Reshaping scientific interdisciplinary collaborations

Rafael S. de Souza

Eotvos Lorand University, Budapest, Hungary
University of Sao Paulo, SP, Brazil

My credentials

Head Cosmostatistics Initiative

Vice-President International Astrostatistics Association

Co-editor Astrostatistics & Astroinformatics Portal

Author Bayesian Models for Astrophysical Data

CAMBRIDGE UNIVERSITY PRESS

Member BIGSKYEARTH

Big Data Era in Sky and Earth Observation

TD COST Action TD 1403

Some Perspective

Joint statistical organizations

The International EnvironMetrics Society - TIES

IBS The International Biometric Society

60 (1954-2014) years IEA International Epidemiological Association

iSCB International Society for Computational Biology

International Association for Mathematical Geosciences 1968

Brief History of International Astrostatistics Association

Jul 2008 - ISI Astrostatistics Interest Group formed (Hilbe)
Dec 2009 - ISI Astrostatistics Committee approved (Hilbe)
Jan 2010 - ISI International Astrostatistics Network formed (Hilbe)
Feb 2012 - ASAIP started
(Feigelson & Hilbe editors, incl. de Souza, R. from 2014)
Jun 2012 - AAS approved the
Astrostatistics & Astroinformatics Working Group (Ivezic)
Aug 2012 - IAU approved the
Astrostatistics & Astroinformatics Working Group (Feigelson)
Aug 2012 - Aug 30th , IAA officially established (Hilbe)
Mar 2014 - American Statistical Assoc. approved the
Astrostatistics interest group (Cisewski)
Apr 2014 - IAA Cosmostatistics Initiative formed (de Souza, R.)

An acronym made by astronomers...
…outdated but we keep using it anyway

Cosmostatistics

Initiative

An acronym made by astronomers...

…outdated but we keep using it anyway

Astro

**~~Geo~~statistic
s**

Initiative

An acronym made by astronomers...
…outdated but we keep using it anyway

Astro

~~Geo~~statistics

+ computer science

~~Initiative~~

High-end education
Social experiment

The Cosmostatistics Initiative (COIN), an international working group built under the umbrella of the International Astrostatistics Association (IAA), aims to create an interdisciplinary environment where collaborations between astronomers, statisticians and machine learning experts can flourish.

Projects are defined given the people interested in collaborating, and not the other way around!

What is its goal?

Long term goal:

Establish Astrostatistics as a discipline on its own.

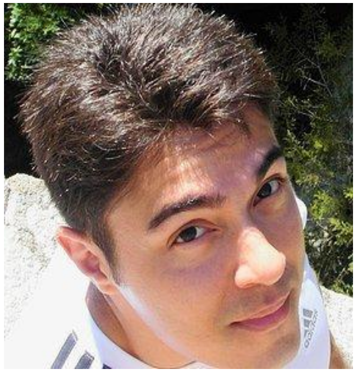
Short term goal:

Make astronomers, statisticians, computer scientists and machine learning experts understand each other ... WHILE doing science......

Why is it different?

Collaboration as a goal by itself

A Brazilian approach to science development



Rafael S. de Souza
(head) - statistics

Alberto Krone-Martins
astrometry

Emille E. O. Ishida
SN cosmology

2014

STATISTICAL CHALLENGES in 21st CENTURY COSMOLOGY
IAU SYMPOSIUM 306 Lisbon Portugal 25-29 May 2014

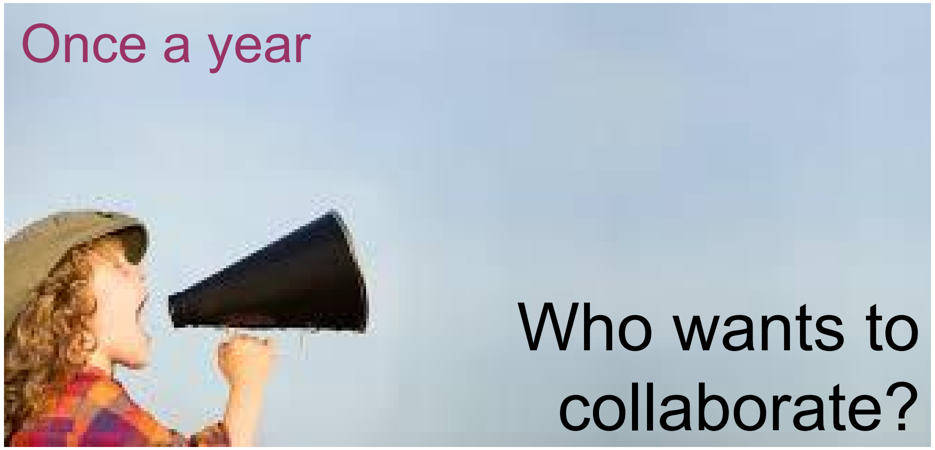
Monday

Session: CMB (Chair: Graca Rocha)

16h15 – Anomalies – Hiranya Peiris
16h50 – Transforming Data into Science: Planck data and the CMB non-Gaussianity – Anna Mangilli
17h10 – Applications of the Gaussian Kinematic Formula in Cosmology – Yabebal Fantaye
17h30 – Detectability of multi-connected topologies – Ophélia Fabre
17h50 – Cosmology with photometric quasars – Boris Leistedt
18h10 – Session ends
***18h10 to 18h40** – Meeting of the IAA Working Group on Cosmostatistics – Hosted by Rafael de Souza*

How does it work?

The COIN Residence Program



What we can NOT guarantee up front

Authorship

Talks

1 house

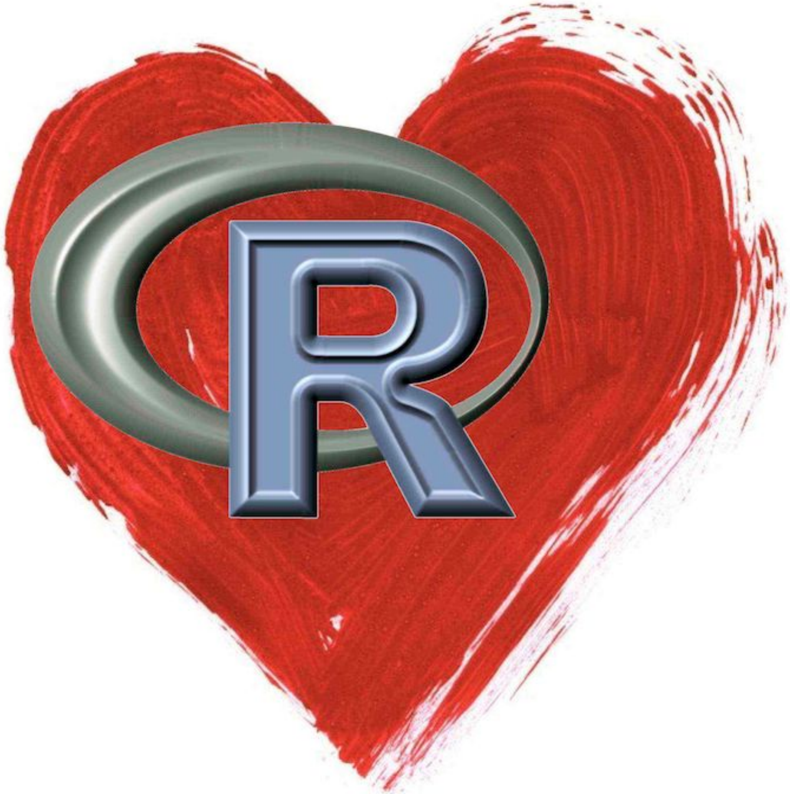
Lots of coffee

paper

What we can guarantee up front

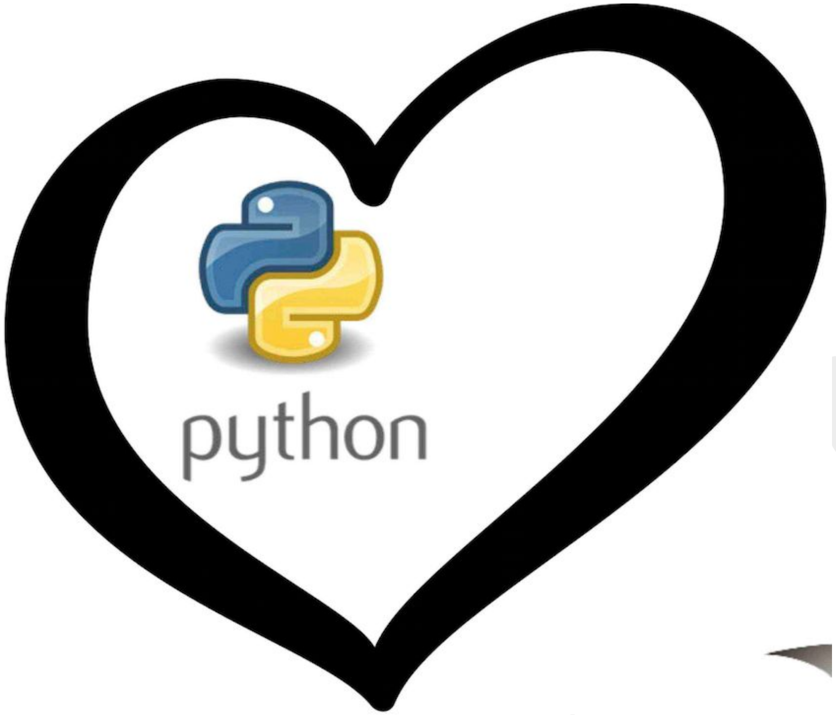
What we require from participants

Statisticians love R….



DiveRSITY

...as astronomers love Python



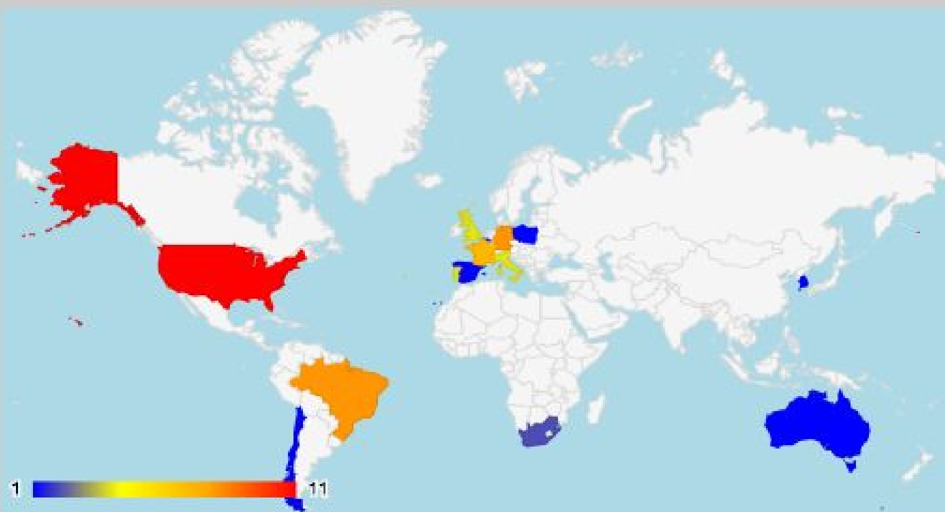
COIN tries to be bilingual as much as possible.

What it has achieved so far?

A Group of People

Lead by Rafael de Souza (ELTE, Hungary)
60 researchers from **15** countries

Nearly half actively involved in an ongoing project

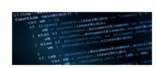


Country	Members
Australia	1
Belgium	1
Brazil	8
Chile	1
France	7
Germany	8
Italy	4
Netherland	1
Poland	1
Portugal	4
South Africa	2
South Korea	1
Spain	1
United Kingdom	4
USA	11

Scientific outcomes

In 3 years

1	GLM I	de Souza *et al.*, 2015
2	GLM II	Elliott *et al.*, 2015
3	GLM III	de Souza *et al.*, 2015
4	AMADA	de Souza & Ciardi, 2015
5	CosmoABC	Ishida *et al.*, 2015
6	DRACULA	Sasdelli *et al.*, 2016
7	AGNlogit	de Souza *et al.*, 2016
8	PhotoZ	Beck *et al.*, 2017



1	CosmoPhotoZ	de Souza *et al.*, 2014,
2	AMADA	de Souza & Ciardi, 2015
3	CosmoABC	Ishida *et al.*, 2015
4	DRACULA	Aguena *et al.*, 2015

+ 1 paper
+ 2 photoz catalogs

CRP #1 – Lisbon, Aug/2014

The Overlooked Potential of Generalized Linear Models in Astronomy - I: Binomial Regression and Numerical Simulations

R S. de Souza[a], E. Cameron[b], M. Killedar[c], J. Hilbe[d,e], R. Vilalta[f], U. Maio[g,h], V. Biffi[i], B. Ciardi[j], J. D. Riggs[k], for the COIN collaboration

Astronomy and Computing 12 (2015) 21–32

The Overlooked Potential of Generalized Linear Models in Astronomy-II: Gamma regression and photometric redshifts

J. Elliott[a], R. S. de Souza[b], A. Krone-Martins[c], E. Cameron[d], E. E. O. Ishida[e], J. Hilbe[f,g], for the COIN collaboration

Astronomy and Computing 10 (2015) 61–72

[ascl:1408.018]

cosmoabc: Likelihood-free inference via Population Monte Carlo Approximate Bayesian Computation

E. E. O. Ishida[1], S. D. P. Vitenti[2], M. Penna-Lima[3,4], J. Cisewski[5], R. S. de Souza[6], A. M. M. Trindade[7,8] E. Cameron[9] and V. C. Busti[10] for the COIN collaboration

Astronomy and Computing 13 (2015) 1–11 [ascl:1505.013]

CRP #2 – Isle of White, UK – Oct/2015

Exploring the spectroscopic diversity of type Ia supernovae with DRACULA: a machine learning approach

M. Sasdelli[1,2★], E. E. O. Ishida[2,3], R. Vilalta[4], M. Aguena[5], V. C. Busti[5], H. Camacho[5], A. M. M. Trindade[6,7], F. Gieseke[8], R. S. de Souza[9], Y. T. Fantaye[10], and P. A. Mazzali[1,2], for the COIN collaboration

[ascl:1512.009]

MNRAS (2016), 461 Issue 2, p.2044-2059

Is the cluster environment quenching the Seyfert activity in elliptical and spiral galaxies?

