Industrial Research Chair Control of Oilsands Processes

Process Data Analytics
State of the art and applications in oil sands industry

Biao Huang

Department of Chemical and Materials Engineering
University of Alberta

Sept. 2nd, 2016

Presentation @ IFAC MMM Vienna - BH

Outline

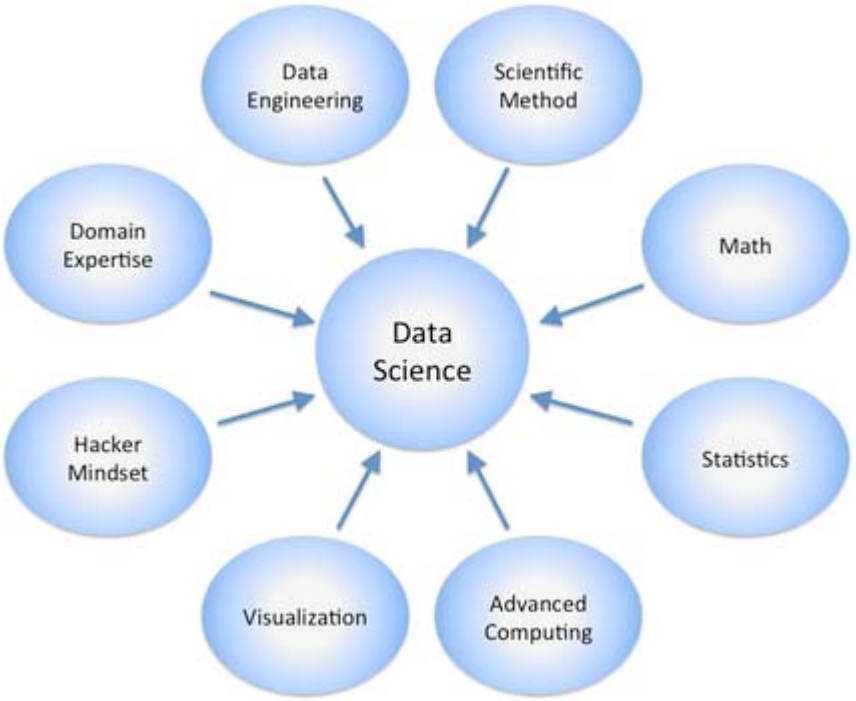
- Data Analytics – State of the art
- Oil Sands Industry
- Process Data Analytics in Applications
- Analytics Toolboxes in Progress
- Conclusion

Data Analytics

Data Analytics

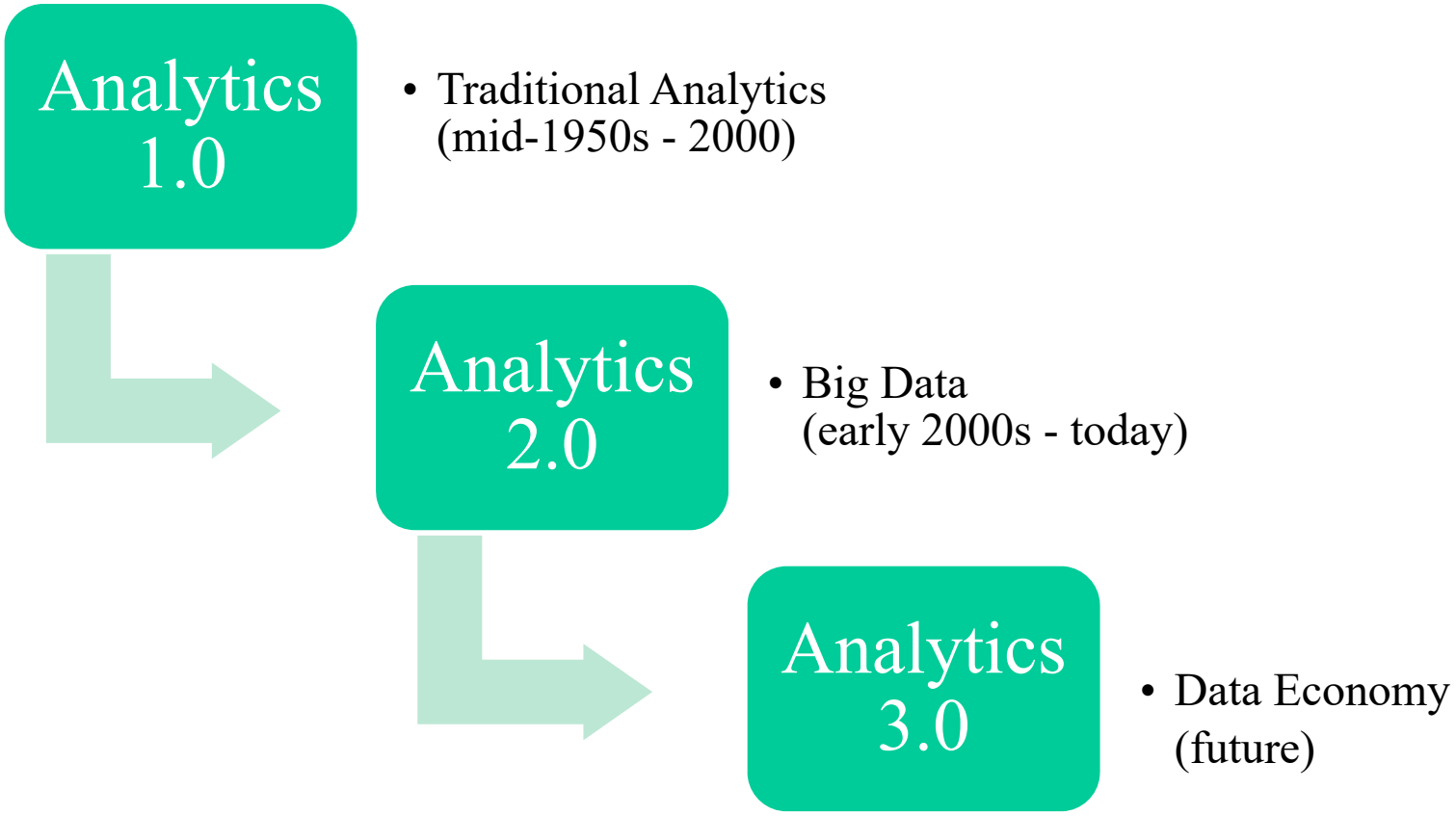
Data Era

Data Science

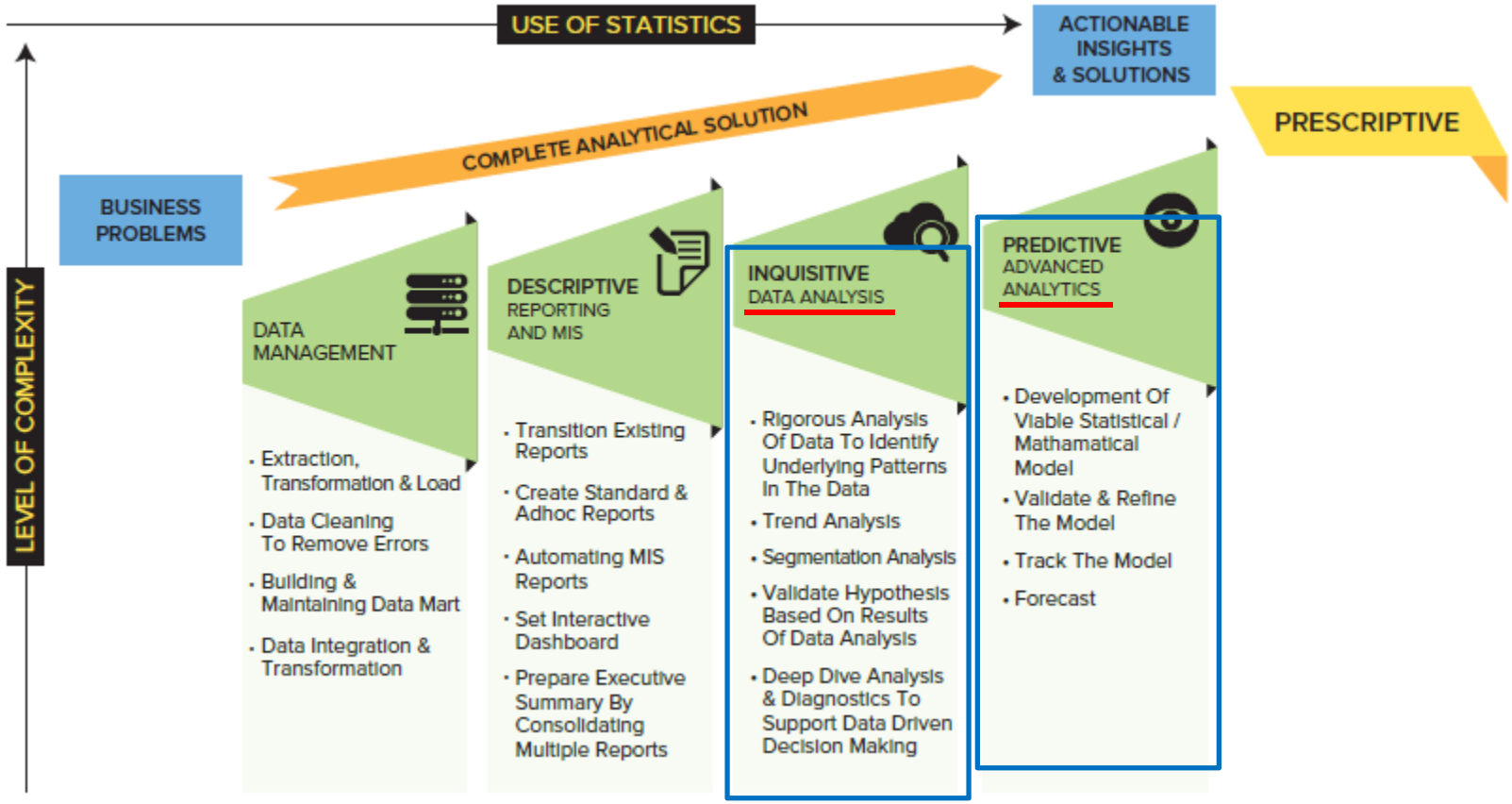


https://www.quora.com/What-is-the-difference-between-Data-Analytics-Data-Analysis-Data-Mining-Data-Science-Machine-Learning-and-Big-Data-1

History of Data Analytics

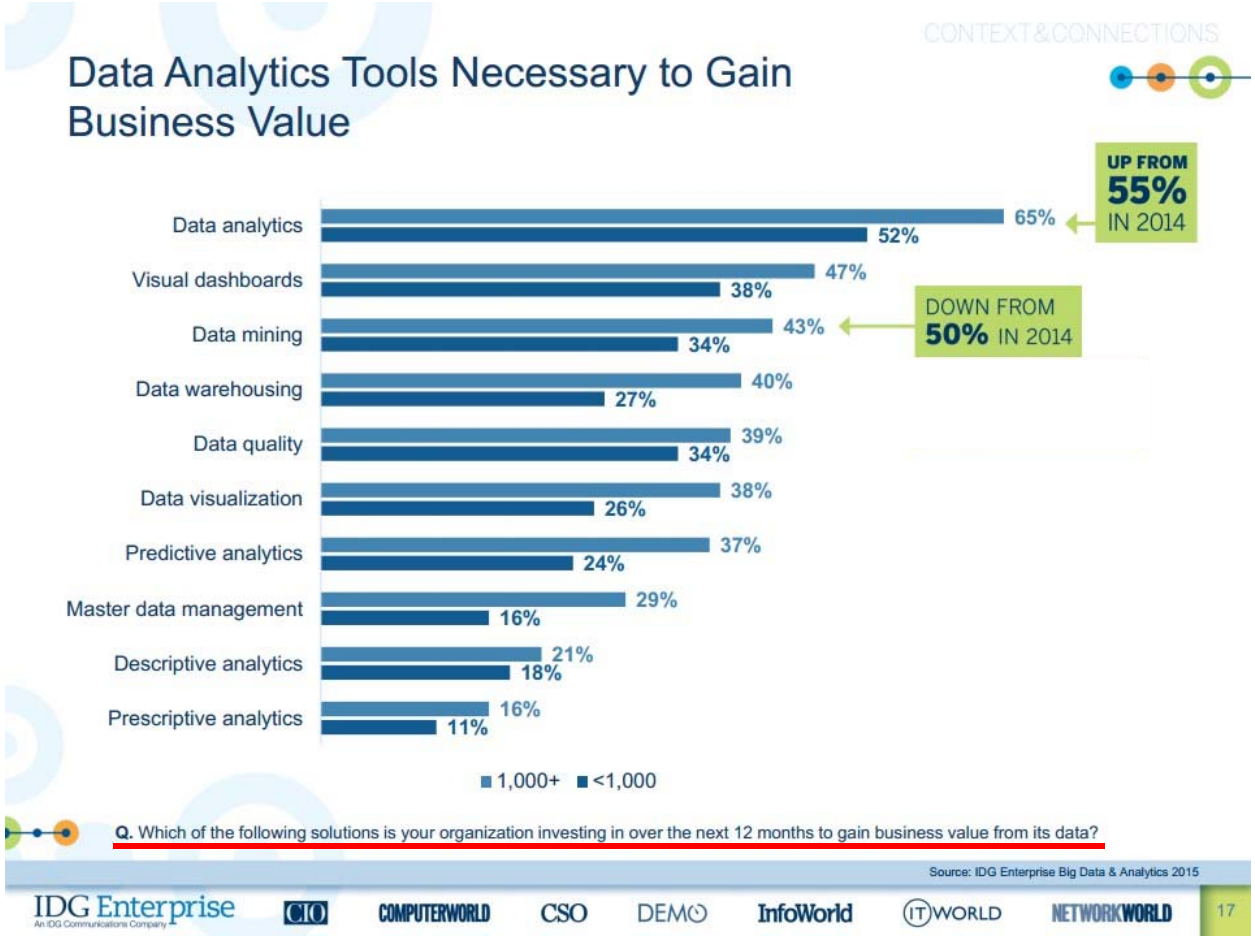


Data Analytics vs Data Analysis



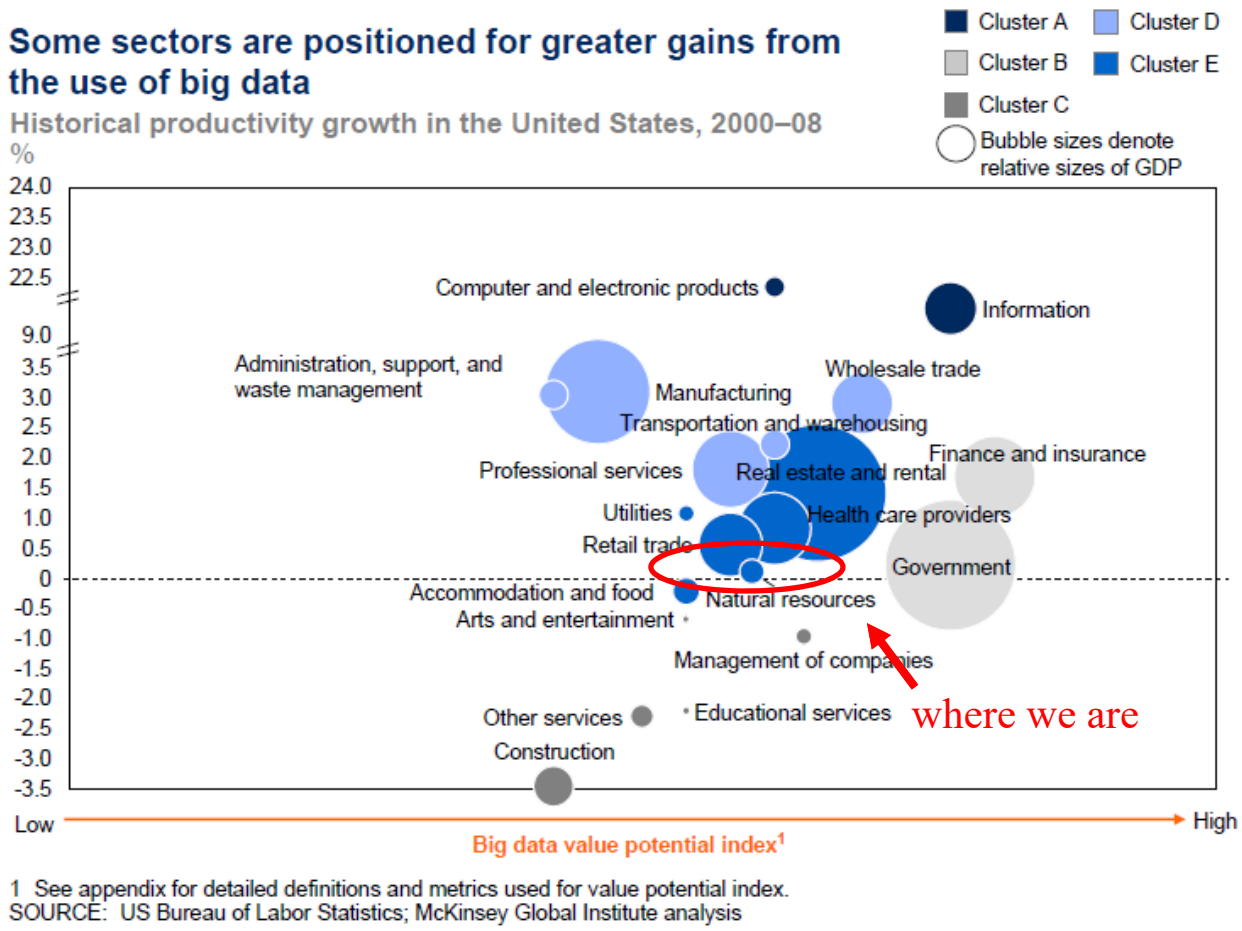
http://www.blueoceanmi.com/big-data-analytics-overview

Increasing Trend of Data Analytics



http://www.forbes.com/sites/louiscolumbus/2015/03/15/data-analytics-dominates-enterprises-spending-plans-for-2015/#5df926dc3eb4

Application of Data Analytics: Engineering



1 See appendix for detailed definitions and metrics used for value potential index.
SOURCE: US Bureau of Labor Statistics; McKinsey Global Institute analysis

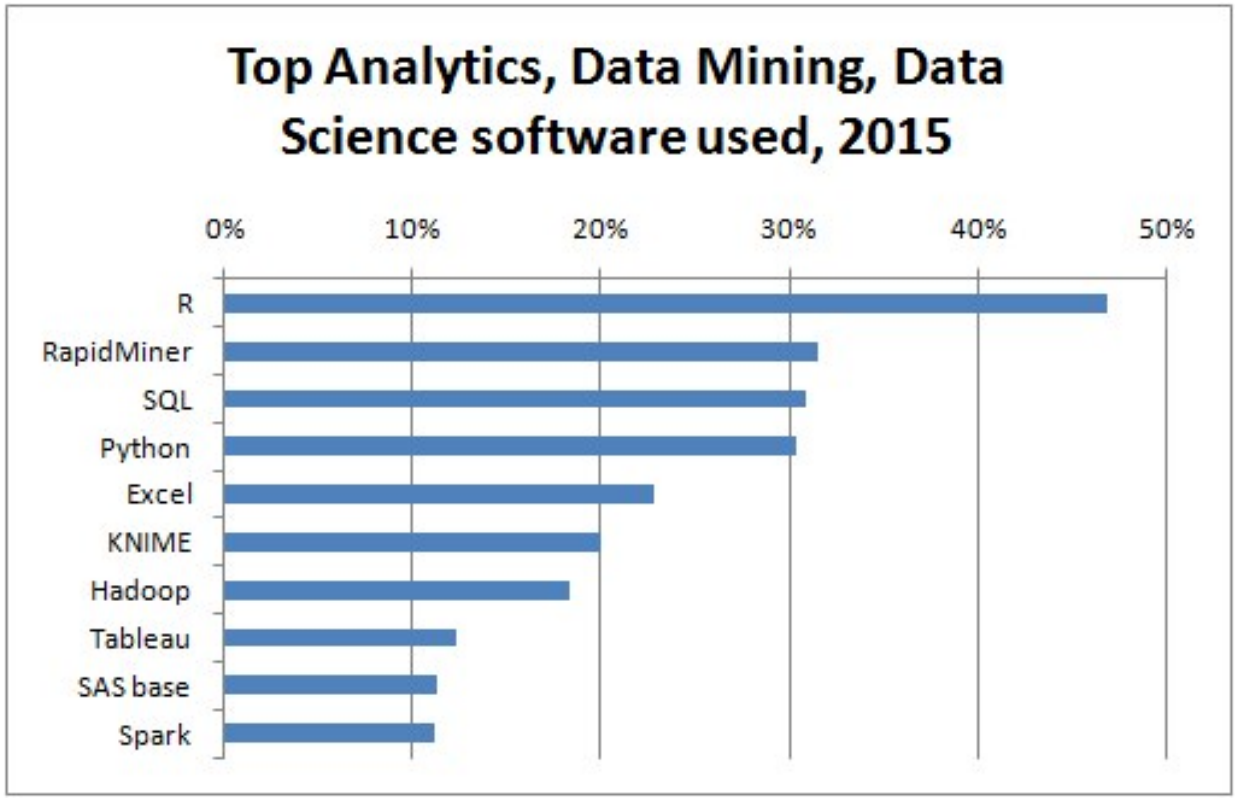
Conclusion:

- Growth of natural resources industry stagnated;
- Big data in natural resources industry has great potential.

