256  0.011538  0.006410  0.023075  0.008974
97  0.030023  0.007506  0.022517  0.007506  0.015012  0.035027  0.002502
98  0.051194  0.000000  0.015998  0.009599  0.015998  0.047994  0.003200
99  0.028961  0.011139  0.006683  0.028961  0.017822  0.033417  0.004456

          X7        X8        X9  …       X32       X33       X34       X35  \
0   0.048060  0.168211  0.076095  …  0.023088  0.040040  0.008774 -0.045788
1   0.060116  0.060116  0.054105  …  0.020824  0.024052 -0.017308 -0.011172
2   0.028876  0.017770  0.019991  …  0.001470  0.035289  0.040876  0.052135
3   0.029726  0.054048  0.027024  …  0.013181  0.026930  0.020476  0.022721
5   0.050026  0.071466  0.059555  …  0.039835  0.026933  0.007172  0.009079
..       …         …         …  …         …         …         …         …
95  0.080351  0.095416  0.120526  … -0.001153 -0.005857 -0.018664  0.040327
96  0.016665  0.025639  0.025639  …  0.021306  0.018779  0.043689  0.031970
97  0.022517  0.027521  0.025019  …  0.035698  0.034458  0.029946  0.009651
98  0.031996  0.009599  0.031996  …  0.030078  0.012167 -0.009185  0.011758
99  0.033417  0.037872  0.028961  …  0.019347  0.023954  0.032441  0.040017
```

```
          X36       X37       X38       X39       X40        X41
0    0.031213  0.043926 -0.027391 -0.038804  0.017931  efectores
1    0.007770  0.032816  0.007845  0.017950 -0.001396  efectores
2    0.043505  0.036778  0.029692  0.030770  0.029755  efectores
3    0.038526  0.029827  0.029568  0.024482  0.031242  efectores
5    0.019174  0.002615  0.006795  0.027835  0.008214  efectores
..        …         …         …         …         …         …
95 -0.011422  0.045603 -0.004344  0.011523 -0.007532  efectores
96   0.038839  0.041756  0.026733  0.032170  0.033829  efectores
97   0.035498  0.022932  0.024234  0.012556  0.025153  efectores
98   0.025093 -0.001299  0.016508  0.015096  0.035283  efectores
99   0.021038  0.021369  0.026250  0.038284  0.023144  efectores

[87 rows x 42 columns]
```

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

```
              X0         X1         X2         X3         X4         X5  \
count  87.000000  87.000000  87.000000  87.000000  87.000000  87.000000
mean    0.029819   0.013853   0.027224   0.036762   0.027030   0.039306
std     0.017987   0.013674   0.021658   0.037476   0.022918   0.017702
min     0.000000   0.000000   0.000000   0.000000   0.001479   0.006094
25%     0.018012   0.003799   0.011207   0.011219   0.011320   0.029927
50%     0.027405   0.011139   0.022900   0.021854   0.020582   0.034489
75%     0.040001   0.021031   0.038309   0.057411   0.032188   0.047615
max     0.097944   0.062016   0.095416   0.180483   0.113193   0.097734

              X6         X7         X8         X9  …        X31        X32  \
count  87.000000  87.000000  87.000000  87.000000  …  87.000000  87.000000
mean    0.009112   0.035539   0.049657   0.045729  …   0.017966   0.024309
std     0.008307   0.022621   0.033437   0.035107  …   0.026137   0.021445
min     0.000000   0.002973   0.001692   0.001692  … -0.059173  -0.029379
25%     0.002721   0.020360   0.024832   0.021091  …   0.008228   0.009279
50%     0.006483   0.032717   0.041224   0.030536  …   0.026700   0.021669
75%     0.015453   0.045193   0.076278   0.066142  …   0.034519   0.038179
max     0.029341   0.130607   0.168211   0.174142  …   0.068542   0.096380

             X33        X34        X35        X36        X37        X38  \
count  87.000000  87.000000  87.000000  87.000000  87.000000  87.000000
mean    0.020188   0.024193   0.022875   0.022538   0.022482   0.019145
std     0.021852   0.023869   0.021492   0.017875   0.019963   0.022258
min    -0.036745  -0.020031  -0.045788  -0.036384  -0.030777  -0.048699
25%     0.006287   0.007606   0.010704   0.013809   0.011454   0.008261
50%     0.025772   0.027336   0.023107   0.025093   0.022932   0.021435
75%     0.035093   0.038923   0.039514   0.036303   0.035677   0.030360
```

```
max     0.056602   0.104173   0.070558   0.061125   0.086040   0.080786

             X39        X40
count  87.000000  87.000000
mean    0.021387   0.020639
std     0.022265   0.025834
min    -0.042773  -0.043462
25%     0.011268   0.007132
50%     0.024852   0.027250
75%     0.035362   0.033673
max     0.074633   0.133728

[8 rows x 41 columns]
```

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

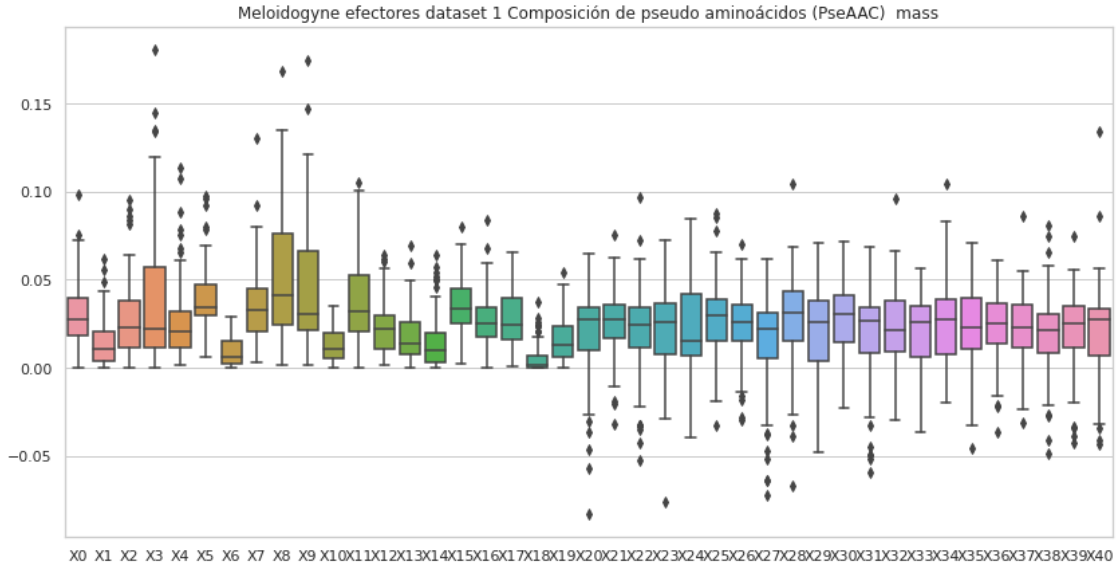
	X0	X1	X2	X3	X4	X5	X6	\
0	0.040282	0.018127	0.040282	0.052366	0.022155	0.026183	0.016113	
1	0.037920	0.018960	0.018960	0.069519	0.031600	0.044240	0.025280	
2	0.027369	0.027369	0.027369	0.000000	0.043790	0.032843	0.005474	
3	0.031086	0.010362	0.016579	0.035922	0.031777	0.026251	0.010362	
4	0.026074	0.052149	0.032593	0.117335	0.052149	0.052149	0.026074	
..	…	…	…	…	…	…	…	
492	0.037580	0.008221	0.022313	0.029360	0.017616	0.021139	0.021139	
494	0.037221	0.011754	0.033303	0.060729	0.035262	0.041139	0.017631	
496	0.034479	0.003831	0.042142	0.074705	0.011493	0.036395	0.017240	
497	0.026948	0.034345	0.036987	0.047026	0.032760	0.048612	0.013738	
499	0.055038	0.005694	0.043651	0.060732	0.032264	0.030366	0.007591	

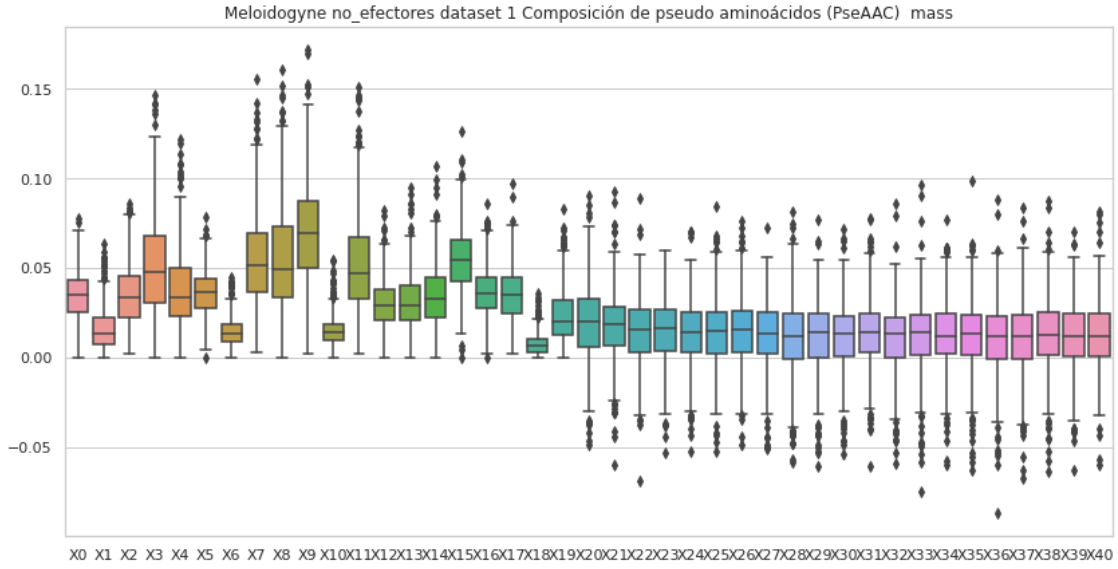
	X7	X8	X9	…	X32	X33	X34	\
0	0.062437	0.066465	0.100705	…	0.016491	0.008064	-0.000392	
1	0.056879	0.088479	0.050560	…	-0.006368	-0.046941	0.015923	
2	0.027369	0.038317	0.054738	…	0.043934	0.029153	0.008909	
3	0.031777	0.022797	0.053192	…	0.022062	0.027379	0.017751	
4	0.058667	0.078223	0.084742	…	0.009360	0.002659	-0.016512	
..	…	…	…	…	…	…	…	
492	0.025836	0.030534	0.050498	…	0.016355	0.012046	0.014732	
494	0.027426	0.054852	0.072484	…	0.032542	0.002385	0.034876	
496	0.059381	0.045973	0.063212	…	0.006448	-0.008978	-0.005994	
497	0.047555	0.039101	0.044913	…	0.016559	0.014200	0.004912	
499	0.036060	0.098689	0.045549	…	0.011292	0.017977	0.021060	

	X35	X36	X37	X38	X39	X40	X41
0	0.021723	0.009806	0.023432	0.021341	0.023350	0.018648	no_efectores

```
1    0.022599 -0.019049 -0.001007  0.034691  0.026019  0.013226  no_efectores
2    0.049489  0.013194 -0.006174 -0.031467  0.034315  0.048842  no_efectores
3    0.028778  0.014893  0.010081  0.011782  0.025925  0.020586  no_efectores
4    0.016366 -0.002329  0.020651 -0.014711  0.007880  0.016996  no_efectores
..        …        …        …        …         …         …
492  0.025066  0.027367  0.011436  0.014726  0.018776  0.018923  no_efectores
494  0.036823  0.015176  0.020986 -0.003106  0.005242  0.009765  no_efectores
496  0.008998  0.014549  0.021219  0.014871 -0.008108  0.008320  no_efectores
497  0.018309  0.005712  0.021756  0.036676  0.012469  0.012060  no_efectores
499  0.013486  0.008780  0.011943  0.012896  0.005154  0.014969  no_efectores

[402 rows x 42 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  402.000000  402.000000  402.000000  402.000000  402.000000  402.000000
mean     0.034825    0.016110    0.034622    0.050644    0.037910    0.036357
std      0.013612    0.012092    0.016143    0.027110    0.021078    0.012622
min      0.000000    0.000000    0.002037    0.000000    0.000000    0.000000
25%      0.025209    0.007477    0.022371    0.030410    0.023360    0.027706
50%      0.035355    0.013130    0.033584    0.048116    0.033775    0.036916
75%      0.043588    0.022066    0.045353    0.067795    0.050200    0.043747
max      0.077969    0.063693    0.085835    0.146086    0.121711    0.078549

                X6          X7          X8          X9  …         X31  \
count  402.000000  402.000000  402.000000  402.000000  …  402.000000
mean     0.014317    0.055475    0.054231    0.069877  …    0.013447
std      0.008297    0.026772    0.028685    0.028705  …    0.019538
min      0.000000    0.002813    0.000000    0.001876  …   -0.060527
25%      0.008696    0.036635    0.033788    0.050192  …    0.002822
50%      0.013396    0.051574    0.049174    0.069315  …    0.014114
75%      0.018917    0.069594    0.073096    0.087261  …    0.024969
max      0.044621    0.155082    0.160740    0.171725  …    0.077583

               X32         X33         X34         X35         X36         X37  \
count  402.000000  402.000000  402.000000  402.000000  402.000000  402.000000
mean     0.011071    0.012308    0.012415    0.012118    0.010356    0.011493
std      0.020250    0.020085    0.019365    0.019682    0.019875    0.020158
min     -0.059076   -0.074732   -0.060086   -0.062539   -0.086715   -0.067501
25%     -0.000319    0.001694    0.001910    0.001092   -0.000504   -0.000899
50%      0.013273    0.014101    0.011876    0.013277    0.011727    0.011978
75%      0.022609    0.023929    0.024252    0.023638    0.023075    0.024173
max      0.085545    0.095995    0.076570    0.098658    0.088191    0.083369
```

```
              X38         X39         X40
count  402.000000  402.000000  402.000000
mean     0.012061    0.011359    0.011454
std      0.020467    0.019360    0.020019
min     -0.063508   -0.062562   -0.059967
25%      0.001627    0.000380    0.000755
50%      0.012893    0.012039    0.011739
75%      0.025083    0.024332    0.024315
max      0.087639    0.070350    0.081638

