

## NEAT Memo

NEAT-sci-001

Orig. : *A. Crouzier*  
Dest. : NEAT consortium  
Date : 16-09-2014  
Version : v4

Subject : **Creation and exploitation of the NEAT catalogs**

---

### Memo version history

- 25-Jan-2012 (v1): initial writing
- 29-May-2012 (v2): addition of the section for HZ calculation
- 01-Jun-2012 (v3): reorganization of the sections
- 16-09-2014 (v4): major update, change of main sequence selection criteria (broader tolerances), addition of observational strategies

### Scope

The targets of the NEAT mission are the Sun-like stars in our close neighborhood. They correspond more precisely to the main sequence stars of spectral types A, F, G, K and M that are located at a distance of less than 20 pc from the Sun.

The goal of this memo is to describe the sequence of commands I used to create the **NEAT catalog of targets** with the program Topcat. This catalog is defined as the A, F, G, K and M main sequence stars, located at less than 20 pc from the Sun, in the Hipparcos catalog. We will also define the **NEAT catalog of references** stars from the **NEAT catalog of targets**. Other sections will show how one can use Topcat and the newly defined NEAT catalogs to check for sample completeness and to obtain statistics about the target and reference stars.

Topcat is a spreadsheet application for astronomical catalogs, developed by Mark Taylor. It offers the possibility to define catalog subsets and has a nice set of visualization tools. It can be downloaded for free from this webpage: <http://www.star.bris.ac.uk/~mbt/topcat/>.

### Acronyms

BC Bolometric Correction



HZ Habitable zone

NEAT Nearby Earth Astrometric Telescope

SciDAVis Scientific Data Analysis and Visualization

SNR Signal to Noise Ratio

TOPCAT Tool for OPERations on Catalogues And Tables

## Reference documents, files, catalogs and papers

- NEAT: a spatial telescope to detect nearby exoplanets using astrometry (PhD manuscript of Antoine Crouzier)
- NEAT Proposal of 2010
- TOPCAT Reference manual
- Hiparcos catalog - available on Vizier
- Tycho2 catalog - available on Vizier
- The exoplanet catalog of “The Extrasolar Planets Encyclopaedia”
- Transformations from Theoretical Hertzsprung-Russell Diagrams to Color-Magnitude Diagrams: Effective Temperatures, B-V Colors, and Bolometric Corrections (Flower, 1996)
- The B-V / BC / Teff relation for main sequence stars, subgiants, and giants (Flower, 1996) - available on Vizier
- Target star catalogue for Darwin Nearby Stellar sample for a search for terrestrial planets (Kaltenegger et al., 2010).
- Habitable Zones around Main Sequence Stars (Kasting et al., 1993)
- Physical Parameters of the Stars (Schmidt-Kaler, 1982)
- NEAT error budget (Renaud Goullioud, priv comm.)
- Spreadsheet for calculation of NEAT photon noise limit: `neat_photon_limit.xls`
- Planet population synthesis model (Alibert et al., 2013).
- List of synthetic exoplanet systems: `CD2133_1tev.dat`.
- Matlab script (calculation of exoplanet yields): `observational_strategy_yields.m`
- Python script (calculation of reference equivalent magnitude): `process_ref_stars.py`



## Contents

<b>1</b>	<b>Creation of the NEAT catalogs of target and reference stars</b>	<b>4</b>
1.1	Extraction of the NEAT catalog of targets from the Hipparcos catalog . . . . .	4
1.2	Creation of the catalog of reference stars. . . . .	5
1.3	Availability of the reference stars . . . . .	7
1.4	Calculation of the reference star total flux . . . . .	8
<b>2</b>	<b>Creation of the NEAT columns</b>	<b>9</b>
2.1	Determination of the astrometric signal in HZ . . . . .	9
2.2	Time allocation schemes and science yields . . . . .	14
2.3	Astrometric signal in HZ versus stellar mass . . . . .	22
2.4	Crossmatch with "The Extrasolar Planets Encyclopaedia" . . . . .	23
2.5	Checking the sample completeness . . . . .	27
<b>A</b>	<b>Summary of the Topcat operations</b>	<b>30</b>
<b>B</b>	<b>List of the NEAT columns</b>	<b>30</b>
<b>C</b>	<b>Python script for summing reference stars flux: source code</b>	<b>32</b>



## 1 – Creation of the NEAT catalogs of target and reference stars

### 1.1 – Extraction of the NEAT catalog of targets from the Hipparcos catalog

The **NEAT catalog of targets** will consist of all the main sequence stars of spectral types A, F, G, K and M located at a distance of less than 20 pc from the sun (see Fig. 1). What I call the **final list of targets** will be a sub sample of this pre selection, relative to the exercise of observation time allocation. The final selection will prioritize the targets with the highest astrometric signals, so depending on the observational strategy, a fraction of the stars is likely to be excluded.

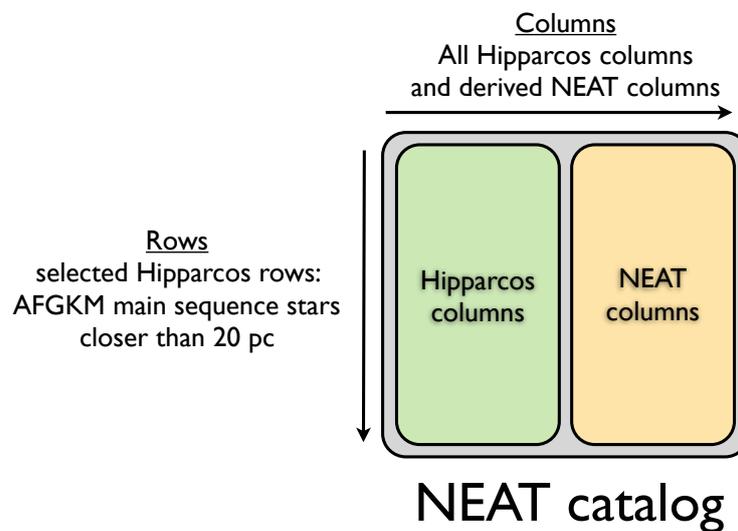


Figure 1: *Schematic view of the NEAT catalog of target stars.* The NEAT targets are selected within the Hipparcos catalog. The part called NEAT columns are computed values from the Hipparcos original columns.

Unfortunately Topcat does not offer the possibility to save a list of commands. Instead the full session, including all the Hipparcos data, has to be saved in a fits-plus file, which can have a substantial size (hundreds of mega bytes in our case), this constraint limits the possibility to share the Topcat sessions. To mitigate this problem, I will describe in details the commands I used to define the NEAT catalog.

To begin, the NEAT catalog is created from a subset of stars from the Hipparcos catalog. In order to manipulate the data more easily, several subsets will be defined. Additionally, we have to download the Hipparcos catalog from the vizier portal because we need to add RA2000 and DEC2000 columns.

This is done with the following sequence of actions:

- Search for the Hipparcos catalog in Vizier.
- In the catalog list select **All** (so you will be also downloading the tycho catalog, which is the last table).
- On the **Preferences** window on the left, tick **All columns**, set max to **unlimited**, set the format to **VOTable** and verify that the J2000 box is active in the **Compute** menu.



- Click **Query selected Tables** and save the tables on the hard drive.
- In Topcat, open the **Load New Table** dialog box via the menu **File > Load table**.
- Click on the **System Browser** icon and load the catalogs.
- Select the Hipparcos main catalog: 1/239/hip\_main (it is the first table in the **Table List**).
- Click the tool **Display row subsets**.
- Click the tool **Define a new subset...** in the window showing the list of subsets
- Type a condition in the column **Algebraic expression**

The last two actions are repeated for each new subset. Each subset is defined by its own algebraic expression. Here is the list of all the subsets I chose to create:

- A first subset named `closest_stars` containing all the Hipparcos stars closer than 20 pc, using the following expression: `Plx > 50`
- One subset for each sun-like spectral type (A, F, G, K and M), among the closest stars. For example, for the main sequence F stars, the subset name is: `closest_MS_F_stars` and the expression is:  
`_2 && matches(SpType, "^F.*(IV-V|[^I]V)")`
- One subset named `closest_MS_FGK_stars`, this is the union of the three subsets above. The expression is:  
`_2 && (matches(SpType, "^F.*(IV-V|[^I]V)") || matches(SpType, "^G.*(IV-V|[^I]V)") || matches(SpType, "^K.*(IV-V|[^I]V"))`
- One subset named `closest_MS_AFGKM_stars` with the A,F,G,K and M main sequence stars closer than 20 pc. This is the NEAT catalog of targets. The expression is:  
`_2 && (matches(SpType, "^A.*(IV-V|[^I]V)") || matches(SpType, "^F.*(IV-V|[^I]V)") || matches(SpType, "^G.*(IV-V|[^I]V)") || matches(SpType, "^K.*(IV-V|[^I]V)") || matches(SpType, "^M.*(IV-V|[^I]V"))`

The `(IV-V|[^I]V)` part in the regular expression mean that we accept selected all main sequence stars (or between main sequence and sub-giant). For stars with two components (separated by a “+” sign), we should ideally base the selection on the main component. However, errors will be made if the secondary component only is main sequence: with this regular expression, the star will be included in the list. The problem is that it is not possible to use a “\” sign in the regular expression (it causes an error in Topcat), it this therefore impossible to get the needed escaped “+” in the regex.

“`_2`” refers to the second subset (“`_1`” is referring to the full Hipparcos catalog itself). By doing this, each spectral class can be easily manipulated. You should now have 351 stars in the `closest_MS_FGK_stars` subset and 455 in the `closest_MS_AFGKM_stars` subset.

## 1.2 – Creation of the catalog of reference stars.

One strong requirement of NEAT is the availability of the reference stars. The field of view of NEAT is limited to about  $0.6^\circ$  because of the aberrations. For each target star, NEAT needs at least 6 reference stars of  $V_{\text{mag}} < 11$  in the field of view to reach its specified accuracy. The reference stars are ideally quiet and immobile stars, used for differential astrometry measurements. In practice they are mostly K giants at long distances. Ideally, the reference stars should fill the following list of conditions:



- Magnitude lower than 11
- Not too noisy...

Stars too close, too active or with giant planets could show excessive noise and be unsuitable to be used as references. In practice, we only have knowledge about the magnitude in Tycho and some targets critically lack reference stars. It is important to keep all we can find<sup>1</sup>. Also, the shape of the NEAT detector is square, which authorizes the use of some reference stars located between  $0.6^\circ$  and  $0.84^\circ$ <sup>2</sup>. So we will use the whole Tycho catalog, with a threshold of either  $0.6$  or  $0.84^\circ$ . We will also calculate a magnitude equivalent to the sum of the flux from the reference stars to estimate the reference star photon noise for each target (some targets might have a lot of references stars but only faint ones).

The reference stars are found by cross-matching the NEAT catalog (`closest_MS_AFGKM_stars` subset) with the Tycho catalog. The list of conditions above can be applied to filter out all the stars that are unsuitable (for example to reject stars with  $V_{\text{mag}} < 13$ ). It is necessary to apply at least the condition `!NULL_Vmag`: some stars have unknown/null magnitudes and will crash the script (section 1.4) unless they are filtered out. Here is the list of the corresponding operations in Topcat:

- Select the Tycho catalog ("`1/239/tyc_main`").
- Create a "`reference_stars`" subset of the Tycho catalog with the conditions of your choice (for example `Vmag < 13 && !NULL_Vmag`).
- Crossmatch the NEAT catalog ("`closest_MS_AFGKM_stars`" subset) with the "`reference_stars`" subset. Use the following parameters:
  - **Max Error**  $0.3$  or  $0.42^\circ$
  - **Table 1**: Hipparcos main catalog
  - **RA column** (for Table 1): `_RAJ2000` - degrees
  - **Dec column** (for Table 1): `_RAJ2000` - degrees
  - **Table 2**: Tycho main catalog
  - **RA column** (for Table 2): `_RAJ2000` - degrees
  - **Dec column** (for Table 2): `_RAJ2000`- degrees
  - **Match selection**: Best match for each Table 2 row
  - **Join type**: 1 and 2
- Eventually plot the results on a sphere (to visually verify the RAs and DECs)

You should obtain a crossmatch table containing 6466 reference stars (for  $0.84^\circ$  FoV and with  $V_{\text{mag}} < 13$ ). Click on the tool **Display Table Metadata** and in front of the raw **Name**,

<sup>1</sup>Tycho is magnitude limited to about  $V_{\text{mag}} = 12 - 13$ . It is not complete after  $V_{\text{mag}} = 11$ , but still have a lot of stars between 11 and 13. These stars will be our backup reference is we lack brighter one and we will not be flooded with stars too faint to be useful.

<sup>2</sup>With only  $0.6^\circ$  of FoV, the first result indicated that a significant proportion of target have few reference stars (4 or less). Using  $0.84^\circ$  improves the result quite dramatically for some stars. We will overestimate the number of available references stars in some rare cases (not all references close to  $0.84^\circ$  can be used). But seeing how to best place a square FoV can not be solved with Topcat, it would require tricky programming instead.



change the value `Joined to Tycho reference stars crossmatch`. This will help distinguish all the different crossmatches.

