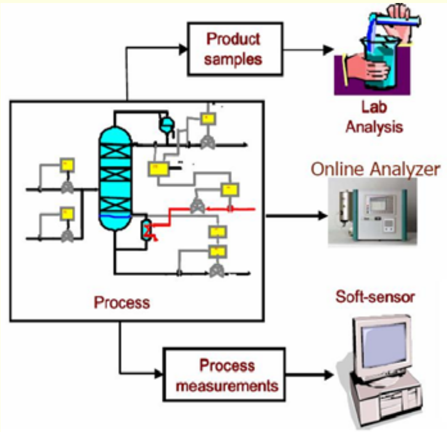
Soft Sensors and Its Applications

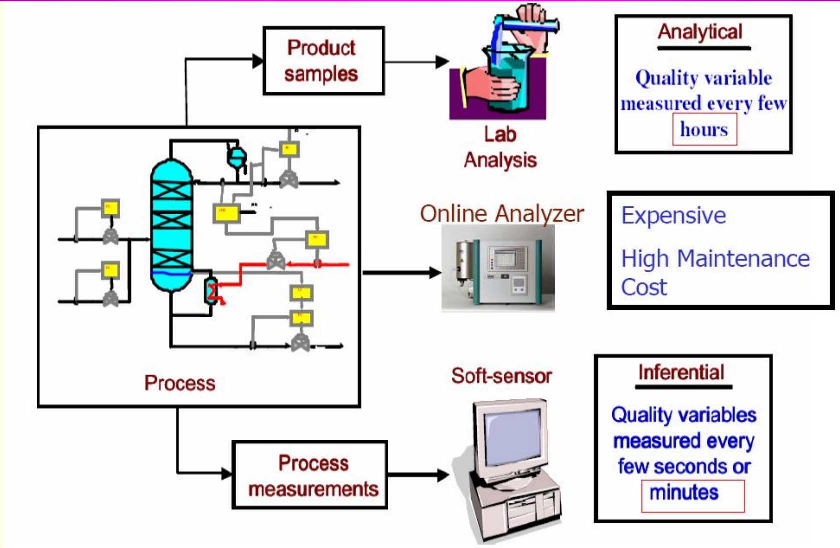


劉毅 & 陳榮輝
台灣中原大學化工系
計算機程序系統工程研究室
ail : jason@wavenet.cycu.edu.tw
http://wavenet.cycu.edu.tw/

Outline

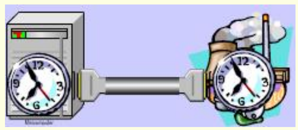
1. Background and Motivation
2. Current methods: a simple review
3. Some Challenges
4. Recent results
5. Concluding remarks

What is a Soft Sensor?



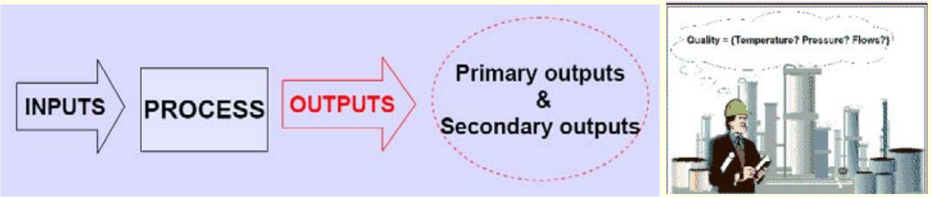
Measurement Problems

- Difficult to provide reliable, fast, on-line measurements to control quality
- The quality measure may only be available as a laboratory analysis or very infrequently on-line
  - » Lead to excessive off-specification of products
- Reliability of on-line instruments
  - » Drafting, fouling, sample system failure etc.
- Automatic control and optimization schemes cannot be implemented.



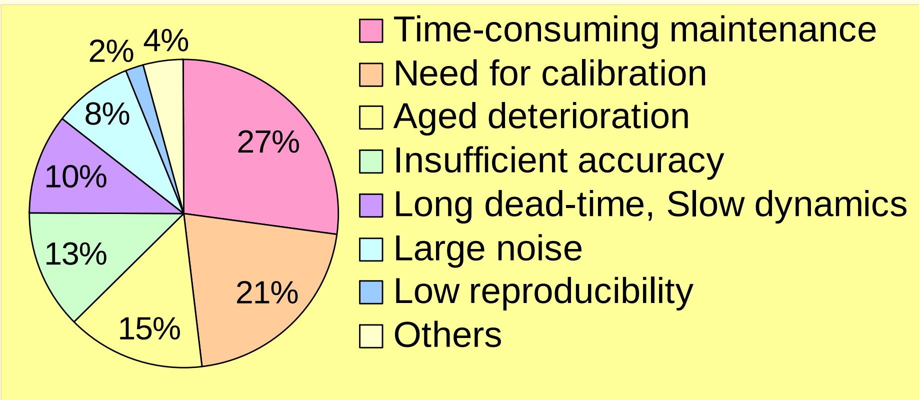
Measurement Problems

- Productivity is quantified by specification upon which the product is sold, eg. purity, physical or chemical properties
  - » Primary variables: difficult to measure on-line



- The other output (eg. temperature, flow and pressure)
  - » Secondary variables: easily measured on-line

Hardware Sensor Problems



Questionnaire to 26 companies (*JSPS PSE 143 committee, 2003*)

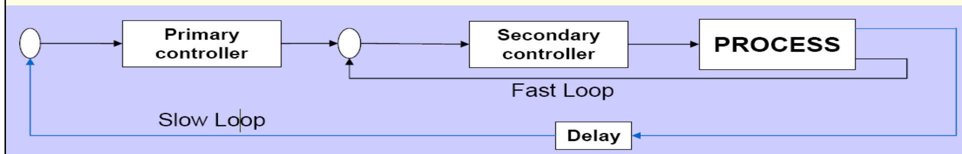
Solutions 1: Manual Control

- Common approach to effecting control on process is to control it manually

- Return of information for control purpose from laboratory is slow and irregular

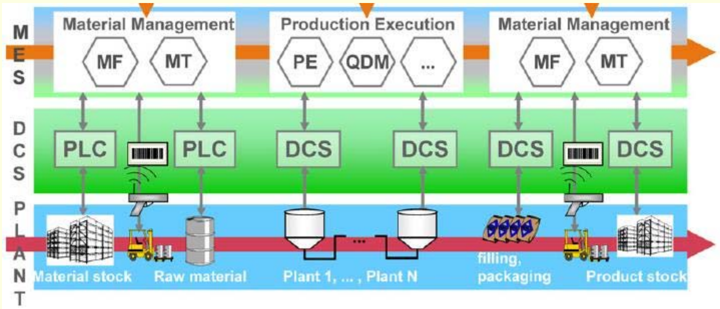
- Its success depends on the operator's training and experience

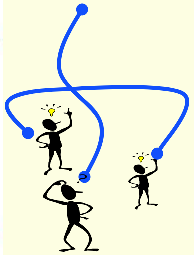
Solutions 2: Parallel Cascade Control



- The slow loop dictated by the delay of a measurement device
- Assumption: If the secondary variable is kept at some value, then the primary variable also be kept at some level!
  - » Often employed in the product composition control of distillation columns
  - » Two control loops:
    - – Fast secondary loop: a tray temperature
    - – Slow primary loop: product composition
- Slow loop is used to correct the set-point of the secondary loop
- Drawbacks:
  - » Disturbances affecting the secondary variable may not affect the primary variable
  - » The relationship between the primary variable and the secondary variable is non-linear, and can change depending on the operating conditions

Motivation: Rich Data



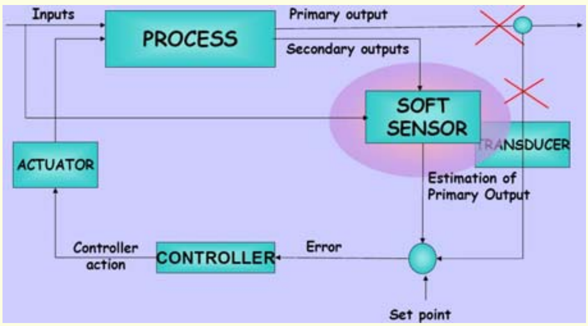


- Huge data
- Little information or knowledge

Klatt and Marquardt. Computers and Chemical Engineering, 2009, 33: 536-550.

