Big Data in Astrophysics

Highlights from current research

Benjamin Joachimi

Physics & Astronomy

b.joachimi@ucl.ac.uk

MAPS Research Festival

April 20th, 2016

Astronomical(ly) Big Data

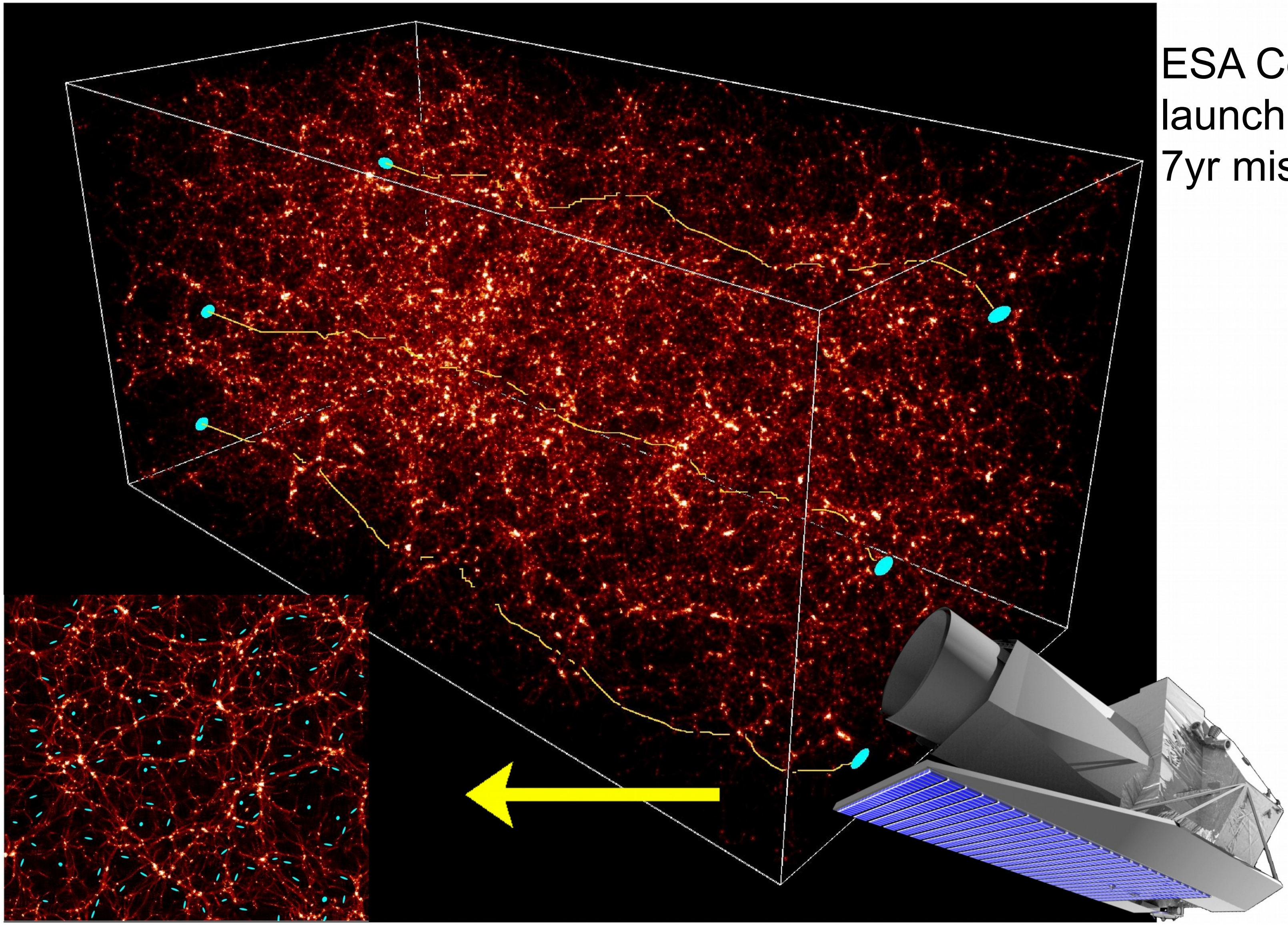
- ***volume:*** e.g. Euclid imaging survey
  15,000 deg^2 imaged on average 4 times with 0.1“ pixels
  + 8-band photometry over same area
  + extensive calibration & 2 deep fields
  → 10 PB raw space data (all of HST imaging in a few weeks)

- ***velocity:*** e.g. Large Synoptic Survey Telescope (LSST)
  ~20,000 deg^2 in 6 bands in a 10yr 'video'
  5.5 x 10^6 3.2 Gpx images in total
  → 15 TB/ day raw data rate

- ***variety:*** datasets generally heterogeneous
  missing observations; variable depth; combination of ground and space; different cameras and calibration for different bands; evolving systematics; etc. etc.

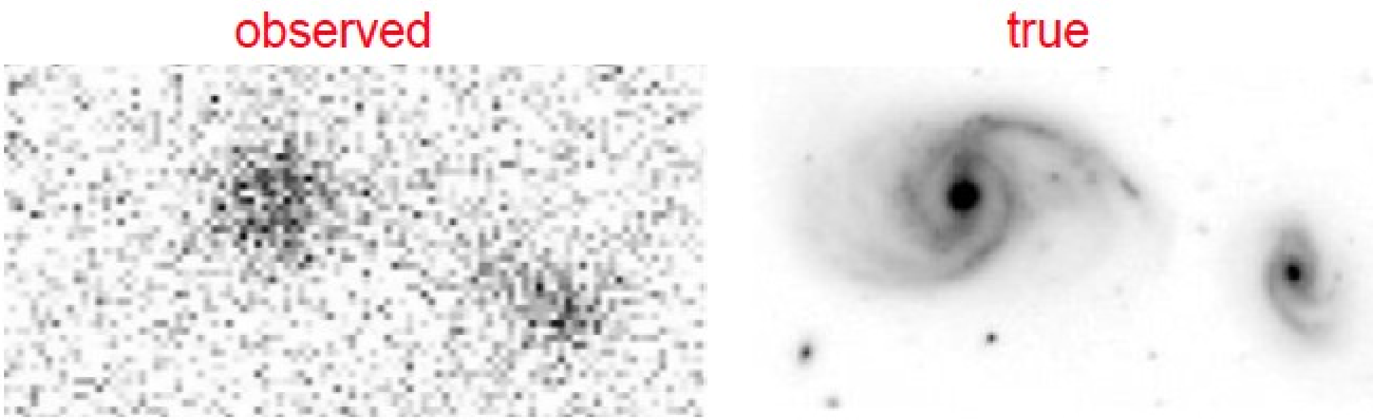
- ***veracity:*** fishing for sub-per cent variations in very low S/N data

