

Data-calibrated point spread function prediction

General description of the method and demonstration on MUSE-NFM

Arseniy Kuznetsov^{1,2,3,*}, Benoit Neichel², Sylvain Oberti¹, and Thierry Fusco^{2,3}

¹ European Southern Observatory, Karl-Schwarzschild-str-2, 85748 Garching, Germany

² Aix Marseille Univ, CNRS, CNES, LAM, 13013 Marseille, France

³ DOTA, ONERA, Université Paris Saclay, 91123 Palaiseau, France

Received 17 September 2025 / Accepted 19 February 2026

ABSTRACT

Context. Precise knowledge of the point spread function (PSF) underpins many data analysis steps in astronomy, from photometry and astrometry to source de-blending and deconvolution. In adaptive optics (AO) observations, however, the PSF is highly variable with wavelength, field position, and observing conditions, making it difficult to model. Traditional PSF reconstruction (PSF-R) requires full AO telemetry and complex infrastructures, limiting its routine use, especially for tomographic systems.

Aims. We present a practical framework for fast, accurate, and data-calibrated PSF modeling that captures the spatial and spectral variability of AO-corrected PSFs without relying on complete AO telemetry.

Methods. Our approach builds on a Fourier-based PSF model inspired by astro-TIPTOP. As inputs, our model uses only a compact set of physically meaningful parameters retrievable from the ESO archive. A lightweight neural network corrects these inputs to achieve the best match with real data. It is trained end to end with the PSF model, allowing it to learn any miscalibrations directly from on-sky data.

Results. The framework achieves high accuracy on on-sky data. On a test set of MUSE-NFM standard stars, it yields median errors of 13.5% in the Strehl ratio and 10.9% in the core full width at half maximum (FWHM). In crowded MUSE-NFM observations of ω Centauri, the method predicts dozens of off-axis, wavelength-dependent PSFs with a Strehl error of $<5\%$ and a FWHM error of 4.6%, enabling source separation without per-star PSF extraction.

Conclusions. Our compact, physics-informed, and data-calibrated model delivers accurate, polychromatic, and field-varying PSFs without relying on full AO telemetry. While demonstrated on MUSE-NFM, the method is still transferable to other AO-assisted instruments.

Key words. instrumentation: adaptive optics – instrumentation: spectrographs – methods: analytical – methods: data analysis – methods: numerical – techniques: high angular resolution

1. Introduction

In astronomical imaging, the accuracy of scientific analysis depends strongly on the point spread function (PSF). A broadened PSF reduces the clarity of information that can be extracted from observations. For ground-based telescopes, atmospheric turbulence is the main cause of PSF degradation. Adaptive optics (AO) systems counteract this by partially correcting atmospheric turbulence-induced wavefront errors, pushing the attainable angular resolution toward the diffraction limit. However, AO correction is inherently imperfect due to the engineering limitations of the real systems, leaving residual wavefront errors (AO residuals) that prevent the PSF from reaching the diffraction limit. Consequently, in many astronomical applications, the accuracy of scientific analysis remains strongly constrained by PSF quality, but can be greatly enhanced through precise knowledge of the PSF structure, which allows it to be disentangled from the astrophysical signal. For example, when observing the Galactic center, the precise centroiding of stars facilitated by an accurate PSF model can improve the precision for kinematic measurements and thus help one study the interaction of Sgr A* with the surrounding stellar environment. However, the PSF

estimation error of only a few percent can induce astrometric biases that exceed the signal (Neichel et al. 2014), necessitating the accurate characterization of the PSF morphology. Similarly, when studying the stellar kinematics within the dense cores of globular clusters, the precise tracking of stellar trajectories can reveal signatures of intermediate-mass black holes. However, attaining radial-velocity uncertainties below 3 km/s requires PSF full width at half maximum (FWHM) errors of under 2% (Kamann et al. 2013). In such crowded environments, detailed PSF characterization is crucial for de-blending adjacent sources and thereby enhancing both astrometric and photometric accuracy. Extragalactic studies require one to resolve morphological and spectral details to probe internal dynamics, yet errors of over 20% in the PSF FWHM can mimic a galaxy's inclination and bias rotation velocity measurements (Bouché et al. 2015). In integral-field spectroscopy (IFS), imperfect knowledge of PSF elongation can distort metallicity gradients (Epinat et al. 2012). Similarly, imperfect PSF shape knowledge can obfuscate the true shape of compact star-forming clumps, skewing spatially resolved star-formation-rate measurements (Cibinel et al. 2015). Hence, accurately characterizing the PSF's variation over wavelength and field position is vital to separate it from true astrophysical features. When talking about Solar System observations, PSF deconvolution markedly enhances the contrast of

* Corresponding author: akuznets@eso.org

surface features on extended bodies (Fétick et al. 2019a). To ensure reliable restoration, however, PSF shape errors must be below 5% (Lau et al. 2023), as accurate knowledge of the PSF structure can regularize the inherently ill-posed deconvolution problem.

