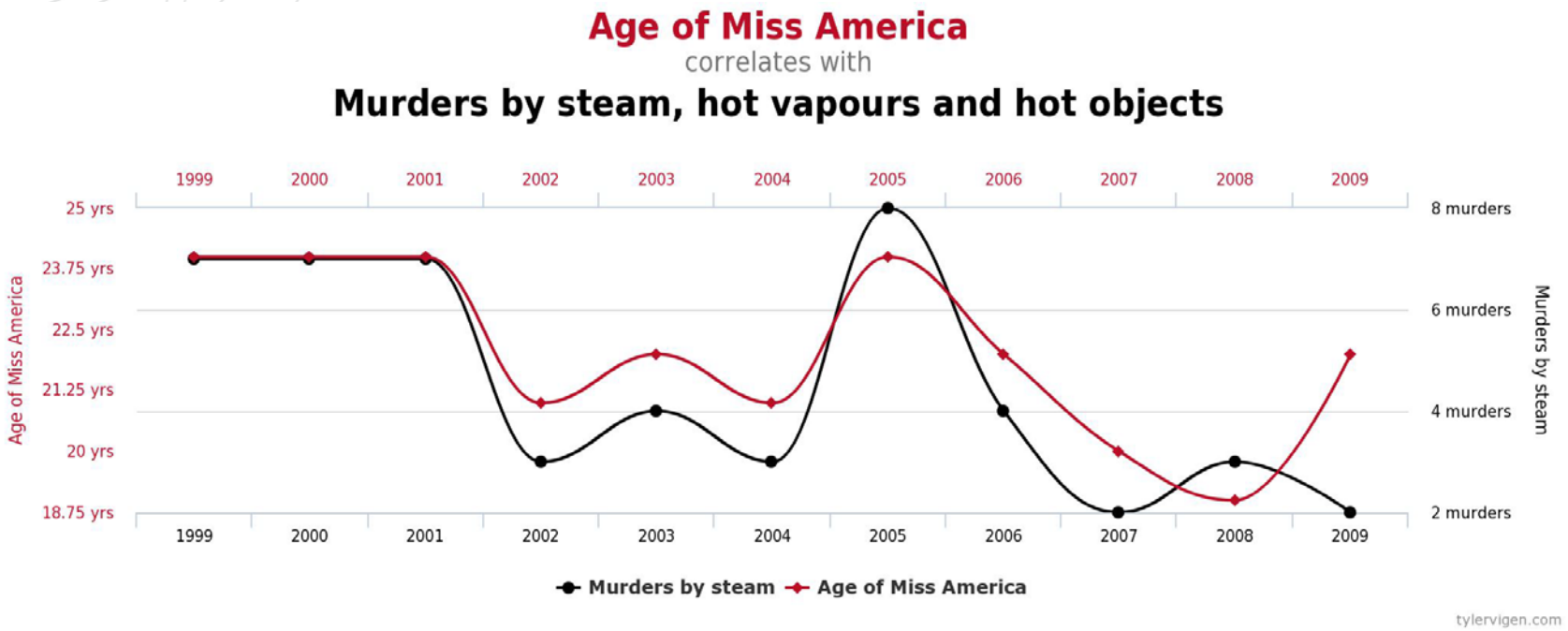
Visual Analytics to support research & applications for a sustainable future

Thomas Tamisier
thomas.tamisier@list.lu

LUXEMBOURG INSTITUTE OF SCIENCE AND TECHNOLOGY

Evidence in data – case 1



correlation ≠ causation

Evidence in data – case 2

Anscombe's example*

Let's consider datasets I- IV with:

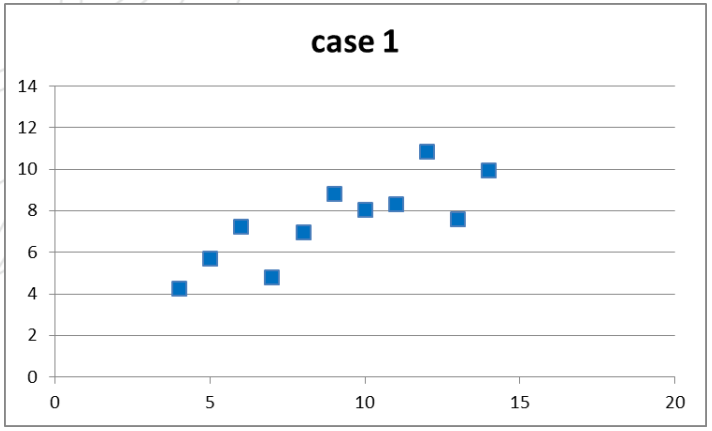
- ✓ same mean (*x*) = 9.0,
- ✓ same mean (*y*) = 7.5
- ✓ same x variance
- ✓ nearly same y variance
- ✓ identical (x,y) correlation
- ✓ identical regression lines

