SYM INTELLISWITCH 2017: Reliable condition assessment of switches and crossings

IntelliSwitch WP5: *On the comparison of manual measurements of switches and crossings and data from a measuring car*

Pernille Maren Jøndrup, Andre Filipe da Silva Rodrigues, René Fongemie;

Banedanmark

Emil Hovad, Camilla Thyregod, Bjarne Ersbøll;

DTU Compute

Outline

- **Overview of the WP5 data**

- The WP5 data
  - Maintenance information
  - Manual data
  - Loaded data
  - Sensor data

- Statistical modelling
  - Reducing the dimensionality
  - Time series analyses
  - Statistical Process Control (SPC) and Multivariate Statistical Process Control (MSPC)

- Conclusion

WP5: Modelling of Maintenance Performance Indicators

- To develop models for Maintenance Performance Indicators (MPIs) for the condition of the S&C

- Issues to be addressed with the WP5 data:
  - Handling of complex data from various sources: Manual measurements, conventional measurement vehicles (loaded data) and sensors from WP1+WP2 are implemented in selected S&Cs. Data describing time of maintenance and reason for maintenance
  - A data report with the data will be created and uploaded to all the WP partners in IntelliSwitch
  - Misalignment of e.g. the loaded data may occur and thereby techniques (e.g. dynamic time warping) for aligning the data before analysis might be necessary
  - Lowering the dimensionality of data by taking advantage of the correlation structure among the variables. The aim is to extract important features

- Methods in statistical modelling which address the issues:
  - Dimensional reduction and multivariate feature extraction
  - Time series analysis
  - Statistical Process Control (SPC) and Multivariate Statistical Process Control (MSPC)

WP5 – data overview

Background information:

Type of S&C

Other

(e.g. tonnage, age etc.)

Date of the measurements

Data parameters (y):

Manual data (y_m)

Loaded data (y_l)

Sensor data (y_s)

Feature extraction $f(y)$:

Creating functions, $f_m(y_m)$, $f_l(y_l)$, $f_s(y_s)$ and an overall function $f(y_m, y_l, y_s)$

Maintenance info

Compare data and extracted features to the maintenance info

1) Select MPIs from data analysis and feature extraction
2) Model the selected MPIs
3) Based on the detailed modelling of the MPI dynamics, predict S&C behaviour

Decisions from MPI modelling are future work

Outline

- Overview of the WP5 data

- **The WP5 data**
  - **Maintenance information**
  - Manual data
  - Loaded data
  - Sensor data

- Statistical modelling
  - Reducing the dimensionality
  - Time series analyses
  - Statistical Process Control (SPC) and Multivariate Statistical Process Control (MSPC)

- Conclusion

Maintenance data - overview

Background information:

Type of S&C

Other

(e.g. tonnage, age etc.)

Date of the measurements

Data parameters (y):

Manual data (y_m)

Loaded data (y_l)

Sensor data (y_s)

Feature extraction $f(y)$:

Creating functions, $f_m(y_m)$, $f_l(y_l)$, $f_s(y_s)$ and an overall function $f(y_m, y_l, y_s)$

Maintenance info

Compare data and extracted features to the maintenance info

1) Select MPIs from data analysis and feature extraction
2) Model the selected MPIs
3) Based on the detailed modelling of the MPI dynamics, predict S&C behaviour

Decisions from MPI modelling are future work

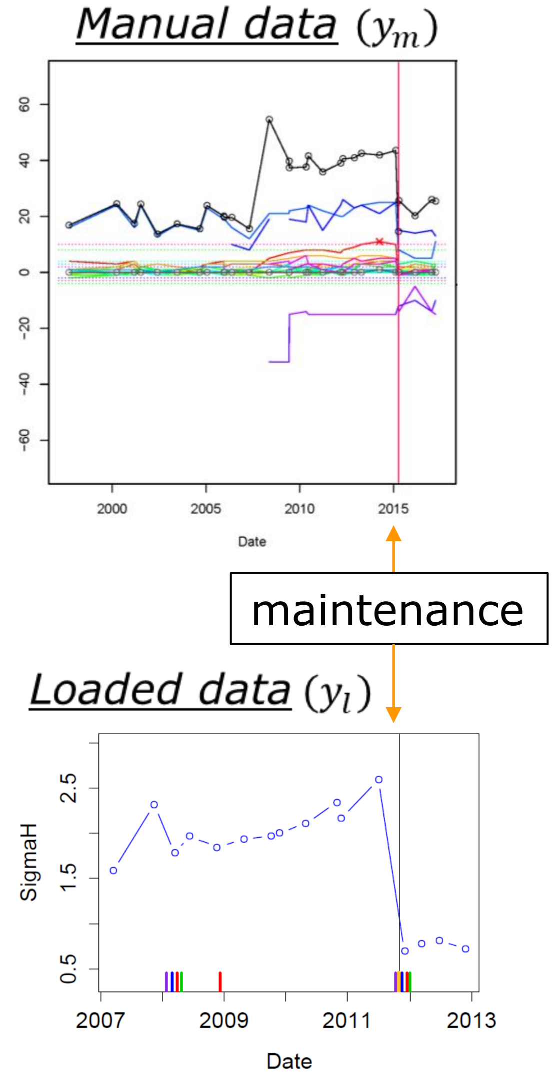
Maintenance information

- Detailed information of the S&C state has been reported and changes that are made is reported

- Data describing time of maintenance and reason for maintenance as e.g.
  - Welding of S&C
  - Switch grinding
  - Tamping
  - Change of S&C
  - Etc…

Compare data to maintenance information

- Note that maintenance information is important
  - Type of maintenance
  - How does maintenance effect the measured data (manual data, loaded data, sensor data)
  - How to model the selected MPIs with respect to the maintenance information



Maintenance data - conclusion

- Conclusion:
  - Detailed information of the S&C state together with the changes are reported

- Outlook:
  - From the maintenance information quantify and find the changes in the measured S&C data (manual, loaded, sensor)

  - Finding correlations between all the collected data and the maintenance information with statistical modelling methods

Outline

- Overview of the WP5 data

- **The WP5 data**
  - Maintenance information
  - **Manual data**
  - Loaded data
  - Sensor data

- Statistical modelling
  - Reducing the dimensionality
  - Time series analyses
  - Statistical Process Control (SPC) and Multivariate Statistical Process Control (MSPC)

- Conclusion

Manual data - overview

Background information:

Type of S&C

Other

(e.g. tonnage, age etc.)

