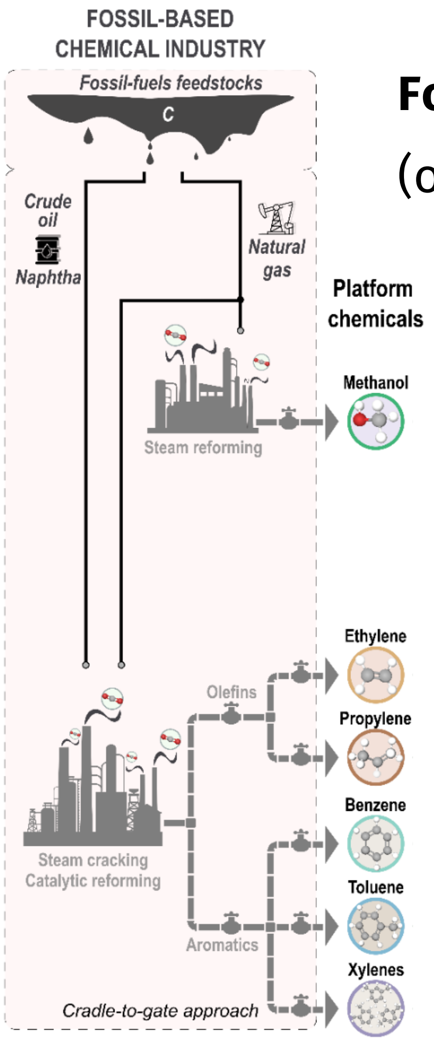
Recent applications of machine learning and life cycle assessment to carbon capture

Gonzalo Guillén-Gosálbez, ETH Zurich

24th March 2025, Edinburgh

Current fossil-based chemical industry is unsustainable

ETH zürich



Fossil carbon

(oil and natural gas)

C

CO_2

CO_2

3 Gt CO_2eq/yr

Renewable C:

- **CO_2 (CCU)**
- **Biomass**
- **Chemical products waste**

?

To what extent can technologies contribute to sustainable chemicals production?

One Earth **2021**, 4, 565-583
PNAS **2019**, 116, 11187-11194

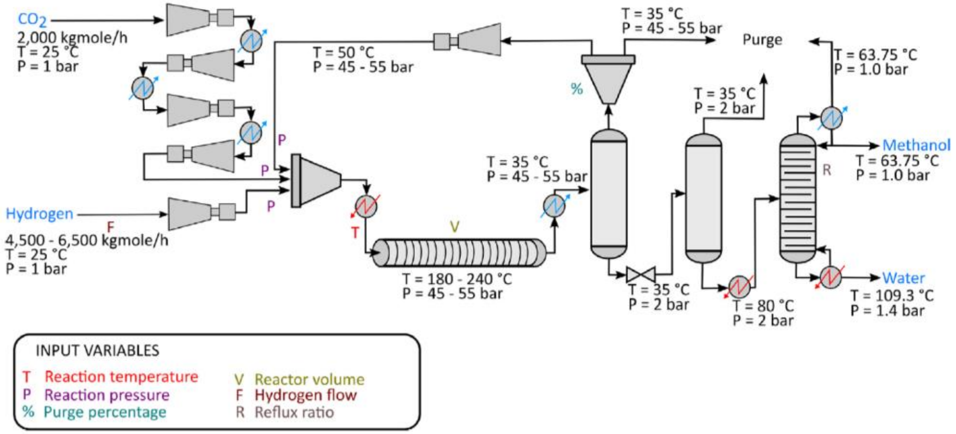
How to quantify sustainability performance?

- **Life cycle assessment of chemical systems**:

Chemical reaction

$CO_2 + 3H_2 \rightarrow CH_3OH + H_2O$

Process simulation



Electricity

Heat

Reactants
X kg CO_2
Y kg H_2

Direct emissions

ecoinvent

Emissions per unit of flow

Fossil MeOH: **0.72 kg CO_2eq/kg**

Green MeOH: **-0.68 kg CO_2eq/kg**

Technology A

Technology B

Technology ...

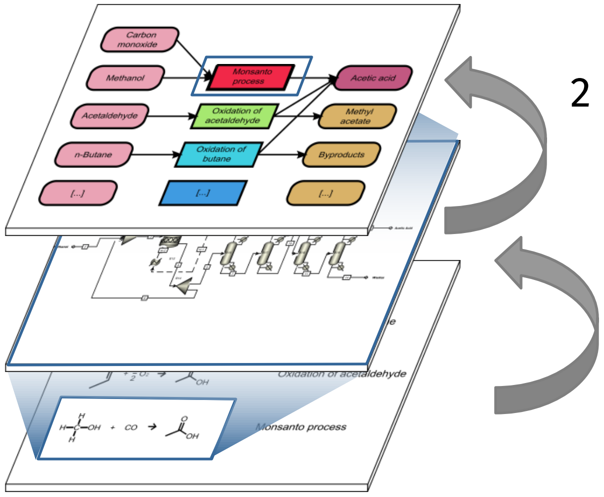
$$
\begin{aligned}
min \quad & f_1(x,y), \dots, f_k(x,y) \\
s.t. \quad & h(x,y) = 0 \\
& g(x,y) \leq 0 \\
& x \in \Re^n, y \in \{0,1\}^m
\end{aligned}
$$

Energy & Environmental Science **2019**, 12, 3425-3436
Chemical Engineering Science **2021**, 246, 116891

Challenges in sustainable engineering

- **How to compare emerging technologies?**

3. Supply chain/life cycle
2. Process flowsheets
1. Experimental work



2 → 3 Life cycle optimization

1 → 2 Generation of process flowsheets

Experimental group

Technology A

Technology B

Technology ...

