The performance of GSP-Photwith SVM and ILIUM

prepared by: C. Liu and C. A.L. Bailer-Jones
approved by:
reference: GAIA-C8-TN-MPIA-CHL-005-1
issue: 1
revision: 0
date: 2011-07-27
status: Issued

Abstract

This Note is about the performance assessment of the latest verion of GSP-Phot.

Document History

Issue	Revision	Date	Author	Comment
1	0	2010-07-27	CHL	Creation

Contents

1 Introduction **4**

1.1 Applicable Documents 4

1.2 References 4

1.3 Definitions 5

1.4 Acronyms 5

2 An overview of the algorithms **6**

3 The simulated data **7**

4 Results for cycle5 data **8**

4.1 SVM results 9

4.2 ILIUM results 12

4.3 More analysis on SVM and ILIUM results 14

5 Results for cycle7 Phoenix library **16**

5.1 SVM results 17

5.2 ILIUM results 19

5.3 More analysis on SVM and ILIUM results 22

6 The degeneracy between $T_{\rm eff}$and A_0 23

7 Comparison between BaSeL and Phoenix 24

8 Apply BaSeL trained algorithms to MARCS data 24

9 ILIUM trained with different density of the discrete grid 25

10 Other tests 26

10.1 Test of the binned BP/RP 26

10.2 The influence of the radial velocities 29

11 Conclusion and discussions 29

1 Introduction

GSP-Photwill estimate stellar astrophysical parameters for all stellar objects with Gaia BP/RP spectra. Several algorithms will be applied in GSP-Photin order to deal with all kind of objects in a better way. CHL-001 has been published the first performance report of GSP-Phot, which was based on four mixed stellar libraries (MARCS, BaSeL, A and OB) from cycle 3, meaning that the slightly difference between synthetic libraries are not taken into account. And both training and test dataset in that TN are only fixed at $G = 15$ mag, which gives much accurate but unrealistic estimates. Finally the tests made in CHL-001 is only with SVM. In this TN, we assess the performance of two algorithms (SVM and ILIUM) been implemented in the current version of GSP-Phot(v10). The training and test datasets are firstly constrained in the same synthetic library and then from different libraries (BaSeL and MARCS). The former test will avoid the noises due to the difference of the synthetic libraries and the later test will focus on the effect of the difference of the various libraries. All test dataset are evenly distributed from $G = 6.8$ mag $G = 20$ mag. There is a third algorithm, q-method, is also included in the current version pof GSP-Phot. The performance of q-method will be reported in another TN soon.

1.1 Applicable Documents

CHL-001
CHL-002
CHL-004

1.2 References

[CBJ-042], Bailer-Jones, C., 2009, *ILIUM. An iterative local interpolation method for parameter estimation*,
GAIA-C8-TN-MPIA-CBJ-042,
URL http://www.rssd.esa.int/llink/livelink/open/2872359

[CBJ-043], Bailer-Jones, C., 2009, *Application of ILIUM to the estimation of the Teff-[Fe/H] pair from BP/RP*,
GAIA-C8-TN-MPIA-CBJ-043,
URL http://www.rssd.esa.int/llink/livelink/open/2873690

[CBJ-046], Bailer-Jones, C., 2009, *ILIUM III. Further observations, tests and developments*,
GAIA-C8-TN-MPIA-CBJ-046,
URL http://www.rssd.esa.int/llink/livelink/open/2890802

[CBJ-048], Bailer-Jones, C., 2009, *ILIUM IV. Three-dimensional forward model and demonstration of a strong and ubiquitous Teff-Av degeneracy in BP/RP*,

GAIA-C8-TN-MPIA-CBJ-048,
URL http://www.rssd.esa.int/llink/livelink/open/2918589

[CBJ-049], Bailer-Jones, C., 2010, *Probabilistic combination of stellar astrophysical parameter estimates based on spectra, astrometry, photometry and the HR Diagram*,
GAIA-C8-TN-MPIA-CBJ-049,
URL http://www.rssd.esa.int/llink/livelink/open/2970453

[CBJ-050], Bailer-Jones, C., 2010, *ILIUM V. Further degeneracy mapping, application to the semi-empirical library and the use of the forward model and MCMC to build likelihood maps*,
GAIA-C8-TN-MPIA-CBJ-050,
URL http://www.rssd.esa.int/llink/livelink/open/2968731

Bailer-Jones, C., 2010, The ilium forward modelling algorithm for multivariate parameter estimation and its application to derive stellar parameters from gaia spectrophotometry., MNRAS, 403, 96

[CHL-001], Liu, C., 2009, *GSP-Phot performance as a function of APs*,
GAIA-C8-TN-MPIA-CHL-001,
URL http://www.rssd.esa.int/llink/livelink/open/2906427

[CHL-002], Liu, C., 2011, *GSP-Phot Cycle 8 STR*,
GAIA-C8-TR-MPIA-CHL-002,
URL http://www.rssd.esa.int/llink/livelink/open/3057873

[CHL-004], Liu, C., Janotto, A.M., Bailer-Jones, C., et al., 2011, *CU8 Scientific Algorithms Software Design Description*,
GAIA-C8-SP-MPIA-CHL-004,
URL http://www.rssd.esa.int/llink/livelink/open/3062016

[AV-007], Vallenari, A., Sordo, R., 2008, *Interpolation of synthetic stellar spectral libraries in Cycles 2 t o 5*,
GAIA-C8-TN-OAPD-AV-007,
URL http://www.rssd.esa.int/llink/livelink/open/2863940

Vapnik, V., 1995, *The Nature of Statistical Learning Theory*, Springer verlag, New York

1.3 Definitions

1.4 Acronyms

The following is a complete list of acronyms used in this document.

Acronym	Description

AP	Astrophysical Parameter
BP	Bed Photometer
HR	Hertzsprung-Russell (diagram)
ILIUM	An iterative local interpolation method
RP	Red Photometer
SVM	Support Vector Machine
SVR	Support vector regression
TN	Technical Note

2 An overview of the algorithms

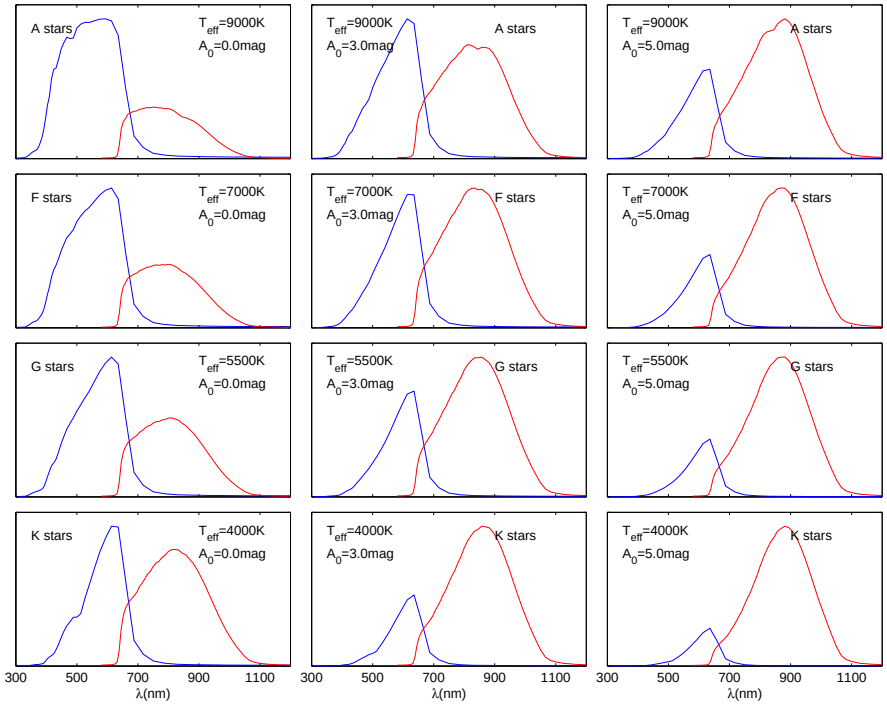

FIGURE 1: Nine samples of the Gaia BP/RP spectra showing how the strong APs $A_0$and $T_{\rm eff}$change the shape of the spectra. The effective temperature changes from 4000 K to 9000 K from bottom to top, while the interstellar extinction varies from 0 to 5 from left to right.

GSP-Photis a software package in Agsis to parametrize the stellar physical parameters, e.g., effective temperature, metallicity, surface gravity etc. Three algorithms are implemented in GSP-Photand used in the performance assessment, SVM, ILIUM and q-method. SVM (Vapnik 1995), which is broadly used in classification, is used as a regression method here in GSP-Phot. ILIUM is a forward model method with special treatment of the strong and weak APs (Bailer-Jones 2010; CBJ-042; CBJ-043; CBJ-046; CBJ-048; CBJ-050). Q-method calculates the posterior probability density function of A_0(extinction parameter) and $T_{\rm eff}$ given the ob-

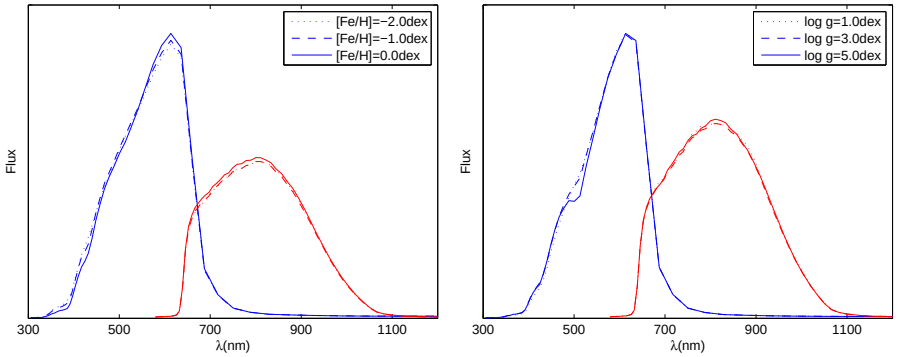

FIGURE 2: Samples of Gaia BP/RP spectra showing how the weak APs log$g$and [Fe/H]change the shape of the spectra given $T_{\rm eff}$and A_0. In the left panel the APs are fixed at $T_{\rm eff}$= 5000 K, A_0= 0 mag and log$g$=4 dex, the [Fe/H]changes from -2 to 0 dex. In the right panel the APs are fixed at $T_{\rm eff}$= 5000 K, A_0= 0 mag and [Fe/H]=0 dex, while log$g$changes from 1 to 5 dex.

served BP/RP spectra and the parallaxes (CBJ-049; Bailer-Jones 2011). In this TN we will concentrate on the results from the first two algorithms.