The current method used to select the reference stars is very crude. For a real mission, other criteria should be used to filter out unsuitable reference stars. A minimum distance (for example 1 kpc) and a condition on the activity of the stars could be added. The parallax of the Tycho catalog can not be used as the errors are typically several dozen of mas (as a result, distances of more than one hundred parsec are not measured properly). An alternative to estimate the distance would be to use the H-R diagram... or to wait for Gaia. The statistics we will obtain are close to a best case scenario for the number of reference stars because Tycho is 99% complete to magnitude 11.

### 1.3 – Availability of the reference stars

Now that we have a catalog of the reference stars, we want to know **the distribution of the size of the groups of reference stars**, each group being all the reference stars associated with one NEAT target. This is a bit tricky to do this with Topcat: using directly the information of the column `count` in the crossmatch results to plot an histogram doesn't yield correct results. This is because for a group of  $n$  stars, each reference star will contribute one count. Consequently, in the final distribution, a group of size  $n$  will be counted  $n$  times.

To tackle this problem, do the following sequence of actions:

- Open the menu **Joins > Internal match..**
- Select the reference stars crossmatch in **Table**.
- Chose the option **Eliminate all but first of each group**.
- Leave all the other options unchanged and click ok.
- Change the name of the new crossmatch table to:  
**Histogram group size reference stars**

Now you can plot the histogram of the distribution of the number of reference stars for each NEAT target stars. Click on the **Histogram** tool, then change the X-axis to **GroupSize-1**. The -1 eliminates the target stars from the group (it was counted with the reference stars). The result is shown in the fig. 2.

Remark: The internal match discards the stars that are not in groups, i.e. for which no reference star is available. They do not appear in Fig. 2. By doing an internal match with the option **Eliminate All Grouped Rows** you will find how many targets are without reference (0 for  $0.84^\circ$  FoV and with  $V_{\text{mag}} < 13$ ).

### Conclusions

Though in average the number of reference is good, there is a significant fraction of the NEAT targets (about 100 of 340) that have only 4 available references stars or less (with a field of  $0.6^\circ$  and a  $V_{\text{mag}}$  limit of 11). This could affect the ability of NEAT to find planets around these stars.

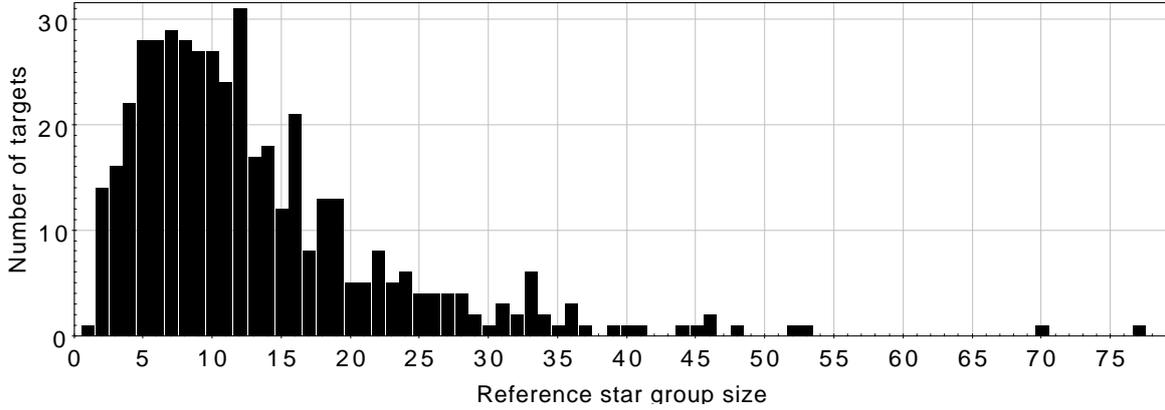
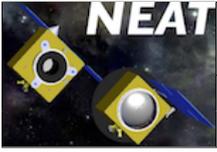


Figure 2: *Distribution of the number of reference stars per NEAT target star.*

### 1.4 – Calculation of the reference star total flux

The total flux of the reference stars is useful to estimate the integration time requirement for each target star and for a given astrometric signal and mass detection threshold.

It is not straightforward to obtain this quantity with Topcat: we have a catalog of the reference stars, which are organized in groups that also include the target star. We have the visible magnitude of all the stars of each group and we want to “sum” the magnitudes to find an equivalent magnitude which flux corresponds to the sum of the flux of the reference stars. The problem is that Topcat does not provide a way to perform this specific operation. So we will export the catalogs of the reference stars and the target stars as .csv files and run a python script through them. The catalog of the reference stars is called `reference_stars_topcat.csv` and the catalog of the subset of main sequence AFGKM NEAT targets is called `closest_MS_AFGKM_stars_topcat.csv`. To create these files, select the appropriate table and subset, then use the menu **File > Save Table(s)/Session**, select the .csv format, enter the name and hit save.

The operations applied in the python script is summarised here:

- Get all the targets’ hipparcos IDs from the catalog of targets
- Loop though the catalog of references and identify the targets stars among them
- Loop though the catalog of reference stars, sum the flux of all the reference stars (excluding the targets) and calculate the equivalent magnitudes
- Add the equivalent magnitudes into the catalog of targets

The equivalent magnitude is the magnitude of a star which flux would be the sum of the reference star flux. We note  $F_0$  the reference star flux for the visible magnitude. We have:

$$M = -2.5 \log (F/F_0) \Leftrightarrow F = F_0 \times 10^{-M/2.5}$$

$$\Rightarrow \sum_i F_i = F_0 \sum_i 10^{-M_i/2.5} \Rightarrow M_{eq} = -2.5 \log \left( \sum_i 10^{-M_i/2.5} \right)$$



After running the script a new csv file is created: `closest_MS_AFGKM_targets_topcat_out.csv`. Import the file back into the list of table via the menu **File > Load**. The new column `EqRefMag` has been added.

Figure 3 is an histogram of the equivalent magnitudes. The NEAT error budget assumed 6 stars of magnitude 11 (equivalent magnitude of 9.05). The histogram confirms that this is a sound assumption: nearly all the target have enough bright reference stars to verify “equivalent”  $V_{\text{mag}} < 9.05$ .

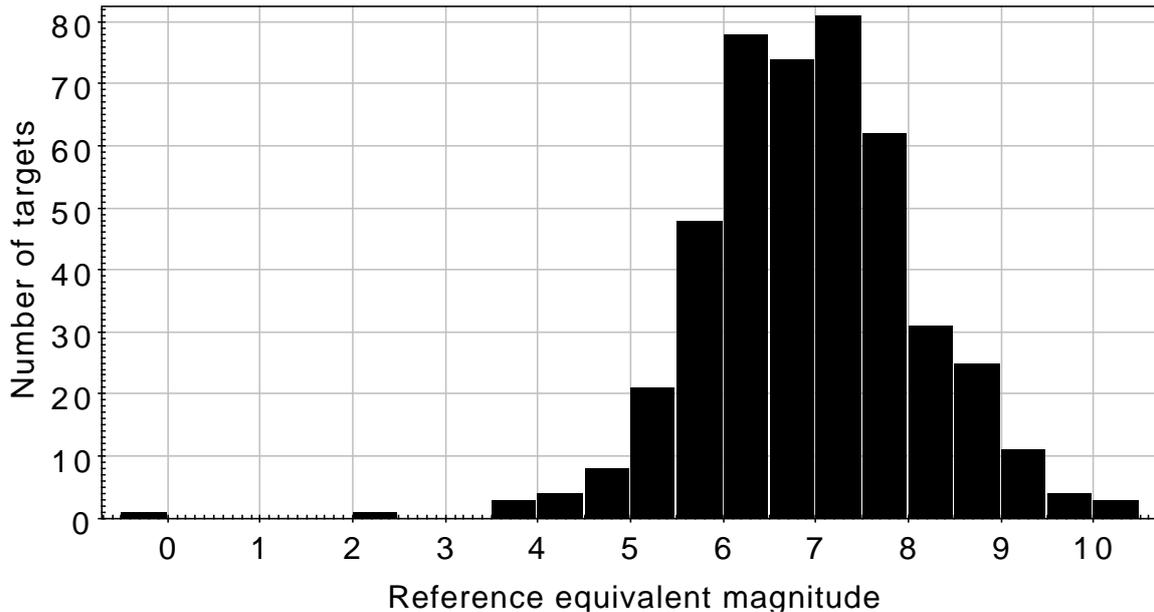


Figure 3: *Distribution of the equivalent magnitudes of groups of reference stars (one group per target).*

## 2 – Creation of the NEAT columns

In this second part we will add a lot of NEAT columns to the catalog of targets. Some of them can in principle be added before the column `eq_ref_mag` has been added. But the latter is integrated into a new table (`closest_MS_AFGKM_targets_topcat_out.csv`, different from the original Hipparcos subset). So any column created before has to be created again, or a crossmatch must be performed to merge the two tables. The simplest way to avoid the problem is to create the `closest_MS_AFGKM_targets_topcat_out.csv` table first (with the `eq_ref_mag` column), and to create all other NEAT columns later.

### 2.1 – Determination of the astrometric signal in HZ

#### 2.1.1 – Introduction

The goal of this subsection is to add a new column specifying the astrometric signal of an Earth mass planet in the HZ of the target stars. The amplitude of the astrometric signal



yielded by an exoplanet at a distance  $D$ , of mass  $M_{\text{Planet}}$ , with a semi major axis  $R$  around a star of mass  $M_{\text{Star}}$  is given by:

$$A = 3\mu as \times \frac{M_{\text{Planet}}}{M_{\text{Earth}}} \times \left(\frac{M_{\text{Star}}}{M_{\text{Sun}}}\right)^{-1} \times \frac{R(\text{AU})}{1\text{AU}} \times \left(\frac{D(\text{pc})}{1\text{pc}}\right)^{-1} \quad (1)$$

From that formula one can see that a star mass increase will have the following effects on the signal:

- Greater star/planet mass ratio (decreases the signal)
- Greater star luminosity  $\Rightarrow$  further away HZ  $\Rightarrow$  greater semi major axis (increases the signal)

Note that the mass and the spectral type are strongly correlated, so changes in the stars emission spectrum have to be taken into account. The additional effects, implicit in the formula above, that happen when the mass increase are listed below:

- Blue shifted spectrum  $\Rightarrow$  Greater normalized solar flux factor  $\Rightarrow$  closer HZ  $\Rightarrow$  smaller semi major axis (decreases the signal).
- Greater semi major axis  $\Rightarrow$  Period increase
- Increase of the target stellar flux.

The normalized solar flux factor is a coefficient taking into account the overlap of the stellar emission and the planetary albedo. The values of this coefficient for different spectral types are found in the literature (Kasting et al., 1993). The overall effect is a moderate negative feedback on the variation of the HZ location with the mass: for more blue and luminous stars, the albedo is higher, so the net energy input is a bit smaller than expected.

The period also has an influence on the detection threshold. Because of the nature of the mission, for periods longer than the mission during the minimum planet mass required for detection increases, despite an increasing astrometric signal: in this case only a portion of the orbit will be sampled so a larger SNR is required. But in our case study this parameter doesn't matter much: even for F stars the planet period at the HZ outer edge barely exceed 5 years.

Finally, if the stellar flux is larger, the precision of the astrometric measurements will increase more rapidly with the integration time (noted T): the location error on the measured centroid of the target will be  $\sigma_{\text{measure}} \propto \frac{1}{\sqrt{TF_{\text{star}}}}$ . However in the case of NEAT which is differential astrometry, the photon noise of the much fainter reference stars dominates in most cases. This means that the final precision will be almost insensitive of the target star flux, but will depend on: the astrometric signal, the photon noise of the reference stars and the integration time.

The overall effect of the host star mass on the astrometric signal in HZ is a bit tricky to determine. The result is that the astrometric signal increase with the stellar mass, see section 2.3 for details.

Another important parameter I haven't mentioned yet is the distance, but it is easily derived from the Hipparcos parallax. The last parameter, the planet mass, can not be known before the observations. We will normalize by the Earth mass: the signal for any mass is easily obtained by scaling by the ratio:  $\frac{M_{\text{planet}}}{M_{\text{Earth}}}$ .



2.1.2 – Determination of Dist, L, HZin, HZout, and Mass

To determine the astrometric signal in habitable zone, we will determine the following intermediate quantities: the distance, the bolometric luminosity, the HZ inner and outer boundaries, and the star mass for each NEAT target star (respectively noted Dist, L, HZin, HZout, and Mass). Figure 4 shows how L, HZin, HZout, and Mass are derived. The detailed descriptions of the process for each quantity are presented below.

