

Rapport final de stage de Recherche au LPNHE

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

Contents

1	Synthèse	5
2	Introduction	7
2.1	Contexte	7
2.1.1	Analyse du clustering des galaxies pour mesurer $f\sigma_8$	7
2.1.2	DESI DR1	8
2.1.3	Analyse cosmologique avec des SN Ia pour mesurer $f\sigma_8$	9
2.1.4	Le projet LEMAÎTRE	9
2.1.5	Analyse jointe	11
2.2	Objectifs	13
2.3	Enjeux	13
2.4	Reformulation du problème dans son contexte	13
3	Approche et analyse	15
3.1	État de l’art	15
3.1.1	Mesurer $f\sigma_8$ avec le clustering des galaxies et les SN de type Ia	15
3.2	Approche de résolution du problème	17
3.2.1	Génération des SN Ia	17
3.2.2	Reconstruction des vitesses particulières	18
3.2.3	Meilleure reconstruction	18
3.3	Résultats intermédiaires et finaux visés	19
3.3.1	Résultats intermédiaires	19
3.3.2	Résultats finaux	19
3.4	Difficultés rencontrées	19
3.5	Gestion des risques réalisée	20
4	Résultats	21
4.1	Validation de LEMAITRE sur simulations	21
4.1.1	Génération de SNe à partir d’Uchuu	21
4.1.2	Filtrage par PETS	24
4.1.3	Reconstruction avec SALT2.4	26
4.1.4	Analyse cosmologique sur les résultats SALT2.4	26
4.1.5	Reconstruction et vitesses particulières avec NaCl	26
4.2	Problèmes identifiés dans le pipeline LEMAÎTRE et corrections apportées	26
4.2.1	Problèmes liés à des valeurs de méta-paramètres	26
4.2.2	Problèmes liés aux méthodes	27
4.3	Extraction des vitesses particulières	29

5	Travail futur	31
5.1	Finalisation du travail effectué lors du stage	31
5.2	Postérité pour l'entreprise	31
6	Bilan personnel	33
7	Enjeux éthiques	35
7.1	Éthique de la recherche	35
7.2	Égalité F/H	35
7.3	Impact environnemental	35
A	Annexes	i
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 $f\sigma_8$	iv

Chapter 1

Synthèse

Une synthèse d'une page (ou executive summary) présentant le problème posé et ses enjeux pour le client, les éléments clés de la démarche, les solutions apportées et les résultats obtenus, et ce, en français et en anglais

La cosmologie moderne repose sur un modèle nommé Λ CDM, construit à partir de la relativité générale d'Albert Einstein, et supposant 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 l'explique 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 possèdent 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 quand à lui un programme d'observation visant à détecter les phénomènes transitoires, en particulier les *supernovae* (SN) de type Ia. Ces SNe sont particulièrement intéressante puisqu'elles constituent des chandelles standards, c'est-à-dire que leurs luminosités et leurs distances sont corrélées, ce qui permet de reconstruire avec précision la vitesse de leur galaxies hôtes. Leur utilisation 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 possédant des vitesses particulières sont générées selon un modèle cosmologique fixé, et des observations sont simulées à l'aide des logs d'observations de ZTF. La chaîne d'analyse permet alors de reconstruire les paramètres de ces SNe ainsi que leurs vitesses particulières, afin d'en déduire le taux de croissance des structures.

Chapter 2

Introduction

2.1 Contexte

Le modèle standard de la cosmologie, nommé Λ 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 l'énergie noire (sous la forme d'une constante cosmologique Λ). Ces deux inconnues sont nécessaires pour décrire l'évolution de l'Univers ainsi que la croissance des structures aux grandes échelles, 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 redshift¹ $f(z)$, et l'écart-type du champ de densité dans une sphère de rayon $8h^{-1}\text{Mpc}$, σ_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 Λ 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 ², mais il existe également une tension sur σ_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 Distorsion (RSD)** (Kaiser, 1987), des distorsions liées aux transformations des coordonnées comobiles vers l'espace des redshifts. Le redshift mesuré, par exemple par spectroscopie, ne contient pas uniquement le redshift cosmologique dû à l'expansion de l'univers, mais

¹En français, *décalage vers le rouge*, effets de dilatation des longueurs d'ondes des photons entre leur émission et leur observation sur Terre du à l'expansion de l'Univers. Le terme anglais sera utilisé plutôt que sa traduction pour alléger les propos.

²les mesures directes utilisant le fond diffus cosmologique donne $H_0 \sim 67 \text{ km.s}^{-1}.\text{Mpc}^{-1}$, tandis que les mesures indirectes reposant sur des supernovae ou des phénomènes de lentillage gravitationnel tendent vers $H_0 \sim 73 \text{ km.s}^{-1}.\text{Mpc}^{-1}$.

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 σ_8 avec différentes sondes

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 σ_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, telle que la relation de Tully-Fisher pour les galaxies spirales (corrélacion entre la vitesse radiale de ses étoiles et la luminosité totale de la galaxie, Tully and Fisher, 1977) ou la méthode du Plan Fondamental (Djorgovski and 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 (Adams and Blake, 2020; Qin et al., 2019; Turner et al., 2023).

