Astronomy & Astrophysics manuscript no. ordieres
October 2, 2016

Physical parameter estimates of M-type stars: a machine learning perspective.

L. M. Sarro[1], J. Ordieres-Mere[2], A. Bello-Garcia[3], A. Gonzalez-Marcos[4], and M.B. Prendes-Gero[3]

[1] [1] Universidad Nacional de Educación a Distancia,
Department of Artificial Intelligence. e-mail: lsb@uned.es
[2] [2] Universidad Politécnica de Madrid (UPM), PMQ Research Group,
José Gutiérrez Abascal 2, 28006 Madrid, Spain. e-mail: j.ordieres@upm.es
[3] [3] Universidad de Oviedo, Construction and Manufacturing Engineering Department,
Campus de Viesques s/n, Gijón, Asturias, Spain. e-mail: {abello,mbprendes}@uniovi.es
[4] [4] Universidad de la Rioja, P2ML Research Group,
Luis de Ulloa 20, 26004 Logroño, La Rioja, Spain. e-mail: ana.gonzalez@unirioja.es

Received October 2, 2016; accepted

ABSTRACT

Key words. class M stars – dynamic feature selection – physical parameter identification – Temperature, gravity and metalicity Modelling – Learning from BT-Settl spectra library

TODO: Luis, cambiar spectral libraries por stellar atmosphere models o synthetic spectral libraries.

1. Introduction

2. Methodology.

The objective addressed in this Section is to develop a procedure to identify spectral bands that yield good temperature, gravity and metallicity diagnostics. Given the lack of a calibration set of observed spectra with homogeneous coverage of the space of physical parameters, we turn to synthetic libraries of spectra. Furthermore, only temperatures and gravities can be calibrated independently of the spectra: all metallicity estimates in the literature are based on collections of synthetic spectra, and therefore spectral synthesis codes are the only resource to construct regression models.

The atomic or molecular line/band parameters could in principle indicate the spectral features that are more sensitive to changes in the physical parameters. The suitability of spectral features as diagnostics ofthe stellar atmospheric properties depends not only on the individual behaviour of each line/band, but also on the relative properties of neighbouring features in the same spectral region, that may overlap depending on the spectral resolution. Furthermore, good spectral diagnostics at a given signal-to-noise ratio (SNR) may show a severely degraded predictive power in the low SNR regime. In the following we adopt the BT-Settl library of synthetic spectra (Allard et al. (2013)) as the framework where spectral diagnostics will be searched for. These synthetic spectra were pre-processed in several steps as described below.

2.1. Spectral preprocessing

First, and in order to define good temperature diagnostics, spectra between 2000 and 4200K in steps of 100 K were selected, with $\log(g)$ in the range between 4 and 6 dex (when g is expressed in cm/s^{-2}), in steps of 0.5 dex. The metallicity of the representative spectra was restricted to the set 0, 0.5 and -1 dex. This yields a total set size of 535 available spectra.

A series of preprocessing steps were then carried out in order to match the spectral resolution and wavelength coverage and sampling of the synthetic library to that of the collection of observed spectra (IPAC or IRTF, see below). This required the definition of a common wavelength range present in all available observed spectra, and the subsequent trimming to match that range. A unique wavelength sampling was also defined and all spectra (synthetic and observed) interpolated to match the sampling. Finally, all spectra, both synthetic and observed were divided by the integrated flux in order to factor out the stellar distance.

In order to avoid selecting spectral features that are good predictors only in the unrealistic SNR=∞ regime, the search for optimal diagnostics of the atmopheric paramters of M stars was carried out for three SNR values (10, 50 and ∞) by degrading the synthetic spectra with Gaussian noise of zero mean. These values were found to be sufficient in a wide range of experiments carried out in parallel and described in González et al. (submitted).

2.2. Feature definition and selection

As mentioned in Sect. 1, it is well known the difficulty in defining good spectral diagnostics for M stars in the infrared.

The work in Cesetti et al. (2013) defined wavelength regions in the I and K bands optimal for the diagnostic of physical parameters based on the sensitivity exhibited by the flux emitted in these segments to changes of the physical parameters. The sensitivity was measured in terms of the derivative of the flux with respect to the physical parameter. The approach adopted in this work is to select spectral features that yield the best accuracy when used as predictive variables in a regression model that estimates the stellar atmospheric physical parameters (T_{eff}, log(g) and metallicity). The evaluation of the accuracy of the estimates produced from a subset of features is described further below. We consider the effective temperature as the dominant parameter influencing changes in the stellar spectra (a strong feature) and thus, it was estimated first, and then used as in the regression models for the gravity and metalicity.

Here, a feature F is defined as

$$F = \int_{\lambda_1}^{\lambda_2} (1 - \frac{f(\lambda)}{F_{cont}} \cdot \mathrm{d}\lambda) \quad (1)$$

where $f(\lambda)$ denotes the normalized flux from the star at wavelength λ, and where F_{cont} is the average flux in a spectral band between $\lambda_{cont;1}$ and $\lambda_{cont;2}$. We explain below how we search for the band definitions that produce physical parameter predictions with the smallest errors.

Another type of features defined as

$$F' = \frac{\int_{\lambda_1}^{\lambda_2} f(\lambda) \cdot \mathrm{d}\lambda}{\int_{\lambda_3}^{\lambda_4} f(\lambda) \cdot \mathrm{d}\lambda} \quad (2)$$

were considered, where $\lambda_1, \lambda_2, \lambda_3$, and λ_4 delimit two spectral bands such that the ratio of the integrated fluxes in the two bands is hoped to be a good predictor (alone or in combination with other features) of the star atmospheric physical parameters. The results obtained with this alternative feature definition did not differ significantly on average from the ones observed with the one adopted in Eq. 1, and including them here would result in an excessively lengthy paper. In view of the equivalent global performances, we prefered the former because it allows direct comparison with the features proposed by Cesetti et al. (2013).

We used Genetic Algorithms to solve the optimization problem described above, that is, the problem of finding the features (band boundaries) that minimize the prediction error of a regression estimate of the physical parameters. We used the implementation of genetic algorithms publicly available as the R (R Core Team 2013) GA package. The concept of using in-silico evolution for the solution of optimization problems was introduced by Holland (1975). Although its application is now reasonably widespread (Goldberg et al. 1989, see e.g.)), they became very popular only when sufficiently powerful computers became available. **Aquí hay que citar trabajos en astrofísica que utilicen GA y, en particular, un artículo de Charbonneau http://adsabs.harvard.edu/abs/1995ApJS..101..309C en 1995 que fue como la presentación en sociedad.**

For the sake of simplicity let us define Genetic Algorithms (GAs) as search algorithms that are based on the principle of evolution by natural selection. The procedure works by evolving (in the sense explained below) sets of variables (chromosomes) from an initial random population. Evolution proceeds via cycles of differential replication, recombination and mutation of the fittest chromosomes. The concept of fittest is context dependent, but in our case fitness is defined in relation with the accuracy with which a given chromosome (set of spectral features F_i) predicts the physical parameters.

The implementation of the GA comprises the following steps:

Stage 1: Definition of the population of potential features (chromosomes).

Stage 2: Each chromosome in the population is evaluated by its ability to predict the physical paramters of each star in the dataset (fitness function).

Stage 3: Chromosome selection, when a chromosome has a score higher than a predefined value.

Stage 4: The population of chromosomes is replicated. Chromosomes with higher fitness scores will generate more numerous offspring.

Stage 5: The genetic information contained in the replicated parent chromosomes is combined through genetic crossover. Two randomly selected parent chromosomes are used to create two new chromosomes.

Stage 6: Mutations are then introduced in the chromosome randomly. These mutations produce new genes used in chromosomes. Steps 5 and 6 are applied over the chromosomes established at Step 4.

Stage 7: This process is repeated from Stage 2 until a target accuracy is achieved or the maximum number of iterations is attained.

We test features (both for the numerator and denominator of Eq. 1) that comprise ten consecutive spectral bins of the spectrum. These features may overlap by as much as 5 consecutive bins (which in practice implies that we define the first feature as the spectral chunk between wavelength bins $i = 1$ and $i = 10$, the second feature between bins $i = 6$ and $i = 15$, the third feature between bins $i = 11$ and $i = 20$, etc). We do not test for feature ratios that overlap in wavelength.

