CHAPTER 3

PRE-PROCESSING AND FEATURE EXTRACTION

In this chapter it is discussed how to provide feature signals. The primary objective is to find a method that gives proper feature signals, i.e. signals in which the engine condition changes are detectable. The point of development is that the human eye must be able to see the engine condition changes in the feature signals in order to make the change detection automatic. If a change is hidden for the human eye, then it is also hidden for the automatic change detection. Two approaches are described, the mean value and the standard deviation of the cycles, and the residual error when applying principal components analysis. Also a tempting approach, which might improve the feature signals, will be described.

As mentioned in the above section, the primary objective is to find adequate feature signals. This chapter will not include optimization of the methods in order to improve the feature signals, because chapter 4 will have a main focus on the segmentation task. Of course, some considerations have to be done, but as soon as a proper feature signal is determined, investigation on the method stops.

3.1 Mean value and standard deviation

In the introduction the motivations behind choosing this method were given. The hypothesis was that when the engine changes condition, the number of AE's increases, causing the AE-RMS signals also to increase. If the hypothesis is true, then the engine condition changes will be directly observable in the mean value of the cycles. Figure 3.1-4, show the mean values of the cycles from the four AE-RMS sensors. From this it can be concluded that all experiments, but the unstable region "experiment" can be observed, thus there is a good chance of making an automatic detection of the experiments. However, sensor no. 2 and 4 show clearer changes.

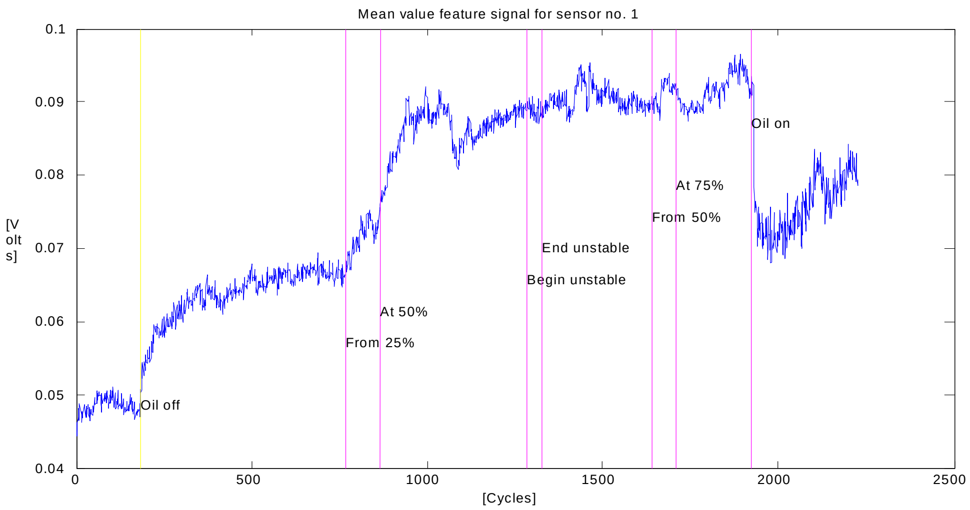

Figure 3.1: Mean value of cycles from sensor no. 1. All the changes but the unstable region can be observed.