3 The simulated data

Three synthetic libraries are used in this work. Cycle5 BaSeL library, covering from 3000 K to 15,000 K in $T_{\rm eff}$, and Cycle7 Phoenix libraries, covering from 3000 K to 10,000 K, are used as both training and test datasets, separately. Cycle5 MARCS, which covers $T_{\rm eff}$from 4000 K to 8000 K and hence overlaps with BaSeL , is only used as a test dataset for the test of inter-library performance.

The random grids of synthetic simulated dataset are split out into two part, one for the training of SVM and the other for testing all three algorithms. The nominal grids of the simulated dataset are only used as the discrete AP grid to build forward models in ILIUM and q-method.

The value of G magnitude for the SVM training dataset are fixed at 9, 15, 17, 18, 18.5, 19, 19.5 and 20 mag, while the ones for the test dataset are evenly distributed between 6.8 and 20 mag. The nominal grid, which is used both in ILIUM and q-method, is fixed at $G = 15$ mag because the forward model will normalize the input spectra to this magnitude.

As demonstrations, figure 1 and 2 show the variety of the BP/RP spectra with strong and weak APs in Phoenix library.

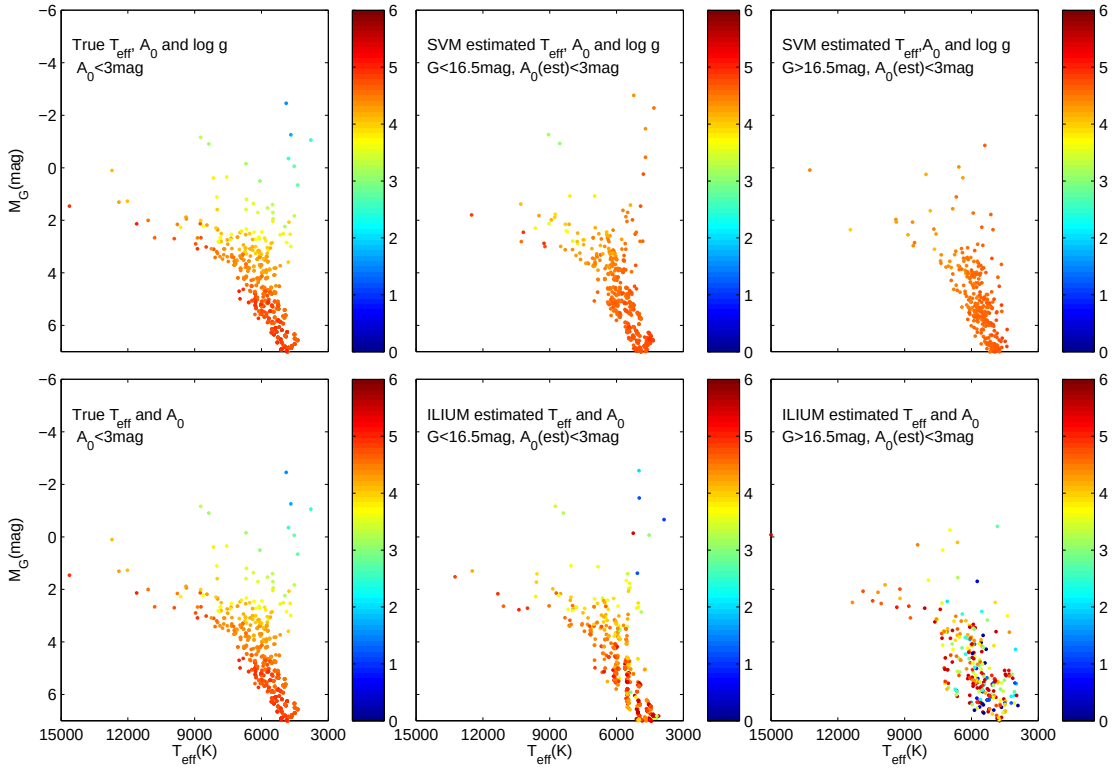

FIGURE 3: Results of the three algorithms in the HR diagram using the BaSeL library as training and test dataset. All the data have $A_0 < 3$ mag. The first column is the HR diagram with true $T_{\rm eff}$, log$g$and A_0values. The color codes the log$g$. The second column shows results for bright samples. The right column shows the results for the faint samples. The first row presents the results from SVR, the second row from ILIUM.

4 Results for cycle5 data

Using 2000 test spectra from the BaSeL library we ran GSP-Photwith 2 algorithms simultaneously. In this section we analyse the results qualitatively and quantitatively.

For a qualitative view of the result we display the estimated APs in the HR diagrams and compare them with the true APs. Figure 3 shows the HR diagram of the true APs of the test dataset (left column) and the HR diagrams of the results of the two algorithms. The test dataset is separated into two groups: the bright sample with $G < 16.5$ mag (the middle column panels in figure 3) and the faint sample with $G > 16.5$ mag (the right column panels in figure 3).

The position of a star in the HR diagram is determined by the estimated $T_{\rm eff}$, A_0, apparent magnitude and parallax. The estimated $T_{\rm eff}$and $A_0$from GSP-Phot, combined with the G magnitude and the parallax measured by Gaia can reconstruct the HR diagram. Figure 3 shows that both SVM and ILIUM estimated $T_{\rm eff}$and A_0reconstruct the HR diagram quite accurately, in particular for the bright sample. The colors in figure 3 indicate the true (left column) and esti-

mated log$g$(middle and right column). It is found that neither SVM or ILIUM gives a realistic estimation of log$g$.

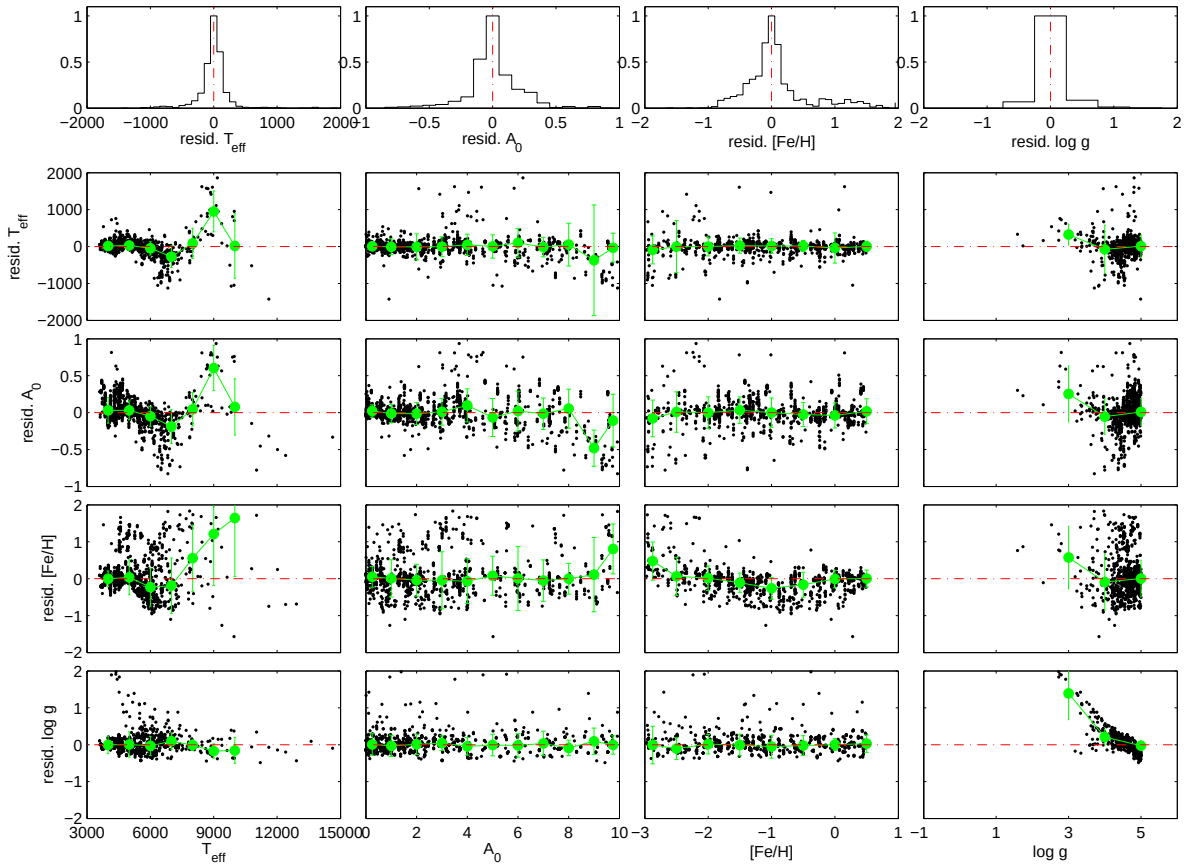

FIGURE 4: SVM AP residuals as functions of the true APs at $G < 16.5$ mag for the BaSeL dataset. The black dots are for individual stars. The big green dots with error bars are the median values.

4.1 SVM results

The performance of the SVM is displayed in table 3 and figure 4 and 5. Table 3 lists the absolute residuals as a measurement of the performance of SVM for both bright and faint samples. The absolute residuals are calculated not only for the whole sample but also for four spectral types: A (7500–10000 K), F(6000–7500 K), G(5250–6000 K) and K(3750–5250 K) stars for better understanding of the performance.

The best estimation of $T_{\rm eff}$is for late type stars, e.g., G and K stars as shown in table 3. For the bright sample the mean absolute residual of $T_{\rm eff}$for A stars is 564 K, 7 times higher than that for K stars. The estimation of $A_0$is similar, for A stars the mean absolute residual of A_0estimates reaches 0.3 mag, while for G/K stars it is only 0.08–0.11 mag. The trends of the performance of $T_{\rm eff}$and $A_0$are similar for faint sample.

