Thank you for your attention!

SCI 1.719 IMPACT FACTOR 2016

Sponsored by
Chinese Academy of Sciences
National Natural Science Foundation of China

Published by
《中国科学》杂志社 SCIENCE CHINA PRESS
Springer

SCIENCE CHINA
Technological Sciences

- A monthly peer-reviewed academic journal
- Indexed by SCI, EI, etc.
- Manuscript types:
  Review, Article, Brief report, News & views
- Rapid review and
  timely publication (online immediately)

Free full text available on http://tech.scichina.com

CityU 香港城市大學 City University of Hong Kong
Professional·Creative For The World

City University of Hong Kong
Institute for Advanced Study

IAS Distinguished Lecture

Materials Genome Initiative and Materials Informatics

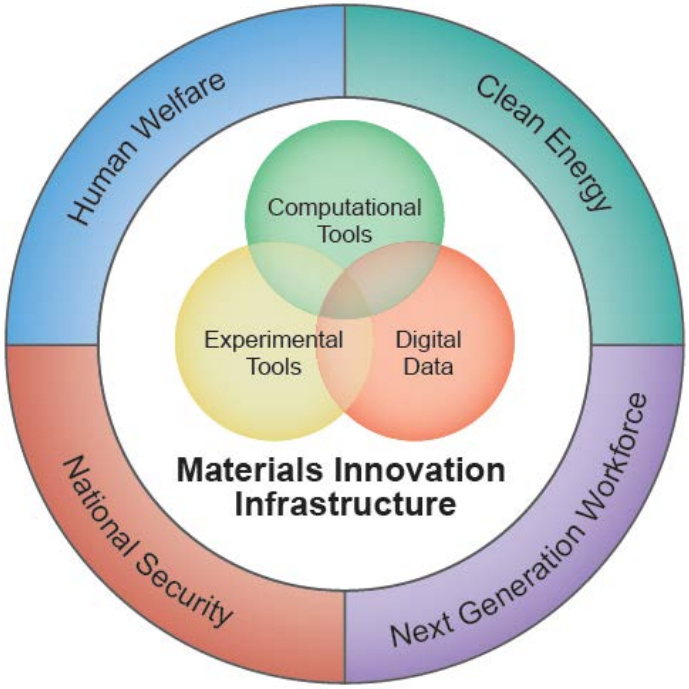
Tong-Yi Zhang

上海大学材料基因组工程研究院

SHANGHAI UNIVERSITY

Gi

MGI



In June 2011, the US President Obama announced AMP(Advanced Manufacturing Partnership), in which "**Materials Genome Initiative"(MGI)** is a key component. The core is to integrate computation, experimentation and database, "to discover, develop, manufacture, and deploy advanced materials at twice the speed than is possible today" and at a half of the current cost.

Half Time, Half Cost

MGI in China

On Dec 21-23, 2011, under the leadership of Profs. **Changxu Shi** and **Kuangdi Xu**, CAE and CAS jointly held the S14 Xiangshan Science Conference: "**Systematic Materials Science and Engineering**"

Four Topics:

(1) MGI databases;

(2) Development of computational methods and software;

(3) High-throughput material preparation & characterization tools;

(4) Key materials research and breakthrough.

Four Strategic Goals of MGI

- **Enable a Paradigm Shift in Culture**

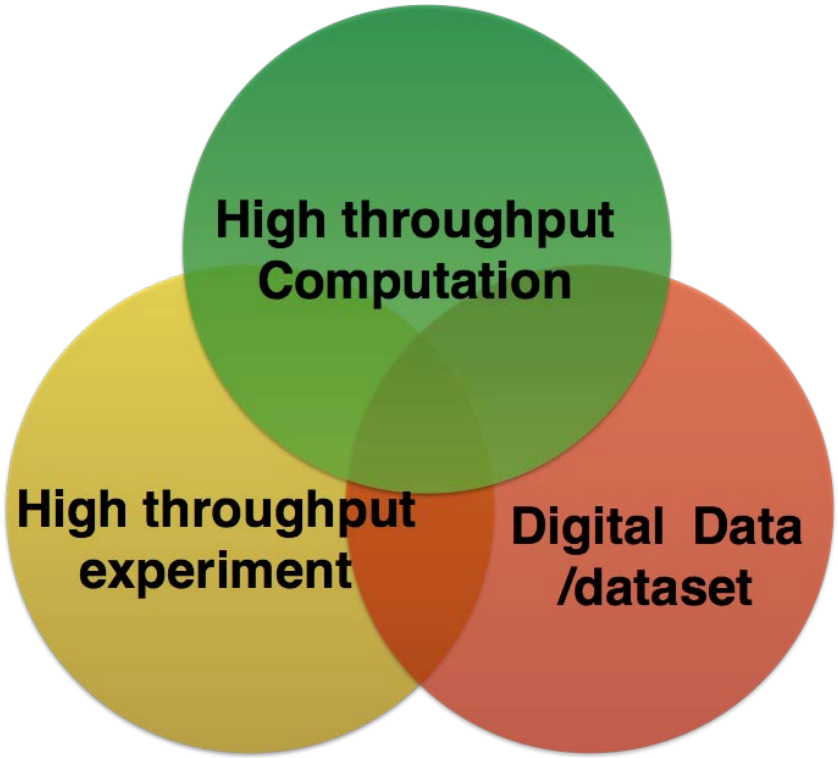
  A Culture Shift in Materials Research, Development, and Deployment

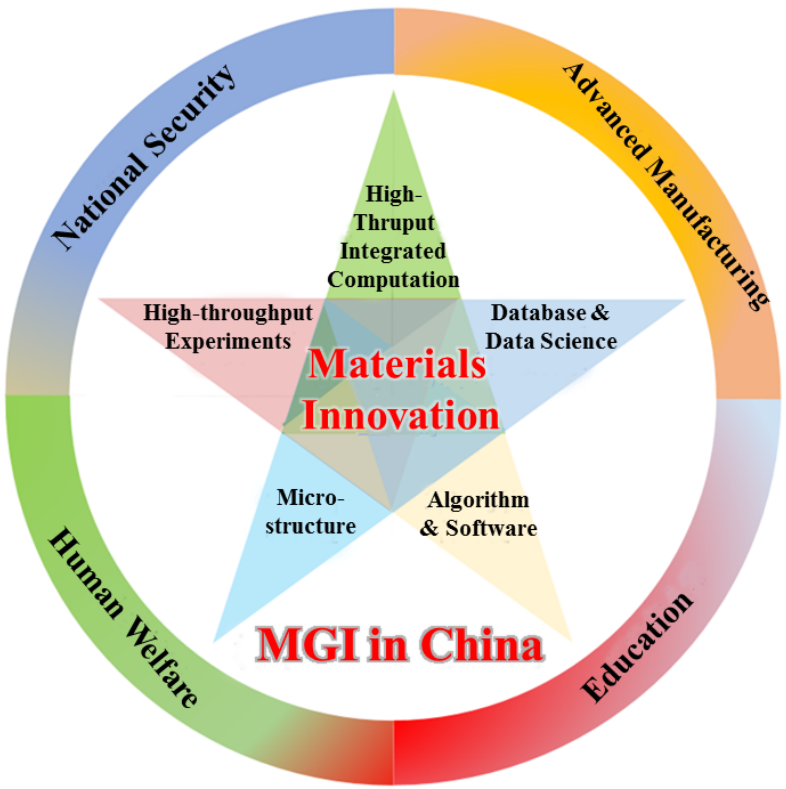
- **Integration of Experiments, Computation, and Theory**

- **Facilitate Access to Materials Data**

- **Equip the Next-Generation Materials Workforce**

MGI & Materials Informatics

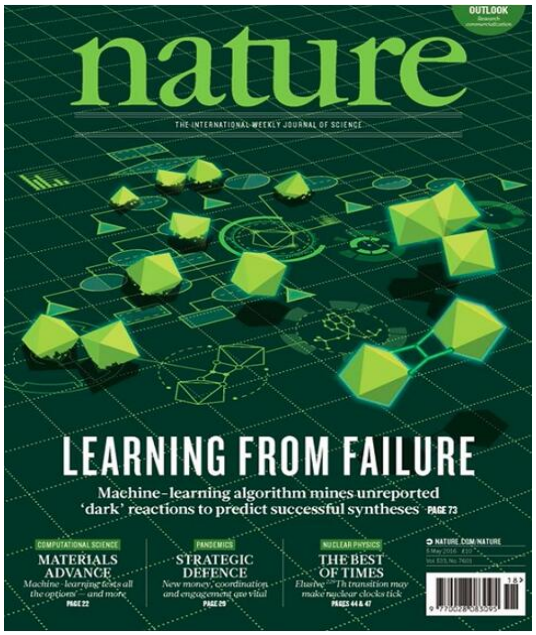




Materials Informatics employs techniques, tools, and theories drawn from the emerging fields of data science, internet, computer science and engineering, and digital technologies to the materials science and engineering to accelerate materials, products and manufacturing innovations.

Artificial intelligence

Fast development of data science, artificial intelligence, machine learning, and materials informatics



ARTICLE

doi:10.1038/nature24270

Mastering the game of Go without human knowledge

David Silver[1]*, Julian Schrittwieser[1]*, Karen Simonyan[1]*, Ioannis Antonoglou[1], Aja Huang[1], Arthur Guez[1], Thomas Hubert[1], Lucas Baker[1], Matthew Lai[1], Adrian Bolton[1], Yutian Chen[1], Timothy Lillicrap[1], Fan Hui[1], Laurent Sifre[1], George van den Driessche[1], Thore Graepel[1] & Demis Hassabis[1]

Nature 542 (2017), 115
doi:10.1038/nature21056

Nature 533 (2016), 73
doi:10.1038/nature17439

MACHINE LEARNING

New tool in the box

A recent burst of activity in applying machine learning to tackle fundamental questions in physics suggests that associated techniques may soon become as common in physics as numerical simulations or calculus.

Lenka Zdeborová

nature physics

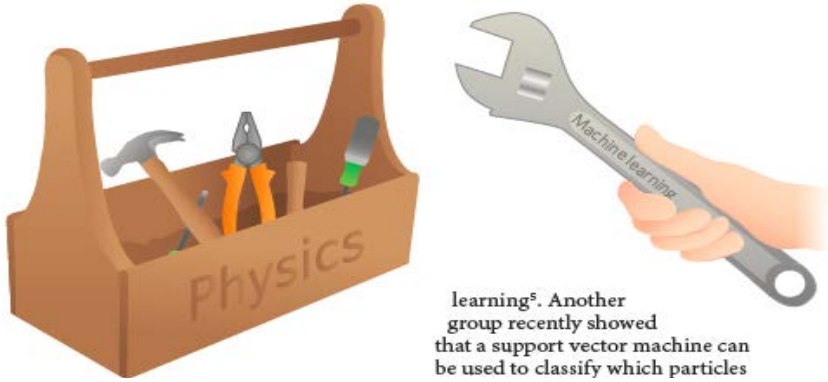
LETTERS

PUBLISHED ONLINE: 13 FEBRUARY 2017 | DOI: 10.1038/NPHYS4035

Machine learning phases of matter

Juan Carrasquilla[1]* and Roger G. Melko[1,2]

learning[2]. Despite this curse, the machine learning community has developed techniques with remarkable abilities to recognize, classify, and characterize complex sets of data. Here, we show that modern machine learning architectures, such as fully connected and convolutional neural networks[3], can identify phases and phase transitions in a variety of condensed-matter Hamiltonians. Readily programmable through modern software libraries[4,5], neural networks can be trained to detect multiple types of order parameter, as well as highly non-trivial states with no conventional order, directly from raw state configurations sampled with Monte Carlo[6,7].



learning[5]. Another group recently showed that a support vector machine can be used to classify which particles

nature physics

LETTERS

PUBLISHED ONLINE: 13 FEBRUARY 2017 | DOI: 10.1038/NPHYS4037

Learning phase transitions by confusion

Evert P. L. van Nieuwenburg*, Ye-Hua Liu and Sebastian D. Huber

large Hilbert space. With modern computing power and access to ever-larger data sets, classification problems are now routinely solved using machine-learning techniques[1]. Here, we propose a neural-network approach to finding phase transitions, based on the performance of a neural network after it is trained with data that are deliberately labelled incorrectly. We demonstrate the success of this method on the topological phase transition in the Kitaev chain[2], the thermal phase transition in the classical Ising model[3], and the many-body-localization transition in a disordered quantum spin chain[4]. Our method does not depend on order parameters, knowledge of the topological content of the phases, or any other specifics of the transition at hand. It therefore paves the way to the development of a generic tool for identifying unexplored phase transitions.

Shanghai University
上海大学材料基因组工程研究院

APPLIED MATHEMATICS

Data-driven discovery of partial differential equations

Samuel H. Rudy,[1]* Steven L. Brunton,[2] Joshua L. Proctor,[3] J. Nathan Kutz[1]

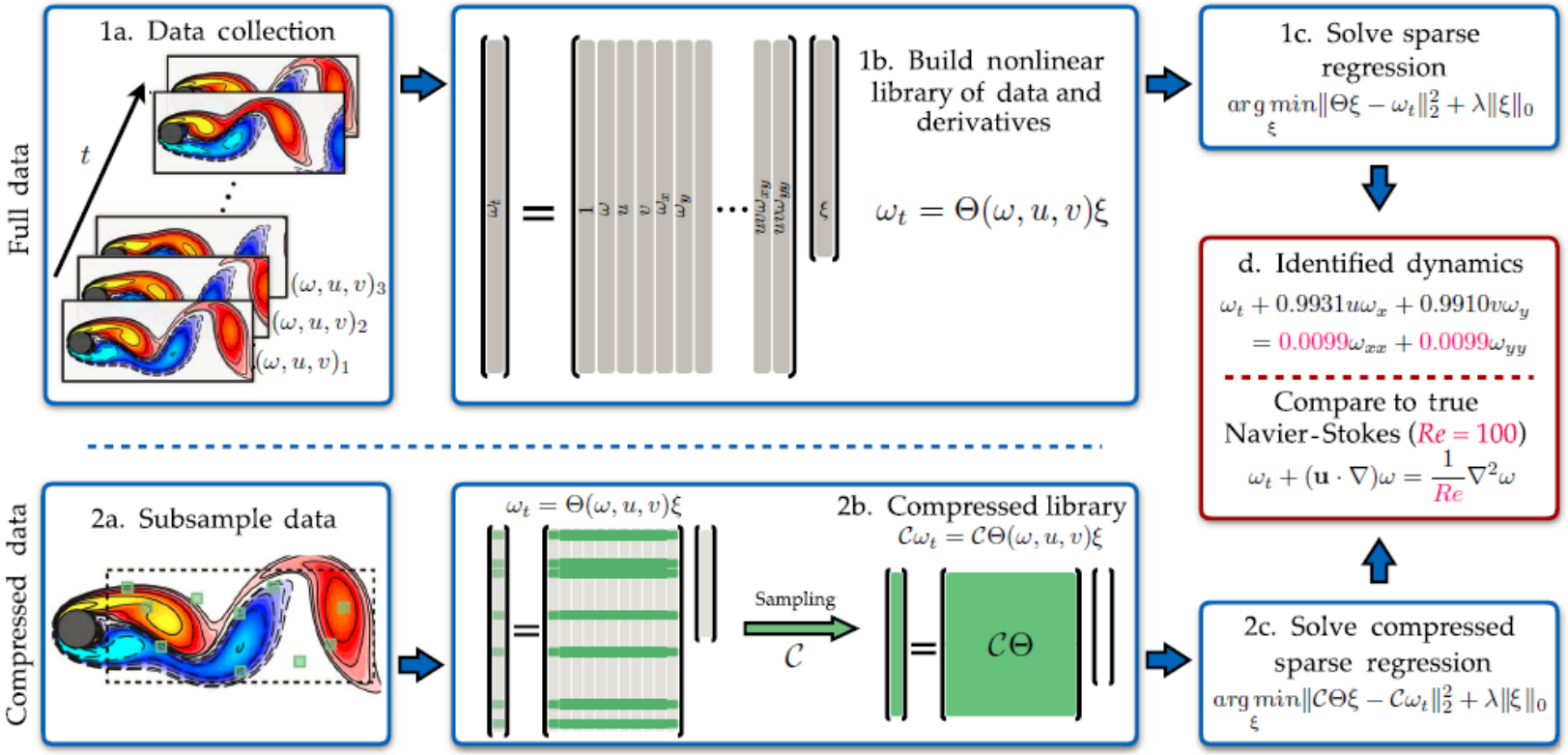


Fig. 1. Steps in the PDE functional identification of nonlinear dynamics (PDE-FIND) algorithm, applied to infer the Navier-Stokes equations from data. (1a) Data are collected as snapshots of a solution to a PDE. (1b) Numerical derivatives are taken, and data are compiled into a large matrix Θ, incorporating candidate terms for the PDE. (1c) Sparse regressions are used to identify active terms in the PDE. (2a) For large data sets, sparse sampling may be used to reduce the size of the problem. (2b) Subsampling the data set is equivalent to taking a subset of rows from the linear system in Eq. 2. (2c) An identical sparse regression problem is formed but with fewer rows. (d) Active terms in ξ are synthesized into a PDE.