An obvious conceptual limitation of a univariate approach (considering chromosomes that code a single predictive feature) would be the lack of consideration that features work in the context of interconnected pathways and, therefore, it is their behavior as a group that has to be evaluated in terms of the predictive accuracy. Multivariate selection methods thus seem more suitable for the analysis of the regressors since variables are tested in combination to identify interactions between features. In this work we define a chromosome as a set of ten individual genes, and each gene codes a pair of non-overlapping spectral bands, the ratio of which is used as predictor of the physical parameters according to (1).

The population size was set to 8000 individuals and the maximum number of accepted iterations set to 4000. We produced three randomly started populations so as to provide enough initial variety. The crossover and mutation probabilities were set to 0.85 and 0.35 respectively. Elitism was fixed to 0.15 **No hemos mencionado elitismo; hay que mencionarlo y definirlo antes**.

Feature fitness was defined in terms of the Akaike Information Criterion (AIC) for linearity between the potential feature against the physical parameter.

The most frequent and efficient features were selected as candidates to predictive variables of the physical parameters in regression models. We used a binary codification of the chromosomes and a parallel implementation of the GA in a farm of fifteen computers per physical parameter. The used architecture was the CESVIMA power7 HPC which involve processors with 8 cores and four threads per core, running by 3,3 GHz and with 32Gb of RAM each.

The GA procedure provides us with a large collection of chromosomes. Although these are all potential solutions of the problem, it is not immediately clear which one should be selected for the final regression model.

It is relevant to say that when genes of a chromosome induced as ratio of two gaussian variables have extremely wide wings if any of the features is centered arounf zero, the fitness criteria removes it as a cadidate feature as it is not able to explain the physical parameter. Therefore, the single regression model should, to some extent, be descriptive of the population. The simpler strategy would be to use the frequency of the chromosome in the population as criterion for inclusion in a forward selection strategy. However we prefered to select the features based on their highest fitness as it enhances the value of the direct contribution to explain the physical parameter. As we have accepted that compex interaction between individual features are possible, it was selected a fixed number of features allowing both, enough room for developping those complex dependency relationships but to keep the complexity of the comming regression models under control. Therefore, it was selected ten as the suitable number of features being considered per physical parameter.

Once the GA has generated a proposal set of features for predicting each of the physical parameters, the next step consists in training the regression model to predict them based in these features. The GA generates a large set of proposals that they are ordered by fitness and frequency of participation in the final poluation. Features arising less than five times were discarded as they can not be relevant by themselves and just arising randomly by combination with other stronger features.

In order to assess the performance of the regression models, we compare their predictions with i) values of the physical parameters from the literature (when available); ii) the predictions from the popular $minimum\chi^2$ distance to spectra in the BT-Settl library. In the case of features proposed by Cesetti et al. (2013), such ranges were directly provided by their paper and the GA based selection of features was, therefore, not needed.

It is worthy to said that because of the impact on the emission spectrum for all the pohysical parameters are the same, several strategies were implemented to verify to what end using hierarchical knowledge becomes helpful to the modelling procedure. Therefore, extensive set of trials have been conducted to derive gravity and metallicity features with and without temperature knowledge. Models were trained to determine the impact of such temperature hierarchical contribution and it was found that, no matter of the model considered, the features determined plus the estimated star temperature outperform the same model without the temperature knowledge by about 20% in case of gravity models and 12% in case of metallicity models. Based on those results, GA features for gravity and netallicity were reinforced with the estimated star temperature as an additional feature.

2.3. Models considered.

For the models to be built, the same strategy was used for all the three physical parameters ($T_{eff}, log(g), met$) and it was to use regression techniques over the selected set of features. As a classical regression problem several modelling techniques, with specific selection of adequate parameters per method when required, were considered:

- Bagging with Multiadaptative Spline Regression Models *(MARS)*.
- Random Forest Regression Models *(RF)*.
- K Nearest Neighbours *(KNN)*.
- Generalized Boosted Regression Models *(GBM)*.
- Support Vector Regression with Gaussian Kernel *(SVR)*.
- MLP Neural Networks *(NNET)*.
- Kernel Partial Least Squares Regression *(KPLS)*.
- Rule Regression Models *(RR)*.

Including here a sufficient description of each and every regression model that we trained would render the manuscript excessively lengthy but interested readers can find additional information in Baraud (2002); Geman et al. (1992); Elith et al. (2008); Meyer et al. (2003); Svetnik et al. (2003). Suffice it to say that each one of them can be thought of as a parametric model that predicts one physical parameter from an input vector. The input vector can be the full normalised spectrum, the ICA lower-dimensional representation of the full spectrum, or the spectral features selected by Cesetti et al. (2013) or by the GA. The model parameters are infered (using algorithms that differ from one regression model to the other) from a set of examples. This set of examples (spectra of stars for which we know the physical parameters) is called the training set, and the process by which the model parameters are determined from the training set, is called training of the model. In the next paragraph we give minimal details of each regression model trained, and references for the interested reader.

As every type of model has its own set of tunable parameters as well as its own training procedure, the authors have selected a common wrapper named caret (short for C_lassification A_nd RE_gression T_raining) from Jed Wing et al. (2016). This wrapper enables a common interface as well as the use of the same set of samples for the adopted five-folder cross-validation training technique used. The adopted strategy for selection of parameters was to search throughout the grid of values. Therefore, generally speaking the adopted procedure for learning models is presented in Algorithm 2.1.

Algorithm 2.1: MODEL LEARNING($DataSet, ParRanges$)

$\mathcal{S}_{ModelParameters} \leftarrow ParRanges$
$\mathcal{S}_{DataFolders} \leftarrow Preprocess(DataSet)$
for each $x \in \mathcal{S}_{ModelParameters}$

$$\textbf{do} \begin{cases} \textbf{for each } z \in \mathcal{S}_{DataFolders} \\ \quad \textbf{do} \begin{cases} HDS(z) \leftarrow \text{Hold-out specific samples} \\ Model(z) \leftarrow Fits(\mathcal{S}_{DataFolders} \setminus HDS(z)) \\ Perf(z) \leftarrow Predicts(Model(z), HDS(z)) \end{cases} \\ Perf(x) \leftarrow Average(Perf(z)) \quad \forall z \in \mathcal{S}_{DataFolders} \end{cases}$$

$OPS \leftarrow argmax(Perf(y)) \quad \forall y \in \mathcal{S}_{Modelparameters}$
$Model \leftarrow Fits(DataSet, POS)$

As mentioned above, the training set was constructed from the BT Settl library of stellar spectra. The interested reader may find different approaches in the literature to the problem of finding an optimal set of training examples. Ness et al. (2015) for example prefer to use real observed spectra rather than synthetic libraries to create a generative model in which the individual spectral fluxes are modelled as second degree polynomials with the physical parameters as arguments. The real observed spectra have physical parameters taken from the literature, which in turn are almost always inferred using synthetic spectral libraries. In our opinion, this approach does not solve the dependence of the predicted parameters on the necessarily imperfect synthetic libraries, but has the advantage that the relative frequencies of examples in the training set represents better the biases naturally encountered in surveys than the uniform sampling of parameter space found in synthetic libraries. Recently, Heiter et al. (2015) have started a program to compile a set of stars with accurate physical parameter determinations infered independently of spectroscopic measurements and atmospheric models (as much as possible). Unfortunately, this ambitious program only contains 34 stars of spectral types F, G, and K. In the M regime we find similar approaches in **?**, **?**, and **?**, where the atmospheric parameters are derived using interferometric measurements of stellar radii. Again, this only amounts to a very small number of examples and a very sparse sampling of the parameters space.

We believe that all efforts to compile training sets of stars with accurate, homogeneous, and reliable physical parameters derived independently of spectroscopic measurements are valuable not only because they allow for the improvement of the stellar atmospheric models but also because they help increase the reliability of the regression models by making them independent of these same atmospheric models. But until these training sets with sufficient and homogeneous sampling of the parameter space are available, we turn to the use of synthetic libraries.

3. Physical parameters of the IRTF collection of spectra.

3.1. Spectral bands

During the preprocessing stage (described in Sect. 2) the spectral resolution of the BT-Settl library was degraded to the IRTF resolution (R 2000) by convolving with a Gaussian. Then, the spectra were trimmed to produce valid segments between 8145.92 and 24106.85Å, which is the spectral range common to all M stars in the IRTF library. Finally, all spectra were divided by the total integrated flux in this range in order to factor out the stellar distance.

