THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

COIN Cosmostatistics Initiative

Cosmostatistics Initiative

Catalyzing Interdisciplinarity

Rafael S. de Souza, on behalf of COIN

Artificial Intelligence in Astronomy,
ESO, Garching, Germany, 22–26 July 2019


context
retention
customer
research
sales
customer
plan
management
analysis
planning
implementation
monitor
risk
project
impact
complex
data
probability
assessment
identification
risk
process
impact
sales
opportunity
retention
strategy
impact
assessment
sales
data
performance
management
cost
evaluation
process
performance
cost
research
cost
research
evaluation
data
treatment
impact
project
plan
strategy
data
research
resources
risk
performance
assessment
context
evaluation
scope
performance
important
data
project

Collaboration as a goal by itself

Rafael S. de Souza

Alberto Krone-Martins

Emille E. O. Ishida

LISBOA | UNIVERSIDADE DE LISBOA

The first analytical expression to estimate photometric redshifts suggested by a machine

A. Krone-Martins,[1★] E. E. O. Ishida[2,3] and R. S. de Souza[4,5]

Why not more people?

MNRASL **443,** L34–L38 (2014)

doi:10.1093/mnrasl/slu067

Symbolic Regression

See also Ivan's slides.

Time Before Time - Pre-COIN

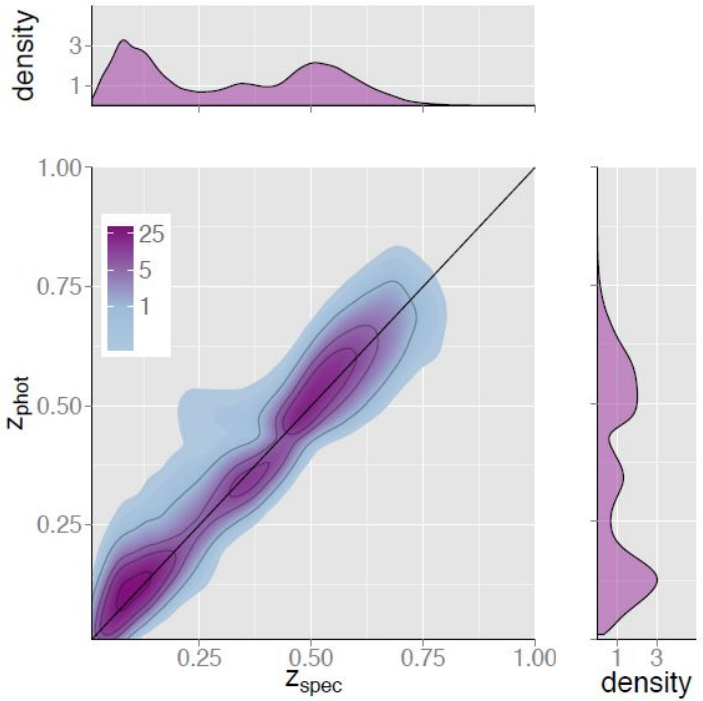
Atomic operators:
+, - , x , / , pow

1 - Construction of an **analytical** expression from a dictionary

2 - Optimization via **genetic algorithm** with a **sparsity** enforcing principle

Derived expression:

$$z_{\text{phot}} = \frac{0.4436r - 8.261}{24.4 + (g-r)^2(g-i)^2(r-i)^2 - g} + 0.5152(r-i).$$



Krone-Martins, Ishida & de Souza, MNRASL 443 (2014)

STATISTICAL CHALLENGES in 21st CENTURY COSMOLOGY

IAU SYMPOSIUM 306 Lisbon Portugal 25-29 May 2014

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*

Team Science Learning

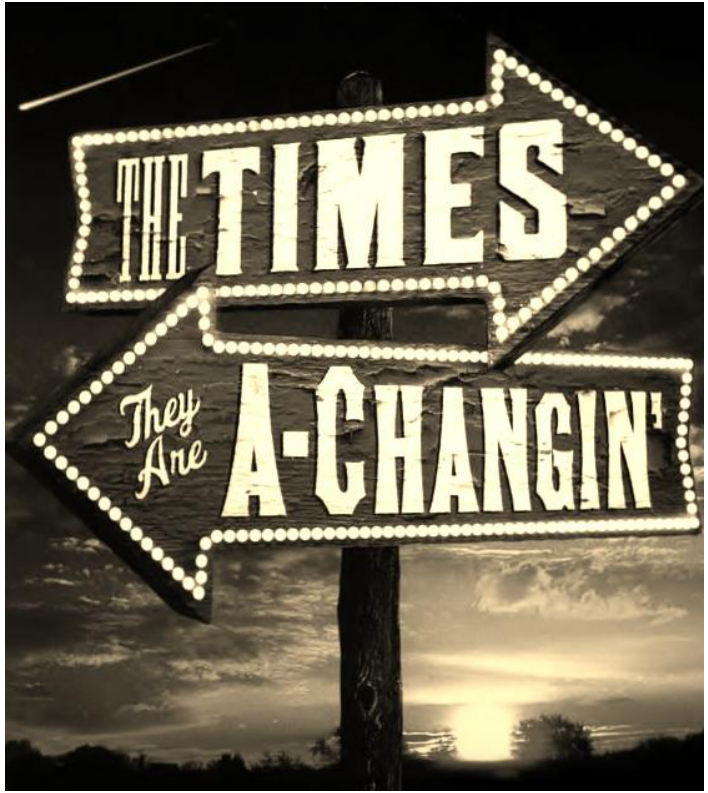
A worldwide endeavour aimed to foster interdisciplinary collaborations around Astronomy.

The Cosmostatistics Initiative

Interdisciplinary science development

<u>The COIN Residence Program (CRP)</u>

- ✓ Interdisciplinarity
- ✓ Unstructured organization
- ✓ People-centric
- ✓ Ambidextrous:
  personal + community needs



Interdisciplinary cross-fertilization

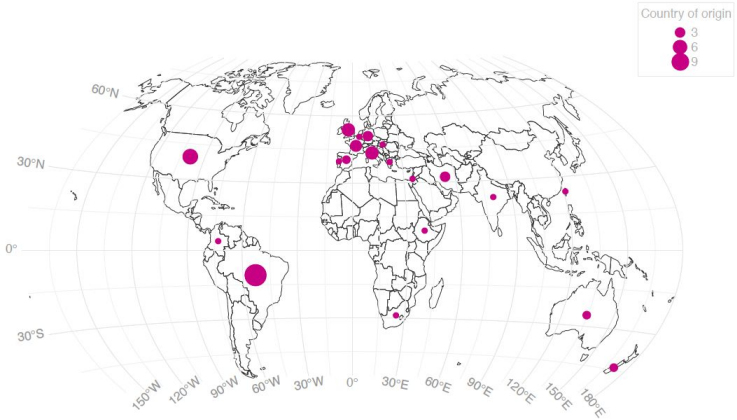
The COIN Residence Program (CRP)

Step 1 - Choose the people

Step 2 - Let the project emerge

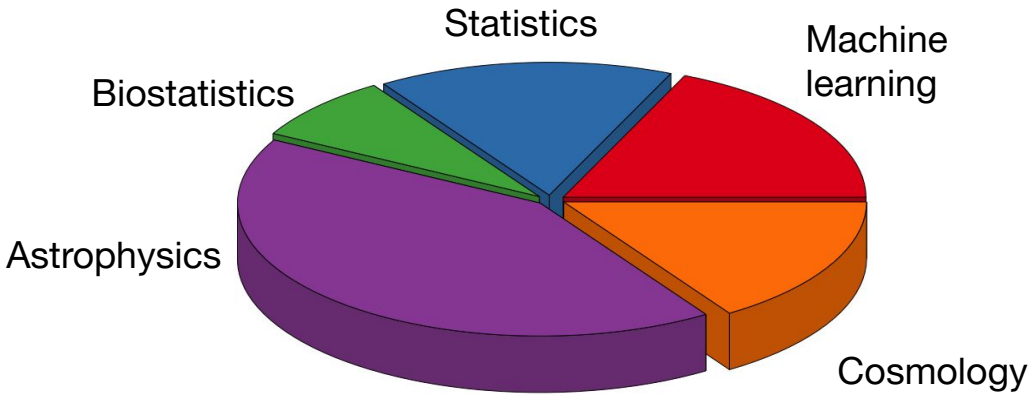
Step 3 - Make the project converge

A worldwide task force



Over **60** researchers from **15** countries

Management model : concepts of startups and meta-studies of interdisciplinary teams.



