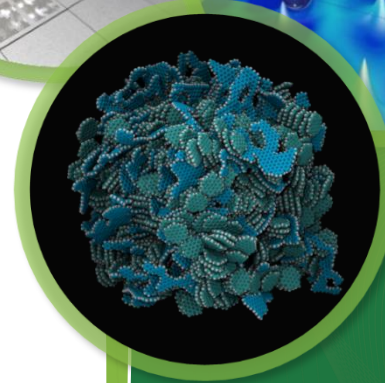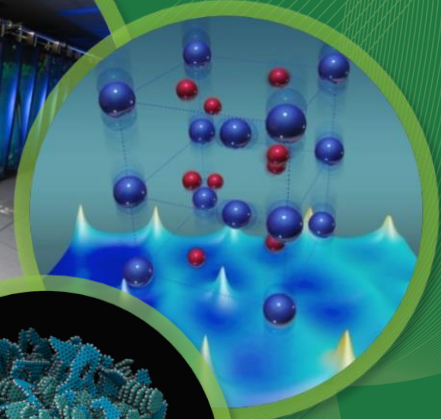# Thoughts on Materials and Big Data/Machine Learning

Bruce Pint

Group Leader and Distinguished Research Staff

Corrosion Science and Technology Group

Materials Science and Technology Division

Oak Ridge National Laboratory

OAK RIDGE
National Laboratory

Perspective

- I believe in outcomes: use all the tools available

- I believe in competition to accelerate innovative research

- High temperature corrosion specialist
  - Ph.D. 1992 in Ceramic Science and Engineering
  - Google Scholar: >10,000 citations, h index = 55
  - Fellow of NACE International and ASM International
  - Impacts: gas turbine recuperators and accident tolerant fuel cladding

- Worked for DOE Fossil Energy for ~25 years (Feb. 1994)
  - Manage four current FE projects (plus others for NE, Solar, Fusion, EERE)
    - Steam oxidation
    - Supercritical carbon dioxide (sCO_2) compatibility
    - Thermal/environmental barrier coatings for gas turbine hot section
    - ASME BPV code qualification for Haynes 282

OAK RIDGE
National Laboratory

Experience with machine learning

- My experiments not really "big" data

- DOE EERE Vehicle Technologies Project (in progress)
  - ICME for high temperature exhaust valves
  - Input to alloy developers on oxidation behavior of Ni-Cr-Fe-Al-Ti alloys
  - Model Ni-Cr-Fe-Al-Ti alloys did not yield sufficient guidance
  - Environment: exhaust gas, 750-1050°C, cyclic operation
  - ML: significant improvement in fit from ~30 to ~75 data sets
 - Pathway: Composition => Thermodynamic Phases => Performance
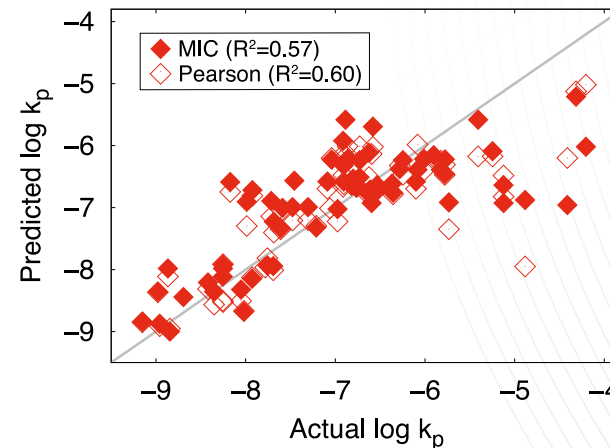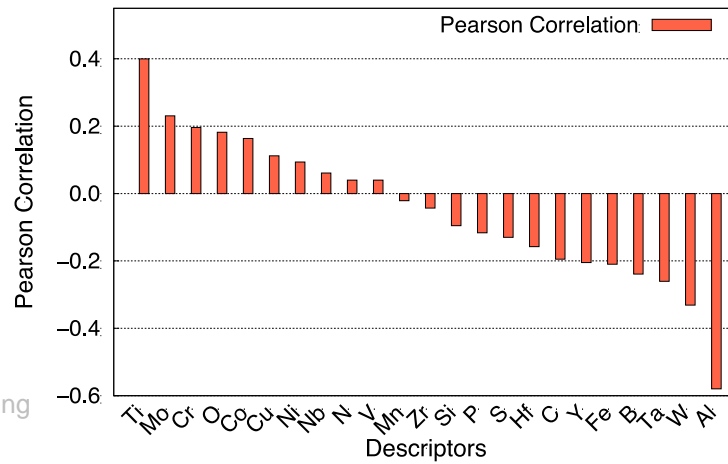
BD/ML Opportunities (where find complexity + data?)

