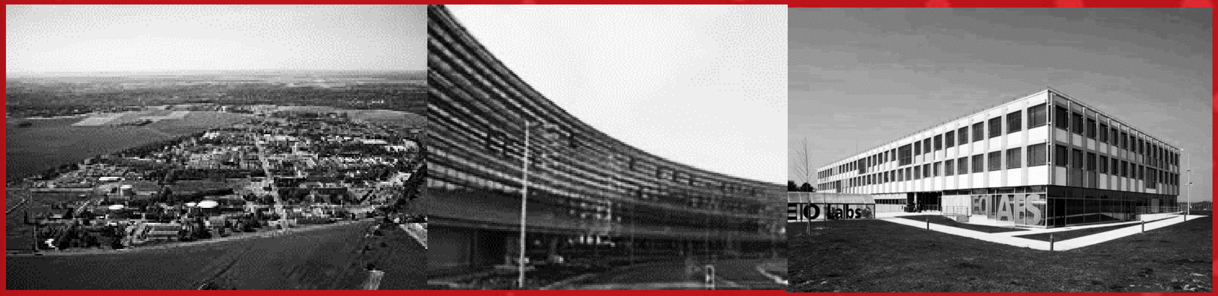


Symbolic artificial intelligence for new material design

From research to industry

31/05/2022

Aurore Lomet, Jean-Philippe Poli

INTRODUCTION: DESIGN OF NEW MATERIALS

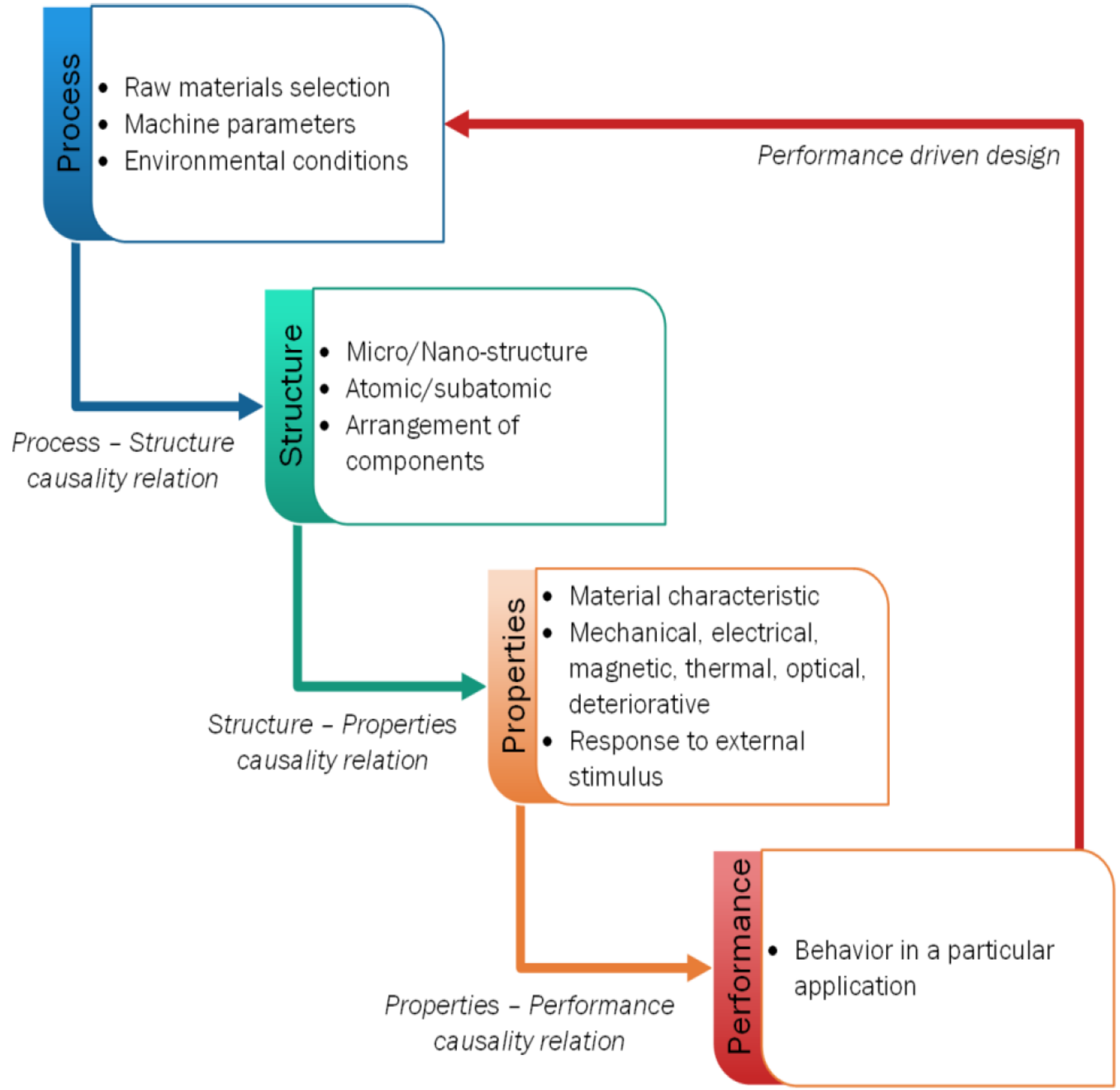
- Design under constraints to obtain some properties:
    - Resistivity, conductivity, toxicity, weight
    - Best performances for a specific use

- Exploratory approach based on the empiric method **‘Trial and Error’** → expensive and low performances due to:
    - the combinatory search space (chemical elements, proportions, process parameters)
    - The irregularity (non linearity, discontinuity, non isolated solutions)

- Alternative approach based on model predictions (physical or numerical) → difficulty to develop due to the complexity of the phenomenon especially the reactions of the mixtures

