

Validation du pipeline HyperGal

Sommaire

8.1	Génération des simulations	2
8.1.1	Méthode	2
8.1.2	Cube de galaxies isolées	2
8.1.3	Modèles de supernovae	2
8.1.4	Marginalisations et paramètres d'étude	4
8.1.5	Création des cubes de simulation	6
8.2	Résultats et précision	8
8.2.1	Précision spectrophotométrique	9
8.2.2	Précision avec correction de continuum	9
8.2.3	Efficacité de classification	9

Nous avons présenté et détaillé dans le chapitre précédent le fonctionnement du pipeline **HYPERGAL**. Après s'être assuré de sa stabilité numérique, nous avons cherché une méthode de validation de son efficacité. L'objectif est ainsi de quantifier la précision d'extraction des spectres de supernovae en fonction des conditions d'observation, et la capacité d'**HYPERGAL** à les classifier. Nous avons pour cela choisi de procéder à des simulations de cube d'observation avec la SEDm.

Dans ce chapitre nous présenterons dans un premier temps la procédure de génération des simulations, puis nous présenterons les résultats ainsi obtenus de l'utilisation d'**HYPERGAL** sur ces cubes simulés.

8.1 Génération des simulations

8.1.1 Méthode

Afin de se rapprocher au plus près des conditions d’observation, nous avons profité de quelques périodes de mise hors service de la caméra principale ZTF (entre fin novembre 2021 et fin janvier 2022) : nous avons ainsi pu utiliser occasionnellement la SEDm pour observer des galaxies hôtes isolées, dans lesquelles une supernova a été observée dans le passé.

Nos simulations sont ainsi basées sur une dizaine de ces cubes, extraits avec l’instrument pour lequel nous souhaitons tester **HYPERGAL**, et contenant dans le champ de vue une galaxie et un fond réels.

Le but est ainsi de rajouter une composante de supernova dans ces cubes en marginalisant sur les conditions d’observation habituelles comme le seeing, ou la proportion de chaque type de supernova, tout en explorant les conditions influants sur la robustesse d’**HYPERGAL** comme la distance entre la source ponctuelle et le centre galactique, et le rapport signal sur bruit.

Pour notre étude nous avons créé un jeu de 5000 cubes de simulations, et nous détaillons dans cette section leur conception.

8.1.2 Cube de galaxies isolées

La base de nos simulations proviennent donc d’observations réelles avec la SEDm de galaxies ayant accueilli au moins un an dans le passé une supernova. Ces cubes sont donc naturellement dans l’espace de l’instrument pour lequel **HYPERGAL** a été conçu.

Les effets d’ADR sont également présents, et il faut donc bien les caractériser avant d’inclure une composante de supernova pour que celle ci soit soumise aux mêmes effets chromatiques. Bien que nous connaissons à priori la masse d’air et l’angle parallactique au début de l’observation, nous ne connaissons pas ces paramètres effectifs, car ils varient au cours de l’exposition (de l’ordre d’une demi-heure de temps de pose).

Nous avons pour cela inclus dans **HYPERGAL** la flexibilité de prendre en compte ou non n’importe laquelle des composantes de scène, et avons procéder à l’ajustement de scène avec uniquement la galaxie hôte dans le MLA. Tout comme détaillé au chapitre précédent, l’ajustement du centroïde à chaque méta-tranche nous permet d’ajuster les paramètres effectifs d’ADR. Nos cubes présentent dans notre simulation une masse d’air allant de 1.01 à 2.04, ce qui nous permet de couvrir les conditions idéales d’observations, les conditions habituelles et les conditions dégradées.

Nous montrons dans la Figure 8.1 les cubes intégrés des galaxies hôtes utilisés pour les simulations.

8.1.3 Modèles de supernovae

Afin de tester la précision d’extraction de spectre avec **HYPERGAL**, il nous faut inclure dans les cubes une source ponctuelle dont le spectre est connu a priori. Cette étude est indépendante de la forme du spectre, et donc du type de la supernova. Cependant nous souhaitons également avoir une estimation de l’efficacité d’**HYPERGAL** à classifier les

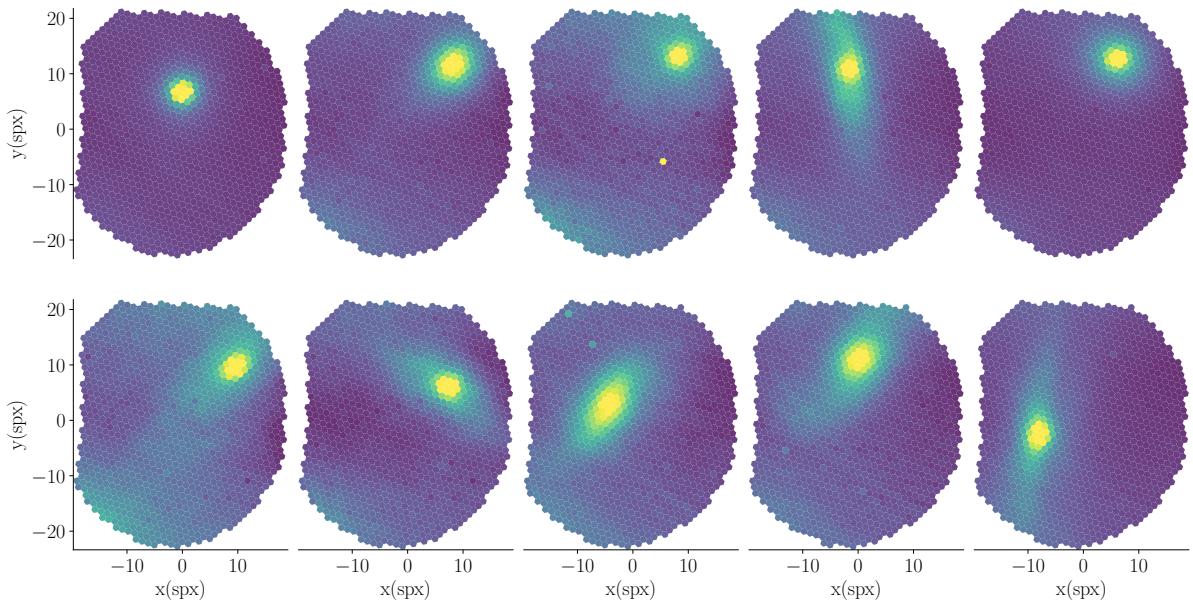


FIGURE 8.1 – Cubes intégrés des galaxies hôtes utilisés pour les simulations. Bien que nous n’ayons pas eu l’opportunité d’avoir un grand nombre d’observations de galaxies isolées avec la SEDm, nous avons fait l’hypothèse que ces morphologies et localisations variées de ces galaxies étaient suffisamment représentatives des observations pour constituer la base des simulations.

supernovae. Pour analyser ces deux aspects (précision et classification), il faut donc que le spectre de la source ponctuelle simulée soit connu a priori et que nous connaissons sa classification.

