"*How can we reach out to the **creative entrepreneur types** with our statistics, mathematics and physics education?*"

Harald Martens, dr.techn.

Prof. II, Dept. Engineering Cybernetics, NTNU Trondheim

CEO IDLETechs AS

Biggest science scandal in the last 30 years?

Innsikt 35

Aftenposten Torsdag 5. mars 2015

Aftenposten Torsdag 5. mars 2015

Norske forskere er for dårlige i statistikk

Viten-debatt:

Sigrid Bratlie Thoresen postdoktor, Institutt for kreftforskning, Radiumhospitalet

Det er feil i en mye større andel av vitenskapelige arbeider enn man tidligere har vært klar over.

Professor John Ioannidis ved Stanford-universitetet skapte furore i 2005 da han publiserte en artikkel som proklamerte at brorparten av medisinske forskningsresultater er feil. Med tiden er dette problemet blitt stadig tydeligere, blant annet gjennom studier som viser at kun en brøkdel av forskningsfunn kan gjentas med samme resultat.

Da forskere ved det amerikanske legemiddelfirmaet Amgen for få år siden forsøkte å gjenskape resultatene fra 53 publiserte studier de mente var landemerker innen grunn-

dermed som «statistisk signifikant».

På bakgrunn av dette vil man altså anta at det sannsynligvis er en sammenheng mellom vaksinasjon og færre tilfeller av meslinger. Dette utelukker på ingen måte at observasjonen faktisk er tilfeldig. Av 100 tenkte studier vil fem av dem, med en grenseverdi for p på 0,05, gi et statistisk signifikant resultat selv om det egentlig

vann dør. Betyr det at drikkevann er en ledende dødsårsak? Nei, selvfølgelig ikke. Korrelasjon, altså statistisk sammenheng, betyr ikke nødvendigvis at det ene fenomenet skyldes det andre, altså kausalitet. Og slike feilslutninger dukker opp rett som det er i forskningsartikler.

For eksempel konkluderte forfatterne av en Nature-artikkel i 1999 med at kunstig belys-

Norwegian researchers are too bad in statistics

«There are errors in a much higher fraction of scientific works than known earlier»

38

AKTUELT FORSKNING 27. mars–2. april 2015 | MORGENBLADET

Vitenskapens pinlige sykdommer

TALLKNUSING *De fleste publiserte forskningsfunn er feil, og 85 prosent av pengene som går til biomedisinsk forskning, er bortkastet, sier John Ioannidis.*

JON KÅRE TIME

Vi må omregulere insentivsystemene slik at de sikter seg inn mot å få troverdige resultater og betydningsfulle oppdagelser snarere enn bare flere artikler og mer prosjektstøtte, sier John Ioannidis, professor ved Stanford.

«Superhelten» som skal redde medisinsk forskning – slik er han blitt beskrevet. Tallknuseren fra Hellas leder i dag Metrics – et senter for «meta-forskning» som har som mål å gjøre vitenskapen verden over både bedre og mer effektiv. Han er en sentral skikkelse i arbeidet for å rette opp i systematiske skjevheter innen vitenskapelig publise-

MORGENBLADET 20. MARS

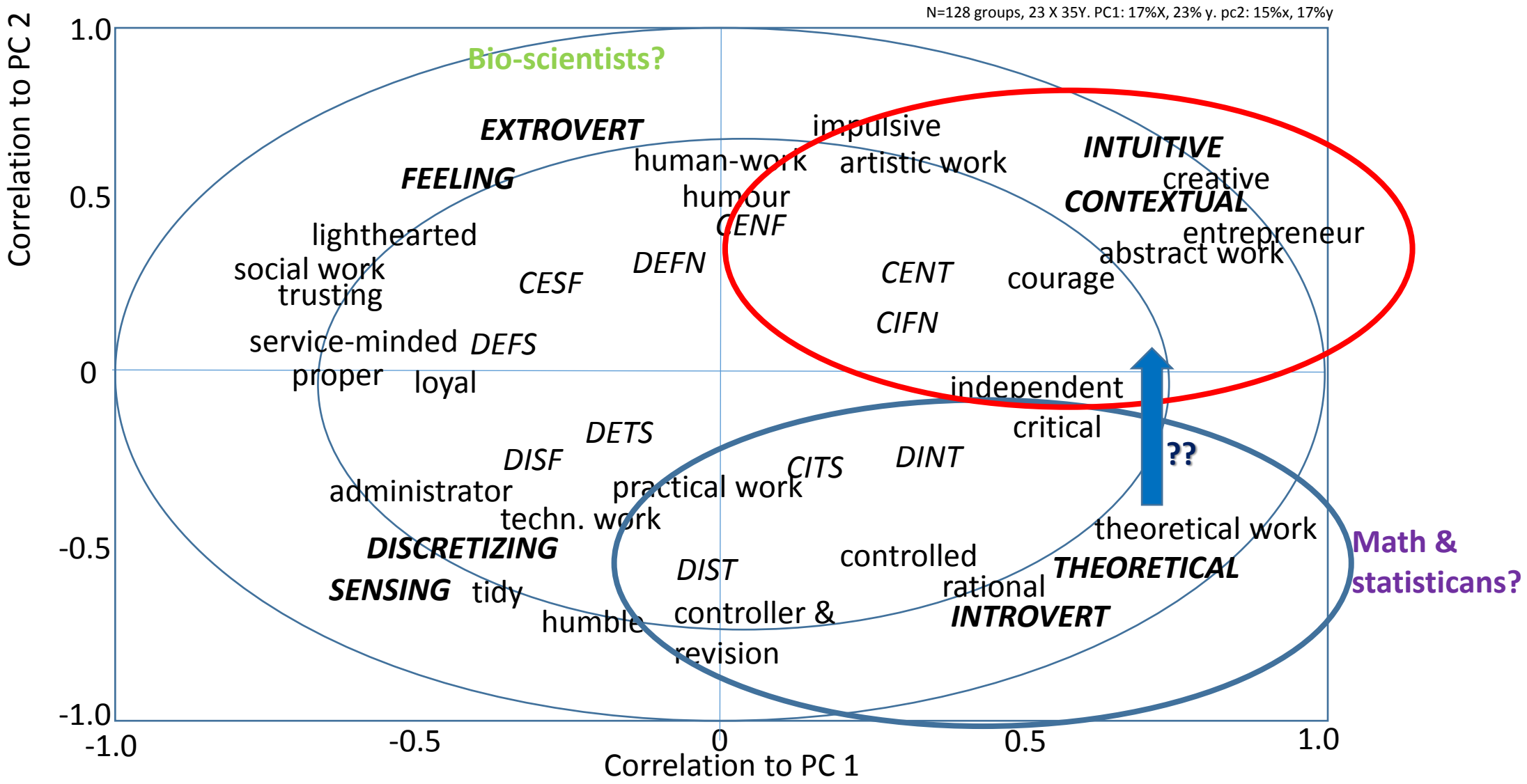
AKADEMISK ALARM

Publiseringspress fører til en opphopning av dårlig forskning, advarte professor Erik Boye i Morgenbladet forrige uke. Han fikk støtte av nobelprisvinner Edvard Moser.

Science's embarassing illnesses

«Most of published research findings are wrong, and 85% of the money to biomedical research is wasted, says John Ioannidis»