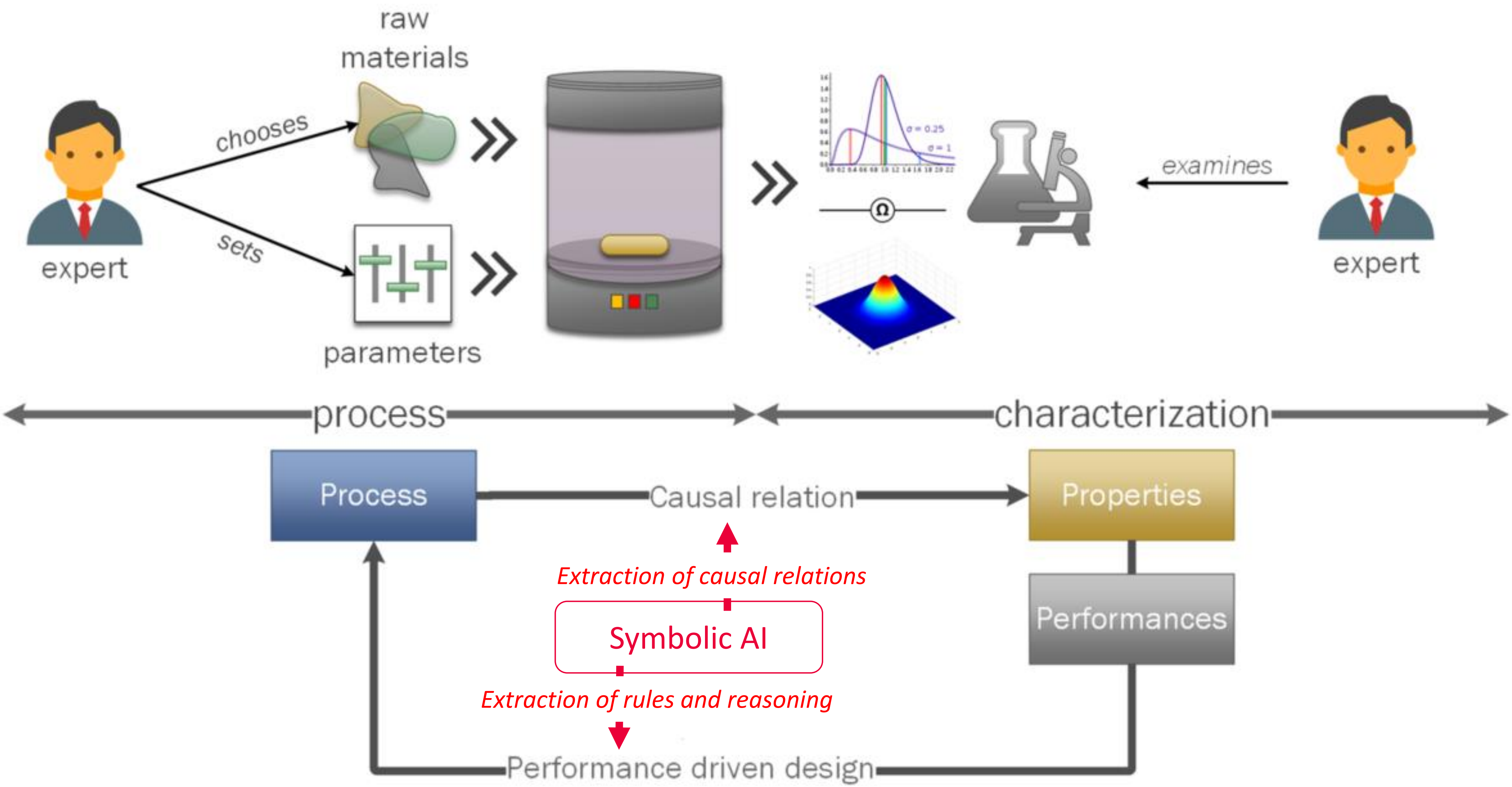
Context

Relations PSPP



Context

Performance-driven design



Objectives

- **Support researchers in a faster discovery of new materials**
- **Understand the phenomena that are modeled during the learning step**
- **Explain the decisions, increase the confidence of the researcher in the AI**

Objectives

- **Support researchers in a faster discovery of new materials**
- **Understand the phenomena that are modeled during the learning step**
- **Explain the decisions, increase the confidence of the researcher in the AI**

Develop a new Symbolic Artificial Intelligence:

- Extract the knowledge in a intelligible way for humans
- With few data and the expert knowledge
- Reproduce the human reasoning