Euclid & weak lensing



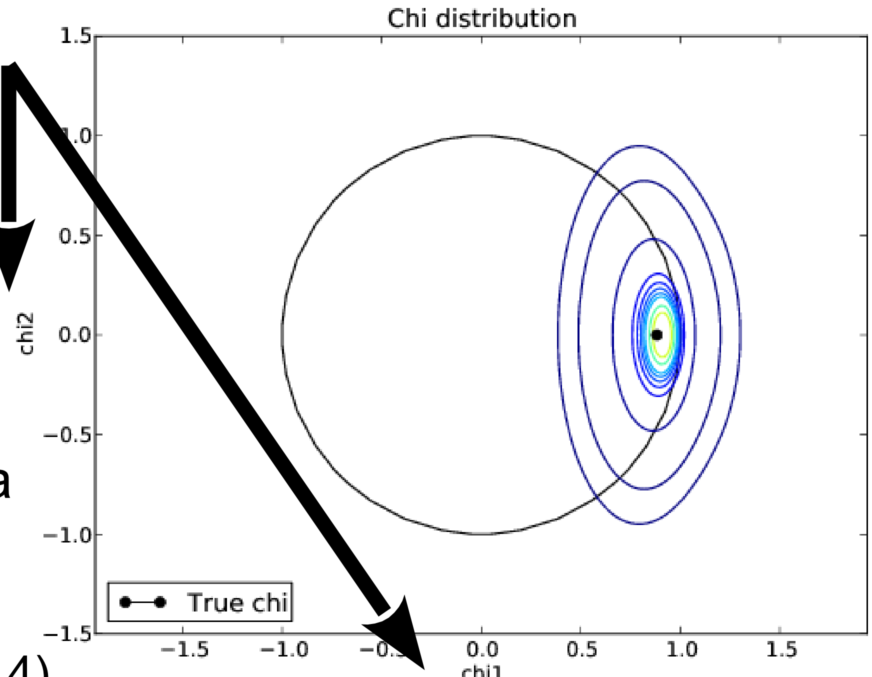
ESA Cosmic Vision
launch end 2020
7yr mission

Measuring gravitational shear



from H. Hoekstra

Cartesian components of galaxy ellipticity



- 'noise bias' in ellipticity estimates

→ need to calibrate on simulations

→ simulations require realistic galaxy parameters

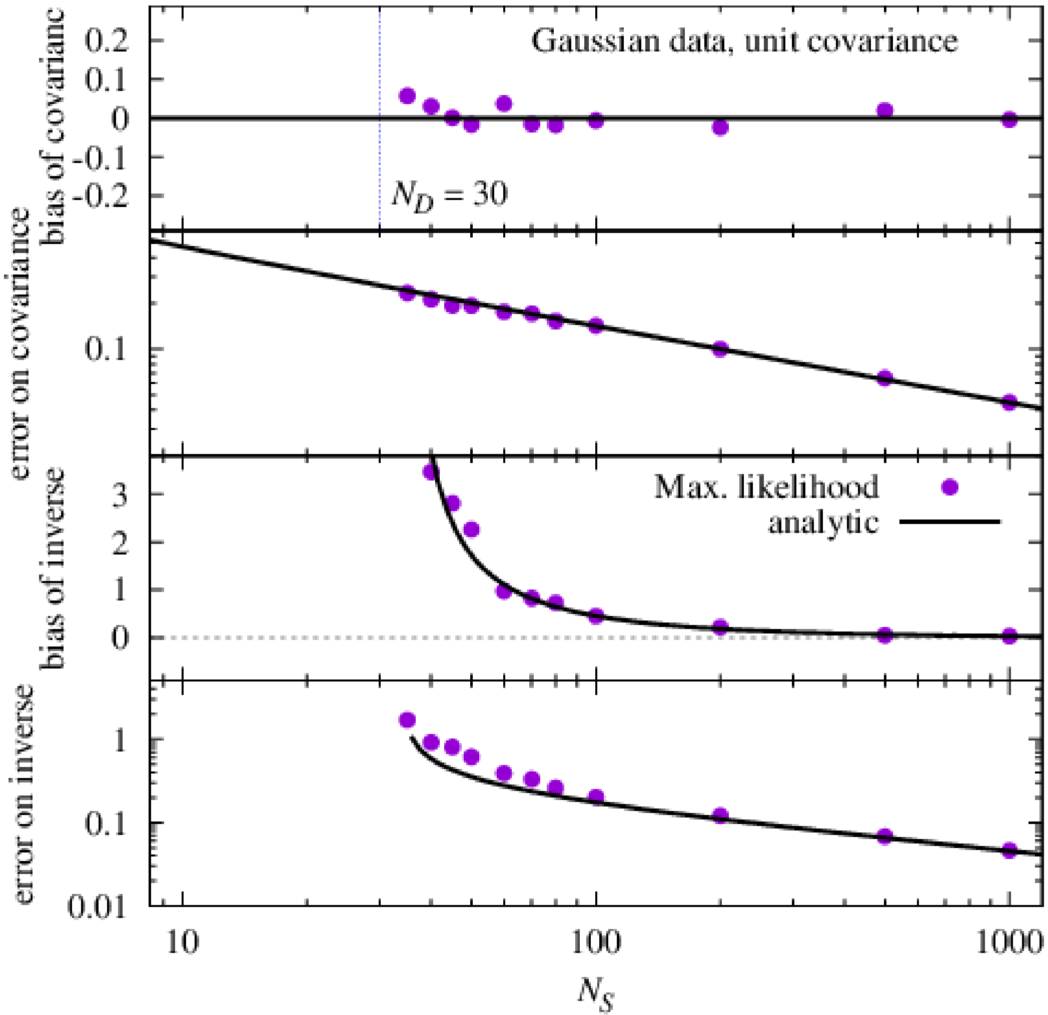
→ get these from low-noise deep data

→ drives Euclid deep fields

Viola, Kitching & BJ (2014)

