

HyperGal : Modéliseur de scène pour l'extraction de sources ponctuelles

La première partie de ce manuscrit était dédiée à la présentation du contexte scientifique dans lequel ce travail de recherche est effectué.

Nous avons dans un premier temps introduit les notions de cosmologies nécessaires pour comprendre l'environnement scientifique de travail, ainsi que la nature et le rôle des supernovae de type Ia en tant que sondes cosmologiques.

Dans un second temps nous avons présenté la collaboration Zwicky Transient Facility, ses différents groupes de recherches et plus particulièrement la place qu'occupe l'étude des SNeIa dans ce relevé astronomique nouvelle génération. Après avoir introduit la nécessité d'une méthode de classification spectroscopique des événements transitoires détectés par la caméra ZTF, nous avons présenté la Spectral Energy Distribution machine, un spectrographe 3D que possède la collaboration et conçu pour la classification.

Le pipeline de réduction de données actuel, **PYSEDm**, permet également une extraction des sources ponctuelles observées par la caméra de l'IFU de la SEDm. La méthode implémentée est toutefois rudimentaire, et ne permet pas de palier aux nombreuses situations de contamination de la source ponctuelle par sa galaxie hôte.

Non seulement cela induit une perte statistique de supernovae classifiables non négligeable, mais de surcroît cela induit un biais environnemental dans l'échantillon des SNeIa de ZTF.

C'est pour répondre à cette problématique que nous introduisons **HYPERGAL**, un modéliseur de scène pour l'extraction de sources ponctuelles.

4.1 Idée générale

4.1.1 Problématique

Le champ de vue de la SEDm étant étroit (28×28 arcsec), nous avons en général 3 composantes qui composent la scène, à savoir le fond du ciel, la galaxie hôte et la source ponctuelle.

La difficulté majeure d'une modélisation de scène spectrale provient de la chromaticité de chacune de ces composantes, et plus particulièrement de la galaxie qui est une source structurée de forme et chromaticité variable.

Une première idée serait d'attendre l'extinction de l'évènement transitoire, réobserver l'hôte, et projeter cette seconde acquisition dans l'espace de la première observation afin d'isoler la source ponctuelle. Une telle approche est envisageable pour une extraction de quelques cibles, mais en aucun cas à notre époque où les relevés grands champs et à haute cadence deviennent légions et observent des milliers de supernovae par an.

Le but d'**HYPERGAL** est de pouvoir modéliser la scène observée par la SEDm après réduction des données, c'est à dire le cube 3D, à la volée des observations.

Il va donc non seulement falloir trouver un moyen de modéliser chacune des composantes, mais également de les projeter dans l'espace des observations de la SEDm. Cela implique une étude profonde des caractéristiques de l'instrument mais également de prendre en compte les conditions environnementales comme l'atmosphère le long de la ligne de visée.

4.1.2 La composante galactique

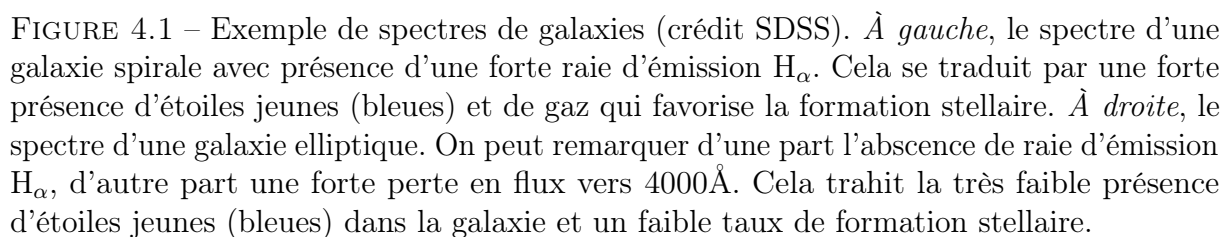
La motivation principale de ce modéliseur de scène est le fait que nous avons des informations sur la galaxie hôte avant l'apparition de l'évènement transitoire. En effet, d'autres relevés astronomiques comme le Sloan Digital Sky Survey (SDSS ; ?) ou Panstarrs (?) ont couvert des portions communes de ciel avec ZTF, et permettent donc de remonter à des informations photométriques de la galaxie encore isolée de la supernova.