Solutions 3: Inferential Measurement

- Inferential measurement allows process quality, or a difficult to measure process variables, to be inferred from other easily made plant measurements such as pressure, flow or temperature.
- Soft(ware) sensor: Model which estimates immeasurable process states based on easy to measure input and output variables
  - » Software sensor, virtual soft sensor, smart sensor, virtual analyzer, inferential control

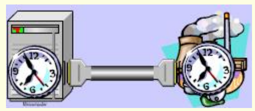


Motivation: What Can a Soft Sensor Do?

- Using a good soft sensor, we can do many things in PSE:
  - » Online Quality Prediction
  - » Process Fault Detection
  - » Process Monitoring
  - » Sensor Fault Detection
  - » Inferential Control
  - » …

Types of Models for Soft Sensor Developing

- First principle models
  - » Based on balance and phenomenological equations
  - » Process knowledge required
  - » Comparatively expensive
- Data driven models
  - » Identified using process data
  - » Comparatively inexpensive



- 13

First Principle Models vs. Data Driven Models

	First Principle Models	Data Driven (Empirical) Models
Method	Fundamental Principles	Experiments
Advantages 	Excellent relationships between parameters in physical systems and the transient behavior of the systems	Good for process design since it is easy to use (Easy, less effort)
Disadvantages	• Complex. ex. distillation column, 10 compounds, 50 trays 500 diff. Eqs • Large engineering effort	Less accuracy; do not provide enough information to satisfy all process design and analysis requirement

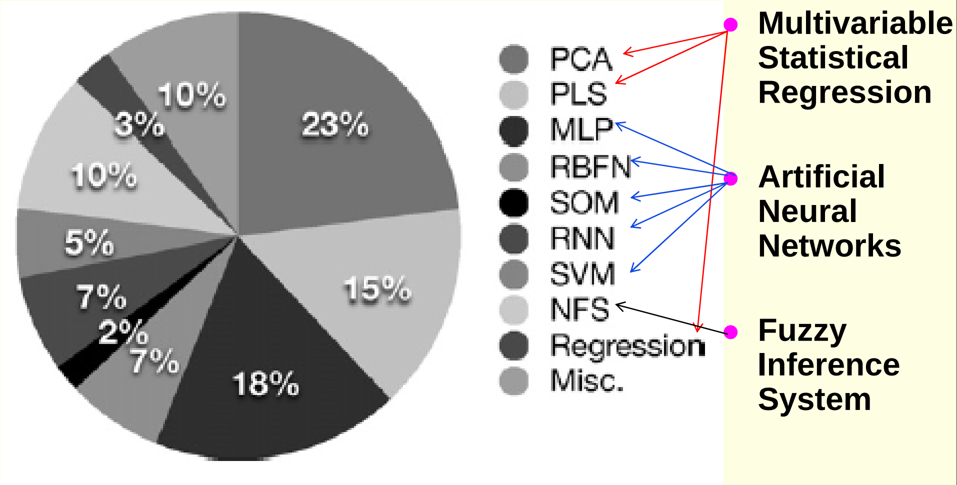
- 14

Outline

1. Background and Motivation
2. Current methods: a simple review
3. Some Challenges
4. Recent results
5. Concluding remarks

- 15

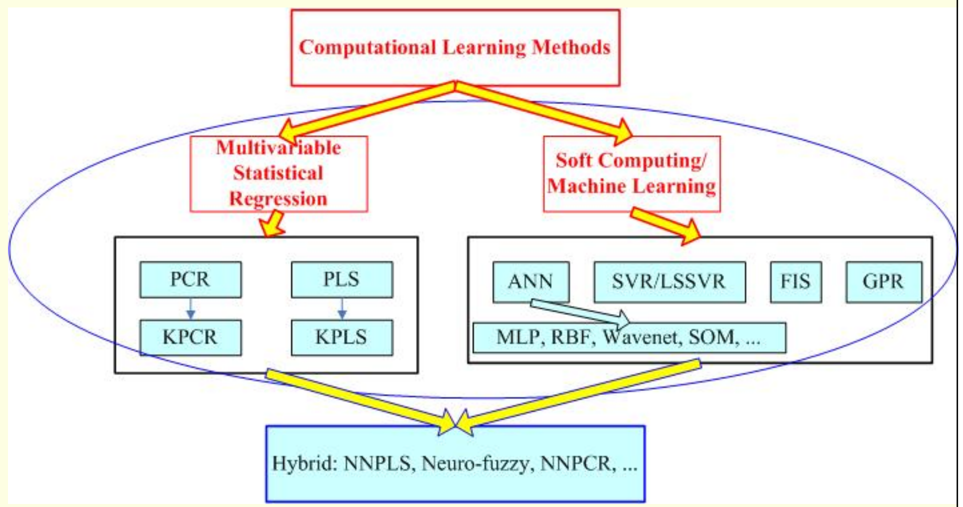
Current Popular Methods



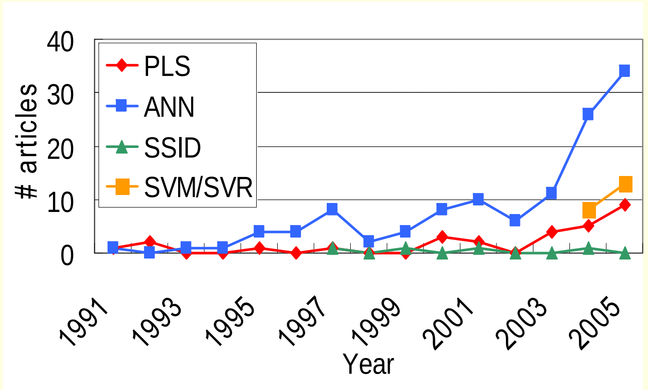
Kadlec P, et al. *Comput. Chem. Eng.*, **2009**, 33(4): 795-814.

- 16

Current Popular Methods



Methods with Applications



ANN is dominant in the literature while PLS is popular in industry.

(Kano and Nakagawa, *Comput. Chem. Eng.*, 2008)

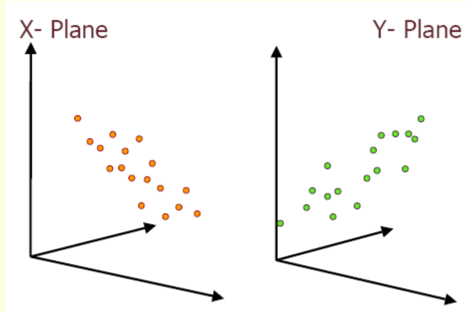
Methods with Applications

	methodology							
process	Phys	MRA	PLS	O.L.	ANN	JIT	Gray	total
distillation	20	256	41	6	0	5	3	331
reaction	5	32	43	0	0	5	1	86
polymerization	0	4	8	0	3	0	5	20
others	0	1	1	0	0	0	0	2
total	25	293	93	6	3	10	9	439

Phys: physical model
MRA: multiple regression analysis
PLS: partial least squares regression
O.L.: other linear regression
ANN: artificial neural network
JIT: just-in-time model
Gray: gray-box model or hybrid model between physical model and statistical model

(Kano and Ogawa, *J. Process Control*, 2010)

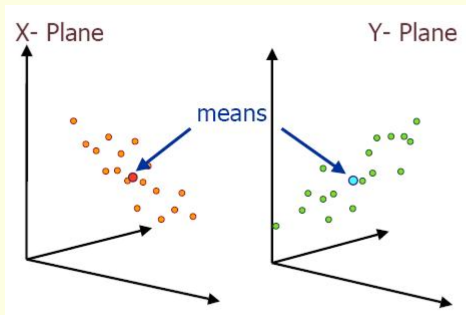
PLS : A Geometric Interpretation



Each object is one point in the X-space and one point in the Y-space.

The data matrices **X** and **Y**, thus they are connected swarms of points in these two spaces

PLS : A Geometric Interpretation

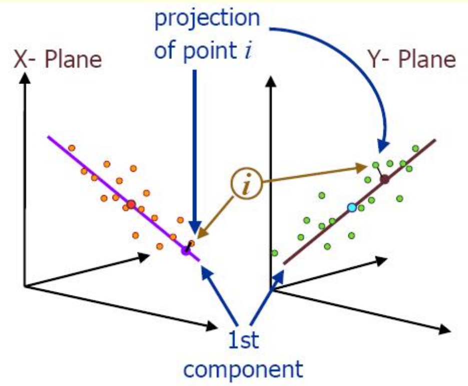


Calculate the average of each variable. The vectors of variable averages are points in the X and Y-spaces.