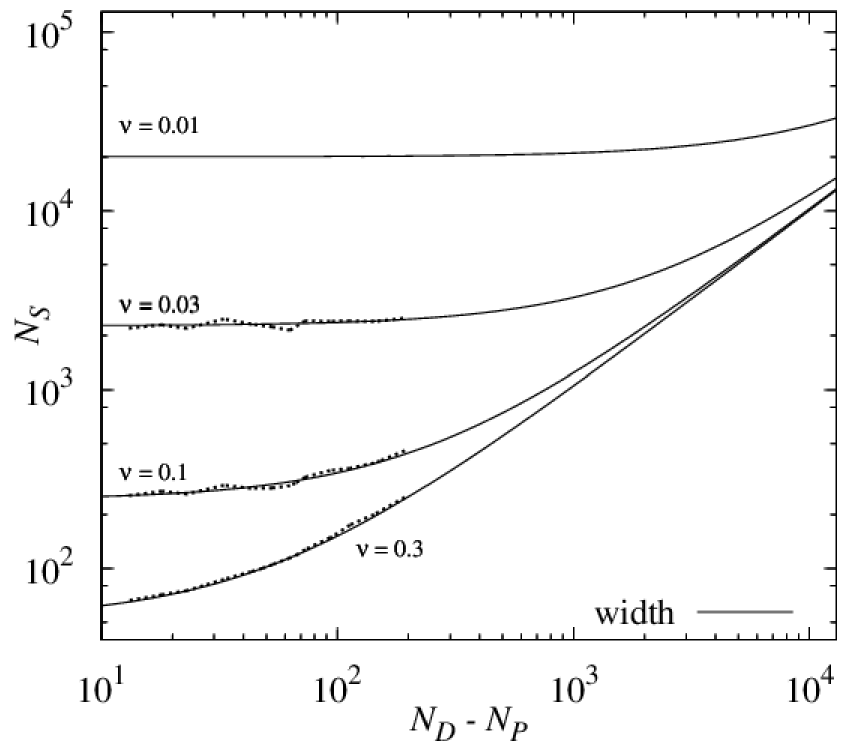
Errors on the errors

scaling of errors/biases with no. of realisations



noise bias in sample covariance matrices

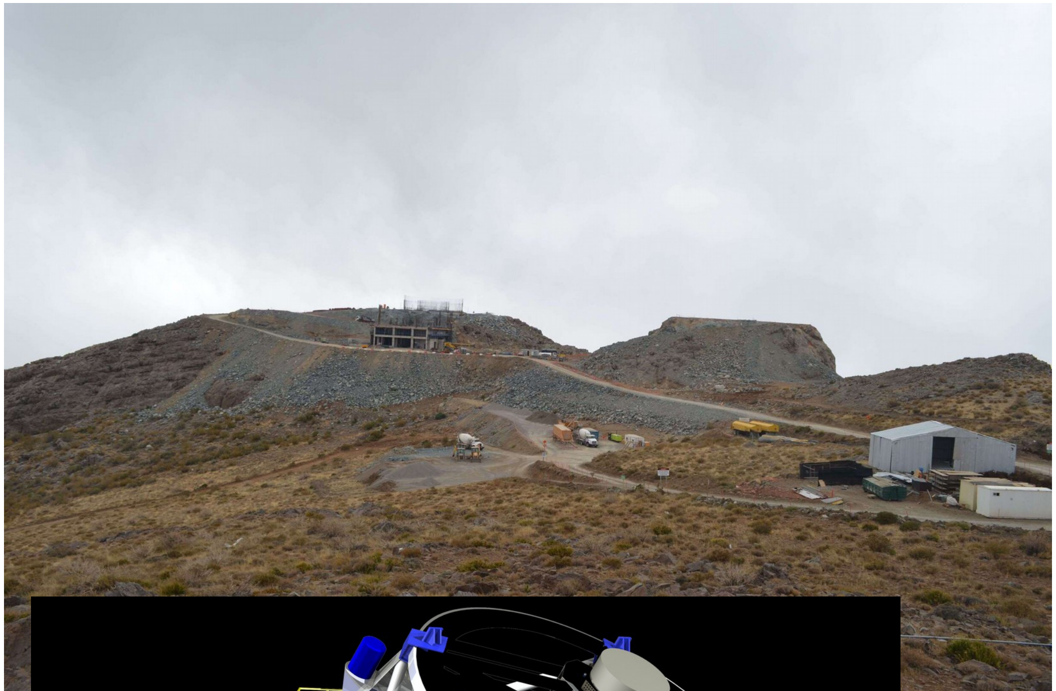
relative error in parameter covariance



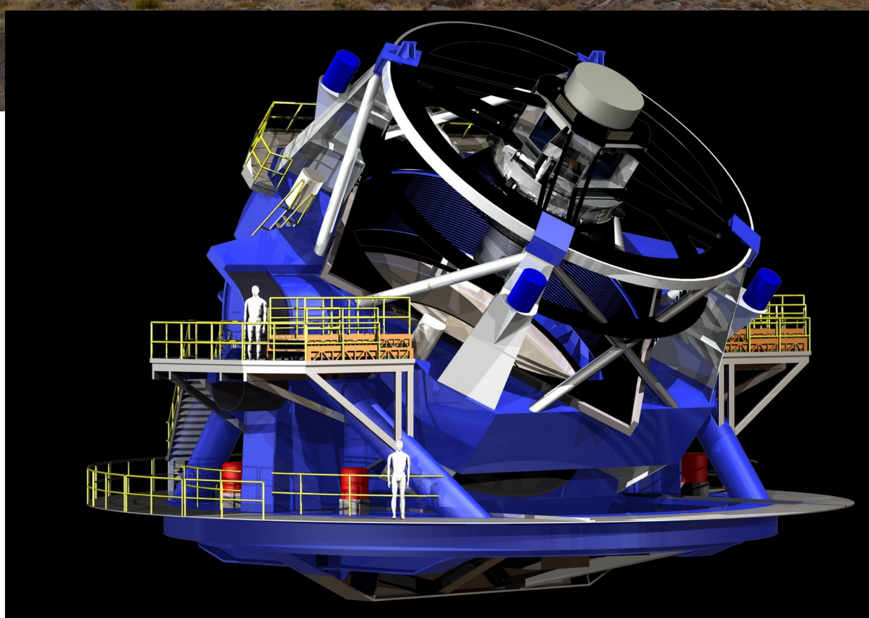
Taylor & BJ 14

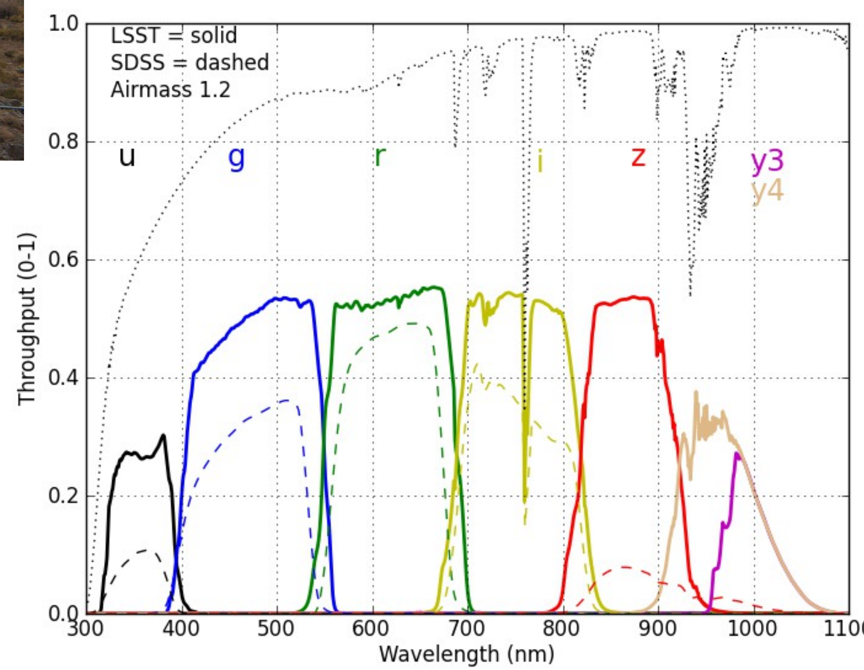
- bias and variance of inverse covariance diverge

