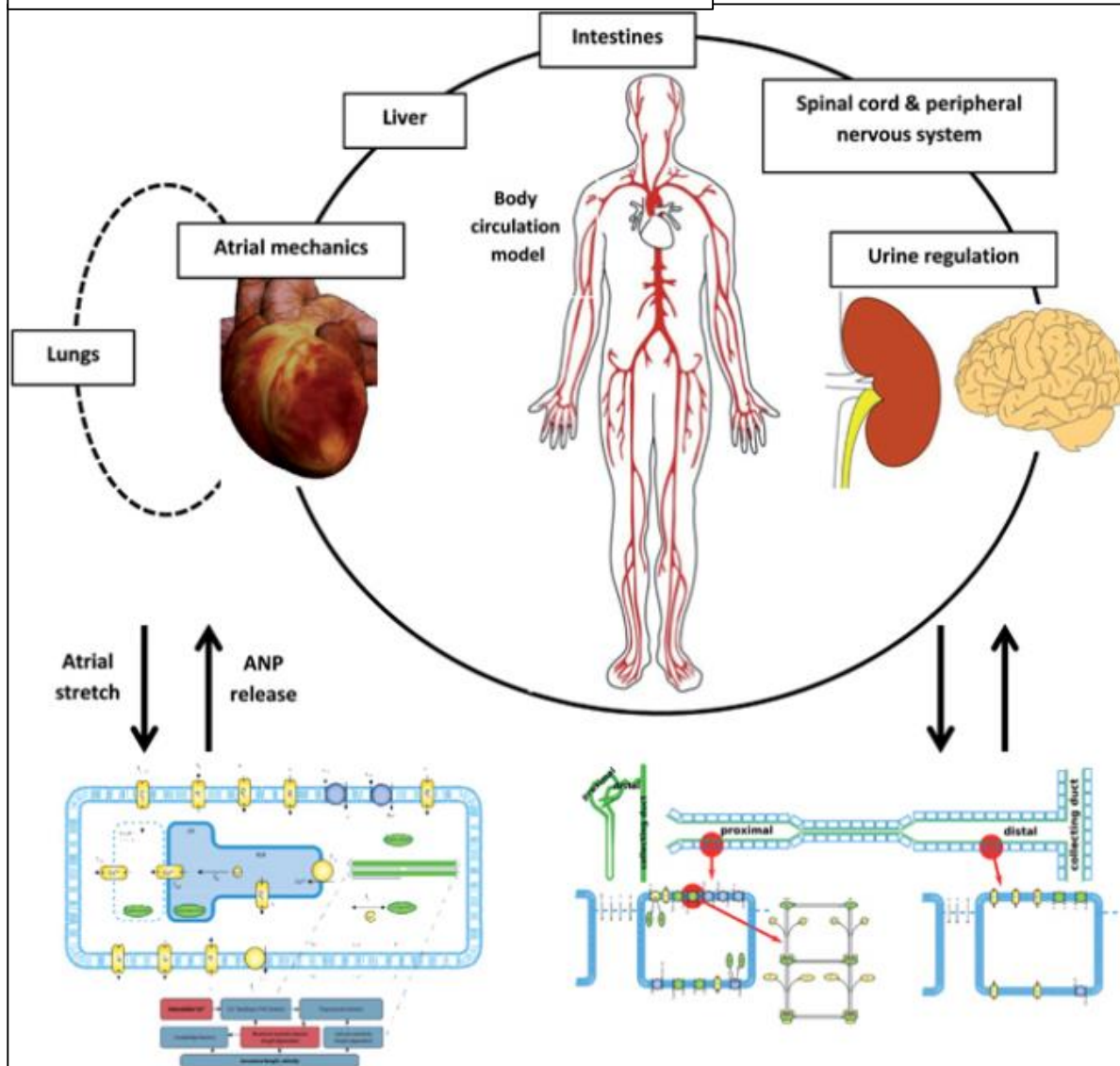


Høydimensjonale datastrømmer i helsesektoren:  
Ikke Gucci-statistikk, ikke Macho-matematikk,  
ikke Blind Informatikk,  
men heller *Rhythm & Blues fra data*

Harald Martens<sup>1,2</sup>, Jan Otto Reberg<sup>2</sup>, Joa Fortuna<sup>1</sup> og Nils Kristian Skjærvold<sup>3</sup>

<sup>1</sup>Inst. teknisk kybernetikk, IME, NTNU, <sup>2</sup>Idletechs AS, <sup>3</sup>Inst. sirkulasjon og bildediagnostikk, Med. Fak, NTNU, Trondheim.



**Models Home**

**Exposures**

**Documentation**

You are here: [Home](#) / [FieldML](#)

## FieldML



**Home**

**About CellML**

**Getting started**

**Tools**

**Models**

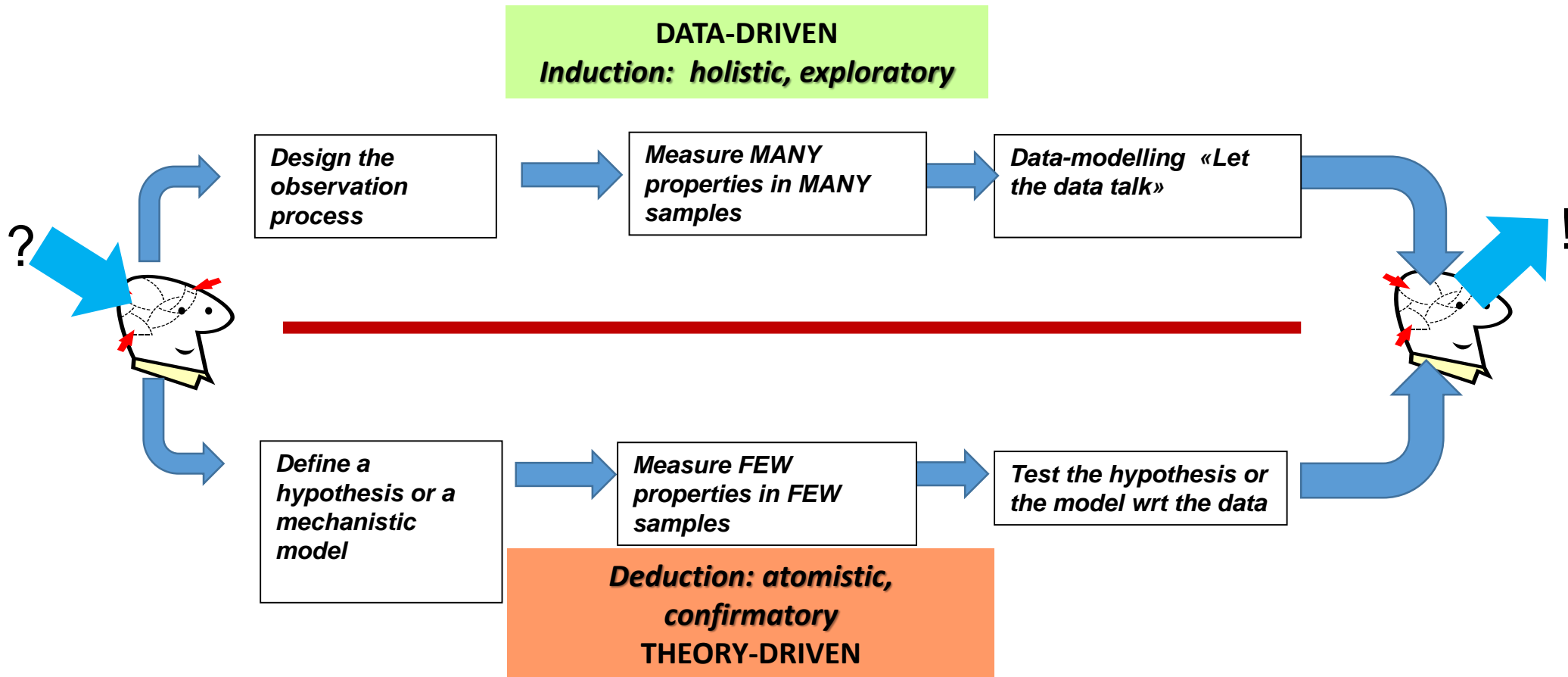
**Community**

You are here: [Home](#) / [Tools](#) / Visualisation and annotation tools

### Visualisation and annotation tools

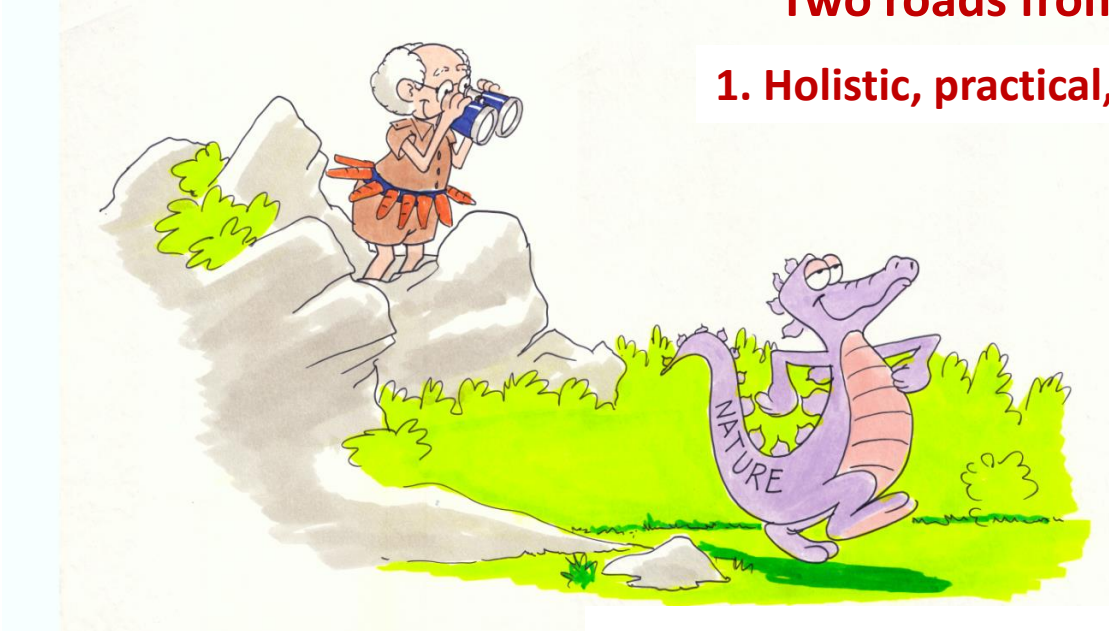
#### Model Visualisation

# Two roads from question to answer



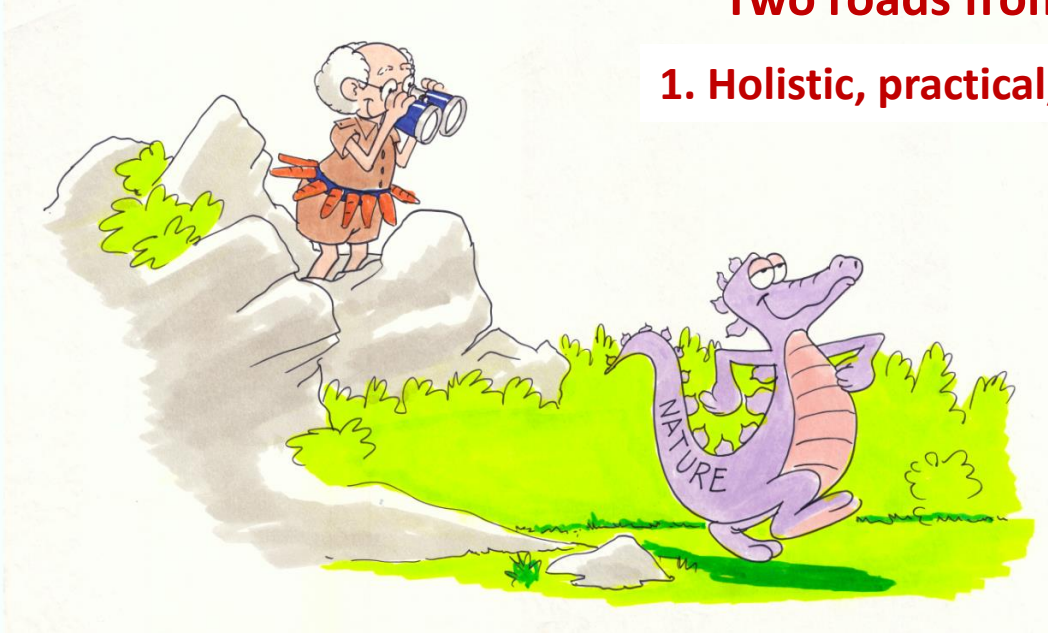
## Two roads from question to answer in science

