Supplementary materials to

GEOMIND: A hybrid generative artificial intelligence for geopolymer design and optimization

Sébastien Rousseau[1,2], Assil Bouzid[1], Sylvie Rossignol[1] and Ameni Gharzouni*[1]

[1] Institut de Recherche sur les Céramiques, UMR 7315 CNRS - Université de Limoges, Centre Européen de la Céramique, 12 rue Atlantis, 87068, Limoges Cedex, France.

[2] XLIM, UMR CNRS 7252, 16 Rue Jules Vallès, 19100 Brive La Gaillarde, France

*Corresponding author: ameni.gharzouni@unilim.fr

Table S1. Simulator hyperparameters

Hidden Layers	**Units**	**Activation**	
Dense	128	Parametric Rectified Linear Unit (PReLU)	
Dense	128	PReLU	
Dense	64	PReLU	
Dense	64	PReLU	
Dense (mean)	32 (latent dimension)	Linear	
Dense (log variance)	32 (latent dimension)	Linear	
Dense	64	PReLU	
Dense	64	PReLU	
Dense	128	Rectified Linear Unit (ReLU)	
Dense	128	ReLU	
Output Layers			**Loss function**
Dense (viscosity)	3	Linear	MAE
Dense (mixture density)	1	Linear	MAE
Dense (compressive strength)	1	Linear	MAE
Dense (material density)	1	Linear	MAE

Optimizer	**Learning rate**	**Epochs**	**Batch size**
RMSprop	0.0005	250	16

Table S2. Formulator hyperparameters

Hidden Layers	Units	Activation	
Dense	128	PReLU	
Dense	128	PReLU	
Dense	64	PReLU	
Dense	64	PReLU	
Dense (mean)	32 (latent dimension)	Linear	
Dense (log variance)	32 (latent dimension)	Linear	
Dense	64	PReLU	
Dense	64	PReLU	
Dense	128	ReLU)	
Dense	128	ReLU	
Output Layers			**Loss function**
Dense (precursors)	11	Sigmoid	MAE
Lambda (sums)	1		MAE
Lambda (Si/M)	1		MAE
Lambda (Si/Al)	1		MAE
Lambda (solid/liquid)	1		MAE

Optimizer	Learning rate	Epochs	Batch size
RMSprop	0.0005	310	16

Table S3. Final loss values of Formulator model training methods

Formulator model loss	Training	Validation
MAE with GEOMIND	0.092	0.094
MAE without GEOMIND	0.118	0.123

Table S4. Final loss values of different models of Formulator in GEOMIND framework

Model	Training MAE on precursors	Validation MAE on precursors	Training MAE on properties	Validation MAE on properties
MLP	0.087	0.096	0.275	0.261
Gaussian VAE	0.092	0.094	0.282	0.249
Student's t VAE	0.091	0.098	0.284	0.291