2.1.2 DESI DR1

PRESENTER DESI

Cette méthode a été utilisée à partir de la simulation Uchuu (Prada et al., 2023) à 2.1 milliards de particules, peuplée avec des galaxies du *DESI Bright Galaxy Survey* (BGS) qui est un set complet de galaxies lumineuses ($r < 19.5$ pour le sous-échantillon *Bright*) à bas redshift ($z < 0.4$) (Hahn et al., 2023), pour produire un univers statistiquement fidèle au notre. Le catalogue obtenu reproduit donc à la fois le champ de densité et le champ de vitesse, à partir duquel des SNe avec vitesses particulières peuvent être tirées.

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 and Fowler, 1960). Ces SNe sont des explosions thermonucléaire de naine blanche 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 5 milliards de 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*³ 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, définies par $M = \log_{10}(\{)$. La standardisation des SN Ia est alors décrite par la formule de Tripp (Tripp, 1998)

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

où $M_{b,SN}^*$ est la magnitude absolue⁴ en bande B, $x_{1,SN}$ est le *stretch*, c_{SN} est la couleur, et 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 σ_{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.

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 champs de vitesses à bas redshift que les galaxies, et permettent de meilleurs contraintes sur la croissance des structures (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), SuperNovae

³Comme 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.

⁴*i.e.* dans le référentiel propre

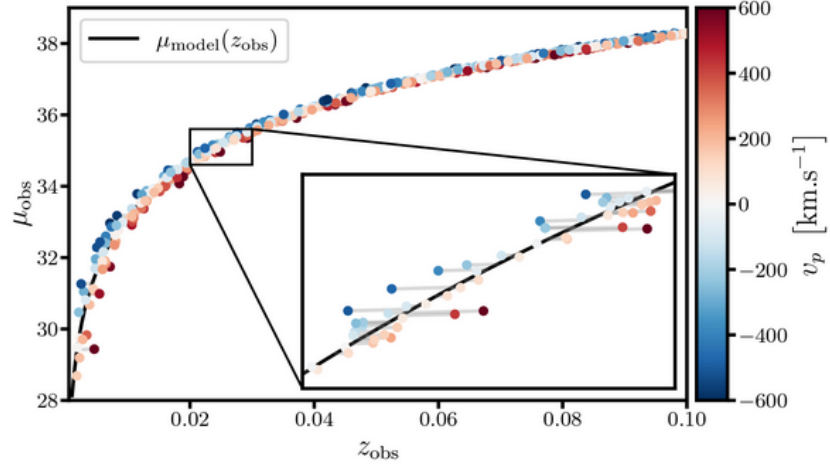


Figure 2.1: Diagramme de Hubble de SNe simulés. Les vitesses particulières selon la ligne de visée provoque un écart entre leur redshift cosmologique et le redshift observé, ce qui se traduit par un décalage horizontal des SNe par rapport à la ligne de base du diagramme de Hubble. Les SNe ayant une vitesse 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

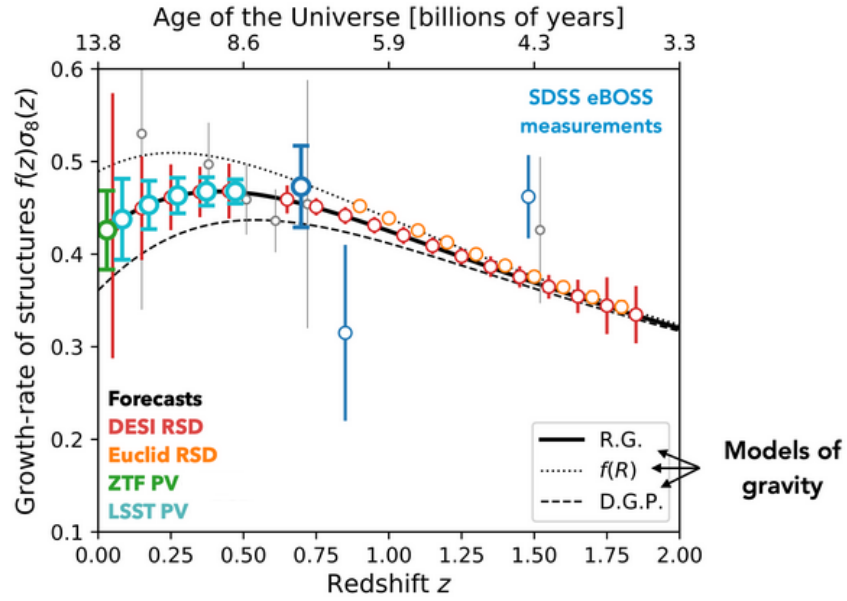


Figure 2.2: Prévisions des contraintes sur $f\sigma_8$ pour les relevés DESI (Hahn et al., 2023), Euclid (Euclid Collaboration et al., 2024) et la combinaison de DESI avec des vitesses particulières de ZTF (Carreres et al., 2023) ou LSST (Howlett et al., 2017). Credit : J. Bautista

Legacy Survey (SNLS, Pritchett and SNLS Collaboration, 2004) et Hyper Suprime-Cam Subaru Strategic Program (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). 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 LEMAÎTRE⁵ (voir également la Figure 2.3) est composé des modules suivants :

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 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 standardisations en entraînant un modèle empirique. Il vise à ré-implémenter un modèle type **SALT2.4** (adapté à l'analyse des volumes de données des relevés futurs). 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 standardisations 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.