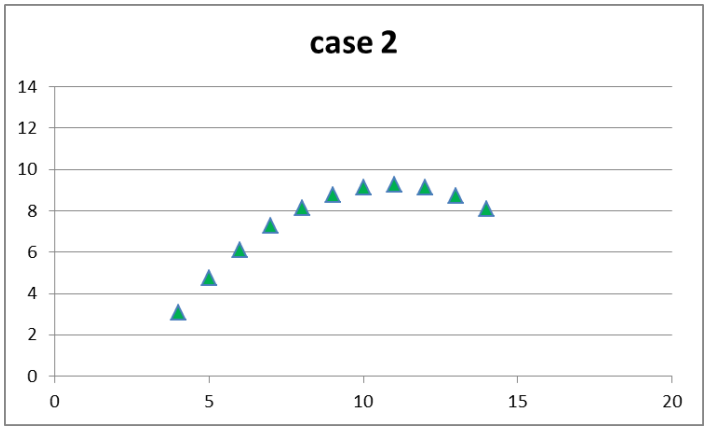
I		II		III		IV	
x	**y**	**x**	**y**	**x**	**y**	**x**	**y**
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

similarity between the 4 cases could be deducted

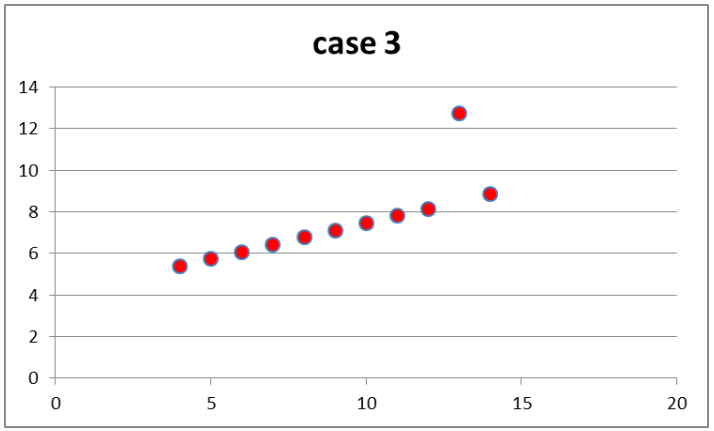
*** The American Statistician, Vol. 27 (1), pp. 17-21 (1973)**

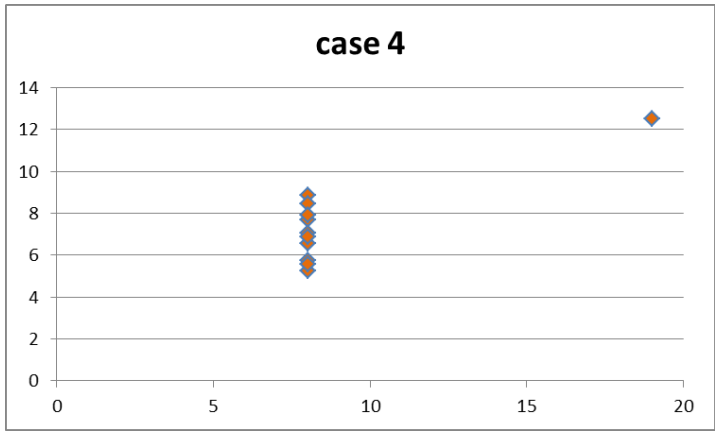
Case 2 continued





Seeing the graphics makes the differences crystal clear



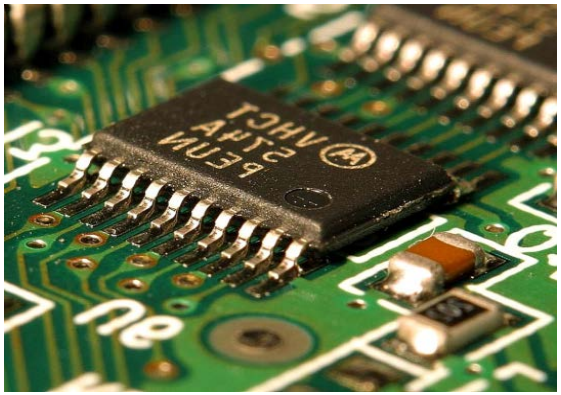


Data to knowledge

Making sense of data needs perspectives & domain knowledge (case 1)