Personality types and work types are correlated



Ontology: position and intensity variation in time, space and properties

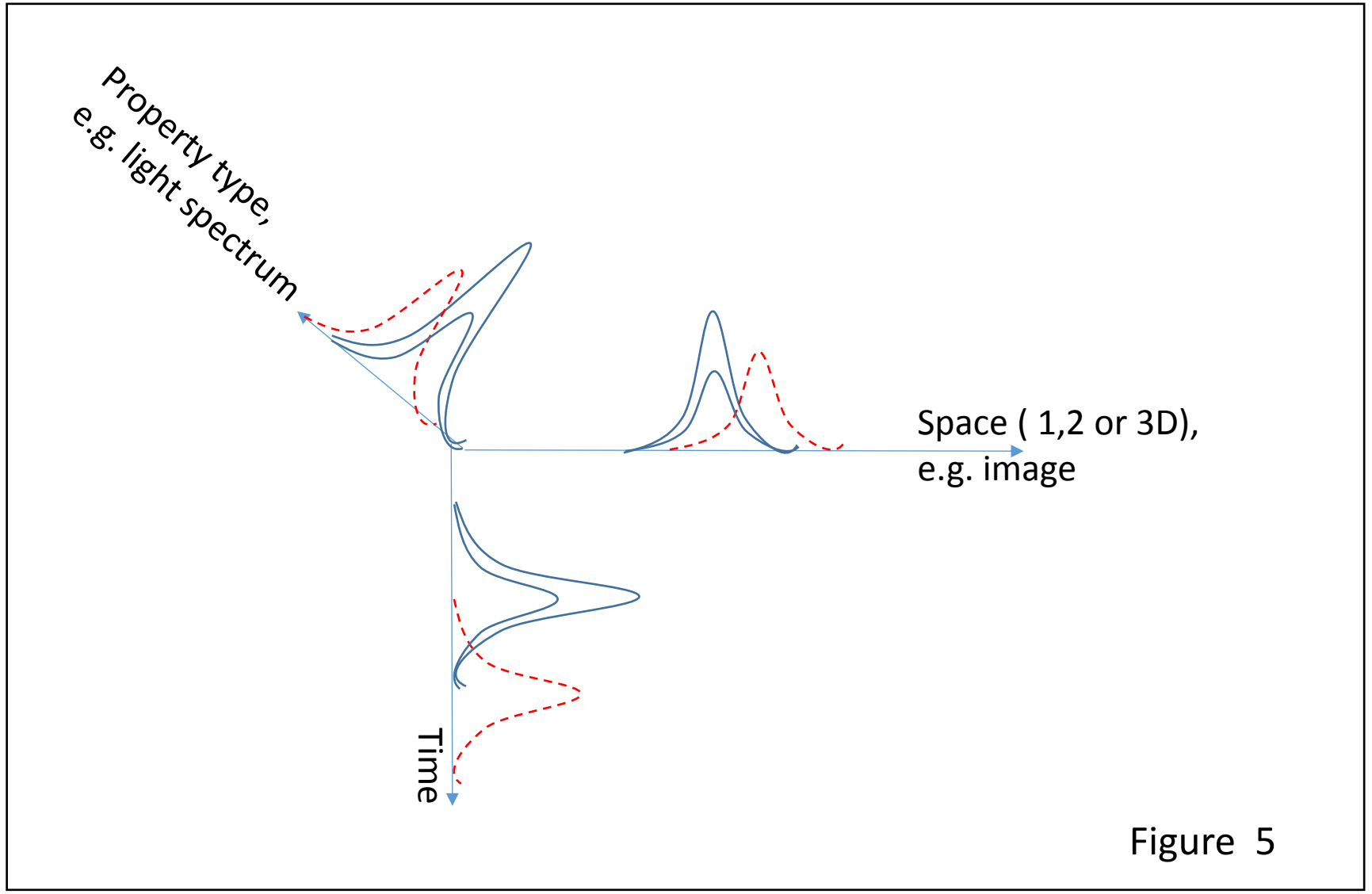


Figure 5

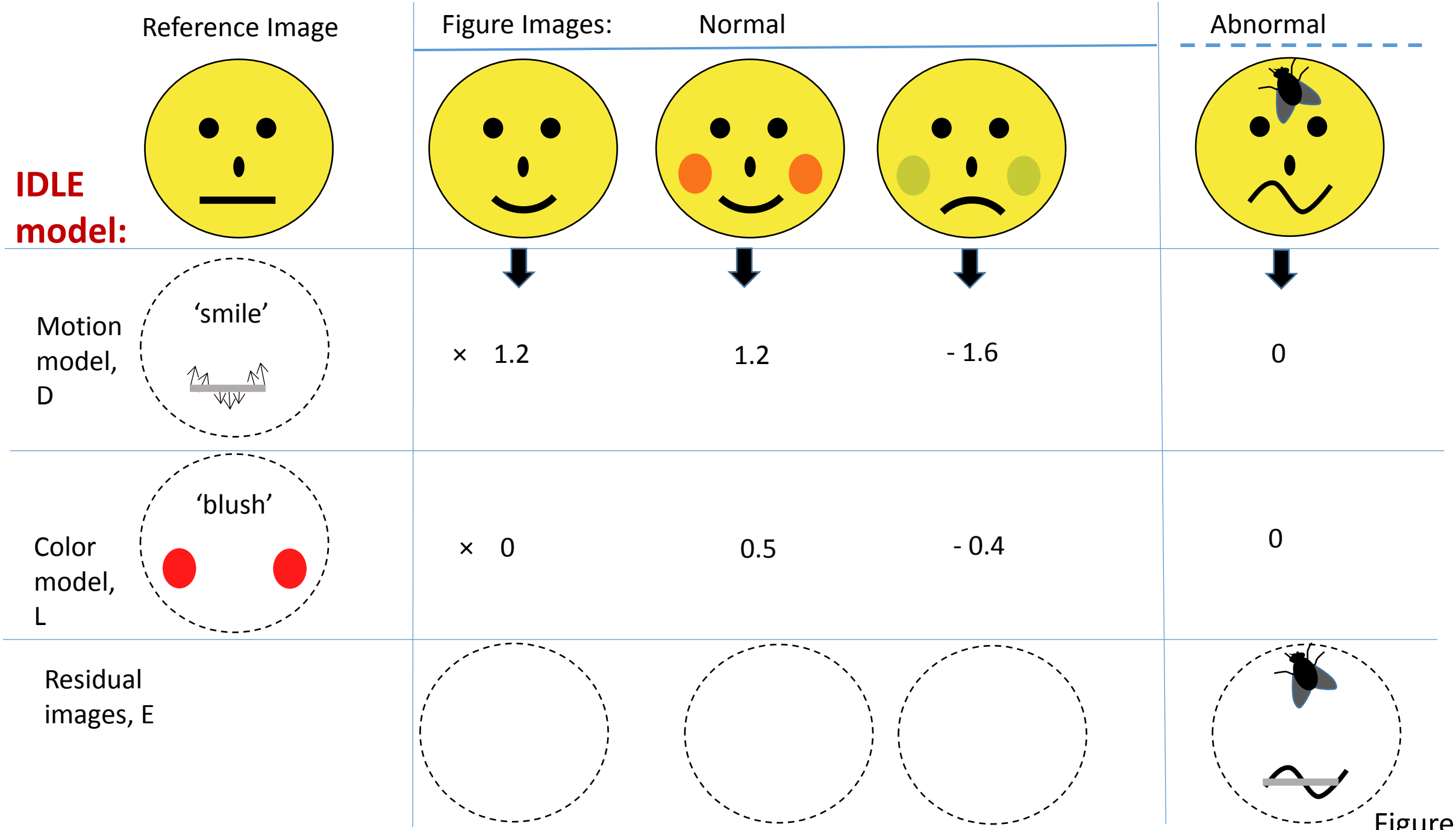

Reference Image
Figure Images:
Normal
Abnormal
IDLE model:
Motion model, D
'smile'
× 1.2
1.2
- 1.6
0
Color model, L
'blush'
× 0
0.5
- 0.4
0
Residual images, E

Figure 8

Example of IDLE modelling

Motion of babies in RGB video:

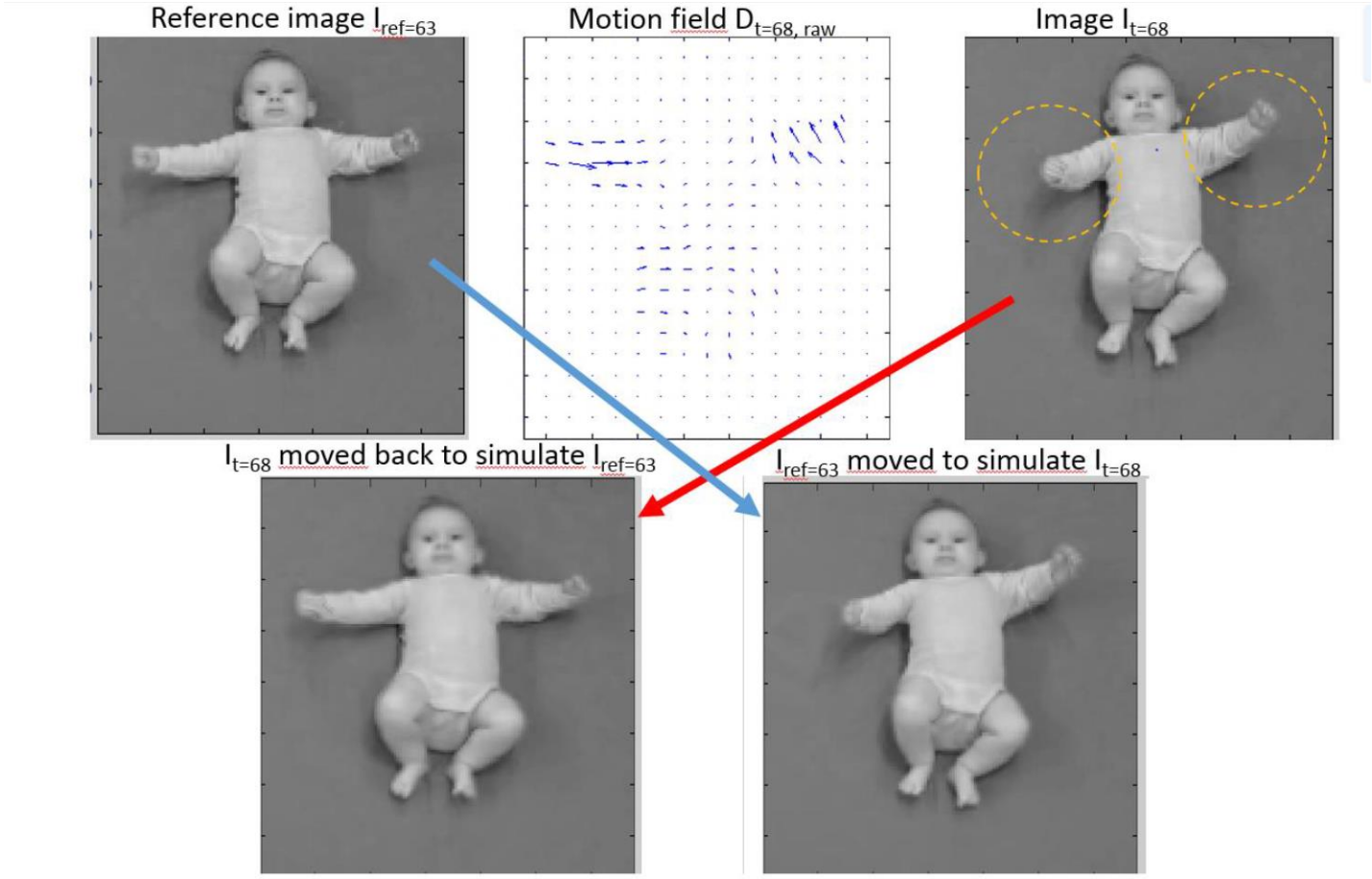
Early detection of cerebral palsy?

Acknowledgements:

Lars Adde, StOlavs Hospital

Hodjatolah Rahmati, Ole Morten Aamo, NTNU

IDLE modelling
Motion between frames 63 and 68



The Beluzov-Zhabotinsky reaction as a reproducable metaphor of life?

Medicine against Macho Math and Gucci Statistics: Try to model the real Belusov-Zhabotinsky reaction! (Anna Zhyrova & Dalibor Stys, Cz)

Malonic acid, inorganic salts, sulfuric acid and Ferroin (iron-based red/ox indicator) dissolved in water at low pH:

In Petri-dish: spontaneously forms a variety of attractor types!

