

1 Background and Data

The assignment material consists of several cases of industrial process data, as produced by the Tennessee Eastman process simulator (see the provided article by Russel et al.). There are 22 different data sets, each consisting of 960 observations of 52 process variables. One of them, the file `d00_te.dat` contains observations from the process under normal conditions, while all the rest (`d01_te.dat`, ..., `d21_te.dat`) present a fault, some behavior that deviates from the reference case in `d00_te.dat`.

2 Problem Statement

The task is to analyze the fault situations by PCA methods (but naturally other methods are allowed as well). Your task is to try to identify the following:

- When the fault behavior starts.
- Does the process return to normal behavior or stay abnormal.
- Which process variables cause the fault behavior.

The data sets give time series observations, so the consecutive data points are correlated. Use both the PCA and DPCA projection methods. Another option is to use, for instance, moving averages to further take the time correlations into account. Test different PCA dimensions, and time lags in DPCA.

Each group should study 3 data sets, starting from case 4. The other two cases can be freely chosen among the fault cases 1-21. However, one of them must be either case 3, 9 or 15, considered as ‘difficult’ in the Russel et al. article.

Note: from the Russel article, only the PCA and DPCA approaches are needed, together with T2 and Q statistics, otherwise it is background reading (e.g., the threshold statistics of Eq. (4) or (6) need not be used).

3 General instructions

The general instructions are given as a separate document. They include the requirements, deadline and information on the grading.

If there are any problems with the assignment description and/or data, contact the person supervising the practical assignment. This should be

done before inventing your own interpretations or making (too) radical assumptions.