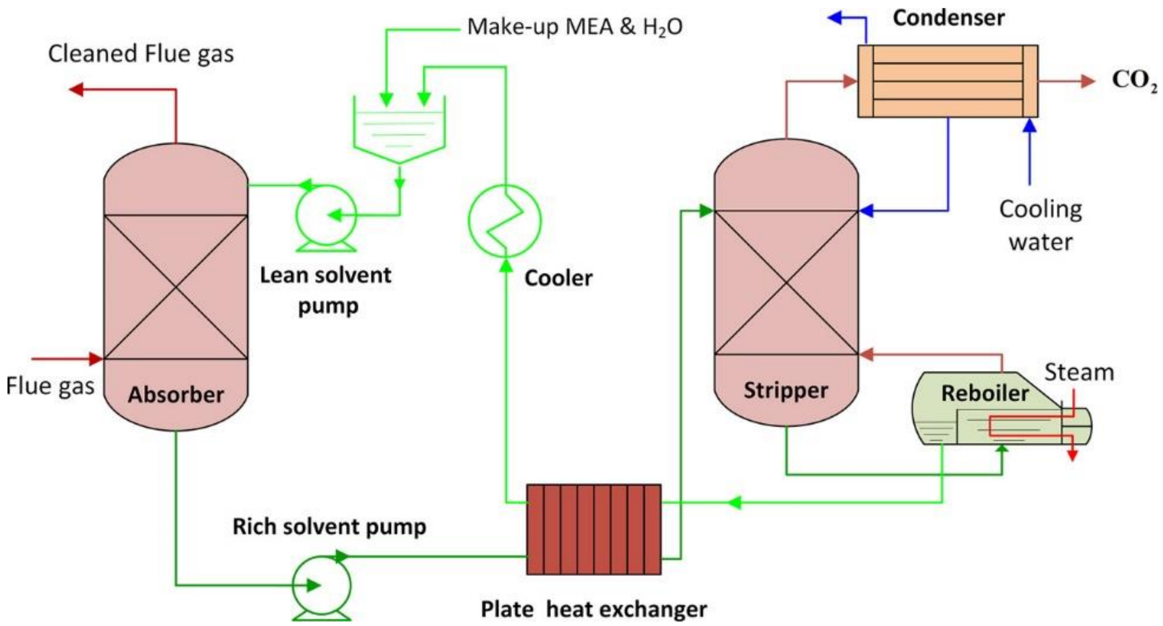
Data

Insights

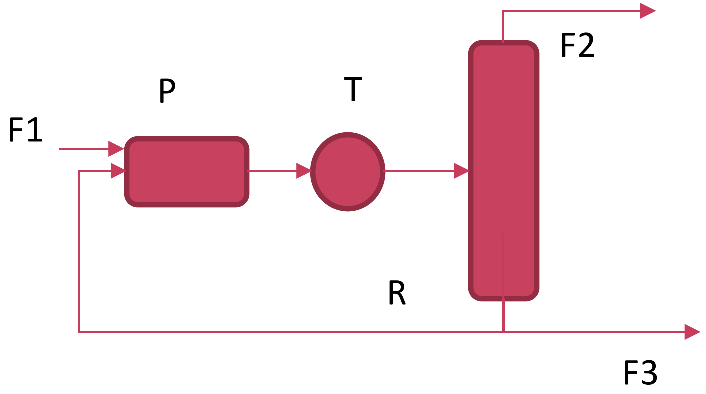
Modeling group
(techno-economic & environmental assessment)

Challenges in sustainable engineering: Process simulation

- **Key idea:** Build analytical correlations from detailed simulations



- Hard to develop
- License/version issues



$$F_2 = f_1(F_1,P,T,R)$$

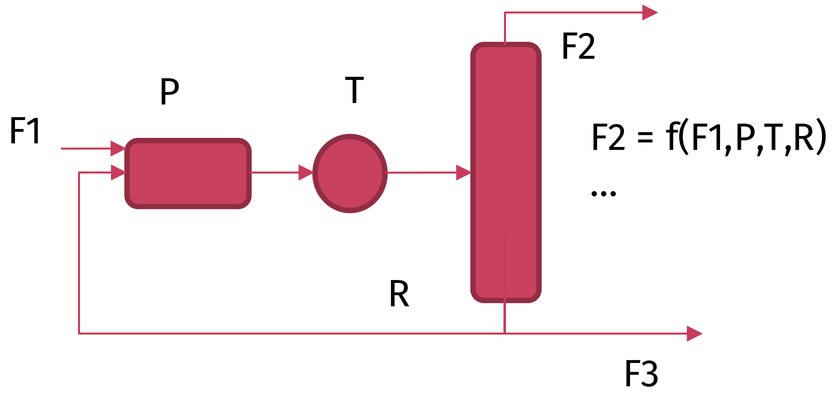
...

- Easy to reproduce
- Can be directly used by experimental groups

Industrial & Engineering Chemistry Research **2021**, 60 (14), 5176-5193

Challenges in sustainable engineering: Process simulation

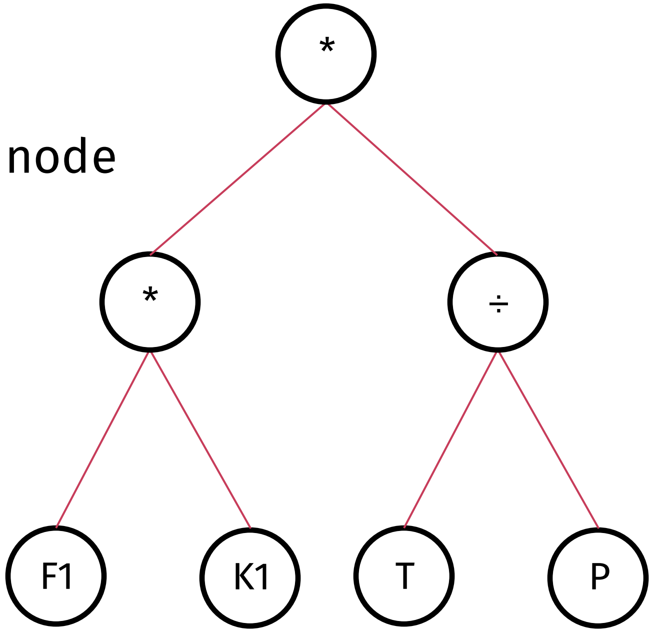
- **Symbolic regression[1,2]:** Mathematical expressions built using expression trees



Disjunction over each intermediate node
{+} V {-} V {*} V {÷} V ...

Disjunction over each leaf node
{F1} V {P} V {T} V {K1} V ...

F2 = F1*K1*(T/P)



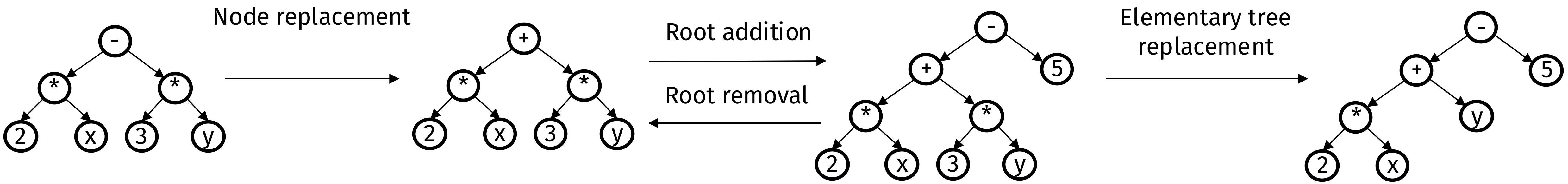
1. *Mathematical Programming* **2018**, 170, 97–119
2. *MIT Press, Cambridge, MA* **1992**

Challenges in sustainable engineering: Process simulation

The Bayesian Machine Scientist[1]