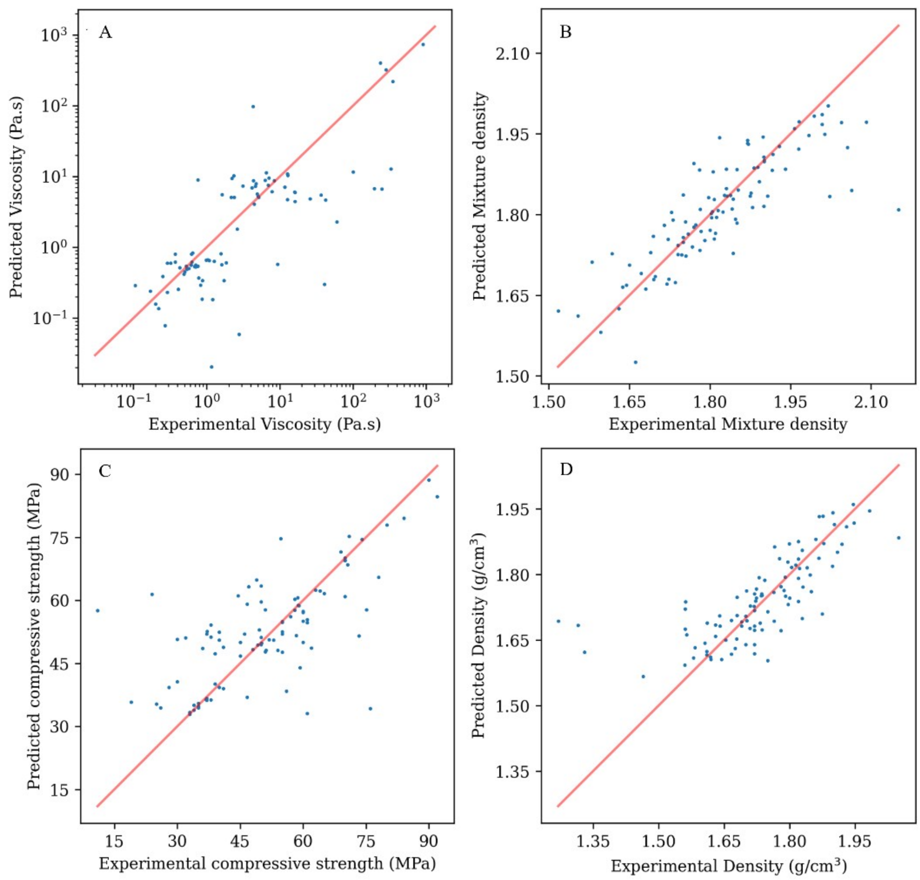

Figure S1. Predicted vs. experimental results for 98 samples

Table S5. Target, predicted and experimental properties of the fifteen test samples

Sample	Properties	Initial viscosity (Pa.s)	Mixture density (g/cm³)	Compressive strength (MPa)	Material density (g/cm³)
1	target	1.62	1.56	82	1.60
	predicted	0.75	1.70	51	1.76
	experimental	1.44	1.79	57	1.80
2	target	79.62	1.55	32	1.72
	predicted	4.95	1.75	54	1.77
	experimental	9.78	1.79	56	1.84
3	target	14.15	1.91	45	1.98
	predicted	7.18	1.77	58	1.83
	experimental	109.2	1.82	75	1.83
4	target	0.50	1.69	64	1.5
	predicted	0.54	1.78	46	1.69
	experimental	0.73	1.74	58	1.72
5	target	2.11	1.64	50	1.55
	predicted	5.42	1.71	44	1,66
	experimental	7.08	1.66	52	1,69
6	target	4.00	1.31	45	1.50
	predicted	9.05	1.73	48	1.66
	experimental	13.36	1.65	47	1.70
7	target	6.00	1.85	62	1.83
	predicted	8.55	1.78	53	1.77
	experimental	12.80	1.80	52	1.79
8	target	2.00	1.65	30	1.69
	predicted	0.29	1.68	47	1.61
	experimental	1.45	1.60	33	1.60
9	target	1.00	1.31	41	1.80
	predicted	0.34	1.73	32	1.60
	experimental	0.10	1.65	27	1.59
10	target	2.40	1.90	80	2.00
	predicted	8.28	1.88	53	1.79
	experimental	10.94	1.80	60	1.79
11	target	0.30	1.60	38	1.90
	predicted	0.41	1.71	35	1.61
	experimental	0.23	1.64	39	1.61
12	target	1.20	1.10	25	1.50
	predicted	0.07	1.54	29	1.60
	experimental	0.02	1.51	14	1.55
13	target	1.70	1.43	41	1.65
	predicted	0.27	1.66	34	1.60
	experimental	0.12	1.63	34	1.55
14	target	0.19	2.10	70	2.30
	predicted	7.76	1.86	56	1.76
	experimental	12.00	1.82	76	1.76
15	target	0.13	1.90	32	1.60
	predicted	0.58	1.74	45	1.67
	experimental	0.48	1.74	42	1.63