Sauf que nous souhaitons une modélisation spectrale. Il faut donc un moyen de passer de l'espace photométrique à l'espace spectral, autrement dit estimer la Spectral Energy Distribution (SED) de la galaxie.

Plus précisément encore, l'objectif est de pouvoir recréer un cube 3D contenant uniquement la galaxie, afin de pouvoir le projeter dans l'espace de la SEDm en prenant en compte les propriétés de l'instrument et les conditions d'observation. Donc non seulement il faut être en mesure de passer de l'espace photométrique à l'espace spectral, mais il faut le faire localement pour que chaque spaxel du cube modèle ait son propre spectre associé.

Une approche triviale de ce problème serait de partir d'images de plusieurs bandes photométriques de la galaxie, et interpoler un spectre pour chaque pixel de ces images à l'aide d'un simple polynôme. Cela permettrait de créer un cube 3D avec une source structurée. Mais grâce à l'avènement de nombreux instruments terrestres et spatiaux lancés au cours des dernières décennies, nous avons une certaine connaissance de la composition d'une galaxie, et ainsi des différentes contributions qui forment sa SED.

L'idée est donc d'utiliser un SED Fitter appliqué aux données photométriques de la galaxie, afin de construire un cube 3D qui servira de base pour le modéliseur de scène.



4.2.1 SED d'une galaxie

Des rayons γ au domaine radio, la SED d’une galaxie est définie par sa composition en matière baryonique, et de leurs interactions physiques complexes. En mettant de côté la matière noire, une galaxie est composée d’étoiles de tout âge, de gaz atomiques moléculaires ou ionisés, de poussières et potentiellement d’un trou noir supermassif (?). Les étoiles qui composent la galaxies (entre 10^8 et 10^{14}) émettent la lumière qui nous permet de la détecter. D’autre part, le gaz interstellaire et la poussière vont quant à eux principalement altérer la SED : le gaz en ajoutant des raies d’émission et d’absorption, la poussière en provoquant une atténuation par absorption dans l’UV/proche infrarouge et une diffusion des radiations, puis en ré-émettant dans l’infrarouge moyen/lointain.

La SED d’une galaxie contient ainsi l’empreinte de tous ces ingrédients et phénomènes physiques complexes, évoluant au cours du temps et traçant l’histoire de la galaxie. Deux exemples de spectres de galaxies obtenus avec le relevé SDSS dans l’optique sont présentés dans la Figure 4.1.

Malgrès tout, certaines corrélations entre plusieurs paramètres rendent cette tâche très

difficile, comme par exemple la dégénérescence entre l'âge et la métallicité (?), ou encore l'âge et l'atténuation (?).

Ces deux dernières décennies ont été extrêmement riches en développement de modèles et observations panchromatiques, permettant une compréhension de plus en plus fine de la formation et l'évolution d'une galaxie.

Ont vu ainsi le jour des modèles de populations stellaires grâce à ?, ? et ?. D'un autre côté, différentes lois d'atténuation par la poussière ont été développées, comme par ?? via l'étude de SED de galaxies proches ayant un fort taux de formation stellaire, ou encore avec une approche plus théorique de modèles de transferts radiatifs (?). Comme mentionné précédemment, la poussière ré-émet dans l'infrarouge et l'étude et la modélisation de ce phénomène est un domaine actif de recherche (????).

La manipulation de modèles pour chaque processus physique en oeuvre dans une galaxie a permis l'émergence de nombreuses méthodes pour fitter une SED. Ces nouvelles techniques permettent ainsi d'inférer les propriétés intrinsèques des galaxies observées (locales, globales ou les deux), de pouvoir interpoler un spectre à partir d'informations photométriques ou encore d'en estimer le redshift.