Citation report for 1,977 results from All Databases between 1950 and 2018 Go

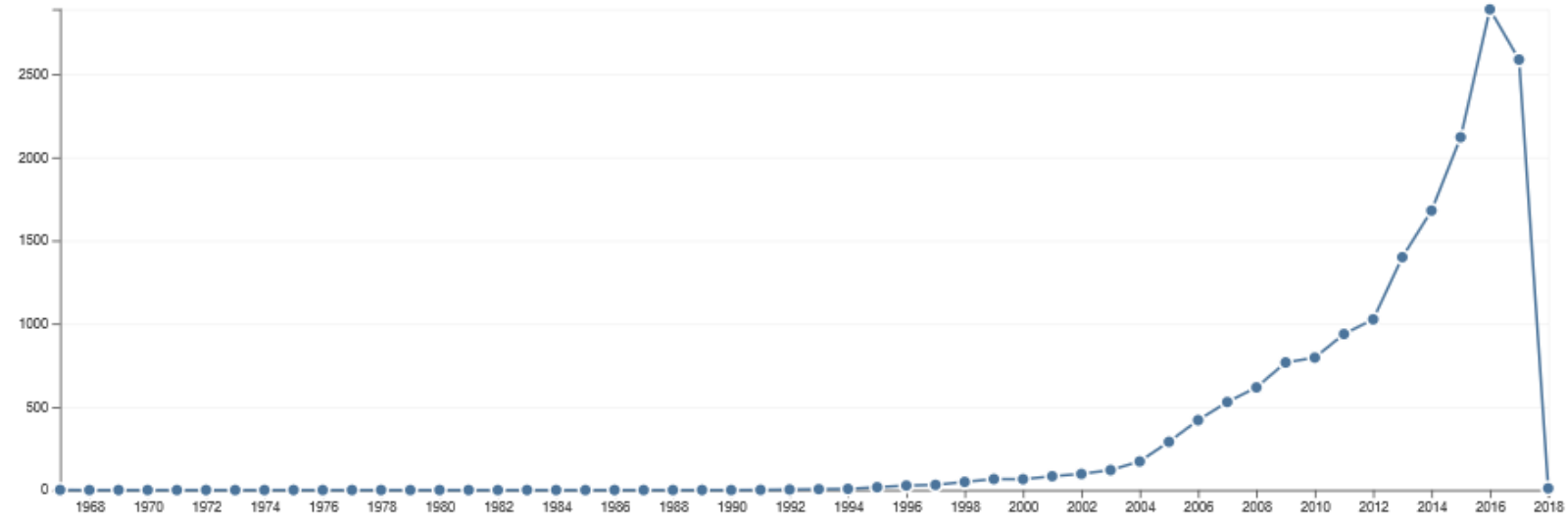
You searched for: TOPIC: (materials informatics) ...More

This report reflects citations to source items indexed within All Databases.

Export Data: Save to Text File

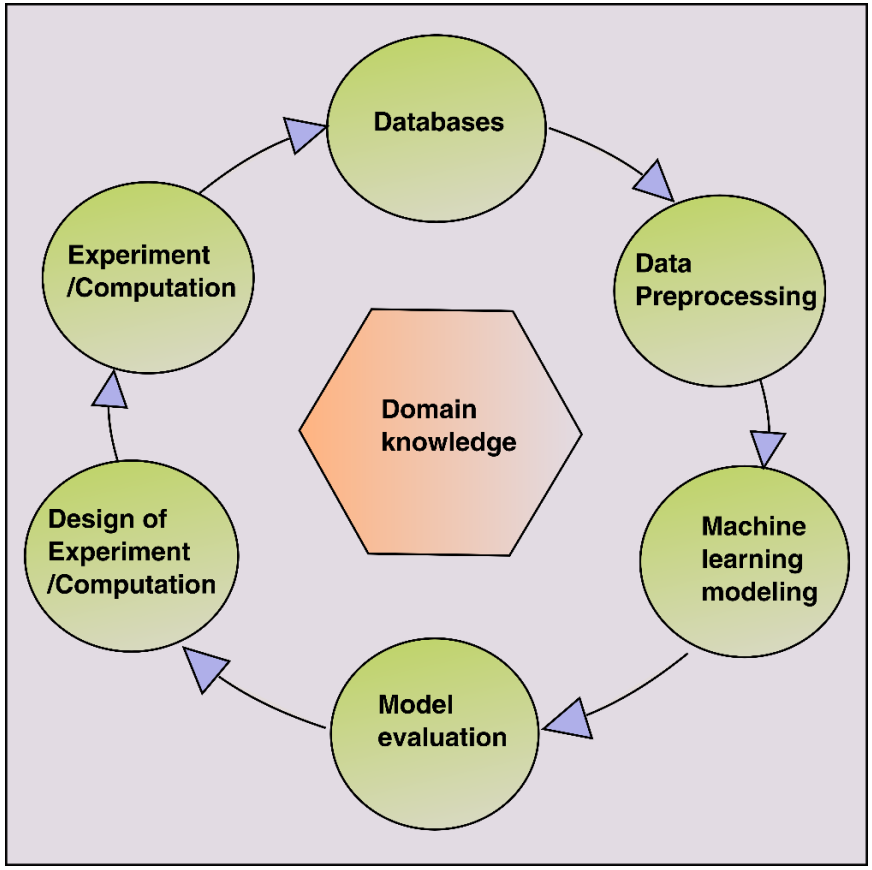
Total Publications	*h*-index	Sum of Times Cited	Citing articles
1,977	55	16,816	15,246
1998 – 2017	Average citations per item: 8.51	Without self citations: 16,152	Without self citations: 14,946

Sum of Times Cited per Year



Materials Informatics

Materials informatics integrates artificial intelligence and materials science and engineering.



Domain knowledge means the knowledge of materials science and engineering, which is the hub of materials informatics.

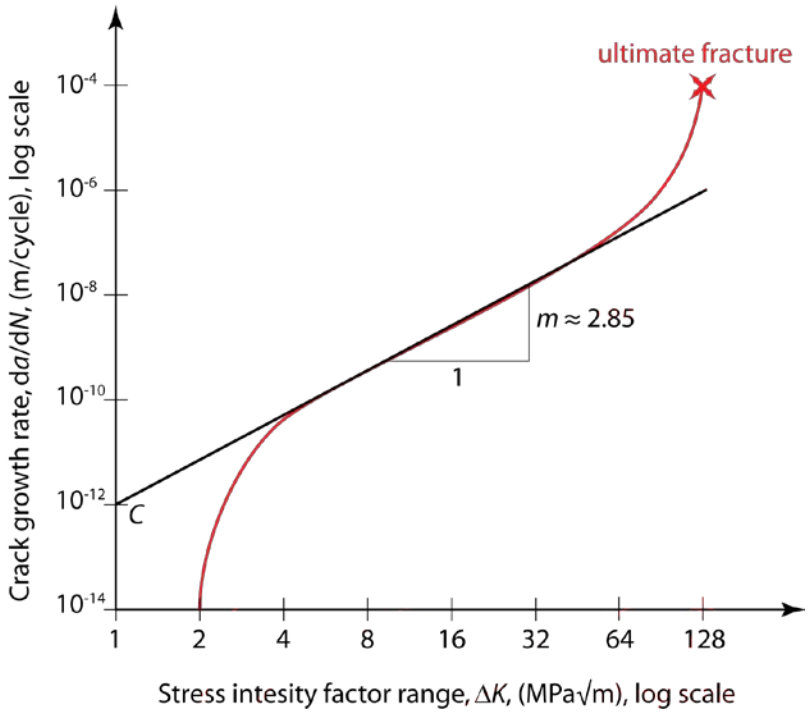
Knowledge from data

- To understand why things happened (explanation)
- To predict what will happen (prediction)
- To manage things that are happening (control)

Leaning can be done by individual human being, especially genius, or/and by machine.

The development of Natural Science is the continuous and endless progress that human being observe nature behaviors, develop learning models, and then gain knowledges.

Paris' law for fatigue crack growth



$$\frac{da}{dN} = C\Delta K^m$$

$$N_f = \frac{2\left(a_c^{\frac{2-m}{2}} - a_i^{\frac{2-m}{2}}\right)}{(2-m)\,C(\Delta\sigma Y\sqrt{\pi})^m}$$

Paris' law is widely used in the prediction of fatigue life.

P.C. Paris, M.P. Gomez, and W.E. Anderson. A rational analytic theory of fatigue. *The Trend in Engineering*, 1961, **13**: p. 9-14.

Paris' law is purely a data-driven formula and no theoretical derivation has been yet published so far.

Hall-Petch Equation for grain size dependent strength

$$\sigma_y = \sigma_0 + \frac{k_y}{\sqrt{d}}$$

How to determine the form of $1/\sqrt{d}$

where σ_y is the yield stress, σ_o is a materials constant for the starting stress for dislocation movement (or the resistance of the lattice to dislocation motion), k_y is the strengthening coefficient (a constant specific to each material), and d is the average grain diameter.

In the early 1950s two groundbreaking series of papers were written independently on the relationship between grain boundaries and strength.

In 1951, E. O. Hall wrote three papers which appeared in volume 64 of the Proceedings of the Physical Society. In his third paper, Hall[6] showed that the length of slip bands or crack lengths correspond to grain sizes and thus a relationship could be established between the two. Hall concentrated on the yielding properties of mild steels.

Based on his experimental work carried out in 1946–1949, N. J. Petch published a paper in 1953 independent from Hall's. Petch's paper[7] concentrated more on brittle fracture. By measuring the variation in cleavage strength with respect to ferritic grain size at very low temperatures, Petch found a relationship exact to that of Hall's. Thus this important relationship is named after both Hall and Petch.

Hume-Rothery's Rules for Solid Solution

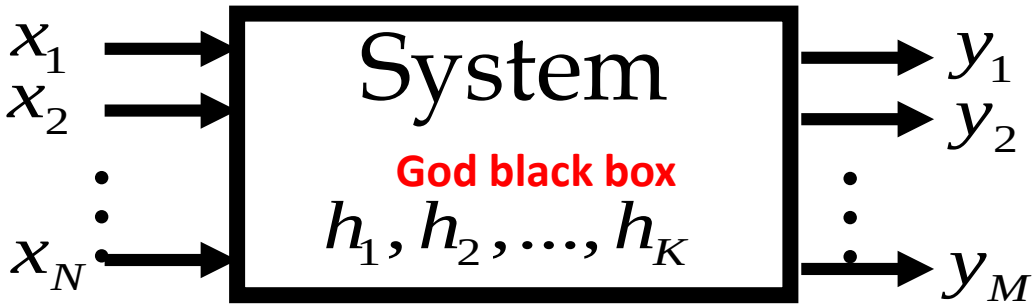
(1) **The atomic size factor. If the atomic diameters of the solute and solvent differ by more than 14%, the solubility is likely to be restricted because the lattice distortion is too great for substitutional solubility.**

(2) **The electrochemical factor. Strongly electropositive components are more likely to form compounds with electronegative components than to form solid solutions.**

(3) **The relative valency factor: Other factors being equal, a lower valence metal is more likely to dissolve one of higher valency than vice versa, i.e., the tendency for two metals to form solid solutions is not necessarily reciprocal. This has been found to be valid mainly for alloys of copper, silver or gold combined with metals of higher valency.**

Hume-Rothery W, Mabbott GW, Channel-Evans KM. Philos Trans Soc A 1934;233:1.

Machining leaning



Input Variables: $\mathbf{x} = (x_1, x_2, \ldots, x_N)$

Hidden Variables: $\mathbf{h} = (h_1, h_2, \ldots, h_K)$

Output Variables: $\mathbf{y} = (y_1, y_2, \ldots, y_K)$

$$y = f(x) + \varepsilon$$

There is an intrinsic error ε0

Stock market: $\frac{\varepsilon}{f(x)} \approx 50\% \text{ -- } 200\%$

Well controlled experiments or computations $\frac{\varepsilon}{f(x)} \approx 0$

Error analysis

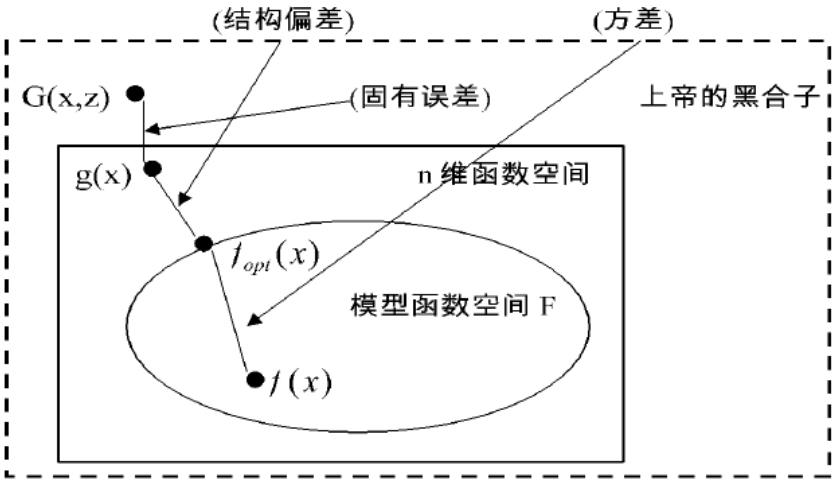
$$f_{opt}(x) = \arg\min_{f \in F} \int_{R^n} \| g(x) - f(x) \| dx$$

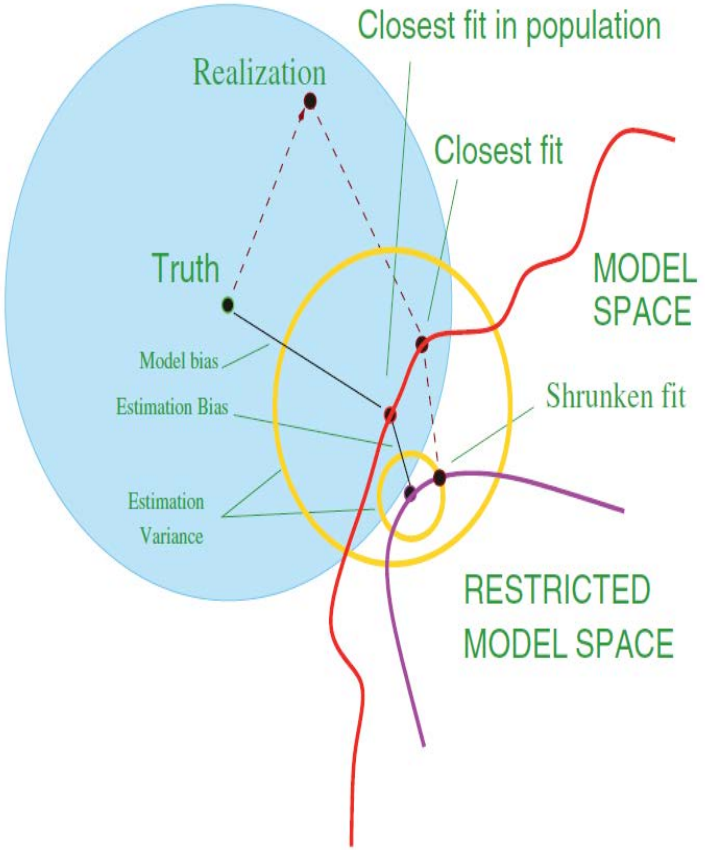
g(x) is an approximate function obtained from machine learning.

The error in machine learning contains two parts

$$\varepsilon_1(f) = \int_{R^n} | g(x) - f(x) | dx$$

$$\leq \underbrace{\int_{R^n} | g(x) - f_{opt}(x) | dx}_{bias} + \underbrace{\int_{R^n} | f_{opt}(x) - f(x) | dx}_{variance}$$

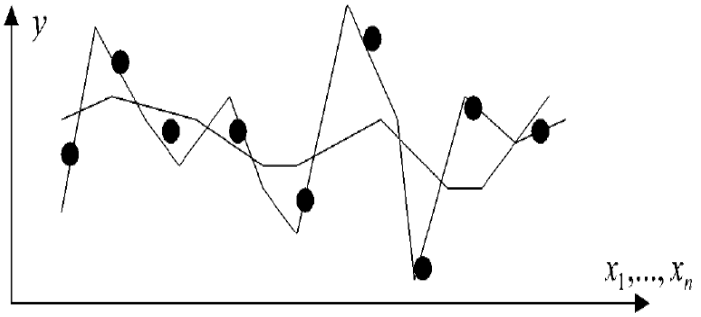




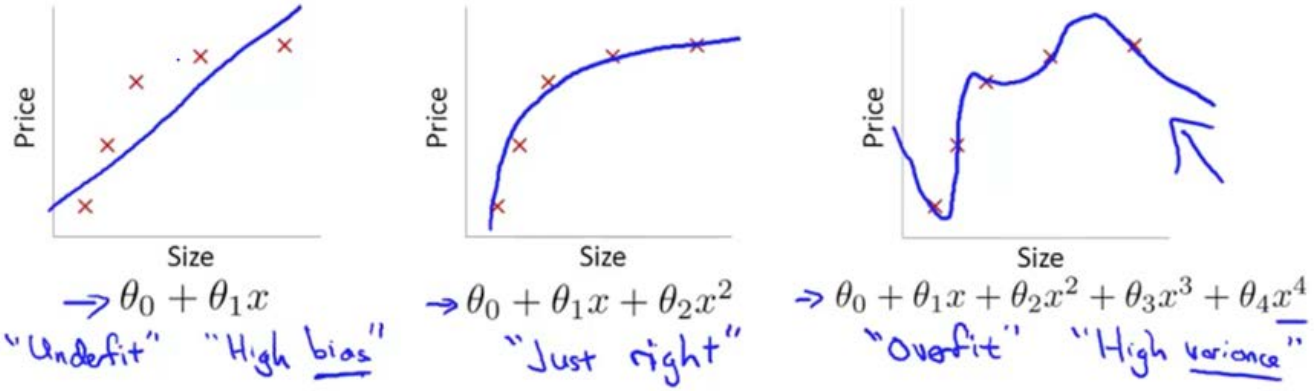
Error = intrinsic error + bias + variance

Difficulties in Machining Leaning

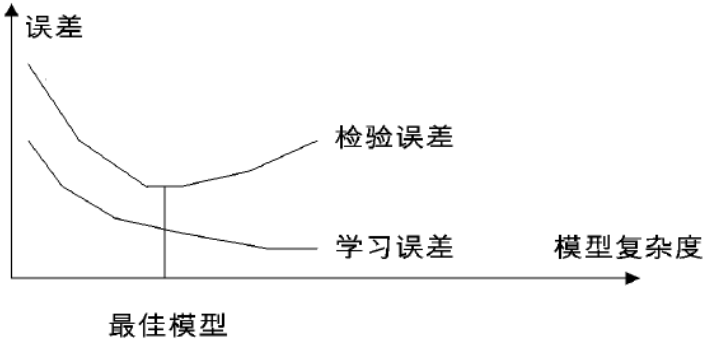
The regressed expression for the continuous function $y = f(x), x \in R^n$, is **not unique** , due to the fact that the sampling data of y and $x_1, ..., x_n$ *are* finite and discrete.