We use this system do make generic methods & software for data-driven model development in complex systems

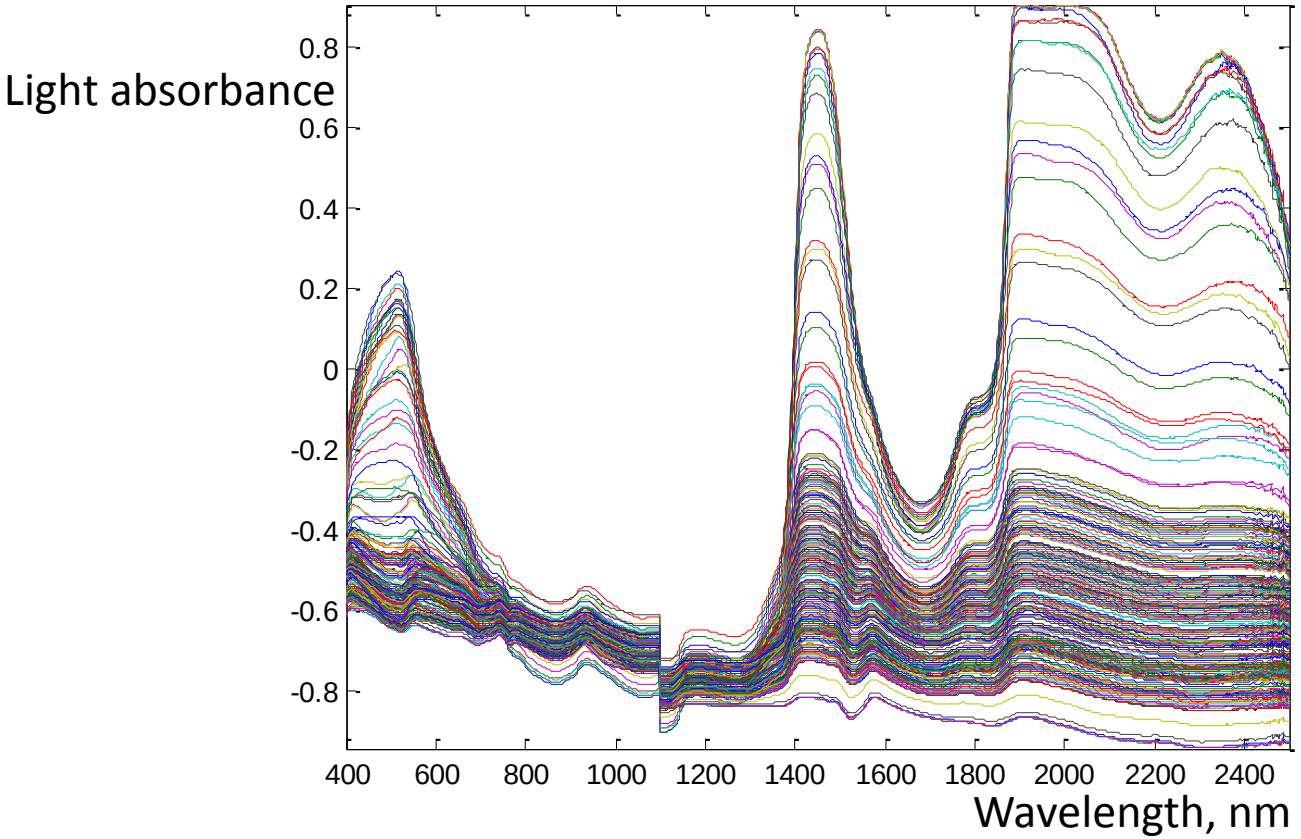
Our OnTheFlyCompression technique for generic modelbased data compression combines soft and hard modelling from high-dimensional biochemometrics, dynamic modelling from cybernetics and dual-domain subspace modelling from IDLE gestalt registration.

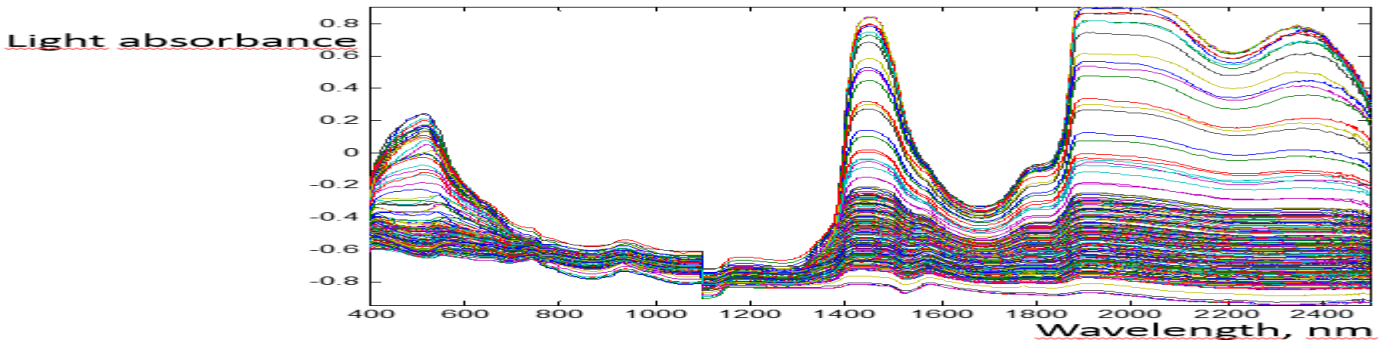
ENTHALPY: malonic acid -> CO2. Boring! A complex ENTROPY-driven pattern formation process. Data-driven modelling?

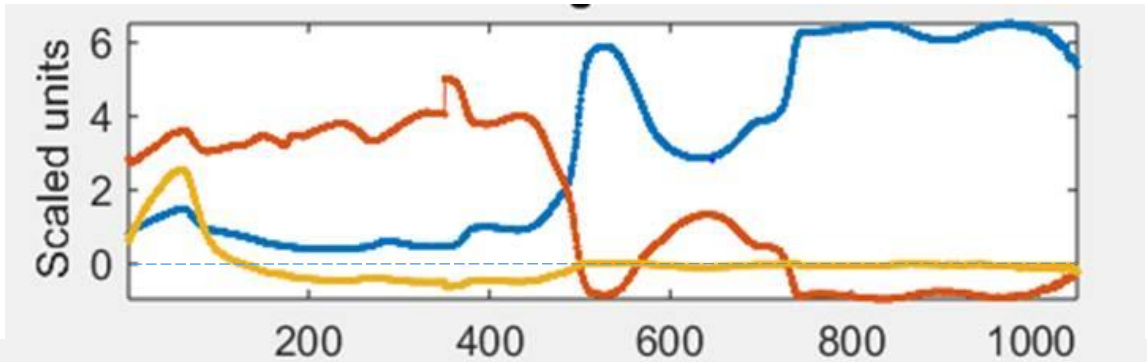
NIR/vis. spectral measurements of Belusov-Zhabotinsky reaction

(Anna Zhyrova & Dalibor Stys, Cz/ Bjørn Alsberg & Harald Martens, NTNU)

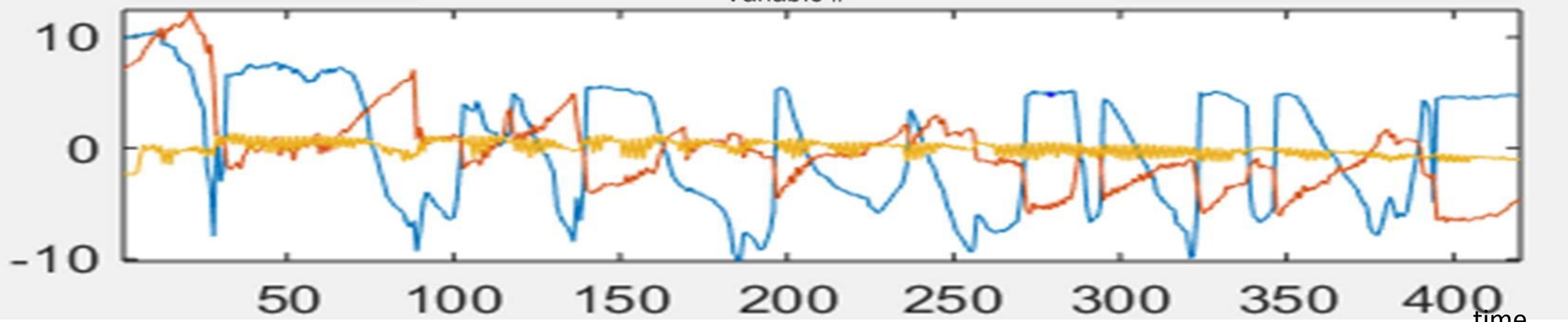
Raw transflectance spectra (T), linearised as absorbance (log(1/T)) at different wavelengths (nm).

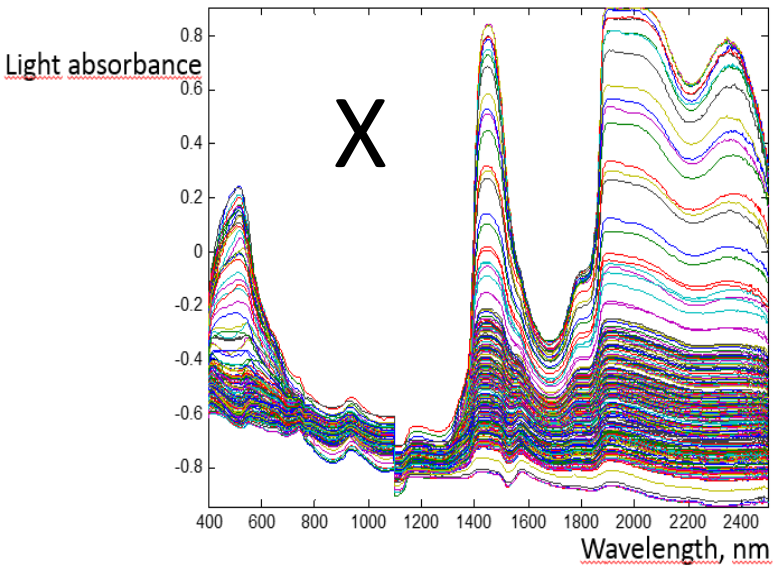




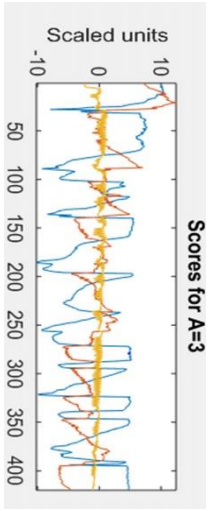


BZ reaction in stirred vessel, monitored by multichannel vis/NIR spectroscopy.
"Continuous PCA":
Three first subspace dimensions= > 99% explained variance. File reduction: 3:1000, i.e.

