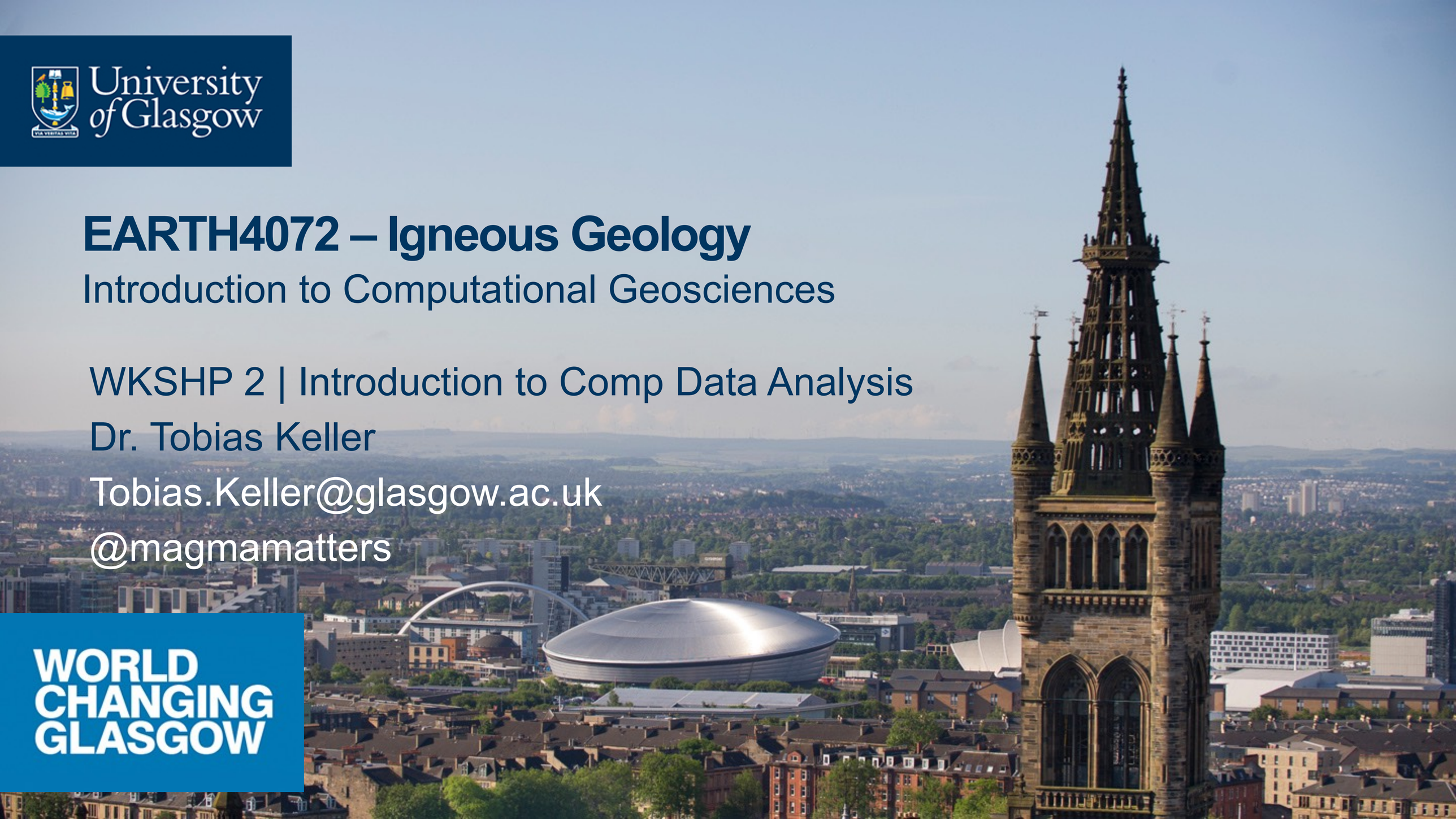


University of Glasgow

EARTH4072 – Igneous Geology

Introduction to Computational Geosciences

WKSHP 2 | Introduction to Comp Data Analysis

Dr. Tobias Keller

Tobias.Keller@glasgow.ac.uk

@magmamatters

WORLD CHANGING GLASGOW

Intro Comp Geosci | Programme

Week	WKSHP I	WKSHP II	WKSHP III	WKSHP IV
20/10/2020	First Steps Coding	Comp Data Analysis	Comp Modelling I	Comp Modelling II

Comp Geosci | Intended Learning Outcomes

Introduction to Comp Data Analysis

- understand the relationships between observations, experiments, models, and theory
- understand complementary roles of Machine Learning and Process Modelling in Computational Geosciences
- expand your scientific programming skills with Python
- take first steps in data analysis and machine learning
- learn how to plot and fit lines and curves to time series data