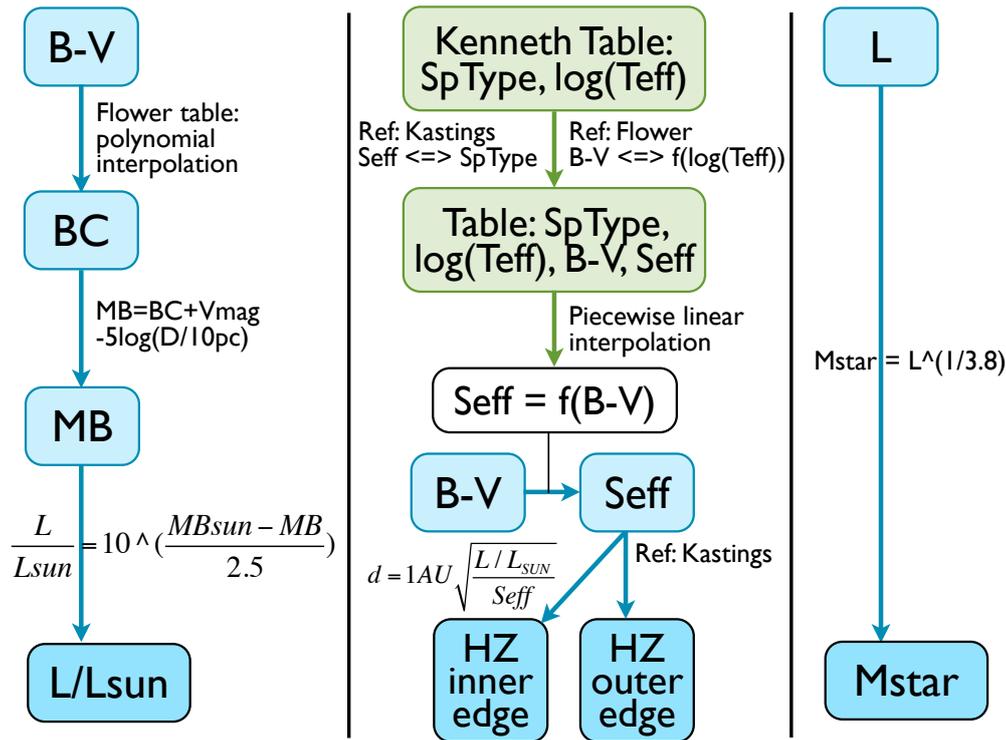


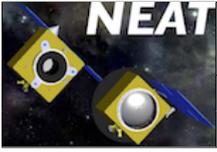
Figure 4: *How the luminosity, HZ boundaries and the stellar mass are derived. The blue boxes represent added columns to the catalogues and the green boxes represent data tables.*

□ How to obtain the distance

The distance is obtained directly from the hipparcos parallax. Add a new column for the distance by doing the following actions in Topcat:

- Select the Hipparcos main catalog (1 239 hip main, closest\_stars\_AFGKM subset) OR closest\_MS\_AFGKM\_targets\_topcat\_out.csv in the Table List.
- Click the tool **Display column metadata**
- Click the tool **Add a new column**
- Name it Dist and enter the following expression: 1000/P1x

□ How to obtain the luminosity



This work is based mostly on the following publication: Transformations from Theoretical Hertzsprung-Russell Diagrams to Color-Magnitude Diagrams: Effective Temperatures, B-V Colors, and Bolometric Corrections (Flower, 1996). In order to complete the first step, download the main sequence "Teff, B-V and BC" table from the same author (the data is accessible via Vizier). It is necessary to download the data and perform an interpolation manually because the  $BC = P(B-V)$  relation doesn't exist in the publication (it only gives the  $\log(\text{Teff}) = P(B-V)$  and the  $\log(\text{Teff}) = P(BC)$  polynomial relations).

I downloaded the data in a csv format and loaded it in SciDAVis (a free application for Scientific Data Analysis and Visualization). I performed with the latter a polynomial interpolation (order 10 i.e. 11 DoF). The resulting fit looks very good, the  $\chi^2$  value of the fit is 0.0004. The table 1 gives the coefficients of the polynomial ( $P = a_0 + a_1X + a_2X^2 + \dots$ ).

Coefficient	Value	Error
a0	-0,1784538504862980	0,0040439753600725
a1	2,9023636120332208	0,0358593758682318
a2	-12,3188141929402839	0,1468192147338562
a3	20,9310264552908833	0,7901640483937037
a4	-22,5026837177971153	1,7789678865741207
a5	63,6149604337794230	3,6915037287900767
a6	-177,5417213564442136	11,9138718837156965
a7	249,9150285256258428	16,1324092451378540
a8	-182,6519997569876352	11,0043353138590163
a9	67,4220567552053751	3,7755649838561576
a10	-9,9870014580361062	0,5193285281379360

Table 1: Polynomial coefficients of the relation  $BC = P(B - V)$ .

With these coefficients I can now complete the first step and add a BC column to the AFGKM subset:

$$\begin{aligned}
 BC &= -0.1784538 + 2.9023636 * \$39 - 12.318814 * \text{pow}(\$39, 2) + 20.931026 * \text{pow}(\$39, 3) \\
 &\quad - 22.50268371 * \text{pow}(\$39, 4) + 63.6149604 * \text{pow}(\$39, 5) - 177.541721356 * \text{pow}(\$39, 6) \\
 &\quad + 249.915028525 * \text{pow}(\$39, 7) - 182.6519997 * \text{pow}(\$39, 8) + 67.42205675 * \text{pow}(\$39, 9) \\
 &\quad - 9.98700145 * \text{pow}(\$39, 10)
 \end{aligned}$$

In the above expression, \$39 refers to the B-V column. It is required that one addresses this column by its ID because it has a "-" in its name.

One last thing to do is to verify that all the targets have a B-V index in the interpolated range. This can be done by creating a new subset with the following expression: `_8 && $39 < 1.8 && $39 > -0.35` and noting that the number of rows is unchanged (`_8` is the FGKM subset). From this point, it is very straightforward to obtain the luminosity in solar units, by creating two other columns:

$$MB = V_{\text{mag}} + BC - 5 * \ln(100 / P_{\text{lx}}) / \ln(10)$$

$$L = \text{pow}(10, (4.75 - MB) / 2.5)$$

Remark: The value of 4.75 in the formula above is the solar absolute bolometric magnitude.



□ How to obtain the boundaries of the habitable zone

First of all I extract some data from a table (Schmidt-Kaler, 1982). I need only the rows F0, G2 and M0 and the following columns: spectral type and log(Teff). Then I add the following columns: B-V, Seff IN, Seff OUT. All this is done manually by feeding the values in a spreadsheet (the final table size is only 3 rows by 7 columns). The table is shown on Fig. 5.

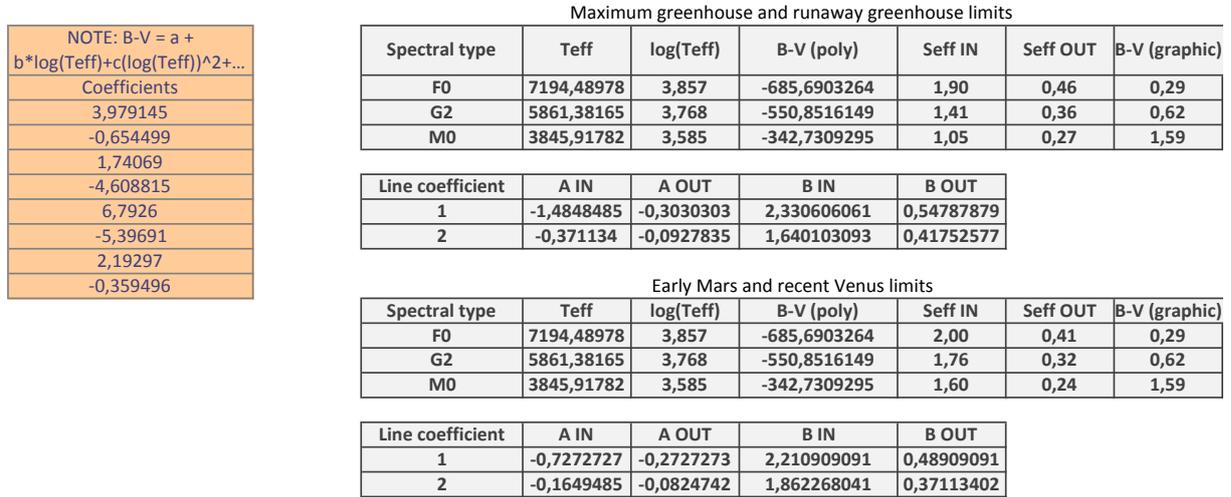


Figure 5: **The Seff versus B-V table.** The spreadsheet also gives the values of the A and B coefficients of the  $S_{eff}^{IN/OUT} = A_{1/2}^{IN/OUT} \times (B - V) + B_{1/2}^{IN/OUT}$  relations.

The values in the B-V column are found by using the relation  $B-V = P(\log(Teff))$  (Flower, 1996). Since there is obviously a problem when using the polynomial coefficients from the paper, I simply used the plots of the publication to graphically obtain the B-V values.

The values in the Seff columns are found in a publication from Kastings (Kasting et al., 1993). They performed climate simulations to determine the width and the locations of the HZ. Note that the concept of HZ remains very fuzzy, as it depends strongly on which model and hypothesis one relies. Here we use two sets of limits: the runaway greenhouse/maximum greenhouse limits and the recent venus/early mars ones.

These limits were calculated for F0, G2 and M0 spectra and should be updated for all spectral sub classes (Kaltenegger, Segura and Kasting in prep). So when more values for Seff will become available, it will be possible to add rows to the table and perform a polynomial interpolation. In the meantime, with only 3 data points available to calibrate the relation  $(Seff IN, Seff OUT) = f(B-V)$  I chose to use a piecewise linear function  $(Seff = A \times (B - V) + B)$ .

Assuming the following notations for Seff IN in the table:

B-V	Seff IN
$x_1$	$y_1$
$x_2$	$y_2$
$x_3$	$y_3$

the coefficients A et B are given by:  $A_1 = (y_2 - y_1)/(x_2 - x_1)$  and  $B_1 = (y_1x_2 - y_2x_1)/(x_2 - x_1)$ , for the first segment. Similarly we have  $A_2 = (y_3 - y_2)/(x_3 - x_2)$  and  $B_2 = (y_2x_3 - y_3x_2)/(x_3 - x_2)$  for the 2nd segment. The coefficient for Seff OUT are calculated in the same way. After replacing  $A_{1/2}$  and  $B_{1/2}$  by their numerical values (see Fig. 5), we obtain the following expressions:



SeffIn1  $\$39 < 0.62?$   $-1.484 * \$39 + 2.330$  :  $-0.371 * \$39 + 1.640$

SeffOut2  $\$39 < 0.62?$   $-0.273 * \$39 + 0.489$  :  $-0.082 * \$39 + 0.371$

SeffIn1  $\$39 < 0.62?$   $-1.484 * \$39 + 2.330$  :  $-0.371 * \$39 + 1.640$

SeffOut2  $\$39 < 0.62?$   $-0.273 * \$39 + 0.489$  :  $-0.082 * \$39 + 0.371$

The columns indexes 1 and 2 are for different types of habitable zone limits (they are not related to A1/A2 B1/B2 which simply results from the piecewise nature of the fits). For all the columns related to the HZ, the index 1 is for the runaway greenhouse/maximum greenhouse limits whereas the index 2 is for the recent venus/early mars limits ((Kasting et al., 1993)).

Finally the inner and outer edges of the habitable zones are given by the following formula(Kasting et al., 1993):

$$d = 1AU \times \left( \frac{L/L_{\odot}}{S_{\text{eff}}} \right)^{0.5} \tag{2}$$

So the expression for the edges of the HZ columns will be:

HZIn1  $\text{pow}(L/\text{SeffIn1}, 0.5)$

HZIn2  $\text{pow}(L/\text{SeffIn2}, 0.5)$

HZOut1  $\text{pow}(L/\text{SeffOut1}, 0.5)$

HZOut2  $\text{pow}(L/\text{SeffOut2}, 0.5)$

### □ How to obtain the star mass

The star mass can be obtained in at least two ways: the first way is to use the information contained on the exoplanet encyclopedia. The drawback of this method is that the mass will only be available for the NEAT target stars that have already known exoplanets around them (92 stars out of 385). The other method is to use the relation  $L = M^{3.8}$ . The problem with this second method is that its very approximative. It is nevertheless necessary to estimate the mass of all the target stars in order to know how much integration time must be allocated to reach the habitable zone. The column is created as follows:

Mass  $\text{pow}(L, 1.0/3.8)$

We know have all the columns needed to derive the astrometric signal at the HZ boundaries and the Earth flux equivalent position, the inner working angles of the respective circular orbits and their orbital periods. All columns are listed in appendix B. Only the astrometric signals are useful for the next section (observation strategy and science yields), but the other columns can be useful to explore various aspect by plotting the data, such as inner working angle for direct imaging or deriving RV amplitude signals.