Aux paramètres de standardisations (x_1, c) précédemment introduits, on ajoute également le paramètre d'amplitude $x_0 = 10^{0.4M_b^*}$, car la magnitude intervient sous cette forme dans le modèle **SALT2.4**. Les paramètres d'amplitude et de stretch utilisés par **NaCl** seront en revanche notés X_0 , X_1 et C , car du fait des différences dans les contraintes et du ré-entraînement (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 demeurent eux bien constants entre **SALT2.4** et **NaCl**.

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 pour obtenir des mesures de précision.

⁵Un module de calibration est présent en amont, mais n'est pas l'objet de ce stage et ne sera donc pas abordé.

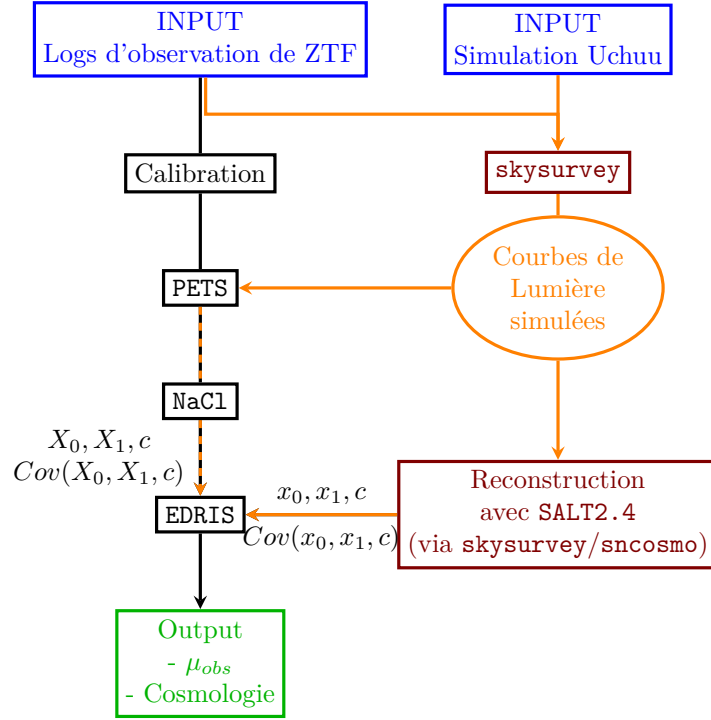


Figure 2.3: Chaîne d’analyse LEMAÎTRE développée au LPNHE (représentée en noire), et travail d’intégration effectué (en orange). L’étape de calibration contient en réalité plusieurs modules, mais elle n’a pas été utilisée pour ce stage car les données simulées n’incluaient pas d’effets de calibration.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

2.2 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 Λ 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.

Chapter 3

Approche et analyse

3.1 État de l’art

méthodes, raisons des choix faits, analyses menées (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 and 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.

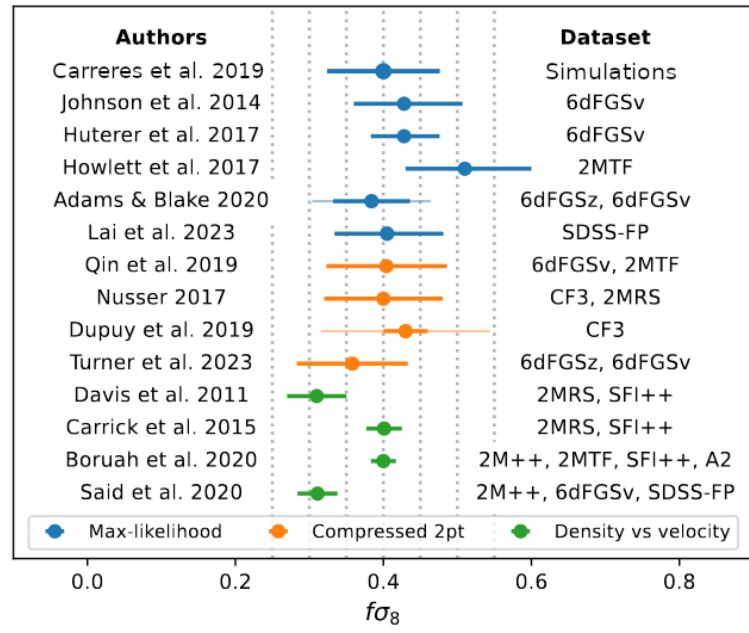


Figure 3.1: Mesures du coefficient de croissance des structures $f\sigma_8$ à partir de vitesses particulières et de catalogues de galaxies. Les barres d'erreurs en trait fin inclues les erreurs systématiques, à l'exception de Dupuy et al. 2019, pour lequel la contribution supplémentaire vient de la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

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
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.