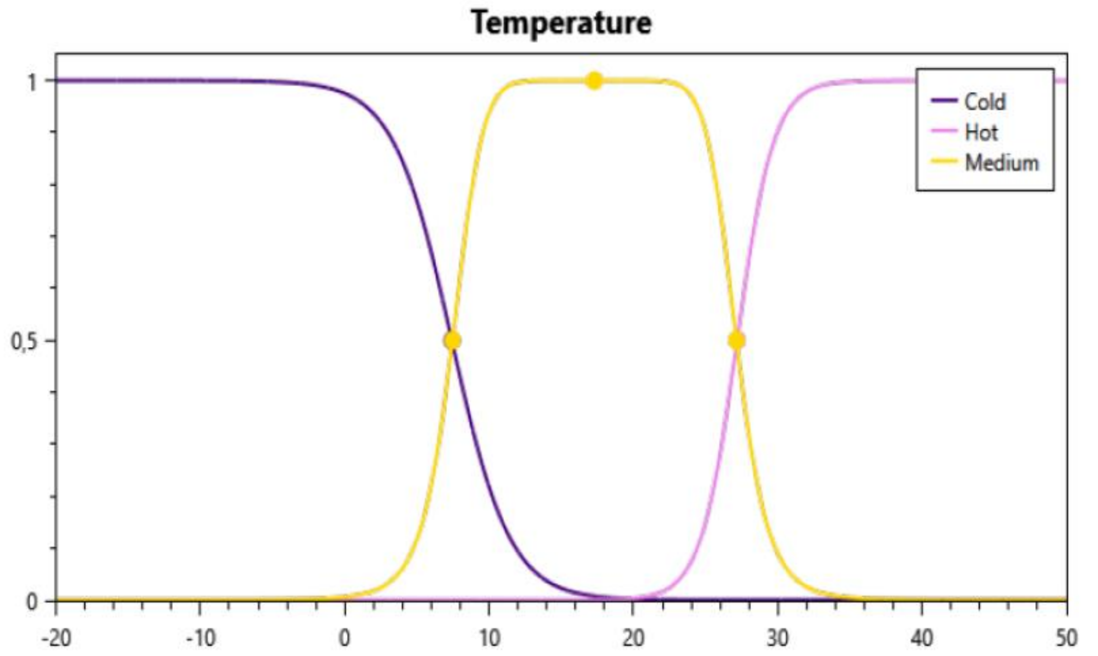
Symbolic AI

- **Fuzzy logic:**
  - Extension of classical logic, which allows to handle uncertainty
  - Get closer to the human reasoning
  - Avoid the threshold effects pertaining to classical logic approaches
  - IF condition THEN conclusion

“IF temperature is low THEN sweating is low”

“IF temperature is low THEN sweating=f(temperature)”

- **Fuzzy inference**
  - Human reasoning through a set of rules



Example of a fuzzy linguistic variable "Temperature”

Application to new material design

Causality detection

- Between process parameters and the observed properties
- Select the parameters that impact the properties

Characterization of the causal link

- Quantification and qualification of the causal link
- Positive or negative causality
- Gradual link

Rules

- Simple rules: “IF the laser power is high THEN the material is thick”
- Gradual rules: ”The higher the laser power, the thicker the material”

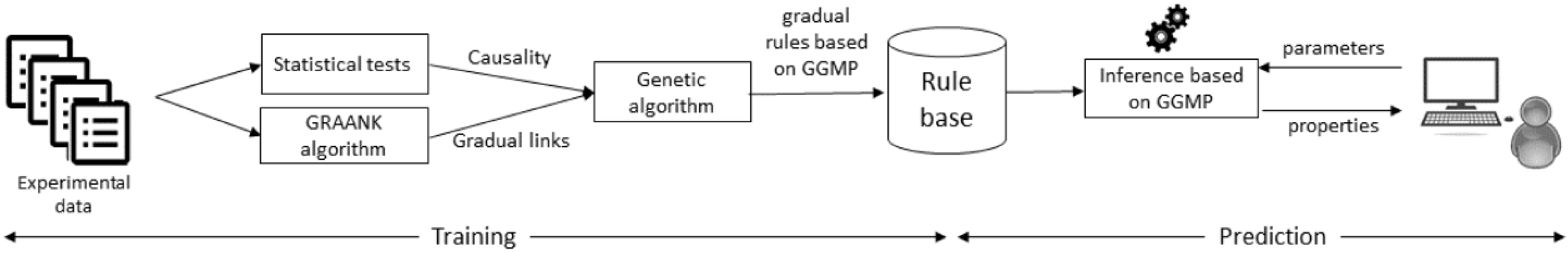
Prediction

- Material properties
- Prediction explanation
- Confidence degree of the prediction
- Experimental plan design support

3 applications of the symbolic AI for new material design

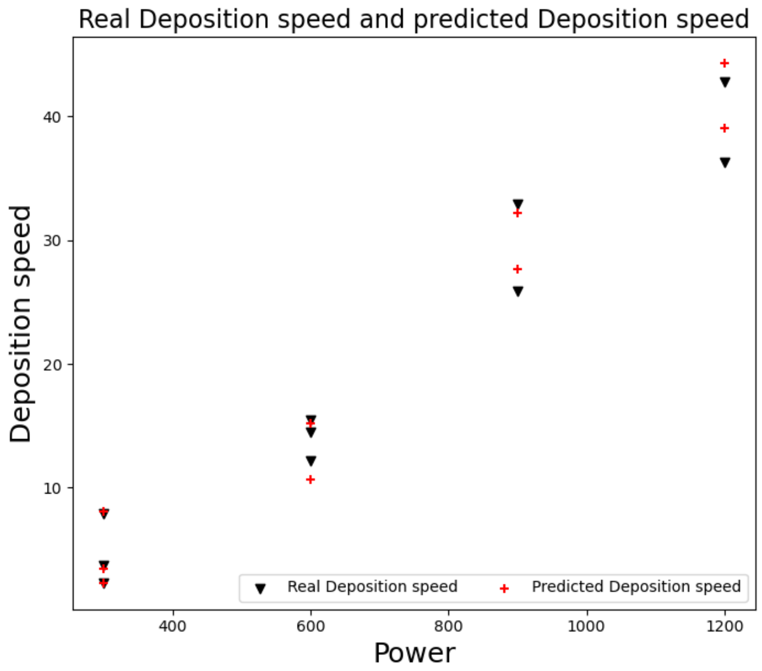
- **Fuzzy gradual rule generation between process parameters and material properties**
  - Causal link detection between parameters and properties by parametric et non parametric tests
  - Qualification of the gradual links between parameters and properties: monotony (Algorithm of gradual pattern detection)