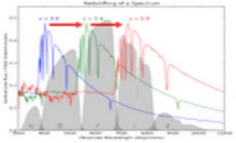
See e.g. Am J Prev Med. 2008, 35, S96-115. *The ecology of team science understanding contextual influences on transdisciplinary collaboration*

Innovation for interdisciplinary science development

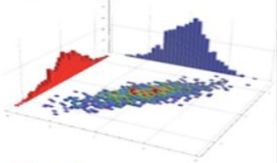
COIN
Cosmostatistics Initiative

COIN software

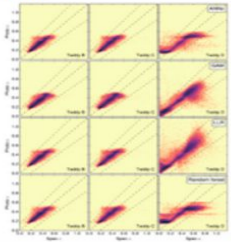
CosmoPhotoZ
Fast photo-z estimation via GLMs.



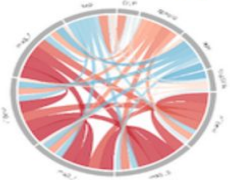
Cosmoabc
Likelihood free inference for cosmology



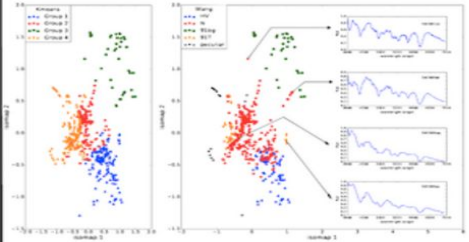
Happy and Teddy Catalogues for realistic photo-z validation



AMADA
Analysis of Muldimensional Astronomical DAtasets



DRACULA
Dimensionality Reduction And Clustering for Unsupervised Learning in Astronomy



https://cosmostatistics-initiative.org

Scientific outcomes

In 5 years

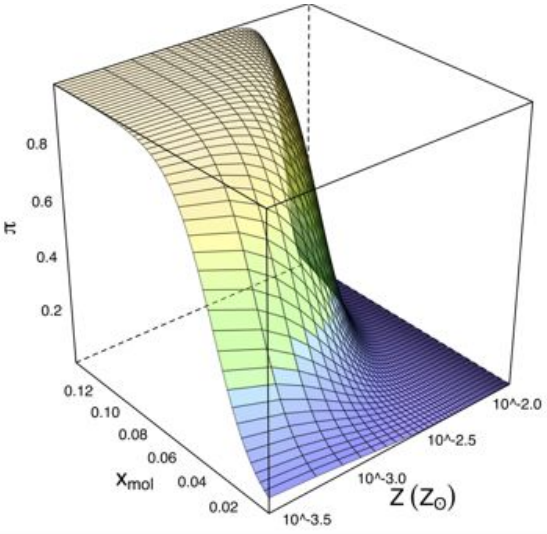
	Paper	Citation
1	GLM I	de Souza *et al.*, 2015
2	GLM II	Elliott *et al.*, 2015
3	GLM III	de Souza *et al.*, 2015
4	AMADA	de Souza & Ciardi, 2015
5	CosmoABC	Ishida *et al.*, 2015
6	DRACULA	Sasdelli *et al.*, 2016
7	AGNlogit	de Souza *et al.*, 2016
8	PhotoZ	Beck *et al.*, 2017
9	AGNgmm	de Souza *et al.*, 2017
10	GalINLA	Gonzalez-Gaitan *et al.*, 2018
11	ActSNclass	Ishida *et al.*, 2018
12	COIN-Gaia	Cantat-Gaudin *et al.*, 2018
13	Hurdle	Hattab *et al.*, 2019
14	SNCosmo	Moews *et al.*, 2018
15	Deblending	Boucaud *et al.*, 2019

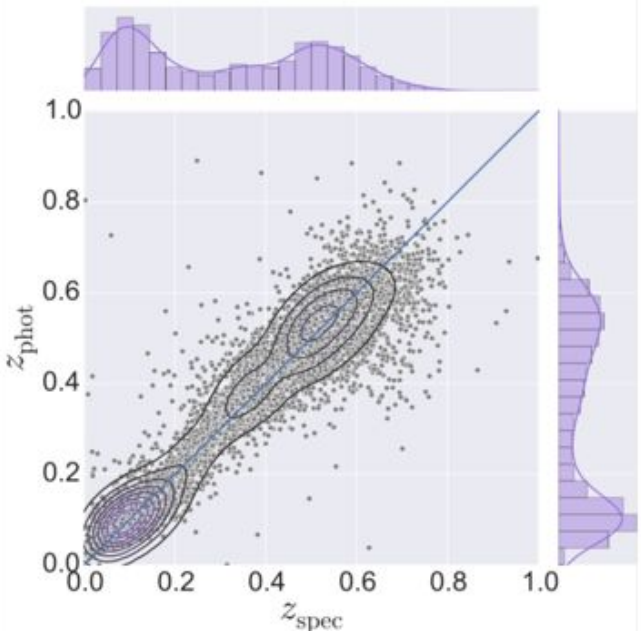
	Code	Citation
1	CosmoPhotoZ	de Souza *et al.*, 2014,
2	AMADA	de Souza & Ciardi, 2015
3	CosmoABC	Ishida *et al.*, 2015
4	DRACULA	Aguena *et al.*, 2015
5	CoinINLA	Gonzalez-Gaitan *et al.*, 2018

+ 1 galaxy catalog
+ 1 GMM tutorial
+ 2 photo-z catalogs

1 - GLM I: Binomial Regression

Logistic regression applied to cosmological simulations. Similar to ANN, but more interpretable.



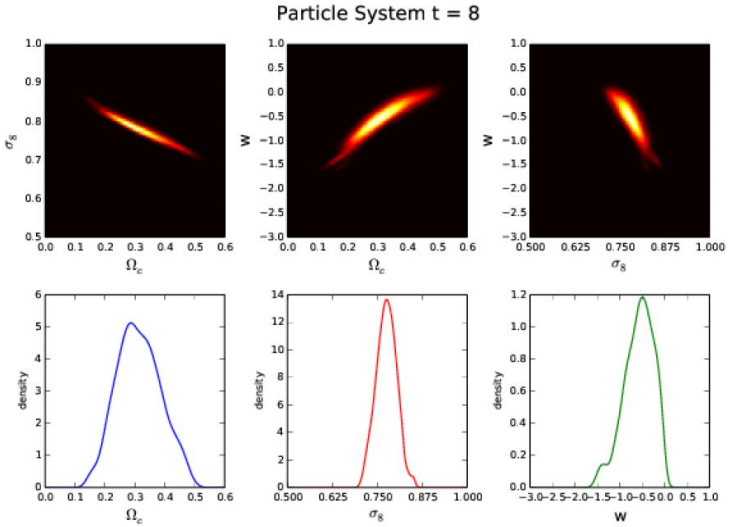


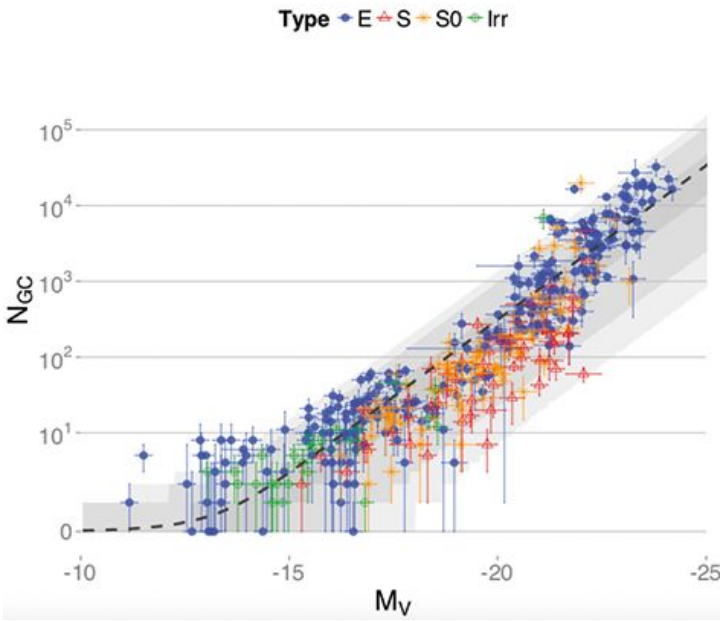
2 - GLM II: Gamma Regression

Semi-supervised Learning, PCA, KPCA, GLMs applied to photometric redshifts.