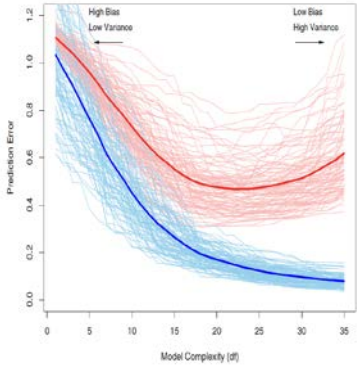


Example: Linear regression (housing prices)



A data set is usually divided into two subsets: training set and testing set.





Features (Descriptors, Attributes)

Two critical issues in machine learning:

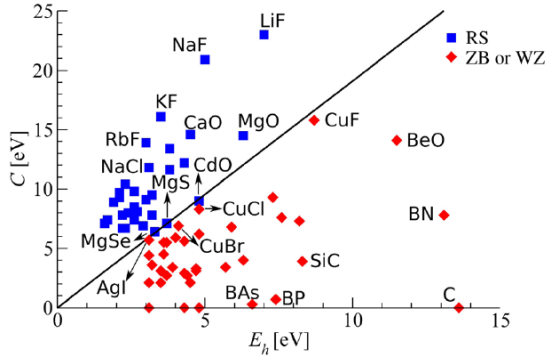
Feature selection and Search space selection

PRL 114, 105503 (2015) PHYSICAL REVIEW LETTERS week ending 13 MARCH 2015

Big Data of Materials Science: Critical Role of the Descriptor

Luca M. Ghiringhelli,[1,*] Jan Vybiral,[2] Sergey V. Levchenko,[1] Claudia Draxl,[3] and Matthias Scheffler[1]

- **In mechanism-based analysis of data, descriptors are well defined and targeted properties are explicitly or implicitly expressed in terms of descriptors, and the expressions appear in analytic form or others.**
- **Descriptors are called variables in mechanism-based formulation and modeling, where it is very much straightforward to identify whether these variables are independent variables or not.**
- **In mechanism-independent machine leaning, the situation is completely different from the mechanism-based data analysis, especially in the machine learning of first-principles calculation data.**



RS: rocksalt
ZB: zinc blende
WZ: wurtzite

The original classification.
Eh: the real part of a complex energy gap
C: the imaginary part of the complex energy gap

The descriptor was started with atomic features. The initial atomic features were combined in physical sense manner to form sums and absolute differences of homogeneous quantities. The combined features were combined further, which finally formed about **4500** feature candidates. Then, **lasso (least absolute shrinkage and selection operator)** was used to select the best 1D, 2D and 3D features.

$$\frac{\mathrm{IP}(B)-\mathrm{EA}(B)}{r_p(A)^2}, \quad \frac{|r_s(A)-r_p(B)|}{\exp[r_s(A)]}, \quad \frac{|r_p(B)-r_s(B)|}{\exp[r_d(A)]}.$$

PRL 114, 105503 (2015) PHYSICAL REVIEW LETTERS week ending 13 MARCH 2015

Big Data of Materials Science: Critical Role of the Descriptor

Luca M. Ghiringhelli,[1,*] Jan Vybiral,[2] Sergey V. Levchenko,[1] Claudia Draxl,[3] and Matthias Scheffler[1]

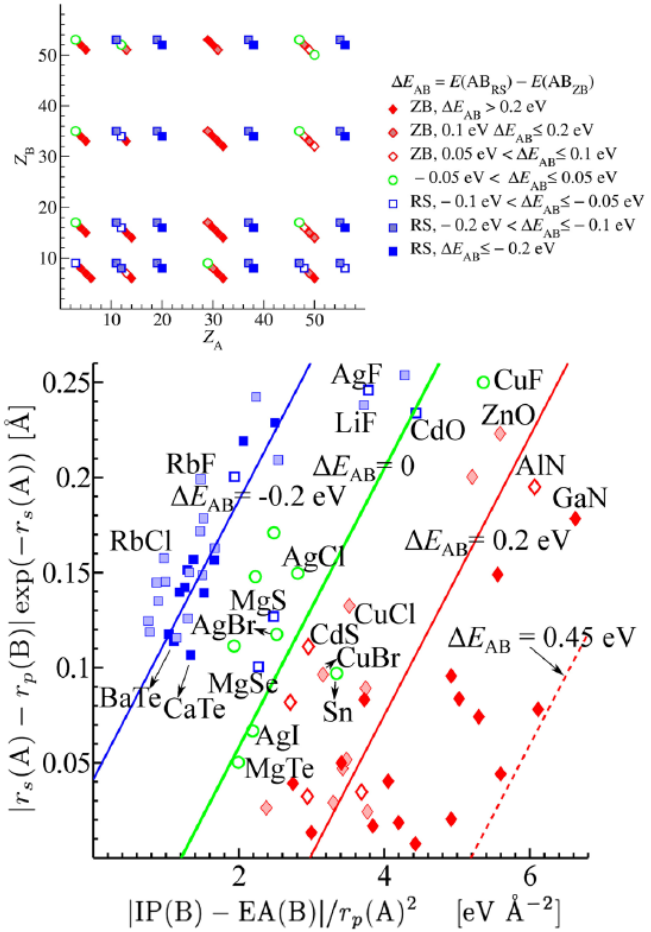


FIG. 2 (color online). Calculated energy differences between RS and ZB structures of the 82 octet binary AB materials, arranged by using the nuclear numbers (Z_A, Z_B) as descriptor (top) and according to our optimal two-dimensional descriptor (bottom). In the bottom panel, seven ZB materials with predicted $\Delta E_{AB} > 0.5$ eV are outside the shown window (see Supplemental Material [6]).

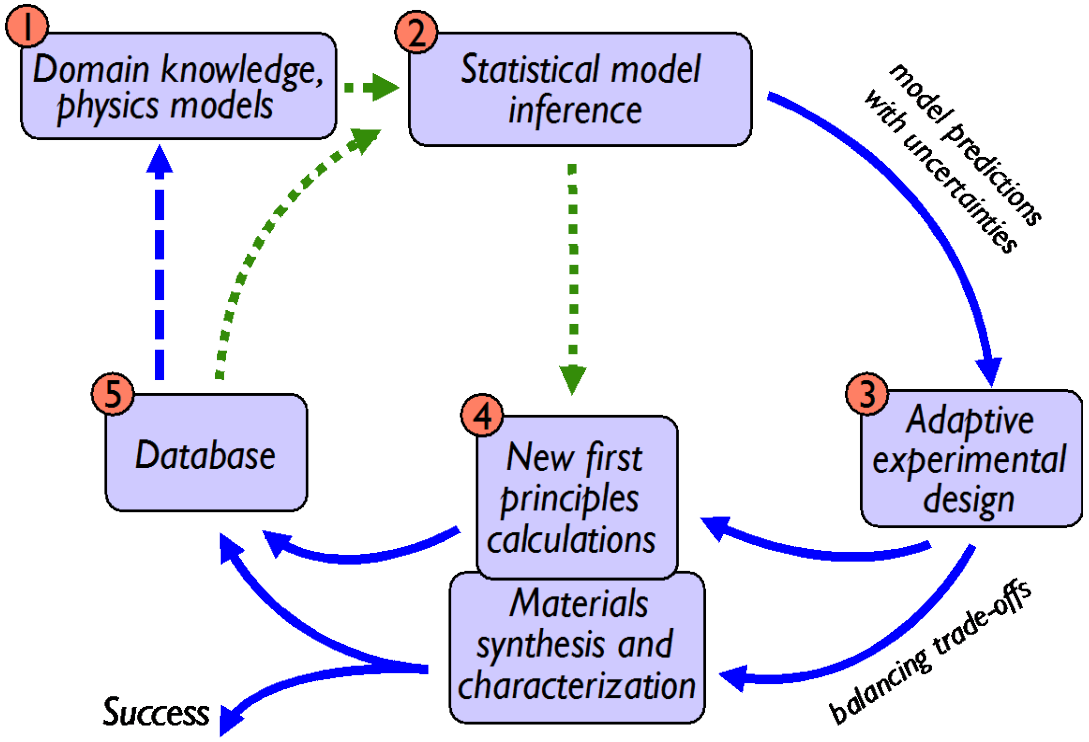
ARTICLE

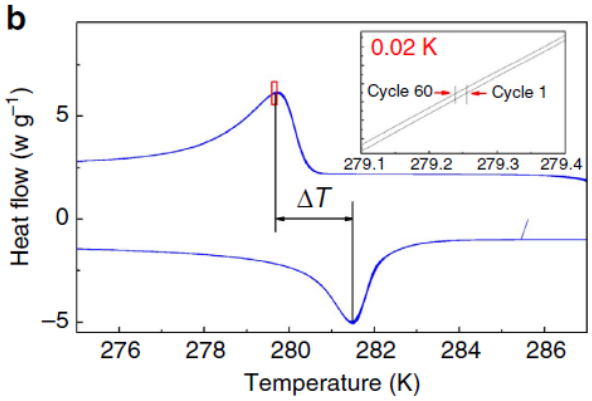
Received 2 Sep 2015 | Accepted 4 Mar 2016 | Published 15 Apr 2016

DOI: 10.1038/ncomms11241 OPEN

Accelerated search for materials with targeted properties by adaptive design

Dezhen Xue[1,2], Prasanna V. Balachandran[1], John Hogden[3], James Theiler[4], Deqing Xue[2] & Turab Lookman[1]



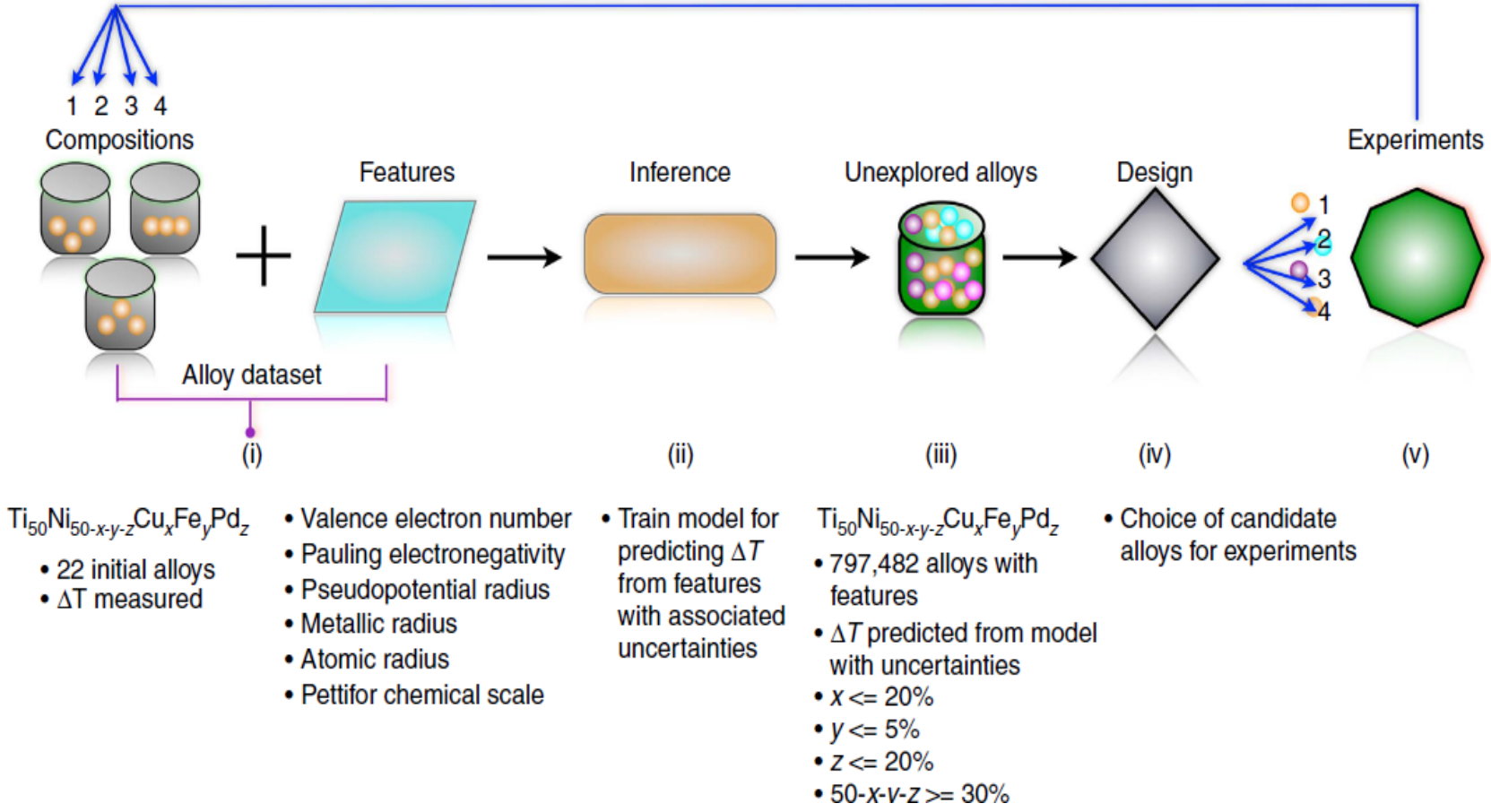


Differential Scanning Calorimetry (DSC)

Design alloy with the smallest ΔT

An adaptive design loop integrating data mining, design, and experimental feedback.

Adaptive design



Regressor: Data mining

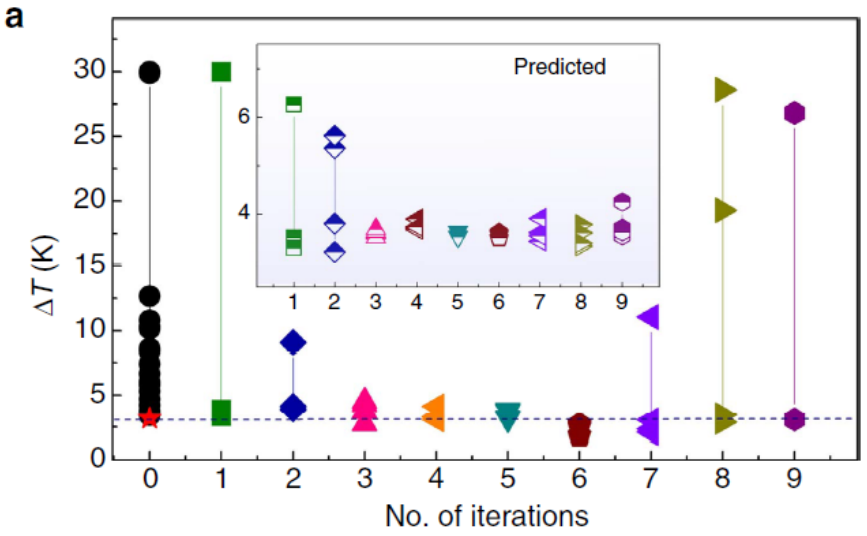
Gaussian Process Model (GPM), which gives means and uncertainties naturally, and Support Vector Regressions with a radial basis function kernel (SVR_{rbf}) and with a linear kernel (SVR_{lin}).

Selector: Experiment design

Min, efficient global optimization (EGO), and Knowledge Gradient (KG)

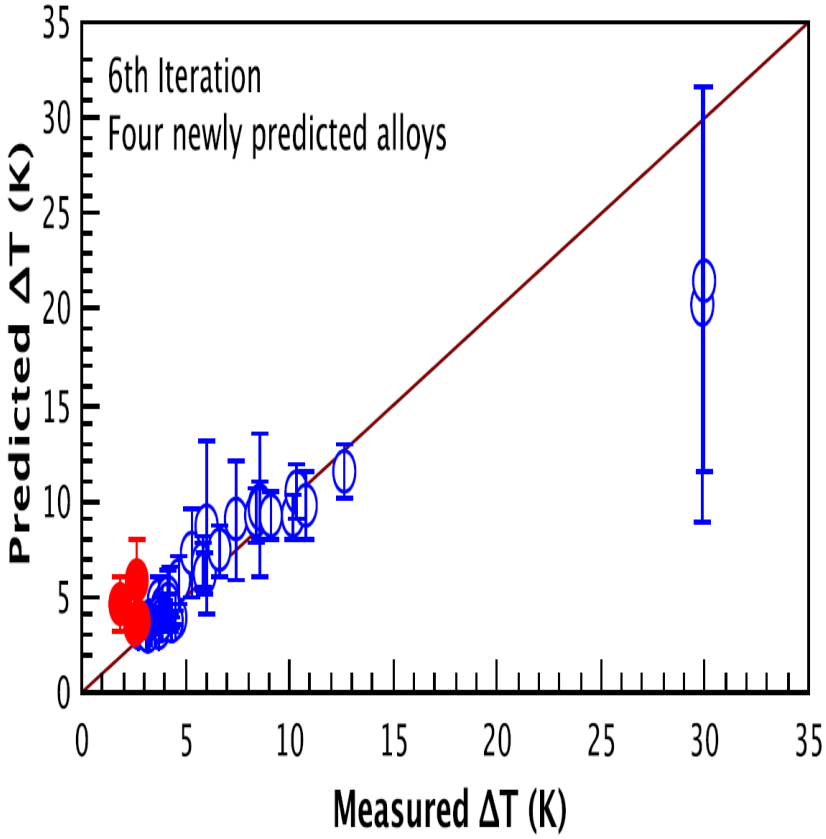
Regressor and selector combination

The lowest ΔT appears in 6th iteration



The experimental measurements for thermal hysteresis ΔT as a function of the number of iterations compared with the predictions (inset).