R. S. de Souza[1★], M. L. L. Dantas[2], A. Krone-Martins[3], E. Cameron[4], P. Coelho[2], M. W. Hattab[5], M. de Val-Borro[6], J. M. Hilbe[7], J. Elliott[8] and A. Hagen[9], for the COIN Collaboration

MNRAS (2016) 461, Issue 2, p.2115-2125

CRP #3 – Budapest – Aug/2016

Paper 1 :

On the realistic validation of photometric redshifts, or why Teddy will never be Happy

arXiv:1701.08748v2

R. Beck[1⋆], C.-A. Lin[2,3], E. E. O. Ishida[4], F. Gieseke[5], R. S. de Souza[6,7], M. V. Costa-Duarte[7,8], M. W. Hattab[9], A. Krone-Martins[10], for the COIN Collaboration

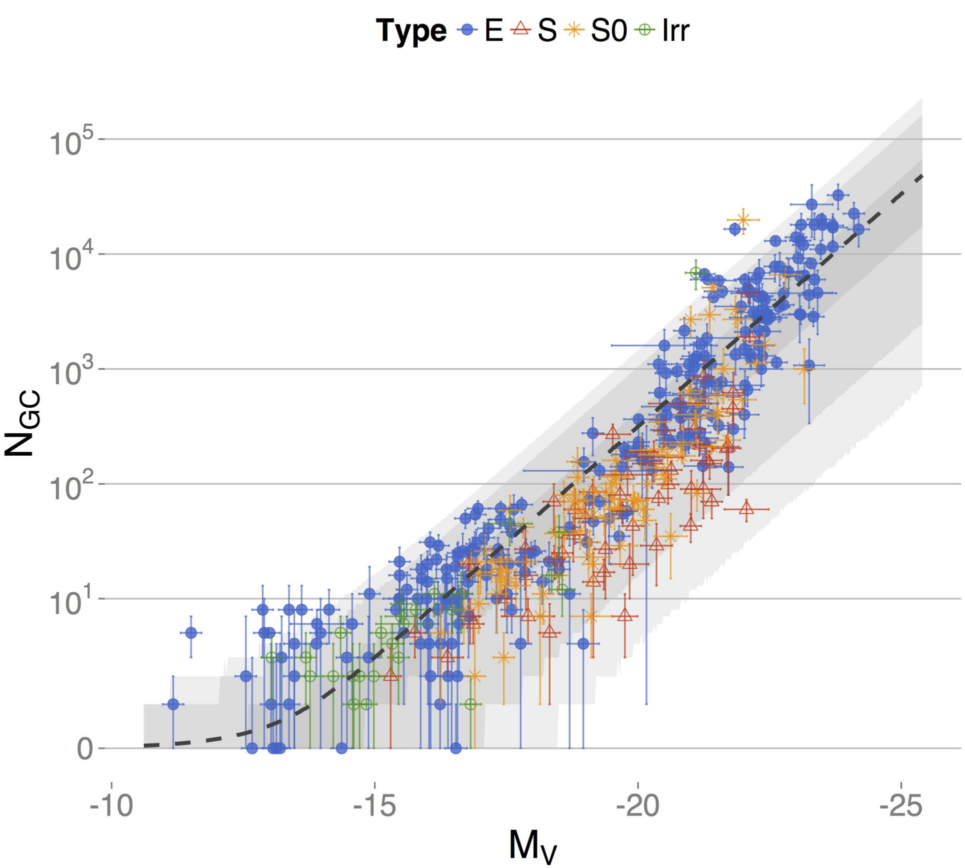
Paper 2:

A data-driven probabilistic approach for emission-line galaxy classification

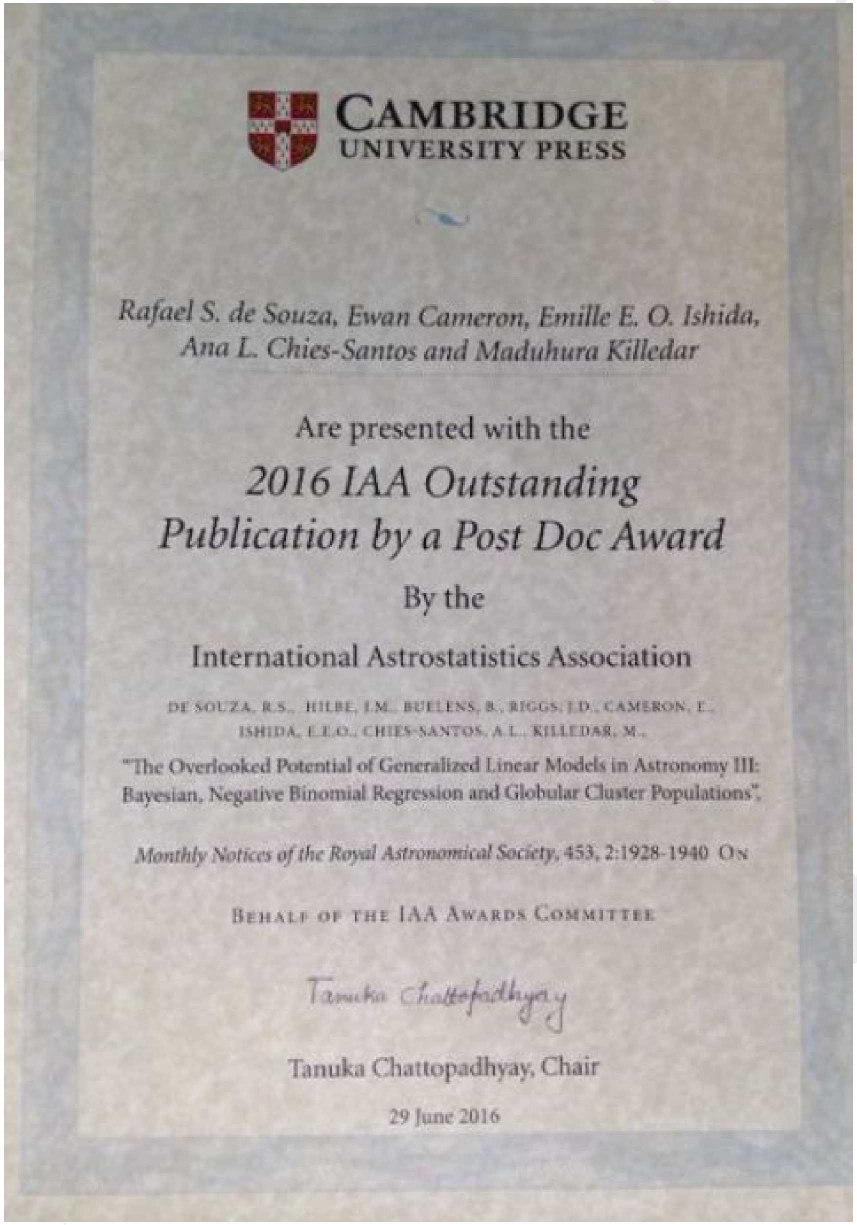
R. S. de Souza[1,2 ⋆], M. L. L. Dantas[1], M. V. Costa-Duarte[1,3], P.-Y. Lablanche[4,5], M. Killedar[6], E. Feigelson[7], R. Vilalta[8], A. Krone-Martins[9], R. Beck[10], F. Gieseke[11], for the COIN Collaboration

Beyond gaussian: Negative binomial regression

Bayesian negative binomial regression for discrete response variables.



Globular cluster population vs absolute visual magnitude (de Souza, et al. *in prep*)



CAMBRIDGE
UNIVERSITY PRESS

Rafael S. de Souza, Ewan Cameron, Emille E. O. Ishida, Ana L. Chies-Santos and Maduhura Killedar

Are presented with the

2016 IAA Outstanding Publication by a Post Doc Award

By the

International Astrostatistics Association

DE SOUZA, R.S., HILBE, J.M., BUELENS, B., RIGGS, J.D., CAMERON, E., ISHIDA, E.E.O., CHIES-SANTOS, A.L., KILLEDAR, M.,

"The Overlooked Potential of Generalized Linear Models in Astronomy III: Bayesian, Negative Binomial Regression and Globular Cluster Populations",

Monthly Notices of the Royal Astronomical Society, 453, 2:1928-1940 On

Behalf of the IAA Awards Committee

Tanuka Chattopadhyay

Tanuka Chattopadhyay, Chair

29 June 2016

Fork me on GitHub

COINtoolbox

DOI 10.5281/zenodo.16376

Cosmostatistics Initiative COIN

Methodology and software for cosmology

The COsmostatistics INitiative (COIN), a working group built within the International Astrostatistics Association (IAA), aims to create a friendly environment where hands-on collaboration between astronomers, cosmologists, statisticians and machine learning experts can flourish. COIN is designed to promote the development of a new family of tools for data exploration in cosmology.

Generalized Linear Models in Astronomy

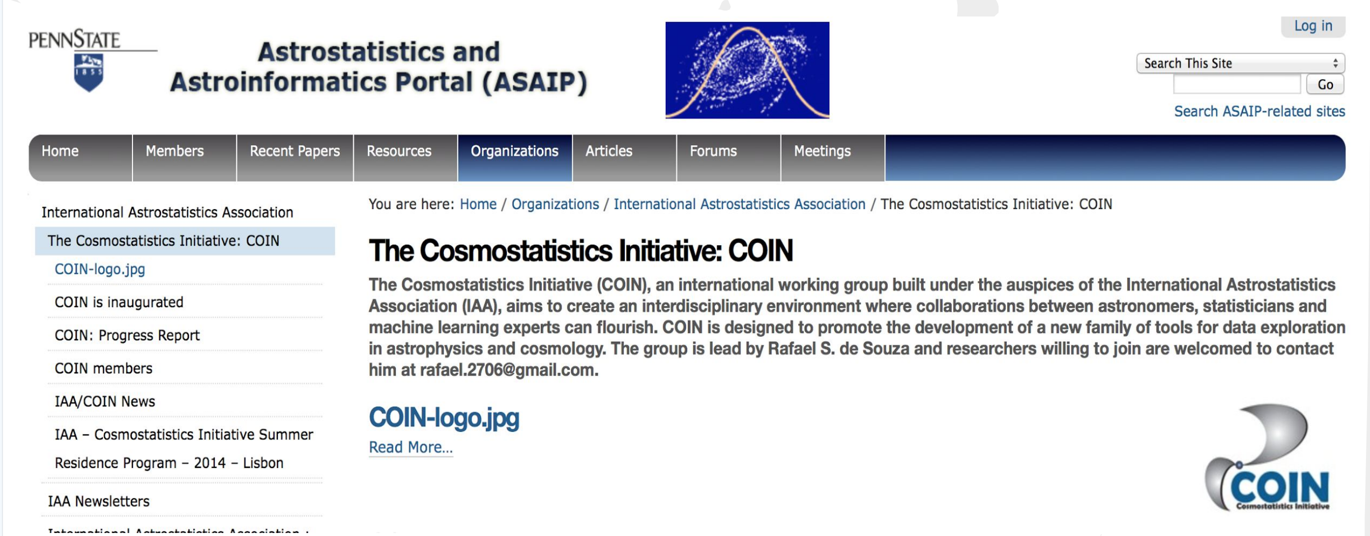
Statistical methods play a central role to fully exploit astronomical catalogues and an efficient data analysis requires astronomers to go beyond the traditional Gaussian-based models. This projects illustrates the power of generalized linear models (GLMs) for astronomical community, from a Bayesian perspective. Applications range from modelling star formation activity (logistic regression), globular cluster population (negative binomial regression), photometric redshifts (gamma regression), exoplanets multiplicity (Poisson regression), and so forth.

Binomial Regression

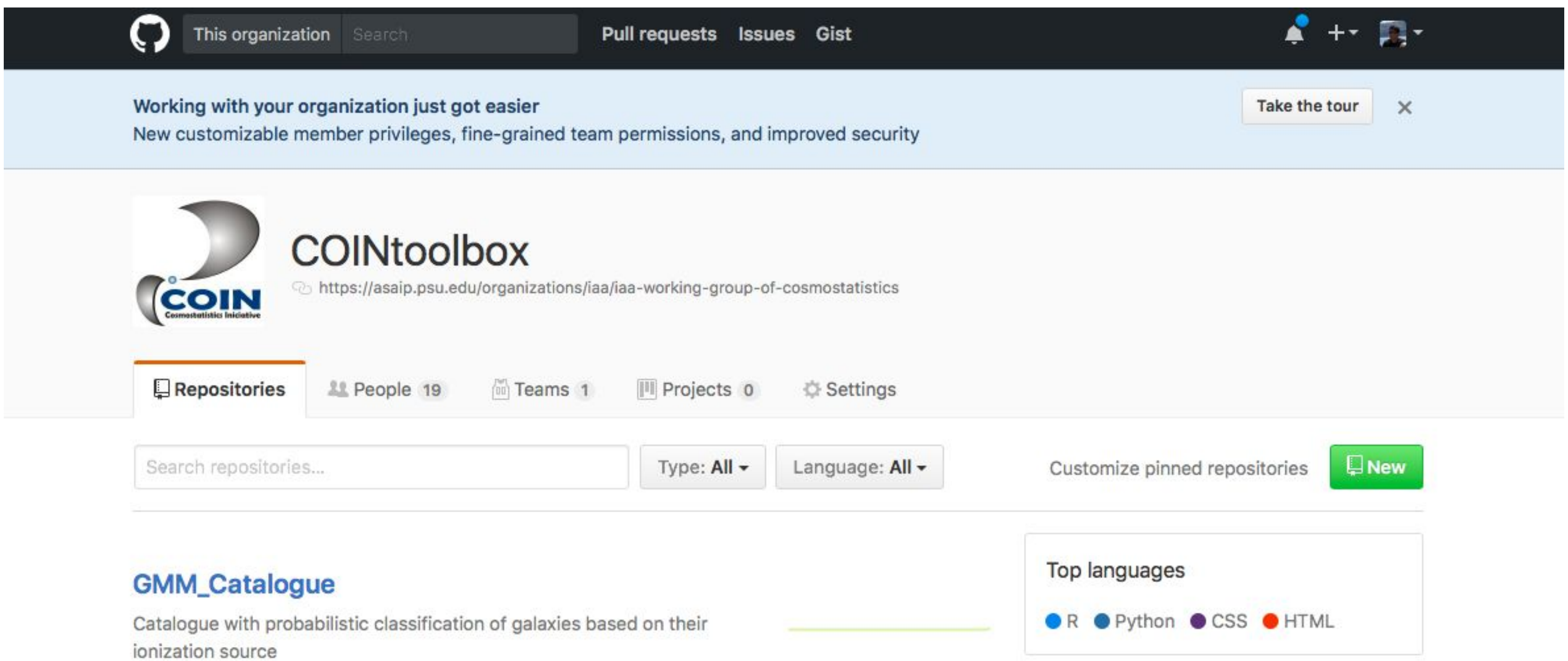
Suited to handle binary or proportional data, also called absence and presence data. For example AGN activity, star-galaxy separation, fraction of bars in a galaxy, scape fraction, etc.