Subtracting averages from the two data matrices, corresponds to moving their origins to the centre of their respective data swarms.

- 21

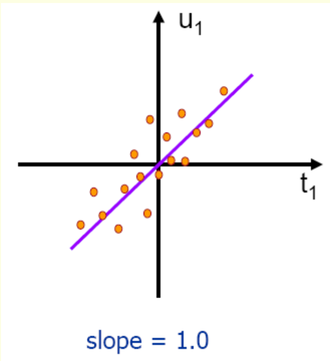
PLS : A Geometric Interpretation



- The first PLS component is a line in X-space and a line in Y-space, through the average points, such that:
  - » the lines well approximate the data (maximize variance explained), and
  - » the projections (**t**$_1$ and **u**$_1$) are well correlated (maximize correlation)

- 22

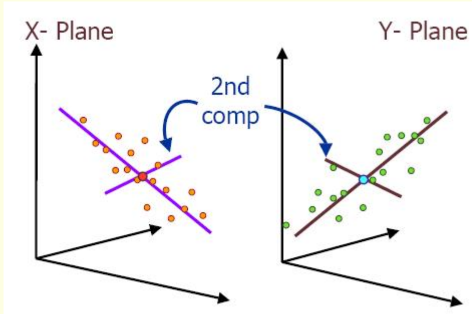
PLS : A Geometric Interpretation



The projected coordinates in the two spaces (u_1 and t_1 in **Y** and **X**) are correlated in the inner relation

- 23

PLS : A Geometric Interpretation

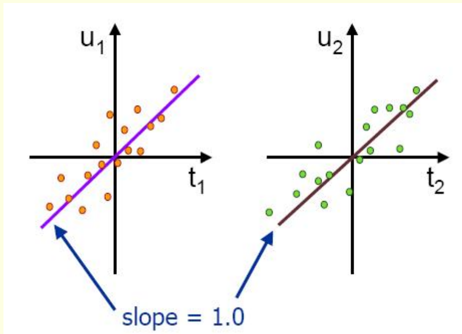


The 2nd PLS lines in the X and Y-spaces pass through the average points.

These lines improve the approximation and the correlation as much as possible.

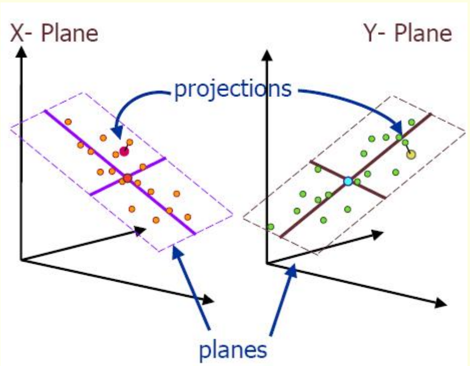
- 24

PLS : A Geometric Interpretation



The second projection coordinates (**u**$_2$ and **t**$_2$) correlate, but usually less well than the first.

PLS : A Geometric Interpretation



- The PLS components together form planes in the X and Y-space.
  - » Scores: t_1, t_2 and u_1, u_2 location on planes
  - » Loadings: **P** and **Q** orientation of planes
  - » DModX – distance of observation from plane in the X-space
  - » DModY – distance of observation from plane in the Y-space

PLS *is* Simultaneous Dimension Reduction and Regression

max Var(**scores**) Corr^2(**response**,**scores**)

Dimension Reduction (PCA)

Regression

Using PLS Model

- Note: if you use the regression form of PLS (Y=XB), then
  - » cannot check data for consistency
  - » cannot handle missing data
- Solution:
  - » Use latent variable (LV) form of PLS model: calculate t's from x_{new} and then get predictions from the t's

1. Is new data vector, x_{new}, consistent with past operation?
   - » calculate the t-scores, SPE, and T^2 values coming from x_{new}
   - » if SPE and T^2 are below their limits, then make prediction
   - » if not, then:
     - – flag bad values in x_{new} as missing and retry the PLS model
     - – or, discontinue soft sensor until process problems are corrected

Using PLS model

2. Handling missing data
   - » Must use the LV form of model to handle missing data in x_{new}
   - » Calculate the new t's with missing data algorithm: $\mathbf{t}_{new}$

$$y_{pred} = t_{1new}c_1 + t_{2new}c_2 + \cdots$$

   - » Do not use the regression form of the PLS model:

$$y_{pred} = \mathbf{x}_{new}^T \mathbf{b}$$

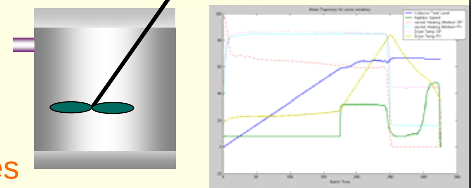
Types of Processes and Data Structures

Continuous Processes

Data structures

Hybrid Processes

Batch Processes



Data structures

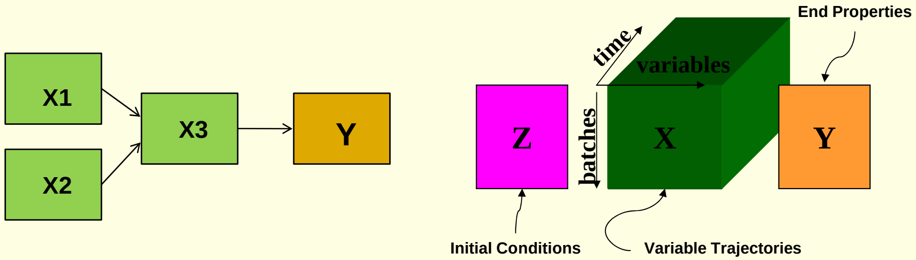
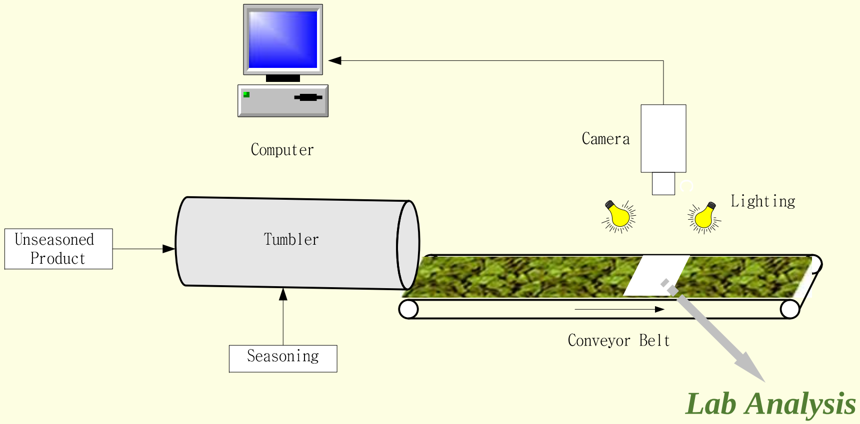
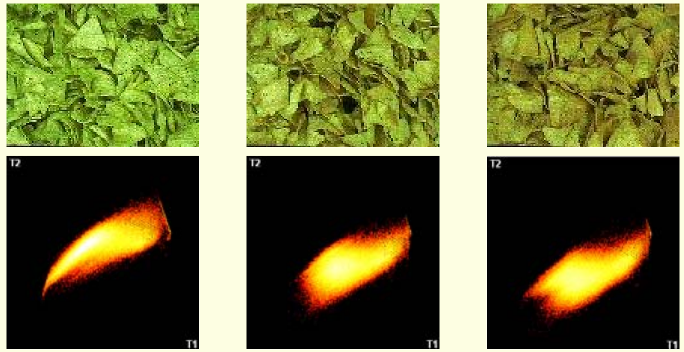


Image-based Soft Sensor for Monitoring and Feedback Control of Snack Food Quality