Date of the measurements

Data parameters (y):

Manual data (y_m)

Loaded data (y_l)

Sensor data (y_s)

Feature extraction $f(y)$:

Creating functions, $f_m(y_m)$, $f_l(y_l)$, $f_s(y_s)$ and an overall function $f(y_m, y_l, y_s)$

Maintenance info

Compare data and extracted features to the maintenance info

1) Select MPIs from data analysis and feature extraction
2) Model the selected MPIs
3) Based on the detailed modelling of the MPI dynamics, predict S&C behaviour

Decisions from MPI modelling are future work

Manual data - Information

- The manual data:
  - For the 17 parameters tolerance limits are defined in Banedanmark railway standards
  - A nominal value, max value and min value.
  - Note, sometimes just one limit is needed: Min or max

	a	i	p	i1	p1	y	y1	b	b1	c	c1	d	e	h	d1	e1	h1
	1400+	min	max	min	max	1400+	1400+	1400+	1400+	1400+	1400+	1400+	1300+		1400+	1300+	
Nominelle mål	36	60	78	60	78	35	35	35	35	35	35	35	94	41	35	94	41
Vedligeholdelsestolerance	44		78		78	43	45	43	45	43	45	41	97	44	41	98	44
max./ min.	33	60		60		32	31	32	31	32	31	32	92	40	31	92	40
Sikkerhedstolerance	46		80		80	45	48	45	48	45	48	43	98	44	43	99	44
max./ min.	33	58		58		32	31	32	31	32	31	31	92	39	31	92	39

  - Two types of tolerance exists: The “maintenance intervention limit” and the “the safety intervention limit”

The manual data - plots

- Example of the individual manual parameters (usually values of gauge and wheel passage) are plotted and compared to the maintenance data

- Note the manual data is collected at a specific reference position (km)

The manual data – example of initial modelling

- Investigating the underlying structures in the manual data
  - By e.g. creating a function $f_m(y_m)$ from only the manual values y_m initially trying the Euclidian norm.

  - The Euclidian norm is defined as the overall distance from all the individual manual parameters y_i to their respective nominal values n_i at a specific measurement data,

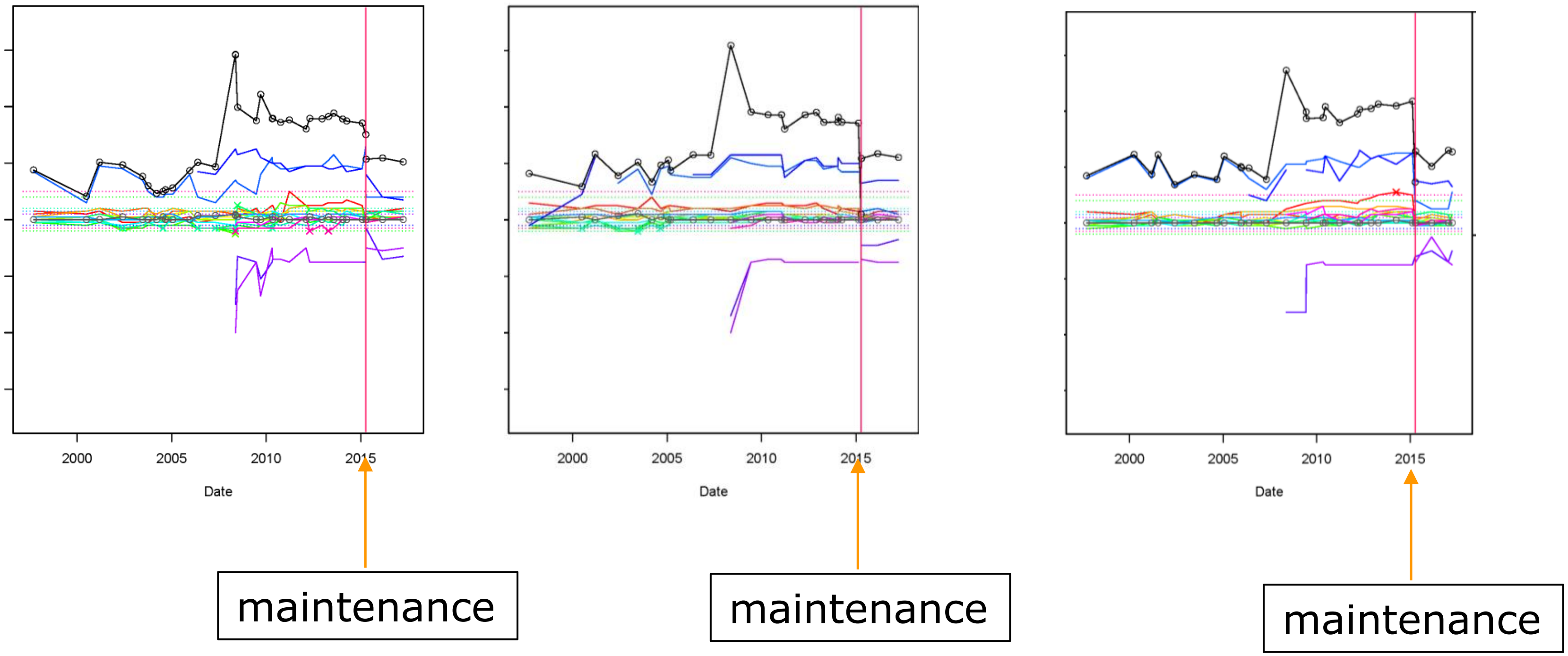
$$f_m(\boldsymbol{y}_m) = \sqrt{\sum_{i=1}^{17} (y_i - n_i)^2}$$

  - This is a naïve approach since the Euclidian norm $f_m(y_m)$ should consider the distance to the min values and the maximum values. This is currently ignored

  - Weighting the different parameters in the manual data might be of importance

The manual data – example with the norm

- Investigating the effect of the maintenance on the manual data
  - The Eucledian norm descreases (black line decreases) after the maintenance date for the three S&C's shown



The manual data - conclusion

- Conclusion:
  - A norm was suggested to give one overall value describing the manual data at a specific date

- Outlook:
  - Correlate the maintenance information to the manual data by finding causality in the S&C maintenances
  - Next correlate the manual data to the loaded data
  - Next correlate the manual data to the sensor data in WP2
  - Finding correlations between the all the collected data with statistical modelling

Outline

- Overview of the WP5 data

- **The WP5 data**
  - Maintenance information
  - Manual data
  - **Loaded data**
  - Sensor data

- Statistical modelling
  - Reducing the dimensionality
  - Time series analyses
  - Statistical Process Control (SPC) and Multivariate Statistical Process Control (MSPC)

- Conclusion

Loaded data - overview

Background information:

Type of S&C

Other

(e.g. tonnage, age etc.)

Date of the measurements

Data parameters (y):

Manual data (y_m)

Loaded data (y_l)

Sensor data (y_s)

Feature extraction f(y):

Creating functions, $f_m(y_m)$, $f_l(y_l)$, $f_s(y_s)$ and an overall function $f(y_m, y_l, y_s)$

Maintenance info

Compare data and extracted features to the maintenance info

1) Select MPIs from data analysis and feature extraction
2) Model the selected MPIs
3) Based on the detailed modelling of the MPI dynamics, predict S&C behaviour

Decisions from MPI modelling are future work

Loaded data - information

- The loaded data are collected from a track geometry measurement vehicle
- A standard procedure for checking the quality of the S&Cs includes the use of the loaded data measurements
- The purpose of investigating the loaded data is to find dependencies of the parameters from the loaded data to the manual data parameters
- Next find dependencies of the parameters from the loaded data to the maintenance information and the sensor data
- Misalignment of e.g. the loaded data may occur and thereby techniques (e.g. dynamic time warping) for aligning the data before analysis might be necessary

