UNIVERSITY OF CAMBRIDGE

Data validation and imputation with artificial intelligence

Gareth Conduit

TCM Group, Department of Physics

Exploiting materials databases

Experiment

Simulation

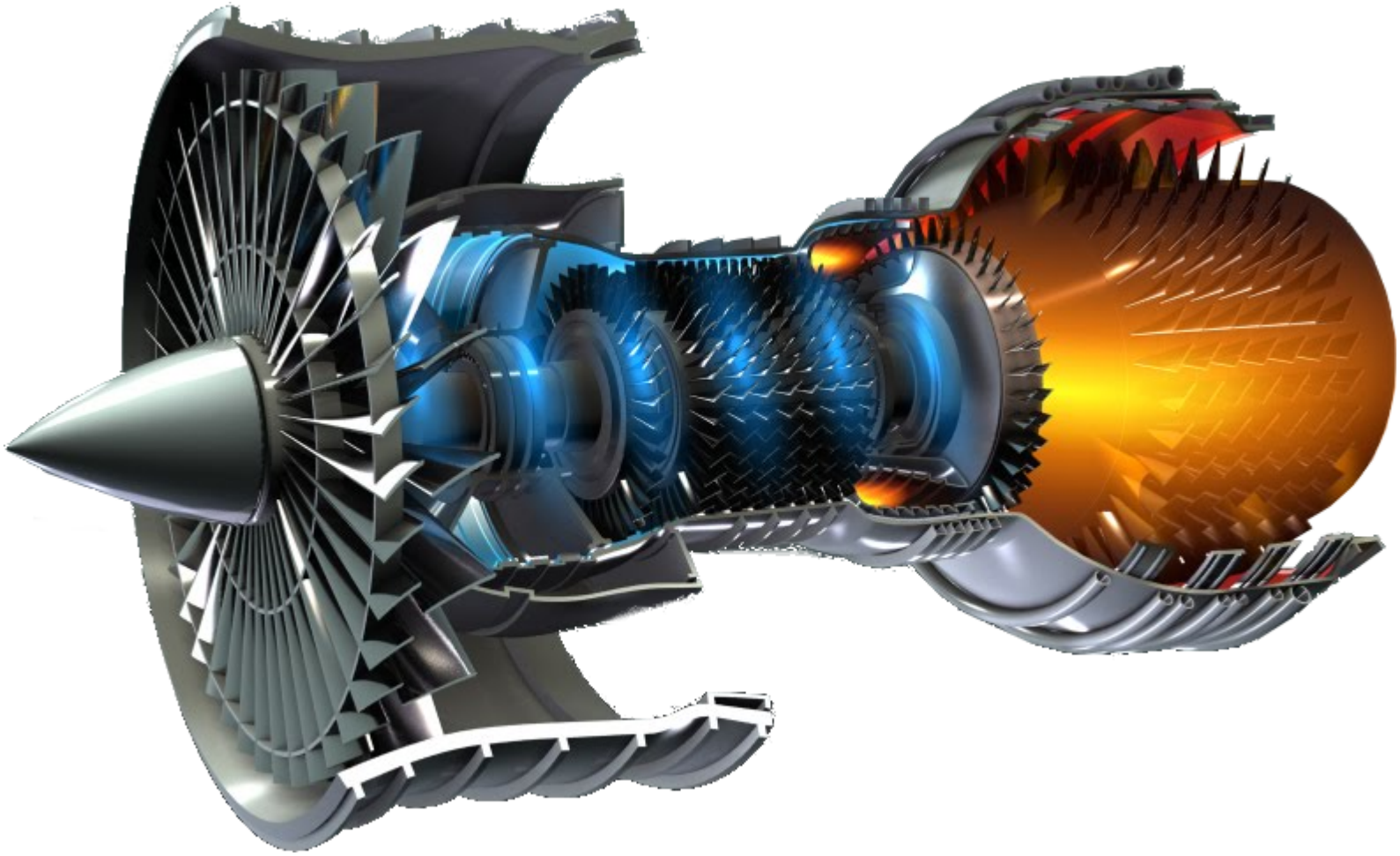
Materials modeling

Exploiting materials databases

Experiment

Materials modeling

Schematic of a jet engine



Artificial intelligence

Composition

Artificial intelligence

Yield stress

Hardness

Melting point

Oxidation resistance

Cost

Density

Fatigue life

Fracture toughness

Creep