### 1. Holistic, practical, curious exploration

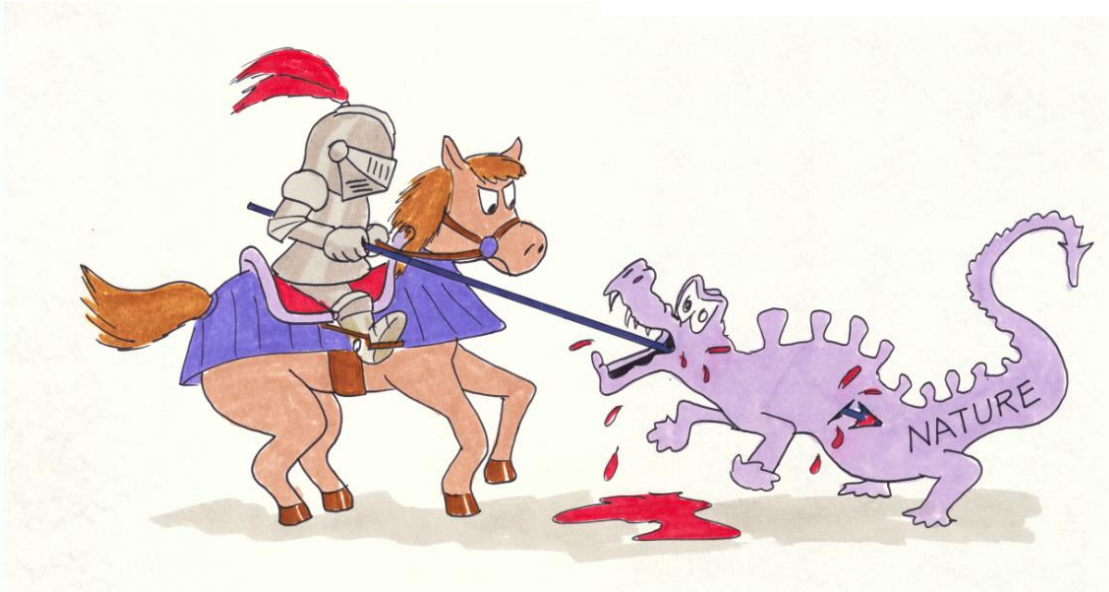


## Two roads from question to answer in science

### 1. Holistic, practical, curious exploration



### 2. Atomistic, theoretical hypothesis testing



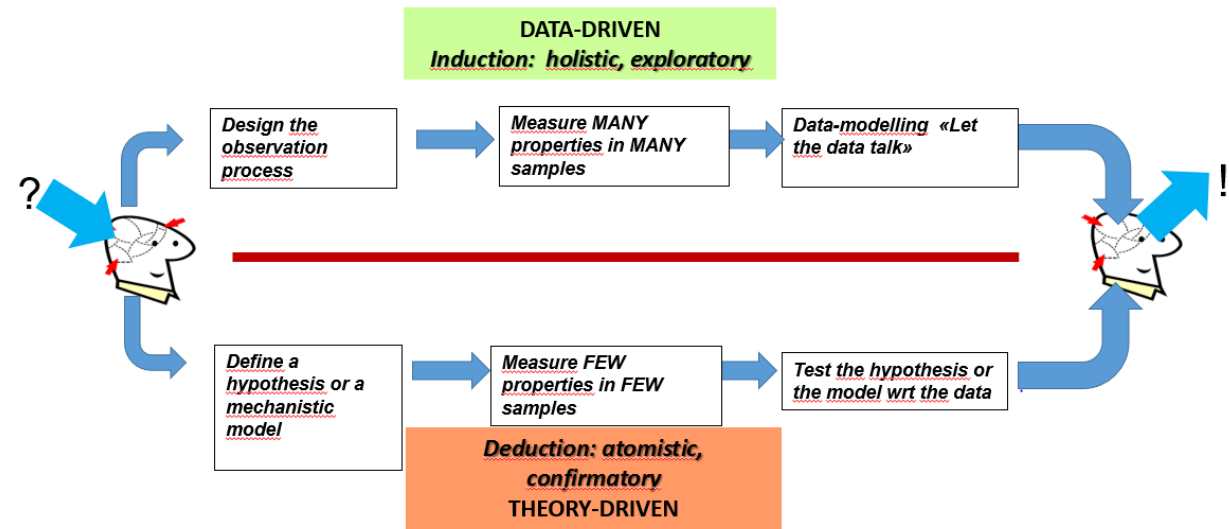
With permission from prof. Lars Munck

*Rhythm & Blues fra data :*  
*Matematikk uten tårer*

1: Verden har mønstre – finn dem

2: Utvid sansene

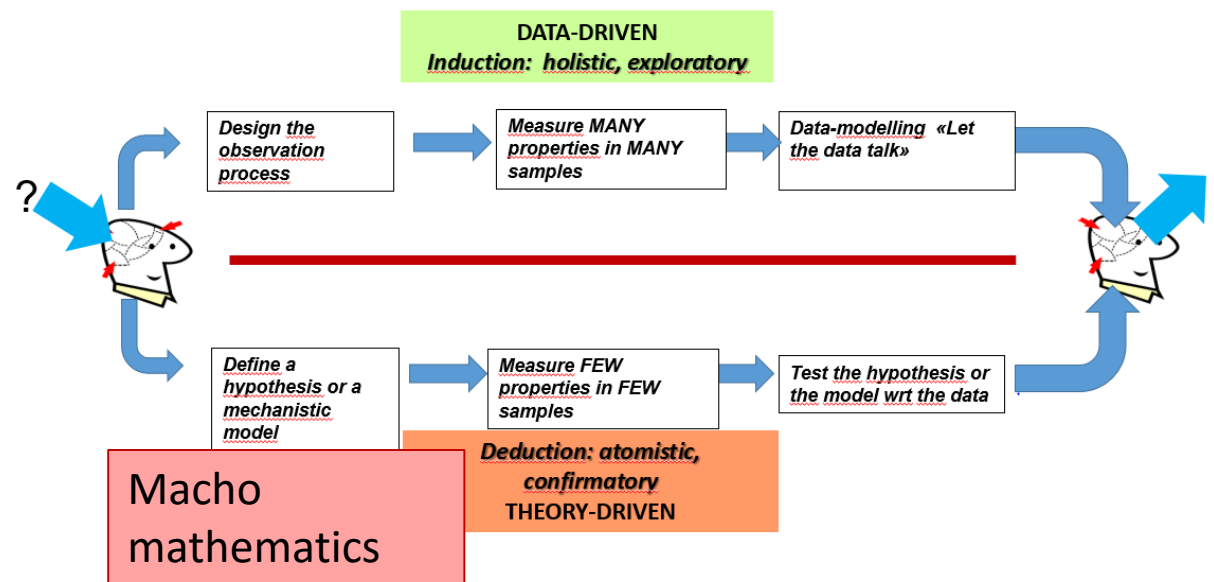
3: Mektig matte uten tårer



Matematisk modellering,  
Statistisk evaluering og  
Data modellering:  
Viktige verktøy!

# Macho-matematikk?

Vi trenger: Mer og bedre matematisk modellering for å samle opp og binde sammen detaljkunnskapene.  
Men uten: «**macho-matematikk**» - dvs urealistisk og arrogant reduksjonisme ('my Model, right or wrong!').

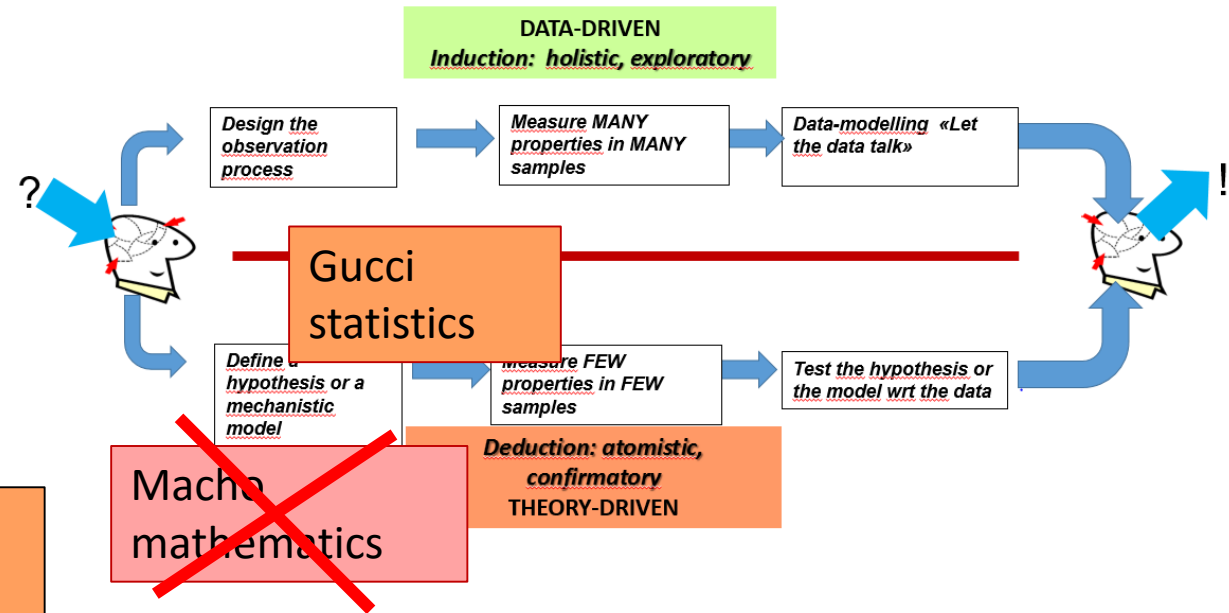




# Gucci-statistikk?

Vi trenger: Mer og bedre matematisk modellering for å samle opp og binde sammen detaljkunnskapene.  
Men uten: «**macho-matematikk**» - dvs urealistisk og arrogant reduksjonisme ('my Model, right or wrong!').

Vi trenger: Mer og bedre bruk av statistisk vurdering av pålitelighet.  
Men uten: «**Gucci-statistikk**» - dvs forfengelig, ukyndig fokus på univariate p-verdier ('permission to publish').

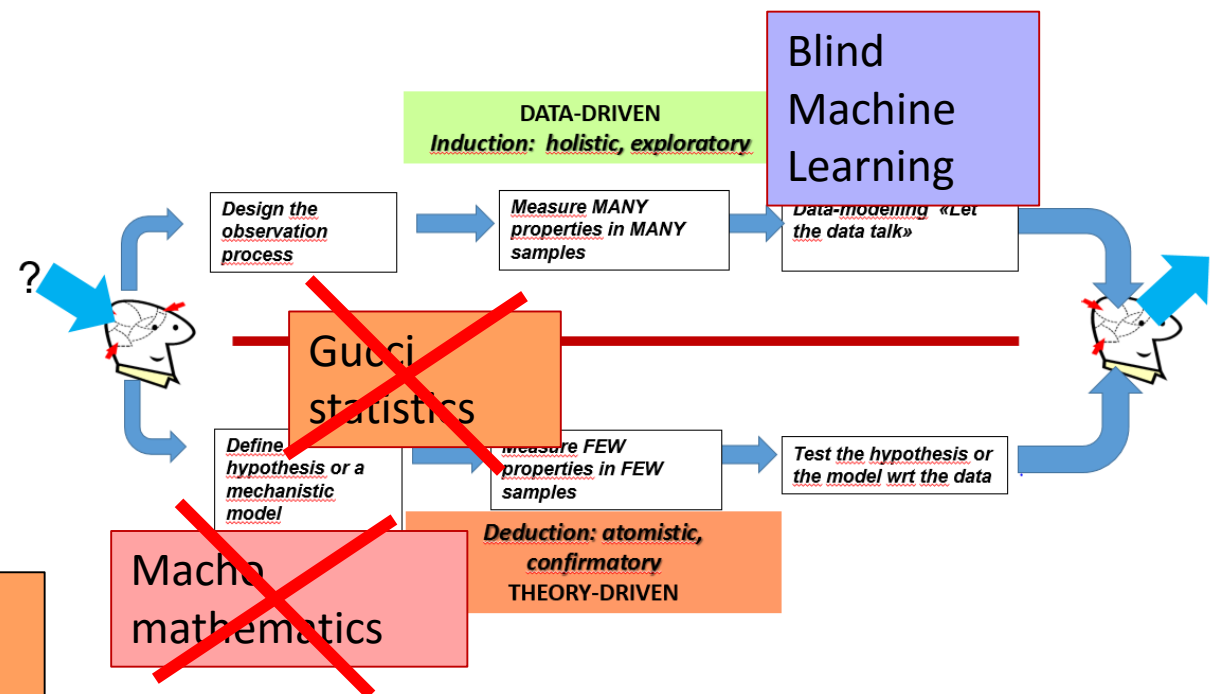


# Blind maskinlæring?

Vi trenger: Mer og bedre matematisk modellering for å samle opp og binde sammen detaljkunnskapene.  
Men uten: «**macho-matematikk**» - dvs urealistisk og arrogant reduksjonisme ('my Model, right or wrong!').

Vi trenger: Mer og bedre bruk av statistisk vurdering av pålitelighet.  
Men uten: «**Gucci-statistikk**» - dvs forfengelig, ukyndig fokus på univariate p-verdier ('permission to publish').

Vi trenger: Mer og bedre bruk av beregningsalgoritmer til å trekke ut skjult essens ifra overveldende datastrømmer.  
Men uten: «**Blind maskinlæring**», d.v.s. Black box løsninger (ANN osv) implementert slik at INGEN mennesker forstår.



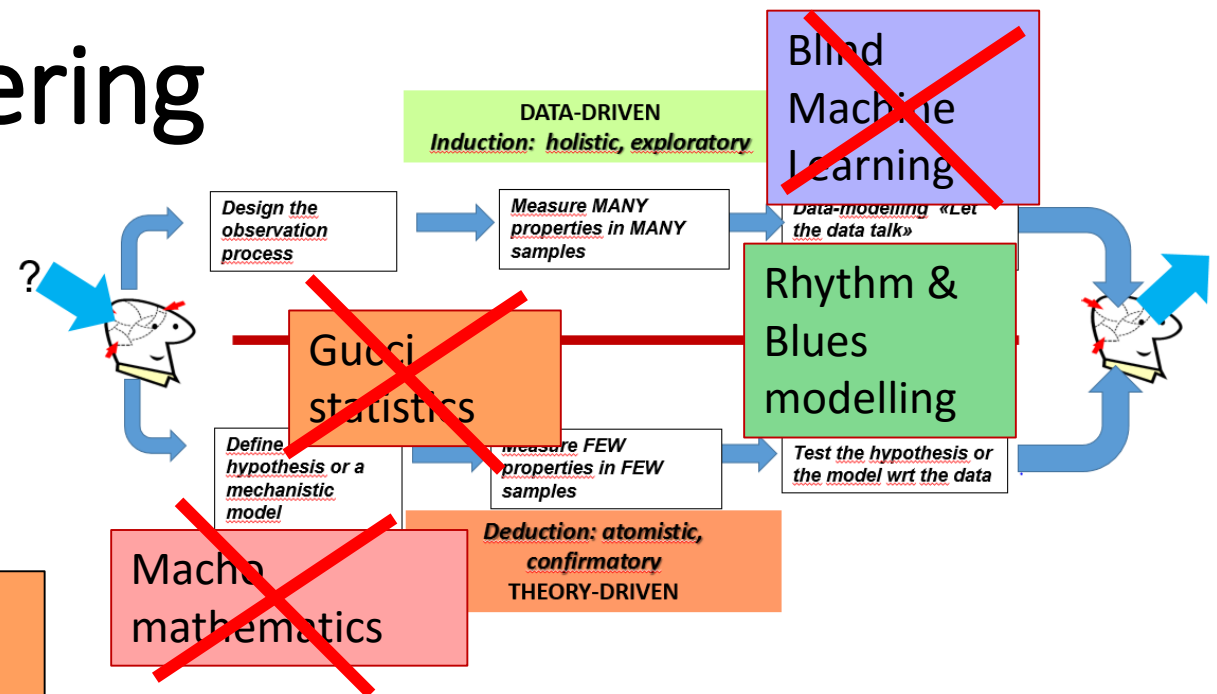
# Multivariat myk modellering

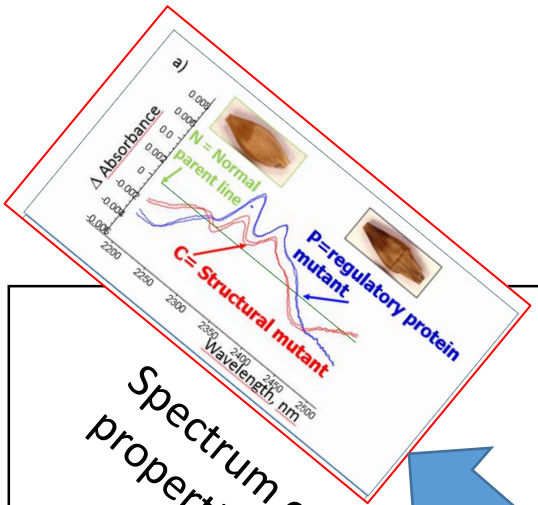
Vi trenger: Mer og bedre matematisk modellering for å samle opp og binde sammen detaljkunnskapene.  
Men uten: «**macho-matematikk**» - dvs urealistisk og arrogant reduksjonisme ('my Model, right or wrong!').

Vi trenger: Mer og bedre bruk av statistisk vurdering av pålitelighet.  
Men uten: «**Gucci-statistikk**» - dvs forfengelig, ukyndig fokus på univariate p-verdier ('permission to publish').

