Multivariate Evolutionary Classification in Astrophysics

(Astrocladistics)

Didier Fraix-Burnet

Biologists P. Choler, E. Douzery

Astrophysicists E. Davoust, T. Chattopadhyay, C. Charbonnel, F. Lamareille

Statisticians A.K. Chattopadhyay, (T. Chattopadhyay)

« Soft-Computing » M. Thuillard

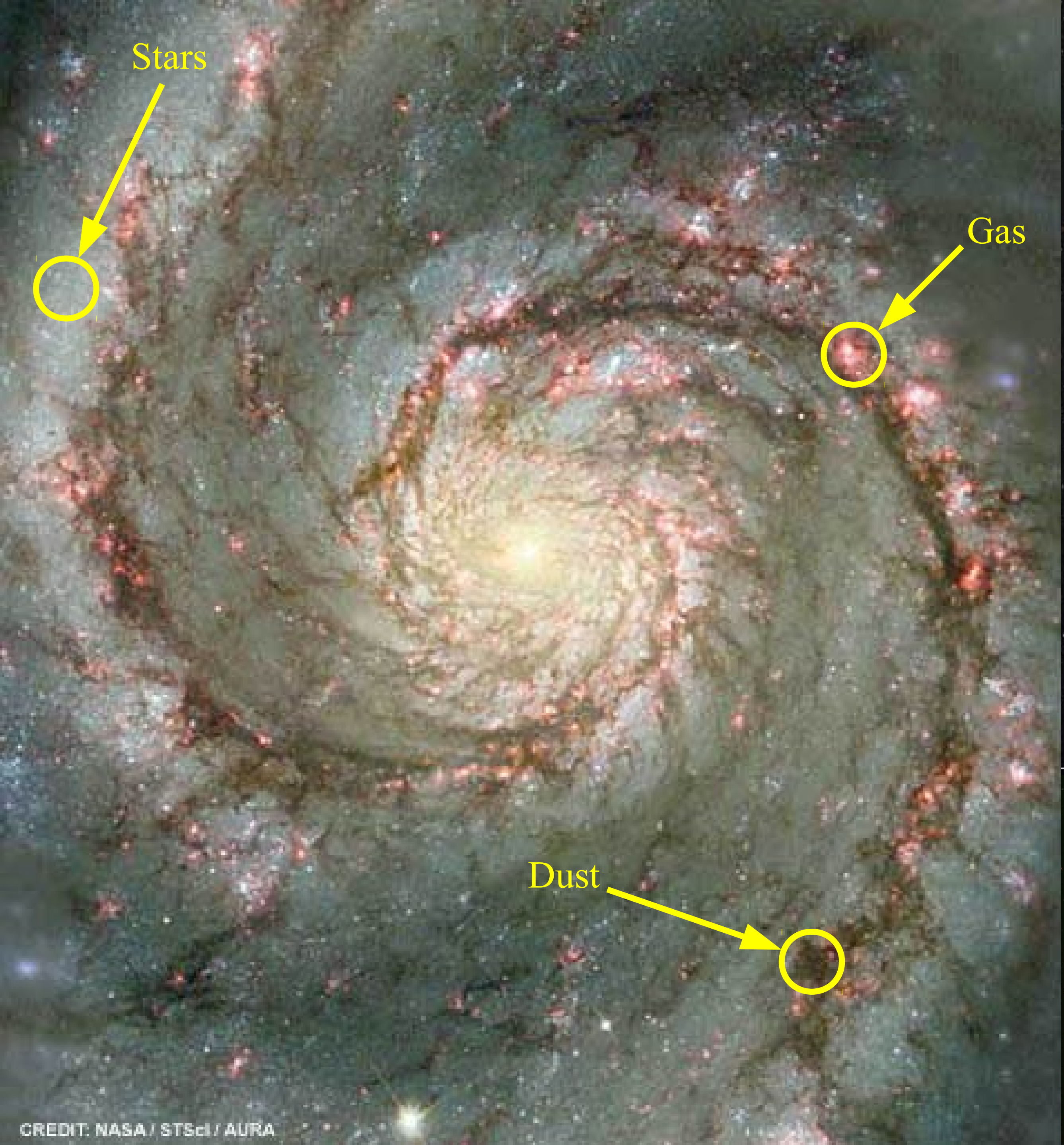
Students A. Verhamme, M. Dugué

IPAG

http://ipag.osug.fr/~fraix/ http://astrocladistics.org/

