CU8 products in DR3: Source classification and physical parameters

REG-RIA meeting: Expanding the Gaia Legacy.
The role of Spanish ground-based facilities.
A celebration of the research career of Jordi Torra

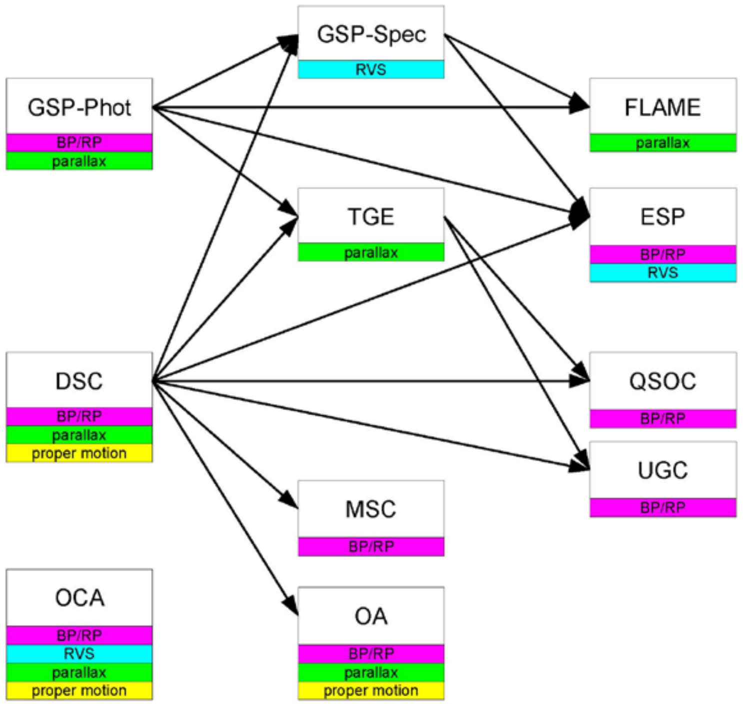
Minia Manteiga
DPAC-CU8
UDC

ICCUB, Barcelona 17-19 February 2020

1.- Outline

- CU8 contribution to DR2
- CU8 in DR3:
  - astronomical classes
  - astrophysical parameters
  - Stellar APs from RVS spectra
- Description of APSIS modules: dependencies, algos, products
- CU8 products in DR3

APSIS



→ From Gaia data

2.- CU8 contribution to DR2

- **Stellar Parameters**: Teff - A_G - $E(G_{BP}$- $G_{BRP})$ – Radii - Luminosities
  161-77 million stars with G≤17
- Based on **integrated photometry** in 3-bands + **parallaxes**. Strongs degeneracies between Teff and A_G, and $E(G_{BP}$- $G_{RP})$
- **Machine learning algorithm used: EXTRATREES** regression with an ensemble of 201 trees trained with Gaia observations of sources with literature data (Teff 3K-10K) or synthetic models (A_G, $E(G_{BP}$- $G_{BRP})$).
- Luminosities from G, bolometric corrections and parallaxes assuming A_G=0. And Radii from Teff and L.

Most used algorithm: **EXTRATREES** regression

Supervised (trained) algorithm based on a decision tree
Parameters:

- max_depth, min_samples_split, min_samples_left,max_leaf_nodes, min_impurity_split
- Impurity function: Entropy, Gini index, classification error
- Splitter criterio

Random Forest: random subsamples to train, uses a subsample of splitting criteria, performs statistics over the trees

Extremely Randomized Trees: No boostrapping in sampling and random selection of thresholds, MCMC samples

3.- CU8 products in DR3

- Only Gaia information:
  - parallaxes (sky position, RUWE, proper motions,..)
  - BP/RP spectrophotometry
  - RVS spectra
- Algorithms trained with semiempirical or synthetic models
- Three types of products:
  - **Astronomical classes** (including outlier analysis)

    stars– binary stars – WDs – unresolved galaxies – quasars

  - **Astrophysical parameters** (not only for stars)
  - **Multidimensional parameters** (Extinction map, SOM map)

