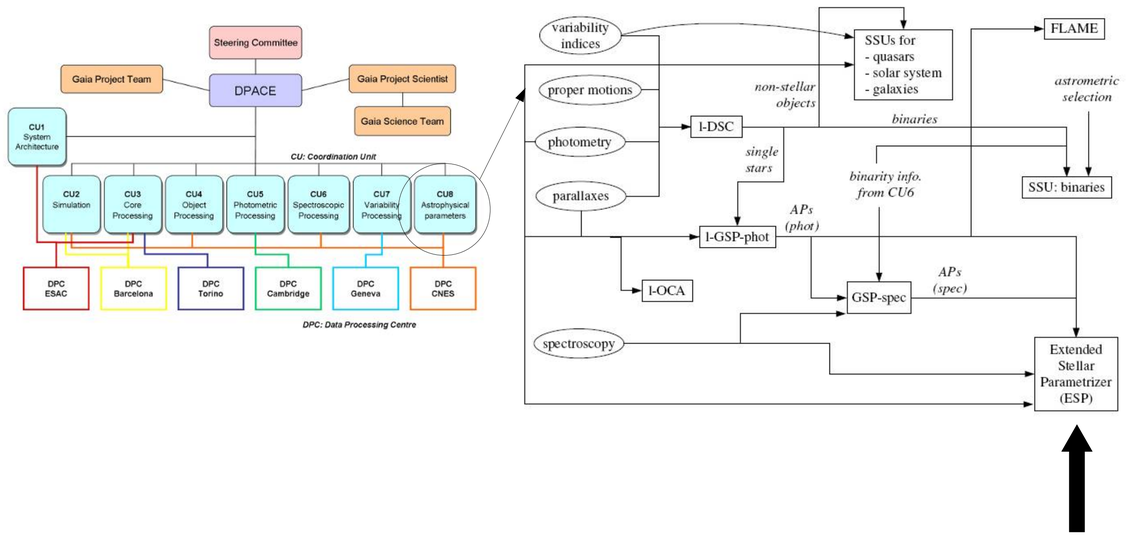
Astrophysical Parameters of the GAIA Ultracool Dwarves



ULTRA COOL DWARVES (ESP_UCDS)

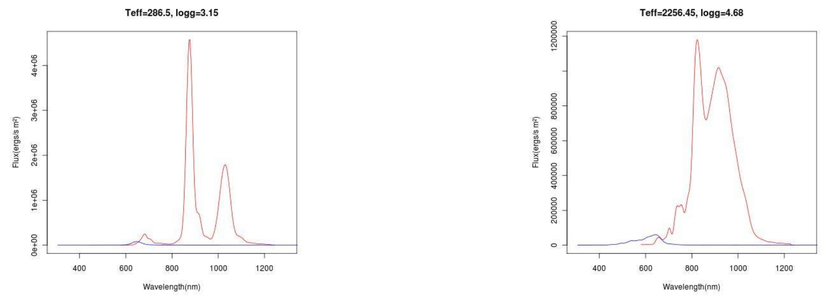
We use France Allard's models (http://perso.ens-lyon.fr/france.allard/) based on the Phoenix model atmospheres code. Phoenix is a static and radial (1D) code which is however general, modeling as well novae and supernovea envelopes in relativistic expansion phases and extrasolar planets irradiated by a star within the hydrostatic equilibrium and spherical symmetry approximations.

Amongst the model atmosphere grids that France Allard has constructed, we use

- AMES-Dusty for 2500K > Teff > 1500K (brown dwarfs/extrasolar planets without irradiation, with dust opacity)
- AMES-Cond for Teff < 1600K (brown dwarfs/extrasolar planets without irradiation, no dust opacity)

DPAC-CU2/CU8 have simulated BP/RP spectra from these models for different apparent magnitudes (G=8, 11, 15; fainter simulations soon to be available). For each model physics (Cond or Dusty), we have two differents sets og GAIA simulations: NOM stands for a so called nominal grid with equispaced values in Teff and log(g) that reflects the initial availability of model atmospheres. RAN stands for a grid with values of effective temperature and gravities generated randomly from a set of evolutionary tracks, and interpolated in Allard's nominal grid.

G=8.0



Our mision is to determine and predict effective temperatures (Teff) and gravities (logg) for GAIA Ultra Cool Dwarves candidates. The analysis consists in an iterative scheme that combines a preprocessing stage whereby optimal attributes are selected, a regression stage that aims at predicting the dependent variables (Teff, logg) based on the predicting variables, and finally, the assessment stage where the various solutions are quantitatively evaluated according to the errors in the predictions.

PREPROCESSING:

Preprocessing consists in modifying the source data in to a different format which

(i) enables data mining algorithms to be applied easily
(ii) improves the effectiveness and the performance of the mining algorithms
(iii) represents the data in easily understandable format for both humans and machines
(iv) supports faster data retrieval from databases
(v) makes the data suitable for a specific analysis to be performed.

