Electronic Supplementary Material (ESI) for Materials Advances.

Scalable Gaussian Processes for Predicting the Properties of Inorganic Glasses with Large Datasets

Suresh Bishnoi[1], R. Ravinder[1], Hariprasad Kodamana[2,], N. M. Anoop Krishnan[1,3,*]*

[1]Department of Civil Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India
[2]Department of Chemical Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India
[3]Department of Materials Science and Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India
**Corresponding authors: H. Kodamana (kodamana@iitd.ac.in), N. M. A. Krishnan (krishnan@iitd.ac.in)*

Supplementary Material

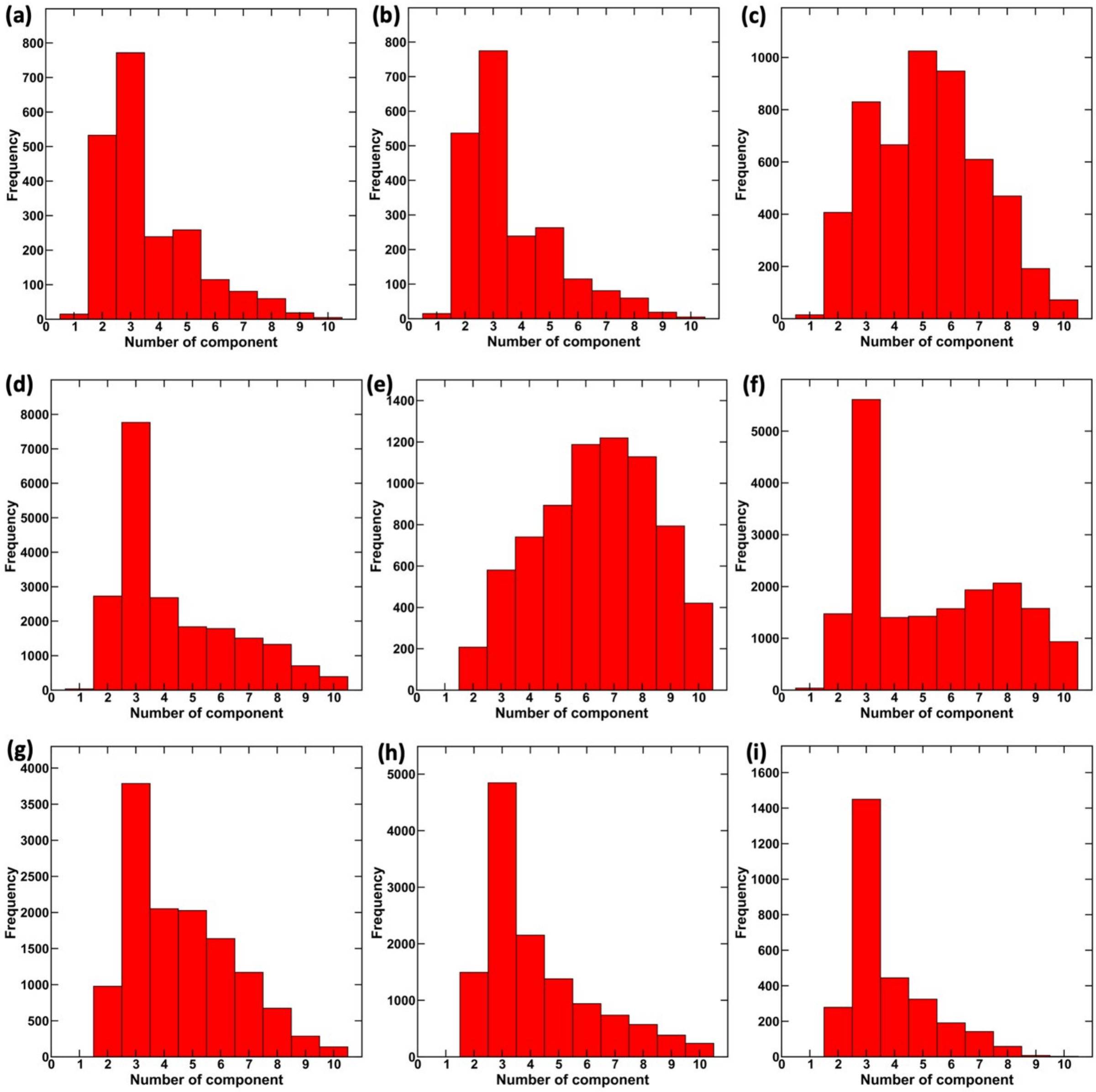
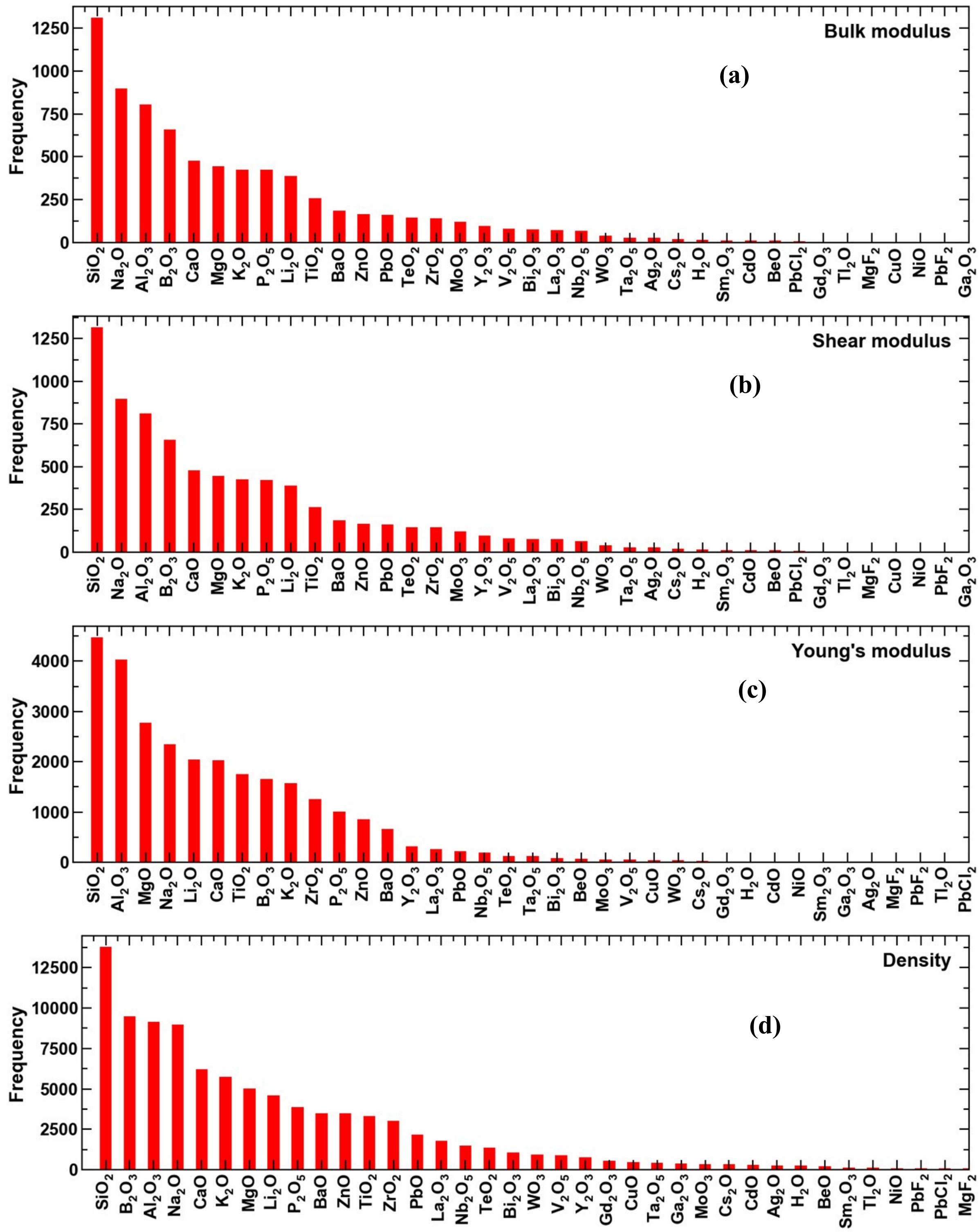
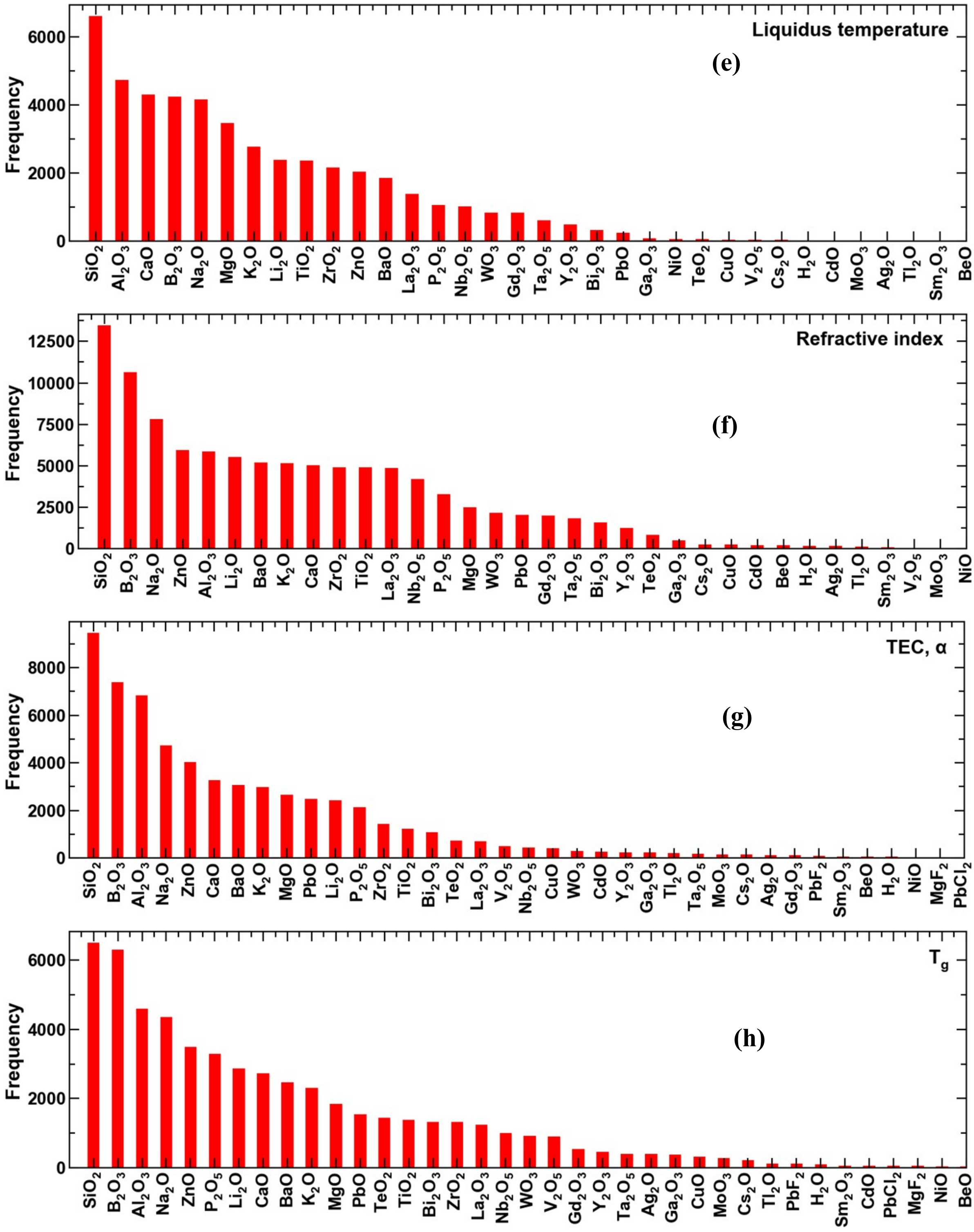


