



isec



Approche empirique : apport des bases de données et des modèles statistiques

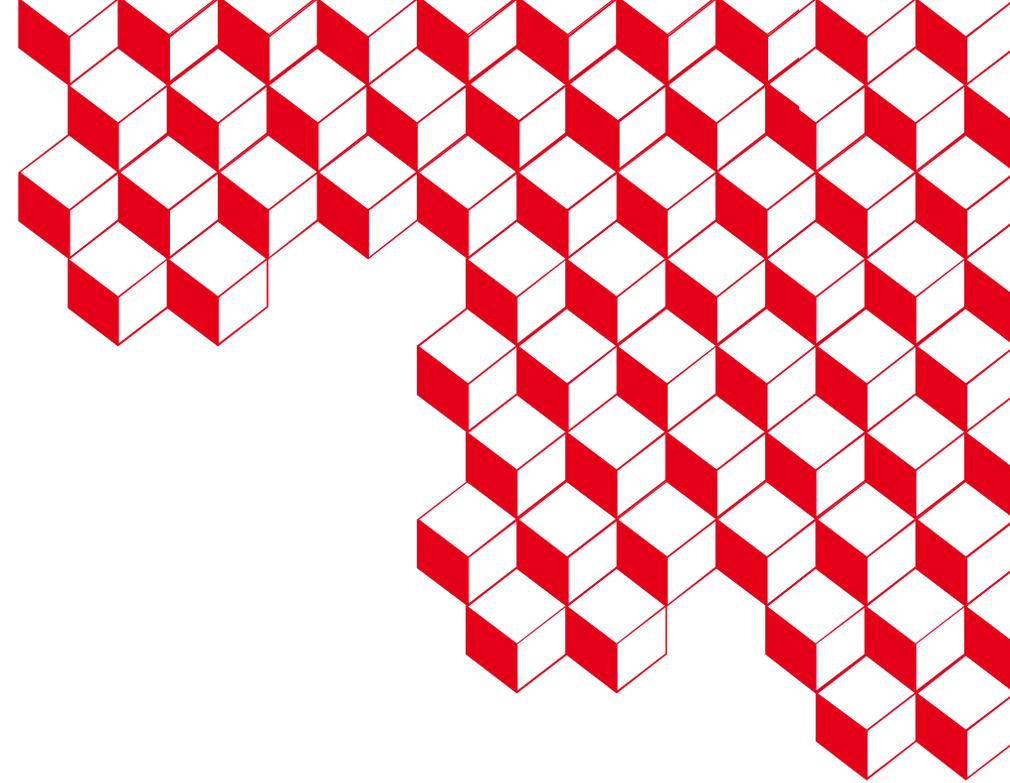
D. Perret, N. Bisbrouck, M. Neyret, C. Chabal

CEA Marcoule

Atelier USTV

Propriétés physiques des verres de la basse à la haute température

Avignon, 12-14 avril 2023





Différentes approches pour la modélisation

- ❑ Approche théorique, cognitive

Basée sur notre connaissance intrinsèque du phénomène, sur les lois fondamentales de la physique et de la chimie (conservation de l'énergie et du moment, équations de la diffusion, principes de la thermodynamique,...)

- ❑ Approche empirique

Basée sur des données expérimentales (« data-driven models »), approche mathématique, statistique, qui ignore toute connaissance physico-chimique du phénomène

- ❑ Approche mixte

Combinaison des deux approches précédentes

Pour chacune de ces trois approches, différents types de modèles : linéaires ou non-linéaires, statiques ou dynamiques, déterministes ou stochastiques, continus ou discrets,...

Différentes approches pour la modélisation

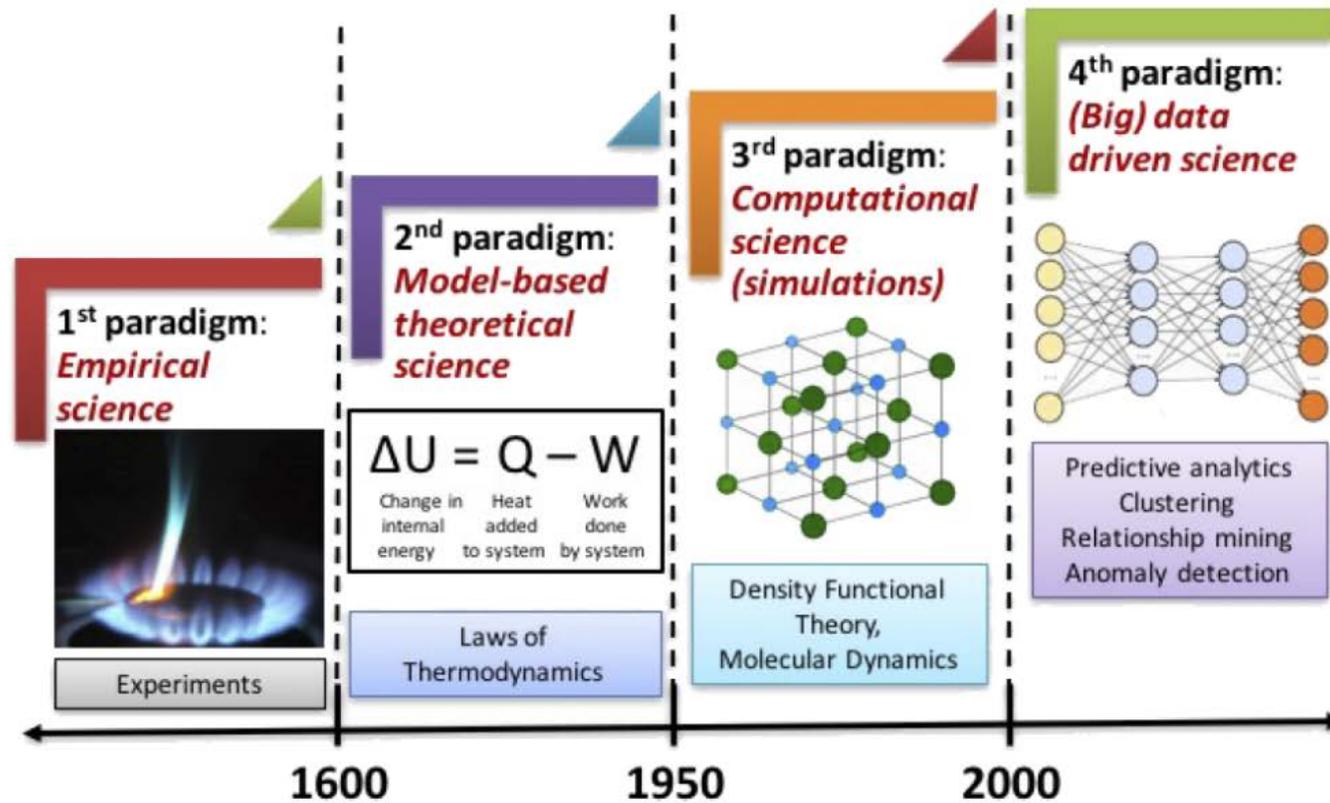
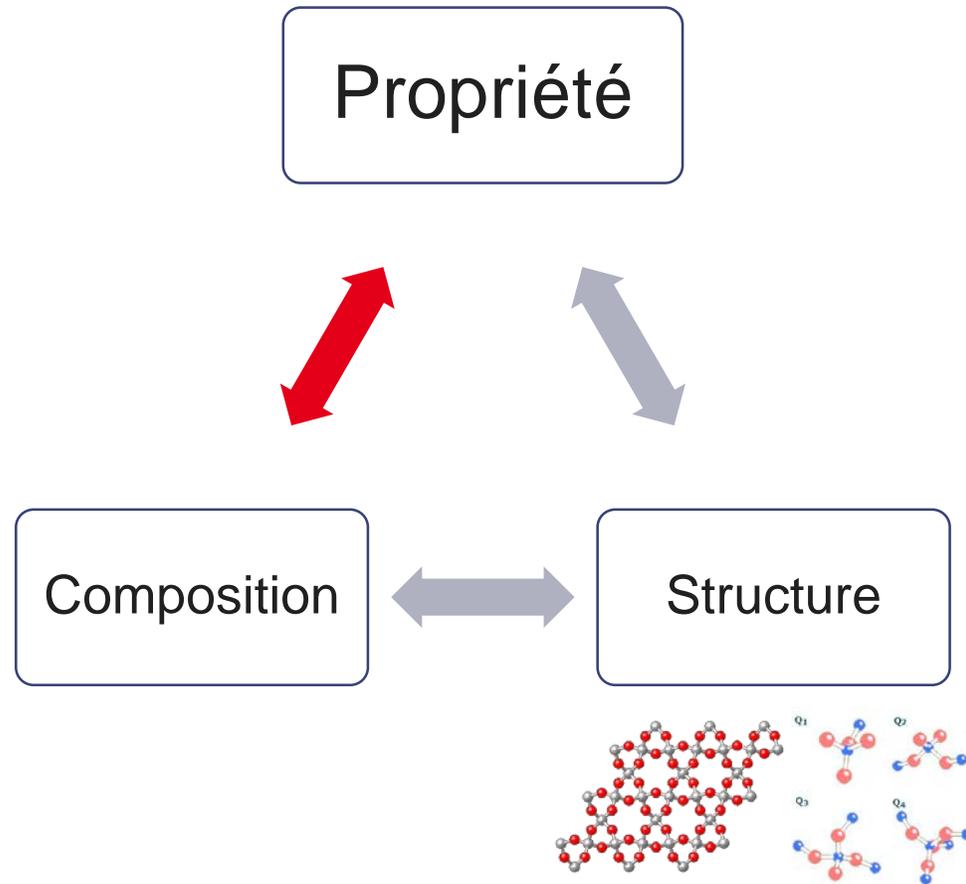


