Thursday, September 26, 2019

Explainable AI and dimensionality reduction | AI Village learning session #2

Explainable AI, Cognitive Science and Culture:

Towards a transparent, democratic and secular AI

Harald Martens

Founder & research leader, Idletechs AS,

www.idletechs.com, Pirsenteret 2.etc.

Prof. emerit. Big Data Cybernetics, NTNU

Guest prof. Macau U. of Science & Technology

Thursday, September 26, 2019

Explainable AI and dimensionality reduction | AI Village learning session #2

meetup

Explainable AI, Cognitive Science and Culture:

Towards a transparent, democratic and secular AI

PCA and bi-linear data modelling

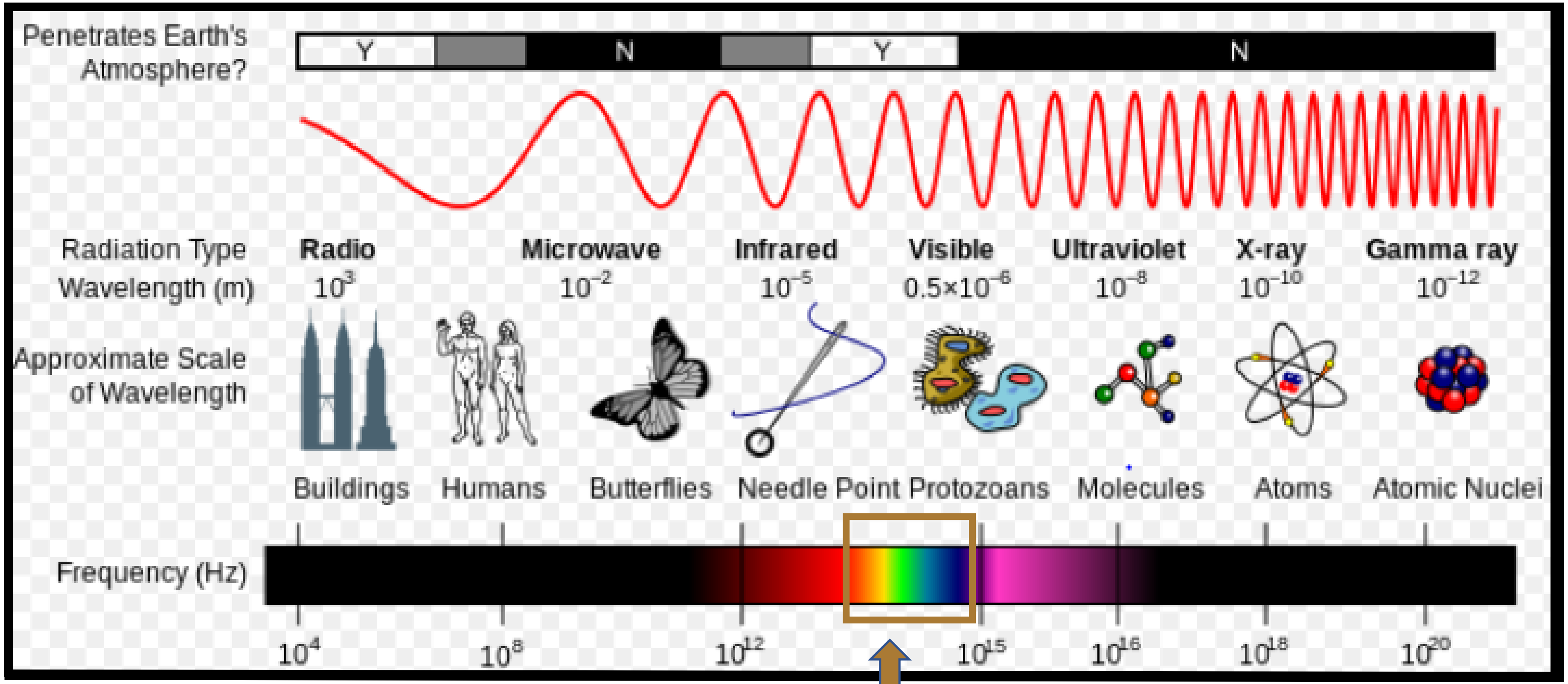
A toolbox for discovering the real world

idletechs

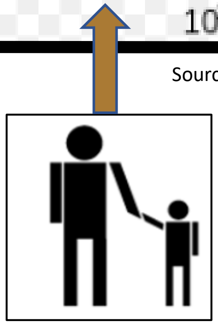
Outline:

- 30 min:
  - What is soft modelling?
  - What is PCA, and how can it be used
- 10 min break
- 30 min
  - Explainable AI
  - Bilinear PCA extensions
- Discussion

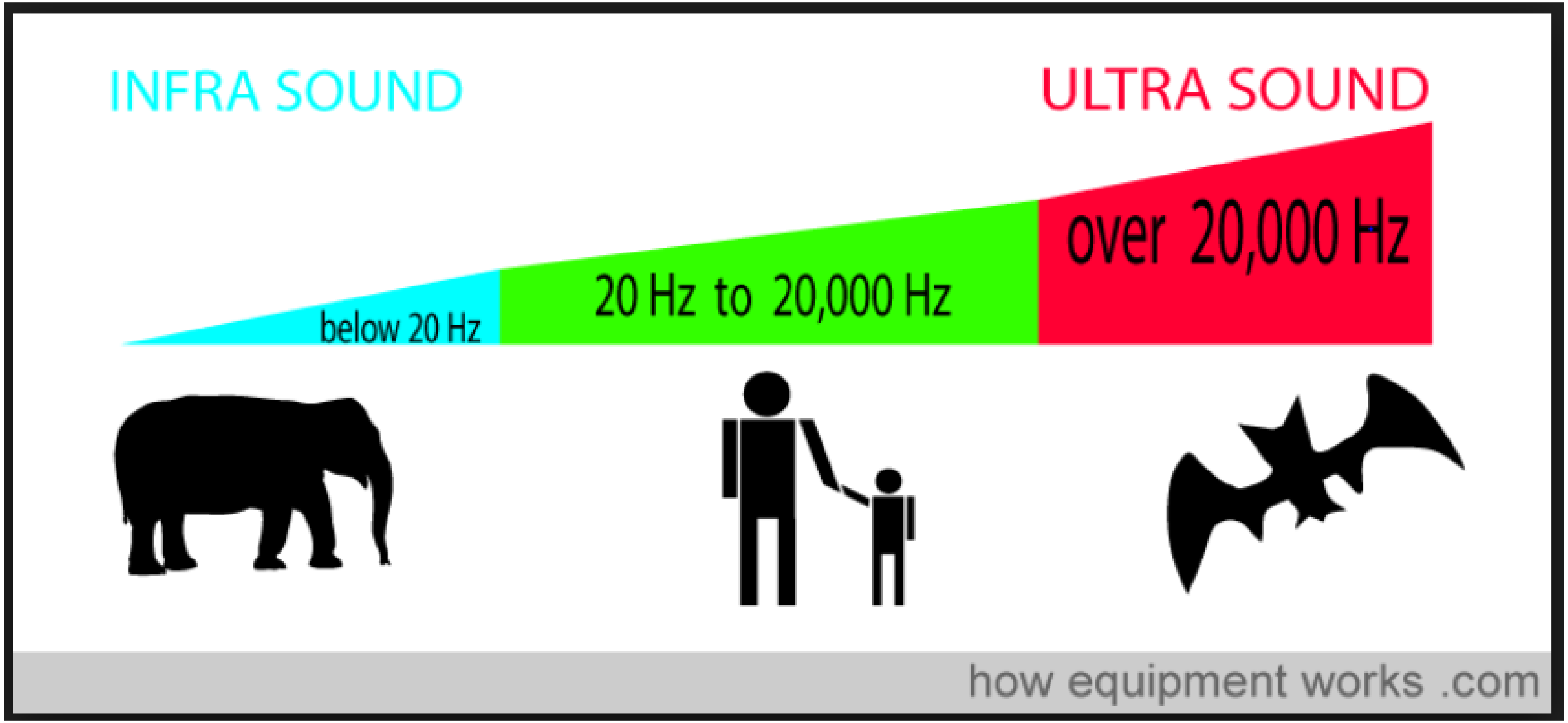
Vårt fargesyn er bra, men begrenset



Source: http://hubblesite.org/reference_desk/faq/answer.php.id=70&cat=light



Vår hørsel er bra, men begrenset



Utvid sansene

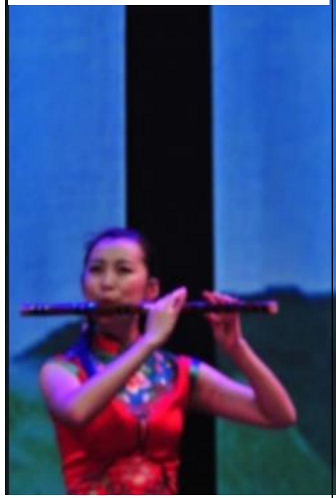
Trad. instruments:
1 channel

Utvid sansene

Trad. instruments:
1 channel

2 channels

Utvid sansene



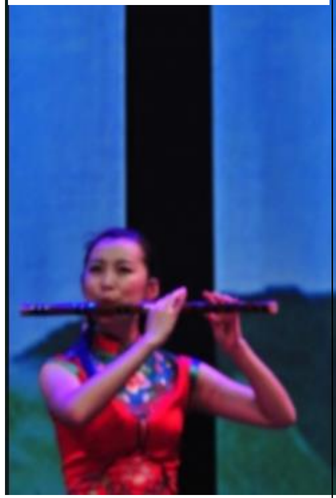
Trad. instruments:
1 channel

2 channels

A guitar:
6 string

12 strings

Utvid sansene



Trad. instruments:
1 channel

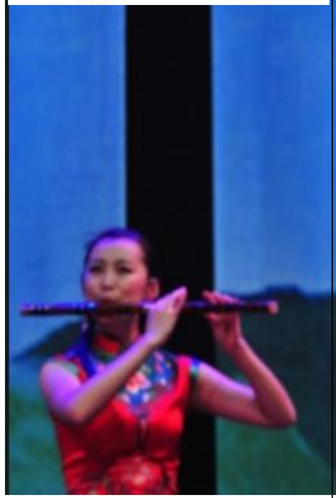
2 channels

A guitar:
6 string

12 strings

A grand piano : Lots of keys

Utvid sansene



Trad. instruments:
1 channel

2 channels

A guitar:
6 string

12 strings

A grand piano : Lots of keys

A band: 7 instruments

Utvid sansene

Trad. instruments:
1 channel

2 channels

A guitar:
6 string

12 strings

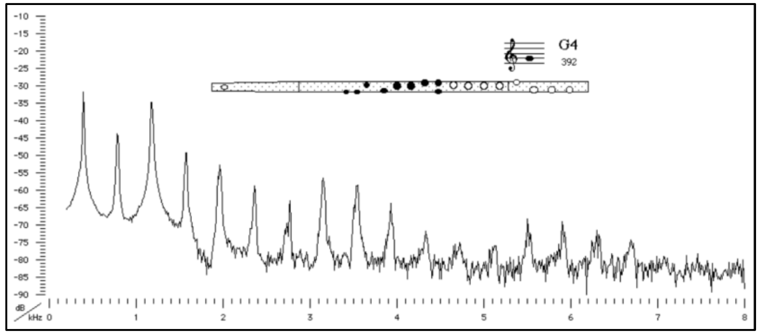
A grand piano : Lots of keys

A band: 7 instruments

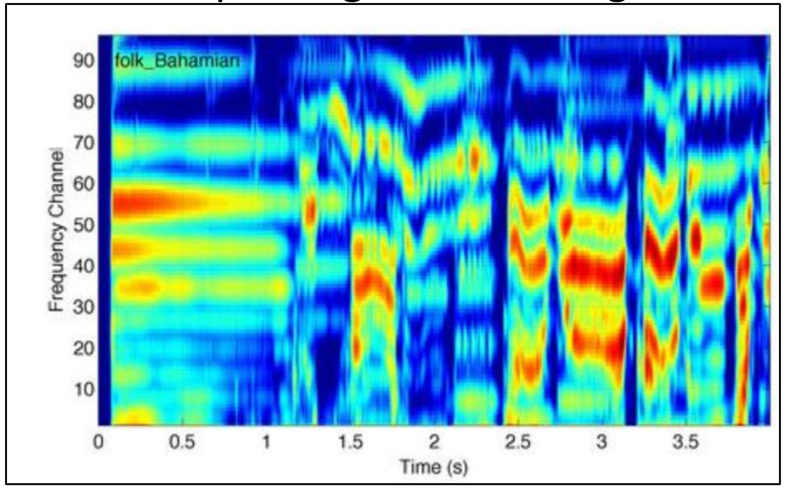
A symphony orchestra: 100 instruments

Utvid sansene

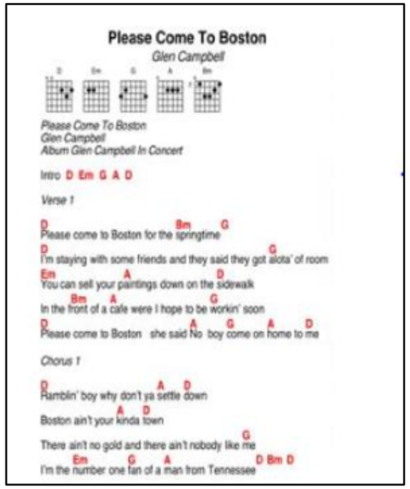
A frequency spectrum of flute, playing G



A spectrogram of a song



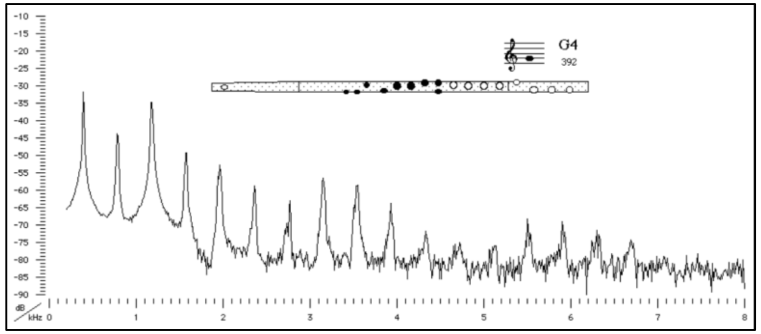
Compressed into written music and chords



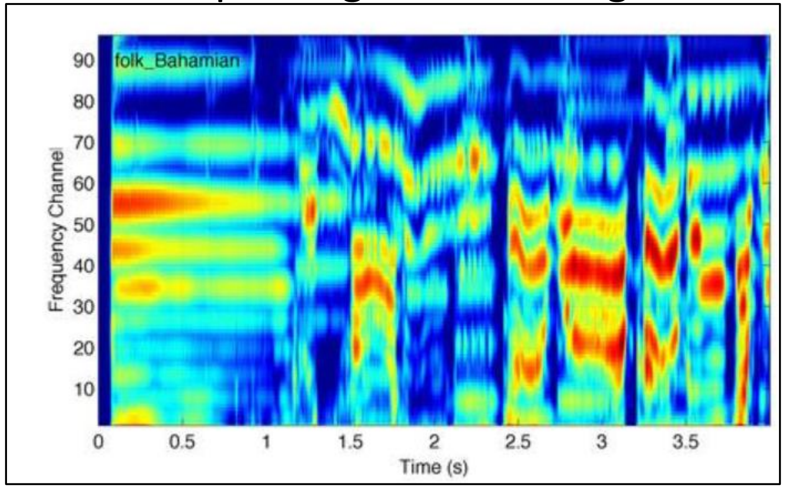
Further compressed into chords only

Utvid sansene

A frequency spectrum of flute, playing G



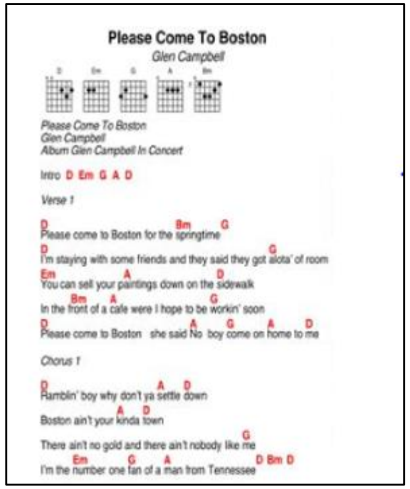
A spectrogram of a song



≈

Compressed into written music and chords

≈



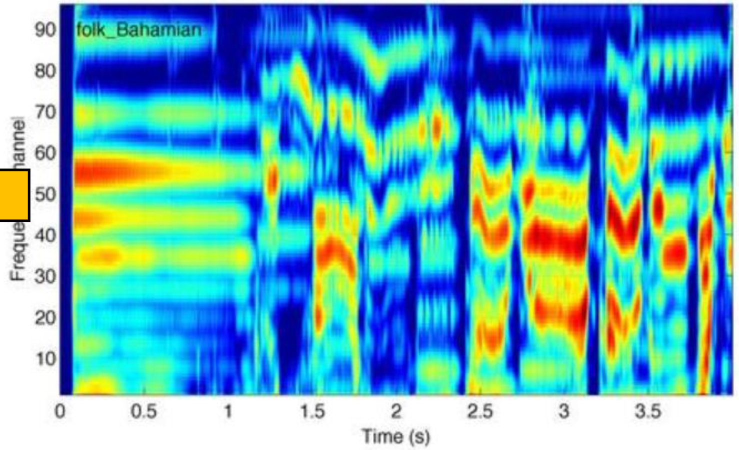
Further compressed into chords only

Utvid sansene

A cacophony of sounds

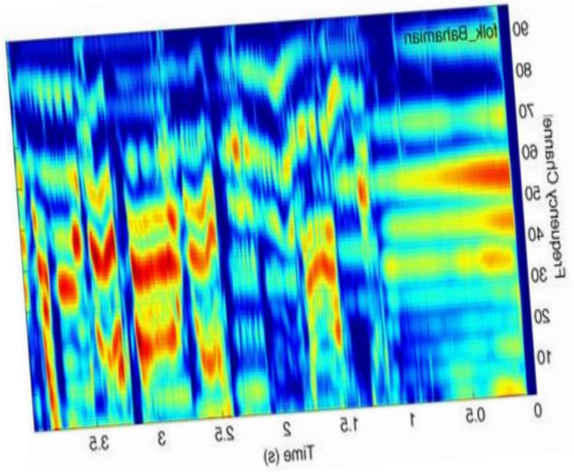
Human Perception and Interpretation

Data processing



Utvid sansene

A cacophony of sounds



Data processing

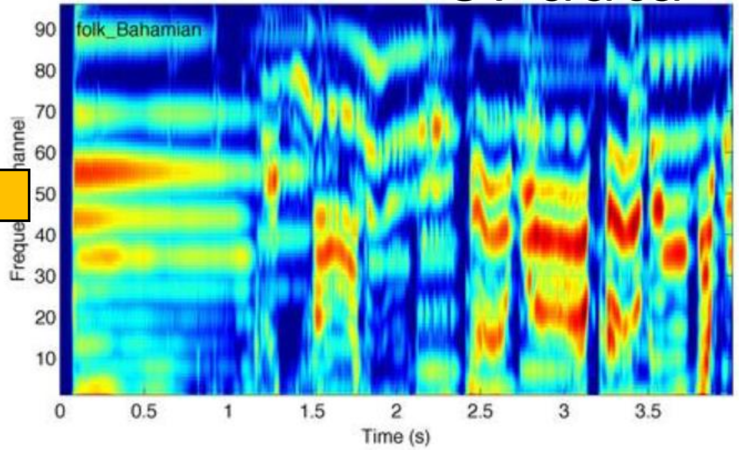
Human Perception and Interpretation

Utvid sansene

A cacophony of data

Human Perception and Interpretation

Data processing

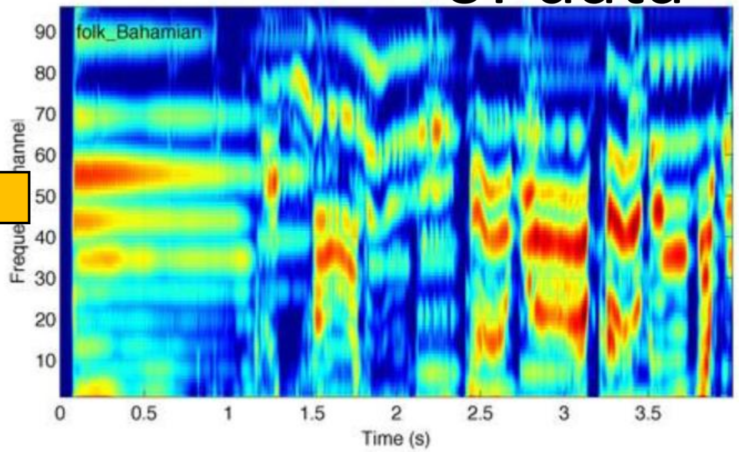


Utvid sansene

A cacophony of data

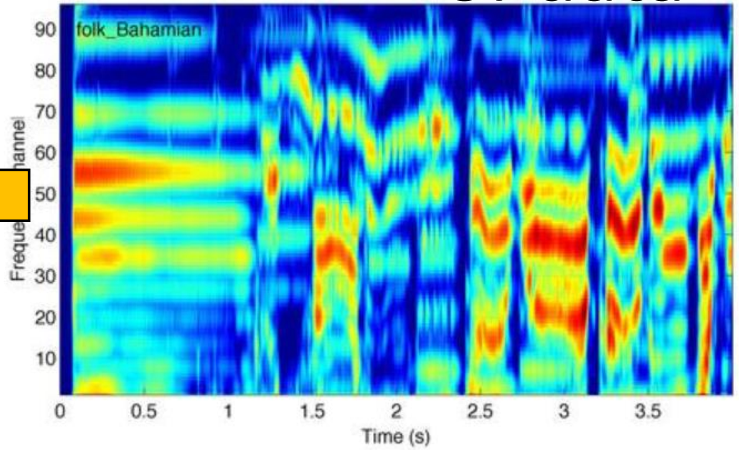
Human Perception and Interpretation

Data processing



Utvid sansene

A cacophony of data



Data processing

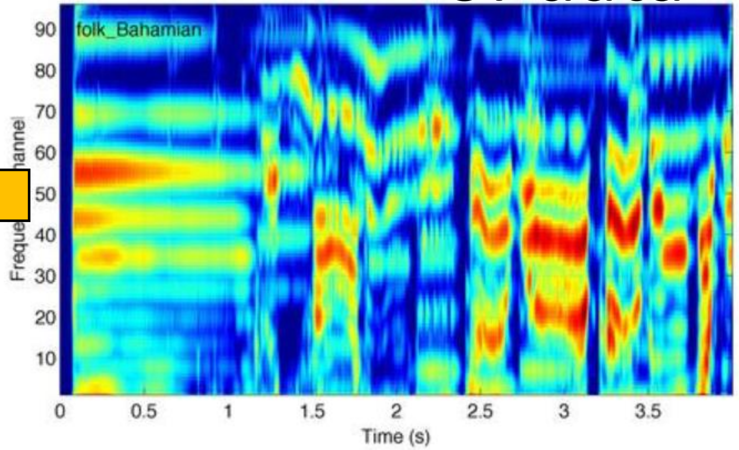
Human Perception and Interpretation

Utvid sansene

A cacophony of data

Human Perception and Interpretation

DIFFERENT Data processing

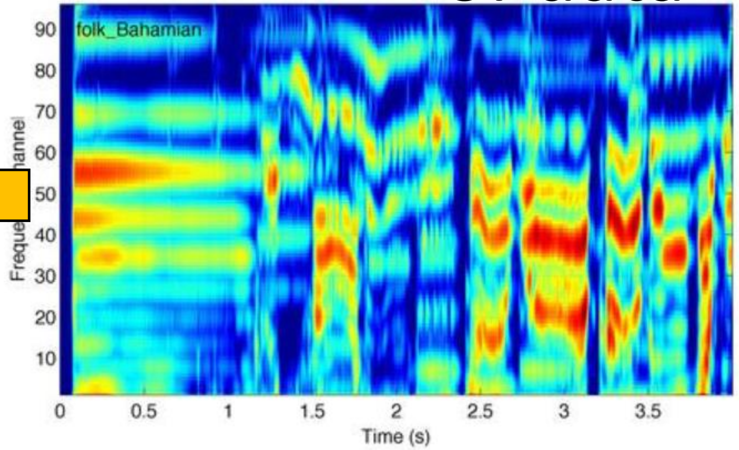


Utvid sansene

A cacophony of data

Human Perception and Interpretation

DIFFERENT Data processing

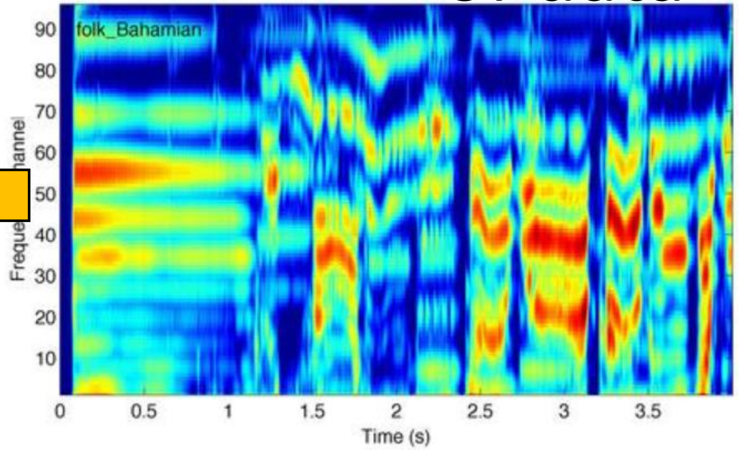


Utvid sansene

A cacophony of data

Human Perception and Interpretation

DIFFERENT Data processing

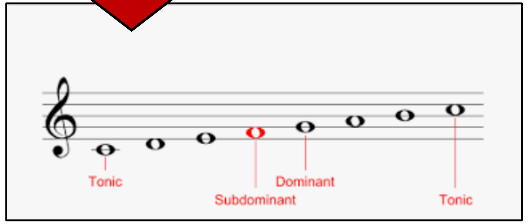


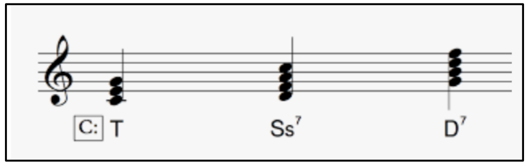
Utvid sansene

Human Perception and Interpretation

DIFFERENT Data processing

Expected structures

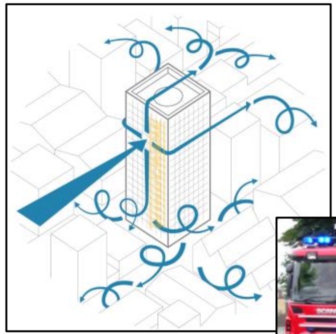




Prior, known phenomena,
Forming theory-based models
(e.g.: harmonic analysis in classical music)

+

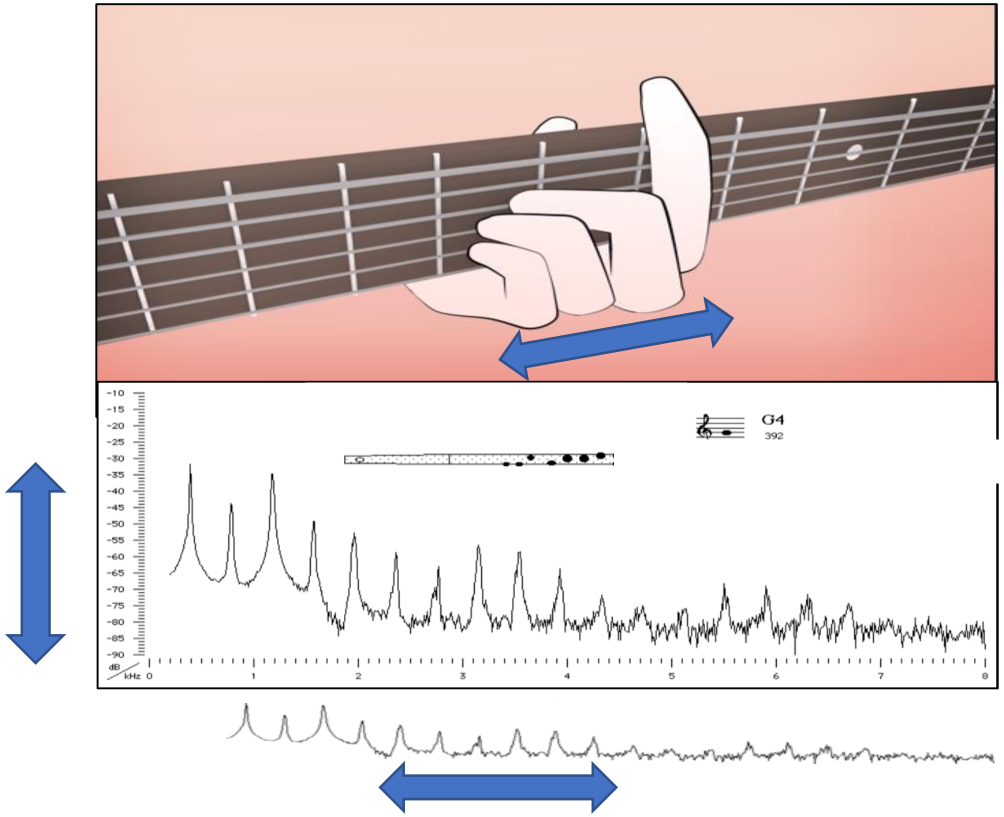
Unexpected structures



New phenomena
forming data-based models
(e.g. sound of hiss, mess & hickups…)

(+ «random» noise)

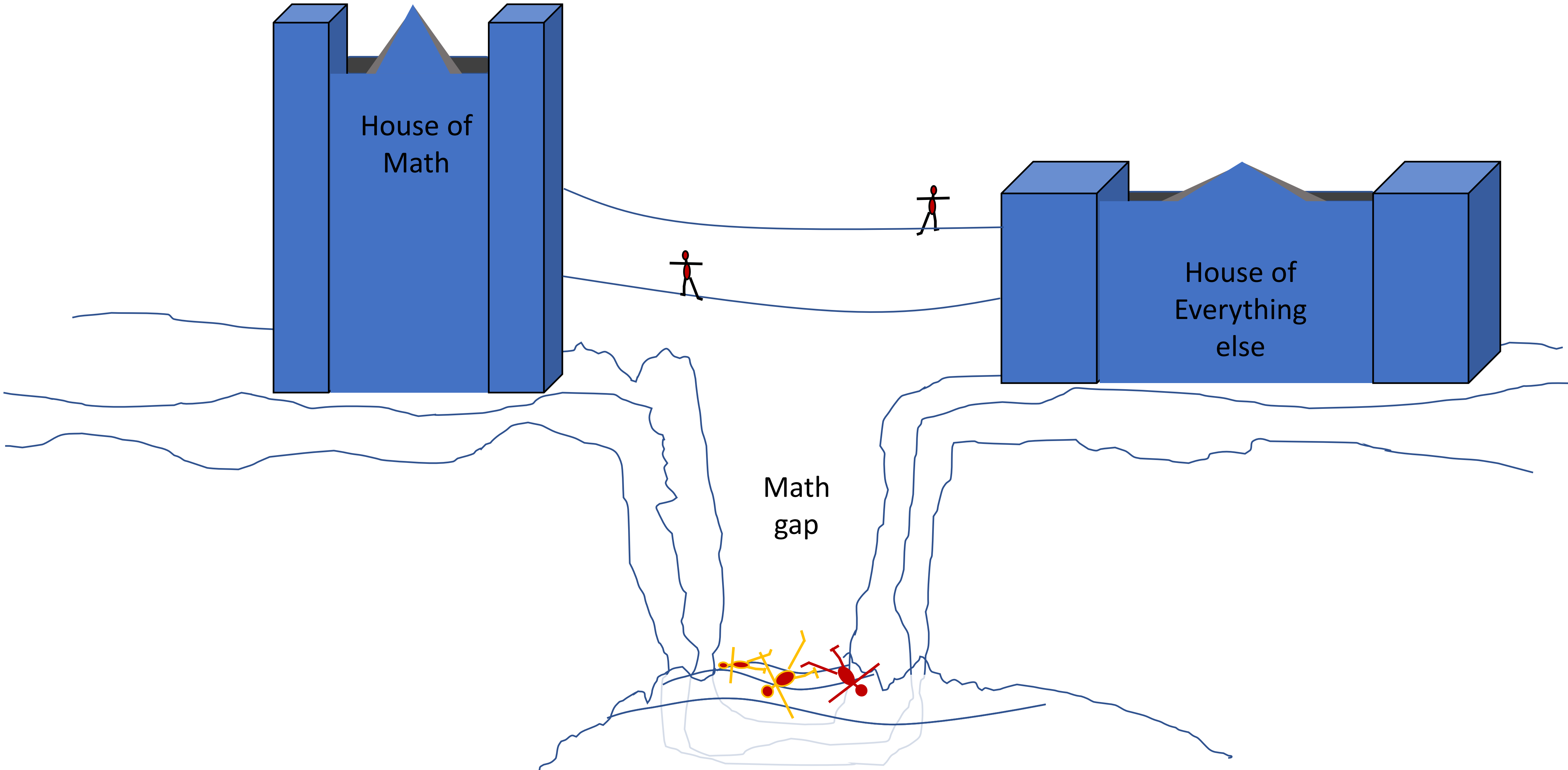
Two types of info in data from any instrument:



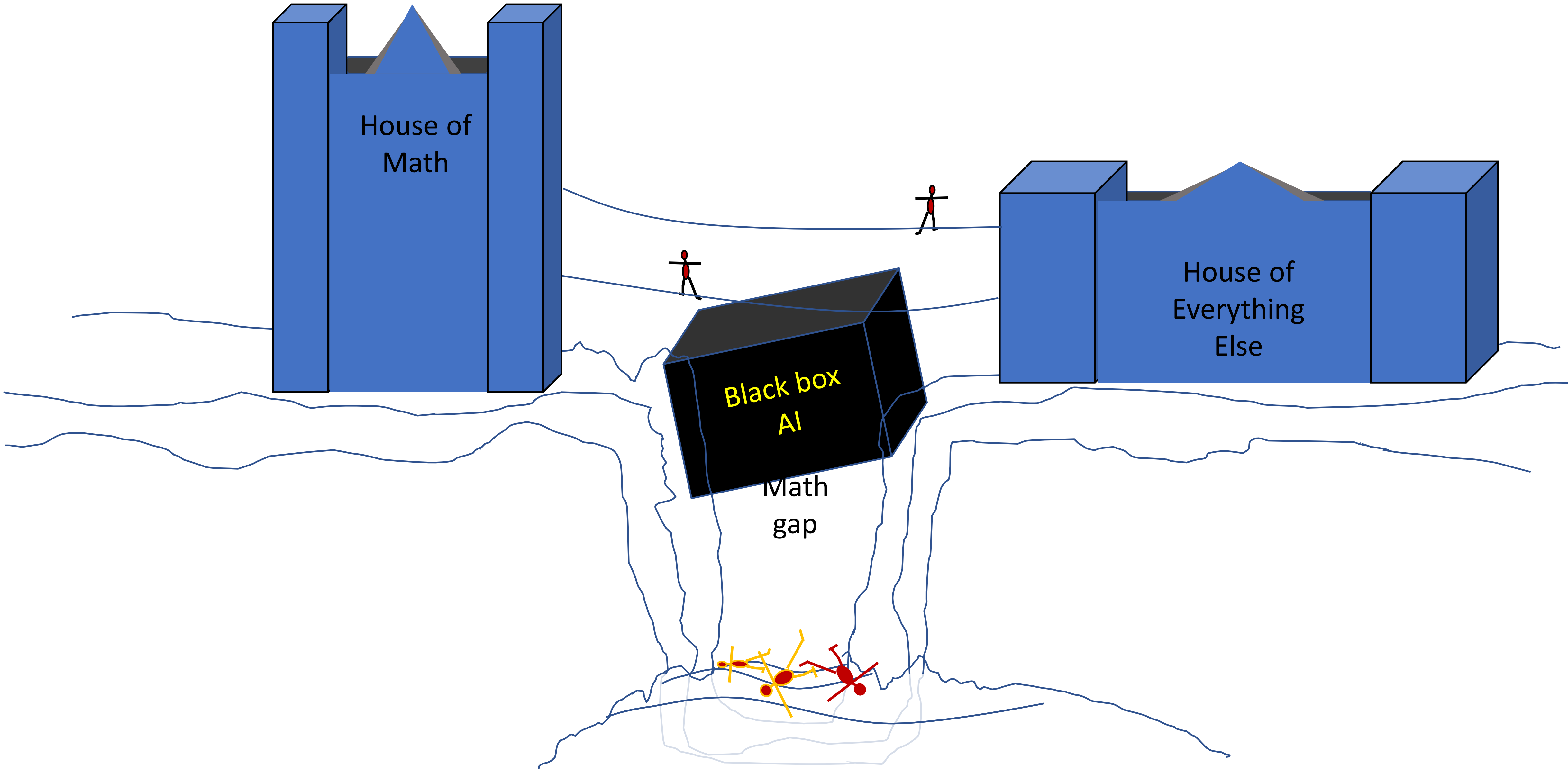
Ordinate: Amplitude:
How strong sound, which type of sound?

Abscissa: Pitch:
Which tone?
Which rotation rate of machinery?

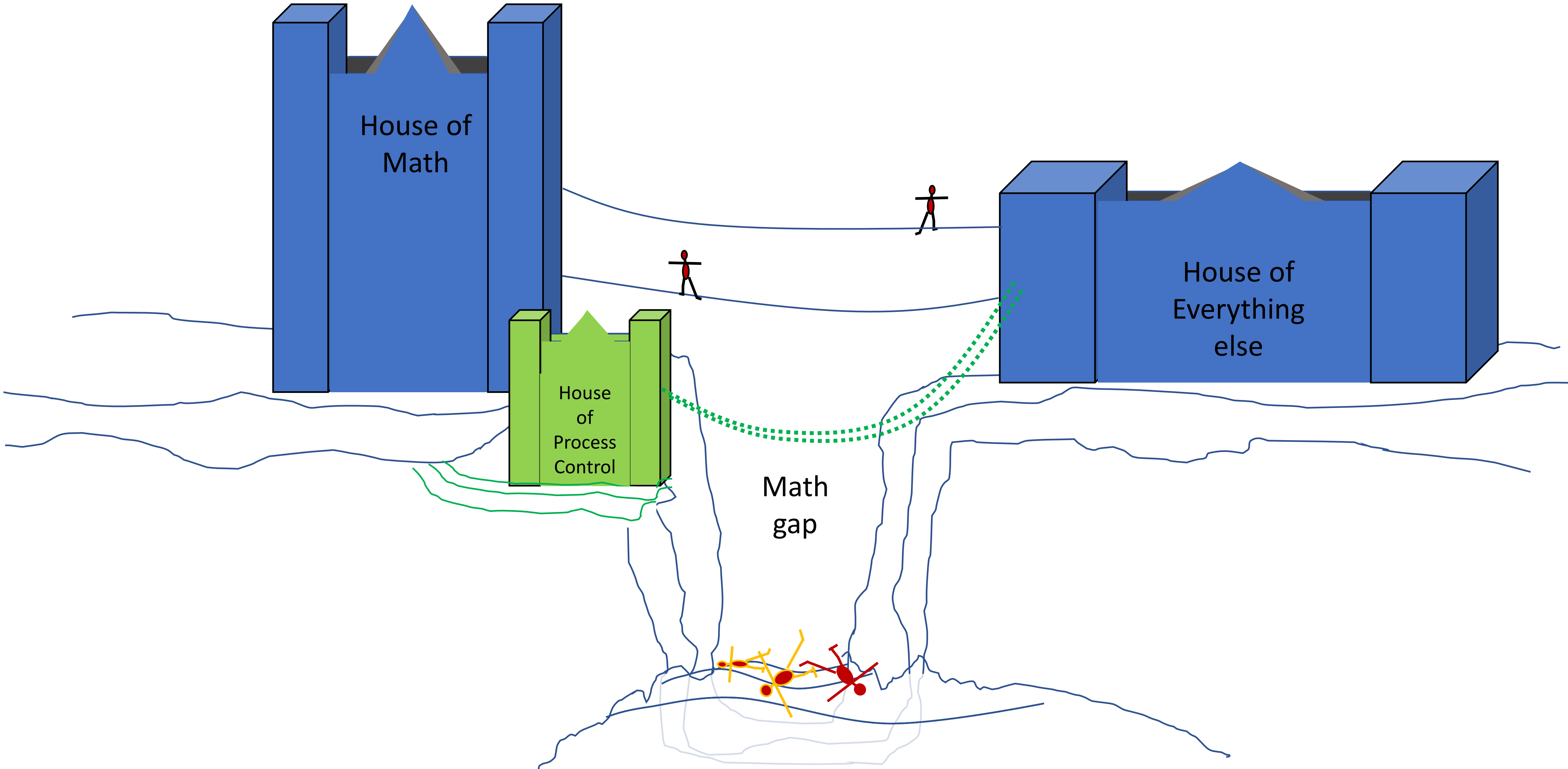
The Math Gap



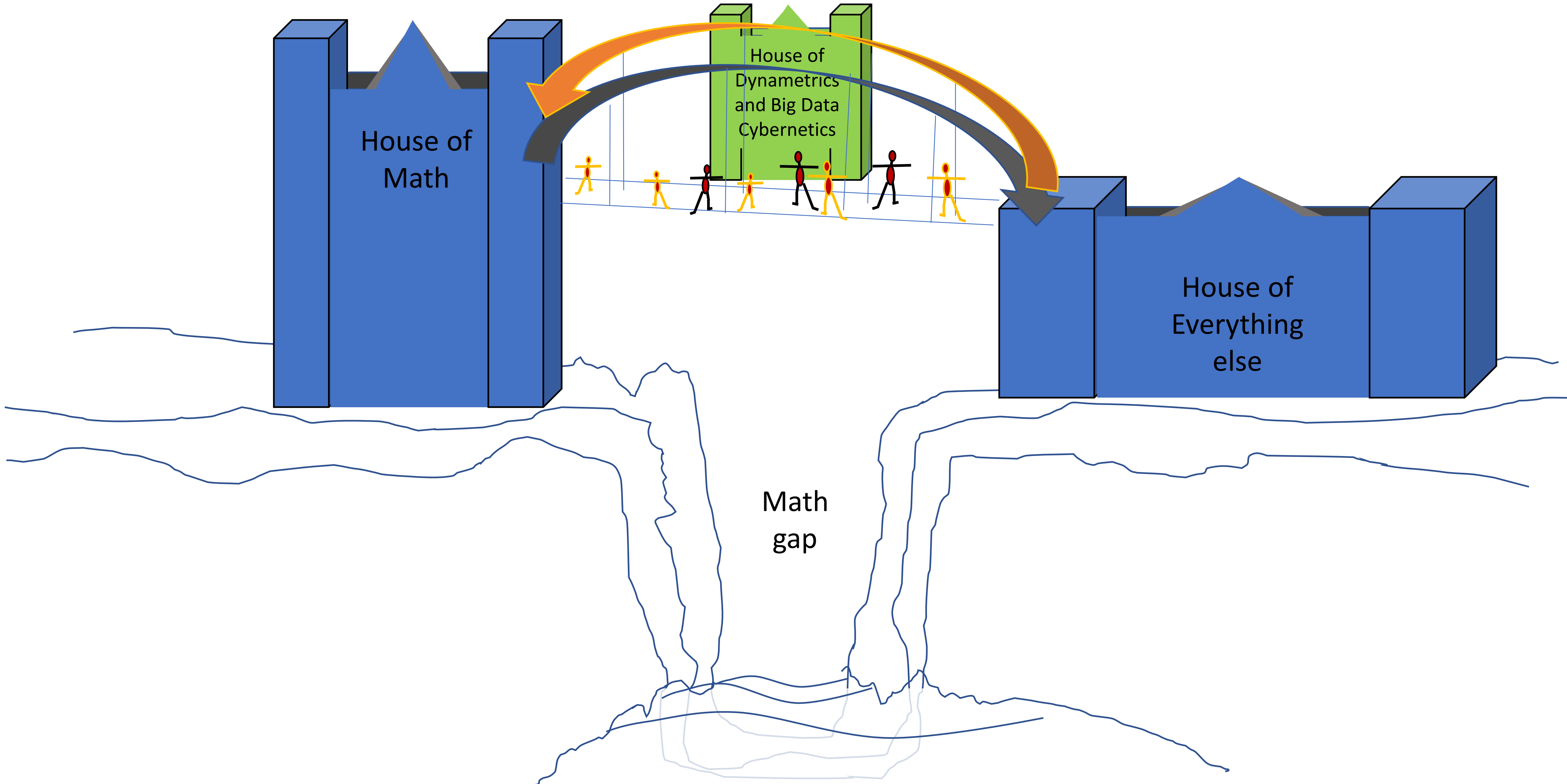
The Math Gap



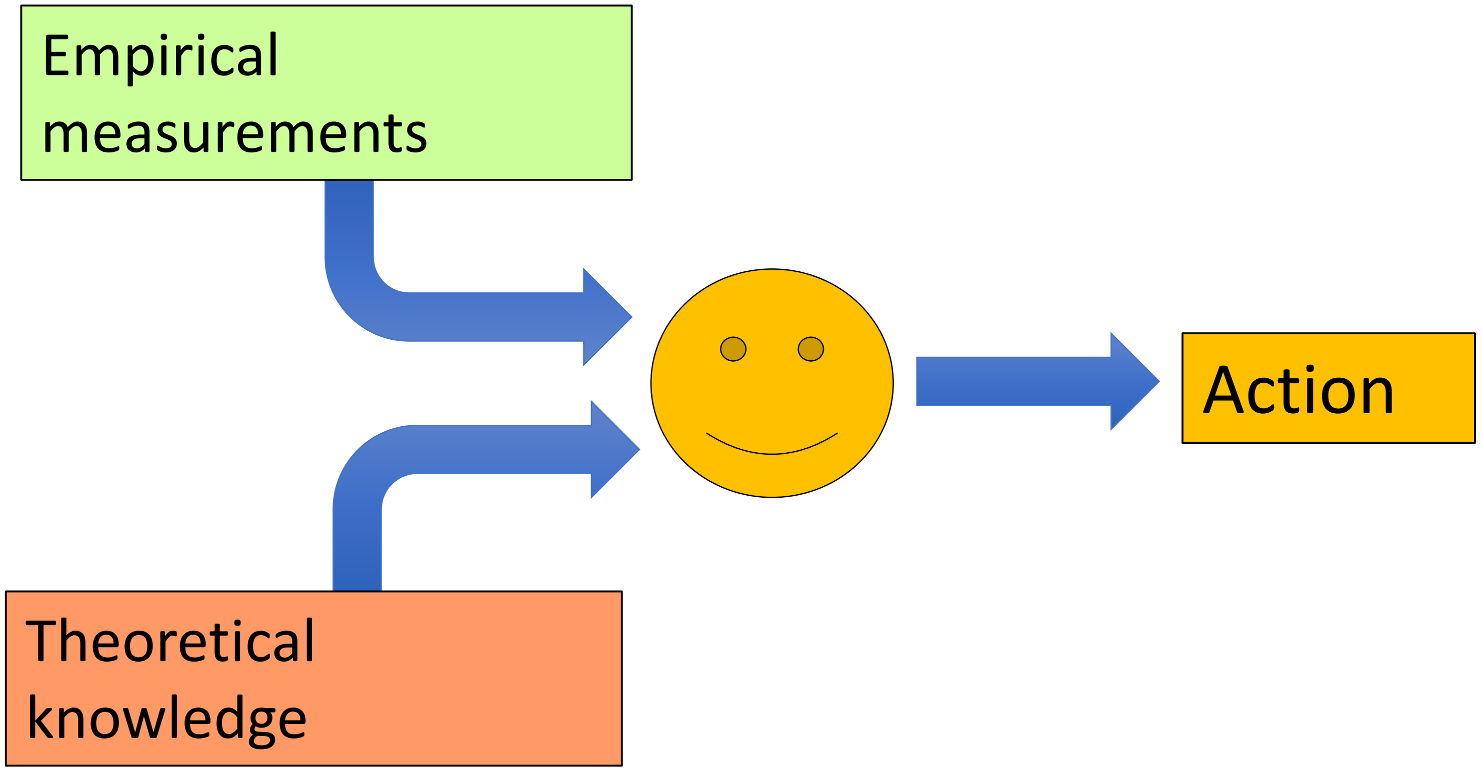
The Math Gap


House of
Math
House of
Everything
else
House
of
Process
Control
Math
gap

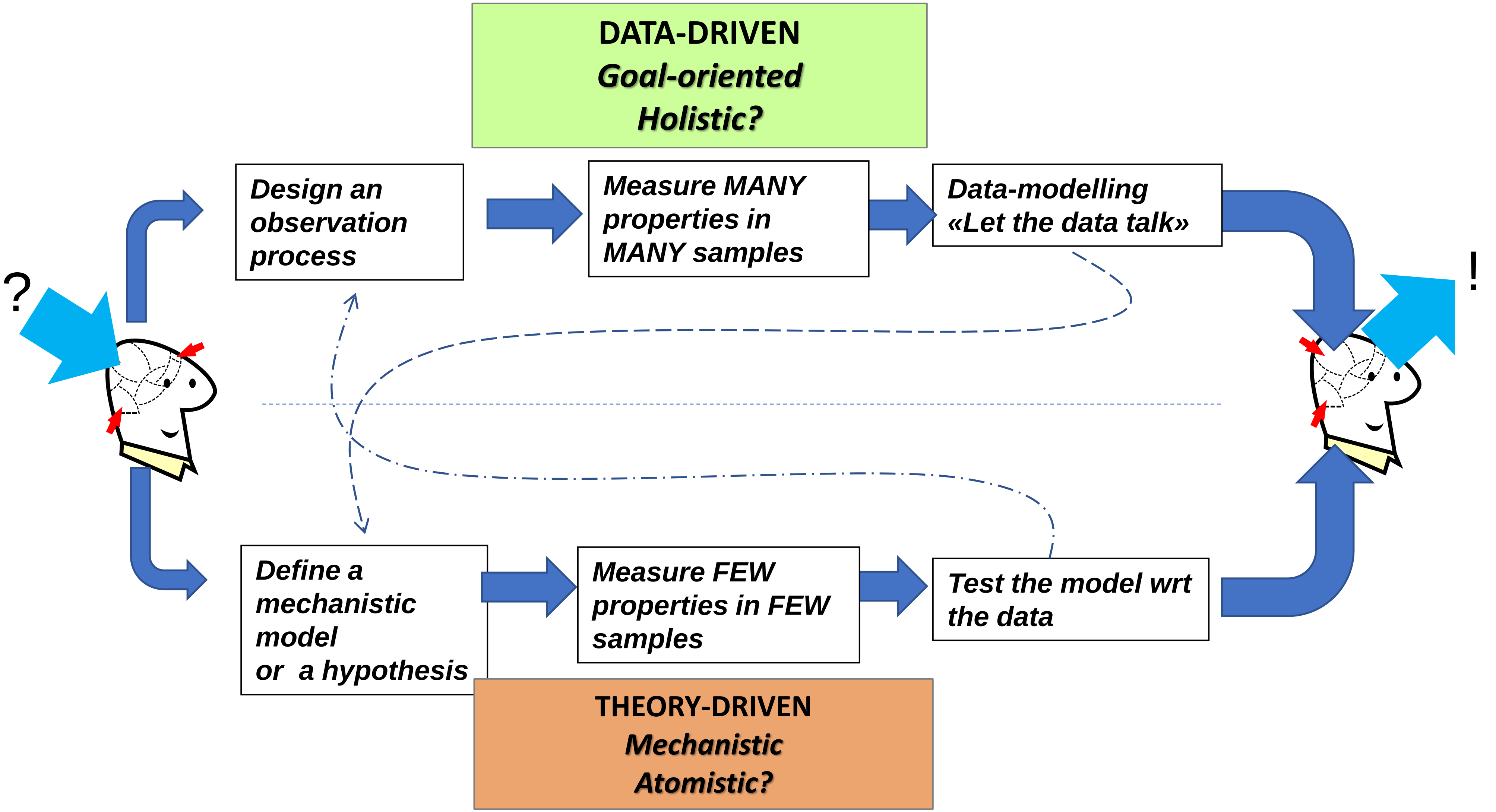
A two-way bridge across the math gap



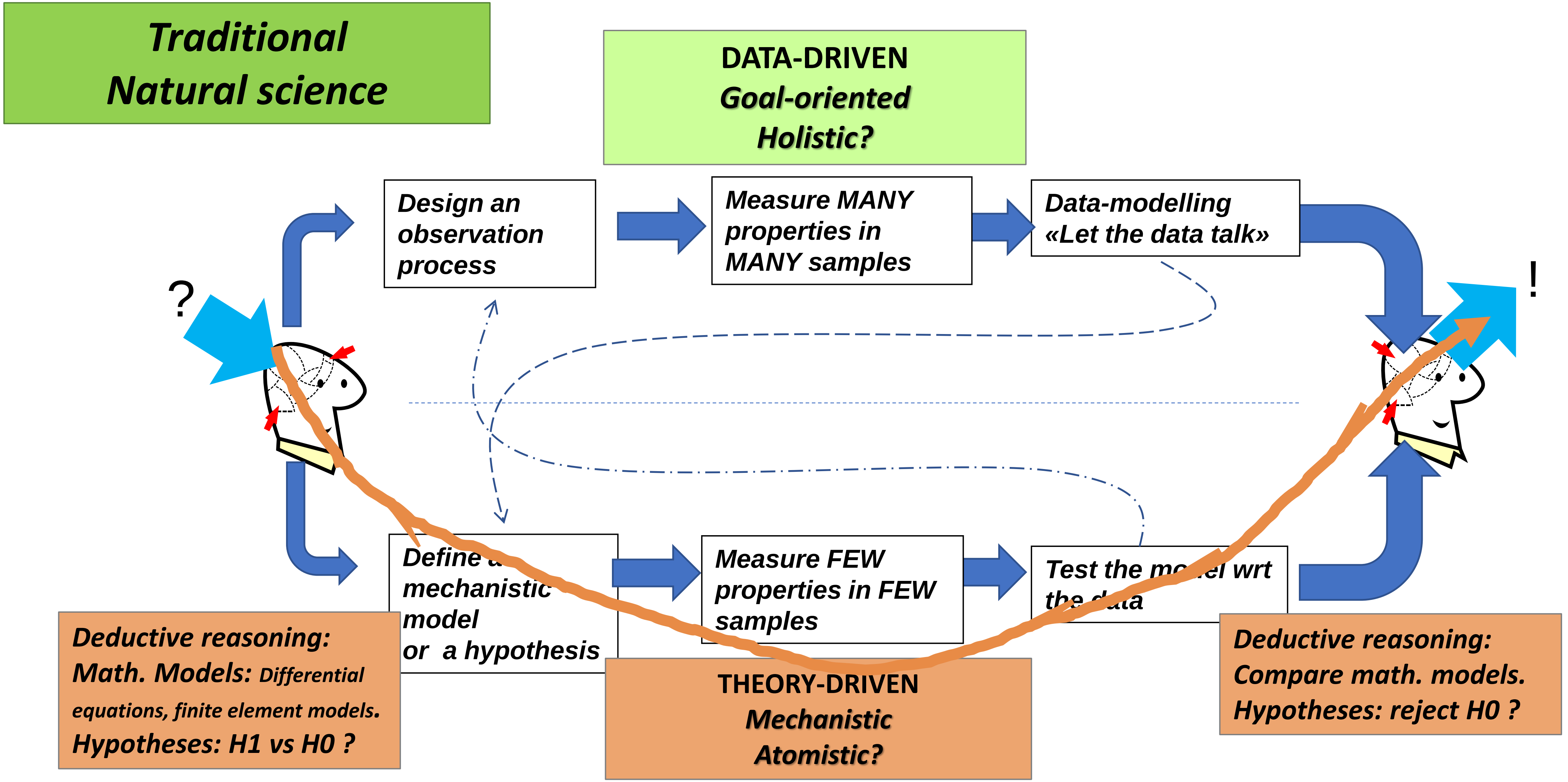
Source of knowledge at any given moment



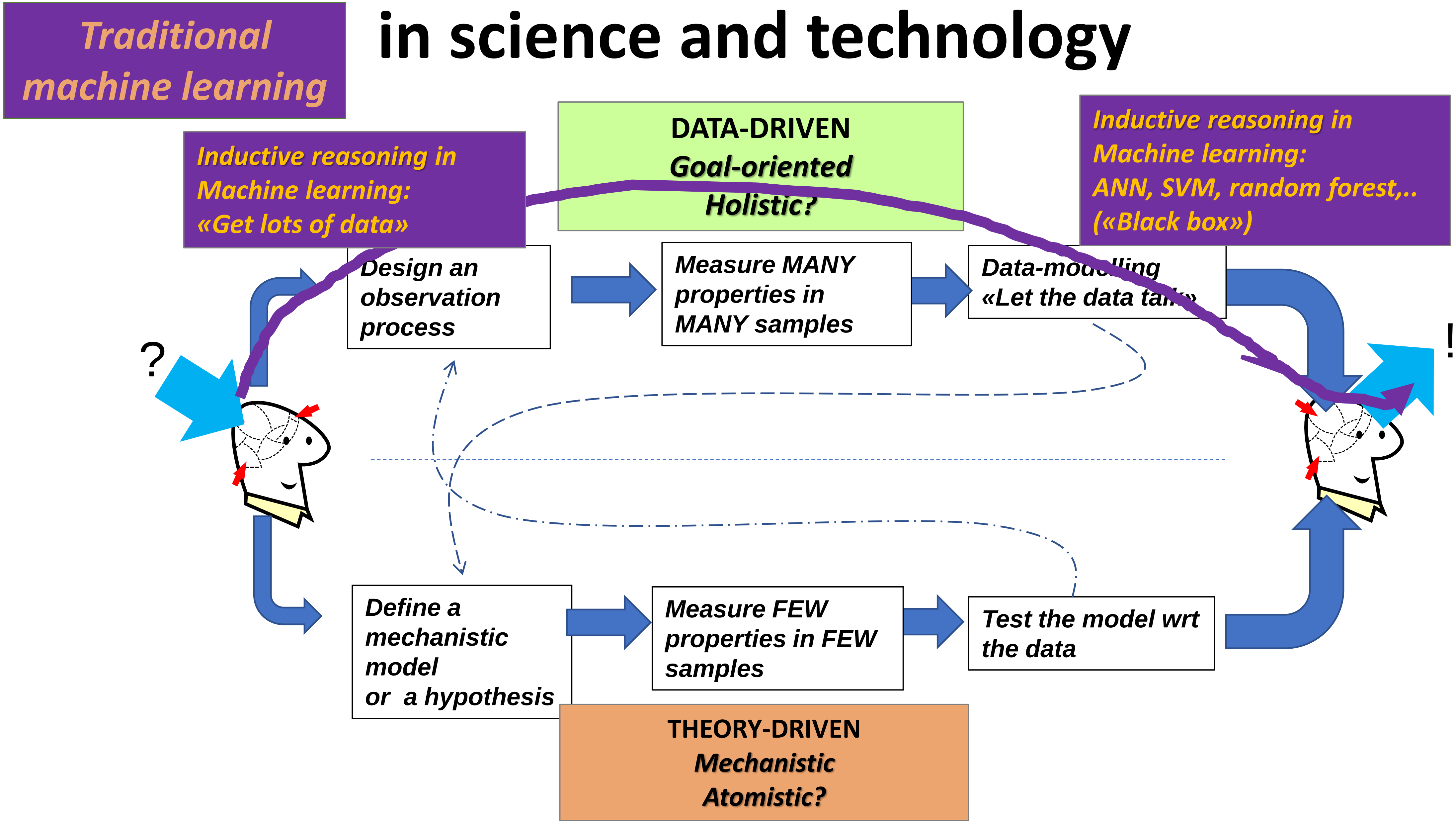
Two roads from question to answer in science and technology