- Explores symbolic trees to provide **closed-form mathematical expressions**

- **Markov chain Monte Carlo (MCMC)** seeks the best expressions

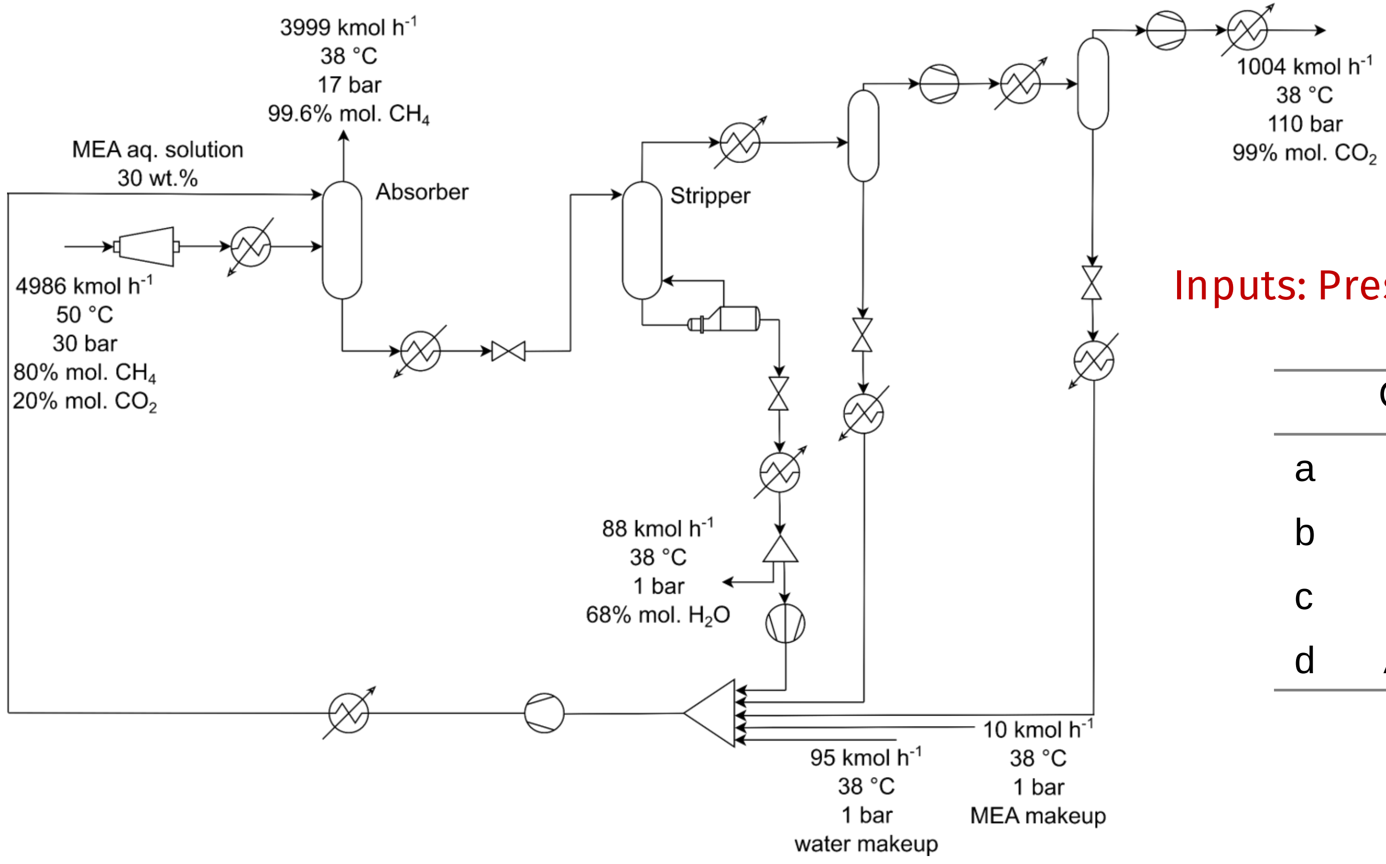


- Mathematical expressions evaluated via the **description length**
  - $L(f_i) \approx \frac{BIC}{2} - \log(p(f_i))$
  - $p(f_i)$: Probability of prior over expressions (corpus of **~5000 equations** from **Wikipedia**)

1. *Science advances* ***2020***, 6(5), eaav6971

Challenges in sustainable engineering: Process simulation

The Bayesian Machine Scientist applied to carbon capture



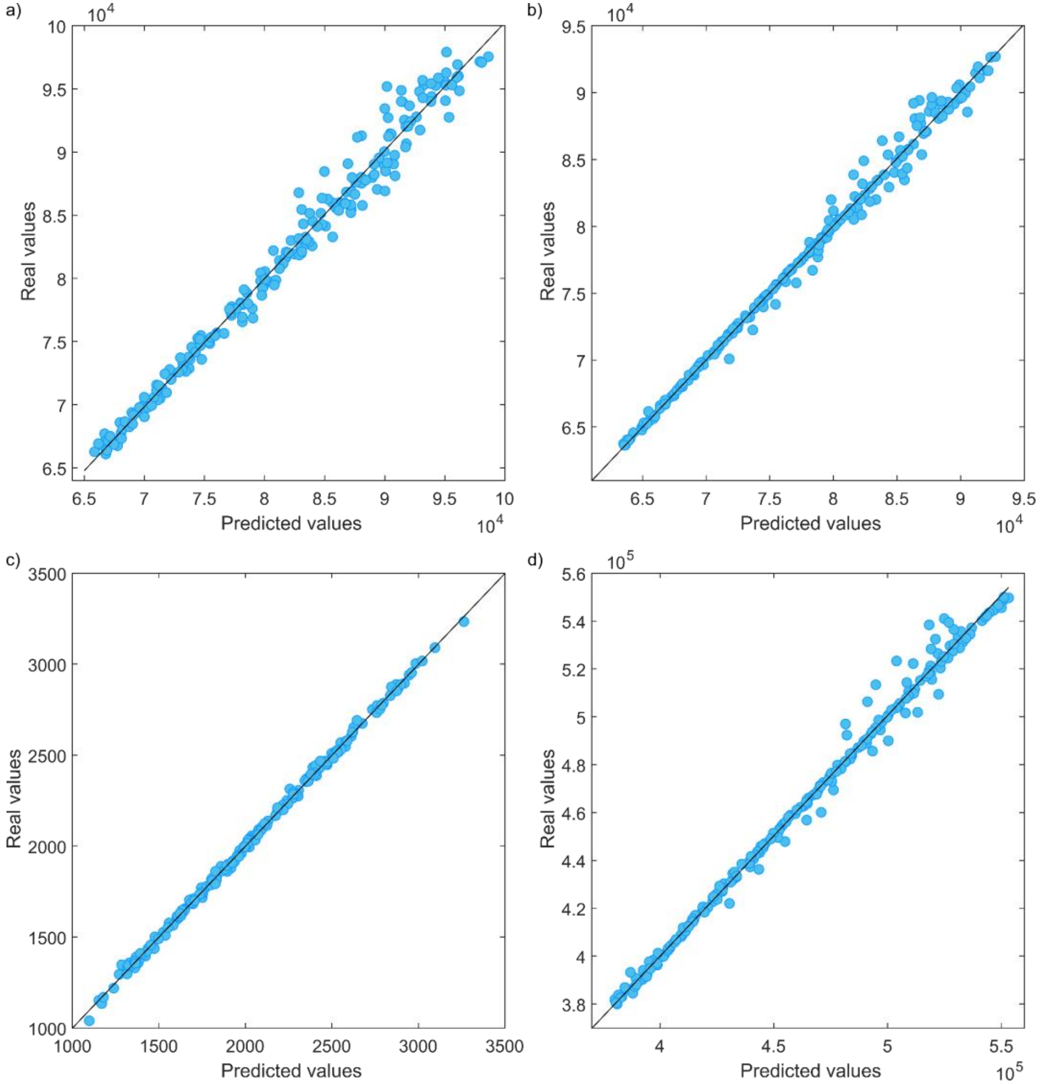
Inputs: Pressure, temperature, input feed and CO_2 concentration