4.2.2 SED Fitter

Le fit d'une distribution énergétique spectrale d'une galaxie est la méthode première permettant d'inférer ses propriétés physiques intrinsèques à partir d'observations. Ces propriétés peuvent ensuite être confrontés aux prédictions provenant de théories d'évolution et formations de galaxies. De ce fait, l'utilisation de SED Fitter est une pratique très fréquente lorsqu'il s'agit de tester des hypothèses en astronomie extragalactique (?????).

Trois composantes sont nécessaires pour procéder à un fit de SED : un modèle physique qui décrit les différentes contributions qui la composent, des données d'observations de la galaxie (photométriques et/ou spectroscopiques) et le fitter lui-même qui va inférer la combinaison adéquate entre les modèles physiques et les observations.

De nombreuses techniques de SED Fitting ont été développées, certaines basées sur la simple optimisation de maximum de vraisemblance, parfois appelée code d'inversion, comme dans ULYSS¹(?), FIREFLY²(?) ou LEPHARE³(?) plus axé sur la détermination de redshift photométrique.

Cette technique est très populaire de part sa rapidité de calcul et une certaine simplicité à mettre en place. Néanmoins ces avantages sont bridés par certaines limites. Par exemple, de part la haute non-Gaussianité de certains espaces de vraisemblance, un léger changement dans les données d'entrées (comme un bruit dans une image photométrique de galaxie) peut conduire à de grands écarts dans les paramètres inférés (?). Par ailleurs, une méthode de maximum de vraisemblance peut être difficile à adapter à des modèles hautement non-linéaires comme l'émission par la poussière.

Dans l'optique de résoudre ces problèmes, des techniques de forward-modeling Bayésien ont à leur tour été développées. Avec cette méthode, des grilles de paramètres sont pré-calculées puis comparées aux observations. Le calcul de vraisemblance est alors très rapidement déterminée, malgré le fait que le nombre de modèle à calculer au préalable croît

1. <http://ulyss.univ-lyon1.fr>

2. <http://www.icg.port.ac.uk/firefly/>

3. <https://www.cfht.hawaii.edu/~arnouts/LEPHARE/lephare.html>

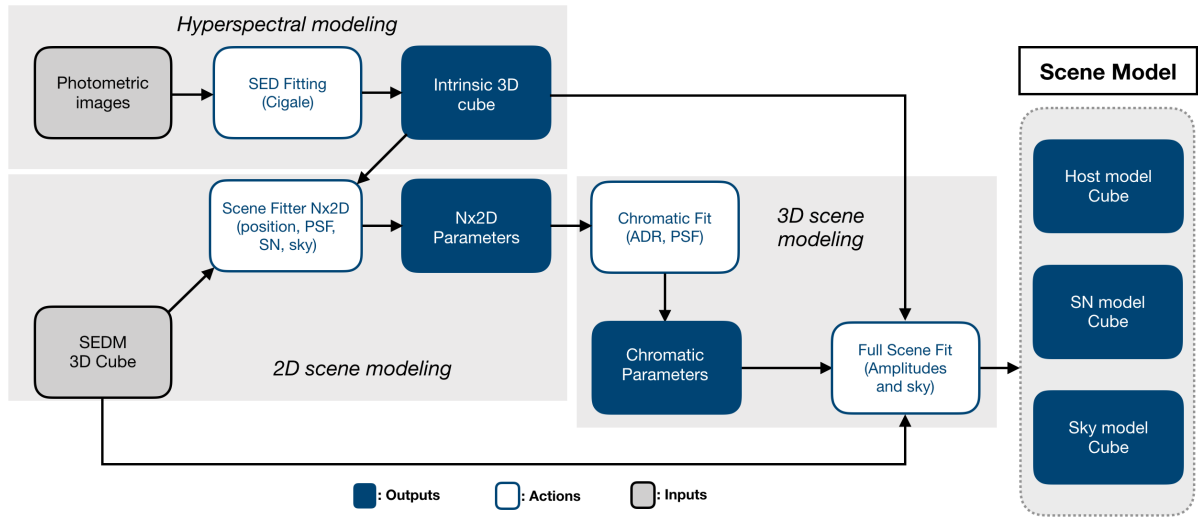


FIGURE 4.2 – Présentation du pipeline HYPERGAL .

