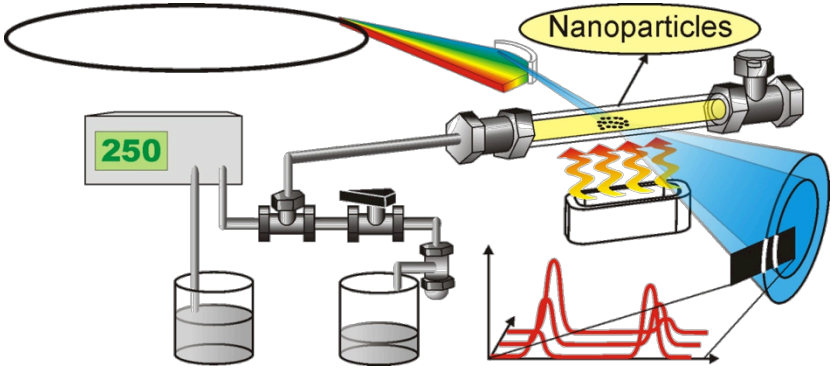
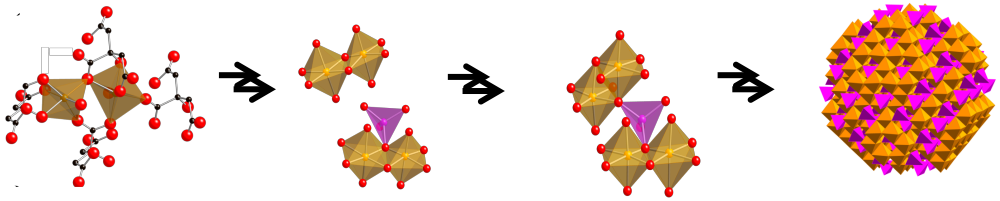
Applications of principal component analysis to pair distribution function