Loaded data - information

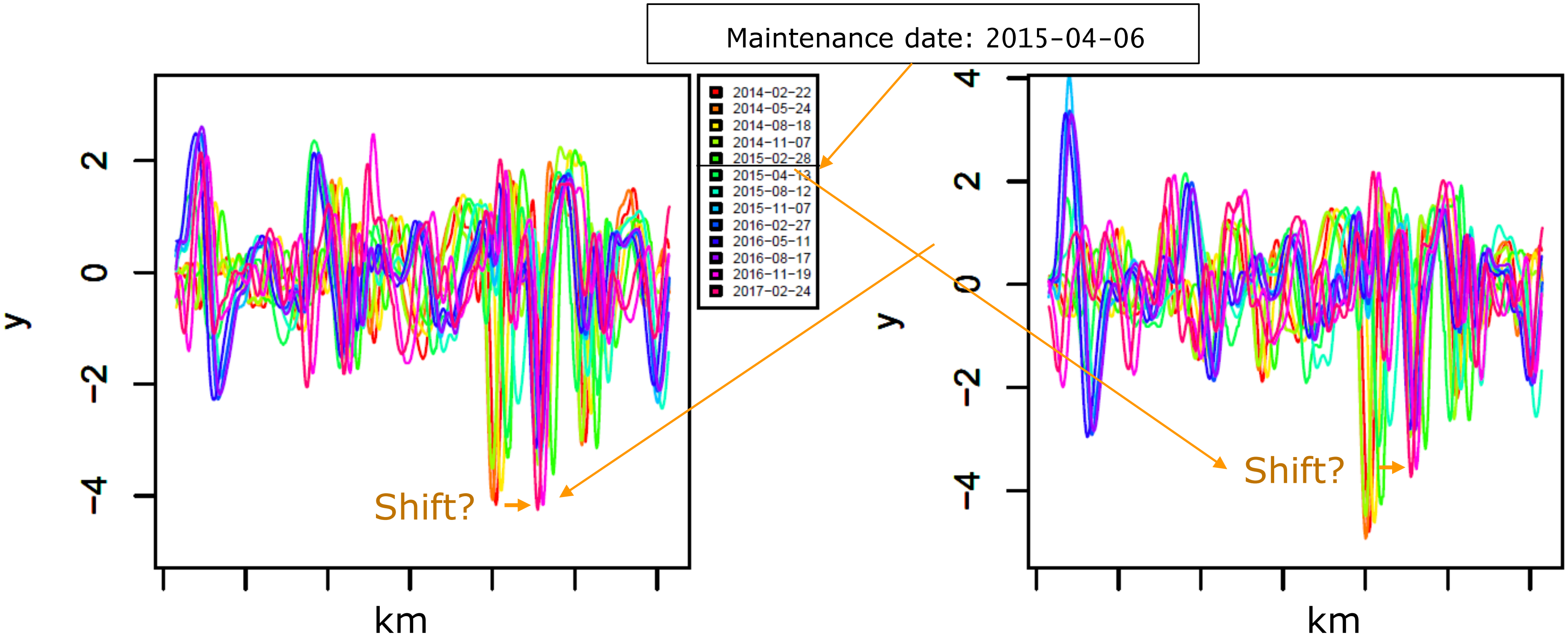
- The loaded data parameters are:
- Level and alignment data:
- Level data H 3-25m = Højd_R_D1
- Level data V 3-25m = Højd_L_D1
- Level data H 25-70m = Højd_R_D2
- Level data V 25-70m = Højd_L_D2
- Alignment data H 3-25m = Side_R_D1
- Alignment data V 3-25m = Side_L_D1
- Alignment data H 25-70m = Side_R_D2
- Alignment data V 25-7 = Side_L_D2

- Track gauge:
- Twist 2m = Vridning 2m
- Twist 3m = Vridning 3m
- Twist 6m = Vridning 6m
- Twist 9m = Vridning 9m
- Twist 12m = Vridning 12m
- Twist 15m = Vridning 15m

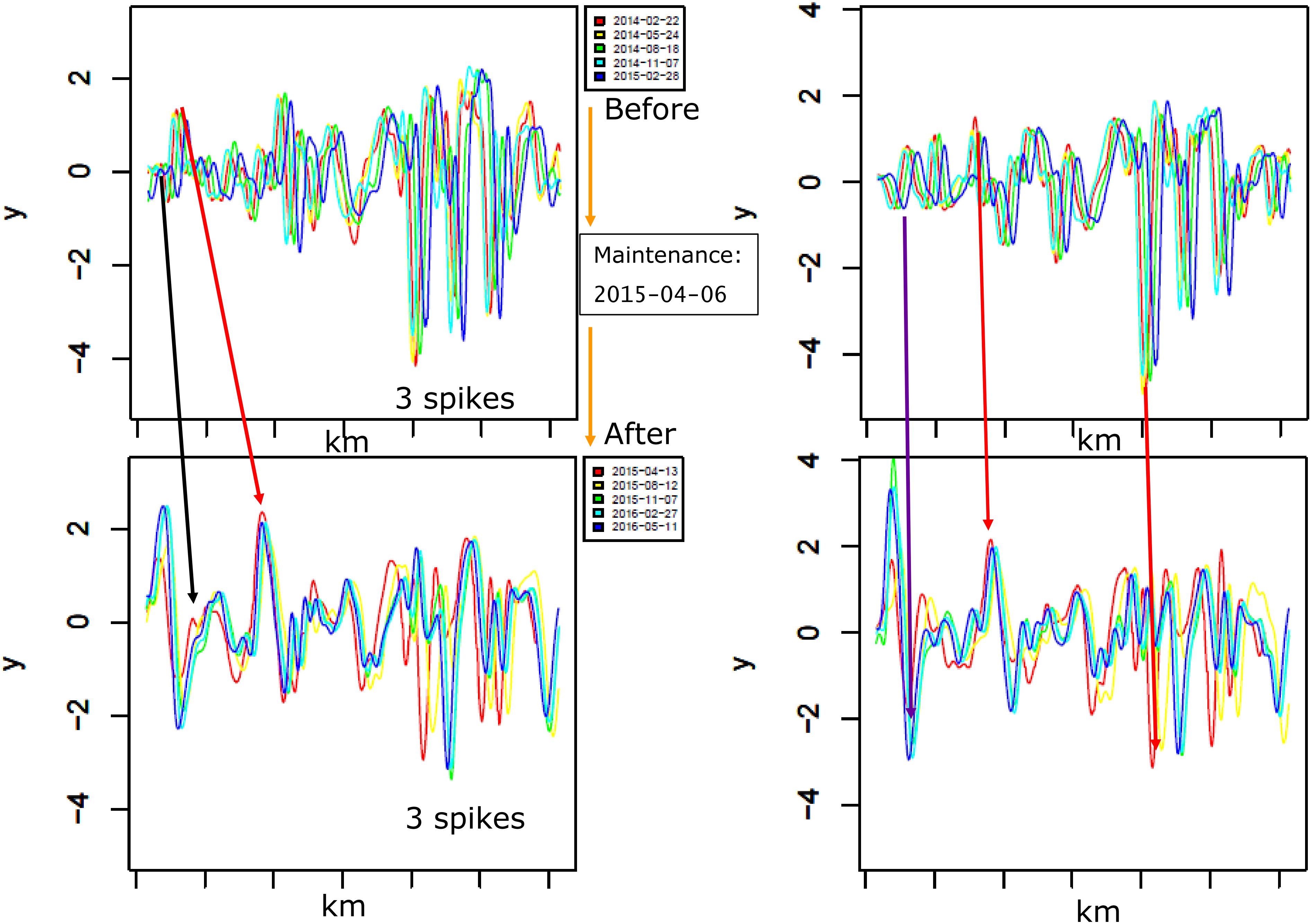
- Wear = Overhøjde
- Corrugation_30-300mm_DepthRight = Rifler og Bølger 30-300mm dybde H
- Corrugation_30-300mm_DepthLeft = Rifler og Bølger 30-300mm dybde V
- Rail inclination R = Skinnehældning H
- Rail inclination L = Skinnehældning V
- Rail_HeightWear_R = Vertical wear of rail profiles = Højdeslid H
- Rail_HeightWear_L = Vertical wear of rail profiles = Højdeslid V
- DK_RW45dgR_W2 = The lateral wear of the rail profiles (mesured 45 degrees on the gauge corner) = Sideslid (45 grader) H
- DK_RW45dgL_W2 = The lateral wear of the rail profiles (mesured 45 degrees on the gauge corner) = Sideslid (45 grader) V