The blue circles show the estimated vs. actual ΔT values in the training set. The red solid points compare the predicted and experimentally measured ΔT values after each iteration.



Machine learning Xue et al.'s data

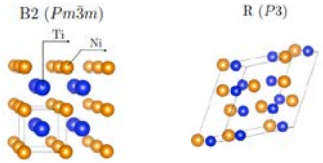
ARTICLE
Received 2 Sep 2015 | Accepted 4 Mar 2016 | Published 15 Apr 2016
DOI: 10.1038/ncomms11241 OPEN
Accelerated search for materials with targeted properties by adaptive design
Dezhen Xue[1,2], Prasanna V. Balachandran[1], John Hogden[3], James Theiler[4], Deqing Xue[2] & Turab Lookman[1]

- **Original Dataset**

  58 NiTi-based shape memory alloys

  $Ti_{50}Ni_{50-x-y-z}Cu_xFe_yPd_z$

- **Preliminary data selection by considering only B2-R transformation leads to 53 alloys**



- **Features**

  x(Cu), y(Fe) and z(Pd)

- **Target Value**

  Low thermal hysteresis (ΔT)

Unpublished work

Machine learning Xue et al.'s data

ARTICLE
Received 2 Sep 2015 | Accepted 4 Mar 2016 | Published 15 Apr 2016
DOI: 10.1038/ncomms11241 OPEN

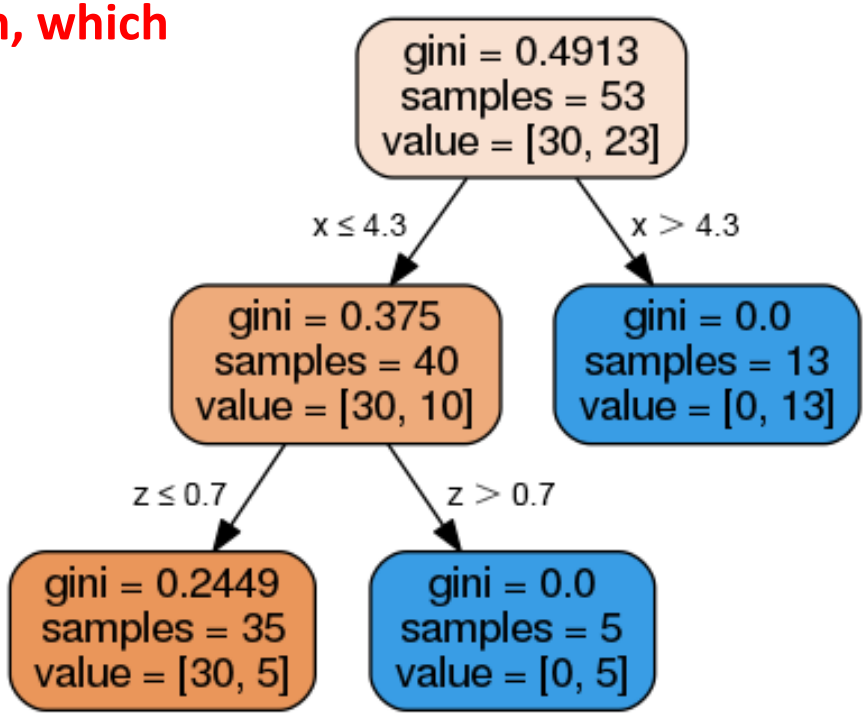
Accelerated search for materials with targeted properties by adaptive design

Dezhen Xue[1,2], Prasanna V. Balachandran[1], John Hogden[3], James Theiler[4], Deqing Xue[2] & Turab Lookman[1]

Decision tree is based on information gain, which partitions the feature space.

Decision Tree Classifier (CART): Class 1: $\Delta T < 4.0$ K

Class 2: $\Delta T \geq 4.0$ K yields a subset of data $\{x \leq 4.3, z \leq 0.7$, and y no restrict$\}$.



The Gini index is decreased from 0.4913 to 0.2449, which indicates the information gain.

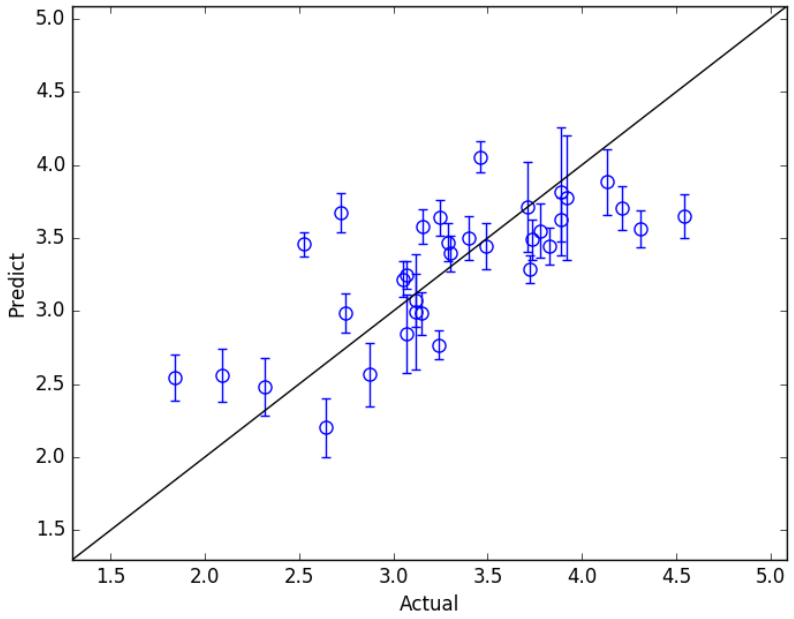
There is an outlier of $Ti_{50}Ni_{48.2}Cu_{0.6}Fe_{0.9}Pd_{0.3}$ with $\Delta T = 11.05$ K, which is removed and the data number is 34 now.

ARTICLE
Received 2 Sep 2015 | Accepted 4 Mar 2016 | Published 15 Apr 2016
DOI: 10.1038/ncomms11241 OPEN

Accelerated search for materials with targeted properties by adaptive design

Dezhen Xue[1,2], Prasanna V. Balachandran[1], John Hogden[3], James Theiler[4], Deqing Xue[2] & Turab Lookman[1]

Bootstrap + SVR (Leave-two-out validation)



Measured ΔT versus predicted mean ΔT with variance

MRE	R	RMSE
11.06%	0.7229	0.4271

ARTICLE
Received 2 Sep 2015 | Accepted 4 Mar 2016 | Published 15 Apr 2016
DOI: 10.1038/ncomms11241 OPEN

Accelerated search for materials with targeted properties by adaptive design

Dezhen Xue[1,2], Prasanna V. Balachandran[1], John Hogden[3], James Theiler[4], Deqing Xue[2] & Turab Lookman[1]

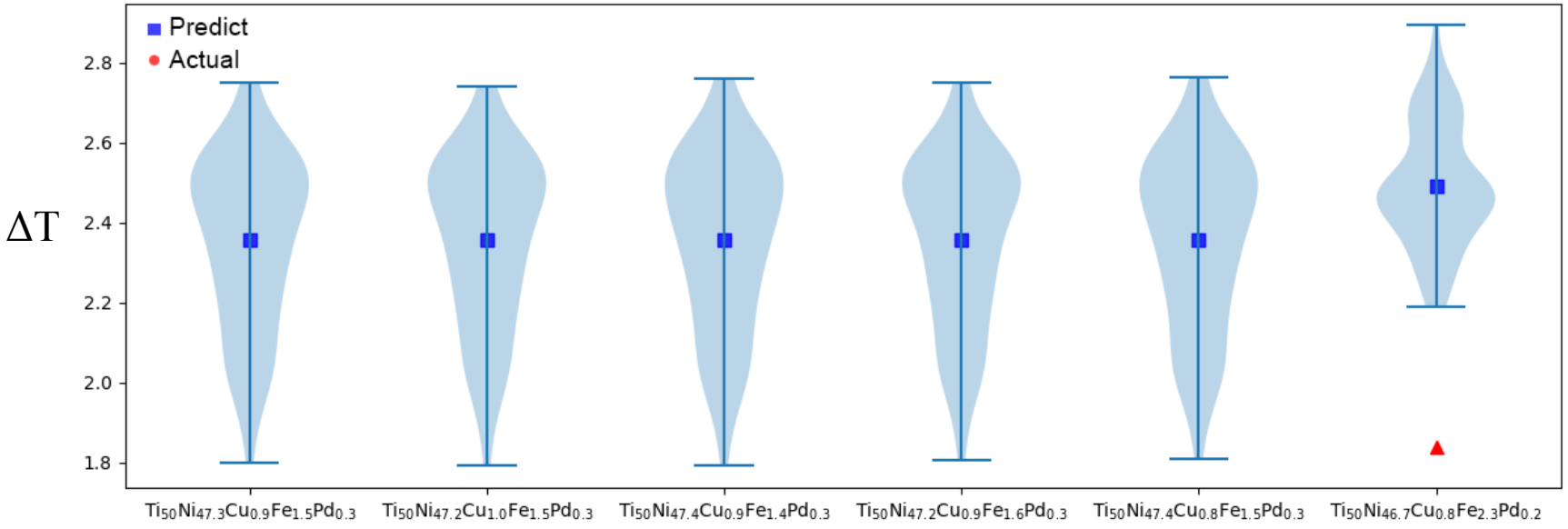
SVR + Bootstrap search in the search space

$$Ti_{50}Ni_{50-x-y-z}Cu_xFe_yPd_z$$

$$\begin{cases} x \le 4.3\% \\ z \le 0.7\% \\ y \le 5.0\% \\ 50 - x - y - z \ge 30\% \end{cases}$$

Total 17952 alloys

Violin plot of the best five predicated alloys



S. Ning et al./Ultramicroscopy 184 (2018) 274–283

Scanning distortion correction in STEM images

Image matching is similar to face recognition, but much simple.

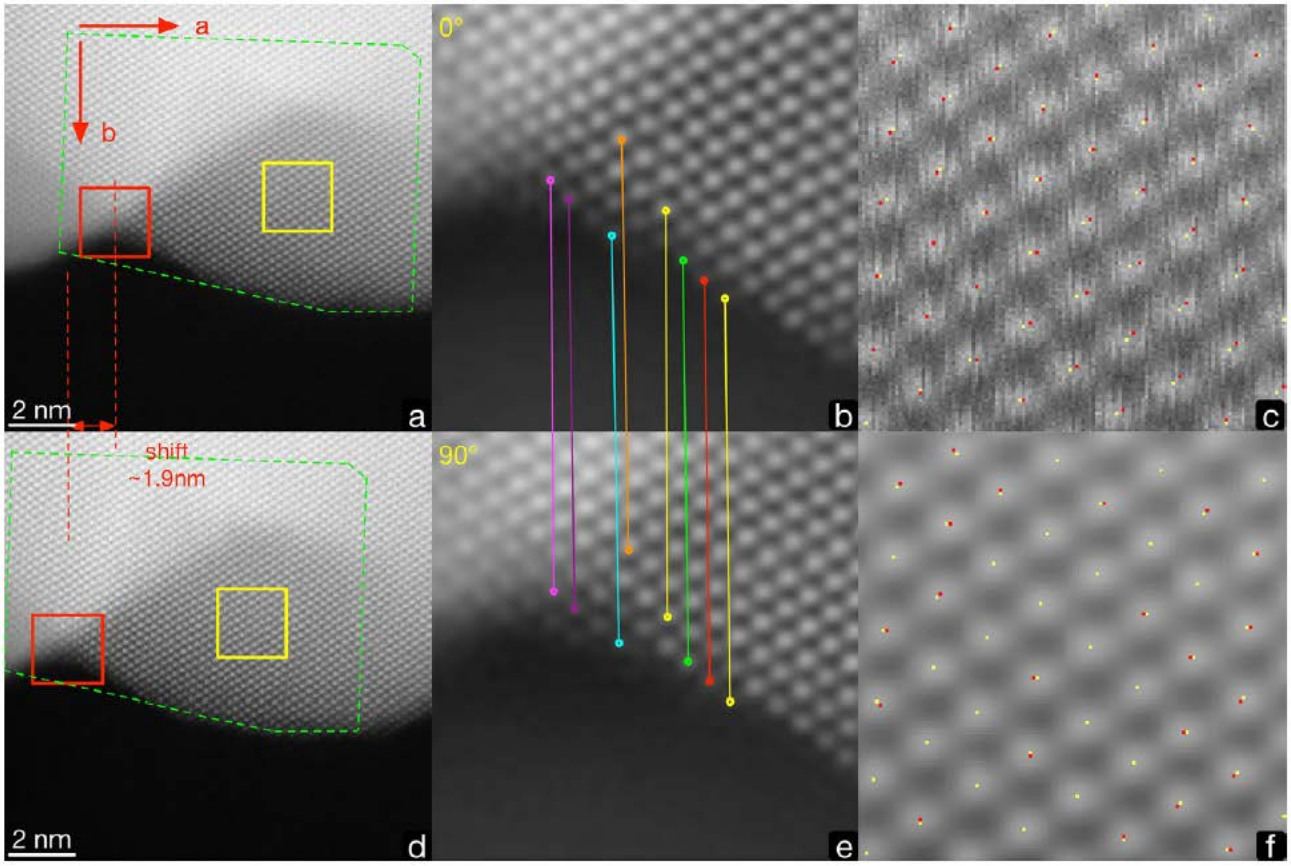


Fig. 5. (a) and (**d**) are two STEM images of a gold sample obtained by scanning the sample from the horizontal and perpendicular directions, i.e., 0° and 90°, respectively, where the red square boxes highlight the feature zone. Comparison of the feature zone between (**d**) and (**a**) indicates that (**d**) image shifts towards left by 1.9 nm. In the present work, the atomic column positions are represented by the intensity centers, see text for detailed description. Some characteristic column positions in the feature zone are used as the starting match points. As an example, (**b**) and (**e**) show the cross correlation match of the feature points in the 0° and 90° images, where each of the color lines links the images of one same real atomic column. Extending the starting matched atomic column positions with the consideration of only atom column positions will result in one-to-one match of some regions in the two images without any corrections of distortions in the positions. The matched regions of the 0° and 90° images are taken from a same area of the probed sample. After that, it is straightforward to find two smaller matched square areas in (**a**) and (**d**), which are highlighted by the yellow boxes in the 0° and 90° images. Then, the 0° squared area is enlarged in (**c**), where a common origin is set on the atomic image at the left bottom of each image, meaning this atomic image is set to be the common reference point. In (**c**), the yellow dots and red dots mark the atomic column positions in the 0° and 90° images, respectively. After the distortion correction, (f) shows the atomic column positions with yellow and red dots for the two correction methods, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Scanning distortion correction in STEM images

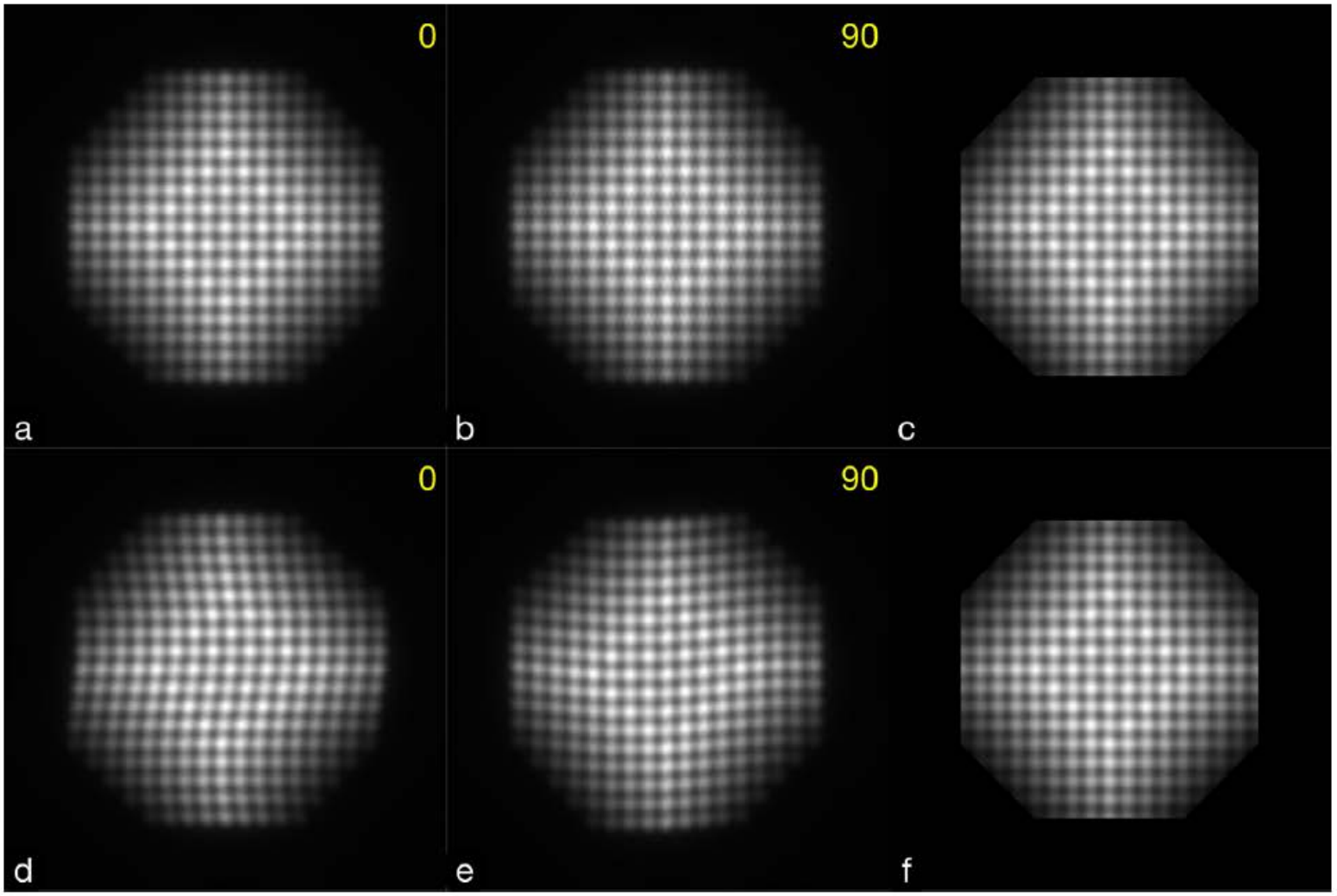


Fig. 7. Simulated HAADF STEM images of the gold nano-particle, where (**a**) and (**d**), and (**b**) and (**e**) are 0° and 90° images, respectively, and (**c**) and (**f**) are corrected images by using the (**a**) (**b**) and (**d**) (**e**) pairs, respectively. The distortion wavelengths in (**a**) and (**b**) are 150 and 230 pixels, respectively, and the amplitudes are both 2 pixels. The distortion wavelengths in (**d**) and (**e**) are 150,000 and 200,000 pixels, respectively, and the amplitudes are both 6 pixels. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Size dependent strength of concrete (unpublished)