To summarize this discussion, a well-characterized shape of long-exposure PSF can be instrumental in various post-processing applications, including precise photometry and astrometry, source subtraction, deconvolution, and more. The comprehensive list of PSF accuracy requirements for different applications can be found in Neichel et al. (2024), Fig. 9.

A direct approach to characterizing the PSF is to extract it from science exposures using point sources across the field. In practice, however, this is rarely feasible since AO-corrected long-exposure PSFs usually exhibit complex structures that vary significantly with field position, wavelength, and atmospheric conditions. On ELT-class telescopes, for which advanced AO systems will be essential, the PSF complexity will be even greater. Consequently, reference PSFs derived from field reference stars can de-correlate significantly from the science target's PSF, limiting their usefulness for accurate post-processing (Fétick et al. 2019b). The problem is compounded by the frequent absence of isolated point sources, particularly in narrow-field instruments such as MUSE-NFM (Leibundgut et al. 2019) and HARMONI (Thatte et al. 2022), or in deep extragalactic fields.

In summary, advanced post-processing of AO-assisted observation data requires a PSF-modeling framework that captures the full spatial and spectral variability of long-exposure PSFs across the field. In addition, it needs to predict the PSF morphology even in regions without isolated point sources.

2. Comparing approaches to PSF modeling

2.1. PSF reconstruction

To address this challenge, several techniques can be used (see Beltramo-Martin et al. 2020b, for a comprehensive overview). Among these, PSF reconstruction (PSF-R) is particularly prominent. Proposed in Véran et al. (1997), PSF-R exploits AO telemetry, specifically long sequences of wavefront sensor (WFS) measurements and deformable-mirror (DM) commands, to reconstruct the long-exposure AO-corrected PSF. However, despite several demonstrated successful applications (see Harder & Chelli 2000; Martin et al. 2016; Massari et al. 2020; Beltramo-Martin et al. 2019), systematic on-sky deployment of PSF-R has remained elusive over the past three decades, with practical implementations largely confined to experiments on single-conjugate AO (SCAO) systems. This limitation mainly arises from several practical hurdles.

- One of the key practical limitations is the lack of systematic access to raw AO telemetry required for PSF-R. Archiving and transmitting these data require specialized data infrastructure and large storage capacities, which are not feasible on all existing systems. For example, while Keck Observatory routinely records AO telemetry (Ragland et al. 2016) for PSF-R needs, ESO's VLT stores such data only intermittently. The issue of the required volume will be exacerbated on ELT-class systems, for which the sheer scale of AO systems will generate significantly larger telemetry volumes.
- Calibrating PSF-R is another major challenge. For example, PSF-R algorithms are sensitive to errors in determining the C_n^2 profile and WFS gains (see Beltramo-Martin et al. 2019), necessitating calibration to achieve the best match to on-sky data. Furthermore, since PSF-R depends exclusively on

WFS measurements, it cannot recover any phase aberrations that the AO system does not directly sense. Therefore, contributors such as high-order uncorrected atmospheric aberrations, phasing errors on a segmented primary, non-common path aberrations (NCPAs), and the low-wind effect must be incorporated via external simulation or measurement and carefully adjusted against real data to reproduce observed PSFs. Although techniques such as PRIME (Beltramo-Martin et al. 2019) have been proposed to facilitate PSF-R calibration, the process remains labor- and computationally intensive. Moreover, a robust and generalizable PSF-R calibration requires an extensive on-sky dataset that spans various atmospheric conditions and AO-system operating regimes. In practice, however, the limited volume of archived AO telemetry impedes the collection of such a representative dataset. Thus, it forces the initiation of dedicated data-collection campaigns, which are complicated by the scarcity of available observing time on operational telescopes.

- Point spread function reconstruction for tomographic AO is significantly more challenging than for SCAO. Although theoretical extensions of PSF-R exist for tomographic AO (Wagner et al. 2022), the required algorithms and numerical routines are far more complex and computationally demanding. Moreover, tomographic systems that employ multiple WFSs and deformable mirrors generate large volumes of telemetry, which brings us back to the first point. Future instruments such as MAVIS on the VLT and HARMONI/MORFEO on the ELT will rely heavily on this type of AO correction.