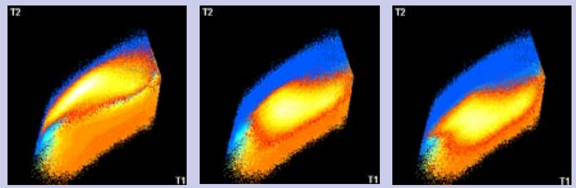


PCA Score Plot Histograms

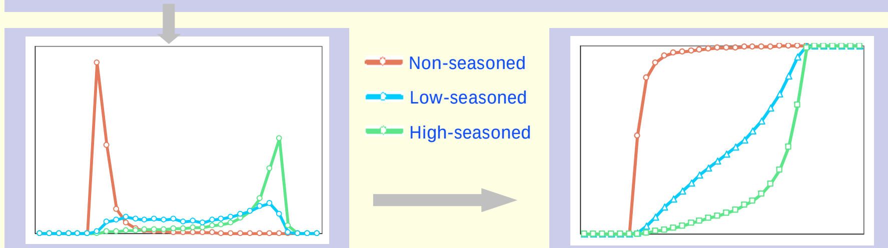


Non-seasoned Low-seasoned High-seasoned

Distribution of Pixels



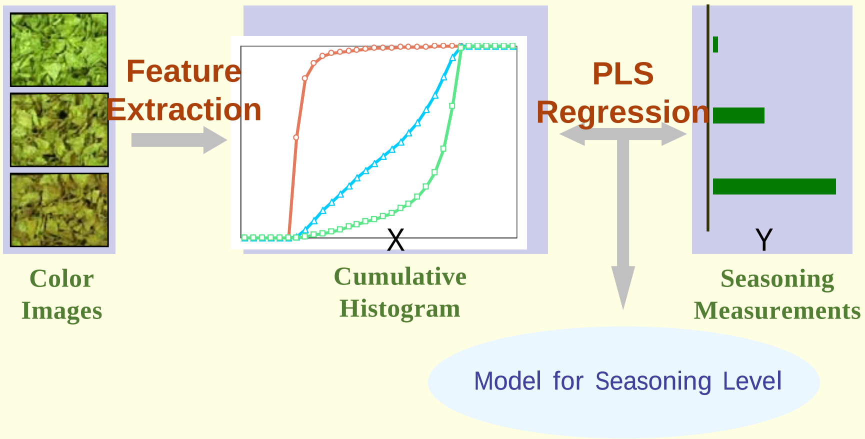
Superposition of the score plots from 3 sample images on top of the mask



Distribution of pixels

Cumulative distribution of pixels

Model to Predict Seasoning Concentration



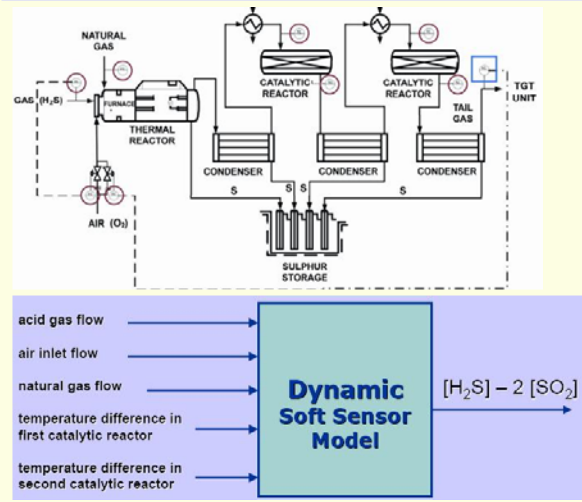
Example : Distillation Column

- Predicting a lab-measured property: X_D (distillate composition)
- Data used: 45 tags around the unit, and its ancillary equipment:
  - » flows, temperatures, pressure
  - » calculated variables from a previous model:
    - – $\log(X_D)$ = function of process measurements
    - – some of the terms in the above equation are non-linear, so add these into the model as new variables

$$\log(x_D) = \cdots + \log(T_{84}) - \frac{1}{P_{12}} + (T_{84})^{3/2} + \cdots$$

Example: Sulphur Recovery Unit

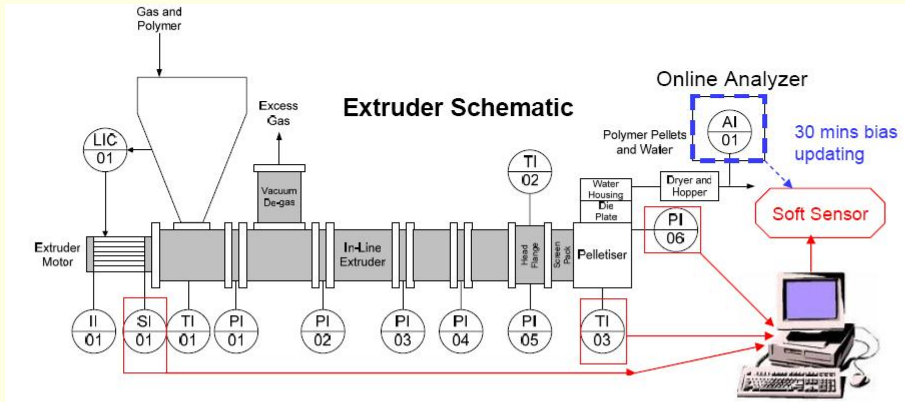
$H_2S + 1.5O_2 \longrightarrow H_20 + SO_2$ $\qquad 2H_2S + SO_2 \longrightarrow 3S + 2H_2O$



H_2S and SO_2 Analyzer

Example: Extruder

- Melt index of polymer is observed to be nonlinearly related to variables monitored in the extruder.



Caution 1: Collinear Variables

- Correlation versus Causation
- Modeling with collinear variables

Example:

Output: y Inputs: x_1, x_2, x_3, x_4, x_5

"True" cause and effect relationship (unknown): $y = 1.0x_2 + e$

Data: the 5 x-variables are collinear (in practice often nearly so)

y	x_1	x_2	x_3	x_4	x_5
-1.092	-1	-1	-1	-1	-1
2.074	2	2	2	2	2
-0.581	-0.5	-0.5	-0.5	-0.5	-0.5
0.888	0.9	0.9	0.9	0.9	0.9
...	...	...	...	...	...

Example: Collinear Variables

Assuming a linear relationship, we fit the model of the form:

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 + b_5 x_5 + \varepsilon$$

1. Least squares (linear regression)
   $X^T X$ matrix : no solution if singular, OR poor estimates if nearly singular
2. Step-wise linear regression
   $$y = 1.0 x_1 + \varepsilon \qquad k = 1, 2, 3, 4, \text{or } 5$$
3. PLS
   $$y = 0.2x_1 + 0.2x_2 + 0.2x_3 + 0.2x_4 + 0.2x_5 + \varepsilon$$
4. Neural networks
   $$y = 0.63x_1 + 0.36x_2 + 0.09x_3 + 0.22x_4 + 0.30x_5 + \varepsilon$$

Can These Models be Used ?

- They are all wrong from a cause and effect point of view! They just model the correlation structure in the historical data set - Correlation models only
- But each model provides equally good predictions of *y* as long as the process remains the same (correlation structure stays constant).
- Models cannot be used to change the process in a way that would lead to a different correlation structure among the variables (e.g. optimization).
  - » Since predictions are not valid outside the correlation structure in the data used to build the model.

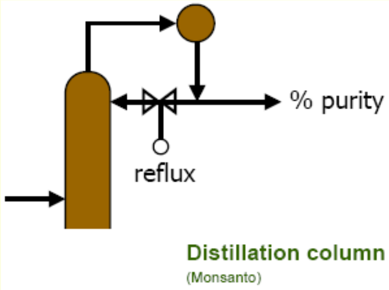
Can These Models be Used ?

- The problem is not the model but the data – **collinearity** is a result of the mode of operation. Real process data are often close to being collinear.
- To imply cause and effect for each variable, one needs designed experiments to generate the data.

Caution 2: Feedback Control

"True" cause and effect model:

% purity = $b_0 + \underbrace{b_1}_{\text{positive}}$ · reflux



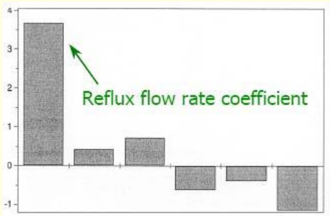
Regression model from data:

% purity = $c_0 - \underbrace{c_1}_{\text{negative!}}$ · reflux

Models the effect of the feedback control (i.e. the natural correlation structure in the data) not of the causal effect in the process!

