Photo by Stefan Els

Stellenbosch
UNIVERSITY
IYUNIVESITHI
UNIVERSITEIT

Recommended for you:
What chemical engineering can learn from Netflix

Dr Jamie Cripwell

2024 Stellenbosch Engineering Industry Showcase

The Netflix Prize

The Netflix Prize

Stellenbosch
UNIVERSITY
IYUNIVESITHI
UNIVERSITEIT

The Netflix Prize

The Netflix Prize

$\equiv$

$$\boldsymbol{R} \in (m \times n)$$

$\times$

The Netflix Prize

n movies

m users

4	2	1	5		
1		5		3	
	5		4		2
		4			

$\equiv$

$$\boldsymbol{R} \in (m \times n)$$

$\times$

$r_{ij} \ni$ rating of user i for movie j

The Netflix Prize

n movies

m users

4	2	1	5		
1		5		3	
	5		4		2
		4			

$\equiv$

$\boldsymbol{U} \in (m \times k)$

$\boldsymbol{V}^{\boldsymbol{T}} \in (k \times n)$

$$\boldsymbol{R} = \boldsymbol{U}\boldsymbol{V}^{\boldsymbol{T}}$$

The Netflix Prize

Stellenbosch
UNIVERSITY
IYUNIVESITHI
UNIVERSITEIT

n movies

m users

4	2	1	5		
1		5		3	
	5		4		2
		4			

$\equiv$

m users

$U \in (m \times k)$

n movies

$V^T \in (k \times n)$

$$R = UV^T$$

$k \equiv latent\ features$

The Netflix Prize

Stellenbosch
UNIVERSITY
IYUNIVESITHI
UNIVERSITEIT

“Matrix Completion”

n movies

m users

4	2	1	5	2	3
1	3	5	4	3	5
3	5	2	4	3	2
1	4	4	1	3	5

$\equiv$

n movies

m users

$\boldsymbol{U} \in (m \times k)$

$\boldsymbol{V}^{\boldsymbol{T}} \in (k \times n)$

$$\boldsymbol{R} = \boldsymbol{U}\boldsymbol{V}^{\boldsymbol{T}}$$

$k \equiv latent\ features$

The Netflix Prize

"Matrix Completion"

n movies

m users

4	2	1	5	2	3
1	3	5	4	3	5
3	5	2	4	3	2
1	4	4	1	3	5

$\equiv$

m users

$\boldsymbol{U} \in (m \times k)$

n movies

$\boldsymbol{V}^{\boldsymbol{T}} \in (k \times n)$

$$\boldsymbol{R} = \boldsymbol{U}\boldsymbol{V}^{\boldsymbol{T}}$$

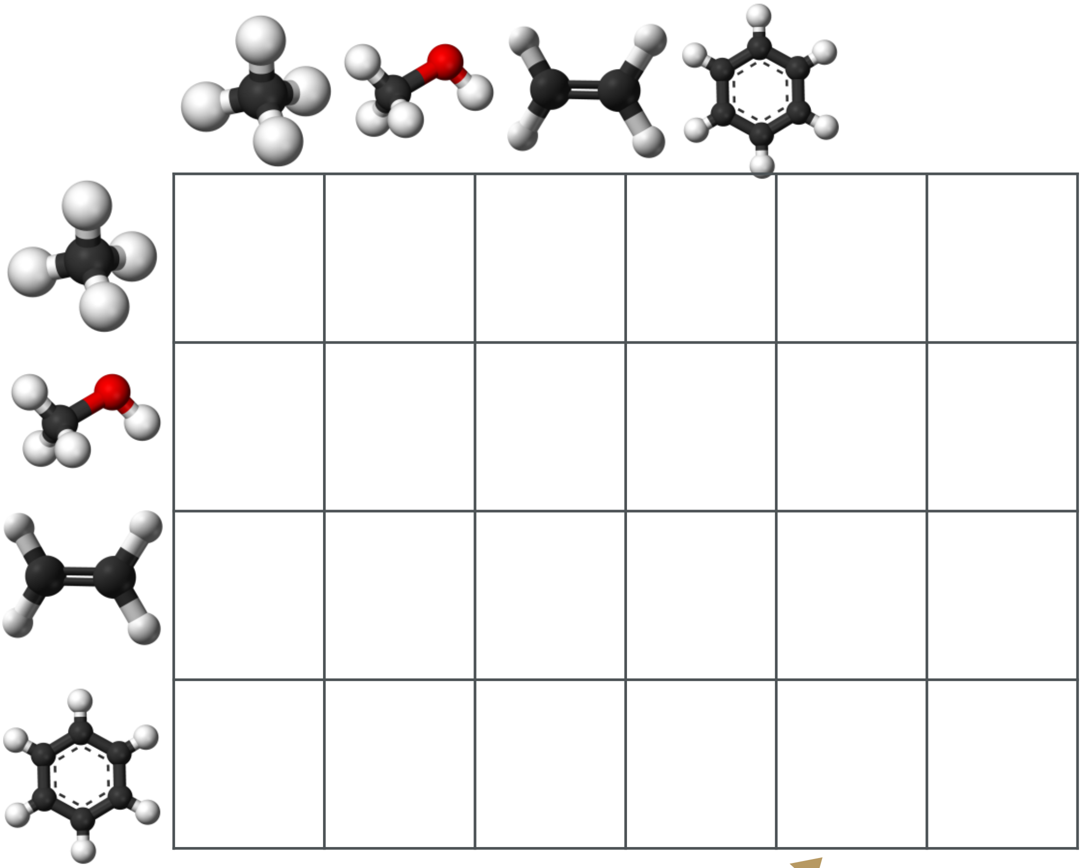
$$\underbrace{k \equiv latent\ features}$$

"Collaborative Filtering"

Relevance to Chemical Engineering?

Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT

"Matrix Completion"



n species

$\equiv$ m species

$\boldsymbol{U} \in (m \times k)$

$\boldsymbol{V}^{\boldsymbol{T}} \in (k \times n)$

$\boldsymbol{R} = \boldsymbol{U}\boldsymbol{V}^{\boldsymbol{T}}$

$\underbrace{k \equiv latent\ features}$

"Collaborative Filtering"

Relevance to Chemical Engineering?

$$\begin{array}{cc} & n\ \text{species} \\ m\ \text{species} & \begin{bmatrix} d_{11} & \cdots & \cdots & d_{14} \\ \vdots & \ddots & \cdots & \vdots \\ \vdots & \cdots & \ddots & \vdots \\ d_{41} & \cdots & \cdots & d_{44} \end{bmatrix} \end{array}$$

$\boldsymbol{D} \ni$ mixture property data

Relevance to Chemical Engineering?

Stellenbosch
UNIVERSITY
IYUNIVESITHI
UNIVERSITEIT

$$\begin{array}{cc} & n \text{ species} \\ m \text{ species} & \begin{bmatrix} d_{11} & \cdots & \cdots & d_{14} \\ \vdots & \ddots & \cdots & \vdots \\ \vdots & \cdots & \ddots & \vdots \\ d_{41} & \cdots & \cdots & d_{44} \end{bmatrix} \end{array}$$

$\boldsymbol{D} \ni$ mixture property data

$\boldsymbol{d}_{ii} \ni$ pure component

$\boldsymbol{d}_{ij} \ni$ binary mixture

Success stories…

- Compared to UNIFAC Dortmund: "…error is below 0.1 for 37.4% of the predictions with UNIFAC, whereas the proposed MCM achieves the same accuracy for 50.0% of the predictions."

n solutes

m solvents

$$\begin{bmatrix} d_{11} & \cdots & \cdots & d_{14} \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ d_{41} & \cdots & \cdots & d_{44} \end{bmatrix}$$

