



## Research Networking Programmes

### Science Meeting – Scientific Report

Scientific report (one single document in WORD or PDF file) should be submitted online within two months of the event. It should not exceed seven A4 pages.

***Proposal Title:*** The 2013 Data Mining and Astrostatistics School  
***Application Reference N°:*** 4725

## 1) Summary (up to one page)

The GREAT Astrostatistics School 2013 took place between June 17<sup>th</sup> and 21<sup>st</sup> in the Pueblo Acantilado facilities in Alicante (Spain). In December 2012, the first announcement of the Astrostatistics School was released. The SOC received 47 applications (as compared to 50 in the 2011 edition of the school) before the deadline, set to February 24<sup>th</sup> 2013. Acceptance letters were sent during March and 35 applicants eventually registered in the school. The selection of the applicants was carried out by the organising committee, based on the relevance of the school topics for the applicants research projects.

### School material

In the following, we will often refer to school materials made available to the students through the School web pages

<http://camd08.ast.cam.ac.uk/Greatwiki/Astrostats2013>

### Software & readiness tests.

In the application form, applicants committed to having a laptop ready for coding with the software recommended by the school organisers. The SOC together with the lecturers decided to use python as the programming language of the school. Applicants had to run a test python code in their computers and copy the result into the application form. Thus, students were aware by that time that python was a requirement and had 3,5 months for getting acquainted with the python language.

Furthermore, during May/June several other software readiness tests were set up in the School wiki pages hosted by GREAT to check for the correctness of the software installations before the school. These are still available in the school wiki under section [Prerequisites/code tests](#). The python packages needed for the school, together with the required versions, were (and still are) listed in the school wiki pages under sections [Prerequisites/Python](#) and [Prerequisites/AstroML](#).

Two spreadsheets were then made available to the students in order to pair students with balanced python programming skills, such that inexperienced students could work through the examples with the aid of a more advanced student. Also, we had a first hand assessment of the degree to which the various packages had been installed by the students.

Finally, a 2 hours pre-school session was set up on Sunday afternoon to help students finalise the complete installation of python packages in their laptops.

### Materials and wiki

All the information relative to the Astrostatistics school was provided through the school web pages in the GREAT wiki with specific information on individual lectures.

This includes information on

- i) the application and registration processes, deadlines, forms
- ii) logistics (Venue, hotels, booking, getting there...)
- iii) scope and committees
- iv) lecturers and the program

The GREAT wiki pages were used to provide lecturers with an interactive space that could be tailored to their respective areas. The goal was to have a detailed program with links to relevant bibliography, python code and datasets. We recommend a visit to the program wiki pages

- [http://great.ast.cam.ac.uk/Greatwiki/GreatItn/AstroStats2013#The\\_Lecturers](http://great.ast.cam.ac.uk/Greatwiki/GreatItn/AstroStats2013#The_Lecturers)
- [http://great.ast.cam.ac.uk/Greatwiki/GreatItn/AstroStats2013#Format\\_and\\_Programme](http://great.ast.cam.ac.uk/Greatwiki/GreatItn/AstroStats2013#Format_and_Programme)

All presentations used in the lectures were made available immediately as links in the school wiki. The objective of the school was (to provide attendants with both, a wide overview of the body of statistical techniques widely applied to astronomical problems, and a specialized primer to the latest developments in this field, occurred in the past decade) was successfully achieved.