The Large Synoptic Survey Telescope

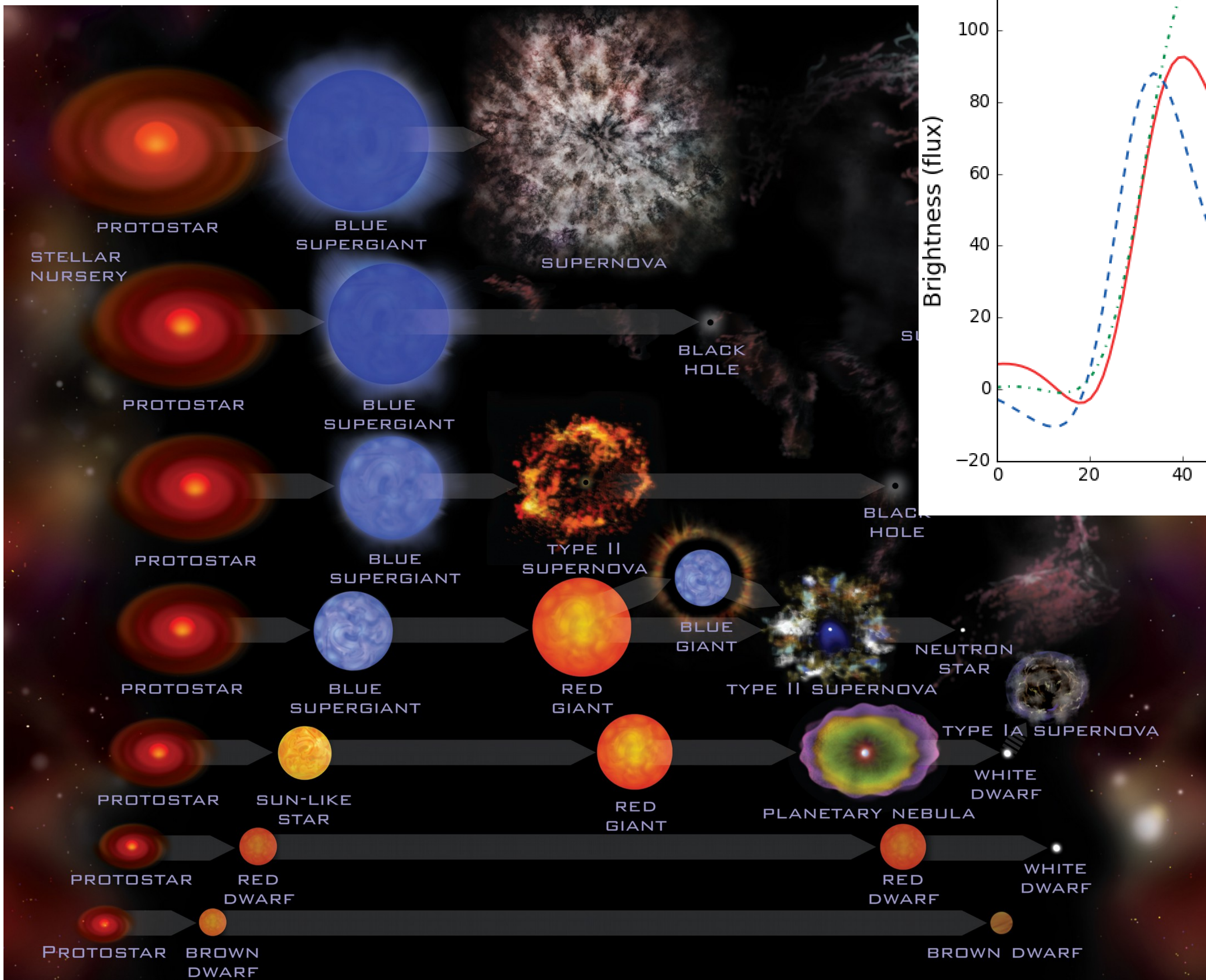


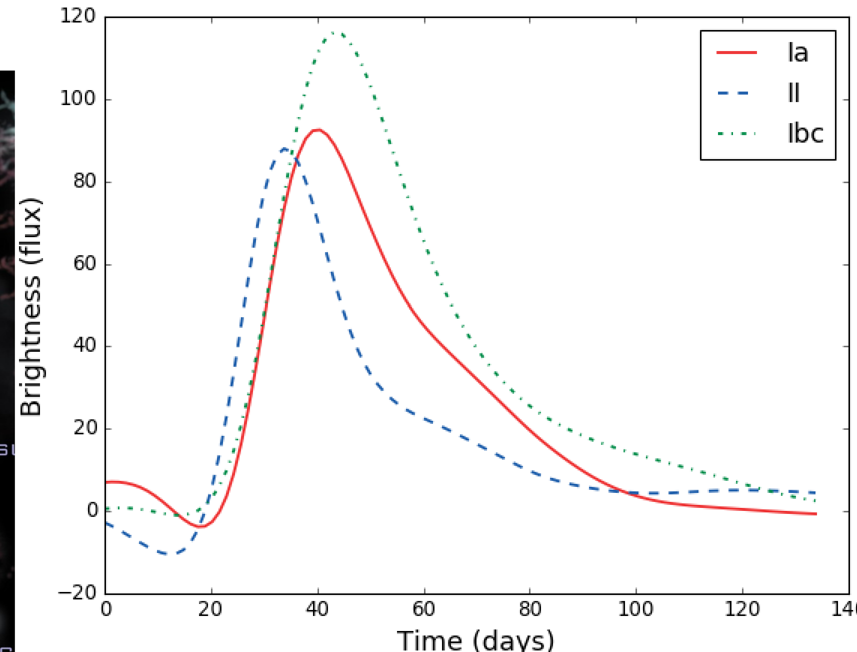
on Cerro Pachon (Chile)
full operations: 2022
8.4 m primary diameter
9.6 deg^2 field of view





