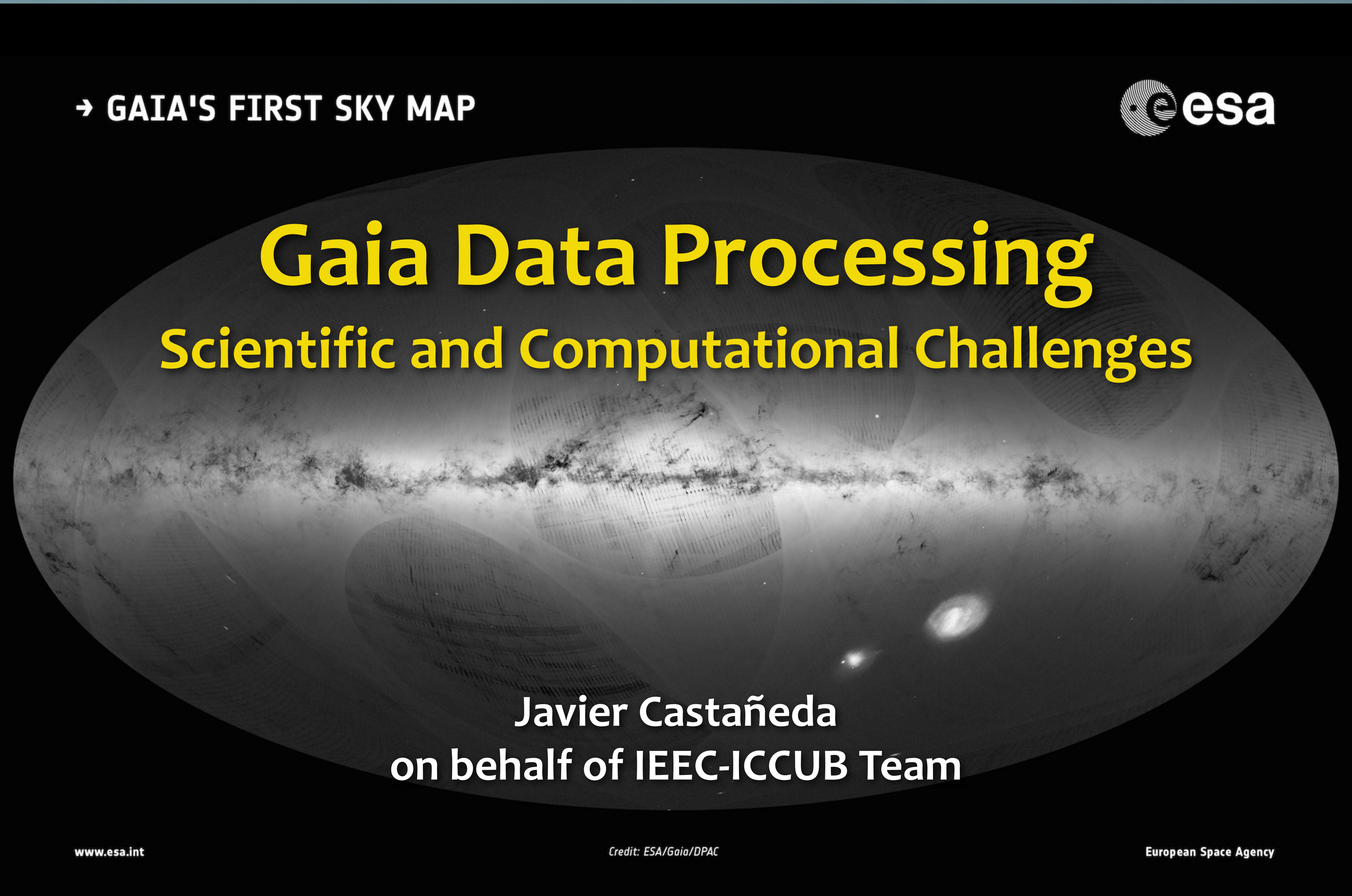
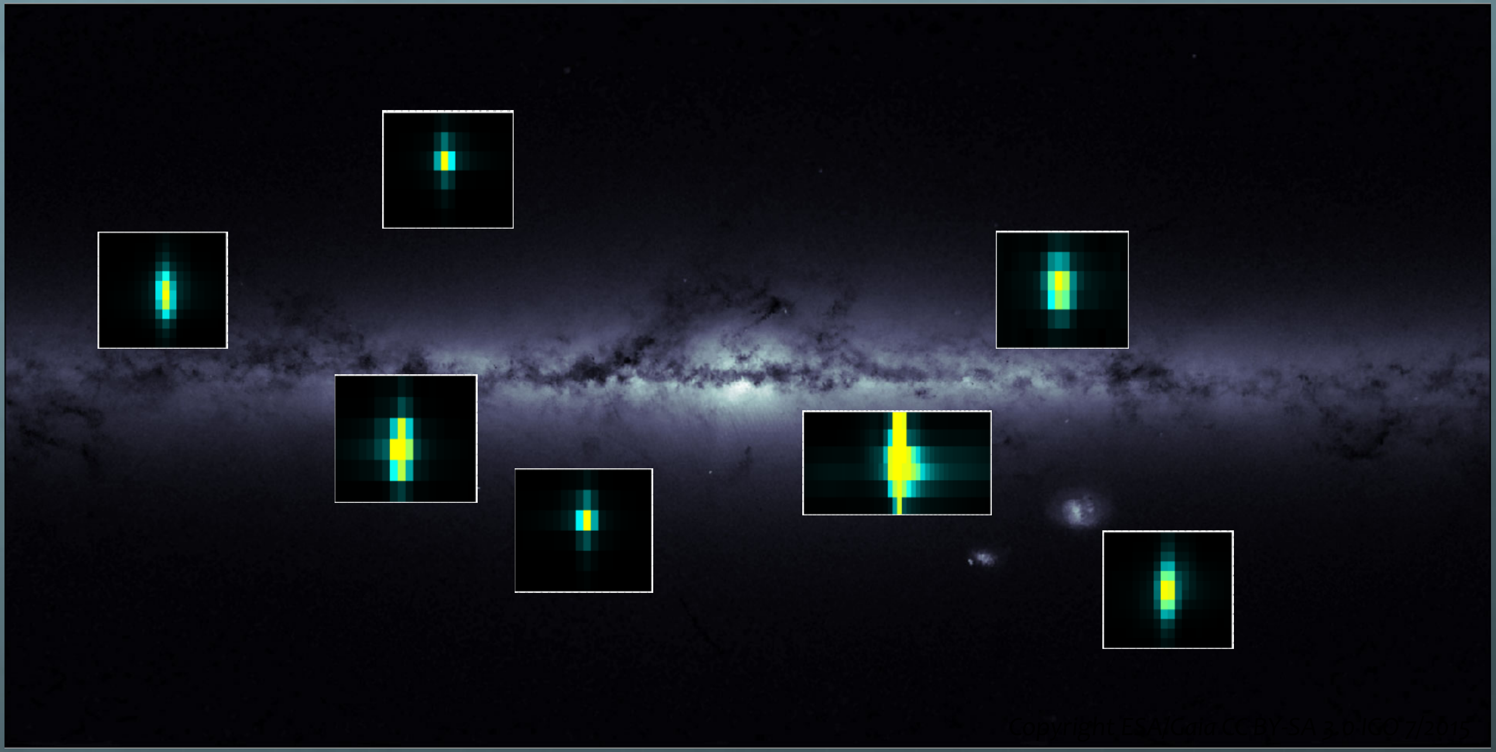
Supercomputación y eCiencia

Grand Challenges in Astrophysics and Cosmology

Gaia Data Processing
Scientific and Computational Challenges



Javier Castañeda
on behalf of IEEC-ICCUB Team



The main challenge of the Gaia mission is the translation of the thousand billion tiny images gathered by the approximately hundred CCDs on-board Gaia into the positional and physical parameters of more than one billion sources to the scientific community

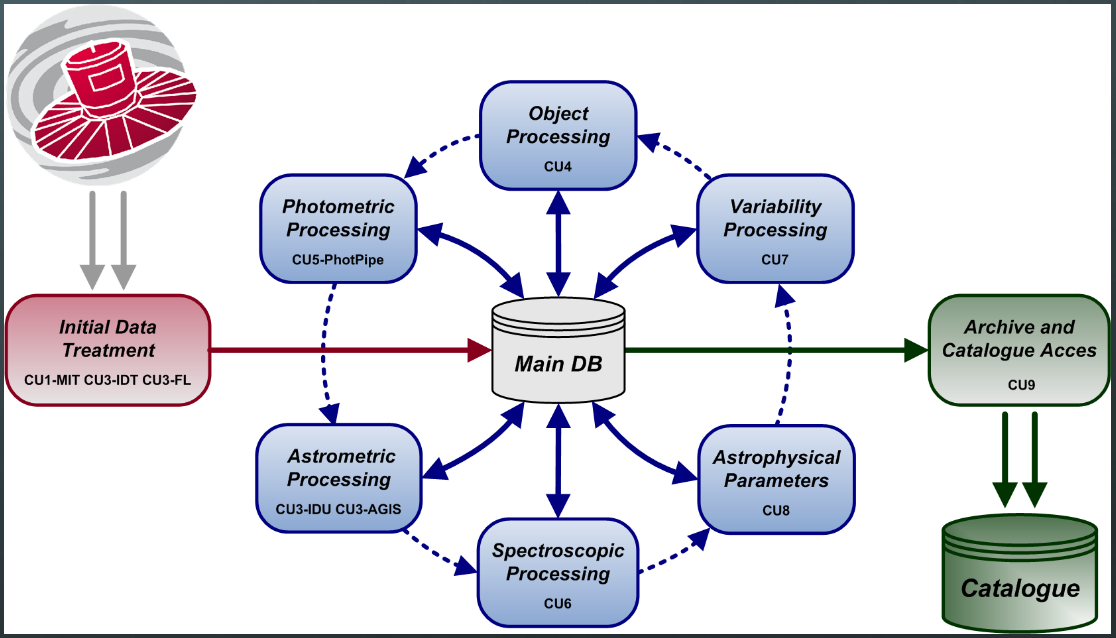
Our Responsibilities and Team

- We participate in three different areas:
  - Design of the Gaia Daily data processing
  - Design and Operation of Gaia Cyclic data processing
  - Coordination, Design and Development of the Gaia Data Archive