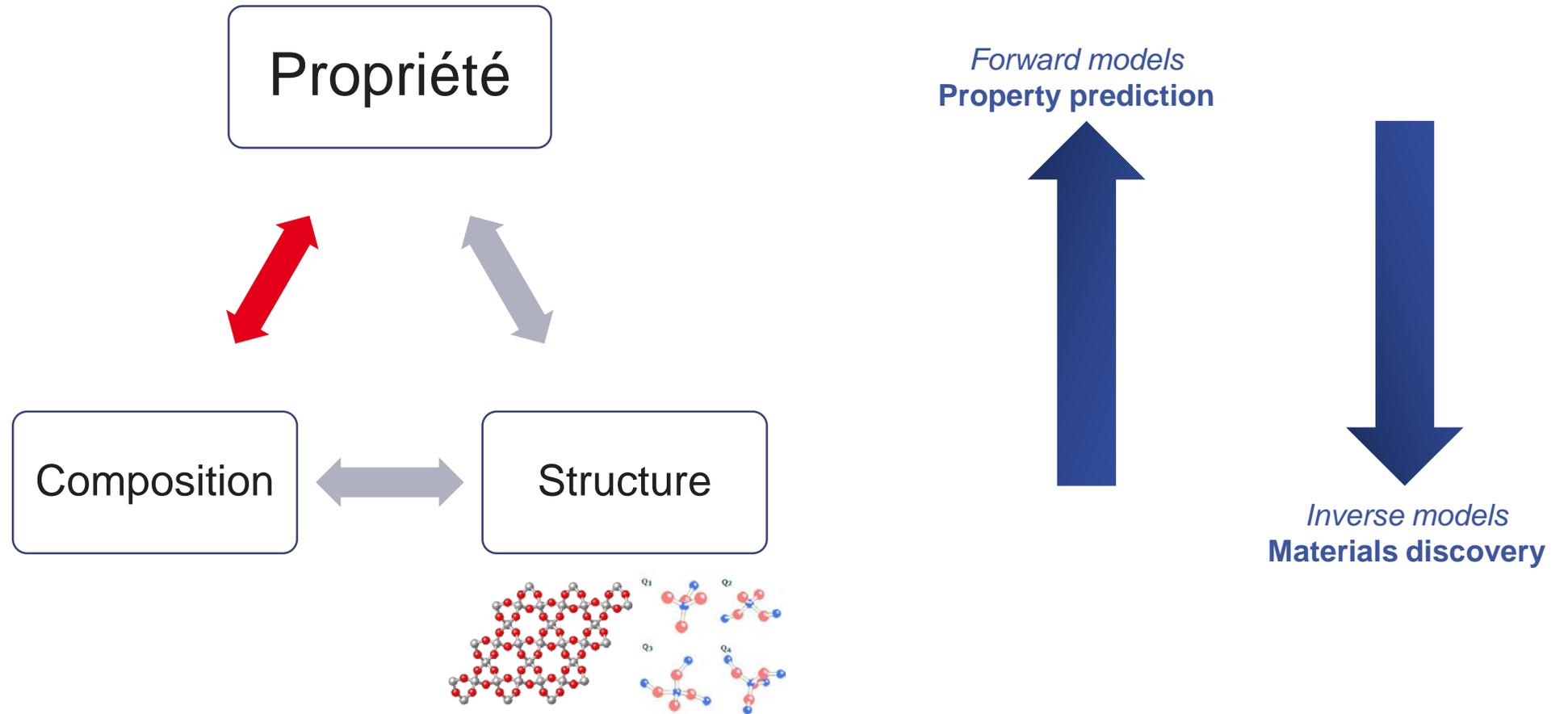
FIG. 1. The four paradigms of science: empirical, theoretical, computational, and data-driven.

From A. Agrawala and A. Choudhary. APL Mater. 4, 053208 (2016); <https://doi.org/10.1063/1.4946894>

Différentes approches pour la modélisation



Différentes approches pour la modélisation



Différentes approches pour la modélisation

- ❑ Premières tentatives de calcul des propriétés des verres à partir de leur composition par Winckelmann et Schott (fin du 19^{ème} siècle)

❑ Principe d'Additivité

M.B. Volf, *Mathematical Approach to Glass*, Elsevier Science Publishers, 1988

$$G = \sum g(G)_i x_i$$

G is the property of the glass
 $g(G)_i$ is the additive factor for oxide i and property G
 x_i is the amount of oxide i

- Applicable sur des domaines restreints de composition
 - Non valable si présence de mécanismes complexes (cristallisation, séparation de phase, anomalies dans la polymérisation du réseau vitreux, interactions ioniques,...)
- ❑ Dans le cas de la viscosité, différentes méthodes de calcul proposées depuis 1988 (Lakatos, Lyon, Mazurin, Hrma, Priven, Okhotin, Fluegel,...)

La modélisation statistique dans les verres

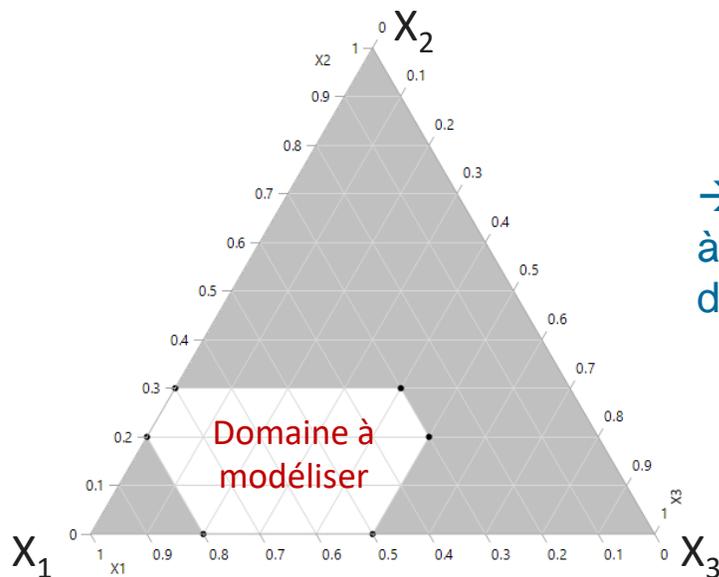


L'approche par
« plans d'expériences »

L'approche « Bases de données,
Machine Learning »

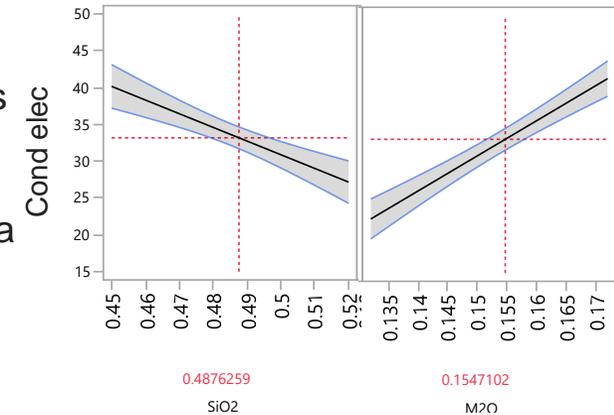
L'approche « plans d'expériences »

Principe de la méthodologie des plans d'expériences



→ Obtenir le **maximum d'informations** à partir d'un nombre **minimum** d'essais expérimentaux

Identification des constituants ayant un **effet significatif** sur la réponse

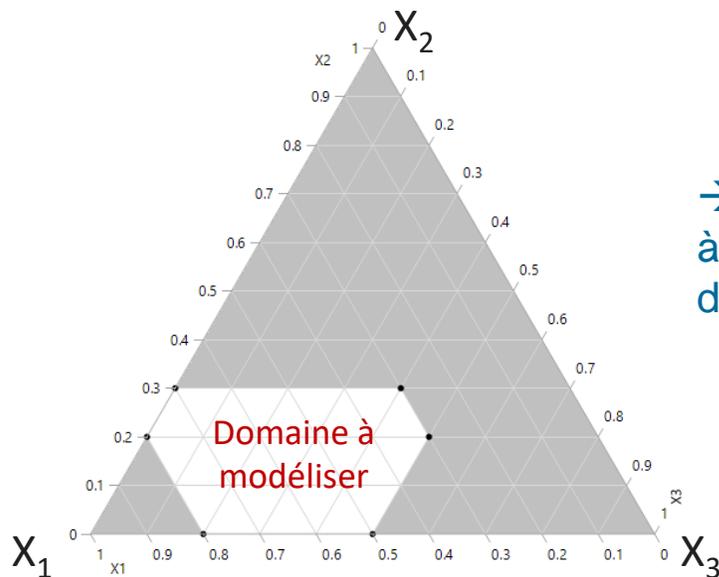


Modèle de prédiction de la réponse en tout point du domaine de composition

$$E(Y) = \sum_{i=1}^p a_i x_i + \text{selected} \left\{ \sum_{i=1}^{p-1} \sum_{j=1}^p a_{ij} x_i x_j \right\}$$

L'approche « plans d'expériences »

