KU LEUVEN


# Optimal Experimental Designs for Process Robustness Studies

**Ying Chen**

(Research funded by a grant from GlaxoSmithKline Biologicals SA, Belgium)

**Supervisor: Prof. dr. Peter Goos**
(KU Leuven, University of Antwerp)

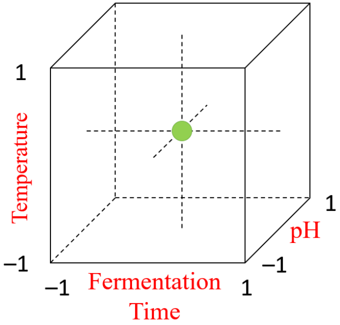
**Co-supervisor: Dr. Bernard G. Francq**
(An employee of the GSK group of companies)

# Overview

- Robust process
- Design selection criteria
- Design examples and evaluation
- Summary

# Robust Process

● **Target operating condition**: a specific combination of settings of the process parameters

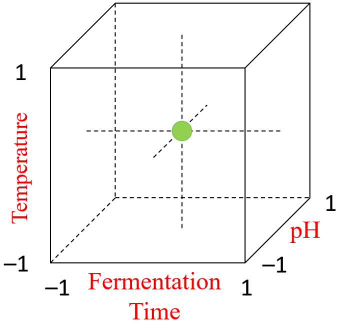




KU LEUVEN

# Robust Process

● **Target operating condition**: a specific combination of settings of the process parameters



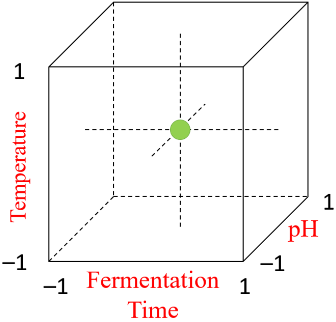

- Unavoidable deviations around the target condition during routine production



KU LEUVEN

# Robust Process

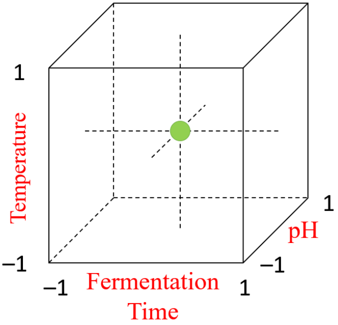
● **Target operating condition**: a specific combination of settings of the process parameters




- Unavoidable deviations around the target condition during routine production
- Are the responses very different from the response at the target operating condition? If not, the process is robust

# Robust Process

**Target operating condition**: a specific combination of settings of the process parameters




- Unavoidable deviations around the target condition during routine production
- Are the responses very different from the response at the target operating condition? If not, the process is robust

⇒ Precise prediction of the difference in responses is important

# Design selection criteria

**I-Optimality Criterion (Integrated Variance)**. I-optimal designs minimize the average prediction variance:

$$\frac{\int_{\chi} \mathbf{f}'(\mathbf{x}) \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{f}(\mathbf{x}) d\mathbf{x}}{\int_{\chi} d\mathbf{x}}$$

- $\mathbf{x}' = [x_1, x_2, \ldots, x_k]$: a point in the experimental region
- $\mathbf{f}(\mathbf{x})$: model expansion of $\mathbf{x}$
- $\mathbf{X}$: model matrix
- $\mathbf{f}'(\mathbf{x}) \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{f}(\mathbf{x})$: prediction variance at $\mathbf{x}$ relative to $\sigma_{\epsilon}^2$
- $\int_{\chi} d\mathbf{x}$: volume of the experimental region

# Design selection criteria

**I-Optimality Criterion (Integrated Variance).** I-optimal designs minimize the average prediction variance:

$$\frac{\int_{\chi} \mathbf{f}'(\mathbf{x}) \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{f}(\mathbf{x}) d\mathbf{x}}{\int_{\chi} d\mathbf{x}} = \mathrm{tr}\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{M}\right]$$

- $\mathbf{x}' = [x_1, x_2, \ldots, x_k]$: a point in the experimental region
- $\mathbf{f}(\mathbf{x})$: model expansion of $\mathbf{x}$
- $\mathbf{X}$: model matrix
- $\mathbf{f}'(\mathbf{x}) \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{f}(\mathbf{x})$: prediction variance at $\mathbf{x}$ relative to $\sigma_{\epsilon}^2$
- $\int_{\chi} d\mathbf{x}$: volume of the experimental region
- $\mathbf{M}$: moments matrix

