Statistical methods and data-driven models to predict glass properties

The case of glass melt viscosity

Damien Perret, Alexandra Garcin, Adil Zomahoun, Caroline Chabal

CEA Marcoule

DES/ISEC/DPME

Traditional approach for glass property modeling

- First attempt for the calculation of glass properties from their composition proposed by Winckelmann and Schott at the end of the 19th century

- Theoretical Principle of Additivity

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

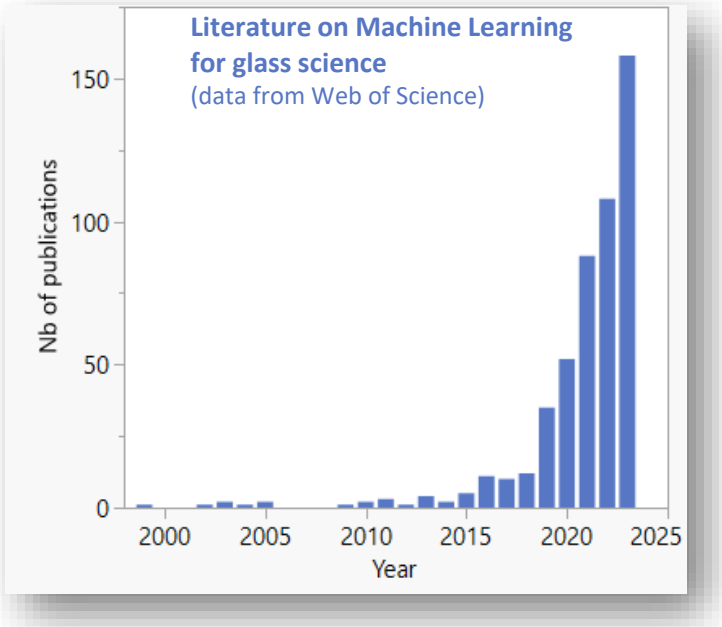
G is the property of the glass
$g(G)_i$ is the additive factor for oxide i and property G
x_i is the amount of oxide i

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

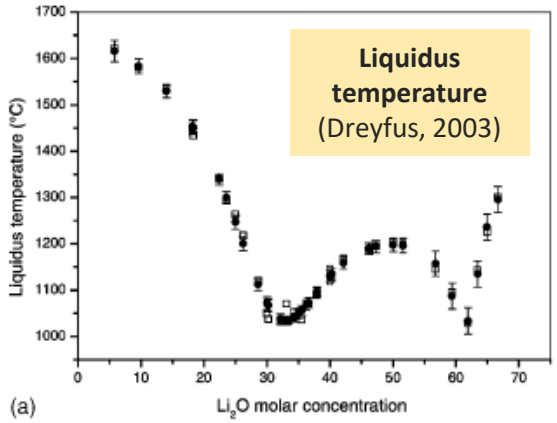
  - Generally valid when investigating suitably narrow composition range
  - Errors in additive calculation could be due to phase separation, crystallization, degree of cross-linking, anomalies in the cross-linked structure, interaction between ions,…

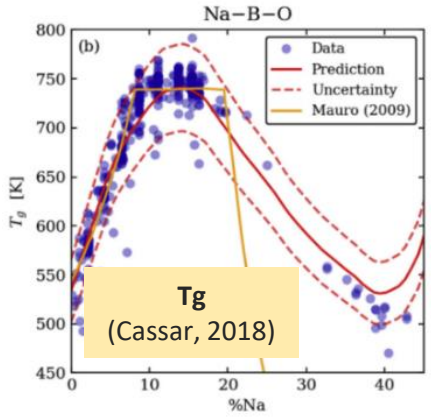
Machine Learning for glass property modeling

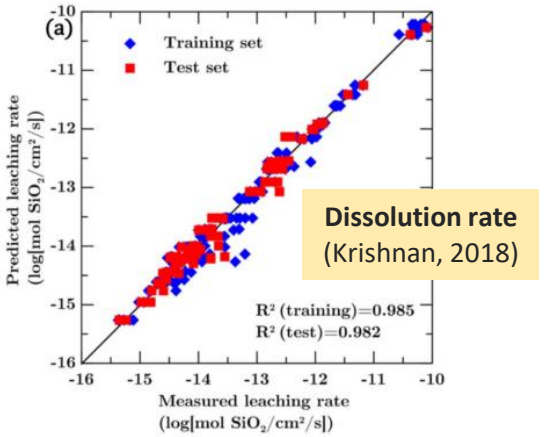
- Some examples of ML use in glass science

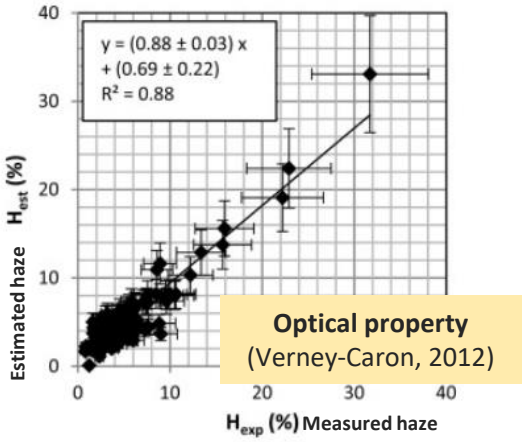


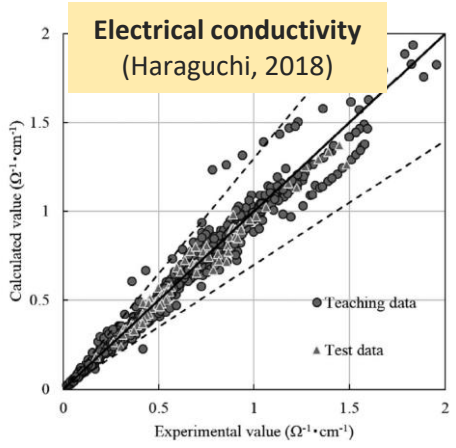
2019 : SciGlass in open access (GlassPy)

