

# Rapport final de stage de fin d'études

Mesure de la croissance des structures avec les galaxies du  
DESI BGS et les supernovae de type Ia de ZTF : vers une  
analyse jointe

Antoine Gilles-Lordet  
encadré par Pauline Zarrouk et Nicolas Regnault

2 octobre 2024





# Remerciements

Je tiens à remercier l'ensemble du LPNHE et tout particulièrement le groupe cosmologie, pour l'accueil qu'ils m'ont réservé pour mon projet de troisième année puis mon stage de fin d'études. Ce fut un véritable plaisir de travailler avec mes encadrants, Pauline Zarrouk et Nicolas Regnault, et bien que je parte effectuer ma thèse à Stockholm, j'espère que j'aurai d'autres occasions de collaborer avec eux. Je remercie également Thomas De Jaeger, Dylan Kuhn et Mahmoud Osman pour leurs aides multiples sur l'utilisation des modules **PETS**, **EDRIS** et **NaCl**, et Marc Betoule pour ses explications très claires sur toutes les méthodes d'analyses utilisées.

# Table des matières

<b>1 Synthèse</b>	<b>3</b>
<b>2 Introduction</b>	<b>5</b>
2.1 Contexte . . . . .	5
2.1.1 Analyse du clustering des galaxies pour mesurer $f\sigma_8$ . . . . .	5
2.1.2 DESI . . . . .	6
2.1.3 Analyse cosmologique avec des SN Ia pour mesurer $f\sigma_8$ . . . . .	7
2.1.4 Le projet LEMAÎTRE . . . . .	8
2.1.5 Analyse jointe . . . . .	10
2.2 Objectifs . . . . .	10
2.3 Enjeux . . . . .	11
2.4 Reformulation du problème dans son contexte . . . . .	11
<b>3 Approche et analyse</b>	<b>13</b>
3.1 État de l’art . . . . .	13
3.1.1 Mesurer $f\sigma_8$ avec le clustering des galaxies et les SN de type Ia . . . . .	13
3.2 Approche de résolution du problème . . . . .	13
3.2.1 Génération des SN Ia . . . . .	14
3.2.2 Reconstruction des vitesses particulières . . . . .	15
3.2.3 Meilleure reconstruction . . . . .	16
3.2.4 Analyse $f\sigma_8$ . . . . .	16
3.3 Résultats intermédiaires et finaux visés . . . . .	17
3.3.1 Résultats intermédiaires . . . . .	17
3.3.2 Résultats finaux . . . . .	17
3.4 Difficultés rencontrées . . . . .	17
3.5 Gestion des risques réalisée . . . . .	18
<b>4 Résultats</b>	<b>21</b>
4.1 Validation du pipeline LEMAITRE sur simulations . . . . .	21
4.1.1 Génération de SNe à partir d’Uchuu . . . . .	21
4.1.2 Filtrage par PETS . . . . .	22
4.1.3 Reconstruction avec SALT2.4 . . . . .	23
4.1.4 Analyse cosmologique sur les résultats SALT2.4 . . . . .	24
4.1.5 Reconstruction et vitesses particulières avec NaCl . . . . .	25
4.2 Problèmes identifiés dans le pipeline LEMAÎTRE et corrections apportées . . . . .	27
4.2.1 Problèmes liés à des valeurs de méta-paramètres . . . . .	27
4.2.2 Problèmes liés aux méthodes . . . . .	31

<b>5</b>	<b>Travail futur</b>	<b>35</b>
5.1	Finalisation du travail effectué lors du stage . . . . .	35
5.2	Postérité pour l'entreprise . . . . .	35
<b>6</b>	<b>Bilan personnel</b>	<b>37</b>
<b>7</b>	<b>Enjeux éthiques</b>	<b>39</b>
7.1	Éthique de la recherche . . . . .	39
7.2	Égalité des genres . . . . .	39
7.3	Impact environnemental . . . . .	40
<b>A</b>	<b>Annexes</b>	<b>i</b>
A.1	Fonctionnement des modules . . . . .	i
A.1.1	PETS . . . . .	i
A.1.2	NaCl . . . . .	ii
A.1.3	EDRIS . . . . .	iv
A.2	Méthode de l'analyse jointe $f\sigma_8$ . . . . .	vi

# Chapitre 1

## Synthèse

### Version française

La cosmologie moderne repose sur un modèle nommé  $\Lambda$ CDM, construit à partir de la relativité générale d'Albert Einstein, qui suppose que l'univers contient environ 5% de matière ordinaire, 25% de matière noire qui n'interagit que gravitationnellement, et de 60% d'énergie noire dont le rôle est d'accélérer l'expansion de l'Univers. Cette accélération est encore mal comprise, et des modèles alternatifs de gravitation propose de l'expliquer sans faire intervenir cette dernière composante. Dans ce contexte, ce stage vise à déterminer le taux de croissance des structures dans l'Univers, car l'évolution temporelle de cette grandeur dépend directement du modèle considérée. Elle constitue donc un test de la relativité générale aux échelles cosmologiques.

Ce stage se base sur les données de deux relevés, le relevé *Dark Energy Spectroscopic Instrument* (DESI) et le relevé *Zwicky Transient Facility* (ZTF). DESI a pour objectif de cataloguer et cartographier les galaxies de notre Univers proche, en mesurant leurs positions angulaires et leur redshift par spectroscopie. Ces galaxies s'éloignent de nous à cause de l'expansion de l'Univers, mais possèdent également une vitesse propre par rapport à l'expansion de l'Univers car elles sont constamment attirées par les puits de potentiels gravitationnels. Ces vitesses particulières déforment alors la distribution des redshift par effet Doppler, et il est possible de mesurer le taux de croissance des structures à partir de cette distorsion. ZTF est quant à lui un programme d'observation visant à détecter les phénomènes transitoires, en particulier les *supernovae* (SN) de type Ia. Ces SNe constituent ce qu'on appelle des chandelles standards, c'est-à-dire que leurs luminosités et leurs distances sont corrélées. On est alors capable de reconstruire les redshifts théoriques correspondant à ces distances, puis d'en déduire la contribution des vitesses particulières au redshifts observés. Cela permet de reconstruire avec précision la vitesse de leur galaxies hôtes. L'utilisation des SNe de type Ia en combinaison avec un catalogue de galaxies permet d'améliorer considérablement la précision sur la mesure du taux de croissance des structures.

Lors de ce stage, j'ai assemblée une chaîne d'analyse cosmologique et je l'ai testée sur des simulations. Pour cela des SNe de type Ia possédant des vitesses particulières sont générées selon un modèle cosmologique fixé et le catalogue de galaxies Uchuu BGS, et des observations sont simulées à l'aide des logs d'observations de ZTF. La chaîne d'analyse LEMAÎTRE actuellement en développement au LPNHE est alors utilisée afin de reconstruire les paramètres de ces SNe ainsi que leurs vitesses particulières, ce qui constitue la base d'une future mesure du taux de croissances des structures dans l'Univers à l'aide de SNe.

## English version

Modern cosmology is based on a model build from Albert Einstein's general relativity known as  $\Lambda$ CDM, which assumes that the Universe contains around 5% ordinary matter, 25% dark matter that only interacts gravitationally, and 60% dark energy whose role is to accelerate the expansion of the Universe. This acceleration is still poorly understood, and alternative models of gravitation have been proposed to explain it without involving this last component. In this context, this internship aims to determine the growth rate of structures in the Universe, as the temporal evolution of this quantity depends directly on the underlying cosmological model. It is therefore a test of general relativity on cosmological scales.

This internship uses data from two surveys, *Dark Energy Spectroscopic Instrument* (DESI) and *Zwicky Transient Facility* (ZTF). DESI aims to catalogue and map the galaxies in our nearby Universe, by measuring their angular positions and redshifts using spectroscopy. These galaxies move relatively to us due to the expansion of the Universe and also possess their own velocity relative to the expansion of the Universe, as they are constantly falling in gravitational potential wells. These peculiar velocities distort the redshift distribution by Doppler effect and it is possible to measure the growth rate of structures from this distortion. ZTF is an observation programme designed to detect transient phenomena, in particular type Ia *supernovae* (SNe). These SNe are called standard candles which means that their luminosities and distances are correlated. We can then reconstruct the theoretical redshifts according to their distances and deduce the component arising from peculiar velocities in the observed redshifts. This allows to accurately reconstruct the velocity of their host galaxies. Using type Ia SNe in combination with a galaxy catalogue considerably improves the accuracy of the growth rate of structures.

During this internship, I assembled a cosmological analysis pipeline and tested it on simulations. Type Ia SNe with particular velocities are generated according to a fixed cosmological model and the Uchuu BGS galaxy catalogue and observations are simulated using ZTF observation logs. The LEMAÎTRE analysis pipeline currently under development at LPNHE is then used to reconstruct the parameters of these SNe as well as their peculiar velocities, laying the groundwork for a future measurement of the growth rate of structures in the Universe using SNe.

# Chapitre 2

## Introduction

### 2.1 Contexte

Le modèle standard de la cosmologie, nommé  $\Lambda$ CDM, suppose que la gravité est décrite à toutes les échelles par la relativité générale, et que les principales contributions à la gravité viennent de la matière noire froide (*Cold Dark Matter*) et de l'énergie noire (sous la forme d'une constante cosmologique notée  $\Lambda$ ). Ces deux inconnues sont nécessaires pour décrire l'évolution de l'Univers et permettent à ce jour de reproduire fidèlement les observations, en particulier l'expansion accélérée de l'Univers. Cependant des modèles alternatifs de gravité ont également été proposés pour expliquer cette accélération sans faire intervenir de constante cosmologique, et prédisent des évolutions temporelles différentes de celles de la relativité générale pour certaines quantités.

Parmi ces quantités, on trouve notamment le taux de croissance des structures en fonction du redshift <sup>1</sup>  $f(z)$ , et l'écart-type du champ de densité dans une sphère de rayon  $8h^{-1}\text{Mpc}$ ,  $\sigma_8$ . Ces deux quantités sont intéressantes car on peut montrer les vitesses particulières des galaxies, c'est-à-dire leurs vitesses par rapport à l'expansion de l'univers, sont proportionnelles au produit  $f(z)\sigma_8(z)$  et caractérisent ainsi la formation des structures observées dans l'univers.

En plus de la nature inconnue de la matière noire et de l'énergie noire, le modèle  $\Lambda$ CDM présente également des tensions sur certains paramètres. Une tension survient lorsque des analyses utilisant différents types d'observation obtiennent une différence significative de valeur de paramètres. La plus connue est la tension sur la constante de Hubble  $H_0$  <sup>2</sup>, mais il existe également une tension sur  $\sigma_8$  (voir les valeurs en Table 2.1. Il est donc particulièrement important de déterminer sa valeur avec précision.

#### 2.1.1 Analyse du clustering des galaxies pour mesurer $f\sigma_8$

Pour mesurer la croissance des structures, la méthode la plus commune se base sur les *Redshift Space Distortion* (RSD) (KAISER, 1987), des distorsions liées aux transformations des coordonnées comobiles vers l'espace des redshifts. Le redshift mesuré, par exemple par spectroscopie, ne

---

1. En français, *décalage vers le rouge*, effets de dilatation des longueurs d'ondes des photons entre leur émission par une source et leur observation sur Terre causé par l'expansion de l'Univers. Le terme anglais sera utilisé plutôt que sa traduction française par habitude et pour ne pas surcharger le texte.

2. les mesures directes utilisant le fond diffus cosmologique donne  $H_0 \sim 67.3 \pm 0.5 \text{ km.s}^{-1}.\text{Mpc}^{-1}$  (PLANCK COLLABORATION et al., 2020), tandis que les mesures indirectes reposant sur des supernovae ou des phénomènes de lentillage gravitationnel tendent vers  $H_0 \sim 73.2 \pm 1.3 \text{ km.s}^{-1}.\text{Mpc}^{-1}$  (RIESS et al., 2022).



Relevé	valeur
Planck2018 (PLANCK COLLABORATION et al., 2020)	$\sigma_8 = 0.811 \pm 0.006$
DES Y3 + KIDS-1000 (DARK ENERGY SURVEY et al., 2023)	$\sigma_8 = 0.825^{+0.067}_{-0.073}$
SDSS IV (EBOSS COLLABORATION et al., 2021)	$\sigma_8 = 0.85 \pm 0.03$
DESI (CHEN et al., 2024)	$\sigma_8 = 0.878^{+0.089}_{-0.080}$

TABLE 2.1 – Mesures de  $\sigma_8$  avec différentes sondes

contient pas uniquement le redshift cosmologique dû à l’expansion de l’univers, mais inclut également une contribution par effet Doppler due aux vitesses particulières des galaxies, c’est-à-dire leurs vitesses propres par rapport à l’expansion de l’Univers.

Ce terme supplémentaire déplace les galaxies dans l’espace des redshifts par rapport à l’espace comobile, et les corrélations spatiales du champ de densité deviennent alors anisotropes : le long de la ligne de visée, les galaxies semblent plus regroupées qu’orthogonalement à la ligne de visée. A grande échelle, l’amplitude de cette anisotropie étant proportionnelle au facteur de croissance des structures  $f$  et à l’amplitude des fluctuations du champ de densité, représenté communément par  $\sigma_8$  (l’écart-type du champ de densité dans une sphère de rayon  $8h^{-1}\text{Mpc}$ ), ces analyses du clustering des galaxies contraignent le paramètre composite  $f(z)\sigma_8(z)$ . De nombreuses analyses des distorsions dans l’espace des redshifts ont été effectuées à l’aide de relevés spectroscopiques, tels que 6dFGRS (BEUTLER et al., 2012), SDSS-MGS (HOWLETT et al., 2015), FastSound (OKUMURA et al., 2016), SDSS-III BOSS (ALAM et al., 2017) ou SDSS-IV eBOSS (EBOSS COLLABORATION et al., 2021). Les dernières mesures atteignent une précision de l’ordre de 10%, et sont compatibles avec la relativité générale.

Une autre méthode pour mesurer  $f\sigma_8$  est de dériver le paramètre des mesures directes des vitesses particulières des galaxies. Les vitesses particulières peuvent être mesurées directement à condition de pouvoir mesurer indépendamment les redshifts et les distances absolues des galaxies. Des mesures précises des redshifts peuvent être obtenues par spectroscopie et des estimations des distances peuvent être obtenues en utilisant les corrélations entre les distances et d’autres observables, telles que la relation de Tully-Fisher pour les galaxies spirales (corrélation entre la vitesse radiale de ses étoiles et la luminosité totale de la galaxie, TULLY et FISHER, 1977) ou la méthode du Plan Fondamental (DJORGOVSKI et DAVIS, 1987). Ces corrélations sont utilisables uniquement pour des galaxies à des redshifts relativement faibles ( $z < 0.1$ ) car les incertitudes augmentent rapidement avec le redshift. Les propriétés statistiques d’un échantillon de vitesses particulières peuvent ensuite être utilisées seules ou en combinaison avec un champ de densité de galaxies afin de déterminer  $f\sigma_8$  (ADAMS et BLAKE, 2020 ; QIN et al., 2019 ; TURNER et al., 2023).

### 2.1.2 DESI

De telles analyses font parties des objectifs du relevé *Dark Energy Spectroscopic Instrument* (DESI). Ce programme d’observation utilise 5000 fibres optiques positionnées à l’aide de robots pour réaliser des spectres des galaxies nous entourant, afin de construire une cartographie 3D de l’univers proche, jusqu’à  $z = 1.7$ . Ce relevé constitue le relevé de galaxies le plus dense jamais réalisées, les premières données publiées (nommées DESI EDR, *Early Data Release*) représentent 700 000 objets, et seulement 1% du volume final couvert. La publication des données DESI Y1 (*Year 1*), dont la prise s’est arrêtée en juin 2022 est prévue pour 2025, mais a déjà permis la réalisation d’analyses cosmologiques (DESI COLLABORATION et al., 2024c, 2024a, 2024b) ainsi que la production de catalogues simulés fidèles à l’univers observé, tel que le catalogue Uchuu.

Ce catalogue a été réalisé à partir de la simulation éponyme Uchuu<sup>3</sup> (PRADA et al., 2023), une simulation à N-corps utilisant 2.1 milliards de particules, peuplée avec des galaxies du *DESI Bright Galaxy Survey* (BGS) qui est un set complet de galaxies lumineuses<sup>4</sup> à bas redshift ( $z < 0.4$ ) (HAHN et al., 2023), pour produire un univers statistiquement fidèle au notre. Le catalogue ainsi obtenu reproduit le champ de densité de la simulation Uchuu mais aussi le champ de vitesse. Il donne donc accès aux redshifts cosmologiques et aux redshifts observés, et peut être utilisé comme base d’une analyse  $f\sigma_8$  sur des simulations.

### 2.1.3 Analyse cosmologique avec des SN Ia pour mesurer $f\sigma_8$

À bas redshift, un autre moyen d’obtenir les distances vient des supernovae (SNe) de type Ia (HOYLE et FOWLER, 1960). Ces SNe sont des explosions thermonucléaires de naines blanches dans un système binaire, riches en carbone et oxygène, et présentent la particularité d’avoir des conditions d’explosions très similaires les unes aux autres. En particulier, les évolutions temporelles de leur luminosité dans leurs référentiels propres sont très similaires. Le maximum de flux émis, environ  $10^{10}$  fois plus que le soleil, est corrélé à leurs durées et leurs couleurs intrinsèque selon deux lois, communément appelées *brighter-slower* et *brighter-bluer*. Les SNe plus brillantes présentent une décroissance plus lente (donc durent plus longtemps), et présentent un décalage de leurs longueurs d’ondes émises vers le bleu, qui n’est pas du à un effet de redshift.

Grâce à ces deux corrélations, il est possible de décrire leurs flux maximaux à l’aide de deux paramètres, nommés *stretch*<sup>5</sup> et couleur, et d’exprimer leur flux maximal en fonction d’un flux moyen et de contributions liées au stretch et à la couleur.

Cette caractéristique rend les SNe Ia standardisables et en fait d’excellentes chandelles standard : des événements astrophysiques dont la luminosité peut être modélisée, ce qui permet d’en déduire directement leurs distances. Les SNe Ia ont ainsi permis la découverte de l’accélération de l’expansion de l’Univers (PERLMUTTER et al., 1998 ; RIESS et al., 1998).

Pour exprimer cette standardisation, il est plus commode de travailler avec des magnitudes plutôt que des flux. Les magnitudes sont définies par  $M = \log_{10}(\{ \})$ , et les magnitudes absolues par  $M_{abs} = \log_{10}(\{ / \{_{ref} \})$ , où  $\{_{ref}$  est le flux qu’aurait l’objet s’il se situait à une distance de 10 parsecs. La standardisation des SN Ia est alors décrite par la formule de Tripp (TRIPP, 1998)

$$M_{b,SN}^* = M_b - \alpha x_{1,SN} + \beta c_{SN} + n(\sigma_{int})\sigma \quad (2.1)$$

où  $M_{b,SN}^*$  est la magnitude absolue en bande B de la SN,  $x_{1,SN}$  son stretch et  $c_{SN}$  sa couleur.  $M_b$  est la magnitude absolue moyenne à  $c = 0$  et  $x_1 = 0$  dans la bande B. Expérimentalement, on mesure  $M_b = 19.3$ , ainsi que les coefficients  $\alpha = 0.14$  et  $\beta = 3.15$ . En réalité, la standardisation n’est pas parfaite et une dispersion intrinsèque subsiste, représentée par  $\sigma_{int}$ . Cette dispersion est modélisée par une distribution gaussienne d’écart type 0.10.

L’outil principal pour exploiter les SNe est le diagramme de Hubble, la représentation de la relation entre les distances des SNe, représentées par leurs magnitudes, avec leurs vitesses, représentée par leurs redshifts. Un tel diagramme est représenté en Figure 2.1. Un modèle peut alors être ajusté, et les résidus au modèle donnent accès aux vitesses particulières des SNe (DAVIS et al., 2011). En effet, les magnitudes des SNe dépendent de leurs redshift cosmologiques, dus à l’expansion de l’univers, et non de leurs redshift observé, qui inclue les vitesses particulières. Les SNe apparaissent alors comme décalées horizontalement par rapport au diagramme de Hubble, et leurs vitesses particulières sont contenues dans les résidus au modèle.

3. <https://skiesanduniverses.org/Simulations/Uchuu/>

4. ce sous-échantillon est défini comme ayant une magnitude en bande r inférieure à 19.5

5. Comme pour le redshift, le paramètre de *stretch* n’a pas d’équivalent français utilisé dans la communauté, le terme anglais sera donc employé à la place.

