Aix*Marseille université
Socialement engagée

NNT : 2024AIXM0555

THÈSE DE DOCTORAT

Soutenue à Aix-Marseille Université
le 10 décembre 2024 par

Arseniy Kuznetsov

Towards the next generation of tomographic AO-assisted instruments : Self-learning techniques for system optimization & science exploitation

Discipline
Physique et Sciences de la Matière

Spécialité
Instrumentation

École doctorale
ED 352 / PHYSIQUE ET SCIENCES DE LA MATIÈRE

Laboratoire/Partenaires de recherche
European Southern Observatory (ESO)

Laboratoire d'Astrophysique de Marseille / Groupe R&D Optique et Instrumentation

HRA / Département d'Optique et Techniques Associées (DOTA) de l'ONERA

Composition du jury

Laurent JOLISSAINT Professeur – HES-SO	Rapporteur
Jean-Pierre VÉRAN Professeur adjoint – UVic	Rapporteur
Joël VERNET Astronome – ESO	Examinateur
David MARY Professeur – Université Côte d'Azur	Examinateur
Jessica LU Professeure associé – UC Berkley	Examinateurice
Annie ZAVAGNO Professeure – Aix Marseille Université	Présidente du jury
Benoît NEICHEL Chargé de recherche – CNRS	Directeur de thèse
Thierry FUSCO Directeur de recherche – ONERA	Codirecteur de thèse
Sylvain OBERTI Ingénieur de recherche – ESO	Membre invité

ESO

LAM LABORATOIRE D'ASTROPHYSIQUE DE MARSEILLE

ONERA THE FRENCH AEROSPACE LAB

Affidavit

I undersigned, Arseniy Kuznetsov, hereby declare that the work presented in this manuscript is my own work, carried out under the scientific supervision of Benoît Neichel, Thierry Fusco, and Sylvain Oberti, in accordance with the principles of honesty, integrity, and responsibility inherent to the research mission. The research work and the writing of this manuscript have been carried out in compliance with both the French national Charter for Research Integrity and the Aix-Marseille University Charter on the fight against plagiarism.

This work has not been submitted previously either in this country or in another country in the same or in a similar version to any other examination body.

Garching by München, 8th of October 2024.



This work is licensed under Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International Public License

Liste de publications et participation aux conférences

Liste des publications et/ou brevets réalisées dans le cadre du projet de thèse :

1. **Arseniy Kuznetsov, Benoît Neichel, Sylvain Oberti, Thierry Fusco, Olivier Beltramo-Martin**, "Prediction of AO corrected PSF for AOF NFM/SPHERE," *Proc. SPIE 12185, Adaptive Optics Systems VIII*, 121853I (August 2022) ; https://doi.org/10.1117/12.2628616
2. **Arseniy Kuznetsov, Sylvain Oberti, Cedric Taïssir Heritier, Cédric Plantet, Benoît Neichel, Thierry Fusco, Stefan Ströbele, Carlos Correia**, "Study of the LIFT focal-plane wavefront sensor for GALACSI NFM," *Proc. SPIE 12185, Adaptive Optics Systems VIII*, 1218531 (August 2022) ; https://doi.org/10.1117/12.2628588
3. **Guido Agapito, Cedric Plantet, Fabio Rossi, Giulia Carlà, Anne-Laure Cheffot, Daniele Vassallo, Arseniy Kuznetsov, Simon Conseil, Benoit Neichel**, "TIP-TOP : cone effect for single laser adaptive optics systems," *Adaptive Optics for Extremely Large Telescopes 7th Edition*, Avignon, France, (June 2023) ; https://doi.org/10.13009/AO4ELT7-2023-012
4. **Arseniy Kuznetsov, Benoit Neichel, Sylvain Oberti, Thierry Fusco**, "Prediction of AO corrected PSF for SPHERE / AOF NFM," *Adaptive Optics for Extremely Large Telescopes 7th Edition*, Avignon, France, (June 2023) ; https://doi.org/10.13009/AO4ELT7-2023-106
5. **Arseniy Kuznetsov, Sylvain Oberti, Benoit Neichel, Thierry Fusco**, "Striving towards robust phase diversity on-sky. Implementing LIFT for VLT/MUSE-NFM," *Astronomy & Astrophysics*, Vol. 687, A221 (July 2024) ; https://doi.org/10.1051/0004-6361/202449860
6. **Benoit Neichel, Guido Agapito, Arseniy Kuznetsov, Fabio Rossi, Cédric Plantet, Carlo F. Manara, Romain Fétick, Alice Concas, Joël Vernet, Olivier Hainaut, Anne-Laure Cheffot, Giulia Carlà, Jean-François Sauvage, Michele Cirasuolo, Paolo Padovani, Carlos Correia, Cedric Taïssir Héritier, Thierry Fusco**, "Tip-Top : toward a single tool for all ELT instrument's PSF prediction," *Proc. SPIE 13097, Adaptive Optics Systems IX*, 130972Y (August 2024) ; https://doi.org/10.1117/12.3015061

Participation aux conférences et écoles d'été au cours de la période de thèse :

1. **Talk** : "Extending the limiting magnitude with LIFT for GALASCI NFM", *Wavefront Sensing in the VLT/ELT Era VI Workshop 2021*
2. **Talk** : "Study of the LIFT focal-plane wavefront sensor for GALACSI NFM", *SPIE 2022 Astronomical Telescopes + Instrumentation* (presented by Benoît Neichel)
3. **Poster** : "Prediction of AO corrected PSF for AOF NFM/SPHERE", *SPIE 2022 Astronomical Telescopes + Instrumentation* (presented by Benoît Neichel)
4. **Talk** : "Study of the LIFT focal-plane wavefront sensor for GALACSI NFM", *Wavefront Sensing in the VLT/ELT Era VII Workshop 2022*
5. **Talk** : "Prediction of AO corrected PSF for SPHERE / AOF NFM", *AO4ELT7 conference 2023*
6. **Poster** : "Study of the LIFT focal-plane wavefront sensor for GALACSI NFM", *AO4ELT7 conference 2023*

Abstract

Several key areas of modern astronomy, such as studies of galaxies and the Galactic Center, exoplanet detection and characterization, imaging of the bodies within the solar system, and more, rely heavily on adaptive optics (AO) correction. These research fields demand the highest possible angular resolution, along with extremely precise astrometry and photometry. In this context, AO significantly enhances the angular resolution achievable with modern large ground-based observatories by compensating for atmospheric turbulence, thus pushing the attainable resolution towards the diffraction limit. For the upcoming extremely large telescopes, such as ESO's Extremely Large Telescope (ELT), AO systems will play a critical role, assisting nearly all observations.

The quality of scientific results is intrinsically linked to the quality of the images produced, which, in turn, depends on the point spread function (PSF). Therefore, precise knowledge of the PSF is essential for interpreting scientific data and enabling advanced post-processing techniques, such as deconvolution.

However, using AO introduces a complex structure to the PSF in the scientific focal plane, making accurate modeling challenging, especially when real PSFs are unavailable for reference within the scientific field (e.g. when observing extragalactic targets).

This thesis proposes a method for fast and realistic modeling of AO-corrected PSFs using a hybrid approach that combines an analytical model with the machine-learning-based component. This approach learns the complex behavior of the telescope and AO system from the data they produce, thereby improving the accuracy of PSF modeling. The data-driven model is trained on the fused data inputs collected from multiple sources, including reduced AO telemetry, site-monitoring data, and scientific / wavefront sensor PSFs. The method is tested on two ESO Very Large Telescope (VLT) facility instruments: SPHERE and MUSE.

Because PSF accumulates the effects of all possible contributions along the optical imaging path, it can be used to diagnose and characterize these aberrations. The second part of this thesis addresses this problem, complementing the conclusions and findings presented in the other part of the manuscript.

Keywords: adaptive optics, PSF, machine learning, PSF modeling, PSF prediction, instrumentation for large ground-based telescopes.

Résumé

Plusieurs domaines clés de l'astronomie moderne, tels que l'étude des galaxies et du centre galactique, la détection et la caractérisation des exoplanètes, l'imagerie des corps du système solaire, et bien d'autres, dépendent fortement de la correction par optique adaptative (OA). Ces domaines de recherche exigent la plus haute résolution angulaire possible, ainsi qu'une astrométrie et une photométrie extrêmement précises. Dans ce contexte, l'OA améliore considérablement la résolution angulaire des grands observatoires terrestres modernes en compensant la turbulence atmosphérique, ce qui permet d'atteindre une résolution proche de la limite de diffraction. Pour les prochains télescopes de très grande taille, tels que l'Extremely Large Telescope (ELT) de l'ESO, les systèmes d'OA joueront un rôle essentiel en assistant presque toutes les observations.

La qualité des résultats scientifiques est intrinsèquement liée à la qualité des images produites, qui, à son tour, dépend de la fonction d'étalement du point (PSF). Par conséquent, une connaissance précise de la fonction d'étalement du point est essentielle pour interpréter les données scientifiques et permettre des techniques avancées de post-traitement, telles que la déconvolution.

Cependant, l'utilisation de l'OA introduit une structure complexe de la PSF dans le plan focal scientifique, ce qui rend une modélisation précise difficile, en particulier lorsque les PSF réelles ne sont pas disponibles pour référence dans le champ scientifique (par exemple, lors de l'observation de cibles extragalactiques).

Cette thèse propose une méthode de modélisation rapide et réaliste des PSF corrigées par l'OA en utilisant une approche hybride qui combine un modèle analytique avec un composant basé sur l'apprentissage automatique. Cette approche apprend le comportement complexe du télescope et du système d'OA à partir des données qu'ils produisent, améliorant ainsi la précision de la modélisation de la PSF. Le modèle basé sur les données est entraîné sur les entrées de données fusionnées collectées à partir de sources multiples, y compris la télémétrie réduite de l'OA, les données de surveillance du site et les PSF des capteurs scientifiques / de front d'onde. La méthode est testée sur deux instruments du Very Large Telescope (VLT) de l'ESO : SPHERE et MUSE.

Parce que la PSF accumule les effets de toutes les contributions possibles le long du chemin d'imagerie optique, elle peut être utilisée pour diagnostiquer et caractériser ces aberrations. La deuxième partie de cette thèse aborde ce problème, en complétant les conclusions et les résultats présentés dans la première partie.

Mots clés : optique adaptative, PSF, apprentissage automatique, modélisation de PSF, prédiction de PSF, instrumentation pour les grands télescopes au sol.

Résumé en français

L'optique adaptative dans la science

Plusieurs domaines de recherche en astronomie moderne exigent la plus haute résolution angulaire possible, ainsi qu'une astrométrie et une photométrie extrêmement précises. Dans ce contexte, l'Optique Adaptive (OA) améliore considérablement la résolution angulaire que l'on peut atteindre avec les grands observatoires terrestres modernes en compensant la turbulence atmosphérique, poussant ainsi la résolution atteignable vers la limite de diffraction. Pour les prochains télescopes de très grande taille, tels que l'Extremely Large Telescope (ELT) de l'ESO, les systèmes d'OA joueront un rôle essentiel, en assistant presque toutes les observations :

— Étude des galaxies ;
— Étude sur les centres galactiques ;
— Détection et caractérisation des exoplanètes ;
— Imagerie des corps du système solaire ;

Le lien entre l'OA, la qualité des données scientifiques et la PSF

La qualité des résultats scientifiques est intrinsèquement liée à la qualité des images produites, qui dépend à son tour de la fonction d'étalement du point (PSF). La fonction d'étalement du point représente la façon dont les systèmes optiques voient un objet ponctuel, par exemple une étoile. Par conséquent, une fonction d'étalement du point moins nette se traduit par des images scientifiques plus ambiguës. Une connaissance précise de la PSF est donc essentielle pour interpréter les données scientifiques. Une bonne connaissance de la PSF peut également permettre l'utilisation de techniques avancées de post-traitement, telles que la déconvolution.

Cependant, l'utilisation de l'OA introduit une structure complexe de la PSF dans le plan focal scientifique, ce qui rend une modélisation précise difficile, en particulier lorsque des PSF réelles ne sont pas disponibles pour référence dans le champ de vue scientifique (par exemple lors de l'observation de cibles extragalactiques).

Modélisation PSF basée sur les données

Cette thèse introduit une nouvelle méthode pour la modélisation rapide et réaliste des PSF corrigées par OA en utilisant une approche hybride qui intègre un modèle de PSF analytique basé sur la méthode de Fourier (Rigaut et al. (1998); Jolissaint (2010))

avec un composant basé sur l'apprentissage automatique (Figure 1). En incorporant un module basé sur les données, le modèle hybride peut apprendre le comportement complexe du télescope et du système d'OA à partir des données observées, en affinant les prédictions du modèle analytique pour mieux correspondre aux conditions du monde réelles.

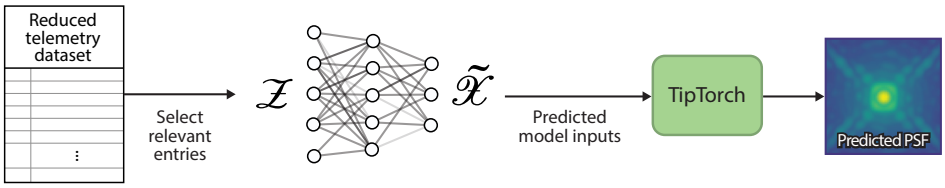
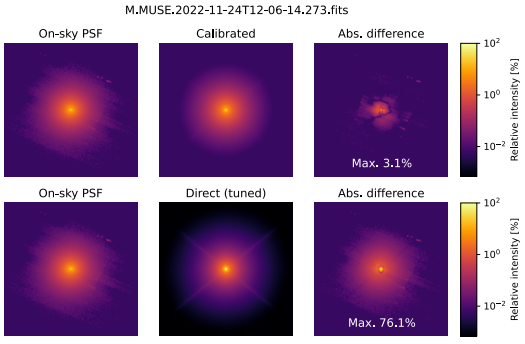
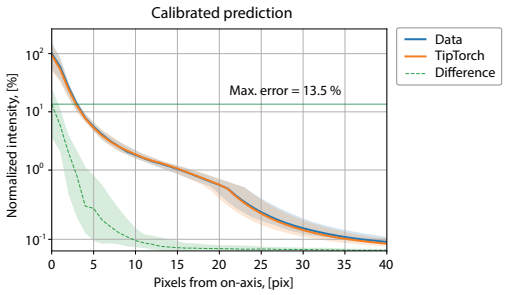


FIGURE 1. – Représentation schématique de l'approche de modélisation de PSF proposée. Les données fusionnées sont représentées par l'ensemble de données "reduced telemetry", tandis que le modèle de PSF développé est désigné par "TipTorch".

Le modèle est entraîné à l'aide de données d'entrée fusionnées obtenues à partir de sources multiples, y compris telles que la télémétrie réduite de l'OA, les données de surveillance du site, et les PSF scientifiques et du capteur de front d'onde. La méthode a été validée sur deux instruments du Very Large Telescope (VLT) de l'ESO : SPHERE et MUSE. Les résultats pour MUSE en mode champ étroit sont présentés dans les Figures 2(a) et (b).



(a)



(b)

FIGURE 2. – (a) : Exemple de la PSF prédite avec l'approche proposée (rangée du haut) par rapport à l'approche conventionnelle entièrement analytique (rangée du bas). (b) : Comparaison des profils radiaux prédits à l'aide des techniques proposées avec les profils de PSF correspondants sur le ciel.

Ces figures comparent les performances du modèle de PSF hybride proposé (étiqueté "calibrated") à celles d'un modèle de PSF analytique classique (étiqueté "direct") dépourvu de toute composante d'apprentissage automatique. Les deux modèles sont évalués à l'aide de données d'observation sur le ciel. La figure de gauche montre un

exemple spécifique de PSF scientifique prédite, tandis que les profils radiaux dans la figure de droite présentent les résultats de la modélisation pour un ensemble de données de 78 PSFs. Comme nous l'avons démontré, l'approche hybride permet d'obtenir une précision nettement supérieure dans la modélisation des PSF par rapport au modèle analytique traditionnel, en s'alignant étroitement sur les données observées.

Utilisation de la PSF pour détecter les aberrations du front d'onde

Une autre propriété intéressante de la PSF est qu'elle intègre les effets de toutes les aberrations optiques le long du trajet d'imagerie, ce qui en fait un outil précieux pour diagnostiquer les erreurs quasi-statiques dans le système optique. Certaines de ces erreurs peuvent se manifester sous la forme d'aberrations de front d'onde quasi-statiques qui dégradent la netteté de la PSF et, par conséquent, réduisent la qualité des images scientifiques. Le processus d'extraction de ces aberrations du front d'onde à partir de la morphologie de la PSF du plan focal est connu sous le nom de détection du front d'onde du plan focal.

Cette thèse explore la détection du front d'onde dans le plan focal en utilisant la technique du plan focal linéarisé (LIFT). Ce manuscrit propose également des améliorations à cette méthode pour permettre une récupération de phase plus robuste et plus précise dans les opérations réelles. Les résultats sont illustrés à la figure 3, où l'approche proposée a permis de reconstruire avec succès des cartes d'aberration du front d'onde cohérentes à partir de plusieurs PSF enregistrées sur le ciel à différents angles de rotation du télescope.

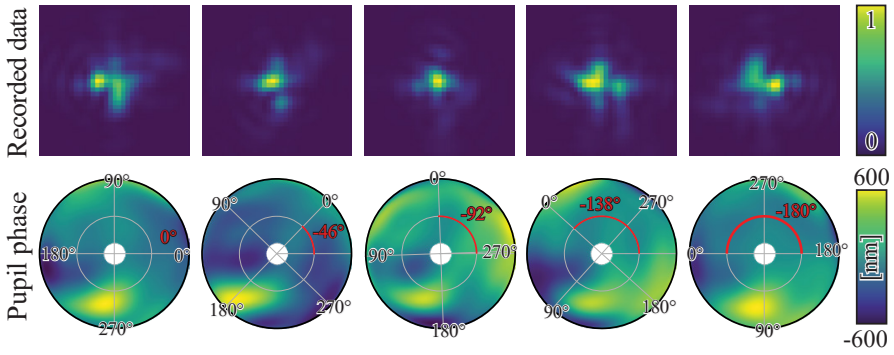


FIGURE 3. – Résultats de l'estimation des aberrations de chemin non-commun (NCPA) sur le ciel en utilisant la méthode d'estimation proposée pour différentes positions du dérotateur sur l'UT4 de l'ESO. Dans toutes les positions, la méthode permet d'obtenir des cartes NCPA similaires.

En outre, la technique proposée peut être utilisée pour l'étalonnage du système, ce qui permet une récupération plus fiable des aberrations de basse fréquence spatiale (de bas ordre) à l'aide de la méthode LIFT (Figure 4).

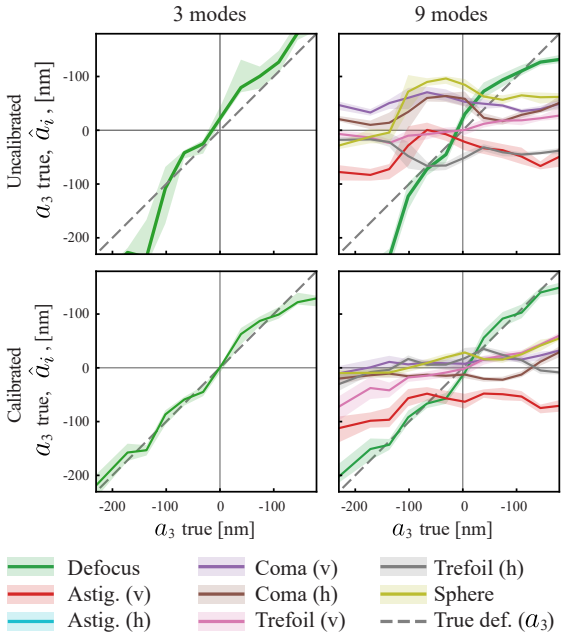


FIGURE 4. – La rangée du haut montre la récupération des modes de bas ordre avec LIFT lorsqu'aucun étalonnage n'est appliqué. Le mode de défocalisation est balayé linéairement ici, tandis que les autres aberrations restent fixes. La rangée du bas montre l'estimation lorsque le LIFT est précalibré avec la technique proposée. Dans ce dernier cas, la précision est bien meilleure.

En résumé, la méthode LIFT améliorée présentée dans ce travail sert d'outil d'étalonnage efficace sur le ciel, complétant le système actuel de contrôle de l'OA mis en œuvre sur MUSE. Les résultats de cette étude sont documentés dans la publication suivante (Kuznetsov et al. (2024)).

En résumé, ce doctorat s'est concentré sur l'utilisation plus efficace des données générées par les télescopes assistés par OA, avec un accent particulier sur les PSF. Les algorithmes d'estimation et de prédiction basés sur les PSF développés dans le cadre de ce travail ont permis l'extraction des aberrations quasi-statiques de la forme des PSF et ont permis la prédiction précise de la morphologie des PSF sur la base de la télémétrie OA réelle sur le ciel. Ce travail ouvre également la voie à de nouvelles idées concernant l'extension des approches proposées aux nouveaux instruments et leur amélioration par l'incorporation de nouvelles données télémétriques sur le ciel, ce qui les rendra plus précises et applicables à un plus large éventail d'applications scientifiques.

Mots clés : optique adaptative, PSF, apprentissage automatique, modélisation de PSF, prédiction de PSF, instrumentation pour les grands télescopes au sol.

Table des matières

Affidavit 2

Liste de publications et participation aux conférences 3

Abstract 5

Résumé 6

Résumé en français 7

Table des matières 11

Table des figures 14

Liste des tableaux 19

1. Introduction 20
1.1. Motivation, objectives, and methods . . . 22
1.2. The structure of this manuscript . . . 26

2. General background and theory 28
2.1. Basics of image formation . . . 28
2.1.1. Optical imaging . . . 28
2.1.2. Point spread function . . . 31
2.1.3. Angular resolution . . . 32
2.1.4. Image formation . . . 33
2.2. Atmospheric turbulence . . . 35
2.2.1. Structure function . . . 35
2.2.2. Fried's parameters (r_0) . . . 37
2.2.3. Temporal behavior of the turbulence (τ_0) . . . 38
2.2.4. Atmospheric scintillation . . . 39
2.2.5. Angular anisoplanatism . . . 39
2.2.6. Cone effect (focal anisoplanatism) . . . 41
2.2.7. Turbulent PSD . . . 42
2.2.8. From PSD to PSF . . . 42
2.2.9. Seeing-limited PSF . . . 43
2.2.10. Strehl ratio . . . 44
2.3. Adaptive optics . . . 45
2.3.1. Basic scheme . . . 45

2.3.2. AO correction strategies . . . 46
2.3.3. Main AO system components . . . 46
2.3.4. Types of AO correction . . . 54
2.4. Instruments . . . 60
2.4.1. SPHERE . . . 60
2.4.2. MUSE . . . 63
2.5. Error sources . . . 67
2.5.1. AO-related errors . . . 67
2.5.2. Telescope/instrument-related errors . . . 68
2.5.3. Optical Errors . . . 72
2.5.4. NCPAs . . . 74

3. Focal-plane wavefront sensing **77**
3.1. Striving towards robust phase diversity on-sky Implementing LIFT for VLT/MUSE-NFM . . . 77
3.2. Modal cross-talk . . . 92
3.2.1. Selection of the diversity phase amplitude . . . 96
3.2.2. The effect of low flux . . . 99
3.3. Negative log-likelihood and loss function . . . 100
3.3.1. MSE vs. MAE . . . 103
3.4. Non-linear optimization . . . 104
3.4.1. First-order methods : Gradient descent (GD) . . . 104
3.4.2. Second-order methods : Newton's method . . . 105
3.4.3. Quasi-Newton Methods . . . 106
3.5. From focal-plane WFSing to PSF prediction . . . 108

4. PSF reconstruction and prediction **110**
4.1. Motivation for PSF-R . . . 110
4.2. Approaches to PSF modeling . . . 113
4.2.1. Classical PSF-R . . . 114
4.2.2. Fourier-based modeling . . . 114
4.2.3. Comparing both approaches . . . 115
4.3. Theory behind Fourier-based PSF models . . . 117
4.3.1. Error due to imperfect AO-correction . . . 121
4.4. PSF model implementations . . . 133
4.4.1. Code implementations . . . 133
4.4.2. Initialization and usage . . . 134
4.5. Data . . . 135
4.5.1. Data types . . . 135
4.5.2. Science data . . . 138
4.5.3. Meteorological data . . . 139
4.5.4. Data lake . . . 140
4.5.5. Collecting and processing the data . . . 141
4.5.6. Why not PSF-R? . . . 148

4.6. Direct PSF prediction . . . 148
4.6.1. SPHERE results . . . 149
4.6.2. MUSE-NFM results . . . 150
4.6.3. Conclusion . . . 152
4.7. PSF model validation . . . 153
4.7.1. SPHERE . . . 154
4.7.2. MUSE . . . 159
4.7.3. Conclusions . . . 163
4.8. Calibrated PSF prediction . . . 164
4.8.1. The concept of model calibration . . . 164
4.8.2. Approaches to model calibration . . . 165
4.8.3. Training a data-driven calibrator . . . 168
4.8.4. Final prediction results for SPHERE . . . 169
4.8.5. Final prediction results for MUSE . . . 178
4.8.6. Conclusion . . . 184

5. Conclusion and Future Perspectives **186**
5.1. Focal-Plane wavefront sensing . . . 186
5.1.1. Conclusions . . . 186
5.1.2. Future perspectives . . . 187
5.2. PSF prediction . . . 188
5.2.1. Conclusion . . . 188
5.2.2. Future perspectives . . . 189

Remerciements **204**

Annexes **206**
A. Coupling between modes in LIFT . . . 206
B. PSF models . . . 207
B.1. Code . . . 207
B.2. Model initialization . . . 208
C. Data processing . . . 209
C.1. IRDIS . . . 209
C.2. Downloading IRDIS exposure cubes . . . 209
C.3. MUSE-NFM . . . 214
D. NN-calibrator architecture . . . 217
E. Computing PSF profiles . . . 217
F. Tuned direct prediction values . . . 219
F.1. IRDIS . . . 219
F.2. MUSE . . . 219
G. Distributions of calibrated model inputs . . . 220
G.1. SPHERE . . . 220
G.2. MUSE . . . 220

Table des figures

1. Représentation schématique de l'approche de modélisation de PSF proposée. Les données fusionnées sont représentées par l'ensemble de données "reduced telemetry", tandis que le modèle de PSF développé est désigné par "TipTorch". 8
2. (a) : Exemple de la PSF prédite avec l'approche proposée (rangée du haut) par rapport à l'approche conventionnelle entièrement analytique (rangée du bas). (b) : Comparaison des profils radiaux prédits à l'aide des techniques proposées avec les profils de PSF correspondants sur le ciel. 8
3. Résultats de l'estimation des aberrations de chemin non-commun (NCPA) sur le ciel en utilisant la méthode d'estimation proposée pour différentes positions du dérotateur sur l'UT4 de l'ESO. Dans toutes les positions, la méthode permet d'obtenir des cartes NCPA similaires. 9
4. La rangée du haut montre la récupération des modes de bas ordre avec LIFT lorsqu'aucun étalonnage n'est appliqué. Le mode de défocalisation est balayé linéairement ici, tandis que les autres aberrations restent fixes. La rangée du bas montre l'estimation lorsque le LIFT est précalibré avec la technique proposée. Dans ce dernier cas, la précision est bien meilleure. 10

1.1. For a considerable time, the ELT will remain the only extremely large telescope in operation. Meanwhile, almost 100% of observations conducted with ELT will be AO-assisted. Credit : ESO. 21
1.2. ESO's very large telescope (VLT) facility consists of four 8-meter class unit telescopes (UTs) and four movable 1.8-meter auxiliary telescopes (ATs). Either all UTs or ATs combined can be utilized as a VLT interferometer (VLTI). Credit : ESO. 22

2.1. Airy pattern demonstrated in 1D and 2D for the square root of intensity. This diffraction pattern represents a PSF produced by an optical system with the aperture of diameter D and a perfectly circular aperture. . . . 33
2.2. (a) Diffraction-limited PSF, where FWHM is defined by λ/D; (b) Short-exposure PSF affected by atmospheric turbulence, where separate speckles still have FWHM λ/D, while the overall shape results from the distorted wavefront; (c) Long-exposure PSF affected by the turbulence. The FWHM, in this case, is defined by λ/r_0. 37

2.3. Anisoplanatism illustrated for the off-zenith target. The guide star used for the AO correction and the science target sample different turbulence volumes at high altitudes. 40
2.4. The illustration of the cone effect. An artificial guide star placed at a finite distance from the observer cannot measure the turbulence at high altitudes since it samples the conic section of the turbulence volume. For clarity, the science target is not depicted in this figure. But without the loss of generality, one can imagine it as the place of the natural guide star. 41
2.5. The scheme of a basic AO system with a post-focal DM. The telescope collects the light from a scientific object and a guide star. Part of this light is redirected to WFS, while another part is sent to the science detector. The WFS measurements are processed by an RTC to compute the corrective commands on DM. 45
2.6. Basic working principle of SH WFS. Taken from Mauch & Reger (2014). 49
2.7. Illustration of the concept of generic Fourier filtering optical system. Taken from Fauvarque et al. (2016). 49
2.8. Illustration of the photonic lantern. Taken from Norris et al. (2020). . . 50
2.9. Different types of AO correction. The blue beam represents the light coming from the scientific field; the red beam shows the light coming from an NGS; the yellow cones demonstrate the light from LGSs. . . . 54
2.10. Examples of real PSFs (log-scale) significantly influenced by strong LWE recorded with SPHERE IRDIS. The prominent LWE-induced distortions are notable in the region of the first Airy ring. 74

3.1. *Corner plot :* log-likelihood function sampled in the vicinity of the true modal coefficient values. The couplings between different modes can be seen as elongations in correlation plots. *Top-right :* the correlation matrix between the modal coefficients. *Right-middle :* the ground-truth simulated high-flux PSF used to compute the likelihood. No AO residuals were introduced in the simulation. 94
3.2. Coupling between modes under varying conditions : high flux (a), low flux (b), and high flux with reduced phase diversity (50 nm RMS instead of 170 nm RMS) (c). The ground-truth coefficient values are equivalent to those in Figure 3.1. 96
3.3. 2D visualization of the optimization criterion when both a_3 and a_5 modes are estimated simultaneously with -50 nm RMS phase diversity (left plot) and 200 nm phase diversity (right plot). The error metric visualized in this figure is defined as $|I_0 - I(\mathbf{A})|$. 97
3.4. Optimization criterion when only a_3 is estimated as a function of the introduced astigmatic diversity strength; (a) represents the 3D visualization of the criterion landscape, while (b) shows the 2D projection of the plot for different astigmatic diversity strengths (0, -100, and -200 nm RMS). The criterion here is the same as in Fig. 3.2. 99

4.1. Different approaches to extract and simulate PSFs. The figure is taken from Beltramo-Martin et al. (2020b). . . . 113
4.2. Differential refraction effect. The apparent on-sky position of the object for blue and red wavelength differ by angle θ_λ due to the atmospheric dispersion. It also implies that "red" and "blue" wavefronts sample slightly different columns of atmospheric turbulence. . . . 129
4.3. All PSD contributors are summed and re-normalized to get the final turbulent-phase PSF $W_{\text{total}}(\mathbf{k})$. . . . 131
4.4. Total PSD is then used to compute the final infinite-exposure PSF. In this figure and the figure above, the example of IRDIS is used. The shift term D_v was omitted here. . . . 132
4.5. The diagram of various telemetry levels. Each successive level applies additional post-processing steps to the initial engineering data generated by different telescope subsystems, including the AO system. . . . 138
4.6. Depiction of the post-processing procedure for raw IRDIS PSF cubes to extract centered and clean calibration PSFs. IRDIS splits the image into "left" and "right" channels, where each channel can have its own spectral or polarization filter. Note that these PSFs are influenced by the apodizer, whereas no coronagraph was inserted during their acquisition. 142
4.7. Diagram illustrates the sources of IRDIS datasets and their storage location in ESO archives. . . . 144
4.8. Diagram illustrates the process of collecting the dataset of IRDIS PSFs with the associated reduced telemetry. . . . 144
4.9. Left cube illustrates raw spectral PSF data obtained during on-sky calibration routines for MUSE-NFM. In contrast, the left cube shows the final processed and spectrally binned PSF with $N_{\text{bins}} = 14$. . . . 145
4.10. Diagram illustrates the sources of MUSE-NFM datasets, as well as their storage location. . . . 147
4.11. Diagram illustrates the process of collecting the dataset of MUSE-NFM PSFs with the associated reduced telemetry. . . . 147
4.12. Schematic representation of the direct PSF prediction. After the PSF model is initialized with the default instrument's configuration file, a subset of parameters is re-initialized with the values taken from the reduced telemetry dataset. These values are used without any modifications to generate PSF. . . . 148
4.13. Direct prediction results for the dataset of "left" and "right" IRDIS PSFs. The solid lines demonstrate median profiles computed for the datasets of on-sky PSFs, predicted PSFs, and their absolute difference. The error bars show the 1σ deviation. More details on profiles computation can be found in Appendix E. . . . 150
4.14. Median profiles for the dataset of directly predicted MUSE-NFM PSFs compared with on-sky profiles. The algorithm used to compute these profiles is identical to the one in Figure 4.13. The profiles are calculated for "white" PSFs, which are averaged over all wavelengths. . . . 151

4.15. Orthogonal modal basis that represents the LWE-induced OPD in the pupil plane of IRDIS. The basis contains 4 pistons, 4 tips, and 4 tilt modes for each island of the IRDIS pupil. 156
4.16. Fitting TipTorch model to on-sky PSFs to test if the model can accurately recreate their morphology. The resulting optimal values of model inputs $\hat{\mathscr{X}}$ are stored in the dataset $\mathscr{D}_{\text{fit}}$. 158
4.17. The radial profiles derived from fitting the TipTorch model to the IRDIS PSF dataset. The results show minuscule discrepancies from the on-sky PSFs, indicating very high fitting accuracy. 159
4.18. (a) Effect of the "phase bump" on MUSE-NFM PSFs (image in log-scale) ; "phase bump" seen by IRLOS [1] ; (c) "phase bump" pattern reproduced by DSM. 160
4.19. Radial profiles derived from fitting the TipTorch model to the MUSE-NFM PSF dataset. Similar to IRDIS, the results here demonstrate minimal discrepancies from the on-sky PSFs. 162
4.20. Schematic view of a hypothetical full ML-driven PSF predictor that can be trained on the extensive realistic E2E simulations and then fine-tuned on the real data. Such an approach can offer ultimate flexibility at the price of low interpretability. 166
4.21. Scheme of the proposed calibrated (hybrid) PSF prediction approach. The ML component, trained on on-sky reduced telemetry, transforms the dataset $\mathscr{Z}$) into model inputs $\widetilde{\mathscr{X}}$ before passing them to the PSF model. Unlike Figure 4.12, where reduced telemetry is used unaltered. 167
4.22. Detailed schematics of the calibrated PSF model architecture for IRDIS. Note that the predictor is complemented with 2 models that exclusively infer LWE patterns and LWE-induced WFE. 170
4.23. First step : training the LWE-predictors based on the $\hat{\mathbf{c}}$ values extracted from the earlier obtained $\mathscr{D}_{\text{fit}}$ dataset. Model (a) predicts the coefficient values for the modes from Figure 4.15. Model (b) re-normalizes the prediction of the first model by the WFE values it infers. Both models are trained separately and before the calibrator-NN. Here, $\hat{\mathbf{c}}^*, \tilde{\mathbf{c}}^*$ are the PCA-compressed representation of ground truth and predicted LWE coefficients. 171
4.24. Second step : training of the NN-calibrator. This time, the PSF model is included in the training loop. The LWE predictors have already been trained. 172
4.25. The results of "tuned" direct prediction for "left" and "right" IRDIS PSFs. The main difference with the pure direct prediction is that here, some model inputs are assigned to constant values that, on average, give a smaller discrepancy of predicted PSF with on-sky data. 174
4.26. Calibrated prediction results were obtained using the techniques outlined in Figure 4.22. This approach demonstrates the highest prediction accuracy. 175

4.29. Examples of predicted IRDIS PSFs. The top row shows direct calibrated predictions, the middle row presents tuned direct predictions, and the bottom row compares fitted (GT) and predicted LWE wavefronts. GT LWE values are taken from $\hat{\mathbf{c}} \in \mathscr{D}_{\text{fit}}$. 176
4.30. FWHM prediction error for IRDIS PSFs. The left plot is the absolute, and the right is the relative error. All values above the threshold, both for absolute and relative errors, are compressed into a single bar. 177
4.31. The design of the calibrated predictor for MUSE-NFM. Compared to the approach from Figure 4.24, it has a much simpler design due to the absence of LWE in MUSE-NFM data. 179
4.32. Scheme of training the NN-calibrator for MUSE-NFM. In contrast to IRDIS, the predictor is fully trained with the PSF model in the loop this time. 179
4.33. Left figure recaps the profiles obtained for the direct prediction demonstrated earlier, right figure demonstrates the results obtained using tuned direct prediction. The profiles shown in this figure are computed for "white" PSFs in the same fashion as earlier in Fig. 4.19. 181
4.34. The results of the calibrated prediction for the dataset of MUSE-NFM PSFs. Similar to the case of IRDIS, the calibrated model again demonstrates the best prediction accuracy. 182
4.38. Examples of predicted MUSE PSFs from the validation set. The top row shows calibrated predictions, while the bottom row presents tuned direct predictions. 183
4.39. FWHM prediction error for "white" MUSE-NFM PSFs. The left plot is the absolute, and the right is the relative error. All values above the threshold, both for absolute and relative errors, are compressed into a single bar. 184

1. The effect of low flux ($m_J = 16$) on modal cross-talk. Increased noise introduces numerous small parasitic local minima, which can trap the optimizer and hinder convergence to the true solution. 206
2. The effect of astigmatic small diversity (50 nm RMS instead of 170 nm RMS) on modal cross-talk. The second local minimum appears. 207
3. The SPHERE-client interface. This software was used to download the IRDIS PSF dataset from the ESO archive. 209
4. Assumed tip-tilt jitter values for tuned direction prediction. 219
5. The distribution of values predicted by the NN-calibrator ($\widetilde{\mathscr{X}}$) for IRDIS. 220
6. The distribution of values predicted by the NN-calibrator ($\widetilde{\mathscr{X}}$) for MUSE-NFM. 220

Liste des tableaux

2.1. Characteristics of the Narrow-Field and Wide-Field modes of MUSE enabled by the GALACSI module. 64
2.2. The list of scientific cases for both MUSE WFM and NFM regimes. . . . 65
2.3. The list of different error sources common to AO-assisted telescopes. "-//-" indicates "same as in the previous row". 73

4.1. Median FVU computed for the PSFs from the IRDIS validation dataset. 177
4.2. Median FVU computed for the PSFs from the MUSE validation dataset. 184

1. The data samples in the IRDIS dataset are categorized based on the type of wavelength filters applied in the left and right channels during acquisition. 210
4. Architecture of the calibrator neural network. 217

1. Introduction

The unresolved questions in astronomy

Modern astronomy is experiencing unprecedented growth, driven by the deployment of advanced instruments and observatories over the past decades. This progress has significantly enhanced our understanding of the universe. Yet, many fundamental questions still remain unresolved, such as the nature of dark matter and dark energy, the evolution of the early universe, and the habitability of nearby exoplanets.

Although definitive answers to these questions are still elusive, several approaches have been proposed to address them. For instance, a precise measurement of the redshift and its drift can help determine the acceleration rate of the universe, providing insights into dark energy. Similarly, studying galactic kinematics and stellar dynamics can improve our understanding of galactic halos, potentially composed of dark matter, thereby shedding light on its nature. Additionally, obtaining direct light from exoplanets is crucial for understanding their atmospheric chemistry and assessing their habitability, which, in turn, requires a thorough understanding of stellar chemistry and the formation of planetary systems, which involves resolving the inner regions of protoplanetary disks.

Need for new powerful instruments

Although not comprehensive, this list highlights several key research directions. It is crucial to emphasize that these investigations rely heavily on observations in the visible and near-infrared regions of the electromagnetic spectrum. Furthermore, these observations require exceptional spectral and spatial resolution, which the current generation of astronomical instruments does not fully provide. For instance, with the scope of exoplanetary science, the spectral and spatial resolutions achieved by current instruments through Doppler shift measurements (e.g., HARPS, ESPRESSO, HIRES) or direct imaging (e.g., SPHERE, NIRC2, SCExAO, GPI) are insufficient for studying rocky Earth-sized planets in the habitable zones around nearby stars. Similarly, the precise measurement of the redshift drift of distant galaxies, essential for assessing the universe's expansion and validating cosmological models, largely remains beyond the reach of existing technology.

This illustrates the need for more sensitive astronomical instruments with enhanced spatial and spectral resolution. In this scope, one way to increase spatial resolution and sensitivity is to increase the aperture of the telescope. This is precisely the motivation behind the upcoming generation of giant ground-based telescopes, including the ESO Extremely Large Telescope (ELT) (see Figure 1.2), the Giant Magellan Telescope

(GMT), and the Thirty Meter Telescope (TMT). However, ESO's ELT (later referred to simply as "the ELT") is expected to be the only extremely large telescope in operation for a considerable time. Therefore, every minute of observation with the ELT will be incredibly valuable. This underscores the necessity for delivering the highest quality scientific data to make the most of the limited observation time and fully leverage the ELT's capabilities.

Motivation to use adaptive optics

However, achieving superior imaging quality requires a precise understanding of the imaging process, including all factors that influence the propagation of light from the observed target to the scientific detector.

In ground-based astronomy, the primary source of error is atmospheric turbulence. Atmospheric turbulence limits the achievable angular resolution by blurring images, reducing the advantages of having a large aperture. The impact of the atmosphere on imaging will be discussed in detail later in this manuscript in Section 2.2. Compensation for such atmosphere-induced phase disturbances is the goal of *adaptive optics (AO)* systems. Therefore, the scientific goals achievable with giant observatories like the ELT will be severely limited without using the AO. In other words, almost 100% of the observations with the ELT must be AO-assisted. This underscores the critical importance of this technology in the operation of giant telescopes.

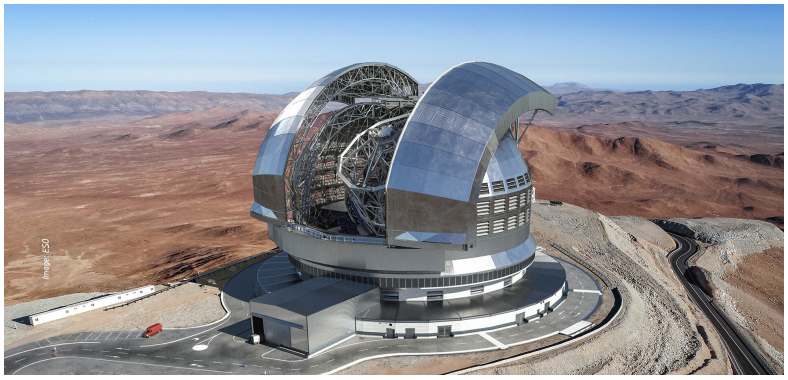


FIGURE 1.1. – For a considerable time, the ELT will remain the only extremely large telescope in operation. Meanwhile, almost 100% of observations conducted with ELT will be AO-assisted. Credit : ESO.

AO will be covered in greater detail later in Section 2.5. However, before that, it is essential to point out that the giant size of the ELT introduces additional error sources beyond atmospheric turbulence. For instance, a larger aperture necessitates a larger primary mirror, which requires massive support structures. These structures can mechanically deform, vibrate, and bend. Another challenge is that the ELT's primary mirror will be segmented due to its size, leading to potential issues with co-phasing the mirror segments.

Of course, many more error contributors can be listed here. While many are already well-known and thoroughly studied through simulations, others still lack complete understanding or characterization. Moreover, some error sources may still be simply unknown and will only be discovered when the telescope starts its operations.

Need for new control strategies

Given that conventional AO control strategies are specifically optimized for atmospheric compensation, controlling these telescope/instrument-induced errors simultaneously with atmospheric disturbance presents a significant challenge. This is because, as detailed in Section 2.5, these internal errors manifest at spatial and temporal frequencies very different from those of the atmosphere. In addition, at specific spatial and temporal frequencies, errors coming from the telescope can even surpass atmospheric errors, further degrading image quality. Therefore, to effectively mitigate internal errors, the ELT might need to rely on more sophisticated control strategies that utilize various sensors and data inputs, going beyond conventional AO control schemes that must be rethought in this case.

1.1. Motivation, objectives, and methods

Abundance of data

When talking about data, it is worth noting that current-generation ground-based observatories, such as ESO's Very Large Telescope (VLT) (see Figure 1.2) or W. M. Keck Observatory (WMKO), generate terabytes of data each week. Future-generation telescopes like the ELT are expected to produce similar volumes every night. This vast amount of data includes not only scientific exposures but also site monitoring measurements, AO system telemetry, and various engineering data recorded by the instruments and telescope subsystems.



FIGURE 1.2. – ESO's very large telescope (VLT) facility consists of four 8-meter class unit telescopes (UTs) and four movable 1.8-meter auxiliary telescopes (ATs). Either all UTs or ATs combined can be utilized as a VLT interferometer (VLTI). Credit : ESO.

In contrast to AO telemetry, which is involved in real-time AO system control, most of these additional engineering data are used only for occasional diagnostics or are not used at all. Nevertheless, these underutilized engineering datasets capture the behavior of the telescope as a whole, as they contain real measurements produced by the interconnected subsystems. This makes this type of data potentially valuable as a complementary source of information that can enable adaptive telescope control that goes beyond conventional AO control.

Among the data that the telescope generates, it is important to emphasize one specific type of data : the images of the point spread function (PSF). These images can be generated as a byproduct of scientific observations and calibration routines or recorded at AO wavefront sensors' focal planes. In essence, the PSF represents the impulse response of the telescope. Therefore, it can also be perceived as the final outcome of the imaging process, accumulating all the effects from all contributors along the imaging path. Therefore, PSFs can serve as a key diagnostic tool and help characterize telescope performance, facilitate a deeper understanding of scientific results, and even improve AO control, making PSFs a crucial source of information.

In this thesis, the term *"fused data"* will be used to refer to a combination of relevant data entries collected from various sources, including scientific images, AO system telemetry, external atmospheric measurements, instrument settings, logs, and, most importantly, PSF images. A more detailed explanation of what constitutes "fused data" will be provided in Section 4.5.

The role of PSF

To emphasize the importance of PSF, it is crucial to list several applications where the use of this kind of data can be highly beneficial.

PSF for science

First, in many scientific applications, the quality of scientific data is directly related to the quality of PSF. Therefore, predicting the PSF based on weather forecasts can be essential in predicting the instrument's performance and estimating the necessary exposure time. Furthermore, accurately predicted PSF morphology can help in better observation planning by inferring the attainable angular resolution under specific conditions, which helps prioritize observations based on their angular resolution requirements.

Furthermore, a well-characterized PSF can be utilized in various post-processing applications. These include bright object subtraction, deconvolution, photometry, astrometry, and more. A detailed understanding of the PSF's morphology is also crucial to avoid misinterpreting its features as genuine features of an astrophysical object. A more detailed discussion of this topic will be given in Section 4.1.

PSF for system diagnostics

Second, understanding the morphology of the PSF is essential to analyze the factors that contribute to image formation. As already mentioned, PSFs collect information from each contributor along the optical path. Thus, theoretically, PSFs contain all the information necessary to characterize the imaging system [1]. This property suggests that the morphology of the PSF can be utilized as a diagnostic tool.

Third, scientific PSFs can capture information that conventional AO wavefront sensors cannot detect. For example, it can happen because AO wavefront sensors usually utilize a separate optical path from science instruments. Meanwhile, as discussed in Section 2.5.4, the optical path responsible for forming scientific images may have its own set of aberrations. The utilization of PSF to address this issue is discussed in Chapter 3.

In summary, all these points underscore the critical importance of understanding and using PSFs as a valuable source of information. However, extracting precise and unambiguous information from PSFs is a complex task. To address this challenge, this research project employs advanced modeling techniques to generate realistic focal-plane PSFs. These models are applied to fit on-sky data, enabling the retrieval of information about aberrations. In addition, they are used within a predictive framework to infer the morphology of PSF based on the fused data inputs.

The problem of PSF modeling

It is important to note that this project exclusively considers AO-corrected PSFs. These PSFs exhibit a highly intricate structure, which emerges from the complex interplay of numerous telescope systems, including the effects resulting from the imperfect AO correction, as detailed in Sections 2.5 and 4.3.1. Classically, the task of modeling AO-corrected PSFs is approached through synthetic end-to-end simulations or methods such as PSF reconstruction (see Véran et al. (1997)), where on-sky AO telemetry is utilized to generate PSF. This telemetry comprises an extensive array of data recorded at each iteration of the AO correction.

One of the main methods applied in this work is the utilization of *Fourier-based PSF models*, which differ from the classical PSF reconstruction methods described earlier. Unlike classical PSF reconstruction, which is directly based on telemetry data, Fourier-based models utilize statistical modeling. As will be discussed later in Section 4.2.3, these models are much more efficient in terms of data consumption, allowing the entire PSF to be described with just a few parameters, thus eliminating the need to process bulky telemetry data arrays.

Although the concept of Fourier models is not new (see Rigaut et al. (1998)), this work introduces a novel hybrid approach that relies on a machine-learning component

1. Provided the information about the phase of the electromagnetic field is preserved.

to complement analytical PSF models with data-driven techniques to achieve more accurate predictions and better alignment with real on-sky data. The machine learning component is designed to correct the inputs of the analytical model, using real on-sky data for training. Chapter 4 will be dedicated to exploring this approach.

In summary, the primary objective of this thesis is to investigate the use of AO telemetry in conjunction with less traditional data sources (including PSFs collected from various sources) to enhance the scientific output of AO-assisted telescopes and also potentially improve current AO control strategies. In this context, particular emphasis is placed on leveraging PSFs to address the following aspects :

— **Benefit for scientific applications :** The accurate knowledge of PSF is invaluable for various scientific post-processing applications, enabling a more unambiguous interpretation of the structure of astrophysical objects in general.

— **PSFs for diagnostics :** In next-generation telescopes, the error contributors can be complex and not solely originate from atmospheric turbulence. These errors may arise from the telescope's size and structural complexity. Focal-plane PSFs, which encapsulate effects from all contributors along the optical path, can be used to identify potential problems.

— **PSFs for calibration :** The morphology of the PSF can contain information that cannot be seen by WFSs employed in AO control. These might be path-specific or unseen aberrations that other AO wavefront sensors fail to measure, such as tip/tilt and focus drifts of laser guide stars.

My research addresses these challenges through comprehensive PSF modeling using physically accurate approaches, covering PSFs obtained from both wavefront sensors and scientific instruments. A key emphasis of this thesis is the use of real data from on-sky observations to ensure realistic and robust modeling. The first part of this research is dedicated to wavefront sensing using focal-plane PSFs. In this context, I propose a novel estimation method that enables calibration of the quasi-static aberrations in the AO system directly on-sky. The other part of this research is dedicated to the realistic modeling of PSFs acquired at the focal planes of scientific instruments. Here, I introduce a novel hybrid PSF modeling technique that augments conventional analytical models with a machine learning component trained on fused data inputs and real on-sky PSFs. This approach demonstrates higher accuracy compared to traditional analytical models alone. Although this PSF modeling method has not yet been directly applied to the post-processing of specific scientific observations, it establishes a flexible and robust, instrument-oriented foundation that can be further tailored to meet specific scientific requirements in the future.

Astronomical instruments used in this project

For further discussion, it is crucial to specify the instruments to which the proposed methods will be applied. Until now, the discussion has addressed this issue in a rather

abstract manner. Howver, the main two instruments considered in this thesis are the Spectro-Polarimetric High-contrast Exoplanet Research (SPHERE) and the Narrow-Field Mode (NFM) of the Multi-Unit Spectroscopic Explorer (MUSE) (see Sections 2.4.1 and 2.4.2 respectively). Both SPHERE and MUSE are current-generation AO-assisted instruments installed on Unit Telescopes (UTs) 3 and 4 at the ESO VLT facility. Although the focus of this thesis on current-generation instruments might seem to contradict the initial emphasis on the ELT, there is no inconsistency. The scope of this thesis is indeed driven by the upcoming challenges posed by the ELT. However, due to the reliance on actual data and the applied nature of this work, verification can only be performed using existing instruments. Therefore, SPHERE and MUSE-NFM serve as the pathfinders for the ideas proposed in this research. In fact, the closest "scale model" to the ELT currently available is the UT4. For instance, both telescopes have a pre-focal adaptive deformable mirror integrated into their structures. Consequently, the experience gained with MUSE will be highly relevant for future ELT instruments, such as HARMONI [2] or MICADO [3].

1.2. The structure of this manuscript

Outlining the structure of this thesis and identifying which parts are dedicated to specific topics is essential to gain a clearer picture of the topics mentioned previously. This will also help clarify transitions and highlight the main components of this work.

— Following this general introduction, the basics of image formation will be discussed, including the theory behind atmospheric turbulence and its impact on astronomical imaging.

— Next, a detailed discussion on adaptive optics will be provided, covering the main components and types of AO systems.

— Building on this foundation, an introduction to the scientific instruments, their regimes, and subsystems used in this work will be given. The selection of these particular instruments will also be justified.

After these foundational sections, the manuscript is divided into two main parts, each with its own introduction, methods, and results sections.

— The third chapter focuses on focal plane wavefront sensing. It describes the method of extracting information about quasi-static phase aberrations from PSFs obtained at the focal plane of a wavefront sensor. In addition, this chapter also covers how information that is not visible to the main AO wavefront sensors can be extracted from the morphology of the focal plane PSF obtained on another auxiliary wavefront sensor.

— In contrast to the PSF model discussed in the previous chapter, which is tailored for short-exposure wavefront sensor PSFs, the approach presented in this chap-

2. https://elt.eso.org/instrument/HARMONI/
3. https://elt.eso.org/instrument/MICADO/

ter focuses on modeling the infinite-exposure scientific PSFs produced by the SPHERE and MUSE instruments. This PSF model is employed to fit the on-sky data and predict the shape of scientific PSFs using machine learning techniques trained on fused data inputs.

Despite the different approaches and applications discovered in these two sections, they are closely related as both focus on utilizing PSFs and rely on physics-based PSF modeling.

— The final section concludes the results from the previous two sections and discusses potential directions for further development of these projects.

With this structure in mind, we now proceed to the main content of this thesis.

2. General background and theory

2.1. Basics of image formation

This manuscript will focus on instruments that primarily function as imagers. Therefore, understanding the process of image formation is crucial. Introducing the necessary mathematical framework will allow the derivation of key quantities and the outlining of concepts that will later be heavily used in this research work.

2.1.1. Optical imaging

Fresnel approximation

The mathematical description of the imaging process can be started by introducing the Fresnel integral (see (Goodman 2005, Chapter 4)), which is formulated as

$$U(x', y') = \frac{e^{ikz}}{i\lambda z} \iint_{-\infty}^{\infty} U_0(x, y)\, e^{i\frac{k}{2z}[(x-x')^2+(y-y')^2]}\, dx\, dy. \tag{2.1}$$

This integral describes the propagation of light between two planes in the Fresnel approximation. Here, x, y and x', y' are the coordinates in the first and second planes, respectively, z is the propagation distance, $U_0(x, y)$ is the incident field observed on the exit of some aperture stop, $k = \frac{2\pi}{\lambda}$ is a wavenumber, and λ is the wavelength of light.

This approximation is valid for near-field diffraction when the light wave is not too far from the aperture and the propagation distance z is not infinite and also not too close to the aperture ($\frac{(D/2)^2}{\lambda z} \sim 1$, where D is the diameter of the aperture stop).

To make Eq.(2.1) more relevant for the purposes of this thesis, it must be applied in the context of imaging optics, since astronomical telescopes are fundamentally imaging systems. The primary goal of telescopes then is to gather light from distant objects and focus rays arriving from different directions onto the corresponding points on the detector, thereby forming a detailed image.

Incoming light is collected in the *pupil plane*, which is defined by the aperture stop of the telescope, which is usually a primary mirror. The aperture defines the amount of light collected by the telescope as well as the maximum attainable resolution, as will be shown later. Meanwhile, the image of the object is created in the *focal plane*. In an ideal imaging system, each point in the object (e.g. a star) maps to a point in the focal plane. Therefore, this is the plane where the detectors are placed. The distance between the aperture of the telescope and the focal plane is denoted as the focal distance f.

Since telescopes are imaging quasi-infinitely far objects, the incoming rays are parallel, which means that when propagated undisturbed, these parallel beams correspond to planar incoming electromagnetic waves.

In the context of imaging, it is possible to define a transfer function $L(x,y)$ that would transform an incident planar wave into a spherical wave converging into a point at distance f :

$$L(x,y) = e^{-i\frac{k}{2f}(x^2+y^2)},$$

where x, y are the 2D coordinates in the pupil plane. In essence, this function describes the effect of a lens or a mirror that focuses light in the focal plane. The incident wave $U_{\text{inc}}(x,y)$ is assumed to be a flat planar wave defined as

$$U_{\text{inc}}(x,y) = A\,e^{-ikz_0},$$

where z_0 is some reference propagation distance. Since it is a planar wave, the exact value of z_0 will later play no role. The incident field $U_0(x,y)$ then can be described as :

$$U_0(x,y) = U_{\text{inc}}(x,y)\,\mathscr{P}(x,y)\,L(x,y) = A\,e^{-ikz_0}\,\mathscr{P}(x,y)\,e^{-i\frac{k}{2f}(x^2+y^2)}.$$

Here, $\mathscr{P}(x,y) \in \mathbb{C}^2$ is a complex aperture function.

Fraunhofer formalism

Substitution of $U_0(x,y)$ to Eq.(2.1) gives :

$$U(x',y') = A\frac{e^{ik(z-z_0)}}{i\lambda z}\iint_{-\infty}^{\infty}\mathscr{P}(x,y)\,e^{-i\frac{k}{2f}(x^2+y^2)}\,e^{i\frac{k}{2z}[(x-x')^2+(y-y')^2]}\,dx\,dy,$$

where x', y' are the coordinates in the image plane. To shorten the expression, it is useful to define $\phi(x,y,x',y')$,

$$\phi(x,y,x',y') = \frac{k}{2z}(x-x')^2+(y-y')^2-\frac{k}{2f}(x^2+y^2) = \left(\frac{k}{2z}-\frac{k}{2f}\right)(x^2+y^2)-\frac{k}{z}(xx'+yy')+\frac{k}{2z}(x'^2+y'^2). \quad (2.2)$$

Then,

$$U(x',y') = A\frac{e^{ik(z-z_0)}}{i\lambda z}\iint_{-\infty}^{\infty}\mathscr{P}(x,y)\,e^{-i\,\phi(x,y,x',y')}\,dx\,dy.$$

The special case of this equation is imaging in the focal plane. In this case, $z = f$, then $\frac{k}{2z}-\frac{k}{2f} = 0$. Thus, the phase term $\phi(x,y,x',y')$ simplifies to :

$$\phi(x,y,x',y') = -\frac{k}{f}(xx'+yy')+\frac{k}{2f}(x'^2+y'^2)$$

Substituting the simplified phase term gives :

$$U(x',y') = A\frac{e^{ik(f-z_0)}}{i\lambda f}e^{-ik\frac{x'^2+y'^2}{2f}}\iint_{-\infty}^{\infty}\mathscr{P}(x,y)\,e^{-i\frac{2\pi}{\lambda f}(xx'+yy')}\,dx\,dy.$$

In this equation, one can recognize the Fourier transform, which can be defined as :

$$\mathscr{F}\{\mathscr{P}(x,y)\} = \iint_{-\infty}^{\infty}\mathscr{P}(x,y)\,e^{-i\frac{2\pi}{\lambda f}(xx'+yy')}\,dx\,dy.$$

Inserting the lens transfer function $L(x,y)$ in the equation cancels out the quadratic term of the Fresnel integral if the field is considered in the focal plane. In other words, with the addition of the lens term, the Fraunhofer equation naturally emerges and, in this case, is not merely an approximation but a fully valid expression within the framework of scalar diffraction theory, resulting in

$$U(x',y') = A\frac{e^{ik(f-z_0)}}{i\lambda f}e^{-ik\frac{x'^2+y'^2}{2f}}\mathscr{F}\{\mathscr{P}(x,y)\}. \tag{2.3}$$

Intensity in the focal plane

Equation 2.3 can now be used to describe light propagation in imaging systems. Having this equation of the complex field, it is now possible to compute the intensity distribution in the focal plane. The intensity $I(x',y')$ is defined as :

$$I(x',y') = |U(x',y')|^2.$$

Meanwhile, in Eq.(2.3), the $\frac{e^{ik(f-z_0)}}{i\lambda f}$ term is a constant phase shift that does not affect the final intensity distribution. The same is valid for the $e^{-ik\frac{x'^2+y'^2}{2f}}$ term. As a result, they can be neglected in the final expression for the intensity distribution, which is defined as :

$$I(x',y') = \left|\frac{A}{\lambda f}\mathscr{F}\{\mathscr{P}(x,y)\}\right|^2. \tag{2.4}$$

However, this expression can be modified further. The amplitude can be renormalized for convenience so that the λf denominator is removed. Moreover, without loss of generality, it is also convenient to assume $A = 1$ in the following derivations. In addition, note that $\mathscr{P}$ is a complex function, so it describes not only the amplitude but also the phase in the pupil plane. Thus, it can be divided into the amplitude and phase components. Then, Eq.(2.4) becomes

$$I(x',y') = \left|\mathscr{F}\{P(x,y)\,e^{i\frac{2\pi}{\lambda}\Phi(x,y)}\}\right|^2, \tag{2.5}$$

where $P(x,y) \in \mathbb{R}^2$ is a binary mask of the amplitude in the pupil plane. $\Phi(x,y) \in \mathbb{R}^2$ is a 2D function that describes the wavefront shape in the pupil plane. With this definition, it becomes possible to insert wavefront aberrations by modifying the $\Phi(x,y)$ term.

These aberrations may result from the imperfections in the optical system itself or be induced by the propagation of a planar wavefront through the inhomogeneous refractive medium (like Earth's atmosphere) before arriving at the pupil plane. These aspects will be mentioned and explored in detail later.

Note that modeling $P(x, y)$ with a flat intensity distribution is an assumption. In reality, wave amplitude can vary across the pupil due to various factors, such as, for example, atmospheric scintillation discussed later in Section 2.2.4. However, in the context of AO and this work, such variations are neglected since their impact is usually minimal. Thus, the only instance where $P(x, y)$ is considered non-uniform in this work is when modeling the effect of *apodizer*, which is a technique that employs specially designed amplitude pupil masks to shape the morphology of the focal-plane PSF, as detailed in Vigan et al. (2013).

To summarize, the Eq.(2.5) is the imaging equation that describes optical imaging under the following primary assumptions :

— **Paraxial approximation**. As mentioned previously, it assumes that the light rays propagate at small angles relative to the optical axis.

— **Flat pupil-plane amplitude.** The only exception is the apodizer;

— **Scalar diffraction theory.** Polarization is ignored;

— **Thin optical elements.** This implies that the intensity in the focal plane is solely defined by the aperture mask and the wavefront in the pupil plane;

— **Monochromaticicity.** Equation 2.5 is only valid for the monochromatic light with the wavelength of λ;

In fact, the discussed equation is very general and hence can be applied to various imaging systems, including astronomical telescopes. The Eq.(2.5) will be extensively used later in different parts of this manuscript.

2.1.2. Point spread function

According to previously derived expressions, a plane wave is never imaged as a single point but rather as a spot with limited dimensions because of diffraction. This is precisely what is described by Eq.(2.5). From this perspective, this expression defines the impulse response of the optical system or, in other words, its response to a single incoming plane wave. This is also known as the *point spread function (PSF)*. The PSF is fundamental in optics, as it captures the effects of diffraction and aberrations within the system. Moreover, the concept of PSF plays a central role in this work, as this research project aims to extract information from the PSF and correlate its morphology with the system's behavior and performance.

In this context, Eq.(2.5) can be rewritten as follows :

$$\mathrm{PSF}(x', y') = \left| \mathscr{F}\{P(x, y)\, e^{i\frac{2\pi}{\lambda}\Phi(x,y)}\} \right|^2 . \tag{2.6}$$

Therefore, if there are wavefront aberrations inside the optical system, they affect the $\Phi(x, y)$ term and thus alter the morphology of the PSF. This also implies that PSF inherently encodes all the information about the imaging system itself and about the optical signal that was collected by this system. However, information about the phase is lost because of the $|\cdot|^2$ operation. This means that special techniques are required to manipulate the PSF in such a way that information about the phase can be partially recovered from the intensity distribution. This aspect will be discussed later in Chapter 3.

2.1.3. Angular resolution

When discussing the characterization of optical systems, the PSF can be useful in determining the limiting resolution of an imaging optical system. In case there are no wavefront aberrations, the size of the PSF is determined by the diameter of the aperture [1].

To define angular resolution, it is helpful to assume a telescope with a circular aperture of diameter D that can be described by the aperture function $P(x, y)$ in the pupil plane :

$$P(x, y) = \text{circ}\left(\frac{2R}{D}\right) = \begin{cases} 1, & \text{if } R = \sqrt{x^2 + y^2} \leq \frac{D}{2}; \\ 0, & \text{otherwise.} \end{cases}$$

Applying Eq.(2.6) to this aperture function to compute PSF gives :

$$I_{\text{Airy}}(r) = I_0 \left[\frac{2J_1\left(kr\frac{D}{2}/f\right)}{kr\frac{D}{2}/f}\right]^2,$$

where J_1 is the first-order Bessel function of the first kind, I_0 is the intensity normalization factor and $r = \sqrt{x'^2 + y'^2}$ is the radial coordinate in the image plane. The relation $\frac{r}{f} = \sin\theta \approx \theta$ can be expressed in terms of angular coordinates, leading to :

$$I_{\text{Airy}}(\theta) = I_0 \left[\frac{2J_1\left(\frac{\pi D}{\lambda}\sin\theta\right)}{\frac{\pi D}{\lambda}\sin\theta}\right]^2. \tag{2.7}$$

The equation obtained illustrates the *Airy pattern*, which describes how a circular aperture is imaged by a diffraction-limited system (see Figure 2.1). This formulation is crucial because it helps easily determine the highest possible angular resolution the system can achieve.

1. This can also be interpreted through Heisenberg's uncertainty principle ($\Delta x \Delta p_x \geq \hbar/2$). The larger aperture is worse at exactly localizing the photon arrival position Δx, meaning that their transverse momentum Δp_x is better localized, leading to a smaller angular spread. This means that photons that arrived from the same direction tend to focus in smaller spots.

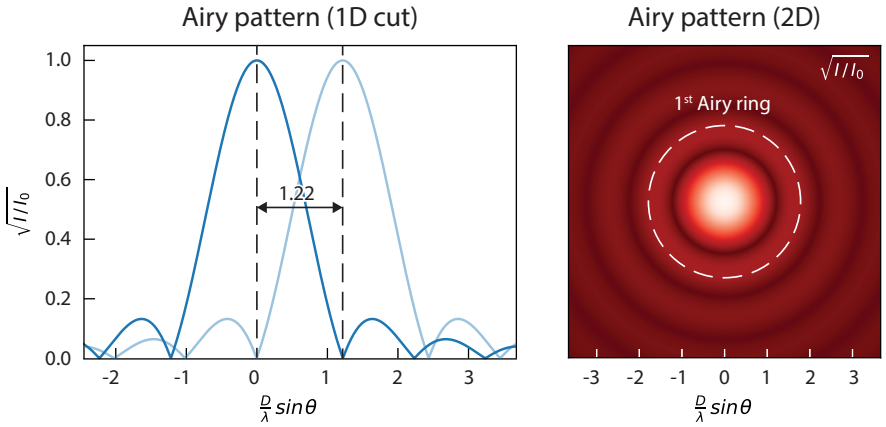


FIGURE 2.1. – Airy pattern demonstrated in 1D and 2D for the square root of intensity. This diffraction pattern represents a PSF produced by an optical system with the aperture of diameter D and a perfectly circular aperture.

The resolution is usually evaluated using the *Rayleigh criterion*, which describes the minimum angular separation θ_{min} at which two point sources can be resolved :

$$\theta_{\text{min}} = 1.22\,\frac{\lambda}{D}.$$

The factor 1.22 in this expression comes from the first zero of J_1, which is the first dark ring in the Airy pattern. This equation essentially indicates that two points are considered resolved if the angular separation between the peak of one PSF and the first minimum of the other PSF is more than $1.22\,\lambda/D$ (refer to Figure 2.1). This expression demonstrates that the only way to increase the (theoretical) angular resolution of the telescope is to increase the size of its aperture [2] provided the wavelength stays the same. In reality, however, the attainable resolution is determined not solely by diffraction limit but also by atmospheric turbulence and other optical aberrations. This will be discussed in detail later in this manuscript.

2.1.4. Image formation

Following the approximation listed above for deriving Eq.(2.6), it is now possible to apply this equation to image formation using the following assumption :

— **Linear optical systems.** It implies that the system can be fully characterized by its impulse response (PSF). It also implies that the output response can be represented by the weighted sum of the individual responses;

— **Shift-invariant optical systems.** It means that regardless of the coordinate in the focal plane, the PSF stays unchanged;

2. Of course, interferometry should not be overlooked in this context. However, this method still involves creating a virtual aperture, which is determined not by the aperture size of individual telescopes but by the separation between them.

In astronomical imaging, light from celestial objects can be decomposed into a series of plane waves due to their quasi-infinite distance from the observer. Without diffraction, these plane waves would be perfectly focused as δ-functions (or point sources) in the focal plane, represented by $O(x', y')$. Since the optical system is assumed to be linear and shift-invariant, the image of an object can be modeled as the convolution of the object $O(x', y')$ with the system's PSF $\mathrm{PSF}(x', y')$. Then the image $I(x', y')$ is given by :

$$I(x', y') = O(x', y') \circledast \mathrm{PSF}(x', y') = \int\int O(x, y)\, \mathrm{PSF}(x' - x, y' - y)\, dx\, dy, \tag{2.8}$$

where $\circledast$ denotes convolution operation, $O(x', y')$ denotes the image of an object as captured by a perfect optical system, free from aberrations and diffraction. This convolution integral implies that each point of the object contributes to the image according to the PSF. The overall image is thus a superposition of shifted and weighted PSFs corresponding to each point in the object. It is important to note that the shift-invariance of the optical system is a strong approximation that, in practice, holds only within a limited isoplanatic region of the field of view. In reality, PSFs observed in actual telescopes can exhibit significant variability across the field due to the optical design or the type of AO correction employed. Nevertheless, this assumption is adopted here to streamline the derivation of key concepts that will be essential for the subsequent discussion.

The convolution model is powerful because it allows the use of Fourier analysis to study image formation. By the *convolution theorem*, the Fourier transform of the image is the product of the Fourier transforms of the object and the PSF :

$$\mathscr{F}\{I(x', y')\} = \mathscr{F}\{O(x', y')\}\, \mathscr{F}\{\mathrm{PSF}(x', y')\}.$$

Here, the term $\mathscr{F}\{\mathrm{PSF}(x', y')\}$ is also known as *optical transfer function (OTF)*. This relationship is instrumental in understanding how the optical system modifies the frequency content of the object.

However, a more accurate form of expression 2.8 would incorporate the PSF's variability across the field, thereby violating the shift-invariance property. Nevertheless, the field variability can be integrated into the model "on top", allowing Eq.(2.8) to be utilized in a broad spectrum of applications, such as image deconvolution and imaging modeling over a wide field of view.

In summary, several key points were highlighted for further discussion.

— Increasing the aperture diameter enhances the resolving power of a diffraction-limited system ;

— The PSF characterizes an optical system and thus can be used for measuring aberrations and assessing system performance ;

— The PSF is determined by the aperture mask and the phase at the pupil plane ;

This section assumes ideal, error-free light propagation from the object to the optical system, meaning that the propagation medium does not introduce any distortions (as would be the case in a vacuum). Although this is true for space-based observatories, where all aberrations are intrinsic to the imaging system, ground-based telescopes must deal with strong and highly dynamic aberrations introduced by the Earth's atmosphere. The atmosphere causes wavefront distortions that accumulate as the light travels from space to the ground-based telescope, introducing aberrations on top of those inherent to the optical system. These effects will be explored in the next section.

2.2. Atmospheric turbulence

The theory of atmospheric turbulence has been extensively explored in numerous sources (Kolmogorov (1991); von Kármán (1948); Fried (1965), (Roddier 1999b, Chapter 2), (Dainty 2000, Chapter 1)). Therefore, this section will only cover the key quantities and concepts necessary for the subsequent discussions. Atmospheric turbulence refers to the irregular and chaotic motion of air, characterized by random fluctuations in velocity, pressure, and temperature. These turbulent motions can be caused by various factors, including convective processes, where the Earth's surface absorbs solar heat during the day and releases it back at night, or by temperature inversions. Turbulence can also arise from wind interaction with terrain or differences in wind speed and direction at different altitudes, such as the influence of high-altitude jet streams.

Turbulent flow comprises so-called "eddies," which are swirling motions or vortices within the fluid or air. Eddies represent regions where the local air movement deviates from the overall flow. These eddies exist at different scales, each containing varying amounts of energy. In a turbulent system, larger eddies break down into smaller ones through a process called the energy cascade, where energy is transferred from larger to smaller scales until it dissipates.

The mixing of cold and hot air in the turbulent atmosphere causes variations in air density, and because air density directly affects the refractive index, light passing through the atmosphere is subject to random retardations based on local refractive index fluctuations. In turn, these variations in the air refractive index affect the imaging through the atmosphere. Consequently, atmospheric turbulence plays a significant role in degrading the performance of ground-based telescopes by distorting incoming light.

2.2.1. Structure function

A widely accepted model of these refractive index fluctuations is based on Kolmogorov's model (see Kolmogorov (1991)) and is defined as :

$$D_n(\Delta\mathbf{r}) \equiv \left\langle |n(\mathbf{r}) - n(\mathbf{r}+\Delta\mathbf{r})|^2 \right\rangle = C_n^2 \, \Delta\mathbf{r}^{2/3}. \tag{2.9}$$

Here, $\mathbf{r}$ is a 3-dimensional vector, while $\Delta\mathbf{r}$ represents the displacement between two points in the space. The structure function (SF) D_n measures the expected value of the squared difference between the refractive index $n(\mathbf{r})$ measured at two points separated by a distance $\Delta\mathbf{r}$. It characterizes the spatial variation of the refractive index due to turbulence. The refractive index structure constant, C_n^2, quantifies the intensity of these fluctuations. This parameter is essential in understanding and modeling the impact of atmospheric turbulence on optical wave propagation through the turbulent medium.

By introducing the refractive index structure function $D_n(\Delta\mathbf{r})$, it is now possible to compute the phase structure function D_ϕ, which is highly practical to quantify the impact of turbulence on optical wavefronts. The phase structure function D_ϕ measures the variance in phase differences between two points in a turbulent medium, which helps to quantify how atmospheric turbulence affects the phase of an optical wavefront.