## Example moments matrix **M**: 3 quantitative factors

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1/3 | 1/3 | 1/3 |
| 0 | 1/3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 1/3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 1/3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 1/9 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 1/9 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 1/9 | 0 | 0 | 0 |
| 1/3 | 0 | 0 | 0 | 0 | 0 | 0 | 1/5 | 1/9 | 1/9 |
| 1/3 | 0 | 0 | 0 | 0 | 0 | 0 | 1/9 | 1/5 | 1/9 |
| 1/3 | 0 | 0 | 0 | 0 | 0 | 0 | 1/9 | 1/9 | 1/5 |

## $\mathbf{I_D}$-Optimality Criterion (Integrated Variance for Differences)

$$\frac{\int_{\chi}[\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{0})]'\left(\mathbf{X}'\mathbf{X}\right)^{-1}[\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{0})]d\mathbf{x}}{\int_{\chi}d\mathbf{x}}$$

- $\mathbf{0}$: center of the experimental region, typically where the target operating condition is located

**$I_D$-Optimality Criterion** (Integrated Variance for Differences)

$$\frac{\int_{\chi} [\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{0})]' \left(\mathbf{X}'\mathbf{X}\right)^{-1} [\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{0})] d\mathbf{x}}{\int_{\chi} d\mathbf{x}}$$

- $\mathbf{0}$: center of the experimental region, typically where the target operating condition is located

$\Rightarrow$ $I_D$-optimal designs minimize the average prediction variance of differences in responses

**$I_D$-Optimality Criterion** (Integrated Variance for Differences)

$$\frac{\int_{\chi} [\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{0})]' (\mathbf{X}'\mathbf{X})^{-1} [\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{0})] d\mathbf{x}}{\int_{\chi} d\mathbf{x}}$$

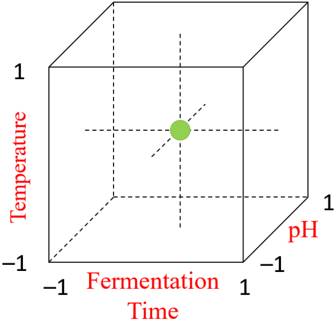
- **0**: center of the experimental region, typically where the target operating condition is located

⇒ $I_D$-optimal designs minimize the average prediction variance of differences in responses

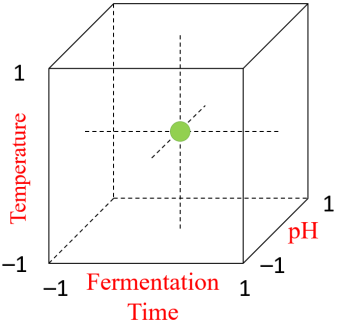
Trinca, L., & Gilmour, S. (2015). Improved Split-Plot and Multi-Stratum Designs. Technometrics, 57, 145–154. https://doi.org/10.1080/00401706.2014.915235



KU LEUVEN


1
Temperature
–1
–1
Fermentation Time
1
–1
pH
1

What if the target setting is *not* at the center?



What if the target setting is *not* at the center?

- Fermentation time

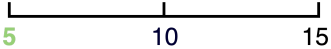



KU LEUVEN

What if the target setting is *not* at the center?

- Fermentation time


- Temperature of a material



**GI$_{\mathrm{D}}$-Optimality Criterion** (Generalized Integrated Variance for Differences)

$$\frac{\int_{\chi}[\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{t})]'\left(\mathbf{X}'\mathbf{X}\right)^{-1}[\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{t})]d\mathbf{x}}{\int_{\chi} d\mathbf{x}}$$

- **t**: target point

**$\mathbf{GI}_{\mathrm{D}}$-Optimality Criterion** (Generalized Integrated Variance for Differences)

$$\frac{\int_{\chi}[\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{t})]'\left(\mathbf{X}'\mathbf{X}\right)^{-1}[\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{t})]d\mathbf{x}}{\int_{\chi}d\mathbf{x}} =$$

$$\mathrm{tr}\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{M}\right]+\mathbf{f}'(\mathbf{t})\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{f}(\mathbf{t})-2\mathbf{m}'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{f}(\mathbf{t})$$