Typical Algorithms in Data Analytics

- Supervised learning
  - Regression: LASSO, Decision tree, PLS, MLR
  - Classification: Logistic regression
  - Hybrid: Gaussian Process, Neural Network, SVM/SVR
  - …
- Unsupervised learning
  - Dimension Reduction: PCA
  - Clustering: k-means
  - …
- Inference
  - Maximum Likelihood, Expectation Maximization
  - Bayesian Method, Variational Bayesian, Bayesian Network
  - …

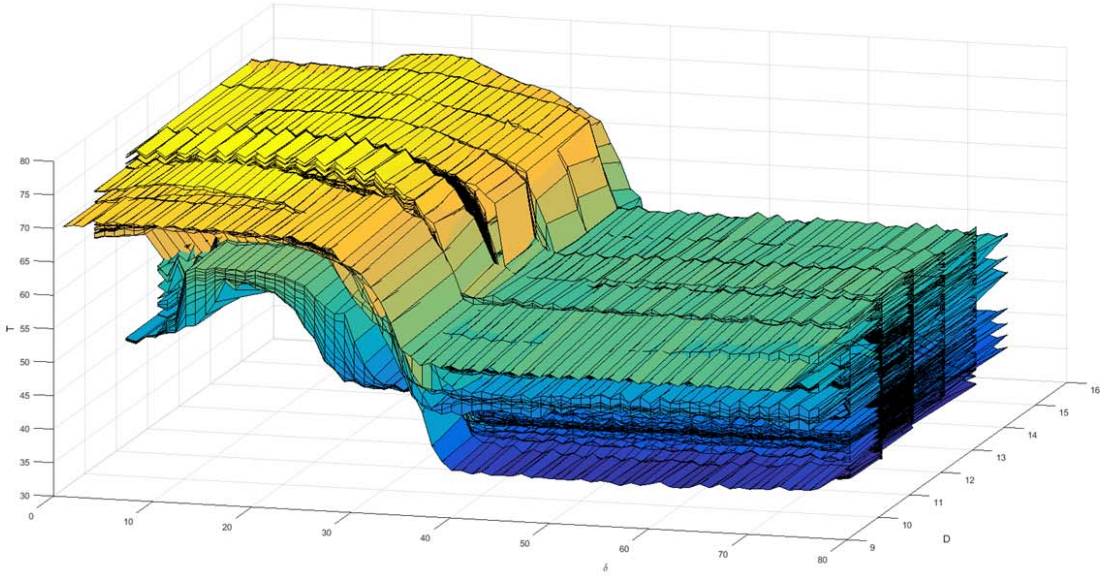
Data Analytics Software Platform and Toolboxes

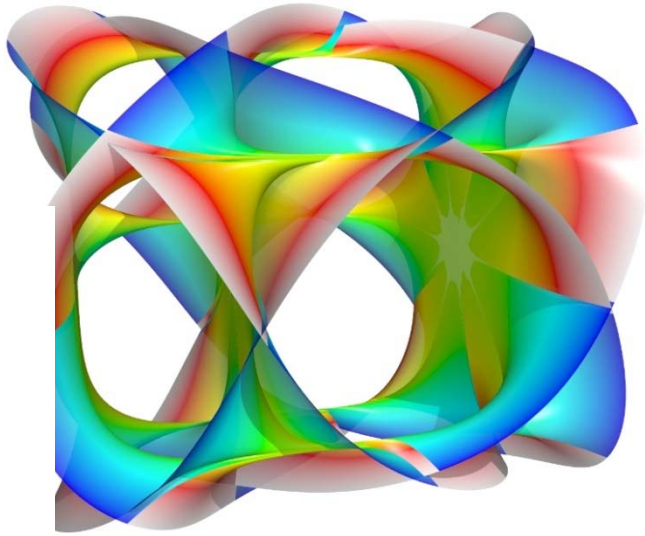


http://www.kdnuggets.com/polls/2015/analytics-data-mining-data-science-software-used.html

Process Data Analytics(Engineering)

High Dimensionality

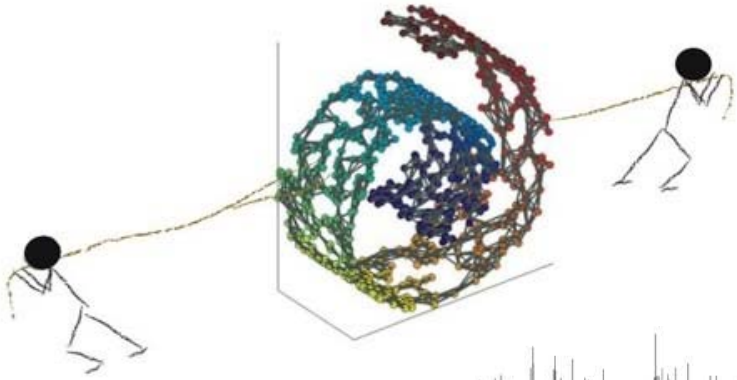




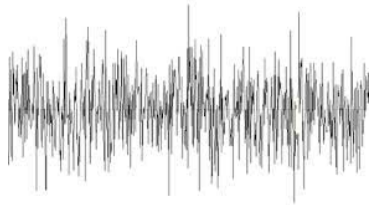
High Dimensionality of Data - “Decoding”

PCA/PLS/ICA and Applications:

•Dimensionality Reduction



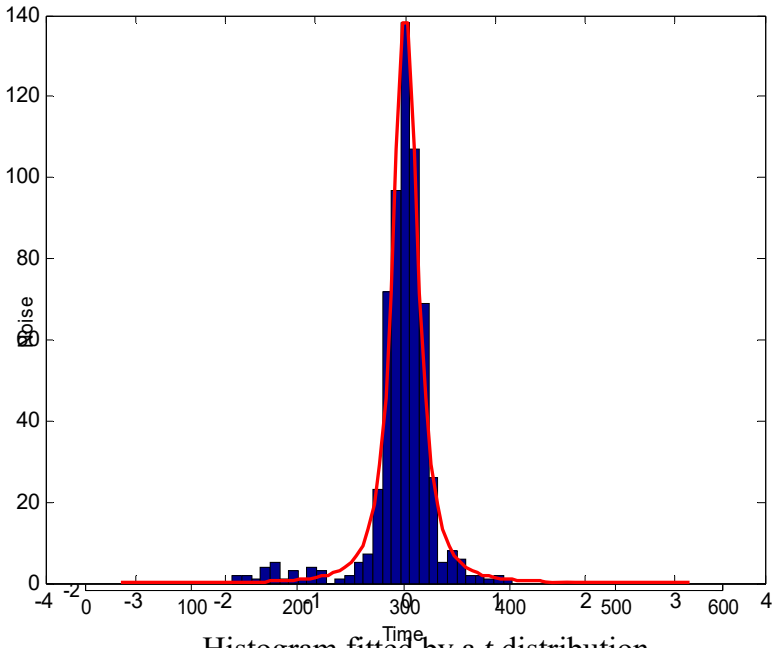
•Pre-processing for many data mining tasks (Noise Reduction)



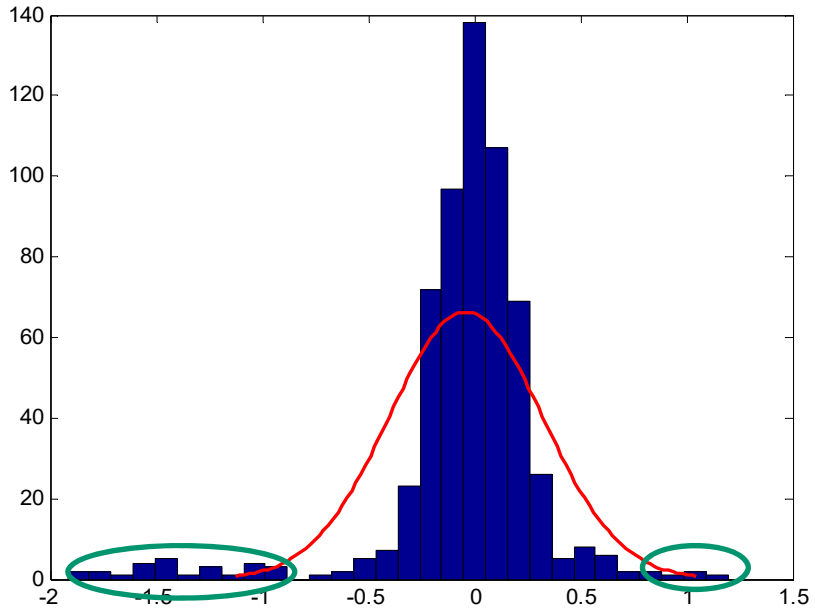
•Analyze data and to find patterns

Robustness – Dealing with Irregular Data

- Modeling the noise: Gaussian distribution vs others

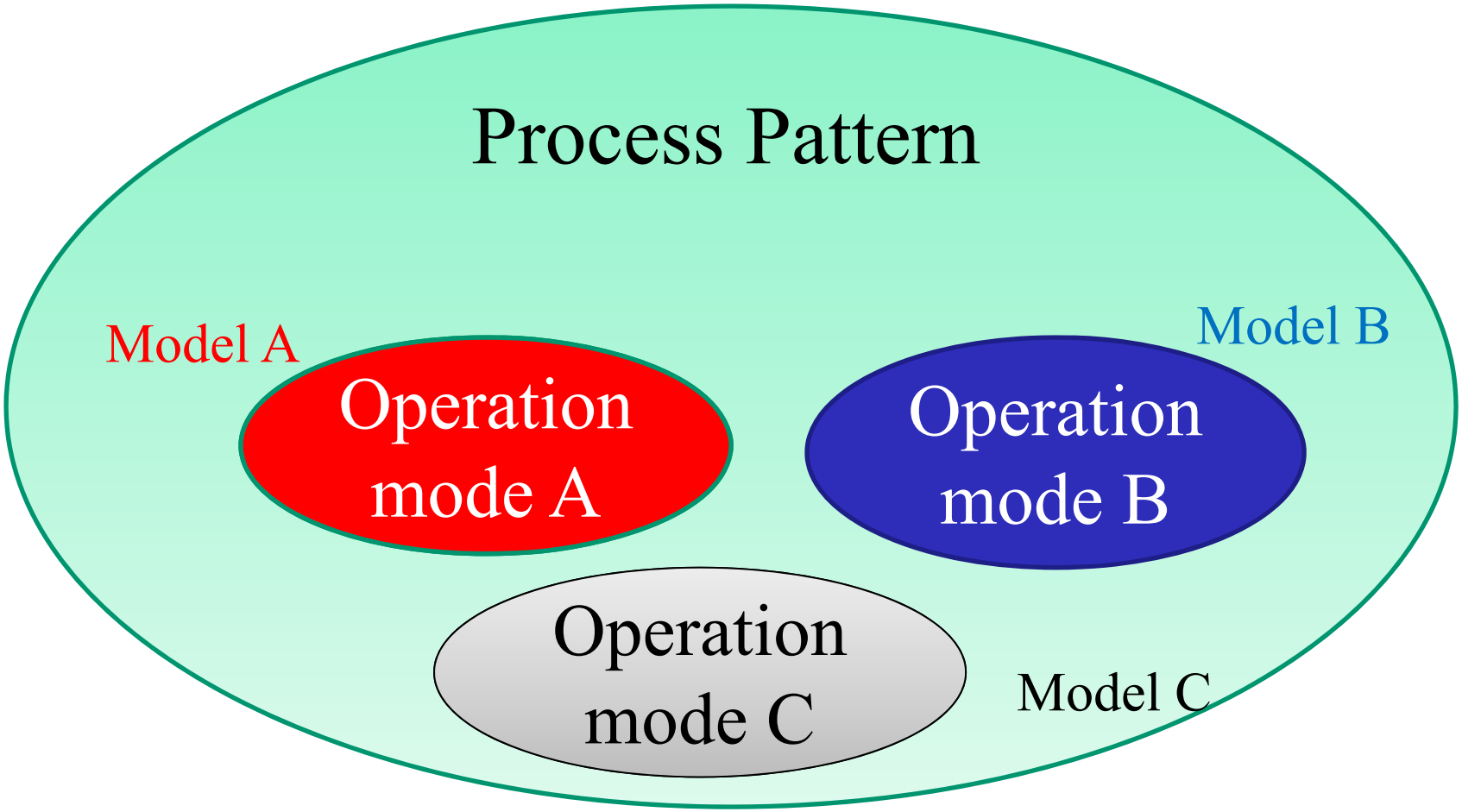


Histogram fitted by a *t* distribution



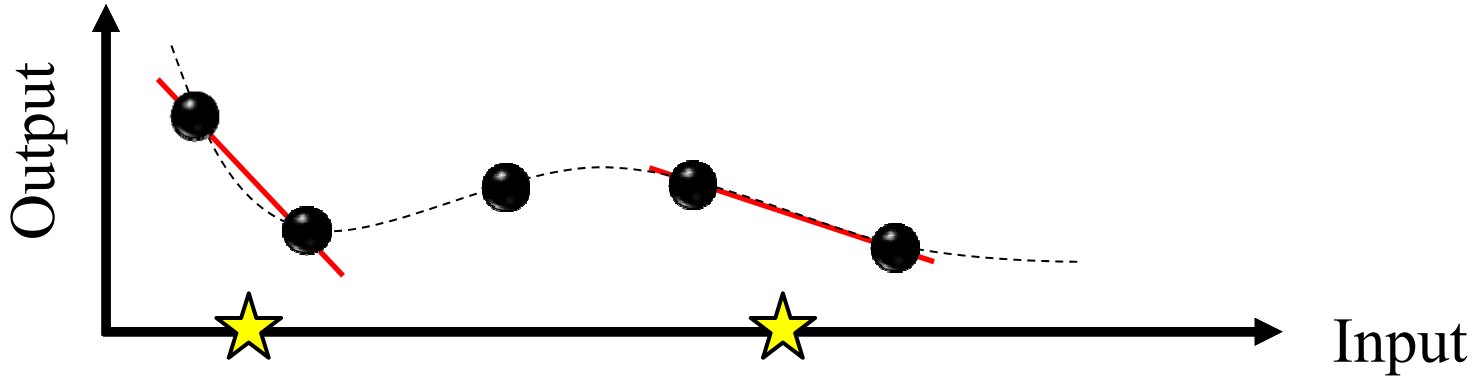
Histogram fitted by a Gaussian distribution

Multimodality



Nonlinearity – Local Solution

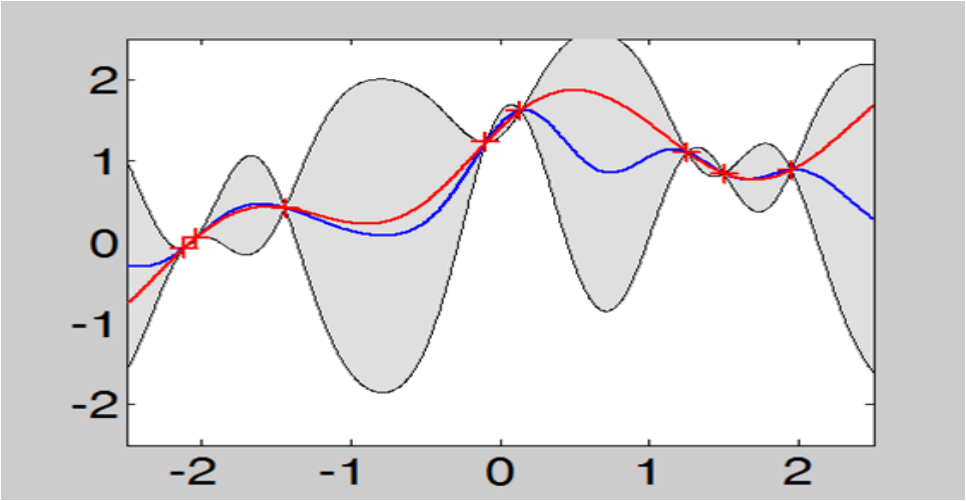
JIT modeling
= Locally weighted modeling
= Relevance-In-Space modeling
= Lazy modeling



Nonlinearity – Gaussian Process

$$p(\boldsymbol{y}|\boldsymbol{X}, l, \sigma^2) = \int p(\boldsymbol{y}|\boldsymbol{f}, \boldsymbol{X}, \sigma^2) p(\boldsymbol{f}|\boldsymbol{X}, l)\, d\boldsymbol{f}$$

Analytical expression of likelihood exists



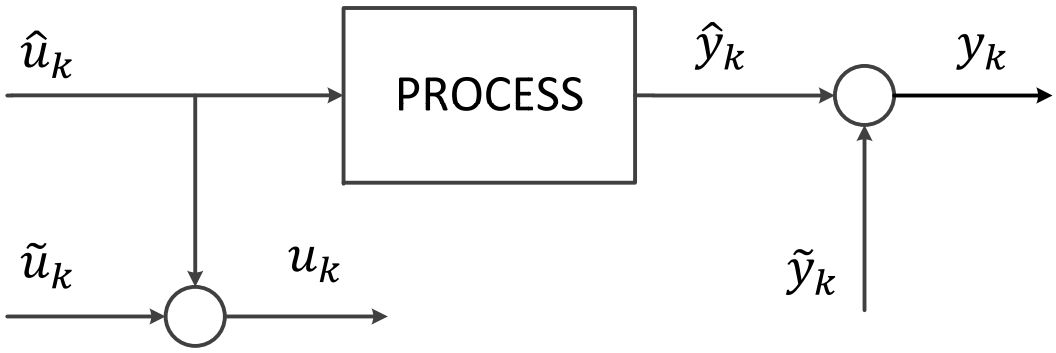
Time varying time delays

Multirate sampling with time Varying time-delays:

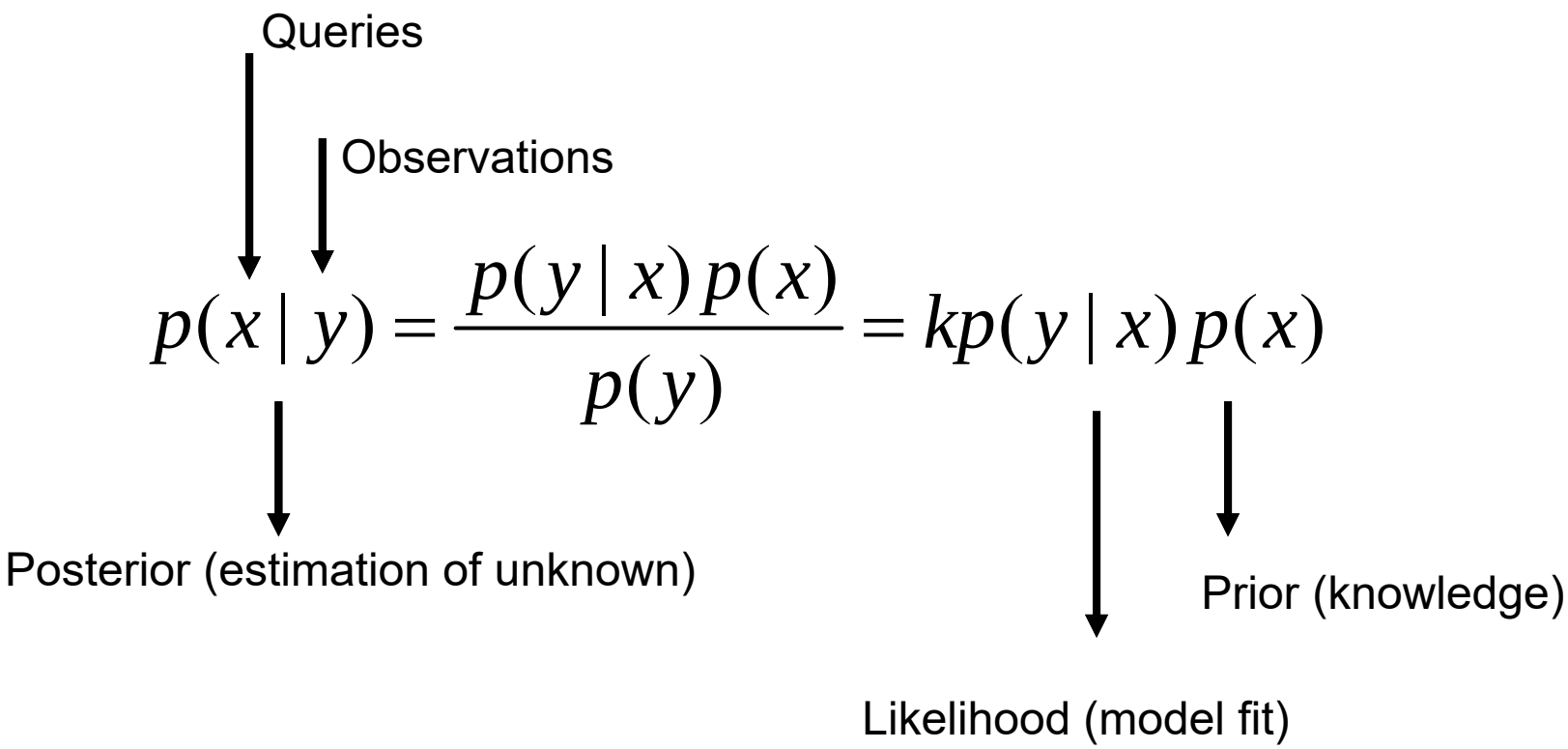
- Dual rate: fast rate input while slow rate output
- Time delay is varying at every sample

Errors-In-Variables (EIV)

- Noise-corrupted measurements: u_k ,y_k
- Additive noise: $\tilde{u}_k$, $\tilde{y}_k$
- Unknown noise-free input and output: $\hat{u}_k$,$\hat{y}_k$