Figure S1. Number of component with respect to frequency plot for (a) bulk modulus, (b) shear modulus, (c) Young's modulus, (d) density, (e) liquidus temperature, (f) refractive index, (g) thermal expansion coefficient (TEC), (h) glass transition temperature (T_g) and (i) hardness.





Liquidus temperature
(e)
Refractive index
(f)
TEC, α
(g)
T_g
(h)
Frequency

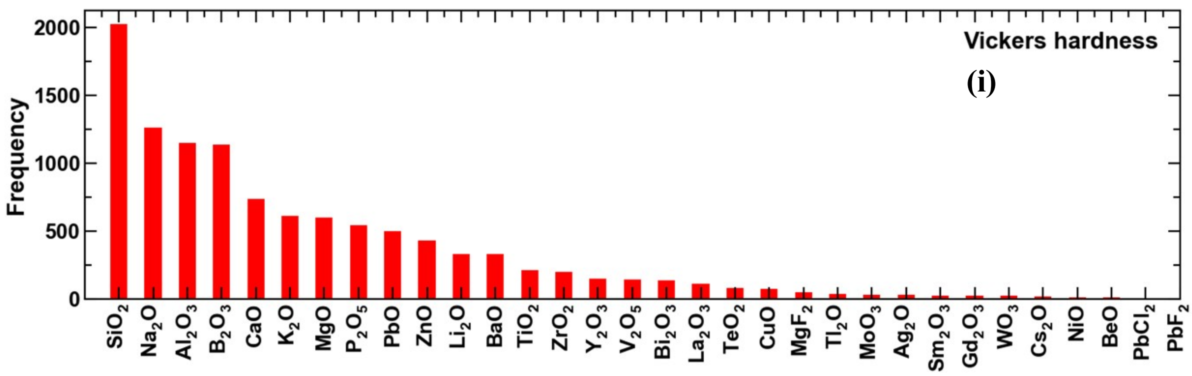
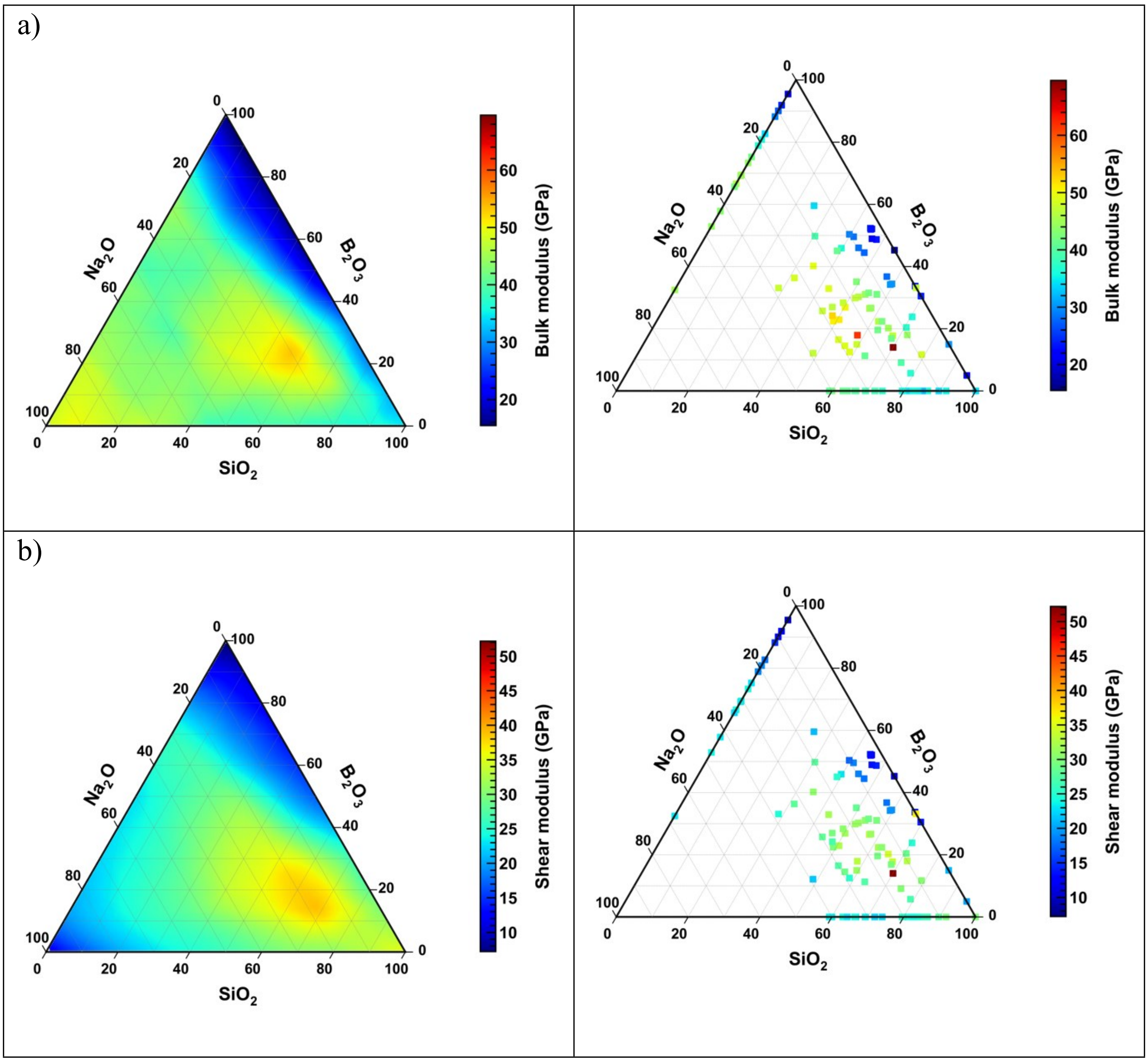
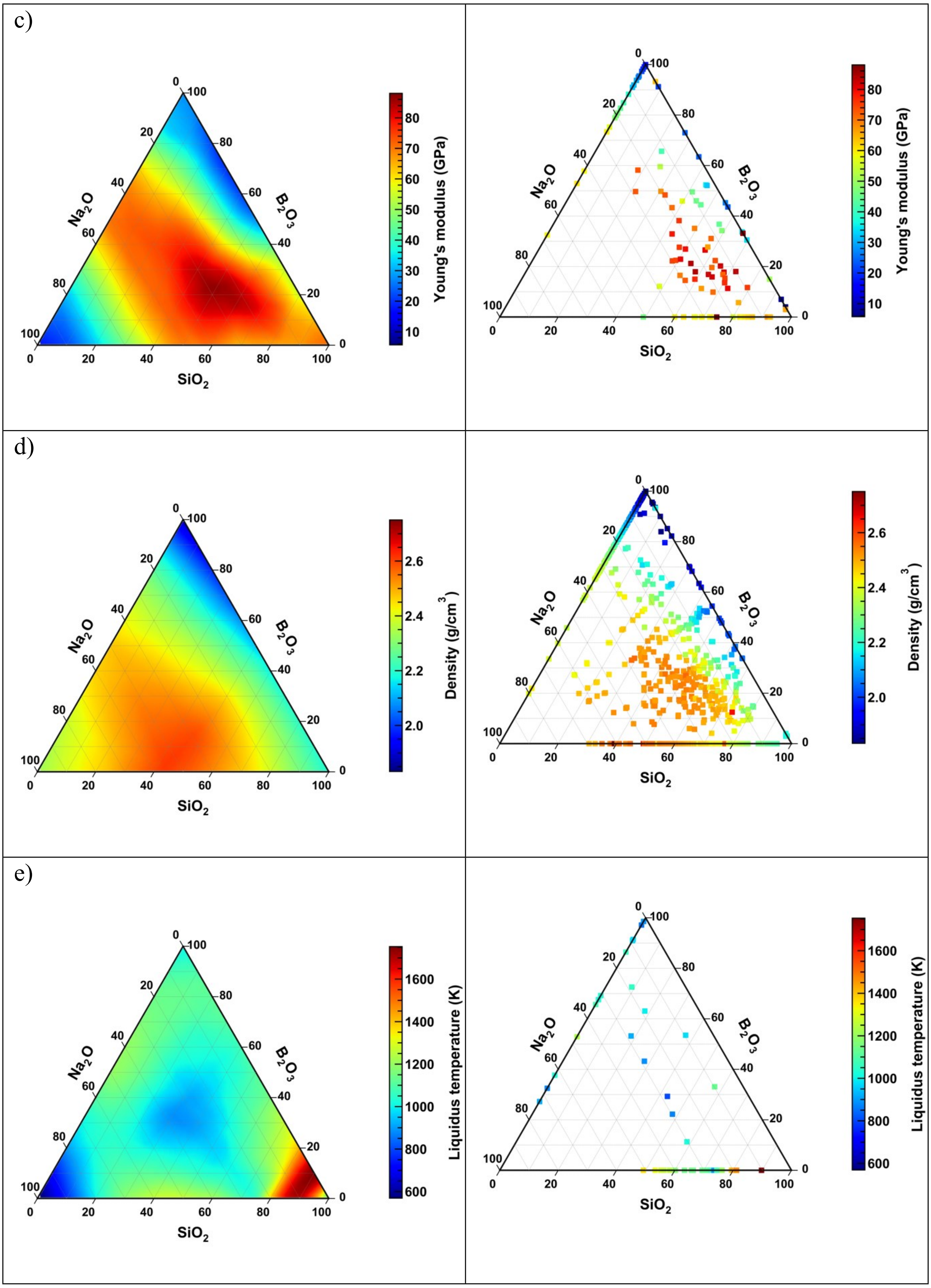
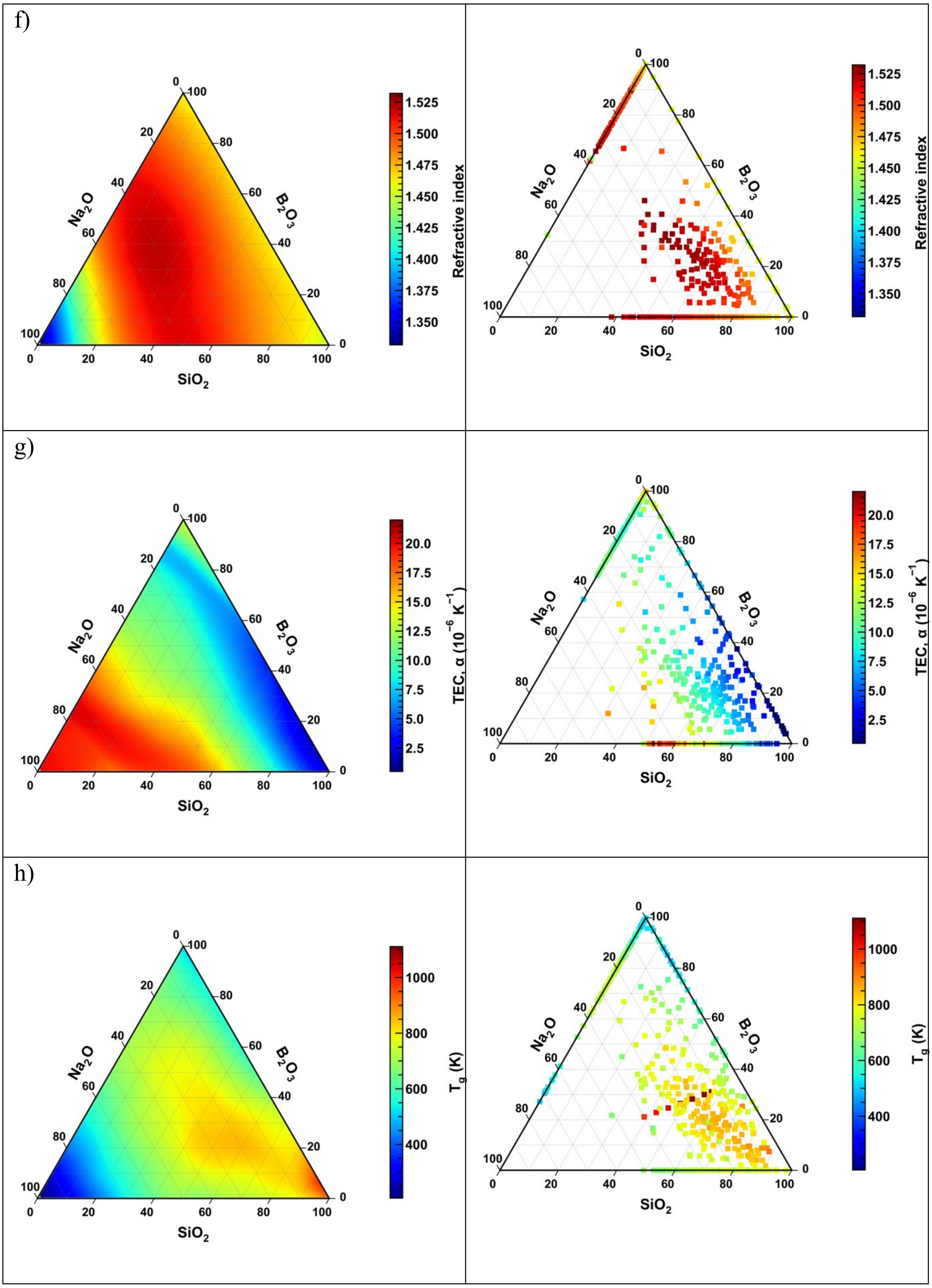