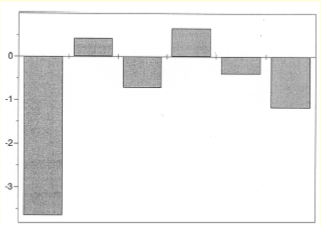
Effect of Feedback Control: Coefficient Plots

Open loop DOE data



The true cause and effect model

Closed-loop (feedback control) data



The correlation model

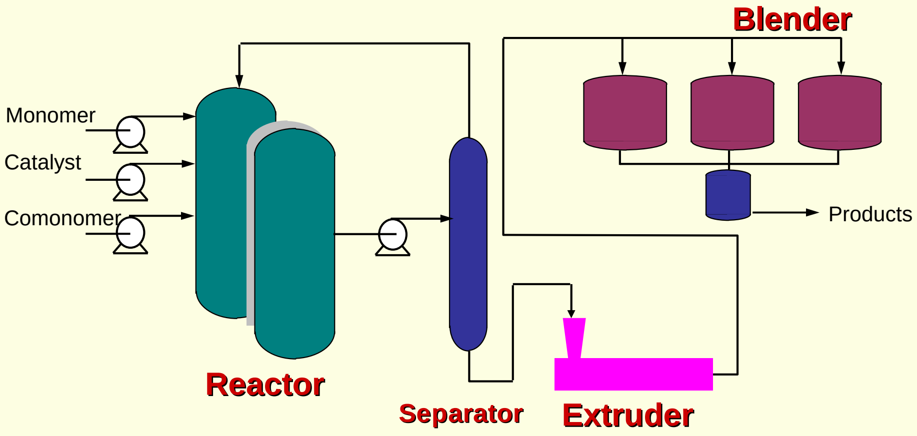
To Get Causal Models: Need DOE

- Historical data is good for
  - » Building model for monitoring
  - » Soft sensors
  - » Exploratory analysis of process operating problems
    - – eg. Score and loading plots, contribution plots, to drill-down to variables related to the problem.
- For causal model, eg. for control or optimization
  - » Need independent effects of all manipulated variables
  - » Generally need designed experiments in these variables
  - » Use statistical design (eg. factorials)
  - » Do not change one factor at a time

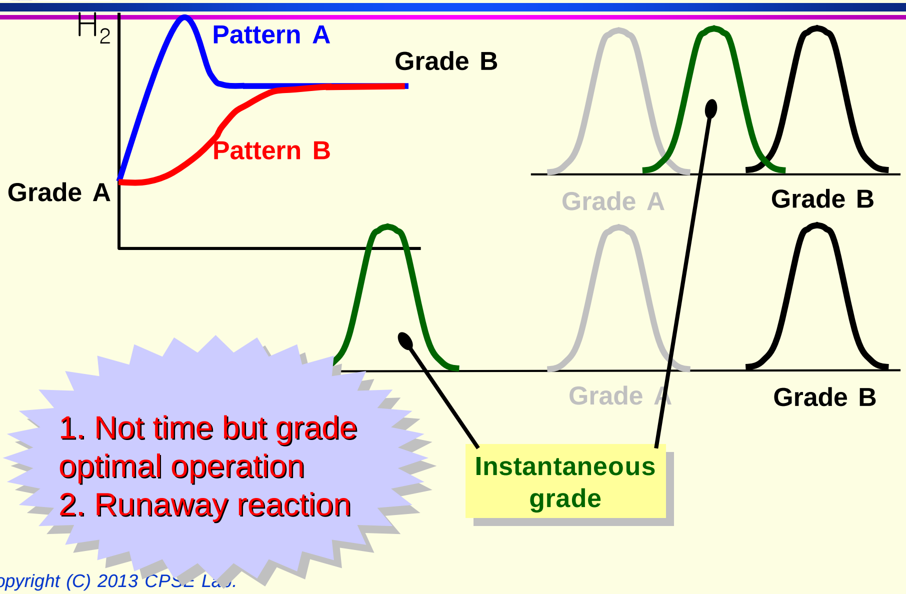
Example: Polymer Production Plant

- A polymerization process
  - » Around 50 available X-variables
  - » Melt index is the property of interest (Y)
  - » Company fit large amount of data with neural network regression model
  - » Good fit to the data but very poor predictions with new data
- Four grades (four Y set-points); 5, 10, 15, 20
- Operating data collected for each grade
  - » Appears to be simple problem:
    - – fit model to data, then
    - – use model for prediction
- But what data should be used? This is the key issue.

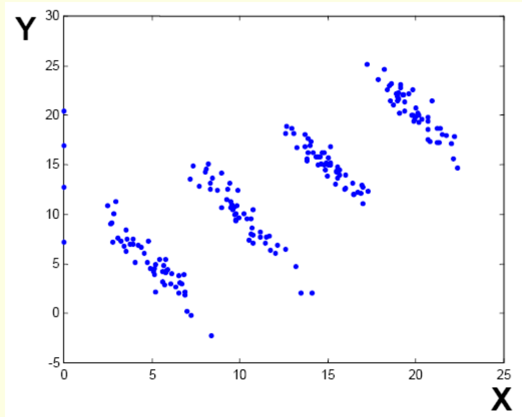
Polymer Production Plant



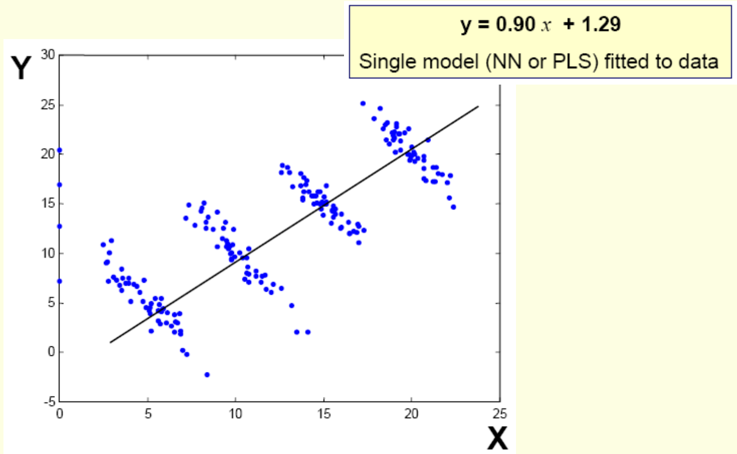
Grade Changeover Operation



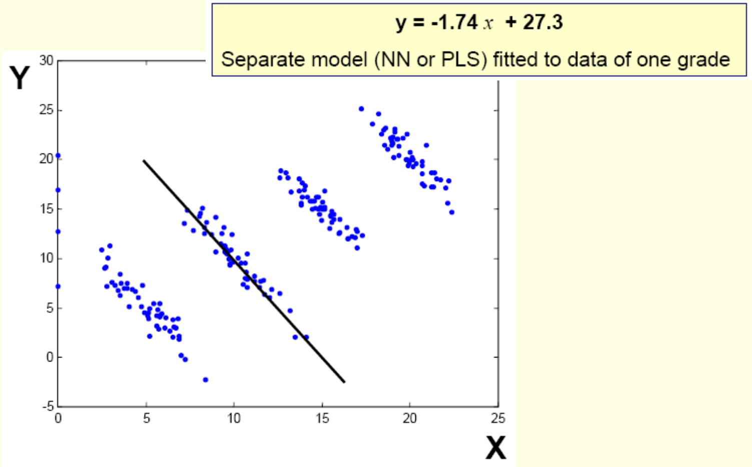
Y vs X Data for 4 Grades



Regression Fit of All Data



Regression Fit of Data from One Grade



Outline

1. **Background and Motivation**
2. **Current methods: a simple review**
3. **Some Challenges**
4. **Recent results**
5. **Concluding remarks**

Problems of Current Soft Sensors

[illegible]	[illegible]
burden of modeling itself	14 %
burden of data preprocessing	7 %
inadequate accuracy since installation	7 %
inadequate accuracy due to changes in operating conditions	7 %
difficulty in evaluating reliability	7 %
unjustifiable cost performance	7 %

Not Always Good?