3.2 Approche de résolution du problème

présentation détaillée de l’approche de résolution de problème retenue, des ressources scientifiques, techniques et humaines mises en oeuvre, de l’organisation du travail

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 standardisations de distributions expérimentales, puis 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`¹.

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 standardisations, à la dispersion intrinsèque et aux vitesses particulières. Ainsi, à haut redshift, les SNe les moins brillantes ne sont pas observées, ce qui biaiserait 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 ZTF Rigault 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

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

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 ΛCDM 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, λ 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 standardisations de stretch et couleur. M_0 et M_1 sont appelés les modèles et CL la loi de couleur.

Avec les résultats d'EDRIS, on peut remonter aux modules de distances

$$\mu = m_b - M_b^* \quad (3.2)$$

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 $\mu = \mu(z)$ et tout les points devraient s'aligner sur le modèle (à une dispersion Gaussienne introduite par σ_{int} près). En réalité comme mentionné en section 3.1.1, les vitesses particulières ont pour effet de décaler les redshifts observés. Les points de données sont décalés horizontalement par rapport à la courbe théorique de $\mu(z)$ (voir Fig. 2.1).

On peut alors inverser cette courbe pour obtenir une relation $z(\mu)$, c'est-à-dire obtenir le redshift auquel correspond un module de distance, puis remonter aux vitesses particulières en utilisant :

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

avec c la vitesse de la lumière².

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,

²à 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

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

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

Avec réentraînement du modèle par NaCl, et intégration

3.3 Résultats intermédiaires et finaux visés

présentation des 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 α et β . 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 branches de reconstruction 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

exposé des difficultés rencontrées et de la façon dont elles ont été surmontées

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 leurs bons fonctionnements. 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.

Les accords pour les données DESI/ZTF/BORG

3.5 Gestion des risques réalisée

gestion des risques réalisée

Temps de calcul -> optimisation, NERSC mais visée pour un ordi personnel

Disponibilité des données

Chapter 4

Résultats

présentation détaillée des résultats effectivement obtenus en les reliant au problème qui était à traiter et mettant en exergue :

- la validité des résultats par rapport au problème
- le caractère innovant / neuf des résultats
- l'utilité des résultats, l'usage effectif qui sera fait du travail de l'élève ingénieur
- la mise en exergue des limites du travail et des suites à donner à ce travail
- la quantification et/ou qualification de la valeur ajoutée

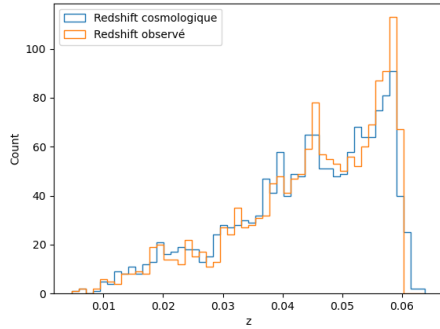
4.1 Validation de LEMAITRE sur simulations

4.1.1 Génération de SNe à partir d'Uchuu

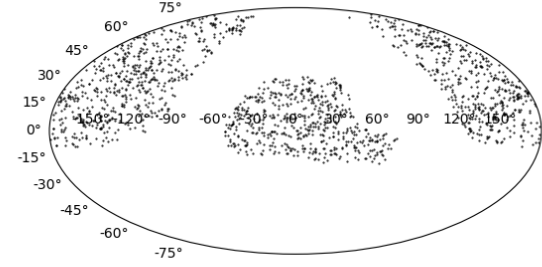
La distribution spatiale des SNe tirées est présentée en Fig. 4.1a et 4.1b. La distribution en redshift observés (incluant les vitesses particulières) diffère bien des redshifts cosmologiques, et la distribution angulaire reproduit la footprint DESI Y5, qui est similaire à la footprint ZTF.

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 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.



(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

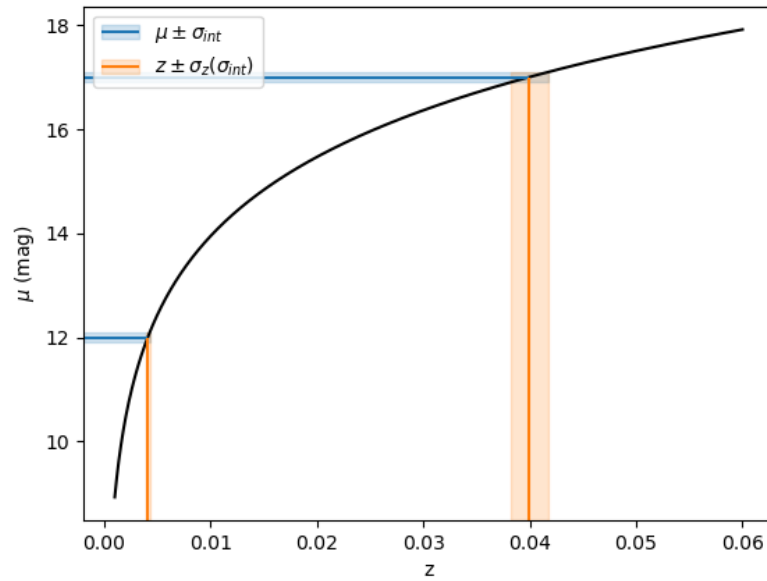


Figure 4.2: Illustration de l'erreur commise à 1σ sur les redshifts cosmologiques reconstruits due à la dispersion intrinsèque.