As a by-product of special relevance to the Gaia community, we have identified

- i. the areas of astronomy where the analysis of the Gaia data will benefit more from a statistical treatment (via the research interest of the applicants) and
- ii. the population of young scientists in the position to carry out these advances as part of their PhDs

2) Description of the scientific content of and discussions at the event (up to four pages)

The scientific content can be summarised in the following descriptions by the lecturers (with links pointing to the school web hosted at the GREAT wiki):

- **Pieter Degroote** - Introduction to Python as a tool for scientific computing. Contents: basic Python programming, Numerical Python (Numpy), Scientific Python (Scipy), scientific plotting with Python, Monte Carlo simulations with Python. [Further Details](#).
- **Mike Irwin** - Advanced Bayesian statistics. Foundations of Bayesian statistical inference for classification and regression, handling uncertainties, hypothesis testing and model selection, handling of incomplete data, hierarchical models, sampling techniques. [Further Details](#).
- **Coryn Bailer-Jones** - Bayesian parameter estimation and model comparison (MCMC for sampling the posterior PDF; Bayesian evidence as marginal likelihood; K-fold cross validation likelihood; examples of fitting and comparing linear and quadratic models). 2. Bayesian time series analysis and stochastic processes (fitting sinusoidal models; approximation leading to a simple Bayesian periodogram related to the Schuster periodogram; Markov processes; Ornstein-Uhlenbeck process; comparison of the OU process and sinusoidal models on quasar light curves). [Further details](#).
- **Berry Holl** - (What you should know when doing) Statistics with Gaia Data. First the basic properties of the Gaia satellite and how it works are explained. Next, how the data can be modeled based on the expected errors affecting the observations. The self-calibrating forward modeling approach of Gaia is discussed and its implementation for the astrometric and photometric calibration explained. The error properties of the data and the propagation of systematic and random errors into the astrometric catalog are discussed. The last part is about the use of Gaia data: when will it be available, how can one make optimal use of the data, and what pitfalls should one be aware of. . [Further Details](#).
- **Zeljko Ivezić** Statistical analysis of large scale surveys. Dimensionality reduction, classification, regression. [Further Details](#).

	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday
08:50-09:00		Luis Sarro - welcome address				
09:00-10:30		<u>Python:</u> <b>Pieter DeGroot Lecture 1</b>	<u>Bayesian Techniques:</u> <b>Coryn Bailer-Jones Lecture + hands-on 1</b>	<u>Bayesian Techniques:</u> <b>Coryn Bailer-Jones Lecture + hands-on 3</b>	<u>Statistics for Gaia:</u> <b>Berry Holl Lecture 1</b>	<u>Advanced Statistics:</u> <b>Zeljko Ivezic Lecture 1</b>
10:30-11:00		Coffee break	Coffee break*	Coffee break*	Coffee break	Coffee break
11:00-12:30		<u>Python:</u> <b>Pieter DeGroot Lecture 1</b>	<u>Bayesian Techniques:</u> <b>Coryn Bailer-Jones Lecture + hands-on 2</b>	<u>Bayesian Techniques:</u> <b>Coryn Bailer-Jones Lecture + hands-on 4</b>	<u>Statistics for Gaia:</u> <b>Berry Holl Lecture 2</b>	<u>Advanced Statistics:</u> <b>Zeljko Ivezic Lecture 1</b>
12:30-14:00		Lunch	Lunch	Lunch	Lunch	Lunch
14:00-15:30		<u>Statistics:</u> <b>Mike Irwin Lecture 1</b>	<u>Statistics:</u> <b>Mike Irwin Lecture 2</b>	Visit to Alicante	<u>Statistics for Gaia:</u> <b>Berry Holl Hands on 1</b>	<u>Advanced Statistics:</u> <b>Zeljko Ivezic Hands on 1</b>
15:30-16:00		Coffee break	Coffee break	Visit to Alicante	Coffee break	Coffee break
16:00-17:30		<u>Statistics:</u> <b>Mike Irwin Hands on 1</b>	<u>Statistics:</u> <b>Mike Irwin Hands on 2</b>	Visit to Alicante	<u>Statistics for Gaia:</u> <b>Berry Holl Hands on 2</b>	<u>Advanced Statistics:</u> <b>Zeljko Ivezic Hands on 1</b>
18:00-19:30	Registration and help with installation			Visit to Alicante		
19:30-	Welcome barbecue					

### 3) Assessment of the results and impact of the event on the future directions of the field (up to two pages)

We include here a summary of the results of a questionnaire formulated to the school attendants after its celebration.