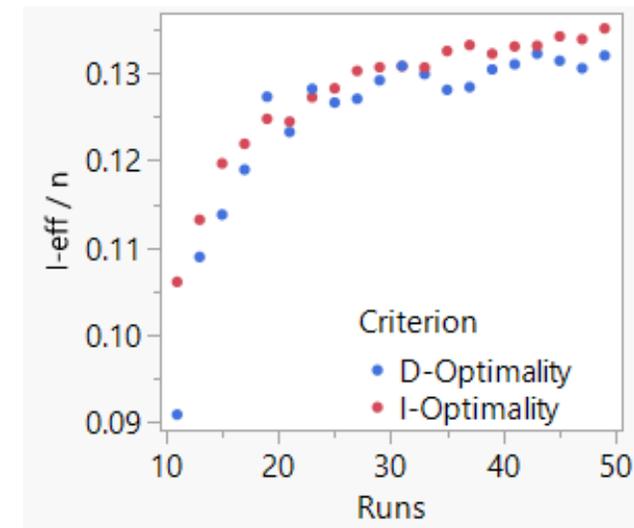
Principe de la méthodologie des plans d'expériences



→ Obtenir le maximum d'informations à partir d'un **nombre minimum** d'essais expérimentaux



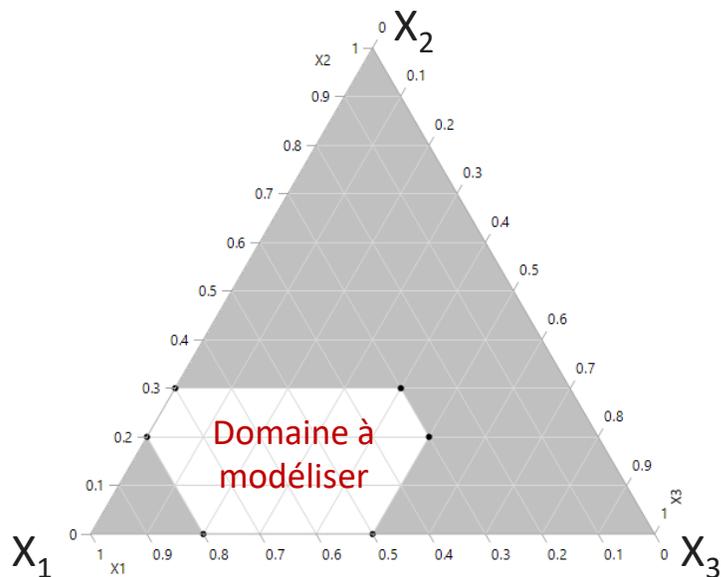
Evolution des **critères d'optimalité** en fonction du nombre d'essais



→ Environ 20 runs pour un PEX à 8 variables avec contraintes

L'approche « plans d'expériences »

Principe de la méthodologie des plans d'expériences



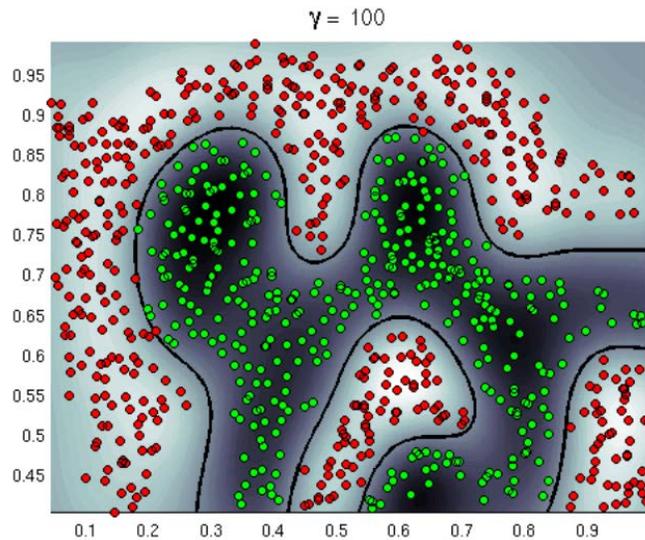
- Points clés :
 - Définition des bornes du domaine
 - Détermination du nombre optimal d'essais
 - Sélection du meilleur modèle (risque overfitting)
 - Validation du modèle
- Méthode applicable pour des verres jusqu'à 10-12 oxydes
- Modèles très robustes, mais pour des étendues modérées en composition

L'approche Base de données, Machine Learning

☐ Quelques techniques courantes de ML

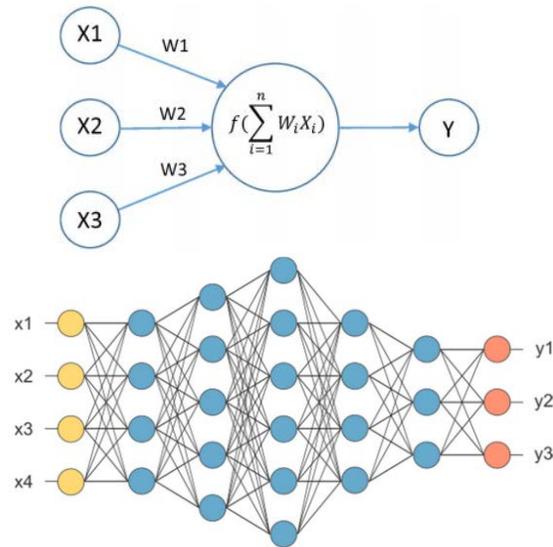
Exemples de méthodes non-linéaires :

Support vector machines, random forests, neural networks, ...

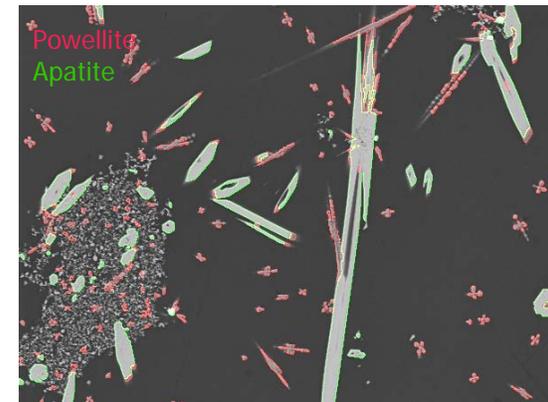


<http://openclassroom.stanford.edu/>

Description schématique d'un réseau de neurones



Exemple d'application en analyse d'images
(identification de cristaux)



Test	
Superclasse	Measures
Generalized RSquare	0,9647809
Entropy RSquare	0,8870834
RMSE	0,1757289
Mean Abs Dev	0,0775361
Misclassification Rate	0,0348259
-LogLikelihood	199,39034
Sum Freq	1608

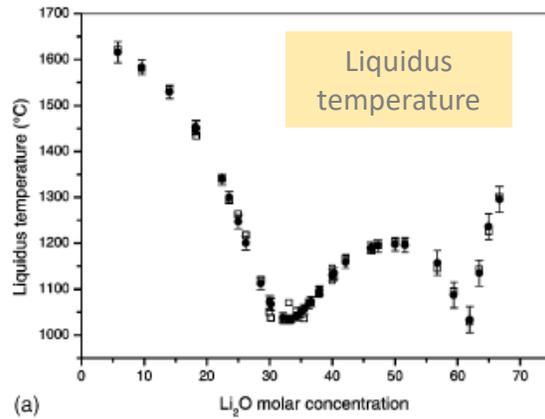
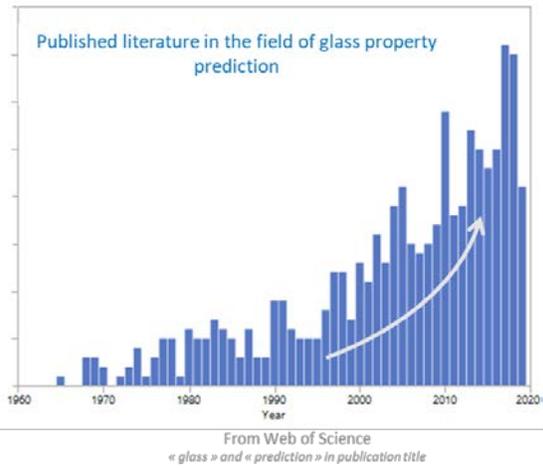
Confusion Matrix			
Actual	Predicted		
Superclasse	Apatite	Platinoide	Powellite
Apatite	541	1	7
Platinoide	6	527	13
Powellite	1	28	484

Confusion Rates			
Actual	Predicted		
Superclasse	Apatite	Platinoide	Powellite
Apatite	0,98543	0,00182	0,01275
Platinoide	0,01099	0,96520	0,02381
Powellite	0,00195	0,05458	0,94347

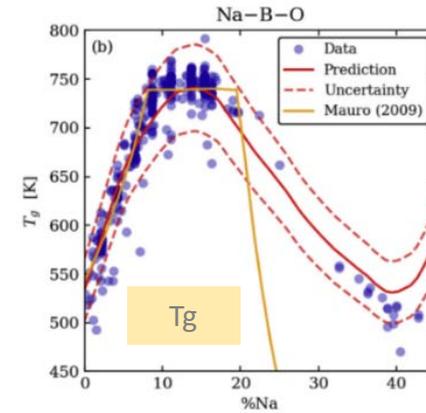


L'approche Base de données, Machine Learning

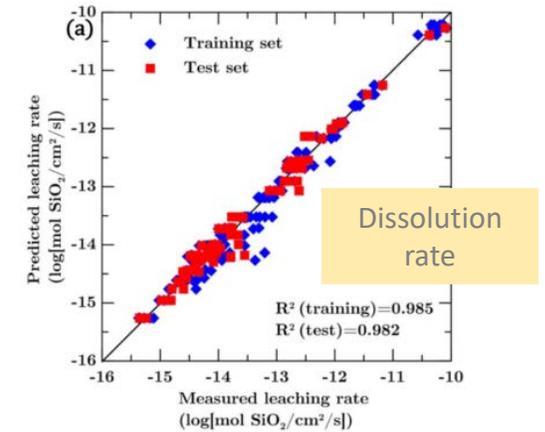
Quelques exemples d'application du ML dans les verres