Link to ADS | Tutorial

COIN in the media



Python products by COIN



Example of a Machine Learning Application

DRACULA - Dimensionality Reduction And Clustering for Unsupervised Learning in Astronomy

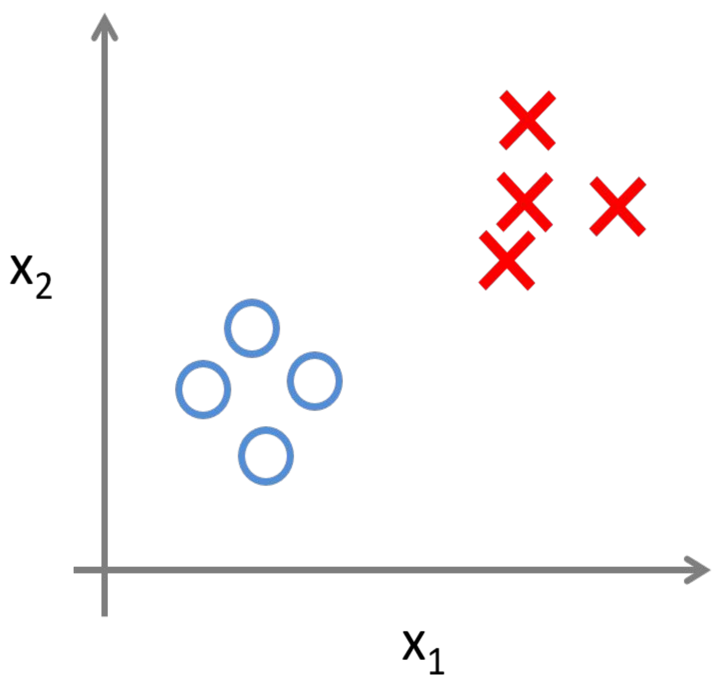
DRACULA is distributed under GPL3 or latter. It is one of the products of the second edition of the COIN Residence Program and is maintained by Michel Aguena (University of Sao Paulo).

If you have any questions, suggestions or just want to be updated about the development of the code, please send an email to coin_dracula+subscribe@googlegroups.com .

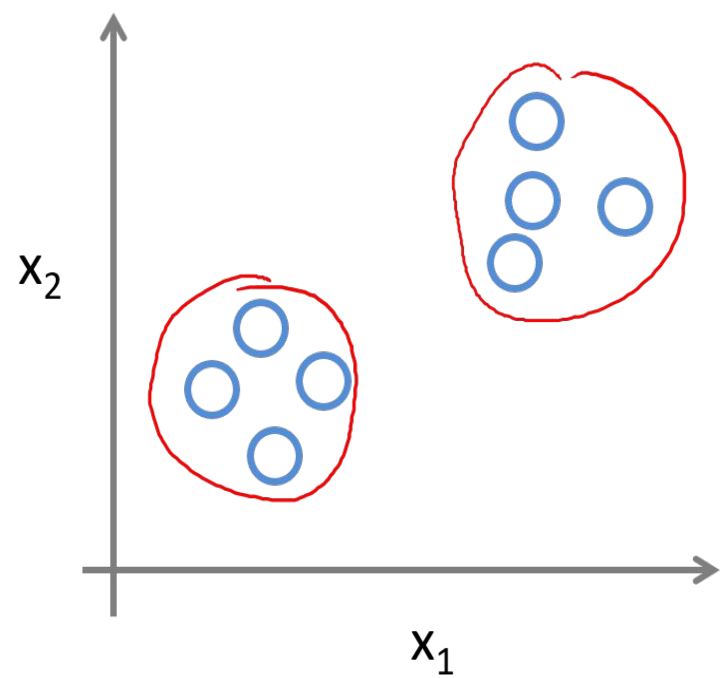
If you use DRACULA in your research, please cite Sasdelli et al, 2015 and Aguena et al., 2015.

Supervised versus Unsupervised

Supervised Learning



Unsupervised Learning

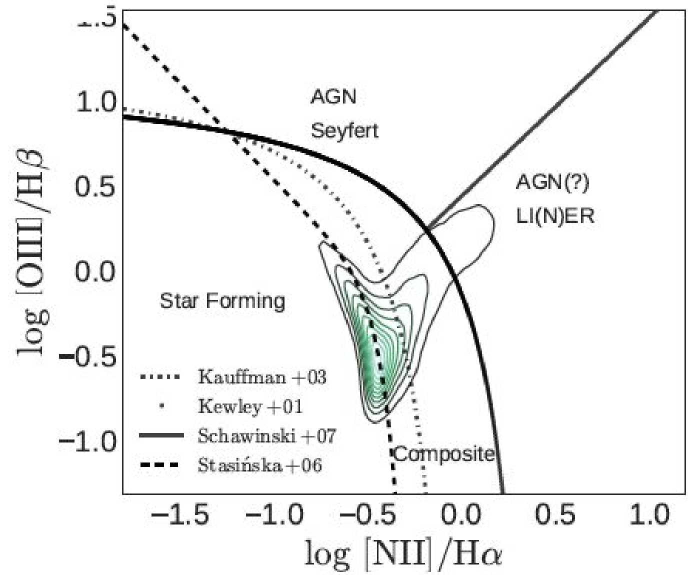


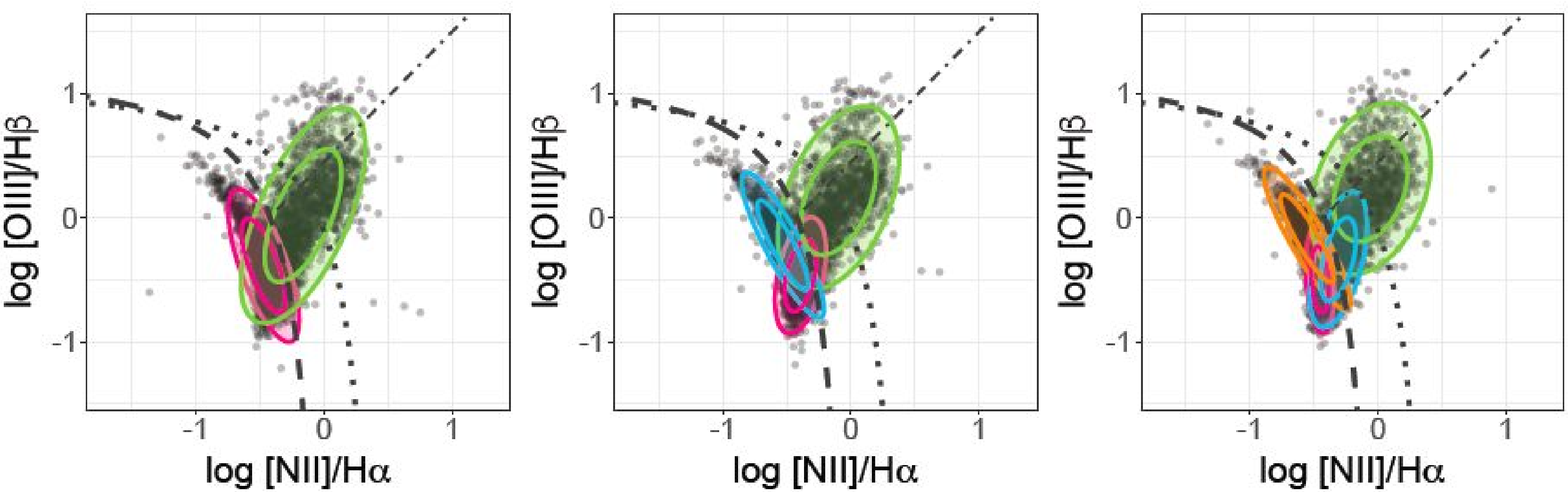
A data-driven probabilistic approach for emission-line galaxy classification

R. S. de Souza[1,2] ⋆, M. L. L. Dantas[1], M. V. Costa-Duarte[1,3], P.-Y. Lablanche[4,5], M. Killedar[6], E. Feigelson[7], R. Vilalta[8], A. Krone-Martins[9], R. Beck[10], F. Gieseke[11],
for the COIN Collaboration

The first step in wisdom is to know the things themselves. This notion consists in having a true idea of the objects; objects are distinguished and known by classifying them methodologically and giving them appropriate names. Therefore, classification and name-giving will be the foundation of our science.

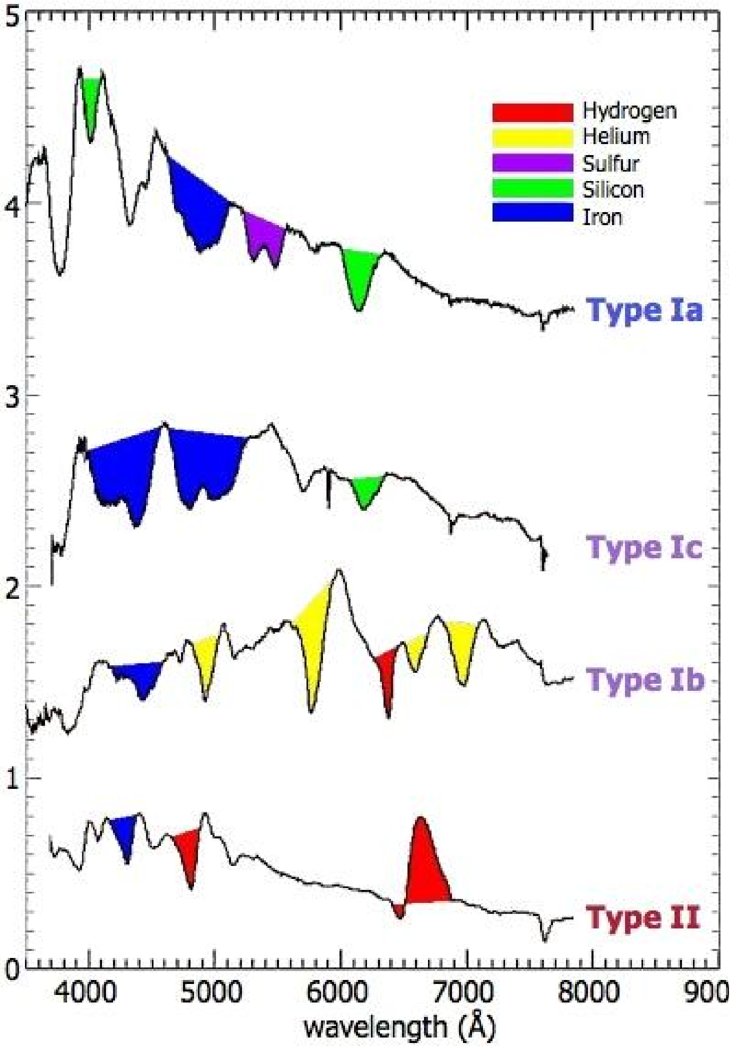
Carl Linnaeus, 1735





Can we use machine learning to distinguish multiple sub-classes of SN Ia ?

If so, are the resulting classes physically meaningful?



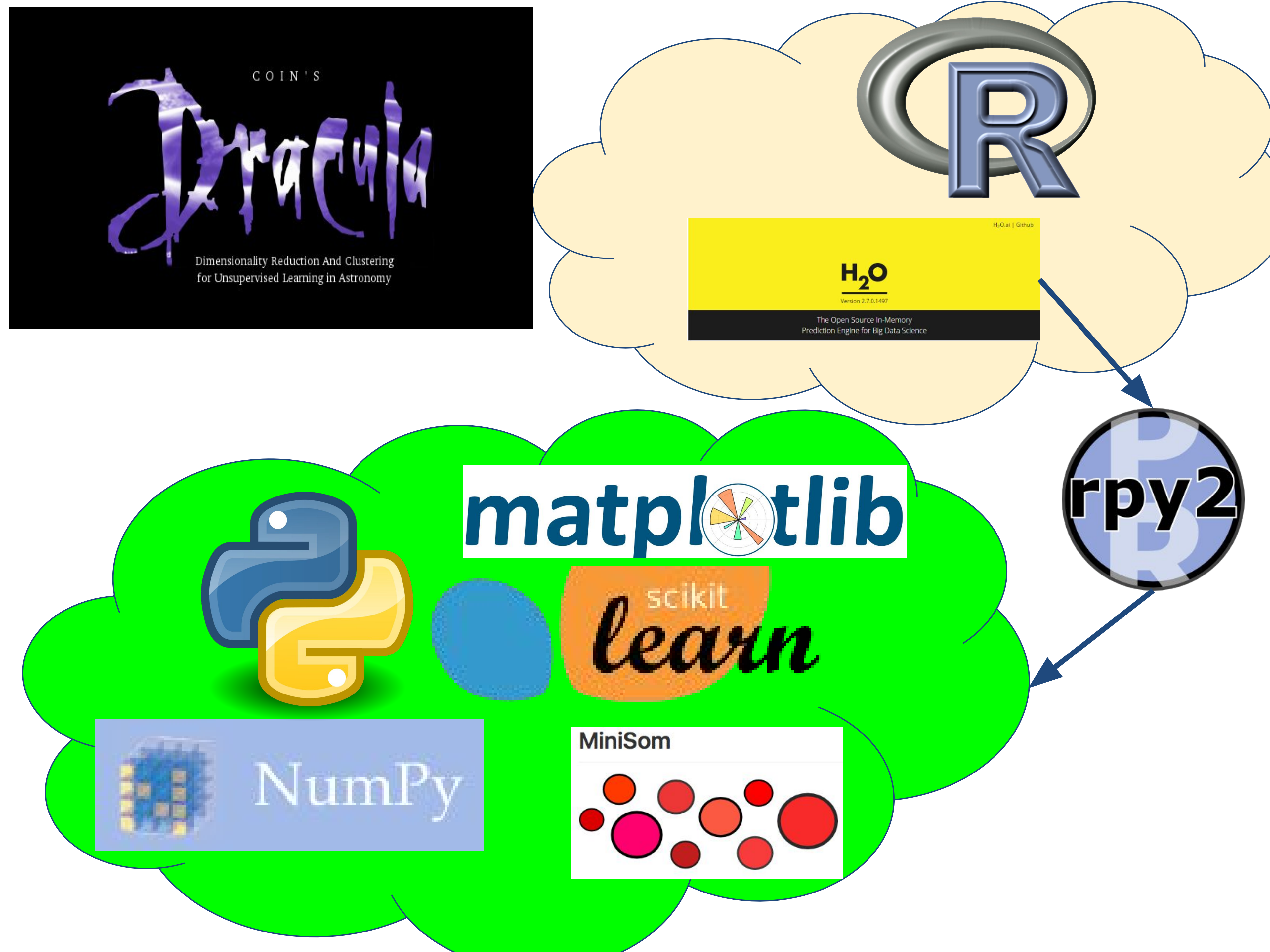
Initial strategy:

1. Apply a machine learning algorithm

2. Compare with subgroups found by humans

3. Move beyond existing sub-groups

4. Guide current classification system and simulation studies


COIN'S
Dracula
Dimensionality Reduction And Clustering
for Unsupervised Learning in Astronomy
R
H_2O.ai | Github
H_2O
Version 2.7.0.1497
The Open Source In-Memory
Prediction Engine for Big Data Science
rpy2
matplotlib
scikit
learn
NumPy
MiniSom