Within this framework, the primary objective of this paper is to introduce a robust and practical PSF-modeling strategy that overcomes the listed limitations of conventional PSF-R. Our approach is designed to be data-efficient, computationally lightweight, applicable across diverse AO regimes, and readily calibratable with existing on-sky data. This article presents the general methodology and validates it on ESO's MUSE instrument in the laser tomography AO (LTAO) regime (Oberti et al. 2016). MUSE provides an ideal test case: it is a widely used instrument with a demanding tomographic AO system that is challenging to model, and its broad spectral coverage requires accurate characterization of the polychromatic PSF. These features make MUSE both scientifically important and technically interesting for evaluating the proposed method. Future papers will investigate scientific applications and extend the framework to other instruments, such as SPHERE (Beuzit et al. 2019).

2.2. Fourier-based PSF modeling

As an alternative to PSF-R, Fourier-based PSF modeling was introduced in Rigaut et al. (1998) and later applied in Jolissaint (2010), Fusco et al. (2020), Fétick et al. (2019b), and Neichel et al. (2009). It relies on modeling the power spectral density (PSD) of the AO residuals in the pupil plane, defined in the spatial frequency domain. Instead of reconstructing the AO residuals' PSD from the full telemetry (WFS slopes and DM commands), it computes a theoretical residual-phase PSD from a compact set of integrated atmospheric and system parameters. It includes seeing, coherence time (τ_0), wind profile, guide star positions, and photon flux per WFS subaperture. This makes Fourier-based models far more parsimonious compared to PSF-R, as they rely on a compact set of physically meaningful parameters while retaining broad descriptive power. Crucially, most of these integrated parameters are routinely stored in the

headers of observation FITS files and observatory logs, making them directly accessible from archives and enabling the use of vast amounts of existing data for validation and calibration. Calibration itself is also greatly simplified for Fourier models: a compact parameter space allows these models to be easily optimized to match on-sky data. In addition, the computational cost is modest, as synthetic residual PSDs can be generated far more efficiently with simple analytical laws. Given its efficiency, accessibility, and practicality, Fourier-based modeling provides an ideal foundation for the methods developed in this study.

3. Theory

3.1. PSF formation

To better understand the input space and the calibration of Fourier-based models, it is useful to first review their theoretical foundation. The focal plane PSF at wavelength λ can be computed via the inverse Fourier transform, $\mathcal{F}^{-1}(\cdot)$, of the optical transfer function (OTF):

$$I(\mathbf{x}', \lambda) = F(\lambda) \mathcal{F}^{-1}(\text{OTF}(\mathbf{v}, \lambda)) + b(\lambda). \quad (1)$$

Here, $\mathbf{v} = (u, v)^\top \in \mathbb{R}^2$ is the spatial frequency vector in OTF space ($u^2 + v^2 \leq 1$), $\mathbf{x}' \in \mathbb{R}^2$ denotes image-plane coordinates, F is the flux normalization factor, and b is the constant background offset. The OTF itself can be factorized into distinct components:

$$\text{OTF}(\mathbf{v}) = D_{\text{xy}}(\mathbf{v}) \text{OTF}_{\text{static}}(\mathbf{v}) \text{OTF}_{\text{turb}}(\mathbf{v}) \text{OTF}_{\text{TT}}(\mathbf{v}). \quad (2)$$

For clarity, λ is omitted here but will be reintroduced later. The term D_{xy} represents a focal-plane sub-pixel shift, expressed as

$$D_{\text{xy}}(\mathbf{v}) = \exp\left(-2\pi i \frac{D \alpha_{\text{pix}}}{\lambda} (u \delta x + v \delta y)\right), \quad (3)$$

where δx and δy are sub-pixel displacements in the image plane, α_{pix} is the angular pixel scale on-sky, and D is the telescope aperture diameter. Adjusting this term enables accurate modeling of astrometric shifts.

The term $\text{OTF}_{\text{static}}$ in Eq. (2) accounts for all quasi-static aberrations in the system and is defined as

$$\text{OTF}_{\text{static}}(\mathbf{v}) = \mathcal{P}(\mathbf{x}) \otimes \mathcal{P}^*(-\mathbf{x}), \quad \mathcal{P}(\mathbf{x}) := P(\mathbf{x}) \exp(-i \Phi(\mathbf{x})), \quad (4)$$

where \otimes denotes the autocorrelation, $\mathbf{x} \in \mathbb{R}^2$ are pupil-plane coordinates, P is the pupil mask, and Φ is the pupil-plane phase, which can represent NCPAs or other quasi-static aberrations. This term is revisited in Sect. 4.2.1. The rationale for treating $\text{OTF}_{\text{static}}$ as a separate contributor is outlined in Jolissaint (2010, see Eqs. (1)–(5)).

In contrast, OTF_{turb} represents the impact of AO residuals on the infinite-exposure PSF. As is shown in Appendix A.1, it was derived from the PSD of the residual phase W , defined in spatial-frequency space $\mathbf{k} = (k_x, k_y)^\top \in \mathbb{R}^2$. Modeling W is therefore central to the Fourier-based framework and is examined in detail in Sect. 3.2.