Karena W. Chapman, Saul H. Lapidus and Peter J. Chupas

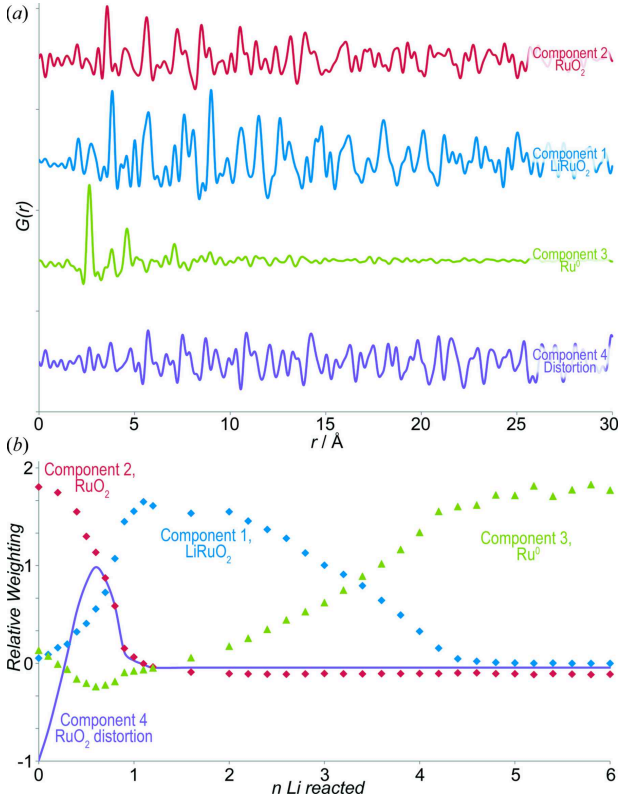


Why do we do pair distribution function analysis



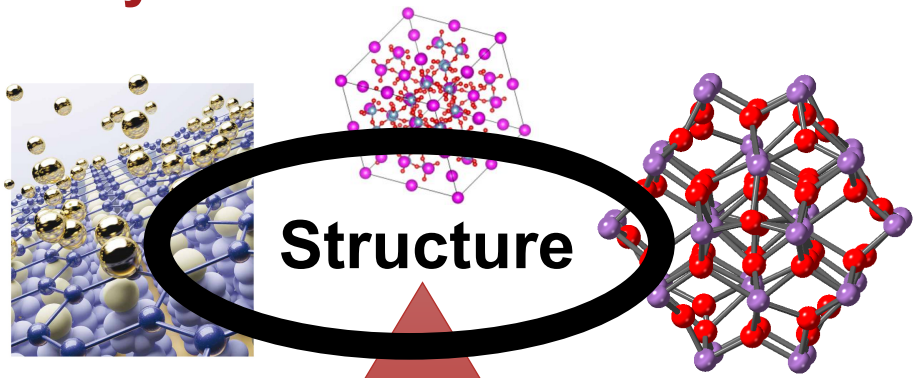
How does we use PCA on pair distribution function data?