(Kano and Ogawa, *J. Process Control*, 2010)

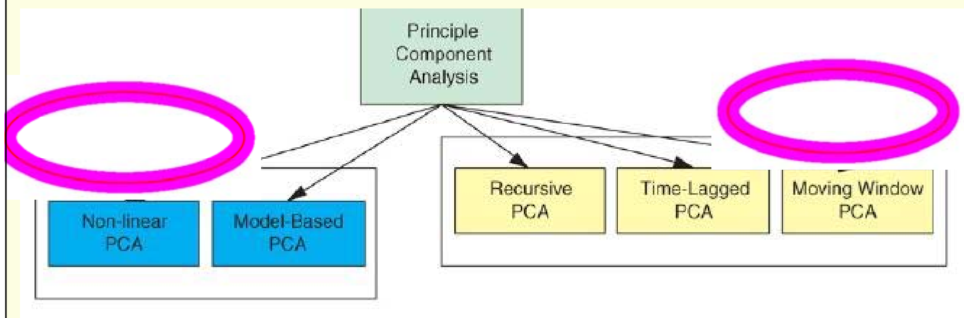
Reliability of a Soft Sensor

- **A fixed soft sensor is not enough (Reliability)**
  - » Changes in process characteristics and operating conditions
  - » Data are insufficient to build a global and good soft sensor for complex processes
  - » The most serious, practical problem is how to keep the high estimation performance
- **Model Maintenance is important but difficult**

Soft Sensor Adaptation

- **Adaptation of a soft sensor**
  - » Reconstruction of a global fixed soft sensor is difficult in practice
- **Recursive approaches**
  - » Update a model recursively
    - – Exponentially Weighted PLS (Dayal and MacGregor, 1997)
    - – Recursive PLS (Qin, 1998)
    - – Recursive LSSVR (Liu et al., 2009)
    - – Extensions and Applications
  - » Unable to cope with abrupt changes
  - » Insufficient for multi-mode processes

Improved Methods: PCA as an Example



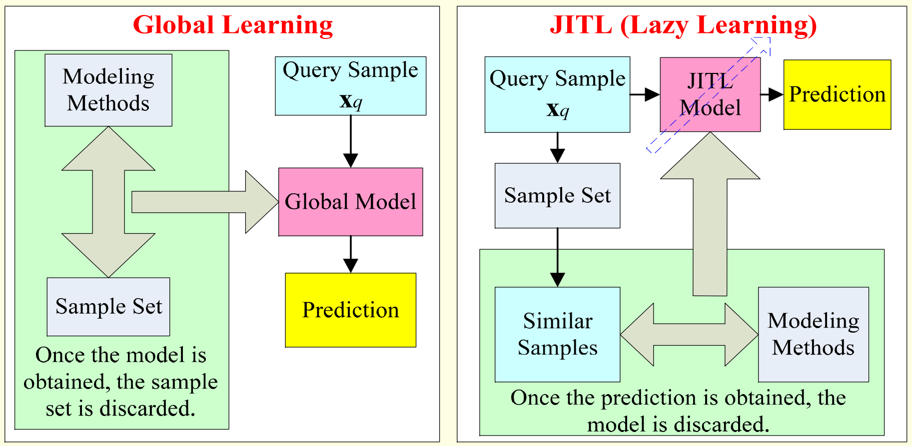
◆**Nonlinear:** $y = f(u)$ is essentially nonlinear for most complex chemical processes.
◆**Adaptive:** to capture different complicated characteristics

Kadlec P, et al. *Comput. Chem. Eng.*, **2009**, 33(4): 795-814.

Soft Sensor Adaptation

- **Alternative: Just-in-time learning (JITL) methods**
  - » Build a local model from neighbor samples stored in a database on demand (in a JIT manner)
  - » Similar input for similar output
- **Advantages of JITL methods**
  - » Treat changes in process characteristics
  - » Without model maintenance
  - » Local fitting ability for better prediction

Two Modeling Manners: Global Learning vs Just-in-time Learning



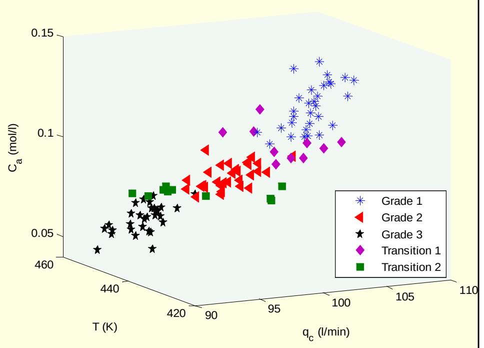
Yi Liu, et al. ***Ind. Eng. Chem. Res.***, **2012**, 51(11): 4313-4327.

Outline

1. **Background and Motivation**
2. **Current methods: a simple review**
3. **Some Challenges**
4. **Recent results**
5. **Concluding remarks**

Multi-Grade Polymerization Processes

- Increasing demand for product diversification results in more stringent specifications requirements in properties.
- Production of multi-grade products requires frequently changing operating conditions of reactors.
- Polymerization is a highly nonlinear process, which usually produces products, e.g., melt index (MI), with multiple quality grades.



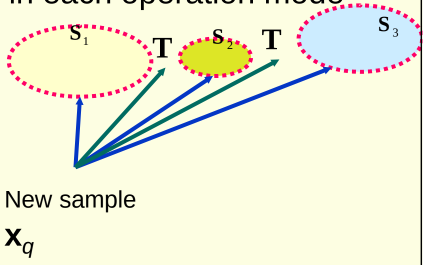
- Steady grades (a relatively simple task)
- Transitions (difficult)
- Whole processes (difficult)

Probabilistic Modeling Strategy

The probability of the new sample $\mathbf{x}_q$ in each operation mode

$$P\left(g\left|\mathbf{x}_q\right.\right)=\frac{n_g}{D^2_{q,\mathbf{x}_g}\sum_{i=1}^{G}\left(n_i/D^2_{i,\mathbf{x}_g}\right)}$$

$$g=1,\cdots,G$$



$$P\left(g\left|\mathbf{x}_q\right.\right)>\delta, g=1,\cdots,G$$ Steady grade: LSSVR model

$$\max\left\{P\left(g\left|\mathbf{x}_q\right.\right)\right\}\le\delta, g=1,\cdots,G$$ Transitions: JLSSVR model

Liu, Y. and Chen, J. (2013) "Dynamic Process Fault Monitoring Based on Neural Network and PCA, Integrated Soft Sensor Using Just-in-time Support Vector Regression and Probabilistic Analysis for Quality Prediction of Multi-Grade Processes" *J. Process Contr.* (In Press).

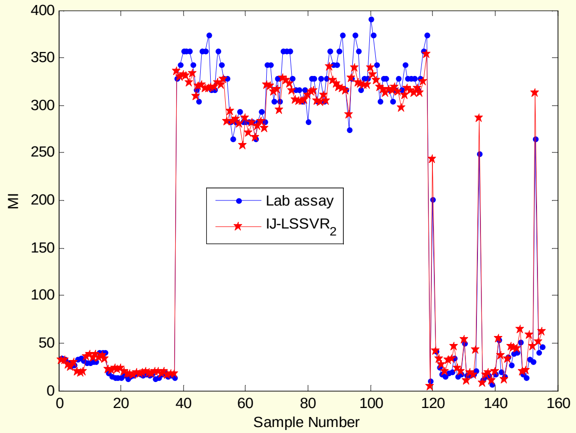
Prediction Performance Comparisons for Steady and Transitional Modes Using Different Soft Sensors

Model	Mode	RMSE	RE (%)
Proposed method	$\mathbf{S}$	**17.91**	**25.86**
	$\mathbf{S}_1$	**7.44**	22.61
	$\mathbf{S}_2$	4.54	32.47
	$\mathbf{S}_3$	22.05	6.35
	$\mathbf{S}_t$	**15.64**	**43.34**
WLSSVR $\hat{y}_q=\sum_{g=1}^{G}P\left(\mathbf{S}_g\left|\mathbf{x}_q\right.\right)\mathrm{LSSVR}_{\mathbf{S}_g}$	$\mathbf{S}$	18.42	40.75
	$\mathbf{S}_1$	7.58	**22.01**
	$\mathbf{S}_2$	5.55	39.50
	$\mathbf{S}_3$	**21.90**	**6.29**
	$\mathbf{S}_t$	18.17	75.81
WPLS $\hat{y}_q=\sum_{g=1}^{G}P\left(\mathbf{S}_g\left|\mathbf{x}_q\right.\right)\mathrm{PLS}_{\mathbf{S}_g}$	$\mathbf{S}$	20.18	41.11
	$\mathbf{S}_1$	8.88	26.85
	$\mathbf{S}_2$	**4.34**	**31.09**
	$\mathbf{S}_3$	23.55	6.79
	$\mathbf{S}_t$	21.21	78.18