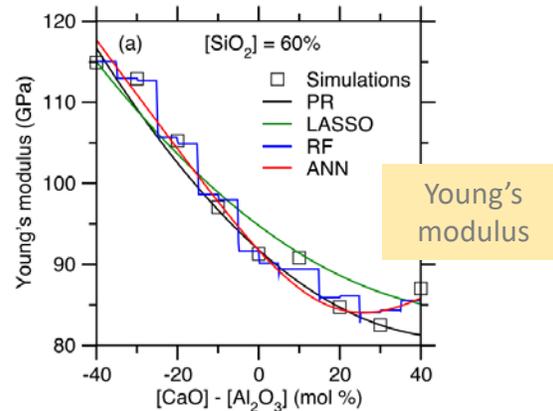
C. Dreyfus, G. Dreyfus / *Journal of Non-Crystalline Solids* 318 (2003) 63–78



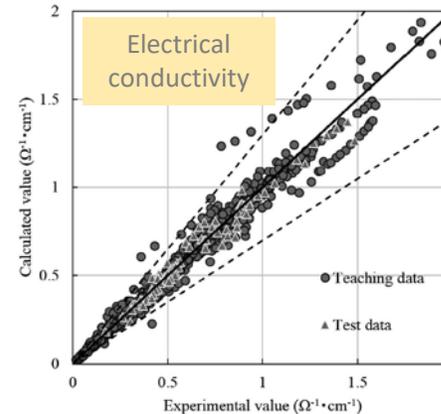
D.R. Cassar et al., *Acta Materialia*, 159 249-256 (2018)



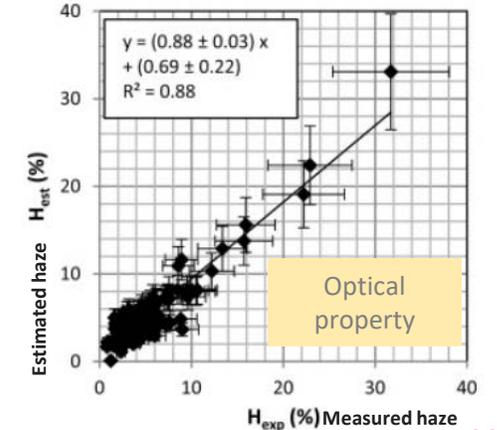
N.M. Anoop Krishnan et al.
Journal of Non-Crystalline Solids 487 (2018) 37–45



K. Yang et al., *Scientific Reports*, 8739, 9 (2019)



Y. Haraguchi et al., *ISIJ International*, Vol. 58 (2018), No. 6, pp. 1007–1012



A. Verney-Caron et al., *Atmospheric Environment*, 54 141-148 (2012)

L'approche Base de données, Machine Learning

❑ Les bases de données commerciales

The screenshot shows the SciGlass Professional 7.10 interface. On the left is a sidebar with various query and analysis tools. The main window is titled "Search for Glasses (Experimental Data)" and contains a table with columns for "Included", "Min", "Max", "Excluded", and "System". The "System" column lists various chemical formulas like SiO₂, B₂O₃, Na₂O, Al₂O₃, CaO, MgO, ZnO, MoO₃, and Fe₃O₄. Below the table is a "Property" section with columns for "Property", "Unit", "Min", and "Max". The SciGlass logo is visible at the bottom of the window.

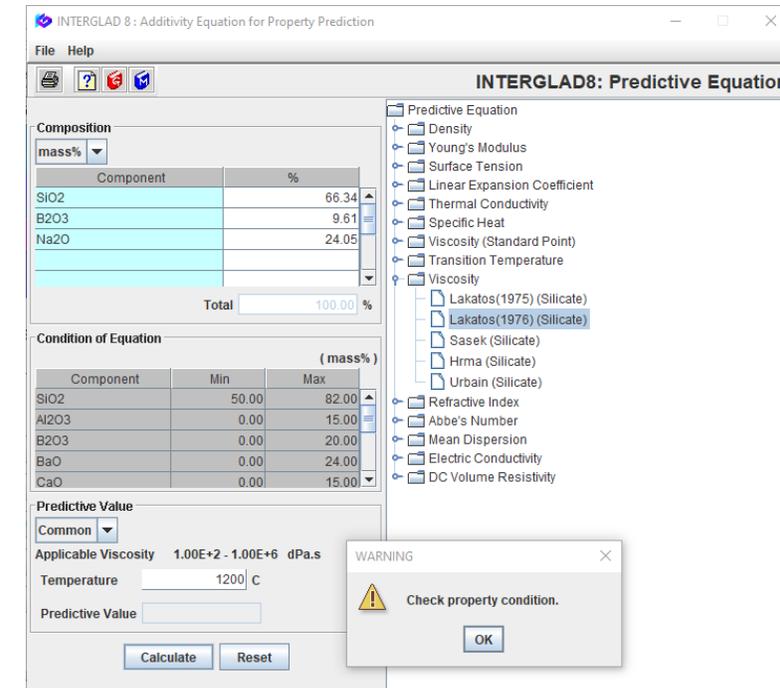
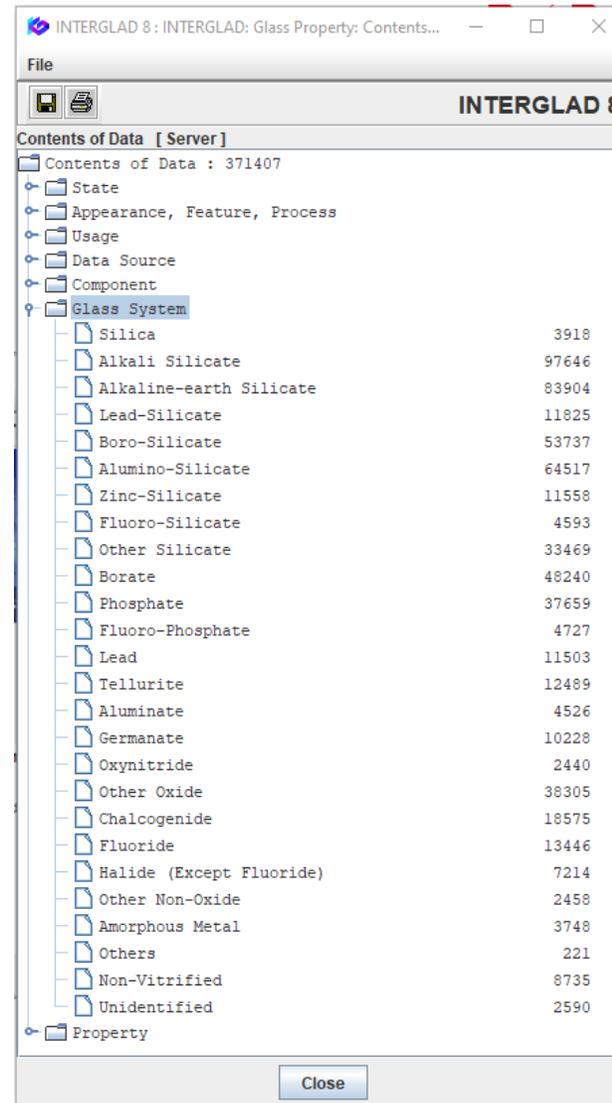
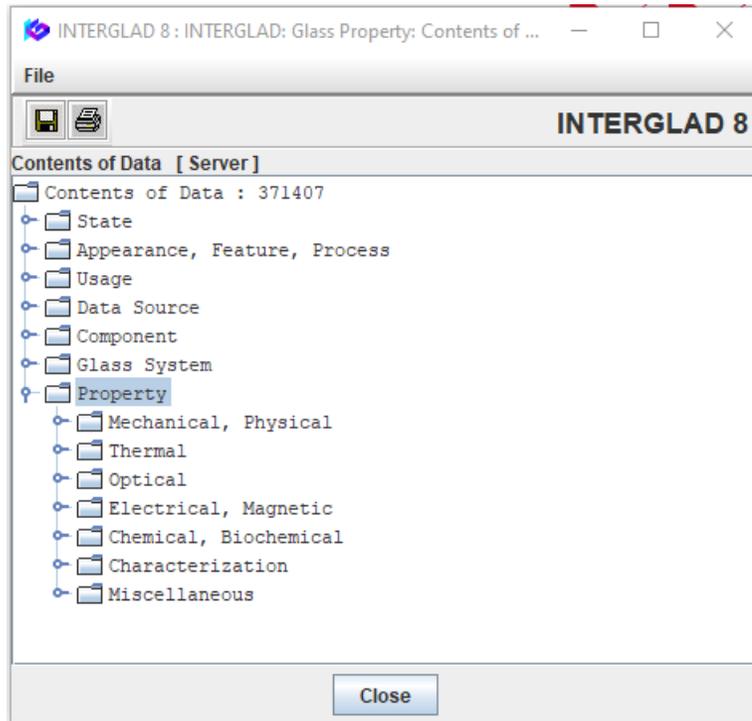
The screenshot shows the INTERGLAD 8: Main interface. It features a header with the "International Glass Database System" logo and the text "INTERGLAD ver.8" and "NEW GLASS FORUM". Below the header is the title "Glass Property & Structure Database" and the version "INTERGLAD 8 (ver.8.3.1.0.03)". On the left side, there are five buttons: "Search Property Data", "Property Prediction", "User Data (Property Data)", "Search Structure Data", and "Exit". On the right side, there is a text box containing the following information:

The international glass database system INTERGLAD was released in 1991 for the first time in the world by the New Glass Forum. Newly released Ver.8 has composition/property data of approximately 340,000 glasses and includes a glass structure database newly developed with structural data of approximately 10,000 glasses. Besides, new functions for increasing prediction accuracy and tools for easier search and prediction have been added. Main functions are

- 1) Search of glass data,
- 2) Analysis of searched glasses,
- 3) Prediction of properties and design of glass composition,
- 4) Registration and utilization of user data.