Probably FILTERS will be applied after VALIDATION
What I am showing here is for the most favorable case

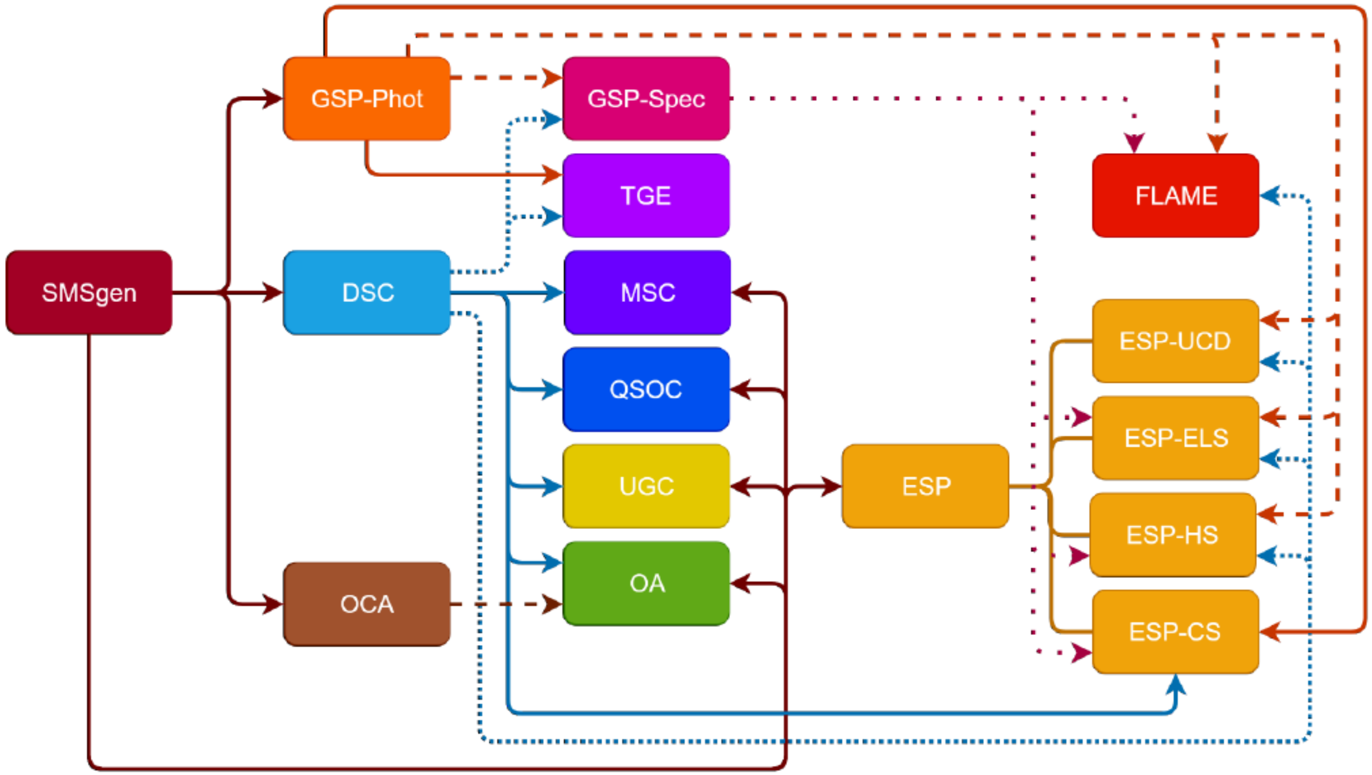
4.- APSIS modules and interconexions

General modules: DSC, GSP-Phot, TGE, FLAME

Specific modules: QSOC, UGC, MSC and OA

Extended parameterization for stellar classes: UCD, ELS, HS, CS

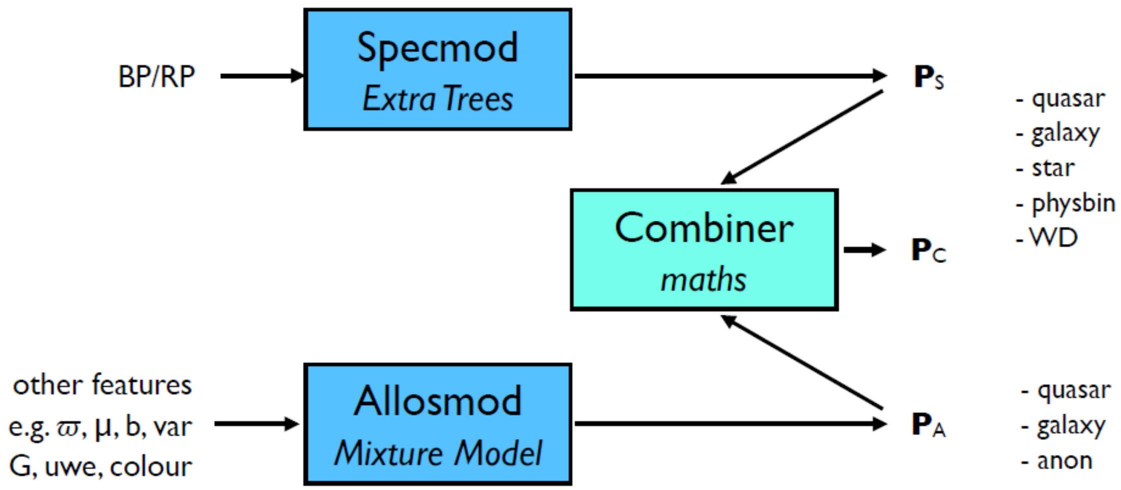
Based on RVS data: GSP-Spec



4.- APSIS modules

- **DSC: "Discrete source classifier" (C. Bailer-Jones et al. , Heidelberg)**

2 supervised algorithms:



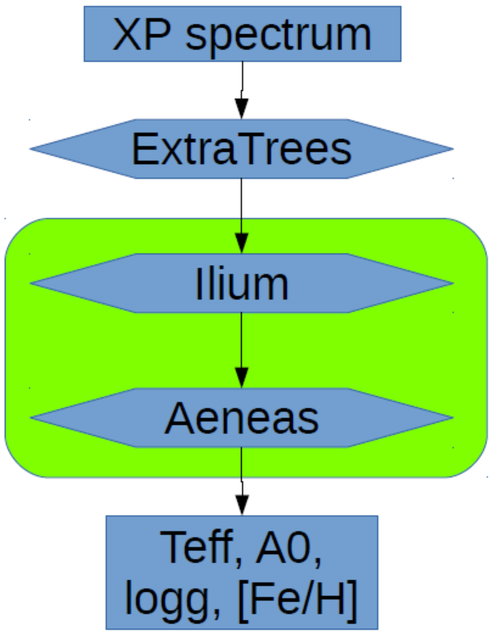
- Algorithm trained with data from SDSS (galaxies and qsos), Kareen El-Badry catalogue of WDs, synthetic binary pairs and a random sample (stars).
- **Outliers:** max(Pc)<0.5 → input of OA module

4.- APSIS modules

- **GSP-phot: General Stellar Parametrizer by spectrophotometry (R. Andrae et al., Heidelberg)**

Forward model fitting labelled XP. Uses Newton-Rapson minimization

ExtraTrees algo. Fully Bayesian method



DR3 expected products:

- Teff, A0, logg, [Fe/H]
- possibly A_G, M_G, E(BP-RP)
- values + 1D uncertainties (e.g. percentiles)
- no symmetric "Gaussian" errors
- no "Gaussian" covariance matrices
- MCMC chains or ExtraTree ensemble

Requires synthetic XP spectra and uses different stellar libraries: MARCS, PHOENIX, A, OB

4.- APSIS modules

- **QSOC: Quasar Classifier (L. Delchambre et al., Lieja)**

DR3 products:

- redshift
- quasar type (Type I qso or BAL qso)