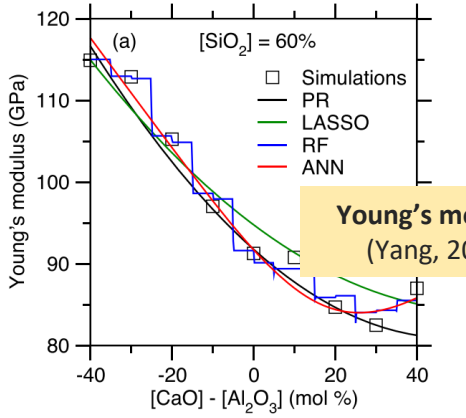








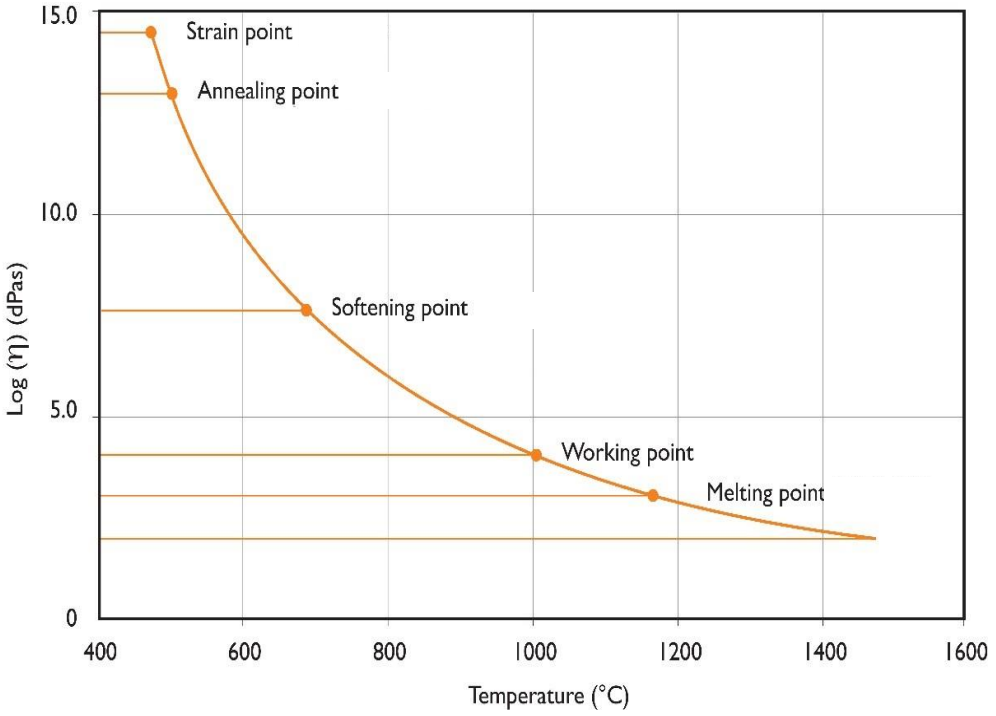




The viscosity case

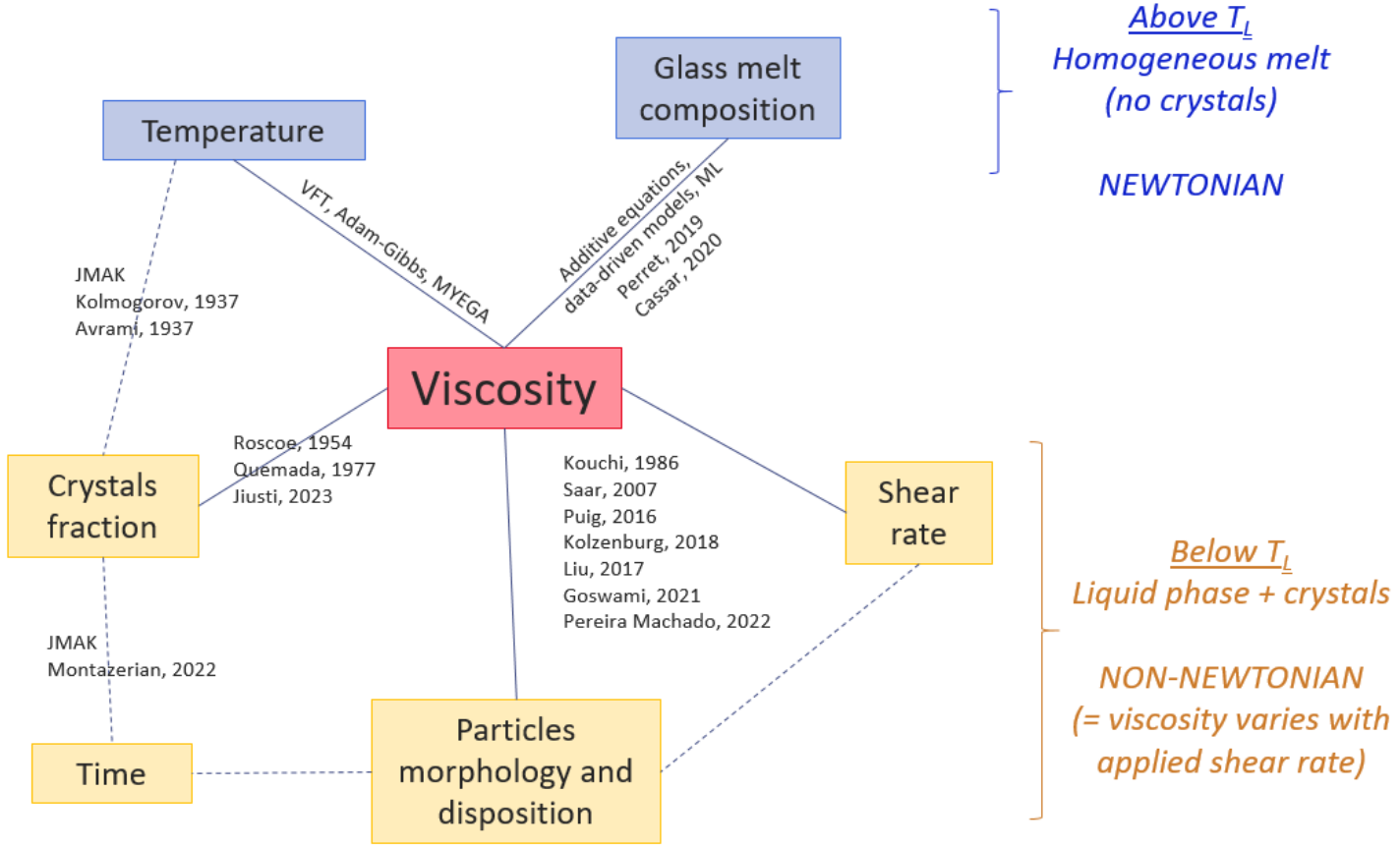
Why is glass melt viscosity prediction so challenging?

→ Range of viscosity values is very wide vs temperature and vs composition (~ 13 orders of magnitude)



→ Chemical dependence of viscosity is complex

→ Viscosity temperature dependence is highly sensitive to phase separation and crystallization