Let us now define the turbulent wavefront (WF) $\psi(\mathbf{x})$ as

$$\psi(\mathbf{x}) = \exp\big(i\phi(\mathbf{x})\big),$$

with the phase shift caused by refractive index fluctuations is defined as

$$\phi(\mathbf{x}) = k \int_h^{h+\delta h} n(\mathbf{x}, z)\, dz, \quad k = \frac{2\pi}{\lambda}.$$

Here, k is the wavenumber, λ is the wavelength. These equations define a WF observed in the pupil plane of a telescope. Given that in the case of astronomy, light travels from space to Earth, it is practical to express the atmospheric influence of turbulent layers of thickness δh at various altitudes h.

To characterize the statistical behavior of this wavefront, the coherence function $B_h(\boldsymbol{\rho})$ of the wavefront passing through an atmospheric layer at height h can be expressed as

$$B_h(\boldsymbol{\rho}) \equiv \left\langle \psi(\mathbf{x})\, \psi^*(\mathbf{x}+\boldsymbol{\rho}) \right\rangle = \left\langle \exp\left[i\left(\phi(\mathbf{x}) - \phi(\mathbf{x}+\boldsymbol{\rho})\right)\right] \right\rangle, \tag{2.10}$$

where $\langle\cdot\rangle$ is a spatial averaging operation and $\boldsymbol{\rho}$ is the spatial separation between two points in the pupil-plane phase screen. Assuming that term $\left[\phi(\mathbf{x}) - \phi(\mathbf{x}+\boldsymbol{\rho})\right]$ has Gaussian statistics with zero mean, it is possible to write

$$B_h(\boldsymbol{\rho}) = \exp\left(-\frac{1}{2} D_\phi(\boldsymbol{\rho})\right),$$

where D_ϕ is the phase structure function. After inserting Eq.(2.9) into Eq.(2.10) equation, it is possible to derive the phase structure function for a turbulent layer of thickness δh :

$$D_\phi(\boldsymbol{\rho}) = 2.914\, k^2 C_n^2\, \delta h\, \|\boldsymbol{\rho}\|^{5/3}. \tag{2.11}$$

Equations (11-17) in Quirrenbach (2004) provide more detailed derivations of these expressions. By plugging the structure-function equation into the coherence function, we derive the coherence function for a single atmospheric layer.

$$B_h(\boldsymbol{\rho}) = \exp\left[-\frac{1}{2}\left(2.914\,k^2 C_n^2\,\delta h\,\|\boldsymbol{\rho}\|^{5/3}\right)\right].$$

Integrating over all atmosphere layers gives the final expression for the mutual coherence function :

$$B(\boldsymbol{\rho}) = \exp\left[-\frac{1}{2}\left(2.914\,k^2\sec(\gamma)\,\|\boldsymbol{\rho}\|^{5/3}\int C_n^2(h)\,dh\right)\right]. \tag{2.12}$$

Here, γ represents the angle of observation from the zenith, effectively accounting for the elongated column of turbulence encountered by light as it travels through the atmosphere at non-zenith angles.

2.2.2. Fried's parameters (r_0)

Using the equations above, it is possible now to introduce a new quantity r_0, known as *Fried's parameter*, which is one of the most critical quantities in describing atmospheric turbulence (see Fried (1965)). It has two main interpretations :

1. **Coherence length :** r_0 defined as the spatial scale over which the atmospheric turbulence causes phase variations of roughly 1 rad root mean square (RMS). In this sense, it defines the coherence length of turbulence for a given wavelength.
2. **Resolution** : It can be interpreted as the diameter of the circular pupil for which the diffraction-limited image and the seeing-limited image have the same angular resolution. When the telescope aperture $D > r_0$, atmospheric turbulence defines the limiting angular resolution, resulting in a seeing-limited PSF with a full width at half-maximum (FWHM) proportional to λ/r_0 (see Figure 2.2).

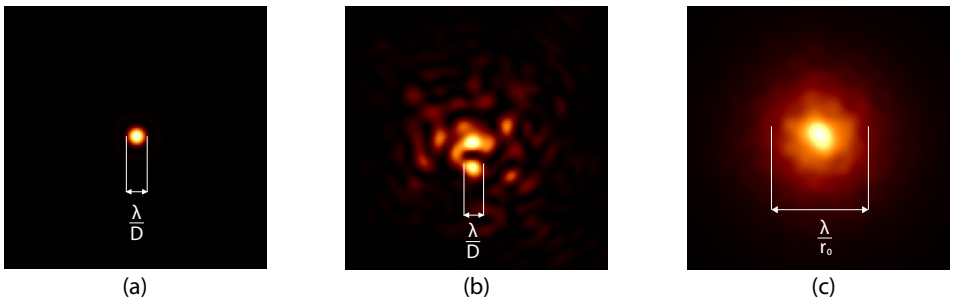


FIGURE 2.2. – (a) Diffraction-limited PSF, where FWHM is defined by λ/D; (b) Short-exposure PSF affected by atmospheric turbulence, where separate speckles still have FWHM λ/D, while the overall shape results from the distorted wavefront; (c) Long-exposure PSF affected by the turbulence. The FWHM, in this case, is defined by λ/r_0.

The Fried's parameter can also be related to the atmospheric turbulence profile C_n^2 :

$$r_0 = \left[0.423\, k^2 \sec(\gamma) \int C_n^2(h)\, dh\right]^{-3/5}. \tag{2.13}$$

Then, equations (2.12) and (2.11) transform into

$$B(\boldsymbol{\rho}) = \exp\left[-3.44\left(\frac{\|\boldsymbol{\rho}\|}{r_0}\right)^{5/3}\right], \tag{2.14}$$

$$D_\phi(\boldsymbol{\rho}) = 6.88\left(\frac{\|\boldsymbol{\rho}\|}{r_0}\right)^{5/3}. \tag{2.15}$$

Importantly, it is possible to derive from Eq.(2.13) that r_0 is a wavelength-dependent parameter, increasing with the 6/5 power of the wavelength :

$$r_0 \propto \lambda^{6/5}.$$

Typical values of r_0 in the visible spectrum range from 10 to 20 cm.

In Fried (1965) and Noll (1976) it is also shown that the turbulence-induced wavefront error σ_1 for the telescope with a diameter D can be calculated as

$$\sigma_1^2 = 1.03\left(\frac{D}{r_0}\right)^{5/3}.$$

The key takeaway from this relation is that the wavefront error caused by turbulence grows almost quadratically with increasing telescope diameter. This highlights that imaging is affected by atmospheric turbulence more substantially for larger telescopes.

2.2.3. Temporal behavior of the turbulence (τ_0)

The Eq.(2.9) can be reformulated as a function of both space and time, allowing us to account for temporal variations in atmospheric turbulence.

$$D_n(\tau) \equiv \left\langle |n(\mathbf{r}, t) - n(\mathbf{r}, t+\tau)|^2 \right\rangle.$$

Experience indicates that temperature inhomogeneities in the atmosphere typically persist longer than the time required for wind-driven inhomogeneities to cross the line of sight. This phenomenon holds for most wind conditions that affect open-air turbulence. If $\mathbf{v}$ denotes the wind velocity vector, then

$$n(\mathbf{r}, t+\tau) = n(\mathbf{r} - \mathbf{v}\tau,\, t).$$

This approximation is known as the Taylor (*frozen-flow*) approximation, where the turbulence is "frozen" as it moves along with the wind. Essentially, the temporal evolution of turbulence is dominated by the advection of spatial turbulence structures

by the wind. Then :

$$D_n(\tau) = \left\langle |n(\mathbf{r}, t) - n(\mathbf{r} - \mathbf{v}\tau, t)|^2 \right\rangle = C_n^2 \|\mathbf{v}\tau\|^{2/3}.$$

Assuming a mean propagation velocity with $\overline{v}$, the temporal structure function of the wavefront phase can be derived by substituting $\overline{v}\tau$ for the spatial separation $\Delta\mathbf{r}$ in the definition of the spatial structure function definition in Eq.(2.15). This leads to

$$D_\phi(\tau) = 6.88 \left(\frac{\overline{v}\tau}{r_0} \right)^{5/3}.$$

Thus, the characteristic atmospheric coherence time τ_0, which represents the time scale over which 1 rad of WF error accumulates, can be expressed as :

$$\tau_0 = 0.314 \frac{r_0}{\overline{v}}.$$

More detailed explanations can be found in (Roddier 1999b, Chapter 2).

As will be demonstrated in Section 4.3.1, the frozen-flow approximation significantly simplifies the derivations by assuming that turbulence remains spatially unchanged within a short time as it moves with the wind. Therefore, this approximation is fundamental in many applications, such as predictive wavefront control and turbulent layer profiling.

2.2.4. Atmospheric scintillation

Another phenomenon worth mentioning in the context of atmospheric turbulence is the rapid intensity fluctuations of light as it passes through the turbulent atmosphere, known as *scintillation*. This effect occurs because strong refractive index fluctuations can locally focus and defocus light that passes through the turbulent atmosphere. This locally induced wavefront curvature propagates through the space, resulting in spatial intensity fluctuations (see Osborn et al. (2015)).

Scintillation is particularly pronounced in systems with small apertures. A typical example of this phenomenon is the twinkling of stars observed by the human eye. However, the scintillation averages out as the aperture diameter increases, making it negligible for big aperture systems such as modern large astronomical telescopes. Therefore, scintillation effects will not significantly influence the subsequent discussions and derivations and are mentioned here only for completeness.

2.2.5. Angular anisoplanatism

The expressions discussed previously apply only to a wavefront observed in a single direction. However, light coming from different directions interacts with different columns of the turbulent atmosphere. This leads to different wavefront aberrations accumulated along each line of sight, as illustrated in Figure 2.3.

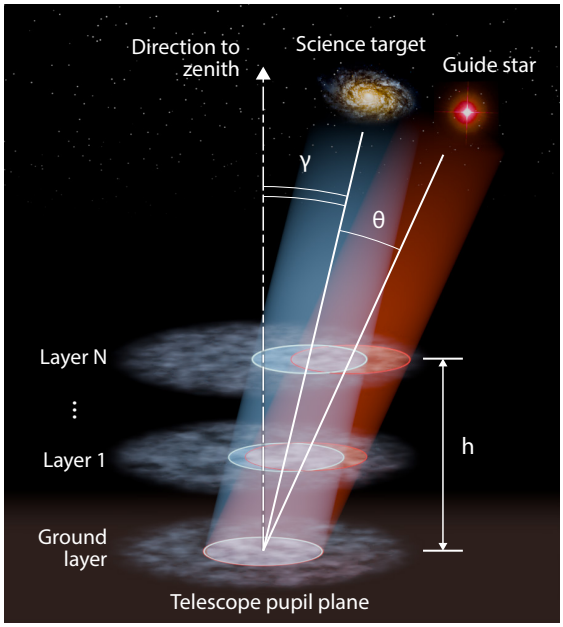


FIGURE 2.3. – Anisoplanatism illustrated for the off-zenith target. The guide star used for the AO correction and the science target sample different turbulence volumes at high altitudes.

This figure shows two light directions observed at angles γ and $\gamma+\theta$. By expressing the height h in terms of these angles and substituting them into Eq.(2.11), it becomes possible to quantify this effect, resulting in the following expression :

$$D_{\phi,\text{aniso}}(\theta) = 2.91k^2\sec(\gamma)\int C_n^2(h)\left(\theta h\sec(\gamma)\right)^{5/3}dh = \left(\frac{\theta}{\theta_0}\right)^{5/3}.$$

Here, $D_{\phi,\text{aniso}}$ represents the anisoplanatism structure function, which quantifies the degree of decorrelation between two wavefronts observed in adjacent directions.

Similarly to the Fried parameter r_0, we can define the isoplanatic angle θ_0, which describes the angular range over which the variance of the relative phase between wavefronts remains within 1 rad^2 (isoplanatic patch) :

$$\theta_0 \equiv \left[2.91k^2\sec(\gamma)^{8/3}\int C_n^2(h)h^{5/3}dh\right]^{-3/5}.$$

It can also be expressed as :

$$\theta_0 = 0.314\cos(\gamma)\frac{r_0}{H} \quad \text{if} \quad H \equiv \left(\frac{\int C_n^2(h)h^{5/3}\,dh}{\int C_n^2(h)\,dh}\right)^{3/5}.$$

As will be shown later, anisoplanatism is an essential factor in adaptive optics. It defines the angular range within which the AO corrections remain effective.

2.2.6. Cone effect (focal anisoplanatism)

Previously, the observed target was considered to be at an infinite distance. However, if light crosses the turbulent atmosphere from a finite distance, the effect known as the *"focal anisoplanatism"* or the *"cone effect"* occurs. This effect is particularly relevant in adaptive optics systems that use laser guide stars positioned at a finite altitude in the atmosphere (this will be discussed again in Section 2.3.3.1). Figure 2.4 illustrates the basic principle of the cone effect.

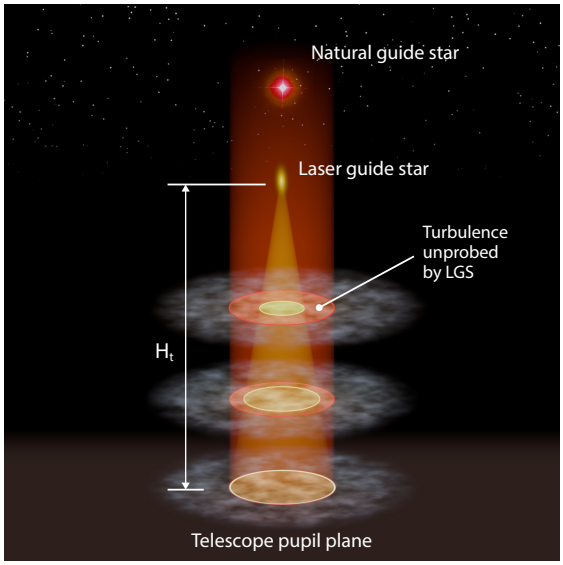


FIGURE 2.4. – The illustration of the cone effect. An artificial guide star placed at a finite distance from the observer cannot measure the turbulence at high altitudes since it samples the conic section of the turbulence volume. For clarity, the science target is not depicted in this figure. But without the loss of generality, one can imagine it as the place of the natural guide star.

The figure illustrates that an artificial (laser) guide star positioned at a finite distance H_t samples a conical section of the turbulent atmospheric volume, unlike a natural guide star, which probes a cylindrical column of the atmosphere (will be discussed later in Section 2.3.3.1). Consequently, the turbulence outside the guide star's light cone remains unmeasured, although it continues to impact the image quality. This introduces an error, which can be expressed as

$$\sigma^2_{\text{cone}} = \left(\frac{D}{d_0}\right)^{5/3}, \quad d_0 = \lambda^{6/5} \cos^{3/5}(\gamma) \left[\int C_n^2(h)\, F\left(\frac{h}{H_t}\right) dh\right]^{-3/5}. \tag{2.16}$$

Here, H_t is the altitude of the target, and the function $F(h/H_t)$ is detailed in Tyler (1994). Further explanations of the cone effect can be found in Foy (2000).

The key insight from Eq.(2.16) is that the cone effect increases almost quadratically with the telescope's diameter. This effect is negligible for smaller telescopes (meter-

class) but begins to pose challenges for 8-meter-class telescopes. Meanwhile, for ELTs, the cone effect becomes unmanageable. Additionally, the cone effect strongly depends on wavelength. Although it can still be tolerated in infra-red on VLT-sized telescopes ($\sigma_{\text{cone}} \propto 10^1$ nm RMS), it severely affects observations in the visible spectrum.

2.2.7. Turbulent PSD

Another crucial aspect of describing and quantifying atmospheric turbulence involves expressing the statistical properties of turbulent wavefronts through the phase spectral density (PSD). For Kolmogorov turbulence, the phase spectral density $W_n(\mathbf{k})$ can be derived by utilizing the relationship between the structure function and the covariance, along with the Wiener-Khinchin theorem. Then, it takes the form

$$W_n(\mathbf{k}) = 0.033\, C_n^2\, \|\mathbf{k}\|^{-11/3}, \tag{2.17}$$

where $\mathbf{k}$ is the 2D spatial frequency vector. The derivation of this expression is quite complex and will not be included here. More details on the derivation can be found in Dainty (2000); Quirrenbach (2004); Noll (1976). Equation 2.17 can be extended by incorporating the so-called inner and outer scales, resulting in

$$W(\mathbf{k}) = \frac{0.033\, C_n^2}{\left(\|\mathbf{k}\|^2 + \kappa_0^2\right)^{11/6}} \exp\left(-\frac{\|\mathbf{k}\|^2}{\kappa_m^2}\right), \tag{2.18}$$

where $\kappa_0 = 2\pi/L_0$ and $\kappa_m = 5.92/\ell_0$. This modified PSD represents the von Kármán turbulence model. Mathematically, the rationale for introducing the outer scale accounts for the fact that, under the Kolmogorov model, the average refractive index difference between two points grows indefinitely with increasing separation, which is unrealistic. The outer scale limits this growth at large separations, rendering the formulation more physically accurate. Furthermore, it prevents PSD from growing infinitely as $\|\mathbf{k}\| \to 0$, where low spatial frequencies dominate.

Physically, the outer scale L_0 represents the largest size of the atmospheric eddies, while the inner scale ℓ_0 corresponds to the smallest eddies. As discussed in Dainty (2000), the outer scale is particularly relevant for adaptive optics since it influences the lowest spatial frequencies, which contain most of the energy according to Eq.(2.18). Consequently, the inner scale ℓ_0 will be neglected in this work.

2.2.8. From PSD to PSF

As will be demonstrated later, the PSD in Eq.(2.18) plays a crucial role in all the derivations in Section 4.3.1. Meanwhile, as previously discussed in Section 2.1.2, under certain assumptions, the PSF morphology is determined by the complex phase in the pupil plane. At the same time, since the PSD represents the wavefront variance of the phase, it can be directly linked to the morphology of the focal-plane PSF.

As shown in Neichel (2008a), multiplying the PSD with white noise can generate

instantaneous pupil-plane phase screens, producing a short-exposure PSF. Meanwhile, using the PSD as it is within a statistical framework enables calculating a long-exposure PSF. In this case, the first step in this process is to compute the phase covariance function $B(\boldsymbol{\rho})$ by taking the inverse Fourier transform of the PSD :

$$B(\boldsymbol{\rho}) = \mathscr{F}^{-1}\{W(\mathbf{k})\}.$$

Note that the definition of $W(\mathbf{k})$ here is not limited to von Kármán or Kolmogorov PSDs and can be general. The phase structure function D_ϕ is then related to the covariance $B(\cdot)$ by :

$$D_\phi(\boldsymbol{\rho}) = 2\left[B(0) - B(\boldsymbol{\rho})\right].$$

Once the structure function is determined, we can compute the OTF, which quantifies how the optical system itself and atmospheric turbulence affect the transfer of different spatial frequencies.

$$B(\mathbf{x} = \lambda f_T \boldsymbol{\nu}) = \exp\left(-\frac{1}{2} D_\phi(\lambda f_T \boldsymbol{\nu})\right),$$

where $\boldsymbol{\nu}$ is the vector of spatial frequencies in the image plane; D and f_T are the aperture diameter and focal length of the telescope, respectively; λ is the wavelength of light. Finally, the PSF is obtained by taking the inverse Fourier transform of the total OTF :

$$I(\mathbf{x}') = \mathscr{F}^{-1}\{B(\boldsymbol{\nu})\, T_0(\boldsymbol{\nu})\}.$$

Here, $T_0(\boldsymbol{\nu})$ is the OTF of the telescope optical system with the maximal transmitted spatial frequency $\|\boldsymbol{\nu}\|_{\max} = D/(\lambda f_T)$. In the simplest case, $T_0(\boldsymbol{\nu})$ is just a covariance of the aperture function (diffraction-limited OTF). The more complex case may also include static aberrations, as discussed in Section 4.3. $I(\mathbf{x}')$ is the PSF and $\mathbf{x}'$ is the vector of coordinates in the image plane. A comprehensive summary of this mathematical framework is provided in Fétick et al. (2018).

2.2.9. Seeing-limited PSF

Applying the previously derived expressions and Eq.(2.14), it is possible to formulate an analytical expression for the long-exposure OTF :

$$\mathrm{OTF}_{\mathrm{seeing}}(\boldsymbol{\nu}) = B(\lambda f_T \boldsymbol{\nu})\, T_0(\boldsymbol{\nu}) = T_0(\boldsymbol{\nu}) \exp\left(-3.44\left(\frac{\lambda f_T \|\boldsymbol{\nu}\|}{r_0}\right)^{5/3}\right).$$

Conceptually, the mutual coherence function and temporally averaged long-exposure OTF are connected since both are the pupil field's autocorrelation function.

A PSF produced by $\mathscr{F}^{-1}\{\mathrm{OTF}_{\mathrm{seeing}}(\boldsymbol{\nu})\}$ is called seeing-limited. In this formulation, the strength of atmospheric turbulence, characterized by r_0, determines the broadening of the PSF. This expression also indicates that, regardless of the telescope's aperture diameter, atmospheric turbulence will constrain the FWHM of a long-exposure PSF as long as $D > r_0$.

However, this expression does not apply to PSFs obtained using the AO correction. The details of modeling long-exposure PSFs in the presence of AO correction will be covered extensively in Section 4.3.

2.2.10. Strehl ratio

To explore the connection between phase aberration and image quality, it is helpful to introduce the *Strehl ratio* (SR), defined as the ratio of the central intensities of the aberrated PSF to that of the diffraction-limited PSF :

$$\mathrm{SR} = \frac{I(\|\mathbf{x}'\| = 0)}{I_{\mathrm{DL}}(\|\mathbf{x}'\| = 0)},$$

where I_{DL} is the diffraction-limited PSF. The Strehl ratio indicates how much energy is concentrated in the central peak of the PSF. It can also be expressed as the ratio of the integrals of OTFs :

$$\mathrm{SR} = \frac{\int \mathrm{OTF}(\boldsymbol{\nu})\, d\boldsymbol{\nu}}{\int \mathrm{OTF}_{\mathrm{DL}}(\boldsymbol{\nu})\, d\boldsymbol{\nu}},$$

where $\mathrm{OTF}_{\mathrm{DL}}$ is the diffraction-limited OTF. As mentioned in Dainty (2000), for small phase perturbations, the Strehl ratio can be related to the phase variance with Maréchal approximation :

$$\mathrm{SR} \approx 1 - \sigma_\varphi^2, \quad \sigma_\varphi^2 \ll 1\ \mathrm{rad}^2, \quad \mathrm{SR} \geq 0.8.$$

Meanwhile, for larger phase aberrations, the Strehl ratio depends on the specific nature of the aberration. However, as also mentioned in Dainty (2000), in the case of strong random Gaussian aberrations caused by atmospheric turbulence, and when $D \gg r_0$, the relationship becomes

$$\mathrm{SR} \approx \exp\left(-\sigma_\varphi^2\right).$$

In conclusion, this discussion on turbulence and its impact on image quality underscores the critical importance of minimizing turbulence effects on astronomical images. This can be achieved by actively correcting the pupil-plane phase to compensate for the atmospheric distortions and "flatten" the wavefront. The closer the wavefront approaches flatness, the more the resulting PSF resembles the diffraction-limited case, which represents the theoretical resolution limit of the system. This is exactly the primary goal of AO, which will be explored in more detail in the following section.

2.3. Adaptive optics

Adaptive optics (AO) technology has already been mentioned in this thesis, but a brief overview of its fundamental principles is necessary for the following discussion, as all results presented further apply only to AO-assisted observations.

As discussed in the previous section, the primary objective of an AO system applied to astronomy is to correct for the rapidly changing phase aberrations caused by atmospheric turbulence. In Section 2.1.3, it was shown that a perfectly flat wavefront produces a diffraction-limited PSF that results in the highest possible image quality. The AO system aims to "flatten" the wavefront in the telescope's pupil plane to improve image quality.

2.3.1. Basic scheme

The discussion of AO begins by describing the basic structure of an AO system, as shown in Figure 2.5.

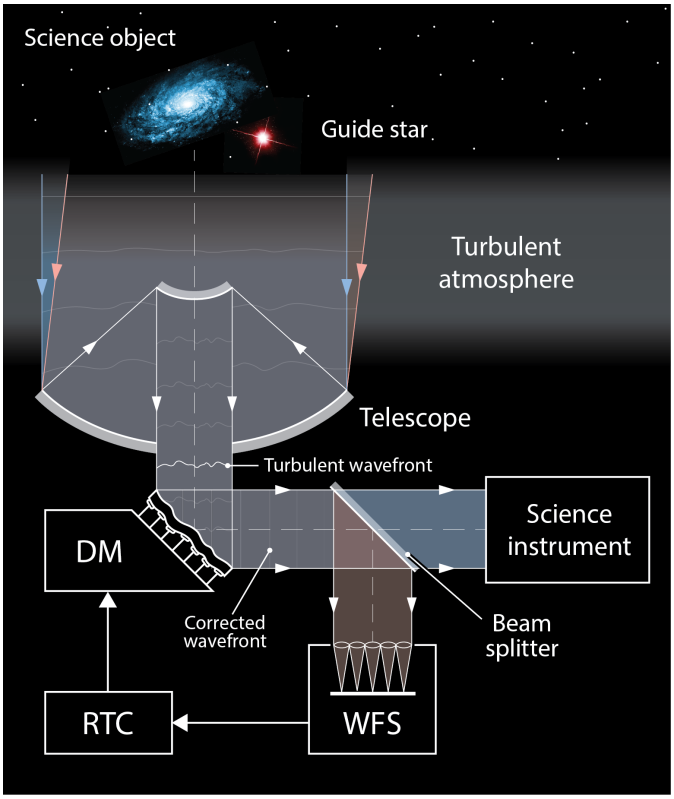


FIGURE 2.5. – The scheme of a basic AO system with a post-focal DM. The telescope collects the light from a scientific object and a guide star. Part of this light is redirected to WFS, while another part is sent to the science detector. The WFS measurements are processed by an RTC to compute the corrective commands on DM.

A typical AO system consists of three main components : a *wavefront sensor* (WFS), a *deformable mirror* (DM), and a *real-time computer* (RTC). Incoming light is divided into two parts : one is directed toward the science detector, and another is sent to the WFS. The WFS measures wavefront distortions caused by atmospheric turbulence, while the DM compensates for these distortions. The RTC processes the WFS data to reconstruct the phase and computes corrective commands, which are then sent to the DM to perform the correction. Since the atmosphere is constantly evolving, the AO operates iteratively. However, to measure atmospheric turbulence, the system requires a reference object or a guide star (GS), as illustrated in Figure 2.5.

2.3.2. AO correction strategies

The iterative process of AO correction is usually referred to as an "AO loop". When discussing AO correction strategies, two primary approaches exist to implement AO loops (Mu et al. (2008)) :

— **Closed-loop correction :** The system measures the residual wavefront error and uses this to adjust the DM. The residual WF (later simply "residuals") is the WF that remained after the previous DM correction iteration. Measuring residuals on every subsequent iteration creates a feedback mechanism since every current measurement depends on the system's previous state. This allows for gradual iterative compensation of errors and enhanced stability of the loop. This feedback mechanism also makes the system more tolerant of model inaccuracies (in this context, the "model" maps WFS measurements to DM commands). However, errors that are not properly detected and compensated can propagate within the loop and accumulate there, eventually destabilizing the system over time and even causing the loop to open.

— **Open-loop correction :** In contrast, in this regime, the system measures the incoming wavefront error before DM correction and directly adjusts the DM shape based on these measurements. Therefore, there is no feedback from the corrected wavefront. Although more reactive to current conditions, open-loop correction requires a highly accurate system model since there is no feedback loop to iteratively compensate for errors "inherited" from previous correction steps. In addition, since the atmospheric turbulence is measured directly in this correction scheme, the WFSs used in this case must be able to sense large wavefront aberration (to have a large linearity range).

For most astronomical instruments, closed-loop correction is preferred due to its stability and higher correction accuracy. Therefore, only a closed-loop correction will be discussed later in this manuscript.

2.3.3. Main AO system components

As previously stated, a basic AO system comprises WFS, DM, and RTC. However, each of these components requires a more in-depth discussion.

2.3.3.1. Guide stars

As mentioned, in the astronomical AO, a reference GS must be observed to reconstruct the wavefront distorted by atmospheric turbulence along a specific line of sight.

Natural Guide Star (NGS). In this context, a bright star in the sky serves as an ideal reference object. It is known as a natural guide star (NGS). Figure 2.5 illustrates the AO system that utilizes an NGS for wavefront sensing. To minimize photon noise in WFSs, this object must be sufficiently bright to ensure accurate wavefront reconstruction as elaborated in (Rousset 1999, Section 5.5).

Laser Guide Star (LGS). However, due to the anisoplanatism effect (discussed in Section 2.2.5), an NGS must be located close to the target object, which is often not feasible, as suitable NGSs are sparsely distributed in the sky[3]. Laser guide stars (LGSs) are used to overcome this limitation. In this case, artificial reference stars are created in desired directions by projecting a laser beam into the Earth's upper atmosphere. There are two main types of LGS :

— **Rayleigh LGS :** Utilizes Rayleigh scattering in the lower atmosphere (10–20 km altitude) by emitting laser light in the mid-visible range (Thompson & Gardner (1989)).

— **Sodium LGS :** A laser excites the sodium layer at around 90 km altitude, causing sodium atoms to fluoresce at 589.6 nm, creating a bright, visible spot on-sky Happer et al. (1994).

LGSs enable AO-assisted observations in regions without an NGS, such as the galactic pole, significantly expanding the sky coverage of the AO system. However, LGSs come with limitations :

— The laser beam is deflected by the atmosphere twice (upwards and downwards), unlike a stellar beam, which experiences only one deflection. As a result, LGSs cannot measure tip-tilt errors induced by atmospheric turbulence.

— LGS images are elongated, which complicates the WFS measurements since the LGS does not behave as a point source. This effect is particularly pronounced for Rayleigh LGS, which appears line-like because of scattering along the entire laser beam path. In such cases, a pulsed laser combined with fast electro-optical shutters can be used to isolate portions of the laser beam (gating), making it appear more point-like to WFS. Meanwhile, Na-LGS is isolated on specific altitude ranges and thus looks more point-like when observed from the ground. However, the elongation increases the more off-axis an LGS target is observed.

— Higher-order aberrations occur because the laser beam propagates as a cone on its return path, unlike the more uniform deflection of stellar light.

— Na-LGS altitudes can fluctuate due to density perturbations in the sodium layer.

3. Provided the scientific object is not used as an NGS.

— Cone effects. As mentioned in Section 2.2.6, since LGSs are at a finite distance, they do not sample high-altitude atmospheric turbulence as effectively as NGSs.
— LGS-emitted wavefronts are spherical, unlike planar wavefronts incoming from stellar objects. This complicates some mathematical models.
— Finally, LGS systems add complexity to AO design, requiring powerful lasers, projection systems, and safety protocols to prevent interference with aircraft and satellites.

Due to these limitations, real AO systems often combine LGS and NGS to enable a more complete correction, as will be discussed in more detail in Section 2.3.4.1.

2.3.3.2. Wavefront sensor (WFS)

The wavefront sensor measures the WF induced by atmospheric turbulence with appropriate spatial and temporal sampling. A typical WFS comprises an optical system that converts the phase information into the intensity distribution in the focal plane of a detector or a set of detectors. Every WFS is constrained by at least these limiting factors :

— **Spatial wavefront sampling error :** Due to the finite number of sampling points, real WFSs can only measure WF in a discrete fashion.
— **Sensitivity/linearity trade-off :** Every WFS, like most sensors in general, must balance the ability to detect small perturbations with the need to address larger distortions without non-linear effects (linearity range). Different WFS designs offer varying characteristics in this regard. It should be noted that sensitivity can be partially regained via non-linear reconstruction techniques.
— **Detector noise error :** The photodetectors used to measure intensity are subject to intrinsic noise, which can affect the accuracy of wavefront measurements.

In astronomical AO, some of the most commonly used WFS types include the following.

— **Shack-Hartmann WFS (SH WFS) :** One of the most widely used WFS types, the SH WFS consists of an array of lenslets (sub-apertures) in the pupil plane (Mauch & Reger (2014)). When a planar wavefront passes through the lenslet array, it gets split, forming an image of spots whose positions are determined by the lenslet array geometry. A distorted wavefront shifts the spots according to the local wavefront slope over each lenslet (see Figure 2.6). By computing the centroids of the spots, the wavefront slopes can be estimated, allowing the SH WFS to effectively measure the first derivative of the wavefront. SH WFSs balance moderate sensitivity with a large linearity range. Smaller lenslets and higher-resolution detectors improve sensitivity but may introduce noise due to reduced flux per sub-aperture. The system's linearity depends on the maximum allowable spot displacement, with denser lenslet arrays reducing the linearity range (given the same number of pixels). SH WFSs are well-suited for conditions with strong atmospheric turbulence since they can be designed to achieve high

linearity.

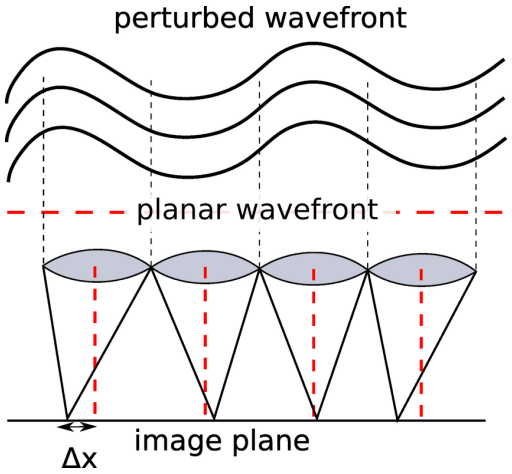


FIGURE 2.6. – Basic working principle of SH WFS. Taken from Mauch & Reger (2014).

— **Fourier-based WFSs :** These WFSs leverage Fourier optics principles to measure wavefront errors by analyzing intensity patterns produced by specific optical transformations (see Fauvarque et al. (2016) and Figure 2.7). Among the notable examples in the family of Fourier WFSs, the following can be listed :

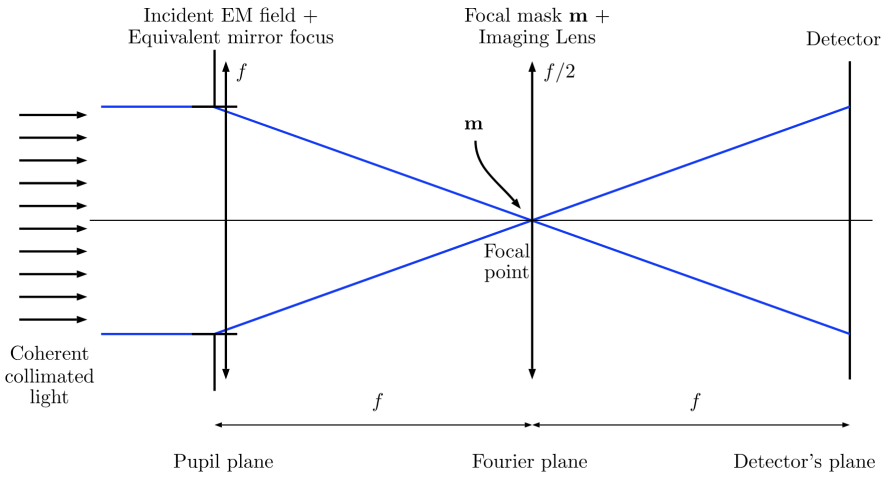


FIGURE 2.7. – Illustration of the concept of generic Fourier filtering optical system. Taken from Fauvarque et al. (2016).

— **Pyramid WFS (PyWFS) :** The PyWFS uses a refractive pyramid to split incoming light into 4 beams (it also can be less or more), performing the phase retrieval by computing the differences between the beams' images (Ragazzoni (1996)). The light is focused on the pyramid's tip for splitting. PyWFS is highly sensitive at the price of a limited linearity range. The latter can be addressed by modulating the incoming beam, i.e., by circularly shifting it

over the pyramid tip. This approach extends the linearity range but reduces sensitivity.

- **Zernike WFS :** This WFS introduces a phase-shifting element at the focal plane to create intensity variations proportional to the wavefront phase. It is highly sensitive to low-spatial (low-order) frequency aberrations but has a very limited linearity range, making it effective mostly for detecting small-amplitude aberrations (N'Diaye et al. (2016)).
- **Bi-O edges WFS :** A recent concept introduced in Vérinaud, C. et al. (2024), it operates similarly to PyWFS or Foucault knife but uses semi-reflective gradual transmission mask for beam splitting instead of refractive pyramids or sharp amplitude masks. This design results in higher optical throughput and increased sensitivity.

- **Photonic lanterns (PLs) :** A promising concept for low-order wavefront sensing, photonic lanterns consist of a tapered waveguide where a multimode fiber transitions into an array of single-mode fibers (see Norris et al. (2020) and Figure 2.8). The spatial modes of the multimode input are redistributed into the fundamental modes of the single-mode outputs. This structure allows efficient coupling to devices like spectrographs. PLs offer high sensitivity to small wavefront aberrations, which can even be tuned per mode. However, photonic lanterns usually exhibit a largely non-linear response. In this scope, machine learning techniques are highly prospective for reconstructing the phase from PL measurements.

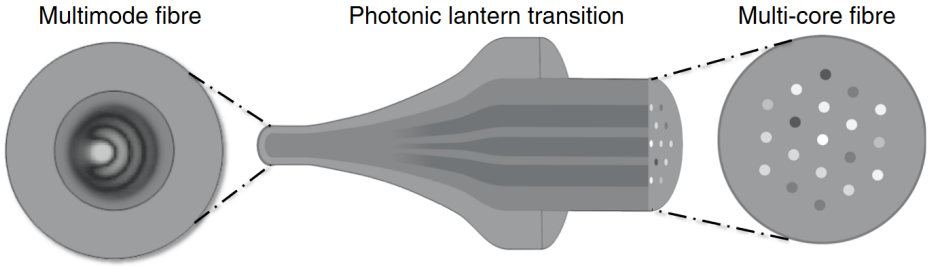


FIGURE 2.8. – Illustration of the photonic lantern. Taken from Norris et al. (2020).

- **Curvature WFS (CWFS) :** A low-order wavefront sensor that measures the curvature (second derivative) of the wavefront by analyzing intensity variations between two images with different defocus (Roddier (1988)). CWFS is particularly sensitive to low-order aberrations. It reconstructs the wavefront by solving the Poisson equation, which results from the transport-of-intensity equation. Under certain conditions, CWFS can achieve high sensitivity. Meanwhile, non-linear reconstruction algorithms can be applied to extend the linearity range of CWFS (Mateen et al. (2017)).
- **Focal-plane wavefront sensing (FP WFSing) techniques :** A broad family of wavefront sensing methods that retrieve phase information by analyzing the PSF morphology in the focal plane. These methods are discussed further in detail

in Chapter 3. FP WFSing approaches are convenient as they require minimal modification to system optics (Gonsalves (1982); Meimon et al. (2010); Korkiakoski et al. (2014)). In some cases, even scientific PSFs can be used for wavefront sensing. Since light is concentrated in the PSF core, these techniques provide excellent sensitivity, especially for low-order aberrations. However, phase extraction from PSF is a highly non-linear process, which necessitates intricate reconstruction algorithms.

This list does not encompass all possible wavefront sensing methods or applications. For example, interferometry-based techniques were not covered here. However, it offers a brief overview of the most widely used approaches. In fact, only the Shack-Hartmann WFS and focal-plane wavefront sensing techniques will be directly relevant to this thesis.

2.3.3.3. Deformable mirror (DM)

A deformable mirror (DM) is an optical element with a surface that can be reshaped precisely and rapidly to correct the incoming distorted wavefronts. In basic terms, it usually consists of a thin reflective surface mounted on an array of actuators, which can be piezoelectric, magnetic, or electrostatic. By adjusting the actuators as measured WF $\times(-2)$, the DM introduces a counteracting distortion to compensate for atmospheric-induced phase aberrations, which flattens the incident wavefront and thus restores the image quality.

When talking about hardware implementations of DMs, the following technologies can be listed :

— **Electrostatic DMs :** Operate by applying electric potential to generate a repelling electrostatic force between the actuators and the mirror surface, moving the mirror (Nam et al. (2016)).

— **Piezoelectric DMs :** Use piezoelectric actuators that expand or contract with the applied voltage, causing the mirror to deform (Srivastava et al. (2022)).

— **Electromagnetic DMs :** Utilize actuators with magnets or coils. An electric current generates a magnetic field, moving the actuators to deform the mirror surface.

— **Voice Coil DMs :** Use voice coil actuators, similar to those found in loudspeakers, to rapidly adjust the mirror surface (Jiang et al. (2023)). In a way, these mirrors are a special case of electromagnetic DMs. An example of such a mirror is the deformable secondary mirror (DSM) installed on UT4.

— **MEMS DMs :** Microfabricated mirrors with tiny actuators, often based on electrostatic or piezoelectric principles, suitable for miniaturized applications (Bifano (2011)).

— **Spatial Light Modulators (SLM) :** Although not mirrors, these devices control wavefronts by modulating the light phase using liquid crystals (see Schmidt et al. (2006)).

Integrating DM into telescope design

Apart from the hardware implementation, it is useful to cover DM integration into the optical design of the AO system. DMs can be integrated in pre-focal or post-focal configurations, each with distinct implications for the optical design. For example, ESO instruments like MUSE and ERIS installed on ESO UT4 utilize pre-focal DMs, such as the DSM, while SPHERE and GRAVITY+ employ post-focal DMs. The schematic AO system presented in Figure 2.5 depicts an AO system with a post-focal DM as well. However, there are several advantages to using a DSM in a pre-focal configuration :

- **Simplified instrument design :** Post-focal DMs, typically small mirrors within the instrument's AO module, require additional relay optics, complicating the design. In contrast, pre-focal DMs are usually implemented as deformable secondary mirrors and, therefore, correct early in the optical path using a larger mirror (e.g., UT4's DSM), allowing the telescope, not an instrument, to manage the AO correction, simplifying the instrument design.
- **Shared AO correction :** A pre-focal DM can serve multiple instruments simultaneously, eliminating the need for separate AO modules in each instrument.
- **Wider field of view :** Post-focal DMs require complex relay optics to reproject the pupil for wide-field coverage, while pre-focal DMs, integrated directly into the telescope, can correct turbulence over a large FOV without additional optics on the instrument side.
- **Improved throughput :** Pre-focal DMs reduce the number of reflective and refractive elements, increasing an AO system's optical throughput and sensitivity by minimizing light loss.
- **Better thermal properties :** DMs installed inside instruments can generate significant heat in the tightly packed space, which complicates thermal management. Meanwhile, pre-focal mirrors external to the instrument enclosure can be cooled efficiently.

Therefore, installing a pre-focal deformable mirror to implement the AO correction is better if the design allows it. For example, the pre-focal configuration is adopted in the ELT, which has a deformable M4 mirror integrated into the telescope design.

The drawbacks of using pre-focal DMs include their higher cost and reduced stability in maintaining the alignment between the DM and the WFS. This instability arises because the DM is positioned farther away from the WFS, making precise optical alignment more challenging. In addition, a key benefit of post-focal DMs is their ability to implement AO correction on telescopes that were not initially designed with integrated AO systems. Moreover, some configurations of the AO systems (see Section 2.3.4.3) require multiple DMs. This can only be implemented with several additional post-focal DMs.

2.3.3.4. Real-time computer (RTC)

The computational core of an AO system is the real-time (controller) computer (RTC), which receives the measurements from the WFS, performs the reconstruction

of the phase, and finally calculates the corresponding corrective commands for the DM. Due to rapidly evolving atmospheric turbulence, these computations must be performed at high rates $\propto 10^2 - 10^3$ Hz.

The relationship between the measured wavefront distortions and the corrective commands is determined via the control law. The control law implemented in the most current-generation RTCs is a simple integrator (see (Madec 1999, Section 6.4)). However, more advanced and more optimal control laws (Kulcsár et al. (2006)), such as linear-quadratic-Gaussian (LQG) control (Petit et al. (2009)), are also possible. For example, LQG is implemented by the ITTM controller on the SPHERE instrument (see Beuzit et al. (2019)).

The RTC introduces temporal errors due to the computation time needed for phase reconstruction and the calculation of corrective commands. Additionally, because the WFS operates at a finite frequency, it samples phase information at discrete time intervals and requires signal integration time, which further contributes to the overall system bandwidth limitations.

More advanced control algorithms can mitigate these delays. In a basic setup, the RTC calculates corrections based on the most recent WFS measurements. However, RTC can also work in the predictive control framework and compensate for temporal lag by predicting corrections based on past data and atmospheric evolution models. In this scope, LQG control and machine learning-based control strategies (see Nousiainen et al. (2022); Landman et al. (2021)) offer more flexibility by incorporating models of system dynamics, noise statistics, and temporal correlations to predict and correct wavefront errors more accurately than traditional integrator-based controllers. This allows these advanced methods to handle more complex scenarios, adapt to changing conditions, and achieve higher performance in AO control systems.

Another important aspect of RTCs is the hardware implementation. It determines the speed, latency, and overall performance of the control loop. The implementation of the RTC hardware can vary depending on performance needs. Current-generation RTCs often use high-performance Field-Programmable Gate Arrays (FPGAs) (for example, SPARTA (Valles et al. (2012))). Recently, solutions based on Graphics Processing Units (GPUs) like COSMIC RTC (Ferreira et al. (2020)) have gained popularity because of GPUs' high parallel computing capabilities. Moreover, GPU-based RTCs are well-suited to integrate machine learning control strategies, as GPUs are highly efficient for such computational tasks.

To summarize, this section presented the fundamental principles of AO and offered a concise overview of the key components that make up nearly any astronomical AO system. However, in practice, the AO systems employed by modern ground-based observatories such as VLT, Keck, Gemini, LBT, and others are typically much more intricate. These advanced systems incorporate multiple DMs, WFSs, and even RTCs to implement different types of AO corrections tailored to specific scientific applications. The subsequent section will expand on the subject of AO correction, discussing various

types of AO correction types implemented in actual astronomical AO systems.

2.3.4. Types of AO correction

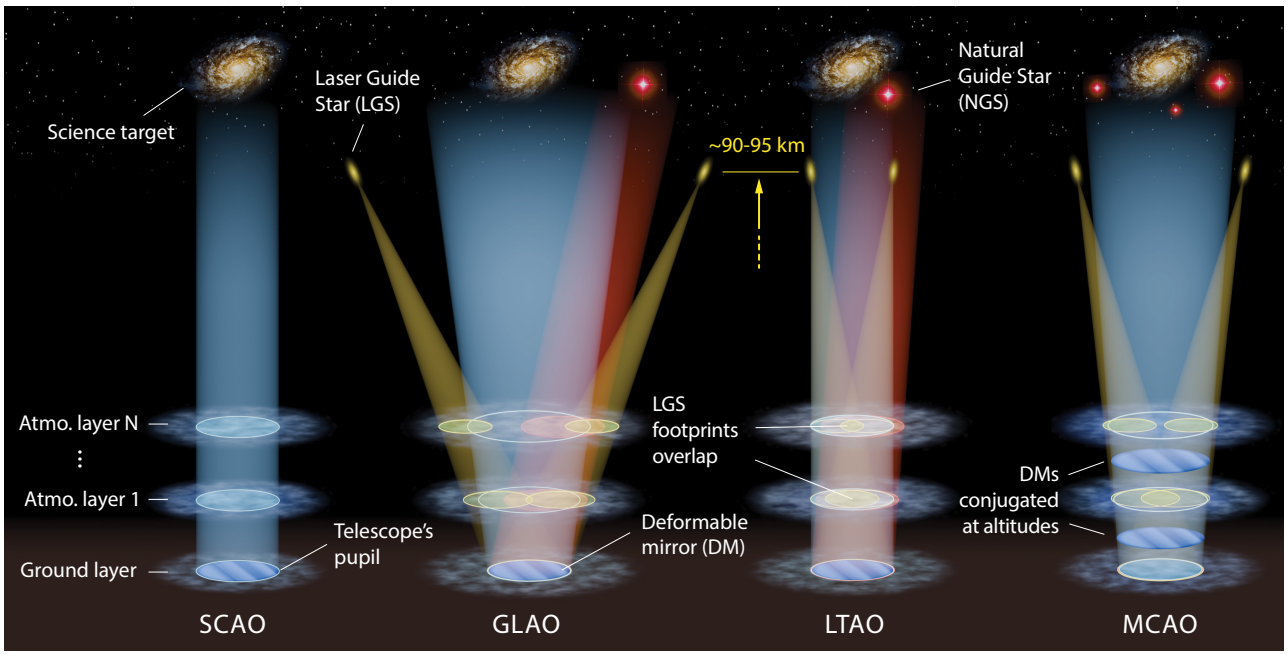


FIGURE 2.9. – Different types of AO correction. The blue beam represents the light coming from the scientific field; the red beam shows the light coming from an NGS; the yellow cones demonstrate the light from LGSs.

2.3.4.1. Single-Conjugate Adaptive Optics (SCAO)

SCAO represents the most classic AO configuration (apart from simple tip-tilt correction). It relies on a single relatively bright target to measure the turbulent wavefront. The SCAO is shown schematically in Figure 2.9. It was also previously depicted in Figure 2.5 to illustrate the fundamental principles behind AO.

However, due to anisoplanatism (refer to Section 2.2.5), SCAO is primarily effective in the direction of the target. Thus, SCAO is employed for "moderate FOV" observations within or near the isoplanatic patch, which usually ranges from a few arcseconds in the visible spectrum to a few tens of arcseconds in the infrared. Among instruments that use this type of correction, it is possible to list NACO (Rousset et al. (2003)) and SINFONI (Eisenhauer et al. (2003)) instruments of VLT, MICADO (Davies et al. (2010)) instrument of ELT, Keck II NIRC2 instrument (Wizinowich et al. (2000)), and many more.

Extreme Adaptive Optics (XAO)

Thus, the observed target is typically placed closer to the center of the FOV. However, this type of correction is optimal for specific applications, such as eXtreme

Adaptive Optics (XAO). XAO aims to achieve high angular resolution within the FOV of several arcseconds while providing excellent AO correction for an on-axis target and its vicinity. To accomplish this, XAO operates at very high loop frequencies and corrects for high spatial frequencies using high-resolution WFS and DMs. XAO can also function in a two-stage configuration (see KPIC (Mawet et al. (2016)), SAXO+ (Goulas et al. (2023)), PCS (Kasper et al. (2021))). Combined with a coronagraph, XAO is ideal for applications like direct imaging of exoplanets or protoplanetary disks around nearby stars. The SPHERE instrument of ESO, one of the instruments studied in this manuscript (Beuzit et al. (2019)), relies on the principles of XAO (see Section 2.4.1 for details and references). Other prominent examples of XAO instruments are the Gemini Planet Imager (GPI) (Macintosh et al. (2014)) at the Gemini Observatory, the SCExAO instrument of the Subaru telescope (Lozi et al. (2018)), KPIC at the Keck (Mawet et al. (2016)), and the upcoming PCS at the ELT (Kasper et al. (2021)).

Single-Laser Adaptive Optics (SLAO)

Another special case of SCAO is Single-Laser Adaptive Optics (SLAO) (not shown in Figure 2.9). In this regime, a single LGS is used instead of an NGS. As mentioned earlier in Section 2.3.3.1 the accuracy of the AO correction depends on the brightness of the target. Therefore, not all on-sky sources can serve as reliable NGSs, leading to limitations in sky coverage. SLAO addresses this problem by using an LGS, which can be created in almost any observing direction. As highlighted in Section 2.3.3.1, LGS are bright artificial targets that can be generated almost anywhere in the sky. This flexibility allows for selecting any desired viewing direction without having to point near a bright NGS to enable AO correction.

This is particularly useful for observing extragalactic targets, as regions containing such objects often lack sufficiently bright NGSs. The ERIS instrument in UT4 (see Davies et al. (2023)) is a prime example of using SLAO for this type of scientific observation.

However, employing SLAO presents certain limitations.

— This correction regime is affected by the cone effect. As discussed in Section 2.2.6, this effect arises because the LGS is positioned at a finite distance from the telescope, which leads to the inability to sample the highest layers of turbulence adequately. Unsampled turbulence results in deteriorated correction and image degradation. Thus, in this case, the trade-off for improved sky coverage is a decreased SR.

— Moreover, SLAO continues to be affected by anisoplanatism effects that limit the FOV.

— Another drawback, as discussed in Section 2.3.3.1, is the insensitivity of LGSs to turbulence tip-tilt. Consequently, these measurements must be complemented with fast tip-tilt measurements ($\propto 10^2 - 10^3$ Hz) using an NGS.

— Furthermore, as also highlighted in Section 2.3.3.1, the sodium layer employed to generate Na-LGS experiences altitudinal density fluctuations over time (Pfrom-

mer & Hickson (2014)), leading to biased defocus measurements with LGS. As a result, this defocus drift must be tracked using an NGS. However, unlike the fast-evolving tip-tilt, the sodium layer drifts much slower, allowing correction at lower frequencies ($\propto 10^0 - 10^1$). In fact, these two points will be thoroughly discussed in the paper on p. 77.

Considering all these constraints, it is reasonable to say that SLAO addresses the sky coverage limitations only to a certain extent.

2.3.4.2. Laser-Tomography Adaptive Optics (LTAO)

Although SLAO addresses the issue of limited sky coverage, it is still restrained by the cone effect and anisoplanatism. In this scope, LTAO can be considered as an approach to overcome these constraints of SLAO correction. LTAO generalizes the concept of SLAO by using multiple LGS, each measured by its own WFS, to achieve optimal performance in a limited field of view (FOV). Using LGSs also facilitates broad sky coverage.

LTAO provided corrections to a FOV similar to that of SCAO. However, unlike SCAO, LTAO features a more sophisticated wavefront reconstruction technique known as tomographic reconstruction. Tomography involves computing the reconstruction across the 3D turbulence volume, which helps mitigate the anisoplanatism and cone effects. A comprehensive explanation of the principles of AO tomography can be found in Fusco et al. (2001a).

Tomography involves reconstructing the volumetric structure of the turbulence by sampling the atmospheric volume from various partially overlapping view directions (see Figure 2.9). This diversity in viewing angles allows for disentangling the turbulent WFs corresponding to different altitudes and enables their localization in three dimensions.

When applied to LTAO, tomographic reconstruction involves finding a solution for the turbulent WFs at different altitudes, so that when they are summed and projected on a single DM, the discrepancy between them and the corrective DM shape is minimal in the direction of the science target. Solving this problem can then be approached within the framework of regularized least-squares minimization.

Prior information about the layers' strengths, statistical parameters, and altitudes is provided externally to the tomographic reconstructor through atmospheric profiling, which can be performed offline using techniques such as SLODAR. Profiling can be performed by the same AO system that is used for correction or by external instruments (see Cortés et al. (2012); Haguenauer et al. (2020)).

LTAO-enabled observations benefit scientific applications such as monitoring solar system bodies, studying active galactic nuclei (AGN), and observing globular clusters and young stellar objects. A more comprehensive list of LTAO-enabled science applications will be given later in Table 2.2.

A prominent example of an instrument that uses LTAO is the MUSE Narrow Field Mode (NFM) (see Section 2.4.2.2), which benefits from LTAO correction provided by the AO system of ESO VLT-UT4 (Oberti et al. (2018)). The available scientific field in

this regime is only 7.5" × 7.5", but it can achieve a Strehl ratio of more than 5% at 650 nm (R band). MUSE-NFM will be discussed in more detail later in Section 2.4.2. Another example is the narrow-field mode of the Gemini North Adaptive Optics system (GNAO), which provides LTAO correction over 20" × 20" in the scientific field with an expected SR of 30% in the H band. Among the upcoming LTAO systems, the LTAO regime of ELT's High Angular Resolution Monolithic Optical and Near-infrared Integral field spectrograph (HARMONI) should be mentioned (see Thatte et al. (2022)).

Given the limitations of using LGS mentioned earlier, which include insensitivity to tip-tilt and defocus bias due to variations in the Na-layer density, LTAO additionally utilizes an NGS for tip-tilt and defocus correction, typically situated within the scientific field. The defocus correction using NGS will be discussed later in Chapter 3.

2.3.4.3. Multi-Conjugate Adaptive Optics (MCAO)

Although the LTAO addresses some limitations of the SLAO regime, it still corrects within a very narrow FOV. In this framework, MCAO can be seen as a generalization of LTAO that addresses this issue. Like LTAO, it relies on measurements from multiple WFSs that probe several GSs, typically LGSs. However, unlike LTAO, which uses a single DM, MCAO projects the reconstructed turbulent layers onto multiple DMs. These DMs are optically conjugated to corresponding atmospheric layers at different altitudes, predetermined by the system design. The tomographic reconstruction in MCAO is then formulated similarly to LTAO, but with the key difference that the turbulent layers are projected onto multiple DMs to minimize the residual phase.

By simultaneously correcting for turbulence at multiple atmospheric layers, MCAO can address a larger fraction of the turbulent volume, resulting in a correction over a larger FOV than LTAO. As a result, MCAO can achieve more uniform correction within more than 30" × 30". In MCAO, the LGS constellation is positioned slightly outside the science field. MCAO can also utilize NGSs to perform tip-tilt estimation in a tomographic manner, with NGSs positioned inside or outside the LGS field.

A prominent example of an MCAO system is the Gemini Multi-Conjugate Adaptive Optics System (GeMS) (see Rigaut et al. (2013)), which is the first *operational* LGS-based MCAO system implemented on-sky. Currently, it provides a uniform correction over a 85" × 85" FOV with SR ranging from 15% to 50% (in the range of 1 to 2.5 μm). Another example of the MCAO system is the ESO's Multi-conjugate Adaptive-optics Demonstrator (MAD), which is the first MCAO system to be implemented on-sky (see Melnick et al. (2012)). When talking about the upcoming MCAO systems, the MCAO-Assisted Visible Imager and Spectrograph (MAVIS) must be mentioned (McDermid et al. (2020)). It is the future UT4 instrument that will have a science field of 30" × 30" and achieve a Strehl ratio of more than 10% in the V band (551 nm). For the ELT, instruments like the Multi-AO Imaging Camera for Deep Observations (MICADO) (see Davies et al. (2010)) will benefit from the MCAO correction provided by the Multi-conjugate Adaptive Optics Relay for ELT Observations (MORFEO) (see Diolaiti et al. (2016)), with an expected field of up to 50" × 50".

Scientific applications that benefit from MCAO include the study of stellar evolution

and star formation, the physical composition of mid-redshift galaxies, the assembly of early galaxies, and observations of the Sun, as well as planets and moons within our solar system.

A comprehensive overview of the principles of MCAO is provided in Fusco et al. (2001b) and Rigaut & Neichel (2018), which also covers aspects of AO in more detail.

2.3.4.4. Ground-Layer Adaptive Optics (GLAO)

When discussing wide-field AO correction regimes, it is crucial to highlight the GLAO type of AO correction. This approach is the simplest and most robust AO configurations that use multiple GSs (see Rigaut (2002); Hubin et al. (2005)).

Similarly to LTAO and MCAO, GLAO relies on multiple GSs probed by several dedicated WFSs. Guide stars can be both NGS or LGS, with LGS providing more flexibility in selecting the LGS constellation configuration (asterism). As shown in Figure 2.9, in GLAO, the GSs are usually placed far from the scientific field and also far apart from each other. As a result, they probe different, mostly non-overlapping, columns of the atmosphere. The only region where measurements overlap is near the ground. The turbulent wavefront is then reconstructed by averaging the measurements from all WFSs, making the signal from the ground layer (GL) of turbulence the strongest, while the non-overlapping measurements average out to a low value as their turbulent content is statistically decorrelated [4]. Even though uncompensated upper layers still cause wavefront distortion, their combined impact is typically weaker than that of GL.

As a result, GLAO delivers suboptimal correction compared to what could be achieved by optimizing AO performance for a given direction. However, GLAO still offers a uniform [5] improvement of the encircled energy over a fairly wide FOV with respect to non-AO operation. Generally, it decreases the FWHM of a seeing-limited PSF by a factor of 2-3, thus acting as a "seeing improver."

Scientific applications benefiting from GLAO include wide-field and deep-field surveys and observations of galaxy morphology and kinematics. A notable example of a GLAO-enabled instrument is the MUSE Wide Field Mode (WFM) (see 2.4.2.1), where the correction covers a $1' \times 1'$ science field. Another example is GNAO in wide-field mode, which achieves correction over a scientific field of up to $2' \times 2'$. When talking about the upcoming GLAO systems, the ELT's multi-object and multi-integral field spectrograph MOSAIC can be mentioned, which is expected to have a FOV of $40' \times 40'$ (refer to Evans et al. (2015)).

Similar to LTAO and MCAO, LGS-based GLAO utilizes an NGS for tip-tilt sensing. The NGS can be located within or outside the LGS field. This is because GLAO correction is less demanding in terms of performance requirements and can tolerate the errors introduced by the relatively large angular separation of the NGS from the center of FOV.

4. Tomographic reconstruction in GLAO framework is also possible.
5. Since the GL is seen as the same in all view directions.

2.3.4.5. Multi-Object Adaptive Optics (MOAO)

When discussing wide-field AO correction regimes, it is helpful to refer to the MOAO correction (see Rigaut & Neichel (2018)). This regime is not illustrated in Figure 2.9. Essentially, MOAO involves performing SCAO/SLAO but for multiple targets simultaneously. Correction is achieved by operating several deformable mirrors, each conjugated with a different FOV direction. As in the case of SCAO/SLAO, MOAO provides excellent correction only within the isoplanatic patches corresponding to the chosen targets.

In summary, it is possible to recap this discussion as follows :

- **Narrow-field AO correction :**
  - **SCAO :** A single NGS is used to correct in a single direction using one DM. Provides a good correction within the isoplanatic patch of sufficiently bright guide stars
  - **XAO :** Like SCAO, but typically faster and more high-resolution, thus specifically suitable for high-contrast imaging. It can also be two-staged
  - **SLAO :** Like SCAO, LGS is used as a target. It provides good sky coverage at the price of a lower SR due to the cone effect. However, it still suffers from anisoplanatism.
  - **LTAO :** Generalizes SLAO to multiple LGSs. Addresses anisoplanatism and the cone effect by using tomographic reconstruction. However, it is still a narrow field.
- **Medium-field AO correction :**
  - **MCAO :** Basically, the extension of LTAO. It also uses the same tomographic principle. But, unlike LTAO, it utilizes multiple DMs conjugated with different altitudes to correct over larger FOV.
- **Wide-Field AO correction :**
  - **GLAO :** In a way, it is a mix of LTAO and SCAO. Like LTAO, it probes several LGS, but like SCAO, it uses only a single DM in the pupil to correct for the GL. GLAO averages out the anisoplanatic effects at the price of being blind to high-altitude layers, resulting in a quite uniform but very partial correction.
  - **MOAO :** It is equivalent to performing SCAO/SLAO simultaneously in multiple specified directions. The correction is performed by multiple deformable mirrors conjugated in the pupil for different GS directions.

This brief overview of the AO regimes aims to provide a foundation for future discussions on AO systems, their modeling, and potential error sources. Most importantly, the discussion on different types of AO correction sets the stage for a detailed discussion of the two instruments considered in this manuscript : SPHERE and MUSE-NFM, which will be provided in the following section.

2.4. Instruments

2.4.1. SPHERE

Spectro-Polarimetric High-Contrast Exoplanet Research (SPHERE) is an advanced AO and coronagraphic instrument designed for high-resolution, high-contrast observations at the Very Large Telescope (VLT) UT3 (see Beuzit et al. (2019)). Optimized for near-infrared (NIR) and visible wavelengths, SPHERE achieves almost diffraction-limited resolution and superior image quality within a narrow field of view.

2.4.1.1. Science cases

When talking about scientific applications of SPHERE, it excels in high-contrast imaging, enabling the following key types of studies :

1. Direct imaging of exoplanets around bright, nearby stars ($m_R < 10$), focusing on young, giant planets in wide orbits (Chauvin et al. (2017)) ;
2. Investigations of protoplanetary and debris disks around young stars (Ginski et al. (2024)) ;
3. Polarimetric imaging of scattered light to detect faint dust around stars (Engler et al. (2018)) ;
4. Direct imaging of the closest giant stars (European Southern Observatory (2019)) ;
5. Characterization of exoplanet atmospheres (Samland et al. (2017)) ;
6. Research on Solar System objects (European Southern Observatory (2022)), including the minor bodies (European Southern Observatory (2018)) ;

To achieve its scientific goals, SPHERE relies on 3 post-focal instruments and a common extreme AO system, which will be discussed below.

2.4.1.2. Adaptive optics of SPHERE (SAXO)

Due to SPHERE's strong emphasis on exoplanetary observations, achieving very high angular resolution is essential. Direct observation of exoplanets involves studying regions close to their host stars, where the angular separation between the star and the planet is extremely small, while the contrast in luminosity between them is significant (typically more than $1 : 10^6$ depending on the wavelength range and planet type). In this context, AO is absolutely necessary for minimizing image degradation caused by atmospheric turbulence, thereby enabling high-contrast imaging and the direct detection of exoplanets near their host stars.

A major challenge in high-contrast imaging is the presence of bright, variable speckles resulting from atmospheric turbulence and instrumental aberrations. These speckles can mimic or obscure the signal of planets or circumstellar disks. AO helps to reduce the intensity and number of speckles by correcting them in real-time, leading to clearer images and more reliable detection of faint planets.

Since speckles are often caused by relatively high-frequency wavefront aberrations, their correction requires a high-speed, high-resolution AO system, such as SPHERE's XAO system, SAXO (see Fusco et al. (2006)).

SAXO provides near-diffraction-limited images to scientific instruments of SPHERE, both for coronagraphic [6] and imaging systems.

SAXO's main loop (VIS loop) compensates for short-term optical aberrations at frequencies up to 1380 Hz resulting from atmospheric turbulence. It is also capable of correcting defects originating from the telescope or the instrument itself. It uses a 40 × 40 SH WFS operating in the R band, paired with a 41 × 41 actuator high-order deformable mirror (HODM). Additionally, SAXO includes a fast (>800 Hz) tip-tilt mirror (ITTM) to correct rapid tip-tilt variations (see Fusco et al. (2014)).

SPHERE also includes several slow control loops, including the differential tip-tilt sensor (DTTS) loop, which sends commands to a differential tip-tilt plate (DTTP). This compensates for slow movements between the coronagraph's focus and the visible wavefront sensor, caused by thermal shifts, differential dispersion, etc., ensuring precise centering of the star on the coronagraph in the NIR.

In addition, the pupil loop stabilizes the VLT-UT3 pupil at a frequency of 0.1 Hz using a pupil tip-tilt mirror (PTTM) located at the system entrance. This mirror ensures that the pupil image remains stable on the Lyot stop of the coronagraph.

2.4.1.3. Zurich IMaging POLarimeter (ZIMPOL)

ZIMPOL provides classical and differential polarimetric imaging (DPI) with an angular resolution of 22 - 28 mas in the visible spectral range (500 - 900 nm) (see Schmid et al. (2018)). It benefits from the diffraction-limited imaging provided by SAXO and enables imaging through two (spectral) channels equipped with differential polarization capabilities, making it ideal for studying scattered light from circumstellar disks and reflected light from the (atmospheres) of evolved cold exoplanets. Since scattered and reflected light is slightly polarized, it can be isolated from the light of the star by polarization filters, allowing one to study the structure and composition of circumstellar disks or to search for planetary companions. Although optimized for polarimetric imaging, ZIMPOL also offers excellent general imaging capabilities and can also operate in a coronagraphic mode.

2.4.1.4. The InfraRed Dual-band Imager and Spectrograph (IRDIS)

The primary objective of IRDIS is high-contrast imaging for the direct detection and characterization of exoplanets or brown dwarfs, as well as for the study of circumstellar disks. It has a FOV of 11" × 12.5" with a pixel scale of 12.25 mas (see Dohlen et al. (2008)). Unlike ZIMPOL, which operates in the visible range, IRDIS works mainly in NIR, specifically within the atmospheric windows of the Y, J, H, and Ks (0.95 - 2.32 μm)

6. Within this context, coronagraphs refer to both Lyot stops, which obstruct light from a luminous central star, and apodizers, which minimize diffraction patterns generated by the telescope's optics (see Beuzit et al. (2019)).

bands. Observing in this spectral range is valuable for detecting molecular signatures such as water and methane in exoplanetary atmospheres and circumstellar disks. Like ZIMPOL, IRDIS can also perform coronagraphic observations.

IRDIS implements simultaneous differential imaging (SDI) and offers multiple imaging modes, including classical imaging (CI), dual-band imaging (DBI), and dual-polarization imaging (DPI). This is achieved by projecting two parallel beams onto a 2k × 2k cryogenic IR detector. In these configurations, the detector captures images from two channels, each corresponding to slightly distinct wavelengths or two orthogonal polarizations.

In DBI, starlight, which has a relatively smooth spectrum, is captured inside and outside the planet's absorption band (see Vigan et al. (2010)). Subtracting these two simultaneous images removes most of the starlight, leaving primarily the planetary signal in the residuals.

In DPI mode, IRDIS employs a simpler approach to measuring polarization compared to ZIMPOL's complex polarization modulation and demodulation technique. IRDIS splits the incoming light into two orthogonal polarization states using a polarization beam splitter, capturing both polarization states simultaneously in two separate channels on one detector (see de Boer et al. (2020)).

In addition to these modes, IRDIS can operate in long-slit spectroscopy (LSS) mode, which is crucial for exoplanet characterization (see Vigan et al. (2008)). LSS covers the 0.95 - 2.32 μm range at low spectral resolution ($R \approx 50$) and the 0.95 - 1.65 μm range at medium spectral resolution ($R \approx 350$). This mode takes advantage of the fact that speckles vary with wavelength while the exoplanet's position remains constant, allowing for the separation of planetary signals from starlight speckles.

2.4.1.5. Integral Field Spectrograph (IFS)

IFS is a low-resolution NIR spectro-imager unit within SPHERE (Claudi et al. (2008)). It enables to capture the spectrum in every pixel within its field of view, providing simultaneous spatial and spectral data. IFS offers a spectral resolution of $R \approx 50$ in the 0.95 - 1.35 μm range (Y-J) or $R \approx 30$ in the 0.95 - 1.65 μm range (Y-H), which is ideal for detecting planetary features. IRDIS can be operated simultaneously with IFS, making the latter highly complementary to IRDIS. However, IFS focuses only on the inner region around a star, providing detailed spectral information over a smaller field of view (1.73" × 1.73") (see Mesa et al. (2015)).

In summary, it is important to highlight that IRDIS will play a central role in the upcoming discussion, as the non-coronagraphic calibration PSF datasets, recorded specifically by IRDIS, will be extensively utilized in Sections 4.6.1, 4.7.1, and 4.8.4. Therefore, when referencing SPHERE later in this thesis, the focus will primarily be on IRDIS.

The primary motivation for selecting this instrument is the availability of high-quality PSF and relevant telemetry datasets, which perfectly align with the data-oriented strategy adopted for PSF modeling in this research. Furthermore, IRDIS

offers high angular resolution, allowing the resulting PSFs to be well sampled in the NIR, achieving Nyquist sampling at 0.95 μm. Moreover, modeling SPHERE's XAO correction serves as a valuable complement to the LTAO-based MUSE NFM modeling discussed later. Exploring diverse types of AO correction affirms the universality of the approach to PSF modeling utilized in this work. Therefore, these factors collectively make IRDIS an ideal instrument for studying the methods discussed in Chapter 4.

2.4.2. MUSE

The Multi Unit Spectroscopic Explorer (MUSE) is an integral field spectrograph installed on VLT-UT4, known for its high spatial resolution and extensive spectral coverage in visible and NIR (see Bacon et al. (2010)). One of MUSE's key strengths is its ability to capture a spectrum for every pixel (called "spaxel") within its FOV, providing simultaneous spatial and spectral data in the visible range. This capability enables detailed studies of the composition, kinematics, and physical properties of various astronomical targets. Its unique design and versatility have made MUSE one of the most highly requested instruments in the astronomical community.

The hyperspectral functionality of MUSE is achieved by dividing its field of view into 24 image segments (channels), which are further sliced into 48 mini slits. These slits are fed into a spectrograph that uses Volume Phase Holographic Gratings (VPHGs) to generate spectra for each slit. The resulting spectral resolution per pixel ranges from R=1770 at 465 nm to R=3590 at 930 nm. MUSE covers a spectral range from 465 nm to 930 nm.

However, the full potential of MUSE is unlocked when using AO correction, which is enabled by the Adaptive Optics Facility (AOF) of UT4 (Oberti et al. (2018)). AOF utilizes four 22 W sodium lasers as LGSs, which greatly enhance the sky coverage available for AO correction. Additionally, UT4 is equipped with a 1.12-meter adaptive deformable secondary mirror (DSM) with 1156 useful actuators. The AO correction for MUSE is managed through the GALACSI module (Hartke et al. (2020)).

MUSE operates in two modes : Wide-Field Mode (WFM) and Narrow-Field Mode (NFM), which are described below. Note that each mode can also benefit from the AO correction.

2.4.2.1. WFM

In WFM mode, MUSE offers a field of view of $1' \times 1'$ with a spatial resolution of 200×200 mas per spaxel. The AO correction in this mode is implemented as GLAO (see Section 2.3.4.4). Since GLAO uniformly improves the FWHM across the entire field, WFM is well-suited for observing extended objects and relatively broad regions of the sky.

2.4.2.2. NFM

In contrast, NFM provides a smaller field of view, about 7.5"×7.5", with enhanced spatial resolution (25 × 25 mas per spaxel), making it ideal for detailed observations of compact objects and for studying fine structures. NFM can benefit from the LTAO correction (see Section 2.3.4.2), which provides a much stronger improvement of SR compared to WFM.

It is worth noticing that when AO is enabled, NFM also relies on an NGS observed in NIR for tip-tilt and focus correction. This aspect will be covered in more detail in Section 3.1.

A quantitative comparison of both regimes is provided in Table 2.1.

	Wide field mode (WFM)	**Narrow field mode (NFM)**
MUSE field of view	1' × 1'	7.5" × 7.5"
Size of an angular spaxel	0.2" × 0.2"	0.025" × 0.025"
AO performance	× 2 ensquared energy in 1 spaxel at 750 [nm]	> 5% Strehl ratio at 650 [nm]
Type of AO correction	GLAO	LTAO
NGS star brightness	< 17.5 magnitude (R-band)	< 18.5 magnitude (J+H bands)
NGS star location	52" to 105" off-axis → 70% sky coverage	On-axis, within 13.3" diameter
LGS separation	4 LGS 64" off-axis	4 LGS 10" off-axis

TABLEAU 2.1. – Characteristics of the Narrow-Field and Wide-Field modes of MUSE enabled by the GALACSI module.

2.4.2.3. Science cases

MUSE has contributed to a wide range of scientific investigations (Bacon et al. (2023)). However, the type of science performed with MUSE depends on the selected regime. For example, the strength of NFM lies in its high resolution, while WFM excels in offering a large FoV. Hence, WFM is designed for broader surveys and observations of large-scale structures such as galaxies, galaxy clusters, and extended regions of space. It is worth mentioning that using a secondary deformable mirror and an AO design in which the NGS is selected outside the scientific FOV has enabled an optical design that minimizes the number of optical surfaces between the telescope and the instrument and thus maximizes optical throughput. This makes MUSE a very sensitive instrument suited for ultra-deep fields, which allows it to probe very faint and distant galaxies. Meanwhile, NFM is better suited for resolving fine-scale features and conducting detailed observations of small regions, such as galaxy cores, star clusters, protoplanetary disks, and environments near black holes.