Finally, the last term in Eq. (2) accounts for the residual tip-tilt (TT) jitter, which is the random, rapid displacement of the PSF across the focal plane. A 2D Gaussian convolution kernel can approximate its long-exposure effect:

$$\text{OTF}_{\text{TT}}(\mathbf{v}) = \exp\left(-\frac{\pi^2}{4 \ln 2} \left(\frac{D}{\lambda}\right)^2 (J_x^2 u^2 + J_y^2 v^2)\right), \quad (5)$$

$$\mathbf{v}_r := (u_r, v_r)^\top = \mathbf{T}(\theta_J) (u, v)^\top.$$

Here, J_x and J_y denote the FWHM of the tip and tilt jitter, respectively. Unlike the formulation in Eq. (10) of Beltramo-Martin et al. (2020a), the kernel is expressed in a rotated, axis-aligned coordinate frame rather than through the TT covariance matrix, with θ_J defining the rotation angle and \mathbf{T} the corresponding rotation matrix (see Eq. (A.2)).

Although both OTF_{TT} and OTF_{turb} arise from imperfect AO correction, separating them is well justified. First, in many instruments (including MUSE-NFM), high-order correction is enabled by laser guide stars (LGSs), which are insensitive to TT. Tip and tilt are instead measured by dedicated low-order (LO) WFSs, whose contribution must be modeled separately. Second, Fourier-based models reproduce high spatial frequencies well but poorly capture the low-frequency region of the PSD, which contains information about TT. Finally, isolating the TT component allows for it to be straightforwardly calibrated by tuning J_x and J_y independently from OTF_{turb} .

3.2. Simulating AO residuals

The key remaining question is how to compute the term W that characterizes the statistics of AO residuals. This derivation begins by identifying the main contributors to imperfect AO correction, as outlined in Appendix A of Jolissaint et al. (2006) and Sect. 2 of Correia et al. (2017):

- Fitting error: the DM can only correct spatial frequencies, \mathbf{k} , below its cutoff frequency $k_c = (2 d_{\text{act}})^{-1}$, which is set by the actuator spacing, d_{act} .
- WFS noise: arising from photon noise (limited photon flux per integration time) and detector noise.
- Servo-lag error: introduced by the temporal delay between wavefront measurement and DM correction, due to WFS integration and the computation time of the corrective commands.
- Anisoplanatic error: occurring when the wavefront in the science direction de-correlates from that measured along the guide-star direction with increasing angular separation.
- Aliasing error: the WFS samples the wavefront at finite resolution, limited by its Nyquist frequency $k_{\text{WFS}} = (2d)^{-1}$, where d is the subaperture size. Higher spatial frequencies cannot be sensed, and therefore remain uncorrected, folding back into the lower-frequency domain.
- Chromatic error: caused when the WFS operates at a wavelength, λ_{WFS} , different from the science wavelength, leading to dispersion-related mismatches.

In Fourier-based PSF modeling, these contributions are treated as statistically independent error terms. This framework was first introduced in Rigaut et al. (1998), extended to two-dimensional and polychromatic formulations in Jolissaint et al. (2006), Jolissaint (2010), and Jolissaint & Kendrew (2010), adapted to XAO systems in Correia & Teixeira (2014) and Correia et al. (2017), and further applied in Beltramo-Martin et al. (2020a). Its generalization to tomographic systems was developed in Neichel et al. (2009).

Building on this extensive theoretical groundwork, we provide the explicit expressions of the simulated error terms without derivations. The more detailed explanations are included in Appendix A.3.

3.2.1. Fitting error

Since DM cannot correct the spatial frequencies above its cutoff frequency, k_c , it is possible to split the residuals PSD into two

distinct regions containing corrected ($\|\mathbf{k}\| \leq k_c$) and uncorrected ($\|\mathbf{k}\| > k_c$) spatial frequencies, where a pure open-loop atmospheric spectrum, W_φ , governs the latter region (see Eq. (A.3)). Then, W_{fitting} can be expressed as

$$W_{\text{fitting}}(\mathbf{k}) = \begin{cases} 0 & \text{if } \|\mathbf{k}\| \leq k_c, \\ \tilde{\Pi}(\mathbf{k}) W_\varphi(\mathbf{k}) & \text{otherwise.} \end{cases} \quad (6)$$

Here, $\tilde{\Pi}$ is the piston mode filter (see Appendix A.3.1 for more details).