exponentiellement à mesure que l'on rajoute des paramètres. Parmi les codes développés à partir de cette méthode, on peut citer *?*, *?*, le framework *CIGALE*¹ (???) ou encore *MAGPHYS*² (?).

Le succès du forward-modeling Bayésien a rapidement été adopté, et étendu à un couplage avec des algorithmes de Monte-Carlo par chaînes de Markov (MCMC) pour plus efficacement explorer l'espace des posterior. Cette extension, initiée par *?* avec *GALMC* (retiré du domaine public par faute de maintenance), puis rapidement suivi de codes plus récents tels que *BEAGLE*³ (?), *BAGPIPES*⁴ (??) ou encore plus récemment *PROSPECTOR*⁵ (?) et *PIXEDFIT*⁶ (?).

Nous terminerons la présentation des SED Fitter en mentionnant le site sedfitting.org, maintenu par Tamas Budavari, Daniel Dale, Brent Groves et Jakob Walcher qui regroupe la grande majorité des codes et set de modèles disponibles publiquement.

4.3 Présentation du Pipeline

Nous allons à présent introduire le modéliseur de scène HYPERGAL . Les étapes principales de ce pipeline sont présentées dans la Figure 4.2, et tracerons l'organisation de cette Partie du manuscrit.

4.3.1 Cube intrinsèque

Comme abordé dans la section 4.1, le cœur d'HYPERGAL repose sur la conception d'un cube 3D contenant la galaxie hôte isolée de sa supernova : c'est la modélisation

1. <https://cigale.lam.fr>
2. <http://www.iap.fr/magphys/>
3. <http://www.jacopochevallard.org/beagle/>
4. <https://github.com/ACCarnall/bagpipes>
5. <https://github.com/bd-j/prospector>
6. <https://github.com/aabdurrouf/PIXEDFIT>

hyperspectrale de la galaxie. Le but n'est pas de remonter aux propriétés intrinsèques de la galaxie, mais de simplement être en mesure d'interpoler un spectre cohérent à l'échelle local.

Cette étape, entièrement indépendante des observations de la SEDm, va nécessiter l'utilisation d'un SED Fitter, que nous avons introduit dans la section 4.2. Dans un premier temps, nous allons récupérer des images de différentes bandes photométriques de la galaxie hôte de la supernova détectée par ZTF. On procèdera ensuite à un fitting de SEDs de portions locales de la galaxie, ce qui permettra d'obtenir une multitude de spectres propre à chaque région de la galaxie. Avec un échantillonnage spectral adéquat, nous serons ainsi en mesure de reconstruire un cube 3D, dont les deux dimensions spatiales (x, y) seront définis par les images photométriques, et la dimension spectral par le SED Fitter. Le cube résultant ne contiendra ainsi que la galaxie hôte, et sera appelé dans la suite de ce manuscrit *cube intrinsèque*. Cette étape de modélisation hyperspectrale est détaillée dans le Chapitre ??.

4.3.2 Modélisation de scène 2D

Dans cette seconde étape, les observations de la SEDm deviennent nécessaires : le but ici est de projeter le cube intrinsèque de la galaxie dans l'espace des observations. Pour faire cela, nous allons de façon indépendante caractériser la réponse impulsionnelle spatiale et spectrale de la SEDm (Chapitre ??).