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

          X0        X1        X2        X3        X4        X5        X6  \
0   0.018277  0.002150  0.016127  0.027954  0.007526  0.008601  0.004301
1   0.008166  0.012248  0.021775  0.036745  0.020414  0.031301  0.009527
2   0.041872  0.057574  0.099445  0.031404  0.068042  0.125615  0.036638
3   0.045142  0.028214  0.056428  0.039499  0.028214  0.067713  0.016928
4   0.011651  0.011651  0.006991  0.011651  0.037284  0.009321  0.004661
..       …         …         …         …         …         …         …
95  0.029139  0.014569  0.069205  0.021854  0.021854  0.010927  0.003642
96  0.041387  0.031040  0.027591  0.031040  0.017244  0.062080  0.024142
97  0.057701  0.014425  0.043276  0.014425  0.028850  0.067318  0.004808
98  0.064255  0.000000  0.020080  0.012048  0.020080  0.060239  0.004016
99  0.063137  0.024283  0.014570  0.063137  0.038853  0.072850  0.009713

          X7        X8        X9  …       X53       X54       X55       X56  \
0   0.012902  0.045156  0.020428  …  0.038048  0.004979  0.025071 -0.009413
1   0.027219  0.027219  0.024497  …  0.018668  0.014771  0.011573  0.000397
2   0.068042  0.041872  0.047106  … -0.043038 -0.019285 -0.045342 -0.006446
3   0.062070  0.112855  0.056428  …  0.021276  0.001224  0.015255  0.050500
4   0.046606  0.020972  0.048936  … -0.015240  0.034482  0.017423 -0.017249
..       …         …         …  …         …         …         …         …
95  0.058278  0.069205  0.087416  … -0.015870  0.036447  0.009691  0.033816
96  0.044836  0.068978  0.068978  … -0.017791 -0.008684  0.016652  0.005073
97  0.043276  0.052892  0.048084  …  0.024147 -0.020967 -0.019653 -0.017877
98  0.040159  0.012048  0.040159  … -0.023844 -0.015367  0.000702  0.019234
99  0.072850  0.082563  0.063137  …  0.013279 -0.002846  0.002789  0.034997

          X57       X58       X59       X60       X61        X62
0   0.025744  0.006863  0.029744  0.013973  0.031792  efectores
1   0.015191 -0.015618  0.007395  0.007897  0.005754  efectores
2  -0.025726 -0.022121 -0.014818 -0.021167  0.008946  efectores
3   0.084067  0.009914  0.035966  0.048298  0.057136  efectores
4  -0.020685  0.025240  0.007409  0.010190  0.017668  efectores
..       …         …         …         …         …         …
95  0.014090 -0.025375 -0.019853 -0.029026  0.002324  efectores
96 -0.003802 -0.023932 -0.015085 -0.009725 -0.004879  efectores
97  0.027297 -0.045948 -0.023252 -0.047877 -0.029207  efectores
```

```
98  0.031143  0.022588  0.016427 -0.009138 -0.021995  efectores
99  0.042919  0.028950 -0.015335 -0.066631 -0.028355  efectores

[100 rows x 63 columns]
```

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

```
                X0          X1          X2          X3          X4          X5  \
count  100.000000  100.000000  100.000000  100.000000  100.000000  100.000000
mean     0.032215    0.018662    0.028036    0.030296    0.031139    0.055443
std      0.032432    0.020514    0.022477    0.029173    0.030050    0.055895
min     -0.180299   -0.045075   -0.045075   -0.135224   -0.135224   -0.180299
25%      0.012416    0.001831    0.012267    0.015345    0.017199    0.017922
50%      0.030105    0.015176    0.024079    0.027946    0.029545    0.049950
75%      0.049674    0.028122    0.043370    0.041656    0.041696    0.073536
max      0.101589    0.082428    0.099445    0.141710    0.149583    0.244697

                X6          X7          X8          X9  …         X52  \
count  100.000000  100.000000  100.000000  100.000000  …  100.000000
mean     0.010000    0.038357    0.047281    0.049886  …   -0.000890
std      0.012378    0.047211    0.039278    0.048364  …    0.077740
min     -0.045075   -0.360598   -0.225374   -0.270448  …   -0.094341
25%      0.002294    0.024241    0.027146    0.028351  …   -0.028634
50%      0.007770    0.042829    0.050154    0.048567  …   -0.004073
75%      0.014885    0.060331    0.069227    0.064879  …    0.014273
max      0.048061    0.101589    0.149583    0.174153  …    0.715199

               X53         X54         X55         X56         X57         X58  \
count  100.000000  100.000000  100.000000  100.000000  100.000000  100.000000
mean     0.004279    0.004419    0.010401    0.005041    0.010896    0.005783
std      0.066951    0.089677    0.071084    0.061679    0.064882    0.090757
min     -0.077318   -0.085583   -0.076342   -0.117104   -0.124497   -0.094272
25%     -0.020315   -0.018299   -0.014094   -0.013850   -0.010412   -0.018630
50%     -0.001916   -0.002956    0.008739    0.004341    0.011888    0.000402
75%      0.022480    0.014870    0.025233    0.022205    0.033513    0.011759
max      0.614241    0.855159    0.661286    0.491680    0.519834    0.844470

               X59         X60         X61
count  100.000000  100.000000  100.000000
mean     0.005805    0.001305    0.008003
std      0.089897    0.066096    0.054342
min     -0.140903   -0.092203   -0.086615
25%     -0.015364   -0.030410   -0.014781
50%      0.002289    0.003869    0.005992
75%      0.020136    0.020018    0.024560
```

```
max      0.844235    0.553482    0.442533

[8 rows x 62 columns]
```

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.051485	0.023168	0.051485	0.066930	0.028317	0.033465	0.020594	
1	0.033662	0.016831	0.016831	0.061713	0.028051	0.039272	0.022441	
2	0.070985	0.070985	0.070985	0.000000	0.113576	0.085182	0.014197	
3	0.040729	0.013576	0.021722	0.047065	0.041634	0.034393	0.013576	
4	0.031955	0.063911	0.039944	0.143799	0.063911	0.063911	0.031955	
..	…	…	…	…	…	…	…	
495	0.031926	0.007512	0.013146	0.020658	0.028170	0.018780	0.011268	
496	0.021354	0.002373	0.026100	0.046268	0.007118	0.022541	0.010677	
497	0.032088	0.040897	0.044043	0.055998	0.039010	0.057885	0.016359	
498	0.013881	0.006941	0.034703	0.062465	0.062465	0.020822	0.020822	
499	0.029392	0.003041	0.023311	0.032432	0.017230	0.016216	0.004054	

	X7	X8	X9	…	X53	X54	X55	\
0	0.079801	0.084950	0.128712	…	-0.048189	-0.006500	-0.012422	
1	0.050493	0.078544	0.044882	…	0.000500	-0.029194	-0.012513	
2	0.070985	0.099379	0.141970	…	-0.124881	0.062352	0.067650	
3	0.041634	0.029868	0.069692	…	-0.009274	-0.010327	-0.010012	
4	0.071900	0.095866	0.103855	…	-0.025192	0.017769	0.047870	
..	…	…	…	…	…	…	…	
495	0.013146	0.022536	0.039438	…	0.012852	0.000255	0.011296	
496	0.036777	0.028473	0.039150	…	0.024733	-0.004952	0.020188	
497	0.056627	0.046560	0.053481	…	-0.000557	0.005748	-0.000452	
498	0.048584	0.055525	0.076346	…	-0.073613	-0.032176	-0.001108	
499	0.019257	0.052702	0.024324	…	0.009586	-0.002238	0.019271	

	X56	X57	X58	X59	X60	X61	X62
0	0.068816	0.040016	-0.008085	0.038232	-0.051698	-0.005664	no_efectores
1	0.048154	-0.018722	0.022243	0.064726	0.008273	0.035618	no_efectores
2	0.061600	0.008776	0.006592	-0.008908	-0.036864	0.037273	no_efectores
3	-0.013077	-0.007994	0.005501	0.006554	0.016648	0.005918	no_efectores
4	0.012125	-0.025344	0.021842	0.088815	-0.011940	-0.002441	no_efectores
..	…	…	…	…	…	…	…
495	0.014542	0.009288	0.015997	0.015419	0.006757	0.012635	no_efectores
496	0.020483	0.016439	-0.004266	0.014547	0.000690	0.005644	no_efectores
497	-0.011013	-0.005374	0.008868	0.009760	0.010399	0.015071	no_efectores
498	0.053303	-0.022884	-0.005257	-0.012702	0.033253	0.020123	no_efectores

```
499 -0.019173  0.014669 -0.011952  0.011015 -0.011081  0.014802  no_efectores

[500 rows x 63 columns]
```

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

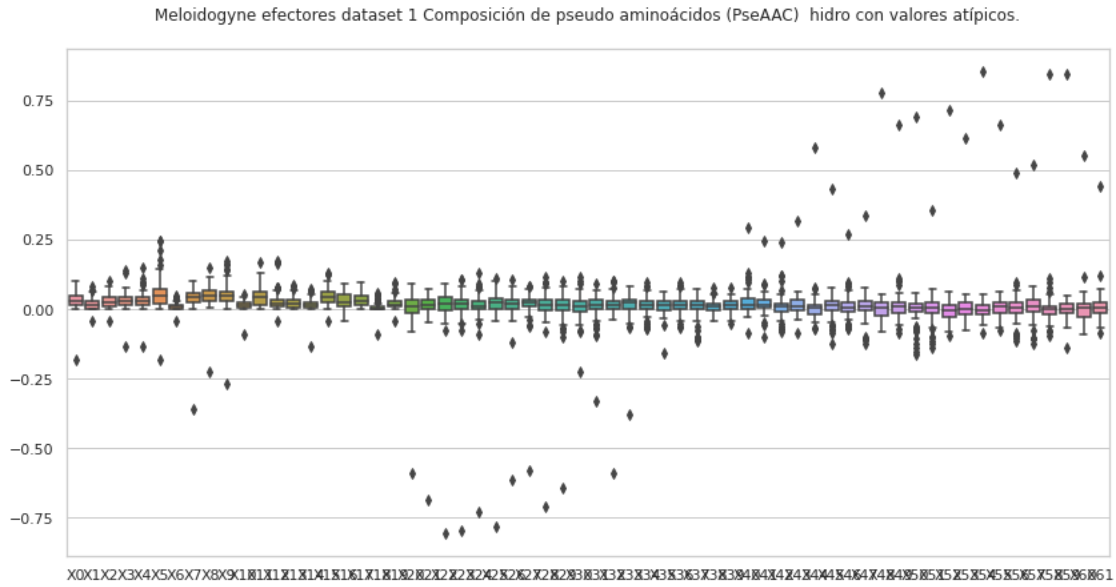
	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.034949	0.016828	0.034066	0.045834	0.039127	0.039534	
std	0.025242	0.016563	0.024817	0.027306	0.030549	0.032230	
min	-0.000000	-0.000000	-0.083784	-0.083784	-0.083784	-0.167568	
25%	0.015763	0.005497	0.016753	0.026675	0.020853	0.018485	
50%	0.029691	0.011626	0.029035	0.044144	0.033355	0.031492	
75%	0.046571	0.024144	0.047068	0.060466	0.050514	0.054563	
max	0.146669	0.119078	0.241320	0.175132	0.329073	0.227432	

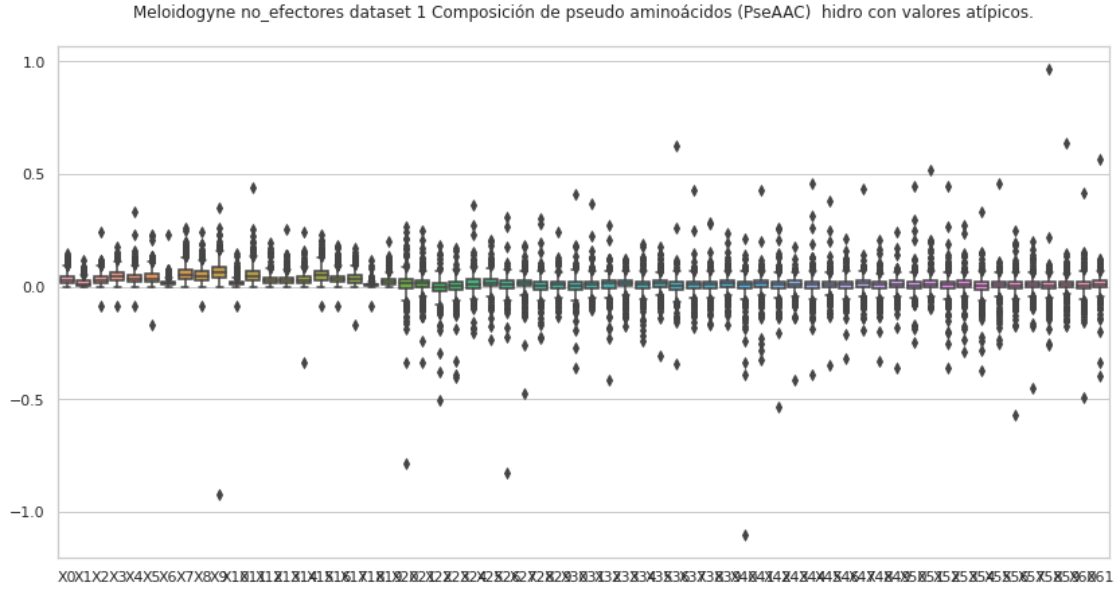
	X6	X7	X8	X9	…	X52	\
count	500.000000	500.000000	500.000000	500.000000	…	500.000000	
mean	0.015045	0.056251	0.051358	0.066374	…	0.003817	
std	0.015472	0.037050	0.032794	0.060888	…	0.047622	
min	-0.000000	-0.000000	-0.083784	-0.921626	…	-0.364160	
25%	0.006384	0.032039	0.029809	0.040485	…	-0.011133	
50%	0.012094	0.048495	0.044899	0.061837	…	0.006281	
75%	0.019548	0.071832	0.065855	0.087862	…	0.019752	
max	0.231016	0.259893	0.241320	0.351012	…	0.447390	

	X53	X54	X55	X56	X57	X58	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.008384	0.001107	0.006593	0.002089	0.006734	0.004038	
std	0.040347	0.045566	0.041983	0.046219	0.038732	0.057415	
min	-0.286867	-0.374990	-0.251893	-0.572112	-0.453824	-0.262256	
25%	-0.003290	-0.014274	-0.004139	-0.010788	-0.004658	-0.009204	
50%	0.011469	0.005763	0.009738	0.004983	0.008829	0.005223	
75%	0.024396	0.020082	0.023393	0.019235	0.022050	0.019587	
max	0.273513	0.332627	0.458681	0.246041	0.197321	0.964720	

	X59	X60	X61
count	500.000000	500.000000	500.000000
mean	0.008587	0.004128	0.009323
std	0.044763	0.046173	0.047934
min	-0.175628	-0.491607	-0.396963
25%	-0.001734	-0.010363	-0.004858
50%	0.009705	0.005927	0.010586
75%	0.021833	0.023205	0.024747
max	0.638838	0.418182	0.563564

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

	X0	X1	X2	X3	X4	X5	X6	\
0	0.018277	0.002150	0.016127	0.027954	0.007526	0.008601	0.004301	
1	0.008166	0.012248	0.021775	0.036745	0.020414	0.031301	0.009527	
3	0.045142	0.028214	0.056428	0.039499	0.028214	0.067713	0.016928	
4	0.011651	0.011651	0.006991	0.011651	0.037284	0.009321	0.004661	
5	0.040904	0.000000	0.027269	0.020452	0.036359	0.018179	0.018179	
..	…	…	…	…	…	…	…	
93	0.015147	0.005049	0.015147	0.010098	0.030294	0.055539	0.030294	
95	0.029139	0.014569	0.069205	0.021854	0.021854	0.010927	0.003642	
97	0.057701	0.014425	0.043276	0.014425	0.028850	0.067318	0.004808	
98	0.064255	0.000000	0.020080	0.012048	0.020080	0.060239	0.004016	
99	0.063137	0.024283	0.014570	0.063137	0.038853	0.072850	0.009713	

	X7	X8	X9	…	X53	X54	X55	X56	\
0	0.012902	0.045156	0.020428	…	0.038048	0.004979	0.025071	-0.009413	
1	0.027219	0.027219	0.024497	…	0.018668	0.014771	0.011573	0.000397	
3	0.062070	0.112855	0.056428	…	0.021276	0.001224	0.015255	0.050500	
4	0.046606	0.020972	0.048936	…	-0.015240	0.034482	0.017423	-0.017249	
5	0.047721	0.068173	0.056811	…	-0.002151	0.015165	0.005178	0.028081	
..	…	…	…	…	…	…	…	…	
93	0.030294	0.015147	0.045441	…	-0.013228	0.015413	0.021468	0.015679	
95	0.058278	0.069205	0.087416	…	-0.015870	0.036447	0.009691	0.033816	
97	0.043276	0.052892	0.048084	…	0.024147	-0.020967	-0.019653	-0.017877	
98	0.040159	0.012048	0.040159	…	-0.023844	-0.015367	0.000702	0.019234	
99	0.072850	0.082563	0.063137	…	0.013279	-0.002846	0.002789	0.034997	

	X57	X58	X59	X60	X61	X62
0	0.025744	0.006863	0.029744	0.013973	0.031792	efectores
1	0.015191	-0.015618	0.007395	0.007897	0.005754	efectores
3	0.084067	0.009914	0.035966	0.048298	0.057136	efectores
4	-0.020685	0.025240	0.007409	0.010190	0.017668	efectores
5	0.005864	0.022716	0.000189	-0.009404	-0.020793	efectores
..	…	…	…	…	…	…
93	0.008355	-0.002688	0.009749	0.023439	0.015619	efectores
95	0.014090	-0.025375	-0.019853	-0.029026	0.002324	efectores
97	0.027297	-0.045948	-0.023252	-0.047877	-0.029207	efectores
98	0.031143	0.022588	0.016427	-0.009138	-0.021995	efectores
99	0.042919	0.028950	-0.015335	-0.066631	-0.028355	efectores