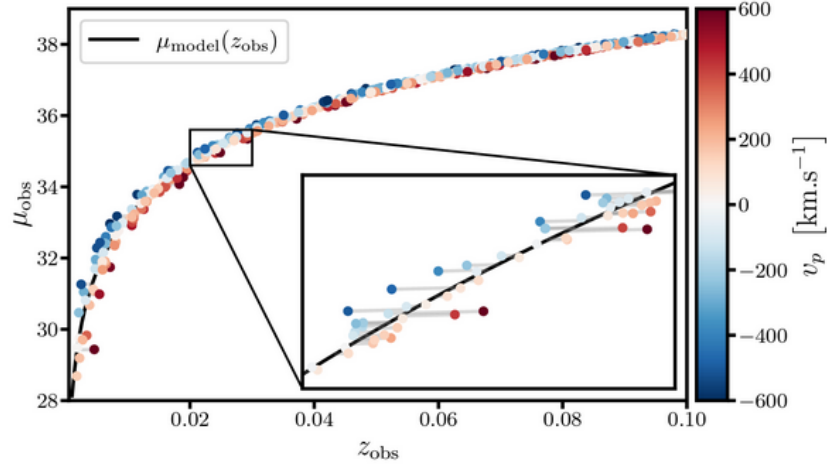


FIGURE 2.1 – Diagramme de Hubble de SNe simulés. Les vitesses particulières selon la ligne de visée provoque un écart entre leur redshift cosmologique et le redshift observé, ce qui se traduit par un décalage horizontal des SNe par rapport à la ligne de base du diagramme de Hubble. Les SNe ayant une vitesse particulière positive s'éloignent de nous et nous apparaissent plus rouges, tandis que celles ayant une vitesse particulière négative nous apparaissent plus bleues. Crédit : B. Carreres

Les analyses de clustering de galaxies sont moins performantes à bas redshift, car le volume étudié ne contient qu'un faible nombre de galaxies et cette statistique est insuffisante pour produire des analyses précises. Les SNe Ia constituent alors un bien meilleur traceur du champ de vitesses à bas redshift que les galaxies, mais leur statistique est trop faible pour décrire le champ de densité. Une analyse jointe reposant sur les deux types de traceurs permet d'obtenir de meilleurs contraintes sur la croissance des structures qu'une analyse uniquement des galaxies (cf Figure 2.2).

#### 2.1.4 Le projet Lemaître

La réalisation d'un diagramme de Hubble de nouvelle génération utilisant des lots encore non exploités de SN Ia des relevés *Zwicky Transient Facility* (ZTF, BELLM et al., 2018), SuperNova Legacy Survey (SNLS, PRITCHET et SNLS COLLABORATION, 2004) et Hyper Suprime-Cam Subaru Strategic Programm (HSC-SSP, AIHARA et al., 2018 ; MIYAZAKI et al., 2012) est actuellement l'objet du projet LEMAÎTRE au LPNHE. Ce projet vise à développer un nouveau pipeline d'analyse cosmologique des SN Ia à partir de zéro, adapté au traitement des volumes de données attendus des prochains programmes d'observation tels que LSST (THE LSST DARK ENERGY SCIENCE COLLABORATION et al., 2021). Le lot ZTF est appelé DR2.5, pour *Data Release 2.5*, car il couvre les deux premières années de ZTF en prenant en compte des effets des capteurs jusqu'à présent inconnus. Ce projet utilisant des lots non exploités, la mesure des paramètres du modèle sera donc indépendante des mesures précédentes.

La partie analyse cosmologique du pipeline LEMAÎTRE<sup>6</sup> (voir également la Figure 2.3) est composé des modules suivants :

6. Un module de calibration est présent en amont, mais n'est pas l'objet de ce stage et ne sera donc pas abordé.

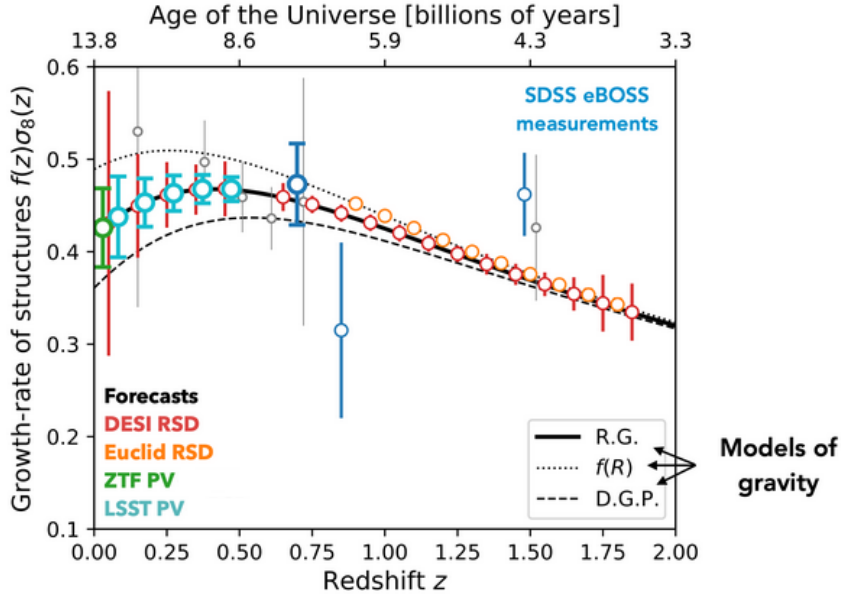


FIGURE 2.2 – Prévisions des contraintes sur  $f\sigma_8$  pour les relevés DESI (HAHN et al., 2023), Euclid (EUCLID COLLABORATION et al., 2024) et la combinaison de DESI avec des vitesses particulières de ZTF (CARRERES et al., 2023) ou LSST (HOWLETT et al., 2017). Credit : J. Bautista

1. Le module PETS (Preprocessing and sElection of a Training Sample) effectue une première sélection en éliminant les points de photométrie ainsi que les SNe mal échantillonnées. Il donne également une première évaluation des paramètres de standardisation des SNe, qui sert de point de départ pour NaCl. La principale difficulté que doit résoudre ce module est de proposer une sélection dont l'efficacité est indépendante du redshift et du rapport signal sur bruit, car cela introduirait des biais dans la suite de l'analyse. Plus de détails sont disponibles en Annexe A.1.1.
2. Le module NaCl (Nouvel algorithme de Courbe de lumière) évalue ensuite les luminosités des SNe dans leurs référentiels propres, ainsi que leurs paramètres de standardisation en entraînant un modèle empirique. Il vise à ré-implementer un modèle type SALT2.4 adapté à l'analyse des volumes de données des relevés futures. De plus, les incertitudes sont propagées lors de l'entraînement, et ne nécessitent pas leurs caractérisations a posteriori sur des simulations. Le fonctionnement détaillé de ce module est présenté en Annexe A.1.2.
3. Le module EDRIS (Estimateur de Distance pour les Relevés Incomplets de Supernovae) utilise les paramètres de standardisation et leur matrice de covariance pour ajuster un modèle cosmologique, en prenant en compte des effets de sélection de manière analytique. Là encore, l'estimation des biais introduit par des effets de sélections sur les modèles est habituellement traitée a posteriori sur des simulations. Le détail de ce module est présenté en Annexe A.1.3.

Un des objectifs à court terme du projet LEMÂTRE qui guide le développement des modules est de disposer d'un pipeline d'intégration continue. Cela signifie que dès qu'un module est modifié ou que les données changent, une nouvelle analyse cosmologique complète est automatiquement réalisée. L'ambition est d'ainsi pouvoir continuer à développer les modules existants, en prenant en compte différents effets ou biais et d'obtenir rapidement une estimation des gains et des biais

potentiels sur les paramètres cosmologiques.

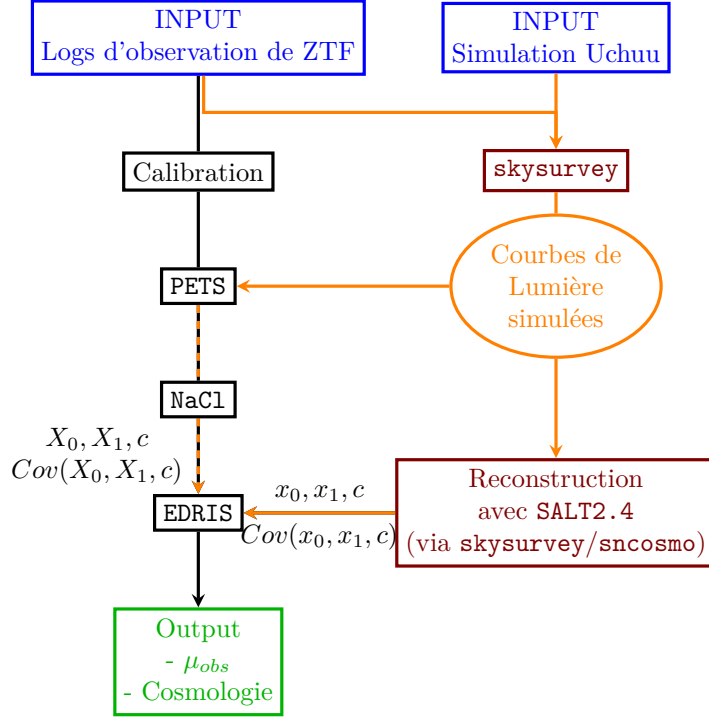


FIGURE 2.3 – Chaîne d’analyse LEMAÎTRE développée au LPNHE (représentée en noire), et travail d’intégration effectué (en orange). L’étape de calibration contient en réalité plusieurs modules, mais elle n’a pas été utilisée pour ce stage car les données simulées n’incluaient pas d’effets de calibration. `skysurvey` ne fait pas partie intégrante de LEMAÎTRE, il s’agit d’un software indépendant

### 2.1.5 Analyse jointe

Il est également possible de cumuler les deux analyses, en introduisant dans le clustering des galaxies les vitesses particulières des SNe. Une telle analyse  $f\sigma_8$  a par exemple été menée dans BORUAH et al., 2022 ; STAHL et al., 2021. Ces analyses jointes permettent de meilleures contraintes sur  $f\sigma_8$  à bas redshift que pour les analyses de RSD des galaxies seules (voir Fig. 2.2), et sont donc particulièrement intéressantes obtenir des mesures de précision.

## 2.2 Objectifs

L’objectif du stage est double :

1. Produire des échantillons simulés de SNe basés sur les positions des galaxies de la simulation Uchuu qui reproduit les données spectroscopiques du DESI BGS, ainsi que les observations de ZTF, et les utiliser pour tester le pipeline LEMAÎTRE actuellement en développement au LPNHE pour produire les diagrammes de Hubble et reconstruire la cosmologie sous-jacente des échantillons. Le pipeline étant encore en développement, je développe l’intégration des différents modules et je les teste.

2. Comparer les résultats de ce pipeline, *i.e.* les positions ainsi que les vitesses particulières reconstruites, aux données d’entrée et les utiliser pour
  - (a) Estimer  $f\sigma_8$  à partir des vitesses particulières de SNe de ZTF et comparer ces résultats avec ceux de B. Carreres (CARRERES et al., 2023).
  - (b) Produire une analyse jointe du DESI BGS et des SNe de ZTF pour quantifier le gain sur les contraintes de  $f\sigma_8$ .

## 2.3 Enjeux

L’enjeu majeur de ce stage est la validation du pipeline LEMAÎTRE sur des simulations. En particulier, l’intégration faisant partie des tâches de ce stage, les différents modules n’ont été testés qu’individuellement sur des données simplifiées ou approximées. Les simulations de SNe utilisées et passées à travers tous les modules sont plus réalistes et permettront de mettre en lumière les points critiques dans l’exécution du pipeline. Ces points critiques incluent (sans être restreints à) le paramétrage des modules, des différences éventuelles de convention ou de nommage des variables utilisées, des erreurs dans les modèles, approximations ou dans leur implémentation et la caractérisation des biais potentiels. Le fait de travailler sur des simulations permet un plus fin contrôle des conditions de test que les données réelles, et de facilement prendre en compte ou retirer différents effets.

Un second enjeu est l’estimation du gain offert par LEMAÎTRE, et plus généralement des SNe de ZTF, dans l’estimation de  $f\sigma_8$  par rapport à une analyse du seul clustering du DESI BGS. Comme évoqué précédemment, la détermination précise de ce paramètre est essentielle pour pouvoir valider ou invalider différents modèles de gravitation.

## 2.4 Reformulation du problème dans son contexte

L’estimation précise du paramètre  $f\sigma_8$  est essentielle à la cosmologie actuelle. Elle permet a minima de contraindre les paramètres du modèle  $\Lambda$ CDM, et de valider ou invalider différents modèles de gravités dont la relativité générale afin d’améliorer notre compréhension de l’Univers. Cette estimation nécessite, à bas redshift, l’utilisation de traceurs plus performants que les seuls catalogues de galaxie du fait du faible volume exploré. Les données de SNe Ia peuvent fournir les vitesses particulières nécessaires à cette estimation, mais les relevés récents et futurs nécessiteront de traiter des volumes de données plus importants qu’auparavant. Dans ce contexte, le développement d’une nouvelle chaîne d’analyse adaptée est fondamental, et est l’objectif du pipeline LEMAÎTRE. Ce pipeline nécessite cependant d’être testé sur des simulations afin de diagnostiquer d’éventuels problèmes ou biais, et de quantifier la précision sur les vitesses particulières reconstruites.



# Chapitre 3

## Approche et analyse

### 3.1 État de l’art

#### 3.1.1 Mesurer $f\sigma_8$ avec le clustering des galaxies et les SN de type Ia

Plusieurs méthodes ont été développées pour extraire des mesures du taux de croissance à partir de vitesses particulières :

- la méthode dite du maximum de vraisemblance, où les champs de vitesse (et de densité) sont supposés être tirés de distributions Gaussiennes multivariées et sur lesquels un ajustement de  $f\sigma_8$  est réalisé par maximum de vraisemblance (ADAMS et BLAKE, 2020; CARRERES et al., 2023; HOWLETT et al., 2017; HUTERER et al., 2017; JOHNSON et al., 2014; LAI et al., 2022);
- l’analyse de statistiques compressées à deux points telles que la fonction de corrélation à deux points, le spectre de puissance, ou les vitesses moyennes par paire sur lesquelles  $f\sigma_8$  est ajusté (DUPUY et al., 2019; NUSSER, 2017; QIN et al., 2019; TURNER et al., 2023);
- la comparaison entre les vitesses observées et celles reconstruites à partir du champ de densité (BORUAH et al., 2020; CARRICK et al., 2015; DAVIS et al., 2011; SAID et al., 2020; STAHL et al., 2021). En utilisant un modèle de gravitation, on peut exprimer le champ de vitesse à partir du champ de densité et du taux de croissance des structures. Cette relation peut alors être utilisée pour obtenir  $f\sigma_8$  en faisant concorder les vitesses particulières observées et celles obtenues à partir du champ de densité;
- l’inférence dite *fieldlevel*, qui consiste à retracer l’évolution complète de l’univers à partir de conditions initiales à l’aide d’un modèle, puis à comparer le résultat obtenu aux observations des champs de densité et vitesses, et rétro-propager les erreurs jusqu’au conditions initiales (BORUAH et al., 2022; PRIDEAUX-GHEE et al., 2023).

Un résumé des valeurs obtenues par ces diverses analyse est représenté en Fig. 3.1. J’utilise pour l’analyse  $f\sigma_8$  la méthode décrite dans BORUAH et al., 2020; STAHL et al., 2021, qui est de comparer les vitesses observées reconstruites avec des SN Ia aux vitesses déduites du champ de densité. Une description plus détaillée est donnée en Annexe A.2.

### 3.2 Approche de résolution du problème

La méthode employée pour la génération suit celle de CARRERES et al., 2023 pour les simulations, c’est à dire de tirer les paramètres de standardisation de distributions expérimentales, puis



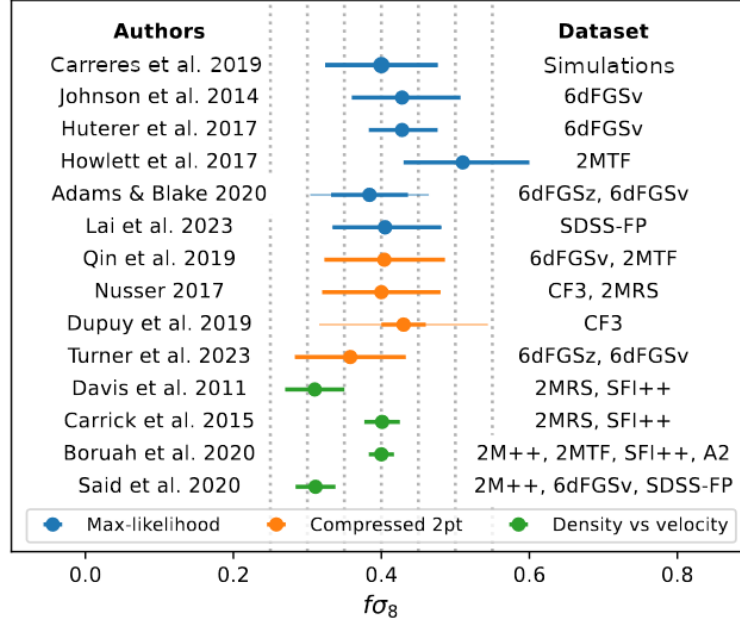


FIGURE 3.1 – Mesures du coefficient de croissance des structures  $f\sigma_8$  à partir de vitesses particulières et de catalogues de galaxies. Les barres d'erreurs en trait fin inclues les erreurs systématiques, à l'exception de Dupuy et al. 2019, pour lequel la contribution supplémentaire vient de la variance cosmique. Credit : CARRERES et al., 2023.

de générer des observations à partir de ces paramètres et des logs d'observations réels de ZTF. La reconstruction de vitesses particulières des SNe utilisera elle les nouveaux modules développés pour le projet LEMAÎTRE. L'analyse  $f\sigma_8$  prévue suit quand à elle celle de BORUAH et al., 2020 ; STAHL et al., 2021.

### 3.2.1 Génération des SN Ia

Les vitesses particulières des SNe simulées ne peuvent pas être tirées aléatoirement, puisqu'elles doivent être en accord avec les structures de galaxies et les unes avec les autres. Une possibilité est de simuler la dynamique de la matière noire d'un univers via une simulation à N-corps, et de la peupler de galaxies en fonction de la distribution de densité obtenue.

Dans un premier temps, un échantillon de SN Ia est généré selon les distributions observées des paramètres  $x_0$ ,  $x_1$ ,  $c$  et  $t_0$  (voir Table 3.1), et les positions et redshifts des galaxies BGS simulées dans Uchuu à l'aide de `skysurvey`<sup>1</sup>.

Seules les galaxies à redshift  $z < 0.06$  sont utilisées afin de s'affranchir du biais de Malmquist (voir notamment CARRERES et al., 2023 ou BOYD et al., 2024). Ce biais observationnel vient du fait que les télescopes ont une magnitude limite d'observation, au-delà de laquelle les SNe ne sont plus détectées. Or, à un redshift donné, les SNe couvrent une plage de magnitude du fait de la dispersion due aux paramètres de standardisation, à la dispersion intrinsèque et aux vitesses particulières. Ainsi, à haut redshift, les SNe les moins brillantes ne sont pas observées, ce qui

1. <https://skysurvey.readthedocs.io/en/latest/>

Paramètre	Distribution
Temps de maximum du flux en bande B $t_0$	Uniforme entre le 02/03/2018 et le 01/01/2021
Stretch $x_1$	Gaussienne bimodale (NICOLAS et al., 2021)
Couleur $c$	Loi exponentielle convoluée avec une gaussienne (GINOLIN et al., 2024)
Magnitude absolue $m_{abs}$	Obtenue par la formule de Tripp (eq 2.1)
Magnitude observée $m_{obs}$	$m_{obs} = 5 \log(d_L(z_{cosmo})) + 25 + m_{abs}$
Amplitude $x_0$	$x_0 = 10^{0.4(m_b - m_{obs})}$

TABLE 3.1 – Distributions utilisées pour le tirage des paramètres des SN Ia. Il est important de noter que le redshift utilisé pour le calcul des magnitudes observées utilise le redshift cosmologique, et non le redshift observé.

biaiseraient une estimation des paramètres cosmologique. ZTF est considéré complet jusqu'à  $z = 0.06$ , ce qui signifie que l'impact du biais de Malmquist en dessous de ce redshift est négligeable.

La période du 02/03/2018 au 01/01/2021 utilisée pour le tirage du temps de maximum correspond à la plage temporelle de la DR2 de ZTFRIGAULT et al., 2024. Cette restriction sert à la fois à avoir un volume de donné proche de celui que devra traiter le pipeline LEMAÎTRE pour estimer sa rapidité, et à pouvoir quantifier le gain qui pourrait être obtenu sur  $f\sigma_8$  avec des données réelles. Le taux d'explosions de SN Ia est pris égal à  $2,35.10^4 \text{Gpc}^{-3}$  (PERLEY et al., 2020). Pour le calcul de la magnitude observée,  $d_L(z)$  représente la distance de luminosité, qui est calculée à l'aide d'un modèle  $\Lambda\text{CDM}$  contraint par les données de Planck 2015 (PLANCK COLLABORATION et al., 2016) pour correspondre au modèle cosmologique utilisé par la simulation Uchuu.

Une simulation du relevé SNLS est également réalisée afin d'avoir un échantillon de SNe à plus haut redshift pour mieux contraindre le modèle cosmologique. Comme les redshifts observés par SNLS vont environ de  $z = 0.1$  à  $z = 1.2$ , ils sont trop élevés pour le catalogue Uchuu BGS. Les positions des SNe sont donc tirées uniformément dans la footprint SNLS, et n'incluent pas d'effets de vitesses particulières. Seule la simulation de ZTF sera utilisée pour la reconstruction des vitesses particulières.

