Queen's University Belfast

Virtual Engineering Centre

Development and Application of Nonlinear PCA for Fault Diagnosis in Internal Combustion Engines

Uwe Kruger

School of Electronics, Electrical Engineering
& Computer Science
Ashby Building
Stranmillis Road
Belfast, BT9 5AH
Northern Ireland, United Kingdom

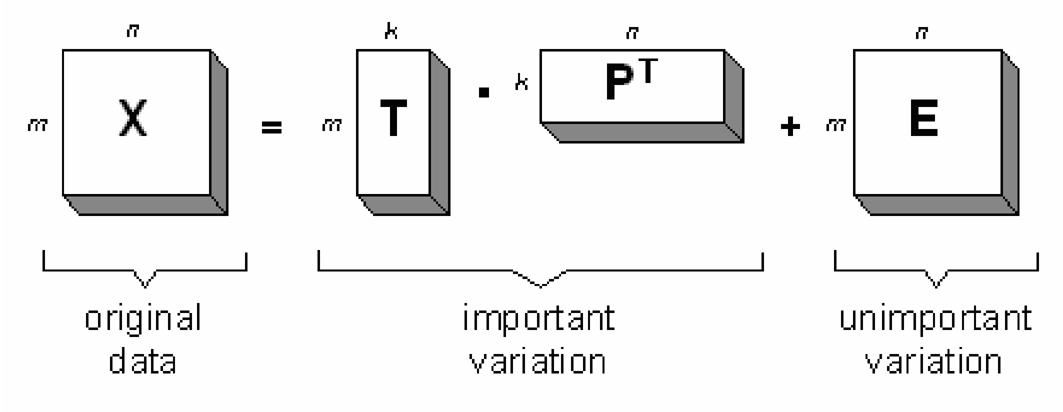
Presentation Outline

- **Introduction into Multivariate Statistical Process Control (MSPC)**
- **Linear MSPC for Condition Monitoring**
- **Limitations and Nonlinear Principal Component Analysis (NLPCA)**
- **Fault Detection using NLPCA**
- **Fault Diagnosis using NLPCA**
- **Application Study to an Internal Combustion Engine**
- **Conclusions and Future Challenges**

- **Principal Component Analysis (PCA) was originated by C. Person (Pearson 1901) and later refined by H. Hotelling in the 1930s.**
- **Partial Least Squares (PLS) algorithm was developed by H. Wold in 1960s (Wold 1966a,1966b)**
- **Because of its simplicity, PCA is predominantly used for condition monitoring**

- **Initial work was done by J. E. Jackson in 1950s (Jackson 1956, 1957, 1959).**
- **Refinements were introduced in the 1980s and 1990s by J. E. Jackson (1980), MacGregor and co-worker (1989, 1991, 1995).**
- **Surveys are available in Kourti and MacGreogor (1995) and Wise and Gallagher (1996)**

- PCA produces the following decomposition of a given data matrix $\mathbf{X}\in\mathbb{R}^{m\times n}$



where $\mathbf{T}\in\mathbb{R}^{m\times n}$ is a score matrix, $\mathbf{P}\in\mathbb{R}^{n\times k}$ is a loading matrix, $\mathbf{E}\in\mathbb{R}^{m\times n}$ is the residual matrix and $k < n$ is the number of retained principal components (PCs)

The PCA Decomposition (2)

- For condition monitoring, significant variation is encapsulated in the *k* retained PCs, whilst insignificant and/or redundant information manifest itself in the residuals.
- The decomposition is based on an eigenvector-eigenvalue decomposition of the data covariance matrix $\mathbf{S}_{XX}$:

$$\mathbf{S}_{XX} = \frac{\mathbf{X}^T\mathbf{X}}{m-1} \in \mathbb{R}^{n \times n} = \begin{bmatrix} \mathbf{P} & \mathbf{P}_0 \end{bmatrix} \begin{bmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_0 \end{bmatrix} \begin{bmatrix} \mathbf{P}^T \\ \mathbf{P}_0^T \end{bmatrix}$$

after the recorded data are mean centred and appropriately scaled.

- Given a data vector $\mathbf{x} \in \mathbb{R}^n$, the score vector can be obtained as follows:

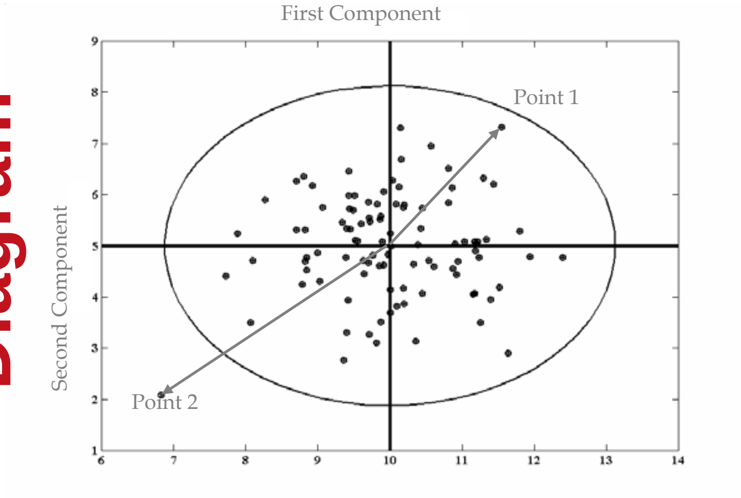
$$\mathbf{t} \in \mathbb{R}^k = \mathbf{P}^T\mathbf{x}$$

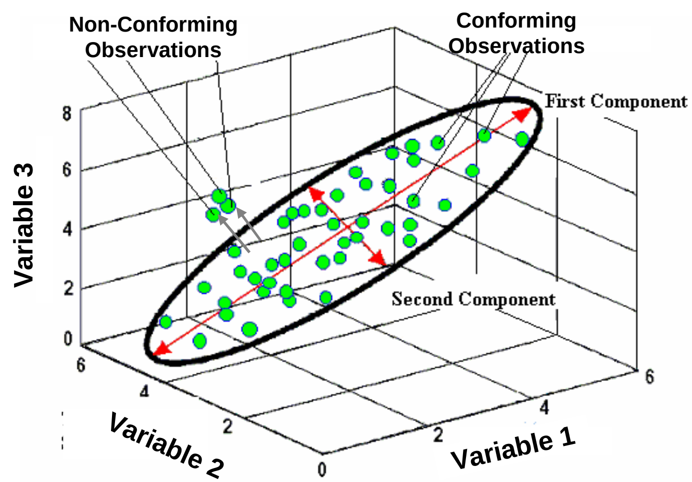
and the residuals are:

$$\mathbf{e} \in \mathbb{R}^n = \mathbf{x} - \mathbf{P}\mathbf{t} = \left[\mathbf{I} - \mathbf{P}\mathbf{P}^T\right]\mathbf{x}$$

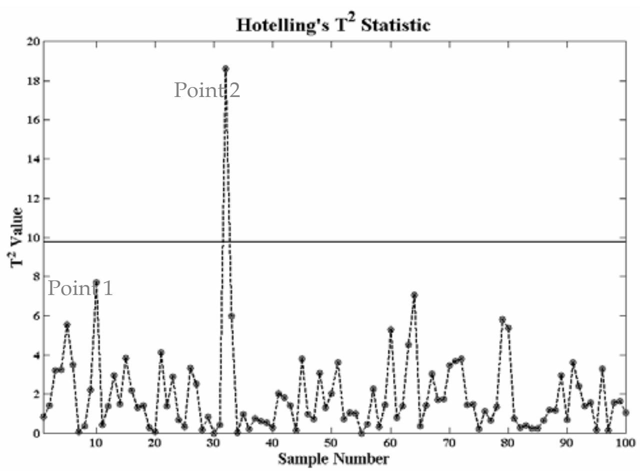
Geometry of the PCA Decomposition

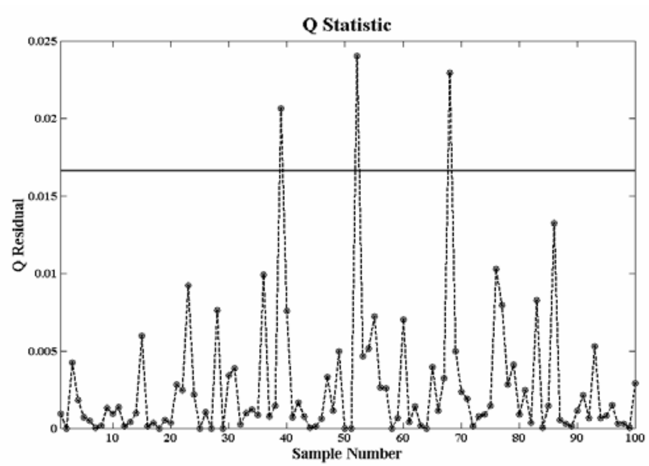
Scatter Diagram





Univariate Statistics





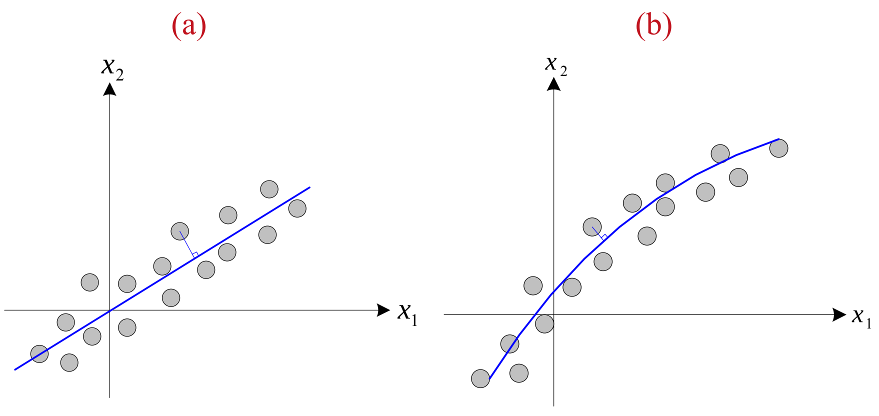
Uwe Kruger

Condition Monitoring Using PCA

- Condition monitoring entails the detection and diagnosis of fault conditions.
- Fault diagnosis can be further divided into fault isolation and identification.
- Fault detection is a Boolean decision, that is there is a fault present in the monitored system or not. This includes the use of scatter diagrams and univariate statistics (shown on the previous slide).
- Fault isolation relates to the determination of the kind and location of a detected fault condition and includes the application of contribution charts
- Fault identification determines size and time-varying behaviour of the fault and includes contribution charts and variable reconstruction.