L'approche Base de données, Machine Learning

Les bases de données commerciales



L'approche Base de données, Machine Learning

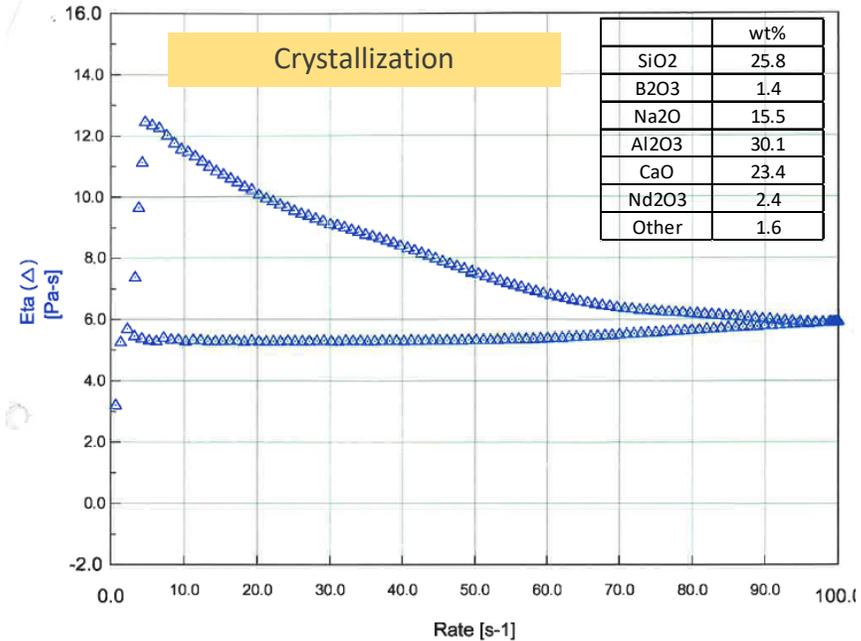
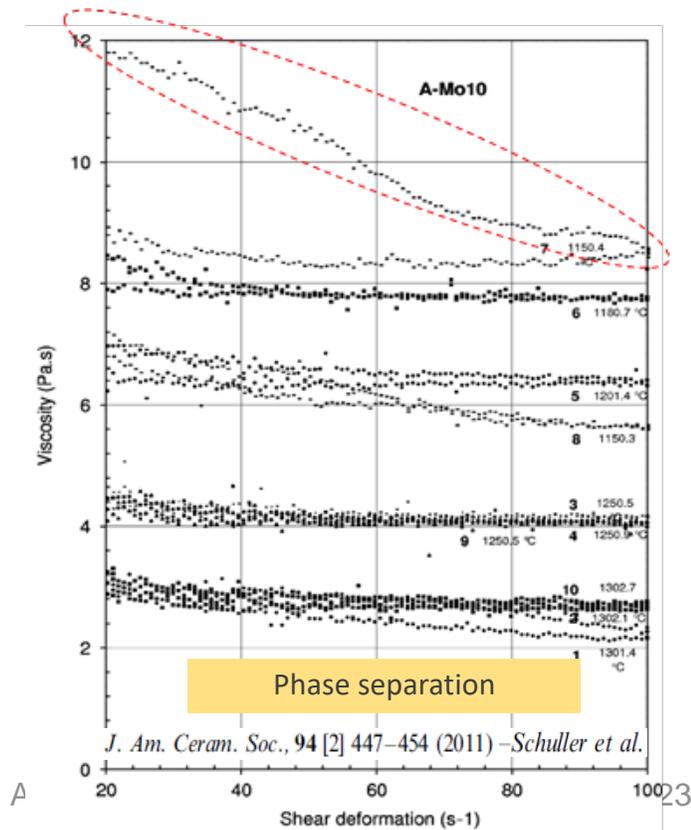
□ Cas d'application : modélisation prédictive de la viscosité

➤ Pourquoi la viscosité des fontes verrières est-elle si difficile à prédire ?

→ Gammes de variation de η vs T et compo (~ 13 ordres de grandeur)

→ Complexité des mécanismes qui peuvent se produire (séparation de phases, cristallisation, modifications structurales,...)

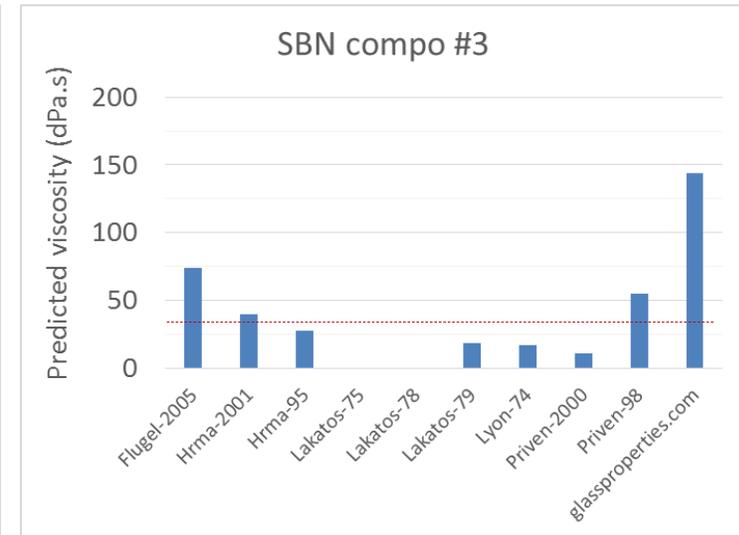
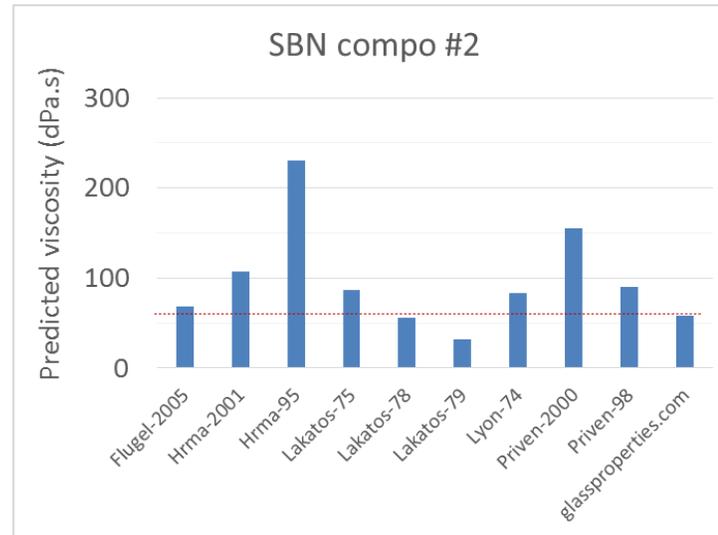
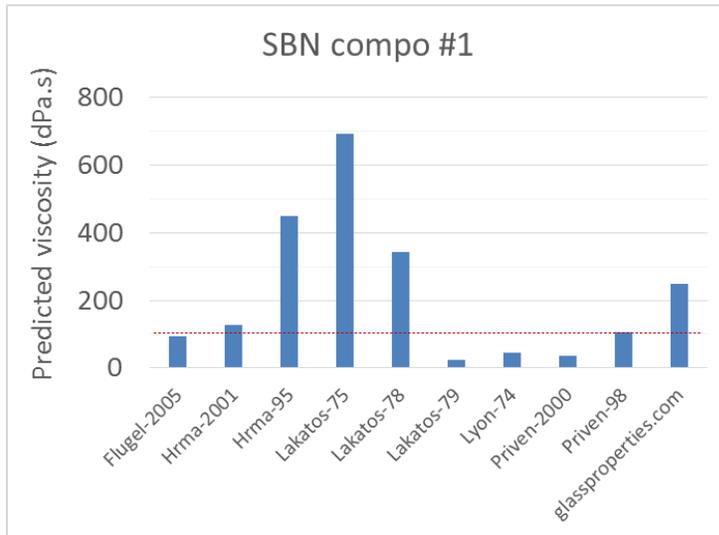
Oxide mol%	AMo10 batch no. 5052
SiO ₂	46.5
Na ₂ O	11.0
B ₂ O ₃	14.4
Li ₂ O	0.2
CaO	7.9
MoO ₃	5.1
P ₂ O ₅	1.6
Al ₂ O ₃	5.4
ZnO	5.4
ZrO ₂	2.0
Nd ₂ O ₃	0.2
Minors	0.3
Sum	100.0