ASCL.net
Astrophysics Source Code Library

Home About Resources Browse Submissions News Forum Dashboard

ASCL Code Record

[ascl:1512.009] DRACULA: Dimensionality Reduction And Clustering for Unsupervised Learning in Astronomy

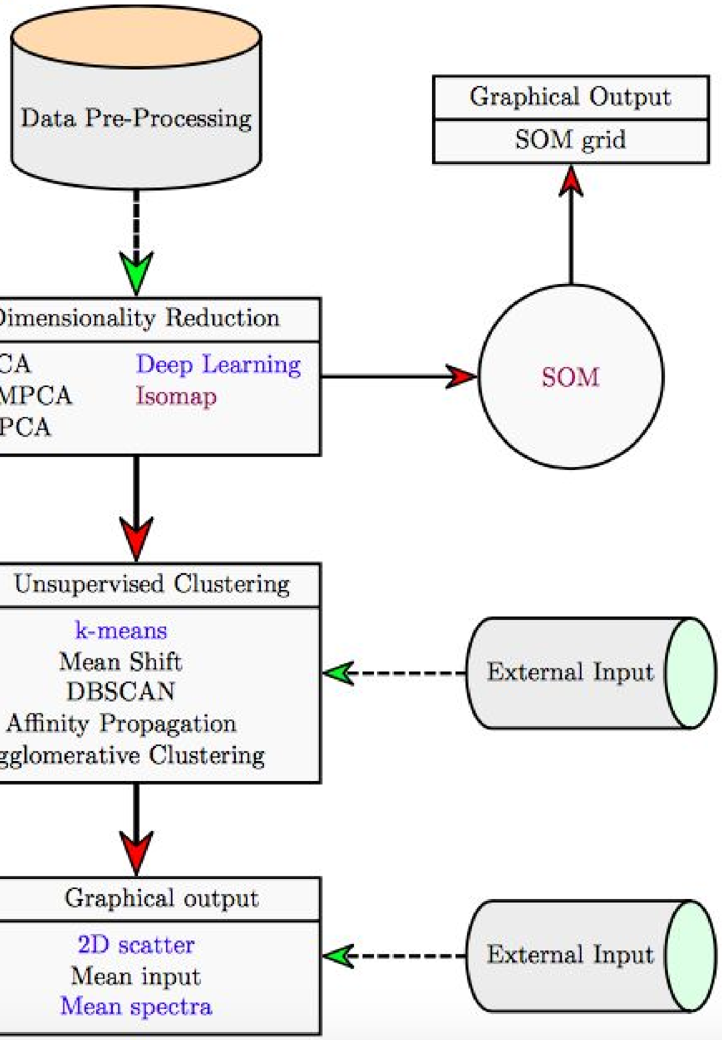
Aguena, Michel; Busti, Vinicius C.; Camacho, Hugo; Sasdelli, Michele; Ishida, Emille E. O.; Vilalta, Ricardo; Trindade, Arlindo M. M.; Gieseke, Fabien; de Souza, Rafael S.; Fantaye, Yabebal T.; Mazzali, Paolo A.

DRACULA classifies objects using dimensionality reduction and clustering. The code has an easy interface and can be applied to separate several types of objects. It is based on tools developed in scikit-learn, with some usage requiring also the H2O package.

Code site: https://github.com/COINtoolbox/DRACULA
Appears in: http://arxiv.org/abs/1512.06810
Bibcode: 2015ascl.soft12009A

Explain these fields?

ascl 1512.009



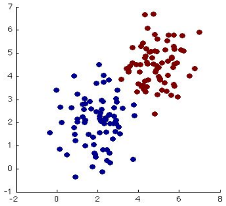
Exploring the spectroscopic diversity of Type Ia supernovae with dracula: a machine learning approach

M. Sasdelli; E. E. O. Ishida; R. Vilalta; M. Aguena; V. C. Busti; H. Camacho; A. M. M. Trindade; F. Gieseke; R. S. de Souza; Y. T. Fantaye; ... Show more

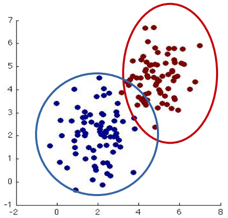
Mon Not R Astron Soc (2016) 461 (2): 2044-2059. DOI: https://doi.org/10.1093/mnras/stw1228
Published: 24 May 2016 Article history ▾

IDEALLY

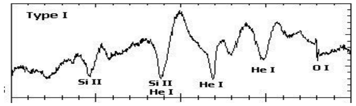
Build data matrix

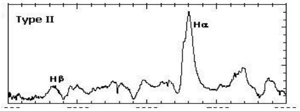


Dimensionality Reduction



Unsupervised clustering



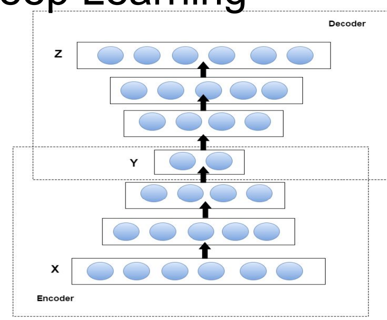


subclasses

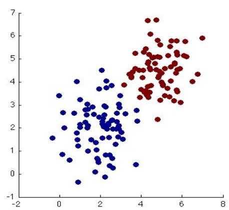
-REALISTICALLY-

Build data matrix ← Transfer learning

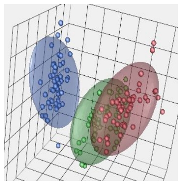
Dimensionality Reduction ← Deep Learning



Isomaps



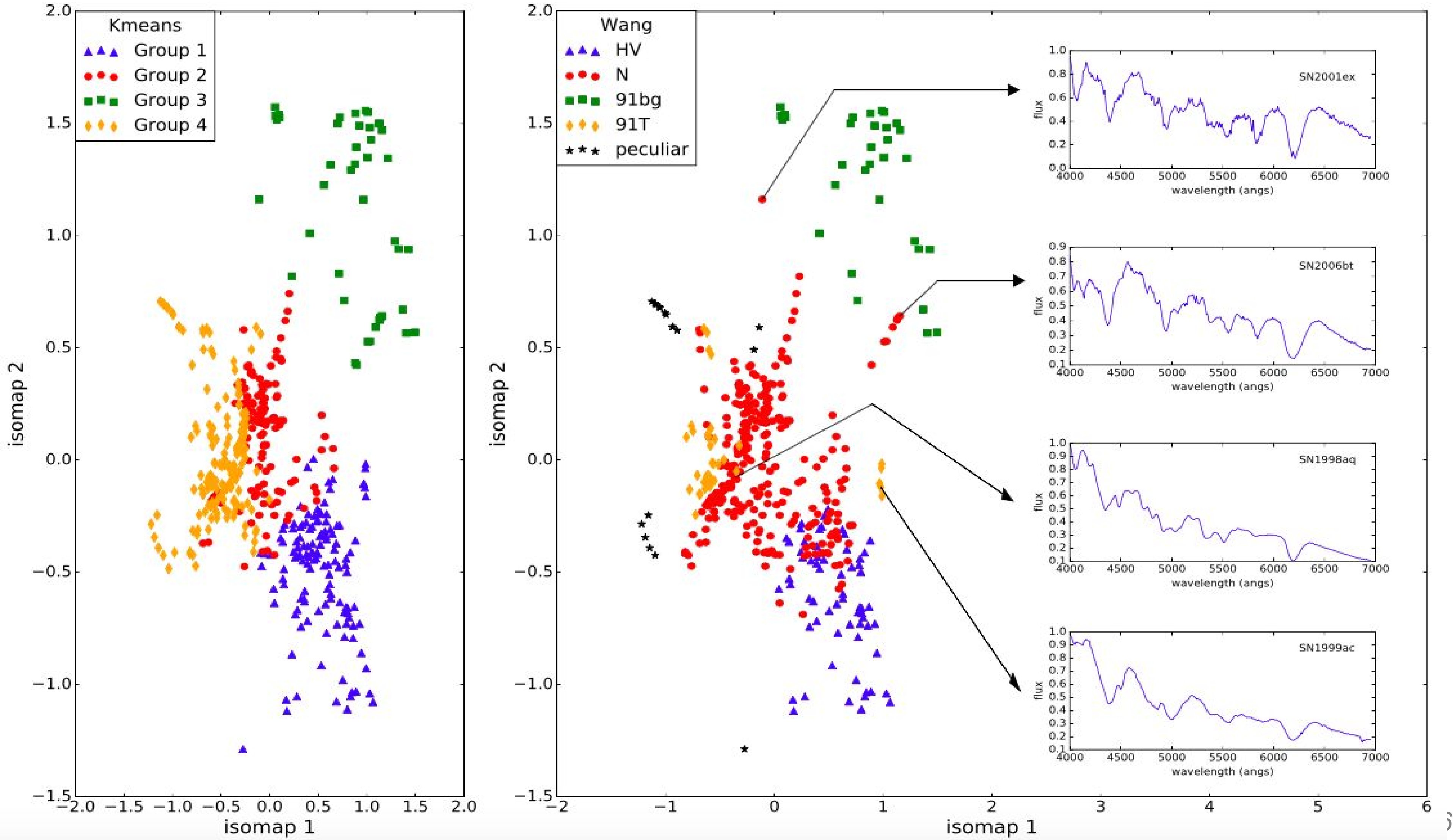
Dimensionality Reduction ← Unsupervised clustering ← K-means



subclasses

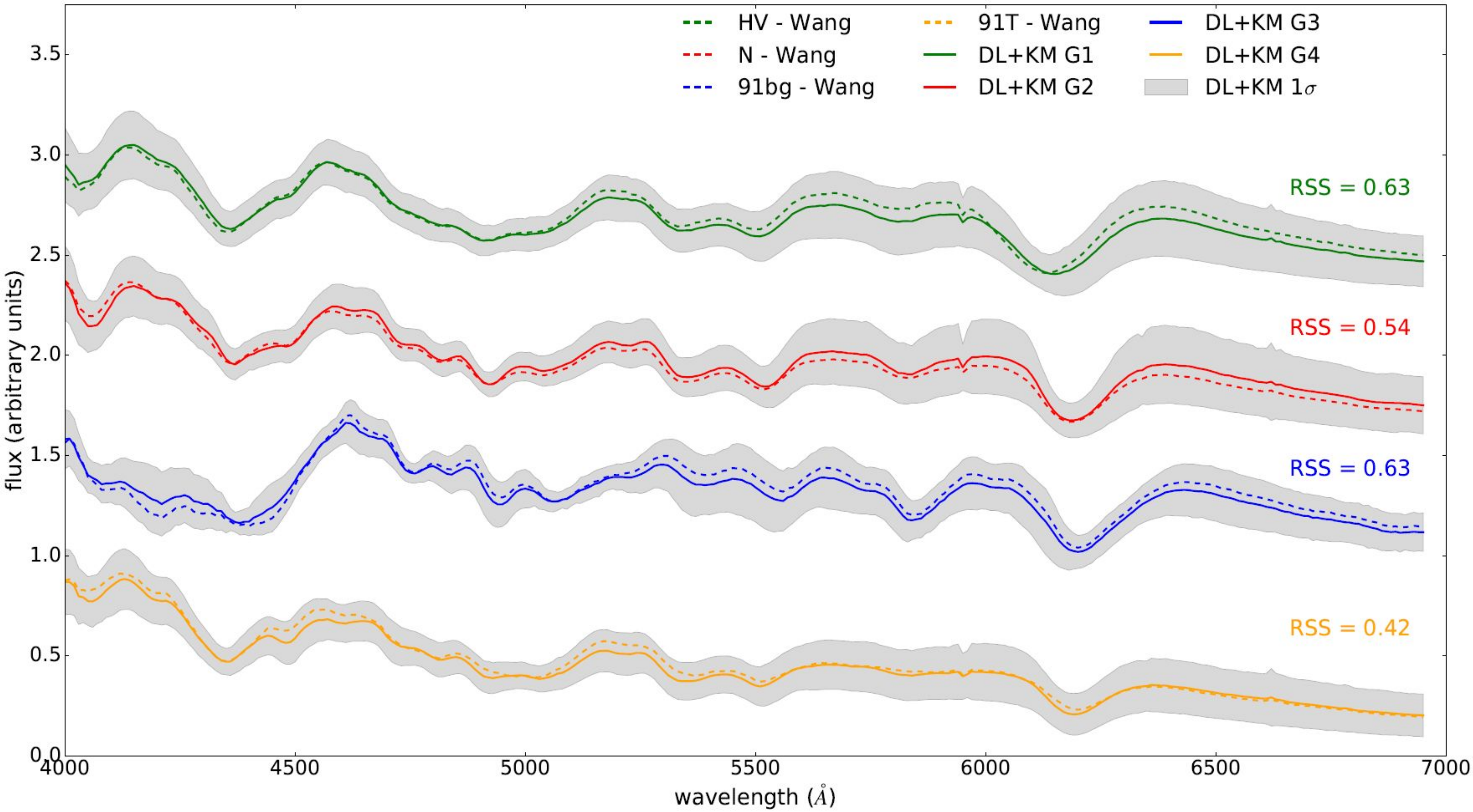
2D visualization of 4D Deep Learning parameter space

Results from K-means



Mean spectra by ML x human

K-means with 4 groups

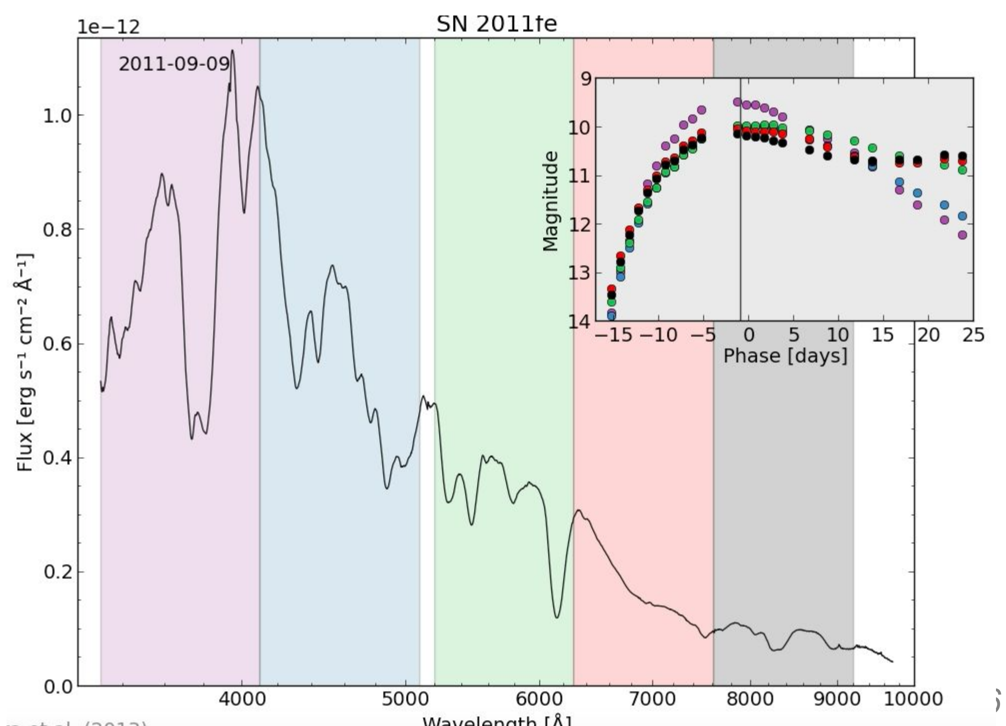


Sasdelli *et al.*, 2016, MNRAS, *461, Issue 2,* 2044-2059

The Nearby Supernova Factory

The **Nearby Supernova Factory** (*SNfactory*) is an experiment to develop Type Ia supernovae as tools to measure the expansion history of the Universe and explore the nature of Dark Energy. It is the largest data volume supernova search currently in operation.