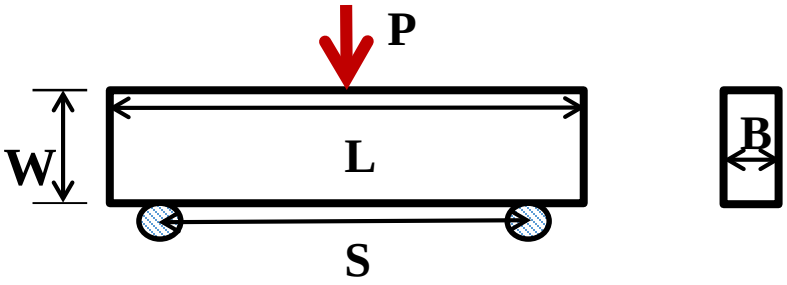
In the literature there is academic argument regarding the size-dependent strength of concrete. Hu et al. published a paper entitled "Comparison of boundary and size effect models based on new development " in Engineering Fracture Mechanics 175 (2017) 146–167.

A B S T R A C T

Hoover, Bazant and colleagues have published a number of papers in recent years (Bazant and Yu, 2009; Yu, 2010; Hoover and Bazant, 2013a, 2013b, 2014) on comparisons between Bazant size effect model (SEM) and Hu-Duan boundary effect model (BEM) for quasi-brittle fracture of concrete. With the recent developments of BEM (Wang et al., 2016; Guan et al., 2016; Wang and Hu, 2017) on irregular and discrete crack growth in concrete shaped by coarse aggregate structures, it is time to clarify issues on the SEM and BEM comparison raised by Bazant and Yu (2009), Yu (2010), Hoover and Bazant (2013a, 2013b, 2014). The experimental results of Hoover and Bazant (2013a, 2013b, 2014) are analyzed again using BEM, and new findings and in-depth understandings that have not been achieved by SEM are presented in this study. BEM is one concise equation, containing only two fundamental material constants, tensile strength f_t and fracture toughness K_{IC}, applicable to both notched and un-notched concrete specimens. Most importantly, BEM explains the inevitable influence of coarse aggregate structures on quasi-brittle fracture of concrete through modeling irregular and discrete crack formations and by considering the critical role of the maximum aggregate d_{max}. In contrast, SEM has three different equations, one for notched, one for un-notched, and one for shallow-notch specimens, containing total 18 empirical parameters to be determined from curve fitting. Despite with the staggering 18 parameters, the three SEM equations still overlook the crucial role of coarse aggregate structures in concrete fracture; d_{max} and discrete crack formation are not considered. After establishing the relation between discrete fictitious crack formation Δa_{fic} and d_{max} at the peak load P_{max} based on four different sets of independently obtained experimental results of concrete and rock with d_{max} from 2 to 10 and 19 mm, BEM becomes a predictive design model which only needs strength f_t and toughness K_{IC}.

Size dependent strength of concrete (unpublished)

There point bending test on concrete beams

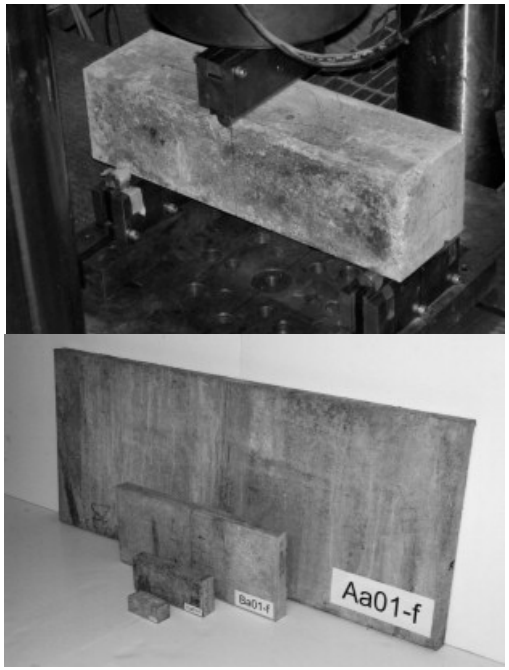


Loading configuration

Max bending stress

$$\sigma_N = \frac{3PS}{2BW^2}$$

L=2.4W S=2.176W B=40mm



W (mm)	Peak stress (MPa)	W (mm)	Peak stress (MPa)	W (mm)	Peak stress (MPa)	W (mm)	Peak stress (MPa)
40	7.17	93	6.74	215	5.99	500	5.68
	7.32		6.95		6.21		5.83
	7.37		7.02		6.23		6.36
	7.56		7.37		6.75		4.92
	7.62		7.4		8.28(Failed)		5.63(Failed)
	8.57		7.69		8.46(Failed)		5.65(Failed)
	8.68		7.78				
			7.86				

Dataset of three point bending concrete beams.[Bazant, 2013]

Size dependent strength of concrete (unpublished)

Weibull distribution

Failure probability

(Equivalent volume V_N)

$$P_f(\sigma_N) = 1 - \exp\left[-\frac{V_N}{V_0}\left(\frac{\sigma_N}{\sigma_0}\right)^m\right]$$

Failure probability density

$$f(\sigma_N) = \frac{\partial P_f}{\partial \sigma_N} = \frac{mV_N}{\sigma_0 V_0}\left(\frac{\sigma_N}{\sigma_0}\right)^{m-1} \exp\left[-\frac{V_N}{V_0}\left(\frac{\sigma_N}{\sigma_0}\right)^m\right]$$

The EM (Expectation Maximization) algorithm, which simplifies difficult maximum likelihood problems, is used to determine m and σ_0 for a given V_0

$$L(m,\sigma_0) = \prod_{i=1}^{n} f(\sigma_i, V_i; m, \sigma_0) = \prod_{i=1}^{n}\left(\frac{mV_i}{\sigma_0 V_0}\right)\left(\frac{\sigma_i}{\sigma_0}\right)^{m-1} \exp\left[-\sum_{i=1}^{n}\left(\frac{V_i}{V_0}\right)\left(\frac{\sigma_i}{\sigma_0}\right)^m\right]$$

$$\ln L(m,\sigma_0) = \sum_{i=1}^{n} \ln\left[\left(\frac{mV_i}{\sigma_0 V_0}\right)\left(\frac{\sigma_i}{\sigma_0}\right)^{m-1}\right] - \sum_{i=1}^{n}\left(\frac{V_i}{V_0}\right)\left(\frac{\sigma_i}{\sigma_0}\right)^m$$

Size dependent strength of concrete (unpublished)

Derivation

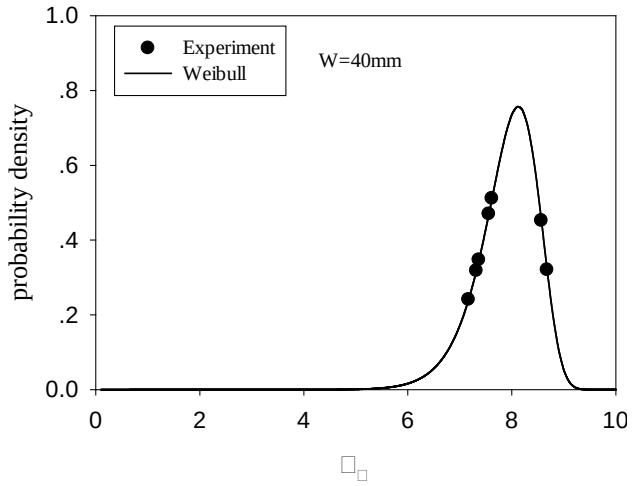
$$\frac{\partial}{\partial m}\ln L(m,\sigma_0)=\sum_{i=1}^{n}\left[\frac{1}{m}+\ln\left(\frac{\sigma_i}{\sigma_0}\right)\right]-\sum_{i=1}^{n}\left[\frac{V_i}{V_0}\left(\frac{\sigma_i}{\sigma_0}\right)^m\ln\frac{\sigma_i}{\sigma_0}\right]$$

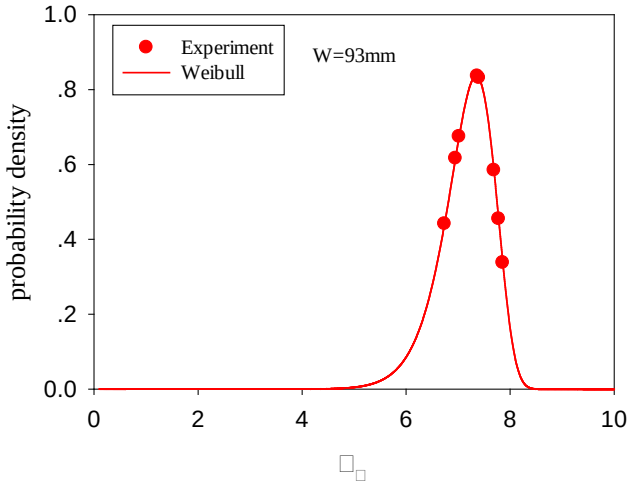
$$\frac{\partial}{\partial\sigma_0}\ln L(m,\sigma_0)=-\frac{nm}{\sigma_0}+\sum_{i=1}^{n}\frac{V_i m}{\sigma_0 V_0}*\left(\frac{\sigma_i}{\sigma_0}\right)^m$$

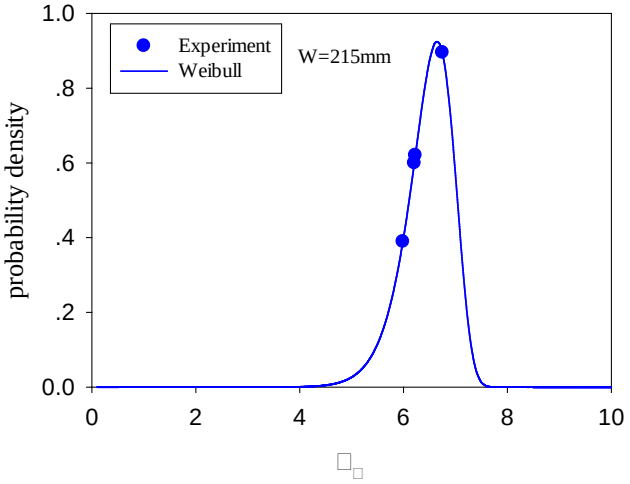
extremum

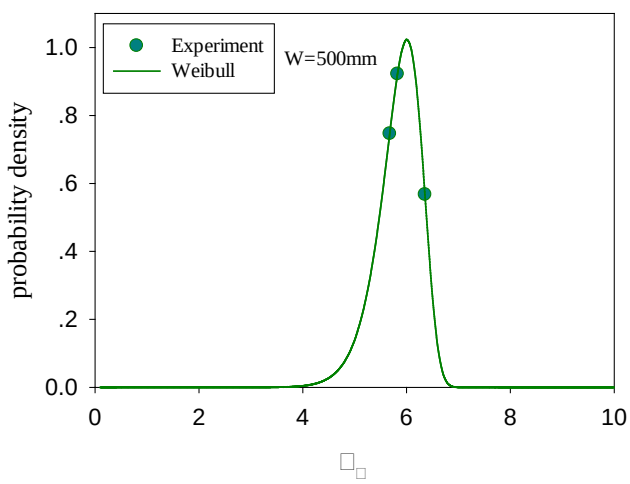
$$\begin{cases}\dfrac{\partial}{\partial m}\ln L(m,\sigma_0)=0\\ \dfrac{\partial}{\partial\sigma_0}\ln L(m,\sigma_0)=0\end{cases}$$

Probability density of different sizes









Size dependent strength of concrete (unpublished)

Survival probability

$$P_s(\sigma_N) = 1 - P_f(\sigma_N)$$

$$= \exp\left[-\frac{V_N}{V_0}\left(\frac{\sigma_N}{\sigma_0}\right)^m\right]$$

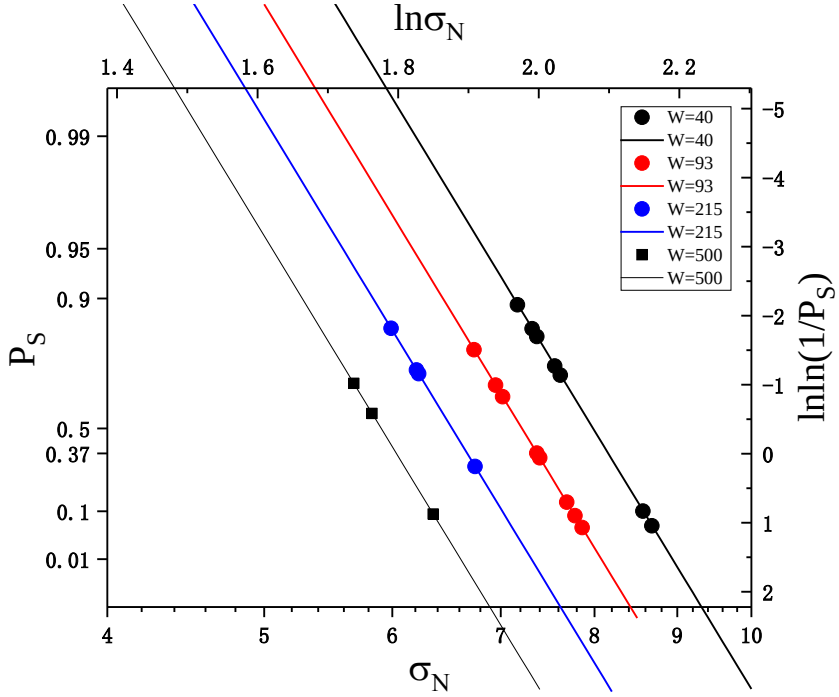
V_0	m	σ_0
V_{40}	16.76	8.16
V_{93}	16.76	7.38
V_{215}	16.76	6.67
V_{500}	16.76	6.03

Taking logarithm twice

$$\ln\ln\frac{1}{P_s(\sigma_N)} = \ln\left[\frac{V_N}{V_0}\left(\frac{\sigma_N}{\sigma_0}\right)^m\right]$$

$$= m\ln\sigma_N + \ln V_N - m\ln\sigma_0 - \ln V_0$$

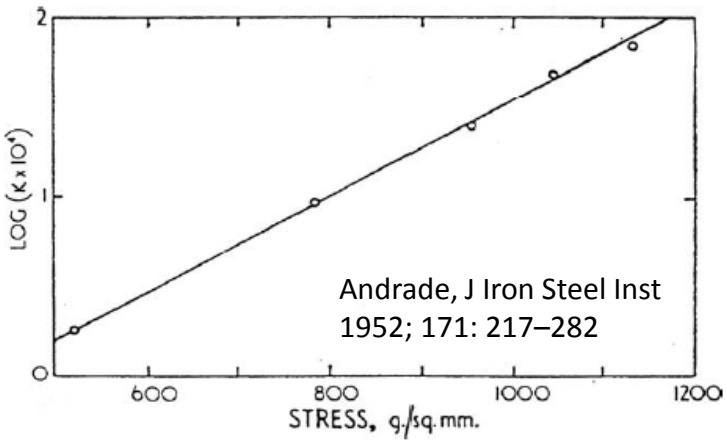
With survival probability of 99%, no failure will occur if the max bending stress is lower than 6.2 MPa for W=40mm, 5.6 MPa for W=93mm, 5.1 MPa for W=215mm, and 4.6 MPa for W=500mm.



The smaller the stronger!

Data driven development of formula for creep and stress relaxation

Experimental data curves of lnσ versus ln $\dot{\varepsilon}$ during the tensile creep tests appear approximately linear and linearly fitting versus determines the value of *m*. Then, integrating the above equation gives a power-law relation of $\dot{\varepsilon} \propto \sigma^n$.



Logarithm of steady-state strain rate as a function of logarithm of stress for copper at 683 K.