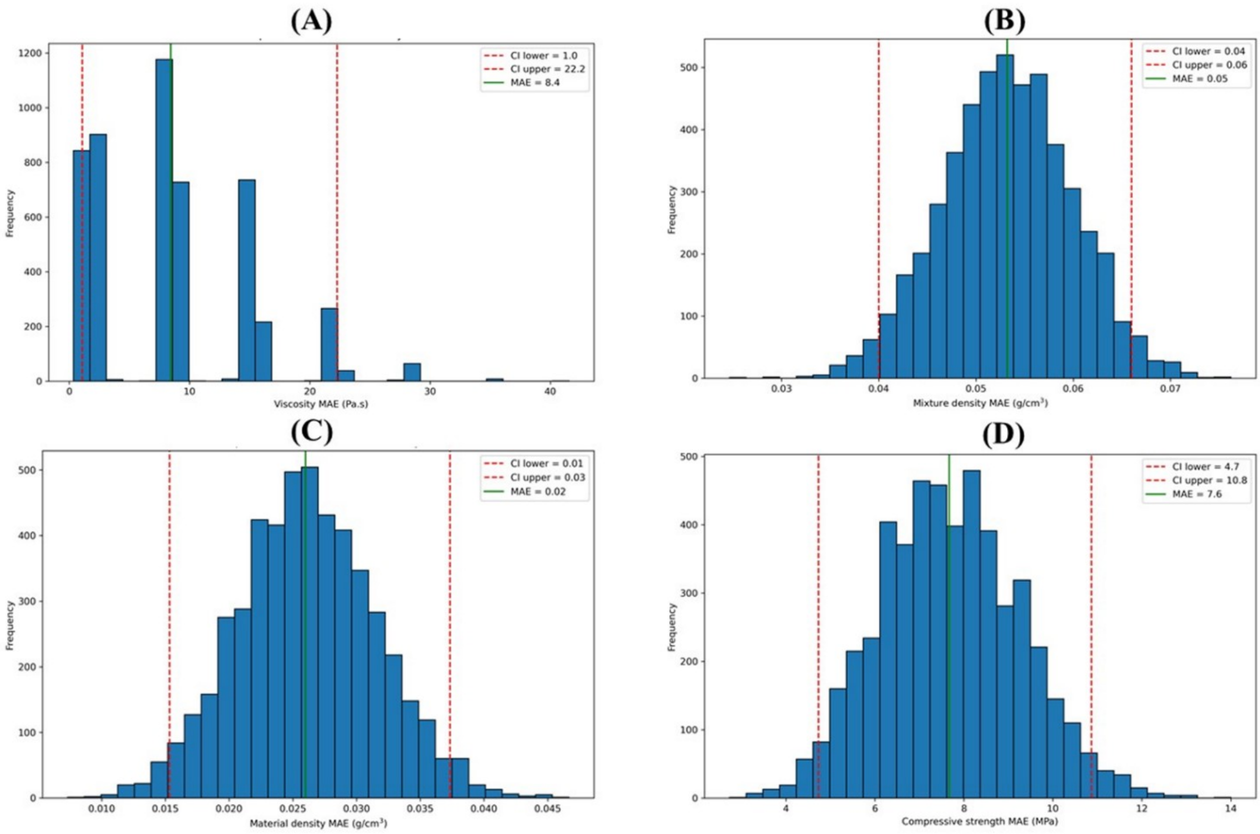

Figure S2. Bootstrap distribution of (A) initial viscosity, (B) mixture density, (C) compressive strength and (D) material density mean absolute errors. MAE (solid green line), CI lower (dashed red line) and upper (dashed red line) are also provided

Table S6. Target, predicted and experimental properties of the five feasible test samples proposed by GEOMIND without the feasibility controller block

Sample	**Properties**	**Initial viscosity (Pa.s)**	**Compressive strength (MPa)**	**Material density (g/cm³)**
1	target	1.20	70	1.55
	predicted	3.99	63	1.71
	experimental	40.86	50	1.82
2	target	1.30	67	1.37
	predicted	4.69	47	1.61
	experimental	1,85	50	1.75
3	target	79.00	51	1.20
	predicted	1.03	33	1.37
	experimental	2.78	39	1.56
4	target	0.50	85	1.60
	predicted	2.73	41	1.69
	experimental	0.99	53	1.82
5	target	0.80	75	1.30

	predicted	5.00	56	1.76
	experimental	1.03	30	1.72

Bayesian optimization method

An alternative way to achieve materials design through formulation prediction can be achieved by resorting to Bayesian optimization on top of the Simulator model in a way to search for the combination of precursor mixtures to achieve a given target property. As a grid-based search is inefficient and computationally expensive, the Bayesian offers a cost-efficient way to approach the solution in a straightforward manner. In practice, one uses the precursors mass ratios as the variables to be optimized (the formulation to be obtained) and the MAE between target and predicted properties by the simulator model as the cost to be minimized. The working principle of this scheme is illustrated in Figure S3. The main advantage of this method, in comparison with GEOMIND, is the use of a single VAE model, while the main disadvantage is that it cannot process a batch of target properties in a single run. As such, identification epochs are dedicated to a single set of target properties at once.