Par manque de temps et pour éviter de devoir générer des spectres avec des outils inconnus, puis les projeter dans l’espace des observations de la SEDm (transmission, LSF, échantillonnage ...), nous avons choisi d’utiliser des spectres de supernovae déjà obtenus avec la SEDm, et classifiés avec SNID.

Afin de s’assurer de la classification, nous n’avons sélectionné que des spectres avec un très haut *r_{lap}* (paramètre de qualité/confiance de SNID, considéré comme bon si *r_{lap}* > 5). Pour les spectres de supernovae de type Ia (les plus nombreuses), nous avons sélectionné 70 spectres avec un *r_{lap}* > 25 pour le meilleur modèle, et un *r_{lap}* > 15 pour les 30 premiers modèles.

Sur un raisonnement similaire, nous avons sélectionné 7 spectres de supernova de type II avec un *r_{lap}* > 12. Pour les types Ic et Ib, plus rarement observés ($\approx 5\%$ des observations), nous avons préféré prendre seulement 1 spectre de chaque mais avec une très forte confiance de classification (*r_{lap}* ≈ 18 pour la Ib et *r_{lap}* ≈ 15 pour la Ic).

Nous procédons ensuite sur chacun de ces spectre à un lissage en appliquant un filtre de Savitzky-Golay (SAVITZKY et GOLAY, 1964). Afin de ne pas casser les structures des spectres, nous utilisons un lissage léger avec un polynôme d’ordre 3 sur une fenêtre de 5 pixels.

Nous montrons dans la Figure 8.2 un exemple de spectre après lissage pour chaque type de supernova, ainsi que le meilleur modèle de classification SNID et le *r_{lap}* associé.

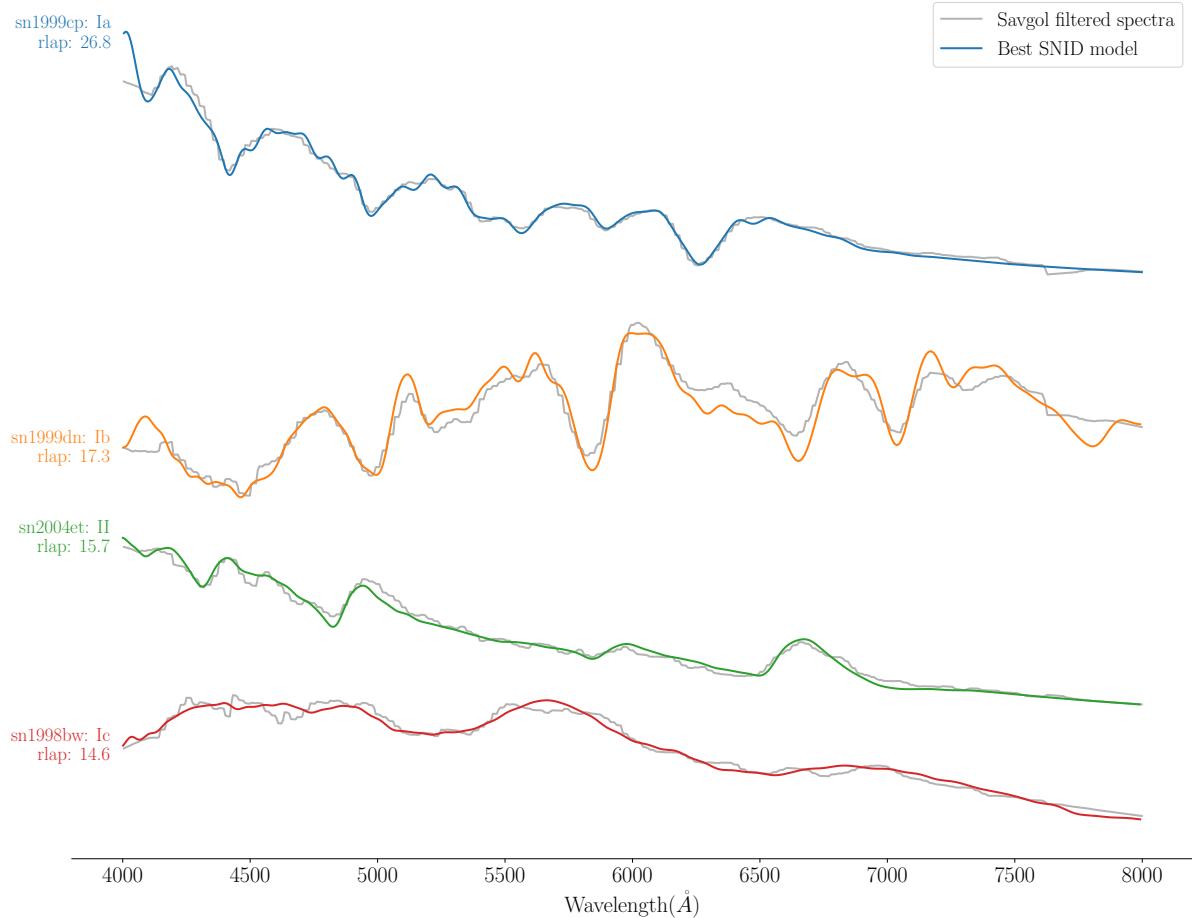


FIGURE 8.2 – Exemple de spectre pour chaque type de supernova pour les simulations. Nous y montrons en gris le spectre après lissage par un filtre de Savitzky-Golay, qui provient d'une observation de la SEDm. En couleur nous montrons le meilleur modèle de classification SNID et le *rlap* associé, avec de haut en bas : une SNIa, une SNIb, une SNII, et une SNIc.