Page: School Logistics/ Organisation

#### 5. How did you find the organisation of the School?

	answered question				20 5 Rating Count
	Very Good	Good	Reasonable	Poor	
Information provided before the meeting	25.0% (5)	60.0% (12)	15.0% (3)	0.0% (0)	20
Information provided during the meeting	45.0% (9)	55.0% (11)	0.0% (0)	0.0% (0)	20
Post meeting information	30.0% (6)	65.0% (13)	5.0% (1)	0.0% (0)	20
The School wiki pages	50.0% (10)	45.0% (9)	5.0% (1)	0.0% (0)	20
School meeting rooms	63.2% (12)	36.8% (7)	0.0% (0)	0.0% (0)	19
Coffee and Lunch breaks	73.7% (14)	21.1% (4)	5.3% (1)	0.0% (0)	19
Accommodation	73.7% (14)	10.5% (2)	15.8% (3)	0.0% (0)	19
Wednesday Alicante visit	42.9% (6)	28.6% (4)	28.6% (4)	0.0% (0)	14
Welcome B-B-Q	61.1% (11)	33.3% (6)	5.6% (1)	0.0% (0)	18
Any other comments - especially suggestions for improvements. <sup>9</sup>					
<a href="#">Show replies</a>					

Page: School Lectures

#### 6. How well did the lecturer present the material and explain the concepts? The programme is at [http://great.ast.cam.ac.uk/Greatwiki/GreatItN/AstroStats2013#Format\\_and\\_Programme](http://great.ast.cam.ac.uk/Greatwiki/GreatItN/AstroStats2013#Format_and_Programme) (In this and the following, answer 'NA' (Not Applicable) - if you didn't do or complete that particular element of the school. )

	answered question					19 6 Rating Count
	Excellent	Good	Reasonable	Poor	NA	
Python	52.6% (10)	42.1% (8)	0.0% (0)	0.0% (0)	5.3% (1)	19
Statistics	15.8% (3)	42.1% (8)	15.8% (3)	26.3% (5)	0.0% (0)	19
Baysian Techniques	42.1% (8)	42.1% (8)	10.5% (2)	0.0% (0)	5.3% (1)	19
Statistics for Gaia	15.8% (3)	68.4% (13)	10.5% (2)	0.0% (0)	5.3% (1)	19
Advanced Statistics	78.9% (15)	10.5% (2)	5.3% (1)	0.0% (0)	5.3% (1)	19
Any further comments? <sup>5</sup>						
<a href="#">Show replies</a>						

#### 7. How would you rate your learning experience from the lectures?

	answered question					17 8 Rating Count
	Excellent	Good	Reasonable	Poor	NA	
Python	17.6% (3)	70.6% (12)	0.0% (0)	5.9% (1)	5.9% (1)	17
Statistics	17.6% (3)	29.4% (5)	17.6% (3)	35.3% (6)	0.0% (0)	17
Baysian Techniques	29.4% (5)	47.1% (8)	5.9% (1)	17.6% (3)	0.0% (0)	17
Statistics for Gaia	5.9% (1)	47.1% (8)	47.1% (8)	0.0% (0)	0.0% (0)	17
Advanced Statistics	47.1% (8)	35.3% (6)	17.6% (3)	0.0% (0)	0.0% (0)	17
Any further comments? <sup>4</sup>						
<a href="#">Show replies</a>						