Nonlinear Process Behaviour

- Most industrial processes exhibit nonlinear relationships between monitored variables.
- The relationship between the process variables may therefore only be approximated by a linear model in a small area around the origin.
- The “model plane” changes to “model surface”.



Linear PCA Model   Nonlinear PCA Model

How to decide whether to use a linear or nonlinear PCA model for condition monitoring?

Uwe Kruger

Development and Application of Nonlinear PCA

August 25., 2006

Virtual Engineering Centre

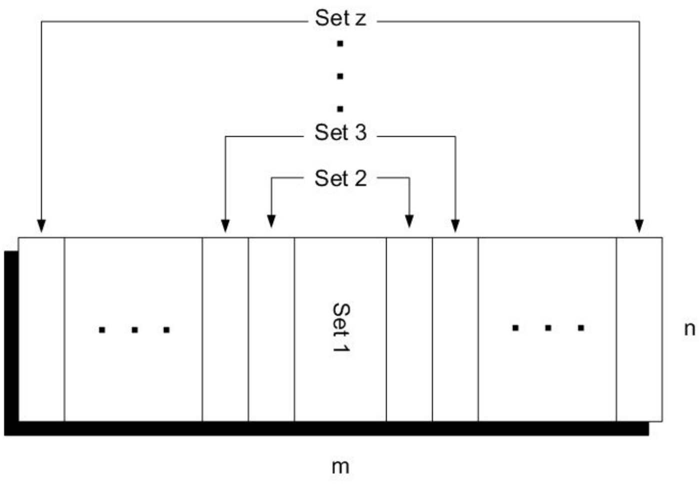
By Applying a Nonlinearity Test?

- A nonlinearity test has been developed for PCA models (Kruger *et al.*, 2004)
- Test relies on dividing data space into disjunct regions, the definition of accuracy bounds and is based on following five assumptions:
  1. ***The variables are mean centred and scaled to unit variance with respect to disjunct regions for which the accuracy bounds ought to be determined.***
  2. ***Each of the disjunct regions contains the same number of observations.***
  3. ***A PCA model is available for one of the disjunct regions for which the accuracy bounds for the sum of the discarded eigenvalues can be obtained.***
  4. ***PCA models are obtained for the remaining disjunct regions.***
  5. ***The same number of principal components is retained for each of the PCA models.***

Disjunct Regions

Dividing of data space can be accomplished using

1. *a priori* knowledge of the process, e.g. knowledge of existing data clusters;
2. a direct analysis of the recorded data by analysing the entire data set using PCA; and
3. an iterative approach if the above 2 do not reveal distinctive regions:
   (i) initiated by dividing data space into 2 regions first and then applying non linearity test
   (ii) increment the number of regions to 3, 4 and so on to assess the impact of the increasing number of regions upon test outcome.



Confidence Limits for Correlation Matrix

- The correlation matrix is now determined for each region wrt mean and variance of the individual region.
- Using standard calculation for determining confidence limits for non-diagonal elements in covariance matrices for a confidence of 95 or 99%:

$$\mathbf{S}_{XX} = \begin{bmatrix} 1 & s_{12_L} \le s_{12} \le s_{12_U} & \cdots & s_{1n_L} \le s_{1n} \le s_{1n_U} \\ s_{21_L} \le s_{21} \le s_{21_U} & 1 & \cdots & s_{2n_L} \le s_{2n} \le s_{2n_U} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1_L} \le s_{n1} \le s_{n1_U} & s_{n2_L} \le s_{n2} \le s_{n2_U} & \cdots & 1 \end{bmatrix}$$

$$\mathbf{S}_{XX_L} \le \mathbf{S}_{XX} \le \mathbf{S}_{XX_U}$$

where the indices U and L refer to upper and lower limit, respectively

Determination of Accuracy Bounds (1)

- The sum of variances of the PCA model residuals, σ, are equal to the sum of the discarded eigenvalues of the correlation matrix:

$$\sigma = \sum_{j=1}^{n} \sigma_j = \frac{1}{m^* - 1} \sum_{i=1}^{m^*} \sum_{j=1}^{n} e_{ij}^2 = \sum_{l=k+1}^{n} \lambda_l ,$$

where m* is the number of samples inside one of the disjunct regions

- Using the above equation, two optimisation functions can be constructed:

$$\sigma_{MAX} = \arg \max_{\Delta \mathbf{S}_{XX_{MAX}}} \sigma\left(\mathbf{S}_{XX} + \Delta \mathbf{S}_{XX_{MAX}}\right)$$

$$\sigma_{MIN} = \arg \min_{\Delta \mathbf{S}_{XX_{MIN}}} \sigma\left(\mathbf{S}_{XX} + \Delta \mathbf{S}_{XX_{MIN}}\right)$$

which is subject to the following constraints:

$$\mathbf{S}_{XX_L} \leq \mathbf{S}_{XX} + \Delta \mathbf{S}_{XX_{MAX}} \leq \mathbf{S}_{XX_U}$$

$$\mathbf{S}_{XX_L} \leq \mathbf{S}_{XX} + \Delta \mathbf{S}_{XX_{MIN}} \leq \mathbf{S}_{XX_U}$$

Determination of Accuracy Bounds (2)

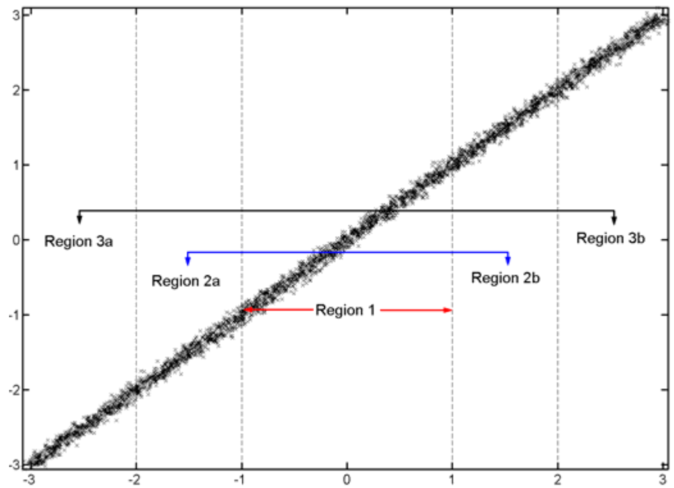
- The values of σ_{MAX} and σ_{MIN} are accuracy bounds and determine how much the sum of the error variances can vary within a linear PCA model based on the size of the recorded data set.
- If the number of samples goes to infinity, the upper and lower accuracy bound are identical.
- The nonlinearity test is therefore reduces to a check whether the sum of the residual variances of each region falls within these boundaries and this for accuracy bounds determined for each region, which characterises a linear PCA model.
- If at least one of these sums is outside the accuracy bounds, a nonlinear PCA model needs to be used.

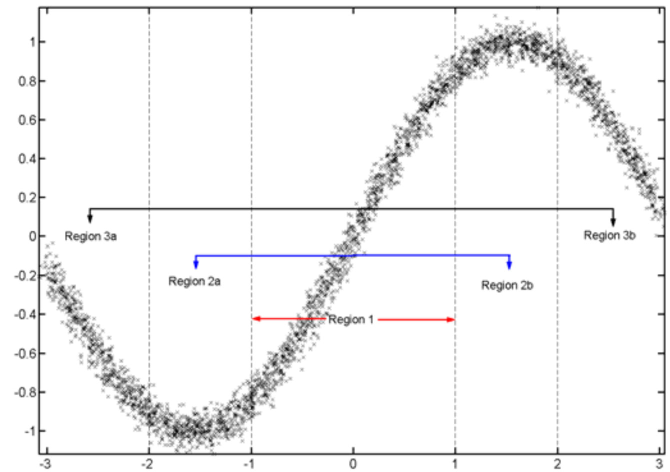
Simulation Example (1)

- Consider the following examples:

Example 1	Example 2
$x_1 = x + 0.05e_1$	$x_1 = x + 0.05e_1$
$x_2 = x + 0.05e_2$	$x_2 = \sin(x) + 0.05e_2$

where $x \in \mathcal{N}\{0,1\}$ and e_1 and e_2 are normally distributed i.i.d. sequences of zero mean and variance 1.