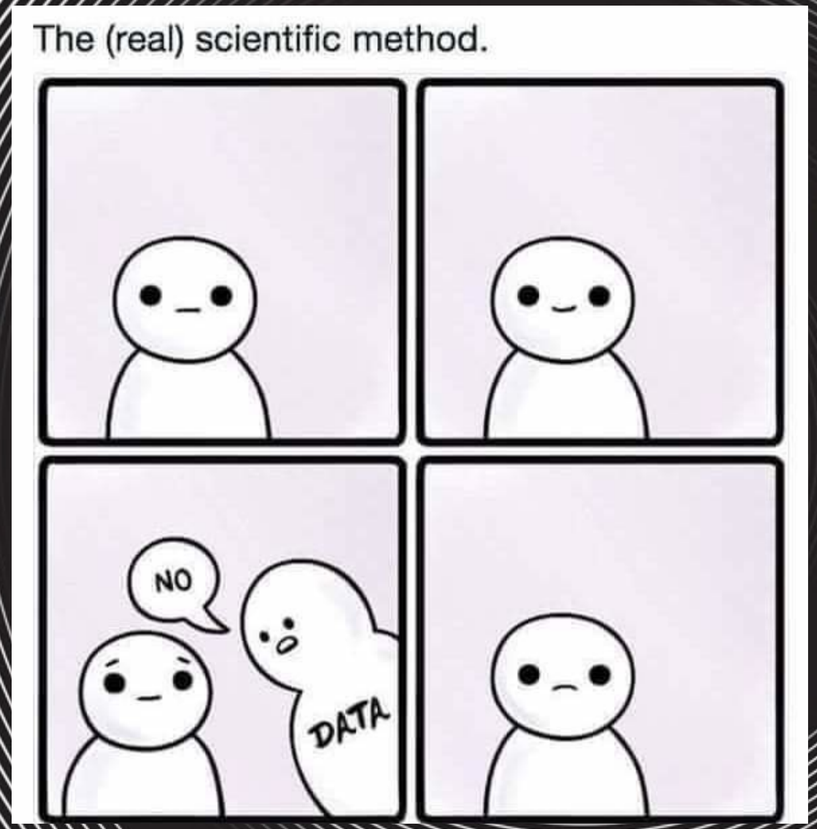
```
% update constitutive relations
txx = eta .* exx + chi .* txxo;       % x-normal stress
tzz = eta .* ezz + chi .* tzzo;       % z-normal stress
txz = etac.* exz + chic.* txzo;       % xz-shear stress

p   = - zeta .* Div_V + xi .* po;     % compaction pressure
p([1 end],:) = p([end-1 2],:);        % periodic boundaries
p(:,[1 end]) = p(:,[end-1 2]);

w   = - (K(1:end-1,:).*K(2:end,:)).^0.5 .* (diff(P,1,1)./h + 1);
w(:,[1 end]) = w(:,[end-1 2]);

u   = - (K(:,1:end-1).*K(:,2:end)).^0.5 .* (diff(P,1,2)./h);
u([1 end],:) = u([end-1 2],:);

% update z-reference velocity
Div_tz = diff(tzz(:,2:end-1),1,1)./h + diff(txz,1,2)./h;

res_W(:,2:end-1) = - Div_tz + diff(P(:,2:end-1),1,1)./h + diff(p(:,

res_W([1 end],:) = [sum(res_W([1 end],:),1)./2;sum(res_W([1 end],:)
res_W(:,[1 end]) = res_W(:,[end-1 2]);

W = Wi - alpha.*res_W.*dtW + beta.*(Wi-Wii);

% update x-reference velocity
Div_tx  = diff(txx(2:end-1,:),1,2)./h + diff(txz,1,1)./h;

res_U(2:end-1,:) = - Div_tx + diff(P(2:end-1,:),1,2)./h + diff(p(2:

res_U([1 end],:) = res_U([end-1 2],:);
res_U(:,[1 end]) = [sum(res_U(:,[1 end]),2)./2,sum(res_U(:,[1 end])

U = Ui - alpha.*res_U.*dtU + beta.*(Ui-Uii);

% update reference pressure
Div_V(2:end-1,2:end-1) = diff(U(2:end-1,:),1,2)./h + diff(W(:,2:end
Div_v(2:end-1,2:end-1) = diff(u(2:end-1,:),1,2)./h + diff(w(:,2:end

res_P = Div_V + Div_v;

res_P([1,end],:) = res_P([end-1,2],:);
res_P(:,[1,end]) = res_P(:,[end-1,2]);

P = Pi - alpha.*res_P.*dtP + beta.*(Pi-Pii);

% update liquid evolution equation (enforce min/max limits on f)

flxdiv_fromm;  % upwind-biased advection/compaction term for liquid

res_f = (f-fo)./dt - (theta.*Div_fV + (1-theta).*Div_fVo);

res_f([1,end],:) = res_f([end-1,2],:);
res_f(:,[1,end]) = res_f(:,[end-1,2]);

if ~mod(step,nop); res_f = res_f - mean(res_f(:)); end

f = fi - alpha.*res_f.*dt/50;
f = max(0.001/f0,min(0.999/f0, f ));

% check and report convergence every nup iterations
if ~mod(it,nup); report; end
```

What is Science?

pursuit of understanding by observations, experiments, analysis, models, theory

The Scientific Method

- make observations
- design experiments
- analyse data
- develop models
- formulate theory



Observations

qualitatively or quantitatively record events or processes in natural world

Observations in Geosciences

- collect rock samples
- map geological units
- measure water temperature
- measure topography
- analyse rock composition
- monitor volcanic gas emissions
- …