Table 2.2 lists some examples of scientific applications that can be addressed with MUSE.

Scientific application	WFM	NFM
Star clusters and stellar populations	Observing entire galaxies or globular clusters (Kamann et al. (2021)), making it suitable for broad surveys of stellar populations and tracking star formation across large regions.	Effective for resolving individual stars in densely populated regions, such as globular clusters or galaxy cores, allowing for detailed studies of their dynamics and composition (Pechetti et al. (2024)).
Protoplanetary disks	Insufficient angular resolution for resolving protoplanetary disks.	Enables detailed studies of envelopes or protoplanetary disks around young stars (Girard et al. (2020); Aru et al. (2024)).
Supernovae remnants	Useful for mapping supernova remnants and their surrounding environments (Gúrpide et al. (2022)).	Allows observation of fine structures within supernova remnants, such as shock waves, to understand their interaction with the environment and chemical element distribution.
AGN and SMBHs	Useful for observing large-scale structures in and around active galactic nuclei (AGN) and supermassive black holes (SMBHs), including jets and surrounding gas.	Enables detailed mapping of regions near SMBHs and AGN, helping to understand how these black holes influence their host galaxies (Knapen et al. (2019)), including the Milky Way.
Galaxy studies	Effective for mapping large-scale galactic structures, facilitating the study of galactic kinematics, galaxy interactions, and overall evolution (Bacon et al. (2017)).	Allows for detailed studies of galaxy kinematics, the dynamics of stars and gas, and resolving individual stars within globular clusters or nearby galaxies to observe internal kinematics and dynamics (Usher et al. (2021)).
Planetary nebulae	Useful for studying the general structure and surroundings of planetary nebulae, characterizing the overall motion of the gas, interactions with the environment, and approximate chemical distributions (Monreal-Ibero & Walsh (2020)).	Enables the study of fine structures, such as filaments and shock waves, revealing localized chemical gradients and understanding the interaction of ejected material with nearby stars (Haworth et al. (2023)).

TABLEAU 2.2. – The list of scientific cases for both MUSE WFM and NFM regimes.

Of course, the list of scientific applications is not limited to those mentioned in Table 2.2. However, these examples are sufficient to demonstrate the power and versatility of MUSE as an instrument.

As will be discussed in Section 4.1, many of these applications can benefit significantly from precise knowledge of the PSF morphology during the post-processing stage (as will be discussed in Section 4.1), especially in the case of observations as-

sisted by AO. This makes modeling MUSE PSFs highly valuable, which is a strong motivation for selecting this instrument for future analyses in this thesis.

In Fusco et al. (2020), a technique for modeling AO-corrected PSFs for WFM was successfully demonstrated. However, my work will focus on modeling AO-corrected PSFs for NFM, which presents a distinct challenge. The key difference compared to WFM is that NFM-driven PSFs undergo a much higher degree of AO correction, resulting in more complex morphologies. Accurate modeling of these PSFs largely depends on how well the AO system behavior is modeled, a task that differs greatly from the previously discussed IRDIS, where the PSF morphology is predominantly determined by the diffraction pattern. Moreover, unlike IRDIS, MUSE NFM PSFs are undersampled, adding another layer of complexity to the problem. This combination makes modeling NFM PSFs a uniquely challenging task.

In summary, this research focuses on two distinct subsystems and regimes : SPHERE IRDIS and MUSE NFM. These instruments support different scientific objectives and produce images influenced by two very different types of AO correction. Therefore, modeling PSFs for both systems using one approach will serve as a valuable test of the universality of the approaches which are extensively studied in Chapter 4.

However, AO correction is never perfect. Thus, it is crucial to account for all imperfections and errors in AO-assisted imaging to accurately model PSFs. To achieve this, it is vital to understand all potential sources of error that may arise in telescopes. Therefore, the next section will provide an overview of these potential error sources.

2.5. Error sources

As mentioned, a thorough understanding of the telescope's behavior with all its subsystems is essential for precise PSF modeling. Consequently, this necessitates comprehending potential sources of error that may exist within a telescope. Given the complexity of modern ground-based observatories and their instruments, the list of potential sources of error can be extensive. Moreover, the significance and prioritization of these errors can be done differently depending on the specific context. In this discussion, we focus on errors that can affect image quality. When determining these errors, it is also important to address the following aspects :

— What is their origin?
— How powerful are they?
— At what temporal frequencies do they occur?
— In which optical path do they arise?
— What spatial frequencies in the wavefront do they affect?
— Can they be detected and corrected by an AO system?
— Where in the optical path do they occur?
— Are they well understood so that they can be effectively described and predicted using an analytical model?

These factors will be briefly addressed in this section, with further elaboration in Section 4.3.1. For clarity, in this section, the error sources are organized into three big categories :

— **AO-related errors :** These result from the imperfect compensation of atmospheric turbulence by the AO system, including errors related to AO-system calibration.
— **Telescope/instrument-related errors :** Primarily caused by structures within the telescope and instruments, these errors can exist independently of AO.
— **Optical errors :** These arise from the optical components within the telescope, AO system, and instruments. While the previous categories are also, in essence, optical errors, this category refers primarily to quasi-static, usually path-specific aberrations.

Note that alternative categorizations may be valid, as many error sources span multiple categories. However, the discussion will proceed using this framework.

2.5.1. AO-related errors

2.5.1.1. Imperfect AO Correction.

These errors arise from the inherent limitations of real AO systems in fully correcting wavefront distortions caused by atmospheric turbulence due to engineering and physical limitations. In this context, the following common AO-related errors can be listed :

— Fitting error;
— Spatio-temporal error;
— Anisoplanatism;
— Differential refraction;
— Chromatic aberration;
— Aliasing;
— WFS noise;

As already mentioned, a more detailed analysis of each error will be provided in Section 4.3.1. For now, these are grouped under the term "Imperfect AO correction".

These errors are intrinsic to all AO systems and cannot be fully eliminated. It also spans all spatial and temporal frequencies covered by an AO system. The specific contribution of each error varies based on the design of the AO system and the type of correction applied. These errors manifest themselves in the residual phase, which, in turn, plays a critical role in determining the closed-loop PSF morphology. Consequently, accurately modeling the effects of imperfect AO correction is essential but challenging, as it requires replicating the AO system's behavior very realistically. This complexity is a key reason why PSF reconstruction methods, discussed later in Section 4.2, have not always been successful in the past.

2.5.1.2. Misregistration

Misregistration refers to errors associated with the misalignment between the WFS and DM actuators grid. More generally, it includes imperfections in the optical transformations between the DM plane and the WFS pupil sampling plane (for example, the lenslet array for an SH). Moreover, it is not limited to misalignments regarding the shift of the image grid. It may also be rotation, magnification, and low-order distortion misalignments. This misalignment can introduce biases that result in less accurate AO correction and less stable closed-loop operations. Although this error source can vary with the telescope pointing, it remains relatively static and reproducible, making it amenable to calibration. Techniques have been developed to calibrate misregistration, as shown in Heritier et al. (2021). A more detailed discussion of misregistration as an error source can also be found in Heritier (2019).

2.5.2. Telescope/instrument-related errors

2.5.2.1. Phasing Errors

This type of error is characteristic of all observatories with segmented mirrors, such as Keck, JWST, and the upcoming ELT, whose primary mirror (M1) will be segmented, as it is currently not technologically feasible to manufacture a monolithic optical mirror of such a large diameter. Phasing errors arise because each segment is physically separated from the others. However, to approach close to a diffraction-limited PSF, all segments must form a smooth and continuous optical surface together.

Aligning segments to achieve this goal is challenging. For example, inductive edge sensors installed on each M1 segment can align the segments with a wavefront error (WFE) of 50 nm across the pupil. However, finer control must rely on more sensitive optical techniques. Phasing errors manifest themselves as discontinuous wavefront patterns similar to those caused by LWE in the closed-loop. Like LWE, these discontinuities cannot be measured with classical SH WFS, necessitating the use of more advanced techniques such as those mentioned for LWE. However, unlike LWE, phasing errors generally exhibit lower WFE and are more stable temporally. In addition, in certain systems, phasing errors can be calibrated using artificial calibration sources, which makes them generally easier to control.

2.5.2.2. Scalloping

Scalloping is an error specific to segmented telescopes. In the context of the ELT, this error arises because M4 cannot fully compensate for defocus offloaded from the primary mirror (M1) (see Louarn et al. (2016)). Each segment of M1 has defocus as a degree of freedom. Due to the limited number of actuators on M4, it cannot completely replicate the shape of M1, resulting in residual defocus per segment. The resulting error manifests itself as periodic wavefront distortions, appearing as a series of small, scallop-like deviations across the mirror surface. However, this error is generally not expected to exceed 50 nm RMS WFE, and thus, it is not likely to significantly impact the performance of the ELT.

2.5.2.3. Missing Segments

A potential source of error in the AO system is the absence of some segments from the segmented primary mirror of the telescope. This error can become significant if ten or more segments are missing and if the AO system has not been calibrated to account for these missing segments (see Louarn et al. (2016)). However, the absence of only a few segments should not significantly affect the overall performance of the system.

Depending on the area of missing segments or their size, they can induce quasi-static high-frequency speckles in focal-plane PSF. In this scope, the effects of missing segments can be more noticeable in coronagraphic high-contrast observations. Nevertheless, the impact of missing segments can be mitigated with apodization (Vigan et al. (2013); Leboulleux et al. (2022)).

2.5.2.4. Vibrations

In AO systems, vibrations can originate from various sources, such as wind-induced telescope shake, mechanical movements of the telescope structures, or components within instruments. Consequently, their temporal spectrum may cover a wide range of frequencies depending on the source. For example, wind-induced vibrations can be significantly slower than the AO loop temporal bandwidth, while motorized components within instruments can cause vibrations at frequencies exceeding the AO

loop rate. In the spatial frequencies domain, vibrations typically manifest as low-order aberrations, such as tip-tilt, defocus, astigmatism, coma, and trefoil. This occurs because mechanical vibrations propagate through the optical system, deforming the mirror along its resonant modes. These deformations result in low-order aberrations, as seen from the perspective of the WFS.

Depending on the AO system in question, vibrations can be a fairly notable error contributor. For example, in the context of XAO, vibrations can be particularly influential, as they affect the centering of the coronagraph (Beuzit et al. (2019)).

Vibrations can be mitigated by mechanical damping or with advanced control strategies (see Kulcsár et al. (2012); Guesalaga et al. (2013); Muradore et al. (2014)). However, vibrations that are fast or strong enough are not sufficiently rejected by the AO loop

2.5.2.5. Dome Seeing

Dome seeing typically arises from thermal turbulence caused by free and weakly mixed convection within the dome of the telescope (see Woolf (1979)). Dome seeing can be detected and corrected using AO. However, its amplitude can be significant, causing DM saturation. As a result of its large amplitude, this error term can also significantly deteriorate the image quality.

Dome seeing is usually mitigated by insulating the wall and floor surfaces, adding vents, enforcing air circulation, and applying active cooling to various heat sources, such as electronics. For example, the ELT dome has been designed to allow sufficient ventilation, thereby minimizing dome seeing. It is also equipped with ventilation louvers and a windscreen that together enable the management of airflow inside the dome.

Although dome seeing can be relatively easily prevented through design measures, it remains challenging to model accurately. Modeling typically requires detailed finite element aerodynamic simulations (see Sebag & Vogiatzis (2014)). Direct measurement is also non-trivial, though there are studies dedicated to characterizing it (see Lai et al. (2019)).

2.5.2.6. Low Wind Effect (LWE)

LWE can be regarded as a special case of dome seeing. However, this effect exhibits some very distinctive features that make it unique and worthy of special attention. LWE occurs when warm air, preheated by the surrounding ground during the day, is blown into the telescope's dome during night observations. This air is cooled upon contact with the colder spider structures supporting the secondary mirror, which remains cool from being in a closed dome during the day prior to observations.

LWE is most pronounced when the wind is slow, allowing sufficient time for thermal exchange between the spiders and the air before it is blown away. The resulting temperature gradient at the edges of the spider induces abrupt changes in air density,

causing phase discontinuities to appear in the pupil plane. A more detailed overview of the physics behind LWE is provided in Martins et al. (2022).

Before propagating in the AO loop, the LWE introduces an additional disturbance on top of the atmosphere, which includes not only sharp WF discontinuities but also optical path difference (OPD) spanning various spatial frequencies, including high spatial frequencies, especially if the wind flow is turbulent (Holzlöhner & Brinkmann (2020)).

When propagated through the AO loop, these sharp discontinuities along the spiders create specific phase patterns resembling per-island tip-tilt and piston, with "islands" referring to wavefront regions separated from each other by the spiders. As a result, this effect leads to the apparent "splitting" of the focal-plane PSF, reducing the quality of the image. Further discussion of these effects can be found in Pourré et al. (2022).

LWE has been observed on multiple modern telescopes, including the VLT, with initial observations on UT3 using the SPHERE instrument (see Sauvage et al. (2016); Milli et al. (2018)). However, in the case of ELT, LWE is expected to be much more pronounced due to the scale and thickness of the spiders. Thicker spiders have more surface area for contact with air, resulting in a larger accumulation of OPD when light is propagated along the spider's edge. But most importantly, the ELT spiders will be thicker than the coherence length of the wavefront distortions, resulting in the wavefront islands being vastly decorrelated from each other [7]. When propagated through the AO loop, it may result in a severe differential piston appearing between islands (petaling effect). Although, in general, these petals may appear without LWE just due to the WF decorrelation, the presence of LWE makes them more powerful (see Sauvage et al. (2022); Levraud et al. (2024a)).

LWE can be partially mitigated. For example, on the VLT, LWE was significantly reduced by applying a special thermal coating to the spiders to equalize the temperature gradient (see Milli et al. (2018)) between air and the spiders. The ELT may require additional measures, such as active heating of the spiders, to reduce thermal difference. Moreover, if measured accurately in a slower secondary AO loop, LWE can be efficiently corrected for by DM.

However, accurate measurement of LWE is challenging. The SH WFS, as well as any differentiating WFS (for example, modulated pyramid WFS), is insensitive to the differential piston features produced by LWE since such sensors can only sense the first-order spatial derivative of the wavefront, which almost zeros out localized constant piston-like phase shifts. In this context, several recent WFS solutions have been proposed to address this issue, including :

- Fourier-based wavefront sensors (Fauvarque et al. (2016))
  - Unmodulated pyramid WFS (PyWFS) (Engler et al. (2022); Levraud et al. (2024a)) ;
  - Zernike WFS (N'Diaye et al. (2016)) ;
  - Bi-O-edge (Vérinaud, C. et al. (2024)) ;
- Photonic lantern (Norris et al. (2020)) ;

7. Mostly applies to ELT systems like the SCAO regime of HARMONI.

— Modified Shack-Hartmann WFS (Mazzoleni et al. (2008));
— Dispersed-fringe wavefront sensors (van Dam et al. (2016));
— Focal-plane-based WFS techniques (Agapito et al. (2022); Korkiakoski et al. (2014));

Although each of these WFSing techniques deserves a detailed discussion, it is beyond the scope of this manuscript. However, for slightly more details on focal-plane-based techniques, refer to the Introduction section of the embedded paper on page 77. Furthermore, Section 4.7.1.2 presents a method to retrieve LWE from PSF morphology and predict LWE-induced wavefront patterns based on reduced telemetry.

In summary, LWE is a complex aberration that arises under specific conditions and can significantly impact scientific imaging. In the context of upcoming giant telescopes, this error contributor will become even more critical. Since the telescope's design cannot fully prevent it, adaptive compensation is necessary, and accurately measuring it presents a unique challenge, making LWE a very important error contributor to consider.

2.5.3. Optical Errors

This is a broad category of error sources. In the context of this discussion, it generally refers to such error sources as misalignment of optical components, OPD shifts caused by thermal expansion of components, or mechanical stress that alters the shape of optical surfaces. In addition, polishing errors, physical damage to optical components, and chromatic aberrations caused by dispersion can also be added to this list.

The impact of these errors depends significantly on the placement of the optical component within the optical path and the size of the light beam footprint on the component. The optical component's position in either collimated or convergent/divergent beams of light also plays a role in this case. Consequently, the resulting WFE can range from a few nanometers to a few microns RMS, depending on the specific circumstances. The spatial frequencies affected by optical errors can also span a wide range. For instance, a misplaced or decentred lens will primarily introduce low-order wavefront aberrations, while mechanical stress at the edges where a component is fixed can cause rather high spatial-frequency aberrations.

Despite their diversity, these errors are generally static or quasi-static, well understood and therefore do not pose a significant threat to overall performance. They can often be calibrated using artificial calibration sources. However, addressing these errors may occasionally require intervention in the system.

All this discussion on error sources is summarized in Table 2.3. The objective of this table is to compile all the errors mentioned above, categorize them, and detail their spatial and temporal characteristics. Furthermore, the table recaps the questions posed at the beginning of this section, specifically regarding the origin of each error source, the ability of the AO system to compensate for it, and to which extent this error contributor can be understood and modeled.

Type	Example	Common/non-common path	Correctable by classical AO?	Affected spatial freqs.	Affected temp. freqs.	Approximate induced WFE (OPD RMS)	Can be modeled analytically?	How good can be measured?	Can be compensated /prevented?
AO-related errors									
	Imperfect AO correction	Yes/No	No, results from AO limitations	All frequencies	0.1 - 2000 Hz	< 200 nm < 500 nm	E.g., Fourier-based models (Sec.(4.3))	From residual phase	Always present
	Misregistrations	Yes/No	No, calibration needed	All frequencies	Quasi-static	< 20 nm (Heritier et al. (2021)) < 50 nm		E.g., using SPRINT	Calibratable
Telescope/instrument-related errors									
	Co-phasing	Yes/No	No, needs a separate loop or calibration	High	Minutes/hours	< 50 – 100 nm (Lamb et al. (2021)) (for Keck)		Needs special WFS	
	Scalloping	Yes/No	Yes, but correction is limited by DM resolution	High	Minutes/hours	< 50 nm (Louarn et al. (2016))		Limited by WFS resolution	Compensation limited by DM
	Missing segments	Yes/No	No, calibration / intervention needed	High	Quasi-static	< 50 – 100 nm (Louarn et al. (2016))			
	Windshake	Yes/No	No, solved by mechanical design / additional loop	Low	< 100 Hz, depends on the vibrating structure's size	< 100 nm if suppressed (Courtney-Barrer et al. (2022))			
	Motorized components	Both possible	No, solved by mechanical design / additional loop	Low	> 100 Hz, sometimes faster than atmosphere	< 100 nm if suppressed (Courtney-Barrer et al. (2022))			Damping
	Dome seeing	Yes/No	Yes, also solved by dome design	Similar to the atmosphere	< 10 Hz	VLT : < 5 μm ELT : < 15 μm	Needs CFD simulations		Ventilation
	Low wind effect	Yes/No	Partially, still needs an additional loop	Low + very high	~1 - 100 Hz	VLT : < 3 μm, < 200 nm closed-loop (Milli et al. (2018)) ELT : < 10 μm, < 500 nm closed-loop (Sauvage et al. (2022))	CFD + E2E simulations are better	Needs special WFS	Coatings / active thermal control / AO compensation
Optical errors									
	Components misalignment	Both possible	Yes, in common path. In non-common, calibration / intervention is needed	Mostly low, can be high on the edges	Quasi-static	Largely depends on the position of a component, system type, cannot be generalized	E2E simulations		May require intervention
	Chromatic aberrations	Both possible	No, solved by optical design	Low	Quasi-static	-//-	-//-		
	Thermal expansion & contraction	Both possible	No, solved by optical design / thermal control	Low	Minutes/hours	-//-	-//-		

TABLEAU 2.3. – The list of different error sources common to AO-assisted telescopes. "-//-" indicates "same as in the previous row".

When discussing the table above, it is important to note that while this table is comprehensive, it is not exhaustive. Numerous second-order error sources specific to certain AO systems are not mentioned here. Furthermore, the magnitude of the listed errors may vary between AO systems due to different AO correction types, target angular resolutions, spectral bandwidths used for observations or WFSing, and other factors.

However, the table still outlines general trends and describes the typical temporal and spatial behavior of the listed errors. It also provides a rough estimate of how well some of these errors can be prevented or compensated for, helping to sort them in terms of severity.

For example, the table shows that LWE, while it evolves slower than atmospheric turbulence, contains a very specific mixture of low and high spatial frequencies (see Figure 8 in Levraud et al. (2024b)). In addition, it can introduce significant WFE, especially on the ELT, and therefore seriously affect the quality of images (see Figure 2.10).

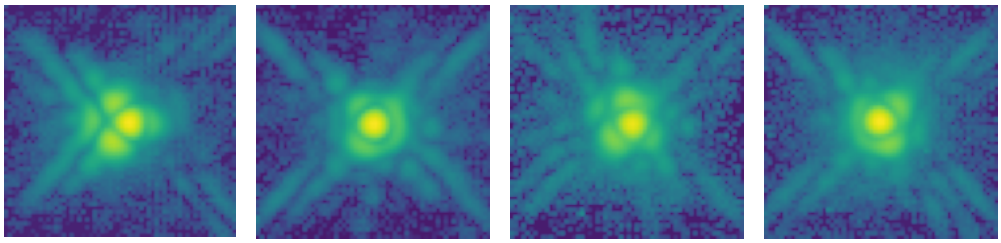


FIGURE 2.10. – Examples of real PSFs (log-scale) significantly influenced by strong LWE recorded with SPHERE IRDIS. The prominent LWE-induced distortions are notable in the region of the first Airy ring.

Since it is impossible to entirely prevent LWE and still non-trivial to measure it precisely, LWE can be considered a significant error contributor. In the context of this work, LWE needs to be included in models, retrieved from on-sky data, or even predicted, as will be shown in Section 4.8.4.3.

Another example is the imperfect AO correction. Although it is not the dominant source of error in most cases, it is inherent in every AO system and affects all spatial and temporal frequencies. Imperfect correction significantly influences the PSF morphology, making it a critical factor to include in PSF modeling to ensure that it is realistic and physically accurate.

In this context, it is important to note that Chapter 4 will focus primarily on two error sources : the LWE and imperfect AO correction, both of which will be extensively modeled.

2.5.4. NCPAs

The notion of NCPAs

When discussing Table 2.3, it is important to highlight another crucial error term. Many of the listed error sources can be path-specific, meaning that they originate

in a specific optical path. In this context, "path" refers to the optical channel that propagates light towards any detector, which can be part of a scientific instrument, a WFS, or even an auxiliary technical imager (see Hartung et al. (2003)). As discussed in Section 2.3.1, AO systems divide incident light and then direct a part of it toward the WFSs for measurement purposes. As a result, an AO system contains multiple optical paths, the (non-common) WFSing path and (common) scientific path, each characterized by its own distinct set of intrinsic aberrations.

Through the feedback inside the closed AO loop, the uncalibrated aberrations from the non-common path can propagate through the AO correction and affect a common path. In other words, the WFS measures an aberration within its optical path and sends corrective commands to the DM to compensate for it. As a result, the science image experiences unnecessary compensation for an aberration that is not present in the scientific optical path, resulting in error. The opposite situation is also possible : some aberrations that can be visible in the science path remain completely unseen by the WFSs and thus unmeasured and uncompensated. This type of errors is usually referred to as *Non-Common Path Aberrations (NCPAs)*[8]. NCPAs might arise from AO calibration inaccuracies or from optical aberrations. In addition, vibrations can also be path-specific.

Impact of NCPAs on imaging

NCPAs can have a substantial impact on scientific images. Therefore, they need to be accurately measured or realistically modeled to enhance imaging quality and facilitate the post-processing of scientific data. The precise knowledge of NCPAs is crucial to a comprehensive understanding of the AO system. This also means that since NCPAs affect the imaging, information about them is inherently present in focal-plane PSFs collected in science or WFSing optical paths. Therefore, multiple studies have attempted to extract NCPA directly from PSF (e.g., Hartung et al. (2003); Mugnier et al. (2008); Robert et al. (2008); Lamb et al. (2018); Beltramo-Martin et al. (2020)).

NCPAs in systems with multiple WFSs

It is also important to note that NCPAs can arise not only between WFS and scientific optical paths but also between multiple WFSs. For instance, in the case of MUSE-NFM, where LGS WFSs are complemented by a separate NGS WFS. The NGS-based WFS can detect aberrations within the LGS loop that are invisible to the LGS loop. These include not only tip-tilt and defocus drift but also NCPAs. In this configuration, one WFS can be used to sense aberrations coming from another WFS. In this context, it is also convenient to turn one WFS into an imager to enable the retrieval of NCPAs and unseen modes directly from the PSF morphology.

8. When dealing with multiple off-axis WFS, common path aberrations occur that do not affect the WFS and the science path in the same way. These may be called non-common field aberrations (NCFAs).

This is exactly what will be covered in the next chapter of this manuscript, where I will cover the issue of measurement and compensation of NCPAs (in the context of the AOF) using PSFs in more detail. I also propose an improvement of the technique to measure NCPAs using PSFs collected at the focal plane of a low-order WFS used by MUSE-NFM. The chapter starts with a brief introduction, followed by an embedded paper, in which I am the first author.

3. Focal-plane wavefront sensing

The previous chapters reviewed various sources of error, including NCPAs. Although AO-related errors will be explored in more detail in Chapter 4, this chapter focuses more on NCPA estimation. In addition, it addresses unseen errors that are invisible to one type of WFS but detectable by another, specifically the defocus drift caused by density fluctuations in the sodium (Na) layer. As a main tool to achieve this goal, this chapter utilizes PSFs captured in the focal plane of the WFS, operated as an imager. This will be further elaborated below.

Following the paper, this chapter delves into topics such as the selection of non-linear optimization algorithms, parameter coupling, and uncertainties in parameter estimation. The forthcoming discussion primarily covers the theoretical foundations, serving a dual purpose : justifying decisions made in the paper and laying the background for Section 4.7, which will heavily rely on the non-linear optimization methods.

3.1. Striving towards robust phase diversity on-sky Implementing LIFT for VLT/MUSE-NFM

This section includes the paper on focal-plane wavefront sensing, which discusses the use of linear focal-plane techniques (LIFT) for measuring slow-changing low-order aberrations applied to the MUSE-NFM instrument in an on-sky setting. Additionally, the paper introduces a method to increase the robustness of estimation by calibrating NCPAs using a focal-plane approach similar to LIFT but based on a different estimation principle. The proposed techniques were validated in real on-sky experiments, demonstrating their effectiveness. This paper is referred to as *"embedded paper"* within this manuscript. With this being said, it is now possible to proceed with the paper itself.

A&A, 687, A221 (2024)
https://doi.org/10.1051/0004-6361/202449860
© The Authors 2024

Astronomy & Astrophysics

Striving towards robust phase diversity on-sky

Implementing LIFT for VLT/MUSE-NFM

Arseniy Kuznetsov[1,2,3], Sylvain Oberti[1], Benoit Neichel[2], and Thierry Fusco[2,3]

[1] European Southern Observatory, Karl-Schwarzschild-str-2, 85748 Garching, Germany
e-mail: akuznets@eso.org
[2] Aix Marseille Univ, CNRS, CNES, LAM, 13013 Marseille, France
[3] DOTA, ONERA, Université Paris Saclay, 91123 Palaiseau, France

Received 5 March 2024 / Accepted 23 May 2024

ABSTRACT

Context. The recent IRLOS upgrade for VLT/MUSE narrow field mode (NFM) introduced a full-pupil mode to enhance sensitivity and sky coverage. This involved replacing the 2 × 2 Shack-Hartmann sensor with a single lens for full-aperture photon collection, which also enabled the engagement of the linearized focal-plane technique (LIFT) wavefront sensor instead. However, initial on-sky LIFT experiments have highlighted a complex point spread function (PSF) structure due to strong and polychromatic non-common path aberrations (NCPAs), complicating the accurate retrieval of tip-tilt and focus using LIFT.
Aims. This study aims to conduct the first on-sky validation of LIFT on VLT/UT4, outline challenges encountered during the tests, and propose solutions for increasing the robustness of LIFT in on-sky operations.
Methods. We developed a two-stage approach to focal-plane wavefront sensing, where tip-tilt and focus retrieval done with LIFT is preceded by the NCPA calibration step. The resulting NCPA estimate is subsequently used by LIFT. To perform the calibration, we proposed a method capable of retrieving the information about NCPAs directly from on-sky focal-plane PSFs.
Results. We verified the efficacy of this approach in simulated and on-sky tests. Our results demonstrate that adopting the two-stage approach has led to a significant improvement in the accuracy of the defocus estimation performed by LIFT, even under challenging low-flux conditions.
Conclusions. The efficacy of LIFT as a slow and truth focus sensor in practical scenarios has been demonstrated. However, integrating NCPA calibration with LIFT is essential to verifying its practical application in the real system. Additionally, the proposed calibration step can serve as an independent and minimally invasive approach to evaluate NCPA on-sky.

Key words. instrumentation: adaptive optics – methods: numerical

1. Introduction

Adaptive optics (AO) represents a significant advancement in the field of optical imaging, enabling the correction of wavefront distortions that occur as light passes through turbulent media or imperfect optical systems. The wavefront is corrected using deformable mirrors (DM) based on measurements performed by wavefront sensors (WFSs).

In astronomical AO systems, incoming light is usually divided into distinct optical paths: one fraction is channelled towards scientific instruments (science path), while another fraction is directed to WFSs (AO path). However, this scheme introduces a possibility for path-specific aberrations termed NCPAs (see Mugnier et al. 2008; Lee et al. 1997). For instance, aberrations that occur outside the AO optical path remain unseen by WFSs and, thus, they cannot be compensated for. Conversely, aberrations that come from WFSs are not inherent to the scientific path but can still affect the science image being propagated through the AO feedback loop. The uncalibrated NCPAs can prominently deteriorate the quality of the resulting scientific PSFs, preventing the AO system from delivering a high Strehl ratio (SR). This underscores the importance of measuring and compensating for NCPAs.

In this framework, leveraging the focal-plane point spread function (PSF) for wavefront sensing (WFSing) and phase retrieval emerges as an appealing option. This approach facilitates extracting information directly from post-AO PSFs obtained close to the science path. As a result, it allows for the measurement of specific NCPAs that impact the quality of scientific images.

In addition, post-AO PSFs imprint the information about phase discontinuities (Lamb et al. 2017). These effects are typically associated with co-phasing errors Chanan et al. (1998) or with the low wind effect (LWE) Sauvage et al. (2016); Pourré et al. (2022). In turn, the latter can introduce wavefront errors with root-mean-square (RMS) wavefront error (WFE) reaching hundreds of nanometers; this is manifested as combinations of the differential piston and tip-tilt errors that significantly deteriorate the image quality. It is important to note that this effect will be amplified on the upcoming extremely large telescopes, making the question of its measurement increasingly important. Currently, multiple methodologies have been proposed to address petaling (island) effects (Esposito et al. 2003; Shi et al. 2004). Among them, focal plane-based techniques can also be applied (see Agapito et al. 2022; Wilby et al. 2018).

To summarise, leveraging phase retrieval in the focal plane proves to be helpful for measuring LWE, co-phasing segmented telescopes (Lamb et al. 2021; Acton et al. 2012), characterising NCPAs in AO systems (Lamb et al. 2018; Robert et al. 2008), and image post-processing (Wilding et al. 2017; Wu et al. 2018).

Open Access article, published by EDP Sciences, under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
This article is published in open access under the Subscribe to Open model. Subscribe to A&A to support open access publication.

Focal-plane WFSing can successfully complement other WFSing approaches and thus unlock more nuanced AO control strategies, while also enabling more detailed measurements of optical wavefronts. Moreover, depending on the specific AO system, PSFs can be abundant, appearing not only in wavefront sensor images but also in the science images themselves. We highlight a few of the existing focal-plane phase retrieval techniques in the section below.

2. A concise overview of focal-plane WFSing techniques

2.1. Phase diversity (PD)

Phase diversity is a versatile family of focal-plane techniques used for WFSing. It involves capturing images at single or multiple focal planes, each corresponding to a known phase diversity introduced into the optical path. This diversity helps to resolve sign ambiguity and optimises the distribution of information about PSF and object morphologies across the detector pixels. While the classical implementation from Gonsalves (1982) employs defocus diversity, other aberrations and types of diversities can also be used (Campbell et al. 2004; Brady et al. 2009; Almoro & Hanson 2008).

Among PD techniques, the estimation of the phase in the pupil plane usually turns into a problem of finding the optimal set of modal coefficients that minimises the error between the model and data. In this approach, expressing phase as a superposition of pupil-space modes entails implicit regularisation, increasing robustness to noise. However, a solution for the phase in the pixel space is also possible, as outlined in Robert et al. (2008). The solution can be found by iteratively fitting the PSF model to the data. It usually involves first or second-order (Smith et al. 2013) approximations of the PSF model in the vicinity of phase estimates. Meanwhile, having multiple diverse images also allows for the decoupling of the unknown object morphology from wavefront aberrations (Mocœur et al. 2009; Blanc et al. 2003). Overall, PD is a valuable tool for measuring wavefront aberrations and improving image quality. In AO, PD methods are commonly employed due to their minimally invasive implementation and strong performance in the presence of noise. It can effectively handle a wide range of aberrations, making it a resilient and versatile technique for focal-plane WFSing.

2.2. Gerchberg-Saxton-based (GS) algorithm

The GS algorithm described in Gerchberg (1972), (see also Fienup 1982; Huang et al. 2021) is a classical forward-reverse phase retrieval algorithm that has given rise to such techniques as PD. It alternates between the spatial and Fourier domains, enforcing the amplitude with known measurements while updating the phase. The process iterates until the phase information converges, providing a phase estimate consistent with amplitude constraints applied in both domains. It usually requires many steps to converge and may easily become trapped in local minima due to the vast space of parameters to optimise. Therefore, it does not guarantee global convergence to the true phase, making this method unsuitable for real-time applications.

2.3. Linearized focal-plane technique (LIFT)

LIFT is a PD-based technique that employs a known amount of astigmatism instead of defocus to create phase diversity (see Meimon et al. 2010; Plantet et al. 2014, 2017). Consequently, there is no need to employ additional extra-focal intensity planes to resolve the sign ambiguity, meaning that only one image can be used for phase retrieval. This greatly simplifies the hardware implementation of LIFT and prevents flux splitting between images, leading to even greater sensitivity. However, it also makes it difficult to decouple the object morphology from wavefront aberrations. To summarise, LIFT is an ideal choice for low-order mode estimation using a point source in low-flux conditions.

2.4. Fast & furious (F&F)

Fast & Furious is a very performant PD-based method designed to run in closed-loop conditions (Korkiakoski et al. 2014; Bos et al. 2020). It is essentially derived from the small-phase approximation outlined in Gonsalves (2001) and it employs the difference between the DM shapes on current and preceding loop iterations as phase diversity. It also uses focal-plane PSFs acquired on the current and previous iterations to reconstruct the wavefront. Unlike LIFT, F&F does not disturb the morphology of the in-focus PSF. However, it also requires persistent access to the DM commands to extract the diversity phase.

2.5. Machine learning (ML) methods

Recently, ML-based methodologies have gained popularity due to their flexibility and versatility. In contrast to conventional methods, ML approaches can effectively integrate complex nonlinear relationships without relying on explicit analytical models (see Lloyd-Hart et al. 1992; Wang et al. 2021; Terreri et al. 2022). Therefore, they can receive inputs from various sources and be applied to virtually any kind of AO system. Such models can be trained both offline and online (in-loop) on actual data, enabling them to adjust to evolving conditions or changing system configurations. Nonetheless, in data-driven approaches, estimation accuracy directly depends on the quality and size of the training dataset. Despite the abundance of PSFs in AO systems, collecting a large and diverse, real or artificial dataset can prove challenging in practice. Moreover, running ML solutions in-loop necessitates significant computational overheads. In addition, the opaque nature of ML-based approaches may be viewed as a disadvantage in certain applications. Nevertheless, ML presents promising prospects in many applications, including focal-plane WFSing.

Selecting the appropriate method for focal-plane WFSing depends on the specific system or application in question. In the context of this work, our primary focus is the InfraRed Low-Order wavefront Sensor (IRLOS). It is an integral component of the Adaptive Optics Facility (AOF) (see Oberti et al. 2018), which enables the narrow field mode (NFM) of Multi Unit Spectroscopic Explorer (MUSE) instrument. The primary goal of IRLOS is to accurately sense low-order (LO) modes, specifically tip and tilt (TT) and defocus. This capability is crucial as TT modes are beyond the sensing capability of laser guide stars (LGSs), whereas defocus retrieval is compromised by the altitudinal drift of the sodium layer. To enable tip-tilt and focus (TTF) measurements, IRLOS employs a single natural guide star (NGS). The baseline configuration of IRLOS is a 2×2 Shack-Hartmann (SH) WFS, operating in J and H optical bands, with a current limiting NGS magnitude of $m_J < 19$.

A proposal has been made to enhance the sensitivity of IRLOS by substituting the current SH WFS configuration with a

full-pupil mode. The full-pupil mode entails the removal of the 2×2 lenslet array from the pupil plane of IRLOS and replacing it with a single lens, thereby eliminating the per-aperture flux splitting and enabling the collection of photons from the entire pupil and focusing them into a single PSF. In this regime, IRLOS would operate as a fast TT sensor at the frequency of 200 Hz or 500 Hz (depending on the magnitude of a target) and as a slow truth sensor for defocus, accumulating a PSF over 1–2 s of exposure. Due to the broadband sensitivity of IRLOS and the necessity to accumulate as much flux as possible under low-flux conditions, PSFs recorded in the full-pupil mode are polychromatic.

In this context, LIFT was chosen as the optimal slow defocus sensing method due to its inherent noise robustness, ease of hardware implementation, and the ability to accurately estimate defocus using a single image in a single focal plane. Furthermore, a slow defocus loop operates at a frequency of 0.5–1 Hz, providing LIFT with sufficient time to perform computations. The question of sensitivity gains when using LIFT compared to the 2×2 SH WFS is covered in detail in Plantet et al. (2014) and Meimon et al. (2010). However, the main objective of this paper is to discuss the challenges faced while applying LIFT on-sky and suggest methods to enhance its robustness. However, it is useful to briefly investigate the theoretical foundations of this method first to provide a background for later discussions.

3. Theory

The focal-plane PSF with introduced phase diversity can be approximated using the following expression:

$$\mathbf{I}(\mathbf{r}, \Phi(\mathbf{u})) = \mathbf{O}(\mathbf{r}) \circledast \sum_{\lambda} F(\lambda) \left| \mathcal{F}\{\mathbf{P}(\mathbf{u}) e^{i\frac{2\pi}{\lambda}\mathbf{\Phi}(\mathbf{u})}\} \right|^2 \Delta\lambda \; + \; \mathbf{n}(\mathbf{r}), \quad (1)$$

where $\mathbf{P}(\mathbf{u})$ is the aperture mask, $\mathbf{\Phi}(\mathbf{u})$ is the optical path difference (OPD) in the pupil and $\mathbf{r}$, $\mathbf{u}$ are the 2D pixel coordinates in the focal and pupil planes, respectively. In subsequent expressions, we omitted the pixel coordinates. Also, $\mathbf{n}(\mathbf{r})$ is the per-pixel noise term in image space, $\mathbf{O}(\mathbf{r})$ is the convolution kernel that represents the shape of an NGS in case it is an extended source, $\circledast$ is the 2D convolution operation, and $\mathcal{F}\{\cdot\}$ is the Fourier transform, while λ is the wavelength. Quantities in bold are the vectors and matrices. Then, $\mathbf{\Phi}(\mathbf{u})$ can be described as the superposition of modes in the pupil space:

$$\mathbf{\Phi} = \sum_i a_i \mathbf{Z}_i + \phi_\mathrm{d} = \mathbf{Z}\mathbf{A} + \phi_\mathrm{d}, \quad (2)$$

where $\mathbf{Z}_i(\mathbf{u})$ is the i-th pupil mode of modal basis $\mathbf{Z}$. In this work, we adopt the Zernike modal basis. However, the choice of the basis can be arbitrary under the condition of orthogonality in the pupil plane. The vector $\mathbf{A} = [a_0, a_1, \ldots, a_M]$ comprises the modal coefficients a_i, $\phi_\mathrm{d}(\mathbf{u})$ denotes the introduced diversity phase. We specifically employ astigmatic phase diversity using the mode Z_5[1]. This choice is connected to the hardware implementation of LIFT, details of which will be expanded in the next section. However, any predetermined wavefront aberration that avoids sign ambiguity and significantly impacts the PSF morphology can function as diversity. The question of choosing the optimal amplitude of phase diversity is covered in Plantet et al. (2013), Dean & Bowers (2003), and Polo et al. (2013).

If $P(\mathbf{u})$, $Z_i(\mathbf{u})$, and $\mathbf{O}(\mathbf{r})$ are pre-defined, the PSF morphology described using Eq. (1) is a function of modal coefficients $\mathbf{A}$; thus, we can say that $\mathbf{I}(\mathbf{r}, \Phi(\mathbf{u})) = \mathbf{I}(\mathbf{A})$. Therefore, the goal of the focal-plane wavefront estimator is to find the optimal vector of modal coefficients $\hat{\mathbf{A}}$ that minimises the difference $\Delta\mathbf{I}(\mathbf{A})$ between the observed PSF $\mathbf{I}_\mathrm{data}(\mathbf{r})$ and the simulated PSF $\mathbf{I}(\mathbf{A})$, as:

$$\Delta\mathbf{I}(\mathbf{A}) = \Delta\mathbf{I} = \mathbf{I}_\mathrm{data} - \mathbf{I}(\mathbf{A}). \quad (3)$$

We can simplify further derivations in this section by omitting the convolution kernel $\mathbf{O}(\mathbf{r})$ and assuming a PSF at a single wavelength.

The derivation of the LIFT estimator is covered in detail in Meimon et al. (2010) and Plantet (2014, chap. A.2). Therefore, we do not delve into the derivations here. The best estimate $\hat{\mathbf{A}}$ for the modal coefficients can be derived within the maximum likelihood (ML) or maximum a posteriori probability (MAP) frameworks (Eqs. (4) and (5), respectively) as:

$$\hat{\mathbf{A}}_\mathrm{ML} = \left(\mathbf{H}^\top \mathbf{R}_\mathrm{n}^{-1} \mathbf{H}\right)^{-1} \; \mathbf{H}^\top \mathbf{R}_\mathrm{n}^{-1} \Delta\mathbf{I}, \quad (4)$$

$$\hat{\mathbf{A}}_\mathrm{MAP} = \left(\mathbf{H}^\top \mathbf{R}_\mathrm{n}^{-1} \mathbf{H} + \mathbf{C}_\varphi^{-1}\right)^{-1} \left(\mathbf{H}^\top \mathbf{R}_\mathrm{n}^{-1} \Delta\mathbf{I} + \mathbf{C}_\varphi^{-1} \bar{\mathbf{A}}\right). \quad (5)$$

Here, $\mathbf{H} = \mathbf{H}(\mathbf{A}) = \left[\frac{\partial \mathbf{I}(\mathbf{A})}{\partial a_0} \cdots \frac{\partial \mathbf{I}(\mathbf{A})}{\partial a_M}\right]$ is the interaction matrix (Jacobian), $\mathbf{R}_\mathrm{n} = \langle \mathbf{n}\mathbf{n}^\top \rangle$ is the covariance matrix of detector pixels, assuming they are uncorrelated with each other and that the distribution of photo-electrons per pixel is Gaussian.

Equations (4) and (5) provide the one-shot estimations derived by linearising the $\Delta\mathbf{I}(\mathbf{A})$ in the vicinity of $\mathbf{A} = [0, 0, \ldots, 0]$, resulting in a very limited linearity range. To overcome this limitation, the ML and MAP estimators can be approached iteratively when every subsequent estimate $\hat{\mathbf{A}}_{i+1}$ is computed in the vicinity of the previous estimate $\hat{\mathbf{A}}_i$ as follows:

$$\hat{\mathbf{A}}_{i+1}^\mathrm{ML} = \left(\mathbf{H}(\hat{\mathbf{A}}_i)^\top \mathbf{R}_\mathrm{n}^{-1} \mathbf{H}(\hat{\mathbf{A}}_i)\right)^{-1} \mathbf{H}(\hat{\mathbf{A}}_i)^\top \mathbf{R}_\mathrm{n}^{-1} \Delta\mathbf{I}(\hat{\mathbf{A}}_i) + \hat{\mathbf{A}}_i, \quad (6)$$

$$\hat{\mathbf{A}}_{i+1}^\mathrm{MAP} = \left(\mathbf{H}(\hat{\mathbf{A}}_i)^\top \mathbf{R}_\mathrm{n}^{-1} \mathbf{H}(\hat{\mathbf{A}}_i) + \mathbf{C}_\varphi^{-1}\right)^{-1} \left(\mathbf{H}(\hat{\mathbf{A}}_i)^\top \mathbf{R}_\mathrm{n}^{-1} \Delta\mathbf{I}(\hat{\mathbf{A}}_i) + \mathbf{C}_\varphi^{-1} \bar{\mathbf{A}}\right) + \hat{\mathbf{A}}_i. \quad (7)$$

Here, $\mathbf{C}_\varphi^{-1}$ is the covariance matrix of the modal coefficients, and $\bar{\mathbf{A}}$ is their mean. However, determining the exact statistics of the modal coefficients is challenging since LIFT operates in closed-loop conditions, and very few on-sky measurements have been performed so far. At the same time, the imprecise knowledge of coefficient statistics can lead to biased solutions. Therefore, within the scope of this work, we computed $\hat{\mathbf{A}}$ using an unbiased iterative ML estimator. However, in practice, it is possible to assume that $\mathbf{C}_\varphi^{-1}$ follows some power law while $\bar{\mathbf{A}} = [0, 0, \ldots, 0]$ when utilising MAP estimation framework.

In essence, the iterative estimators shown in Eqs. (6) and (7) are similar to the Gauss-Newton optimisation method (see Björck 1996, chap. 9), while the minimisation criterion is equivalent to the mean squared error (MSE) weighted by $\mathbf{R}_\mathrm{n}^{-1}$. From this point of view, LIFT is a least-squares minimisation problem of fitting the PSF model to the data that can be decomposed into three main components: PSF model, non-linear solver (optimisation method), and optimisation criterion (loss function). Depending on the required application, each component can be replaced or adjusted in this context. This provides an essential framework for the further enhancement of LIFT. For example, the PSF model can be modified. Furthermore, the Gauss-Newton optimiser can be substituted with other approaches, such as the

[1] Following the indexing from Noll (1976).

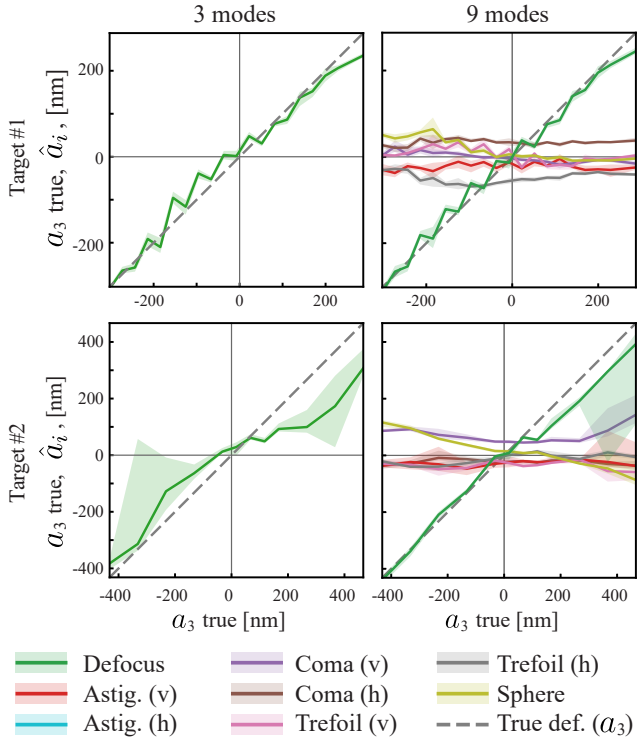


Fig. 1. LIFT estimations for the dataset of PSFs with ramping defocus. Results were obtained for two on-sky targets (top and bottom rows). Estimations were done for three modes (a_1, a_2, a_3) (left column) and nine modes $(a_1, \ldots a_{10}) \setminus \{a_5\}$ (right column). For target #2, including modes beyond TTF allowed us to retrieve the defocus in an accurate way. Here, TT modes are estimated but not displayed. Lines represent median values, while error bars show 1σ confidence interval.

L-BFGS method described Liu & Nocedal (1989). Such modifications also suggest a notable deviation from the expressions in Eqs. (6) and (7). The issue of modifying LIFT is further examined in Sect. 5.3. The upcoming section will discuss the practical application of the unaltered LIFT technique on real-life sky data.

4. First on-sky tests

Following the mid-2021 upgrade of IRLOS, the feasibility of LIFT was experimentally tested in on-sky conditions on Unit Telescope 4 (UT4) of ESO's Very Large Telescope (VLT) facility. In the hardware implementation of LIFT within IRLOS, a 2 × 2 lenslet array was replaced with a single lens and a cylindrical lens was added to induce astigmatism in the optical path. This lens determines the amplitude of introduced astigmatic diversity, resulting in roughly -170 nm RMS. However, since the astigmatic lens occupies a slot in the filter wheel, it cannot be used simultaneously with a narrow-band filter. As a result, LIFT needs to operate with strongly polychromatic PSFs that cover both *J* and *H* bands. Additionally, the SAPHIRA detector (see Finger et al. 2014) was introduced with the IRLOS upgrade. It yielded an angular resolution of ∼9.7 mas per pixel in full-pupil mode[2], resulting in a sampling of ∼3.2 and ∼4.4 pixels per $\frac{\lambda}{D}$ at the J and H bands, respectively. Consequently, LIFT has access to PSFs that are above the Nyquist sampling and have well-resolved morphological features.

The first experimental validation of LIFT was conducted on two bright natural targets (see Fig. 1). These tests aimed to introduce the range of known defocus values and retrieve them utilising LIFT. Defocus was introduced by shifting the focusing stage of MUSE. Therefore, the introduced defocus offsets are the only accessible ground truth values in this experiment. The selected range of defocus values was based on simulations which showed that the linearity of LIFT in the H-band is confined to about ±410 nm RMS[3]. Since defocus was the sole variable changed during the experiment, other coefficients are expected to exhibit relative stability. Table 1 lists the exact data acquisition parameters.

Table 1. Acquisition parameters for the first on-sky test.

Parameter	Values
m_J of NGS targets	9, 12.5
Exposure time per PSF (s)	1
Seeing (")	0.5, 1.5
PSFs recorded per defocus offset	10, 20
Introduced defocus (nm RMS)	–432 . . . 467, –302 . . . 288
a_5 diversity (lens) (nm RMS)	–170
Target wavelength λ (nm)	1215 . . . 1625
σ_{obj} (mas)	12, 14.5

The results presented in Fig. 1 demonstrate the successful retrieval of defocus with LIFT alongside the values of additional coefficients. The saw-tooth behaviour seen in coefficient estimates for target #1 is caused by the fact that every consequent PSF in the scan was recorded with alternating defocus sign, which required moving the focusing stage of MUSE to opposite directions in ping-pong fashion after every recorded sample, causing the introduction of the unmeasured bias into the reference values due to the hysteresis of the focal stage. The constant bias in defocus estimates for the second target is caused by the fact that two parts of the same scan were recorded in two separate runs, potentially biasing the reference.

4.1. Lessons learned

Although initial on-sky experiments clearly demonstrated the capability of LIFT to operate in realistic conditions, achieving these results practically demanded multiple manual modifications to the estimation process. Therefore, discussing these adjustments is essential for the interpretation of the results.

First, LIFT required an initial approximation of the defocus. Providing LIFT with a defocus prior yielded faster and more accurate convergence, especially when dealing with strongly defocused PSFs. The initial guess for a_3 was extracted from the diversity-induced PSF elongation by fitting a 2D Gaussian to each PSF and computing the $A(\sigma_x/\sigma_y - 1) + B$ ratio, where σ_x, σ_y are the fitted parameters of a 2D Gaussian, while A, B are the normalising constants for calibration.

Second, adding a convolution kernel σ_{obj} to the PSF model was required. In the PSF model defined in Eq. (1), the convolution kernel $\mathbf{O}(\mathbf{r})$ represents an extended object. However, for this experiment, we selected only point-source targets, meaning that this kernel absorbed all the effects that are not accounted for by the PSF model used in LIFT. The primary contributors may include the residual TT jitter, the polychromatic smearing of PSFs, and the potential cross-talk between the detector's pixels (although the latter effect is negligible in the SAPHIRA detector). The atmospheric dispersion can also have an impact on

[2] With the LIFT lens introduced.

[3] Assuming that the initial guess for defocus is zero.

the IRLOS PSFs because the dispersion compensator situated upstream of IRLOS is designed to optimise wavelengths in the visible spectrum for MUSE and may produce residual elongation of the spot due to chromatism at J and H bands. The initial shape of the 'object' can be derived from the expected TT jitter under specific seeing conditions and then adjusted via fitting. Then, $\mathbf{O}(\mathbf{r})$ is modelled as a 2D Gaussian. However, adding σ_{obj} as an optimisable parameter is undesirable due to the single-image nature of LIFT, as it can cause coupling between the kernel parameters and modal coefficients. Therefore, we had to fine-tune the convolution kernel parameters manually.

Third, modes beyond focus had to be estimated. While target #1 indicated that retrieving defocus by estimating only TTF is feasible, target #2 showed that estimating additional modes is sometimes essential for accurate defocus retrieval. It was determined through experimentation that the most precise results were obtained by estimating the first ten Zernike modes (excluding a_5, the diversity astigmatism). Similarly, although LIFT was proposed for slow defocus sensing, incorporating TT modes into the estimation enhanced the accuracy of estimation by compensating for sub-pixel shifts of the PSF relative to the centre of the region of interest, even though TT estimates from LIFT are not used for control. Meanwhile, the necessity of estimating orders higher than TTF can be understood when examining the recorded on-sky PSFs (refer to Fig. 2), which reveal highly complex morphological features. We attribute this to the strong presence of the uncompensated NCPAs. We further suggest that these NCPAs can be modelled as phase aberrations in the pupil plane. Additionally, their quasi-static behaviour is indicated by the temporal consistency of speckle patterns observed on PSF images and the relative stability of estimated coefficients, as demonstrated in Fig. 1. Thus, including only TTF modes or even only focus in the estimation is generally insufficient for capturing this complexity. Adding more modes improves the descriptiveness of the model. However, it also introduces more variables, weakening the inherent regularisation induced by the modal decomposition of the pupil phase. This makes the model more susceptible to noise and overfitting, leading to cross-talk between modal coefficients. For example, the astigmatic diversity mode was excluded from the estimation due to its coupling with other modal coefficients. This scenario underscores a trade-off between complexity and precision. Reducing the number of estimated modes regularises the solution, yet a model that is too simplistic may result in unrepresented higher-order modes aliasing into the TTF. We offer additional information on cross-talk in Appendix B.

Out of all the points mentioned, the last one needs the most attention. Estimating excessive modes restricts the robustness of the method, particularly in low-flux scenarios, as later demonstrated in Fig. 9. This contradicts the initial promise of using LIFT as a highly sensitive TTF sensor. Addressing this issue is therefore critical.

4.2. Need for LIFT calibration

We propose addressing this problem by separating TTF retrieval from NCPA estimation. This can be achieved in a configuration divided into two consecutive steps. The first step is the model calibration. It "absorbs the model complexity" by retrieving NCPAs, assuming their quasi-static behaviour. This step is done offline before closing the slow TTF loop with LIFT. The second step is the focus estimation (online), which estimates only TTF, leveraging the preceding NCPA calibrations as a prior.

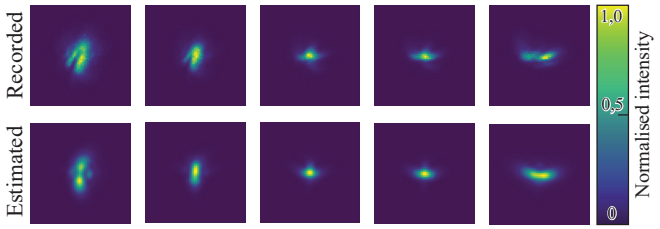


Fig. 2. Examples of recorded PSFs for target #1 (top row) and their estimated counterparts (bottom row), captured with varying defocus levels and inserted diversity. The prominent effect of NCPAs on PSF morphology is evident. The data was acquired in the focal plane of IRLOS after closing the HO and fast TT control loops, with a seeing of ≈0.5″.

Calibration for NCPA can be conducted using either an artificial source or directly an on-sky target. However, using the latter can yield more realistic results due to the ability to capture all contributors missed by the optical path of the artificial calibration source[4]. Our study thus focuses on developing a methodology that leverages on-sky observations for retrieving quasi-static NCPAs, utilising the same information channel as LIFT for minimal invasiveness and ease of execution. The next section will discuss further details on implementing this calibration approach.

It should be noted that within the scope of this paper, the term 'quasi-static' refers to the assumption that the measured NCPAs remain constant during the calibration process and continue to be applicable when the loop is closed on defocus. However, the detailed investigation of the NCPAs' exact temporal dynamics to determine how long the calibration estimates remain valid is not addressed in this paper.

5. Two-stage approach

The following section focuses on implementing and analysing the two-stage focal-plane WFSing technique. All the results presented in this section are derived from realistic and representative IRLOS simulations.

5.1. The problem of NCPA calibration

We start the discussion with a more detailed description of the calibration stage. The primary objective and essential criteria are as follows. It must accurately estimate the quasi-static component of NCPAs and reconstruct as many modes as possible to provide enough degrees of freedom to describe the complex PSF morphology accurately. The cross-coupling between estimated modes should be minimised. Similar to LIFT, calibration should utilise PSF images acquired in the focal plane of IRLOS. Calibration must be executed after closing the high order (HO) loop and before performing in-loop focus measurements. Calibration is conducted offline and thus is not subject to strict hardware and time constraints encountered during in-loop operations.

To accurately estimate wavefront aberrations and access higher-order modes, it is essential to retrieve high spatial frequencies in the pupil plane and thus preserve the structure of speckles. This requires recording monochromatic PSFs to eliminate the effects associated with polychromaticity. Additionally, using short exposure times can reduce the blurring associated with the uncompensated HO and TT residuals. Therefore, if a point source is used, $\mathbf{O}(\mathbf{r})$ can be excluded from the PSF

[4] At least, in the case of AOF.

Table 2. Notations used in this and subsequent sections.

Notation	Meaning
N	Number of PSFs in the cube,
M	Number of simulated or estimated modes,
W, H	Width and height of the PSF image,
$\mathcal{I}_{\text{data}}$	Cube of recorded PSFs,
$\mathbf{I}_{\text{data},i}$	ith PSF in the cube (sample), $i = 1, \dots N$,
$\mathbf{A}_i$	ith vector of modal coefficients
$\mathcal{A}$	Matrix of stacked coefficient vectors
$\mathbf{I}(\mathbf{A})$	PSF simulated using modal coefficients vector $\mathbf{A}$
$\mathcal{I}(\mathcal{A})$	Cube of simulated PSFs
$\mathbf{1}_N$	Vector of ones of length N
$\otimes$	Kronecker product
$\mathcal{L}(\cdot)$	Optimisation criterion/loss function
$\|\cdot\|_1$	L_1-norm.

model because the main contributors, such as TT jitter and polychromatic blurring, are mitigated in this case (see Sect. 4.1). Therefore, the PSF can now be modelled using solely wavefront aberrations. In this case, a bright target must be used, which is necessitated by the combination of short exposures combined with the narrow spectral bandwidth. These factors mean that the calibration and defocus-retrieval stages may utilise different on-sky targets. This also implies that the data acquisition parameters for the calibration stage must differ from those used for LIFT.

The IRLOS WFS utilises SPARTA (see Valles et al. 2012) as a real-time computer (RTC). In the current configuration, SPARTA allows for the recording and storing of the detector pixel readings from the SAPHIRA detector at a maximum frequency of 10 Hz. Consequently, it enables access to frames with 2–5 ms exposure time at 10 Hz or to averaged images with a 0.1 second exposure at 10 Hz. In terms of exposure duration, theoretically, an infinitely short exposure would yield the most accurate, unaveraged snapshot of speckle patterns, thereby giving access to the highest number of spatial frequencies. Conversely, our simulations have demonstrated that extending the average exposure beyond 0.15–0.2 s leads to a predominance of HO residuals. Therefore, within the framework of this study, we have selected this upper limit as a compromise to optimise the trade-off between minimising exposure time and maximising the signal-to-noise ratio (S/N).

5.2. Different approaches to modal coefficients optimisation

Considering the requirements mentioned earlier in this paper and leveraging the hardware capabilities, it becomes possible to capture PSF cubes. For example, with a sampling rate of 10 Hz, capturing an entire PSF dataset within 1 min of exposure time is feasible. This effectively increases the quantity of available data samples, better defining the problem of searching for optimal modal coefficients. Therefore, we propose leveraging the PSF cubes instead of averaged long-exposure PSFs for phase retrieval. In this scope, several non-linear optimisation approaches can be considered. Before elaborating on this, a few notations must be introduced in the table below.

The entities introduced in Table 2 have the following dimensions:

$$
\begin{aligned}
\mathbf{1}_N &= [1, 1, \dots 1], & \mathbf{1}_N &\in \mathbb{R}^N \\
\mathbf{A}_i &= [a_{i,0}, a_{i,1}, \dots a_{i,M}], & \mathbf{A}_i &\in \mathbb{R}^M, \\
\mathcal{A} &= [\mathbf{A}_0; \mathbf{A}_1; \dots \mathbf{A}_N], & \mathcal{A} &\in \mathbb{R}^{N\times M}, \\
\mathcal{I}_{\text{data}} &= [\mathbf{I}_{\text{data},0}; \mathbf{I}_{\text{data},1}; \dots \mathbf{I}_{\text{data},N}], & \mathcal{I}_{\text{data}} &\in \mathbb{R}^{N\times W\times H}, \\
\mathcal{I}(\mathcal{A}) &= [\mathbf{I}(\mathbf{A}_0); \mathbf{I}(\mathbf{A}_1); \dots \mathbf{I}(\mathbf{A}_N)], & \mathcal{I}(\mathcal{A}) &\in \mathbb{R}^{N\times W\times H}, \\
\mathcal{L}(x, \hat{x}) &= \frac{1}{N} \|x - \hat{x}\|_1 .
\end{aligned}
$$

Here, $\mathcal{L}(x, \hat{x})$ is the loss function or optimisation criterion, $x, \hat{x}$ are some variables, $\mathcal{I}_{\text{data}}$ represents a cube composed of PSFs captured sequentially at a sampling frequency of 10 Hz. With all the required terms defined, we are ready to introduce a few approaches to coefficient optimisation.

Method 1 involves averaging the recorded PSF cube $\mathcal{I}_{\text{data}}$ along the temporal dimension and fitting a single modal coefficients vector to the averaged image, which is akin to the standard implementation of LIFT discussed earlier. This provides minimal computational overhead and rapid convergence due to the relatively small scale of the problem. The optimisation can then be formulated as:

$$
\hat{\mathbf{A}} = \text{argmin}_{\mathbf{A}} \mathcal{L}\left(\frac{1}{N}\sum_i^N \mathbf{I}_{\text{data},i},\ \mathbf{I}(\mathbf{A})\right).
$$

However, averaging the data cube can lead to the blurring of the fine morphological features of the PSFs and generating a residuals-driven "halo". It is possible to incorporate the residual-related component into the PSF model to address this issue, as demonstrated in Mugnier et al. (2008). However, implementing this technique is beyond the scope of the current work. Therefore, we limit our discussion to PSF models driven solely by modal coefficients.

Method 2. In contrast to the previous method, this optimisation approach implies fitting a single coefficient vector to the PSF cube $\mathcal{I}_{\text{data}}$. This forces the optimiser to find a coefficient vector $\hat{\mathbf{A}}$ that simultaneously satisfies all PSFs in the cube, thus preventing overfitting to a particular PSF sample. Given the utilisation of a single coefficient vector, the gradient updates for each coefficient in the vector are collected from all samples and then averaged. Therefore, the part of the gradients common to all samples will dominate, while the noisy component will be averaged out. This estimation method can be expressed as follows, noting that the Kroeneker product $\otimes$ clones the same coefficient vector for each data sample:

$$
\hat{\mathbf{A}} = \text{argmin}_{\mathbf{A}} \mathcal{L}\left(\mathbf{1}_N \otimes \mathbf{I}(\mathbf{A}),\ \mathcal{I}_{\text{data}}\right).
$$

Method 3. Another possible approach is to independently fit a unique coefficient vector $\mathbf{A}_i$ to every individual sample $\mathbf{I}_{\text{data},i}$. No gradients between coefficients are shared in this case. While this approach allows for a more precise fit for each individual sample by increasing the parameter space, it also increases the likelihood of overfitting. This approach be expressed as follows:

$$
\hat{\mathcal{A}} = \begin{bmatrix} \text{argmin}_{\mathbf{A}_0} \mathcal{L}\left(\mathbf{I}_{\text{data},0},\ \mathbf{I}(\mathbf{A}_0)\right) \\ \text{argmin}_{\mathbf{A}_1} \mathcal{L}\left(\mathbf{I}_{\text{data},1},\ \mathbf{I}(\mathbf{A}_1)\right) \\ \vdots \\ \text{argmin}_{\mathbf{A}_N} \mathcal{L}\left(\mathbf{I}_{\text{data},N},\ \mathbf{I}(\mathbf{A}_N)\right) \end{bmatrix}.
$$

The same can be written down in a slightly shorter notation as fitting a matrix of coefficients, $\hat{\mathcal{A}}$, to the PSF cube, $\mathcal{I}_{\text{data}}$.

$$
\hat{\mathcal{A}} = \text{argmin}_{\mathcal{A}} \mathcal{L}\left(\mathcal{I}_{\text{data}},\ \mathcal{I}(\mathcal{A})\right), \quad \hat{\mathbf{A}} = \frac{1}{N}\sum_i^N \hat{\mathbf{A}}_i. \tag{8}
$$

This notation is equivalent to the simultaneous estimation of the ensemble of coefficient vectors, $\mathbf{A}_i$. All coefficient vectors inside the $\mathcal{A}$ can then be averaged to get the final estimate for the quasi-static component.

In summary, the optimisation methods discussed above require substantial computational resources due to the increased scale of the optimisation problems. The iterative optimisation process involves repeatedly simulating hundreds of PSFs and computing Jacobians, which becomes particularly computationally demanding when using methods 2 and 3. Therefore, it is crucial to implement an efficient and high-performing code to address this issue before performing the analysis. This problem will be approached in the subsequent section.

5.3. Introducing DIP

Here, we present the differentiable PSF model (DIP), the code[5] designed to accelerate and facilitate extensive PSF simulations and large-scale PSF fitting tasks based on the PyTorch framework. Given it is designed for machine learning, PyTorch is well-optimised for handling massive data batches both on CPU and GPU backends. Furthermore, PyTorch offers a diverse range of optimisation algorithms that can be seamlessly integrated with DIP. Finally, DIP takes advantage of the automatic differentiation functionality of PyTorch, which allows for gradients to be computed with almost analytical precision for the PSF models of arbitrary complexity. In turn, this eliminates the necessity to derive the analytical expressions for Jacobian and Hessian matrices or to approximate them numerically with finite differences. Harnessing these advantages allows DIP to go beyond the initial mathematical formalism of LIFT, resulting in a versatile framework where the PSF model, optimiser, and loss function are distinct interchangeable components, implementing a concept elaborated upon earlier in Sect. 3.

5.4. Selecting the estimation approach

By leveraging DIP, it becomes feasible to implement the optimisation approaches 1–3 proposed in the previous subsection. To quantitatively compare them, we conducted a test shown in Fig. 3. The test assesses the ability of estimation methods to retrieve the quasi-static phase component associated with NCPAs accurately. The figure depicts the distribution of RMS WFE, which is computed for the difference between the temporally averaged simulated wavefront cube and the estimated static LO wavefront (see Eq. (9) for more details). To do this test, a dataset of simulated PSF cubes was generated and subsequently analysed using the techniques derived from estimation methods 1–3 covered earlier in the previous section.

In Fig. 3, "LIFT avg." represents the standard implementation of LIFT detailed earlier in Sect. 3. "DIP avg." operates similarly to classical LIFT and implements the estimation approach 1. However, it is powered by the DIP framework, meaning it takes the same average PSF as input, while the optimisation algorithm and loss function differ. For DIP, we adopted the L-BFGS optimisation algorithm described in Liu & Nocedal (1989). As an optimisation criterion, DIP uses the Maximum Absolute Error (MAE or L_1-loss). "DIP joint" implements the estimation approach 2, where a single coefficient vector is shared across all PSF samples within the cube. "LIFT slices" refers to applying LIFT in the fashion outlined in method 3, where the fitting process is conducted on a per-sample basis. Finally, "DIP global" is also based on method 3. However, it simultaneously fits all coefficient vectors, essentially implementing Eq. (8).

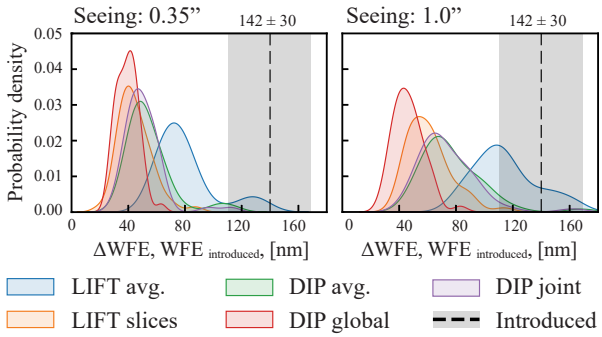


Fig. 3. Comparison of different methods to retrieve the quasi-static LO modes associated with NCPAs. The "DIP global" method shows the highest accuracy, especially in the case of larger seeing (right picture). The curves represent the distribution of RMS of the averaged delta wavefront error (ΔWFE). The black dashed line and grey fill show the mean RMS of introduced NCPA realisations ($\mathrm{WFE}_{\mathrm{introduced}}$) with 1σ confidence interval.

The analysis indicates that the "LIFT slices" and "DIP global" methods exhibit superior precision in retrieving quasi-static LO modes compared to other listed methods. This can be explained by the fact that when the PSF model is fitted to a set of short-exposure PSFs, each PSF sample preserves the information about high spatial frequencies in its speckles, consequently facilitating the retrieval of higher-order modes. In this context, "higher-order modes" refer to modes higher than TTF that are included in the estimation. In this test, 29 first Zernike modes were estimated. The question of choosing the number of estimated modes will be covered in the next section. Meanwhile, accurately retrieving higher orders also facilitates the accurate estimation of LO modes, as previously mentioned in Sect. 4.1. In contrast, this information is lost in averaged long-exposure PSFs. In the same fashion, sharing one coefficient vector over all PSF samples in "DIP joint" effectively acts as averaging the PSF. It is important to note that there is a risk of overfitting when using "LIFT slices" and "DIP global". However, this is counterbalanced by the final averaging across the ensemble of modal vector estimates. This is connected to the fact that PSF samples are captured sequentially in the presence of HO residuals. Therefore, all PSF samples slightly differ from each other. Thus, only the static component of corresponding estimated modal vectors remains after the averaging.

It is important to note that all DIP-based approaches demonstrated overall higher estimation accuracy compared to LIFT-based approaches. For example, "DIP avg." shows superior accuracy compared to classical LIFT implementation, even though the same data inputs are used in these two cases. In turn, "DIP global", although it is essentially equivalent to "LIFT slices", shows better accuracy for the larger seeing values. This can be attributed to utilising the L-BFGS instead of the Gauss-Newton optimisation algorithm intrinsically used in LIFT. Based on these results, we adopt the "DIP global" as our baseline solution for NCPA calibration further in this paper.

To generate Fig. 3, realistic simulations were performed to recreate on-sky conditions. The dataset of 100 samples was generated, where each sample features a PSF cube with $N = 300$ sequential IRLOS-like short-exposure PSFs and a corresponding cube of true phase. We selected this number of samples because there was no noticeable improvement in estimation quality when using $N > 200$. However, to ensure robustness, we

[5] https://github.com/EjjeSynho/DIP

Table 3. Parameters for the calibration-stage simulations.

Parameter	Value
Target m_J	7
Sampling time (s)	0.1
PSFs in a cube	300
a_5 diversity (nm)	–170
Wavelength λ (nm)	1600
Seeing (")	0.35, 1.0
Introduced NCPAs (nm RMS)	142 ± 30

opted for a slightly higher number. Each phase screen in such a cube is the sum of tip-tilt ($\mathbf{\Phi}_{\mathrm{TT}}$) residuals, low-order NCPAs ($\mathbf{\Phi}_{\mathrm{LO}}$), HO residual phase ($\mathbf{\Phi}_{\mathrm{HO}}$), and phase diversity ($\phi_{\mathrm{d}}$); $\mathbf{\Phi}_{\mathrm{LO}}$ is based on the random set of the first 50 Zernike coefficients, which remain constant for each PSF within one given cube. However, every sample in the dataset features different realisations of LO coefficients. For further information on simulating the LO component, we refer to Appendix A. The residual phase screens are sampled from actual recorded AOF telemetry with the addition of the simulated atmospheric residue. The fast $\mathbf{\Phi}_{\mathrm{TT}}$ is randomly generated for each screen based on residual TT statistics observed on IRLOS. The simulations incorporate noise and model a point source with a magnitude of $m_J = 7$ for all samples in the dataset, matching the magnitude used for the on-sky samples discussed in this paper. However, the simulations indicate that, with an exposure time of 0.1 s per sample, calibration targets with $m_J < 15$ are viable. Further details on the simulation parameters can be found in Table 3.

Further details on the computation of the WFE presented in Fig. 3 are provided in Eq. (9), along with clarifications on previously introduced notations. Here, $\mathbf{\Phi}$ is the average of simulated wavefronts $\mathbf{\Phi}_i$; then $\mathit{\Phi}$ is the cube of simulated wavefronts; $\hat{\mathbf{\Phi}}$ is the estimated static NCPA wavefront; ΔWF is the difference between introduced and retrieved averaged wavefronts $\mathrm{WF}_{\mathrm{introduced}}$ and $\mathrm{WF}_{\mathrm{estimated}}$; WFE$(\cdot)$ is the RMS of a wavefront computed over the pupil pixels; N_{pupil} is the number of valid pixels in the pupil mask, $\mathbf{P}$; the term $\mathbf{P}(\mathbf{\Phi} - \hat{\mathbf{\Phi}})$ is summed over the pupil-plane pixels; $\odot$ is the Hadamard product; and $\mathbf{X}$ is a matrix variable. Thus, we have:

$$
\begin{aligned}
\mathbf{\Phi}_i &= \mathbf{\Phi}_{\mathrm{LO}} + \mathbf{\Phi}_{i,\mathrm{TT}} + \mathbf{\Phi}_{i,\mathrm{HO}} + \phi_{\mathrm{d}},\\
\mathit{\Phi} &= [\mathbf{\Phi}_0; \mathbf{\Phi}_1; \ldots \mathbf{\Phi}_N],\\
\mathbf{\Phi} &= \frac{1}{N}\sum_i^N \mathbf{\Phi}_i,\\
\hat{\mathbf{\Phi}} &= \mathbf{Z}\hat{\mathbf{A}} + \phi_{\mathrm{d}}\\
\mathrm{WFE}\,(\mathbf{X}) &= \sqrt{\frac{1}{N_{pupil}}\sum\sum(\mathbf{P}\odot\mathbf{X})^2},\\
\mathrm{WF}_{\mathrm{introduced}} &= \mathbf{\Phi} - \phi_{\mathrm{d}},\\
\mathrm{WF}_{\mathrm{estimated}} &= \hat{\mathbf{\Phi}} - \phi_{\mathrm{d}},\\
\Delta\mathrm{WF} &= \mathrm{WF}_{\mathrm{introduced}} - \mathrm{WF}_{\mathrm{estimated}},\\
\mathrm{WFE}_{\mathrm{introduced}} &= \mathrm{WFE}\,(\mathbf{\Phi} - \phi_{\mathrm{d}}),\\
\Delta\mathrm{WFE} &= \mathrm{WFE}\,(\Delta\mathrm{WF}).
\end{aligned}
\tag{9}
$$

5.5. Verifying calibration approach

After selecting the "DIP global" approach as the baseline method for NCPA calibration, it is now possible to illustrate its ability to correct for the quasi-static LO phase. Figure 4 highlights the performance improvement achieved when using "DIP global" as the NCPA calibrator.