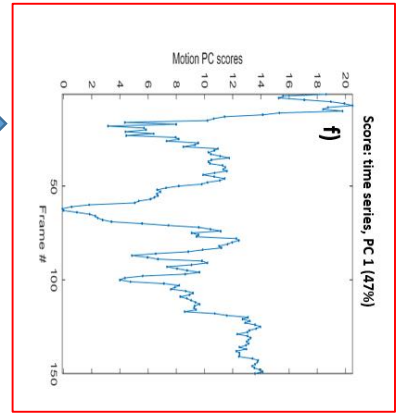
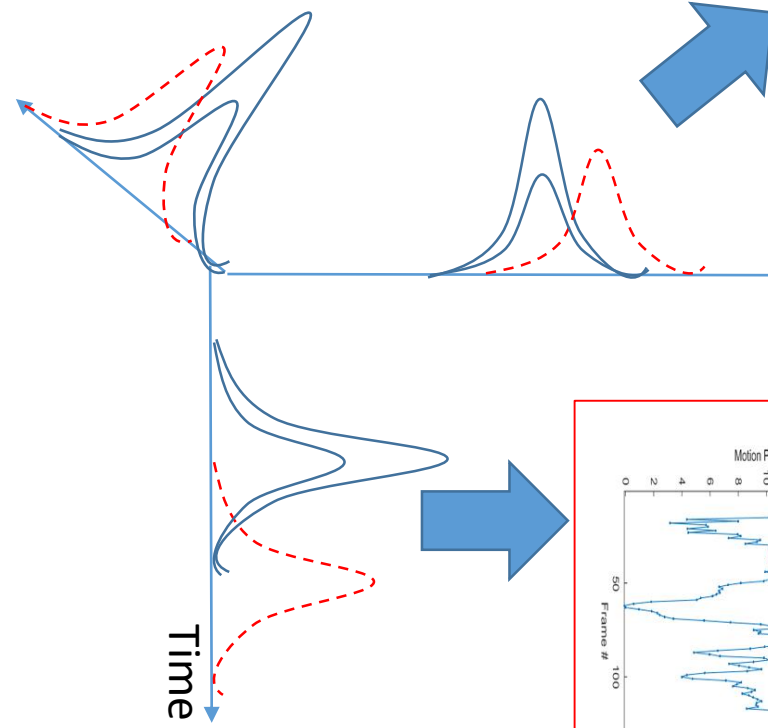
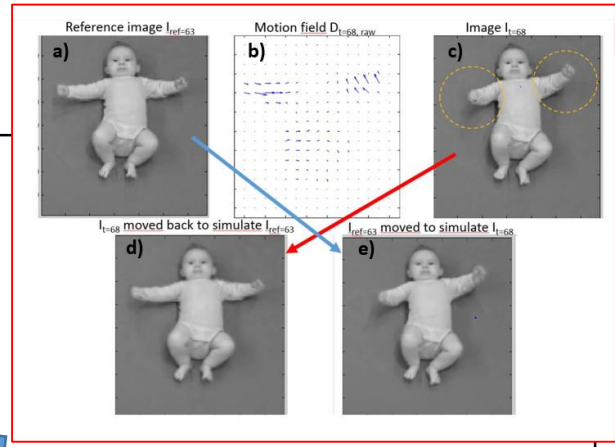
Vi trenger: Mer og bedre bruk av beregningsalgoritmer til å trekke ut skjult essens ifra overveldende datastrømmer.  
Men uten: «**Blind maskinlæring**», d.v.s. Black box løsninger (ANN osv) implementert slik at INGEN mennesker forstår.

Flerfaglig brobyggende bruk av Big Data: med mennesket i 'loopen'.





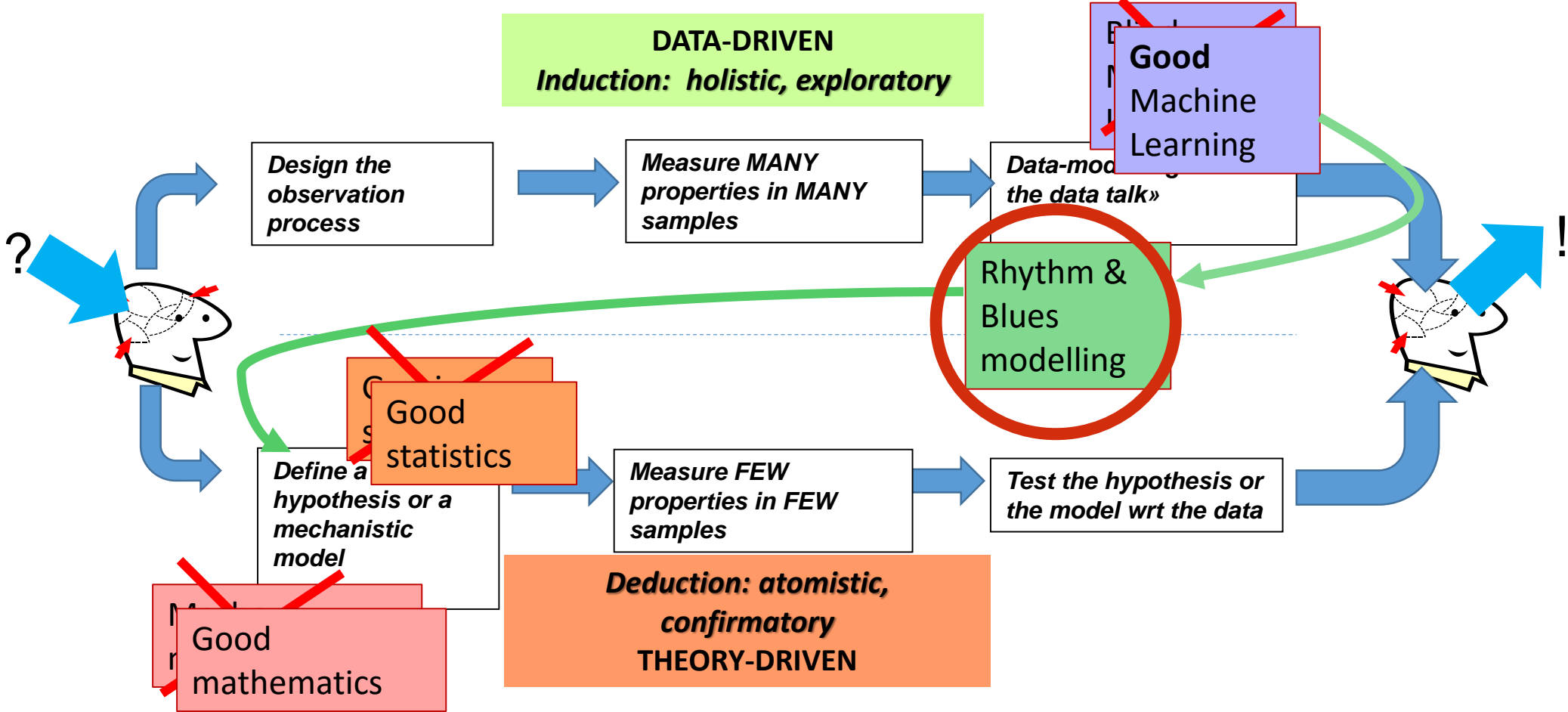
Spectrum of properties



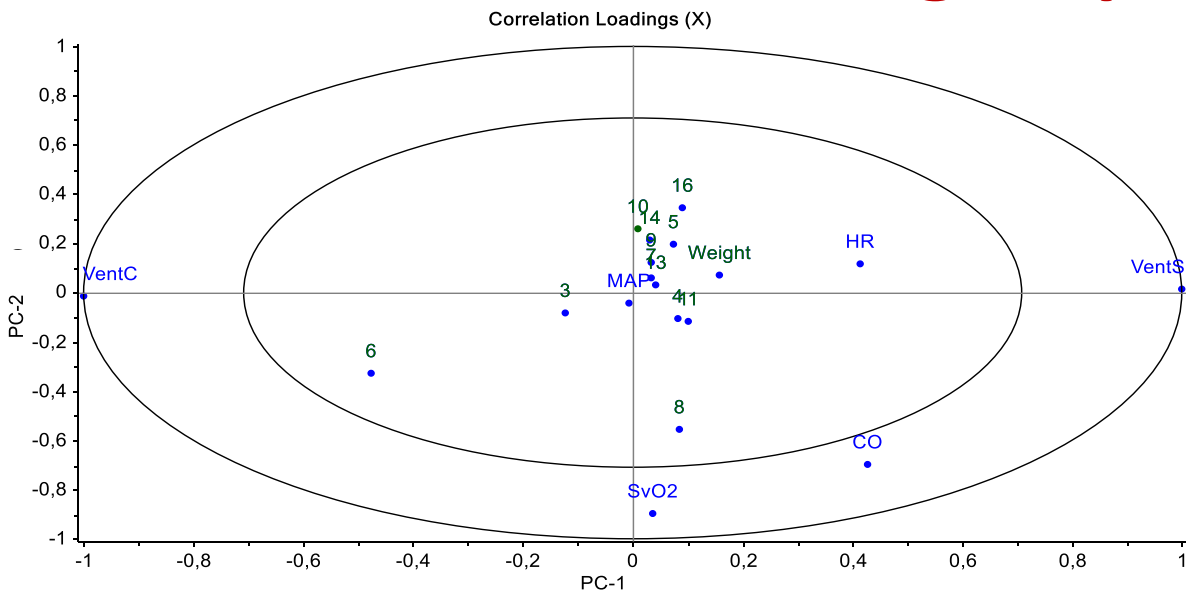
# Rhythm & Blues i BIG DATA: *Matematikk uten tårer*

1: Verden har mønstre – finn dem

# Multivariat Myk Data-Modelling:



# Intensiv-overvåking av pasienter, St.Olavs:

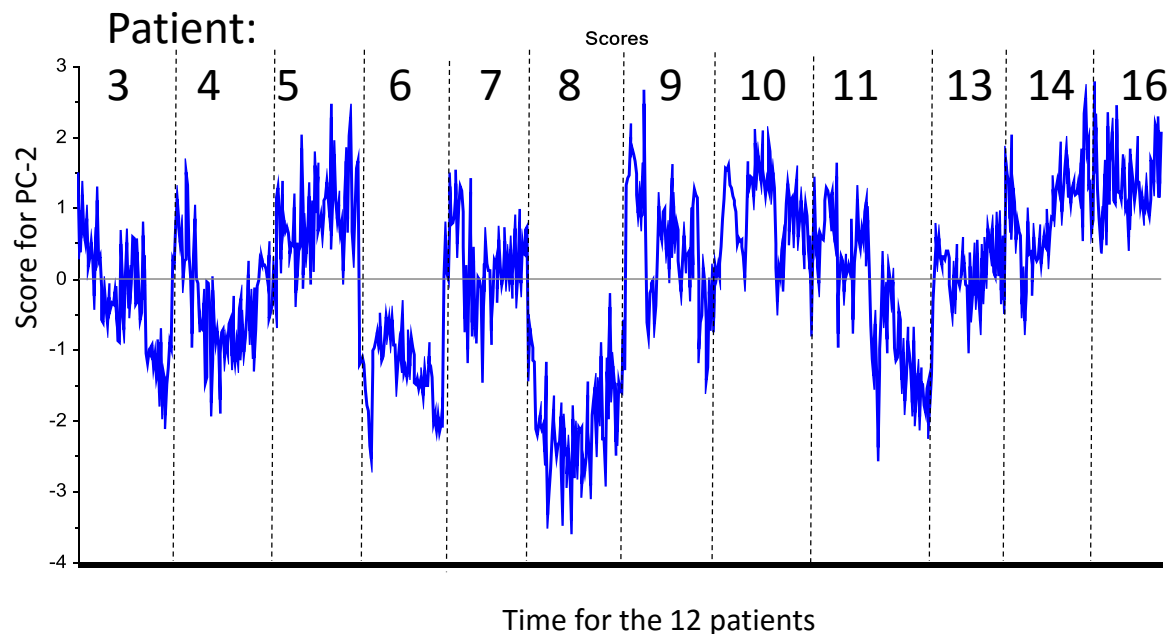


For mange og for dårlige alarmer

Oppsummer alle sensorenes signaler.

Mål:

- Få bedre medisinsk oversikt
- Få ned antall alarmer



Dr. Nils Kristian Skjærvold m. fl.,  
St.Olavs 2016

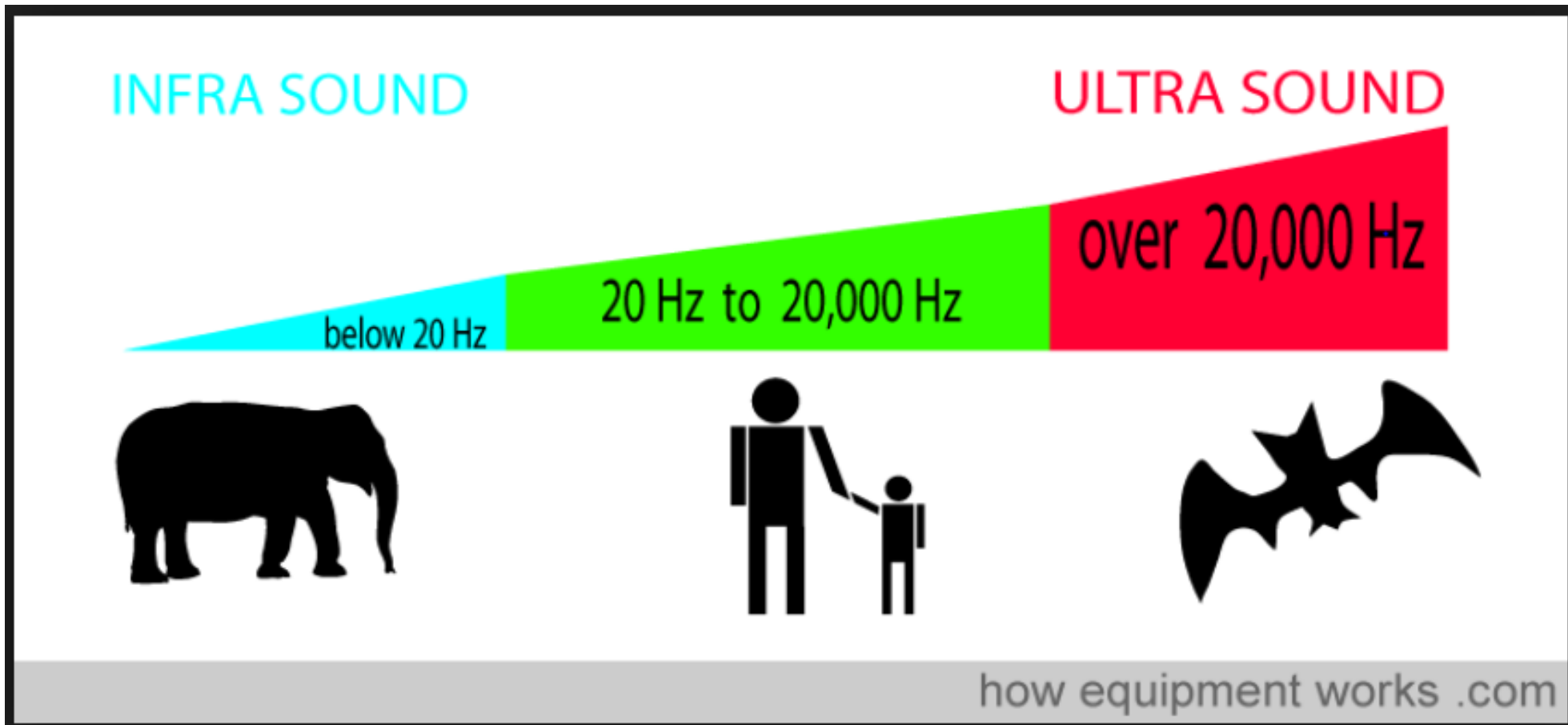
# Rhythm & Blues i BIG DATA: *Matematikk uten tårer*