Process Knowledge - Bayes Methods



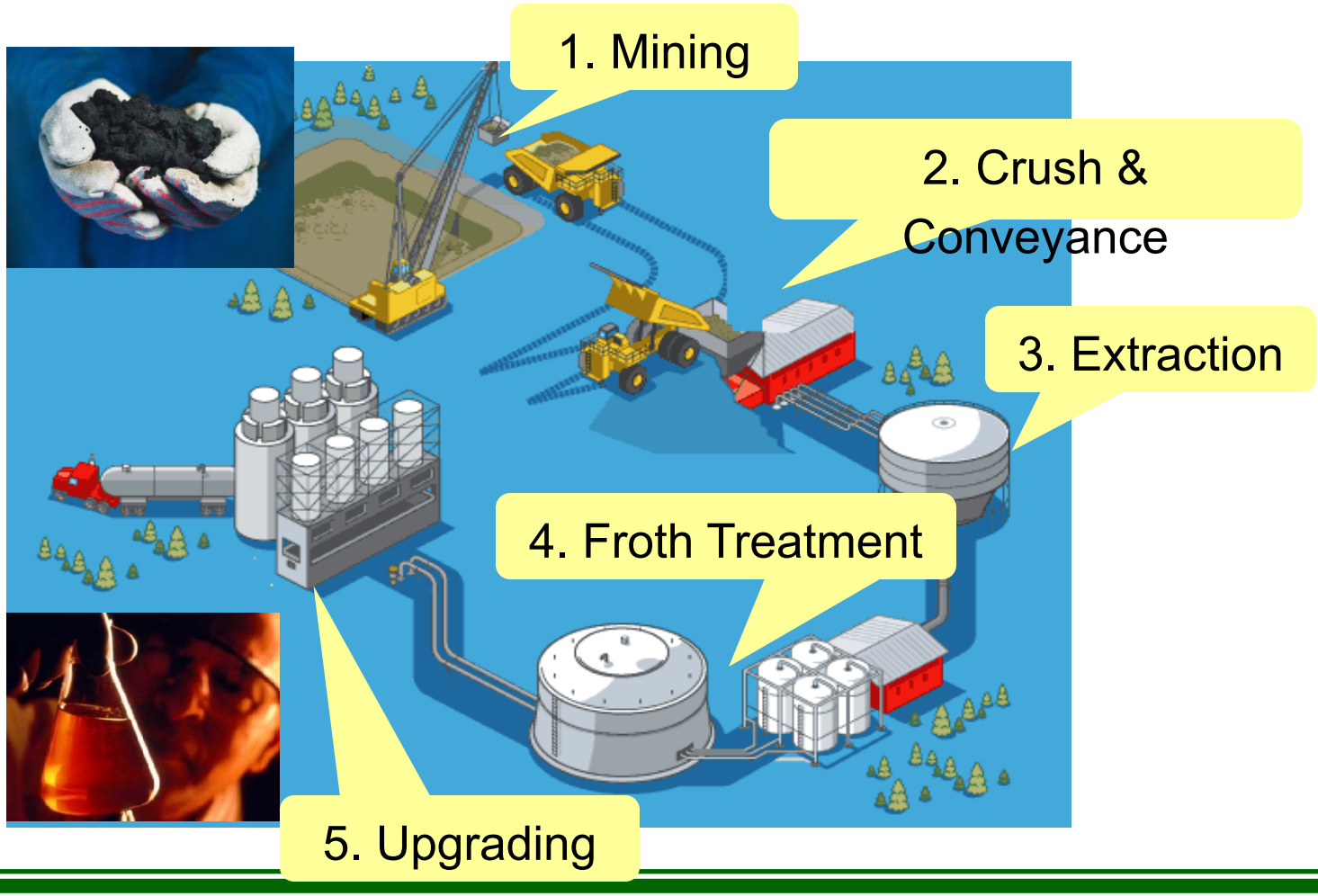
Oil Sands

Canada's Oil Sands

- 141,000 square kilometres deposit
- 1.7 trillion barrels of bitumen
- 170 billion barrels recoverable
- second largest oil reserve
- 1.3 million barrels crude oil per day

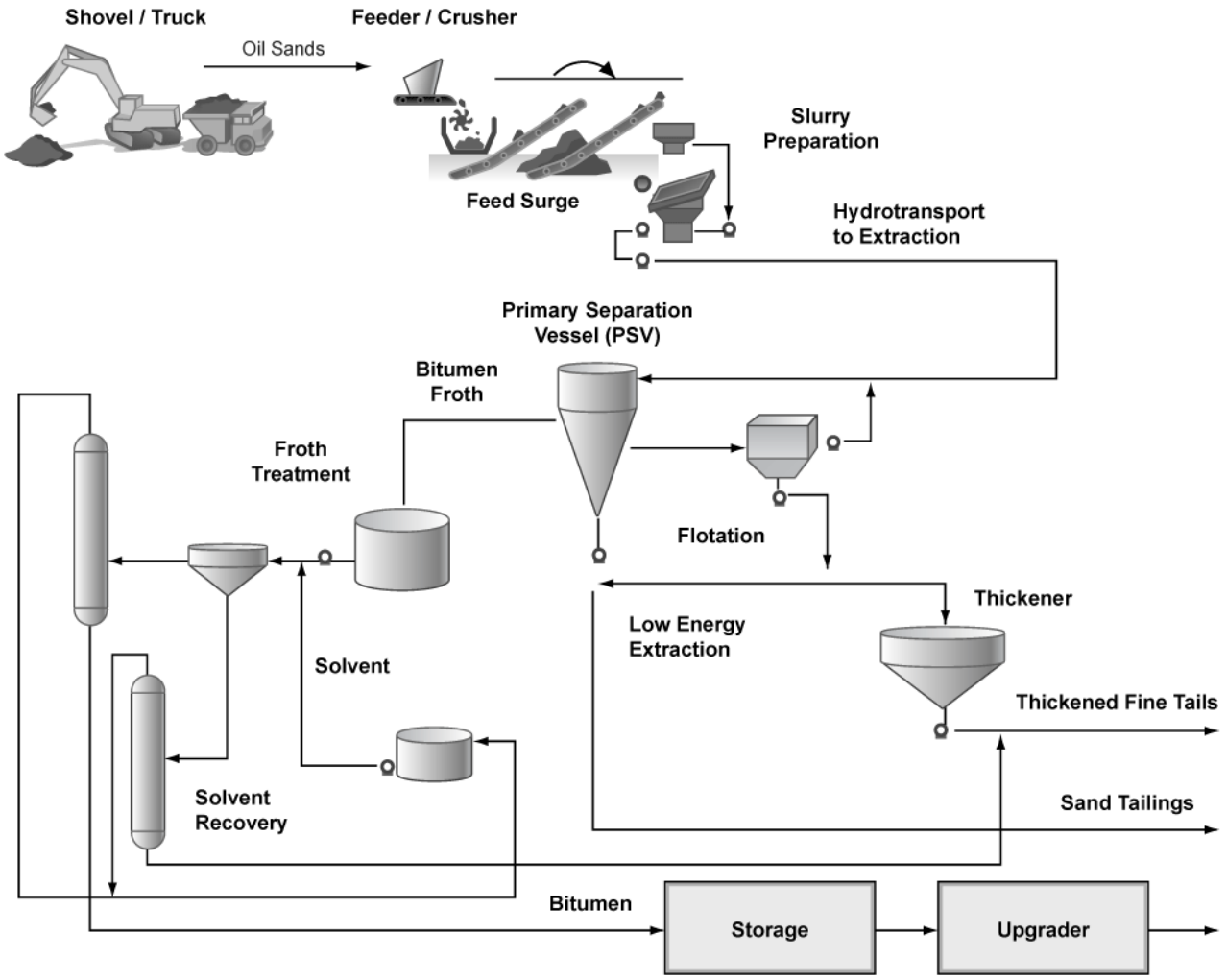
Source: http://en.wikipedia.org/wiki/Athabasca_oil_sands

From Oil Sands to Sweat Crude Oil

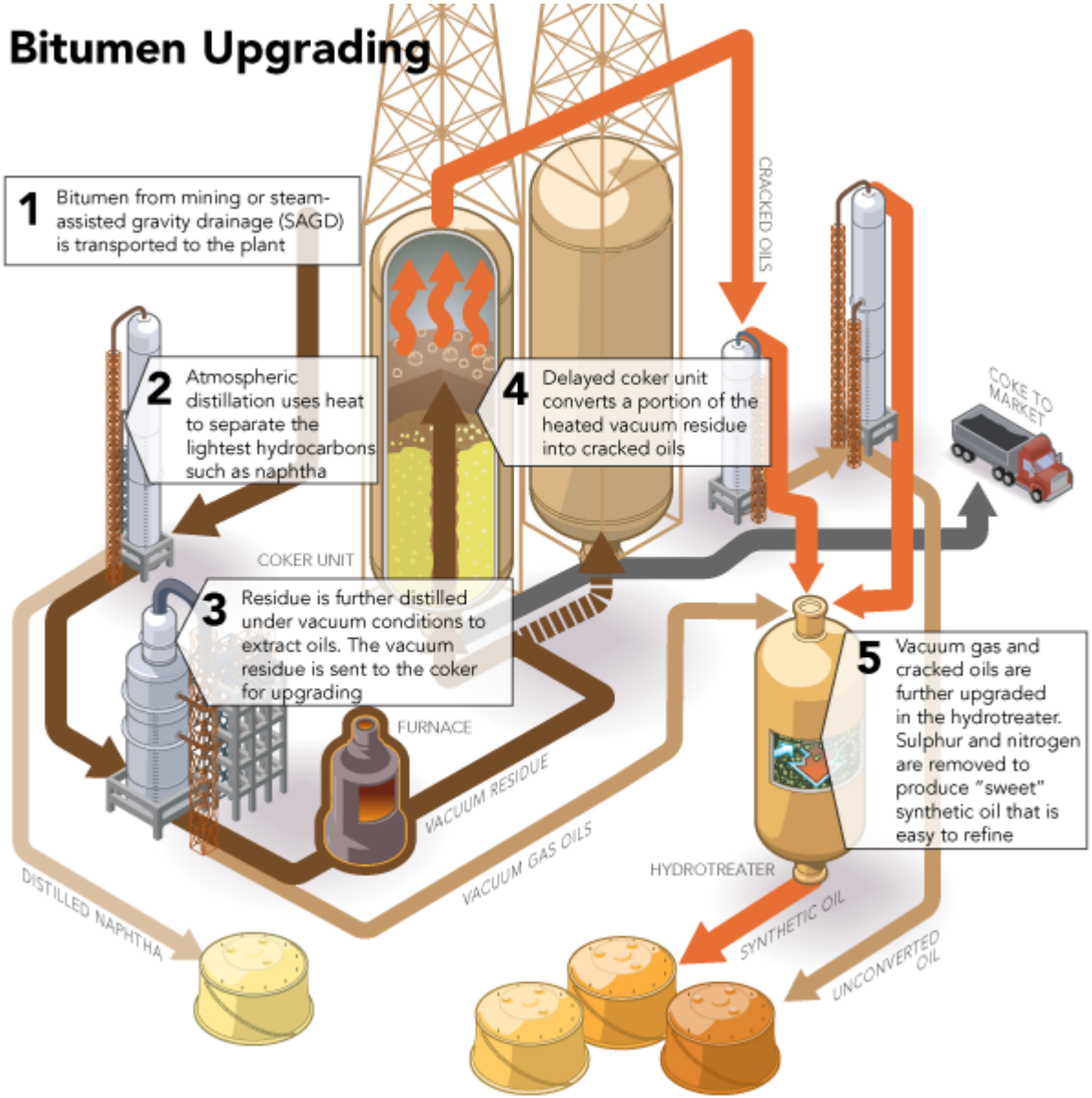


Source: http://www.ems.psu.edu/~pisupati/ACSOutreach/Oil_Sands.html

Extraction

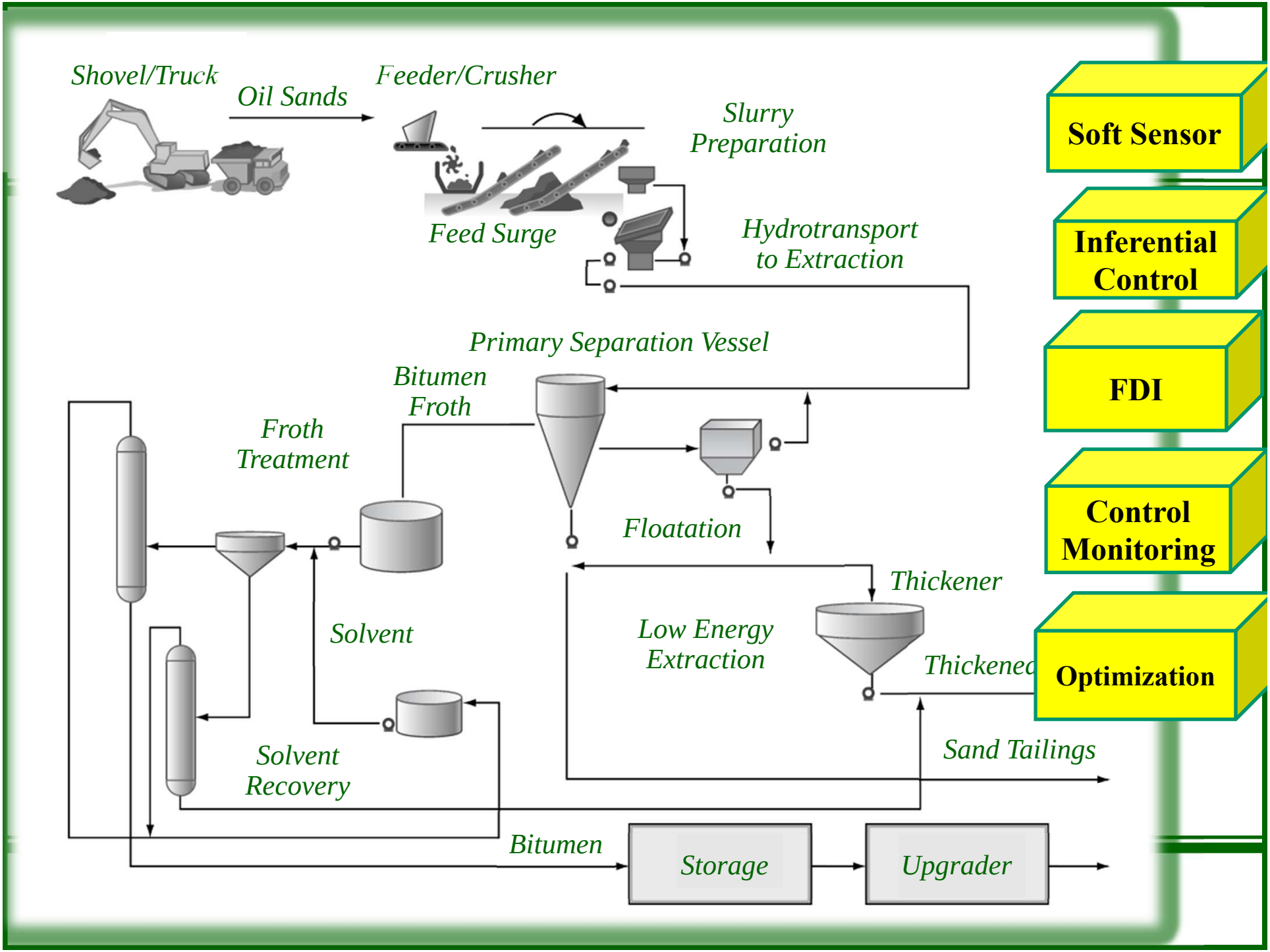


Source: http://www.ontime.methanetomarkets.org/m2mtool/oilsands.html

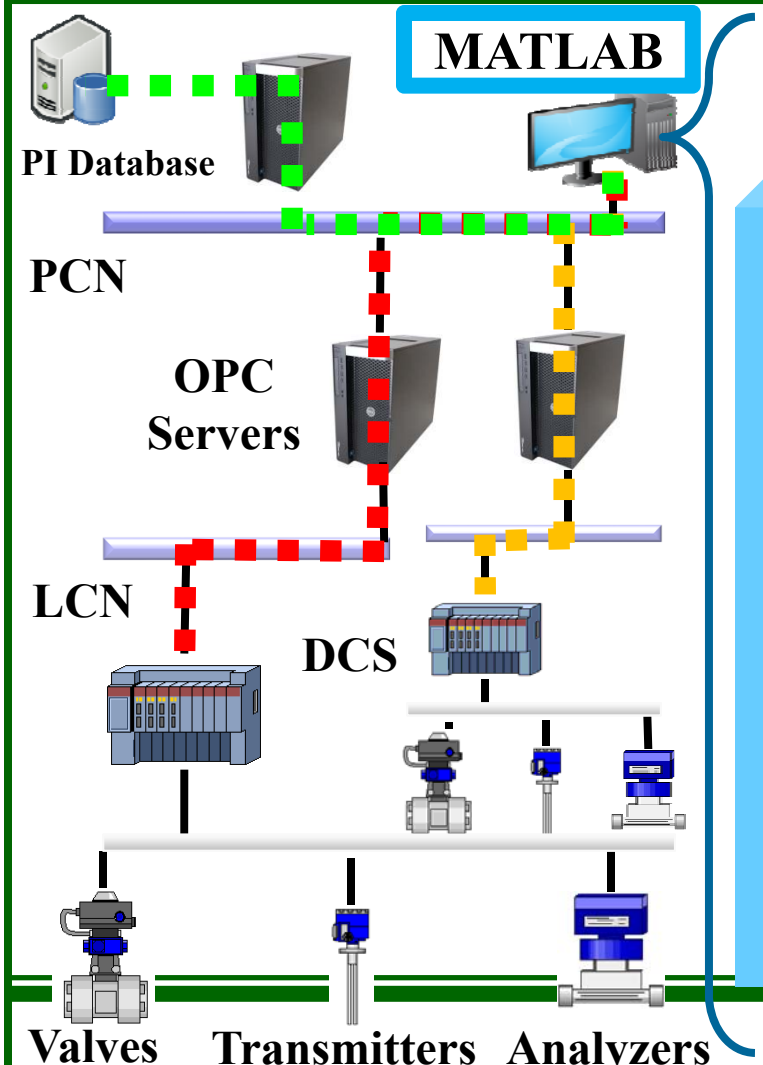


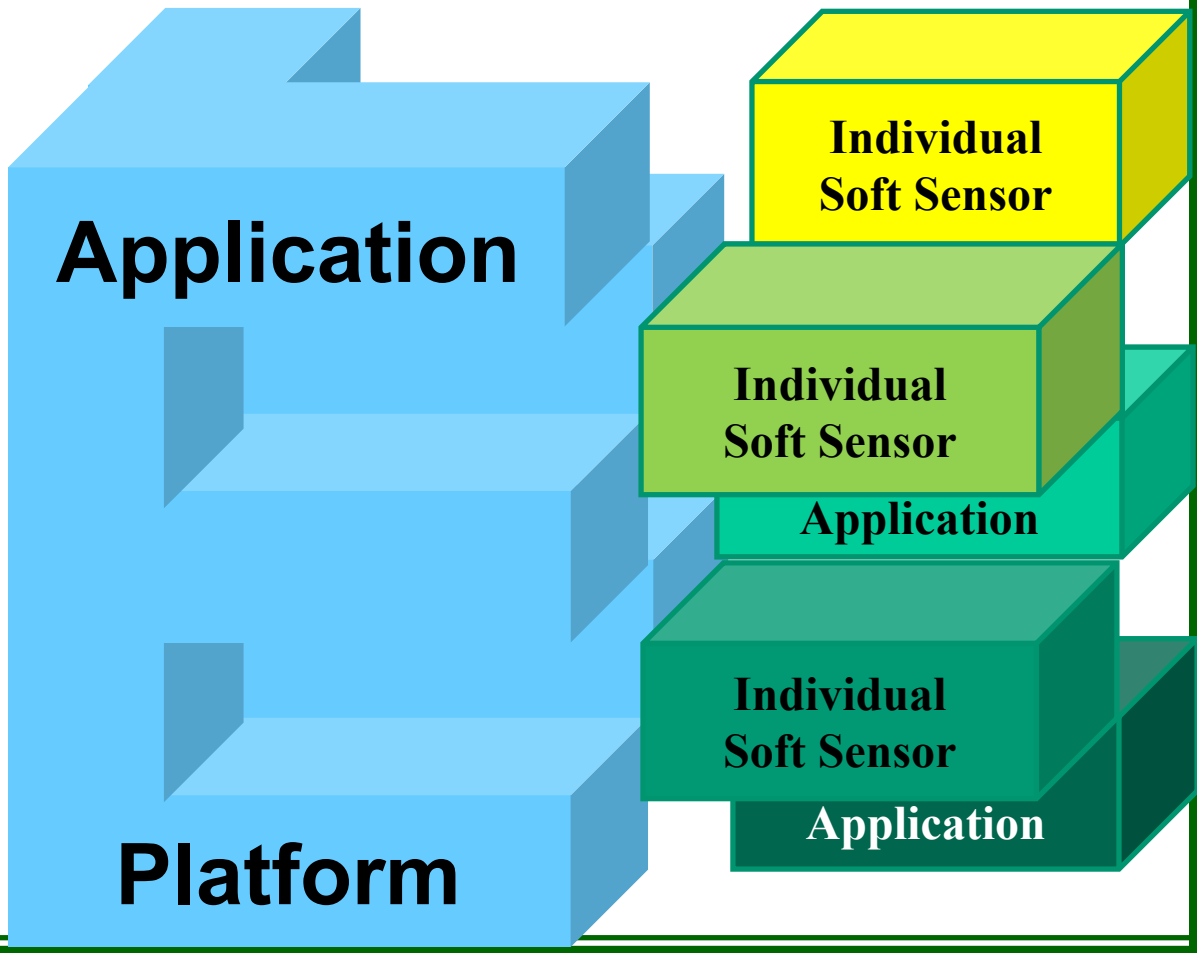
Source: http://www.bantrel.com/markets/downstream.aspx