Processibility

Artificial intelligence

Composition

Artificial intelligence

Yield stress

Hardness

Melting point

Oxidation resistance

Cost

Density

Fatigue life

Fracture toughness

Creep

Processibility

Artificial intelligence

Composition

Artificial intelligence

Yield stress

Hardness

Melting point

Oxidation resistance

Cost

Density

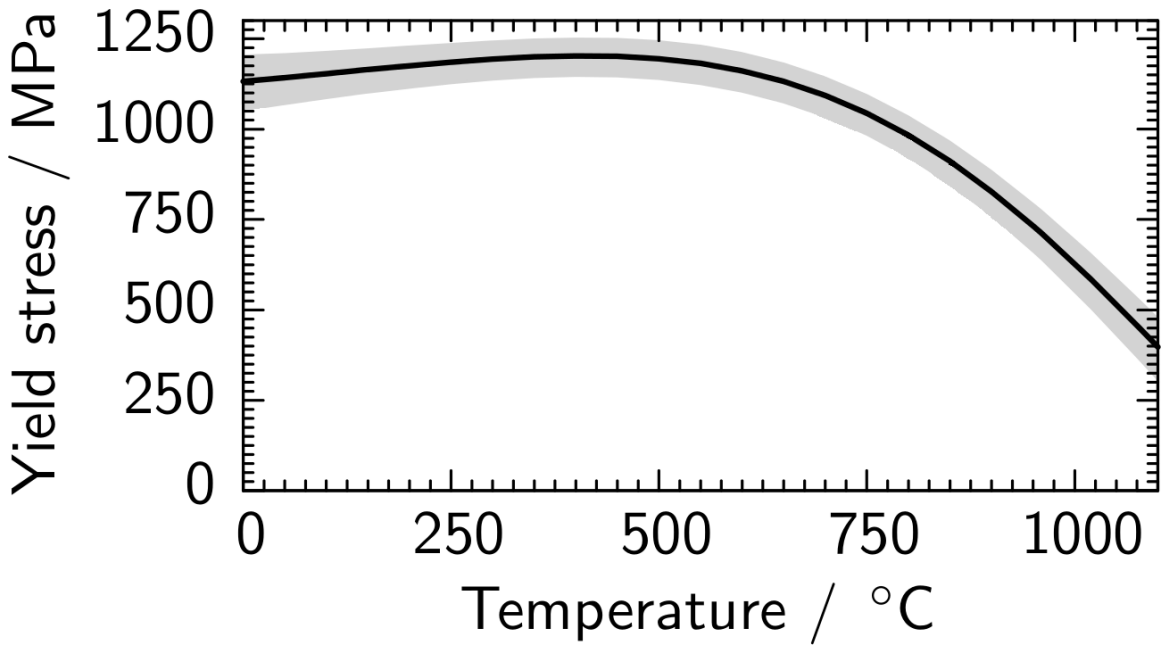
Fatigue life

Fracture toughness

Creep

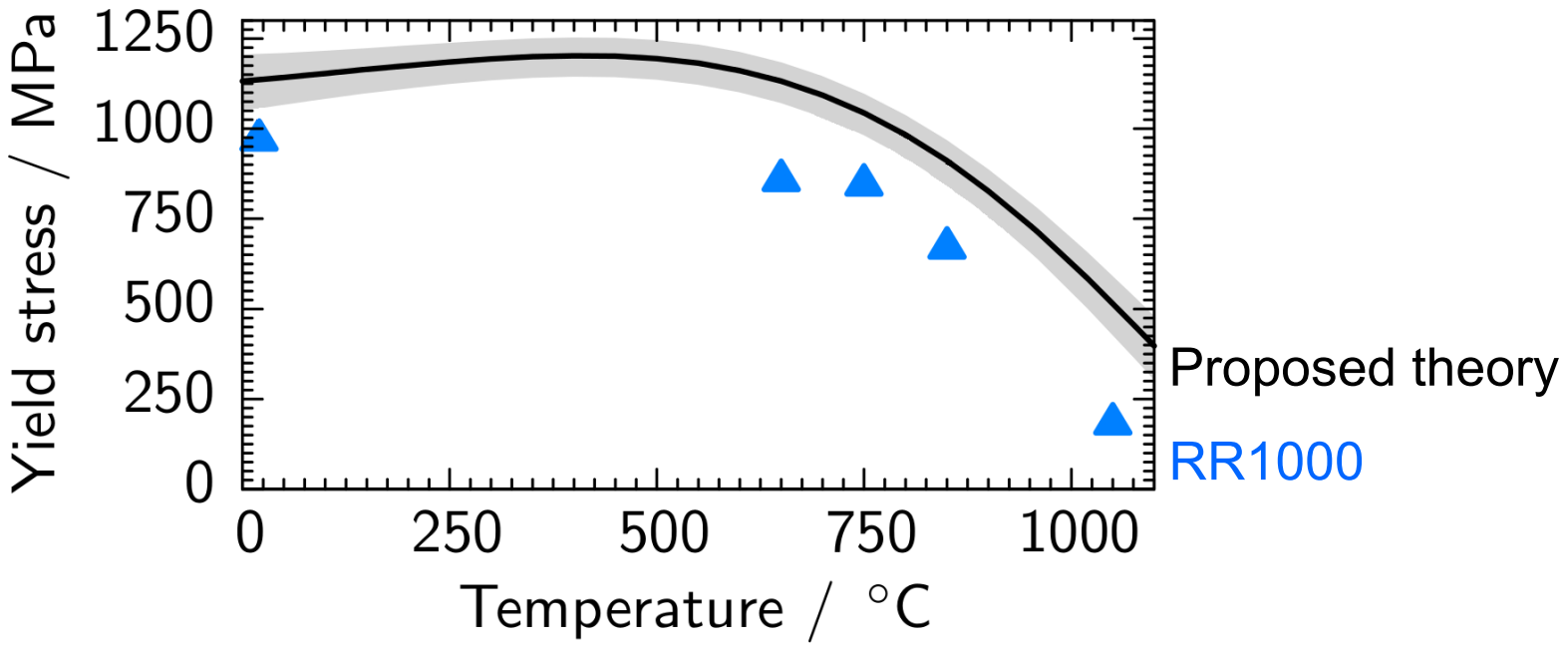
Processibility

Testing the yield stress