3.2.2. Wavefront sensor noise error

This error term originates from noise in the WFS detectors, which propagates through the wavefront reconstruction. Since MUSE-NFM relies on tomographic correction, the expression here can be adapted from the second term of Eq. (23) in Neichel et al. (2009), yielding

$$W_\eta(\mathbf{k}) = \tilde{\Pi}(\mathbf{k}) (\mathbf{P}_\theta^{\text{DM}} \mathbf{W}) \mathbf{C}_b (\mathbf{P}_\theta^{\text{DM}} \mathbf{W})^\top, \quad (7)$$

where $\mathbf{P}_\theta^{\text{DM}}$ is the vector¹ that projects the DM phases seen in the direction, θ , of the science object in the pupil plane, in which the tomographic solution is optimized (see Eq. (10) from Neichel et al. 2009), \mathbf{W} is the phase volume reconstruction matrix defined in Eq. (13) in Neichel et al. (2009), and \mathbf{C}_b is the WFS noise variance matrix (see Eq. (A.4)).

3.2.3. Spatio-temporal error

This error term represents a combination of anisoplanatic² and servo-lag errors. As was noted in Jolissaint (2010), these two contributions are coupled. Under the frozen-flow approximation (Taylor hypothesis), a temporal shift of the wavefront can be partially compensated for by a spatial shift of the opposite sign. The first term of Eq. (23) in Neichel et al. (2009) can therefore be adapted to describe this error in terms of a PSD:

$$W_{\text{ST}}(\mathbf{k}) = \tilde{\Pi}(\mathbf{k}) \mathbf{U} \mathbf{C}_\varphi \mathbf{U}^\top, \quad \mathbf{U} := \mathbf{P}_\theta^L - \mathbf{P}_\theta^{\text{DM}} \mathbf{W} \mathbf{M} \mathbf{P}_\alpha^L. \quad (8)$$

Here, \mathbf{P}_θ^L projects turbulent layers in the direction, θ , onto the pupil plane, \mathbf{P}_α^L projects turbulent layers in the directions, α , of the guide stars (GSs), \mathbf{M} is the WFS measurement operator matrix, originally defined in Rigaut et al. (1998) and Jolissaint et al. (2006) and later extended to tomographic systems in Sect. 6 of Neichel (2008). Further details on these quantities are provided in Appendix A.3.3. Note that \mathbf{P}_θ^L , \mathbf{P}_α^L , $\mathbf{P}_\theta^{\text{DM}}$, \mathbf{M} , \mathbf{W} , and \mathbf{C}_φ are all functions of \mathbf{k} , though the notation is omitted here for brevity.

3.2.4. Aliasing

As was noted earlier, this error arises when the WFS under-samples the high-spatial-frequency component of the wavefront. Because the wavefront is sampled by a finite number of subapertures or pixels, spatial frequencies above the WFS Nyquist limit are not measured and instead fold back into lower frequencies during discretization. The control system then attempts to correct

¹ For on-axis LTAO correction, where a single DM is conjugated to the ground layer, $\mathbf{P}_\theta^{\text{DM}} = \mathbf{1}$.

² In tomographic AO, there is no anisoplanatism in the strict SCAO sense. Instead, it is incorporated into the tomographic reconstruction.

these spurious low-frequency signals, introducing an aliasing error.

For this term, we departed from the tomographic framework and adopted an SCAO formulation. Although aliasing derivations for tomographic reconstruction are presented in (Neichel 2008, Sect. 7.6.4 and Annexe B.1), the SCAO treatment is computationally more efficient.

Accordingly, we adopted Eq. (19) from Correia et al. (2017), re-expressing its terms with the notation of Correia & Teixeira (2014), to obtain

$$W_A(\mathbf{k}) = \sum_m \sum_n |Q(\mathbf{k}) A_\varphi(\mathbf{k})|^2 \tilde{\Pi}(\mathbf{k} + \mathbf{s}/d) W_\varphi(\mathbf{k} + \mathbf{s}/d), \quad (9)$$

where $\mathbf{s} = (n, m)^\top \in \mathbb{N}^2 \setminus \{(0, 0)^\top\}$; Q and A_φ functions are defined in Appendix A.3.4.

3.2.5. Chromatism error

The PSD terms described above are assumed to be wavelength-independent (neglecting atmospheric dispersion) for computational simplicity. In practice, however, weak dispersion is always present and must be considered, particularly for strongly multi-spectral instruments such as MUSE.

One such effect is the chromatism error. Following Eqs. (12)–(13) of Fusco et al. (2006) and Eq. (3) of Jolissaint & Kendrew (2010), it can be written as

$$W_{\Delta n}(\mathbf{k}, \lambda) = \left(1 - \frac{\Delta n(\lambda)}{\Delta n(\lambda_{\text{GS}})}\right)^2 \tilde{\Pi}(\mathbf{k}) W_\varphi(\mathbf{k}). \quad (10)$$

Here, λ_{GS} is the WFS wavelength, $\Delta n(\lambda) := n(\lambda) - 1$ is the refractivity, and $n(\lambda)$ is the refractive index of air according to Ciddor (1996). See Appendix A.3.5 for a more detailed description.