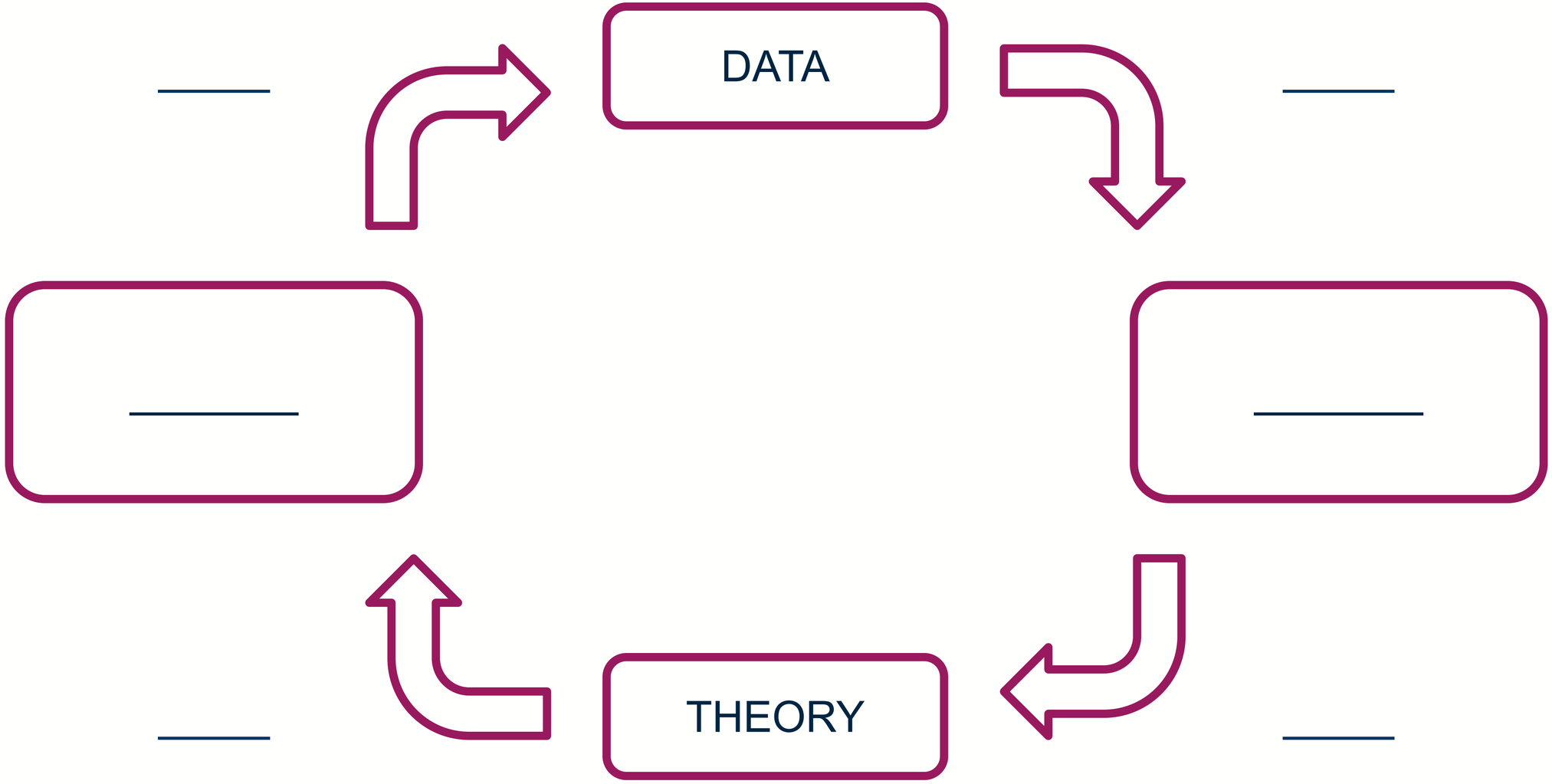
Experiments

repeatedly record event or process under controlled conditions

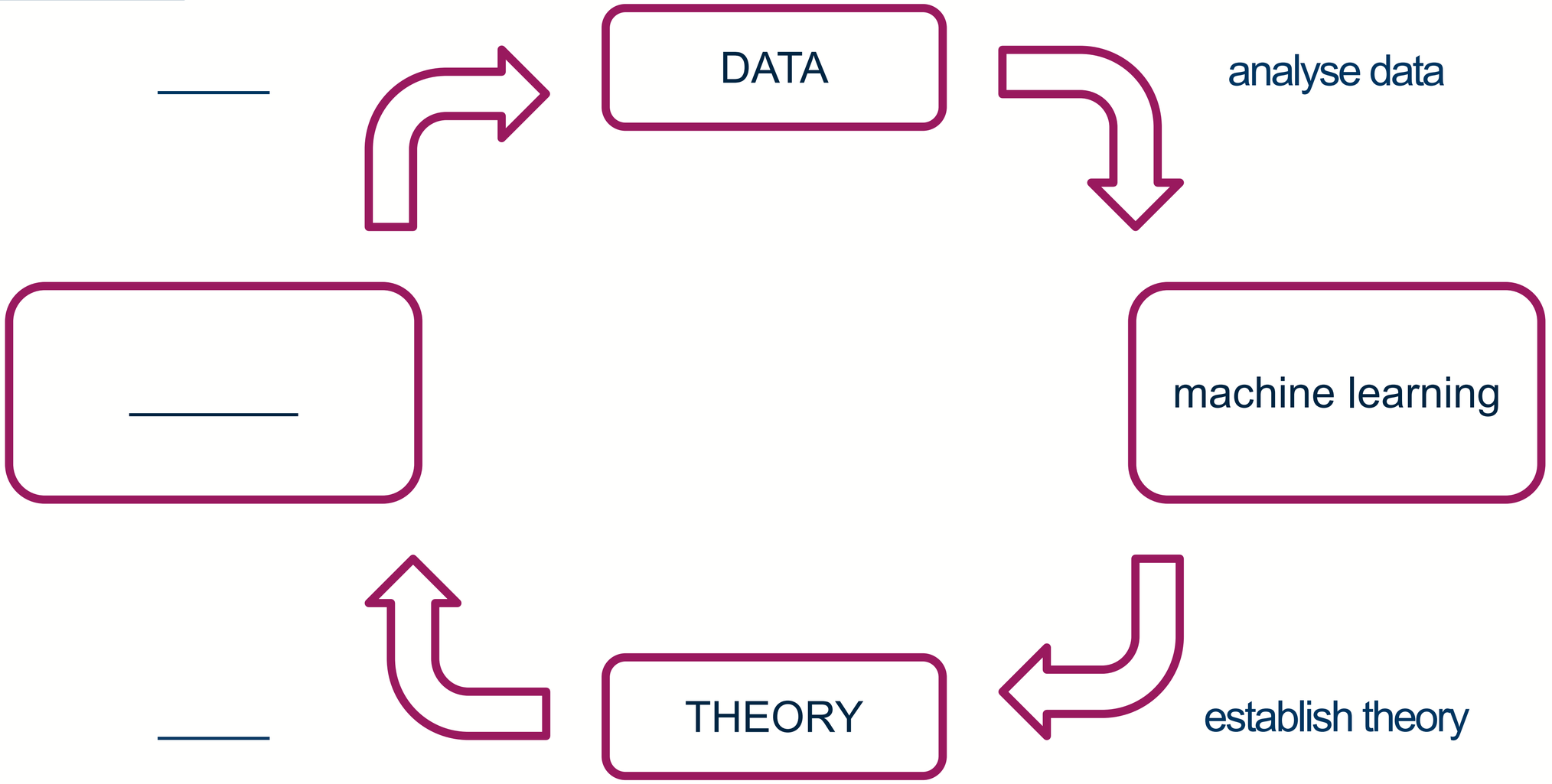
Experiments in Geosciences

- melt rock in furnace
- deform rock in piston cylinder
- make waves in a tank
- create river on a table
- …

