Methanex Plant Optimization

Towards Responsible care and Sustainability

Deepak Nangamuthu Chanthiramathi

Why Optimize Efficiency?

Company

- Largest producer of Methanol in the world
- Uses 45% of New Zealand's natural gas output
- Contributes $84 million annually to The New Zealand economy

Context

- Constrained Natural Gas
- Global competitors producing Methanol more efficiently
- Sub-Optimal Current Production

Solution

- A reliable model to predict efficiency
- Find optimal values for pv to maximize Efficiency
- Provide insights on the impact of process variables

Challenges deep-dive

Challenge 1

Multicollinearity

The process variables are not independent from each other

Challenge 2

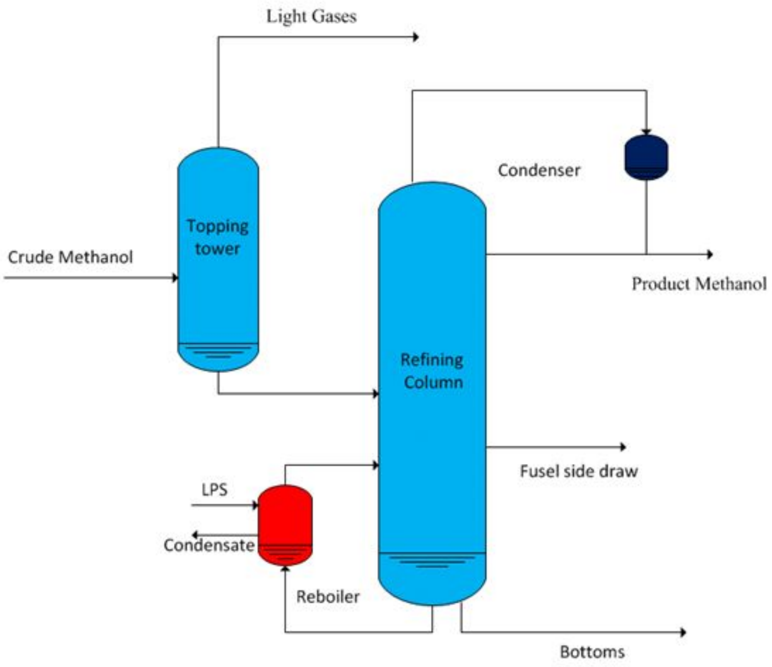
Reliability

A model which is reliable on all datasets across the plant

Challenge 3

Understanding the impact

Insights between the process variables and efficiency

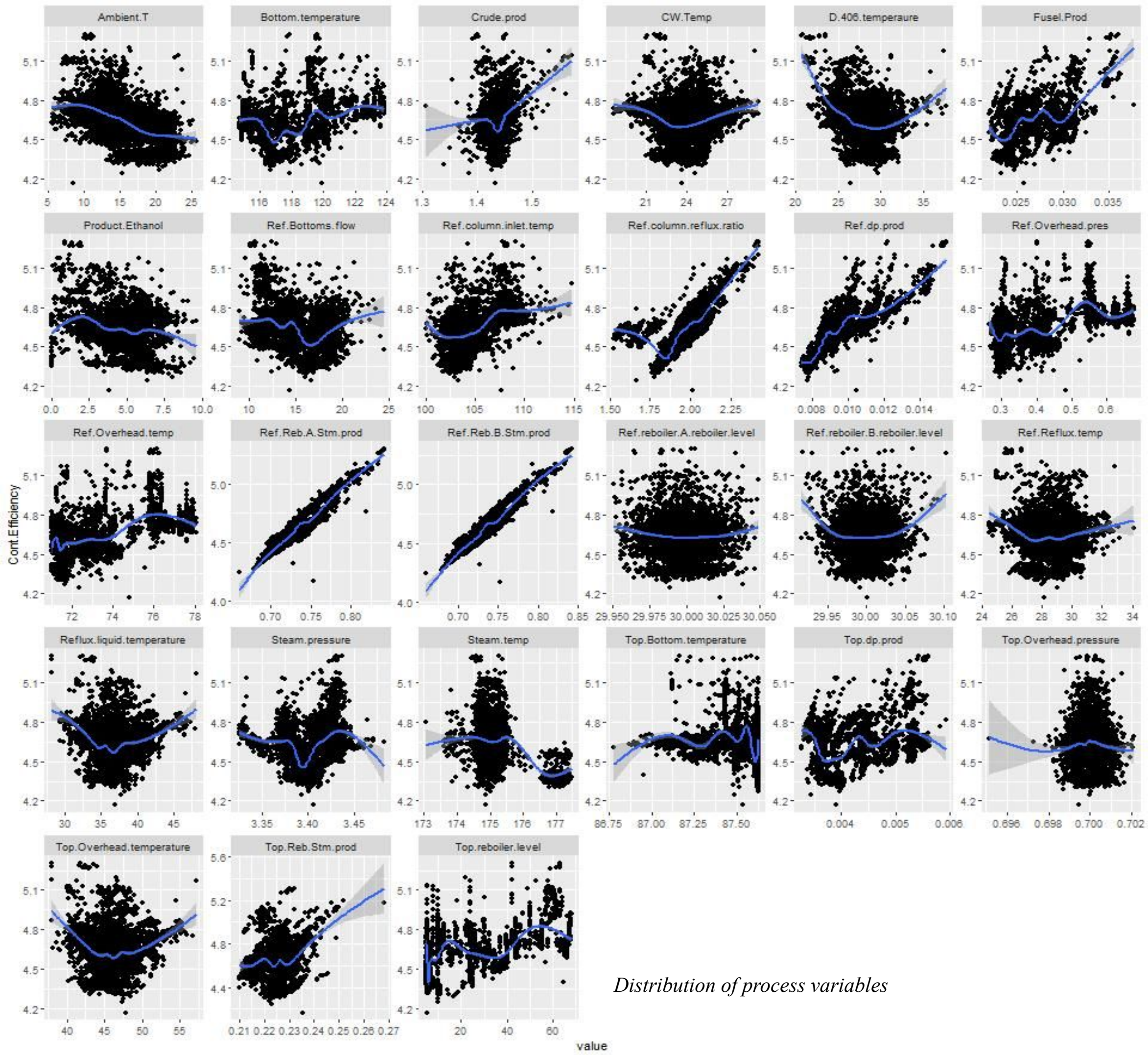


Methanol Distillation process, picture courtesy of Methanex Corporation

Distillation Control

- Process variables like temperature, pressure, flow rate and level within a distillation column affect one another
- The sensor measures the process variable from the plant and the transmitter sends the information to the DCS
- Historian records the real time data from DCS

