Supporting information for:

Predicting the Mechanical Properties of Zeolite Frameworks by Machine Learning

Jack D. Evans* and François-Xavier Coudert

Chimie ParisTech, PSL Research University, CNRS, Institut de Recherche de Chimie Paris, 75005 Paris, France

E-mail: jack.evans@chimie-paristech.fr

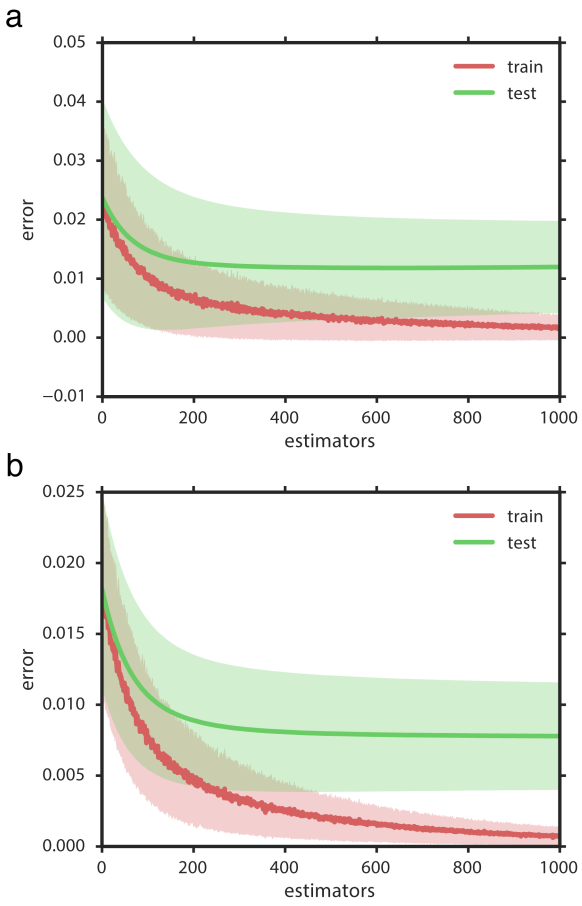

Figure S1: Variance plot illustrating the relative error attributed to cross-validation for prediction of K and G (a and b, respectively). Curves are diplayed as the mean (line) and standard deviation (shading) for 100 testing iterations.

Table S1: Descriptors used in this study

	descriptor	
local	SiOSi average	°
	SiOSi geometric mean	°
	SiOSi harmonic mean	°
	SiOSi maximum	°
	SiOSi mean	°
	SiOSi minimum	°
	SiOSi skew	-
	SiOSi standard deviation	°
	SiOSi variance	°
	SiO average	Å
	SiO geometric mean	Å
	SiO harmonic mean	Å
	SiO maximum	Å
	SiO mean	Å
	SiO minimum	Å
	SiO skew	-
	SiO standard deviation	Å
	SiO variance	Å
structural	density	$g.cm^3$
	volume	$Å^3$ per SiO_2 unit
	space group	-
porosity	accessible surface area	$m^2.g^{-1}$, 1.82 Å probe radius
	nonaccessible surface area	$m^2.g^{-1}$, 1.82 Å probe radius
	accessible volume	$cm^3.g^{-1}$, 1.3 Å probe radius
	nonaccessible volume	$cm^3.g^{-1}$, 1.3 Å probe radius
	volume fraction	1.3 Å probe radius
	largest free sphere	Å
	largest included sphere	Å
	largest included free sphere	Å
	maximum pore dimensionality	-
	minimum pore dimensionality	-
	mode pore dimensionality	-