8.1.4 Marginalisations et paramètres d'étude

8.1.4.1 Types et phases

Dans le but de représenter dans nos simulations les proportions observées de chaque type de supernova, nous utilisons les statistiques de la Data Release 1 du groupe Bright Transient Survey de ZTF (BTS; [FREMLING et al., 2020](#)). Nous choisissons ainsi de répartir dans nos simulations 80% de SNIae, 15% de SNeII, 2.5% de SNeIb et 2.5% de SNeIc. Ces deux derniers types sont habituellement regroupés, nous considérerons donc par la suite un groupe de 5% de SNeIbc.

Nous choisissons également de procéder à une marginalisation des phases des spectres de SNIea, en se basant sur les statistiques de la DR1 du groupe SNeIa de ZTF ([DHAWAN et al., 2022](#)). Pour les 70 spectres utilisés, nous déduisons la phase en comparant le jour d'observation de la supernova dont est issu le spectre avec le pic de luminosité ajusté par ZTF avec SALT2 ([GUY et al., 2005, 2007, 2010; BETOULE et al., 2014](#)) sur la courbe de

lumière.

La distribution de phase de notre échantillon s'étend de -15 à +15 jours, avec une médiane à -2 jours. Nous pouvons ainsi sélectionner aléatoirement les spectres de SNIa, sachant leur phase et suivant une distribution équivalente à celle relevée dans ([DHAWAN et al., 2022](#)). Nous marginalisons nos simulations suivant une distribution de phase gaussienne, centrée sur -3 jours et d'écart type 4 jours.

8.1.4.2 Seeing

Les supernovae étant des sources ponctuelles à ajouter dans nos cubes de simulations, elles sont entièrement caractérisées par leur profil de PSF.

Nous faisons l'hypothèse d'un profil connu a priori, et utilisons le profil radial développé au chapitre ?? avec l'étude des étoiles standards. Afin de représenter une distribution en seeing similaire à celle observée par la SEDm, nous marginalisons nos simulations sur le seeing en utilisant les distributions conjointes des paramètres de forme de PSF ajustés des 2202 étoiles standards extraites pour l'étude de la calibration en flux (section ??).

Nous faisons donc la supposition que la distribution en seeing des étoiles standards est représentative de celle des supernovae. Bien que la contribution de l'optique du télescope soit indépendante de l'objet observé, il faut noter que les étoiles standard le sont habituellement avec une masse d'air comprise entre 1 et 1.2. Nos simulations ayant une masse d'air comprise entre 1 et 2, cela implique potentiellement une sous-estimation que nous n'avons pas caractérisé de la distribution en seeing utilisée pour nos simulations.

8.1.4.3 Distance supernova/centre galactique

HYPERGAL a été conçu pour répondre à la problématique de la contamination par la galaxie hôte. Nous voulons donc explorer la précision d'extraction de spectre des SNe et l'efficacité de classification suivant la distance séparant la source ponctuelle du centre galactique. Dans ces simulations nous ne nous intéressons pas aux cas où la supernova est complètement isolée dans le champ de vue, ayant déjà entraîné le pipeline avec les étoiles standards.

Nous utilisons une distribution uniforme comprise entre 0 et 10 spaxels de distance, ce qui correspond à un intervalle entre 0 et $\approx 5''.6$. Cette distance seuil représente généralement environ 2 à 3 largeur à mi-hauteur suivant le profil radial des sources ponctuelles, ce qui nous semble suffisant pour explorer un large intervalle de séparation angulaire jusqu'à la limite d'une isolation totale de la supernova.

Nous prenons également en compte que lors des observations réelles, les supernovae sont habituellement situées vers le centre du MLA. Ainsi afin d'éviter de simuler une cible dans un des coins du cube, nous restreignons la localisation possible de la source ponctuelle dans un disque de 12 spaxels de rayon au centre du MLA. Pour les cas où la galaxie est très excentrée et que nous simulons une source ponctuelle proche du centre galactique, nous privilégions de la positionner dans le quart de cercle en direction du centre du MLA.

8.1.4.4 Contraste

Le dernier paramètre que nous utilisons pour explorer la robustesse d'**HYPERGAL** correspond à l'intensité du flux de la supernova par rapport à ce qui se situe à sa localisation :

nous introduisons ainsi le contraste c_r , défini dans la bande photométrique équivalente ZTF_r afin de pouvoir plus aisément comparer les résultats des simulations avec un cas réel d'observation, exprimé comme :

$$c_r = \frac{S_r}{S_r + B_r} \quad (8.1)$$

avec S_r le signal de la supernova et B_r le signal de tout ce qui se situe en fond (ciel + galaxie).

Afin de déterminer la quantité B_r qui contamine le signal de la supernova, il faut prendre en compte le profil de PSF utilisé pour simuler la source ponctuelle. En effet, si on suppose que la SN est centrée (pour une longueur d'onde donnée) à la position (x_0, y_0) , alors le signal de fond à la même position aura un plus grand impact de contamination que le fond à la position $(x_0 + dx_0, y_0 + dy_0)$.

Pour prendre cela en compte et plutôt que de considérer une ouverture fixe autour de la localisation de la SN simulée pour définir B_r , nous multiplions le cube de simulation sans la SN par un cube ne contenant que le profil de PSF (normalisé avec un pic à 1) à la localisation de simulation de la SN.

Le contraste est ainsi défini dans l'intervalle $]0, 1[$, 0 impliquant que la supernova n'existe pas, et 1 qu'elle est infiniment plus intense que le fond (ou que le fond est à zéro ce qui n'est pas notre cas ici).

Nous pouvons également relier le contraste au rapport $R = \frac{S_r}{B_r}$:

$$c_r = \frac{R}{1 + R} \quad (8.2)$$

Les simulations sont ainsi générées suivant une distribution uniforme du contraste c_r entre 0 et 1.