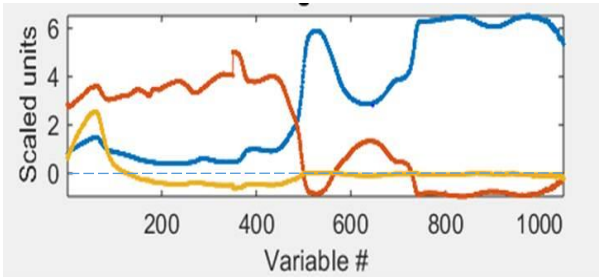




$X = \bar{x} +$



$*$



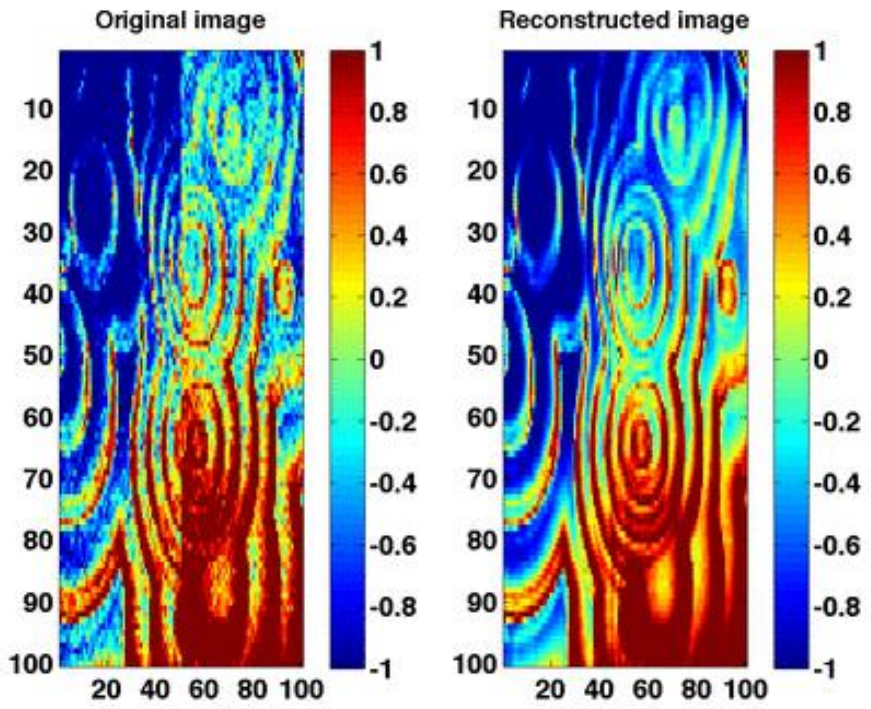
$+ \varepsilon$

time

«Never-ending» processes generate huge amounts of data

Hyperspectral NIR/vis video of Belusov-Zhabotinsky reaction (Anna Zhyrova & Dalibor Stys, Cz/ Lise Lyngsnes, Harald Martens, NTNU)

- Original Data Size 44.4Mb, compressed 925kb
- Compression ratio: ~2% of the original.
- Loss of information: 1%
- Compression time: ~33 seconds

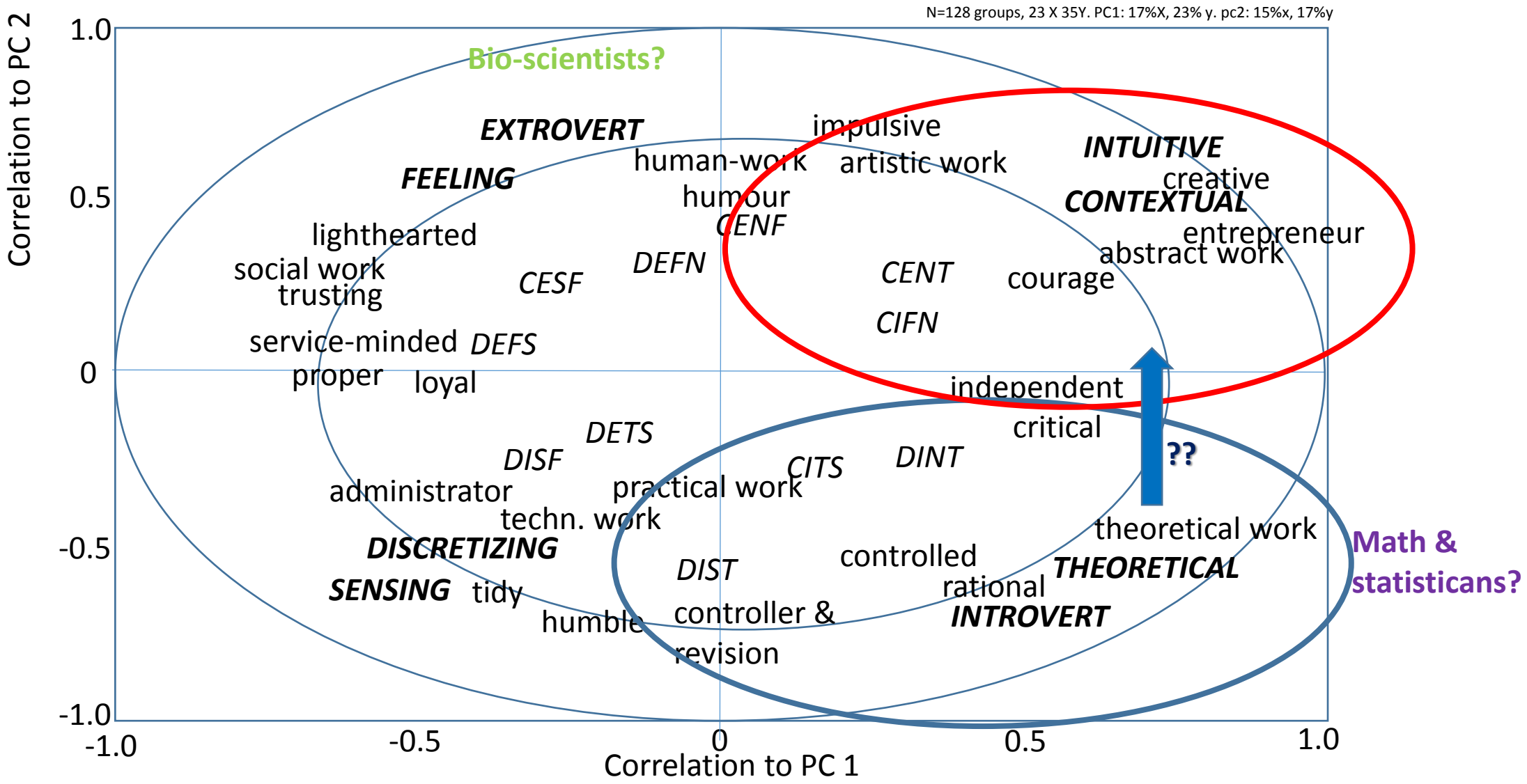


Need for model-based compression of data:

Hyperspectral NEO NIR video
1-2 GB of data/hr!
One single image from video recording, even after strongly reduced image size: 44.4 Mb

Combines soft and hard modelling from high-dimensional biochemometrics, dynamic modelling from cybernetics and dual-domain subspace modelling from IDLE gestalt registration.

Personality types and work types are correlated



Ongoing experiment, NTNU Dept. Engineering Cybernetics

Course (MSc, PhD) on «how to discover the real world»

1) Use e.g. microphone of students' PCs (FFT) as high-dimensional sensors
2) Student groups choose project for a semester
3) Students discuss experimental design
4) Students perform measurements according to chosen design
5) Students analyse data by cross-valildated, graphically oriented bilinear modelling (PCA, PLSR) in The Unscrambler (www.camo.com)
6) Students interpret bilinear model from physics
7) Report=exam

ISee:
Students' curiosity-driven discovery of real-world physics, math and statistics

IDLETechs AS

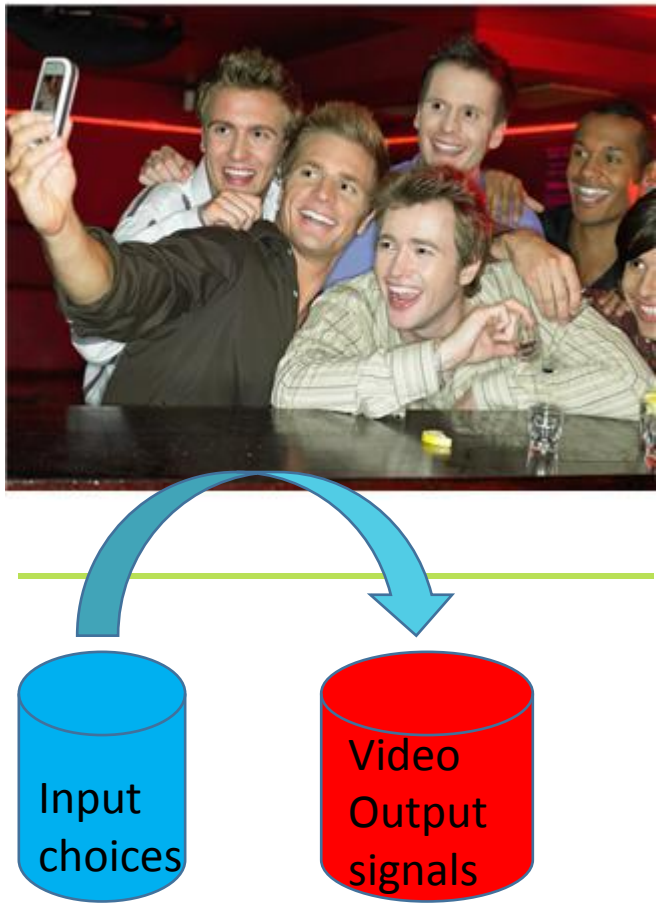
Purpose:

- Excite high school and university students' curiosity about their own physical world
- Let them feel how they can have fun being in actual control over physics, math and statistics, instead of just suffering to receive theoretical instruction from teacher.
- Motivate for learning laws of physics, tools of statistics and methods of mathematics
- Allow teacher to demonstrate physics, math and statistics in reality, via students' video filming

Basic method for students

- The students makes video sequences of whatever they like, e.g. via their mobile phones.
- Each video sequence is represented by a «soft modelling»: compact, simple and data-driven model of motion and colour changes of each independent object in the image sequence.
- Sound is likewise submitted to model-based compression in both the time- and frequency domains. This data modelling is automatic, hidden to the students, and yields simple graphs.
- The students can playfully modify the compact bilinear video data model via the interactive graphs, and thereby reconstruct video sequences of «alternative realities»