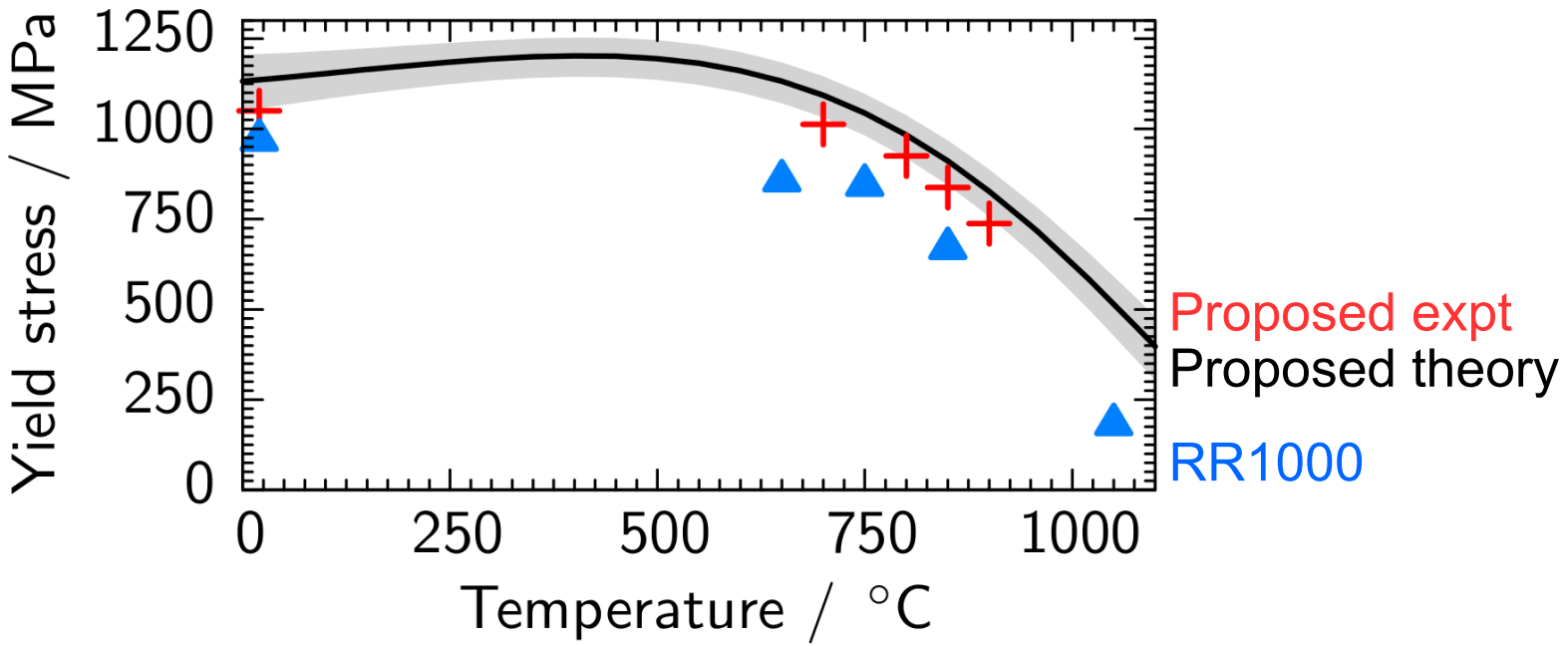


Proposed theory

Testing the yield stress

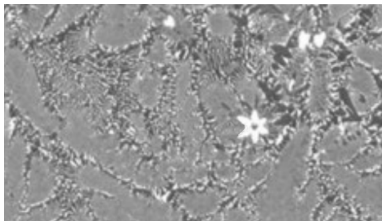


Testing the yield stress

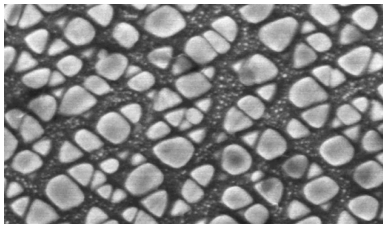


Alloys discovered

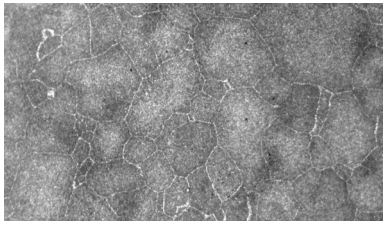
Cr-Cr_2Ta alloys
Intermetallics, 48, 62



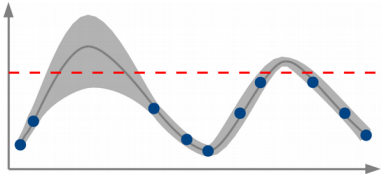
Combustor alloy
GB1408536



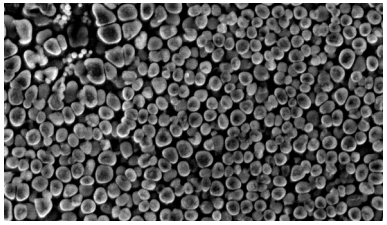
RR1000 grain growth
Acta Materialia, 61, 3378



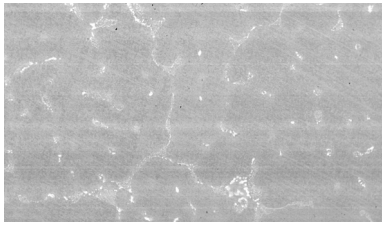
Discovery algorithm
EP14153898
US 2014/177578