3 - GLM III: Negative Binomial Regression

Bayesian LASSO NB under discrete measurement errors probe overdispersed globular cluster data.





4 - Approximate Bayesian Computation (CosmoABC)

First package for likelihood-free inference in Cosmology.

Non-exhaustive list of techniques employed by COIN

- Binomial, gamma, negative binomial, Hurdle Models, GPs,
- Integrated Nested Laplace Approximation,
- Approximate Bayesian Computation, HMC,
- Convolutional Neural Networks, Transfer Learning, Active Learning
- Hierarchical Bayesian Models, Symbolic Regression,
- Random Forests, Support Vector Machines, GAMs
- PCA, KPCA, ISOMAPs, Self-Organized Maps, Minimum Spanning Trees, Autoencoders
- Gaussian Mixture Models, K-means, KD-trees
- LASSO regularization, Partial Pooling, Population Monte-Carlo, Propensity Score Matching.

COIN'S

Dracula

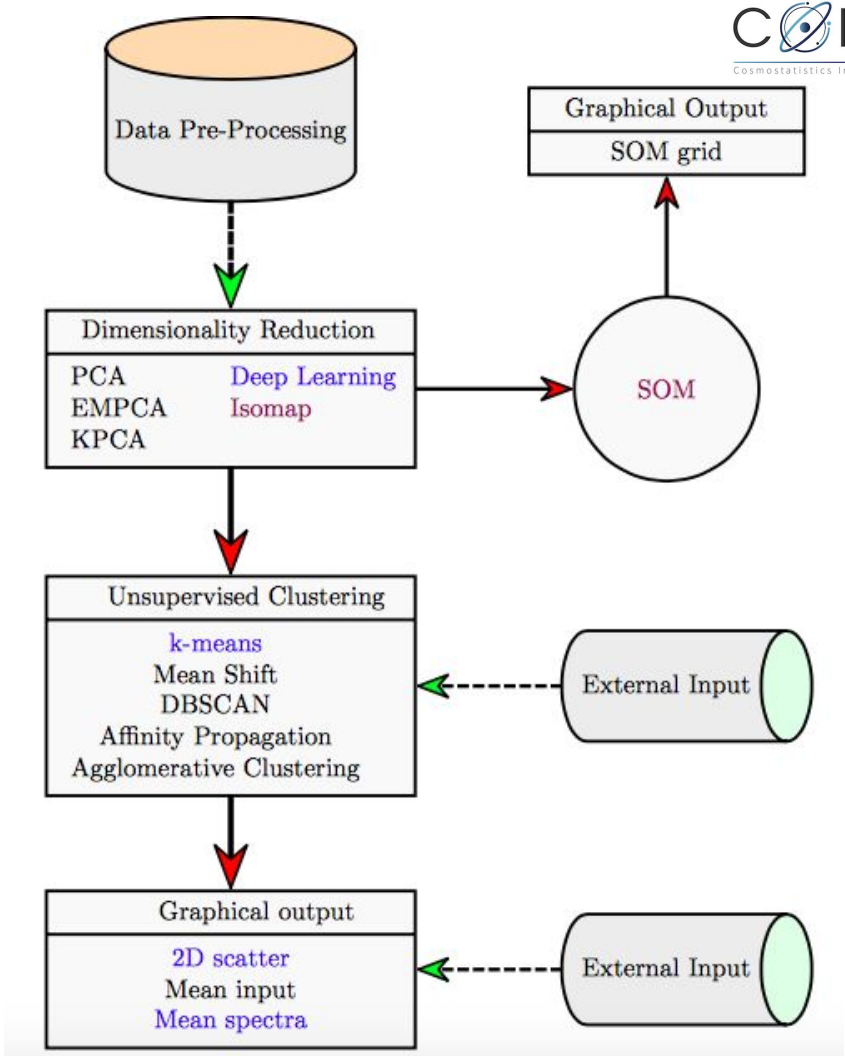
Dimensionality Reduction And Clustering
for Unsupervised Learning in Astronomy