Basic method for students

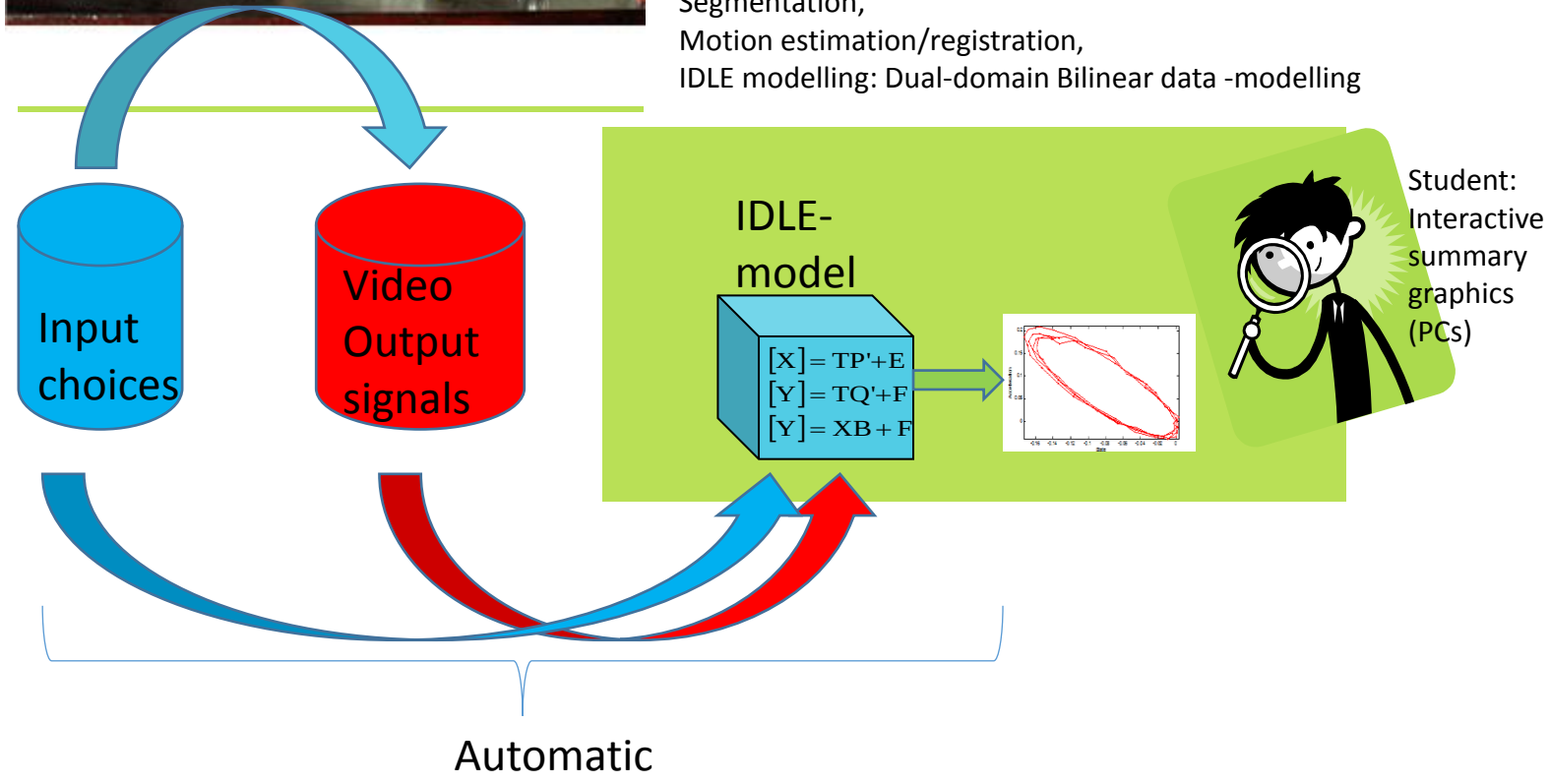


Basic method for students

IDLE model:
For each independent object in video:

Intensity = DisplacementModel(LocalIntensityModel)+Error

Requires:
Segmentation,
Motion estimation/registration,
IDLE modelling: Dual-domain Bilinear data -modelling



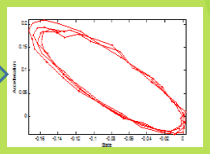
Basic method for students

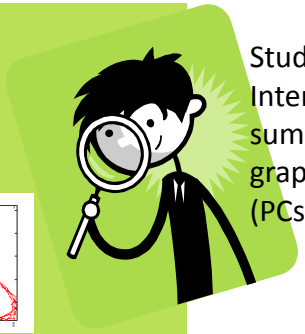
Input choices

Video Output signals

IDLE-model

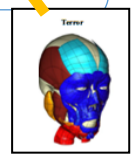
$[X] = TP' + E$
$[Y] = TQ' + F$
$[Y] = XB + F$





Student: Interactive summary graphics (PCs)

Metamodel



Automatic

Interactive

Basic method for students

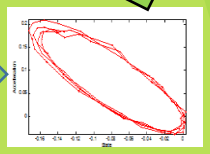
Morph to funny new video sequences

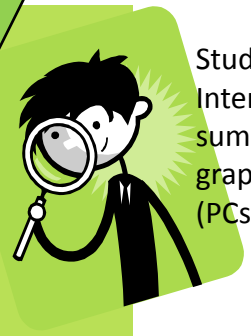
Input choices

Video Output signals

IDLE-model

$[X] = TP' + E$
$[Y] = TQ' + F$
$[Y] = XB + F$





Student: Interactive summary graphics (PCs)

Metamodel

Automatic

Interactive

Basic method for students

Morph to funny new video sequences

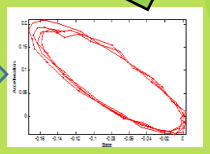
Generate video game etc

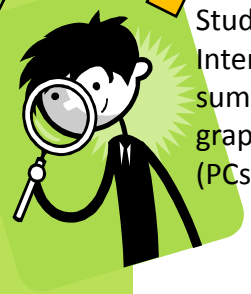
Input choices

Video Output signals

IDLE-model

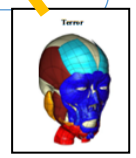
$[X] = TP' + E$
$[Y] = TQ' + F$
$[Y] = XB + F$





Student: Interactive summary graphics (PCs)

Metamodel



Automatic

Interactive

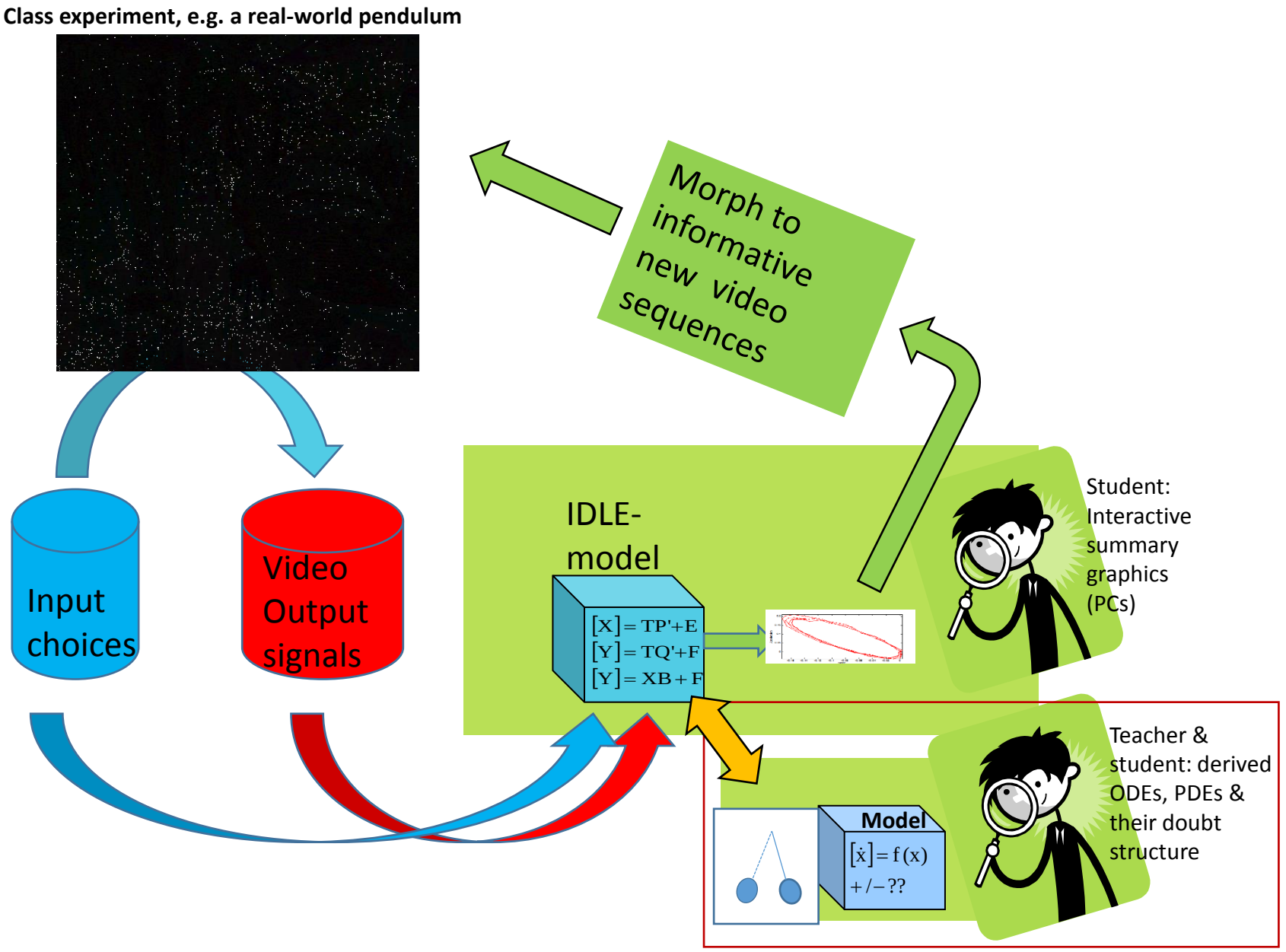
Classroom method for teachers

- The soft-to-hard model conversion may be purely data-driven, but will then yield «Ptolomus-type » rather than «Kepler-type» mathematical models – with good *ad hoc* descriptive ability of the actual video content, but without a simple physical interpretation.
- Hard models with simpler and more meaningful interpretation may be obtained for well recognized objects in the video sequence, e.g. hard objects with affine transformation, human faces talking and showing emotions, flowing water etc. The soft-to-hard conversion is aided by multivariate metamodels of causal-cohesive mechanistic models, while the purely data-driven model development is only used for quantifying unexpected phenomena in the video sequence.
- The teacher can thus demonstrate to the students that the laws of physics actually control the students' daily activities, and show that mathematics is useful, intuitively understandable and something that they can learn to control.