## 2.2 – Time allocation schemes and science yields

In this section we explore different ways of distributing the observation time between the targets and the effect on the science yield of the mission.



## 2.2.1 – Relation between minimum detectable signal and integration time

In order to carry on our analysis we need a relation between the minimum astrometric signal detectable ( $A_{\min}$ ) and the integration time ( $T$ ) for each target. We will need the relation in both ways:

- With  $T = f(A_{\min})$ : we can rank the targets by order of increasing “difficulty”. The difficulty is defined as the integration time required to detect a terrestrial mass planet at the inner edge of the habitable zone.
- With  $A_{\min} = f(T)$ : for a given integration time, we know the detection limit. For any pair of target star/simulated planetary system, we can easily determine which planets would be detected.

An error model for NEAT, with the knowledge of the visible magnitude of the target star and the magnitude equivalent to the sum of associated reference stars allows us to find for each target a relation between its observation time and the minimum detectable signal. The model is mainly based on the photon noise limit of the NEAT instrument. The photon noise limit can be estimated with a spreadsheet, using fundamental laws (see reference document: `neat_photon_limit.xls`). The steps used to estimate the photon limit are:

1. The total visible photon flux of the Sun ( $\text{photon.s}^{-1}.\text{m}^{-2}$ ) is calculated. This is done using the plank function and a numerical integration among 41 wavelength channels, from 400 to 810 nm. After the stellar spectral irradiances ( $\text{W.m}^{-1}$ ) are converted into a photon flux for each channel, the sum of all channels gives the total photon flux.
2. The photon flux is expressed as a function of distance (in pc) from the Sun and the mirror size.
3. The flux is converted into a photo-electron flux ( $\text{e}^{-}.\text{s}^{-1}$ ).
4. The fundamental limit on the (angular) accuracy of the centroid is given by:  $\frac{0.42\lambda_{\text{eff}}}{D\sqrt{N_e}}$ . The factor 0.42 is the standard deviation of the best Gaussian fit to an Airy spot central lobe (this is an optimistic assumption, we ignore the wings). In practice the PSF is polychromatic: longer wavelengths have larger standard deviations and vice versa. The photon budget per channel indicates that red channels have more photons. To safely take into account out these effects we use the maximum wavelength (810 nm) as a reference.

To verify the method, I have used the same star brightness than the NEAT error budget: I assume a G2 target star of  $V_{\text{mag}} = 6$  and 6 references stars of  $V_{\text{mag}} = 11$ . The reference stars total flux is equivalent of one star of  $V_{\text{mag}} = 9.05$ . As a common reference I also assume an integration time of one hour. I find that the photon noise of the target star ( $\sigma_{\text{target}}$ ) and the reference stars ( $\sigma_{\text{ref}}$ ) are respectively  $0.09 \mu\text{as.h}^{-0.5}$  and  $0.36 \mu\text{as.h}^{-0.5}$ .

There are significant differences with the NEAT error budget, they can explain why the photon noise limits we find are smaller. Indeed, the error budget yields:  $\sigma_{\text{target}} = 0.12 \mu\text{as.h}^{-0.5}$ ,  $\sigma_{\text{ref}} = 0.61 \mu\text{as.h}^{-0.5}$  for references, plus  $\sigma_{\text{systematics}} = 0.44 \mu\text{as.h}^{-0.5}$  for instrument systematics. The notable differences are:

- K0 stars (target and references) are used instead of G2



- Wavelength integration is over 8 channels, from 450 to 950 nm
- CCD quantum efficiency is specified per channel, not only averaged
- A lot of other error sources are taken into account, instrumental and astrophysical errors like metrology calibration, stellar aberration...
- The method to derive the centroid accuracy from the number of collected photon is different

For our purposes, we will use the more conservative values of the NEAT error budget. We will adjust the multi epoch accuracy  $\sigma_f$  for the brightness of target (baseline flux  $F_{\text{target}}^0$ ) and reference stars (baseline flux  $F_{\text{ref}}^0$ ) and for the integration time  $T$  (in hours) in the following way:

$$\sigma_f^2 = \left[ \sigma_{\text{target}}^2 \frac{F_{\text{target}}^0}{F_{\text{target}}} + \sigma_{\text{ref}}^2 \frac{F_{\text{ref}}^0}{F_{\text{ref}}} + \sigma_{\text{systematics}}^2 \right] \times \frac{1}{T} = \sigma_{\text{target/h}}^2 \frac{1}{T} \quad (3)$$

The fundamental point is the square root increase of the accuracy with the number of collected photons, either because we have more flux or we integrate longer. Moreover, we consider that most systematics can be averaged out by taking many data points, eventually several per epoch. We can now determine for each target the sum of errors  $\sigma_{\text{target/h}}$ , normalized for a one hour integration. In Eq. 3: the value of  $\sigma_{\text{target/h}}$  as a function of the target and reference equivalent magnitudes is:

$$\sigma_{\text{target/h}} = \sqrt{10^{\frac{V_{\text{mag,target}}-6}{2.5}} \sigma_{\text{target}}^2 + 10^{\frac{V_{\text{mag,ref}}-9.05}{2.5}} \sigma_{\text{ref}}^2 + \sigma_{\text{systematics}}^2} \quad (4)$$

The column is created as follows:

SigmaH  $\text{pow}(0.0144 * \text{pow}(10, (V_{\text{mag}}-6)/2.5) + 0.372 * \text{pow}(10, (EqRefMag-9.05)/2.5) + 0.194, 0.5)$

The minimum detectable astrometric signal is for an SNR of 6:

$$A_{\text{min}} = 6\sigma_f = 6\sigma_{\text{target/h}} \sqrt{\frac{1}{T}} \quad (5)$$

The inverse relation is:

$$T = 36 \frac{\sigma_{\text{target/h}}^2}{A_{\text{min}}^2} \quad (6)$$

The column for the target “difficulty” (integration time needed for an exo-Earth at the inner edge of HZ) is created with:

TexoEarth  $36 * \text{pow}(\text{SigmaH}, 2) / \text{pow}(A_{\text{In1}}, 2)$

Note that I have made implicit assumptions to obtain this simple result. I have forgotten the complex considerations about how to optimize the number of visits, their duration and the number of slews required (these are addressed in the manuscript). I retain only the total time of effective observation per target (excluding slew time from this total). The approximation can be made because the number of epochs per target, number of data points per epoch or integration time per data point all increase the accuracy roughly as the power 0.5 when the number of independent measurements is larger than the number of free parameters. Concerning systematic errors, we assume that we have enough epochs per target and data points per epoch to verify the number given by the error budget. There are a lot of possible caveats to this simplistic approach, here is a list of important ones I can think of:



- We assume that systematic errors are averaged by having many epochs. On a real instrument, this will not be the case for all kinds of systematics.
- The NEAT error budget assumes the same number of visits for every target. In a real mission, the number of visits could be different for each target and could also depend on the number of planet detected before the end of the mission. If we find a very interesting system, no doubt we will come back more often before the mission is over. Moreover, in the manuscript we have determined that the optimal number of visits depends on the final accuracy desired.
- We have not considered the impact of giant exoplanets ( $M > M_{\text{Jupiter}}$ ) around reference stars. They would increase the total number of parameters on the global solution. There is also the option of removing noisy reference stars from the global solution (in this case the reference equivalent magnitude increases). We expect to find giant planets around about 10% of stars, so depending on the rejection criteria, the number of useful reference stars might be slightly lower than expected and/or some reference stars will be more noisy.

## 2.2.2 – Impact of the time allocation scheme

Having decoupled the total integration time from the number and durations of visits, from now on I only consider the total effective observing time (slew time not included). Based on the content of the NEAT proposal, I fix the total effective observation time at  $T_{\text{tot}} = 20000$  hours (subtracting about 10% of the total time from the proposal to account for slew time). I analyze different scenarios for the time allocation scheme. This is done via a matlab script: `observational_strategy_yield.m`.

1. All NEAT catalog targets (455 stars) are selected. The observation time is distributed equally.
2. The 200 easiest stars are selected. The observation time is distributed equally.
3. The 100 easiest stars are selected. The observation time is distributed equally.
4. Time allocated equal to TexoEarth: more observation time is allocated for harder targets. Cumulative time stops at  $T_{\text{tot}}$ .
5. The observation time is distributed to reach specific mass detection thresholds in habitable zone. Several categories are defined manually. For the closest stars, a threshold of  $M_{\text{lim}} = 0.5 M_{\oplus}$  at the inner edge of the HZ is used to calculate the time allocated. When the time needed becomes too large, the threshold is changed to  $1 M_{\oplus}$  (2nd category). Same thing for the last category ( $M_{\text{lim}} = 5 M_{\oplus}$ ). The boundaries between the categories are placed so that all the observation time is used after at least 400 targets.
6. Time allocated equal to  $\frac{C_1}{\text{TexoEarth}^\alpha}$ , with a maximum time threshold of  $\frac{T_{\text{tot}}}{C_2}$ . So slightly more time is allocated for easier targets. The constant  $C_1$  determines at which target we stop allocating a constant fraction of time and we start to decrease. The constant  $C_2$  determines the maximum time threshold for easy targets. The constant  $\alpha$  determines the speed at which we decrease the allocated time with increasing difficulty. The numerical values are:  $C_1 = 957$ ,  $C_2 = 250$  and  $\alpha = 0.5$ .

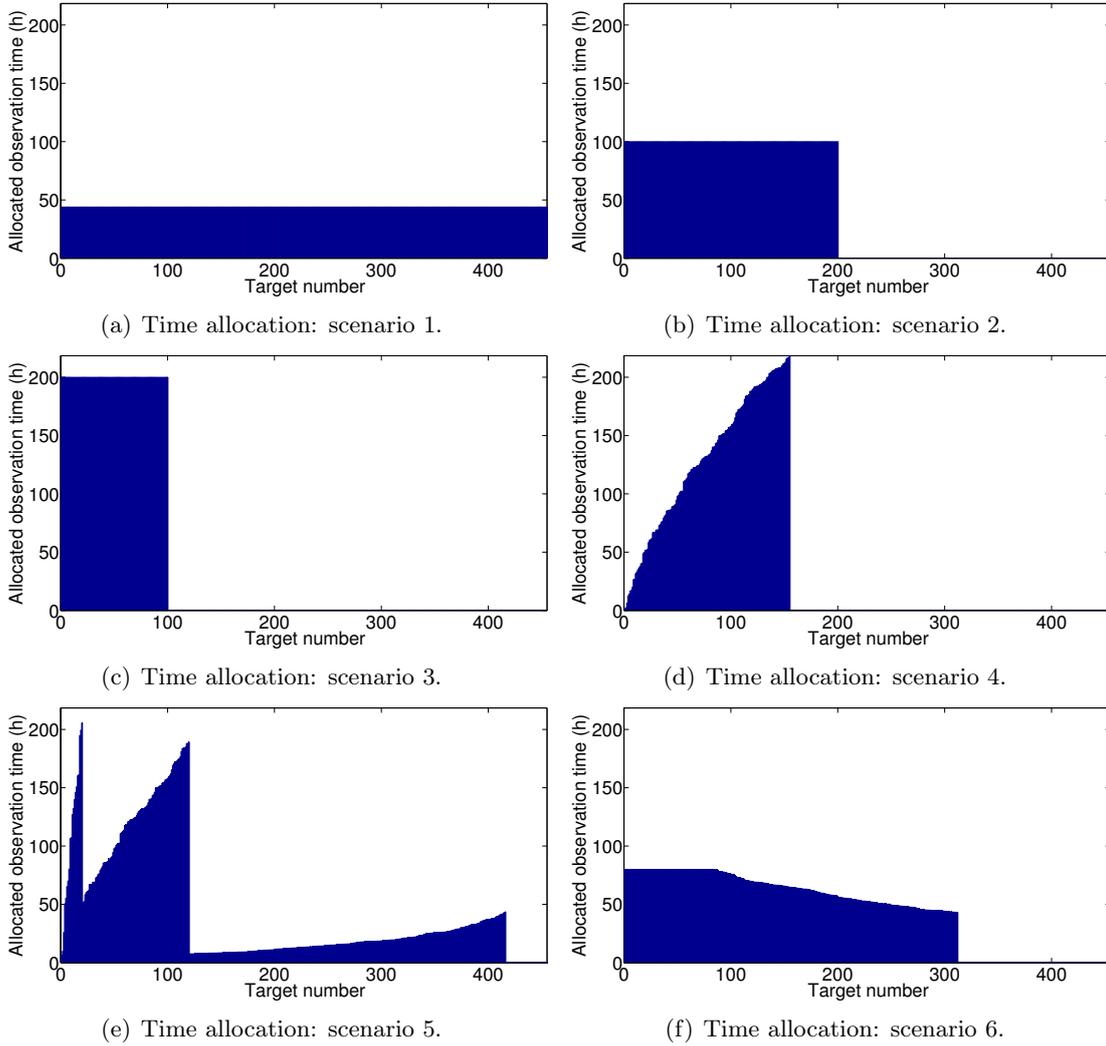
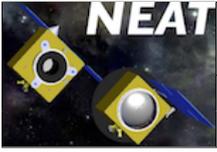


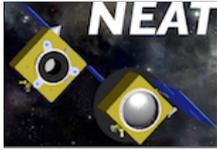
Figure 6: Time distribution per target for 6 different time allocation schemes.