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** ▾

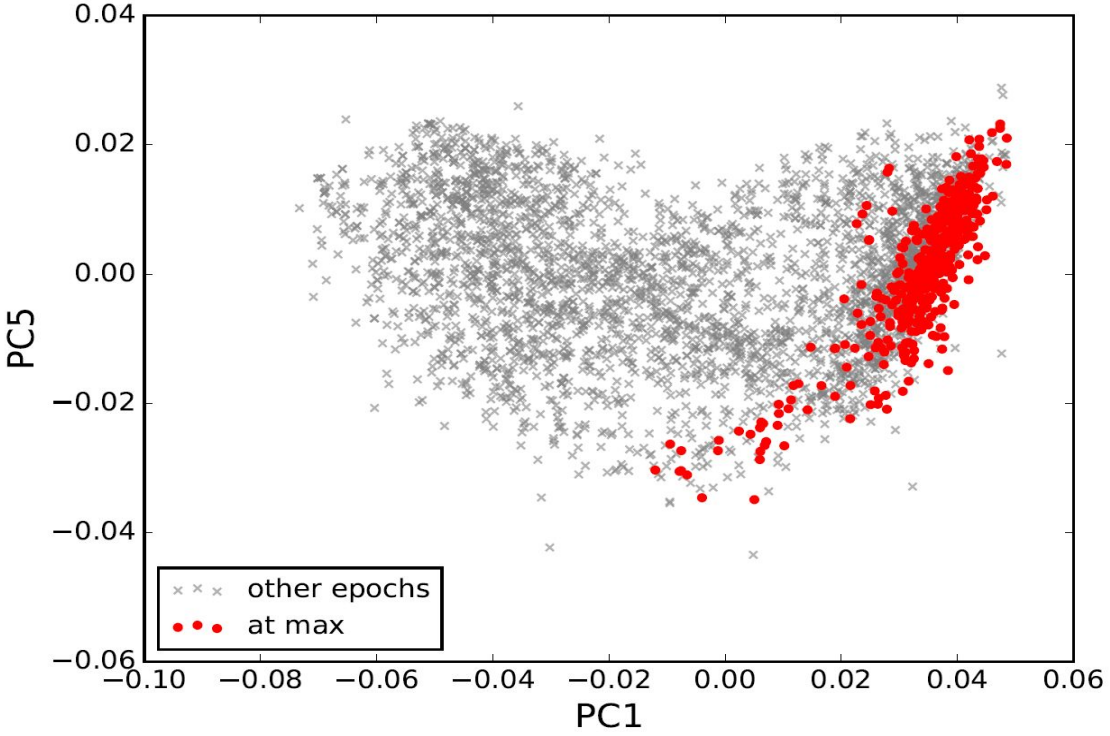
*From **COIN Residence Program #2***



Transfer Learning

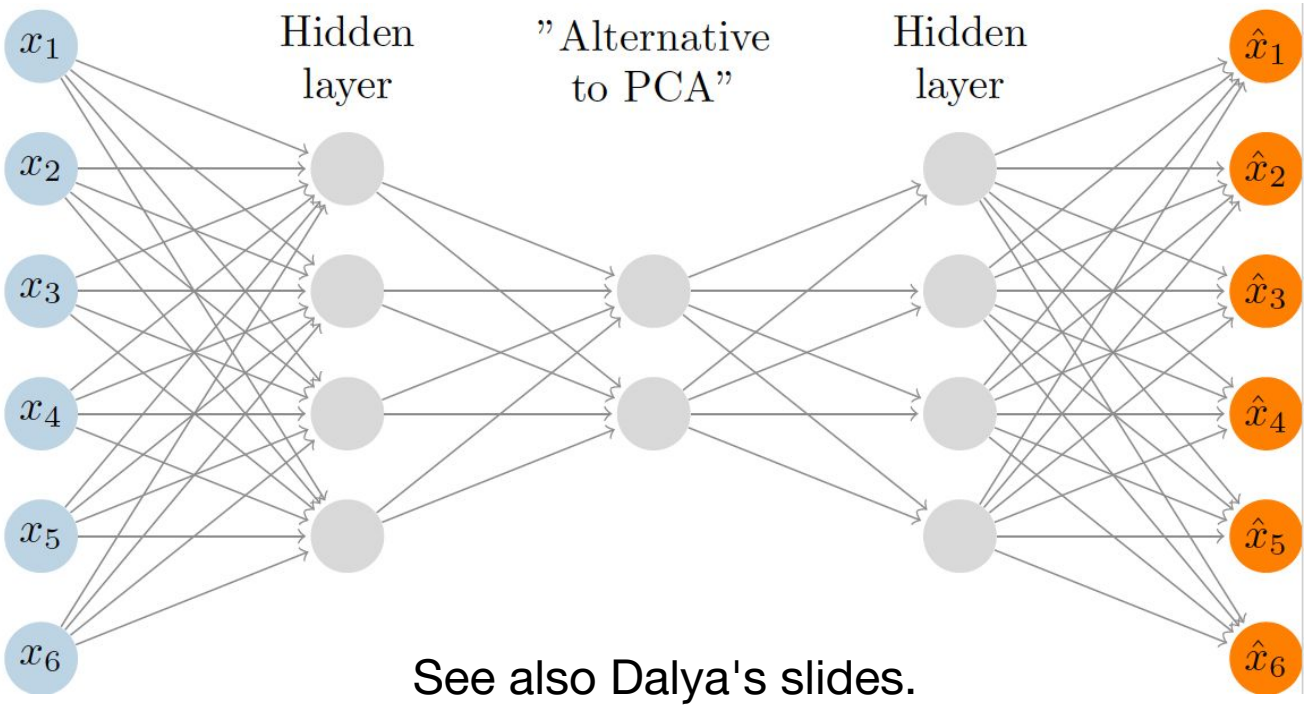
Initial dataset with SNe Ia near peak had more features than objects.

Transfer learning from spectrums taken in other epochs (usually not used)