Invited to be applied to SNfactory by the Berkeley group



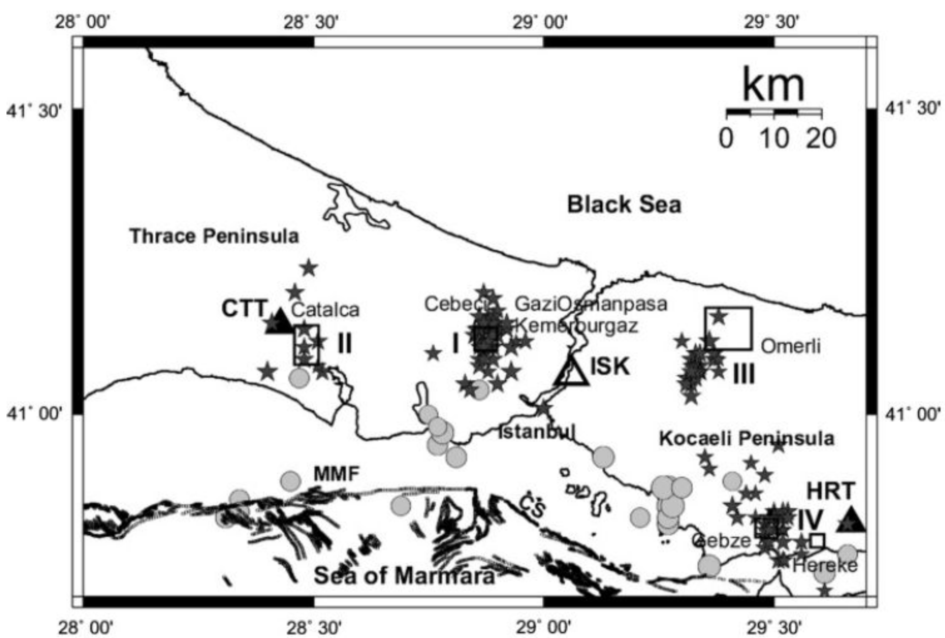
a et al. (2013).

Examples in Geophysics

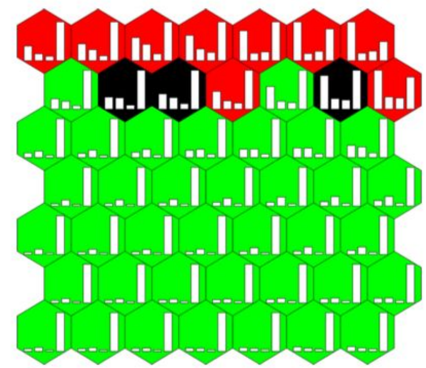
An unsupervised learning algorithm: application to the discrimination of seismic events and quarry blasts in the vicinity of Istanbul

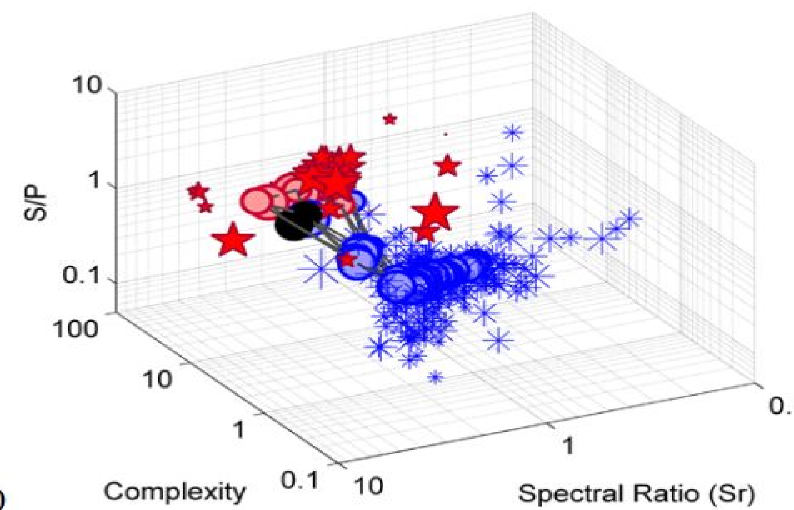
H. S. Kuyuk[1], E. Yildirim[2], E. Dogan[1], and G. Horasan[2]

[1]Department of Civil Engineering, Sakarya University, Sakarya, Turkey
[2]Department of Geophysical Engineering, Sakarya University, Sakarya, Turkey



Self Organizing Map





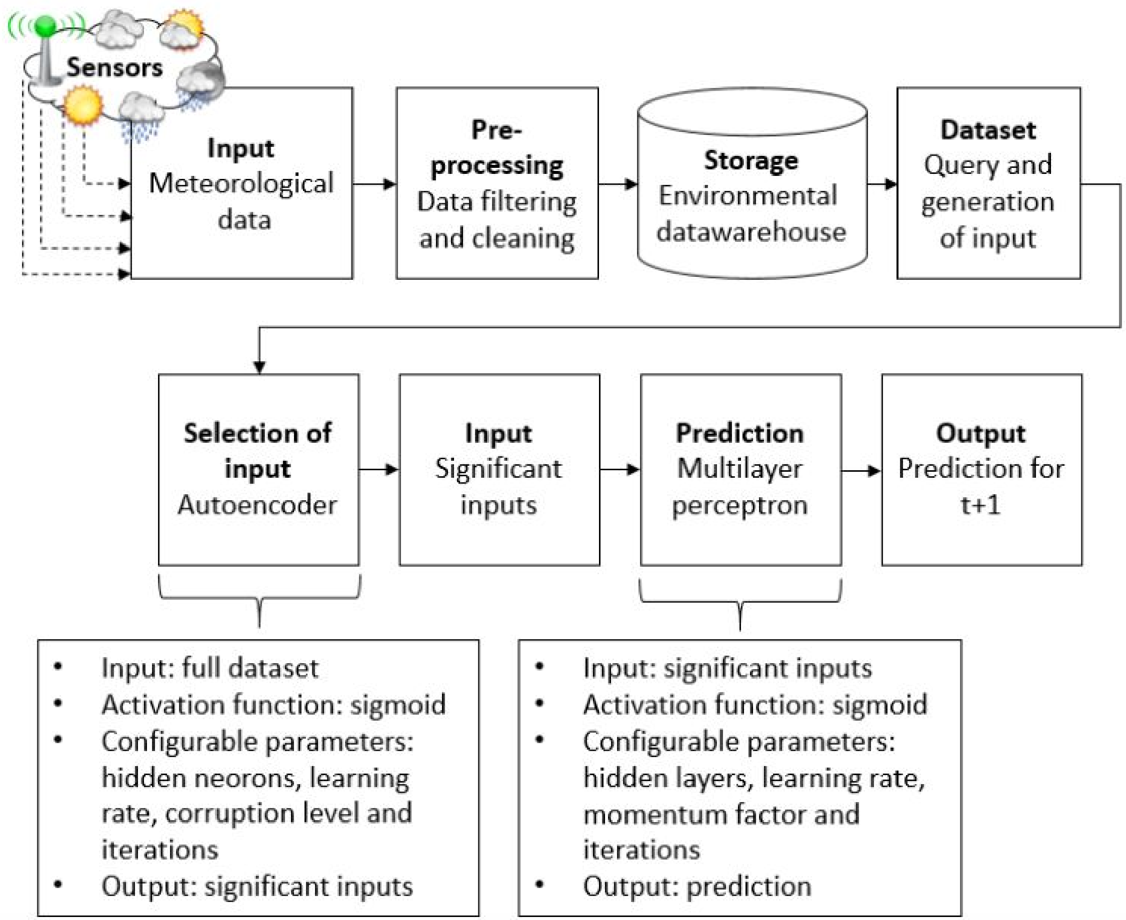
(a)

Fig. 5. Red stars indicate earthquakes and blue asterisks represent the quarry blasts; gray lines show the connections between neighboring map units. Red, blue and black circles located at the nodes indicate earthquakes, quarry blast and unlabeled events, respectively, and are classification results from the SOM. **(a)**

Examples in Meteorology

Rainfall prediction: A Deep Learning approach

Emilcy Hernández[1], Victor Sanchez-Anguix[2], Vicente Julian[3], Javier Palanca[3], and Néstor Duque[4]



Now a *root* Statistical Application

Approximate Bayesian Computation

CosmoABC

Search docs

CosmoABC - Likelihood free parameter estimation for cosmology

Python *code:* Ishida, et al. 2015

`CosmoABC` is a package which enables parameter inference using an Approximate Bayesian Computation (ABC) algorithm, as described in Ishida et al., 2015 [LINK].

The code was originally designed for cosmological parameter inference from galaxy clusters number counts based on Sunyaev-Zel'dovich measurements. In this context, the cosmological simulations were performed using the NumCosmo library .

Nevertheless, the user can easily take advantage of the ABC sampler along with his/her own simulator, as well as test personalized prior distributions, summary statistics and distance functions.

Get it now!

The package can be installed using the PyPI and pip:

```
$ pip install CosmoABC
```

Or if the tarball or repository is downloaded, in the CosmoABC directory you can install and test it:

```
$ python setup.py install
```

You can run a few tests with:

```
$ test_ABC_algorithm.py
```

Summary

1 - Parameter inference

2 – Approximate Bayesian Computation (ABC)

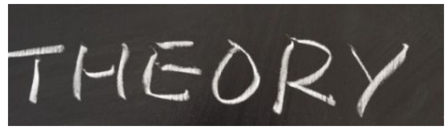
3 – ABC in astronomy

4 - Conclusions

Parameter inference

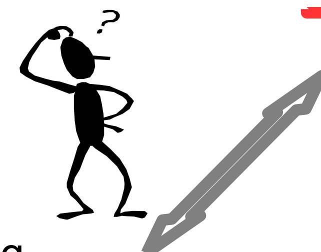
1. Physical phenomenon

2. Physical reasoning



3. Model

$\{z_i, x_i\}$ for $i=1,2\ldots,N \leftarrow M(\Omega_m, w, \ldots)$



4. Collect

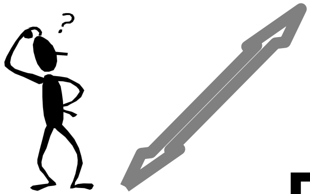
5. Catalog

z	ξ
2.5	0.9
5.6	2.1
7.4	2.9
10.5	4.2
..	...

6. Find parameters

$\Omega_m = \ldots$

$w = \ldots$



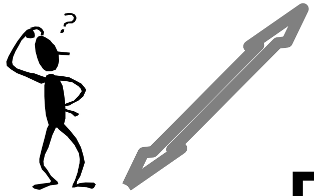
Bayes theorem

$$p(\boldsymbol{\theta}|\mathcal{D}) = \frac{p(\mathcal{D}|\boldsymbol{\theta})\,p(\boldsymbol{\theta})}{p(\mathcal{D})}$$

Bayes theorem

Can be too complicated!

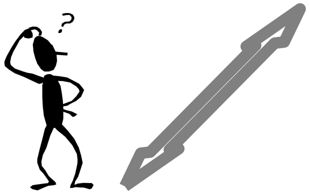
$$p(\boldsymbol{\theta}|\mathcal{D}) = \frac{p(\mathcal{D}|\boldsymbol{\theta})\,p(\boldsymbol{\theta})}{p(\mathcal{D})}$$



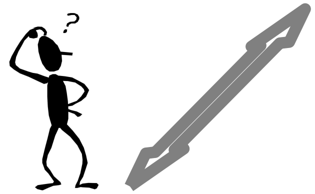
Bayes theorem

$$p(\boldsymbol{\theta}|\mathcal{D}) = \frac{p(\mathcal{D}|\boldsymbol{\theta})\,p(\boldsymbol{\theta})}{p(\mathcal{D})}$$

Can be problematic!



Can simulations help?



Comparing catalogs (or basic ABC)

Model + $\{\Omega_{m1}, w_1\}$

z	ξ
2.0	1.8
11.5	10.5
13.1	12.9
25.4	26.7
..	...

Observed

z	ξ
2.5	0.9
5.6	2.1
7.4	2.9
10.5	4.2
..	...