γ_{ij}^{∞}

240 solutes
250 solvents

$\equiv$

m solvents

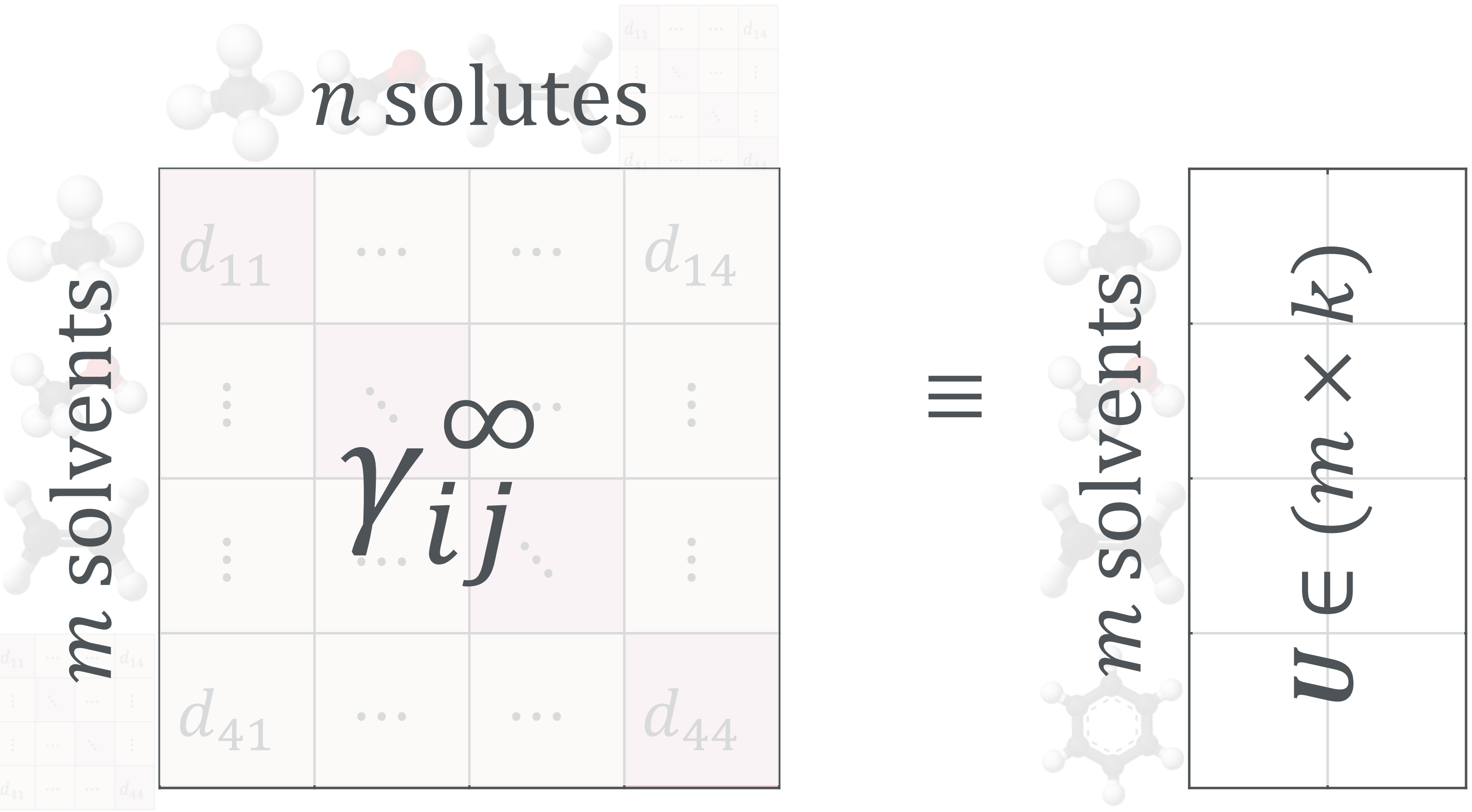
$U \in (250 \times k)$

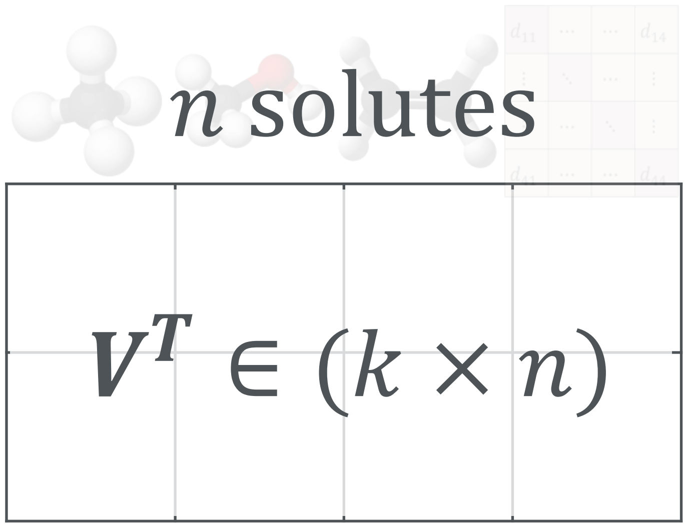
n solutes

$V^T \in (k \times 240)$

Success stories…

- Extended to other scalar valued mixture properties with similar success



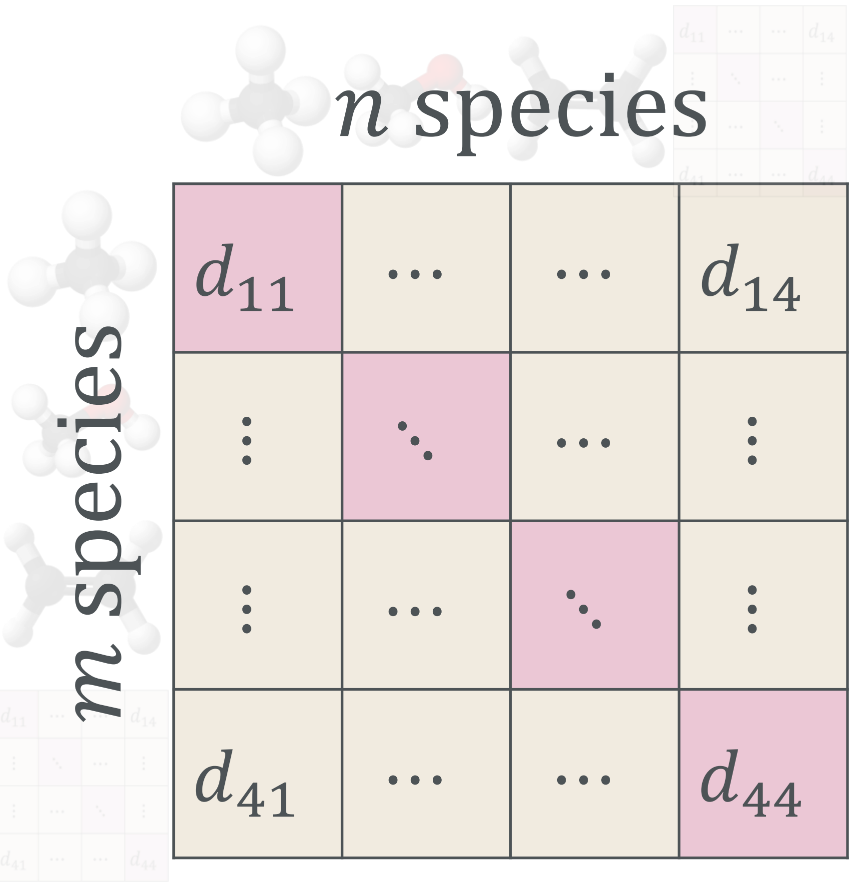


- Henry constants: H_{ij}
- Diffusion coefficient: D_{ij}
- UNIQUAC pair interactions: U_{ij}

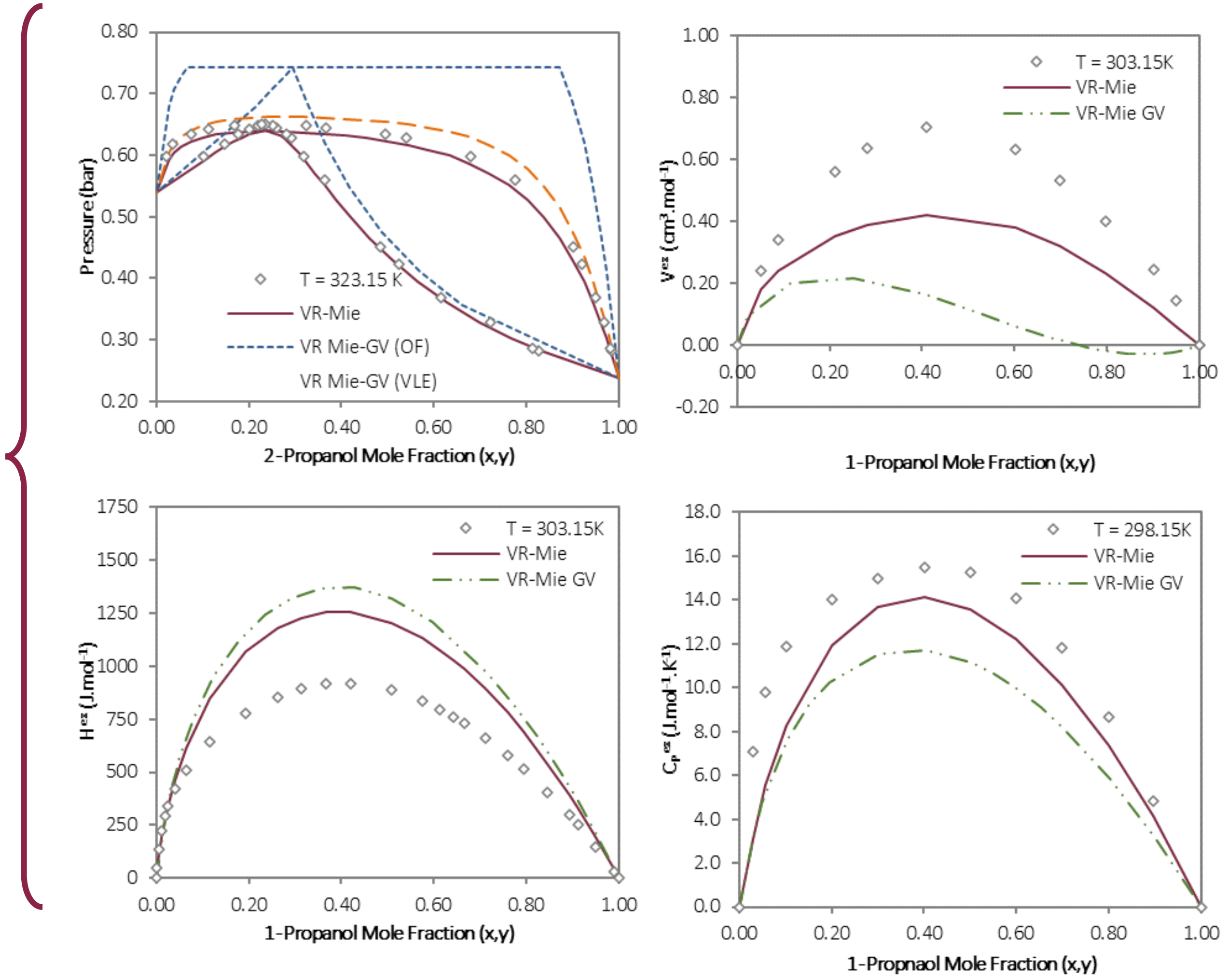
Hayer, N., et al. (2022). AIChE Journal 68(9)
Jirasek, F., et al. (2022). Chemical Science 13(17): 4854-4862
Großmann, O., et al. (2022). Digital Discovery 1(6): 886-897