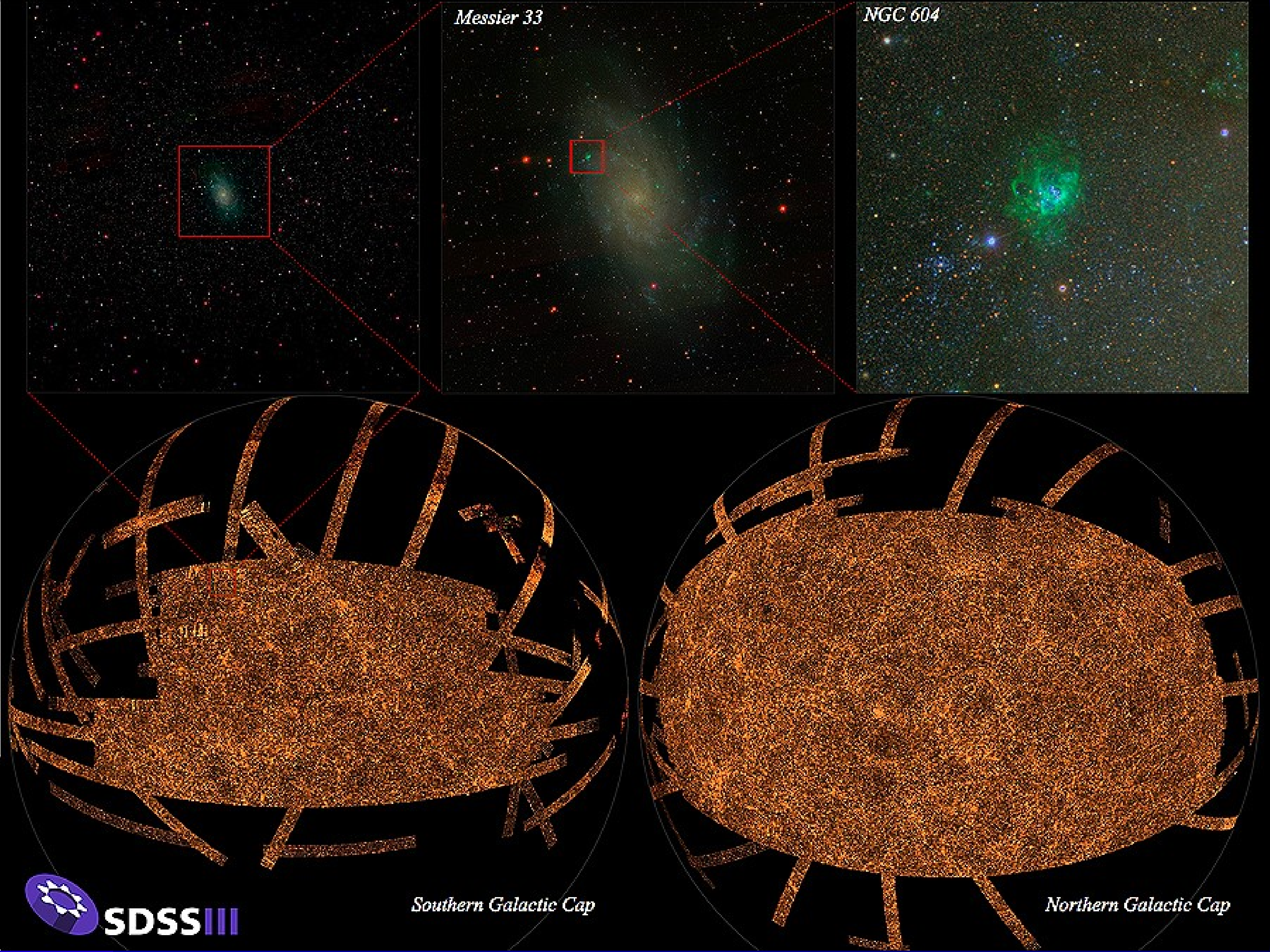
Outline

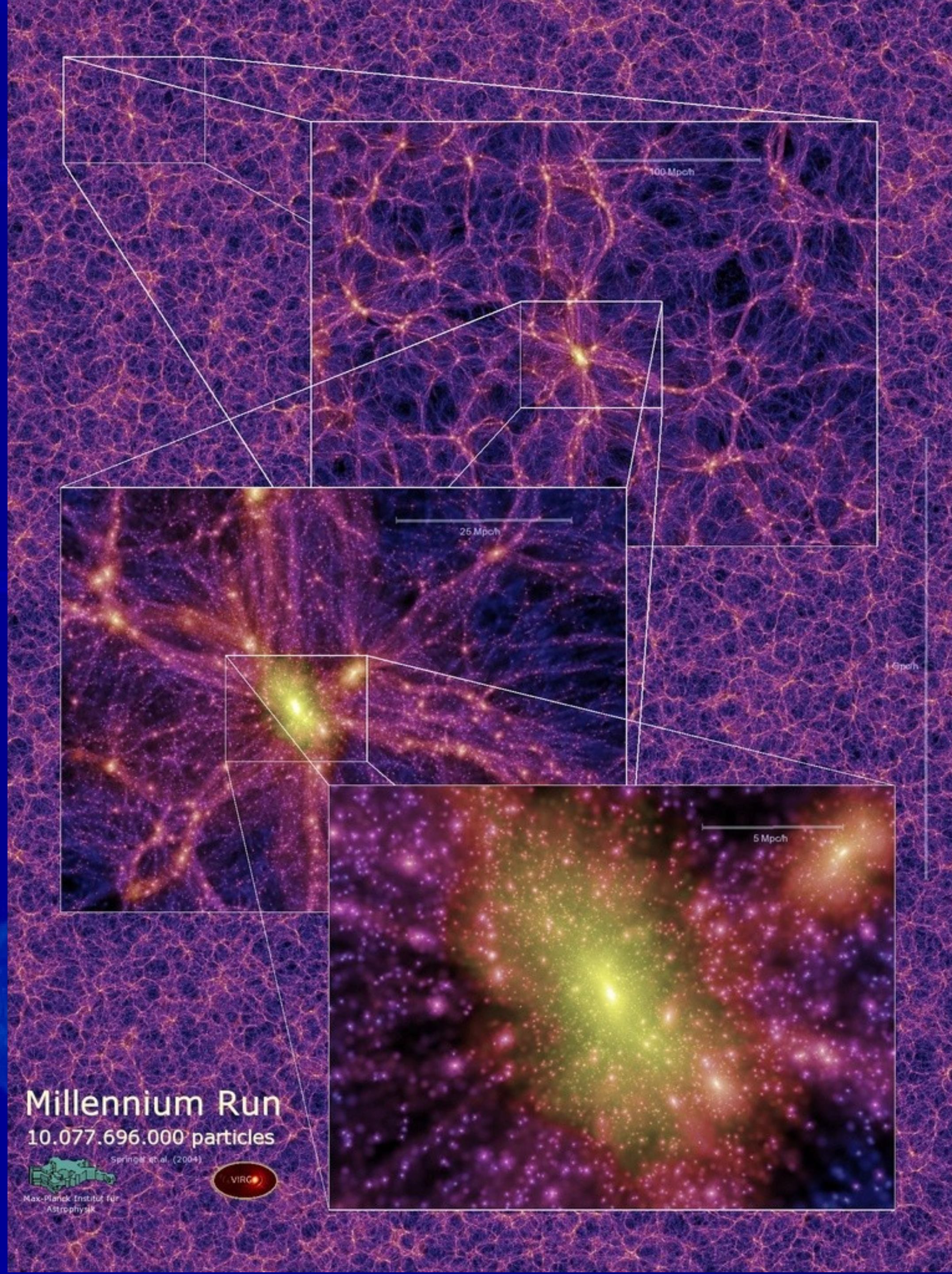
- Why classification ?
- Why be multivariate ?
- Why be evolutive ?
- Why cladistics ?
- Classification = clustering + taxonomy
- Confounding correlations
- Cladistics with continuous characters
- Spectra
- Classification and dimensionality reduction
- Astrostatistics : what's next ?