Process Data Analytics in Oil Sands

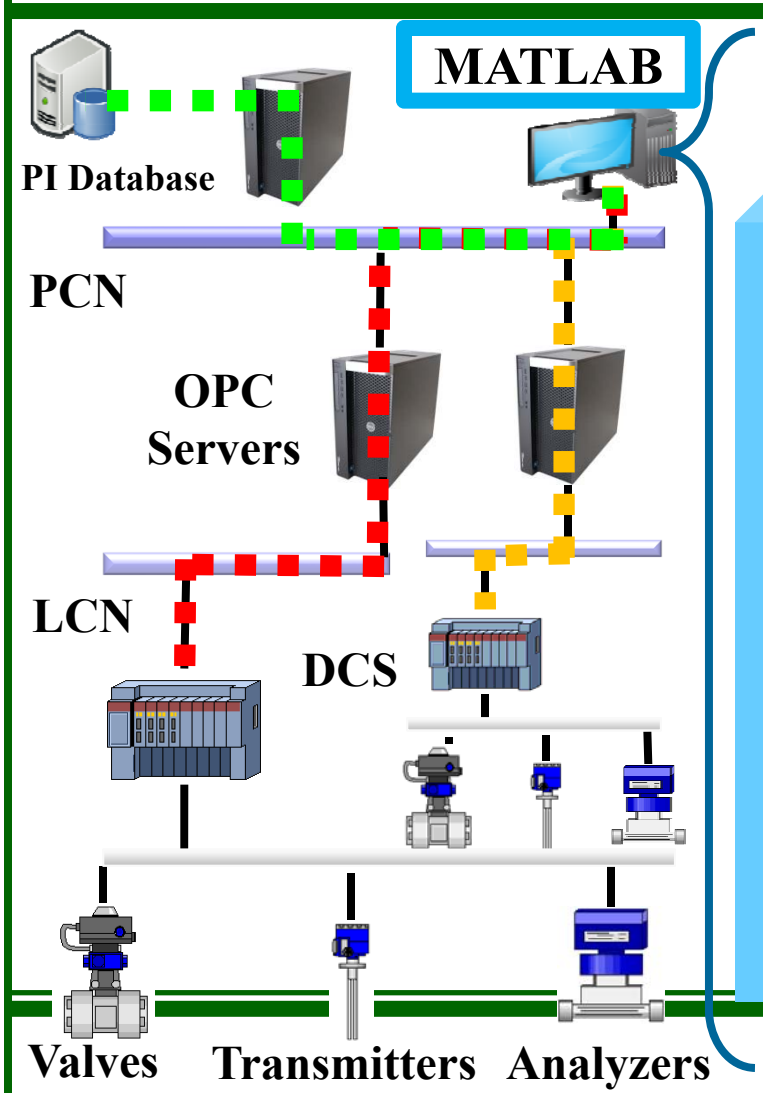

Shovel/Truck
Oil Sands
Feeder/Crusher
Slurry Preparation
Feed Surge
Hydrotransport to Extraction
Primary Separation Vessel
Bitumen Froth
Froth Treatment
Floatation
Thickener
Low Energy Extraction
Thickened
Solvent
Solvent Recovery
Sand Tailings
Bitumen
Storage
Upgrader
Soft Sensor
Inferential Control
FDI
Control Monitoring
Optimization

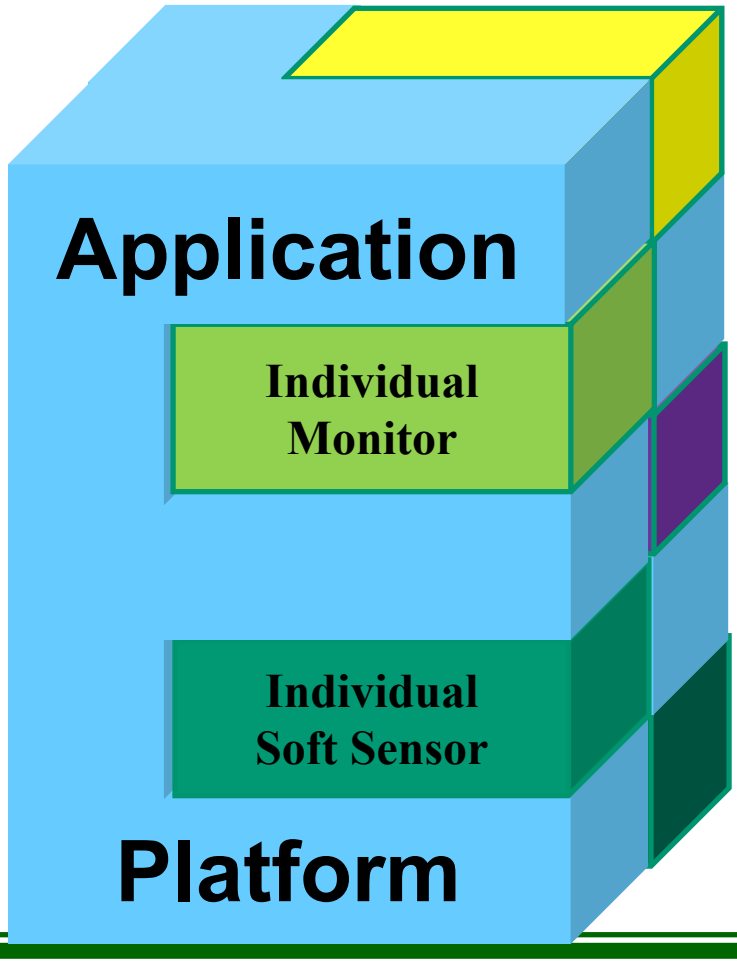
Rapid Technology Transfer Platform





Application Platform



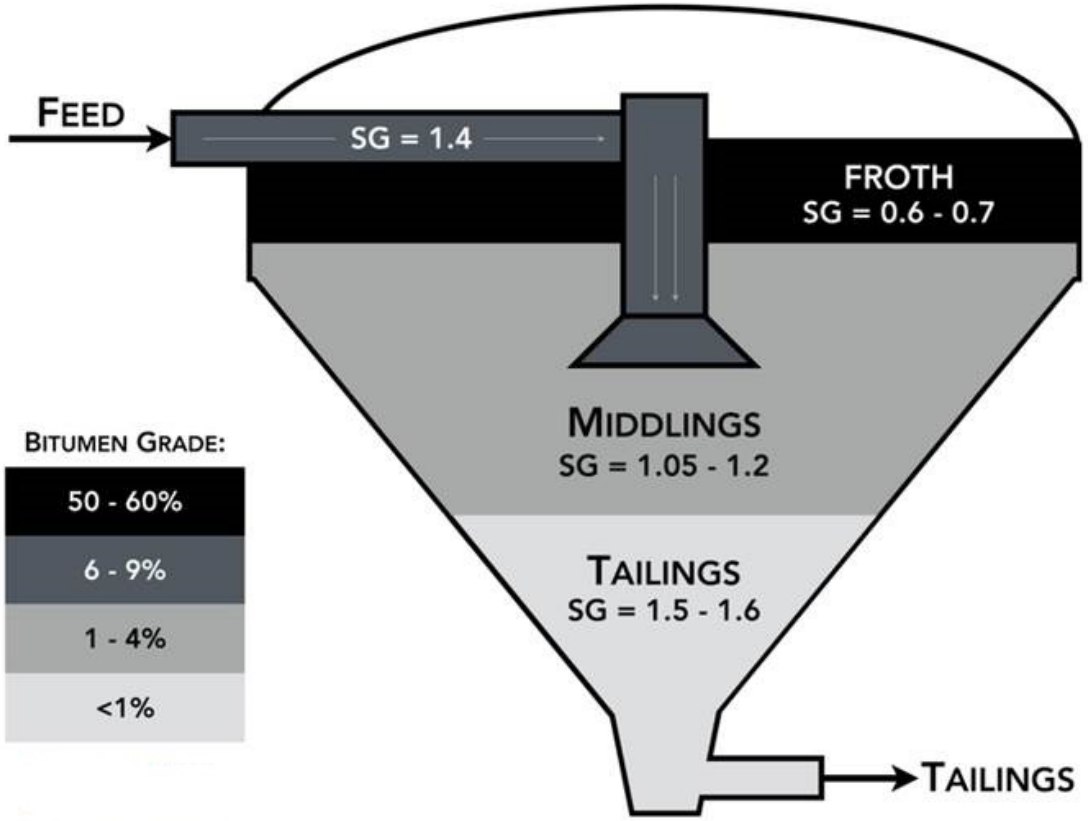


- Only One Platform
- Multiple OPC
- Multiple Individual Applications

- ❑ Rapid Technology Transfer
- ❑ Efficient Implementation
- ❑ Convenient Maintenance

Data Analytics in Image Processing

Introduction of PSV



- Three layers due to density difference
- Froth/Middling interface level is the **most important** control variable

Image source: http://www.oilsandsmagazine.com/oilsands-primary-extraction-gravity-separation-process-bitumen-production/

PSV interface measurements

Interface measurements
- Density profiler
- D/P Cell
- Camera

↓

Best performance, however, only reliable when the performance index (PI) is higher than a given constant.

Objective: Improve the camera reading reliability when the performance index is low.

↓

Image Analysis

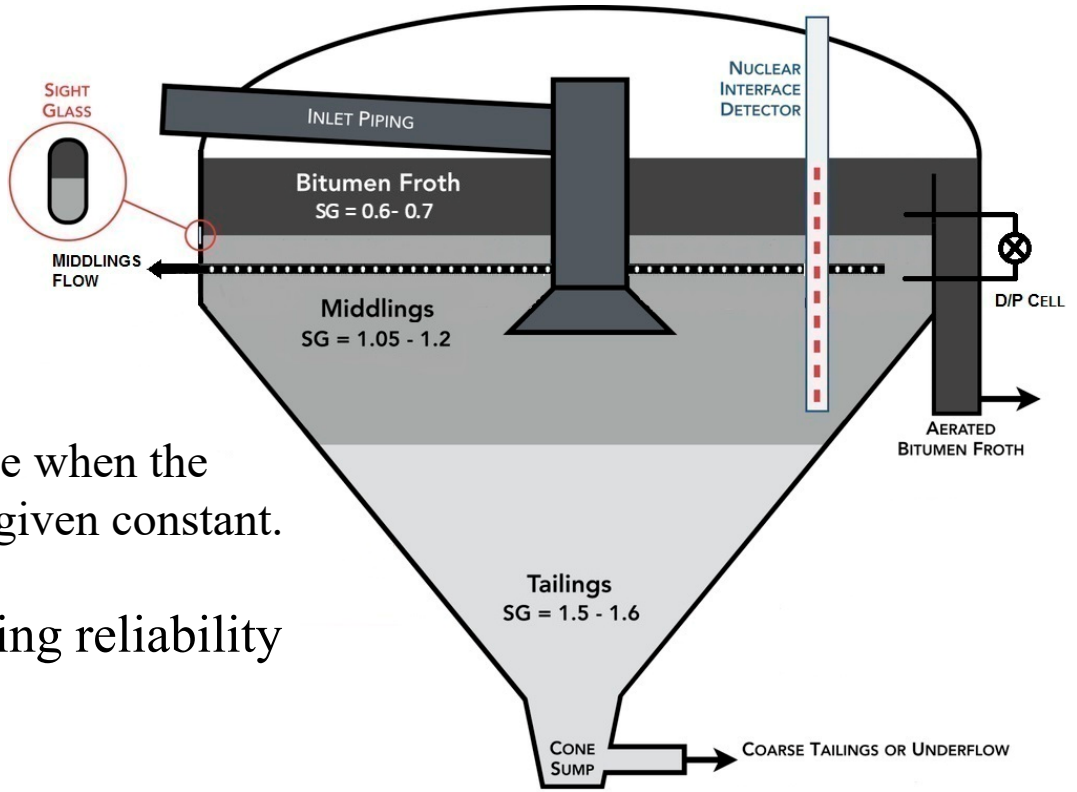
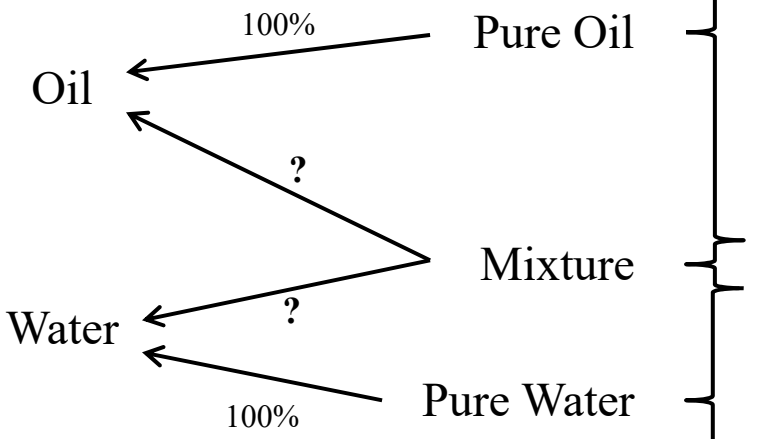


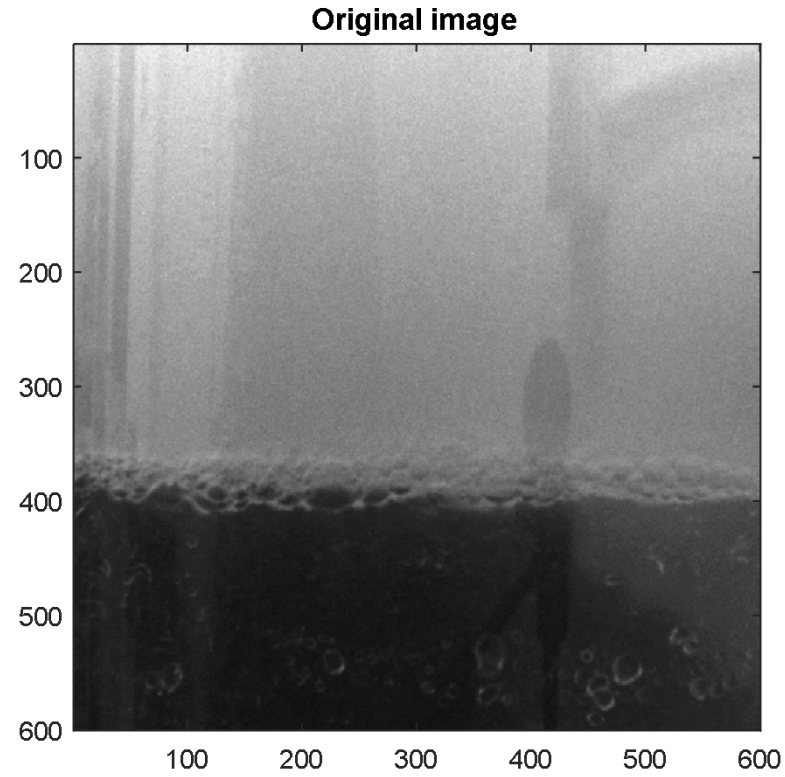
Image source: http://www.oilsandsmagazine.com/oilsands-primary-extraction-gravity-separation-process-bitumen-production/

Experimental Design

Image Captured by Camera

- The image shown on the right is the original image observed by the camera.





Objective: Segment the captured image as a binary image (+1 for oil/-1 for water/ 0 for interface)

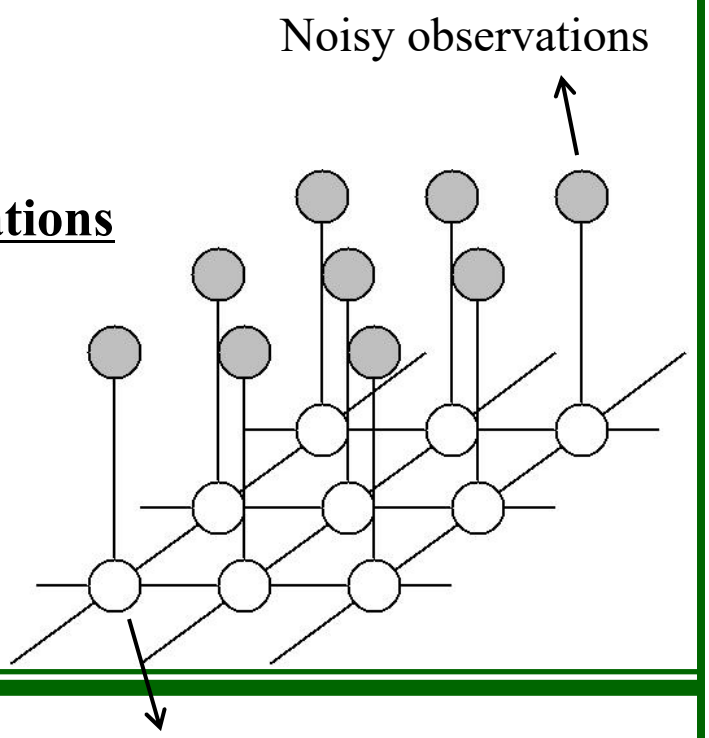
Note: the image size for all images is (600 pixels × 600 pixels)

Theory of Data Based Image Analysis

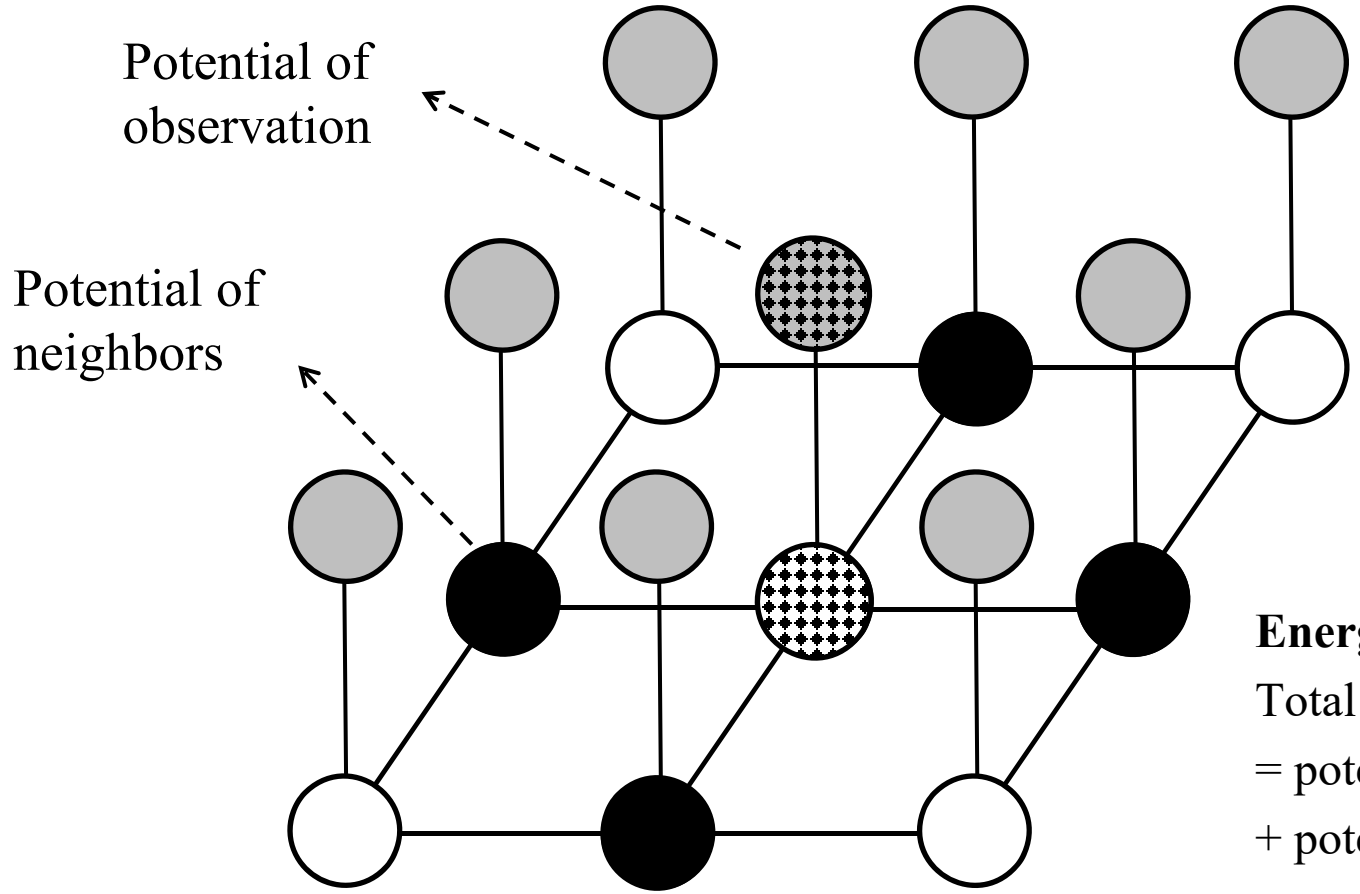
- Images can be modeled using Markov random field (MRF).
  - Each pixel is considered as a random variable (RV)
  - Each random variable (pixel) has a corresponding observation (corrupted with noise)

Aim: to recover clean pixels from noisy observations

MRF is employed to perform image segmentation and classification.