#### 8. How do you rate the level (thus too hard, about right, too easy) of the lectures?

	answered question					18 7 Rating Count
	A little too hard, but that was OK.	About the right level	Too Hard	Too Easy	NA	
Python	0.0% (0)	100.0% (18)	0.0% (0)	0.0% (0)	0.0% (0)	18
Statistics	33.3% (6)	22.2% (4)	44.4% (8)	0.0% (0)	0.0% (0)	18
Baysian Techniques	23.5% (4)	52.9% (9)	23.5% (4)	0.0% (0)	0.0% (0)	17
Statistics for Gaia	29.4% (5)	64.7% (11)	5.9% (1)	0.0% (0)	0.0% (0)	17
Advanced Statistics	16.7% (3)	83.3% (15)	0.0% (0)	0.0% (0)	0.0% (0)	18
Any further comments? <sup>3</sup>						
<a href="#">Show replies</a>						

Page: The Practical Exercise Sessions

#### 9. Specifically for the exercise sessions: how useful were the exercises for learning the concepts covered in the lectures? The programme is at [http://great.ast.cam.ac.uk/Greatwiki/GreatItN/AstroStats2013#Format\\_and\\_Programme](http://great.ast.cam.ac.uk/Greatwiki/GreatItN/AstroStats2013#Format_and_Programme) with setup material at <http://great.ast.cam.ac.uk/Greatwiki/GreatItN/AstroStats2013#Prerequisites>

	answered question					18 7 Rating Count
	Very Useful	Useful	Reasonable	Poor	NA	
Python	22.2% (4)	55.6% (10)	11.1% (2)	0.0% (0)	11.1% (2)	18
Statistics	33.3% (6)	11.1% (2)	22.2% (4)	33.3% (6)	0.0% (0)	18
Baysian Techniques	38.9% (7)	44.4% (8)	11.1% (2)	5.6% (1)	0.0% (0)	18
Statistics for Gaia	5.6% (1)	50.0% (9)	38.9% (7)	5.6% (1)	0.0% (0)	18
Advanced Statistics	33.3% (6)	33.3% (6)	22.2% (4)	0.0% (0)	11.1% (2)	18
Any other comments? <sup>2</sup>						
<a href="#">Show replies</a>						

#### 10. Specifically for the exercise sessions: how was the sufficient technical and scientific support for the exercises?

	answered question					18 7 Rating Count
	Very Useful	Useful	Reasonable	Poor	NA	
Python	44.4% (8)	38.9% (7)	0.0% (0)	0.0% (0)	16.7% (3)	18
Statistics	27.8% (5)	33.3% (6)	16.7% (3)	22.2% (4)	0.0% (0)	18
Baysian Techniques	55.6% (10)	44.4% (8)	0.0% (0)	0.0% (0)	0.0% (0)	18
Statistics for Gaia	38.9% (7)	38.9% (7)	22.2% (4)	0.0% (0)	0.0% (0)	18
Advanced Statistics	50.0% (9)	33.3% (6)	5.6% (1)	5.6% (1)	5.6% (1)	18

**10. Specifically for the exercise sessions: how was the sufficient technical and scientific support for the exercises?**

Any other comments?<sup>2</sup>  
[Show replies](#)

**11. Specifically for the exercise sessions, do you agree that 'the exercise sessions were set at an appropriate level of complexity'?**

	answered question					18	
	skipped question					7	
	Strongly Agree	Agree	Neither Agree or Disagree	Disagree	NA		Rating Count
Python	33.3% (6)	38.9% (7)	11.1% (2)	0.0% (0)	16.7% (3)	18	
Statistics	11.1% (2)	27.8% (5)	27.8% (5)	27.8% (5)	5.6% (1)	18	
Baysian Techniques	11.1% (2)	55.6% (10)	33.3% (6)	0.0% (0)	0.0% (0)	18	
Statistics for Gaia	0.0% (0)	66.7% (12)	27.8% (5)	5.6% (1)	0.0% (0)	18	
Advanced Statistics	16.7% (3)	38.9% (7)	33.3% (6)	0.0% (0)	11.1% (2)	18	

Any other comments?<sup>4</sup>  
[Show replies](#)