En utilisant ces informations, nous projetterons dans l'espace de la SEDm le cube intrinsèque préalablement scindé en N méta-tranches (2D). Il faudra pour cela prendre en compte la forme et la taille de l'échantillonnage spatial des deux espaces (source photométrique et MLA de la SEDm) ainsi que la différence de seeing. En plus de la composante galactique, nous caractériserons les composantes de supernova, de fond de ciel et de potentiels artefacts à modéliser pour compléter la scène. La projection de chaque meta-tranche dans l'espace SEDm sera fittée aux meta-tranches correspondantes de l'observation, dont la minimisation permettra de récupérer un set de $N \times 2D$ paramètres.

4.3.3 Modélisation chromatique et projection 3D

Les $N \times 2D$ paramètres sont ensuite utilisés pour fixer la chromaticité des composantes de la scène, comme la réponse impulsionnelle spatiale de la SEDm (fonction d'étalement de point ; PSF) ou la variation de la position des objets dans le MLA due à la réfraction de la lumière par l'atmosphère (ADR). Les modèles de chromaticités sont déterminés a priori, et les paramètres de ces modèles sont fittés à partir des $N \times 2D$ paramètres obtenus de l'étape précédente.

Une fois la chromaticité fixée, l'ensemble des paramètres de projection de chaque tranche du cube intrinsèque dans l'espace SEDm devient fixe, et seuls les paramètres d'amplitudes (fond de ciel, supernova...) sont fittés pour chaque longueur d'onde. Cette étape permet ainsi d'extraire les trois composantes de la scène d'observation de la SEDm, à savoir le background, la galaxie hôte et la source ponctuelle.

4.4 Cas pédagogique de présentation

Les chapitres suivants de cette partie du manuscrit seront consacrés à la description des différentes étapes du pipeline. Pour une illustration appropriée, nous utiliserons un cas pédagogique de modélisation de scène où la supernova est suffisamment distinguable de sa galaxie hôte.

La cible choisie est ZTF18accrorf, dont le cube extrait à partir d'une observation de la SEDm est présenté dans la Figure 4.3.

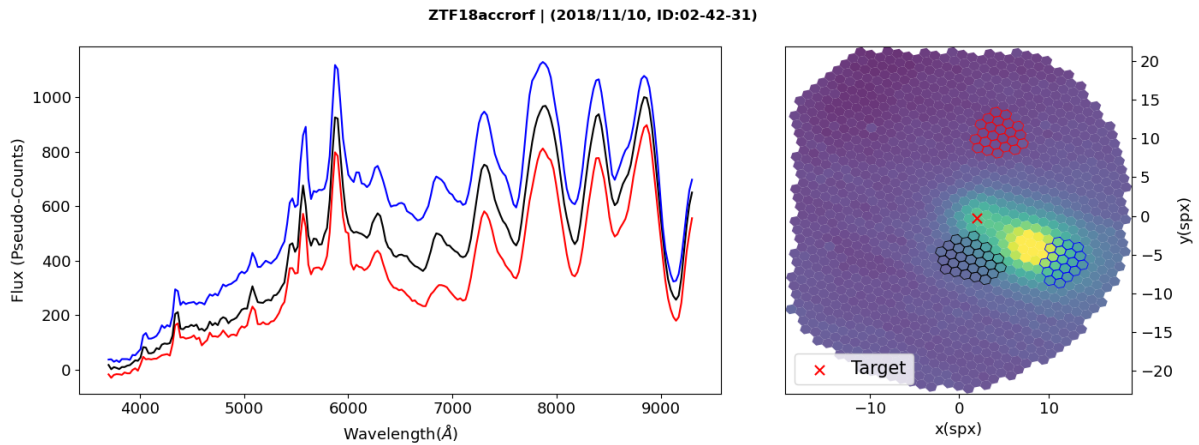


FIGURE 4.3 – Cube 3D d'une observation de ZTF18accrorf avec la SEDm. La figure de gauche montre les spectres en unité de pseudo-ADU, dont le code couleur correspond aux spaxels sélectionnés dans la figure de droite. La forme similaire des trois spectres malgré une localisation différente des spaxels sous-jacents dans le MLA est due à la présence du spectre du ciel sur l'ensemble du champ de vue. La croix rouge indique la position estimée de la supernova à partir des informations de guidage avec la Rainbow Camera. Dans ce cas ci, la supernova est assez aisément distinguable de sa galaxie hôte.