Supernovae for cosmology and astrophysics



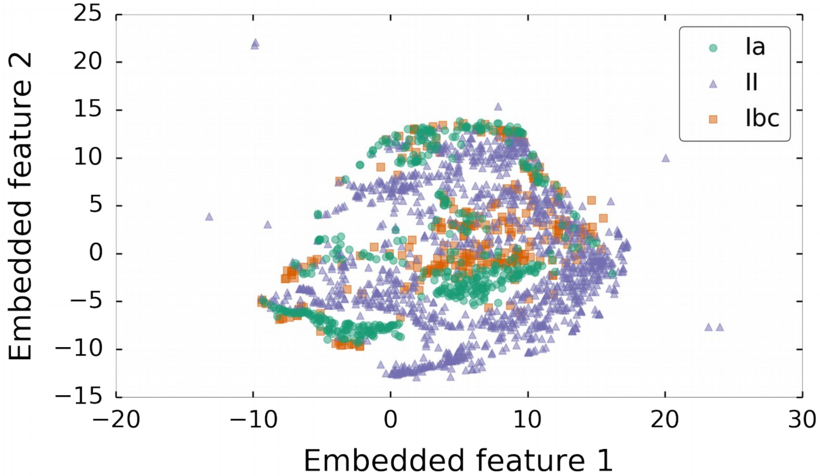


supernova lightcurves

10M 'alerts'/ night
→ ID and classification

NASA/CXC/M. Weiss

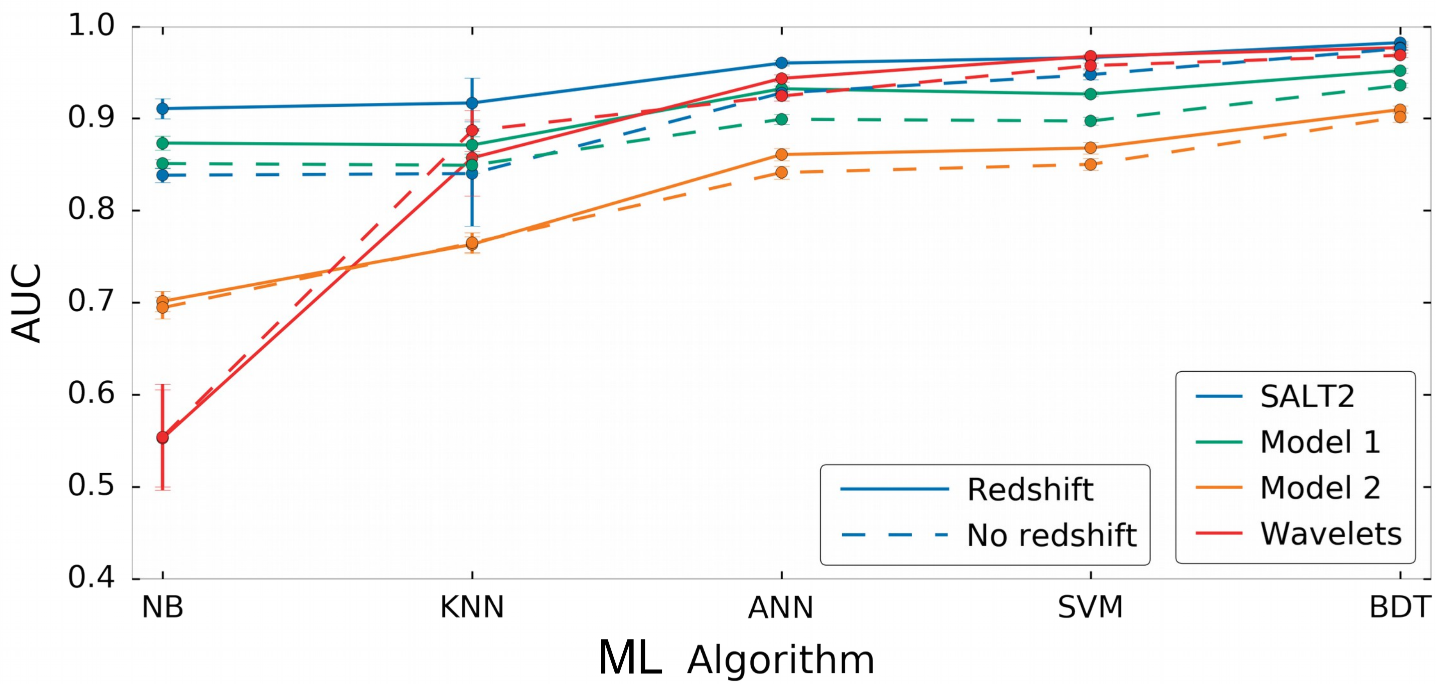
Supernova classification for LSST



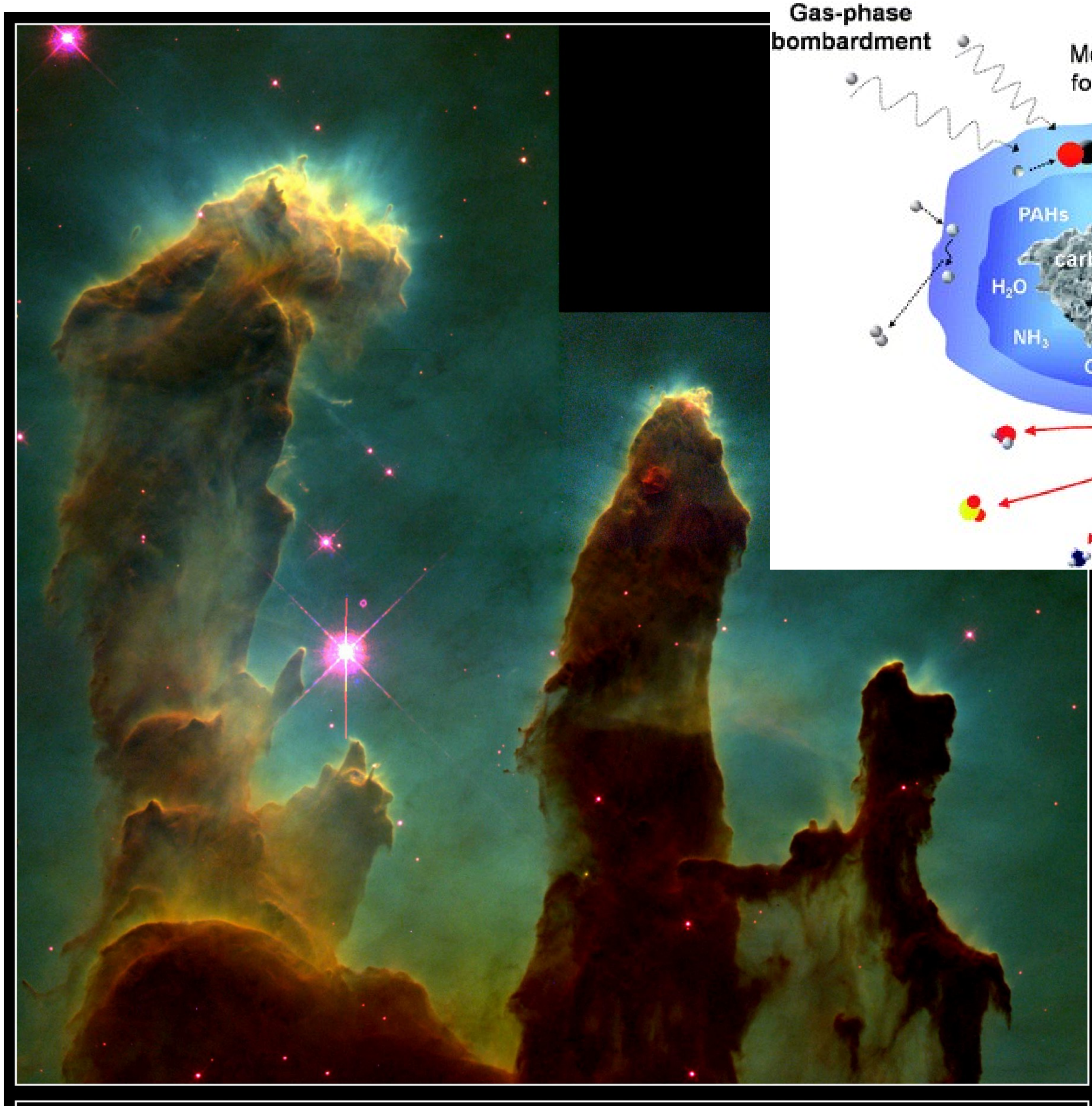
distance-preserving feature space derived from lightcurve parameters