3.2.6. Differential refraction error

A second chromatic error contributor is the differential refraction error. This effect can be formulated as an anisoplanatism-like term by simplifying Eqs. (7)–(9) of Jolissaint & Kendrew (2010). As with W_A , we adopted an SCAO formulation for computational efficiency, yielding

$$W_{\text{DR}}(\mathbf{k}, \lambda) = \sum_{l=1}^{N_L} 2 \mathbf{w}[l] (1 - \cos(\mathbf{k} \cdot \boldsymbol{\Theta}_{\text{DR}, l}(\lambda))) \tilde{\Pi}(\mathbf{k}) W_\varphi(\mathbf{k}). \quad (11)$$

For more details, see Appendix A.4.

3.2.7. Computing the total AO residuals' PSD

Assuming the statistical independence of the individual PSD terms described above, the total PSD W_{total} was computed as the sum of the individual contributors:

$$W(\lambda) = \left(\frac{\lambda_{\text{atm}}}{2\pi}\right)^2 \left(W_c(\lambda) \Big|_{\|\mathbf{k}\| \leq k_c} + W_{\text{fitting}} \Big|_{\|\mathbf{k}\| > k_c}\right), \quad (12)$$

$$W_c(\lambda) := W_A + W_{\text{ST}} + W_\eta + W_{\Delta n}(\lambda) + W_{\text{DR}}(\lambda) + \Delta W.$$

In Eq. (12), \mathbf{k} is omitted for clarity. The prefactor $(\lambda_{\text{atm}}/2\pi)^2$ converts the residual phase PSD from $\text{rad}^2 \text{m}^{-2}$ into an optical path difference (OPD) PSD in square nanometers per square meter. The additional term, ΔW , represents unexplained residual variance. Its role is discussed later, in Sect. 4.2.2.

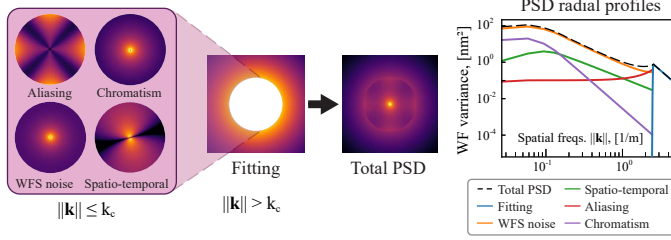


Fig. 1. Illustration of Eq. (12). W_{DR} and ΔW are omitted for brevity.

With this, the theoretical framework is complete. The identification of the main contributors to AO residuals, together with their corresponding analytical expressions, now provides the basis for defining the input space of the PSF model.

3.3. Identifying model inputs

Based on the theoretical framework outlined in Sect. 3 and Appendix A, we identified the set of variable dynamic parameters, \mathcal{Y}_d , that define the PSF under different observing conditions for a given instrument and AO regime. These parameters vary from one observation to another:

$$\mathcal{Y}_d = \left\{ \begin{array}{l} J_x(\lambda), J_y(\lambda), \delta x(\lambda), \delta y(\lambda), F(\lambda), b(\lambda), \\ \Phi(\mathbf{x}, \lambda), \theta_J, \theta_{zen}, \boldsymbol{\theta}, \sigma_{ph}^2, \Delta\sigma^2, \\ r_0(\lambda), L_0, N_L, \mathbf{h}, \mathbf{w}, \mathbf{V} \end{array} \right\}. \quad (13)$$

A second set, \mathcal{Y}_s , contains static parameters that are intrinsic to MUSE-NFM and remain fixed across observations. These are well-characterized system properties:

$$\mathcal{Y}_s = \left\{ \begin{array}{l} d_{act}, d, \lambda_{atm}, \lambda_{GS}, N_{GS}, \boldsymbol{\alpha}, P(\mathbf{x}), \\ \alpha_{pix}, \sigma_{det}^2, D, \Delta t, \zeta \end{array} \right\}. \quad (14)$$

Assuming $\Phi = \mathbf{0}$ and $\Delta W = \mathbf{0}$, the union $\mathcal{Y}_d \cup \mathcal{Y}_s$ fully determines the PSF. Notably, almost all elements of these sets are integrated quantities that can be derived from raw AO telemetry. In practice, model calibration and tuning to specific atmospheric conditions require adjusting only \mathcal{Y}_d given the same instrument and AO regime. The glossary with all parameter definitions can be found in Table B.1.

This discussion further highlights both the data efficiency and the flexibility of the Fourier modeling approach. Moreover, clarifying the input space establishes the basis for the calibration methodology presented in the following section.