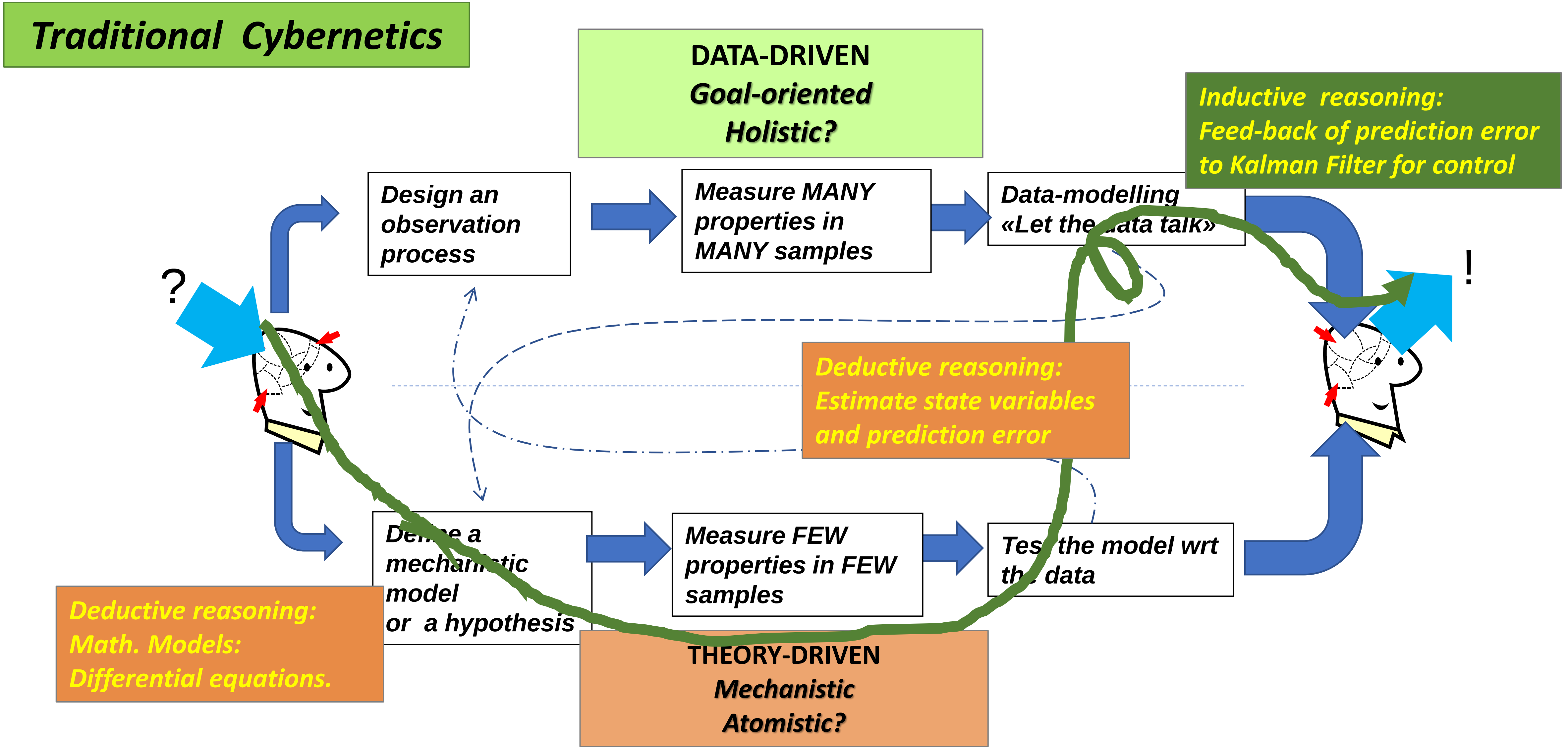
Two roads from question to answer in science and technology



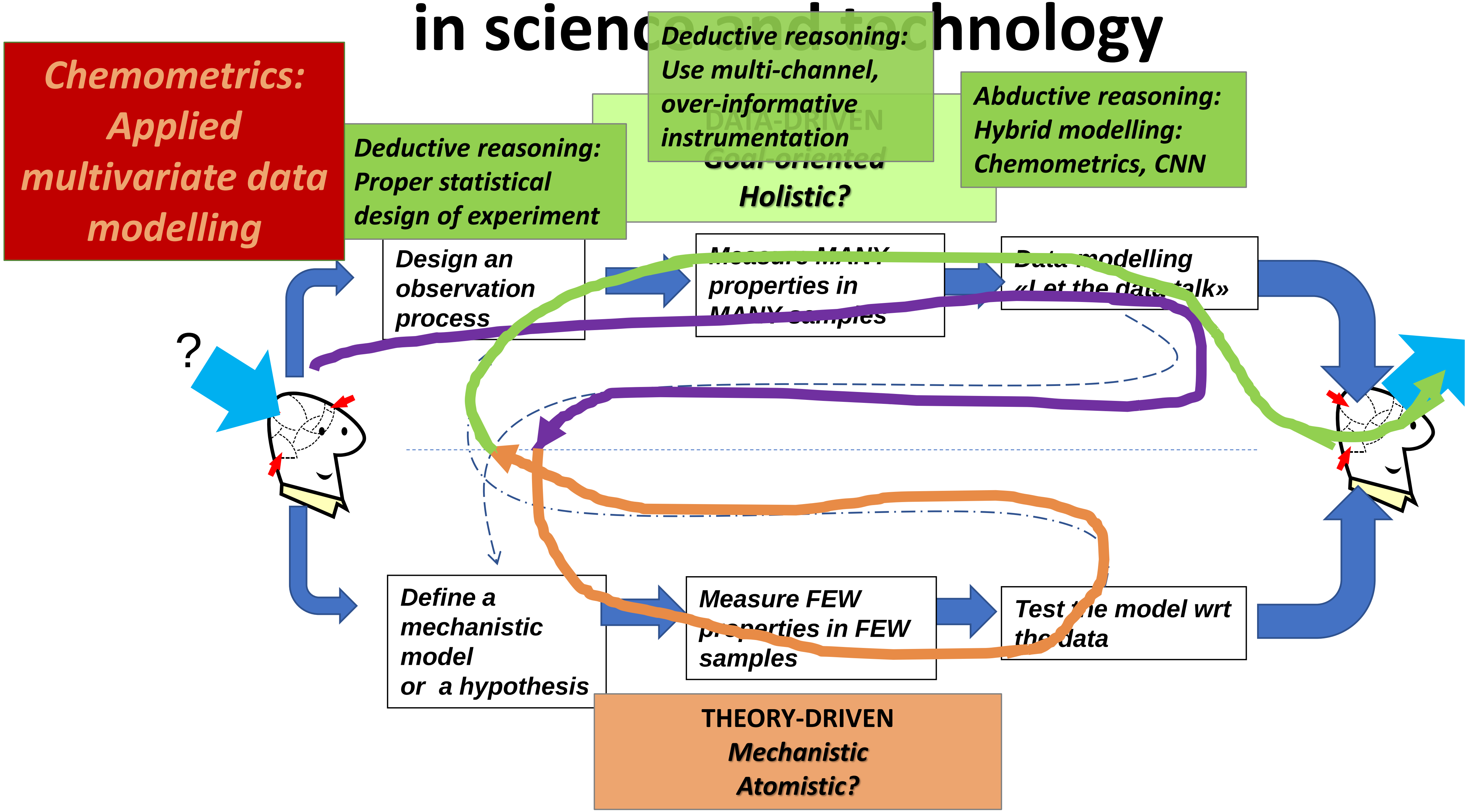
Two roads from question to answer in science and technology



Two roads from question to answer in science and technology

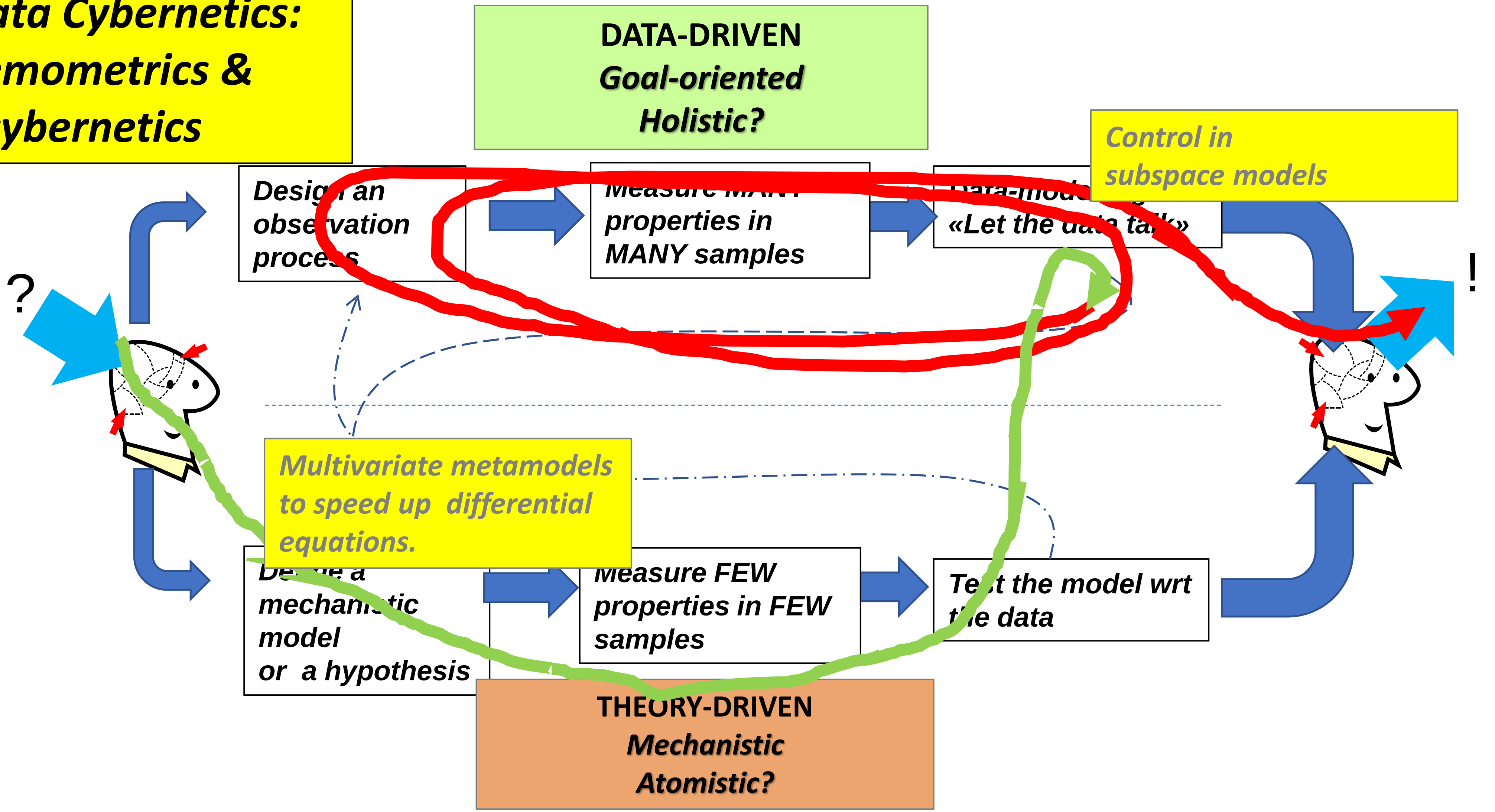


Two roads from question to answer in science and technology



Two roads from question to answer in science and technology

Big Data Cybernetics: Chemometrics & cybernetics



Big Data Cybernetics:

Combining advanced process control and chemometrics

Dept. Engineering Cybernetics, NTNU Trondheim 2018/2019:

5 +1 professors, lots of students:

«How to discover the real world»

equinor

KONGSBERG

CAMO
Bring data to life

SYKEHUSET ØSTFOLD

DNB

+ banks, metallurgy & other industry …

Big Data Cybernetics:
Combining advanced process control and chemometrics

equinor

KONGSBERG

CAMO
Bring data to life

SYKEHUSET ØSTFOLD

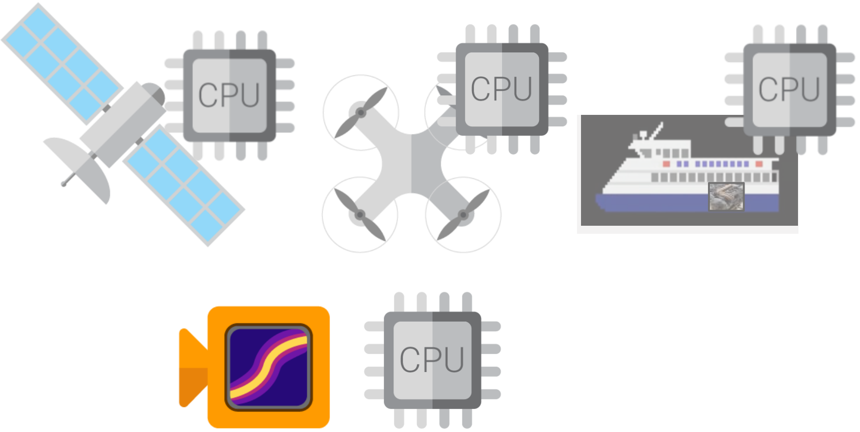
DNB

+ banks, metallurgy & other industry …

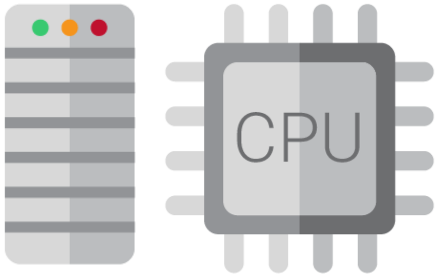
<u>Dept. Engineering Cybernetics, NTNU Trondheim</u>
<u>2018/2019:</u>
5 +1 professors, lots of students:
«How to discover the real world»

idletechs

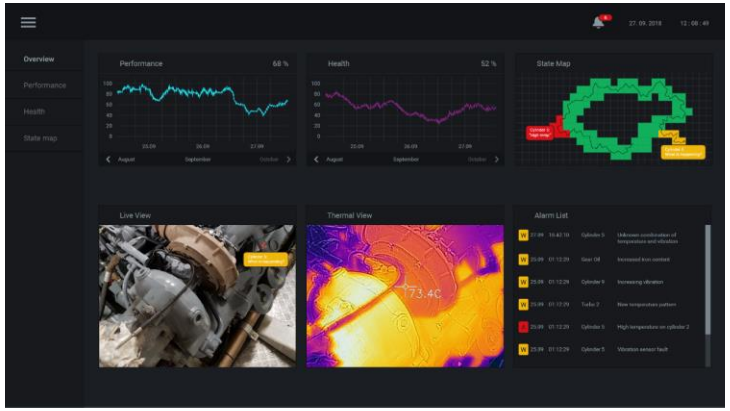
Mektig matte uten tårer



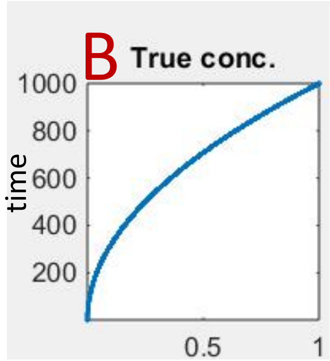
idletechs Edge



idletechs Analytics

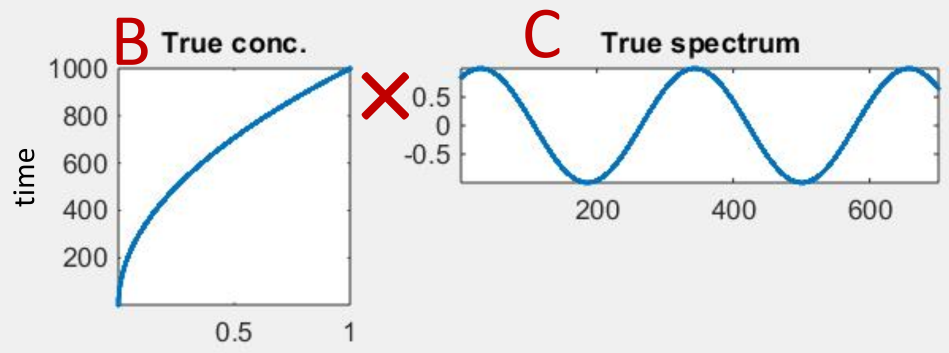


idletechs Application



A causal phenomenon's time-dependent development

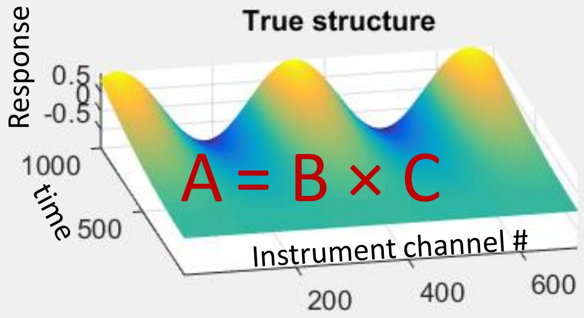
How Quantitative Big Data are often generated



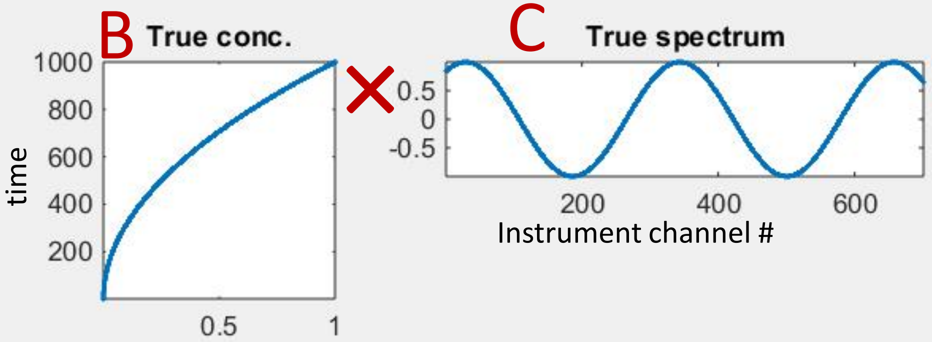
A causal phenomenon's time-dependent development

Its multi-channel property profile

How Quantitative Big Data are often generated



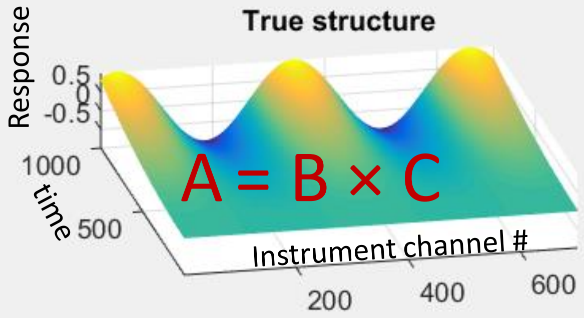
True properties

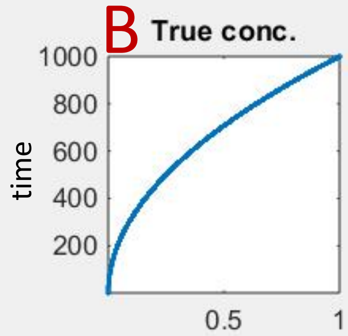


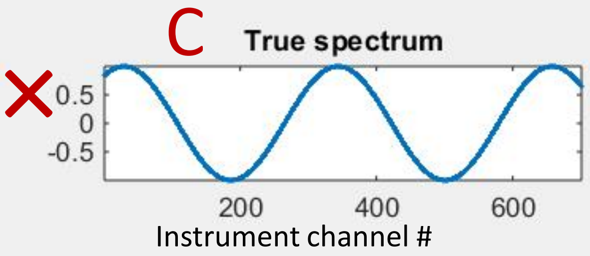
A causal phenomenon's time-dependent development

Its multi-channel property profile

How Quantitative Big Data are often generated







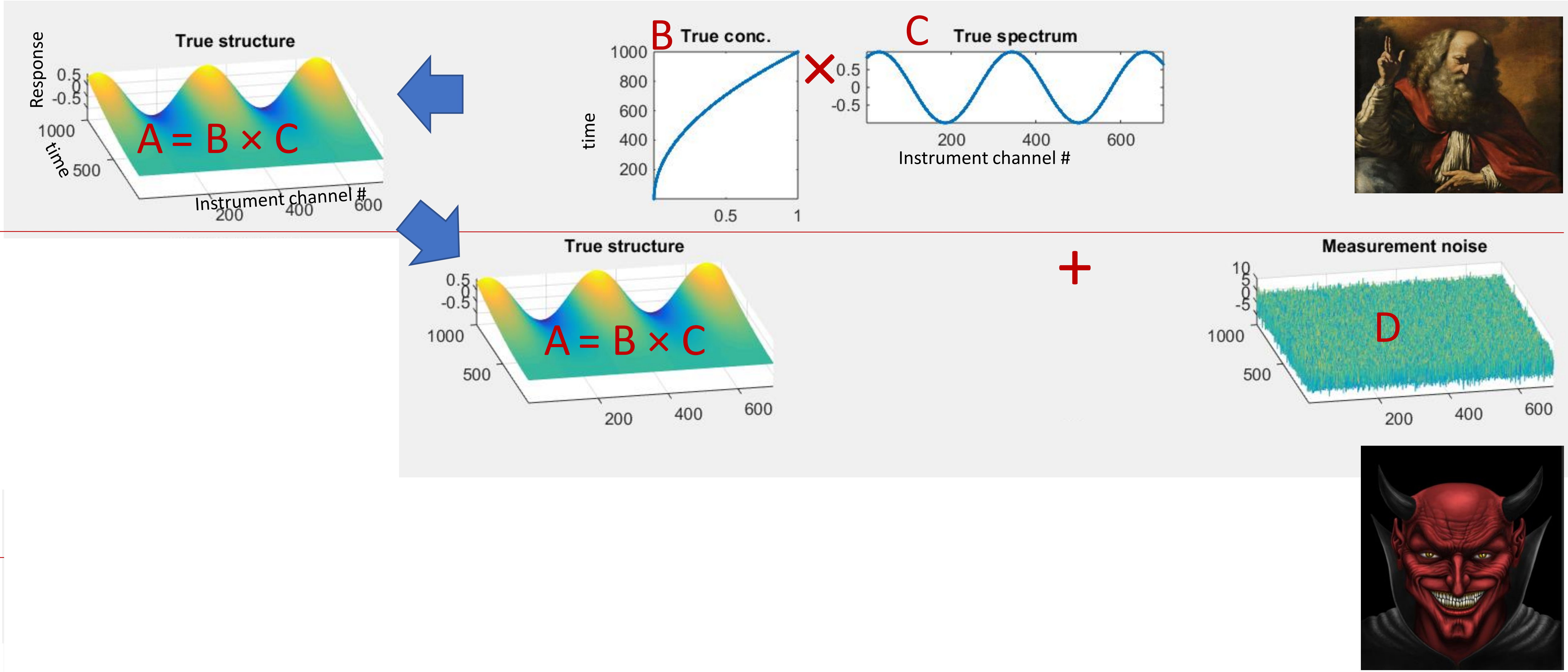
True properties

A causal phenomenon’s time-dependent development

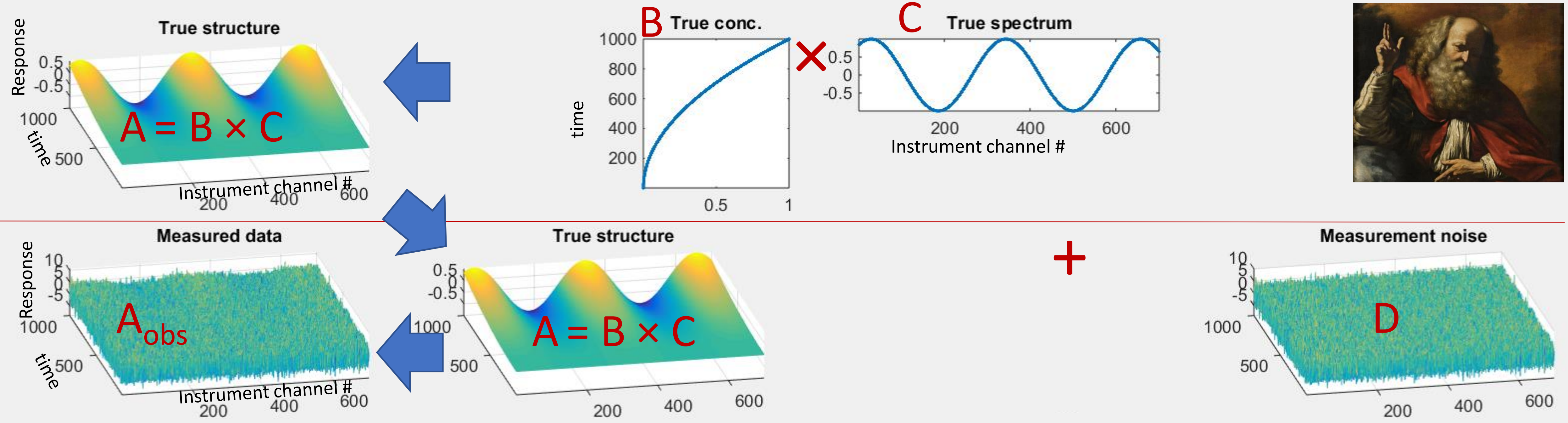
Its multi-channel property profile

Vector algebra
First: Caspar Wessel from Vestby, 1797

How Quantitative Big Data are often generated



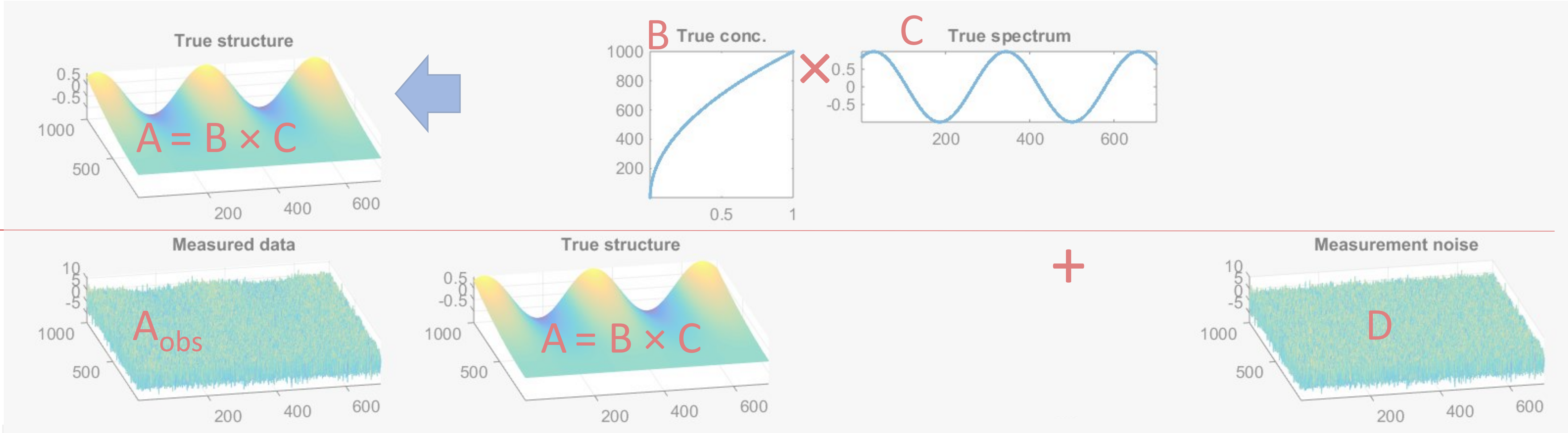
How Quantitative Big Data are often generated



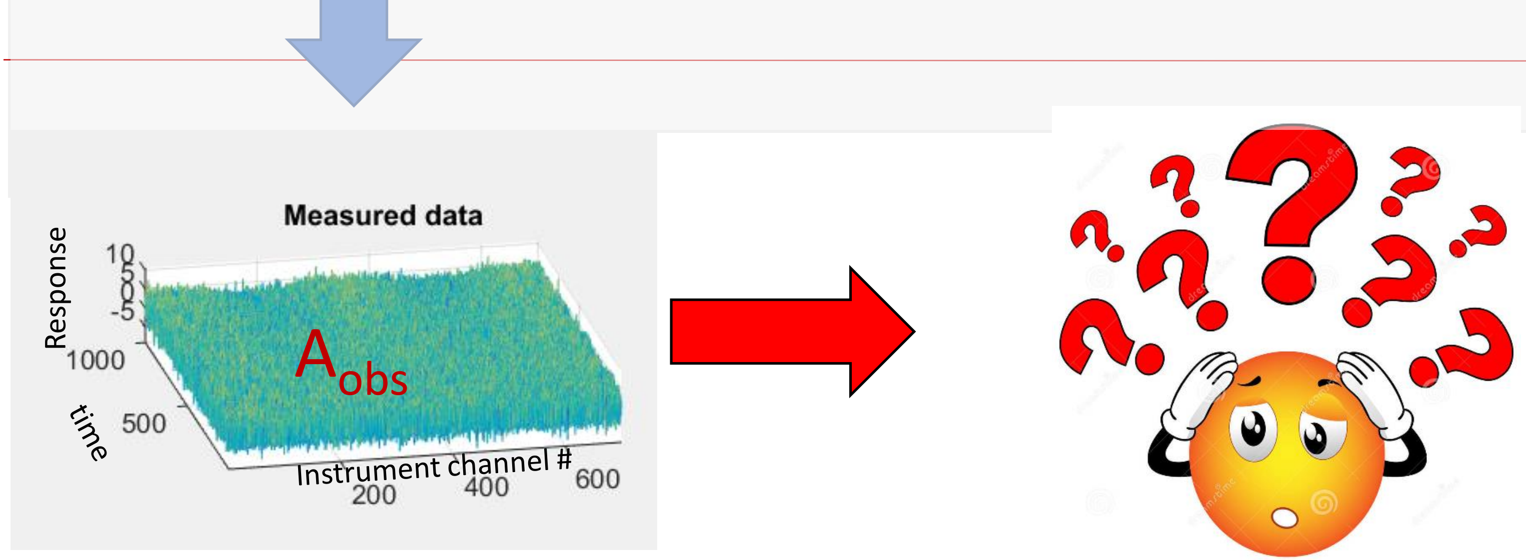
CAUSAL MATH MODEL:

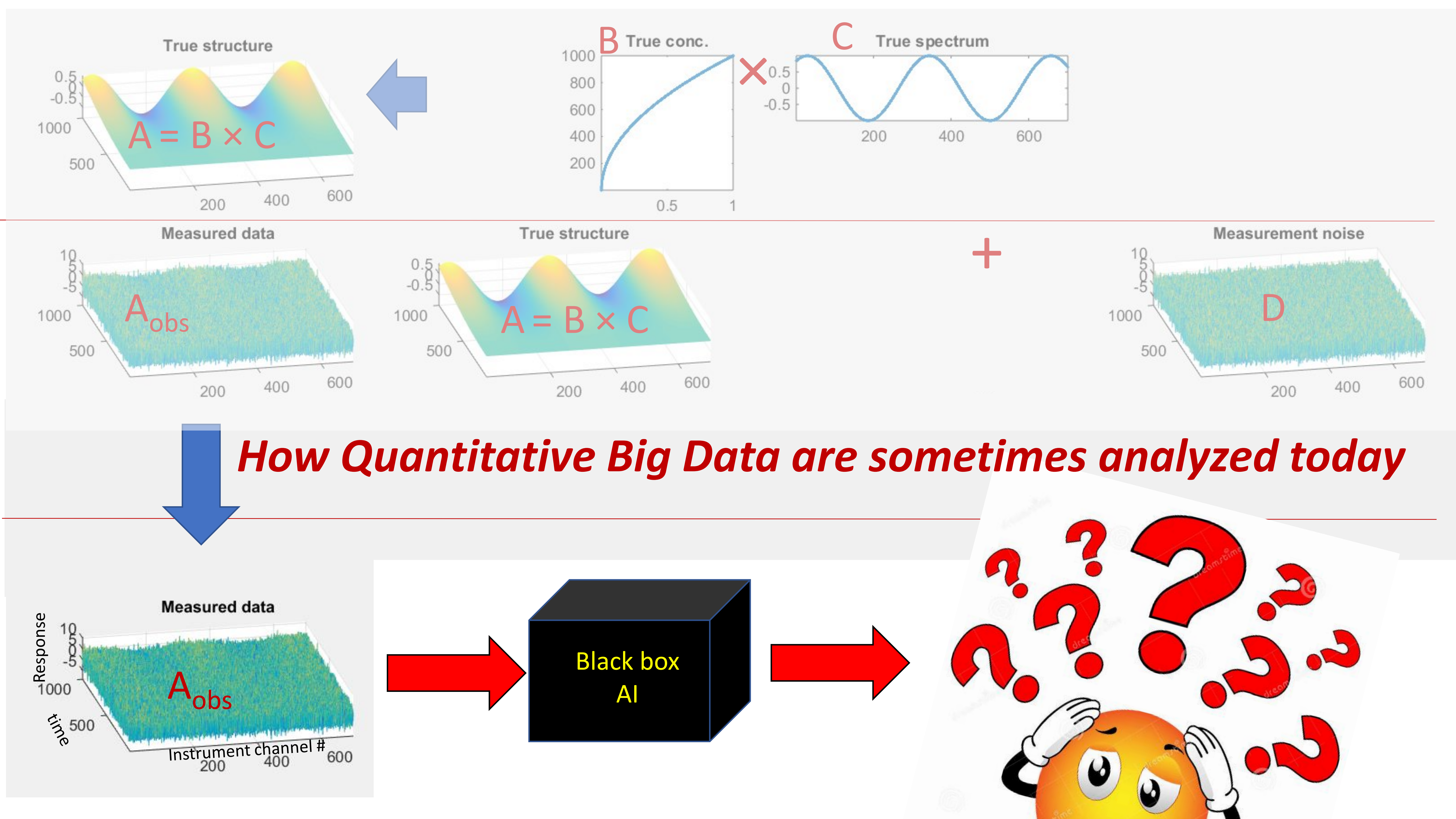
$A_{obs} = B \times C + D$

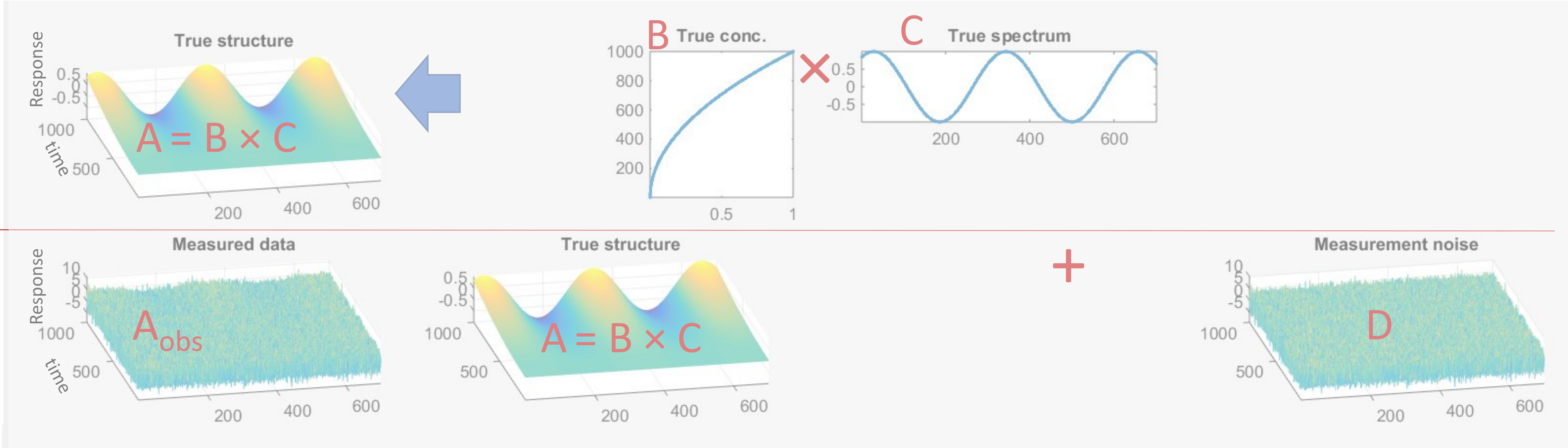
How Quantitative Big Data are often generated



How Quantitative Big Data are NOT analyzed

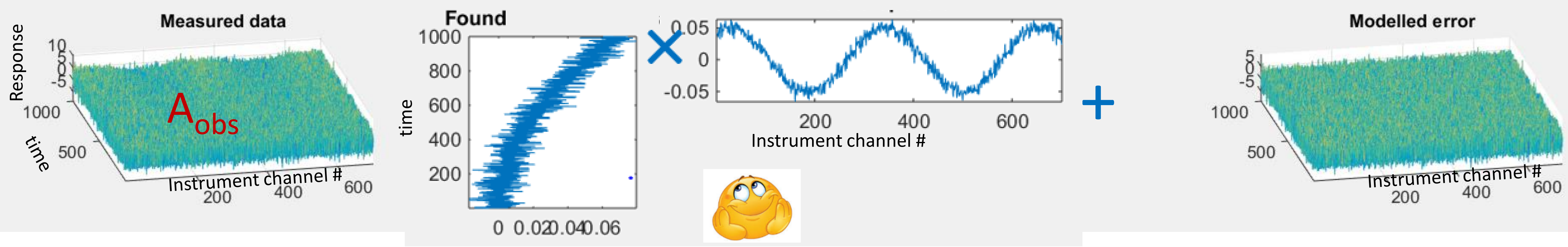



True structure
A = B × C
B True conc.
C True spectrum
Measured data
Aobs
True structure
A = B × C
Measurement noise
D
+
How Quantitative Big Data are sometimes analyzed today
Measured data
Response
time
Instrument channel #
Aobs
Black box
AI



How Quantitative Big Data may be analyzed

Multivariate analysis:

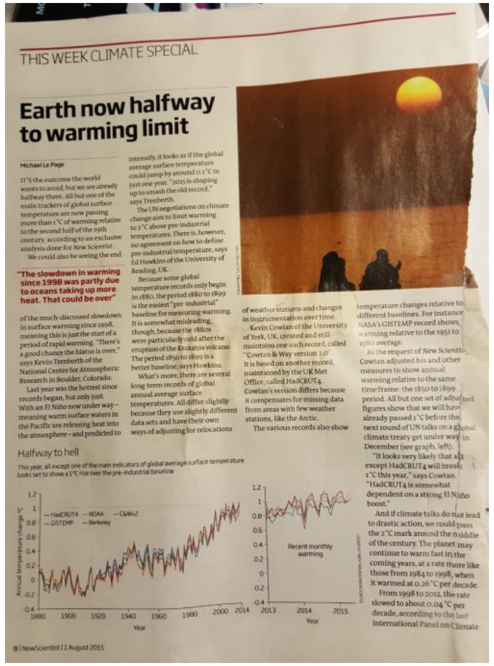


Naturens rytmer og harmonier (eller mangel på sådan)

Fem ulike forskningsgruppers estimat av jordklodens gjennomsnitts-temperatur

1880-2014

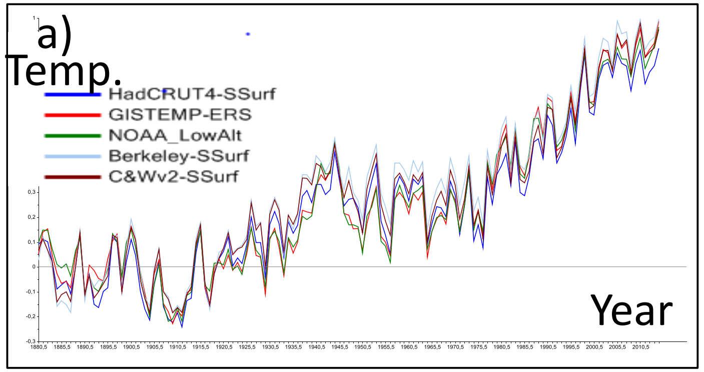
(New Scientist August 2015)

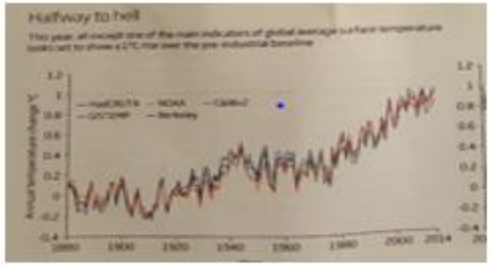

THIS WEEK CLIMATE SPECIAL

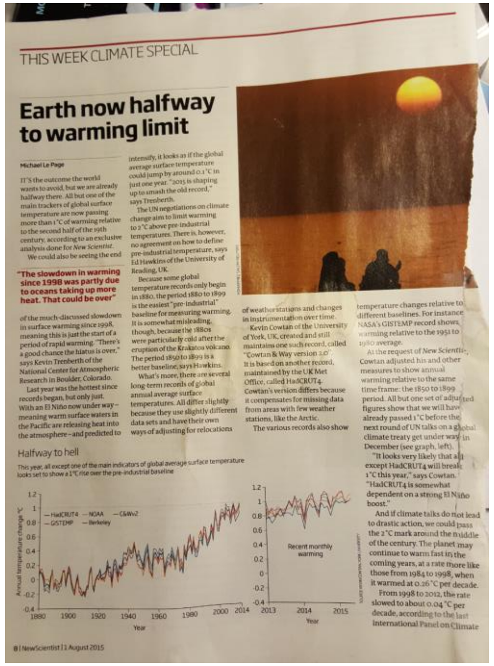
Earth now halfway to warming limit

Halfway to hell

The earth's average temperature from 1880 till 2014, as estimated by five different laboratories (from New Scientist 2015)







THIS WEEK CLIMATE SPECIAL

Earth now halfway to warming limit

Michael Le Page

IT'S the outcome the world wants to avoid, but we are already halfway there. All but one of the main trackers of global surface temperature are now passing more than 1°C of warming relative to the second half of the 19th century, according to an exclusive analysis done for *New Scientist*.

We could also be seeing the end of the much-discussed slowdown in surface warming since 1998, meaning this is just the start of a period of rapid warming. "There's a good chance the hiatus is over," says Kevin Trenberth of the National Center for Atmospheric Research in Boulder, Colorado.

"The slowdown in warming since 1998 was partly due to oceans taking up more heat. That could be over"

Last year was the hottest since records began, but only just. With an El Niño now under way – meaning warm surface waters in the Pacific are releasing heat into the atmosphere – and predicted to intensify, it looks as if the global average surface temperature could jump by around 0.1°C in just one year. "2015 is shaping up to smash the old record," says Trenberth.

The UN negotiations on climate change aim to limit warming to 2°C above pre-industrial temperatures. There is, however, no agreement on how to define pre-industrial temperature, says Ed Hawkins of the University of Reading, UK.

Because some global temperature records only begin in 1880, the period 1880 to 1899 is the easiest "pre-industrial" baseline for measuring warming. It is somewhat misleading, though, because the 1880s were particularly cold after the eruption of the Krakatoa volcano. The period 1850 to 1899 is a better baseline, says Hawkins.

What's more, there are several long-term records of global annual average surface temperatures. All differ slightly because they use slightly different data sets and have their own ways of adjusting for relocations of weather stations and changes in instrumentation over time.

Kevin Cowtan of the University of York, UK, created and still maintains one such record, called "Cowtan & Way version 2.0". It is based on another record, maintained by the UK Met Office, called HadCRUT4. Cowtan's version differs because it compensates for missing data from areas with few weather stations, like the Arctic.

The various records also show temperature changes relative to different baselines. For instance NASA's GISTEMP record shows warming relative to the 1951 to 1980 average.

At the request of *New Scientist*, Cowtan adjusted his and other measures to show annual warming relative to the same time frame: the 1850 to 1899 period. All but one set of adjusted figures show that we will have already passed 1°C before the next round of UN talks on a global climate treaty get under way in December (see graph, left).

"It looks very likely that all except HadCRUT4 will break 1°C this year," says Cowtan. "HadCRUT4 is somewhat dependent on a strong El Niño boost."

And if climate talks do not lead to drastic action, we could pass the 2°C mark around the middle of the century. The planet may continue to warm fast in the coming years, at a rate more like those from 1984 to 1998, when it warmed at 0.26°C per decade.