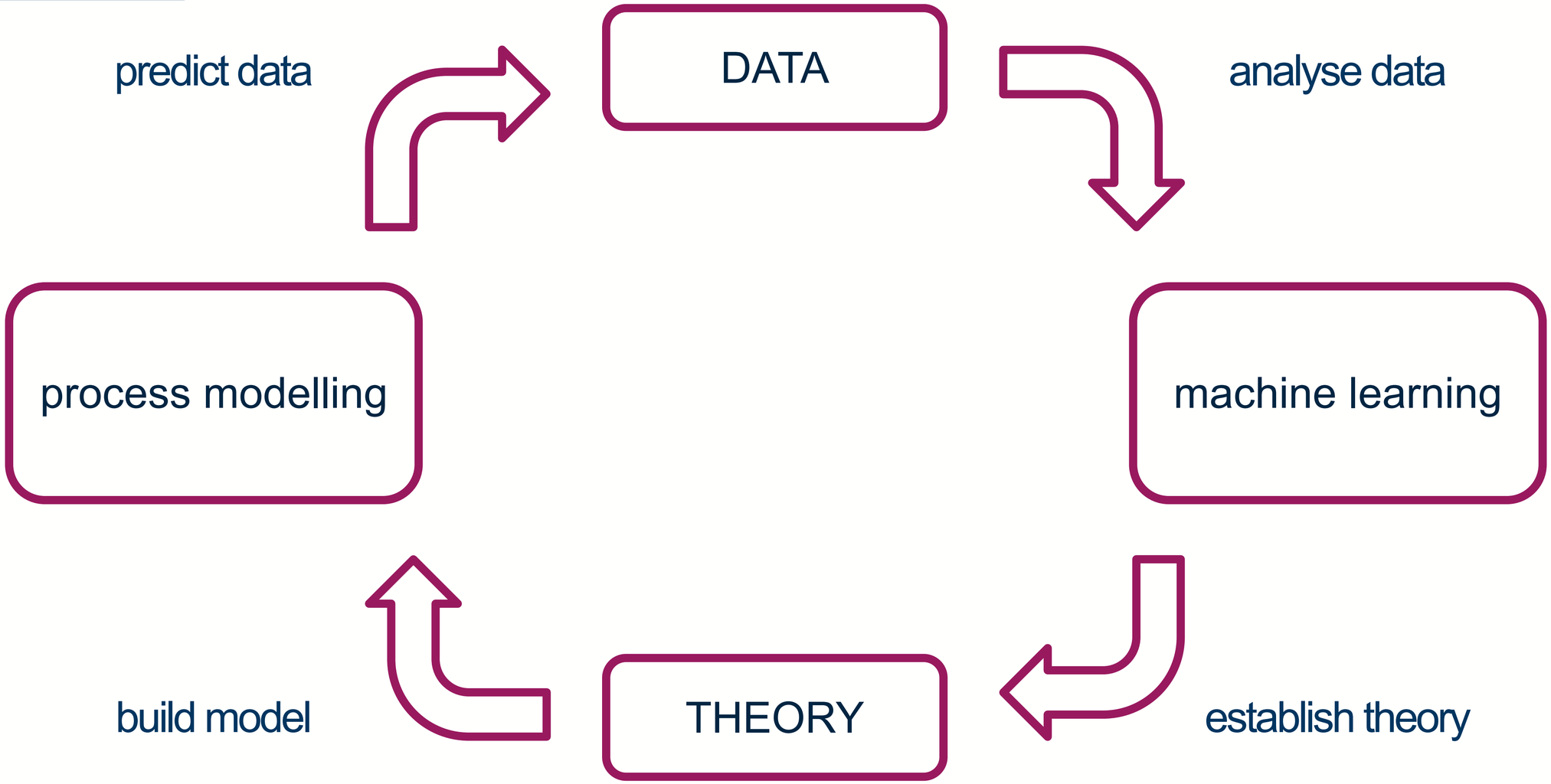
Comp Data Analysis | Data vs. Theory



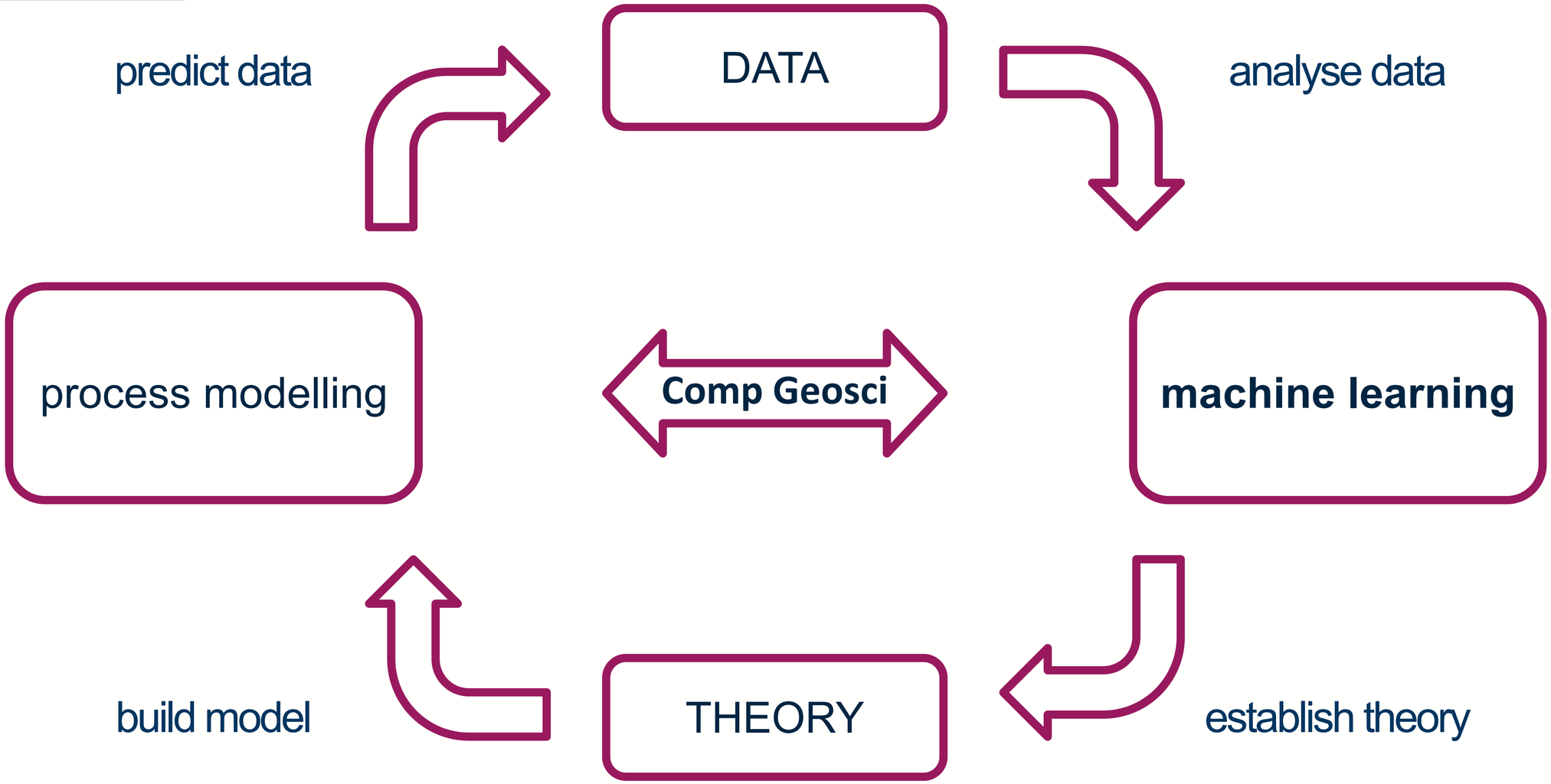
Comp Data Analysis | Data vs. Theory



Comp Data Analysis | Data vs. Theory



Comp Data Analysis | Data vs. Theory



Comp Data Analysis | Machine Learning

Data-driven

- start with observations, find model to fit data