Up until now, we have only tested two simple preprocessing techniques: Principal Components Analysis and Partial Least Squares combined with areal normalization. These will serve as a basis for evaluating more complex techniques if needed.

Principal component analysis (PCA) involves a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. Depending on the field of application, it is also named the discrete Karhunen–Loève transform (K.L.T.), the Hotelling transform or proper orthogonal decomposition (POD).

Partial least squares regression (PLS-regression) is a statistical method that bears some relation to principal components regression; instead of finding hyperplanes of maximum variance between the response and independent variables, it finds a linear regression model by projecting the predicted variables and the observable variables to a new space. Because both the X and Y data are projected to new spaces, the PLS family of methods are known as bilinear factor models. It is used to find the fundamental relations between two matrices (X and Y), i.e. a latent variable approach to modeling the covariance structures in these two spaces. A PLS model will try to find the multidimensional direction in the X space that explains the maximum multidimensional variance direction in the Y space. PLS-regression is particularly suited when the matrix of predictors has more variables than observations, and when there is multicollinearity among X values. By contrast, standard regression will fail in these cases. Here we intend to apply **Kernel Partial Least Squares** in order to account for the nonlinearity of the mapping.

REGRESSION STAGE:

The training grid is now prepared to feed various regression approaches. The benchamark techniques that we have used are

•**Support vector machine**s (SVMs) are a set of related supervised learning methods used for classification and regression. In simple words, given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that predicts whether a new example falls into one category or the other. Intuitively, an SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. More formally, a support vector machine constructs a hyperplane or set of hyperplanes in a high or infinite dimensional space, which can be used for classification, regression or other tasks. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training datapoints of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier.

•An **artificial neural network** (ANN), usually called "neural network" (NN), is a mathematical model or computational model that tries to simulate the structure and/or functional aspects of biological neural networks. It consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. Modern neural networks are non-linear statistical data modeling tools. They are usually used to model complex relationships between inputs and outputs or to find patterns in data.

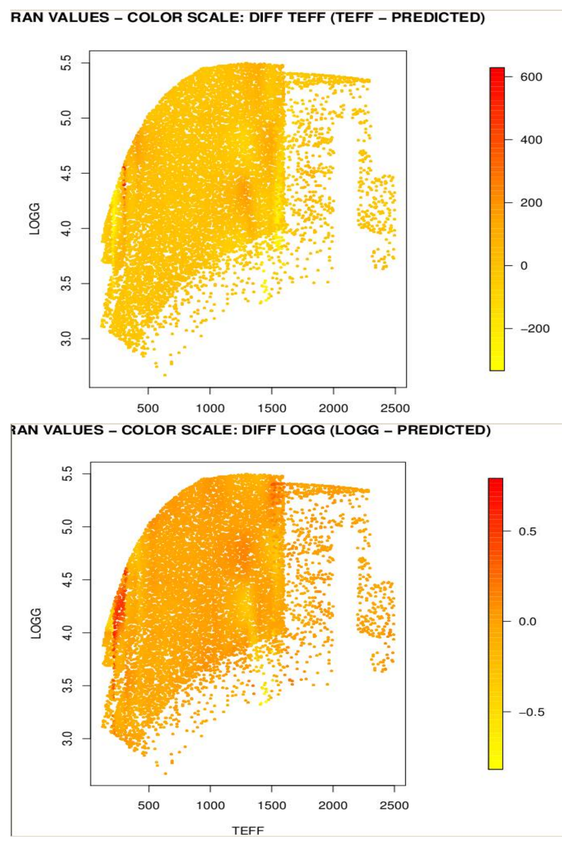
In the following we describe the present situation and results obtained in our first rounds of training and testing. We have used Weka 3.6.1 as simulation software and R Statistical Computing 2.11.1 for grid data manipulation (running in a Intel Core 2 Duo computer with 4 Gb RAM and with O.S. Ubuntu 10.04). We use NOM data grids for training and RAN data grids for testing (to minimize overfitting). We obtain best results with Support Vector Machines and RBF kernel (outperforming Support Vector Machines with Polynomial kernel and Neuronal Networks). Preprocessing includes areal normalization and PCA.

The figures to the right show the absolute difference between the predicted values of the effective temperatures (top) and logg (bottom) and the reference values in the NOM(INAL) and RAN(DOM) grids. The densest parts of the plot correspond to the COND models (100 < Teff < 1600). Dusty models overlap in the temperature range between 1500 and 1600, and extend up to Teff=2500, the limit adopted for the definition of ultracool dwarves.

The first tests indicate that Support Vector Machines with gaussian (RBF) kernels outperform significantly polynomial kernels and Artificial Neural Networks.