❖ **Strain rate sensitivity *m*, reflects the strain rate dependence on stress *σ*, which is defined as:**

$$m = \left.\frac{\partial \log \sigma}{\partial \log \dot{\varepsilon}}\right|_T = \left.\frac{\partial \ln \sigma}{\partial \ln \dot{\varepsilon}}\right|_T$$

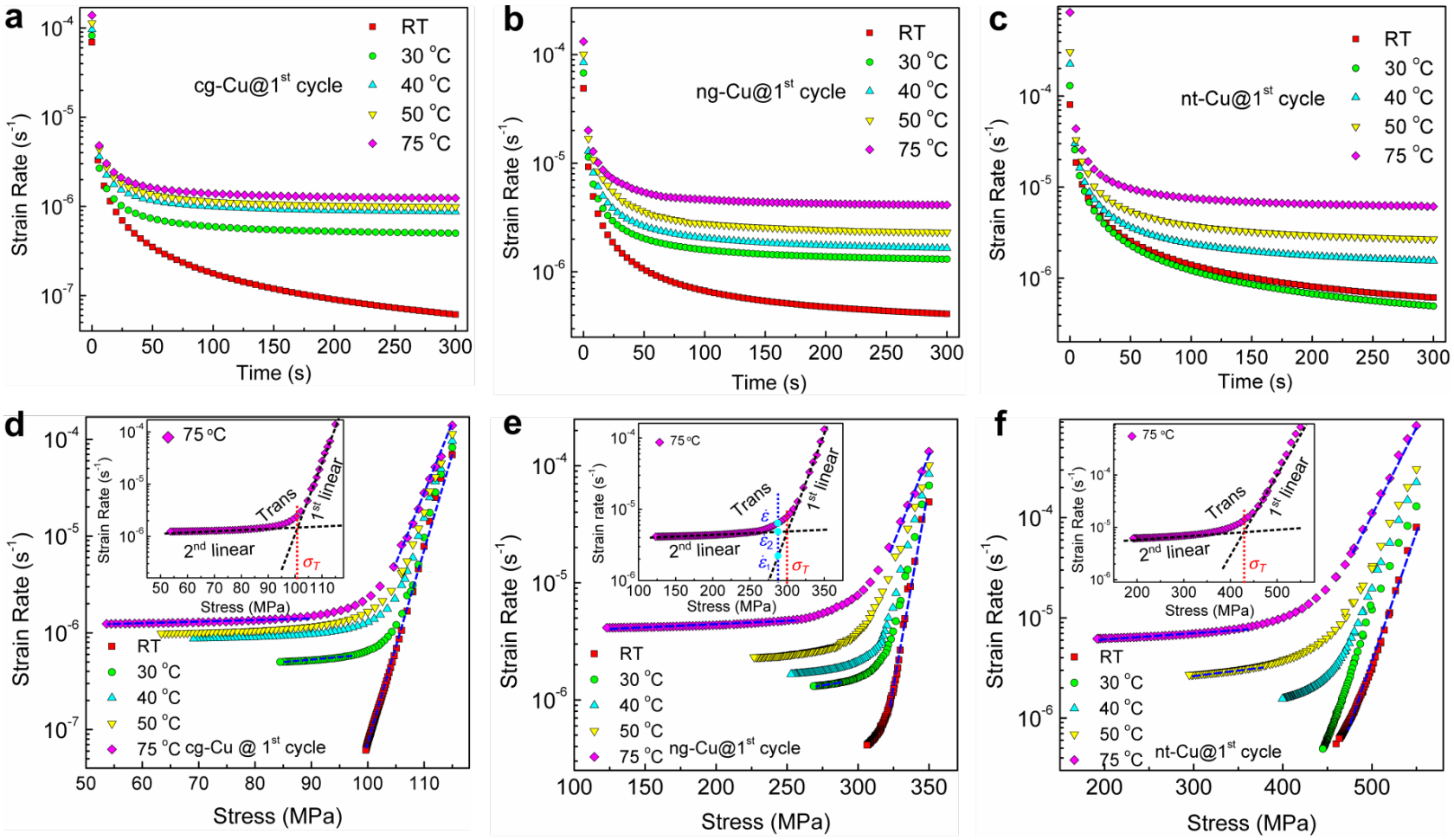
❖ **Stress exponent *n* is based on the power-law relation between stress and strain rate:**

$$n = \left.\frac{\partial \ln \dot{\varepsilon}}{\partial \ln \sigma}\right|_T \qquad \frac{1}{m} = n$$

The temperature effect has not been considered yet.

Yang et al., Acta Mat 2016; JMPS 2016

Two linear stress relaxation regions

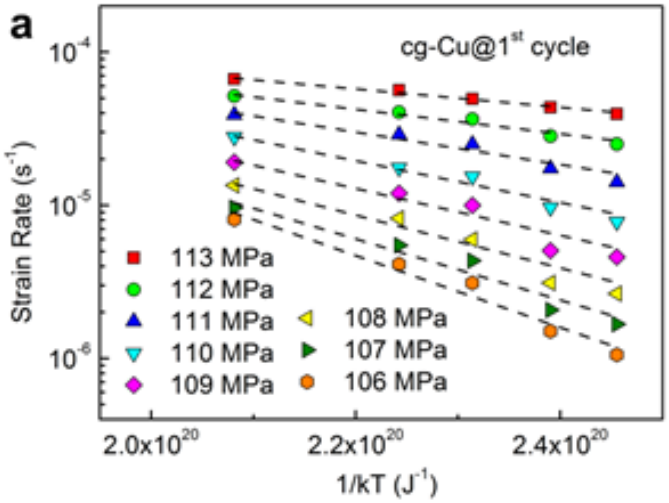


❖ **from a linear high stress relaxation region to a subsequent nonlinear stress relaxation region and finally to a linear low stress relaxation region**

Yang et al., Acta Mat 2016; JMPS 2016

Data driven development of formula for creep and stress relaxation

❖ **Temperature dependency is usually described by the Arrhenius equation:**



$$\dot{\varepsilon}_p(\sigma, T) = \dot{\varepsilon}_0 \exp\left(-\frac{\Delta G}{kT}\right)$$

$\dot{\varepsilon}_0$: prefactor

ΔG: apparent activation energy

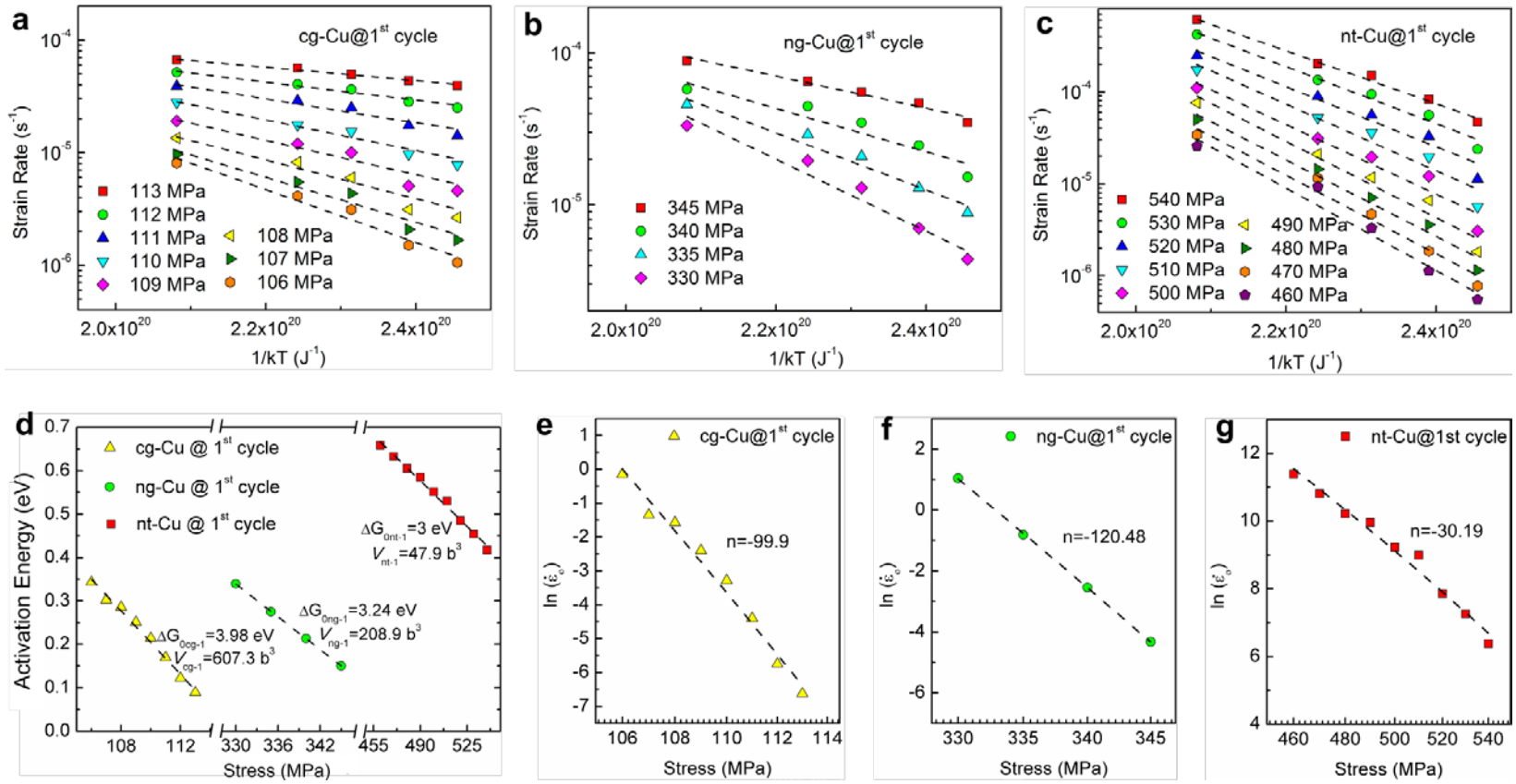
Hypothesis: prefactor and apparent activation energy do not depend on temperature.

❖ **The apparent activation energy is the real activation energy reduced by the applied stress and the activation volume *V*, as a conjugate of applied stress, is introduced to count the reduction in activation energy.**

$$\Delta G = \Delta G_0 - \frac{\sigma}{\sqrt{3}} V$$

where ΔG_0 is called the intrinsic activation energy and $1/\sqrt{3}$ is the numerical factor to convert shear stress to tensile stress.

linear high stress relaxation regions



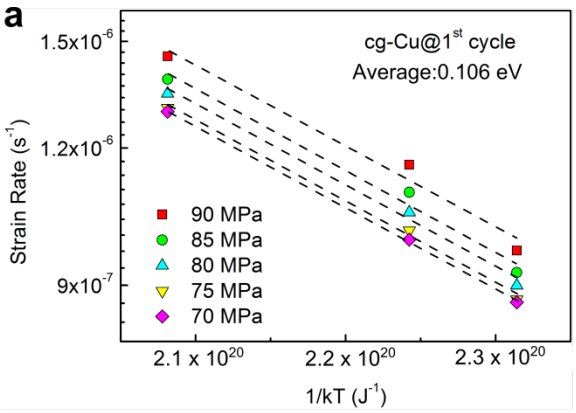
Arrhenius plotting

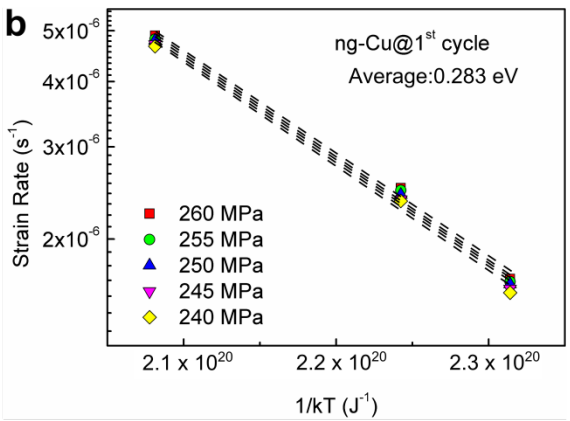
$$\Delta G(\sigma) \longrightarrow \Delta G(\sigma) = \Delta G_0 - \sigma\tilde{V} = \Delta G_0 - \frac{\sigma}{\sqrt{3}}V$$

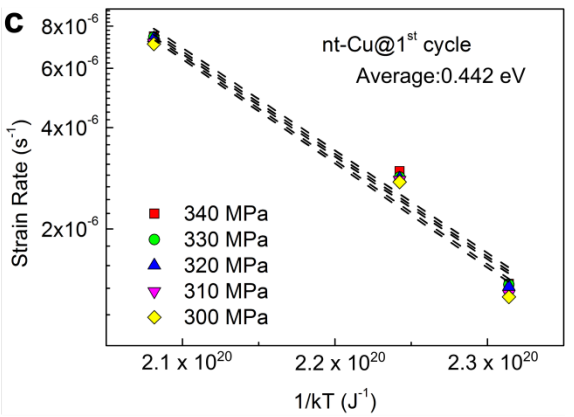
$$\ln(\dot{\varepsilon}_0(\sigma)) \longrightarrow \dot{\varepsilon}_0(\sigma) = \dot{\varepsilon}_{00}\left(\frac{\sigma}{\sigma_0}\right)^n$$

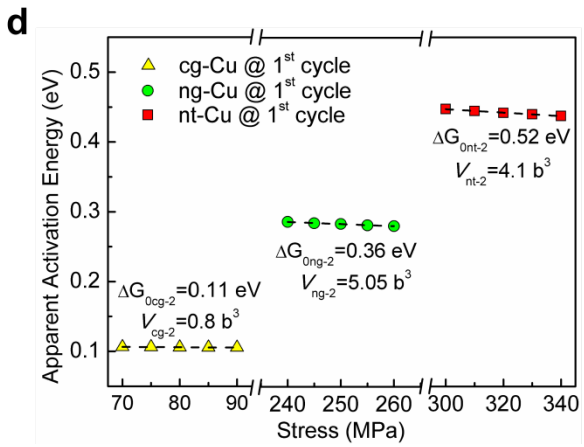
Yang et al., Acta Mat 2016; JMPS 2016

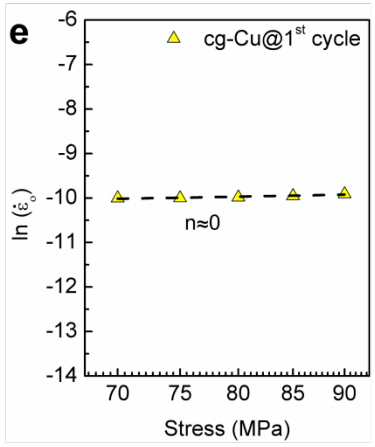
linear low stress relaxation regions

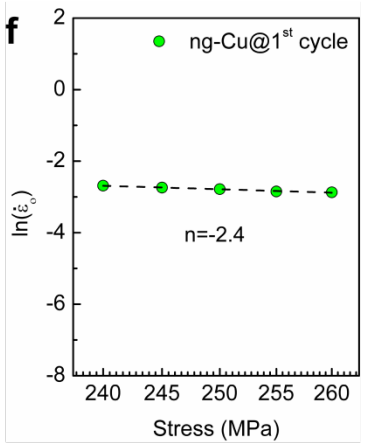


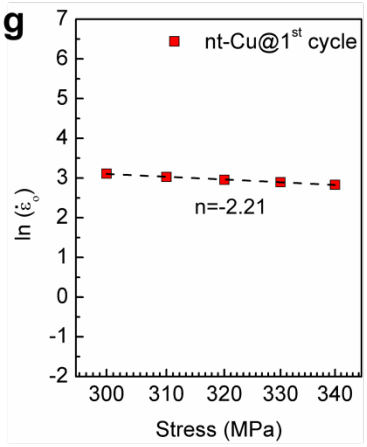












Arrhenius plotting

$$\Delta G(\sigma) \longrightarrow \Delta G(\sigma) = \Delta G_0 - \sigma\tilde{V} = \Delta G_0 - \frac{\sigma}{\sqrt{3}}V$$

$$\ln(\dot{\varepsilon}_0(\sigma)) \longrightarrow \dot{\varepsilon}_0(\sigma) = \dot{\varepsilon}_{00}\left(\frac{\sigma}{\sigma_0}\right)^n$$

Yang et al., Acta Mat 2016; JMPS 2016

Data driven development of formula for creep and stress relaxation

❖ **In the multi-temperature approach, the prefactor is given by a stress power law and the apparent activation energy is the difference between the intrinsic activation energy and mechanical work.**

$$\dot{\varepsilon}_p(\sigma,T)=\dot{\varepsilon}_{00}\left(\frac{\sigma}{\sigma_0}\right)^{n'}\exp\left(-\frac{\Delta G_0}{k_BT}+\frac{\sigma V}{k_BT\sqrt{3}}\right)$$

one has the relationship among the Strain rate sensitivity m, the athermal stress component n', and the activation volume V:

$$\frac{1}{m}=n'+\frac{\sigma V}{\sqrt{3}k_BT}$$

- The equation shows that only when the ratio of mechanical work over the thermal energy approaches zero, i.e., $(\sigma V)/(k_BT\sqrt{3})\to 0$, we shall have $\frac{1}{m}=n'$. If $n'\to 0$, $\frac{1}{m}=\frac{\sigma V}{\sqrt{3}k_BT}$

- If the value of $(\sigma V)/(k_BT\sqrt{3})$ is large and positive, the *m* value is still positive even the value n' is negative.

Data driven development of formula for creep and stress relaxation

❖ **One-temperature approach, in which experiments are conducted at only one temperature, determines the stress component n=1/m from the isothermal data of curves of** lnσ versus ln $\dot{\varepsilon}$**, i.e., from**

$$m = \left.\frac{\partial \log \sigma}{\partial \log \dot{\varepsilon}}\right|_T = \left.\frac{\partial \ln \sigma}{\partial \ln \dot{\varepsilon}}\right|_T \qquad n = \left.\frac{\partial \ln \dot{\varepsilon}}{\partial \ln \sigma}\right|_T$$

With the same isothermal data, **one-temperature approach** determines the value of activation volume *V with*

$$\frac{1}{m} = \frac{\sigma V}{\sqrt{3} k_B T}$$

This implies that one-temperature approach simultaneously uses the power law equation and the Arrhenius equation. **It is mathematically inconsistent.** The fundamental issue is that one-temperature could not provide sufficient data to develop a formula for the stress- and temperature-dependent strain rate.