### 3.2.2 Reconstruction des vitesses particulières

Une fois les paramètres tirés, des points de données d'observations photométriques et spectroscopiques sont générés en utilisant le modèle de SNe SALT2.4 (GUY et al., 2007; RIGAULT et al., 2024) et les logs d'observations de ZTF. Ce modèle consiste en trois fonctions  $M_0(p, \lambda)$ ,  $M_1(p, \lambda)$  et  $CL(\lambda)$  de manière à décrire les flux émis par les SNe dans leurs référentiels propres par :

$$F(SN, p, \lambda) = x_0 \times [M_0^{\text{SALT2.4}}(p, \lambda) + x_1 M_1^{\text{SALT2.4}}(p, \lambda)] \times \exp[c CL^{\text{SALT2.4}}(\lambda)] \quad (3.1)$$

où  $p$ , nommé phase, est le temps dans le référentiel de la SN depuis la date du maximum de luminosité dans la bande B,  $\lambda$  est la longueur d'onde dans le référentiel de la SN,  $x_0$  est l'amplitude du flux, et  $x_1$  et  $c$  sont les paramètres de standardisation de stretch et couleur.  $M_0$  et  $M_1$  sont appelés les modèles et  $CL$  la loi de couleur.

Une fois les paramètres de standardisation obtenus, il est alors possibles d'en déduire  $M_b^*$  à l'aide de la formule de Tripp (eq 2.1), puis d'obtenir les modules de distances

$$\mu = m_b - M_b^* = m_b - M_b + \alpha x_{1,SN} - \beta c_{SN} \quad (3.2)$$

où  $m_b$  est la magnitude maximale dans la bande B. Comme les modules de distances ne dépendent que du redshift via le modèle cosmologique, en théorie  $\mu = \mu(z)$ . Comme mentionné en section 3.1.1 les vitesses particulières ont pour effet de décaler les redshifts observés. Pour obtenir le redshift cosmologique, on inverse donc la relation  $\mu(z)$  par interpolation pour obtenir une relation  $z(\mu)$ , puisque comme les SNe sont des chandelles standards, connaître leurs modules de distance revient à connaître leurs distances, auxquelles on peut associer un unique redshift. On obtient alors les vitesses particulières en utilisant :

$$v_p = c(z_{obs} - z(\mu)) \quad (3.3)$$

avec  $c$  la vitesse de la lumière<sup>2</sup>.

Il est important de noter que du fait de la modélisation de la relation  $\mu(z)$  dans le cadre du modèle  $\Lambda$ CDM, il existe une dégénérescence complète entre  $m_b$  et la constante de Hubble  $H_0$  qui ne peut pas être levée. Cela signifie qu’une analyse cosmologique n’utilisant que des SNe Ia ne peut pas contraindre  $H_0$ , et que tout biais additif sur les modules de distances peut-être absorbé dans  $m_b$ . Autrement dit, seule la pente d’un diagramme de Hubble tel que celui représenté en Figure ?? est importante, les valeurs exactes de  $\mu$  ne le sont pas.

### 3.2.3 Meilleure reconstruction

Une meilleure reconstruction est possible en ré-entraînant le modèle lors de l’ajustement des paramètres. En effet, le modèle SALT2.4 a été entraîné, puis les incertitudes évaluées sur des simulations. Or le modèle n’était pas nécessairement bien contraint sur l’intégralité de sa surface, et les incertitudes ne sont donc pas toujours fiables. Un ré-entraînement incluant la propagation des incertitudes devrait donc permettre d’améliorer la reconstruction du flux

$$F(SN, p, \lambda) = X_0 \times [M_0^{\text{NaCl}}(p, \lambda) + X_1 M_1^{\text{NaCl}}(p, \lambda)] \times \exp[0.4C CL^{\text{NaCl}}(\lambda)] \quad (3.4)$$

Les paramètres d’amplitude, de stretch et de couleurs utilisés par NaCl sont notés  $(X_0, X_1, C)$  et non  $(x_0, x_1, c)$  comme SALT2.4, car du fait des différences dans les contraintes et du ré-entraînement des modèles (voir Annexe A.1.2), rien ne garantit que ces paramètres soient exactement les mêmes. Dans les faits, ce n’est d’ailleurs pas le cas, mais les modules de distance doivent eux rester consistants entre SALT2.4 et NaCl pour assurer qu’aucun biais n’est introduit dans la cosmologie.

### 3.2.4 Analyse $f\sigma_8$

L’analyse  $f\sigma_8$  initialement prévue n’a pas encore pu être réalisée, et ce principalement pour une raison de disponibilité des données. La méthode prévue, de comparaison entre les vitesses particulières déduites du champ de densité de matière noire et celles des SNe, a besoin d’un champ de densité décrivant le même univers que les SNe. Ces champs de densité sont disponibles sur des données réelles (CARRICK et al., 2015), mais la simulations Uchuu ne propose actuellement pas un tel champ. Puisqu’il s’agit d’une simulations à N-corps à partir desquels des halos de matière noire sont déterminés et peuplés avec des galaxies, le champ de densité n’est pas sauvegardé et il serait trop coûteux de le déduire directement de la simulation à N-corps pour ce stage. Une solution intermédiaire est de l’approcher en suivant la méthodologie décrite dans CARRICK et al., 2015, ou encore de le déduire du champ de galaxies à l’aide de méthode d’inférences du champ comme *Bayesian Origin Reconstruction from Galaxies* (BORG, JASCHE et LAVAUX, 2019). Ces

<sup>2</sup>. à ne pas confondre avec la couleur. Comme il s’agit de l’unique utilisation de la vitesse de la lumière dans ce travail, la couleur continuera à être noté  $c$

deux solutions seront explorées dans les mois à venir, et sorte du cadre de ce stage. À la place, je me suis focalisé sur le diagnostiques du pipeline LEMAÎTRE et l'intégration de mon travail dans les outils en développement.

### 3.3 Résultats intermédiaires et finaux visés

#### 3.3.1 Résultats intermédiaires

##### Simulations avec vitesses particulières

Le premier résultat visé est de générer des SNe à partir de la simulation Uchuu contenant l'information des vitesses particulières avec `skysurvey`. Ce tirage doit reproduire la distribution en redshift des galaxies hôtes, et permettre une reconstruction des vitesses particulières à partir du diagramme de Hubble. Une fois les paramètres tirés, les courbes de lumière associées doivent être générées en utilisant les logs d'observations des relevés ZTF, SNLS et HSC.

##### Reconstruction des SNe avec SALT2.4

Le second résultat visé est de reconstruire les paramètres des SNe à partir de leurs courbes de lumière à l'aide de SALT2.4, et de caractériser cette reconstruction.

##### Reconstruction de la cosmologie avec EDRIS et extraction des vitesses particulières

Une fois les paramètres SALT2.4 reconstruits, EDRIS sera exécuté afin d'obtenir la cosmologie sous-jacente et les coefficients de standardisation  $\alpha$  et  $\beta$ . Cela permettra alors d'obtenir les vitesses particulières via les résidus au diagramme de Hubble, et de caractériser les biais éventuels.

##### Reconstruction des SNe avec Lemaître

En parallèle, le même résultat est attendu en utilisant PETS pour la sélection d'un lot d'entraînement puis l'entraînement du modèle et la reconstruction des paramètres des SNe avec NaCl. Cette deuxième reconstruction sera également injectée dans EDRIS afin d'obtenir les vitesses particulières, et caractériser les biais éventuels de NaCl.

#### 3.3.2 Résultats finaux

Les résultats finaux visés contiennent la caractérisation des deux reconstructions SALT2.4 et NaCl et leur comparaison, ainsi que l'exploitation des vitesses particulières reconstruites dans une analyse  $f\sigma_8$  jointe Uchuu x ZTF.

### 3.4 Difficultés rencontrées

La principale difficulté que j'ai rencontré a été de suivre le développement des différents modules de LEMAÎTRE tout en les utilisant et en remplissant mes tâches. Puisque l'intégration des modules n'était pas encore réalisée, j'ai pu identifier différents problèmes qui y étaient liés tels que des différences de convention ou de notation pour des variables. Mes tests ont également permis de révéler certains comportements anormaux des modules lors de leurs utilisations sur des données réalistes (par exemple des erreurs liées à des paramètres sortant des plages de définition du modèle dans NaCl), et de mettre en lumière des étapes de filtrages des données entre les modules afin d'assurer leur bon fonctionnement. Cela a nécessité de faire des points

pour discuter de l'avancement des modules en plus des points hebdomadaires prévus avec les autres membres de l'équipe, afin de remonter et résoudre les problèmes rencontrés, et être à jour sur l'utilisation des modules.

Une autre difficulté notable a été la compréhension fine du domaine et des tâches, afin de proposer une analyse correcte et robuste. J'ai parfois eu du mal à différencier certains concepts et outils, mais avec des explications et l'aide de mes encadrants et des autres chercheurs de l'équipe, j'ai pu surmonter cette difficulté.

La principale difficulté nouvelle, dont je n'avais pas forcément conscience en début de stage, est l'accès et le partage des données entre différentes collaborations. Il est naturel qu'une collaboration ait l'exclusivité de l'analyse sur les données qu'elle produit, mais cela pose de nombreux problèmes pour des analyses transverses. Au début du stage, les données DESI, y compris les simulations, étaient réservées à leur utilisation au sein de la collaboration. Si papiers de la collaboration sur les données DESI Y1 ont été publiés courant avril, les données elles-mêmes ne seront partagées qu'en 2025. Cela signifie que le catalogue BGS simulé à partir d'Uchuu n'est actuellement pas publique, puisqu'il est basé sur ces données DESI Y1. De même, les données de la DR2.5 de ZTF ne seront publiques qu'après publication des résultats cosmologiques obtenus avec le pipeline LEMAÎTRE, et ne peuvent être partagés au sein de la collaboration DESI. Des accords sont actuellement en cours de rédaction pour définir les données réelles et simulées partagées entre les deux collaborations et les études que chaque groupe pourra mener dessus. Mon travail portant sur des simulations, il a été moins contraint que s'il avait porté sur des données réelles, mais il a tout de même fallu que je m'assure de ne pas partager de données critiques dans mes interactions avec les autres chercheurs.

### 3.5 Gestion des risques réalisée

Les analyses cosmologiques souffrent principalement de deux risques : la disponibilité des données, et le temps nécessaires à leur traitement.

La disponibilité des données frappe particulièrement les travaux préparant des analyses sur des données réelles. Une panne du télescope, telle que la rupture d'une pièce, peut considérablement ralentir la prise de données et retarder les analyses qui en dépendent. Dans mon cas précis, je n'ai pas eu à gérer un tel risque pour plusieurs raisons. La première est que j'ai travaillé majoritairement sur des simulations, et que les rares paramètres issus de données réels (*i.e.* les logs d'observations de ZTF et SNLS, et les contraintes de la simulation Uchuu) utilisés étaient déjà acquis bien avant le début du stage. La prise de données pour la DR2.5 de ZTF a été finalisée début 2021, celle pour DESI Y1 en juin 2022.

En revanche rien n'assure qu'un problème sur les données n'ait pas eu lieu plus en amont de la chaîne. Entre un capteur CCD de ZTF et une valeur de flux, utilisée par le pipeline LEMAÎTRE, de nombreuses étapes de calibration, traitement, et corrections ont lieu. Des effets d'accumulations de charges dans les CCD détériorant les images ont ainsi été découverts après la DR2 de ZTF, ce qui a motivé la production de la DR2.5. De même, une partie du pipeline LEMAÎTRE rapidement évoquée dans ce rapport mais absolument cruciale est l'étape de calibration. Il faut assurer que des télescopes différents, avec des optiques et des capteurs différents, puissent être accordés les uns avec les autres pour que les observations soient consistantes. Pour le pipeline LEMAÎTRE, par exemple, cela s'est traduit par un soin particulier apporté à la détermination des fonctions de transmissions des filtres, notamment ceux de SNLS, qui constituera une publication à part entière.

Ces risques concernant les données étaient hors de ma portée, mais ont eu pour effet un suivi assidu du travail réalisé pour corriger ces problèmes, par exemple en utilisant les données les plus

récentes disponibles pour la transmission des filtres.

Un autre risque majeur en cosmologie est le temps de calcul nécessaire au traitement et à l'analyse des données. Si mes simulations ont été effectuées au NERSC, centre de calcul absolument titanesque, l'objectif de mon travail est que les scripts utilisés puissent tourner sur un ordinateur personnel. C'est également un des prérequis du pipeline LEMAÎTRE, puisque le but principal est de disposer d'un pipeline d'analyse en intégration continue dans lequel des biais variés peuvent être introduits, et leur effets sur la reconstruction des paramètres cosmologiques rapidement quantifiés. Cette philosophie a conduit à de nombreuses optimisations pour accélérer le traitement, tels que l'utilisation de la parallélisation du traitement des SNe par **PETS** ou encore l'utilisation de **JAX** pour **EDRIS**. Une ré-implémentation de **NaCl** en **JAX** est également prévue, afin de réduire d'avantage le temps nécessaire à l'entraînement du modèle.



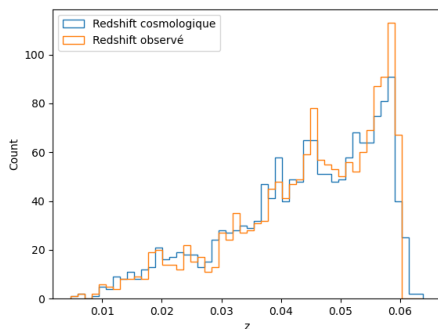
# Chapitre 4

## Résultats

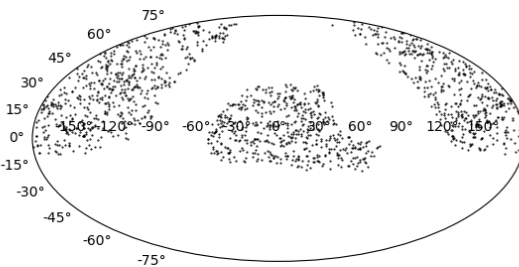
### 4.1 Validation du pipeline LEMAITRE sur simulations

#### 4.1.1 Génération de SNe à partir d’Uchuu

La distribution spatiale des SNe tirées est présentée en Fig. 4.1a et 4.1b. La distribution en redshift observés (incluant les vitesses particulières) diffère bien des redshifts cosmologiques, et la distribution angulaire reproduit la footprint DESI Y5, qui est similaire à la footprint ZTF.



(a) Distribution des redshift tirés à partir du catalogue de galaxies Uchuu. Le redshift cosmologique correspond au redshift de la position de la galaxie, le redshift observé inclue l'effet Doppler des vitesses particulières.



(b) Distribution angulaire des SNe tirées à partir du catalogue de galaxies Uchuu.

FIGURE 4.1 – Distribution spatiale des SNe tirées

De plus il est déjà possible de quantifier l'effet de la distribution intrinsèque sur la reconstruction des vitesses particulières. En effet, une erreur commise sur les modules de distances impact directement le redshift cosmologiques reconstruit, et donc les vitesses particulières (cf. Fig 4.2). Cette erreur est corrélée avec le redshift, et devient plus importante à haut redshift car  $\mu(z)$  y croît plus lentement.

La distribution des vitesses particulières ainsi reconstruites est représentée en Fig. 4.3. Les vitesses particulières sont reconstruites avec une erreur moyenne de 18.4km/s et une dispersion



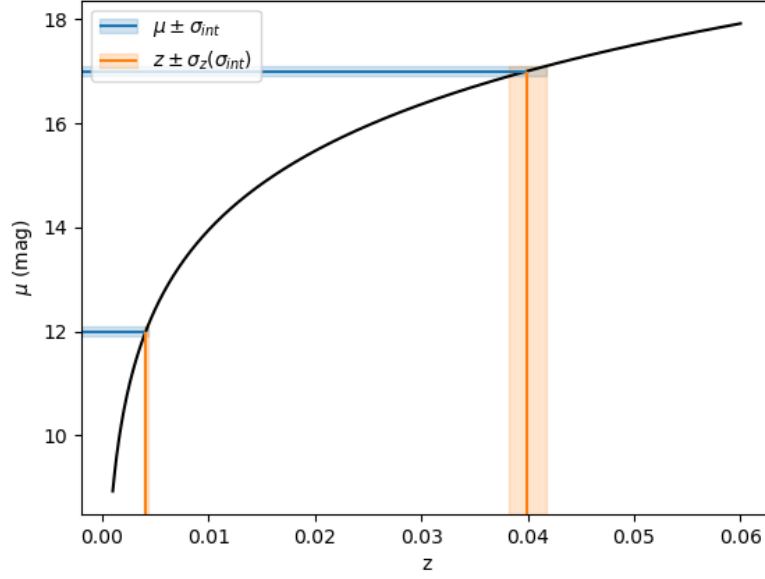


FIGURE 4.2 – Illustration de l’erreur commise à  $1\sigma$  sur les redshifts cosmologiques reconstruits due à la dispersion intrinsèque.

de 596.8km/s. On observe bien une augmentation de l’erreur commise sur les vitesses particulières reconstruites, et cette augmentation suit l’erreur théorique due à la dispersion intrinsèque. De plus, si on calcule la moyenne des erreurs théoriques due à la dispersion intrinsèque sur l’échantillon complet, on obtient  $\sigma_{v_{pec}}(\sigma_{int}) = 592.4\text{km/s}$ , ce qui est proche de la dispersion observée.

Une fois les paramètres des SNe tirés, les observations photométriques et spectroscopiques sont générées sur la base du modèle **SALT2.4** et des logs d’observation de ZTF. De telles courbes de lumière et spectres sont représentés en Figure ??.

#### 4.1.2 Filtrage par PETS

Le filtrage effectué par PETS permet d’enlever les SNe qui sont mal échantillonnées. Le nombre de SNe retirées à chaque étape de sélection est indiqué en Table 4.1, et suivent la démarche décrite en Annexe A.1.1.

Des cas de figure typiques sont représentés en Figures 4.5, 4.6 et 4.7.

La première représente le cas d’une SNe bien échantillonnée : le minimum de la fonction  $\chi^2(t_{max})$  est unique, bien défini, et présente une incertitude de 0.012 jours, ce qui est extrêmement faible. Cette précision est telle que les intervalles à  $1\sigma$  et  $3\sigma$  sont indiscernables du trait plein du  $t_{max}$  déterminé, qui correspond parfaitement avec celui du fit initial. Les paramètres  $x_1$  et  $c$  ont également des valeurs raisonnables, la SN est donc conservée.

Le second cas de figure (Figure 4.6), est celui d’une SNe mal échantillonnée : le minimum de la fonction  $\chi^2(t_{max})$  est toujours unique, mais mal défini car le  $\chi^2$  varie peu autour du minimum, ce qui a pour conséquence une incertitude de  $^{+0.764}_{-0.828}$  jours et une asymétrie du  $\chi^2$ . Cette aplatissement du  $\chi^2$  autour du minimum a pour conséquence majeure d’entraîner de fortes variations

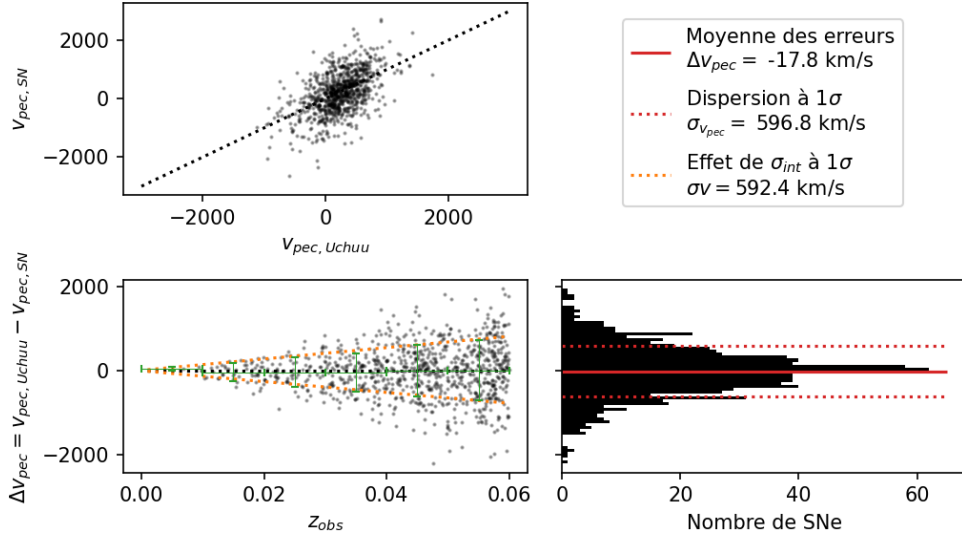


FIGURE 4.3 – Distribution des vitesses particulières reconstruites directement à partir des SNe tirées à partir du catalogue de galaxies Uchuu. La ligne en pointillés noire (graphiques de gauche) représente la ligne de base  $v_{pec,SN} = v_{pec,Uchuu}$ , les lignes oranges (en bas à gauche) représente l’erreur théorique à  $1\sigma$  due à la dispersion intrinsèque, et les lignes rouges pleine et en pointillés (en bas à droite) représentent respectivement la moyenne et la déviation standard des erreurs. Les barres d’erreurs en vert (en bas à gauche) montre l’évolution de la moyenne et de la dispersion des erreurs en fonction du redshift.

des fonctions  $x_1(t_{max})$  et  $c(t_{max})$ , notamment vers  $t_{max} = 58810$  où les deux paramètres sortent des plages acceptables. Ce comportement est à éviter pour l’entraînement de NaCl, car l’ajustement des paramètres du modèle conjointement à ceux des SNe a plus de chance de se bloquer et d’obtenir des paramètres de standardisation anormalement élevés. Cette SN est donc rejetée.