```
[82 rows x 63 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	82.000000	82.000000	82.000000	82.000000	82.000000	82.000000	
mean	0.031563	0.016199	0.026670	0.029872	0.029013	0.047343	
std	0.021666	0.017049	0.019321	0.018328	0.020811	0.038565	
min	0.000000	0.000000	0.000000	0.000000	0.000308	0.001846	
25%	0.011949	0.000315	0.011803	0.015581	0.016337	0.016196	
50%	0.028615	0.014499	0.021784	0.027917	0.028009	0.042543	
75%	0.047056	0.023904	0.042988	0.039622	0.037998	0.064976	
max	0.083817	0.068314	0.080793	0.075241	0.098048	0.210455	

	X6	X7	X8	X9	…	X52	X53	\
count	82.000000	82.000000	82.000000	82.000000	…	82.000000	82.000000	
mean	0.009176	0.039301	0.048564	0.045813	…	-0.003904	0.003048	
std	0.008982	0.022097	0.026685	0.028818	…	0.025334	0.023787	
min	0.000000	0.001013	0.004974	0.001846	…	-0.062589	-0.046626	
25%	0.003123	0.024880	0.027001	0.023374	…	-0.024787	-0.014982	
50%	0.007588	0.039937	0.049694	0.046707	…	-0.001239	0.003023	
75%	0.013691	0.057997	0.068324	0.059471	…	0.014916	0.024100	
max	0.040128	0.083817	0.117657	0.143686	…	0.063435	0.054273	

	X54	X55	X56	X57	X58	X59	\
count	82.000000	82.000000	82.000000	82.000000	82.000000	82.000000	
mean	-0.001148	0.008466	0.005908	0.011306	-0.001113	0.003800	
std	0.024815	0.023290	0.030481	0.034520	0.026137	0.024695	
min	-0.085583	-0.061558	-0.093812	-0.105953	-0.076047	-0.068114	
25%	-0.012038	-0.005868	-0.009382	-0.006023	-0.014301	-0.009596	
50%	0.001890	0.009910	0.006276	0.014669	0.002766	0.007335	
75%	0.015351	0.027995	0.022962	0.034124	0.011523	0.022231	
max	0.052728	0.060562	0.086135	0.084067	0.078807	0.046444	

	X60	X61
count	82.000000	82.000000
mean	-0.003522	0.004096
std	0.032666	0.030344
min	-0.092203	-0.086615
25%	-0.027513	-0.015052
50%	0.006158	0.006056
75%	0.019199	0.025103
max	0.063835	0.072234

[8 rows x 62 columns]

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

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.051485  0.023168  0.051485  0.066930  0.028317  0.033465  0.020594
1    0.033662  0.016831  0.016831  0.061713  0.028051  0.039272  0.022441
3    0.040729  0.013576  0.021722  0.047065  0.041634  0.034393  0.013576
5    0.033178  0.015799  0.042658  0.044237  0.034758  0.041078  0.009479
6    0.075941  0.022047  0.047769  0.066142  0.044095  0.045319  0.017148
..        …         …         …         …         …         …         …
494  0.053573  0.016918  0.047933  0.087408  0.050753  0.059212  0.025377
495  0.031926  0.007512  0.013146  0.020658  0.028170  0.018780  0.011268
496  0.021354  0.002373  0.026100  0.046268  0.007118  0.022541  0.010677
497  0.032088  0.040897  0.044043  0.055998  0.039010  0.057885  0.016359
499  0.029392  0.003041  0.023311  0.032432  0.017230  0.016216  0.004054

           X7        X8        X9  …       X53       X54       X55  \
0    0.079801  0.084950  0.128712  … -0.048189 -0.006500 -0.012422
1    0.050493  0.078544  0.044882  …  0.000500 -0.029194 -0.012513
3    0.041634  0.029868  0.069692  … -0.009274 -0.010327 -0.010012
5    0.053717  0.048977  0.074256  …  0.003907  0.006852  0.021792
6    0.071041  0.061242  0.133509  … -0.014680 -0.002393  0.015491
..        …         …         …  …       …         …         …
494  0.039475  0.078949  0.104326  …  0.048054 -0.007767 -0.017099
495  0.013146  0.022536  0.039438  …  0.012852  0.000255  0.011296
496  0.036777  0.028473  0.039150  …  0.024733 -0.004952  0.020188
497  0.056627  0.046560  0.053481  … -0.000557  0.005748 -0.000452
499  0.019257  0.052702  0.024324  …  0.009586 -0.002238  0.019271

          X56       X57       X58       X59       X60       X61           X62
0    0.068816  0.040016 -0.008085  0.038232 -0.051698 -0.005664  no_efectores
1    0.048154 -0.018722  0.022243  0.064726  0.008273  0.035618  no_efectores
3   -0.013077 -0.007994  0.005501  0.006554  0.016648  0.005918  no_efectores
5    0.013780  0.024834 -0.003858  0.015828  0.015844  0.021924  no_efectores
6    0.039760  0.055427  0.009991  0.004287 -0.014923 -0.003600  no_efectores
..        …         …         …         …         …         …           …
494 -0.005978  0.005620 -0.024219 -0.011462 -0.024834 -0.041322  no_efectores
495  0.014542  0.009288  0.015997  0.015419  0.006757  0.012635  no_efectores
496  0.020483  0.016439 -0.004266  0.014547  0.000690  0.005644  no_efectores
497 -0.011013 -0.005374  0.008868  0.009760  0.010399  0.015071  no_efectores
499 -0.019173  0.014669 -0.011952  0.011015 -0.011081  0.014802  no_efectores