Preliminary Analysis



Distribution of process variables

Data collection

- Obtained in Excel format from the PI DataLink

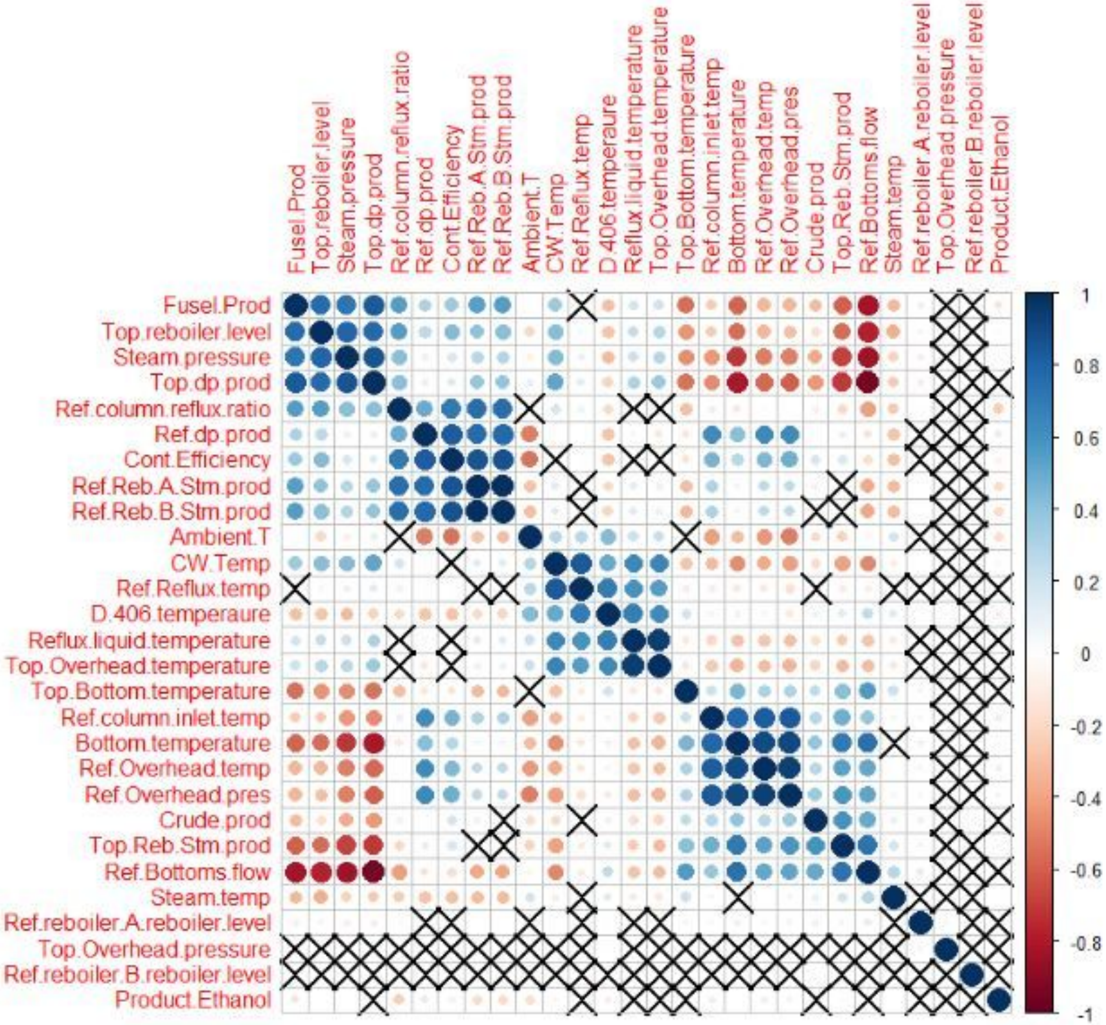
Cleaning

- Data set comprises of plant running in normal operating conditions
- Partial Deletion: Abnormal ranges of process variables are removed manually

Data Distribution

- No non-linear relationship detected. Plenty of no relationships are seen.

Correlation Matrix



- Why?

  Summarize data with an aim of revealing patterns

- A correlation coefficient shows how much one variable change in function of the other
- Correlation does not mean causation
- Pearson correlation: comparing process variables to find a linear relationship
- Data multicollinearity: Independent process variables in data are correlated