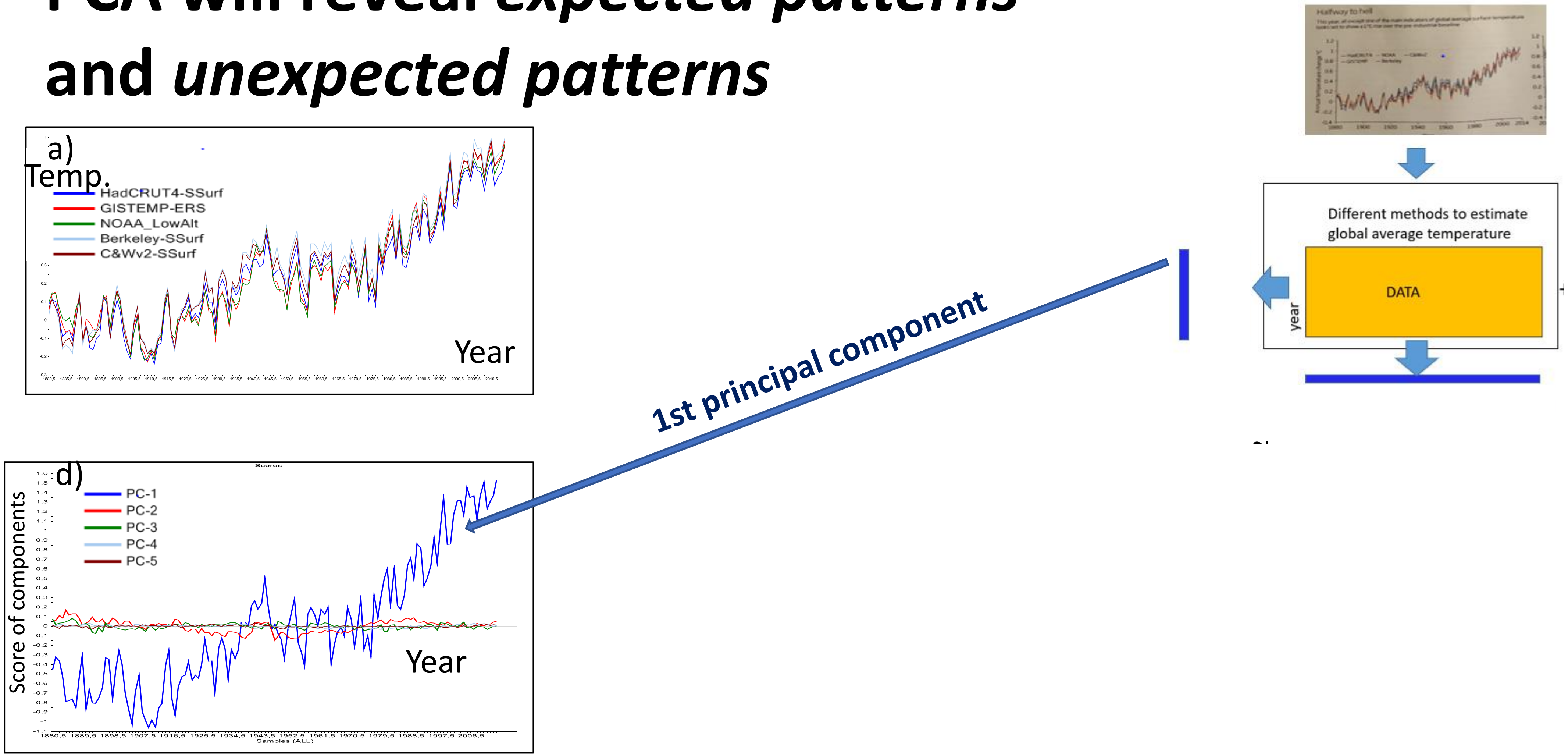
From 1998 to 2012, the rate slowed to about 0.04°C per decade, according to the last International Panel on Climate

Halfway to hell

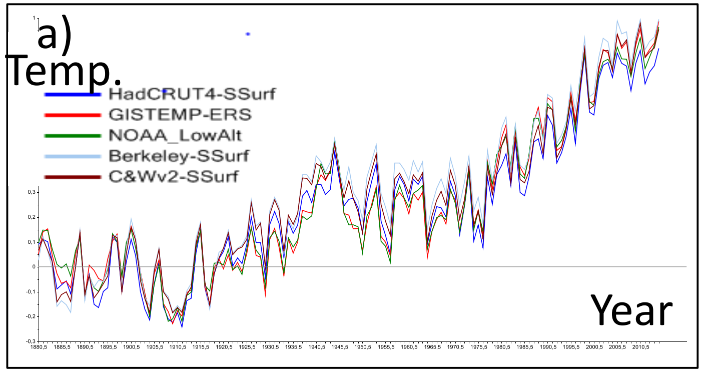
This year, all except one of the main indicators of global average surface temperature looks set to show a 1°C rise over the pre-industrial baseline

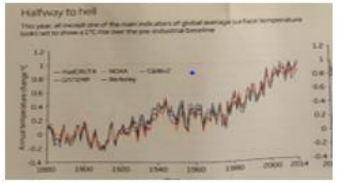
(New Scientist August 2015

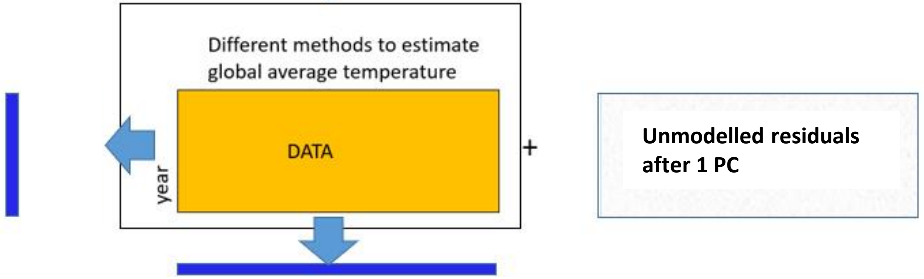
Data driven multivariate modelling by PCA will reveal *expected patterns* and *unexpected patterns*



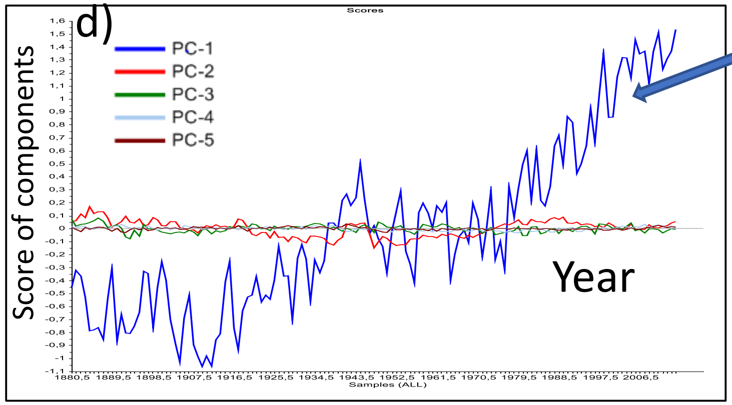
Data driven multivariate modelling by PCA will reveal *expected patterns* and *unexpected patterns*



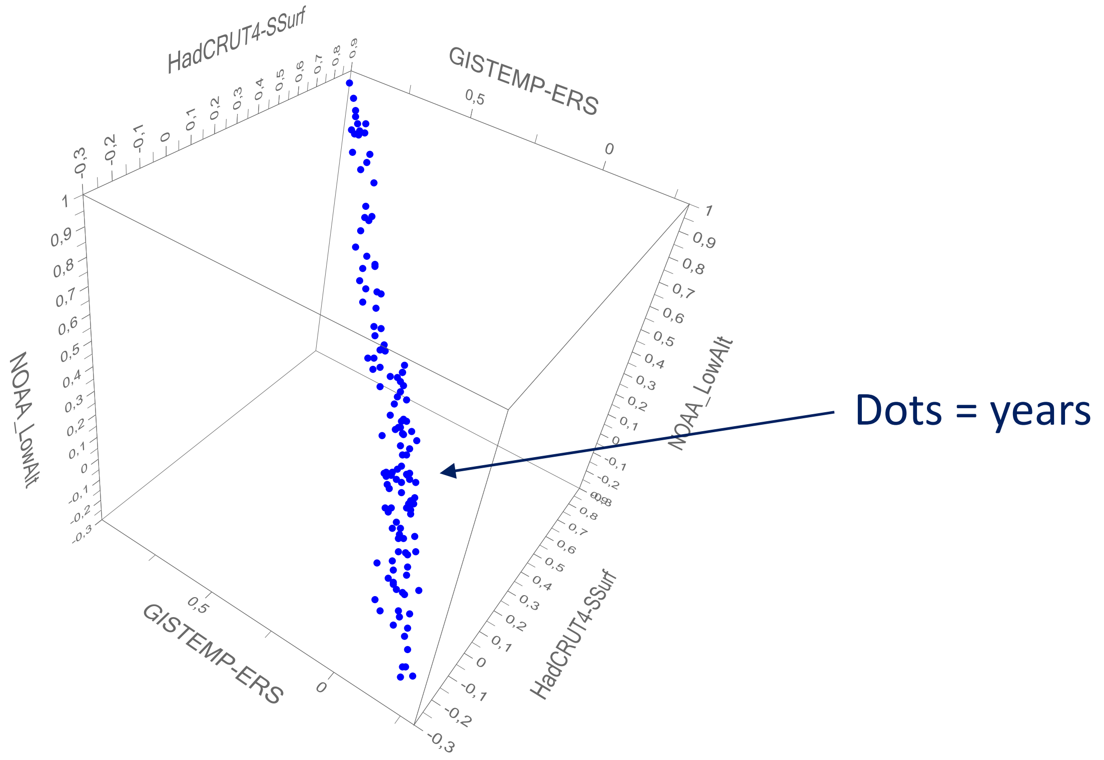




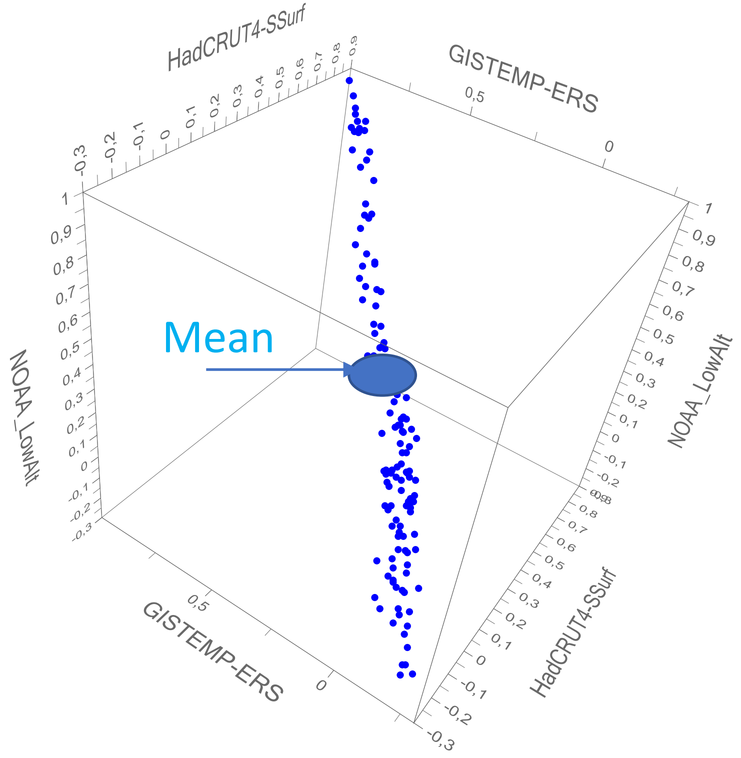
1st principal component



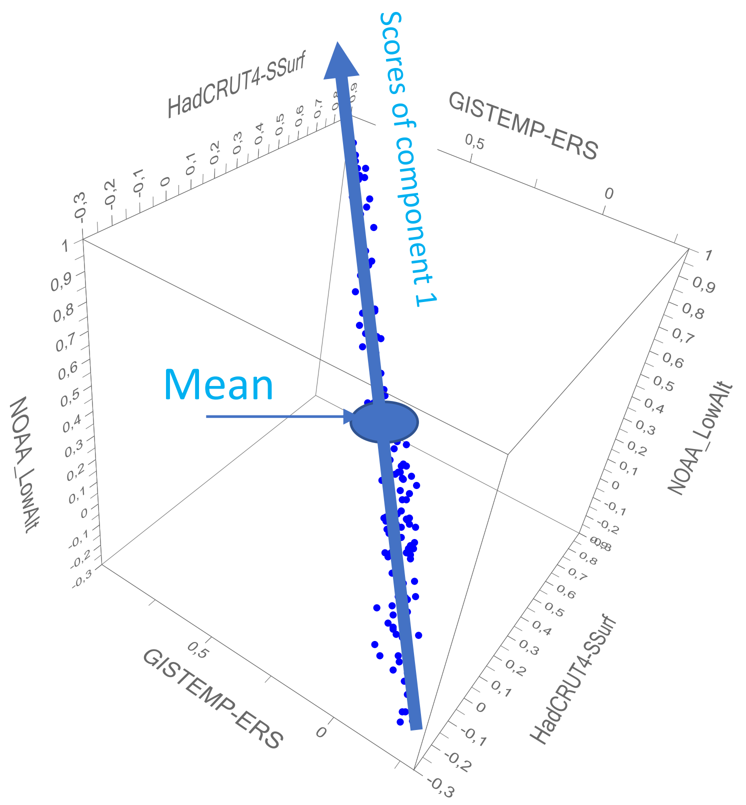
Three of the 5 (or e.g. 50 000?) available variables



Mean centering:

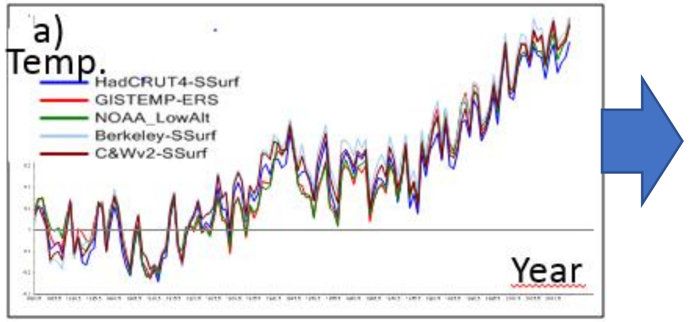


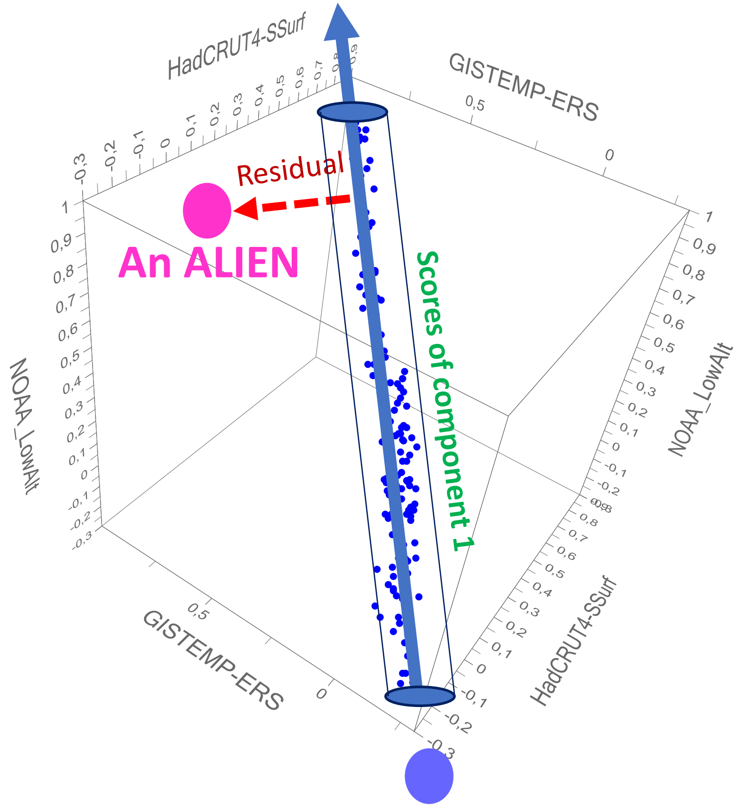
First principal component:



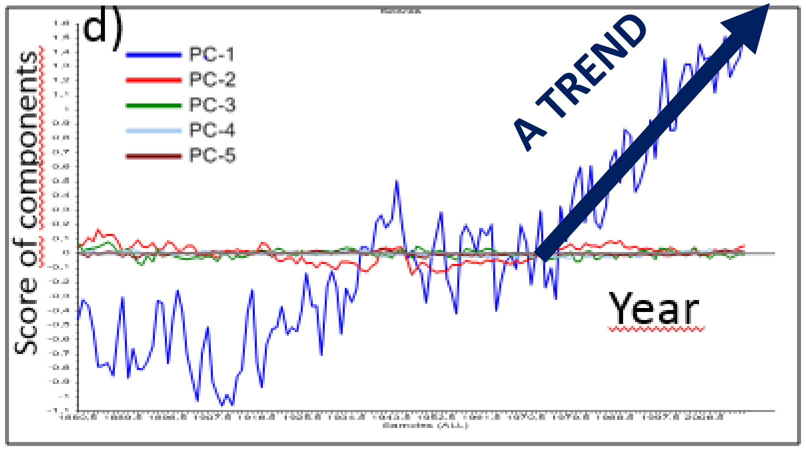
Multivariate soft data-modelling gives **fewer, but more sensitive alarms,** for **three different types of abnormalities:**

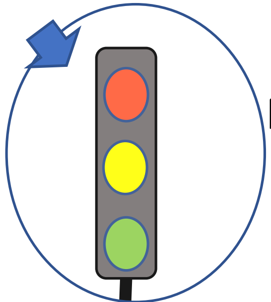
Raw data
Many variables (temperatures)





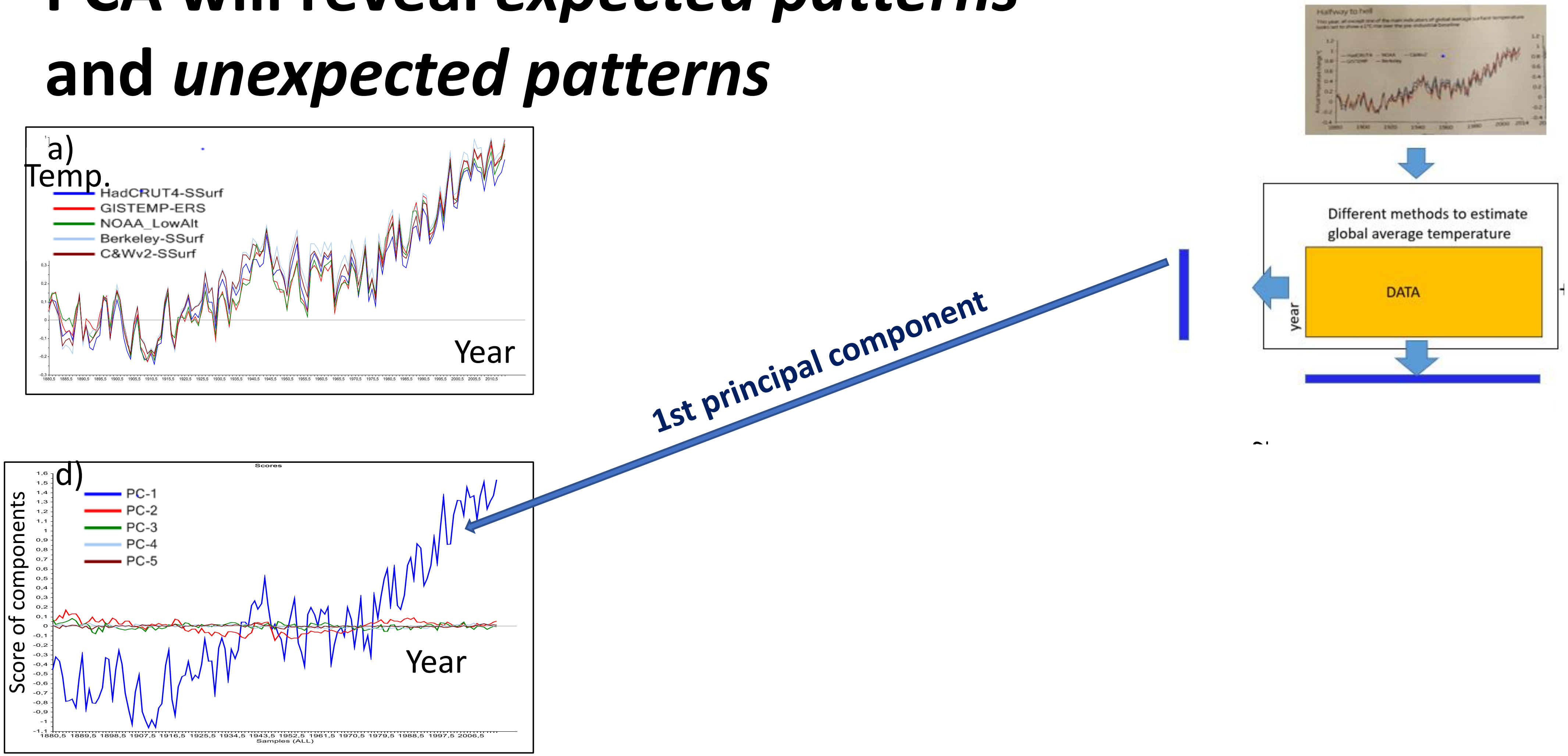
Few informative combinations of the many variables



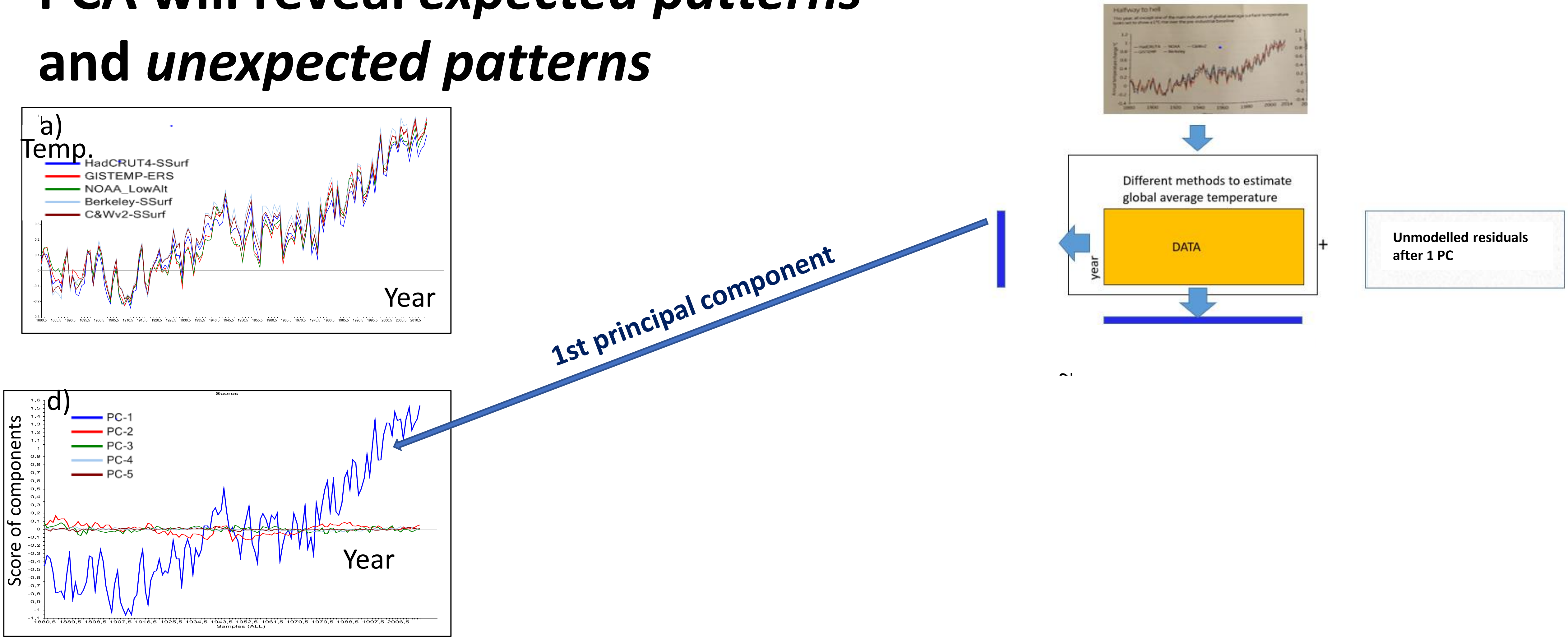


Færre, bedre alarmer!

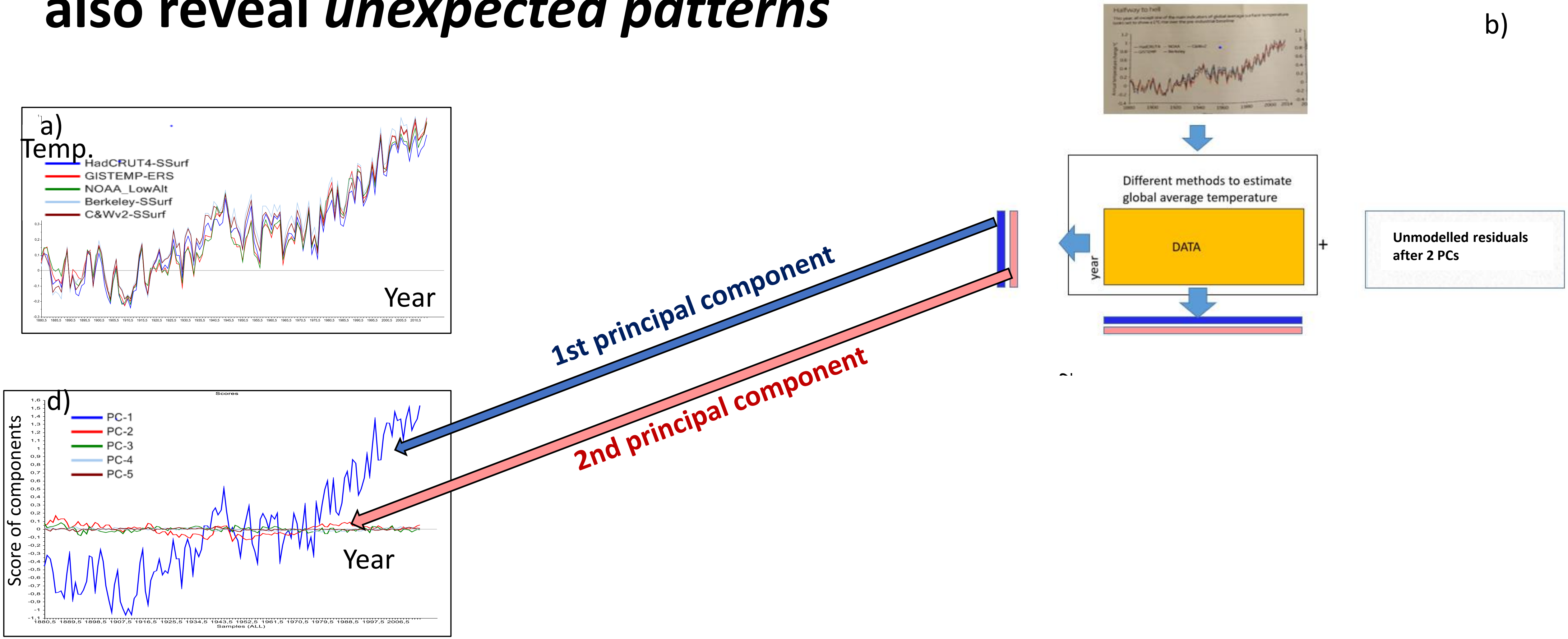
Data driven multivariate modelling by PCA will reveal *expected patterns* and *unexpected patterns*



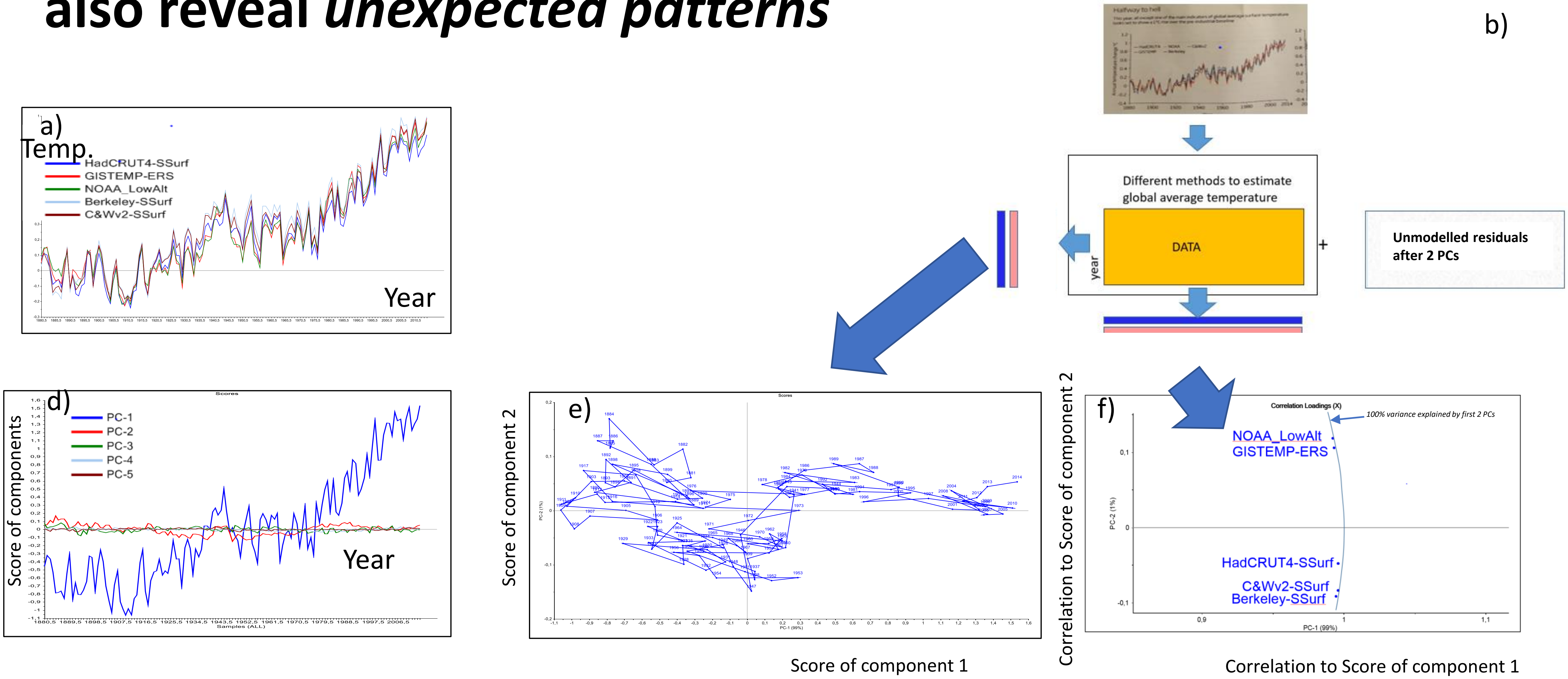
Data driven multivariate modelling by PCA will reveal *expected patterns* and *unexpected patterns*



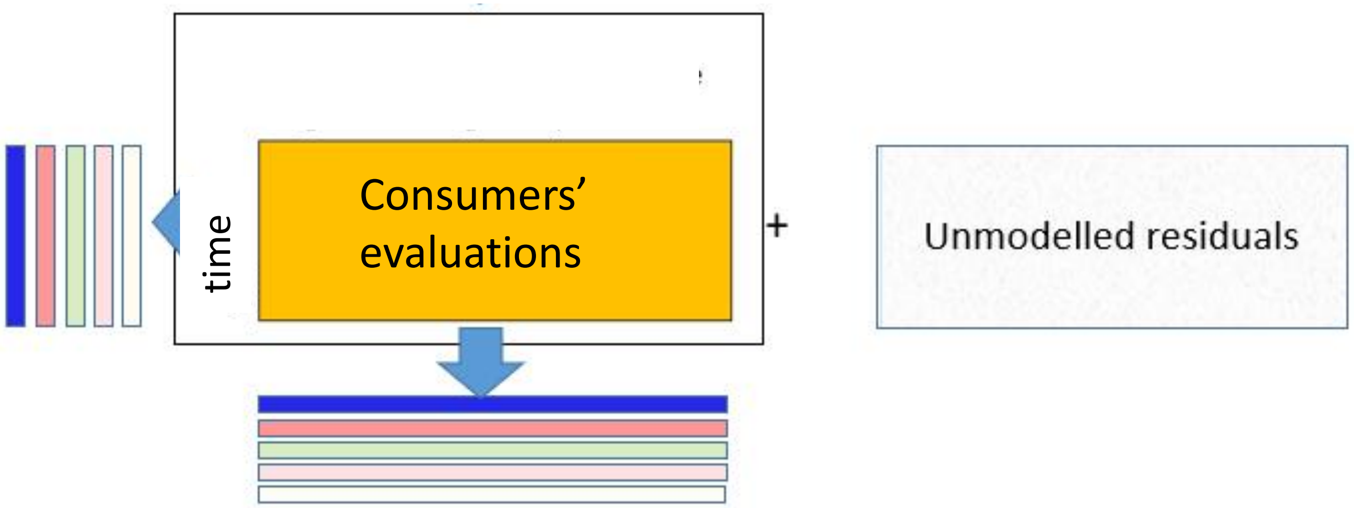
Data driven multivariate modelling will also reveal *unexpected patterns*



Data driven multivariate modelling will also reveal *unexpected patterns*



Find patterns in
a data set, e.g. consumer' purchase of products
by PCA



The bilinear data-model as
numbers, vectors and matrices:

Simple but mighty math

Principal Component Analysis illustrated

Simple but mighty math:

- 21 $= 3 \times 7$ $a = b \times c$

Simple but mighty math:

- $21 \quad = 3 \times 7$ $\qquad a = b \times c$
- $21.1 = 3 \times 7 + 0.1$ $\qquad a = b \times c \quad + d$

Simple but mighty math:

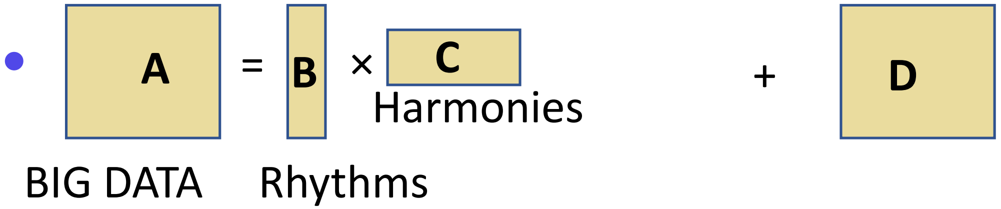
- $21 = 3 \times 7$ $\quad a = b \times c$
- $21.1 = 3 \times 7 + 0.1$ $\quad a = b \times c + d$
- $41.1 = 3 \times 7 + 2 \times 10 + 0.1$ $\quad a = b_1 \times c_1 + b_2 \times c_2 + d = \mathbf{b} \times \mathbf{c} + d$

Vector-algebra, published
av Caspar Wessel 1797

Simple but mighty math:

- $21 \quad = 3 \times 7$ $\qquad a = b \times c$
- $21.1 = 3 \times 7 + 0.1$ $\qquad a = b \times c \quad + d$
- $41.1 = 3 \times 7 + 2 \times 10 + 0.1$ $\qquad a = b_1 \times c_1 + b_2 \times c_2 + d = \mathbf{b} \times \mathbf{c} + d$

Vector-algebra, published av Caspar Wessel 1797

- 

A= B × C + D

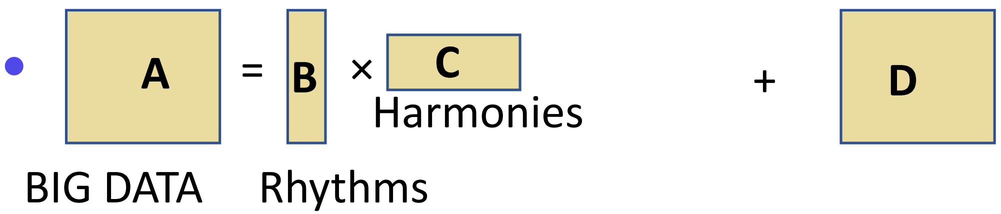
Matrix-algebra, published 1835

From data-table **A**, discover the unknown causal structure **B × C** and noise **D**

Simple but mighty math:

- $21 \quad = 3 \times 7$ $\qquad$ $a = b \times c$
- $21.1 = 3 \times 7 + 0.1$ $\qquad$ $a = b \times c + d$
- $41.1 = 3 \times 7 + 2 \times 10 + 0.1$ $\qquad$ $a = b_1 \times c_1 + b_2 \times c_2 + d = \mathbf{b} \times \mathbf{c} + d$

Vector-algebra, publisherd av Caspar Wessel 1797

- 

A= B × C + D

Matrix-algebra, published 1835

From data-table **A**, discover the unknown causal structure **B × C** and noise **D**

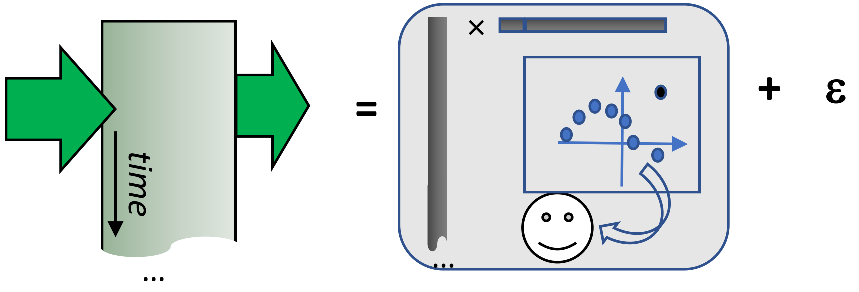
Principal component-analyse (PCA or SVD):
All multivariate methods' mother!

Automatic modelling of “ever-lasting” data streams

From raw streams of data, systematic patterns and relationships are automatically discovered and modelled. The data is stored in a highly compressed format:

Complex real-world system

- **Sensor data:**
  - **Hyperespectral video**
  - **Thermal video**
  - **Recordings of light, sound, vibrations, images, temperatures, moisture, etc.**
- **SCADA data**
- **Computer simulation outputs**
- **...**

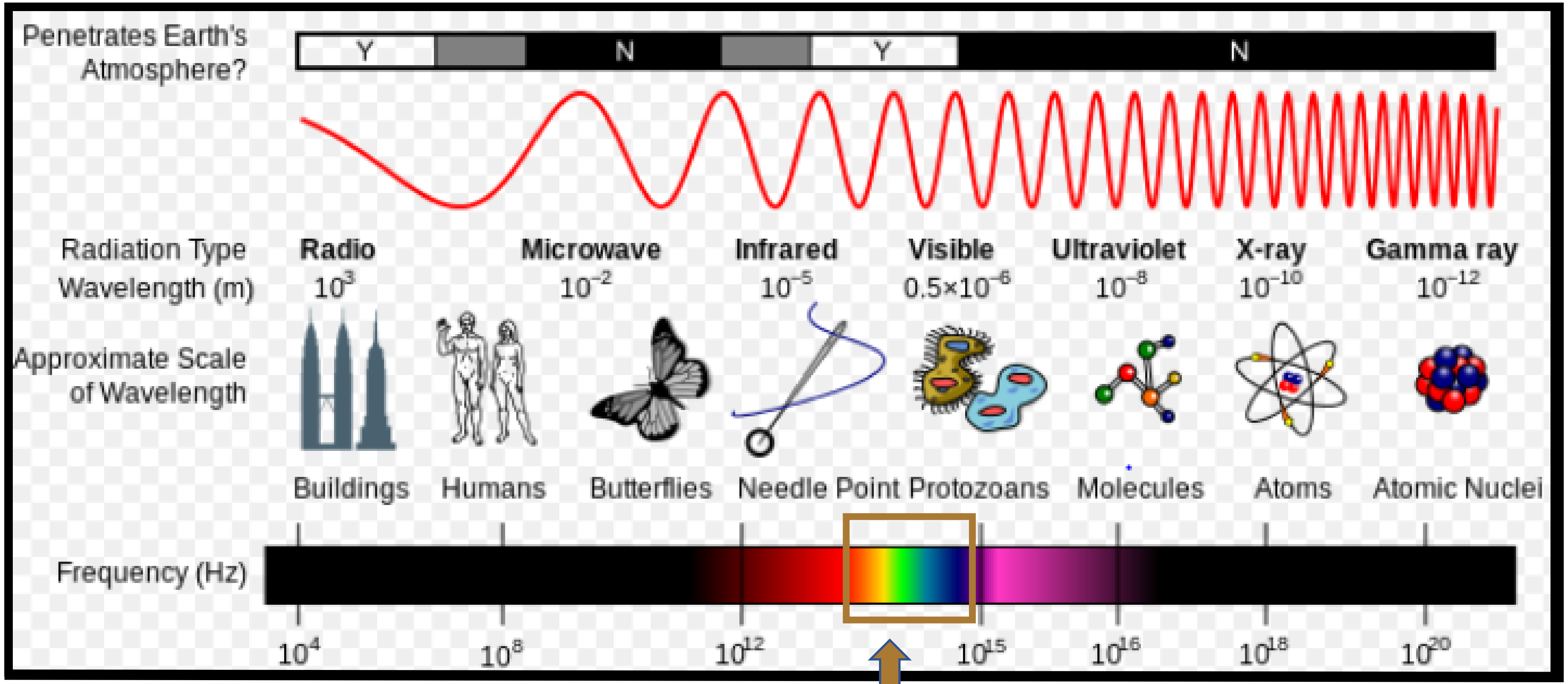


Raffaele Vitale, Anna Zhyrova, João F. Fortuna, Onno E. de Noord, Alberto Ferrer, Harald Martens: On-The-Fly Processing of continuous high-dimensional data streams Chemometrics and Intelligent Laboratory Systems 161 (2017) 118–129

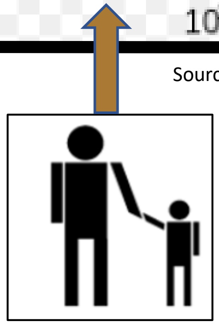
Idletechs AS provides new software tools for automatic modelling & compression and for human interpretation, of multidimensional, dynamic data.

idletechs

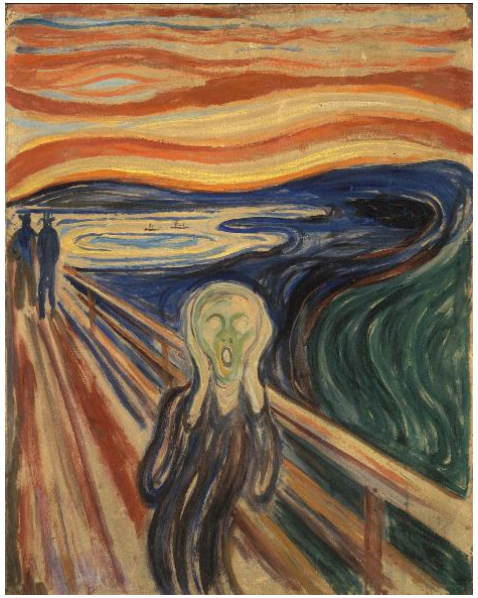
Vårt fargesyn er bra, men begrenset



Source: http://hubblesite.org/reference_desk/faq/answer.php.id=70&cat=light



Looking at art



Photographed in 3 colours (R,G,B)

Looking at nature

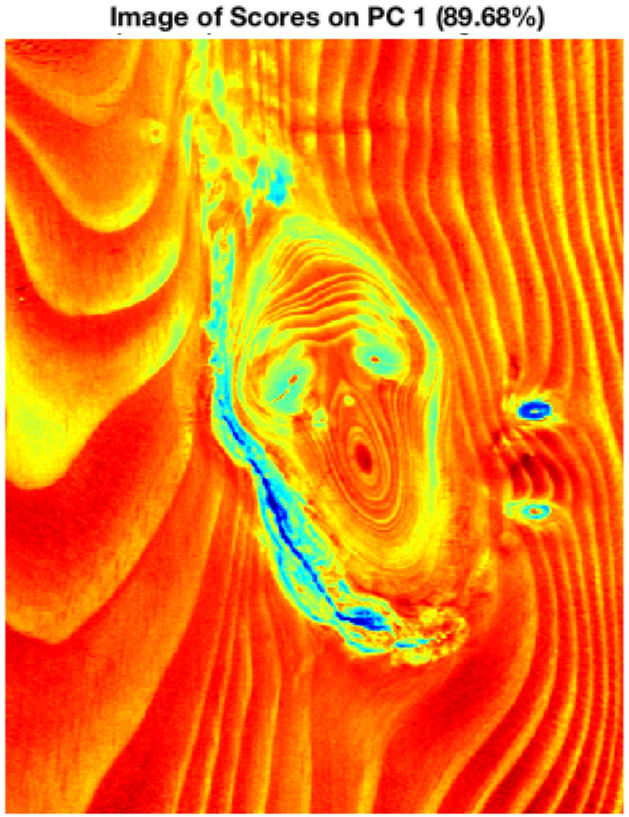
Else Tronstad, Leksvika

Photographed in 3 colours (R,G,B)

Looking at nature

Else Tronstad, Leksvika

Photographed in 3 colours (R,G,B)

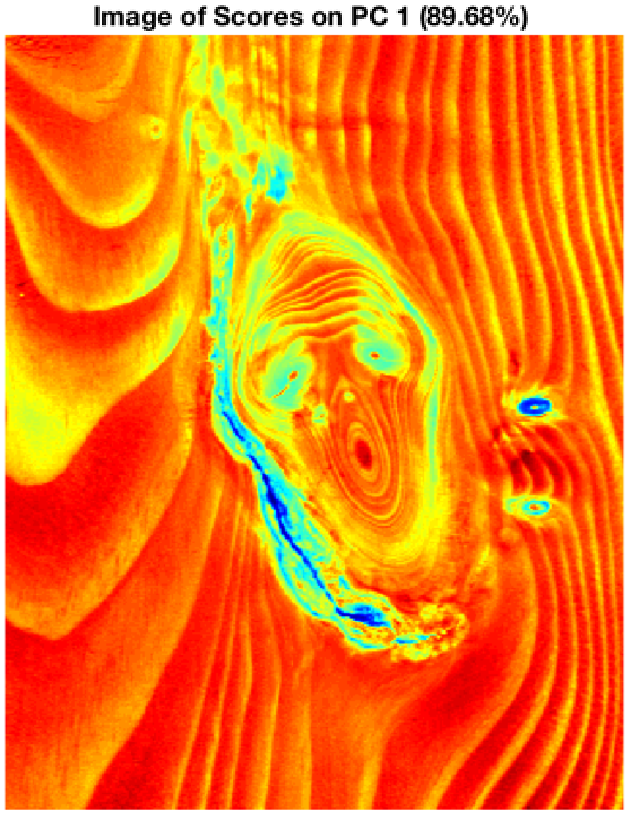


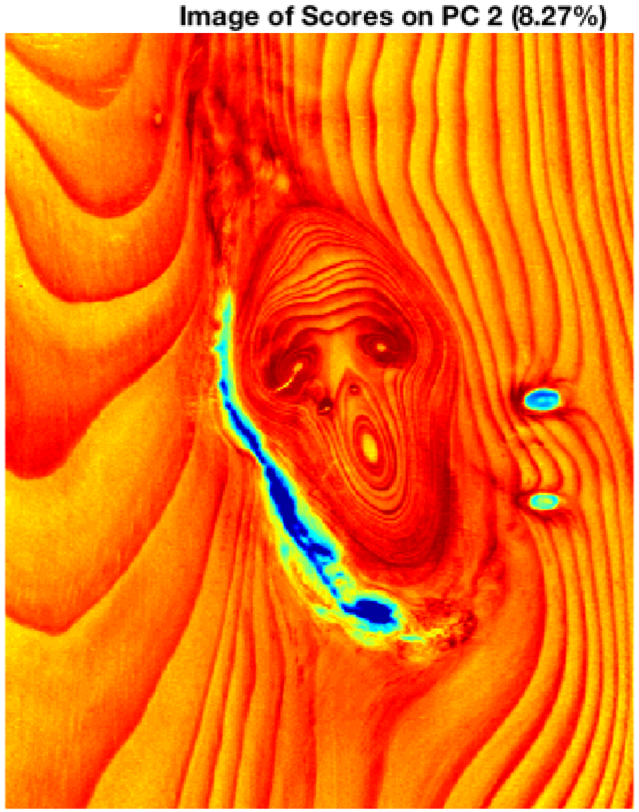
Photographed in several hundred «colours» (wavelength channels in vis. & NIR)
Courtesy of NMBU (Ingunn Burud)

Looking at nature

Else Tronstad, Leksvika

Photographed in 3 colours (R,G,B)

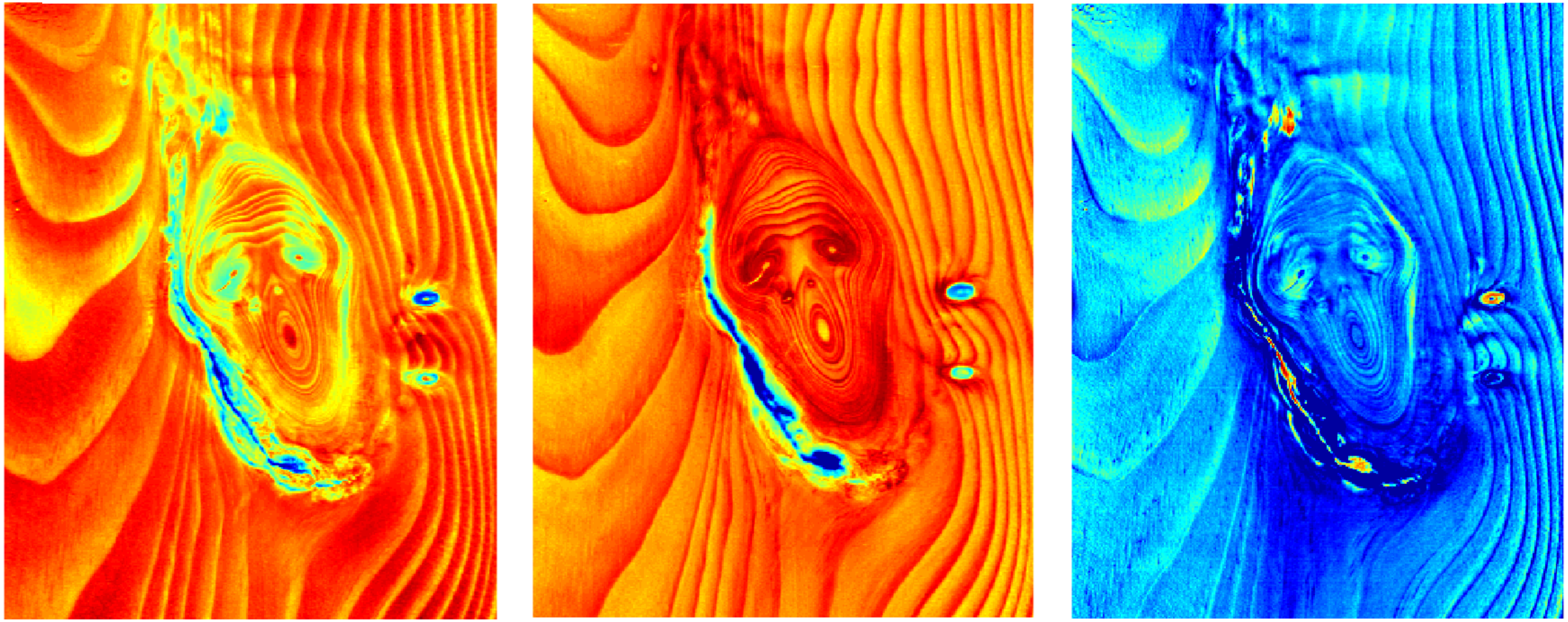




Photographed in several hundred «colours» (wavelength channels in vis. & NIR)
Courtesy of NMBU (Ingunn Burud)

Looking at nature

Image of Scores on PC 1 (89.68%) | Image of Scores on PC 2 (8.27%) | Image of Scores on PC 3 (0.94%)



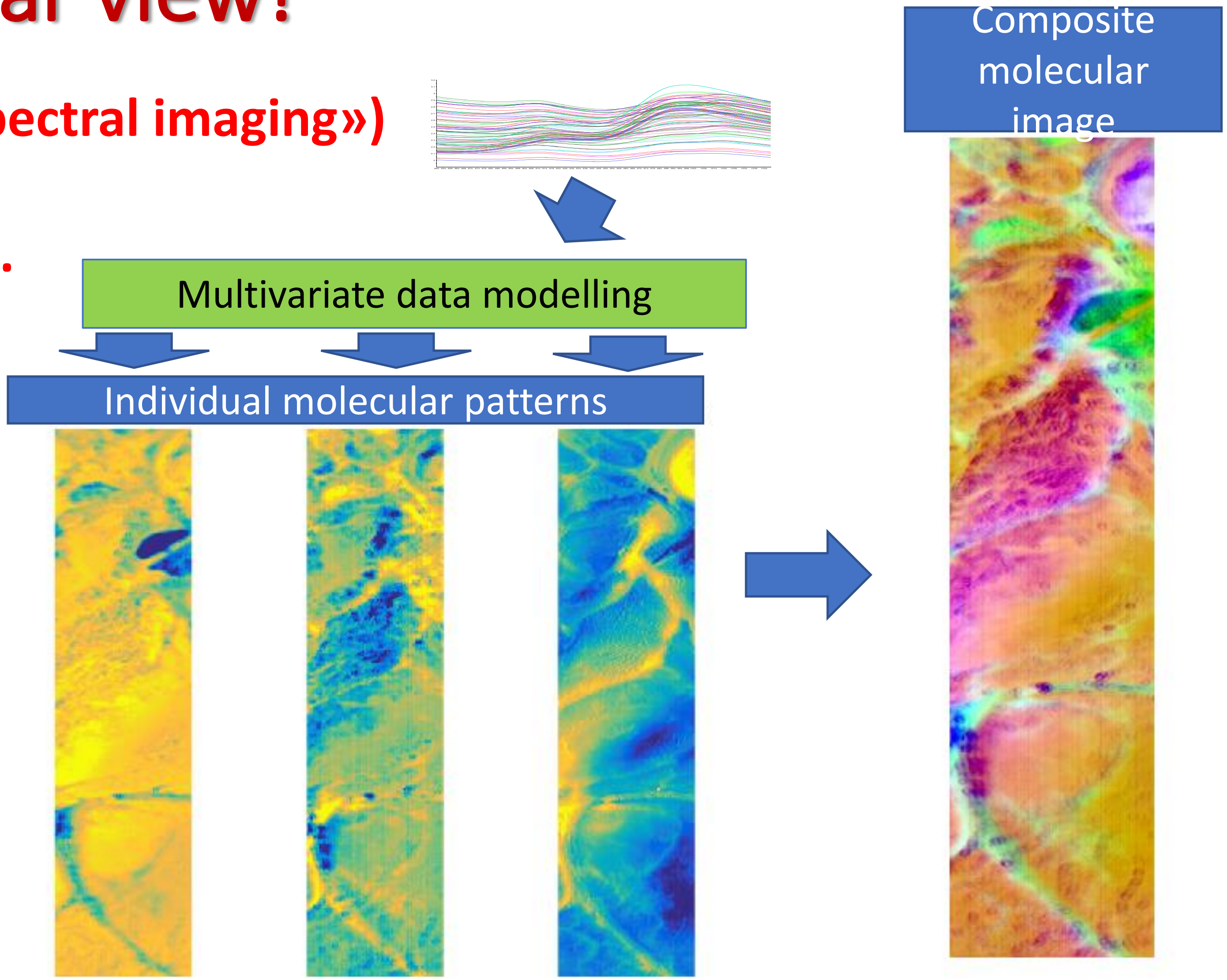
Else Tronstad, Leksvika

Photographed in 3 colours (R,G,B)

Photographed in several hundred «colours» (wavelength channels in vis. & NIR)
Courtesy of NMBU (Ingunn Burud)

Giving surgeons molecular view!