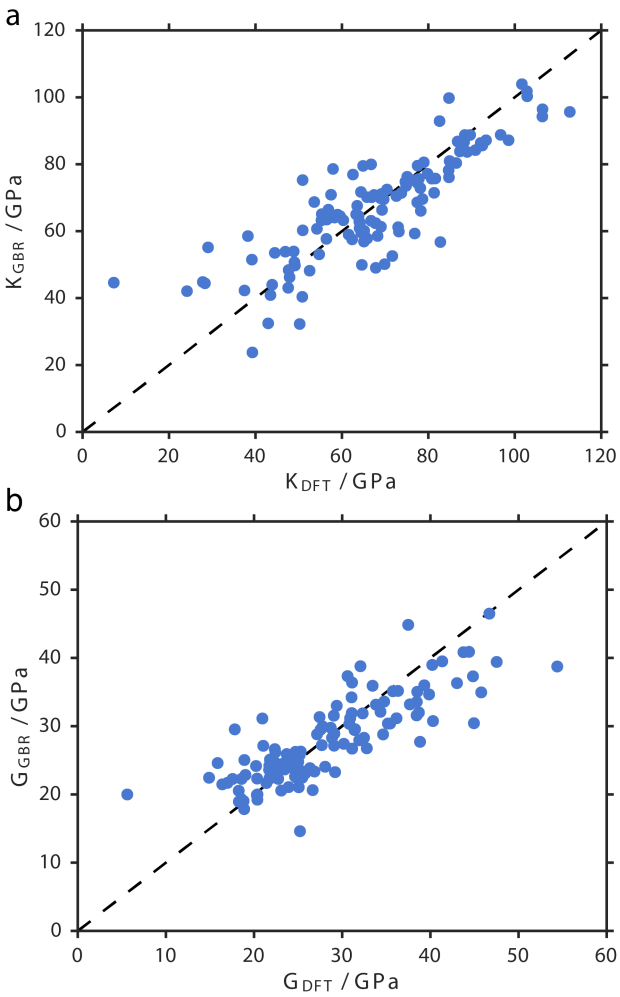

Figure S2: Comparison of cross-validated DFT training set with GBR predictions for bulk modulus K (a) and shear modulus G (b) of 121 pure silica zeolites where the model was trained on K and G not $\log(K)$ and $\log(G)$ as presented in the article.

Table S2: RMSE for the GBR model trained for K and G, not $\log(K)$ and $\log(G)$, and conventional model potentials in comparison to the DFT dataset

method	K **RMSE**	G **RMSE**
GBR	10.00 ± 1.64	4.74 ± 0.69
BKS	22.7	36.1
Catlow	18.8	11.7
Gale	20.0	12.6
Sastre	18.1	141
Matsui	16.8	29.4

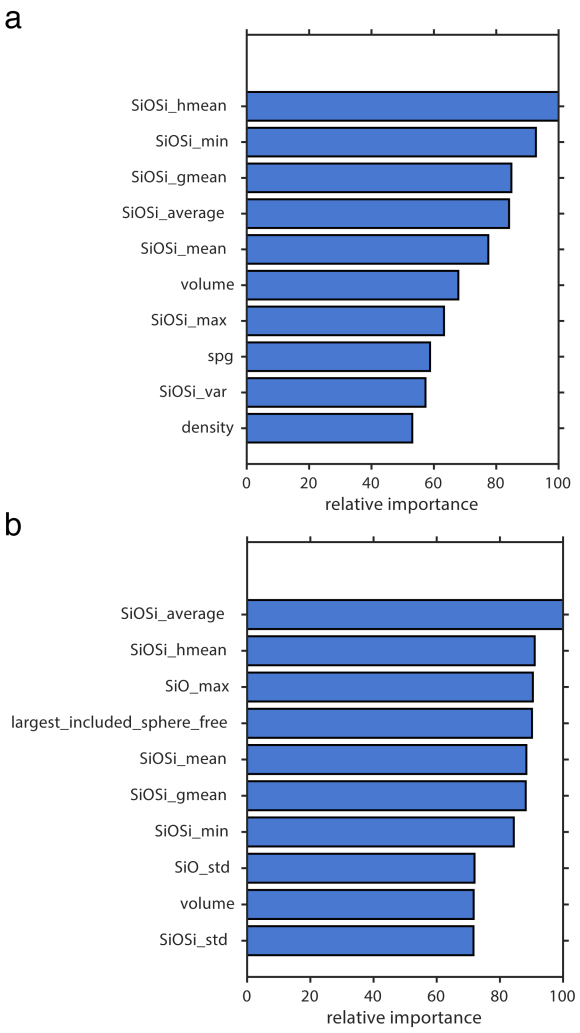

Figure S3: Relative importances of the descriptors used for production of the GBR model for log(K) (a) and log(G) (b).

Table S3: List of IZA zeolites investigated

ABW	ASV	CDO	ESV	ITH	LTF	MTW	RRO	SFN	TUN
ACO	ATN	CFI	ETR	ITR	LTJ	MVY	RSN	SFO	UEI
AEI	ATO	CGF	EUO	ITW	LTL	MWW	RTE	SFS	UFI
AEL	ATS	CGS	EZT	IWR	LTN	NAB	RTH	SGT	UOS
AEN	ATT	CHA	FAR	IWS	MAR	NAT	RUT	SIV	UOZ
AET	ATV	CON	FER	IWV	MAZ	NES	RWR	SOD	USI
AFG	AVL	CZP	FRA	IWW	MEI	NON	RWY	SOF	UTL
AFN	AWO	DAC	GIS	JBW	MEL	NPO	SAF	SOS	UWY
AFO	AWW	DDR	GIU	JOZ	MEP	NPT	SAO	SSF	VET
AFR	BCT	DFO	GME	JRY	MER	NSI	SAS	SSY	VFI
AFS	BEC	DFT	GON	JSN	MFI	OBW	SAT	STF	VNI
AFT	BIK	DOH	GOO	JST	MFS	OFF	SAV	STI	VSV
AFV	BOF	DON	HEU	JSW	MON	OSI	SBE	STT	WEI
AFX	BOG	EAB	IFR	KFI	MOR	OWE	SBN	STW	YUG
AFY	BOZ	EDI	IHW	LAU	MOZ	PAU	SBS	SZR	ZON
AHT	BPH	EEI	IMF	LEV	MSE	PCR	SBT	TER	
ANA	BRE	EMT	IRR	LIO	MSO	PHI	SFE	THO	
APC	BSV	EON	ISV	LOS	MTF	PON	SFF	TOL	
APD	CAN	EPI	ITE	LOV	MTN	PUN	SFG	TON	
AST	CAS	ERI	ITG	LTA	MTT	RHO	SFH	TSC	