Our interest…

- Higher dimensional data: composition (x_{ij}), temperature (T), pressure (P),…

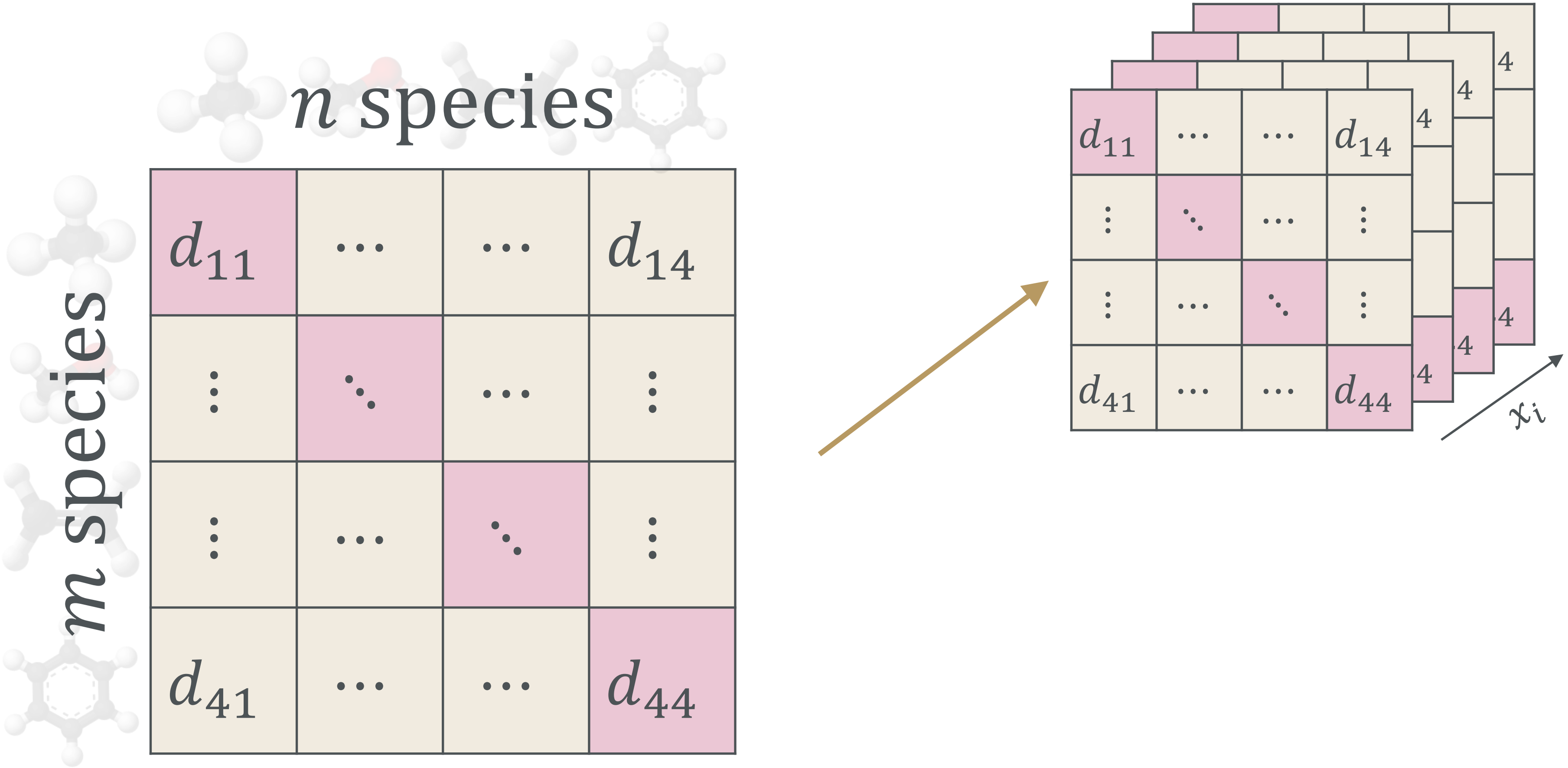


$\overset{?}{\longleftarrow}$



Our interest…

- **Higher dimensional** data: composition (x_{ij}), temperature (T), pressure (P),…