Applying ML Models

```
Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                  7.311e-01  1.108e+00   0.660 0.509236
Top.Reb.Stm.prod             3.598e+00  1.202e-01  29.923  < 2e-16 ***
Top.reboiler.level           2.657e-03  5.990e-05  44.353  < 2e-16 ***
D.406.temperaure            -1.979e-03  6.219e-04  -3.182 0.001469 **
Crude.prod                  -1.226e-01  3.140e-02  -3.903 9.61e-05 ***
Ref.column.inlet.temp       -8.445e-04  4.090e-04  -2.065 0.038982 *
Ref.column.reflux.ratio      4.427e-02  5.835e-03   7.587 3.78e-14 ***
Ref.Reb.A.Stm.prod           2.240e+00  2.084e-01  10.750  < 2e-16 ***
Ref.reboiler.A.reboiler.level -4.279e-02  2.704e-02  -1.583 0.113541
Ref.Reb.B.Stm.prod           2.837e+00  2.099e-01  13.516  < 2e-16 ***
Ref.reboiler.B.reboiler.level 7.079e-04  1.945e-02   0.036 0.970973
Steam.temp                   7.964e-04  8.579e-04   0.928 0.353307
Steam.pressure               2.955e-01  2.811e-02  10.513  < 2e-16 ***
CW.Temp                      7.092e-03  9.722e-04   7.295 3.37e-13 ***
Ref.Bottoms.flow             1.605e-03  4.691e-04   3.422 0.000626 ***
Fusel.Prod                  -5.727e-01  2.660e-01  -2.153 0.031384 *
Reflux.liquid.temperature   -9.717e-04  6.324e-04  -1.536 0.124484
Top.Overhead.temperature     1.163e-03  5.272e-04   2.206 0.027414 *
Top.Overhead.pressure       -1.075e+00  6.637e-01  -1.620 0.105240
Top.Bottom.temperature       1.346e-02  2.337e-03   5.761 8.80e-09 ***
Top.dp.prod                 -6.060e+01  4.151e+00 -14.600  < 2e-16 ***
Product.Ethanol              3.425e-03  2.862e-04  11.970  < 2e-16 ***
Ref.Reflux.temp             -8.393e-03  8.962e-04  -9.365  < 2e-16 ***
Ref.Overhead.temp           -5.918e-04  2.298e-03  -0.258 0.796788
Ref.Overhead.pres            1.768e-01  4.263e-02   4.147 3.42e-05 ***
Bottom.temperature          -4.328e-03  1.228e-03  -3.525 0.000427 ***
Ref.dp.prod                  2.874e+00  1.797e+00   1.599 0.109860
Ambient.T                   -2.897e-03  1.833e-04 -15.807  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03274 on 5940 degrees of freedom
Multiple R-squared:  0.9582,    Adjusted R-squared:  0.9581
F-statistic:  5048 on 27 and 5940 DF,  p-value: < 2.2e-16
```

LINEAR REGRESSION
Summary

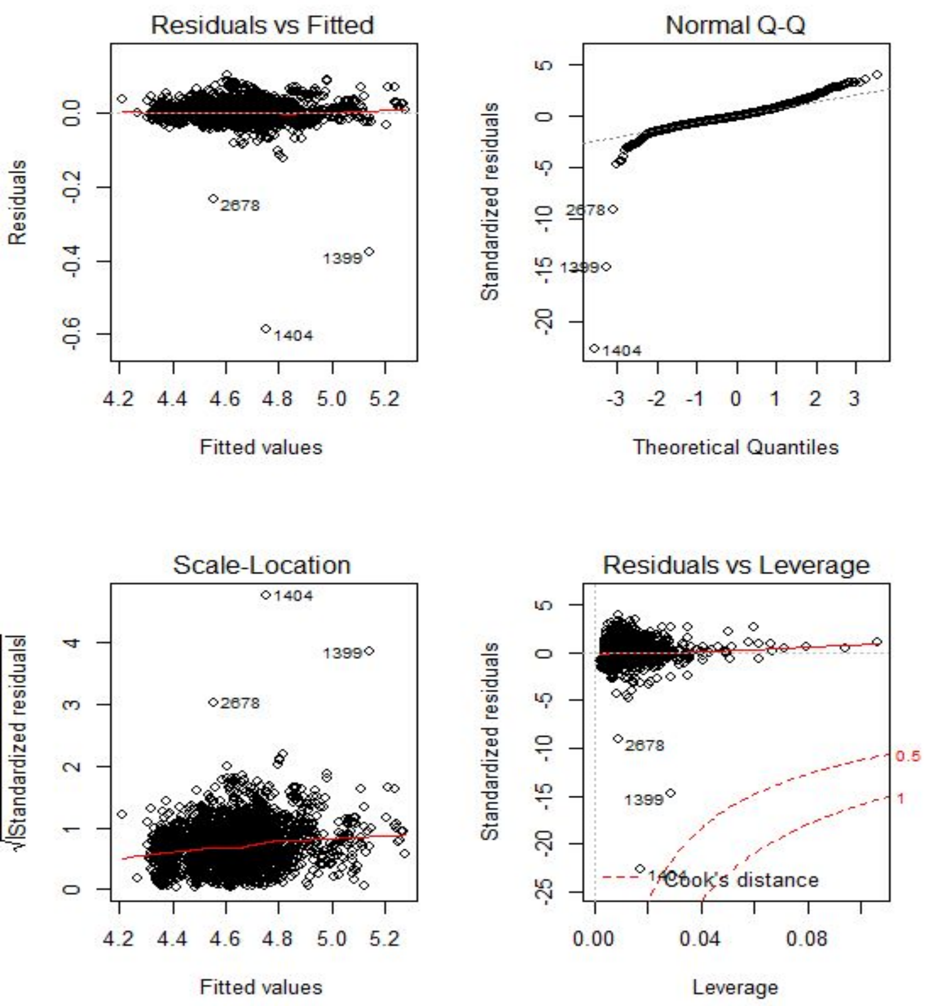
- A linear regression is amongst the simplest possible model. It tries to approximate a variable as a sum of the other variables multiplied by a coefficient, added to a constant