- Large, experienced and multidisciplinary team:

X.Luri
J.Torra
C.Fabricius
J.Portell
J.Castañeda
M.Clotet
F.Julbe

N.Garralda
J.J.González
S.Soria
E.Masana
F.Torra
R.Borrachero
and BSC Team



Daily Data Processing Challenges

- Real time processing of the incoming data from the satellite
  - Light data processing of a small data volume - processing 40 million objects (40 GB) a day
  - Essential to monitor spacecraft health and correct operation
- Running in a dedicated cluster at ESAC
  - Running Java code against an InterSystems Cache DB
- Event driven processing approach:
  - Triggered by the incoming data types and predefined data processing flow dependencies.
  - Flexible system able to adapt the processing flow

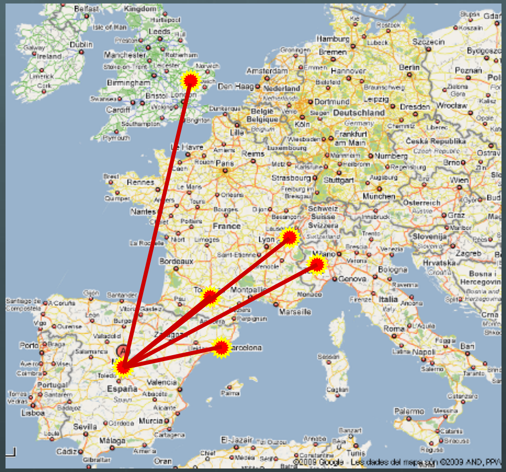
Running continuously for the last 1000 days,

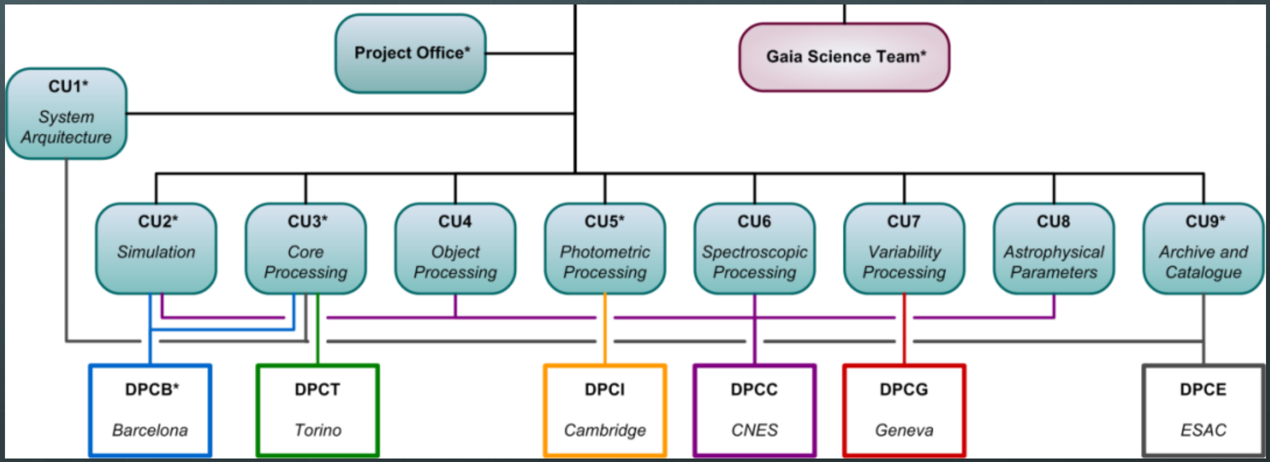
up to total 70 billion objects observed and 200 TB generated so far ☺

See https://www.cosmos.esa.int/web/gaia/mission-numbers

Cyclic Data Processing Challenges

- Distributed among six data processing centers
- Three main difficulties to solve:
  - Data distribution: transfers
  - Data handling: storage and access
  - Iterative data scientific reduction: scientific dependencies





Data Distribution Challenge

- Probably quite specific to Gaia, adopting a partitioned and distributed processing tasks in different processing centers
- Daily products transferred on daily basis, from 20 to 80 GB
- Cyclic products produced periodically at least once a year, up to few TBs
- Data transfers through Aspera, initial rates of 300Mpbs, now incremented to 0.5-1Gpbs. From latest activities, we clearly need to optimize:
  - Cyclic data transfers: reducing data size and file granularity and/or incrementing transfer rates
  - Improve automatic data consistency checks and accountability
- The data flow and dependencies among the different systems imply additional complications
  - All systems wants the latest results from the rest, which is of course not feasible

Data Handling Challenge

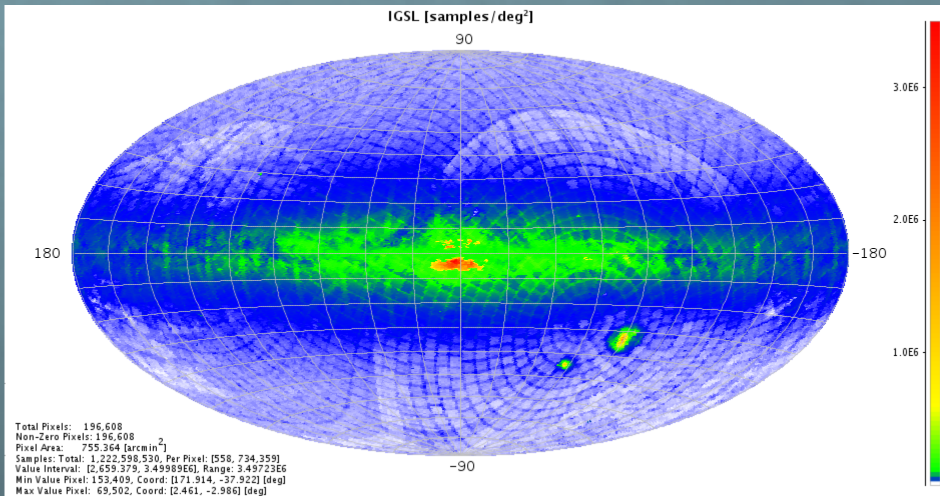
- Large set of data types, some tables reaching huge number of records with relatively large data volume, currently:
  - Working catalogue up to 3 billion records in approximately 1TB
  - Observations up to 70 billion records in approximately 25TB

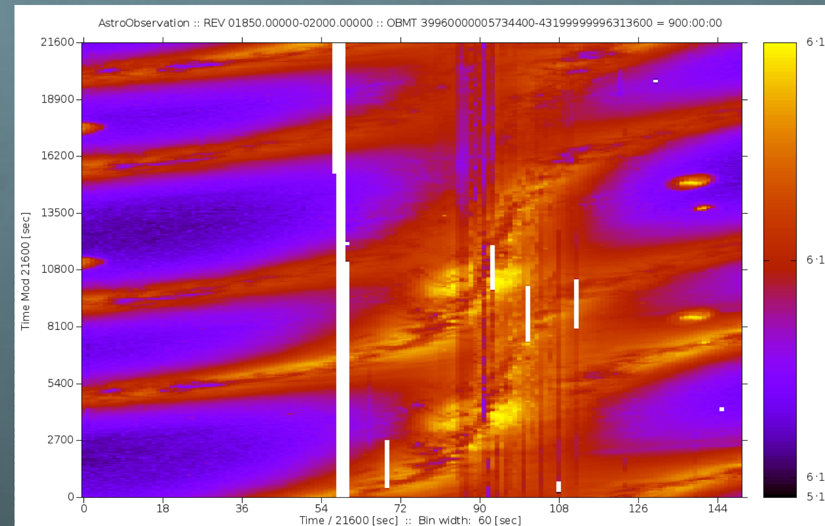
- At BSC, main storage based on files but using DB for the metadata
  - Other DPCs started with only DB but gradually moving to file-based solutions

- Dual data access requirements: time-based and spatial-based
  - Use of the HealPix sky tessellation scheme
  - Tailored file format based on HDF5 to integrate both access types

- Store efficiency and performance boost by proper data equalization
  - Tools to retrieve time and spatial data density profiles
  - Tools for data and job equalization

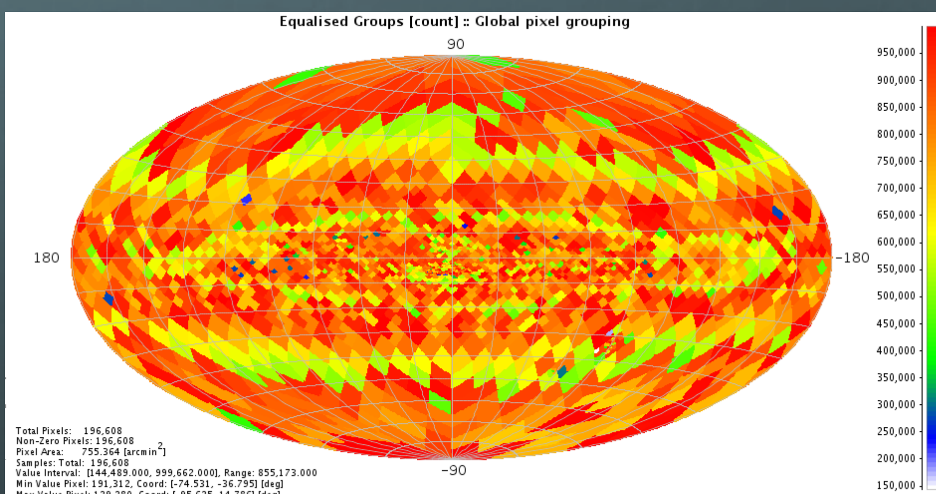
- Adopt of a human-readable file system structure easy to be exported outside the cluster

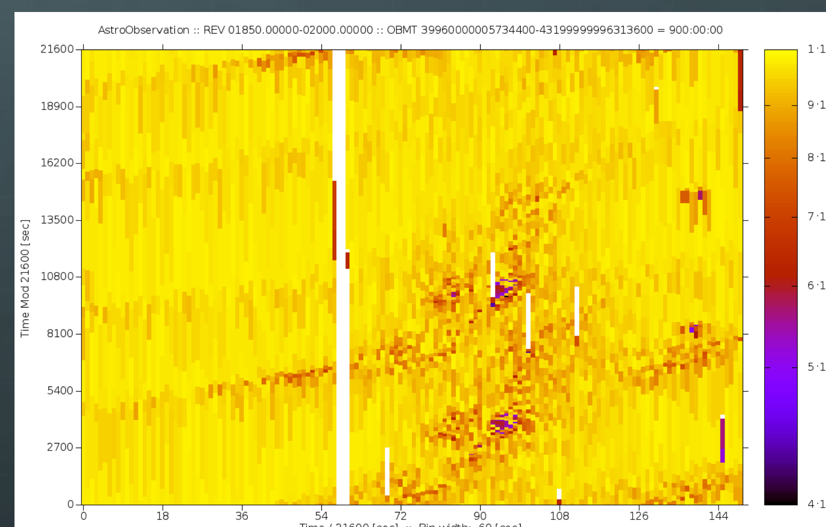




Spatial-based (HEALPix)