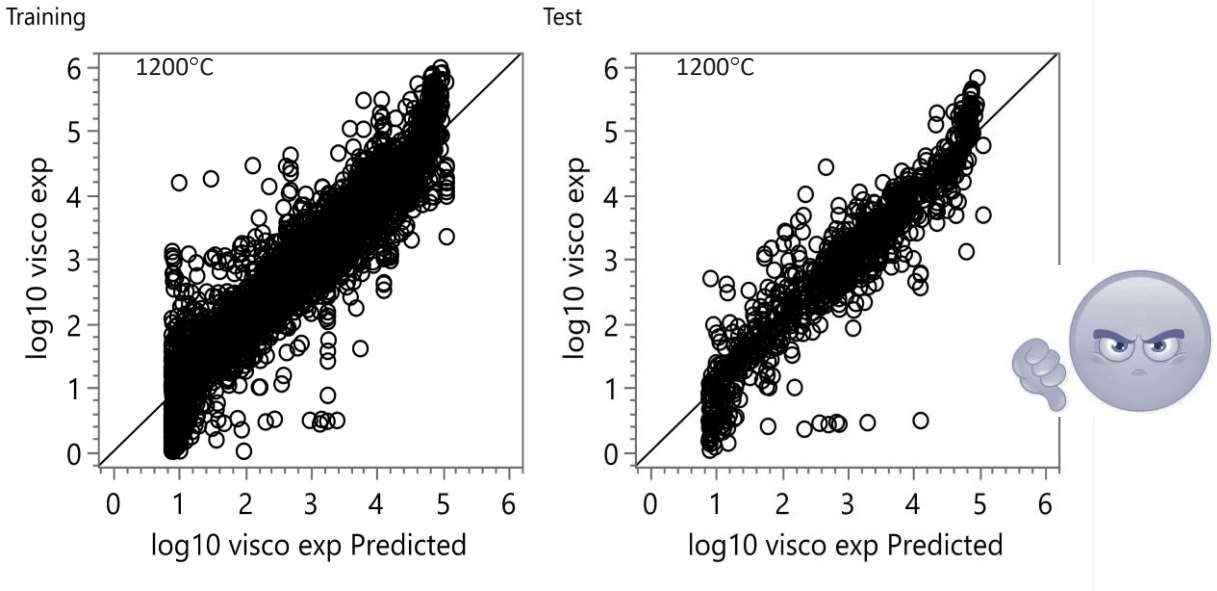


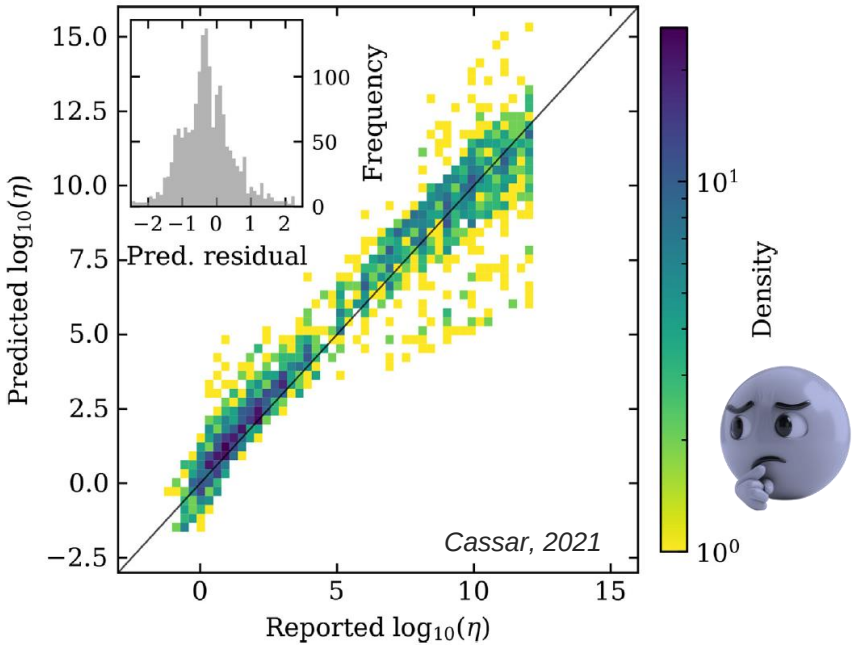
The viscosity case

Why is glass melt viscosity prediction so challenging?

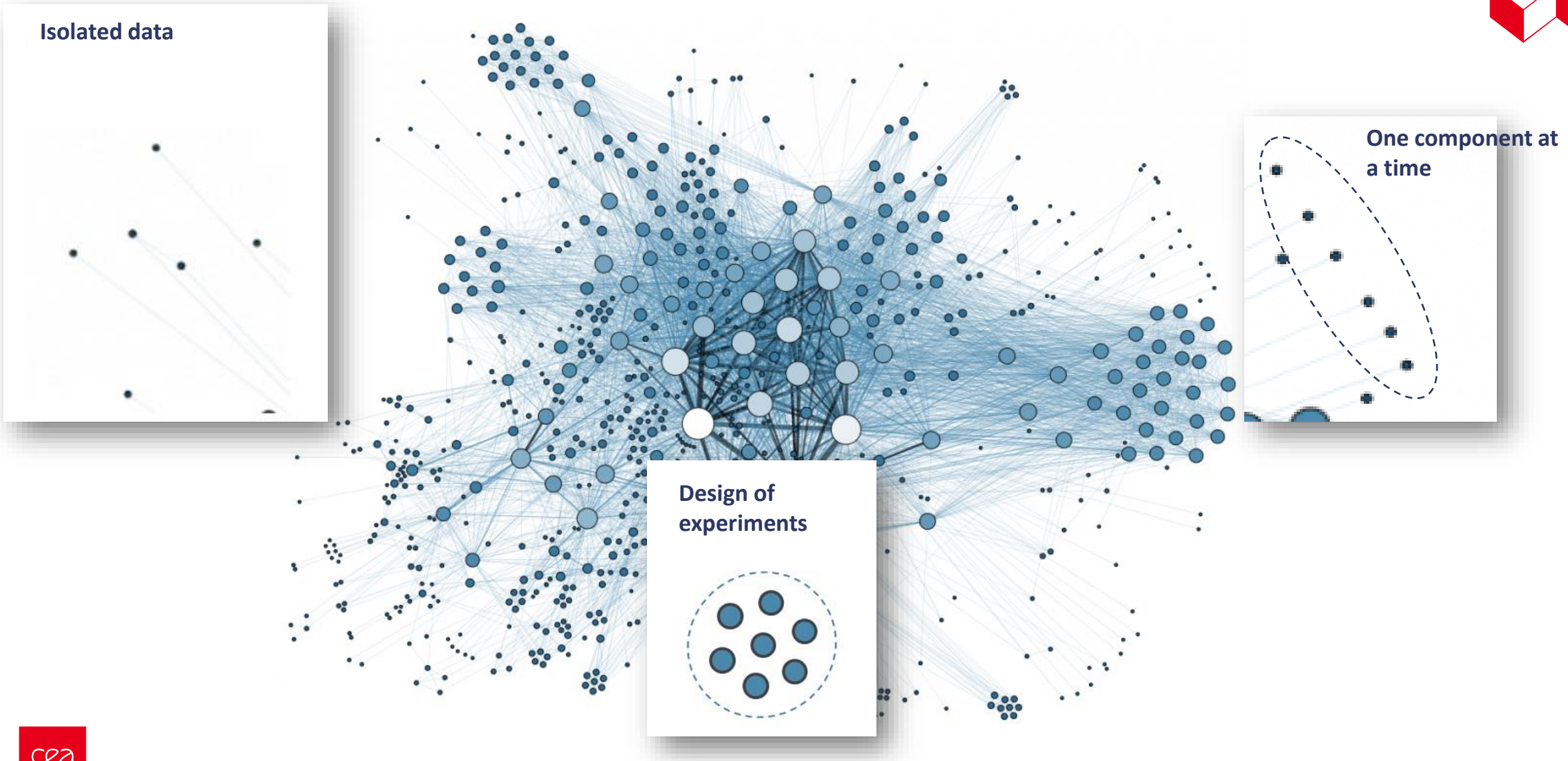
- Recent ML models applied to the global glass database can be up to 6 orders of magnitude far from the reported values