Autoencoders

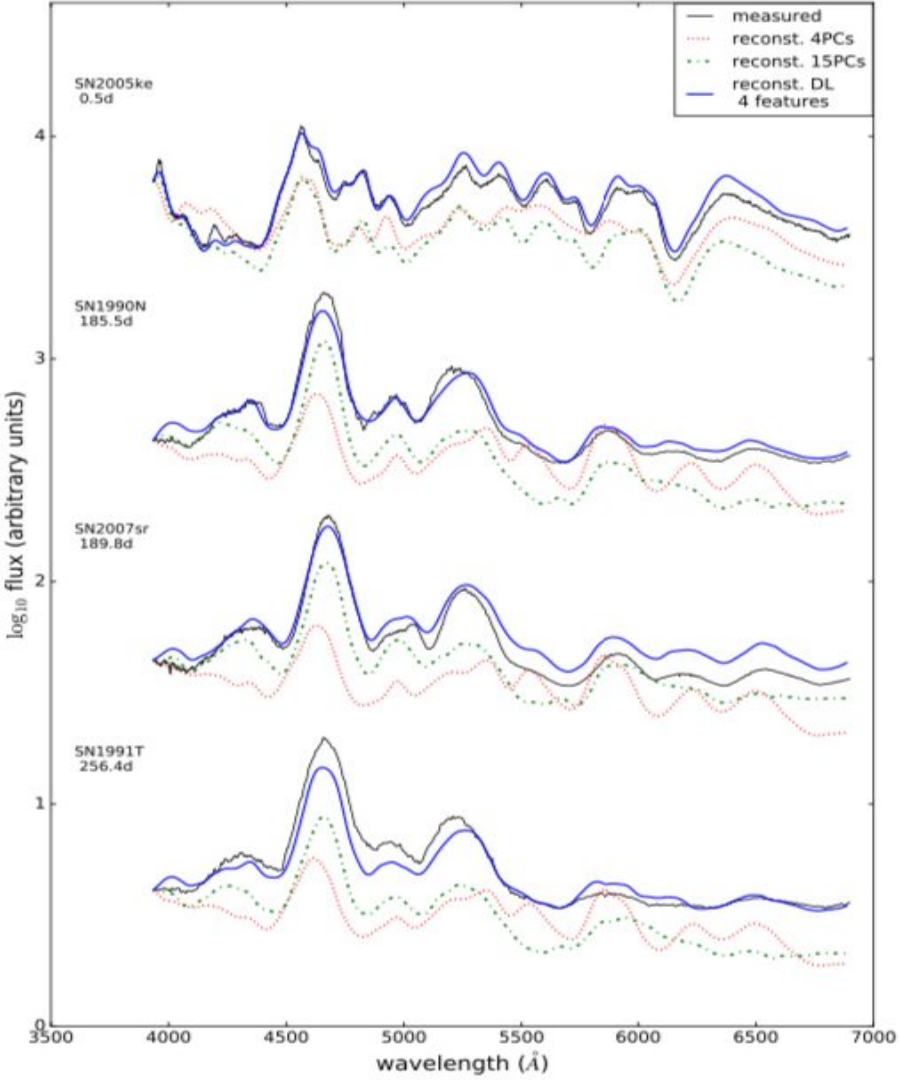
Dimensionality reduction



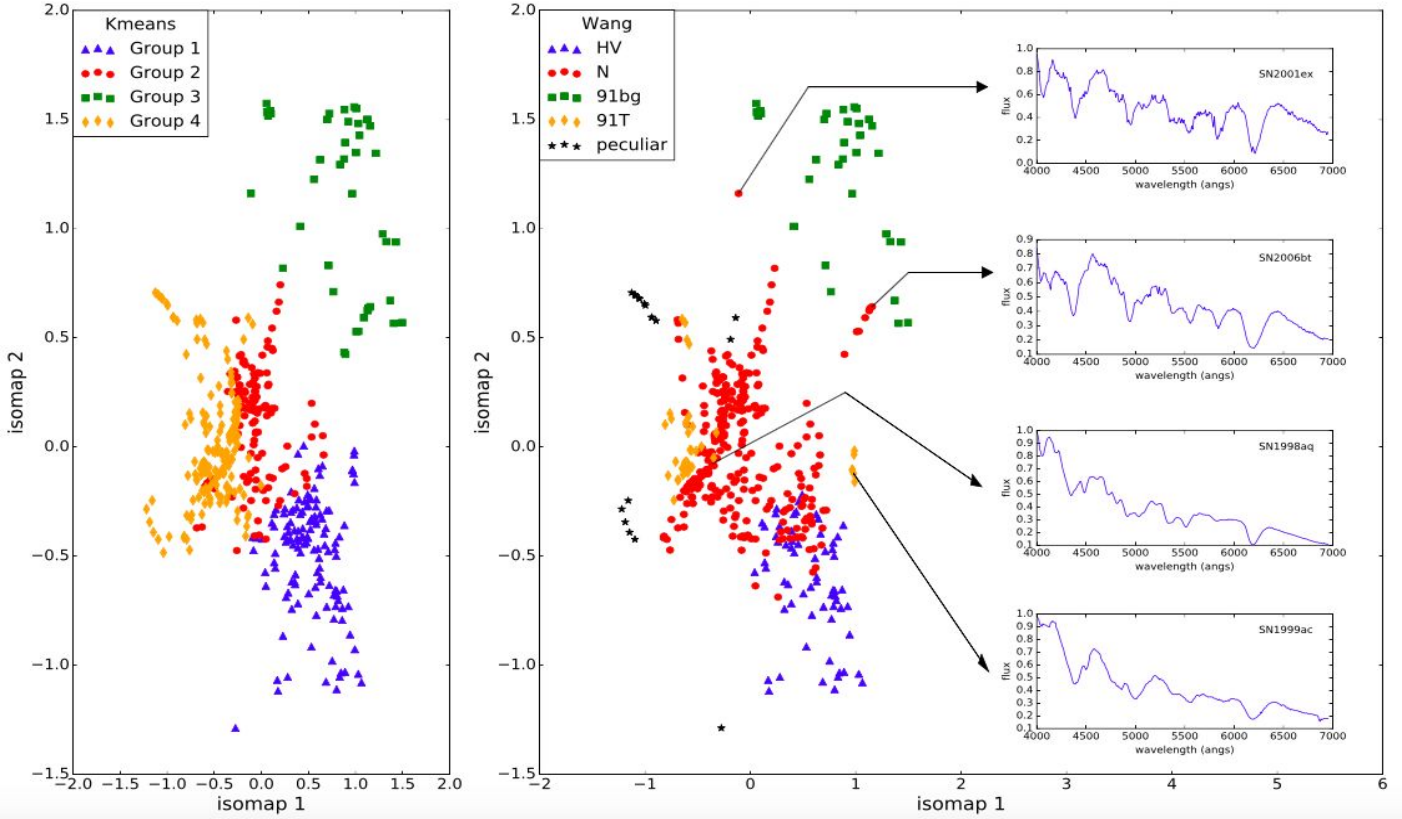
See also Dalya's slides.

COIN
Cosmostatistics Initiative

Deep Autoencoder vs Principal Component reconstruction

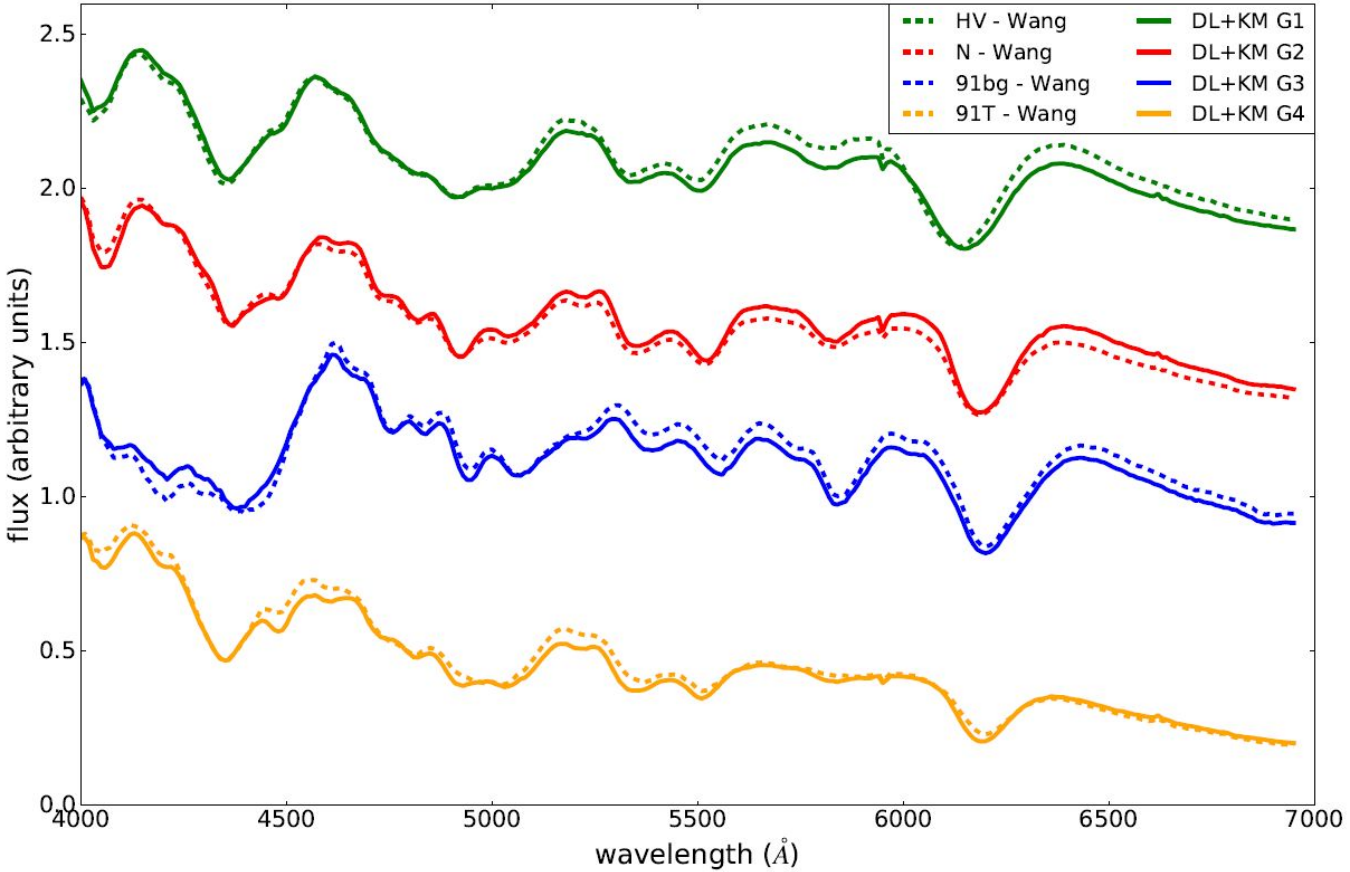


2D visualization of 4D Deep Learning parameter space



Mean spectra by ML x human

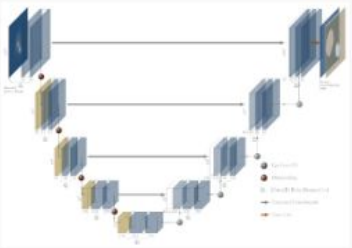
K-means with 4 groups x Wang



Recent highlights

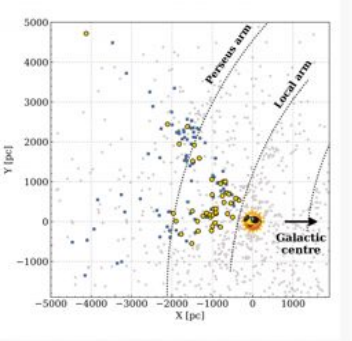
COIN and LSST-DESC join forces in astro-wise machine learning research