	Output	R^2	MRE
a	Min CU	0.9818	0.0103
b	Min HU	0.9921	0.0051
c	Net power	0.9986	0.0072
d	Amount of MEA	0.9922	0.0050

Appl. Energy **2017**, 204, 353–361
ACS omega **2022**, 7 (45), 41147-41164

Challenges in sustainable engineering: Process simulation

Case study



$$\text{MinCU} = \left(\left(exp\left(a_6^{x_{CH4P}^{a_5}} + x_{CO2F} + \exp\left(\frac{x_{CH4P}}{a_2}\right)\right) * a_3\right) + a_4\right) * a_1\Big)^{a_2}$$

$$\text{MinHU} = \left(\sinh\left(a_4^{x_{CH4P}^{a_3}} * x_{CH4P}^{a_5} * a_2^2\right) + a_0 * \frac{x_{CO2F}}{a_2 * x_{CH4P}}\right) * a_1 + \frac{a_0}{a_6 * x_{CH4P}} + a_2$$

Net power

$$= a_7 * \left(a_1^{a_3 + x_{CO2F}} + a_5 * x_{CH4P} + x_P^{a_4} + \left(\frac{a_7 + a_2}{a_6 + x_T}\right)\right)$$

Amount of MEA

$$= a_0 * a_3^{x_{CH4P}} * a_7 * x_{CO2F} * \left(\left(\frac{x_4}{a_2}\right)^{a_6} + \left(a_4 * x_{CH4P} + \frac{a_5 * tan\left(a_5 + x_{CH4P}^{a_0}\right) * (a_1 + 2 * a_6 + x_{CH4P})}{x_3}\right)^2\right)$$

Challenges in sustainable engineering: Process simulation

Symbolic regression in process optimization

Use COM interface Aspen Plus® Python → Generate surrogate models → Process model in GAMS → Solution evaluated in the rigorous simulation

'Black-box' surrogate model
- BMS
- Kriging

'Hybrid' surrogate model
- BMS
- Kriging

BARON

$$\begin{aligned} \min \quad & f(x,y) \\ \text{s.t.} \quad & h(x,y) = 0 \\ & g(x,y) \leq 0 \\ & x \in \mathbb{R}^n, y \in \{0,1\}^n \end{aligned}$$

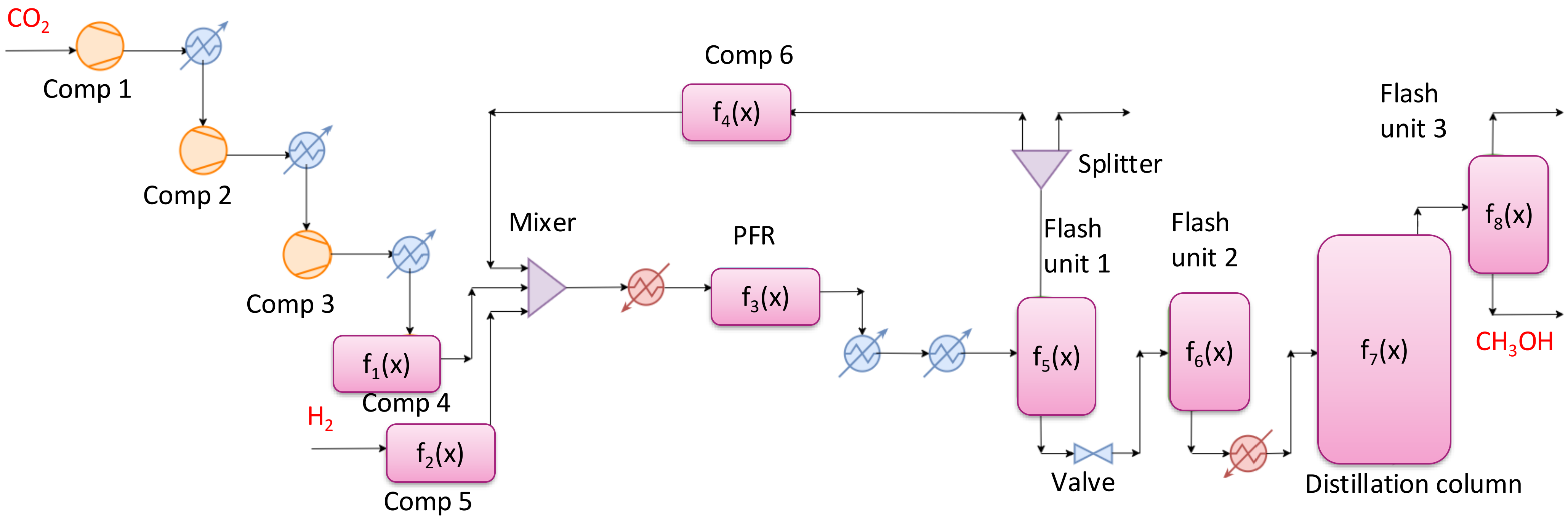
Symbolic regression

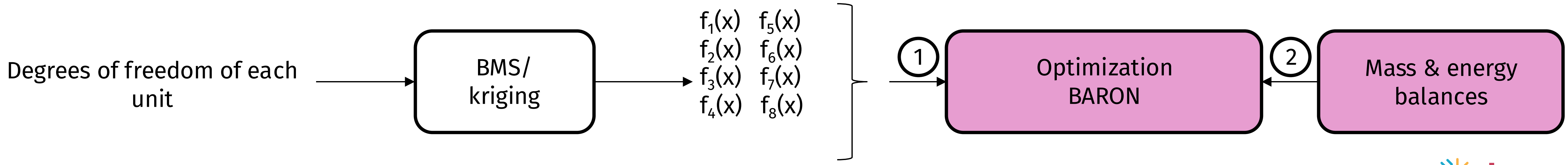
Aspen Plus® Python MATLAB®

Computers & Chemical Engineering ***2024***, 182, 108563

Challenges in sustainable engineering: Process simulation

Symbolic regression in process optimization: Methanol production





Challenges in sustainable engineering: Process simulation

Symbolic regression in process optimization: Methanol production

Objective: Cost minimization

	Model	Training (s)	Optimization time (s)	OF – Surrogate ($/kg)	OF – Aspen Plus® ($/kg)	% Error
1	Aspen Plus®		8	-	1.55	-
2	BMS Hybrid	~14400	0.10	1.42	1.40	0.98%
3	BMS Black-box	~14400	14400*	1.53	1.66	-7.81%
4	Kriging Hybrid	30.2	14400*	No feasible solution found		
5	Kriging Black-box	1.3	14400*	1.34	1.41	-5.62%