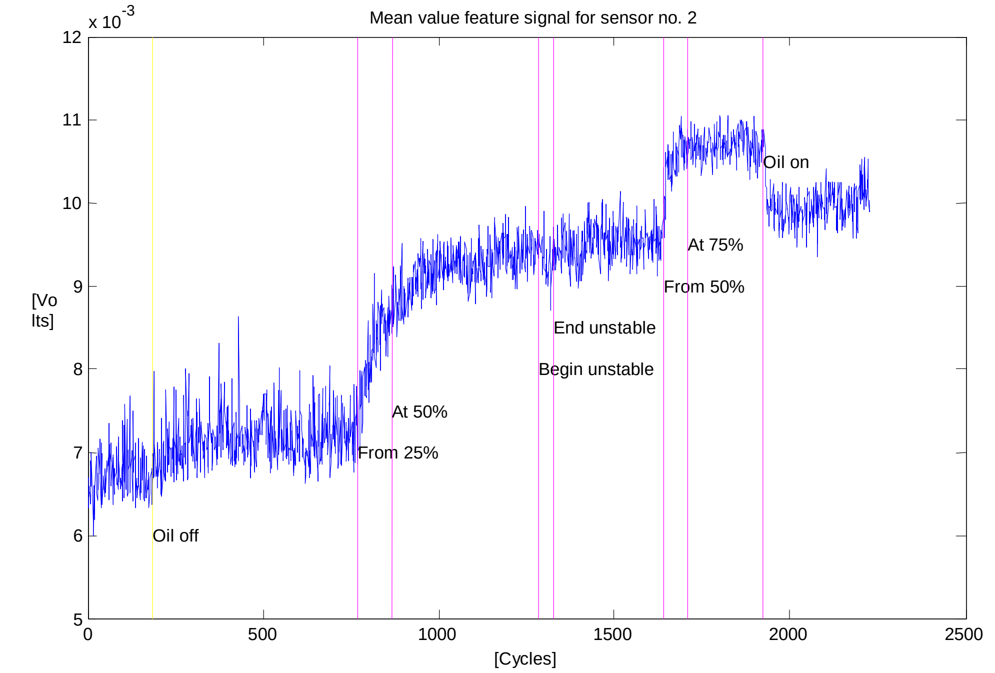

Figure 3.2: *Mean value of cycles from sensor no. 2. All the changes but the unstable region can be observed. This is one of the clearest mean value feature signals.*

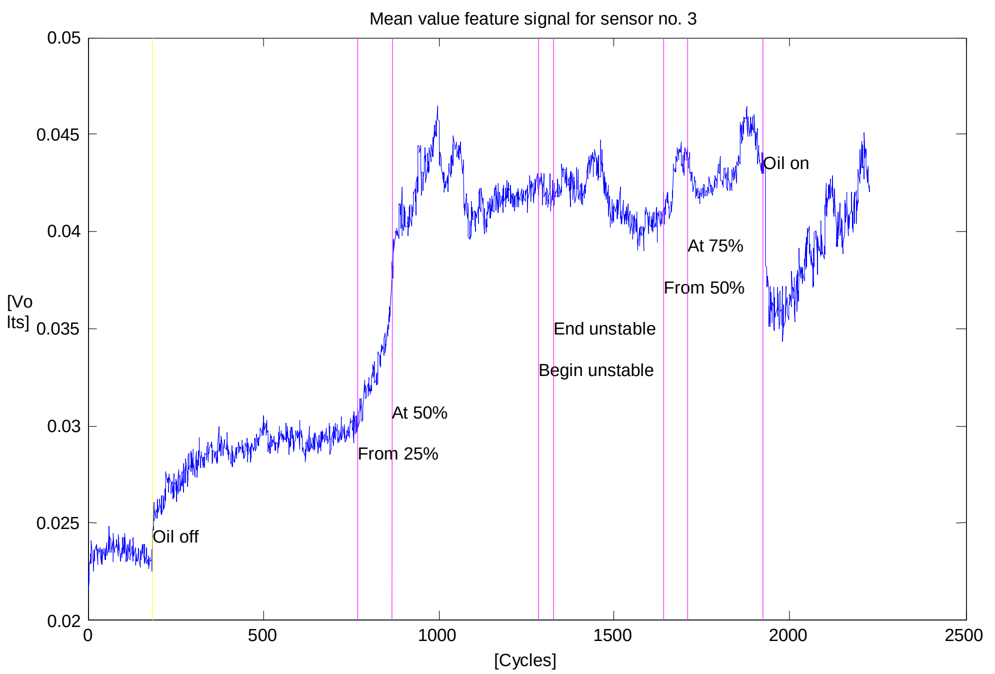

Figure 3.3: *Mean value of cycles from sensor no. 3. All the changes but the unstable region can be observed.*