Recent attempts to predict glass melt viscosity with Neural Nets…



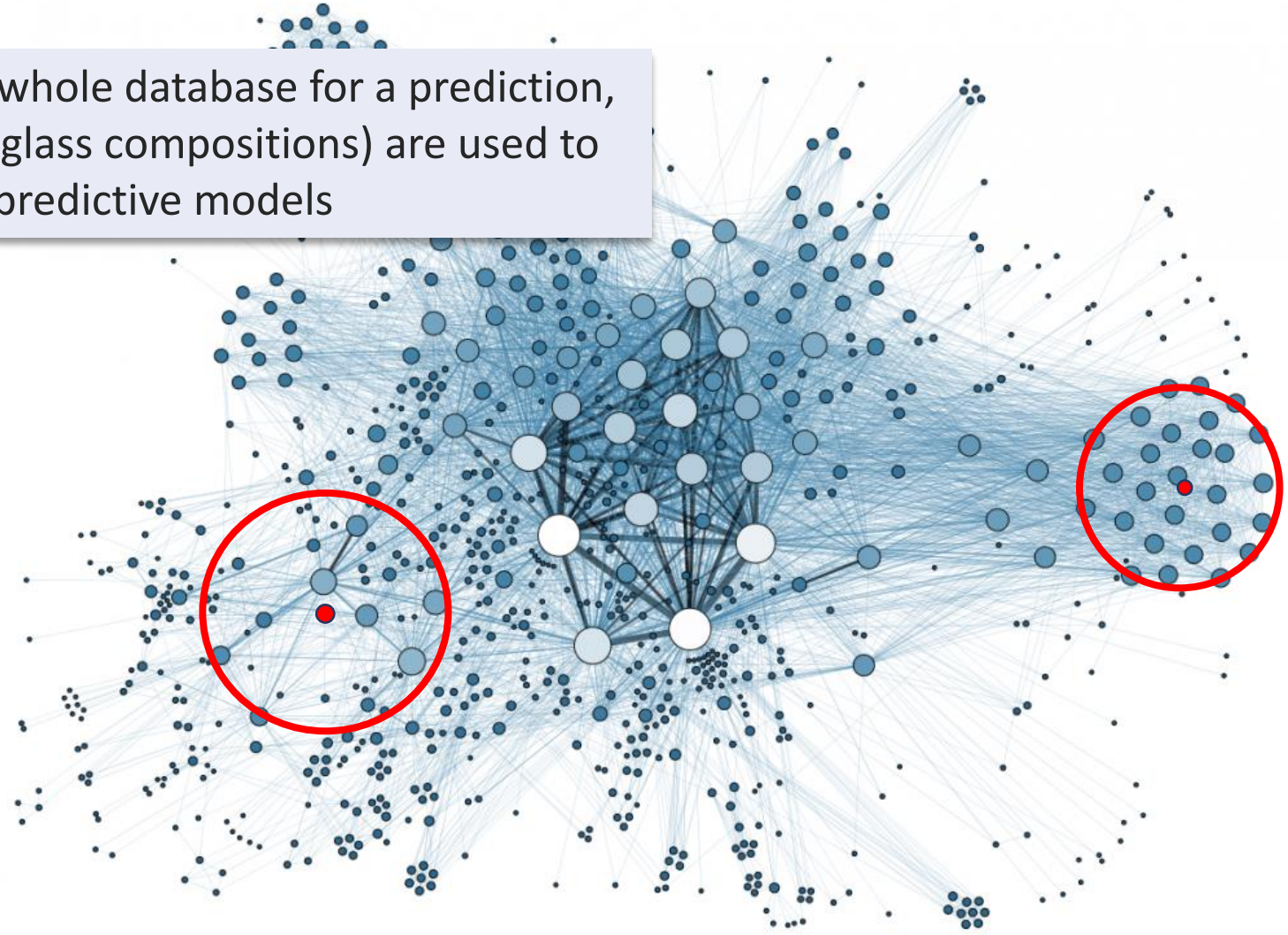


Unevenly populated database

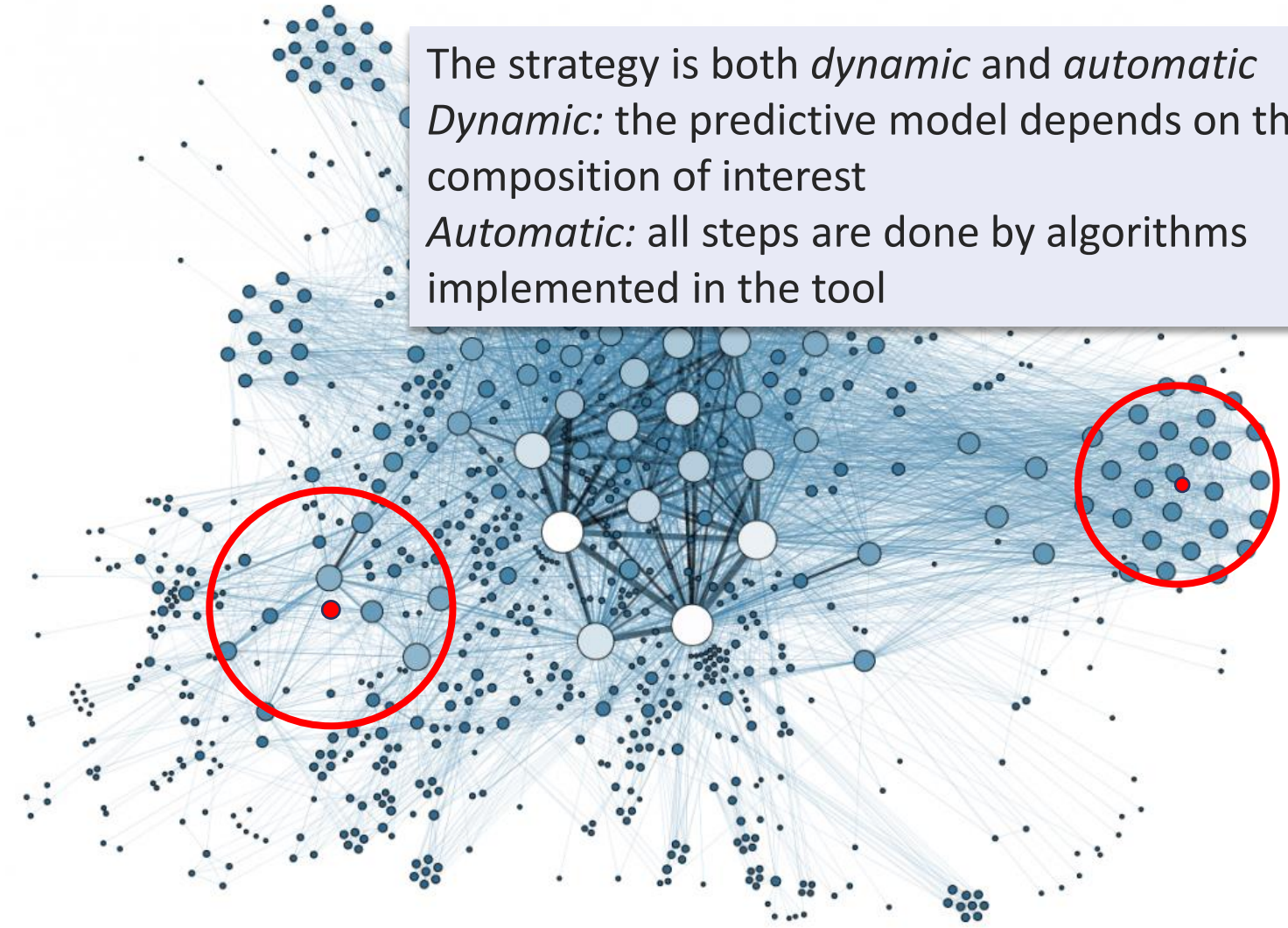


Smart subsampling strategy

Instead of using the whole database for a prediction, only *similar* data (= glass compositions) are used to train predictive models