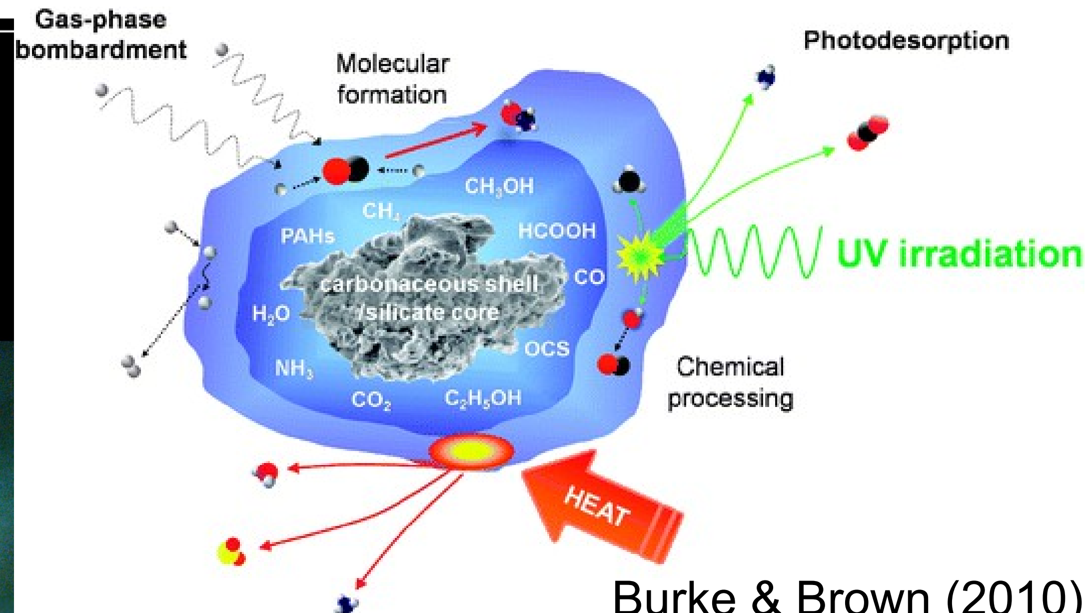
Lochner et al. (2016)

1 = perfect classification
0.5 = random assignment



A Bayesian approach to astrochemistry



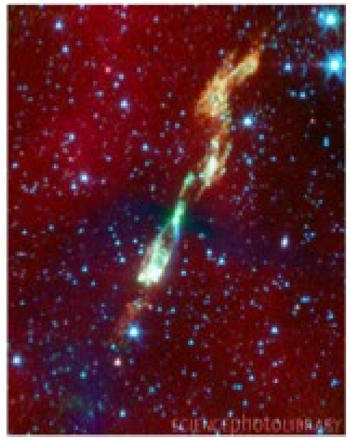


Burke & Brown (2010)

Uncover chemistry that is

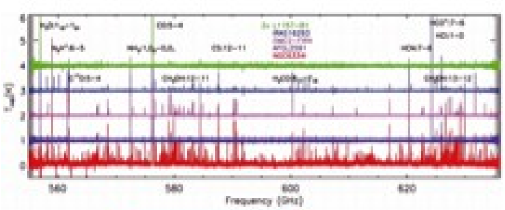
- hard to do in the lab
- key to the formation of stars
- key to the formation of life

Inference from molecular spectra

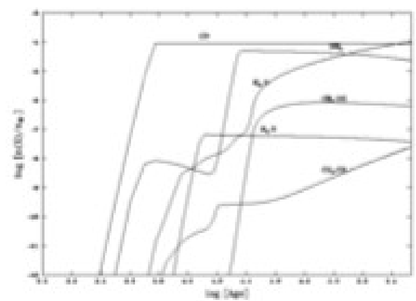


Astrophysical object

depends on density, thermodynamic state, relative abundances, etc.
→ 'Big Models'