Figure 6 shows the time distribution corresponding to each scenario.

Figures 7 and 8 are bar graphs of predicted exoplanet yields. The yield simulation is based on a list of 344 synthesized planetary systems kindly provided by Yann Alibert, they were originally generated for the NEAT double blind test (for non-blind test runs). For detailed informations about the population synthesis model, refer to (Alibert et al., 2013). The fundamental assumptions of the planet synthesis model are:

- Maximum of 10 planets per system
- Star mass of  $1M_{\odot}$
- Alpha disk model (1+1D model, azimuthal symmetry,  $\alpha = 2 \times 10^{-3}$ )
- Migration model based on (Mordasini et al., 2011) and (Dittkrist et al., 2014)
- Disk properties (mass and lifetime) should be similar to observations

The yields are obtained by averaging 100 random permutations. With only one permutation the results vary too much (by several percents for a yield of about 8% for terrestrial planet in HZ, for example). The variability is introduced when we randomly assign the planetary



systems to the target stars: the permutation will be more favorable if more planets/more interesting planets end-up around easy/close stars). With 100 permutations, the deviation between two runs is less than 0.1%.

The 6 bar plots highlight the trade-offs between the allocation schemes. The difference between **1**, **2** and **3** is that we increasingly concentrate more time on easy targets. **1** yields more giant planets than **2**, but slightly less terrestrial ones. However when we reduce the number of targets too much (**3**), we obtain less giant but also less terrestrial planets. In this case we have already spent a fair amount of time on the easy targets in **2**, so the additional observation time around these do not compensate for the planet loss of the “medium” targets (ranked 100 to 200).

The idea behind schemes **4**, **5** and **6** is to see whether having an unequal distribution time can yield better results. We test several possibilities:

- In **4**, I choose to spend more time on hard targets, by having for each star an observation time that allows the detection of an Earth at the inner edge of the habitable zone. The advantage is that we avoid using more time than necessary on easy targets. But this also causes the allocation time to increase very quickly (see Fig. 6(d)) and I run out of time around the target number 160. The overall result is not very good compared to **5**.
- In **5** the approach is similar, but I mitigate the problems encountered in **4** by having several mass detection thresholds. When the allocated time has become too large, the threshold is changed to a higher mass value. The final yield difference with **4** is sensible. However if we look at the distribution of allocated time, the jumps when we change threshold feel unnatural: we are using a lot of time for targets just before the peak and not enough for targets just after, which are only slightly harder. This approach (**5**) was the one used in the NEAT proposal.
- In **6**, the idea is to mainly spend more time on easy targets, but with some caps to avoid wasting too much time. The first time cap is flat, the second one is equal to the time needed to detect a  $0.5 M_{\oplus}$  planet in HZ. In practice the second cap make almost no difference (very few targets concerned), so I ignore it and only apply the first one. For harder targets, the time spent is proportional to  $\frac{1}{T_{exoEarth}^{\alpha}}$ . Around harder targets, we reduce the time spent: we only hope to detect massive super-Earth and giant planets.

In the end the simulation indicates that **6** finds the most terrestrial exoplanets. It seems to be a good compromise for splitting time between easy and hard targets, because a significant fraction of the giant planets are detected (55%), not very far from the maximum obtained with **1** (75%). Figure 9 is a mass/period diagram of the detected exoplanets for scenario **6**. The plot is obtained for a single permutation (to avoid having fractions of planets !).

In the case of the new ESA call for proposals (M4, 2015), the former NEAT consortium will propose a new mission called Theia, with revisited science objectives. What was called the main science program in NEAT will shrink to a smaller portion of the total available telescope time and more alternative science programs will be proposed. Also, the previous goal of detecting gas giants, down to about  $10 M_{\oplus}$  (so called mini-Neptunes), and up to 20pc, has become less interesting as the number of known exoplanets has increased to more than 2000 and other missions already in preparation will detect a lot of these planets. The critical point remains the exo-Earths (and habitable super Earths) around nearby stars, especially for stars latter than M. This yield simulator will be very useful to determine to which extent the observation time can be reduced (for a given objective in number of detections) and what is the best way to distribute the observation time among the targets.

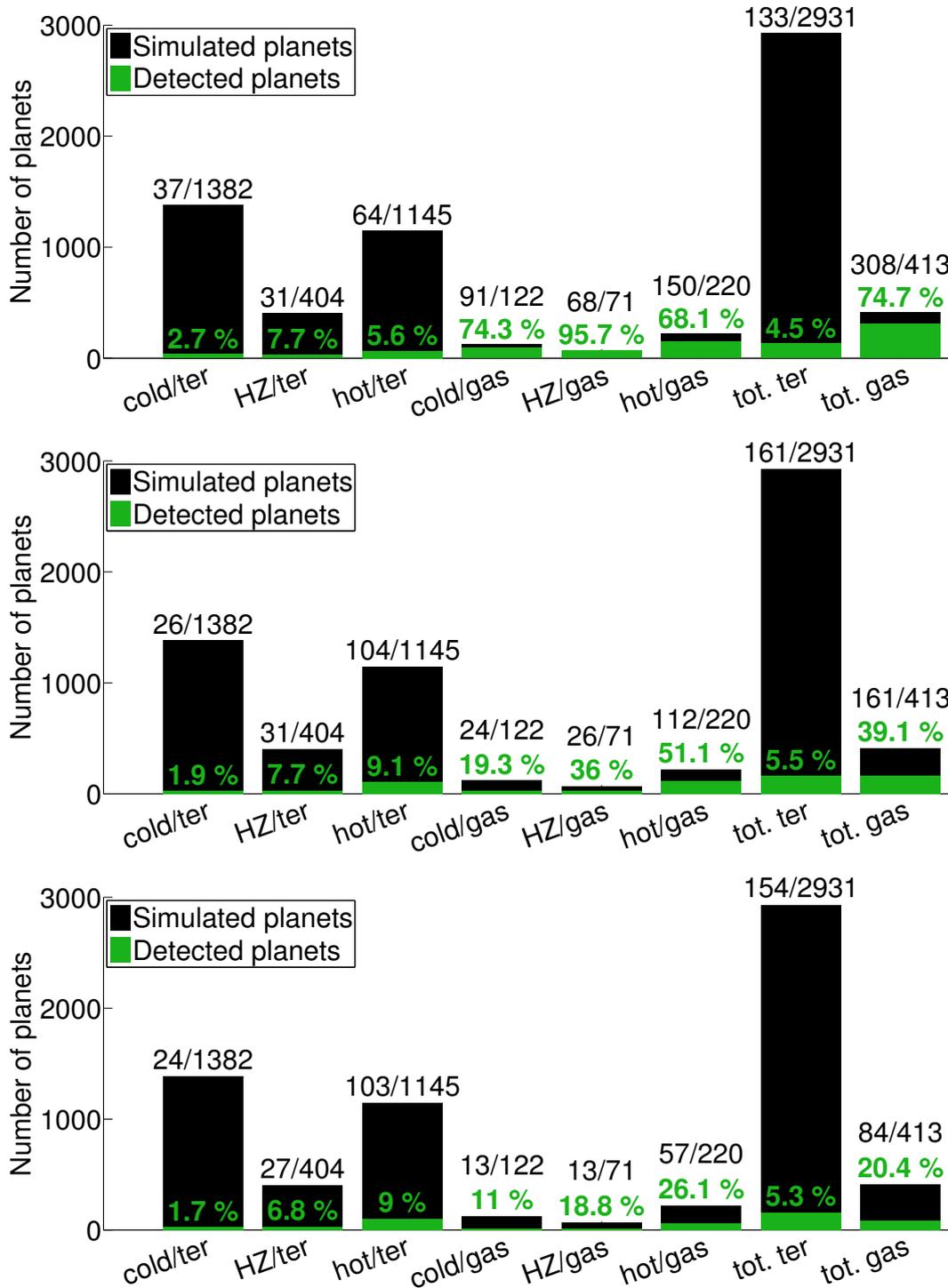
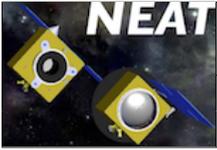


Figure 7: Exoplanet yields for allocation schemes 1, 2 and 3 (from top to bottom).

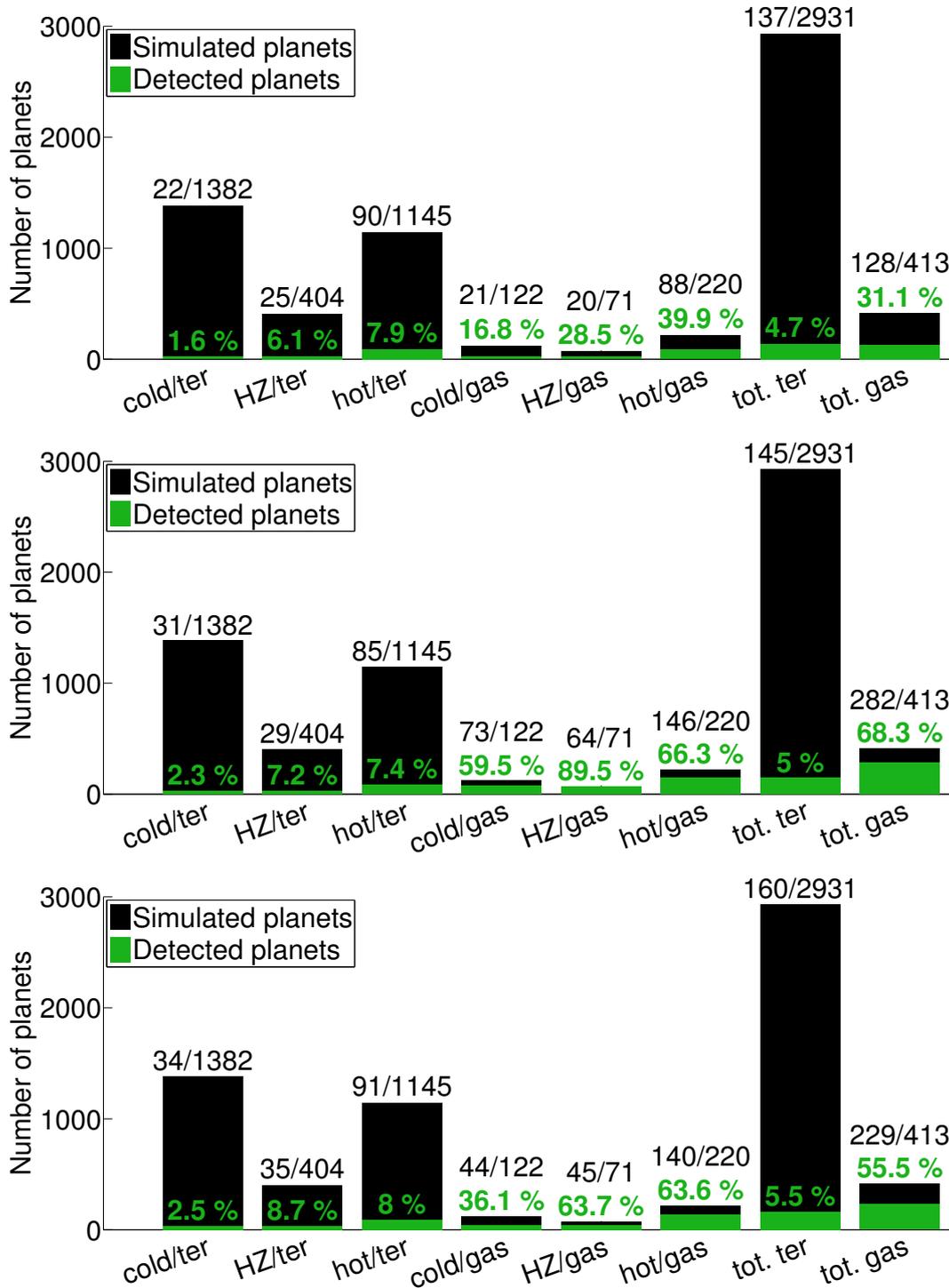
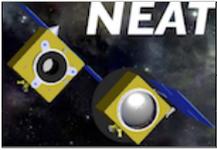


Figure 8: Exoplanet yields for allocation schemes 4, 5 and 6 (from top to bottom).