Time-based





Cyclic Data Processing Challenge

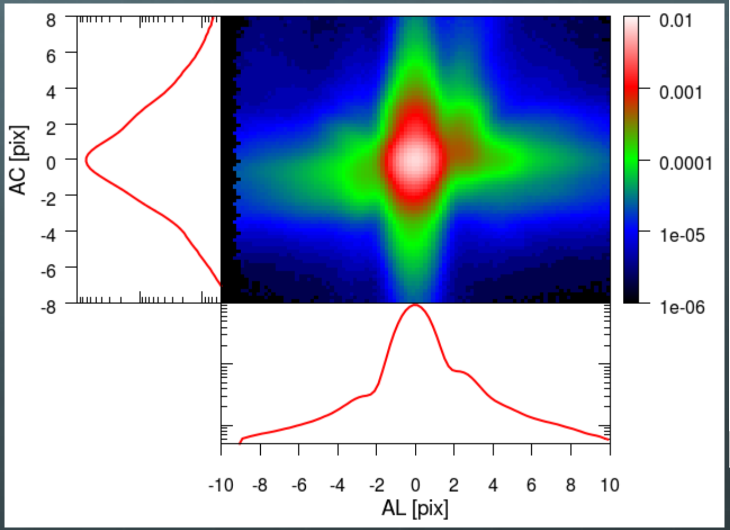
- Large dataset, but in record counts and data types rather than data volume
- Complex calibration system to account for the spacecraft events, instrument response and the data features
- Heavy computational load for the image data processing
  - Calibration of the instrument optical and electronic response and the computation of the image parameters.
- Applying heterogeneous algorithmic techniques:
  - Clustering, classification, equation resolution (LeastSqueare, HouseHolder, Cholesky Decomposition, MaximumLikelihood, etc) among others

Cyclic Data Processing Challenge

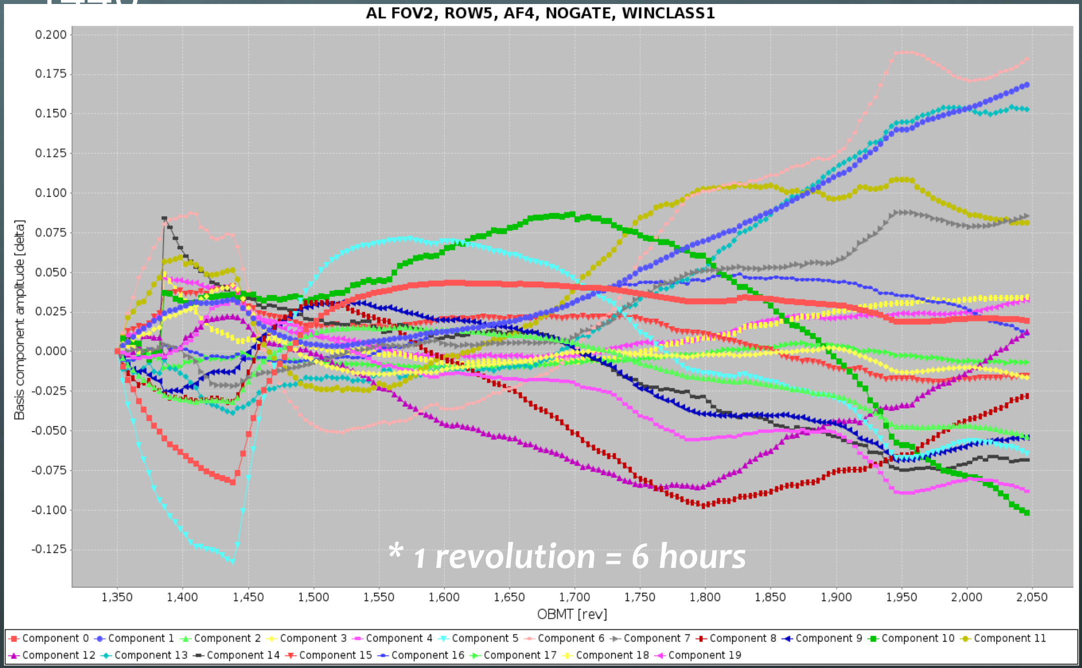
- Our main activities are focused in:
  - Development of a versatile and tailored execution framework and DAL for the integration of all processing tasks at BSC
  - Design and development of some instrument calibration tasks
  - Image processing of all Gaia astrometric measurements
  - Design and development of the Gaia detection classification and cross-match with the working source catalogue.

Instrument Response Profile Calibration

Instrument response profile using a 2D numerical mapping.
Parametrized by observation time, FoV, CCD, magnitude, colour, window geometry, etc.



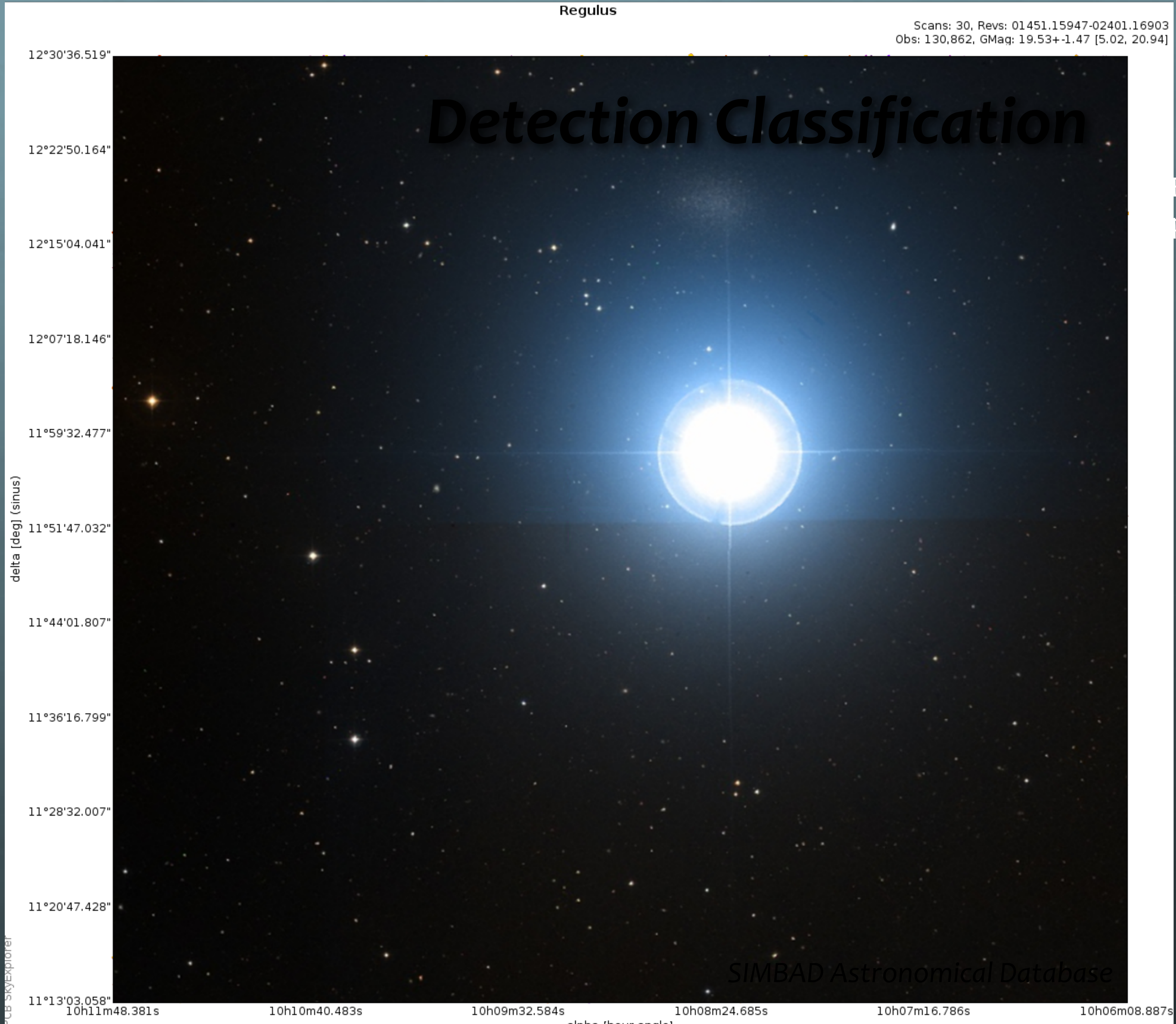
Basis coefficient Evolution during ~6 months of data including refocus operation at revolution* 1440