Stars
Gas
Dust
CREDIT: NASA / STScI / AURA


Interacting or merging galaxy


Dwarf galaxy


Messier 33
NGC 604
Southern Galactic Cap
Northern Galactic Cap
SDSS III

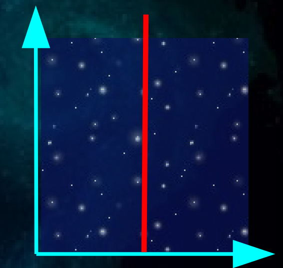

100 Mpc/h
25 Mpc/h
5 Mpc/h
1 Gpc/h
Millennium Run
10.077.696.000 particles
Springel et al. (2004)
Max-Planck Institut für Astrophysik
VIRGO

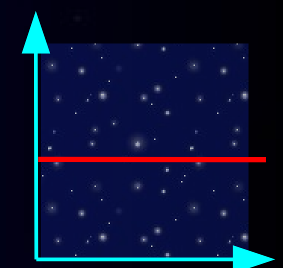
Why be multivariate?

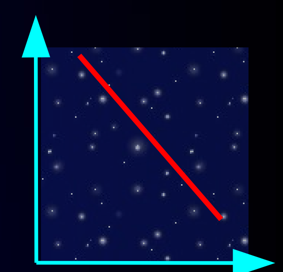
Globular clusters of our Galaxy

Table 2. Comparison between Table 1.6 of Harris (2001) and our grouping by number of clusters for each class.

		Total	G1	G2	G3
MRC	All [Fe/H] > -1	14	2	0	12
MRC	R_{gc} = 0 - 4 kpc	9	2	0	7
MRC	R_{gc} = 4 - 9 kpc	5	0	0	5
MPC	All [Fe/H] < -1	40	23	11	6
MPC	R_{gc} = 0 - 4 kpc	10	6	1	3
MPC	R_{gc} = 4 - 8 kpc	11	5	4	2
MPC	R_{gc} = 8 - 12 kpc	7	6	1	0
MPC	R_{gc} = 12 - 20 kpc	7	4	2	1
MPC	R_{gc} > 20 kpc	5	2	3	0
MPC	-2.30 < [Fe/H] ≤ -1.85	9	1	8	0
MPC	-1.85 < [Fe/H] ≤ -1.65	6	4	2	0
MPC	-1.65 < [Fe/H] ≤ -1.50	8	6	1	1
MPC	-1.50 < [Fe/H] ≤ -1.32	8	5	0	3
MPC	-1.32 < [Fe/H] ≤ -1.00	9	7	0	2
MPC	All [Fe/H] < -1.70	27	20	1	6
MPC	HBR>0, R_{gc} > 8kpc	13	7	5	1
MPC	HBR<0, R_{gc} > 8kpc	5	5	0	0
MPC	HBR<0	18	6	0	12





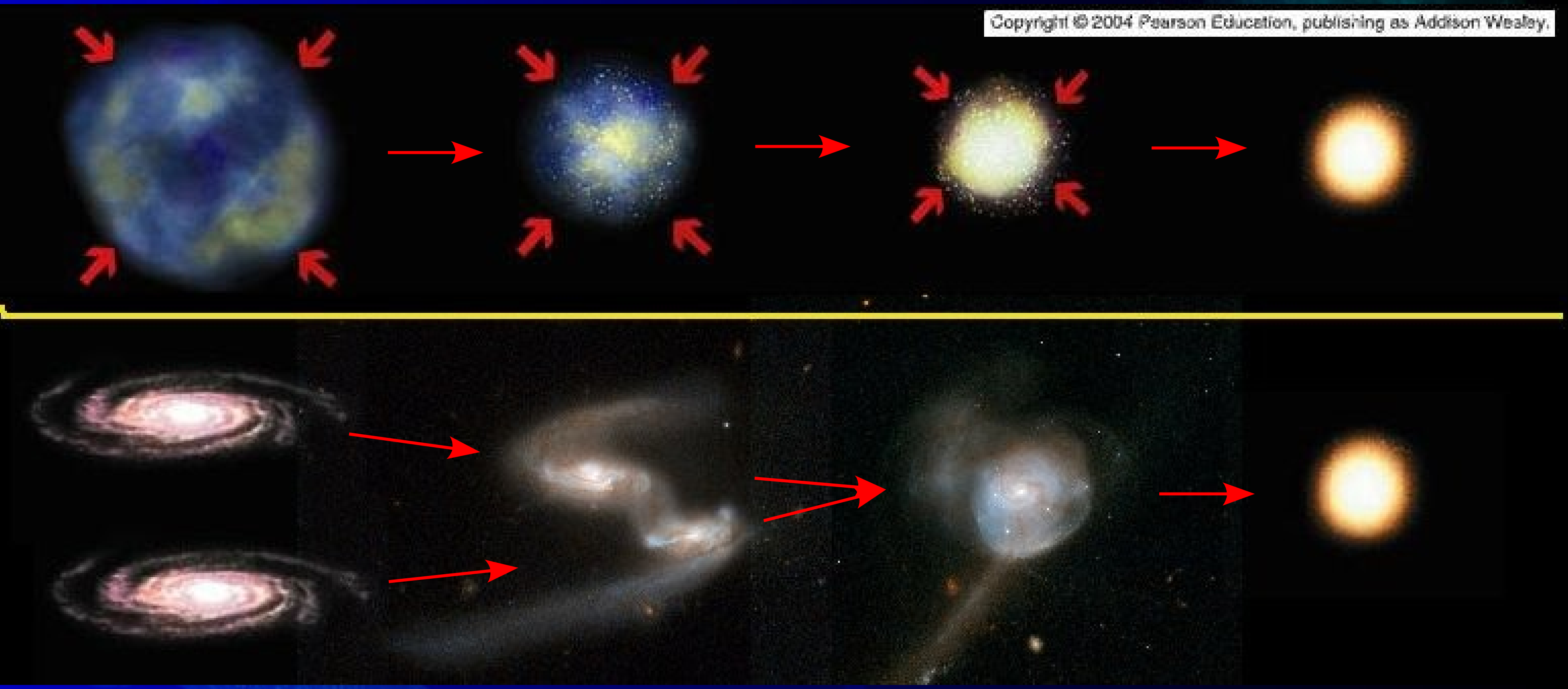


Metallicity Fe/H

R (distance from galactic center)

Horizontal Branch

Why be multivariate and evolutive?