Le troisième cas de figure concerne un cas pathologique, qui a été diagnostiqué lors de ce stage et est détaillé en Section 4.2.1. En effet, la SNe ne présente des points de photométrie que sur la fin de sa décroissance, ce qui a pour effet de focaliser la première estimation de  $t_{max}$  autour de ces points. PETS ignorant alors les minimums aux extrémités de l’intervalle de  $t_{max}$  utilisés, cette SNe ne présentait qu’un minimum bien défini, avec une incertitude faible de  $\pm 0.111$  jours, et était donc satisfaisante. Ce comportement a donc été modifié pour prendre en compte les minimums aux extrémités, et la détection de minimums potentiellement problématiques remontée à  $\Delta\chi^2 = 64$  (équivalent à une détection à  $8\sigma$ ) plutôt que  $\Delta\chi^2 = 9$  (respectivement  $3\sigma$ ).

#### 4.1.3 Reconstruction avec SALT2.4

La première reconstruction des paramètres de standardisation possible une fois les courbes de lumières simulées et traitées avec PETS est obtenue avec le modèle SALT2.4. Les résidus sur les paramètres reconstruits sont présentés en Figures 4.8 et 4.9

Sans grande surprise, les paramètres de standardisation sont bien reconstruits puisque seules des SNe bien échantillonnées sont utilisées. On peut toutefois noter le biais qui survient pour les grandes amplitudes (cf Figure 4.8a), et les incertitudes étonnamment élevées pour certains stretch (cf Figure 4.8b) et couleurs (cf Figure 4.8c), alors que les résidus des valeurs reconstruites

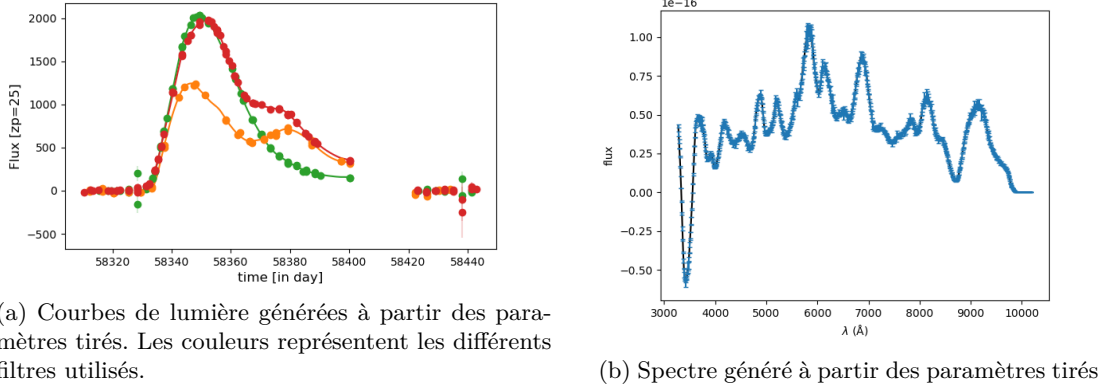


FIGURE 4.4 – Exemple d’observations simulées pour les échantillons de SNe

Cut	Discarded	Remaining
Total	-	1591
<b>SALT2.4 converge</b>	442	1149
$\sigma_{t_{max}} < 1$	31	1118
$\chi^2(t_{max})$ est symétrique autour de $t_{max}$	77	1041
Pas d’autres minimums à $8\sigma$	9	1032
$ x_1  < 4$	0	1032
$ c  < 2$	0	1032

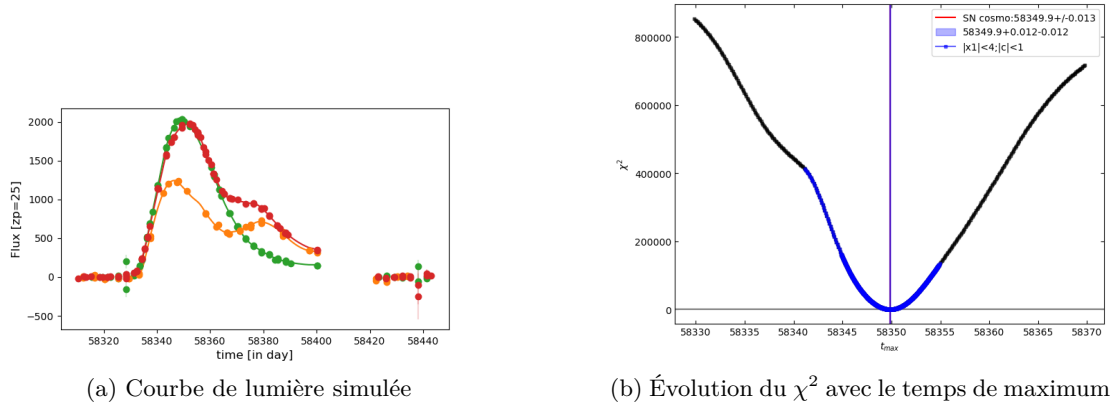
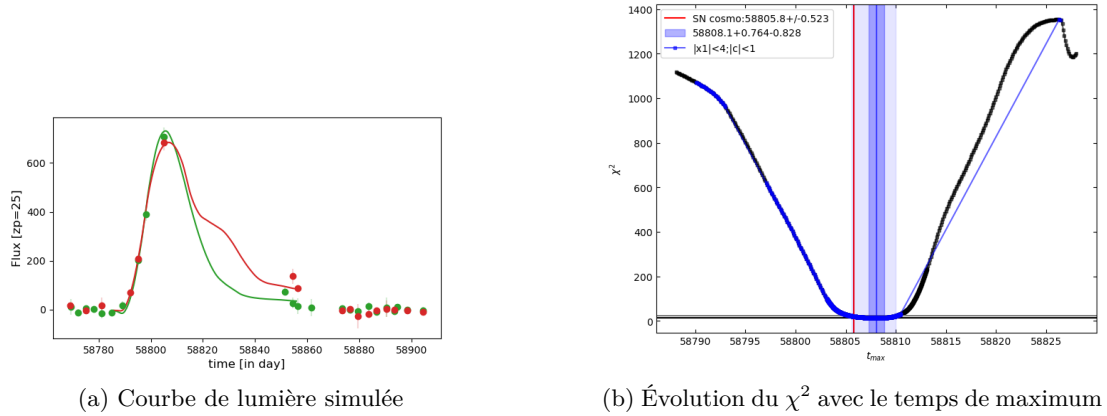
TABLE 4.1 – SNe rejetées par PETS à chaque étape de sélection pour le lot ZTF simulé

sont faibles. Cela est du en grande partie au fait que le modèle d’erreur a été ajusté à l’aides de simulations après entraînement pour **SALT2.4**. Pour la reconstruction des temps de maximum (cf Figure 4.8d), on peut clairement voir la distinction temporelle entre SNLS et ZTF, et il n’est pas surprenant que les résidus de SNLS soient plus importants. En effet, SNLS observe des SNe plus lointaine, donc les flux reçus sont bien plus faibles que ceux de ZTF. Cela a pour effet d’avoir des incertitudes relatives sur les flux plus importantes, et le temps de maximum est moins bien contraint.

Les modules de distances sont très biens reconstruits, l’erreur moyenne commise n’est que de l’ordre de  $10^{-2}$  mag. De plus comme seules les SNe de ZTF seront utilisées pour le calcul des vitesses particulières, les incertitudes élevées des SNe SNLS (moins lumineuses, donc ayant une magnitude plus élevées) ne sont pas gênantes.

#### 4.1.4 Analyse cosmologique sur les résultats **SALT2.4**

Les vitesses particulières obtenues sont représentées en Fig. 4.10. Les vitesses reconstruites sont en bon accord avec les vitesses injectées, mais présentent un léger biais. L’erreur moyenne n’est que de  $270\text{km.s}^{-1}$  et l’écart type de l’erreur commise est de  $791\text{km.s}^{-1}$ , ce qui est plus élevé de 36% que l’effet attendu de la dispersion intrinsèque, à  $581\text{km.s}^{-1}$ . On peut également noter le fait que le biais sur la moyenne des erreurs évolue avec le redshift, qui peut être à l’origine d’une part du biais non nul observé sur l’ensemble de l’échantillon, et d’autre part de l’élargissement de la dispersion. Un tel effet est intéressant puisqu’il pourrait s’agir d’un effet de flux cohérent, c’est-à-dire que les galaxies hôtes présentent toutes une même composante de vitesse par rapport

FIGURE 4.5 – SNe validée par PETS. Le minimum du  $\chi^2$  est unique et bien défini.FIGURE 4.6 – SNe rejetée par PETS. Le minimum du  $\chi^2$  est unique mais mal défini.

à l'observateur liée à la formation d'une structure commune.

Ce biais peut également être causé par une mauvaise estimation des paramètres du modèle par EDRIS : un des principal défaut des analyses cosmologiques basées sur les SNe Ia est qu'elles n'imposent pas de contraintes sur la constante de Hubble  $H_0$  car elle intervient dans le calcul de la magnitude selon le modèle sous la forme d'un terme additif. Cela signifie qu'une variation du à la calibration des flux ou de la magnitude de référence en bande B  $m_B$  peut être à tort interprétée comme une modification de  $H_0$ . Pour les vitesses particulières, un tel effet se traduirait par un déplacement vertical artificiel des modules de distance, et introduirait donc un biais sur l'ensemble des modules de distance reconstruits par EDRIS.

#### 4.1.5 Reconstruction et vitesses particulières avec NaCl

En exécutant la même chaîne avec le résultat de NaCl, on obtient les résultats suivants.

Les résultats de l'entraînement de NaCl sont présentés en Figure 4.11 pour les paramètres des SNe, en Figure 4.13 pour les modèles, et en Figure 4.14 pour les modules de distances.

Les paramètres reconstruits présentent plusieurs biais, mais qui ne sont pas problématiques. En effet, le biais sur les amplitudes  $X_0$  n'apparaît que pour les grandes amplitudes, qui ne

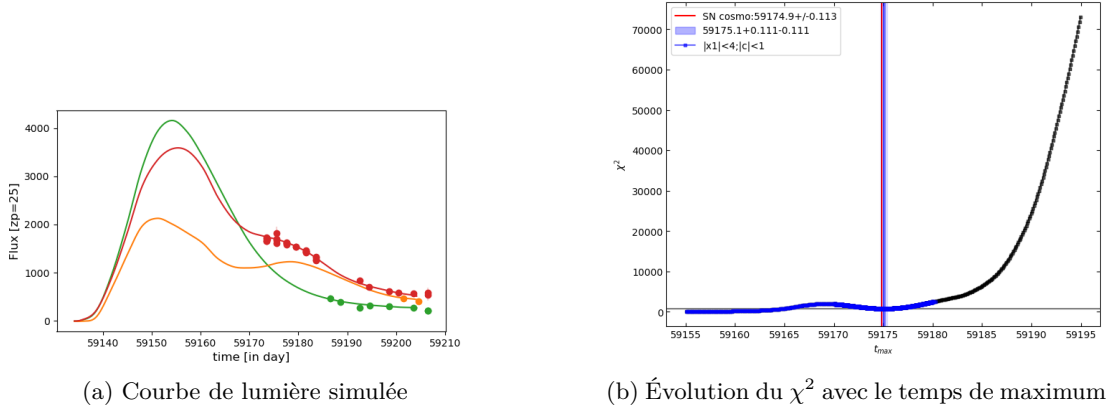


FIGURE 4.7 – SNe anormale validée par PETS avant correction. Le minimum du  $\chi^2$  trouvé initialement est un minimum local, et l’intervalle exploré ne couvre pas la vraie valeur.

concernent que les SNe à haut redshift de SNLS et dont on ne peut pas mesurer la vitesse particulière à cause de la dispersion intrinsèque. Les biais sur le stretch  $X_1$  et la couleur  $C$ , sont principalement dus à la renormalisation des modèles  $M_0^{\text{NaCl}}(p, \lambda)$  et  $M_1^{\text{NaCl}}(p, \lambda)$ . De même, le biais d’environ 0.6 jours sur le temps de maximum, est causé par le ré-entraînement de  $M_0^{\text{NaCl}}(p, \lambda)$ .

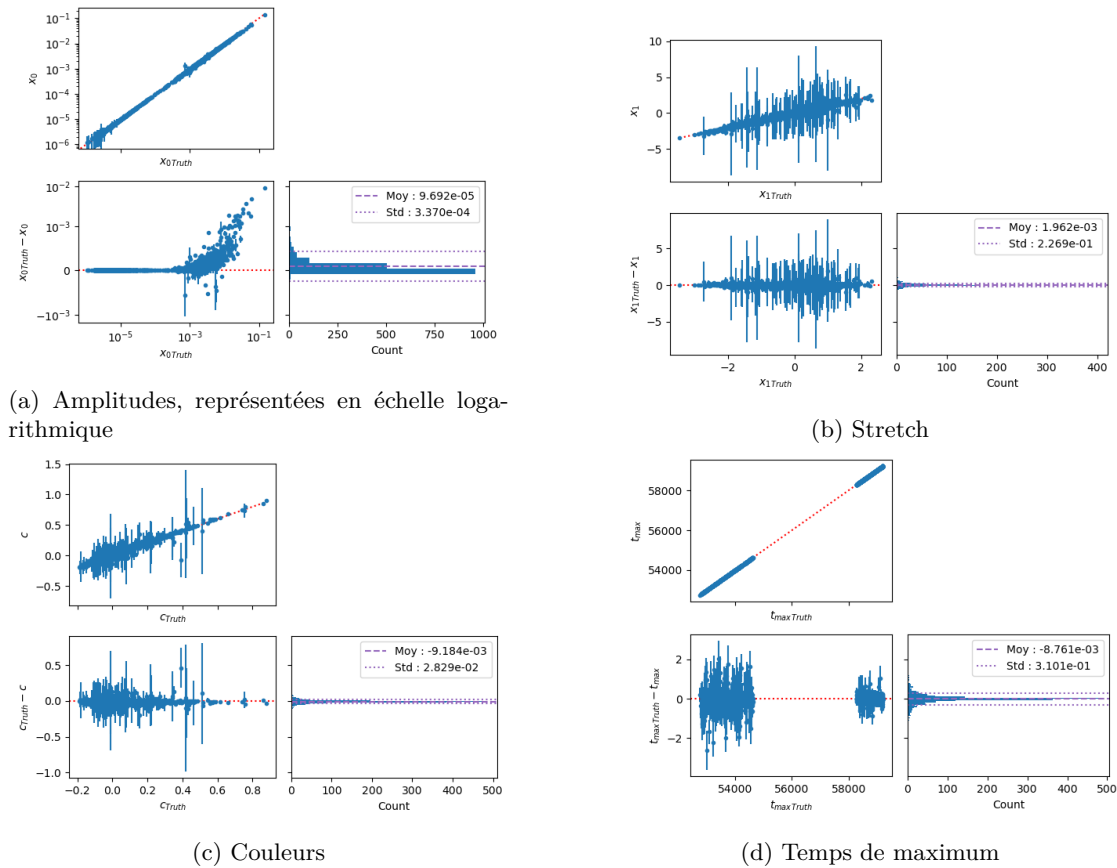
Les biais présents ne gênent en rien la modélisation des courbes de lumière ni des spectres (cf Figure 4.12) et sont purement dus au ré-entraînement des modèles. Les incertitudes rapportées ne sont pas non plus aberrantes comparées à l’amplitude des paramètres.

Les modèles ré-entraînés semblent diverger et fluctuent anormalement aux bornes de leurs ensembles de définition (cf Figure 4.13a, 4.13b), mais cela n’est pas un problème. En effet, ce sont des zones sur lesquels le modèle SALT2.4 initial était défini, mais qui n’était pas contraintes par les données utilisées pour l’entraînement (cf. Figure 4.13d). Le modèle ajusté n’est donc pas fiable dans ces régions. Ces fortes fluctuations sont toutefois surprenantes, d’une part car la contrainte de régularisation (voir Annexe A.1.2) vise précisément à empêcher ces fluctuations, d’autre part car c’est habituellement la marque d’un sur-entraînement, pour lequel le modèle a trop de paramètres libres et manque de données pour les contraindre. En revanche, les modèles SALT2.4 et NaCl s’accordent dans les régions contraintes par les données.

Les modules de distances reconstruits par NaCl présentent un biais d’environ  $-0.2$  mag (cf Figure 4.14), mais ce biais n’est pas gênant pour la suite de l’analyse. En effet, comme mentionné précédemment, les magnitudes ne sont définies que relativement les unes aux autres.

Les vitesses particulières obtenues à l’aide d’EDRIS sont représentées en Fig. 4.15. Les vitesses reconstruites sont en bon accord avec les vitesses injectées, mais présentent un léger biais. L’erreur moyenne n’est que de  $265 \text{ km.s}^{-1}$  et l’écart type de l’erreur commise est de  $686 \text{ km.s}^{-1}$ , ce qui correspond à une augmentation de la dispersion de 18% par rapport à l’effet attendu de la dispersion intrinsèque.

Le biais entre les vitesses reconstruites par SALT2.4 et NaCl étant semblables, son origine ne provient pas de l’algorithme utilisé pour la reconstruction des paramètres de standardisation. En revanche le ré-entraînement du modèle pour NaCl a permis une meilleure estimation des vitesses particulières, puisque la dispersion additionnelle sur l’effet de la dispersion intrinsèque a été réduit de 38% à 18%.

FIGURE 4.8 – Résidus des paramètres ( $x_0, x_1, c, t_{max}$  reconstruits par SALT2.4

## 4.2 Problèmes identifiés dans le pipeline Lemaître et corrections apportées

J'ai identifié et traité différents problèmes et erreurs au sein du pipeline LEMAÎTRE. Certains étaient triviaux à traiter une fois identifiés, tel que des erreurs dues à des différences d'encodage (sur 32 ou 64 bits) de certaines valeurs lors de sauvegardes et chargements de données entre des modules, d'autres plus compliqués, et quelques uns restent encore non résolus. Les problèmes rencontrés peuvent être regroupés en deux catégories : ceux causés par l'utilisation de mauvaises valeurs de méta-paramètres, qui ne remettent pas en cause les méthodes mais nécessitent un ajustement de ces paramètres, et ceux liés directement aux méthodes utilisées.

### 4.2.1 Problèmes liés à des valeurs de méta-paramètres

#### Minimums multiples dans PETS

Un premier problème est apparu lors de l'utilisation de PETS concerne la caractérisation des SNe pour l'entraînement de NaCl. Certaines SNe ne présentent des points de mesure que sur la croissance ou la fin de la décroissance de leur luminosité pour des raisons observationnelles

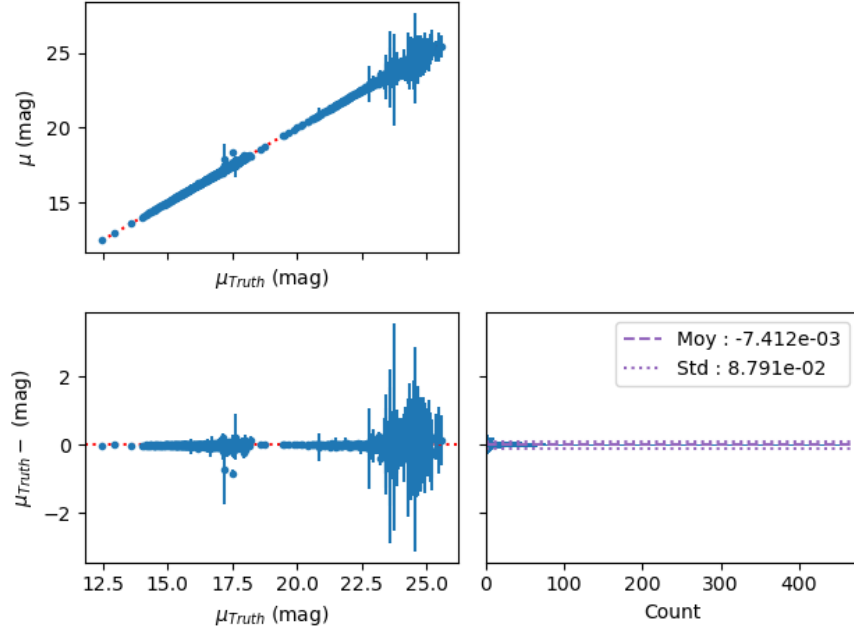


FIGURE 4.9 – Résidus des modules de distances  $\mu = -2.5 \log_{10} x_0 - 0.14 \times x_1 + 3.15 \times c$  reconstruits par SALT2.4.

tel que leur concordance avec le début ou la fin d’une période d’observation. Ces SNe posent alors le problème d’être difficiles à caractériser, bien que parfaitement décrites par SALT2.4. En particulier, les SNe qui n’ont des points de photométrie que sur la fin de leur décroissance présentent souvent deux minimums locaux (voir Figure 4.7). Comme PETS utilise le point de données de flux le plus élevé comme point de départ pour ajuster le  $t_{\text{max},0}$  à partir duquel la grille est générée, il peut converger vers un minimum local bien marqué et générer une grille ne couvrant pas le minimum global au vrai temps de maximum.