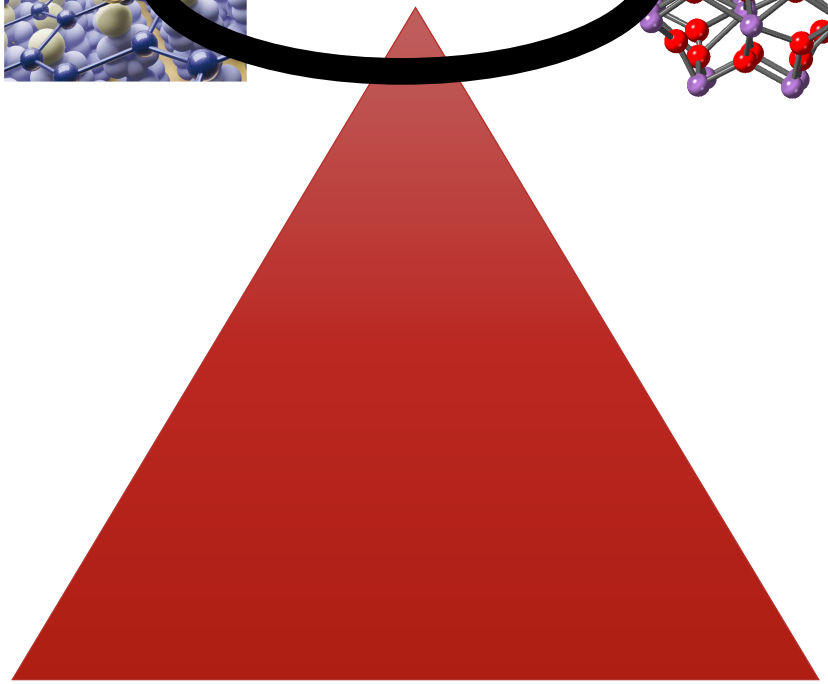
doi:10.1107/S1600576715016532

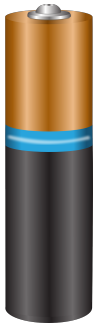


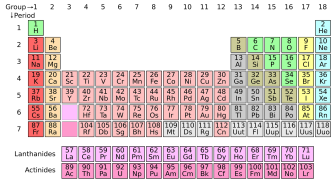
What is principal component analysis

The Holy Grail of Materials Chemistry



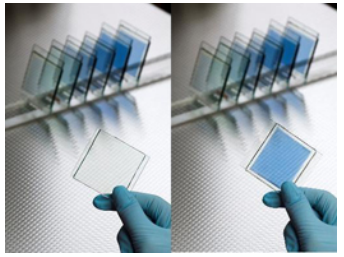




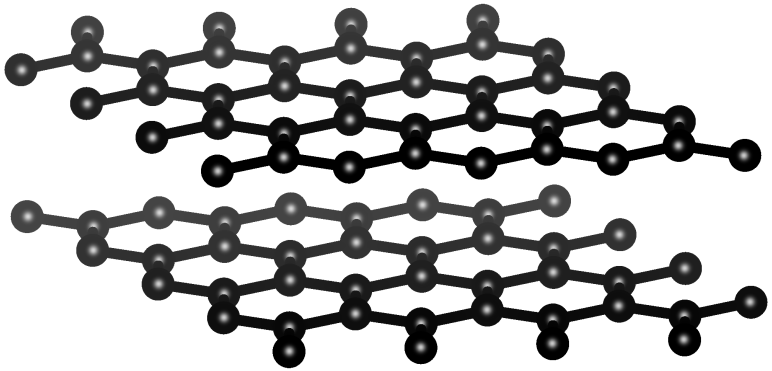


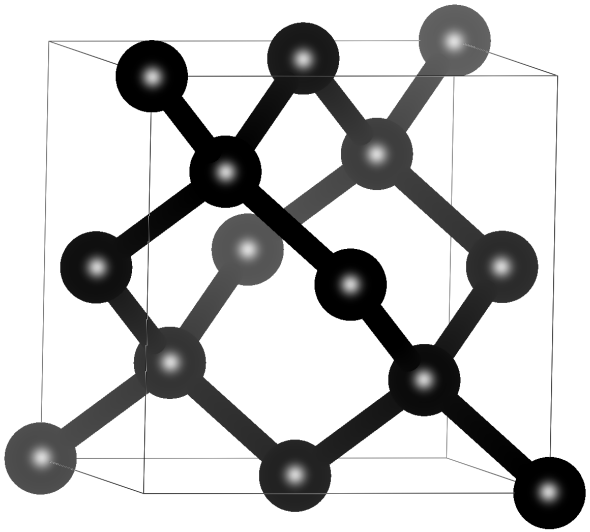