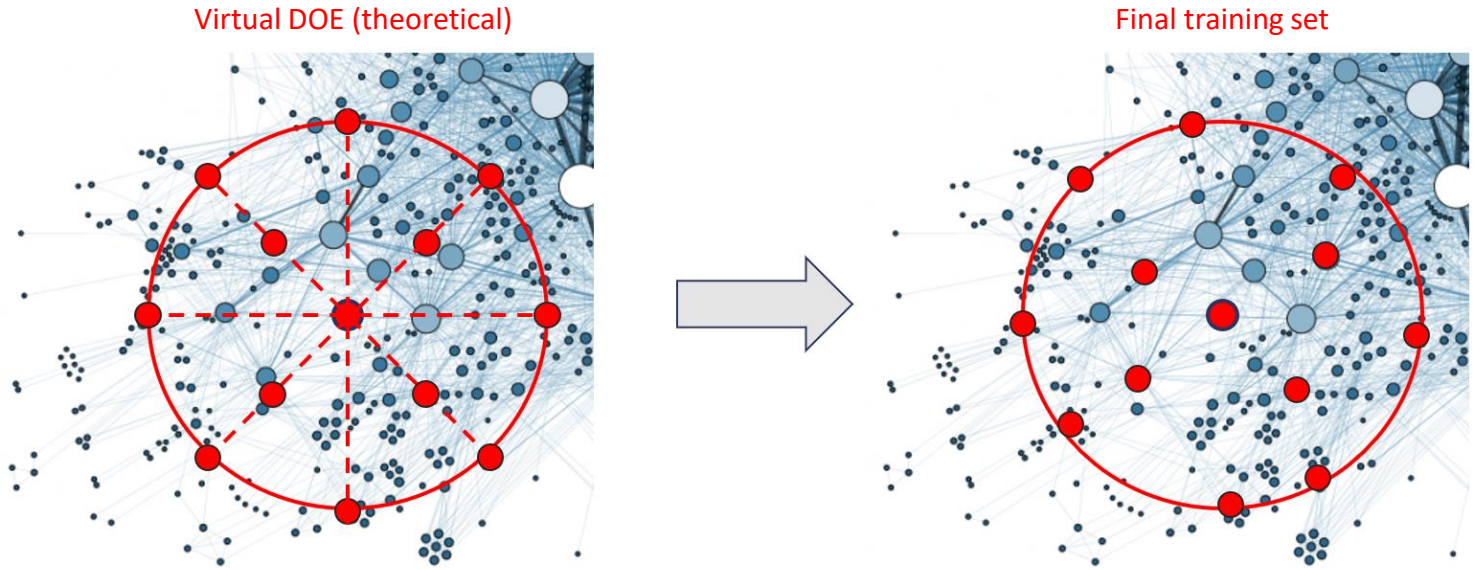
Smart subsampling strategy



The strategy is both *dynamic* and *automatic*
Dynamic: the predictive model depends on the composition of interest
Automatic: all steps are done by algorithms implemented in the tool

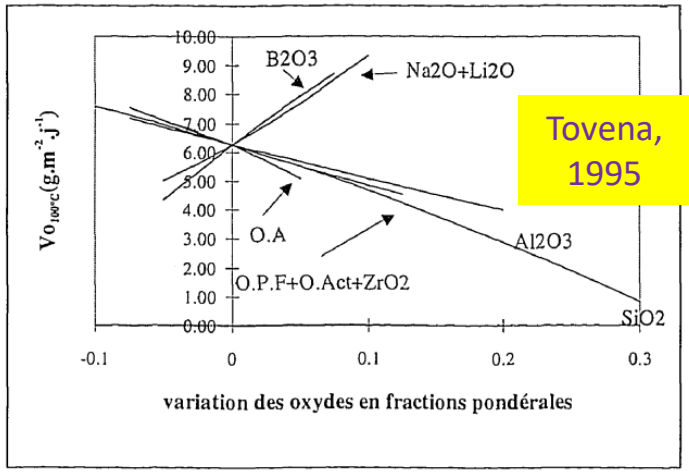
Smart subsampling strategy

- Build a virtual optimal DOE (Mixture Design) around the composition of interest
- Replace DOE runs by similar experimental data present in the database



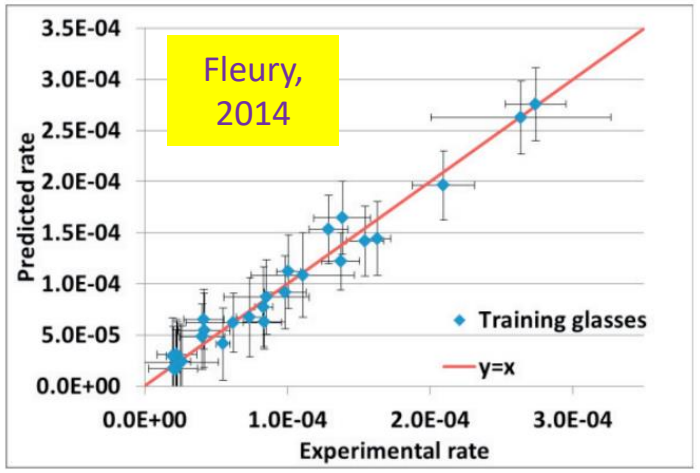
Examples of DOE applied to glass property prediction at CEA

Initial dissolution rate (R7T7 glasses)



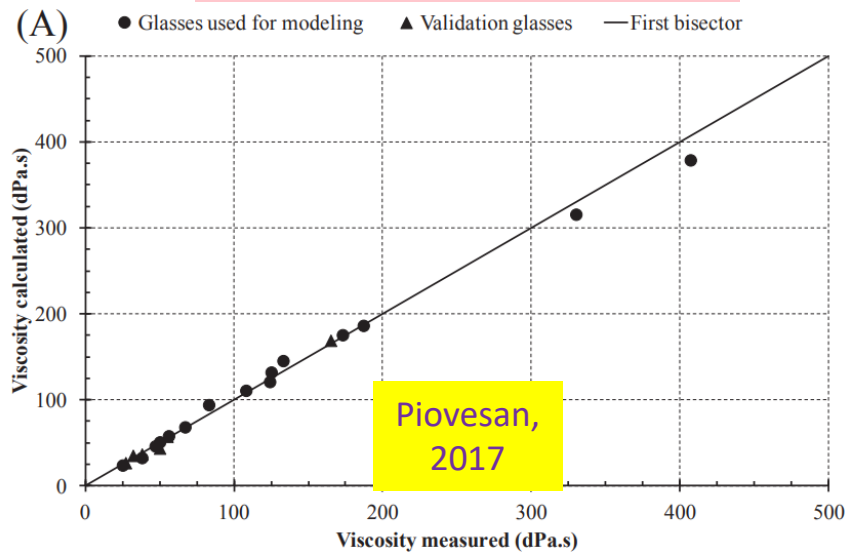
Tovena, 1995

Residual dissolution rate (R7T7 glasses)