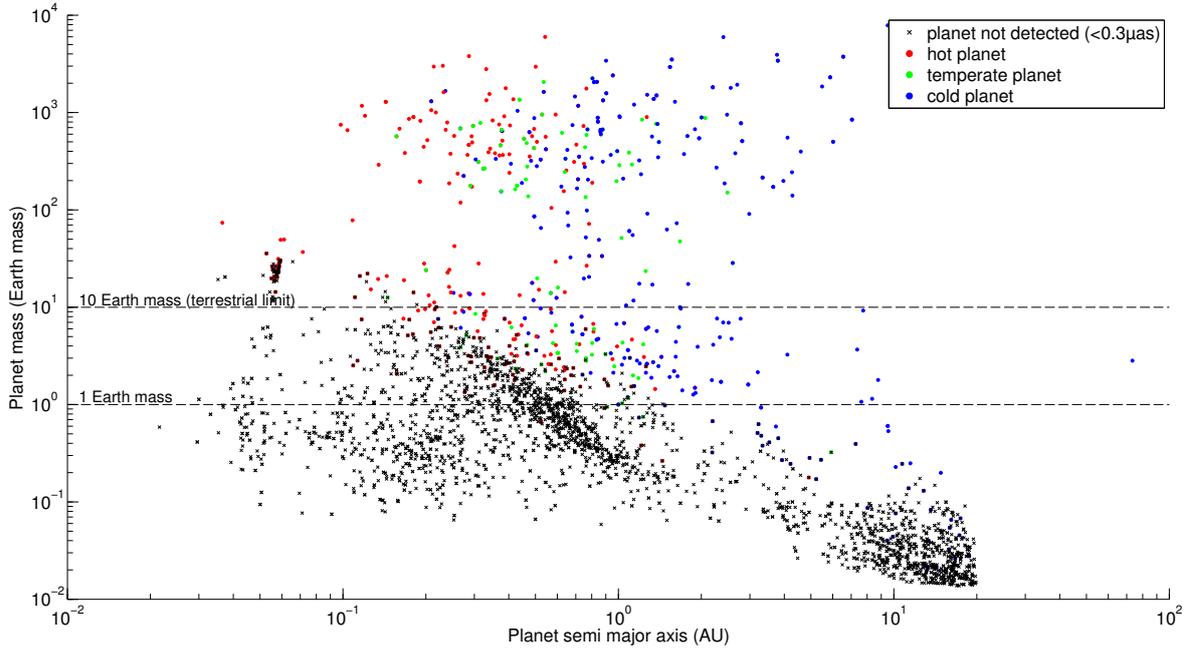
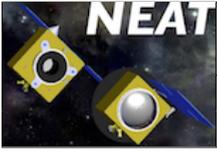


Figure 9: *Mass period diagram of the NEAT mission simulated exoplanet yield.*

### 2.3 – Astrometric signal in HZ versus stellar mass

Overall, the astrometric signal ( $A$ ) in HZ increases with the host stellar mass ( $M$ ). This is the result of several competing effects:

1. The star luminosity increase with the mass:  $L \propto M^{3.8}$
2. The orbital distance of the HZ increase with the luminosity:  $R(\text{HZ}) \propto L^{0.5}$
3. The albedo of the planet is higher for bluer light (it moves the HZ slightly closer to the star)
4. The ratio  $\frac{M_{\text{planet}}}{M_{\text{star}}}$  is smaller

The net effect of 1 and 2 is an increase of the astrometric signal ( $A \propto M^{1.9}$ ). This is somewhat attenuated by 3 (but it is a weak effect). We can take the albedo into account by defining an effective luminosity  $L_{\text{eff}} \approx L^{-\alpha}$  and we obtain for 1, 2 and 3:  $A \propto M^{1.9-\alpha/2}$ . The effect of 4 is to subtract 1 to the total power. So we finally obtain:  $A \propto M^{0.9-\alpha/2}$ . This relation is approximative, as we have used several approximated power laws: the  $L = f(M)$  relationship breaks down for M stars (the limit is around  $0.43 M_{\odot}$ ) and the albedo effect is in our case modeled by a piecewise linear function. In reality the exact albedo effect will be non linear and specific to each planet.

Figure 10 shows the astrometric signal in HZ versus the stellar mass. In order to separate the effects of the distance and the stellar mass, Fig. 11 shows the astrometric signal in HZ ( $\mu\text{as}$ ) multiplied by the distance to the Sun (pc) versus the stellar mass. This is equivalent to moving all the targets at a distance of 1 pc from the Sun.

Figure 11 confirms that the spectral effect is weak and that the astrometric signal in HZ increase with stellar mass, roughly as  $M^{0.63}$  ( $\Rightarrow \alpha = 0.54$ )<sup>3</sup>. This result is important, it tells

<sup>3</sup>Extrapolating from the following data points: (M=0.42, A=2.0) and (M=2.0, A=6.0).

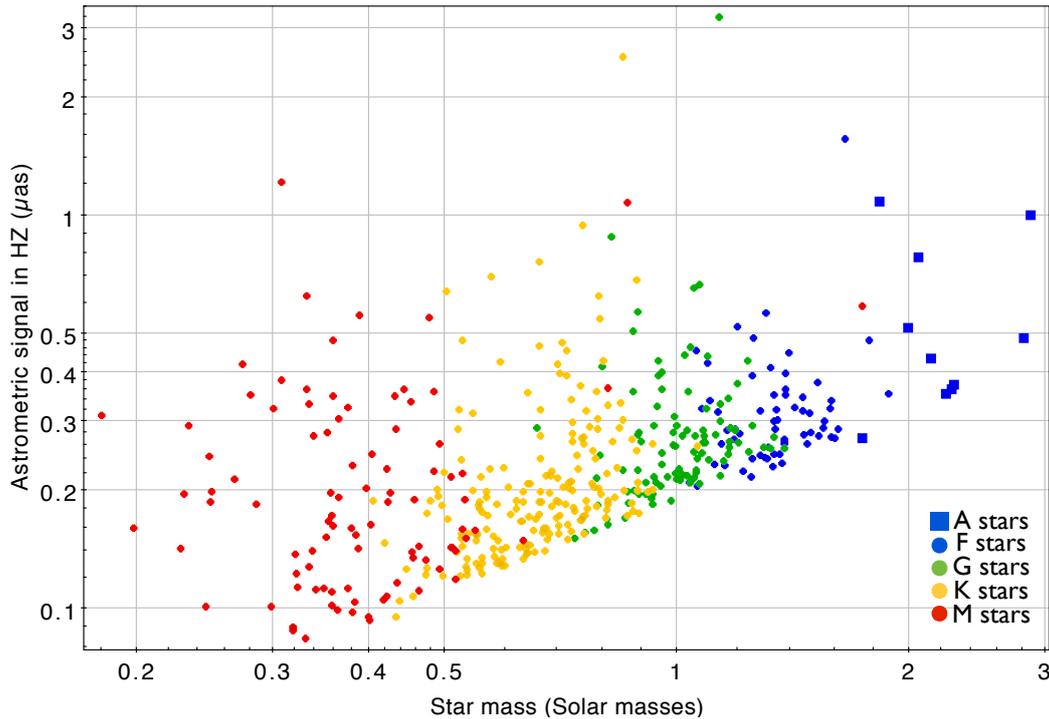
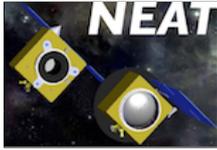


Figure 10: *Astrometric signal in HZ versus star mass for the NEAT targets, by spectral type.* The different areas corresponding to spectral type from M to A are easily distinguishable. The bottom line is for stars close to 20 pc, the vertical dispersion is caused by the distance. Low hanging fruits (and their spectral types) are easy to spot on this plot. The two stars at the very top (G and K) are the Alpha Centauri system. A few blatant spectral miss classifications (for example 2 M stars with  $M > 0.8M_{\odot}$ ) are visible on the plot.

us that astrometry is complementary with radial velocities when searching for planets in the habitable zone. Indeed, for radial velocities, exoplanet detection in HZ becomes harder when the stellar mass increases. There is an important caveat that must be noted here: the increase in astrometric signal with stellar mass goes with a period increase. The temporal baseline must be larger than the period to fully take advantage of the increased signal amplitude. For stars later than A, the orbital period at the inner edge of the HZ becomes larger than 5 years (the nominal duration of the NEAT mission).

## 2.4 – Crossmatch with “The Extrasolar Planets Encyclopaedia”

Now we will load the “Interactive Extra-solar Planets Catalog”, from this website: <http://www.exoplanet.eu/>. This catalog contains all the known exoplanet candidates, we will use it to perform a crossmatch with the NEAT catalog and pinpoint which NEAT targets already have exoplanet candidates.

- Go to the website, then to the “All catalogs” page, then click on the “All fields” link
- Download the data as a **VOTABLE**
- Open the **Load New Table** dialog box via the menu **File > Load table**
- Click on the **System Browser** icon and load the table
- Create a subset `closest_exoplanets` with the condition: `star_distance < 20`

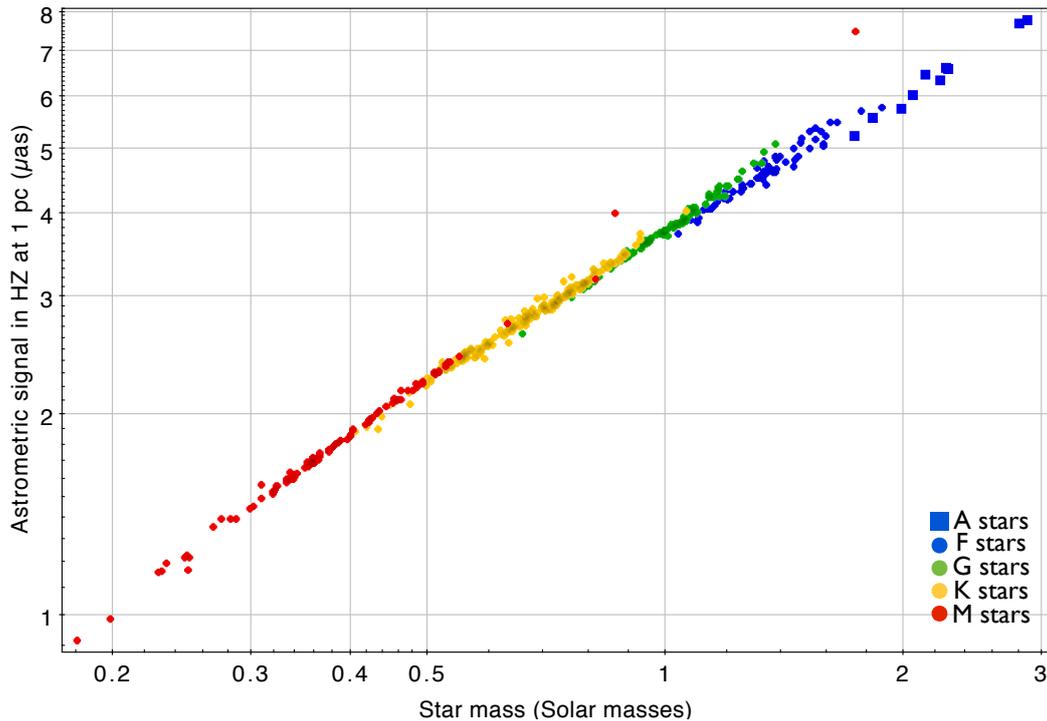
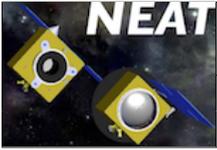


Figure 11: *Astrometric signal in HZ times distance versus star mass for the NEAT targets, by spectral type.* Multiplying by the distance has normalized the astrometric signal to the signal seen at a distance of 1 pc.

- Select the `closest_exoplanets` subset in the exoplanets catalog.
- Select the `closest_stars` subset in the Hipparcos main catalog.
- Crossmatch the catalogs via the menu **Joins > Pair Match**, using the following parameters:
  - **Max Error** of 1 arcmin.
  - **Table 1**: Hipparcos main catalog
  - **RA column** (for Table 1): RA(ICRS) - degrees
  - **Dec column** (for Table 1): DE(ICRS) - degrees
  - **Table 2**: Exoplanet catalog
  - **RA column** (for Table 2): St\_Right\_Asc - degrees
  - **Dec column** (for Table 2): St\_Decl - degrees
  - **Match selection**: Best match for Table 2 row
  - **Join type**: 1 and 2
- Call the table `Exoplanet.eu closest stars crossmatch`

BEFORE performing the crossmatch, make sure you have selected the `closest_stars` and `closest_exoplanets` subsets in their respective catalogs in the **Table List**. The result of the crossmatch are the exoplanets closer then 20 pc that orbit nearby stars of the Hipparcos catalog. The crossmatch result has only 113 rows instead of the 126 expected most likely



because some host stars are absent of the Hipparcos catalog (which is not complete up to 20 pc for K and M stars).

Repeat the crossmatch operation with the hipparcos `closest_stars_AFGKM` subset and name the new table `Exoplanet.eu closest MS AFGKM stars crossmatch`.

Tip: if you change the **Match selection** parameter to "Best match for Table 1 row", you will obtain the number of stars with known exoplanets candidates in the subset (instead of the number of exoplanets candidates around stars of the subset). Table 2 sums up the row counts in all the cases.

