

ANALYSE DES DONNEES DU JMDC POUR LE CALCUL DU JSDC

Alain Chelli le 11 août 2025

Objectif: Produire un JMDC robuste avec le maximum de données indépendantes.

Fichier analysé: JMDC-2025-07-17_final_lddUpdated_pubOlbinTags.fits

Nombre d'entrées: 2328 (numérotées de 0 à 2327)

Pour le calcul du JSDC, on supprime du JMDC les entrées sans références (35), les diamètres non finies ou avec une erreur non finie ou sans barre d'erreur (85), puis les Mira, les étoiles variables, etc... (414 entrées). Il nous reste 1809 entrées qui se décomposent comme suit:

Nombre d'entrées: 1809

Sources distinctes: 918

Sources avec une seule entrée: 515

Sources avec plusieurs entrées: 403 (totalisant: 1294 entrées)

Les entrées uniques sont ce quelles sont, on ne peut rien y faire pour l'instant. Afin de produire un JSDC robuste, il est nécessaire selon moi de remplacer les entrées multiples du JMDC par leurs moyennes statistiques respectives, si les mesures sont cohérentes bien sûr. Cela permet déjà de supprimer les corrélations photométriques. Cela m'a amené à examiner en détail les entrées multiples. Je me suis ainsi aperçu que beaucoup d'entrées étaient corrélées entre elles par le biais des mesures de diamètre. En effet, certains groupes ont l'habitude d'inclure dans leurs dernières mesures les mesures faites précédemment. Cela implique que les derniers diamètres publiés sont les seuls à retenir, rendant obsolètes les mesures précédentes, car on ne connaît pas les corrélations entre diamètres. C'est une option, cela ne serait pas la mienne. Pour les groupes effectuant des 'surveys', cela fait beaucoup de mesures corrélées. NPOI est l'interféromètre gagnant avec 472 entrées potentiellement utiles, mais 233 sont corrélées donc inutiles pour notre étude.

On supprime donc les entrées corrélées (360), les entrées calculées (48), celles répétées (50), ou reprocessées (10), les entrées ayant une erreur supérieure ou de l'ordre de 20% (32), et pour finir, les entrées spéciales (44), comprenant la mesure des diamètres des composantes de système double (8), de sources allongées (8), d'une source variable (1), ainsi que toutes les mesures de deux papiers (28) et autres (2). Pour le détail, on pourra se reporter au fichier `jmdc_revisited_annex.txt`. Finalement, après avoir analysé les mesures multiples une à une et ce plus d'une fois, on supprime aussi les outliers (22) définis comme étant à une distance de l'ordre de 3 sigmas ou plus de la moyenne statistique après une correction de l'erreur (biais de 1% rajouté quadratiquement à l'erreur sur le diamètre, voir ci-après). Les outliers ne concernent que les sources ayant au moins quatre entrées. Après ces opérations, il nous reste 1306 entrées qui se décomposent comme suit:

Nombre d'entrées: 1306

Sources distinctes: 901

Sources avec une seule entrée: 646

Sources avec plusieurs entrées: 255 (totalisant: 660 entrées)

Ces données doivent encore être filtrées autant que faire se peut des entrées multiples en raison des corrélations photométriques. Toutefois, parmi les 660 entrées multiples, 225 diamètres ont une erreur inférieure ou très inférieure à 1%. Ces erreurs sont le plus souvent déraisonnablement faibles. Elles vont dominer les moyennes statistiques et produire en général des diamètres moyens non robustes, ce que l'on veut éviter. Afin de produire des diamètres moyens fiables, j'ai systématiquement rajouté une erreur de 1 % sur le diamètre, ce qui est l'erreur standard communément admise pour les interféromètres fibres. L'erreur minimale sur le diamètre moyen est alors $1/\sqrt{n}$ pourcent, où n est le nombre de mesures indépendantes.

Pour identifier les outliers, il n'est pas possible de se baser sur le Chi2 à cause du faible nombre de mesures de diamètre par source. A la place, je calcule la moyenne statistique et l'erreur sur la différence avec un diamètre donné. A partir de 4 mesures de diamètre, j'élimine les diamètres situés à une distance supérieure ou de l'ordre de 3 sigmas de la moyenne statistique. De plus, j'élimine aussi les sources n'ayant pas magnitude visible ou d'erreur associée (1), ainsi que celles n'ayant pas de type spectral. Cela donne :

Nombre d'entrées: 951

Sources distinctes: 862

Sources avec une seule entrée: 803 (dont 196 diamètres moyens totalisant 512 entrées)

Sources avec plusieurs entrées: 59 (totalisant: 148 entrées)

Conclusions et actions

1/ Le gros du travail est fait. Je prévois de publier cette étude dont le produit phare sera le premier catalogue de diamètres angulaires moyens avec des d'erreurs robustes, calculés avec au moins trois mesures indépendantes et cohérentes entre elles.

2/ Eliminer du JMDC (voir annexe)

- a) Les 85 entrées du JMDC avec un diamètre non fini et/ou avec une erreur non finie, ou sans barre d'erreur.
- b) Les 35 mesures sans référence, dont une référence de 1997 (voir tableau NOREF) sont aussi à éliminer du JMDC, sauf si on trouve les références associées.
- c) Eliminer du JMDC les 48 entrées calculées.