- **t**: target point
- **m**: first column of **M**

**$\text{GI}_\text{D}$-Optimality Criterion** (Generalized Integrated Variance for Differences)

$$\frac{\int_\chi [\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{t})]' \left(\mathbf{X}'\mathbf{X}\right)^{-1} [\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{t})] d\mathbf{x}}{\int_\chi d\mathbf{x}} =$$

$$\text{tr}\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{M}\right] + \mathbf{f}'(\mathbf{t})\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{f}(\mathbf{t}) - 2\mathbf{m}'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{f}(\mathbf{t})$$

Three components:

- the average prediction variance
- the prediction variance at the target point
- average covariance between the predictions at each point in the experimental region and the prediction at the target point

**$\text{GI}_\text{D}$-Optimality Criterion** (Generalized Integrated Variance for Differences)

$$\frac{\int_\chi [\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{t})]' \left(\mathbf{X}'\mathbf{X}\right)^{-1} [\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{t})] d\mathbf{x}}{\int_\chi d\mathbf{x}} =$$

$$\text{tr}\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{M}\right] + \mathbf{f}'(\mathbf{t}) \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{f}(\mathbf{t}) - 2\mathbf{m}' \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{f}(\mathbf{t})$$

Three components:

- the average prediction variance
- the prediction variance at the target point
- average covariance between the predictions at each point in the experimental region and the prediction at the target point

**$\text{GI}_\text{D}$-Optimality Criterion** (Generalized Integrated Variance for Differences)

$$\frac{\int_\chi [\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{t})]' \left(\mathbf{X}'\mathbf{X}\right)^{-1} [\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{t})] d\mathbf{x}}{\int_\chi d\mathbf{x}} =$$

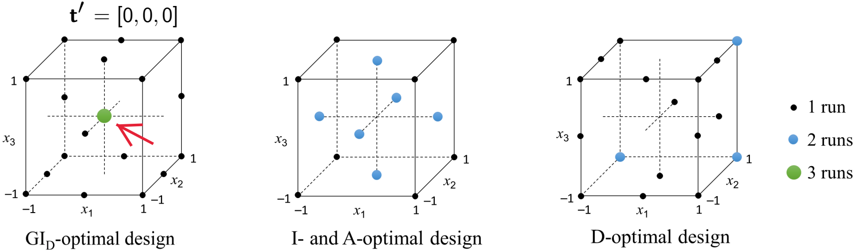
$$\text{tr}\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{M}\right] + \mathbf{f}'(\mathbf{t})\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{f}(\mathbf{t}) - 2\mathbf{m}'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{f}(\mathbf{t})$$

Three components:

- the average prediction variance
- the prediction variance at the target point
- average covariance between the predictions at each point in the experimental region and the prediction at the target point

# Design examples

## 20-Run Designs for a Response Surface Model




- D-optimal designs maximize the determinant of the information matrix $\mathbf{X}'\mathbf{X}$
- A-optimal designs minimize the sum, or the average, of the diagonal elements of $(\mathbf{X}'\mathbf{X})^{-1}$


KU LEUVEN

# Design Evaluation

1. **Relative $\mathbf{GI}_{\mathrm{D}}$-Efficiencies**

$$\frac{\text{GI}_{\mathrm{D}}\text{-optimality criterion value of design 1}}{\text{GI}_{\mathrm{D}}\text{-optimality criterion value of design 2}}$$

# Design Evaluation

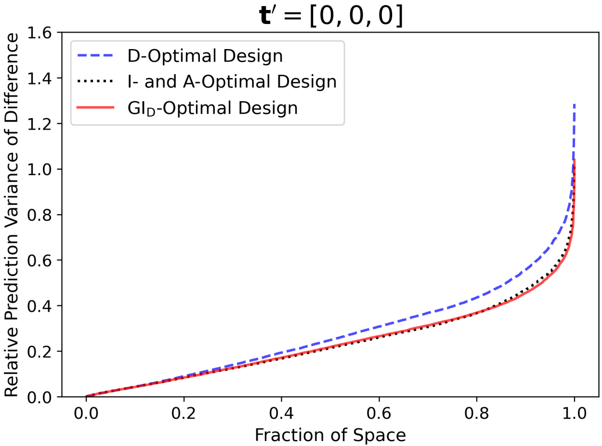
**1. Relative $\mathbf{GI_D}$-Efficiencies**

$$\frac{\text{GI}_\text{D}\text{-optimality criterion value of design 1}}{\text{GI}_\text{D}\text{-optimality criterion value of design 2}}$$