Algorithm: weighted phase correlation (WPC) on synthetic templates

- **UGC: Unresolved Galaxiy classifier (I. Bella-Velidis et al., Athens)**

DR3 products:

- gal_type (spectral)
- sfr_pars star-forming-rate parameters (different per class, totally 12)
- z (redshift 0.0-0.6)

Algorithm: SVM, supervised binary algorithm

4.- APSIS modules

- **MSC: Multiple Star classifier (J. Rybicky, Heidelberg)**

DR3 products:
Stellar parameters for both components of a coeval, non-interacting binary system: Teff12, logg12, A0, [Fe/H], distance [Phoenix models]
Algorithm: Extratrees: – Empirical training – GUMS APs adding fluxes of single star, forward model

- **TGE: Total integrated extinction to the edge of the Galaxy in a field (important for extragalactic objects) (R. Drimmel et al., Torino)**

DR3 products: all-sky HEALPix maps of the total Galactic extinction and related uncertainties at 4 separated HEALPix levels, 6-9. Based on Cardelli et al. 89.
Mostly single red giant stars as tracers + parallax (+GSP-Phot params)

4.- APSIS modules

- **FLAME: Final Luminosity Age Mass Estimator (O. Crevey, Niza)**

Inputs: from GSP-phot: Teff, [Fe/H], A_G + G + π + table bolometric corrections

Products: Radius (R, 0.1-200), Luminosity (L, 0.05-1000), Mass (M, 0.5-2.5), Age (A), Evolstate (ES, 'MS/sG/G'), Grav Redshift (GR - tbd), also publish: BC value, + log (models)
Bayesian treatment of uncertainties

- **ESP: Extended stellar parameters for : (* use both XP and RVS)**

  - Ultra cool stars, (L. Sarro, Madrid)
  - Cool stars, (A. Lanzafame, Catania) *
  - Hot stars (Y. Fremat, Brussels)
  - Emission line stars (Y. Fremat, Brussels) *

4.- APSIS modules

- **OA: Analysis of classification outliers using unsupervised clustering (M. Manteiga et al, GGG Galicia)**

  Method: Self-Organizing Maps.
  Visualization tool: GUASOM

- **GSP-Spec:General Stelar Parameters fron spectroscopy**

(A Recio-Blanco et al., Niza, M.A. Alvarez et al GGG Galicia))

Input: RVS calibrated spectra + grid of synthetic models (MARCS)

Products: Teff, logg, [Fe/H], [M/H], [α/Fe], [Fe/H], [Ca/H], [Si/H], [Ti/H], DIB properties + uncertainties

Several algorithms with supervised training

Limited to rather bright stars (G≈16, may be brighter). Same parameters will only be available for a the brightest stars

Astrophysical Parameter Groups in Main parameters

- **Spectroscopic:** teff logg mh vsini spectraltype activityindex
- **Global/Fundamental:** mg, radius, lum, gravredshift, mass, age, evolstate, accrfillfactor, massloss
- **Interstellar:** distance azero ag abp arp ebpminrp dib
- **Abundance:** alphafe feh sih, cah, tih ch, nh, index, ew
- **Class:** classprob, classlabel, spectraltype, neuronId, neurondist, neurondistquan
- **Binary:** lumratio teff, logg, e.g. teffBinary1, teffBinary2

Auxiliary: bc, flags, gof, mcmcaccept, libname, autocorr, algoId,

5.- CU8 tables in DR3

CU8 publication in 5 Tables and MDD data

- GaiaSource: GspPhot standard params (like DR2), DSC prob, flags
- AstrophysicalParameters & Supp table: many of CU8 WPs
- QSO table: DSC (classification), QSOC (quasar) + CU3
- GALAXY table: DSC, UGC (galaxy) + other CU4

- Multi-dimensional data
    - GSPPhot's / MSC's MCMC chains
    - TotalGalacticExtinctionMap & TotalGalacticExtinctionMapOpt
    - OaNeuronInformation & OaNeuronXpSpectra