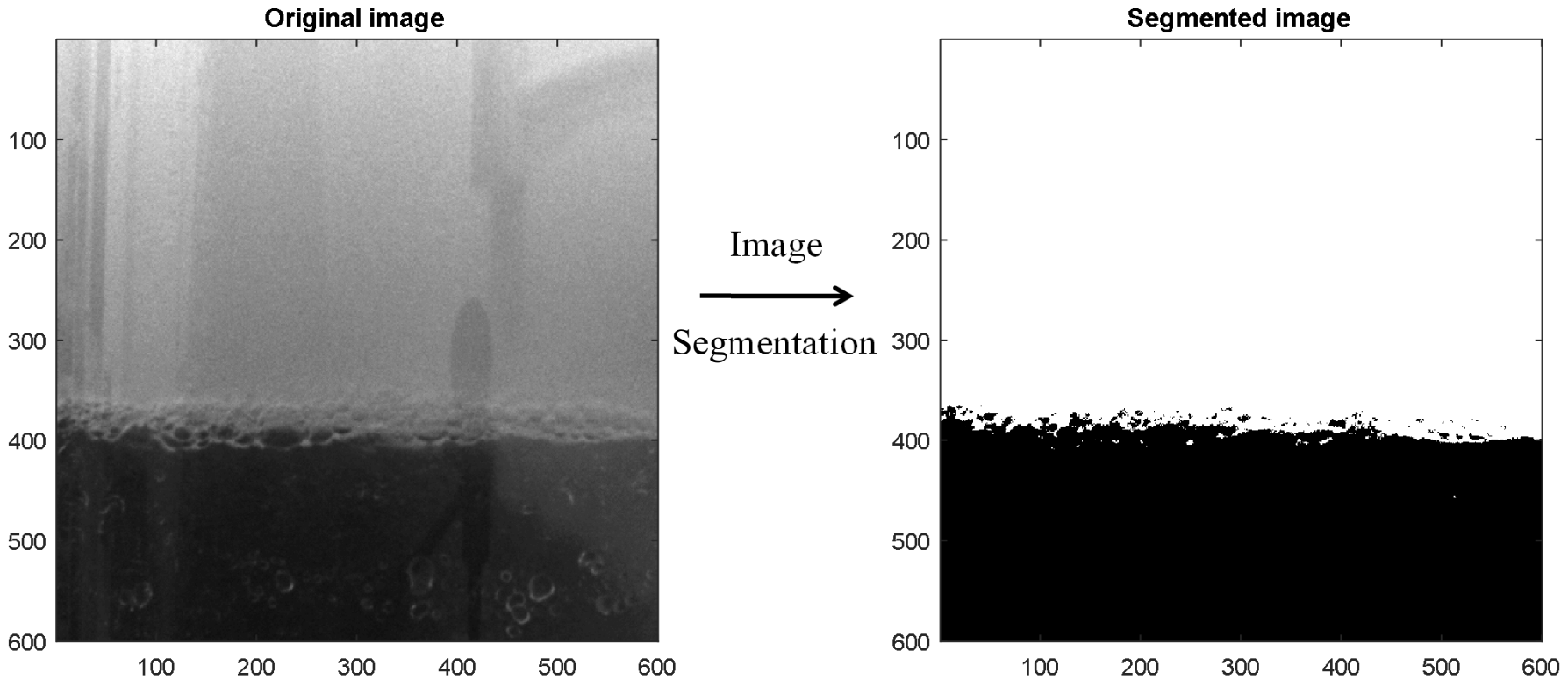


Principle of MRF Estimation

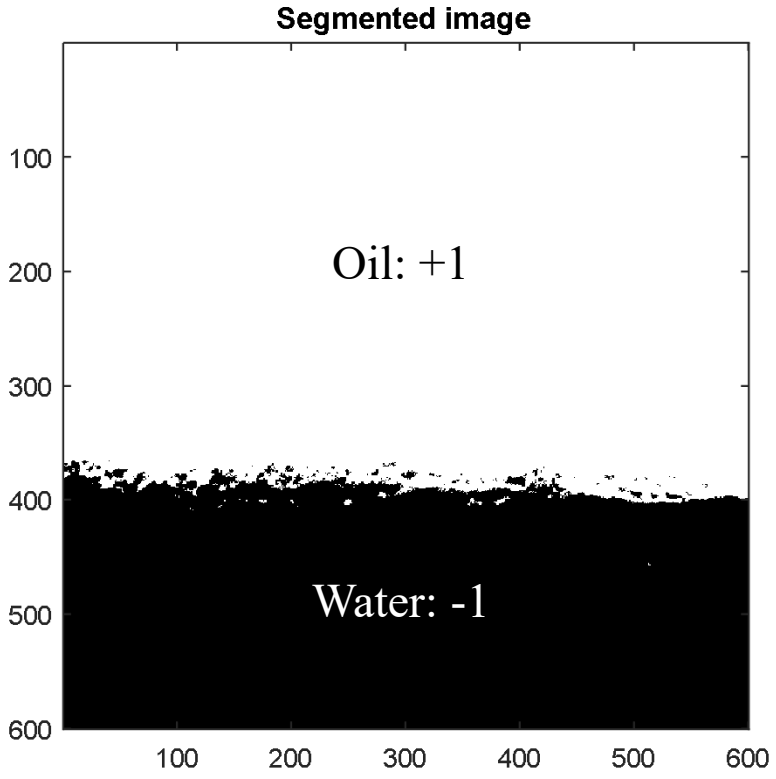


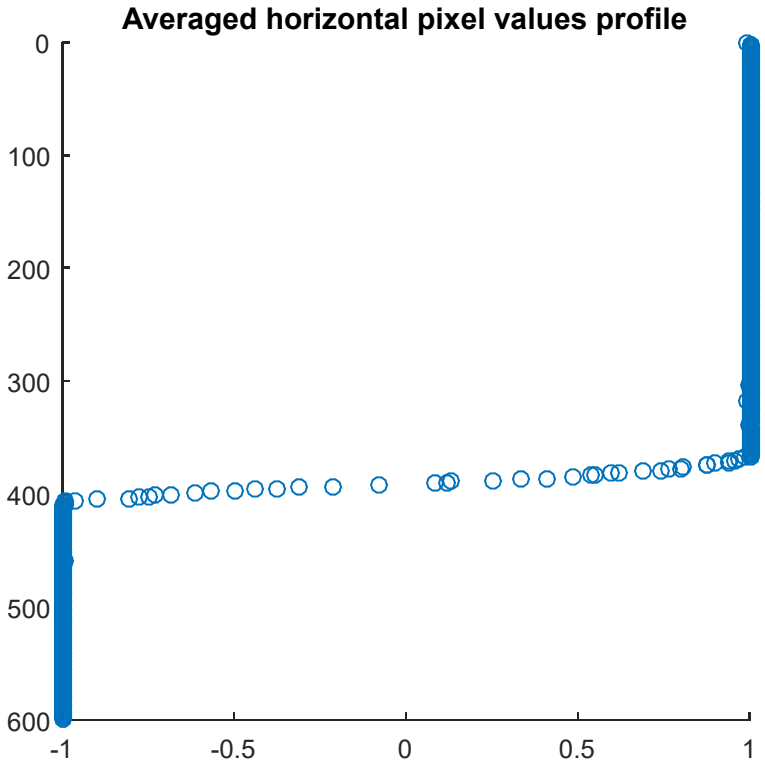
Energy minimization:
Total energy
= potential of observation
+ potential of neighbors