L'approche Base de données, Machine Learning

□ Cas d'application : modélisation prédictive de la viscosité

- Pourquoi la viscosité des fontes verrières est-elle si difficile à prédire ?
 - Gammes de variation de η vs T et compo (~ 13 ordres de grandeur)
 - Complexité des mécanismes qui peuvent se produire (séparation de phases, cristallisation, modifications structurales,...)
- Exemple : prédiction de la viscosité de verres SBN à 1200°C

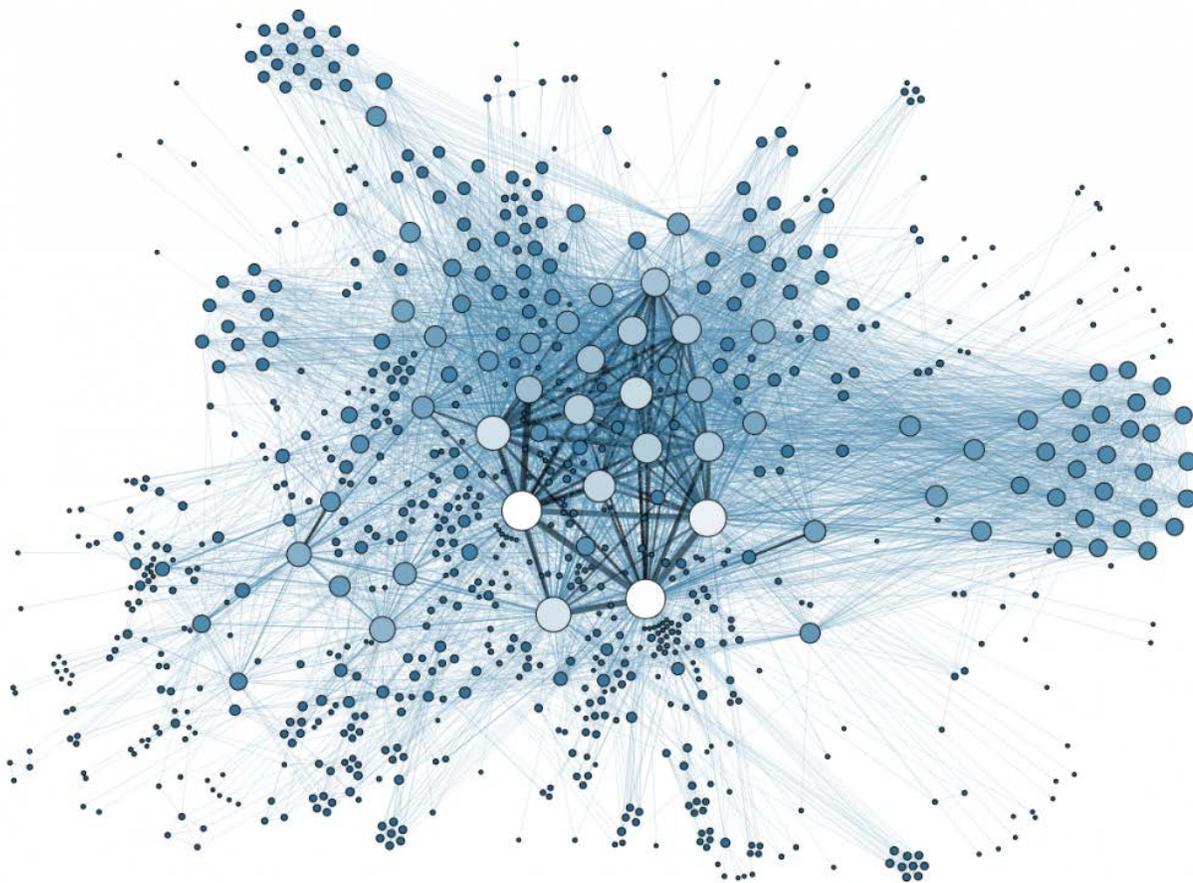


..... Experimental data

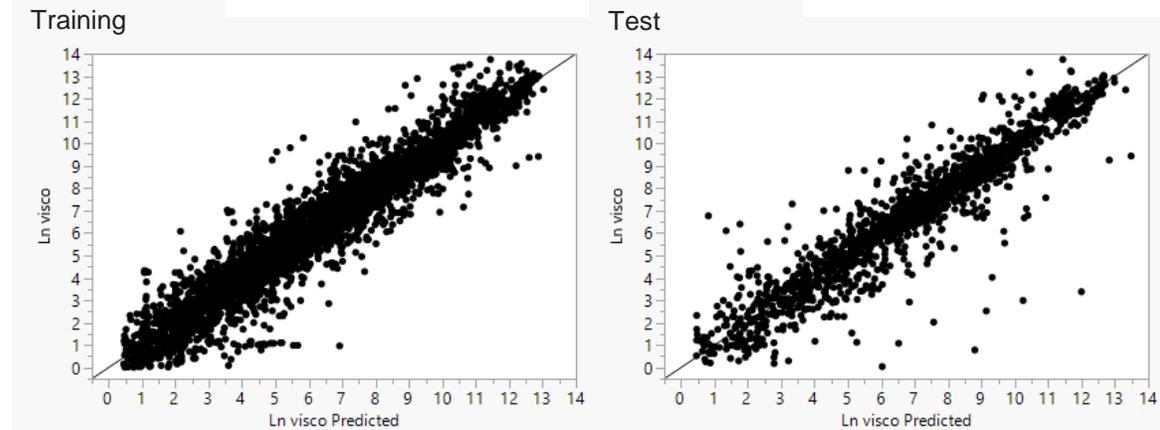
L'approche Base de données, Machine Learning

□ Cas d'application : modélisation prédictive de la viscosité

Notre base de données de viscosité (N=16,000)



Tentative de modélisation par réseau de neurones en considérant l'ensemble de la base de données



Ln visco	
Measures	Value
RSquare	0.9287269
RMSE	0.7518628
Mean Abs Dev	0.496624
-LogLikelihood	6435.0916
SSE	3208.6293
Sum Freq	5676

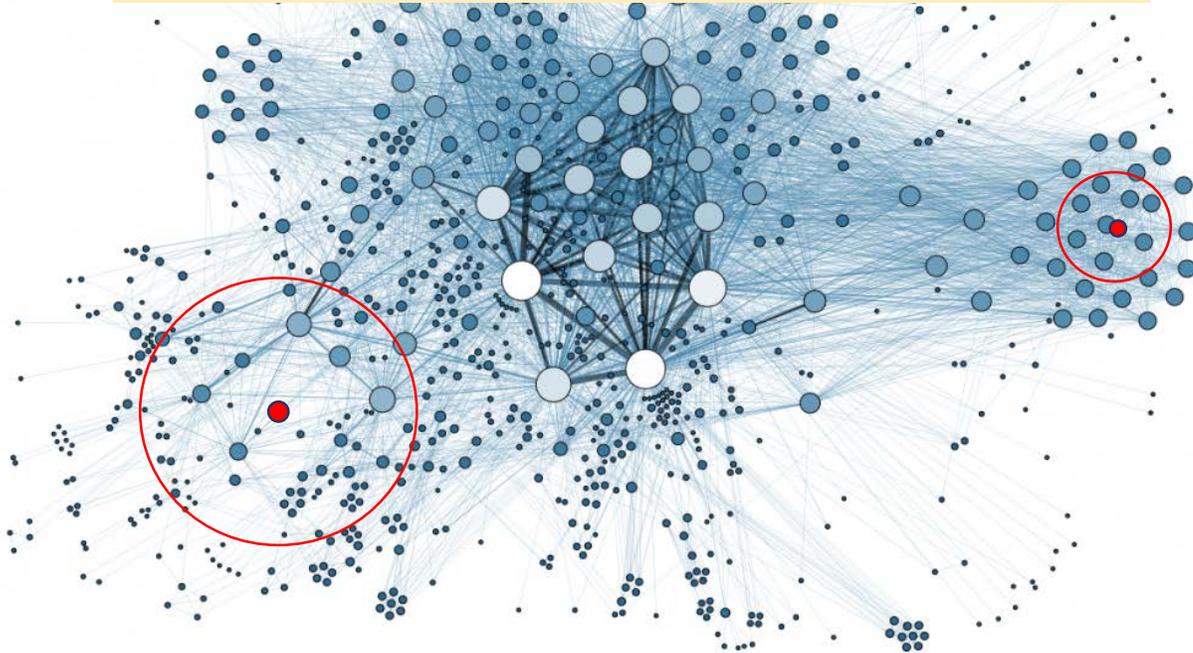
Ln visco	
Measures	Value
RSquare	0.8822247
RMSE	0.9682362
Mean Abs Dev	0.5912263
-LogLikelihood	2545.9066
SSE	1721.2158
Sum Freq	1836

L'approche Base de données, Machine Learning

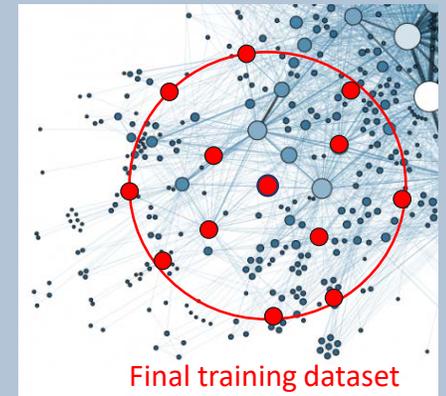
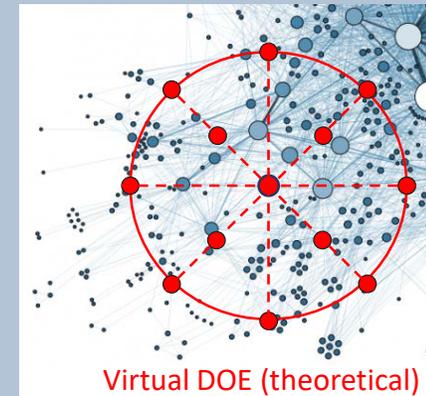
□ Cas d'application : modélisation prédictive de la viscosité