Spectra

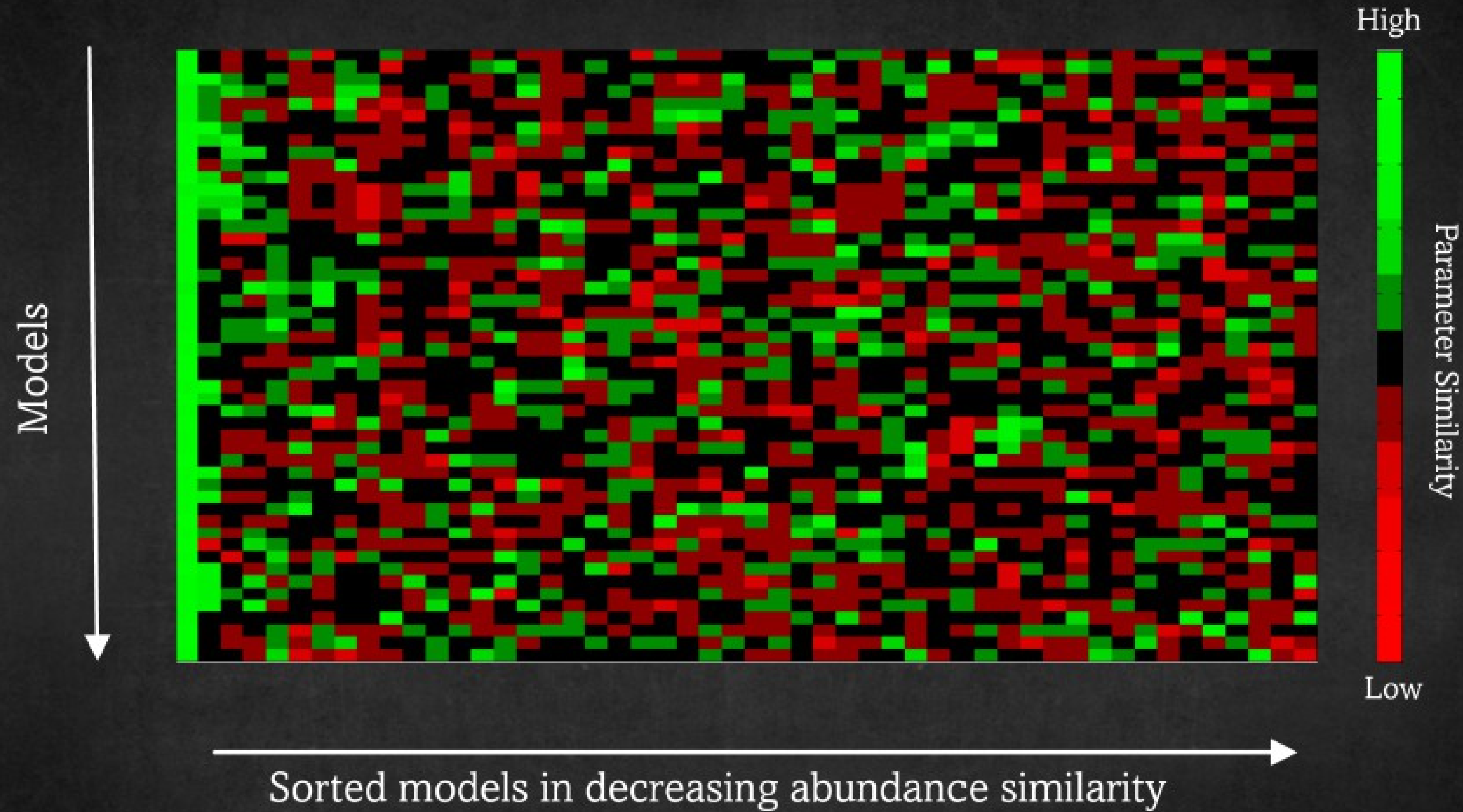


Best fit model

Ceccarelli et al. 2010

→ non-linear ill-posed inverse problem

The challenge of the inverse problem



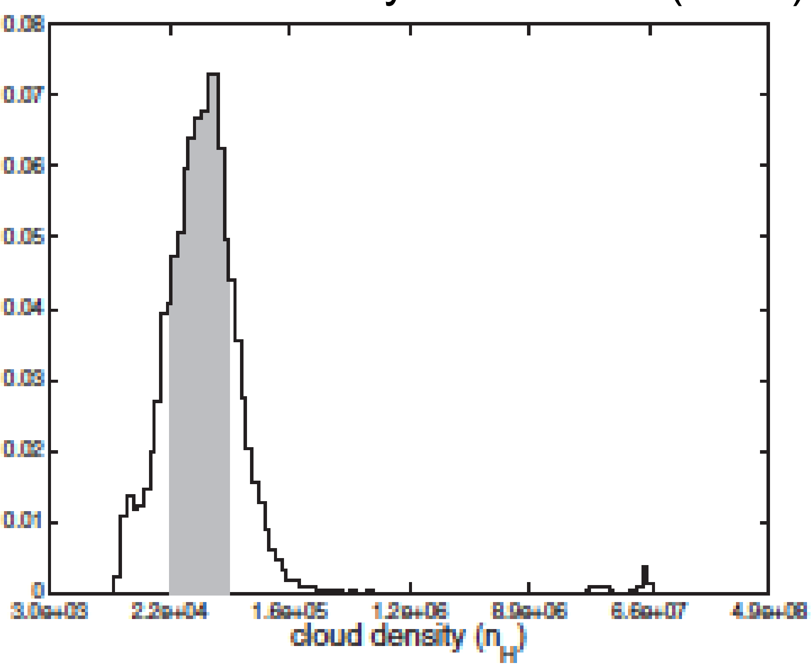
S. Viti

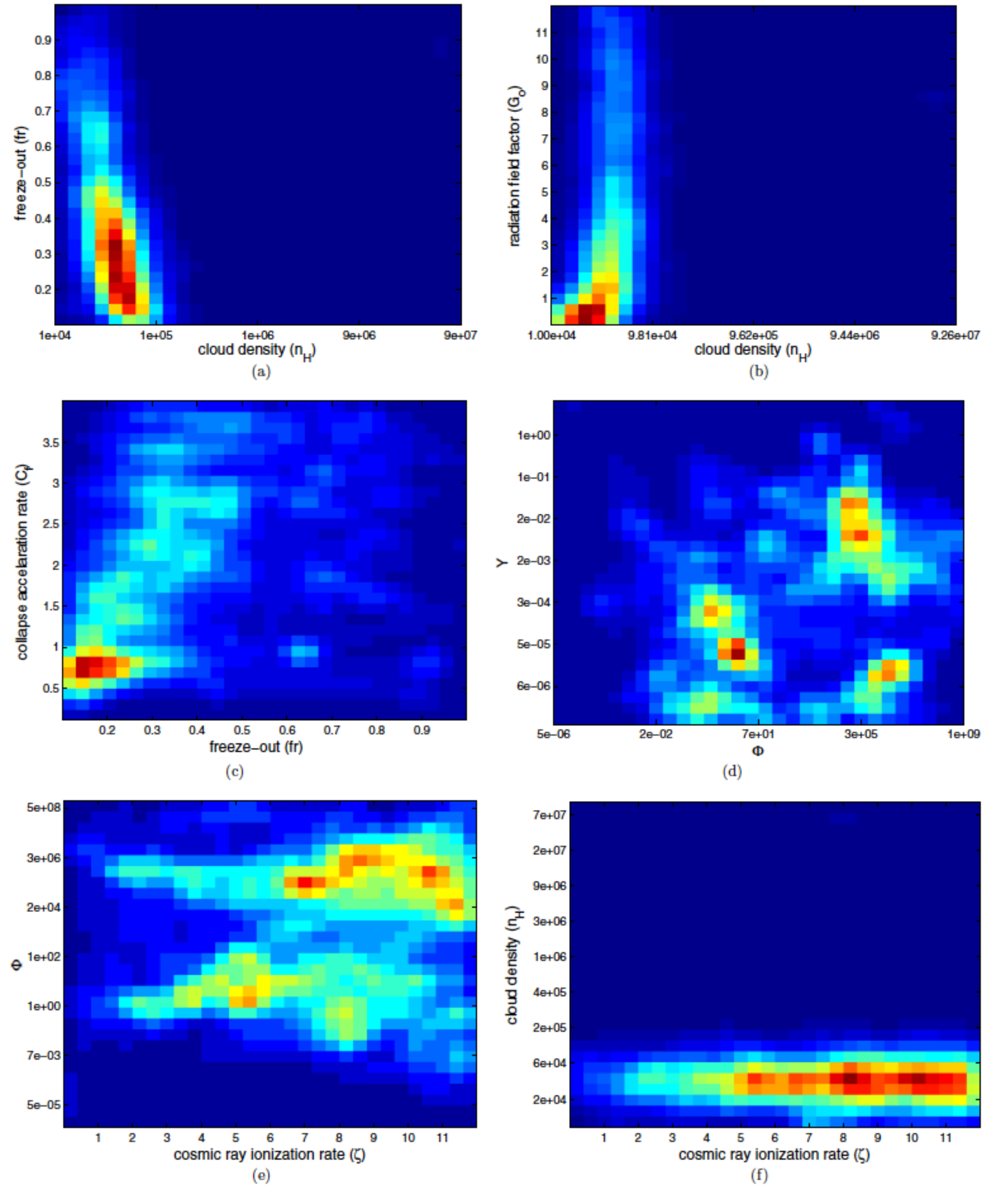
Similar parameters might not give similar abundances
OR
Similar abundances might be produced by very different parameters

