# Data-based Industrial Soft-sensor Design via Optimal Subset Selection

**M. Mojto[1], K. Ľubušký[2], M. Fikar[1], R. Paulen[1]**

[1]Faculty of Chemical and Food Technology, Slovak University of Technology in Bratislava, Slovakia

[2]Slovnaft, a.s., Bratislava, Slovakia

## Goals

Inferential (or soft) sensors infer rarely measured or completely unmeasured variables. The main challenge in designing an inferential sensor is to select a correct structure represented by sensor input variables. This work is focused on the design of an inferential sensor for an industrial depropanizer column. We study the effectiveness of various subset selection (SS) methods that consider different model-overfitting criteria.

## Plant and Data Description

### Industrial Data

- 2 years of refinery data
- Online sensors
  - 9 measured variables
  - 38,360 measurements
- Infrequent lab analysis
  - bottom composition $x_\mathrm{B}$
  - 177 measurements

### Reference (Ref) Inferential Sensor

- Current sensor in the refinery
- $\hat{x}_\mathrm{B} = a_1 T_{37} + a_2 p_\mathrm{B} + a_3 Q_\mathrm{B}/F$
- $a_{1-3}$ are model parameters

Is this sensor optimal?

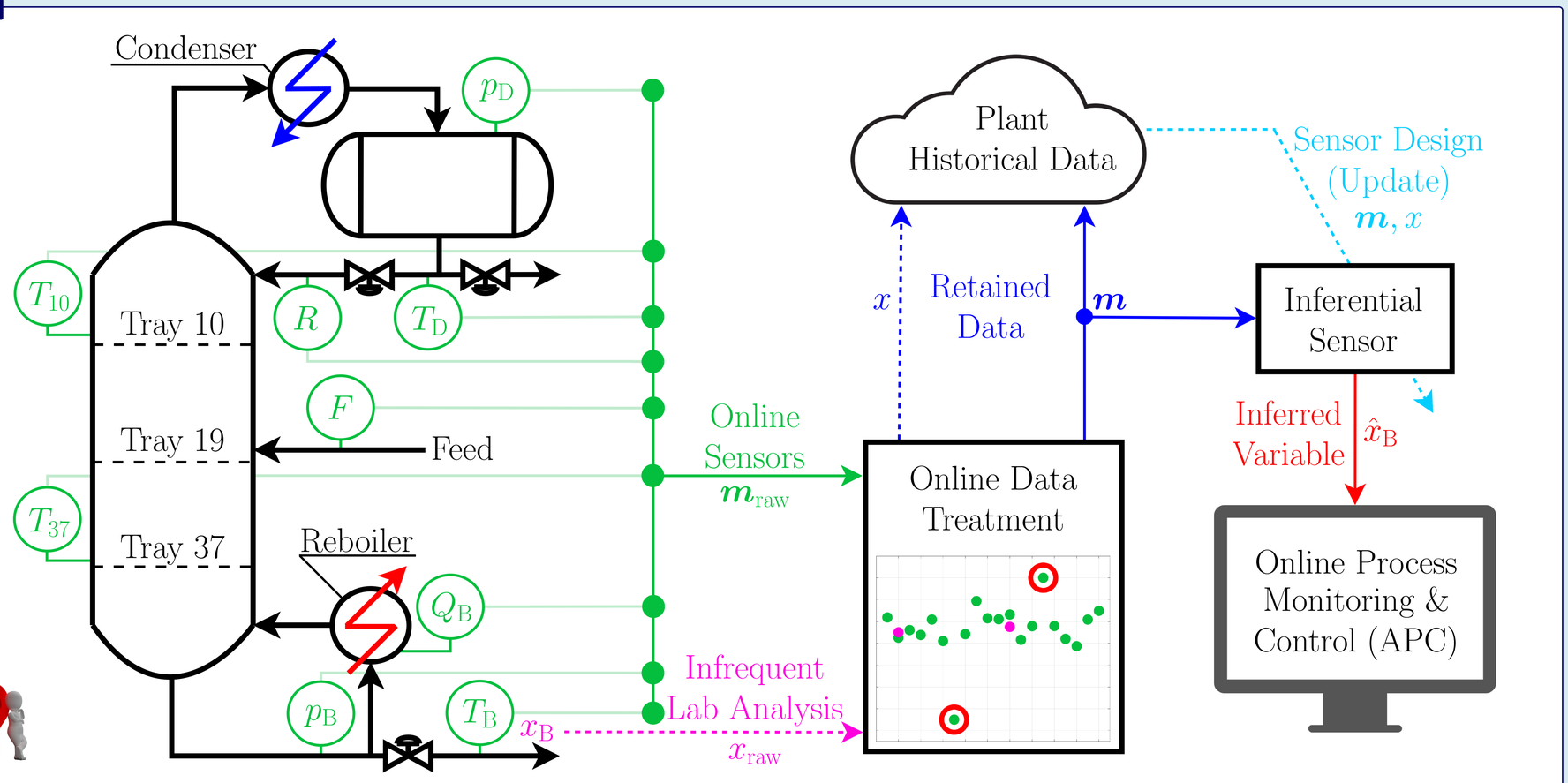


Fig. 1: Online process monitoring of a depropanizer column (part of the FCC unit) using an inferential sensor.

## Design of Inferential Sensors by Subset Selection

Studied model $\hat{x}_\mathrm{B} = \boldsymbol{a}^\intercal \boldsymbol{m}$

Input vector candidates $\boldsymbol{m} = \big(R, F, Q_\mathrm{B}, T_{10}, T_{37}, T_\mathrm{D}, T_\mathrm{B}, p_\mathrm{D}, p_\mathrm{B}, R/F, Q_\mathrm{B}/F\big) \in \mathbb{R}^{n_\mathrm{p}=11}$

**1. Part:** Design of the sensor structure

Iteration through MIQP

$$\min_{\boldsymbol{m}} f_\mathrm{overfit}(\boldsymbol{m}, \boldsymbol{a}^*) = \left\{ \begin{array}{c} R^2_\mathrm{adj},\ \mathrm{AIC_c},\ \mathrm{BIC} \\ \text{or} \\ \text{cross-validation} \end{array} \right. \implies \boldsymbol{m}^*$$

Number of input variables: $n_\mathrm{p}^*$

**2. Part:** Calculation of the sensor parameters

$$\min_{\boldsymbol{a}} f_\mathrm{LS}(\boldsymbol{m}^*, \boldsymbol{a}) = \sum (x_i - \boldsymbol{a}^\intercal \boldsymbol{m}_i^*)^2 \implies \boldsymbol{a}^*$$