Credit ESA/Gaia/IfA-ROE

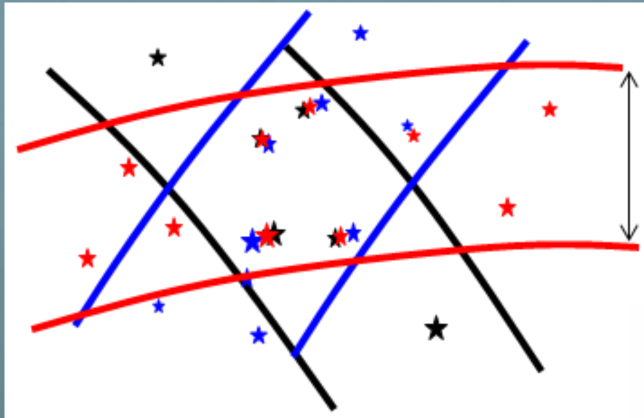
Detection Classifier

Detection Classification

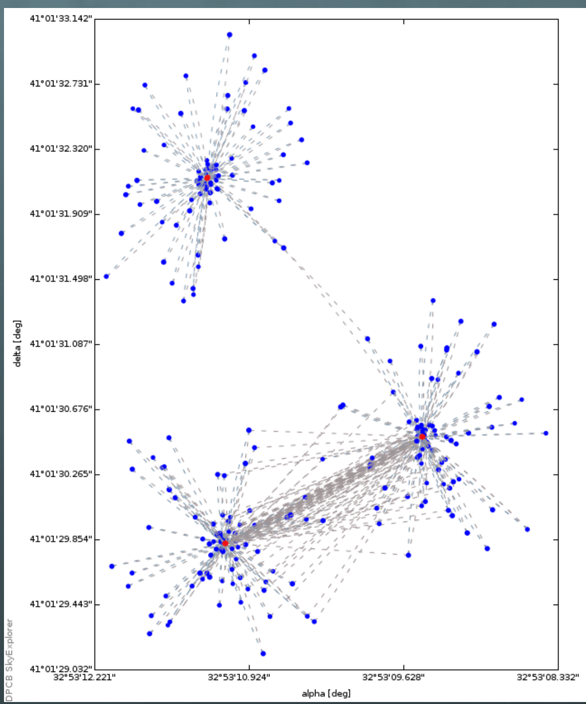


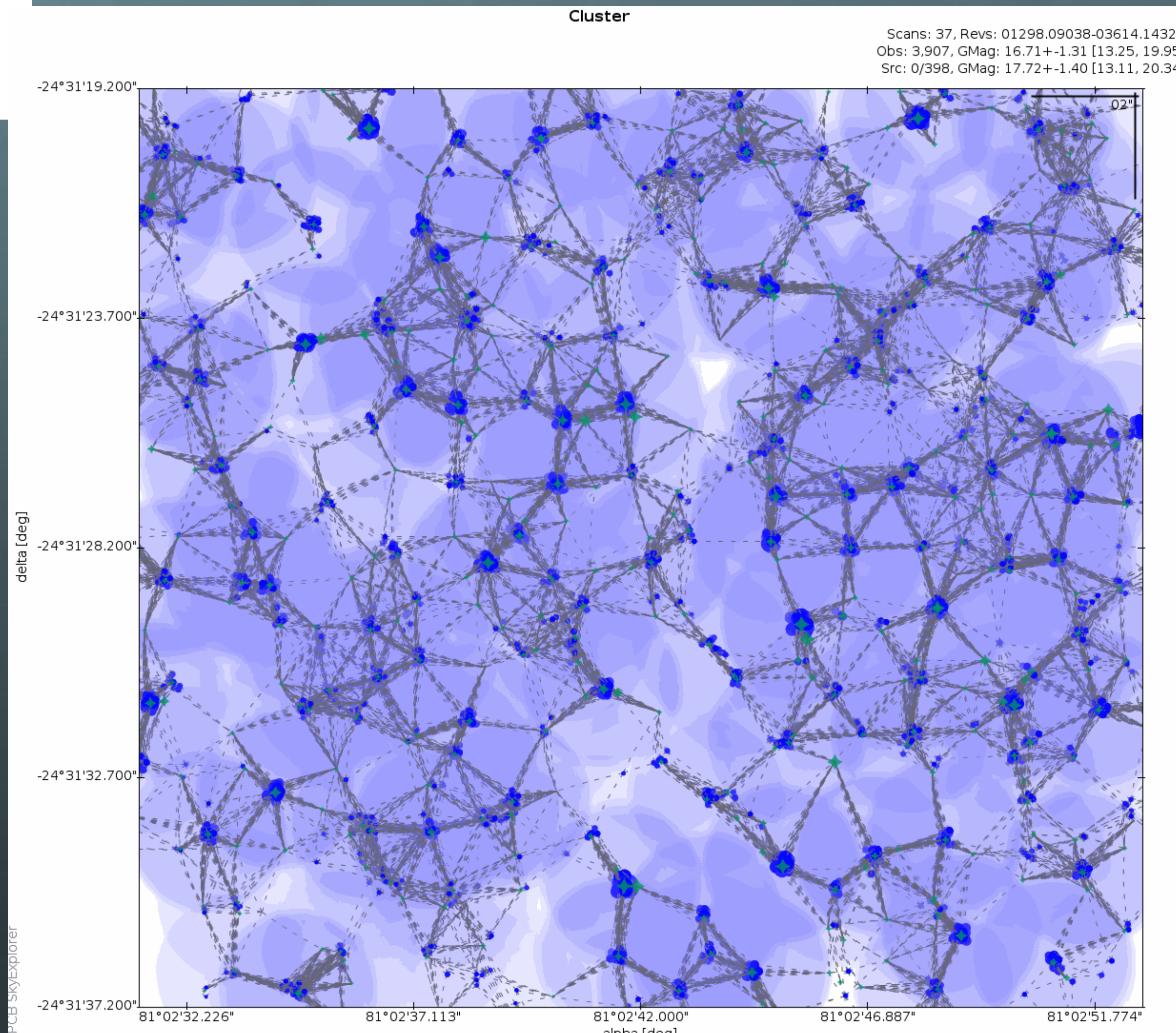
EO I
ph Gal

EGULUS

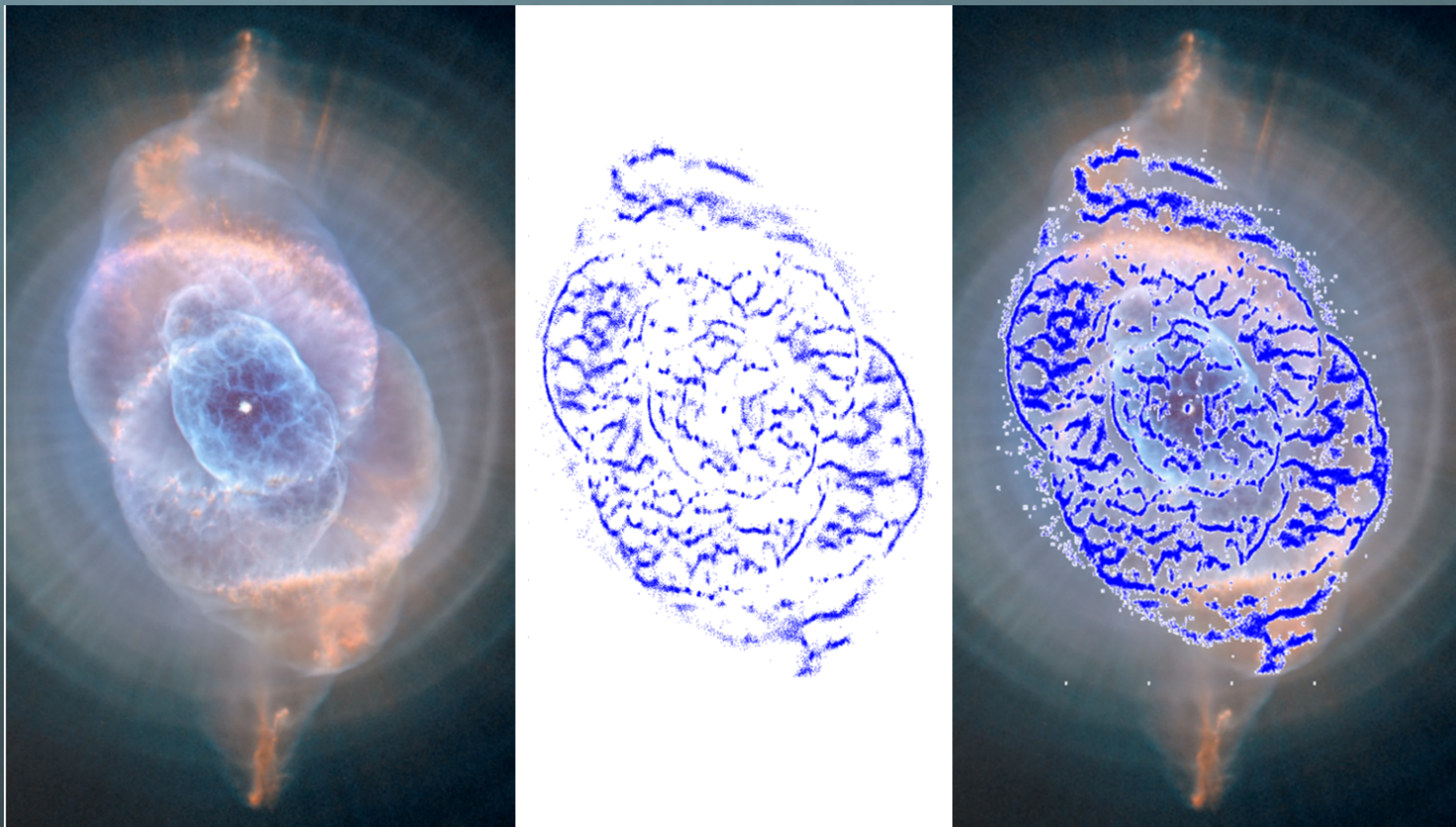


Obs-Src Cross-Match





Extended Objects Detection Examples

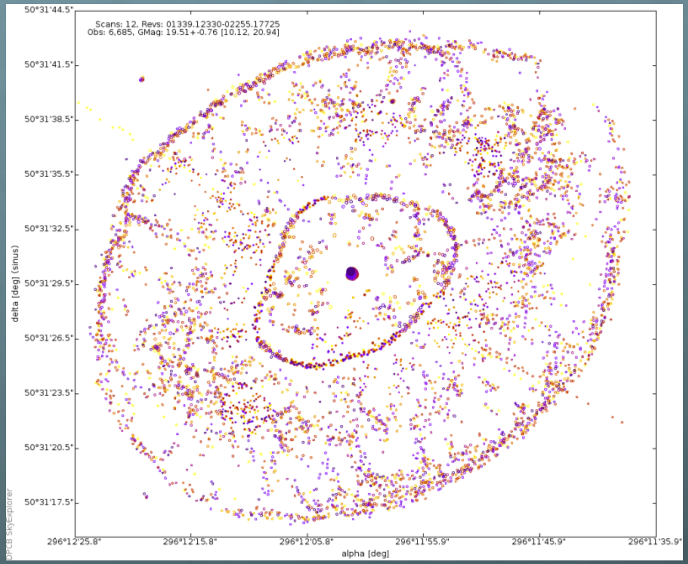


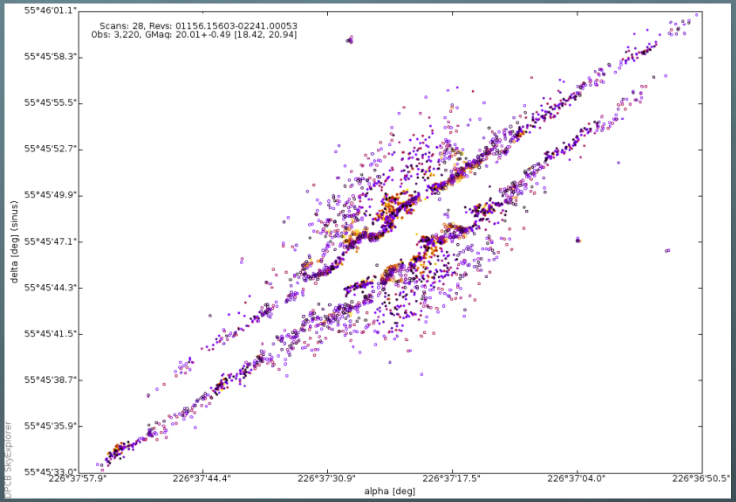
Cat's Eye nebula identified during Operations Rehearsal campaign 2014