3.1.1. Spectral features for the estimation of effective temperatures.

The application of the GAs to the selection of features for the prediction of effective temperature from noiseless spectra with the IRTF wavelength range and resolution results in the features included in Table 13. Features are ordered by the fitness value (the AIC) and we only consider features that are present in at least 5 sets.

TBD by Luis: interpret the features.

When noise is added to the BT-Settl spectra, we obtain

Tables 13 and **??** show a very wide variety of features with very few repetitions. Only spectral features 4, 5, 6, and 9 in the SNR=50 experiment are found too in the SNR=∞ and SNR=10 feature sets (albeit with different continuum definitions). This reinforces the impression that the information useful for the estimation of the effective temperatures is spread over the entire IRTF spectrum.

A closer look at features 4, 5, 6, and 9

As a reference, Table 3 lists the features found by Cesetti et al. (2013) using sensitivity maps.

λ_1	λ_2	$\lambda_{cont;1}$	$\lambda_{cont;2}$
9225.86	9283.94	9736.02	9793.96
11106.48	11193.56	13497.81	13613.95
13438.08	13554.08	12006.54	12093.56
9135.89	9193.91	10002.04	9999.92
9555.93	9614.06	12951.62	13038.62
9466.08	9523.82	13137.94	13253.96
11196.56	11283.24	12546.46	12633.49
8566.08	8624.07	13258.32	13374.32
8266.11	8324.03	9856.06	9913.91
8235.96	8294.04	12366.32	12453.33

Table 1: Features selected by the GA for predicting T_{eff} using BT_Settl noiseless synthetic spectra.

3.1.2. Spectral features for surface gravity estimation.

For gravity estimation (on a logarithmic scale), the GA search procedure produces the features presented in Tables 15 and 18 for the pure synthetic signal and signal-to-noise ratios of 10 and 50, respectively.

3.1.3. Spectral features for metallicity estimation.

Finally, the best features found by the GA for metallicity estimation are listed in Table 17 for the noiseless BT-Settl spectra, and in Table **??** for signal-to-noise ratios equal to 10 and 50.

When signal-to-noise ratios equal to 10 and 50 are considered, the GA finds the features listed in Table **??**.

3.2. Regression models

In the following, we will summarise the results obtained for the IRTF data set. We deal with the different physical paramenters in separate Sections. We start by reporting the Root Mean/Median Square Errors (RMSE/RMDSE) with respect to the parameters gathered from the literature by Cesetti et al. (2013) and included in their Table 3.

3.2.1. Effective temperature models

Table 8 summarises the RMSE/RMDSE for the complete set of models: the minimum χ^2 estimate based on the full spectrum (χ^2), the projection pursuit regression based on the ICA components (PPR-ICA) and models trained on the spectral features proposed by the GA (GA-RF, GA-GBM, GA-SVR, GA-NNET, GA-MARS, GA-KPLS, GA-RR). For each model, we report the RMSE/RMDSE obtained for several noise levels of the training sets. SNR=∞ corresponds to noiseless spectra. In the GA-cases, the model is trained with the spectral features found by the Genetic Algorithms when applied to BT-Settl spectra of the corresponding SNR.

Make sure we always have Rule-Regression models everywhere or discuss why not.

Table 8 shows that the performance of classifiers based on the full spectrum (or in a compressed version in the form of ICA components) and the best classifier based on features derived from limited spectral bands is equivalent. The bartlett test shows that the variances are homogeneous with a Bartlett's K-squared of 8.5 with 2 degrees of freedom and a p-value of0.01426. The Flinger-Killen test shows that homokedascity is verified at the p=0.005886 level. Finally, the F-ANOVA test clearly shows that there is no significant difference between models. Thus, we conclude that the quality of features from the two approaches are equivalent in predictive performance. The difference between the performances of the best classifier ($GA - KNN$; best on average over SNR), the minimum χ^2 classifier, and the $PPR - ICA$ classifiers are not statistically significant. The bartlett test shows that the variances are homogeneous with a Bartlett's K-squared of 8.5 with 2 degrees of freedom and a p-value of0.01426. The Flinger-Killen test shows that homokedascity is verified at the p=0.005886 level. Finally, the F-ANOVA test clearly shows that

SNR = 10				SNR=50			
λ_1	λ_2	$\lambda_{cont;1}$	$\lambda_{cont;2}$	λ_1	λ_2	$\lambda_{cont;1}$	$\lambda_{cont;2}$
8235.96	8294.04	12681.62	12768.68	8145.92	8204.03	12636.48	12723.57
8505.89	8563.93	13378.12	13494.13	8895.95	8953.95	11331.57	11418.65
9376.07	9433.92	12951.62	13038.62	8176.03	8234.13	10611.36	10698.46
8145.92	8204.03	12366.32	12453.33	13438.08	13554.08	12546.46	12633.49
9195.86	9253.93	9135.89	9193.92	8235.96	8294.04	11961.44	12048.54
9585.95	9644.12	10002.04	9999.92	9376.07	9433.92	10002.04	9999.92
8385.99	8443.94	11826.48	11913.28	9406.09	9463.96	13258.32	13374.32
9135.89	9193.92	9225.86	9283.94	9346.13	9403.92	13086.46	13194.09
13618.20	13734.15	11376.63	11463.51	11106.48	11193.56	13438.08	13554.08
9105.87	9163.91	8865.98	8923.94	9255.86	9314.01	8865.98	8923.94

Table 2: Recommended features and continuum bandpasses for predicting T_{eff} by using BT_Settl with SNR= 10 and 50.

Index	Element	Signal_from	Signal_To	Cont1_From	Cont1_To	Cont2_From	Cont2_To
Pa1	H I	8461	8474	8474	8484	8563	8577
Ca1	Ca II	8484	8513	8474	8484	8563	8577
Ca2	Ca II	8522	8562	8474	8484	8563	8577
Pa2	H I	8577	8619	8563	8577	8619	8642
Ca3	Ca II	8642	8682	8619	8642	8700	8725
Pa3	H I	8730	8772	8700	8725	8776	8792
Mg	Mg I	8802	8811	8776	8792	8815	8850
Pa4	H I	8850	8890	8815	8850	8890	8900
Pa5	H I	9000	9030	8983	8998	9040	9050
FeClTi	Fe I, Cl I, Ti I	9080	9100	9040	9050	9125	9135
Pa6	H I	9217	9255	9152	9165	9265	9275
Fe1	Fe I	1.9297	1.9327	1.9220	1.9260	2.0030	2.0100
Brδ	H I (n=4)	1.9425	1.9470	1.9220	1.9260	2.0030	2.0100
Ca1	Ca I	1.9500	1.9526	1.9220	1.9260	2.0030	2.0100
Fe23	Fe I	1.9583	1.9656	1.9220	1.9260	2.0030	2.0100
Si	Si I	1.9708	1.9748	1.9220	1.9260	2.0030	2.0100
Ca2	Ca I	1.9769	1.9795	1.9220	1.9260	2.0030	2.0100
Ca3	Ca I	1.9847	1.9881	1.9220	1.9260	2.0030	2.0100
Ca4	Ca I	1.9917	1.9943	1.9220	1.9260	2.0030	2.0100
Mg1	Mg I	2.1040	2.1110	2.1000	2.1040	2.1110	2.1150
Brγ	H I (n=4)	2.1639	2.1686	2.0907	2.0951	2.2873	2.2900
Na_d	Na I	2.2000	2.2140	2.1934	2.1996	2.2150	2.2190
FeA	Fe I	2.2250	2.2299	2.2133	2.2176	2.2437	2.2479
FeB	Fe I	2.2368	2.2414	2.2133	2.2176	2.2437	2.2479
Ca_d	Ca I	2.2594	2.2700	2.2516	2.2590	2.2716	2.2888
Mg2	Mg I	2.2795	2.2845	2.2700	2.2720	2.2850	2.2874
^{12}CO	^{12}CO(2,0)	2.2910	2.3070	2.2516	2.2590	2.2716	2.2888

Table 3: Features and continuum bandpasses defined in Cesetti et al. (2013) as relevant for the estimation of the effective temperature in bands I and K.

there is no significant difference between models. Thus, we conclude that the quality of features from the two approaches are equivalent in predictive performance. In any case, it is evident that the RMSE is significantly above the grid spacing in temperature. We interpret the small differences as an indication that there is as much information spread over the entire spectrum shape as can be distilled from a few spectral bands.