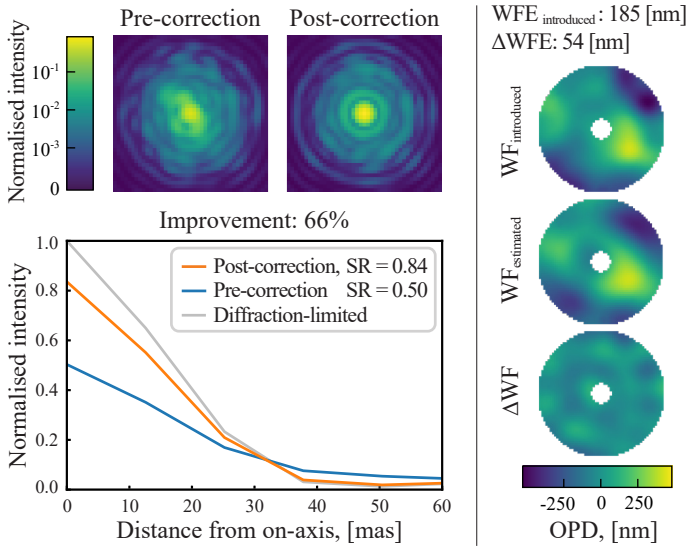


Fig. 4. Averaged PSF cubes simulated with (post-correction) and without (pre-correction) compensating for the NCPAs (see left column, top row). The improvement observed in the "post-correction" case indicates the successful retrieval of the quasi-static phase component associated with NCPAs. The right column demonstrates the corresponding wavefronts and their difference. See Eq. (9) for notation details. Simulated with $\lambda = 1600$ nm.

The top left part of the figure presents averaged PSF cubes illustrating the impact of NCPAs on a focal-plane PSF before and after applying the DIP-measured correction. The PSFs shown are simulated under identical conditions to those that produced the data for Fig. 3 (see Table 3). However, presented PSFs exclude the effect of the astigmatic diversity to improve visual clarity, although diversity was used in estimation. Specifically, the "pre-correction" PSF cube is defined by the wavefront $\mathit{\Phi} - \mathbf{1}_N \otimes \phi_{\mathrm{d}}$, while the "post-correction" PSF cube was computed using the $\mathit{\Phi} - \mathbf{1}_N \otimes (\phi_{\mathrm{d}} + \hat{\mathbf{\Phi}})$ as input wavefronts. In other words, for this plot, the correction was simulated by subtracting the estimated wavefront from the simulated wavefront cube.

The right column of this figure demonstrates the simulated and estimated averaged wavefront cubes ($\mathrm{WF}_{\mathrm{introduced}}$ and $\mathrm{WF}_{\mathrm{esimated}}$, respectively) associated with the depicted PSFs. We note that 29 modes were selected for estimation due to specific SPARTA RTC limitations. As elaborated on in Sect. 6.1, compensation of NCPAs via the DSM can be implemented by adding constant modal offsets to the WFS slopes. Currently, SPARTA allows us to set up to 29 Zernike modal offsets. Although this limitation can be technically overcome, for simplicity, we standardise on $M = 29$ for the entirety of this paper. Further details on selecting the optimal number of estimated modes are available in Appendix B.

5.6. Calibrated LIFT

After demonstrating the ability of the calibration stage to retrieve quasi-static NCPA components, we combine it with the LIFT estimation stage. The plot in Fig. 5 compares the performance of LIFT by using a calibrated prior (right) against its operation without such a prior (left).

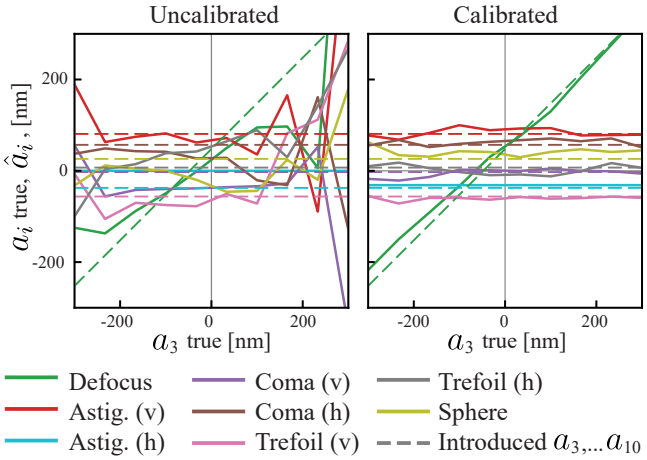


Fig. 5. Estimation of LO modes with LIFT on a simulated dataset of PSFs with ramping defocus. This test is similar to the one shown in Fig. 1. This time, defocus was estimated for the same PSF dataset with and without introducing the pre-calibrated NCPAs prior. A substantial increase in LIFT estimation accuracy is observed when calibration is applied. Solid lines represent the estimated modal coefficients $(\hat{a}_1, \ldots \hat{a}_{10})$, dashed lines represent corresponding introduced values. Simulated seeing = 0.35″, m_J = 12, $\mathrm{WFE}_{\mathrm{introduced}} \approx 140$ nm RMS. See Appendix C for more information on seeing selection.

Essentially, this figure replicates the experiment initially depicted in Fig. 1 in a simulated environment. To produce the PSFs for the second stage, we generated a series of PSFs with varying defocus values within the range $a_3 = -300 \ldots 300$ nm RMS. The other simulation parameters were adopted from Table 1. In addition to PSFs for LIFT, a PSF cube was generated and preemptively estimated with DIP to retrieve the quasi-static NCPAs. In this case, simulation parameters from Table 3 were utilised. Notably, the same set of NCPAs was employed for both the LIFT PSFs and the DIP PSF cube, albeit different realisations of HO and TT residuals were added in both cases. In scenarios where calibration was applied, the prior information from DIP was incorporated into LIFT as an additional phase diversity term, effectively leading to $\phi_{\mathrm{d,calibrated}} = \phi_{\mathrm{d}} + \hat{\mathbf{\Phi}}$. Figure 5 indicates a marked improvement in the precision of defocus estimation, as well as the estimation of other modes, upon utilising calibrated prior.

The results show that LIFT provides more accurate and stable estimates when calibrated. Most importantly, calibrations eliminate the need to estimate modes beyond TTF with LIFT since the higher orders associated with quasi-static NCPAs are 'absorbed' by the prior provided by DIP. Overall, notwithstanding their rather illustrative nature, these tests support our proposal for the two-stage approach in the simulated environment and ensure the following on-sky verification.

6. Second on-sky tests

In this section, we describe how we verified the two-stage approach with a second set of on-sky tests. The organisation of this section is similar to that of the preceding one, beginning with an examination of the calibration stage in isolation, followed by an assessment of its integration with LIFT. Moreover, an additional sensitivity test complements these results. The key distinction is that all the results discussed in this section are exclusively derived from on-sky observations obtained using the VLT UT4 telescope.

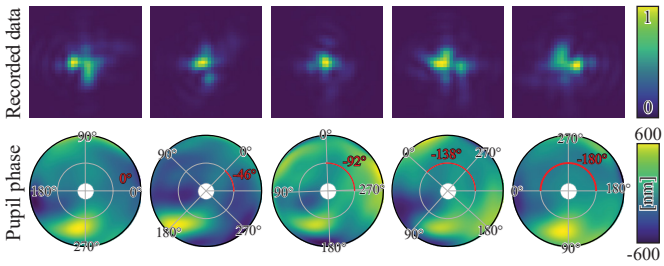


Fig. 6. Wavefronts (bottom row) estimated using on-sky PSFs (top row) with different manually set field derotator angles. A notable 'phase bump' can be consistently seen across all estimates. Note that the images of wavefronts are rotated in accordance with the derotator angle. The samples were recorded under the seeing ≈1.2″, $\lambda = 1600 \pm 20$ nm.

Table 4. Acquisition parameters for the derotator test.

Parameter	Value
Target m_J	8
Exposure time per sample (s)	0.1
PSFs per cube N	600
Number of introduced angles	9
Angles range (deg)	0, 23, ..., 161, 180
a_5 diversity (DSM) (nm)	170
Wavelength $\lambda \pm \Delta\lambda$ (nm)	1600 ± 20

6.1. Calibration stage on-sky

The significant challenge encountered during on-sky testing is the lack of reliable ground truth. As discussed in Sect. 4, controlling the defocus introduced by shifting the focal stage of MUSE is feasible. However, it is hard to characterise the NCPAs met in the IRLOS optical path independently with another method. Furthermore, NCPAs can be induced not only by phase errors but also by amplitude distortions. However, we assume that all NCPA-related effects observed in the recorded on-sky PSFs are exclusively associated with phase.

Due to the absence of reliable ground truth, we first conducted qualitative tests as presented in Fig. 6. This involved recording a series of measurements using IRLOS. The procedure involved manually rotating the MUSE field derotator and recording a PSF cube for each angle value. Here, the consistency of the estimated phase over different derotator angles serves as a proxy for veracity. The parameters of the acquisition process are detailed in Table 4. It is important to remember that the optical design of IRLOS does not support the simultaneous use of a narrow band filter and the astigmatic lens. Consequently, to facilitate data collection in a monochromatic regime, we introduced astigmatic diversity using the deformable secondary mirror (DSM) of UT4 instead of the lens.

It is worth noting that using a narrow-band filter can introduce NCPAs (for example, those encountered when placing a glass plate in a convergent beam). These aberrations will then be present during the calibration but absent during the focus retrieval stage, potentially leading to biased estimations. To alleviate this issue, it is possible to perform NCPA calibration using multiple narrow-band filters and compare the results. However, in these experiments, we assume that the narrow-band filter is free of aberrations.

The PSF cubes shown in Fig. 6 underwent estimation using the "DIP global" calibration method. The retrieved phase screens

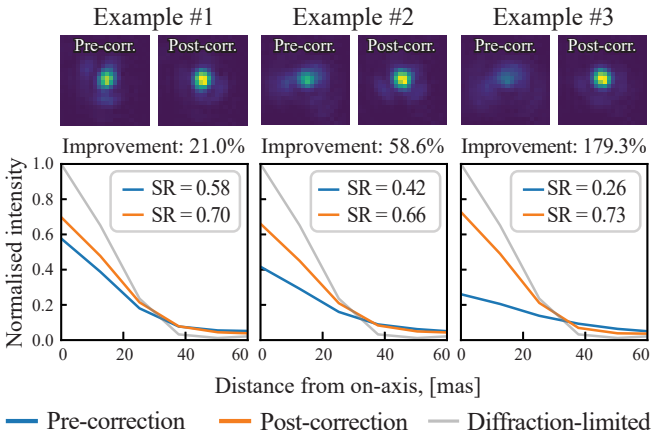


Fig. 7. Improvement in on-sky PSF quality after correcting NCPAs with DSM. The top row shows the recorded in-focus PSFs, while the corresponding radial profiles are displayed in the bottom row. For visual clarity, the astigmatic diversity used during estimation was removed while recording these PSFs. Seeing is ≈0.5", λ = 1600 ± 20 nm.

depicted in Fig. 6, demonstrate the consistency across various derotator angles. The observed variability in the measurements can be attributed to imprecise estimations and the existence of NCPAs that remain non-rotating due to their origin beyond the derotator in the optical path. Notably, a distinct bump-like feature is consistently observed as shifting around the field with different derotator angles. Even though our method is agnostic to the origin of this anomaly, it still offers a means to quantify it.

While primarily qualitative, this experiment demonstrates the capacity to retrieve relatively consistent phase estimates using DIP. Further tests were conducted to examine if DIP-measured NCPAs could be compensated for by using DSM to improve the quality of the in-focus IRLOS PSFs. The results are illustrated in Fig. 7.

Here, the focal-plane PSFs captured before and after correction are displayed. The idea behind this figure is similar to Fig. 4, but with an on-sky execution. The phase retrieved by DIP was sent to DSM as modal offsets to perform the correction. The leftmost result was obtained when DIP was applied to the system 'as is' and without introducing artificial errors. In this case, IRLOS was optimally refocused beforehand to achieve the best PSF quality. The middle and the right plots represent the cases affected by some uncontrolled defocus bias, which was accurately measured by DIP and subsequently corrected using DSM commands. In the same fashion as in Fig. 4, astigmatic diversity was employed during the data acquisition used for DIP estimation, but was not present in the optical path while recording these PSFs to enhance visual clarity.

In summary, these findings validate the ability of DIP to accurately retrieve meaningful information about the quasi-static phase component in the complex on-sky environment, further supporting the hypothesis that some measured NCPAs may indeed relate to phase aberrations, at least to some extent. The forthcoming investigation will involve integrating the calibration process with LIFT for on-sky testing.

6.2. Two-stage approach on-sky

An experiment presented in this subsection is very similar to the one explained earlier in Sect. 4. However, this time, the PSF cube for NCPA estimations was recorded before recording the dataset of defocused PSFs for LIFT, which is the key difference.

Table 5. Acquisition parameters for NCPAs retrieval test.

Parameter	Value
Target m_J	7.1
Exposure time per sample (s)	0.1
PSFs per cube N	300
a_5 diversity (DSM) (nm)	170
Wavelength $\lambda \pm \Delta\lambda$ (nm)	1600 ± 20

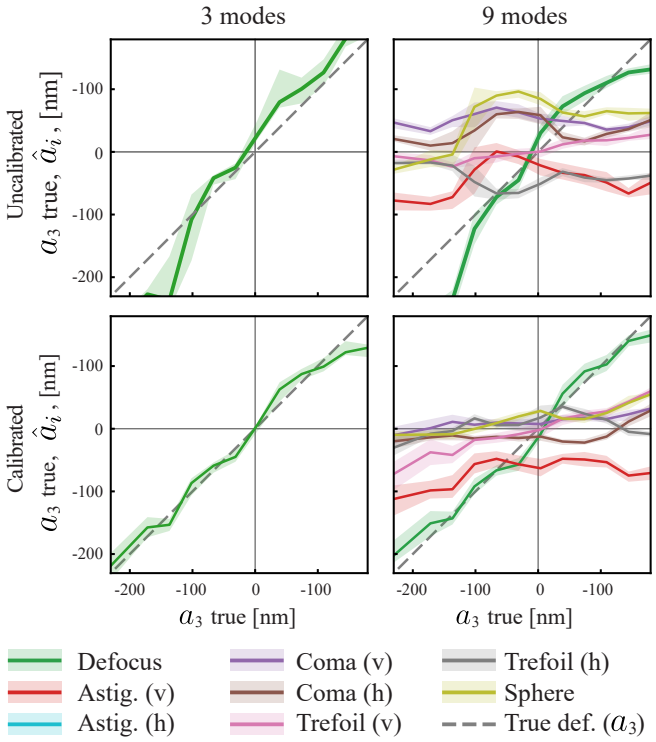


Fig. 8. LIFT estimates for the second set of defocused on-sky PSFs. The experiment is similar to that shown in Fig. 1, but this time, pre-calibration was also performed, resulting in higher estimation accuracy for defocus and higher-order modes (bottom row) compared to the uncalibrated case (upper row). The error bars are the 1σ confidence interval. NGS is m_J = 7.1, seeing is ≈0.5".

Another difference is the monochromatic nature of the recorded LIFT PSFs, which were captured using the H-band filter and with diversity introduced by DSM. This deviation was necessitated by the limited on-sky time available. Nevertheless, these changes are minor and do not alter the fundamental conditions of the experiment. The acquisition parameters for this test are listed in Table 5. It is crucial to highlight that for this particular test, the NCPAs identified through DIP were not physically corrected using the DSM, unlike in the experiment conducted in the previous subsection. Instead, these measurements were added to the phase diversity term inside LIFT. This approach aligns with the procedures discussed earlier in the simulated experiment shown in Fig. 5. Although it was technically feasible to pre-compensate for NCPAs using DSM and record an additional dataset of corrected PSFs with ramping defocus, it would require more on-sky time.

6.3. Two-stage approach in the low-flux conditions

It is important to note that all the results previously detailed in this paper were conducted in high-S/N conditions. However, as

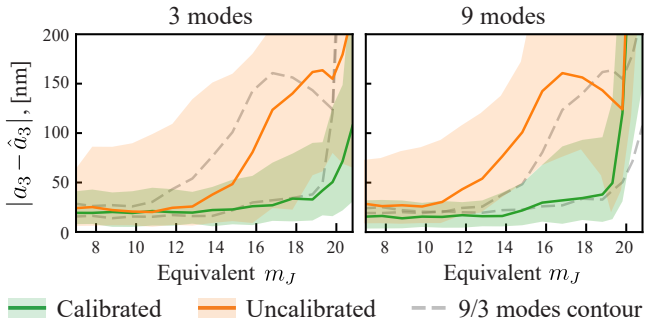


Fig. 9. Sensitivity scans for the on-sky PSFs dataset. Synthetic noise was added to simulate ranging flux levels. The results again demonstrated a notable improvement in accuracy and stability defocus retrieval when NCPA pre-calibration (green line) compared to the direct estimation with LIFT.

mentioned earlier, one of the main motivations for using LIFT is its robustness to high noise. Hence, we conducted an additional sensitivity test demonstrated in Fig. 9. This plot presents the absolute error of defocus computed for each sample as $|\, a_3 - \hat{a}_3 \,|$, where $a_3, \hat{a}_3$ are introduced and estimated defocus values. The introduced defocus spans from –170 to 170 nm RMS.

This test uses the same dataset as in Fig. 8. Essentially, it is a defocus ramp estimation akin to the one from the previous subsection. However, here, it is performed under varying S/Ns. The PSFs are re-normalised to different flux levels with the addition of simulated noise to achieve the desired S/N. The semi-synthetic noisy PSFs are subsequently estimated with LIFT, both with and without employing the calibrated prior. Both three (TTF) and nine modes $(a_1, \ldots a_{10}) \setminus \{a_5\}$ are estimated. Ten semi-synthetic PSFs with different noise realisations are generated for each magnitude and defocus value. The flux is normalised to 150 000 photons per UT4 aperture per second for $m_J = 15$. An equivalent m_J is obtained by analysing the photons collected in the H-band, namely, the band in which PSFs were originally recorded. The acquisition parameters of the original on-sky PSFs are consistent with the ones listed in Table 5.

Figure 9 illustrates that, in practice, employing calibration significantly enhances the accuracy and stability of LIFT estimates, particularly under low-flux conditions. It also confirms that when calibration is applied, it is sufficient to retrieve only TTF modes without sacrificing the quality of the defocus measurements due to model incompleteness, which aligns with findings from earlier simulated experiments. Moreover, the TTF-only (three modes) estimation fails at slightly higher magnitudes compared to the case with additional modes estimated (nine modes). These results reinforce the efficacy of incorporating a calibration strategy and support the core promise of this paper, demonstrating its validity even under the challenges met in the real system in realistic on-sky conditions.

7. Conclusion

In this study, we conducted one of the first on-sky validations of the LIFT technique, exploring its application challenges in real-world settings using IRLOS, the low-order WFS of MUSE-NFM. Our findings show that LIFT can successfully operate as a LO WFS on the real system. However, our investigation also highlights that its accurate and consistent operation necessitates estimating additional modes beyond defocus. This requirement stems from the complex and polychromatic nature of real on-sky PSFs influenced by the strong uncompensated NCPAs. To circumvent these challenges and enhance the accuracy and stability of LIFT, we proposed a two-stage strategy, where the estimation of additional NCPA modes is offloaded to a distinct (offline) calibration step that provides a prior that is later utilised by LIFT during the (online) focus retrieval stage. Both the simulated and experimental verification of this approach demonstrates that adopting this method improves the accuracy of LIFT, including the low-flux scenario. In summary, these findings indicate that LIFT is a very flexible and capable solution for LO WFSing. It can be effectively implemented in real systems and can operate with suitable precision when it is properly pre-calibrated. Furthermore, the calibration technique itself can be utilised as a minimally invasive method for direct on-sky NCPA characterisation. Most importantly, while our study focuses on IRLOS, the methodologies discussed here are adaptable and potentially beneficial for a wide range of focal-plane imaging systems beyond the specific context of IRLOS. To date, applying LIFT within a pre-calibrated framework has not been tested using the astigmatic lens. Although the experiments described in Sect. 6.3 largely reflect the operational conditions anticipated for LIFT, validating the method under conditions close to the real operations remains crucial. Therefore, more on-sky tests are necessary. In addition, this work primarily studies LIFT in a 'bootstrapping' regime, where large defocus values are retrieved. However, closing the slow defocus loop on LIFT and testing it in this operational regime remains a topic for future research. Moreover, the issue of the temporal stability of NCPAs requires thorough characterisation and rigorous consideration in future studies.

The prospect of integrating DIP as an alternative to LIFT in the focus retrieval stage must also be considered. The present adoption of LIFT can be attributed to its relative simplicity and reduced computational requirements in contrast to the more resource-intensive implementation of DIP. Moreover, ongoing efforts to apply LIFT and DIP within the context of petaling and LWE sensing, as initiated in Agapito et al. (2022), present a promising avenue for expanding this research focus.

Acknowledgements. We would like to thank Johann Kolb and Pavel Shchekaturov for their assistance in preparing and conducting the on-sky experiments. This work benefited from the support of the European Southern Observatory, French National Research Agency (ANR) with WOLF (ANR-18-CE31-0018), APPLY (ANR-19-CE31-0011) and LabEx FOCUS (ANR-11-LABX-0013); the Programme Investissement Avenir F-CELT (ANR-21-ESRE-0008), the Action Spécifique Haute Résolution Angulaire (ASHRA) of CNRS/INSU co-funded by CNES, the ECOS-CONYCIT France-Chile cooperation (C20E02), the ORP-H2020 Framework Programme of the European Commission's (Grant number 101004719), STIC AmSud (21-STIC-09), the Région Sud and the french government under the France 2030 investment plan, as part of the Initiative d'Excellence d'Aix-Marseille Université A*MIDEX, program number AMX-22-RE-AB-151.

References

Acton, D. S., Knight, J. S., Contos, A. R., et al. 2012, in Other Conferences
Agapito, G., Busoni, L., Carlà, G., et al. 2022, in Adaptive Optics Systems VIII, 12185, eds. L. Schreiber, D. Schmidt, & E. Vernet, International Society for Optics and Photonics (SPIE), 1218556
Almoro, P. F., & Hanson, S. G. 2008, Appl. Opt., 47, 2979
Björck, Å. 1996, Nonlinear Least Squares Problems (SIAM: Society for Industrial and Applied Mathematics), 339
Blanc, A., Mugnier, L. M., & Idier, J. 2003, J. Opt. Soc. Am. A, 20, 1035
Bos, S. P., Vievard, S., Wilby, M. J., et al. 2020, A&A, 639, A52
Brady, G. R., Guizar-Sicairos, M., & Fienup, J. R. 2009, Opt. Express, 17, 624
Campbell, H. I., Zhang, S., Greenaway, A. H., & Restaino, S. 2004, Opt. Lett., 29, 2707
Chanan, G., Troy, M., Dekens, F., et al. 1998, Appl. Opt., 37, 140
Dean, B. H., & Bowers, C. W. 2003, J. Opt. Soc. Am. A, 20, 1490

Esposito, S., Pinna, E., Tozzi, A., Stefanini, P., & Devaney, N. 2003, Proc. SPIE, 5169
Fétick, R. J. L., Fusco, T., Neichel, B., et al. 2019, A&A, 628, A99
Fienup, J. R. 1982, Appl. Opt., 21, 2758
Finger, G., Baker, I., Alvarez, D., et al. 2014, in Adaptive Optics Systems IV, 9148, eds. E. Marchetti, L. M. Close, & J.-P. Véran, International Society for Optics and Photonics (SPIE), 914817
Gerchberg, R. W. 1972, Optik, 35, 237
Gonsalves, R. A. 1982, Opt. Eng., 21, 829
Gonsalves, R. A. 2001, Opt. Lett., 26, 684
Huang, G., Wu, D., Luo, J., et al. 2021, Photon. Res., 9, 34
Korkiakoski, V., Keller, C. U., Doelman, N., et al. 2014, Appl. Opt., 53, 4565
Lamb, M. P., Correia, C., Sauvage, J.-F., et al. 2017, J. Astron. Telescopes Instrum. Syst., 3, 039001
Lamb, M., Norton, A., Macintosh, B., et al. 2018, SPIE Conf. Ser., 10703, 107035M
Lamb, M. P., Correia, C., Sivanandam, S., Swanson, R., & Zavyalova, P. 2021, MNRAS, 505, 3347
Lee, D. J., Welsh, B. M., Roggemann, M. C., & Ellerbroek, B. L. 1997, Opt. Lett., 22, 952
Liu, D. C., & Nocedal, J. 1989, Math. Program., 45, 503
Lloyd-Hart, M., Wizinowich, P., McLeod, B., et al. 1992, ApJ, 390, L41
Meimon, S., Fusco, T., & Mugnier, L. M. 2010, Opt. Lett., 35, 3036
Mocœur, I., Mugnier, L. M., & Cassaing, F. 2009, Opt. Lett., 34, 3487
Mugnier, L. M., Sauvage, J.-F., Fusco, T., Cornia, A., & Dandy, S. 2008, Opt. Express, 16, 18406
Noll, R. J. 1976, J. Opt. Soc. Am., 66, 207
Oberti, S., Kolb, J., Madec, P.-Y., et al. 2018, SPIE Conf. Ser., 10703, 107031G
Plantet, C. 2014, PhD Thesis, Observatoire de Paris
Plantet, C., Meimon, S., Conan, J.-M., & Fusco, T. 2013, Opt. Express, 21, 16337
Plantet, C., Meimon, S., Conan, J.-M., Neichel, B., & Fusco, T. 2014, SPIE Conf. Ser., 9148, 91482D
Plantet, C., Wizinowich, P., Fusco, T., et al. 2017, in Proceedings of the Adaptive Optics for Extremely Large Telescopes 5 (Instituto de Astrofísica de Canarias (IAC))
Polo, A., Pereira, S. F., & Urbach, P. H. 2013, Opt. Lett., 38, 812
Pourré, N., Le Bouquin, J. B., Milli, J., et al. 2022, A&A, 665, A158
Robert, C., Fusco, T., Sauvage, J.-F., & Mugnier, L. 2008, SPIE Conf. Ser., 7015, 70156A
Sauvage, J.-F., Fusco, T., Lamb, M., et al. 2016, SPIE Conf. Ser., 9909, 990916
Shi, F., Chanan, G., Ohara, C., Troy, M., & Redding, D. C. 2004, Appl. Opt., 43, 4474
Smith, C. S., Marinică, R., den Dekker, A. J., et al. 2013, J. Opt. Soc. Am. A, 30, 2002
Terreri, A., Pedichini, F., Kasper, M., et al. 2022, SPIE Conf. Ser., 12185, 1218587
Valles, M. S., Fedrigo, E., Donaldson, R. H., et al. 2012, in Adaptive Optics systems III, 8447, eds. B. L. Ellerbroek, E. Marchetti, & J.-P. Véran, International Society for Optics and Photonics (SPIE), 84472Q
Wang, M., Guo, W., & Yuan, X. 2021, Opt. Express, 29, 3465
Wilby, M. J., Keller, C. U., Sauvage, J.-F., et al. 2018, A&A, 615, A34
Wilding, D., Pozzi, P., Soloviev, O., Vdovin, G., & Verhaegen, M. 2017, in Optics in the Life Sciences Congress (Optica Publishing Group), BoTu2A.2
Wu, D., Yang, C., Zhang, P., et al. 2018, J. Astron. Telescopes Instrum. Syst., 4, 049007

Appendix A: Statistics of LO modes

In the left plot on Fig. A.1, the standard deviation (STD) of low-order (LO) modes was calculated using the unbiased 'DIP global' method on a dataset of real short-exposure on-sky PSFs (indicated by pale blue dots), where 100 Zernike modes were estimated. An empirical exponential law was then fitted to this data, shown by the green line.

After 50-60 modes (left plot), the STD of coefficients reaches a plateau. We attribute this plateau primarily to estimation noise due to cross-talk with HO residuals rather than to NCPAs. Consequently, in our simulations, we have limited the number of LO modes modelled to 50.

The right plot on Fig. A.1 shows the STD of modes (represented by orange dots) estimated from synthetically generated PSFs, for which 100 Zernike modes were again estimated. These PSFs were created based on the law (blue line) that derives from the best fit to the on-sky data (green line). However, it was truncated to include only 50 modes, and the standard deviations of the tip and tilt were equalised.

The magenta line represents the best fit for the synthetic distribution of coefficients STD. From the distribution of the estimated coefficients, we can see that although only 50 modes were introduced into the simulated LO, the estimates also exhibit plateauing behaviour. Similarly to on-sky, it can also be attributed to the coupling with HO residuals, which are simulated, pointing out that the truncation approach is valid.

The magenta line illustrates the best fit for the synthetic distribution of coefficient STDs. Observations from this distribution reveal that, although only 50 modes were incorporated into the simulated LO, the estimates still display a plateauing trend for higher orders. This is similar to the on-sky results and is likely due to the coupling with HO residuals, indirectly confirming the validity of the truncation approach.

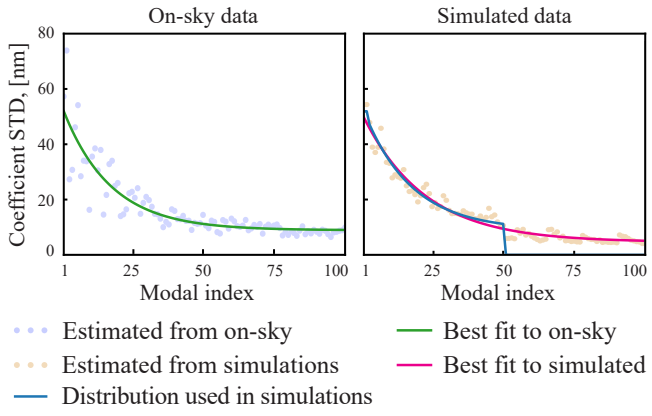


Fig. A.1: Standard deviation of the LO coefficients obtained using on-sky PSFs (left) and realistically simulated PSFs (right). The synthetic PSFs were generated using the LO distribution law (blue line), derived from the distribution previously estimated from on-sky data (green line).

Appendix B: Modal cross-talk and overfitting

To illustrate the modal cross-talk (coupling) issue, we conducted a set of tests (see Fig. B.1) in which the PSF dataset was simulated with a fixed number of introduced LO modes. Meanwhile, the number of estimated modes varied.

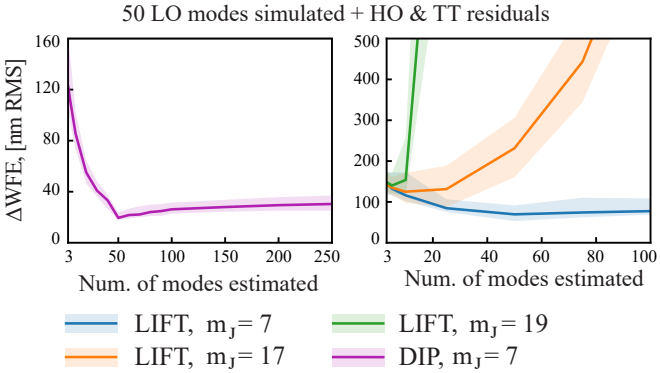


Fig. B.1: Impact of the number of estimated modes on estimation error. The results are shown for 'DIP global' (left) and 'LIFT avg.' (right) methods. The right plot demonstrates that reducing the number of estimated modes enhances accuracy in low-flux scenarios.

For the DIP (left plot), the optimal number of estimated modes equals the number of modes introduced. Below 50 modes, the model is not representative enough, leading to high estimation error. In contrast, sensing more than 50 modes causes overfitting, which leads to the modal cross-talk that causes deviations from the true values.

These results can be related to the practical question of determining the optimal number of modes for calibration. The findings illustrated in the left plot of Fig. A.1 suggest that the most pertinent information from LO is contained within the first 50-60 orders. Thus, this number of estimated modes can be considered optimal for the calibration stage, although in practice, it was limited to 29 to ease the hardware implementation.

Figure B.1 shows the relationship between the estimation error of LIFT and the number of estimated modes for different NGS magnitudes. For bright targets, the highest accuracy is achieved with 50 modes. However, with a rising target magnitude, including fewer modes helps to regularise the problem better, prevents coupling, and reduces noise propagation.

Appendix C: Seeing selection

The simulation depicted in Fig. C.1 illustrates the distribution of seeing measured by SPARTA, based on WFS data from AOF, and by MASS-DIMM. Both methods independently estimate the seeing, although a noticeable shift is observed, with MASS-DIMM consistently being more pessimistic. The same results were earlier presented in Fétick et al. (2019).

It is crucial to highlight that the seeing values presented in this paper with on-sky data are measured using MASS-DIMM. In contrast, the simulated seeing values are more related to SPARTA. This is because the simulated TT and HH residuals were normalised based on their expected values, which were mapped to the measurements taken by SPARTA. In this case, a seeing value of 0.35" indicates quite good conditions, whereas 1" suggests rather poor conditions. This paper utilises these two values in simulations to reflect both cases.

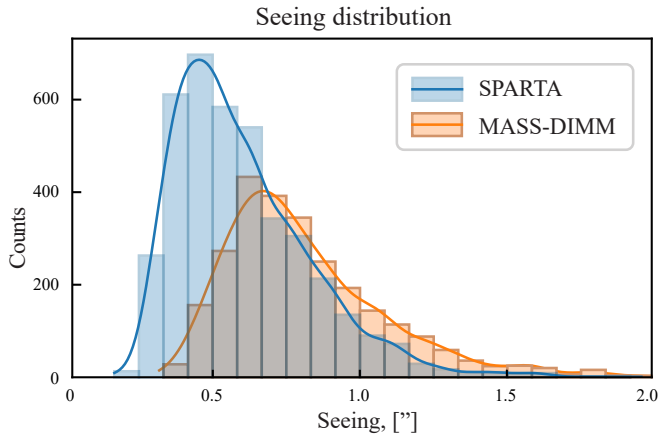


Fig. C.1: Seeing distribution at Paranal from 2015 to 2020 as measured by MASS-DIMM and SPARTA.

In summary, this paper underscores the efficacy of the proposed and validated approaches. However, the paper raised some important points that lacked detailed discussion :

— **Parameter coupling :** This concept needs to be clarified. Additionally, it raises the question of how to quantify this coupling and understand its impact on the coefficient estimation process.

— **Choice of loss function :** The rationale behind choosing a specific optimization criterion (loss function) with DIP needs further exploration. Understanding how this choice affects the estimation of coefficients is also important.

— **Optimization Approaches :** The reasons behind the superior performance of L-BFGS (BFGS) compared to LIFT in the scope of this problem require investigation. Reviewing the underlying mathematical principles of these optimization methods is important to explain the differences in the estimation results.

Addressing these questions not only complements the results, but also introduces concepts that will be useful later in Section 4.7.

3.2. Modal cross-talk

To better understand this phenomenon, it is worth revisiting Eq.(1) in the paper that defines the PSF model used in LIFT. As mentioned previously, the parameters of this model are the Zernike coefficients. Although the Zernike modal basis is orthogonal in the pupil plane by definition, this property is lost after propagation to the focal plane. This occurs because the complex phase term $\exp\left(i\frac{2\pi}{\lambda}\Phi(\mathbf{u})\right)$ mixes information, making the coefficients difficult to disentangle. Furthermore, Fourier propagation spreads the effect of each Zernike mode across the entire focal plane, with stronger localization in certain regions depending on the mode's spatial frequencies. Finally, conversion to intensity destroys the phase information. These factors result in coupling between different modes, which means that different modes exhibit similar effects on the morphology of the PSF. For example, a weak tip/tilt resembles adding a bit of coma since both modes slightly shift the PSF, while adding weak astigmatism is very similar to adding a small amount of defocus aberration.

In this context, one could consider using a modal basis different from Zernike to minimize cross-talk in the focal plane. An intuitive example of such a basis would be a set of pure 2D harmonics in the pupil plane. A pure sine wave propagated to the focal plane results in a pair of δ-functions with the negative sign convolved with an Airy pattern. However, when converted to intensity, this transformation yields two identical δ-functions, indicating that a $\pi/2$-shifted wave would produce the same intensity pattern in this case. Additionally, convolution with the Airy pattern "spreads" information across adjacent pixels. Therefore, modal coupling is inevitable when using this definition of the PSF model without additional extensions, assumptions, or

simplifications [1].

In other words, parameter coupling refers to a situation where parameters are interdependent, meaning that the adjustment of one parameter influences (or is influenced by) the adjustment of another, often in a strongly non-linear manner. As a result, coupling can lead to non-linear effects on the objective function, making the solution more complex. This is closely related to the concept of the convexity of the objective function. In this context, a *"convex"* function means that any local minimum is also the global minimum, whereas a *"non-convex"* function may have multiple local minima, maxima, and saddle points, making it challenging to find the global solution directly. Consequently, non-convex problems often require iterative approaches to find the solution. This is the reason why the phase retrieval problem in LIFT is approached via non-linear optimization. However, this is more general than the problem of coefficient retrieval and applies to many non-linear problems. In this scope, various advanced optimization algorithms, regularization techniques, constraints, and reparameterization are used to mitigate coupling effects.

In the context of LIFT, it is helpful to visualize the modal coupling. Figure 3.1 provides this visualization by showing the correlation plots computed for pairs of modes for a specific PSF (depicted in the middle right of the figure). The diagonal plots represent the marginalized posterior distributions for the corresponding modes.

1. In theory, principal component analysis (PCA) applied to PSF images can create a sort of focal-plane modal basis. However, this basis will only approximate the PSF morphology, while mapping it back to the pupil plane unambiguously would be a challenge.

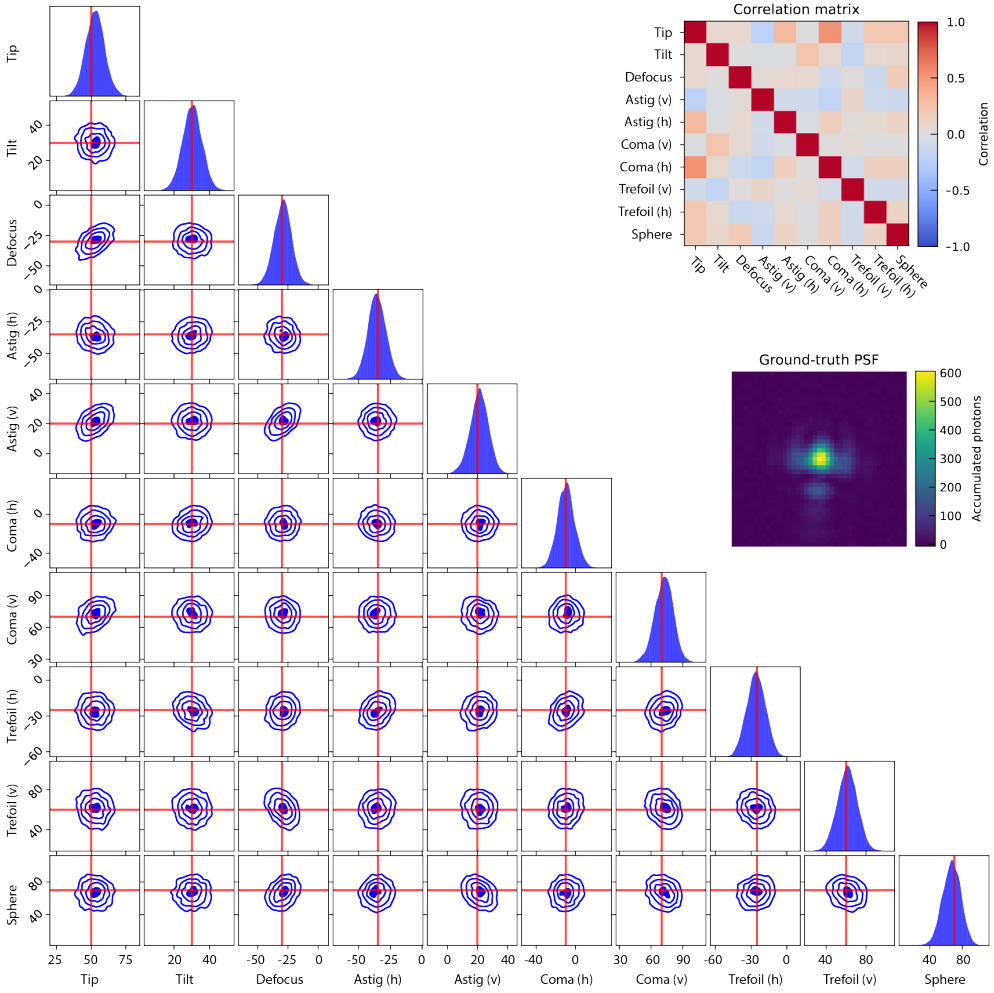


FIGURE 3.1. – *Corner plot :* log-likelihood function sampled in the vicinity of the true modal coefficient values. The couplings between different modes can be seen as elongations in correlation plots. *Top-right :* the correlation matrix between the modal coefficients. *Right-middle :* the ground-truth simulated high-flux PSF used to compute the likelihood. No AO residuals were introduced in the simulation.

The figure is computed for a toy example in which the synthetic PSF is defined only by the first ten Zernike modes $\mathbf{A} = [50, 30, -30, -35, 20, -10, 70, -25, 60, 70]$ nm without AO residuals introduced. The introduced diversity is 170 nm RMS (vertical astig., a_5), and the simulated J-band magnitude $m_J = 14$. Other simulation parameters are taken from Table 3 in the embedded paper. This graph was generated by sampling the log-likelihood function similar to that utilized to derive the mathematical formalism of LIFT :

$$\log L(\mathbf{A}) = -\sum_{i}^{W}\sum_{j}^{H}\left(\frac{1}{R_{\mathrm{n},ij}}(I_{ij}(\mathbf{A}) - I_{\mathrm{data},ij})^2\right). \tag{3.1}$$

Here, $I_{ij}(\mathbf{A})$ is the output of the PSF model at i-th and j-th pixel, driven by the vector

of modal coefficients $\mathbf{A}$; $I_{\text{data},ij}$ is the preemptively simulated synthetic PSF with noise, W, H are the width and the height of PSF images; $\mathbf{R}_\text{n}$ is the pixels' covariance matrix that acts as a weighting map for pixels with high noise. For simplicity, in this plot $\mathbf{R}_\text{n} = \mathbf{1}_{W\times H}$.

The values of this function were sampled using the Nested Sampling approach implemented in the `dynesty`[2] package (see Speagle (2020)). Using a sampler here allows for covering a large space of possible coefficient values, presented as a corner plot. Essentially, Figure 3.1 shows the 2D slices of the multidimensional landscape of the log-likelihood function. This can also be understood as a landscape of the (negative) optimization criterion, meaning that the minimum of the optimization criterion corresponds to the maximum likelihood in this plot. This will be elaborated on in Section 3.3. For now, the terms *"likelihood"* and *"optimization criterion"* are used synonymously.

From 3.1, it is possible to see the correlation between some coefficients, which indicates coupling. Quantitatively, it can be estimated with covariance.

The covariance for two coefficients a_l and a_m is computed using :

$$\text{cov}(a_l, a_m) = \frac{1}{N_s - 1} \sum_{k=1}^{N_s} \left(a_l^{(k)} - \bar{a}_l\right)\left(a_m^{(k)} - \bar{a}_m\right),$$

where N_s is the total number of sampled values produced by the sampler, $a_l^{(k)}$ is the k-th sample of the l-th coefficient in the modal vector $\mathbf{A} = [a_1, a_2, \ldots, a_M]$, where $l, m = 1 \ldots M$ are the coefficient indices (in Noll's notation, see Noll (1976)). $\bar{a}_l$ is the mean of the samples for the l-th parameter :

$$\bar{a}_l = \frac{1}{N_s} \sum_{k=1}^{N_s} a_l^{(k)}.$$

The full covariance matrix Σ_A between all pairs of coefficients (a_l, a_m) is defined as

$$\Sigma_A = \begin{pmatrix} \text{var}(a_1) & \text{cov}(a_1, a_2) & \ldots & \text{cov}(a_1, a_M) \\ \text{cov}(a_2, a_1) & \text{var}(a_2) & \ldots & \text{cov}(a_2, a_M) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(a_M, a_1) & \text{cov}(a_M, a_2) & \ldots & \text{var}(a_M) \end{pmatrix},$$

where $\text{var}(a_l)$ is the variance of parameter a_l, defined as :

$$\text{var}(a_l) = \frac{1}{N_s - 1} \sum_{k=1}^{N_s} \left(a_l^{(k)} - \bar{a}_l\right)^2.$$

2. https://dynesty.readthedocs.io/en/stable/

The element $r_{A,lm}$ of the correlation matrix $\mathbf{r}_A$ is then defined as :

$$r_{A,lm} = \frac{\Sigma_{A,lm}}{\sigma_l \sigma_m}, \quad \sigma_l = \sqrt{\Sigma_{A,ll}}.$$

This correlation matrix is shown in the top-left part of Figure 3.1.

Visualizing and understanding parameter coupling is not only of illustrative interest but also critical to understanding the behavior of nonlinear optimizers. As demonstrated in the following, this understanding is vital for selecting appropriate algorithms and strategies to determine the optimal solution efficiently.

3.2.1. Selection of the diversity phase amplitude

As indicated in Section 2 of the paper, LIFT is a phase diversity algorithm. Hence, the selection of an appropriate amplitude of diversity is essential to ensure convergence to the true solution. The choice of amplitude diversity, along with its impact on coefficient coupling, can be demonstrated using the same log-likelihood sampling method. The results are presented in Figure 3.2(c). To avoid displaying the entire corner plot, this figure shows only the defocus (Z_3) and vertical astigmatism (Z_5) modes with their corresponding marginalized posteriors. The full versions of this graph can be found in Appendix A, Figure 2.

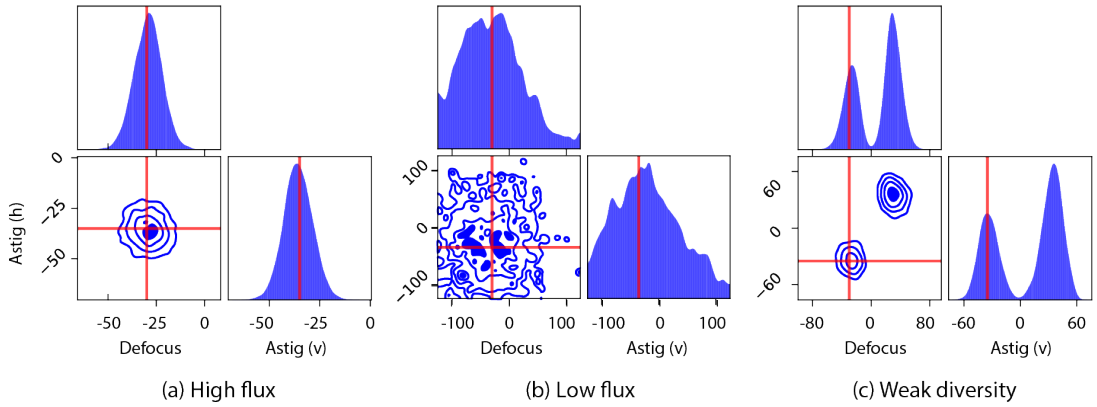


FIGURE 3.2. – Coupling between modes under varying conditions : high flux (a), low flux (b), and high flux with reduced phase diversity (50 nm RMS instead of 170 nm RMS) (c). The ground-truth coefficient values are equivalent to those in Figure 3.1.

As shown in Figure 3.2(c), choosing a weak diversity value leads to the emergence of a "parasitic" local minimum for certain modes. In this simulated PSF, the coefficient values are the same as in Figure 3.1, but the astigmatic diversity is set to 50 nm instead of 170 nm RMS. This minimum duality arises because there are always two wavefronts that produce the same intensity pattern in the focal plane, according to Eq.(1) in the paper. Consequently, within the scope of our PSF model, there are always two

sets of modal coefficients that result in the same PSF morphology. This issue is also directly connected to the sign ambiguity problem, where certain (even) modes with the same amplitude but different signs can produce the same or similar effects on the focal-plane PSF.

Just like modal coupling, the existence of double local minima represents a highly non-convex problem. To resolve the ambiguity, an astigmatic diversity phase is introduced into the PSF model in LIFT. This diversity biases the problem by shifting the zero point of the PSF model toward a region in the parameter space near the diversity value, effectively displacing the parasitic local minimum away. By adding enough diversity and providing a good initial parameter estimate, the parasitic minimum moved far enough away so that the solution is unlikely to converge to it. This is discussed in more detail below.

Toy example of the double minimum problem

To illustrate this principle, consider another PSF (denoted as $\mathbf{I}_0$) influenced solely by defocus and vertical astigmatism aberrations, where the true values for modal coefficients $a_3 = -100$ nm and $a_5 = 100$ nm RMS as depicted in Figure 3.3 (left). Furthermore, this PSF is affected by a vertical astigmatism (Z_5) diversity of -50 nm RMS. Compared to Figure 3.1, different coefficient values are introduced in this example for illustrative sake. AO residuals are not added, meaning $a_i = 0$, $\forall i \notin \{3,5\}$.

Then, determining the actual values of a_3 and a_5 becomes a two-dimensional optimization challenge. However, due to the intensity conversion in Eq.(1), a wavefront affected by $a_3 = -100$ nm and $a_5 = 100$ nm will produce exactly the same PSF as a wavefront with $a_3 = 100$ nm and $a_5 = 0$ nm. Since the diversity value is weak, both solutions $a_5 = 100$ nm and $a_5 = 0$ nm are close to each other and thus are likely to be reached by the optimizer if optimization starts from $\hat{\mathbf{A}} = \mathbf{0}$ point.

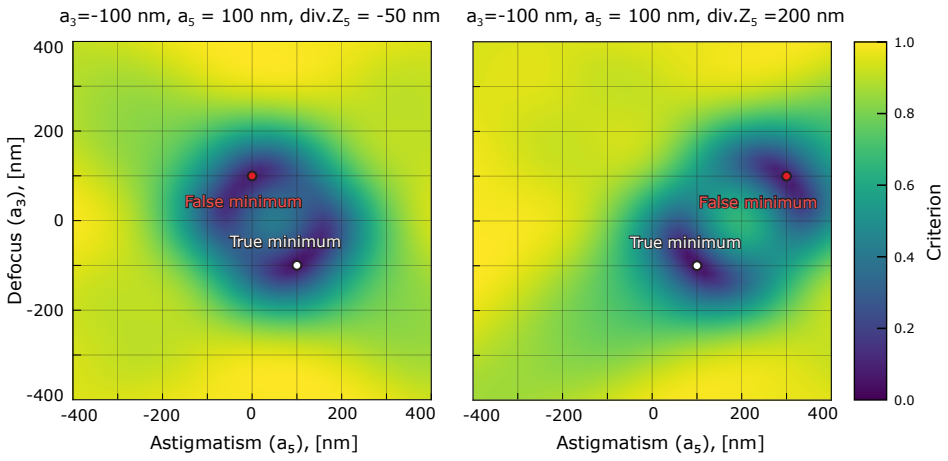


FIGURE 3.3. – 2D visualization of the optimization criterion when both a_3 and a_5 modes are estimated simultaneously with -50 nm RMS phase diversity (left plot) and 200 nm phase diversity (right plot). The error metric visualized in this figure is defined as $|I_0 - I(\mathbf{A})|$.

However, as Figure 3.3 (right) demonstrates, if the astigmatic diversity is strong and equals -200 nm RMS, morphologically, the same PSFs will be produced by $a_3 = -100$ nm, $a_5 = 100$ nm, and $a_3 = 100$ nm, $a_5 = 300$ nm. In this case, the initial guess for the parameters will determine which local minimum the optimizer converges to. For example, if the initial guess for the non-linear optimizer is $\hat{\mathbf{A}} = \mathbf{0}$, then the solution $a_5 = 300$ nm will be too far from $\hat{\mathbf{A}} = \mathbf{0}$, and thus it is less likely to be reached by the optimizer. Therefore, only one minimum will dominate the local criterion landscape within a certain vicinity. This is exactly what is shown in Figures 3.1 and 3.2(a).

Moreover, the value of phase diversity cannot be increased without limits. At a certain point, it will begin to dominate the PSF morphology, making the impact of other modes on it almost unnoticeable. Moreover, if PSF is obtained under low-flux conditions, the high diversity can spread the flux over too many pixels, thus making the estimation more sensitive to noise.

To summarize, the example above shows that the problem is fundamentally non-convex due to two local minima always being globally present. Therefore, increasing diversity and accurately choosing the optimization starting point can help, although they will not fully resolve the problem.

Avoiding double minima

One way to address this limitation is to introduce regularization, such as the Maximum A Posteriori (MAP) estimation. Essentially, formulating the log-likelihood in the MAP framework "warps" the landscape of the optimization criterion so that the values of the criterion increase the further they are from the mean point $\overline{\mathbf{A}}$, with the slope of this increase regulated by $\mathbf{C}_\phi$ (see Eq.(5) from the paper for more details).

However, as mentioned in the paper, MAP can bias the estimate. Therefore, as was also mentioned in Section 2 of the paper, in an unbiased framework, another regularization method can be applied : to reduce the number of estimated modes or exclude specific modes. In this case, vertical astigmatism aa_5 can be excluded from the estimation, turning it into a one-dimensional defocus retrieval problem within the scope of this simplified example. Then, with increasing diversity amplitude, the parasitic local minimum for defocus becomes weaker than the true global minimum. This is demonstrated in Figure 3.4, where the landscape of the optimization criterion is mapped as a function of defocus aberration and astigmatic diversity.

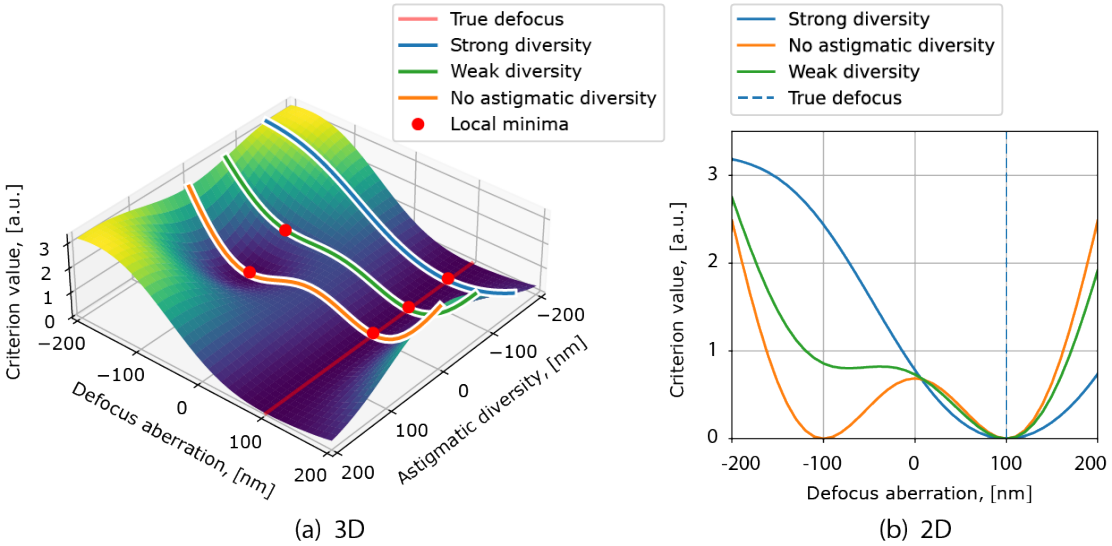


FIGURE 3.4. – Optimization criterion when only a_3 is estimated as a function of the introduced astigmatic diversity strength; (a) represents the 3D visualization of the criterion landscape, while (b) shows the 2D projection of the plot for different astigmatic diversity strengths (0, -100, and -200 nm RMS). The criterion here is the same as in Fig. 3.2.

Obviously, the trade-off for this regularization is the reduced representability of the PSF model, which now includes vertical astigmatism only as fixed phase diversity but not as a degree of freedom. Nevertheless, this explains why a_5 was truncated to obtain the results presented in Figure 1 in the paper. The a_5 coefficient was removed from the estimation to increase the stability of the non-linear optimization process.

In summary, this example illustrates how modal coupling can be understood by examining the landscape of optimization criterion. It demonstrates that parameter coupling can lead to highly complex, non-convex objective function landscapes with multiple local minima and maxima (saddle points), thereby complicating the optimization problem.

3.2.2. The effect of low flux

Another effect linked to parameter coupling and relevant to LIFT is the issue of low flux. In this case, when the criterion in Eq.(3.1) is computed relative to a noisy PSF, the global minimum of the optimization criterion becomes less defined since noise results in the loss of information in the PSF pixels, relative to which the discrepancy between the data and the model is computed. Additionally, noise splits a global minimum into numerous small local minima where the optimizer can become trapped. Figure 3.2(b) illustrates how the global minimum of the optimization criterion is distorted. A full version of this plot can be found in Figure 1 of Appendix A. As mentioned in the

paper, one way to mitigate this issue is to truncate the number of estimated modal coefficients. This reduces the dimensionality the optimizer must navigate, decreasing the number of local minima where it could get potentially stuck. This regularization method was applied in the embedded paper to achieve the results shown in Figure 9.

Another approach to solving this problem is to assign lower priority to pixels according to the amount of information they retain. For example, when the flux is low, more weight is assigned to brighter pixels, while noise-dominated pixels are weighted close to zero. As discussed in Section 2 of the paper, LIFT intrinsically handles this with the $\mathbf{R}_\mathrm{n}$ matrix, which contains the flux variance per pixel.

The effect of low SNR on the optimization criterion landscape is shown in Figures 3.2(b) and 1 in Appendix. For illustrative purposes, $\mathbf{R}_\mathrm{n} = \mathbf{1}_{W\times H}$ in these Figures to make it comparable with Figure 3.1.

In summary, the discussion in this section aims to show that many aspects of LIFT's behavior can be interpreted through the lens of non-linear optimization. From this perspective, improving LIFT's accuracy can be tackled computationally by reformulating the optimization problem and minimizing cross-talk/coupling between modal parameters. These principles extend beyond LIFT and apply to various phase-diversity-like problems, as demonstrated within the DIP framework discussed in the paper.

This section also demonstrated that understanding the (non-convex) landscape of the optimization criterion is crucial for gaining insights into the behavior of non-linear optimizers and predicting potential failure points. Therefore, the concept of the optimization criterion will be explored in more detail in the next section.

3.3. Negative log-likelihood and loss function

It has already been discussed that the log-likelihood and optimization criterion (loss function) play similar roles and are thus both mentioned synonymously in this discussion. However, their mathematical definitions differ, and this point needs some clarification.

Log-likelihood function

The likelihood function quantifies how well a statistical model with parameters $\mathbf{A}$ explains the observed data $\mathbf{I}_\mathrm{data} = \{I_{\mathrm{data},1}, \ldots, I_{\mathrm{data},N}\}$. Here, N is the number of observations or pixels in the PSF image in this context. The *likelihood function* $L(\mathbf{A})$ represents the joint probability of observing the data ($\mathbf{I}_\mathrm{data}$) under the assumed model with parameters $\mathbf{A}$. It is expressed as

$$L(\mathbf{A}\,|\,\mathbf{I}_\mathrm{data}) = \prod_{i=1}^{N} p(I_{\mathrm{data},i}\,|\,\mathbf{A}).$$

Here, $p(I_{\text{data},i}\,|\,\mathbf{A})$ is the probability of explaining the realization $I_{\text{data},i}$ given certain values of $\mathbf{A}$. Assuming that the data follows a Gaussian distribution with variance σ_i^2, the likelihood for each observation is :

$$p(I_{\text{data},i}\,|\,\mathbf{A}) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(I_{\text{data},i} - I_i(\mathbf{A}))^2}{2\sigma_i^2}\right).$$

Here, $I_i(\mathbf{A})$ is the i-th modeled sample. In many cases, Gaussian error distribution is valid due to the central limit theorem. However, when considering a PSF, per-pixel noise follows Poisson statistics for photons and Gaussian statistics for the detector noise. However, approximating the photon noise with Gaussian is still valid if the photon count is high enough. Thus, the joint likelihood can then be expressed as :

$$L(\mathbf{A}\,|\,\mathbf{I}_{\text{data}}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(I_{\text{data},i} - I_i(\mathbf{A}))^2}{2\sigma_i^2}\right). \tag{3.2}$$

Calculating the *log-likelihood* helps to reformulate the problem in a more computationally convenient way. Taking the log of the likelihood function :

$$\log L(\mathbf{A}\,|\,\mathbf{I}_{\text{data}}) = \sum_{i=1}^{N} \log p(I_{\text{data},i}\,|\,\mathbf{A})$$

Assuming that σ is the property of the data, which is the same for all data samples ($\sigma_i = \sigma, \forall i \in 1,\ldots,N$), it is possible to write down the log-likelihood for Gaussian distribution like :

$$\log L(\mathbf{A}\,|\,\mathbf{I}_{\text{data}}) = -\frac{N}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(I_{\text{data},i} - I_i(\mathbf{A})\right)^2.$$

The use of the log-likelihood function is particularly advantageous in this context. First, applying the logarithm preserves the locations of the likelihood function's extrema. Second, it transforms the expression into a sum of terms, making it significantly more manageable for computation and analysis. The constants here can also be dropped since they do not affect the optimization. Then, it is also possible to further eliminate the $1/\sigma^2$ multiplier from the equation, assuming that the error is the same for all samples. Thus, the expression simplifies to minimizing the sum of squared residuals :

$$\hat{\mathbf{A}} = \arg\max_{\mathbf{A}} \sum_{i=1}^{N} \log p(I_{\text{data},i}\,|\,\mathbf{A}) \;\rightarrow\; \hat{\mathbf{A}} = \arg\min_{\mathbf{A}} \sum_{i=1}^{N} (I_{\text{data},i} - I_i(\mathbf{A}))^2.$$

However, if one assumes a Laplacian distribution instead of the Gaussian one in

Eq.(3.2), the expression for joint likelihood takes another form :

$$L(\mathbf{A}\,|\,\mathbf{I}_{\text{data}}) = \prod_{i=1}^{N} \frac{1}{2b_i} \exp\left(-\frac{\left|I_{\text{data},i} - I_i(\mathbf{A})\right|}{b_i}\right), \tag{3.3}$$

where b is the diversity parameter that defines the width of the Laplacian curve, with standard deviation $\sigma = \sqrt{2}b$. Despite the universality of the central limit theorem, the Laplace distribution may sometimes be chosen over the Gaussian if it better describes the discrepancy between the data and the model.

Following a similar logic as in the expressions above for the Gaussian distribution, it is possible to derive the expressions for the optimal estimate of $\hat{\mathbf{A}}$ that maximizes the joint likelihood or minimizes the negative log-likelihood (NLL) for the case of the Laplacian distribution :

$$\hat{\mathbf{A}} = \arg\max_{\mathbf{A}} \left(-N\log(2b) - \frac{1}{b}\sum_{i=1}^{N}\left|I_{\text{data},i} - I_i(\mathbf{A})\right|\right) \rightarrow \hat{\mathbf{A}} = \arg\min_{\mathbf{A}} \sum_{i=1}^{N}\left|I_{\text{data},i} - I_i(\mathbf{A})\right|.$$

As will be described below, the expression derived in this section is crucial for understanding the loss functions later.

Loss Function

By definition, a loss function quantifies the discrepancy between the predicted (modeled) and observed values, typically by assigning a numerical penalty to deviations between them. The goal of a non-linear optimizer, in this case, is to reduce this discrepancy. For example, one of the most commonly used loss functions in machine learning, non-linear optimization, and statistical analysis in general is Mean Squared Error (MSE), defined as :

$$\text{MSE}\,(\mathbf{I}_{\text{data}}, \mathbf{I}(\mathbf{A})) = \frac{1}{N}\sum_{i=1}^{N}\left(I_{\text{data},i} - I_i(\mathbf{A})\right)^2. \tag{3.4}$$

This expression is mathematically equivalent to the NLL function derived earlier from the Gaussian error distribution. In turn, the maximum absolute error (MAE) is defined as :

$$\text{MAE}(\mathbf{I}_{\text{data}}, \mathbf{I}(\mathbf{A})) = \frac{1}{N}\sum_{i=1}^{N}|I_{\text{data},i} - I_i(\mathbf{A})|. \tag{3.5}$$

Again, this expression is equivalent to NLL, but this time, it is derived from the assumption of a Laplacian distribution of discrepancies between the data and the model.

These two examples demonstrate the connection between the NLL and loss functions, showing how the latter can be derived from different statistical distributions (see Hodson (2022)). But this goes beyond just an MAE and MSE. For example, in the context of classification, the cross-entropy loss can be derived following a similar approach (see Zhu et al. (2018)). Thus, iteratively maximizing NLL in LIFT within the

maximum likelihood framework is equivalent to minimizing the weighted MSE.

3.3.1. MSE vs. MAE

As noted previously, MSE is a widely used loss function. However, Section 5.4 of the paper mentions that MAE (or L1) loss was preferred over the MSE for DIP. Understanding this choice requires discussing the characteristic features of the MAE and MSE loss functions.

MSE

As shown during the derivation of MSE from NLL, it has strong statistical justification because it directly results from assuming the Gaussian distribution of the error. Meanwhile, this assumption is valid in many cases due to the central limit theorem.

MSE is characterized by its quadratic nature, which heavily penalizes large error values. MSE is also continuous and easily differentiable, which eases the computation of gradients.

However, the quadratic nature of MSE makes it overly sensitive to outliers. Consequently, when estimating modal coefficients from the focal-plane PSF, the intensity values in the core of the PSF will have a much greater influence on the MSE than values from regions further from the core due to their higher brightness (assuming a high-SNR scenario).

Meanwhile, the core of the PSF is predominantly influenced by the lowest-order modes. In contrast, information related to higher-order modes is concentrated outside the core, as higher-order modes correspond to higher spatial frequencies of the pupil plane WF. Consequently, the information distribution in (AO-corrected) PSF is strongly non-uniform. Thus, the high values towards the core dominate the error function.

Thus, preserving the information in the speckles outside the bright PSF core is crucial to extracting the maximum number of modes from the PSF morphology. Hence, it is necessary to assign noticeable weight to the pixels in these peripheral regions. This can be accomplished by using pixel weighting maps, such as $\mathbf{R}_\mathrm{n}$ in LIFT, or by simply selecting a loss function that inherently prioritizes large error values less than MSE.

MAE

In turn, Equation 3.4 shows that MAE weights all errors equally by computing the absolute difference. Consequently, MAE imposes a milder penalty on outliers (such as bright PSF regions) than MSE and shows less bias towards them. This can be advantageous when a more balanced consideration of all errors is desired[3]. This balanced approach is ideal for retrieving modal coefficients due to the uneven energy distribution in the pixels of PSF. This also explains the observed higher estimation accuracy when using MAE in the context of DIP.

3. Worth noting that, for example, Huber loss function (see Huber (1964)) balances MAE and MSE by treating small error values quadratically and large values linearly.

3.4. Non-linear optimization

In the above discussion, significant emphasis was placed on the convexity analysis of the optimization criterion landscape and selecting the appropriate loss function. However, little attention was given to the non-linear optimization process itself, which has been mentioned numerous times with the assumption that its behavior is well understood.

First, it is useful to briefly outline the difference between samplers and non-linear optimizers. As shown in Figures 3.1 and 3.2, extensive sampling of the criterion landscape can be used to identify saddle points and find the global maximum or minimum (if it exists). The approach behind samplers such as `dynesty` or Monte-Carlo Markov-Chain (MCMC) (Foreman-Mackey et al. (2013)) involves computing numerous estimates of the likelihood function at various optimally chosen locations. However, while these methods can provide a comprehensive overview of the behavior of the criterion, they are computationally expensive and time-consuming.

In contrast, the goal of a non-linear optimizer is to find the global minimum of a loss function as quickly as possible, with the fewest function evaluations and minimal computational resources utilized. However, this approach may result in finding a local minimum instead of a global one. This is because, while navigating the loss function landscape, the non-linear optimizer experiences only a limited representation of the function in the vicinity of the current estimate.

Numerous non-linear optimization algorithms are available. Therefore, the following discussion will focus only on several types of non-linear optimizers directly relevant to this project.

3.4.1. First-order methods : Gradient descent (GD)

Gradient descent is one of the simplest and most widely used optimization algorithms. It is an iterative method employed to minimize a function by iteratively "moving" a current parameter estimate in the direction of the steepest descent, corresponding to the negative gradient of the function. The algorithm can be summarized as follows.

— GD begins with an initial approximation for the estimated parameters.

— Each subsequent estimate $\hat{\mathbf{A}}^{(k+1)}$ is then computed based on the gradient of the loss function $\mathscr{L}(\cdot)$ computed at the current parameter estimate $\hat{\mathbf{A}}^{(k)}$ at iteration k. Then, the update rule for GD is calculated as follows :

$$\hat{\mathbf{A}}^{(k+1)} = \hat{\mathbf{A}}^{(k)} - \alpha\, \mathbf{J}(\hat{\mathbf{A}}^{(k)}), \quad \mathscr{L}(\hat{\mathbf{A}}^{(k)}) = \mathscr{L}\left(\mathbf{I}_{\text{data}}, \mathbf{I}(\hat{\mathbf{A}}^{(k)})\right) = \mathscr{L}^{(k)}$$

Here, $\hat{\mathbf{A}} \in \mathbb{R}^M$ is the vector of estimated parameters (modal coefficients in our case). Note that in this case, $\mathscr{L} : \mathbb{R}^M \to \mathbb{R}$ is a scalar-valued function. This process is repeated iteratively until convergence. In this context, the Jacobian vector

$\mathbf{J}(\hat{\mathbf{A}}^{(k)})$ is computed as :

$$\mathbf{J}(\hat{\mathbf{A}}^{(k)}) = \nabla \mathscr{L}(\hat{\mathbf{A}}^{(k)}) = \left(\frac{\partial \mathscr{L}^{(k)}}{\partial \hat{a}_1^{(k)}}, \frac{\partial \mathscr{L}^{(k)}}{\partial \hat{a}_2^{(k)}}, \ldots, \frac{\partial \mathscr{L}^{(k)}}{\partial \hat{a}_M^{(k)}} \right).$$

Gradient descent operates on the principle of updating parameters in the direction opposite to the gradient, which signifies the steepest ascent of the loss function. By following the negative of this gradient, GD attempts to find a minimum. However, it can easily converge to a local rather than a global minimum in non-convex scenarios. The choice of the learning rate (step size) α is thus crucial : if it is too large, GD may overshoot the minimum, and if it is too small, convergence becomes slow, and the algorithm is more prone to be trapped in local minima. Consequently, more advanced techniques, such as the Adam algorithm, have been introduced to address these limitations (see Kingma & Ba (2014)).

To improve upon GD, more local information about the objective function can be obtained at each iteration, leading to more informed parameter update steps. A possible solution is to consider the second-order behavior of the objective function.

3.4.2. Second-order methods : Newton's method

Second-order optimization methods address some limitations of first-order methods by incorporating not only the first-order derivatives but also the second-order derivatives, represented with the Hessian matrix $\mathbf{Q}$. The Hessian matrix provides a quadratic approximation of the function in the vicinity of a point, defined as :

$$\mathbf{Q} = \nabla^2 \mathscr{L}(\mathbf{A}) = \begin{pmatrix} \frac{\partial^2 \mathscr{L}}{\partial a_1^2} & \frac{\partial^2 \mathscr{L}}{\partial a_1 \partial a_2} & \cdots & \frac{\partial^2 \mathscr{L}}{\partial a_1 \partial a_M} \\ \frac{\partial^2 \mathscr{L}}{\partial a_2 \partial a_1} & \frac{\partial^2 \mathscr{L}}{\partial a_2^2} & \cdots & \frac{\partial^2 \mathscr{L}}{\partial a_2 \partial a_M} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \mathscr{L}}{\partial a_M \partial a_1} & \frac{\partial^2 \mathscr{L}}{\partial a_M \partial a_2} & \cdots & \frac{\partial^2 \mathscr{L}}{\partial a_M^2}. \end{pmatrix} \tag{3.6}$$

In this equation, the "$\cdot^{(k)}$" and "$\hat{\cdot}$" notations were omitted for brevity. A classic example of a second-order optimization method is Newton's method (see Björck (1996, Chapter 9.2)), where the update rule can be written as :

$$\hat{\mathbf{A}}^{(k+1)} = \hat{\mathbf{A}}^{(k)} - \mathbf{Q}^{-1}(\hat{\mathbf{A}}^{(k)}) \nabla \mathscr{L}(\hat{\mathbf{A}}^{(k)})$$

Integrating the Hessian into optimization facilitates a more accurate representation of the function's local behavior compared to relying solely on first-order (gradient) information. The second derivative offers insights into the function's local curvature,

allowing optimization algorithms to make more informed and precise step updates [4]. Consequently, the gradient step parameter may become redundant here, as the Hessian determines both the direction and the step size for updates based on the function's curvature, preventing overshooting and thus improving convergence rates [5]. Furthermore, the off-diagonal elements of the Hessian matrix contain information about parameter coupling, indicating how changes in one parameter depend on changes in another. This makes second-order methods particularly useful for non-linear problems that involve parameter coupling.

However, the primary practical drawback of second-order methods is that using the pure Hessian matrix leads to high computational and memory demands. Computing the Hessian matrix at every estimation point is computationally expensive, and storing the Hessian requires significantly more memory than storing just the gradient vector. This makes such methods unsuitable for large-scale problems, such as training neural networks, where the number of optimizable parameters can be hundreds of billions, necessitating the computation of a Hessian matrix with the number of elements that is the square of the number of optimized parameters.

3.4.3. Quasi-Newton Methods

Quasi-Newton methods were proposed to address the computational and memory challenges of pure Newton's methods. These methods typically use gradients and estimates from previous iterations to construct an approximation to the Hessian matrix (see Nocedal & Wright (2006d, Chapter 6) and Björck (1996, Chapter 9.3.3)). Within this scope, some relevant optimization techniques can be considered.