Image Segmentation

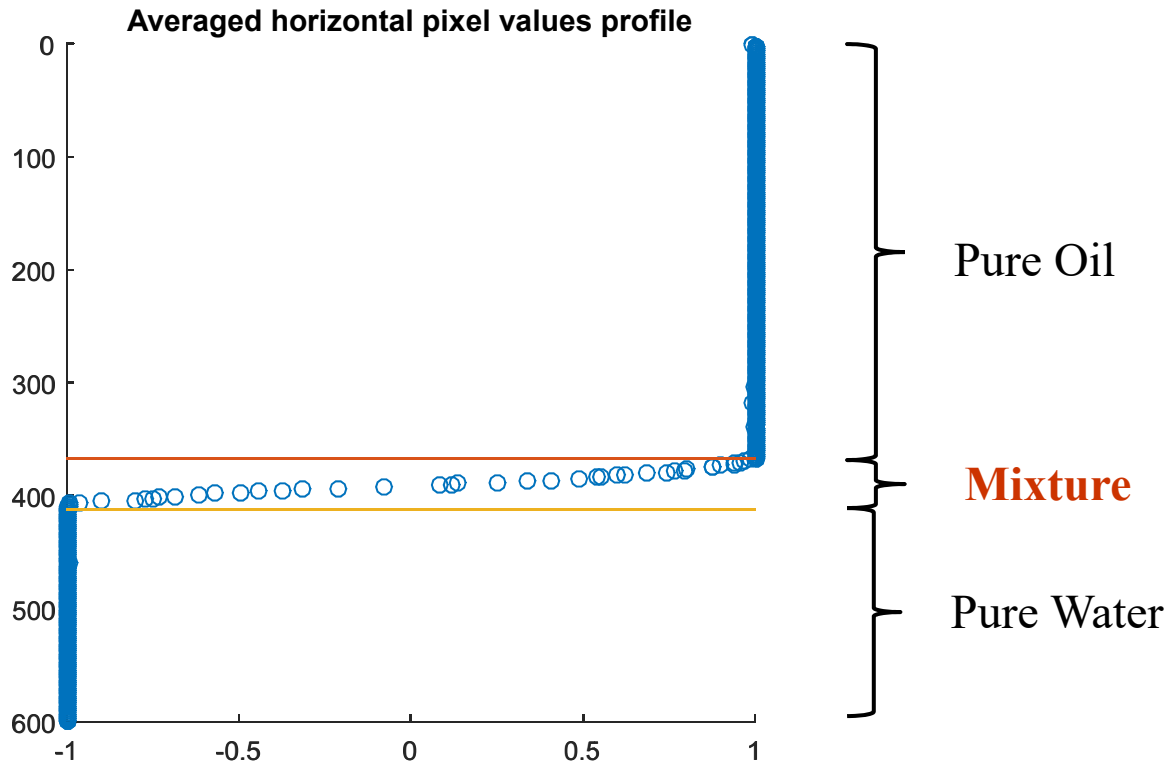


Pixel Values Profile

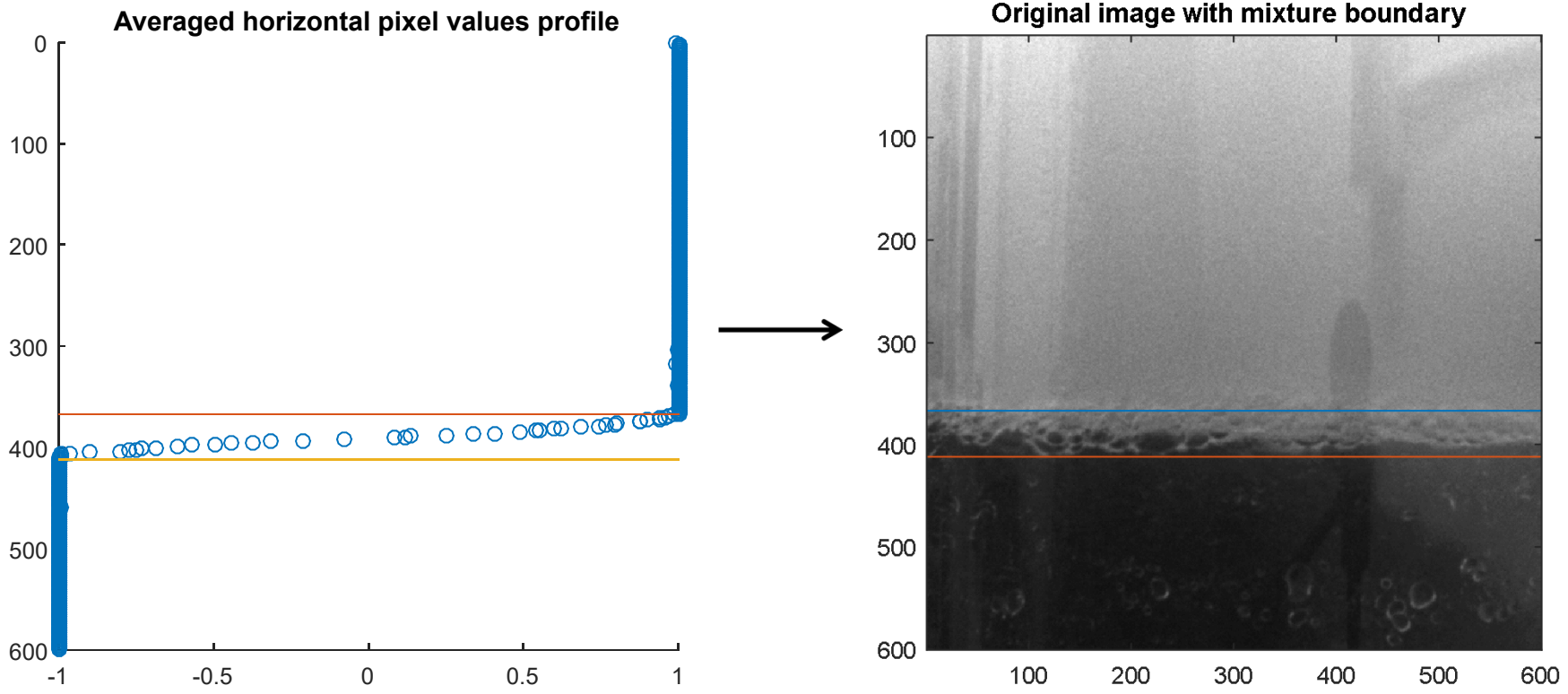




Mixture Boundary Determination

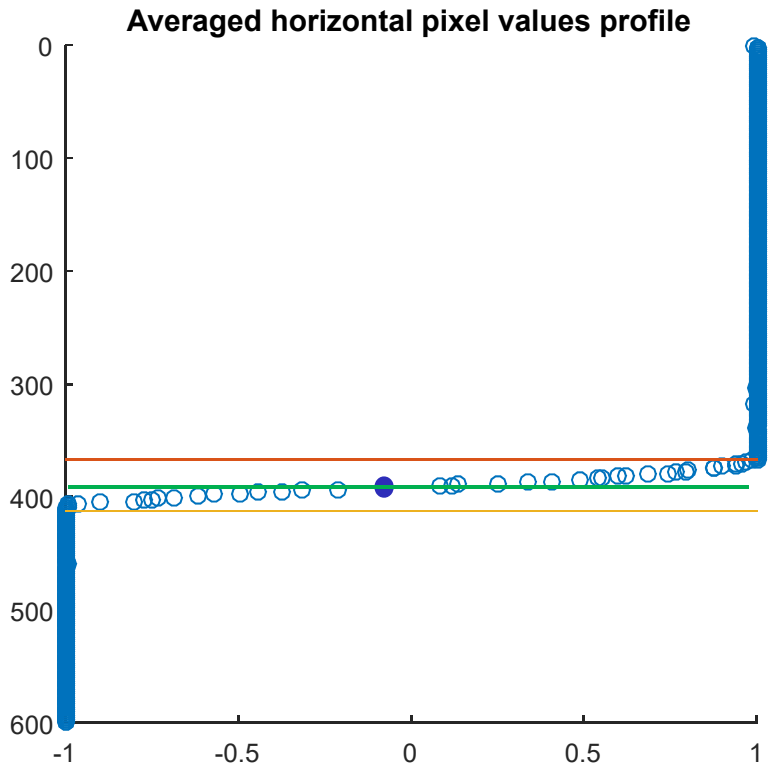


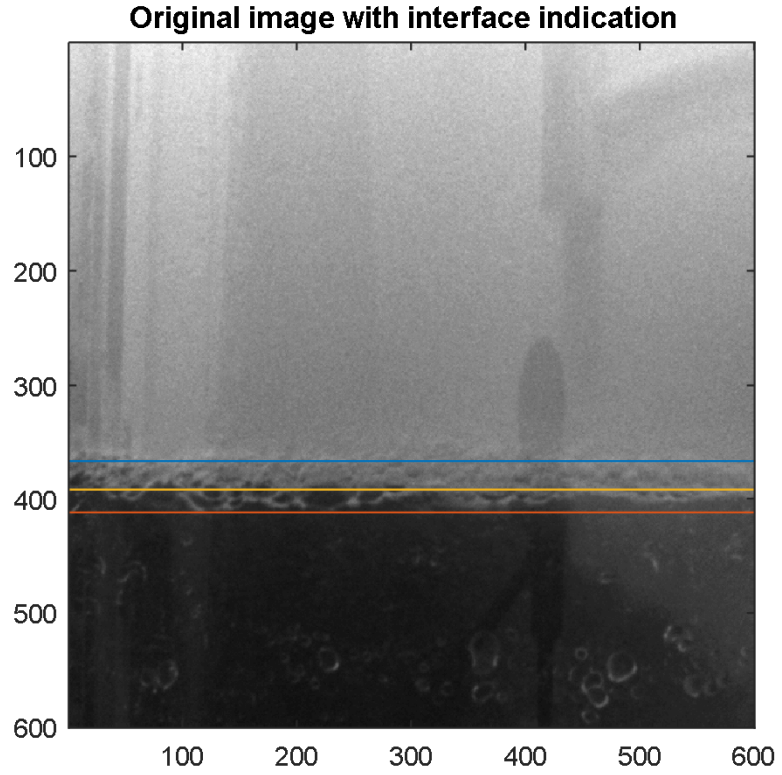
Mixture Boundary Indication



Next step: Identify the interface based on the pixel value close to zero

Interface Estimation



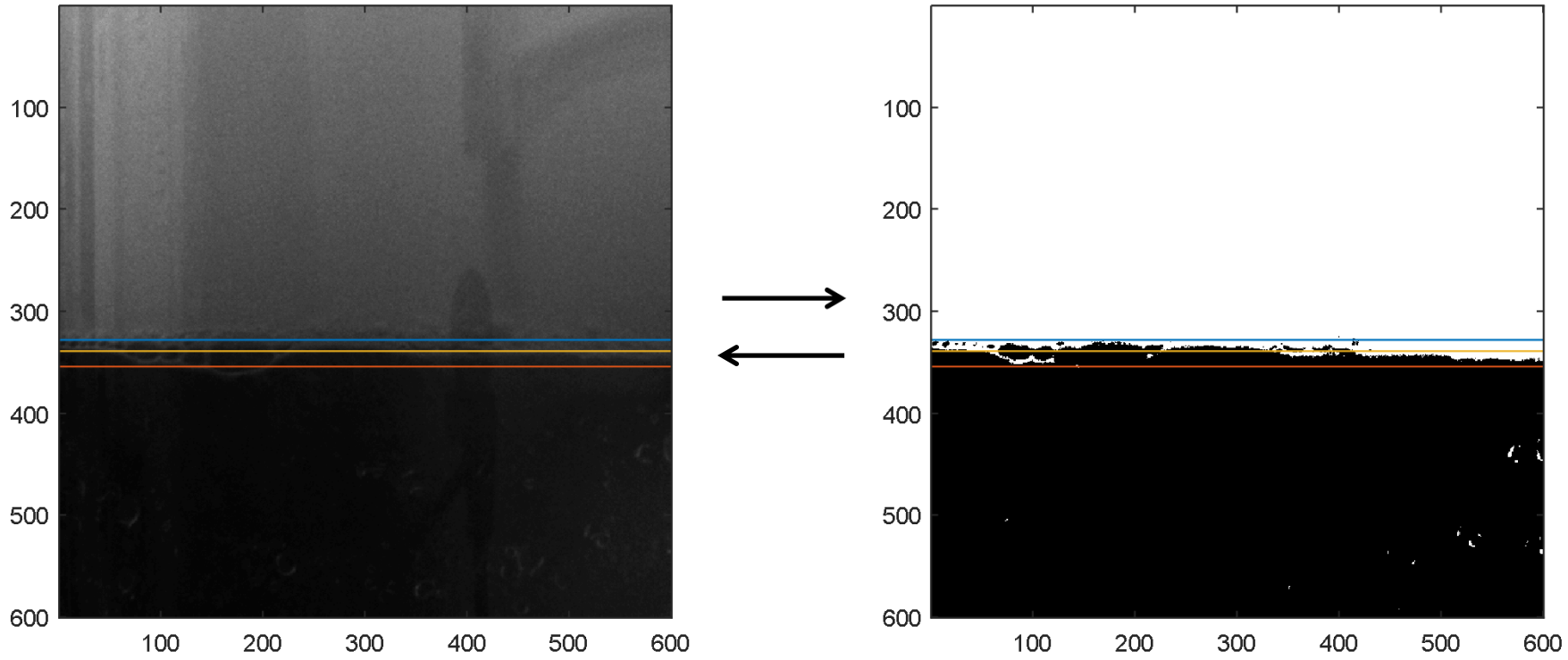


D/P Cell estimation: 35.04 cm

Image Processing estimation: 35.05 cm

} **0.2%**

Image under Different Condition

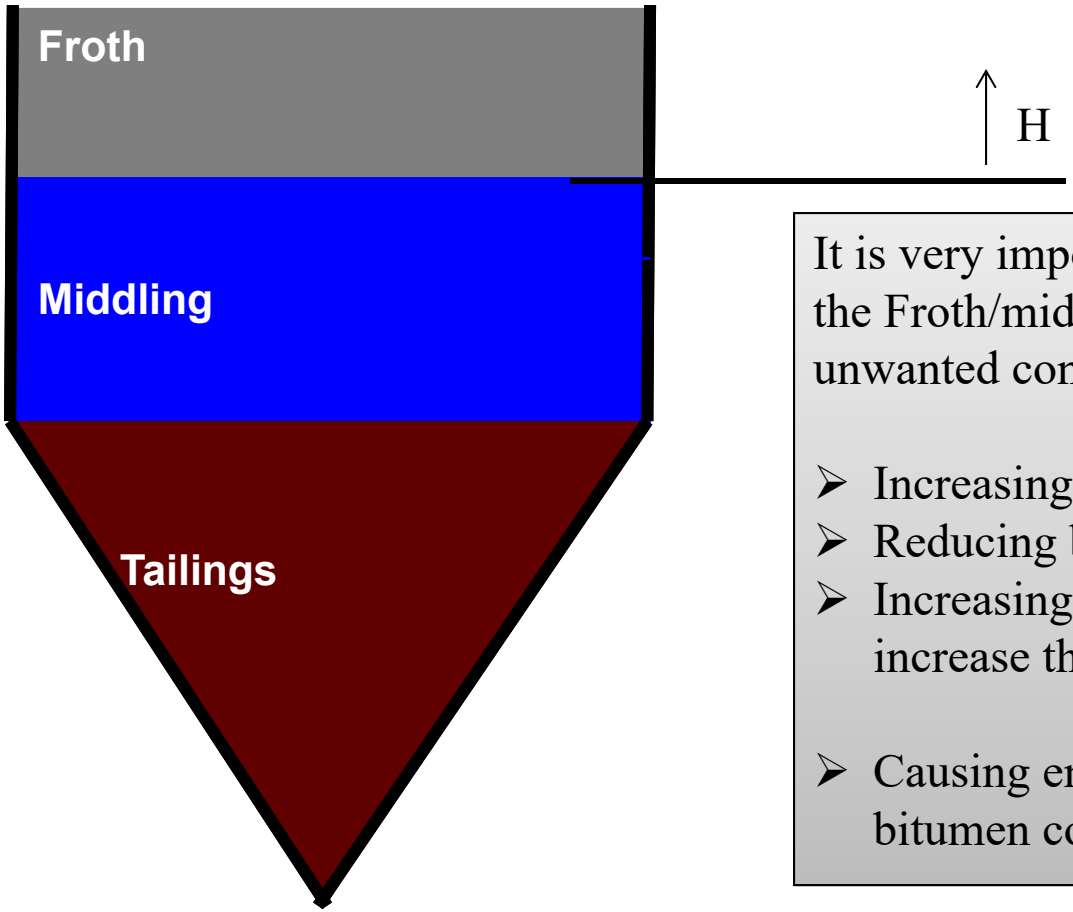


D/P Cell estimation: 35.71 cm
Image Processing estimation: 35.74 cm

0.5%

Data Synthesizing - Field Applications

Process & Motivation

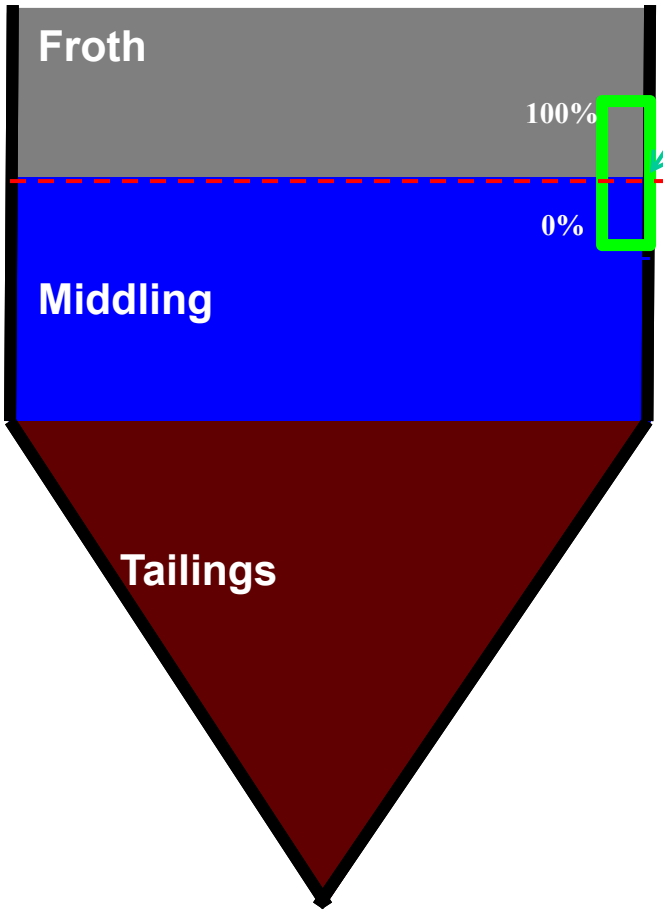


It is very important to correctly control the Froth/middling interface height to avoid unwanted consequences:

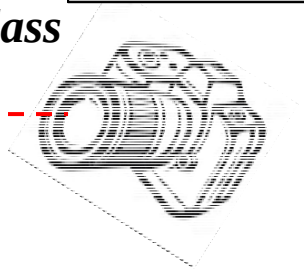
- Increasing the possibility of sanding
- Reducing bitumen recovery
- Increasing water content in Froth → increase the processing load on downstream

- Causing environmental impact due to increased bitumen content in tailings

Process & Motivation



Sight glass

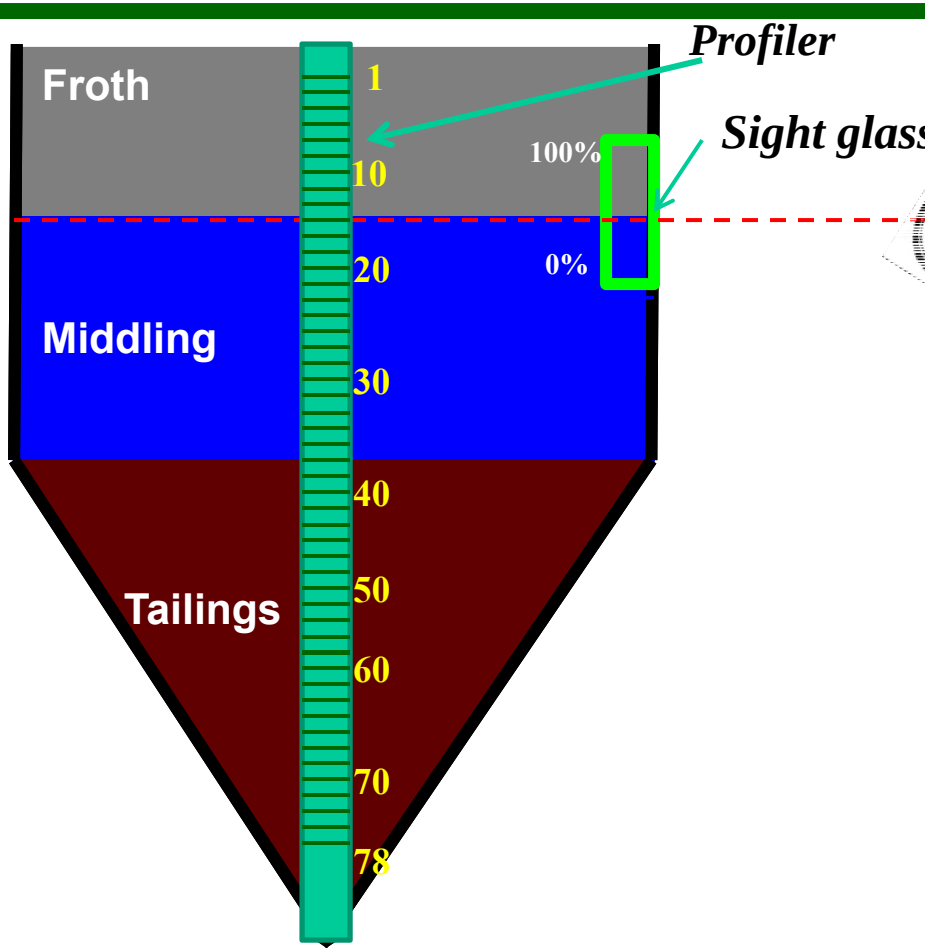


There is a camera that reads the interface level

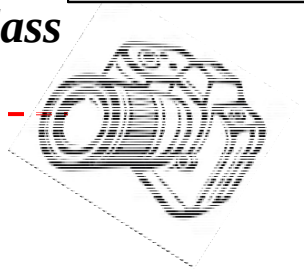
Accurate only if:

- The interface level within the sight glass
- There is no accumulated materials on the sight glass

Process & Motivation



There is a camera that reads the interface level

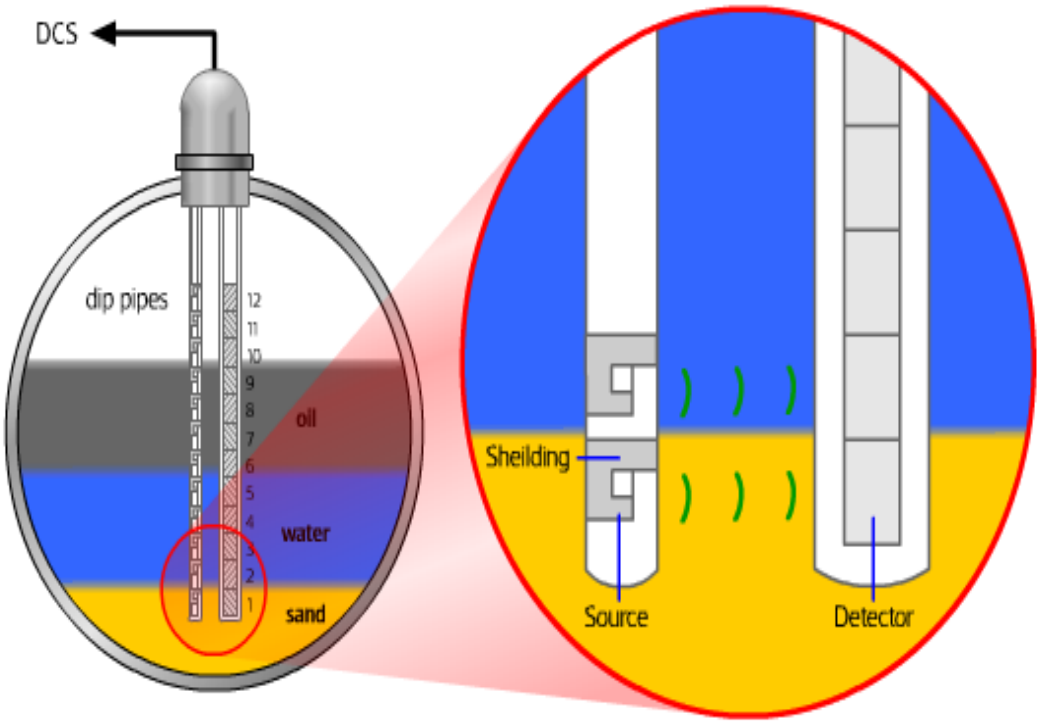


Accurate only if:

- The interface level within the sight glass
- There is no accumulated materials on the sight glass

Therefore, a profiler has been installed to help in measuring the interface when the Camera readings are not available

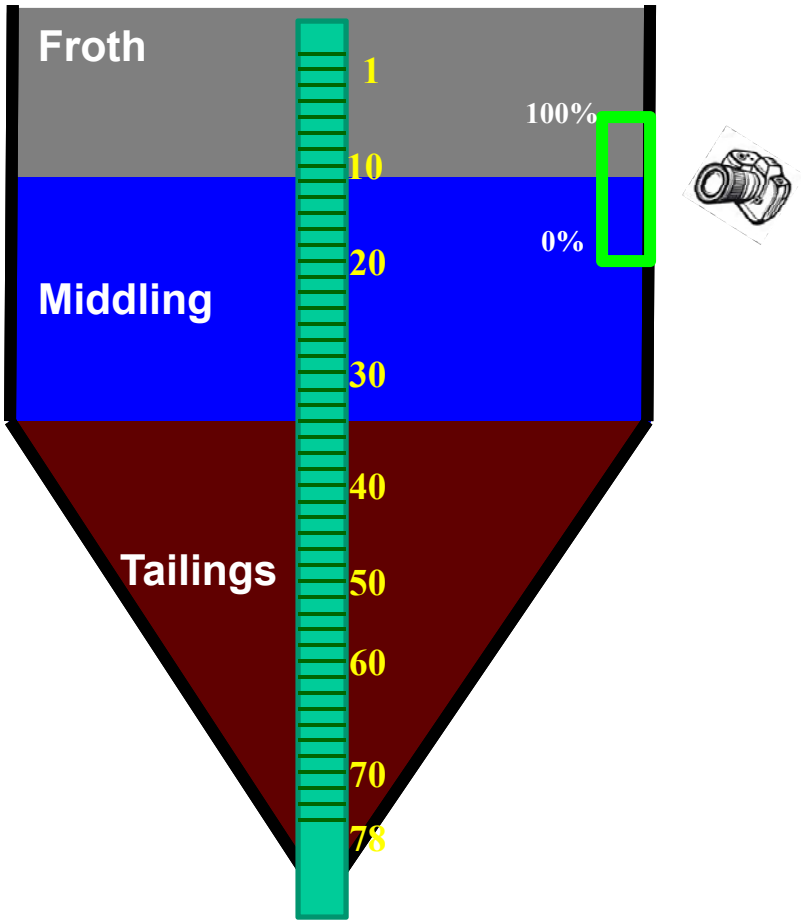
Process & Motivation



The profiler

- Two dip pipes assembly
- A narrow dip pipe emits low energy gamma
- Another dip pipe holds an array of gamma detectors
- Due to difference in density, each phase attenuates the signal by different amounts
- These signals are transmitted to DSC as density measurements

Process & Motivation

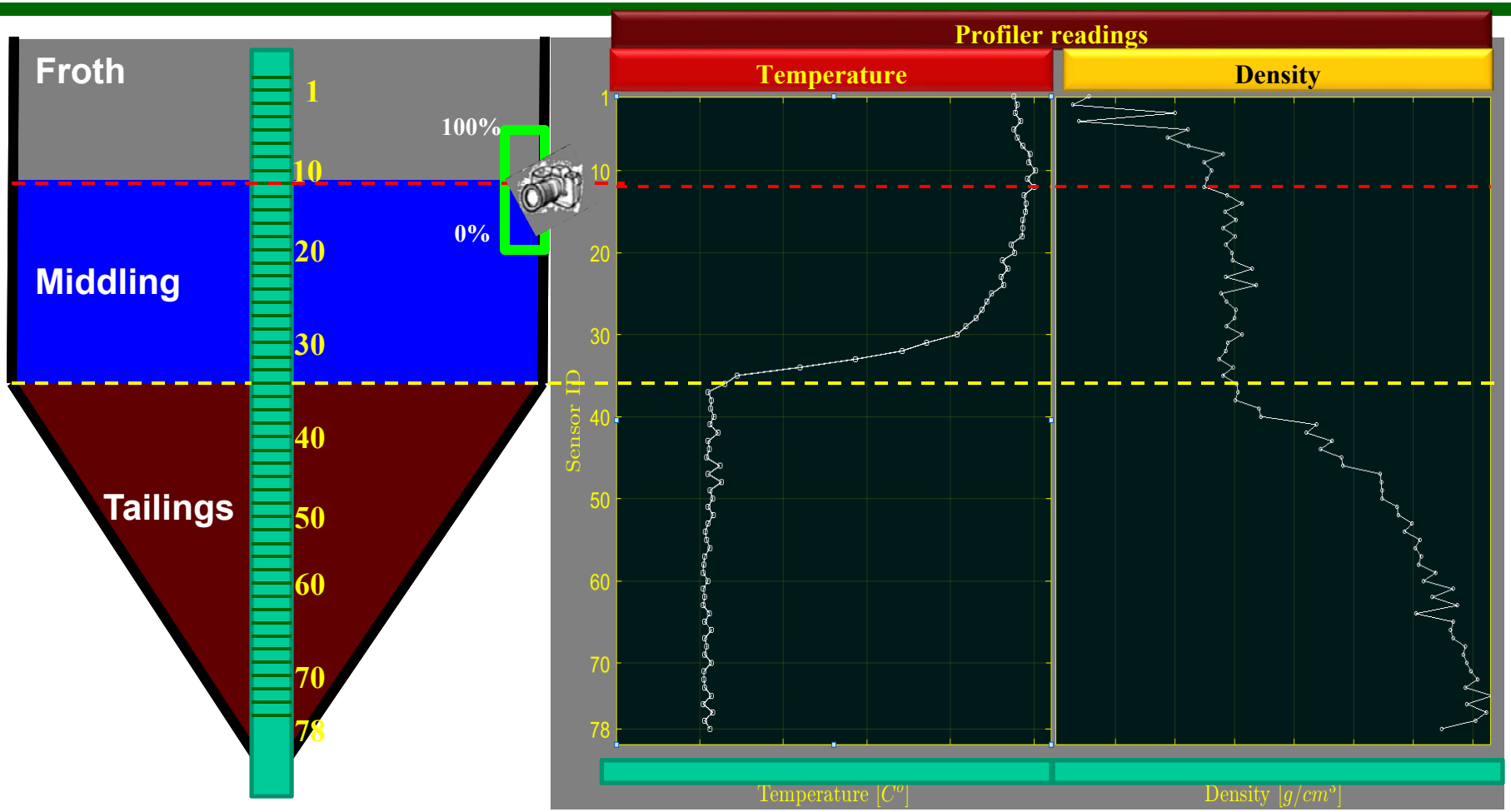


***Objective**:*

Ensure the availability of interface level readings that is:

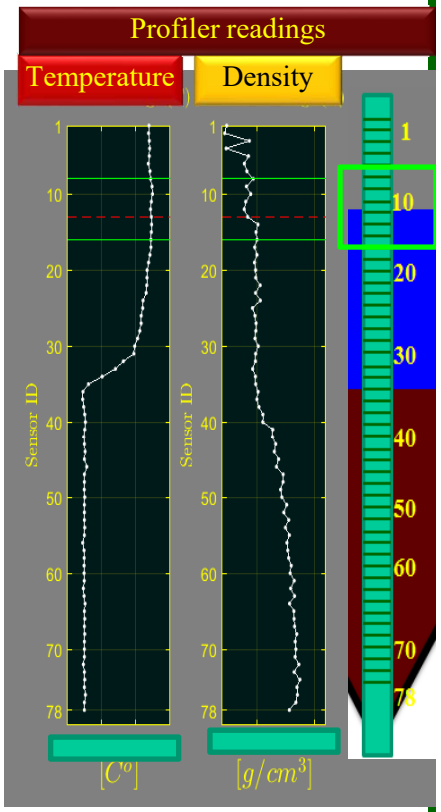
- ***Continuous***
- ***Accurate***
- ***Anywhere** in the PSV.*

Problem description & data visualisation

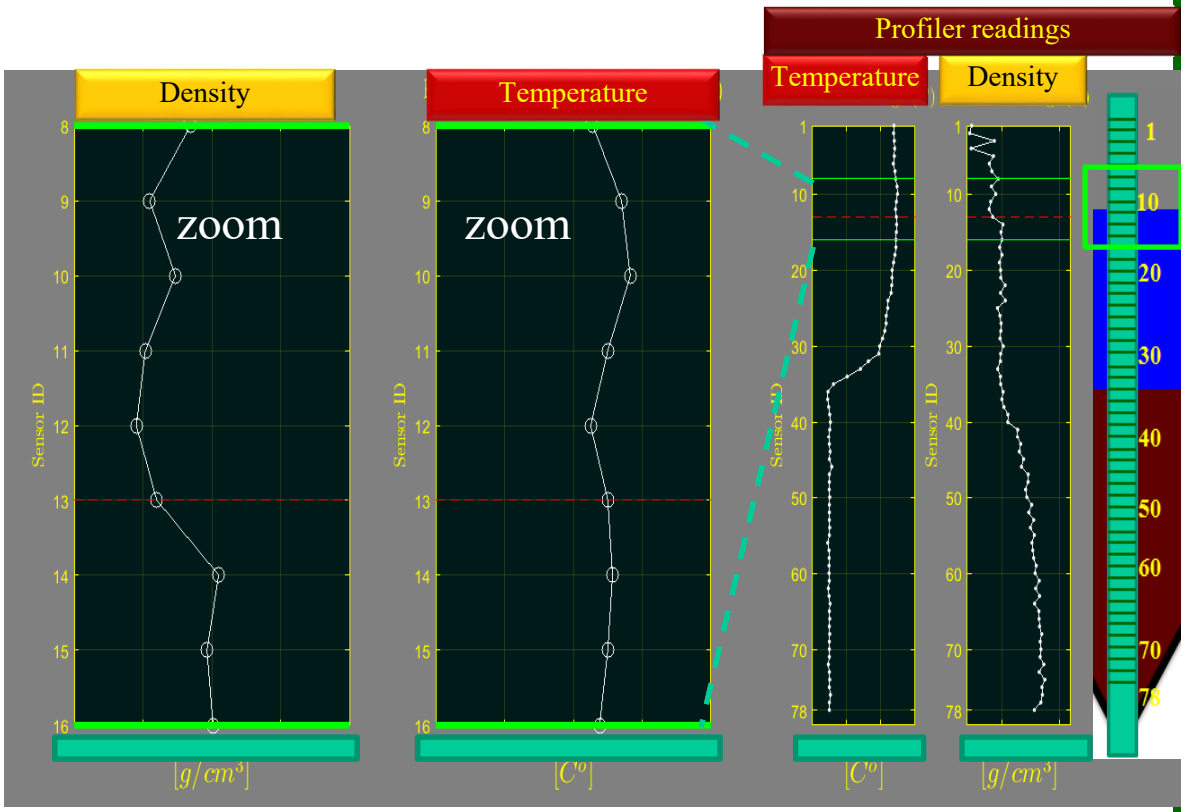


There is no clear characteristic behavior of profiler data around the interface

Problem description & data visualisation



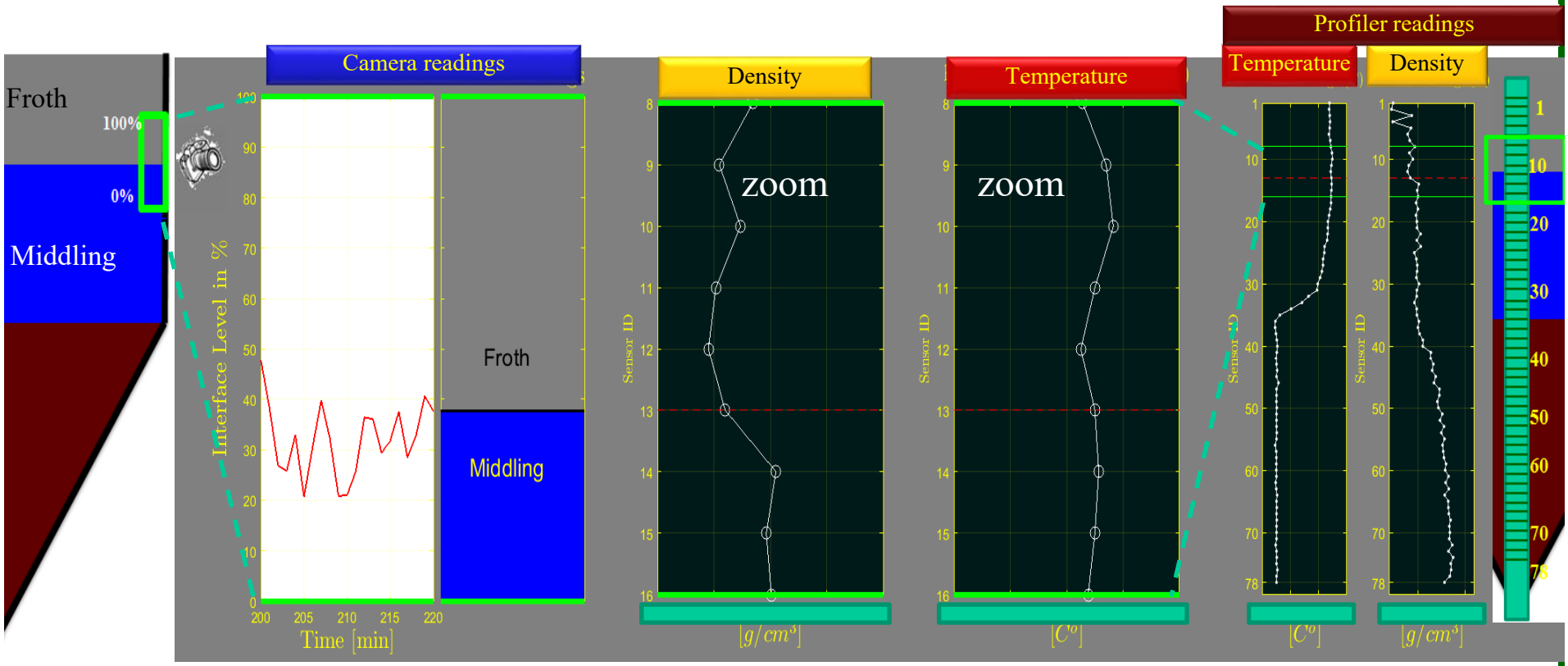
Problem description & data visualisation