Relative to the $\text{GI}_\text{D}$-optimal design:

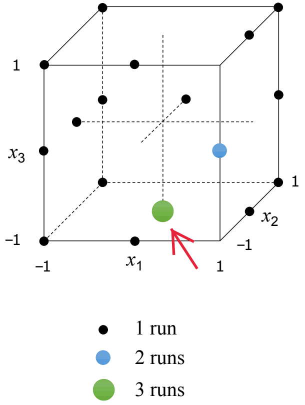
- I- and A-Optimal Design: 99.64%
- D-optimal Design: 84.98%

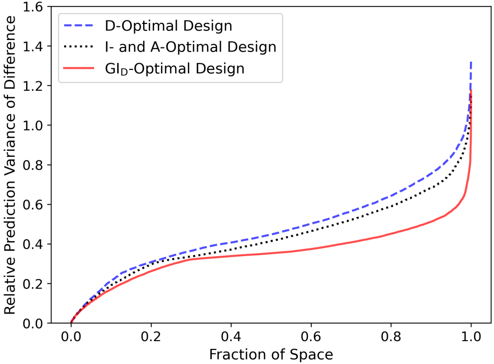
## 2. Difference Fraction of Design Space Plot




$[\mathbf{f}(\mathbf{x}_i) - \mathbf{f}(\mathbf{t})]' (\mathbf{X}'\mathbf{X})^{-1} [\mathbf{f}(\mathbf{x}_i) - \mathbf{f}(\mathbf{t})]$. (de Oliveira et al., 2022)

