



The potential of collecting and analyzing BIG DATA from production

“SESAM Seminar the 7. April 2016 at KMD / Odense”

File name: NordZucker_BigData_SESAM.pptx

Responsible: Nord Zucker

Created by: Kai Vahldiek, Suganya Damodaran

Version: 2016-04-07

Animation: yes

Language: English

Active throughout Europe: Our locations

...one of Europe's
leading sugar
producers

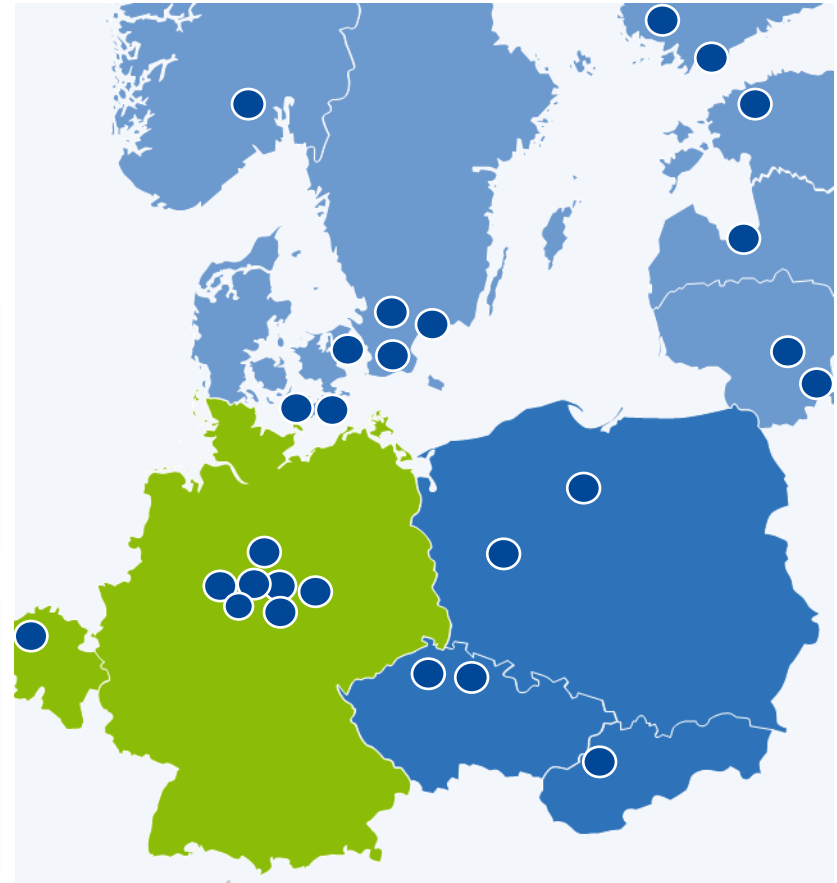
with 18 production sites
(13 sugar factories)

3.300 employees

Production of 2-3 Mio t
sugar out of sugar
beets

Processing up to
6 t beets per minute

Supplied by 15.000
farmers



Sugar production

<http://www.nordicsugar.dk/kend-dit-sukker/fra-roe-til-krystal/>





- 1 Introduction
- 2 Objective
- 3 Examples for quality parameters
- 4 Tool - KNIME
- 5 Workflow for building the prediction model
- 6 Conclusion and Outlook



- 1 Introduction**
- 2 Objective
- 3 Examples for quality parameters
- 4 Tool - KNIME
- 5 Workflow for building the prediction model
- 6 Conclusion and Outlook