Loaded data - plots

- Compare loaded data to the maintenance data



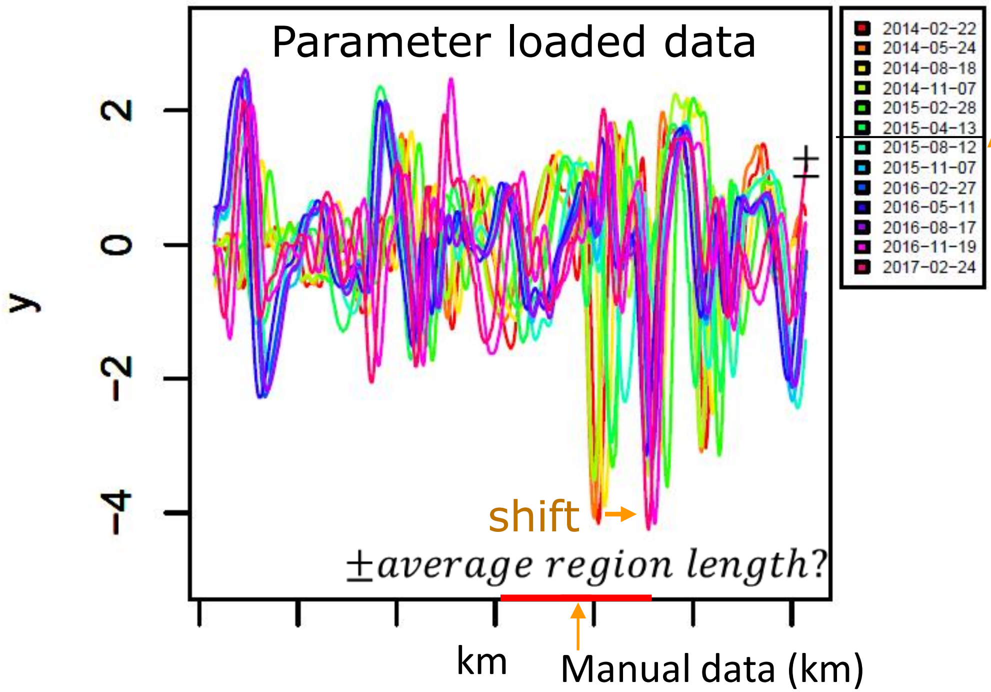
- Note that alignment was not performed on the loaded data

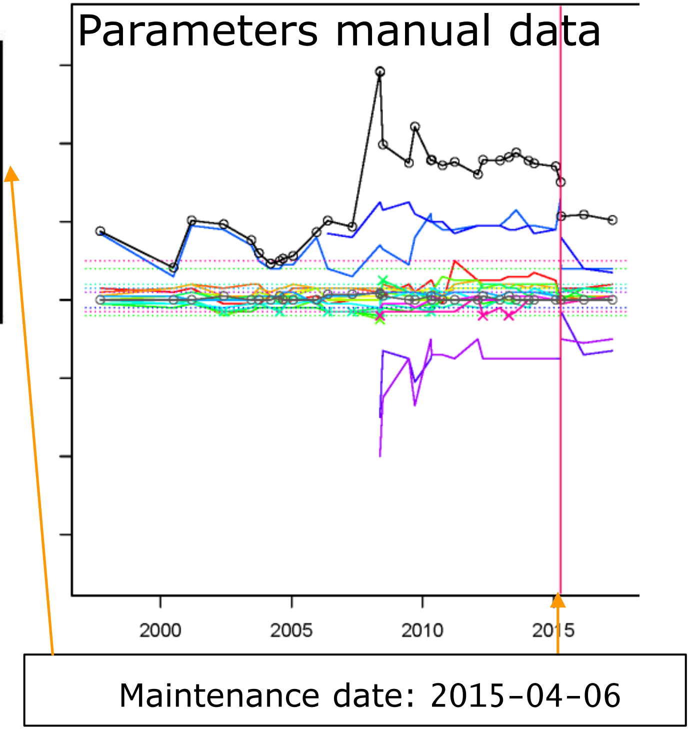


2014-02-22
2014-05-24
2014-08-18
2014-11-07
2015-02-28
Before
Maintenance:
2015-04-06
After
2015-04-13
2015-08-12
2015-11-07
2016-02-27
2016-05-11
3 spikes
3 spikes
y
km

Loaded data – correlate to the manual data

- Specific positions are known for the manual data and therefore can be used to compare to the values of the loaded data





- An e.g. average (max, min, sd) can be calculated from the loaded data in the region where the manual data was monitored (note alignment) for direct numerical comparison

Loaded data - conclusion

- Conclusion:
  - The loaded data was plotted for different measured dates
- Outlook:
  - Finding correlations between the loaded data and the manual measurements
  - Investigate if the loaded data gives good agreement to the maintenance information
  - Finding correlations between the loaded data and the sensor data (WP1+WP2)
  - Use e.g. time series analysis to investigate the loaded data behaviour
  - Finding correlations between all the collected data taking account to the maintenance information with statistical modelling methods

Outline

- Overview of the WP5 data
- **The WP5 data**
  - Maintenance information
  - Manual data
  - Loaded data
  - **Sensor data**
- Statistical modelling
  - Reducing the dimensionality
  - Time series analyses
  - Statistical Process Control (SPC) and Multivariate Statistical Process Control (MSPC)
- Conclusion

Sensor data - overview

Background information:

Type of S&C

Other

(e.g. tonnage, age etc.)