[424 rows x 63 columns]
```

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

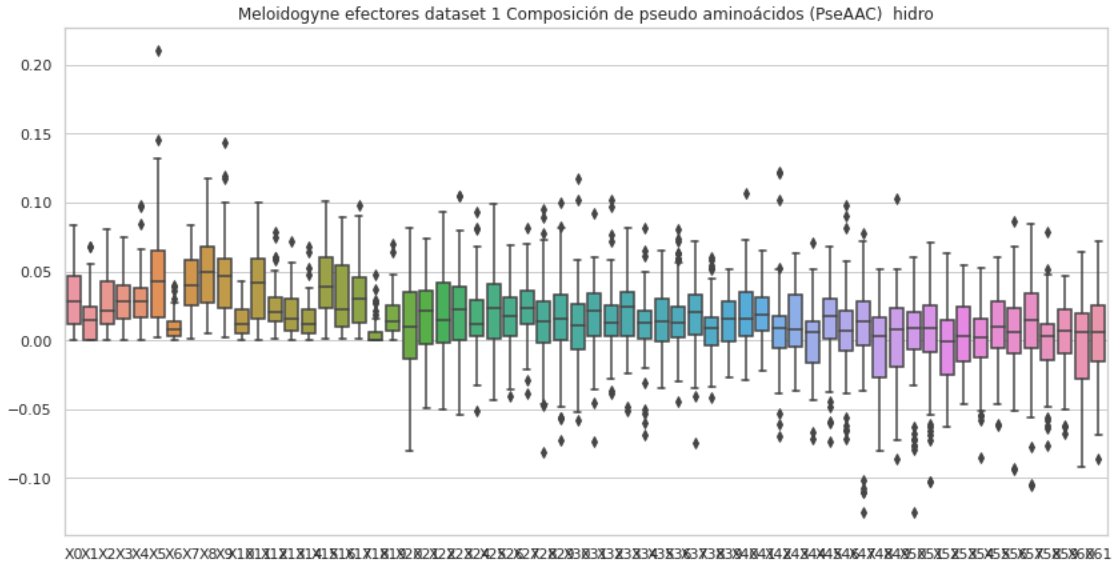
	X0	X1	X2	X3	X4	X5	\
count	424.000000	424.000000	424.000000	424.000000	424.000000	424.000000	
mean	0.031347	0.014422	0.030345	0.042021	0.032787	0.034056	
std	0.020630	0.012614	0.018645	0.021553	0.019080	0.023081	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.015003	0.005283	0.015999	0.025849	0.018938	0.017476	
50%	0.028281	0.010544	0.027231	0.041349	0.030016	0.028327	
75%	0.042786	0.020314	0.042709	0.054528	0.044709	0.045433	
max	0.102869	0.065062	0.098353	0.122387	0.118288	0.133655	

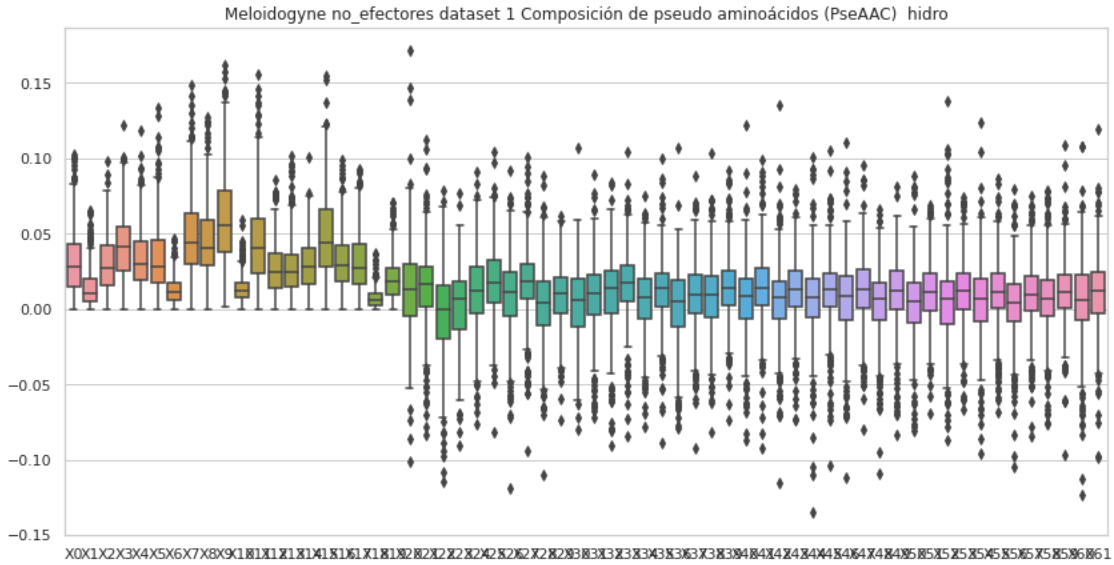
	X6	X7	X8	X9	…	X52	\
count	424.000000	424.000000	424.000000	424.000000	…	424.000000	
mean	0.012653	0.047978	0.044506	0.059976	…	0.005573	
std	0.008750	0.026432	0.023103	0.031424	…	0.026467	
min	0.000000	0.000000	0.000000	0.001575	…	-0.086925	
25%	0.006212	0.030120	0.028647	0.037889	…	-0.009774	
50%	0.011121	0.043975	0.040747	0.055988	…	0.006717	
75%	0.017480	0.063158	0.059188	0.078941	…	0.018493	
max	0.046829	0.149019	0.126950	0.162304	…	0.138049	

	X53	X54	X55	X56	X57	X58	\
count	424.000000	424.000000	424.000000	424.000000	424.000000	424.000000	
mean	0.010591	0.004723	0.010807	0.002802	0.009355	0.005956	
std	0.021607	0.025003	0.022713	0.024008	0.021557	0.023439	
min	-0.066986	-0.095973	-0.068708	-0.104753	-0.084618	-0.076284	
25%	-0.000581	-0.008228	0.000418	-0.008223	-0.001544	-0.005103	
50%	0.011746	0.006842	0.011242	0.004473	0.009906	0.006553	
75%	0.023853	0.019723	0.024125	0.016255	0.022050	0.019023	
max	0.074597	0.124176	0.087001	0.079438	0.075444	0.078468	

	X59	X60	X61
count	424.000000	424.000000	424.000000
mean	0.011739	0.005242	0.010533
std	0.021753	0.027003	0.024318
min	-0.096902	-0.123140	-0.098148
25%	0.000708	-0.007255	-0.002529
50%	0.011187	0.006274	0.011890
75%	0.023236	0.022579	0.024241
max	0.109043	0.108119	0.119302

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

          X0        X1        X2        X3        X4        X5        X6  \
0  -0.016711 -0.038819  0.034747  0.004099  0.000577 -0.042956 -0.075184
1  -0.024730 -0.033735  0.043962  0.059285 -0.029851  0.003876  0.021239
2   0.099254  0.018412  0.113440  0.048564 -0.052121 -0.055317 -0.031070
3   0.007302 -0.010581  0.011657 -0.036138 -0.054003  0.018823 -0.040838
4  -0.043124  0.000099 -0.102402  0.083976 -0.010418 -0.070118 -0.069451
..       …         …         …         …         …         …         …
95  0.012909 -0.018473 -0.079513 -0.022248 -0.016018  0.034589 -0.007860
96  0.034983  0.024450  0.045605  0.002157 -0.034961  0.089501 -0.000895
97  0.074108  0.033914 -0.040629 -0.027014  0.007079 -0.038036 -0.007393
98  0.217638 -0.022910  0.168145  0.035309  0.095543  0.076718  0.141923
99  0.019018 -0.046037 -0.076364  0.038581  0.113805  0.027334  0.024965

          X7        X8        X9       X10       X11       X12        X13
0  -0.078480  0.051717 -0.040286 -0.046466  0.043594  0.034011  efectores
```

```
1  -0.062214  0.007660  0.041061  0.042803 -0.002874  0.033929  efectores
2  -0.006032  0.015766 -0.150838 -0.002576  0.014996 -0.081872  efectores
3  -0.015333  0.101709  0.045478 -0.031599  0.105639 -0.047101  efectores
4   0.013735 -0.056687 -0.074065  0.004819  0.016878  0.079704  efectores
..       …       …       …       …       …       …       …
95  0.000230  0.003870  0.031836 -0.047503 -0.048570 -0.012877  efectores
96  0.010933  0.094244  0.043965  0.071333  0.091996  0.004503  efectores
97 -0.013590  0.062275  0.010002  0.131511  0.008124  0.042546  efectores
98  0.188212  0.061621 -0.089192  0.031964  0.082573  0.037887  efectores
99  0.000714 -0.015229  0.071540  0.134466  0.014104 -0.019980  efectores

[100 rows x 14 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	
mean	-0.003043	0.038740	0.020301	0.013636	0.015350	0.037593	
std	0.137824	0.101201	0.108151	0.086700	0.102013	0.103611	
min	-0.568520	-0.341363	-0.170590	-0.267092	-0.177306	-0.136557	
25%	-0.038868	-0.013518	-0.024807	-0.025291	-0.049862	-0.018496	
50%	0.010243	0.018375	0.014999	0.013177	0.006489	0.018984	
75%	0.057455	0.056850	0.045584	0.057086	0.072971	0.071407	
max	0.366084	0.421241	0.595207	0.251137	0.398630	0.424122	

	X6	X7	X8	X9	X10	X11	\
count	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	
mean	0.029261	-0.002673	0.055870	0.011285	0.040582	0.018197	
std	0.078483	0.097604	0.107336	0.092542	0.088739	0.107474	
min	-0.169134	-0.314314	-0.198825	-0.239777	-0.236089	-0.266382	
25%	-0.021321	-0.033898	-0.007559	-0.046569	-0.018807	-0.030954	
50%	0.012931	0.001064	0.046848	0.018323	0.037763	0.021404	
75%	0.070870	0.030448	0.097163	0.067351	0.080371	0.065388	
max	0.277734	0.366508	0.378650	0.323957	0.400308	0.523388	

	X12
count	100.000000
mean	0.020297
std	0.089523
min	-0.210603
25%	-0.027984
50%	0.009178
75%	0.068892
max	0.342163

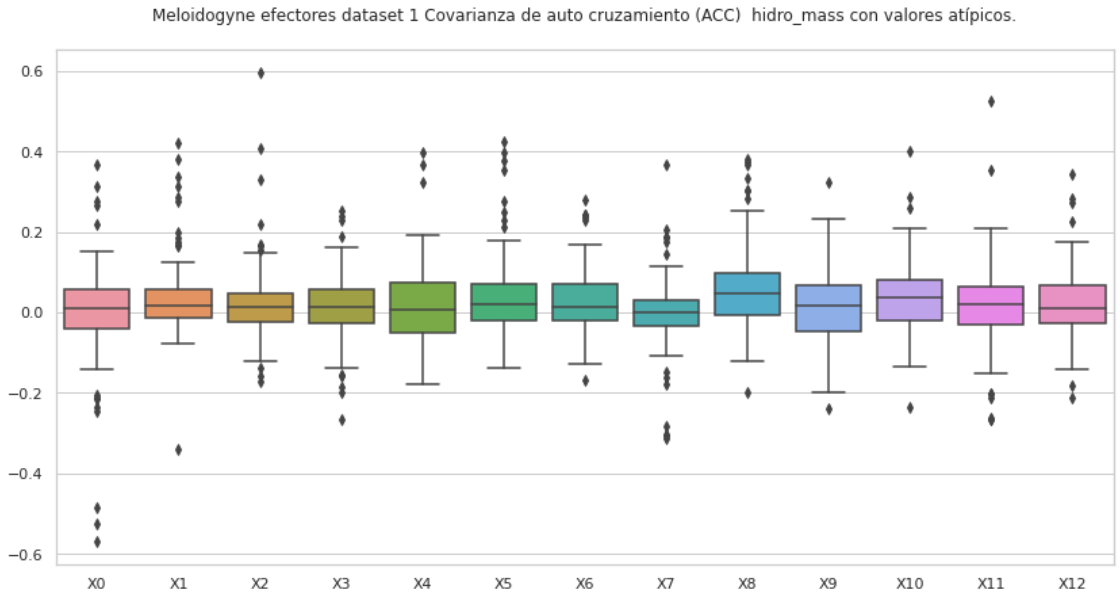
no_efectores
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores Meloidogyne
dataset 1, con valores atípicos.
Valores del documento csv.

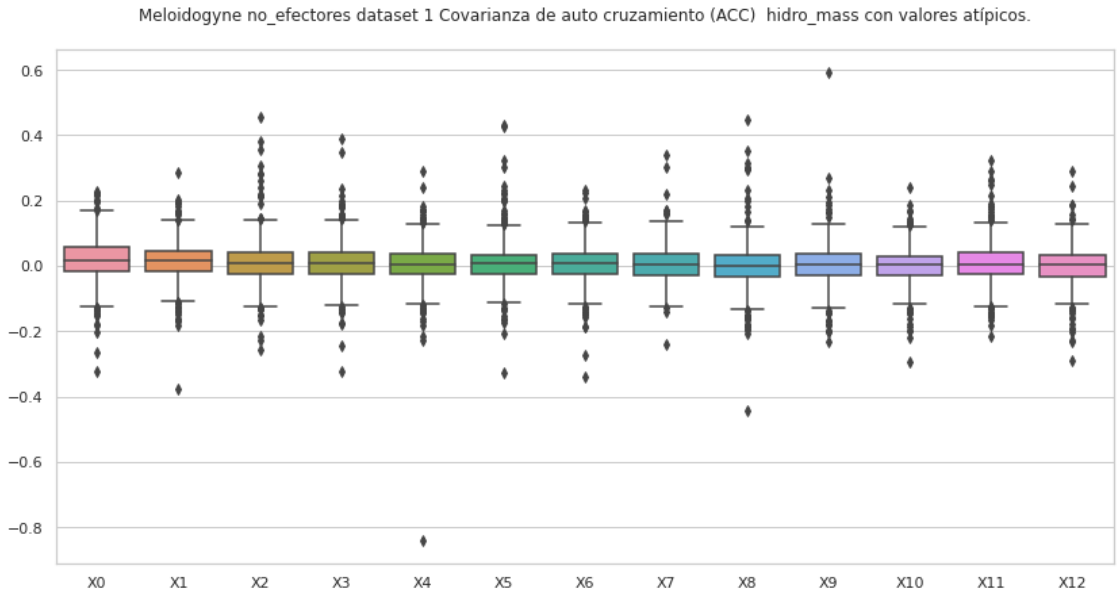
```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.059937 -0.003327 -0.007780  0.039320 -0.019375 -0.037437  0.022728
1    0.131313  0.021171  0.002833 -0.001828  0.087780 -0.050239 -0.004440
2    0.015140  0.031565 -0.072805  0.123689  0.114221 -0.038324  0.168273
3    0.039179  0.026931  0.034465  0.060948  0.046050  0.087206  0.052078
4   -0.006290  0.110345  0.103445 -0.010907  0.058920  0.033896 -0.100157
..        …         …         …         …         …         …
495 -0.135027  0.097400 -0.067243 -0.099122  0.044400 -0.055198 -0.052472
496  0.109624  0.078842  0.078484  0.027877  0.050332 -0.018166 -0.006477
497 -0.021563  0.013939  0.026374 -0.051255 -0.014923  0.022753  0.034066
498  0.033112  0.046003  0.085295  0.032019 -0.053626  0.088083  0.047758
499  0.025865  0.012384 -0.014746 -0.001325  0.030411  0.040879 -0.012147

           X7        X8        X9       X10       X11       X12           X13
0   -0.008882 -0.032152 -0.101270  0.011480  0.006429  0.012883  no_efectores
1   -0.007733 -0.044287  0.018601 -0.010111  0.009493 -0.032125  no_efectores
2    0.129891  0.230137 -0.232525 -0.029253  0.111018  0.115282  no_efectores
3    0.007955  0.051125  0.024718  0.021189  0.020654  0.031834  no_efectores
4   -0.099986 -0.083866 -0.089299  0.083480 -0.123655  0.015975  no_efectores
..        …         …         …         …         …         …
495  0.028444  0.004482  0.086660  0.035851 -0.120097 -0.010234  no_efectores
496 -0.006632 -0.040699 -0.049981 -0.007190 -0.004306 -0.008853  no_efectores
497 -0.019285  0.015876 -0.000186 -0.047112  0.030711  0.000095  no_efectores
498 -0.114249 -0.023544 -0.060544 -0.090824 -0.083271  0.062528  no_efectores
499  0.009902 -0.007093  0.016324 -0.007837 -0.008228  0.006441  no_efectores

[500 rows x 14 columns]
```

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

```
                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.019139    0.016765    0.011676    0.010995    0.004239    0.010394
std      0.068179    0.063456    0.069521    0.063857    0.071132    0.069782
min     -0.324290   -0.376171   -0.255026   -0.324957   -0.840469   -0.327124
25%     -0.015646   -0.015388   -0.024947   -0.023938   -0.024824   -0.024905
50%      0.017275    0.016788    0.007035    0.010324    0.004448    0.008016
75%      0.057932    0.046493    0.042060    0.041928    0.036786    0.035233
```

```
max      0.226635    0.284324    0.454903    0.391143    0.290047    0.429306

                X6          X7          X8          X9         X10         X11  \
count   500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean      0.008762    0.005665    0.000055    0.005638   -0.000988    0.008228
std       0.060785    0.057173    0.072982    0.066862    0.058218    0.063278
min      -0.340819   -0.238420   -0.442546   -0.232525   -0.293703   -0.213525
25%      -0.022744   -0.027583   -0.034581   -0.030875   -0.030501   -0.025925
50%       0.009217    0.005550   -0.000474    0.005962    0.002760    0.005641
75%       0.039301    0.038273    0.031705    0.036065    0.030103    0.040016
max       0.233096    0.338035    0.445625    0.591797    0.241506    0.324271

               X12
count   500.000000
mean     -0.000470
std       0.061271
min      -0.288880
25%      -0.031254
50%       0.004900
75%       0.032414
max       0.291534
```





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

          X0        X1        X2        X3        X4        X5        X6  \
0  -0.016711 -0.038819  0.034747  0.004099  0.000577 -0.042956 -0.075184
1  -0.024730 -0.033735  0.043962  0.059285 -0.029851  0.003876  0.021239
2   0.099254  0.018412  0.113440  0.048564 -0.052121 -0.055317 -0.031070
3   0.007302 -0.010581  0.011657 -0.036138 -0.054003  0.018823 -0.040838
4  -0.043124  0.000099 -0.102402  0.083976 -0.010418 -0.070118 -0.069451
..       …         …         …         …         …         …         …
95  0.012909 -0.018473 -0.079513 -0.022248 -0.016018  0.034589 -0.007860
96  0.034983  0.024450  0.045605  0.002157 -0.034961  0.089501 -0.000895
97  0.074108  0.033914 -0.040629 -0.027014  0.007079 -0.038036 -0.007393
98  0.217638 -0.022910  0.168145  0.035309  0.095543  0.076718  0.141923
99  0.019018 -0.046037 -0.076364  0.038581  0.113805  0.027334  0.024965

          X7        X8        X9       X10       X11       X12        X13
0  -0.078480  0.051717 -0.040286 -0.046466  0.043594  0.034011  efectores
1  -0.062214  0.007660  0.041061  0.042803 -0.002874  0.033929  efectores
2  -0.006032  0.015766 -0.150838 -0.002576  0.014996 -0.081872  efectores
3  -0.015333  0.101709  0.045478 -0.031599  0.105639 -0.047101  efectores
4   0.013735 -0.056687 -0.074065  0.004819  0.016878  0.079704  efectores
..       …         …         …         …         …         …         …
95  0.000230  0.003870  0.031836 -0.047503 -0.048570 -0.012877  efectores
```

```
96  0.010933  0.094244  0.043965  0.071333  0.091996  0.004503  efectores
97 -0.013590  0.062275  0.010002  0.131511  0.008124  0.042546  efectores
98  0.188212  0.061621 -0.089192  0.031964  0.082573  0.037887  efectores
99  0.000714 -0.015229  0.071540  0.134466  0.014104 -0.019980  efectores

[84 rows x 14 columns]
```