3/ Correction des tags du catalogue Olbin Publications. Ci-dessous un résumé non exhaustif:

```
FACILITY([1485+DINDGEN(85),792+DINDGEN(27),2133,2134])='NPOI' au lieu de NPOI,MARK III
FACILITY([1038+DINDGEN(30),1211,1212])='PTI ' au lieu de PTI,CHARA
S=[60,61,62,63] & FACILITY(S)='GSFC ' & INSTRUMENT(S)='Masking' & METHOD=1
S=[107] & FACILITY(S)='Palomar ' & INSTRUMENT(S)='UMAI'
S=[141] & FACILITY(S)='Steward ' & INSTRUMENT(S)='Masking'
S=[144,145] & FACILITY(S)='Steward ' & INSTRUMENT(S)='Masking '
S=[279] & FACILITY(S)='KPNO ' & INSTRUMENT(S)='Speckle '
S=[527+DINDGEN(16)] & FACILITY(S)='NTT ' & INSTRUMENT(S)='Masking ' & METHOD(S)=
S=[1406] --> bibcode(1406)=2014A&A...570A.104C ; et non 2021A&A...652A..26S
```

Corriger les entrées [142,143] de la référence 1977ApJ...217L..55B. Ces deux entrées ont le même ID. En fait il s'agit des deux composantes du système double Capella : Capella A et Capella B → ID(142)='Capella A' & ID(143)='Capella B' → corriger les autres tags de chaque source.

Il y a encore des tags à corriger, on verra plus tard.

4/ Complétude du JMDC

- a) Inclure dans le JMDC les mesures de la référence: 2023AJ....165...41B
- b) Rajouter une entrée dans le JMDC (elle est dans le texte du papier 2021A&A...652A..26S) → ID=HD35468, Id=0.807, e_Id=0.026. Cette entrée de CHARA/VEGA est la mesure reprocessée (diamètre du calibrateur changé) des données de la référence 2014A&A...570A.104C.

5/ Questions

- a) Beaucoup d'entrées de certains papiers sont en double au CDS, mais dans le papier il n'y a qu'une seule donnée, voire pas du tout. Certaines d'entre elles sont aussi parfois statistiquement incompatibles. Quelle est la mesure à retenir quand une seule ou aucune n'apparaît dans le papier? A discuter.
- b) Mais où est passée Vega. L'algorithme d'élimination de sources non désirées a aussi éliminé Vega et Altair. Je proteste, Vega la reine des références photométriques jusque dans les années 80, même si à cette époque on a détecté à 5 μm un très faible excès infrarouge (disque de débris) qui a traumatisé toute une génération de photométristes. Les critères de sélection doivent être revus avec des spécialistes de physique stellaire.
- c) Ce n'est peut-être pas un problème, mais je pense qu'il serait utile que l'on discute aussi des coefficients de conversion diamètre uniforme → diamètre assombri.
- d) La question qui tue : c'est quoi une étoile normale

Annexe : Erreur sur la distance d'un échantillon de la moyenne statistique

Soyons clairs : le χ^2 pour un petit nombre d'échantillons n'a aucun sens. Le filtrage des mesures multiples s'effectue de la manière suivante : à partir de 4 mesures, on élimine les entrées situées à une distance de l'ordre de 3 sigmas ou plus de la moyenne statistique.

On considère n variables aléatoires gaussiennes statistiquement indépendantes $\{x_i\}$ de variance $\{\sigma_i\}$. Leur moyenne statistique x_m s'écrit :

$$x_m = \frac{\sum_1^n \frac{x_i}{\sigma_i^2}}{\sum_1^n \frac{1}{\sigma_i^2}}$$

L'erreur sur la distance entre un échantillon et la moyenne statistique est donnée par :

$$x_j - x_m = x_j - \frac{\sum_1^n \frac{x_i}{\sigma_i^2}}{\sum_1^n \frac{1}{\sigma_i^2}} = x_j - \frac{\frac{x_j}{\sigma_j^2}}{\sum_1^n \frac{1}{\sigma_i^2}} - \frac{\sum_{i \neq j} \frac{x_i}{\sigma_i^2}}{\sum_1^n \frac{1}{\sigma_i^2}} = \frac{\sum_{i \neq j} \frac{x_j}{\sigma_i^2}}{\sum_1^n \frac{1}{\sigma_i^2}} - \frac{\sum_{i \neq j} \frac{x_i}{\sigma_i^2}}{\sum_1^n \frac{1}{\sigma_i^2}}$$
$$\sigma^2(x_j - x_m) = \frac{\sigma_j^2 \left(\sum_{i \neq j} \frac{1}{\sigma_i^2} \right)^2 + \sum_{i \neq j} \frac{1}{\sigma_i^2}}{\left(\sum_1^n \frac{1}{\sigma_i^2} \right)^2}$$

On utilise la formule ci-dessus pour le calcul de $\sigma^2(x_j - x_m)$.