Simple resemblance is not enough
to describe and to understand

Classification, complexity, evolution

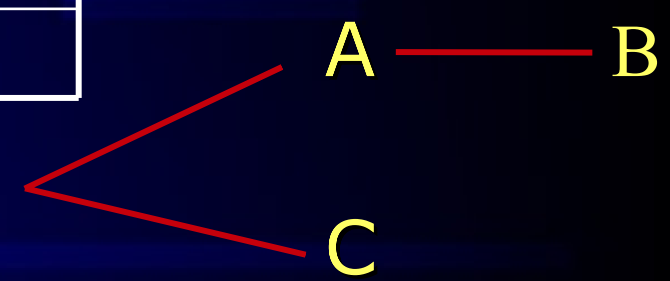
Appearance	*Few parameters*	Traditional
Global similarity	*All parameters*	Distance cluster analyses
Common history	*Evolutive characters*	Cladistics

Global similarity
(distance based)

d^2 (A,B) = 1
d^2 (A,C) = 2
d^2 (B,C) = 3

	P1	P2	P3
A	1	0	0
B	1	1	0
C	0	0	1

Common history
(character based)



Evolved states: 0 → 1 → 2

Astrocladistic analysis of Globular Clusters of our Galaxy

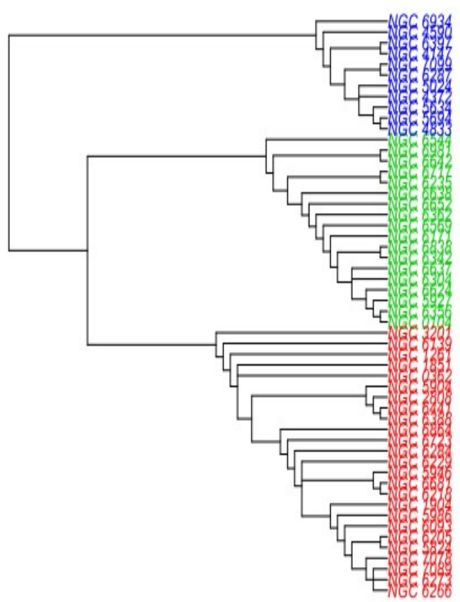
4 parameters:

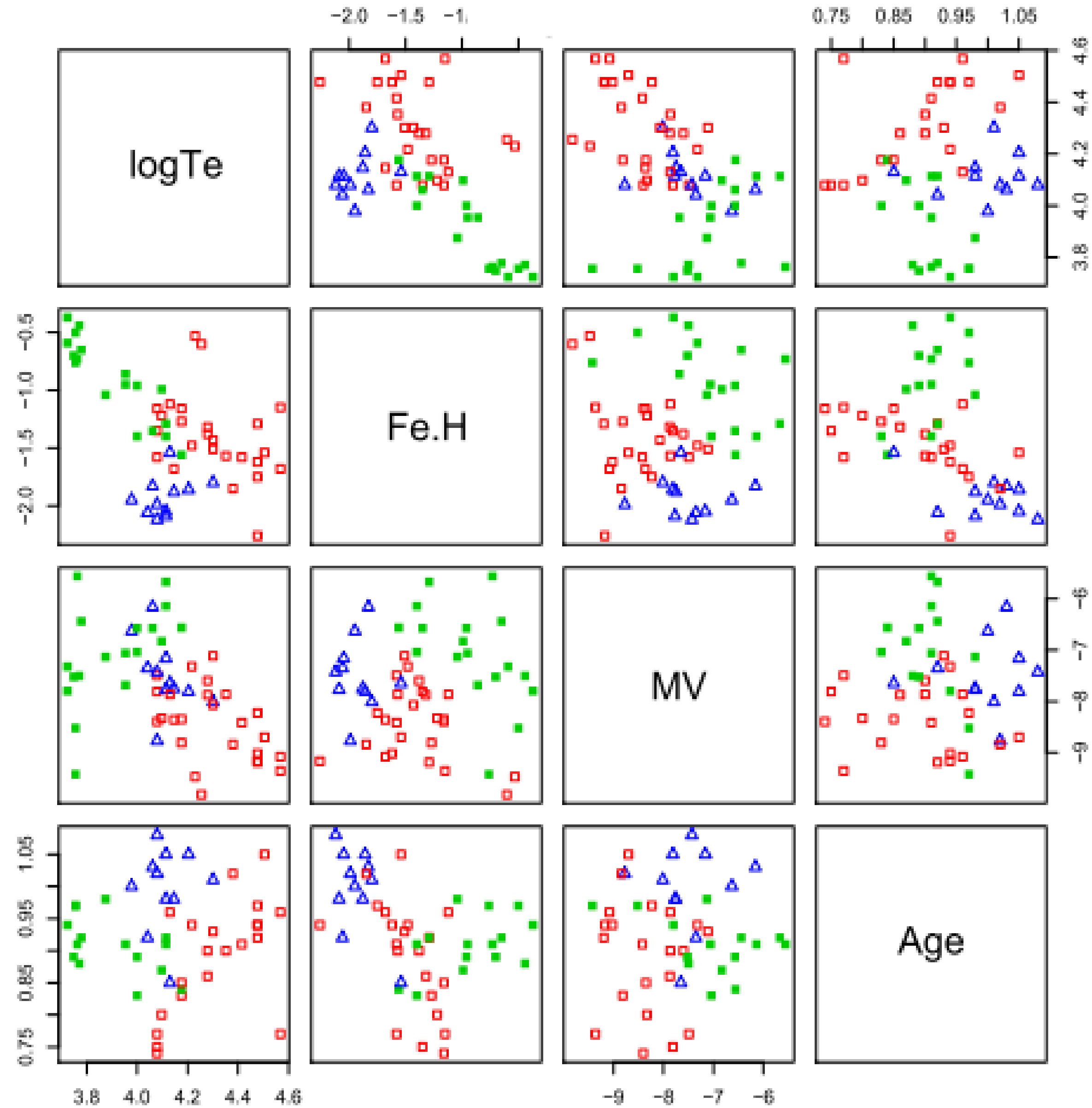
LogTe

Fe/H

Mv

Age (1/2)