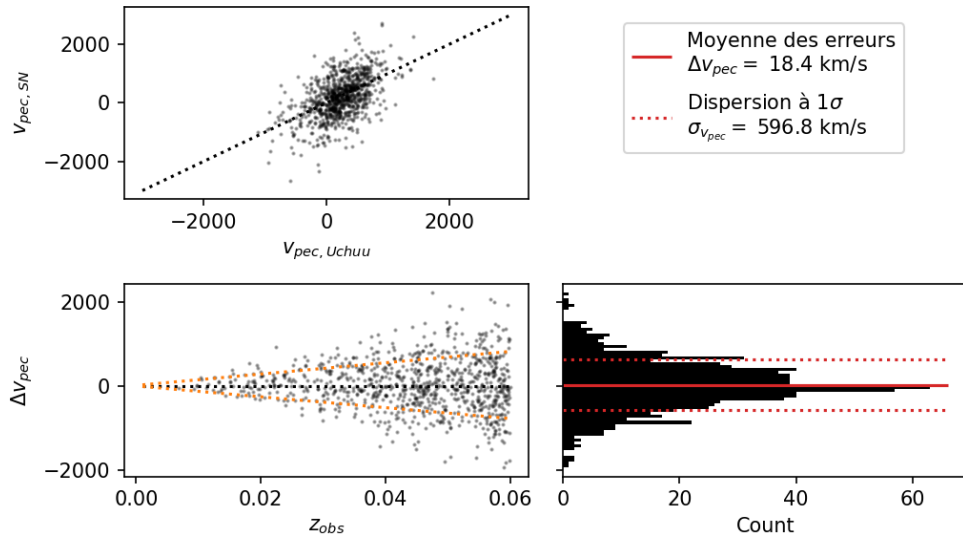


Figure 4.3: Distribution des vitesses particulières reconstruites directement à partir des SNe tirées à partir du catalogue de galaxies Uchuu. La ligne en pointillés noire (graphiques de gauche) représente la ligne de base $v_{pec,SN} = v_{pec,Uchuu}$, les lignes oranges (en bas à gauche) représente l'erreur théorique à 1σ 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.

- (a) Courbes de lumière générées à partir des paramètres tirés. Les couleurs représentent les différents filtres utilisés.
- (b) Spectre généré à partir des paramètres tirés

Figure 4.4: Exemple d'observations simulées pour les échantillons de SNe

Cut	Discarded	Remaining
Total	-	1591
SALT2.4 converge	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 8σ	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

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 ??.

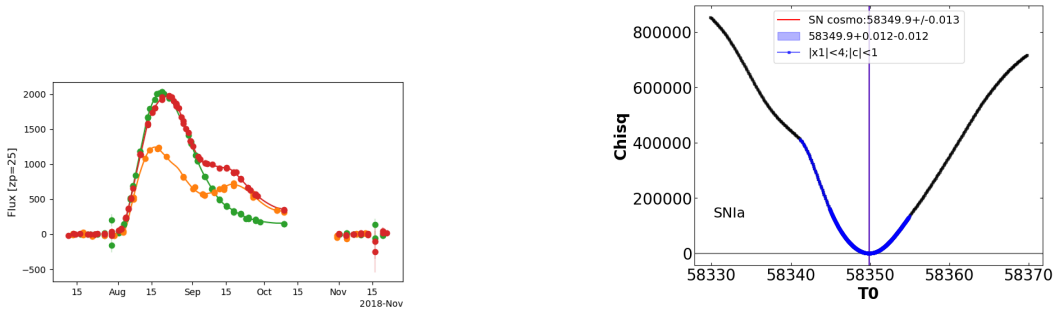
4.1.2 Filtrage par PETS

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.

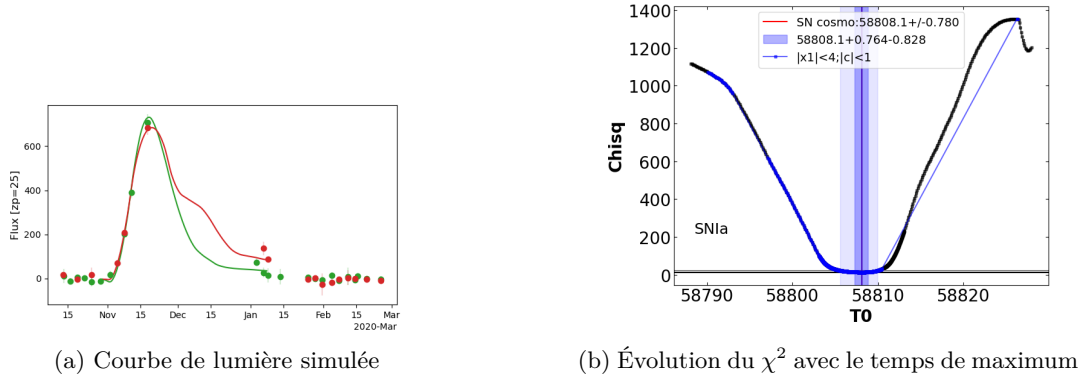
La première représente le cas d'une SNe bien échantillonnée : le minimum de la fonction $\chi^2(t_{max})$ est unique, bien défini, et présente une incertitude de 0.012 jours, ce qui est extrêmement faible. Cette précision est telle que les intervalles à 1σ et 3σ sont indiscernables du trait plein du 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