One-dimensional toy model

$$f(x) = 1.18145 + 0.81x + (-4 + 4x^2)x^2$$

$$x \in [-1.0, 1.0]$$

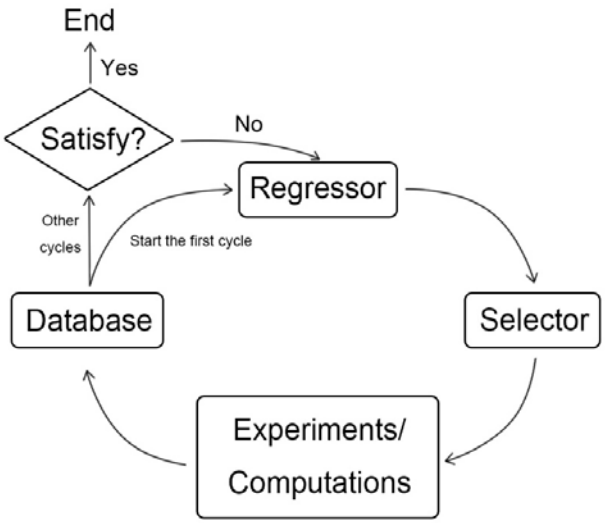
Actual minimum:

$$f(-0.7538) = 0.2226$$

Optimize range:

$$x \in [-1.0, 1.0]$$

feature split 1000 points.



Bootstrap:

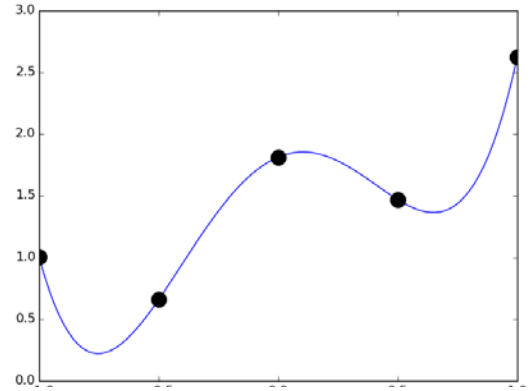
Bootstrap sampling number: 10

Bootstrap sampling feature count between 70% to 90%.

Using SVR regressor with RBF kernel function.

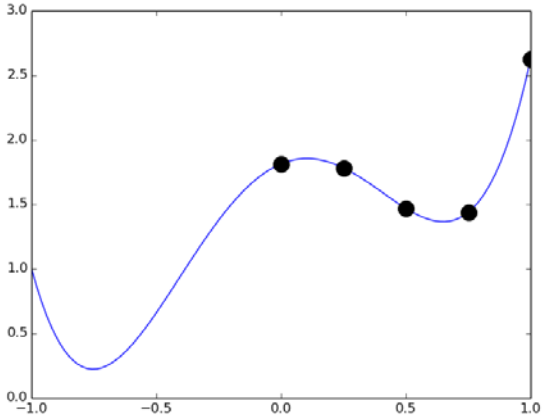
How to find X for min Y with the least number of cycles if the initial data number is 5?

One-dimensional toy model

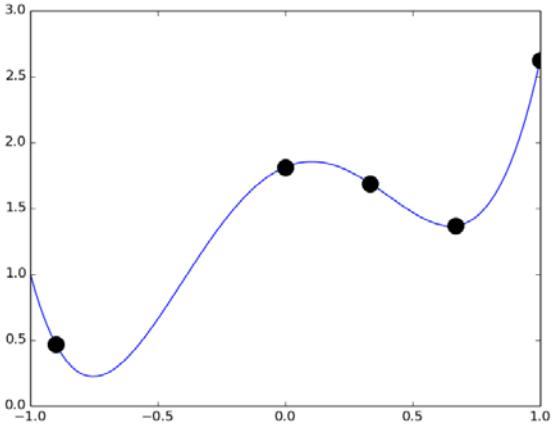


Initial training data distribution

Homogeneous distribution with the minimum of 0.660



Bias distribution with the minimum of 1.438.

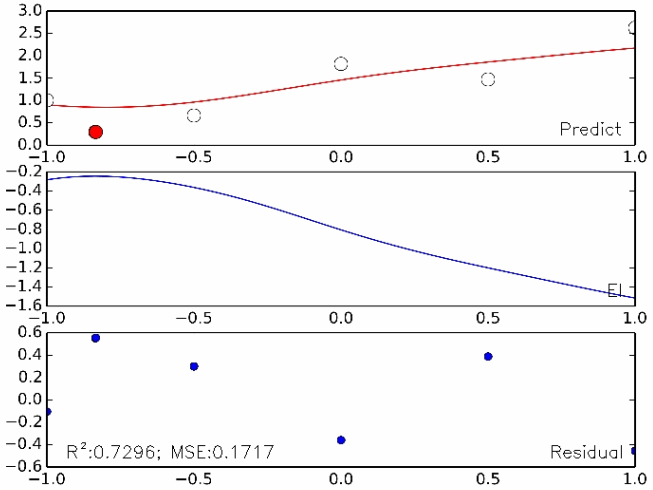


Random distribution with the minimum of 0.470.

SVR Homogeneous distribution of training

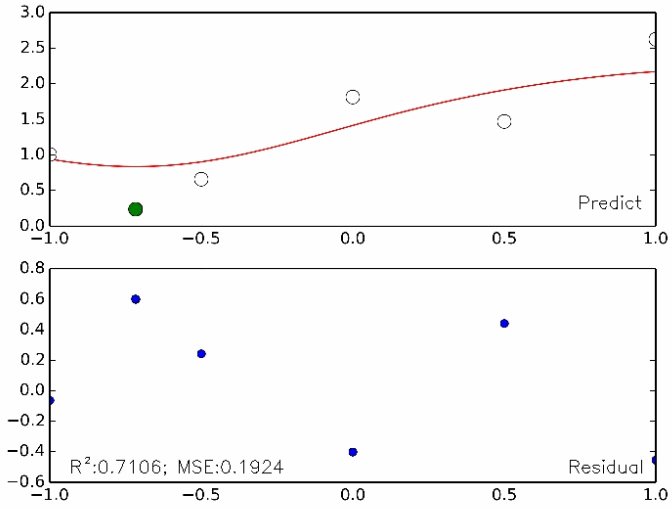
Bootstrap SVR:EGO

iter 1; New point −0.8358 and −0.8338; current best 0.2917



Bootstrap SVR:Min

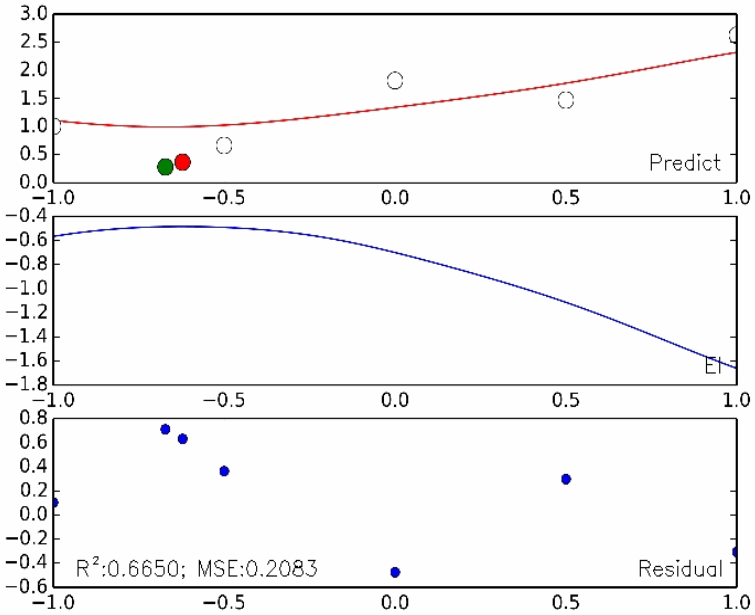
iter 1; New point −0.7157 and −0.7177; current best 0.2341



Residual = $\hat{y}_i - y_i$

Bootstrap SVR:Mix

iter 1; New point −0.6216 and −0.6717; current best 0.2801



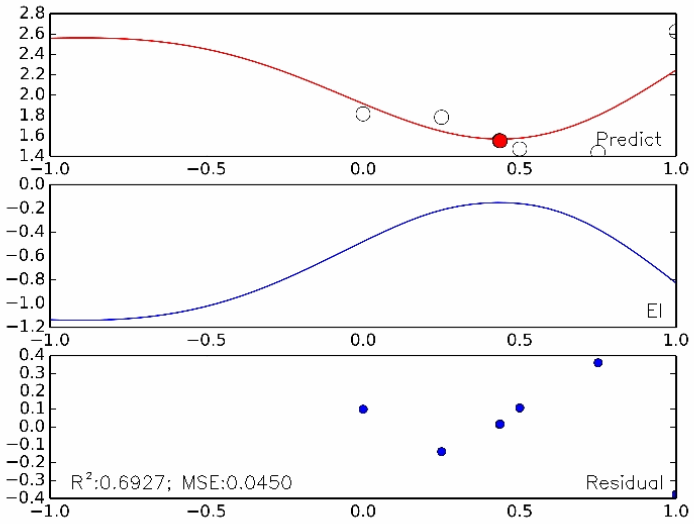
Two data are added in each cycle.

- ✓ EGO: First two largest Eis
- ✓ Min: Two mins predicted by the regressor
- ✓ Mix: The Min predicted by the regressor plus the largest EI

SVR Bias distribution of training data

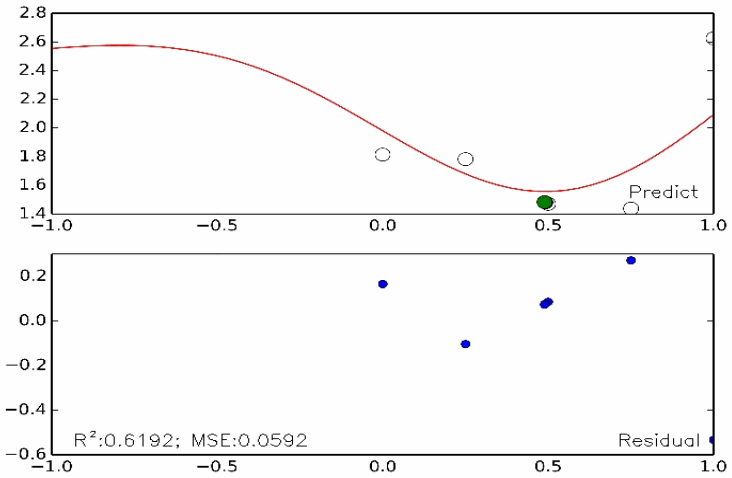
Bootstrap SVR:EGO

iter 1; New point 0.4354 and 0.4374; current best 1.4377



Bootstrap SVR:Min

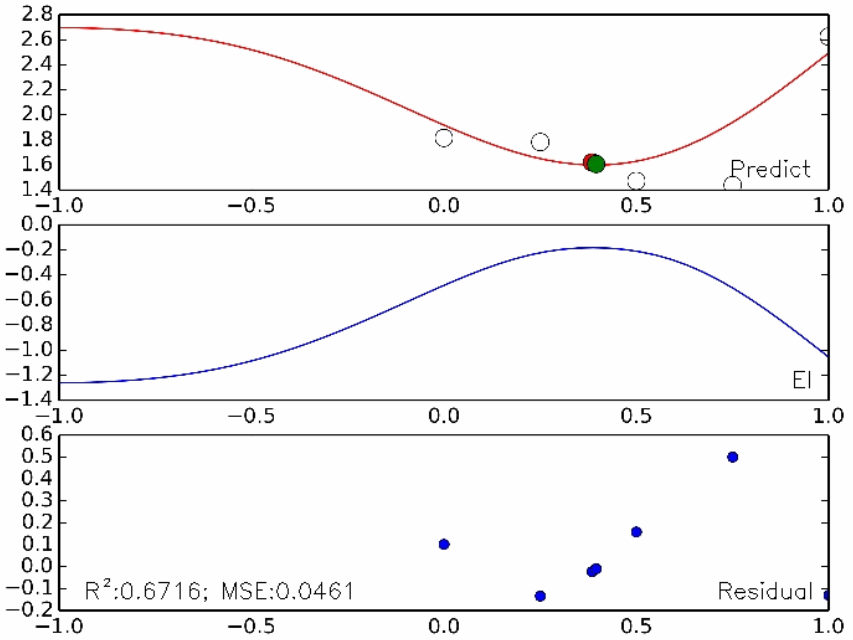
iter 1; New point 0.4915 and 0.4895; current best 1.4377



Only local minimum is reached.

Bootstrap SVR:Mix

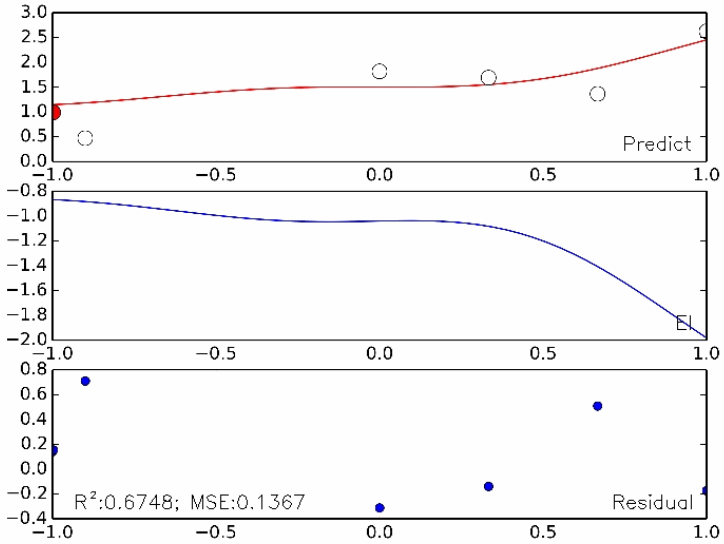
iter 1; New point 0.3854 and 0.3954; current best 1.4377