Combining human & machine is successfully supported by visualisation (case 2)

Data Analytics Approaches
(statistics, machine learning…)



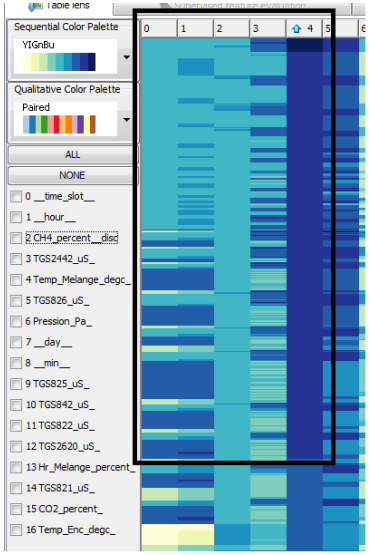
Data Visualisation Approaches
(InfoVis, 3D Visu., Maps…)

How to use Computer + Human Vision in the most efficient way?

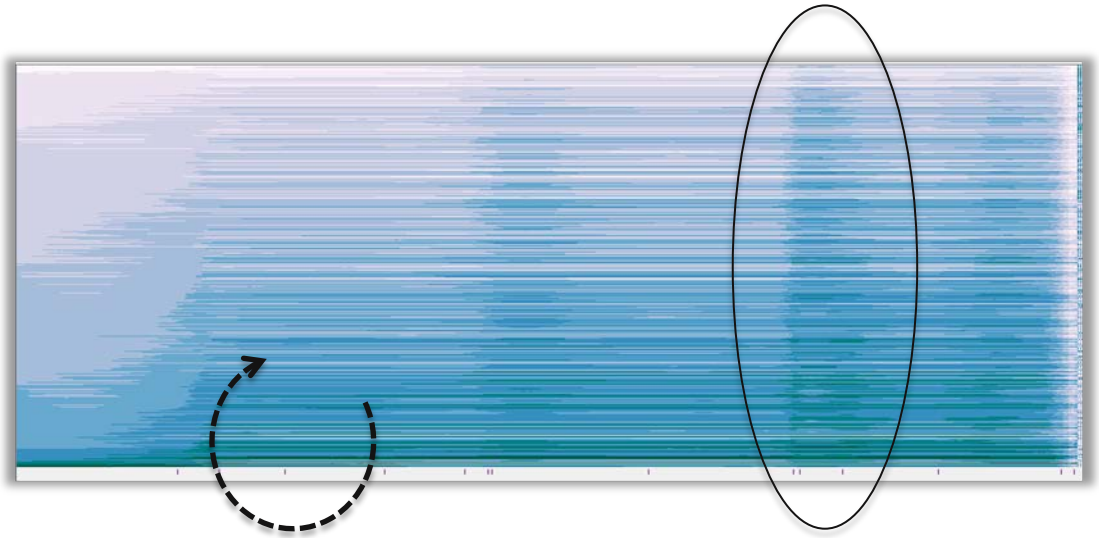
Human in the loop

Visual perception

- Most important input channel to communicate information to the brain
- Very efficient, hard-coded in the human brain
- Parallel processing of visual variables (e.g. colours, shapes, movement…)
- Intuitive, no need for teaching
- Still much more efficient than computer vision



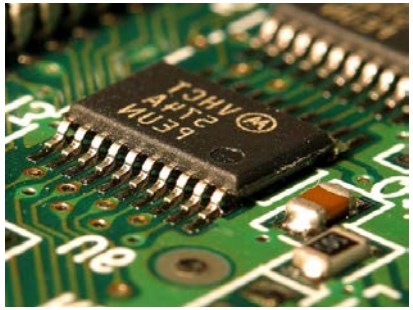
Pattern recognition



Human vision support (real-time) interaction with graphics

Visual Analytics Paradigm

Visual analytics is the science
of analytical reasoning
facilitated by interactive visual interfaces*



* Illuminating the Path. The Research and Development Agenda for Visual Analytics, J.J .Thomas and K.A Cook, IEEE Editions, 2005.