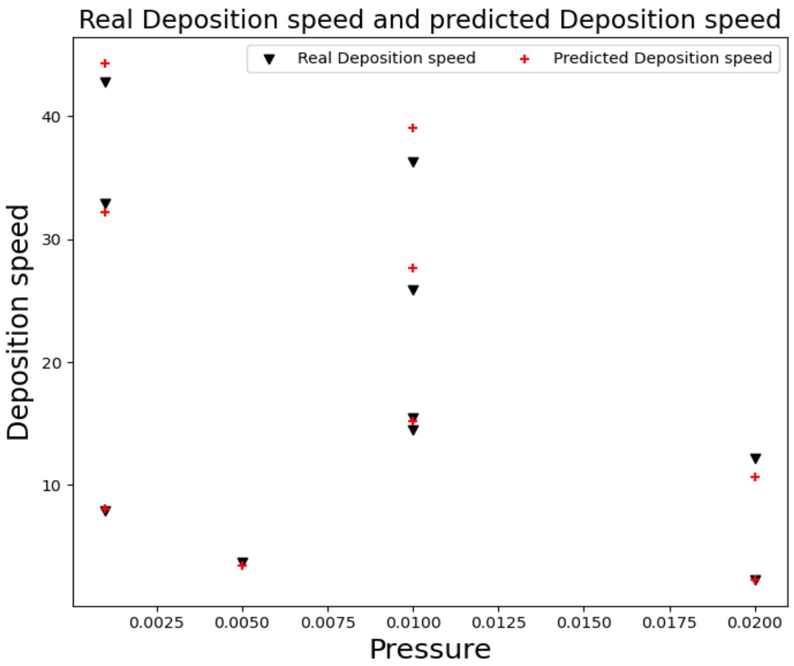
- **Fuzzy inference system based on Generalized Gradual Modus Ponens**
  - Learning the rule parameters by meta-heuristics
  - Test of prediction obtained by the model



1. Physical vapor deposition boosted by Artificial Intelligence

- **Use case: Physical Vapor Deposition**
  - Parameters: pressure, partial pressure, power, speed of passages, number of passages
  - Properties:
    - Physical properties: resistivity, thickness, deposition speed
    - Optical properties: transmittance, optical performance





Prediction and observed values of deposit speed are represented according to the power (left) and the pressure (right) MAPE =9% et RMSE= 1.8

- **Use case: Physical Vapor Deposition**
  - Parameters: pressure, partial pressure, power, speed de passage, number de passages
  - Properties:
    - Physical properties: resistivity, thickness, deposition speed
    - Optical properties: transmittance, optical performance

	RMSE	MAPE (%)
SVR	3.61 ± 1.09	18.75 ± 7.67
Polynomial Regression	1.9 ± 0.5	11.83 ± 3.02
Random Forest	2.83 ± 1.54	12.22 ± 4.15
ANFIS	4.6 ± 3.9	18.10 ± 15.19
XGBOOST	2.62 ± 1.17	14.65 ± 9.45
GGMP	$\mathbf{1.8 \pm 0.31}$	$\mathbf{9.5 \pm 2.23}$

Performance of the prediction of Deposition Speed values based on Power, Scroll Speed and Number of Passages measures using the different selected predictors

	RMSE	MAPE (%)
SVR	108.83 ± 29.09	118.31 ± 43.81
Polynomial Regression	94.19 ± 61.05	43.31 ± 22.67
Random Forest	73.5 ± 20.98	80.12 ± 38.48
ANFIS	96.5 ± 25.87	75.8 ± 32.01
XGBOOST	59.33 ± 23.23	$\mathbf{39.47 \pm 9.42}$
GGMP	$\mathbf{59.12 \pm 19.51}$	57.49 ± 5.77

Performance of the prediction of Thickness values based on Power, Scroll Speed and Number of Passages measures using the different selected predictors

2. Causality between process parameters and properties for optimisation

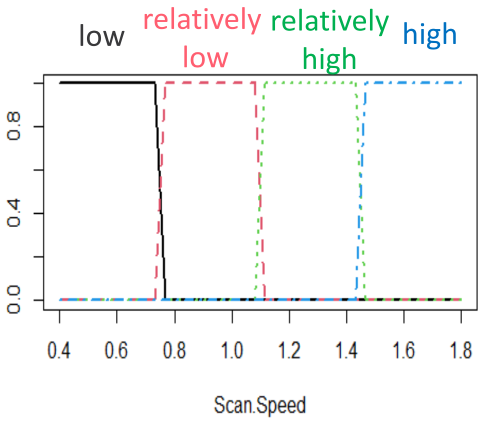
- **Use case:** additive factory (FA #)

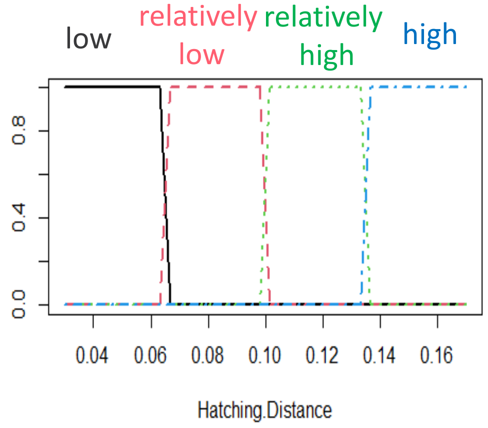
Manufacturing parameters (Causes) ? Materials properties (effects)

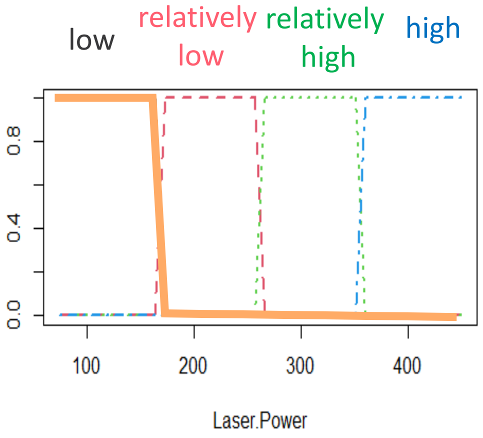
- **Objective :** Propose an automatic method to discover and represent the causal relations (between process parameters and material characteristics), compliant with the **fuzzy logic field:**
  - Causality discovery – by Bayesian graph for fuzzy logic
  - Causal inference – Vocabulary extraction
- **Constraints:**
  - No assumptions about the link type:
    - linearity/non linearity…
    - Data distribution (Gaussian…)
  - Latent variables
  - Missing values
  - Heterogeneous data (categorical or continuous)
  - Few experiences