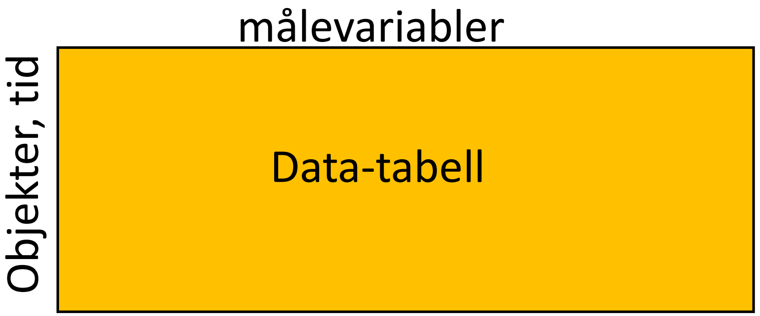
Multi-channel spectroscopy («Hyperspectral imaging») of muscle tissue in the Near-infrared wavelength range.



Hyperspectral NIR measurements of post-rigor porcine *l.dorsi* :
(Ingunn Burud, NMBU/Joao Fortuna NTNU)

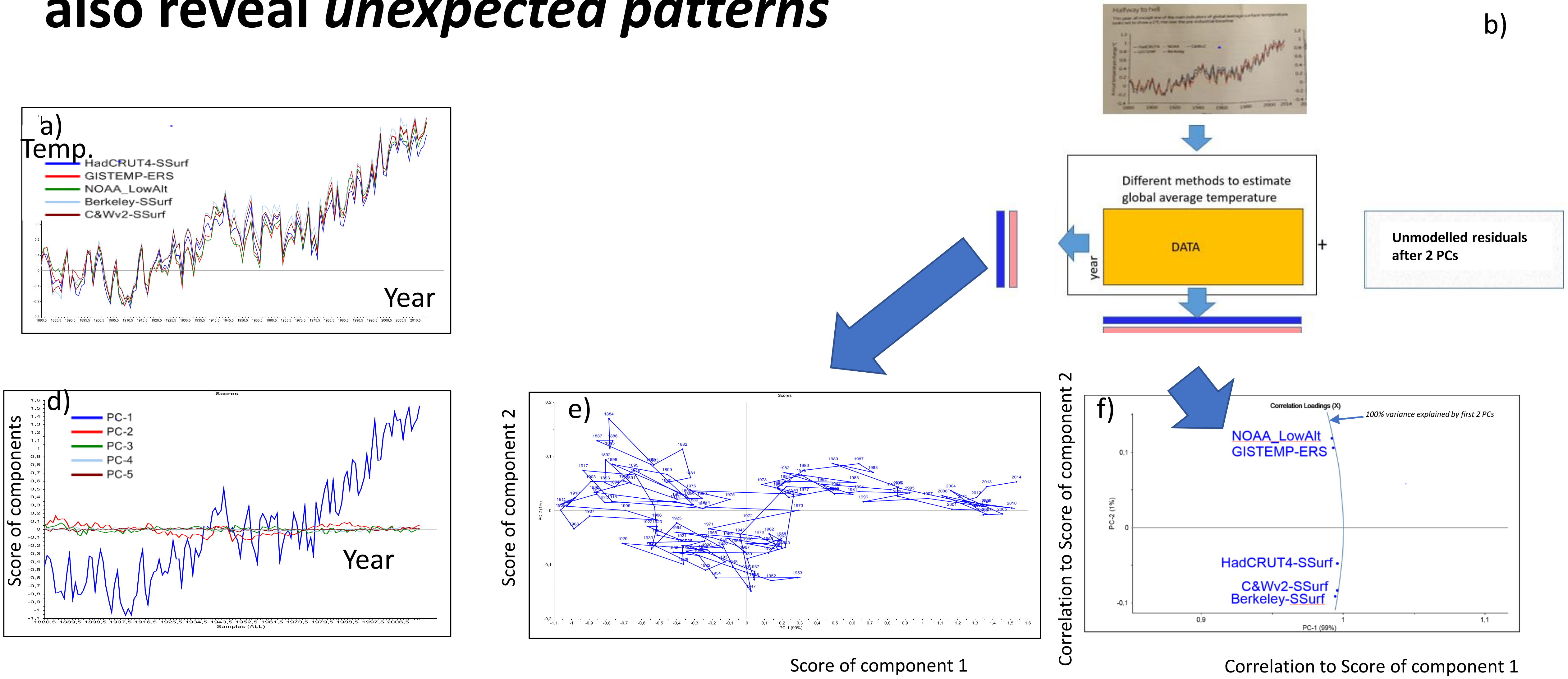
10 min BREAK

Finn sam-variasjonsmønstre i en data-tabell



PCA, MCR, ICA osv

Data driven multivariate modelling will also reveal *unexpected patterns*



Towards a transparent, democratic and secular AI

- Artificial Intelligence (AI): ***A new « religion » ?***

  Automatically building mathematical models from BIG DATA
  ≈ Machine learning
  Previously ANN with sigmoid relations, now often CNN with piecewise linear relations
  For classification of images, language translation, forecasting in time, autonomous cars,…

  Powerful methods. But slow, and difficult to optimize: Need LOTS of GOOD DATA.

  Serious problems: **Black box**, and not always **reliable predictions**.

Towards a transparent, democratic and secular AI

- Artificial Intelligence (AI): ***A new « religion » ?***

  Automatically building mathematical models from BIG DATA

  ≈ Machine learning

  Previously ANN with sigmoid relations, now often CNN with piecewise linear relations

  For classification of images, language translation, forecasting in time, autonomous cars,…

  Powerful methods. But slow, and difficult to optimize: Need **LOTS** of **GOOD** data.

  Serious problems: **Black box**, and not always **reliable predictions**.

- Explainable AI (XAI):

  ≈ *Interpretable* Machine Learning: CNN and hybrid modelling

- Several levels of interpretation:

  1) «Outer level»: Explain why an AI system has made a certain decision, e.g. after an accident

  2) «Inner level»: Discover and interpret new, systematic patterns in data

Towards a transparent, democratic and secular AI

- Artificial Intelligence (AI): ***A new « religion » ?***
  Automatically building mathematical models from BIG DATA
  ≈ Machine learning
  Previously ANN with sigmoid relations, now often CNN with piecewise linear relations
  For classification of images, language translation, forecasting in time, autonomous cars,…

  Powerful methods. But slow, and difficult to optimize: Need **LOTS** of **GOOD** data.

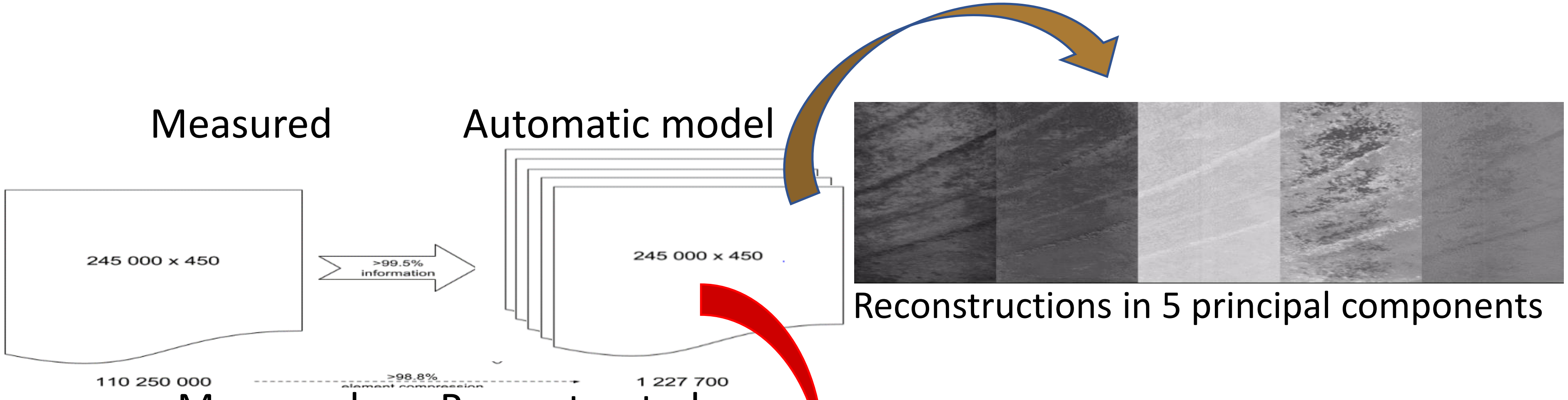
  Serious problems: **Black box**, and not always **reliable predictions**.

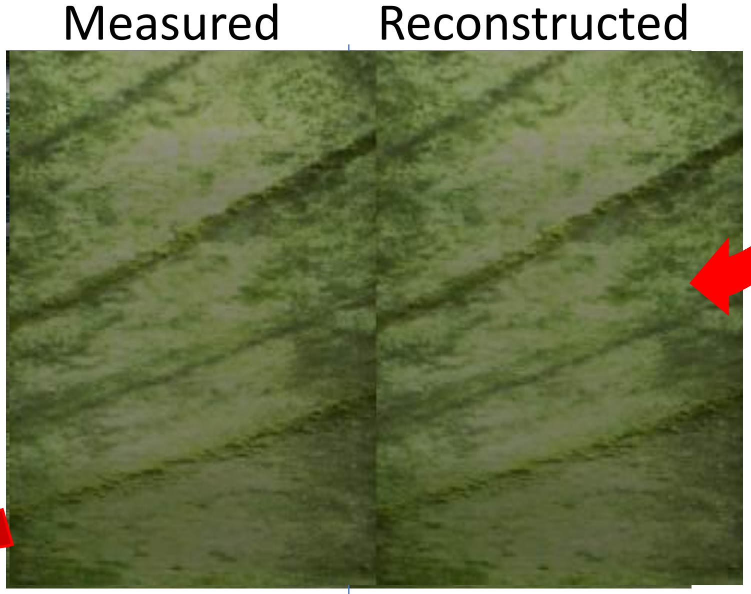
- Explainable AI (XAI):
  ≈ *Interpretable* Machine Learning: CNN and hybrid modelling

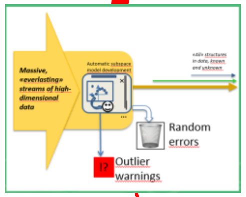
- Several levels of interpretation:
  1) «Outer level»: Explain why an AI system has made a certain decision, e.g. after an accident
  2) «Inner level»: Discover and interpret new, systematic patterns in data
- **My personal research agenda since 1972: Democratic, «secular» data modelling methods: Not mystical!**
  Simple, open to surprise, interpretable in light of prior knowledge

Hyper-spectral camera in small drone for environmental monitoring:

98.8% file reduction, only 0.5% loss (mostly noise)







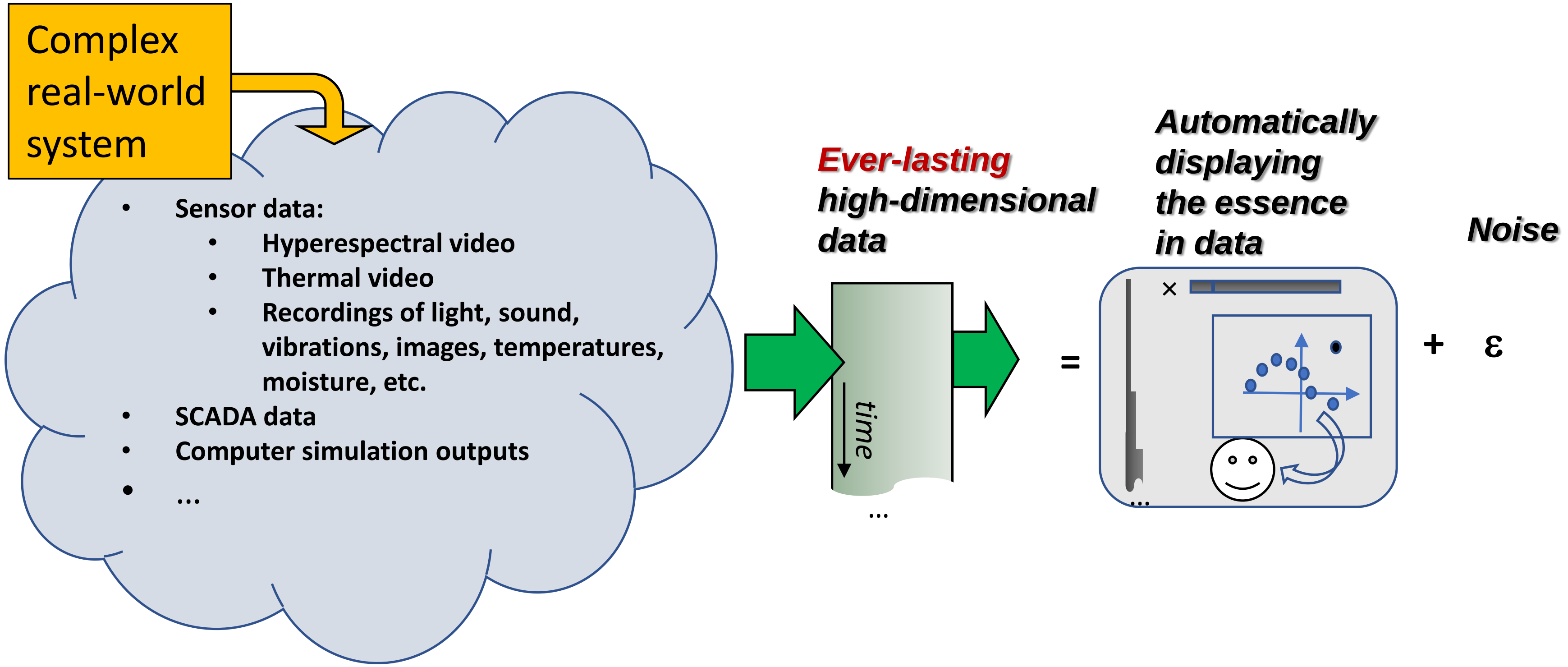
Joao Fortuna, Tor Arne Johansen, Thor Inge Fossen, Harald Martens (2017):

AZORE ISLANDS VEGETATION seen by drone

idletechs

Automatic modelling of “ever-lasting” data streams

From raw streams of data, systematic patterns and relationships are automatically discovered and modelled. The data is stored in a highly compressed format:

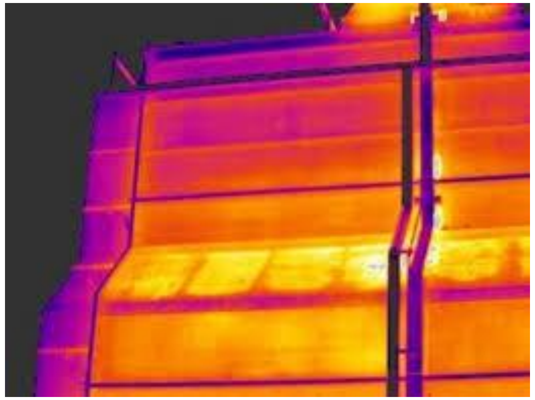


Raffaele Vitale, Anna Zhyrova, João F. Fortuna, Onno E. de Noord, Alberto Ferrer, Harald Martens: On-The-Fly Processing of continuous high-dimensional data streams Chemometrics and Intelligent Laboratory Systems 161 (2017) 118–129

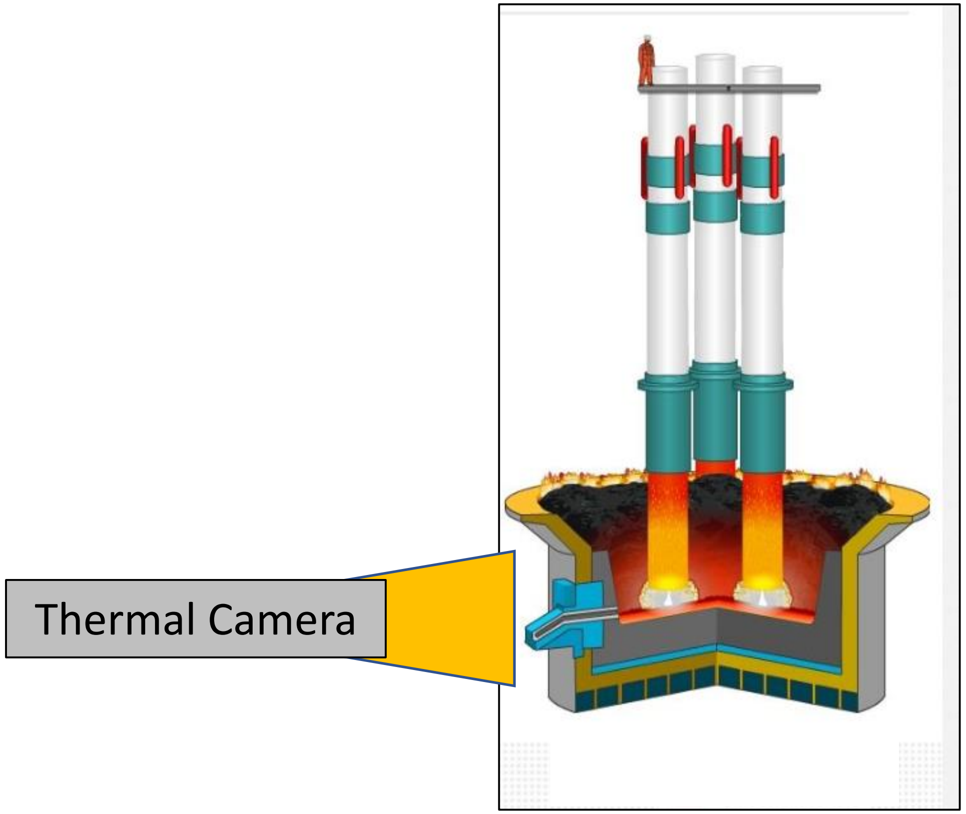
Idletechs AS provides new software tools for automatic modelling & compression and for human interpretation, of multidimensional, dynamic data.

idletechs

Thermal camera



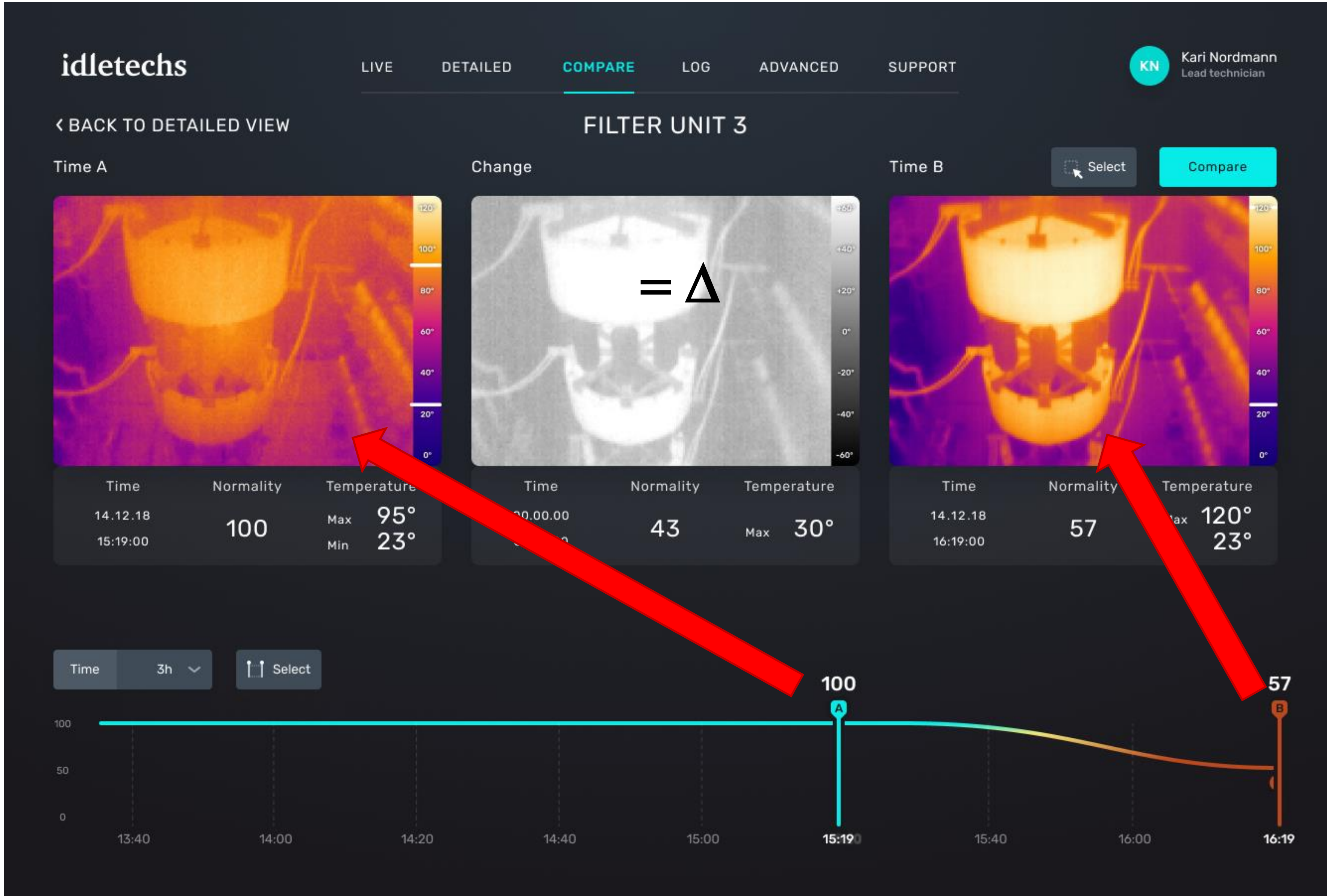
Purpose	Monitor furnace temperatures, e.g. outer surface, electrode or tap hole area, to detect anomalities and unexpected trends



Related example:

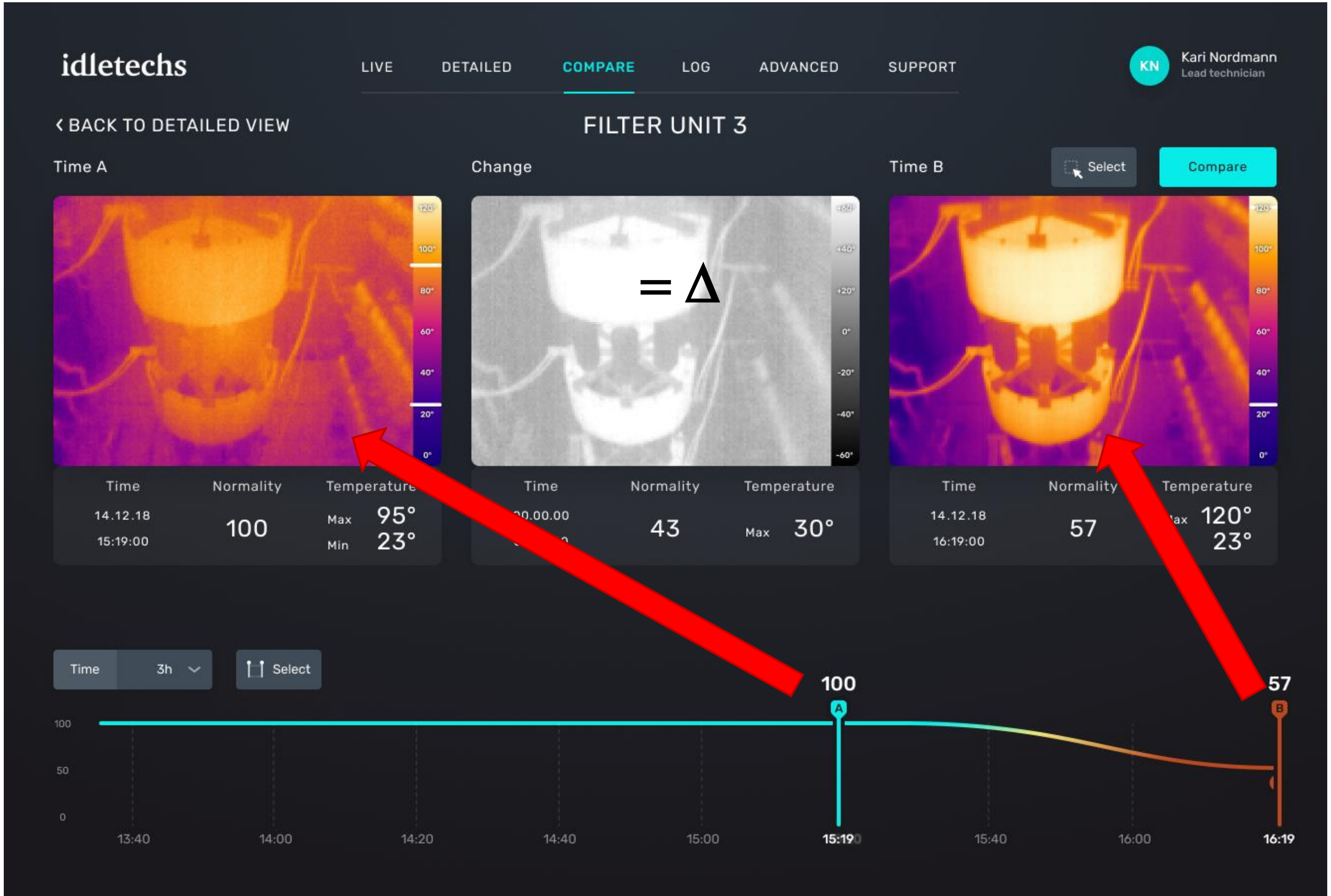
Continuous monitoring of wood ovens:
Heating efficiency experiment at SINTEF 2018

Making quantitative data understandable for ordinary people: e.g. XAI for thermal analysis of high-power electrical equipment



idletechs

Making quantitative data understandable for ordinary people: e.g. XAI for thermal analysis of high-power electrical equipment



idletechs

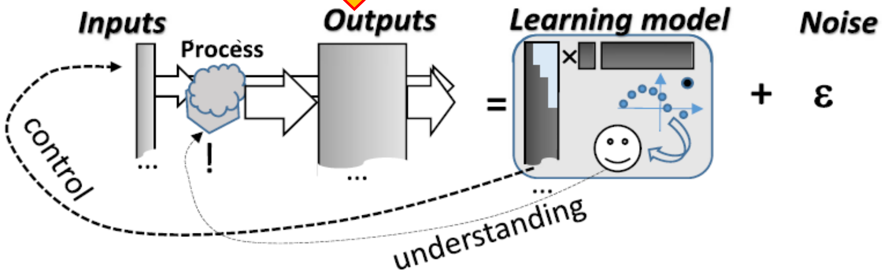
Demo example (non-commercial ☺): Home appliance equipment

idletechs

Experiment setup:

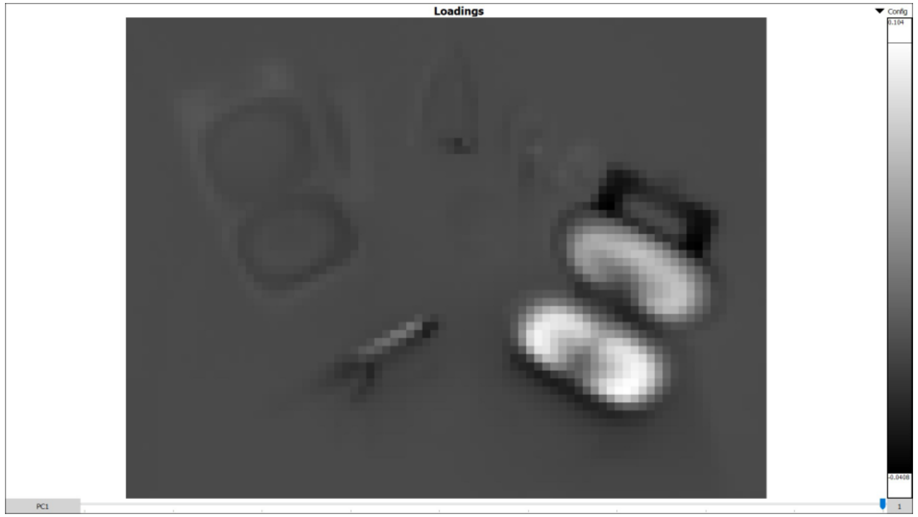
- Instruments:
  - waffle iron
  - burger grill
  - curling iron
  - clothes iron
- Disturbances:
  - water bottle
  - tea cup
  - hair dryer
  - human interf
  - timers

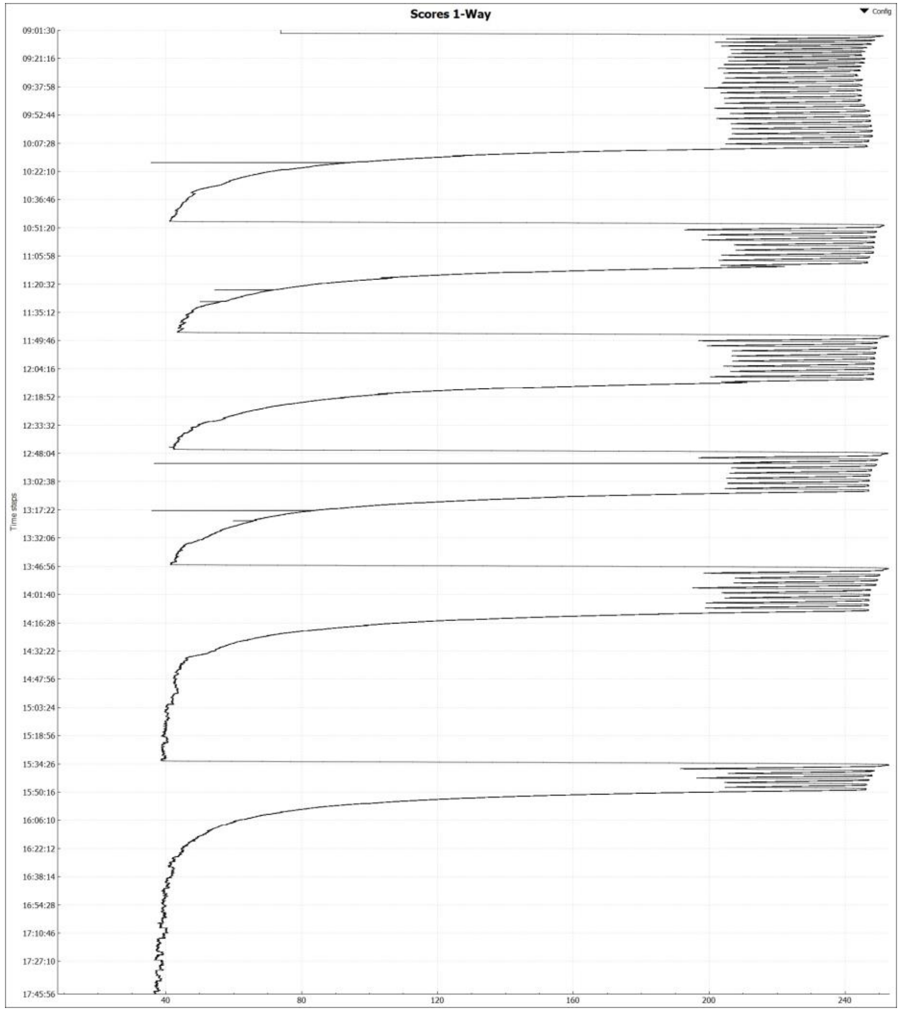
Thermal camera pixels



Discovered State variable: Hamburger grill

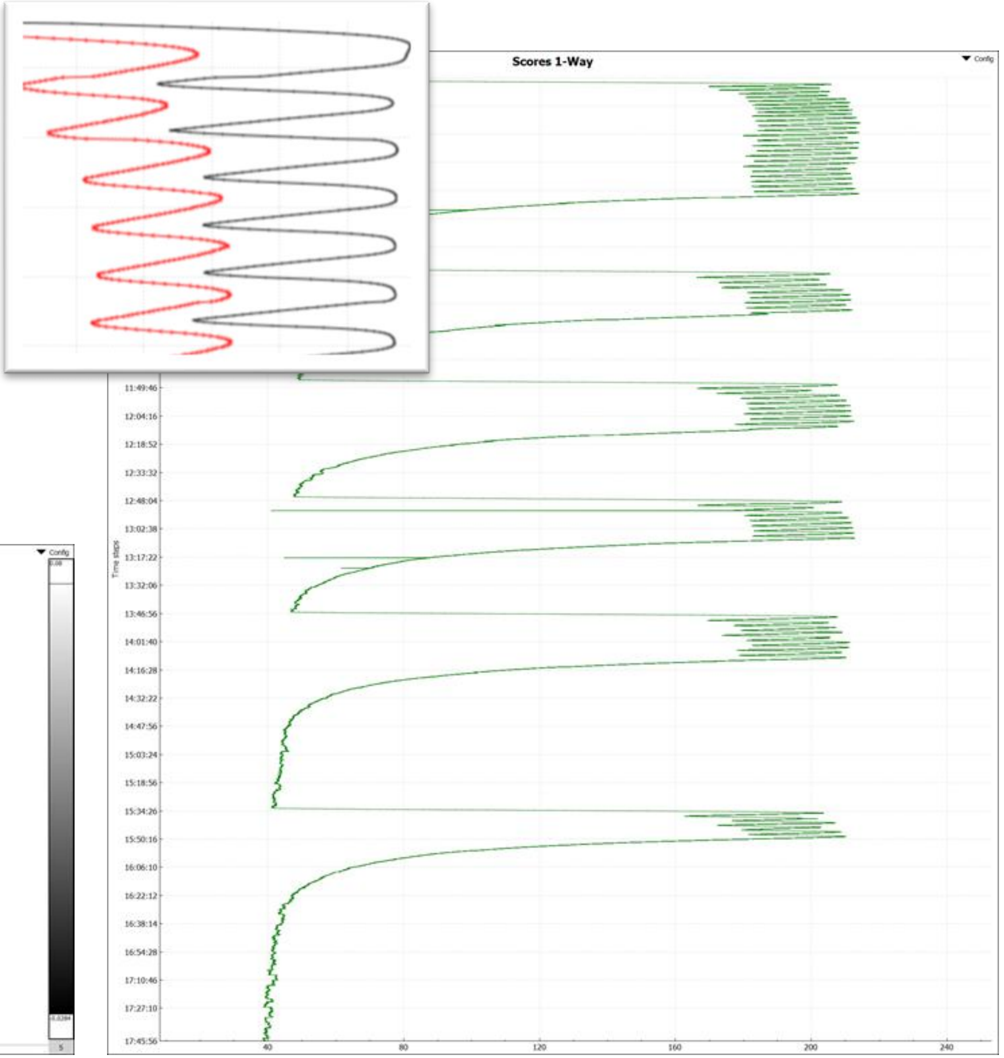
- Trends of two timers
  - Built-in thermostat in equipment
  - 30min timer on equipment
    - Deviation in end, probable mistake in experiment

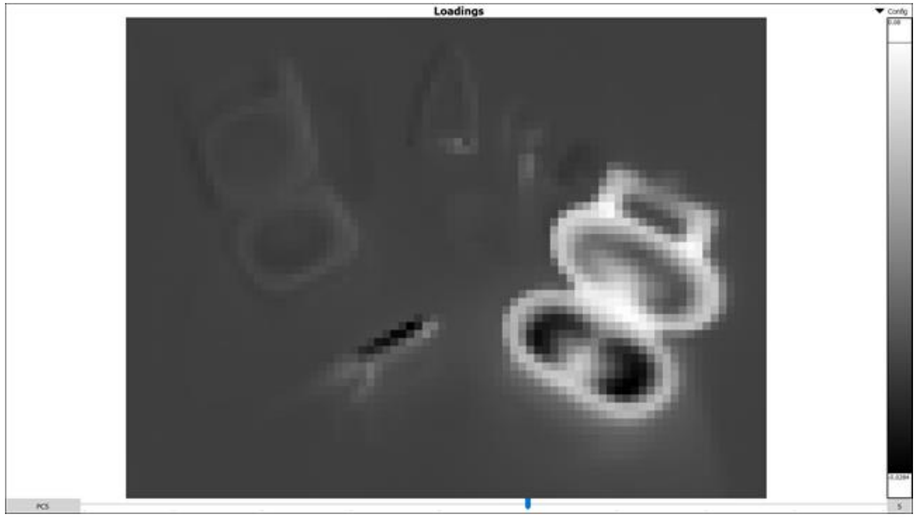




Discovered State variable: Heat dissipation hamburger grill

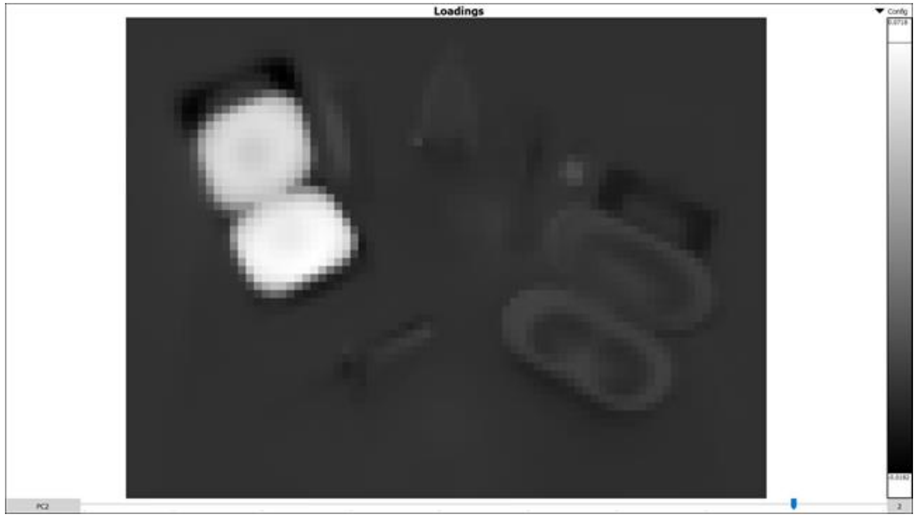
- Same timer trends as hamburger grill
- Small phase offset from heat source, suggests heat dissipation

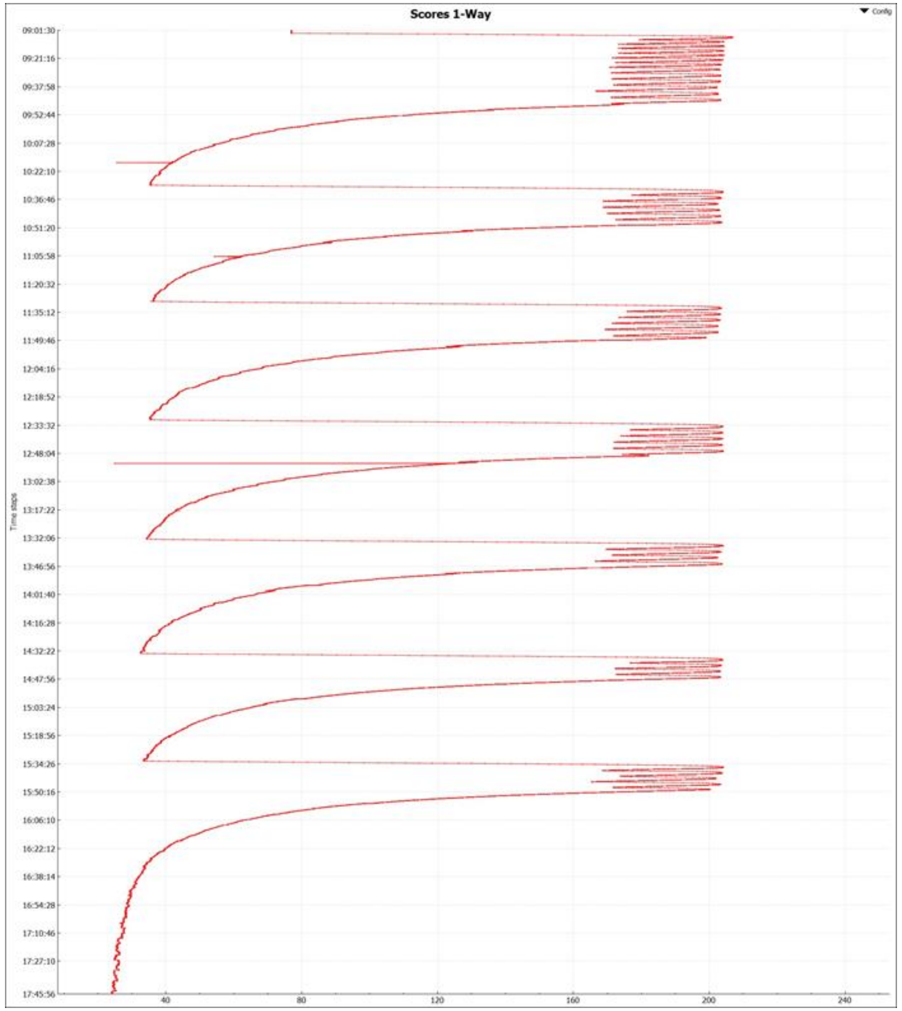




Discovered State variable: Waffle iron

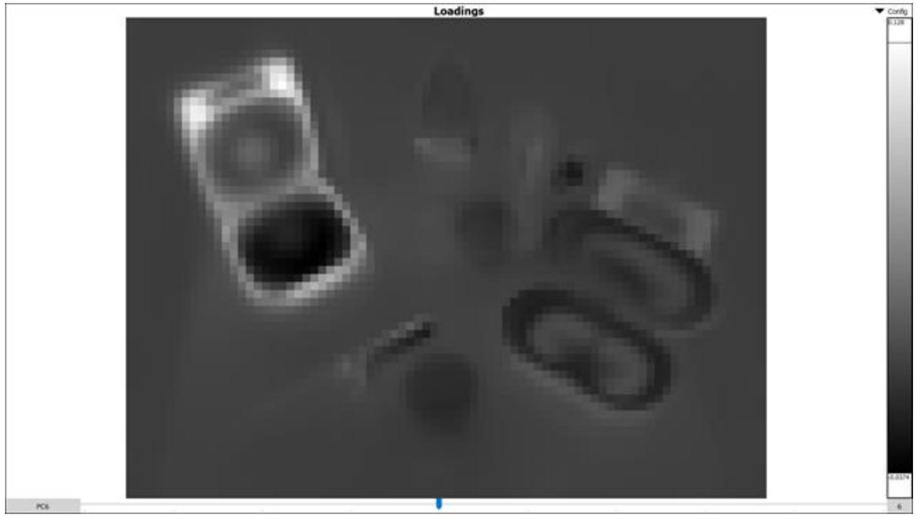
- Trends of two timers
  - Built-in thermostat in equipment
  - 30min timer on equipment

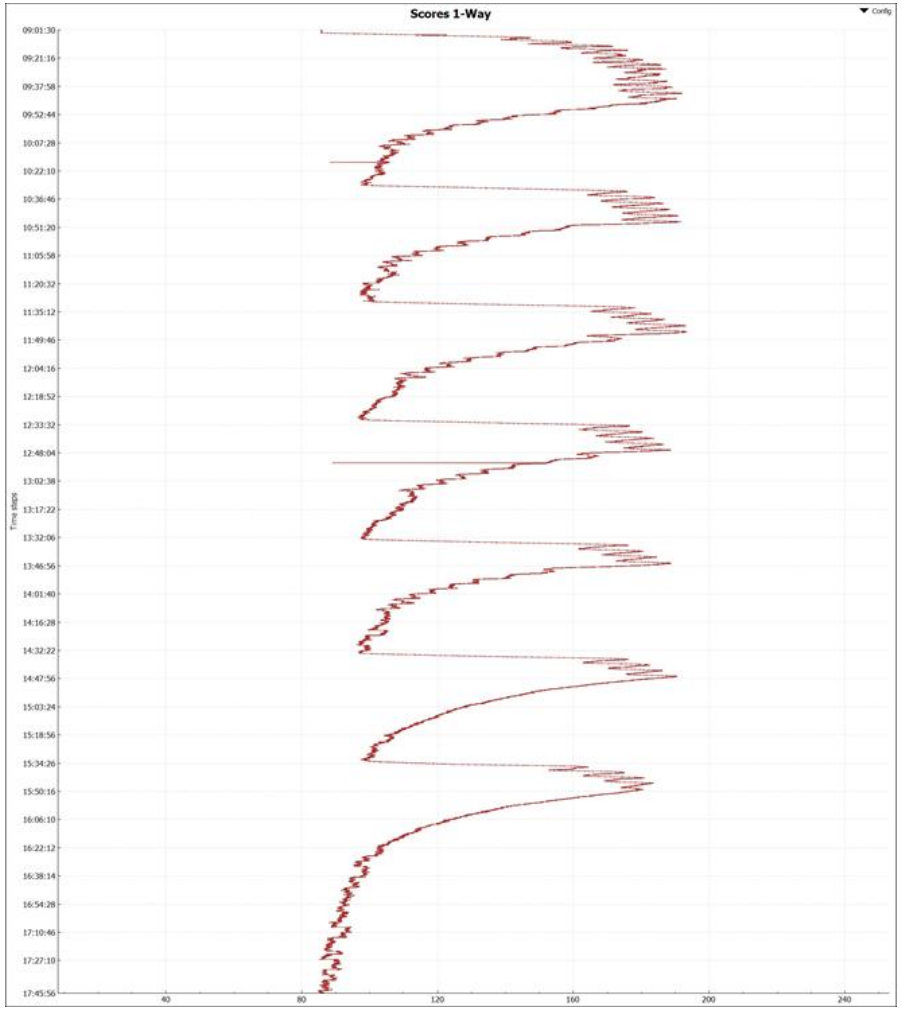




Discovered State variable : Heat dissipation waffle iron

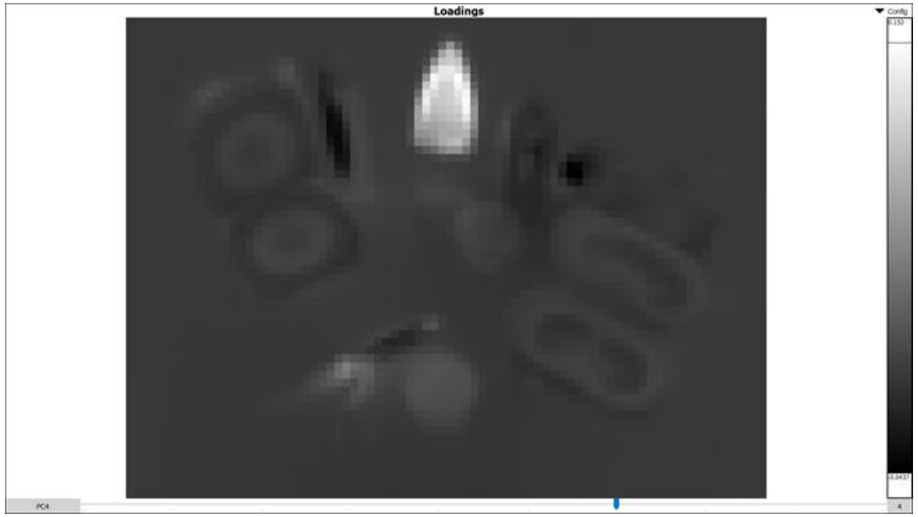
- Same timer trends as waffle iron
- Small phase offset from heat source, suggests heat dissipation

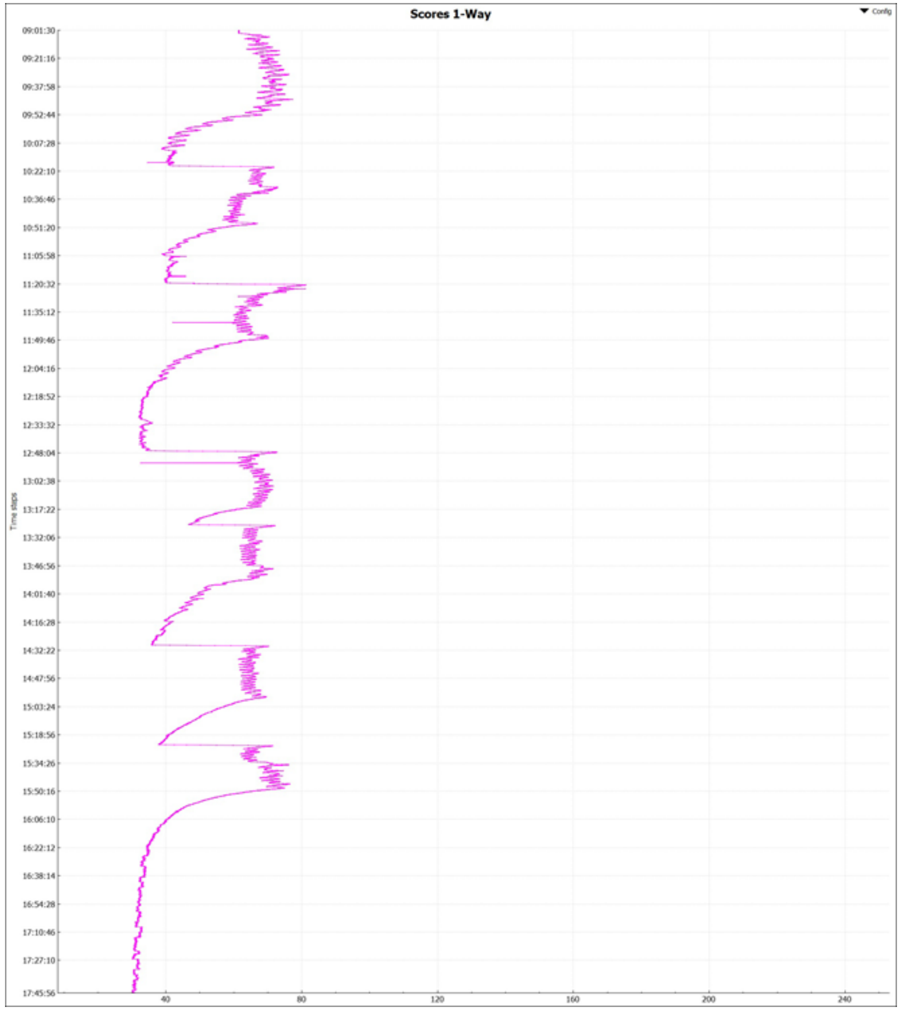




Discovered State variable : Clothes iron

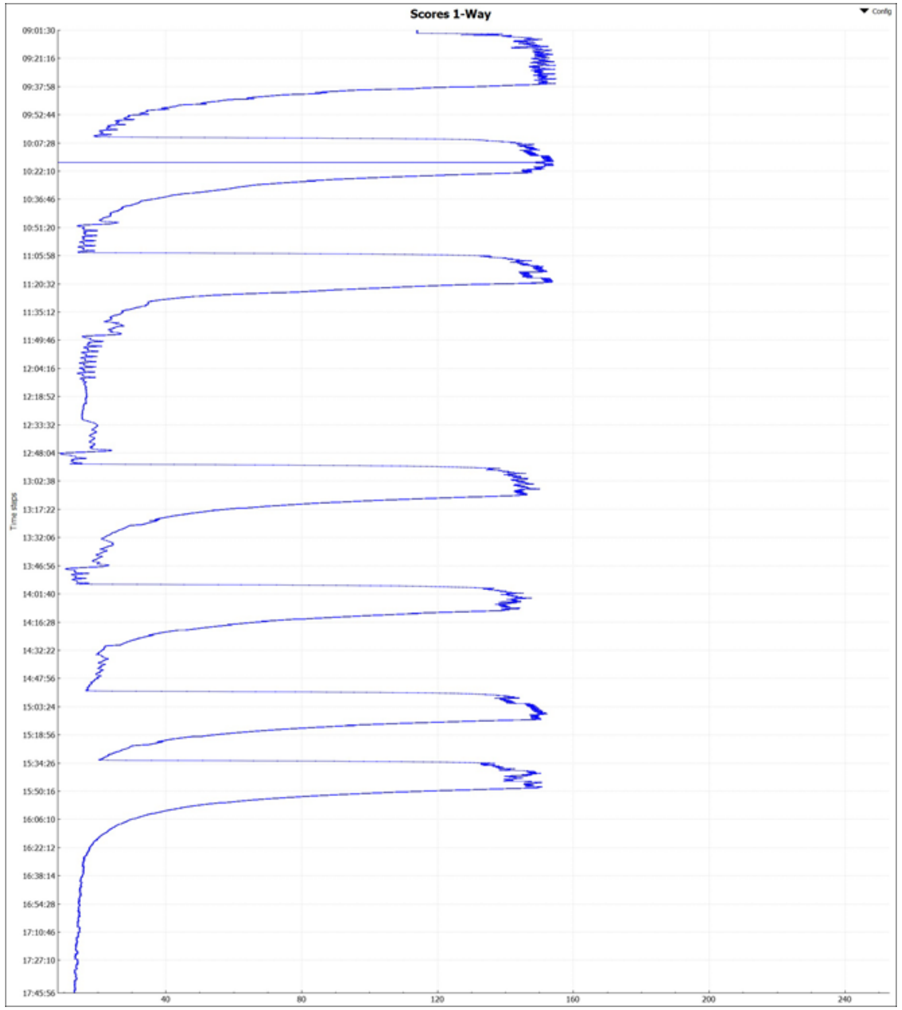
- Trends of two timers
  - Built-in thermostat in equipment
  - Signs of a user manually adjusting timer