Read more



Sight beyond sight: Teasing Galaxies Apart with Deep Learning

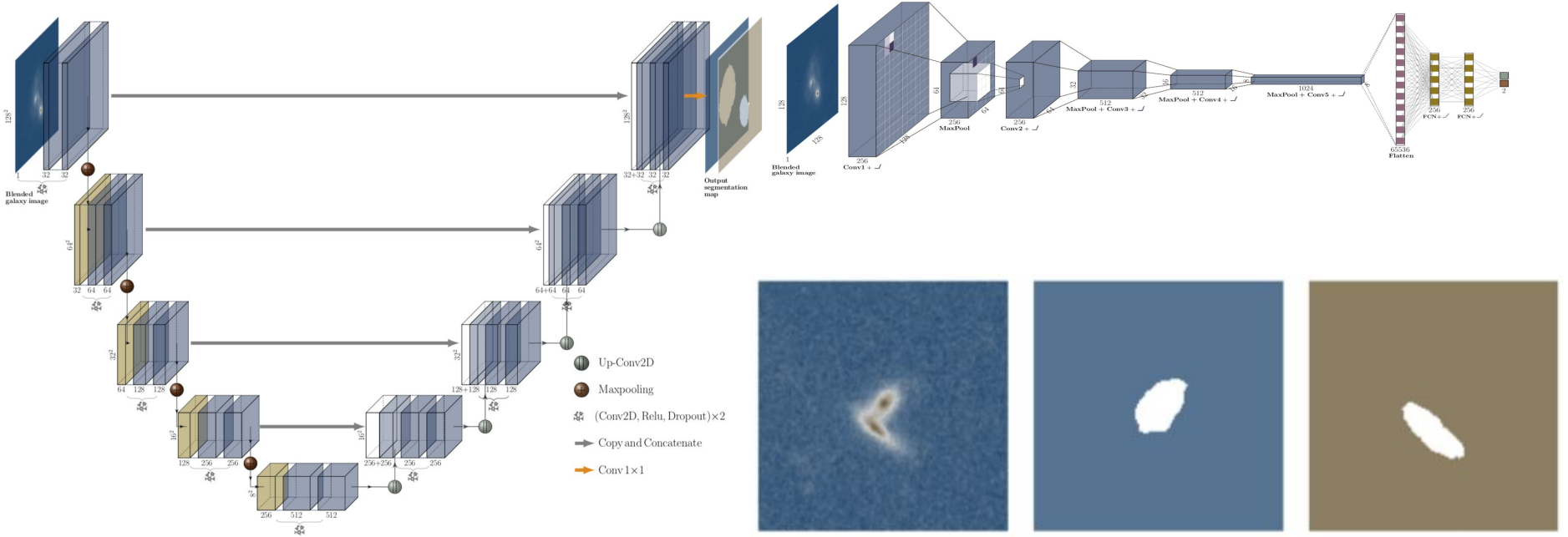
Read more



Now you see me: COIN extends the open cluster census in the solar neighborhood with Gaia DR2

Read more

Deep Learning for Deblending



*From **COIN Residence Program #5**, MNRAS, 2019*

The Cosmostatistics Initiative

The Cosmostatistics Initiative (COIN) is an international network which aims to create an interdisciplinary environment where collaborations between astronomers, statisticians and machine learning experts can flourish. The group utilizes a management model which can find parallel in technological start-ups: based on a dynamic, non-hierarchical and people-centric approach.

The LSST Dark Energy Science Collaboration

The LSST Dark Energy Science Collaboration (DESC) is an international collaboration preparing for a variety of cosmological analyses with the Large Synoptic Survey Telescope (LSST) data. In advance of LSST's first observations, DESC will help prepare for LSST science analysis, make synergistic connections with ongoing cosmological surveys and provide the dark energy community with state of the art analysis tools.

RESSPECT

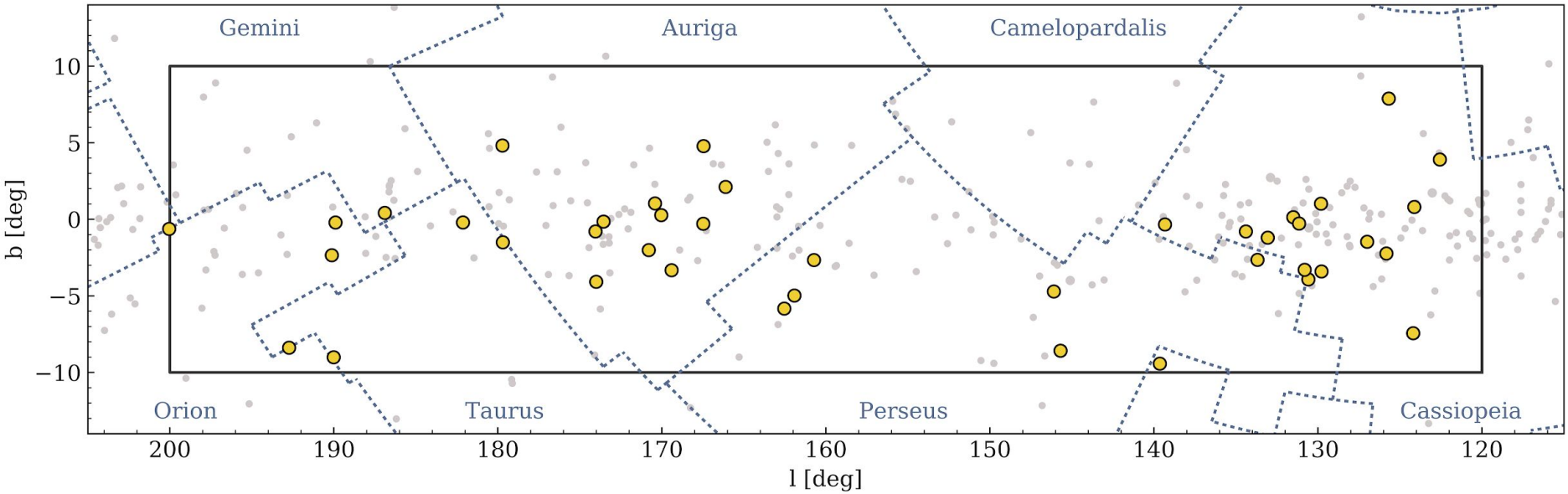
The REcommendation System for SPECTroscopic follow-up (RESSPECT) is a collaboration between COIN and LSST-DESC which aims to adapt active learning strategies for the construction of optimized training samples for supernova photometric classifcation in the context of LSST.

The team is formed by researchers from both collaborations who are working together in the development of a recommendation system which will enable informed decisions regarding the allocation of spectroscopic follow-up resources and consequent optimized scientific results from purely photometric samples.