- OLS regression :
  May yield unstable results in presence of important correlations between explanatory variables

  Do not scale well over a wide range of operating conditions and large disturbances

  Possible to fit data in higher dimension but resulting model poor in new dataset

- Why?
  Interpret the model output and not for fitting the model



Residual Plots

Residuals vs Fitted

The spread of residuals around the horizontal line without any distinct patterns is a good indication that it is linear

Scale-Location

This reveals if the residuals are spread equally along the range of predictors thus checking the assumption of equal variance. A horizontal line with equally spread points is a good indicator.

Normal QQ

This plot is used to show if the residuals are normally distributed. Here, most of the residuals follow the straight line

Residuals vs Leverage

Not all outliers tend to be influential. Here, there are no cases lying outside the Cook's distance

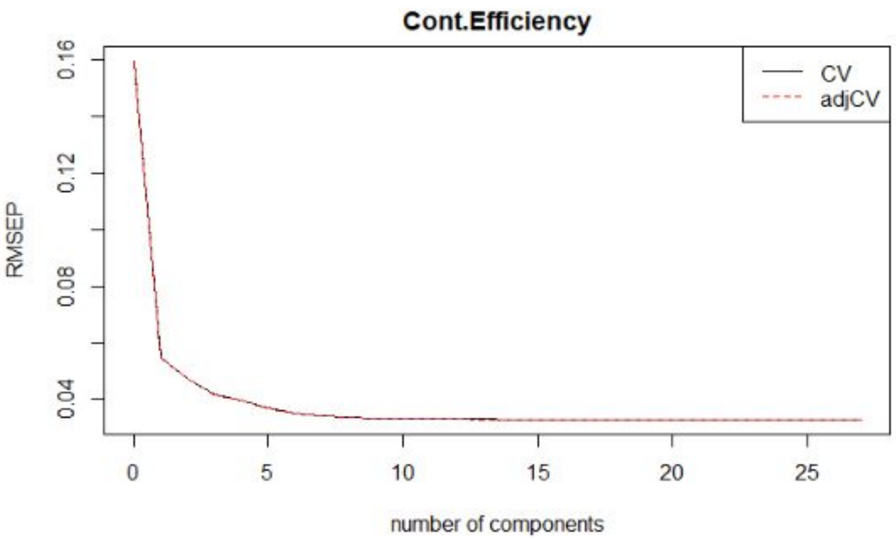
```
Data:    X dimension: 2746 27
         Y dimension: 2746 1
Fit method: kernelpls
Number of components considered: 27

VALIDATION: RMSEP
Cross-validated using 10 random segments.
       (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps  8 comps
CV          0.1594  0.04975  0.03794  0.03508  0.03288  0.03018  0.02860  0.02752  0.02684
adjCV       0.1594  0.04974  0.03791  0.03510  0.03287  0.03016  0.02858  0.02750  0.02682
       9 comps  10 comps  11 comps  12 comps  13 comps  14 comps  15 comps  16 comps
CV     0.02660   0.02653   0.02649   0.02644   0.02641   0.02639   0.02638   0.02635
adjCV  0.02659   0.02651   0.02647   0.02643   0.02639   0.02637   0.02637   0.02633
       17 comps  18 comps  19 comps  20 comps  21 comps  22 comps  23 comps  24 comps
CV      0.02634   0.02631   0.02631   0.02631   0.02631   0.02630   0.02631   0.02631
adjCV   0.02632   0.02629   0.02629   0.02629   0.02629   0.02628   0.02629   0.02629
       25 comps  26 comps  27 comps