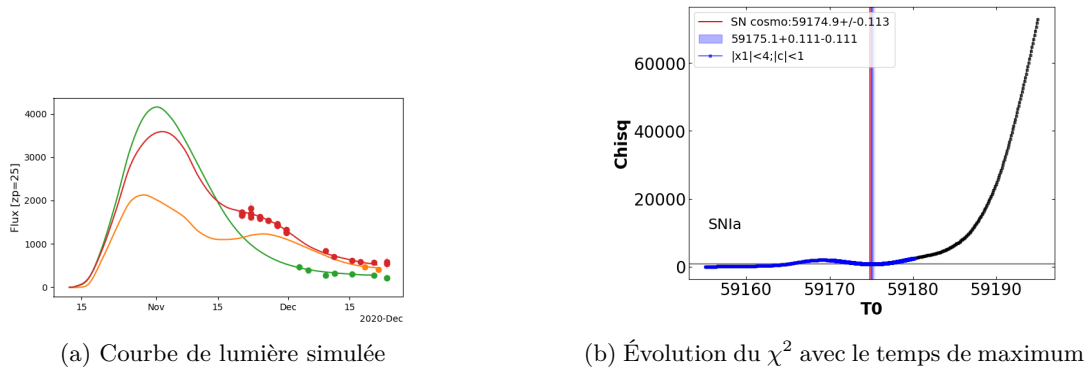


(a) Courbe de lumière simulée

(b) Évolution du χ^2 avec le temps de maximumFigure 4.5: SNe validée par PETS. Le minimum du χ^2 est unique et bien défini.



(a) Courbe de lumière simulée

(b) Évolution du χ^2 avec le temps de maximumFigure 4.6: SNe rejetée par PETS. Le minimum du χ^2 est unique mais mal défini.

(a) Courbe de lumière simulée

(b) Évolution du χ^2 avec le temps de maximumFigure 4.7: SNe anormale validée par PETS avant correction. Le minimum du χ^2 trouvé initialement est un minimum local, et l'intervalle exploré ne couvre pas la vraie valeur.

de la fonction $\chi^2(t_{max})$ est toujours unique, mais mal défini car le χ^2 varie peu autour du minimum, ce qui a pour conséquence une incertitude de $^{+0.764}_{-0.828}$ jours et une asymétrie du χ^2 . Cette aplatissement du χ^2 autour du minimum a pour conséquence majeure d'entraîner de fortes variations 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 standardisations 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 ± 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 8σ) plutôt que $\Delta\chi^2 = 9$ (respectivement 3σ).

4.1.3 Reconstruction avec SALT2.4

La première reconstruction possible une fois les courbes de lumières simulées à l'aide de `skysurvey` est d'utiliser SALT2.4. **Chaîne rapide**

4.1.4 Analyse cosmologique sur les résultats SALT2.4

4.1.5 Reconstruction et vitesses particulières avec NaCl

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 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 $t_{max,0}$ 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 3σ à 8σ , 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. A 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 point de données et ne couvrent qu'une faible fraction des phases. Elles pourraient être fittées après entraînement afin de gagner en statistiques, 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.

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.

Régularisation et contraintes dans NaCl

Un autre problème identifié dans NaCl concerne les coefficients de régularisation μ_{reg} et de contraintes μ_{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 $\mu_{reg} = 100$ et $\mu_{cons} = 10^6$ donnent une bien meilleure convergence, et ne surestiment plus les incertitudes.

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 susceptibles 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 distribution), 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.

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, qui est ici égale la hessienne

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

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 β et σ_{int}

Il existe une dégénérescence entre le coefficient β de la formule de Tripp et le paramètre de dispersion intrinsèque σ_{int} dans **EDRIS**, biais qui est particulièrement marqué lorsque l'incertitude sur la couleur σ_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 ϕ , 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**¹. Ce