Figure S2. Frequency of glasses with respect to each of the input components for (a) bulk modulus, (b) shear modulus, (c) Young's modulus, (d) density, (e) liquidus temperature, (f) refractive index, (g) thermal expansion coefficient (TEC), (h) glass transition temperature (T_g) and (i) hardness.




c)
Na2O
B2O3
SiO2
Young's modulus (GPa)
d)
Na2O
B2O3
SiO2
Density (g/cm3)
e)
Na2O
B2O3
SiO2
Liquidus temperature (K)



f)
Na2O
SiO2
B2O3
Refractive index
g)
TEC, α (10^-6 K^-1)
h)
Tg (K)

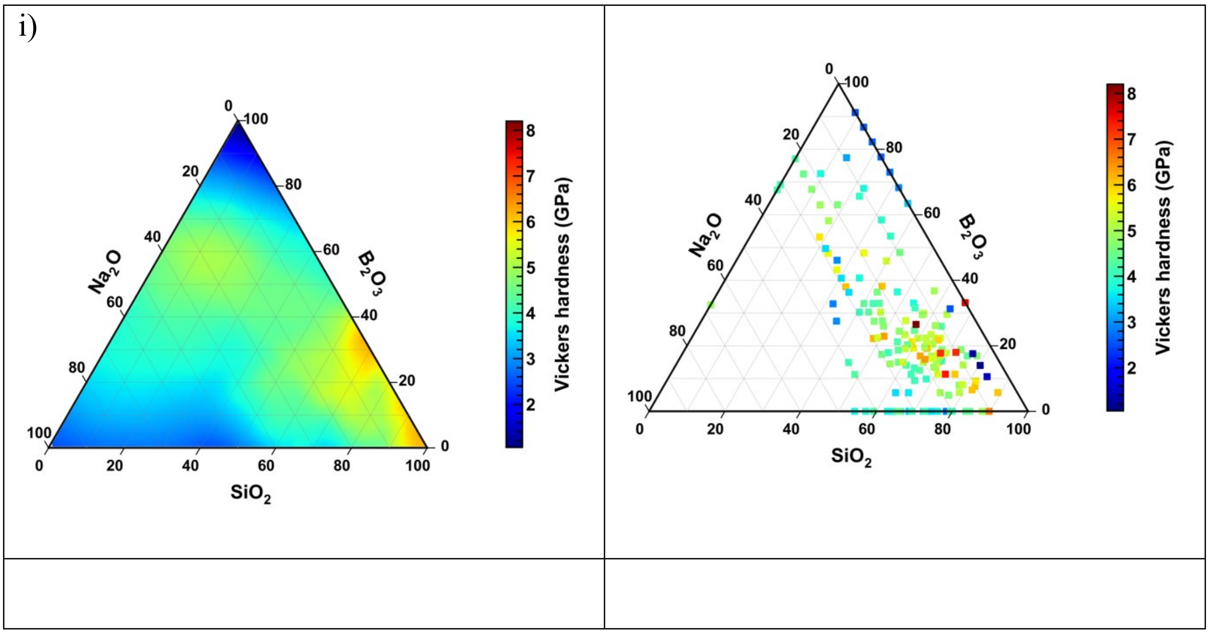
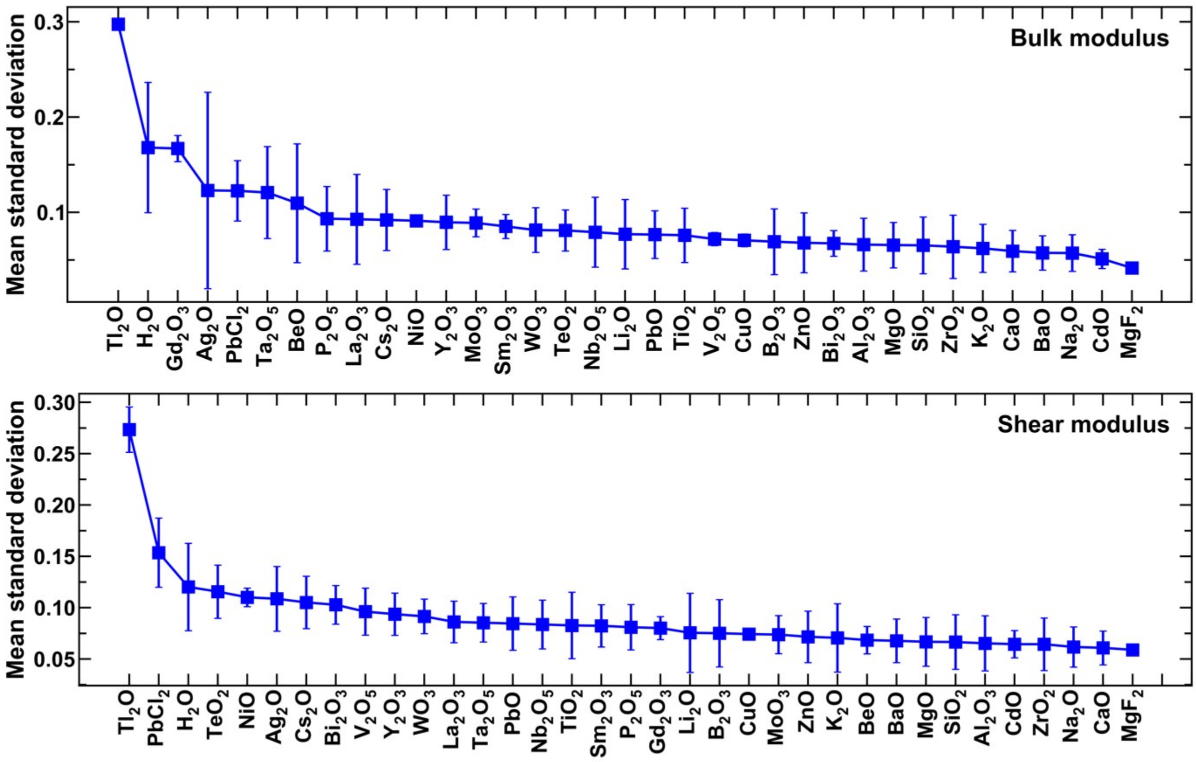
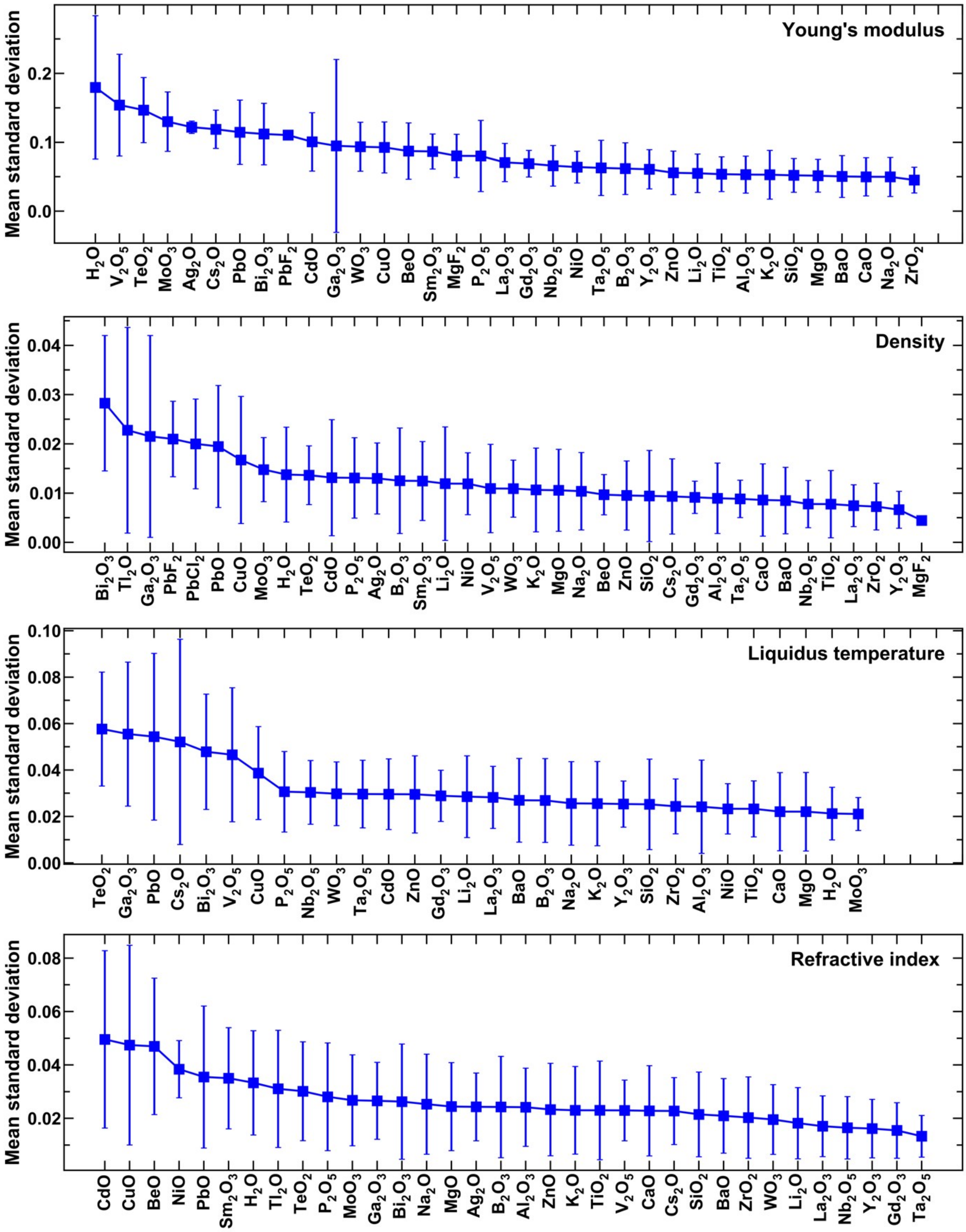