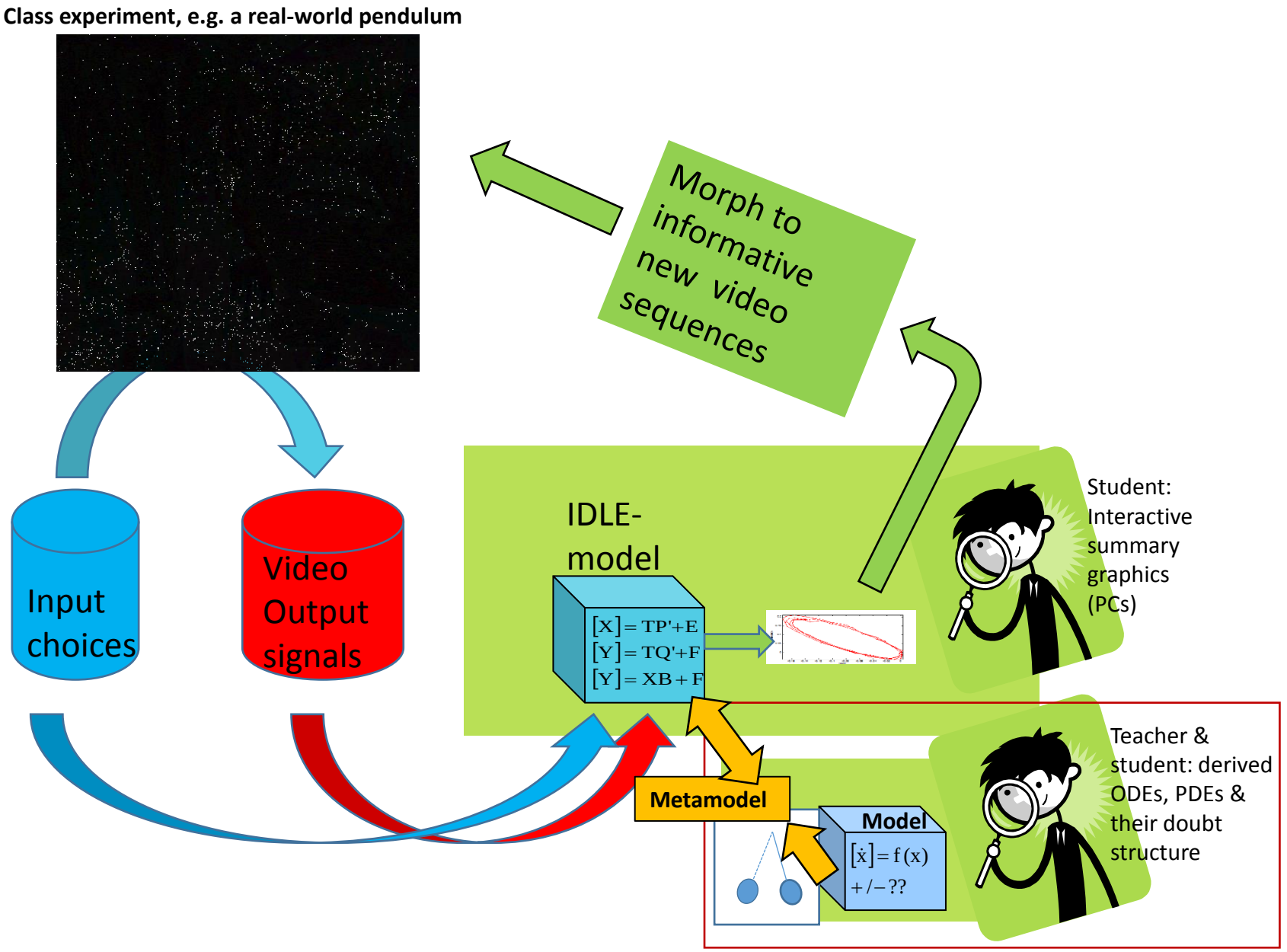
University level students' use

- Used at university level, this combination of «soft», «semi-soft» and «hard» modelling can show the students that while ambitious, mechanistic models of known phenomena are very useful in science, they need to be supplemented by more humble, data-driven models of unexpected phenomena.
- Thus, while our scientific understanding of reality is always imperfect, there is no need for mystification of new phenomena, as long as they are observed in a statistically valid way.
- The system will also at university level be useful for reducing math anxiety and enhancing students' personal interest in science.

Classroom method for teachers



Classroom method for teachers



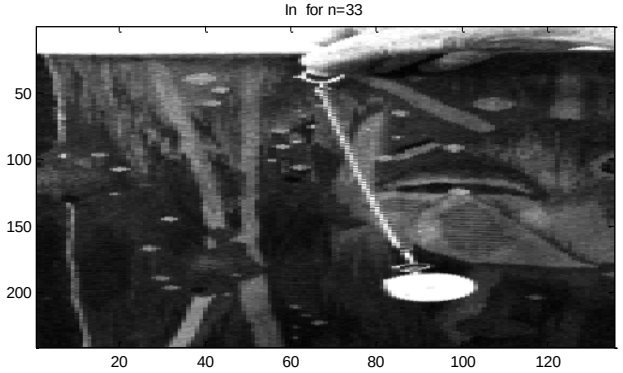
Very preliminary example
(my Matlab software is just a rudimentary demo)

My pocket watch may be a perfect pendulum.

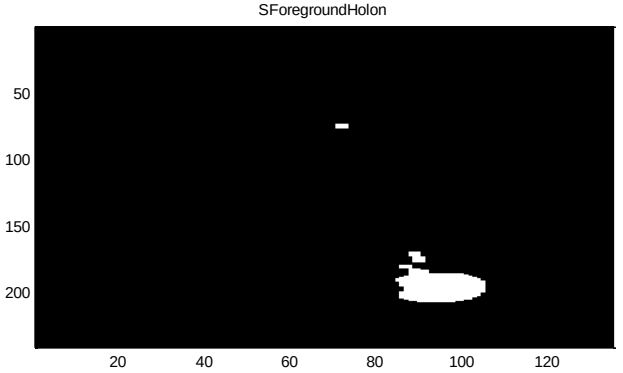
But my i-phone video of it is not perfect, because my camera was moving and so was my hand holding the watch.

Frame # 33 is almost identical to the frame chosen to be reference, # 7

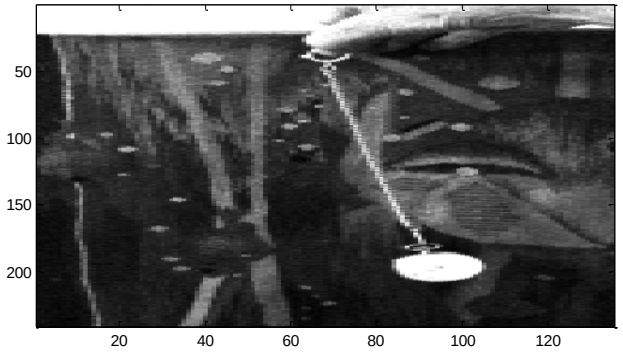
New frame=#33



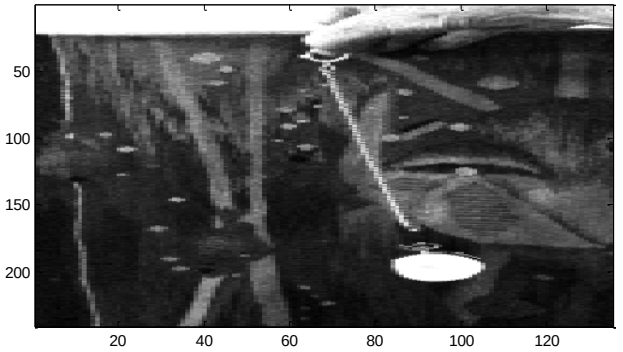
Clearly moving object



Reference frame=# 7

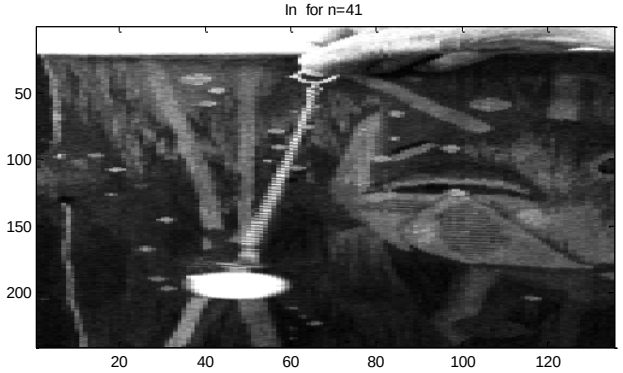


New frame #33 moved to reference position

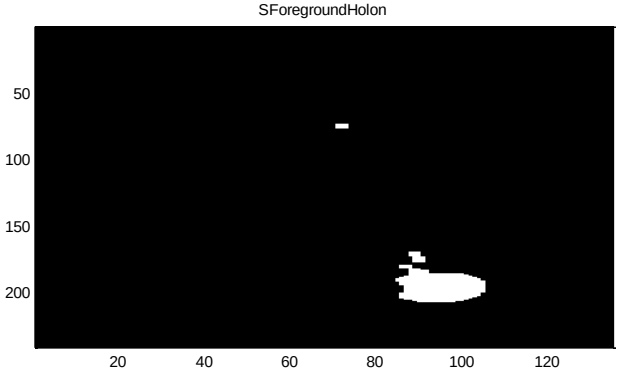


Frame # 41 is quite different from the reference, # 7: the watch pendulum has moved – and so has the camera