Synthesis

Properties



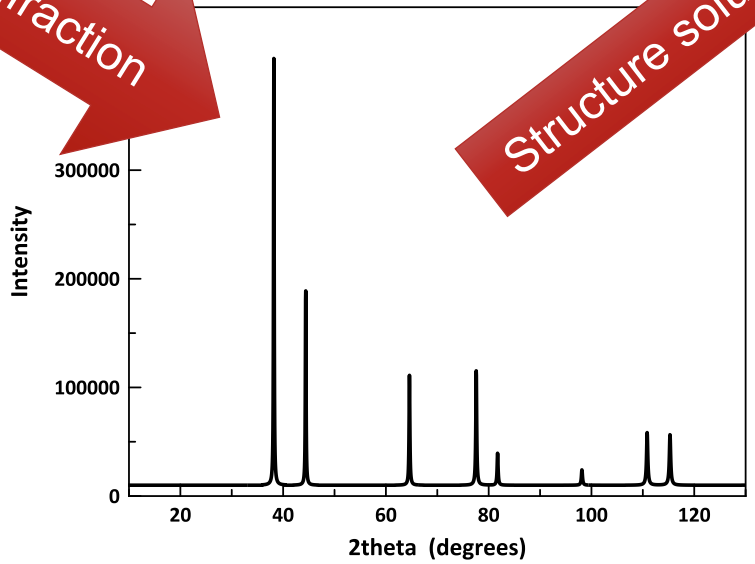
Properties and structure are directly connected



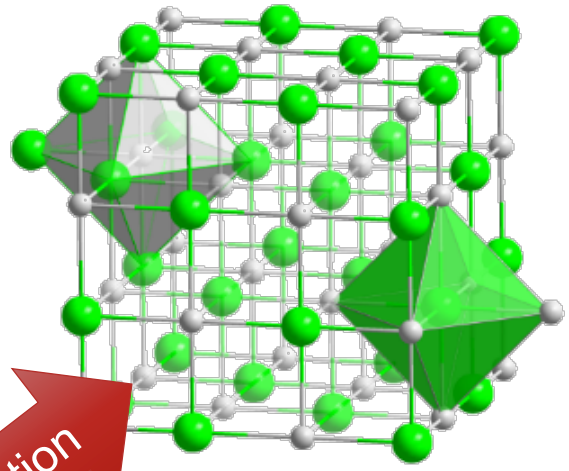


X-ray diffraction from bulk materials

X-ray diffraction



Structure solution



Constructive interference due to periodicity

New possibilities with 3rd generation synchrotron and high energy neutron facilities