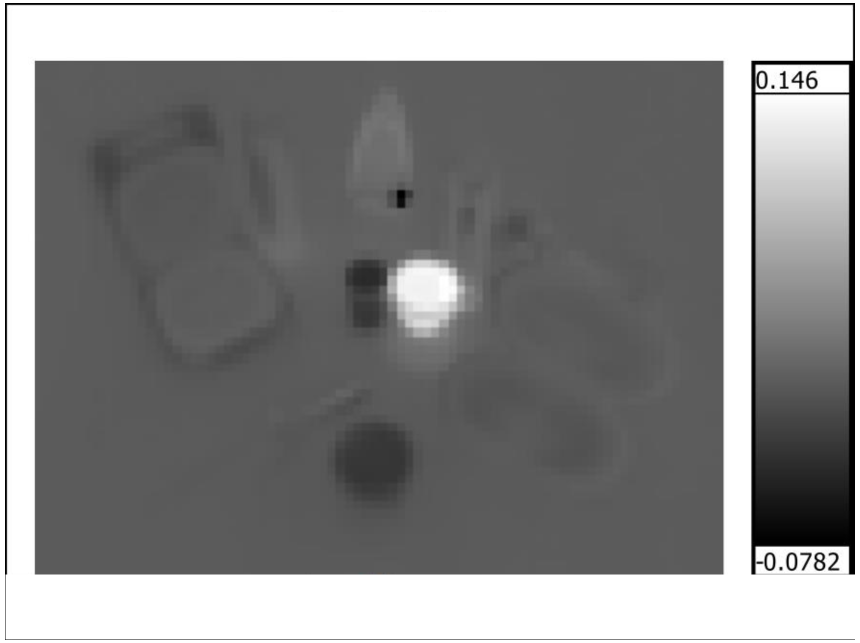
Discovered State variable : Curling iron

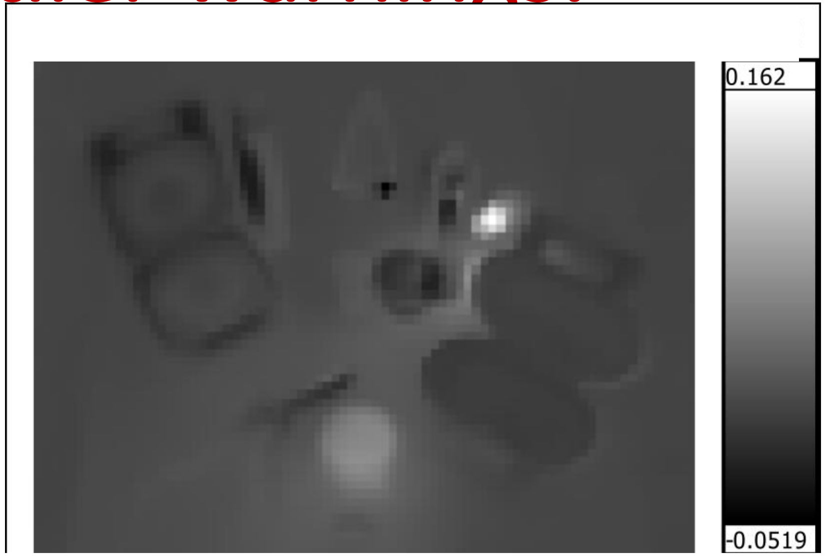
- Trends of two timers
    - Built-in thermostat in equipment
    - 30min timer on equipment
        - Manual timers in end of day
- Deviation around lunch
    - User paused equipment due to potential fire hazard



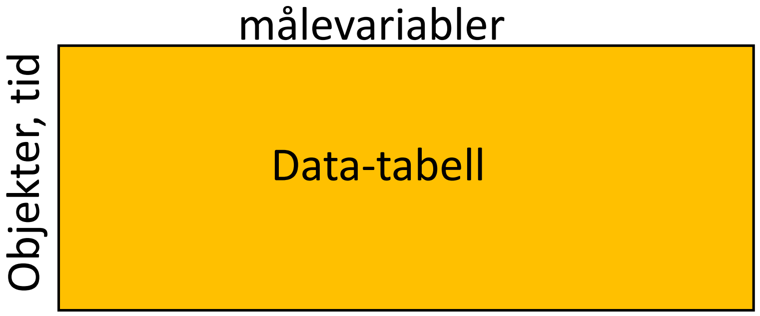
After all “natural” variations discovered, modelled and subtracted: Small but systematic Residuals

→ Fewer, more sensitive outlier warnings:



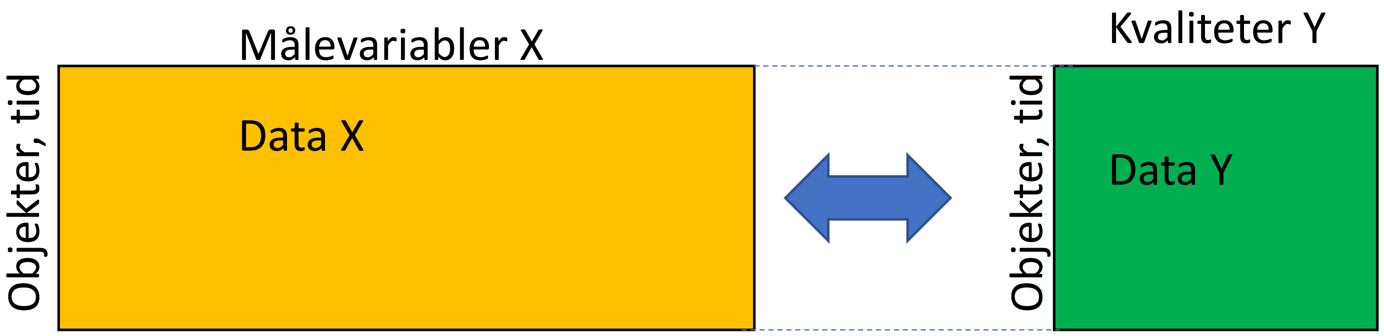


Finn sam-variasjonsmønstre i en data-tabell



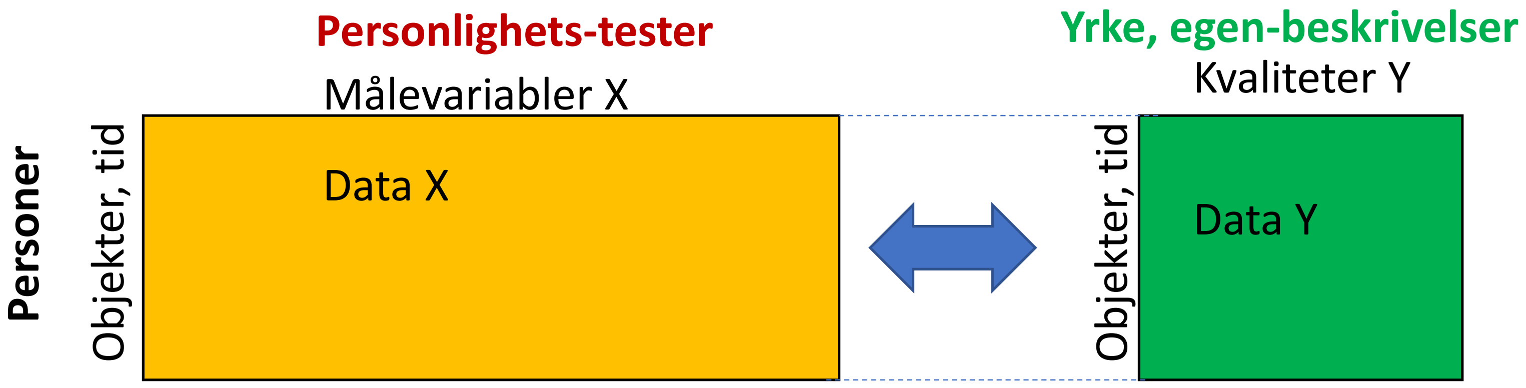
PCA, MCR, ICA osv

Finn sam-variasjonsmønstre mellom to data-tabeller



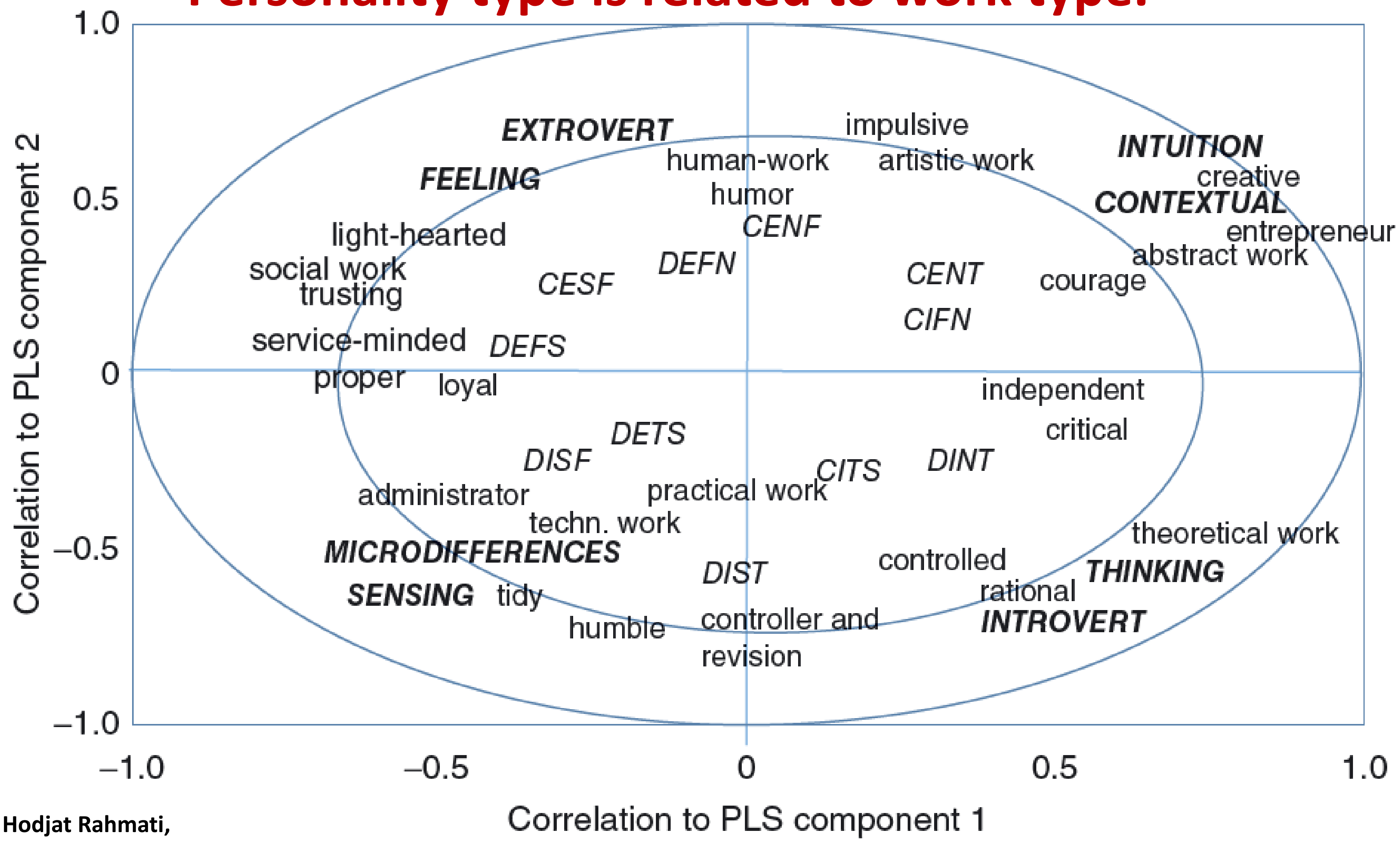
PCR, PLSR, RR osv

Finn sam-variasjonsmønstre mellom to data-tabeller



PCR, PLSR, RR osv

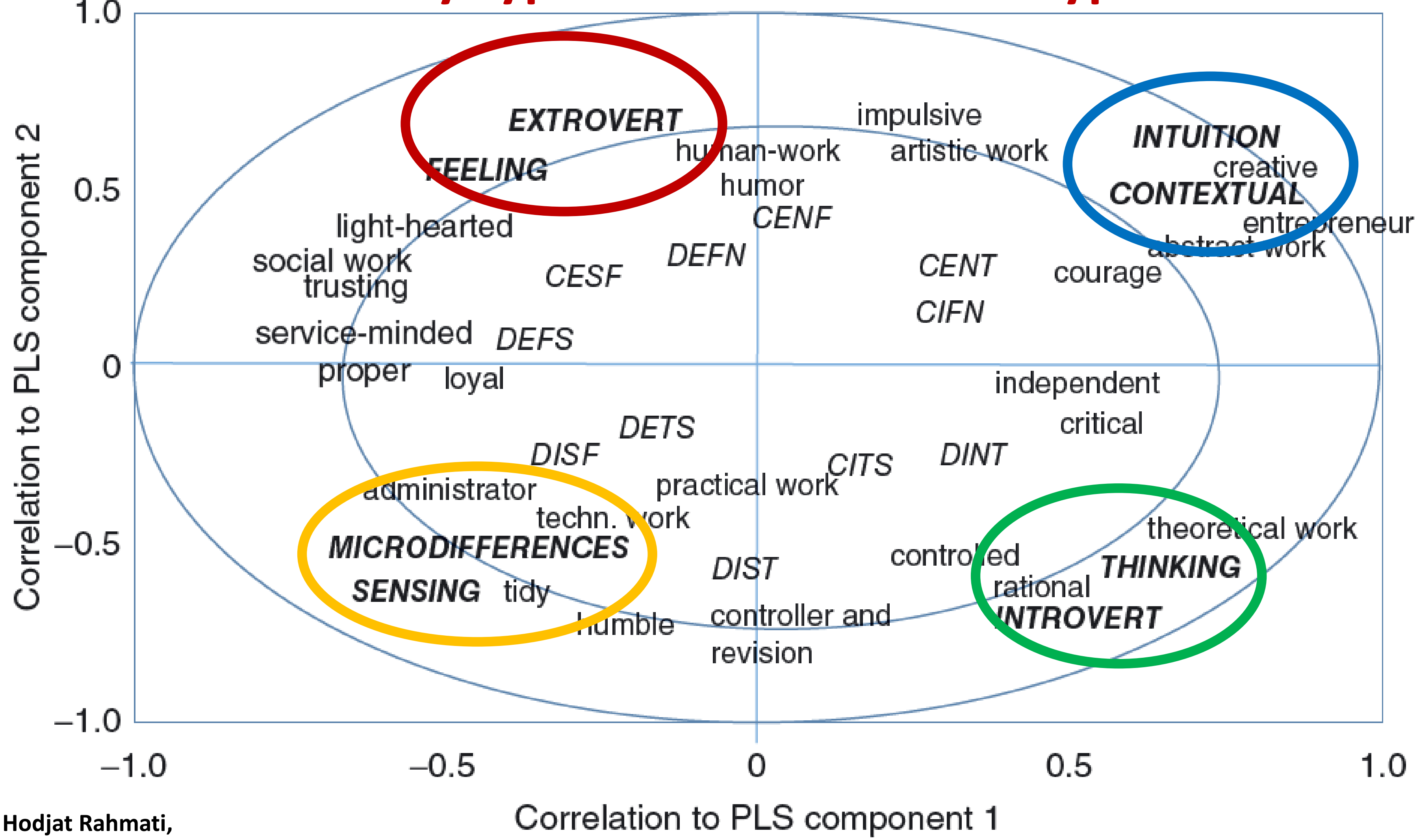
Personality type is related to work type.



**Nils K. Skjærvold, Helge Brovold, Hodjat Rahmati,
Harald Martens, Kristin Tøndel, Gunnar Cedersund, Lars M. Munck (2017)**

Encyclopedia of Analytical Chemistry, Online © 2006–2017 JohnWiley & Sons, Ltd.

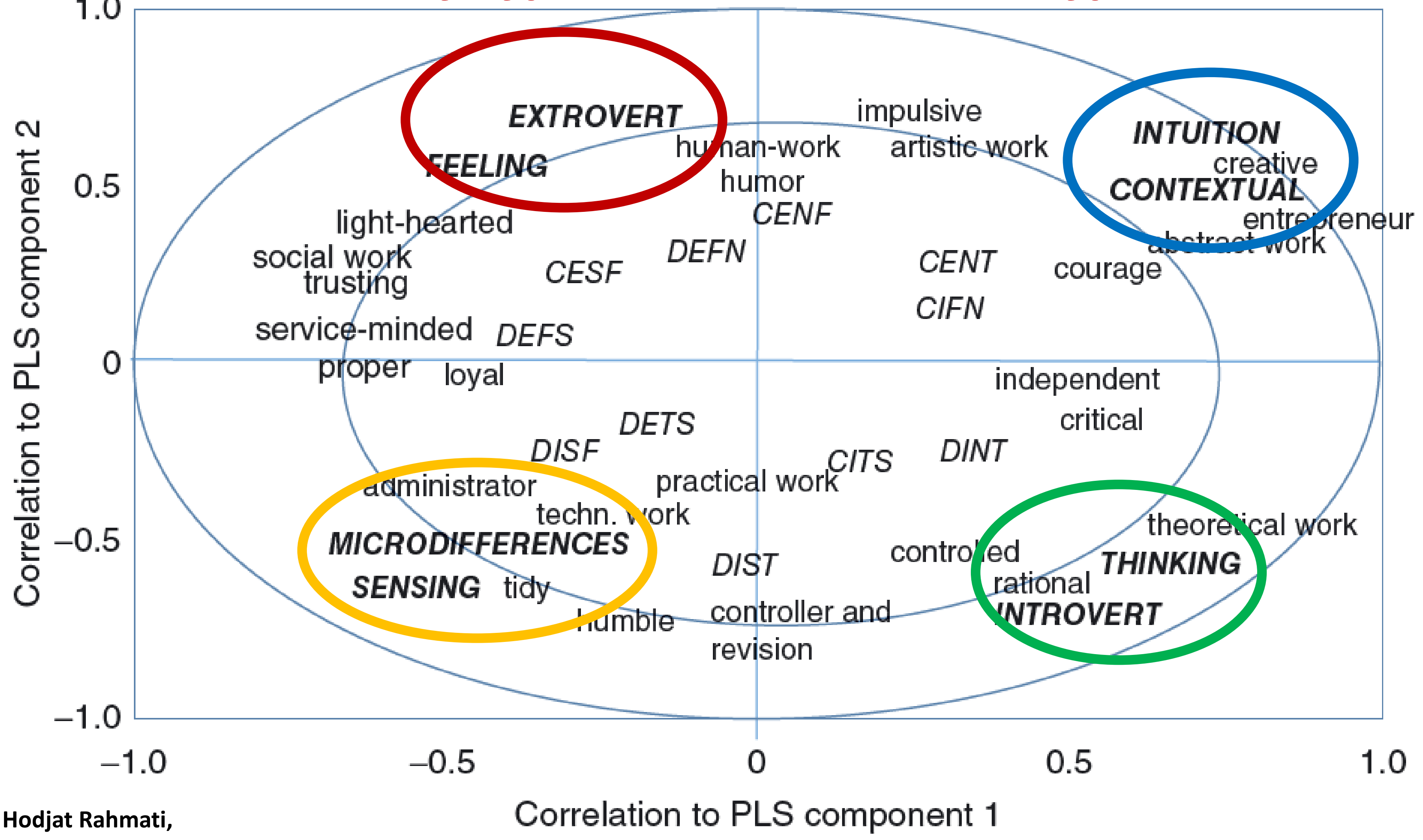
Personality type is related to work type.



Nils K. Skjærvold, Helge Brovold, Hodjat Rahmati,
Harald Martens, Kristin Tøndel, Gunnar Cedersund, Lars M. Munck (2017)

Encyclopedia of Analytical Chemistry, Online © 2006–2017 JohnWiley & Sons, Ltd.

Personality type is related to work type.



Nils K. Skjærvold, Helge Brovold, Hodjat Rahmati,
Harald Martens, Kristin Tøndel, Gunnar Cedersund, Lars M. Munck (2017)

Encyclopedia of Analytical Chemistry, Online © 2006–2017 JohnWiley & Sons, Ltd.

How to teach mathematics to all types of scientists?

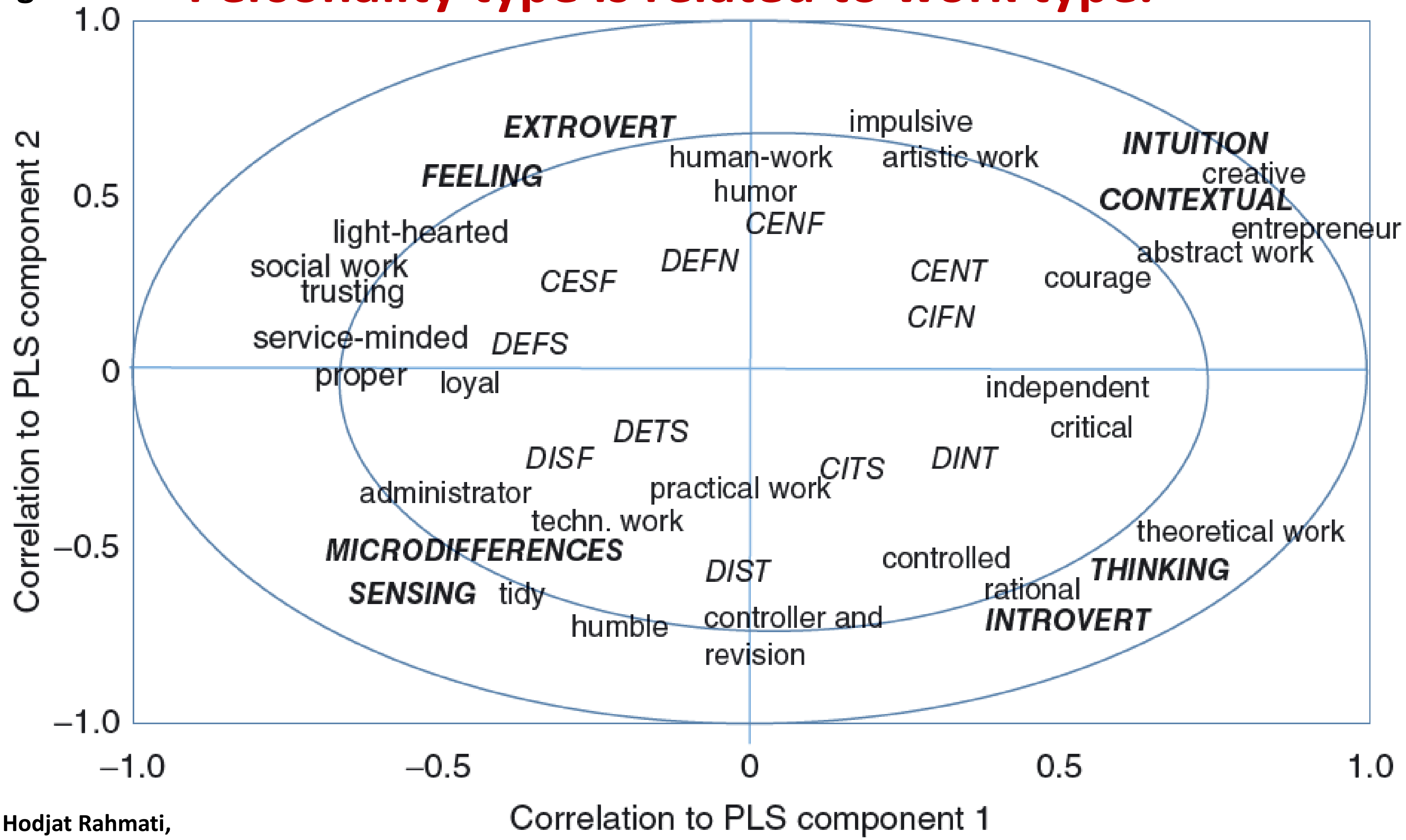
Matematikk uten tårer

Psykolog Helge Brovold, PhD:

Fire veier inn i matematikken.

Data fra 2200 jobbsøkere i Norge:

Personality type is related to work type.



Nils K. Skjærvold, Helge Brovold, Hodjat Rahmati,
Harald Martens, Kristin Tøndel, Gunnar Cedersund, Lars M. Munck (2017)

Encyclopedia of Analytical Chemistry, Online © 2006–2017 JohnWiley & Sons, Ltd.

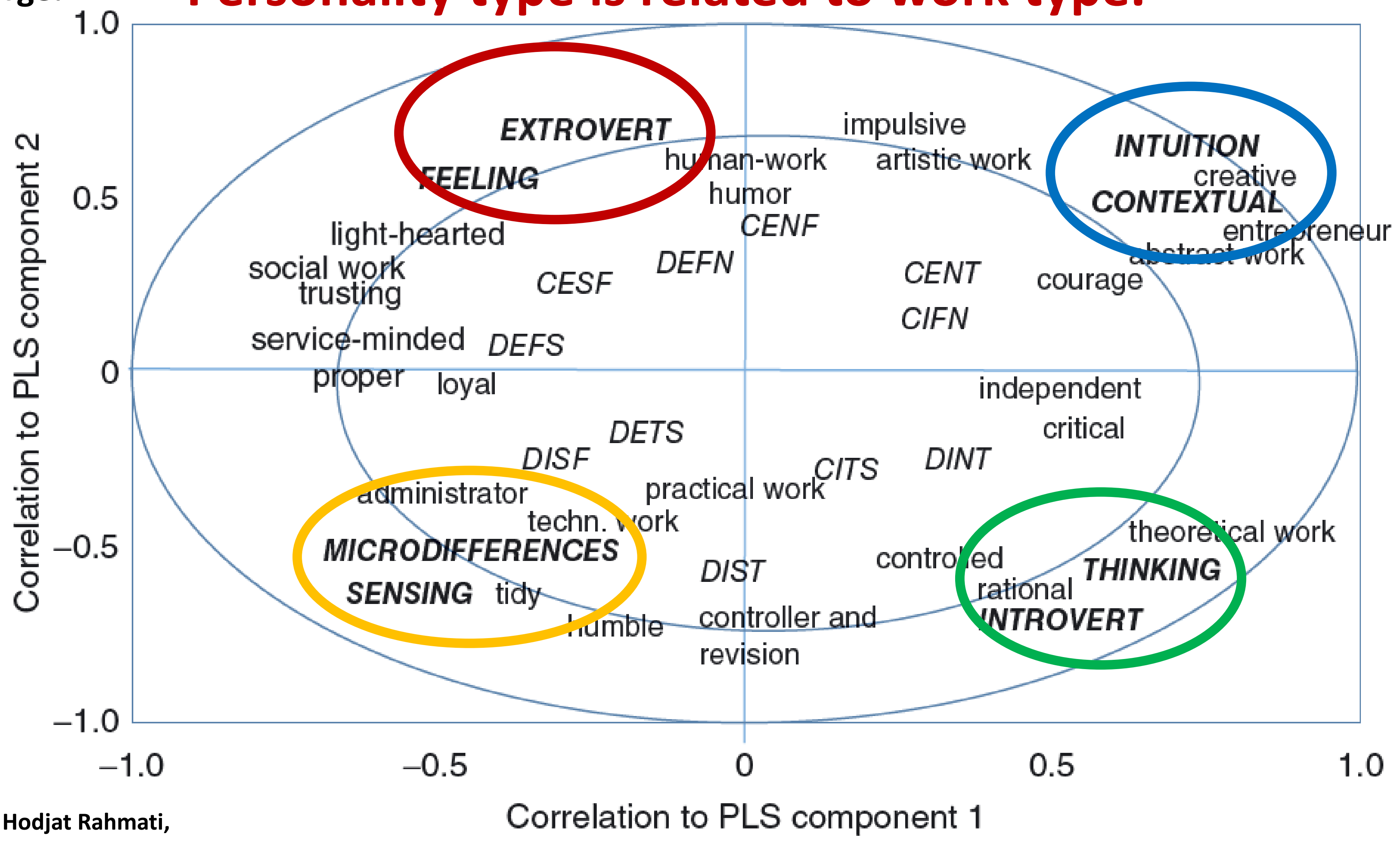
Matematikk uten tårer

Psykolog Helge Brovold, PhD:

Fire veier inn i matematikken.

Data fra 2200 jobbsøkere i Norge:

Personality type is related to work type.



Nils K. Skjærvold, Helge Brovold, Hodjat Rahmati,
Harald Martens, Kristin Tøndel, Gunnar Cedersund, Lars M. Munck (2017)

Encyclopedia of Analytical Chemistry, Online © 2006–2017 JohnWiley & Sons, Ltd.

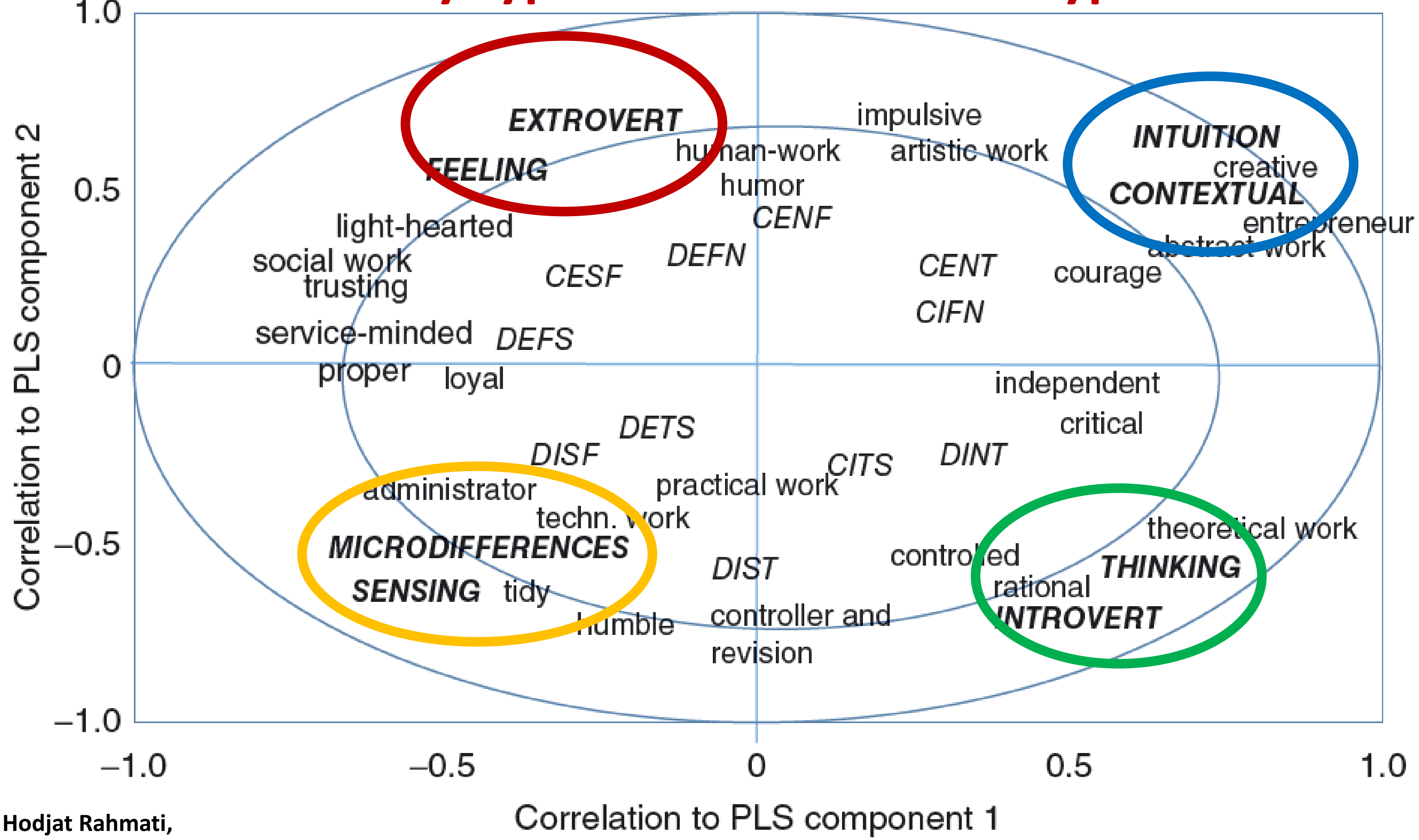
Matematikk uten tårer

Psykolog Helge Brovold, PhD:

Fire veier inn i matematikken.

Data fra 2200 jobbsøkere i Norge:

Personality type is related to work type.

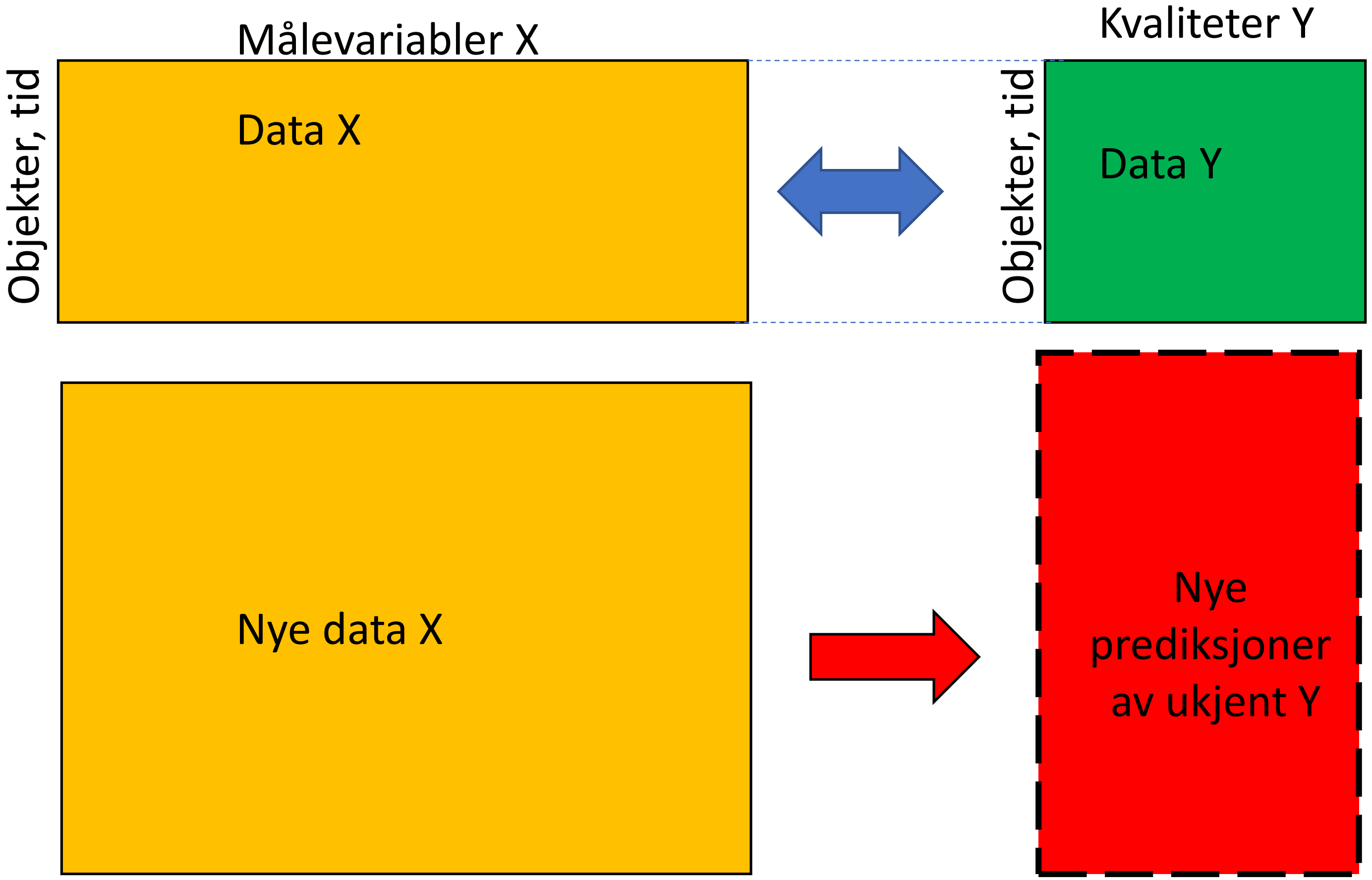


Nils K. Skjærvold, Helge Brovold, Hodjat Rahmati,
Harald Martens, Kristin Tøndel, Gunnar Cedersund, Lars M. Munck (2017)

Encyclopedia of Analytical Chemistry, Online © 2006–2017 JohnWiley & Sons, Ltd.

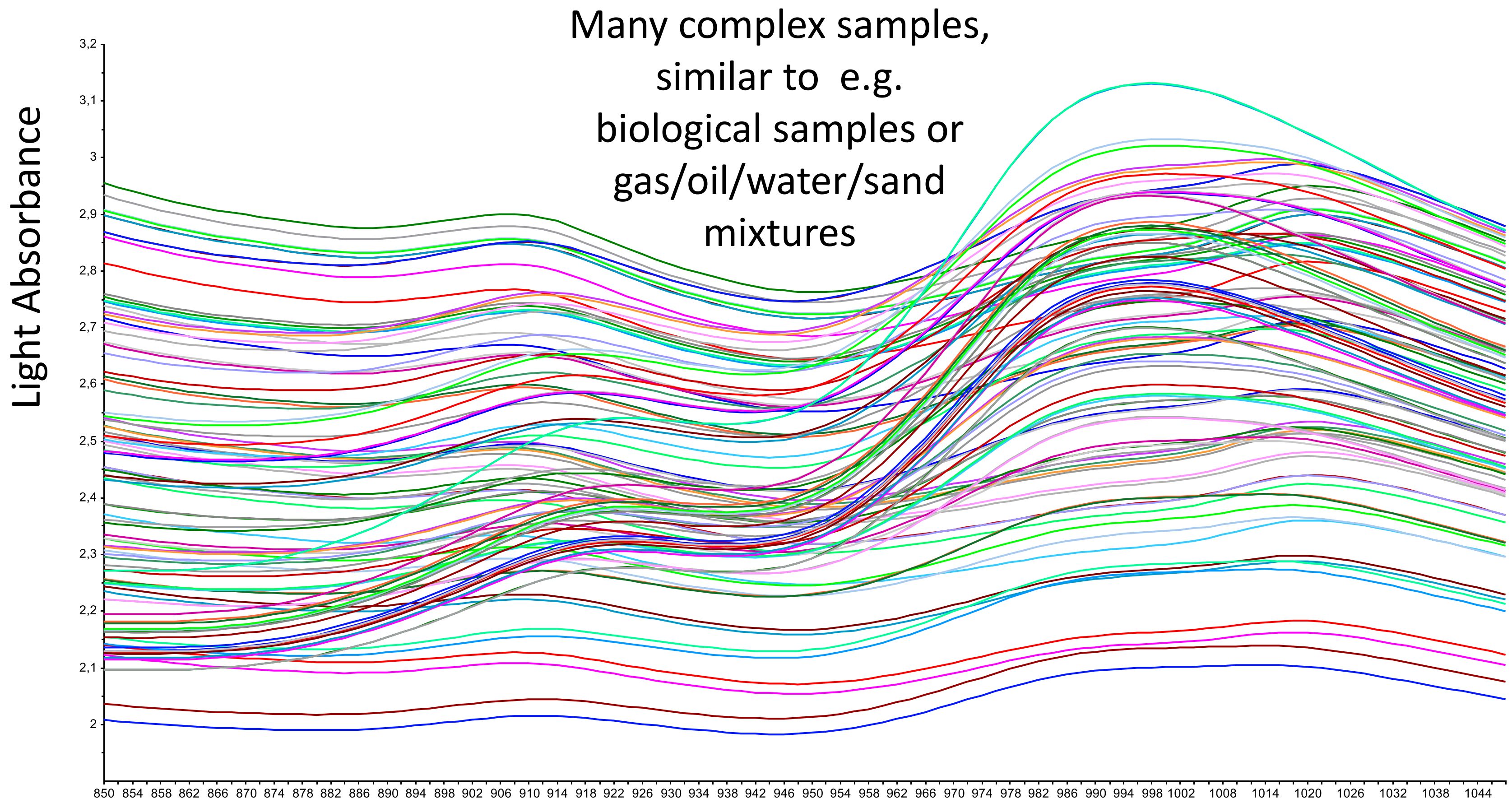
How to teach mathematics to all types of scientists?

Finn sam-variasjonsmønstre mellom to data-tabeller, bruk modell til prediksjon



PCR, PLSR, RR osv

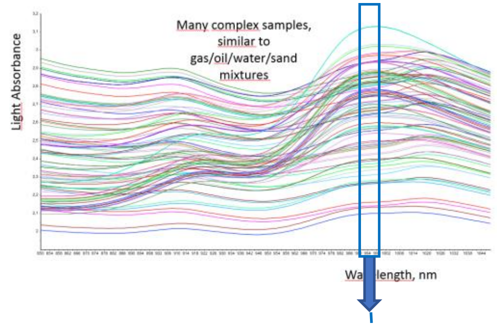
Instrumentation example: Multivariate calibration of multi-wavelength NIR absorbance process spectroscopy

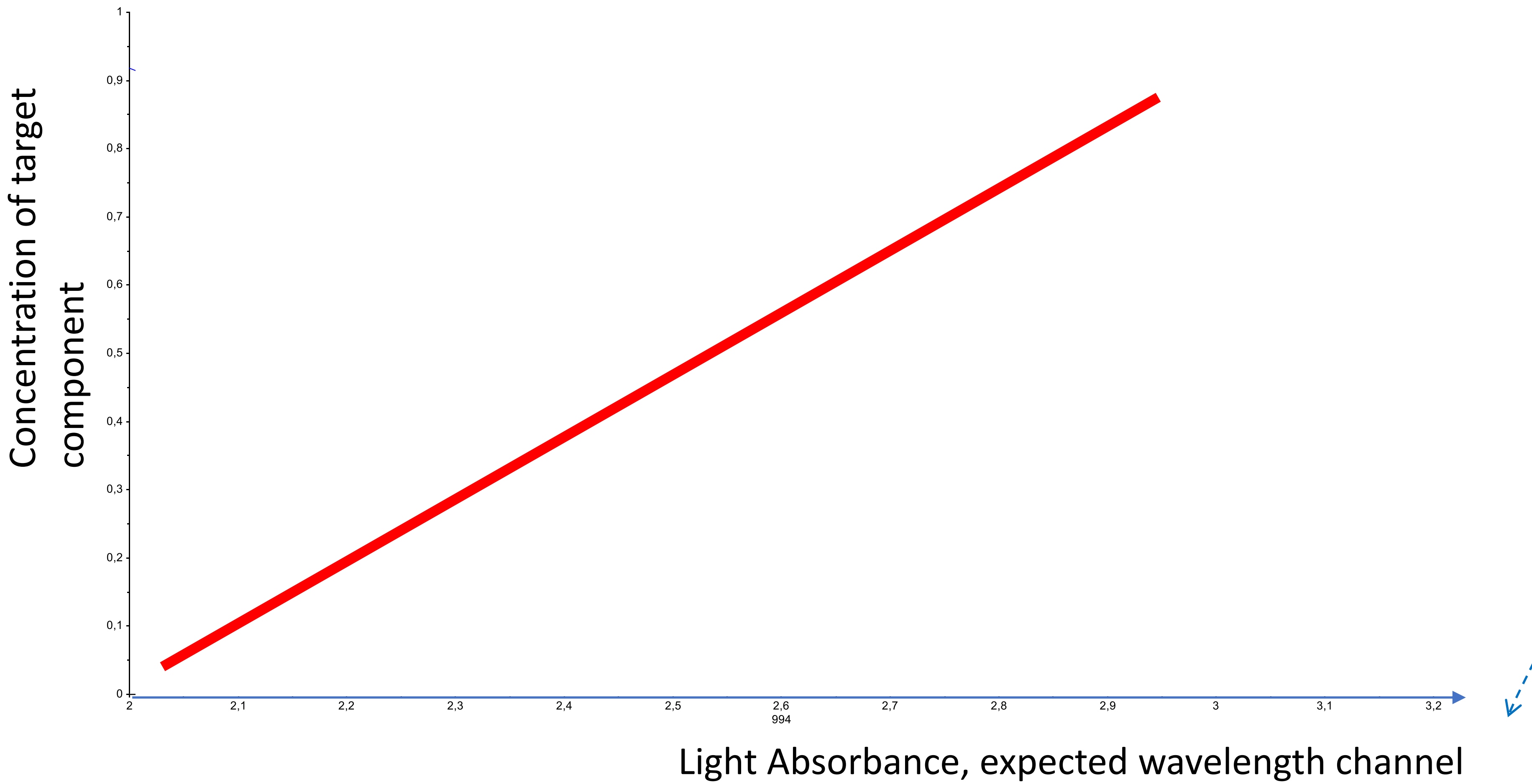


Wavelength, nm
100 channels

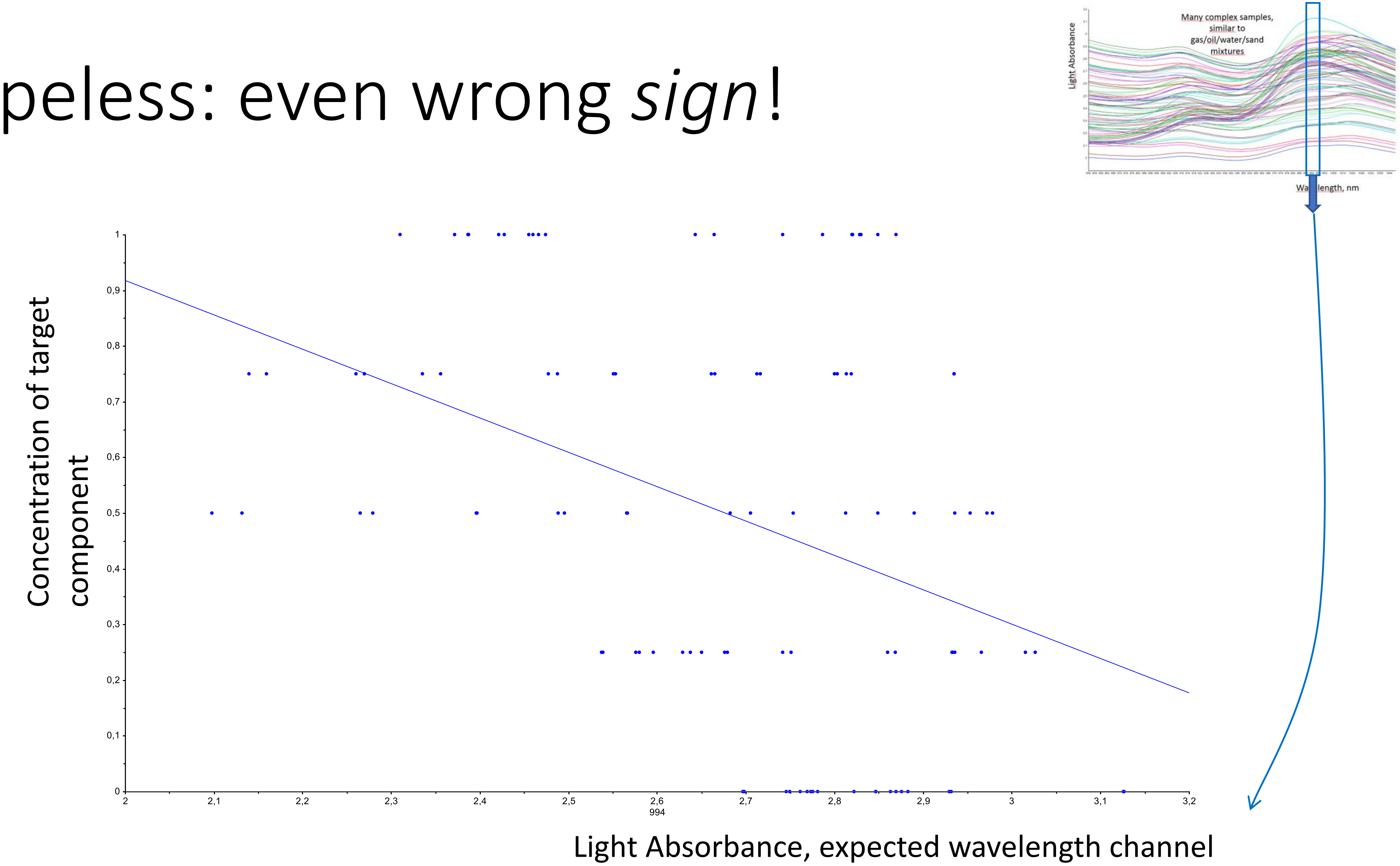
Chemical mixtures of protein and starch powders,
measured under different physical conditions
• streams of wheat flour in a grain mill

Hope: Nice calibration!