✓ Coal ash (fireside) corrosion

x Flexible plant performance (creep-fatigue, etc.) = lack of data

x Steam oxidation = lack of reliable data (correct T, P, water chemistry)

- Alloy Development
  - x Fe-base alloy for supercritical CO_2: lack of data/understanding
  - ✓ Ni-base alloys for 1500°F (816°C) operation (~1400°F current limit)
  - Creep-resistant ferritic-martensitic steels: lack of chemistry with creep data

- Functional materials
  - Sensors, capture, etc.

✓ Pairing data scientists with subject matter experts at institutions

OAK RIDGE
National Laboratory

OAK RIDGE
National Laboratory

# Machine learning – A modern tool for high-dimensional data

Data Collection/Population

- **Consistent 'experimental' dataset with known history**
- Cyclic oxidation behavior represented with pkp model
- Elemental compositions as initial descriptors
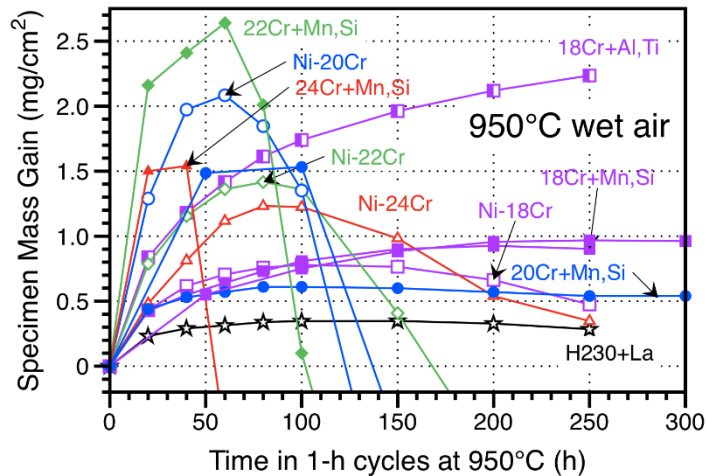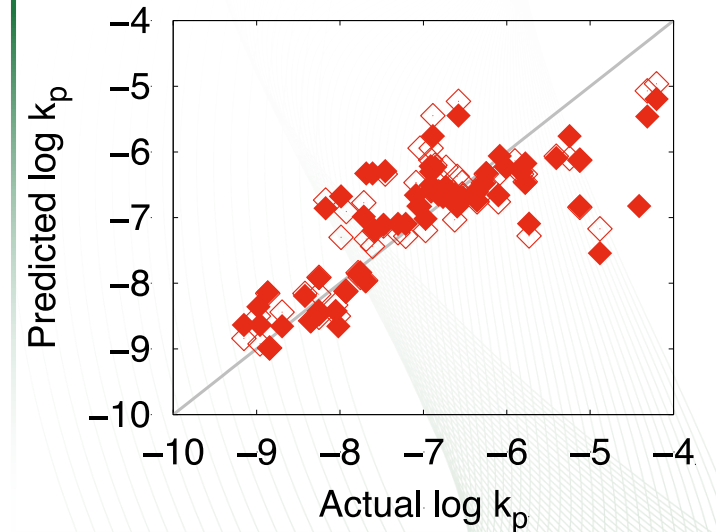

Correlation Analysis

- Selecting features to be used within the machine learning
- High ranking descriptors
- **Domain knowledge for interpretation**
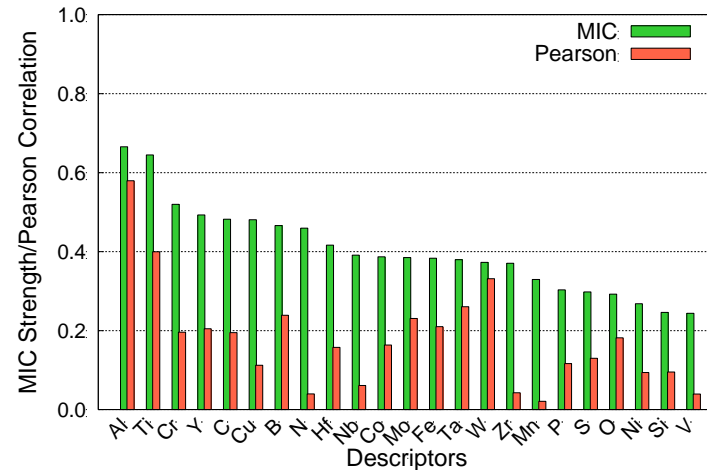
Machine Learning