Simulate

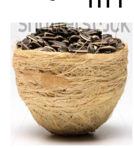
Draw $\{\Omega_{m1}, w_1\}$ from priors

Compare

If close enough

If too different

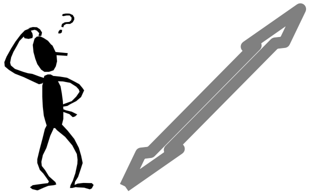
Store $\{\Omega_{m1}, w_1\}$

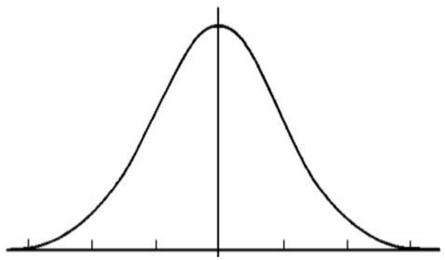


Throw away

Start again

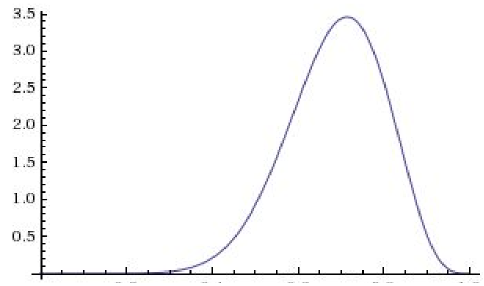
Cosmostatistics Initiative COIN





0.27

Ω_m



-1.0

w

Posterior

Approximate Bayesian Computation (ABC)

Population Monte Carlo ABC

Importance sampling: guiding draws

Draw $\{a_1, b_1\}$ from previous weighted particle system

Repeat until there are N particles stored

if $d < \varepsilon$, store $\{\Omega_{m1}, w_1\}$

Associate a weight with each particle

$$W_t^j = \frac{p(\theta_t^j)}{\sum_{i=1}^{N} W_{t-1}^i \mathcal{N}(\theta_t^j; \theta_{t-1}^i, C_{t-1})},$$

Determine ε

Approximate Bayesian Computation (ABC)

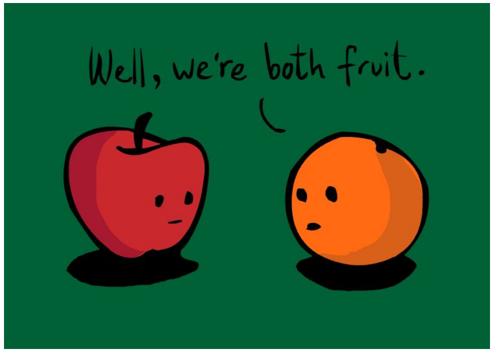
Ingredients

1. Priors
2. Simulator (cheap)
3. Distance function
   (and summary statistics)

Approximate Bayesian Computation (ABC)

Ingredients

1. Priors
2. Simulator (cheap)
3. Distance function
   (and summary statistics)



ABC in Astronomy

2010

cosmopmc

Mon. Not. R. Astron. Soc. **405,** 2381–2390 (2010)

Bayesian model comparison in cosmology with Population Monte Carlo

Martin Kilbinger,[1,2★] Darren Wraith,[1,3] Christian P. Robert,[3] Karim Benabed,[1] Olivier Cappé,[4] Jean-François Cardoso,[1,4] Gersende Fort,[4] Simon Prunet[1] and François R. Bouchet[1]

[1]*Institut d'Astrophysique de Paris, UMR 7095 CNRS & Université Pierre et Marie Curie, 98 bis boulevard Arago, 75014 Paris, France*
[2]*Shanghai Key Lab for Astrophysics, Shanghai Normal University, Shanghai 200234, China*
[3]*CEREMADE, Université Paris Dauphine, 75016 Paris, France*
[4]*LTCI, Telecom ParisTech and CNRS, 46 rue Barrault, 75013 Paris, France*

2010 2012

cosmopmc

Cameron & Petit

Mon. Not. R. Astron. Soc. **425,** 44–65 (2012)

Approximate Bayesian Computation for astronomical model analysis: a case study in galaxy demographics and morphological transformation at high redshift

E. Cameron★ and A. N. Pettitt

School of Mathematical Sciences (Statistical Science), Queensland University of Technology (QUT), GPO Box 2434, Brisbane 4001, QLD, Australia

2010 2012 2013

cosmopmc

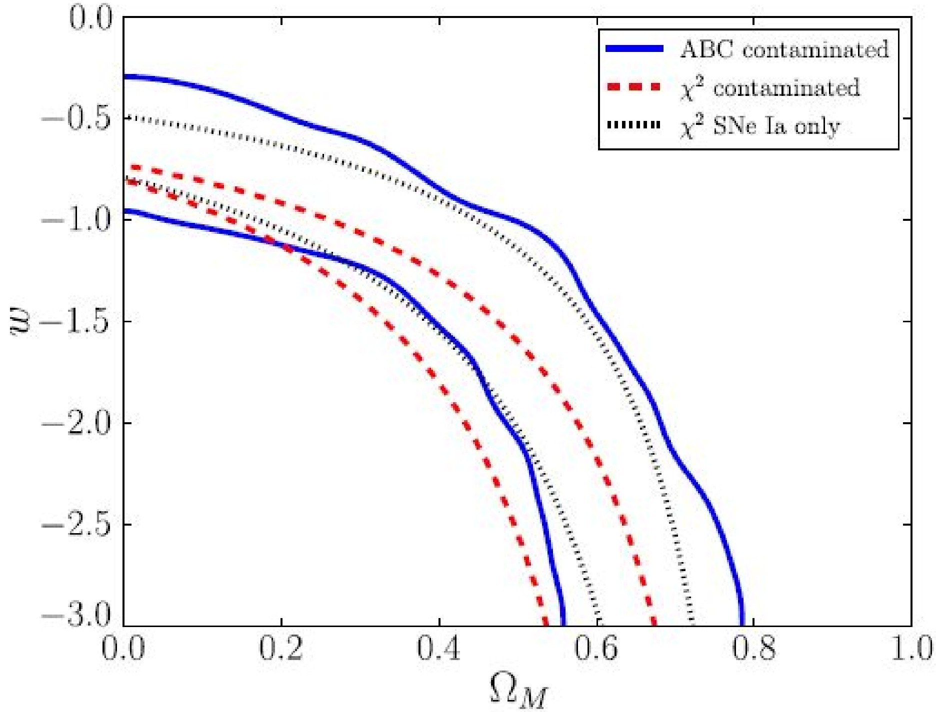
Cameron & Pettitt

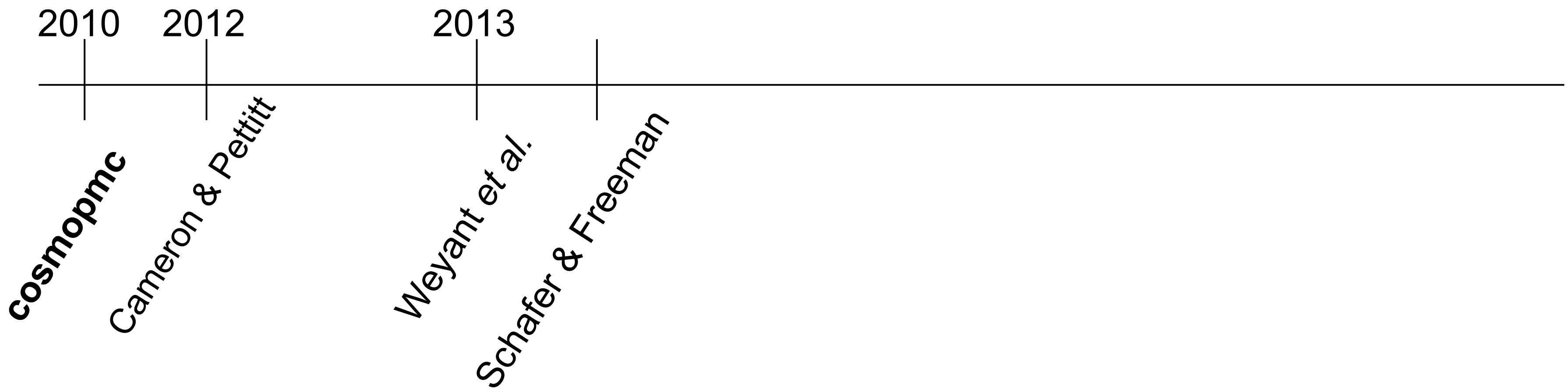
Weyant et al.

THE ASTROPHYSICAL JOURNAL, 764:116 (15pp), 2013 February 20

LIKELIHOOD-FREE COSMOLOGICAL INFERENCE WITH TYPE Ia SUPERNOVAE: APPROXIMATE BAYESIAN COMPUTATION FOR A COMPLETE TREATMENT OF UNCERTAINTY

ANJA WEYANT[1], CHAD SCHAFER[2], AND W. MICHAEL WOOD-VASEY[1]

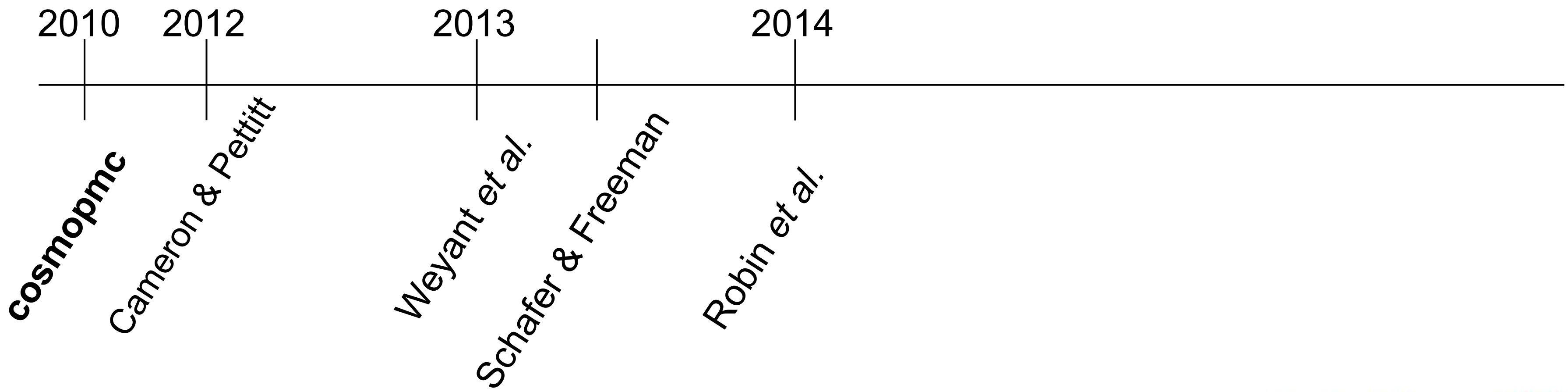




E.D. Feigelson and G.J. Babu (eds.), *Statistical Challenges in Modern Astronomy V*,
Lecture Notes in Statistics 209, DOI 10.1007/978-1-4614-3520-4_1,

Chapter 1
Likelihood-Free Inference in Cosmology: Potential for the Estimation of Luminosity Functions

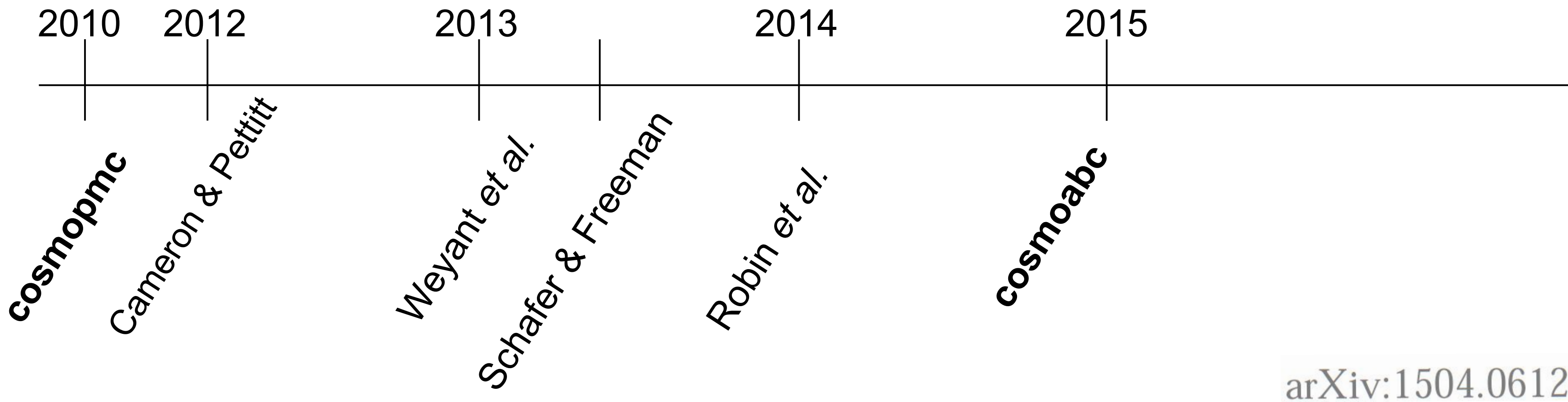
Chad M. Schafer and Peter E. Freeman



A&A 569, A13 (2014)

Constraining the thick disc formation scenario of the Milky Way★

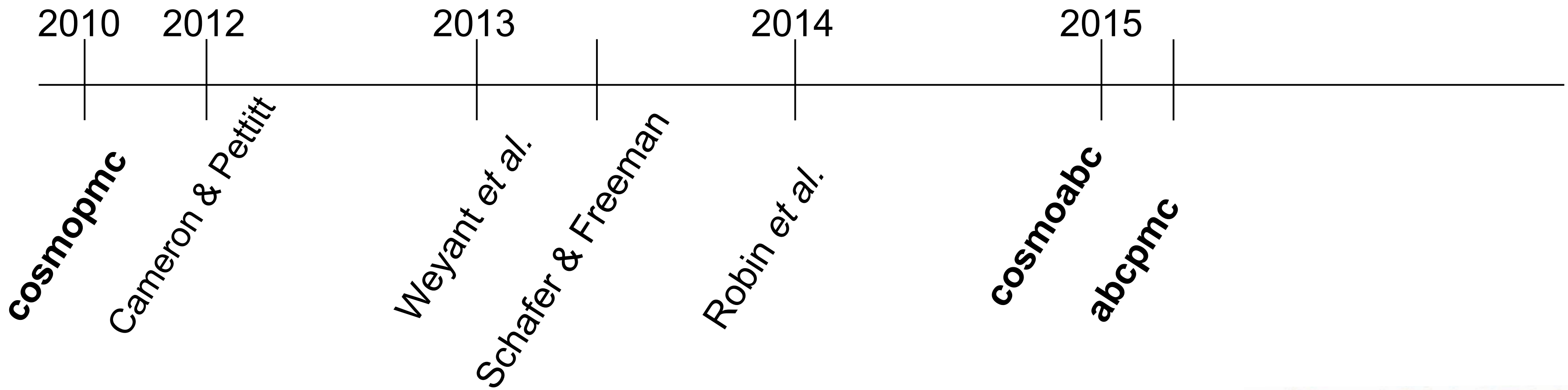
A. C. Robin[1], C. Reylé[1], J. Fliri[2,3], M. Czekaj[4], C. P. Robert[5], and A. M. M. Martins[1]



arXiv:1504.06129v2

cosmoabc: Likelihood-free inference via Population Monte Carlo Approximate Bayesian Computation

E. E. O. Ishida[1], S. D. P. Vitenti[2], M. Penna-Lima[3,4], J. Cisewski[5], R. S. de Souza[6], A. M. M. Trindade[7,8]
E. Cameron[9] and V. C. Busti[10], for the COIN collaboration