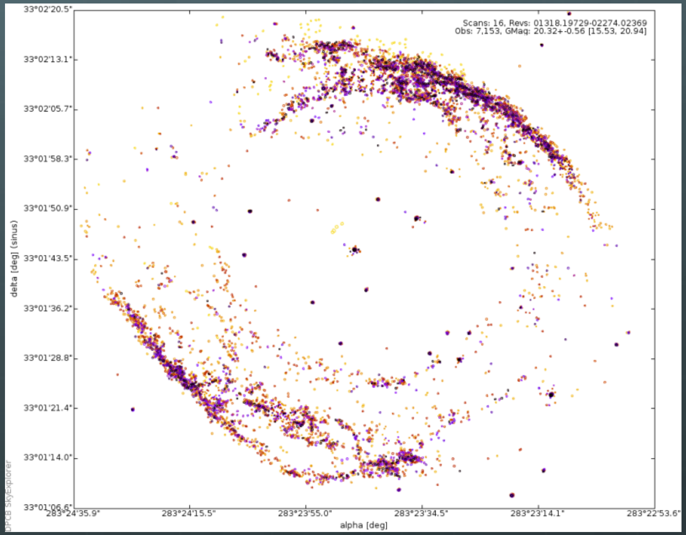
Left: HST image of the nebula taken with the ACS/WFC

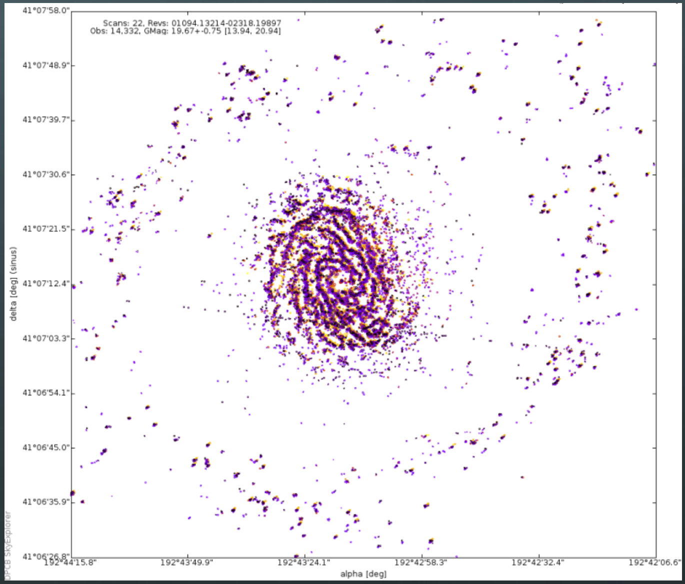
Middle: ~84,000 Gaia detections made from 25 July to 21 August 2014

Extended Objects Detection Examples





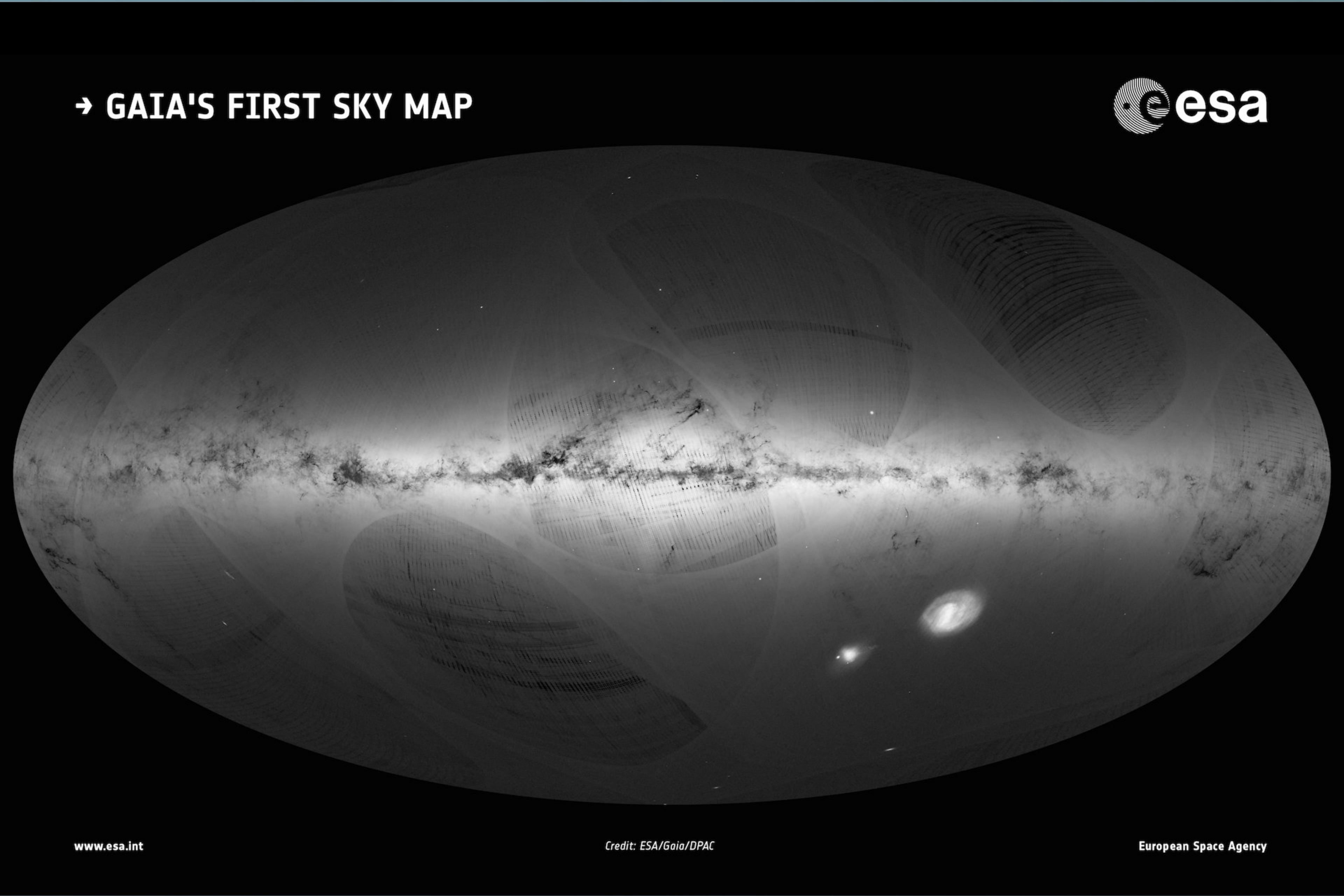




Gaia Data Archive Challenge

- Provision of Gaia data to the community but:
  - Validating the correctness of all its fields
  - Offering the proper infrastructure and middleware upon which scientists will be able to do exploration and modeling with this huge data set

- Our team is currently leading the design of this common processing framework for data exploitation:
  - Usual suspects: Hadoop, Spark, etc.


→ GAIA'S FIRST SKY MAP
esa
www.esa.int
Credit: ESA/Gaia/DPAC
European Space Agency

About A&A
Board of Directors
Author information
Submission process
How to subscribe
Prices
e-Access license
e-Access registration

Reader services
EDPS account
Latest articles FREE
Email-alert
RSS feed
Recommend this journal
Access by vol/page
DOI resolver
Useful links

Events

Scientific Writing for Young Astronomers

NEWS

GREGOR Special Issue published by A&A (November 2016)

A&A publishes Gaia Special Issue (November 2016)

Astronomy & Astrophysics

Gaia Data Release 1

Export the citation of the selected articles Export

Select all

Press Release Free Access

The *Gaia* mission

Gaia Collaboration, T. Prusti, J. H. J. de Bruijne, A. G. A. Brown, A. Vallenari, C. Babusiaux, C. A. L. Bailer-Jones, U. Bastian, M. Biermann, D. W. Evans et al. (616 more)

Published online: 24 November 2016

DOI: https://doi.org/10.1051/0004-6361/201629272

Full HTML | PDF (4.884 MB) | ePUB (2.089 MB) | References | NASA ADS Abstract Service

Press Release Free Access

Gaia Data Release 1 - Summary of the astrometric, photometric, and survey properties

Gaia Collaboration, A. G. A. Brown, A. Vallenari, T. Prusti, J. H.J. de Bruijne, F. Mignard, R. Drimmel, C. Babusiaux, C. A.L. Bailer-Jones, U. Bastian et al. (582 more)

Published online: 24 November 2016

DOI: https://doi.org/10.1051/0004-6361/201629512

Full HTML | PDF (9.221 MB) | ePUB (2.672 MB) | References | NASA ADS Abstract Service

Gaia Data Release 1 - Pre-processing and source list creation

C. Fabricius, U. Bastian, J. Portell, J. Castañeda, M. Davidson, N. C. Hambly, M. Clotet, M. Biermann, A. Mora, D. Busonero et al. (97 more)

Published online: 24 November 2016

DOI: https://doi.org/10.1051/0004-6361/201628643

Full HTML | PDF (4.781 MB) | ePUB (3.037 MB) | References | NASA ADS Abstract Service

Open Access

Gaia Data Release 1 - Astrometry: one billion positions, two million proper motions and parallaxes

L. Lindegren, U. Lammers, U. Bastian, J. Hernández, S. Klioner, D. Hobbs, A. Bombrun, D. Michalik, M. Ramos-Lerate, A. Butkevich et al. (73 more)