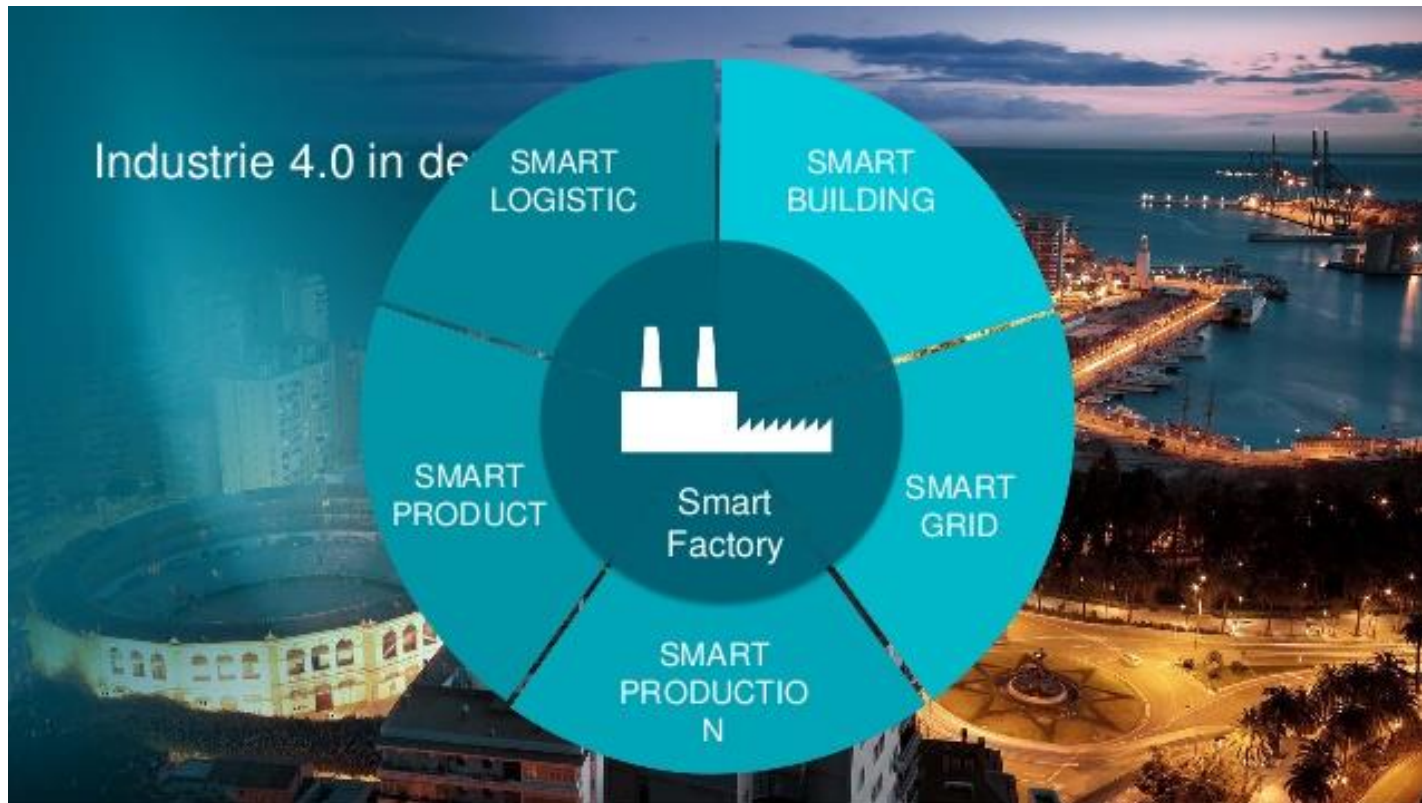
Introduction

Vision: To have a fully automated sugar factory



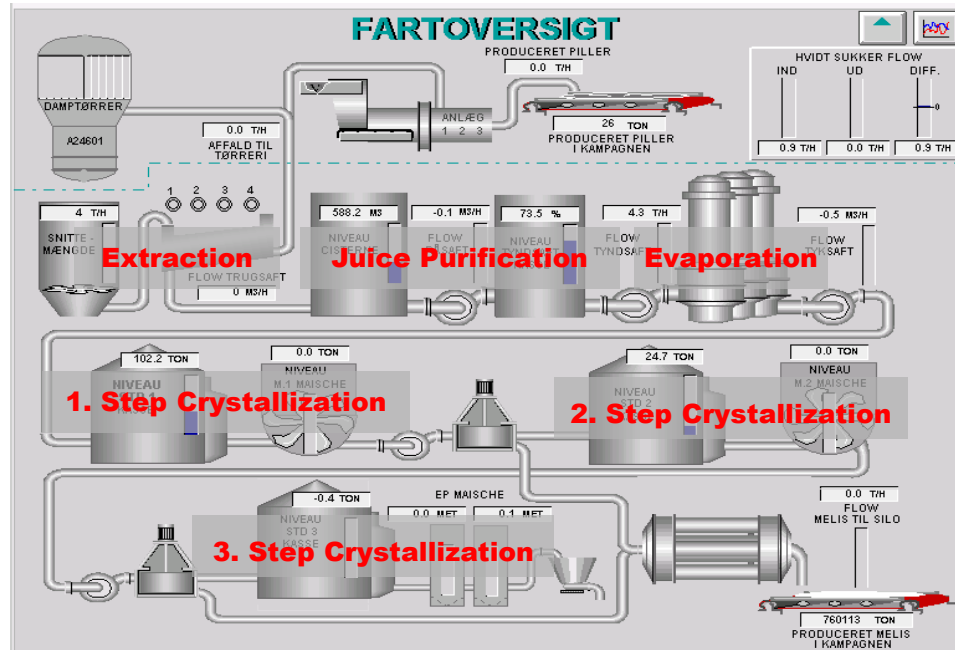
Introduction

Vision: To have a fully automated sugar factory



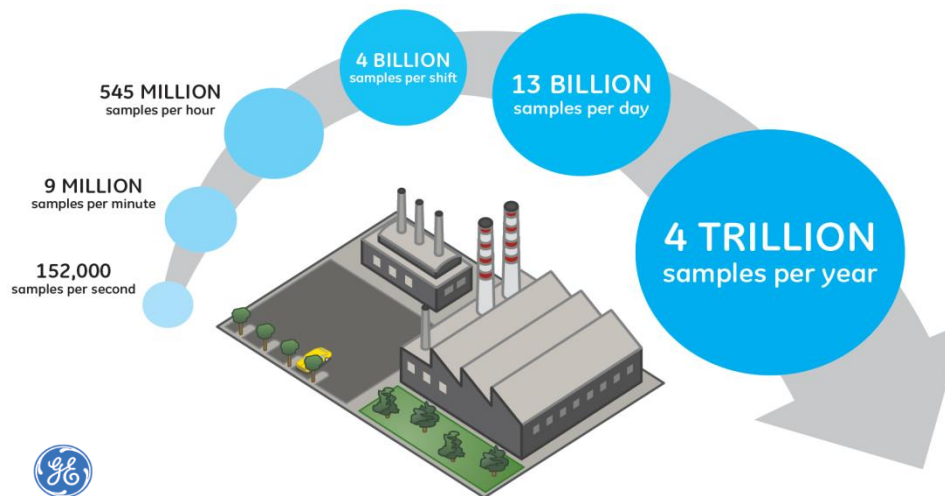
Introduction

- Existing links between Automation and Field devices:
 - Single loop controls
 - Local to one section of the factory
- Can we link different sections of the factory?



Analysis of the current state

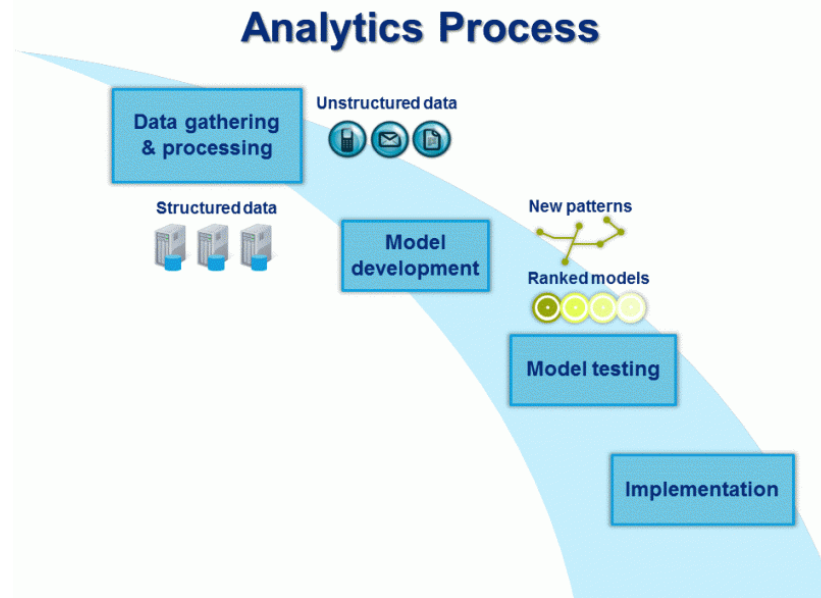
- Model-based technique
 - Complex mathematical representations which require extensive process knowledge.
- Data-based technique
 - Process data from historians combined with process knowledge
 - Big Data Analytics – Examines large data sets to extract useful information



[General Electric: The Rise of Industrial Big Data]

Analysis of the current state

- Model-based technique
 - Complex mathematical representations which require extensive process knowledge.
- Data-based technique
 - Process data from historians combined with process knowledge
 - Big Data Analytics – Examines large data sets to extract useful information





- 1 Introduction
- 2 Objective**
- 3 Examples for quality parameters
- 4 Tool - KNIME
- 5 Workflow for building the prediction model
- 6 Conclusion and Outlook

Objective

- Optimize important process parameters
- Big data approach: Link related sections to optimize process

Case-study:

- Sugar quality parameter prediction

Tasks:

- Identifying variables affecting sugar quality parameter
- Predicting quality parameter
- Cross verification with lab samples
- Identifying potential causes and taking possible corrective actions



- 1 Introduction
- 2 Objective
- 3 Examples for quality parameters**
- 4 Tool - KNIME
- 5 Workflow for building the prediction model
- 6 Conclusion and Outlook

Examples for quality parameters

- Size of the sugar crystal
- Color of the sugar crystal
- Ash content in sugar
- Insoluble residues



- 1 Introduction
- 2 Objective
- 3 Examples for quality parameters
- 4 Tool - KNIME**
- 5 Workflow for building the prediction model
- 6 Conclusion and Outlook

Tool – KNIME (Konstanz Information Miner)