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

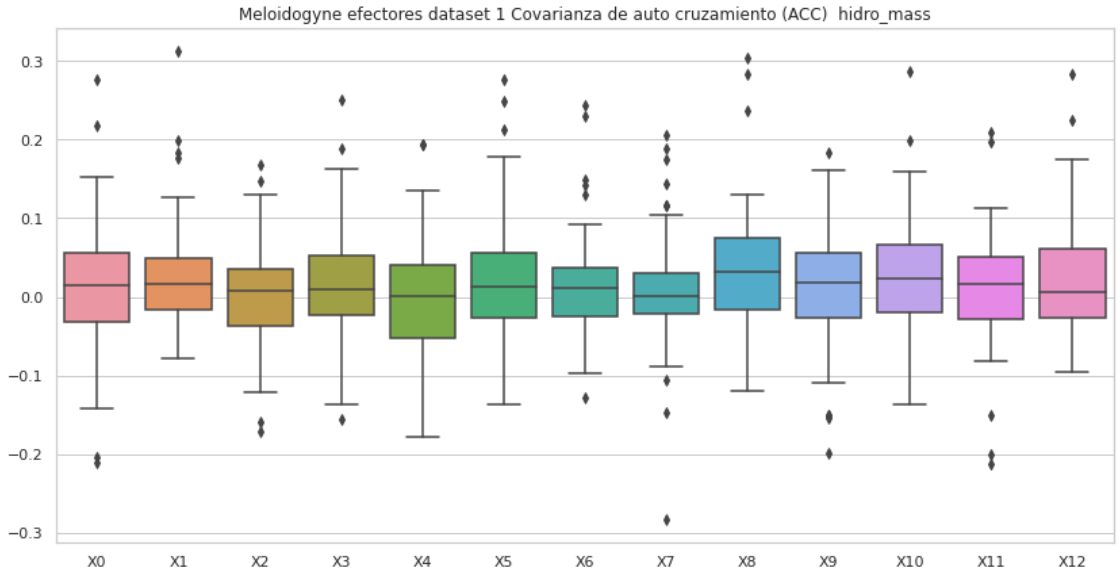
```
              X0         X1         X2         X3         X4         X5  \
count  84.000000  84.000000  84.000000  84.000000  84.000000  84.000000
mean    0.014326   0.022744  -0.002117   0.014598  -0.000289   0.018958
std     0.076734   0.062037   0.068489   0.066372   0.080139   0.073552
min    -0.211447  -0.077600  -0.170590  -0.155161  -0.177306  -0.136557
25%    -0.031026  -0.015875  -0.036457  -0.023776  -0.052919  -0.025997
50%     0.013860   0.016230   0.008591   0.009663   0.000958   0.013172
75%     0.055894   0.048745   0.034872   0.051962   0.040105   0.055826
max     0.275770   0.312105   0.168145   0.251137   0.193364   0.275986

              X6         X7         X8         X9        X10        X11  \
count  84.000000  84.000000  84.000000  84.000000  84.000000  84.000000
mean    0.015966   0.004571   0.032357   0.011648   0.030028   0.013707
std     0.061761   0.069297   0.076737   0.074454   0.070514   0.068456
min    -0.128069  -0.282952  -0.120373  -0.199405  -0.136223  -0.213021
25%    -0.024080  -0.022060  -0.015609  -0.026259  -0.019110  -0.027727
50%     0.010818   0.001064   0.032082   0.018323   0.024039   0.015937
75%     0.036948   0.030223   0.075510   0.056282   0.066129   0.050366
max     0.242701   0.206014   0.304270   0.182667   0.287288   0.209694

             X12
count  84.000000
mean    0.019290
std     0.069218
min    -0.095788
25%    -0.027115
50%     0.005510
75%     0.060535
max     0.283589
```

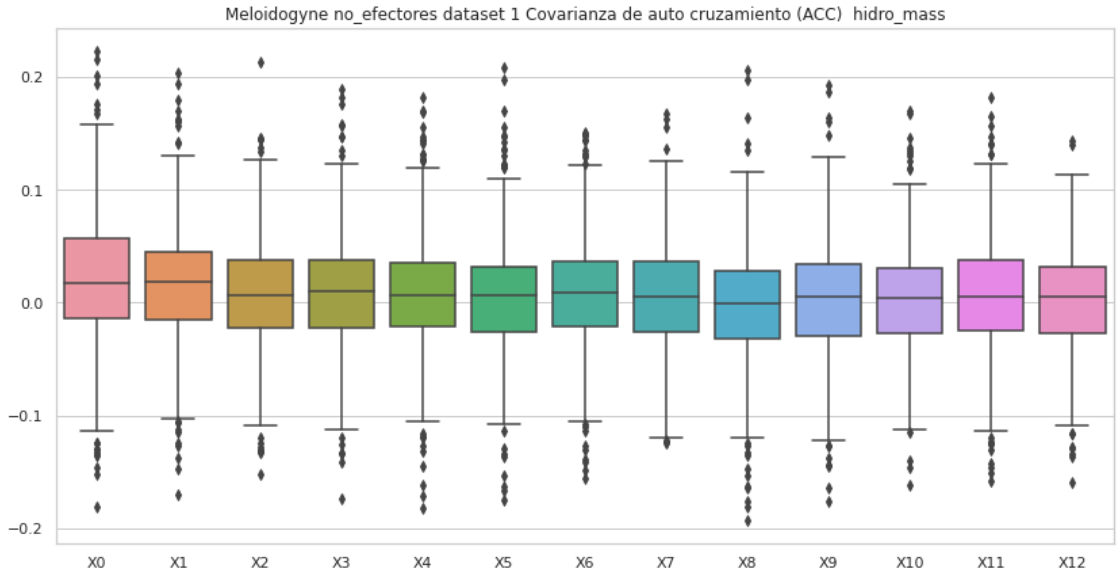
no_efectores
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores Meloidogyne dataset 1, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.059937 -0.003327 -0.007780  0.039320 -0.019375 -0.037437  0.022728
1    0.131313  0.021171  0.002833 -0.001828  0.087780 -0.050239 -0.004440
3    0.039179  0.026931  0.034465  0.060948  0.046050  0.087206  0.052078
4   -0.006290  0.110345  0.103445 -0.010907  0.058920  0.033896 -0.100157
5   -0.007292 -0.010907 -0.012342  0.013719 -0.010260  0.072002  0.056594
..        …         …         …         …         …         …         …
495 -0.135027  0.097400 -0.067243 -0.099122  0.044400 -0.055198 -0.052472
496  0.109624  0.078842  0.078484  0.027877  0.050332 -0.018166 -0.006477
497 -0.021563  0.013939  0.026374 -0.051255 -0.014923  0.022753  0.034066
498  0.033112  0.046003  0.085295  0.032019 -0.053626  0.088083  0.047758
499  0.025865  0.012384 -0.014746 -0.001325  0.030411  0.040879 -0.012147

           X7        X8        X9       X10       X11       X12           X13
0   -0.008882 -0.032152 -0.101270  0.011480  0.006429  0.012883  no_efectores
1   -0.007733 -0.044287  0.018601 -0.010111  0.009493 -0.032125  no_efectores
3    0.007955  0.051125  0.024718  0.021189  0.020654  0.031834  no_efectores
4   -0.099986 -0.083866 -0.089299  0.083480 -0.123655  0.015975  no_efectores
5   -0.078320  0.044407  0.072148 -0.003260  0.017709 -0.007980  no_efectores
..        …         …         …         …         …         …           …
495  0.028444  0.004482  0.086660  0.035851 -0.120097 -0.010234  no_efectores
496 -0.006632 -0.040699 -0.049981 -0.007190 -0.004306 -0.008853  no_efectores
497 -0.019285  0.015876 -0.000186 -0.047112  0.030711  0.000095  no_efectores
498 -0.114249 -0.023544 -0.060544 -0.090824 -0.083271  0.062528  no_efectores
499  0.009902 -0.007093  0.016324 -0.007837 -0.008228  0.006441  no_efectores

[456 rows x 14 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  456.000000  456.000000  456.000000  456.000000  456.000000  456.000000
mean     0.019814    0.018380    0.006982    0.008869    0.007478    0.004227
std      0.060791    0.054054    0.049099    0.050641    0.053235    0.053173
min     -0.181387   -0.170331   -0.152435   -0.174212   -0.182020   -0.174870
25%     -0.014229   -0.014663   -0.022920   -0.022911   -0.021228   -0.025399
50%      0.017430    0.018297    0.006714    0.010187    0.005982    0.006418
75%      0.057422    0.045132    0.037655    0.037792    0.035850    0.032301
max      0.222803    0.203474    0.212712    0.189071    0.181450    0.208386

                X6          X7          X8          X9         X10         X11  \
count  456.000000  456.000000  456.000000  456.000000  456.000000  456.000000
mean     0.008773    0.004404   -0.001186    0.003692    0.001724    0.005768
std      0.050780    0.048687    0.054435    0.053036    0.048976    0.050986
```

min	-0.156282	-0.124486	-0.193017	-0.176600	-0.161668	-0.158056
25%	-0.020935	-0.025858	-0.032437	-0.029603	-0.027084	-0.024131
50%	0.009336	0.005221	-0.000086	0.004800	0.003878	0.005592
75%	0.036527	0.036678	0.028622	0.034134	0.030015	0.037491
max	0.150674	0.168036	0.205748	0.193099	0.170511	0.182588

	X12
count	456.000000
mean	0.000779
std	0.049188
min	-0.159431
25%	-0.027252
50%	0.005030
75%	0.031511
max	0.143565





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

          X0        X1        X2        X3        X4        X5        X6  \
0  -0.016711 -0.038819  0.034747  0.004099  0.000577 -0.042956 -0.075184
1  -0.024730 -0.033735  0.043962  0.059285 -0.029851  0.003876  0.021239
2   0.099254  0.018412  0.113440  0.048564 -0.052121 -0.055317 -0.031070
3   0.007302 -0.010581  0.011657 -0.036138 -0.054003  0.018823 -0.040838
4  -0.043124  0.000099 -0.102402  0.083976 -0.010418 -0.070118 -0.069451
..       …         …         …         …         …         …         …
95  0.012909 -0.018473 -0.079513 -0.022248 -0.016018  0.034589 -0.007860
96  0.034983  0.024450  0.045605  0.002157 -0.034961  0.089501 -0.000895
97  0.074108  0.033914 -0.040629 -0.027014  0.007079 -0.038036 -0.007393
98  0.217638 -0.022910  0.168145  0.035309  0.095543  0.076718  0.141923
99  0.019018 -0.046037 -0.076364  0.038581  0.113805  0.027334  0.024965

          X7        X8        X9       X10       X11       X12        X13
0  -0.078480  0.051717 -0.040286 -0.046466  0.043594  0.034011  efectores
1  -0.062214  0.007660  0.041061  0.042803 -0.002874  0.033929  efectores
2  -0.006032  0.015766 -0.150838 -0.002576  0.014996 -0.081872  efectores
3  -0.015333  0.101709  0.045478 -0.031599  0.105639 -0.047101  efectores
4   0.013735 -0.056687 -0.074065  0.004819  0.016878  0.079704  efectores
..       …         …         …         …         …         …         …
95  0.000230  0.003870  0.031836 -0.047503 -0.048570 -0.012877  efectores
96  0.010933  0.094244  0.043965  0.071333  0.091996  0.004503  efectores
97 -0.013590  0.062275  0.010002  0.131511  0.008124  0.042546  efectores
98  0.188212  0.061621 -0.089192  0.031964  0.082573  0.037887  efectores
99  0.000714 -0.015229  0.071540  0.134466  0.014104 -0.019980  efectores

[100 rows x 14 columns]


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

               X0          X1          X2          X3          X4          X5  \
count  100.000000  100.000000  100.000000  100.000000  100.000000  100.000000
mean    -0.003043    0.038740    0.020301    0.013636    0.015350    0.037593
```

```
std      0.137824    0.101201    0.108151    0.086700    0.102013    0.103611
min     -0.568520   -0.341363   -0.170590   -0.267092   -0.177306   -0.136557
25%     -0.038868   -0.013518   -0.024807   -0.025291   -0.049862   -0.018496
50%      0.010243    0.018375    0.014999    0.013177    0.006489    0.018984
75%      0.057455    0.056850    0.045584    0.057086    0.072971    0.071407
max      0.366084    0.421241    0.595207    0.251137    0.398630    0.424122

                X6          X7          X8          X9         X10         X11  \
count   100.000000  100.000000  100.000000  100.000000  100.000000  100.000000
mean      0.029261   -0.002673    0.055870    0.011285    0.040582    0.018197
std       0.078483    0.097604    0.107336    0.092542    0.088739    0.107474
min      -0.169134   -0.314314   -0.198825   -0.239777   -0.236089   -0.266382
25%      -0.021321   -0.033898   -0.007559   -0.046569   -0.018807   -0.030954
50%       0.012931    0.001064    0.046848    0.018323    0.037763    0.021404
75%       0.070870    0.030448    0.097163    0.067351    0.080371    0.065388
max       0.277734    0.366508    0.378650    0.323957    0.400308    0.523388

               X12
count   100.000000
mean      0.020297
std       0.089523
min      -0.210603
25%      -0.027984
50%       0.009178
75%       0.068892
max       0.342163
```

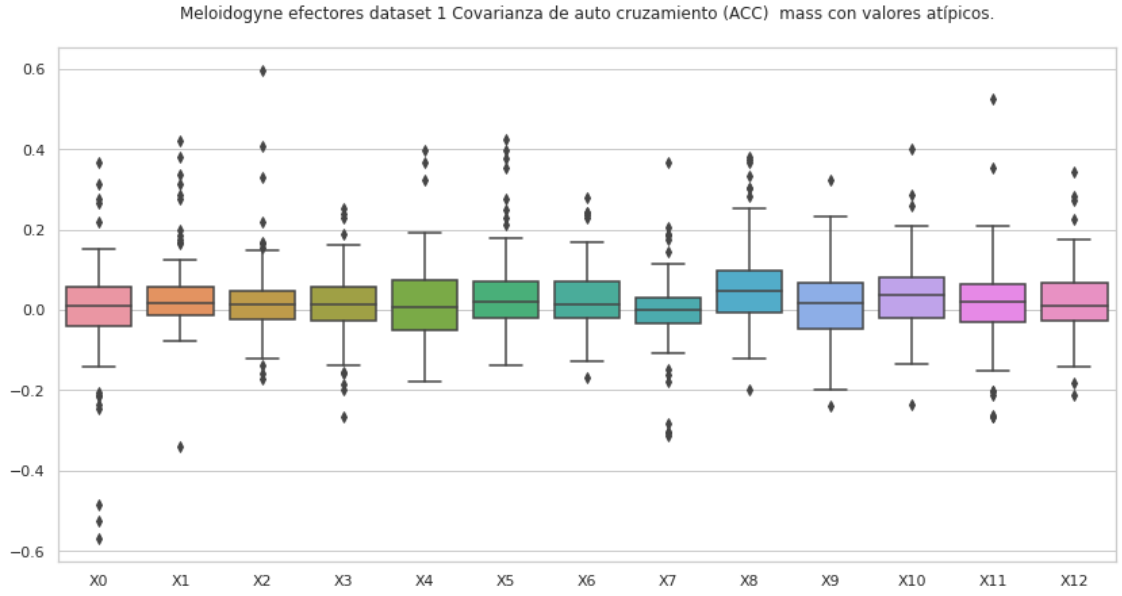
no_efectores
Covarianza de auto cruzamiento (ACC) mass no_efectores Meloidogyne dataset 1,
con valores atípicos.
Valores del documento csv.