Empirically-minded

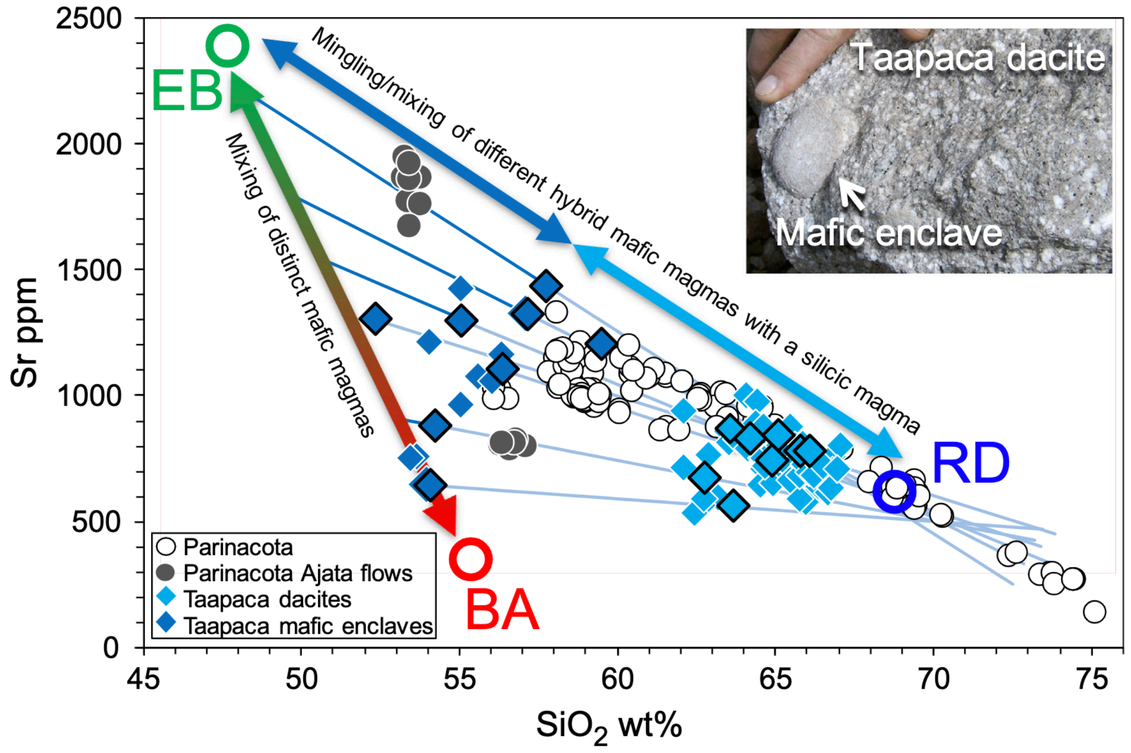
- good fit more important than understanding process

Non-uniqueness

- more than one model fits data, several interpretations reasonable

Examples

- line fitting, clustering analysis, tomographic imaging



End-member extraction for lava compositions (Blum-Oeste & Wörner, Terra Nova, 2016)