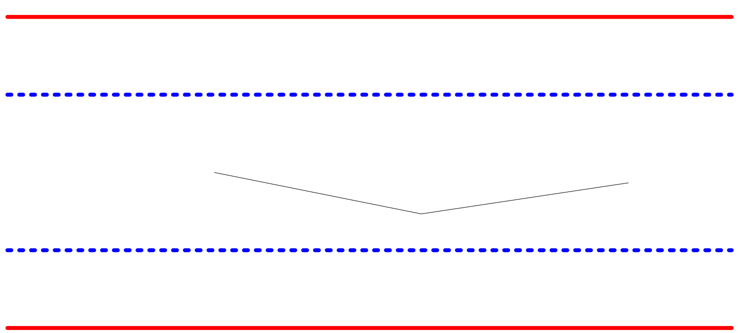


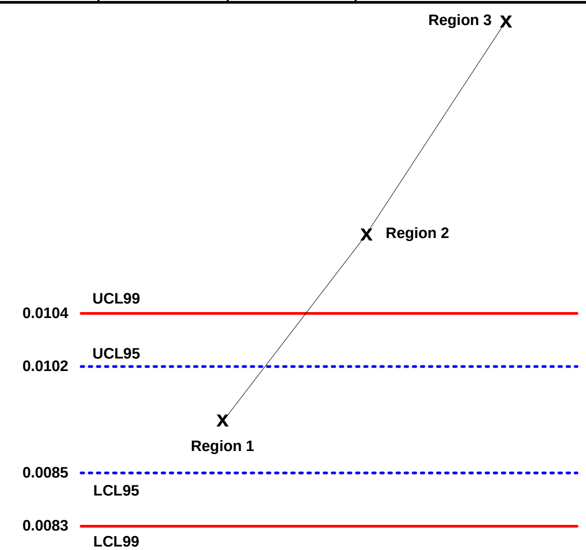
Simulation Example (1)

- Dividing the data space for each example into 3 disjunct regions and applying the nonlinearity test produced the following results (for only one region):

Accuracy bounds of linear process				Sum of discarded eigenvalue		
UCL99	UCL95	LCL95	LCL99	Region 1	Region 2	Region 3
0.0083	0.0081	0.0068	0.0066	0.0074	0.0071	0.0072
0.0090	0.0080	0.0070	0.0060	→ correlation coefficient procedure		

Accuracy bounds of nonlinear process				Sum of discarded eigenvalues		
UCL99	UCL95	LCL95	LCL99	Region 1	Region 2	Region 3
0.0104	0.0102	0.0085	0.0083	0.0093	0.0767	0.2990
0.0110	0.0100	0.0090	0.0080	→ correlation coefficient procedure		





Uwe Kruger
Development and Application of Nonlinear PCA
August 25., 2006

Virtual Engineering Centre

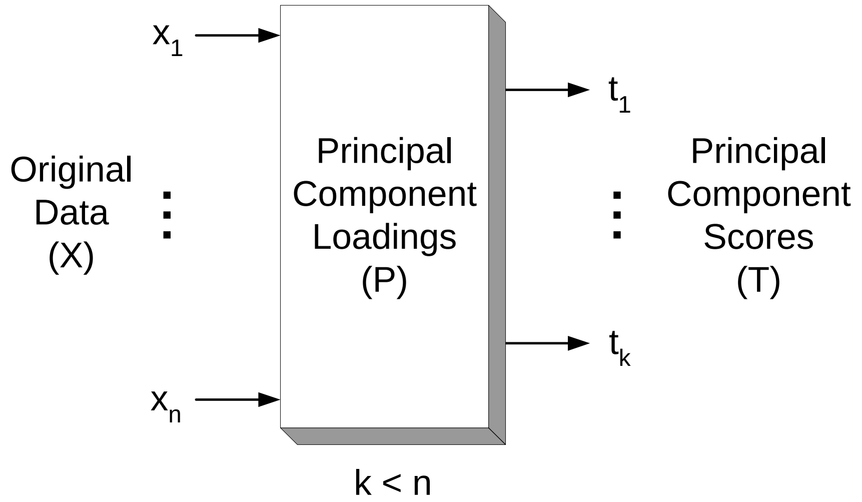
Existing Work on NLPCA

- **Two different approaches have been proposed in the literature:**
  1. **Principal Curves (Hastie and Stuetzle, 1989)**
  2. **Autoassociative Neural Networks (ANNs) (Kramer, 1991)**
- **With increasing dimension (number of variables), ANNs are simpler to use and hence used in subsequent work.**

The T2T ANN Network Structure

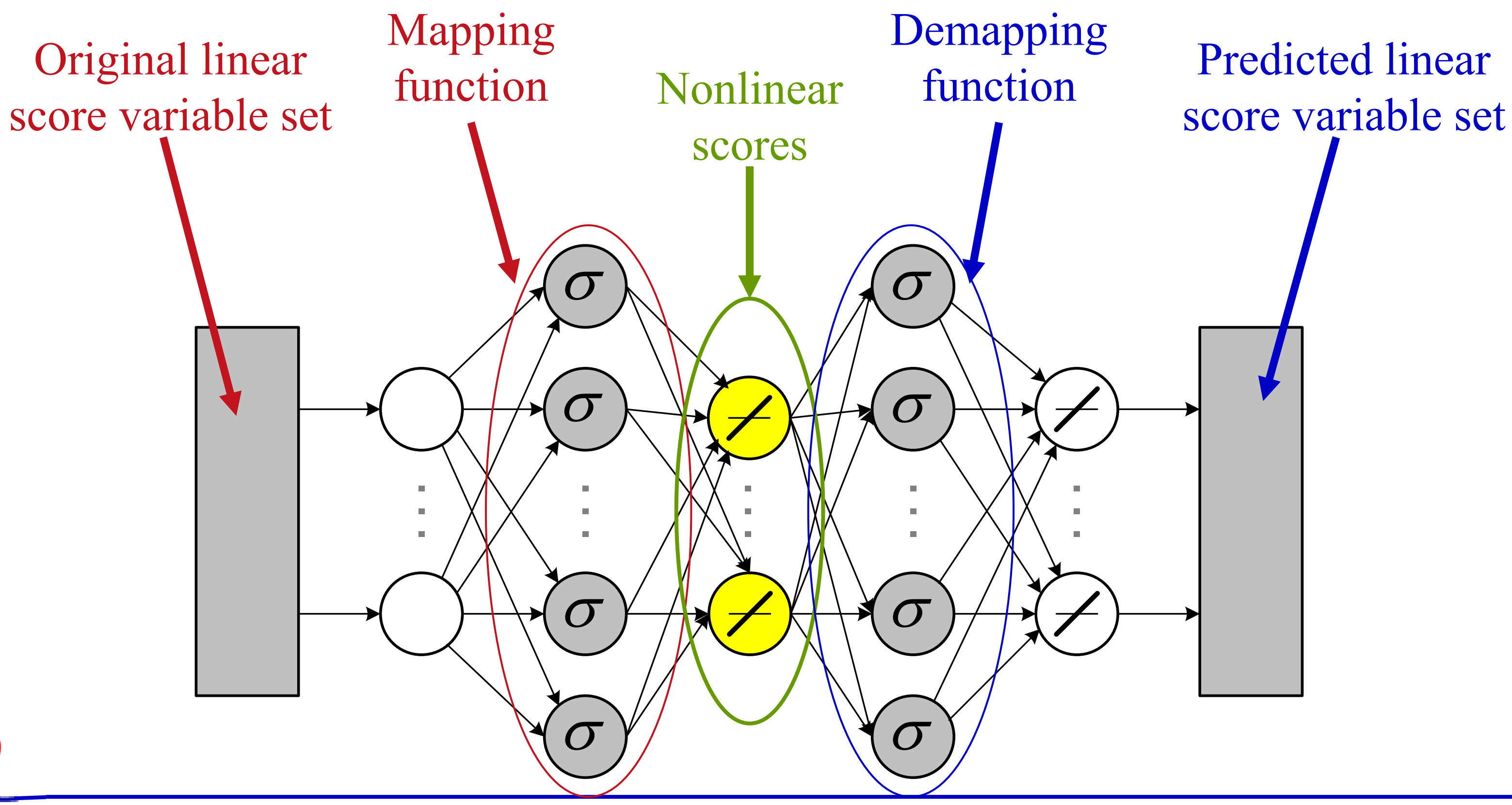
- Introduced by Antory et al. (2003, 2004, 2005)

- Relies on establishing a linear PCA model first in order to remove linear redundancy and/or insignificant process variation



The T2T ANN Structure

- Establish then an auto-associative neural network to predict the retained PCs using a reduced set of nonlinear score variables:



Fault Detection using T2T ANN (1)

- Residual variables

$$\mathbf{e} = \mathbf{x} - \mathbf{P}F\left(G\left(\mathbf{P}^T\mathbf{x}\right)\right) \quad Q = \mathbf{e}^T\mathbf{e}$$

- Explaining how well NLPCA model fits recorded data
- Confidence limit can be obtained as discussed in Nomikos and MacGregor (1995)

- Score Variables

$$\mathbf{t}_{NL}^T = \begin{pmatrix} t_{NL1} & t_{NL2} & \cdots & t_{NLz} \end{pmatrix}$$

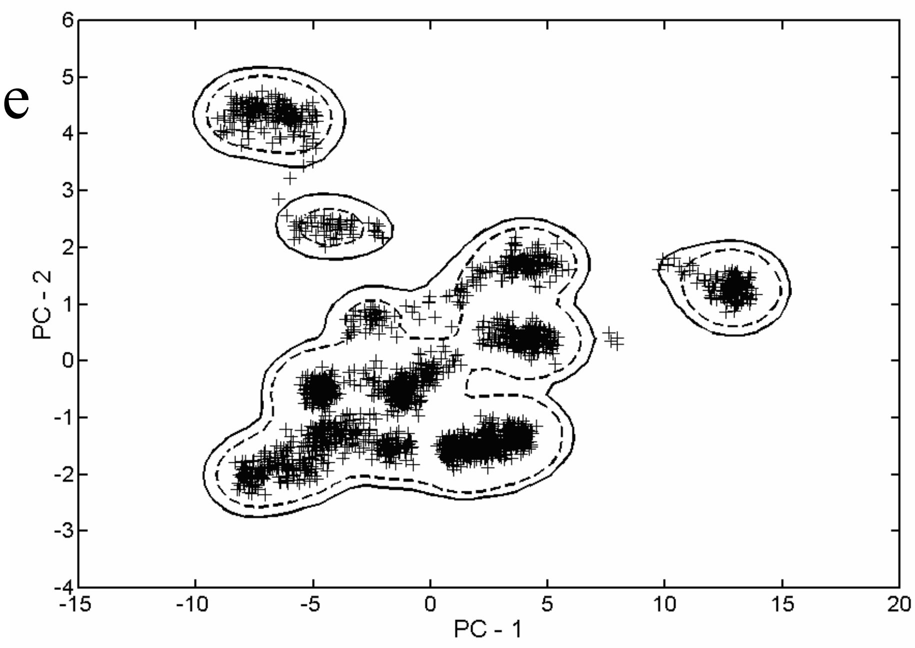
- May <u>not</u> be normally distributed nor statistically independent

⇩

- Process variation is monitored using scatter diagrams and
- confidence regions are built using kernel density estimation

Fault Detection using T2T ANN (2)

- Whilst the univariate residual statistic is relatively straightforward in approach, the kernel density estimation relies on estimating the probability density function for each combination of two nonlinear score variables.
- The determination of the confidence regions, see example to the right, is time consuming and produces (z+1)z/2 plots.
- Hence, the use of scatter diagrams is cumbersome and difficult to implement in practice.