La solution employée ici a été de remonter la détection des autres minimums de  $3\sigma$  à  $8\sigma$ , et de retirer les SNe ainsi identifiées comme problématiques du lot de données à analyser. Une meilleure solution, implémentée pour les utilisations futures de PETS, est d’augmenter l’intervalle couvert par la grille, mais cela nécessite en contrepartie soit de diminuer sa résolution et donc l’estimation des incertitudes, soit d’utiliser plus de points, ce qui augmente le temps de calcul. À noter que si ces SNe sont bien décrites par SALT2.4, elles ne sont pas intéressantes pour entraîner NaCl puisqu’elles ne comportent que peu de points de données et ne couvrent qu’une faible fraction des phases. Elles pourraient être fittées après entraînement afin de gagner en statistique, mais le gain restera minime puisqu’elles demeurent assez rares (environ 1% des SNe respectant par ailleurs les autres critères de PETS).

De plus c’est un problème qui est particulièrement présent dans mes simulations, mais devrait être rare lors de l’utilisation de données réelles. En effet, puisque seules des SNe Ia sont générées, elles sont reconstruites en partant du principe que ce sont des SNe Ia. Sur des données réelles, rien ne garantit qu’un événement transitoire dont on a observé uniquement la décroissance soit une SNe Ia. Il pourrait s’agir d’une SNe Ib, Ic, II, ou d’un autre type d’événement cataclysmique. Dans ce cas, l’événement ne sera pas classé avec certitude comme SN Ia, et ne passera pas donc par PETS.

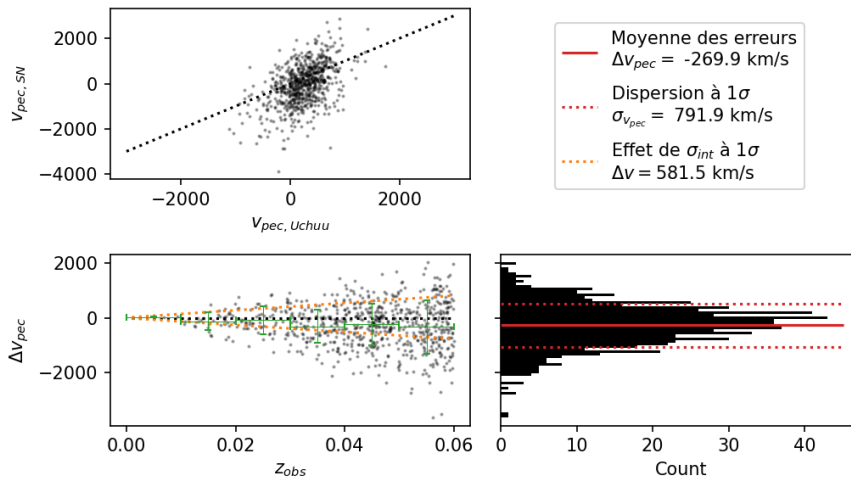


FIGURE 4.10 – Distribution des vitesses particulières reconstruites avec EDRIS à partir de SALT2.4 ( $v_{pec,SN}$ ). La ligne en pointillés noire (graphiques de gauche) représente la ligne de base  $v_{pec,SN} = v_{pec,Uchuu}$ , les lignes oranges (en bas à gauche) représente l’erreur théorique à  $1\sigma$  due à la dispersion intrinsèque, et les lignes rouges pleine et en pointillés (en bas à droite) représentent respectivement la moyenne et la déviation standard des erreurs. Les barres d’erreurs en vert (en bas à gauche) montre un biais de la moyenne des erreurs en fonction du redshift, il peut s’agir d’un effet de flux cohérent.

### Régularisation et contraintes dans NaCl

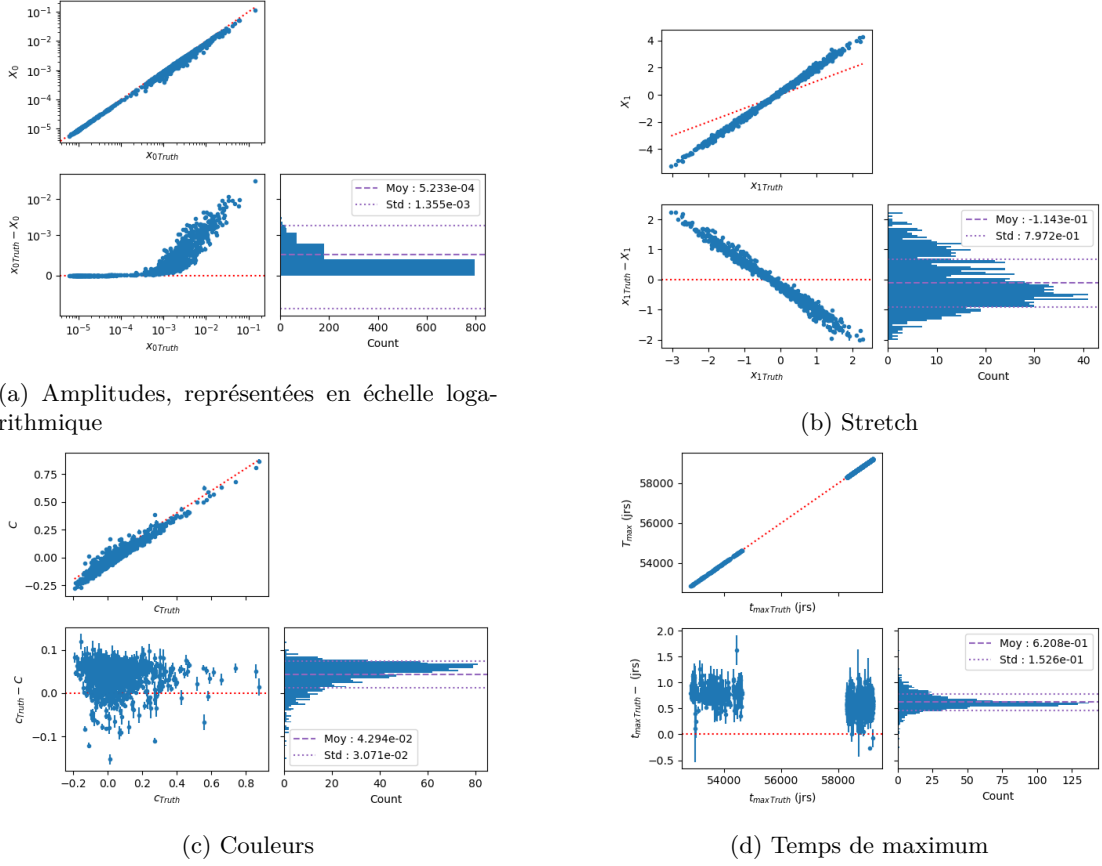
Un autre problème identifié dans NaCl concerne les coefficients de régularisation  $\mu_{reg}$  et de contraintes  $\mu_{cons}$ . Les premiers tests effectués utilisaient  $\mu_{reg} = 1$  et  $\mu_{cons} = 10^{-10}$ , mais ces valeurs étaient trop faibles et ne permettaient pas d’imposer aux modèles les contraintes et la régularisation voulus. Une des conséquences majeures étaient une mauvaise convergence, et une surestimation des incertitudes sur les paramètres reconstruits. Il s’est avéré que prendre des valeurs  $\mu_{reg} = 10^2$  et  $\mu_{cons} = 10^6$  donnent une bien meilleure convergence, et ne surestiment plus les incertitudes.

### Domaine de définition des modèles

Un deuxième problème qui est survenu vient du domaine de définition des modèles dans NaCl. Initialement, les modèles couvraient l’intervalle  $[-20; 50]$  en phase et  $[2000; 9000]$  Ångströms en longueur d’onde. Or, pour le calcul des flux des points de photométrie, il arrive que le support de la transmission du filtre sorte du domaine de définition en longueur d’onde des modèles. Cela revient à couper les filtres et fausse le flux calculé, ce qui en retour biaise la reconstruction des paramètres de standardisation puisque le modèle s’adapte pour compenser cette coupure.

La solution a été simplement d’augmenter le domaine de définition des modèles, de 2 000 à 11 000 Ångströms. Aux redshifts considérés, les filtres ne sortent plus du modèle et l’évaluation redevient correcte.



FIGURE 4.11 – Résidus des paramètres ( $X_0$ ,  $X_1$ ,  $C$ ,  $t_{max}$  reconstruits par NaCl

### Quality cuts pour EDRIS

Dès les premières tentatives de production de diagrammes de Hubble à partir de la reconstruction avec SALT2.4, la convergence d'EDRIS est apparue comme très dépendante des données. En particulier, la présence d'outliers est susceptible de coincer le fit des paramètres et de la cosmologie. Les outliers peuvent apparaître soit lors de la génération (*i.e.* les paramètres sont extrêmes mais probables au vu des distributions), soit lors de la reconstruction (*i.e.* la SNe n'était pas aberrante, mais elle est mal reconstruite). L'amplitude des coefficients de la matrice de covariance des paramètres ( $x_0, x_1, c$ ) a un impact particulièrement important sur la convergence. C'est l'une des raisons qui a motivée l'exploration et la résolution du problème précédent ; une mauvaise convergence de NaCl donnait un point de départ aberrant pour la matrice de covariance pour EDRIS.

Cela a aussi conduit à appliquer des quality cuts semblables à ceux d'Union 3 aux paramètres reconstruits afin de retirer d'éventuels outliers ou SNe mal reconstruites.

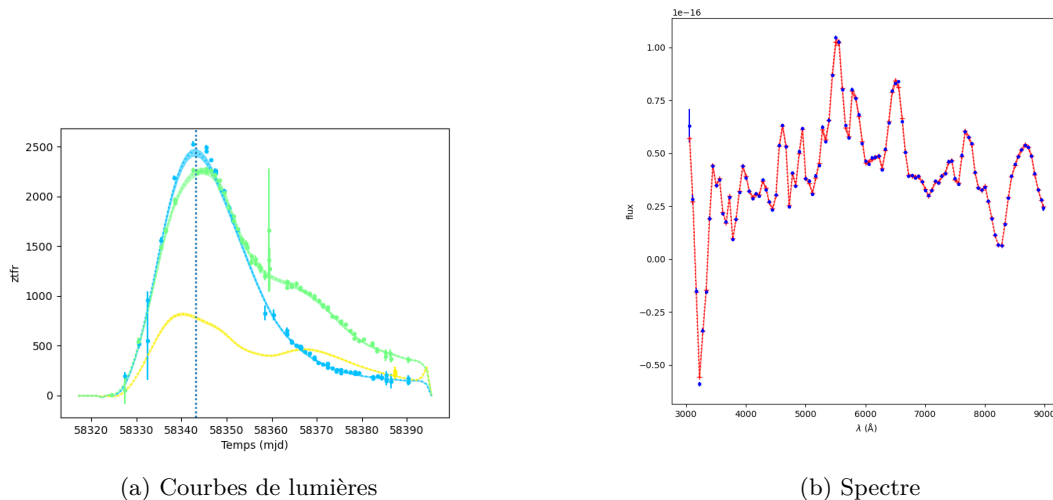


FIGURE 4.12 – Exemple de courbes de lumières et de spectres avec le modèle ajusté par NaCl

### 4.2.2 Problèmes liés aux méthodes

#### Effets d'une mauvaise convergence

J'ai évoqué comme conséquences des problèmes précédents une mauvaise convergence des modules NaCl ou EDRIS. Dans les faits, cette mauvaise convergence est caractérisée par plusieurs symptômes, le plus problématique concernant la hessienne du problème  $\mathcal{H}$ .

Comme il serait trop coûteux d'estimer les incertitudes tout au long de l'optimisation, la solution retenue pour le pipeline LEMAÎTRE est d'utiliser la borne de Cramer-Rao comme estimateur de la covariance. À un coefficient multiplicatif près dépendant de la formulation exacte du problème, la covariance est obtenue en inversant la matrice d'information de Fisher. L'information de Fisher étant dans le cas d'une bonne convergence égale à la Hessienne du problème, on obtient

$$Cov = \mathcal{I}^{-1} = \mathcal{H}^{-1} \quad (4.1)$$

et les incertitudes sont directement obtenues en prenant les racines carrées des coefficients diagonaux

$$\sigma_i = \sqrt{Cov_{ii}} = \sqrt{(\mathcal{H}^{-1})_{ii}} \quad (4.2)$$

Le problème qui survient alors est qu'en cas d'absence de convergence ou de mauvaise convergence, la hessienne peut ne pas être définie positive, ce qui rend les incertitudes indéterminées. Ce problème est survenu à de nombreuses reprises, bien souvent du à un mauvais point de départ ou des outliers.

Dans le cas de NaCl, cela a mis en lumière la nécessité d'avoir une bonne convergence dans la première estimation des paramètres des SNe. En cas de mauvaise estimation, le modèle essaie de surcompenser cette mauvaise estimation pour décrire les données lors du dernier fit. Comme il n'arrive pas à converger correctement, il se bloque généralement dans une région de l'espace des paramètres où la hessienne n'est pas définie positive, ce qui a permis d'identifier ce type de problèmes.

### Dégénérescence entre $\beta$ et $\sigma_{int}$

Il existe une dégénérescence entre le coefficient  $\beta$  de la formule de Tripp (eq 2.1) et le paramètre de dispersion intrinsèque  $\sigma_{int}$  dans EDRIS, biais qui est particulièrement marqué lorsque l'incertitude sur la couleur  $\sigma_c$  est élevée. Ce problème était déjà connu, mais a été confirmé sur les simulations produites lors de ce stage lorsque la reconstruction est effectuée avec SALT2.4. Ce n'est pas le cas pour la reconstruction avec NaCl car les erreurs sont sous-estimées pour une raison encore inconnue, mais ce problème pourrait également survenir à l'avenir. La solution retenue est alors de fixer la dispersion intrinsèque à sa valeur utilisée pour les simulations,  $\sigma_{int} = 0.1$ , puisqu'il s'agit de la dispersion observée expérimentalement.

### Condition d'arrêt du gradient conjugué dans EDRIS

Lors de l'utilisation d'EDRIS, j'ai été amené à me documenter sur le détail de la méthode utilisée. En particulier, l'optimisation utilisée est décrite dans MARTENS, 2010, et mentionne explicitement que pour un problème linéaire  $Ax = b$ , les méthodes de gradient conjugué ne minimisent pas l'erreur quadratique  $\|Ax - b\|^2$  mais la forme quadratique  $\phi(x) = \frac{1}{2}x^T Ax - b^T x$ . Bien que ces deux formulations présentent le même minimum global, une solution acceptable mais sous-optimale pour l'un peut s'avérer être mauvaise pour l'autre. En particulier lors d'une exécution de gradient conjugué, si  $\phi(x)$  est bien décroissante,  $\|Ax - b\|^2$  fluctue violemment et ne commence à tendre vers 0 qu'à la fin de l'optimisation. Pourtant, les conditions d'arrêt des implémentations de gradient conjugué se basent souvent sur  $\|Ax - b\|^2$  et non  $\phi(x)$ , potentiellement car cette dernière n'est pas bornée par 0. L'auteur propose alors de modifier la condition d'arrêt pour la conditionner au taux de réduction de  $\phi$ , moyenné sur  $k$  itérations

$$i > k \quad \text{et} \quad \phi(x_i) < 0 \quad \text{et} \quad \frac{\phi(x_i) - \phi(x_{i-k})}{\phi(x_i)} < k\epsilon \quad (4.3)$$

Le problème pour EDRIS vient alors de son implémentation. Afin d'accélérer l'évaluation des log-vraisemblance et de leurs dérivées, EDRIS a été intégralement implémenté en JAX<sup>1</sup>. Ce package propose une ré-implémentation des packages `python` usuels pour qu'ils soient compatibles avec l'optimisation proposée par JAX, en particulier le package `scipy` et son implémentation du gradient conjugué. Cependant, son implémentation initiale et la ré-implémentation en JAX utilisent toutes les deux la conditions d'arrêt déconseillée par MARTENS, 2010, ce qui peut avoir un impact non négligeable sur la convergence d'EDRIS.

Ce problème demeure à ce jour non résolu d'une part car son effet, s'il est bien présent, n'était pas la cause majeure d'une mauvaise convergence (voir notamment 4.2.1), et d'autre part car sa résolution est compliquée puisqu'elle nécessite d'ajouter une condition d'arrêt non déterministe à la fonction de gradient conjugué, ce dont JAX ne s'accommode pas trivialement.

---

1. <https://jax.readthedocs.io/en/latest/>

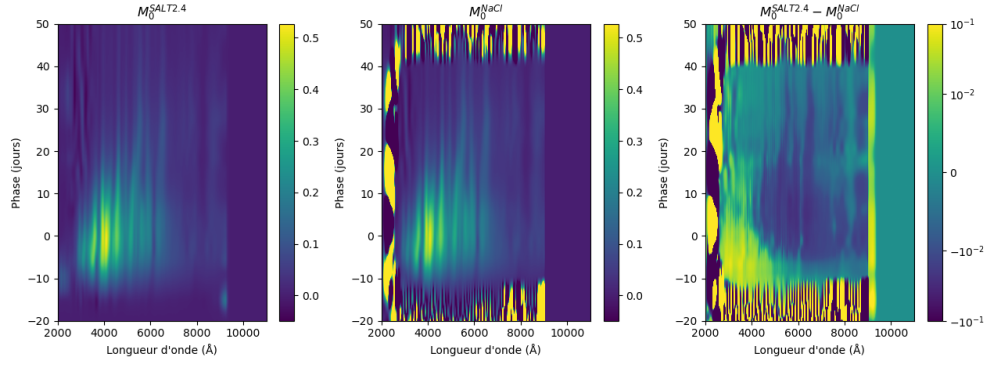
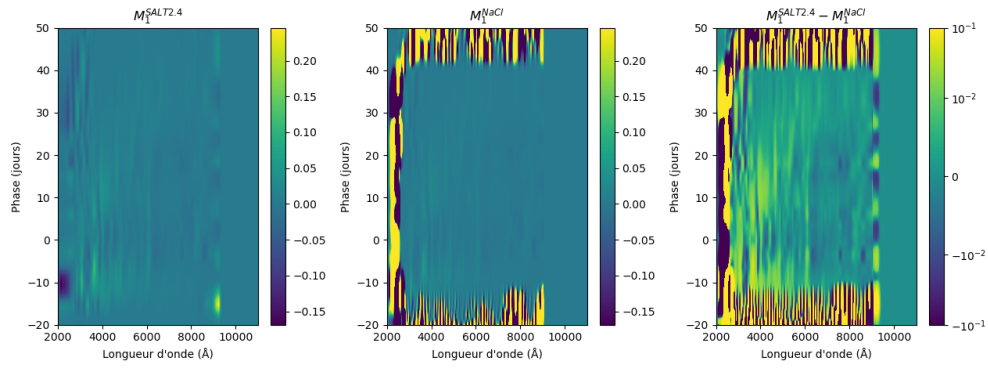
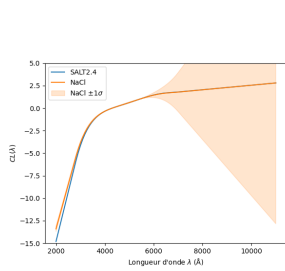
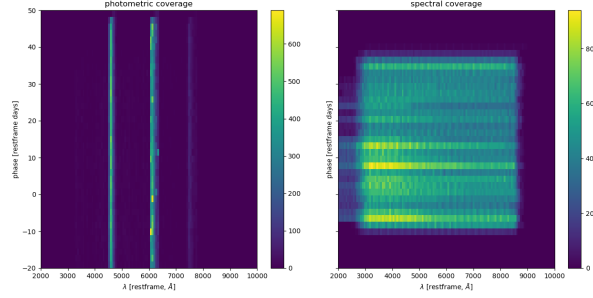

 (a) Modèle  $M_0^{\text{NaCl}}(p, \lambda)$ 

 (b) Modèle  $M_1^{\text{NaCl}}(p, \lambda)$ 

 (c) Loi de couleur  $CL(\lambda)$ 

 (d) Couverture du modèle par les points de photométrie (*i.e.* des courbes de lumières) et les points de spectroscopie

FIGURE 4.13 – Modèles SALT2.4 et modèles NaCl entraînés. Les zones de fortes fluctuations des modèles entraînés, autour des bornes de la surface, est du au manque de points de données contraignant le modèle

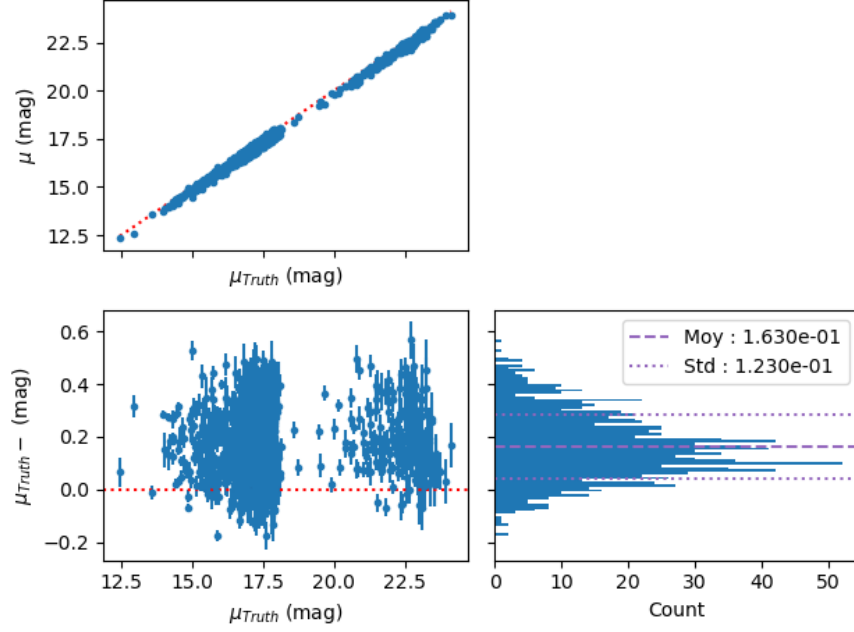


FIGURE 4.14 – Résidus des modules de distances  $\mu = -2.5 \log_{10} X_0 - 0.14 \times X_1 + 3.15 \times C$  reconstruits par NaCl1. Ils présentent un biais d'environ  $-0.2$  mag, qui est causé par le ré-entraînement des modèles, mais n'est pas gênant pour l'analyse cosmologique.