Published online: 24 November 2016

DOI: https://doi.org/10.1051/0004-6361/201628714

arXiv.org Search Results

Back to Search form | Next 29 results

The URL for this search is http://arxiv.org:443/find/all/1/ti:+Gaia/0/1/0/2016,2017/0/1?per_page=100

Showing results 1 through 100 (of 129 total) for ti:Gaia

1. arXiv:1703.00762 [pdf, other]
   The Gaia-ESO Survey: radial distribution of abundances in the Galactic disc from open clusters and young field stars
   L. Magrini, S. Randich, G. Kordopatis, N. Prantzos, D. Romano, A. Chieffi, M. Limongi, P. Francois, E. Pancino, E. Friel, A. Bragaglia, G. Tautvaivsiene, L. Spina, J. Overbeek, T. Cantat-Gaudin, P. Donati, A. Vallenari, R. Sordo, F. M. Jimenez-Esteban, B. Tang, A. Drazdauskas, S. Sousa, S. Duffau, P. Jofre, G. Gilmore, S. Feltzing, E. Alfaro, T. Bensby, E. Flaccomio, S. Koposov, A. Lanzafame, R. Smiljanic, A. Bayo, G. Carraro, A. R. Casey, M. T. Costado, F. Damiani, E. Franciosini, A. Hourihane, C. Lardo, J. Lewis, L. Monaco, L. Morbidelli, G. Sacco, L. Sbordone, C.C. Worley, S. Zaggia
   Comments: 20 pages, 8 figures, 4 table, online tables sent on request
   Subjects: Astrophysics of Galaxies (astro-ph.GA); Solar and Stellar Astrophysics (astro-ph.SR)

2. arXiv:1703.00673 [pdf, ps, other]
   Precise CCD positions of Triton in 2014-2016 using the newest Gaia DR1 star catalogue
   N. Wang, Q. Y. Peng, H. W. Peng, H. J. Xie, S. Ma, Q. F. Zhang
   Comments: 5 pages, 3 figures, accepted for publication on MNRAS
   Subjects: Earth and Planetary Astrophysics (astro-ph.EP); Instrumentation and Methods for Astrophysics (astro-ph.IM)

3. arXiv:1702.07741 [pdf, other]
   Toward a renewed Galactic Cepheid distance scale from Gaia and optical interferometry
   Pierre Kervella, Antoine Mérand, Alexandre Gallenne, Boris Trahin, Nicolas Nardetto, Richard I. Anderson, Joanne Breitfelder, Laszlo Szabados, Howard E. Bond, Simon Borgniet, Wolfgang Gieren, Grzegorz Pietrzynski
   Comments: 5 pages, 4 figures, proceedings of the 22nd Los Alamos Stellar Pulsation Conference "Wide-field variability surveys: a 21st-century perspective" held in San Pedro de Atacama, Chile, Nov. 28-Dec. 2, 2016
   Subjects: Solar and Stellar Astrophysics (astro-ph.SR)

4. arXiv:1702.06393 [pdf, other]
   The SOPHIE search for northern extrasolar planets - XII. Three giant planets suitable for astrometric mass determination with Gaia
   Javiera Rey, Guillaume Hébrard, François Bouchy, Vincent Bourrier, Isabelle Boisse, Nuno C. Santos, Luc Arnold, Nicola Astudillo-Defru, Xavier Bonfils, Simon Borgniet, Bastien Courcol, Magali Deleuil, Xavier Delfosse, Olivier Demangeon, Rodrigo F. Díaz, David Ehrenreich, Thierry Forveille, Maxime Marmier, Claire Moutou, Francesco Pepe, Alexandre Santerne, Johannes Sahlmann, Damien Ségransan, Stéphane Udry, Paul A. Wilson
   Comments: Accepted for publication in A&A. 14 pages, 6 figures, 7 tables
   Subjects: Earth and Planetary Astrophysics (astro-ph.EP)

5. arXiv:1702.06296 [pdf, other]
   Gaia eclipsing binary and multiple systems. Supervised classification and self-organizing maps
   M. Süveges, F. Barblan, I. Lecoeur-Taïbi, A. Prša, B. Holl, L. Eyer, A. Kochoska, N. Mowlavi, L. Rimoldini
   Comments: 20 pages, 22 figures. Accepted for publication in A&A
   Subjects: Instrumentation and Methods for Astrophysics (astro-ph.IM); Solar and Stellar Astrophysics (astro-ph.SR)

6. arXiv:1702.06083 [pdf, other]
   The Gaia-ESO Survey. Mg-Al anti-correlation in iDR4 globular clusters
   E. Pancino (INAF-OAA, ASDC), the GES collaboration
   Comments: 10 pages, 6 figures, accepted by A&A
   Subjects: Solar and Stellar Astrophysics (astro-ph.SR); Astrophysics of Galaxies (astro-ph.GA)

7. arXiv:1702.06046 [pdf, ps, other]
   A TGAS/Gaia DR1 parallactic distance to the sigma Orionis cluster
   J. A. Caballero
   Comments: Astronomische Nachrichten, accepted
   Subjects: Solar and Stellar Astrophysics (astro-ph.SR)

8. arXiv:1702.05621 [pdf, ps, other]
   UCAC5: New Proper Motions using Gaia DR1
   Norbert Zacharias, Charlie Finch, Julien Frouard
   Comments: 15 pages, 15 figures, 3 tables, accepted by AJ
   Subjects: Instrumentation and Methods for Astrophysics (astro-ph.IM)

9. arXiv:1702.04989 [pdf, other]
   The Short Term Stability of a Simulated Differential Astrometric Reference Frame in the Gaia era
   U. Abbas, B. Bucciarelli, M. G. Lattanzi, M. Crosta, M. Gai, R. Smart, A. Sozzetti, A. Vecchiato
   Comments: 14 pages, 7 figures, accepted by Publications of the Astronomical Society of the Pacific
   Subjects: Instrumentation and Methods for Astrophysics (astro-ph.IM)

A mean of one scientific paper every ~2 days!

Next Gaia Data Releases

- Gaia DR2: for April 2018
  - The 3D Galaxy in full colour!
  - Positions, distances, proper motions and integrated photometry in three bands: >1000 millions
  - Radial Velocities: ~1 million
  - Effective temperature estimations: ~10 millions
  - Temporal astrometry for ~10 mil asteroids
- Final Publication: ~2022
  - Full Catalogue with astrometry, photometric and radial velocities
  - Full solutions for variables and binary stars
  - Stellar classification and astrophysics parameters
  - List of exoplanets

Conclusions

- Gaia data processing is really complex:
  - Huge number of records and data types with relatively large data volume
  - Large dimensionality in the instrument calibrations, in practice like having 2000 different instruments (2 FoV, 100 CCDs, Modes)
  - Complicated data flow dependencies
  - Complicated validation due to combined time, spatial and calibration dependencies
- Quite tailored processing approach, execution framework and validation tools
  - Difficult to export the tools directly but the gained expertise may be handy for other teams

Relativistic Astrophysics

Vniversitat de València

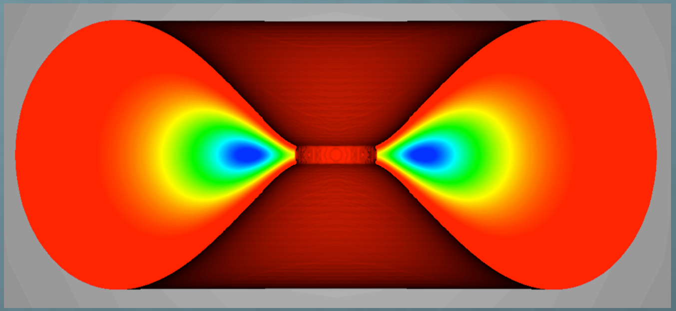
Universidad de Valencia

Relativistic Astrophysics

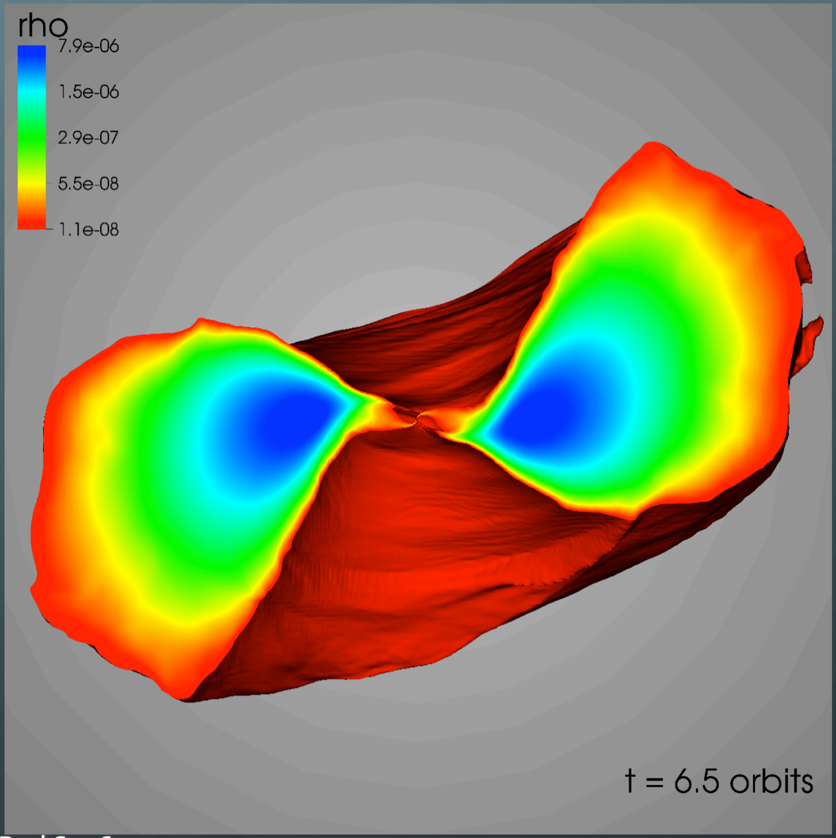
- **Staff:**
  - **P.I.: Miguel Ángel Aloy** (Asoc. Prof., UV)
  - **José Mª. Ibáñez** (Full Prof., UV)
  - **Manel Perucho** (Assit. Prof., UV)
  - **Vicent Quils** (Asoc. Prof., UV)
- **Computer Scientists:**
  - **Carmen Aloy** (SIUV)
- **Faculty:**
  - **Petar Mimica** (UV
  - **Martin Obergaulinger** (UV)
  - **Pablo Cerdá** (UV)
  - **Tomasz Rembiasz** (UV)
  - **Susana Planelles** (UV)
- **Students:**
  - **Carlos Cuesta** (UV)
  - **José E. Adsuara** (UV)
  - **Jesús Rueda** (UV)
  - **Alejandro Torres** (UV)
  - Jens F. Mahlmann (UV)