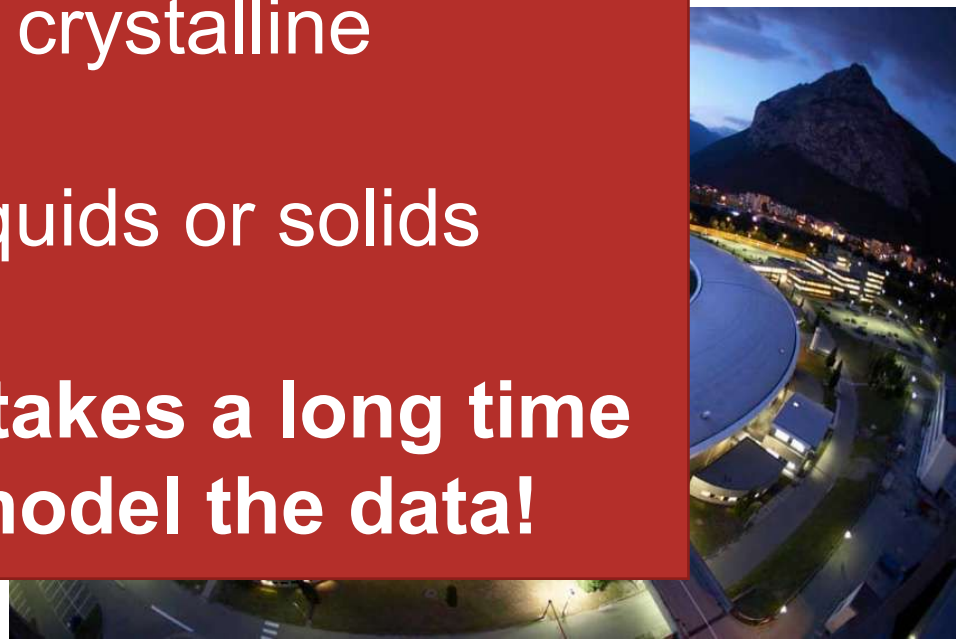
- **High energy**
- **High X-ray flux!**
- **More signal!**
- **Lots of data!**