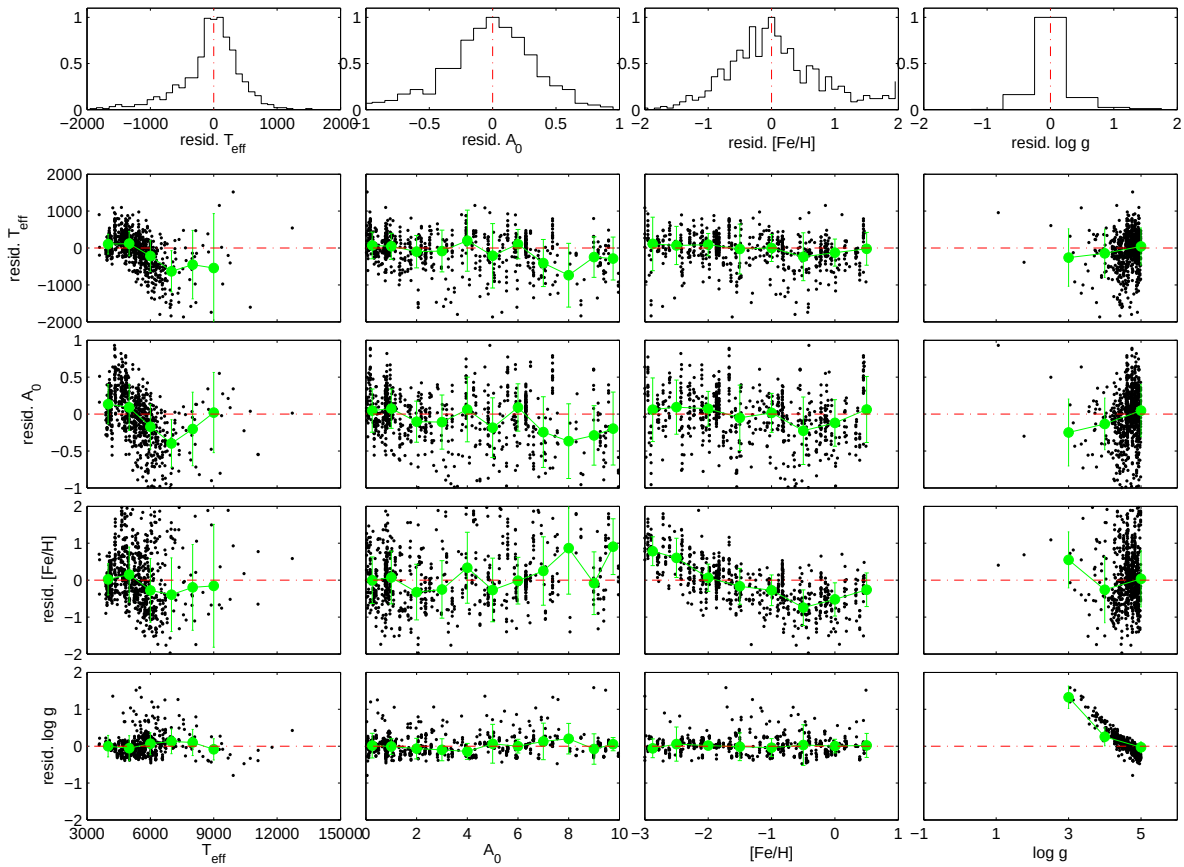

FIGURE 5: SVM AP residuals as functions of the true APs at $G > 16.5$ mag for the BaSeL dataset.

It is known that the [Fe/H]should be easier estimated for F, G and K stars than A stars. This is due to the different sensitivity to the [Fe/H]in spectra for late and early type stars. Table 3 shows the same trend in the third and seventh row, where the absolute residuals of A stars are larger than those of the K stars by a factor of 6 (for the bright sample) and 2 (for the faint sample).

It is also known that the log$g$should be easier estimated for hot stars than cool stars. However, this trend is not found in table 3. As we will see in the next subsection the ILIUM results do show that the log$g$performance of the hot stars is much better than the cool stars, therefore we can rule out that the strange performance SVM estimating log$g$is due to the test dataset. The unoptimized tuning of the SVM model for log$g$might be the reason, although the absolute residuals of log$g$from SVM is quite small. Indeed, the training dataset for SVM does not contain as much giants as dwarfs, meaning that the SVM model trained by this dataset may not give a good estimation of log$g$for giant stars.

Figure 4 shows the distribution of the residuals of $T_{\rm eff}$, A_0, [Fe/H]and log$g$and their correlations with true APs for stars brighter than 16.5 mag. Figure 5 is the counterpart for stars fainter than 16.5 mag.

TABLE 3: Absolute residuals of the estimated APs for the BaSeL library from SVM

AP residual	All stars	A stars	F stars	G stars	K stars
G<16.5 mag					
$<\|dT_{\rm eff}\|>$(K)	160	564	253	89	84
$<\|dA_0\|>$(mag)	0.14	0.30	0.19	0.08	0.11
$<\|d[{\rm Fe/H}]\|>$(dex)	0.36	1.24	0.58	0.35	0.21
$<\|d{\rm logg}\|>$(dex)	0.16	0.16	0.20	0.19	0.13
G>16.5 mag					
$<\|dT_{\rm eff}\|>$(K)	364	939	606	297	249
$<\|dA_0\|>$(mag)	0.28	0.39	0.39	0.24	0.25
$<\|d[{\rm Fe/H}]\|>$(dex)	0.61	1.03	0.92	0.57	0.48
$<\|d{\rm logg}\|>$(dex)	0.19	0.30	0.26	0.23	0.13

For the $T_{\rm eff}$estimation, the second row of figure 4 shows that SVM underestimate the $T_{\rm eff}$at around 6500 K by at most 1000 K, and then overestimate it at around 9000 K by 1000 K. Most of the underestimated values occur for stars with high extinction close to 10 mag, while the overestimated ones correspond to stars with intermediate extinction between 2 and 6 mag. Apart from the two groups of underestimated/overestimated samples $T_{\rm eff}$the most of the $T_{\rm eff}$estimates seem not to be related to the true $A_0$and [Fe/H]. However, it seems that the $T_{\rm eff}$is overestimated by $\sim$300 K for giants, i.e. stars with log$g<3$.

For the A_0estimation, the third row of figure 4 shows that the residual of the estimated $A_0$is similar to those of $T_{\rm eff}$as functions of true APs of $T_{\rm eff}$, [Fe/H], and log$g$. This reflects the intrinsic degeneracy between $T_{\rm eff}$and A_0(CBJ-043). The second column shows that SVM underestimates $A_0$when the true extinction is very high.

For the [Fe/H]estimation, the fourth row of figure 4 shows that the there is not systematic bias for stars colder than 7500 K, while for hot stars it is heavily overestimated. This is due to the different impact of the [Fe/H]in spectra for late and early type stars. The third column panel shows that SVM tends to overestimate [Fe/H]for the metal-poor stars and underestimate it forthe metal-rich ones.

For the estimations of last AP, log$g$, the last row of figure 4 shows that SVM works well for dwarfs (log$g>3$) but overestimates for giants(log$g<3$) by more than 1 dex, which means that it is hard to distinguish giant stars from dwarfs. The steep correlation between the residuals of log$g$and the true values of log$g$implies that the small absolute residuals shown in table 3 are actually the result of over-fitting.

The faint sample has higher dispersion than the bright one, as seen in table 3 and figure 5. For the $T_{\rm eff}$estimation in the second row of the figure, the overestimation for hot stars shown in the bright sample has disappeared. And the overestimation for giant stars in the bright sample is

gone as well. The fourth row of the figure shows similar trend that [Fe/H]estimates is better for cool than hot stars.

4.2 ILIUM results

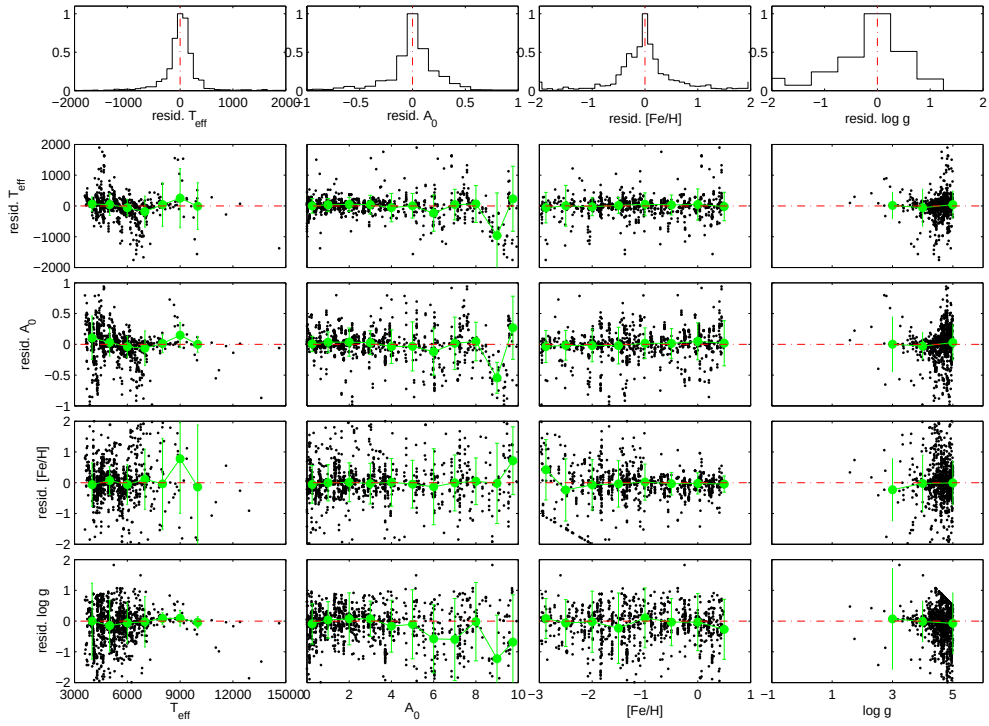

FIGURE 6: ILIUM AP residuals as functions of the true APs at $G < 16.5$ mag for the BaSeL library.