1: Verden har mønstre – finn dem

2: Utvid sansene

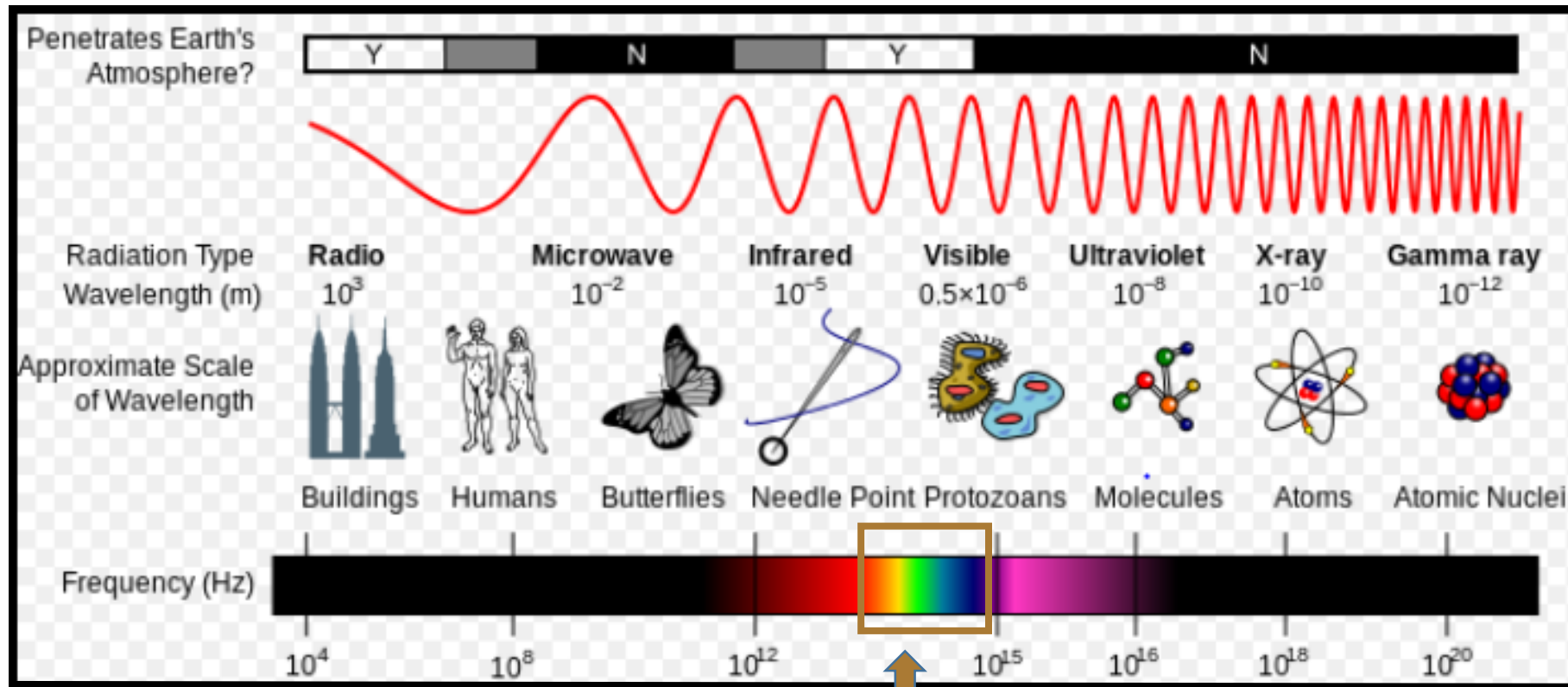


# Vår hørsel er bra, men begrenset



## 2: Utvid sansene

# Vårt fargesyn er bra, men begrenset

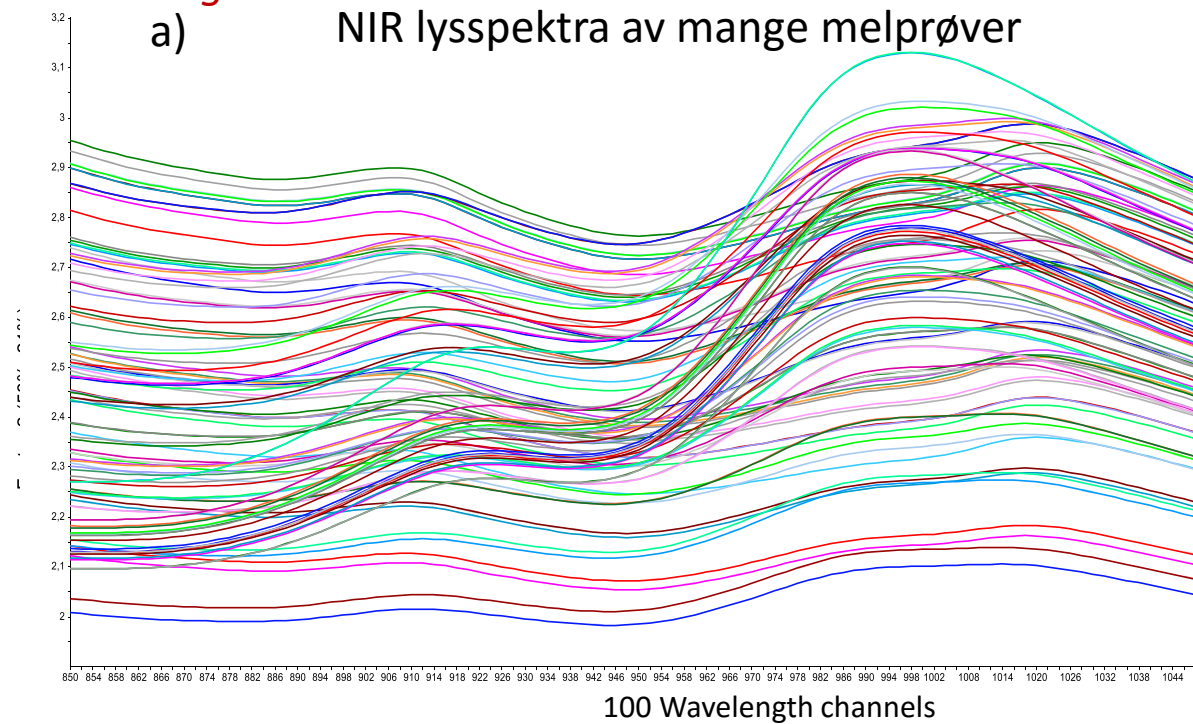


Source: [http://hubblesite.org/reference\\_desk/faq/answer.php.id=70&cat=light](http://hubblesite.org/reference_desk/faq/answer.php.id=70&cat=light)

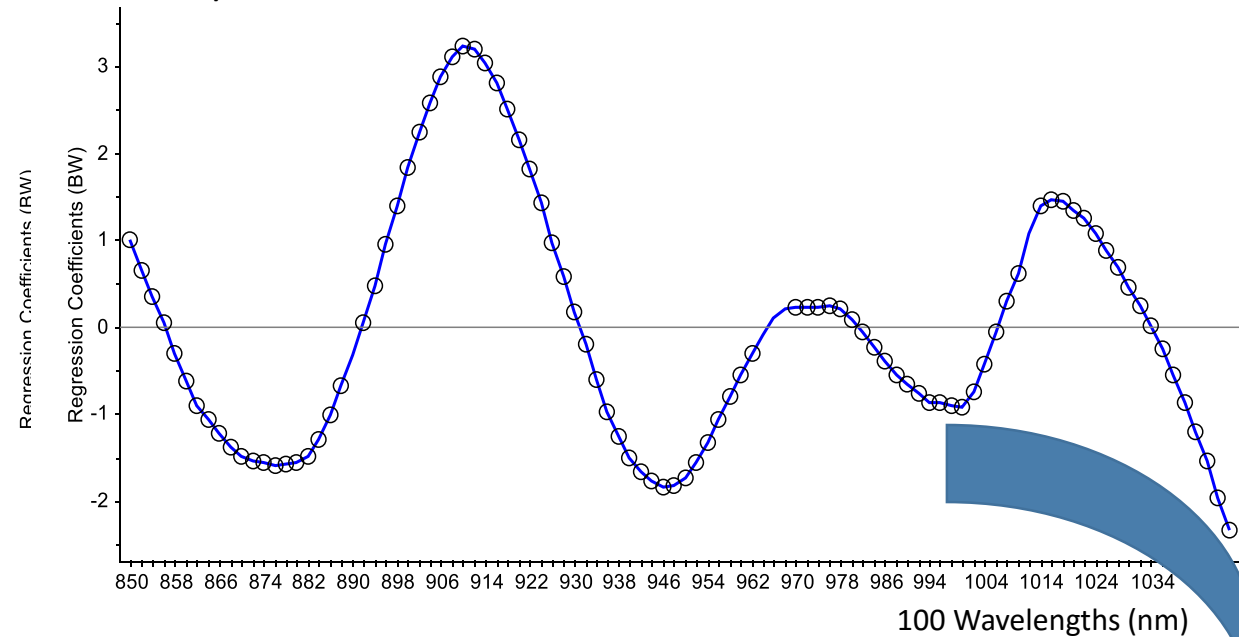


### 3: Mektig matte uten tårer

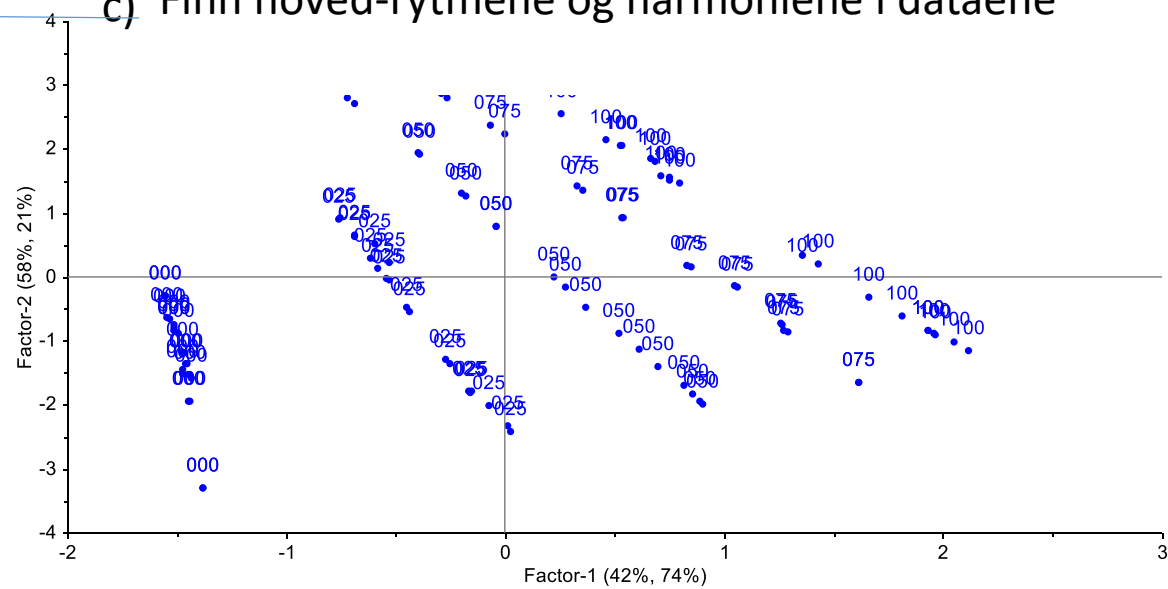
a) NIR lysspektra av mange melprøver



b) Automatisk matematisk modell av kvalitets-variasjon

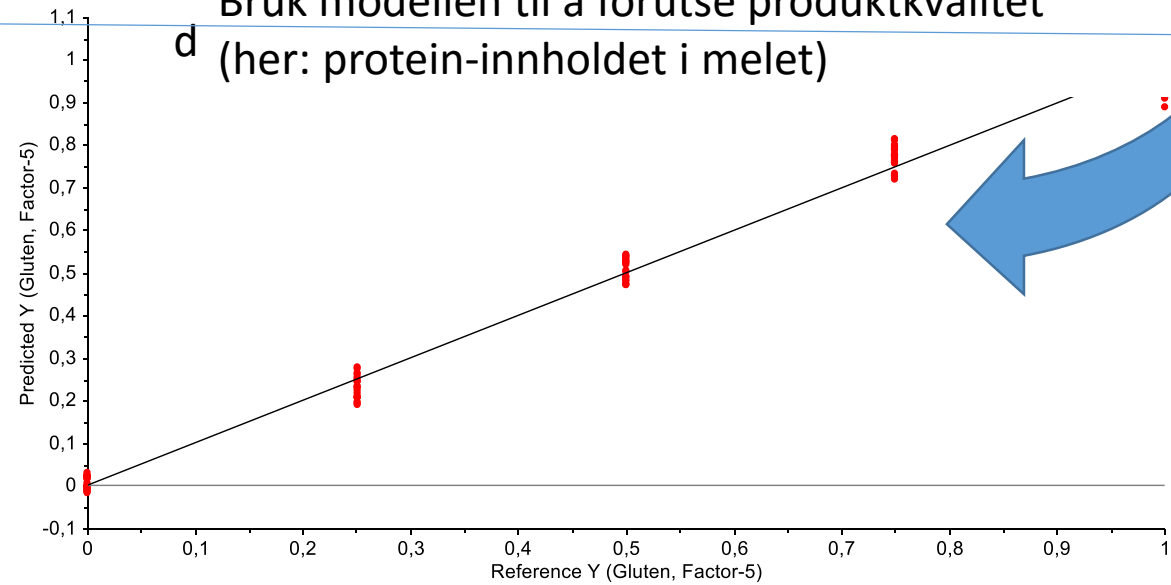


c) Finn hoved-rytmene og harmoniene i dataene



Bruk modellen til å forutse produktkvalitet

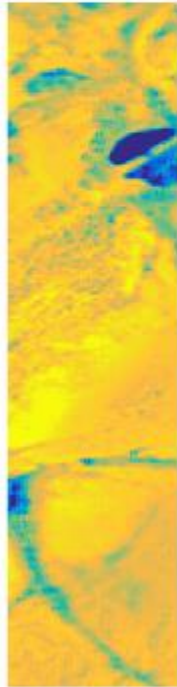
d (her: protein-innholdet i melet)



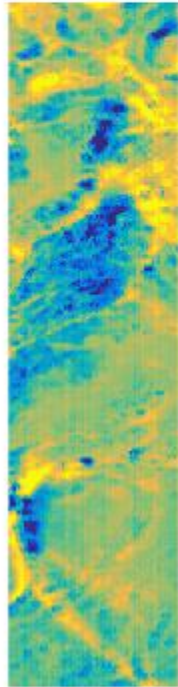
# Near-infrared light, many wavelengths:

Giving the surgeons a 'molecular eye-sight'

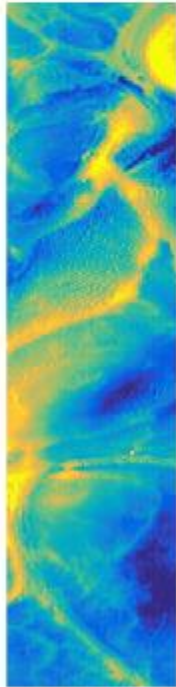
ICA Score 1



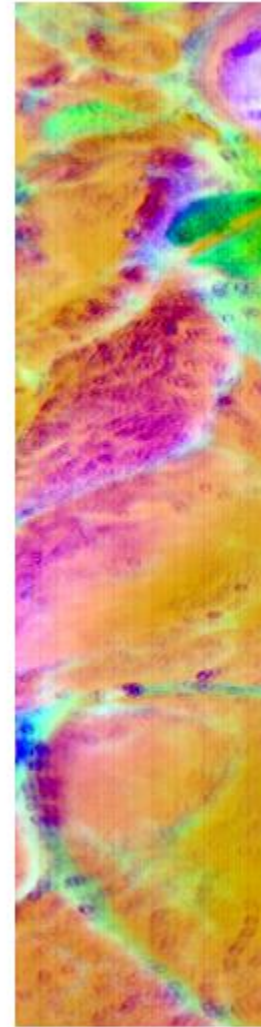
ICA Score 2



ICA Score 3



ICA Scores 1-3



# Rhythm & Blues i BIG DATA: *Matematikk uten tårer*

1: Verden har mønstre – finn dem

2: Utvid sansene

3: Mektig matte uten tårer

# Mektig men enkel matte:

- $21 = 3 \times 7$

- $21.1 = 3 \times 7 + 0.1$

- $a = b \times c + d$  Vector algebra, first published by Norwegian Caspar Wessel 1797