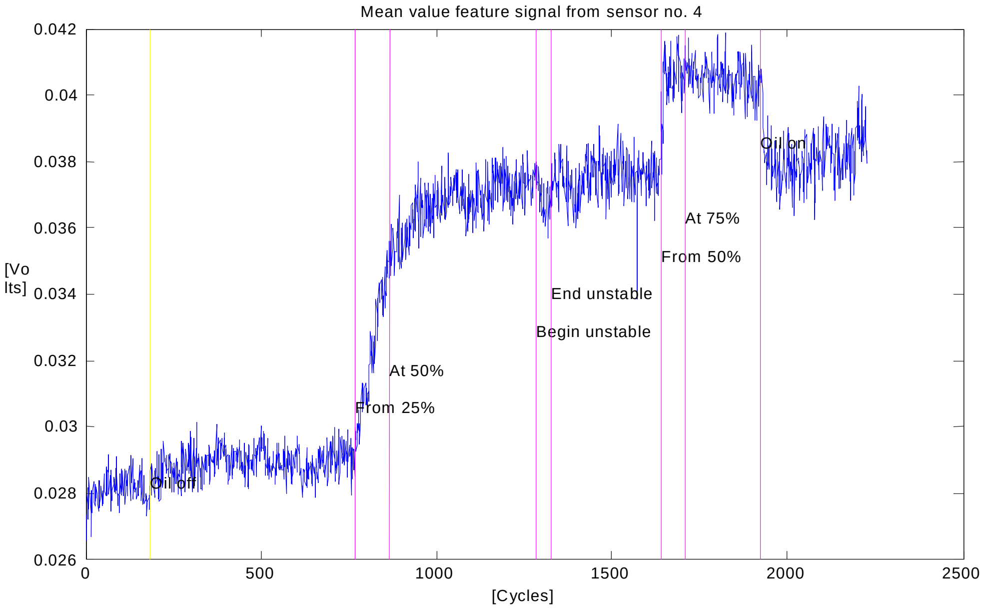

Figure 3.4: *Mean value of cycles from sensor no. 4. All the changes but the unstable region can be observed. This is one of the clearest mean value feature signals.*

The standard deviation of the cycles is shown in figure 3.5-7. Comparing these feature signals with the feature signals in figure 3.1-4 reveals that more engine condition changes are observable in the mean value feature signals. Thus, it could be argued that only the mean values should be used as features, but will not be discussed in this thesis. All eight feature signals are investigated in the test section, chapter 5.

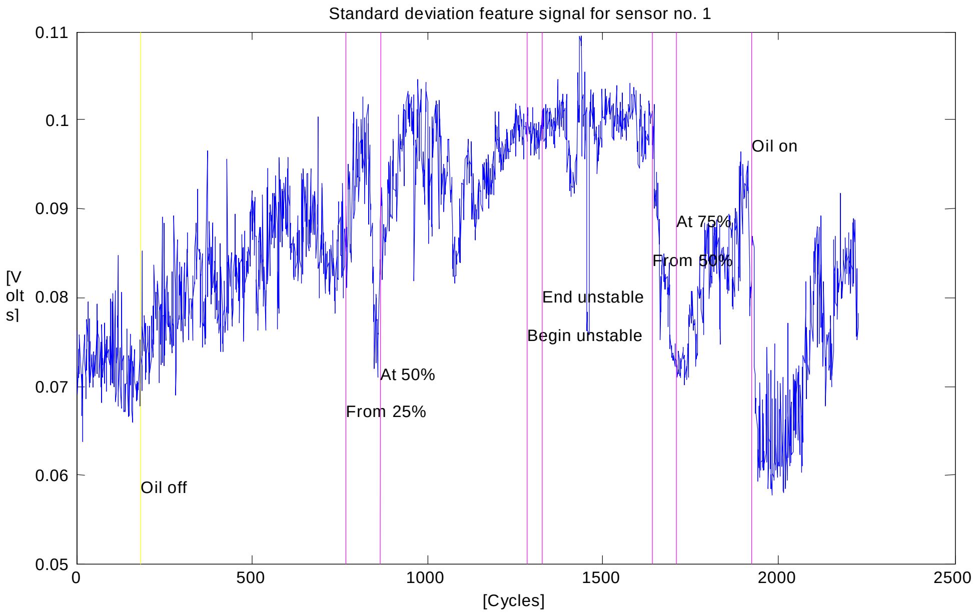

Figure 3.5: *Standard deviation of cycles from sensor no. 1. Not all changes can be observed.*