TABLE 4: Results for the BaSeL library with ILIUM

AP residual	All stars	A stars	F stars	G stars	K stars
G< 16.5 mag					
$< \|dT_{\rm eff}\| >$(K)	230	498	314	159	176
$< \|dA_0\| >$(mag)	0.18	0.11	0.17	0.11	0.20
$< \|d[{\rm Fe/H}]\| >$(dex)	0.47	1.27	0.59	0.39	0.36
$< \|d{\rm logg}\| >$(dex)	0.60	0.17	0.43	0.46	0.71
G> 16.5 mag					
$< \|dT_{\rm eff}\| >$(K)	638	915	775	603	573
$< \|dA_0\| >$(mag)	0.44	0.30	0.47	0.40	0.47
$< \|d[{\rm Fe/H}]\| >$(dex)	1.11	1.71	1.25	0.96	1.10
$< \|d{\rm logg}\| >$(dex)	1.46	0.60	1.19	1.36	1.69

The performance of ILIUM for the bright sample is comparable with that with SVM. Table 4

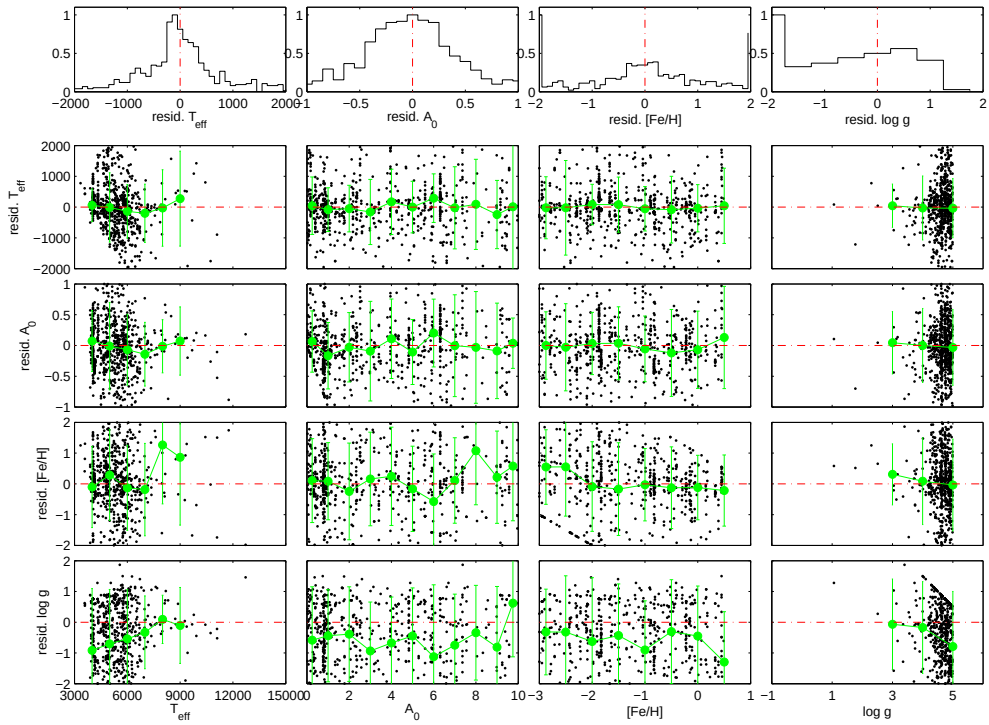

FIGURE 7: ILIUM AP residuals as functions of the true APs at $G > 16.5$ mag for the BaSeL library.

shows that ILIUM results are a little more scattered than the ones of SVM. The total absolute residual of $T_{\rm eff}$from ILIUM is 230 K , $\sim$50% larger than that from SVM (160 K). Similarly, the total absolute residual of $A_0$from ILIUM is only larger than that from SVM by 0.04 mag and the total absolute residual of [Fe/H]from ILIUM is 0.47 dex larger than that from SVM by 0.1 dex. The biggest difference between the two algorithms for the bright sample is log$g$, for which ILIUM is 4 times more disperse than SVM. The best estimates of $T_{\rm eff}$from ILIUM correspond to G/K stars, which is consistent with what we see from SVM. However, the best estimates of $A_0$from ILIUM is A and G stars, which don't show big difference with other spectral types in the results of SVM. Therefore, it is preferable that the performance of $A_0$with ILIUM doesn't show correlation with spectral types.

However, forthe faint sample SVM performs much better than ILIUM. The results for $T_{\rm eff}$, $A_0$and [Fe/H]from ILIUM are worse than those from SVM by a factor of 2. In general, the faint sample has lower signal to noise. SVM works well with low S/N because the training data have S/N in the same level as the test dataset and hence the SVM model has taken the S/N into account. On the other hand, ILIUM uses a noise free spectral grid in AP space. When the test data has very low S/N, the forward model fitting process will return a large uncertainty of the APs.

Figure 6 shows the residuals of APs as functions of the true values for the bright sample. For

the $T_{\rm eff}$, the second row shows that an underestimation occurs around 6000–7000 K with $A_0>$ 8 mag, which is also shown in the SVM results. The [Fe/H]estimates from ILIUM don't show systematic bias for stars with true [Fe/H]> -2 as seen in SVM (see the fourth row), though its residual looks more disperse. The log$g$estimates in ILIUM has a very large absolute residual as shown in table 4. However, the variation of the absolute residuals with spectral type is consistent with the intrinsic characteristics of log$g$, i.e. it is more sensitive in the spectra of hot stars than those of the cool stars.

4.3 More analysis on SVM and ILIUM results

TABLE 5: Completeness and contamination of the spectral type selection based on temperatures estimated with the BaSeL library. All sources have $A_0 < 3$ mag.

Algorithm	G mag		A stars	F stars	G stars	K stars
SVM	< 16.5	completeness	0.679	0.770	0.879	0.929
SVM	< 16.5	contamination	0.077	0.097	0.222	0.023
SVM	> 16.5	completeness	0.517	0.422	0.740	0.922
SVM	> 16.5	contamination	0.250	0.379	0.450	0.149
ILIUM	< 16.5	completeness	0.792	0.682	0.798	0.888
ILIUM	< 16.5	contamination	0.125	0.159	0.323	0.060
ILIUM	> 16.5	completeness	0.621	0.516	0.466	0.839
ILIUM	> 16.5	contamination	0.486	0.609	0.429	0.247

Another analysis of the performance of the two algorithms is to calculate what the completeness and contamination would be when selecting certain spectral type of stars based on SVM/ILIUM estimated $T_{\rm eff}$. The completeness of the selection of a spectral type is defined as the fraction of the stars belonging to the spectral type in the selected sample to the number of that spectral type in the total sample. The contamination is defined as the fraction of the stars in the selected sample that do not belong to the spectral type. The test results are listed in table 5.

We find that the completeness of the spectral type of stars selected based on SVM estimated APs is increasing when the spectral types change from early to late, i.e. the stars get cooler. For A stars with $G < 16.5$ mag the completeness is $\sim$68%, while for K stars it turns out to be $\sim$93%. This gradient is kept also for faint samples. However, the contamination does not show such a trend. For bright samples it suddenly increases to 22% for G stars, while for the other three spectral types it is only 2–9%. The similar jump for G stars is also found in faint sample. Looking back to figure 4 we find that the bigger contamination in G stars is due to the underestimation of the $T_{\rm eff}$for stars with A_0larger than 8 mag.

The completeness based on ILIUM is similar with that based on SVM. However, there is no gradient with spectral types. The contamination based on ILIUM is a little bit higher than that based on SVM, which may be due to the less accurate $T_{\rm eff}$estimation in ILIUM. The highest

contamination occurs again for G stars, same as in the case of SVM results.

5 Results for cycle7 Phoenix library

The two algorithms of GSP-Photwere tested simultaneously with 4000 spectra selected from the Phoenix library. Half are brighter than 16.5 mag and half of them are fainter. We analyse the results analogue to what we did for the BaSeL library.

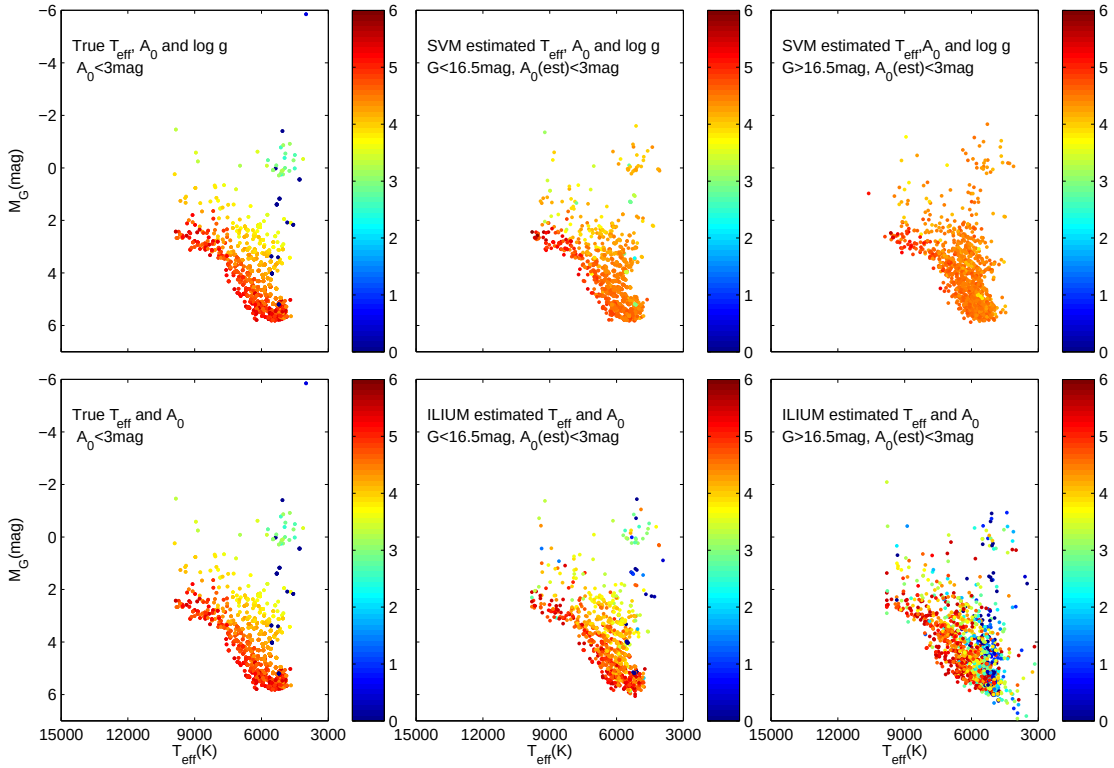

FIGURE 8: Results of the two algorithms in the HR diagram. All the data have $A_0 < 3$ mag. The first column is the HR diagram with true $T_{\rm eff}$, log$g$ and A_0 values. The color codes the log$g$. The second column shows results for the bright sample. The right column shows the results for the faint sample. The first row are the results from SVM, the second row from ILIUM.

Figure 8 shows the HR diagram for the true APs of the test dataset (left column) and the HR diagrams of the results of the two algorithms. The test dataset is separated into two groups: the bright sample with $G < 16.5$ mag (the middle column panels in figure 8) and the faint sample with $G > 16.5$ mag (the right column panels in figure 8).