# Mektig men enkel matte:

- $21 = 3 \times 7$

- $21.1 = 3 \times 7 + 0.1$

- $a = b \times c + d$

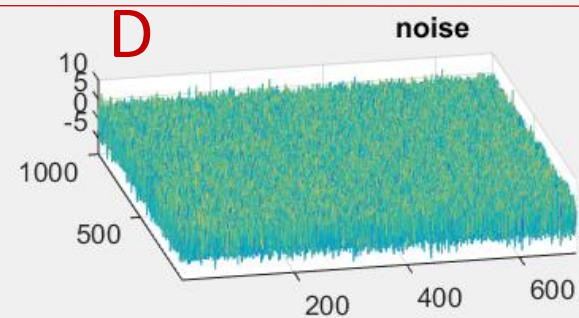
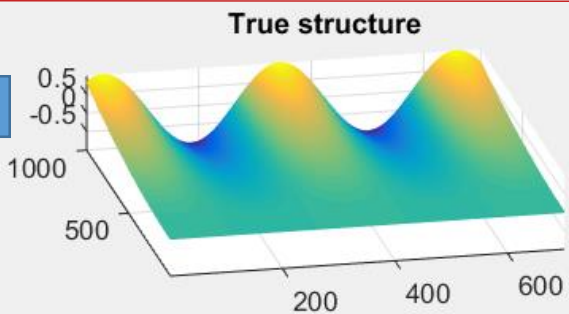
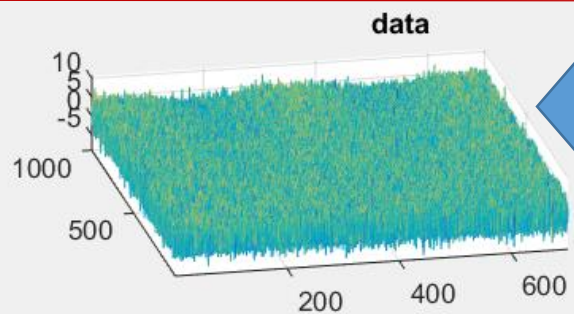
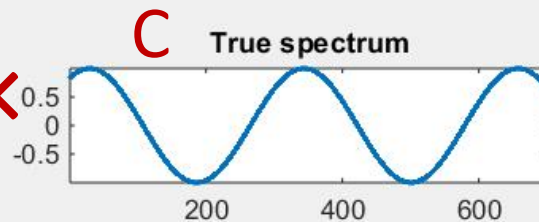
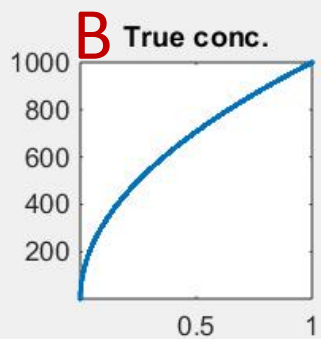
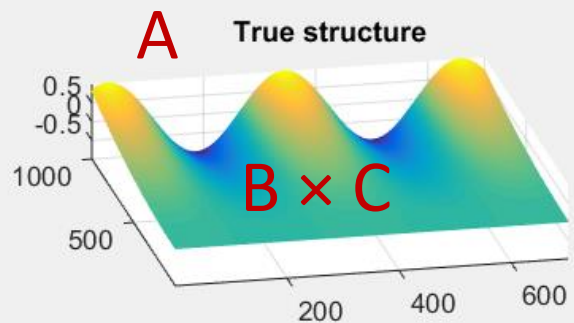
Vector algebra, first published by Norwegian Caspar Wessel 1797

- $\mathbf{A} = \mathbf{B} \times \mathbf{C} + \mathbf{D}$

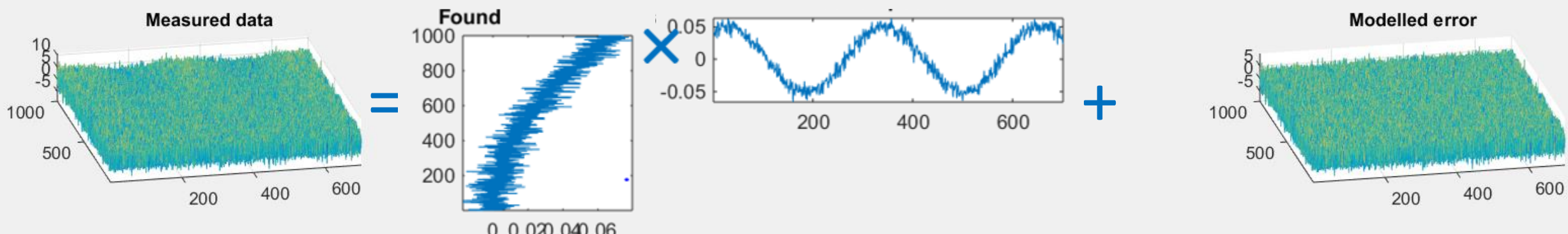
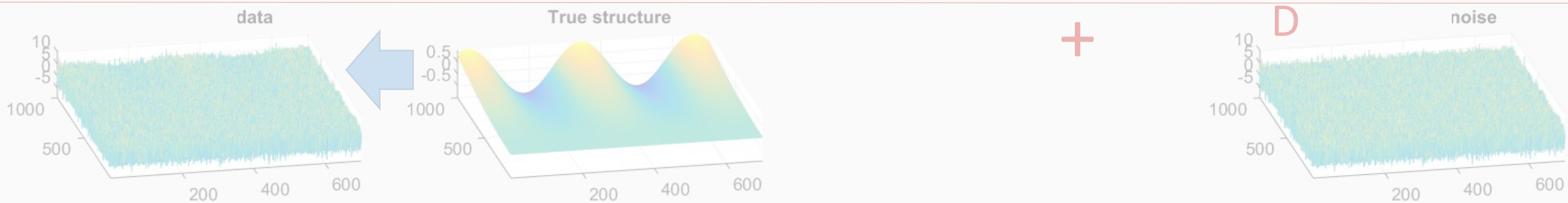
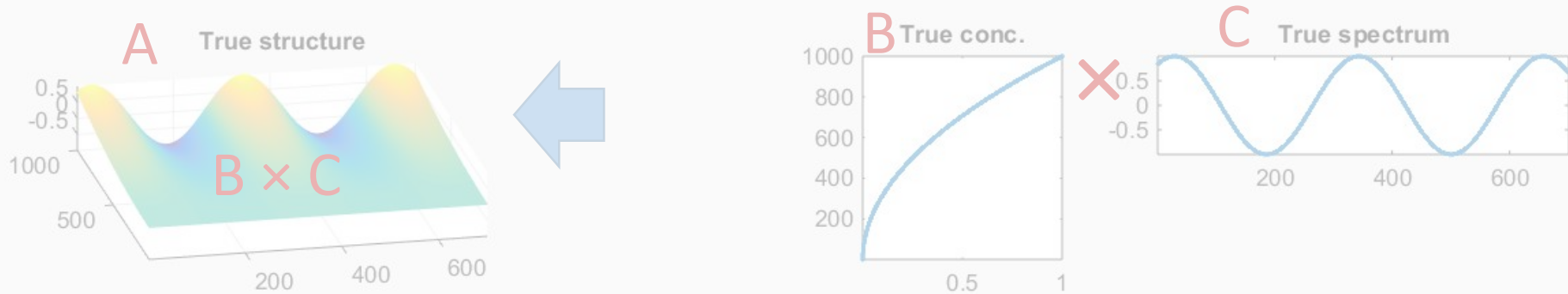
Fra data-tabell  $\mathbf{A}$ , finn strukturen  $\mathbf{B} \times \mathbf{C}$  og residual-tabellen  $\mathbf{D}$

Prinsipal komponent-analyse (PCA):

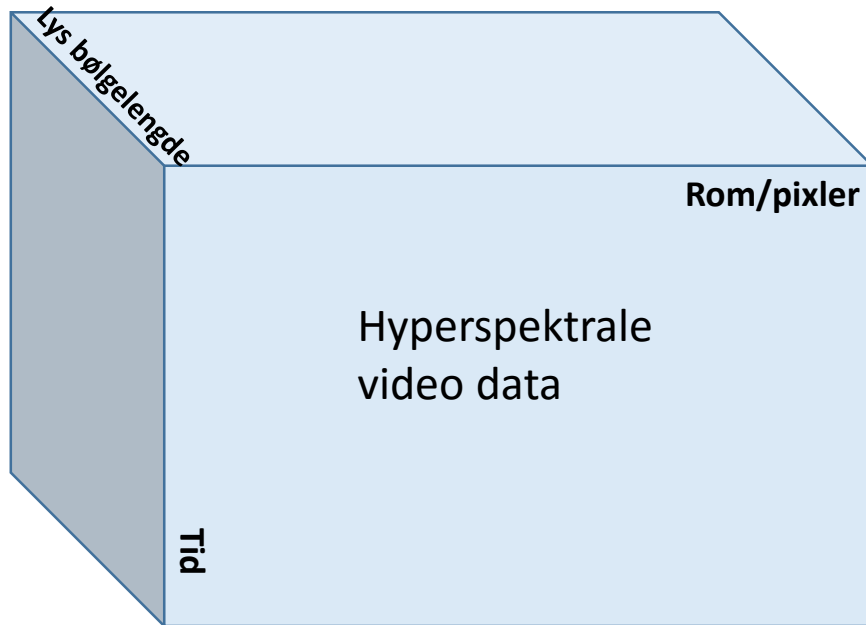
Alle multivariate metoders moder!







# BIG DATA fra moderne måleinstrumenter, f.eks. hyperspektrale kameraer



**Kommer nå, innen f.eks.**

*Medisin*

*Miljøovervåking*

*Ressurs-undersøkelser*

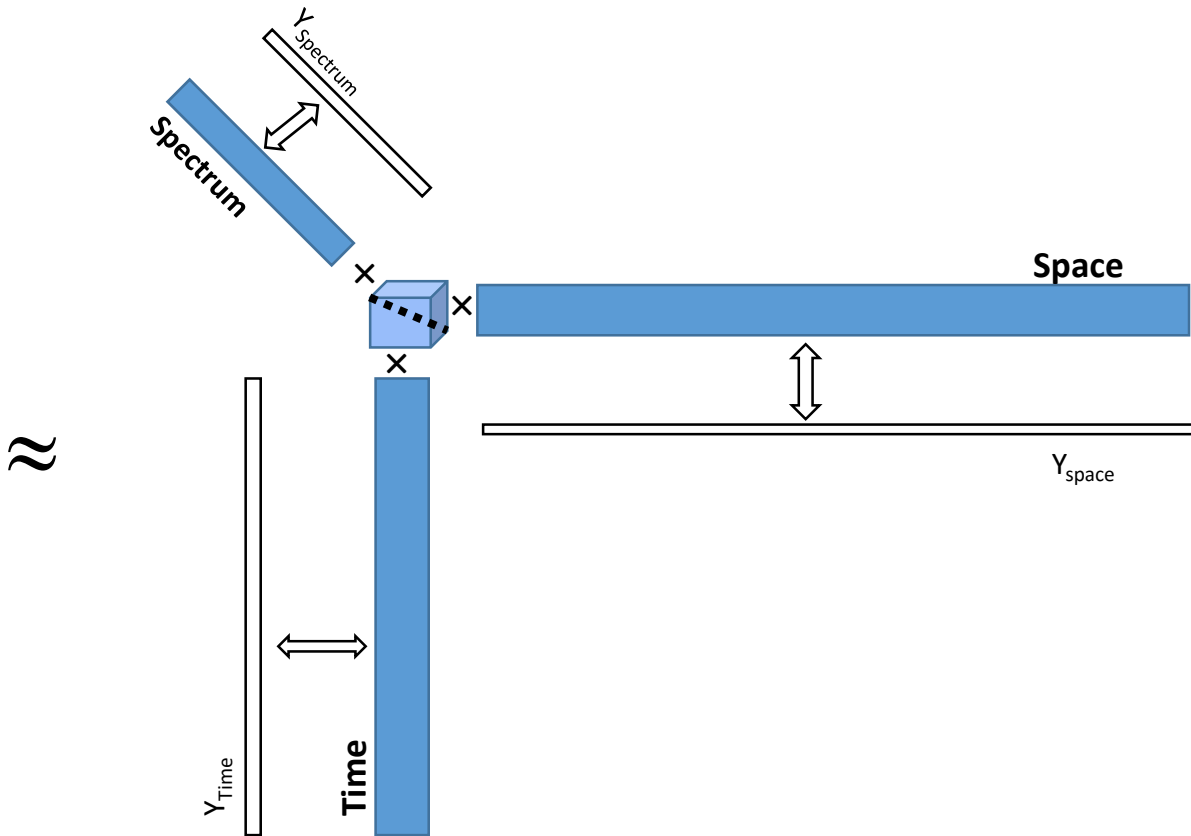
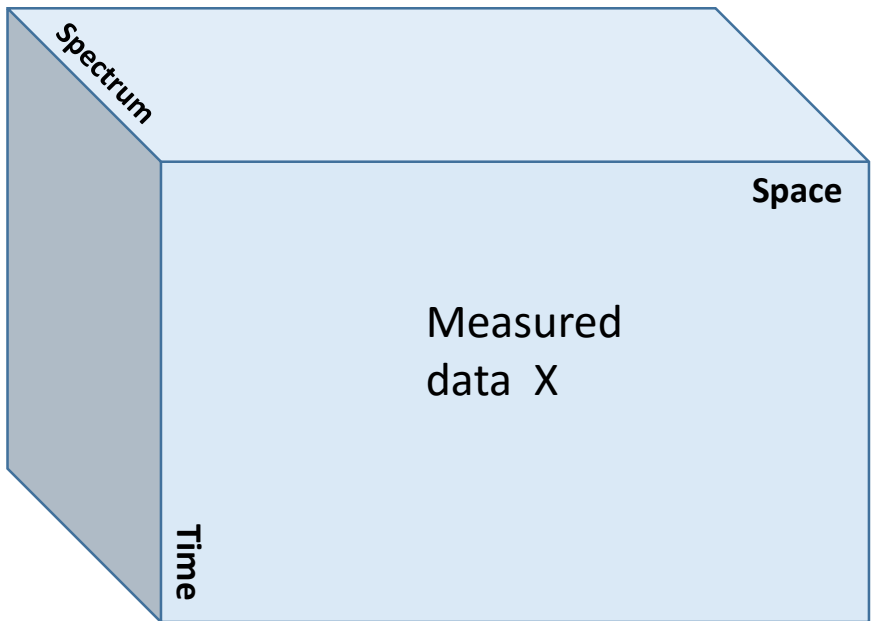
*Prosessindustri*