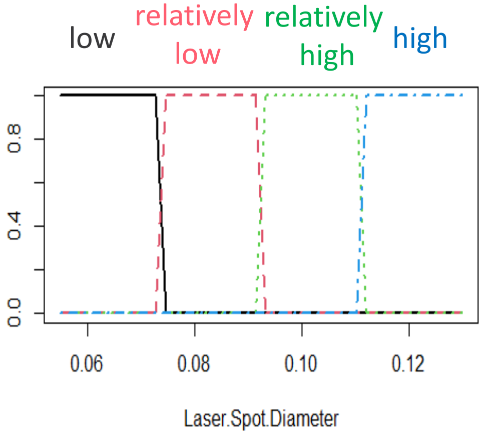
2. Fuzzy causality

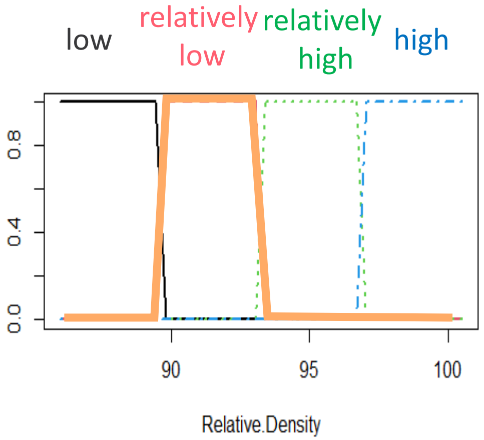
Fuzzy set definition:





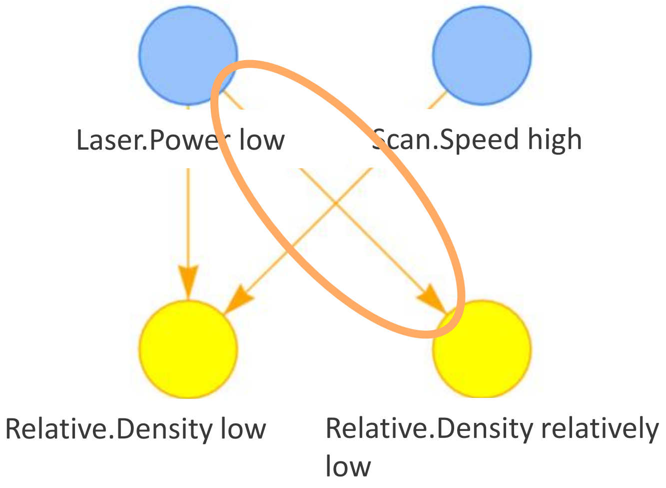






Causality discovery *

between fuzzy sets :



Example :

Belonging to the fuzzy set “Laser power is low” has a causal impact on belonging to the fuzzy set “Relative density is relatively low”

*** With Stochastic Complexity-based Conditional Independence Criterium**

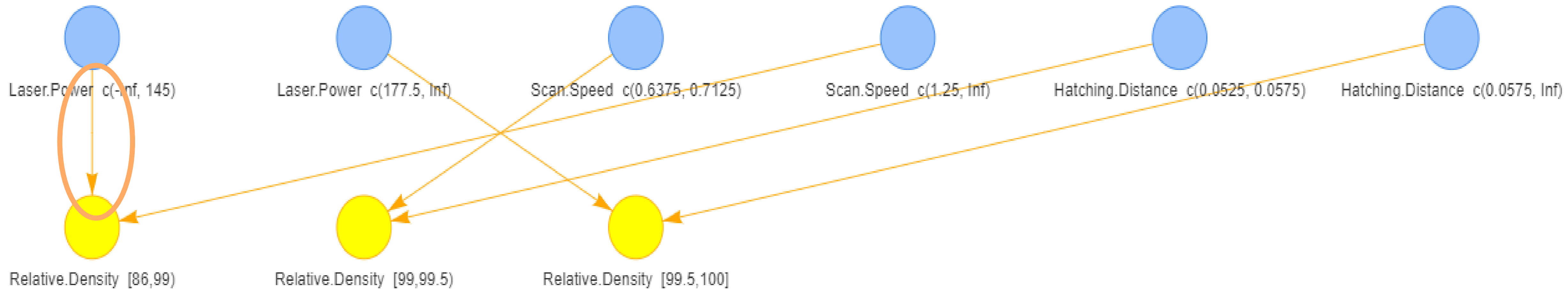
Belonging to the fuzzy set “Laser power is low” has a causal impact on belonging to the fuzzy set “Relative density is relatively low”

Translation in a simple rule

IF « Laser power is low » THEN « Relative density is relatively low ».