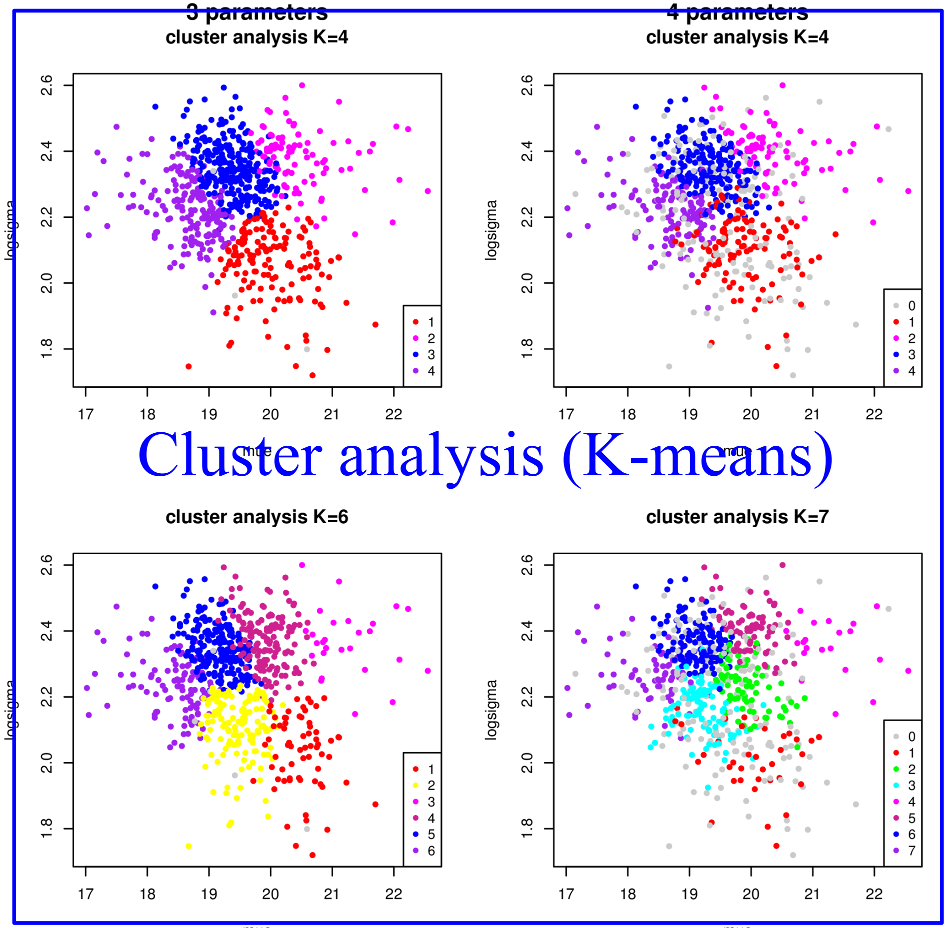


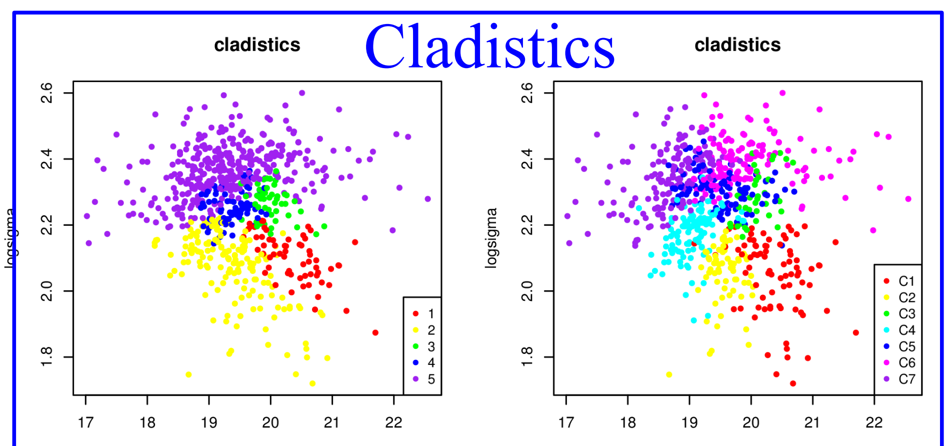


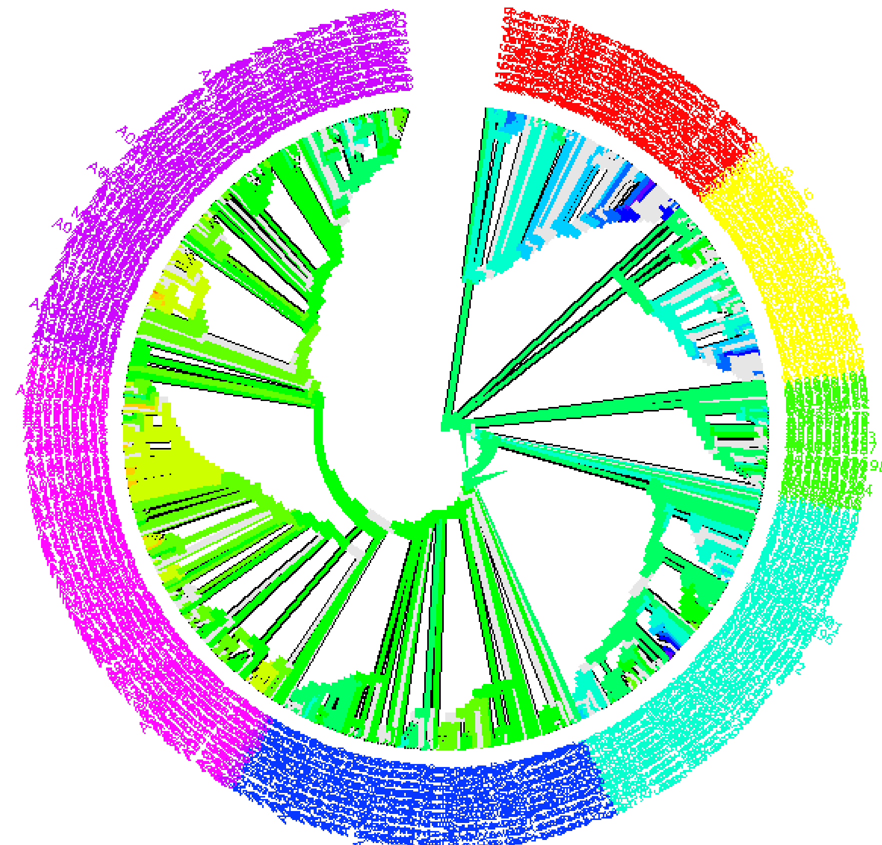
Fraix-Burnet, Davoust, Charbonnel, MNRAS 2009