4. Data-calibrated PSF modeling

4.1. Motivation

The ease of calibration was previously identified as one of the key factors in selecting the Fourier-based PSF modeling framework. Hence, it is essential to clarify the notion of calibration in this context and motivate it.

- Uncertainty in input parameters: in practice, integrated parameters are often imperfectly known and affected by measurement errors (see Fig. 8 of Fétick et al. 2019b).
- Telemetry processing: since these parameters are derived from raw telemetry, this process can introduce errors.
- Model simplifications: the Fourier-based approach necessarily relies on approximations, whereas real AO systems are highly complex and include many poorly characterized,

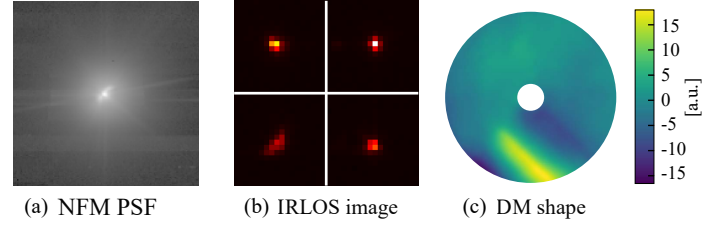


Fig. 2. (a) Effect of the phase bump on MUSE-NFM PSFs (log scale). (b) Phase bump observed by the IRLOS WFS. (c) Measured phase bump.

sometimes dynamic, effects. Calibration allows the PSF model to absorb these discrepancies from data to provide a more realistic PSF representation.

To summarize, in this context, calibration means compensating for systematic biases and unmodeled effects in the data and model to achieve a more realistic representation of PSF morphology. Its role is further demonstrated in Appendix D.2. Here, we first specify which components of the model are subject to calibration.

4.2. Identifying calibratable quantities

From the analysis in Sect. 3, four domains emerge as candidates for calibration:

- the vector of input variables, \mathcal{Y}_d ,
- the static pupil-plane phase, $\Phi(\mathbf{x}, \lambda)$,
- the AO residuals' PSD, $W(\mathbf{k}, \lambda)$,
- the focal-plane intensity distribution, $I(\mathbf{x}')$.

In this case, adjusting \mathcal{Y}_d propagates through all four domains. In addition, a phase term, Φ , can capture NCPAs, and an additive PSD term, ΔW , can represent unexplained residual variance. In principle, a focal-plane correction, ΔI , could account for detector-specific artifacts, but here we assume that no such artifacts are present.

4.2.1. Calibrating pupil phase

In the case of MUSE-NFM, static phase calibration must account for a specific instrument-related effect known as the “phase bump.” This NCPA appeared between late 2021 and early 2024 due to incorrect reference-slope calibration in one of the WFSs, caused by background biases in a specific octant of the LGS WFS detector. It manifested as a quasi-static phase structure (Fig. 2) whose strength varied from observation to observation. The morphology of the phase bump was measured separately (see Fig. 2c) and incorporated in the PSF model in the pupil phase. Then, the term Φ in Eq. (4) can now be defined as

$$\Phi(\mathbf{x}, \lambda) = \frac{2\pi}{\lambda} P_b(\mathbf{x}) a_b,$$

where P_b is the pupil-plane OPD map of the measured phase bump (Fig. 2c), and a_b is a scaling factor that sets its amplitude. Other NCPAs may also contribute to the pupil phase, as is discussed in Kuznetsov et al. (2024). For simplicity, however, we assume here that the phase bump is the sole source of quasi-static phase error.

4.2.2. Calibrating residuals PSD

To account for errors in the residual PSD, we adopted a simplified 2D Moffat function (Moffat 1969) as a flexible and

calibratable ad hoc PSD error absorber. It can be expressed as

$$\Delta W(\mathbf{k}) = \tilde{\Pi}(\mathbf{k}) \left(\frac{A (\beta_M - 1) \alpha_M^{2(\beta_M - 1)}}{\pi (\alpha_M^2 + \|\mathbf{k}\|^2)^{\beta_M}} + B \right), \quad (15)$$

where A and B denote amplitude and background, while α_M and β_M control the function's width and steepness, respectively.