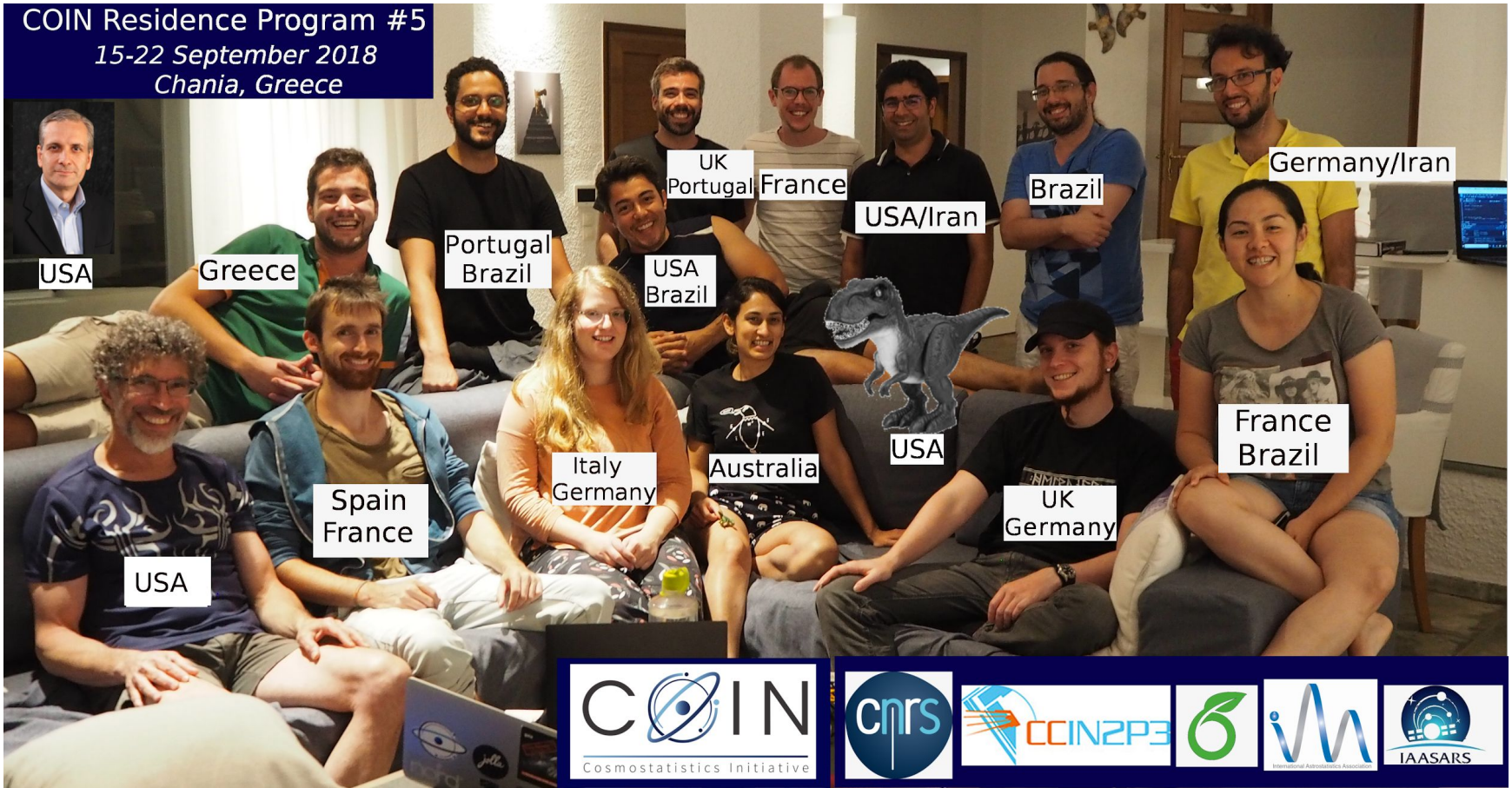
Now you see me: COIN extends the open cluster census in the solar neighborhood with Gaia DR2

See also Alberto's slides

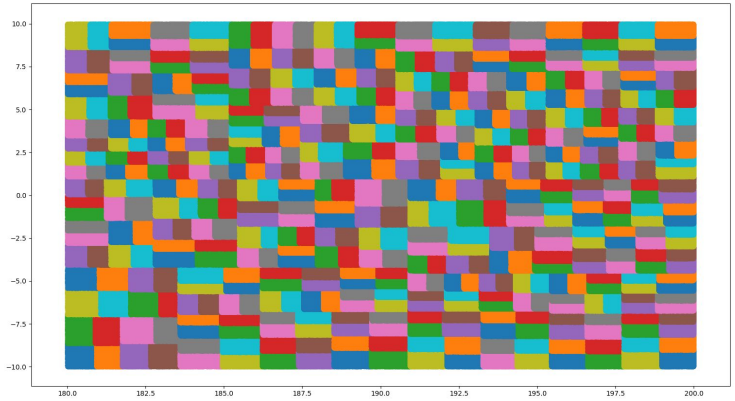
A&A, 624, A126, 17, 2019

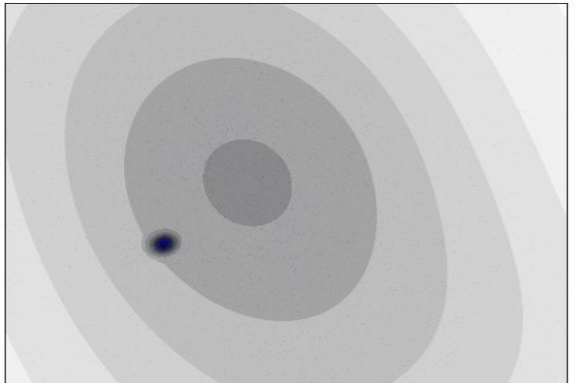


The Team


COIN Residence Program #5
15-22 September 2018
Chania, Greece
USA
Greece
Portugal
Brazil
USA
Brazil
UK
Portugal
France
USA/Iran
Brazil
Germany/Iran
Italy
Germany
Australia
USA
France
Brazil
Spain
France
UK
Germany
USA
COIN
Cosmostatistics Initiative
cnrs
CCIN2P3
IAASARS

Astro-aware statistical learning



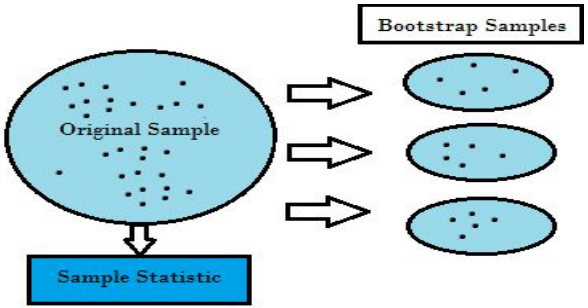


Fast density-aware partitioning of the sky via a *k*-d tree in the spatial domain of Galactic coordinates.

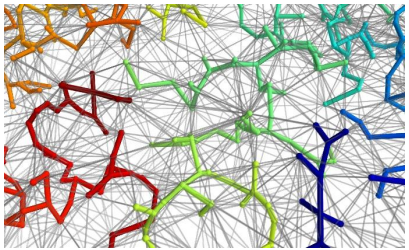
Only meaningful cases are further scrutinized, i.e. low variance in proper motion.

Recommender system

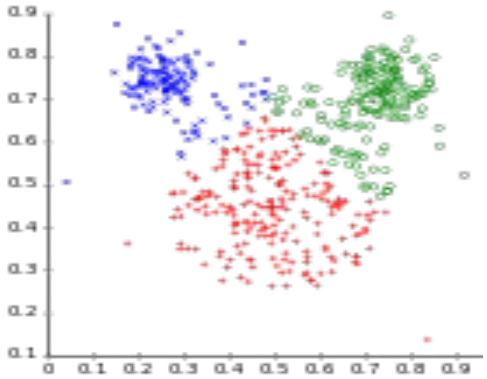
Astro-aware statistical learning



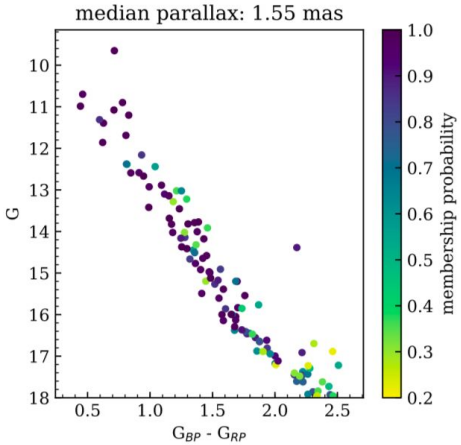
Bootstrap for measurement uncertainty



Sanity check against a random field via minimum spanning trees

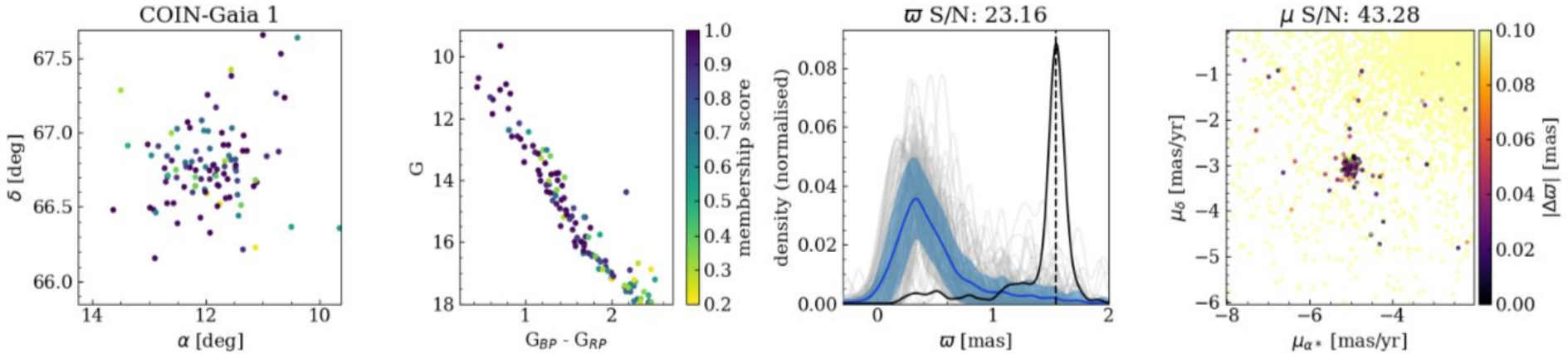


Iterative K-means in the space of proper motions.