The fundamental plane of galaxies

699 galaxies, 4 characters: σ, μ_e, R_e, Mg_2







Fraix-Burnet, Dugué, Chattopadhyay, Chattopadhyay, Davoust
2010 MNRAS

Confounding Correlations

Evolution as a confounding parameter

$$\begin{cases} r_e &= A_1 \widetilde{X}^p \\ \sigma &= A_2 \widetilde{X}^s \\ L &= A_3 \widetilde{X}^t \end{cases}$$

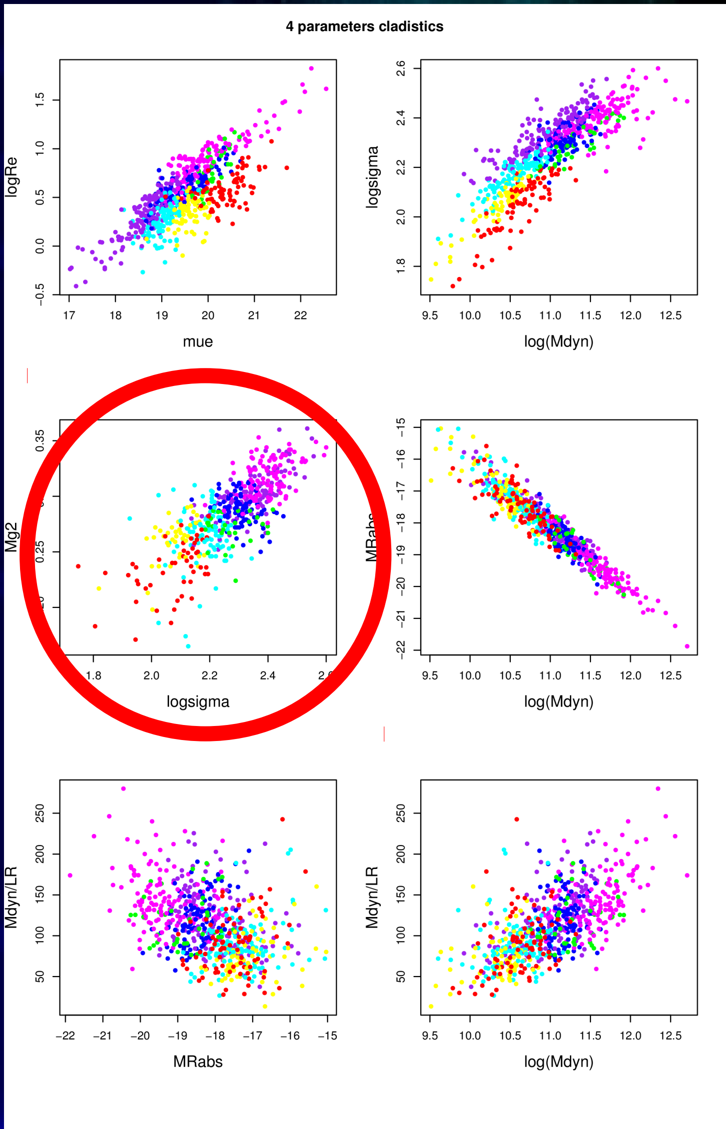
For the observed fundamental plane :

$$\log r_e = a \log \sigma + b\mu_e + c$$

$$\Rightarrow \quad p = sa + (-2.5t + 5p)\, b$$

X can be fraction of starburst, black hole mass, anything related to global evolution

If $2s + p = 1$ **then** « virial plane ».