Figure S3. Predicted and experimental values (a) bulk modulus, (b) shear modulus, (c) Young's modulus, (d) density, (e) liquidus temperature, (f) refractive index, (g) thermal expansion coefficient (TEC), (h) glass transition temperature (T_g) and (i) hardness by the trained GPR models for sodium borosilicate glasses. Experimental values are marked in the right panel, while the predicted values are shown in the left panel.




Young's modulus
Mean standard deviation
0.0
0.1
0.2
H_2O V_2O_5 TeO_2 MoO_3 Ag_2O Cs_2O PbO Bi_2O_3 PbF_2 CdO Ga_2O_3 WO_3 CuO BeO Sm_2O_3 MgF_2 P_2O_5 La_2O_3 Gd_2O_3 Nb_2O_5 NiO Ta_2O_5 B_2O_3 Y_2O_3 ZnO Li_2O TiO_2 Al_2O_3 K_2O SiO_2 MgO BaO CaO Na_2O ZrO_2
Density
Mean standard deviation
0.00
0.01
0.02
0.03
0.04
Bi_2O_3 Tl_2O Ga_2O_3 PbF_2 $PbCl_2$ PbO CuO MoO_3 H_2O TeO_2 CdO P_2O_5 Ag_2O B_2O_3 Sm_2O_3 Li_2O NiO V_2O_5 WO_3 K_2O MgO Na_2O BeO ZnO SiO_2 Cs_2O Gd_2O_3 Al_2O_3 Ta_2O_5 CaO BaO Nb_2O_5 TiO_2 La_2O_3 ZrO_2 Y_2O_3 MgF_2
Liquidus temperature
Mean standard deviation
0.00
0.02
0.04
0.06
0.08
0.10
TeO_2 Ga_2O_3 PbO Cs_2O Bi_2O_3 V_2O_5 CuO P_2O_5 Nb_2O_5 WO_3 Ta_2O_5 CdO ZnO Gd_2O_3 Li_2O La_2O_3 BaO B_2O_3 Na_2O K_2O Y_2O_3 SiO_2 ZrO_2 Al_2O_3 NiO TiO_2 CaO MgO H_2O MoO_3
Refractive index
Mean standard deviation
0.02
0.04
0.06
0.08
CdO CuO BeO NiO PbO Sm_2O_3 H_2O Tl_2O TeO_2 P_2O_5 MoO_3 Ga_2O_3 Bi_2O_3 Na_2O MgO Ag_2O B_2O_3 Al_2O_3 ZnO K_2O TiO_2 V_2O_5 CaO Cs_2O SiO_2 BaO ZrO_2 WO_3 Li_2O La_2O_3 Nb_2O_5 Y_2O_3 Gd_2O_3 Ta_2O_5