Visual Analytics in Environment

Model is not the reality

Model → Abstraction, simplification

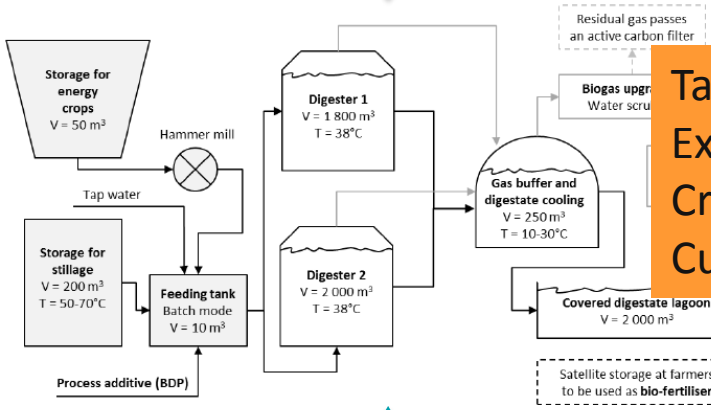


Figure from *

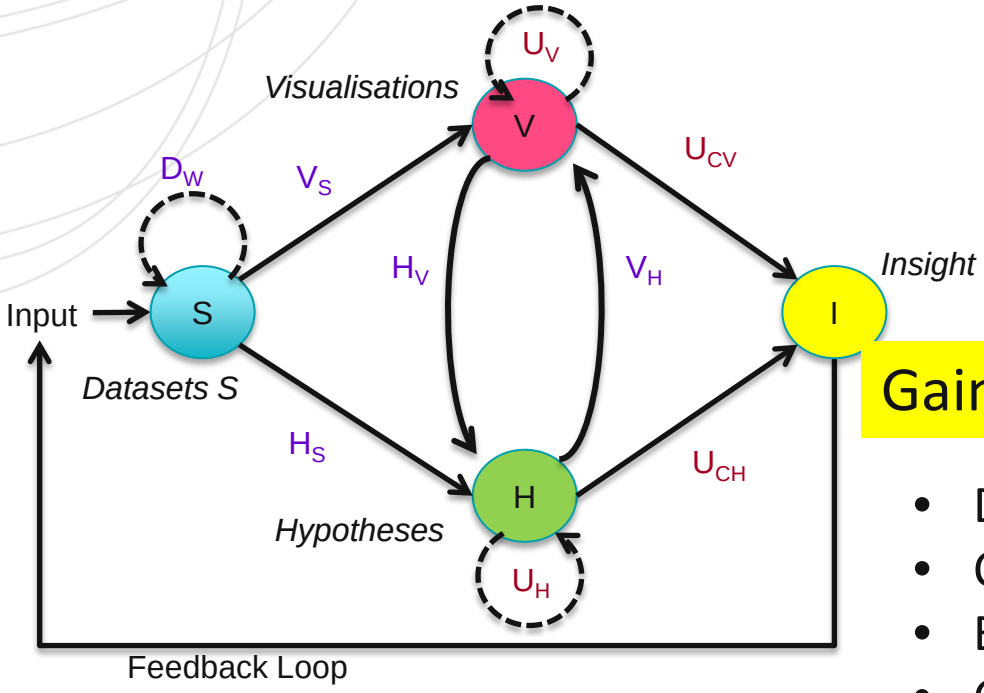
Tacit knowledge
Experience
Critical thinking
Curiosity

"I remember that I've already observed something similar"

"This behaviour contradicts this well-known chemical rule, we should check the data again"

* Biogas Production from Thin Stillage on an Industrial Scale – Experience and Optimisation. By Moestedt et al. In Energies, Oct. 2013