$$\mathbf{t}' = [0, 0, -1]$$

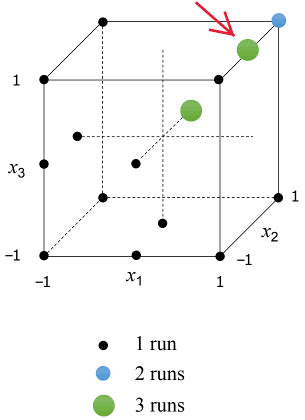


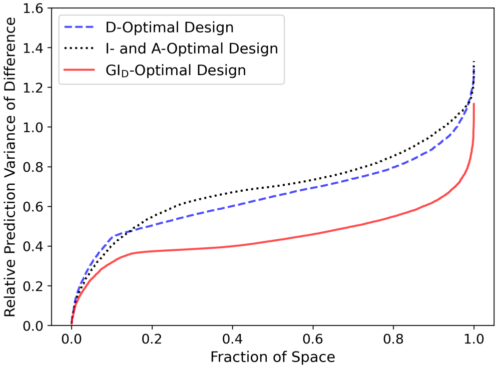





KU LEUVEN

$$\mathbf{t}' = [1, 0, 1]$$

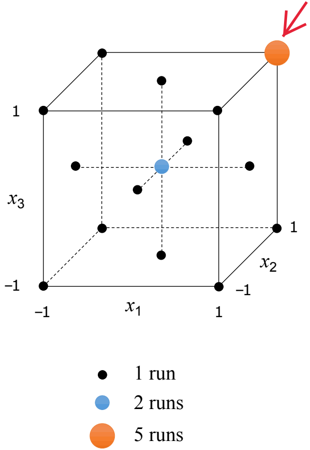


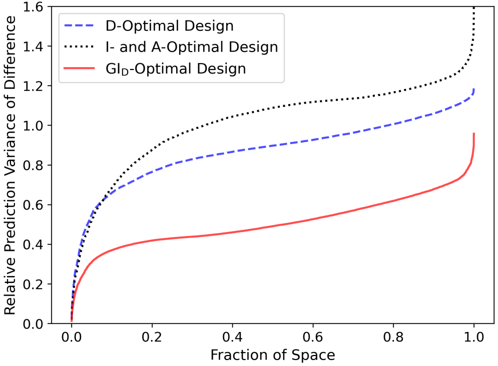





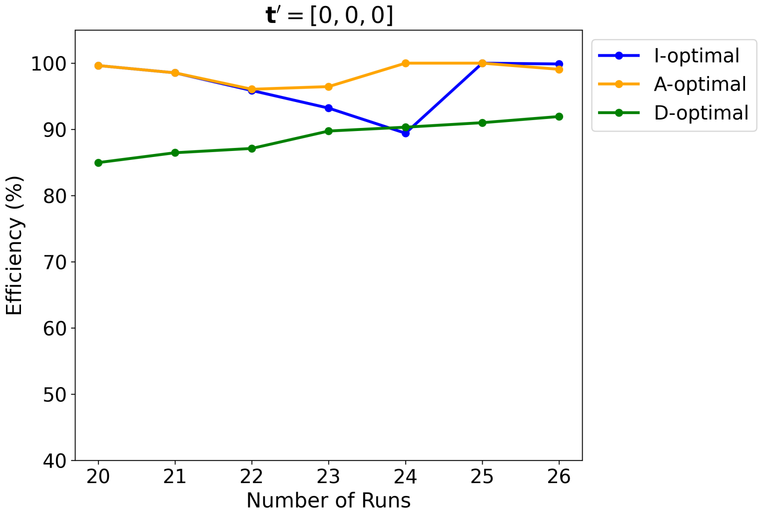
KU LEUVEN

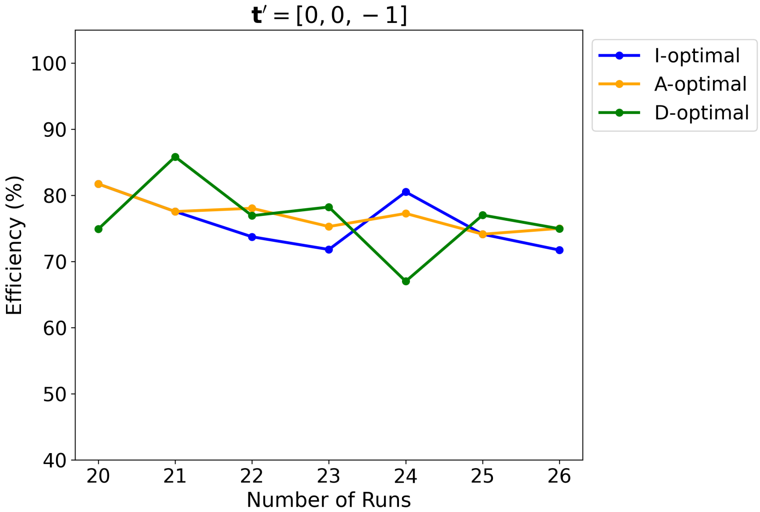
$$\mathbf{t}' = [1, 1, 1]$$


1
$x_3$
−1
−1
$x_1$
1
1
$x_2$
−1
1 run
2 runs
5 runs

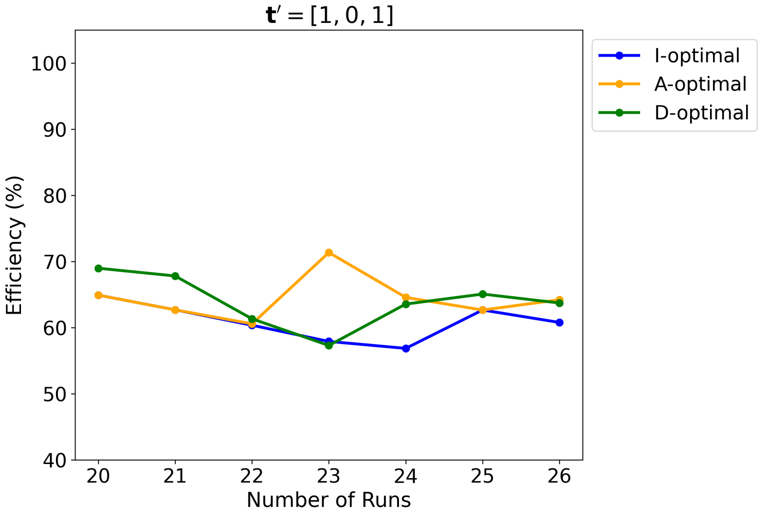

D-Optimal Design
I- and A-Optimal Design
$GI_D$-Optimal Design
Relative Prediction Variance of Difference
Fraction of Space
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