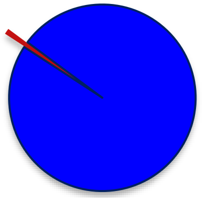
Solved with GAMS 35.2.0 using BARON 21.1.13, sampling time ~2250 s on an Intel® Core i7-10700 CPU @ 2.90 GHz

BMS hybrid model:

- **Best solution**
- **Highest accuracy**: Minimum deviation from the rigorous simulation
- Other approaches unable to close the gap ("*" = 100% gap when maximum time is reached)

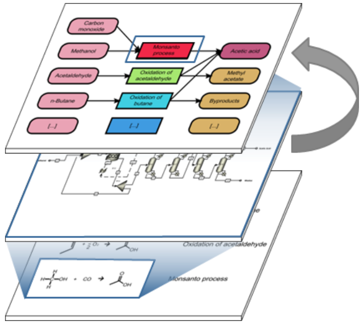
Challenges in sustainable engineering: Lack of LCA data

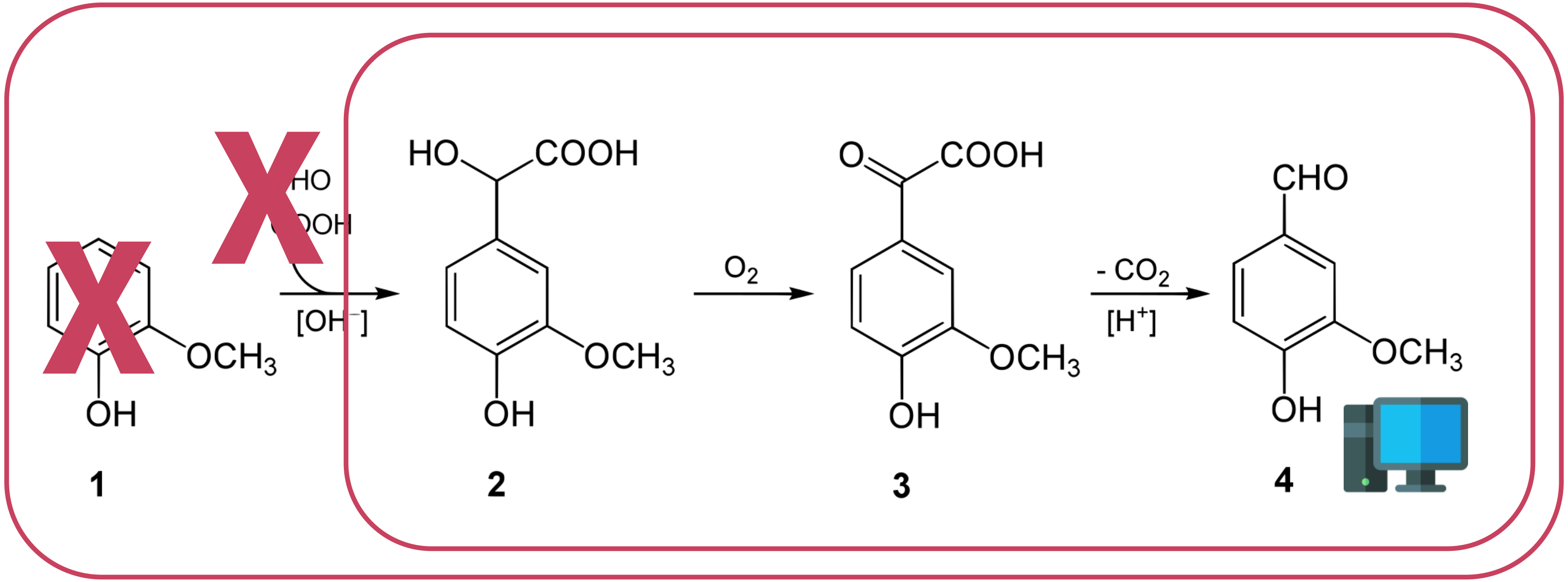
- Life cycle assessment: **LCAs of more complex chemicals may face data gaps**



~1 Million registered chemicals

LCA data of only a few thousand chemicals





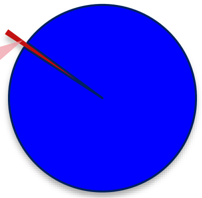
https://en.wikipedia.org/wiki/Vanillin

Challenge: Simple yet accurate methods to assess sustainability performance

Kirk-Othmer Encyclopedia of Chemical Technology **1997**, 24, 812–825.

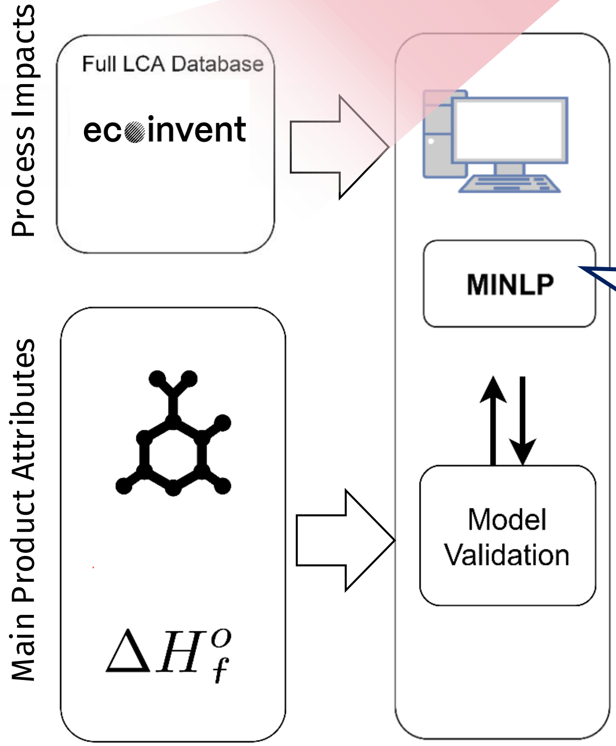
Challenges in sustainable engineering: Lack of LCA data