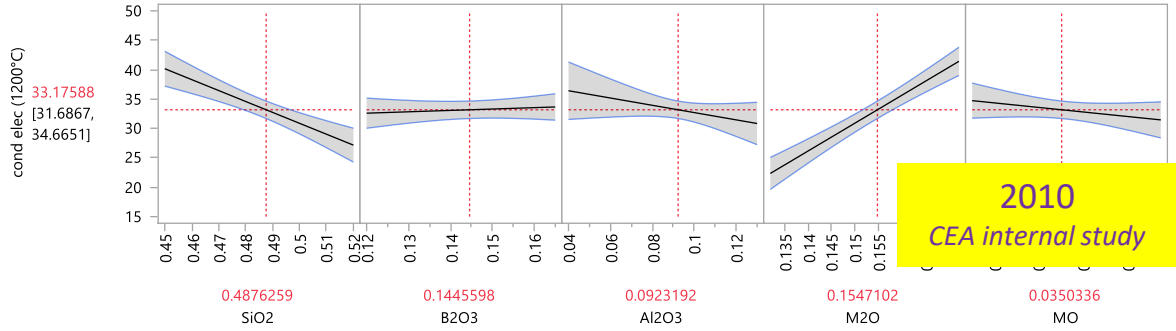
Fleury, 2014

Viscosity (peraluminous glasses)

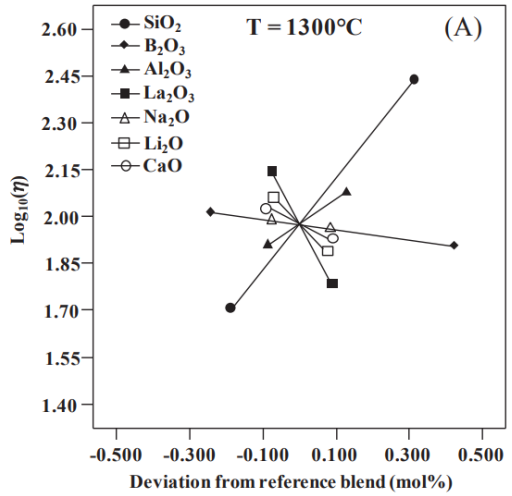


Piovesan, 2017

Electrical conductivity (LAW glasses)



2010
CEA internal study



Predictive modeling

- Four predictive algorithms implemented
  - Polynomial model (BIC-F Partial Quadratic Mixture model)
  - Generalized Linear model (Generalized Regression in JMP)
  - Neural Nets
  - SVR (Support Vector Regression)

  → **Four predicted values** for the response (glass viscosity)

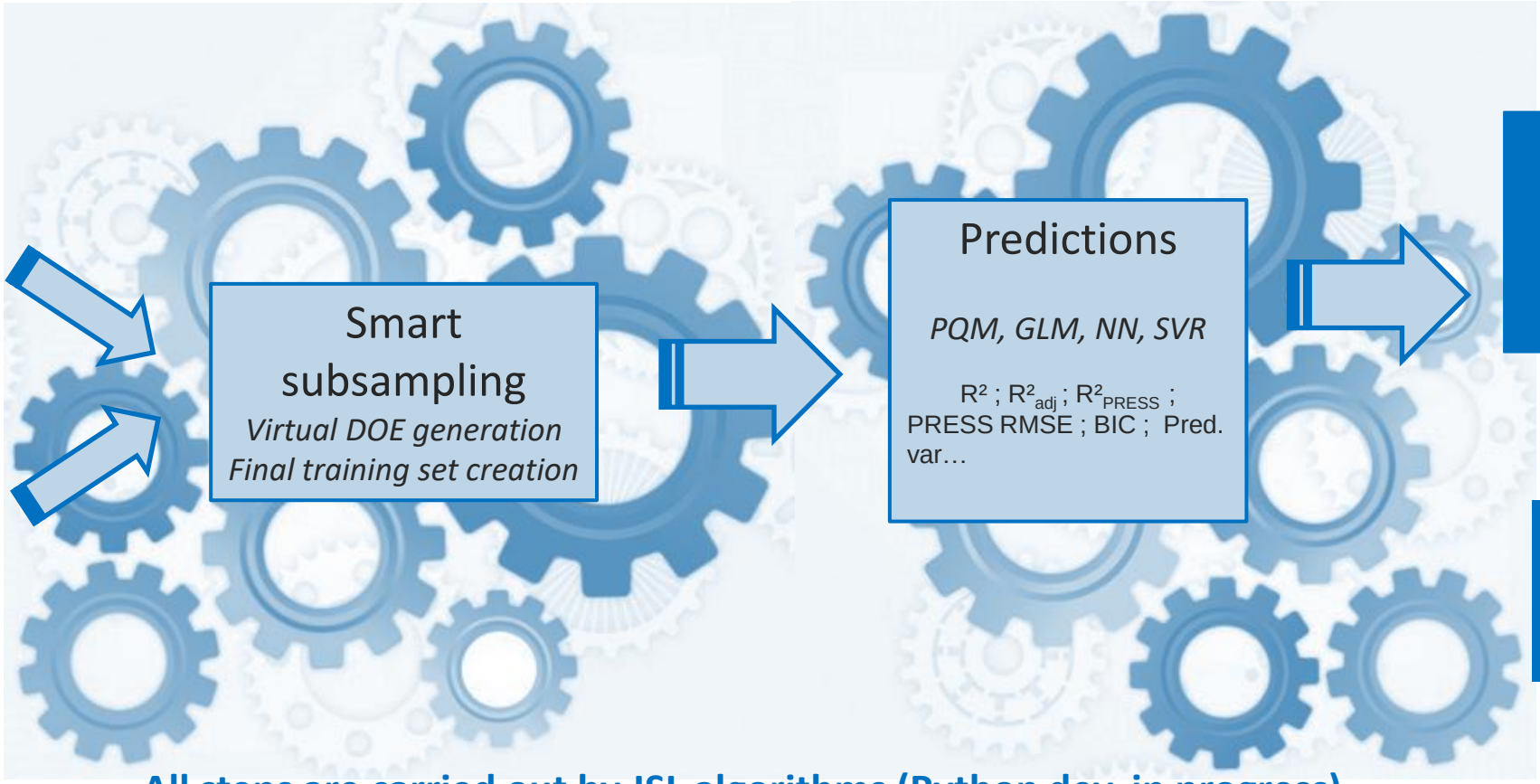
- Statistical criteria (R^2 ; R^2_{adj} ; R^2_{PRESS} ; PRESS RMSE ; BIC ; variance of prediction…) computed and analyzed to compare algorithm predictions