$$\mathrm{RMSE}=\sqrt{\sum_{i=1}^{k}\left(\hat{y}_i-y_i\right)^2/k}$$

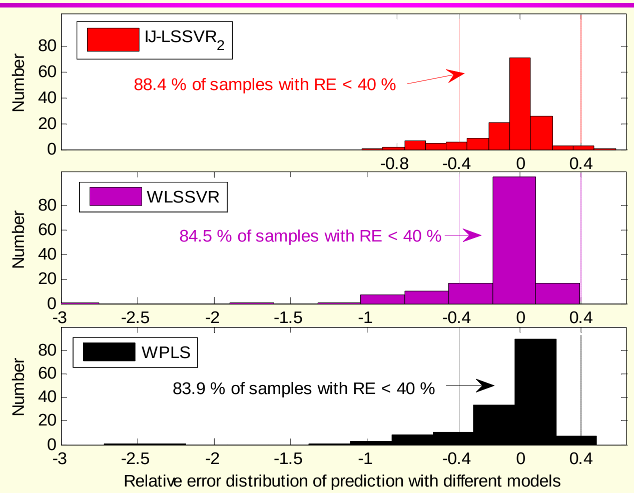
$$\mathrm{RE}=\sqrt{\sum_{i=1}^{k}\left(\frac{\hat{y}_i-y_i}{y_i}\right)^2/k}$$

Online Prediction of MI



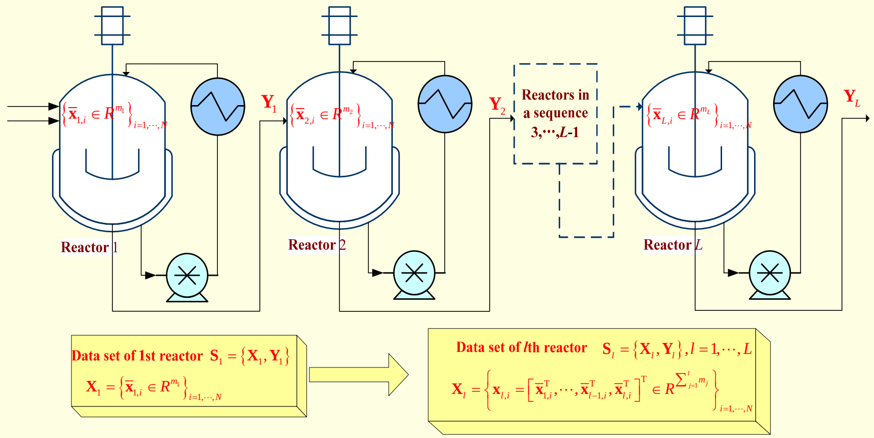
Online prediction of MI with proposed soft sensor method

Comparison of Relative Prediction Errors Distribution of MI



Comparison of relative prediction errors distribution of MI with IJ-LSSVR$_2$, WLSSVR and WPLS methods

Sequential-Reactor-Multi-Grade (SRMG) Processes

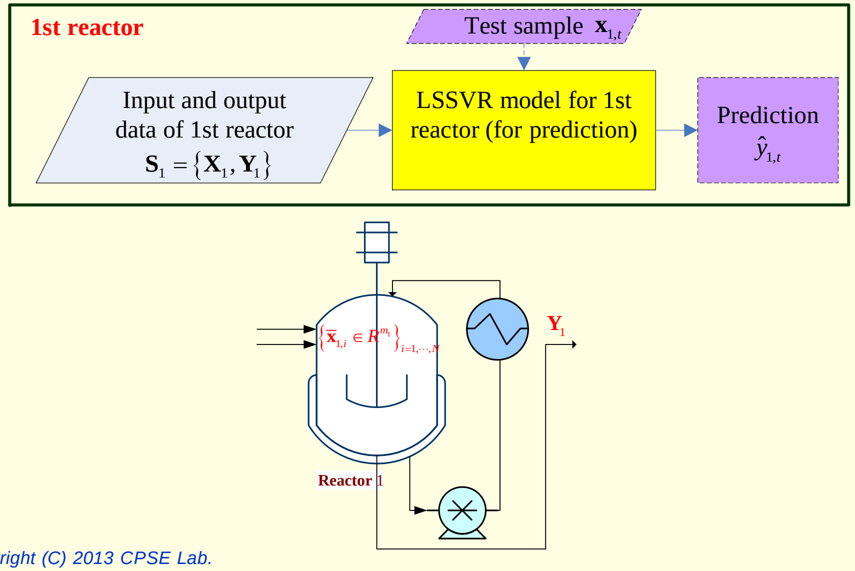


A simplified flowchart of a sequential process with *L* operating reactors and the related modeling data set of the *l*th reactor

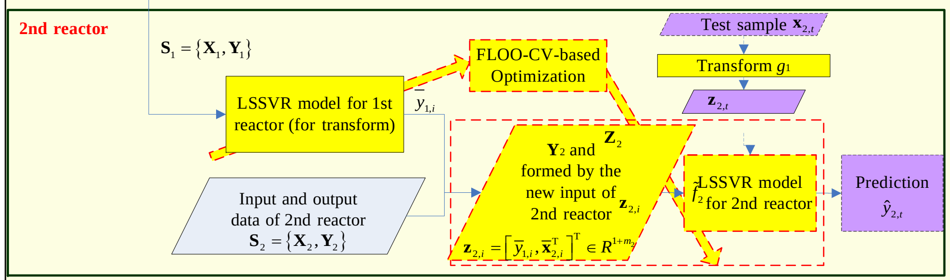
Difficulties in Soft Sensor for SRMG

- Difficult to choose suitable input variables for modeling of an SRMG process, especially for the last reactor.
- Input variables often show co-linearity and are combined with noise.
- Principal component analysis (PCA) can extract latent variables (LVs) as a preprocessing step, and then followed by a soft sensor model. However, the important LVs may capture most of the process variation, but may not necessarily explain quality properties.
- Multi-grade issues

Proposed Soft-Sensor for SRMG

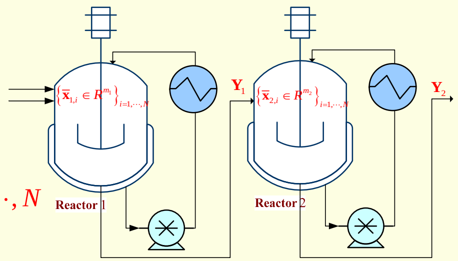


Proposed Soft-Sensor for SRMG

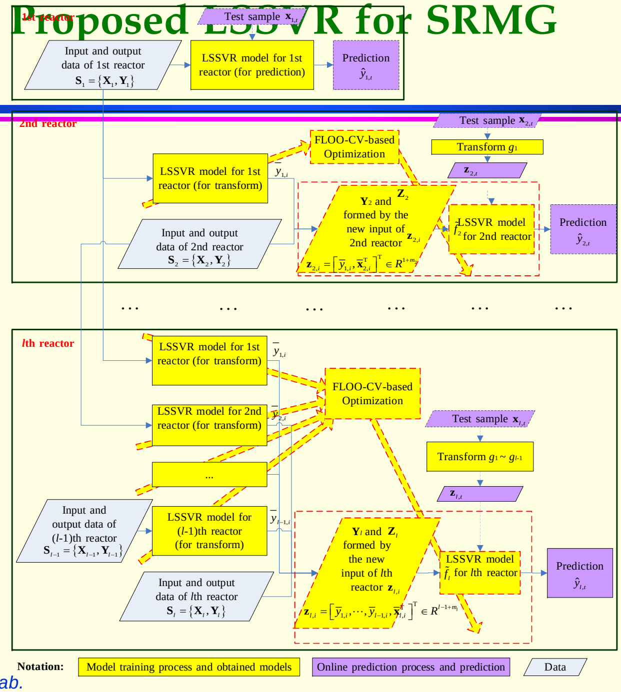


The input variables in the previous reactors can be substituted for "virtual" variables obtained from the transform models