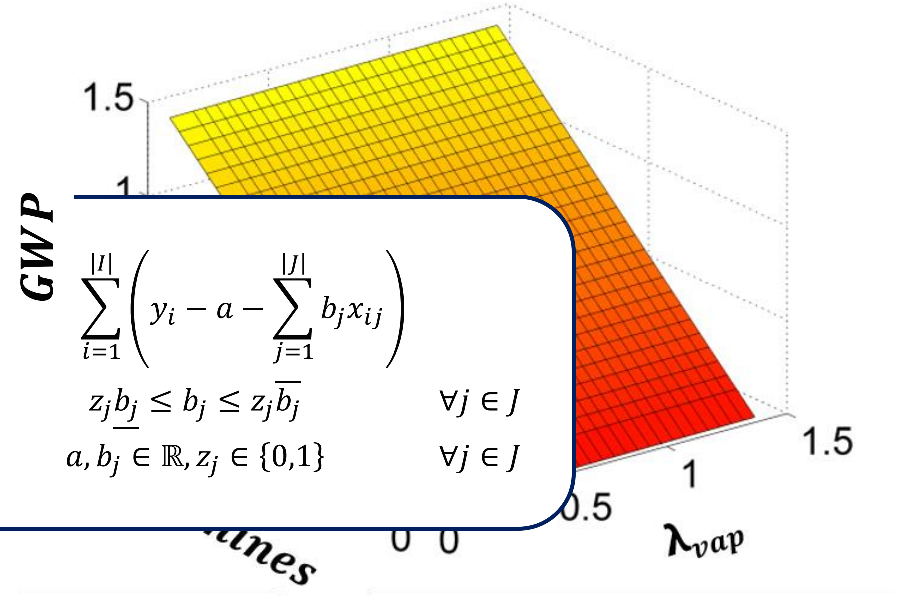
Automatic Data Regression
+ Feature Selection



~1 Million registered chemicals

LCA data of only a few
thousand chemicals





$$
\begin{array}{lll}
Min \; F & \sum_{i=1}^{|I|}\left(y_i - a - \sum_{j=1}^{|J|} b_j x_{ij}\right) & \\
s.t. & z_j \underline{b_j} \leq b_j \leq z_j \overline{b_j} & \forall j \in J \\
 & a, b_j \in \mathbb{R}, z_j \in \{0,1\} & \forall j \in J
\end{array}
$$

MIQCP
200 continuous
32 binary
315 equations
CPLEX

$$\boldsymbol{GWP} = 75.2 + 0.786 \cdot \boldsymbol{amines}$$
$$+3.1 \cdot 10^{-8} \cdot \boldsymbol{\lambda_{vap}}$$

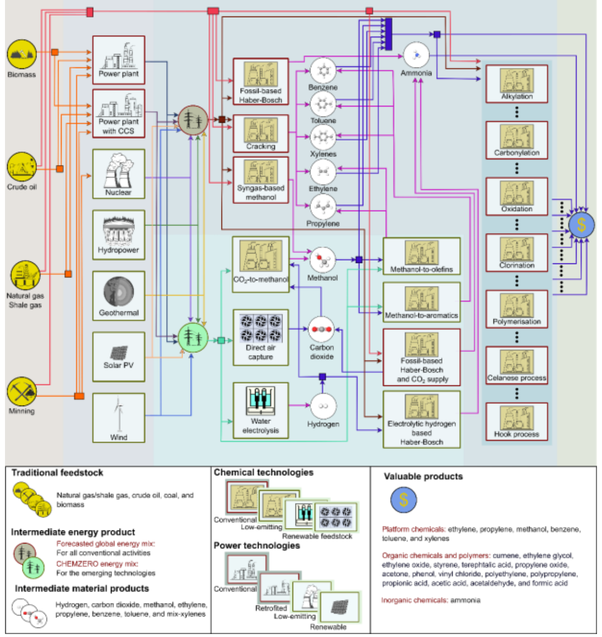
~30% error

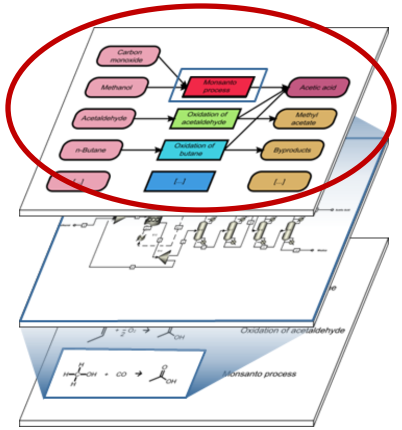
Computers & Chemical Engineering **2018**, 108, 179-193
ACS Sustainable Chemistry and Engineering **2024**, 12, 2700-2708

ETH zürich

Challenges in sustainable engineering: Life cycle optimization

2030 Agenda for Sustainable Development





21 chemicals
Renewable C & power, DAC, CCU and CCS

- **Superstructure** of technologies encompassing thousands of alternatives

$$
\begin{aligned}
\min_x \quad & f(x) && \text{Objective function}\\
\text{s.t.} \quad & h(x) = 0 && \text{Mass \& energy balances}\\
& g(x) \leq 0 && \text{Capacity constraints}
\end{aligned}
$$