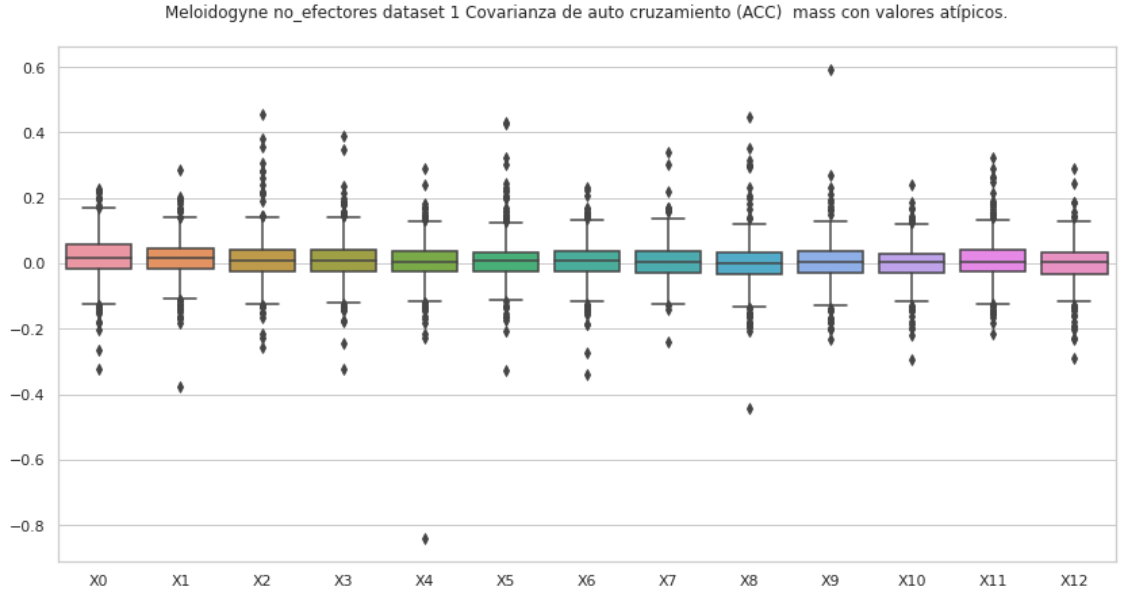
```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.059937 -0.003327 -0.007780  0.039320 -0.019375 -0.037437  0.022728
1    0.131313  0.021171  0.002833 -0.001828  0.087780 -0.050239 -0.004440
2    0.015140  0.031565 -0.072805  0.123689  0.114221 -0.038324  0.168273
3    0.039179  0.026931  0.034465  0.060948  0.046050  0.087206  0.052078
4   -0.006290  0.110345  0.103445 -0.010907  0.058920  0.033896 -0.100157
..        …         …         …         …         …         …         …
495 -0.135027  0.097400 -0.067243 -0.099122  0.044400 -0.055198 -0.052472
496  0.109624  0.078842  0.078484  0.027877  0.050332 -0.018166 -0.006477
497 -0.021563  0.013939  0.026374 -0.051255 -0.014923  0.022753  0.034066
498  0.033112  0.046003  0.085295  0.032019 -0.053626  0.088083  0.047758
499  0.025865  0.012384 -0.014746 -0.001325  0.030411  0.040879 -0.012147

           X7        X8        X9       X10       X11       X12           X13
```

```
0   -0.008882 -0.032152 -0.101270  0.011480  0.006429  0.012883  no_efectores
1   -0.007733 -0.044287  0.018601 -0.010111  0.009493 -0.032125  no_efectores
2    0.129891  0.230137 -0.232525 -0.029253  0.111018  0.115282  no_efectores
3    0.007955  0.051125  0.024718  0.021189  0.020654  0.031834  no_efectores
4   -0.099986 -0.083866 -0.089299  0.083480 -0.123655  0.015975  no_efectores
..        …         …         …         …         …         …           …
495  0.028444  0.004482  0.086660  0.035851 -0.120097 -0.010234  no_efectores
496 -0.006632 -0.040699 -0.049981 -0.007190 -0.004306 -0.008853  no_efectores
497 -0.019285  0.015876 -0.000186 -0.047112  0.030711  0.000095  no_efectores
498 -0.114249 -0.023544 -0.060544 -0.090824 -0.083271  0.062528  no_efectores
499  0.009902 -0.007093  0.016324 -0.007837 -0.008228  0.006441  no_efectores

[500 rows x 14 columns]
```

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

```
                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.019139    0.016765    0.011676    0.010995    0.004239    0.010394
std      0.068179    0.063456    0.069521    0.063857    0.071132    0.069782
min     -0.324290   -0.376171   -0.255026   -0.324957   -0.840469   -0.327124
25%     -0.015646   -0.015388   -0.024947   -0.023938   -0.024824   -0.024905
50%      0.017275    0.016788    0.007035    0.010324    0.004448    0.008016
75%      0.057932    0.046493    0.042060    0.041928    0.036786    0.035233
max      0.226635    0.284324    0.454903    0.391143    0.290047    0.429306

                X6          X7          X8          X9         X10         X11  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.008762    0.005665    0.000055    0.005638   -0.000988    0.008228
std      0.060785    0.057173    0.072982    0.066862    0.058218    0.063278
min     -0.340819   -0.238420   -0.442546   -0.232525   -0.293703   -0.213525
25%     -0.022744   -0.027583   -0.034581   -0.030875   -0.030501   -0.025925
50%      0.009217    0.005550   -0.000474    0.005962    0.002760    0.005641
75%      0.039301    0.038273    0.031705    0.036065    0.030103    0.040016
max      0.233096    0.338035    0.445625    0.591797    0.241506    0.324271

               X12
count  500.000000
mean    -0.000470
std      0.061271
min     -0.288880
25%     -0.031254
50%      0.004900
75%      0.032414
max      0.291534
```


Meloidogyne efectores dataset 1 Covarianza de auto cruzamiento (ACC) mass con valores atípicos.
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12


Meloidogyne no_efectores dataset 1 Covarianza de auto cruzamiento (ACC) mass con valores atípicos.
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
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 Meloidogyne dataset 1, sin
valores atípicos.
Valores del documento csv.
```

	X0	X1	X2	X3	X4	X5	X6	\
0	-0.016711	-0.038819	0.034747	0.004099	0.000577	-0.042956	-0.075184	
1	-0.024730	-0.033735	0.043962	0.059285	-0.029851	0.003876	0.021239	
2	0.099254	0.018412	0.113440	0.048564	-0.052121	-0.055317	-0.031070	
3	0.007302	-0.010581	0.011657	-0.036138	-0.054003	0.018823	-0.040838	
4	-0.043124	0.000099	-0.102402	0.083976	-0.010418	-0.070118	-0.069451	
..	…	…	…	…	…	…	…	
95	0.012909	-0.018473	-0.079513	-0.022248	-0.016018	0.034589	-0.007860	
96	0.034983	0.024450	0.045605	0.002157	-0.034961	0.089501	-0.000895	
97	0.074108	0.033914	-0.040629	-0.027014	0.007079	-0.038036	-0.007393	
98	0.217638	-0.022910	0.168145	0.035309	0.095543	0.076718	0.141923	
99	0.019018	-0.046037	-0.076364	0.038581	0.113805	0.027334	0.024965	

	X7	X8	X9	X10	X11	X12	X13
0	-0.078480	0.051717	-0.040286	-0.046466	0.043594	0.034011	efectores
1	-0.062214	0.007660	0.041061	0.042803	-0.002874	0.033929	efectores
2	-0.006032	0.015766	-0.150838	-0.002576	0.014996	-0.081872	efectores
3	-0.015333	0.101709	0.045478	-0.031599	0.105639	-0.047101	efectores
4	0.013735	-0.056687	-0.074065	0.004819	0.016878	0.079704	efectores
..	…	…	…	…	…	…	…
95	0.000230	0.003870	0.031836	-0.047503	-0.048570	-0.012877	efectores
96	0.010933	0.094244	0.043965	0.071333	0.091996	0.004503	efectores
97	-0.013590	0.062275	0.010002	0.131511	0.008124	0.042546	efectores
98	0.188212	0.061621	-0.089192	0.031964	0.082573	0.037887	efectores
99	0.000714	-0.015229	0.071540	0.134466	0.014104	-0.019980	efectores

```
[84 rows x 14 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	84.000000	84.000000	84.000000	84.000000	84.000000	84.000000	
mean	0.014326	0.022744	-0.002117	0.014598	-0.000289	0.018958	
std	0.076734	0.062037	0.068489	0.066372	0.080139	0.073552	
min	-0.211447	-0.077600	-0.170590	-0.155161	-0.177306	-0.136557	
25%	-0.031026	-0.015875	-0.036457	-0.023776	-0.052919	-0.025997	
50%	0.013860	0.016230	0.008591	0.009663	0.000958	0.013172	
75%	0.055894	0.048745	0.034872	0.051962	0.040105	0.055826	

```
max     0.275770   0.312105   0.168145   0.251137   0.193364   0.275986

              X6         X7         X8         X9        X10        X11  \
count  84.000000  84.000000  84.000000  84.000000  84.000000  84.000000
mean    0.015966   0.004571   0.032357   0.011648   0.030028   0.013707
std     0.061761   0.069297   0.076737   0.074454   0.070514   0.068456
min    -0.128069  -0.282952  -0.120373  -0.199405  -0.136223  -0.213021
25%    -0.024080  -0.022060  -0.015609  -0.026259  -0.019110  -0.027727
50%     0.010818   0.001064   0.032082   0.018323   0.024039   0.015937
75%     0.036948   0.030223   0.075510   0.056282   0.066129   0.050366
max     0.242701   0.206014   0.304270   0.182667   0.287288   0.209694

             X12
count  84.000000
mean    0.019290
std     0.069218
min    -0.095788
25%    -0.027115
50%     0.005510
75%     0.060535
max     0.283589
```

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

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.059937 -0.003327 -0.007780  0.039320 -0.019375 -0.037437  0.022728
1    0.131313  0.021171  0.002833 -0.001828  0.087780 -0.050239 -0.004440
3    0.039179  0.026931  0.034465  0.060948  0.046050  0.087206  0.052078
4   -0.006290  0.110345  0.103445 -0.010907  0.058920  0.033896 -0.100157
5   -0.007292 -0.010907 -0.012342  0.013719 -0.010260  0.072002  0.056594
..        …         …         …         …         …         …         …
495 -0.135027  0.097400 -0.067243 -0.099122  0.044400 -0.055198 -0.052472
496  0.109624  0.078842  0.078484  0.027877  0.050332 -0.018166 -0.006477
497 -0.021563  0.013939  0.026374 -0.051255 -0.014923  0.022753  0.034066
498  0.033112  0.046003  0.085295  0.032019 -0.053626  0.088083  0.047758
499  0.025865  0.012384 -0.014746 -0.001325  0.030411  0.040879 -0.012147

           X7        X8        X9       X10       X11       X12           X13
0   -0.008882 -0.032152 -0.101270  0.011480  0.006429  0.012883  no_efectores
1   -0.007733 -0.044287  0.018601 -0.010111  0.009493 -0.032125  no_efectores
3    0.007955  0.051125  0.024718  0.021189  0.020654  0.031834  no_efectores
4   -0.099986 -0.083866 -0.089299  0.083480 -0.123655  0.015975  no_efectores
5   -0.078320  0.044407  0.072148 -0.003260  0.017709 -0.007980  no_efectores
..        …         …         …         …         …         …           …
```

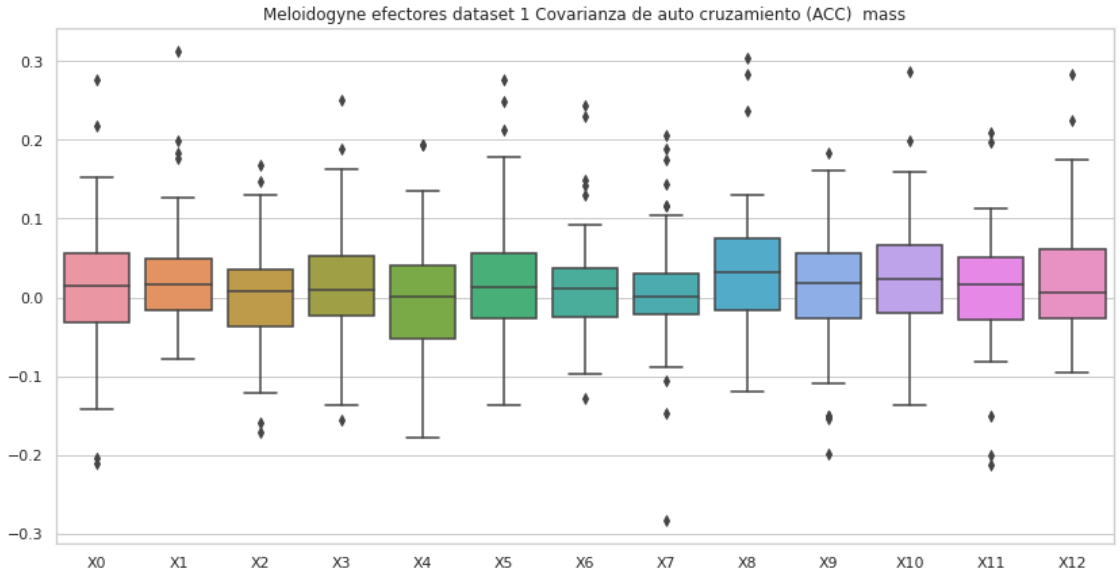
```
495  0.028444  0.004482  0.086660  0.035851 -0.120097 -0.010234  no_efectores
496 -0.006632 -0.040699 -0.049981 -0.007190 -0.004306 -0.008853  no_efectores
497 -0.019285  0.015876 -0.000186 -0.047112  0.030711  0.000095  no_efectores
498 -0.114249 -0.023544 -0.060544 -0.090824 -0.083271  0.062528  no_efectores
499  0.009902 -0.007093  0.016324 -0.007837 -0.008228  0.006441  no_efectores