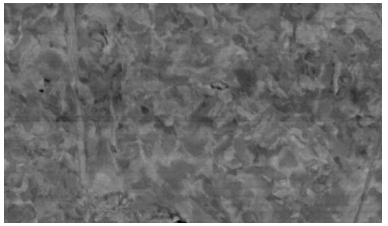
Ni disc alloy
EP14157622
US 2013/0052077 A2



Mo-Hf forging alloy
EP14161255
US 2014/223465



Mo-Nb forging alloy
EP14161529
US 2014/224885



Exploiting materials databases

Experiment

Materials modeling

Exploiting materials databases

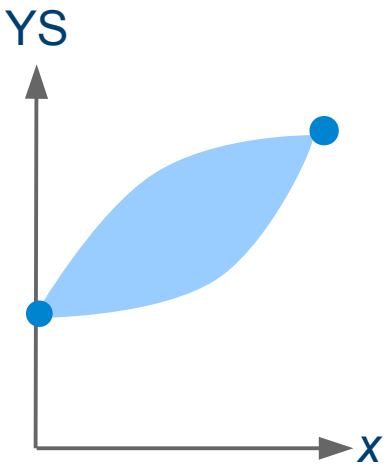
Experiment

Simulation

Materials modeling

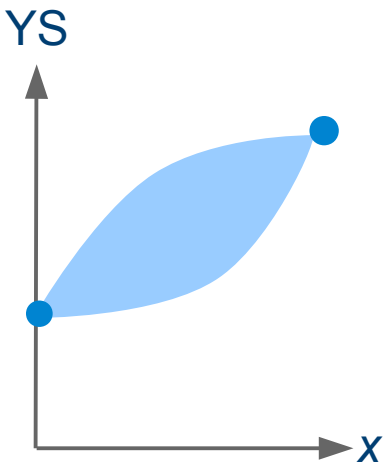
Merging simulation and experiment

Experiment

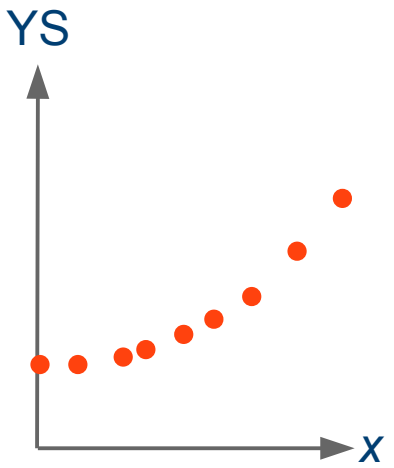


Merging simulation and experiment

Experiment