*Kommunikasjon*

*Forsvar*

*Romfart*

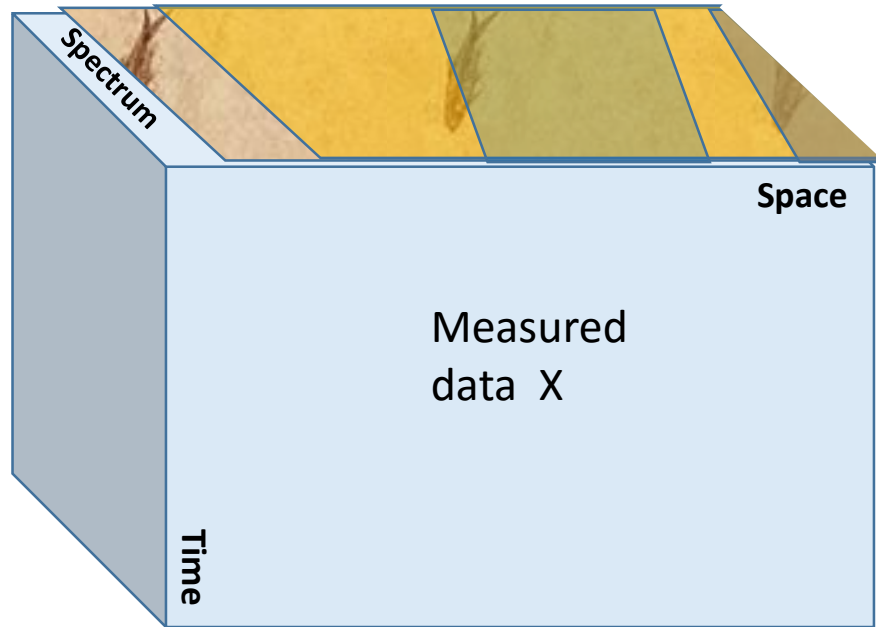
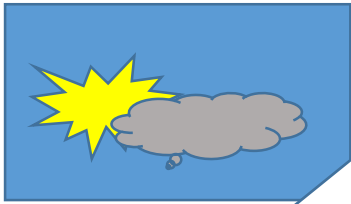
# Tomorrow's medical measurements, and how to use these:



Prototype of future's highdimensional spatiotemporal instrumentation:  
**Hyperspectral NIR video**  
1-2 GB of data/hr

Figure 7

# Hyperspectral NIR video: 2-way PCA-based deshadowing



Prototype of future's highdimensional spatiotemporal instrumentation:

**Hyperspectral NIR video**

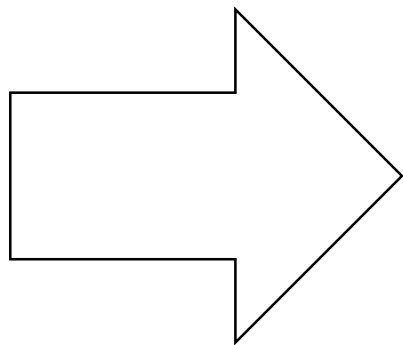
1-2 GB of data/hr

Raw



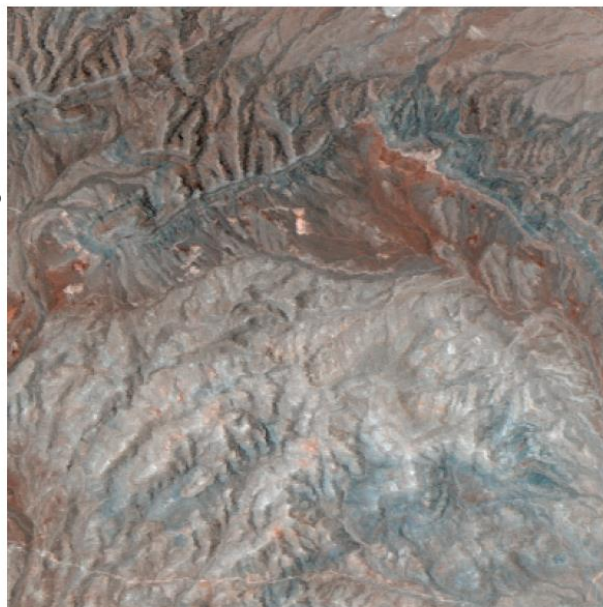
Data borrowed from NASA,

«Deshadowed» images

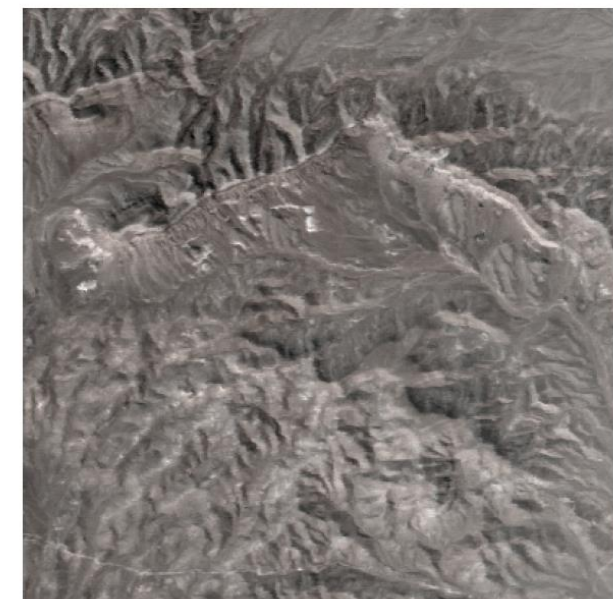
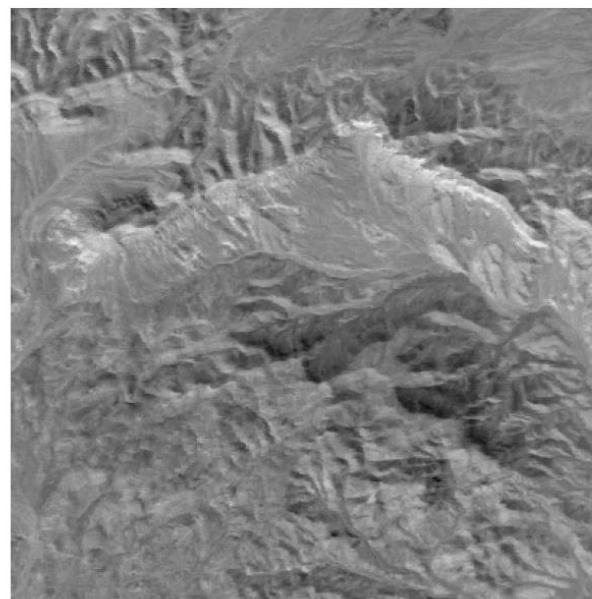
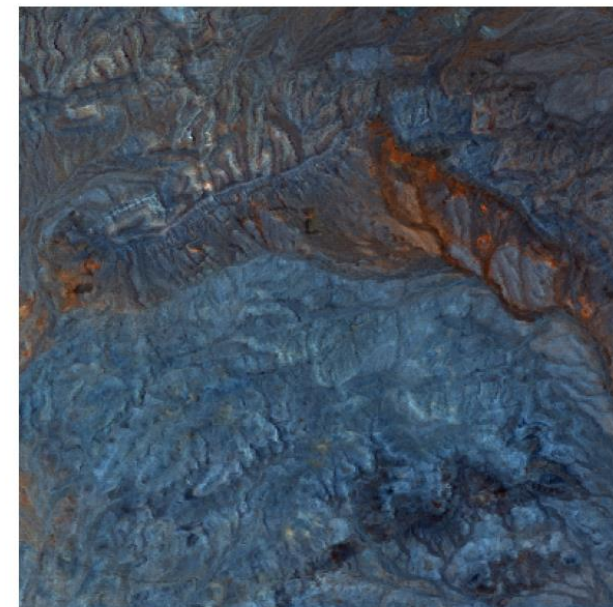


Illumination changes

Traditional



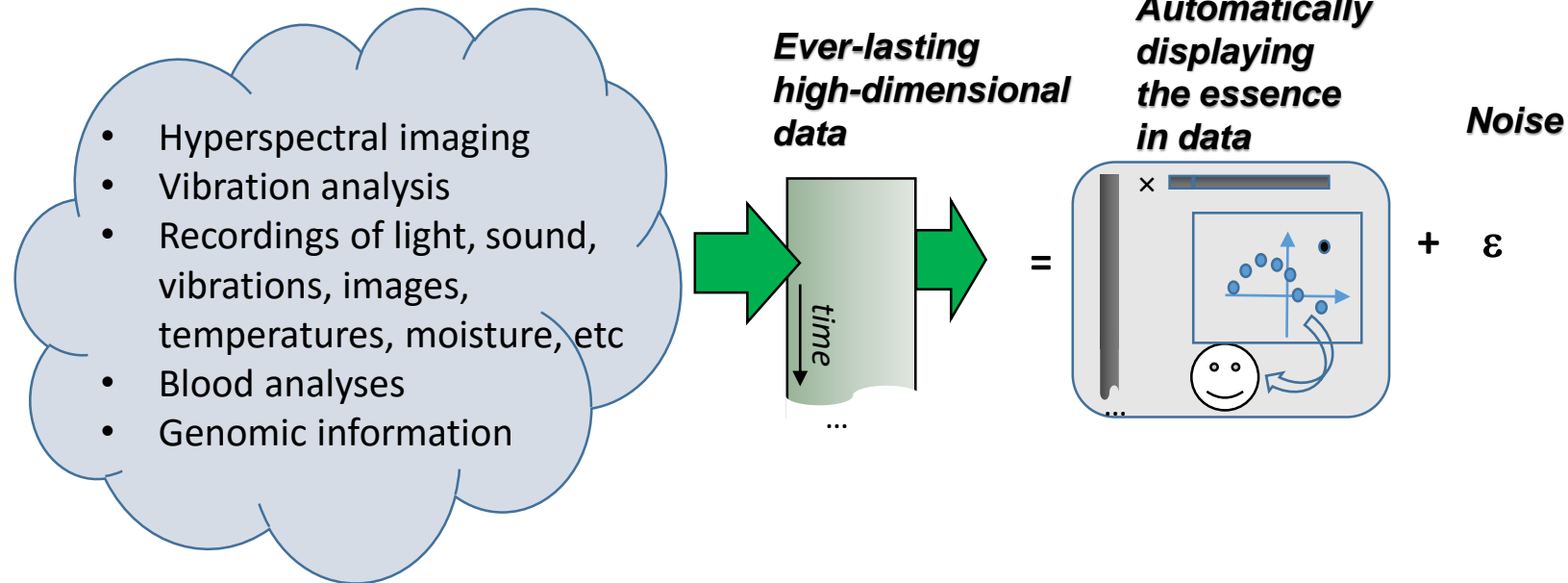
Our



# On-the-fly processing of high-dimensional data streams

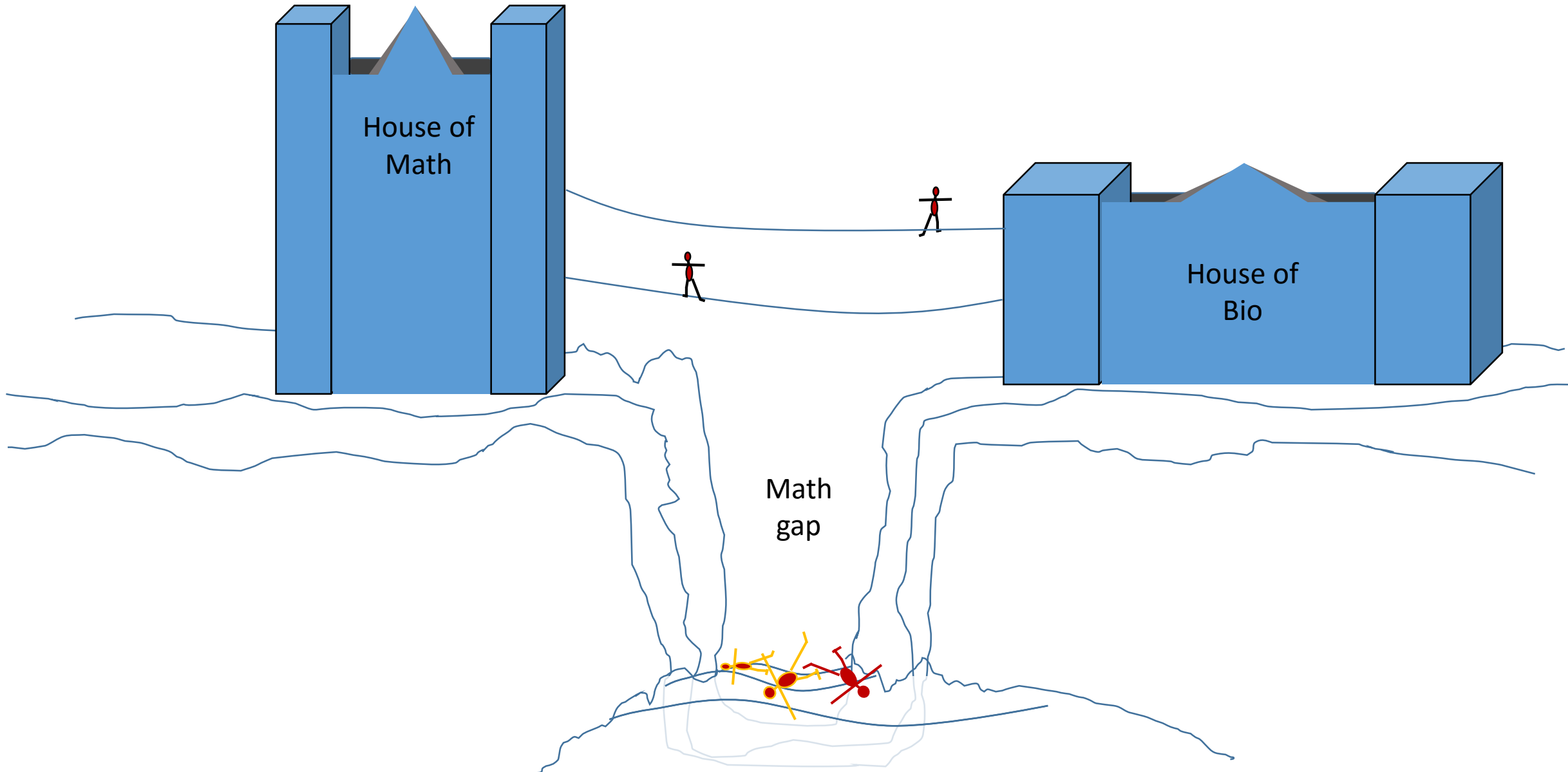
New tools for analysis and compression of streams of multidimensional, dynamic data.