Optimization of crisp sets

From crisp sets discretized to form equal frequency bins:



Example : « Laser power < 145 » has a causal influence on « Relative density ∈ [86,99) »

1- Mathematical formalization: optimization under constraints

$$\underset{x,y,z}{\text{maximise}} \quad F(x,y,z)$$

subject to

$$\begin{aligned} \underline{P} \leq P(x,y,z) &\leq \overline{P} \\ g(x,y,z) &\leq \alpha \\ h(x,y,z) &\leq \beta \end{aligned}$$

$$\left.\begin{aligned} \sum_{i=1:n} y_i &= n_{max} \\ \sum_{i=1:n} x_i &= 1 \\ 0 \leq x_i \leq y_i \; i &= 1,\dots,n \end{aligned}\right\} \text{Mixture constraints}$$

x: solvent percentage in the mixture
y: component presence in the mixture
z: experimental parameters
F: quantity to maximize (Analytic function or isolated values)
P,g,h: constraints
- experimental
- physicochemical
- From experts (preferences, intuition)

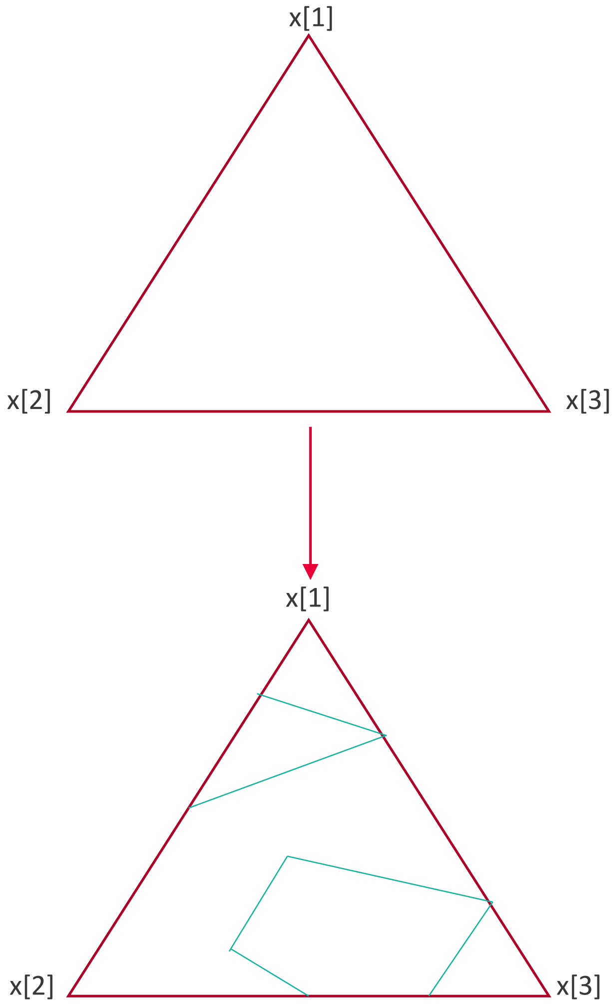
3. EXPERIMENTAL PLAN DESIGN SUPPORT

1- Mathematical formalization: optimization under constraints

$$
\begin{aligned}
&\underset{x,y,z}{\text{maximise}} && F(x,y,z)\\
&\text{subject to} && \\
& && \underline{P} \leq P(x,y,z) \leq \overline{P}\\
& && g(x,y,z) \leq \alpha\\
& && h(x,y,z) \leq \beta\\
& && \left.\begin{array}{r} \sum_{i=1:n} y_i = n_{max}\\ \sum_{i=1:n} x_i = 1\\ 0 \leq x_i \leq y_i \;\; i = 1,\ldots,n \end{array}\right\} \text{Mixture constraints}
\end{aligned}
$$

x: solvent percentage in the mixture
y: component presence in the mixture
z: experimental parameters
F: quantity to maximize (Analytic function or isolated values)
P,g,h: constraints } experimental / physicochemical / From experts (preferences, intuition)

2- Constraint satisfaction to reduce the search space



3. EXPERIMENTAL PLAN DESIGN SUPPORT

1- Mathematical formalization: optimization under constraints

$$\underset{x,y,z}{\text{maximise}} \quad F(x,y,z)$$

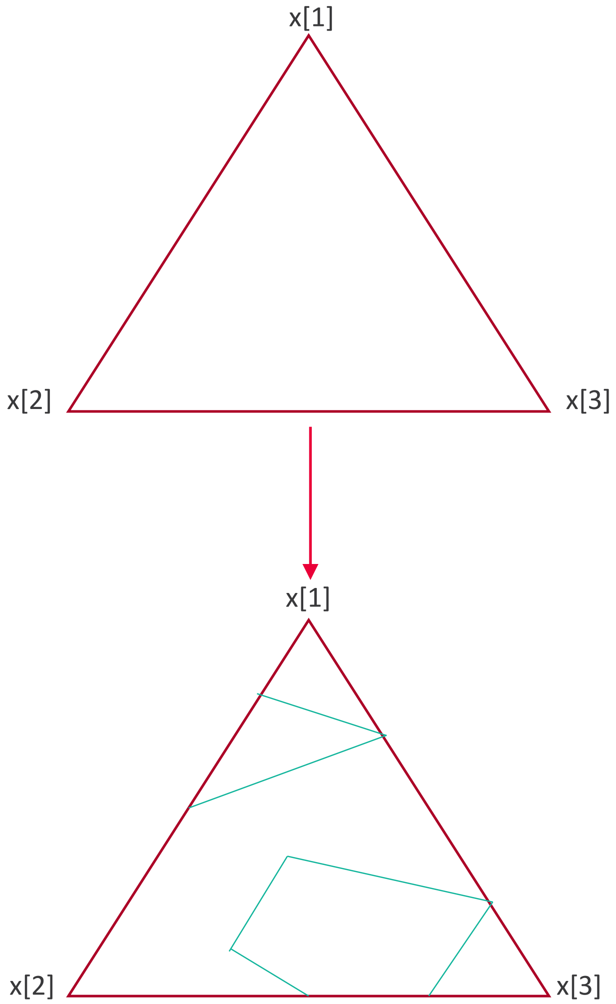
subject to

$$\underline{P} \leq P(x,y,z) \leq \overline{P}$$
$$g(x,y,z) \leq \alpha$$
$$h(x,y,z) \leq \beta$$