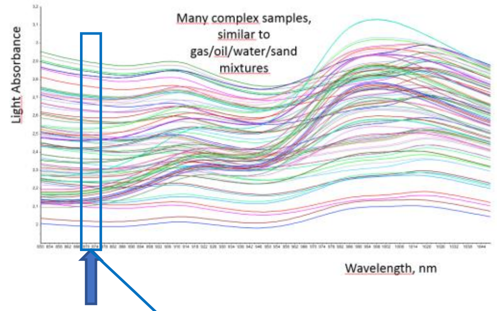


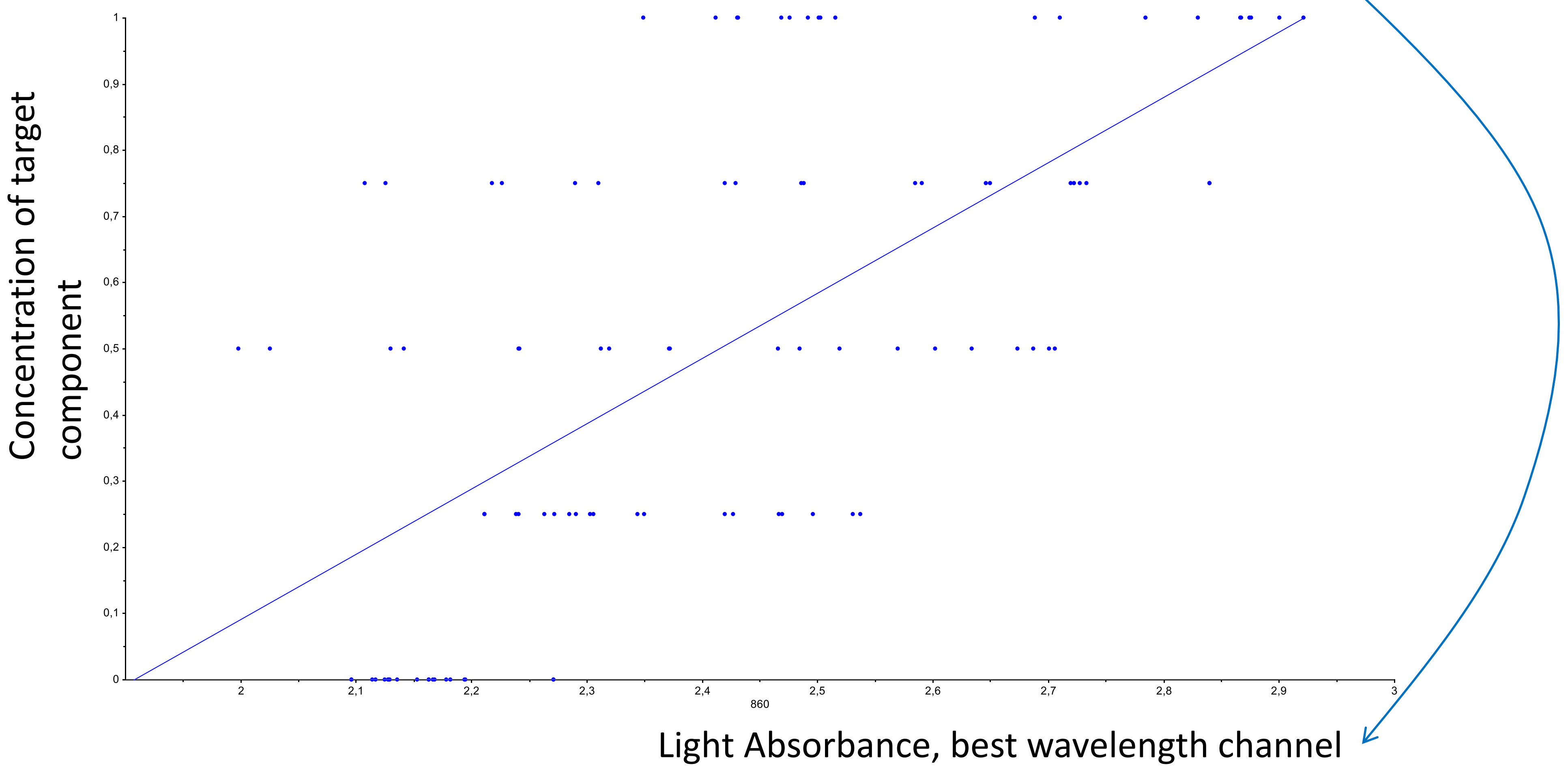


Hopeless: even wrong *sign*!



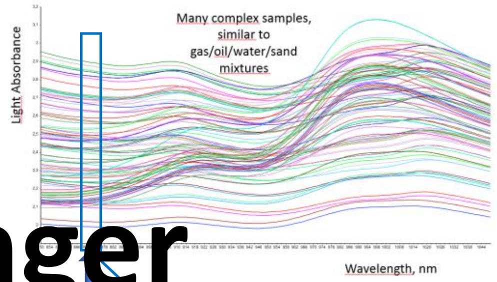
«Best» channel: useless calibration

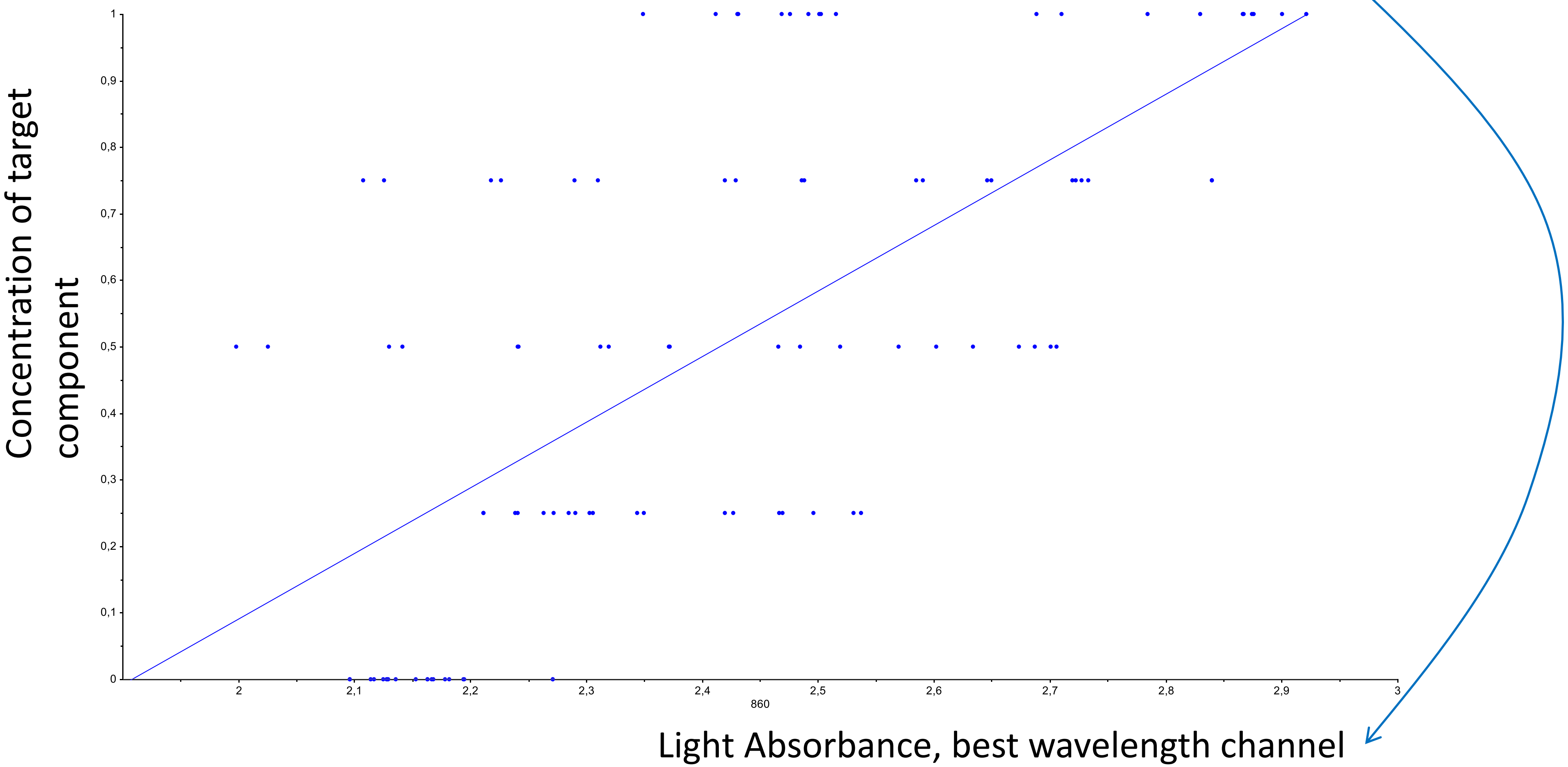




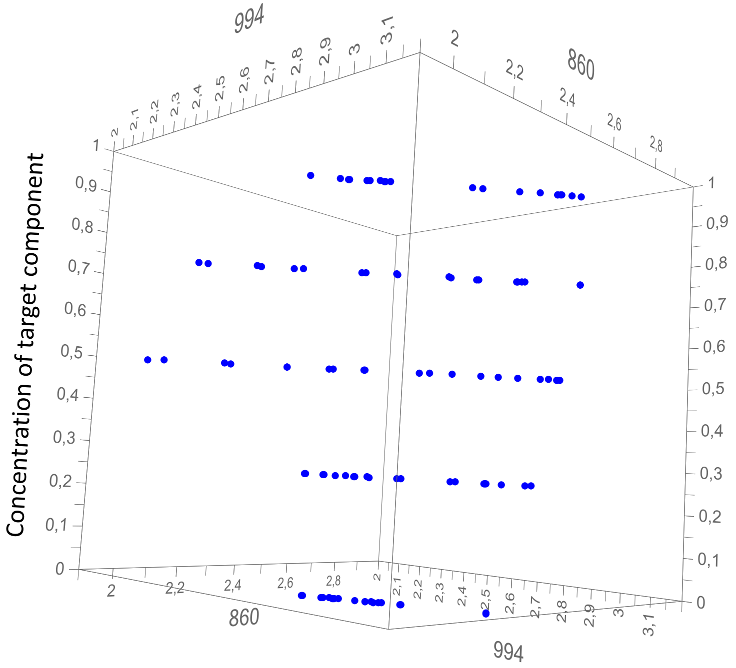
«Best» channel: useless calibration

Playing the piano with only one finger

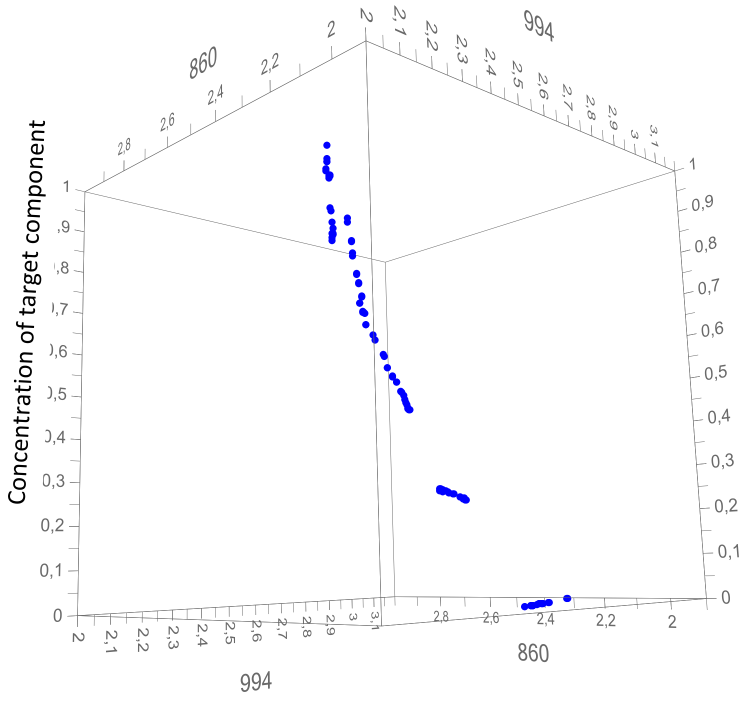




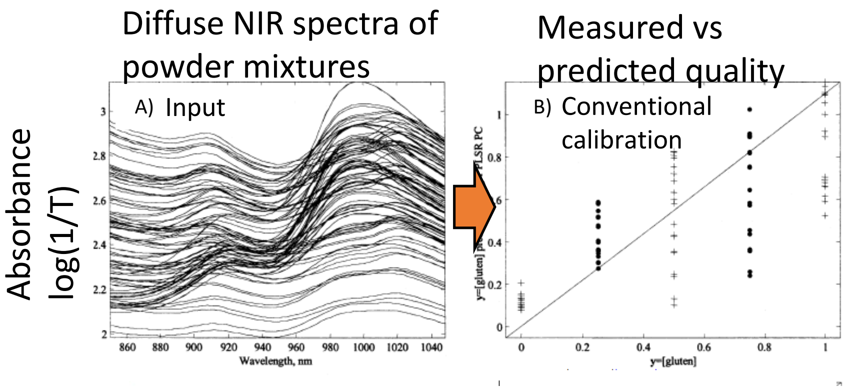
Playing the same piano with two fingers



Playing the same piano with two fingers

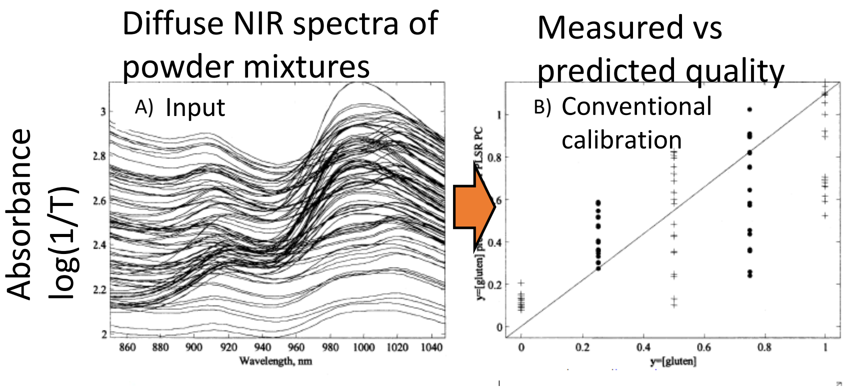


Modern high-speed multichannel instruments: Always record a whole *spectrum* of properties at e.g. many frequencies of light or sound.

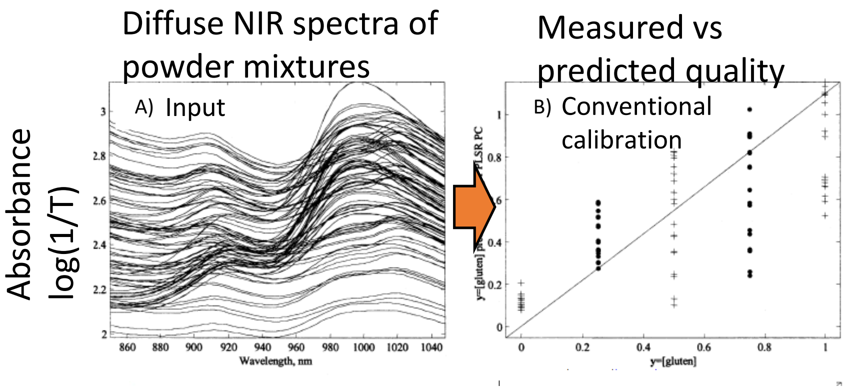


Modern high-speed multichannel instruments:
Always record a whole *spectrum* of properties at e.g. many frequencies of light or sound.

Measure more than you need, for math is cheaper than physics



Modern high-speed multichannel instruments: Always record a whole *spectrum* of properties at e.g. many frequencies of light or sound.



Measure more than you need, for math is cheaper than physics

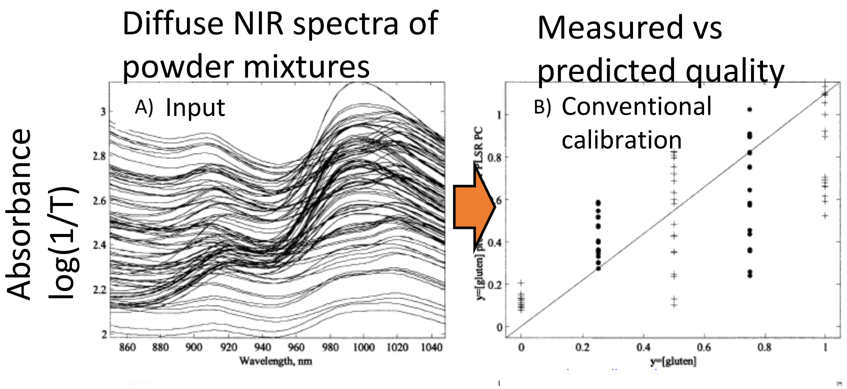
Play your instrument with more than one finger at a time

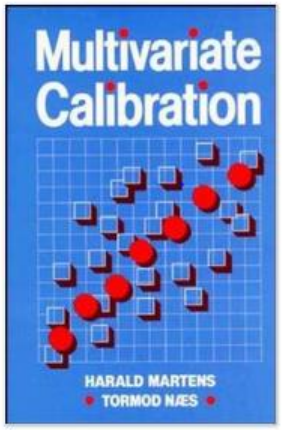
Modern high-speed multichannel instruments: Always record a whole *spectrum* of properties at e.g. many frequencies of light or sound.

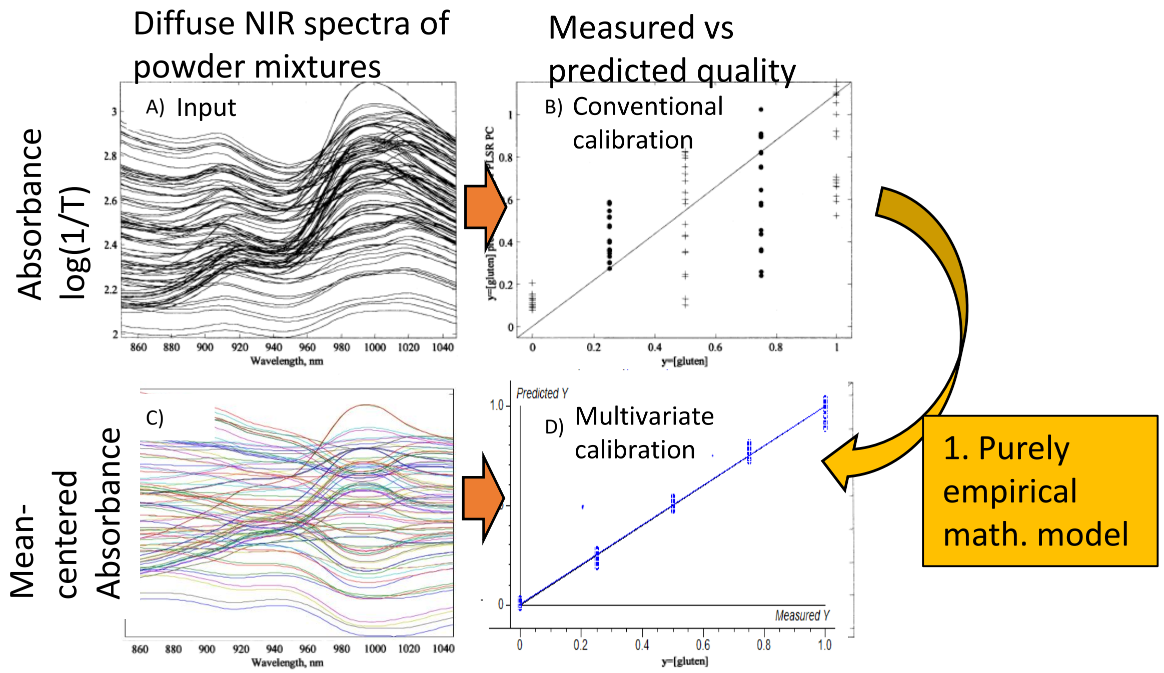
Measure more than you need, for math is cheaper than physics

Play your instrument with more than one finger at a time

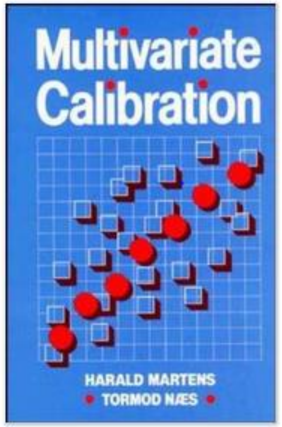
Don’t use a black box if you can avoid it

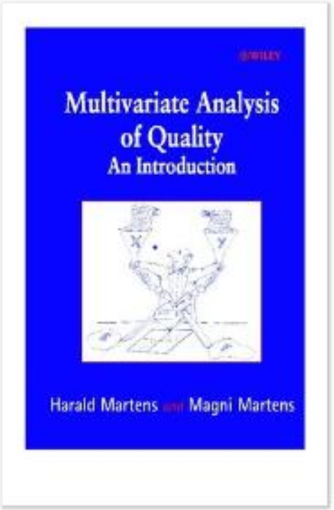


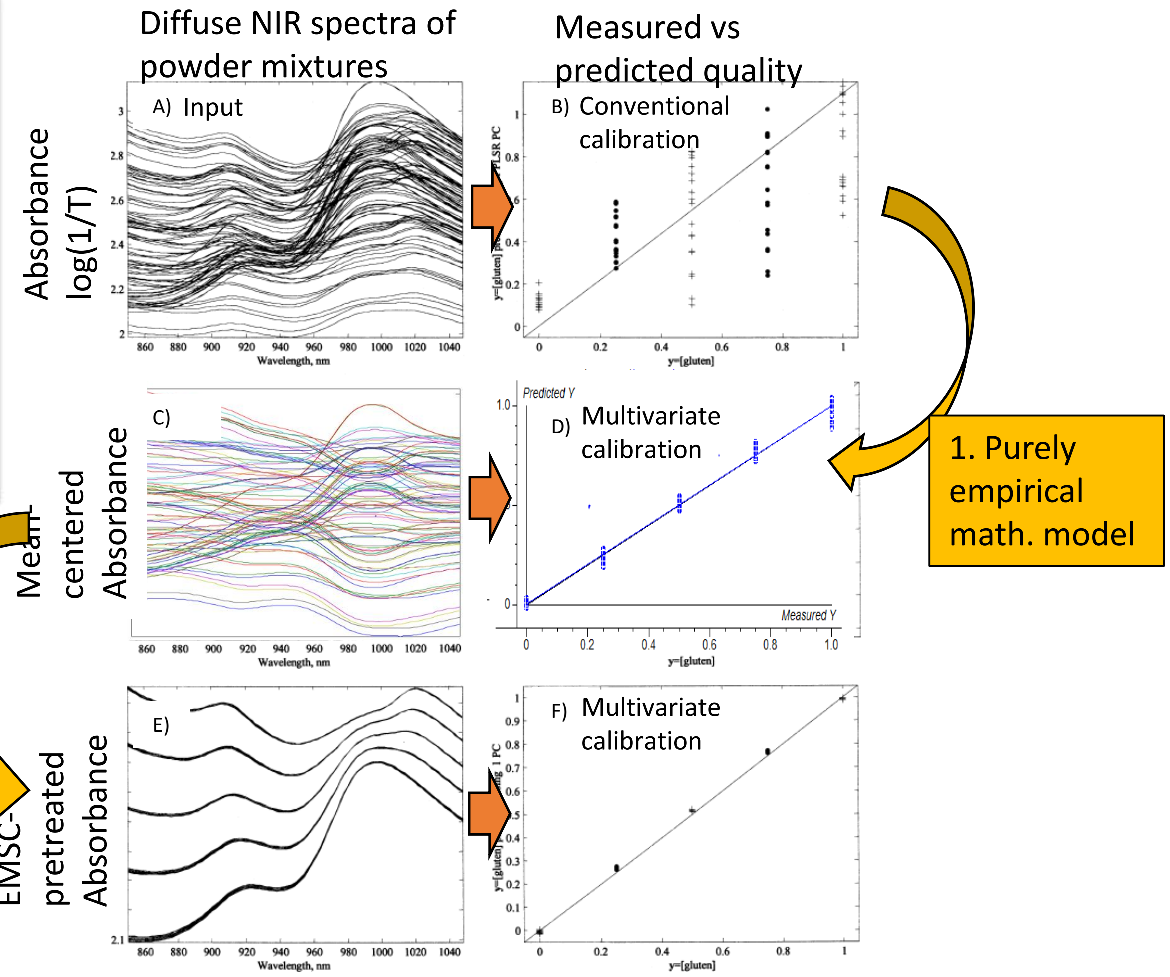




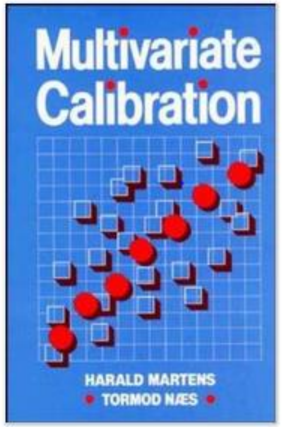
What is CHEMOMETRICS ?
A particular science culture & tool box for «soft multivariate data modelling»: interpretable machine learning

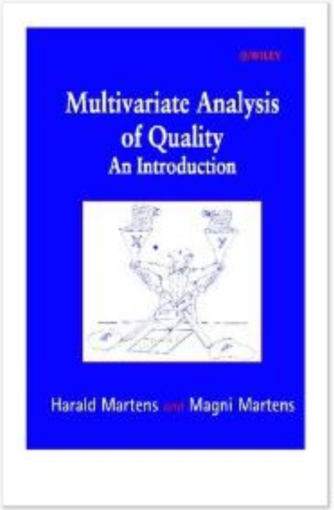


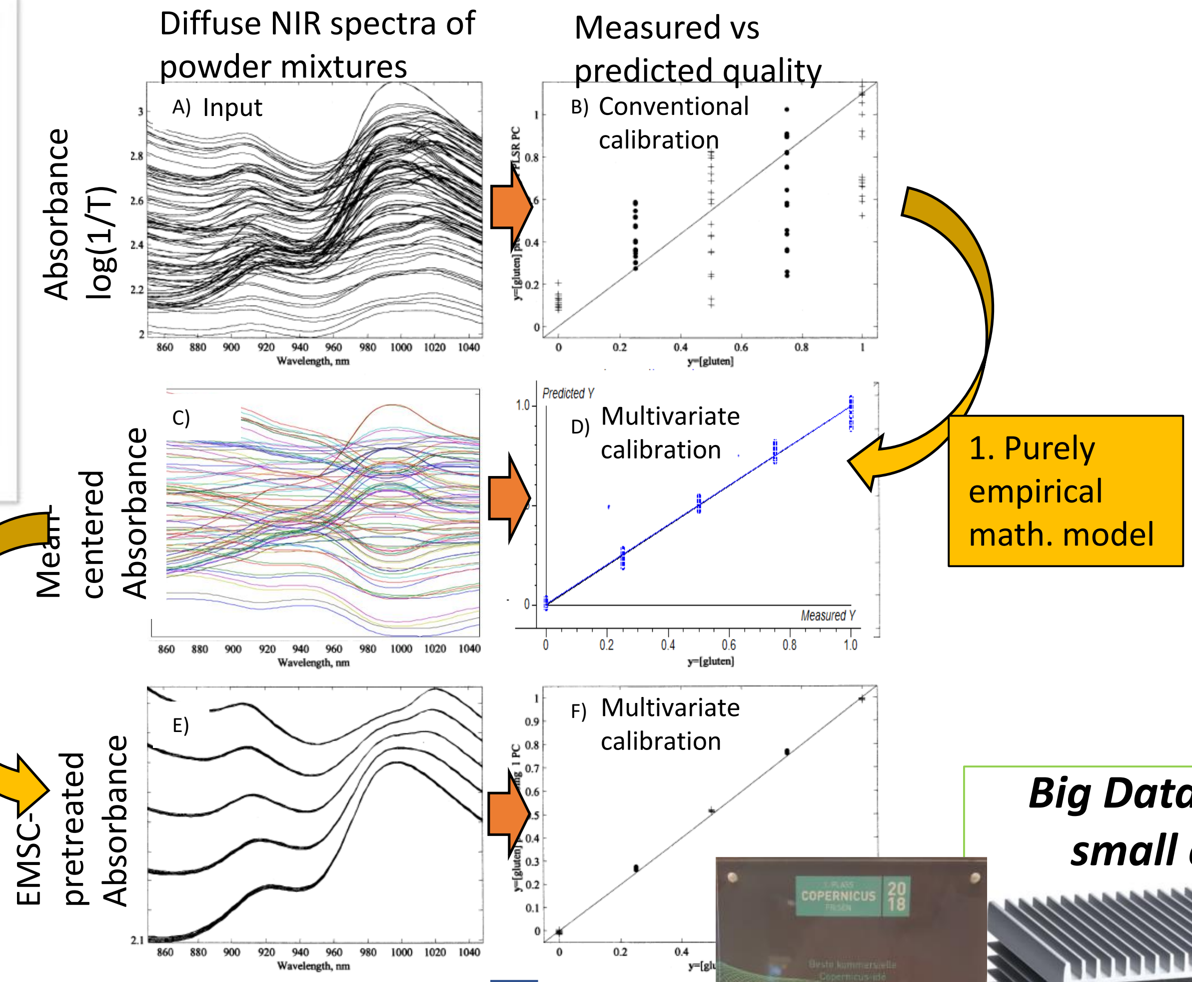




What is CHEMOMETRICS ?
A particular science culture & tool box for «soft multivariate data modelling»: interpretable machine learning







2. Semi-causal math. model

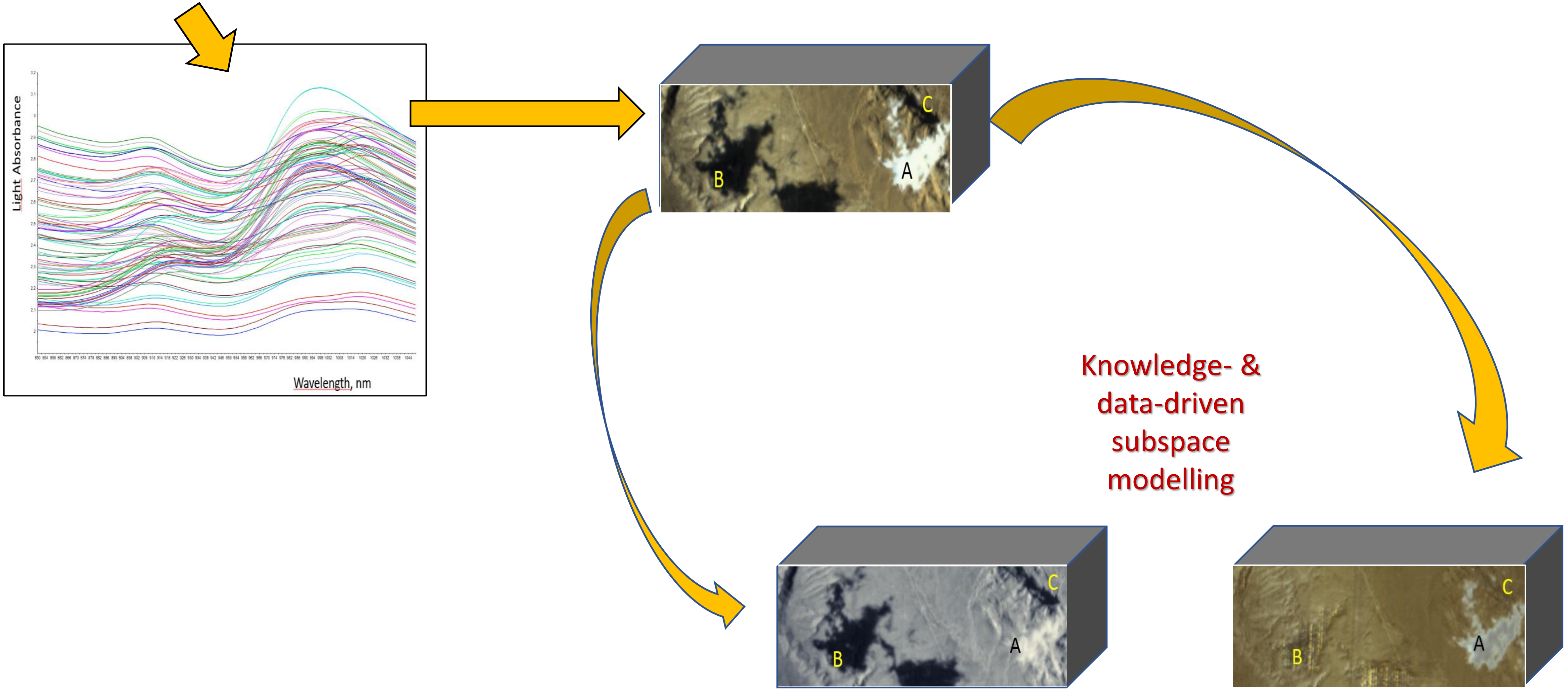
What is CHEMOMETRICS ?
A particular science culture & tool box for «soft multivariate data modelling»: interpretable machine learning

3. Big Data Cybernetics

Big Data Streams in small computer

idletechs

Success story: Multivariate calibration of high-speed non-selective instruments



idletechs

*Deshadowing via **Informative Converse** model:* Separating illumination / ground properties in HSI

Data Model

In order to apply the described method, it is useful to first define a model for the measured data in each pixel:

$$\mathbf{Y} = \mathbf{C}\,\mathbf{S}^{\mathsf{T}} + \mathbf{D}\,\mathbf{Z}^{\mathsf{T}} + \mathbf{F}$$

With:

$$\begin{cases} \mathbf{Z} = \mathbf{A}^{\mathsf{T}}\mathbf{S}^{\mathsf{T}} + \mathbf{Z}^{\mathsf{T}}_{\perp\mathbf{S}} \\ \mathbf{C} = \mathbf{D}\mathbf{B} + \mathbf{C}_{\perp\mathbf{D}} \end{cases}$$

Where:

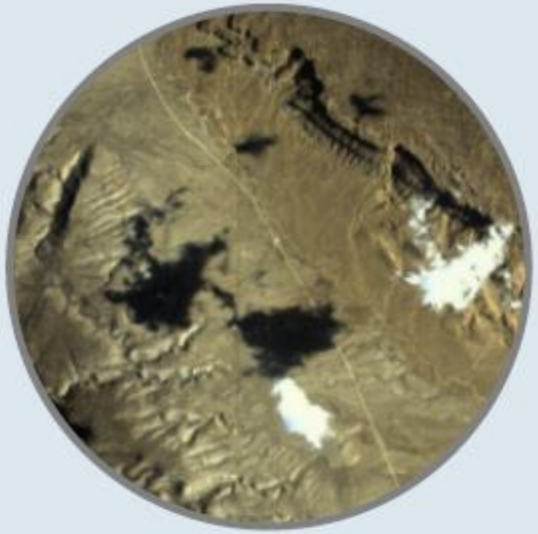
- $\mathbf{Y}$ is absorbance data obtained by sensor
- $\mathbf{CS}^{\mathsf{T}}$ is the contribution of partially known effects (e.g. illumination variations, "shadows")
- $\mathbf{DZ}^{\mathsf{T}}$ is the contribution of unknown effects (e.g. ground geology/biology variations)
- $\mathbf{F}$ is measurement noise, assumed normal
- $\mathbf{A}$ captures the non-orthogonality between $\mathbf{Z}$ and $\mathbf{S}$
- $\mathbf{B}$ captures the non-orthogonality between $\mathbf{C}$ and $\mathbf{D}$

Assumptions

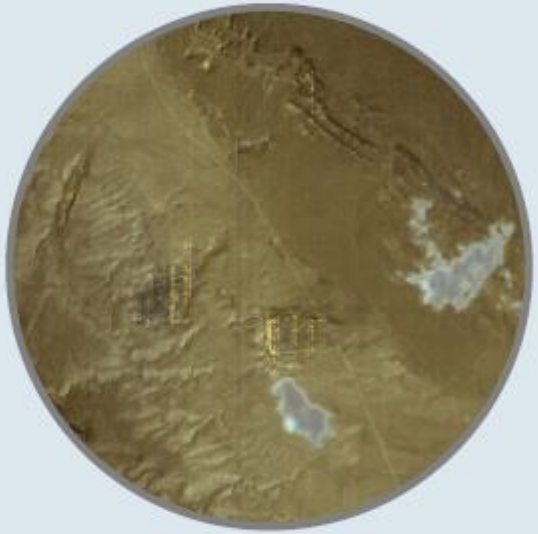
- Illumination effects are multiplicative in Reflectance
- $\mathbf{Y}$ data is given in Absorbance $(-\log_{10}(R))$
- Spectra of different illumination sources $\mathbf{S}$ are known

Earth Observing-1

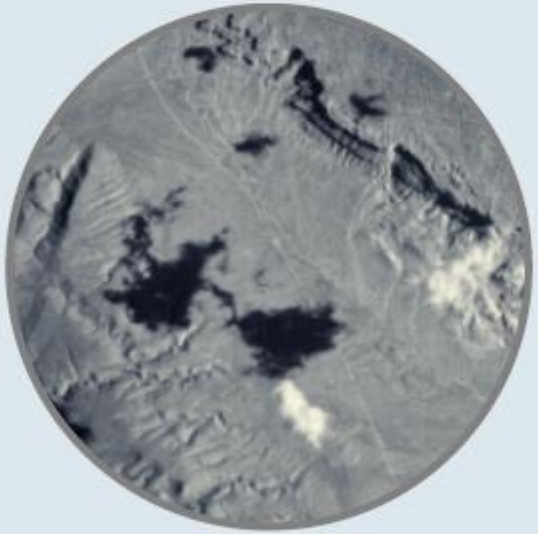
Data from the Hyperion instrument onboard the EO-1 Satellite. Data contains 200 bands in the VIS-NIR region.



Input data Y, in RGB



Deshadowed image, in RGB



"Shadow" (illumination change) image, $\widehat{C}S^{\mathsf{T}}$, in RGB

J.F. Fortuna and H. Martens, *J. Spectral Imaging* **6**, a2 (2017) Multivariate data modelling for de-shadowing of airborne hyperspectral imaging

Fast decomposition of hyperspectral images

Input HSI image, in RGB

How much trees and grass?
Healthy trees?

Hybrid multivariate modelling of causalities in HSI spectra:

Example from arial surveillance of biological resources by NEO HYSPEX camera

Lorenzo Fuzini & Harald Martens, Idletechs 2019
Acknowledgements: Terratec AS, Oslo

Fast decomposition of hyperspectral images

Input HSI image, in RGB

How much trees and grass?
Healthy trees?

Hybrid multivariate modelling of causalities in HSI spectra:

Example from arial surveillance of biological resources by NEO HYSPEX camera

idletechs

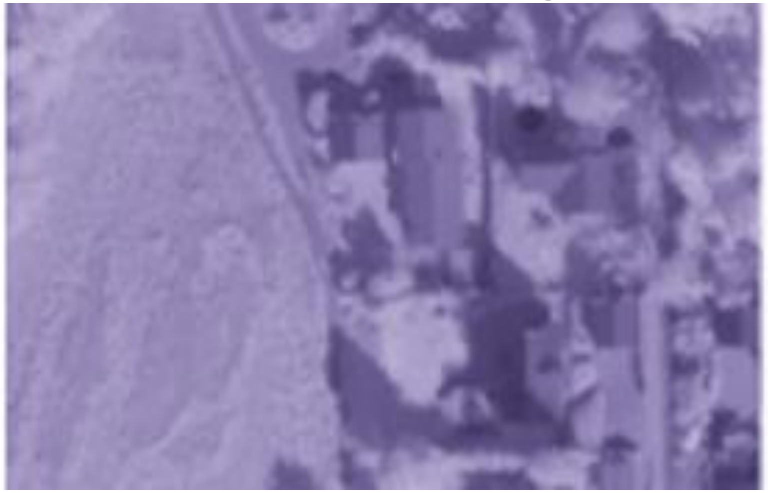
- Chemical variations
- Physical variations
- Instrument variations
- Measurement conditions
- Outliers
- Measurement noise

Lorenzo Fuzini & Harald Martens, Idletechs 2019
Acknowledgements: Terratec AS, Oslo

Fast decomposition of hyperspectral images

Input HSI image, in RGB

Shadow image



How much trees and grass?
Healthy trees?

Hybrid multivariate modelling of causalities in HSI spectra:

Example from arial surveillance of biological resources by NEO HYSPEX camera

idletechs

- **Chemical variations**
- **Physical variations**
- **Instrument variations**
- **Measurement conditions**
- **Outliers**
- **Measurement noise**

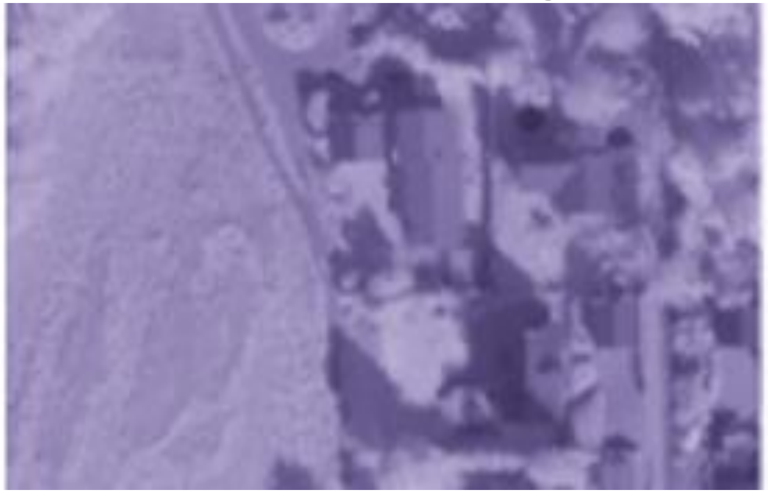
Lorenzo Fuzini & Harald Martens, Idletechs 2019
Acknowledgements: Terratec AS, Oslo

Fast decomposition of hyperspectral images

Input HSI image, in RGB

Deshadowed image

Shadow image



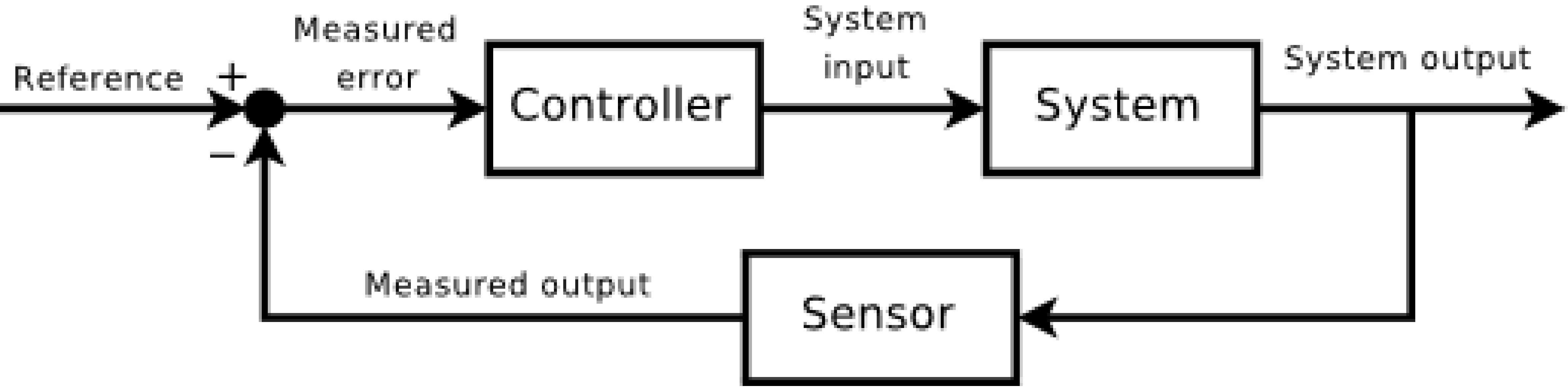
Hybrid multivariate modelling of causalities in HSI spectra:

Example from arial surveillance of biological resources by NEO HYSPEX camera

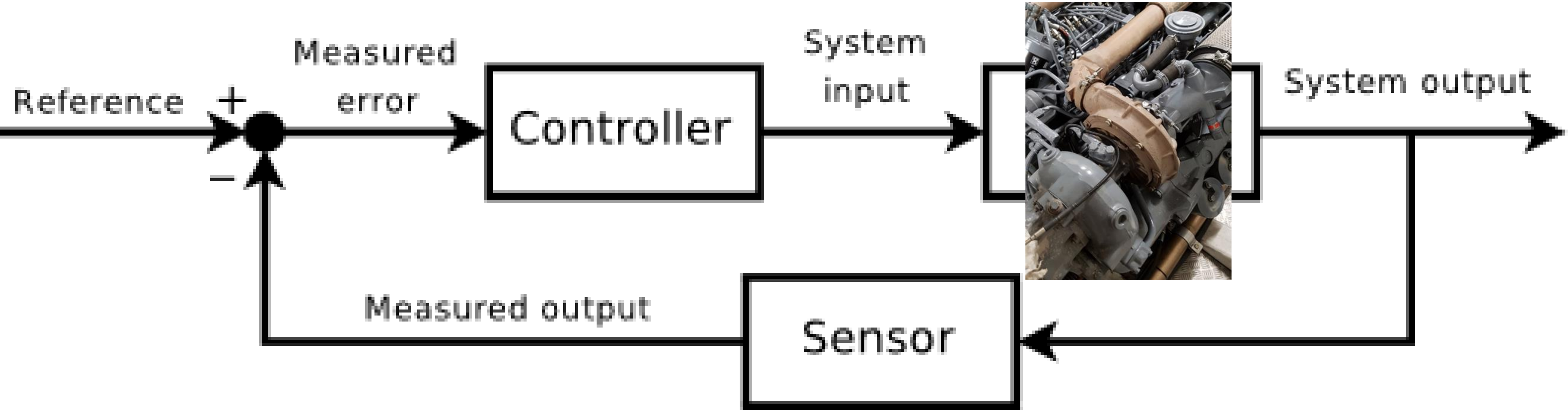
idletechs

- Chemical variations
- Physical variations
- Instrument variations
- Measurement conditions
- Outliers
- Measurement noise

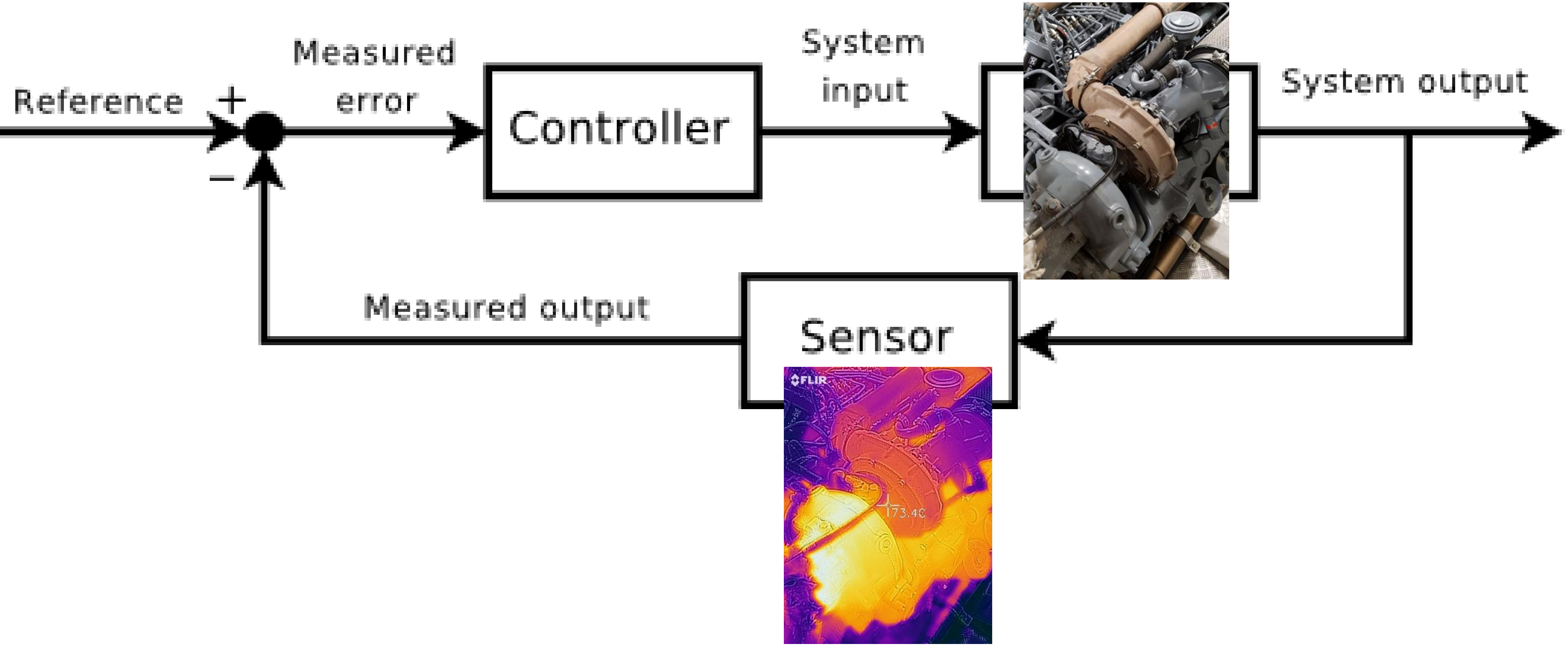
Lorenzo Fuzini & Harald Martens, Idletechs 2019
Acknowledgements: Terratec AS, Oslo



**Big Data Cybernetics for e.g.
Thermal machinery monitoring**

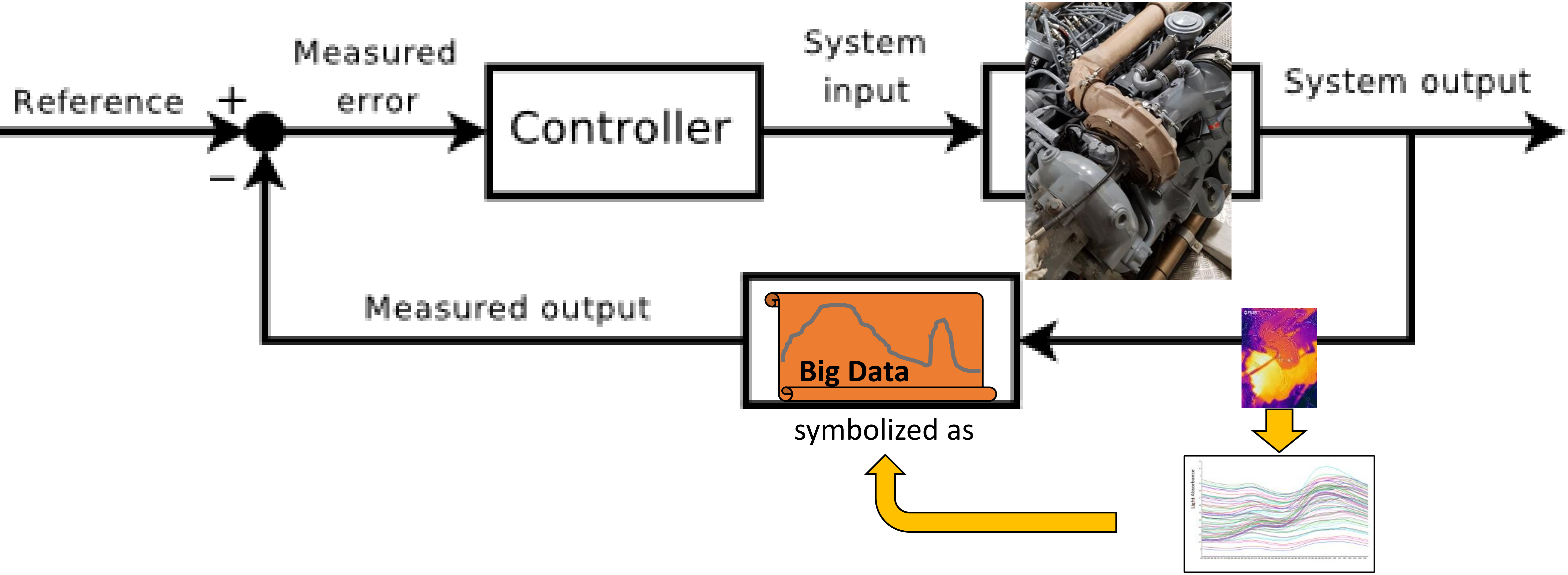


Big Data Cybernetics for e.g.
Thermal machinery monitoring



Big Data Cybernetics for e.g.
Thermal machinery monitoring

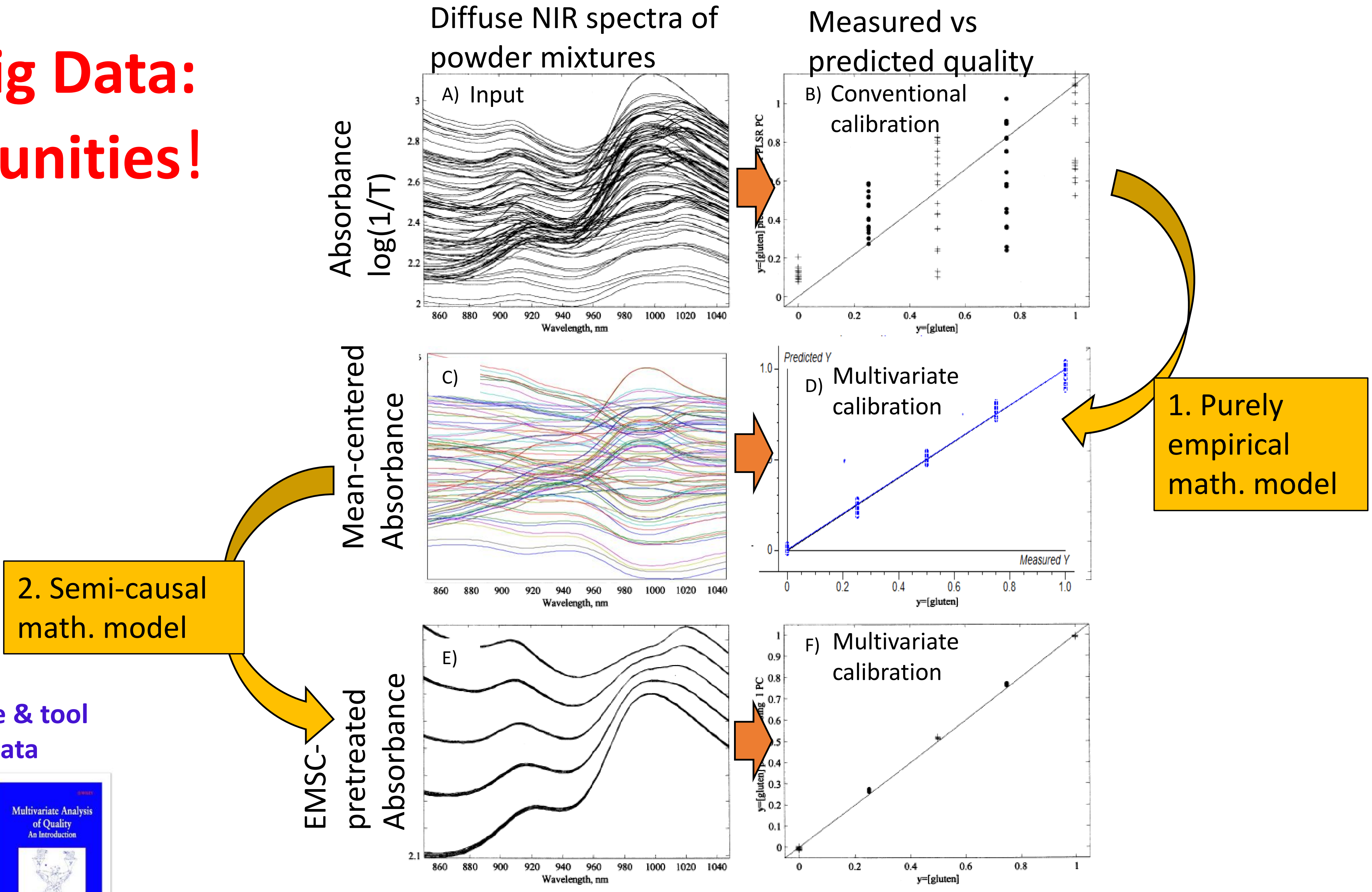
idletechs



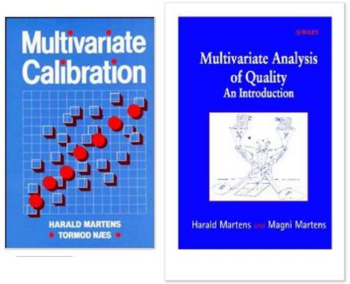
Big Data Cybernetics for e.g.
Thermal machinery monitoring

idletechs

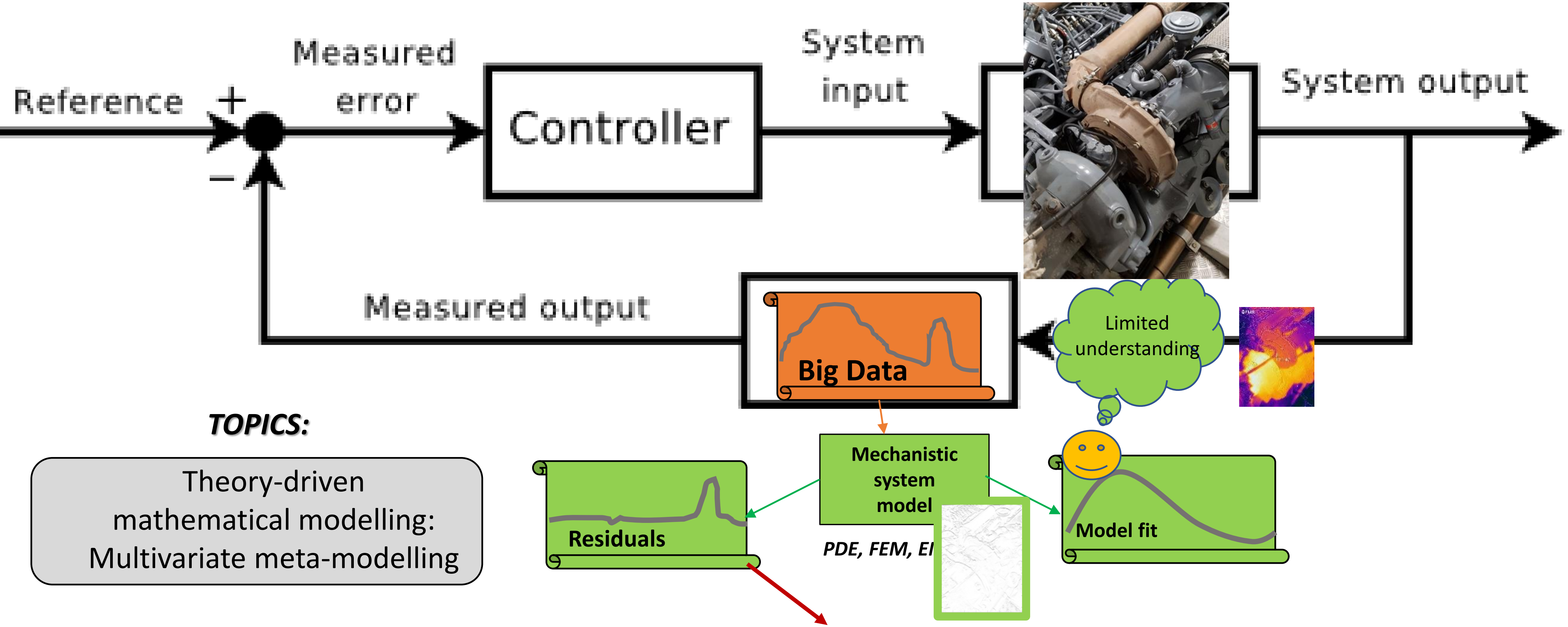
Industrial Big Data: new opportunities!