Visual Analytics in practice

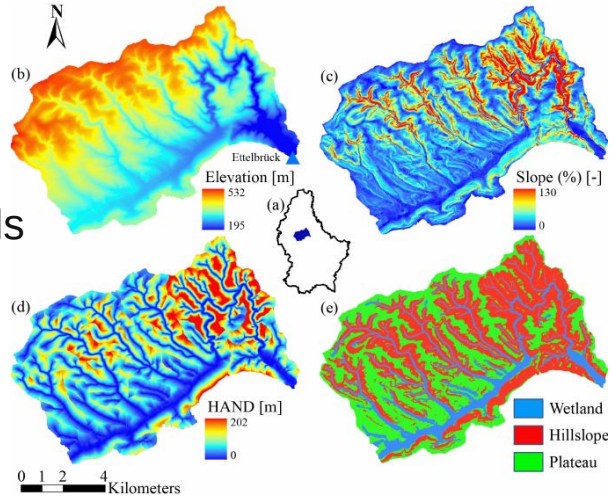


Gaining Insight… for:

- Data pre-processing
- Complex system understanding
- Ensemble exploration (simulation)
- Crisis management, citizen science
- …

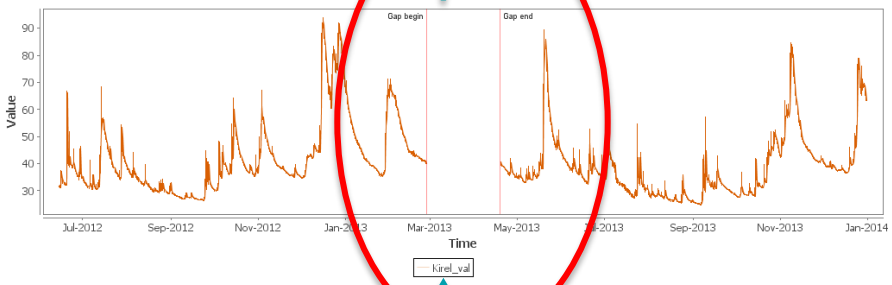
Data pre-processing (hydrology)