Fault Detection using T2T ANN (3)

- To monitor the score variables, a conceptually better approach is to integrate the statistical local approach into the nonlinear PCA monitoring framework (Wang *et al.*, 2006a,b).

- This relies on the definition of *improved residuals* $\zeta_0(\cdot)$:

$$\zeta_0(\boldsymbol{\theta}, j) \square \frac{1}{\sqrt{k}} \sum_{j=1}^{k} \boldsymbol{\theta}\left(\boldsymbol{\rho}_0, \mathbf{x}(j)\right)$$

where *j* and *k* are sample indeces, ρ_0 is a parameter vector, $\mathbf{x}(\cdot)$ is a data vector and $\theta(\cdot)$ is a vector of *primary residuals* that has to follow these assumptions:

$$E\left\{\boldsymbol{\theta}\left(\boldsymbol{\rho}, \mathbf{x}(j)\right)\right\} = 0 \quad \text{if } \boldsymbol{\rho} = \boldsymbol{\rho}_0$$

$$E\left\{\boldsymbol{\theta}\left(\boldsymbol{\rho}, \mathbf{x}(j)\right)\right\} \neq 0 \quad \text{if } \boldsymbol{\rho} \in \boldsymbol{\omega}\left(\boldsymbol{\rho}_0\right), \boldsymbol{\rho}_0 \notin \boldsymbol{\omega}\left(\boldsymbol{\rho}_0\right)$$

$$\boldsymbol{\theta}\left(\boldsymbol{\rho}_0, \mathbf{x}(j)\right) \text{ is differentiable in } \boldsymbol{\rho}$$

$$\boldsymbol{\theta}\left(\boldsymbol{\rho}, \mathbf{x}(j)\right) \text{ exists in the vicinity of } \boldsymbol{\rho}_0, \text{ i.e. } \boldsymbol{\rho} \in \boldsymbol{\omega}\left(\boldsymbol{\rho}_0\right)$$

Fault Detection using T2T ANN (4)

- The parameter vector ρ_0 is associated with normal operating condition, while $\rho \in \omega(\rho_0)$ corresponds to a faulty operating condition.
- Furthermore, since ρ lies in the vicinity of ρ_0, ρ can be written as:

$$\boldsymbol{\rho} = \boldsymbol{\rho}_0 + \frac{\Delta\boldsymbol{\rho}}{\sqrt{k}}$$

with $\Delta\mathbf{p} \neq \mathbf{0}$ being fixed but otherwise unknown.

- The improved residuals follow asymptotically a multinormal distribution with zero mean and covariance

$$\mathbf{S}_{\zeta\zeta}\left(\boldsymbol{\rho}_0\right) = \lim_{k\to\infty}\frac{1}{k}\sum_{i=1}^{k}\sum_{j=1}^{k}\boldsymbol{\theta}\left(\boldsymbol{\rho}_0,\mathbf{x}(i)\right)\boldsymbol{\theta}^T\left(\boldsymbol{\rho}_0,\mathbf{x}(j)\right)$$

Fault Detection using T2T ANN (5)

- For process monitoring, the following hypothesis, H_0, and alternative hypothesis, H_1, can be established using the improved residuals:

$$H_0 \quad : \quad \boldsymbol{\rho} = \boldsymbol{\rho}_0 \qquad H_1 \quad : \quad \boldsymbol{\rho} = \boldsymbol{\rho}_0 + \frac{\Delta\boldsymbol{\rho}}{\sqrt{k}}$$

- The hypothesis H_0 has the following distribution:

$$\zeta_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{S}_{\zeta\zeta})$$

- whereas the alternative hypothesis H_1 possesses the distribution:

$$\zeta_1 \sim \mathcal{N}(\mathbf{M}(\rho_0)\Delta\rho, \mathbf{S}_{\zeta\zeta})$$

with:

$$\mathbf{M}\left(\boldsymbol{\rho}_0\right) \square -\frac{\partial \boldsymbol{\theta}\left(\boldsymbol{\rho}, \mathbf{x}\left(j\right)\right)}{\partial \boldsymbol{\rho}}\Bigg|_{\boldsymbol{\rho}=\boldsymbol{\rho}_0}$$

Definition of Primary Residuals (1)

- For an ANN, the calculation of the *j*th score variables is as follows:

$$t_j = \sum_{p=1}^{N} w_{jp}^{(2)} \psi\left(\sum_{q=1}^{n} w_{pq}^{(1)} x_q + b_p^{(1)} \right) + \sum_{q=1}^{n} w_{jq}^{(5)} x_q + b_j^{(2)}$$

- and the prediction of the *j*th process variables is given by:

$$x_j' = \sum_{p=1}^{N} w_{jp}^{(4)} \psi\left(\sum_{q=1}^{z} w_{pq}^{(3)} t_q + b_p^{(3)} \right) + \sum_{q=1}^{z} w_{jq}^{(6)} t_q + b_j^{(4)}$$

where $\psi(\cdot)$ is the hyperbolic tan function.

- As parameter vector ρ, the parameters of the ANN can be used, i.e..:

$$\boldsymbol{\rho}^T = \begin{pmatrix} w_{11}^{(1)} & \cdots & w_{zn}^{(6)} & b_1^{(1)} & \cdots & b_z^{(4)} \end{pmatrix}$$

Definition of Primary Residuals (2)

- The network weights, stored in ρ, are determined by minimising the least squares cost function:

$$J = \sum_{k=1}^{m} \left\| \mathbf{x}(k) - \mathbf{x}'(k) \right\|_2^2 = J(\boldsymbol{\rho})$$

- The solution of the above cost function is given by:

$$\left. \frac{\partial \sum_{k=1}^{m} \left\| \mathbf{x}(k) - \mathbf{x}'(k) \right\|}{\partial \boldsymbol{\rho}} \right|_{\boldsymbol{\rho}=\boldsymbol{\rho}_0} \left. \frac{\partial \sum_{k=1}^{m} J_k}{\partial \boldsymbol{\rho}} \right|_{\boldsymbol{\rho}=\boldsymbol{\rho}_0} = \mathbf{0}$$

Definition of Primary and Improved Residuals

- Which suggests to define the following primary residuals:

$$\boldsymbol{\theta}\left(\boldsymbol{\rho}_0, \mathbf{x}(k)\right) = \nabla\left(J_k\right)\Big|_{\boldsymbol{\rho}=\boldsymbol{\rho}_0}$$

- Now, the improved residuals can be established:

$$\zeta_0 = \frac{1}{\sqrt{k}} \sum_{j=1}^{k} \nabla\left(J_k\right)\Big|_{\boldsymbol{\rho}=\boldsymbol{\rho}_0}$$

- Finally, a univariate statistic can be established to monitor changes in the mean of the parameter vector $\boldsymbol{\rho}_0$

$$T^2(\zeta, k) = \zeta^T\left(\nabla\left(J_k\right), k\right) S_{\zeta\zeta}^{-1} \zeta\left(\nabla\left(J_k\right), k\right)$$

Definition of Univariate Monitoring Statistic

- The following problem emerges with an increasing number of recorded samples:
  The sensitivity of detecting small changes in the mean of $\theta(\cdot)$ reduces with an increasing number of samples (Wang *et al*., 2005).

- This, however, can be overcome by the use of a moving window technique, which implies that the improved residuals are computed as follows:

$$\zeta_0 = \frac{1}{\sqrt{k_0}} \sum_{j=k-k_0+1}^{k} \nabla\left(J_k\right)\Big|_{\boldsymbol{\rho}=\boldsymbol{\rho}_0}$$

  with a sufficiently large k_0.

Motivation for Fault Diagnosis Problem

- An investigation by Lieftucht *et al.* (2006a) showed that traditional techniques for fault isolation and identification suffer from collinearity (contribution charts) or restrictions concerning the number of variables to be reconstructed (variable reconstruction).
- In a further study, Lieftucht et al. (2006b) outlined that a regression based reconstruction approach overcomes these deficiencies.
- Wang *et al.* (2006b) proposed an extension to the work in Lieftucht et al. (2006b) to nonlinear PCA.
- This extension relies on the following assumptions, discussed on the next slide.

Assumptions for Fault Diagnosis Charts

- The fault signatures are deterministic, that is the signature for a particular process variable is a function of the sample index.
- The fault is superimposed onto the process variables.
- The recorded process variables representing normal process operation are stochastic in nature.

- This can be mathematically described as follows:

$$\mathbf{x}_f = \mathbf{x} + \mathrm{f}$$

where $\mathbf{x}_f$, $\mathbf{x}$ and $\mathbf{f}$ are vectors storing the recorded variables containing the recorded variables including fault information, normal stochastic variation without fault information and the fault signature, respectively.

- Under the assumption that the fault signature takes the form of a step of magnitude **f**, the change in mean of the model residuals $\mathbf{e}_f$ and the primary residuals $\Delta(\nabla J)$ are given by:

$$\mathbf{e}_f = \mathbf{x} + \mathbf{f} - \mathbf{H}\left(\mathbf{G}\left(\mathbf{x}+\mathbf{f}\right)\right)$$

$$\mathbf{e} = \mathbf{x} - \mathbf{H}\left(\mathbf{G}\left(\mathbf{x}\right)\right)$$

$$\Delta\left(\nabla J\right) = \left.\frac{\partial \left\|\mathbf{e}_f\right\|_2^2}{\partial \boldsymbol{\rho}}\right|_{\boldsymbol{\rho}=\boldsymbol{\rho}_0} - \left.\frac{\partial \left\|\mathbf{e}\right\|_2^2}{\partial \boldsymbol{\rho}}\right|_{\boldsymbol{\rho}=\boldsymbol{\rho}_0}$$

  where $\mathbf{H}(\cdot)$ and $\mathbf{G}(\cdot)$ are the mapping and demapping function of the ANN.