What is CHEMOMETRICS ?
A particular science culture & tool set for «soft multivariate data modelling»



idletechs

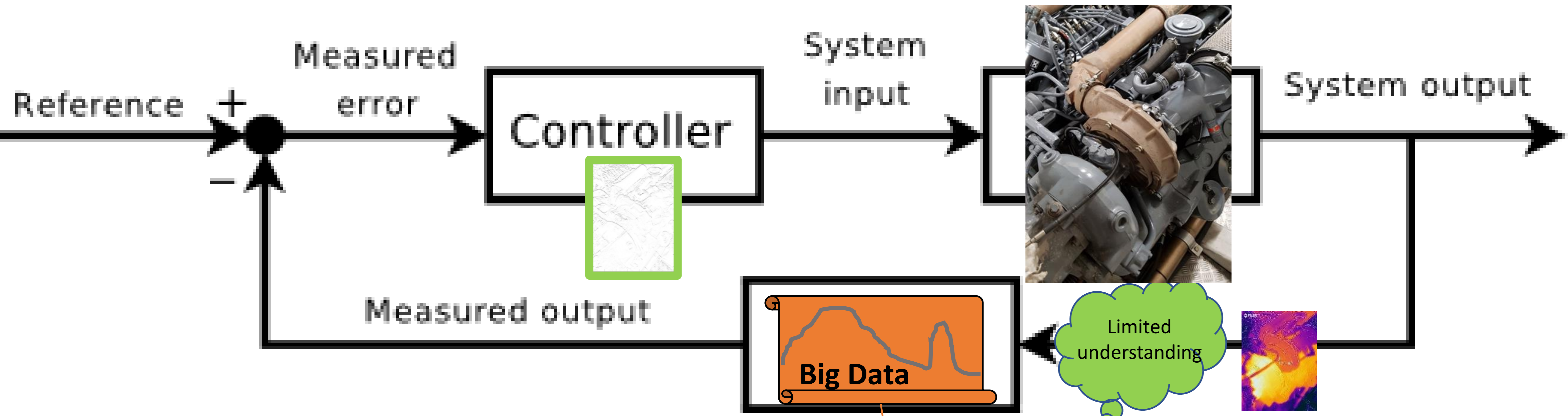


TOPICS:

Theory-driven
mathematical modelling:
Multivariate meta-modelling

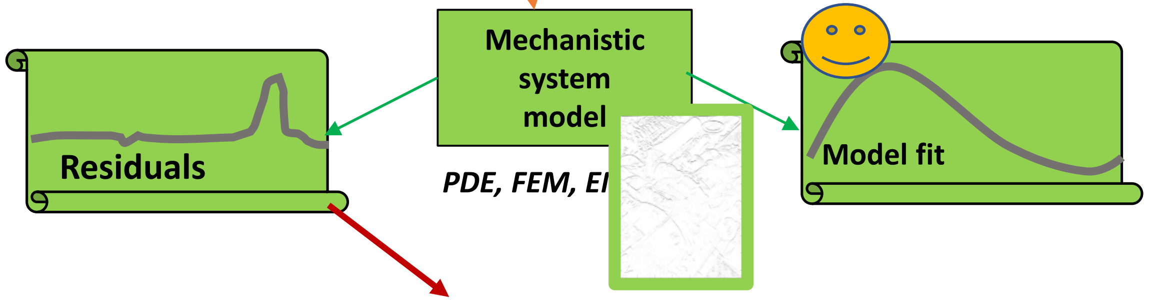
**Big Data Cybernetics for e.g.
Thermal machinery monitoring**

idletechs



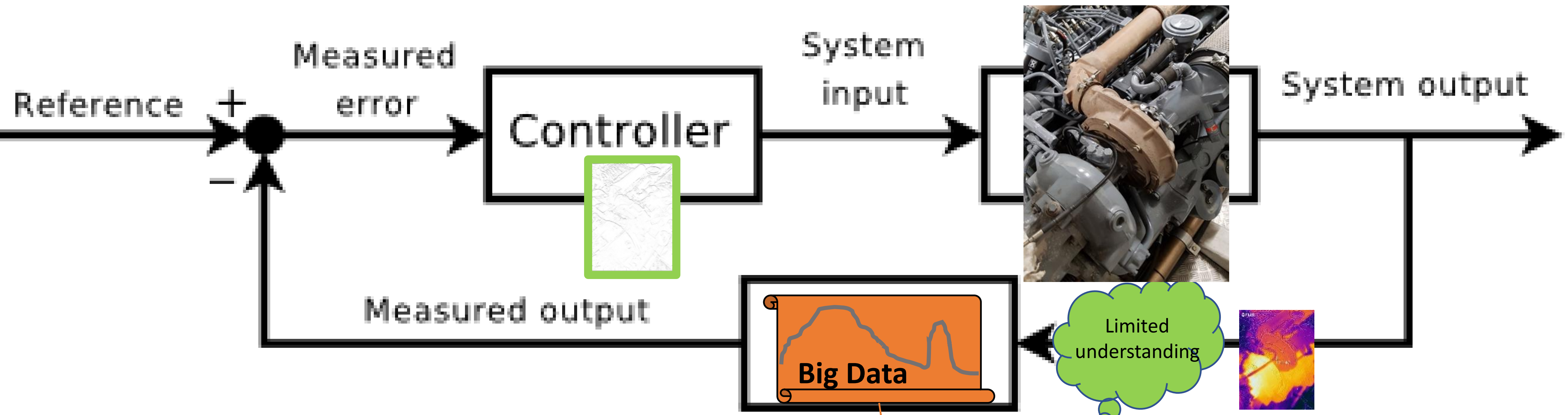
TOPICS:

Theory-driven
mathematical modelling:
Multivariate meta-modelling



Big Data Cybernetics for e.g.
Thermal machinery monitoring

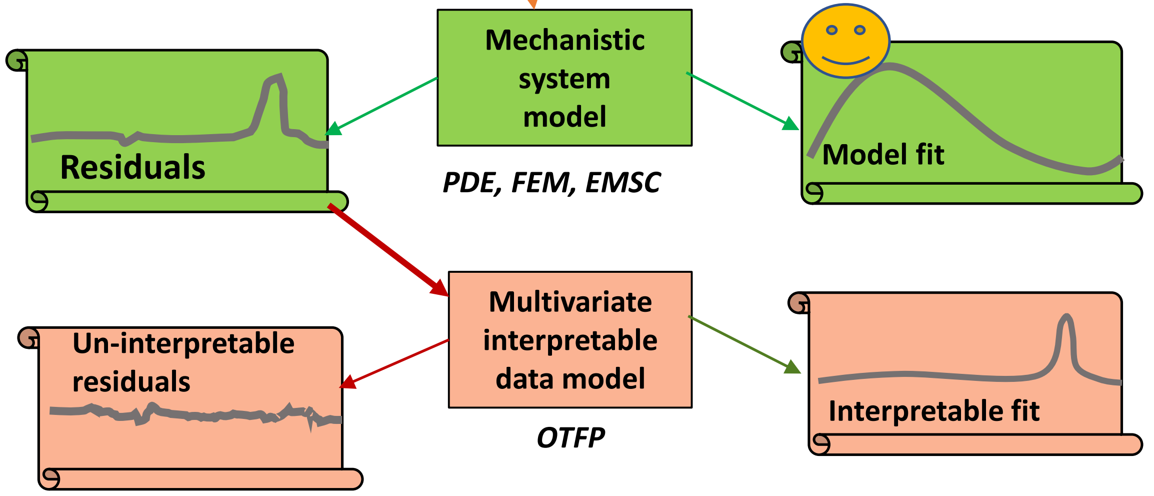
idletechs

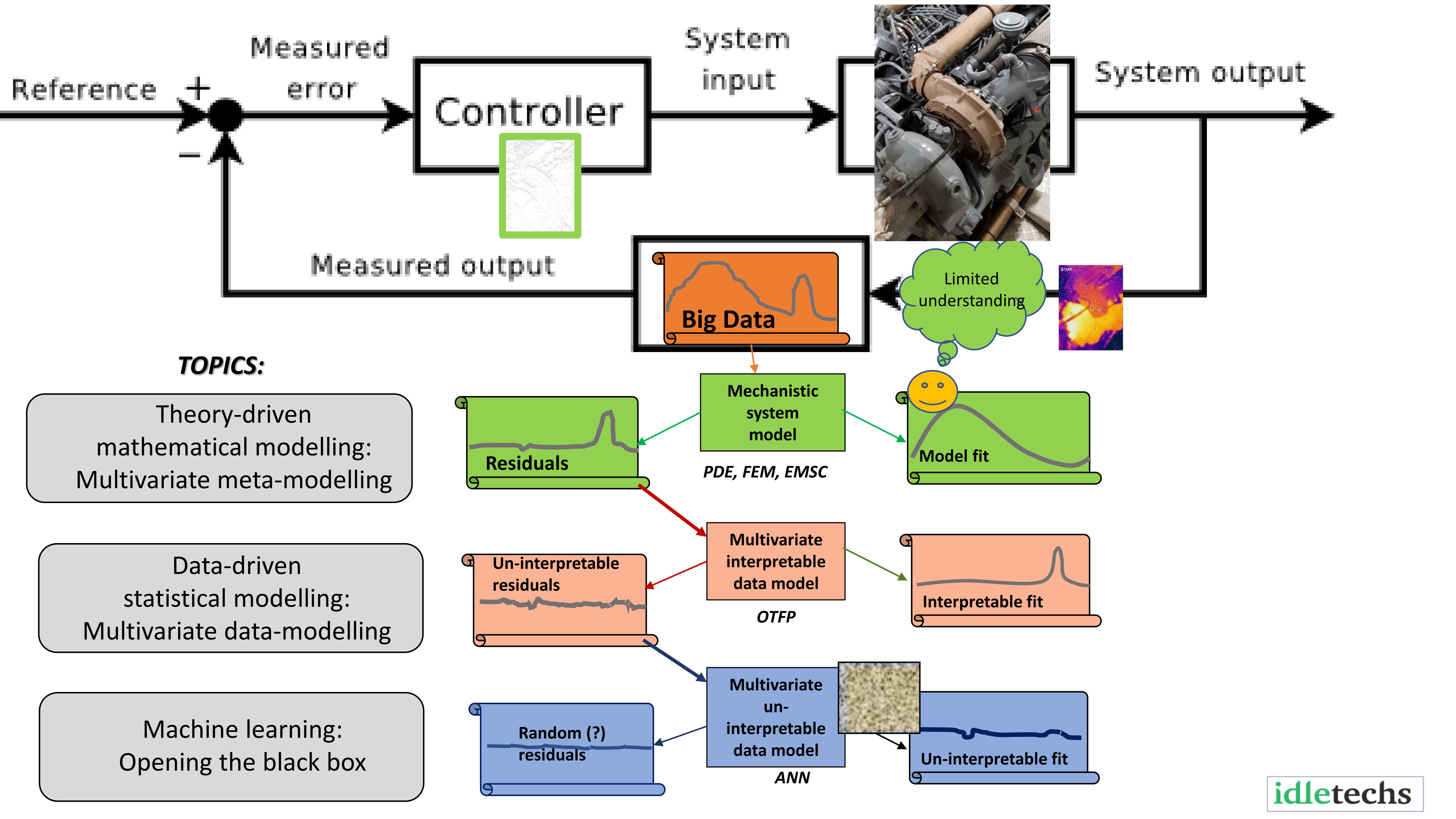


TOPICS:

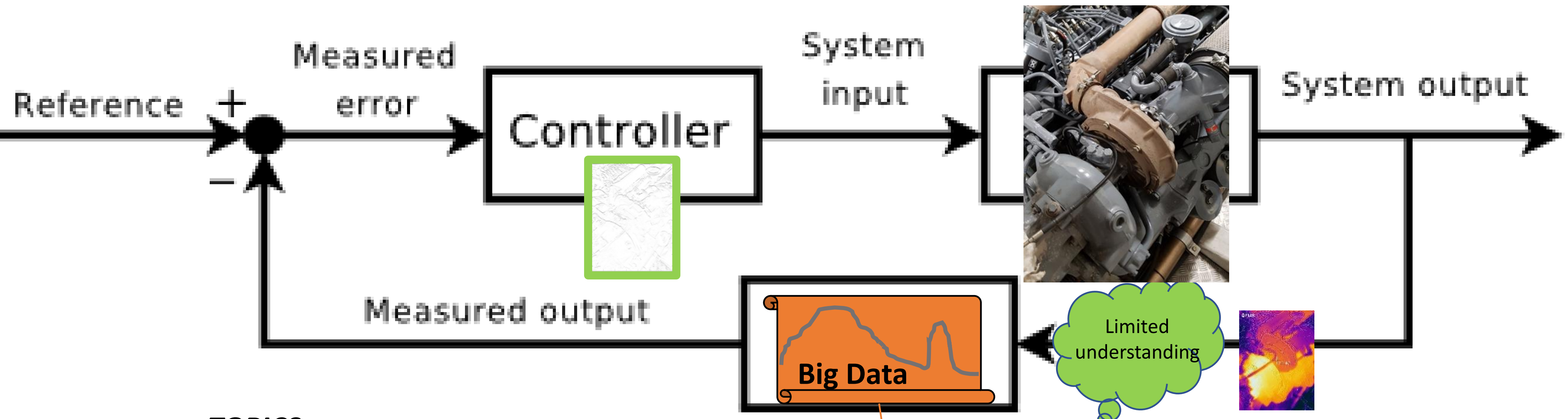
Theory-driven mathematical modelling: Multivariate meta-modelling

Data-driven statistical modelling: Multivariate data-modelling





Reference
+
−
Measured
error
Controller
System
input
System output
Measured output
Big Data
Limited
understanding
TOPICS:
Theory-driven
mathematical modelling:
Multivariate meta-modelling
Data-driven
statistical modelling:
Multivariate data-modelling
Machine learning:
Opening the black box
Residuals
Mechanistic
system
model
PDE, FEM, EMSC
Model fit
Un-interpretable
residuals
Multivariate
interpretable
data model
OTFP
Interpretable fit
Random (?)
residuals
Multivariate
un-
interpretable
data model
ANN
Un-interpretable fit
idletechs

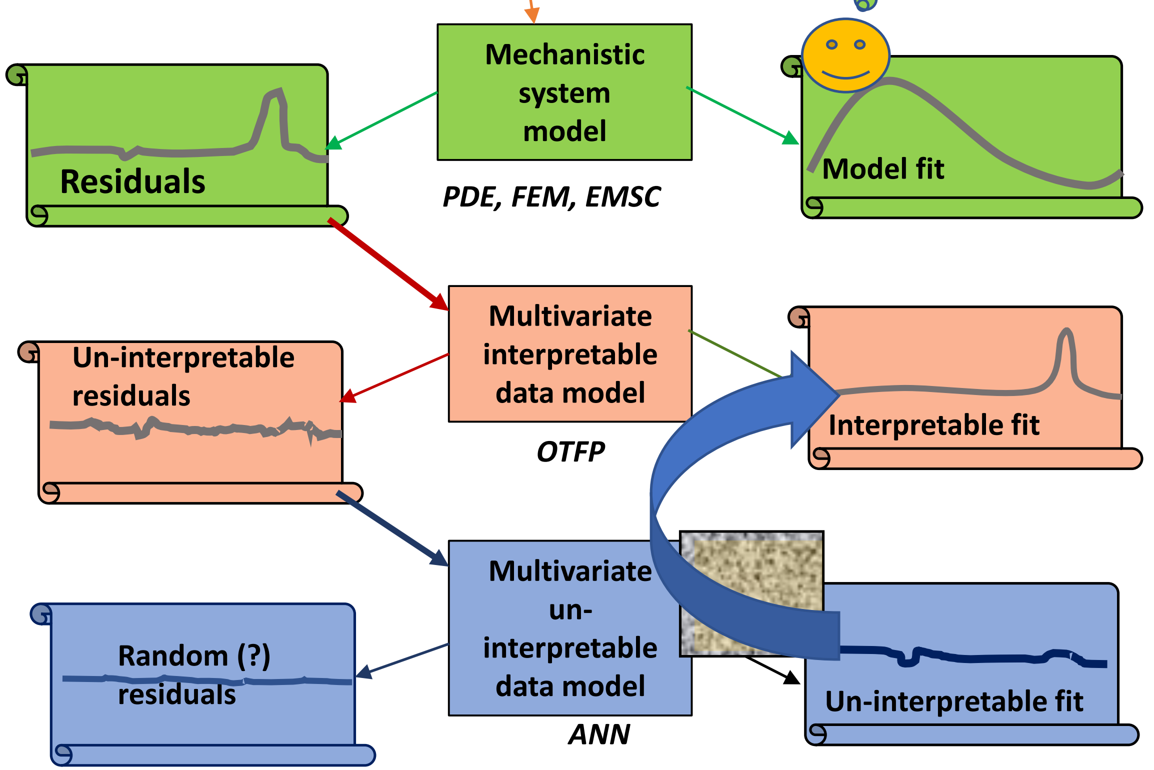


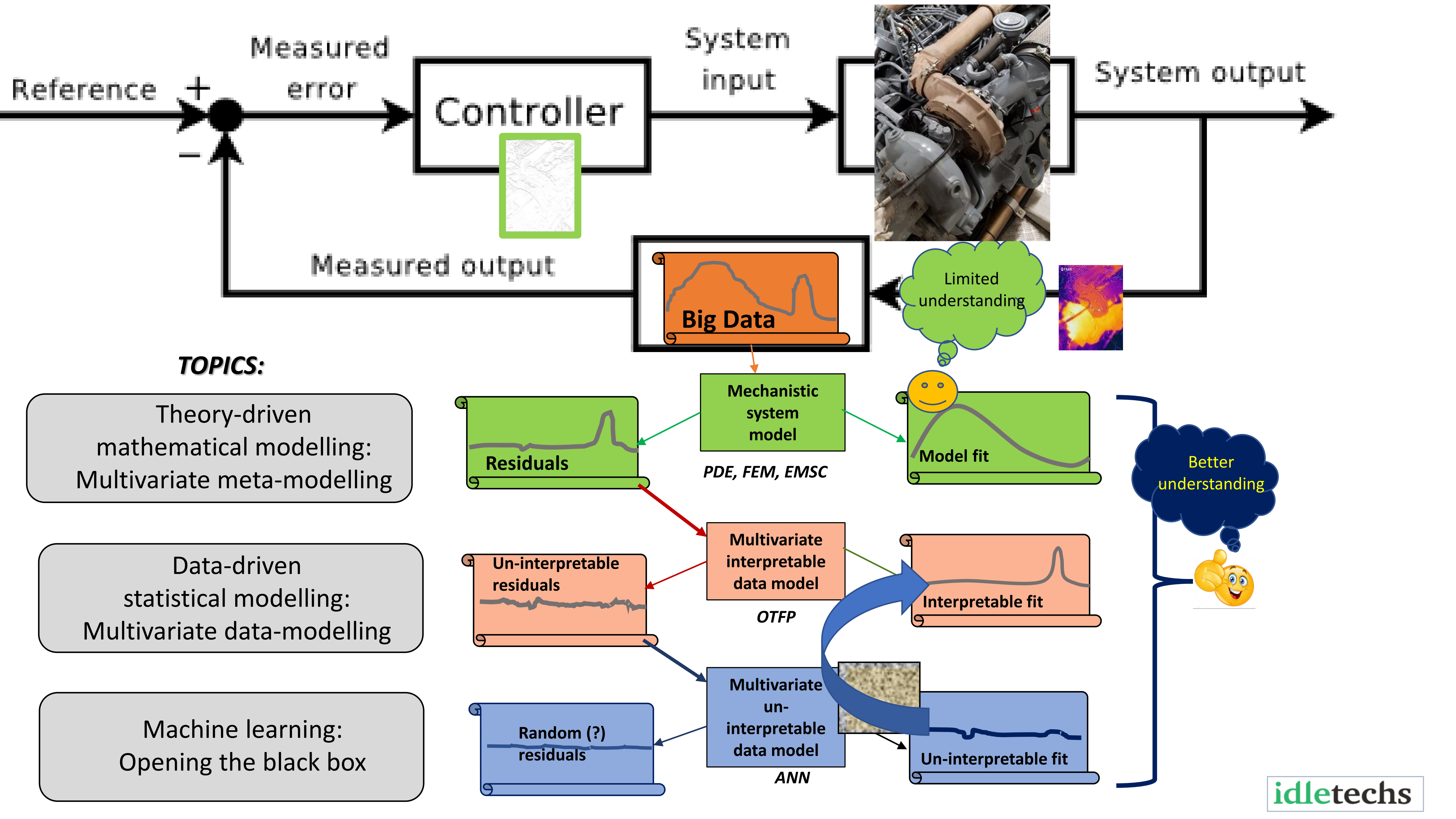
TOPICS:

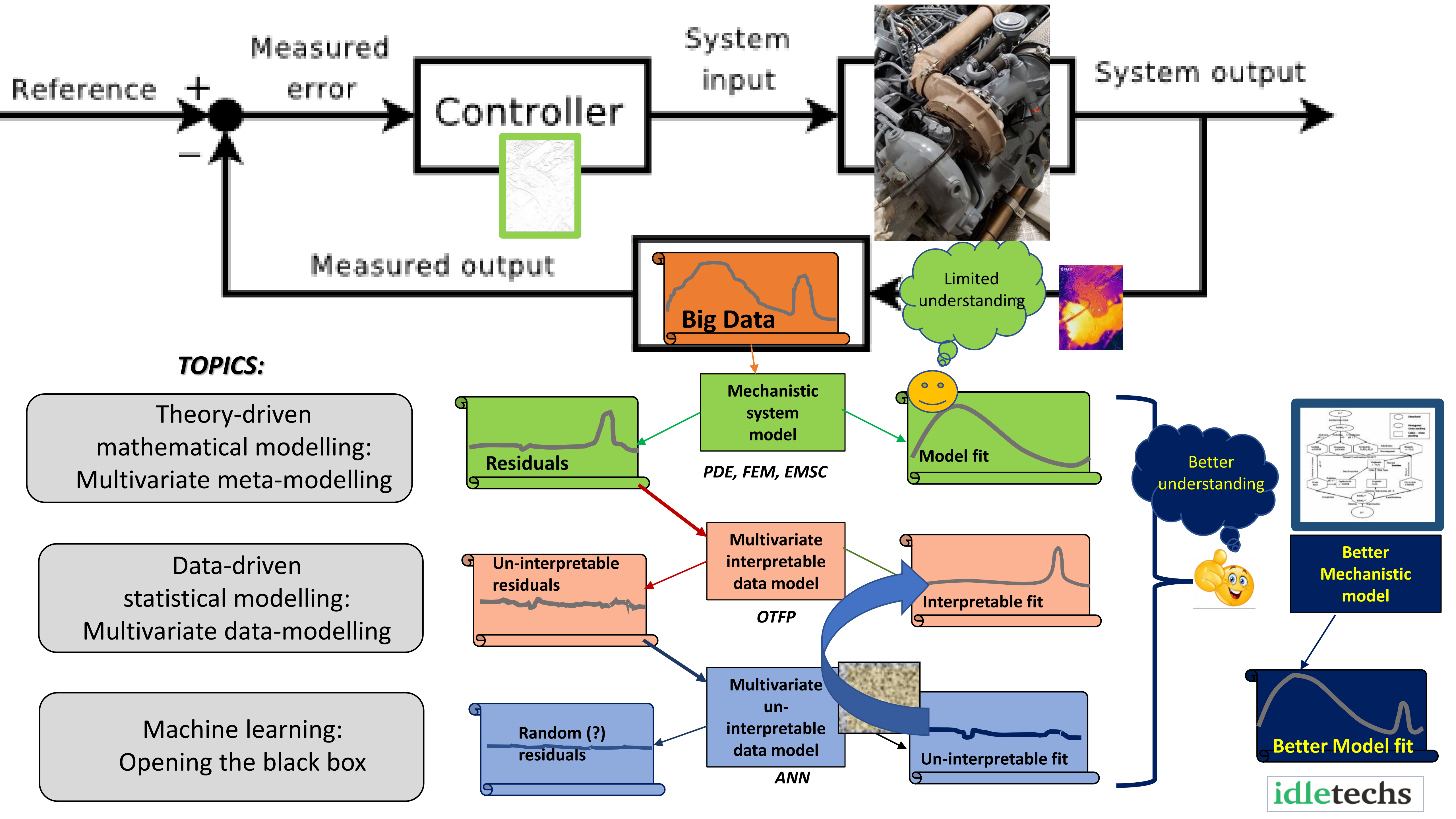
Theory-driven
mathematical modelling:
Multivariate meta-modelling

Data-driven
statistical modelling:
Multivariate data-modelling

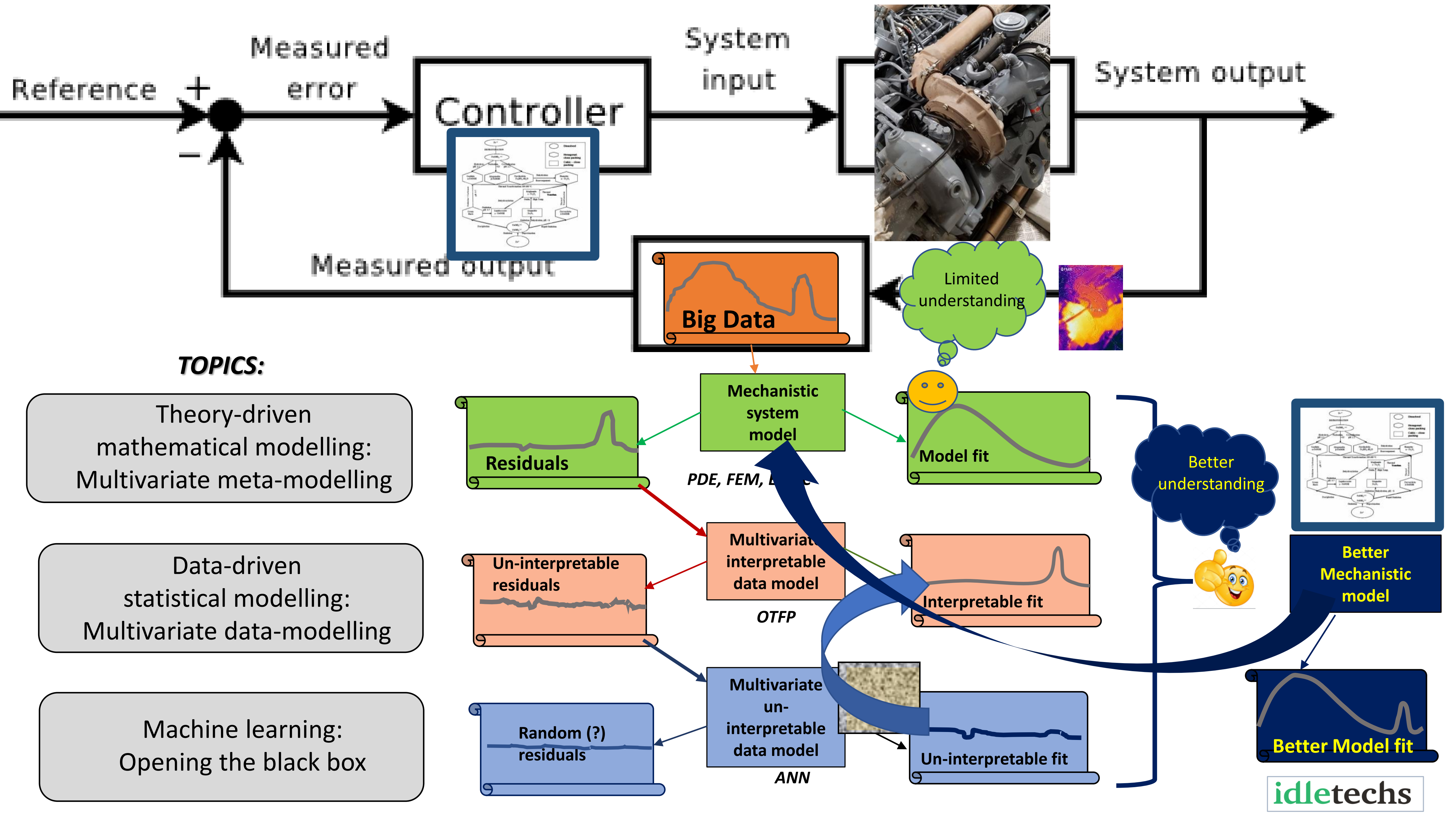
Machine learning:
Opening the black box

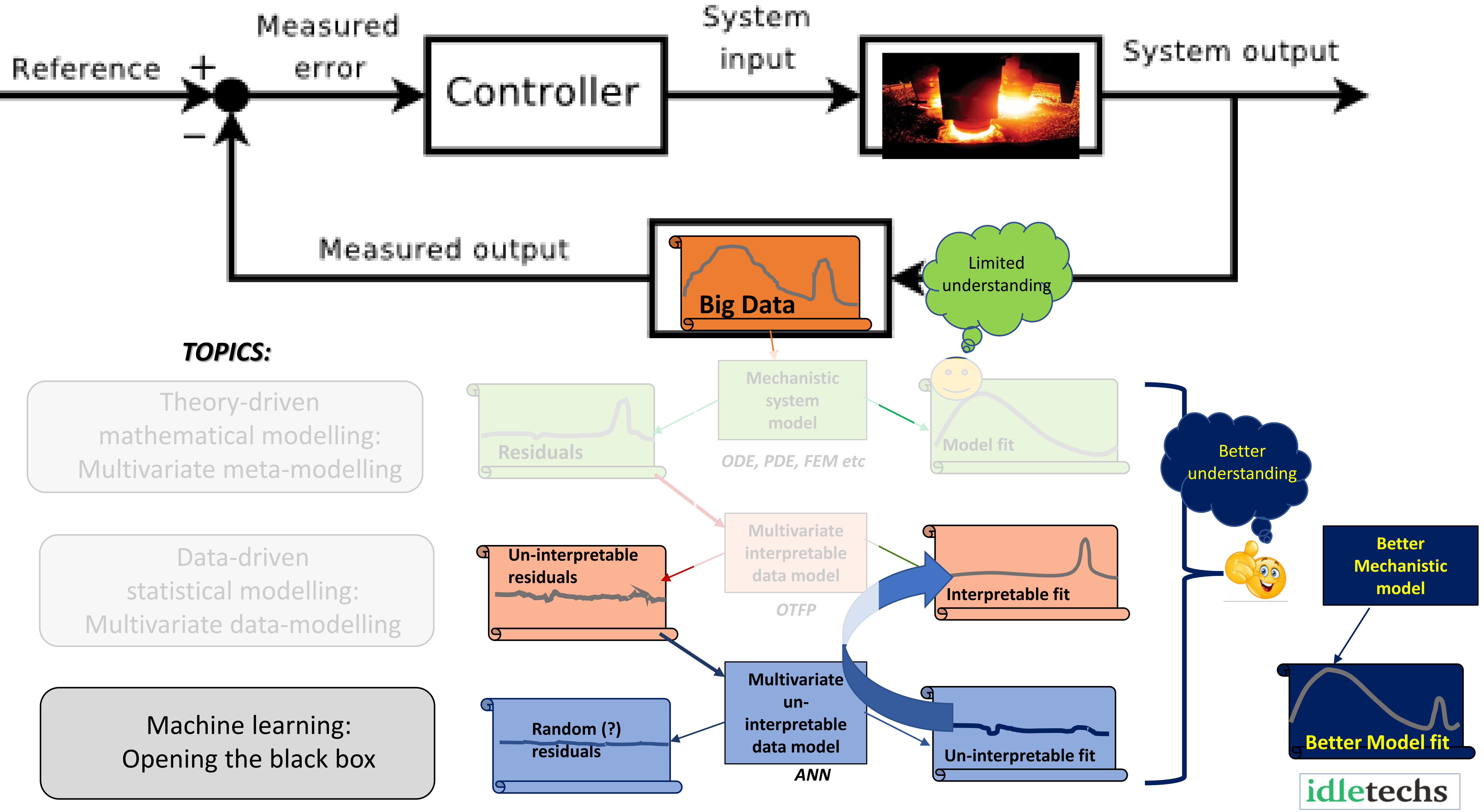



Reference
+
−
Measured
error
Controller
System
input
System output
Measured output
Big Data
Limited
understanding
TOPICS:
Theory-driven
mathematical modelling:
Multivariate meta-modelling
Data-driven
statistical modelling:
Multivariate data-modelling
Machine learning:
Opening the black box
Residuals
Mechanistic
system
model
PDE, FEM, EMSC
Model fit
Better
understanding
Un-interpretable
residuals
Multivariate
interpretable
data model
OTFP
Interpretable fit
Random (?)
residuals
Multivariate
un-
interpretable
data model
ANN
Un-interpretable fit
idletechs



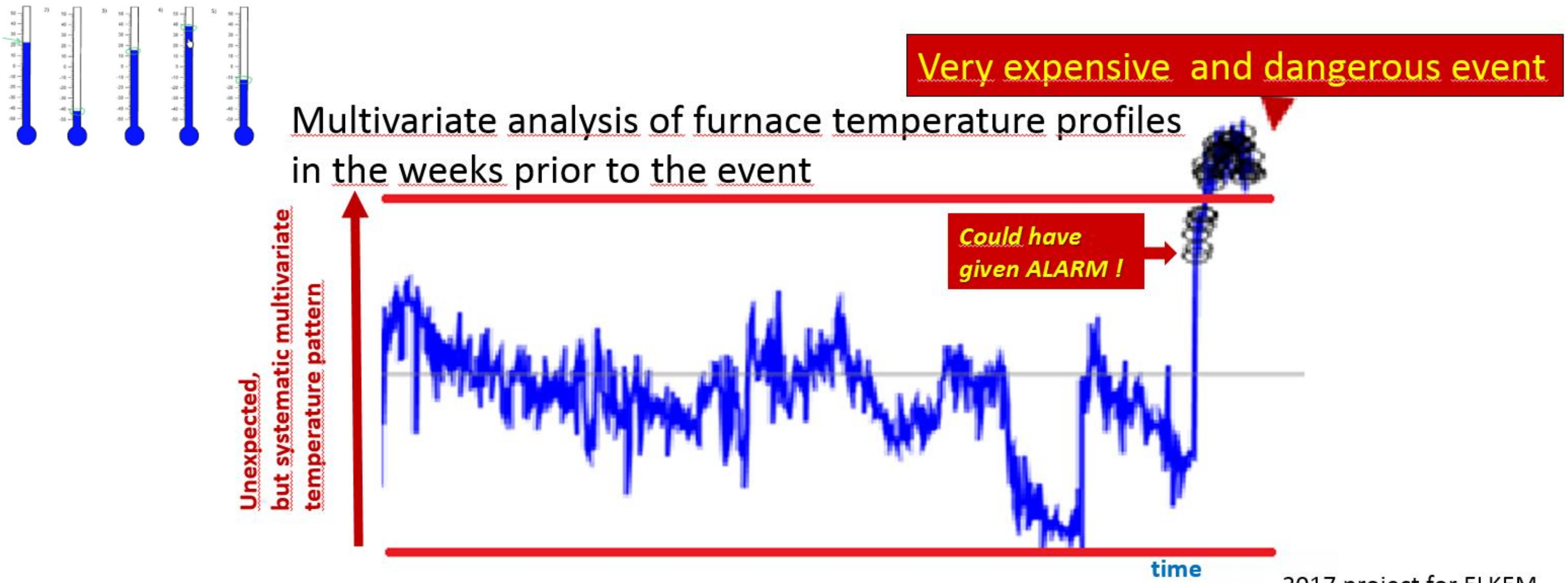
Reference
+
Measured error
Controller
System input
System output
–
Measured output
Big Data
Limited understanding
TOPICS:
Theory-driven mathematical modelling: Multivariate meta-modelling
Data-driven statistical modelling: Multivariate data-modelling
Machine learning: Opening the black box
Residuals
Mechanistic system model
PDE, FEM, EMSC
Model fit
Better understanding
Un-interpretable residuals
Multivariate interpretable data model
OTFP
Interpretable fit
Better Mechanistic model
Random (?) residuals
Multivariate un-interpretable data model
ANN
Un-interpretable fit
Better Model fit
idletechs


Reference
+
−
Measured error
Controller
System input
System output
Measured output
Big Data
Limited understanding
TOPICS:
Theory-driven mathematical modelling: Multivariate meta-modelling
Data-driven statistical modelling: Multivariate data-modelling
Machine learning: Opening the black box
Residuals
Mechanistic system model
Model fit
PDE, FEM,
Un-interpretable residuals
Multivariate interpretable data model
Interpretable fit
OTFP
Random (?) residuals
Multivariate un-interpretable data model
Un-interpretable fit
ANN
Better understanding
Better Mechanistic model
Better Model fit
idletechs


Reference
+
−
Measured
error
Controller
System
input
System output
Measured output
Big Data
Limited
understanding
TOPICS:
Theory-driven
mathematical modelling:
Multivariate meta-modelling
Data-driven
statistical modelling:
Multivariate data-modelling
Machine learning:
Opening the black box
Residuals
Mechanistic
system
model
ODE, PDE, FEM etc
Model fit
Better
understanding
Un-interpretable
residuals
Multivariate
interpretable
data model
OTFP
Interpretable fit
Better
Mechanistic
model
Random (?)
residuals
Multivariate
un-
interpretable
data model
ANN
Un-interpretable fit
Better Model fit
idletechs

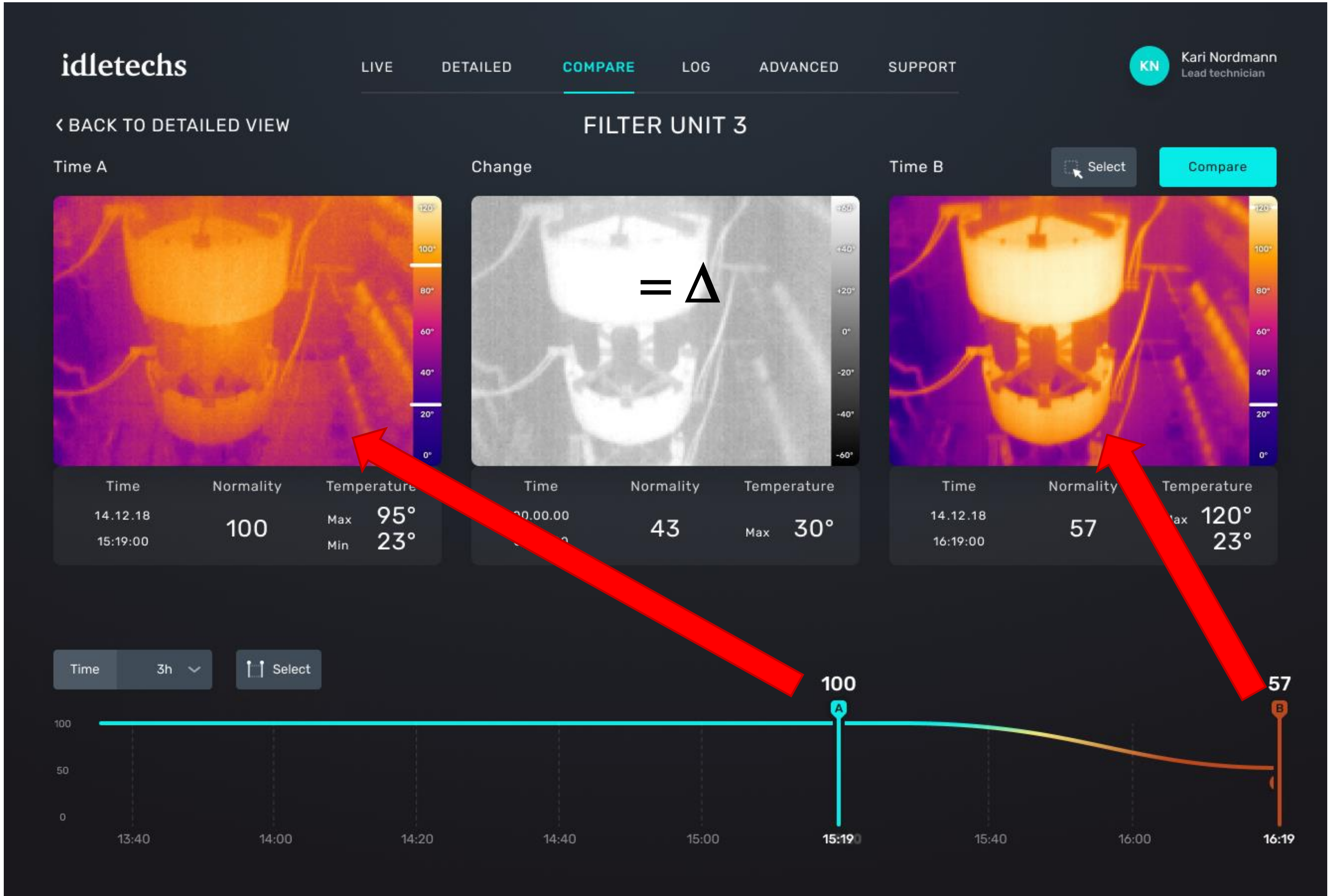
Point temperatures / Cooling water temperatures

Purpose	Monitor a number of point measurements in a process



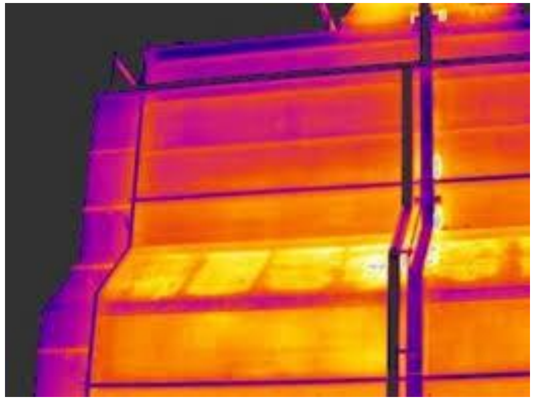
2017 project for ELKEM

Making quantitative data understandable for ordinary people: e.g. XAI for thermal analysis of high-power electrical equipment



idletechs

Thermal camera



Purpose	Monitor furnace temperatures, e.g. outer surface, electrode or tap hole area, to detect anomalities and unexpected trends

Thermal Camera

Related example:

Thermal Camera

Continuous monitoring of wood ovens:
Heating efficiency experiment at SINTEF 2018

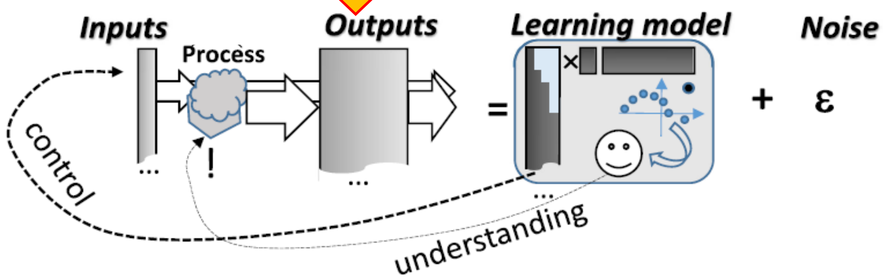
Demo example (non-commercial ☺): Home appliance equipment

idletechs

Experiment setup:

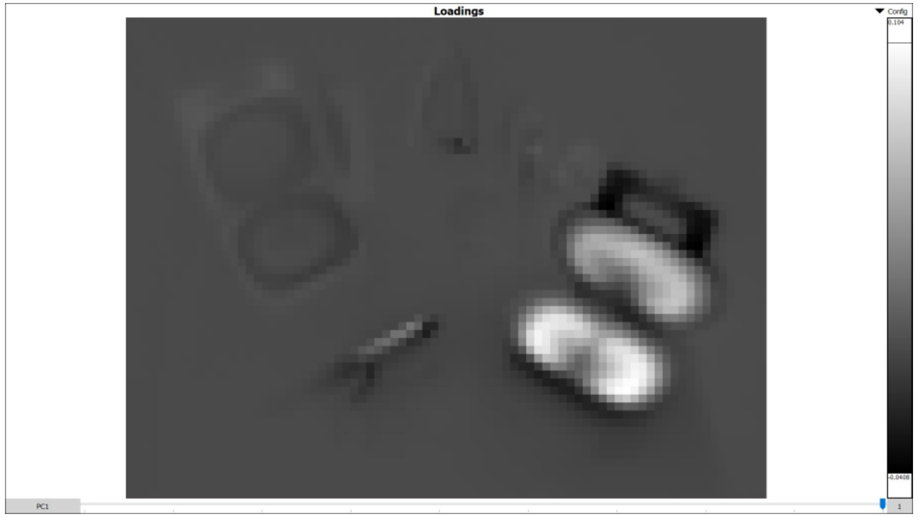
- Instruments:
  - waffle iron
  - burger grill
  - curling iron
  - clothes iron
- Disturbances:
  - water bottle
  - tea cup
  - hair dryer
  - human interf
  - timers