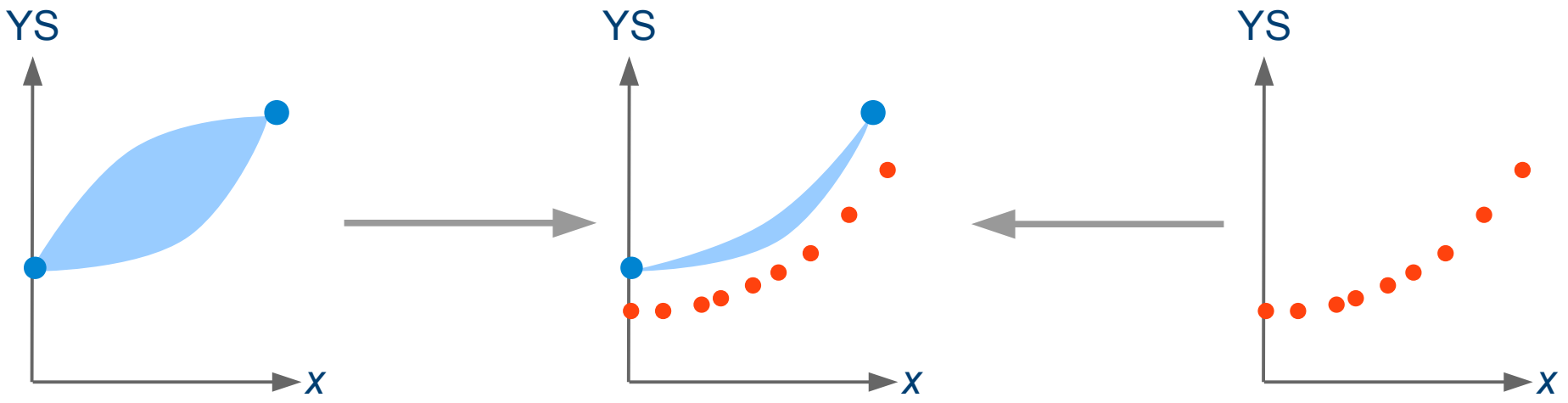
Simulation



Merging simulation and experiment

Experiment

Simulation



Exploiting materials databases

Experiment

Simulation

Materials modeling

Exploiting materials databases

Experiment

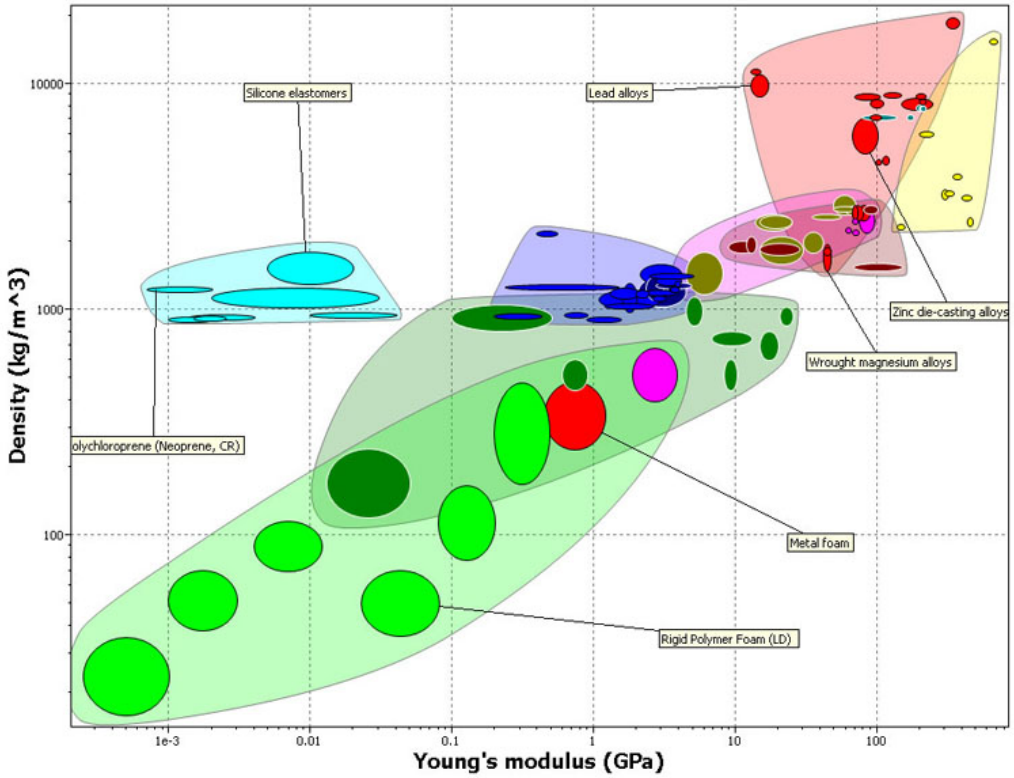
Simulation

Materials modeling

Fragmented databases

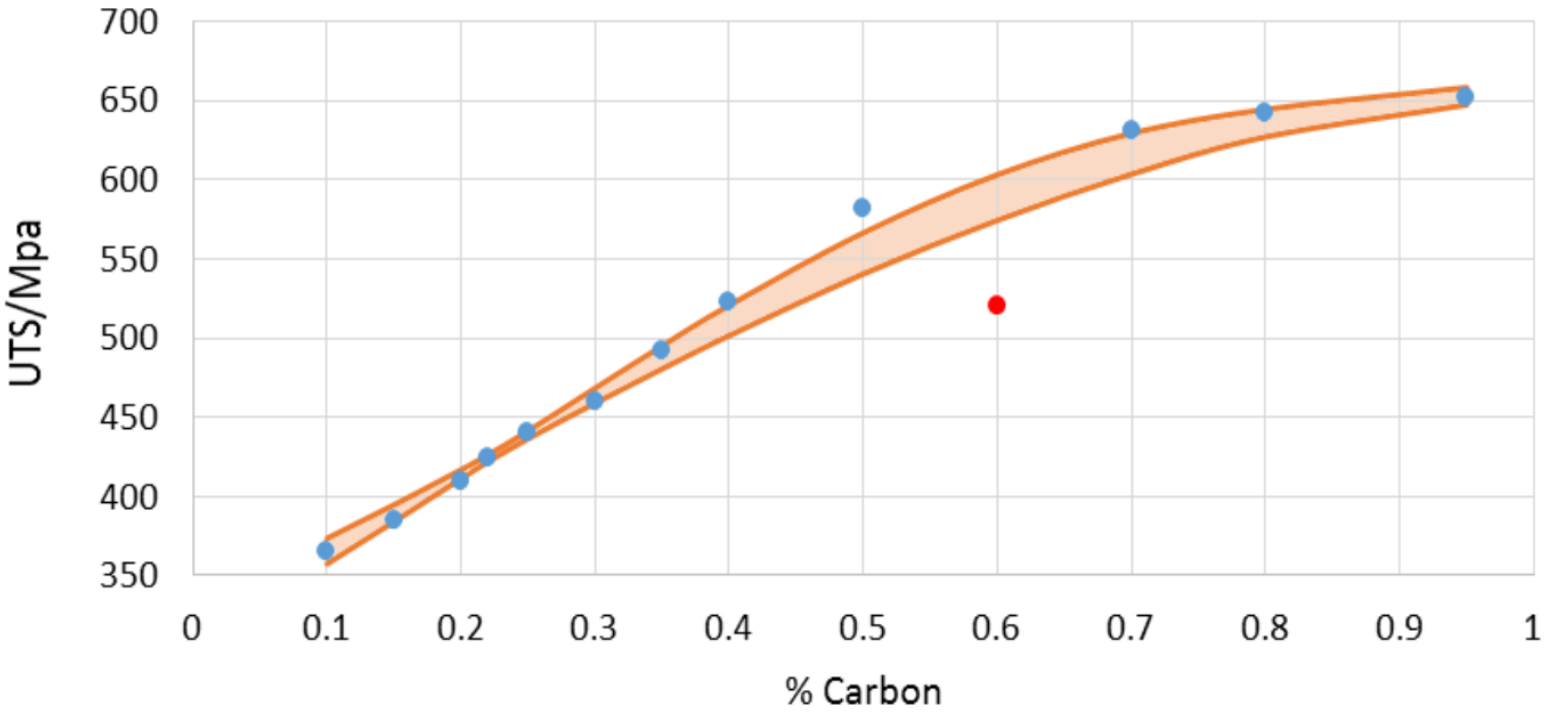
Composition	Computed YS	YS	Hardness	UTS	Charpy	Compressive
✓	✓	✓	✓	x	x	x
✓	✓	x	✓	✓	x	✓
✓	✓	x	x	✓	✓	✓
✓	✓	✓	✓	x	✓	x
✓	✓	x	x	✓	x	x
✓	✓	x	✓	x	✓	x
x	x	✓	x	✓	✓	x
✓	✓	x	✓	x	x	✓
✓	✓	✓	x	x	✓	x