Energy & Environmental Science **2022**, 16, 113-124

Challenges in sustainable engineering: Life cycle optimization

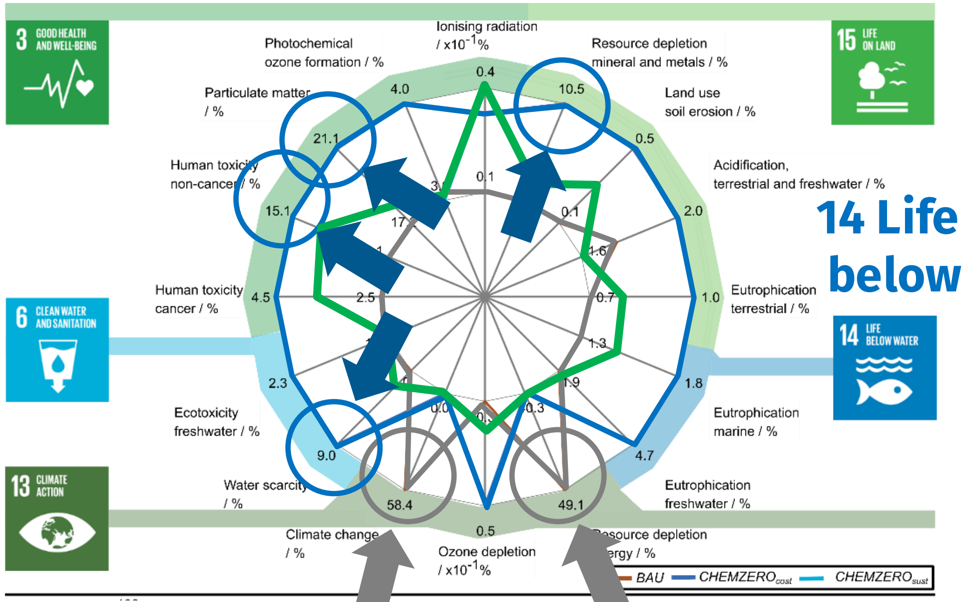
3 Good health and well-being

15 Life on land

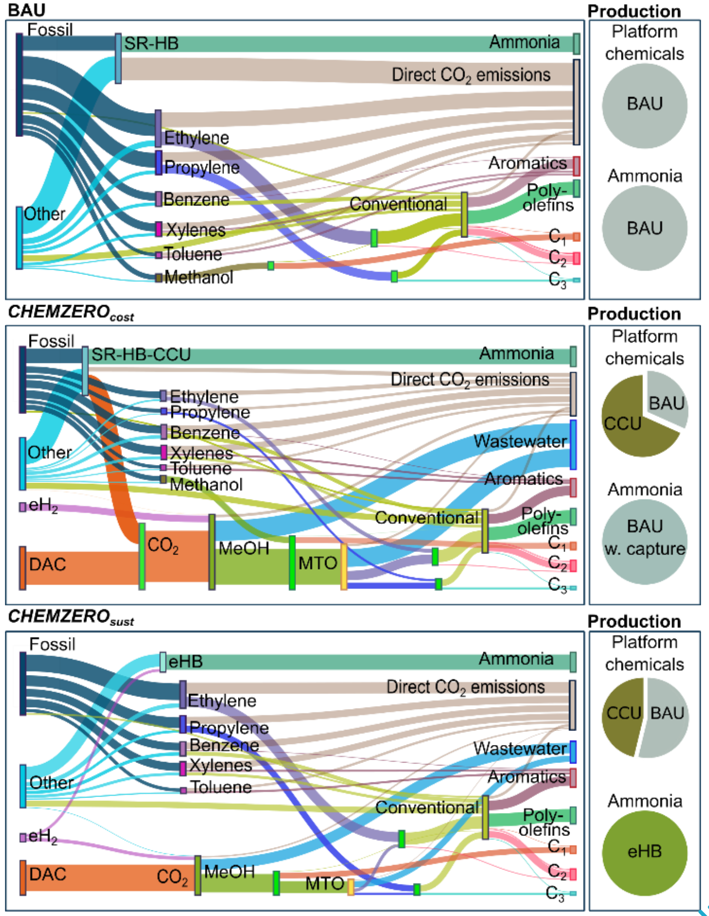
6 Clean water and sanitation

14 Life below water

13 Climate action



- **Hybrid production** patterns are required to produce chemicals sustainably



- Fossil chemicals lead to **large impacts on SDG 13**
- Carbon neutral at minimum cost leads to **burden-shifting**
- Carbon neutral optimizing SDGs performance **reduces the potential collateral damage**