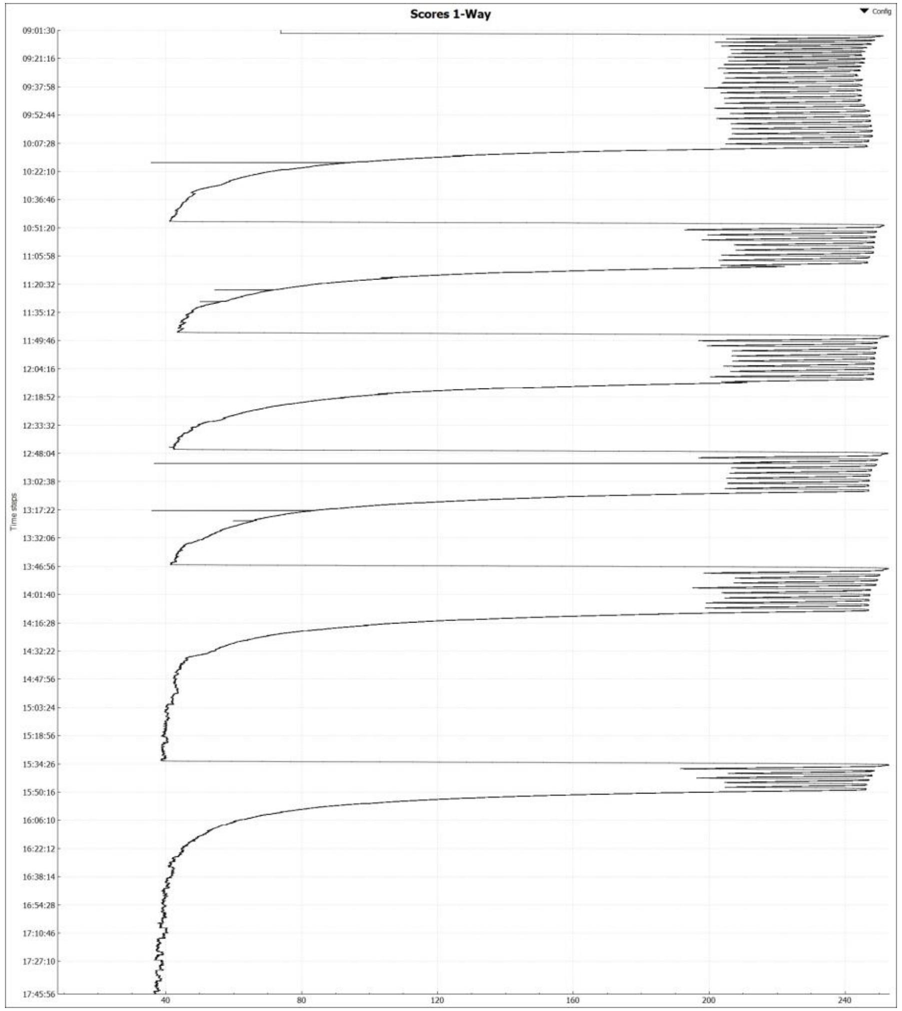
Thermal camera pixels



Discovered State variable: Hamburger grill

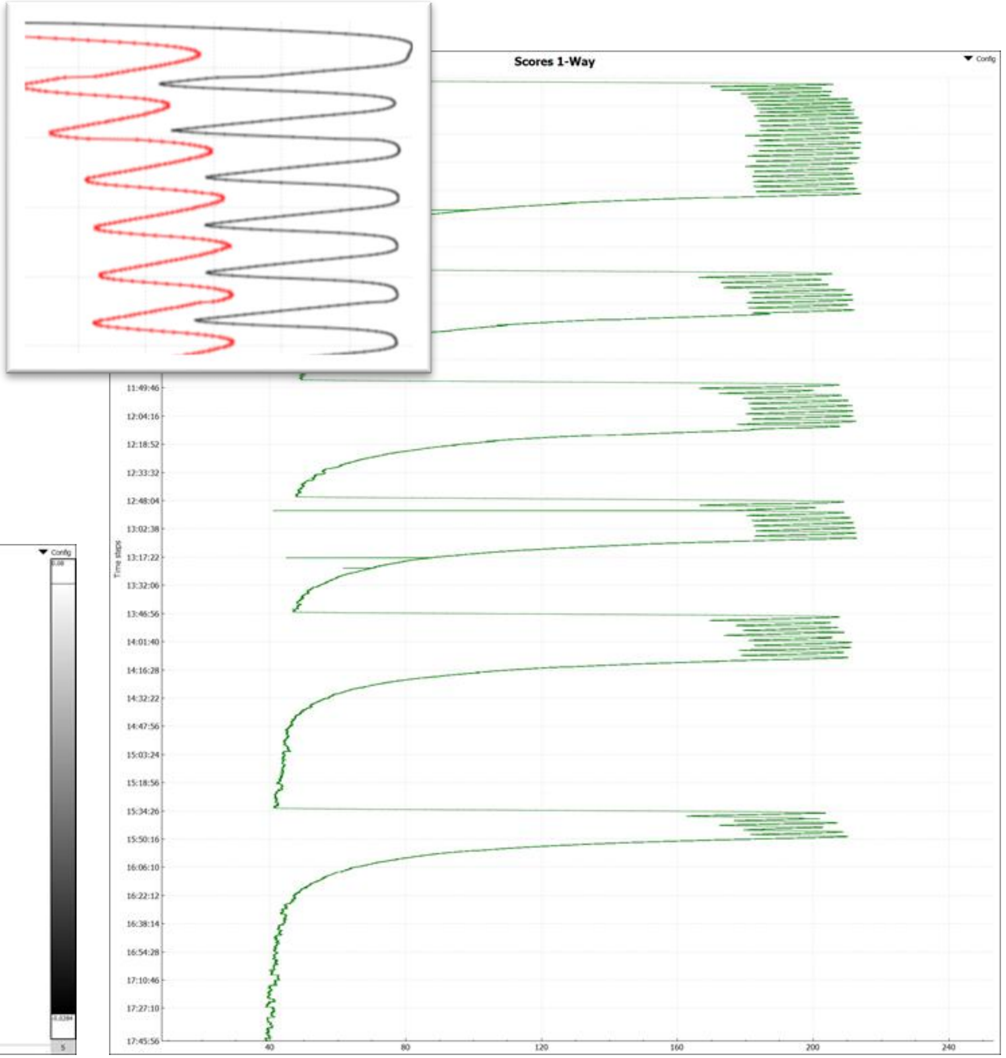
- Trends of two timers
    - Built-in thermostat in equipment
    - 30min timer on equipment
        - Deviation in end, probable mistake in experiment

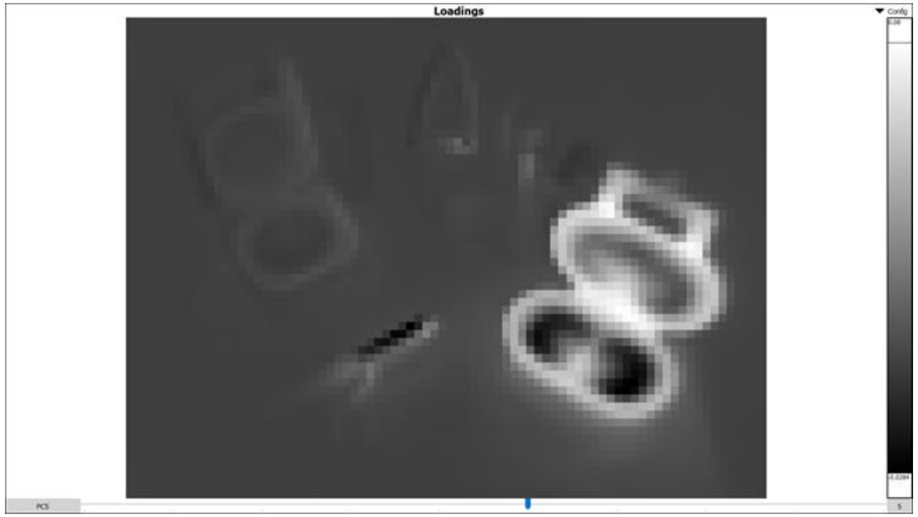




Discovered State variable: Heat dissipation hamburger grill

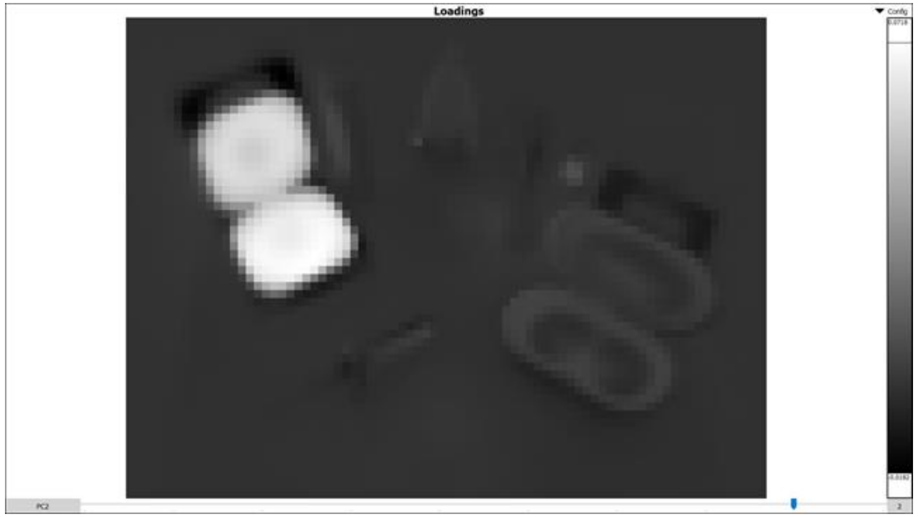
- Same timer trends as hamburger grill
- Small phase offset from heat source, suggests heat dissipation

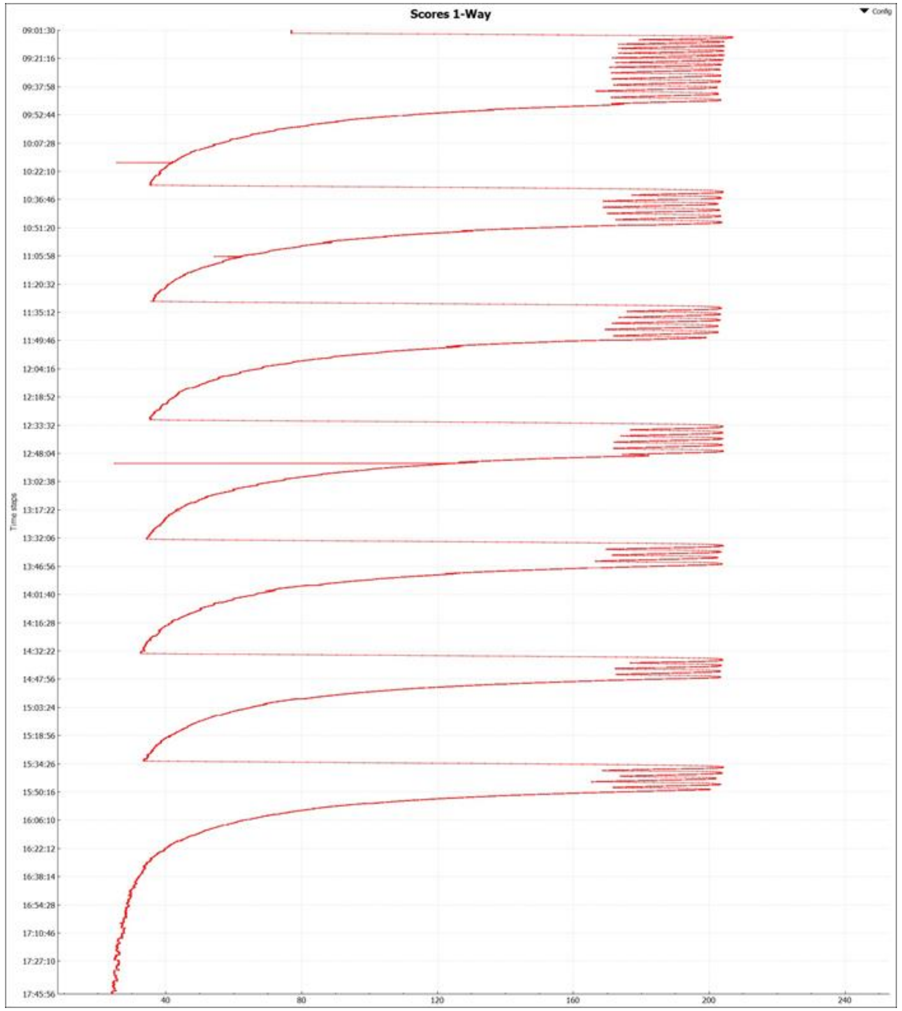




Discovered State variable: Waffle iron

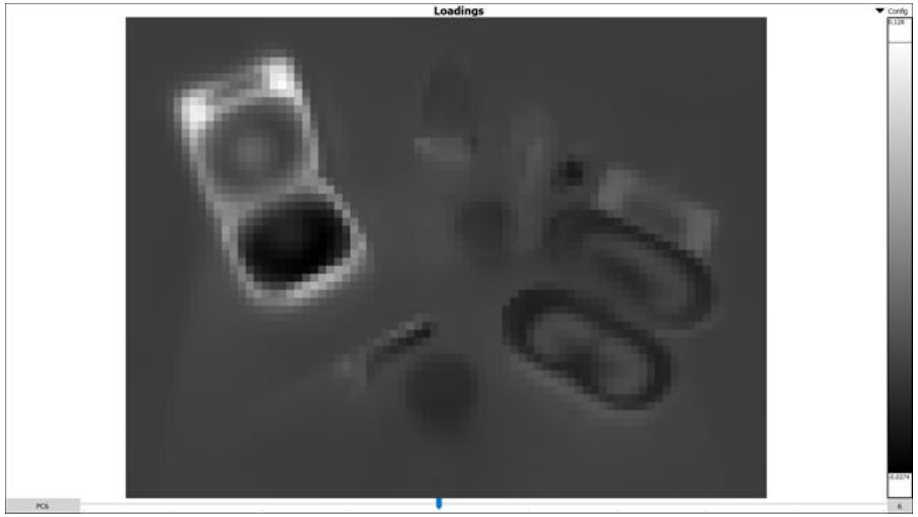
- Trends of two timers
  - Built-in thermostat in equipment
  - 30min timer on equipment

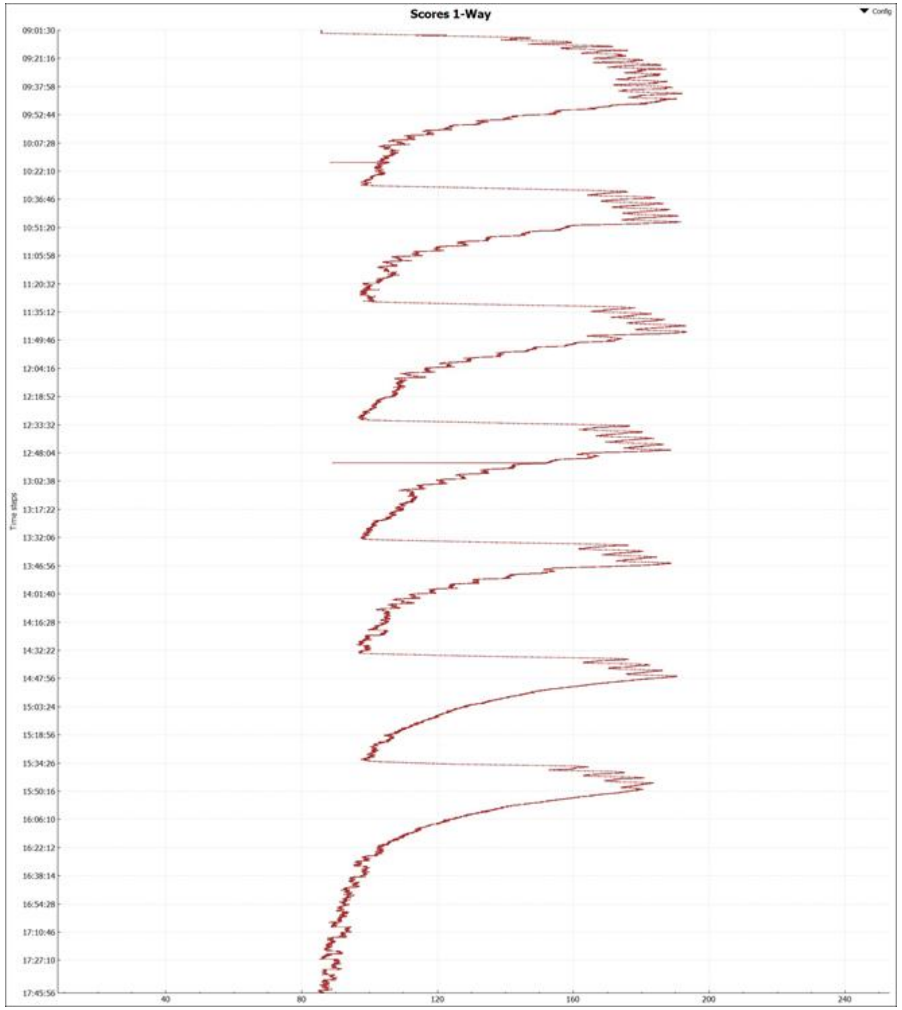




Discovered State variable : Heat dissipation waffle iron

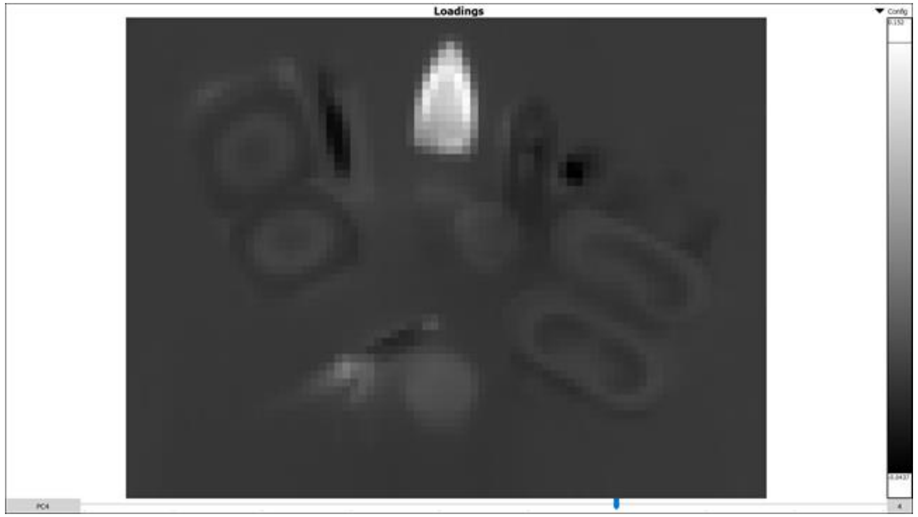
- Same timer trends as waffle iron
- Small phase offset from heat source, suggests heat dissipation

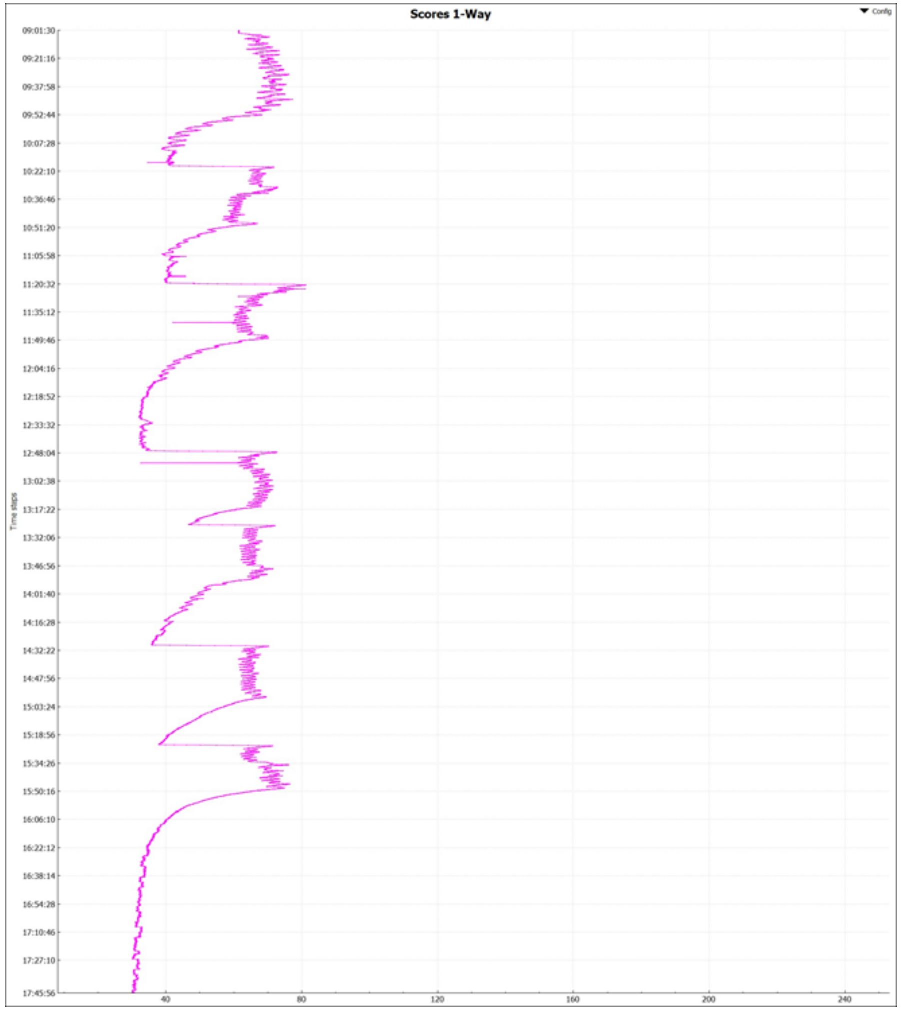




Discovered State variable : Clothes iron

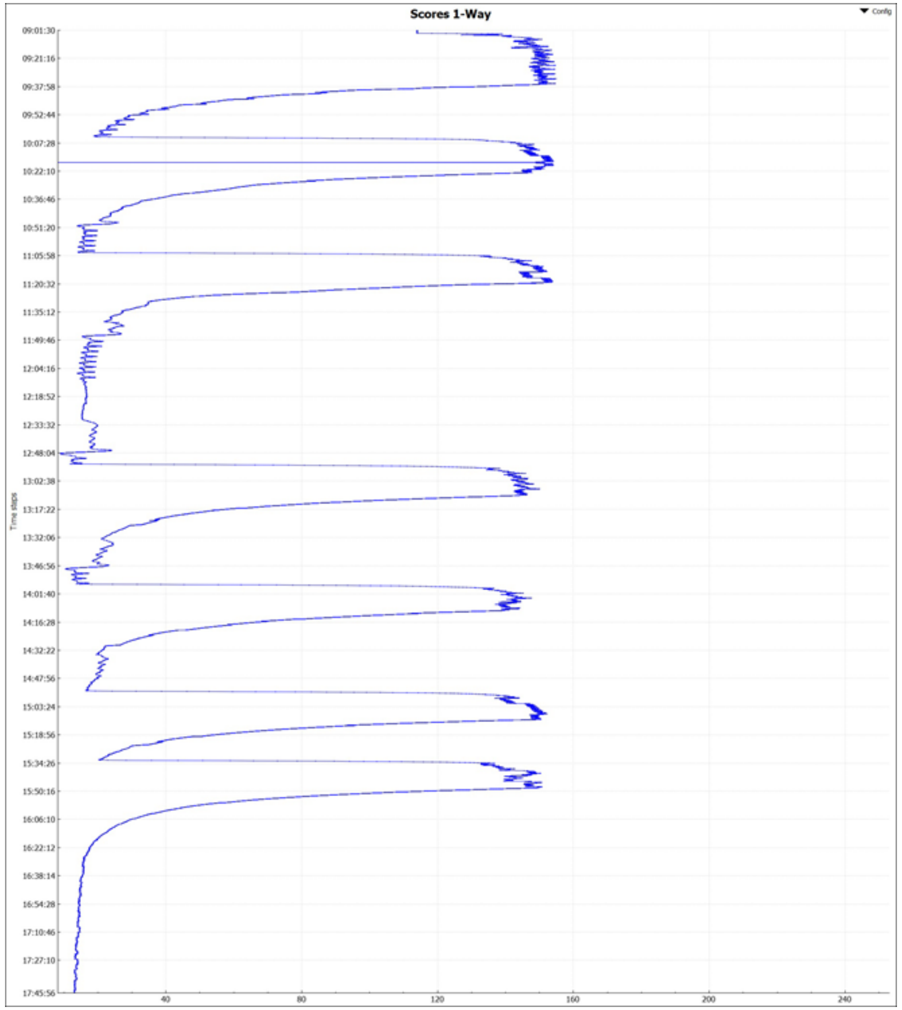
- Trends of two timers
    - Built-in thermostat in equipment
    - Signs of a user manually adjusting timer





Discovered State variable : Curling iron

- Trends of two timers
    - Built-in thermostat in equipment
    - 30min timer on equipment
        - Manual timers in end of day
- Deviation around lunch
    - User paused equipment due to potential fire hazard



idletechs

Example of hybrid modelling:

Hyperspectral monitoring of the drying of wood

idletechs

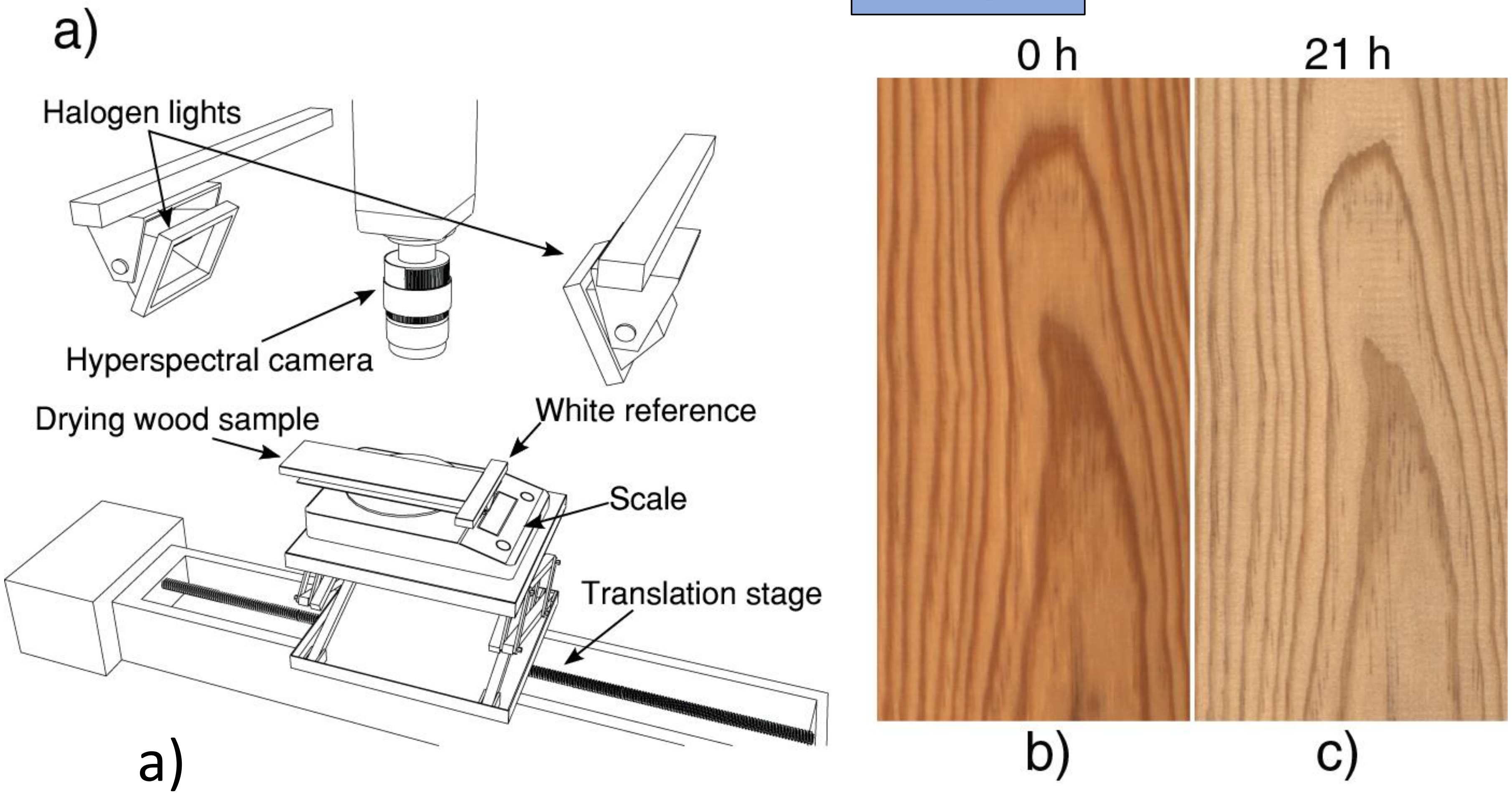
Hyperspectral video analysis: Hyperspectral image data streams interpreted by modeling known and unknown variations

Authors: P. Stefansson[a], J. Fortuna[b, c], H. Rahmati[b], I. Burud[a], T. Konevskikh[a], H. Martens[b, c]

[a]Faculty of Science and Technology, Norwegian University of Life Sciences NMBU, Drøbakveien 31, 1430 Ås
[b]Idletechs AS, Havnegata 9, 7010 Trondheim Norway
[c]Department of Engineering Cybernetics, Norwegian University of Science and Technology NTNU, 7034 Trondheim Norway

(2200 x 1070) pixels
× 159 wavelengths
× 150 time points.

idletechs



What are the causes that control wood drying ?

* Spectra?
* Spatial patterns?
* Time dynamics?

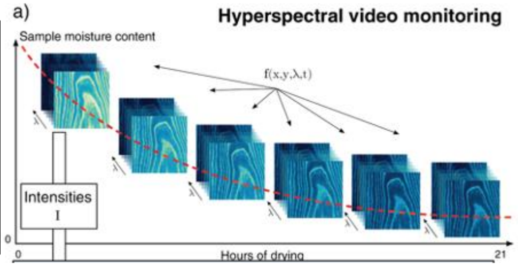
Figure 2.12.1. The experiment. a) Illustration of experimental setup used to measure the spectral reflectance and weight of a drying wood sample. b) RGB rendering of wood sample in wet state (drying time = 0 hours). c) RGB rendering of wood sample in dry state (drying time = 21 hours).

Hyperspectral video analysis: Hyperspectral image data streams interpreted by modeling known and unknown variations

Authors: P. Stefansson[a], J. Fortuna[b, c], H. Rahmati[b], I. Burud[a], T. Konevskikh[a], H. Martens[b, c]

[a]Faculty of Science and Technology, Norwegian University of Life Sciences NMBU, Drøbakveien 31, 1430 Ås
[b]Idletechs AS, Havnegata 9, 7010 Trondheim Norway
[c]Department of Engineering Cybernetics, Norwegian University of Science and Technology NTNU, 7034 Trondheim Norway

(2200 x 1070) pixels
× 159 wavelengths
× 150 time points.



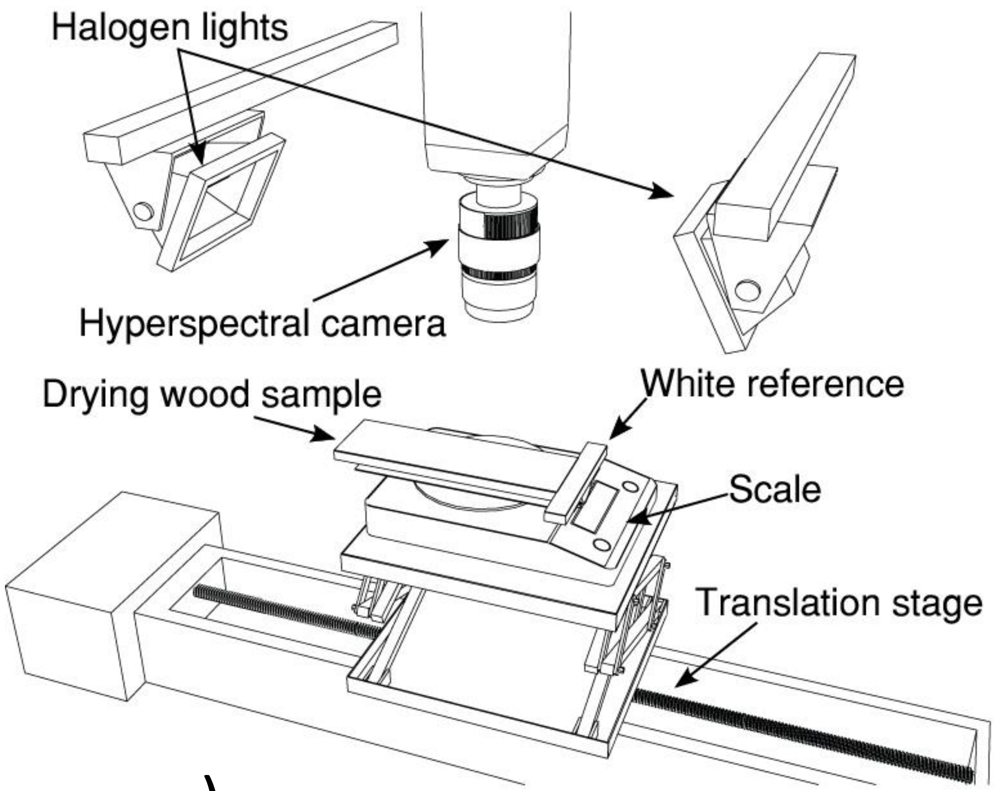
0 h 21 h

>300 GB raw data from one experiment

What are the causes that control wood drying ?

* Spectra?
* Spatial patterns?
* Time dynamics?

a)



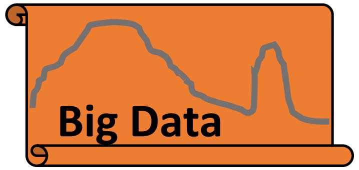
a)

b) c)

Figure 2.12.1. The experiment. a) Illustration of experimental setup used to measure the spectral reflectance and weight of a drying wood sample. b) RGB rendering of wood sample in wet state (drying time = 0 hours). c) RGB rendering of wood sample in dry state (drying time = 21 hours).

Quantitative Big Data in Bio-sciences

>300 GB raw data from one experiment



P. Stefansson, J. Fortuna, H. Rahmati, I. Burud, T. Konevskikh, H. Martens (2019): Hyperspectral video analysis: Hyperspectral image data streams interpreted by modeling known and unknown variations. *In press*.

Quantitative Big Data in Bio-sciences

« AI » :

>300 GB raw data from one experiment

Big Data

Limited understanding

Machine learning:
Opening the black box

No understanding

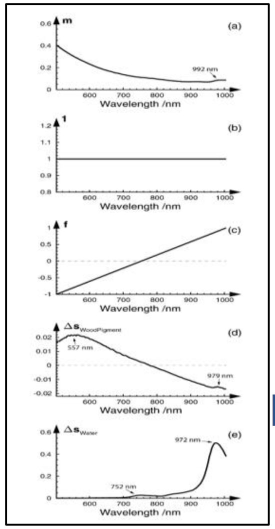
No Learning

Uncertain Control

idletechs

Quantitative Big Data in Bio-sciences

« XAI » : Combine knowledge and data



>300 GB raw data from one experiment

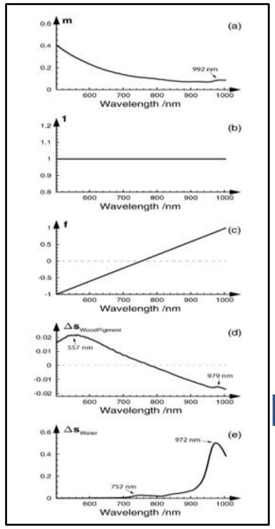
Limited understanding

Theory-driven mathematical modelling: Multivariate meta-modelling

idletechs

Quantitative Big Data in Bio-sciences

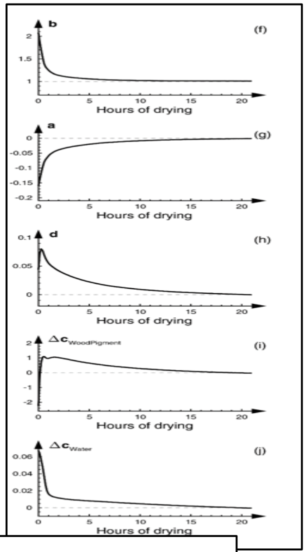
« XAI » : Combine knowledge and data

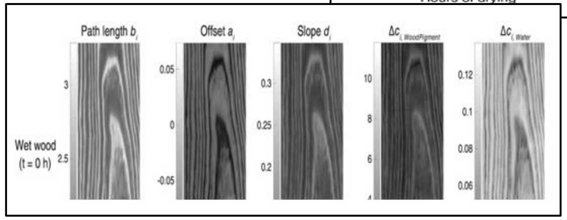


Big Data

Limited understanding

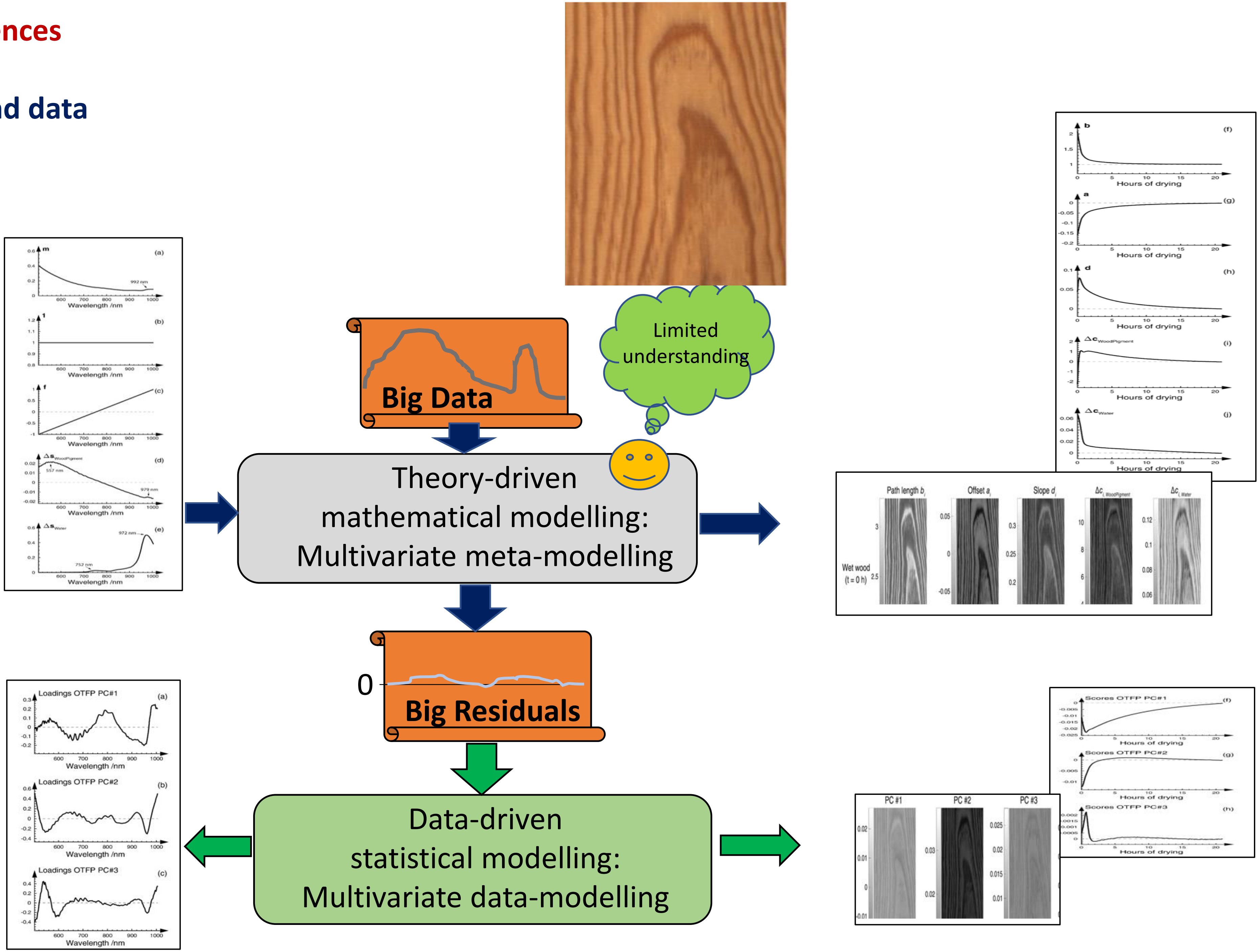
Theory-driven mathematical modelling: Multivariate meta-modelling





Quantitative Big Data in Bio-sciences

« XAI » : Combine knowledge and data

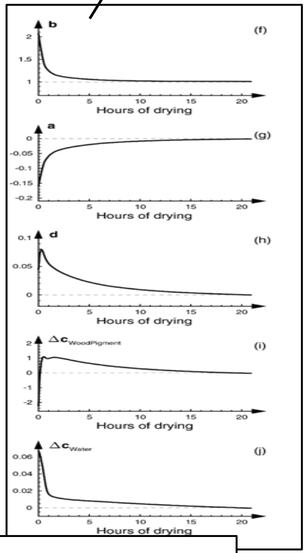


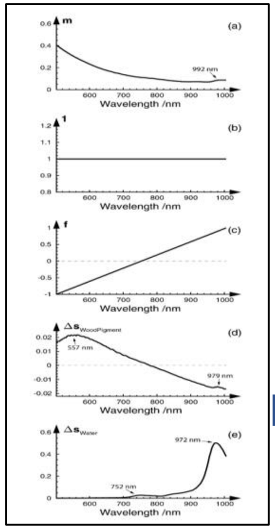
idletechs

Quantitative Big Data in Bio-sciences

« XAI » : Combine knowledge and data

ODEs

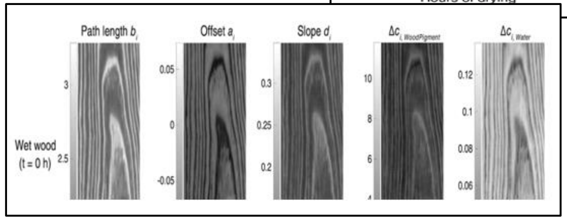




Big Data

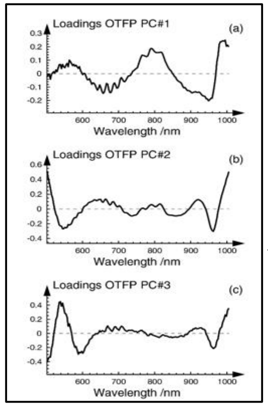
Limited understanding

Theory-driven mathematical modelling: Multivariate meta-modelling

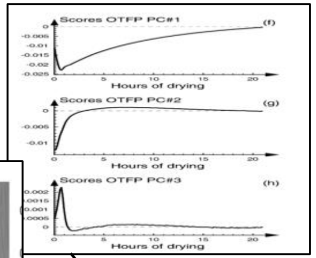


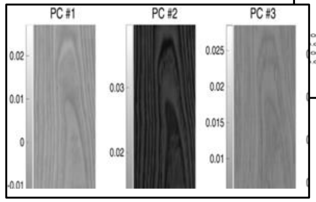
0

Big Residuals



Data-driven statistical modelling: Multivariate data-modelling



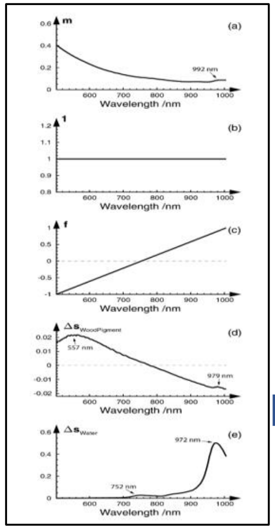


ODEs

idletechs

Quantitative Big Data in Bio-sciences

« XAI » : Combine knowledge and data



Limited understanding

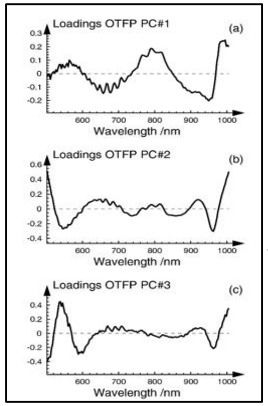
Big Data

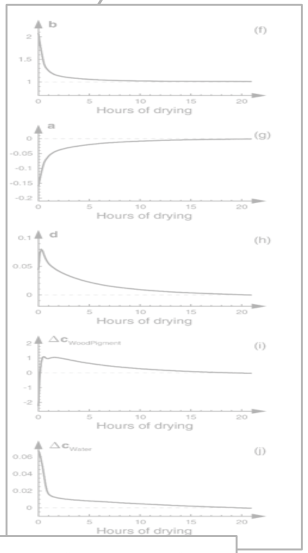
Theory-driven mathematical modelling: Multivariate meta-modelling

0

Big Residuals

Data-driven statistical modelling: Multivariate data-modelling





ODEs

Better understanding

Better Theory-driven model

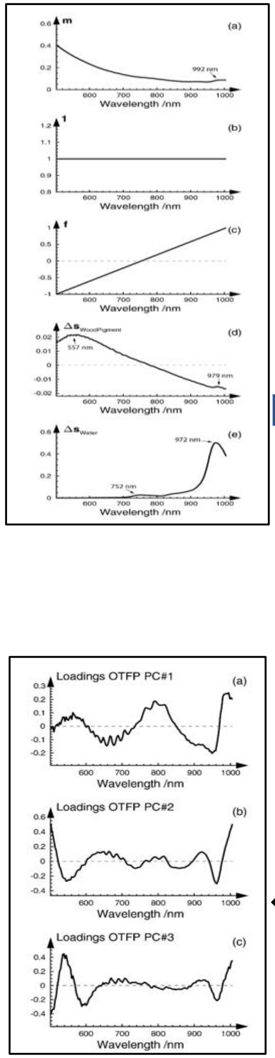
Better Conrol

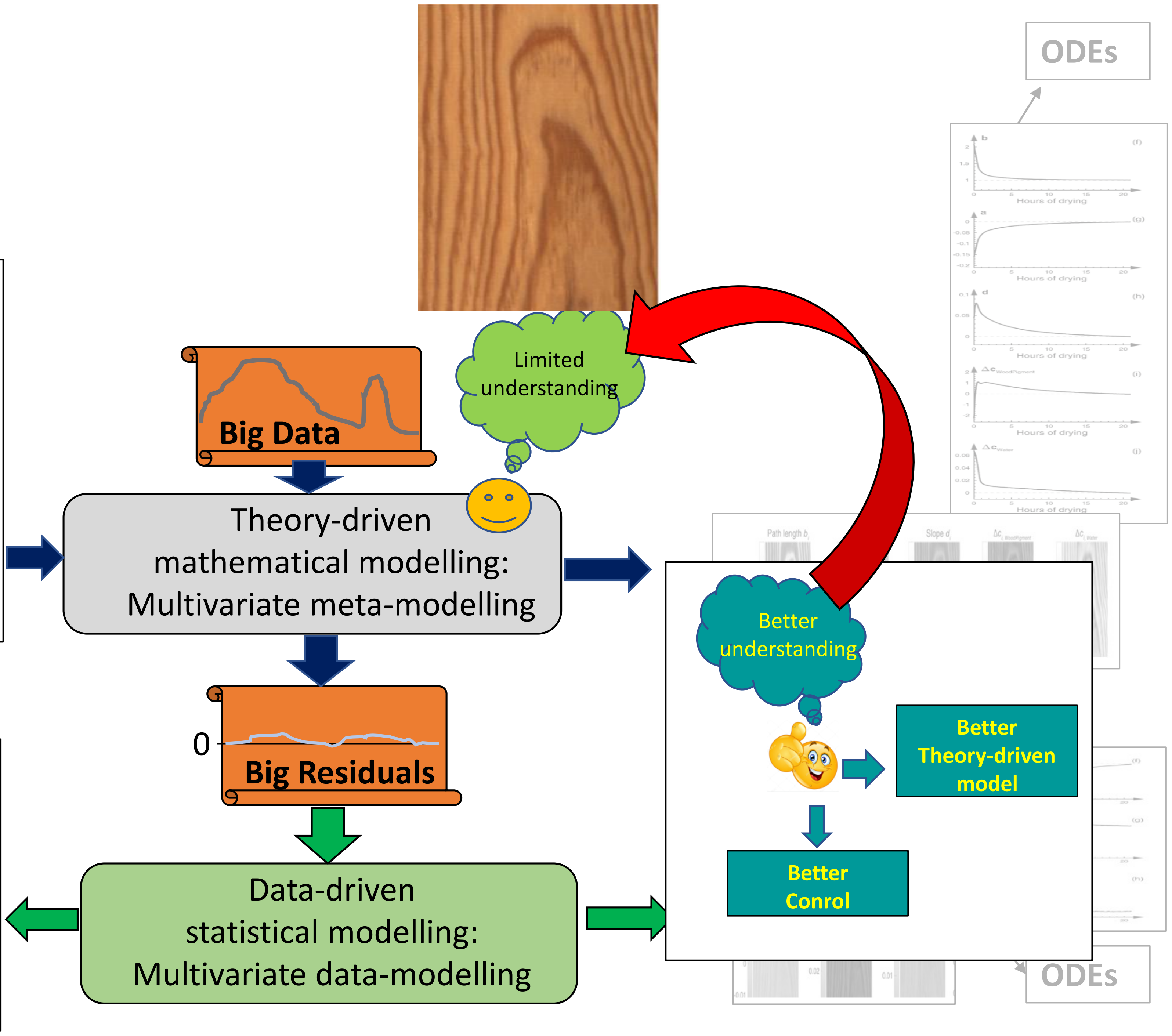
ODEs

idletechs

Quantitative Big Data in Bio-sciences

« XAI » : Combine knowledge and data





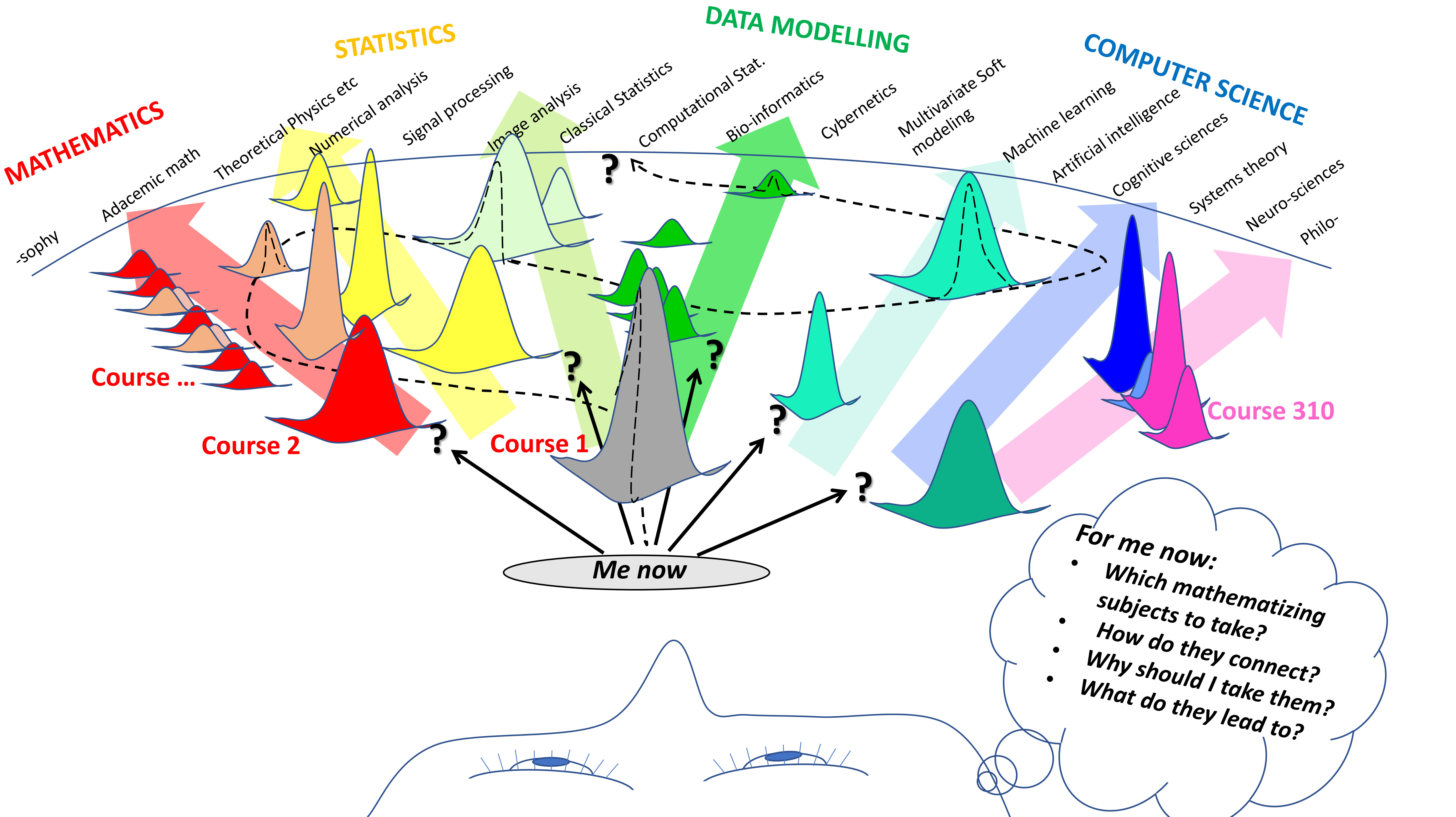
idletechs

Since the Enlightenment: Understand more!
Now, with BIG DATA: Understand less?

BIG DATA:
Keep humans in the loop!

We need more Mathematical Modelling, more Statistical Assessment and more Learning from Data,
but less Macho Mathematics, less Gucci Statistics and less Blind Machine Learning

Thank you!


MATHEMATICS
STATISTICS
DATA MODELLING
COMPUTER SCIENCE
-sophy
Adacemic math
Theoretical Physics etc
Numerical analysis
Signal processing
Image analysis
Classical Statistics
Computational Stat.
Bio-informatics
Cybernetics
Multivariate Soft modeling
Machine learning
Artificial intelligence
Cognitive sciences
Systems theory
Neuro-sciences
Philo-
Course ...
Course 2
Course 1
Course 310
Me now
For me now:
• Which mathematizing subjects to take?
• How do they connect?
• Why should I take them?
• What do they lead to?