The comparison with the effective temperatures compiled by Cesetti et al. (2013) shows however some significant differences across models when evaluated not by the RMSE/RMDSE, but by the average bias (see Table 9).

In general, all classifiers tend to predict lower effective temperatures than those in the literature except in the noiseless scenario. The models trained with noiseless spectra tend to overestimate $T_{\rm eff}$, suggesting that the optimal SNR is between SNR=50 and ∞. The minimum-χ^2 approach and the GA-KNN model systematically underestimate $T_{\rm eff}$ for all SNR regimes. This shared behaviour is not surprising since minimum χ^2 is a single nearest neighbour method applied in the space of the entire spectrum as opposed to the space selected features.

We have found in previous studies that, at least for input spaces constructed from ICA compressions of the spectra, it is not necessary to adapt the training set SNR to match exactly that of the prediction set. On the contrary, we find that two regimes are sufficient to obtain acceptable results. The two regimes are separated at SNR=10. The model trained with SNR=50 spectra gives close to optimal results for spectra with SNRs above 10,

λ_1	λ_2	$\lambda_{cont;1}$	$\lambda_{cont;2}$
10245.88	10304.02	11241.29	11328.54
8415.91	8473.96	11511.51	11598.51
12906.56	12993.61	13041.48	13133.82
8716.00	8773.99	10425.90	10484.13
8805.93	8863.97	12816.72	12903.73
10126.02	10183.93	13086.46	13194.09
8176.03	8234.13	10971.57	11058.46
8626.02	8683.99	10746.43	10833.57
8536.03	8594.06	10215.95	10274.10
12951.62	13038.62	11196.56	11283.24

Table 4: Recommended features and continuum bandpasses for predicting log(g) obtained using noiseless BT_Settl spectra.

while below that limit the same situation holds for the model trained with SNR=10 spectra. **Cite paper by Ana.**

Figure **??** shows the correlation between the $T_{\rm eff}$ estimates of the best (in the RMDSE sense) regression models and the effective temperatures in Table 3 of Cesetti et al. (2013).

We then compare the predicted effective temperatures with the spectral types listed in the IRTF spectral library in order to increase the size of the validation sample beyond the 57 cases with estimated temperatures in Table 3 of Cesetti et al. (2013). We converted the spectral types into effective temperatures using the calibration of Stephens et al. (2009). Both the RMSE and RMDSE were used to evaluate the prediction accuracy (see Table **??**).

Faltan las tablas y figuras.

We have trained the same non linear regression models discussed above using the features suggested by Cesetti et al. (2013). The performace of the models based on these features are included in Table 20.

How do you explain that the best SNR=10 model has the poorest performances for SNR=50 or ∞?

From the comparison of Tables 20 and 8 we can draw the following conclusions:

- the RMSE for SNR=10 and 50 is equivalent for the regression models trained on GA features and those recommended in Cesetti et al. (2013);
- however, the RMDSE is significantly higher in the case of the latter features for all SNR values.
- in the unrealistic case of noiseless spectra, the features proposed by Cesetti et al. (2013) produce RMSE and RMDSE significantly worse than the GA features.

As a summary, we believe that the features found by the GA are to be prefered to the ones proposed by Cesetti et al. (2013).

3.2.2. Surface gravity models

For the validation of our models, we only have 10 literature values of the surface gravity available in Table 3 of Cesetti et al. (2013). Unfortunately, this is too small a number to draw significant conclusions on the comparison of methodologies from external data. Hence, we are left only with plausibility arguments for the selection of models. In this Section we will use $\log(T_{\rm eff}) -- \log(g)$ diagram comparisons to select the most plausible model results. An important difference with respect to the models discussed above is that we use the $T_{\rm eff}$ estimated in the previous stage as input of our models. **do we have some hint whether this was beneficial, neutral or detrimental?**

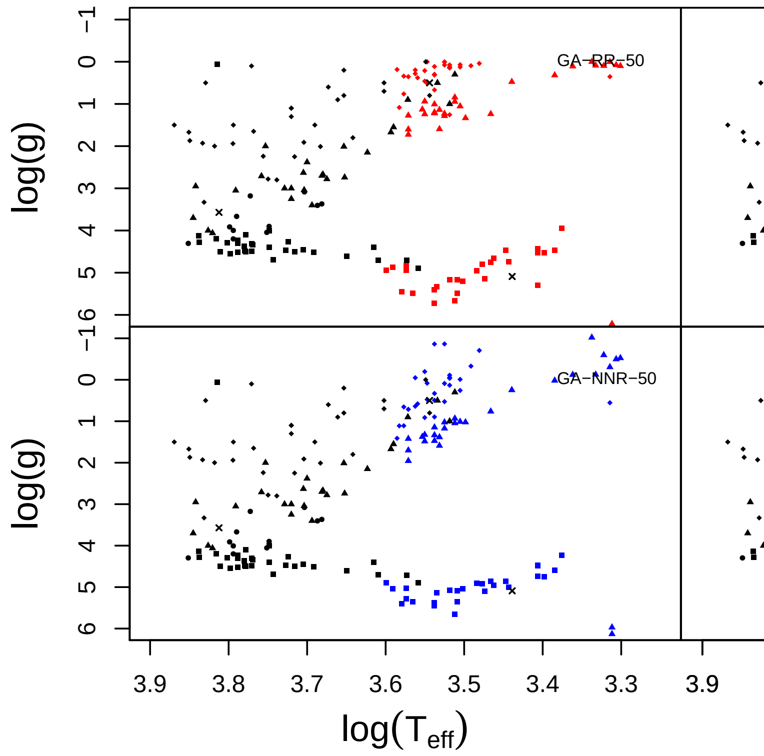

Fig. 1: $\log(T_{eff})$–$\log(g)$ diagrams produced by the GA-KNN (SNR=∞) effective temperatures and gravities derived with the GA-RR (SNR=50), GA-PLS (SNR=50), GA-NNR (SNR=50), and χ^2 models (clockwise, starting from the top left plot).

Table 21 shows the RMSE and RMDSE of the log(g) regression models for the same SNR regimes discussed for the estimation of $T_{\rm eff}$.

Again, as in the case of the effective temperatures, the differences between the various models as measured by the RMSE or RMDSE are not statistically significant. This is not surprising given the extraordinarily small sample of gravity measurements gathered from the literature and used as reference for the computation of errors. However, we can evaluate the models according to plausibility arguments relative to the distribution of the model predictions in $T_{\rm eff}$–log(g) diagrams. Figure 1 shows this distribution for four models selected based on these plausibility criteria: GA-RR, GA-PLS, GA-KNN (the three of them for SNR=50), and PPR-ICA (clockwise, starting at the top left corner).

Is χ^2 much worse now for the weak parameter logg? I guess no. This needs be discussed

Discuss these plots in the case of Cesetti features.

3.2.3. Metallicity models

Finally, the same machine learning models are trained to infer the metallicity, again considering the effective temperature as an input feature as in the log(g) regression models. Table 22 shows the RMSE and RMDSE obtained for each regression model for the only seven M-type stars in Table 3 of Cesetti et al. (2013) with a metallicity estimate in the literature.

SNR = 10				SNR=50			
λ_1	λ_2	$\lambda_{cont;1}$	$\lambda_{cont;2}$	λ_1	λ_2	$\lambda_{cont;1}$	$\lambda_{cont;2}$
8176.03	8234.13	9165.87	9223.91	11151.63	11238.46	13086.46	13194.09
10485.99	10563.41	10002.04	9999.92	8385.99	8443.94	13618.20	13734.14
8656.09	8714.047	10926.46	11013.60	8176.03	8234.13	11241.29	11328.54
9525.89	9584.059	10002.04	9999.92	8536.03	8594.06	13041.48	13133.82
8205.98	8263.967	13041.48	13133.82	12771.70	12858.73	10306.03	10363.88
10275.97	10333.96	11376.63	11463.51	13378.12	13494.13	10002.04	9999.92
10306.03	10363.88	11151.63	11238.46	8626.02	8683.99	10926.46	11013.60
9165.87	9223.91	8385.99	8443.94	9826.05	9883.91	10006.07	10064.01
9645.82	9704.16	13137.94	13253.96	10521.56	10608.46	11736.71	11823.49
8326.00	8383.94	12726.69	12813.71	8205.98	8263.96	9796.09	9853.94

Table 5: Recommended features and continuum bandpasses for predicting $log(g)$ obtained using BT_Settl with SNR= 10 and 50.