Nous pouvons également voir que le rapport signal sur bruit est étroitement lié au contraste. En effet, en supposant que le signal dans le cube est entièrement caractérisée par une loi de Poisson, nous avons alors que :

$$SNR_r \triangleq \frac{S_r}{\sqrt{\sigma_{S_r}^2 + \sigma_{B_r}^2}} \approx \frac{S_r}{\sqrt{S_r + B_r}} = c_r \times \sqrt{S_r + B_r} \quad (8.3)$$

avec S_r et B_r en unités de coups. Avec un raisonnement similaire nous pouvons montrer que $SNR_r \approx R \times \sqrt{B_r}$.

En pratique, nous sommes en mesure de récupérer la quantité σ_B , car présente dans le cube SEDm avec la galaxie hôte isolée. Pour remonter au SNR, nous utilisons directement sa définition en supposant que le bruit à ajouter dans le cube à cause du signal de la supernova simulée est $\sigma_S^2 = S$.

8.1.5 Création des cubes de simulation

Après avoir procéder à la marginalisation des proportions de chaque type de supernova, de la phase des Ia et du seeing, nous générerons un jeu de $N \times m$ paramètres avec N le

nombre de simulations (5000), et m les paramètres de la simulation :

- ◊ Cube de la galaxie hôte ;
- ◊ Spectre de supernova ;
- ◊ Paramètres de PSF décrivant la SN ;
- ◊ Distance entre la SN et le centre galactique ;
- ◊ Contraste.

Pour pouvoir ajouter le signal de la supernova simulée et surtout le bruit associé, nous devons utiliser le cube SEDm en unité de flux ADU et travailler dans ces unités avec le spectre de la SN. Connaissant a priori la calibration en flux qui sera utilisé pour chacun des cubes, nous appliquons une calibration inverse sur le spectre à simuler, qui est initialement en unité de flux physique.

La création d'un cube de simulation sachant les m paramètres se fait ensuite en plusieurs étapes :

- (a) **Détermination de la localisation** (x_{ref}, y_{ref}) **de la supernova** dans le cube à une longueur d'onde de référence ($\lambda_{ref} = 6000\text{\AA}$) : nous prenons aléatoirement une position sur le cercle centré sur la galaxie, avec un rayon égal à la distance simulée SN/galaxie. Nous prenons en compte les contraintes pour éviter les bords du cube expliquées dans la section 8.1.4.3 ;
- (b) **Détermination du signal de fond B** : nous construisons un cube vide dans lequel nous plaçons le profil de PSF à la localisation et longueur d'onde fixée à l'étape précédente. La localisation est propagée pour toutes les tranches avec le modèle d'ADR, sachant les paramètres de masse d'air et d'angle parallactique, et le profil est normalisé à un pic égal à 1 pour chaque longueur d'onde. Nous multiplions alors le cube de galaxie par celui-ci, le résultat étant un cube contenant uniquement le signal de fond B contaminant la SN.
- (c) **Détermination du coefficient multiplicatif à appliquer sur le spectre de la supernova.** Nous déterminons grâce à l'étape précédente le fond contaminant B_r dans la bande équivalente R de ZTF. Connaissant également le spectre de la supernova en ADU, nous en déduisons son signal S'_r dans la bande R avant adaptation au contraste souhaité. Enfin, connaissant B_r , S'_r et c_r , nous appliquons le coefficient multiplicatif nécessaire sur l'ensemble du spectre de la SN (et donc sur S'_r) pour obtenir le contraste souhaité.
- (d) **Ajout du bruit associé à la supernova.** Nous supposons que le flux ajouté de la supernova simulée est entièrement caractérisé par une loi de Poisson, et ajoutons donc au cube SEDm pour chaque spaxel de chaque tranche une variance telle que $\sigma_{S,\lambda,spx}^2 = S_{\lambda,spx}$.
- (e) **Détermination du SNR.** Le SNR n'est pas un paramètre de nos simulations, mais nous pouvons le récupérer et le stocker connaissant B_r , S_r , σ_{B_r} et σ_{S_r} .
- (f) **Construction du cube de simulation.** Tous les ingrédients sont réunis pour la construction du cube : le spectre de la supernova, sa position chromatique, son profil de PSF chromatique et le coefficient multiplicatif pour avoir le contraste désiré.

Nous procédons ainsi à la générations des 5000 cubes de simulations. Dans la Figure 8.3 nous illustrons quelques exemples de ces cubes pour différentes valeurs de contraste, distance, type de SN et SNR.