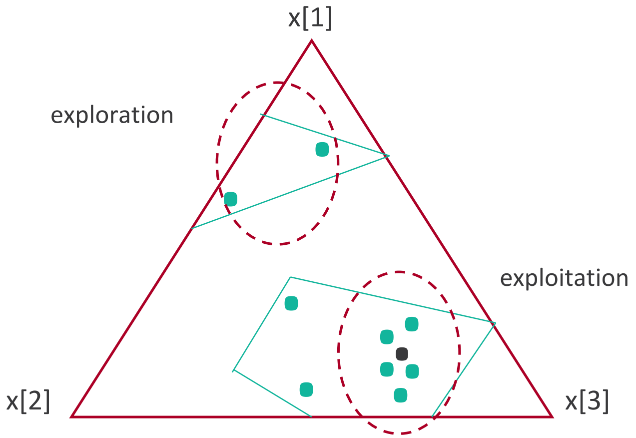
$$\left.\begin{array}{r} \sum_{i=1:n} y_i = n_{max} \\ \sum_{i=1:n} x_i = 1 \\ 0 \leq x_i \leq y_i \;\; i = 1, \ldots, n \end{array}\right\} \text{Mixture constraints}$$

x: solvent percentage in the mixture
y: component presence in the mixture
z: experimental parameters
F: quantity to maximize (Analytic function or isolated values)
P,g,h: constraints } experimental / physicochemical / From experts (preferences, intuition)

2- Constraint satisfaction to reduce the search space



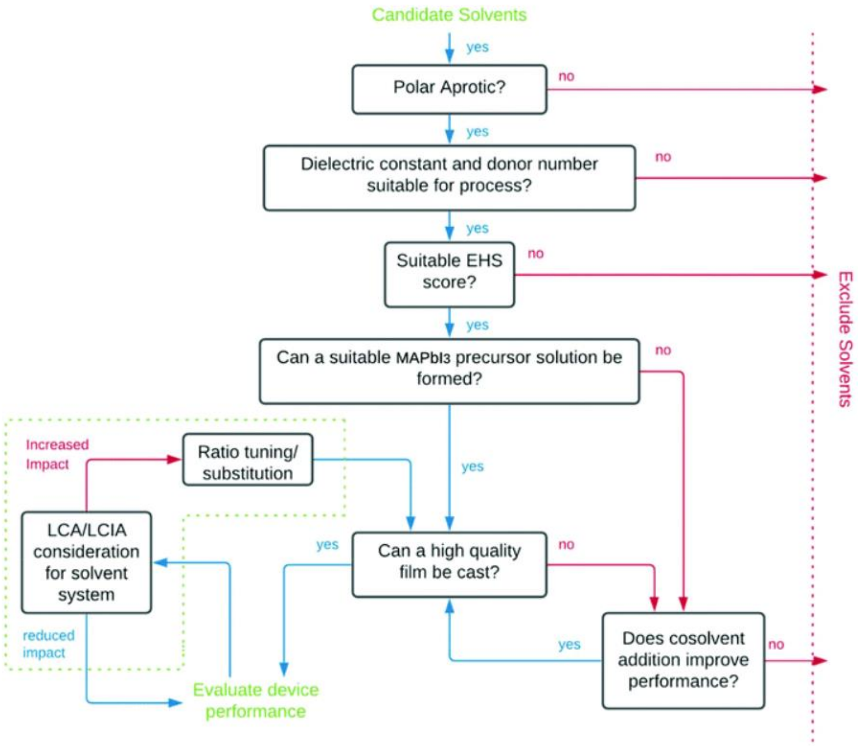
3- Optimum search: adaptive experimental plan



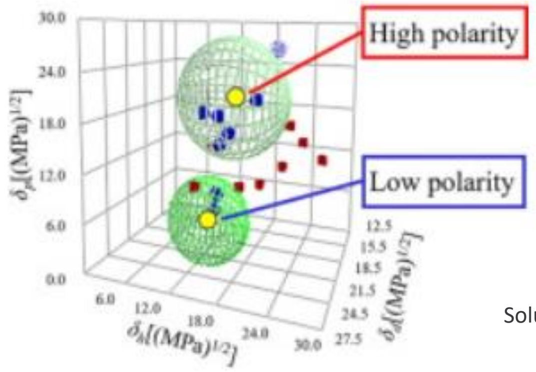
- Sampling methods in search space with complex and uncertain forms
- Genetic algorithm: limitation → not possible to extract knowledge

USE CASE: SOLVENT MIXTURE

In photovoltaic cells with Perovskite : the objective is to replace the solvent DMF used in laboratory by less toxic mixtures



Decision tree for solvent choice (Doolin et al. (2021))



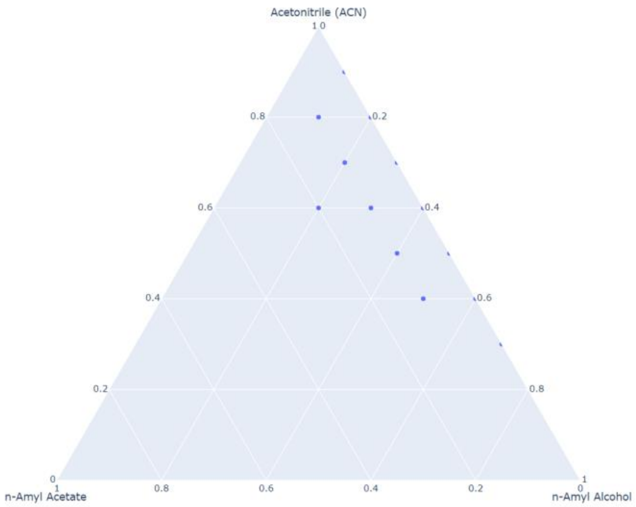
Solubility model of Hansen (Agata et al. (2018))