## $GI_D$-Efficiencies of I-, A-, and D-optimal designs relative to $GI_D$-optimal designs

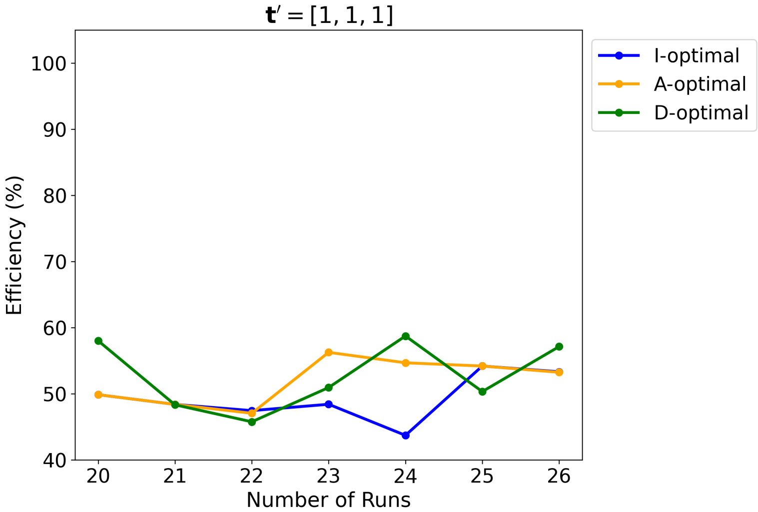




$\mathbf{t}' = [0, 0, -1]$
100
90
80
70
60
50
40
Efficiency (%)
20
21
22
23
24
25
26
Number of Runs
I-optimal
A-optimal
D-optimal



$\mathbf{t}' = [1, 0, 1]$
100
90
80
70
60
50
40
Efficiency (%)
20
21
22
23
24
25
26
Number of Runs
I-optimal
A-optimal
D-optimal



$\mathbf{t}' = [1, 1, 1]$
100
90
80
70
60
50
40
Efficiency (%)
20
21
22
23
24
25
26
Number of Runs
I-optimal
A-optimal
D-optimal

# Summary

- $\mathrm{GI_D}$-optimal designs minimize the average prediction variance of differences in responses, where the target point can be any point in the experimental region



KU LEUVEN

# Summary

- $GI_D$-optimal designs minimize the average prediction variance of differences in responses, where the target point can be any point in the experimental region
- $GI_D$-optimal designs tend to allocate more runs to the target point


KU LEUVEN

# Summary

- $GI_D$-optimal designs minimize the average prediction variance of differences in responses, where the target point can be any point in the experimental region
- $GI_D$-optimal designs tend to allocate more runs to the target point
- $GI_D$-optimal designs outperform other designs for process robustness studies, especially when the target point is not at the center

KU LEUVEN

# Optimal Experimental Designs for Process Robustness Studies


**Ying Chen**

(Research funded by a grant from GlaxoSmithKline Biologicals SA, Belgium)

**Supervisor: Prof. dr. Peter Goos**

(KU Leuven, University of Antwerp)

**Co-supervisor: Dr. Bernard G. Francq**

(An employee of the GSK group of companies)

Relative I-, A- and D-efficiencies of the $GI_D$-optimal designs ($\mathbf{t}' = [0, 0, 0]$) for estimating a response surface model in three quantitative factors

| Run Size | Design Evaluation Criterion | | |
|---|---|---|---|
| | I | A | D |
| 20 | 92.00% | 92.32% | 94.47% |
| 21 | 94.10% | 95.26% | 93.21% |
| 22 | 90.36% | 95.68% | 97.64% |
| 23 | 93.17% | 99.12% | 98.46% |
| 24 | 98.69% | 100% | 96.39% |
| 25 | 100% | 100% | 93.94% |
| 26 | 97.22% | 98.08% | 94.01% |
| Average | 95.08% | 97.21% | 95.45% |