All classes of materials

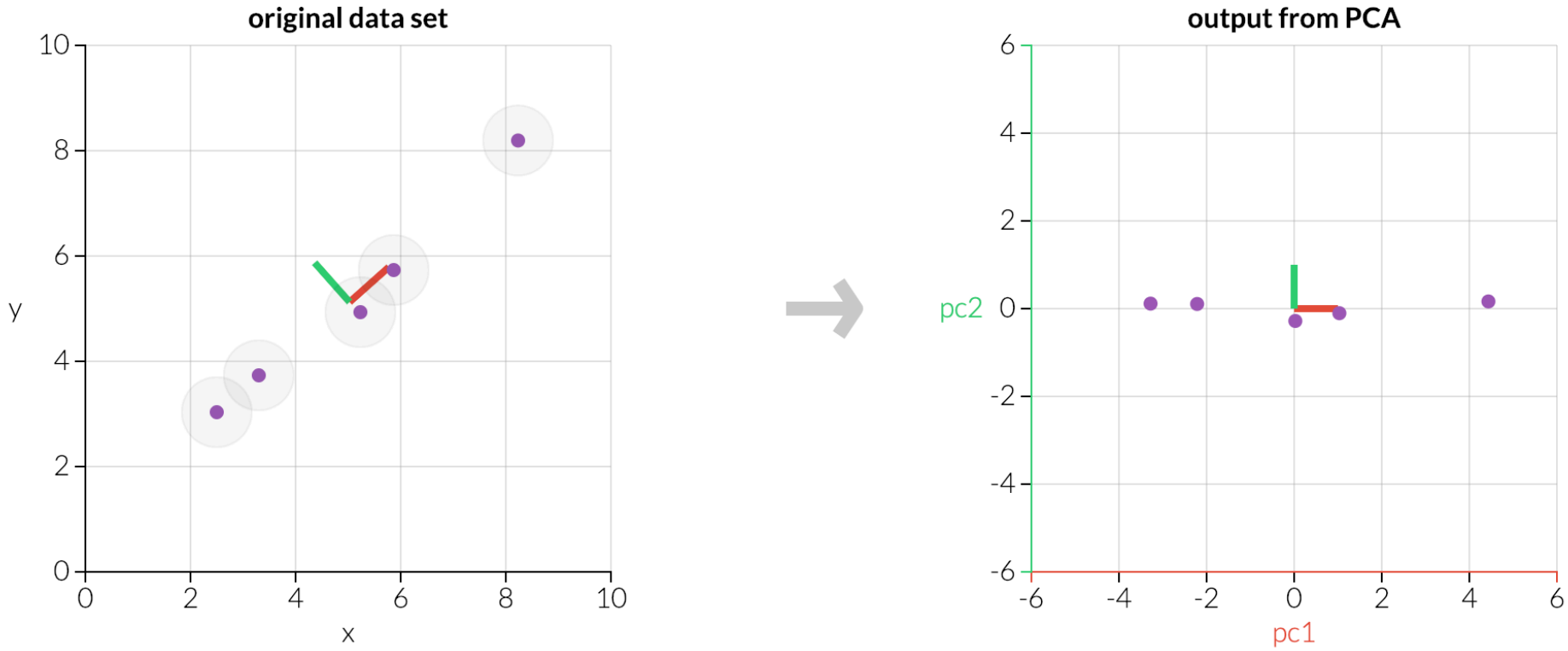
Amorphous, nanostructured and crystalline

Liquids or solids

But it takes a long time to model the data!



Principal Component Analysis in 2 dimensions

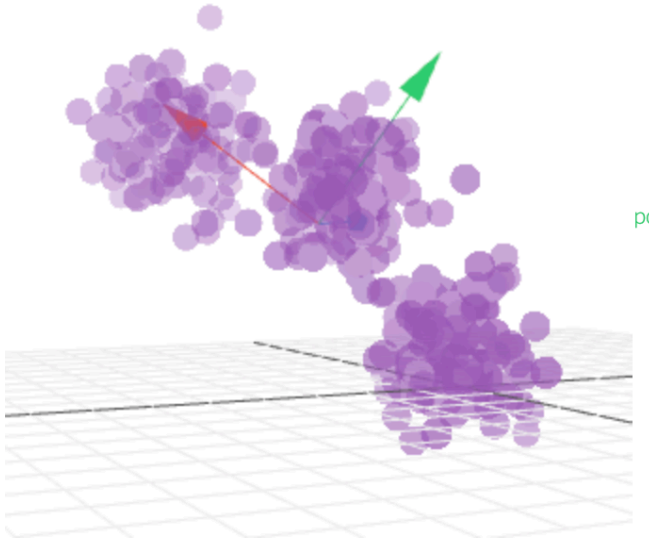


- **2 linear regressions**
- **Scaling – difficulties with chemical systems??**
- **Information in the PC's**