Date of the measurements

Data parameters (y):

Manual data (y_m)

Loaded data (y_l)

Sensor data (y_s)

Feature extraction $f(y)$:

Creating functions, $f_m(y_m)$, $f_l(y_l)$, $f_s(y_s)$ and an overall function $f(y_m, y_l, y_s)$

Maintenance info

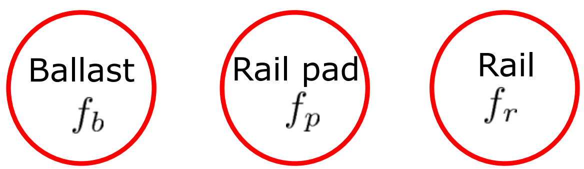
Compare data and extracted features to the maintenance info

1) Select MPIs from data analysis and feature extraction
2) Model the selected MPIs
3) Based on the detailed modelling of the MPI dynamics, predict S&C behaviour

Decisions from MPI modelling are future work

Sensor data - information

- The sensor data from WP1+WP2
  - Measured accelerations and displacements carry information about the health status of the S&C
  - Changes in stiffness and damping at the ballast and rail pad levels are indicators of the ongoing deterioration processes
  - Thereby three y’s from WP2 are given to WP5
    - Ballast (f_b), Rail pad (f_p) and Rail (f_r)



Sensor data - conclusion

- <u>Conclusion:</u>
  - The sensor data (y's) from WP2 are now accessible and can be applied for the statistical modelling

- <u>Outlook:</u>
  - Finding correlations between all the collected data with statistical modelling methods

Outline

- Overview of the WP5 data

- The WP5 data
  - Maintenance information
  - Manual data
  - Loaded data
  - Sensor data

- **Statistical modelling**
  - Reducing the dimensionality
  - Time series analyses
  - Statistical Process Control (SPC) and Multivariate Statistical Process Control (MSPC)

- Conclusion

Statistical modelling - overview

Statistical modelling examples

Background information:

Type of S&C

Other

(e.g. tonnage, age etc.)

Date of the measurements

Data parameters (y):

Manual data (y_m)

Loaded data (y_l)

Sensor data (y_s)

Feature extraction $f(y)$:

Creating functions, $f_m(y_m)$, $f_l(y_l)$, $f_s(y_s)$ and an overall function $f(y_m, y_l, y_s)$

Maintenance info

Compare data and extracted features to the maintenance info

1) Select MPIs from data analysis and feature extraction
2) Model the selected MPIs
3) Based on the detailed modelling of the MPI dynamics, predict S&C behaviour

Decisions from MPI modelling are future work

Statistical modelling – overview

- Reducing the dimensionality
  - Principle Component Analysis (PCA)
  - Sparse Principal Component analysis (SPCA)
  - Factor analysis
  - Other possibilities are dimensional reduction by e.g. variable using correlation structure from the data with methods as e.g. Multi-way Principle Component Analysis

- Time series analysis
  - Co-variance and correlation
  - Autocorrelation

- Statistical Process Control (SPC)
  - Banedanmark has defined specific maintenance levels for the S&Cs
  - Specific univariate models exist in SPC to model process data

- Multivariate statistical process control (MSPC)
  - Combining all the mentioned principles by creating control charts for serially dependent multivariat data including autocorrelation

Outline

- Overview of the WP5 data
- The WP5 data
  - Maintenance information
  - Manual data
  - Loaded data
  - Sensor data
- **Statistical modelling**
  - **Reducing the dimensionality**
  - Time series analyses
  - Statistical Process Control (SPC) and Multivariate Statistical Process Control (MSPC)
- Conclusion

Statistical modelling - reducing the dimensionality

- Variable reduction can show similarities in the data and connect the different measurement types

- Factor analysis
  - Correlation matrix between variables contains information about which variables co-vary. Use this to find groupings of variables

- Sparse Principal Component Analysis (SPCA)
  - Note that SPCA can create reduced dimension with smaller number of parameters in the groupings as compared to the factor analysis
  - Sparse PCA finds linear combinations that contain just a few input variables (sparse loadings). This means that the most important parameters for the different measurements types can be found and linked together
  - This can give clear ideas of the underlying correlation structure
  - Next this can be used to create control charts

Latent Structures-Based Multivariate Statistical Process Control: A Paradigm Shift
Quality Engineering, 26:72–91, 2014
Copyright © Taylor & Francis Group, LLC
ISSN: 0898-2112 print/1532-4222 online
DOI: 10.1080/08982112.2013.846093

- Multi-way Principle Component Analysis can possible be used

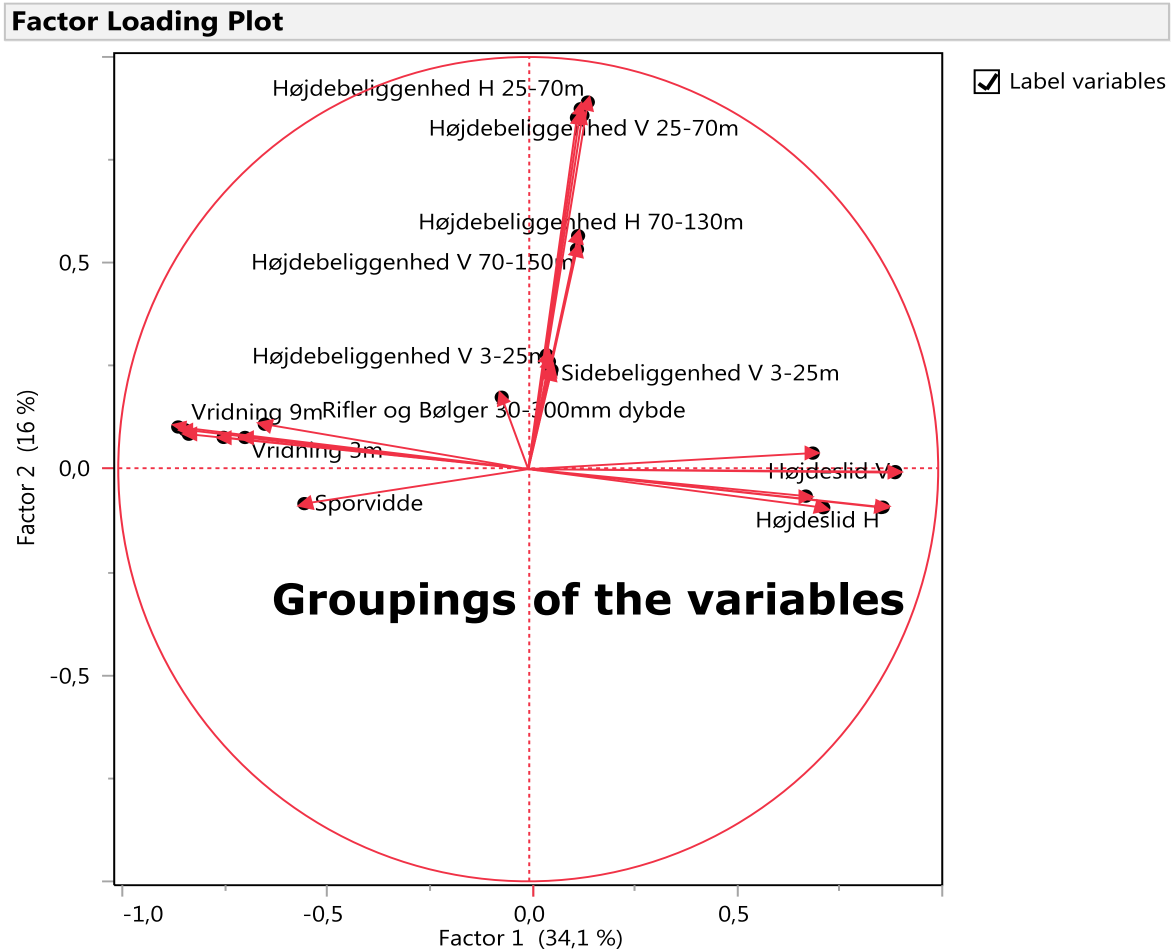
Statistical modelling – Example reducing the dimensionality

- Example of a preliminary factor analysis applying the loaded data

- Correlation matrix between variables contains information about which variables co-vary. Use this to find **groupings of the variables**. Also acheives dimension reduction.