- Under the above assumption, the following holds true:

$$E\left\{\mathbf{e}\right\} = \mathbf{0} \quad E\left\{\mathbf{e}_f\right\} = \boldsymbol{\varepsilon} \quad E\left\{\Delta\left(\nabla J\right)\right\} = \Delta\mathbf{J}$$

- It must generally be assumed that $\mathbf{f} \neq \boldsymbol{\varepsilon}$ (Lieftucht et al., 2006a). Moreover, there is a close link between $\boldsymbol{\varepsilon}$ and $\Delta\mathbf{J}$.

- Assuming a data set of m_f samples that describe a step type fault, the statistical confidence limit for the Q statistic, $Q_{\alpha f}$ can be computed as discussed in Nomikos and MacGregor (1995):

$$Q_{\alpha_f} \approx \eta_f \chi^2\left(\vartheta_f\right) \quad \eta_f = \frac{\sigma_f^2}{2\mu_f} \quad \vartheta_f = \frac{2\mu_f^2}{\sigma_f^2}$$

where μ_f and σ_f are the estimated mean and variance of the sequence $Q_{\alpha_f} = \mathbf{e}_f^T \mathbf{e}_f$.

- Without the fault, the confidence limit of the Q statistic, Q_α, is given by $Q_\alpha \approx \eta \chi^2\left(\vartheta\right)$ where μ and σ are the estimated mean and variance of the sequence $Q = \mathbf{e}^T\mathbf{e}$.
- Now, defining $\hat{\mathbf{x}} = \mathbf{x}_f - \hat{\mathbf{f}}$ gives rise to the following optimisation problem

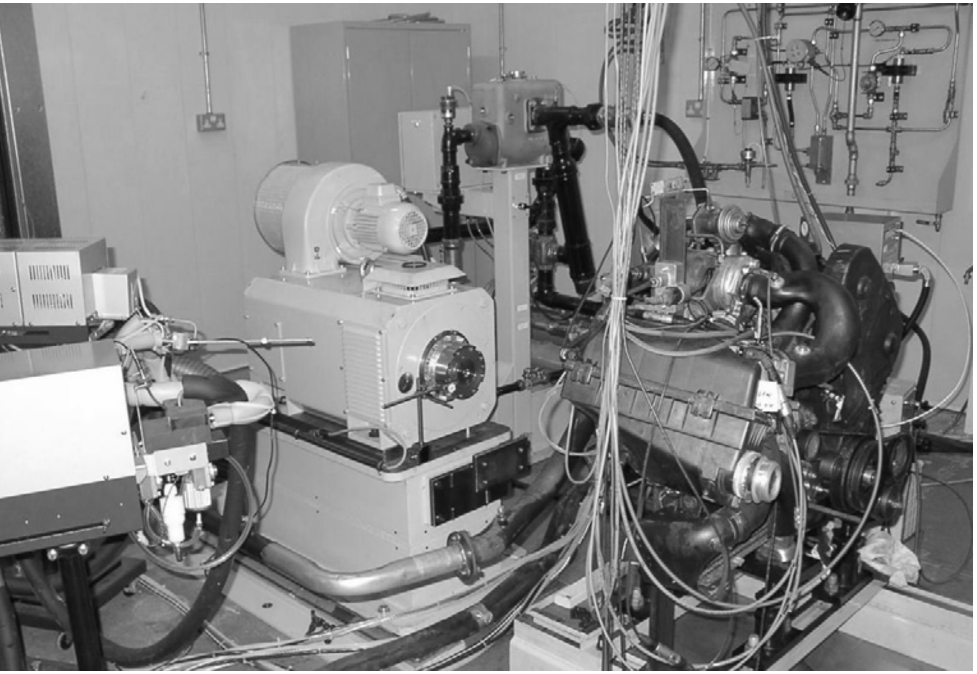
$$\Delta Q_\alpha = Q_{\alpha_f} - Q_\alpha = \arg\min_{\hat{\mathbf{f}}} \left(\frac{\sigma^2\left(\hat{\mathbf{x}}\right)}{2\mu\left(\hat{\mathbf{x}}\right)} \chi^2\left(\frac{2\mu^2\left(\hat{\mathbf{x}}\right)}{\sigma^2\left(\hat{\mathbf{x}}\right)} \right) - Q_\alpha \right)$$

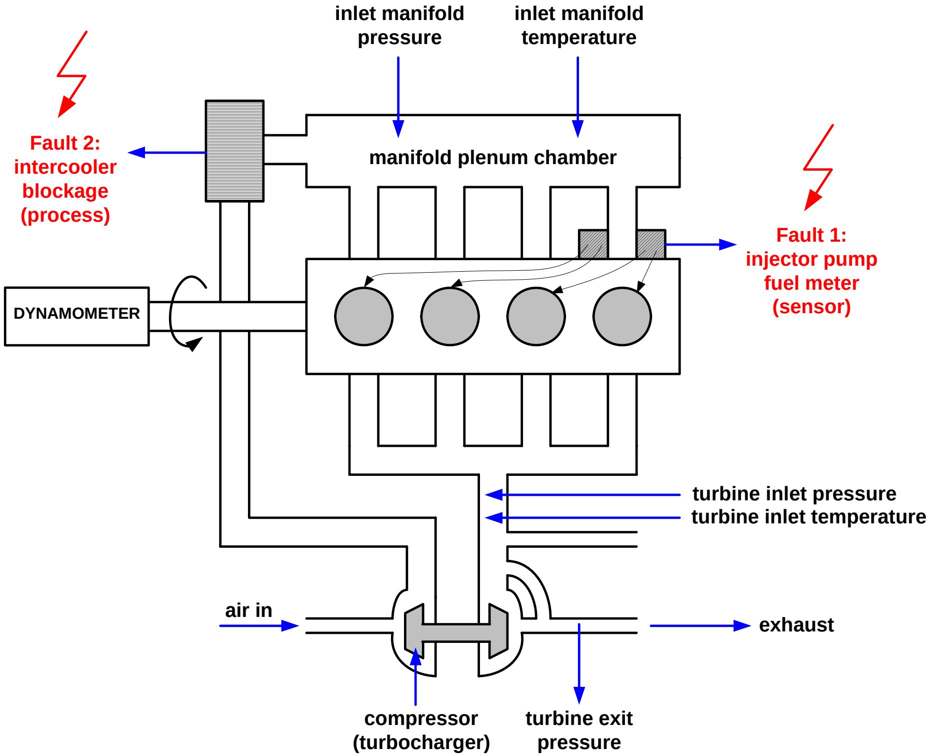
Summary of NLPCA Condition Monitoring

- The following steps are required to establish a NLPCA condition monitoring scheme:
  - Determine a NLPCA model using an ANN structure
  - Use the model residuals of the ANN model to construct a Q statistic and obtain the confidence limit Q_α
  - Define the primary and improved residuals and construct a Hotelling's T^2 statistic
  - Utilise both univariate statistics, i.e. T^2 and Q statistics, to assess the current state of the engine operation.
  - If the engine variables produce a statistically significant number of violations wrt the confidence limits for the T^2 and Q statistics, apply the fault diagnosis scheme to determine **f** if it is a step-type fault and a series of **f**-vectors if it is a general deterministic fault condition , i.e. divide the recorded sequence of fault data into a number of sections and approximate the fault signature by a series of step-type faults.
  - Plot the **f**-vector(s) versus variable index in a bar chart to produce the fault diagnosis chart.

Application Studies (Diesel Engine Used)

- Analysis of data from a Volkswagen 1.9L TDI diesel engine.
- Process variables exhibit non-linear relationships.
- Two recorded fault conditions were diagnosed.





Uwe Kruger
Development and Application of Nonlinear PCA
August 25., 2006

Application Studies (PCA and NLPCA Models)

Variables analysed

No	Engine Variable	Unit	Note
1	Fuel Flow	kg/h	output
2	Air Flow	kg/h	
3	Inlet Manifold Pressure	Bar	
4	Inlet Manifold Temperature	˚C	
5	Turbine Inlet Pressure	Bar	
6	Turbine Inlet Temperature	˚C	

RPM	1500	2500	3500	4500
Pedal Position	30%	49%	57%	62%
	40%	59%	64%	65%
	54%	74%	74%	76%
	62%	78%	80%	83%
	100%	100%	100%	100%

Modelling results

Principal Component	Variance Captured (%)	Variance Total (%)
1	79.5998	79.5998
2	16.4492	96.0490
3	2.4169	98.4659
4	1.0745	99.5404
5	0.4010	99.9414
6	0.0586	100.000

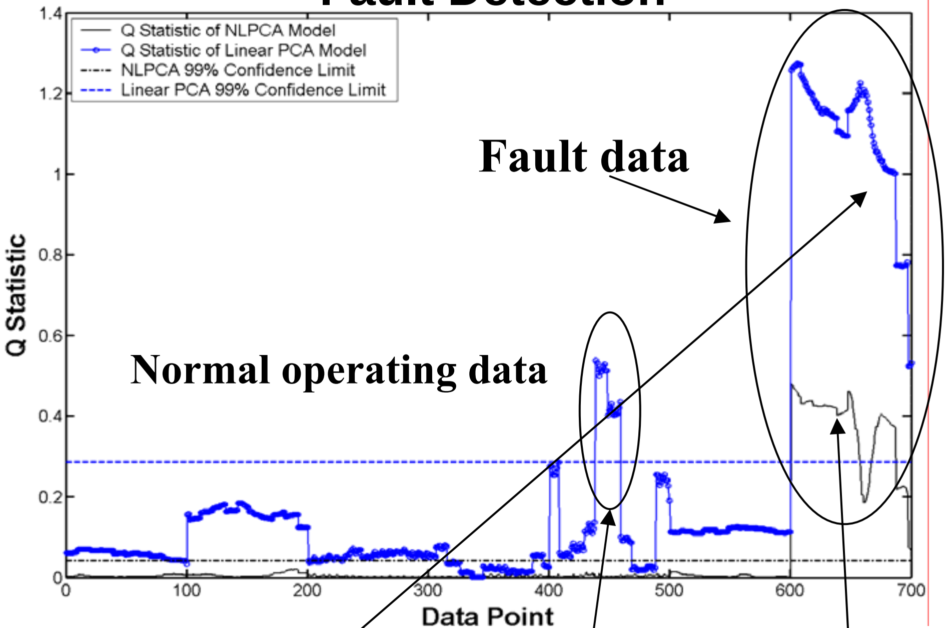
Number of Bottleneck Nodes	Variance Captured (%)	Note
1	97.8160	Important variation
2	99.4212	
3	99.8336	
4	99.8725	Negligible
5	99.9401	
6	99.9414	