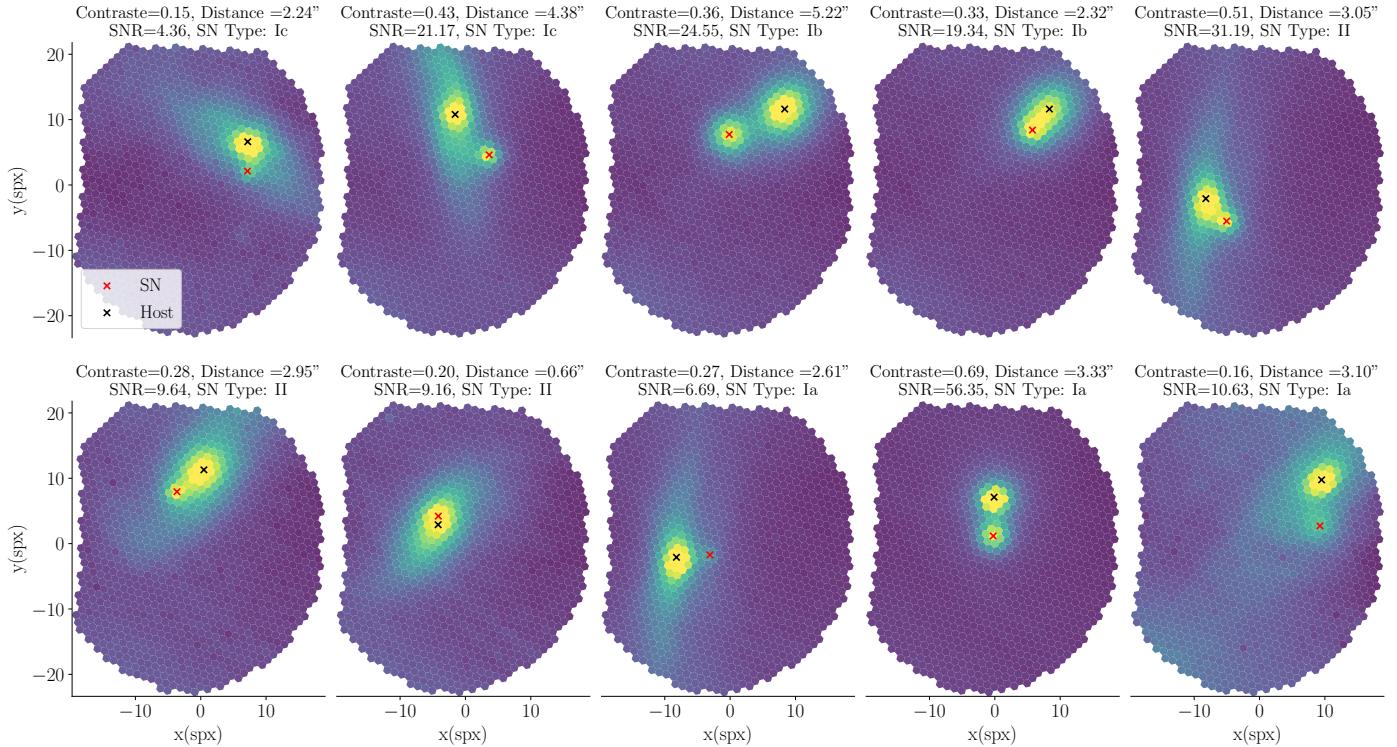


FIGURE 8.3 – Exemples de cubes de simulation pour différentes valeurs de contraste, distance, type de SN et SNR.

8.2 Résultats et précision

Après avoir généré nos cubes de simulation, nous avons fait tourner HYPERGAL suivant 2 méthodes : la première avec une modélisation de scène comprenant toutes les composantes comme détaillée au chapitre précédent. Et une deuxième fois avec la même méthode d'extraction que le pipeline d'origine PYSEDM, sans modélisation de la galaxie hôte.

Nous n'avons pas utilisé directement le pipeline PYSEDM car les modèles de PSF et de fond sont différents de celui d'HYPERGAL, ce qui n'aurait pas permis une comparaison robuste. La méthode d'extraction est cela dit identique, suivant le procédé détaillé dans RIGAULT et al. (2019) et la section ???. Les seules différences avec la modélisation de scène complète étant l'absence de modèle de galaxie, et le fait que l'on ne considère qu'un disque de 10 spaxels de rayon autour de la position de la supernova pour son extraction.

En plus d'une étude de la robustesse absolue d'HYPERGAL, cette confrontation nous permet d'avoir également une idée de l'amélioration apportée avec ce nouvel outil d'extraction de spectre.

Nous dénominerons dans la suite du manuscrit l'indice HG pour la méthode de modélisation de scène HYPERGAL, et RMS_{PS} pour la méthode d'extraction de source ponctuelle basique.

Dans cette section nous allons étudier 3 informations pour chacune des 2 méthodes :

- ◊ **La précision spectrophotométrique**, c'est à dire une comparaison brute du spectre de simulation et du spectre extrait ;
- ◊ **La précision après correction du continuum**, à l'instar de la méthode de pré-traitement utilisé dans SNID (section ??). La SEDm ayant été conçu pour la classification de spectres, ce qui nous importe est la capacité d'**HYPERGAL** à extraire les informations spectrales permettant cette classification, c'est à dire la structure du spectre traduisant les caractéristiques de tel ou tel type.
- ◊ **L'efficacité de classification**. Pour cela nous utiliserons le même classifieur utilisé par ZTF, SNID, et nous comparerons la classification du spectre extrait avec celui connu a priori.