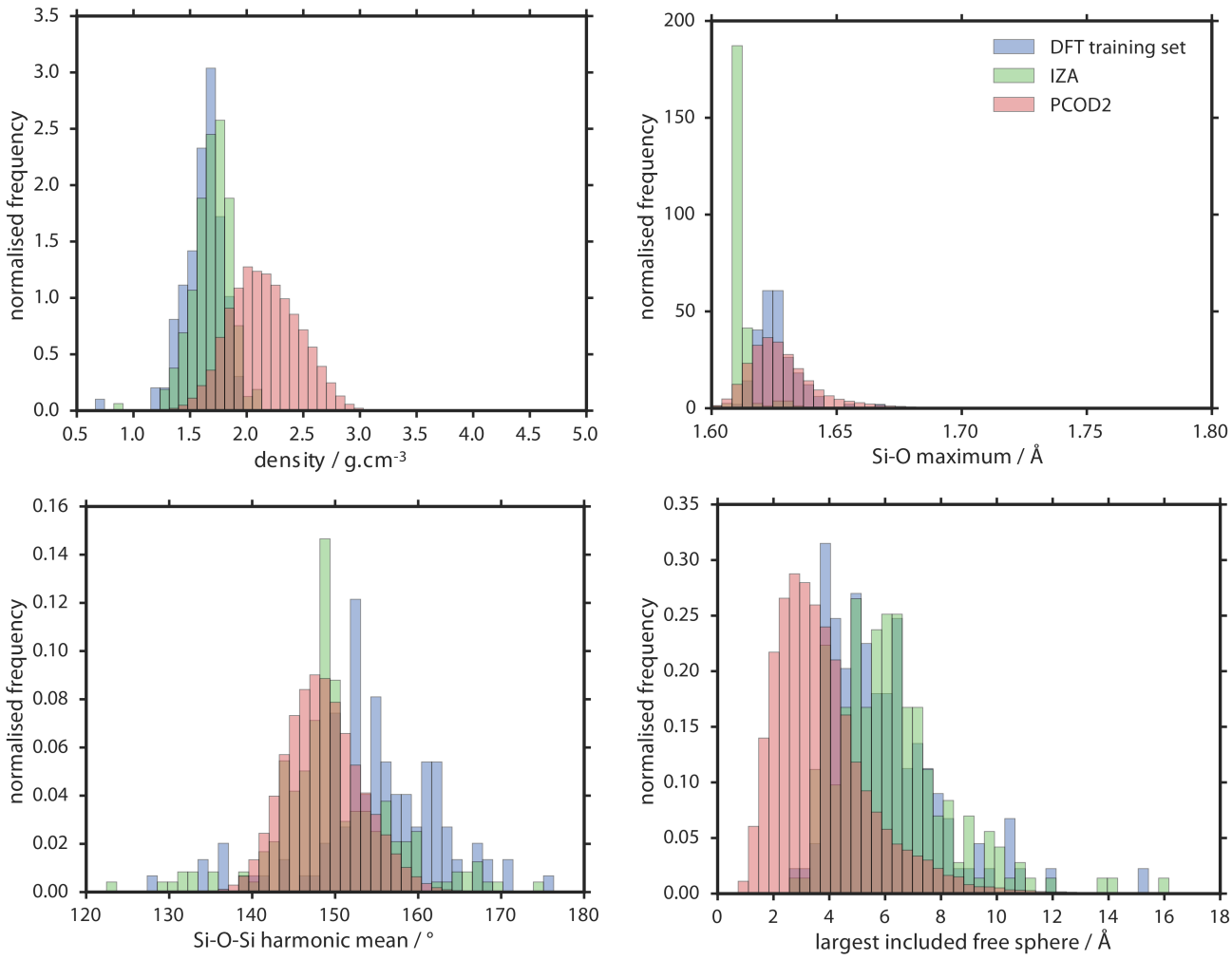

Figure S4: Distributions of density, Si-O maximum, Si-O-Si harmonic mean and largest included free sphere for the three datasets used in this study.