Challenges in sustainable engineering: Life cycle optimization

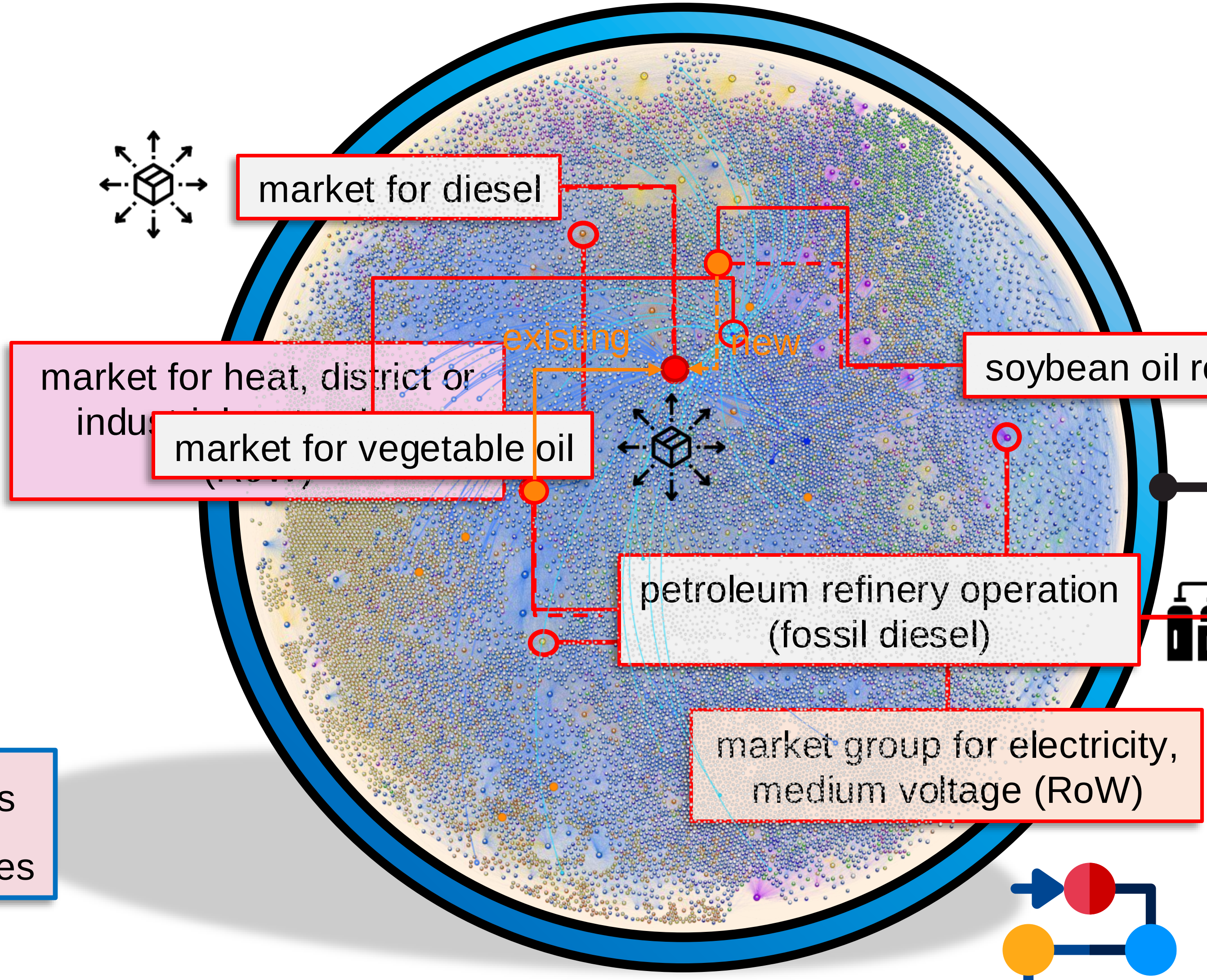
Material

Electricity

Heat

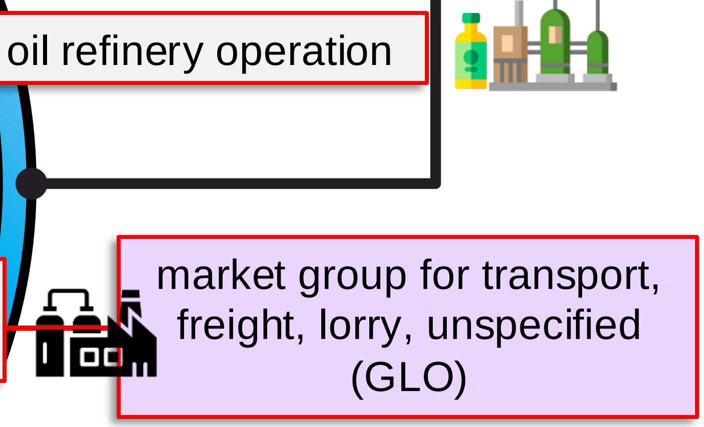
Transport

Land use



ecoinvent 3.8

(cutoff system model)



- 19.000 Nodes
- 200.000 Edges

$$\begin{aligned} min \quad & f_1(x,y), \dots, f_k(x,y) \\ s.t. \quad & h(x,y) = 0 \\ & g(x,y) \leq 0 \\ & x \in \Re^n, y \in \{0,1\}^m \end{aligned}$$

PULPO: An *oracle* to underpin *sustainable* technology development...

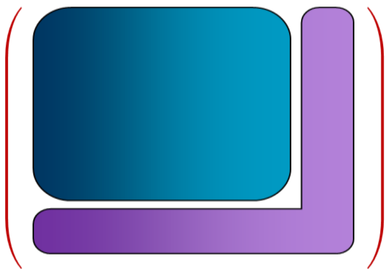
ETH zürich

User Input

(processes)

1. functional unit
2. objective function
3. choices
4. constraints

LCI Database
(+ User Inventories)



Brightway

Python-based User-defined
Lifecycle Product Optimization

PULPO

Optimization Model

PYOMO

Results

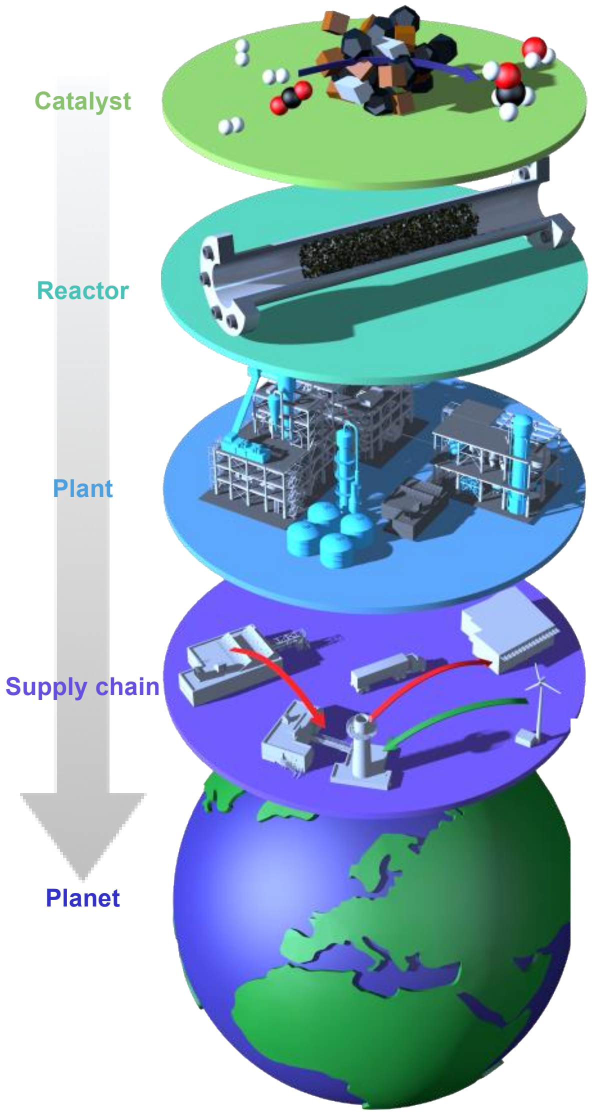
save_results()

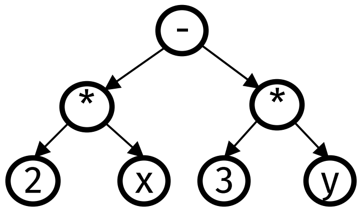
pandas

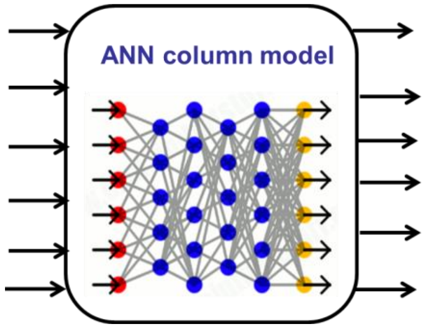
summarize_results()

- Which C feedstocks?
- Which renewable technologies?
- Break-even efficiency and bottlenecks?
- Where, when and how to deploy technologies?
- How to optimally couple the chemical industry with other sectors?

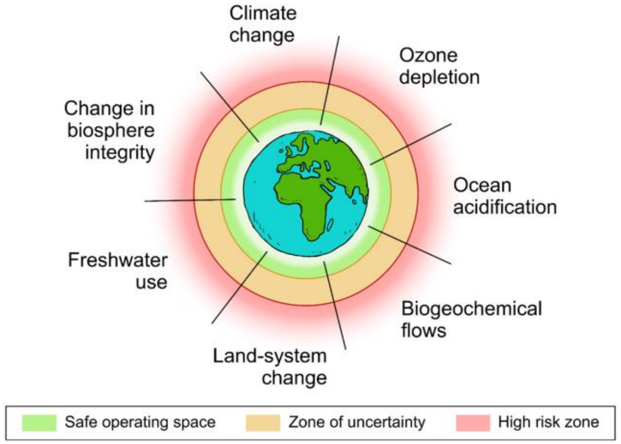
PULPO: An *oracle* to underpin *sustainable* technology development…







PULPO



Generative AI
(for identifying technological and sectoral improvement opportunities)

Conclusions

- Symbolic regression can simplify the process modelling task
- Machine learning algorithms can help cover data gaps in life cycle assessment
- Emerging technologies should be assessed following a multi-scale modelling approach
- Generative AI tools could be used to identify sustainable technological solutions

Take-home message

AI tools could help in guiding research efforts in carbon capture technologies and understanding their future role in sustainable industrial systems