Application Studies (Using Scatter Diagrams)

- We first investigated two fault conditions, (i) a bias on the fuel flow sensor reading and (ii) an intercooler blockage using scatter diagrams and the Q statistic for fault detection and contribution charts and variable reconstruction for fault diagnosis.
- Comparing the PCA model with an NLPCA model using the same number of retained Principal Components (3), the PCA model produces false alarms, which results from an incorrect identification of the interrelationships between the recorded engine variables.
- This work and shows that both fault conditions can be detected.
- However, only the sensor fault can be correctly diagnosed, whilst the process fault is incorrectly diagnosed.

- **First fault condition : fuel flow sensor drift**

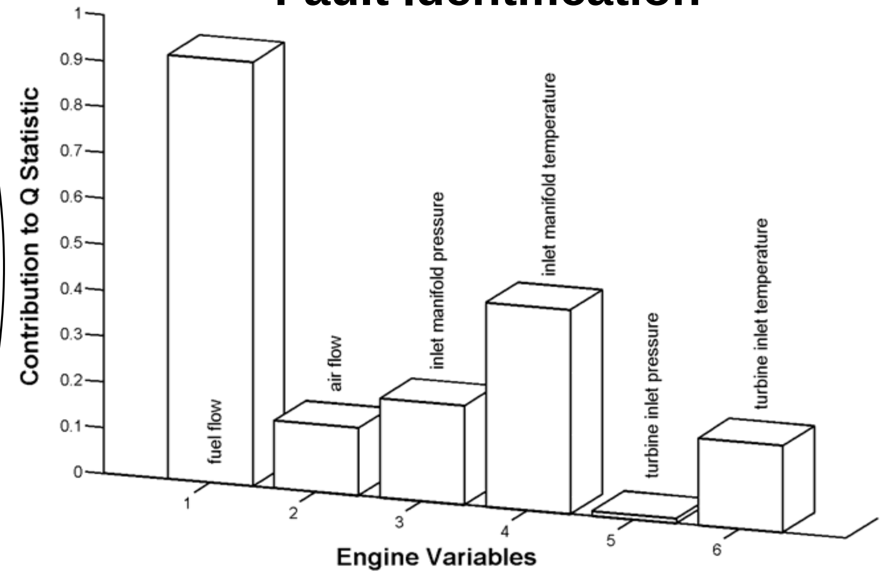
Fault Detection



Linear PCA Model **Nonlinear PCA Model**

False alarm!

Fault Identification

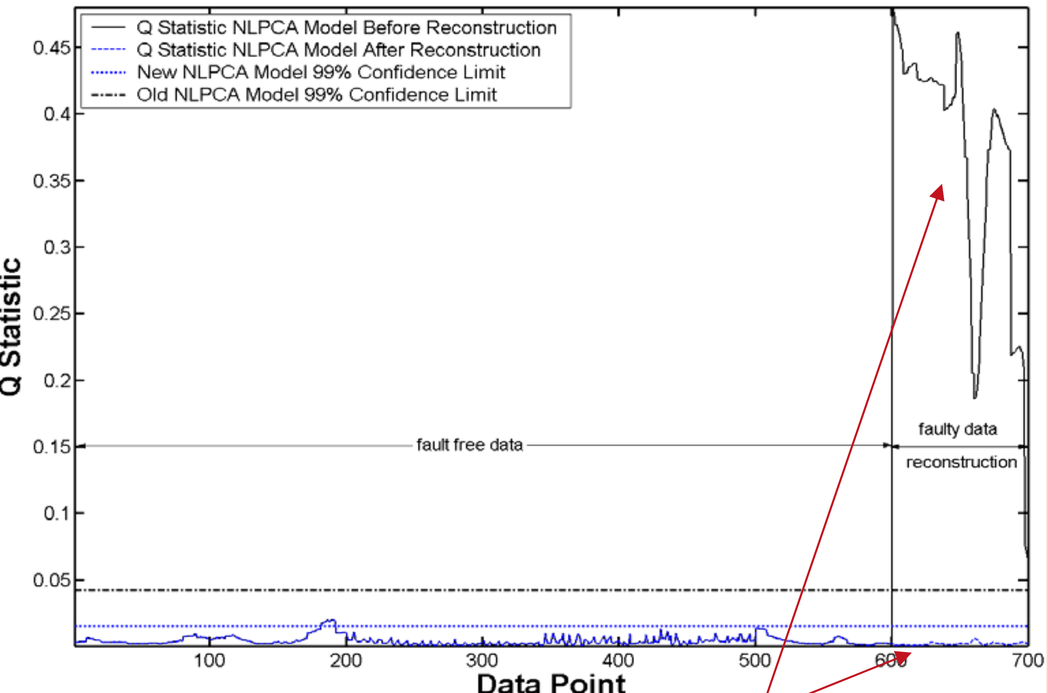


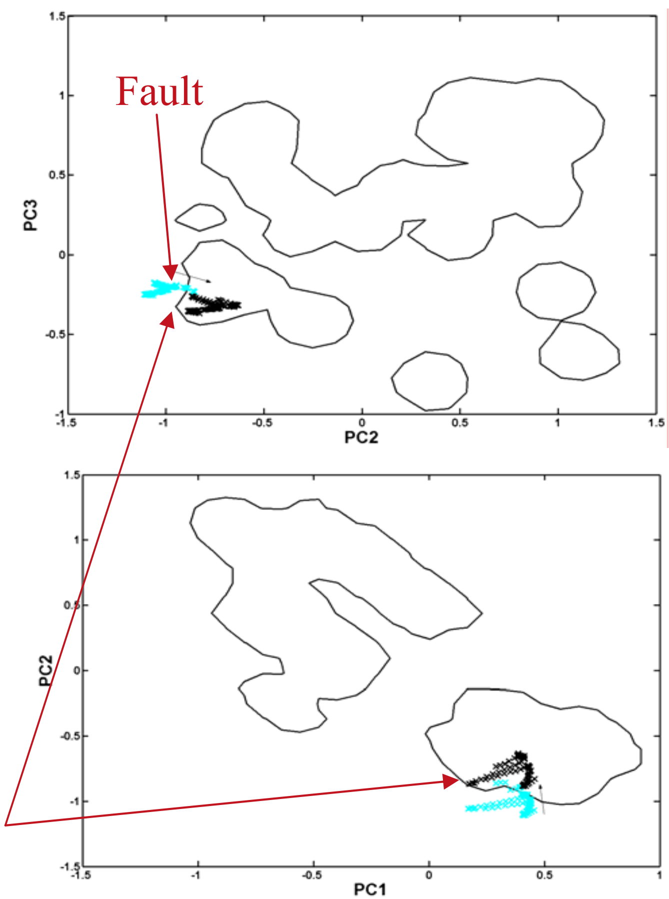
PCA model produced false alarms!