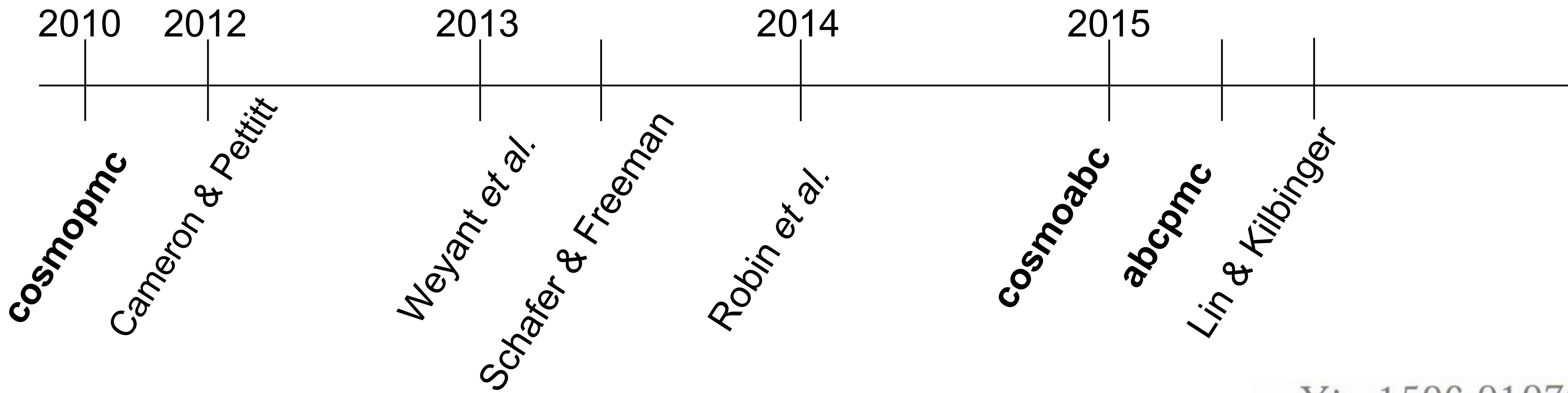


arXiv:1504.07245v2

Approximate Bayesian Computation for Forward Modeling in Cosmology

Joël Akeret[a] Alexandre Refregier[a] Adam Amara[a] Sebastian Seehars[a] Caspar Hasner [a]

[a]ETH Zurich, Institute for Astronomy, Department of Physics, Wolfgang Pauli Strasse 27, 8093 Zurich, Switzerland



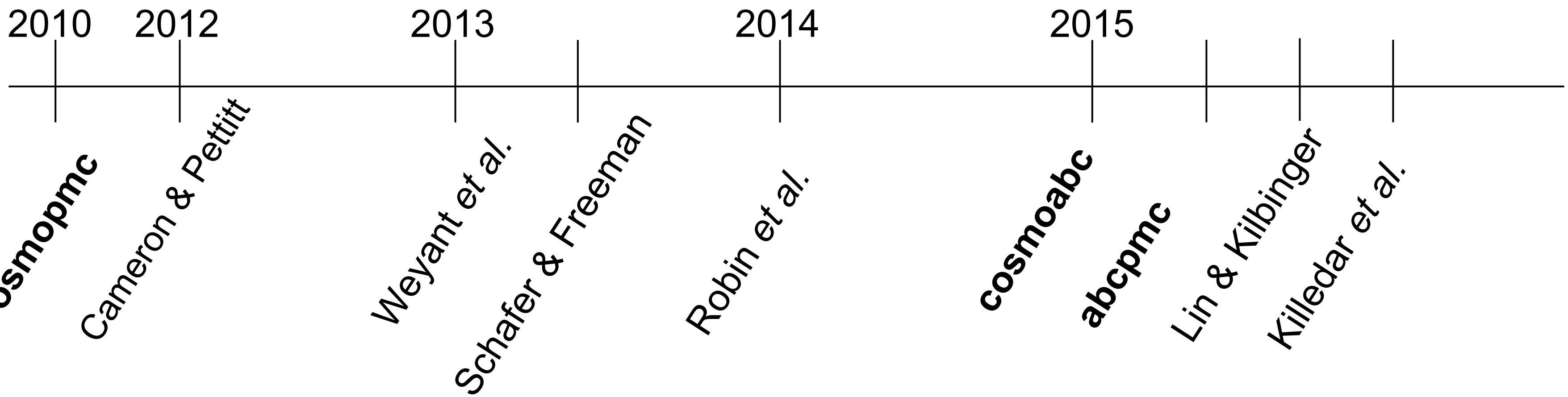
arXiv:1506.01076v1

A new model to predict weak-lensing peak counts

II. Parameter constraint strategies

Chieh-An Lin and Martin Kilbinger

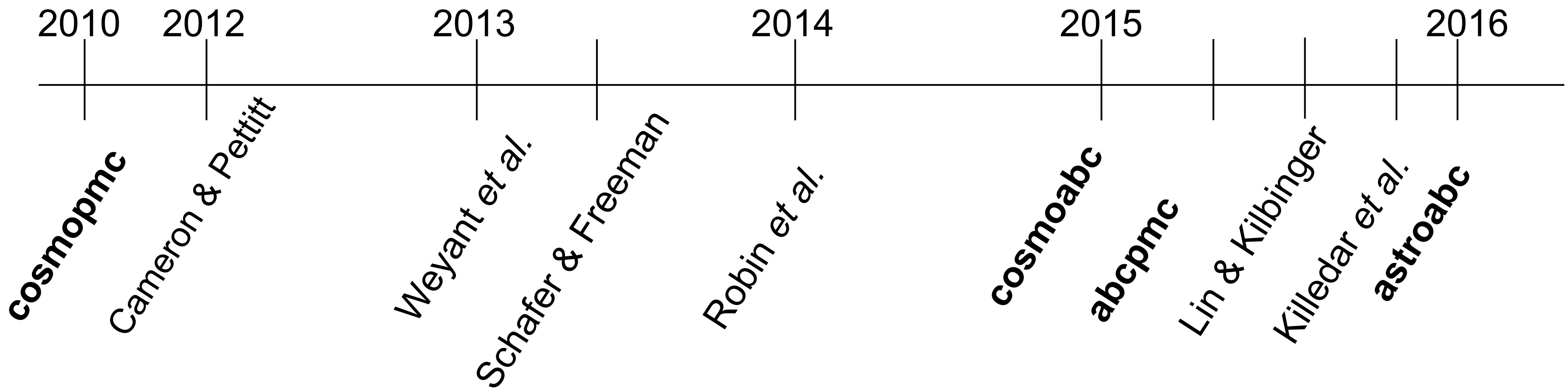
Service d'Astrophysique, CEA Saclay, Orme des Merisiers, Bât 709, 91191 Gif-sur-Yvette, France



arXiv:1507.05617v2

Weighted ABC: a new strategy for cluster strong lensing cosmology with simulations

M. Killedar,[1,2]★ S. Borgani,[2,3,4] D. Fabjan,[5,3] K. Dolag,[1,6] G. Granato,[2,3]
M. Meneghetti,[7,8,9] S. Planelles[2,3,4,10] & C. Ragone-Figueroa[3,11]



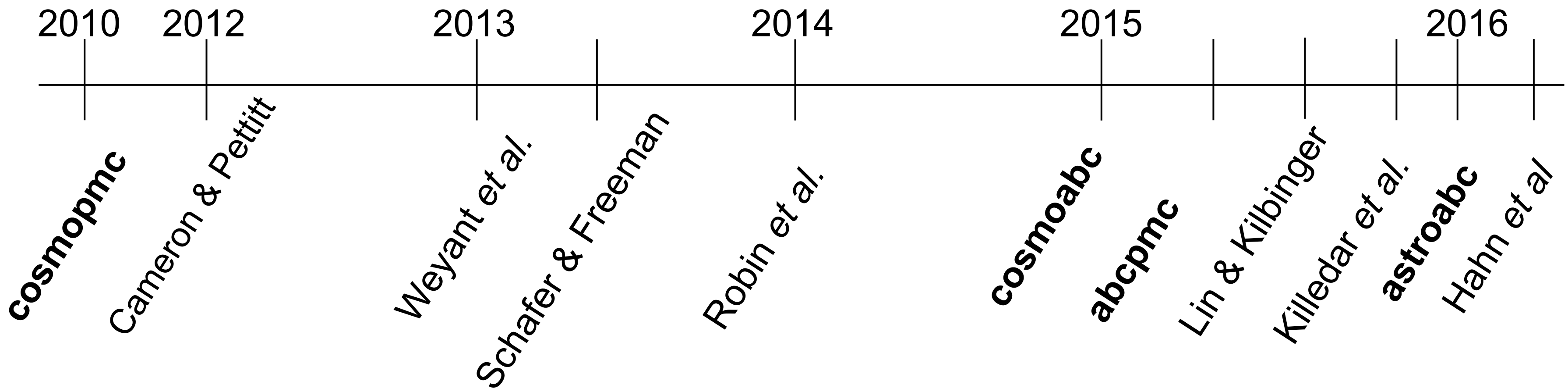
arXiv:1608.07606v1

astroABC: An Approximate Bayesian Computation Sequential Monte Carlo sampler for cosmological parameter estimation

Elise Jennings[a,b], Maeve Madigan[c]

[a] *Center for Particle Astrophysics, Fermi National Accelerator Laboratory MS209, P.O. Box 500, Kirk Rd. & Pine St., Batavia, IL 60510-0500*
[b] *Kavli Institute for Cosmological Physics, Enrico Fermi Institute, University of Chicago, Chicago, IL 60637*
[c] *Department of Theoretical Physics, University of Dublin, Trinity College, Dublin, Ireland*

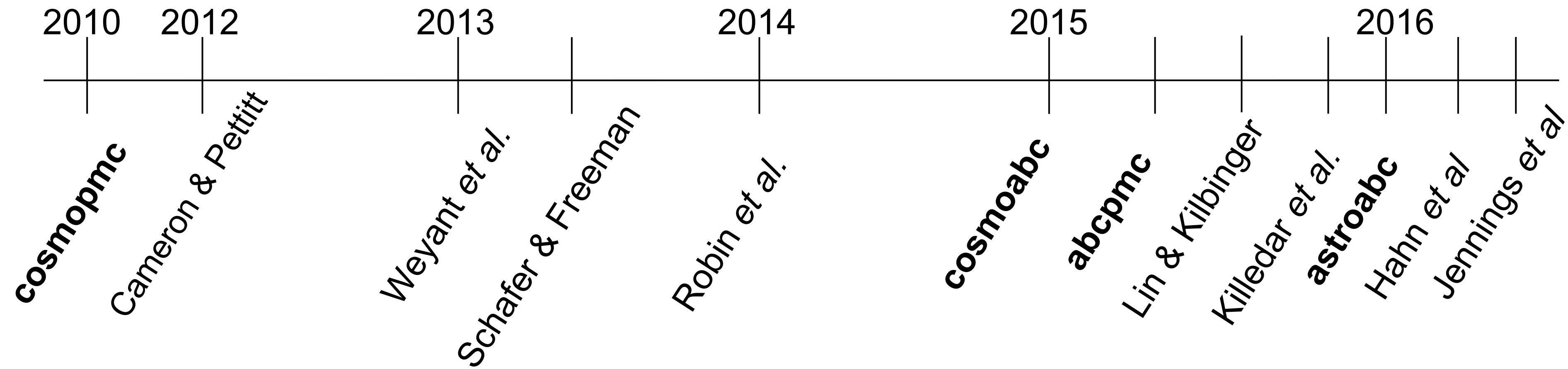
Cosmostatistics Initiative



arXiv:1607.01782v2

Approximate Bayesian Computation in Large Scale Structure: constraining the galaxy-halo connection

ChangHoon Hahn[†1], Mohammadjavad Vakili[†1], Kilian Walsh[1], Andrew P. Hearin[2], David W. Hogg[1,3,4,5], and Duncan Campbell[6]

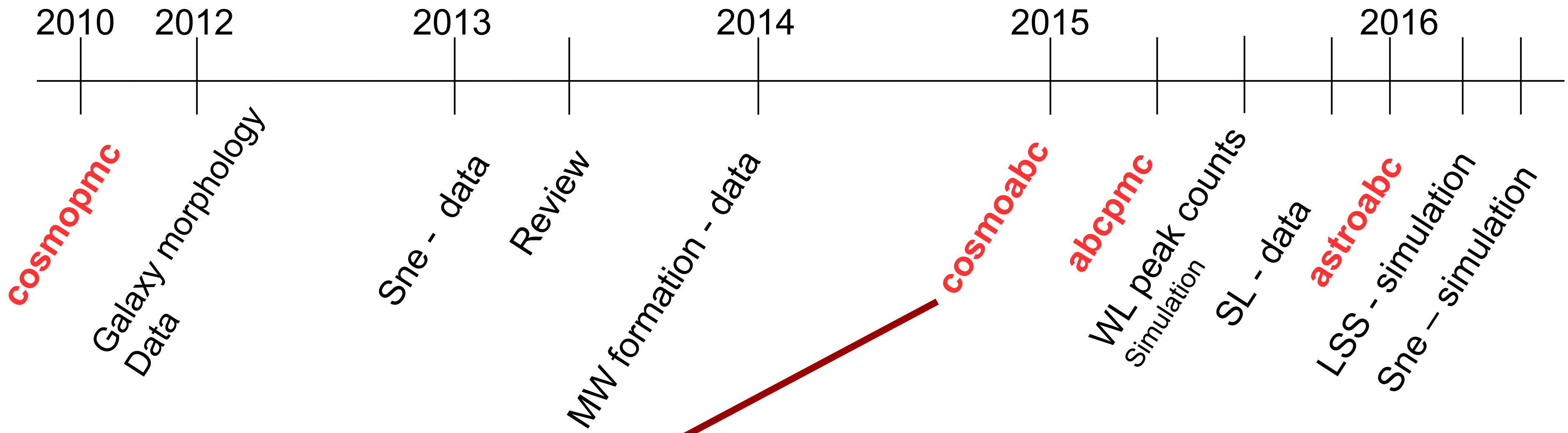


arXiv:1611.03087v1

A NEW APPROACH FOR OBTAINING COSMOLOGICAL CONSTRAINTS FROM TYPE IA SUPERNOVAE USING APPROXIMATE BAYESIAN COMPUTATION

Elise Jennings [1,2,4], Rachel Wolf [3], Masao Sako [3]

[1]Fermi National Accelerator Laboratory MS209, P.O. Box 500, Kirk Rd. & Pine St., Batavia, IL 60510-0500
[2]Kavli Institute for Cosmological Physics, Enrico Fermi Institute, University of Chicago, Chicago, IL 60637
[3]University of Pennsylvania Department of Physics & Astronomy, 209 South 33rd Street, Philadelphia, PA 19104-6396
[4]elise@fnal.gov



The LSST is a new kind of telescope. Currently under construction in Chile, the LSST is designed to conduct a ten-year survey of the dynamic universe. LSST can map the entire visible sky in just a few nights; each panoramic snapshot with the 3200-megapixel camera covers an area 40 times the size of the full moon.