There is no clear characteristic behavior of profiler data around the interface.

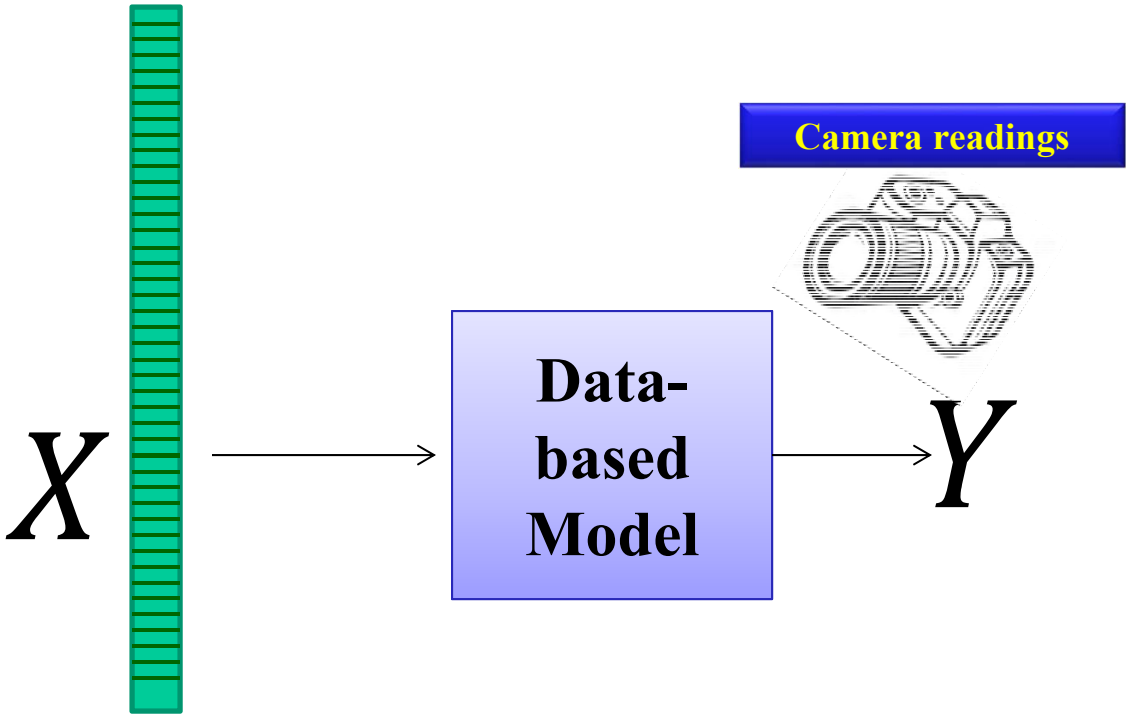
The majority of them move with the interface ⇒ Data-based modeling

Problem description & data visualisation



Method/Regression

Profiler readings: Density & Temperature



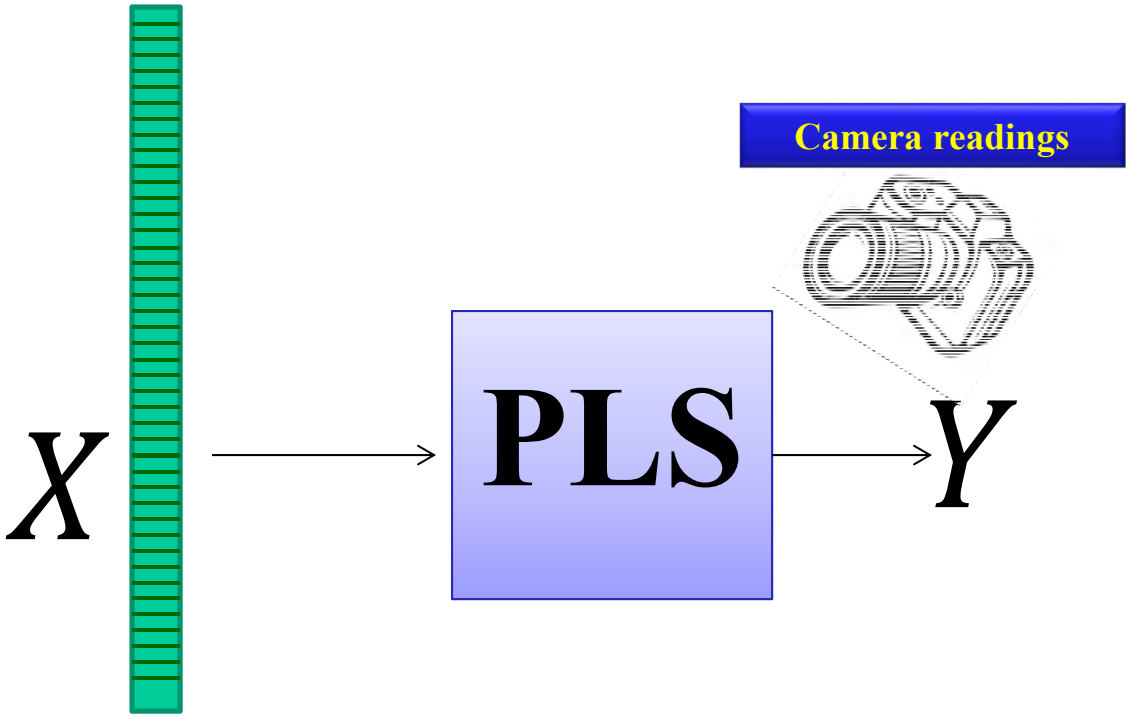
We choose
data-based
modeling technique where:

*model is built to predict interface from profiler (**D** & **T**) readings by learning from the camera as an accurate reference,*
*by means of **regression** between:*

- *Profiler data **X***
- *Camera readings **Y***

Method/Regression/PLS

Profiler readings: Density & Temperature

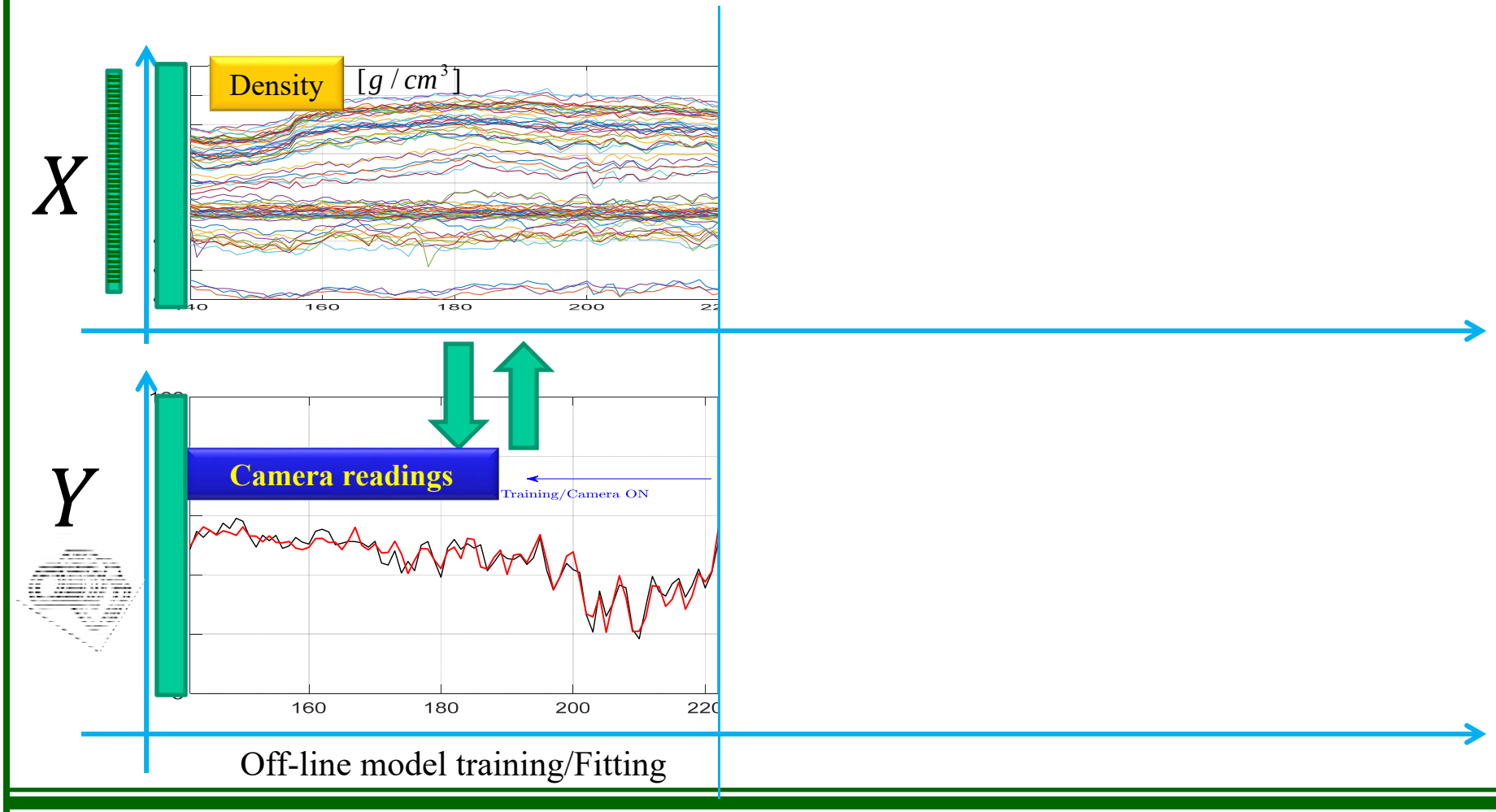


We choose:

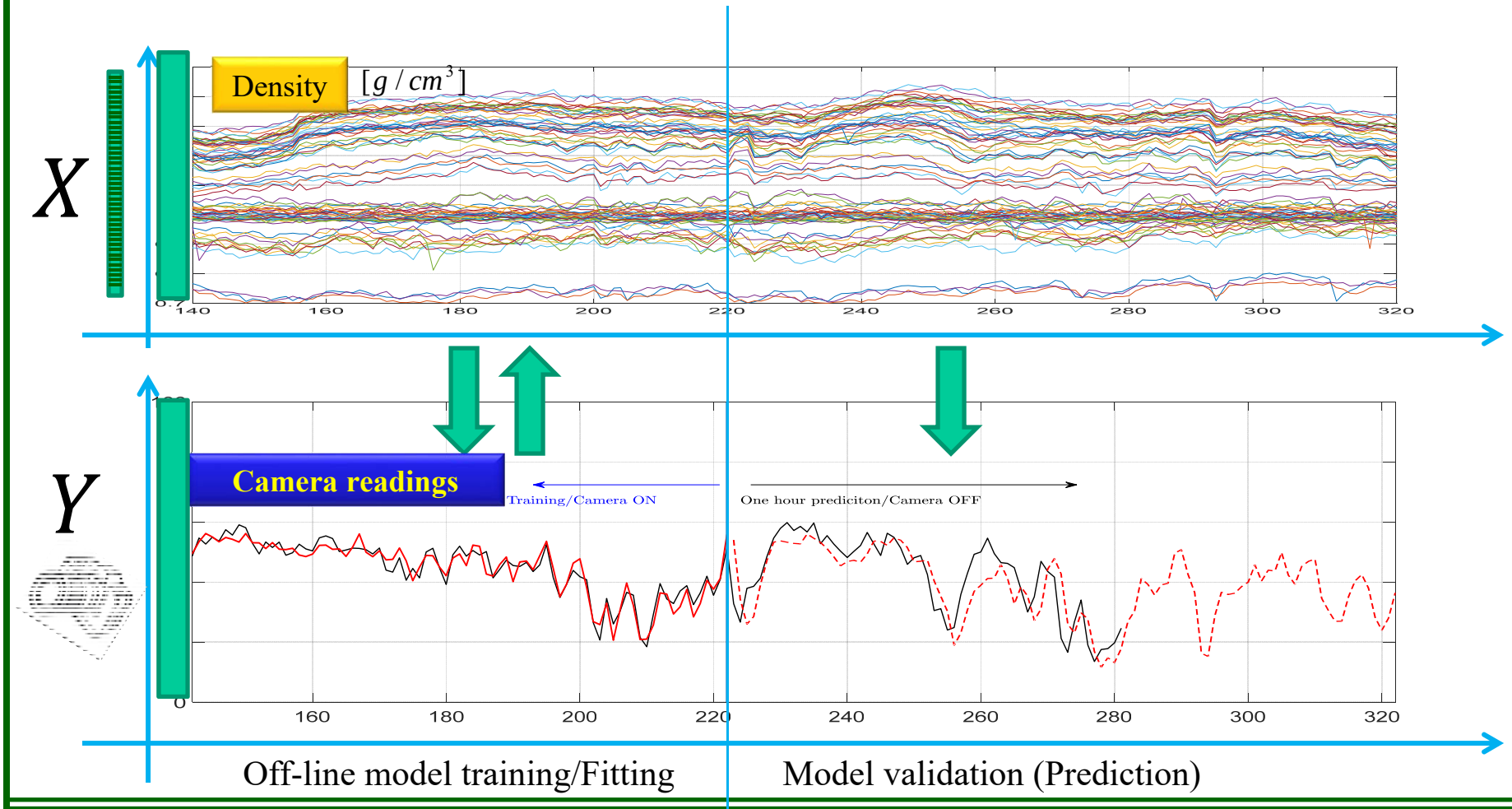
Partial Last Squares regression (PLS)

- *Solves the collinearity issues among the X variables*
- *Avoids inverting covariance matrix $(XX)^{-1}$ compared to OLS ⇨ "RPLS" suitable for online DCS Application*
- *Dimensionality reduction*

Method/Regression/PLS

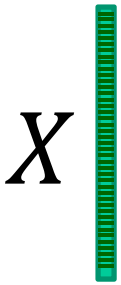


Method/Regression/PLS



However, due to variations in process conditions, the off-line model becomes outdated

Method/Regression/Recursive PLS



- *Choose a representative training set* $(\mathbf{X}_0, Y_0)$
- *Calculate the covariance matrices "offline"*

$$(\mathbf{X}^{\mathrm{T}}\mathbf{X})_0$$

$$(\mathbf{X}^{\mathrm{T}}\mathbf{y})_0$$

- *Update it online whenever a new sample* $(\mathbf{x}_t, y_t)$ *becomes available*

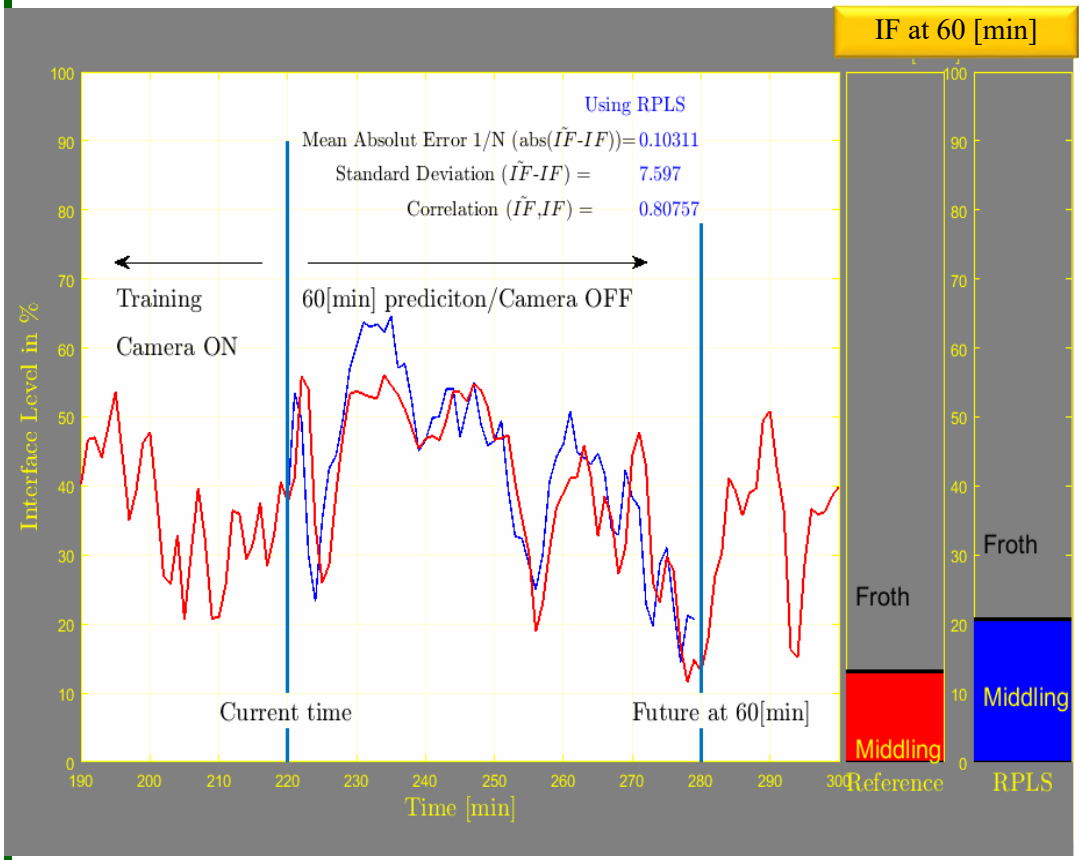
$$(\mathbf{X}^{\mathrm{T}}\mathbf{X})_t = \lambda(\mathbf{X}^{\mathrm{T}}\mathbf{X})_{t-1} + \mathbf{x}_t^{\mathrm{T}}\mathbf{x}_t$$

$$(\mathbf{X}^{\mathrm{T}}\mathbf{y})_t = \lambda(\mathbf{X}^{\mathrm{T}}\mathbf{y})_{t-1} + \mathbf{x}_t^{\mathrm{T}}y_t$$

$1 > \lambda > 0$ Forgetting factor

The model will be updated with the most recent profiler & camera data and be used **when the camera data are not available**