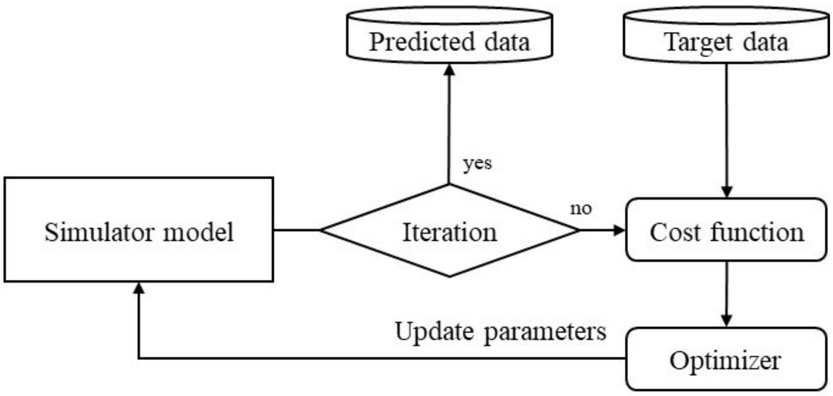

Figure S3. Bayesian optimization of material precursors

Compared to GEOMIND, the Bayesian Optimization method leads to larger errors on the final materials properties (Table S7). Three use-cases are represented on database properties heatmap in Figure S4. Considering sample 1, that obeys a user request of obtaining a final lightweight, viscous and resistant material, we find that while it is located in the bottom left part of the map, it's dark color that reflects a high compressive strength, does not fit with the color of the surroundings. This reflects a non-consistent user request. After prediction of the corresponding formulation, it is passed to the simulator to cross-check its actual predicted properties.

Table S7. MAE and nMAE between experimental and predicted with Bayesian optimization method

Property	Experimental vs. target properties		Experimental vs. predicted properties	
	MAE	nMAE (%)	MAE	nMAE (%)
Initial viscosity (Pa·s)	102	7.7	150	11.4
Mixture density(g/cm³)	0.26	39	0.112	17
Compressive strength (MPa)	18	22	5.7	7
Material density (g/cm³)	0.156	20	0.088	11

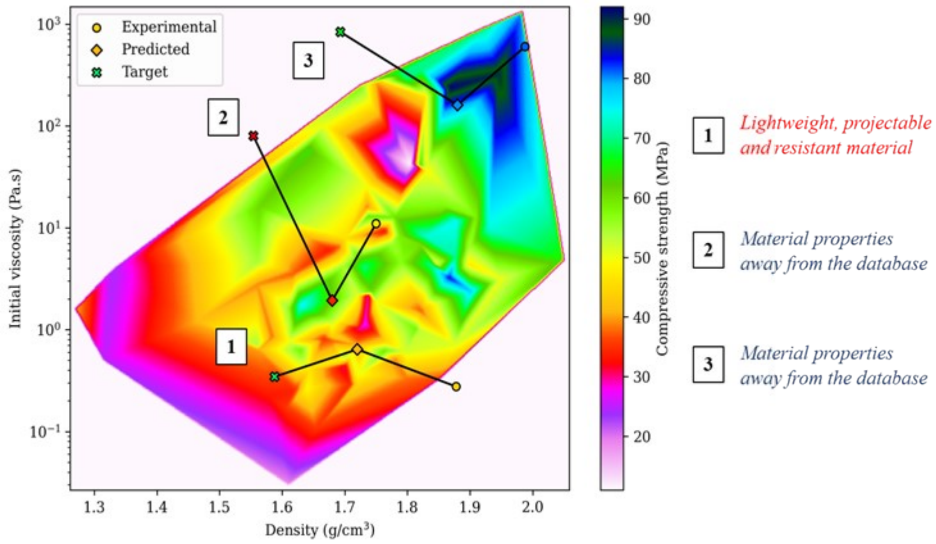

Figure S4. Three use-cases of Bayesian Optimization method samples represented on database properties heatmap