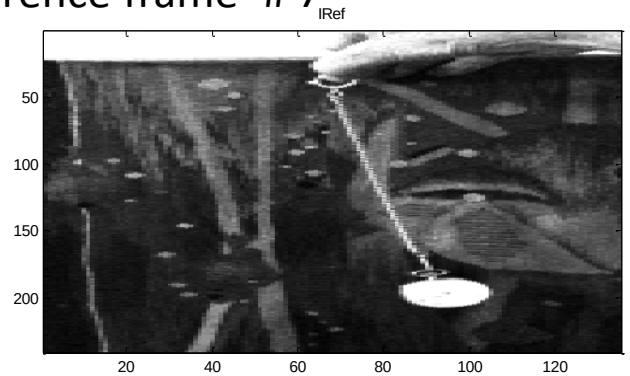
New frame=#41



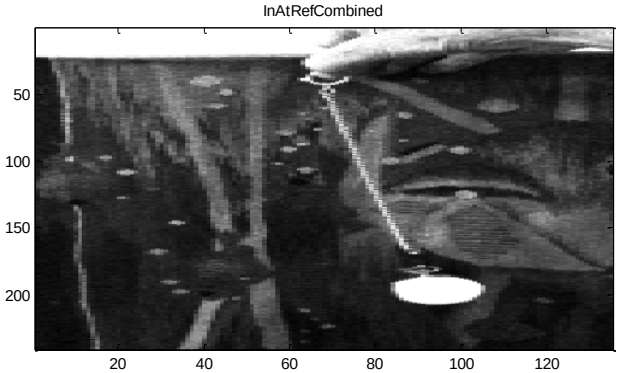
Clearly moving object



Reference frame=# 7

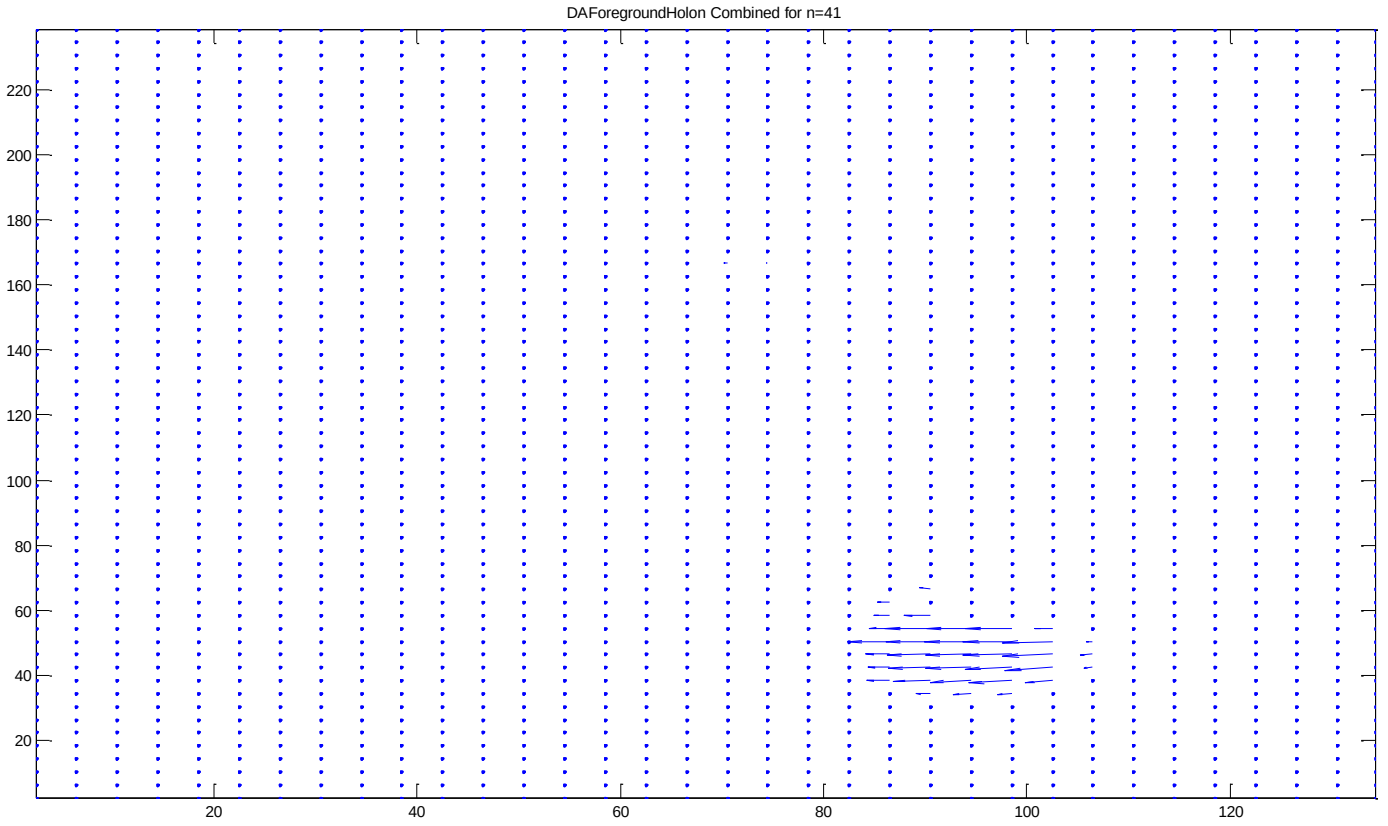


New frame #41 moved to reference position



Camera motion corrected and object motion corrected

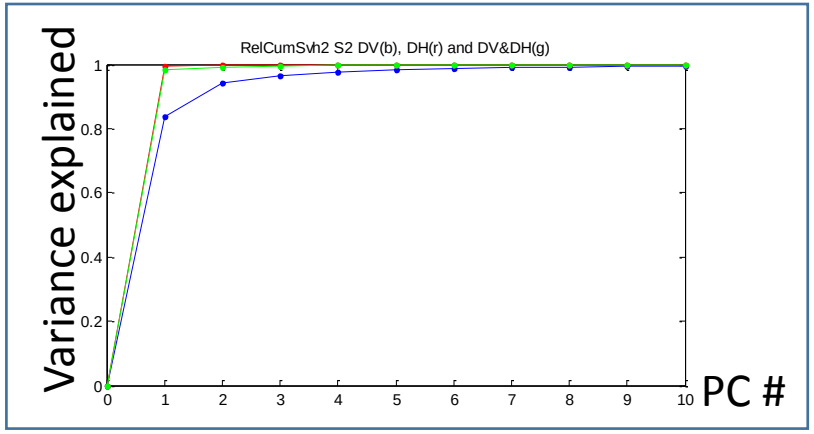
Motion of pixels from reference frame # 7 to frame # 41, determined by optical flow estimation



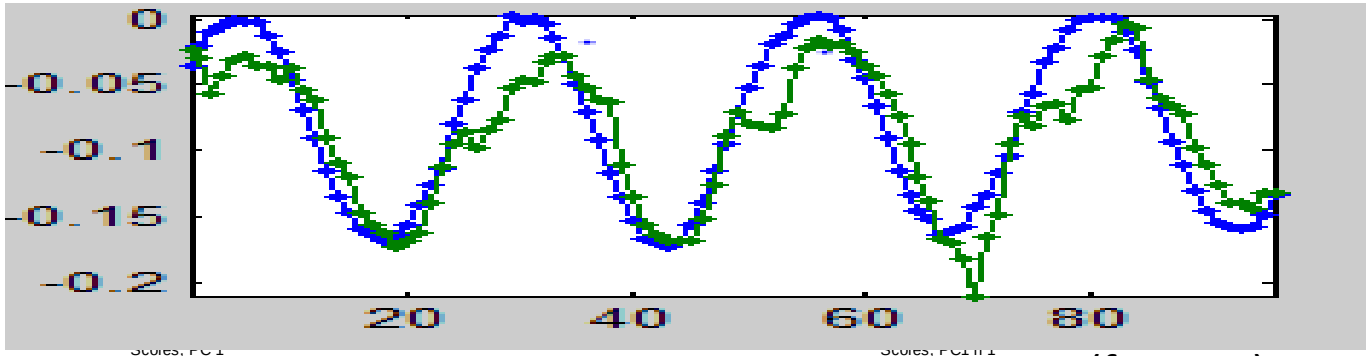
At this first estimation stage, the motion of the thin chain of the watch is not correctly estimated. That is corrected at a later stage of the automatic estimation process.

Automatic model-based compression of motion information:

Principal Component Analysis shows that the horizontal and vertical motions are sufficiently approximated by their first bilinear Principal Component (PC1) :

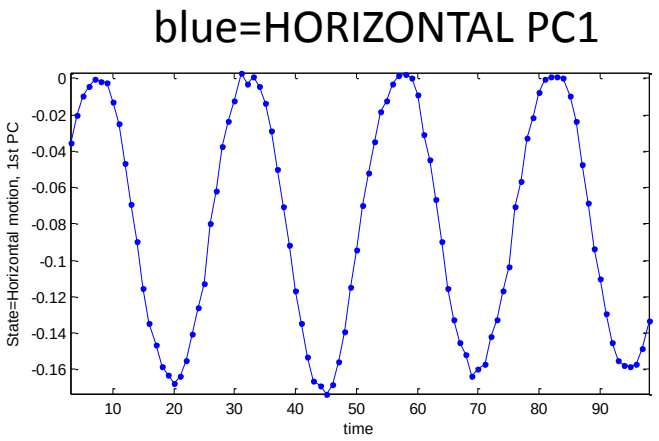


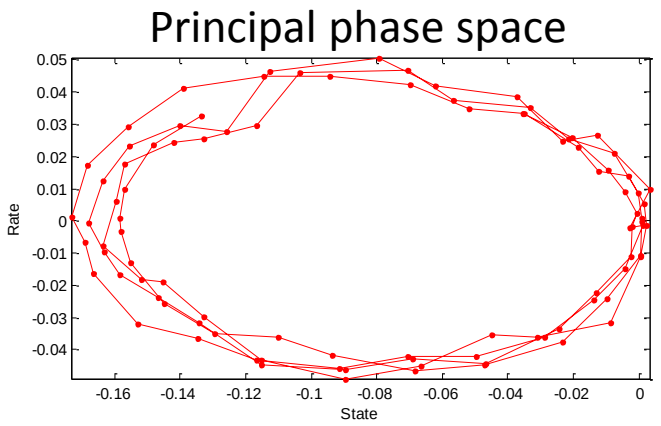
Motion summaries: blue=HORIZONTAL PC1, green = VERTICAL PC1