[456 rows x 14 columns]
```

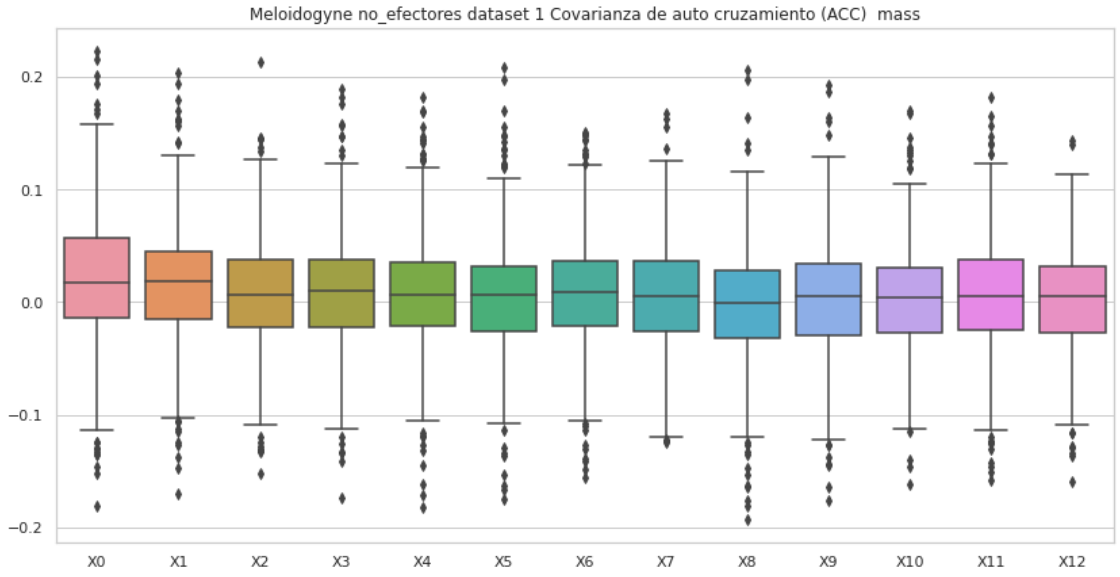
Covarianza de auto cruzamiento (ACC) mass no_efectores Meloidogyne dataset 1, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5
count	456.000000	456.000000	456.000000	456.000000	456.000000	456.000000
mean	0.019814	0.018380	0.006982	0.008869	0.007478	0.004227
std	0.060791	0.054054	0.049099	0.050641	0.053235	0.053173
min	-0.181387	-0.170331	-0.152435	-0.174212	-0.182020	-0.174870
25%	-0.014229	-0.014663	-0.022920	-0.022911	-0.021228	-0.025399
50%	0.017430	0.018297	0.006714	0.010187	0.005982	0.006418
75%	0.057422	0.045132	0.037655	0.037792	0.035850	0.032301
max	0.222803	0.203474	0.212712	0.189071	0.181450	0.208386

	X6	X7	X8	X9	X10	X11
count	456.000000	456.000000	456.000000	456.000000	456.000000	456.000000
mean	0.008773	0.004404	-0.001186	0.003692	0.001724	0.005768
std	0.050780	0.048687	0.054435	0.053036	0.048976	0.050986
min	-0.156282	-0.124486	-0.193017	-0.176600	-0.161668	-0.158056
25%	-0.020935	-0.025858	-0.032437	-0.029603	-0.027084	-0.024131
50%	0.009336	0.005221	-0.000086	0.004800	0.003878	0.005592
75%	0.036527	0.036678	0.028622	0.034134	0.030015	0.037491
max	0.150674	0.168036	0.205748	0.193099	0.170511	0.182588

	X12
count	456.000000
mean	0.000779
std	0.049188
min	-0.159431
25%	-0.027252
50%	0.005030
75%	0.031511
max	0.143565





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

          X0        X1        X2        X3        X4        X5        X6  \
0   0.081814  0.074866  0.230401  0.168680 -0.020038  0.085978  0.067126
1  -0.032596  0.193662  0.162619  0.019286  0.173091  0.175636  0.086643
2   0.056670  0.009894 -0.075014 -0.047856 -0.043665 -0.115383  0.065359
3  -0.058783  0.071121 -0.121116  0.017349 -0.183891 -0.003685 -0.060185
4   0.044512  0.054776  0.080541  0.081847  0.123093 -0.070253 -0.003033
..       …         …         …         …         …         …         …
95  0.050216  0.005412  0.151306  0.141364 -0.027907  0.088667  0.137127
96 -0.047844 -0.044908  0.024577 -0.016167 -0.002997  0.001896  0.013139
97  0.132159  0.068076  0.039767  0.019330  0.092448  0.120548  0.062074
98  0.245521 -0.036898  0.080500  0.127070  0.141525  0.041621 -0.059005
99  0.006297  0.103350 -0.020410  0.011157  0.011883  0.110948  0.019616

          X7        X8        X9       X10       X11       X12        X13
0   0.083150  0.046669  0.091717  0.107026  0.023249  0.092277  efectores
1   0.072397  0.013966  0.123588 -0.032599  0.144821  0.045807  efectores
2   0.033725 -0.054770  0.066835  0.056242  0.009557  0.082798  efectores
```

```
3  -0.011846  0.121214  0.007869  0.073367 -0.086673  0.080207  efectores
4   0.038158  0.007343  0.040385  0.073230 -0.016757  0.126193  efectores
..       …       …       …       …       …       …       …
95 -0.017147  0.142449 -0.007126  0.027827 -0.062784 -0.049186  efectores
96  0.022979  0.015500 -0.011891  0.020419 -0.011234 -0.008283  efectores
97 -0.023000  0.014197 -0.030757  0.019672 -0.009433 -0.111334  efectores
98  0.013783  0.130014  0.090630  0.145048 -0.021508  0.042552  efectores
99 -0.085058  0.044598 -0.062720 -0.022593  0.041408  0.020963  efectores

[100 rows x 14 columns]
```

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

```
                X0          X1          X2          X3          X4          X5  \
count  100.000000  100.000000  100.000000  100.000000  100.000000  100.000000
mean     0.065805    0.073121    0.080853    0.078551    0.054273    0.047635
std      0.134958    0.154316    0.135988    0.115554    0.139850    0.111698
min     -0.141160   -0.190564   -0.121116   -0.106794   -0.183891   -0.168304
25%     -0.017606   -0.020152   -0.016953    0.010876   -0.028434   -0.016414
50%      0.025328    0.047097    0.057195    0.050734    0.036100    0.039599
75%      0.119325    0.116934    0.152420    0.129356    0.111748    0.091576
max      0.647257    0.751686    0.740489    0.569969    0.605469    0.508538

                X6          X7          X8          X9         X10         X11  \
count  100.000000  100.000000  100.000000  100.000000  100.000000  100.000000
mean     0.065069    0.041965    0.038773    0.036822    0.043183    0.024770
std      0.115078    0.104745    0.113575    0.090478    0.091666    0.101368
min     -0.157831   -0.128874   -0.346655   -0.121852   -0.256946   -0.318202
25%     -0.014556   -0.011694   -0.023286   -0.021089   -0.011640   -0.036197
50%      0.050092    0.021485    0.033481    0.021949    0.023094    0.010854
75%      0.115384    0.083178    0.094601    0.079293    0.084069    0.085098
max      0.474328    0.482565    0.394873    0.341708    0.323615    0.333918

               X12
count  100.000000
mean     0.019454
std      0.113268
min     -0.419976
25%     -0.029486
50%      0.019555
75%      0.081229
max      0.398254
```

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

	X0	X1	X2	X3	X4	X5	X6	\
0	-0.010073	-0.129066	0.118455	0.027531	-0.044698	0.022677	0.002645	
1	-0.172111	-0.001042	0.014270	0.016890	-0.109254	-0.115543	-0.096073	
2	-0.074897	-0.036676	0.096713	-0.185106	0.154392	-0.010288	0.016213	
3	0.073114	0.022243	0.050390	0.019120	-0.008882	-0.029968	0.055396	
4	-0.202295	0.066436	-0.069340	-0.233333	-0.094012	0.056514	-0.026992	
..	…	…	…	…	…	…	…	
495	0.151387	0.052575	0.073921	0.173208	0.062970	0.123454	0.194388	
496	0.030683	0.060312	0.124983	0.068563	-0.041006	-0.047499	0.073248	
497	0.038112	0.024989	-0.006509	0.012630	-0.016461	-0.042168	0.010432	
498	-0.064152	0.173932	0.285552	-0.130835	-0.042775	-0.117084	-0.081570	
499	0.134715	0.088043	0.053630	0.070698	0.086241	-0.066103	0.060703	

	X7	X8	X9	X10	X11	X12	X13
0	-0.038301	0.062924	0.057492	0.033098	-0.062308	0.047709	no_efectores
1	0.020725	0.019390	0.209881	-0.000027	0.105658	0.150697	no_efectores
2	0.142422	-0.187000	-0.030916	-0.057814	-0.224040	0.067380	no_efectores
3	0.054556	0.011305	-0.003807	0.088741	0.061733	0.032338	no_efectores
4	0.134596	-0.008161	0.079045	0.043786	-0.014975	-0.049142	no_efectores
..	…	…	…	…	…	…	…
495	0.184787	-0.000140	0.049963	0.074354	0.073359	0.132440	no_efectores
496	-0.014000	-0.008610	0.062009	-0.001693	-0.014176	-0.004159	no_efectores
497	0.005248	-0.009830	0.007287	-0.019873	0.009390	-0.048494	no_efectores
498	-0.229033	0.077036	-0.040822	0.087209	0.033712	0.171807	no_efectores
499	0.023428	0.083978	0.017460	0.044132	-0.026646	0.010579	no_efectores

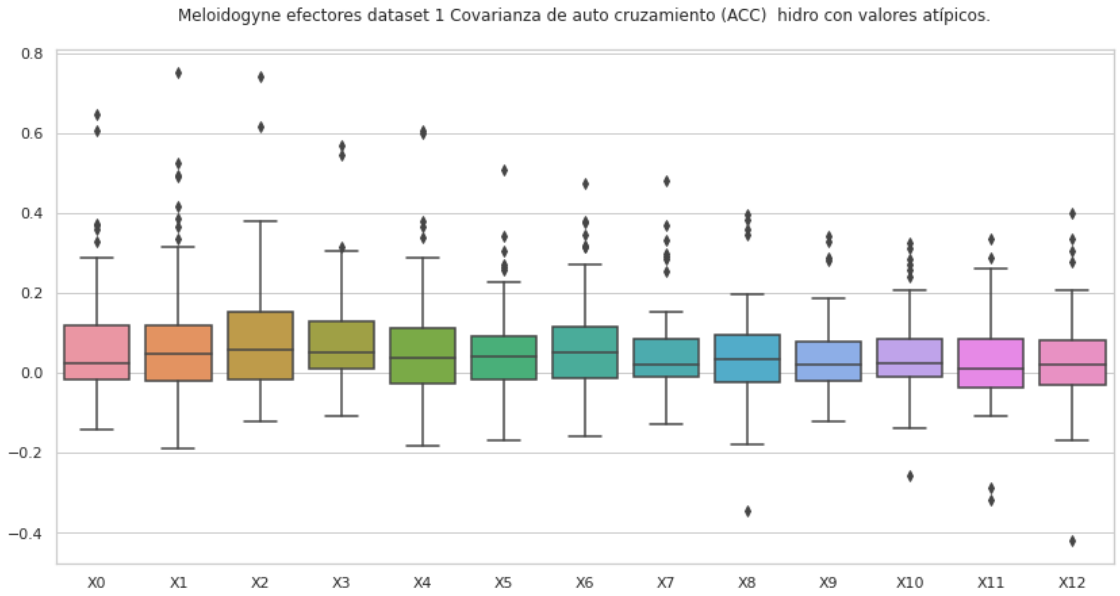
[500 rows x 14 columns]