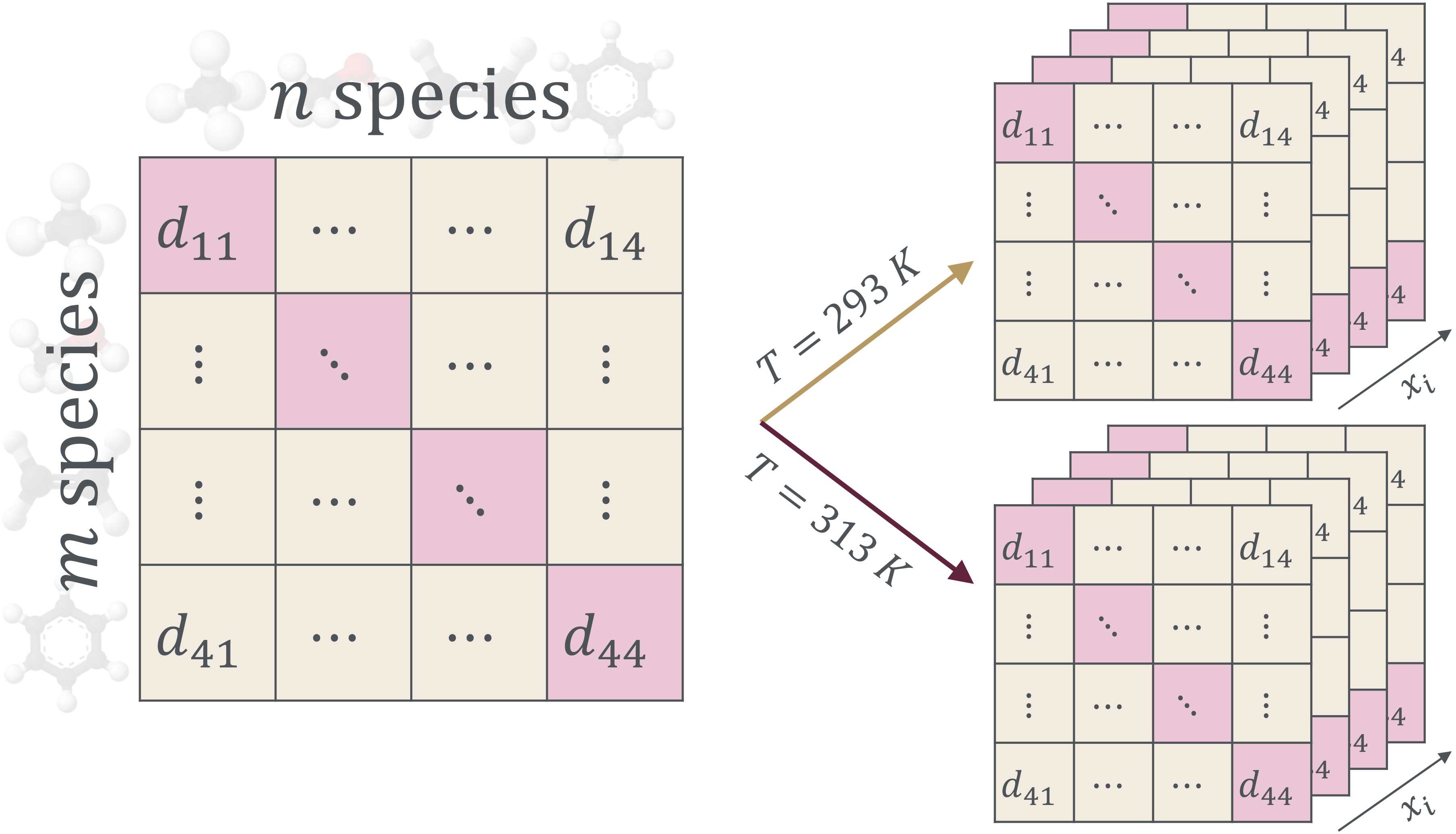


Parallelised Matrix Completion

- Independent "slices" of continuous, higher dimensions
- Same algorithm as MCM

Our interest…

- Higher dimensional data: composition (x_{ij}), temperature (T), pressure (P),…



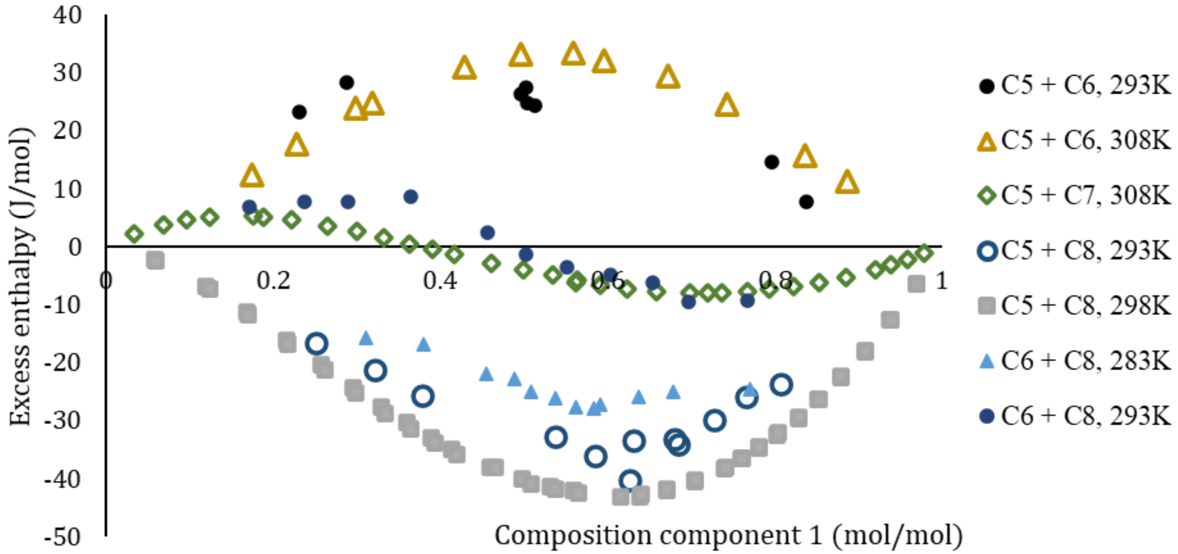
Parallelised Matrix Completion

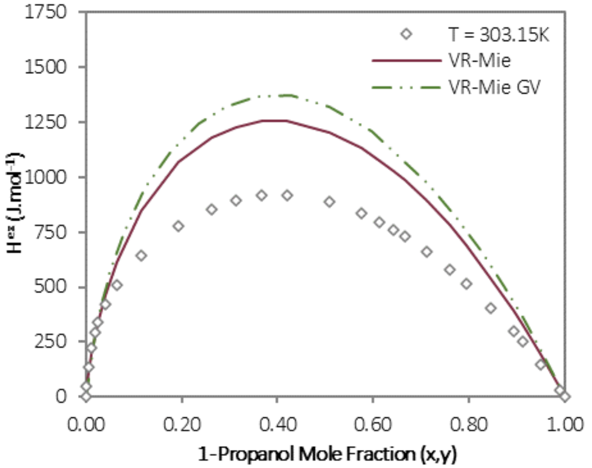
- Independent "slices" of continuous, higher dimensions
- Same algorithm as MCM