**12. Specifically for the exercise sessions - concerning setting up the software.**

	answered question					18	
	skipped question					7	
	Before the start of the School	1st day (Monday)	2nd day (Tuesday)	By the end of the School	I was unable to get a running installation		Rating Count
Did you successfully install the requested software and packages?	66.7% (12)	22.2% (4)	0.0% (0)	11.1% (2)	0.0% (0)	18	

Any other comments?<sup>4</sup>  
[Show replies](#)

Page: Summary Comments on the School

**13. Please rate the usefulness to you as to the School as a whole. The programme is at [http://great.ast.cam.ac.uk/Greatwiki/GreatItn/AstroStats2013#Format\\_and\\_Programme](http://great.ast.cam.ac.uk/Greatwiki/GreatItn/AstroStats2013#Format_and_Programme)**

	answered question				18	
	skipped question				7	
	Very Useful	Useful	Ok	Poor		Rating Count
The school was ...	66.7% (12)	27.8% (5)	5.6% (1)	0.0% (0)	18	
Being shown new statistical analysis techniques was ...	82.4% (14)	11.8% (2)	5.9% (1)	0.0% (0)	17	
I found interacting with the other students and tutors was ...	66.7% (12)	27.8% (5)	5.6% (1)	0.0% (0)	18	
In terms of supporting my current research, attendance at the School was ...	52.9% (9)	23.5% (4)	23.5% (4)	0.0% (0)	17	
In terms of supporting my future research, attendance at the School was ...	76.5% (13)	17.6% (3)	5.9% (1)	0.0% (0)	17	

What would you change to improve the session if it were to be repeated?<sup>6</sup>  
[Show replies](#)

The SOC considers that the objectives of the school have been met satisfactorily, specially in the fluent interaction between students, lecturers and the SOC itself. The choice of the location was perfect in this respect, with not too many opportunities for the group to get dispersed. We found problems with the internet connection, which failed that week despite the back up ADSL lines hired by the organisers. We believe we have learned some lessons to improve the practical sessions and the software setup, which inevitably involve a very early involvement of the lecturers in the preparation.

Finally, the SOC recognises the lack of gender balance in the final selection of lecturers and students. The proportions amongst applicants and students are very similar, and thus the selection process did not introduce any unwanted bias. Still, strategies to encourage applications from female students would need to be applied in future editions of this school, or apparently, any other in the same field.

**Annex 4a: Programme of the meeting**The school comprised 5 main blocks, each with a total of 4 lectures of 90 minutes each (except the introductory course on the python programming language that only took two 90 minutes lectures). The areas and lecturers are listed below:

1. Pieter DeGroote (KU Leuven, Belgium).  
Contents:
- 

***Getting and understanding a working Python installation***

1. Why use Python?
2. Who is using Python?
3. Python installation
4. Understanding packages

***First steps with Python***

1. Running Python code
2. Python built-in types and operations
3. Assignment operator
4. Control flow statements
5. Writing functions
6. Modules
7. Standard library
8. A quick note about Python 3

***Reading and writing files***

1. ASCII data
2. Binary data

***Plotting and images***

1. Matplotlib
2. Publication quality plots
3. Advanced Interactive plots
4. 3D plotting with Mayavi

***Scientific programming with Python***

1. Working with arrays
2. Basic statistics
3. Probability distributions
4. Examples
5. Optimization

## 2. Mike Irwin (Cambridge University, UK). Contents

---

### Lecture 1A

- some illustrative "simple" problems
- foundations of probability theory and statistical techniques;
  - some issues when combining variables;
  - Central Limit theorem to the rescue;
  - noise, covariance, multivariate Gaussians and Rayleigh distribution;
  - aliasing and Shannon sampling
- introduction to Bayes' theorem
  - as a classifier illustrated with DNA example
  - as a way to use prior information to estimate future odds
- Bayesian view of model selection (e.g. Gregory 2005)