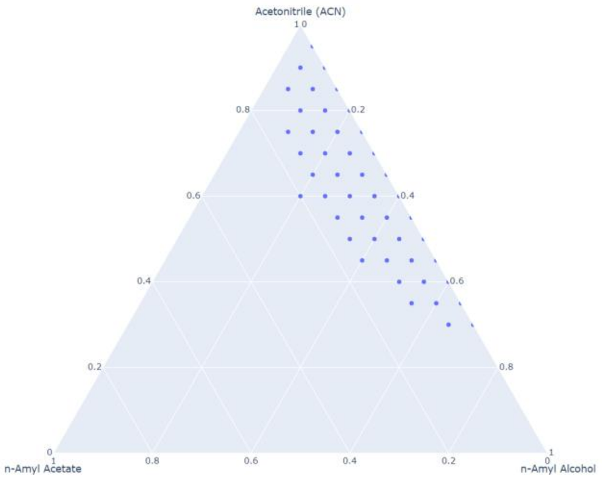
Ink system	Volume [%]	>1 M Solubility	Perovskite formation
GBL	100	Y	Y
AcOH	100	N	N
EtOH	100	N	N
PrOH	100	N	N
PC	100	N	N
DMEA	100	Y	N
GBL/-/AcOH	60/-/40	Y	Y
GBL/EtOH/-	60/40/-	Y	Y
GBL/PrOH/-	60/40/-	Y	Y
GBL/PC/-	60/40/-	Y	N
GBL/DMEA/-	60/40/-	Y	N
GBL/EtOH/AcOH	60/20/20	Y	Y
GBL/PrOH/AcOH	60/20/20	Y	Y
GBL/PC/EtOH	60/20/20	Y	N
GBL/DMEA/EtOH	60/20/20	Y	N
DMF (hazard)	100	Y	Y
DMSO (hazard)	100	Y	Y
NMP (hazard)	100	Y	Y
DMAC (hazard)	100	Y	Y

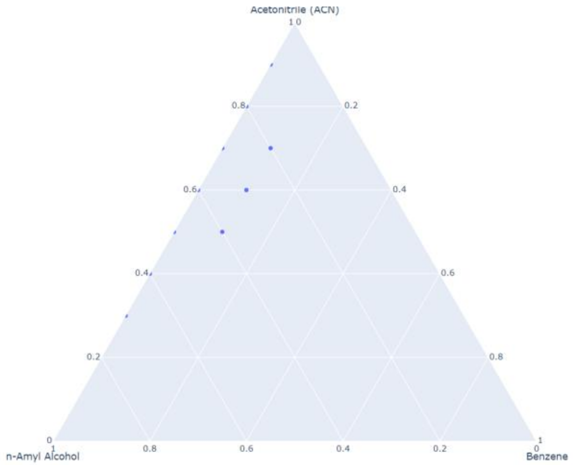
Data example in the state-of-art (Gardner. et al (2016))

USE CASE: SOLVENT MIXTURE

-N=89 solvents
-Solubility constraint (Hansen)
-Step: 0.1, 005

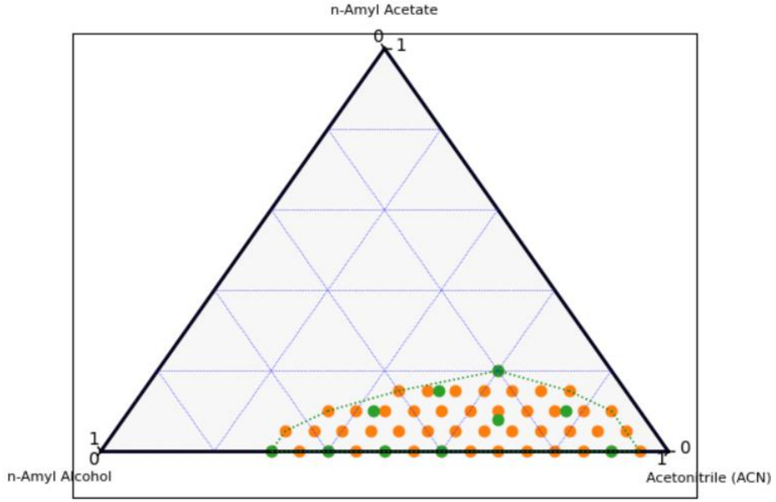






Optimal experimental plan (Expert design 13)

Run	Component 1 A:ACN Percent	Component 2 B:Acetate Percent	Component 3 C:Alcohol Percent	Response 1 R1
1	0,3	0	0,7	
2	0,65957	0,0793631	0,261067	
3	0,428664	0,10483	0,466506	
4	0,599141	0,00882641	0,392032	
5	0,76795	0,101722	0,130329	
6	0,778993	0	0,221007	
7	0,39585	0,0113112	0,592839	
8	0,76795	0,101722	0,130329	
9	0,76795	0,101722	0,130329	
10	0,428664	0,10483	0,466506	
11	0,65957	0,0793631	0,261067	
12	0,50711	0	0,49289	
13	0,599141	0,00882641	0,392032	
14	0,899654	0	0,100346	
15	0,522722	0,146162	0,331116	
16	0,601915	0,199561	0,198524	



ExpressIF® for new material design

- Model the knowledge
- Reproduce human reasoning
- Make decisions
- Justify the decisions

Conclusion: ExpressIF® for new material design

- Model the knowledge
- Reproduce human reasoning
- Make decisions
- Justify the decisions

- Causality discovery and inference between process parameters and properties to select the parameters that induce the expected property
- Several types of rules in natural language for property prediction
- Uncertainty of experimental plan
- Explainable AI to have confidence in the decision or prediction and to understand the role of the process parameter and their interactions for example