Malgré la localisation relativement excentrée de la source ponctuelle, son faible contraste vis à vis de son hôte rend difficile son extraction automatique par le pipeline `PYSEDm`. Nous illustrons cette tentative dans la Figure 4.4. Après la présentation d'`HYPERGAL` à travers le cas pédagogique, nous montrerons un cas d'extraction extrême où la supernova est quasiment confondu avec le coeur de sa galaxie hôte.

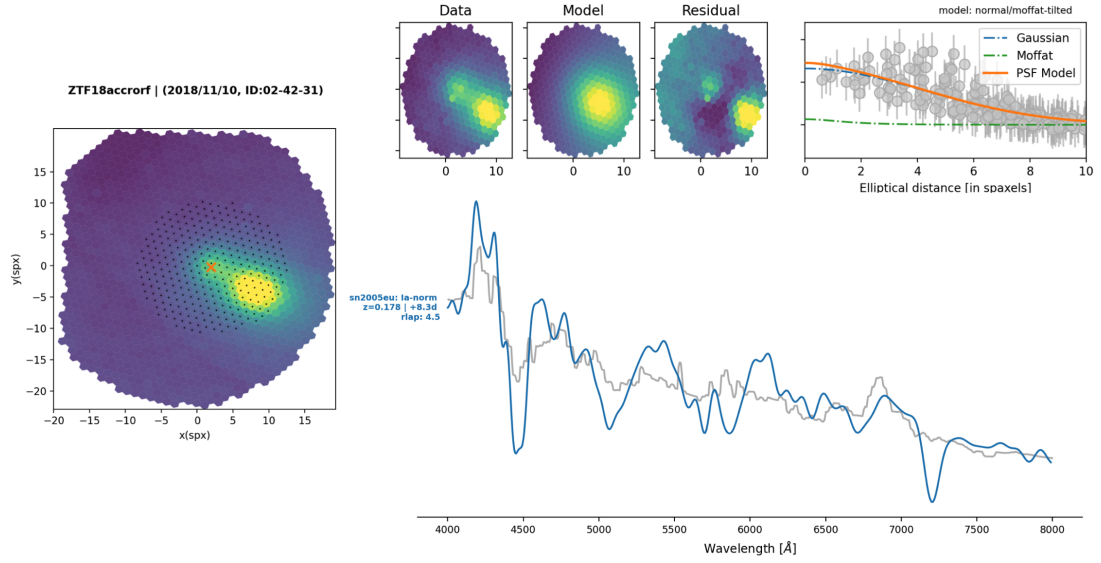


FIGURE 4.4 – Extraction de ZTF18accrofr avec **PYSED**. Sur la *gauche* est représenté une image 2D du cube intégré spectralement, avec la croix rouge indiquant la position estimée de la supernova à partir des informations de guidage avec la Rainbow Camera. Les marqueurs noirs représentent les spaxels sélectionnés pour l'extraction automatique. *En haut à droite* est représenté la modélisation d'une méta-tranche et le profil radial estimé. On voit clairement la composante de la galaxie hôte qui contamine l'extraction de la supernova. *En bas* nous montrons en unité de flux le spectre extrait par **PYSED** en gris et le meilleur modèle estimé par **SNID**. Même si certaines structures semblent rappeler le spectre d'une SNIa, comme les absorptions FeII, FeIII et MgII (entre $[4200 - 4600]\text{\AA}$, et entre $[4800 - 5200]\text{\AA}$), le niveau de confiance pour la classification reste très faible, comme l'atteste le $rlap = 4.5$, et surtout le redshift estimé de $z = 0.178$ bien au dessus de la profondeur en magnitude atteignable par ZTF.