Table 2: *Number of stars with known exoplanets (updated 02/09/2014)*. Two different stars subset are compared: the first one is all Hipparcos stars closer than 20 pc (934 stars) and the second one is the NEAT catalog of targets (455 stars). The "Exoplanet count" tells how many exoplanets have been successfully associated with a Hipparcos host star, the "Star counts" tells how many Hipparcos stars host at least one known exoplanet. There is a small fraction of exoplanets for which no host star is found, they are most likely M stars that are too faint for Hipparcos.

Hipparcos subset	Exoplanet count	Exoplanet count (%)	Star count	Star count (%)
All stars < 20 pc	113/126	90%	64/934	6.9%
Main sequence AFGKM stars < 20 pc	74/126	59%	39/455	8.6%

Remark: apparently the RA and DEC coordinates are from different epochs for each catalog, hence the very wide tolerance used for the crossmatch. This could be another explanation for missing rows. Nevertheless, the results are visually satisfying.

Table 2 tells several interesting facts:

- Among the 1821 exoplanets candidates, only 113 are located at less than 20 pc.
- The fraction of main sequence stars that have known exoplanets is not significantly higher than the fraction for all stars. We have not put a particularly strong emphasis on main sequence stars in past surveys.
- We expect on average several planets per star: many more exoplanets are yet to be discovered in our solar neighborhood.

We will now create a 3D map of the stars and planets, by matching the angular and radial positions the of objects with their real positions in space.

Remark: the following step works best with Topcat version V3. On later version, the 3D plot with spherical grid has been removed.

To plot the data, click the tool **Spherical polar scatter plot**. Select the Hipparcos main catalog in the **Table** field. Chose a style for each subset (each spectral type), add a radial axis and enter the value **1000/Plx**: as the Hipparcos Plx value is the parallax in mas, this will yield the distance of the star in parsecs. Click on the **Add a new data set** button, select the result of the crossmatch with all the stars and process it in a similar way. For the radial value, you can directly enter the distance column: **star\_distance**. You should also deactivate the fog option. The resulting plot is shown in Fig. 12.

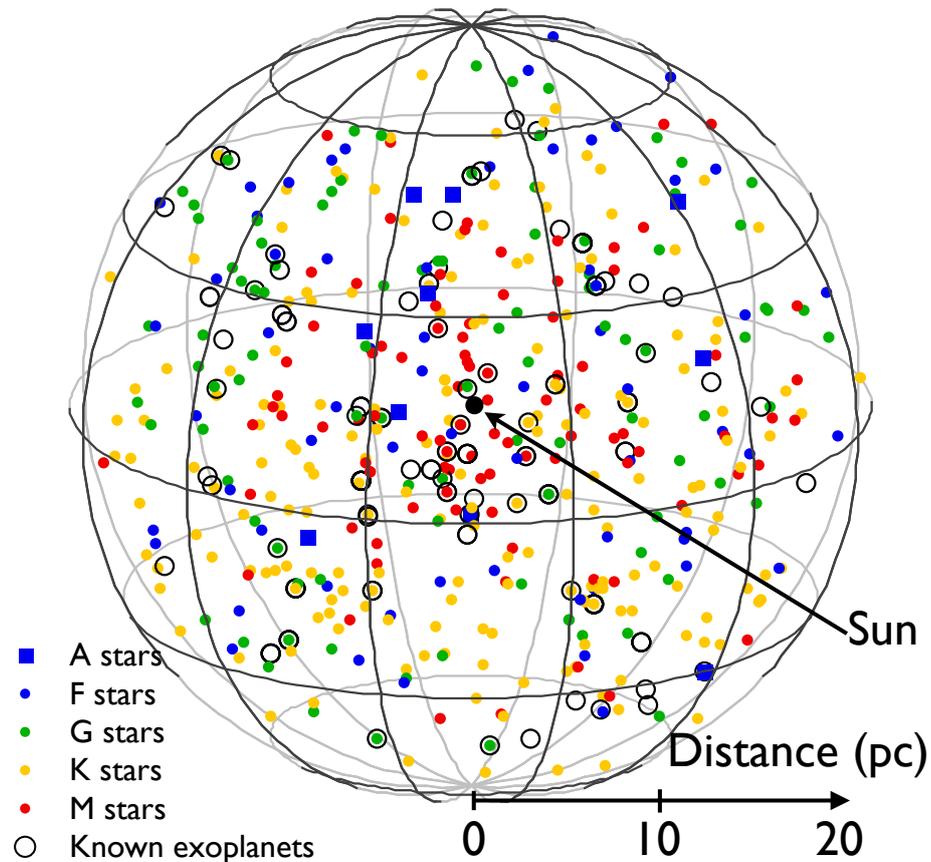
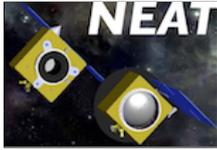


Figure 12: *3D representation of the catalog of the NEAT catalog of targets, by spectral type (updated 02/09/2014).* The stars are crossmatched with the list of known exoplanets, the ones with at least one known exoplanet in orbit are circled. There are 10 A, 62 F, 97 G, 192 K and 94 M stars.

As expected, only a small fraction of the stars are circled. Notice that they are also some circles without star. These are exoplanets around star that are absent of the Hipparcos catalog or they are crossmatch errors.

If you haven't already done it, I advise you to save the full session now: use the menu **File** > **Save Table(s)/Session**. Make sure to select the tab **Session** and the format **fits-plus** as the output format to save the whole session.

**WARNING:** tough saving the full session in a fits-plus format will save the subsets, the algebraic expressions defining each subset will be discarded upon loading of the session. The plots will be lost as well, so you should export them as images before you close Topcat.



## 2.5 – Checking the sample completeness

By looking at the 3D representation of the NEAT catalog, it is clear that the M stars are not evenly distributed. They are mostly located within 10 pc of the sun. There is obviously a problem of completeness for the M stars: the most distant M stars seems to be absent of the Hipparcos catalog because their magnitudes are too high.

So I ask the following: are all the F, G and K stars closer than 20 pc present in the Hipparcos catalog ? To answer this, one can simply plot the histogram of the number of stars versus the distance.

Add new subsets of main sequence F,G,K and M stars closer than 100 pc in the main Hipparcos table. For example for F stars:

```
-name: close_MS_F_stars
```

```
-condition: Plx > 10 && matches(SpType, "^F.*(IV-V|[^I]V)")
```

Assuming an homogeneous distribution of the stars in the solar neighbourhood, the number of stars in a radius  $r$  is:  $n = \alpha \times 4/3\pi r^3$ , where  $\alpha$  is the mean density of stars, thus:  $dn = \alpha \times 4\pi r^2 dr$ . An histogram of the stars count versus the distance (in log axis) is in fact a representation of  $\log \frac{dn}{dr}$  versus  $\log r$ . Since we have:  $\log \frac{dn}{dr} = 2 \log r + 2 \log 4\pi\alpha$ , we expect to see a slope of 2 for the short distances, where the Hipparcos catalog is complete. The breakpoint in the slope indicate where the catalog is not complete anymore.

In order to see the breakpoints, other subsets are defined. The procedure is the same than the one that was used to define the NEAT catalog, except that the distance is limited to 1kpc (as a result almost all the Hipparcos stars are included). The final histograms are shown in Fig. 13 and 14

### □ Conclusion

One can assume near completeness of the Hipparcos catalog for the FGK main sequence stars (the completeness breakpoints are all well further than 20 pc). The same is not true for M stars, because they are much fainter. These results are consistent with the existing literature, a completeness study was performed to assess the completeness of the Darwin catalog (which was also build from the Hipparcos catalog): “Excluding the M stars from the estimate leads to a high completeness at 20 pc and 30 pc of 97% and 76% respectively, based on the estimated number of stars per volume using the 10 pc RECONS data set.” (Kaltenegger et al., 2010).

The incompleteness issue is of little practical importance for the main science program because Hipparcos is 99.9% complete to magnitude 10, and measures around fainter stars would be strongly impaired by high photon noise (it becomes larger than reference photon noise). To summarize: the NEAT catalog is volume limited for F,G and K stars at 20 pc, and magnitude limited for M stars (around  $V_{\text{mag}} = 10$ ).

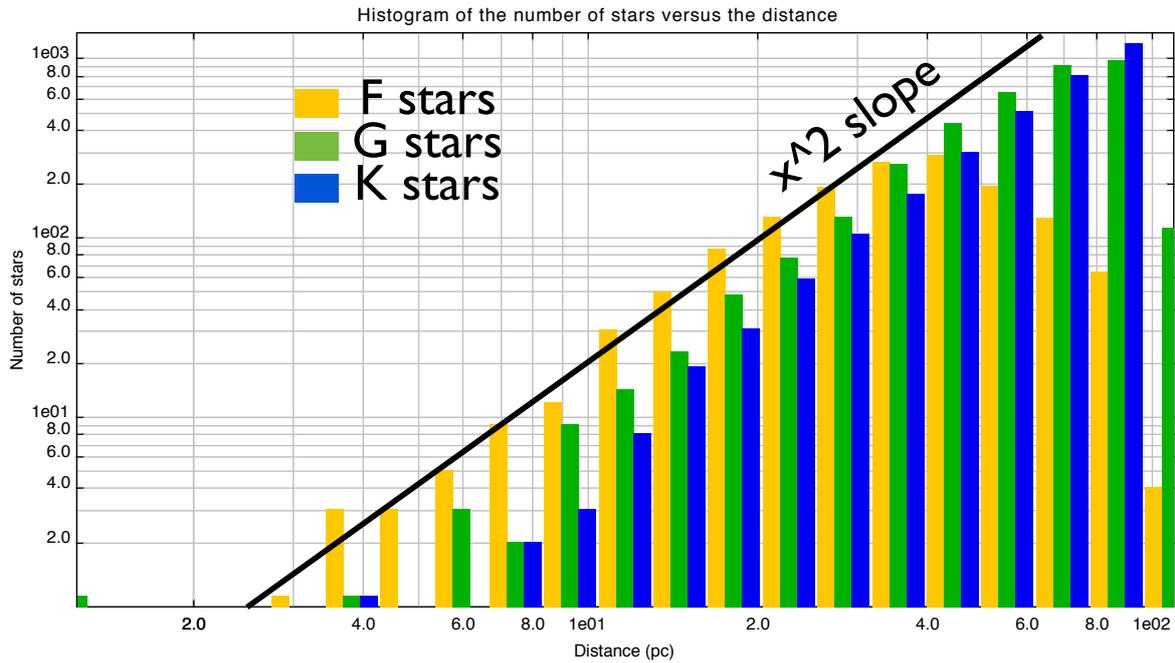
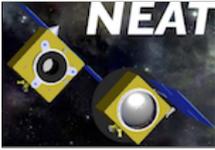


Figure 13: *Histogram of the distribution of the F, G and K stars versus the distance. We can see that the completeness breakpoint locations are different for each spectral type: 40 pc for K stars, 60 pc for G stars and further away for F stars. This pattern was expected since the hotter a star is, the larger it's luminosity is. The values for distances less than 10 pc can be ignored (the number of stars in these bins is too small).*

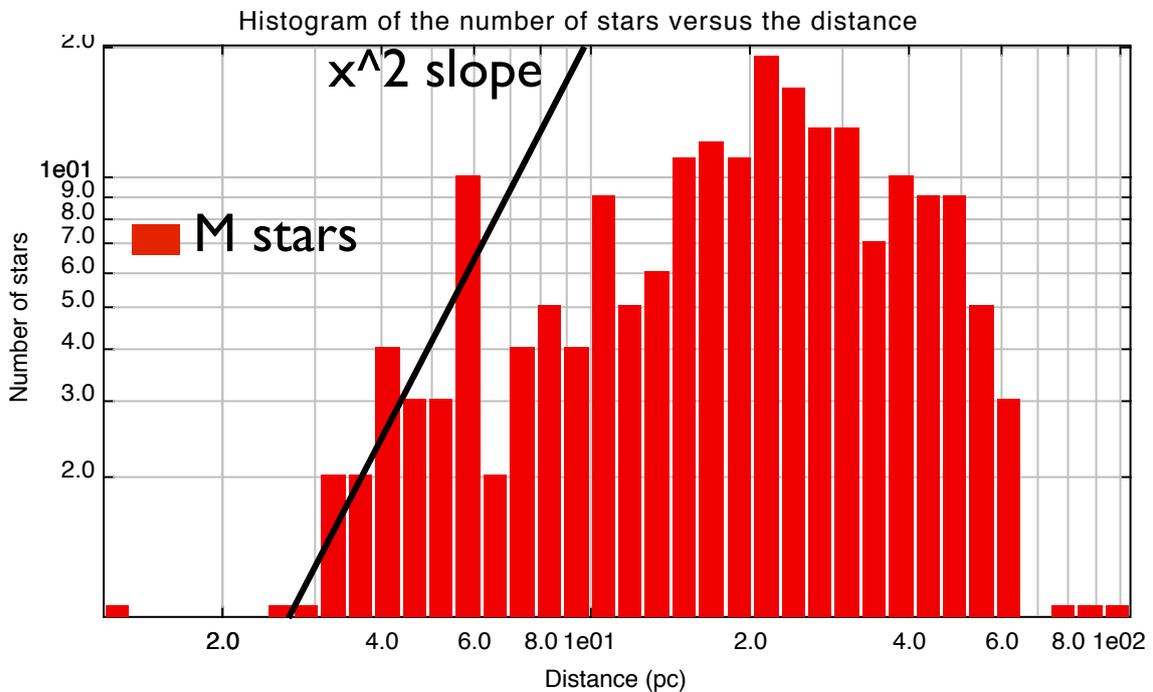


Figure 14: *Histogram of the distribution of the M stars versus the distance. The sample is quite small (75 stars). One can see that the slope is less than 2 for this histogram: although there is an upward trend for distances less than 20 pc, the sample is clearly incomplete (not even at 10 pc!).*