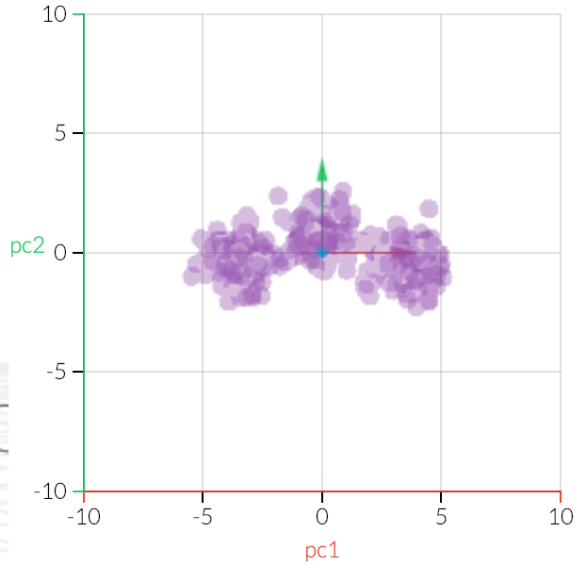
http://setosa.io/ev/principal-component-analysis/

Principal Component Analysis in 3 dimensions

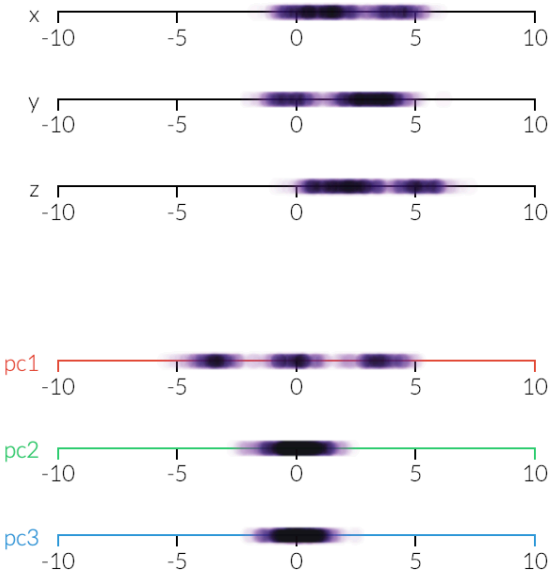
- **3D visualization**



Projected to 2D



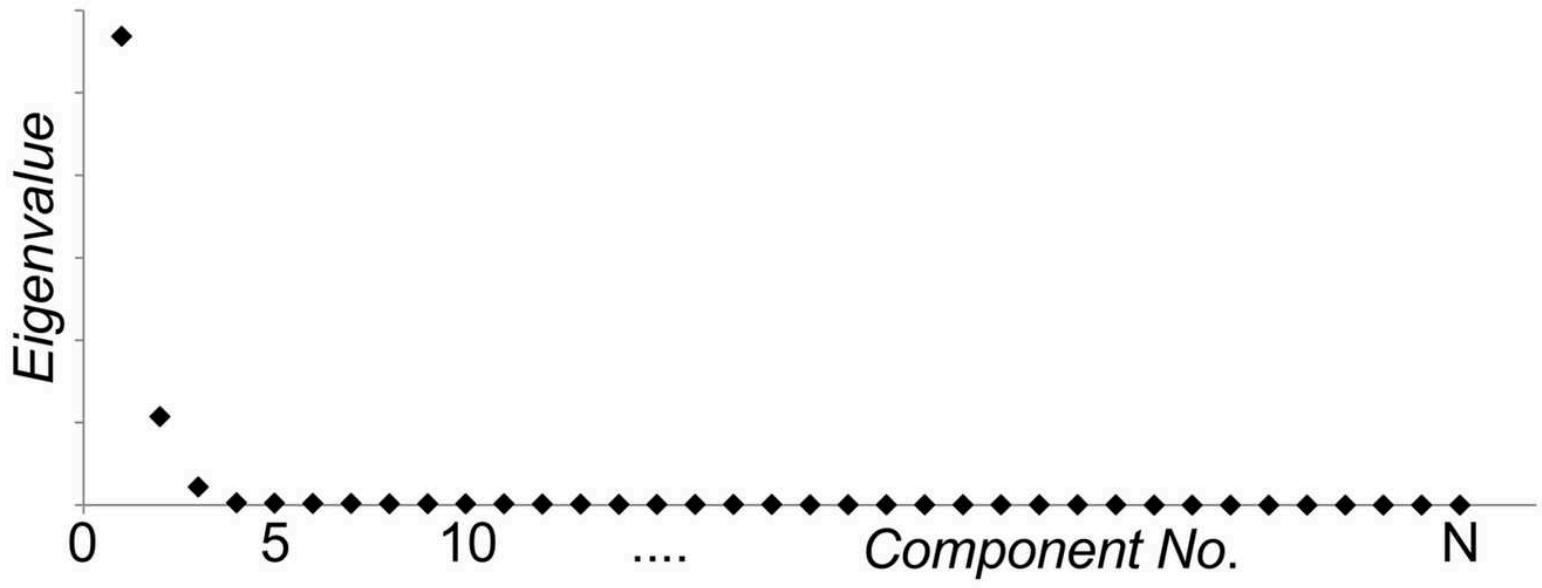
Projected to 1D



- **3 linear regressions**
- **Difficult to visualize multiple dimensions**
- **Clustering**
- **Robustness with pair distribution function data**