λ_1	λ_2	$\lambda_{cont;1}$	$\lambda_{cont;2}$
12096.68	12183.66	12051.50	12096.68
9525.89	9584.05	12321.33	12408.32
8205.98	8263.96	10126.02	10183.93
8566.08	8624.07	12276.52	12363.34
11196.56	11283.24	11151.63	11196.56
11151.639	11238.46	11466.35	11553.33
9555.93	9614.06	8205.98	8263.96
11016.62	11103.37	10791.44	10878.40
9766.16	9823.94	12681.62	12768.68
9942.14	9999.92	9555.93	9614.06

Table 6: Feature and Continuum bandpasses selected for predicting metallicity using noiseless BT_Settl spectra.

Compare the 7 or 6 values available. Discuss. χ^2 is the most popular method by far. We compare predictions of machine learning methods with minimum chi-squared. We first do histogram plots. Then, the same logTeff-logg plots as above but with metallicity coded in colour.

Figure 2 shows the relationships between metalicity predicted by global espectrum estimation and GA feature based estimation against the real values provided by **?** can be observed.

Include table as annex with metallicities from the literature.

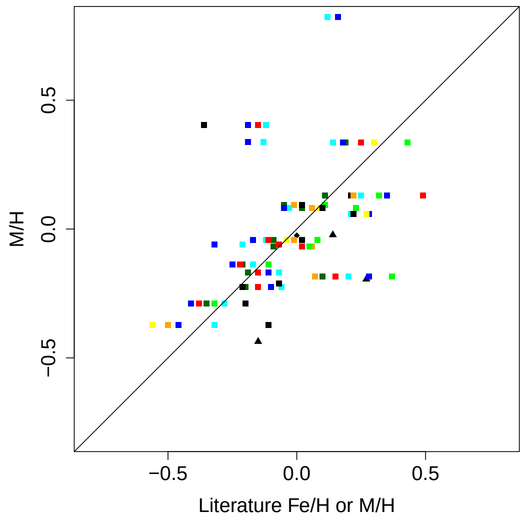

Fig. 2: Comparison between metallicity estimates from the literature and predictions from the PPR-ICA (SNR=10) model. **TBC: Include description of symbols and colours.**

SNR = 10				SNR=50			
λ_1	λ_2	$\lambda_{cont;1}$	$\lambda_{cont;2}$	λ_1	λ_2	$\lambda_{cont;1}$	$\lambda_{cont;2}$
8235.96	8294.04	11331.57	11418.65	9255.86	9314.01	13197.94	13313.92
9376.07	9433.92	10566.33	10653.62	8385.99	8443.94	9376.07	9433.92
10306.03	10363.88	9942.14	9999.92	8716.00	8773.99	9585.95	9644.12
11286.42	11373.45	11241.29	11286.42	8235.96	8294.04	13086.46	13194.09
9676.00	9734.02	13086.46	13194.09	9676.00	9734.02	10791.44	10878.40
8775.95	8833.94	8415.91	8473.96	8415.91	8473.96	12411.34	12498.41
12411.34	12498.41	10245.88	10304.02	8446.03	8503.94	9406.09	9463.96
8476.01	8534.03	12276.52	12363.34	8205.98	8263.96	8955.88	9013.95
12636.48	12723.57	12051.50	12138.72	8985.93	9043.98	12186.62	12273.48
8415.91	8473.96	13618.20	13734.14	9015.98	9073.98	11241.29	11328.54

Table 7: Feature and Continuum bandpasses selected for predicting metallicity using noisy BT_Settl spectra with signal-to-noise ratios equal to 10 and 50.

	$SNR = 10$		$SNR = 50$		$SNR = \infty$	
$Regression Models$	$RMSE$	$RMDSE$	$RMSE$	$RMDSE$	$RMSE$	$RMDSE$
χ^2	232	**100**	235	120	232	**100**
PPR-ICA	242	128	242	99	280	162
GA-RF	308	183	248	136	**167**	135
GA-GBM	287	160	248	149	233	113
GA-SVR	**221**	122	281	151	299	160
GA-NNET	283	192	264	114	326	212
GA-KNN	238	120	**232**	137	219	**100**
GA-MARS	253	113	254	**95**	226	133
GA-KPLS	275	120	300	119	387	218

Table 8: Cross-validation RMSE and RMDSE for the various regression models that predict T_{eff} (K).

	$SNR = 10$	$SNR = 50$	$SNR = \infty$
χ^2	-77	-87	-85
$ICA + ppr$	-104	-55	-130
GA-RR	-102	-39	170
GA-RF	-173	-127	-5
GA-GBM	-141	-109	32
GA-SVR	-58	-3	92
GA-NNET	-147	-36	39
GA-KNN	-76	-110	-67
GA-MARS	-57	-88	98
GA-KPLS	-120	-4	214

Table 9: Average bias in the T_{eff} (K) estimates computed with respect to the reference values in Table 3 of Cesetti et al. (2013).

	$SNR = 10$		$SNR = 50$		$SNR = \infty$	
RegressionModels	*RMSE*	*RMDSE*	*RMSE*	*RMDSE*	*RMSE*	*RMDSE*
CS-RF	234	180	**264**	218	**321**	265
CS-GBM	**232**	195	268	254	325	246
CS-SVR	268	227	293	257	432	364
CS-NNET	357	255	357	**204**	552	435
CS-KNN	249	172	293	256	327	**230**
CS-KPLS	351	**162**	856	456	1086	535

Table 10: Regression model performance based on the features proposed by Cesetti et al. (2013)

	$SNR = 10$		$SNR = 50$		$SNR = \infty$	
RegressionModels	*RMSE*	*RMDSE*	*RMSE*	*RMDSE*	*RMSE*	*RMDSE*
χ^2	0.82	0.45	0.93	0.61	3.5	3.48
$PPR - ICA$	0.54	0.48	**0.3**	**0.17**	0.72	0.57
GA-RF	0.64	**0.38**	0.77	0.72	0.53	0.39
GA-GBM	**0.48**	0.45	0.61	0.47	0.49	0.41
GA-SVR	0.66	0.40	0.63	0.58	**0.46**	**0.21**
GA-NNET	0.78	0.61	0.47	0.44	1.2	0.97
GA-MARS	0.84	0.57	0.54	0.37	0.99	0.76
GA-KNN	1.23	0.83	1.39	1.44	1.60	1.32
GA-KPLS	0.99	0.99	0.51	0.49	0.96	0.77
GA-RR	0.74	0.57	0.50	0.47	0.57	0.41

Table 11: RMSE and RMDSE for the various $\log(g)$ regression models [dex].

	$SNR = 10$		$SNR = 50$		$SNR = \infty$	
RegressionModels	*RMSE*	*RMDSE*	*RMSE*	*RMDSE*	*RMSE*	*RMDSE*
χ^2	0.76	0.22	0.36	0.18	0.36	0.18
$PPR - ICA$	0.24	**0.13**	0.31	0.22	0.43	0.27
$GA - RF$	0.33	0.25	0.73	0.41	0.61	0.36
$GA - GBM$	0.27	0.19	0.70	0.52	0.63	0.35
$GA - SVR$	0.33	0.22	0.45	0.32	0.92	0.89
$GA - NNET$	0.37	0.30	0.33	0.37	0.95	0.81
$GA - KNN$	0.69	0.55	0.23	**0.15**	0.21	**0.15**
$GA - MARS$	0.36	0.16	0.49	0.41	0.83	0.85
$GA - RR$	0.31	0.17	0.30	0.24	0.78	0.23

Table 12: RMSE and RMDSE for the various regression models predicting metallicity [dex].

4. Physical parameters of the IPAC collection of spectra.

4.1. Spectral bands selected

As for the IRTF spectra, the spectral resolution of the BT-Settl library was degraded to match the average resolution of IPAC spectra in the Dwarf Archives[1]. **What is the average resolution?**. Then, the spectra were trimmed to produce valid segments between *** and *** Å, which is the spectral range common to all M stars in the archive. Finally, all spectra were divided by the total integrated flux in this range in order to factor out the stellar distance.