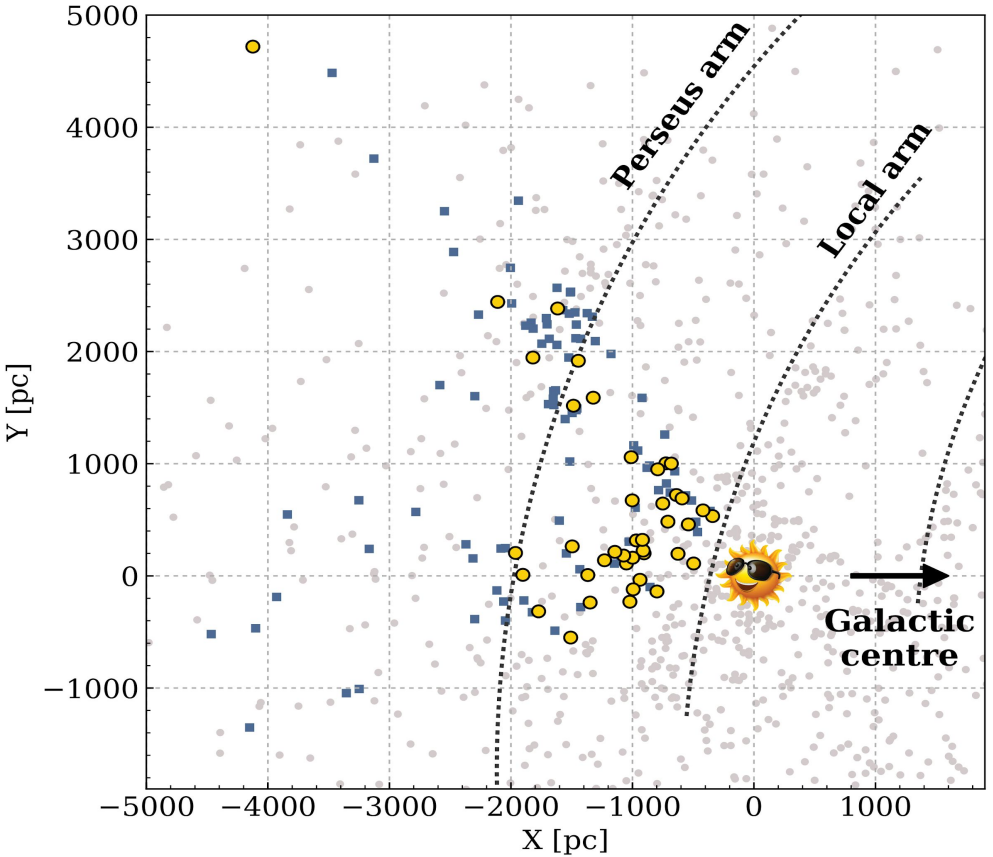
Independent expert validation

COIN-Gaia 1


COIN-Gaia 1
δ [deg]
α [deg]
G
G_BP - G_RP
membership score
ϖ S/N: 23.16
density (normalised)
ϖ [mas]
μ S/N: 43.28
μ_δ [mas/yr]
μ_α* [mas/yr]
|Δϖ| [mas]

COIN-Gaia Open Clusters

We reported the discovery of **41** new stellar clusters



COIN Residence Program #6

Chamonix - France, 24 - 31 August 2019

- COIN was constructed under the lemma *people come first*.
- The scientific project emerges from a shared group interest.
- They are a product of the interaction between a unique group of people, whose materialisation is only possible in an environment which profoundly respects the diversity of their scientific backgrounds, gender, career stages and nationalities.


COIN
Cosmostatistics Initiative
COIN
https://cosmostatistics-initiative.org
About
Chairs
Projects
Residence Programs
Focus
Community
Highlights
Partners

Cambridge Elements

Astrostatistics

The development of new statistical methods and their application in the quest to answer astronomical endeavours has given rise, in the past few decades, to the field of Astrostatistics. This exciting area, that emerges at the intersection of Statistics and Astronomy, provides a fertile ground for both communities to grow. This innovative series from Cambridge University Press is a new medium to communicate advances in Astrostatistics to enable a steady astronomical–statistical dialogue, to develop a common language for the benefit of both communities, and to catalyse the synergy between them.

Elements are short, timely, broadly accessible papers that will appeal to both astronomical and statistical communities, highlighting cutting-edge developments for graduate students and researchers. This Elements series offers a unique platform where statistical methods and their potential applications may be demonstrated, for example, in advance of an important astronomical data release, as well as reviews and tutorials on more general topics in Astrostatistics, accepting both invited and unsolicited contributions, both subject to suitable peer review by astronomers and statisticians. The series also encourages best practice in software/code archiving and distribution through appropriate repositories to ensure their long-term access and scientific reproducibility.

Forthcoming topics in the series

(among many others planned over the coming few years):

- Sparsity in Astronomical Data Analysis and Acquisition
- Astronomical Inference via Forward Modelling and Template Libraries
- Bayesian Inference for Astrophysics
- Poisson Statistics in High-Energy Astrophysics

Series Editors

Rafael de Souza, *University of North Carolina at Chapel Hill*

Emille Ishida, *Université d'Auvergne*

Alberto Krone-Martins, *Universidade de Lisboa*

Jianhua Huang, *Texas A&M University*

Alan Heavens, *Imperial College London*

Benjamin Wandelt, *Université de Paris*

CAMBRIDGE UNIVERSITY PRESS

Cambridge Elements

Cambridge Elements are original, concise, authoritative, and peer-reviewed collections of scholarly and scientific research. Organised into focused series edited by leading scholars, they provide comprehensive coverage of key topics in disciplines spanning the arts and sciences.

Flexibility

With lengths of approx. 20,000 to 30,000 words (~40-70pp.), Elements offer an opportunity to publish concise overviews of techniques and their applications not typically found in the traditional book and journal literature.

Rapid Publication and dissemination

Elements are short; authors will be provided with a standard template, and each manuscript will be copy-edited. Elements will be published within 12 weeks of acceptance of the final manuscript.

Indexing

Relevant metadata for individual elements will be sent to the key abstracting and indexing organisations, and Altmetric data for individual Elements will also be available.

Open Access

In order to comply with the requirements of funding agencies, Cambridge will offer Open Access publication options for Elements.

Functionality

The digital-first format allows:

- Enhanced search
- Linked references
- Citation export
- Ability to link to a variety of associated formats – audio, video
- Richer content – social media, debates, resources for teaching, links to data repositories

Content will be hosted on Cambridge Core, which means Elements will benefit from launch marketing as well as increased discoverability. Functionality will be updated on a regular basis.

Want to find out more?

To write an Element for this series or to find out more information about it, contact:
Rafael de Souza (rafael@cosmostatistics-initiative.org)

For more information about the Cambridge Elements publishing model, contact Vince Higgs (vhiggs@cambridge.org)

www.cambridge.org/elements

Sign up for e-alerts in this subject area! Go to cambridge.org/alerts

CAMBRIDGE UNIVERSITY PRESS


THANK
YOU