3.4.3.1. Gauss-Newton (GN) Method

One such method is the Gauss-Newton method, which is primarily used for non-linear least squares problems (see Björck (1996, Chapter 9.3)). It approximates the Hessian matrix as a product of the Jacobians :

$$\mathbf{Q} \approx \mathbf{J}(\hat{\mathbf{A}}^{(k)})^{\top}\mathbf{J}(\hat{\mathbf{A}}^{(k)})$$

Note, that unlike previously defined Jacobian vector $\mathbf{J}$, in these derivations, $\mathbf{J}$ means *Jacobian matrix*, which is defined as :

4. Assuming the function is twice continuously differentiable.

5. Many practical implementations of (quasi) Newtonian methods require a line search or trust-region step, especially if the Hessian is ill-conditioned or not strictly positive definite. For these applications, the step size parameter is still required (see Nocedal & Wright (2006b, Chapter 3)).

$$\mathbf{J}(\hat{\mathbf{A}}^{(k)}) = \frac{\partial \mathbf{r}(\hat{\mathbf{A}}^{(k)})}{\partial \hat{\mathbf{A}}^{(k)}} = \begin{pmatrix} \frac{\partial r_1}{\partial a_1} & \frac{\partial r_1}{\partial a_2} & \cdots & \frac{\partial r_1}{\partial a_M} \\ \frac{\partial r_2}{\partial a_1} & \frac{\partial r_2}{\partial a_2} & \cdots & \frac{\partial r_2}{\partial a_M} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial r_N}{\partial a_1} & \frac{\partial r_N}{\partial a_2} & \cdots & \frac{\partial r_N}{\partial a_M} \end{pmatrix}$$

Here, the "$\cdot^{(k)}$" and "$\hat{\cdot}$" notations were again omitted inside the matrix definition, $\mathbf{r}(\hat{\mathbf{A}}^{(k)})$ is a vector of residuals [6] defined as

$$\mathbf{r}(\hat{\mathbf{A}}^{(k)}) = \left(r_1(\hat{\mathbf{A}}^{(k)}), r_2(\hat{\mathbf{A}}^{(k)}), \ldots, r_N(\hat{\mathbf{A}}^{(k)})\right)^\top, \quad r_i(\hat{\mathbf{A}}^{(k)}) = I_{\text{data},i} - I_i(\hat{\mathbf{A}}^{(k)})$$

Then, the update rule is defined as follows :

$$\hat{\mathbf{A}}^{(k+1)} = \hat{\mathbf{A}}^{(k)} - [\mathbf{J}(\hat{\mathbf{A}}^{(k)})^\top \mathbf{J}(\hat{\mathbf{A}}^{(k)})]^{-1} \mathbf{J}(\hat{\mathbf{A}}^{(k)})^\top \mathbf{r}(\hat{\mathbf{A}}^{(k)}) \tag{3.7}$$

Since the Gauss-Newton method is a least-squares estimator, the MSE loss is inherently embedded in it. The method aims to minimize the sum of squared residuals defined as :

$$S(\hat{\mathbf{A}}) = \frac{1}{2} \sum_{i=1}^{N} [r_i(\hat{\mathbf{A}})]^2$$

Inspecting Eq.(3.7), one can immediately see the parallels with the update rules for LIFT, defined earlier in Eqs.(4) and (6) in the paper. It essentially indicates that the iterative likelihood maximization method within LIFT effectively corresponds to the Gauss-Newton optimization framework, thereby supporting the previous observation regarding this resemblance mentioned in the article. However, the notation has been adjusted here. Instead of the interaction matrix **H**, the Jacobian **J** is used. In essence, both **H** and **J** have the same meaning.

3.4.3.2. BFGS (Broyden-Fletcher-Goldfarb-Shanno) Algorithm

The BFGS algorithm is one of the most popular quasi-Newton methods. Compared to GN, the Hessian approximation in BFGS is more sophisticated and precise than that in the GN method. However, the update rule for BFGS involves complex mathematical formulations, which are much more elaborate than in the case of GN, which can be found in Yuan (1991) and Nocedal & Wright (2006a, Chapter 7.2). Conceptually, the approximation of Hessian is updated as follows :

— Use the gradients at the current and previous steps to update the Hessian approximation.

6. Not to confuse with AO residuals.

— Ensure that the Hessian approximation remains symmetric and positive definite.

Moreover, the overall mathematical formulation of the BFGS algorithm is more general than that of GN, making its application not limited to non-linear least squares problems. Consequently, BFGS can be combined with an arbitrary loss function. As mentioned in Section 5.4, DIP employed a combination of BFGS (specifically L-BFGS) and the MAE loss function.

3.4.3.3. L-BFGS (limited memory BFGS)

L-BFGS is a variant of the BFGS algorithm designed for large-scale problems. Instead of storing the full Hessian approximation, L-BFGS (Liu & Nocedal (1989)) retains only a few vectors that implicitly represent the Hessian. This reduces memory requirements and computational cost, making it suitable for high-dimensional optimization problems.

Given the points discussed in this section, the benefit of using L-BFGS over GN (LIFT) for coefficient estimation, as demonstrated in Section 5.4 of the paper, can be explained. The advantage of this method over LIFT lies in the fact that L-BFGS and BFGS provide a more accurate Hessian approximation. Meanwhile, the information about coupling between modes is stored in off-diagonal elements of the Hessian since these elements show how the change of one parameter depends on the change of another one (see Eq.(3.6)). This allows for a more precise treatment of the modal coupling, leading to better convergence. In addition, as mentioned previously, L-BFGS can be combined with MAE loss, which helps extract information from PSF morphology more uniformly (if the SNR is sufficiently high).

The discussion on optimization, modal coupling, and criterion convexity is helpful in the context of LIFT. However, this discussion is much more general and can be applied to various problems which involve (non-linear) optimization. For example, the subsequent discussions in Sections 4.7 and 4.8, dedicated to PSF prediction and fitting, will also rely heavily on the principles outlined here.

3.5. From focal-plane WFSing to PSF prediction

In conclusion to this chapter, it is important to say that the information that can be extracted from PSFs extends beyond pupil phase retrieval, as demonstrated in the current chapter. By utilizing PSFs recorded in the focal plane of the scientific detector, it is also possible to extract valuable information on the behavior and performance of the AO system. However, this will require diverging from the PSF modeling principles established earlier. A more complex PSF model is needed, which relies on an extensive theoretical framework to describe various aspects of AO performance and associated errors. A shift in the type of data used is also required for such applications. Instead of short-exposure monochromatic PSFs used in this chapter, the long-exposure polychromatic PSFs recorded directly by the science instruments must be employed.

This transition will be addressed in the next chapter. Similarly to this chapter dedicated to LIFT, it utilizes on-sky focal-plane PSFs as the primary tool for analysis. However, objectives, data sources, models, and analytical approaches differ significantly. The next chapter will also devote much more time to discussing the benefits of precise PSF modeling for scientific applications.

The next part of this thesis introduces a novel data-driven method for PSF prediction that utilizes the Fourier-based modeling principles introduced in Rigaut et al. (1998). At the end of the chapter, the application of this technique to on-sky data from the MUSE-NFM (Section 4.8.5.3) and SPHERE IRDIS (Section 4.8.4) instruments will be demonstrated.

With this said, we can now proceed to the next chapter in this manuscript.

4. PSF reconstruction and prediction

4.1. Motivation for PSF-R

In the previous chapter, I demonstrated how information extracted from focal-plane PSFs can significantly enhance our understanding of system behavior, facilitating the detection of potential error sources. However, this chapter shifts the focus to a different aspect : how a comprehensive understanding of PSF can improve scientific data analysis.

As will be discussed, the modeling and retrieval of focal-plane PSFs for scientific applications involve specific requirements that differ from those addressed in Chapter 3. For example, scientific observations typically involve long exposures, often lasting minutes or even hours, and are conducted across multiple wavelengths. This makes the methods described in Chapter 3 less applicable, thereby necessitating alternative modeling approaches.

To better define the requirements for PSF models, it is crucial to consider examples of scientific observations that could benefit from a better understanding of PSF. This section will explore these aspects in greater depth, focusing on the following scientific applications. This section will also address the top-level requirements for PSF knowledge, which are largely dictated by the specific scientific application.

AO-assisted Galactic Center observations.

One of the most frequently observed targets using AO is the center of the Milky Way. Extensive research in this field ultimately resulted in the award of the 2020 Nobel Prize in Physics [1]. These observations primarily focus on the central SMBH and its interactions with the surrounding environment, with a particular emphasis on studying the stellar dynamics near Sagittarius A* (Lu et al. (2008)).

First, analysis of stellar trajectories has provided evidence for the existence of a compact gravitating object at the center of the Milky Way (Genzel et al. (2010)). Second, examining these trajectories offers an opportunity to test general relativity (GRAVITY Collaboration et al. (2018)). Any observed deviations from the theory could pave the way for new physics. Third, stellar dynamics can be used to measure the mass of the central SMBH (Ghez et al. (2008)), thereby contributing to the development of more accurate models of galactic formation, given the crucial role that the SMBH plays in

1. https://www.nobelprize.org/prizes/physics/2020/press-release/

this process. Of course, potential scientific applications extend beyond these three examples.

In this context, precise photometric and astrometric measurements are essential for studying stellar kinematics. The accuracy of such measurements can be improved through the use of PSF models for deconvolution or PSF-fitting. However, as demonstrated by Neichel et al. (2014), even small errors in the determination of PSF, on the order of a few percent, can introduce astrometric biases larger than the signal itself, underscoring the importance of accurate PSF knowledge for such applications.

Observations of intermediate-mass black holes.

In this scientific case, the kinematics of stars is studied within the dense cores of globular clusters. Precise measurements of stellar trajectories in these regions can provide evidence for the presence of intermediate-mass black holes (IMBHs). Understanding IMBHs is essential for advancing our knowledge of galaxy evolution and, importantly, for bridging the gap between stellar-mass black holes and the SMBHs found in galactic centers, thereby addressing fundamental questions about their formation.

For this purpose, accurate measurements of stellar radial velocities near the central regions of globular clusters are required, with errors not exceeding 3 km/s, along with precise astrometry and photometry.

However, globular clusters are typically crowded environments, making the isolation of individual point sources challenging. However, it is crucial for obtaining accurate photometric, astrometric, and spectroscopic data. This can be effectively achieved through non-blind deconvolution driven by a precise PSF model.

Achieving radial velocity measurements with errors below 2 km/s is only possible when the PSF used for deconvolution has a FWHM error of less than 2% (Kamann et al. (2013)), which is a level of precision that current-generation PSF models have yet to reach.

Study of distant galaxies.

The study of extragalactic targets frequently involves resolving their morphological and spectral features to investigate matter dynamics within these galaxies. Such studies are crucial not only for understanding galactic evolution and star formation but also for revealing the nature of dark matter through the measurement of galaxy rotation curves (Persic et al. (1996)).

However, imperfect knowledge of the PSF can complicate these measurements. For instance, PSF elongation can be misinterpreted as the inclination of a galaxy. Given the coupling between rotational velocity and inclination, it is essential to decouple the PSF morphology from the actual geometry of the galaxy to improve the accuracy of rotational velocity measurements. This requires knowing the PSF FWHM with an error margin of less than 20%.

A similar challenge arises when integral field spectroscopy (IFS) is used to study metallicity gradients within galaxies. Inaccurate PSF knowledge can cause smearing, leading to erroneous metallicity estimates. The ability to predict the PSF across different wavelengths would be invaluable for disentangling PSF effects from the real signal (Epinat et al. (2012)).

Furthermore, when angular resolution permits the resolution of morphological features within a galaxy, it becomes possible to measure the sizes and intrinsic velocity dispersions of clumps, giant star-forming regions, and other structures within galaxies. In these cases, the study is often limited by the contrast of these features, which requires deconvolution to increase it.

It is also important to note that the variability of PSF poses a significant limitation for such observations. Due to the low brightness of extragalactic targets, these observations typically require multiple exposures recorded over several nights, leading to variations in atmospheric conditions and PSF across the different observations of the same target.

Observation of the bodies in the Solar System

AO-assisted observations are also highly valuable for studying various bodies within the Solar System, including planets, dwarf planets, and smaller objects such as asteroids. These observations are critical for advancing our understanding of planetary formation, including that of Earth, and for assessing the potential for life in different environments, such as Mars, Europa, and Enceladus.

AO-assisted ground-based observations of objects in the Solar System can complement space missions (Fétick et al. (2019a)). This will become particularly relevant with the advent of extremely large telescopes like the ELT, which promise to deliver exceptional angular resolution.

In this context, deconvolution techniques can enhance the contrast of observations and better highlight surface features. However, a significant challenge arises from the usual lack of a point source within the field for extracting the PSF during this type of observation. Acquiring star images before or after the primary observation to estimate the PSF can introduce substantial errors due to differences in sky patches and observation timestamps. Meanwhile, for optimal deconvolution results, the error in PSF knowledge must be kept below 5% (Lau et al. (2023)).

To summarize this discussion, several key points relevant to this research project can be highlighted.

— Observations typically involve long exposures, necessitating that the PSFs used for data analysis must also represent long-exposure PSFs [2];

— Accurate knowledge of the FWHM of the PSF is essential, which requires a precise determination of both the peak of the PSF and its elongations;

2. If the SNR is high enough, the sequence of "lucky shots" can be used instead of averaged exposures (see Law et al. (2006); Fried (1978)).

— The same is valid for deconvolution, where the entire PSF morphology must be precisely known, including the PSF's peak and also its wings;

— For some applications, the PSF model should generalize the PSF morphology over a wide range of wavelengths and over the scientific field of view;

— When point sources are available in the field, they can be fitted with a PSF model to extract the PSF parameters. However, in many applications, point sources are absent in the field. This necessitates that the PSF morphology be retrieved jointly with the object or predicted without direct fitting to provide prior for deconvolution;

The last point is particularly important. Thus, the objective appears twofold : first, to develop a PSF model that can accurately fit PSFs when point sources are present in the field, and second, to ensure that the model can accurately infer PSF morphology based on some auxiliary input data rather than relying solely on PSF images within the field.

4.2. Approaches to PSF modeling

Formulating these points naturally leads to the question of which PSF model to select. Several methods are currently available for modeling focal-plane PSFs. A comprehensive overview of these methods is provided in Beltramo-Martin et al. (2020b) and Figure 4.1.

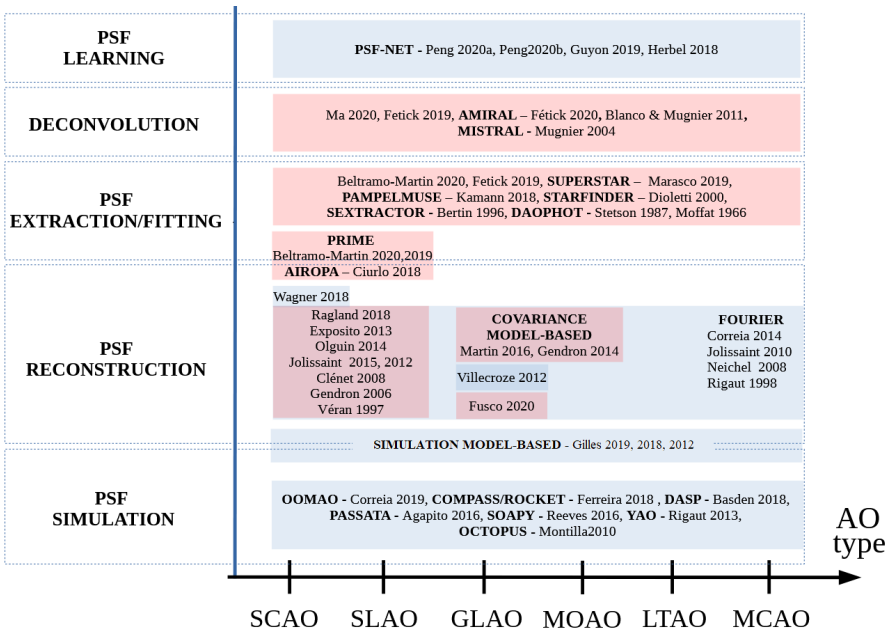


FIGURE 4.1. – Different approaches to extract and simulate PSFs. The figure is taken from Beltramo-Martin et al. (2020b).

In summary, the following types of models can be distinguished :

— Methods that directly simulate PSFs in the focal plane, such as simple extraction from astronomical data or fitting a Moffat/Gaussian function to on-sky sources. These methods focus solely on reproducing PSF morphology without considering the underlying physics or AO behavior;

— End-to-end (E2E) methods that physically simulate the complete light propagation from the sky to the detector at each step of the AO loop, generating sequences of short-exposure focal-plane PSFs. However, running such simulations [3] for minute- or hour-long exposures can be extremely time-consuming even with modern hardware;

— Methods that combine physical simulations with real data (such as AO telemetry) to infer PSF morphology;

PSF reconstruction (PSF-R) falls under the last category. The primary goal of any PSF-R method is to model the morphology of focal-plane PSFs based on AO telemetry data by partially recreating the behavior of the AO system. The following section will delve into a detailed discussion of this type of model.

4.2.1. Classical PSF-R

In 1997, Véran et al. (1997) introduced a PSF Reconstruction (PSF-R) method that estimates atmospheric / AO-corrected PSF using AO telemetry data, such as WFS measurements and deformable mirror shapes.

The core concept of this method involves calculating the covariance matrix of the residual phase based on a long sequence of recorded DM commands and WFS measurements. This covariance matrix is then used to derive the structure function, which in turn is used to calculate the OTF and then the long-exposure PSF.

4.2.2. Fourier-based modeling

Shortly after Véran et al. (1997), Rigaut et al. (1998) proposed a markedly different PSF-R method. Instead of computing the PSF directly from DM commands, Rigaut et al. suggested decomposing the AO system into its individual components and modeling the behavior of each component statistically. This allows modeling the spatial power spectral density of the residual phase in the pupil plane, which is then used to calculate SF, OTF, and, ultimately, infinite-exposure PSF applying a mathematical framework similar to that described earlier in Section 2.2.8. This technique is commonly referred to as *Fourier-based PSF modeling*.

It is crucial to note that this approach differs significantly from end-to-end (E2E) simulations. As insightfully observed by Jolissaint (2010), the "synthetic [Fourier-based] approach models the behavior of the system, while the end-to-end approach

3. Depending on their complexity.

models the structure of the system". Unlike E2E methods, which iteratively accumulate short-exposure PSFs, the Fourier-based model enables a one-shot inference of the infinite-exposure PSF morphology by leveraging the analytical models.

4.2.3. Comparing both approaches

The PSF-R principles introduced by Véran et al. (1997) have led to the development of numerous techniques applied across various systems and even tested in on-sky conditions. For example, applying such a PSF-R method to PUEO (the CFHT AO system) demonstrated that the PSF could be estimated with an accuracy better than 5% for its FWHM. PSF-R is also successfully used as part of the AIROPA package (Witzel et al. (2016); Terry et al. (2023)). Another application of PSF-R for NIRC2 at the Keck Observatory was demonstrated in Ragland et al. (2016). This type of PSF-R will also be adopted for MICADO, the upcoming ELT instrument (Simioni et al. (2020)). Consequently, this approach has become the most widely adopted method for PSF-R within the astronomical community.

However, despite continuous advances over the past 27 years, the practical application of these techniques in scientific data processing remains limited. This indicates the inherent complexity of the entire process, which can be attributed to several factors.

— As outlined earlier in Section 2.5, not all errors factor that affect imaging can be sensed by the WFS of the AO system. In this case, the information about them will not be reflected in the AO telemetry and thus cannot be modeled by the PSF-R algorithm.

— Generalizing the PSF-R algorithm to multi-LGS systems remains limited, as the mathematical framework has only been fully developed for SCAO systems. Therefore, existing PSF-R tools remain narrowly oriented towards specific SCAO systems.

In this context, Fourier-based models offer several advantages.

— The mathematical framework underlying these models is well-developed for nearly all types of AO systems, including multi-GS tomographic systems.

— The modular design of these models allows the flexible incorporation of various components, allowing the modeling of a wide range of error contributors, both static and dynamic.

— Another critical advantage of Fourier-based PSF models is their lower data requirements compared to classical PSF-R methods. As mentioned above, Véran's approach relies on DM commands and WFS slopes as input. For exposures lasting several minutes or hours, handling and storing the vast arrays of noisy, poorly compressible data recorded at kHz rates, with thousands of elements per loop iteration, becomes highly problematic. In this manuscript, the combination of WFS measurements and DM commands recorded at loop rates will be referred to as *full telemetry*. Recording such extensive data for every observation each

night would require substantial data storage capacity. With the advent of the ELT, the volume of data required to store full telemetry will increase significantly due to higher-resolution DMs and WFSs. For example, storing just 2 minutes of M4 DM commands recorded at a 1 kHz rate in half-float precision (16 bits) would generate more than 1 GB of uncompressed data. Although this amount of data might seem modest compared to facilities such as the LHC or ALMA, it could still present storage and transfer challenges for the ESO infrastructure.

The analytical approach underlying Fourier-based models offers a clear advantage in this context. As mentioned above, these models rely on a statistical representation of the residual phase (spatial PSD), which can be defined using only a few key input parameters. For example, only the variables r_0 and L_0 are required to describe the uncorrected atmospheric phase. Similarly, to account for the statistical contribution of WFS noise to the residual phase, knowing the average number of photons per sub-aperture and a set of fixed parameters related to WFS geometry is sufficient. In turn. these parameters can be extracted from the full telemetry. For example, atmospheric parameters can be estimated from DM commands and WFS slopes by approximating the open-loop residual phase. In this manuscript, such compact representations of full telemetry, reduced to a few parameters, will be referred to as *reduced telemetry*.

Thus, Fourier-based models, driven by reduced telemetry, are significantly more efficient and far less data-intensive than classical PSF-R models. The morphology of a PSF can be fully described with the model that inputs just a few parameters, as will be demonstrated in Section 4.4.2. This computational efficiency allows the generation of dozens or even hundreds of PSFs per second on modern (GPU-enabled) desktop-level hardware.

— Another key advantage of the compact input space in Fourier-based models is the ability to efficiently explore it, which makes these models *calibratable*. In the scope, calibration refers to the process of adjusting the model parameters to represent the experimental data better. It can be implemented by providing the PSF model with externally measured data or by exploring the parameter space of the model and identifying adjustments that produce the least discrepancy with the experiment. In other words, the Fourier model parameters can be feasibly fine-tuned through non-linear optimization to fit on-sky data.

When discussing what needs to be calibrated in the context of PSF-R, it is crucial to address quasi-static NCPAs. Classic PSF-R can efficiently handle this. However, calibrating dynamic AO residuals-related errors poses challenges for traditional PSF-R approaches, as it would require optimizing AO telemetry to match sky data. This does not make sense from the computational and physical points of view. Although additional optimizable "error-absorbing" components could be introduced in the PSF-R model in this case, they would still need to rely on some compact statistical representation to remain calibratable.

Although Fourier models offer many advantages, they also come with certain drawbacks. Some of these arise from the mathematical formalism and assumptions under-

lying these models, as discussed below. However, in general, the accuracy of Fourier-based models is inherently limited by the representativeness of their analytical framework. This means that the model must make assumptions about the nature of all potential error sources and, as a result, cannot account for unknown error contributors.

However, given all the listed benefits of the Fourier-based approach, this manuscript exclusively focuses on exploring this family of methods. Hence, it is crucial to dive into the theoretical foundations of Fourier-based models. This will provide the necessary background for future discussions on the limitations of these models and strategies to improve the accuracy of PSF fitting and prediction, which will be addressed in Sections 4.6, 4.7 and 4.8.

4.3. Theory behind Fourier-based PSF models

As mentioned above, the foundation of Fourier PSF modeling lies in its statistical approach. The turbulent component of the residual phase is represented by spatial PSD. However, before computing this PSD, it is necessary to separate the turbulent component of the pupil-plane wavefront from the static component :

$$\varphi(\mathbf{x}, t) = \overline{\varphi}(\mathbf{x}) + \delta\varphi(\mathbf{x}, t),$$

where $\mathbf{x}$ is a spatial coordinate in the pupil plane, t is time, $\varphi(\mathbf{x}, t)$ is a residual phase in the pupil plane, $\overline{\varphi}(\mathbf{x})$ is a constant part and $\delta\varphi(\mathbf{x}, t)$ is a turbulent zero-mean part. In the context of PSF modeling, this separation is performed most effectively in the OTF space. The OTF of the system ($\mathrm{OTF}_{\mathrm{sys}}(\boldsymbol{\nu})$) can then be computed as

$$\mathrm{OTF}_{\mathrm{sys}}(\boldsymbol{\nu}) = \frac{1}{S_p} \iint_{\mathbb{R}^2} \left\langle \exp\left\{ i\left[\delta\varphi(\mathbf{x}, t) - \delta\varphi(\mathbf{x}+\boldsymbol{\rho}, t)\right]\right\}\right\rangle_t \times \exp\left\{ i\left[\overline{\varphi}(r) - \overline{\varphi}(\mathbf{x}+\boldsymbol{\rho})\right]\right\} P(\mathbf{x}) P(\mathbf{x}+\boldsymbol{\rho})\, d^2\mathbf{x}, \quad (4.1)$$

where $\boldsymbol{\nu} = (u, v)$ is the angular frequency vector in the focal plane, associated with the spatial shift $\boldsymbol{\rho} = \lambda\boldsymbol{\nu}$ in the pupil plane, λ is the optical wavelength, S_p is the area of the pupil, $P(\mathbf{x})$ is the transmission of the pupil, $\langle\cdot\rangle_t$ is the temporal averaging operation. Assuming that the phase is spatially stationary and obeys Gaussian statistics (see Roddier (1981)) gives :

$$\left\langle e^{i\left[\varphi(\mathbf{x}, t) - \varphi(\mathbf{x}+\boldsymbol{\rho}, t)\right]}\right\rangle_t = e^{-\frac{1}{2}\left\langle\left[\varphi(\mathbf{x}, t) - \varphi(\mathbf{x}+\boldsymbol{\rho}, t)\right]^2\right\rangle_t}.$$

As noted in Véran et al. (1997), "this approximation is equivalent to assuming that the dispersion along the variable $\mathbf{x}$ is small enough that the exponential of the mean may be approximated by the mean of the exponential." Further assuming that the time-variable component has Gaussian statistics (see Jolissaint et al. (2006), Eq.(3))

allows for simplifying Eq. 4.1 :

$$\mathrm{OTF}_{\mathrm{sys}}(\boldsymbol{\nu}) \approx \exp\left[-D_{\varphi}(\rho)/2\right] \times \frac{1}{S_p} \iint_{\mathbb{R}^2} \exp\left\{i\left[\overline{\varphi}(\mathbf{x}) - \overline{\varphi}(\mathbf{x}+\boldsymbol{\rho})\right]\right\} P(\mathbf{x}) P(\mathbf{x}+\boldsymbol{\rho})\, d^2\mathbf{x},$$

where $D_{\varphi}(\boldsymbol{\rho})$ is the residuals phase structure function. Finally, the expression can be written as

$$\mathrm{OTF}_{\mathrm{sys}}(\boldsymbol{\nu}) = \mathrm{OTF}_{\mathrm{turb}}(\boldsymbol{\nu})\ \mathrm{OTF}_{\mathrm{static}}(\boldsymbol{\nu}), \tag{4.2}$$

where $\mathrm{OTF}_{\mathrm{turb}}(\boldsymbol{\nu}) = \exp\left[-D_{\varphi}(\rho)/2\right]$ is the OTF of the turbulent component of the residuals and $\mathrm{OTF}_{\mathrm{static}}$ is the static OTF component, respectively.

The static component, $\mathrm{OTF}_{\mathrm{static}}$, is determined by the static wavefront aberrations and the Airy pattern, which is governed by the pupil mask $P(\mathbf{x})$ (and the apodizer if it is present). In contrast, the turbulent component, $\mathrm{OTF}_{\mathrm{turb}}$, can be evaluated from the spatial PSD using the expressions introduced in Section 2.2.8. But before doing that, $D_{\varphi}(\boldsymbol{\rho})$ and $\mathrm{OTF}_{\mathrm{static}}(\boldsymbol{\nu})$ must be computed.

Static OTF

When discussing static aberrations, the static phase can be expressed in the same manner as outlined in Eq. 2 in the embedded paper from Chapter 3, meaning that static aberrations can be represented as a weighted sum of modes [4]. This formulation will be further exploited in Section 4.7.1 to model the effects related to LWE. Consequently, $\mathrm{OTF}_{\mathrm{static}}$ can be expressed using the following equation :

$$\mathrm{OTF}_{\mathrm{static}}(\boldsymbol{\nu}) = \mathscr{P}(\mathbf{x}) \circledast \mathscr{P}^{*}(-\mathbf{x}), \tag{4.3}$$

$$\mathscr{P}(\mathbf{x}) = P(\mathbf{x})\, A_p(\mathbf{x}) \exp(-i\Phi(\mathbf{x})). \tag{4.4}$$

Here, $\mathscr{P}(\mathbf{x})$ is the complex phase in the pupil plane, $\mathbf{x}$ is the pupil coordinate, $\circledast$ is the convolution operation, $\Phi(\mathbf{x})$ is the static phase map in the pupil plane, $P(\mathbf{x})$ is the binary pupil mask, $A_p(\mathbf{x})$ is the apodizer transmission map [5] (if an apodizer is present in the system; otherwise, $A_p = 1$).

Limitations

After this discussion, it is possible to highlight other limitations of the Fourier-based approach, which arise directly from the mathematical formalism described above.

The first limitation is the assumption of temporal stationarity in the residual phase, which means that any transient behavior cannot be modeled. Thus, the model's inputs must vary with time according to some predefined law to simulate temporal evolution.

4. In theory, it is possible to mathematically combine the PSF model from the paper with the Fourier-based PSF model into a unified model similar to one discussed in Mugnier et al. (2008).

5. It is important to note that when modeling the SPHERE instrument in this work, the apodizer transmission map also accounts for the real measured amplitude variations throughout the instrument's optical path.

The second limitation is that the Fourier approach assumes spatial stationarity of the corrected phase across the pupil. However, this assumption does not hold near the pupil's edges (as, for example, demonstrated in Fig.2 in Jolissaint et al. (2018)) or for low-order aberrations (see Jolissaint (2010)). These limitations effectively lead to the assumption of an infinite-sized pupil. It is worth noting that if the actuator density is high, this assumption becomes less problematic, as only the actuators near the edge of the DM break the stationarity, as discussed in Jolissaint et al. (2015). This effect is expected to be even less significant for large apertures like one of the ELT.

Therefore, when the PSF is non-stationary and varies across the field of view, modeling it with Fourier-based methods may require simulating a distinct PSF at each field position according to a model of spatial variation.

Tip-tilt jitter OTF

When discussing OTF, it is important to introduce another key contributor, $\mathrm{OTF}_{\mathrm{TT}}$, which accounts for the tip-tilt (TT) jitter. In this context, the PSF model employed in this project adheres closely to the principles outlined in Sections 2.2.1 and 2.2.2 of Beltramo-Martin et al. (2020a). The expression adopted in this work is

$$\mathrm{OTF}_{\mathrm{TT}}(\boldsymbol{\nu}) = \left(\frac{D}{\lambda} \cdot \frac{2\pi\sqrt{2\log 2}}{60^2 \cdot 180 \cdot 10^3} \right)^2 (J_x^2\, u^2 + J_y^2\, v^2),$$

$$\boldsymbol{\nu} = \mathbf{R}(J_{xy})\, \boldsymbol{\nu}_0 = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} u_0 \\ v_0 \end{pmatrix}.$$

Here, J_x and J_y are the tip and tilt jitter, respectively, in mas; the vector $\boldsymbol{\nu}_0 = (u_0,\, v_0)$ stays for the spatial frequency vector before applying the rotation transformation $\mathbf{R}$, where J_{xy} acts as a rotation angle, D is the aperture diameter, λ is the wavelength.

As previously mentioned, TT jitter refers to small, rapid oscillations of the PSF across the focal plane due to uncorrected or residual TT errors. These oscillations blur the image, effectively acting like a Gaussian convolution that reduces the sharpness of the observed PSF over long exposures.

Expressing this kernel as an OTF is more computationally convenient. For instance, if the jitter is weak and the PSF sampling is coarse, the convolution kernel can be undersampled in the pixel space, while it will be sufficiently sampled in the Fourier domain.

Decoupling TT from the rest of the modes offers several advantages. First, modeling TT with just two or three parameters provides better control over this error contributor, which is crucial given the strength of the TT modes. Second, if TT is managed by a separate wavefront sensor (WFS) in a dedicated control loop (like IRLOS in MUSE-NFM), decoupling it from the rest of the residual wavefront reflects the real system's design more accurately.

Consequently, $\mathrm{OTF}_{\mathrm{turb}}$ can be split into two contributors. Then, the Eq. 4.2 trans-

forms into :

$$\mathrm{OTF}_{\mathrm{sys}}(\boldsymbol{\nu}) = \mathrm{OTF}_{\mathrm{turb}}(\boldsymbol{\nu})\ \mathrm{OTF}_{\mathrm{static}}(\boldsymbol{\nu})\ \mathrm{OTF}_{\mathrm{TT}}(\boldsymbol{\nu})\ D_{\nu}(\boldsymbol{\nu}). \tag{4.5}$$

Here, $D_{\nu}(\boldsymbol{\nu})$ represents an additional spatial shift term, which can be useful in centering the PSF in the field to align with the source position, which is defined as :

$$D_{\nu}(\boldsymbol{\nu}) = \exp\left(-\pi i \left(u\,\delta x + v\,\delta y\right)\right),$$

where δx and δy are the PSF shifts along x' and y' axes in the focal plane, respectively, measured in pixels.

It is important to note that the TT contribution is not subtracted from the turbulent component but is instead added on top of the static and turbulent components. A potential concern here is whether a conflict might arise between $\mathrm{OTF}_{\mathrm{turb}}$ and $\mathrm{OTF}_{\mathrm{TT}}$, as the latter could duplicate the tip-tilt contribution already present in $\mathrm{OTF}_{\mathrm{turb}}$. But in practice, this kernel compensates for the underrepresented TT jitter by $\mathrm{OTF}_{\mathrm{turb}}$. It will be used later in Sections 4.7.1.1 and 4.7.2.1.

Turbulent OTF and PSD

Next, it is crucial to discuss how $\mathrm{OTF}_{\mathrm{turb}}$ can be calculated. As mentioned above, $\mathrm{OTF}_{\mathrm{turb}}$ is derived from the structure function D_{φ}, which in turn is calculated from the PSD of the residual phase in the pupil plane (will be later demonstrated in more detail in Eq.(4.8)).

Essentially, this spatial PSD (hereafter referred to simply as PSD) represents the residual phase variance across the pupil plane as a function of spatial frequency. In other words, it indicates how effectively the AO system corrects each spatial frequency.

Before calculating this PSD, certain assumptions must be made. First, the PSD can be divided into two components : the corrected part ($\|\mathbf{k}\| \leq k_c$) and the uncorrected part ($\|\mathbf{k}\| > k_c$), where k_c is the DM cutoff frequency. This distinction arises because any AO system can only correct a limited range of spatial frequencies due to the finite resolution of the DM actuators. As a result, the effects related to partial AO correction influence only the corrected frequency domain, while the open-loop atmospheric spectrum governs the uncorrected frequencies. Hence, the expression for the PSD can be written as the sum of two components :

$$W(\mathbf{k}) = W_c(\mathbf{k})\Big|_{\|\mathbf{k}\|\leq k_c} + W_{\mathrm{fiting}}(\mathbf{k})\Big|_{\|\mathbf{k}\|>k_c},$$

where $W_c(\mathbf{k})$ is the PSD at the frequencies corrected by the AO system, $W_{\mathrm{fiting}}(\mathbf{k})$ is the high-order uncorrected part, $W(\mathbf{k})$ is the total PSD.

The corrected PSD can be further broken down into monochromatic and polychromatic components because of the atmosphere's weak but non-negligible chromatic dispersion. Meanwhile, some PSD contributors can still be considered achromatic to simplify the derivations. Thus, PSD can be further divided into chromatic and

achromatic components as :

$$W(\mathbf{k}) = \Big(W_c(\mathbf{k}) + W_{c,\lambda}(\mathbf{k})\Big)\Big|_{\|\mathbf{k}\|\leq k_c} + W_{\text{fiting}}(\mathbf{k})\Big|_{\|\mathbf{k}\|>k_c},$$

where $W_{c,\lambda}(\mathbf{k})$ is the wavelength-dependent part. Now, the task of modeling the PSF narrows down to computing the PSD_{turb} component, and several methods in the literature address this, which must be discussed.

Brief overview of the literature

The residual PSD can be modeled in different ways. An approach is to assume an empirical function for the PSD shape. This is exemplified in Fétick et al. (2019b), where the PSD is modeled as a 2D Moffat function. Although this method is highly effective for fitting PSF morphology to on-sky data, it cannot be used to infer PSF morphology in the absence of point sources, as there is no explicit mapping between the Moffat function parameters and reduced AO telemetry. In contrast, this research focuses on physics-based modeling so that the real telemetry can be connected with the output PSF.

In terms of physics-based modeling of residual PSD, it is important to review Rigaut et al. (1998), which laid the foundation for this approach. Subsequent works have extended these ideas : Jolissaint et al. (2006) and Jolissaint (2010) generalized the model to 2D, while Correia & Teixeira (2014) and Correia et al. (2017) introduced the formalism necessary to extend the approach to XAO systems. Finally, Neichel (2008b) and Neichel et al. (2009) expanded the mathematical framework to tomographic systems, generalizing Fourier-based models to any AO system. The evolution of this method can be summarized as follows :

Rigaut et al. (1998) → Jolissaint et al. (2006) → Neichel et al. (2009) → Correia & Teixeira (2014) → Correia et al. (2017) → Beltramo-Martin et al. (2020a)

All of the mathematical derivations presented in this and the following sections are grounded in the foundational works cited here.

For the sake of clarity, it is important to note that the "PSF model" in the scope of this entire chapter refers exclusively to the implementation introduced and described in Section 4.4.1.3.

With this essential background established, we can now proceed to discuss the mathematics behind computing the physics-based residual phase PSD.

4.3.1. Error due to imperfect AO-correction

Modeling the residual phase PSD can be reformulated as modeling the impact of partial AO correction. The primary errors that contribute to the imperfect correction are detailed in Section 2.5.1.1. In this scope, the PSD of the residual phase can be decomposed into the sum of these errors, assuming that they are statistically independent. This transforms the problem of modeling the PSD of the residual phase into

the task of modeling each independent error contributor. Although, in practice, this assumption of independence may not fully hold, it remains a useful framework for deriving the required mathematical formalism.

Each of these error contributors stems from partial AO correction and is associated with distinct elements of the AO system's operation. To proceed with the derivation of these contributions, the effect of some AO system components in the Fourier space must be modeled.

Measurement operator

The first and most critical component to model is the WFS, as all subsequent expressions will be built upon it. For the purposes of this discussion, we will focus exclusively on the formalism related to the SH WFS, as it is the type employed by both the AO systems behind SPHERE and MUSE instruments, which are the primary focus of this research. Therefore, all the following expressions are specific to SH WFSs.

In the frequency domain, the SH WFS measurement process can be represented using operators. Essentially, the SH WFSing operator performs a spatial differentiation of the incoming wavefront, discretizes it to a finite number of sub-apertures, and averages the signal over the WFS integration time.

As mentioned previously, the cutoff frequency (k_c) is defined by the maximum spatial frequency that the DM can correct. Thus, it is reasonable to assume that the WFS can sample the wavefront at a spatial frequency that matches or slightly exceeds this cutoff frequency. For an SH WFS, this means that the number of sub-apertures should allow it to resolve wavefront distortions at the same scale as the DM can correct.

In this scope, it is important to note that this operator acts on the entire spatial frequency domain, including frequencies beyond the cutoff ($\|\mathbf{k}\| > k_c$), as all spatial frequencies are propagated to the WFS, but not all are used to extract the signal. The discretization inherent to this operator introduces aliasing, a phenomenon that will be discussed later.

This process is described in Eq. 20 of Rigaut et al. (1998), and the key concepts underpinning the derivation of this operator are further detailed in Eq. 36 and Section G of Jolissaint et al. (2006). Furthermore, a similar definition is presented in Correia & Teixeira (2014) in Eq. 10.

Note that the specific formulation of this operator depends not only on the type of WFS but also on different SH WFS geometries, as discussed in more detail in Equations 24(a,b) and Table 1 of Correia & Teixeira (2014).

To summarize this discussion, the measurement operator $\mathcal{G}(\mathbf{x})$ in the pupil plane can be defined as

$$\mathcal{G}(\mathbf{x}) = \text{III}\left(\frac{\mathbf{x}}{d}\right)\left[\Pi\left(\frac{\mathbf{x}-\frac{1}{2}}{d}\right) \circledast \Pi\left(\frac{\mathbf{x}}{\mathbf{V}T_s}\right) \circledast \nabla\right],$$

where d is the SH WFS sub-aperture diameter; then, $k_c \approx 1/(2d)$ T_s is the WFS sensor integration period (sampling time); $\mathbf{V} = (V_x, V_y)$ is the velocity vector of an atmos-

pheric layer. Note that all the derivations here and below are conducted under the assumption of stratified atmospheric layers, meaning each layer is treated independently from the others. As a result, each layer has a corresponding altitude and velocity; $\Pi(\cdot)$ is the rectangle function, $\mathrm{III}(\cdot)$ is the Dirac comb function, and ∇-operator is defined as $\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right)$.

In the Fourier space, the operator $\widetilde{\mathcal{G}}(\mathbf{k}) = \mathcal{F}\{\mathcal{G}(\mathbf{x})\}$ can be written as

$$\widetilde{\mathcal{G}}(\mathbf{k}) = \mathrm{III}(\mathbf{k}d) \circledast \left[2i\pi d\mathbf{k}\,\mathrm{sinc}(k_x d)\,\mathrm{sinc}(k_y d)\exp\left(j\pi d(k_x+k_y)\right)\mathrm{sinc}(\mathbf{k}\cdot\mathbf{V}T_s) \right],$$

where $\mathbf{k} = (k_x, k_y)$ is spatial frequency in the pupil plane, $\mathcal{F}\{\cdot\}$ is the Fourier transform. The "$\cdot$" operation specifies the dot product when applied to two vectors.

As will be shown later, these operators are defined differently in the case of tomographic reconstruction. For example, the measurement operator in this case is defined in the same fashion as in Rigaut et al. (1998) and Jolissaint et al. (2006) :

$$\widetilde{\mathcal{M}}(\mathbf{k}) = 2i\pi\mathbf{k}\,\mathrm{sinc}\,(k_x d)\,\mathrm{sinc}\left(k_y d\right)\mathrm{sinc}(\mathbf{k}\cdot\mathbf{V}T_s). \tag{4.6}$$

Later, it will be used in the matrix form, denoted as $\mathbf{M}$, following the notations from Neichel et al. (2009).

Reconstruction operator

The next step after performing the measurements is to reconstruct the WF. The purpose of the reconstructor operator, then, is to model this reconstruction process in Fourier space. The reconstructor operator acts almost as an inverse to the measurement operator, though it operates exclusively within the AO-corrected frequency range. The foundational concept of the reconstructor operator is introduced in Eq. 21 of Rigaut et al. (1998) and is further detailed in Section H of Jolissaint et al. (2006).

The reconstructor expression used in the PSF model is derived from Eq. 27 of Correia & Teixeira (2014), which differs from the definition in Eq. 39 of Jolissaint et al. (2006). Correia employs a regularized MMSE (Minimum Mean Square Error) approach, whereas Jolissant's method follows the Least-Squares Error (LSE) framework. The reconstructor expression adopted in the PSF model can be formulated as

$$\widetilde{R} = \frac{\widetilde{\mathcal{G}}^*}{|\widetilde{\mathcal{G}}|^2 + \gamma\frac{W_\eta}{W_\varphi}}, \quad |\widetilde{\mathcal{G}}|^2 = |\widetilde{\mathcal{G}}_x|^2 + |\widetilde{\mathcal{G}}_y|^2.$$

The expression for the reconstruction can be slightly simplified assuming $T_s \to 0$:

$$\widetilde{R}_x = \frac{\widetilde{S}_x^*\widetilde{A}^*}{|\widetilde{S}_x\widetilde{A}|^2 + |\widetilde{S}_y\widetilde{A}|^2 + \gamma\frac{W_\eta}{W_\varphi}}, \quad \widetilde{R}_y = \frac{\widetilde{S}_y^*\widetilde{A}^*}{|\widetilde{S}_x\widetilde{A}|^2 + |\widetilde{S}_y\widetilde{A}|^2 + \gamma\frac{W_\eta}{W_\varphi}},$$

$$\widetilde{A} = \mathrm{sinc}(k_x d)\,\mathrm{sinc}(k_y d)\exp\left(j\pi d(k_x+k_y)\right),$$

$$\widetilde{S}_x = 2i\pi k_x d, \quad \widetilde{S}_y = 2i\pi k_y d.$$

Here $\widetilde{R}(\mathbf{k}) = (\widetilde{R}_x, \widetilde{R}_y)$ – MMSE Wiener filter reconstruction operator, developed in Correia & Teixeira (2014). $W_\varphi(\mathbf{k}), W_\eta(\mathbf{k})$ are the spatial PSDs of the atmospheric turbulent phase and the WFS noise. Noise can be assumed to be white and uncorrelated and therefore constant at all frequencies ($W_\eta(\mathbf{k}) = \sigma_\eta^2$), whereas the turbulent phase PSD will be given later in Eq.(4.7) below; γ is a scalar weighting factor, W_η / W_φ is a regularization term. When the signal is strong, $W_\eta / W_\varphi \to 0$. But for frequencies with weak signal $W_\eta / W_\varphi \to \infty$. However, in all further development, $\widetilde{R}$ is simplified to an LSQ filter by assuming $\gamma = 0$; $\widetilde{A}(\mathbf{k})$, $\widetilde{S}(\mathbf{k}) = (\widetilde{S}_x, \widetilde{S}_y)$ are averaging functions on the phase and slopes, respectively (again, see Correia & Teixeira (2014)).
It should be noted that this formalism is specific to SCAO-type correction and, therefore, was only used for the SPHERE instrument in this work. The definition of a reconstruction operator also differs for tomographic AO systems. It includes modeling the tomographic reconstruction, which is defined in Eq. 13 in Neichel et al. (2009) as

$$\mathbf{W} = \mathbf{P}_{\text{opt}} \mathbf{W}_{\text{tomo}},$$

where $\mathbf{P}_{\text{opt}}$ is a projector of a tomographic solution on DMs. For example, in LTAO, $\mathbf{P}_{\text{opt}} = \mathbf{P}_\beta^L$ as elaborated in Eq.(7.18) in Neichel (2008b), $\mathbf{W}_{\text{tomo}}$ is the tomographic reconstructor. Please refer to Neichel et al. (2009) for the exact definition of both quantities. Note that although all quantities here are still formulated in spatial frequency space, the tilde notation is omitted for brevity.

In this case, the measurement operator is embedded in $\mathbf{W}_{\text{tomo}}$. For computational efficiency, it is based on the simplified model of measurements formulated earlier here in Eq. 4.6. But here, it only takes vector form, so it is denoted as $\mathbf{M}$.

After introducing the measurement and reconstruction operators, it is now possible to advance to the analysis of the various error sources. These two operators provide a foundation for modeling the behavior of an AO system in the Fourier domain. Applying these operators to the residual wavefront results in the derivation of different error sources, which will be discussed below. This mechanism is described in Eq.(11) in Jolissaint et al. (2006) and Eq.(11) in Neichel et al. (2009). However, in the following discussion, we will not dive into the detailed derivations, as they are extensively covered in the existing literature. Instead, the focus will shift to explaining the role and origins of the different error sources that contribute to the PSD of the residual phase.

Fitting error

This is the simplest error contributor to the model. As previously outlined, the frequency domain in this model can be divided into two regions : the AO-corrected frequencies and the non-corrected frequencies. The non-corrected frequencies are those too high to be sensed and compensated by the AO system and thus are driven by the atmosphere, although they may still contribute via aliasing into the corrected

frequency range. In contrast, AO-corrected frequencies are those influenced by the performance of the AO system.

The fitting error captures the behavior of the uncorrected frequencies. It is governed by the open-loop atmospheric spectrum defined by the Von Kármán model, leading to the following expression for this error term :

$$W_\varphi = \Gamma_\varphi \, r_0^{-\frac{5}{3}} \left(k_x^2 + k_y^2 + \frac{1}{L_0^2} \right)^{-\frac{11}{6}}, \quad \Gamma_\varphi = \left(\frac{24\,\Gamma\left(\frac{6}{5}\right)}{5} \right)^{\frac{5}{6}} \left(\frac{\Gamma\left(\frac{11}{6}\right)^2}{2\pi^{\frac{11}{3}}} \right), \tag{4.7}$$

$$W_{\text{fitting}}(\mathbf{k}) = \begin{cases} 0 & \text{if } \|\mathbf{k}\| < k_c, \\ \widetilde{P} W_\varphi & \text{otherwise,} \end{cases}$$

where Γ_φ is a constant in the atmospheric power spectrum definition according to the Von Kármán model, while $\Gamma(\cdot)$ is the gamma function; $\widetilde{P}(\mathbf{k})$ is the piston mode filter. It must be filtered out because the phase average over the pupil of the telescope (which is the piston mode) cannot be seen by the AO system with the differentiating WFS. The piston filter is defined as

$$\widetilde{P}(\mathbf{k}) = 1 - \left[\frac{2J_1(\pi D \|\boldsymbol{k}\|)}{\pi D \|\boldsymbol{k}\|} \right]^2.$$

It is important to note that $W_{\text{fitting}}(\mathbf{k})$ term is computed across the entire spatial frequency domain. In contrast, the subsequent error terms presented below will be computed only within the domain of AO-corrected frequencies.

Aniso-servo-lag

This error term arises due to two main factors : the delays in the AO system (servo-lag) and the angular separation between the guide stars and the science target (anisoplanatism). The term "aniso-servo-lag" reflects the correlation between these two errors, which results from the Taylor frozen-flow hypothesis (see Section 2.2.3), which is intrinsically adopted in this model. Thus, a wavefront shift can be compensated for by a spatial frozen-flow-driven wavefront shift.

Another key assumption adopted in the derivations of this error term is that the atmosphere can be modeled as independent stratified layers, allowing the aniso-servo-lag error to be computed for each layer independently.

As outlined in Rigaut et al. (1998) (Eq. 27), the approach to deriving this error term involves integrating the reconstruction operator over the duration of the AO system's time lag, which encompasses the delay between wavefront measurement and the execution of corrective commands.

Eq. 33 of Jolissaint et al. (2006) and Eq. 43 of Jolissaint (2010) present the PSD expression for this error term, with anisoplanatism included for an SCAO system. In the case of XAO systems employed, the influence of anisoplanatism diminishes, as both the science target and NGS are close to the on-axis, effectively reducing the impact of anisoplanatism close to zero. Thus, anisoplanatism is not included in the

SPHERE PSF model in Sections 4.6.1, 4.7.1, and 4.8.4. However, in this case, it becomes even more crucial to accurately model the AO system's temporal dynamics and account for transfer functions. The derivation for this model, specifically for servolag, is found in Eq. 18 of Correia et al. (2017). This expression adopted in the PSF model for the SPHERE instrument can be written as :

$$W_{\text{STA}} = \left(1 + |\widetilde{R}\widetilde{\mathcal{G}}|^2 \widetilde{H}_2 - 2\,\text{Re}\{\widetilde{R}\widetilde{\mathcal{G}}\widetilde{H}_1\}\right) W_\varphi \widetilde{P}$$

$$\widetilde{H}_1(\mathbf{k}) = \sum_{l=1}^{N_L} w[l]\, h_{\text{cor}}(-2i\pi\,\mathbf{k}\cdot\mathbf{V}[l])$$

$$\widetilde{H}_2(\mathbf{k}) = \sum_{l=1}^{N_L} w[l]\, |h_{\text{cor}}(-2i\pi\,\mathbf{k}\cdot\mathbf{V}[l])|^2$$

where $\mathbf{w} = \left(w_0, w_1, \ldots w_{N_L}\right)$ is the vector of weights of the atmospheric layers, indexed by $l = 1, \ldots, N_L$, with N_L being the number of layers. Then, $w[l]$ is a scalar; $\mathbf{V}[l] = (V_x[l], V_y[l])$ is the wind velocity vector corresponding to the l-th atmospheric layer; h_{cor} is the closed-loop transfer function between corrective and input phases, defined in Eq. 9 in Correia et al. (2017); $\widetilde{H}_1(\mathbf{k})$ and $\widetilde{H}_2(\mathbf{k})$ are auxiliary functions computed as a weighted average of h_{cor} over the atmospheric layers.
A substitution can be made :

$$G \equiv \widetilde{R}\widetilde{\mathcal{G}} = \widetilde{R}_x \widetilde{S}_x \widetilde{A} + \widetilde{R}_y \widetilde{S}_y \widetilde{A}.$$

Then, the expression for W_{STA} can be rewritten as

$$W_{\text{STA}} = \left(1 + |G|^2 \widetilde{H}_2 - 2\,\text{Re}\{G\widetilde{H}_1\}\right) W_\varphi \widetilde{P}.$$

When tomographic reconstruction is introduced, the derivation of this error term undergoes significant changes. In this context, the expression transforms into the one presented in Eq. 23 of Neichel et al. (2009), where the first term on the right-hand side is responsible for the spatio-temporal error, and the second addend is the WFS noise, which will be discussed later. The expression adopted in the PSF model is the following :

$$W_{\text{STA}} = \left(\mathbf{P}_\theta^L - \mathbf{P}_\theta^{\text{DM}}\mathbf{W}\mathbf{M}\mathbf{P}_\alpha^L\right) \mathbf{C}_\varphi \left(\mathbf{P}_\theta^L - \mathbf{P}_\theta^{\text{DM}}\mathbf{W}\mathbf{M}\mathbf{P}_\alpha^L\right)^\top \widetilde{P}$$

$$\mathbf{C}_\varphi = (\mathbf{w}\mathbf{I}_d)\, W_\varphi \widetilde{P},$$

$$\mathbf{P}_\theta^L = \left(\exp\left(2i\pi\,\mathbf{k}\cdot(h[l]\boldsymbol{\theta} - \mathbf{V}[l]\Delta t)\right)\right)_{l=1}^{l\le N_L},$$

$$\mathbf{P}_\alpha^L = \left(\exp\left(2i\pi h[l]\,(\boldsymbol{\alpha}_i\cdot\boldsymbol{k})\right)\right)_{l=i=1}^{l\le N_L,\, i\le N_{GS}},$$

$$\mathbf{M}\mathbf{P}_\alpha^L = \left(2i\pi\,\mathbf{k}\,\text{sinc}\,(k_x d)\,\text{sinc}\left(k_y d\right)\text{sinc}(T_s\,\mathbf{k}\cdot\mathbf{V})\exp\left(2i\pi h[l](\boldsymbol{\alpha}_i\cdot\mathbf{k})\right)\right)_{l=i=1}^{l\le N_L,\, i\le N_{GS}},$$

where $\Delta t = t_d$ is the loop delay time; $\mathbf{C}_\varphi$ is the matrix of Von Kármán spectrum weighted by layer weights vector $\mathbf{w}$; $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_i)$ vector of direction to GSs; $\mathbf{P}_\theta^L$ is the vector that projects the N_L turbulent layers seen in the θ direction in the pupil plane; $\boldsymbol{\theta}$ is the

direction to the target in which the solution is optimized.

It is important to note that in tomographic reconstruction, anisoplanatism does not exist in the same sense in SCAO systems. The closest equivalent in a tomographic setup is when the DM correction is optimized in a direction that deviates from the direction of the scientific target. However, in the scope of the MUSE-NFM data used in this research, the DM correction is always optimized in the direction of the scientific target, so $\boldsymbol{\theta} = \mathbf{0}$. Therefore, only the spatio-temporal component of $\mathbf{P}_\theta^L$ that contains the layers' velocities remains.

Wavefront sensing noise

This error term arises from the inherent noise in all WFS measurements, which propagates into the reconstruction process. The noise originates from two primary sources : photon noise and detector noise (see Equations 47 and 48 in Correia et al. (2017) for detailed definitions).

This creates an irreducible error term that can be derived by applying the reconstructor operator to the noise component of the measurement operator, as demonstrated in Eq. 29 in Rigaut et al. (1998).

The specific expression for this error term, which incorporates the noise transfer function, is given in Eq. 20 in Correia et al. (2017). This is the exact equation implemented in the PSF model. The model assumes that the noise variance remains uniform across all sub-apertures, implying an even flux distribution. This results in the following expressions :

$$W_{\text{noise}} = \widetilde{P} W_\eta = \langle |\widetilde{P}\,\widetilde{R}\,\widetilde{\eta}\,\widetilde{H}_n(\mathbf{k})|^2 \rangle = \widetilde{P} |\widetilde{R}|^2 \sigma_\eta^2 d^2 p,$$

$$p = 2T_s \int_0^{1/2T_s} |h_\eta(s)|^2 ds,$$

where $h_\eta(s)$ is the noise transfer function, explained in Eqs. 20-21 in Correia et al. (2017), $\sigma_\eta^2 = \sigma_{ph}^2 + \sigma_{det}^2$ is the measurement noise variance, photon noise σ_{ph}^2 and detector noise σ_{det}^2 are defined in Eqs.(47) and (48) of Correia et al. (2017).

For the tomographic case, the expression for the noise term is adapted based on Eq. 23 from Neichel et al. (2009), resulting in :

$$W_{\text{noise}} = \left(\mathbf{P}_\theta^{\text{DM}} \mathbf{W}\right) \mathbf{C}_b \left(\mathbf{P}_\theta^{\text{DM}} \mathbf{W}\right)^\top \widetilde{P},$$

where $\mathbf{P}_\theta^{\text{DM}} = \mathbf{I}_d$ for the on-axis LTAO case and $\mathbf{C}_b = \sigma_\eta^2 \mathbf{I}_d$ is the matrix that contains the noise variance.

Aliasing

Aliasing occurs when the WFS spatially undersamples the high-spatial-frequency components of wavefront distortions, which results from dividing the wavefront by a

finite number of sub-apertures. As a result, these high-frequency components "fold" back into lower frequencies during discretization. This creates an error because the control system attempts to correct for this spurious low-frequency component.

Different derivations of the aliasing error term exist. The foundational formalism is presented in Rigaut et al. (1998) (Eq. 33), and this was further developed in Jolissaint et al. (2006) (Eq. 45). Furthermore, aliasing is addressed in Correia et al. (2017) (Eq. 19).

In essence, all these derivations are fundamentally based on filtering the atmospheric spectrum using the WFS reconstructor operator, while simultaneously sampling overlapping spatial frequencies. For an SH WFS, the spacing of the overlapped instances in the frequency space is determined by the reciprocal of the lenslet diameter. These overlapped instances are then summed within the AO-corrected spatial frequency domain, with this computation being carried out independently for each atmospheric layer.

In this work the Eq. 19 from Correia et al. (2017) is adopted, which leads to the following expression :

$$W_{\text{alias}} = \widetilde{P} \sum_{\mathbf{m}\neq 0} \left|\widetilde{R}(\mathbf{k})\widetilde{H}_1(\mathbf{k})\widetilde{\mathcal{G}}(\mathbf{k}+\mathbf{m}/d)\right|^2 W_\varphi(\mathbf{k}+\mathbf{m}/d).$$

where $\mathbf{m}$ is a pair of non-zero integer numbers written down in a vector form, which can be split into $\mathbf{m} = (n, m) \in \mathbb{N}/\{0\}$. Thus, $k_m = k_x - \frac{m}{d}$, $k_n = k_y - \frac{m}{d}$. It is useful to make a few substitutions :

$$Q = 2i\pi d\left(k_m \widetilde{R}_x + k_n \widetilde{R}_y\right)\text{sinc}\left(k_m d\right)\text{sinc}\left(k_n d\right),$$

$$A_\varphi = \widetilde{H}_1 \sum_{l=1}^{N_L} w[l]\,\text{sinc}\left(k_m V_x[l]\,T_s\right)\text{sinc}\left(k_n V_y[l]\,T_s\right)\exp\left(2i\pi\, t_d\left(k_m V_x[l] + k_n V_y[l]\right)\right).$$

Then, the expression for the W_{alias} can be rewritten as

$$W_{\text{alias}} = \widetilde{P} \sum_{\substack{m=-\infty \\ m\neq 0}}^{\infty} \sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} \left|QA_\varphi\right|^2 W_\varphi(k_m, k_n)$$

However, this expression only applies to SCAO systems and does not account for tomographic reconstruction. Despite this, the PSF model employs the same aliasing computations for both SCAO and tomographic cases for simplicity. The derivation of aliasing terms for tomographic reconstructors is discussed in detail in Neichel (2008b), specifically in Section 7.6.4 and Annexe B.1.

Differential refraction

All the error contributors discussed above are treated as monochromatic, meaning each error term is computed for a single wavelength. However, in reality, the atmosphere is weakly dispersive, which introduces additional errors caused by performing scientific observations and WFS measurements at different wavelengths.

One such error term is differential refraction, which arises when the scientific target

and GS are observed at an angle away from the zenith. If the science wavelength differs from the WFSing wavelength, the atmosphere acts as a prism, inducing an additional error (illustrated in Figure 4.2). This effect is most pronounced when the target is near the horizon and diminishes to zero when observing at the zenith.

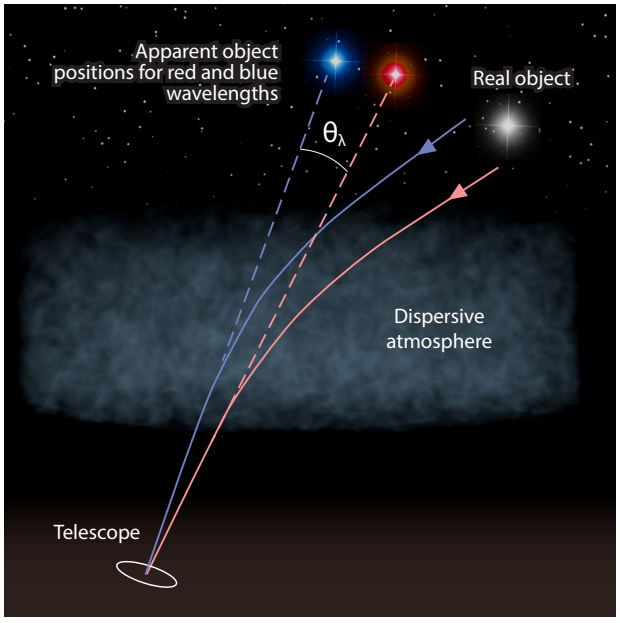


FIGURE 4.2. – Differential refraction effect. The apparent on-sky position of the object for blue and red wavelength differ by angle θ_λ due to the atmospheric dispersion. It also implies that "red" and "blue" wavefronts sample slightly different columns of atmospheric turbulence.

To illustrate this concept, consider a SCAO system where the science target simultaneously serves as the NGS. In this setup, the shorter wavelengths of the target's spectrum are used for WFSing, while the longer wavelengths are allocated for science. When observing at non-zenith angles, the light at shorter wavelengths is refracted more strongly by the atmosphere, resulting in a slight displacement of the target's image at the WFSing wavelength relative to that at the science wavelength. It also implies that the light used by the WFS samples a slightly different column of atmospheric turbulence compared to the science light, introducing an error analogous to anisoplanatism. This phenomenon is described in detail in Section 6 of Sasiela (1992).

As outlined in Jolissaint & Kendrew (2010), this error can be formalized as an aniso-servo-dispersion error, which merges the aniso-servo-lag and chromatism error terms, resulting in complex equations (7-9) in that reference. However, these equations can be drastically simplified and reduced to an additional anisoplanatism-like PSD, which depends on the difference in zenith angles between the WFSing and science wavelengths (Θ_λ). The following simplified expression is implemented in the PSF

model :

$$W_{\text{diff.ref}} = \sum_{l=1}^{N_L} 2w[l]\left(1-\cos\left(2\pi h[l]\,\|\boldsymbol{k}\|\tan(\theta_\lambda)\cos\left(\text{atan2}(k_y,k_x)-\theta_{\text{az}}\right)\right)\right) W_\varphi \widetilde{P}$$

where θ_{az} is the azimuth angle, and θ_λ is the angle between two images of the target can be computed using the Eq.(10) from Fusco et al. (2006) as

$$\theta_\lambda = (n(\lambda_{\text{src}}) - n(\lambda_{\text{GS}}))\,\tan(\theta_{\text{zen}}),$$

where $n(\lambda)$ is the index of refraction of the atmosphere defined in Eq.(8) and (9) from Fusco et al. (2006), θ_{zen} is the zenith angle, λ_{src} is the scientific wavelength, λ_{GS} is the WFSing wavelength. The atan2$(\cdot,\cdot)$ function is defined as :

$$\text{atan2}(k_y,k_x) = \begin{cases} \text{atan}\left(\frac{k_y}{k_x}\right) & \text{if } k_x > 0, \\ \text{sign}(y)\,(\pi - \text{atan}\left(\left|\frac{k_y}{k_x}\right|\right)) & \text{if } k_x < 0, \\ \text{sign}(y)\,\frac{\pi}{2} & \text{if } k_x = 0 \text{ and } k_y \neq 0, \\ 0 & \text{if } k_x = k_y = 0. \end{cases}$$

However, because of its relatively weak second-order influence, the differential refraction error term *was neglected in the PSF model used in this chapter* to simplify the computations.

Chromatism error

This error also originates from the dispersive properties of the atmosphere. Because light waves with shorter wavelengths undergo stronger refraction than their longer counterparts, they accumulate more optical path difference as they propagate through the atmosphere. As a result, the wavefront correction computed at the WFSing wavelength differs in strength from the wavefront disturbances at the science wavelength, resulting in imperfect compensation. It leads to an error, which is defined as :

$$W_{\text{chrom}} = \left(1 - \frac{n(\lambda_{\text{src}})}{n(\lambda_{\text{GS}})}\right)^2 W_\varphi \widetilde{P}.$$

The derivation of this error term is discussed briefly in Eq.(13) from Fusco et al. (2006).

Computing resulting PSD

As mentioned above, all the listed error terms together contribute to the partial AO correction error. This discussion, therefore, concludes the analysis of the various error sources initially introduced in Section 2.5.

Now, these PSD terms must be combined to compute the total PSD of the residual phase. Given the assumed statistical independence of these terms, the total W_{total} can

be calculated as the sum of all these individual PSDs :

$$W_{\text{total}}(\lambda) = \left(\frac{\lambda_{\text{atm}}}{2\pi}\right)^2 \left(W_{\text{fitting}} + W_{\text{alias}} + W_{\text{STA}} + \left(\frac{\lambda_{\text{atm}}}{\lambda_{\text{GS}}}\right)^2 W_{\text{noise}} + W_{\text{chrom}}(\lambda)\right),$$

here, $\mathbf{k}$ was dropped for brevity, $W_{\text{total}}(\mathbf{k},\lambda)$ is the spectral slice of the total PSD $W_{\text{total}}(\lambda)$, $\lambda_{\text{atm}} = 500$ nm is the "atmospheric wavelength", which is adopted by ESO as the standard for normalizing all atmospheric-related parameters, including r_0 and L_0. Then, $(\lambda_{\text{atm}}/2\pi)^2$ is the term that recomputes PSD from rad^2 at λ_{atm} to nm^2 OPD. The $(\lambda_{\text{atm}}/\lambda_{\text{GS}})^2$ term is necessary to recompute the noise PSD from the GS wavelength to the atmospheric one. The final step involves calculating the polychromatic covariance matrix $B \rightarrow$ structure function $D_\phi \rightarrow \text{OTF}_{\text{turb}}$. This turbulent OTF is then multiplied by $\text{OTF}_{\text{static}}$ and OTF_{TT}. This is then used to compute the polychromatic infinite-exposure PSF. The whole process can be summarized with the following equations :

$$\begin{aligned}
B(\boldsymbol{\nu},\lambda) &= \mathscr{F}^{-1}\{W_{\text{total}}(\mathbf{k},\lambda)\},\\
D_\phi(\boldsymbol{\nu},\lambda) &= \left(\frac{2\pi}{\lambda}\right)^2 2\,[B(0,\lambda) - B(\boldsymbol{\nu},\lambda)],\\
\text{OTF}_{\text{turb}}(\boldsymbol{\nu},\lambda) &= \exp\left(-\frac{1}{2} D_\phi(\boldsymbol{\nu},\lambda)\right),
\end{aligned} \tag{4.8}$$

$$\text{PSF}_0(\mathbf{x}',\lambda) = \mathscr{F}^{-1}\{\text{OTF}_{\text{turb}}(\boldsymbol{\nu},\lambda)\,\text{OTF}_{\text{static}}(\boldsymbol{\nu},\lambda)\,\text{OTF}_{\text{TT}}(\boldsymbol{\nu},\lambda)\,D_\nu(\boldsymbol{\nu},\lambda)\},$$

$$\text{PSF}(\mathbf{x}',\lambda) = F_\lambda \frac{\text{PSF}_0(\mathbf{x}',\lambda)}{\int \text{PSF}_0(x,y,\lambda)\,dx\,dy} + b_\lambda, \quad \text{PSF}(\mathbf{x}') = \{\text{PSF}(\mathbf{x}',\lambda_i)\}_{i=1,\ldots,N_\lambda}.$$

Here, $\text{PSF}(\mathbf{x}')$ is the final infinite-exposure polychromatic PSF. Note that later in this chapter, a discretized version of PSF will be used, where instead of focal-plane coordinates $\mathbf{x}' = (x', y')$, the discrete pixel coordinates i, j will be used; N_λ is the number of spectral slices; F_λ is the flux multiplier and b_λ is the constant background flux. Figures 4.3 and 4.4 schematically illustrate the entire process of computing PSF starting from PSD :

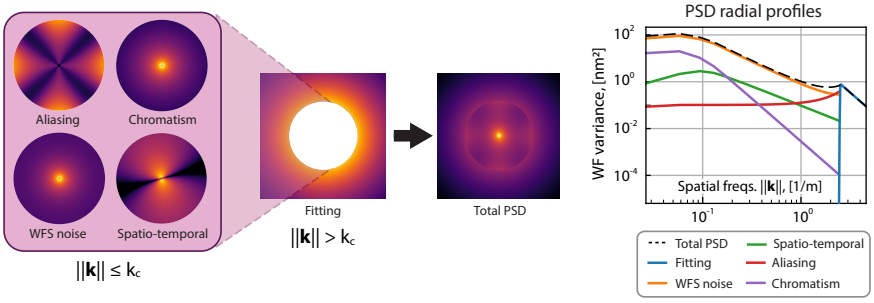


FIGURE 4.3. – All PSD contributors are summed and re-normalized to get the final turbulent-phase PSF $W_{\text{total}}(\mathbf{k})$.

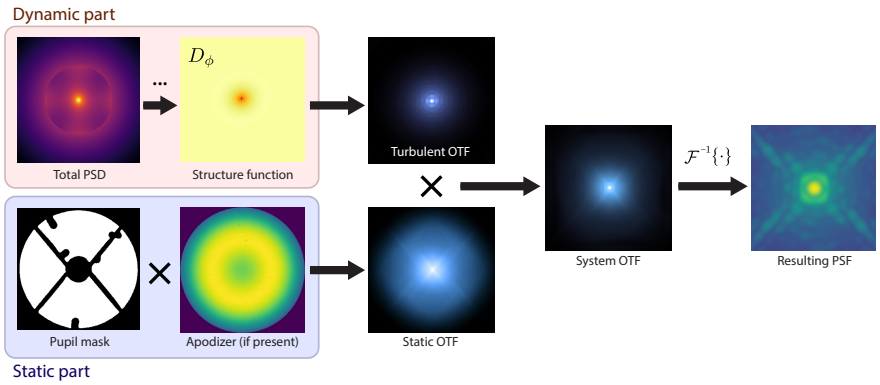


FIGURE 4.4. – Total PSD is then used to compute the final infinite-exposure PSF. In this figure and the figure above, the example of IRDIS is used. The shift term D_v was omitted here.

To summarize this section, the key conclusions are as follows :

— Fourier-based models provide a versatile, physics-driven, and fast PSF modeling framework. However, they cannot capture transient temporal effects and assume PSF invariance across the field, issues that can be mitigated by adjusting inputs based on time or field coordinates;

— The dynamic (turbulent) part of the residual wavefront is represented as a spatial PSD, which is modeled as the sum of (assumingly) statistically independent error terms;

— These error terms can be understood as frequency filters and linear operators applied to the open-loop atmospheric turbulence spectrum, W_φ;

— Error terms require different derivations for SCAO and LTAO systems;

— However, the aliasing error term is formulated the same way for both SCAO and LTAO to simplify computations;

— The differential refraction term is neglected due to its relatively minor effect, simplifying the computations further;

— Anisoplanatism is also excluded from the model, as all targets used in this study are on-axis for both the SPHERE and MUSE datasets;

This concludes the discussion of the theoretical framework used for the model adopted in this research project. The next logical step is to address how these theoretical principles are translated into practical application. Regrettably, this elegant theoretical framework must eventually face practical implementation challenges, which involve developing reliable and efficient code supported by accurate data inputs. This topic is vast and, therefore, deserves a detailed discussion, which will be provided in the next section.

4.4. PSF model implementations

4.4.1. Code implementations

4.4.1.1. Astro-TipTop

When discussing existing implementations, it is essential to discuss the Astro-TipTop [6] project (later referred to as TipTop or TIPTOP for brevity). This open-source, Python-based library specializes in physics-based Fourier PSF modeling and is currently the most comprehensive and versatile tool available. It will be adopted by ESO as an exposure time calculator and AO performance predictor for several instruments, including ERIS, MUSE-NFM, and upcoming MORFEO and HARMONI.

Technically, TipTop is divided into two components : the low-order (LO) and high-order (HO) parts. As detailed in Neichel et al. (2020), the LO component generates jitter maps across the field based on the selected NGS constellation, wind shake/vibrations, and LO WFS measurements, which makes it particularly useful for wide-field simulations and asterism selection. On the other hand, the HO component models the spatial frequencies higher than tip and tilt. The final PSF is obtained by convolving the HO PSF with the jitter kernels at different field positions.

However, as discussed in Sections 4.5.5.2 and 4.5.5.1, in this project, only on-axis targets are considered. As a result, the LO component becomes irrelevant, leaving only the HO component applicable. Thus, it is accurate to say that the mathematical framework outlined in the previous section pertains solely to the HO part.

4.4.1.2. P3

In TipTop, the HO part is managed by the P3 library, a physics-based PSF model that implements the full mathematical formalism outlined in Section 4.3. Like TipTop, P3 is an open-source Python-based library that can also operate independently as a standalone tool. In particular, it is important to highlight the contribution of Olivier Beltramo-Martin to the development of P3, who was instrumental in consolidating the theoretical formulations into a coherent and practical software framework, which laid a background for the work conducted in my thesis.