SVR Random distribution of training data

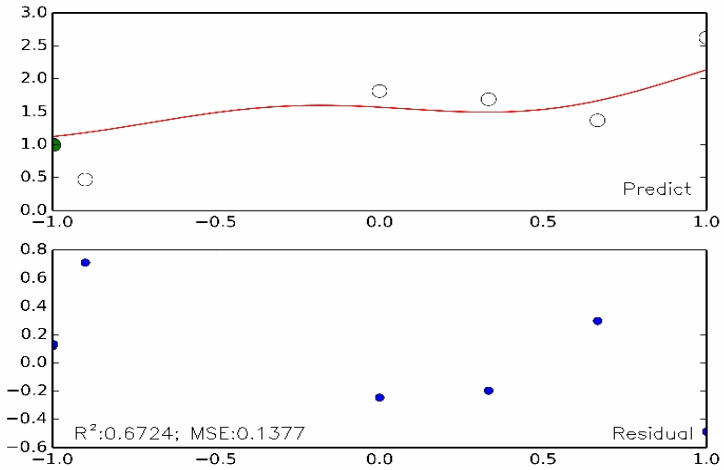
Bootstrap SVR:EGO

iter 1; New point −1.0000 and −0.9980; current best 0.4700



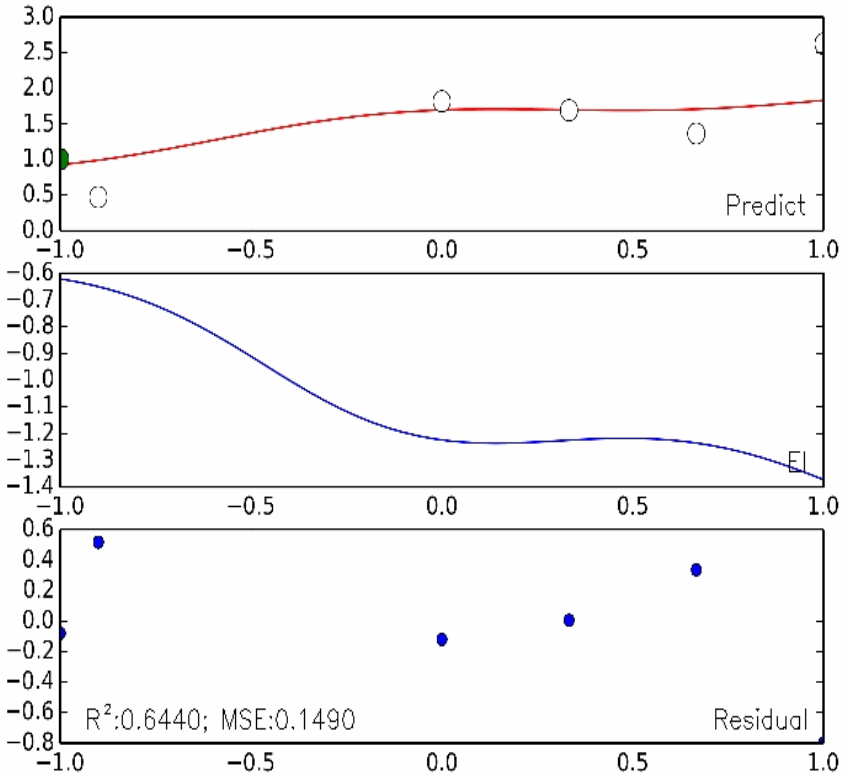
Bootstrap SVR:Min

iter 1; New point −1.0000 and −0.9980; current best 0.4700



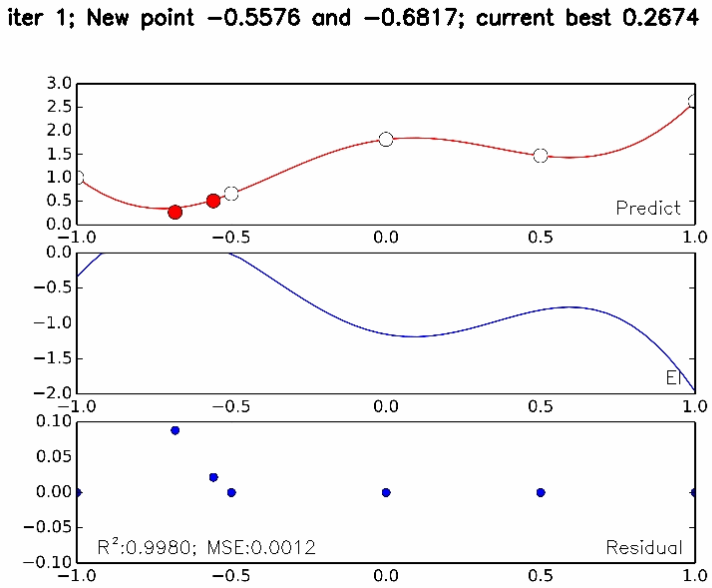
Bootstrap SVR:Mix

iter 1; New point −1.0000 and −1.0000; current best 0.4700

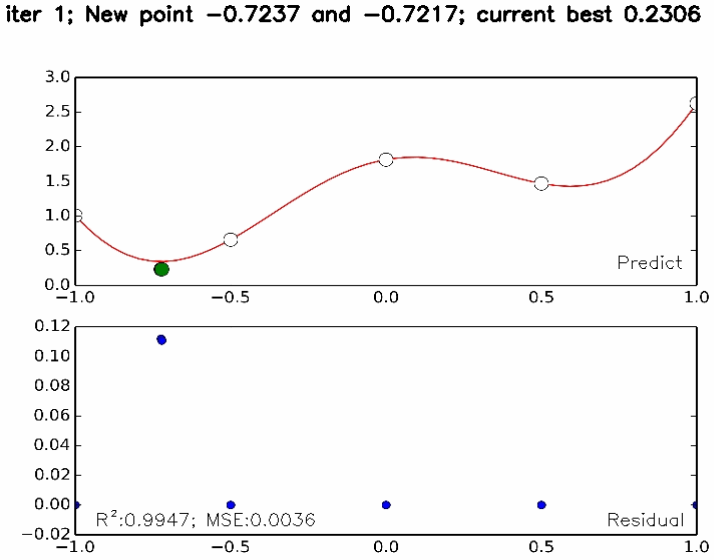


GPR Homogeneous distribution of training data

GPR:EGO

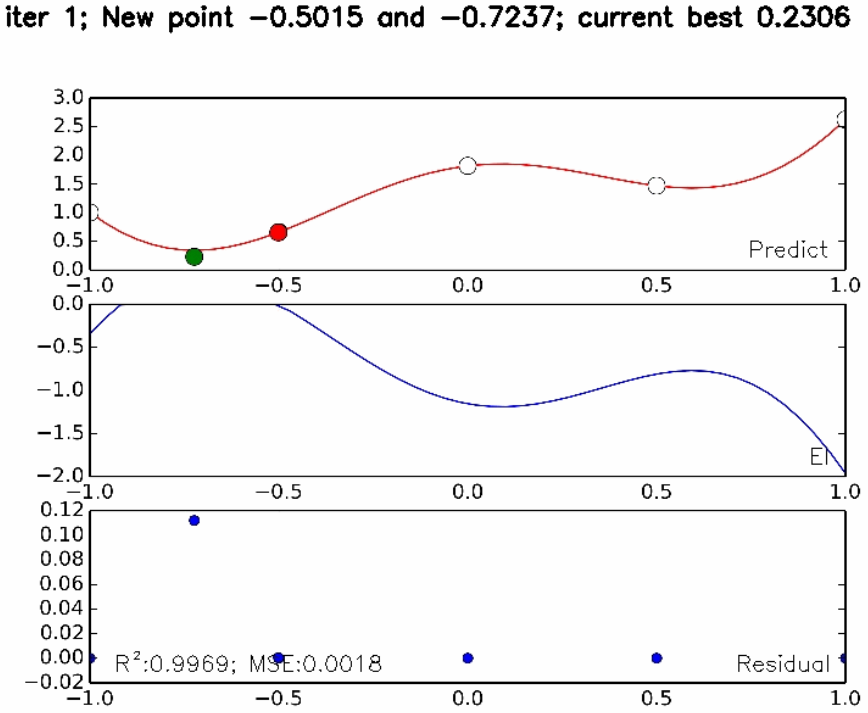


GPR:Min



In a Gaussian process, every point in some continuous input space is associated with a normally distributed random variable. Moreover, every finite collection of those random variables has a multivariate normal distribution. The distribution of a Gaussian process is the joint distribution of all those (infinitely many) random variables, and as such, it is a distribution over functions with a continuous domain. The Gaussian Process Regression (GPR) implements Gaussian processes (GP) for regression purposes.

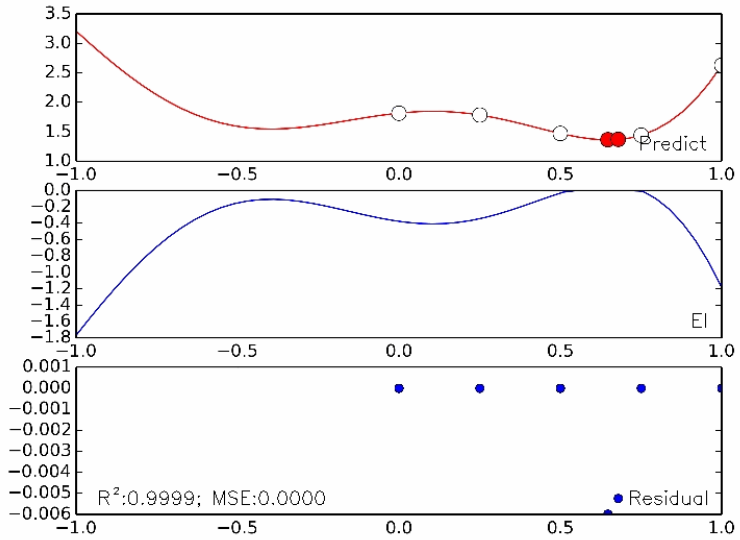
GPR:Mix



GPR Bias distribution of training data

GPR:EGO

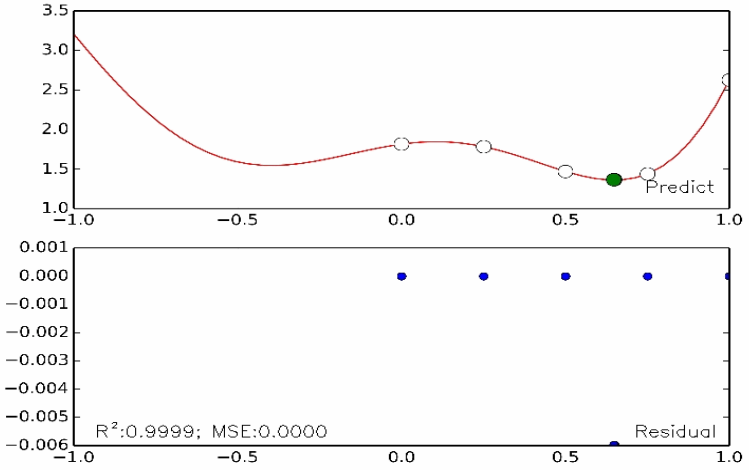
iter 1; New point 0.6476 and 0.6797; current best 1.3651



Global minimum is reached.

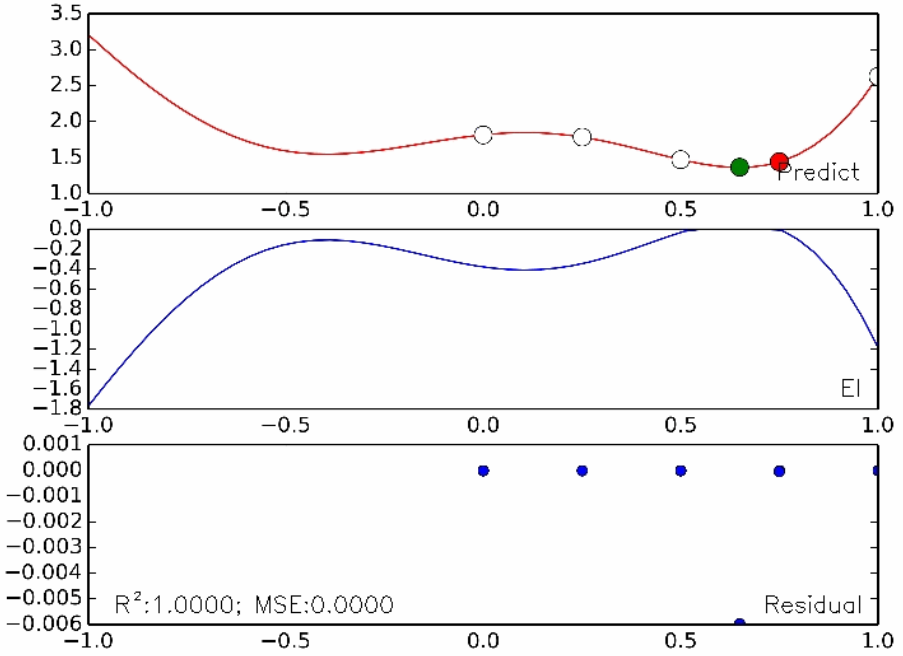
GPR:Min

iter 1; New point 0.6496 and 0.6476; current best 1.3651



GPR:Mix

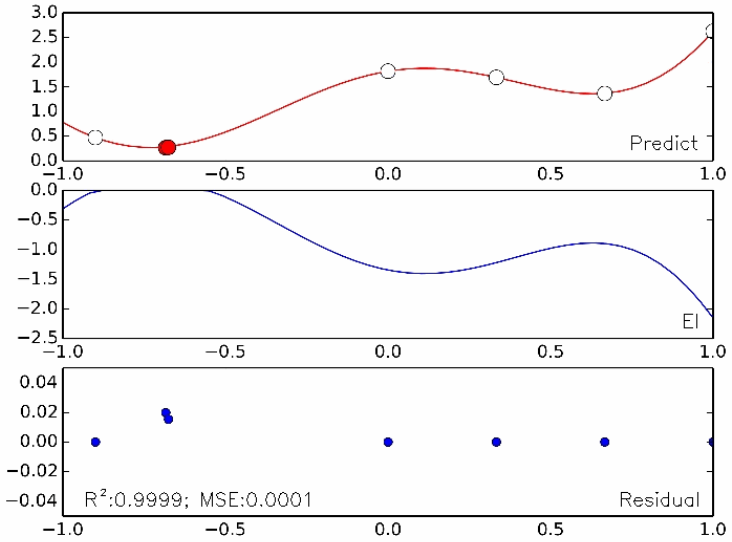
iter 1; New point 0.7497 and 0.6496; current best 1.3651



GPR Random distribution of training data

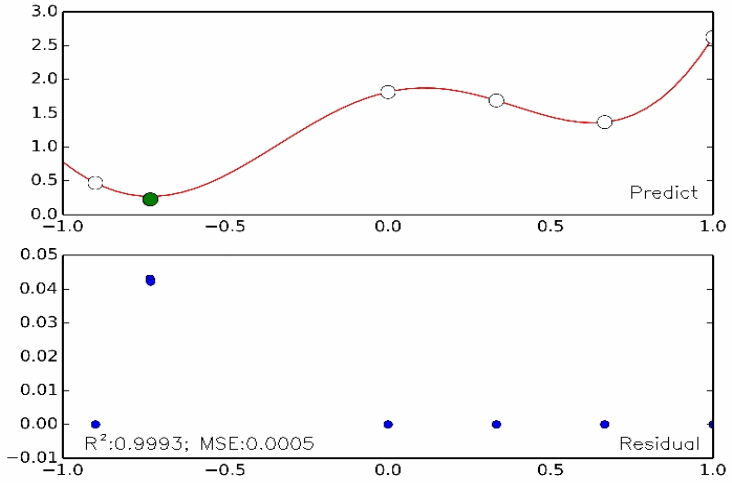
GPR:EGO

iter 1; New point −0.6837 and −0.6757; current best 0.2650



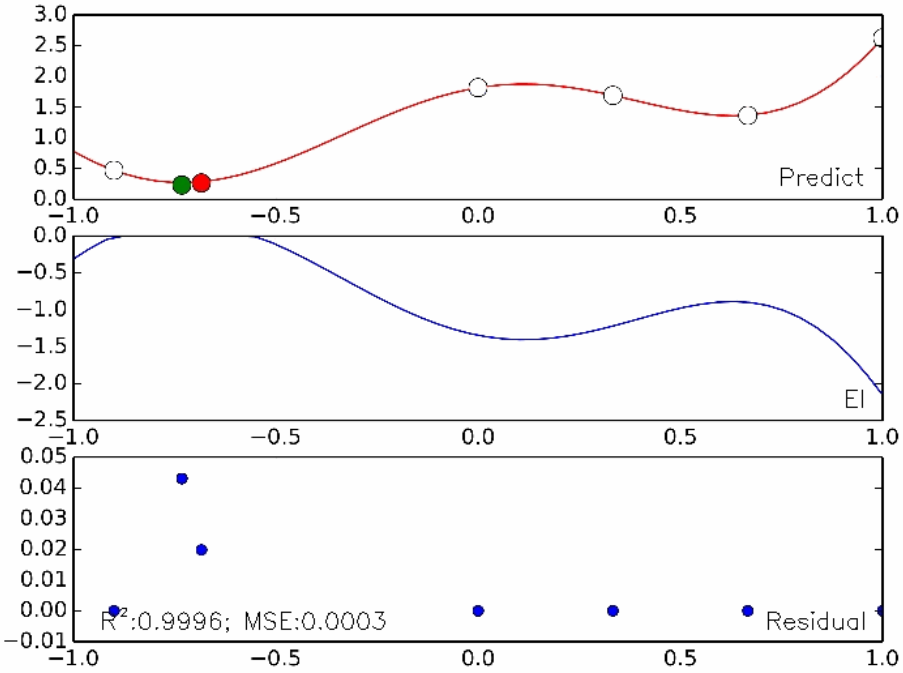
GPR:Min

iter 1; New point −0.7317 and −0.7297; current best 0.2269



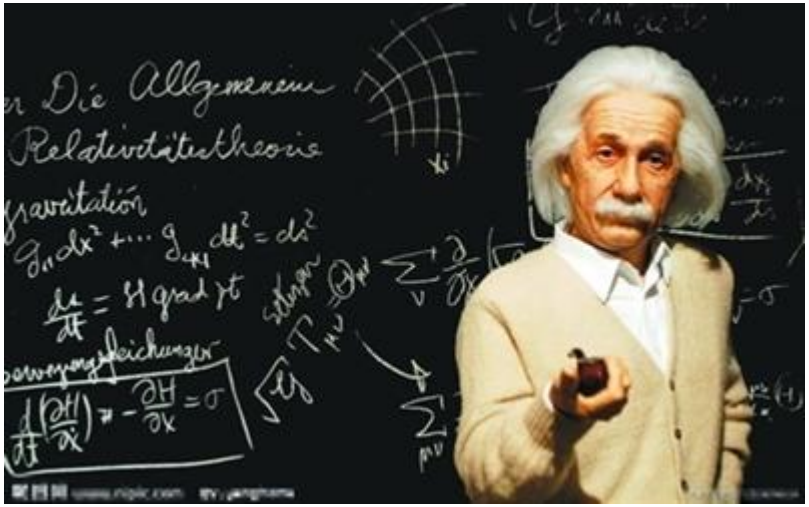
GPR:Mix

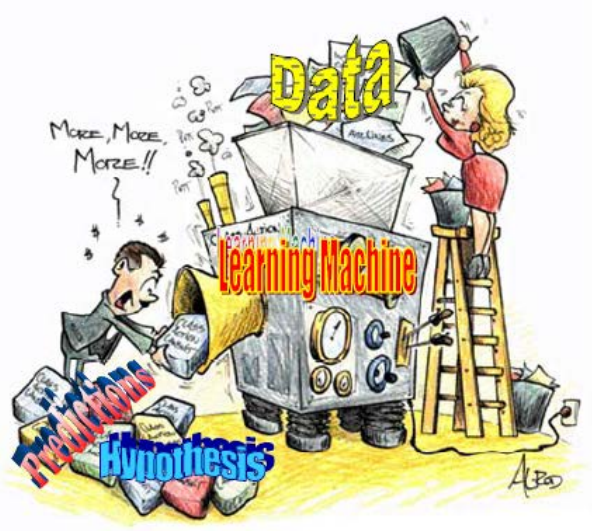
iter 1; New point −0.6837 and −0.7317; current best 0.2269



Concluding Remarks

- **It is urgent to building up materials database.**
- **Integration of Experiments, Computation, Theory and Data Science**
- **Educate the Next-Generation Materials Workforce with artificial intelligence**
- **Knowledge from data is achieved by leaning of human being, especially genius, or/and of machine.**





© Eric Xing @ CMU, 2006-2008

Thanks!!!

Thanks for your attention !

Questions ?

Thank you for your attention!

SCI 1.719 IMPACT FACTOR 2016

Sponsored by
Chinese Academy of Sciences
National Natural Science Foundation of China

Published by
《中国科学》杂志社 SCIENCE CHINA PRESS
Springer

SCIENCE CHINA
Technological Sciences

- A monthly peer-reviewed academic journal
- Indexed by SCI, EI, etc.
- Manuscript types:
  Review, Article, Brief report, News & views
- Rapid review and
  timely publication (online immediately)

Free full text available on http://tech.scichina.com