## References

- Alibert, Y., Carron, F., Fortier, A., Pfyffer, S., Benz, W., Mordasini, C., and Swoboda, D.: 2013, *Astronomy and Astrophysics* **558**, A109
- Dittkrist, K.-M., Mordasini, C., Klahr, H., Alibert, Y., and Henning, T.: 2014, *Astronomy and Astrophysics* **567**, A121
- Flower, P. J.: 1996, *APJ* **469**, 355
- Kaltenegger, L., Eiroa, C., and Fridlund, C. V. M.: 2010, *APSS* **326**, 233
- Kasting, J. F., Whitmire, D. P., and Reynolds, R. T.: 1993, *Icarus* **101**, 108
- Mordasini, C., Dittkrist, K.-M., Alibert, Y., Klahr, H., Benz, W., and Henning, T.: 2011, in A. Sozzetti, M. G. Lattanzi, and A. P. Boss (eds.), *IAU Symposium*, Vol. 276 of *IAU Symposium*, pp 72–75
- Schmidt-Kaler: 1982, in *Astronomy and Astrophysics* (eds.), *Stars and star clusters*, Vol. 2b, SpringerMaterials



## A – Summary of the Topcat operations

- Load the hipparcos and tycho catalog (all the tables), with the computed J2000 coordinates in a VOTable format

- In the main Hipparcos table, define the following subsets

- closest\_stars: `Plx > 50`
- closest\_MS\_A\_stars: `_2 && matches(SpType, "^A.*V$")`
- closest\_MS\_F\_stars: `_2 && matches(SpType, "^F.*V$")`
- closest\_MS\_G\_stars: `_2 && matches(SpType, "^G.*V$")`
- closest\_MS\_K\_stars: `_2 && matches(SpType, "^K.*V$")`
- closest\_MS\_M\_stars: `_2 && matches(SpType, "^M.*V$")`
- closest\_MS\_FGK\_stars:

```
_2 && (matches(SpType, "^F.*V$")  
|| matches(SpType, "^G.*V$") || matches(SpType, "^K.*V$"))
```

- closest\_MS\_AFGKM\_stars:

```
_2 && (matches(SpType, "^A.*V$")  
|| matches(SpType, "^F.*V$") || matches(SpType, "^G.*V$") ||  
matches(SpType, "^K.*V$") || matches(SpType, "^M.*V$"))
```

- Select the Tycho table and perform a crossmatch with Hipparcos to find the reference stars.

Change table name to: **Tycho reference stars crossmatch**

Export table as: **reference\_stars\_topcat.csv**

- Export the closest\_MS\_AFGKM\_stars subset as **closest\_MS\_AFGKM\_stars\_topcat.csv**, run the script and import the new table with the new magnitude column.
- Select the closest\_MS\_AFGKM\_stars subset and create all the "NEAT" columns (see the list in section B)
- For completeness only: create the close main sequence F, G, K and M stars subsets in the main Hipparcos table (stars closer than 100 pc).
- Re create the A, F, G, K and M main sequence stars subsets in the `closest_MS_AFGKM_stars_topcat.csv` table.

## B – List of the NEAT columns

- Columns prior to sum of reference stars flux (by script)** .

These columns can be created either before or after the equivalent reference star magnitude is added.

Dist 1000/Plx



BC  $-0.1784538+2.9023636*\$39-12.318814*\text{pow}(\$39,2)+20.931026*\text{pow}(\$39,3)$   
 $-22.50268371*\text{pow}(\$39,4)+63.6149604*\text{pow}(\$39,5)-177.541721356*\text{pow}(\$39,6)$   
 $+249.915028525*\text{pow}(\$39,7)-182.6519997*\text{pow}(\$39,8)+67.42205675*\text{pow}(\$39,9)$   
 $-9.98700145*\text{pow}(\$39,10)$

MB  $\text{Vmag}+\text{BC}-5*\ln(100/\text{Plx})/\ln(10)$

L  $\text{pow}(10, (4.75-\text{MB})/2.5)$

Mass  $\text{pow}(L, 1.0/3.8)$

SeffIn1  $\$39<0.62? -1.484*\$39+2.330 : -0.371*\$39+1.640$

SeffOut1  $\$39<0.62? -0.303*\$39+0.547 : -0.092*\$39+0.417$

SeffIn2  $\$39<0.62? -0.727*\$39+2.211 : -0.165*\$39+1.862$

SeffOut2  $\$39<0.62? -0.273*\$39+0.489 : -0.082*\$39+0.371$

HZIn1  $\text{pow}(L/\text{SeffIn1}, 0.5)$

HZIn2  $\text{pow}(L/\text{SeffIn2}, 0.5)$

HZOut1  $\text{pow}(L/\text{SeffOut1}, 0.5)$

HZOut2  $\text{pow}(L/\text{SeffOut2}, 0.5)$

AIn1  $3/\text{Mass}*HZIn1/\text{Dist}$

AIn2  $3/\text{Mass}*HZIn2/\text{Dist}$

AOut1  $3/\text{Mass}*HZOut1/\text{Dist}$

AOut2  $3/\text{Mass}*HZOut2/\text{Dist}$

AEarthFlux  $3/\text{Mass}*\text{pow}(L, 0.5)/\text{Dist}$

AMid1  $(\text{AIn1}+\text{AOut1})/2$

AMid2  $(\text{AIn2}+\text{AOut2})/2$

IWAIN1  $HZIn1/\text{Dist}$

IWAIN2  $HZIn2/\text{Dist}$

IWAOut1  $HZOut1/\text{Dist}$

IWAOut2  $HZOut2/\text{Dist}$

IWAEarthFlux  $\text{pow}(L, 0.5)/\text{Dist}$

IWAMid1  $(\text{IWAIn1}+\text{IWAOut1})/2$

IWAMid2  $(\text{IWAIn2}+\text{IWAOut2})/2$

TIn1  $\text{pow}(\text{pow}(HZIn1, 3)/\text{Mass}, 0.5)$

TIn2  $\text{pow}(\text{pow}(HZIn2, 3)/\text{Mass}, 0.5)$

TOut1  $\text{pow}(\text{pow}(HZOut1, 3)/\text{Mass}, 0.5)$

TOut2  $\text{pow}(\text{pow}(HZOut2, 3)/\text{Mass}, 0.5)$

## Columns after sum of reference stars flux (by script) .

These columns can only be created after the equivalent reference star magnitude is added.

SigmaH  $\text{pow}(0.0144*\text{pow}(10, (\text{Vmag}-6)/2.5)+0.372*\text{pow}(10, (\text{EqRefMag}-9.05)/2.5)+0.194, 0.5)$

TexoEarth  $36*\text{pow}(\text{SigmaH}, 2)/\text{pow}(\text{AIn1}, 2)$



## C – Python script for summing reference stars flux: source code

---

```
import sys
print "\n" + "PYTHON VERSION: " + sys.version + "\n"

from math import sin, cos, atan2, pi, sqrt, log

# The path to the include directories has to be added manually...
sys.path.append("/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/site-packages")
sys.path.append("/opt/local/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/site-packages")

import pylab as pl
import scipy as sp
import scipy.integrate as integ
from scipy import optimize
import numpy as np
import matplotlib.pyplot as plt # the tidy way
import matplotlib.mlab as mlab
import matplotlib as mp

# this script processes the target and reference stars from topcat imported csv
# data files. The goal is to compute for each target stars, the flux of each ref
# star and the total flux of the ref stars

import csv
from decimal import Decimal

print "SCRIPT START"

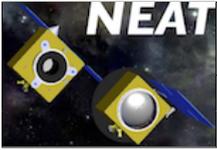
print "Step 1: retrieve magnitudes and IDs of target stars"

target = csv.reader(open('closest_FGK_targets_topcat.csv'))
#data = csv.DictReader(open('test.csv'))

ref = csv.reader(open('reference_stars_topcat.csv'))
fields = ref.next()
# the columns of interest are:
# - 0: RA (HIP)
# - 1: DEC (HIP)
# - 3: HIP ID
# - 81:

# TODO: add a way to find/print the selected columns, to check the correctness of
# the calculations
hip_ids = []
ra_targets = []
dec_targets = []

# retrieve hipparcos ids, RA and DEC coordinates of neat targets
for raw in target:
    ra_targets.append(raw[0])
```



```
        dec_targets.append(raw[1])
        hip_ids.append(raw[3])

ra_targets.pop(0)
dec_targets.pop(0)
hip_ids.pop(0)

print('hip_ids = ')
print(hip_ids)

print "Step 2: find which tychos stars are the targets"
# finds which entry in the tychos ref stars is the target
tyc_ids = [0]*len(hip_ids)
ref_star_mageq = [99]*len(hip_ids) # equivalent magnitude of ref stars

counts = 0

for i in range(len(hip_ids)):
    dist = 3600.0**2
    ref = csv.reader(open('reference_stars_topcat.csv'))
    ref.next()
    for raw in (ref):
        # print('hip_id = ' + hip_ids[i])
        # print('raw = ' + raw[3])
        # print(' ')

        if int(raw[3]) == int(hip_ids[i]):
            counts = counts + 1
            new_dist = (float(raw[81])-float(ra_targets[i]))**2 +
                (float(raw[82])-float(dec_targets[i]))**2
            if new_dist < dist:
                tyc_ids[i] = raw[84]
                dist = new_dist

print('counts = ')
print(counts)
print('tyc_ids = ')
print(tyc_ids)

print('size tyc_ids = ')
print(len(tyc_ids))

vmag_hip = []
vmag_tyc = []
sep = []
print "Step 3: verification of the matching between the hipparcos targets and the
    tychos matches found "
# TODO: add a verification step: verify that the parameters for the identified
    hipparcos targets in the list of references have the same hipparcos/tychos
    parameters. Count the good matches. Plot the values of hip params VS tyc params
for i in range(len(tyc_ids)):
    ref = csv.reader(open('reference_stars_topcat.csv'))
    ref.next()
    for raw in ref:
        if raw[84] == tyc_ids[i]:
```



```
        vmag_hip.append(float(raw[7]))
        vmag_tyc.append(float(raw[88]))
        sep.append(float(raw[142]))

#TODO: switch loop imbrication to speed up, with a counter on the neat targets
      (works only if the neat targets are in the same order than in the reference
      stars file)s
vmag_hip_arr = np.array(vmag_hip)
vmag_tyc_arr = np.array(vmag_tyc)
pl.figure(1)
plt.scatter(vmag_hip_arr,vmag_tyc_arr)
for ids, maghip, magtyc in zip(hip_ids, vmag_hip, vmag_tyc):
    plt.annotate(ids, xy = (maghip, magtyc), xytext = (-20, 20),textcoords =
        'offset points', ha = 'right', va = 'bottom',
        bbox = dict(boxstyle = 'round,pad=0.5', fc = 'yellow', alpha = 0.5))
#arrowprops = dict(arrowstyle = '->', connectionstyle = 'arc3,rad=0') adding arrow
    option => crash
plt.show()
pl.figure(2)
plt.scatter(sep,vmag_hip_arr-vmag_tyc_arr)
plt.show()
pl.figure(3)
plt.scatter(sep,vmag_hip_arr)

print "Step 4: calculate the sums of the reference stars flux for each target"
# calculate the sums of the reference stars flux for each target
for i in range(len(hip_ids)):
    ref = csv.reader(open('reference_stars_topcat.csv'))
    ref.next()
    for raw in ref:
        if raw[3] == hip_ids[i]:
            if raw[84] != tyc_ids[i]:
                ref_star_mageq[i] =
                    -2.5/log(10)*log(10**(-ref_star_mageq[i]/2.5) +
                    10**(-float(raw[88])/2.5))

print(ref_star_mageq)

target_out = csv.writer(open('closest_FGK_targets_topcat_out.csv', 'wb'))

target = csv.reader(open('closest_FGK_targets_topcat.csv'))
fields = target.next()

fields.append('eq_ref_mag')
print(fields)

target_out.writerow(fields)

i=0
for row in target:
    row.append(ref_star_mageq[i])
    target_out.writerow(row)
    i = i + 1

print "SCRIPT ENDS"
```

---