Fraix-Burnet 2011 MNRAS

Minimum Contradiction Analysis

Continuous characters in cladistics

Tree
Split network

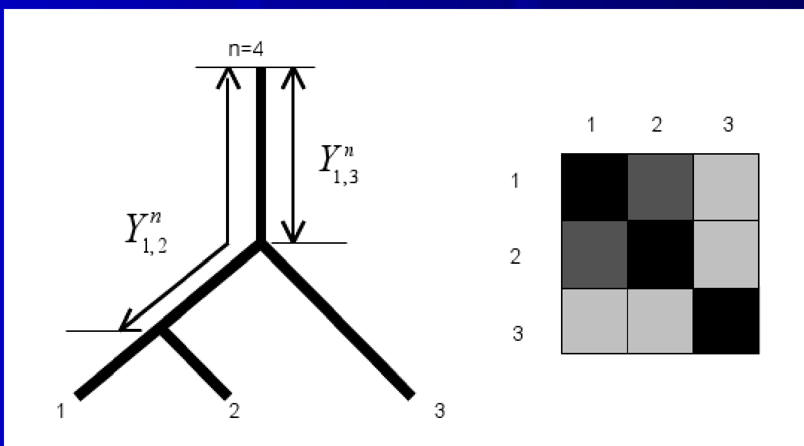
Perfect order

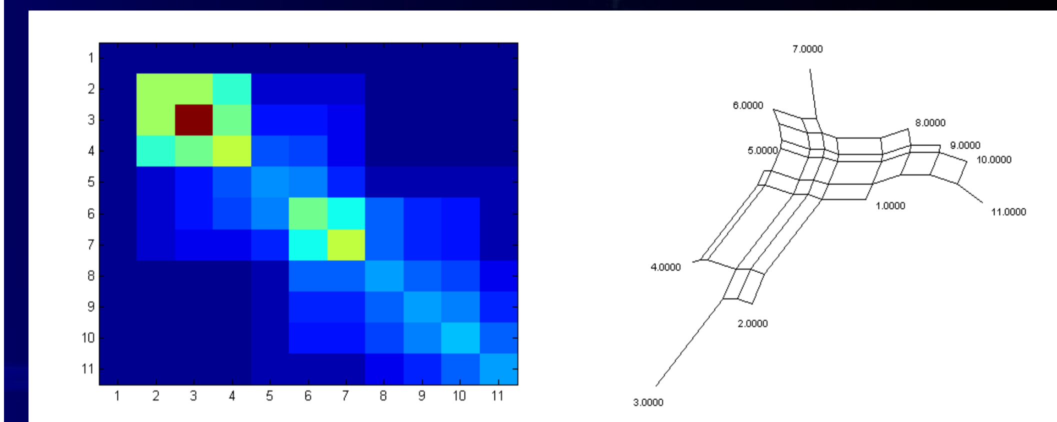
Kalmanson inequalities

$$Y^n_{i,j} \geq Y^n_{i,k},\ Y^n_{k,j} \geq Y^n_{k,i}\ (i \leq j \leq k) \text{ with } Y^n_{i,j} = 1/2 \cdot (d_{i,n} + d_{j,n} - d_{i,j})$$

Contradiction = distance to perfect order

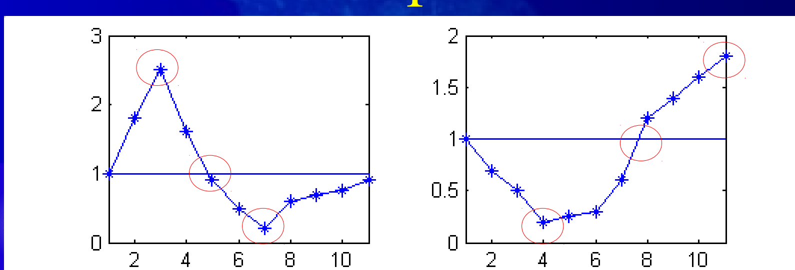
$$C = \sum_{\substack{k > j \geq i \\ i,j,k \neq n}} \left(\max\left(\left(Y^n_{i,k} - Y^n_{i,j}\right), 0\right)\right)^{\beta} + \sum_{\substack{k \geq j > i \\ i,j,k \neq n}} \left(\max\left(\left(Y^n_{i,k} - Y^n_{j,k}\right), 0\right)\right)^{\beta}$$