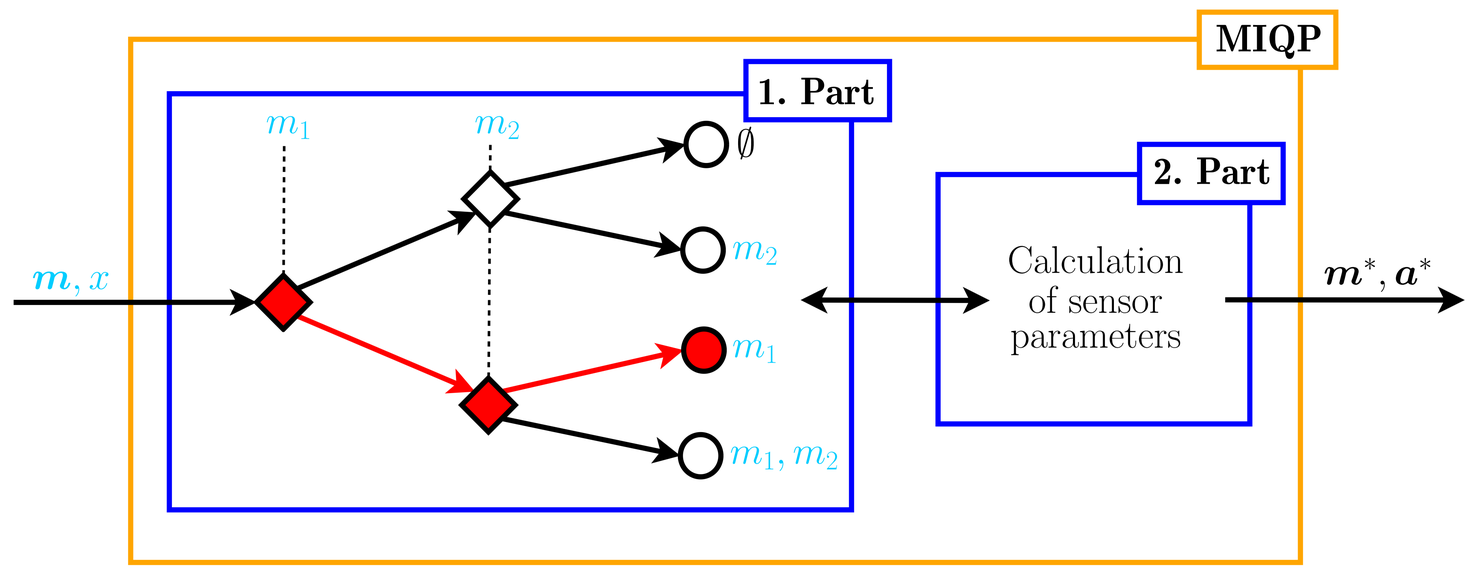


Fig. 2: Principle of subset selection.

## Results

Performance of inferential sensors on testing data

↳ complexity ($n_\mathrm{p}^*$) and accuracy (RMSE)

| | $R^2_\mathrm{adj}$ | $\mathrm{AIC_C}$ | BIC | cross-validation | Ref |
|---|---|---|---|---|---|
| $n_p^*$ | 9 | 4 | 4 | 4 | 3 |
| RMSE | 0.110 | 0.106 | 0.106 | 0.106 | 0.128 |

Soft sensor designed by $\mathrm{AIC_C}$, BIC and cross-validation

↳ $\hat{x}_\mathrm{B} = a_1 T_{10} + a_2 T_{37} + a_3 T_\mathrm{B} + a_4 Q_\mathrm{B}/F$

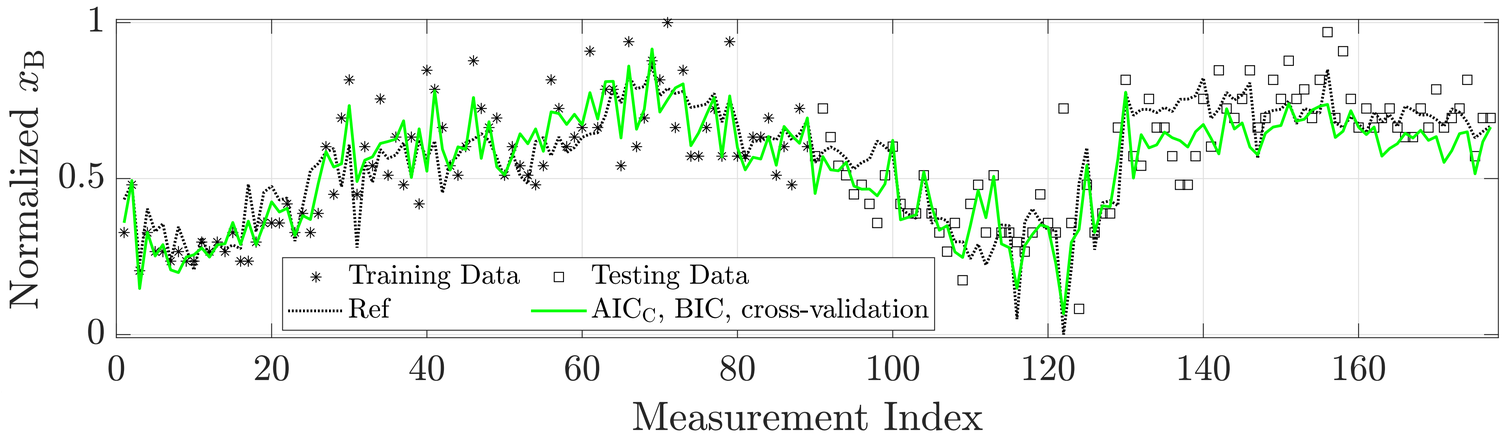


Fig. 3: Inferential sensors design on time series data.

## Conclusions

We analyzed the effectiveness of optimal subset selection to design an inferential sensor. The structure of the inferential sensor suggested by SS with $\mathrm{AIC_C}$, BIC and cross-validation is the same. The results indicate accuracy improvement of these inferential sensors compared to Ref sensor by around 15 %. Our further research confirms that SS suggests a less complex inferential sensor than PCA. The performance of SS appears to be comparable to LASSO.

**Acknowledgements**: This research is funded by the Slovak Research and Development Agency under the project APVV SK-FR-2019-0004 and by the Scientific Grant Agency of the Slovak Republic under the grant 1/0691/21.