CV      0.02630   0.02631   0.02631
adjCV   0.02628   0.02629   0.02629

TRAINING: % variance explained
                 1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps  8 comps
X                  20.75    31.15    52.79    66.52    70.00    72.92    75.05    77.03
Cont.Efficiency    90.30    94.40    95.22    95.81    96.49    96.85    97.08    97.23
                 9 comps  10 comps  11 comps  12 comps  13 comps  14 comps  15 comps
X                  79.25     81.48     83.93     85.52     86.85     89.31     91.22
Cont.Efficiency    97.28     97.29     97.30     97.31     97.32     97.33     97.33
                 16 comps  17 comps  18 comps  19 comps  20 comps  21 comps  22 comps
X                   92.38     95.30     95.78     97.15     98.79     98.99     99.13
Cont.Efficiency     97.34     97.34     97.34     97.34     97.34     97.35     97.35
                 23 comps  24 comps  25 comps  26 comps  27 comps
X                   99.60     99.75     99.81     99.99    100.00
Cont.Efficiency     97.35     97.35     97.35     97.35     97.35
```

PLSR Summary results

Dimension Reduction

Partial Least Square Regression

- The approach involving reducing the number of random variables under consideration by obtaining a set of principal features
- Partial Least Squares (PLS) is useful for constructing models when there are many factors which are collinear
- Reduce overfitting, Solving Multicollinearity, Less Computational cost
- The number of predictors is chosen by looking at the component number which adequately explains both the predictor and response variances



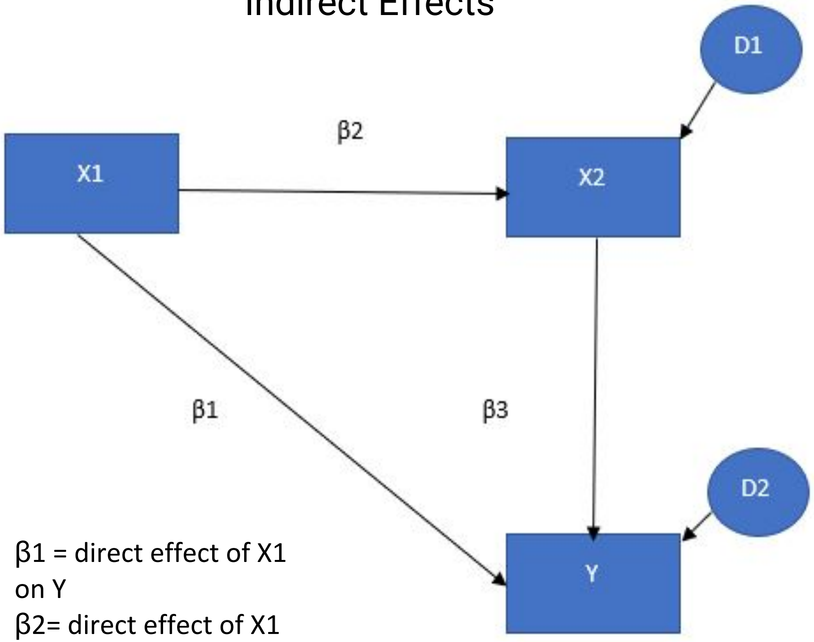

```
Data:    X dimension: 2746 27
         Y dimension: 2746 1
Fit method: kernelpls
Number of components considered: 5
TRAINING: % variance explained
                 1 comps  2 comps  3 comps  4 comps  5 comps
X                  20.75    31.15    52.79    66.52    70.00
Cont.Efficiency    90.30    94.40    95.22    95.81    96.49
```