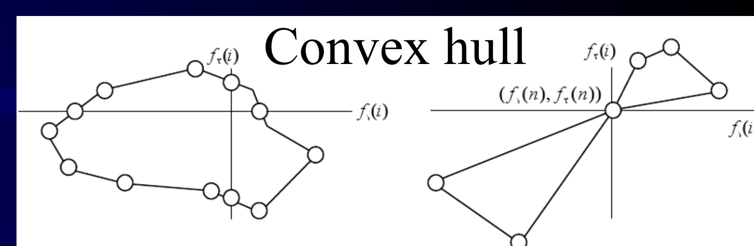




Thuillard, Fraix-Burnet, *Evolutionary Bioinformatics* 2009

Conditions on parameters

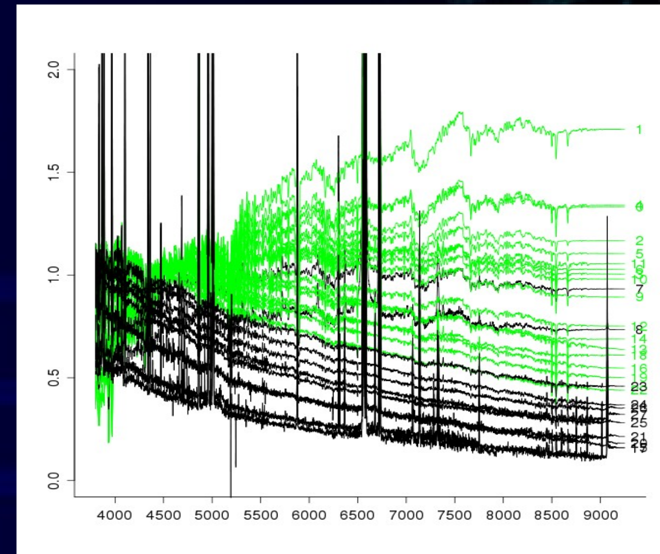




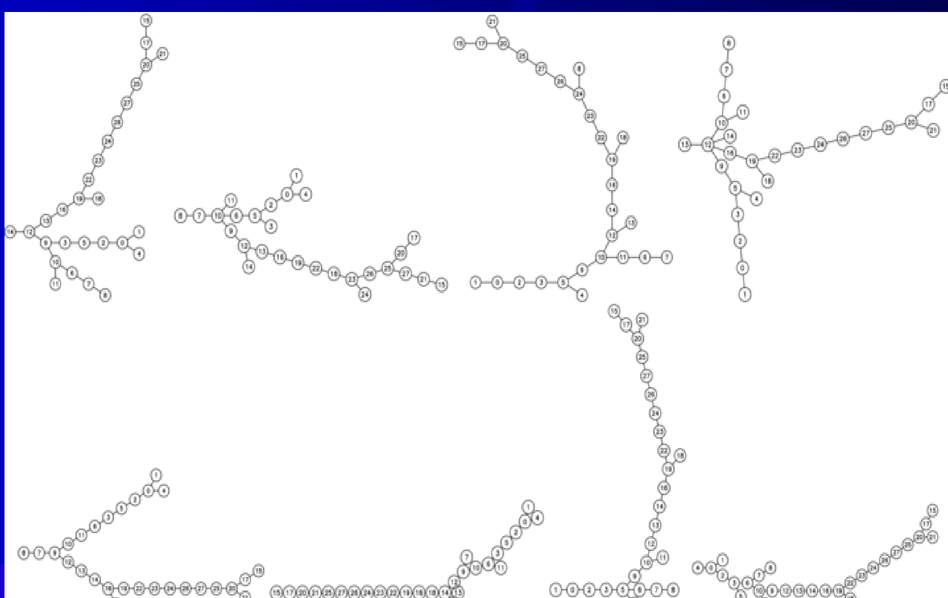
Spectra

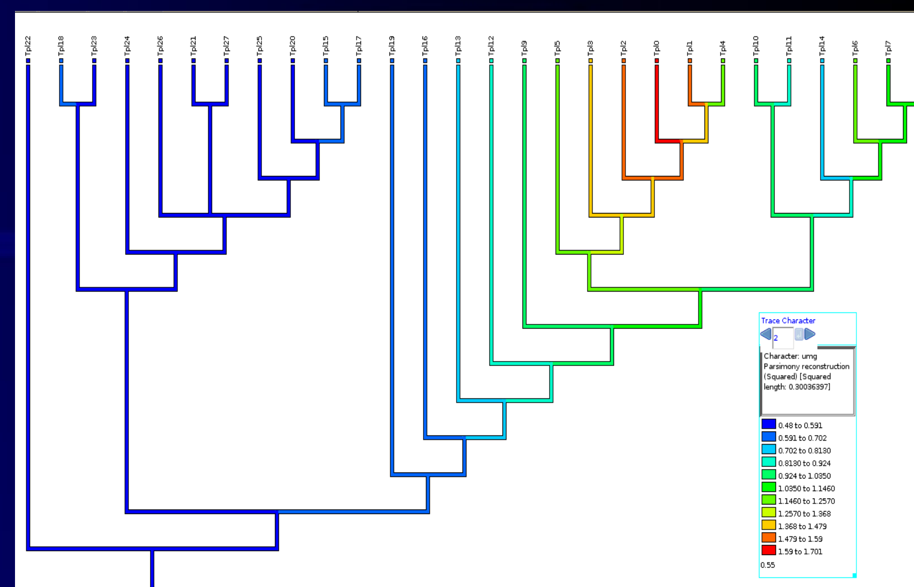
K-means on 700 000 spectra → classes

Sanchez-Almeida et al 2010
Gosh, Chattopadhyay, Fraix-Burnet

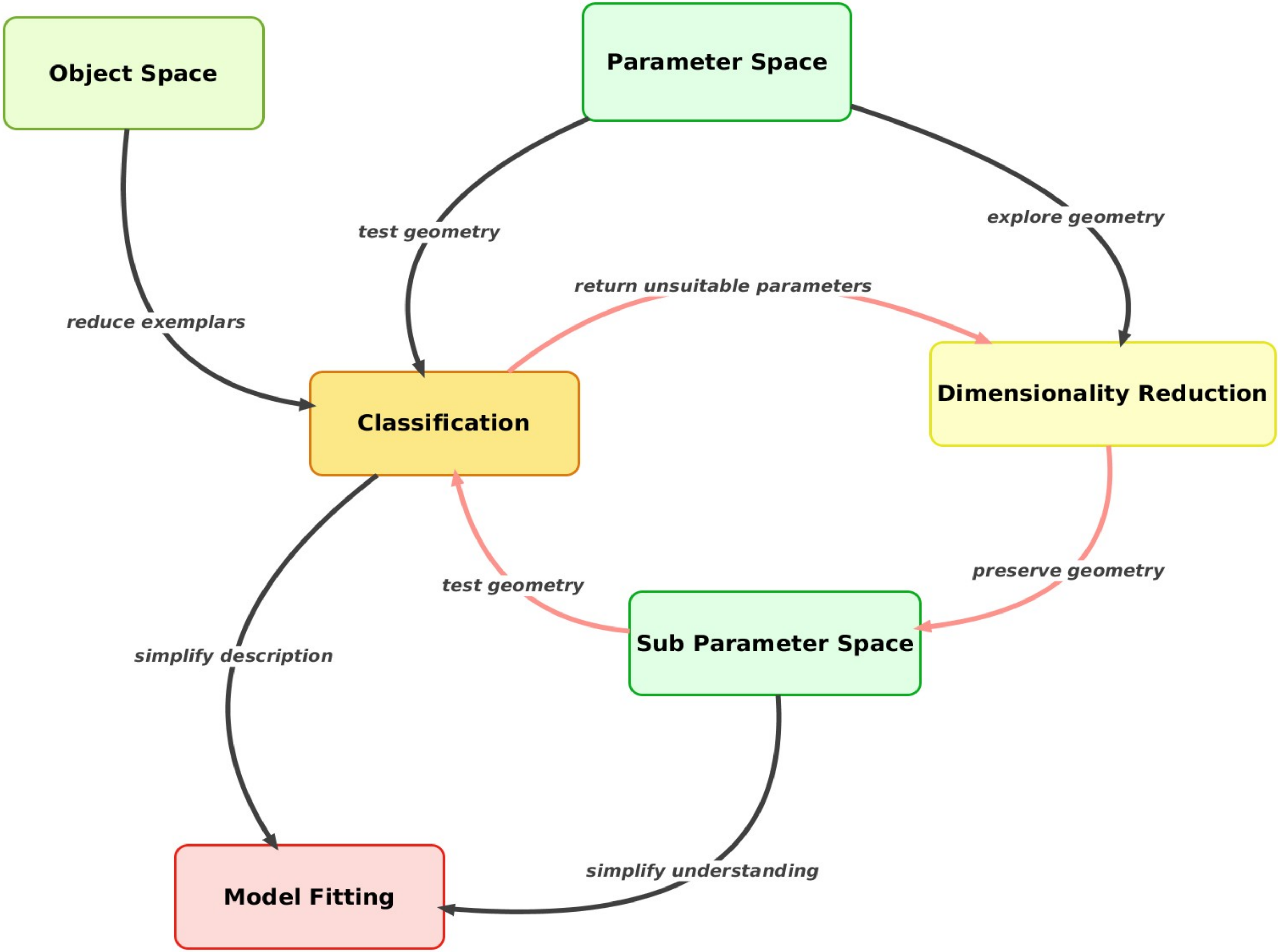


Minimum Spanning Tree or Cladistics





Ascasibar & Sanchez-Almeida 2011


Object Space
Parameter Space
test geometry
explore geometry
return unsuitable parameters
reduce exemplars
Dimensionality Reduction
Classification
preserve geometry
test geometry
Sub Parameter Space
simplify description
simplify understanding
Model Fitting

Astrostatistics : what's next ?

Training and culture:

- Statistical tools
- Statistical inference ≠ astrophysical inference

Visibility/recognition of astrostatistics

- Many works
- Many needs
- Necessity of an official structure

Animation

- Focussed and practical workshops
- Re-use methodologies (not always developments)