Database verification



10,000,000 entries, 20% filled

Erroneous points: 1D example



Erroneous points: density & yield stress

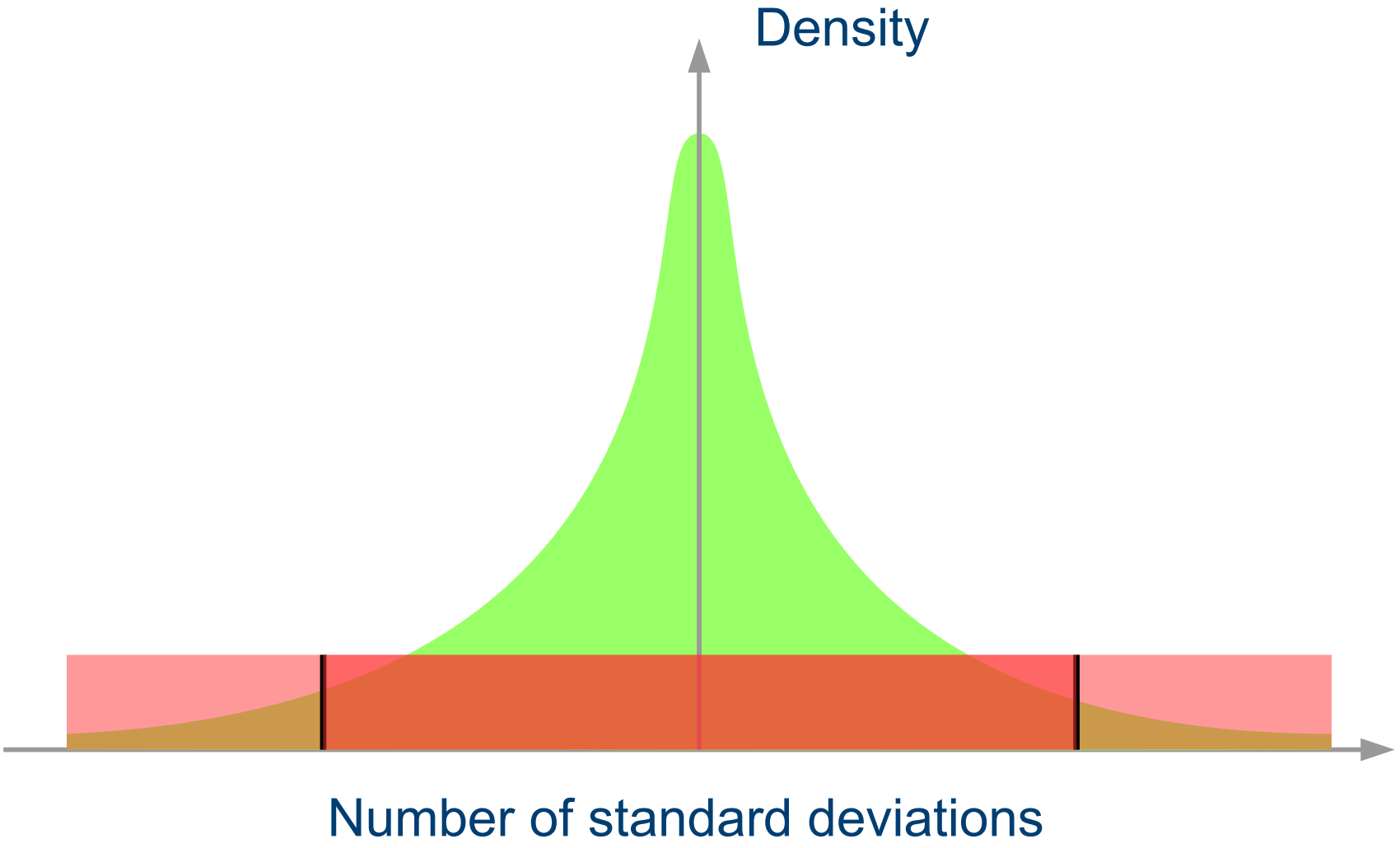
Alloy	Source densit/gcc^{-1}	NN density/gcc^{-1}	#Sigma	Correct density /gcc^{-1}
Stainless steel, duplex, Ilium P, cast	7.60	7.9060	11.6	7.75-8.0
Tool steel, molybdenum alloy, AISI M43 (high speed)	8.44	8.0345	-11.5	7.7-8.03
Copper-nickel alloy, C70400, wrought, half hard (95/5 copper-nickel)	8.53	8.9254	10.7	8.94
Tool steel, AISI A3	8.00	7.7211	-20.2	7.86
Tool steel, AISI A4	7.91	7.80	8.8	8.03
			11.4	7.70-8.03