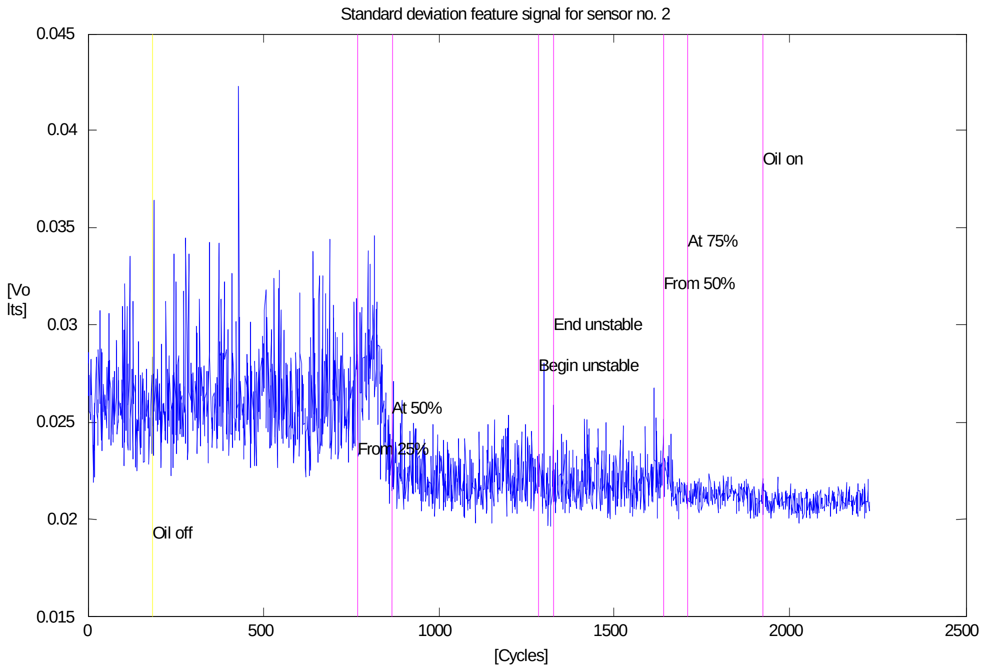

Figure 3.6: Standard deviation of cycles from sensor no. 2. Not all changes can be observed. Again sensor no. 2 provides a feature signal, which is clearer than the feature signals from the other sensors.

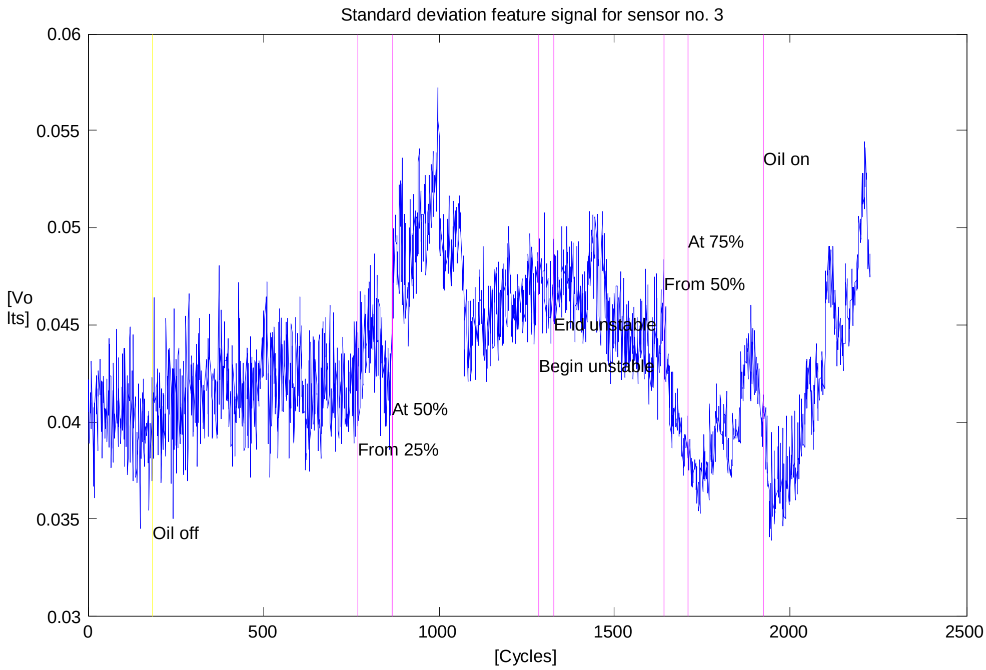

Figure 3.7: Standard deviation of cycles from sensor no. 3. Not all changes can be observed.