4.4.1.3. TipTorch

TipTorch [7] is the work-in-progress PSF model developed specifically for this project. It was designed to address certain limitations of the original P3 code in terms of performance and flexibility. Thus, it shares the same inputs and underlying physics with P3, while its implementation significantly diverges from P3. It is worth noting that all results presented in this manuscript were generated using the TipTorch model exclusively, meaning that throughout this chapter, *any reference to the "PSF model" refers specifically to TipTorch.*

6. https://github.com/astro-tiptop/TIPTOP
7. https://github.com/EjjeSynho/TipTorch

TipTorch is implemented using the PyTorch library. This makes it feasible to implement GPU-accelerated computation in a highly parallel batched manner, enabling simulations of multiple PSFs across multiple wavelengths in a single run. This makes it highly efficient for simulating multiple sources at the same time. Moreover, using the PyTorch framework enables seamless switching between GPU and CPU backends.

Additionally, TipTorch's use of PyTorch enables the utilization of the automatic differentiation (AD) functionality. The benefits of using AD are discussed earlier in Section 5.2 of the embedded paper in the context of DIP. In summary, AD allows for fast and precise computation of gradients and Hessians for (differentiable) functions of almost any complexity (see Baydin et al. (2017)). This approach eliminates the need to express analytical gradients and avoids inaccuracies and instabilities associated with numerical differentiation. AD proved helpful in training artificial neural networks (NNs), but it can also be used to solve various optimization problems.

Finally, using PyTorch enables the seamless integration of TipTorch with NNs, which will be extremely useful later in Section 4.8.

In summary, TipTorch is a compact, efficient, and versatile library for PSF simulations that implements the Fourier-based approach. It can potentially be integrated into the Astro-TipTop project in the future, either complementing or replacing P3 for HO simulations. All subsequent discussions in this chapter will focus exclusively on TipTorch.

All links for the libraries mentioned in this section can be found in Appendix B.1.

4.4.2. Initialization and usage

To provide context for the following discussions, it is important to briefly outline the model initialization process.

Both the P3 and TipTorch models are first initialized using configuration files containing parameters that define the telescope, AO system, science detectors, WFSs, and more. Appendix B.2 provides a complete list of these parameters.

Some parameters, such as sampling rate, pupil mask, and telescope diameter, remain constant as they are system-specific. However, others, such as the Fried parameter (r_0), WFS photon count, and the C_n^2 profiles, must be updated dynamically. These variable parameters are later referred to as *model inputs* and denoted as $\mathscr{X}$. The exact contents of $\mathscr{X}$ will be elaborated later since it is not strictly fixed and varies depending on the application.

The important message here is that once the model is initialized using configuration files for a specific instrument, it can be controlled by a compact set of inputs $\mathscr{X}$ ($\propto 10^1$ elements) to model different observational conditions.

4.5. Data

As discussed earlier, the PSF model requires physics-based inputs to calculate PSF estimates. Since this project focuses on accurately modeling the PSF for potential use in postprocessing real observations, these inputs must also be derived from associated on-sky data. Thus, the model must rely on so-called "reduced telemetry" as inputs. Meanwhile, when evaluating the accuracy of the PSFs generated by the model, the simulated PSFs must be compared with real on-sky PSFs. This necessitates associating reduced on-sky telemetry with the corresponding on-sky PSFs recorded simultaneously with the reduced telemetry. This, in turn, defines the type of dataset required for further analysis : on-sky long-exposure PSFs paired with their corresponding reduced telemetry. In this context, it is crucial to clarify what "reduced telemetry" means, where it comes from, and how it is processed to serve as input for the TipTorch model. The following section will discuss these aspects.

4.5.1. Data types

When discussing data generated by ESO facilities, it is essential to distinguish between *scientific* and *engineering* data. Given that this project focuses on instruments like SPHERE and MUSE, the discussion will be limited to the relevant ESO data infrastructure.

In summary, scientific data encompass everything available through the ESO Science Archive, which external users can access. This includes raw and processed science exposures, standard calibration data, meteorological data from site monitors, and much more. Meanwhile, engineering data consist primarily of telemetry generated by instruments and AO facilities. This type of data is typically inaccessible to external users and is not always available in archives, as it is often stored temporarily locally on internal servers. Engineering data is mainly used for diagnostics, debugging, and other engineering purposes.

The datasets utilized in this project involve both engineering and scientific data. Therefore, it is crucial to explain the specific meaning of each in the context of this project and how they can be combined to improve PSF modeling.

4.5.1.1. Telemetry (full and reduced)

As mentioned above, engineering data refers primarily to various forms of telemetry. Telemetry encompasses all data generated by the observatory's instruments and systems, excluding the actual science exposures. Numerous systems within the telescope, such as detectors, cryostats, translation stages, and accelerometers, continuously log their measurements. Among these, the AO facilities are particularly prominent. The telemetry that AO systems generate includes commands sent to DMs, WFS measurements, WFS images, auxiliary data from WFS detectors, and more. Therefore, telemetry can be broadly understood beyond just AO-related data. However, within the scope of this thesis, specifically, AO telemetry is essential, given its direct connection with

image formation and the substantial volume of data it generates. Consequently, the subsequent discussion will primarily focus on AO telemetry and its derivatives.

Additionally, the concepts of raw and reduced telemetry, introduced earlier, require further clarification. Telemetry can undergo post-processing, and depending on the extent of this processing, it can be categorized into several distinct levels, which can be defined as follows :

Level 0 : Raw telemetry.

Raw telemetry refers to unprocessed engineering data that is time-stamped and stored immediately after being recorded. This type of telemetry provides a direct snapshot of the system's behavior and is crucial for system diagnostics and debugging. Examples of raw telemetry include :

— WFS measurements,
— DM commands,
— unprocessed WFS images,
— LGS telemetry,
— raw sensor data (e.g., from accelerometers or thermometers),
— instrument and system logs,
— and more.

As discussed in Section 4.2.3, storing and systematically transferring level-0 AO telemetry requires substantial storage capacity and high bandwidth, which are not always feasible. For example, while level-0 telemetry data are systematically archived at the Keck Observatory, this is not true for ESO's VLT. In this case, such data are temporarily stored for immediate processing to produce higher-level telemetry and are subsequently discarded. In VLT, the raw telemetry is usually archived only during specific system maintenance or diagnostics, making it inaccessible in a systematic way.

Level 1 : Processed telemetry.

It refers to data that have undergone some postprocessing. This processing may include filtering, statistical reduction, compression, averaging of WFS images, or organizing the data into more convenient and systematic formats, as suggested in Gomes et al. (2024).

This telemetry level also includes entries that capture key statistics from the raw telemetry, such as cross-correlation maps generated by atmospheric profilers or covariance matrices of WFS measurements and DM commands, commonly used in classical PSF-R algorithms.

Despite these processing steps, level-1 telemetry still requires too much storage and transfer bandwidth to be archived for every observation, which presents the same challenges as discussed above. Consequently, like level-0 telemetry, level-1 data is not systematically stored and remains mostly inaccessible to external users.

Level 2 : Integrated Parameters.

This telemetry has undergone further statistical reduction, resulting in data that is compressed into just a few integral parameters that cannot be further reduced. For example :

— Fitting the von Kármán model to the pseudo-open-loop phase allows for the retrieval of r_0 and L_0 parameters;

— Atmospheric turbulence profiles can be distilled to fractional C_n^2 values, layer altitudes, and wind speeds per layer;

— WFS images can be averaged into a few values representing the average flux per sub-aperture;

— Similarly, WFS slopes can be reduced to the average RMS of the slopes across all sub-apertures;

This list can be further extended with a few more examples. Reduced telemetry is the most manageable, as it requires minimal storage space. It is systematically stored and readily accessible, even to external users. Furthermore, reduced telemetry entries, such as Strehl, seeing, wind speed, etc., are often included in the headers of scientific exposures.

Level 3 : Post-Processed and Fused Integrated Parameters.

This telemetry layer is particularly relevant in the context of this manuscript. Although level-2 telemetry can already be used by the P3 and TipTorch models, in practice, further post-processing is often necessary. This includes extracting data from archives, filtering and normalizing it, associating it with corresponding science data by timestamps, and packaging it into ready-to-use datasets. Level-3 telemetry also involves fusing data from multiple sources.

Additionally, unlike the previous levels, which are stored and processed on the ESO side, level-3 telemetry involves custom processing routines developed specifically for this project. This level represents the final stage of the telemetry processing, designed to be directly input into the PSF model. Detailed steps for retrieving and processing level-3 telemetry will be outlined in Sections 4.5.5.1 and 4.5.5.2.

It is important to note that every level of telemetry involves more processing. Meanwhile, the procedures required for this kind of processing are not always transparent for the end user, as reduction down to level-2 is usually done intrinsically by an instrument or AO system. Therefore, it is not always clear how the initial values were collected and aggregated and which raw telemetry data was exactly used as a source. Thus, having the same data stored and available at all telemetry layers would be very useful, at least for specific selected observations. It would allow cross-checking and debugging of the reduction procedures and comparing custom telemetry reduction routines with ones from ESO.

In summary, "full telemetry" typically refers to data at levels 0 and 1, while "reduced telemetry" pertains to levels 2 and 3. This relationship is visually outlined in Figure 4.5.

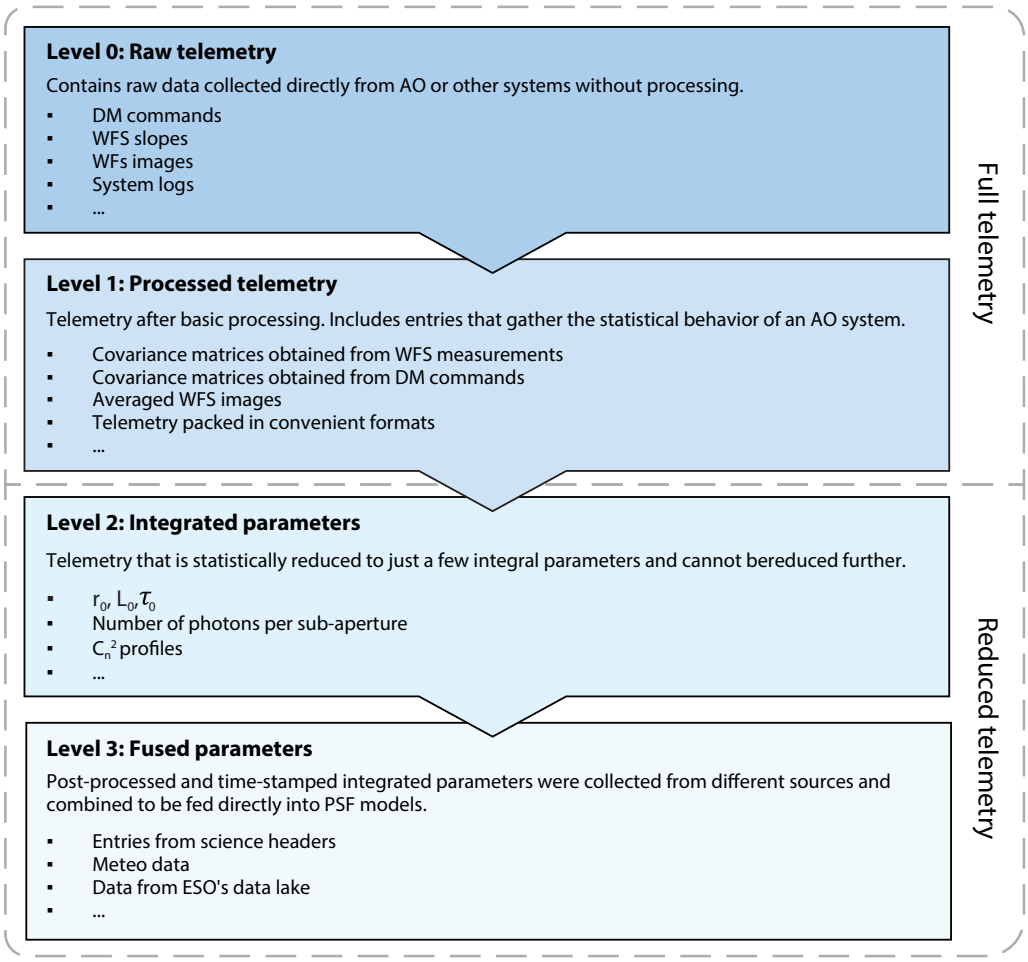


FIGURE 4.5. – The diagram of various telemetry levels. Each successive level applies additional post-processing steps to the initial engineering data generated by different telescope subsystems, including the AO system.

However, this figure requires an explanation for the "meteorological data", "science headers", and "data lake" points referenced in the level-3 telemetry box in this diagram. Understanding these points is also crucial for understanding the inputs of the PSF model later.

4.5.2. Science data

Since the instruments discussed in this work are ESO instruments, the relevant data can be accessed through the ESO Science Archive. This archive is an extensive repository of astronomical data collected by various ESO telescopes and instruments. It stores a wide range of data, including raw and processed images, spectra, catalogs, and survey data. While it is impossible to cover all the archive contents in this discussion, it is important to highlight some elements relevant to this project.

Raw Data : The ESO archive stores unprocessed observational data obtained by ESO telescopes and instruments. This raw data represents the raw exposure files recorded for various scientific programs, which can later be processed and analyzed using ESO instrument pipelines.

Processed Data : In addition to raw data, the archive includes processed data that have already undergone reduction, facilitating easier analysis by researchers without the need to run data reduction pipelines themselves.

Calibrations : The archive also contains calibration data essential for correcting and processing raw observational data. This includes dark frames, flat fields, wavelength calibrations, and atmospheric conditions. For some instruments, calibration involves recording short exposures (up to several minutes long) of standard spectrophotometric targets (usually bright stars) to assess the system's performance. These calibration observations are conducted systematically, resulting in datasets accumulated over time. For example, for instruments like IRDIS and MUSE-NFM, such calibrations often involve recording bright on-axis stars to check AO system performance and the instrument's spectral transmission. These calibration datasets are crucial to this project as they constitute the PSF datasets utilized in the remainder of this chapter. This will be elaborated on further in Sections 4.5.5.1 and 4.5.5.2.

Observational and calibration data from the ESO Science Archive are typically distributed in the Flexible Image Transport System (FITS) file format. Each FITS file contains header data units (storing instrument settings, observation details, and metadata) and pixel blocks. The format specification can be found in Pence et al. (2010).

Of particular importance is that the headers of science or calibration files produced by ESO instruments often store entries related to level-2 telemetry. As demonstrated below, these values extracted from the headers of calibration PSFs are critical for composing the reduced telemetry datasets used in this work.

4.5.3. Meteorological data

Another crucial type of data retrievable from the ESO archives is meteorological data, continuously measured by site monitors. ESO site monitors provide real-time atmospheric data at its observatories, including seeing, humidity, temperature, wind speed, and other environmental conditions. This information is vital for optimizing observation schedules and ensuring that observations are conducted under suitable conditions. Additionally, this data plays a critical role in maintaining the safety and performance of the instruments. For instance, the observatory domes are closed if wind speeds or humidity levels exceed safe thresholds.

At the Paranal Observatory, the Astronomical Site Monitor (ASM) [8] includes several key instruments :

8. https://www.eso.org/sci/facilities/paranal/astroclimate/asm-instruments.html

— **Differential Image Motion Monitor (DIMM) :** As part of the MASS-DIMM instrument, this channel operates across a wide range of atmospheric conditions, often overlapping with periods of scientific observation. DIMM provides reference seeing data, among other parameters.

— **Multi-Aperture Scintillation Sensor (MASS) :** MASS offers vertical turbulence profiles ($C_n^2(h)$) at six altitude layers : 0.5, 1, 2, 4, 8, and 16 km above the telescope. The combination of MASS and DIMM yields both accurate seeing measurements and low-resolution turbulence profiles, with the MASS-DIMM combination generally delivering more precise seeing data than DIMM alone.

— **Low Humidity And Temperature PROfiling (LHATPRO) :** This microwave radiometer measures low levels of precipitable water vapor at Paranal. It includes a humidity profiler, a temperature profiler, and an infrared camera for cloud detection.

— **METEO :** Comprising several meteorological stations from Vaisala, this system monitors wind, rain, temperature, humidity, and pressure. Weather stations are installed at different heights (2 m, 10 m, 20 m, and 30 m) on a weather tower. Data from these stations is crucial for computing atmospheric differential refraction, while temperature, wind, and humidity data are vital for operational safety.

— **SLOpe Detection And Ranging system (SLODAR) :** SLODAR measures the vertical turbulence profile in the surface layer up to 100 meters above the site. This data is instrumental in modeling and understanding the imaging performance of the Very Large Telescope (VLT), both with and without AO.

In the context of this work, the relevant entries produced by MASS-DIMM, METEO, and SLODAR are included in the reduced telemetry datasets. The atmospheric parameters collected by these instruments are stored in the ESO Ambient Conditions Database [9], with some measurements also embedded in the headers of scientific or calibration files.

4.5.4. Data lake

The data lake represents a vast repository of engineering data. At level-0, it primarily stores timestamped logs generated by various observatory systems. Additionally, it offers an interface for extracting specific entries in level-2 format, along with their associated timestamps, using elastic search [10]. However, this database and its interface are better suited for storing and querying small data packets, such as sequences of individual entries, rather than handling full telemetry datasets.

The entries logged in the data lake typically include hardware-oriented engineering data, such as pressure inside cryostats, flux on WFS detectors, RMS of WFS slopes, detector temperatures, VLTI delay line positioning, and more. These data are invaluable for monitoring the health and performance of observatory systems. Consequently,

9. http://archive.eso.org/wdb/help/eso/ambient_paranal.html
10. https://www.elastic.co/elasticsearch

the data lake can be used to retrieve specific data entries that are not available in the headers.

To summarize this discussion, the following data sources are utilized in this project for both MUSE and SPHERE instruments :

— **ESO Science Archive :** Used to retrieve on-sky calibration PSFs and some associated reduced telemetry entries from their headers.

— **Ambient Conditions Database :** Accessed to obtain atmospheric parameters such as wind direction/speed, seeing, and τ_0, generated by various ASM instruments.

— **ESO Data Lake :** Employed to fetch missing values essential for initializing the PSF model that may not be readily available in the headers.

The next section will explore how these data sources were processed to create the reduced telemetry datasets.

4.5.5. Collecting and processing the data

Developing data collection and post-processing routines was a significant aspect of this PhD project due to its data-intensive nature. Although the same sources are used to retrieve data for both SPHERE and MUSE, the procedures and post-processing routines differ between the two instruments. Therefore, each will be discussed separately.

4.5.5.1. SPHERE

Creating the dataset of PSFs.

The dataset collection process begins with downloading the available PSFs (see Appendix C.2.2 for more details). As noted previously, these files contain observations of bright standard targets, ensuring a high SNR for most of these observations. Additionally, all recorded targets are positioned close to the on-axis, eliminating any anisoplanatism as discussed in Section 4.3.1. Furthermore, the coronagraph is removed during these calibration observations, allowing the capture of "pure" IRDIS PSFs. This is one of the key reasons why these PSFs were used in this project. Another key reason for selecting SPHERE is the abundance of available PSFs for this instrument and the convenience of accessing them through the SPHERE data center, which efficiently manages and stores these datasets.

The retrieved PSFs are stored in raw ADU units within $1024 \times 1024 \times N_{DITS}$ cubes. Other data blocks within these files are not relevant to this analysis. The exposures are averaged, and the resulting image is centered relative to the PSF peak using the center of mass and then cropped to 256×256 pixels. It is important to note that each calibration observation provides both left and right PSFs, which, as discussed in Section 4.5.5.1, represent images of the same target recorded with two different spectral or polarization filters, meaning that each sample in the dataset includes two

PSFs. Each PSF is then background-subtracted using the `photutils` [11] library routine. This processing is done separately for the "left" and "right" IRDIS PSFs. The resulting PSFs are then normalized so that the sum of their pixels over the FOV per λ equals one :

$$\mathbf{I}_{\text{data},\lambda} = \frac{\mathbf{I}'_{\text{data},\lambda} - \mathbf{b}'_{\lambda}}{\sum_{i=1}^{W}\sum_{j=1}^{H} I'_{\text{data},ij\lambda} - b'_{ij\lambda}} \rightarrow \sum_{i=1}^{W}\sum_{j=1}^{H}\sum_{\lambda} I_{\text{data},ij\lambda} = 2, \quad \lambda = \{\lambda_L, \lambda_R\}$$

Here, W, H are the width and height of the cropped images respectively, $I'_{\text{data},ij\lambda}$ is a pixel in the non-normalized image, λ_L, λ_R are the wavelengths of the left and right IRDIS images, respectively; $b'_{ij\lambda}$ is a pixel value in the background map $\mathbf{b}'_{\lambda}$, the apostrophe notation "$\cdot'$" is used with data before post-processing.

This data retrieval and processing procedure is repeated for all PSFs available through the SPHERE client (see Figure 3), resulting in a comprehensive PSF dataset. The process of PSFs post-processing is schematically illustrated in Figure 4.6.

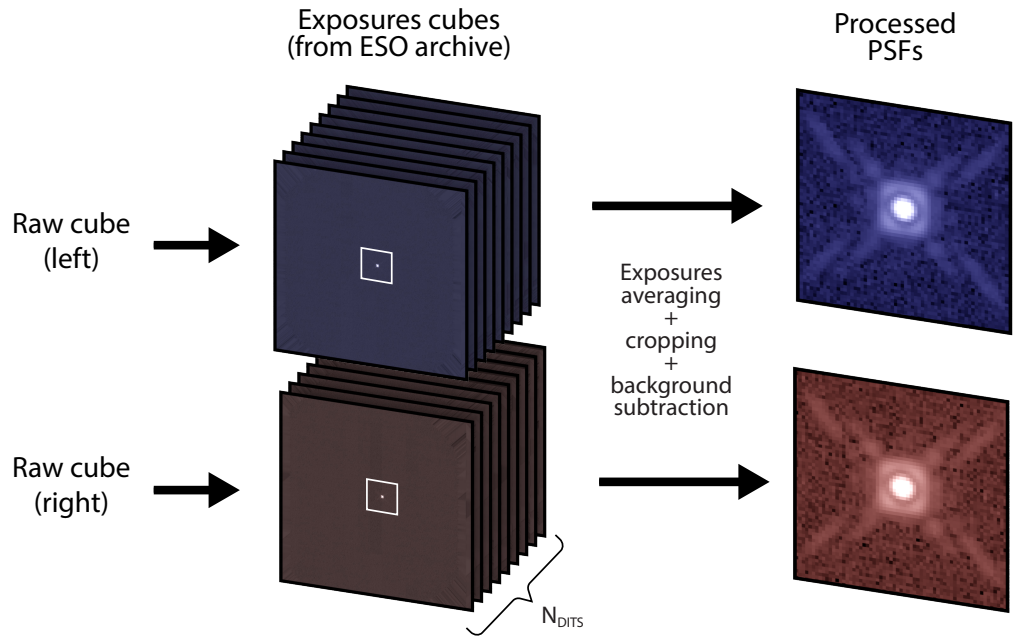


FIGURE 4.6. – Depiction of the post-processing procedure for raw IRDIS PSF cubes to extract centered and clean calibration PSFs. IRDIS splits the image into "left" and "right" channels, where each channel can have its own spectral or polarization filter. Note that these PSFs are influenced by the apodizer, whereas no coronagraph was inserted during their acquisition.

The initial size of the unfiltered dataset obtained this way is 4100 × 2 PSFs. The dataset then undergoes manual filtering to remove corrupted samples, resulting in

11. https://photutils.readthedocs.io/en/stable/

$N_{I,\mathrm{IRDIS}} = 3112$ usable PSF samples [12], each including "left" and "right" PSFs. Filtering improves the quality of the dataset, which is particularly important for later stages of the project where machine learning methods are applied.

Creating a reduced telemetry dataset.

As mentioned before, the datasets used in this project involve PSFs and associated reduced telemetry. However, compared to PSFs, the process of obtaining the reduced telemetry is more intricate. Julien Milli developed the initial procedures [13] to download and process reduced telemetry for SPHERE, which Olivier Beltramo-Martin then complemented. The routines used in this project are based on their developments, which were modified to fit the needs of this project. This involved automatizing some data fetching routines, optimizing the code, getting more entries from the archive, and filtering them more efficiently. More details about the procedure can be found in the code itself [14] [15]. In short, the procedure consists of the following steps :

1. After PSFs are downloaded, it is possible to look for their date and list the nights on which these observations were conducted.
2. All relevant integrated parameters are extracted from the headers of the FITS file with PSFs. These parameters are guaranteed to be temporally matched with the PSFs.
3. Then, the FITS files generated by SPARTA with the reduced telemetry are fetched from the ESO Science Archive. Only the ones with relevant timestamps are downloaded and unpacked. These files contain time-stamped DTTS images, estimates of the atmospheric parameters, and some integrated WFS parameters like flux per sub-aperture, loop parameters, etc.
4. Then, the Ambient Conditions Database is queried to download the atmospheric parameter estimates corresponding to the observations.

All data collected are filtered to match the time frames of the observation. It is worth noting that the collected entries are distributed very unevenly in time. Some entries can be recorded by orders of magnitude more frequently than others. Hence, if multiple data entries are found within the exposure time frame, their median value is computed and stored in the final dataset. Meanwhile, if there is no data record within the exposure time, the value is searched 120 seconds before the exposure starts and 120 seconds after its finish. The closest one of those is then stored in the final dataset [16]. If there are no close enough values found, the record in the dataset is assigned to NaN.

12. The PSF samples that were missing most of the associated reduced telemetry were also filtered out even if the PSF images themselves were good quality.
13. https://github.com/jmilou/sparta
14. https://github.com/EjjeSynho/sparta
15. https://github.com/EjjeSynho/TipTorch/data_processing
16. Perhaps, the better way to do it is to compute the average between the nearest pre- and post-observation values.

This temporal filtering is applied to reduced SPARTA telemetry, DTTS images, and ASM entries. The result of this processing is the level-3 telemetry. It is then stored in the table, where each PSF name is associated with a row filled with singular float values, each associated with a distinct reduced telemetry entry (columns).

If more than 25% of values are absent in a column, it is excluded from the dataset. The resulting dataset content and the origin of each parameter it contains are discussed in Section C.2.2 in more detail.

The following two figures can summarize this whole discussion. Figure 4.7 schematically summarizes how the employed dataset is processed and compiled on the ESO side. Meanwhile, Figure 4.8 shows how the data is retrieved and processed within this project's scope to produce usable PSF plus associated reduced telemetry datasets.

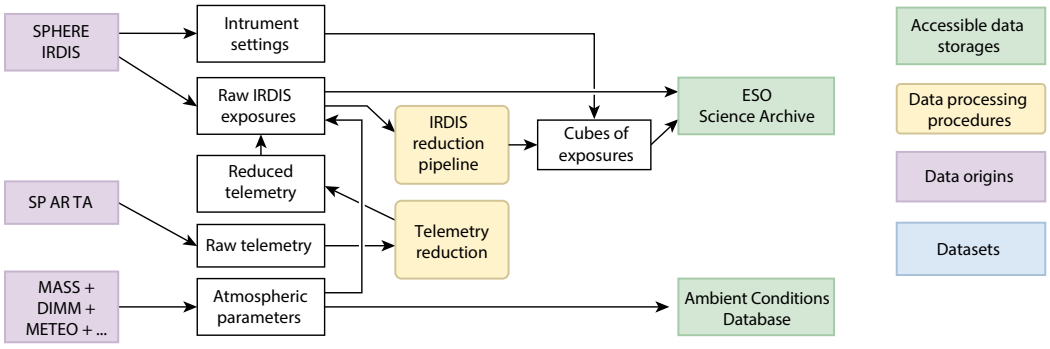


FIGURE 4.7. – Diagram illustrates the sources of IRDIS datasets and their storage location in ESO archives.

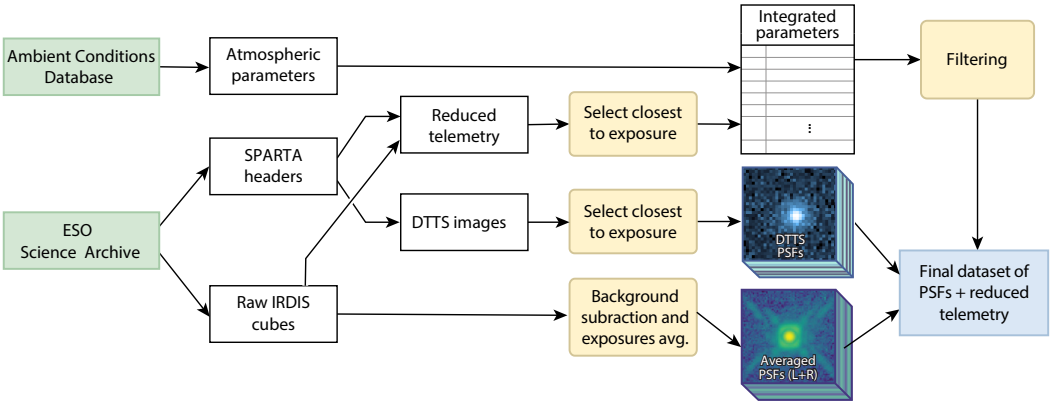


FIGURE 4.8. – Diagram illustrates the process of collecting the dataset of IRDIS PSFs with the associated reduced telemetry.

4.5.5.2. MUSE

Conceptually, the dataset retrieval process for MUSE-NFM is similar to that for IRDIS. Still, the actual data retrieval procedure differs significantly due to the difference

in the structures of the scientific data and telemetry produced by these instruments. For detailed information on how the reduction code functions, refer to the code in the TipTorch repository [17].

Creating the dataset of PSFs.

The initial step involves collecting the PSFs. For MUSE-NFM, the PSFs are also recorded during calibration routines, where bright standard targets are observed on-axis. As discussed earlier on page 2 of the embedded paper in Section 3.1, MUSE-NFM uses a separate NGS for TTF sensing. However, in calibration observations, the target itself serves as the NGS, effectively eliminating anisoplanatism for both the science target and the NGS. This setup allows high-quality, easier-to-simulate PSFs to be recorded, the quality of which is predominantly driven by the AO correction under optimal conditions.

As an integral field spectrograph, MUSE produces hyperspectral images as its science products, with dimensions $315 \times 303 \times 3681$, where the last dimension corresponds to the spectral channels. During processing, the spectral cube is centered relative to the PSF peak, cropped, and spectrally binned to dimensions $200 \times 200 \times 14$ (see Figure 4.9).

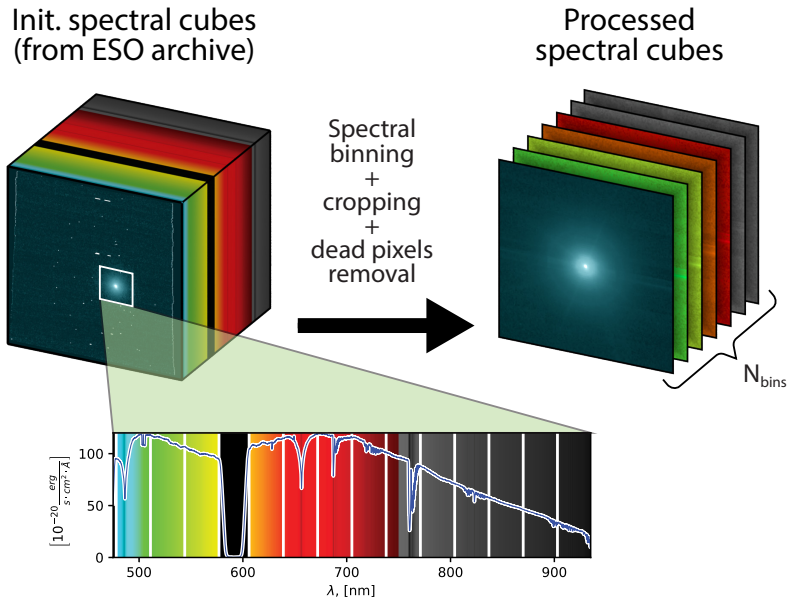


FIGURE 4.9. – Left cube illustrates raw spectral PSF data obtained during on-sky calibration routines for MUSE-NFM. In contrast, the left cube shows the final processed and spectrally binned PSF with $N_{\text{bins}} = 14$.

The wavelength ranges of $\lambda = \{450, \ldots, 478\}$ nm and $\lambda = \{577, \ldots, 606\}$ nm are excluded during the binning process. The first band is erased due to a weak signal in the

17. https://github.com/EjjeSynho/TipTorch/tree/master/data_processing

blue spectral region, while the latter is removed because the sodium filter blocks all light coming from LGSs in this spectral band.

The background is then subtracted from each spectral slice of the cube. Finally, each spectral slice is normalized so that the sum of the pixel values across the field equals one :

$$\mathbf{I}_{\text{data},\lambda} = \frac{\mathbf{I}'_{\text{data},\lambda}}{\sum_{i=1}^{W}\sum_{j=1}^{H} I'_{\text{data},ij\lambda}} \rightarrow \sum_{i=1}^{W}\sum_{j=1}^{H}\sum_{\lambda} I_{\text{data},ij\lambda} = N_{\text{bins}}.$$

In this equation, W, H are the width and height of the cropped spectral cube respectively, $I'_{\text{data},ij\lambda}$ is a pixel in the non-normalized spectral cube $\mathbf{I}'_{\text{data}}$; $N_{\text{bins}} = 14$ is the number of spectral bins in the processed spectral cube.

The processed spectral cubes contain the following wavelengths : $\lambda = \{494, 527, 560, 622, 655, 688, 721, 754, 787, 820, 853, 886, 919\}$ nm. Each wavelength in this range represents the central wavelength of its corresponding spectral bin. The images within these bins are generated by averaging all spectral slices contained within the defined boundaries of each bin. The placement of these boundaries is depicted in Figure 4.9.

Creating a reduced telemetry dataset.

Similarly to the process for IRDIS, the headers of the FITS files containing the MUSE-NFM PSF images are utilized to extract some integrated parameters. However, some parameters are lost during the MUSE data reduction pipeline, so they can only be found in the source raw files. For instance, headers of raw MUSE files contain more information on atmosphere profiles measured by AOF. Additionally, raw files include sequences of 100 IRLOS frames [18], Consequently, retrieving the initial raw MUSE files together with the reduced cubes is useful.

The rest of the level-3 telemetry is obtained in a manner similar to IRDIS, meaning that, in addition to the values from the headers, the DIMM/MASS-DIMM and METEO data are also queried from the ambient conditions database. However, in the case of MUSE, ESO's data lake is additionally employed to fetch certain parameters. For example, the LGS flux or operational status can only be obtained from the data lake. Moreover, when collecting data related to IRLOS, the data lake is used to retrieve any missing values directly from the logs.

The resulting reduced telemetry dataset contains 412 records, with only 382 PSFs being useful. More details on the dataset content can be found in Section C.3.1.

Due to the small size of the dataset, an imputation of missing values was performed, which was not required when processing the IRDIS dataset. The imputation involved training a set of random forest regressors [19], one for each column. During this process, the existing values in the column were used as the target data, while the values from all other columns served as features. The standard `sklearn` [20] procedure was used for

18. Although the DTTS and IRLOS images are part of the final datasets, they are not currently used.
19. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html
20. https://scikit-learn.org/stable/

the imputation [21] [22].

The resulting data retrieval and processing routines are depicted in the figures below. Figure 4.10 schematically illustrates how the data are stored on the ESO side, while Figure 4.11 outlines the main steps of the data post-processing performed in this project.

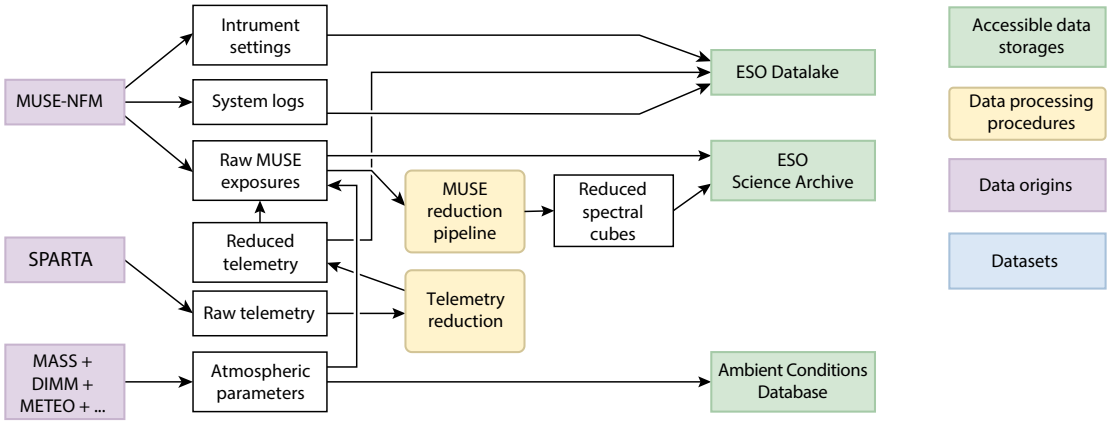


FIGURE 4.10. – Diagram illustrates the sources of MUSE-NFM datasets, as well as their storage location.

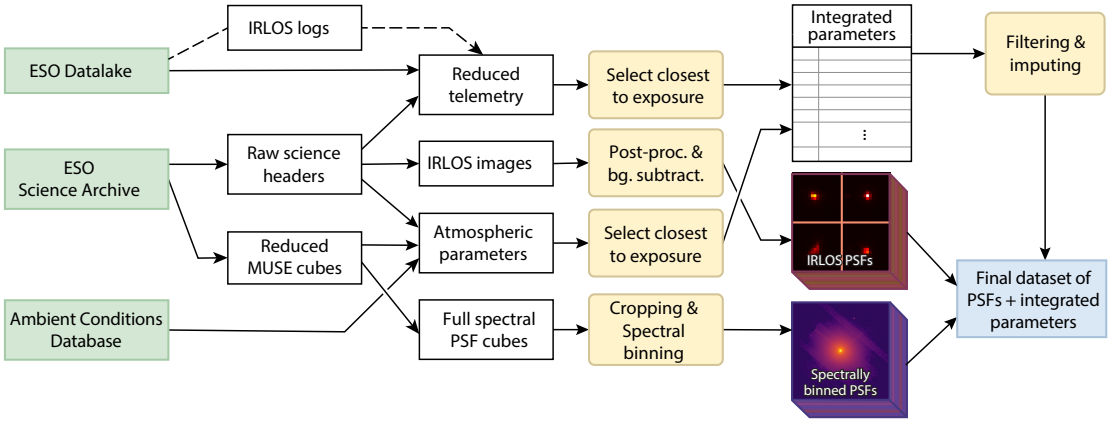


FIGURE 4.11. – Diagram illustrates the process of collecting the dataset of MUSE-NFM PSFs with the associated reduced telemetry.

21. https://scikit-learn.org/stable/modules/impute.html
22. https://scikit-learn.org/stable/modules/generated/sklearn.impute.IterativeImputer.html

4.5.6. Why not PSF-R ?

After discussing the nature and availability of the datasets used in this work, it becomes evident that full telemetry datasets are unavailable for SPHERE and MUSE instruments due to the transient nature of their storage. Consequently, if the objective is to test the methods on as many real on-sky samples as possible, Fourier-based models are preferred over classical PSF-R since they only require systematically stored reduced telemetry. From this perspective, the choice of Fourier-based models in this work is also driven by the nature of the available datasets.

With an understanding of the types of data available, it is now possible to describe how these data are utilized in the PSF model. The following section will focus on how reduced telemetry is "fed" to TipTorch.

Given the variety of retrieved telemetry data entries, one might wonder if all these telemetry entries are necessary to run the PSF model. Of course, not all of them are required. Instead, the model relies only on a small subset. This will be elaborated on in the following section. However, some seemingly redundant data entries will become useful later in Section 4.8, where machine learning methods will be explored.

4.6. Direct PSF prediction

As discussed in Section 4.4.1, the PSF model is initialized using predefined configuration files and then updated with model inputs from the reduced telemetry dataset to model different conditions. In this context, directly using the reduced telemetry without any modifications [23] is referred to as *direct PSF prediction*. Figure 4.12 schematically demonstrates this principle.

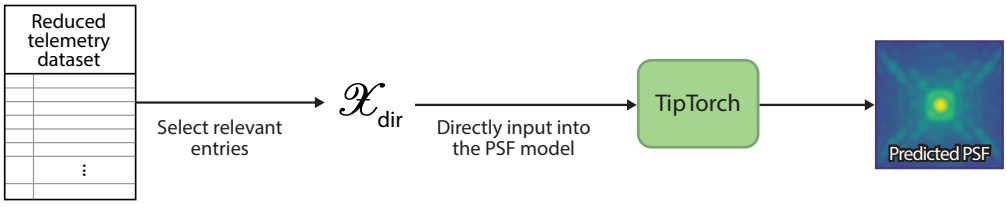


FIGURE 4.12. – Schematic representation of the direct PSF prediction. After the PSF model is initialized with the default instrument's configuration file, a subset of parameters is re-initialized with the values taken from the reduced telemetry dataset. These values are used without any modifications to generate PSF.

23. For example, this will not be the case later in Section 4.8.

It is important to clarify that, in the scope of this chapter, *"prediction" is not used in the sense of forecasting future events*. It simply means inferring outputs based on a given set of inputs. This notion of precision is adopted in machine learning, which is appropriate in the scope of this project since machine learning will be used intensively later in Section 4.8. In this terminology, the classical PSF-R can also be perceived as a method of predicting PSF. However, for clarity, later in this chapter, the PSF prediction will be reserved exclusively for the PSF modeling methods proposed in this thesis.

With these clarifications, the direct PSF prediction results for IRDIS and MUSE-NFM can now be presented.

4.6.1. SPHERE results

Before proceeding with the results, it is important to comment on the model inputs. For IRDIS, the set of model inputs, denoted as $\mathscr{X}_{\text{dir}}$, includes the following entries :

- **Atmosphere**
  - Seeing
  - Wind direction
  - Wind speed
- **Sensor Science**
  - Zenith
  - azimuth
  - Readout noise (σ^2_{RON})
  - Detector gain
- **Telescope**
  - Zenith angle
- **HO WFS sources**
  - Wavelength λ_{WFS}
- **HO WFS sensor**
  - Number of photons
  - Jitter X (J_y)
  - Jitter Y (J_x)
- **Real-Time Controller (RTC)**
  - HO WFS frame Rate
- **Science Sources**
  - Wavelength (λ_{science})

All entries in $\mathscr{X}_{\text{dir}}$ were obtained using the procedures described earlier in Section 4.5.5.1 and also detailed in Appendix C.2.2. However, it is important to emphasize that the values for J_x and J_y retrieved from the archives are always recorded in the archive as zero. This is likely due to a bug in the SPHERE telemetry reduction process. Consequently, J_x and J_y are also set to zero in the PSF model for IRDIS [24].

The results of the direct prediction for IRDIS are shown in Figure 4.13 :

24. Although this remains valid, further work during this thesis revealed that tip-tilt jitter can, in fact, be extracted from ITTM commands.

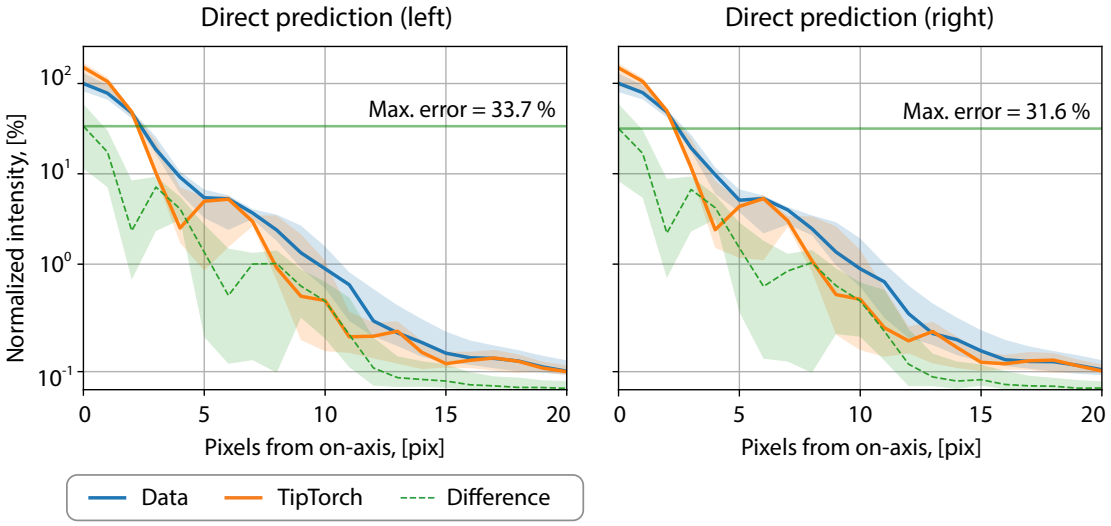


FIGURE 4.13. – Direct prediction results for the dataset of "left" and "right" IRDIS PSFs. The solid lines demonstrate median profiles computed for the datasets of on-sky PSFs, predicted PSFs, and their absolute difference. The error bars show the 1σ deviation. More details on profiles computation can be found in Appendix E.

This figure presents the median profiles for both the on-sky and predicted PSFs, as well as their differences. More details on how these profiles were calculated can be found in Appendix E. All IRDIS profiles here and later were computed by averaging the profiles belonging to different wavelengths in the IRDIS dataset (see Table C.2.1).

In general, *these results indicate poor accuracy*. The model is overly optimistic, which results in an inaccurate prediction of SR and overall PSF morphology. To gain additional insights, it is helpful to examine whether the conclusion can be applied to the direct prediction for MUSE-NFM. If similar results are observed, it could provide clues on the systematic limitations of this direct PSF prediction approach.

More information on how the profiles in Fig. 4.13 were computed can be found in Appendix E.

4.6.2. MUSE-NFM results

The same type of analysis was performed for MUSE-NFM, with the set of model inputs $\mathcal{X}_{\text{dir}}$ including :

- **Telescope**
  - Zenith angle
  - Azimuth
  - Pupil angle
- **Atmosphere**
  - Outer scale (L_0)
  - Seeing
  - C_n^2 Weights
  - C_n^2 Heights
  - Wind speed
  - Wind direction
- **Science sources**

- Wavelength (λ_{science})
- **HO WFSs**
  - Number of photons per WFS
  - Clock rate
- **LO WFS**
  - Number of photons
  - Readout noise σ_{RON}
  - Gain
- **Real-Time Controller (RTC)**
  - LO WFS frame rate
  - HO WFS frame rate
  - Loop gain for HO WFS
  - Loop gain for LO WFS

Figure 4.14 presents the results of the direct prediction. Aside from the difference in inputs, the procedure for obtaining the results here follows the same approach as for IRDIS.

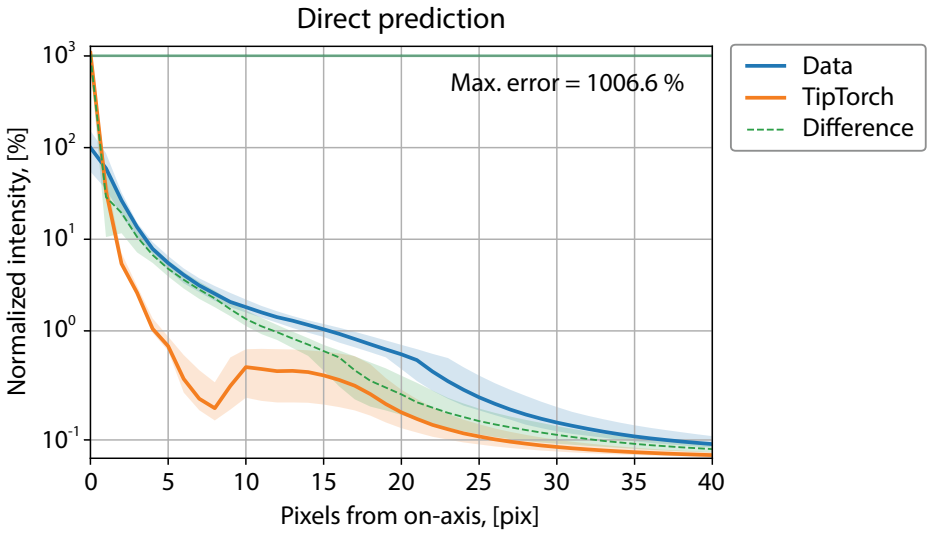


FIGURE 4.14. – Median profiles for the dataset of directly predicted MUSE-NFM PSFs compared with on-sky profiles. The algorithm used to compute these profiles is identical to the one in Figure 4.13. The profiles are calculated for "white" PSFs, which are averaged over all wavelengths.

However, unlike in the case of IRDIS, the values of J_x and J_y could be retrieved from the telemetry, although they were not stored directly. The jitter values were estimated based on the IRLOS photon count, which enabled the calculation of the WFS noise variance using Equations (47) and (48) from Correia et al. (2017). The process of computing J_x and J_y then begins with calculating

$$n_D = \frac{3600 \cdot 180 \cdot \lambda_{\text{WFS}}}{\pi \, \alpha_{\text{pix}} \, d_{\text{sub}}}, \quad n_T = \frac{3600 \cdot 180 \cdot \lambda_{\text{WFS}}}{\pi \, \alpha_{\text{pix}} \, r_0(\lambda_{\text{WFS}})},$$

where n_T is the image FWHM, and n_D is the FWHM of the diffraction pattern of a sub-aperture, λ_{WFS} is the IRLOS WFSing wavelength, which is the average of central wavelengths at J and H bands, d_{sub} is the diameter of the WFS sub-aperture. For IRLOS,

$d_{\text{sub}} = 2$; α_{pix} is the angular size of the detector pixels, $r_0(\lambda_{\text{WFS}})$ is re-normalized to the WFSing wavelength. Then, the variance of the detector noise σ^2_{det} and the photon noise σ^2_{ph} can be computed as

$$\sigma^2_{\text{det}} = \frac{\pi^2}{3}\left(\frac{\sigma_{\text{RON}}}{N_{\text{ph}}}\right)^2\left(\frac{N^2_{\text{pix}}}{n_D}\right)^2, \quad \sigma^2_{\text{ph}} = \frac{\pi^2}{2N_{\text{ph}}}\left(\frac{n_T}{n_D}\right)^2,$$

$$\sigma^2_\eta = \varepsilon\,(\sigma^2_{\text{det}} + \sigma^2_{\text{ph}})\left(\frac{\lambda_{\text{atm}}}{\lambda_{\text{WFS}}}\right)^2, \text{ [rad}^2\text{]},$$

where N_{ph} is the number of photons per sub-aperture per frame; ε is the excessive noise factor that can be assigned to 1 here; λ_{atm} and λ_{WFS} are the atmospheric and WFSing wavelength, respectively; σ_{RON} is the LO WFS detector readout noise measured in photoelectrons. Multiplication by their ratio is necessary to re-normalize the WFSing noise variance σ_η to the atmospheric wavelength. The final expression for the TT jitter then is

$$J_x = J_y = \arctan\left(\frac{2\,\text{TT}_{\text{max}}}{D}\,\frac{\lambda_{\text{WFS}}}{2\pi}\sigma_\eta\right), \text{ [mas]}, \quad J_{xy} = 0.$$

Here, D is the diameter of the telescope aperture, $\text{TT}_{\text{max}} \approx 1.97$ is the maximum absolute values of the tip and tilt modes on the edge of the aperture.

To summarize this discussion, the profiles shown in Figure 4.14 were computed following the same methods detailed in Appendix E. The results demonstrate an even greater discrepancy between the predicted and experimental data for MUSE-NFM compared to IRDIS.

4.6.3. Conclusion

A comparison of the PSF models for both instruments reveals significant discrepancies with the experimental results, affecting both the peak value and the PSF wings. High error on PSF peak values suggests an inability of the model to predict Strehl ratios correctly, while misprediction of the PSF wings indicates errors in capturing overall PSF morphology, rendering the model unsuitable for applications such as deconvolution.

What insights can be drawn from this substantial prediction inaccuracy for both instruments? An inspection of the profiles already hints at potential sources of error.

— For IRDIS, the results suggest that neglecting the TT jitter significantly contributes to poor peak prediction accuracy, implying that TT jitter is an influential factor in reality.

— For MUSE-NFM, the TT jitter appears to be severely underpredicted, leading to poor peak value predictions.

— The real on-sky PSFs exhibit complex behavior around the first Airy ring. This is difficult to see from the profiles alone but can be seen from the PSFs shown earlier in Figure 2.10. The current formulation of the PSF model cannot account for these effects.

— The model underrepresents the WFSing noise for both instruments. This is evident in the regions influenced by AO-corrected spatial frequencies, where the modeled profiles exhibit values lower than the real on-sky PSFs.

— The incorrect prediction of PSF morphology also implies a poor prediction of energy distribution per spatial frequency, leading to errors in the peak values as a consequence.

Several hypotheses can be proposed to explain these errors.

— The model may not accurately model the physics.

— The reduced telemetry may contain intrinsic errors and biases due to incorrect measurement or processing, meaning that it does not accurately correspond to the on-sky PSFs.

— Combination of both.

To address these issues, the PSF model must be validated for both instruments, which will be the focus of the next section.

4.7. PSF model validation

The inaccurate results observed in the previous section with direct prediction highlight the need to identify the potential source of the problem. Therefore, validating the PSF model is essential to ensure it can accurately reproduce the morphology of on-sky PSFs. If fitting the physics-based model to on-sky data results in only a small discrepancy, it suggests that the model itself is capable of capturing the on-sky PSF structure. In this case, the issue is more likely to stem from inaccuracies in the model inputs.

This validation is implemented by fitting the PSF model to data. As was already mentioned, fitting involves finding the optimal combination of model parameters that yields the closest match to the on-sky PSFs. Some of the model inputs must become optimizable to do so. The parameter values from the dataset $\mathscr{X}_{\text{dir}}$ will then serve as an initial guess for the optimized inputs. In many aspects, this task resembles the problem previously discussed throughout Chapter 3 and in the embedded paper in the context of LIFT. However, in this case, the model is defined with physical parameters instead of modal coefficients.

As in the case of LIFT, finding the optimal parameter values can be addressed both through non-linear optimization and sampling. Given that this optimization process must be performed for every sample in the on-sky dataset, non-linear optimization is preferred over sampling due to its typically faster convergence, albeit with a higher

probability of parameter coupling (as discussed in Section 3.2).

Similarly to DIP (see Section 5.3 in the embedded paper), the BFGS optimization algorithm was chosen. In the scope of TipTorch, it demonstrated better convergence compared to L-BFGS or GN. The implementation of the BFGS algorithm was taken from the `pytorch-minimize` [25] library.

Similar to DIP, TipTorch relies on automatic differentiation, enabling fast, flexible, and precise computation of derivatives with respect to nearly any parameter. This capability enhances the efficiency and stability of the nonlinear optimization process. However, for a parameter to be optimizable, it must still satisfy several conditions :

— **Preservation of model differentiability :** Since non-linear optimization techniques are employed, it is crucial to ensure that the chosen parameters do not disrupt the model's differentiability. For example, from the computation point of view, calculating the derivative of the loss function with respect to sampling or wavelength would require re-initializing the PSF model, thus interrupting gradient computation via auto-differentiation.

— **Exclusion of system-defined parameters :** Parameters fixed by the system's design and known with high precision should not be optimized. This category includes WFS gains, loop frequencies, filters, central wavelengths, and other system/regime-specific constants.

— **Minimization of parameter cross-Coupling :** To prevent cross-coupling between parameters, some may need to be fixed and thus excluded from the optimization process. This form of regularization has been discussed previously in the context of LIFT in Section 4.1 of the embedded paper.

Knowing this, it is now possible to identify a set of variables that can be unlocked for optimization. However, this set of optimizable parameters is system-specific and thus differs for IRDIS and MUSE-NFM. Thus, model verification must be discussed separately for both instruments.

4.7.1. SPHERE

4.7.1.1. Modifying the model

While the preceding section emphasized the necessity to work on model inputs, minor adjustments to the PSF model are still required to match the on-sky data precisely. For example, in the case of IRDIS, on-sky PSFs display noticeable quasi-static patterns in the first Airy ring (see Figure 2.10). Furthermore, Section 4.6 suggests that the predictions for WFS noise are overly optimistic, requiring further adjustments. Moreover, although the $\mathrm{OTF}_{\mathrm{TT}}$ component is already incorporated, tip-tilt effects are currently underrepresented due to the absence of relevant entries for IRDIS, as discussed in Section 4.6.1. Addressing these factors requires the following modification to the model.

25. https://github.com/rfeinman/pytorch-minimize

WFSing noise

As discussed in Section 4.3.1, the wavefront sensing (WFS) noise error can be modeled using the WFSing noise PSD (W_{noise}), which in the focal-plane domain manifests itself as an additional background within the correction radius.

However, previous results demonstrated in Section 4.6 revealed that the initial estimates for this PSD contributor were probably overly optimistic, which can potentially be adjusted with more accurate WFSing noise variance values. The WFS noise variance, which drives this PSD contributor, is physically determined by the number of photons and the parameters of the WFS detector. To address the discrepancies in the WFS noise variance, an additional data-driven term Δn is introduced. This term is designed to absorb the error while keeping the number of photons and detector parameters fixed.

By incorporating Δn into the noise variance expression, the overall noise variance is no longer solely defined by the photon count and detector properties but also includes this adjustable component. As a result, the expression for the noise variance is modified to :

$$\sigma_{\eta}^2 = \sigma_{\text{photon}}^2 + \sigma_{\text{detector}}^2 + \Delta n,$$

where σ_{photon} is the photon noise variance and σ_{detector} is the detector noise variance. This Δn term can then be used as a free parameter during the fitting process, allowing it to "absorb" underrepresented noise or suppress excessive noise variance if needed, leading to a more accurate representation of the WFS noise in the model.

LWE

Based on the findings in Sauvage et al. (2016); Milli et al. (2018) and an examination of the PSF morphology in the current dataset (see Figure 2.10), it is reasonable to assume that the aberrations observed in the first Airy ring are due to LWE. This effect can be effectively modeled by introducing a static phase component defined by a basis comprising differential piston [26] and per-island tip-tilt modes [27], as illustrated in Figure 4.15.

26. While one differential piston mode might be omitted, it is retained here for simplicity.
27. A compressed basis, such as the one introduced in Pourré et al. (2022), could also be employed.

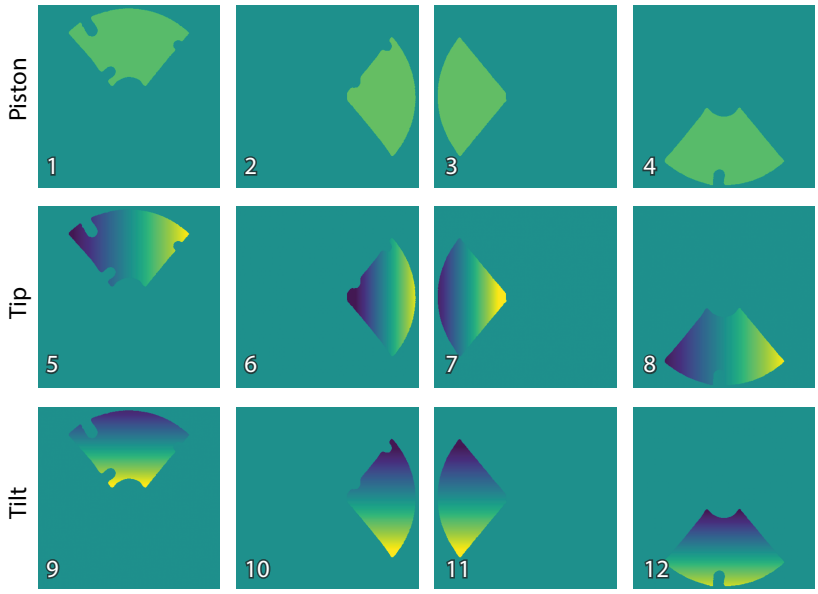


FIGURE 4.15. – Orthogonal modal basis that represents the LWE-induced OPD in the pupil plane of IRDIS. The basis contains 4 pistons, 4 tips, and 4 tilt modes for each island of the IRDIS pupil.

With the addition of LWE, the modified version of the expression 4.4 for the static complex pupil-plane phase in the case of the IRDIS is

$$\mathscr{P}(\mathbf{x}) = P(\mathbf{x})\, A_p(\mathbf{x}) \exp\left(-i \sum_{i=1}^{12} c_i \,\mathrm{LWE}(\mathbf{x})\right),$$

where LWE($\mathbf{x}$) are the LWE modes demonstrated in Figure 4.15, c_i are the corresponding modal coefficients. Other terms denoted the same as in Eq.(4.4).

The coefficients c_i can then be treated as optimized parameters, following a similar approach to the estimation of LWE as described in Beltramo-Martin et al. (2020). This method assumes that LWE remains quasi-static during the exposure. Fundamentally, the estimation of these coefficients is similar to the FP WFSing method discussed earlier in Chapter 3. The key distinction lies in the absence of phase diversity, with the approach instead relying on amplitude diversity introduced by asymmetries in the IRDIS pupil.

Tip-tilt jitter

As previously mentioned, this term does not require any modifications to the PSF model. However, determining the correct values for TT jitter requires fitting.

4.7.1.2. Fitting the model

After incorporating the necessary modifications into the model, the next step is to proceed with the fitting. Essentially, the model fitting process, in this case, can

be formulated similarly to the approach used in LIFT, with the difference being that this time, a set of other physical parameters is optimized jointly with the LWE modal coefficients. The optimal estimate for the modal parameters is determined as follows :

$$\hat{\mathscr{X}} = \text{argmin}_{\mathscr{X}} \, \mathscr{L}\left(\mathbf{I}(\mathscr{X}), \mathbf{I}_{\text{data}}\right), \tag{4.9}$$

where $\mathscr{X}$ is the set of the optimizable parameters, defined as :

$$\mathscr{X} = \{ r_0, \mathbf{F}, \mathbf{b}, \Delta n, \delta x, \delta y, J_x, J_y, J_{xy}, W_{\text{dir}}, \mathbf{c} \}. \tag{4.10}$$

Here, $\hat{\mathscr{X}}$ is the set of optimal parameter values obtained by fitting the PSF model to on-sky data; r_0 is Fried's parameter; $\mathbf{F} = (F_L, F_R)$ is the normalizing multiplier; $\mathbf{b} = (b_L, b_R)$ is a constant background addend; left and right IRDIS PSFs correspond to different flux and background values; δx and δy are the sub-pixel PSF shifts measured in pixels, J_x, J_y, J_{xy} are the tip-tilt jitter coefficients in mas, and W_{dir} is the wind direction angle in degrees.

The minimization criterion (loss function) $\mathscr{L}(\cdot)$ is defined as

$$\mathscr{L}\left(\mathbf{I}(\mathscr{X}), \mathbf{I}_{\text{data}}\right) = \text{MAE}\left(\mathbf{I}(\mathscr{X}), \mathbf{I}_{\text{data}}\right) + \sum_{\mathbf{p}\in\mathscr{P}} A \exp\left(-\frac{\sum_{m=1}^{M} (c_m - p_m)^2}{2\sigma^2} \right) + \epsilon \frac{\|\mathbf{c}\|^2}{M}, \tag{4.11}$$

where $\mathbf{I}(\mathscr{X})$ is the simulated and $\mathbf{I}_{\text{data}}$ is the on-sky PSFs, $\mathbf{c}$ is the vector of LWE modal coefficients (in nm) from the basis of size $M = 12$. Given that LWE modes fitting introduces 12 additional parameters. Hence, it is crucial to mitigate the parameter coupling and overfitting. To address this, regularization terms must be incorporated into the optimization process. In this scope, here, the third addend is the L_2-regularization term, which prevents LWE coefficients from reaching high values during optimization and allows suppressing the pupil-wide piston mode, $\epsilon = 10^{-4}$. The second addend is a Gaussian penalty that tells the optimizer to avoid specific coefficient patterns defined in the set $\mathscr{P}$. This term is larger when the vector $\mathbf{c}$ is close to one of the vectors $\mathbf{p}$. Here, $\sigma = 25$ nm is the standard deviation that defines the "influence range" of each vector $\mathbf{p}$, $A = 5$ is the weighting factor of the error term. The coefficient patterns are defined as

$$\begin{aligned}
\mathbf{p}_1 &= B \cdot (0,0,0,0,0,-1,1,0,-1,0,0,1),\\
\mathbf{p}_2 &= B \cdot (0,0,0,0,0,1,-1,0,1,0,0,-1),\\
\mathbf{p}_3 &= B \cdot (0,0,0,0,1,0,0,-1,0,1,-1,0),\\
\mathbf{p}_4 &= B \cdot (0,0,0,0,-1,0,0,1,0,-1,1,0),\\
\mathbf{p}_5 &= B \cdot (0,0,0,0,0,-1,1,0,1,0,0,-1),\\
\mathbf{p}_6 &= B \cdot (0,0,0,0,0,1,-1,0,-1,0,0,1),
\end{aligned}$$

$$\mathscr{P} = \{\mathbf{p}_k\}_{k=1}^{6},$$

where $B = 50$ nm is the normalizing multiplier. The values of A, B, σ, ϵ constants were

obtained empirically to minimize overfitting and balance the error terms. The vectors in $\mathscr{P}$ represent unrealistic symmetrical patterns that typically tend to emerge during non-linear optimization. More rigorous hyper-parameter searches can be performed to improve the results further.

The first addend in Eq.(4.11) is the MAE loss term that compares estimated and real PSFs, which is defined as :

$$\mathrm{MAE}\left(\mathbf{I}(\mathscr{X}), \mathbf{I}_{\mathrm{data}}\right) = \sum_{\lambda} \sum_{i=1}^{H} \sum_{j=1}^{W} \left| I_{ij\lambda}(\mathscr{X}) - I_{ij\lambda,\mathrm{data}} \right|, \quad \lambda = \{\lambda_L, \lambda_R\}, \tag{4.12}$$

where i and j are the pixel coordinates; W, H are the width and height of the PSF images; λ is the set of wavelengths corresponding to the left and right channels of IRDIS. Note that both left and right PSFs are fitted simultaneously. Since they are influenced by (almost) the same wavefront aberrations, using both PSFs creates data redundancy that better constrains optimization and leads to better parameter estimates. The MAE loss was employed because it provides a more balanced weighting across different regions of the PSF, as discussed earlier in Section 3.3.1.

By applying this optimization procedure to all samples in the IRDIS dataset, the dataset of fitted parameters $\mathscr{D}_{\mathrm{fit}}$ is created :

$$\mathscr{D}_{\mathrm{fit}} = \{\mathscr{X}_k\}_{k=1}^{N_{I,\mathrm{IRDIS}}},$$

where $N_{I,\mathrm{IRDIS}} = 3112$ is the number of available on-sky samples in the entire dataset (will be later discussed in Section 4.8.4.2). The procedure is illustrated in Figure 4.16.

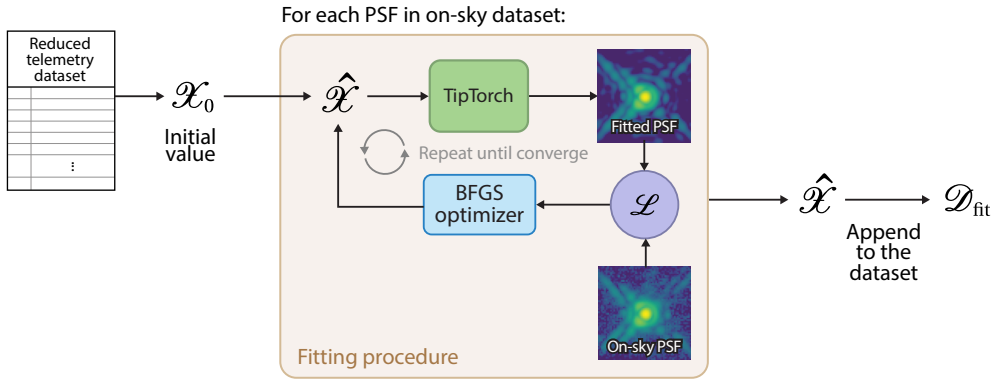


FIGURE 4.16. – Fitting TipTorch model to on-sky PSFs to test if the model can accurately recreate their morphology. The resulting optimal values of model inputs $\hat{\mathscr{X}}$ are stored in the dataset $\mathscr{D}_{\mathrm{fit}}$.

4.7.1.3. Results

It is now possible to evaluate the median accuracy of the fitting process for the entire dataset of fitted PSFs, as shown in Figure 4.17. This radial profile was generated using

the same method as described in Appendix E, with the key difference being that the line here represents the fitted profiles instead of the predicted ones.

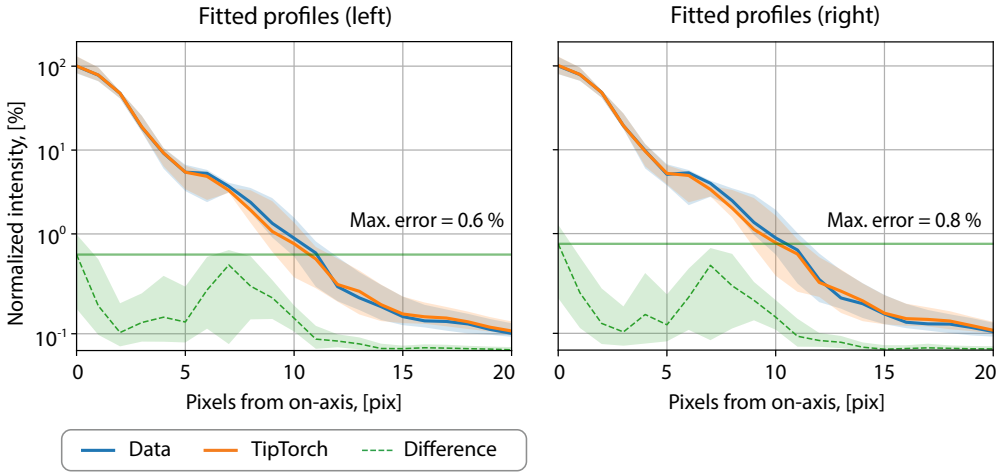


FIGURE 4.17. – The radial profiles derived from fitting the TipTorch model to the IRDIS PSF dataset. The results show minuscule discrepancies from the on-sky PSFs, indicating very high fitting accuracy.

The results demonstrate that the accuracy of the fitted results is exceptionally high. As discussed previously, this suggests that the model is likely physically accurate. However, to confirm this conclusion, it is crucial to apply the same validation procedure to the MUSE-NFM data and verify whether the results are consistent.

4.7.2. MUSE

4.7.2.1. Modifying the model

The model validation process for TipTorch applied to MUSE-NFM closely follows the same steps as for IRDIS. Similarly, before validation, the model needs slight modifications compared to the version used for direct prediction. As discussed in Section 4.6.3, the main discrepancies were attributed to WFS noise, TT jitter, and inaccurate predictions of PSF peak.

WFS Noise

The approach here mirrors the one used for IRDIS. An additional Δn absorber is introduced to the WFS noise variance, resulting in the same expression for the corrected WFS noise variance :

$$\sigma_\eta^2 = \sigma_{\text{ph}}^2 + \sigma_{\text{det}}^2 + \Delta n.$$

The notation for the terms in this section matches the one used in Section 4.7.1.1.

Peak Absorber

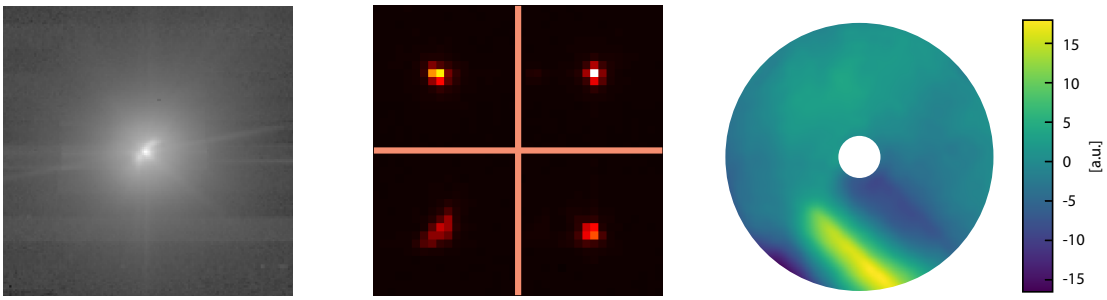
Unlike IRDIS, no distinct LWE patterns were observed in the MUSE-NFM PSF dataset. However, the initial fit of the model struggled to represent the PSF core accurately. To address this, an additional PSD contributor was added to capture the underrepresented energy in the PSF core. This can be modeled using a simplified Moffat [28] function similar to one from Fétick et al. (2019b) :

$$W_{\text{Moffat}}(\mathbf{k}) = \widetilde{P}(\mathbf{k}) \left(\frac{A\,(\beta_M - 1)\,\alpha_M^{2(\beta_M-1)}}{\pi \left(\alpha_M^2 + \|\mathbf{k}\|^2\right)^{\beta_M}} + B \right)$$

Here, A is the amplitude, B is an additional PSD background constant, α_M is a scaling factor, β_M controls the shape of the Moffat profile, $\mathbf{k}$ represents spatial frequencies ($\|\mathbf{k}\| < k_c$), $\widetilde{P}(\mathbf{k})$ is the piston mode filter.

PSF elongation absorber

This error term, specific to MUSE-NFM, can be introduced as a static phase contributor. To gain a clearer understanding of this error, it is best to refer to Figure 6 on page 9 of the embedded paper. There, a noticeable phase bump was described and characterized in the full-pupil measurements of IRLOS. As later tests revealed, this "bump" originates from incorrectly calibrated reference slopes in the AOF. In a way mentioned earlier in Section 2.5.4, it propagated through the AO system via incorrect DM commands and altered the morphology of both IRLOS and MUSE-NFM PSF as shown in Figures 4.18(a,b). From this point of view, this "phase bump" represents a good example of an NCPA, which remained unseen for AO WFSs but could be retrieved from the PSF morphology. Note the similarity in the overall shape of this defect depicted in Figure 4.18(c) and in Figure 6 on page 9 of the embedded paper.



(a) MUSE-NFM PSF (b) IRLOS image (c) DM shape

FIGURE 4.18. – (a) Effect of the "phase bump" on MUSE-NFM PSFs (image in log-scale); "phase bump" seen by IRLOS [29]; (c) "phase bump" pattern reproduced by DSM.

28. The peak absorber, in a way, creates a hybrid model combining elements of TipTorch and PSFAO.
29. Note that IRLOS PSFs shown in Figure 4.18(b) are different from ones studied is Chapter 3. This is

The exact shape of this phase bump in the pupil plane was later characterized and then incorporated into the TipTorch model. The resulting expression for $\mathrm{OTF}_{\mathrm{static}}$ in the case of TipTorch for MUSE-NFM looks as follows :

$$\mathscr{P}(\mathbf{x}) = P(\mathbf{x})\,\exp(-i\,P_b(\mathbf{x})\,a_b).$$

Here, $P_b(\mathbf{x})$ is the term presented in Figure 4.18(c). Other terms are denoted the same as in Eq.(4.4).

The power of this effect varies for different PSFs. Thus, in the model, the bump phase term is multiplied by the variable coefficient a_b. Fitting this coefficient to the data effectively turns the task into a phase-retrieval problem, but only for a single coefficient. To avoid sign ambiguity, the coefficient a_b is always assumed to be positive.