- Extra dimensions ≡ more "cubes"

The case for mixtures

- Liquid phase excess enthalpy (H^E):
  - monotonic, scalar valued, qualitatively varied
  - thermodynamic consistency testing, poorly predicted

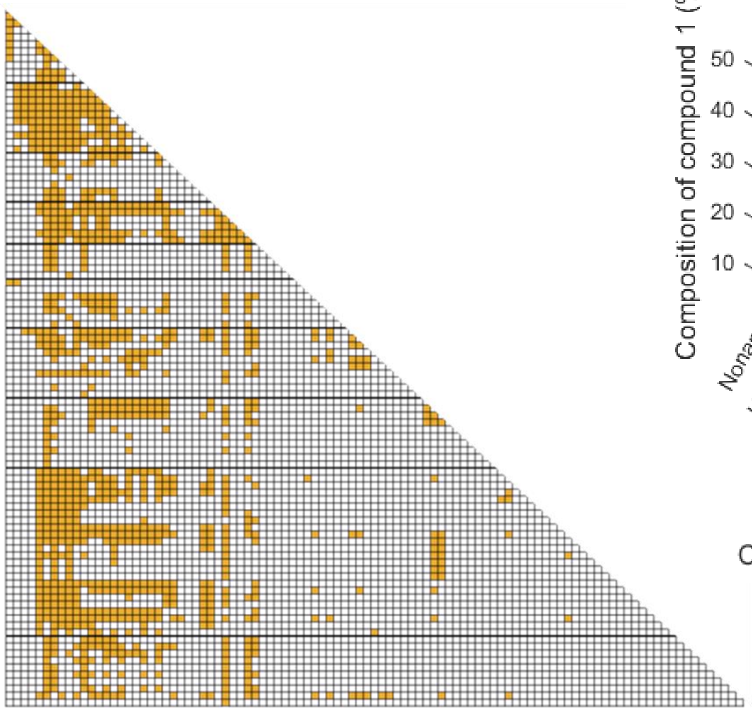


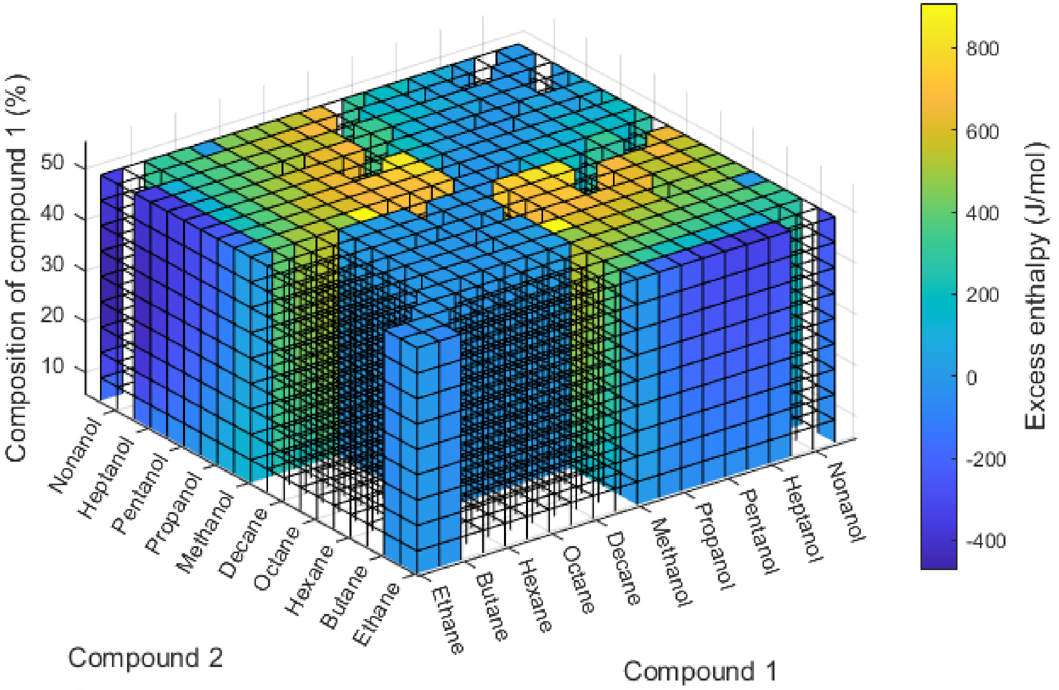


Francesca Middleton

The case for mixtures

- Liquid phase excess enthalpy (H^E):
  - 97 components across 10 chemical families at 7 temperatures
  - 1012 mixtures from 4950 possible combinations → 11% - 20% observed