λ_1	λ_2	$\lambda_{cont;1}$	$\lambda_{cont;2}$
7062	7094.4	7314	7346.4
7116	7148.4	7782	7814.4
7134	7166.4	7872	7904.4
6900	6932.4	7764	7796.4
7170	7202.4	7890	7922.4
7080	7112.4	7926	7958.4
7188	7220.4	7548	7580.4
7800	7832.4	7962	7994.4
6990	7022.4	7008	7040.4
7026	7058.4	6990	7022.4

Table 13: Spectral features and continuum bandpasses selected by the GA for predicting $T_{\rm eff}$ using noiseless BT_Settl spectra.

There is little hope *a priori* for reasonable accuracies with regression models that predict the surface gravity and metallicity from such wavelength-limited, low/intermediate resolution spectra. Anyhow, we provide the results obtained applying the same methodology as in Section **??** to show the limitations.

4.1.1. Spectral features for the estimation of effective temperatures.

The application of the GA to the selection of features for the prediction of effective temperature from noiseless spectra within the IPAC wavelength range and resolution, results in the features included in Table 13. Features are ordered by the fitness value (the AIC)**and we only consider features that are present in at least 5 sets**.

TBD by Luis: interpret the features.

When noise is added to the BT-Settl spectra, we obtain the following features depending on the SNR of the spectra:

Tables 15 and 18 show the spectral features selected by the GA for noiseless BT-Settl spectra and the same spectra with SNR=10 and 50, respectively.

Finally, the best features found by the GA for the estimation of the metallicity are listed in Table 17 for the noiseless BT-Settl spectra, and in Table **??** for signal-to-noise ratios equal to 10 and 50.

4.2. Regression models

In the following, we will summarise the results obtained for the IPAC data set. We deal with the different physical parameters in separate Sections. We start by reporting the cross validation Root Mean Square Errors (RMSE) and Root Median Square Error (RMDSE) for the five-fold cross-validation strategy, and we subsequently discuss the accuracy of the predictions with respect to literature values where available.

[1] http://spider.ipac.caltech.edu/staff/davy/ARCHIVE/index.shtml

SNR = 10				SNR=50			
λ_1	λ_2	$\lambda_{cont;1}$	$\lambda_{cont;2}$	λ_1	λ_2	$\lambda_{cont;1}$	$\lambda_{cont;2}$
7692	7724.4 6936	6968.4	7062	7094.4	7296	7328.4	
6990	7022.4 7998	8030.4	7026	7058.4	7044	7076.4	
6900	6932.4 7548	7580.4	7080	7112.4	7926	7958.4	
7854	7886.4 7710	7742.4	6900	6932.4	7548	7580.4	
7116	7148.4 7908	7940.4	7134	7166.4	7836	7868.4	
7278	7310.4 7926	7958.4	7296	7328.4	7962	7994.4	
7152	7184.4 7746	7778.4	6936	6968.4	7728	7760.4	
7134	7166.4 7764	7796.4	6972	7004.4	6900	6932.4	
6918	6950.4 6900	6932.4	6990	7022.4	7944	7976.4	
7224	7256.4 7962	7994.4	6918	6950.4	7782	7814.4	

Table 14: Spectral features and continuum bandpasses selected by the GA for predicting T_{eff} using BT_Settl spectra with SNR=10 and 50.

λ_1	λ_2	$\lambda_{cont;1}$	$\lambda_{cont;2}$
7134	7166.4	7044	7076.4
6954	6986.4	7152	7184.4
7512	7544.4	7890	7922.4
7062	7094.4	7224	7256.4
6936	6968.4	7854	7886.4
6900	6932.4	7746	7778.4
6918	6950.4	7800	7832.4
7008	7040.4	7134	7166.4
7872	7904.4	7008	7040.4
7962	7994.4	7980	8012.4

Table 15: Spectral features and continuum bandpasses selected by the GA for predicting $\log(g)$ using noiseless BT_Settl spectra.

We train the same regression models applied to the GA selected features, to the features selected in Cesetti et al. (2013), again learning from BT-Settl spectra of various SNRs and predicting over the IPAC set. A summary of the results can be found in Table 20, where we use CS- to indicate that the model was trained using the features by Cesetti et al. (2013).

For SNR=10, the GA best models (GA-KPLS in RMDSE or GA-RF in RMSE) outperform the best CS model (GA-GBM). For SNR=50 the situation depends on the figure-of-merit used to compare the classifiers: in RMSE the best model is CS-GBM while in RMDSE GA-GBM outperforms all CS-models. Finally, for the unrealistic case of noiseless spectra, Table 20 shows an overwhelming degradation of the prediction accuracy from CS-features. **Overfitting?** But even in the only case where the CS features outperform those selected by the GA, the performance is below the one achieved by the minimum-χ^2 approach.

The relationship between the GA predicted Temperature and the one measured by Rojas-Ayala can be found in the Figure 4

4.2.1. Effective temperature models

Table 19 summarises the RMSE/RMDSE for the complete set of models: the minimum χ^2 estimate based on the full spectrum (χ^2), the projection pursuit regression based on the ICA components (PPR-ICA) and some models trained on the spectral features proposed by the GA (GA-RF, GA-GBM, GA-SVR, GA-NNET, GA-MARS, GA-KPLS). For each model, we report the RMSE/RMDSE obtained for several noise levels of the training sets.

Again, as in the IRTF case, we see that the compression of the spectra results in a performance degradation. Figure **??**

Explain the spt-teff calibration used.

Biases?

We do have problems with the prediction at low temperatures when trained with SNR= 10 or 50.

Include plot with 4 models

Having shown that the feature selection with GAs degrades the performance of regression models, one can wonder whether a different feature selection procedure would produce better results. In particular, we investigate the possibility that the features proposed by Cesetti et al. (2013) result in a performance equal to or even better than the one achieved with χ^2.

4.2.2. Surface gravity models

As in the IRTF exercise, we attempt to select features for surface gravity estimation from BT-Settl spectra using GAs despite the much lower spectral resolution and smaller wavelength coverage of the IPAC spectra. Since there is no substantive compilation of surface gravities that we could cross match with the IPAC list of M stars in the Dwarf Archive, we are left with the same plausibility arguments used in the IRTF study which are based on the $\log(T_{\rm eff})$–$\log(g)$ diagram.

We again use the effective temperatures as input of the regression models. Table 21 shows the cross-validation RMSE and RMDSE for the same set of regression models used throughout this article. It shows that the GA-RF model outperforms all other in all SNR regimes, giving a consistent RMDSE of 1.0 dex. Obviously, this is barely enough for classification in luminosity classes.

Figure 5 shows the $\log(T_{\rm eff})$–$\log(g)$ diagram for the GA-RF and GA-NNET models. The latter is, in our opinion, the one that shows the diagram that is most with Fig. **??** in this work, and Fig. 1 in Cesetti et al. (2013). All GA- models predict decreasing surface gravities for main sequence stars below $\log(T_{\rm eff}$ = 3.6. GA-NNET predicts main sequence val-

SNR = 10				SNR=50			
λ_1	λ_2	$\lambda_{cont;1}$	$\lambda_{cont;2}$	λ_1	λ_2	$\lambda_{cont;1}$	$\lambda_{cont;2}$
6990	7022.4	6918	6950.4	6918	6950.4	6936	6968.4
6900	6932.4	7278	7310.4	6936	6968.4	7836	7868.4
7062	7094.4	7242	7274.4	7656	7688.4	7890	7922.4
7692	7724.4	7008	7040.4	6900	6932.4	7872	7904.4
7656	7688.4	7998	8030.4	7008	7040.4	7044	7076.4
6936	6968.4	7836	7868.4	7512	7544.4	7656	7688.4
7206	7238.4	7062	7094.4	7440	7472.4	7332	7364.4
7512	7544.4	7926	7958.4	7800	7832.4	7692	7724.4
7764	7796.4	7710	7742.4	7404	7436.4	7548	7580.4
7404	7436.4	7548	7580.4	7080	7112.4	7152	7184.4

Table 16: Spectral features and continuum bandpasses selected by the GA for predicting log(g) using BT_Settl spectra of SNR=10 and 50.