4.7.2.2. Fitting the model

The problem formulation for MUSE-NFM follows the same structure as for IRDIS outlined in Eq.(4.9). The set of fitted parameters is, however, slightly different :

$$\mathscr{X} = \{\mathbf{F}, \mathbf{b}, r_0, \delta x, \delta y, \mathbf{J}_x, \mathbf{J}_y, J_{xy}, \Delta n, A, B, \alpha_M, a_b\}. \tag{4.13}$$

Note that most of the parameters in the set $\mathscr{X}$ are multi-chromatic, meaning that they are wavelength independent [30]. The exceptions are the vectors of normalization factors $\mathbf{F} = \{F(\lambda)\}$, constant background biases $\mathbf{b} = \{b(\lambda)\}$, and tip-tilt jitter values $\mathbf{J}_{x,y} = \{\mathbf{J}_{x,y,xy}(\lambda)\}$. Note that J_{xy} stays for pupil rotation angle as outlined earlier in Section 4.3.

The definition of the loss function also differs for MUSE-NFM compared to SPHERE and is expressed as :

$$\mathscr{L}(\mathbf{I}(\mathscr{X}), \mathbf{I}_{\mathrm{data}}) = \alpha\,\mathrm{MSE}(\mathbf{I}(\mathscr{X}), \mathbf{I}_{\mathrm{data}}) + \beta\,\mathrm{MAE}(\mathbf{I}(\mathscr{X}), \mathbf{I}_{\mathrm{data}}). \tag{4.14}$$

The MSE and MAE loss functions are normalized slightly differently compared to Eq.(4.12) and expressed as :

$$\mathrm{MAE}(\mathbf{I}(\mathscr{X}), \mathbf{I}_{\mathrm{data}}) = \frac{1}{HW}\sum_{\lambda}\sum_{i=1}^{H}\sum_{j=1}^{W} w(\lambda)\left|I_{ij\lambda}(\mathscr{X}) - I_{ij\lambda,\mathrm{data}}\right|, \tag{4.15}$$

$$\mathrm{MSE}(\mathbf{I}(\mathscr{X}), \mathbf{I}_{\mathrm{data}}) = \frac{1}{HW}\sum_{\lambda}\sum_{i=1}^{H}\sum_{j=1}^{W} w(\lambda)^2\left(I_{ij\lambda}(\mathscr{X}) - I_{ij\lambda,\mathrm{data}}\right)^2. \tag{4.16}$$

In these equations, $\alpha = 200$ and $\beta = 0.4$ are empirical weighting factors that help to balance MAE and MSE loss terms. Here, the MAE component helps to accurately

to remind the reader that chapter IRLOS was used in its experimental full-pupil configuration, while the 2×2 regime of IRLOS presented here is the only one used in real operations.

30. Worth noting, that δx and δy can be expressed as a function of wavelength to model chromatic dispersion. However, this effect is minimal in the reduced MUSE cubes. Thus, these parameters are responsible for the overall sub-pixel shift of the PSF relative to the center of FoV.

weight the off-core PSF regions, while MSE is responsible for fitting the peak value. Meanwhile, the empirical chromatic weighting multiplier $w(\lambda) = 3.162 - \frac{\lambda}{425\,\text{nm}}$ helps to weight each spectral slice differently.

It is important to note that, as in the case of IRDIS, the fitting process here is polychromatic. This means that all 14 spectral slices of the polychromatic PSF are simultaneously fitted to their corresponding slices in the on-sky PSFs. As previously mentioned in Section 4.3.1, most of the PSD error contributors are computed for the λ_{atm} wavelength and then re-normalized for each science wavelength. In this scope, just as for IRDIS, fitting to multiple polychromatic slices creates data redundancy that helps to better constrain the optimization problem. Meanwhile, the simultaneous usage of MAE and MSE losses helps to empirically balance the importance of PSF core and PSF wings, as outlined earlier in Section 3.3.1.

4.7.2.3. Results

Similarly to the approach used for IRDIS, the TipTorch model was fitted to the entire dataset of on-sky PSFs for MUSE-NFM using the procedure outlined in the previous subsection. The resulting median profiles, shown in Figure 4.19, are again calculated using the methodology described in Section 4.6. Note that these profiles represent "white" PSFs, meaning all spectral slices were averaged into one image before computing the profiles.

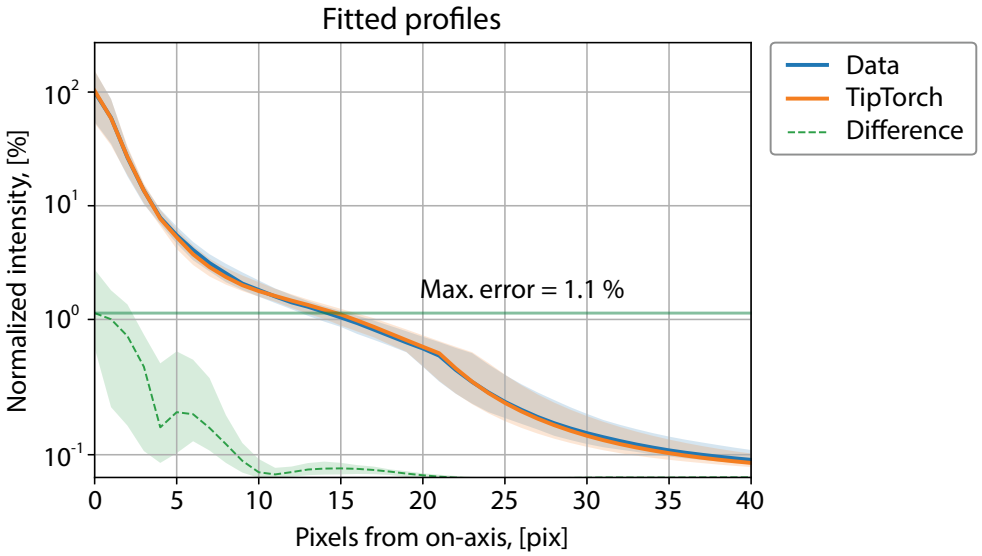


FIGURE 4.19. – Radial profiles derived from fitting the TipTorch model to the MUSE-NFM PSF dataset. Similar to IRDIS, the results here demonstrate minimal discrepancies from the on-sky PSFs.

The fitting quality is exceptionally high, indicating that the model can accurately represent the on-sky data across the entire spectral range of MUSE-NFM, provided the parameters are properly tuned.

4.7.3. Conclusions

To summarize, the fitting results for both instruments revealed several key insights :

— The model demonstrated its ability to accurately represent two distinct instruments that operate on different AO principles, utilizing largely the same mathematical framework, which underscores the model's universality.

— However, to achieve this accuracy, additional, somewhat ad hoc components had to be incorporated into the model.

— Nevertheless, after a few additions, TipTorch can now accurately model PSF morphology when provided with accurate input values, suggesting that prediction inaccuracies likely stem from errors in the input data, especially after the PSF model itself was slightly modified.

These findings suggest that enhancing prediction accuracy could be achieved by improving the accuracy of the model's input predictions. This concept, referred to as the *"calibration"* of the PSF model, will be further explored in the next section.

4.8. Calibrated PSF prediction

As mentioned at the end of the previous section, it was proposed that calibrating the model could enhance prediction accuracy. However, the concept of "calibration" and the methods for implementing it require a detailed examination. This section is dedicated to exploring these aspects in detail.

4.8.1. The concept of model calibration

Biases in the on-sky data ?

As direct prediction shows, the analytical relationships between reduced telemetry and the morphology of the PSF do not work as expected. Validation indicated that the physics underlying the model is likely accurate and sufficiently representative. Thus, the issue appears to stem from the model inputs. Therefore, enhancing prediction accuracy necessitates refining these inputs.

The discrepancies may be due to potential errors and biases in the reduced telemetry data. Such biases can arise from the reduction procedures' specifics or from using different instruments and methods to measure the same quantities. For example, both MASS-DIMM and SPARTA provide seeing estimates, but as illustrated in Figure C.1 of the embedded paper, their estimates can differ significantly. The same point about the biases in the data was already raised in Fétick et al. (2019b).

Learning from the data

The biases present in reduced telemetry can be systematic. The goal of the calibration process, then, is to measure and compensate for them. Since reduced telemetry is obtained from on-sky observations, these biases should be directly extractable from the on-sky data.

It is also important to note that there are correlations between different reduced telemetry entries. For example, the r_0 and L_0 estimates are derived from the same raw telemetry source, while others, such as τ_0 and wind direction, are mathematically related. Additionally, certain parameters, such as the SR estimates produced by SPARTA, may correlate with multiple other parameters, such as seeing and τ_0, since they are connected via the physics of the process. This suggests that some reduced telemetry entries are functions of others. It also implies that information about model input biases might be partially embedded within the values of other related parameters.

Interconnections between data inputs

However, the examples of correlations provided above are relatively straightforward and follow well-established analytical relationships. On the other hand, real data might contain more complex, non-linear relationships between various entries that are not fully understood but are inherently present within the data. This also introduces the idea of using more telemetry parameters than those required by the PSF model as inputs. The rationale is that these additional data could bring more relevant and useful information. With this approach, the model inputs are no longer

taken directly from the reduced telemetry (as done in Section 4.6). Instead, they are represented as functions of other reduced telemetry entries. This approach means that the notions of *"reduced telemetry inputs"* and *"PSF model inputs"* become distinct from this point. The function mapping of these two will be referred to as a *calibrator*. This relation can be summarized in the following expression.

$$f : \mathbf{X} = f(\mathbf{Z}) + \mathbf{E}.$$

Here, f is the calibrator, $\mathbf{E}$ is an error term that captures the inherent noise, random variation, and imperfections in the data that cannot be explained by the calibrator f; $\mathbf{X}$ and $\mathbf{Z}$ denote the vectors of model inputs and reduced telemetry parameters, respectively. From a technical perspective, both $\mathcal{Z}$ and $\mathcal{X}$ are stored as vectors of numerical values. However, in the notation adopted in this chapter, they are represented as sets for convenience, which means $\mathbf{X} = \mathcal{X}$, $\mathbf{Z} = \mathcal{Z}$.

From this point of view, the objective of PSF model calibration is to adjust the values of the model inputs based on reduced telemetry entries to improve the accuracy of PSF predictions. In other words,

$$\min_f \mathscr{L}(\mathbf{I}(\widetilde{\mathcal{X}}), \mathbf{I}_{\text{data}}), \quad \text{where } \widetilde{\mathcal{X}} = f(\mathcal{Z}).$$

Here, $\widetilde{\mathcal{X}}$ is the predicted set of model inputs. Later on, the "$\tilde{\cdot}$" notation will stay for predicted parameters. The methodology for finding the optimal f will be discussed below.

4.8.2. Approaches to model calibration

This data-oriented approach to model calibration naturally implies using machine learning (ML) methods. Meanwhile, there are several ways to address the calibration using ML.

4.8.2.1. Calibrator as multivariate regression

As mentioned, we are now operating under a framework in which model inputs $\mathcal{X}$ are treated as functions of reduced telemetry entries $\mathcal{Z}$, and the relationships between them must be extracted from the data. But how should these functions be formulated? In this case, the simplest ML model is multivariate linear regression. Then, the calibration process involves training this model to map reduced telemetry to model inputs that yield the highest PSF prediction accuracy.

However, assuming a linear relationship between telemetry and model inputs could be overly simplistic. To capture more complexity, non-linear multivariate regression models driven by polynomials or other complex analytical functions could be employed. These models may better represent potential non-linear relationships. However, a strictly parametric approach might restrict the calibrator's flexibility. Therefore, it is worth considering more adaptable methods for implementing the calibrator.

4.8.2.2. Full machine learning

In the previous method, the calibrator and the PSF model are treated as separate entities. In contrast, a fully ML-driven approach that encapsulates both the calibrator and the PSF model into a single unified predictor can be suggested. For example, it can be implemented as a generative neural network (NN) and trained to predict PSFs directly from reduced telemetry inputs. A Generative Adversarial Network (GAN) or a diffusion-based architecture could be particularly well-suited for this task, especially given recent advances in using these generative architectures for image generation (Rombach et al. (2021); Dhariwal & Nichol (2021)). In theory, this method offers maximum flexibility. As the universal approximation theorem states (see Kidger & Lyons (2019)), NNs can approximate a function of any complexity, provided NNs have sufficient training parameters and depth. Thanks to the configurable nature of NNs, such a predictor could learn the behavior of an AO system solely from data without being constrained by predefined assumptions about the relationships between the data, unlike the previous method. This allows for bypassing preconceptions and learning the relationship between PSF morphology and reduced telemetry directly from the data.

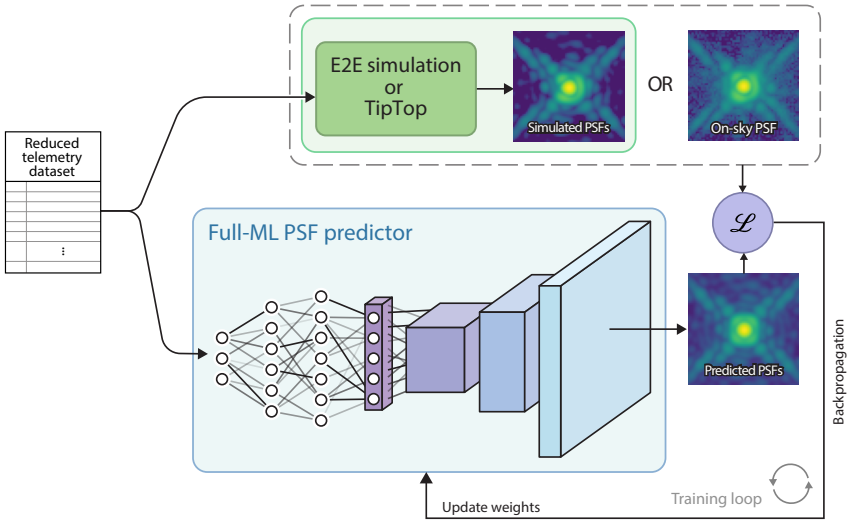


FIGURE 4.20. – Schematic view of a hypothetical full ML-driven PSF predictor that can be trained on the extensive realistic E2E simulations and then fine-tuned on the real data. Such an approach can offer ultimate flexibility at the price of low interpretability.

However, a significant limitation of this approach in the scope of this project is the availability of datasets. As discussed in Section 4.5, the datasets utilized in this work consist of only a few hundred to a few thousand samples, which is scarce for the task discussed. The fewer assumptions a model makes about the data structure, the more flexible it is, but this flexibility comes at the cost of requiring more trainable parameters. Consequently, more data are needed to adequately train the model and

avoid overfitting. In addition, the training data must be diverse enough to span the entire space of possible inputs and outputs to ensure that the model generalizes effectively.

The hypothetical NN proposed in this approach would require a substantial number of trainable parameters to learn the complex physics of the AO system accurately from the data alone. Training a network with a parameter count that greatly exceeds the number of available training samples would likely result in overfitting and a lack of physical accuracy and generalization. However, this limitation can be addressed via the transfer learning approach. In theory, it is possible to pre-train such an NN on virtually unlimited synthetic datasets generated by realistic E2E simulations or uncalibrated P3 or TipTorch models, enabling the network to learn basic physics from simulations and then fine-tune it on a few hundred on-sky samples. This remains an intriguing avenue for future exploration, especially as the dataset of available on-sky PSFs will expand in the future.

However, another drawback of this purely data-driven approach is its black-box nature, which could disadvantage specific applications, such as system diagnostics, where understanding the link between PSF features and input telemetry values is crucial. To date, interpreting the predictions of NNs remains a challenge (Zhang et al. (2021)).

Given these considerations, the calibrated PSF model should be flexible yet not overly large, primarily to avoid overfitting with the scarce available datasets and to somewhat preserve interpretability.

4.8.2.3. Hybrid approach

Considering what was said before, the hybrid approach can be implemented by leaving the PSF model intact while implementing the calibrator as an additional data-driven component. This configuration is similar to the first approach. However, the calibrator here is represented by an NN instead of a linear regressor. The PSF model remains constrained by its physical principles, which is acceptable, assuming that the model is sufficiently physically representative. A schematic of this approach is shown in Figure 4.21.

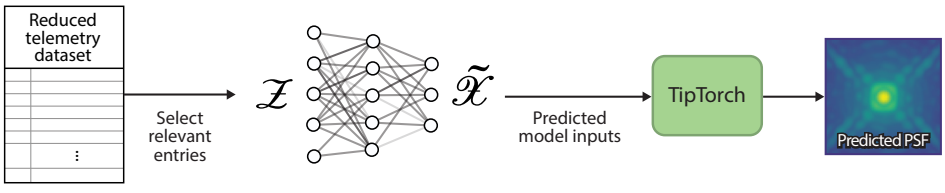


FIGURE 4.21. – Scheme of the proposed calibrated (hybrid) PSF prediction approach. The ML component, trained on on-sky reduced telemetry, transforms the dataset ($\mathcal{Z}$) into model inputs $\widetilde{\mathcal{X}}$ before passing them to the PSF model. Unlike Figure 4.12, where reduced telemetry is used unaltered.

In this setup, the calibrator learns the transformation from a vector of reduced telemetry to a set of model inputs. From the computational point of view, these vectors are compact, consisting of several dozen or fewer values. This implies a compact NN architecture with relatively few trainable parameters. The smaller size reduces the likelihood of overfitting the limited available datasets.

Apart from NNs, a Gradient Boosting Regressor (GBR) can be used to implement the calibrator effectively. GBRs offer several advantages over NNs, particularly in their robustness to outliers and reduced susceptibility to overfitting. This stems from their ensemble nature, where predictions are derived from a sequence of decision trees, each of which is trained to correct the residual errors of its predecessors. GBRs tend to perform well on small to medium-sized datasets, especially when the underlying data structure is not excessively complex. Furthermore, GBRs typically require fewer hyperparameters, simplifying the model initialization and tuning process. While GBRs may not offer the same level of versatility or flexibility as NNs, they can still provide a reliable and computationally efficient solution for basic vector-to-vector prediction tasks. As demonstrated in Section 4.8.4.1, both NN and GBR models can be simultaneously utilized in the calibrator design.

Although this approach turns the calibrator into a black box, whether implemented with an NN or GBR [31], the overall system retains a high degree of physical interpretability since it integrates the analytical PSF model into its architecture. Moreover, smaller networks are typically faster and easier to interpret, although this goes beyond the scope of this thesis.

Given these considerations, this hybrid approach was implemented. The exact details of the implementation differ between IRDIS and MUSE-NFM. However, the general principles for training the calibrator remain the same for both instruments, which will be discussed in the next subsection.

4.8.3. Training a data-driven calibrator

With the architecture of the calibrated PSF model defined and the nature of the datasets understood, the next step is to discuss the training process for the calibrated predictor. Two approaches can be proposed for this purpose.

Naïve approach

The most straightforward approach to training a calibrator involves reusing the data obtained via model fitting during the validation process ($\mathscr{D}_{\text{fit}}$). The calibrator can then be trained to map these fitted parameters $\widetilde{\mathscr{X}}$ to the corresponding reduced telemetry $\mathscr{Z}$. This allows the calibrator to be trained separately and only combined with the PSF model during inference.

However, as detailed in Kuznetsov et al. (2023), such a training approach produced inaccurate results. This inaccuracy likely stems from overfitting during the generation

31. Also, being an ensemble of simple decision trees, GBRs are inherently more interpretable than NNs.

of the dataset with fitted model inputs. As previously discussed in Section 3.2, non-linear optimization can lead to ambiguous estimates in cases where variables are coupled, resulting in overfitting if the optimization problem is not properly regularized or re-parameterized. This introduces a high variance in the fitted inputs, making it difficult for the calibrator to identify reliable dependencies when trained on this dataset. Nevertheless, as will be demonstrated in the following section, this training approach can still be employed in the limited framework for some distinct model inputs.

Training with PSF model in-loop

Another approach proposed in this work, as also discussed in Kuznetsov et al. (2023), involves training the calibrator in conjunction with the PSF model. In this training scheme, the calibrator learns to map reduced telemetry inputs directly to PSF morphology, bypassing the need for a dataset of pre-fitted model parameters. The model inputs then act as latent variables, with the calibrator acting as an encoder and the PSF model acting as a fully analytical decoder. This configuration, in some sense, also resembles a physics-informed neural network (PINN) (Raissi et al. (2019)).

During training, backpropagation is computed through both the calibrator and the PSF model, which is enabled by the PyTorch-based implementation of the TipTorch model. In this training scheme, the PSF model plays a critical role in regularizing the calibrator's training by injecting the underlying physics. Moreover, training the calibrator to map reduced telemetry inputs directly to the PSF shape rather than to model inputs ensures that the model's outputs are precisely aligned with the primary objective of accurately predicting PSF morphology. The illustration of this approach applied to IRDIS and MUSE-NFM will be given later in Figures 4.24 and 4.32.

This work adopts this training approach as the primary method, while the naïve approach is utilized in a more limited capacity, as demonstrated later. With the basic principles established, this approach can now be discussed in the context of the IRDIS and MUSE-NFM instruments more concretely.

Notably, having a trainable calibrator allows it to be fine-tuned and updated as new samples are added to the dataset, provided that a proper data-processing routine is in place to fetch and process data automatically. This means that this method can be utilized in the self-learning framework.

4.8.4. Final prediction results for SPHERE

4.8.4.1. Calibrator design and training

After describing the basic principles of calibrator training procedures, it is now possible to describe the practical applications of the proposed technique to the instruments of interest and quantify the improvement from using the calibrated prediction method. The diagram in Figure 4.22 outlines the architecture of the calibrated PSF model for IRDIS, which comprises the data-driven calibrator and the analytical model (TipTorch). The architecture of the calibrator, as well as its inputs and outputs are de-

picted in the boxes left to the "TipTorch" box. Note that the PSF model simultaneously predicts the "left" and "right" PSFs.

The set of predicted model inputs includes $\widetilde{\mathcal{X}} = \{\mathbf{F}, \Delta n, \mathbf{c}, J_x, J_y, W_{\text{dir}}\}$. Meanwhile, all contents of $\mathcal{I}$ are listed in Figure 4.22 on the left.

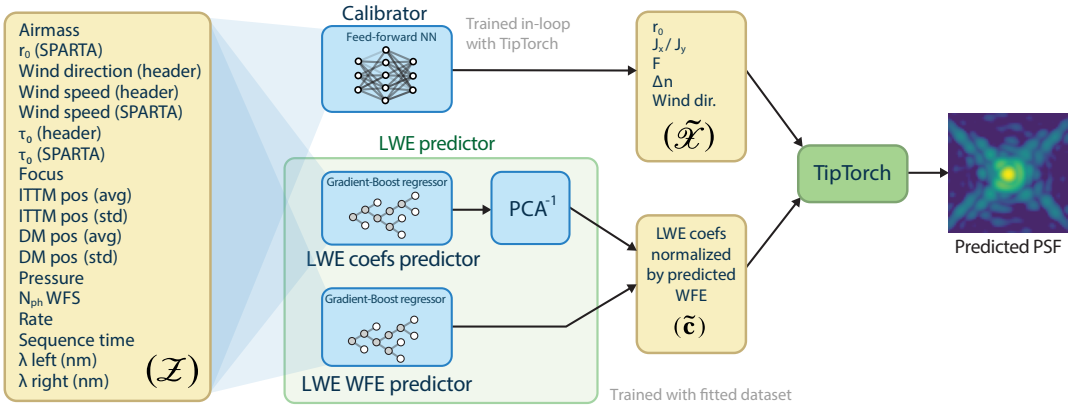


FIGURE 4.22. – Detailed schematics of the calibrated PSF model architecture for IRDIS. Note that the predictor is complemented with 2 models that exclusively infer LWE patterns and LWE-induced WFE.

In this setup, the calibrator is implemented as a NN, with the architecture detailed in the Appendix D. It was trained using the approach discussed earlier, meaning it was trained with the PSF model in-loop. However, it is important to note that the calibrator design incorporates two additional models trained to predict LWE exclusively. One of these models predicts the coefficients for a Principal Component Analysis (PCA) transformation, which are then converted into LWE modal coefficients in the basis described in Figure 4.15. These coefficients are then normalized by a scalar WFE value predicted by the second model. This dedicated handling of LWE is necessitated by its very pronounced effect on the PSF morphology, as discussed earlier in Section 4.6.3. Both predictors utilize the same set of reduced telemetry inputs and both are implemented as GBRs, while the primary calibrator is implemented using a NN. The LWE predictor was trained separately using fitted data (naïve approach). In contrast, the NN-calibrator was trained via backpropagation through the NN (see Section 4.8.3). During the NN training process, the LWE values are predicted first and inserted into the PSF model, and then the NN calibrator predicts the remaining inputs of the model.

However, it is also worth noting that the decision to divide the predictor into specialized components for LWE and the rest of the parameters was motivated by the fact that the initial attempts to predict LWE modes (PCA coefficients) and other integrated PSF parameters jointly using a monolithic NN calibrator were unsuccessful. The predicted LWE tended to some average constant values. This could be potentially fixed by employing more advanced regularization techniques during training. However,

the quickest and most straightforward solution was to train a separate GBR-based predictive model specifically for LWE coefficients. Although this approach produced accurate LWE coefficient patterns, the WFE values computed from these predictions were still inaccurate. To address this, another specialized model was introduced to predict LWE WFE values exclusively and use them to re-normalize the LWE wavefronts inferred by the first model.

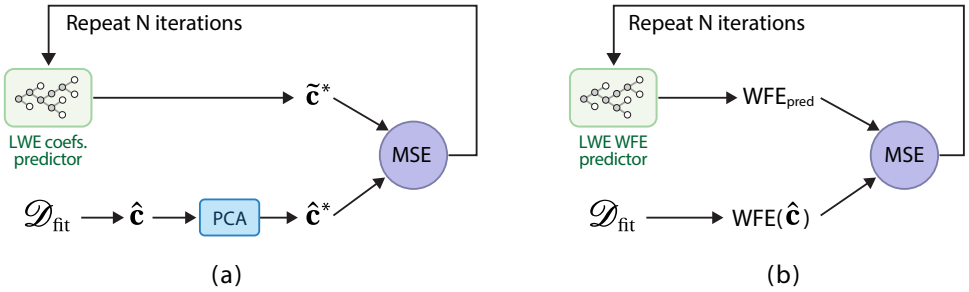


FIGURE 4.23. – First step : training the LWE-predictors based on the $\hat{\mathbf{c}}$ values extracted from the earlier obtained $\mathscr{D}_{\text{fit}}$ dataset. Model (a) predicts the coefficient values for the modes from Figure 4.15. Model (b) re-normalizes the prediction of the first model by the WFE values it infers. Both models are trained separately and before the calibrator-NN. Here, $\hat{\mathbf{c}}^*, \tilde{\mathbf{c}}^*$ are the PCA-compressed representation of ground truth and predicted LWE coefficients.

The loss function for training the LWE WFE and LWE coefficient predictors is a simple MSE, while the loss for the NN-driven calibrator is formulated as follows.

$$\mathscr{L}(\mathbf{I}(\widetilde{\mathscr{X}}), \mathbf{I}_{\text{data}}) = \text{MAE}(\mathbf{I}(\widetilde{\mathscr{X}}), \mathbf{I}_{\text{data}}).$$

Unlike in Eq.(4.11), this loss function has a much simpler formulation. This is because there is no need to regularize the prediction of LWE modes, as their inference is offloaded to separate predictors.

The training process of the calibrated PSF model for IRDIS is summarized in Figures 4.23 and 4.24.

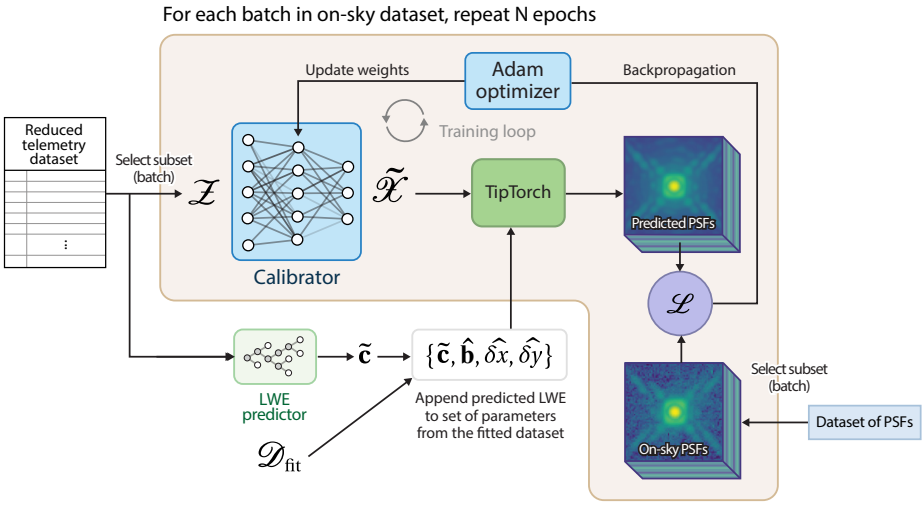


FIGURE 4.24. – Second step : training of the NN-calibrator. This time, the PSF model is included in the training loop. The LWE predictors have already been trained.

In theory, it should be possible to integrate the LWE predictors into a monolithic NN-based predictor. However, this combination of three specialized models proved more effective, serving as the first successful proof-of-concept. In addition, with this approach, the LWE predictor can serve as a separate dedicated tool.

4.8.4.2. Used dataset

Before delving into the calibrated prediction results, it is important to first outline the data used to train the calibrator and clarify which parameters are predicted versus those provided as inputs.

The on-sky PSFs are the most essential part of the dataset. More details on how this dataset was compiled were provided in Section 4.5.5.1. The total count of samples used for training and validation, as well as their count, are presented in the table in Section C.2.1. This table splits the IRDIS dataset into separate wavelength groups corresponding to their respective spectral filter. Note that only the first five groups were used for the training and validation due to the statistically representative amount of samples in them. The group corresponding to `D_J23` filter was ignored due to the generally low quality / low SNR of PSF samples it contains. In total, 3112 samples were selected. From them, 2517 samples were used for training, while 595 samples were used for validation. All prediction results provided later in this chapter are obtained using the validation data only.

As mentioned earlier, the set of the predicted model inputs $\widetilde{\mathcal{X}}$ includes :

$$\widetilde{\mathcal{X}} = \{r_0, \mathbf{F}, \Delta n, J_x, J_y, W_{\text{dir}}, \mathbf{c}\}.$$

Meanwhile, the reduced telemetry inputs set $\mathcal{Z}$ comprises the data presented in

Section C.2.2. The exact selected entries are listed in the leftmost box in Figure 4.22. However, most entries presented in that table were filtered out due to their low importance for the model's prediction accuracy. The significance of specific features was determined by excluding them from the input dataset. The calibrator was then retrained with the updated set of input features after the exclusion. As a result, only those inputs that significantly impacted the prediction accuracy after retraining were included in $\mathcal{I}$. It is essential to acknowledge that this feature selection process was conducted manually, whereas it could be carried out more quantitatively and accurately using various model sensitivity analysis techniques. This should be prioritized in the future to enhance the results further.

It is important to note that not all the data used during the training were taken from the on-sky reduced telemetry dataset. As depicted in Figure 4.22 the PSF flux background $\widehat{\mathbf{b}}$ and PSF shift $\widehat{\delta x}$, $\widehat{\delta y}$ were taken from the fitted values dataset $\mathscr{D}_{\text{fit}}$ (see Section 4.7.1.2).

This is because predicting the target coordinates, for instance, is unnecessary, given the on-axis nature of the PSF data used in this case. However, it is still helpful to account for the precise sub-pixel shift of the PSFs ($\widehat{\delta x}$, $\widehat{\delta y}$) relative to the center of the FoV to minimize the discrepancy between the data and the predicted PSFs during the NN-calibrator training. Consequently, the shifts can be assigned to fitted values obtained in the previous stage of PSF model validation. Since $\widehat{\delta x}$ and $\widehat{\delta y}$ couple weakly with the other fitted parameters, they are less susceptible to overfitting and can thus be used [32].

The same principle applies to $\widehat{\mathbf{b}}$. In reality, this parameter can depend on the sky background, weather conditions, and detector settings. However, in this work, it is treated as a random variable. Thus, its value can also be derived from $\mathscr{D}_{\text{fit}}$.

As shown in Figure 4.23, the LWE predictor is trained separately using the fitted values $hat\mathbf{c}$. It is important to emphasize that there is no ground truth (GT) for LWE coefficient values available in actual on-sky data. Therefore, the only means to obtain GT for LWE coefficients is to use the values from $\mathscr{D}_{\text{fit}}$, as training the LWE predictor alongside the main NN-calibrator was unsuccessful. Thanks to the advanced regularization of LWE coefficients during fitting (see Eq.(4.11), the coupling of $\hat{\mathbf{c}}$ with other PSF model parameters has been minimized, allowing these parameters to be used independently from the other PSF inputs.

4.8.4.3. Calibrated prediction results

Tuned direct prediction

Before diving into the calibrated predictions that involve the intricate ensemble of models described earlier, a rudimentary approach should be tested first. This method essentially involves assigning constant biases and values to some model inputs while retrieving other parameters directly from configuration files and the preserved telemetry data without any modifications. The "constant" means that the

32. In fact, tip-tilt modes of LWE can absorb the effect of PSF shift. To eliminate this coupling, the pupil-wide tip-tilt components were subtracted from fitted LWE WF screens and added to $\widehat{\delta x}$ and $\widehat{\delta y}$.

assumed values stay the same for all predicted PSFs. This simplest form of calibration is later referred to as *"tuned direct prediction"*. The results obtained from this method are shown in Figure 4.25. The algorithm used to calculate the radial profiles in this figure is the same as previously detailed in Appendix E.

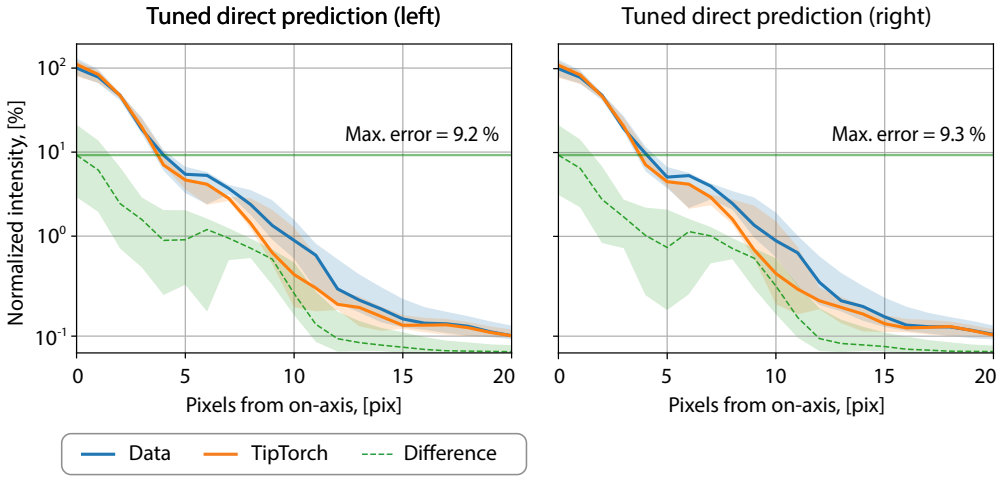


FIGURE 4.25. – The results of "tuned" direct prediction for "left" and "right" IRDIS PSFs. The main difference with the pure direct prediction is that here, some model inputs are assigned to constant values that, on average, give a smaller discrepancy of predicted PSF with on-sky data.

The assumptions made for the model inputs are $\mathbf{F} = \mathbf{1}$, $J_x, J_y = 33$ mas, $J_{xy} = 0$ deg, $\Delta n = 0$ rad^2, $\mathbf{c} = \mathbf{0}$. These values were obtained by simultaneously fitting the same values of J_x, J_y for several on-sky samples. As seen in the left plot, even this simple method significantly improves the results demonstrated earlier in Section 4.6.1.

Calibrated prediction

After trying the simple approach, the next step is to apply the method proposed in the previous section. The results of this test are shown in Figure 4.26. These results were obtained using the validation dataset, which was compiled by selecting ~20% from all the available samples, as mentioned previously.

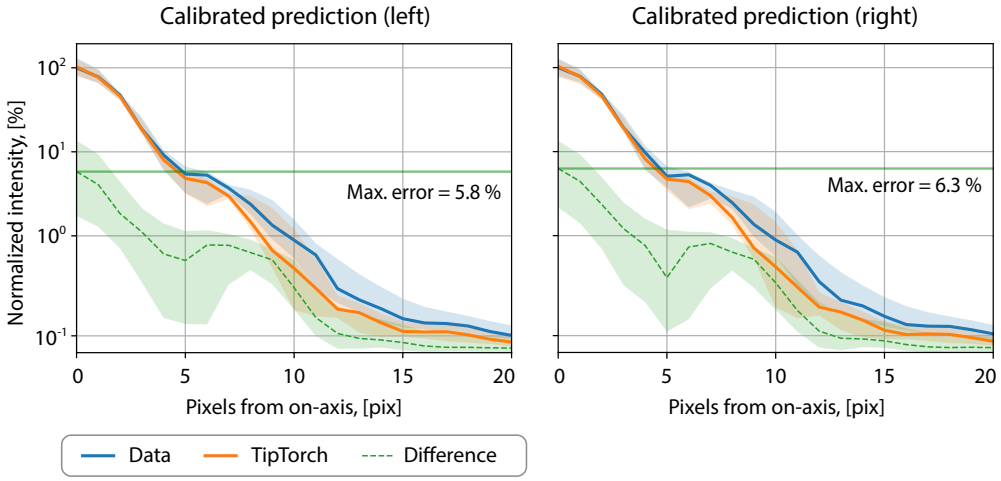
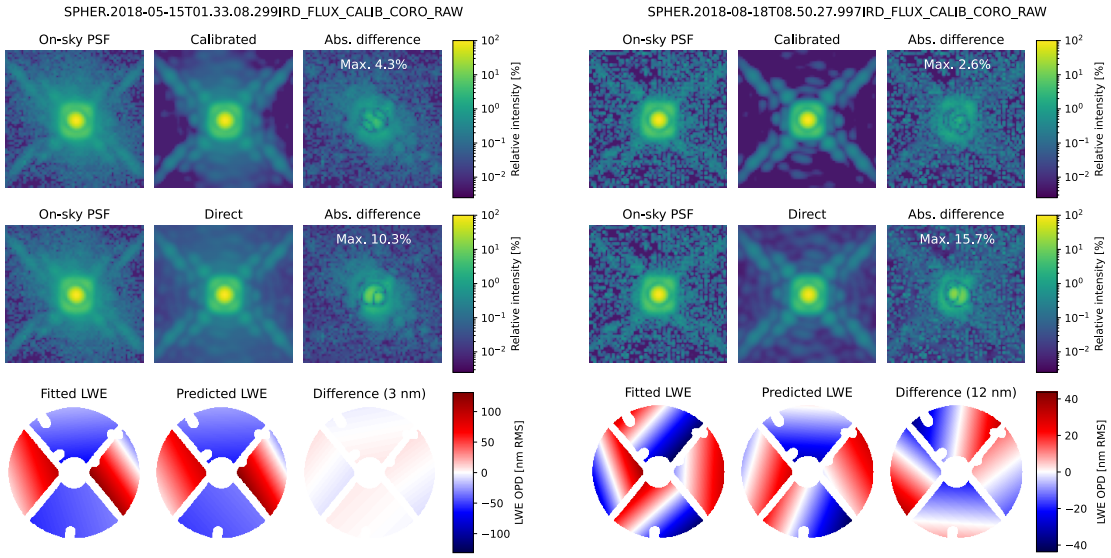


FIGURE 4.26. – Calibrated prediction results were obtained using the techniques outlined in Figure 4.22. This approach demonstrates the highest prediction accuracy.

These results demonstrate an improvement over those presented earlier in Section 4.6.1. However, when comparing the radial profiles of tuned direct and calibrated predictions, the latter may not seem to offer a substantial advantage, particularly given the complexity of data retrieval and training. Nevertheless, the true advantage of the calibration approach becomes evident when examining specific predicted PSFs individually. While radial profiles are useful for evaluating how accurately the PSF's energy distribution and peak value are predicted, rendering the PSF images provides insight into how well the morphology is captured. This is demonstrated in Figure 4.29.



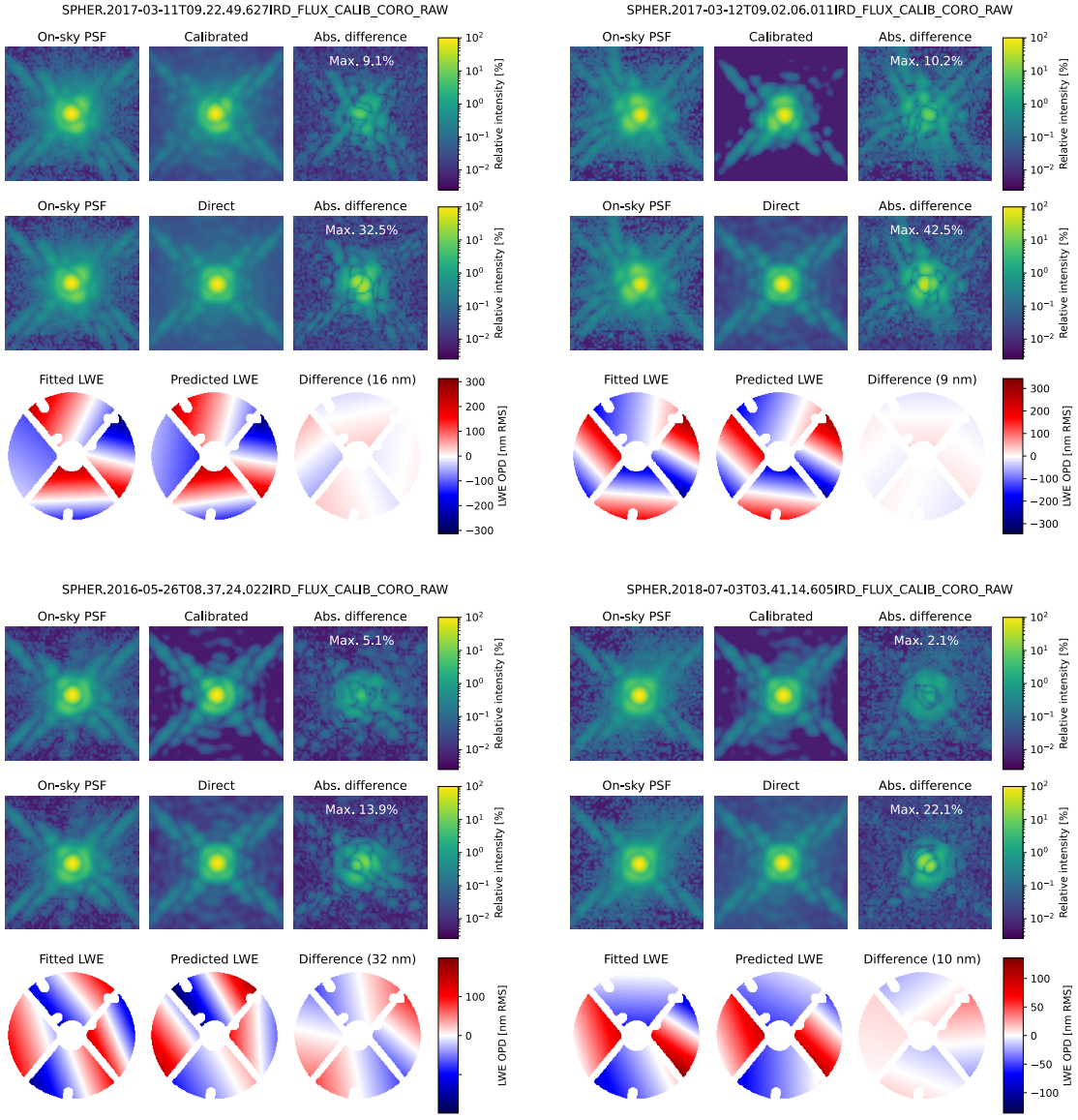


FIGURE 4.29. – Examples of predicted IRDIS PSFs. The top row shows direct calibrated predictions, the middle row presents tuned direct predictions, and the bottom row compares fitted (GT) and predicted LWE wavefronts. GT LWE values are taken from $\hat{\mathbf{c}} \in \mathscr{D}_{\text{fit}}$.

Note that in these figures, the subtitle "direct", in fact, refers to *tuned* direct prediction. The first word was omitted for the sake of brevity.

The results presented above can be further supported by the fraction of variance unexplained (FVU) metric [33] presented in Table 4.1.

33. FVU = 0 means a model that perfectly explains the data; FVU = 100% means that the model does not explain the data at all, so the residual variance equals the variance of the original data; FVU > 100% indicates the poor model's performance and overfitting.

	Tuned direct	Calibrated
FVU, [%]	4.5	0.7

TABLEAU 4.1. – Median FVU computed for the PSFs from the IRDIS validation dataset.

The FVU error for a single pair of data ($\mathbf{I}_{\text{data},k}$) and predicted ($\mathbf{I}(\widetilde{\mathscr{X}_k})$) PSF images can be defined as :

$$\text{FVU}_k = \frac{\text{var}(\mathbf{I}(\widetilde{\mathscr{X}_k}) - \mathbf{I}_{\text{data},k})}{\text{var}(\mathbf{I}_{\text{data},k})} \cdot 100\% \tag{4.17}$$

Here, the metric FVU_k is computed for every k-th sample in the validation dataset; var(·) is the variance computed for pixels of PSF images averaged over the wavelengths. The FVU values presented in Table 4.1 are computed as the median of all FVU_k values.

While the SR error in Figure 4.26 is a more relevant metric for predicting errors in the SPHERE instrument, evaluating FWHM prediction accuracy also provides valuable insight. The corresponding results are shown in Figure 4.30 :

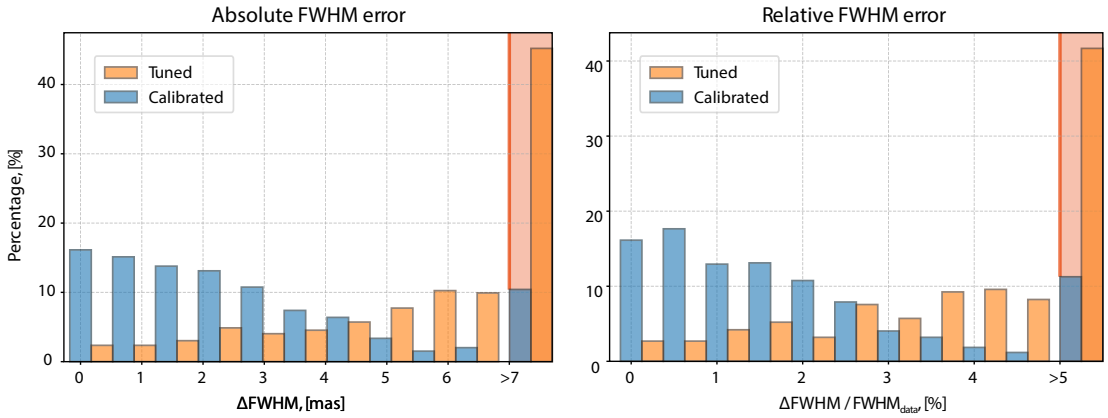


FIGURE 4.30. – FWHM prediction error for IRDIS PSFs. The left plot is the absolute, and the right is the relative error. All values above the threshold, both for absolute and relative errors, are compressed into a single bar.

In this figure, FWHM was computed for the validation dataset using the following expression :

$$\begin{aligned} \text{FWHM} &= \frac{1}{N_\lambda} \sum_\lambda \sqrt{\text{FWHM}_{\lambda,x}^2 + \text{FWHM}_{\lambda,y}^2}, \\ \Delta\text{FWHM} &= \left| \text{FWHM}_{\text{pred}} - \text{FWHM}_{\text{data}} \right|. \end{aligned} \tag{4.18}$$

Here, N_λ is the number of spectral slices; in this case, $N_\lambda = 2$ (for left and right PSFs), $\text{FWHM}_{\lambda,x}$ and $\text{FWHM}_{\lambda,x}$ were obtained by fitting a 2D Moffat [34] function to 1D slices of on-sky $\mathbf{I}_{\text{data}}$ and predicted PSFs $\mathbf{I}(\mathscr{X})$ along the x and y axes respectively. Fitting was done for each λ separately, and the final white FWHM was then computed by

34. See `astropy.modeling.models.Moffat2D`.

averaging over wavelengths; ΔFWHM is the absolute FWHM prediction error, and $\Delta\text{FWHM} / \text{FWHM}_{\text{data}} \cdot 100\%$ is the relative prediction error.

The figure shows the FWHM error distributions for both tuned and calibrated cases. The error values corresponding to calibrated prediction tend to group closer to zero, indicating improved accuracy compared to the tuned approach.

To summarize, the results above show that the ML-driven calibrator leads to overall better prediction accuracy than the tuned approach. However, the most significant advancement of the calibrated approach is its ability to predict LWE patterns using only reduced telemetry entries. This result is particularly notable, given the absence of robust analytical models that relate reduced telemetry to quasi-static LWE wavefronts. In this context, the LWE predictor, even when used independently of the PSF model, serves as a novel and valuable tool.

However, it should also be mentioned that the current LWE predictor relies on telemetry data collected after the observation, as entries derived from AO telemetry were utilized. Consequently, this tool cannot yet forecast the LWE for forthcoming observations solely based on weather prognoses.

4.8.5. Final prediction results for MUSE

The calibrated prediction results for IRDIS showed promising outcomes. Applying the same approach to MUSE-NFM is essential to validate these findings further. This will help demonstrate the universality of the proposed method. This requires discussing the aspects as before but in the context of MUSE-NFM.

4.8.5.1. Calibrator design and training

The MUSE-NFM calibrator has a simpler architecture compared to IRDIS. Since no LWE features were identified in the MUSE-NFM dataset, specialized LWE predictors are unnecessary. As a result, the calibrator consists solely of NN. The architecture is shown in Figure 4.31. The contents of $\widetilde{\mathscr{X}}$ are shown in Figure 4.31. All entries comprising $\mathscr{I}$ are listed in Appendix C.3.1. This will be discussed later in Section 4.8.5.2.

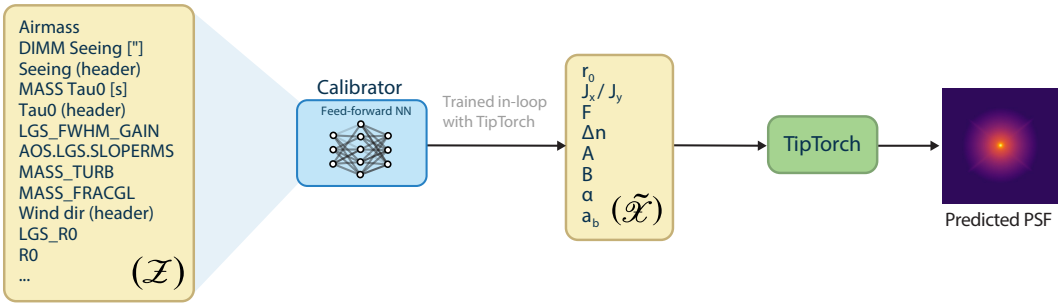


FIGURE 4.31. – The design of the calibrated predictor for MUSE-NFM. Compared to the approach from Figure 4.24, it has a much simpler design due to the absence of LWE in MUSE-NFM data.

The training procedure can then be summarized in Figure 4.32.

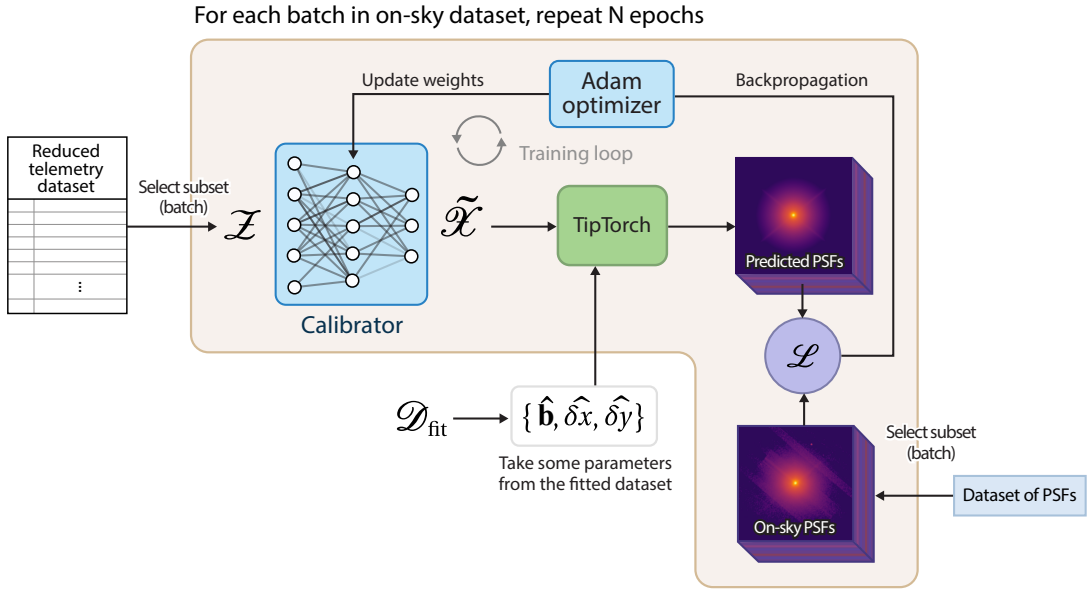


FIGURE 4.32. – Scheme of training the NN-calibrator for MUSE-NFM. In contrast to IRDIS, the predictor is fully trained with the PSF model in the loop this time.

It is important to note that the PSF model is configured to predict PSFs at 7 wavelengths $\lambda = \{494, 560, 655, 721, 787, 853, 919\}$ nm, while each processed PSF cube in the dataset comprises 13 spectral bins. The choice of this wavelength subset showed the best compromise between the number of predicted wavelengths, computational resources, and prediction accuracy. Interestingly, training an ensemble of predictors, with one predictor per spectral bin, resulted in lower overall prediction accuracy. This outcome can be attributed to the fact that the use of multiple spectral slices during training helps to regularize the model in the same fashion as discussed in Section 4.7.2.2

in the scope of the fitting. Since some model inputs (like r_0) influence PSF predictions across multiple wavelengths, collecting gradients from multiple spectral slices and training samples allows for more stable and accurate updates of trainable parameters.

Then the loss function $\mathscr{L}(\cdot)$ employed to train the calibrator model is formulated as :

$$\mathscr{L}(\mathbf{I}(\mathscr{X}), \mathbf{I}_{\text{data}}) = \alpha\,\text{MSE}(\mathbf{I}(\mathscr{X}), \mathbf{I}_{\text{data}}, N) + \beta\,\text{MAE}(\mathbf{I}(\mathscr{X}), \mathbf{I}_{\text{data}}, N) \tag{4.19}$$

Just like Eq.(4.14), this loss function combines the MAE and MSE components. However, the definition of MSE and MAE losses are slightly different in this case compared to Eqs.(4.20) and (4.21) :

$$\text{MAE}(\mathbf{I}(\mathscr{X}), \mathbf{I}_{\text{data}}, N) = \sum_N \frac{1}{HW} \sum_\lambda \sum_{i=1}^{H} \sum_{j=1}^{W} w(\lambda) \left| I_{ij\lambda}(\mathscr{X}) - I_{ij\lambda,\text{data}} \right|, \tag{4.20}$$

$$\text{MSE}(\mathbf{I}(\mathscr{X}), \mathbf{I}_{\text{data}}, N) = \sum_N \frac{1}{HW} \sum_\lambda \sum_{i=1}^{H} \sum_{j=1}^{W} w(\lambda)^2 \left(I_{ij\lambda}(\mathscr{X}) - I_{ij\lambda,\text{data}} \right)^2, \tag{4.21}$$

where $N = 16$ is the number of samples in the training batch. The values for α and β here also differ from Eq.(4.14). In this case, assuming $\alpha = \beta = 1$ was enough to achieve a balanced weighting of both loss terms. As discussed in Section 4.7.2.1, the Moffat PSD absorber was incorporated into the TipTorch model for MUSE-NFM to enhance fitting accuracy. The NN calibrator was then trained to predict the parameters of this absorber.

4.8.5.2. Used dataset

Before proceeding with the prediction results, briefly discussing the dataset used for the training is essential. After filtering all the corrupted data samples for MUSE-NFM, only 382 total PSF cubes with associated reduced telemetry remained, leaving only 304 data samples available for training and 78 for validation. Just like for IRDIS, all prediction results presented below in this section are obtained only for validation samples that were not seen by the calibrator during the training.

The set of predicted model inpts $\widetilde{\mathscr{X}}$ in the case of MUSE-NFM contains the following entries :

$$\widetilde{\mathscr{X}} = \{\mathbf{F}, r_0, \mathbf{J_x}, \mathbf{J_y}, \Delta n, A, B, \alpha_M, a_b\}.$$

The meaning behind each element of this set is discussed in Eq.(4.10) and Section 4.7.2.1. The dimensionality and wavelength dependency of these entries are discussed earlier in Eq.(4.13).

Meanwhile, the set of reduced telemetry inputs $\mathscr{Z}$ associated with every PSF is described in Section C.3.1. In essence, all dataset entries defined in the table of this section were used as calibrator inputs. Unlike in the case of IRDIS, where telemetry inputs were selected based on their importance, the opposite approach was taken here. All entries that could be fetched and associated with PSF were utilized as input features to the calibrator model. The advantage of such an approach is that including

more features can help to discover hidden or unexpected relationships in the data. However, the disadvantage is that too many potentially irrelevant features can lead to overfitting. When too many inputs are added, some might only correlate with the target by chance. The model can thus learn these spurious patterns, which do not generalize. This is especially crucial given the scarcity of training data. Conversely, after excluding too many inputs, the model might underfit and miss meaningful relationships, thus failing to capture the true complexity of the data. Therefore, in an ideal scenario, the inclusion of input features must be carefully balanced based on the model's sensitivity analysis. For now, the existing solution is functional but not ideal and must be improved.

Also note, that the values of $\widehat{\mathbf{b}}, \widehat{\delta x}$, and $\widehat{\delta y}$ were taken from the fitted values dataset $\mathscr{D}_{\text{fit}}$ (see Figure 4.32) in a similar fashion as outlined earlier in Section 4.8.4.2 for IRDIS.

4.8.5.3. Calibrated prediction results

Simple tuned calibration

Following the example of the tuned direct prediction for IRDIS shown in Figure 4.25, the analysis of the calibrated predicted approach for MUSE-NFM begins by testing the tuned direct prediction method (Figure 4.33).

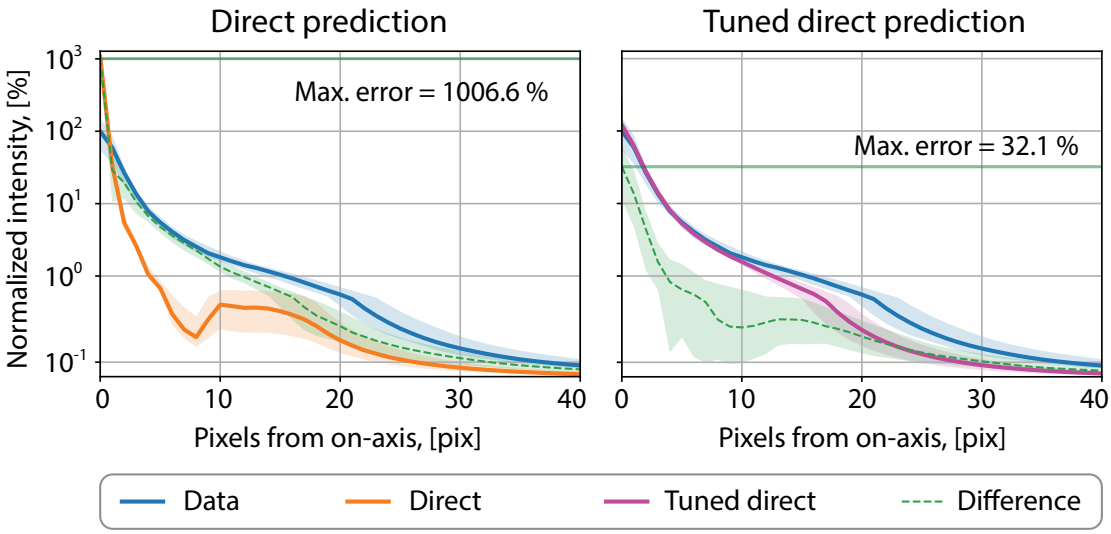


FIGURE 4.33. – Left figure recaps the profiles obtained for the direct prediction demonstrated earlier, right figure demonstrates the results obtained using tuned direct prediction. The profiles shown in this figure are computed for "white" PSFs in the same fashion as earlier in Fig. 4.19.

In this figure, the tuned direct prediction (right profile) is compared with the pure direct prediction results (left profile) presented earlier in Figure 4.14. Similar to the case of IRDIS, a simple technique like tuning can already significantly improve the results. It is implemented by setting specific model inputs to constant empirical values,

which, on average, leads to a smaller discrepancy between the model predictions and the data. More details can be found in Appendix F.2.

NN-based calibrator

The next logical step is to test the approach with the NN-based calibrator. These results are presented in Figure 4.34.

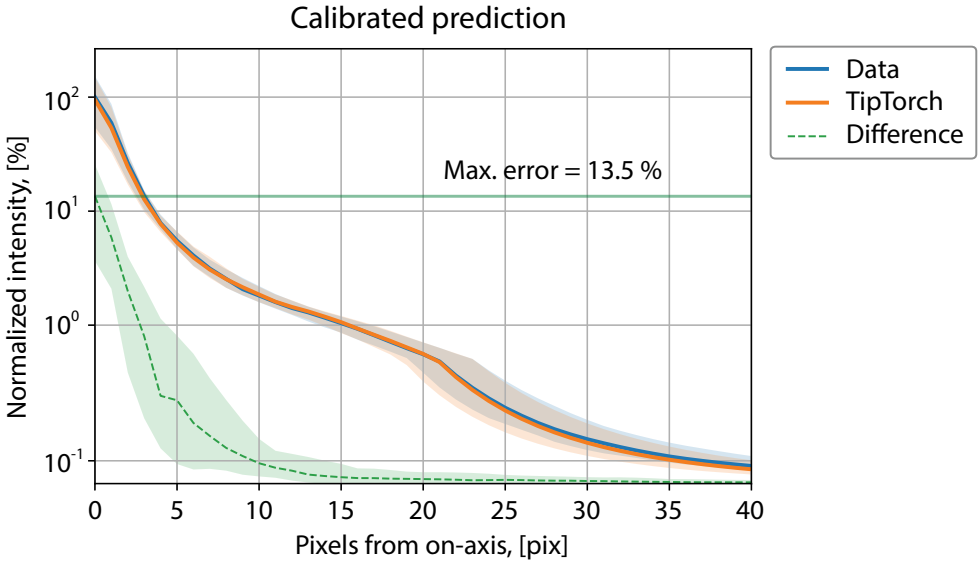


FIGURE 4.34. – The results of the calibrated prediction for the dataset of MUSE-NFM PSFs. Similar to the case of IRDIS, the calibrated model again demonstrates the best prediction accuracy.

Compared to the results previously presented in Section 4.6.2, these profiles demonstrate a significant improvement in prediction accuracy. However, the peak value of the PSF is predicted with an error that still exceeds 10%. This discrepancy may be partially attributed to the difficulty of accurately predicting the strength of the phase bump defect mentioned in Section 4.7.2.1 and Figure 4.18.

Although the level of accuracy obtained using calibrated prediction is not yet sufficient for PSF deconvolution applications, the profiles indicate that the calibrator still successfully predicted the overall shape of the AO-corrected PSF. This is further supported by examining images of individual PSFs from the validation set, as illustrated in Figure 4.38.

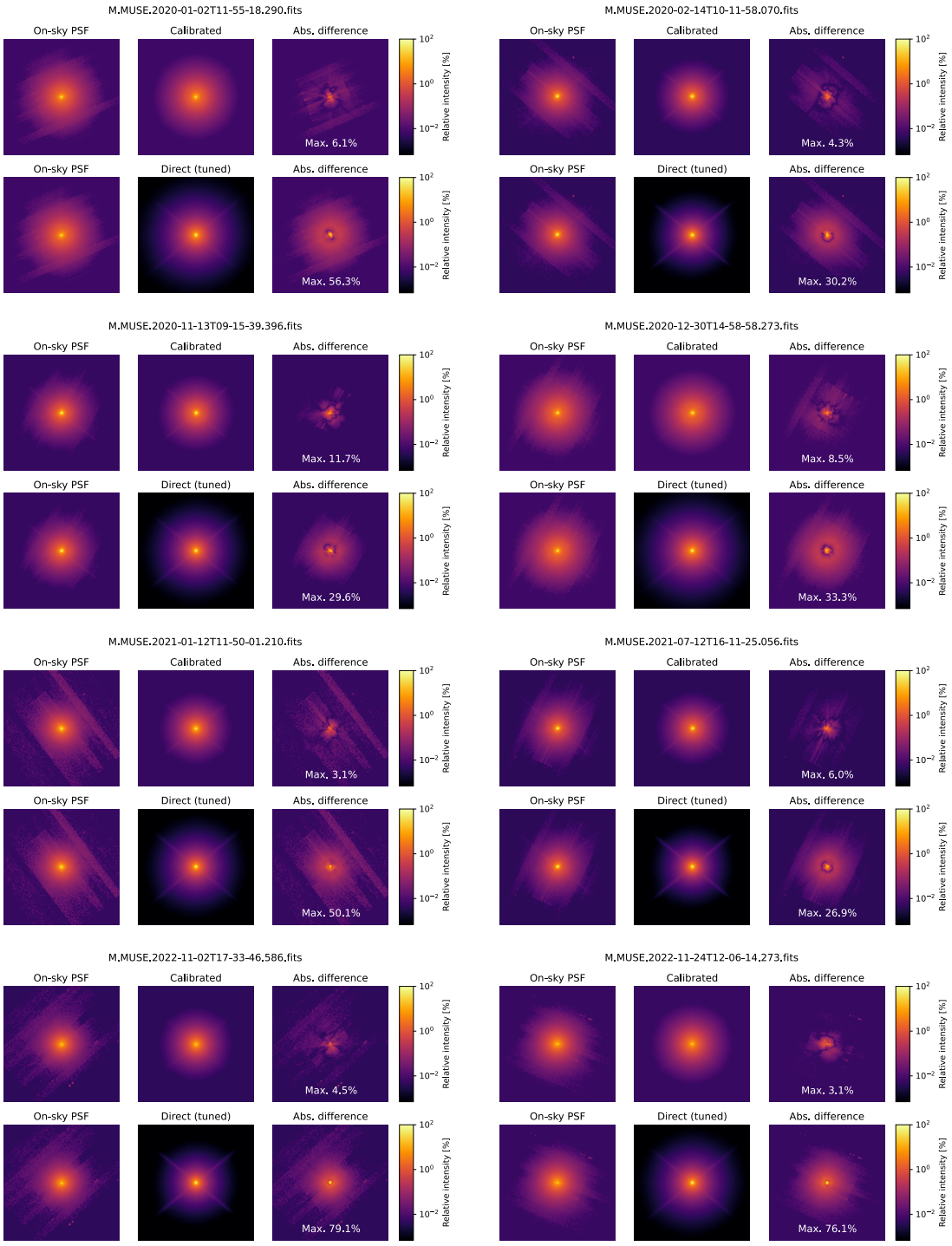


FIGURE 4.38. – Examples of predicted MUSE PSFs from the validation set. The top row shows calibrated predictions, while the bottom row presents tuned direct predictions.

Note that similar to the earlier case for IRDIS, in these figures, the label "Direct prediction", in fact, refers to "*Tuned* direct prediction" since it is futile to compare the

NN-calibrator with the results from Section 4.6.2 because of the highly inaccurate PSF predictions produced with the pure direct approach.

The accuracy of calibrated prediction can be further supported by the FVU metric computed for MUSE-NFM PSFs in the same fashion as described earlier in Eq.(4.17). The results are presented in Table 4.2 :

	Tuned direct	Calibrated
FVU, [%]	17.4	3.2

TABLEAU 4.2. – Median FVU computed for the PSFs from the MUSE validation dataset.

However, when talking about MUSE-NFM, it is very important to consider the accuracy of FWHM prediction since it is directly connected to science applications discussed earlier in Section 4.1. The results of such tests are presented in Figure 4.39.

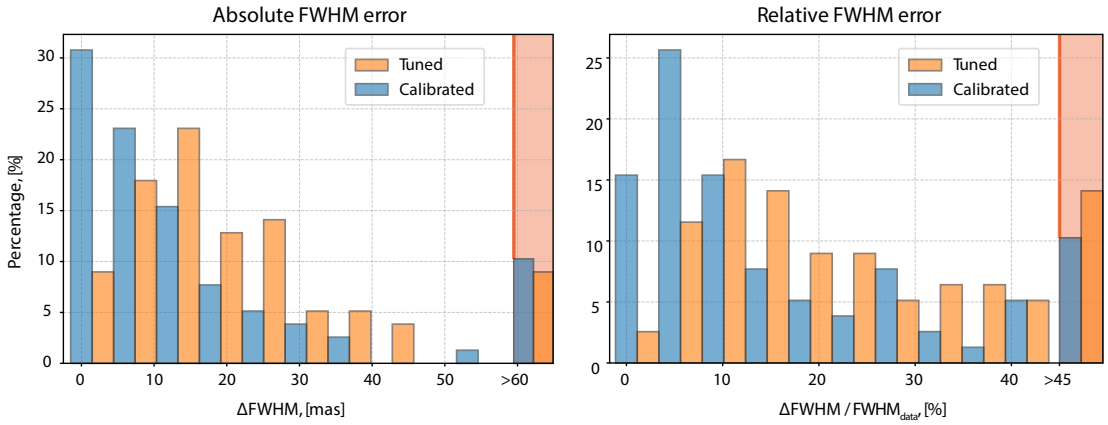


FIGURE 4.39. – FWHM prediction error for "white" MUSE-NFM PSFs. The left plot is the absolute, and the right is the relative error. All values above the threshold, both for absolute and relative errors, are compressed into a single bar.

In this figure, FWHM was computed the same way as for IRDIS using Eq.(4.18) with the only difference being $N_\lambda = 7$.

Like in the case of SPHERE, the FWHM error corresponding to calibrated prediction demonstrates smaller values, again indicating improved accuracy compared to the tuned approach.

4.8.6. Conclusion

Based on the results obtained using different prediction approaches, it can be concluded that the PSF model calibrated with data-driven techniques offers higher prediction accuracy compared to the model relying on direct prediction or using constant values and biases (tuned direct prediction). The results are valid for both

SPHERE and MUSE instruments. The primary advantage of the calibrated model lies in its ability to realistically predict WFS noise, TT jitter, and peak intensity. For IRDIS, this advantage is further enhanced by the capacity of the predictor to accurately model the LWE pattern based on the reduced telemetry inputs in the case of SPHERE. The calibrated approach generally provides a more accurate prediction of the overall energy distribution across the PSF, which is reflected in lower FWHM error. For example, in the case of MUSE-NFM, approximately 56% of the predicted FWHM error remains within 10%, underscoring the effectiveness of the proposed calibrated hybrid predictive approach. According to Section 4.1, this level of predictive accuracy is already enough to use the model with various scientific applications. Consequently, the results presented in this chapter highlight the practical importance of the proposed PSF prediction technique for post-processing scientific data.

Notably, the distributions of predicted model inputs $\widetilde{\mathscr{X}}$ do not exhibit nonphysical values. These distributions for the IRDIS and MUSE-NFM instruments are discussed in Sections G.1 and G.2. This suggested approach did not lead to overfitting, reinforcing the general physical validity of the proposed method.

5. Conclusion and Future Perspectives

In conclusion, this thesis comprehensively explores two primary applications of PSFs in AO : focal-plane WFSing and PSF reconstruction (prediction). It establishes connections between these distinct yet interrelated research areas, addressing them within a partially shared theoretical and practical framework. To provide a detailed summary of each part, the conclusions and future perspectives are organized into separate sections, allowing for a more structured and focused discussion of the findings and implications of each part of the thesis.

5.1. Focal-Plane wavefront sensing

5.1.1. Conclusions

To further support the main conclusions presented in Chapter 3 and in the embedded paper, several additional points need to be elaborated.

Validated on-sky performance

In summary, the PSF-based WFSing approach detailed in Chapter 3 demonstrated its effectiveness for the detection of focus shifts caused by sodium layer drift, for LO wavefront sensing in general, and for the calibration of NCPA. The low-flux performance of defocus sensing using the LIFT technique was shown to be robust, but the highest accuracy was achieved when utilizing a preliminary NCPA calibration step in a two-stage approach. This approach, which integrates LIFT and DIP, has been validated through on-sky testing and has proven feasible for real-world applications.

Although using phase diversity for NCPA calibration has already been extensively studied in the past, applying LIFT within the MUSE-NFM framework is a novel contribution of this work. Moreover, this research represents one of the first on-sky implementations of LIFT and LIFT-like methods such as DIP.

The efficacy of the NCPA measurement technique was further supported by subsequent measurements conducted after publishing the paper. Figure 4.18 presents a "phase bump" observed using AOF WFSs, which, when compared to the earlier measurements derived using the focal plane technique (Fig.6 in the embedded paper), confirms the focal-plane method's ability to accurately retrieve the morphological characteristics of this defect using a sequence of PSFs using a bright on-sky target.

Minimal invasiveness

A key advantage of the proposed methodology is that both defocus drift sensing and NCPA calibration are performed directly on-sky using very similar approaches. Consequently, NCPA calibration can be conducted in a matter of minutes, prior to observations, and after closing the AO loop.

Flexible and accessible implementation

Both the LIFT [1] and the DIP [2] techniques, which constitute the two-stage approach, can function independently as standalone tools. The software I developed for this project is freely available online and can be customized to accommodate various systems or other phase diversity-related problems.

To fully integrate these techniques into real operations, minimal development effort would be required, and given this approach's low invasiveness, no engineering modifications [3] to the AOF are necessary for its implementation.

In summary, the successful on-sky validation, together with the availability of the developed code, demonstrates that the focal-plane WFSing method explored in this research is a practical and robust solution suitable for AO calibration and low-order WFSing in real-life applications.

5.1.2. Future perspectives

Need for additional on-sky experiments

Additional on-sky tests are essential to further support the findings presented in the embedded paper. Although the data analyzed in the paper were collected over different periods and yielded consistent results, the number of the on-sky experiments conducted remains scarce.

This is particularly critical for the calibrated low-flux tests discussed in Section 6.3 of the paper, where synthetic noise was introduced. Conducting true low-flux experiments will provide a more comprehensive validation of the technique and reinforce the findings outlined in the paper.

Application of FP WFSing for LWE and co-phasing

A promising direction for future development is the extension of this focal plane WFSing technique to address co-phasing and LWE sensing.

Of course, the application of focal plane-based methods for co-phasing has already been explored in multiple sources (Paxman & Fienup (1988); Lamb et al. (2018) and

1. https://github.com/EjjeSynho/LIFT
2. https://github.com/EjjeSynho/DIP
3. The astigmatic lens and the full-pupil mode necessary for LIFT operations had already been implemented during the IRLOS upgrade activity.

many more), while the potential of LIFT for LWE sensing has recently been demonstrated in simulations in Agapito et al. (2022).

Nevertheless, it is worthwhile to explore the performance of DIP in such applications, especially under realistic closed-to-on-sky simulation conditions. Research on co-phasing could yield useful insights, particularly in preparation for the next-generation ELT.