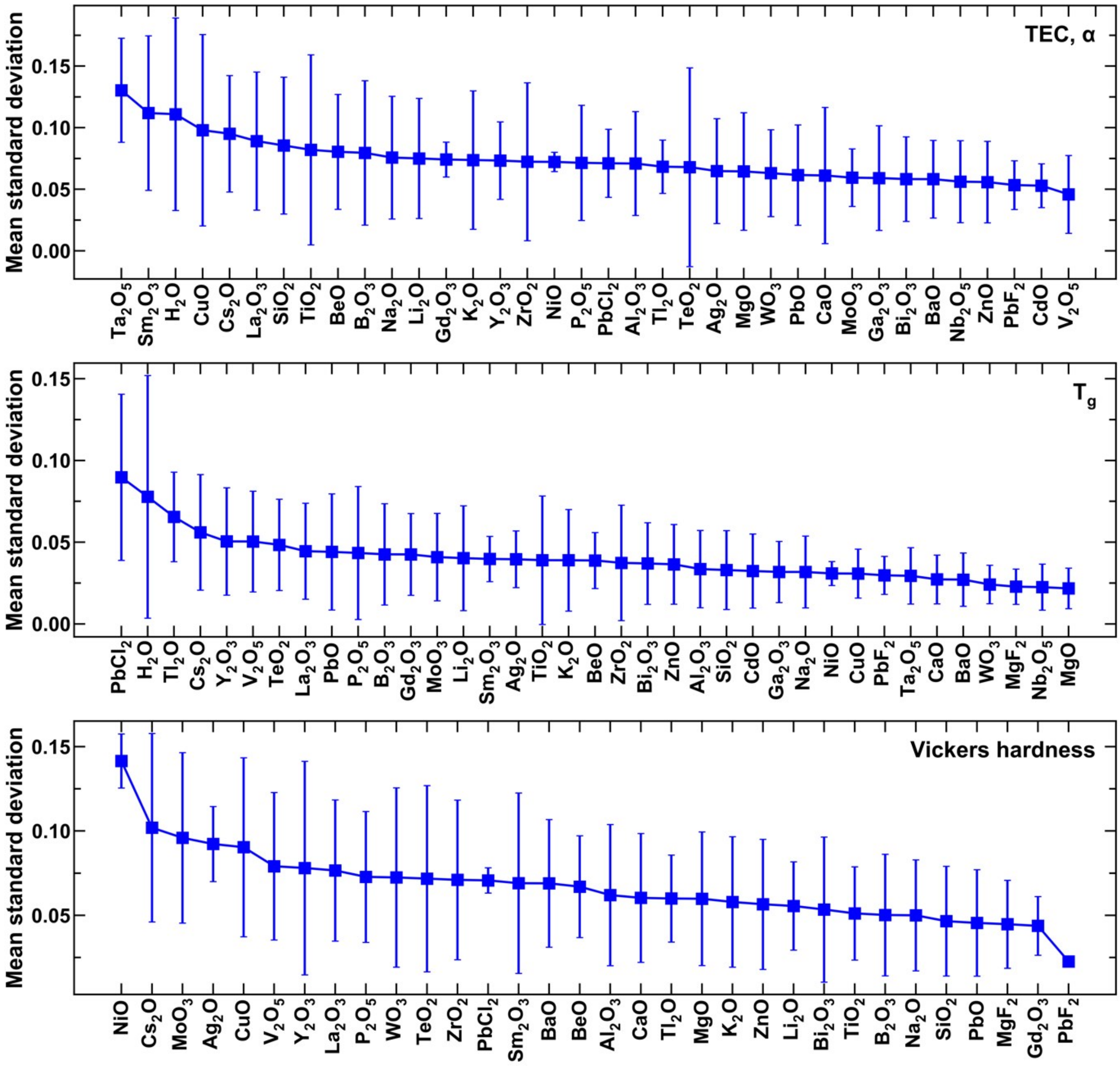


Figure S4: Mean standard deviation for each composition to show how each composition influence the standard deviation of property for (a) bulk modulus, (b) shear modulus, (c) Young's modulus, (d) density, (e) liquidus temperature, (f) refractive index, (g) thermal expansion coefficient (TEC), (h) glass transition temperature (T_g) and (i) hardness.

Least Angle Regression

Figures S5 shows the R^2 and mean squared error (MSE) values with respect to the number of input parameters (that is, oxide components) for a LARS model for (a) Young's modulus and (b) density. We observe that for both Young's modulus and density the R^2 value increases and MSE decreases as number of input parameter increases. Around 30 input components, R^2 value saturates for both the properties. Similar trend is exhibited by other properties as well. Finally, the union of the selected components lead to a total of 37 input components. By doing this, we reduce the dimensionality of the input dataset while ensuring minimal loss in the available information.