The position of a star in the HR diagram is determined by the estimated $T_{\rm eff}$, A_0, apparent magnitude and parallax. The estimated $T_{\rm eff}$and $A_0$from GSP-Phot, combined with the G magnitude and the parallax measured by Gaia can reconstruct the HR diagram. Figure 8 shows that both SVM and ILIUM estimated $T_{\rm eff}$and A_0reconstruct the HR diagram well, particularly for bright samples. The colours in figure 8 indicates the true (left column) and estimated log$g$(middle and right column). It is found that ILIUM does reconstruct the log$g$for bright samples For faint

samples ILIUM produces a lot of fake giants and cold stars.

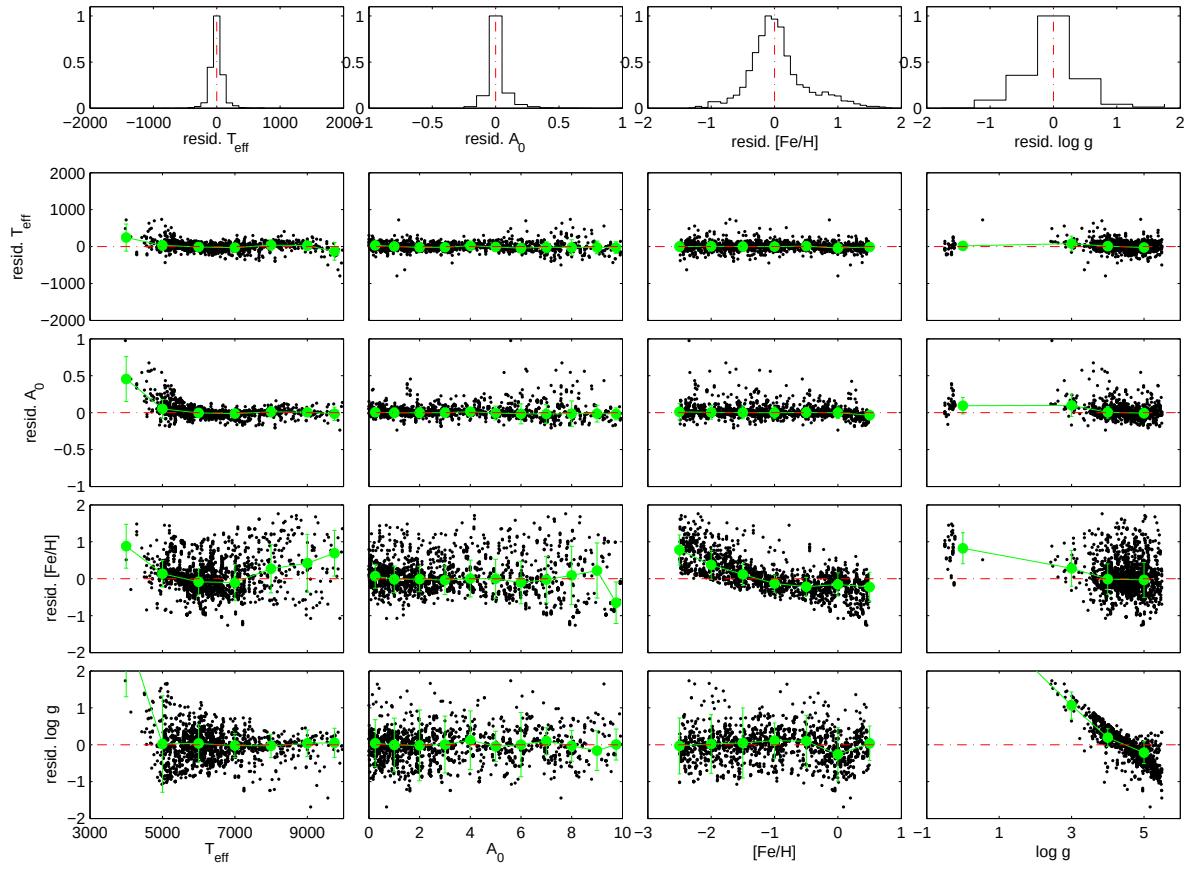

FIGURE 9: SVM AP residuals as functions of the true APs at $G < 16.5$ mag for the Phoenix library.

5.1 SVM results

A quantitative analysis for SVM is presented in table 6 and figure 9 and 10. Figure 9 shows the distribution of the residuals of $T_{\rm eff}$, A_0, [Fe/H]and log$g$as functions of the true APs for stars brighter than 16.5 mag. TFigure 10 is the counterpart for stars fainter than 16.5 mag.

Table 6 shows that the mean absolute residual of the $T_{\rm eff}$estimates is less than 100 K for the bright sample and less than 300 K for the faint sample. With Phoenix library the estimation of $T_{\rm eff}$is particular accurate for F and G stars. The absolute residuals of A_0estimates are larger for K stars than other spectral types by a factor of 2. The intrinsic feature of [Fe/H]estimation that is more sensible for late type than early type stars can be seen in both the bright and faint samples. Additionally the intrinsic trend of log$g$that is easier to be estimated for early type stars have also been seen in the table.

The second row of figure 9 shows that there is a small overestimation for cool stars below

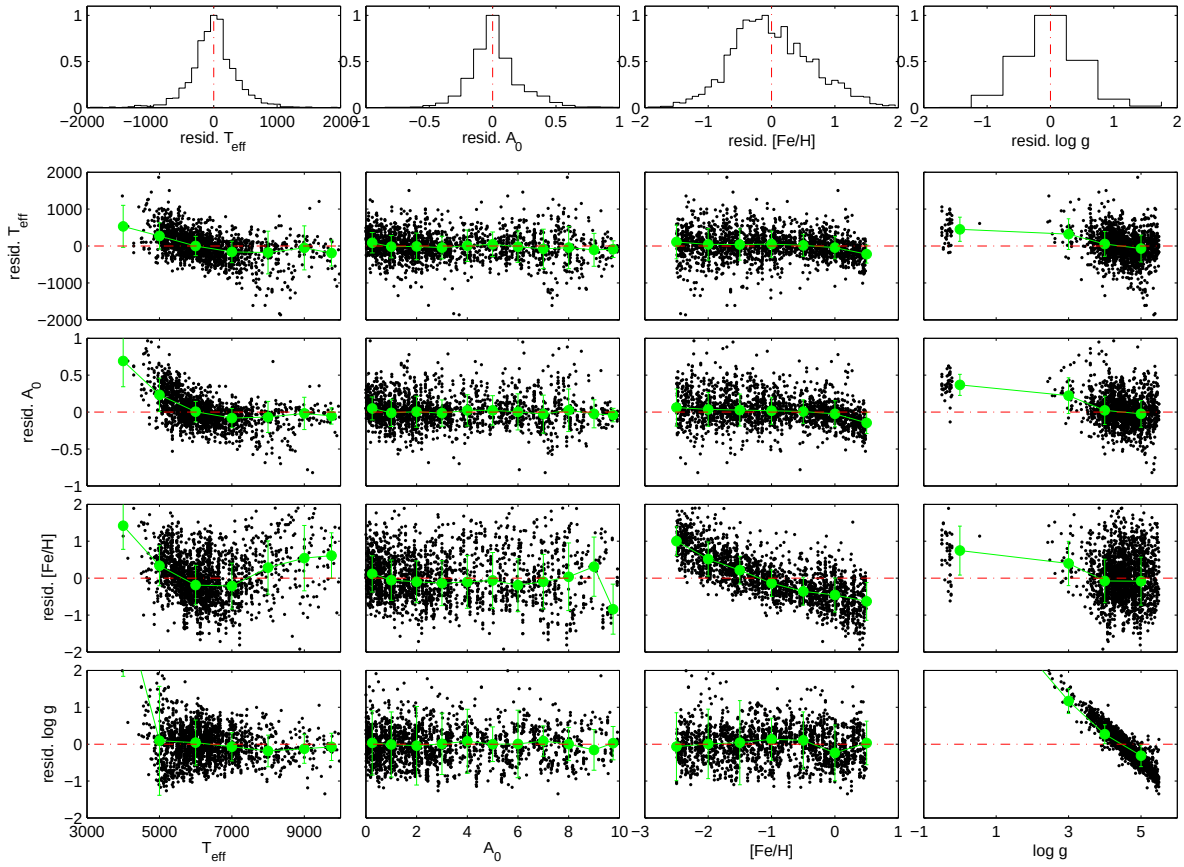

FIGURE 10: SVM AP residuals as functions of the true APs at $G > 16.5$ mag for the Phoenix library.

5000 K and an underestimation for stars hotter than 9000 K in the $T_{\rm eff}$estimation. The residual of $T_{\rm eff}$does not show correlation with the other three APs. These over/under-estimations can explain the fact that the results are more accurate for the A and K stars than for the F and G stars, as seen in table 6.

For the estimation of A_0, the third row shows that there is an overestimation for stars below 5000 K, which suffer extinction higher than 6 mag. For giant stars (log$g$<3) the $A_0$is a little bit overestimated.

For the estimation of [Fe/H], the fourth row shows that there is a systematic bias(an overestimation,) for stars with lower metallicity than -1.5 dex. Also for giant stars (log$g$<3) the metallicity is overestimated by around 1 dex.

Similar to the cycle5 BaSeL result, the residual of log$g$show strong correlation with the true log$g$as shown in the last row. It seems that all giant stars are wrongly classified as dwarf stars.

For the faint sample shown in figure 10 the trend is similar to the bright sample but more scatter.

TABLE 6: The results for the Phoenix library with SVM

AP residual	All stars	A stars	F stars	G stars	K stars
G< 16.5 mag					
$< \|dT_{\rm eff}\| >$(K)	71	111	65	53	117
$< \|dA_0\| >$(mag)	0.05	0.05	0.03	0.04	0.13
$< \|d[{\rm Fe/H}]\| >$(dex)	0.33	0.65	0.35	0.23	0.32
$< \|d{\rm logg}\| >$(dex)	0.39	0.23	0.27	0.43	0.90
G> 16.5 mag					
$< \|dT_{\rm eff}\| >$(K)	265	426	226	226	392
$< \|dA_0\| >$(mag)	0.14	0.16	0.11	0.14	0.30
$< \|d[{\rm Fe/H}]\| >$(dex)	0.51	0.71	0.51	0.41	0.58
$< \|d{\rm logg}\| >$(dex)	0.47	0.35	0.33	0.51	1.02

Note that for metal-rich stars the [Fe/H]is underestimated by -0.5 dex in the faint sample, which is not that abvious in the bright sample.