Harald Martens, Jan Otto Reberg (Idletechs AS, Perleporten/NTNU)

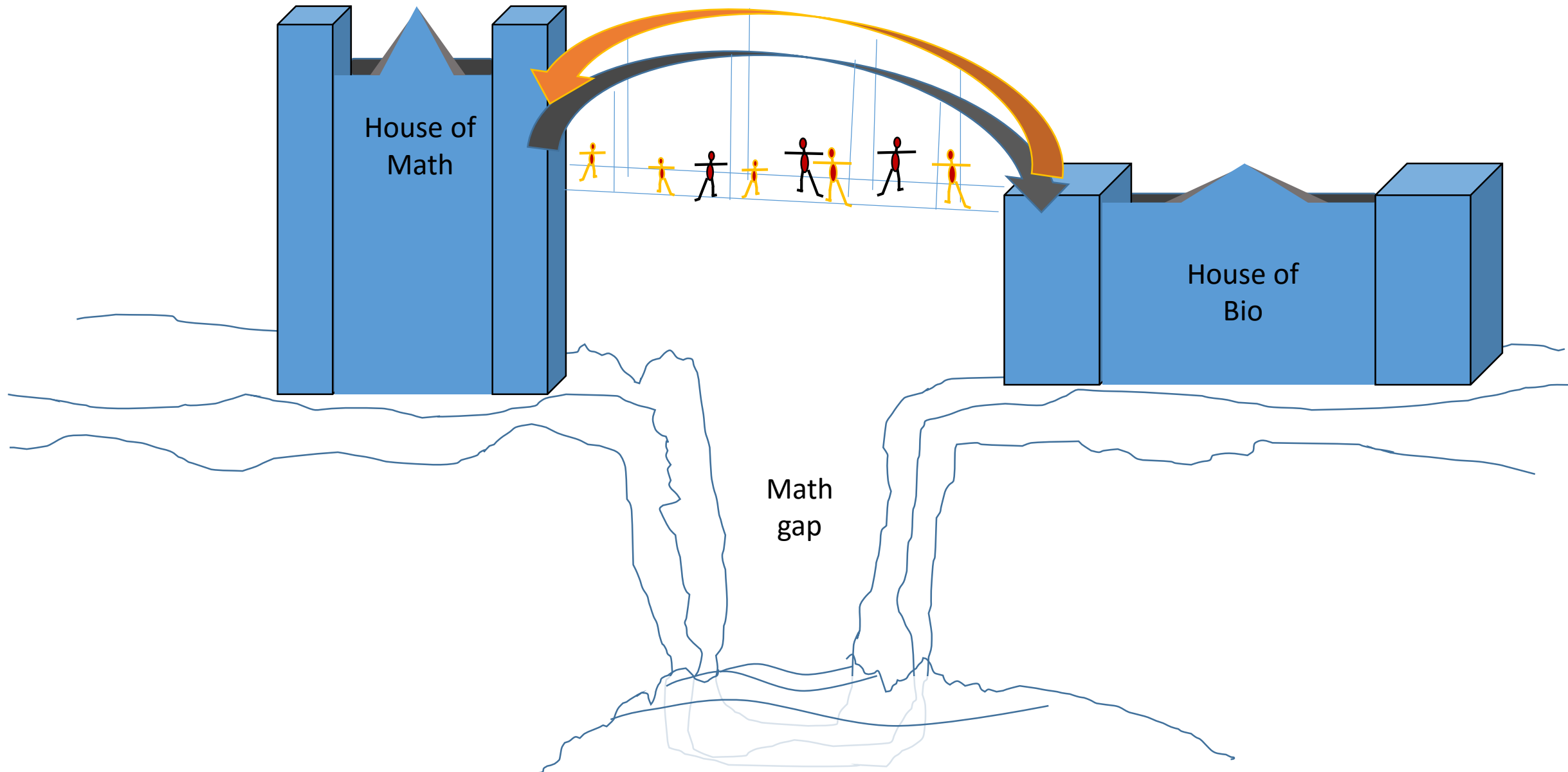


From raw streams of data, systematic patterns and relationships are automatically discovered and modelled. The data is stored in a highly compressed format.

# The Math Gap



# A two-way bridge across the math gap





*Rhythm & Blues fra data :*  
*Matematikk uten tårer*

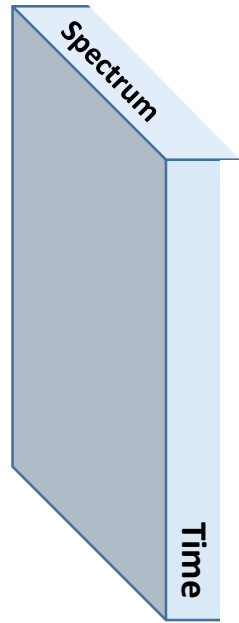
1: Verden har mønstre – finn dem

2: Utvid sansene

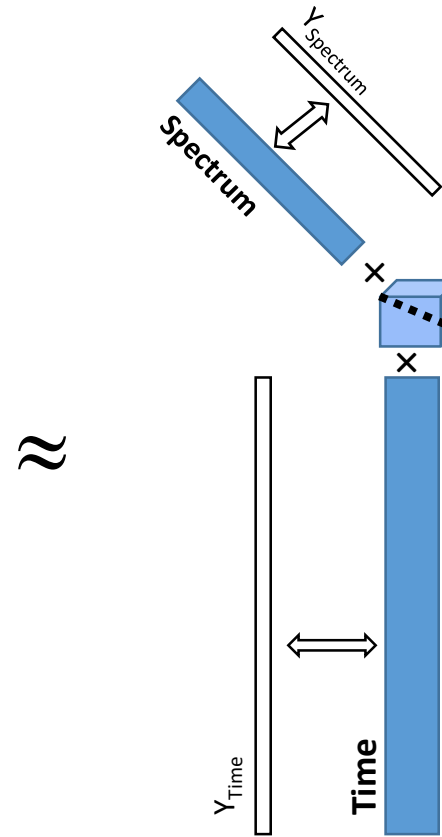
3: Mektig matte uten tårer

***Takk !***

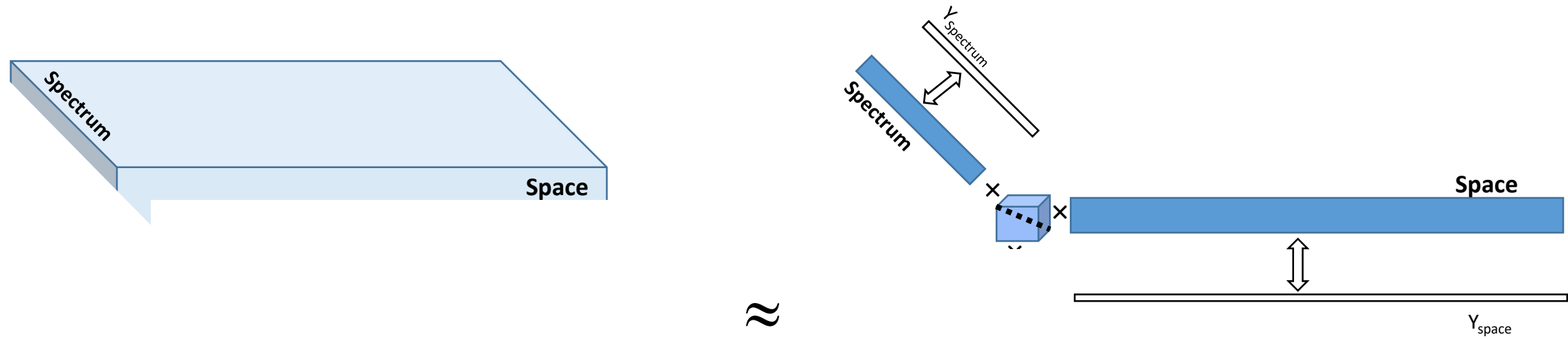
# Multichannel NIR spectroscopy: Bilinear modelling by PCA and PLSR



Multichannel NIR spectroscopy



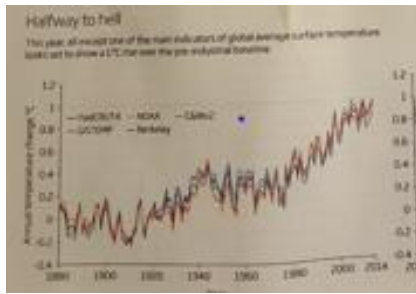
# Hyperspectral NIR imaging: Bilinear modelling by PCA, ICA and PLSR



Hyperspectral NIR imaging

- **Høydimensjonale datastrømmer i helsesektoren:**
- **Ikke Gucci-statistikk, ikke Macho-matematikk, ikke Blind informatikk, men heller Rhythm & Blues fra data**
- Harald Martens<sup>1,2</sup>, Jan Otto Reberg<sup>2</sup>, Joa Fortuna<sup>1</sup> og Nils Kristian Skjærvold<sup>3</sup>
- <sup>1</sup>Inst. teknisk kybernetikk, IME, NTNU, <sup>2</sup>Idletechs AS, <sup>3</sup>Inst. sirkulasjon og bildediagnostikk, Med.Fak, NTNU, Trondheim.
- Mengden medisinske måledata vokser nå så raskt at man mister oversikten, med mindre man bruker matematiske, statistiske og beregningstekniske oppsummeringer av slike data. Det fører til for mange og for ufølsomme alarmer, og for dårlig oversikt for legen / forskeren. Dette kan forbedres, ved å kombinere forkunnskaper og høydimensjonale data på en enkel, forståelig måte.
- Kroppen og enkeltorganene har systematiske variasjons-*mønstre*. Noen av disse mønstrene reflekterer fysiologiske mekanismer som man allerede forstår, mens andre kan reflektere viktige, men ukjente, eller iallfall uventede variasjonsfenomener— enten innen en gruppe pasienter, i en enkelt pasient eller fra måleinstrumentet selv.
- **Kjemometri, dynametri og Big Data Cybernetics:** Dette foredraget forklarer en enkel, generisk måte «eieren» av måledata kan oppdage, komprimere, tolke, teste og bruke informasjonsinnholdet i store biomedisinske datastrømmer, f.eks. fra intensiv-overvåking, medisinsk billedanalyse, metabolomikk osv. Metodikken er en kombinasjon av teknikker og konsepter fra ulike datamodellerings-kulturer.
- **Gucci-statistikk:** Det er vanlig å tolke medisinske måledata i lys av faglige hypoteser. Det kan være fornuftig, for å gardere mot ønsketenkning og tilfeldigheter. Man vil gjerne ha  $p < 0.05$ .
- Men når signifikanstesting primært brukes for det tar seg pent ut, ( $p = \text{«Permission to Publish»}$ ), er det en fordømmende praksis. Det er også problemer med overforenklete dikomier (ja/nei, syk/frisk osv),- virkeligheten er ikke så enkel.
- **Macho-matematikk:** Måledata kan oppsummeres i form av mekanistiske matematiske modeller. Dette kan gi en kompakt og oversiktlig kvantitativ representasjon av slik kunnskap. Når slike modeller tilpasses til måledata, kan man få en faglig god måte å tolke måledata i lys av dagens teoretiske forståelse.
- Men en feilaktig, eller grovt overforenklet mekanistisk modell vil ofte gi feil tolkning av data. Insisterer man likevel på å bruke en slik utilfredsstillende modell, fungerer modellene som et fengsel for tanken.
- **Blind informatikk:** Maskinlæring basert på f.eks. kunstig nevronett, kan automatisk oppdage både kjente og ukjente variasjonsfenomen, og kombinere dem til å gjøre klassifikasjoner og prediksjoner.
- Men den automatisk utviklede formel-løsningen fra en slik Black-box datamodellering er prinsipielt vanskelig å fremstille grafisk for tolkning i lys av pasientens fysiologi og behandlingshistorikk. Informatikk-kulturen - teknologiens svar på postmodernismen, har en tendens til å nedtone behovet for tolkning i lys av bakenforliggende naturvitenskap.

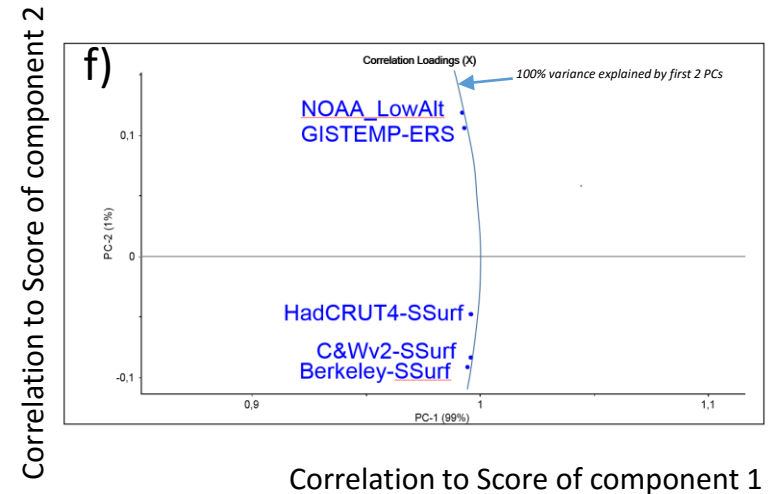
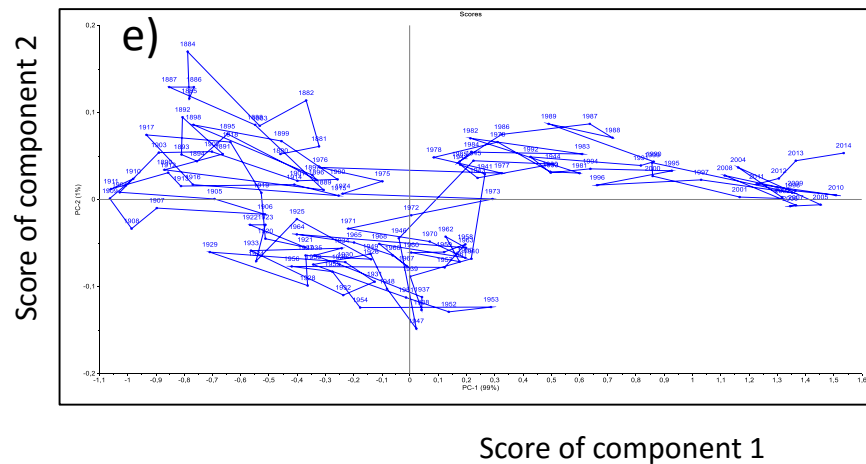
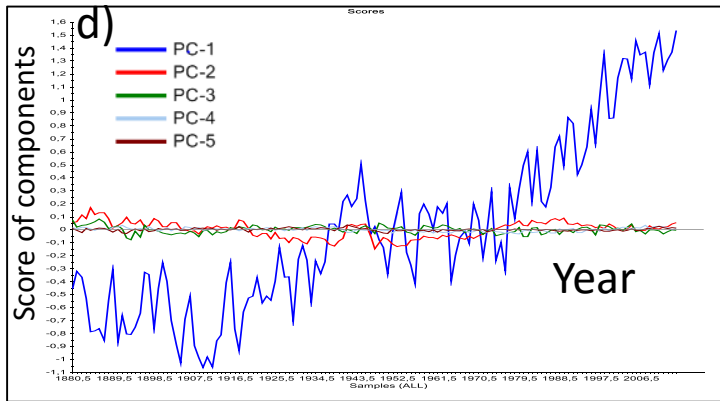
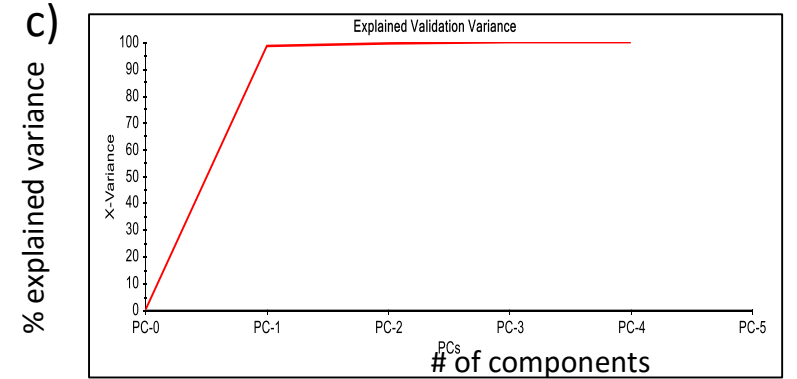
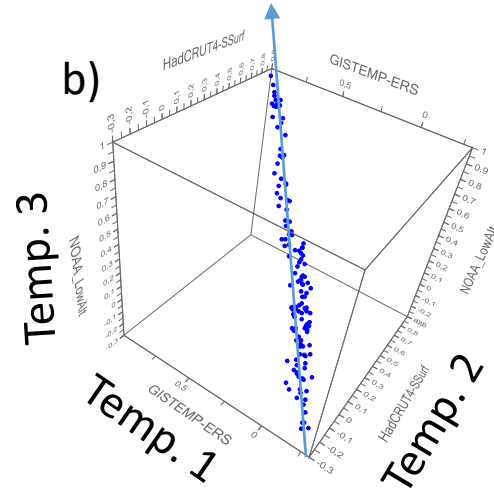
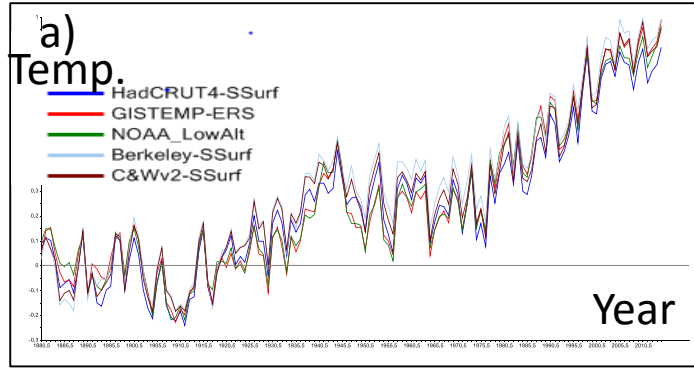
- **Sesjon 4A: Individualiserbar kunnskap 1230:1400 28.9**
- Sesjonsleder: Laura Slaughter
- Helseframtiden er retningslinjerik! EPJ-systemer er lett å integrere! Det er mye å lære av helsedata! Alt er mulig med IT! Eller kanskje vi trenger både forskning, fagutvikling og tung profesjonalitet for å realisere våre Helseinformatiske håp?