Preliminary work in this direction has already been initiated through the efforts of Titouan Delforge, a research intern at the Laboratoire d'Astrophysique de Marseille (LAM), whom I supervised. His work focused on applying LIFT and DIP to estimate LWE-induced aberrations from focal-plane PSFs applied to a MUSE-NFM-like system with IRLOS WFS (in simulations). His results further proved the ability of LIFT and DIP to provide high-accuracy estimates in this framework. While implementing his project, he heavily relied on the code for LIFT and DIP that I developed during this PhD project.

Summary and future directions

In summary, to advance this research further, the following directions can be taken :

— Acquisition of additional on-sky data, particularly under low-flux conditions, is needed to validate the technique comprehensively.

— Extension to co-phasing / LWE sensing, incorporating LIFT and DIP approaches for the ELT-scale simulations.

These steps will solidify the robustness of the proposed methods and potentially extend their applicability to the next generation of astronomical instruments.

5.2. PSF prediction

5.2.1. Conclusion

This research introduces a novel hybrid approach to PSF prediction, leveraging an analytical PSF model combined with a physics-informed neural network (PINN) architecture. This method synergizes the flexibility of neural networks with the interpretability of physics-based models. The compact nature of the proposed architecture not only allows it to be trained with limited on-sky datasets but also facilitates rapid inference times.

Universality

The proposed technique demonstrated robust performance across two distinct astronomical instruments with different AO system principles : SCAO for the SPHERE instrument and LTAO for the MUSE-NFM. This indicates the versatility and broad applicability of the method.

While the data-driven nature of this approach necessitates instrument-specific data processing and aggregation routines, as well as independent training of the data-driven calibrator for each new instrument, the overall PSF model architecture and the NN-calibrator framework can be largely maintained with only minor modifications. Thus, the proposed model can be adapted for various instruments with relatively few adjustments.

Accuracy

Currently, the prediction accuracy is below 10% for IRDIS and above 10% for MUSE-NFM, making the predicted PSFs unsuitable for direct use in high-precision applications such as image deconvolution. However, predictions of such a model can serve as an initial guess for deconvolution methods similar to the one described in Lau et al. (2023).

Moreover, in Section 4.7, it was shown that the model could accurately fit on-sky data for both SPHERE and MUSE, meaning that the model can accurately describe the on-sky data.

The accuracy of the proposed techniques is expected to improve as more diverse datasets become available for training. To date, the training, particularly for MUSE, has been performed using an extremely limited number of samples.

LWE Prediction

This work also introduces the first-ever predictor of LWE-induced wavefront aberrations based solely on reduced telemetry data. The LWE predictor can also operate as a standalone tool, providing reliable LWE pattern predictions without requiring the NN calibrator.

It is essential to clarify that the term "prediction" in this context does not imply forecasting future states. Due to the nature of its input data, this solution predicts the LWE pattern only retrospectively, i.e., for the completed observations. Turning this predictor into the forecasting framework would require modification of its input space.

However, such a tool could already be used in several applications. For example, it can provide initial guesses for some focal-plane-based LWE estimation techniques.

5.2.2. Future perspectives

The future development of this approach can be structured into three main directions :

— Enhancing the dataset quality and quantity.

— Improving the PSF model.

— Adapting the approach to new instruments.

Each direction is discussed in more detail below.

Improving data

The following actions are recommended to improve the dataset :
— Augmenting the current on-sky datasets with newly acquired samples.
— Developing more advanced data processing techniques with more advanced filtering techniques.
— Calculating error bars for all telemetry entries.

The last point is particularly crucial, as the accuracy of PSF prediction is most likely constrained by the quality of the input data. Quantifying these uncertainties would enable better handling of data imperfections, improve model regularization during training, and ultimately improve prediction accuracy.

Model Improvement

To effectively incorporate the uncertainty information mentioned above, the PSF model itself needs to be adapted to manage and propagate these uncertainties through both the estimation and training processes. Due to rapid inference times, the TipTorch model [4] can simulate hundreds of PSFs per second, making it feasible to sample a broad distribution of potential PSF outputs and reflect the data variability more comprehensively.

Multi-source fitting

In addition to its application in the predictive framework, TipTorch has proven to be a powerful tool for fitting scientific PSFs when they are present within the field of view. In this context, TipTorch can simultaneously fit multiple sources in the field. The resulting fit can then be extrapolated to cover the entire science field, facilitating efficient retrieval of off-axis PSF variations and enabling a more accurate representation of the PSF behavior across the full FOV.

Extending off-axis

To make this PSF model applicable for advanced science post-processing tasks such as deconvolution, extending it to the off-axis prediction is necessary. This expansion will require modifying the underlying theoretical framework, particularly the formalism presented in Section 4.3, which currently only addresses on-axis cases. In addition, the nature of the training datasets will need to change. Instead of using bright on-axis PSFs, the model might need to be trained on images of moderately crowded stellar fields with corresponding telemetry data. This idea also complements the previous point about using TipTorch to fit multiple science sources simultaneously across the field.

4. https://github.com/EjjeSynho/TipTorch

More efficient utilization of WFSing data

Both instruments considered in this work, SPHERE IRDIS and MUSE-NFM, offer an extensive dataset of PSFs recorded by LO WFSs, such as DTTS and IRLOS. However, this auxiliary WFS image data has not been fully exploited to its full potential in this project. These PSFs contain valuable information that could be further explored. For instance, they can be used to extract NCPAs (similarly to how IRLOS was used in Chapter 3) and provide this information as input for the TipTorch model. WFS PSFs can be used to retrieve information about LO modes, allowing estimation of the behavior of the lowest spatial frequencies in the residuals PSD. This concept can be generalized and extended further to predict scientific PSFs based on the WFSing PSFs morphology, complementing the complex model of the relationship between the WFSing data and science image quality.

Turning into a forecasting tool

The current model is designed for post-processing applications and cannot forecast AO performance for upcoming observations, as it depends on reduced telemetry entries that are only available post-observation. Converting the model into a true forecasting framework would necessitate modifying its input structure, potentially eliminating certain telemetry entries and developing predictive models for others. This would enable the development of a PSF forecasting model that could be used for exposure time calculation or observation quality forecasting needed to prioritize observations.

Usage for ultra-fast pseudo-closed loop simulations

The PSF simulation framework described in Chapter 4 is currently restricted to infinite-exposure PSFs. However, this limitation can be overcome. During this research, a code was developed to sample PSDs in a temporally consistent manner, utilizing the sparse spectral sampling technique (SSS), which is widely used in computer graphics to simulate the ocean surface (Cha (2011)) and can also be adapted to model atmospheric refractive index variations (Charnotskii (2016)).

By leveraging SSS to generate temporally coherent residual phase screens, the TipTorch model can simulate short-exposure PSFs at very high rates, significantly faster than conventional E2E AO simulations. Although such simulations cannot replace full E2E modeling, they can be used for the rapid generation of large datasets of realistic PSFs, which are invaluable for applications such as machine learning or AO systems performance modeling.

Additionally, the code I developed can generate 100-1000 atmospheric turbulent phase screens per second and allows for the simulation of atmospheric evolution beyond the traditional frozen-flow formalism, enabling the modeling of more complex temporal behaviors of atmospheric turbulence. The code is accessible through this

link [5].

Exploring fully NN-based PSF generation

Section 4.8.2.2 discussed a potential implementation for a fully NN-based PSF generation approach, where a data-driven PSF predictor is pre-trained on outputs from TipTorch / TipTop and subsequently fine-tuned on on-sky data. Although this strategy has not been pursued in the current work, it holds great promise for future research due to its flexibility and adaptability.

Expanding to new instruments

Given that the proposed approach has already demonstrated successful performance across two different instruments with distinct AO systems, it is logical to extend it to additional instruments, such as ERIS, GRAVITY+, or future ELT instruments. However, to harness the full potential of the NN-calibrator, these instruments must accumulate sufficiently large datasets of on-sky samples. This will ensure that the model can be properly trained and validated for each new instrument configuration, facilitating its integration into the operational pipeline.

5. https://github.com/EjjeSynho/FATMOSS

Bibliographie

2011, International Conference on Offshore Mechanics and Arctic Engineering, Vol. Volume 6 : Ocean Engineering, Sparse Spectrum Model of the Sea Surface, 773–778

Agapito, G., Busoni, L., Carlà, G., et al. 2022, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 12185, Adaptive Optics Systems VIII, ed. L. Schreiber, D. Schmidt, & E. Vernet, 1218556

Aru, M. L., Maucó, K., Manara, C. F., et al. 2024, Astronomy and Astrophysics, 687, A93

Bacon, R., Accardo, M., Adjali, L., et al. 2010, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7735, Ground-based and Airborne Instrumentation for Astronomy III, ed. I. S. McLean, S. K. Ramsay, & H. Takami, 773508

Bacon, R., Brinchmann, J., Bouché, N., et al. 2023, The Messenger, 191, 11

Bacon, R., Conseil, S., Mary, D., et al. 2017, Astronomy and Astrophysics, 608, A1

Baydin, A. G., Pearlmutter, B. A., Radul, A. A., & Siskind, J. M. 2017, J. Mach. Learn. Res., 18, 5595–5637

Beltramo-Martin, O., Fétick, R., Neichel, B., & Fusco, T. 2020, Astronomy and Astrophysics, 643, A58

Beltramo-Martin, O., Marasco, A., Fusco, T., et al. 2020a, Monthly Notices of the Royal Astronomical Society, 494, 775

Beltramo-Martin, O., Ragland, S., Fétick, R., et al. 2020b, in Adaptive Optics Systems VII, ed. L. Schreiber, D. Schmidt, & E. Vernet, Vol. 11448, International Society for Optics and Photonics (SPIE), 114480A

Beuzit, J. L., Vigan, A., Mouillet, D., et al. 2019, Astronomy and Astrophysics, 631, A155

Bifano, T. 2011, Nature Photonics, 5, 21

Björck, r. 1996, 9. Nonlinear Least Squares Problems (Society for Industrial and Applied Mathematics), 339–358

Charnotskii, M. 2016, J. Opt. Soc. Am. A, 33, 561

Chauvin, G., Desidera, S., Lagrange, A. M., et al. 2017, Astronomy and Astrophysics, 605, L9

Claudi, R. U., Turatto, M., Gratton, R. G., et al. 2008, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7014, Ground-based and Airborne Instrumentation for Astronomy II, ed. I. S. McLean & M. M. Casali, 70143E

Correia, C. M., Bond, C. Z., Sauvage, J.-F., et al. 2017, J. Opt. Soc. Am. A, 34, 1877

Correia, C. M. & Teixeira, J. 2014, J. Opt. Soc. Am. A, 31, 2763

Cortés, A., Neichel, B., Guesalaga, A., et al. 2012, Monthly Notices of the Royal Astronomical Society, 427, 2089

Courtney-Barrer, B., Woillez, J., Laugier, R., et al. 2022, in Optical and Infrared Interferometry and Imaging VIII, ed. A. Mérand, S. Sallum, & J. Sanchez-Bermudez, Vol. 12183, International Society for Optics and Photonics (SPIE), 1218328

Dainty, J. C. 2000, Optical Effects of Atmospheric Turbulence, ed. N. Ageorges & C. Dainty (Dordrecht : Springer Netherlands), 1–22

Davies, R., Absil, O., Agapito, G., et al. 2023, Astronomy and Astrophysics, 674, A207

Davies, R., Ageorges, N., Barl, L., et al. 2010, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7735, Ground-based and Airborne Instrumentation for Astronomy III, ed. I. S. McLean, S. K. Ramsay, & H. Takami, 77352A

de Boer, J., Langlois, M., van Holstein, R. G., et al. 2020, Astronomy and Astrophysics, 633, A63

Dhariwal, P. & Nichol, A. 2021, in Proceedings of the 35th International Conference on Neural Information Processing Systems, NIPS '21 (Red Hook, NY, USA : Curran Associates Inc.)

Diolaiti, E., Ciliegi, P., Abicca, R., et al. 2016, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9909, Adaptive Optics Systems V, ed. E. Marchetti, L. M. Close, & J.-P. Véran, 99092D

Dohlen, K., Langlois, M., Saisse, M., et al. 2008, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7014, Ground-based and Airborne Instrumentation for Astronomy II, ed. I. S. McLean & M. M. Casali, 70143L

Eisenhauer, F., Abuter, R., Bickert, K., et al. 2003, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 4841, Instrument Design and Performance for Optical/Infrared Ground-based Telescopes, ed. M. Iye & A. F. M. Moorwood, 1548–1561

Engler, B., Louarn, M. L., Vérinaud, C., Weddell, S., & Clare, R. M. 2022, Journal of Astronomical Telescopes, Instruments, and Systems, 8, 021502

Engler, N., Schmid, H. M., Quanz, S. P., Avenhaus, H., & Bazzon, A. 2018, Astronomy and Astrophysics, 618, A151

Epinat, B., Tasca, L., Amram, P., et al. 2012, Astronomy and Astrophysics, 539, A92

European Southern Observatory. 2018, New SPHERE view of Vesta, accessed : 2024-09-27

European Southern Observatory. 2019, ESO Telescope Sees Surface of Dim Betelgeuse, accessed : 2024-09-27

European Southern Observatory. 2022, La lune Ganymède vue par SPHERE, accessed : 2024-09-27

Evans, C., Puech, M., Afonso, J., et al. 2015, arXiv e-prints, arXiv :1501.04726

Fauvarque, O., Neichel, B., Fusco, T., Sauvage, J.-F., & Girault, O. 2016, Optica, 3, 1440

Ferreira, F., Sevin, A., Bernard, J., et al. 2020, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 11448, Adaptive Optics Systems VII, ed. L. Schreiber, D. Schmidt, & E. Vernet, 1144815

Fétick, R. J., Jorda, L., Vernazza, P., et al. 2019a, Astronomy and Astrophysics, 623, A6

Fétick, R. J. L., Fusco, T., Neichel, B., et al. 2019b, Astronomy and Astrophysics, 628, A99

Fétick, R. J. L., Neichel, B., Mugnier, L. M., Montmerle-Bonnefois, A., & Fusco, T. 2018, Monthly Notices of the RAS, 481, 5210

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, Publications of the ASP, 125, 306

Foy, R. 2000, The Cone Effect, ed. N. Ageorges & C. Dainty (Dordrecht : Springer Netherlands), 107–123

Fried, D. L. 1965, J. Opt. Soc. Am., 55, 1427

Fried, D. L. 1978, J. Opt. Soc. Am., 68, 1651

Fusco, T., Bacon, R., Kamann, S., et al. 2020, Astronomy and Astrophysics, 635, A208

Fusco, T., Conan, J.-M., Rousset, G., Mugnier, L. M., & Michau, V. 2001a, J. Opt. Soc. Am. A, 18, 2527

Fusco, T., Conan, J.-M., Rousset, G., Mugnier, L. M., & Michau, V. 2001b, J. Opt. Soc. Am. A, 18, 2527

Fusco, T., Rousset, G., Sauvage, J. F., et al. 2006, Optics Express, 14, 7515

Fusco, T., Sauvage, J. F., Petit, C., et al. 2014, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9148, Adaptive Optics Systems IV, ed. E. Marchetti, L. M. Close, & J.-P. Vran, 91481U

Genzel, R., Eisenhauer, F., & Gillessen, S. 2010, Rev. Mod. Phys., 82, 3121

Ghez, A. M., Salim, S., Weinberg, N. N., et al. 2008, Astrophysical Journal, 689, 1044

Ginski, C., Garufi, A., Benisty, M., et al. 2024, Astronomy and Astrophysics, 685, A52

Girard, J. H., de Boer, J., Haffert, S., et al. 2020, arXiv e-prints, arXiv :2003.02145

Gomes, T., Correia, C. M., Bardou, L., et al. 2024, Astronomy and Astrophysics, 686, A7

Gonsalves, R. A. 1982, Optical Engineering, 21, 829

Goodman, J. W. 2005, Introduction to Fourier optics / Joseph W. Goodman., 3rd edn. (Englewood, Colo : Roberts & Co.)

Goulas, C., Vidal, F., Galicher, R., et al. 2023, arXiv e-prints, arXiv :2310.15765

GRAVITY Collaboration, Abuter, R., Amorim, A., et al. 2018, Astronomy and Astrophysics, 615, L15

Guesalaga, A., Neichel, B., O'Neal, J., & Guzman, D. 2013, Opt. Express, 21, 10676

Gúrpide, A., Parra, M., Godet, O., Contini, T., & Olive, J. F. 2022, Astronomy and Astrophysics, 666, A100

Haguenauer, P., Guesalaga, A., & Butterley, T. 2020, in Adaptive Optics Systems VII, ed. L. Schreiber, D. Schmidt, & E. Vernet, Vol. 11448, International Society for Optics and Photonics (SPIE), 114481K

Happer, W., MacDonald, G. J., Max, C. E., & Dyson, F. J. 1994, J. Opt. Soc. Am. A, 11, 263

Hartke, J., Kakkad, D., Reyes, C., et al. 2020, in Adaptive Optics Systems VII, ed. L. Schreiber, D. Schmidt, & E. Vernet, Vol. 11448, International Society for Optics and Photonics (SPIE), 114480V

Hartung, M., Blanc, A., Fusco, T., et al. 2003, Astronomy and Astrophysics, 399, 385

Haworth, T. J., Reiter, M., O'Dell, C. R., et al. 2023, Monthly Notices of the Royal Astronomical Society, 525, 4129

Heritier, C. 2019, PhD thesis

Heritier, C. T., Fusco, T., Oberti, S., et al. 2021, Monthly Notices of the Royal Astronomical Society, 504, 4274

Hodson, T. O. 2022, Geoscientific Model Development, 15, 5481

Holzlöhner, R. & Brinkmann, M. 2020, in Ground-based and Airborne Telescopes VIII, ed. H. K. Marshall, J. Spyromilio, & T. Usuda, Vol. 11445, International Society for Optics and Photonics (SPIE), 114450Y

Huber, P. J. 1964, The Annals of Mathematical Statistics, 35, 73

Hubin, N., Arsenault, R., Conzelmann, R., et al. 2005, Comptes Rendus. Physique, 6, 1099

Jiang, L., Hu, L., Hu, Q., et al. 2023, Photonics, 10

Jolissaint, L. 2010, Journal of the European Optical Society - Rapid publications, 5

Jolissaint, L. & Kendrew, S. 2010, in Adaptative Optics for Extremely Large Telescopes, 05021

Jolissaint, L., Ragland, S., Christou, J., & Wizinowich, P. 2018, Appl. Opt., 57, 7837

Jolissaint, L., Ragland, S., & Wizinowich, P. 2015

Jolissaint, L., Véran, J.-P., & Conan, R. 2006, J. Opt. Soc. Am. A, 23, 382

Kamann, S., Bastian, N., Usher, C., Cabrera-Ziri, I., & Saracino, S. 2021, Monthly Notices of the Royal Astronomical Society, 508, 2302

Kamann, S., Wisotzki, L., & Roth, M. M. 2013, Astronomy and Astrophysics, 549, A71

Kasper, M., Cerpa Urra, N., Pathak, P., et al. 2021, The Messenger, 182, 38

Kidger, P. & Lyons, T. 2019, arXiv e-prints, arXiv :1905.08539

Kingma, D. P. & Ba, J. 2014, CoRR, abs/1412.6980

Knapen, J. H., Comerón, S., & Seidel, M. K. 2019, Astronomy and Astrophysics, 621, L5

Kolmogorov, A. N. 1991, Proceedings : Mathematical and Physical Sciences, 434, 9

Korkiakoski, V., Keller, C. U., Doelman, N., et al. 2014, Appl. Opt., 53, 4565

Kulcsár, C., Massioni, P., Sivo, G., & Raynaud, H.-F. G. 2012, in Adaptive Optics Systems III, ed. B. L. Ellerbroek, E. Marchetti, & J.-P. Véran, Vol. 8447, International Society for Optics and Photonics (SPIE), 84470Z

Kulcsár, C., Raynaud, H.-F., Petit, C., Conan, J.-M., & de Lesegno, P. V. 2006, Opt. Express, 14, 7464

Kuznetsov, A., Neichel, B., Oberti, S., & Fusco, T. 2023, in Adaptive Optics for Extremely Large Telescopes (AO4ELT7), 104

Kuznetsov, A., Oberti, S., Neichel, B., & Fusco, T. 2024, Astronomy and Astrophysics, 687, A221

Lai, O., Withington, J. K., Laugier, R., & Chun, M. 2019, Monthly Notices of the Royal Astronomical Society, 484, 5568

Lamb, M., Norton, A., Macintosh, B., et al. 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10703, Adaptive Optics Systems VI, ed. L. M. Close, L. Schreiber, & D. Schmidt, 107035M

Lamb, M. P., Correia, C., Sivanandam, S., Swanson, R., & Zavyalova, P. 2021, Monthly Notices of the Royal Astronomical Society, 505, 3347

Landman, R., Haffert, S. Y., Radhakrishnan, V. M., & Keller, C. U. 2021, Journal of Astronomical Telescopes, Instruments, and Systems, 7, 039002

Lau, A., Fétick, R. J. L., Neichel, B., Beltramo-Martin, O., & Fusco, T. 2023, Astronomy and Astrophysics, 673, A72

Law, N. M., Mackay, C. D., & Baldwin, J. E. 2006, Astronomy and Astrophysics, 446, 739

Leboulleux, L., Carlotti, A., & N'Diaye, M. 2022, Astronomy and Astrophysics, 659, A143

Levraud, N., Chambouleyron, V., Sauvage, J. F., et al. 2024a, Astronomy and Astrophysics, 682, A84

Levraud, N., Chambouleyron, V., Sauvage, J. F., et al. 2024b, Astronomy and Astrophysics, 682, A84

Liu, D. C. & Nocedal, J. 1989, Mathematical Programming, 45, 503

Louarn, M. L., Madec, P.-Y., Marchetti, E., Bonnet, H., & Esselborn, M. 2016, in Adaptive Optics Systems V, ed. E. Marchetti, L. M. Close, & J.-P. Véran, Vol. 9909, International Society for Optics and Photonics (SPIE), 990975

Lozi, J., Guyon, O., Jovanovic, N., et al. 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10703, Adaptive Optics Systems VI, ed. L. M. Close, L. Schreiber, & D. Schmidt, 1070359

Lu, J. R., Ghez, A. M., Hornstein, S. D., et al. 2008, The Astrophysical Journal, 690, 1463

Macintosh, B., Graham, J. R., Ingraham, P., et al. 2014, Proceedings of the National Academy of Science, 111, 12661

Madec, P.-Y. 1999, Control techniques, ed. F. Roddier (Cambridge University Press), 131–154

Martins, D., Holzlöhner, R., Verinaud, C., & Kleinclaus, C. 2022, in Modeling, Systems Engineering, and Project Management for Astronomy X, ed. G. Z. Angeli & P. Dierickx, Vol. 12187, International Society for Optics and Photonics (SPIE), 1218712

Mateen, M., Codona, J., & Guyon, O. 2017, in Proceedings of the Adaptive Optics for Extremely Large Telescopes 5 (Instituto de Astrofísica de Canarias (IAC))

Mauch, S. & Reger, J. 2014, IEEE Transactions on Instrumentation and Measurement, 63, 2379

Mawet, D., Wizinowich, P., Dekany, R., et al. 2016, in Adaptive Optics Systems V, ed. E. Marchetti, L. M. Close, & J.-P. Véran, Vol. 9909, International Society for Optics and Photonics (SPIE), 99090D

Mazzoleni, R., Gonté, F., Surdej, I., et al. 2008, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7012, Ground-based and Airborne Telescopes II, ed. L. M. Stepp & R. Gilmozzi, 70123A

McDermid, R. M., Cresci, G., Rigaut, F., et al. 2020, arXiv e-prints, arXiv :2009.09242

Meimon, S., Fusco, T., & Mugnier, L. M. 2010, Opt. Lett., 35, 3036

Melnick, J., Marchetti, E., & Amico, P. 2012, in Adaptive Optics Systems III, ed. B. L. Ellerbroek, E. Marchetti, & J.-P. Véran, Vol. 8447, International Society for Optics and Photonics (SPIE), 84470M

Mesa, D., Gratton, R., Zurlo, A., et al. 2015, Astronomy and Astrophysics, 576, A121

Milli, J., Kasper, M., Bourget, P., et al. 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10703, Adaptive Optics Systems VI, ed. L. M. Close, L. Schreiber, & D. Schmidt, 107032A

Monreal-Ibero, A. & Walsh, J. R. 2020, Astronomy and Astrophysics, 634, A47

Mu, Q., Cao, Z., Li, D., Hu, L., & Xuan, L. 2008, Appl. Opt., 47, 4297

Mugnier, L. M., Sauvage, J.-F., Fusco, T., Cornia, A., & Dandy, S. 2008, Opt. Express, 16, 18406

Muradore, R., Pettazzi, L., Clare, R., & Fedrigo, E. 2014, Control Engineering Practice, 32, 87

Nam, S., Park, S., Yun, S., et al. 2016, Opt. Express, 24, 55

N'Diaye, M., Vigan, A., Dohlen, K., et al. 2016, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9909, Adaptive Optics Systems V, ed. E. Marchetti, L. M. Close, & J.-P. Véran, 99096S

Neichel, B. 2008a, Etudes des galaxies lointaines et optiques adaptatives tomographiques pour les ELTs., 125–127

Neichel, B. 2008b, Theses, Université Paris-Diderot - Paris VII

Neichel, B., Beltramo-Martin, O., Plantet, C., et al. 2020, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 11448, Adaptive Optics Systems VII, ed. L. Schreiber, D. Schmidt, & E. Vernet, 114482T

Neichel, B., Fusco, T., & Conan, J.-M. 2009, J. Opt. Soc. Am. A, 26, 219

Neichel, B., Lu, J. R., Rigaut, F., et al. 2014, Monthly Notices of the Royal Astronomical Society, 445, 500

Nocedal, J. & Wright, S. J. 2006a, Large-Scale Unconstrained Optimization, 2nd edn., Springer Series in Operations Research and Financial Engineering (Springer New York, NY), 164–192, part of the eBook Packages Mathematics and Statistics (R0) ; Copyright Springer-Verlag New York 2006

Nocedal, J. & Wright, S. J. 2006b, Line Search Methods, 2nd edn., Springer Series in Operations Research and Financial Engineering (Springer New York, NY), 30–65, part of the eBook Packages Mathematics and Statistics (R0) ; Copyright Springer-Verlag New York 2006

Nocedal, J. & Wright, S. J. 2006c, Numerical Optimization, 2nd edn., Springer Series in Operations Research and Financial Engineering (Springer New York, NY), XXII + 664, part of the eBook Packages Mathematics and Statistics (R0) ; Copyright Springer-Verlag New York 2006

Nocedal, J. & Wright, S. J. 2006d, Quasi-Newton Methods, 2nd edn., Springer Series in Operations Research and Financial Engineering (Springer New York, NY), 135–163, part of the eBook Packages Mathematics and Statistics (R0) ; Copyright Springer-Verlag New York 2006

Noll, R. J. 1976, J. Opt. Soc. Am., 66, 207

Norris, B. R. M., Wei, J., Betters, C. H., Wong, A., & Leon-Saval, S. G. 2020, Nature Communications, 11, 5335

Nousiainen, J., Rajani, C., Kasper, M., et al. 2022, Astronomy and Astrophysics, 664, A71

Oberti, S., Kolb, J., Madec, P.-Y., et al. 2018, in Adaptive Optics Systems VI, ed. L. M. Close, L. Schreiber, & D. Schmidt, Vol. 10703, International Society for Optics and Photonics (SPIE), 107031G

Osborn, J., Föhring, D., Dhillon, V. S., & Wilson, R. W. 2015, Monthly Notices of the Royal Astronomical Society, 452, 1707

Paxman, R. G. & Fienup, J. R. 1988, J. Opt. Soc. Am. A, 5, 914

Pechetti, R., Kamann, S., Krajnović, D., et al. 2024, Monthly Notices of the RAS, 528, 4941

Pence, W. D., Chiappetti, L., Page, C. G., Shaw, R. A., & Stobie, E. 2010, Astronomy and Astrophysics, 524, A42

Persic, M., Salucci, P., & Stel, F. 1996, Monthly Notices of the Royal Astronomical Society, 281, 27

Petit, C., Conan, J.-M., Kulcsár, C., & Raynaud, H.-F. 2009, J. Opt. Soc. Am. A, 26, 1307

Pfrommer, T. & Hickson, P. 2014, Astronomy and Astrophysics, 565, A102

Pourré, N., Le Bouquin, J. B., Milli, J., et al. 2022, Astronomy and Astrophysics, 665, A158

Quirrenbach, A. 2004, in

Ragazzoni, R. 1996, Journal of Modern Optics, 43, 289

Ragland, S., Jolissaint, L., Wizinowich, P., et al. 2016, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9909, Adaptive Optics Systems V, ed. E. Marchetti, L. M. Close, & J.-P. Véran, 99091P

Raissi, M., Perdikaris, P., & Karniadakis, G. 2019, Journal of Computational Physics, 378, 686

Rigaut, F. 2002, in European Southern Observatory Conference and Workshop Proceedings, Vol. 58, European Southern Observatory Conference and Workshop Proceedings, ed. E. Vernet, R. Ragazzoni, S. Esposito, & N. Hubin, 11

Rigaut, F. & Neichel, B. 2018, Annual Review of Astronomy and Astrophysics, 56, 277

Rigaut, F., Neichel, B., Boccas, M., et al. 2013, Monthly Notices of the Royal Astronomical Society, 437, 2361

Rigaut, F. J., Veran, J.-P., & Lai, O. 1998, in Adaptive Optical System Technologies, ed. D. Bonaccini & R. K. Tyson, Vol. 3353, International Society for Optics and Photonics (SPIE), 1038 – 1048

Robert, C., Fusco, T., Sauvage, J.-F., & Mugnier, L. 2008, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7015, Adaptive Optics Systems, ed. N. Hubin, C. E. Max, & P. L. Wizinowich, 70156A

Roddier, F. 1981, Progess in Optics, 19, 281

Roddier, F. 1988, Appl. Opt., 27, 1223

Roddier, F. 1999a, Adaptive Optics in Astronomy, ed. F. Roddier (Cambridge University Press)

Roddier, F. 1999b, Imaging through the atmosphere, ed. F. Roddier (Cambridge University Press), 9–22

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. 2021, arXiv e-prints, arXiv :2112.10752

Rousset, G. 1999, Wave-front sensors, ed. F. Roddier (Cambridge University Press), 91–130

Rousset, G., Lacombe, F., Puget, P., et al. 2003, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 4839, Adaptive Optical System Technologies II, ed. P. L. Wizinowich & D. Bonaccini, 140–149

Samland, M., Mollière, P., Bonnefoy, M., et al. 2017, Astronomy and Astrophysics, 603, A57

Sasiela, R. J. 1992, J. Opt. Soc. Am. A, 9, 1398

Sauvage, J. F., Bond, C., Fusco, T., et al. 2022, in Adaptive Optics Systems VIII, ed. L. Schreiber, D. Schmidt, & E. Vernet, Vol. 12185, International Society for Optics and Photonics (SPIE), 1218519

Sauvage, J.-F., Fusco, T., Lamb, M., et al. 2016, in Adaptive Optics Systems V, ed. E. Marchetti, L. M. Close, & J.-P. Véran, Vol. 9909, International Society for Optics and Photonics (SPIE), 990916

Schmid, H. M., Bazzon, A., Roelfsema, R., et al. 2018, Astronomy and Astrophysics, 619, A9

Schmidt, J., Whiteley, M., Goda, M., & Duncan, B. 2006, in IEEE Aerospace Conference Proceedings, Vol. 2006, 10 pp.

Sebag, J. & Vogiatzis, K. 2014, in Modeling, Systems Engineering, and Project Management for Astronomy VI, ed. G. Z. Angeli & P. Dierickx, Vol. 9150, International Society for Optics and Photonics (SPIE), 91500R

Simioni, M., Arcidiacono, C., Grazian, A., et al. 2020, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 11448, Adaptive Optics Systems VII, ed. L. Schreiber, D. Schmidt, & E. Vernet, 1144837

Speagle, J. S. 2020, Monthly Notices of the RAS, 493, 3132

Srivastava, V., Mansell, G., Makarem, C., et al. 2022, Opt. Express, 30, 10491

Terry, S. K., Lu, J. R., Turri, P., et al. 2023, Journal of Astronomical Telescopes, Instruments, and Systems, 9, 018003

Thatte, N. A., Melotte, D., Neichel, B., et al. 2022, in Ground-based and Airborne Instrumentation for Astronomy IX, ed. C. J. Evans, J. J. Bryant, & K. Motohara, Vol. 12184, International Society for Optics and Photonics (SPIE), 1218420

Thompson, L. A. & Gardner, C. S. 1989, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 1114, Active telescope systems, ed. F. J. Roddier, 184–190

Tyler, G. A. 1994, J. Opt. Soc. Am. A, 11, 325

Usher, C., Kamann, S., Gieles, M., et al. 2021, Monthly Notices of the RAS, 503, 1680

Valles, M. S., Fedrigo, E., Donaldson, R. H., et al. 2012, in Adaptive Optics Systems III, ed. B. L. Ellerbroek, E. Marchetti, & J.-P. Véran, Vol. 8447, International Society for Optics and Photonics (SPIE), 84472Q

van Dam, M. A., McLeod, B. A., & Bouchez, A. H. 2016, Appl. Opt., 55, 539

Véran, J.-P., Rigaut, F., Maître, H., & Rouan, D. 1997, J. Opt. Soc. Am. A, 14, 3057

Vigan, A., Langlois, M., Moutou, C., & Dohlen, K. 2008, Astronomy and Astrophysics, 489, 1345

Vigan, A., Moutou, C., Langlois, M., et al. 2010, Monthly Notices of the RAS, 407, 71

Vigan, A., N'Diaye, M., & Dohlen, K. 2013, Astronomy and Astrophysics, 555, A49

von Kármán, T. 1948, Proceedings of the National Academy of Sciences of the United States of America, 34, 530

Vérinaud, C., Héritier, C. T., Kasper, M., & Tallon, M. 2024, A&A, 682, A27

Witzel, G., Lu, J. R., Ghez, A. M., et al. 2016, in Adaptive Optics Systems V, ed. E. Marchetti, L. M. Close, & J.-P. Véran, Vol. 9909, International Society for Optics and Photonics (SPIE), 99091O

Wizinowich, P. L., Acton, D. S., Lai, O., et al. 2000, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 4007, Adaptive Optical Systems Technology, ed. P. L. Wizinowich, 2–13

Woolf, N. 1979, Publications of the Astronomical Society of the Pacific, 91, 523

Yuan, Y.-X. 1991, IMA Journal of Numerical Analysis, 11, 325

Zhang, Y., Tiňo, P., Leonardis, A., & Tang, K. 2021, IEEE Transactions on Emerging Topics in Computational Intelligence, 5, 726

Zhu, D., Yao, H., Jiang, B., & Yu, P. 2018, arXiv e-prints, arXiv :1804.10690

Acknowledgement

I would like to thank Olivier Beltramo-Martin for his invaluable support during the early stages of my work on PSF prediction, which was crucial in initiating my research in this area. I want to express my gratitude to Vladimir Guzov for his numerous pieces of advice and insights on various aspects of machine learning, which greatly contributed to the development of my project. I also want to thank Julien Milli for his guidance in understanding IRDIS on-sky data and Joël Vernet for helping me start with MUSE on-sky data processing. I want to acknowledge Cedric Taïssir Heritier and Cédric Plantet for their help in the early development of the LIFT technique implementation.

I want to thank Sylvain Oberti, Johann Kolb, and Pavel Shchekaturov for their support in conducting the on-sky validations of the LIFT and DIP techniques. I am deeply grateful to Carlos Correia, Byron Engler, Olivier Hainaut, Carlo Manara, Christian Stephan, Alain Smette, and Natalie Behara for their assistance in preparing my fellowship application. I am also deeply grateful to Peter Wizinowich, Joël Vernet, Benoît Neichel, and Thierry Fusco for supporting my application.

I am especially thankful to my supervisors, Benoît Neichel, Sylvain Oberti, and Thierry Fusco, for giving me creative freedom but also for giving guidance and encouragement throughout my PhD. In addition, I would like to thank Byron Engler, Alexis Lau, Nicolas Pourré, Marcos Suárez Valles, Romain Fétick, Markus Kasper, and Guido Agapito for their occasional advice and insights that were instrumental in advancing my research.

Finally, I am deeply grateful to my family and friends for their constant moral support, which helped me navigate the challenging times of the post-COVID crisis that accompanied me throughout my entire PhD journey.

This work benefited from the support of the European Southern Observatory, French National Research Agency (ANR) with *WOLF (ANR-18-CE31-0018)*, *APPLY (ANR-19-CE31-0011)* and *LabEx FOCUS (ANR-11-LABX-0013)*; the Programme Investissement Avenir *F-CELT (ANR-21-ESRE-0008)*, the *Action Spécifique Haute Résolution Angulaire (ASHRA)* of CNRS/INSU co-funded by CNES, the *ECOS-CONYCIT* France-Chile cooperation (*C20E02*), the *ORP-H2020* Framework Programme of the European Commission's (Grant number *101004719*), *STIC AmSud (21-STIC-09)*, the Région Sud and the french government under the *France 2030 investment plan*, as part of the *Initiative d'Excellence d'Aix-Marseille Université A*MIDEX, program number AMX-22-RE-AB-151*.

Annexes

A. Coupling between modes in LIFT

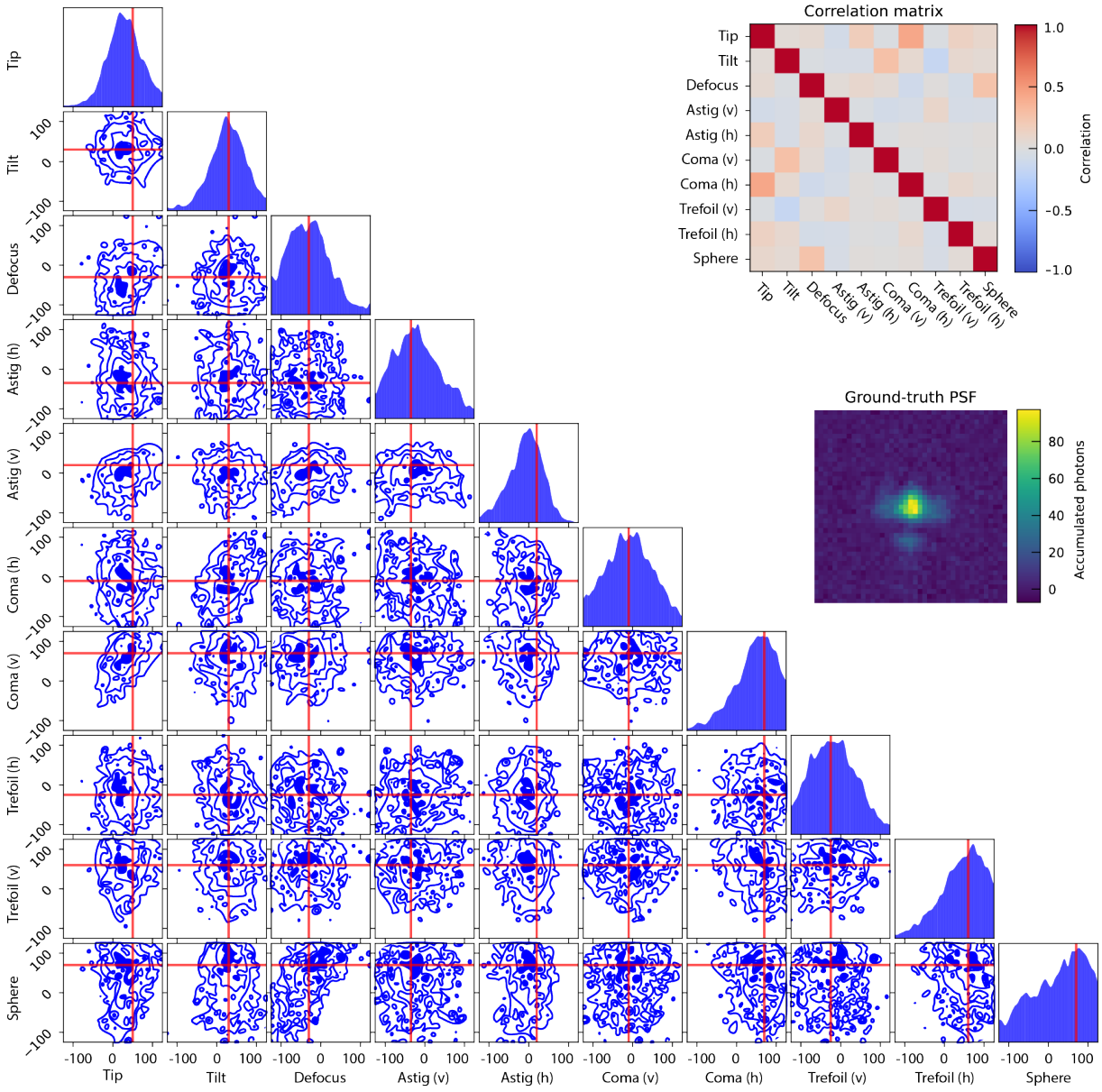


FIGURE 1. – The effect of low flux (m_J = 16) on modal cross-talk. Increased noise introduces numerous small parasitic local minima, which can trap the optimizer and hinder convergence to the true solution.

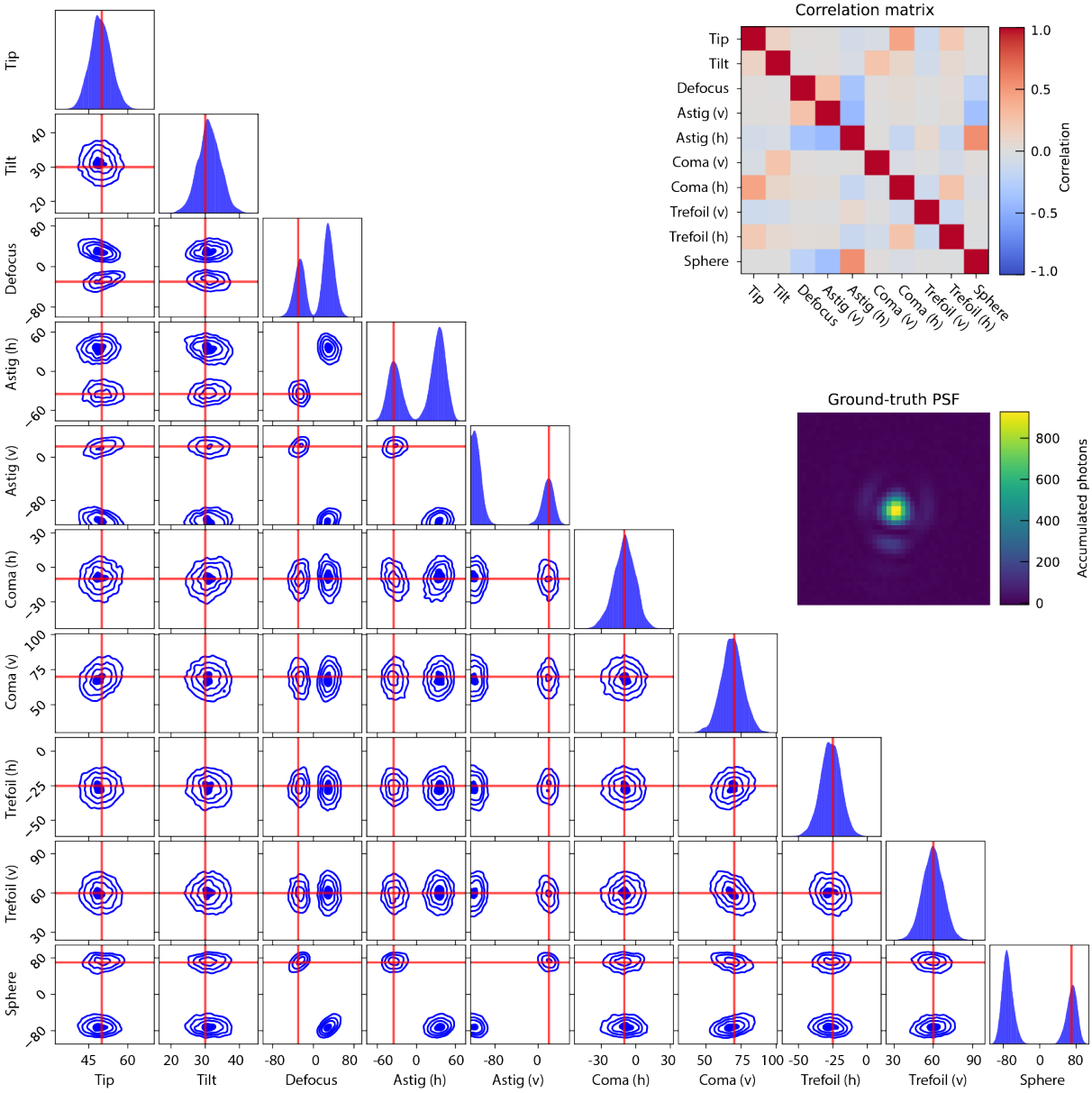


FIGURE 2. – The effect of astigmatic small diversity (50 nm RMS instead of 170 nm RMS) on modal cross-talk. The second local minimum appears.

B. PSF models

B.1. Code

TipTop code can be accessed from : https://github.com/astro-tiptop
The P3 library can be accessed at : https://github.com/astro-tiptop/P3
TipTorch can be accessed through : https://github.com/EjjeSynho/TipTorch

B.2. Model initialization

Typical parameters used to initialize the P3 / TipTorch PSF models are listed below. The examples of config files can be found under the following links : https://github.com/EjjeSynho/TipTorch/tree/master/data/parameter_files, https://github.com/astro-tiptop/P3/tree/main/p3/aoSystem/parFiles.

- **Telescope**
  - Diameter
  - Zenith Angle
  - Obscuration Ratio
  - Pupil mask
  - Apodizer map
  - Pupil Angle
- **Atmosphere**
  - Wavelength
  - Seeing
  - Outer Scale (L_0)
  - Turbulence Layer Weights
  - Turbulence Layer Heights
  - Wind Speed
  - Wind Direction
- **Science Source**
  - Wavelength
  - Zenith Position
  - Azimuth Position
- **High-Order Source**
  - Wavelength
  - Zenith Positions
  - Azimuth Positions
  - Height
- **Low-Order Source**
  - Wavelength
  - Zenith Position
  - Azimuth Position
- **Science Sensor**
  - Pixel Scale
  - Field of View
  - Binning Factor
  - Number of Photons
  - Spot FWHM
  - Spectral Bandwidth
  - Transmittance
  - Dispersion
  - RON
  - Dark Current
  - Sky Background
  - Gain
  - Excess Noise Factor
  - Zenith Position
  - Azimuth Position
- **High-Order Sensor**
  - WFS Type
  - Modulation
  - Pixel Scale
  - Field of View
  - Binning Factor
  - Number of Photons
  - Spot FWHM
  - Spectral Bandwidth
  - Transmittance
  - Dispersion
  - Readout Noise (RON)
  - Dark Current
  - Sky Background
  - Gain
  - Excess Noise Factor
  - Number of Lenslets
  - Lenslet Size
  - Noise Variance
  - Centroiding algorithm
  - Window Radius (WCoG)
  - Threshold (WCoG)
- **Low-Order Sensor**
  - Pixel Scale
  - Field of View
  - Binning Factor
  - Number of Photons
  - Spot FWHM
  - RON

- Dark Current
- Sky Background
- Gain
- Excess Noise Factor
- Number of Lenslets
- Centroiding algorithm
- Window Radius (WCoG)
- Threshold (WCoG)

- **Deformable Mirror (DM)**
  - Number of Actuators
  - Actuator Pitch
  - Influence Model
  - Influence Coupling
  - Heights of Deformable Mirrors
  - Optimization Zenith
  - Optimization Azimuth
  - Optimization Weight
  - Optimization Conditioning
  - Number of Reconstructed Layers
  - Area of Adaptive Optics
- **Real-Time Controller (RTC)**
  - High-Order Loop Gain
  - High-Order Sensor Frame Rate
  - High-Order Loop Delay
  - Low-Order Loop Gain
  - Low-Order Sensor Frame Rate
  - Low-Order Loop Delay

C. Data processing

C.1. IRDIS

C.2. Downloading IRDIS exposure cubes

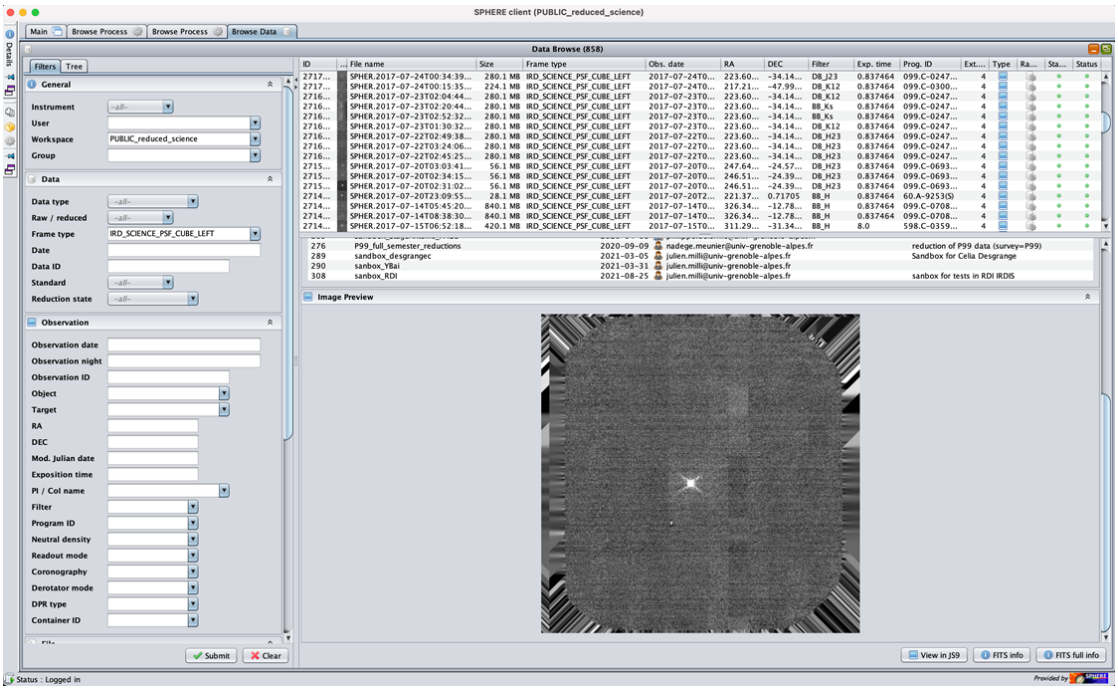


FIGURE 3. – The SPHERE-client interface. This software was used to download the IRDIS PSF dataset from the ESO archive.

C.2.1. PSFs wavelengths groups

Samples count	λ left, [nm]	λ right, [nm]	Filter name
842	1625	1625	`B_H`
781	1593	1667	`D_ND-H23`
588	2110	2251	`D_K12`
488	2182	2182	`B_Ks`
266	1245	1245	`B_J`
79	1190	1273	`D_J23`
30	1043	1043	`B_YB_Y`
14	1573	1573	`N_CntH`
12	1022	1076	`D_Y23`
11	1213	1213	`N_CntJ`

TABLEAU 1. – The data samples in the IRDIS dataset are categorized based on the type of wavelength filters applied in the left and right channels during acquisition.

C.2.2. Reduced telemetry dataset

This section explains the content and the retrieval procedure of reduced telemetry datasets used with the TipTorch model for direct and calibrated prediction in the case of IRDIS. Some important points regarding this data set are discussed below.

— All values are medians over the exposure period;

— If there are no recorded entries during the exposure, the nearest record is searched within 120 seconds before starting the exposure or 120 seconds after finishing it;

— The `DET SEQ UTC` entry in the exposure cube header tells the exposure time stamp. The end time is determined by adding the `EXPTIME` value, which contains the exposure duration in seconds;

— All atmospheric entries, like r_0, seeing, τ_0, turbulence speed, are computed for the atmospheric wavelength $\lambda_{\text{atm}} = 500$ nm;

— Values obtained from the "header" are obtained from the header of the FITS file with PSF cube;

— Values obtained from the "SPARTA" are obtained from the reduced SPARTA telemetry;

— Values with "MASSDIMM" are obtained from the bundle of MASS and DIMM measurements;

— Values with "MASS" are obtained from the MASS telemetry;

— ["] units are arc-seconds;

— [command units] are the units of the commands sent to DM / ITTM;

— The brief description of ESO site monitor, see
https://www.eso.org/sci/facilities/paranal/astroclimate/asm-instruments.html;

— Site monitor manual :
http://archive.eso.org/cms/eso-data/ambient-conditions/Astronomical_Site_Monitor_Data_User_Manual_v20171107.pdf

— Data retrieval and post-processing routines are based on the routines from Julien Milli et al.
https://github.com/jmilou/sparta;

Dataset entry	Archive entry	Data source	Units	Description
Filename	File itself	File itself		Name of the raw PSF cube in the ESO archive.
Date	DATE	Science header		Timestamp of the OB in UTC timezone.[4]
Observation	OBS_NAME	Science header		OB identifier.
RA	RA	Science header	[deg]	Sky coordinates of the target.
DEC	DEC	Science header	[deg]	
Airmass	TEL_AIRM[1]	Science header		Airmass value from header.
Azimuth	TEL_AZ	Science header	[deg]	Telescope pointing from header.
Altitude	TEL_ALT	Science header	[deg]	
r_0 (SPARTA)	R0	DTTS headers	[m]	r_0 value obtained from the reduced SPARTA telemetry.
Seeing (SPARTA)	R0[2]	DTTS headers	["]	Seeing values.
Seeing (MASSDIMM)	tab_fwhm	asm/mass	["]	
Seeing (DIMM)	tab_fwhm	asm/dimm	["]	
FWHM (header)	TEL_IA_FWHMLINOBS	Science header	["]	Estimated PSF FWHM value from the header.
Strehl	StrehlRatio	DTTS headers		Strehl value from SPARTA telemetry.
Turb. speed	tab_turb_speed	asm/mass	[m/s]	Turbulence speed from MASS-DIMM.
Wind direction (header)	TEL_AMBI_WINDDIR	Science header	[deg]	Direction of the wind.
Wind direction (ASM)	tab_wind_dir1	asm/meteo	[deg]	
Wind direction (200 mbar)	ecmwf_200mbar_winddir[3]	ECMWF	[deg]	
Wind speed (header)	WINDSP	Science header	[m/s]	Wind speed.
Wind speed (SPARTA)	WindSpeed	DTTS headers	[m/s]	
Wind speed (ASM)	tab_wind_speed1	asm/meteo	[m/s]	
Wind speed (200 mbar)	ecmwf_200mbar_windspeed[3]	ECMWF	[m/s]	
τ_0 (header)	TEL_AMBI_TAU0	Science header	[s]	Atmospheric coherence time.
τ_0 (SPARTA)	WindSpeed[6]	DTTS headers	[s]	
τ_0 (MASSDIMM)	tab_tau0	asm/mass	[s]	
τ_0 (MASS)	tab_tau	asm/mass	[s]	
Jitter X	TTx_std	DTTS headers	[cmds. units]	Standard deviation ITTM command projection onto the Zernike tip-tilt mode.[5]
Jitter Y	TTy_std	DTTS headers	[cmds. units]	
Focus	Focus_avg	DTTS headers	[cmds. units]	Average projection of HODM commands onto the Zernike focus mode.

ITTM pos (avg)	ITTMPos_avg	DTTS headers	[cmds. units]	Average and STD of the ITTM position.
ITTM pos (std)	ITTMPos_std	DTTS headers	[cmds. units]	
DM pos (avg)	DMPos_avg	DTTS headers	[cmds. units]	Average and STD of the HODM actuators' position averaged over the actuators.
DM pos (std)	DMPos_std	DTTS headers	[cmds. units]	
Pressure	TEL_AMBI_PRES[7]	Science header	[mmHg]	Atmospheric pressure value from the header.
Humidity	TEL_AMBI_RHUM	Science header	[%]	Relative humidity from header.
Temperature	tab_temp1	asm/meteo	[$C°$]	Temperature at 30 meters measured by METEO.
Temperature (200 mbar)	ecmwf_200mbar_temp[3]	ECMWF	[$C°$]	Temperature from ECMWF model.
Nph WFS	Flux_avg[9]	DTTS headers	[ph./sub-ap./fr.]	Photons per sub-aperture per frame.
Flux WFS	Flux_avg[10]	DTTS headers	[ph./s]	Flux per subaperture.
Loop rate	AOS_VISWFS_MODE[11]	Science header	[Hz]	AO loop rate.
WFS gain	AOS_VISWFS_MODE[11]	Science header		Gain of the detector.
Filter common	INS1_FILT_NAME	Science header	[nm]	Wavelength of the common filter used for the science (for both left and right channels).
Filter left	INS1_OPTI2_NAME[12]	Science header	[nm]	Wavelength of the left science channel.
Filter right	INS1_OPTI2_NAME[12]	Science header	[nm]	Wavelength of the right science channel.
Filter WFS	INS4_FILT3_NAME	Science header	[nm]	WFSing wavelength.
DIT Time	DET_SEQ1_DIT	Science header	[s]	Time per DIT.
Num. DITs	DET_NDIT	Science header		Number of DITs.
Sequence time	DET_SEQ1_EXPTIME	Science header	[s]	Exposure time.

1. Computed as `(TEL AIRM START + TEL AIRM END) / 2`;
2. Seeing (SPARTA) = rad2asec ($\frac{500\,\text{nm}}{\text{R0}}$);
3. Entries featuring "200 mbar" were computed using the ECMWF model at 200 mbar atmospheric pressure level, which approximately corresponds to the altitude of 12 km;
4. Not equal to the actual start of the exposure. The exposure start timestamp can be obtained from `DET SEQ UTC`.;
5. Always stored as 0 due to a bug in the system;
6. τ_0 (SPARTA) = $0.314 \cdot r_0$ / `WindSpeed`;
7. Computed as `(TEL AMBI PRES START + TEL AMBI PRES END) / 2`;
8. ADU2photon = $270000/2^{14} \approx 17$ [ph./ADU], $N_{\text{subap}} = 1240$;
9. Nph (WFS) = $\frac{\texttt{Flux_avg} \cdot \text{ADU2photon}}{\text{WFS gain} \cdot N_{\text{subap}}}$;
10. Flux WFS = $\texttt{Flux_avg} \cdot \text{WFS gain} \cdot N_{\text{subap}}$;
11. Loop gain and frequency are extracted from `GAIN_XXX_FREQ_XXX_Hz`, `XXX` means that value varies for different regimes;
12. Visible/Near-IR beam splitter name can be obtained from the `INS4 OPTI16 NAME` from the header. Common filter name can be extracted from `INS1 FILT NAME`;

C.3. MUSE-NFM

C.3.1. Reduced telemetry dataset

This section explains the content and the retrieval procedure of reduced telemetry datasets used with the TipTorch model for direct and calibrated prediction in the case of MUSE-NFM. Some important points regarding this data set are discussed below.

Dataset entry	Archive entry	Data source	Units	Description
RA (science)	RA	Science header	[deg]	Sky coordinates of the science target.
DEC (science)	DEC	Science header	[deg]	
Exp. time	EXPTIME	Science header	[s]	Exposure time.
Tel. altitude	TEL ALT	Science header	[deg]	Telescope pointing.
Tel. azimuth	TEL AZ	Science header	[deg]	
Airmass	TEL AIRM[2]	Science header		Airmass.
Par. angle		Science header	[deg]	Paralactic angle.
Derot. angle		Science header	[deg]	Derotator angle.
Pupil angle		Science header*	[deg]	Angle of derotation of the science target in the pupil plane.
NGS RA	AOS NGS ALPHA / AOS NGS DELTA[4]	Science header	[deg]	Sky coordinates of the NGS target.
NGS DEC				
IRLOS photons	AOS NGSi FLUX	Science header / IRLOS cube*	[ph./s/m^2]	Flux density in the IRLOS focal plane.
NGS mag	AOS NGSi MAG			The apparent magnitude of the NGS.
IRLOS window	AOS MAIN MODES IRLOS / "Main.setMode IRLOS"	Raw header / AOF logs from data lake	[pixels]	IRLOS window size per subaperture in 2×2 mode.
IRLOS loop frequency			[Hz]	IRLOS loop frequency.
IRLOS loop gain				IR loop gain.
IRLOS plate scale		Instrument specifications	[mas/pix]	Angular resolution of the IRLOS detector.
IRLOS conversion			[e-/ADU]	IR detector conversion factor from ADUs to photoelectrons.
IRLOS RON			[e-]	IR detector readout noise.
AOS.LGSi.SLOPERMS	-//-	Data lake	WFS pixels	Average RMS of the WFS slopes.
LGSi_SLOPERMS	-//-	Raw header	WFS pixels	
LGSi_R0	-//-	Raw header	[m]	r_0 estimate from residual phase by AOF
LGSi_SEEING	-//-	Raw header	["]	Seeing estimate from residual phase by AOF
LGSi_STREHL	-//-	Raw header		Strehl estimate from the residual phase by AOF
LGSi photons	AOS.LGSi.FLUX	Data lake	[ph./m^2/s]	LGS flux per WFS for all subapertures
DIMM Seeing	tab_fwhm	asm/dimm	["]	Seeing as measured by DIMM
MASS-DIMM Seeing	tab_fwhm	asm/mass	["]	Seeing as measured by MASS-DIMM
Seeing (header)	TEL AMBI FWHM[3]	Science header	["]	Seeing as measured by MASS-DIMM
MASS-DIMM Tau0	tab_tau0	asm/mass	[s]	τ_0 as measured by MASS-DIMM

Free Atmosphere Seeing	tab_fwhm	asm/mass	["]	Free atmosphere seeing at 500 nm at zenith from MASS stand-alone integrated profile.
MASS Tau0	tab_tau	asm/mass	[s]	τ_0 as measured by MASS
Tau0 (header)	TEL AMBI TAU0	Science header	[s]	τ_0 as measured by MASS-DIMM.
MASS-DIMM Turb Velocity	tab_turb_speed	asm/mass	[m/s]	Characteristic velocity of the turbulence in the whole atmosphere from MASS-DIMM combined profile.
MASS_TURB	-//-[5]	Science header	[m/s]	Average atmospheric turbulence speed measured by MASS.
Wind Speed at 30m	tab_wind_speed1	asm/meteo	[m/s]	Wind speed measured by METEO.
Wind speed (header)	TEL AMBI WINDSP	Science header	[m/s]	
Wind Direction at 30m	tab_wind_dir1	asm/meteo	[deg]	Wind direction measured by METEO.
Wind dir (header)	TEL AMBI WINDDIR	Science header	[deg]	
R0	-//-	Raw headers	[m]	r_0 as measured by MASS-DIMM.
DIMM_SEEING	-//-	Raw headers	["]	seeing as measured by DIMM.
Temperature (header)	TEL AMBI TEMP	Science header	[$C°$]	Temperature measured by METEO.
MASS-DIMM Cn2 fraction at ground	tab_fracgl	asm/mass		Fraction of the total turbulence in the ground layer
Cn2 fraction below 2000m	CN2_FRAC_ALTi / ALTi[6]	Raw headers		Fraction of turbulence below 2000m
Relative Flux RMS	tab_rfl	asm/dimm	[%]	Relative Flux RMS variations as measured by DIMM along the line of sight normalized by the average flux.
MASS_FRACGL	-//-	Raw headers		Fraction of the total turbulence in the ground layer
ASM_RFLRMS	-//-	Raw headers	[%]	Relative Flux RMS variations measured by DIMM.
ASM_WINDSPEED_10	-//-	Raw headers	[m/s]	Wind speed and direction measured by METEO at 10m and 30m altitude
ASM_WINDSPEED_30	-//-	Raw headers	[m/s]	
ASM_WINDDIR_10	-//-	Raw headers	[deg]	
ASM_WINDDIR_30	-//-	Raw headers	[deg]	

1. "-//-" means that the dataset entry is spelled the same as an entry from the archive;
2. Computed as `(TEL AIRM START + TEL AIRM END) / 2` ;
3. Computed as `(TEL AMBI FWHM START + TEL AMBI FWHM END) / 2` ;
4. Computed from alpha [hms] and delta [dms] coordinates for the `NGS AOS NGS ALPHA`, `AOS NGS DELTA` entries, respectively;
5. $\texttt{MASS_TURB} = \sum_{i=0}^{7} \texttt{MASS_TURB}i$;
6. C_n^2 weights, $\text{GL} = \sum_{i=1}^{9} \texttt{CN2_FRAC_ALT}i\,(h) \quad \text{if} \quad h = 100 \cdot \texttt{ALT}i < h_{GL} = 2000\text{m}$

D. NN-calibrator architecture

Layer index	Description	Size / settings	Activation
Input	Input layer	N_{in}	-
1	Fully connected	$N_{\text{in}} \rightarrow N_{\text{hidden}}$	tanh
2	Fully connected	$N_{\text{hidden}} \rightarrow N_{\text{hidden}}$	tanh
3	Fully connected	$N_{\text{hidden}} \rightarrow 2N_{\text{hidden}}$	tanh
4	Dropout	p_{dropout}	-
5	Fully connected	$2N_{\text{hidden}} \rightarrow N_{\text{out}}$	tanh
Output	Output layer	N_{out}	-

TABLEAU 4. – Architecture of the calibrator neural network.

Here, $N_{\text{in}} = \dim(\mathbf{Z})$ is the size of the input layer that corresponds to the reduced telemetry inputs, $N_{\text{out}} = \dim(\mathbf{X})$ is the size of the output layer that corresponds to the PSF model inputs, $N_{\text{hidden}} = 200$ is the number of hidden parameters, $p_{\text{dropout}} = 0.1$ is the fraction of elements that are dropped out during the training.

The same NN architecture is used both for IRDIS and MUSE=NFM. However, N_{in} and N_{out} are different for both caibrators.

E. Computing PSF profiles

The same method is used to calculate the profiles throughout Chapter 4. In this scope, note that :

— The displayed profiles represent the median profile of the dataset, where each individual profile was computed for every predicted PSF and its corresponding on-sky PSF.

— Error bars represent the 1σ deviation from the *median* profile, indicating the variability across the dataset.

— Since all PSFs were centered during the data processing step, profiles were centered relative to the exact center of the images.

— All PSF images were processed to ensure they had an odd number of pixels, ensuring the PSF peak is aligned with a single pixel at the center.

— All profiles are computed for the validation datasets, which comprise 20% of all available PSF samples. The same validation PSFs are used in Sections 4.7, 4.6, and 4.8 to ensure consistency of the results. This split of the dataset is a result of the use of the ML algorithm in Section 4.8, where separate training and validation datasets were required for calibrated prediction.

— For IRDIS, separate profiles are provided for the left and right channels.

— The radial profiles are computed for PSFs preemptively averaged along the spectral dimensions. This is valid for both IRDIS and MUSE.

Profiles computation sequence

1. The `RadialProfile` routine from the `photutils` [6] library was used to compute individual profiles for each PSF in the dataset. This results in the set of on-sky profiles $\{p_{\text{data},i}\}$ and simulated profiles $\{\tilde{p}_i\}$, where :

$$\begin{aligned} p_{\text{data},i} &= \texttt{RadialProfile}(\mathbf{I}_{\text{data},i}), \\ p_i &= \texttt{RadialProfile}(\mathbf{I}_i(\mathscr{X})), \\ i &= 1, \ldots, N_{\text{PSFs}}. \end{aligned}$$

   Where $\mathscr{X}$ are the PSF model inputs, N_{PSFs} is the number of PSFs averaged in the dataset.

2. Each individual profile was divided by the peak value of the on-sky profile and then multiplied by 100% :

$$p'_{\text{data},i} = \frac{p_{\text{data},i}}{\max(p_{\text{data},i})} \cdot 100\%, \quad p'_i = \frac{p_i}{\max(p_{\text{data},i})} \cdot 100\%.$$

3. This normalization allows for calculating the error relative to the on-sky data. The error is computed as the difference between the on-sky and predicted (or fitted) profiles :

$$\Delta p'_i = \mid p'_i - p'_{\text{data},i} \mid$$

   In essence, this error metric corresponds to an error on SR.

6. https://photutils.readthedocs.io/en/stable/

F. Tuned direct prediction values

This section outlines the average assumed values for model inputs in Sections 4.8.4 and 4.8.5.3 for the IRDIS and MUSE-NFM instruments.

F.1. IRDIS

As already elaborated in Section 4.8.4, the assumptions on parameter values made for IRDIS are $F = 1$, $J_x, J_y = 33$ mas, $J_{xy} = 0$ deg, $\Delta n = 0$ rad^2, $\mathbf{c} = \mathbf{0}$. All other parameters were set from the configuration files and reduced telemetry.

F.2. MUSE

Jitter is represented with the empirical function shown below in Figure 4.

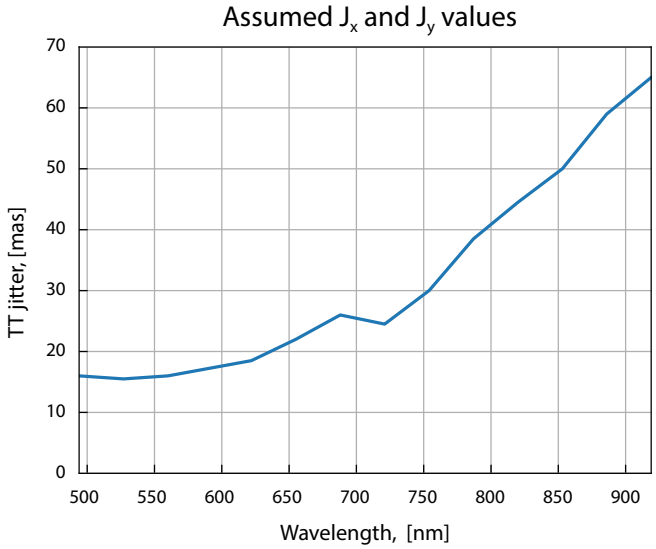


FIGURE 4. – Assumed tip-tilt jitter values for tuned direction prediction.

Other parameters were maintained as constants : $\Delta n = 5.6$ [rad], $F = 0.8, J_{xy} = 0$ [mas]. The Moffat absorber's contribution was canceled by setting $A = 0$. All additional parameters were derived from the configuration files and the reduced telemetry data.

G. Distributions of calibrated model inputs

G.1. SPHERE

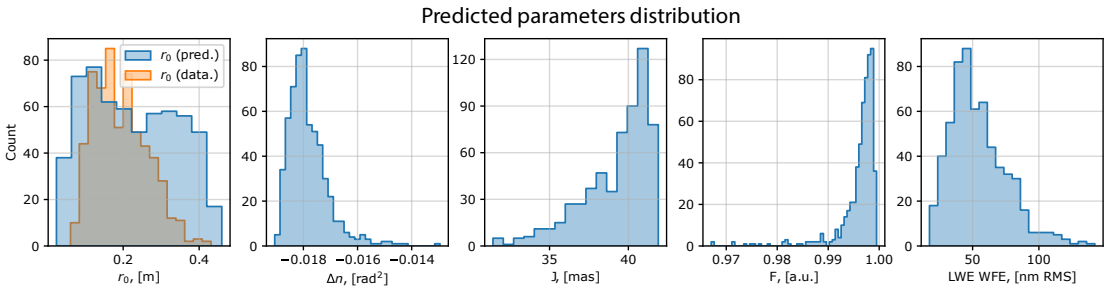


FIGURE 5. – The distribution of values predicted by the NN-calibrator ($\widetilde{\mathscr{X}}$) for IRDIS.

Figure 5 shows the histograms for parameters predicted by the NN calibrator for the validation dataset. Note that the blue histogram on the left corresponds to the predicted r_0 values, while the orange one shows the initial r_0 values taken from the on-sky dataset. The flux normalizing constant F plotted here was computed as $F = (F_L + F_R)/2$, while the LWE-associated WFE is computed as $\|\widetilde{\mathbf{c}}\|$.

The values within distributions indicate that the values for the predicted parameters stay within the physically meaningful ranges.

G.2. MUSE

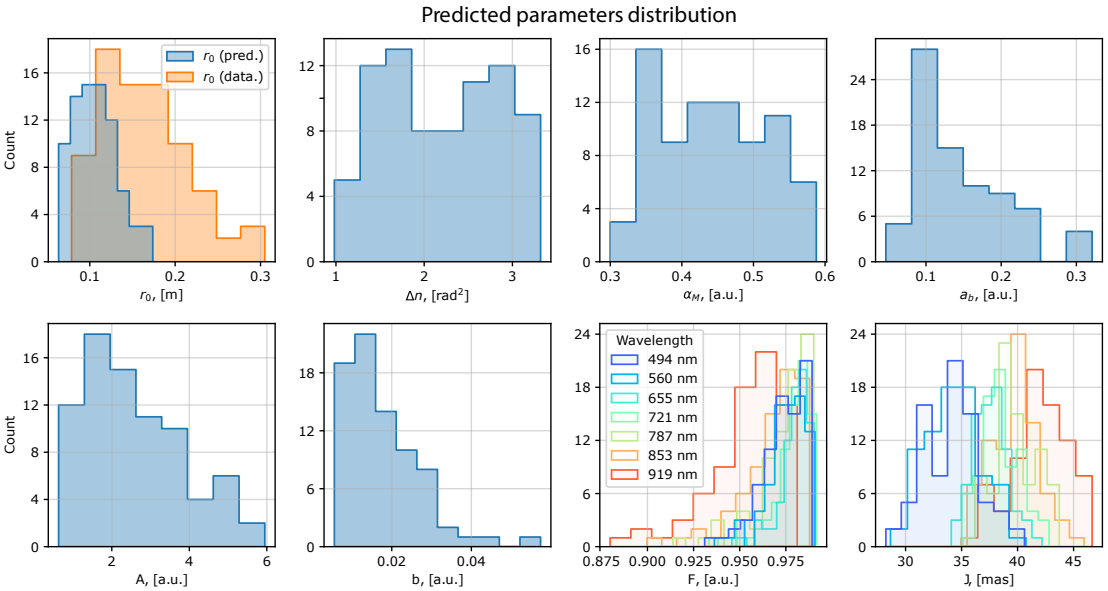


FIGURE 6. – The distribution of values predicted by the NN-calibrator ($\widetilde{\mathscr{X}}$) for MUSE-NFM.

Figure 6 shows the histograms for parameters predicted by the NN calibrator for the validation dataset, this time for MUSE-NFM. Again, the blue histogram on the left top corresponds to the predicted r_0 values, while the orange one shows the initial r_0 values taken from the on-sky dataset. Note that predicted values tend to be more pessimistic than data values computed by SPARTA RTC. The same observation regarding r_0 and seeing values on MUSE has also been previously reported in Fétick et al. (2019b), Figure 8.

The flux normalizing constants F_λ were plotted for each predicted wavelength. The same applies to tip-tilt jitter J_λ, which was computed here as $J_\lambda = (J_{x,\lambda} + J_{y,\lambda}) / 2$.

Like in the previous case of IRDIS, the histograms in Figure 6 indicate that the values for the predicted parameters stay within the physically meaningful ranges.