Hydrological models
Simulation results



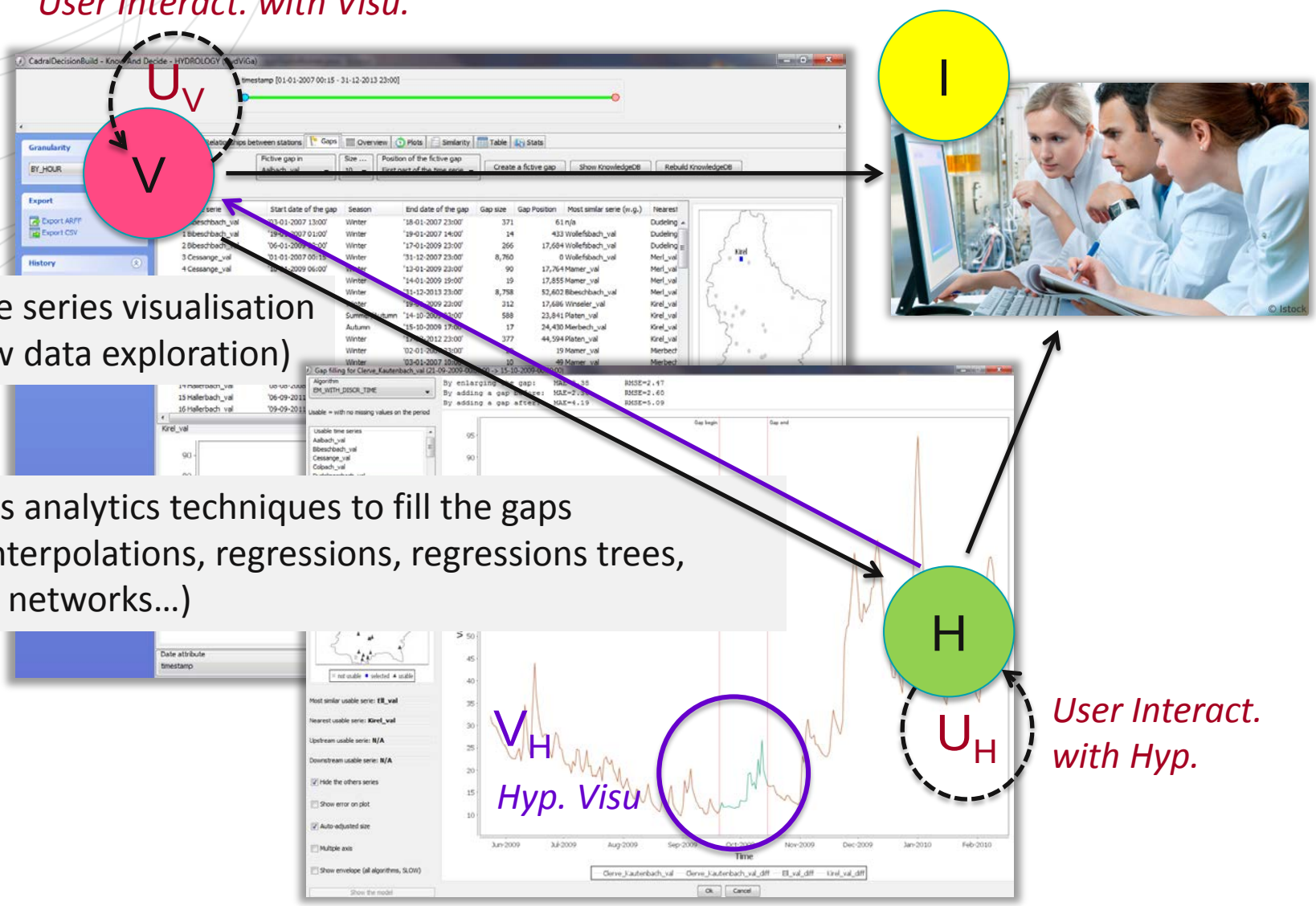
Pre-processed data

Missing Values



Raw Data
Collected from in-situ sensors

Hydrology continued



* A user-driven case-based reasoning tool for infilling missing values in daily mean river flow records. By Giustarini et al. In Environmental Modelling & Software, Aug. 2016

Complex systems (viticulture)

Better understanding the relationship between the wine and the terroir where it is produced

Heterogeneous raw data: 108 parameters

Grape features

Soil characteristics

Wine-making parameters

Scientific analyses of wine

Wine tasting scores

Viticulture continued

- Heterogeneous raw data: 108 parameters

Better characterization of various Lux. terroirs

User Interact. with Visu.

U_V

V

I

Generate Hyp. From Visu

Parallel coordinates visualisation

H_V

Interactive setting of analytics parameters

H

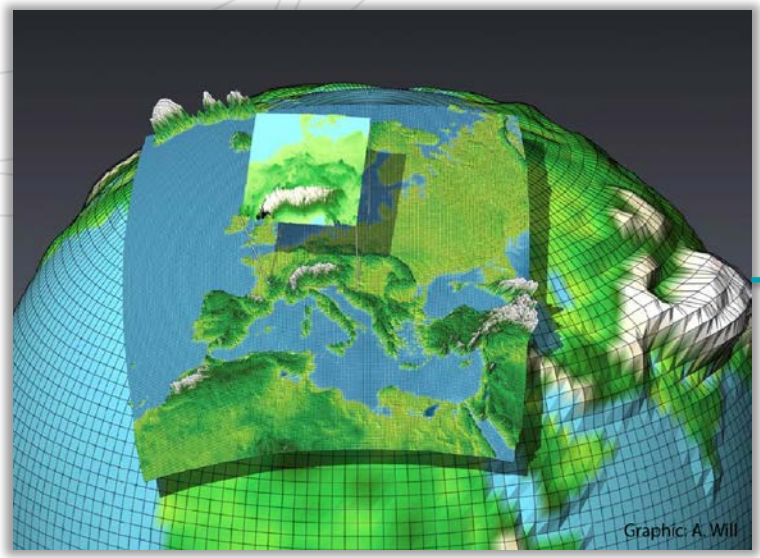
U_H

User Interact. with Hyp.

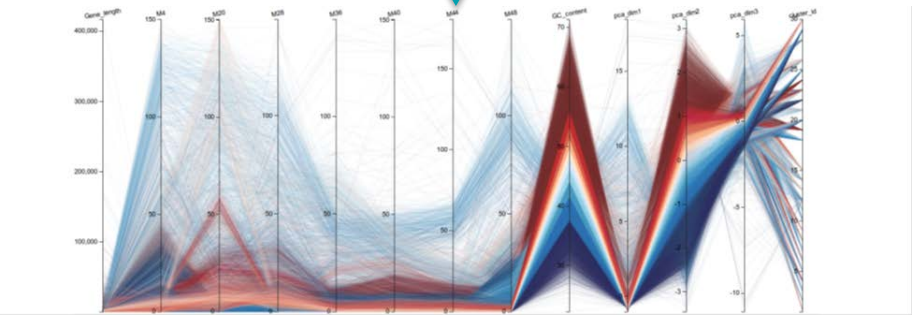
Dimensionality reduction

Simulation , exploration (climatology)