New Scientist, August 2015

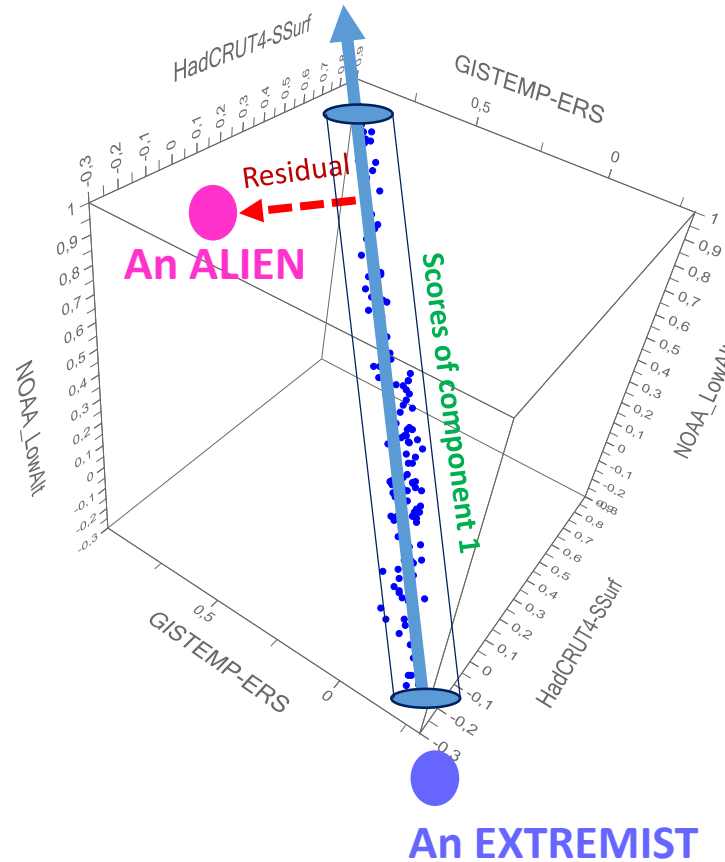
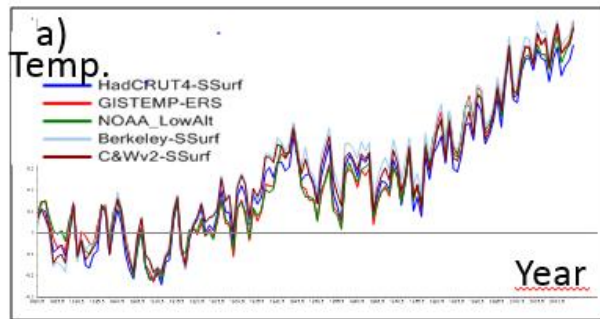
# Global average temperature: different research groups.

Data driven multivariate modelling will also reveal *unexpected patterns*

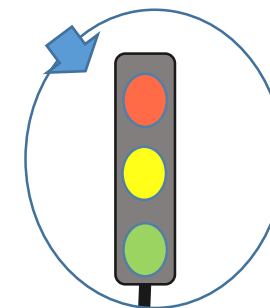
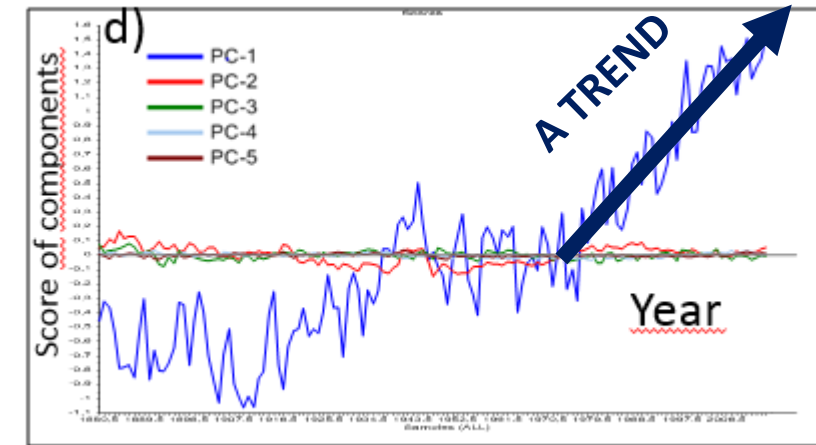


# Multivariate soft data-modelling gives fewer, but more sensitive alarms, for three different types of abnormalities:

Raw data  
Many variables (temperatures)



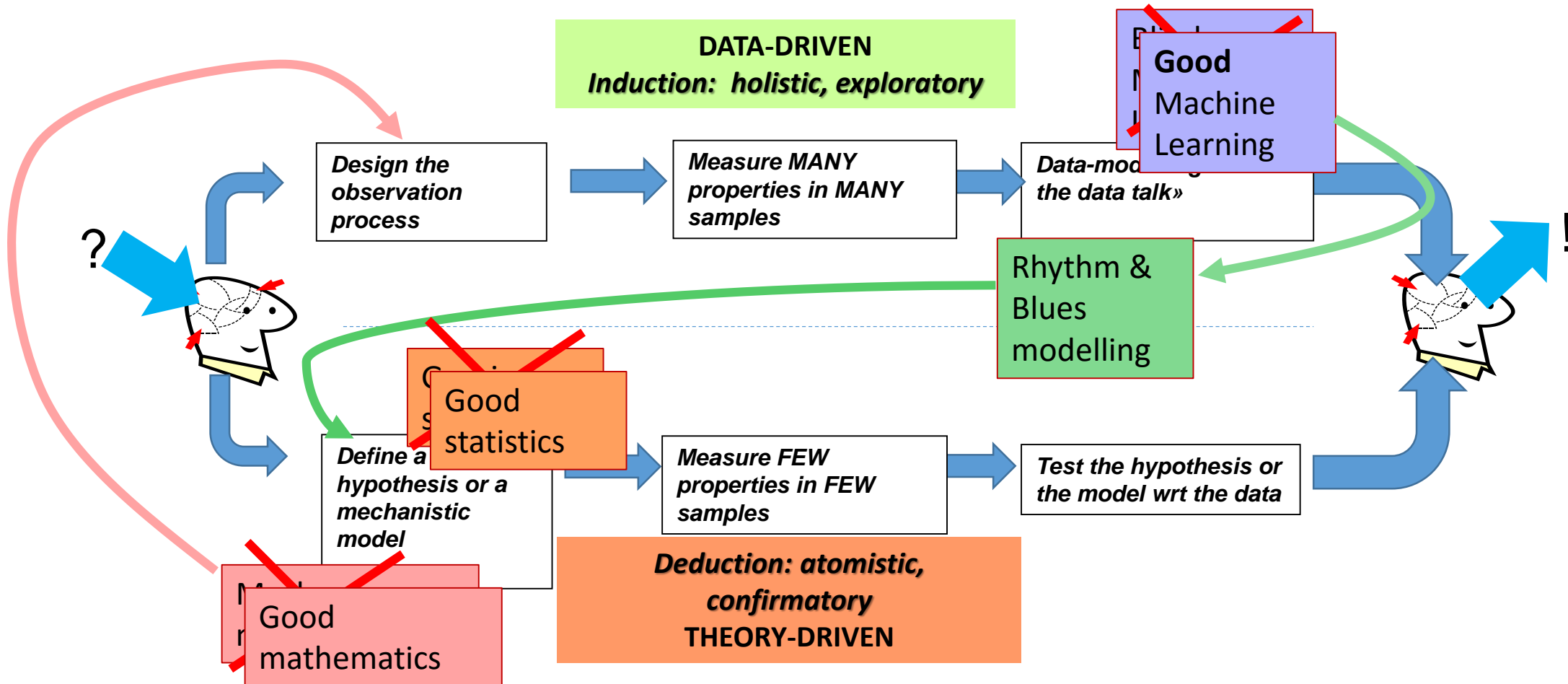
Few informative combinations of the many variables



Færre, bedre alarmer!



# Multivariat Myk META-Modelling:



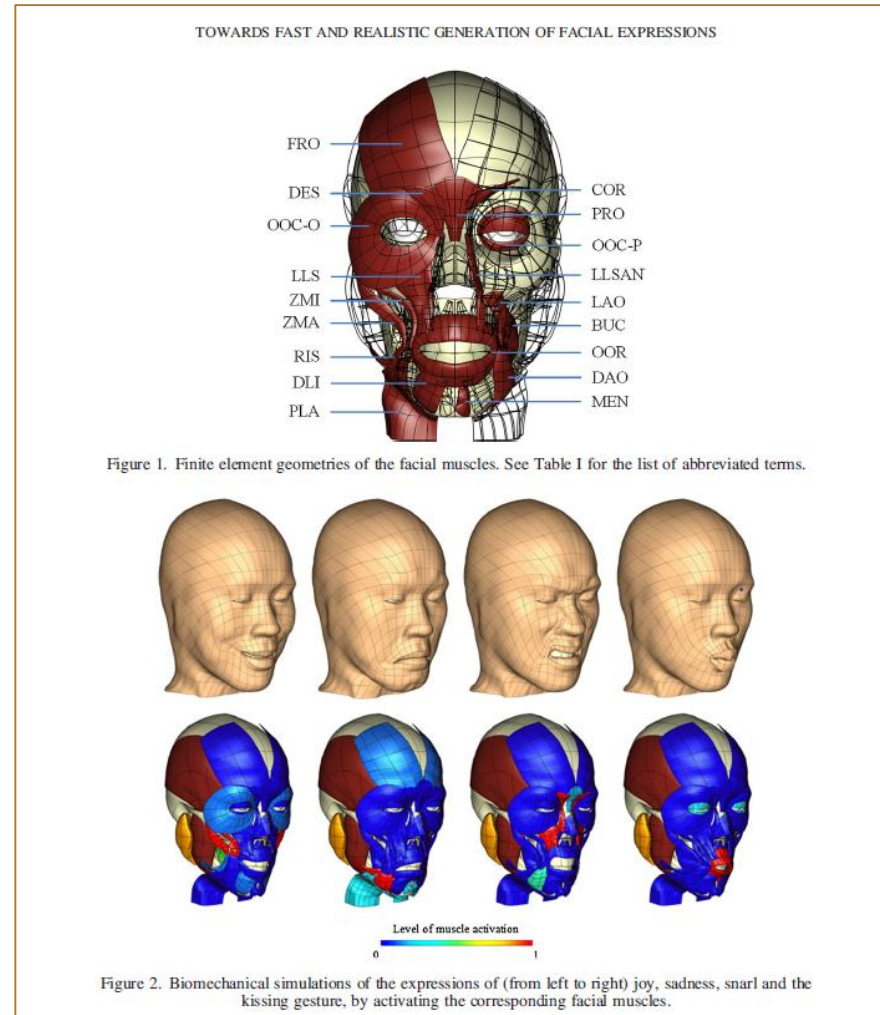
1: Verden har mønstre – finn dem

# Multivariat metamodellering for enklere beskrivelse av menneskets bevegelser

Eksempel: Speed-up av matematisk formulering av ansikts-stemninger

CPU beregningstid per ansiktsuttrykk:

> 2 timer  $\Rightarrow$  < 0.01 sekund



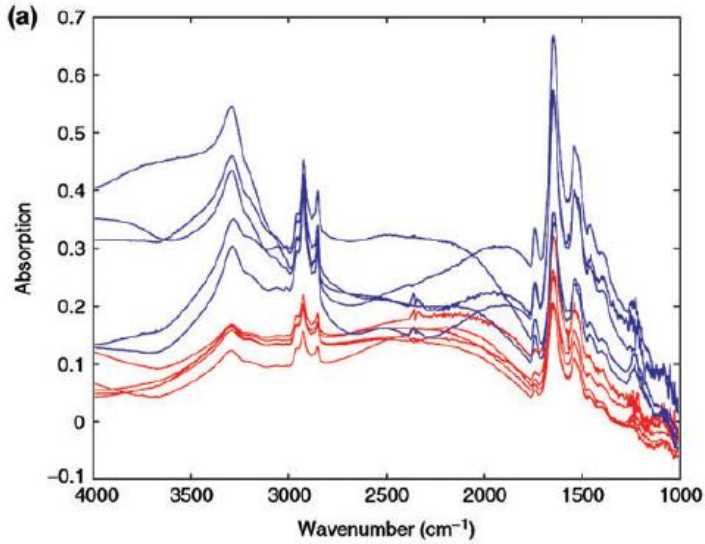
INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN BIOMEDICAL ENGINEERING  
*Int. J. Numer. Meth. Biomed. Engng.* (2014)  
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/cnm.2646

Emulating facial biomechanics using multivariate partial least squares surrogate models

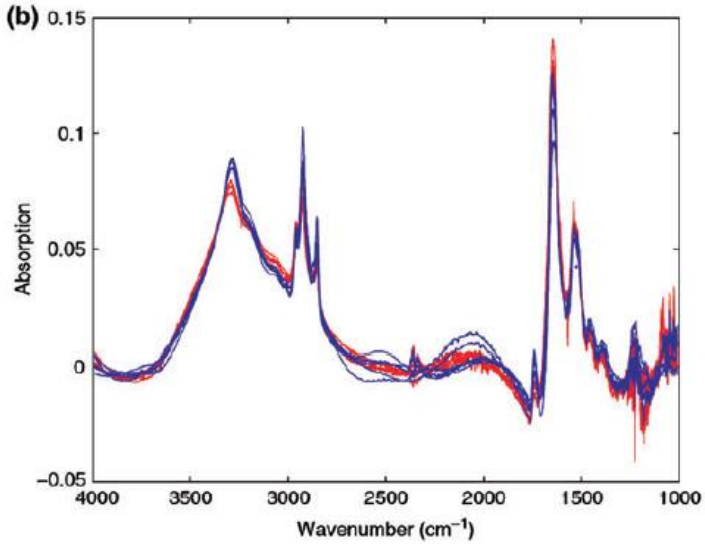
Tim Wu<sup>1,\*</sup>, Harald Martens<sup>2</sup>, Peter Hunter<sup>1</sup> and Kumar Mithratne<sup>1</sup>

# BIO-SPEKTROSKOPI: Mål intakte prostata kreft-celler med Infrarødt lys ved mange bølgelengder

## Bruk multivariat kalibrering til å rense opp dataene matematisk



Matematisk forbehandling av lysmålings-spektrene



Kreft-diagnose, utvikling av bedre kreftmedisin

m/ Achim Kohler et al., NMBU 2014