- Method:
  - Eigen-value / eigen-vector analysis of the correlation matrix. (Also called a principal components analysis - PCA)
  - Select the ”relevant” (small) number of components (!)
  - (Perhaps) perform a matrix rotation (Also called the (varimax) factor solution)
  - See if this makes sence by interpreting the results
  - Perhaps repeat for other number of components

Statistical modelling – Example reducing the dimensionality

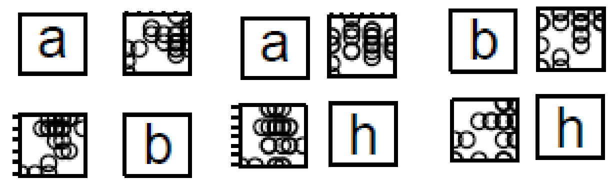


Outline

- Overview of the WP5 data
- The WP5 data
  - Maintenance information
  - Manual data
  - Loaded data
  - Sensor data
- **Statistical modelling**
  - Reducing the dimensionality
  - **Time series analyses**
  - Statistical Process Control (SPC) and Multivariate Statistical Process Control (MSPC)
- Conclusion

Statistical modelling - Time series analysis

- Variable correlation
  - Plot variable against each other



  - Covariance: Cov(X, Y)

$$\mathrm{Cov}(X,Y) = E[(X - E(X))(Y - E(Y))]$$

  - Correlation: Corr(X, Y)

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\sigma_X^2}\sqrt{\sigma_Y^2}}$$

  - Autocovariance:

$$\gamma(k) = E[(z_{t+k} - \mu)(z_t - \mu)]$$

  - The autocorrelation function:

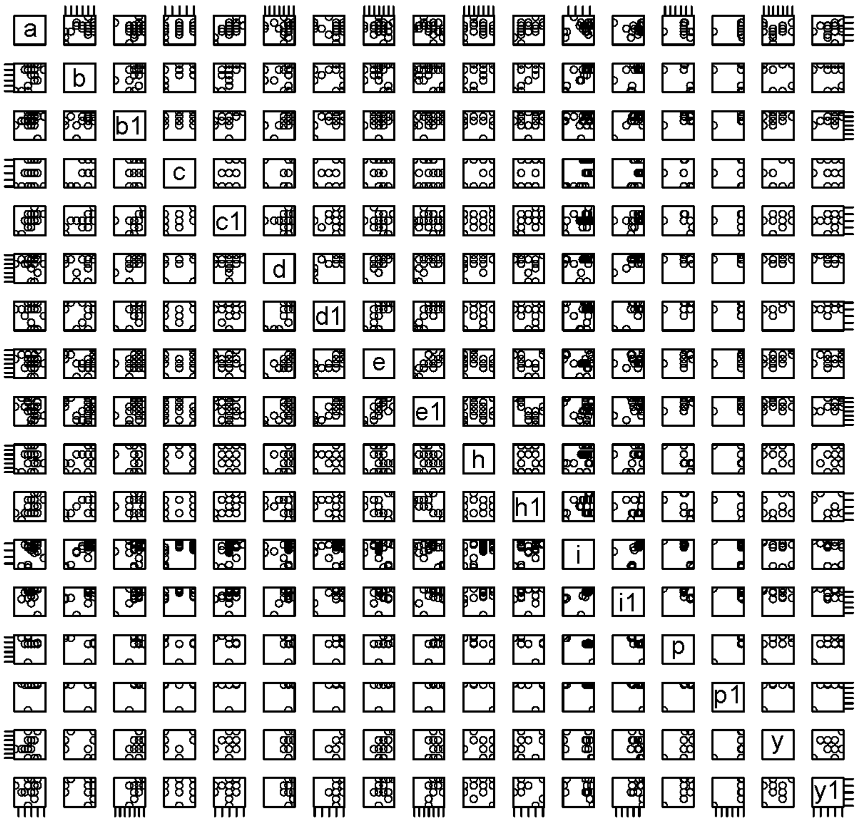
$$\rho(k) = \frac{\gamma(k)}{\sqrt{\gamma(0)}\sqrt{\gamma(0)}} = \frac{\gamma(k)}{\gamma(0)}$$

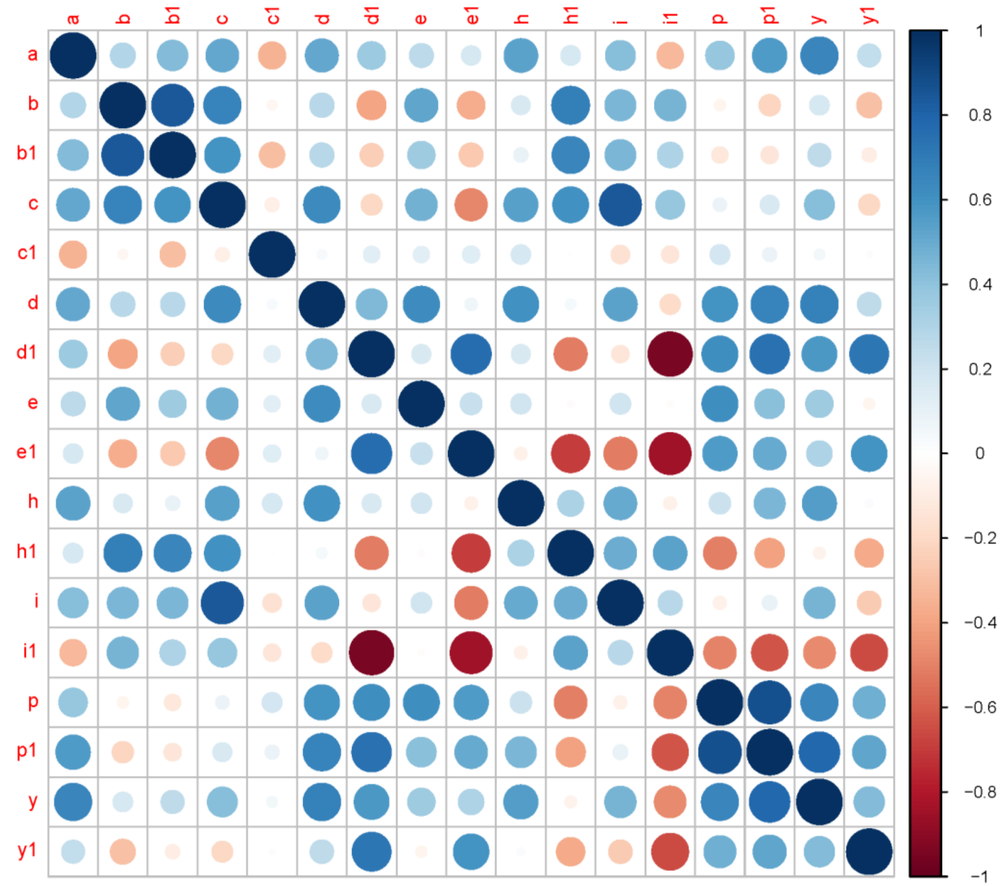
$$\hat{\gamma}(k) = \frac{1}{T}\sum_{t=1}^{T-k}(z_{t+k} - \overline{z})(z_t - \overline{z})$$