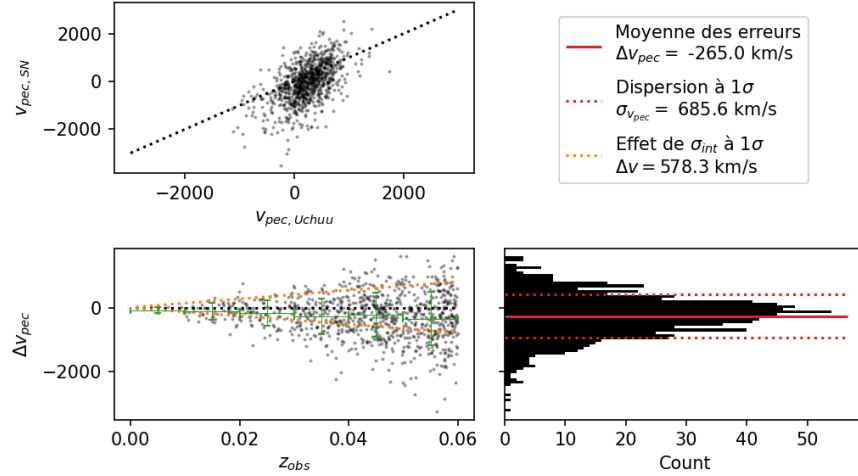


FIGURE 4.15 – Distribution des vitesses particulières reconstruites avec EDRIS à partir de SALT2.4 ( $v_{pec,SN}$ ). La ligne en pointillés noire (graphiques de gauche) représente la ligne de base  $v_{pec,SN} = v_{pec,Uchuu}$ , les lignes oranges (en bas à gauche) représente l'erreur théorique à  $1\sigma$  due à la dispersion intrinsèque, et les lignes rouges pleine et en pointillés (en bas à droite) représentent respectivement la moyenne et la déviation standard des erreurs. Les barres d'erreurs en vert (en bas à gauche) montre un biais de la moyenne des erreurs en fonction du redshift, il peut s'agir d'un effet de sélection ou d'un flux cohérent.

# Chapitre 5

## Travail futur

### 5.1 Finalisation du travail effectué lors du stage

Comme évoqué en Section 3.2.4, l’analyse  $f\sigma_8$  n’a pas encore pu être réalisée, et constituera une partie de mon travail sur les prochains mois, en parallèle du début de ma thèse. Une première estimation est prévue à l’aide d’une approximation du champ de densité de la simulation Uchuu reposant sur la méthode décrite dans CARRICK et al., 2015 afin de tester et me familiariser avec la méthode de comparaison densité-vélocité pour estimer  $f\sigma_8$ . L’objectif est d’ensuite refaire cette analyse en utilisant cette fois un champ de densité inféré par BORG, afin d’en avoir une meilleure reconstruction.

### 5.2 Postérité pour l’entreprise

Mes deux contributions majeures aux modules et outils existants sont d’une part la génération des SNe suivant un catalogue de galaxies, et l’intégration des modules du pipeline LEMAÎTRE. Les simulations que j’ai produites sont actuellement utilisées par d’autres doctorants afin de tester individuellement les modules NaCl et EDRIS.

La génération des SNe est effectuée à l’aide du module `skysurvey`, mais il s’agit d’une nouvelle fonctionnalité qui n’était pas disponible auparavant. Si le tirage utilise actuellement les positions des SNe suivant le catalogue Uchuu BGS, il est possible d’utiliser n’importe quelle autre catalogue de galaxies ou plus généralement n’importe quelle table de coordonnées  $(RA, DEC, Z)$ . Mon implémentation est actuellement en cours d’intégration au module `skysurvey` comme méthode potentielle de tirage des positions, afin de pouvoir être réutilisée pour des analyses futures.

L’intégration des modules du pipeline LEMAÎTRE a eu lieu progressivement tout au long de mon stage, sous le nom du module GEORGE. Il réutilise le code que j’ai développé afin de faire communiquer les différents modules, et permet également une gestion des versions des modules ainsi que le lancement d’une analyse complète lorsqu’un des modules est mis à jour.



## Chapitre 6

# Bilan personnel

Au cours de ce stage, j'ai été amené à réaliser un travail de recherche à plein temps au sein d'une équipe. Cette expérience était plutôt nouvelle, car si j'avais déjà effectué un stage de recherche au cours de ma césure à l'Oskar Klein Center, à Stockholm, j'avais moins de contacts et contraintes avec une équipe travaillant sur un sujet commun. Il s'agissait d'estimer le gain sur la localisation de sources d'ondes gravitationnelles à l'aide du champ de densité des galaxies, et constituait également une analyse impliquant plusieurs traceurs. J'ai dû pour cela utiliser des packages python déjà développés tels que BILBY<sup>1</sup> et les modifier pour mon utilisation, mais mes développements étaient assez distants des équipes de recherche à l'origine de ces packages. Lors du présent stage, j'ai été en revanche complètement impliqué dans le développement du pipeline LEMAÎTRE. Mes tâches m'ont amené à discuter avec les chercheurs et doctorants travaillant sur toutes les étapes de l'analyse cosmologique, et à devoir comprendre le fonctionnement de chaque module pour les utiliser correctement. Cette dynamique était donc très différente de celle de mon stage de césure.

Ce stage a également été ma première occasion d'assister à une conférence, puisque j'ai eu la chance d'assister au meeting DESI 2024 qui a eu lieu début juillet à Marseille. Cette expérience a été très enrichissante car j'ai pu discuter avec des chercheurs du monde entier travaillant sur des problématiques diverses liées aux données DESI, et découvrir leurs travaux. Cela m'a aussi permis de rencontrer physiquement certains chercheurs qui travaillaient également sur les analyses jointes entre DESI et ZTF, et avec qui j'avais pu échanger par messages. Lors de ce meeting, j'ai également pu présenter mon travail lors des sessions *Spotlight Talks*, qui consistaient en une rapide succession de présentations par des doctorants, stagiaires ou post-doctorants, d'une minute chacune avec une unique slide comme support visuel. Cet exercice était particulièrement instructif puisqu'il a nécessité de sélectionner les informations clés de mon travail et de les synthétiser.

Je me sens bien plus confiant pour débiter ma thèse à l'issue de ce stage qu'avant. Je pense qu'il m'a permis de développer des compétences de communication et d'échange au sein d'une équipe comme des techniques de programmation et d'analyse, ce qui est essentiel à tout travail de recherche moderne.

---

1. <https://lscsoft.docs.ligo.org/bilby/index.htm>





## Chapitre 7

# Enjeux éthiques

### 7.1 Éthique de la recherche

La recherche contient nécessairement une dimension éthique dans la manière de la mener. Un chercheur doit faire preuve d'intégrité et d'honnêteté intellectuelle, même si ses résultats vont à l'encontre de ses croyances. Cette dimension a pleinement été prise en compte lors de mon stage puisque ma tâche principale était justement de faire remonter les problèmes rencontrés dans le pipeline LEMAÎTRE, et de m'assurer que tout comportement étaient soit compris, soit explicable par un effet précis, auquel cas l'effet doit être quantifié.

Il est également important de mentionner une pratique devenue courante, le *blinding* des données. Cela consiste en une altération volontaire des données avant réalisation de l'analyse, afin de s'assurer que l'analyse n'est pas biaisée par un a priori sur ces données ou le modèle choisi pour les décrire. Un bon modèle de *blinding* doit respecter plusieurs critères : la qualité des données et leurs propriétés statistiques doit être conservée, il doit être difficile de l'inverser – pour éviter que les données réelles puissent être extraites et utilisées telles quelles – et il doit viser des paramètres précis, sans affecter les variables et diagnostics de l'analyse. Dans le cas de DESI (ANDRADE et al., 2024), le *blinding* est effectué en décalant les redshifts à l'aide d'une cosmologie et d'un facteur de croissance des structures fixés et inconnus des chercheurs chargés de l'analyse. La méthode du *blinding* du lot de données réel utilisé par le pipeline LEMAÎTRE n'a pas encore été communiquée, mais intervient sur les flux mesurés pour les décaler.

### 7.2 Égalité des genres

Concernant l'égalité des genres, il existe toujours au sein de la recherche une grosse disparité dans les proportions d'hommes et de femmes, en partie due aux proportions déjà inégales en universités et écoles d'ingénieurs. Il est également intéressant de noter la sur-représentation des femmes aux postes administratifs et de gestion. Je n'ai pas noté de discriminations lors de mon stage, en revanche j'ai parfois entendu des propos déplacés ou sexistes venant de chercheurs plus âgés dans des contextes informels. Ces comportements relèvent du sexisme ordinaire et demeurent problématiques, mais ne donnent pas lieu à une différence de traitement entre hommes et femmes pour l'accès aux postes, la parole lors des réunions ou des présentations, ou encore l'attribution du premier auteur lors de la publication des papiers.

Sorbonne Université a mis en place une cellule d'écoute et d'aide pour lutter contre les VSS, et j'ai pu voir des affiches de sensibilisations sur les comportements sexistes dans les bâtiments

pour alerter sur ces questions.

### 7.3 Impact environnemental

L’impact environnemental de la recherche fondamentale est non négligeable. Si on peut pointer du doigt les accélérateurs, télescopes et supercalculateurs nécessaires à l’avancée de la science, il ne s’agit pas du point le plus critique pour la réduction des émissions de gaz à effet de serre et de consommation d’énergie. En effet, lors du meeting DESI 2024, une présentation lors d’une session plénière a été consacrée à l’étude de l’impact environnemental de la collaboration. Et loin devant l’utilisation du NERSC pour les simulations et analyses, l’activité la plus polluante était les meetings internationaux eux-mêmes puisqu’ils engendraient de nombreux déplacements, généralement en avion. Cela engendre également un coût que tous les chercheurs ne peuvent pas couvrir, et contraint fortement les emplois du temps.

Ces problèmes n’ont pas encore résolu à une échelle globale mais des solutions commencent à être proposés. Par exemple le *International Centre for Radio Astronomy Research* a organisé le *2024 Astronomical Society of Australia Annual Scientific Meeting* selon une politique *online-first*, sur la base de l’article SARABIPOUR et al., 2021. Cela signifie que le principal canal de communication n’est pas le présentiel, mais en ligne. Les présentations sont également retransmises, et les chercheurs sont invités à participer à des *local Hub Day*, des rencontres entre scientifiques d’une même région ne nécessitant que peu de transports. Cela réduit donc les émissions polluantes d’une part, les coûts d’autre part, et permet également une meilleure inclusion des personnes handicapées ou ne pouvant se déplacer internationalement pour diverses raisons. Cette initiative est particulièrement intéressante, puisqu’elle essaye de concilier les besoins des rencontres internationales de chercheurs aux problèmes environnementaux et sociaux à l’aide des outils en ligne perfectionnés pendant la pandémie de COVID-19.

# Annexe A

## Annexes

### A.1 Fonctionnement des modules

#### A.1.1 PETS

PETS a pour but de sélectionner les SN qui sont suffisamment bien échantillonnées pour l'entraînement du modèle par NaCl. Le fonctionnement de PETS est découpé en 3 étapes, tous utilisant le modèle SALT2.4 déjà entraîné :

1. Une sélection des points de données utilisables.
2. La réalisation d'une grille de temps de maximums et des fits associés à chacun de ces temps
3. La sélection des SNe

La sélection des points de données est effectuée en rejetant itérativement les points de données à plus de  $3\sigma$  du fit SALT2.4. Ensuite, en partant du  $t_{max,0}$  d'un fit complet des points restants, une grille de temps de maximum est réalisé. Elle contient 150 points dans les intervalles  $[t_{max,0} - 20; t_{max,0} - 5]$  et  $[t_{max,0} + 5; t_{max,0} + 20]$  et 1000 points dans l'intervalle  $[t_{max,0} - 5; t_{max,0} + 5]$ . Le modèle est alors fitté sur les données en fixant le temps de maximum à un de ces points, puis PETS analyse la courbe  $\chi^2(t_{max})$  afin de garder ou rejeter la SN. Une SN est gardée si :

1.  $\chi^2(t_{max})$  présente un minimum global à  $t_{min}$ , qui n'est pas aux bords de l'intervalle  $[t_{max,0} - 20; t_{max,0} + 20]$
2. L'incertitude sur  $t_{min}$  est inférieure à 1 jour.
3. L'écart entre les incertitudes positives et négatives à  $3\sigma$  est inférieur à 0.3, ce qui caractérise la symétrie de  $\chi^2(t_{max})$  autour de  $t_{min}$ .
4.  $\chi^2(t_{max})$  ne présente pas d'autres minimums à  $8\sigma$  de ce minimum global, *i.e.* il n'y a pas d'autres minimums proches susceptibles de coïncider le fit de NaCl.
5. Les valeurs de  $x_1$  et  $c$  fittées sont respectivement inférieures à 4 et 2, ce qui impose que le fit ne décrit pas une SN extrêmement distordue (bien qu'elle soit toujours décrite par le modèle).

Une version mise à jour de PETS utilisant un modèle NaCl pré-entraîné à la place de SALT2.4 est en cours de développement, mais je ne l'ai pas utilisée.

### A.1.2 NaCl

#### Modélisation des SNe

**NaCl** consiste en une implémentation d'un modèle type **SALT**, avec quelques améliorations. Tout comme **SALT**, le flux émis par les SNe dans leurs référentiels propres est modélisé par une surface spectrale  $\mathcal{S}(\lambda, p)$ , où  $\lambda$  est la longueur d'onde d'émission dans le référentiel de la SN,  $p$ , la phase, est le temps depuis le maximum d'émission en bande B. De cette surface de flux, on peut déduire dans notre référentiel à la fois le flux photométrique, c'est-à-dire le flux observé dans un filtre de bande passante  $T(\lambda)$  à une date donnée, et les spectres émis en redshiftant cette surface :

$$\phi_{phot} = X_0 \frac{1}{1+z} \int \mathcal{S}\left(\frac{\lambda}{1+z}, p\right) \frac{\lambda}{hc} T(\lambda) d\lambda \quad (\text{A.1})$$

$$\phi_{spec} = X_0 \frac{1}{1+z} \mathcal{S}(\lambda, p) \quad (\text{A.2})$$

où  $X_0 = \Phi_B \frac{(10\text{pc})^2}{d_L(z)^2}$  est l'amplitude, avec  $\Phi_B$  la luminosité maximale en bande B normalisée pour une SN à 10 pc, et  $d_L(z)$  est la distance de luminosité. Toute l'information sur la cosmologie est donc encodée dans  $X_0$  via la distance de luminosité. Pour réduire le temps de calcul, l'équation A.1 est ré-exprimée en intégrant dans le référentiel de la SN, ce qui équivaut à blueshifter les filtres plutôt que de redshifter la surface spectrale :

$$\phi_{phot} = X_0 (1+z) \int \mathcal{S}(\lambda, p) \frac{\lambda}{hc} T((1+z)\lambda) d\lambda \quad (\text{A.3})$$

La surface spectrale est décrite de la même manière que pour **SALT2.4** avec la paramétrisation par SN suivante

$$\mathcal{S}(p, \lambda) = [M_0(p, \lambda) + X_1 M_1(p, \lambda)] \times \exp[0.4cCL(\lambda)] \quad (\text{A.4})$$

où  $X_1$  et  $c$  sont les paramètres de stretch et couleurs propre à la SN considérée.

Les modèles  $M_0(p, \lambda)$  et  $M_1(p, \lambda)$  sont exprimés sur une base de splines, et représentés par un vecteur de coefficients  $\theta$ . Cette compression facilite leurs évaluations et les calculs de leurs dérivées. Les modèles sont ainsi évaluable en tout point ce qui permet de s'affranchir des effets indésirables introduits par une discrétisation (nécessité d'interpoler, moins bonne précision sur les dérivées,...) .

#### Différences entre NaCl et SALT2.4

Les différences entre **NaCl** et **SALT2.4** ne sont pas liées au modèle de SNe lui-même, mais à des choix dans les paramètres ajustés lors de l'entraînement :

- Le temps du maximum  $t_{max}$  n'est plus déterminé avant entraînement et fixé, mais est laissé comme paramètre libre. Autrement dit, les points de données dans l'espace des phases  $(\lambda, p)$  sont susceptibles de bouger selon  $p$ , puisque décaler la date de maximum revient à décaler d'autant tous les points de mesures.
- La calibration des filtres est intégrée dans le modèle ce qui permet d'estimer directement son impact sur les paramètres reconstruits, plutôt que d'entraîner le modèle sur différents lots avec des calibrations variables.
- En conséquence de la différence précédente, le modèle d'erreur, c'est à dire l'incertitude du modèle sur les valeurs de flux, est également ajusté lors de l'entraînement plutôt qu'estimé a posteriori.

### Contraintes

En conséquence de ces choix, les contraintes imposées sur les modèles diffèrent de celles utilisées pour entraîner **SALT2**. Une des idées fondamentales de ces contraintes est que  $M_0$  doit décrire la SN moyenne, et que tous les termes autour ne doivent être que des corrections. Cela se traduit par exemple dans la normalisation des modèles, dégénérée avec  $X_0$ , en imposant

$$\int M_0(p=0, \lambda) \frac{\lambda}{hc} B(\lambda) d\lambda = 1 \quad (\text{A.5})$$

$$\int M_0(p=0, \lambda) \frac{\lambda}{hc} B(\lambda) d\lambda = 0 \quad (\text{A.6})$$

De même, les couleurs et stretches sont contraints pour être de moyenne nulle, par

$$\sum_{i=1}^{N_{SN}} c_i = 0 \quad (\text{A.7})$$

$$\sum_{i=1}^{N_{SN}} X_{1,i} = 0 \quad (\text{A.8})$$

$$(\text{A.9})$$

À cela s'ajoute une dégénérescence liée aux normalisations des  $X_1$  relativement à  $M_1$ , qui est levée en imposant une contrainte sur la variance des  $X_1$

$$\frac{1}{N_{SN}} \sum_{i=1}^{N_{SN}} (X_{1,i} - \langle X_1 \rangle)^2 = 1 \quad (\text{A.10})$$

La variabilité du temps de maximum introduit également une dégénérescence avec les modèles  $M_0$  et  $M_1$ . Cette dégénérescence est brisée en imposant aux modèles de présenter un maximum en bande B à  $p=0$  :

$$\left. \frac{d}{dp} \int M_0(p, \lambda) \frac{\lambda}{hc} B(\lambda) d\lambda \right|_{p=0} = 0 \quad (\text{A.11})$$

$$\left. \frac{d}{dp} \int M_1(p, \lambda) \frac{\lambda}{hc} B(\lambda) d\lambda \right|_{p=0} = 0 \quad (\text{A.12})$$

Toutes ces contraintes sont exprimées sous la forme

$$C(\theta) - \gamma = 0 \quad (\text{A.13})$$

où  $\theta$  représente le vecteur des paramètres des modèles, et intégrées dans la vraisemblance dans un terme

$$\mu_{cons} C(\theta)^T C(\theta) \quad (\text{A.14})$$

Enfin, un terme de régularisation est introduit sous la forme

$$\mu_{reg} \left( \sum_{k,l} \theta_{k,l}^2 + \sum_{k,l} (\theta_{k+1,l} - \theta_{k,l})^2 \right) \quad (\text{A.15})$$

afin de lever d'éventuelles dégénérescences non identifiées, en faisant tendre les paramètres du modèle vers 0 lorsqu'ils ne sont pas contraint par des points de données via premier terme et en limitant l'amplitude des fluctuations des modèles via le second terme.

## Entraînement

L'entraînement a lieu en 3 étapes après initialisation des modèles  $M_0$  et  $M_1$  sur les modèles de SALT2.4. Une première estimation des paramètres de chaque SN ( $x_0, x_1, c, t_{max}$ ) est réalisée à modèles et modèle d'erreur fixés. Autrement dit, cette première estimation revient à ré-estimer les paramètres des SNe en utilisant les mêmes modèles que SALT2.4. Ensuite, les paramètres des SNe sont fixés aux valeurs obtenues, et seul le modèle d'erreur est laissé libre et ajusté. Pour finir, tous les paramètres sont relâchés et le modèle est ajusté en même temps que les paramètres des SNe et le modèle d'erreur. C'est lors de cette dernière étape que les contraintes décrites précédemment sont effectivement appliquées au modèle, puisque les modèles de SALT2.4 utilisaient d'autres contraintes.

Ce choix vient du fait qu'un entraînement sur tous les paramètres serait trop coûteux et qu'il pourrait se bloquer plus facilement dans un minimum local.

### A.1.3 EDRIS

Le module EDRIS réalise un diagramme de Hubble en ajustant un modèle cosmologique aux modules de distances de SNe en prenant en compte le biais de Malmquist (cf. Figure A.1) et en ré-estimant les paramètres de standardisation, de manière à obtenir directement la matrice de covariance complète. Contrairement aux analyses précédemment employées qui quantifient l'effet du biais de Malmquist par des simulations et propagent ensuite une correction, l'objectif est d'inclure un terme correctif directement dans le modèle. Cette solution est bien plus puissante, car elle s'affranchit de biais potentiels des simulations elles-mêmes et nécessite un temps de calcul bien plus faible.