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

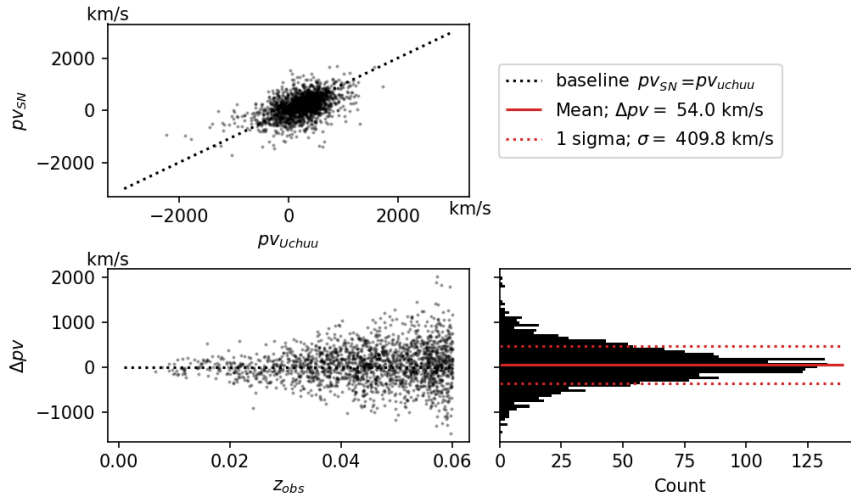


Figure 4.8: Comparaison entre les vitesses particulières injectées (dénommée pv_{uchuu}) et les vitesses particulières reconstruites (pv_{edris}). L’augmentation de l’erreur commise avec le redshift (en bas à gauche) est explicable par le fait qu’à haut redshift, l’effet des vitesses particulières est dominé par la dispersion intrinsèque. Par conséquent la dispersion statistique est à tort considérée comme une contribution des vitesses particulières.

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.

4.3 Extraction des vitesses particulières

Les vitesses particulières ainsi obtenues sont représentées en Fig. 4.8. Les vitesses reconstruites sont en très bon accord avec les vitesses injectées : l’erreur moyenne n’est que de 54 km.s^{-1} et l’écart type de l’erreur commise est de 410 km.s^{-1} , ce qui est faible sachant que les vitesses particulières valent en moyenne $\sim 200 \text{ km.s}^{-1}$ et présentent une dispersion de $\sim 300 \text{ km.s}^{-1}$. L’effet de la dispersion intrinsèque sur l’estimation des vitesses particulières est particulièrement visible à haut redshift, car la contribution des vitesses particulières au redshift observé est alors plus faible, et le module de distance évolue plus lentement avec le redshift.

Chapter 5

Travail futur

prolongements possibles du travail de stage : exposé de ce qu'il reste éventuellement à faire pour que les résultats soient complètement obtenus, et de la façon dont le travail pourra être poursuivi par l'entreprise dans le futur

5.1 Finalisation du travail effectué lors du stage

FS8 -> modélisation de la densité de Uchuu, autre simulation basée sur 2M++

5.2 Postérité pour l'entreprise

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.

Mahmoud utilise des simus semblables, intégration implémentée dans george par Jérémy, skysurvey

Chapter 6

Bilan personnel

bilan des principaux apprentissages professionnels et leur mise en perspective pour le futur professionnel

TODO

Chapter 7

Enjeux éthiques

prise de recul sur les enjeux d'ordre éthique

7.1 Éthique de la recherche

7.2 Égalité F/H

Toujours une grosse disparité, mais messages/affiches de sensibilisation.
Pas de discriminations observées lors du stage.

7.3 Impact environnemental

Sensibilisation lors du meeting DESI 2024, supercalculateurs

Appendix 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 3σ 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 réalisée. 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 3σ 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 8σ 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 λ 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 Φ_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 θ . Cette compression facilite leurs évaluations et les calculs de leurs dérivées. Les modèles sont ainsi évaluables en tout point ce qui permet de s'affranchir des effets indésirables introduits par une discrétisation (nécessitée 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 (λ, 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.

Le lecteur attentif aura noté que les amplitudes et stretch de SALT étaient notées avec des minuscules x_0 et x_1 , tandis que celles de NaCl sont notées avec des majuscules X_0 et X_1 . Cette différence est intentionnelle, puisque du fait des différences dans les contraintes et l'entraînement, rien ne garantit que ces paramètres soient exactement les mêmes, ce qui n'est d'ailleurs pas le cas dans les faits (cf [Résultats](#)).

Contraintes

En conséquence de ces choix, les contraintes imposées sur les modèles diffèrent de celles utilisées pour entraîner SALT2.4. 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 θ 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 les effets de sélections tel que le biais de Malmquist, et en estimant directement la matrice de covariance des paramètres. Il utilise pour cela la méthode d'optimisation dite *Hessian-free* décrite dans Martens, 2010. 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.

MORION

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 :

$$\mathbf{v}(\mathbf{r}) = \frac{H_0 \mathcal{B}}{4\pi} \int_0^{R_{max}} d^3 \mathbf{r}' \delta_g(\mathbf{r}') \frac{\mathbf{r}' - \mathbf{r}}{|\mathbf{r}' - \mathbf{r}|} + \mathbf{V}_{ext} \quad (\text{A.16})$$

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 Θ (notamment les paramètres

du modèle de Tripp α , β , M_b et σ_{int} , voir Eq. 2.1) sont reliés aux paramètres observables des SN $\{\mathbf{x}_i\}$ via la probabilité conditionnelle $\mathcal{P}(\mathcal{B}, \mathbf{V}_{ext}, \boldsymbol{\Theta}|\mathbf{x}_i)$, qui peut être exprimée à l'aide du théorème de Bayes (Stahl et al., 2021)

$$\mathcal{P}(\mathcal{B}, \mathbf{V}_{ext}, \boldsymbol{\Theta}|\mathbf{x}_i) \propto \mathcal{P}(\mathbf{x}_i|\mathcal{B}, \mathbf{V}_{ext}, \boldsymbol{\Theta})\mathcal{P}(\mathcal{B}, \mathbf{V}_{ext}, \boldsymbol{\Theta}) \quad (\text{A.17})$$