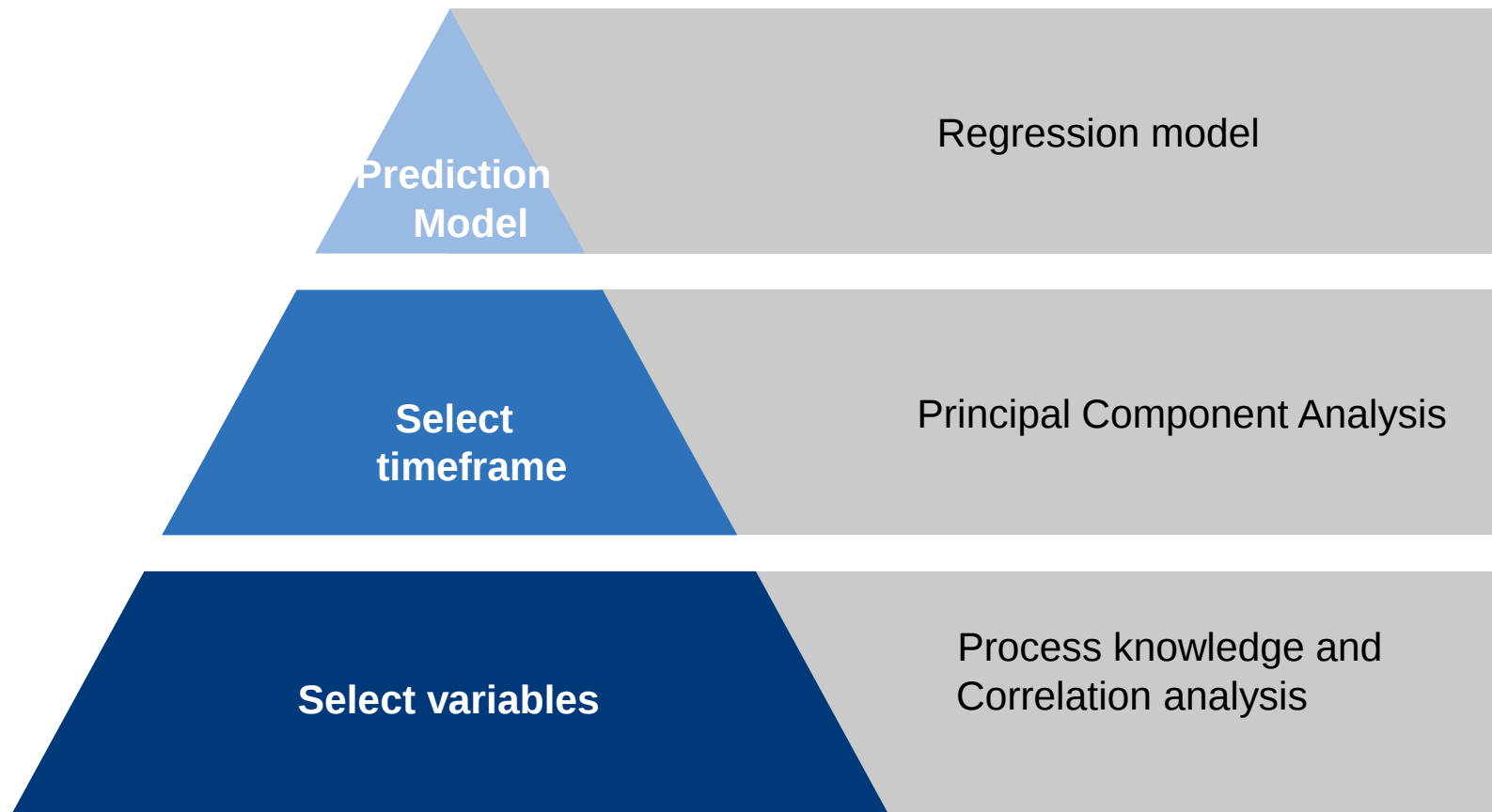
Reasons for choosing KNIME:

- It is a open source software
- KNIME can be used for data mining, analysis and manipulation
- Ability to operate on large datasets: scalability
- Scripting languages like R, Perl, Matlab can be used
- Written in JAVA, based on Eclipse SDK platform



- 1 Introduction
- 2 Objective
- 3 Examples for quality parameters
- 4 Tool - KNIME
- 5 Workflow for building the prediction model**
- 6 Conclusion and Outlook

Workflow for building the prediction model



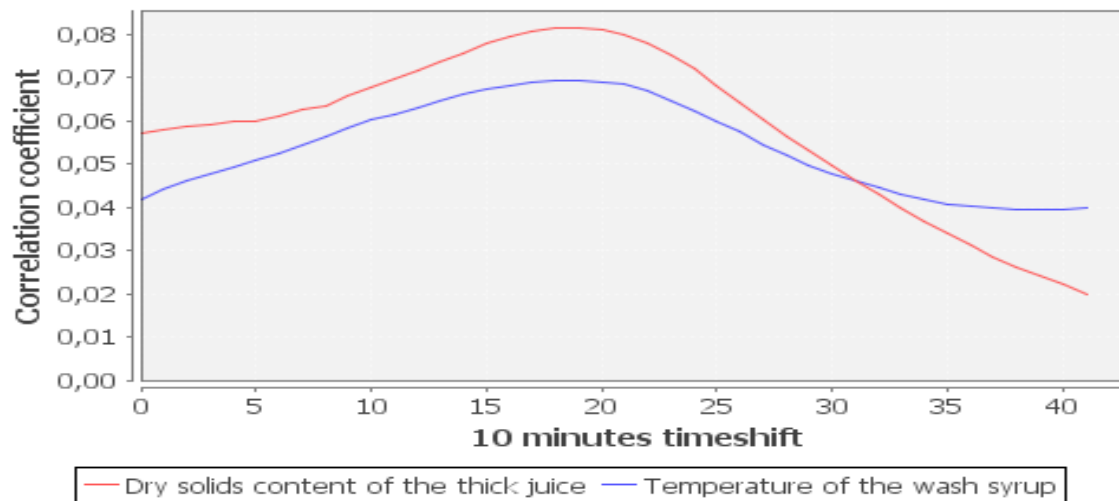
Selecting the process variables

Initial set of parameters:

- Based on process knowledge - 130 process variables

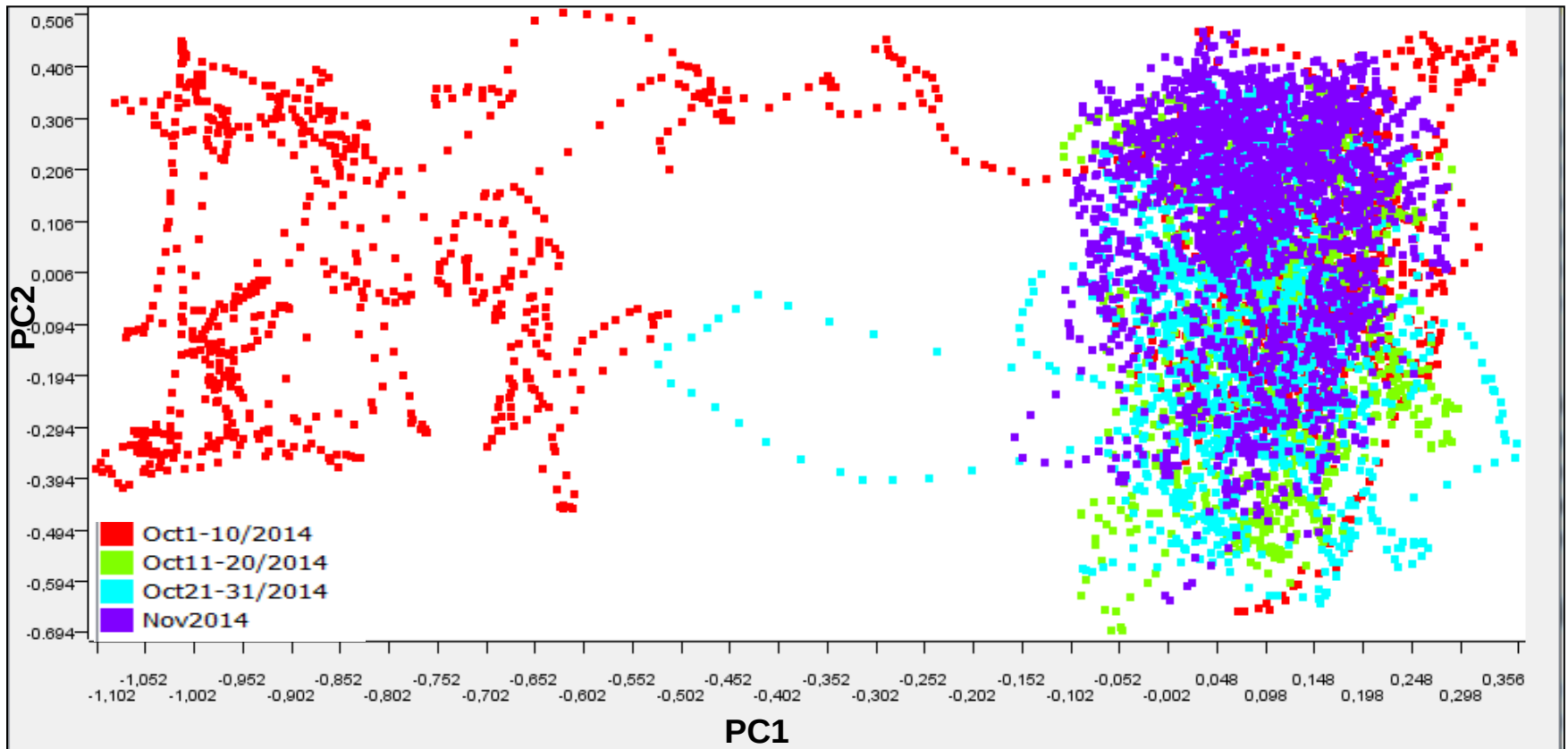
Further reductions:

1. Correlation between the parameters and the quality parameter was analyzed



2. Parameters associated with the batch processes are neglected
3. 40 parameters instead of 130 parameters for modeling

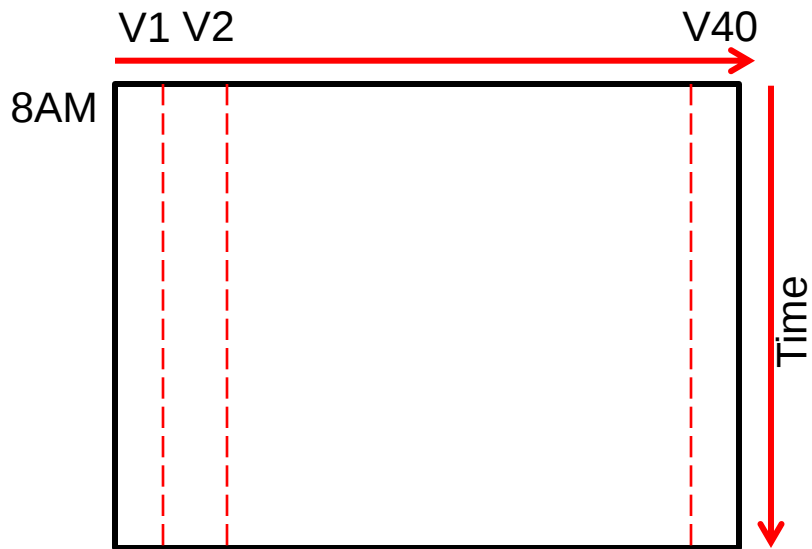
Selecting time frame - PCA



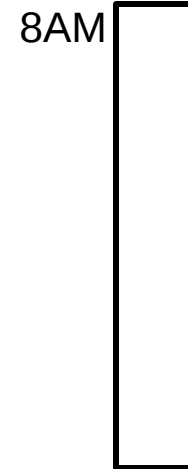
- Map all selected variables in two dimension
- Understand more about the data → time dependency

Prediction model

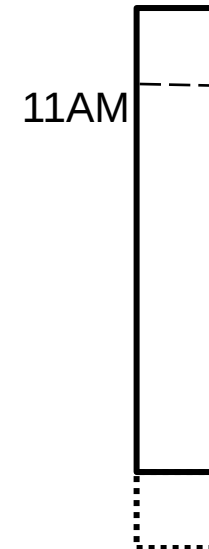
Process variables



Quality Parameter

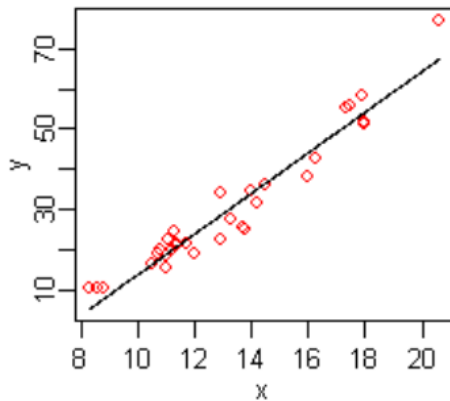


Time shifted 3 hrs

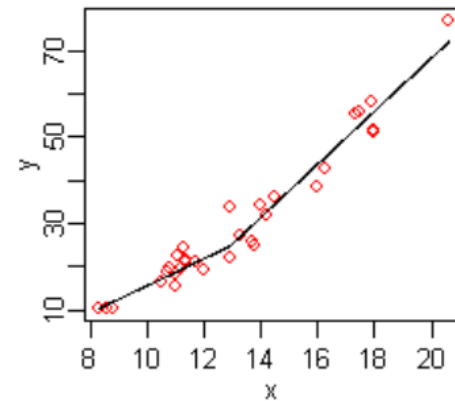


Prediction model

- Model is built using MARS (Multivariate Adaptive Regression Splines)
 - Flexible method
 - Extension of linear models
 - Nonlinearities and interactions between variables



Normal Regression

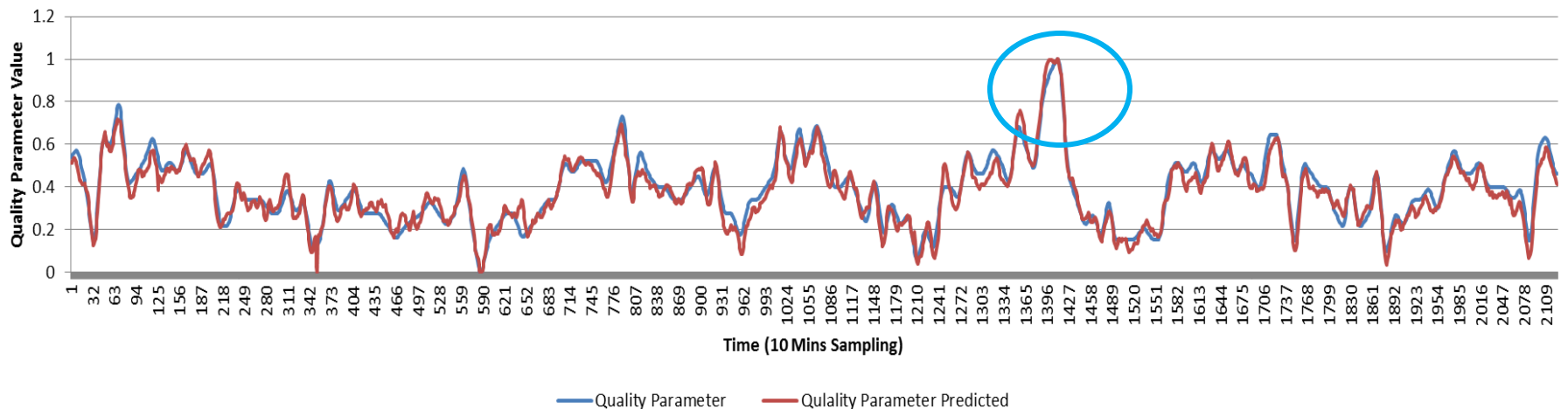


MARS

Workflow for building the prediction model

Data Set 1:

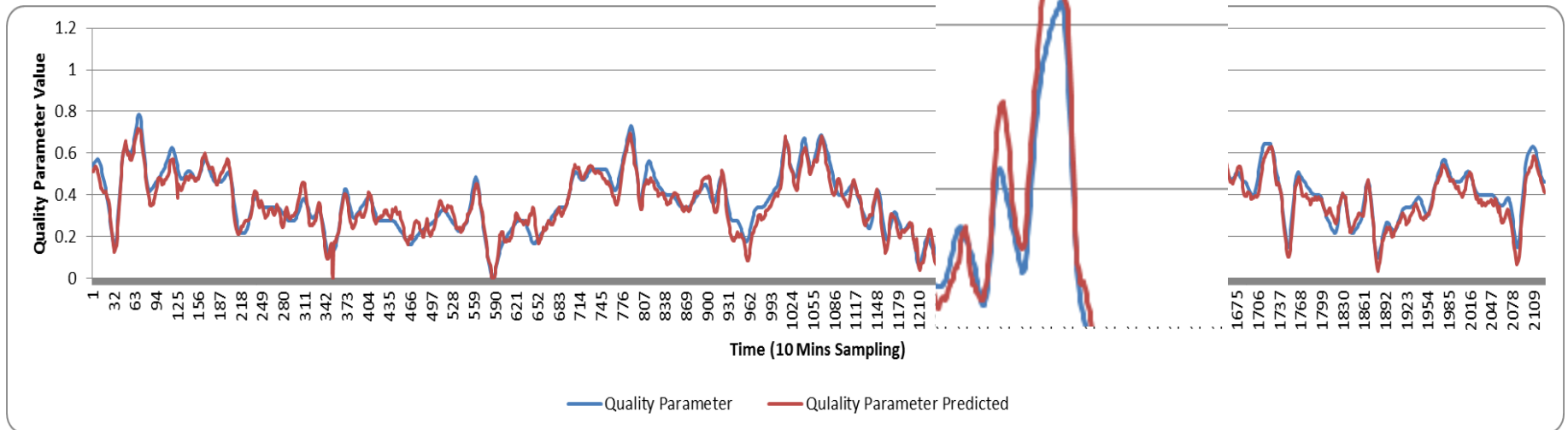
- Data set : 01-OCT-2014 to 15-NOV-2014
- Model is built using data set between 01-OCT-2014 to 31-OCT-2014
- Quality parameter is predicted for the data set 01-NOV-2014 to 15-NOV-2014
- Mean absolute percentage error: 2.8 %



Workflow for building the prediction model

Data Set 1:

- Data set : 01-OCT-2014 to 15-NOV-2014
- Model is built using data set between 01-OCT-2014 to 31-OCT-2014
- Quality parameter is predicted for the data set 01-NOV-2014 to 15-NOV-2014
- Mean absolute percentage error: 2.8 %



Workflow for building the prediction model

- Prediction model identifies important parameters
- The model is built using these 10 parameters

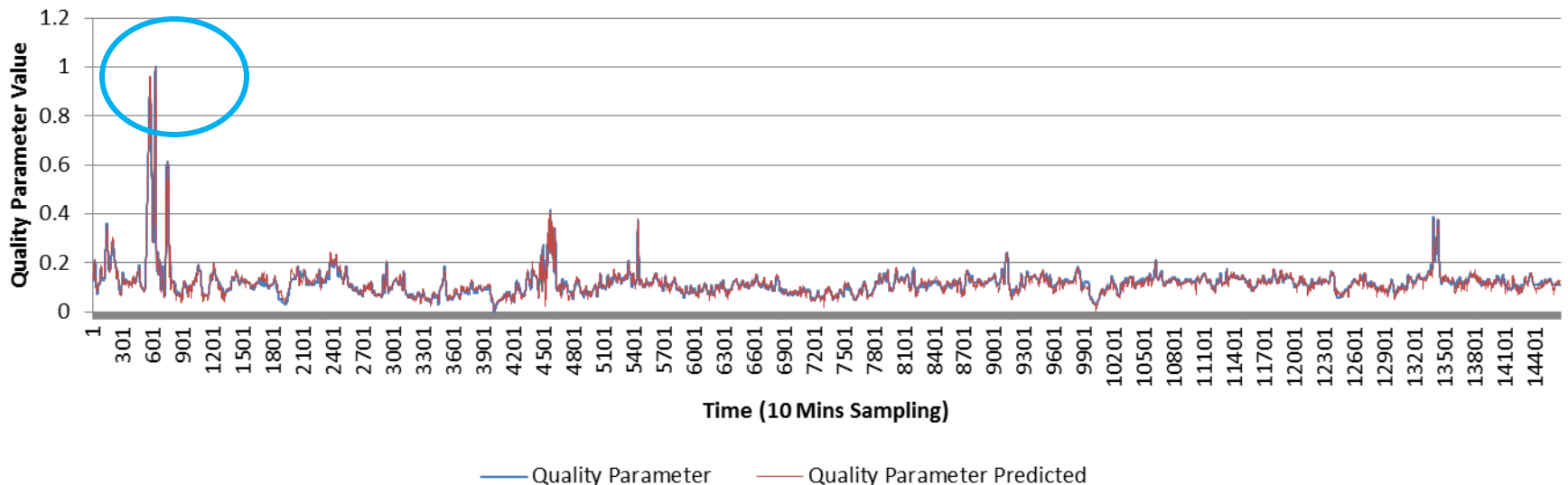
Parameter	Nsubsets	RSS	GCV
Parameter 1	14	100.00	100.00
Parameter 2	14	100.00	100.00
Parameter 3	12	92.60	93.1
...	...	...	...
Parameter 10	2	5.8	6.4