Sensitivity analysis of COSMO-CLM model



- Exploring the results of multiple simulation runs
- Find best range for model parameters
- …



- VISUAL CLASSIFICATION METHOD FOR A LARGE SET OF SENSITIVITY RUNS IN THE FRAMEWORK OF CCLM 5.0 MODEL EVALUATION.
- By Ferrone and Broeksema. In CLM Community Assembly, 2015

Unstructured data (events)

Crisis management, citizen science

V

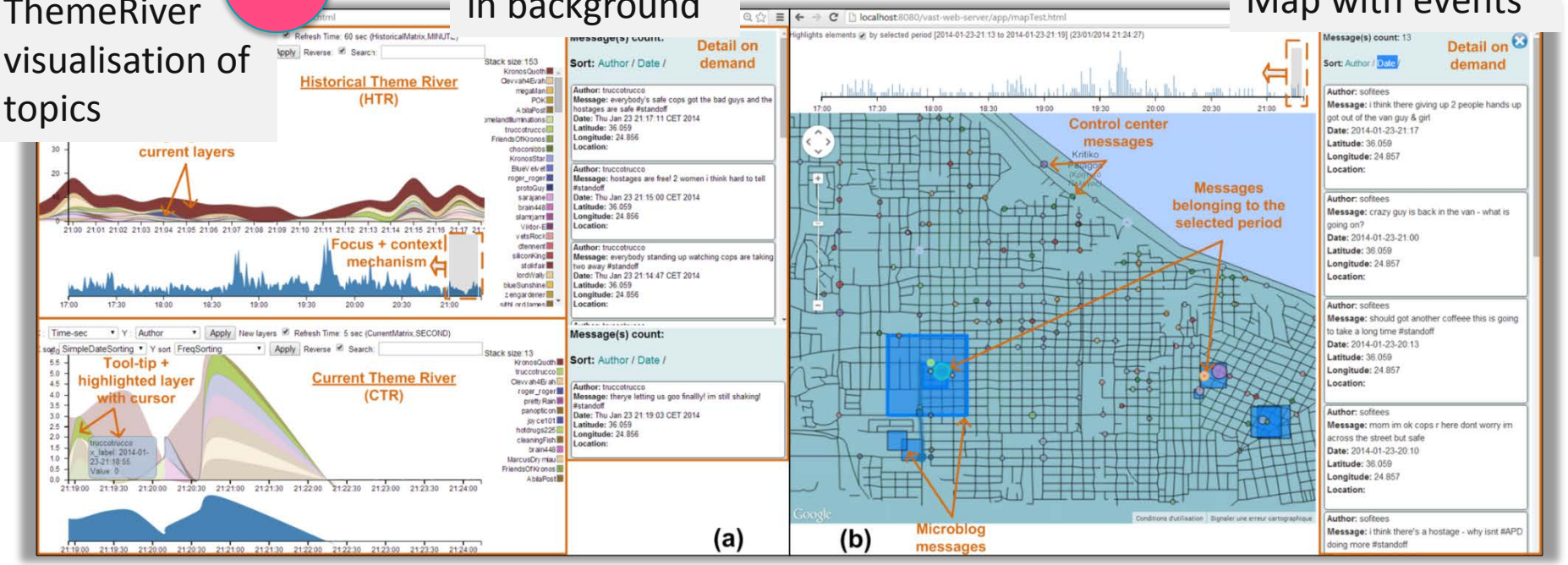
ThemeRiver visualisation of topics

H

Text analytics in background

V

Map with events



Raw Data: Tweets, geo-located messages and events

* Presented at VAST Challenge 2014. By Médoc et al.

Conclusion & Perspectives

Visual Analytics = Analytics + Visualisation + Human Factors

- Great potential for larger use in Environmental Informatics

Generic method

- Energy, Water, Agriculture, Remote Sensing, Biodiversity…
- Data pre-processing, Simulation, System analysis…

Challenges

- Interdisciplinary teams with expertise in Analytics AND Visualisation AND Human Factors together with Domain experts
- User Acceptance depends on background and mind openness

Contact

E-Science Research Unit
Environmental Research & Innovation Department (ERIN)
Luxembourg Institute of Science and Technology

Head of Unit:
Dr Ir Benoît Otjacques
benoit.otjacques@list.lu