Bayesian analysis

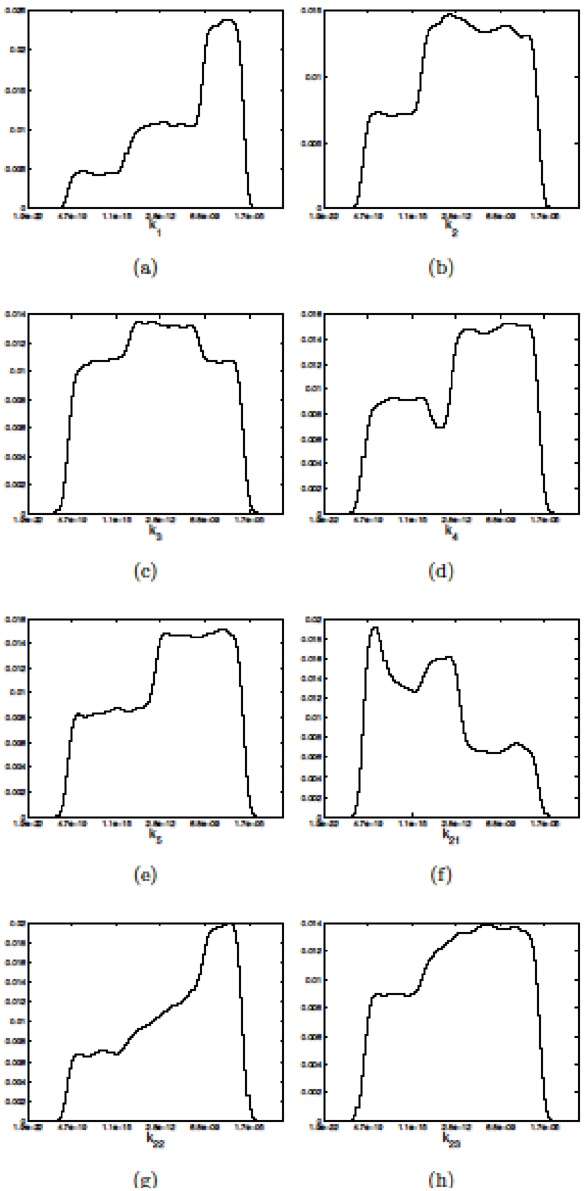
9D parameter space grid of chemical models analysed via Bayesian analysis

Makrymallis & Viti (2014)





Surface reactions in icy mantles



Bayesian analysis of
> 1M chemical models

posteriors of reaction
rate coefficients

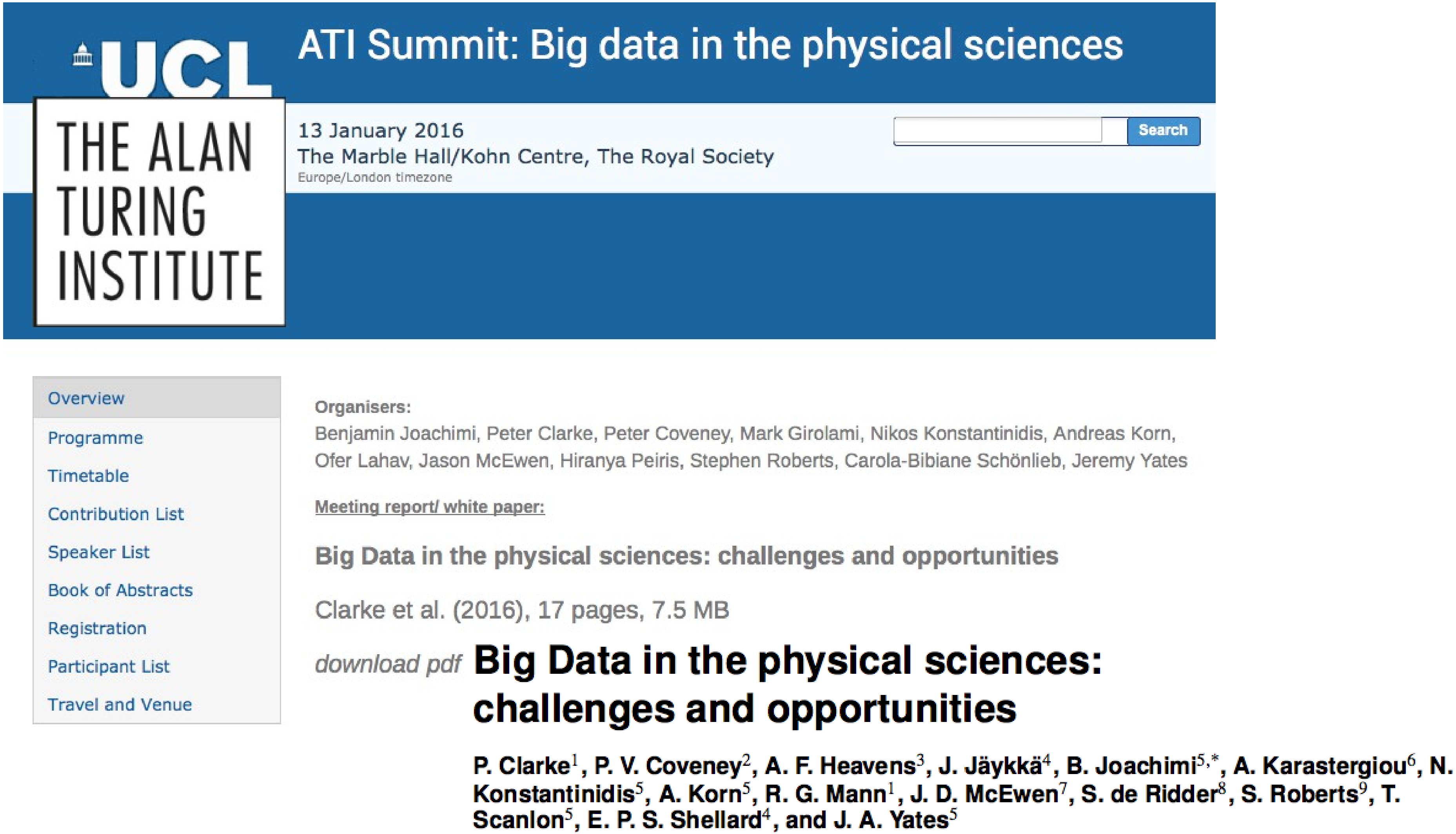
→ prioritise reactions whose
parameters vary strongly

- 8 with significant variability
- discard remaining 15

Makrymallis & Viti (2016)

No.	Reactions				
1.	O	+	H	→	OH
2.	OH	+	H	→	H_2O
3.	CO	+	OH	→	CO_2
4.	S	+	H	→	HS
5.	HS	+	H	→	H_2S
6.	H_2S	+	S	→	H_2S_2
7.	CS	+	H	→	HCS
8.	HCS	+	H	→	H_2CS
9.	CO	+	S	→	OCS
10.	OCS	+	H	→	HOCS
11.	H_2S	+	CO	→	OCS
12.	H_2S	+	H_2S	→	H_2S_2
13.	H_2S_2	+	CO	→	CS2 + O
14.	H_2S	+	O	→	SO_2
15.	CS_2	+	O	→	OCS + S
16.	CO	+	HS	→	OCS
17.	S	+	O	→	SO
18.	SO	+	O	→	SO_2
19.	SO	+	H	→	HSO
20.	HSO	+	H	→	SO
21.	CO	+	H	→	HCO
22.	HCO	+	H	→	H_2CO
23.	H_2CO	+	H	→	CH_3OH

Big Data in the physical sciences



https://indico.cern.ch/event/449964