	Granta Melting Point/K	NN Melting Point /K	#Sigma	Actual Value/K
Wrought iron	1973.0			
Nickel-Fe-Cr alloy, INCOLOY 840, annealed	1419.0			
Titanium, alpha-beta alloy, Ti-4.5Al-3V-2Fe-2Mo, annealed	1593.2			
Carbon steel, AISI 1095	1650.2			

Material	UTS /Mpa	Yield stress /Mpa	Neural Network prediction	# Sigma
Low alloy steel, AISI 4150, annealed	726.64	377.89	520	26.6861
Low alloy steel, AISI 4150, normalized	1149.26	731.16	918	21.5008
Low alloy steel, AISI 4150, tempered at 650°C & oil quenched	952.52	835.69	729	-15.1901

Found 156 erroneous points confirmed against primary sources

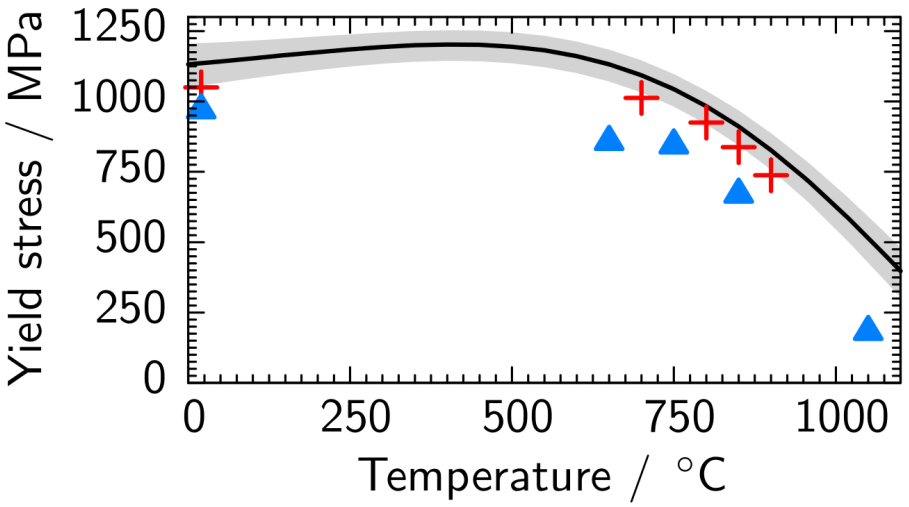
How many erroneous points remain?



Polymers

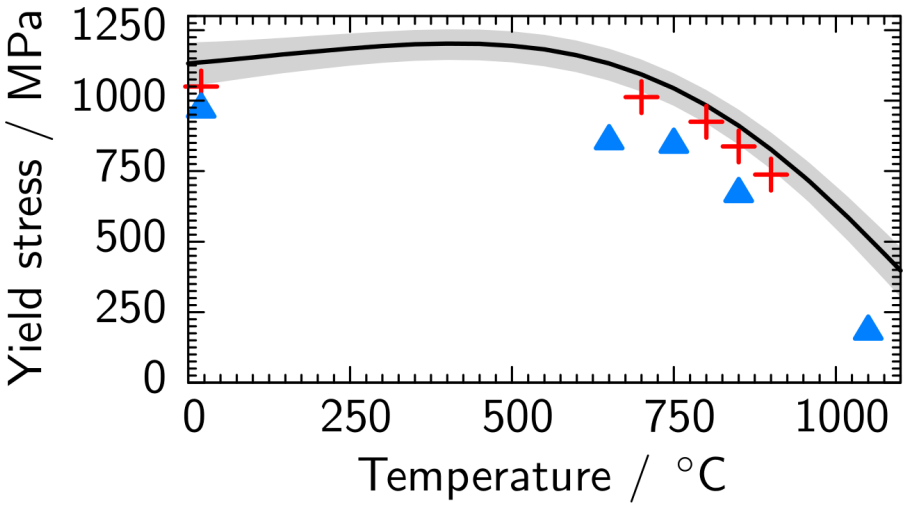
Name	*Property*	*Value*	*Comment*
4PROP®25C21120	Flexural Modulus (MPa)	2300	out by factor 10^3
AZDEL™U400-B01N (Longitudinal)	Flexural Modulus (MPa)	8000	out by factor 10^3
Hyundai EP PPF HT340	Flexural Strength (MPa)	46.9	out by factor 10
Borealis PP NJ201AI	Filler	Mineral: 20%	prediction: $19.9 \pm 5.6\%$
Daplen™EE168AIB	Filler	Mineral: 10%	prediction $11.3 \pm 3.0\%$
Hostacom M2 R03/2 105558	Filler	Mineral: 20%	prediction $14.8 \pm 4.2\%$
Maxxam™NM-818.H001-1049	Filler	Glass Fibre: 20%	prediction $17.8 \pm 4.4\%$
Beetle®PPC120M 9250	Filler	Mineral: 20%	prediction $9.9 \pm 6.2\%$
EMOPLEN®CP GFR 20	Filler	Mineral: 20%	prediction $10.3 \pm 2.7\%$
FORMULA P COMP 5220	Filler	Mineral: 20%	prediction $15.5 \pm 3.2\%$
4PROP®9C13100	Filler	Mineral: 10%	prediction $13.5 \pm 3.0\%$