$$\hat{\rho}(k) = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)}$$

Statistical modelling - Correlation

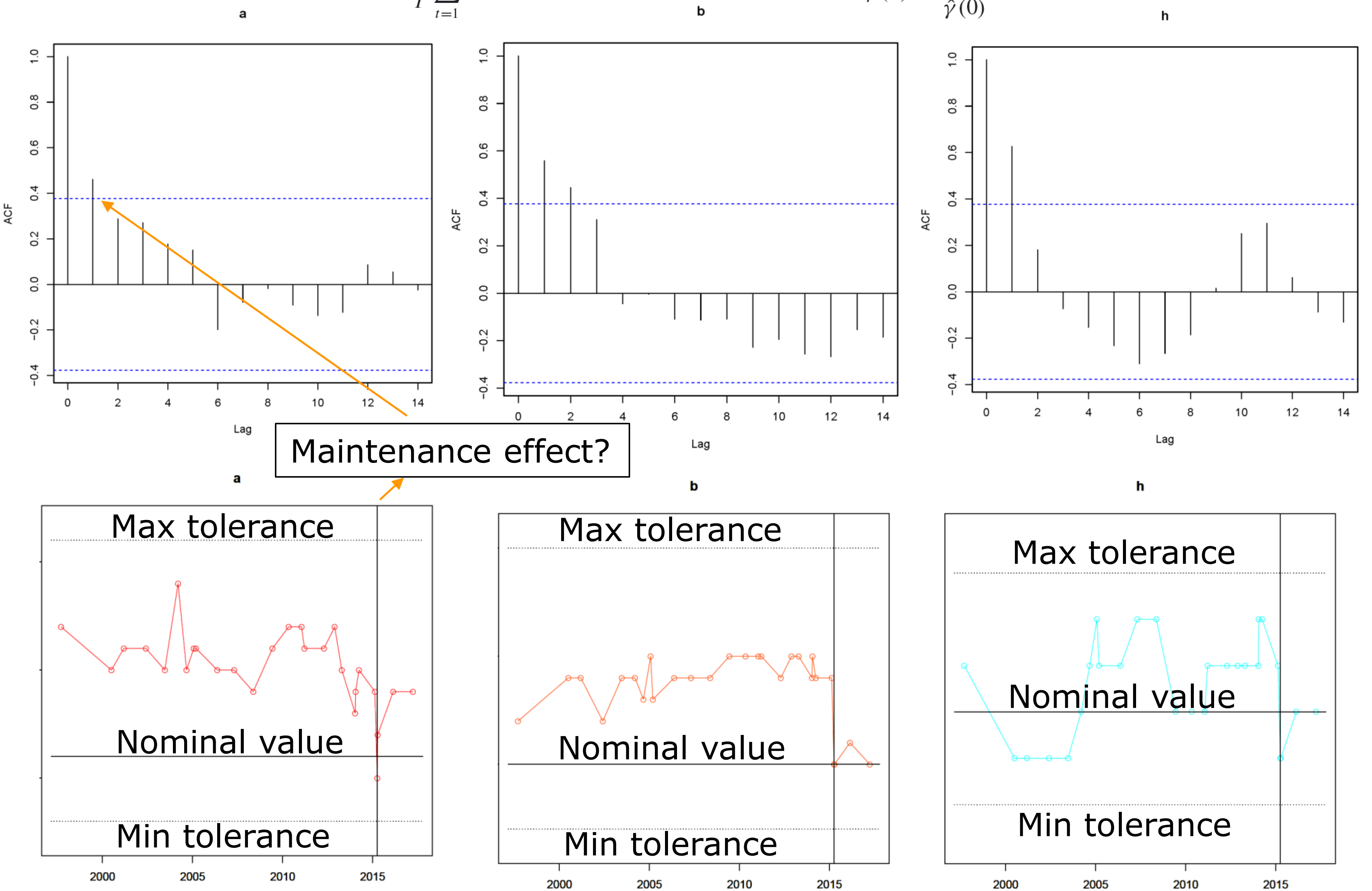
$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\sigma_X^2}\sqrt{\sigma_Y^2}}$$





Statistical modelling – Autocorrelation

$$\hat{\gamma}(k) = \frac{1}{T}\sum_{t=1}^{T-k}(z_{t+k}-\overline{z})(z_t-\overline{z}) \qquad \hat{\rho}(k) = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)}$$



Outline

- Overview of the WP5 data

- The WP5 data
  - Maintenance information
  - Manual data
  - Loaded data
  - Sensor data

- **Statistical modelling**
  - Reducing the dimensionality
  - Time series analyses
  - **Statistical Process Control (SPC) and Multivariate Statistical Process Control (MSPC)**

- Conclusion

Statistical modelling – Statistical Process Control (SPC)

- Banedanmark has defined specific maintenance levels for the S&Cs
  - Banedanmark has set specific nominal values for the manual data and the loaded data
  - Banedanmark has set specific max limits (upper limits) and min limits (minimum limits) for the manual data and the loaded data.
  - The "maintenance intervention limit" and the "the safety intervention limit"

- In traditional SPC a single parameter is used to investigate a process
  - Where an average $\bar{x}$ is calculated from a controlled process and this average is used as the nominal value (assumption of stationarity)
  - Normally the control limits are set by a multiple of the standard deviation around the average values $\bar{x} \pm ns$ (assumption of a constant standard deviation with time)
  - Thereby control charts can be created

- How is the relationship between all the parameters across the different measurement types with respect to creating complex control charts?
  - Can be investigated with e.g. reducing the dimensionality together with time series analysis combined with Multivariate statistical process control (MSPC) gives more complex ways to create improved control charts for all the data in

Statistical modelling – Statistical Process Control (SPC)

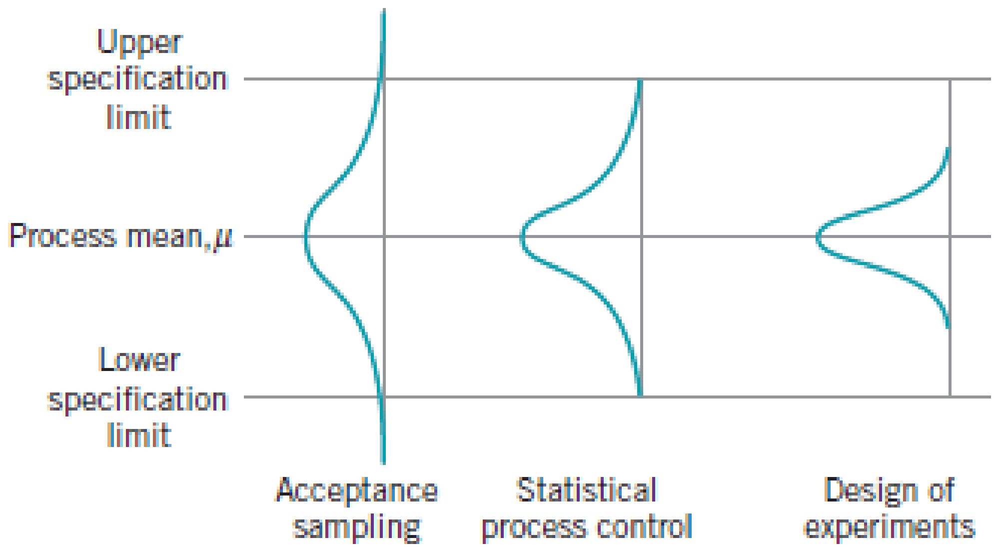
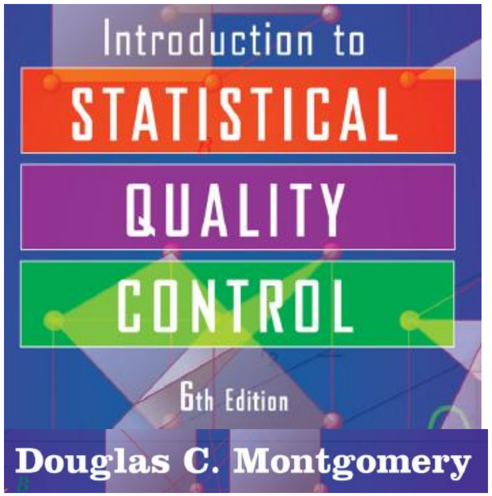