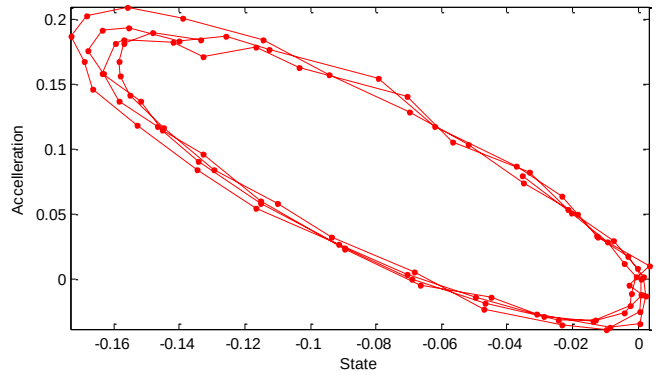


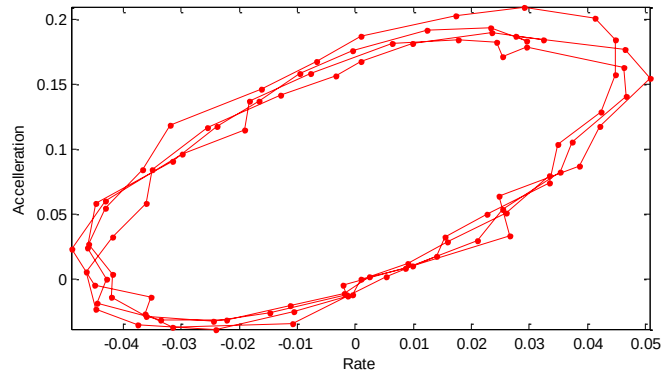
time (frame #)

Horizontal motion fields is automatically approximated by its first PC (its first latent state varible). Plots of this state variable vs its time derivative and accellerations shows the laws of nature that govern this video sequence :

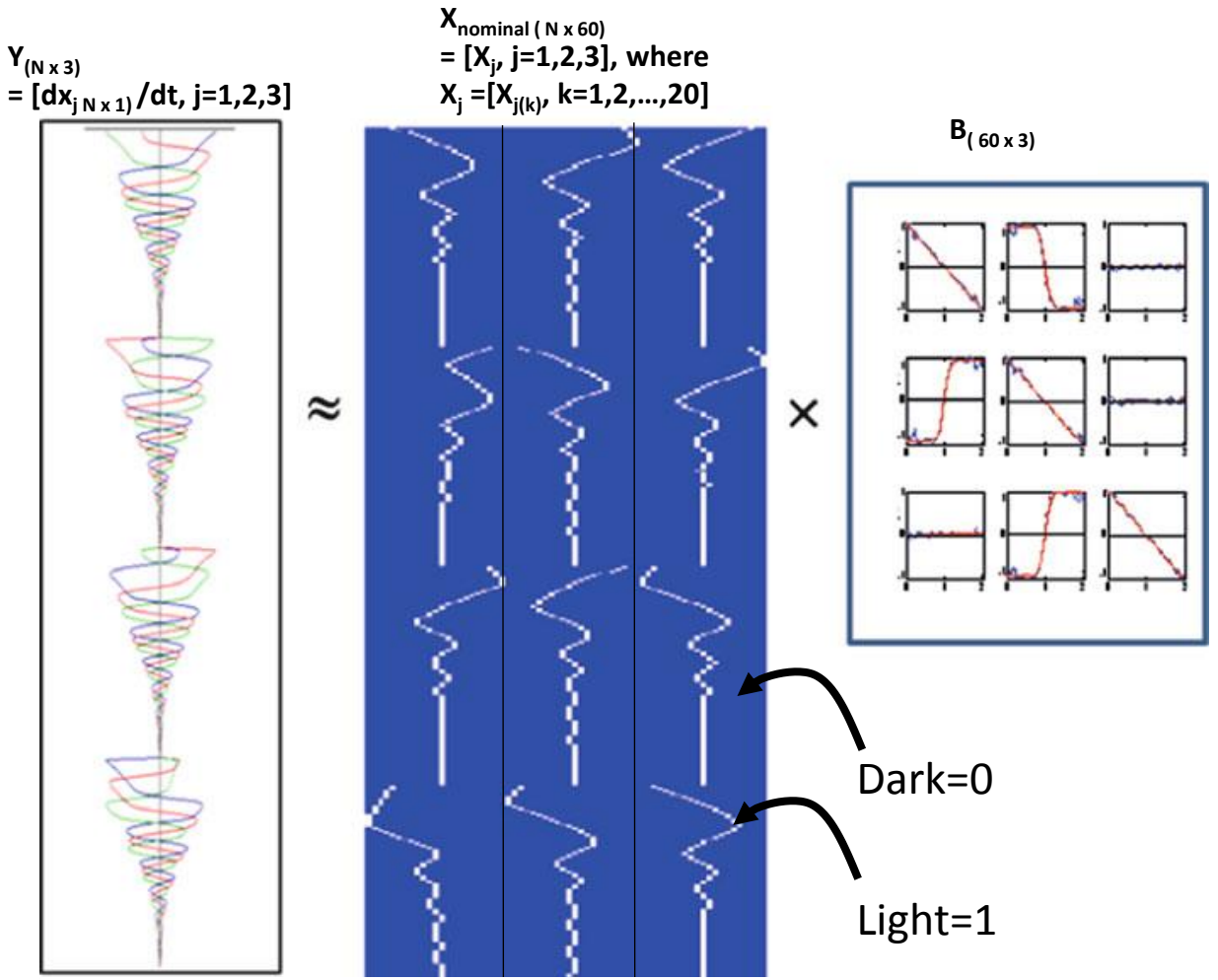








Non-linear data-driven ODE development, e.g. in bilinear Scores

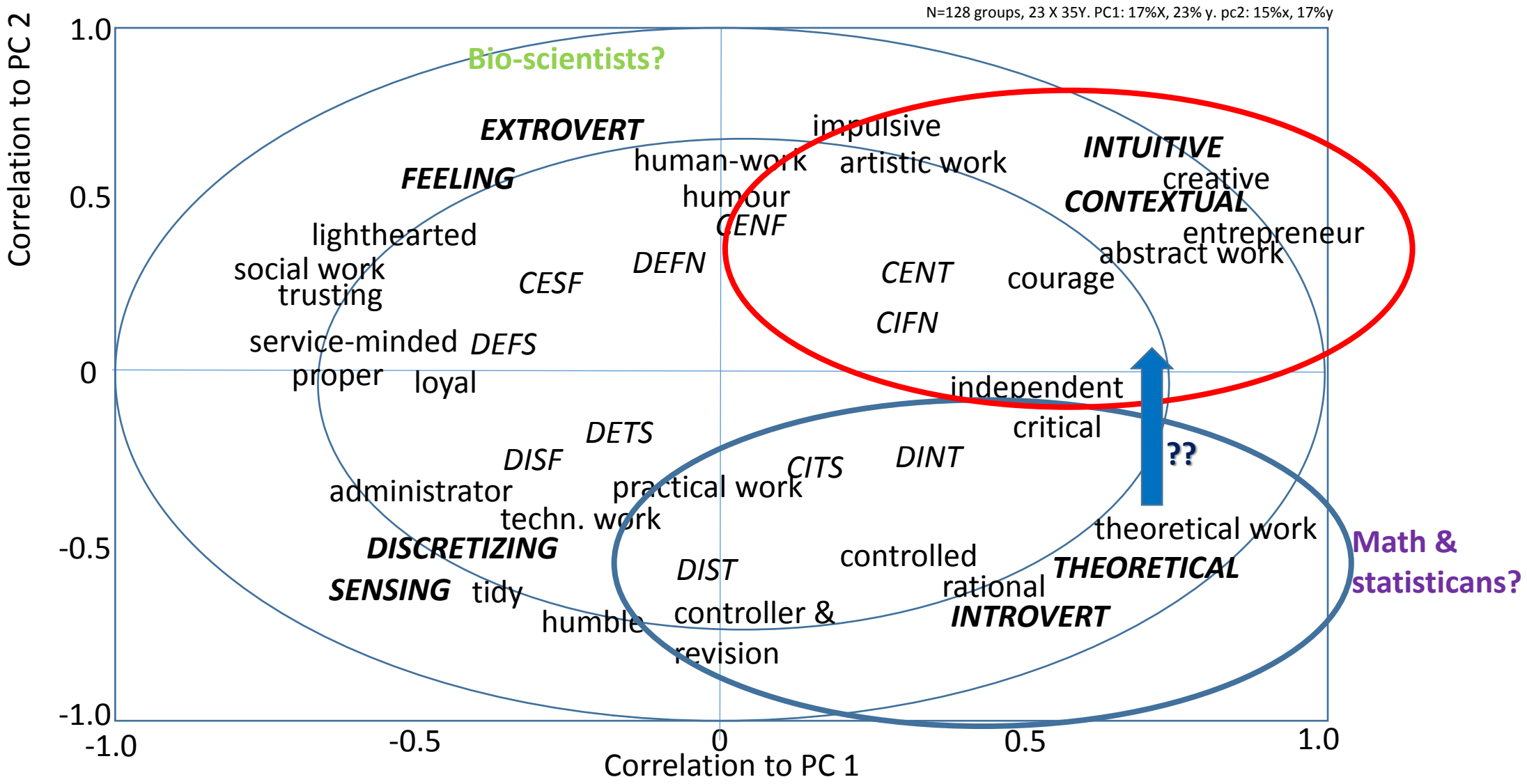


The Estimated Jacobian B of a dynamic system, changing over the state space:
Low-rank regression coefficients from PLSR of rates Y vs states X (20 nominal 0/1 state variables for each of 3 quantitative state variables).

Martens, H, Tøndel K, Tafintseva V, Kohler A, Plahte E, Vik JO, . Gjuvsland AB, . Omholt SW (2013) PLS-Based Multivariate Metamodeling of Dynamic Systems. New Perspectives in Partial Least Squares and Related Methods . Springer Proceedings in Mathematics & Statistics 56, pp 3-30. DOI 10.1007/978-1-4614-8283-3_1

Developing nonlinear ODE models from state variable time series data by nominal-level dynamic PLSR.

Personality types and work types are correlated



THANK YOU!