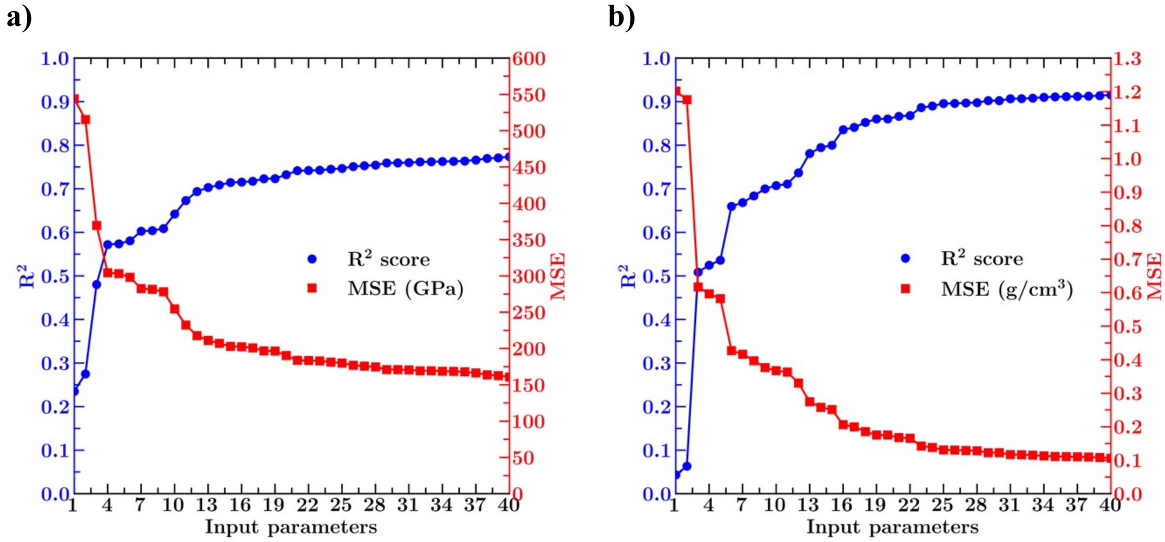
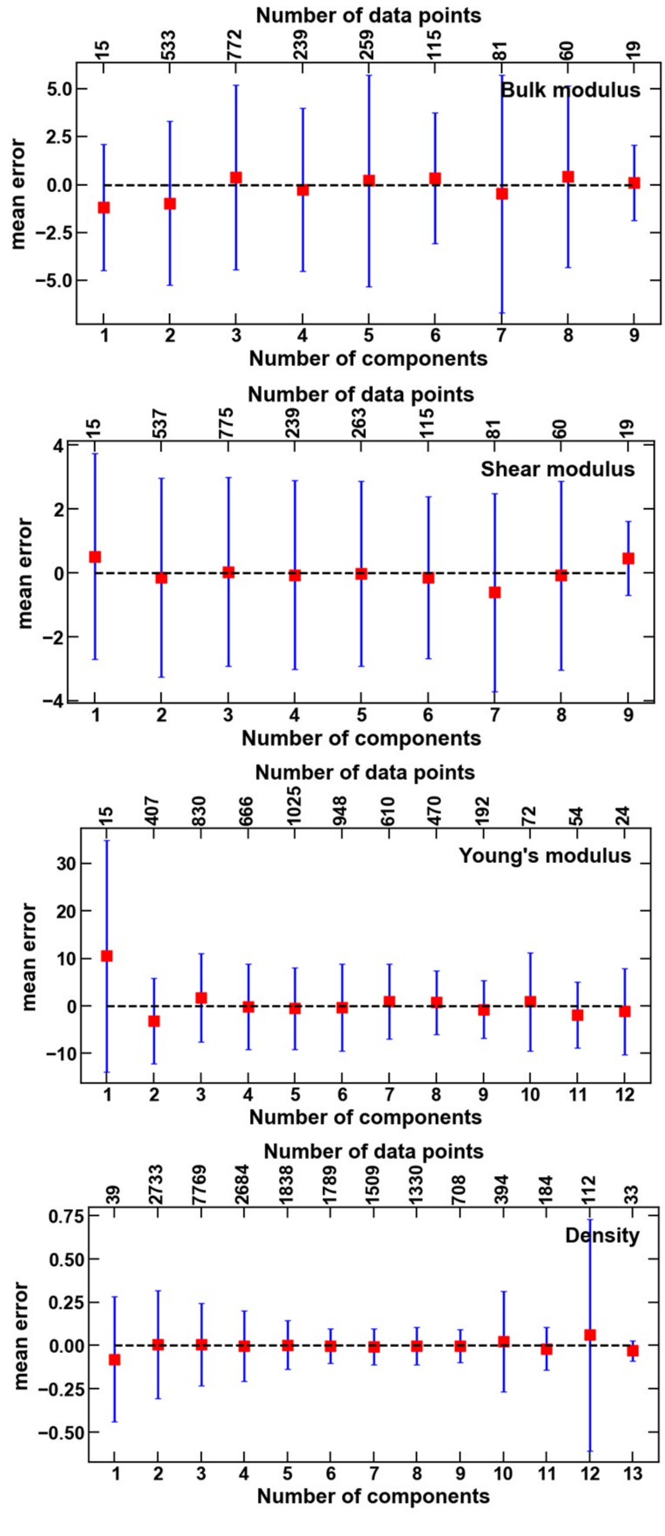
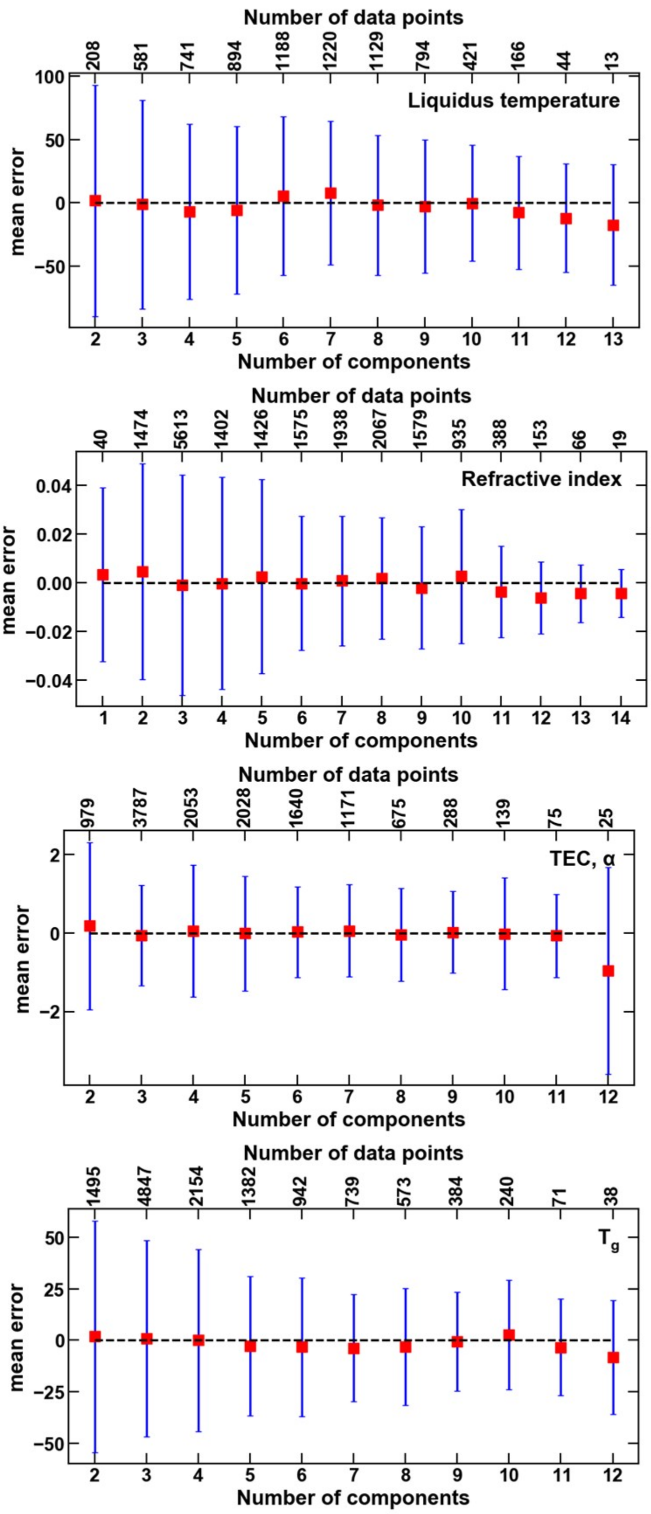
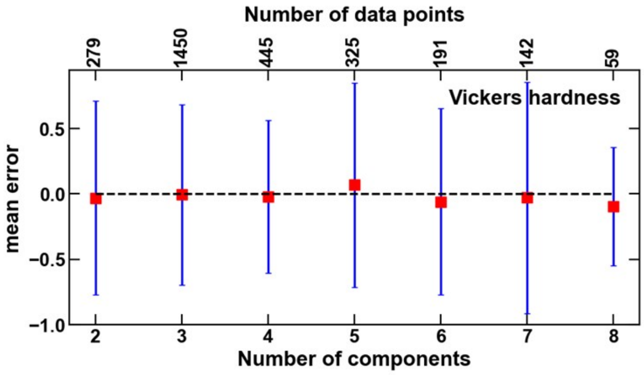


Figure S5. R2 and MSE obtained from LARS with respect to input parameters (number compositions which are having good covariance with output labels) for **a)** Young's modulus and **b)** Density


Number of data points
15 533 772 239 259 115 81 60 19
Bulk modulus
mean error
5.0 2.5 0.0 −2.5 −5.0
1 2 3 4 5 6 7 8 9
Number of components
Number of data points
15 537 775 239 263 115 81 60 19
Shear modulus
mean error
4 2 0 −2 −4
1 2 3 4 5 6 7 8 9
Number of components
Number of data points
15 407 830 666 1025 948 610 470 192 72 54 24
Young's modulus
mean error
30 20 10 0 −10
1 2 3 4 5 6 7 8 9 10 11 12
Number of components
Number of data points
39 2733 7769 2684 1838 1789 1509 1330 708 394 184 112 33
Density
mean error
0.75 0.50 0.25 0.00 −0.25 −0.50
1 2 3 4 5 6 7 8 9 10 11 12 13
Number of components


Number of data points
208 581 741 894 1188 1220 1129 794 421 166 44 13
Liquidus temperature
mean error
100 50 0 −50
2 3 4 5 6 7 8 9 10 11 12 13
Number of components
Number of data points
40 1474 5613 1402 1426 1575 1938 2067 1579 935 388 153 66 19
Refractive index
mean error
0.04 0.02 0.00 −0.02 −0.04
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Number of components
Number of data points
979 3787 2053 2028 1640 1171 675 288 139 75 25
TEC, α
mean error
2 0 −2
2 3 4 5 6 7 8 9 10 11 12
Number of components
Number of data points
1495 4847 2154 1382 942 739 573 384 240 71 38
T_g
mean error
50 25 0 −25 −50
2 3 4 5 6 7 8 9 10 11 12
Number of components



FigureS6: Mean error along with standard error w.r.t. number of input components.