5.2 ILIUM results

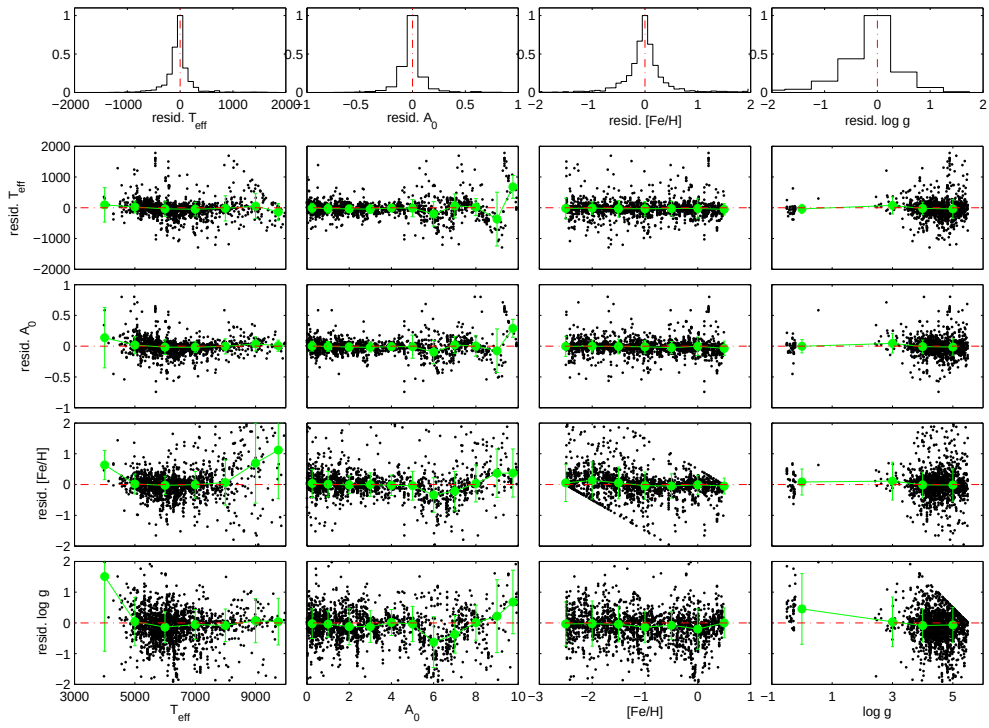

FIGURE 11: ILIUM AP residuals as functions of the true APs at $G < 16.5$ mag for the Phoenix library.

By comparing table 7 and 6 we find that the estimation of $T_{\rm eff}$from ILIUM is less accurate

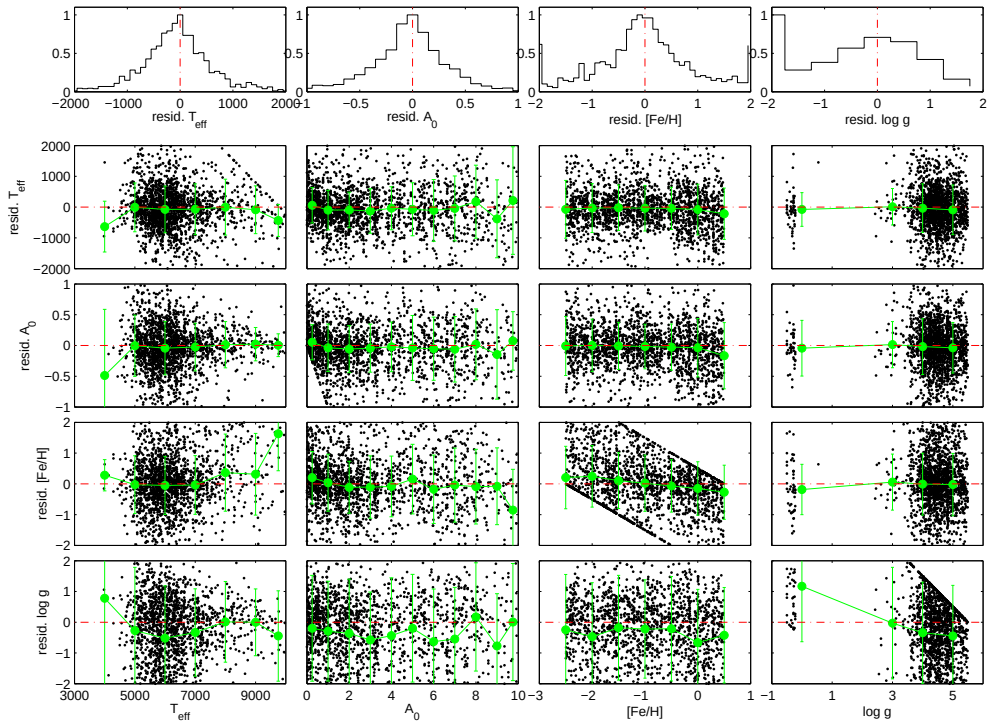

FIGURE 12: ILIUM AP residuals as functions of the true APs at $G > 16.5$ mag for the Phoenix library.

TABLE 7: The results for the Phoenix library with ILIUM

AP residual	All stars	A stars	F stars	G stars	K stars
G<16.5 mag					
$<\|dT_{\rm eff}\|>$(K)	150	262	151	114	152
$<\|dA_0\|>$(mag)	0.07	0.08	0.07	0.07	0.12
$<\|d[{\rm Fe/H}]\|>$(dex)	0.28	0.84	0.23	0.18	0.21
$<\|d{\rm logg}\|>$(dex)	0.40	0.39	0.39	0.36	0.57
G>16.5 mag					
$<\|dT_{\rm eff}\|>$(K)	553	615	585	515	484
$<\|dA_0\|>$(mag)	0.30	0.21	0.29	0.33	0.34
$<\|d[{\rm Fe/H}]\|>$(dex)	0.74	1.12	0.72	0.68	0.61
$<\|d{\rm logg}\|>$(dex)	1.32	0.84	1.25	1.50	1.46

than that from SVM by a factor of two for the whole sample of stars. The worst performance of ILIUM is for A stars, which is not the case for SVM. On the other hand, the performance of ILIUM for $A_0$is quite similar to that of SVM and they have the same trend, that is the performance gets worse for cool stars (K stars). By analysing the results only for the bright sample the performance of [Fe/H]and log$g$with ILIUM is comparable to the ones with SVM. However, for the faint sample ILIUM gives more scattered results than SVM.

For the bright sample shown in figure 11, the estimation of $T_{\rm eff}$in the second row does not show any correlation with the true APs, though the residuals of $T_{\rm eff}$start to oscillate when the true $A_0$is larger than 5 mag. Note that the discrete AP grid used for training the forward model in ILIUM suffers from very sparse distribution of values in A_0. Especially at larger extinctions (it only has spectra at 5,8 and 10 mag), which might degrade the fitting of stars with such values of extinction.

Because of the intrinsic degeneracy between $T_{\rm eff}$and A_0, the third row shows a similar trend of the estimation of A_0as the one of $T_{\rm eff}$.

For the estimation of [Fe/H], the fourth row shows that although the [Fe/H]is overestimated for hot stars and the estimated [Fe/H]is oscillate in high extinction its correlation with the true [Fe/H]and log$g$values is less apparent than in the result of SVM.

For the estimation of log$g$, the last row shows that in ILIUM the giant stars are well distinguished from the dwarfs, though the systematic bias for stars with log$g\sim$ 0 reaches 0.5 dex. Note that in the region of large extinction the residuals of log$g$also oscillate as the other three APs.

For the faint sample shown in figure 10, we find that the dispersion due to the low signal to noise ratios in the test spectra smears out a lot of correlations between the residuals of estimated APs and their true values, which we observe in the bright sample. The oscillations, for instance, of the estimated APs have disappeared in the high extinction region. The only correlation in the faint sample is found between the residual of log$g$and its true values. This implies that the estimation of log$g$turns out to be hard for faint samples in ILIUM.

TABLE 8: Completeness and contamination of the spectral type selection based on temperatures estimated for the Phoenix library. All samples have $A_0 < 3$ mag.

Algorithm	G mag		A stars	F stars	G stars	K stars
SVM	< 16.5	completeness	0.952	0.969	0.968	0.846
SVM	< 16.5	contamination	0.029	0.030	0.076	0.031
SVM	> 16.5	completeness	0.742	0.860	0.751	0.407
SVM	> 16.5	contamination	0.077	0.214	0.308	0.231
ILIUM	< 16.5	completeness	0.919	0.867	0.926	0.819
ILIUM	< 16.5	contamination	0.040	0.056	0.185	0.130
ILIUM	> 16.5	completeness	0.842	0.627	0.476	0.611
ILIUM	> 16.5	contamination	0.362	0.278	0.421	0.647

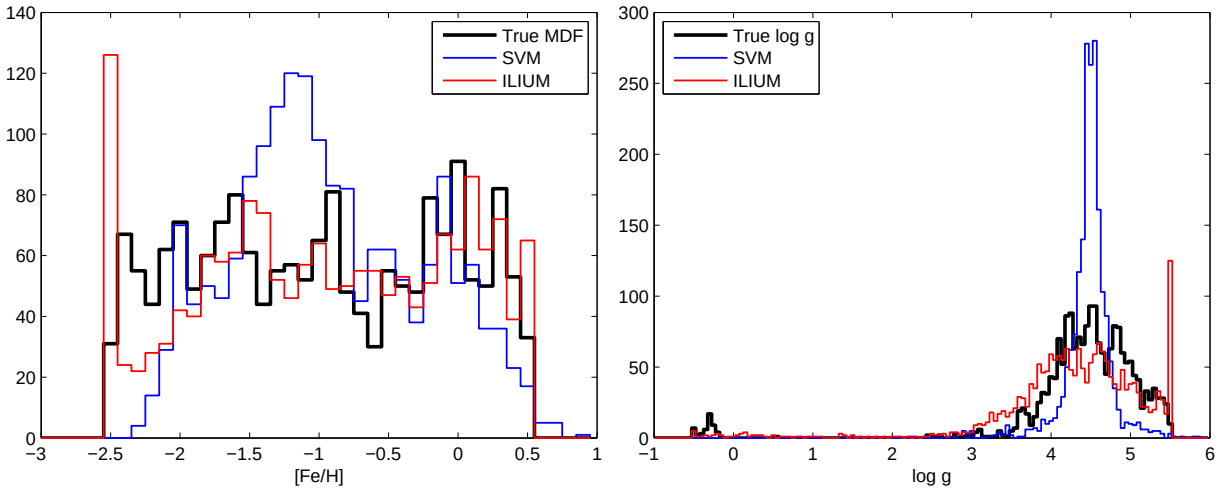

FIGURE 13: Reconstruction of the metallicity (left panel) and surface gravity distribution (right panel) with the Phoenix library.