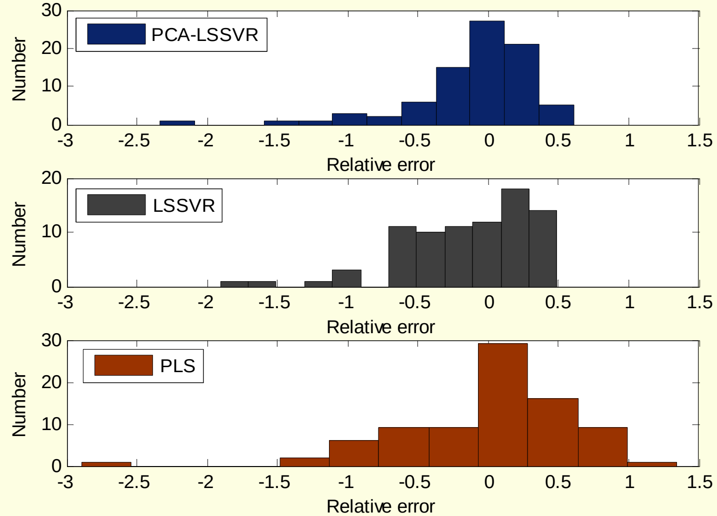
$$\mathbf{z}_{l,i} = \left[\bar{y}_{1,i}, \cdots, \bar{y}_{l-1,i}, \bar{\mathbf{x}}_{l,i}^{\mathrm{T}}\right]^{\mathrm{T}} \in R^{l-1+m_l}, \quad i = 1, \cdots, N$$



Proposed LSSVR for SRMG

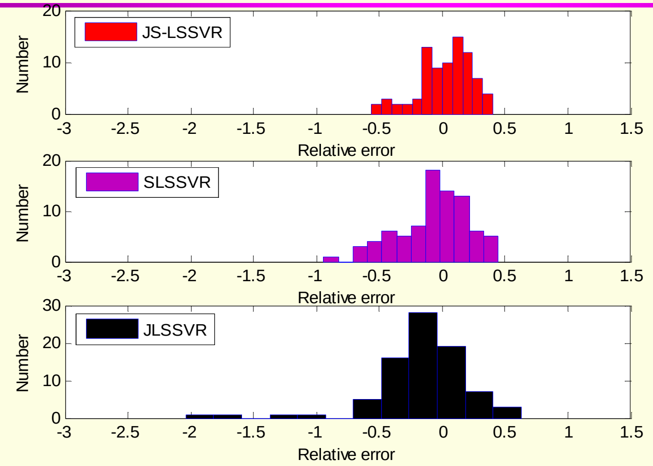


Industrial SRMG Example



Traditional methods
- **PCA-LSSVR**
- **LSSVR**
- **PLS**

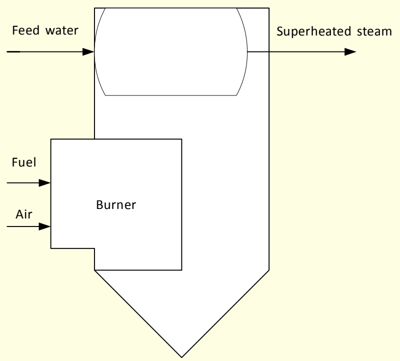
Industrial SRMG Example

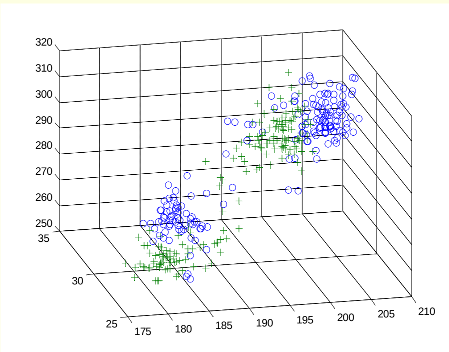


Proposed methods
- ➢JS-LSSVR
- ➢SLSSVR
- ➢JLSSVR

Liu, Y. and Chen, J. (2013) "Development of Soft-Sensors for Online Quality Prediction of Sequential-Reactor-Multi-Grade Industrial Processes" ***Chem. Eng. Sci.*** **(Revised).**

Model with Confidence Intervals



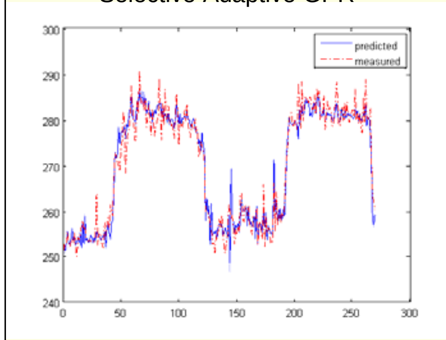


The cogeneration plant produces high temperature, pressurized steam for power generation, and exhaust (low pressure steam) is later used for downstream processes.

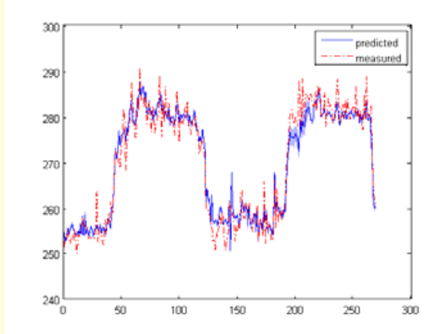
Two Operation Modes

- The plant is operated at the high load from 8 in the morning till 8 in the evening; the other time it is operated at the low load.

Selective Adaptive GPR

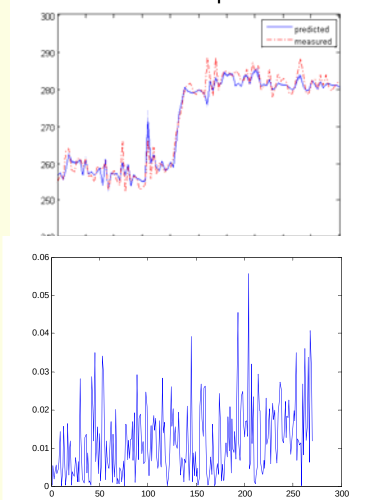


Conventional GPR

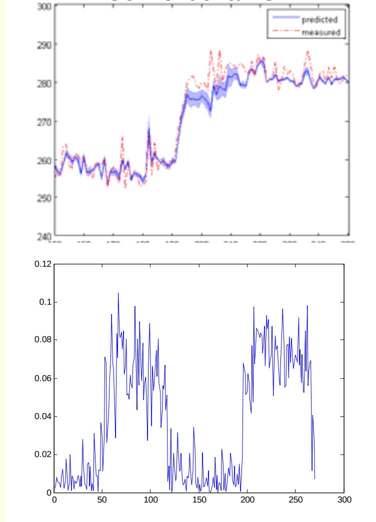


Enlarged part of the sample (200 to 300)

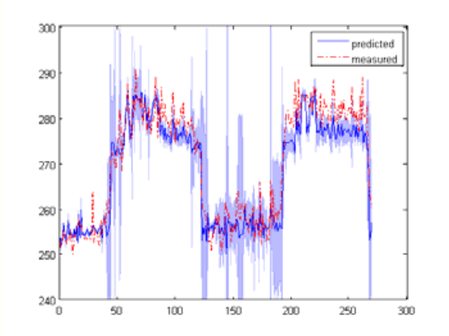
Selective Adaptive GPR

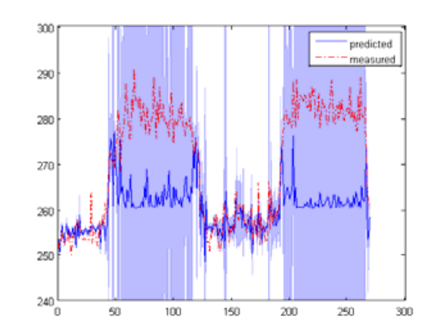


Conventional GPR



Operating modes (from low and high loads)





Chan, L.L.T., Liu, Y. and Chen, J. (2013) "Nonlinear System Identification with Selective Recursive Gaussian Process Models " *Ind. Eng. Chem. Res.* (Revised).

Conclusions

- Data-driven soft sensors is available
  - » Widely accepted in industries
  - » Crucial for inferential control
- Reliability and adaptation
- Just-in-time learning is alternative and promising
- Incorporating process characteristics is important
  - » Multi-grade
  - » Sequential-reactor-multi-grade
- Remaining problems

Nature of Process Data

- High dimensional
  - – Many variables measured at many times
- Non-causal in nature
  - – No cause and effect information among individual variables
- Non-full rank
  - – Process really varies in much lower dimensional space
- Missing data
  - – 10 – 30 % is common (with some columns/rows missing 90%)
- Low signal to noise ratio
  - – Little information in any one variable
- Outliner
- Non-Gaussian
- Multi-Rate