Graphical data



Composition	Temperature	YS
A	200	1100
A	400	1150
A	600	1100
A	800	1000
A	1000	650
B	100	1000
C	100	1050

Graphical data



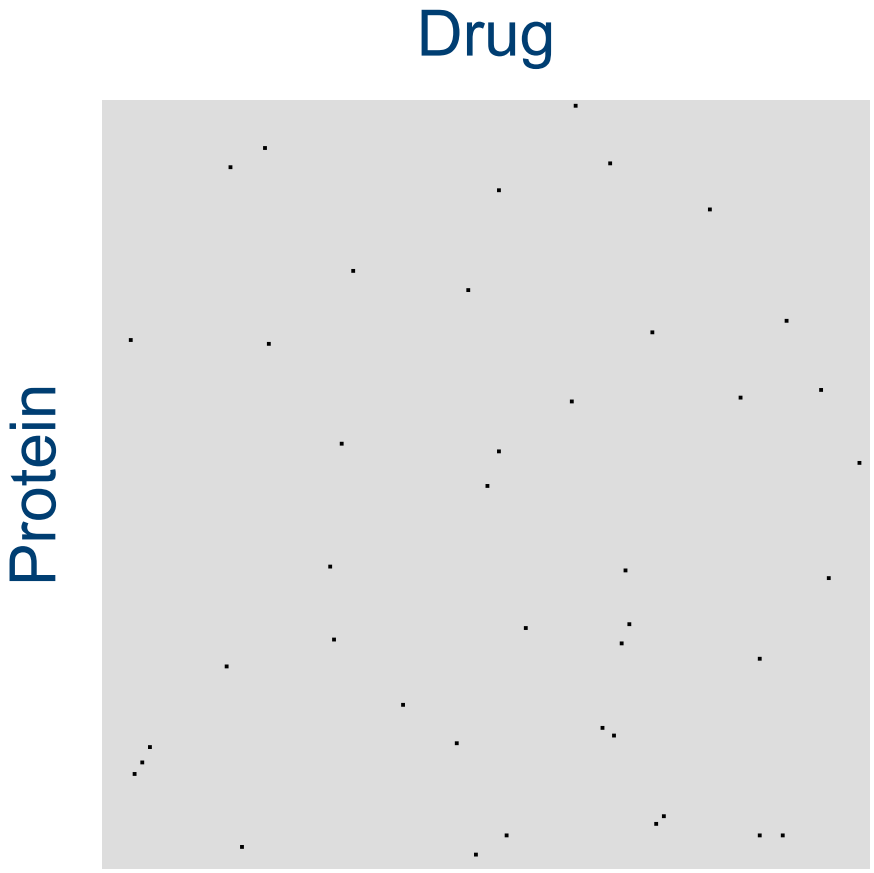
Composition	Temperature	YS
A	200,…,1000	1100,…,650
B	100	1000
C	100	1050

Exploiting experiment and simulations

Lithium cathode materials

Protein activity data

ChEMBL database has protein activity for 0.1% of compounds

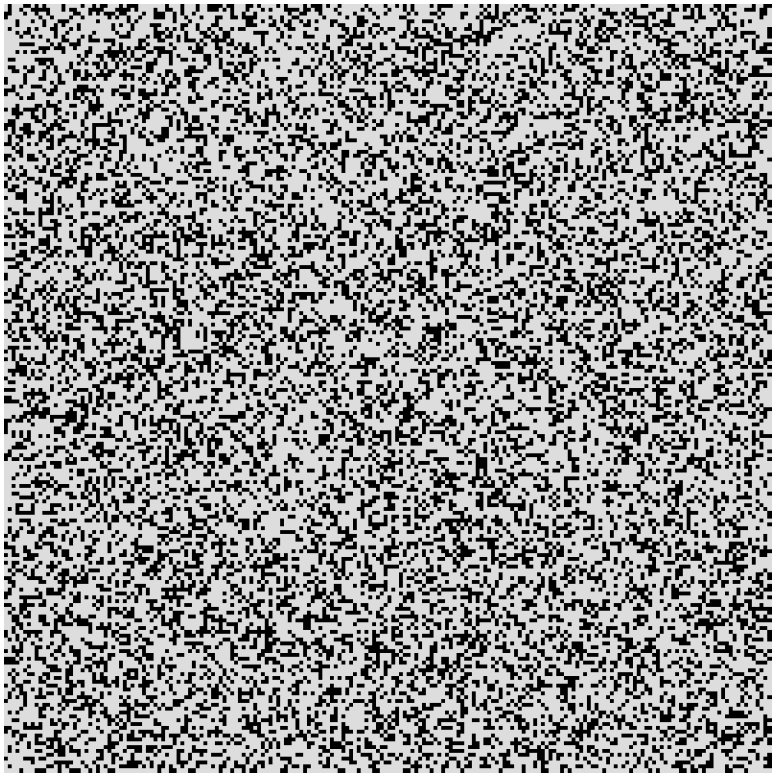


Protein activity data

Filled in 32% of the data points with 99.3% accuracy

Drug

Protein



Summary

Used artificial intelligence to handle fragmented data

Merge simulations and experiments into a holistic tool

Data validation and imputation