FIGURE 1.8 Application of quality-engineering techniques and the systematic reduction of process variability.



Statistical modelling - Multivariate statistical process control (MSPC)

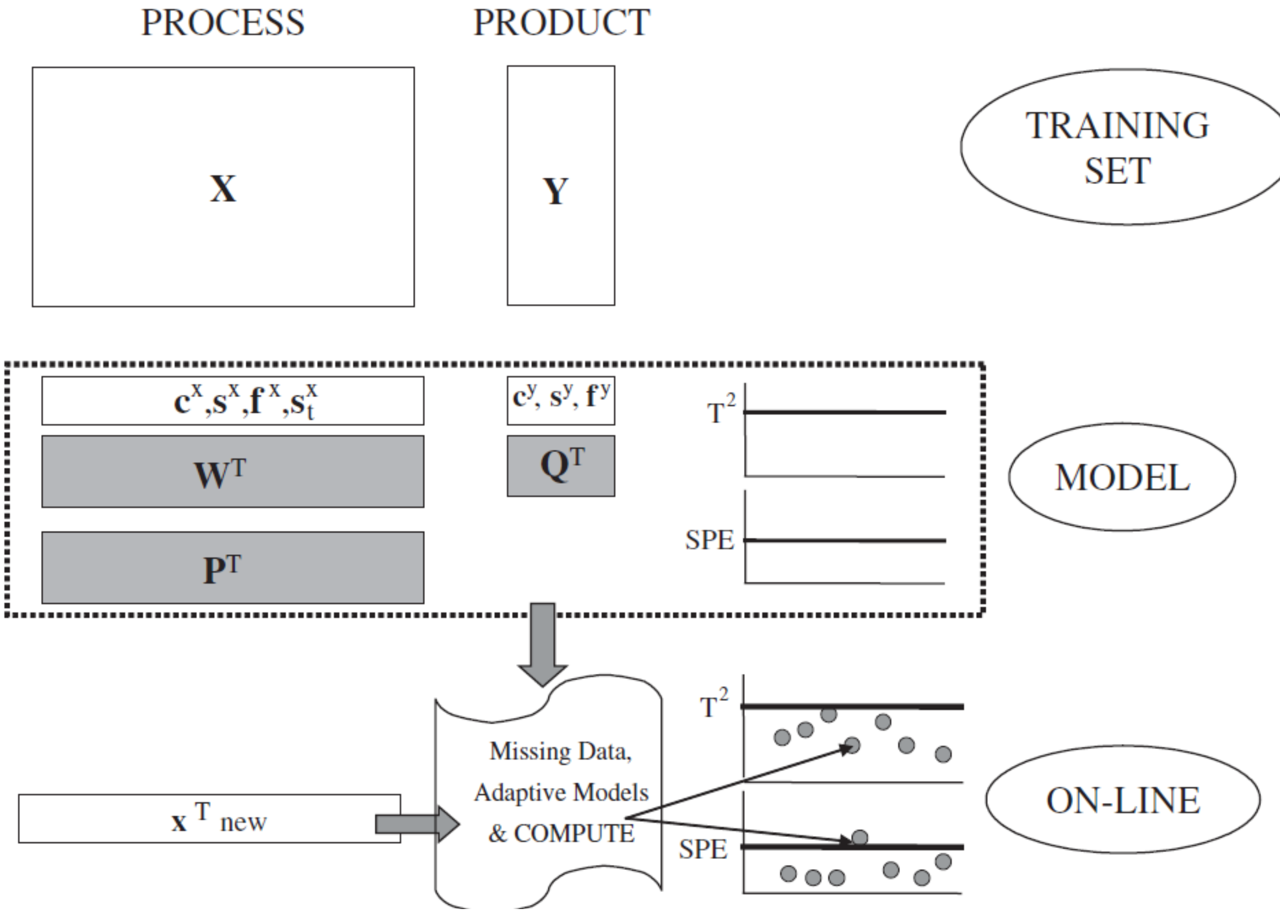
1. Data

    x: S&C type etc.

    y: Measurements

2. Dimensional reduction

3. MSPC



Application of latent variable methods to process control and multivariate statistical process control in industry

Theodora Kourti*,†

INTERNATIONAL JOURNAL OF ADAPTIVE CONTROL AND SIGNAL PROCESSING
Int. J. Adapt. Control Signal Process. 2005; 19:213–246
Published online 27 January 2005 in Wiley InterScience (www.interscience.wiley.com) DOI:10.1002/acs.859

Outline

- Overview of the WP5 data
- The WP5 data
  - Maintenance information
  - Manual data
  - Loaded data
  - Sensor data
- Statistical modelling
  - Reducing the dimensionality
  - Time series analyses
  - Statistical Process Control (SPC) and Multivariate Statistical Process Control (MSPC)
- **Conclusion**

Conclusion - overview

Background information:

Type of S&C

Other

(e.g. tonnage, age etc.)

Date of the measurements

Data parameters (y):

Manual data (y_m)

Loaded data (y_l)

Sensor data (y_s)

Feature extraction $f(y)$:

Creating functions, $f_m(y_m)$, $f_l(y_l)$, $f_s(y_s)$ and an overall function $f(y_m, y_l, y_s)$

Maintenance info

Compare data and extracted features to the maintenance info

1) Select MPIs from data analysis and feature extraction
2) Model the selected MPIs
3) Based on the detailed modelling of the MPI dynamics, predict S&C behaviour

Decisions from MPI modelling are future work

Conclusion

- The data is available and is being investigated
- The maintenance information and the measured data can be investigated in many ways e.g. separate, as groups or all the data with dimensional reduction techniques
- Hopefully dimensional reduction techniques can give insight to the “hidden structures” between the different measurement types and their respective parameters and the maintenance information
- Thereby these “hidden structures” and the underlying dynamics can be found. The knowledge of the data structure can give qualified suggestions for e.g. control charts creating in the reduced dimension for earlier fault detection and the effect of maintenance
- Because **decisions** are easier to take with a decreased amount of information to consider from the reduced dimensions