Universidad de Valencia

Reseach Highlights



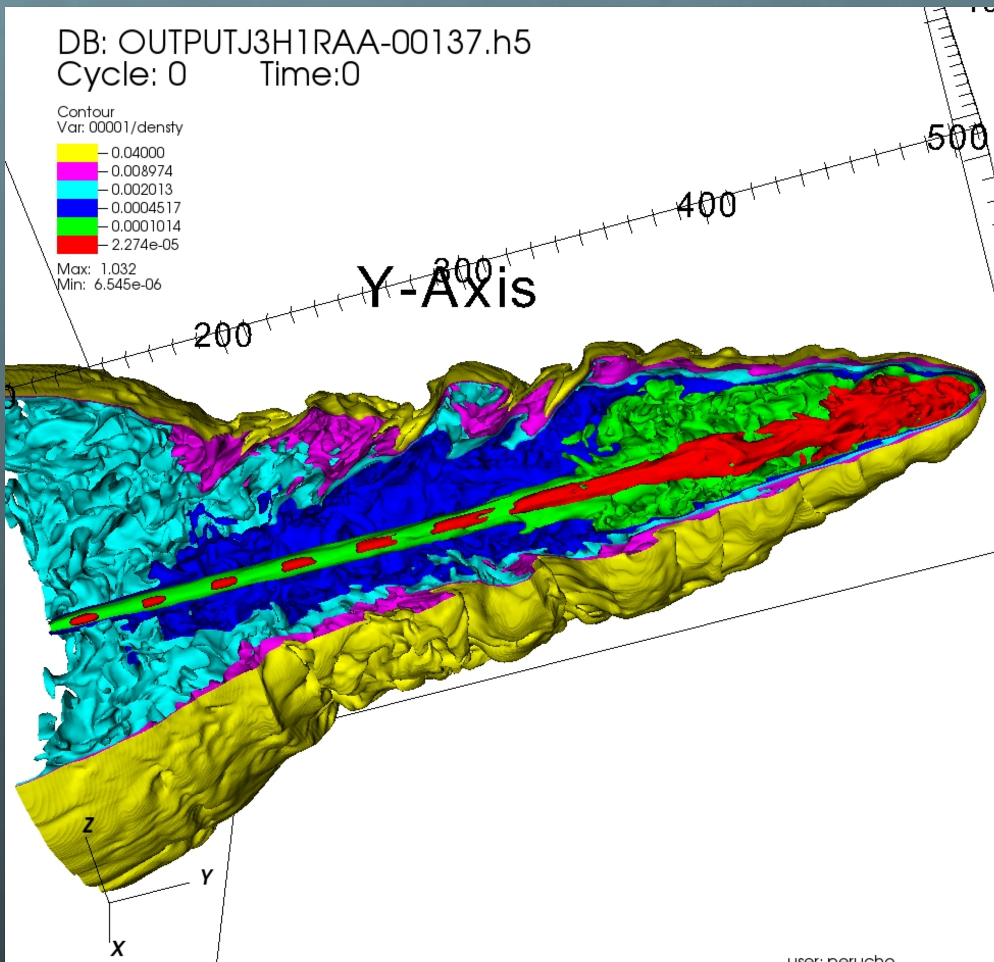
T=0



3D Density evolution of a numerical relativity simulation of a tilted accretion disc surrounding a Kerr black hole.

- CODE: Einstein Toolkit.
- Computer: HYDRA supercomputer at the Max-Planck-Institute for Astrophysics, Germany, running on 1000 cores.
- AMR with 13 box-in-box mesh refinement levels were used in order to be able to achieve sufficiently high resolution inside the BH horizon while having a large computational domain.

Reseach Highlights



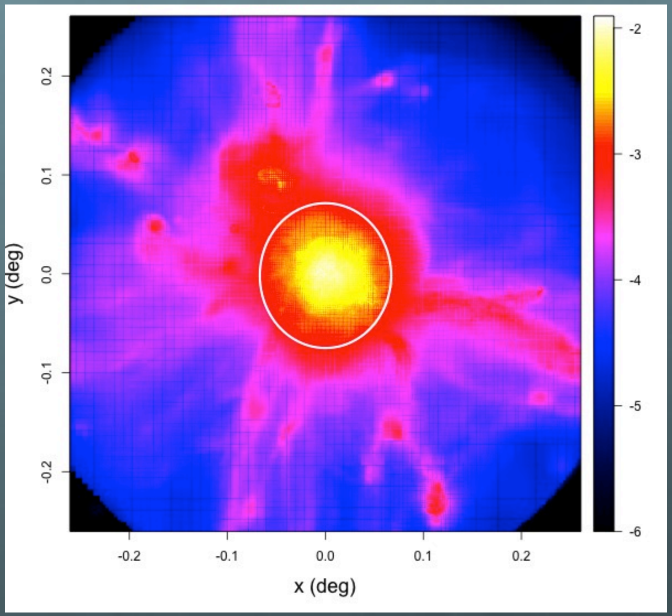
3D evolution of extragalactic jets at kiloparsec scales.

- CODE: RATPENAT.
- Computer: Tirant at the University of Valencia running on 256 cores and on Marenostrum on 1024 cores.
- PPM simulations with 512^3 cells (Tirant) and 1024x1024x2048 cells (Marenostrum).

The simulations show relativistic jets evolving in realistic galactic atmospheres, including inhomogeneities in the ambient medium and mass-load by stellar winds. We have run a number of jets with different properties and ambient media. The aim of the project is to study both the jet evolution and the influence of jets in the host galaxies.

Reseach Highlights

Column density of a cluster



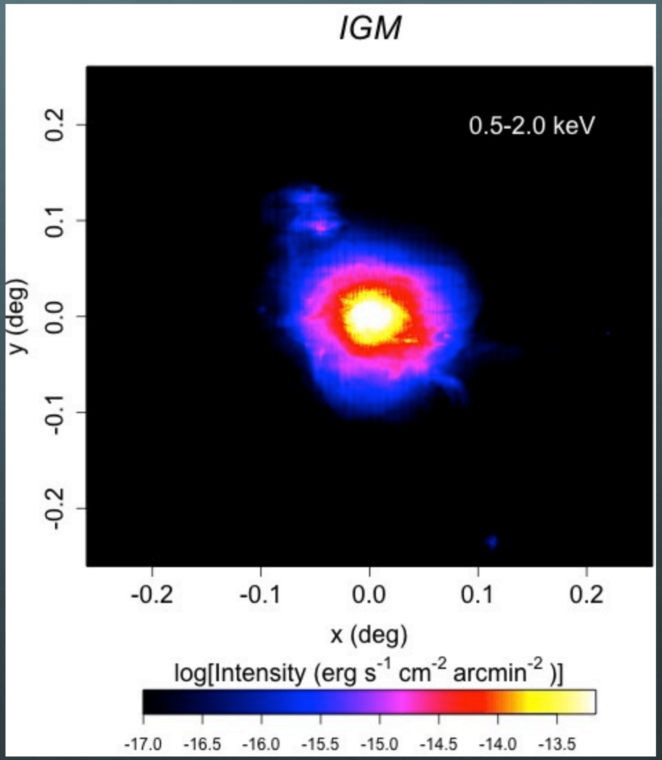
Simulation by MASCLET image computed with SPEV.

Shown: line integral of the volume density of the cells observed simultaneously with the coordinate origin when it is located at z=0.36.

Soft X-ray emission from galaxy clusters

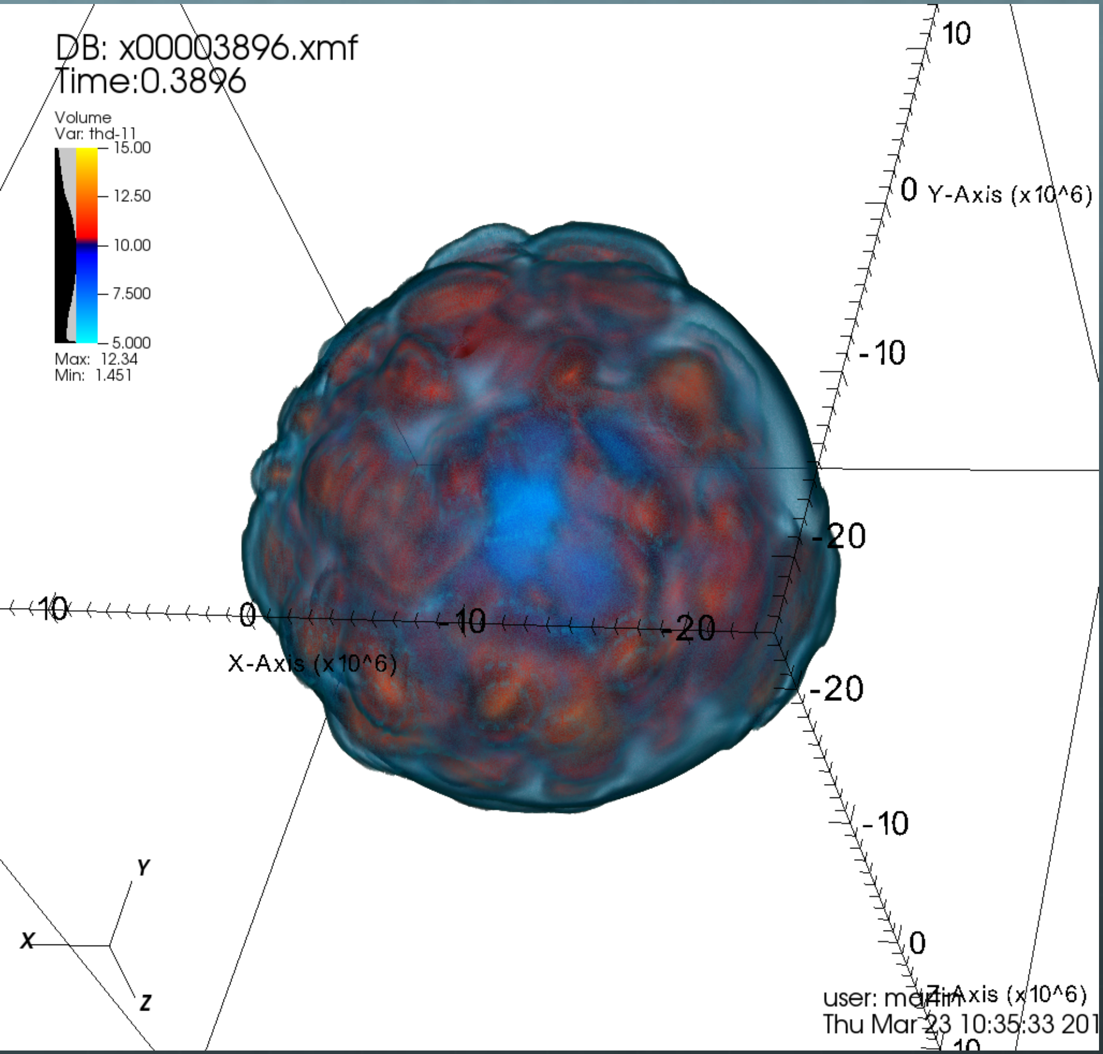
- CODE: MASCLET + SPEV.
- Computer: Lluisvives at the University of Valencia, running on 354 cores.
- Cubic comoving box of 40 Mpc covered with 9 AMR refinement levels + dark matter particles