### Lecture 1B

- Bayesian application to source identification likelihoods (e.g. Prestage & Peacock 1983)
- introducing Maximum Likelihood estimators (e.g. Woodward & Davies 1958)
  - using optimal detection of signals with cross-correlation as a limiting case as one example
  - and estimating velocity dispersions in clusters/galaxies as another
- Maximum Likelihood estimators in general and some numerical considerations

### Exercises 1

1. comparison of cross-correlation and likelihood methods on a real CaT region spectrum using idealised templates and model atmosphere spectra illustrating practicalities and how to estimate radial velocities with realistic errors, the effect of correlated sample noise due to e.g. rebinning the spectra
2. given a set of radial velocities with errors estimate the systemic velocity and line-of-sight velocity dispersion of a dwarf galaxy and hence estimate its mass-to-light ratio, what is the impact of using priors ?
3. some of the "simple" problems can also be used in the examples

### Lecture 2A

- Maximum Likelihood estimators and Fisher information (Fisher 1925)
  - confidence intervals and morphing the parameter space
  - predicting parameter errors and fine tuning the parameter model
  - realistic error models and k-sigma clipping
- examples of estimators
  - optimal spectral extraction (Hewett et al. 2005)
  - low count rates and Lynden-Bells C-statistic;

### Lecture 2B

- further examples of Maximum Likelihood estimators
  - sparse samples, null results and the Press-Schechter method (Schechter & Press 1976)
  - optimal astrometry in a single pass (Irwin 1985)
  - structural analysis, or a brief discourse on curve fitting with errors on both axes (e.g. Hodgkin et al. 2009)
- classical hypothesis testing
- multivariate analysis using PCA as an example of dimensional reduction and noise suppression

### Exercises 2

1. given the distances and coordinates of a population of Galactic Halo dwarf galaxy satellites, what is the best fit power law model for this distribution, what is the effect of selection bias ?
2. given a list of Lyman limit detections, null results etc.. for a sample of QSOs, find the best fit simple power law model of the form  $N(z) = N_0(1+z)^\gamma$  and estimate the error on the predicted number of Lyman limit systems at redshift  $z=3$
3. PCA example using model atmosphere spectra and for noise suppression on real CaT data



### 3. Coryn Bailer-Jones (MPIA, Germany) Contents

---

#### Lecture 1A: Bayesian parameter estimation

- recap of principles (following Mike's lecture) using a simple one parameter problem: coin tossing example
  - estimating the probability,  $p$ , of heads, given a sequence of coin toss results
  - Binomial likelihood, uniform or beta prior
- why we need efficient sampling methods
- fitting a straight line to  $(x,y)$  data with unknown noise
  - use of MCMC (Metropolis algorithm with Gaussian proposal distribution) for sampling from the posterior

#### Lecture 1B: Bayesian model comparison

- fitting quadratic model to data with unknown noise
- Monte Carlo integration
- principle of model comparison using the evidence
  - evidence estimated as marginal likelihood by sampling the prior
  - why you may not use maximum likelihood (e.g. minimum chisq) for model comparison
- problems with the evidence, and alternatives
  - k-fold CV likelihood
- model comparison of straight line vs. quadratic using evidence and k-fold CV likelihood

#### Lecture 2A: Bayesian time series analysis

- time series analysis as model fitting in the time domain
- fitting a sinusoidal model (after Bretthost)
  - fitting a model with a single sinusoidal component
  - approximations
  - cases of given and unknown noise
  - the Bayesian periodogram and its relation to the Schuster periodogram
- parameter estimation (posterior sampling using MCMC) with the sinusoidal model
- overview of a general approach for arbitrary times series models and measurement models

#### Lecture 2B: Stochastic time series models

- stochastic processes
  - Markov models
  - Langevin equation
- The Ornstein-Uhlenbeck process
  - simulating and calculating Markov models
  - likelihood of the OU process
  - parameter estimation (posterior sampling using MCMC)
- application to quasar light curves
  - parameter estimation
  - model comparison with the evidence k-fold CV likelihood

## 4. Berry Holl (INTEGRAL Science Data Centre, Switzerland) Contents

---

### **LECTURE 1 - 4**

#### **Introduction to Gaia**

- Scientific goals
- Scanning of the sky
- The observations process
- The scientific instruments
- (Expected) scientific performances

#### **Modeling of the data**

- Forward modelling
- Self calibration
- The observations process
- Instrument calibration
- From observations to catalogues
- Error characterization
- Astrometric data modeling

#### **Using the data**

- What you get
- Optimal usage (& going beyond the standard catalogues)
- Pitfalls with examples (non-uniform priors, non-linear transformations, correlations, etc.)