Comp Data Analysis | Machine Learning

Supervised Learning

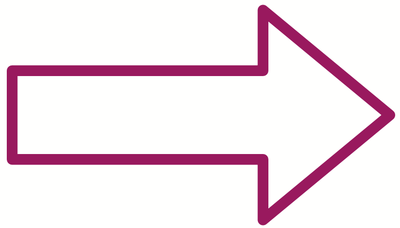
- *deals with labelled data*
- identifies how data are correlated with labels
- example:

labelled images

label: kitten

label: puppy

supervised



learning

predict label for images

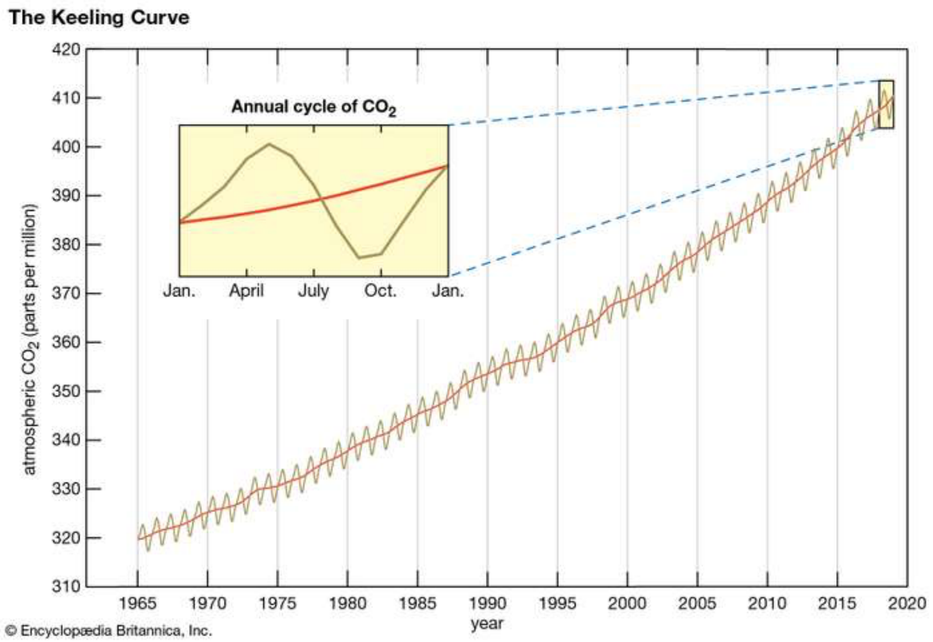
=> puppy

=> kitten

Comp Data Analysis | Machine Learning

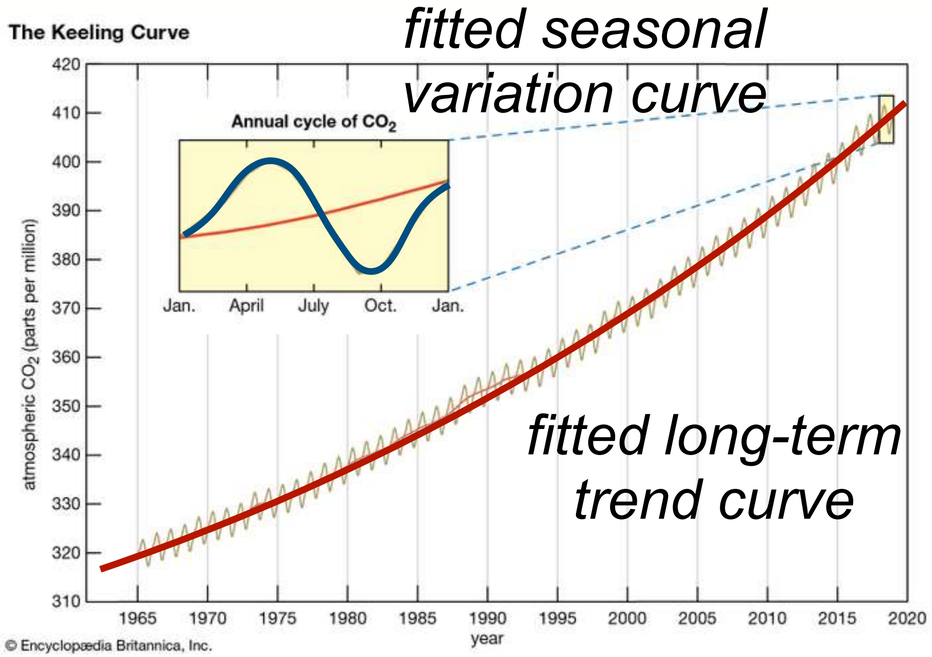
Supervised Learning

- *deals with labelled data*
- identifies how data are correlated with labels
- example:



supervised

learning

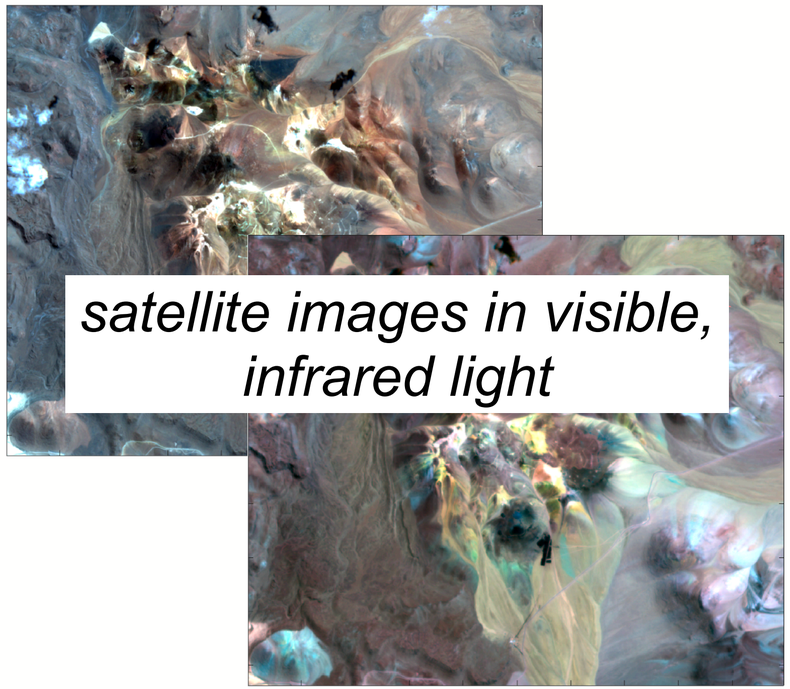


data = CO_2; label = date

Comp Data Analysis | Machine Learning

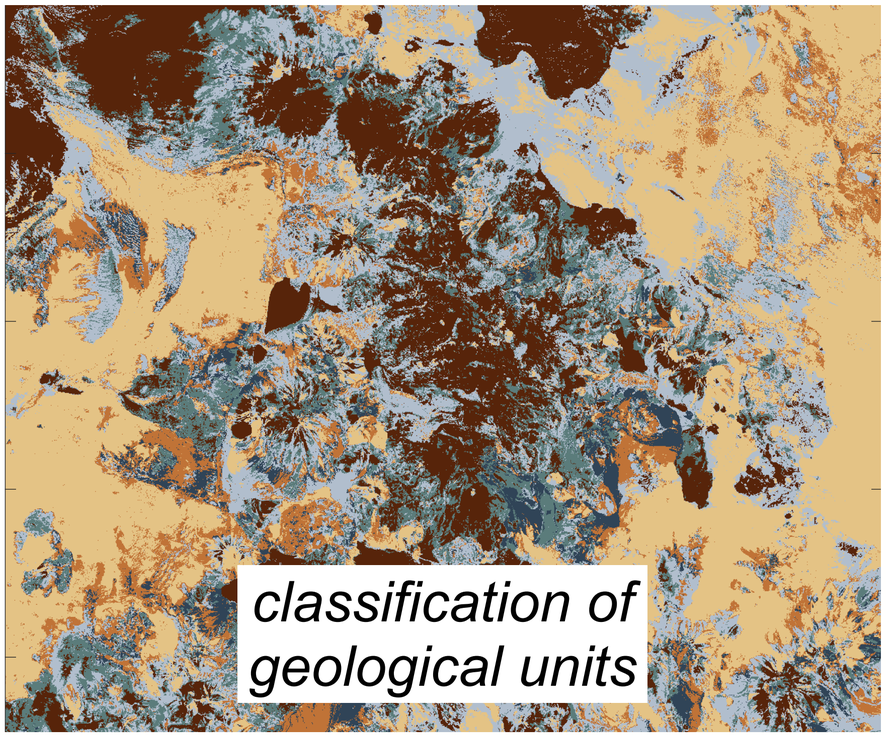
Unsupervised Learning

- *deals with unlabelled data*
- identifies similarities in data by which it can be classified
- example:



unsupervised

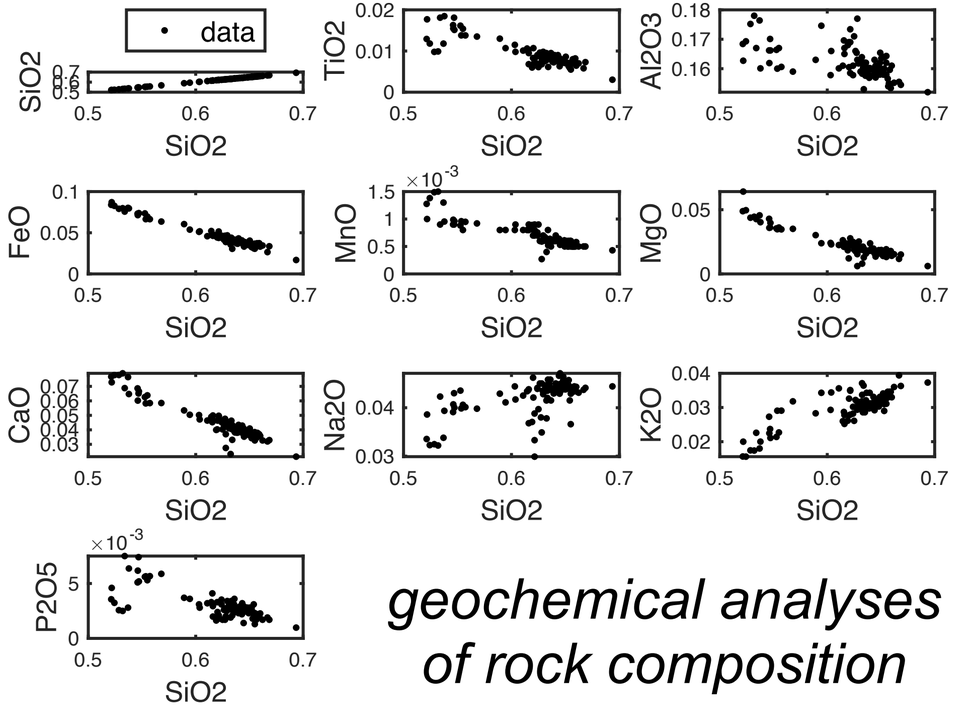
learning



Comp Data Analysis | Machine Learning

Unsupervised Learning

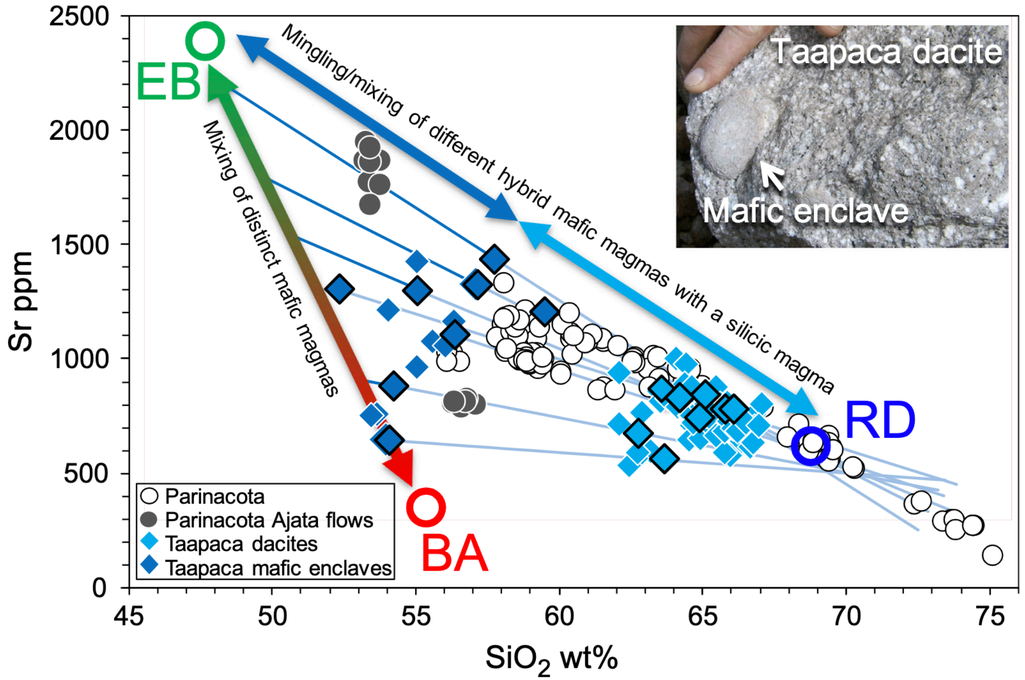
- *deals with unlabelled data*
- identifies similarities in data by which it can be classified
- example:



geochemical analyses of rock composition

unsupervised

learning



end-member compositions

Comp Data Analysis | Machine Learning

Training Data

- part of a data set used to ***train*** or ***calibrate*** a machine learning model

Trained Model

- the output of running a machine learning routine over training data, identifies labels, correlations, similarity clusters, etc. in large data sets

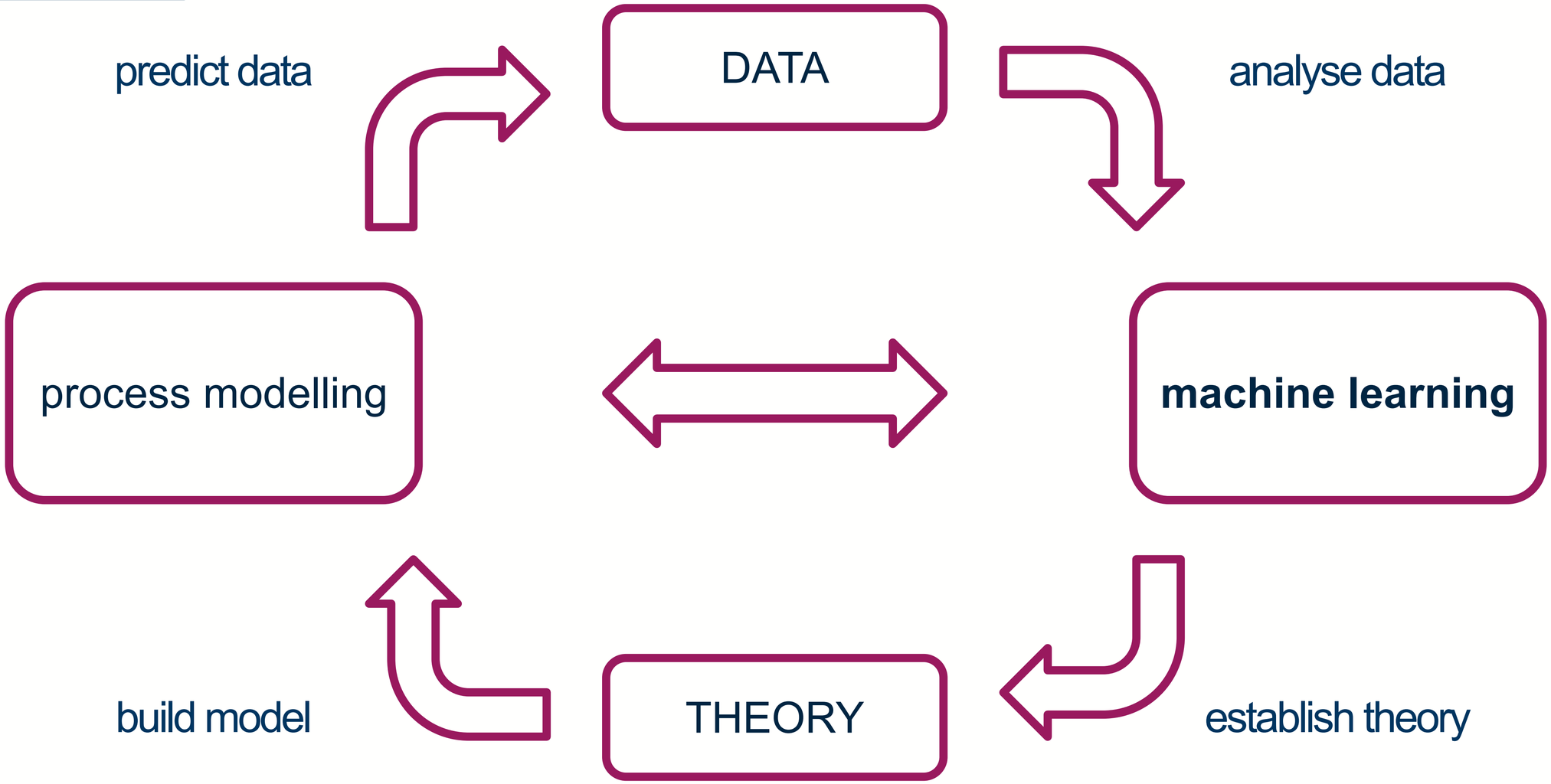
Testing Data

- separate part of a data set used to ***validate*** a trained machine learning model

Model Prediction

- apply a trained model on an entirely new data set to predict labels, classifications, etc.

Comp Data Analysis | Summary



WKSHP II | Intro to Comp Data Analysis

Activity | Introduction to Data Analysis with Python

Load a data set into Python code

- CO2 record from Mauna Loa observatory
- load text file containing data table
- handle, plot, and analyse data as Python array

Get started with basic machine learning

- fit a straight line through the data: does it fit?
- fit more complex model to data to improve fit
- decompose model: seasonal vs. long-term trend
- use calibrated model to predict future data