Key parameters

- Inputs from glass science experts
  - Nature and roles of glass oxides on viscosity
  - Origin and reliability of the data (year, experimental set up,…)

- Weights and type of distance calculation between *similar* glasses
- DOE size (number of runs, upper and lower boundaries)
- All these inputs were implemented in the tool algorithms

Schematic diagram of the tool

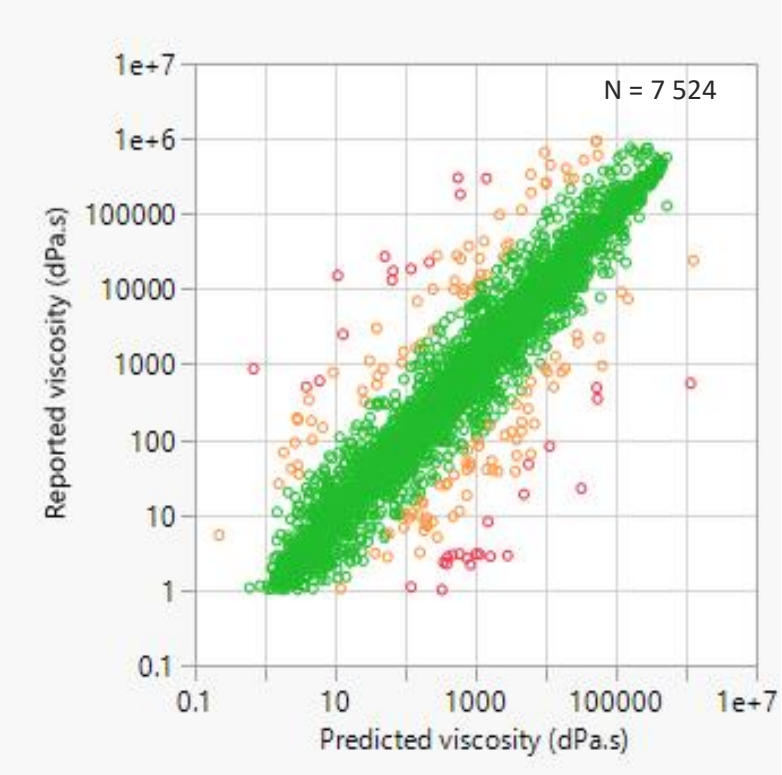


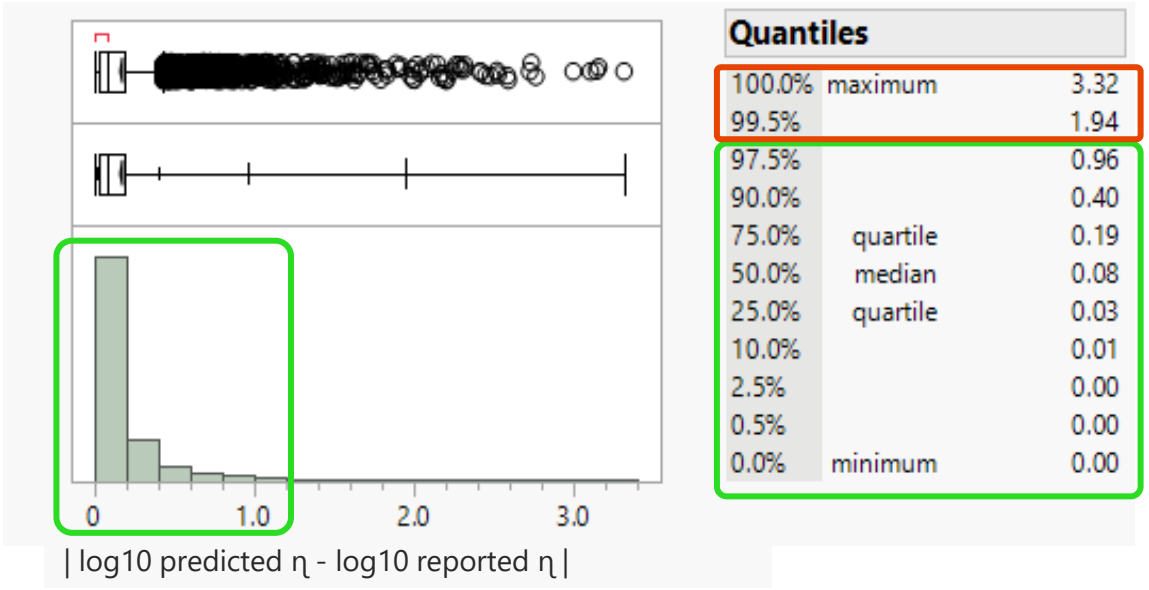
All steps are carried out by JSL algorithms (Python dev. in progress)

Prediction capability evaluation

- Prediction accuracy evaluated on a test set of 7500 data
- Metric of interest is the absolute error **| log10 predicted η - log10 reported η |**
- This indicator reflects how far the predicted viscosity is from the reported value in terms of order of magnitude
- In glass material science, an accuracy of one order of magnitude in viscosity prediction, within the range of 1 dPa.s to 10^6 dPa.s, is considered highly effective

Prediction capability evaluation





Quantiles		
100.0%	maximum	3.32
99.5%		1.94
97.5%		0.96
90.0%		0.40
75.0%	quartile	0.19
50.0%	median	0.08
25.0%	quartile	0.03
10.0%		0.01
2.5%		0.00
0.5%		0.00
0.0%	minimum	0.00

- Predicted viscosity is obtained from the median of the predictions given by the 4 algorithms
- For 97.5% of the data, error on predicted viscosity is less that one order of magnitude
- For only 0,5% of the data (= 38 compositions) error is between 2 and 3.3 orders of magnitude

Prediction capability evaluation

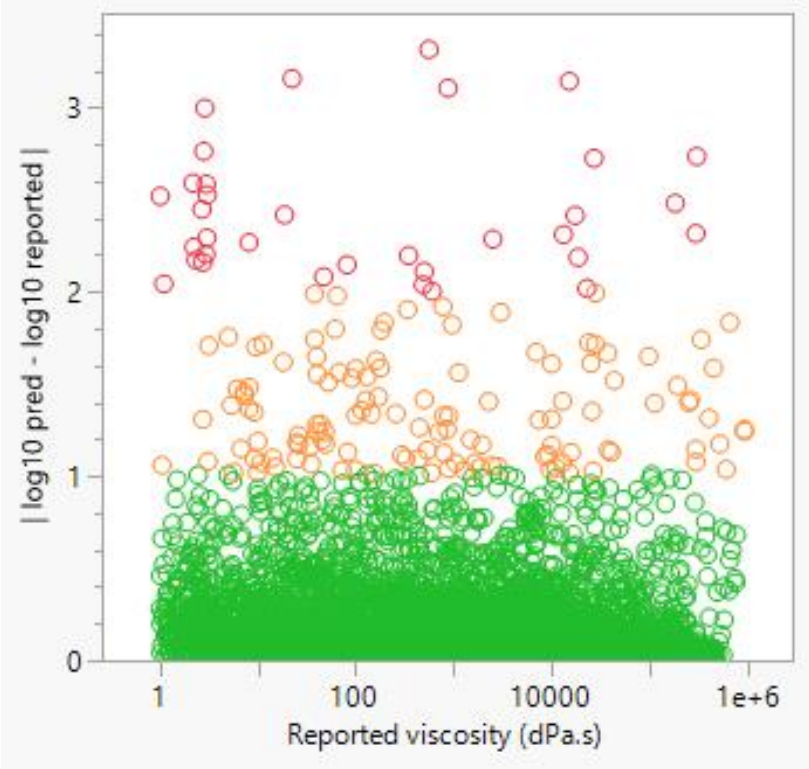
- An analysis of the largest deviations is currently in progress
- It appears that for the least accurately predicted compositions:
  - The data all come from SciGlass (publications or patents) and contain no borosilicate compositions
  - Some viscosity measurements are quite old (1960, 1985)
  - The content of certain oxides reaches unusual values