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

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 σ_{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

$$\mathcal{P}(\mathbf{r}|\boldsymbol{\Theta}) = \frac{1}{\mathcal{N}(\boldsymbol{\Theta})} r^2 \exp\left\{-\frac{[\mu(\mathbf{r}) - \mu(\boldsymbol{\Theta})]^2}{2\sigma_\mu^2(\boldsymbol{\Theta})}\right\} [1 + \delta_g(\mathbf{r})] \quad (\text{A.18})$$

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

$$\mathcal{P}(\mathbf{x}_i|\mathcal{B}, \mathbf{V}_{ext}, \boldsymbol{\Theta}) = \int_0^{R_{max}} dr \mathcal{P}(\mathbf{x}_i|\mathbf{r}, \mathcal{B}, \mathbf{V}_{ext}, \boldsymbol{\Theta})\mathcal{P}(\mathbf{r}|\boldsymbol{\Theta}) \quad (\text{A.19})$$

avec

$$\mathcal{P}(\mathbf{x}_i|\mathbf{r}, \mathcal{B}, \mathbf{V}_{ext}, \boldsymbol{\Theta}) = \frac{1}{\sqrt{2\pi\sigma_\nu^2}} \exp\left\{-\frac{[cz_{obs} - cz_{pred}(\mathbf{r}, \mathcal{B}, \mathbf{V}_{ext})]^2}{2\sigma_\nu^2}\right\} \quad (\text{A.20})$$

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

$$1 + z_{pred}(\mathbf{r}, \mathcal{B}, \mathbf{V}_{ext}) = (1 + z_{cosmo}(\mathbf{r})) \left[1 + \frac{1}{c}(\mathcal{B}\mathbf{v}(r) + \mathbf{V}_{ext}) \cdot \hat{\mathbf{r}}\right] \quad (\text{A.21})$$

σ_ν 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

$$\mathcal{L} = \ln \mathcal{P}(\mathcal{B}, \mathbf{V}_{ext}, \boldsymbol{\Theta}) + \sum_i \mathcal{P}(\mathbf{x}_i|\mathcal{B}, \mathbf{V}_{ext}, \boldsymbol{\Theta}) \quad (\text{A.22})$$

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.1, 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.

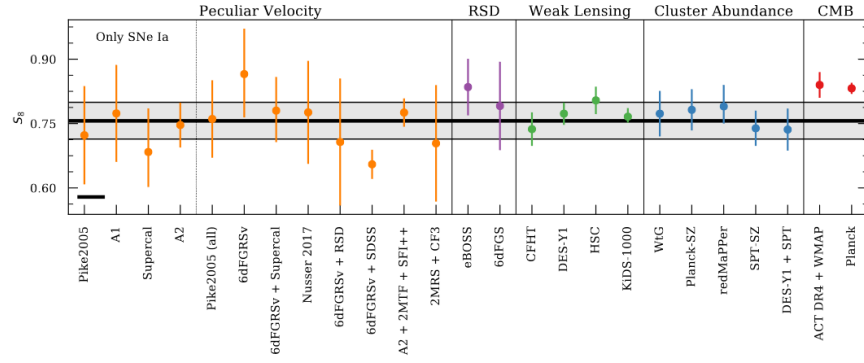


Figure A.1: Paramètre $S_8 = f\sigma_8 / (0.3)^{0.55}$ reconstruit par diverses méthodes. Les lignes noires représentent la valeur et l'incertitude de la valeur reconstruite par Stahl et al., 2021. Crédit : B. Stahl

Bibliography

- 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>
- 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/aaecbe>
- 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: H_0 measurements of the growth rate and σ_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, July). 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., Fouchez, 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, July). Not all lensing is low: An analysis of DESI's 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., Agüena, 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]]. *The OJA*, 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., Mörtzell, 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>
- 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., Baland, C., Bautista, J. E., Bershadsky, 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., Agüena, 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, May). 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., Car-

- rerer, B., Dhawan, S., Deckers, M., Feinstein, F., Fouchez, D., Galbany, L., ... van der Walt, S. (2024, June). 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., Fouchez, D., Hook, I. M., Howell, D. A., Perrett, K., Pritchett, 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: 2017MNRAS.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 Λ 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>
- 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., Uruguchi, 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., Rigault, 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., Dahiwal, A. S., Sharma, Y., Bellm, 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., Goldhaber, 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, December). 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, June). 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>

- Pritchett, C. J., & SNLS Collaboration. (2004, June). 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>
- Rigault, M., Smith, M., Regnault, N., Kenworthy, D. W., Maguire, K., Goobar, A., Dimitriadis, G., Amenouche, M., Aubert, M., Barjou-Delayre, C., Bellm, C. E., Burgaz, U., Carreres, B., Copin, Y., Deckers, M., de Jaeger, T., Dhawan, S., Feinstein, F., Fouchez, D., ... Townsend, A. (2024, June). 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>
- 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, September). 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. Retrieved July 30, 2024, from <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. Retrieved July 29, 2024, from <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>