Pour modéliser les SNe, EDRIS se base sur la formule de Tripp (cf 2.1) et utilise le vecteur de paramètres suivant

$$Y_i = \begin{pmatrix} m_i \\ x_{1,i} \\ c_i \end{pmatrix} = \begin{pmatrix} M_b + \mu(z_i, \Theta) - \alpha x_{1,i} + \beta c_i \\ x_{1,i} \\ c_i \end{pmatrix} + \begin{pmatrix} \varepsilon_i \\ 0 \\ 0 \end{pmatrix} \quad (\text{A.16})$$

où  $m_i$ ,  $x_{1,i}$ ,  $c_i$  et  $z_i$  sont la magnitude apparente, le stretch, la couleur et le redshift de la SN  $i$ ,  $\mu(z, \theta)$  est le module de distance en fonction du redshift et de la cosmologie de paramètres  $\Theta$ ,  $M_b$  est la magnitude absolue et  $\varepsilon_i \sim \mathcal{N}(0, \sigma_{int}^2)$  est le bruit causé par la dispersion intrinsèque.

L'effet de la coupure du biais de Malmquist s'exprime alors sous la forme

$$\begin{aligned} Y_i^{obs} &= Y_i + \eta_i \quad \text{si} \quad m_i \leq m_{lim} + \kappa_i \\ Y_i^{obs} &\quad \text{n'est pas observé sinon} \end{aligned} \quad (\text{A.17})$$

où  $C$  est la matrice de covariance complète des paramètres  $(m_i, x_i, c_i)$ ,  $\eta_i \sim \mathcal{N}(0, C_i)$  est le bruit de mesure,  $m_{lim}$  est la magnitude limite du relevé, et  $\kappa_i \sim \mathcal{N}(0, \sigma_d^2)$  est la fluctuation due aux conditions d'observations.

L'effet du biais de Malmquist est alors équivalent à l'introduction d'une fonction de sélection sigmoïde en fonction de la magnitude limite sur les données observées, de paramètres  $m_{lim}$  et  $\sigma_d$ .

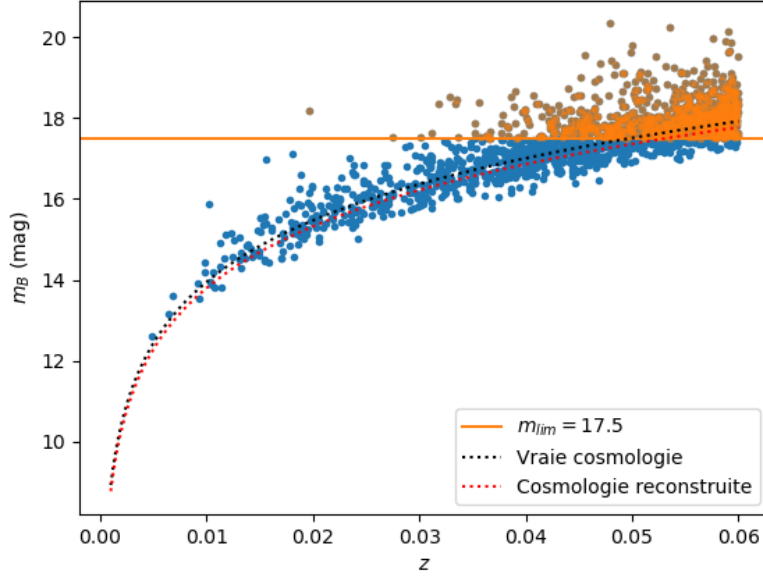


FIGURE A.1 – Illustration d’un biais de Malmquist sur une distribution simulée. Un télescope présentant une magnitude limite à 17.5 n’observera pas les SNe marqués en orange, et une analyse ne prenant pas cette effet en compte reconstruira une cosmologie biaisée négativement par rapport à la cosmologie sous-jacente.

La log-vraisemblance s’exprime alors sous la forme

$$\begin{aligned}
 \Gamma &= -\ln(|C^{-1}|) + \mathbf{r}^\dagger C^{-1} \mathbf{r} \\
 &+ \sum_i 2 \ln \left[ \Phi \left( \frac{m_{lim} - M_b - \mu_i + \alpha x_i - \beta c_i}{\sqrt{\sigma_{int}^2 + \sigma_d^2}} \right) \right] - 2 \ln \left[ \Phi \left( \frac{m_{lim} - m_i^{obs}}{\sqrt{\sigma_d^2 + f(C_i)}} \right) \right] \\
 \text{avec} \quad \Phi(x) &= \frac{1}{2} (1 + \operatorname{erf}(\frac{x}{\sqrt{2}})) \quad \text{et} \quad \mathbf{r} = \begin{pmatrix} m^{obs} \\ x_1^{obs} \\ c^{obs} \end{pmatrix} - \begin{pmatrix} M - \mu + \alpha x_1 - \beta c \\ x_1 \\ c \end{pmatrix}
 \end{aligned} \tag{A.18}$$

$\Phi(x)$  est la fonction de répartition d’une Gaussienne centrée réduite et  $\mathbf{r}$  est le vecteur des résidus entre les paramètres ré-estimés et les paramètres observés. Ce sont les deux derniers termes de la log-vraisemblance qui incluent le biais de Malmquist dans l’ajustement des paramètres.

Pour ajuster le modèle  $\mu(z, \Theta)$  et réévaluer les paramètres, EDRIS a nécessité de nombreuses optimisations. Une de ces optimisations a été une implémentation intelligente de la matrice de covariance utilisant le complément de Schur pour l’inverser efficacement. Une deuxième optimisation est l’utilisation de la méthode d’optimisation dite *Hessian-free* décrite dans MARTENS, 2010, implémentée en JAX. Cette méthode présente l’avantage de ne nécessiter que des calculs de gradients afin de déterminer une approximation de la hessienne, et non de la hessienne complète.



## A.2 Méthode de l'analyse jointe $f\sigma_8$

Comme mentionné précédemment, l'analyse  $f\sigma_8$  se base sur la méthode de comparaison entre les vitesses particulières et celles obtenues à partir du champ de densité, plus précisément sur la méthodologie de STAHL et al., 2021.

La relation entre le champ de vitesse et le contraste de densité des galaxies  $\delta_g = \frac{\rho_g}{\rho_g} - 1$  est donné par :

$$\mathbf{v}(\mathbf{r}) = \frac{H_0 \mathcal{B}}{4\pi} \int_0^{R_{max}} d^3 \mathbf{r}' \delta_g(\mathbf{r}') \frac{\mathbf{r}' - \mathbf{r}}{|\mathbf{r}' - \mathbf{r}|} + \mathbf{V}_{ext} \quad (\text{A.19})$$

où  $\mathcal{B} = \frac{f}{b}$ , avec  $b$  le biais entre le contraste de densité des galaxies et celui de la matière noire ( $\delta_g = b\delta_{DM}$ ),  $R_{max}$  est la distance maximale couverte par le catalogue, et  $\mathbf{V}_{ext}$  est le flux cohérent causé par des structures à l'extérieur du catalogue.

Les vitesses obtenues à partir du champ de densité sont alors ajustées sur celles des vitesses particulières issues des SN à partir d'une méthode *Forward Likelihood*. Les paramètres  $\mathcal{B}$  et  $\mathbf{V}_{ext}$  ainsi que d'autres paramètres de nuisance liés aux observations  $\Theta$  (notamment les paramètres du modèle de Tripp  $\alpha$ ,  $\beta$ ,  $M_b$  et  $\sigma_{int}$ , voir Eq. 2.1) sont reliés aux paramètres observables des SN  $\{\mathbf{x}_i\}$  via la probabilité conditionnelle  $\mathcal{P}(\mathcal{B}, \mathbf{V}_{ext}, \Theta | \mathbf{x}_i)$ , qui peut être exprimée à l'aide du théorème de Bayes (STAHL et al., 2021)

$$\mathcal{P}(\mathcal{B}, \mathbf{V}_{ext}, \Theta | \mathbf{x}_i) \propto \mathcal{P}(\mathbf{x}_i | \mathcal{B}, \mathbf{V}_{ext}, \Theta) \mathcal{P}(\mathcal{B}, \mathbf{V}_{ext}, \Theta) \quad (\text{A.20})$$

On cherche donc à reconstruire les distributions contraintes par les observations des paramètres de biais galactique, de flux cohérent, et de nuisance ( $\mathcal{P}(\mathcal{B}, \mathbf{V}_{ext}, \Theta | \mathbf{x}_i)$ ) à l'aide de la distribution des SN de paramètres  $\mathbf{x}_i$  compte tenu du biais galactique, du flux cohérent, et des paramètres de nuisance ( $\mathcal{P}(\mathbf{x}_i | \mathcal{B}, \mathbf{V}_{ext}, \Theta)$ ) et d'un prior sur la distributions de ces paramètres ( $\mathcal{P}(\mathcal{B}, \mathbf{V}_{ext}, \Theta)$ ).

Les priors sur les paramètres de nuisance du modèle de Tripp sont non restrictifs, ils imposent simplement  $\alpha, \beta > 0$  et  $M < 0$ , tandis que la dispersion intrinsèque  $\sigma_{int}$  (cf. Eq. 2.1) est log-normale avec un pic à 0.15 mag.

Pour corriger d'éventuelles inhomogénéités selon différentes lignes de visées, on doit introduire de plus une distribution radiale

$$\mathcal{P}(\mathbf{r} | \Theta) = \frac{1}{\mathcal{N}(\Theta)} r^2 \exp \left\{ -\frac{[\mu(\mathbf{r}) - \mu(\Theta)]^2}{2\sigma_\mu^2(\Theta)} \right\} [1 + \delta_g(\mathbf{r})] \quad (\text{A.21})$$

où  $\mathbf{r}$  est un vecteur comobile pointant sur la sphère céleste,  $\mathcal{N}(\Theta)$  est un coefficient de normalisation,  $\mu$  est le module de distance, et  $\sigma_\mu$  est l'erreur sur le module de distance causée par la dispersion intrinsèque des SNe. On obtient alors la vraisemblance suivante

$$\mathcal{P}(\mathbf{x}_i | \mathcal{B}, \mathbf{V}_{ext}, \Theta) = \int_0^{R_{max}} dr \mathcal{P}(\mathbf{x}_i | \mathbf{r}, \mathcal{B}, \mathbf{V}_{ext}, \Theta) \mathcal{P}(\mathbf{r} | \Theta) \quad (\text{A.22})$$

avec

$$\mathcal{P}(\mathbf{x}_i | \mathbf{r}, \mathcal{B}, \mathbf{V}_{ext}, \Theta) = \frac{1}{\sqrt{2\pi\sigma_\nu^2}} \exp \left\{ -\frac{[cz_{obs} - cz_{pred}(\mathbf{r}, \mathcal{B}, \mathbf{V}_{ext})]^2}{2\sigma_\nu^2} \right\} \quad (\text{A.23})$$

$z_{obs}$  est le redshift observé de la SN, et  $z_{pred}$  est la prédiction de ce redshift à partir du redshift cosmologique et des vitesses reconstruites grâce à

$$1 + z_{pred}(\mathbf{r}, \mathcal{B}, \mathbf{V}_{ext}) = (1 + z_{cosmo}(\mathbf{r})) \left[ 1 + \frac{1}{c} (\mathcal{B} \mathbf{v}(\mathbf{r}) + \mathbf{V}_{ext}) \cdot \hat{\mathbf{r}} \right] \quad (\text{A.24})$$

$\sigma_\nu$  peut être soit ajusté avec les autres paramètres de nuisance comme dans STAHL et al., 2021, auquel cas il suit une dispersion gaussienne, soit fixé comme dans BORUAH et al., 2020.

Finalement, la log-vraisemblance s'écrit

$$\mathcal{L} = \ln \mathcal{P}(\mathcal{B}, \mathbf{V}_{ext}, \boldsymbol{\Theta}) + \sum_i P(\mathbf{x}_i | \mathcal{B}, \mathbf{V}_{ext}, \boldsymbol{\Theta}) \quad (\text{A.25})$$

Cette vraisemblance est alors explorée avec une Chaîne de Markov Monte Carlo, et produit des résultats comme ceux de STAHL et al., 2021 représentés en Fig. A.2, qui sont en très bon accord avec d'autres analyses  $f\sigma_8$  et ont une meilleure précision que les analyses  $f\sigma_8$  utilisant seulement les RSD.

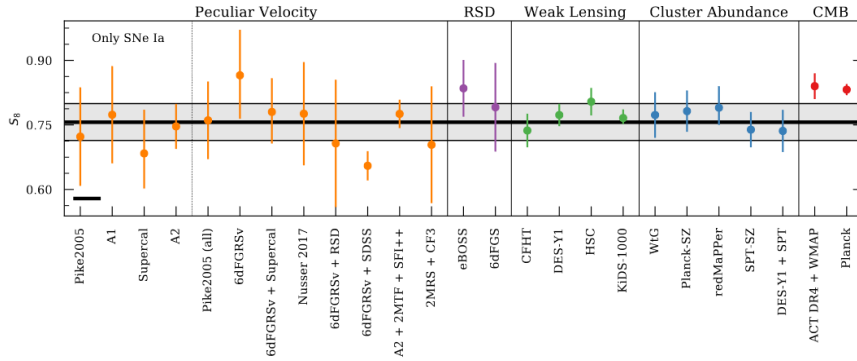


FIGURE A.2 – Paramètre  $S_8 = f\sigma_8/(0.3)^{0.55}$  reconstruit par diverses méthodes. Les lignes noires représentent la valeur et l'incertitude de la valeur reconstruite par STAHL et al., 2021. Crédit : B. Stahl



# Bibliographie

- ADAMS, C., & BLAKE, C. (2020). Joint growth-rate measurements from redshift-space distortions and peculiar velocities in the 6dF Galaxy Survey [Publisher : OUP ADS Bibcode : 2020MNRAS.494.3275A]. *Monthly Notices of the Royal Astronomical Society*, 494, 3275-3293. <https://doi.org/10.1093/mnras/staa845>
- AIHARA, H., ARIMOTO, N., ARMSTRONG, R., ARNOUTS, S., BAHCALL, N. A., BICKERTON, S., BOSCH, J., BUNDY, K., CAPAK, P. L., CHAN, J. H. H., CHIBA, M., COUPON, J., EGAMI, E., ENOKI, M., FINET, F., FUJIMORI, H., FUJIMOTO, S., FURUSAWA, H., FURUSAWA, J., ... YUMA, S. (2018). The Hyper Suprime-Cam SSP Survey : Overview and survey design [Publisher : OUP ADS Bibcode : 2018PASJ...70S...4A]. *Publications of the Astronomical Society of Japan*, 70, S4. <https://doi.org/10.1093/pasj/psx066>
- ALAM, S., ATA, M., BAILEY, S., BEUTLER, F., BIZYAEV, D., BLAZEK, J. A., BOLTON, A. S., BROWNSTEIN, J. R., BURDEN, A., CHUANG, C.-H., COMPARAT, J., CUESTA, A. J., DAWSON, K. S., EISENSTEIN, D. J., ESCOFFIER, S., GIL-MARÍN, H., GRIEB, J. N., HAND, N., HO, S., ... ZHAO, G.-B. (2017). The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey : cosmological analysis of the DR12 galaxy sample [Publisher : OUP ADS Bibcode : 2017MNRAS.470.2617A]. *Monthly Notices of the Royal Astronomical Society*, 470, 2617-2652. <https://doi.org/10.1093/mnras/stx721>
- ANDRADE, U., MENA-FERNÁNDEZ, J., AWAN, H., ROSS, A. J., BRIEDEN, S., PAN, J., de MATTIA, A., AGUILAR, J., AHLEN, S., ALVES, O., BROOKS, D., BUCKLEY-GEER, E., CHAUSSIDON, E., CLAYBAUGH, T., COLE, S., de la MACORRA, A., DEY, A., DOEL, P., FANNING, K., ... WEAVER, B. A. (2024, avril). Validating the Galaxy and Quasar Catalog-Level Blinding Scheme for the DESI 2024 analysis [arXiv:2404.07282 [astro-ph]]. <https://doi.org/10.48550/arXiv.2404.07282>
- BELLM, E. C., KULKARNI, S. R., GRAHAM, M. J., DEKANY, R., SMITH, R. M., RIDDLE, R., MASCI, F. J., HELOU, G., PRINCE, T. A., ADAMS, S. M., BARBARINO, C., BARLOW, T., BAUER, J., BECK, R., BELICKI, J., BISWAS, R., BLAGORODNOVA, N., BODEWITS, D., BOLIN, B., ... ZOLKOWER, J. (2018). The Zwicky Transient Facility : System Overview, Performance, and First Results [Publisher : The Astronomical Society of the Pacific]. *PASP*, 131(995), 018002. <https://doi.org/10.1088/1538-3873/aacbe>
- BEUTLER, F., BLAKE, C., COLLESS, M., JONES, D. H., STAVELEY-SMITH, L., POOLE, G. B., CAMPBELL, L., PARKER, Q., SAUNDERS, W., & WATSON, F. (2012). The 6dF Galaxy Survey :  $\sim 0\%$  measurements of the growth rate and  $\sigma_8$  [Publisher : OUP ADS Bibcode : 2012MNRAS.423.3430B]. *Monthly Notices of the Royal Astronomical Society*, 423, 3430-3444. <https://doi.org/10.1111/j.1365-2966.2012.21136.x>
- BORUAH, S. S., HUDSON, M. J., & LAVAUX, G. (2020). Cosmic flows in the nearby Universe : new peculiar velocities from SNe and cosmological constraints [arXiv:1912.09383 [astro-ph]]. *Monthly Notices of the Royal Astronomical Society*, 498(2), 2703-2718. <https://doi.org/10.1093/mnras/staa2485>

- BORUAH, S. S., LAVAUX, G., & HUDSON, M. J. (2022). Bayesian reconstruction of dark matter distribution from peculiar velocities : accounting for inhomogeneous Malmquist bias [Publisher : OUP ADS Bibcode : 2022MNRAS.517.4529B]. *Monthly Notices of the Royal Astronomical Society*, 517, 4529-4543. <https://doi.org/10.1093/mnras/stac2985>
- BOYD, B. M., GRAYLING, M., THORP, S., & MANDEL, K. S. (2024, juillet). Accounting for Selection Effects in Supernova Cosmology with Simulation-Based Inference and Hierarchical Bayesian Modelling [arXiv:2407.15923 [astro-ph]]. <https://doi.org/10.48550/arXiv.2407.15923>
- CARRERES, B., BAUTISTA, J. E., FEINSTEIN, F., FOCHEZ, D., RACINE, B., SMITH, M., AMENOUCHE, M., AUBERT, M., DHAWAN, S., GINOLIN, M., GOOBAR, A., GRIS, P., LACROIX, L., NUSS, E., REGNAULT, N., RIGAULT, M., ROBERT, E., ROSNET, P., SOMMER, K., ... PURDUM, J. (2023). Growth-rate measurement with type-Ia supernovae using ZTF survey simulations. *A&A*, 674, A197. <https://doi.org/10.1051/0004-6361/202346173>
- CARRICK, J., TURNBULL, S. J., LAVAUX, G., & HUDSON, M. J. (2015). Cosmological parameters from the comparison of peculiar velocities with predictions from the 2M++ density field [arXiv:1504.04627 [astro-ph]]. *Monthly Notices of the Royal Astronomical Society*, 450(1), 317-332. <https://doi.org/10.1093/mnras/stv547>
- CHEN, S., DEROSE, J., ZHOU, R., WHITE, M., FERRARO, S., BLAKE, C., LANGE, J. U., WECHSLER, R. H., AGUILAR, J., AHLEN, S., BROOKS, D., CLAYBAUGH, T., DAWSON, K., de la MACORRA, A., DOEL, P., FONT-RIBERA, A., GAZTAÑAGA, E., GONTCHO, S. G. A., GUTIERREZ, G., ... WEAVER, B. A. (2024, juillet). Not all lensing is low : An analysis of DESI $\times$ DES using the Lagrangian Effective Theory of LSS [arXiv:2407.04795 [astro-ph]]. <https://doi.org/10.48550/arXiv.2407.04795>
- DARK ENERGY SURVEY, KILO-DEGREE SURVEY COLLABORATION, ABBOTT, T. M. C., AGUENA, M., ALARCON, A., ALVES, O., AMON, A., ANDRADE-OLIVEIRA, F., ASGARI, M., AVILA, S., BACON, D., BECHTOL, K., BECKER, M. R., BERNSTEIN, G. M., BERTIN, E., BILICKI, M., BLAZEK, J., BOCQUET, S., BROOKS, D., ... ZUNTZ, J. (2023). DES Y3 + KiDS-1000 : Consistent cosmology combining cosmic shear surveys [arXiv:2305.17173 [astro-ph]]. *TheOJA*, 6, 10.21105/astro.2305.17173. <https://doi.org/10.21105/astro.2305.17173>
- DAVIS, T. M., HUI, L., FRIEMAN, J. A., HAUGBØLLE, T., KESSLER, R., SINCLAIR, B., SOLLERMAN, J., BASSETT, B., MARRINER, J., MÖRTSELL, E., NICHOL, R. C., RICHMOND, M. W., SAKO, M., & SCHNEIDER, D. P. (2011). The effect of peculiar velocities on supernova cosmology [arXiv:1012.2912 [astro-ph]]. *ApJ*, 741(1), 67. <https://doi.org/10.1088/0004-637X/741/1/67>
- DESI COLLABORATION, ADAME, A. G., AGUILAR, J., AHLEN, S., ALAM, S., ALEXANDER, D. M., ALVAREZ, M., ALVES, O., ANAND, A., ANDRADE, U., ARMENGAUD, E., AVILA, S., AVILES, A., AWAN, H., BAHR-KALUS, B., BAILEY, S., BALTAY, C., BAULT, A., BEHERA, J., ... ZOU, H. (2024a, avril). DESI 2024 VI : Cosmological Constraints from the Measurements of Baryon Acoustic Oscillations [arXiv:2404.03002 [astro-ph]]. <https://doi.org/10.48550/arXiv.2404.03002>
- DESI COLLABORATION, ADAME, A. G., AGUILAR, J., AHLEN, S., ALAM, S., ALEXANDER, D. M., ALVAREZ, M., ALVES, O., ANAND, A., ANDRADE, U., ARMENGAUD, E., AVILA, S., AVILES, A., AWAN, H., BAILEY, S., BALTAY, C., BAULT, A., BAUTISTA, J., BEHERA, J., ... ZOU, H. (2024b, septembre). DESI 2024 IV : Baryon Acoustic Oscillations from the Lyman Alpha Forest [arXiv:2404.03001 [astro-ph]]. <https://doi.org/10.48550/arXiv.2404.03001>
- DESI COLLABORATION, ADAME, A. G., AGUILAR, J., AHLEN, S., ALAM, S., ALEXANDER, D. M., ALVAREZ, M., ALVES, O., ANAND, A., ANDRADE, U., ARMENGAUD, E., AVILA, S., AVILES, A., AWAN, H., BAILEY, S., BALTAY, C., BAULT, A., BEHERA, J., BENZVI,