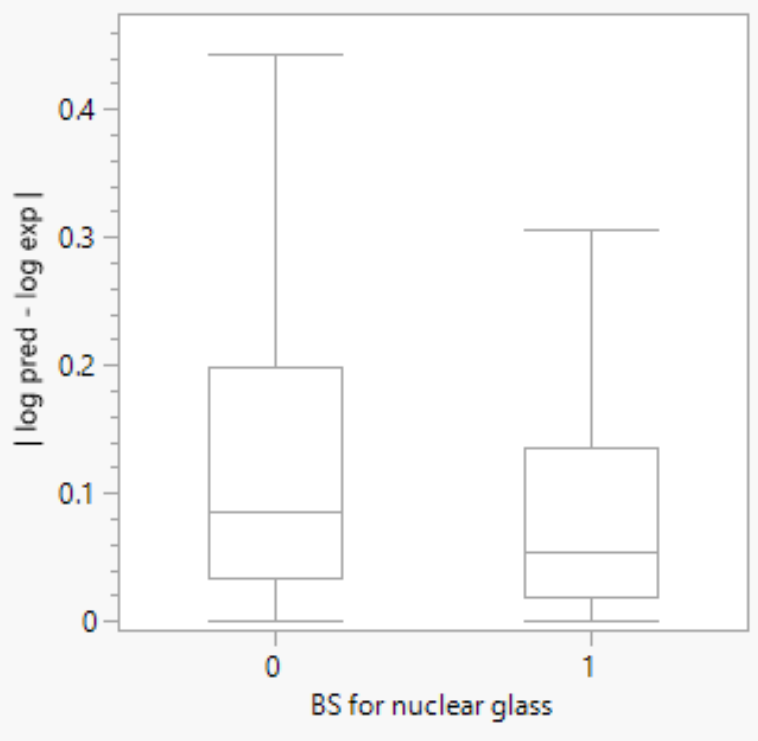
Origin	Year	Meas Visco (dPa.s)	Pred Visco (dPa.s)	Al2O3	As2O3	BaO	CaO	CeO2	Fe2O3	K2O	Li2O	MgO	Na2O	P2O5	SiO2	SO3	TiO2	ZrO2
SCIGLASS	1960	1	413	0.0	0.0	40.0	10.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	50.0	0.0	0.0	0.0
SCIGLASS	2001	3	1105	5.0	0.6	20.0	0.0	0.0	0.0	8.0	0.0	0.0	7.4	0.0	59.0	0.0	0.0	0.0
SCIGLASS	2002	876	2	0.5	0.0	0.0	6.6	0.0	0.0	0.3	0.0	0.6	55.8	0.0	36.0	0.1	0.0	0.0
SCIGLASS	2010	2	1230	9.5	0.0	0.0	0.0	0.4	0.0	2.0	5.0	13.0	4.0	1.0	58.1	0.0	2.0	2.0
SCIGLASS	1985	15137	16	6.5	0.0	0.0	41.8	0.0	16.3	0.0	0.0	1.3	3.9	0.0	29.4	0.0	0.7	0.0

Glass compositions with the poorest prediction accuracy

Prediction capability evaluation

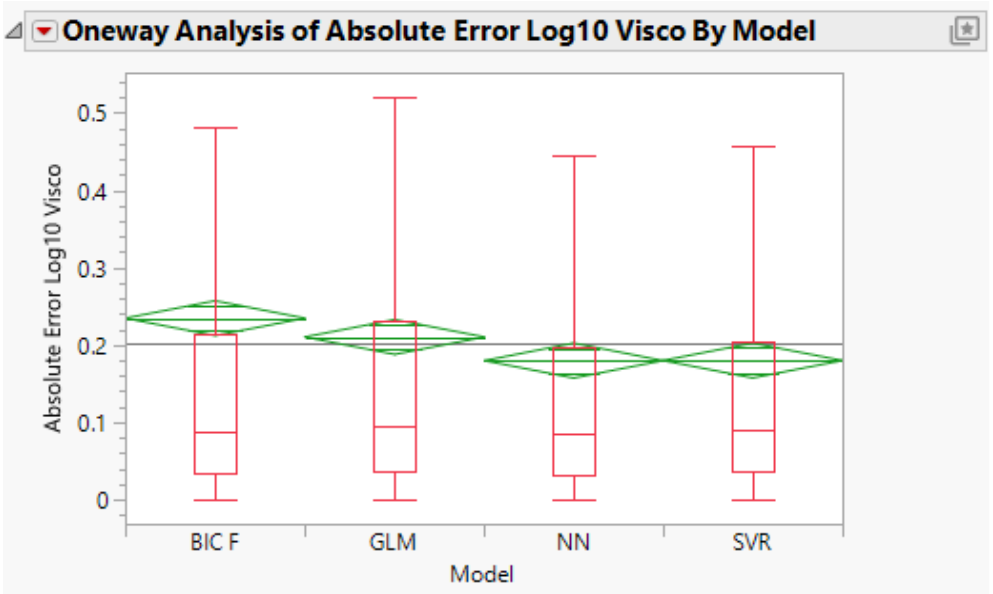
- Good prediction accuracy for both low (1 dPa.s) and high (10^6 dPa.s) viscosity range of values
- Accuracy is even better for glass compositions of interest





Prediction capability evaluation

- MLR, GLM, NN and SVR predictive abilities are very close
- SVR provided smallest max of absolute error
- **Keeping the median of the four algorithms helps ensure highly robust predictions**



Quantiles

Level	Minimum	10%	25%	Median	75%	90%	Maximum
BIC F	6.45e-13	0.012706	0.034561	0.088469	0.213602	0.446479	122.473
GLM	9.469e-6	0.014012	0.037075	0.094388	0.230209	0.469173	54.47181
NN	1.33e-15	0.012338	0.032187	0.0852	0.197774	0.416534	9.812305
SVR	1.732e-5	0.014212	0.035922	0.08937	0.204689	0.427189	3.37347

Conclusion

- Predictive approach based on a dynamic subsampling of the data
- 4 models (MLR, GLM, NN and SVR) are trained on a subset of similar data, generated by an optimal DOE
- Each procedure is fully automatic. All steps are carried out by JSL algorithms (Python code dev in progress)
- Predictive ability evaluated on 7,500 glass compositions
- High prediction accuracy, especially for glass compositions of interest

Future work

- Evaluate and implement Boosted Tree and XGBoost models using the full population
- Complete and test the Python code

Thank you for your attention