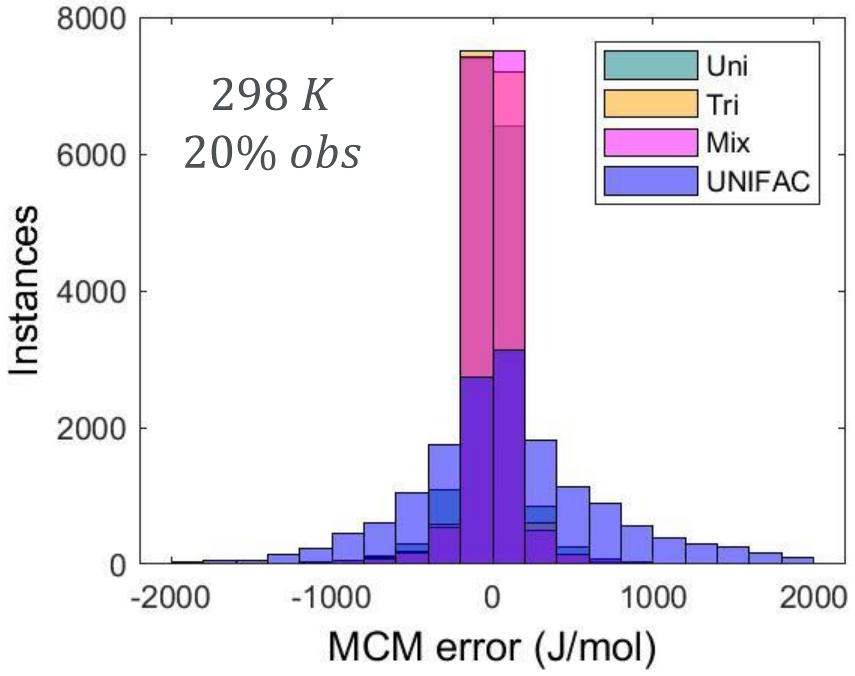


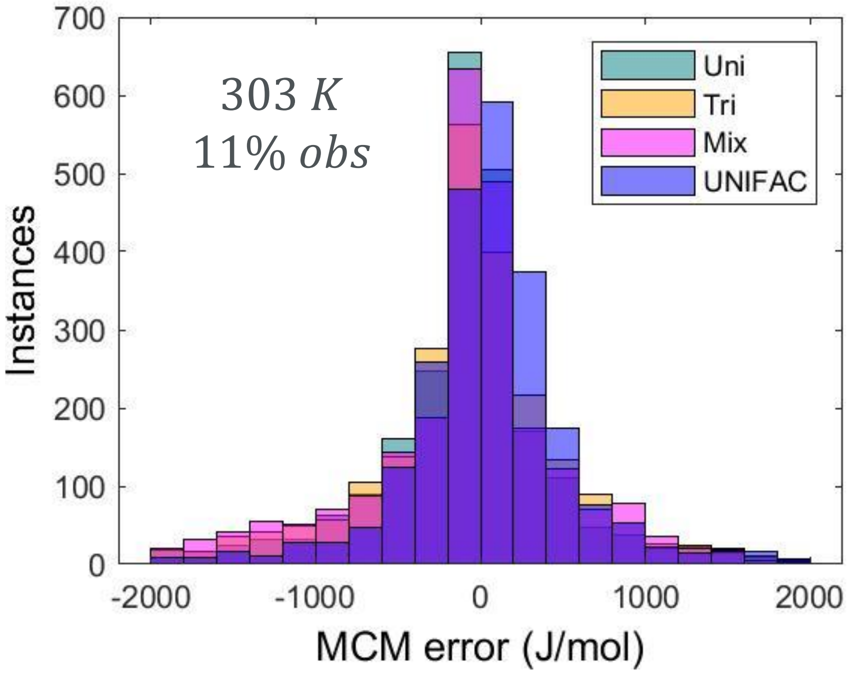


Francesca Middleton

The case for mixtures

- More data ≡ better results
- More generalisable than UNIFAC Dortmund

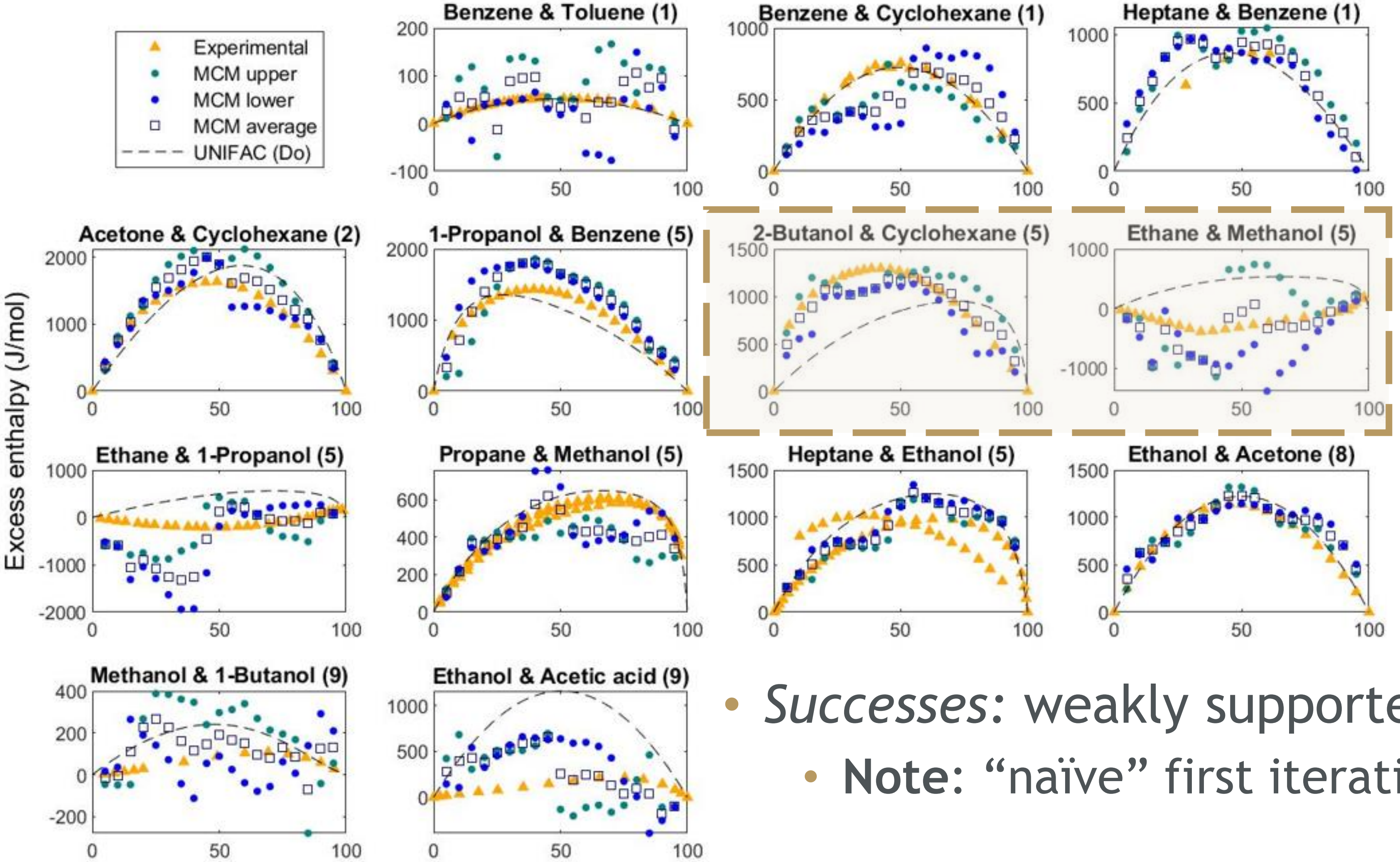




Francesca Middleton

The case for mixtures

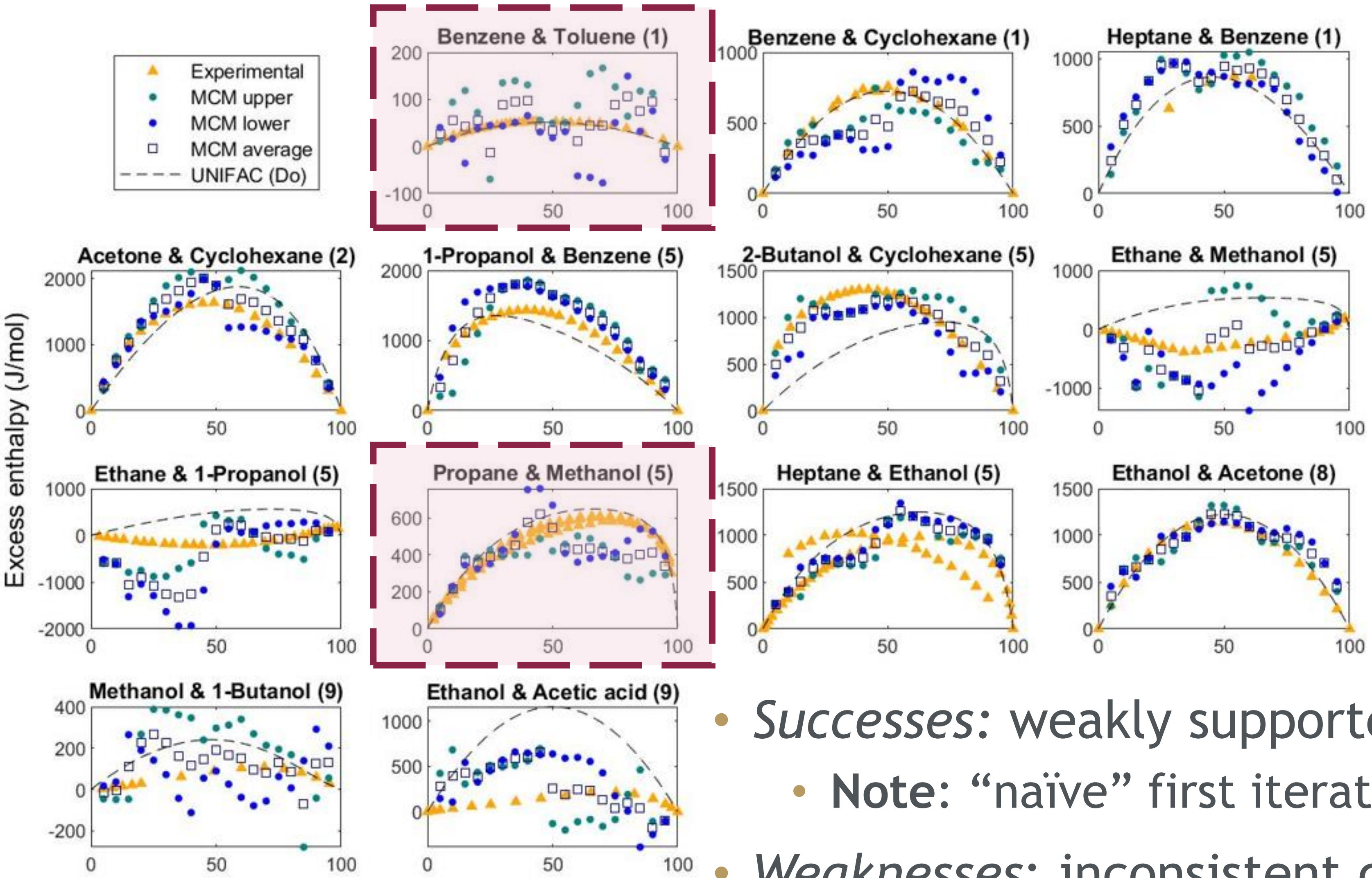
Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT



Francesca Middleton

- *Successes*: weakly supported components; niche cases
  - **Note**: “naïve” first iteration

The case for mixtures



Francesca Middleton

- *Successes*: weakly supported components; niche cases
  - **Note**: “naïve” first iteration
- *Weaknesses*: inconsistent quality for “standard” systems; explicit trends in quality

Recommended for you…

Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT

- Predictions not consistently quantitative
  - Won’t replace fundamental models

- Room for improvement
  - Hybrid models: incorporating predictive fundamental models
  - Alternatives: feature information, clustering
  - Quantifying uncertainty: probabilistic approach*

*Garren Hermanus

- Fundamentals first
  - Many hammers in machine learning toolbox, not all problems are nails…
  - Identify and understand problem → explore solutions
    - Chemical engineering, *then* machine learning…

Structured M.Eng (Chemical) with focus area: Data Analytics

Stellenbosch
UNIVERSITY
IYUNIVESITHI
UNIVERSITEIT

- Equip engineers with fundamentals of data science enabling application in the context of integrated industrial processes
- Collaboratively developed with industry
- Aimed at working engineers studying part-time
- Fully hybrid presentation
- 2024 intake ~ 20 students

Part-time students	Term 1	Term 2	Term 3	Term 4	*Full-time students*
Year 1	Data Science	Applied Machine Learning	Plantwide Dynamics andControl		Year 1
Year 2	Numerical Methods	Optimisation		Data Analysis for Dynamic Processes	
Year 3	Advanced Topics in Eng. Management Integrated Process Data Analysis	Research project			Year 2

Stellenbosch
UNIVERSITY
IYUNIVESITHI
UNIVERSITEIT

Thank you
Enkosi
Dankie

Photo by Stefan Els