#### **Exercise 1**

##### **Simple improved parallaxes in clusters**

In this exercise we will statistically explore the answer to the simple question: Given that we have the measured parallaxes of stars in a cluster, can we improve the accuracy of individual parallaxes by combining the individual and mean cluster parallax?

#### **Exercise 2**

##### **Bayesian luminosity calibration**

For this exercise we will interpret the data from a fictitious parallax survey of all stars of a particular luminosity class, which are distributed uniformly throughout a certain volume around the Sun. In addition we know that the luminosities of the stars have a normal distribution. In this exercise we will use Bayesian modeling to answer the seemingly simple question: what is the mean absolute magnitude of this class of stars and what is the variance around the mean?

The exercise will be given together with Anthony Brown who developed this exercise as part of his chapter 16 in [van Altena \(2013\)](#). The Python code and instruction can be [downloaded from here](#).

## 5. Zeljko Ivezic (Washington University, USA) Contents

---

### **LECTURE 1: Intro and warm-up with astroML**

- making a Hess diagram and coding by a third quantity:
- how to estimate parameters for a 2D covariant gaussian when outliers are present?
- how to apparently beat CLT? Bayesian approach applied to uniform distribution:
- how to account for truncation bias: Cminus method:
- matched filter analysis of a low-SNR signal
- wavelet analysis of a signal with varying power spectrum:
- if time allows: discussion homework for the coffee break: the Monty Hall problem

### **LECTURE 2: Density estimation**

#### **1D introduction:**

- Knuth histograms (fixed bin width, Bayesian optimization)
- Bayesian Blocks algorithm (variable bin width)
- Gaussian Mixture models (Extreme Deconvolution in 1D)
- kernel density estimates (KDE) in 1D
- smoothing data with a Wiener filter and connection to KDE
- Bayesian nearest neighbor method

#### **Density estimation in high-D cases**

- high-D KDE
- Bayesian nearest neighbor method
- Extreme Deconvolution in high-D

### **LECTURE 3: Clustering and Classification**

- 1D hypothesis testing and classification
  - unsupervised vs. supervised classification
- potpourri of classification methods (on the same dataset):
- the ROC curves for comparing classification methods
- grand finale:

### **LECTURE 4: Dimensionality Reduction**

- PCA and friends: ICA, NMF, LLE
- fast positional matching with kd trees
- regression example

## Annex 4b: Full list of speakers and participants

### Lecturers:

Pieter Degroote from KU Leuven University (Belgium)  
Mike Irwin from the Institute of Astronomy in Cambridge (UK)  
Coryn Bailer-Jones, from the Max Planck Institut für Astronomie in Heidelberg (Germany)  
Berry Holl, from the University of Geneva (Switzerland)  
Zeljko Ivezić from the University of Washington, (USA)

### Students:

Ms. HODA ABEDI BARCELONA, (ES) Institut d'Estudis Espacials de Catalunya (IEEC), Institut de Ciències de Cosmos (ICC), Dept. Astronomia i Meteorologia, Universitat de Barcelona (UB), Martí i Franquès 1, E08028 Barcelona, Spain.