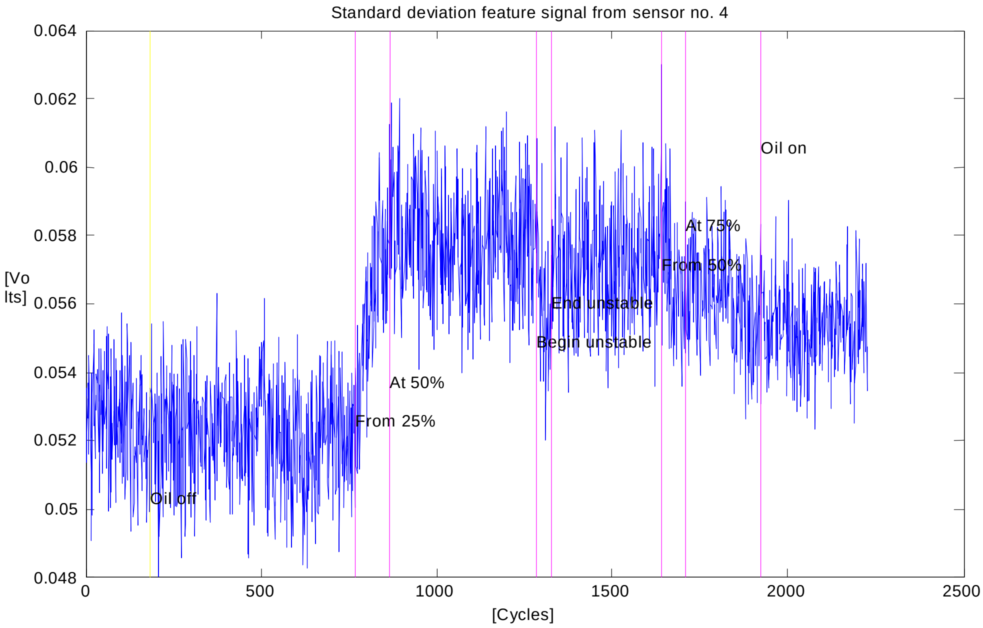

Figure 3.8: *Standard deviation of cycles from sensor no. 4. Not all changes can be observed. Again sensor no. 4 provides a clear feature signal.*