8.2.1 Précision spectrophotométrique

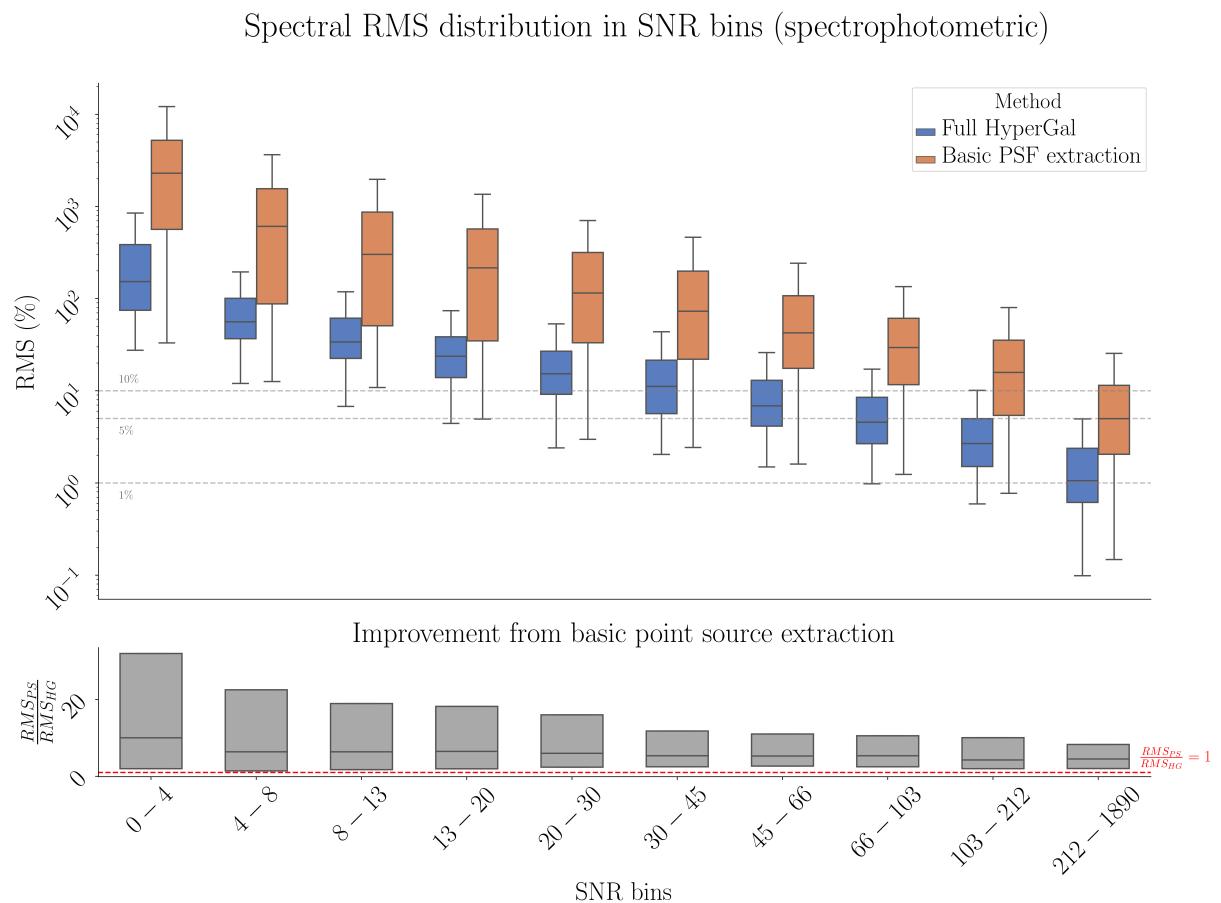


FIGURE 8.4

8.2.2 Précision avec correction de continuum

8.2.3 Efficacité de classification