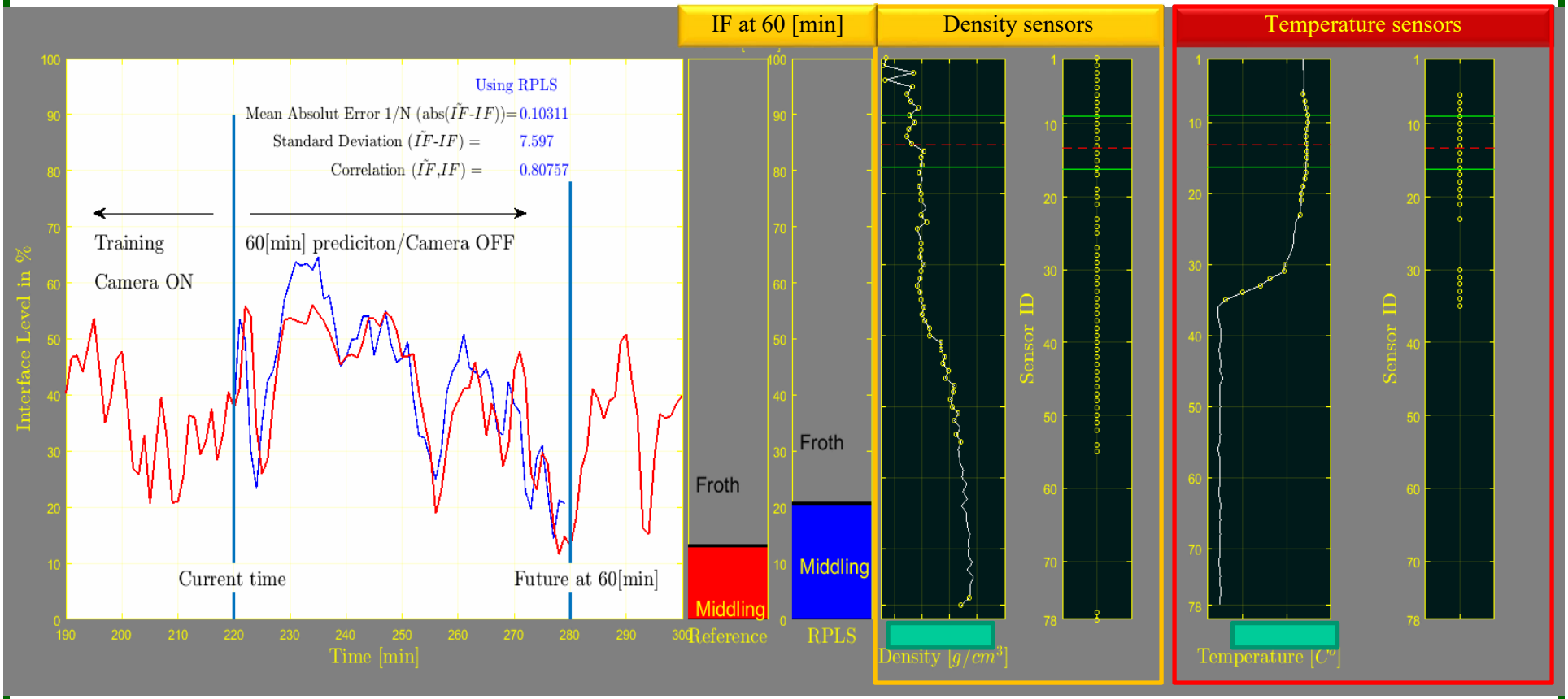
Results



The RPLS prediction is able to track the reference **when the camera readings are not available**

Results

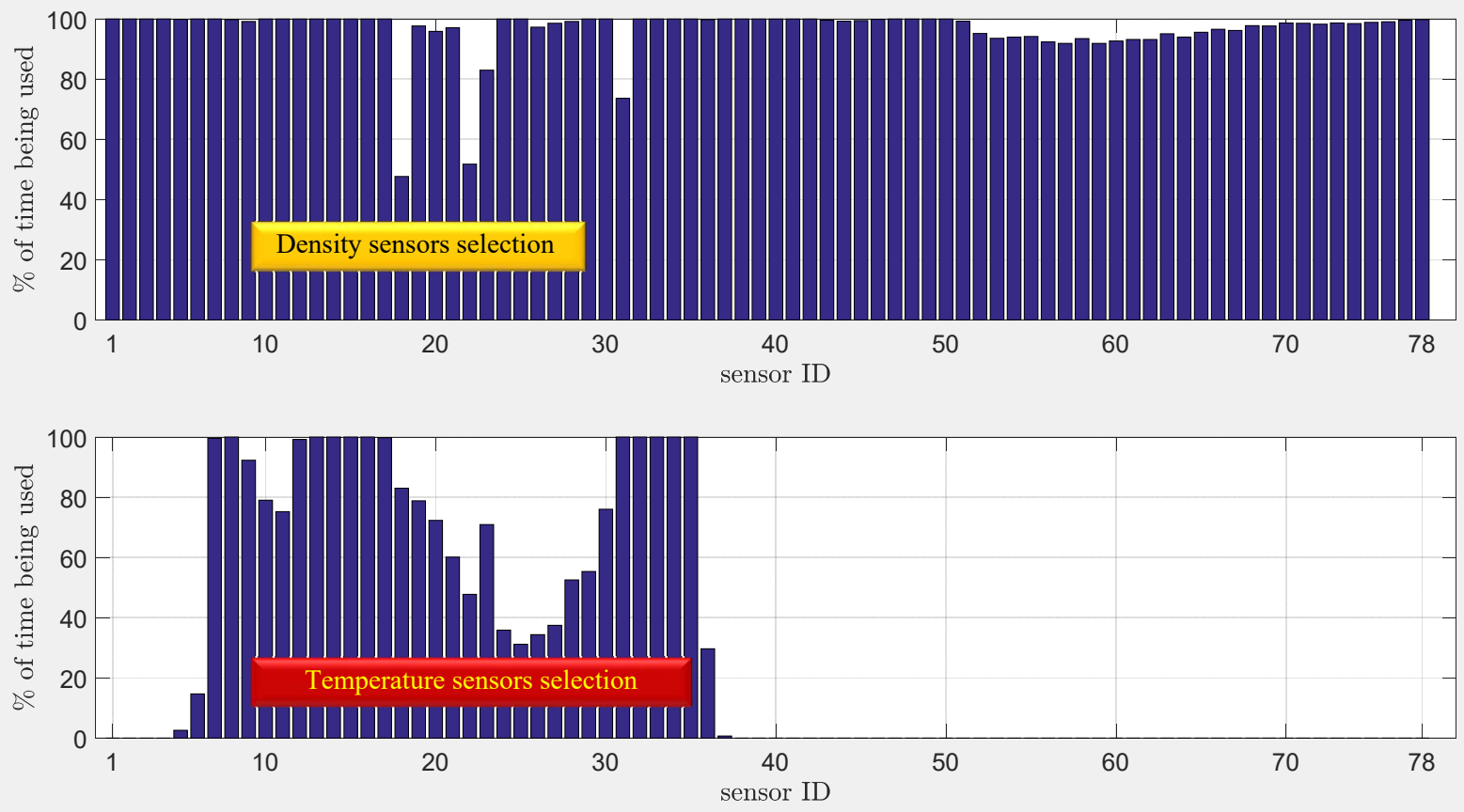
PLS helps in dimensionality reduction in X



The RPLS algorithm allows to select the input variables that have the highest "importance"

➡ ***Reduce dimensionality***

Results

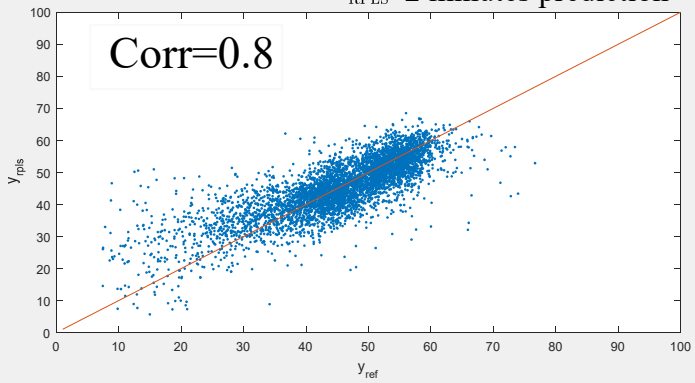


The RPLS algorithm allows to select the input variables that have the highest "importance"

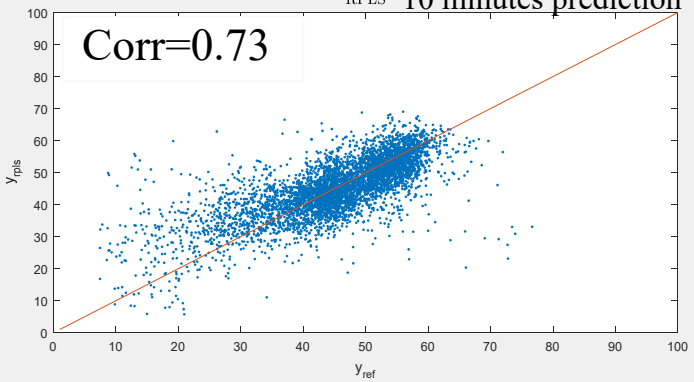
➡ ***Reduce dimensionality***

Results

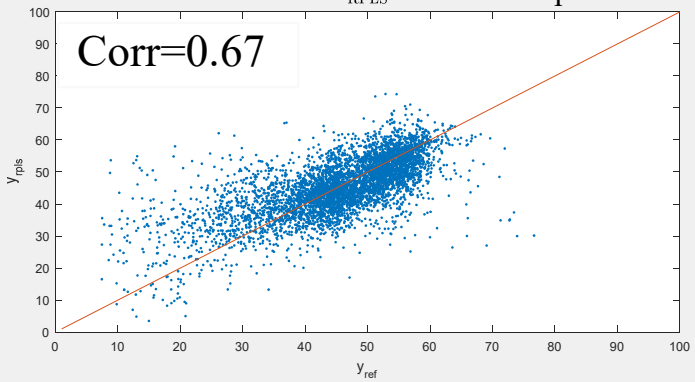
RPLS 2 minutes prediction



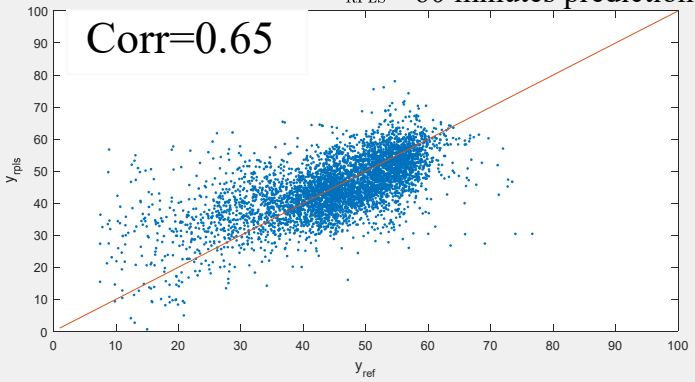
RPLS 10 minutes prediction



RPLS 20 minutes prediction



RPLS 60 minutes prediction



Closer we are to the current (update) time, better the interface prediction performance is

Analytics Toolboxes in Progress

Soft Sensor Analytics

Data preprocess	Cumbersome

Resample, outlier detection, rearrange, normalize, detrend….

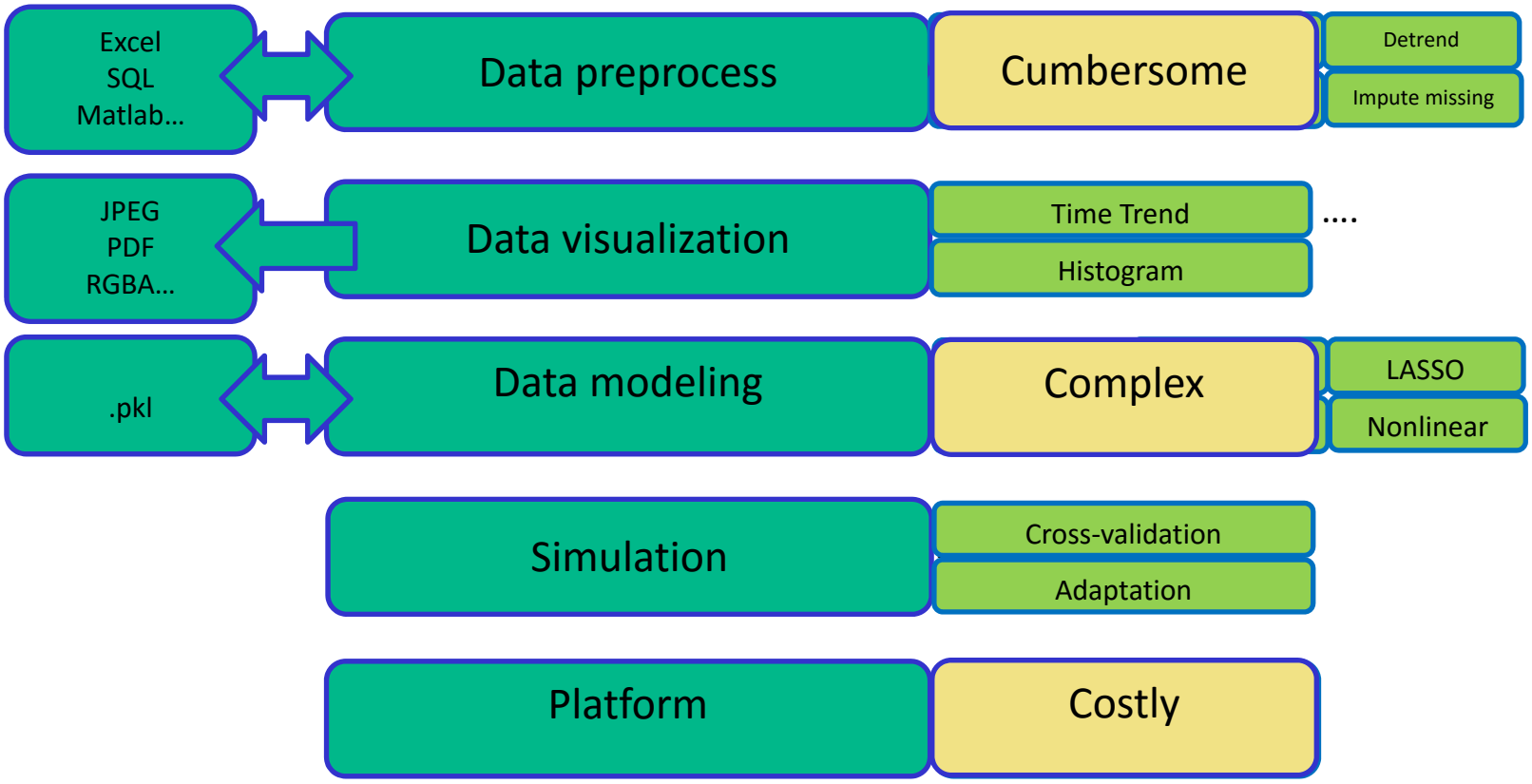
Data modeling	Complex

OLS, LASSO, RR, PCA, PLS, nonlinear regression….

Platform	Costly

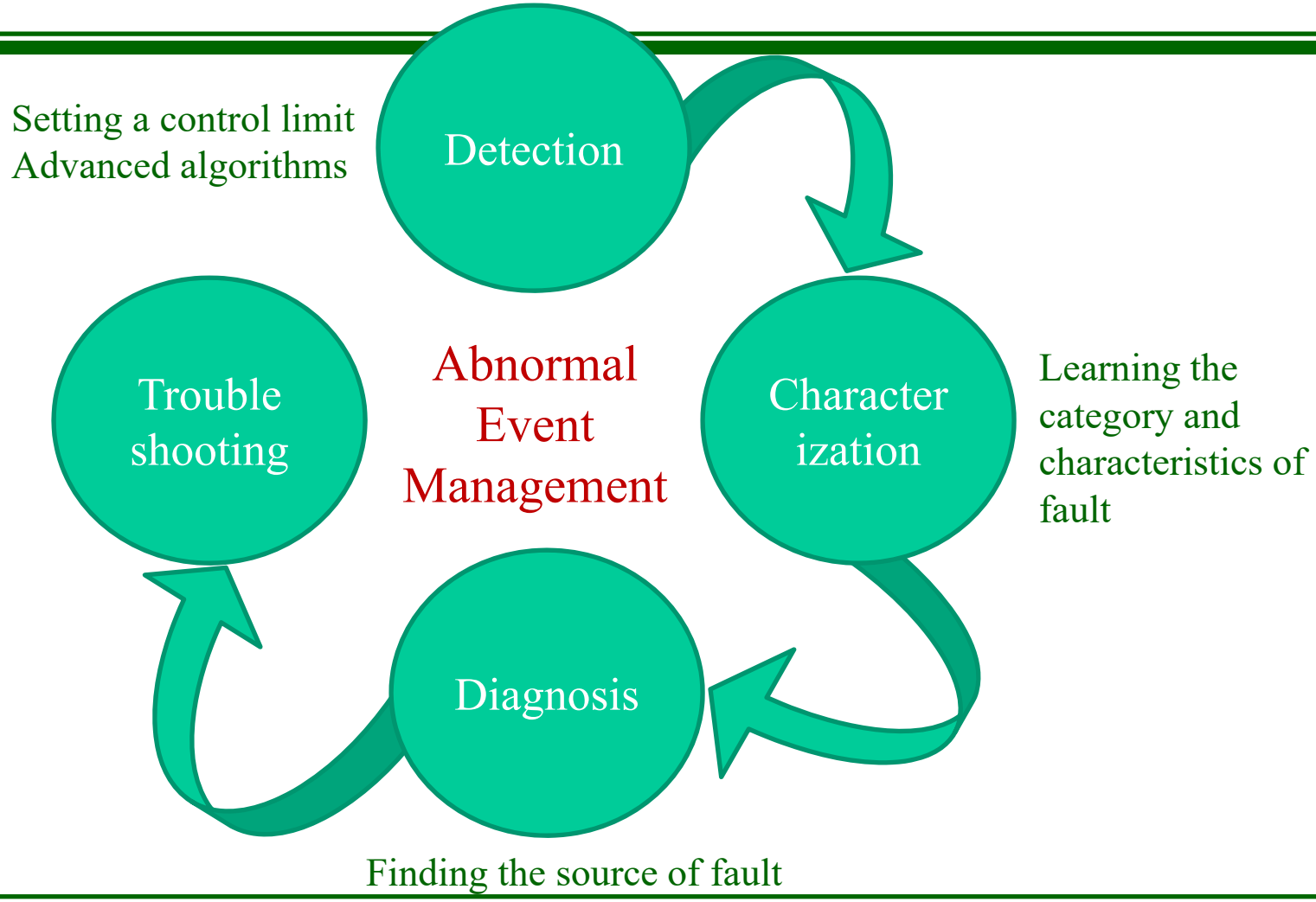
MATLAB, Unscrambler….

Soft Sensor Analytics

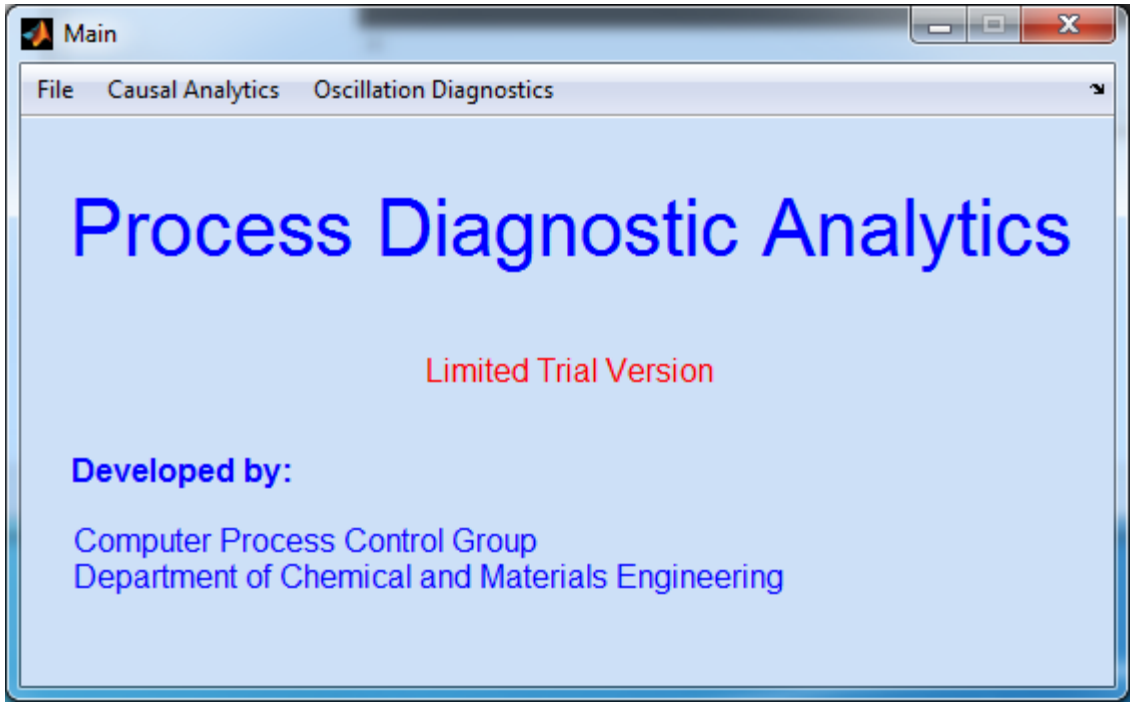


MATLAB, Unscrambler….

Process Diagnosis Analytics - Toolbox



Process Diagnosis Analytics - Toolbox



- Causal Analytics: Extracts causality relations among the variables from data
- Oscillation Diagnostics: Detects and characterizes oscillatory type of faults

Conclusion

- Data analytics is an emerging area of research and applications
- Great potential, demands and opportunities
- Applicable in every sector
- Opportunity for everyone

Acknowledgments

- Zheyuan Liu, Fadi Ibriham, Ruomu Tan, Anahita Sadeghian, Ming Ma, Elham Naghoosi
- NSERC Industrial Research Chair in Control of Oil Sands Processes
- AITF Industry Chair in Process Control
- University of Alberta