http://setosa.io/ev/principal-component-analysis/

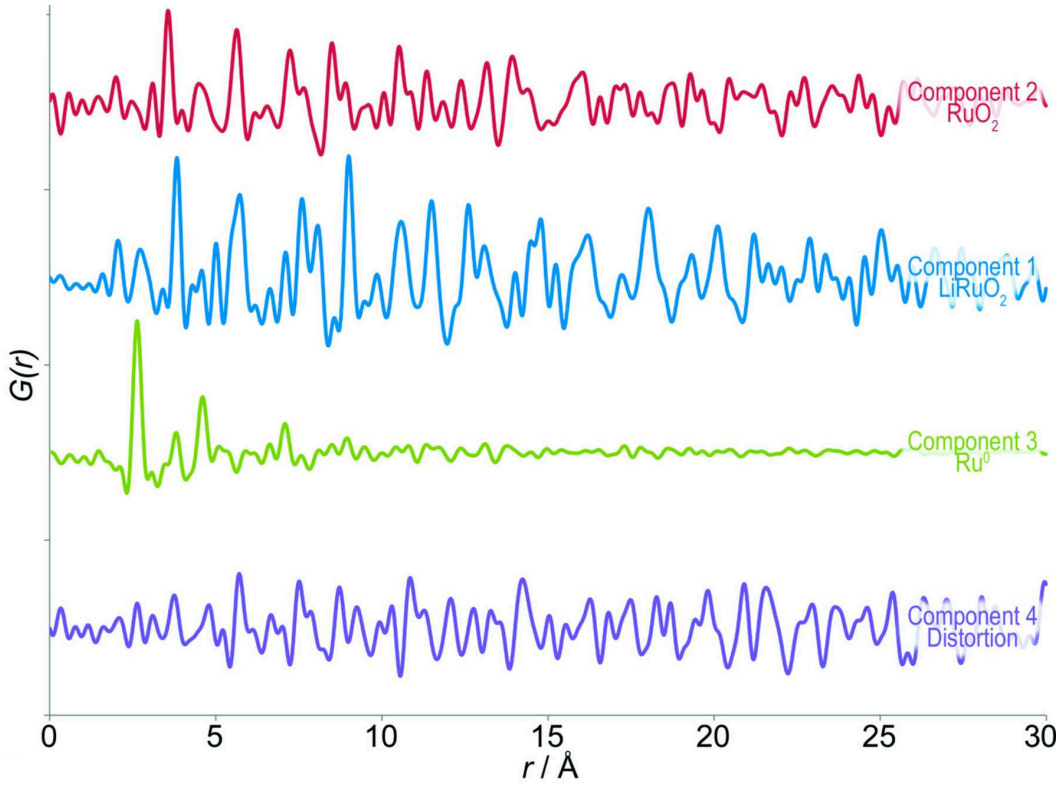
PCA Scree plot



- **How many PC's describes the data?**
- **How much information in the PC?**

doi:10.1107/S1600576715016532

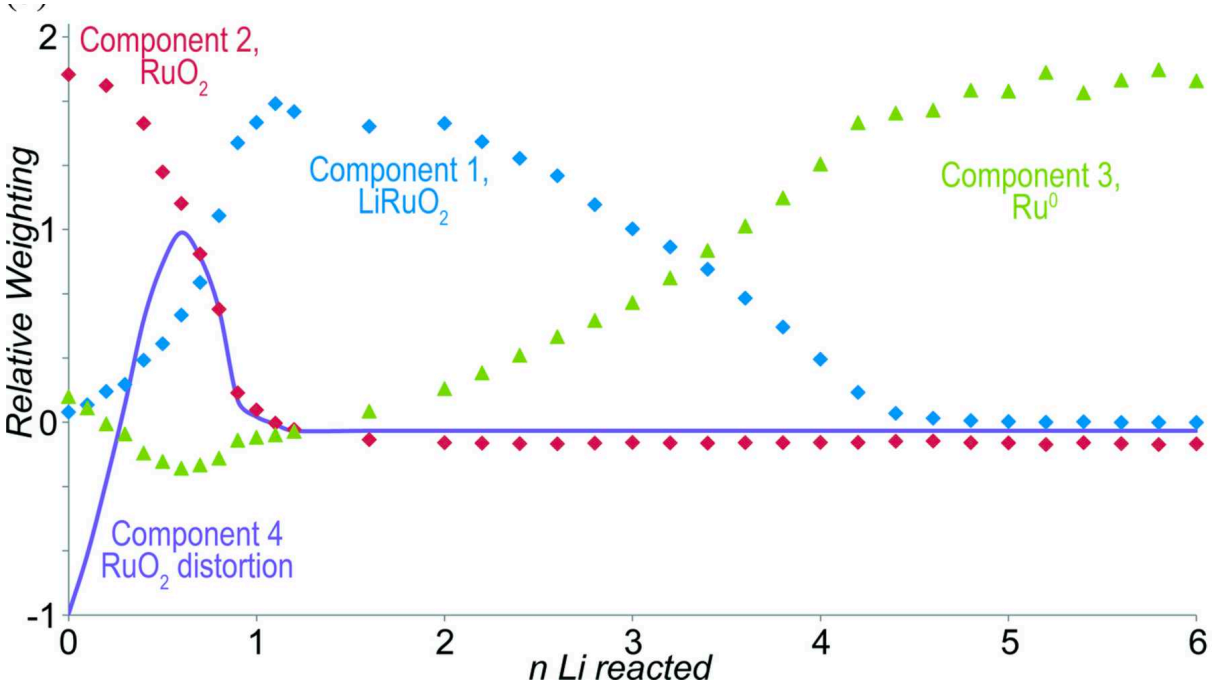
Principal Components plot



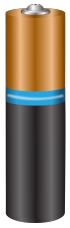
- How does the PC look like?
- Are they physical?
- Does a structure relate to 1 or multiple PC's?

doi:10.1107/S1600576715016532

Ratio of Principal Components



- **How does the ratio of PC's change during the reaction?**



RuO_2 → Li_xRuO_2 → Ru + Li_2O

doi:10.1107/S1600576715016532

Conclusion:
Principal component analysis to pair distribution function

- **Modelling of PDF data by conventional techniques are time-consuming.**
- **PCA on pair distribution function data is fast.**
- **PCA does not need any prior knowledge (bias) of the chemical structure.**
- **PCA can identify trends that the human eye cannot – peak overlap, large amounts of data, complicated changes.**
- **This article shows that PCA is efficient in modelling pair distribution function data.**