- **'Just fit' the curve…**
- Different ML models
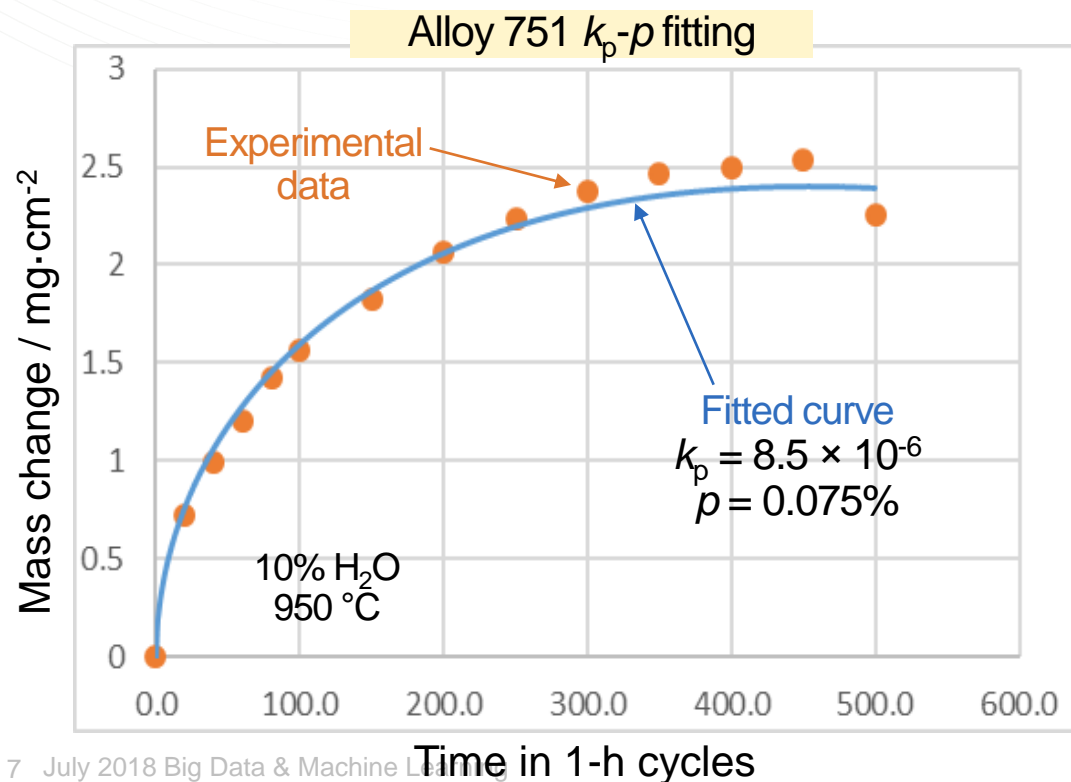 - Random Forest
  - Linear regression

OAK RIDGE
National Laboratory

k_p-p model is used to fit mass change curves and quantify oxidation behavior.

k_p	p
Parabolic oxide growth rate (mg^2·cm^{-4}·s^{-1})	Oxide spallation probability (%)
Mass gain	Mass loss

D. Poquillon and D. Monceau, Oxidation of Metals, Vol. 59, Nos. 3/4 (2003) 409-431.

Net mass change = Mass gain by k_p + Mass loss by p

Alloy 751 k_p-p fitting

Experimental data

Fitted curve
$k_p = 8.5 \times 10^{-6}$
$p = 0.075\%$

10% H_2O
950 °C

Mass change / mg·cm^{-2}

Time in 1-h cycles

Total 75 fitted dataset (10% H_2O & 950°C condition)

Fitting quality by Pearson Correlation Coefficient (PCC)

~40% good (PCC > 0.95)
~30% fair (PCC 0.8~0.9)
~30% poor (PCC < 0.8)

Still used all 75 fitted k_p & alloy composition for correlation.

OAK RIDGE
National Laboratory

Quantified influence of elements on k_p & p values were compared using Pearson correlation analysis.

Pearson Correlation (P²) from manual fitting result

OAK RIDGE
National Laboratory