Mr. RICHARD ANDERSON geneva, (CH) **Geneva Observatory**

Ms. NADEJDA BLAGORODNOVA CAMBRIDGE, (UK) Institute of Astronomy

Mr. SERGI BLANCO CUARESMA FLOIRAC, (FR) Laboratoire d'Astrophysique de Bordeaux

Mr. MIKOLA BRITAVSKIY ATHENS, (GR) Institute of Astronomy, Astrophysics, Space Applications and Remote Sensing

Mr. TRISTAN CANTAT-GAUDIN Padova, (IT) Osservatorio Astronomico di Padova. Vicolo Osservatorio 5 - 35122 - PADOVA. Cod. Fis. 97220210583 - P.IVA 06895721006

Mr. RICARDO DORDA SAN VICENTE DEL RASPEIG, (ES) UNIVERSIDAD DE ALICANTE

Mr. FABO FENG HEIDELBERG, (DE) Max-Planck-Institute for Astronomy

Mr. DIEGO FUSTES VILLADONIGA A CORUÑA, (ES) University Of Coruña

Mr. GUILLAUME GUIGLION NICE, (FR) Laboratoire Lagrange, UMR 7293

Mr. LEO HUCKVALE GARCHING BEI MUENCHEN, (DE) European Southern Observatory

Mr. ZDENĚK JANÁK BRNO, (CZ) Department of Theoretical Physics and Astrophysics

Mr. JENS-KRISTIAN KROGAGER SANTIAGO, (CL) ESO Santiago

Mr. JORGE LILLO-BOX VILLANUEVA DE LA CAÑADA, (ES) Astrobiology Center (INTA-CSIC)

Miss CHENG LIU LUND, (SE) Lund Observatory

Mr. BEIBEI LIU BEIJING, (CN) Peking University, Astronomy and Astrophysics Department,

Mr. PABLO MARCOS ARENAL HEVERLEE, (BE) Instituut voor Sterrenkunde, KU Leuven

Mr. ANDRE MARTINS BESANCON, (FR) Besançon Observatoire

Mr. JABULANI PAUL MASWANGANYE KRUGERSDORP, (ZA) Hartebeesthoek Radio Astronomy Observatory (HartRAO)

Mr. ABBAS MOE HEIDELBERG, (DE) Astronomisches Rechen-Institut (ARI)

Mr. MATTHEW MOLLOY BEIJING, (CN) Kavli Institute for Astronomy & Astrophysics

Mr. THOMAS NORDLANDER UPPSALA, (SE) Uppsala University

Mr. LOVRO PALAVERSA SAUVERNY, (CH) Geneva Observatory, 51 chemin des Maillettes, 1290 Sauvery, Switzerland

Mr. MAX PALMER BARCELONA, (ES) University of Barcelona

Mr. VICTOR PEREIRA BLANCO MADRID, (ES) Universidad Complutense de Madrid

Mr. ALVARO RIBAS MADRID, (ES) ESAC/INSA, P.O. Box, 78, 28691 Villanueva de la Cañada, Madrid, Spain

Mr. ALVARO ROJAS NICE, (FR) Observatoire de la Côte d'Azur Boulevard de l'observatoire BP4229, 06304 Nice Cedex 4

Mr. JAN RYBIZKI HEIDELBERG, (DE) Astronomisches Rechen-Institut, Zentrum fuer Astronomie, Universitaet Heidelberg

Ms. LAURA MARIA SAMPEDRO HERNANDEZ GRANADA, (ES) INSTITUTO DE ASTROFÍSICA DE ANDALUCÍA IAA-CSIC

Ms. SARA ALEJANDRA SANS FUENTES LEUVEN, (BE) KU Leuven

Mr. TONI SANTANA POZNAN, (PL) Instytut Obserwatorium Astronomiczne

Ms. IULIA TEODORA SIMION CAMBRIDGE, (UK) University of Cambridge, Institute of Astronomy, Madingley Road, Cambridge, CB3 0HA

Mr. JOHN VICKERS HEIDELBERG, (DE) ARI

Ms. DANIEL VILJOEN POTCHEFSTROOM, (ZA) North-West University