Outil prédictif MProV développé au CEA

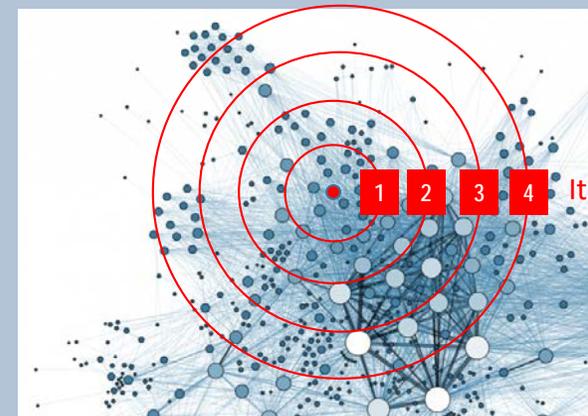
Méthodologie de création de la base d'apprentissage de façon dynamique et automatisée



Méthode 1



Méthode 2



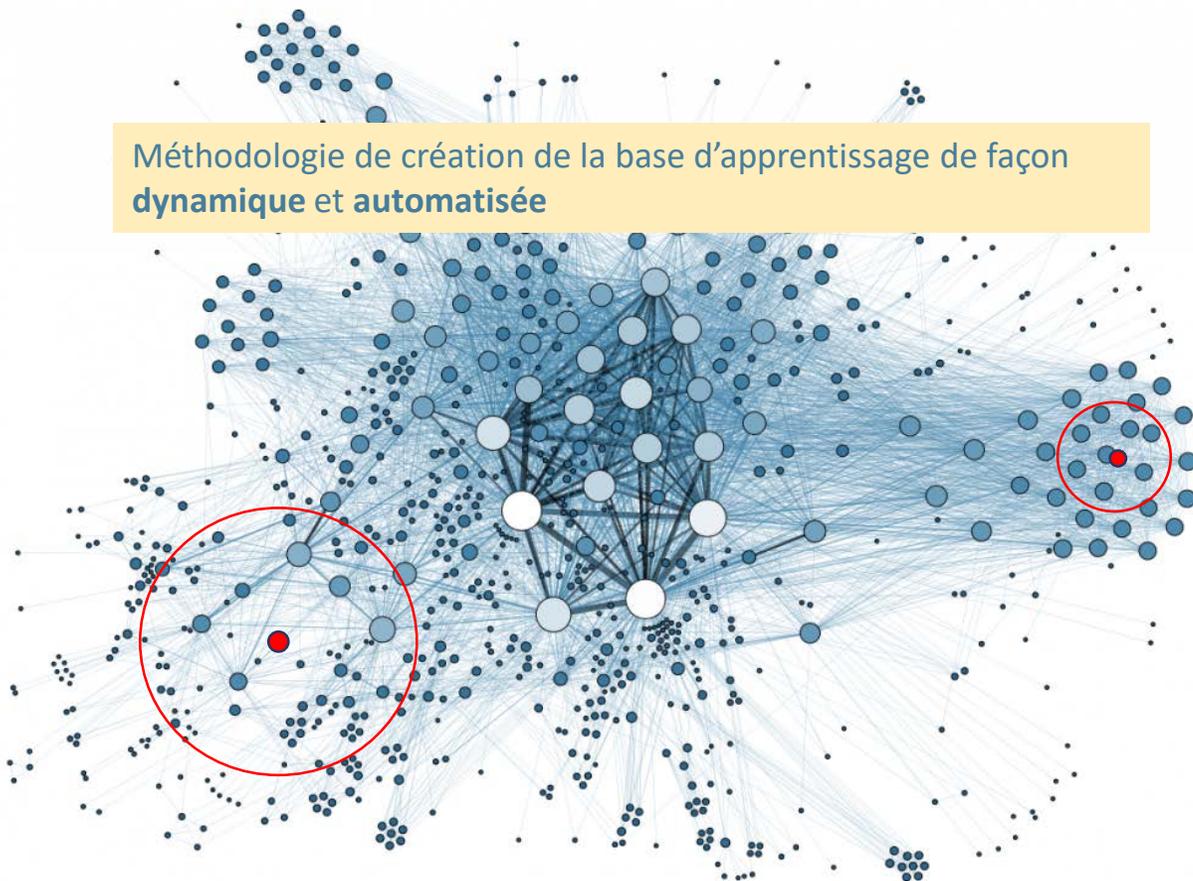
Iterative enlargement

L'approche Base de données, Machine Learning

□ Cas d'application : modélisation prédictive de la viscosité

Outil prédictif MProV développé au CEA

Méthodologie de création de la base d'apprentissage de façon dynamique et automatisée



Pour chacune des deux méthodes :

3 algorithmes de prédiction implémentés dans l'outil

→ Modèle classique (polynomial, régression multilinéaire)

→ Régression linéaire généralisée

→ Réseaux de neurones

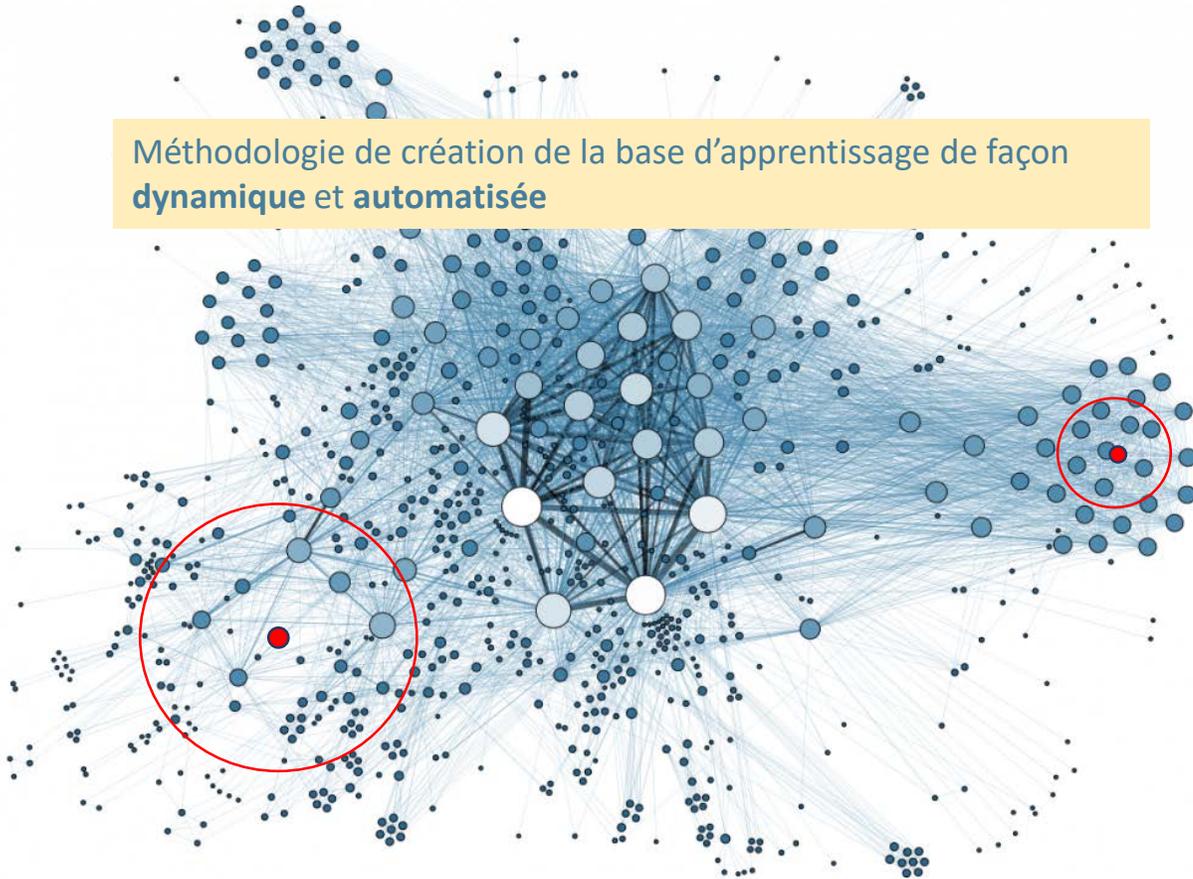
⇒ 6 valeurs de viscosité prédites

L'approche Base de données, Machine Learning

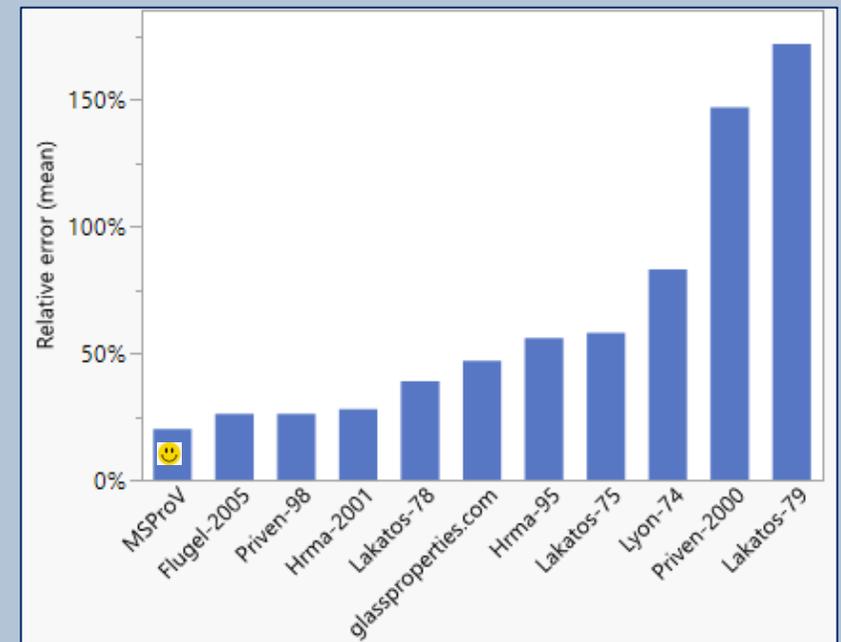
□ Cas d'application : modélisation prédictive de la viscosité