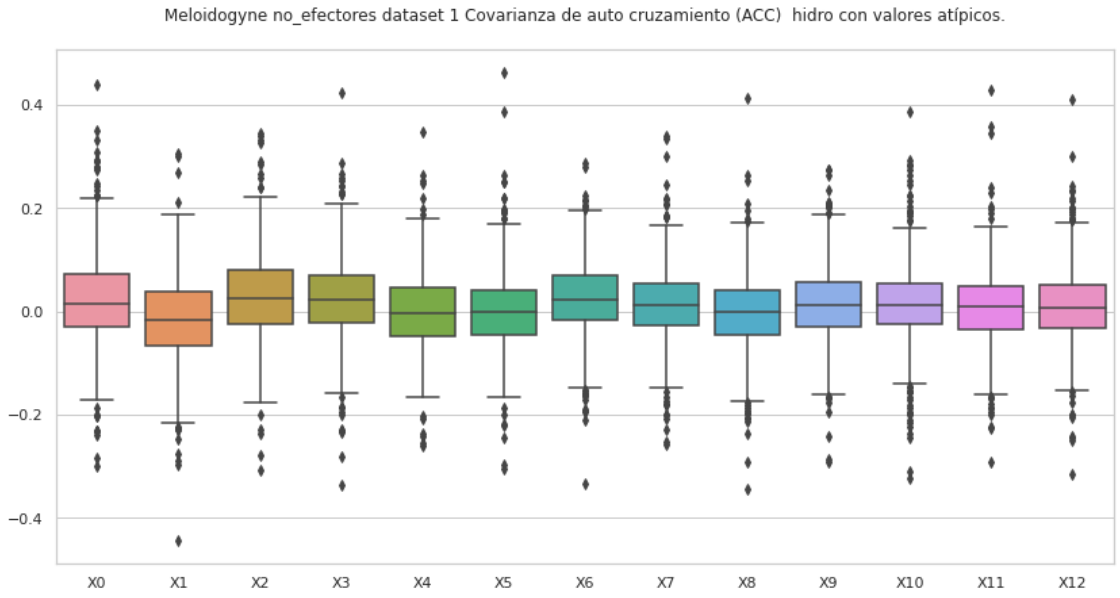
Covarianza de auto cruzamiento (ACC) hidro no_efectores Meloidogyne dataset 1,
con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.022599	-0.017354	0.031054	0.023766	0.001245	0.001740	
std	0.094571	0.086576	0.086169	0.085503	0.078183	0.083362	
min	-0.300005	-0.444183	-0.308136	-0.336379	-0.259764	-0.306090	
25%	-0.029255	-0.067213	-0.024752	-0.022584	-0.048317	-0.046176	
50%	0.014869	-0.016703	0.023835	0.022566	-0.002777	-0.000370	
75%	0.071999	0.037504	0.079889	0.070443	0.045991	0.041566	
max	0.439024	0.306366	0.344032	0.422439	0.348385	0.462460	

	X6	X7	X8	X9	X10	X11	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.023207	0.013011	-0.003201	0.016167	0.014367	0.007220	
std	0.077777	0.077407	0.081725	0.078672	0.083988	0.079133	
min	-0.334932	-0.256983	-0.343519	-0.292531	-0.322486	-0.293294	
25%	-0.018125	-0.027815	-0.045380	-0.031229	-0.024139	-0.035542	
50%	0.022548	0.011196	-0.000262	0.012854	0.012405	0.010108	
75%	0.068288	0.052669	0.041645	0.057330	0.054545	0.049277	
max	0.286149	0.339951	0.412893	0.273447	0.385723	0.427832	

	X12
count	500.000000
mean	0.008985
std	0.080316
min	-0.315091
25%	-0.032045
50%	0.007198
75%	0.051033
max	0.409959





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

          X0        X1        X2        X3        X4        X5        X6  \
0   0.081814  0.074866  0.230401  0.168680 -0.020038  0.085978  0.067126
1  -0.032596  0.193662  0.162619  0.019286  0.173091  0.175636  0.086643
2   0.056670  0.009894 -0.075014 -0.047856 -0.043665 -0.115383  0.065359
3  -0.058783  0.071121 -0.121116  0.017349 -0.183891 -0.003685 -0.060185
4   0.044512  0.054776  0.080541  0.081847  0.123093 -0.070253 -0.003033
..       …         …         …         …         …         …         …
95  0.050216  0.005412  0.151306  0.141364 -0.027907  0.088667  0.137127
96 -0.047844 -0.044908  0.024577 -0.016167 -0.002997  0.001896  0.013139
97  0.132159  0.068076  0.039767  0.019330  0.092448  0.120548  0.062074
98  0.245521 -0.036898  0.080500  0.127070  0.141525  0.041621 -0.059005
99  0.006297  0.103350 -0.020410  0.011157  0.011883  0.110948  0.019616

          X7        X8        X9       X10       X11       X12        X13
0   0.083150  0.046669  0.091717  0.107026  0.023249  0.092277  efectores
1   0.072397  0.013966  0.123588 -0.032599  0.144821  0.045807  efectores
2   0.033725 -0.054770  0.066835  0.056242  0.009557  0.082798  efectores
3  -0.011846  0.121214  0.007869  0.073367 -0.086673  0.080207  efectores
4   0.038158  0.007343  0.040385  0.073230 -0.016757  0.126193  efectores
..       …         …         …         …         …         …         …
95 -0.017147  0.142449 -0.007126  0.027827 -0.062784 -0.049186  efectores
```

```
96  0.022979  0.015500 -0.011891  0.020419 -0.011234 -0.008283  efectores
97 -0.023000  0.014197 -0.030757  0.019672 -0.009433 -0.111334  efectores
98  0.013783  0.130014  0.090630  0.145048 -0.021508  0.042552  efectores
99 -0.085058  0.044598 -0.062720 -0.022593  0.041408  0.020963  efectores

[94 rows x 14 columns]
```

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

	X0	X1	X2	X3	X4	X5	\
count	94.000000	94.000000	94.000000	94.000000	94.000000	94.000000	
mean	0.050064	0.052103	0.062618	0.064175	0.038053	0.031668	
std	0.099871	0.124507	0.098348	0.088386	0.112536	0.090521	
min	-0.141160	-0.190564	-0.121116	-0.106794	-0.183891	-0.168304	
25%	-0.018833	-0.034307	-0.018021	0.010313	-0.029488	-0.033202	
50%	0.023122	0.040502	0.052706	0.040777	0.032853	0.036252	
75%	0.108581	0.104154	0.149771	0.116814	0.106777	0.082328	
max	0.369266	0.493866	0.379847	0.301968	0.377851	0.271424	

	X6	X7	X8	X9	X10	X11	\
count	94.000000	94.000000	94.000000	94.000000	94.000000	94.000000	
mean	0.054611	0.026507	0.034795	0.027796	0.040098	0.024852	
std	0.097174	0.081889	0.089058	0.076150	0.076718	0.081904	
min	-0.106601	-0.128874	-0.179100	-0.121852	-0.137637	-0.110051	
25%	-0.016676	-0.015822	-0.019491	-0.023928	-0.009049	-0.035941	
50%	0.045850	0.018890	0.033481	0.021445	0.023094	0.010854	
75%	0.101689	0.071927	0.089826	0.066503	0.077116	0.082106	
max	0.377863	0.330375	0.343878	0.287837	0.310060	0.287351	

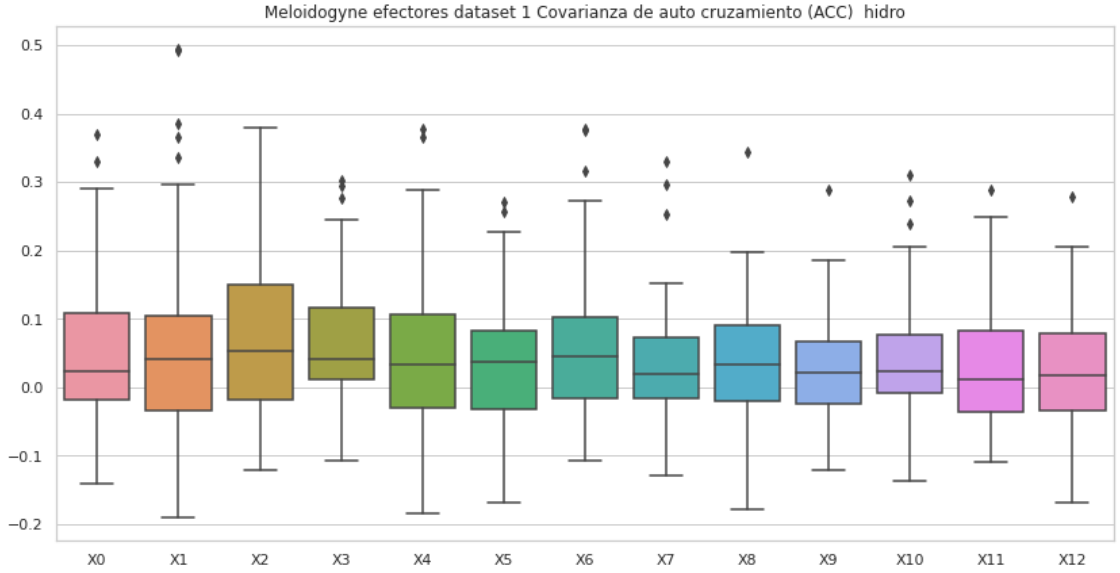
	X12
count	94.000000
mean	0.013849
std	0.089650
min	-0.169147
25%	-0.033692
50%	0.017390
75%	0.079198
max	0.279077

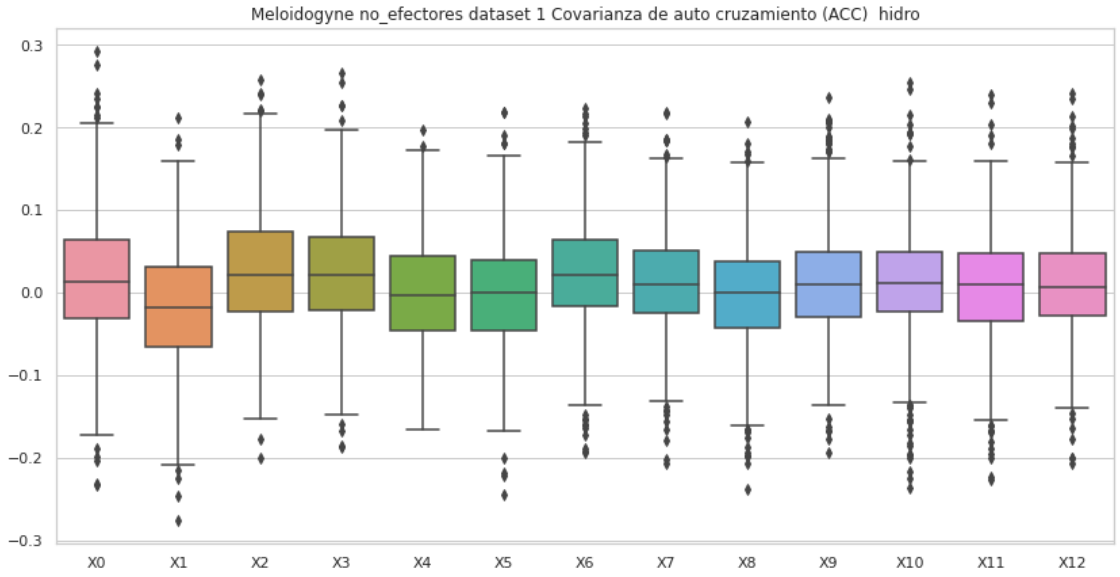
no_efectores
Covarianza de auto cruzamiento (ACC) no_efectores Meloidogyne dataset 1, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0   -0.010073 -0.129066  0.118455  0.027531 -0.044698  0.022677  0.002645
1   -0.172111 -0.001042  0.014270  0.016890 -0.109254 -0.115543 -0.096073
2   -0.074897 -0.036676  0.096713 -0.185106  0.154392 -0.010288  0.016213
3    0.073114  0.022243  0.050390  0.019120 -0.008882 -0.029968  0.055396
5   -0.024853  0.048625 -0.005367  0.043654 -0.012008  0.059723  0.109996
..        …         …         …         …         …         …         …
494 -0.020326 -0.053657 -0.059371  0.016445 -0.024973 -0.051257 -0.072410
495  0.151387  0.052575  0.073921  0.173208  0.062970  0.123454  0.194388
496  0.030683  0.060312  0.124983  0.068563 -0.041006 -0.047499  0.073248
497  0.038112  0.024989 -0.006509  0.012630 -0.016461 -0.042168  0.010432
499  0.134715  0.088043  0.053630  0.070698  0.086241 -0.066103  0.060703

           X7        X8        X9       X10       X11       X12           X13
0   -0.038301  0.062924  0.057492  0.033098 -0.062308  0.047709  no_efectores
1    0.020725  0.019390  0.209881 -0.000027  0.105658  0.150697  no_efectores
2    0.142422 -0.187000 -0.030916 -0.057814 -0.224040  0.067380  no_efectores
3    0.054556  0.011305 -0.003807  0.088741  0.061733  0.032338  no_efectores
5   -0.024919  0.107208 -0.108130  0.049811  0.063279  0.049930  no_efectores
..        …         …         …         …         …         …           …
494 -0.069856 -0.012177 -0.039956  0.041874 -0.029973  0.019629  no_efectores
495  0.184787 -0.000140  0.049963  0.074354  0.073359  0.132440  no_efectores
496 -0.014000 -0.008610  0.062009 -0.001693 -0.014176 -0.004159  no_efectores
497  0.005248 -0.009830  0.007287 -0.019873  0.009390 -0.048494  no_efectores
499  0.023428  0.083978  0.017460  0.044132 -0.026646  0.010579  no_efectores

[454 rows x 14 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  454.000000  454.000000  454.000000  454.000000  454.000000  454.000000
mean     0.016788   -0.019047    0.027186    0.023005    0.000611   -0.001222
std      0.080304    0.075320    0.072802    0.072195    0.066945    0.071249
min     -0.232597   -0.275795   -0.200167   -0.187164   -0.165898   -0.245313
25%     -0.030449   -0.065304   -0.023244   -0.020838   -0.045902   -0.045862
50%      0.012247   -0.018465    0.021647    0.021922   -0.003413   -0.000829
75%      0.064076    0.030760    0.073866    0.067567    0.043551    0.039847
max      0.291928    0.211442    0.257937    0.266204    0.197408    0.218165

                X6          X7          X8          X9         X10         X11  \
count  454.000000  454.000000  454.000000  454.000000  454.000000  454.000000
mean     0.021053    0.012354   -0.003565    0.013185    0.011431    0.005200
std      0.071191    0.066225    0.070103    0.069597    0.072165    0.071481
```

min	-0.194181	-0.207581	-0.238305	-0.194153	-0.236082	-0.226125
25%	-0.016886	-0.024661	-0.043427	-0.030201	-0.023820	-0.034816
50%	0.021752	0.010296	-0.000086	0.009447	0.012061	0.010108
75%	0.063514	0.050085	0.037605	0.049302	0.049795	0.047515
max	0.223786	0.218310	0.207455	0.235809	0.254224	0.239789

	X12
count	454.000000
mean	0.008252
std	0.068662
min	-0.206710
25%	-0.028775
50%	0.007198
75%	0.047644
max	0.241325




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