5.3 More analysis on SVM and ILIUM results

Another test of the performance of the two algorithms is to calculate what the completeness and contamination would be when selecting certain spectral type of stars based on SVM/ILIUM estimated $T_{\rm eff}$. The test results are listed in table 8.

Both the SVM and ILIUM estimation separate well the certain spectral type of stars from their estimated $T_{\rm eff}$for the bright sample. The completeness in SVM is from 85% to 95%, while in ILIUM from 82% to 92%. For K stars, the completeness selected from both SVM and ILUM estimated $T_{\rm eff}$are the worst for the bright sample. From figure 9 and 11 we find that this may be due to the high extinction, which affects the estimation of $T_{\rm eff}$for cool stars. This is also found in the faint sample using SVM estimated $T_{\rm eff}$, but is not true for ILIUM estimated $T_{\rm eff}$. This is again because of the low signal to noise ratio of the spectra for the faint sample which influence the forward model fitting in ILIUM much more than SVM, which models both the spectra and their noise implicitly.

The contamination of the selection based on SVM is fairly small. It is less than 8% for the bright sample and less than 30% for the faint sample. It is smaller than that based on ILIUM. It shows again that G stars are the ones that have largest contamination, in both SVM and ILIUM, as shown for the cycle5 BaSeL library. It seems that this is not due to the algorithm we use to estimate $T_{\rm eff}$. However, looking back to the residual figures shown in previous sections we find that for the BaSeL library the contamination is mostly from the hotter stars which $T_{\rm eff}$is underestimated (figure 4 – 7), while for Phoenix the contamination in G stars are actually from the opposite direction, i.e. the overestimation of the cool stars (figure 9–12). Therefore, although for both synthetic libraries we find the G stars suffer the most contamination, the source of the contamination in the two synthetic libraries is different.

Another test shown in figure 13 is to reconstruct the distributions of the [Fe/H]and log$g$from the estimated APs. It is exhibited that the ILIUM estimated [Fe/H]and log$g$reconstruct well the distributions of the metallicity and surface gravity of the test dataset, while SVM estimated [Fe/H]and log$g$are affected by the systematic bias more than ILIUM. From the metallicity distributions we see that SVM predicts values around the mean, i.e. [Fe/H]$= -1 \sim -1.5$ dex, for more stars. Although the ILIUM predicted [Fe/H] show a spark at -2.5 dex, in most values that are higher than -1.5 dex, the distribution of the ILIUm estimated [Fe/H]follows well the true distribution. For the log$g$distribution, ILIUM does predict the the small peak around log$g$<0 dex, while SVM shrinks the distribution and makes a high fake peak around 4.5 dex.

6 The degeneracy between $T_{\rm eff}$and A_0

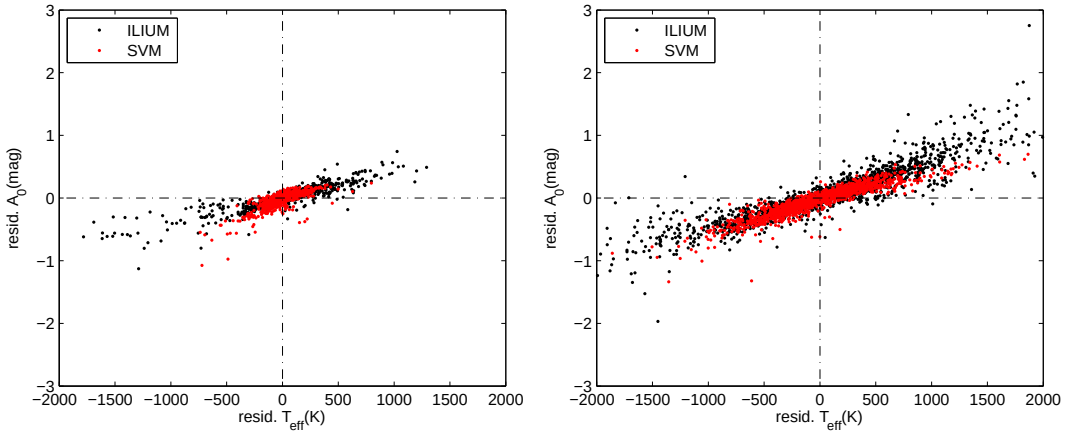

FIGURE 14: The correlations between residuals of $T_{\rm eff}$and $A_0$from both SVM (red dots) and ILIUM (black dots) with Phoenix dataset. The left panel is for the bright sample with $G < 16.5$ mag and the right is for $G > 16.5$ mag

It is known that reports that there is a degeneracy between $T_{\rm eff}$and A_0(Bailer-Jones (2010)). This degeneracy is intrinsic since the spectra with different combinations of $T_{\rm eff}$and A_0can be very similar, as shown in figure 21 in Bailer-Jones (2010). Figure 14 shows the difference of the degeneracy in SVM and ILIUM. The degeneracy is measured via correlation coefficient of the residuals of the estimated $T_{\rm eff}$and A_0. For the bright sample the correlation coefficient is $\sim$0.7 for SVM and $\sim$0.9 for ILIUM, while for the faint sample these two values are turn out to be $\sim$0.9 for both algorithms.

According to the implementation of SVM in GSP-Phot, the APs are modelled and estimated independently. Therefore, the correlation between the residuals of $T_{\rm eff}$and A_0in SVM is essentially due to the intrinsic degeneracy plus random and independent noise. However, the correlation coefficient in ILIUM is composed by the intrinsic degeneracy, the random but independent noise, and the covariance between the estimated $T_{\rm eff}$and A_0due to the simultaneous

calculation of $T_{\rm eff}$and A_0using a 2-dimensional forward model. Hence, for bright sample, in which random noise is not large, the correlation looks stronger in ILIUM because the forward model contribute an extra term above the intrinsic degeneracy. On the other hand, for the faint sample, in which random noise is very large, the correlation is dominated by the random noise and the forward model's contribution can be ignored.

7 Comparison between BaSeL and Phoenix

It is of interest to compare the degeneracy between the BaSeL and the Phoenix library. The correlation coefficients based on the bright sample of the BaSeL test data are 0.7 for both SVM and ILIUM, and both are 0.8 in the faint sample. It may be due to the higher noise in the BaSeL data. The random noise in BP/RP spectra are contributed both from the interpolation of the spectra from discrete APs (AV-007) and Gaussian noise added to simulate the photon noise, read-out noise and sky background noise. Since the Gaussian noise are added on both the BaSeL and the Phoenix data in same way, it seems that the interpolation contribution plays more important role in BaSeL than in Phoenix. Indeed, looking at the synthetic libraries in details we find that the discrete grid of the BaSeL library is sparser than that of the Phoenix.

By comparing table 3 and 6, we see that the performance of both SVM and ILIUM on the Phoenix is better than that on the BaSeL. For instance, the estimation of $T_{\rm eff}$is twice better for bright sample of Phoenix than that of BaSeL. The different qualities of interpolation in BaSeL and Phoenix may cause the difference in the performance. Note that the ranges of $T_{\rm eff}$, [Fe/H], and log$g$are also different for the two synthetic libraries. It is not clear that this difference may afect the performance of SVM and ILIUM. In principle, the different ranges of APs would play different role in SVM and ILIUM. Consequently, it is hard to investigate this unless doing further test using the datasets selected from the two libraries with exactly the same range of APs.

8 Apply BaSeL trained algorithms to MARCS data

It is not feasible to test the BaSeL dataset with Phoenix trained models, since the libraries are from two different cycles and the spectra are not compatible, e.g. the wavelength scale and the instrument model between cycle5 and cycle7 are different. However, we can use the MARCS library as a test dataset with BaSeL trained algorithms to test the performance of the overlapping part of different libraries. This test is important for understanding the performance of the algorithms when applying to different libraries, since the simulation data is combined with several different libraries to cover the whole range of APs space for normal stars.

Table 9 shows the results of the MARCS test dataset with BaSeL trained SVM. It shows that in SVM all 4 APs estimates are less accurate with MARCS than with BaSeL by a factor of 1.5.

This reflects the extent of the difference of the spectra between the two libraries.

TABLE 9: SVM results for the MARCS and BaSeL datasets. The absolute residuals of APs.

AP residual	All stars	A stars	F stars	G stars	K stars
MARCS as test dataset					
$< \|dT_{\rm eff}\| >$ (K)	351	667	460	272	347
$< \|dA_0\| >$ (mag)	0.28	0.25	0.29	0.23	0.30
$< \|d[Fe/H]\| >$ (dex)	0.64	0.68	0.64	0.53	0.70
$< \|d\log g\| >$ (dex)	0.21	0.17	0.23	0.26	0.18
BaSeL as test dataset					
$< \|dT_{\rm eff}\| >$	239	696	384	184	145
$< \|dA0\| >$	0.19	0.33	0.26	0.15	0.16
$< \|d[Fe/H\| >$	0.45	1.16	0.71	0.45	0.31
$< \|d\log g\| >$	0.17	0.21	0.22	0.21	0.13

TABLE 10: ILIUM results for the Phoenix full discrete grid in the forward model.

AP residual	All stars	A stars	F stars	G stars	K stars
G<16.5 mag					
$< \|dT_{\rm eff}\| >$(K)	363	483	419	279	291
$< \|dA_0\| >$(mag)	0.25	0.18	0.29	0.21	0.24
$< \|d[\mathrm{Fe/H}]\| >$(dex)	0.35	0.78	0.29	0.32	0.29
$< \|d\log g\| >$(dex)	0.61	0.49	0.59	0.62	0.77
G>16.5 mag					
$< \|dT_{\rm eff}\| >$(K)	753	971	803	640	706
$< \|dA_0\| >$(mag)	0.47	0.41	0.48	0.45	0.55
$< \|d[\mathrm{Fe/H}]\| >$(dex)	0.75	1.24	0.68	0.70	0.74
$< \|d\log g\| >$(dex)	1.45	0.99	1.40	1.65	1.51

9 ILIUM trained with different density of the discrete grid

ILIUM relies on the discrete grid to support the forward model. The density of the discrete grid affects the performance of the AP estimation. All results of ILIUM provided in previous sections are obtained from a sparse grid by a factor of 2 in $T_{\rm eff}$. Table 10 shows the Phoenix results with the full discrete grid. Surprisingly, for the strong APs, i.e., $T_{\rm eff}$and A_0, the sparse grid results much smaller absolute residuals than the full grid. This contradicts with the experience that denser grid gives better performance in fitting. The reason why the performance is better with the sparse grid is still unclear and more tests are needed to investigate this issue.