PLSR Summary for 5 component

Latent variables

- Hidden variables that are not observed directly but inferred from models that are observed
- Unmeasured pure variables or true scores that are free from errors

Indirect Effects



β1 = direct effect of X1 on Y
β2= direct effect of X1 on X2
β3 = direct effect of X2 on Y
β2*β3= indirect effect of X1 on Y
β1+(β2*β3)= total effect of X1 on Y

Latent variable modelling

- Inferential sensors can be built using the latent variables obtained from PLS models

Potential Applications

- Improved process understanding and monitoring
- Troubleshooting Process problems
- Optimizing: new operating point
- Dealing with higher dimensional data

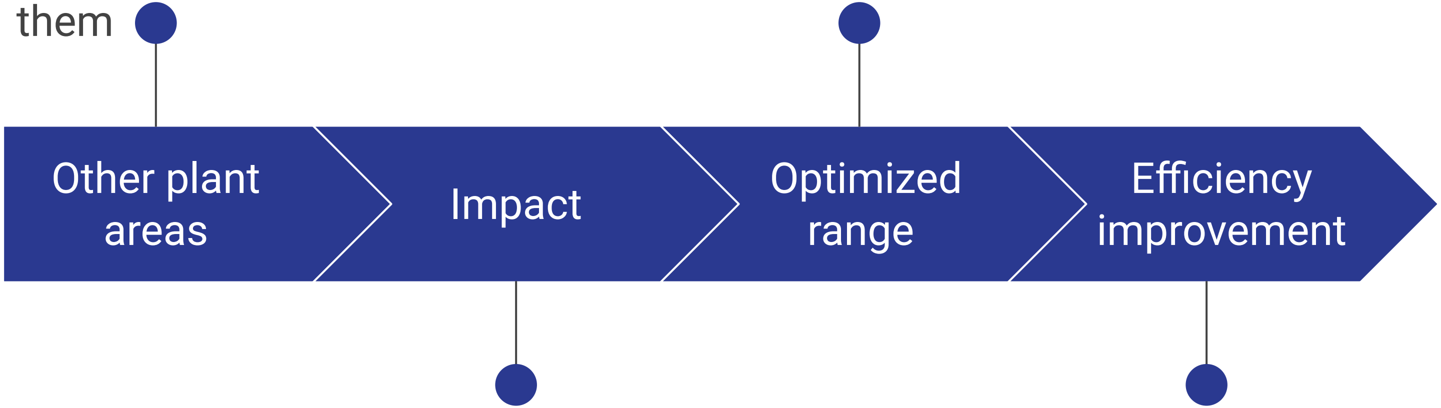
Summary - Modelling a soft sensor pathway

Selection of historical data → Filtering & Outlier detection → PLS → Latent variable modelling → Model Validation

Moving Forward

Fitting datasets from different plant areas and finding a model which best fits across them

Finding Optimal ranges for process variables which improves efficiency



Understanding impact of Process variables on Efficiency

Improve efficiency at lower cost

Thank you!

Nick Ward, Lecturer
Franco, Process Engineer
Bronwyn, Energy Efficiency Engineer

“Energy costs are getting higher and the cheapest energy is the energy you don't use”