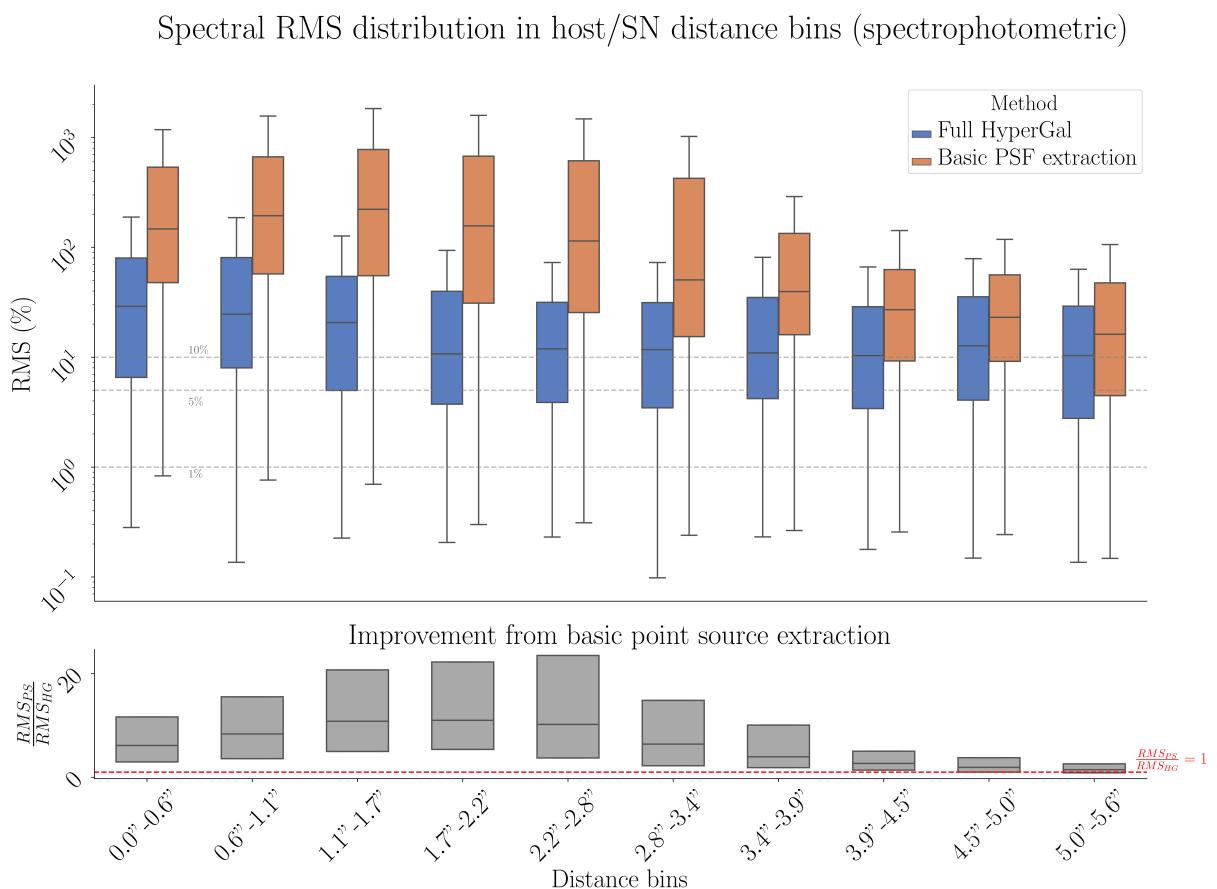


FIGURE 8.5

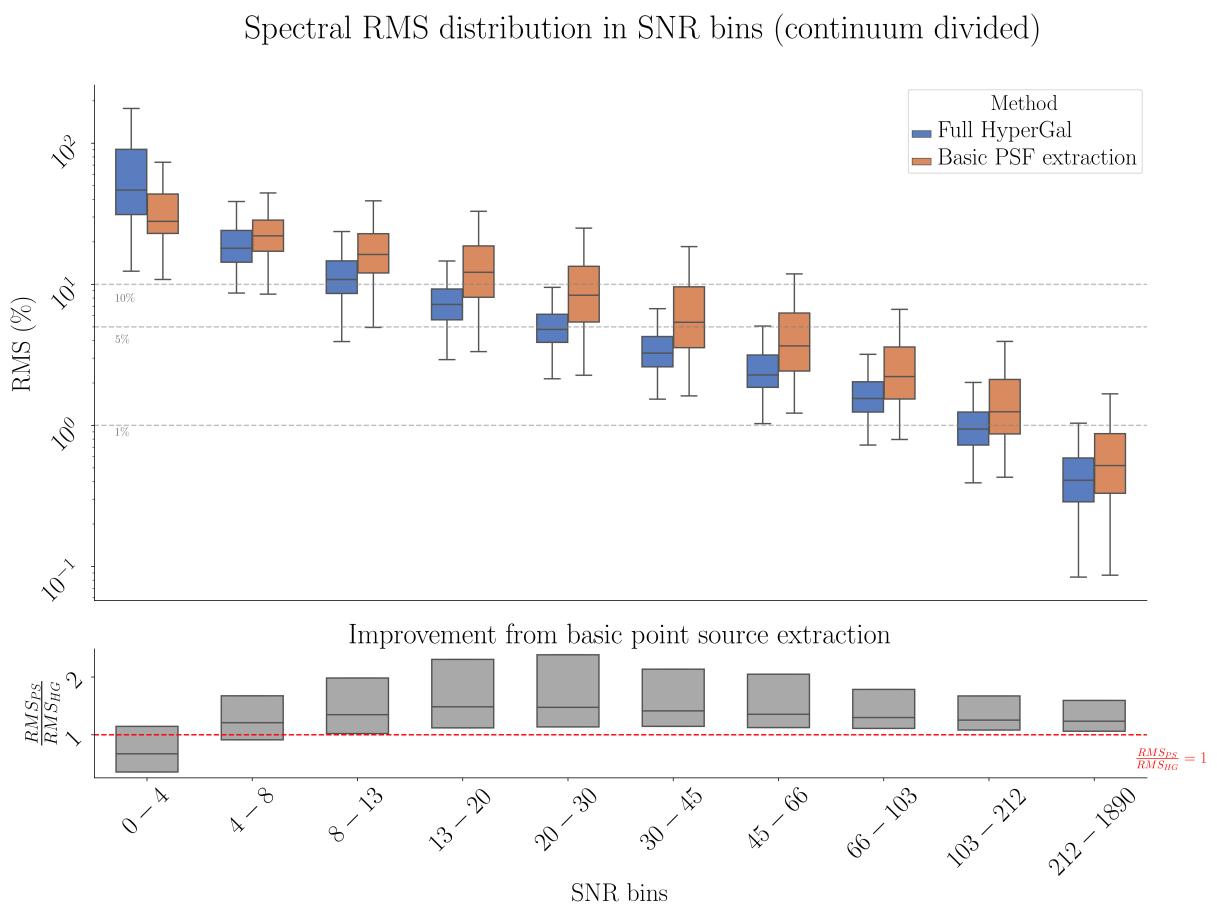


FIGURE 8.6

Spectral RMS distribution in host/SN distance bins (continuum divided)