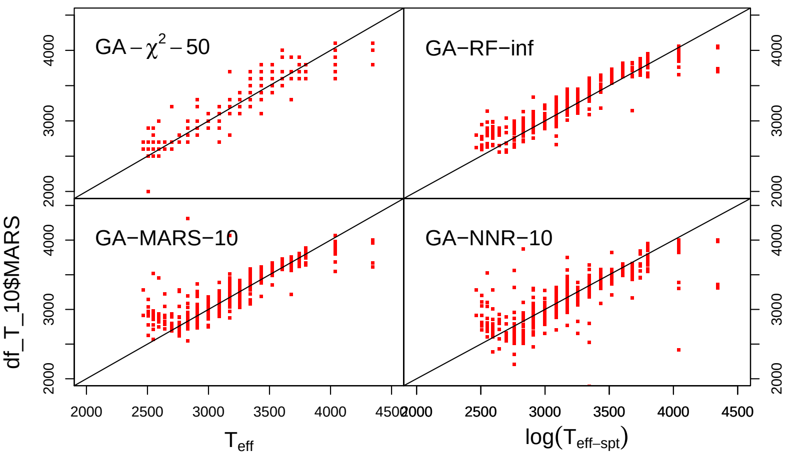

Fig. 3: Comparison between Temperature estimations from Theoretical Temperature in x axis and the modeled ICA based estimation at SNR=∞ on y-axis

ues between $4 \leq \log(g) \leq 6$, while luminosity classes III-I appear clearly separated from the main sequence with values concentrated in the 4-6 range except for the hottest cases with $\log(T_{rmeff} > 3.55$. The GA-RF results, despite showing the best cross-validation errors (RMSE/RMDSE), result in unrealistic main sequence gravities. We interpret this as the result of overfitting to the training examples.

Right now, it appears that feature selected models are worse than χ^2, judging only from the 10 available estimates (mail sent to jbmere). If so, the conclusion is clear: we should not do feature selection at these resolutions. This is useful as Cesseti et al do not question the utility of feature selection. For the IRTF (which is the dataset used by Cesseti et al), we should check this: are the models with feature selection better than χ^2?

4.2.3. Metallicity models

Finally, the same analysis is performed for metalicities, again using the previously inferred temperature as a fixed input feature.

Table 22 shows a summary of the cross-validation performance of the different models.

λ_1	λ_2	$\lambda_{cont;1}$	$\lambda_{cont;2}$
7188	7220.4	7854	7886.4
7080	7112.4	7926	7958.4
7116	7148.4	7098	7130.4
7422	7454.4	7836	7868.4
7350	7382.4	7998	8030.4
7224	7256.4	7818	7850.4
7710	7742.4	7062	7094.4
7476	7508.4	7944	7976.4
7134	7166.4	7584	7616.4
7836	7868.4	7278	7310.4

Table 17: Spectral features and continuum bandpasses selected by the GA for predicting metallicity using noiseless BT_Settl spectra.

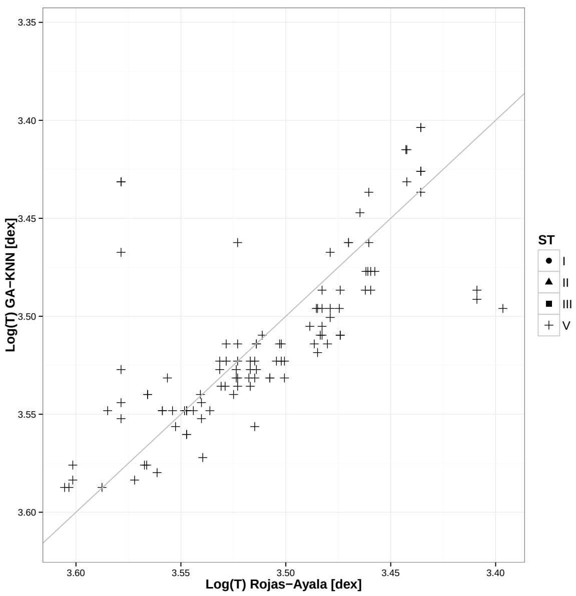

Fig. 4: Relationship between $log(T) from Rojas - Ayala$ in the x axis and $log(T)$ as predicted by KNN with SNR=10

SNR = 10				SNR=50			
λ_1	λ_2	$\lambda_{cont;1}$	$\lambda_{cont;2}$	λ_1	λ_2	$\lambda_{cont;1}$	$\lambda_{cont;2}$
7692	7724.4	7026	7058.4	7098	7130.4	7926	7958.4
6900	6932.4	7008	7040.4	7188	7220.4	7962	7994.4
7350	7382.4	7908	7940.4	7368	7400.4	7980	8012.4
6918	6950.4	6900	6932.4	7116	7148.4	7872	7904.4
7098	7130.4	7314	7346.4	7062	7094.4	7206	7238.4
7440	7472.4	7872	7904.4	7584	7616.4	7170	7202.4
7134	7166.4	7962	7994.4	6936	6968.4	6918	6950.4
7368	7400.4	7926	7958.4	7692	7724.4	7890	7922.4
7080	7112.4	7044	7076.4	7134	7166.4	7548	7580.4
7044	7076.4	7980	8012.4	7494	7526.4	7998	8030.4

Table 18: Spectral features and continuum bandpasses selected by the GA for predicting metallicities using BT_Settl spectra of SNR=10 and 50.

	$SNR = 10$		$SNR = 50$		$SNR = \infty$	
$RegressionModels$	$RMSE$	$RMDSE$	$RMSE$	$RMDSE$	$RMSE$	$RMDSE$
χ^2	**147**	79	**121**	**56**	**126**	**57**
$PPR - ICA$	188	126	164	95	191	130
GA-RF	160	97	196	103	145	94
GA-GBM	175	105	225	99	185	94
GA-SVR	203	112	285	106	368	154
GA-NNET	221	84	313	111	395	202
GA-KNN	183	119	193	109	224	110
GA-MARS	222	76	361	103	374	157
GA-KPLS	227	**72**	331	123	409	208

Table 19: RMSE and RMDSE for the various regression models that predict T_{eff} (K).

	$SNR = 10$		$SNR = 50$		$SNR = \infty$	
$RegressionModels$	$RMSE$	$RMDSE$	$RMSE$	$RMDSE$	$RMSE$	$RMDSE$
CS-RF	203	140	243	**121**	**306**	**172**
CS-GBM	**188**	**120**	**161**	138	337	222
CS-SVR	197	135	379	194	840	688
CS-NNET	207	135	514	296	719	489
CS-MARS	252	124	789	186	3464	784
CS-KNN	235	158	246	137	314	175
CS-KPLS	250	201	741	361	2247	1424
CS-RR	211	128	400	239	828	774

Table 20: Performances of regression models trained on the features selected by Cesetti et al. (2013) applied to BT-Settl spectra.

	$SNR = 10$		$SNR = 50$		$SNR = \infty$	
RegressionModels	*RMSE*	*RMDSE*	*RMSE*	*RMDSE*	*RMSE*	*RMDSE*
χ^2	2.2	1.6	2.2	1.4	2.2	1.6
PPR-ICA	2.1	1.8	1.8	1.4	4.3	4.2
GA-RF	**1.3**	**1.0**	**1.6**	**1.1**	**1.4**	**0.9**
GA-GBM	1.6	1.1	1.7	1.4	1.7	1.2
GA-SVR	2.0	1.8	2.1	1.9	2.3	1.6
GA-NNET	2.0	1.8	2.2	1.9	3.2	2.8
GA-MARS	1.8	1.5	2.0	1.7	2.0	1.5
GA-KNN	2.0	1.5	2.2	1.7	1.7	1.2
GA-KPLS	1.8	1.4	2.0	1.7	2.7	2.3
GA-RR	2.0	1.8	2.1	1.8	3.7	3.2

Table 21: RMSE and RMDSE for the various regression models predicting $Log(G)$ [dex].