Virtual Engineering Centre

- **First fault condition : fuel flow sensor drift**

Fault Isolation

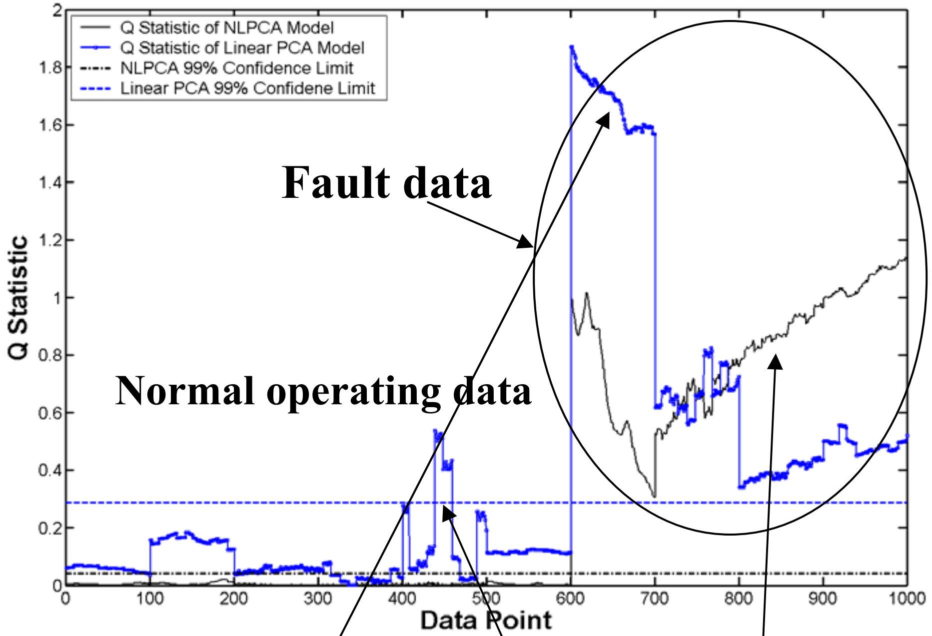




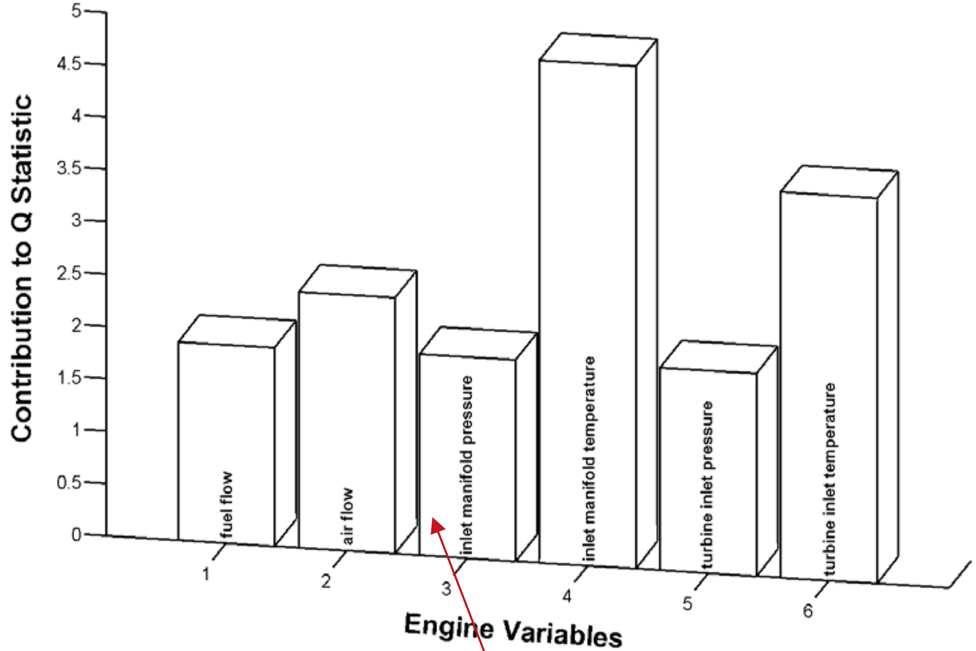
Reconstructing fuel flow and inlet manifold temperature worked for simple sensor fault

- **Second fault condition : Intercooler blockage**

Fault Detection



Fault Identification



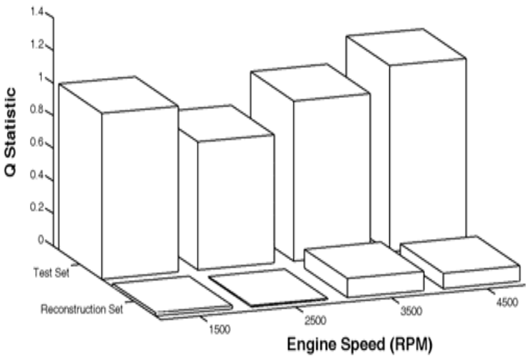
Linear PCA Model **Nonlinear PCA Model**

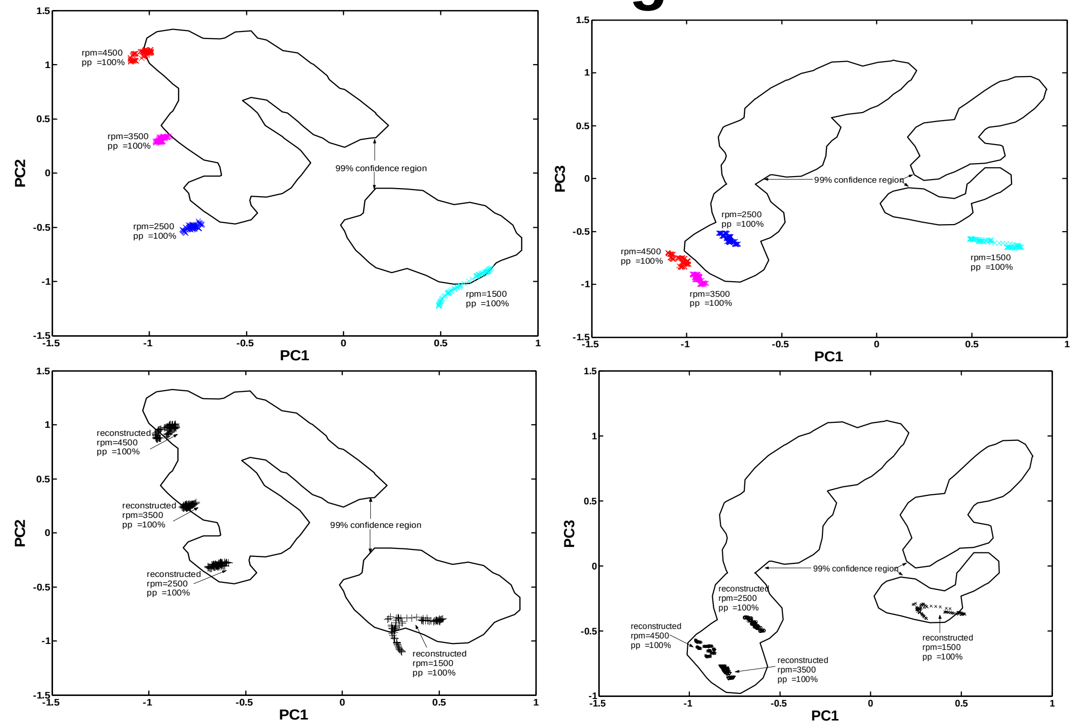
False alarms again!

Knowing the impact of this fault *a priori* the correct variable contribution should be different!

- **Second fault condition : Intercooler blockage**

Fault Isolation





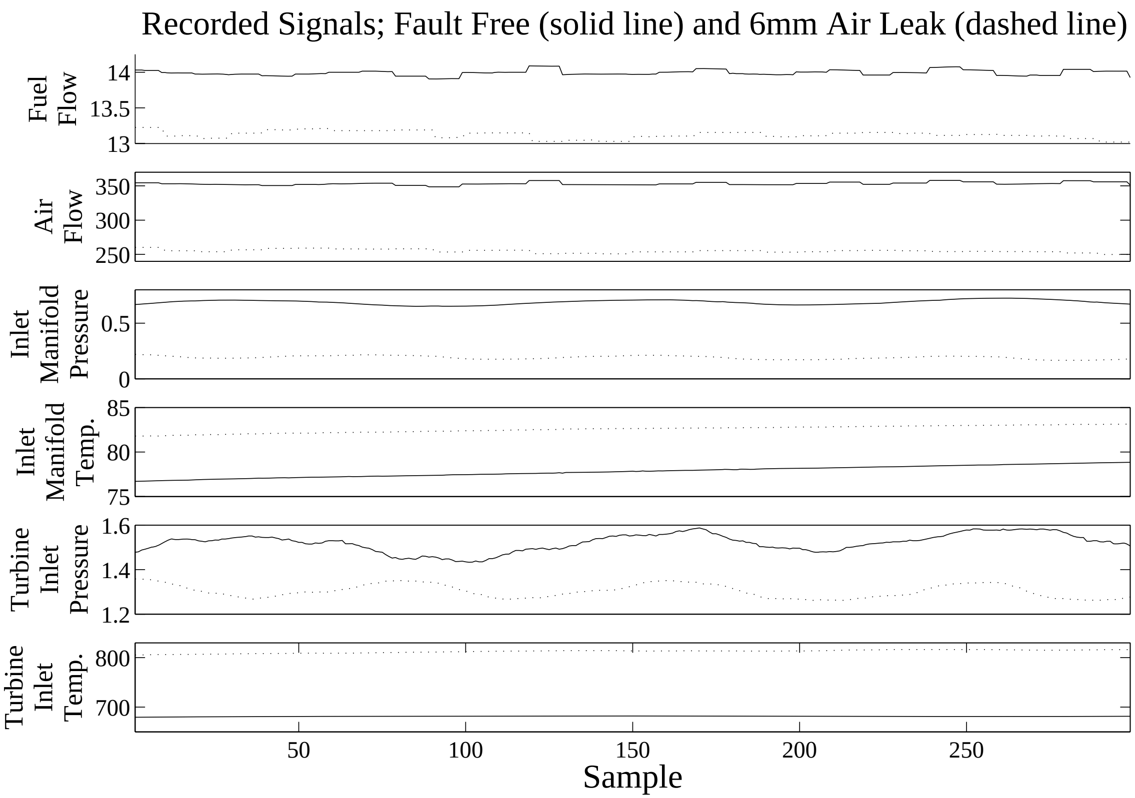
Although the residual variance reduces significantly (plot above), the scatter points (plots to the right) cannot be moved inside the statistical confidence regions. This implies that a better and more general fault diagnosis scheme needs to be developed. In addition, the use of scatter diagrams is cumbersome.