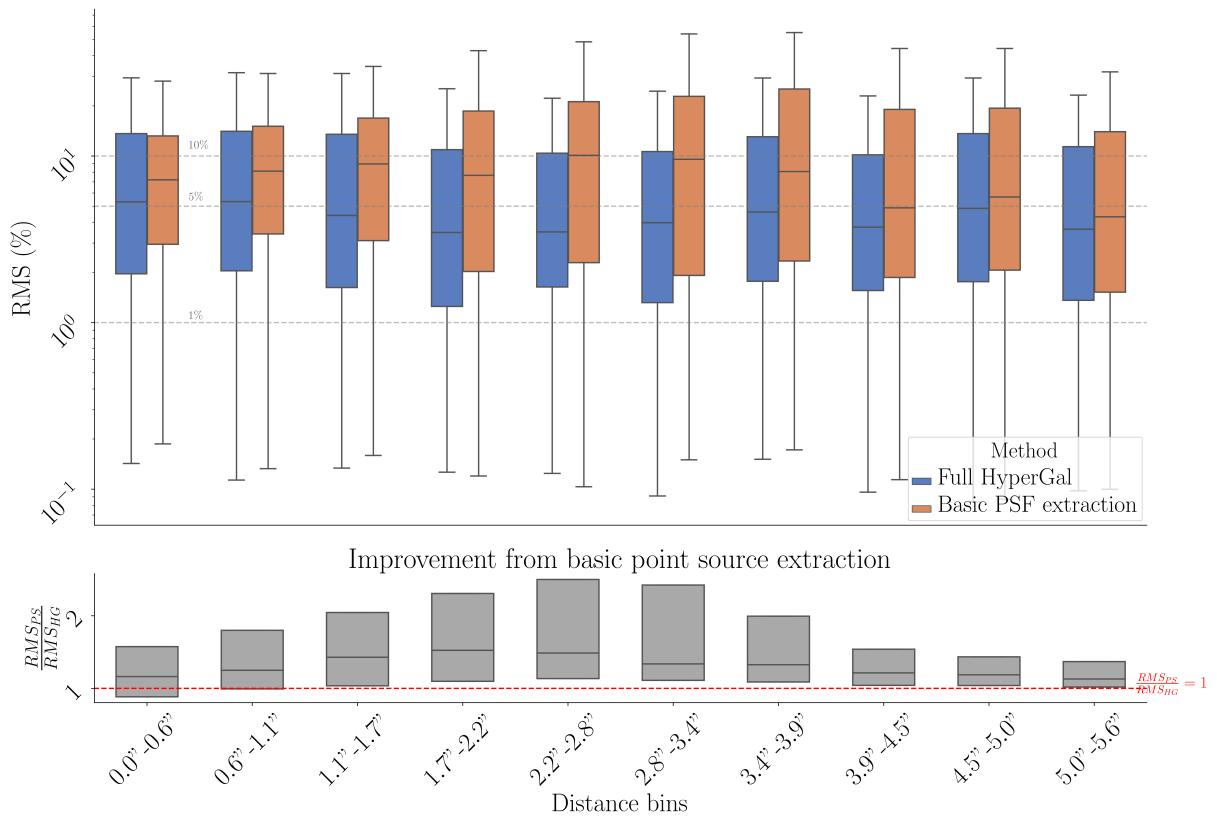


FIGURE 8.7

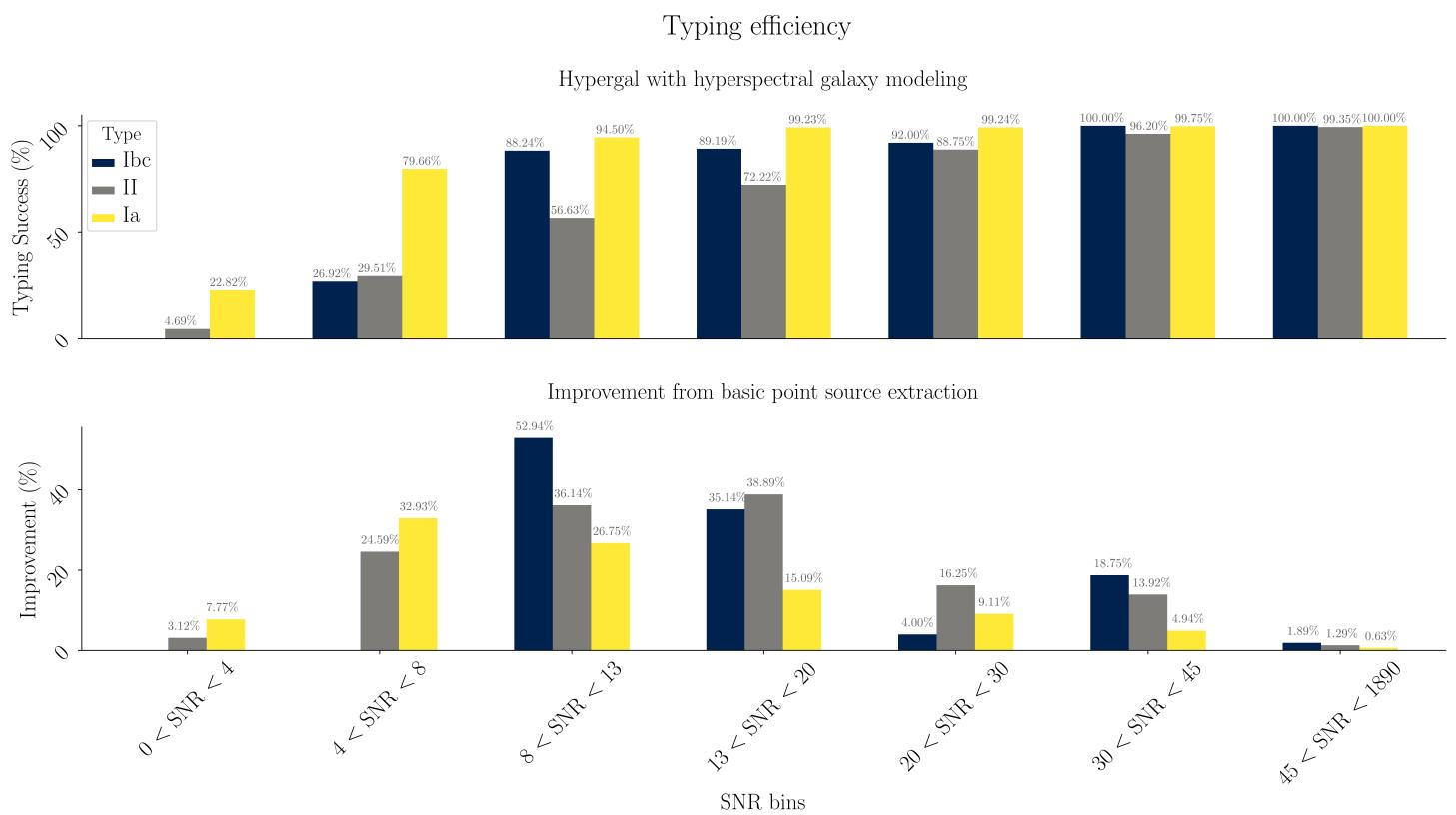


FIGURE 8.8

Bibliographie

- BETOULE M., KESSLER R., GUY J. et al. 2014, « Improved cosmological constraints from a joint analysis of the SDSS-II and SNLS supernova samples », *Astronomy and Astrophysics*, 568, A22 [↑ Section 8.1.4.1](#)
- DHAWAN S., GOOBAR A., SMITH M. et al. 2022, « The Zwicky Transient Facility Type Ia supernova survey : first data release and results », *Monthly Notices of the Royal Astronomical Society*, 510, 2228 [↑ Section 8.1.4.1](#), [↑ Section 8.1.4.1](#)
- FREMLING C., MILLER A. A., SHARMA Y. et al. 2020, « The Zwicky Transient Facility Bright Transient Survey. I. Spectroscopic Classification and the Redshift Completeness of Local Galaxy Catalogs », *The Astrophysical Journal*, 895, 32 [↑ Section 8.1.4.1](#)
- GUY J., ASTIER P., BAUMONT S. et al. 2007, « SALT2 : using distant supernovae to improve the use of type Ia supernovae as distance indicators », *Astronomy and Astrophysics*, 466, 11 [↑ Section 8.1.4.1](#)
- GUY J., ASTIER P., NOBILI S., REGNAULT N. et PAIN R. 2005, « SALT : a spectral adaptive light curve template for type Ia supernovae », *Astronomy and Astrophysics*, 443, 781 [↑ Section 8.1.4.1](#)
- GUY J., SULLIVAN M., CONLEY A. et al. 2010, « The Supernova Legacy Survey 3-year sample : Type Ia supernovae photometric distances and cosmological constraints », *Astronomy and Astrophysics*, 523, A7 [↑ Section 8.1.4.1](#)
- RIGAULT M., NEILL J. D., BLAGORODNOVA N. et al. 2019, « Fully automated integral field spectrograph pipeline for the SEDMachine : pysedm », *Astronomy and Astrophysics*, 627, A115 [↑ Section 8.2](#)
- SAVITZKY A. et GOLAY M. J. E. 1964, « Smoothing and differentiation of data by simplified least squares procedures », *Analytical Chemistry*, 36, 1627 [↑ Section 8.1.3](#)