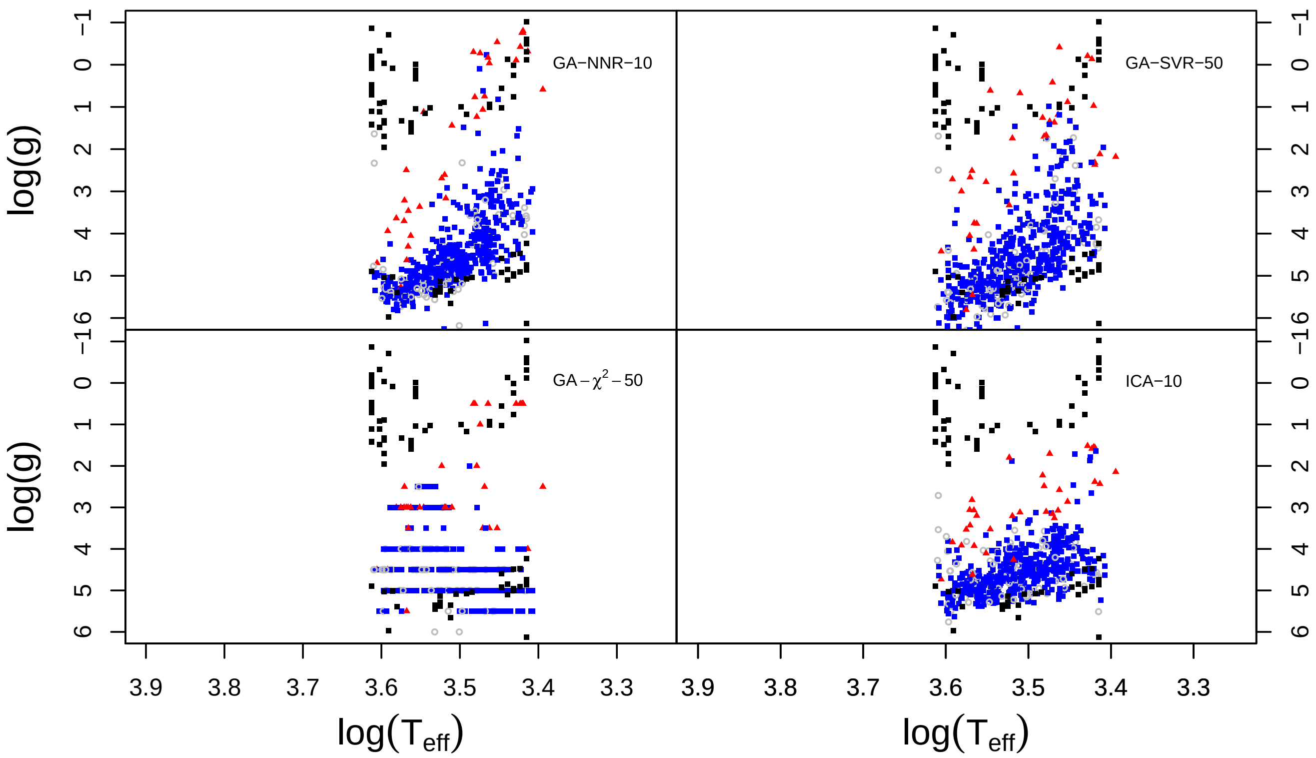

Fig. 5: Relationship between $log(T)$ (x axis) and $log(g)$ (y axis) for several regression models.

	$SNR = 10$		$SNR = 50$		$SNR = \infty$	
$RegressionModels$	$RMSE$	$RMDSE$	$RMSE$	$RMDSE$	$RMSE$	$RMDSE$
$\chi^2 BTSettl$	0.55	0.27	0.51	0.29	0.43	0.29
$ICA + ppr$	0.48	0.27	0.70	0.39	0.85	0.71
rf	0.55	0.38	0.71	0.61	0.23	0.16
gbm	0.64	0.43	0.87	0.84	0.31	0.23
svr	0.46	0.26	0.57	0.44	3.38	2.33
$nnet$	0.52	0.45	0.66	0.54	2.03	1.88
knn	0.37	0.28	0.99	0.78	0.56	0.32
$mars + bagging$	0.71	0.47	0.80	0.69	1.15	0.68
pls	0.67	0.61	0.63	0.55	1.17	1.02
$RuleRegression$	0.47	0.29	0.50	0.36	1.18	1.18

Table 22: RMSE and RMDSE for the various regression models predicting Met [dex].

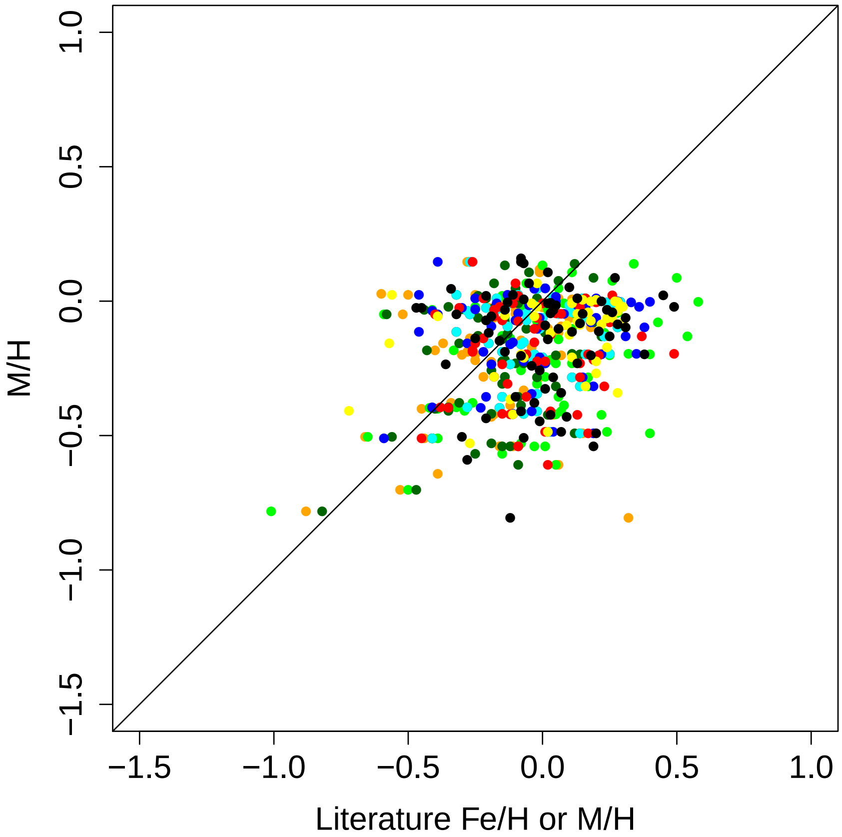

Fig. 6: Relationship between $TfromNevesIII$ in the x axis and Met as predicted by Regression Rules with SNR=10

5. Conclusions

Acknowledgements. This research has benefitted from the M, L, T, and Y dwarf compendium housed at DwarfArchives.org. The authors also thanks to the Spanish Ministery for Economy and Innovation because of the support obtained through the project with ID: AyA2011-24052. IRTF library provided by the University of Hawaii under Cooperative Agreement no. NNX-08AE38A with the National Aeronautics and Space Administration, Science Mission Directorate, Planetary Astronomy Program.

References

Allard, F., Homeier, D., Freytag, B., et al. 2013, Memorie della Societa Astronomica Italiana Supplementi, 24, 128
Baraud, Y. 2002, ESAIM: Probability and Statistics, 6, 127
Cesetti, M., Pizzella, A., Ivanov, V. D., et al. 2013, A&A, 549, A129
Elith, J., Leathwick, J. R., & Hastie, T. 2008, Journal of Animal Ecology, 77, 802
from Jed Wing, M. K. C., Weston, S., Williams, A., et al. 2016, r package version 6.0-71
Geman, S., Bienenstock, E., & Doursat, R. 1992, Neural computation, 4, 1
Goldberg, D. E. et al. 1989, Genetic algorithms in search, optimization, and machine learning, Vol. 412 (Addison-wesley Reading Menlo Park)
González, A., Bello, A., Ordieres-Meré, J., & Sarro, L. submitted, MNRAS
Heiter, U., Jofré, P., Gustafsson, B., et al. 2015, ArXiv e-prints
Holland, J. H. 1975, Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence. (U Michigan Press)
Meyer, D., Leisch, F., & Hornik, K. 2003, Neurocomputing, 55, 169
Ness, M., Hogg, D. W., Rix, H.-W., Ho, A. Y. Q., & Zasowski, G. 2015, ApJ, 808, 16
R Core Team. 2013, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria
Stephens, D. C., Leggett, S. K., Cushing, M. C., et al. 2009, ApJ, 702, 154
Svetnik, V., Liaw, A., Tong, C., et al. 2003, Journal of chemical information and computer sciences, 43, 1947