Thermal free-free X-ray emission



Synthetic image in units of typical X-ray satellites.

Reseach Highlights



3D evolution of progenitors of Gamma-Ray Bursts.

- CODE: AENUS.
- Computer: Marenostrum+ at the BSC, running on 1024 cores.
- 9th order simulation with 420x180x64 zones (covering a domain of $(250\ \text{km})^3$ + relativistic MHD + accurate neutrino transport + approx. GR gravity.

Volume rendering of the specific entropy of the post-shock region of core z35 at t ≈ 150 ms after bounce. Convective plumes appear as high-entropy bubbles (red) located between low-entropy downflows (blue). The shock wave can be identified as the outermost surface of the rendering.

Challenges in Computational Relativistic Astrophysics

- Fast and massive parallel solvers for elliptic systems of PDEs (Poisson, CFC).
  - Adapted to the hybrid nature of current and future architectures: mixed GPU+CPUs model.
  - Must be 3D: likely excludes direct methods (memory-limited). [UV poster]
- Accuracy in multidimensions: *ultrahigh* order (>3rd) in an unsplit fashion.
- Stiffness:
  - Including nuclear reaction networks: treatment of (very) stiff ODEs involving HUGE number of isotopes and reaction channels.
  - Non-ideal relativistic MHD in the limit of high conductivity.
  - Including subgrid turbulence models for the unresolved scales
- Machine learning tools for:
  - Processing:
    - Where are *now* / will be *later* shocks and contact discontinuities?
    - Sensitivity of the results to "small" parameter changes.
  - Post-processing:
    - "interpolation" of the locations of shocks between snapshots.
    - Pattern recognition: shocks as coherent structures.
- Coupled transport (both of photons and of neutrinos) to dynamics.

SyeC
Supercomputación y eCiencia

UAM UNIVERSIDAD AUTONOMA DE MADRID

Grupo de Astrofísica y Cosmología Computacional

PROFESORES PERMANENTES

Gustavo Yepes Alonso (UAM, P. Titular)
Rosa Dominguez Tenreiro (UAM, Catedratica)
Alexander Knebe (UAM, Contratado Doctor)
Chris Brook (UAM, Ramon y Cajal Fellow, now at IAC)

Postdocs y PhD students

Federico Sembolini (Postdoc, now at GMV)
Jesus Vega Ferrero (Postdoc, now at Philadelfia)
Sergio Rodriguez (PhD)
Santiago Avila (Postdoc, now at UK.)
Isabel Santos (Phd)
Robert Mostoghiu (phd)
Doris Stoppacher (PhD)

CFD en Cosmología

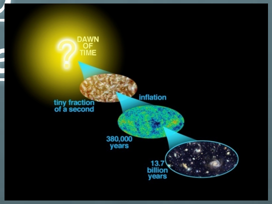
- Cosmo CFD tiene un problema añadido. LA GRAVEDAD.
- En la mayor parte de los problemas astrofísicos es necesario resolver simultáneamente las ecuaciones de la dinámica de gases (Euler) junto con la ecuación de Poisson (potencial gravitatorio).
- Los códigos numéricos son una combinación de métodos de N-cuerpos para simular la dinámica gravitatoria del fluido, acoplados con metodos de resolución de las ecuaciones hidrodinamicas.
- La física del colapso gravitatorio hace que los contrastes de densidad aumenten con el tiempo de evolución, dando lugar a enormes rangos dinámicos (10^5 o mayores)

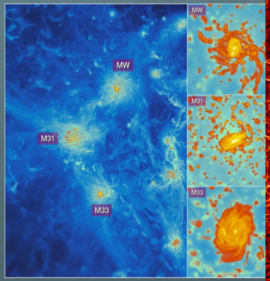
Numerical Methods

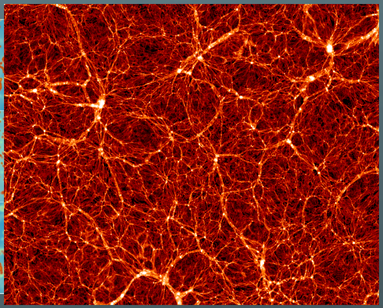
- Lagrangian Methods: most often used in Computational Astrophysics and Cosmology:
  - *Gravitational Processes*: N-body methods:
    - Particle-Mesh (FFTW) + Tree algorithms
  - *Gasdynamics:*
    - Smoothed Particle Hydrodynamcs (SPH)
  - Very useful because they keep the number of computational elements constant during the simulation. Great help controlling memory allocation.
  - Considerably less memory comsumption than Eulerian AMR methods.
- Most often used codes:
- GADGET (Springel 2005) is the most popular code in Computational cosmology. Full OpenMP + MPI parallel with many physics modules.
- RAMSES (Teyssier, 2002) is also very popular. Eulerian AMR MPI +OpenMP
- AREPO (Springel 2009). Moving voroni mesh + Godunov hydro solver.MPI+OMP
- Multiscale Radiative processes: (sub-resolution modeling, cooling, heating, star formation, SN and Black Hole feedbacks, change in number of computational elements

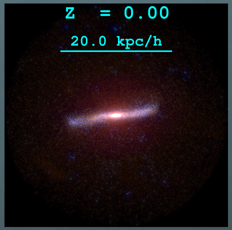
TEMAS DE INVESTIGACIÓN UAM

- *Simulaciones numéricas de la formación de estructuras en e Universo a distintas escalas*









Supercúmulos, filamentos, vacíos etc.

(> 10 Mpc)

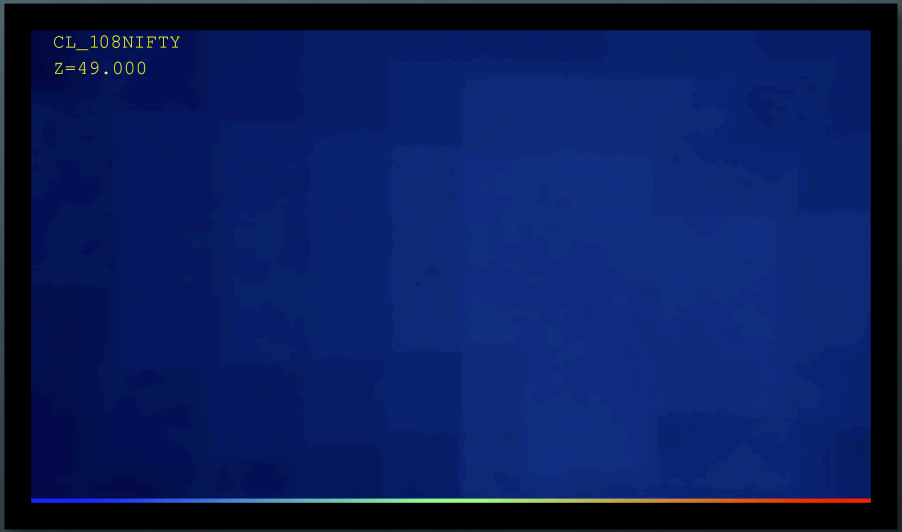
Cúmulos y Grupos de Galaxias (1-10 Mpc)

Galaxias (< 1 Mpc)

Reproducir el Universo Local: *CLUES*

Retos:* *Aumentar rango dinamico:

Gran volumen Y alta resolución



Video: Formacion de un cúmulo de galaxias, el objeto más masivo de universo. Simulacion hidrodinamica con gas formacion estelar y materia oscura

Improvements for Future HPC

- **SOFTWARE DEVELOPMENTS:**
  - Improvement in domain decomposition algorithms for better work Load-Balance of N-body + SPH codes.
  - Optimization of codes for hybrid HPC platforms (GPU's or other accelerators)
  - New interactive visualization techniques for 3D datasets (Virtual Reality).
  - Use of BIG DATA tools to datamining the 5D databases from simulations.
- **HARDWARE DEVELOPMENTS:**
  - Help needed in porting the codes to upcoming 100+ PFLOPS--Exascale platforms with multiples nodes of many core chips each.
  - Very likely need to re-engeineering all our simulation software to make efficient use of the new infrastructures.
    - Example SWIFT code: task-based parallelism of a SPH kernel.
    - http://icc.dur.ac.uk/swift/

PROBLEMS TO BE SOLVED

- Main problem is how to link the dark sector (halos) with the visible universe (galaxies, gas, stars)
- Apply semianalytical models to "paint" dark matter halos with galaxies. Need to measure the merger trees (time evolution patterns of dark matter halos)
  - Codes are very poorly optimized to deal with trillion particles simulations
- Need more sophisticated methods from Big-data techniques to do that.
  - Machine learning methods starting to be used in cosmology. Use results from hydro simulations or semianalytics models to learn how to assign galaxies to halos. Then use them in very large computational volumes of dark matter only simulations.
- Data products have to be made publicly available to the community.
  - Datamining of huge multidimensional databases with halos + galaxy properties. (e.g. 3D searching, online analysis tools, etc)