Outil prédictif MProV développé au CEA

Méthodologie de création de la base d'apprentissage de façon dynamique et automatisée



Résultats obtenus sur la série des verres simples SBN

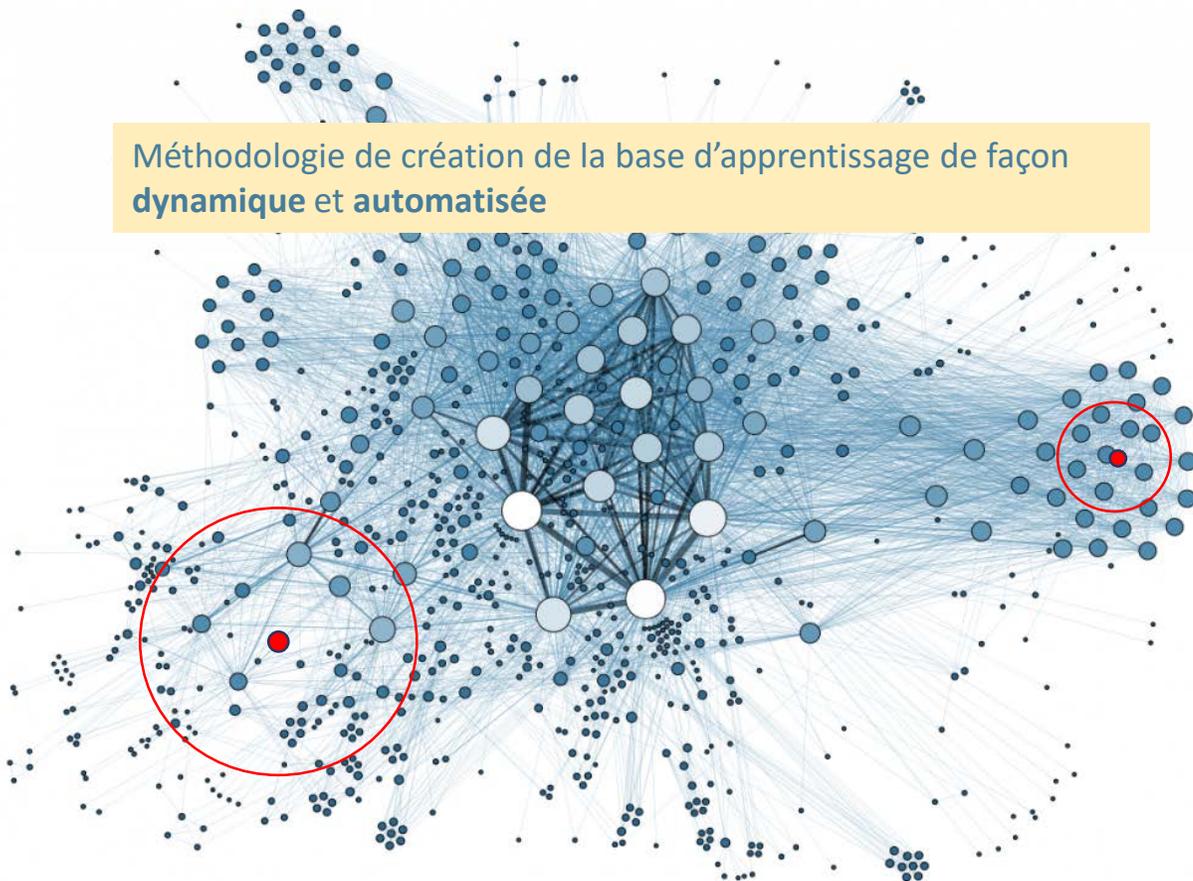


L'approche Base de données, Machine Learning

□ Cas d'application : modélisation prédictive de la viscosité

Outil prédictif MProV développé au CEA

Méthodologie de création de la base d'apprentissage de façon dynamique et automatisée



Résultats obtenus sur 230 verres test

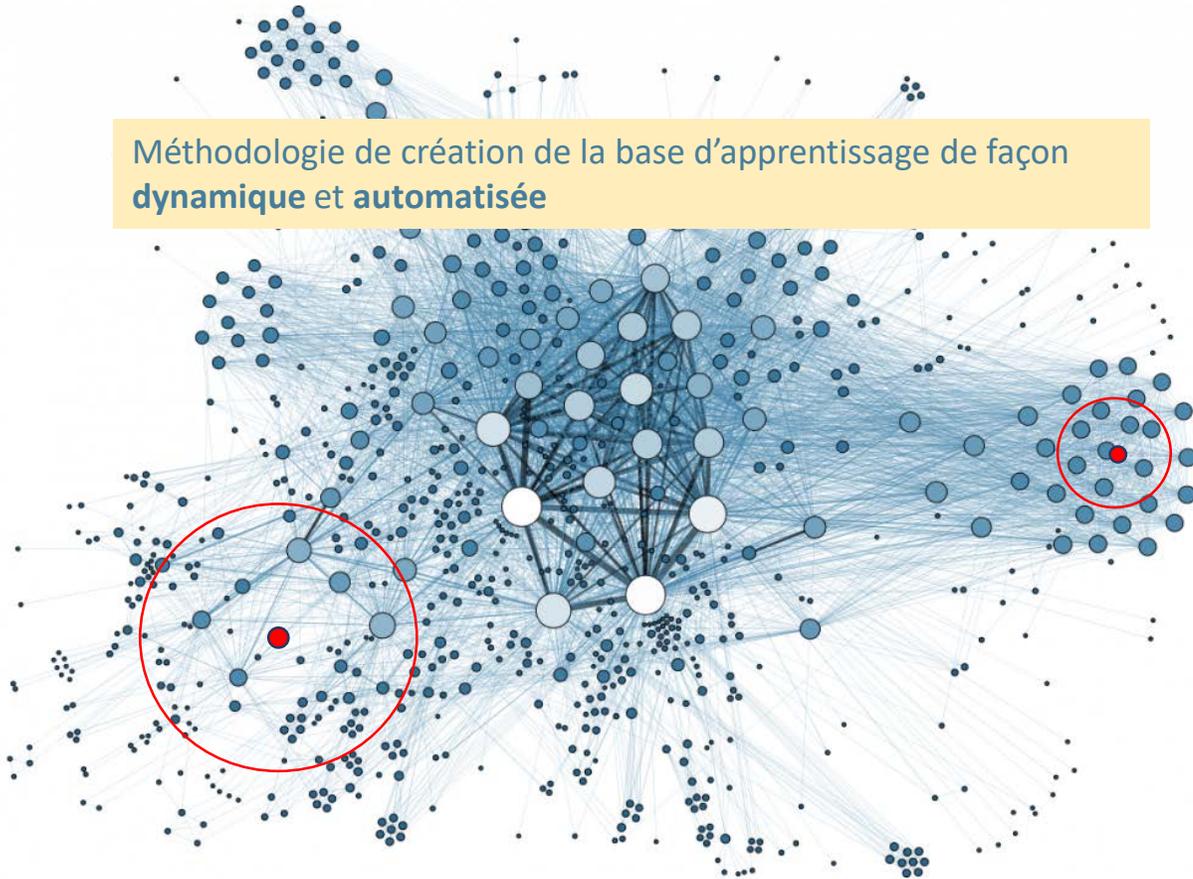
Viscosity prediction relative error	Borosilicate glass for nuclear waste	Sodo alumino silica glass	Overall
	N=73	N=55	N=230
Quantile 50% (median)	11%	18%	17%
Quantile 75%	19%	35%	34%
Quantile 90%	37%	73%	77%

L'approche Base de données, Machine Learning

□ Cas d'application : modélisation prédictive de la viscosité

Outil prédictif MProV développé au CEA

Méthodologie de création de la base d'apprentissage de façon dynamique et automatisée



Résultats obtenus sur 100 verres test
(prédiction de Tg)

Tg prediction error	Borosilicate glass		Sodo alumino silica glass		Overall	
	N=80		N=20		N=100	
	Rel.	Abs.	Rel.	Abs.	Rel.	Abs.
Quantile 50% (median)	1,4%	7°C	1,7%	10°C	1,5%	7°C
Quantile 75%	2,8%	13°C	3,2%	17°C	2,9%	14°C
Quantile 90%	5,1%	29°C	4,2%	19°C	4,7%	26°C



Merci pour votre attention