Uwe Kruger

Application Studies (Using the Local Approach)

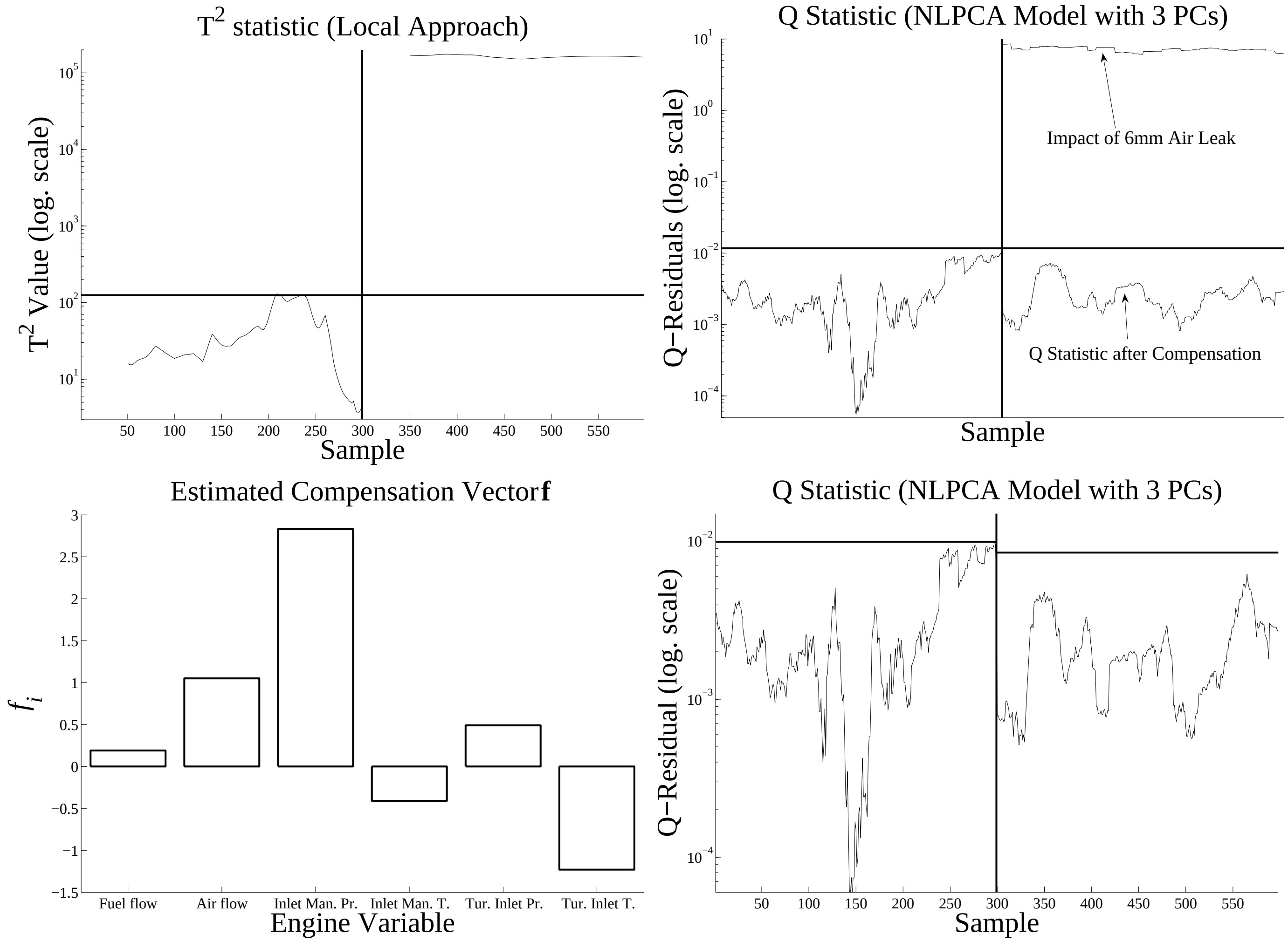
- Here, we investigate the use of the Local Approach to detect and diagnose an air leak in the plenum manifold chamber.
- The detection relies on the use of the Q statistic as well as the Hotelling's T2 statistic constructed from the improved residuals using a window length k_0 = 50.
- For fault diagnosis, the fault diagnosis chart is utilised.
- This process fault could be detected and correctly diagnosed by applying the conceptually more complex local approach.

- The sequences of the 6 recorded variables are shown in the figure to the right.
- The direct comparison between normal operating data for 4500rpm / 100% pedal position and data that describe a 6mm air leak indicate that the air flow, the inlet manifold pressure and the turbine inlet temperature are most significantly affected.



Application Studies (Analysis of Air Leak 1)

The figures to the right show that this process fault could be detected and correctly diagnosed. The air flow, inlet manifold pressure and turbine inlet temperature are identified as the most significantly contributors to this event. Subtracting this fault vector brought the Q statistic below the 99% confidence limit



Uwe Kruger

Development and Application of Nonlinear PCA

August 25., 2006

Virtual Engineering Centre

Conclusions (1)

- Whilst the state-of-the-art in engine fault diagnosis is firmly based on the application of mechanistic first-principal (white-box) models, our research investigates the usefulness of an intelligent analysis of recorded data.
- To analyse steady-state data, as part of a routine annual check (MOT), has been conducted using nonlinear principal component analysis as the core component technology
- The detection of fault conditions relies (i) on the use of scatter diagrams and the Q statistic (earlier work) and (ii) the incorporation of the statistical local approach to construct a Hotelling's T^2 statistic (current work).
- Although the latter approach is conceptually more complex, it permits an easier monitoring, as a large number of scatter diagrams may need to be examined.

Conclusions (2)

- Fault diagnosis, to determine the potential root cause responsible for anomalous engine behaviour, is based on (i) the use of contribution charts and variable reconstruction (earlier work) and (ii) the development of a fault diagnosis chart (current work).
- Problems with contribution charts and variables reconstruction for linear PCA are reported in Lieftucht *et al.* (2006a) and similar problems have been experienced in our earlier work if complex process faults were analysed.
- The fault diagnosis chart performed best in our study if the assumptions, on which this chart is designed upon, are met.
- A total of 3 engine faults were successfully detected and diagnosed and outline the potential for our data driven approach to form part of the engine management system.

Future Work

- This work conducted thus far merely relies on the analysis of steady-state data and is applicable to routine MOT-based checks.
- The aim of future work is the analysis of dynamic data for on-line fault detection and diagnosis, which will be the requirement of future On-Board-Diagnostic (OBD) legislation introduced in the near future.
- This, however, requires more efficient algorithms to identify models that can be benchmarked against the current engine performance
- Since, the training of ANNs is time-consuming, alternative approaches such as principal manifolds to construct NLPCA and dynamic NLPCA models will be investigated.
- Besides engine faults, our research will also focus on catalyst degradation which is significantly contributing to increased emissions. This, again, will be an important point for future OBD legislations.

Thank you

Dr Uwe Kruger
e-mail: uwe.kruger@ee.qub.ac.uk

School of Electronics, Electrical Engineering
& Computer Science
Queen's University Belfast
Ashby Building, Stranmilis Road
Belfast BT9 5AH
Northern Ireland, United Kingdom

References

- Antory D, Kruger U, Irwin GW, McCullough G. Fault diagnosis in internal combustion engines using principal component analysis. Proceedings of the 4th International Conference on Control and Diagnostics in Automotive Applications (CDAUTO 03), Sestri Levante (Genova), Italy, June 18 – 20., paper no: 03A2029, 2003.
- Antory D, Kruger U, McCullough G, Irwin GW. Detecting air leaks in an engine intake manifold using nonlinear pca. Proceedings of the IFAC Symposium on Advances in Automotive Control, Salerno, Italy, April 19 – 23., pp. 404-409, 2004.
- Antory D, Kruger U, Irwin GW, McCullough G. Fault diagnosis in internal combustion engines using nonlinear multivariate statistics. *Proceedings of the Institution of Mechanical Engineers, Part I Journal of Systems and Control Engineering*. 219(4): 243-258, 2005.
- Hastie T, Stuetzle W. Principal curves. *Journal of the American Statistical Association*. 84(406): 502-517, 1989.

References

- Jackson JE. Quality control methods for two related variables, *Industrial Quality Control*, XII(7): 2-6, 1956.
- Jackson JE, Morris RH. An application of multivariate quality control to photographic processing, *Journal of the American Statistical Association*, LII: 186-199, 1957.
- Jackson JE. Quality control methods for several related variables. *Technometrics*. 1(4): 359-377, 1959.
- Jackson JE. Principal components and factor analysis: part I-principal components. *Journal of Quality Technology*. 12(4), 201-213, 1980.
- Kourti T, MacGregor JF. Process analysis, monitoring and diagnosis using multivariate projection methods. *Chemometrics & Intelligent Laboratory Systems*. 28(1): 3-21, 1995.
- Kramer MA. Nonlinear principal component analysis using auto-associative neural networks. *AIChE Journal*, 37(2): 233-243, 1991.
- Kresta JV, MacGregor JF, Marlin TE. Multivariate Statistical Monitoring of Process Performance. AIChE Annual Meeting, San Francisco CA, 1989.

References

- Lieftucht D, Kruger U, Irwin GW. Improved reliability in diagnosing faults using multivariate statistics. *Computers & Chemical Engineering*. 30: 901-912, 2006a.
- Lieftucht D, Kruger U, Irwin GW. Regression based variable reconstruction in multivariate systems. Proceedings of the 6th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes, Beijing, P.R. China, August 30. – September 1., 2006b.
- MacGregor JF, Marlin TE, Kresta JV, Skagerberg B. Multivariate statistical methods in process analysis and control. 4th International Conference on Chemical Process Control, pp. 79-99, 1991.
- MacGregor JF, Kourti T. Statistical Process Control of Multivariate Processes. *Control Engineering Practice*. 3(3): 403-413, 1995.
- Nomikos P, MacGregor JF. Monitoring of batch processes using multiway principal component analysis. *Technometrics*. 40: 1361-1375, 1995.

References

- Pearson K. On lines and planes of closest fit to systems of points in space. *Phil. Mag., Series B*. 2(11): 559-572, 1901.
- Wang X, Kruger U, Irwin GW. Process monitoring approach using moving window pca. *Industrial & Engineering Chemistry Research*, 44(15): 5691-5702, 2005.
- Wang X, Kruger U, Irwin GW, McCullough G, McDowell N. Nonlinear statistical process control for fault detection and diagnosis in automotive engines, submitted to IEEE Transactions on Control Systems Technology, 2006a.
- Wang X, Kruger U, Irwin GW, McDowell N, McCullough G. Nonlinear pca for process monitoring using the local approach, Proceedings of the 6th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes, Beijing, P.R. China, August 30. – September 1., 2006b.
- Wise BM, Gallagher NB. The process chemometrics approach to process monitoring and fault detection. *Journal of Process Control*. 6(6): 329-348, 1996.

References

- Wold H. Estimation of principal components and related models by iterative least squares. In *Multivariate Analysis*, ed. by Krishnaiah PR. Academic Press, New York, pp. 391-420, 1966a.
- Wold H. Non-linear estimation by iterative least squares procedures. In *Research Papers in Statistics*, ed. by David F. John Wiley & Sons, New York, 1966b.