10 Other tests

Some other tests in GSP-Photare summarized here: i) the binned spectra test to investigate if binning of the spectra helps to improve the performance; ii) the radial velocity test to investigate whether the radial velocity, in particular the high velocity stars, will affect the estimation of APs in GSP-Phot.

10.1 Test of the binned BP/RP

For the binned spectra test we bin the spectra by 2, 3 and 5 pixels for both training and test datasets and compare the APs estimates with SVM and ILIUM. Table 11 show the results of SVM for $T_{\rm eff}$, A_0, log$g$and [Fe/H]at 3 fixed magnitudes: 15, 18.5 and 20 mag. No correlation is found between the bin size and the rms of the APs. Similar results with ILIUM are shown in table 12.

This simple test proves that binning BP/RP spectra does not help to improve the AP estimation.

TABLE 11: SVM-Binned results. The rms of the unbinned, binned by 2, 3 and 5 pixels is listed here.

G	$\sigma(1)$	$\sigma(2)$	$\sigma(3)$	$\sigma(5)$
		$T_{\rm eff}$		
150	0.0143	0.0148	0.0156	0.0198
185	0.0209	0.0222	0.0224	0.0270
200	0.0300	0.0310	0.0308	0.0340
		A_0		
150	0.6862	0.6351	0.6586	0.9988
185	0.8016	0.8446	0.8581	1.2316
200	1.1656	1.1638	1.2120	1.4953
		[Fe/H]		
150	0.8626	0.9022	0.9280	1.0761
185	1.1602	1.1913	1.2201	1.3424
200	1.4240	1.4128	1.4309	1.4932
		log$g$		
150	1.1089	1.1369	1.1985	1.5627
185	1.6680	1.7387	1.8258	2.1523
200	2.2822	2.3393	2.3816	2.5809

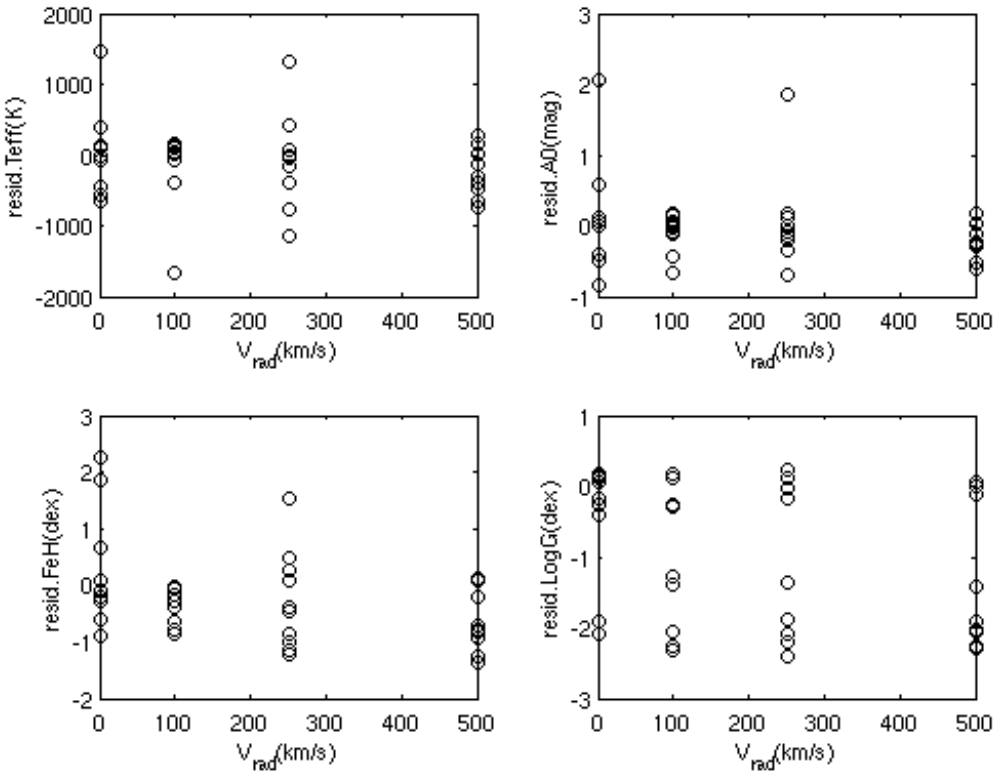

FIGURE 15: AP residuals as functions of the radial velocity. APs are estimated with SVM trained with the BaSeL libraray.

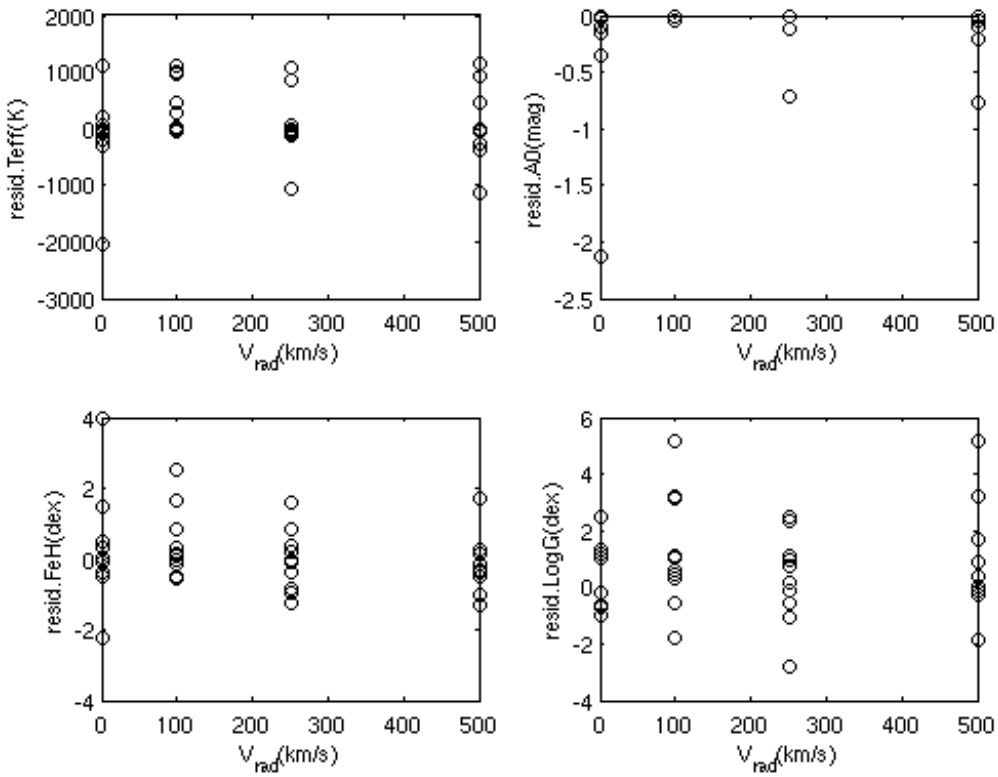

FIGURE 16: AAP residuals as functions of the radial velocity. APs are estimated with ILIUM trained with the BaSeL libraray.

TABLE 12: ILIUM-Binned results. The rms of the unbinned, binned by 2, 3 and 5 pixels are listed here.

G	$\sigma(1)$	$\sigma(2)$	$\sigma(3)$	$\sigma(5)$
		$T_{\rm eff}$		
150	0.0158	0.0176	0.0172	0.0197
185	0.0242	0.0262	0.0264	0.0289
200	0.0325	0.0327	0.0349	0.0373
		A_0		
150	0.7467	0.8732	0.8115	0.9321
185	0.9046	1.0428	0.9828	1.1536
200	1.3059	1.2566	1.3703	1.6353
		[Fe/H]		
150	0.8187	0.8561	0.8635	0.9824
185	1.2135	1.2395	1.2681	1.3195
200	1.4532	1.4472	1.4808	1.5214
		$\log g$		
150	1.0815	1.1433	1.1857	1.5765
185	1.8826	1.9859	2.1056	2.3139
200	2.4675	2.5393	2.5495	2.7774

10.2 The influence of the radial velocities

For stars with high radial velocity, their spectra will shift a little bit in wavelength. Does this affect the AP estimation? A simple test based on only 40 stars with various radial velocities (0, 100, 250 and 500 km/s) is performed. The $T_{\rm eff}$of these stars are around 5000 K without extinction. The SVM and ILIUM used in the test are trained by the cycle5 BaSeL dataset. Figure 15 and 16 show the residuals of the SVM and ILIUM estimated APs as functions of the radial velocities. It seems that there is no obvious correlation between the performance of the AP estimates and the radial velocities. Notice that the test sample are quite few and does not cover a wide range of APs, therefore the conclusion may change if we apply the algorithm to a large test sample covering the whole range of $T_{\rm eff}$.

11 Conclusion and discussions

In this TN we performed an array of tests: two algorithms for two synthetic libraries. from these test we see that first, in sense of the algorithms, SVM gives slightly better estimation for strong APs than ILIUM, while ILIUM gives unbiased and reasonable estimations for weak APs. With SVM estimated $T_{\rm eff}$, stars can be reliably classified in different spectral types with high completeness (more than 85%) and lower contamination than 8% for stars brighter than 16.5 mag. With ILIUM the accuracy of [Fe/H]estimates for cool stars is around 0.3 dex for

bright samples, which is more or less comparable with the quality of the current multi-band photometry surveys, e.g. SDSS. The reconstructed [Fe/H]and log$g$distributions is consistent with the true distributions, which is promising. SVM is more accurate than ILIUM for the faint samples, which has lower signal to noise. It is consistent with our expectation that SVM can work with more noisy data while the forward model is more sensitive to the signal to noise.

Second, the comparison between the cycle5 BaSeL and the cycle7 Phoenix in the same algorithm shows that the performance with the Phoenix is significantly better than that with the BaSeL library. We find from the degeneracy between $T_{\rm eff}$and A_0that the interpolation of the random spectra in the two synthetic libraries may play an important role in the difference of the performance.

Moreover, the overlap test between BaSeL and MARCS shows how big the difference would be in the APs estimation due to the different synthetic spectra. For the random grid of the MARCS the difference of the BP/RP spectra with the BaSeL for the same APs may lead to averagely a factor of 1.5 worse in AP estimation.

Binned spectra test concludes that binning of the BP/RP spectra does not help to improve the performance of the AP estimation. Additionally a small radial velocity test sample shows that radial velocity does not affect the AP estimation in either SVM or ILIUM. However, it is noted that the test dataset is only around $T_{\rm eff} = 5000$ K, which is not complete the AP range.