- S., ... ZOU, H. (2024c, avril). DESI 2024 III : Baryon Acoustic Oscillations from Galaxies and Quasars [arXiv:2404.03000 [astro-ph]]. <https://doi.org/10.48550/arXiv.2404.03000>
- DJORGovski, S., & DAVIS, M. (1987). Fundamental Properties of Elliptical Galaxies [Publisher : IOP ADS Bibcode : 1987ApJ...313...59D]. *The Astrophysical Journal*, 313, 59. <https://doi.org/10.1086/164948>
- DUPUY, A., COURTOIS, H. M., & KUBIK, B. (2019). An estimation of the local growth rate from Cosmicflows peculiar velocities [Publisher : OUP ADS Bibcode : 2019MNRAS.486..440D]. *Monthly Notices of the Royal Astronomical Society*, 486, 440-448. <https://doi.org/10.1093/mnras/stz901>
- EBOSS COLLABORATION, ALAM, S., AUBERT, M., AVILA, S., BALLAND, C., BAUTISTA, J. E., BERSHADY, M. A., BIZYAEV, D., BLANTON, M. R., BOLTON, A. S., BOVY, J., BRINKMANN, J., BROWNSTEIN, J. R., BURTIN, E., CHABANIER, S., CHAPMAN, M. J., CHOI, P. D., CHUANG, C.-H., COMPARAT, J., ... ZHENG, Z. (2021). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey : Cosmological Implications from two Decades of Spectroscopic Surveys at the Apache Point observatory [arXiv:2007.08991 [astro-ph]]. *Phys. Rev. D*, 103(8), 083533. <https://doi.org/10.1103/PhysRevD.103.083533>
- EUCLID COLLABORATION, MELLIER, Y., ABDURRO'UF, BARROSO, J. A. A., ACHÚCARRO, A., ADAMEK, J., ADAM, R., ADDISON, G. E., AGHANIM, N., AGUENA, M., AJANI, V., AKRAMI, Y., AL-BAHLAWAN, A., ALAVI, A., ALBUQUERQUE, I. S., ALESTAS, G., ALGUERO, G., ALLAoui, A., ALLEN, S. W., ... ZUMALACARREGUI, M. (2024, mai). Euclid. I. Overview of the Euclid mission [arXiv:2405.13491 [astro-ph]]. <https://doi.org/10.48550/arXiv.2405.13491>
- GINOLIN, M., RIGAULT, M., COPIN, Y., POPOVIC, B., DIMITRIADIS, G., GOOBAR, A., JOHANSSON, J., MAGUIRE, K., NORDIN, J., SMITH, M., AUBERT, M., BARJOU-DELAYRE, C., BURGAZ, U., CARRERES, B., DHAWAN, S., DECKERS, M., FEINSTEIN, F., FOCHEZ, D., GALBANY, L., ... van der WALT, S. (2024, juin). ZTF SN Ia DR2 : Colour standardisation of Type Ia Supernovae and its dependence on environment [arXiv:2406.02072 [astro-ph]]. <https://doi.org/10.48550/arXiv.2406.02072>
- GUY, J., ASTIER, P., BAUMONT, S., HARDIN, D., PAIN, R., REGNAULT, N., BASA, S., CARLBERG, R. G., CONLEY, A., FABBRO, S., FOCHEZ, D., HOOK, I. M., HOWELL, D. A., PERRETT, K., PRITCHET, C. J., RICH, J., SULLIVAN, M., ANTILOGUS, P., AUBOURG, E., ... RUHLMANN-KLEIDER, V. (2007). SALT2 : using distant supernovae to improve the use of Type Ia supernovae as distance indicators [arXiv:astro-ph/0701828]. *A&A*, 466(1), 11-21. <https://doi.org/10.1051/0004-6361:20066930>
- HAHN, C., WILSON, M. J., RUIZ-MACIAS, O., COLE, S., WEINBERG, D. H., MOUSTAKAS, J., KREMIN, A., TINKER, J. L., SMITH, A., WECHSLER, R. H., AHLEN, S., ALAM, S., BAILEY, S., BROOKS, D., COOPER, A. P., DAVIS, T. M., DAWSON, K., DEY, A., DEY, B., ... ZOU, H. (2023). DESI Bright Galaxy Survey : Final Target Selection, Design, and Validation [arXiv:2208.08512 [astro-ph]]. *AJ*, 165(6), 253. <https://doi.org/10.3847/1538-3881/acff8>
- HOWLETT, C., ROSS, A. J., SAMUSHIA, L., PERCIVAL, W. J., & MANERA, M. (2015). The clustering of the SDSS main galaxy sample - II. Mock galaxy catalogues and a measurement of the growth of structure from redshift space distortions at  $z = 0.15$  [Publisher : OUP ADS Bibcode : 2015MNRAS.449..848H]. *Monthly Notices of the Royal Astronomical Society*, 449, 848-866. <https://doi.org/10.1093/mnras/stu2693>
- HOWLETT, C., STAVELEY-SMITH, L., ELAHI, P. J., HONG, T., JARRETT, T. H., JONES, D. H., KORIBALSKI, B. S., MACRI, L. M., MASTERS, K. L., & SPRINGOB, C. M. (2017). 2MTF - VI. Measuring the velocity power spectrum [Publisher : OUP ADS Bibcode : 2017MN-

- RAS.471.3135H]. *Monthly Notices of the Royal Astronomical Society*, 471, 3135-3151. <https://doi.org/10.1093/mnras/stx1521>
- HOYLE, F., & FOWLER, W. A. (1960). Nucleosynthesis in Supernovae. [Publisher : IOP ADS Bibcode : 1960ApJ...132..565H]. *The Astrophysical Journal*, 132, 565. <https://doi.org/10.1086/146963>
- HUTERER, D., SHAFER, D. L., SCOLNIC, D., & SCHMIDT, F. (2017). Testing  $\Lambda$ CDM at the lowest redshifts with SN Ia and galaxy velocities [arXiv:1611.09862 [astro-ph, physics:gr-qc]]. *J. Cosmol. Astropart. Phys.*, 2017(05), 015-015. <https://doi.org/10.1088/1475-7516/2017/05/015>
- JASCHE, J., & LAVAUX, G. (2019). Physical Bayesian modelling of the non-linear matter distribution : New insights into the nearby universe [Publisher : EDP Sciences]. *A&A*, 625, A64. <https://doi.org/10.1051/0004-6361/201833710>
- JOHNSON, A., BLAKE, C., KODA, J., MA, Y.-Z., COLLESS, M., CROCCE, M., DAVIS, T. M., JONES, H., MAGOULAS, C., LUCEY, J. R., MOULD, J., SCRIMGEOUR, M. I., & SPRINGOB, C. M. (2014). The 6dF Galaxy Survey : cosmological constraints from the velocity power spectrum [Publisher : OUP ADS Bibcode : 2014MNRAS.444.3926J]. *Monthly Notices of the Royal Astronomical Society*, 444, 3926-3947. <https://doi.org/10.1093/mnras/stu1615>
- KAISER, N. (1987). Clustering in real space and in redshift space [Publisher : OUP ADS Bibcode : 1987MNRAS.227....1K]. *Monthly Notices of the Royal Astronomical Society*, 227, 1-21. <https://doi.org/10.1093/mnras/227.1.1>
- LAI, Y., HOWLETT, C., & DAVIS, T. M. (2022). Using peculiar velocity surveys to constrain the growth rate of structure with the wide-angle effect [arXiv:2209.04166 [astro-ph]]. *Monthly Notices of the Royal Astronomical Society*, 518(2), 1840-1858. <https://doi.org/10.1093/mnras/stac3252>
- MARTENS, J. (2010). Deep learning via Hessian-free optimization. *Proceedings of the 27th International Conference on International Conference on Machine Learning*, 735-742.
- MIYAZAKI, S., KOMIYAMA, Y., NAKAYA, H., KAMATA, Y., DOI, Y., HAMANA, T., KAROJI, H., FURUSAWA, H., KAWANOMOTO, S., MOROKUMA, T., ISHIZUKA, Y., NARIAI, K., TANAKA, Y., URAGUCHI, F., UTSUMI, Y., OBUCHI, Y., OKURA, Y., OGURI, M., TAKATA, T., ... TAKESHI, K. (2012). Hyper Suprime-Cam [Conference Name : Ground-based and Airborne Instrumentation for Astronomy IV ADS Bibcode : 2012SPIE.8446E..0ZM], 8446, 84460Z. <https://doi.org/10.1117/12.926844>
- NICOLAS, N., RIGALT, M., COPIN, Y., GRAZIANI, R., ALDERING, G., BRIDAY, M., KIM, Y.-L., NORDIN, J., PERLMUTTER, S., & SMITH, M. (2021). Redshift evolution of the underlying type Ia supernova stretch distribution [Publisher : EDP Sciences]. *A&A*, 649, A74. <https://doi.org/10.1051/0004-6361/202038447>
- NUSSER, A. (2017). Velocity-density correlations from the cosmicflows-3 distance catalogue and the 2MASS Redshift Survey [Publisher : OUP ADS Bibcode : 2017MNRAS.470..445N]. *Monthly Notices of the Royal Astronomical Society*, 470, 445-454. <https://doi.org/10.1093/mnras/stx1225>
- OKUMURA, T., HIKAGE, C., TOTANI, T., TONEGAWA, M., OKADA, H., GLAZEBROOK, K., BLAKE, C., FERREIRA, P. G., MORE, S., TARUYA, A., TSUJIKAWA, S., AKIYAMA, M., DALTON, G., GOTO, T., ISHIKAWA, T., IWAMURO, F., MATSUBARA, T., NISHIMICHI, T., OHTA, K., ... YOSHIDA, N. (2016). The Subaru FMOS galaxy redshift survey (FastSound). IV. New constraint on gravity theory from redshift space distortions at  $z \sim 1.4$  [Publisher : OUP ADS Bibcode : 2016PASJ...68...38O]. *Publications of the Astronomical Society of Japan*, 68, 38. <https://doi.org/10.1093/pasj/psw029>
- PERLEY, D. A., FREMLING, C., SOLLERMAN, J., MILLER, A. A., DAHIWALE, A. S., SHARMA, Y., BELL, E. C., BISWAS, R., BRINK, T. G., BRUCH, R. J., DE, K., DEKANY, R.,



- DRAKE, A. J., DUEV, D. A., FILIPPENKO, A. V., GAL-YAM, A., GOOBAR, A., GRAHAM, M. J., GRAHAM, M. L., ... YAN, L. (2020). The Zwicky Transient Facility Bright Transient Survey. II. A Public Statistical Sample for Exploring Supernova Demographics [arXiv:2009.01242 [astro-ph]]. *ApJ*, 904(1), 35. <https://doi.org/10.3847/1538-4357/abbd98>
- PERLMUTTER, S., ALDERING, G., DEUSTUA, S., FABBRO, S., GOLDBERGER, G., GROOM, D. E., KIM, A. G., KIM, M. Y., KNOP, R. A., NUGENT, P., PENNYPACKER, C. R., GOOBAR, A., PAIN, R., HOOK, I. M., LIDMAN, C., ELLIS, R. S., IRWIN, M., MCMAHON, R. G., RUIZ-LAPUENTE, P., ... COUCH, W. J. (1998, décembre). Cosmology from Type Ia Supernovae [arXiv:astro-ph/9812473]. <https://doi.org/10.48550/arXiv.astro-ph/9812473>
- PLANCK COLLABORATION, ADE, P. A. R., AGHANIM, N., ARNAUD, M., ASHDOWN, M., AUMONT, J., BACCIGALUPI, C., BANDAY, A. J., BARREIRO, R. B., BARTLETT, J. G., BARTOLO, N., BATTANER, E., BATTYE, R., BENABED, K., BENOIT, A., BENOIT-LEVY, A., BERNARD, J.-P., BERSANELLI, M., BIELEWICZ, P., ... ZONCA, A. (2016). Planck 2015 results. XIII. Cosmological parameters [arXiv:1502.01589 [astro-ph]]. *A&A*, 594, A13. <https://doi.org/10.1051/0004-6361/201525830>
- PLANCK COLLABORATION, AGHANIM, N., AKRAMI, Y., ASHDOWN, M., AUMONT, J., BACCIGALUPI, C., BALLARDINI, M., BANDAY, A. J., BARREIRO, R. B., BARTOLO, N., BASAK, S., BATTYE, R., BENABED, K., BERNARD, J.-P., BERSANELLI, M., BIELEWICZ, P., BOCK, J. J., BOND, J. R., BORRILL, J., ... ZONCA, A. (2020). Planck 2018 results. VI. Cosmological parameters [arXiv:1807.06209 [astro-ph]]. *A&A*, 641, A6. <https://doi.org/10.1051/0004-6361/201833910>
- PRADA, F., EREZA, J., SMITH, A., LASKER, J., VAISAKH, R., KEHOE, R., DONG-PÁEZ, C. A., SIUDEK, M., WANG, M. S., ALAM, S., BEUTLER, F., BIANCHI, D., COLE, S., DEY, B., KIRKBY, D., NORBERG, P., AGUILAR, J., AHLEN, S., BROOKS, D., ... ZHOU, Z. (2023, juin). The DESI One-Percent Survey : Modelling the clustering and halo occupation of all four DESI tracers with Uchuu [arXiv:2306.06315 [astro-ph]]. <https://doi.org/10.48550/arXiv.2306.06315>
- PRIDEAUX-GHEE, J., LECLERCQ, F., LAVAUX, G., HEAVENS, A., & JASCHE, J. (2023). Field-based physical inference from peculiar velocity tracers [Publisher : OUP ADS Bibcode : 2023MNRAS.518.4191P]. *Monthly Notices of the Royal Astronomical Society*, 518, 4191-4213. <https://doi.org/10.1093/mnras/stac3346>
- PRITCHET, C. J., & SNLS COLLABORATION. (2004, juin). SNLS – the Supernova Legacy Survey [arXiv:astro-ph/0406242]. <https://doi.org/10.48550/arXiv.astro-ph/0406242>
- QIN, F., HOWLETT, C., & STAVELEY-SMITH, L. (2019). The Redshift Space Momentum Power Spectrum II : measuring the growth rate from the combined 2MTF and 6dFGSv surveys [arXiv:1906.02874 [astro-ph]]. *Monthly Notices of the Royal Astronomical Society*, 487(4), 5235-5247. <https://doi.org/10.1093/mnras/stz1576>
- RIESS, A. G., FILIPPENKO, A. V., CHALLIS, P., CLOCCHIATTIA, A., DIERCKS, A., GARNAVICH, P. M., GILLILAND, R. L., HOGAN, C. J., JHA, S., KIRSHNER, R. P., LEIBUNDGUT, B., PHILLIPS, M. M., REISS, D., SCHMIDT, B. P., SCHOMMER, R. A., SMITH, R. C., SPYROMILIO, J., STUBBS, C., SUNTZEFF, N. B., & TONRY, J. (1998). Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant [arXiv:astro-ph/9805201]. *The Astronomical Journal*, 116(3), 1009-1038. <https://doi.org/10.1086/300499>
- RIESS, A. G., YUAN, W., MACRI, L. M., SCOLNIC, D., BROUT, D., CASERTANO, S., JONES, D. O., MURAKAMI, Y., BREUVAL, L., BRINK, T. G., FILIPPENKO, A. V., HOFFMANN, S., JHA, S. W., KENWORTHY, W. D., ANAND, G., MACKENTY, J., STAHL, B. E., & ZHENG, W. (2022). A Comprehensive Measurement of the Local Value of the Hubble Constant



- with 1 km/s/Mpc Uncertainty from the Hubble Space Telescope and the SH0ES Team [arXiv:2112.04510 [astro-ph]]. *ApJL*, 934(1), L7. <https://doi.org/10.3847/2041-8213/ac5c5b>
- RIGAULT, M., SMITH, M., REGNAULT, N., KENWORTHY, D. W., MAGUIRE, K., GOOBAR, A., DIMITRIADIS, G., AMENOCHE, M., AUBERT, M., BARJOU-DELAYRE, C., BELL, C. E., BURGAS, U., CARRERES, B., COPIN, Y., DECKERS, M., de JAEGER, T., DHAWAN, S., FEINSTEIN, F., FOCHEZ, D., ... TOWNSEND, A. (2024, juin). ZTF SN Ia DR2 : Study of Type Ia Supernova lightcurve fits [arXiv:2406.02073 [astro-ph]]. <https://doi.org/10.48550/arXiv.2406.02073>
- SAID, K., COLLESS, M., MAGOULAS, C., LUCEY, J. R., & HUDSON, M. J. (2020). Joint analysis of 6dFGS and SDSS peculiar velocities for the growth rate of cosmic structure and tests of gravity [Publisher : OUP ADS Bibcode : 2020MNRAS.497.1275S]. *Monthly Notices of the Royal Astronomical Society*, 497, 1275-1293. <https://doi.org/10.1093/mnras/staa2032>
- SARABIPOUR, S., KHAN, A., SEAH, Y. F. S., MWAKILILI, A. D., MUMOKI, F. N., SÁEZ, P. J., SCHWESSINGER, B., DEBAT, H. J., & MESTROVIC, T. (2021). Changing scientific meetings for the better [Publisher : Nature Publishing Group]. *Nat Hum Behav*, 5(3), 296-300. <https://doi.org/10.1038/s41562-021-01067-y>
- STAHL, B. E., de JAEGER, T., BORUAH, S. S., ZHENG, W., FILIPPENKO, A. V., & HUDSON, M. J. (2021). Peculiar-velocity cosmology with Types Ia and II supernovae [arXiv:2105.05185 [astro-ph]]. *Monthly Notices of the Royal Astronomical Society*, 505(2), 2349-2360. <https://doi.org/10.1093/mnras/stab1446>
- THE LSST DARK ENERGY SCIENCE COLLABORATION, MANDELBAUM, R., EIFLER, T., HLOŽEK, R., COLLETT, T., GAWISER, E., SCOLNIC, D., ALONSO, D., AWAN, H., BISWAS, R., BLAZEK, J., BURCHAT, P., CHISARI, N. E., DELL'ANTONIO, I., DIGEL, S., FRIEMAN, J., GOLDSTEIN, D. A., HOOK, I., IVEZIĆ, Ž., ... TROXEL, M. A. (2021, septembre). The LSST Dark Energy Science Collaboration (DESC) Science Requirements Document [arXiv:1809.01669 [astro-ph]]. <https://doi.org/10.48550/arXiv.1809.01669>
- TRIPP, R. (1998). A two-parameter luminosity correction for Type Ia supernovae [ADS Bibcode : 1998A&A...331..815T]. *Astronomy and Astrophysics*, 331, 815-820. Récupérée juillet 30, 2024, à partir de <https://ui.adsabs.harvard.edu/abs/1998A&A...331..815T>
- TULLY, R. B., & FISHER, J. R. (1977). A new method of determining distances to galaxies. [ADS Bibcode : 1977A&A....54..661T]. *Astronomy and Astrophysics*, 54, 661-673. Récupérée juillet 29, 2024, à partir de <https://ui.adsabs.harvard.edu/abs/1977A&A....54..661T>
- TURNER, R. J., BLAKE, C., & RUGGERI, R. (2023). A local measurement of the growth rate from peculiar velocities and galaxy clustering correlations in the 6dF Galaxy Survey [Publisher : OUP ADS Bibcode : 2023MNRAS.518.2436T]. *Monthly Notices of the Royal Astronomical Society*, 518, 2436-2452. <https://doi.org/10.1093/mnras/stac3256>