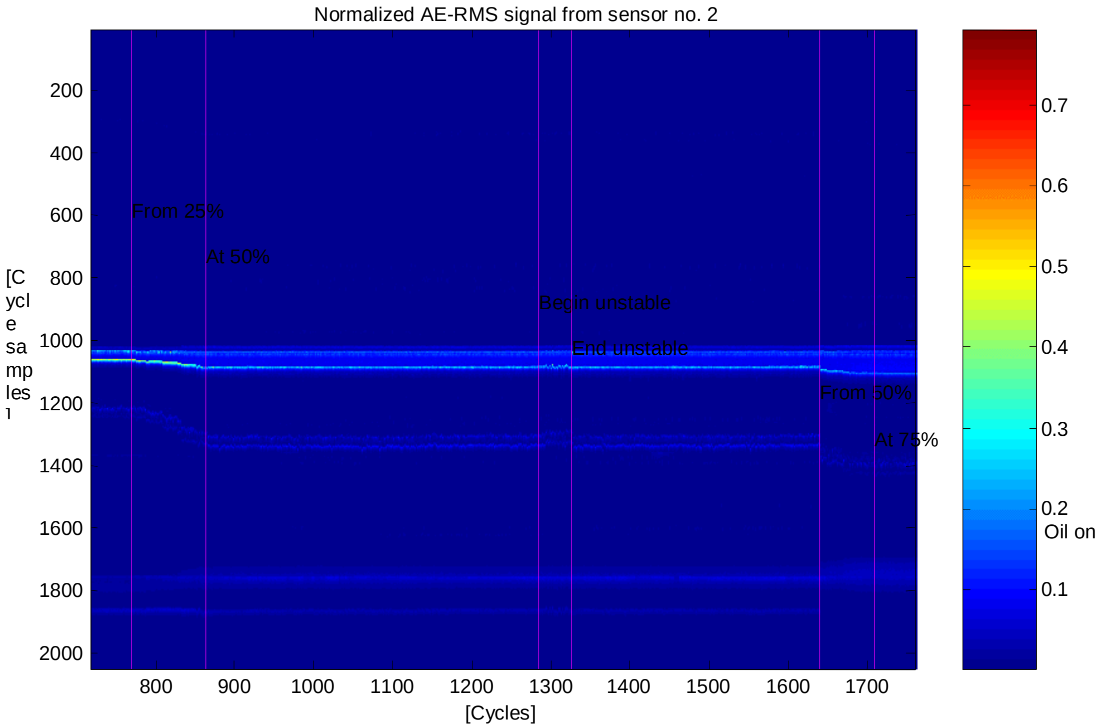

Figure 3.9: *Time plot of the AE-RMS cycles from sensor no. 2. Notice that the peaks seem to shift in the unstable region, but keep their magnitudes. Thus the mean value and the standard deviation remains almost constant.*

The mean value and standard deviation approach is a very simple approach, which reduces the input dimensionality from 2,048 to 2. However, since it fails to ”respond” to the unstable region, which in fact is an error that can occur in reality, then another approach is wanted. From figure 3.9, it can be suggested why the approach fails. The figure shows the AE-RMS cycles from sensor no. 2. When the unstable region is reached, the cycles undergo some kind of shift, but they keep their magnitudes. If the mean value and the standard deviation are calculated in these cycles, they will be quite similar to the cycles just before the unstable region. In this way the unstable region will be unobservable[1].

3.2 Principal components analysis – residual error

PCA is a classical dimension reducing method. It shows good performance when there is a high correlation between the observations (cycles). Usually PCA is used to set up a model for the observations and optimizing this model by minimizing the residual error[2]. However, in this thesis only a *change* in the residual error is sought. Thus, optimization is not a crucial part, but of course realistic models must be applied. Next, a short description on how PCA is applied in this thesis.

The principle steps in PCA are from [2]:

1. Select N examples (cycles) as the training set $\boldsymbol{X}$.
2. Center the training set by subtracting the mean value of the training set. The centered training set is denoted $\tilde{\boldsymbol{X}} \in \boldsymbol{R}^{2,048\times N}$.
3. Calculate the covariance matrix of the centered training set.
4. Calculate the eigenvectores of the covariance matrix. The eigenvectors are denoted as the column vectors in the matrix $\boldsymbol{U} \in \boldsymbol{R}^{2,048\times 2,048}$.
5. Calculate the eigenvalues of the covariance matrix. The eigenvalues are the diagonal of the matrix $\boldsymbol{\Lambda}^{½} \in \boldsymbol{R}^{2,048\times N}$.
6. From the eigenvalues, choose the number c of principal components.
7. Project the cycles on the c eigenvectores corresponding to the c largest eigenvalues.
8. The outcome of step 7 is called the principal components (PC’s).

In fact, what we are looking for are the eigenvalues and the eigenvectors. These can be found by applying a very simple procedure called the Singular Value Decomposition (SVD). This approach is described in [25], and used for this thesis. SVD decomposes the centered training set into the following,

$$\tilde{\boldsymbol{X}} = \boldsymbol{U}\boldsymbol{\Lambda}^{½}\boldsymbol{V}^{\mathrm{T}}, \tag{3.1}$$

where T denotes matrix transposition and $\boldsymbol{V} = \tilde{\boldsymbol{X}}^{\mathrm{T}}\tilde{\boldsymbol{X}} \in \boldsymbol{R}^{N\times N}$ can be ignored in this content. Then the principal components are given by,

[1] The shift is not observed in the corresponding time plots of sensor no. 1 and no. 3, but here no change is observable at all. The second sensor being nearer to the point of AE generation in the unstable region could explain this.

[2] The residual error is the difference between the true observations and the observations given by the model.