ALGORITHM	RELAT ABS ERR TEFF	ROOT REL SQU ERR TEFF	RELAT ABS ERR LOGG	ROOT REL SQU ERR LOGG
Unnormalized – PCA - SVM – Polykernel degree 1	66,56%	77,55%	81,25%	89,3251%
normalized – PCA - SVM – Polykernel degree 1	35,21%	41,84%	62,25%	67,96%
Normalized – PCA NOM SVM – Polykernel degree 1	29,54%	31,12%	41,44%	48,77%
Normalized – PCA NOM SVM – Polykernel degree 2	21,12%	25,51%	24,12%	28,32%
Normalized – PCA NOM SVM – RBF Kernel	6.76 %	9.62 %	16,48%	22,45%
Neuronal Network – Perceptron Multilayer	18,15%	21,05%	58,39%	61,81%

We include the best algorithm performance after exploration of the various parameter spaces



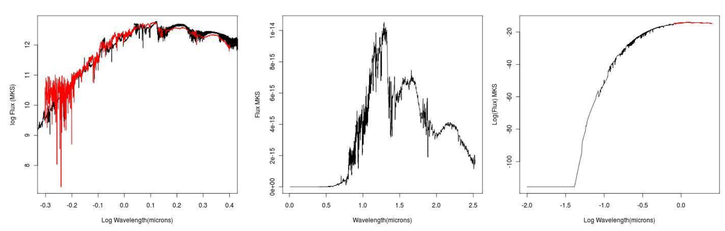
In order to assess the robustness against decreasing signal-to-noise ratios, we have introduced realistic GAIA noise as provided by CU2, and scaled it to magnitudes G=18, 19 and 20. In the following table we summarize the noise free performances for the optimal SVM model with gaussian (RBF) kernels and the evolution of the relative absolute errors with decreasing brightness.

CONDITIONS	RELAT ABS ERR	ROOT REL SQU ERR
NOM – CROSS VAL - TEFF	3.1347 %	6.521 %
DUST – G180 - TEFF	19,257%	26,57%
DUST – G190 - TEFF	30,24%	41,70%
DUST – G200 - TEFF	47,53%	66,41%
COND – G180 - TEFF	19,16%	22,58%
COND – G190 - TEFF	27,65%	33,51%
COND – G200 - TEFF	40,67%	50,68%
NOM – CROSS VAL - LOGG	7,97%	19,22%
DUST . G180 - LOGG	191%	213%
DUST – G190 - LOGG	291%	321%
DUST – G200 - LOGG	439%	473%
COND – G180 - LOGG	37,5%	42,5%
COND – G190 - LOGG	54,8%	63,2%
COND – G200 - LOGG	80,7%	94,6%

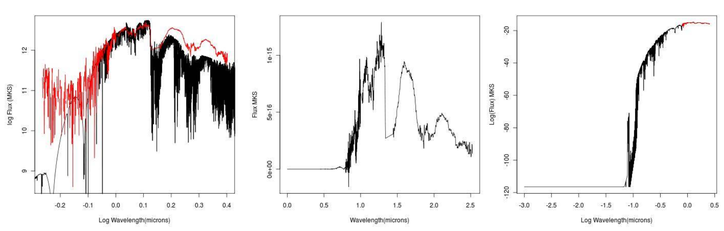
All plots contain completions of ground based spectra (from Legget's compilation http://staff.gemini.edu/~sleggett/LTdata.html) with models from Allard's grid. Left plots represent the SED (wavelength in microns), central ones the SED in logarithmic units; right plots, a close-up in the region around 1 micron. Red is used for the original ground based data.

Further tests with real UCDs and ground based telescopes

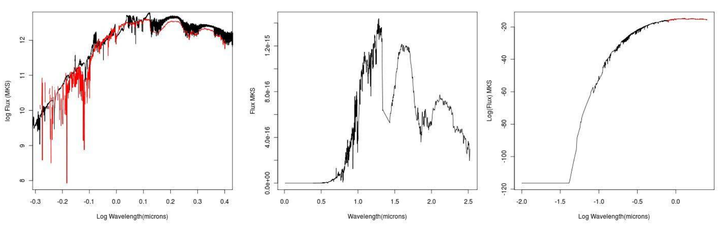
L1_2MASS0345+25



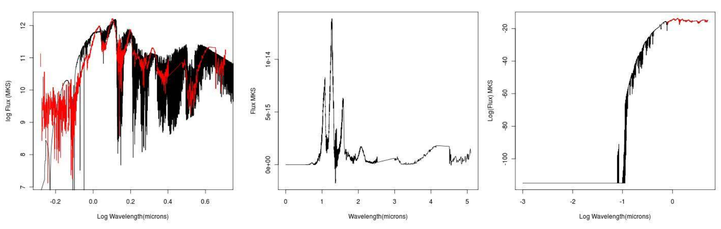
T0.5_SDSS0151+12



L6.5_SDSS0236+00



T7_GI229B



Gaia DPAC
Data Processing & Analysis Consortium