RSS – Residual Sum of Squares
GCV – Generalized Cross Validation

Workflow for building the prediction model

Data Set 2:

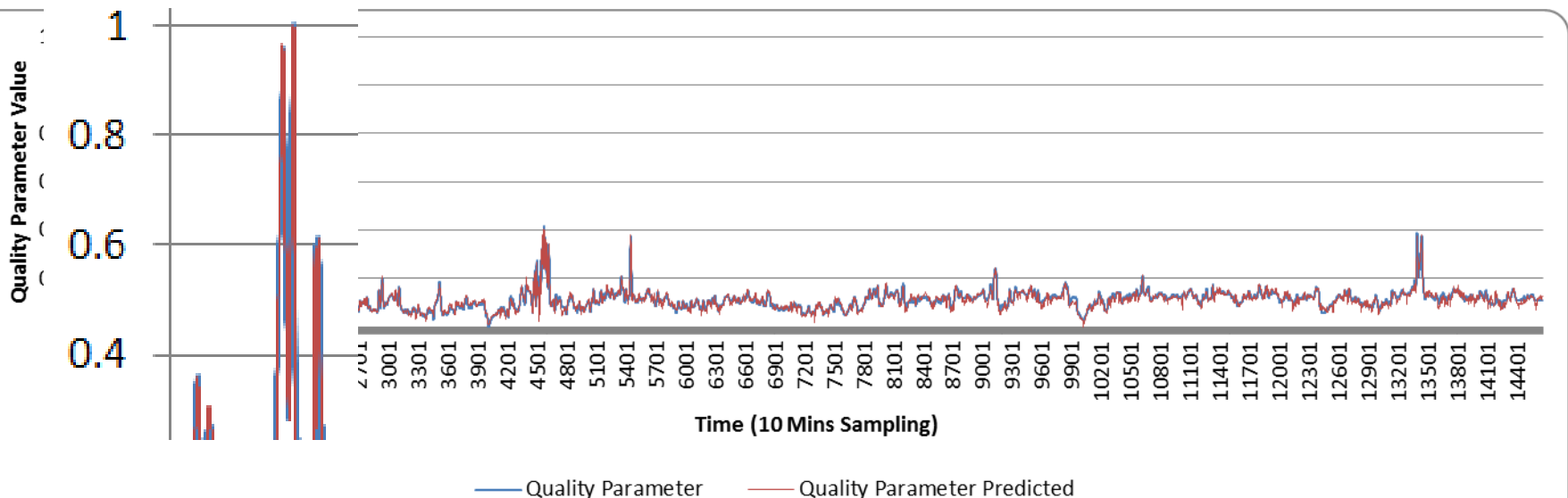
- Data set : 01-OCT-2014 to 31-OCT-2014 and 01-OCT-2013 to 12-JAN-2014
- Model is built using data set between 01-OCT-2014 to 31-OCT-2014
- Quality Parameter is predicted for the data set 01-OCT-2013 to 12-JAN-2014
- One of the important variable was excluded from prediction
- Mean absolute percentage error: 7.5 %



Workflow for building the prediction model

Data Set 2:

- Data set : 01-OCT-2014 to 31-OCT-2014 and 01-OCT-2013 to 12-JAN-2014
- Model is built using data set between 01-OCT-2014 to 31-OCT-2014
- Quality Parameter is predicted for the data set 01-OCT-2013 to 12-JAN-2014
- One of the important variable was excluded from prediction
- Mean absolute percentage error: 7.5 %





- 1 Introduction
- 2 Objective
- 3 Examples for quality parameters
- 4 Tool - KNIME
- 5 Workflow for building the prediction model
- 6 Conclusion and Outlook**

Conclusion and outlook

- Prediction of Quality parameter is possible with error range of 3% (MAPE)
- Identify faulty instruments using operating ranges of instruments
- Cost savings by taking corrective action

Next steps:

- Online implementation
- Automatic identification of parameters causing bad sugar quality
- Compare results with other prediction approaches like Neural Network

Thank You
for Your Attention!