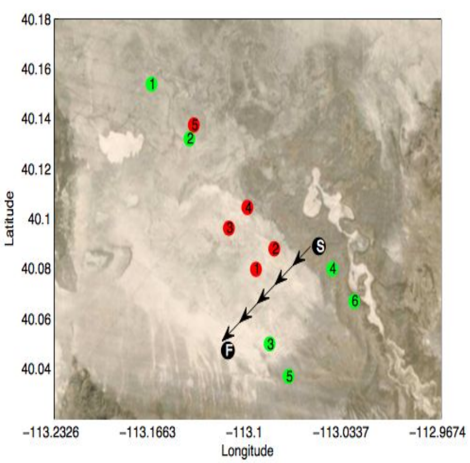
ABC in Meteorology and Geophysics

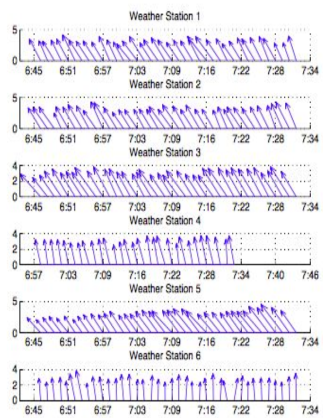
The Approximate Bayesian Computation methods in the localization of the atmospheric contamination source

P Kopka[1], A Wawrzynczak[1,2], M Borysiewicz[1]
[1] National Centre for Nuclear Research, Andrzeja Soltana 7, 05-400 Otwock,Poland
[2] Institute of Computer Sciences, Siedlce University, Poland
E-mail: `piotr.kopka@ncbj.gov.pl`

Using Approximate Bayesian Computation by Subset Simulation for Efficient Posterior Assessment of Dynamic State-Space Model Classes

Majid K. Vakilzadeh[a,b], James L. Beck[a,*], Thomas Abrahamsson[b]
[a] *Division of Engineering and Applied Science, California Institute of Technology, CA, USA*
[b] *Department of Applied Mechanics, Chalmers University of Technology, Gothenburg, Sweden*





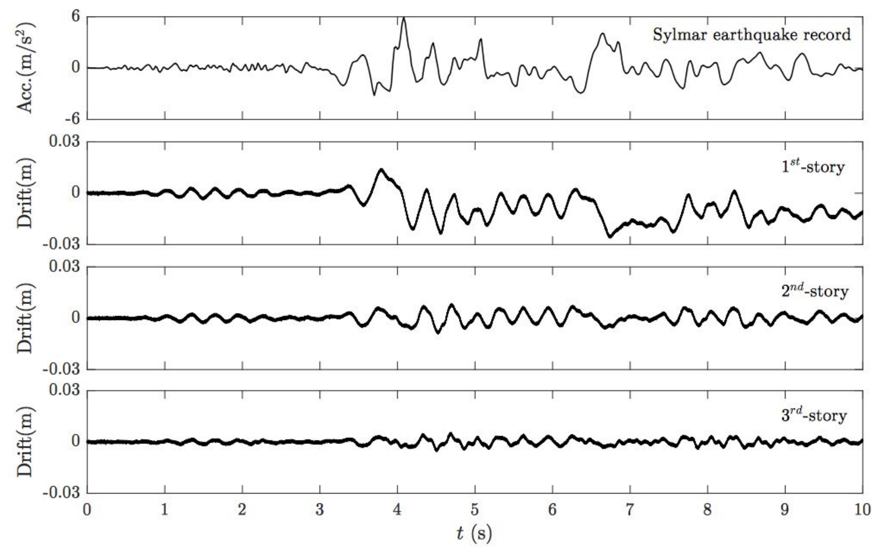
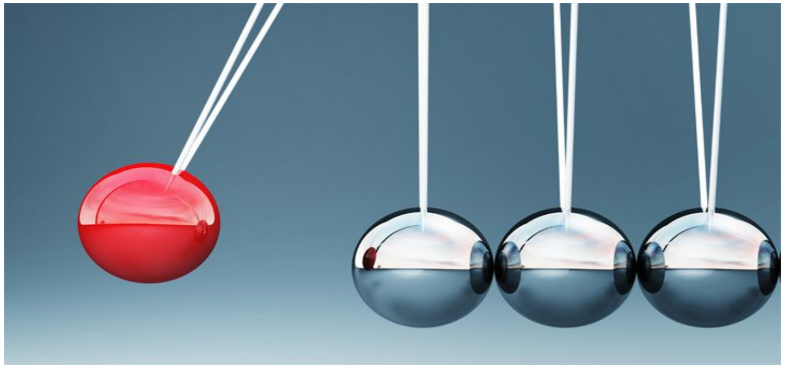


Fig. 8 *Inter-story drift time histories and the Sylmar ground-motion record (Example 2).*

Summary

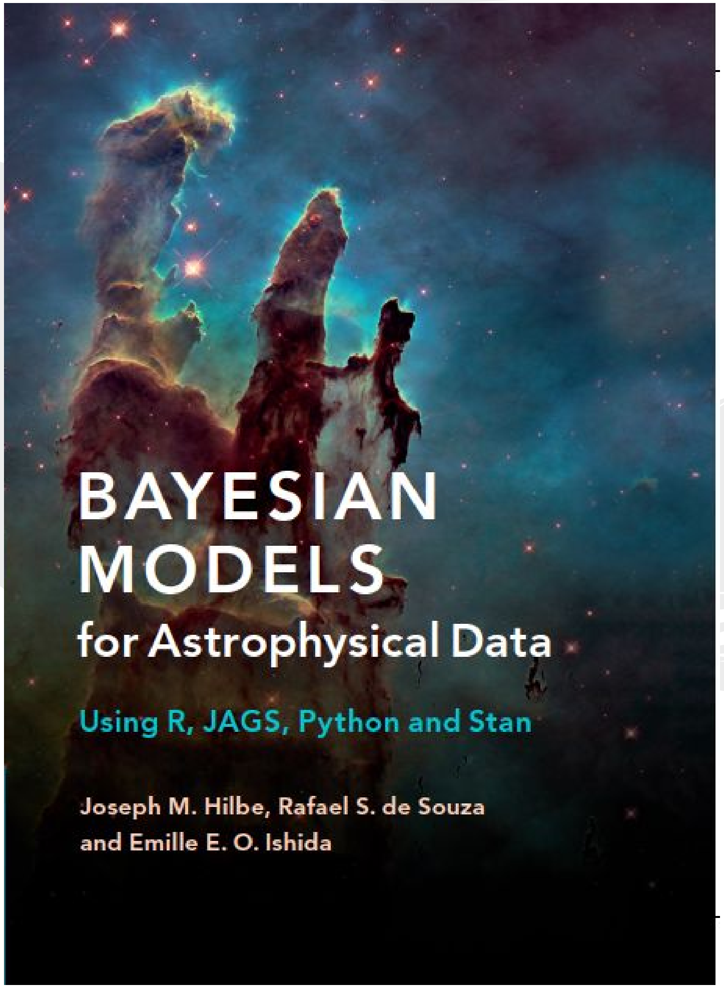
✓ Good alternative when likelihood is not available.
✓ Becomes more attractive with faster simulations.

✗ Definition of distance function/summary statistics



Promising perspective as it gains momentum inside the astronomical community – it is time to focus on data?

The Book



This volume is a very welcome addition to the small but growing library of resources for advanced analysis of astronomical data. Astronomers are often confronted with complex constrained regression problems, situations that benefit from computationally intensive Bayesian approaches. The authors provide a unique and sophisticated guide with tutorials in methodology and software implementation. The worked examples are impressive. Many astronomers use Python and will benefit from the less familiar capabilities of R, Stan and JAGS for Bayesian analysis. I suspect the work will also be useful to scientists in other fields who venture into the world of Bayesian computational statistics.

Eric D. Feigelson
Pennsylvania State University

Encyclopaedic in scope, a treasure trove of ready code for the hands-on practitioner"

Ben Wandelt
IAP - Lagrange Institute
Sorbonne University

This informative book is a valuable resource for astronomers, astrophysicists and cosmologists at all levels of their career. From students starting out in the field to researchers at the frontiers of data analysis, everyone will find insightful techniques accompanied by helpful examples of code. With this book, Hilbe, de Souza and Ishida are firming taking astrostatistics into the 21st century.

Dr Roberto Trotta, Reader in Astrophysics, Imperial College London

Table of Contents

Table 2.1: List of main R packages which should be installed in order to run the examples shown in subsequent chapters.

Name of package	Description
JAGS[1]	Analysis of Bayesian hierarchical models
R2jags[2]	R interface to JAGS
rstan[3]	R interface to Stan
lattice[4]	Graphics library
ggplot2[5]	Graphics library
MCMCpack[6]	Functions for Bayesian inference
mcmcplots[7]	Plots for MCMC Output

Table 2.2: List of Python packages which should be installed in order to run the examples shown in subsequent chapters.

Name of package	Description
matplotlib[1]	plotting library
numpy[2]	basic numerical library
pandas[3]	data structure tools
pymc3[4]	full Python probabilistic programming tool
pystan[5]	Python interface to Stan
scipy[6]	advanced numerical library
statsmodels[7]	statistics library

[1] http://matplotlib.org/, [2] http://www.numpy.org/, [3] http://pandas.pydata.org/ [4] https://pymc-devs.github.io/pymc3/
[5] https://pystan.readthedocs.org/en/latest/, [6] http://www.scipy.org/, [7] http://statsmodels.sourceforge.net/

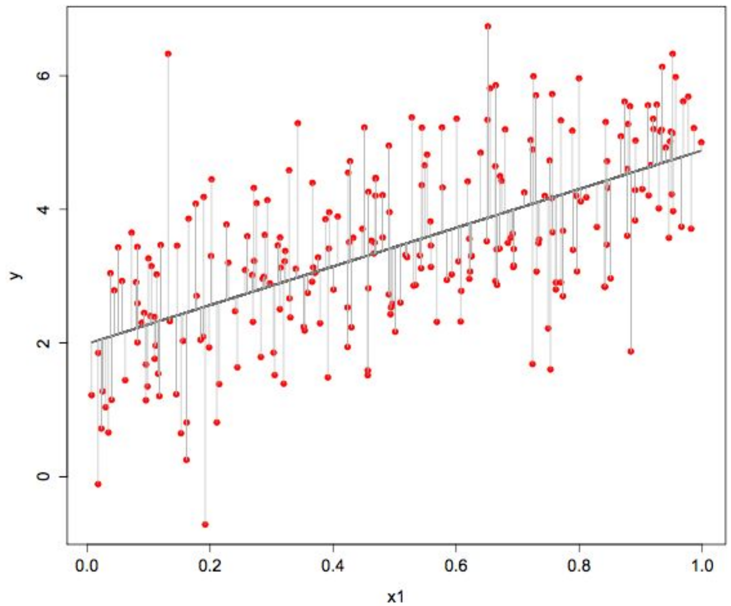
Code examples in both languages, from very simple linear models

Code 3.1: Basic linear model in R.

```
############### Data
set.seed(1056)              # set seed to replicate example
nobs = 250                  # number of obs in model
x1 <- runif(nobs)           # random uniform variable
alpha = 2                   # intercept
beta = 3                    # angular coefficient

xb <- alpha + beta* x1      # linear predictor, xb
y <- rnorm(nobs, xb, sd=1)  # create y as adjusted random normal variate

############### Fit
summary(mod <- lm(y ~ x1))   # model of the synthetic data.
```



The equivalent routine in Python can be written as

Code 3.2: Ordinary Least Square regression in Python without formula

```
import numpy as np
import statsmodels.api as sm
############### Data
np.random.seed(1056)                # set seed to replicate example
nobs= 250                           # number of obs in model
x1 = uniform.rvs(size=nobs)         # random uniform variable

alpha = 2.0                        # intercept
beta = 3.0                         # slope

xb = alpha + beta * x1             # linear predictor, xb
y = norm.rvs(loc=xb, scale=1.0, size=nobs) # create y as adjusted
                                            # random normal variate
############### Fit
unity_vec = np.full((nobs,),1, np.float)  # unity vector
X = np.column_stack((unity_vec, x1))      # build data matrix with intercept
results = sm.OLS(y, X).fit()
```

To more complicated ones, such as Hurdle Models

$$M_{\star;i} \sim \begin{cases} Bernoulli(p_i) & \text{if } M_{\star;i} = 0 \\ LogNormal(\mu_i, \sigma) & \text{otherwise} \end{cases}$$

$$logit(p_i) = \gamma_1 + \gamma_2 \times M_{\mathrm{dm};i}$$

$$\mu_i = \beta_1 + \beta_2 \times M_{\mathrm{dm};i}$$

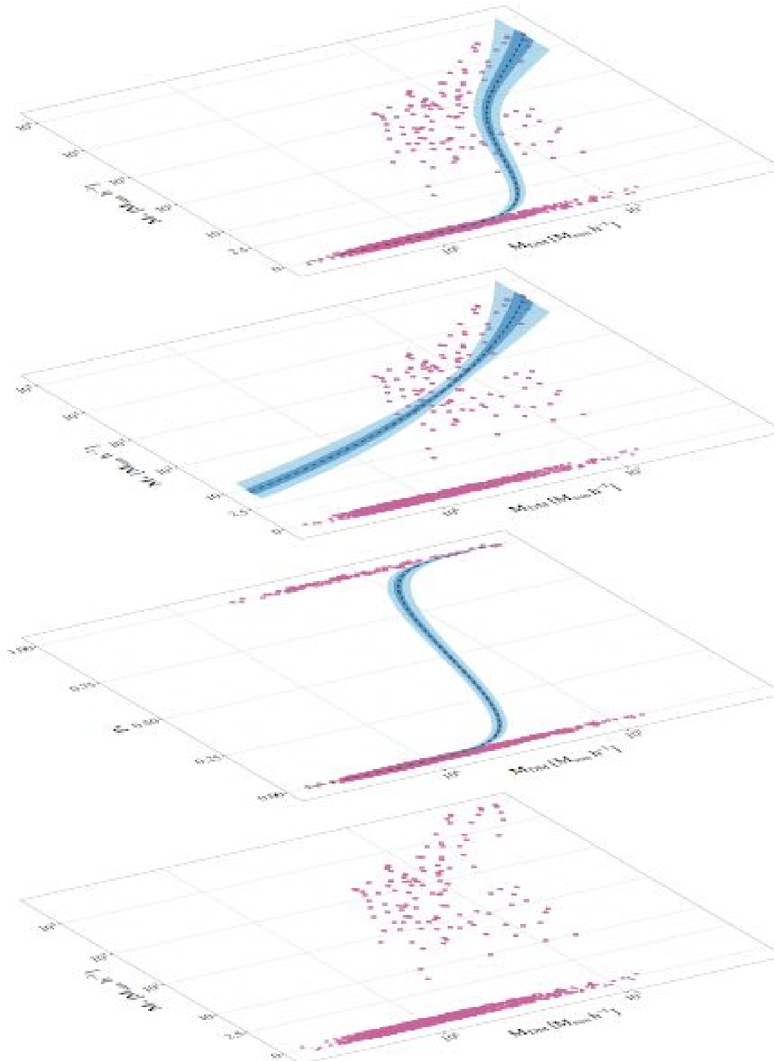
$$\beta_j \sim Normal(0, 10^3)$$

$$\gamma_j \sim Normal(0, 10^3)$$

$$\sigma \sim Gamma(0.001, 0.001)$$

$$i = 1, \cdots, N$$

$$j = 1, \cdots, K$$



COIN
Cosmostatistics Initiative