4.2.3. Calibrating model inputs

Absorbers act as subtle, ad hoc extensions to the otherwise fully physics-based PSF model, enhancing its representativeness. However, like the primary model inputs, they must also be calibrated. Because both Φ and ΔW are expressed in parametrized form, their corresponding parameters can be incorporated into the model's input space. Thus, the entire problem of PSF model calibration reduces to calibrating the model's input domain. We therefore defined the set of calibratable parameters, \mathcal{Y}_c , as

$$\mathcal{Y}_c = \{F(\lambda), J_x(\lambda), J_y(\lambda), \Delta\sigma^2, r_0, \mathbf{w}, A, B, \alpha_M, a_b\}.$$

This set includes only a subset of inputs from \mathcal{Y}_d . For instance, β_M and θ_j are not included, as they are degenerate with other inputs and cannot be uniquely recovered. Likewise, θ , $\delta x(\lambda)$, $\delta y(\lambda)$, and $b(\lambda)$ are not calibratable, since they vary unsystematically for different targets. The systematic shifts in δx , δy due to dispersion and field distortion are effectively corrected by the MUSE data reduction pipeline. The error in the determination of photon noise variance, σ_{ph}^2 , is absorbed by the $\Delta\sigma^2$ term. Telescope pointing is assumed to be perfectly known. Varying the number of turbulence layers, N_L , breaks the model's differentiability, and optimizing the height of the atmospheric layers, \mathbf{h} , introduces strong coupling with \mathbf{w} . Finally, we excluded \mathbf{V} and L_0 , whose effect on the PSF is minor in the case of MUSE-NFM and could lead to overfitting.

4.3. Calibration dataset

After defining the model input space and calibratable variables, it is necessary to identify the sources of parameter values. As was mentioned previously, some quantities can be found in the reduced telemetry archive, which encompasses the following data sources:

- FITS headers of observations (after running ESO's MUSE data reduction pipeline). This is the primary source of information on observation conditions and parameters.
- Raw FITS files of observations before running the pipeline. They contain additional information that is lost during the data reduction pipeline. For example, raw headers contain C_n^2 profiles or LO WFS images.
- Logs generated by the UT4 telescope and the MUSE instrument. They are used to complement information missing from FITS headers, such as LO AO loop parameters, natural guide star (NGS) parameters, and other relevant entries.

To implement the calibration, a set of real on-sky PSFs is required as a reference. These PSFs can then be paired with reduced telemetry to form the calibration dataset. For this purpose, we used spectrophotometric standard stars, which have been routinely observed with MUSE-NFM since 2018 and are archived in the ESO science database. These observations capture bright stars under a wide range of atmospheric conditions. Their abundance, high signal-to-noise ratio (S/N), and strictly

point-source nature make them an excellent reference for PSF model calibration. The resulting dataset, \mathcal{D} , was then defined as

$$\mathcal{D} = \{(\mathcal{X}^{(k)}, I_{\text{data}}^{(k)}(\mathbf{p}, \lambda))\}_{k=1}^{N_I}.$$

Each k -th sample in the dataset consists of an observed PSF, I_{data} , and a corresponding set of reduced telemetry, \mathcal{X} , recorded during the exposure. In total, $N_I = 392$ samples were collected, excluding those with corrupted PSFs or missing telemetry. The continuous image-space coordinates, $\mathbf{x}' \in \mathbb{R}^2$, are represented here by discrete focal-plane pixel indices, $\mathbf{p} \in \Omega$, where $\Omega = 1, \dots, H \times 1, \dots, W$ denotes the pixel grid of an image with a height, H , and width, W . All I_{data} images were background-subtracted and spectrally binned. The original $N_\lambda = 3681$ chromatic slices from MUSE-NFM were averaged into $N_\lambda = 7$ evenly spaced bins of $\Delta\lambda \approx 65$ nm, improving the S/N and reducing computational cost.

Here, we also introduced an additional set of parameters, \mathcal{X} . In contrast to \mathcal{Y}_s and \mathcal{Y}_d , which define the PSF model input space, \mathcal{X} comprises integrated telemetry recorded by various instrument subsystems ($\mathcal{X} \neq \mathcal{Y}_d \cup \mathcal{Y}_s$). These include

- atmospheric parameters (r_0 , L_0 , τ_0 , full C_n^2 profile, \mathbf{V} at different altitudes), which are measured both by the AO system of UT4 and by DIMM/MASS-DIMM instruments;
- LO and LGS loop parameters (loop gain and frequency, guide star fluxes);
- telescope pointing (θ_{zen} , airmass);
- ambient conditions (pressure, temperature, humidity);
- other auxiliary quantities (such as the RMS of WFS slopes and more).

At the same time, \mathcal{X} excludes some model-specific and non-systematic quantities such as δx , δy , and b , as well as parameters that are calibrated ($\mathcal{X} \cap \mathcal{Y}_c = \emptyset$). As a result, \mathcal{X} is redundant in the sense that it contains more logged parameters than is strictly required by the PSF model, yet it is incomplete with respect to the full set of model inputs.

4.4. Calibrated PSF prediction

The structure of the problem, with its data redundancy and the use of ad hoc error absorbers, naturally implies a data-driven calibration strategy based on machine learning (ML). In this framework, the calibration of the PSF model can be formulated as the training of a calibrator function that predicts³ the values of calibrated inputs, \mathcal{Y}_c , that minimize the discrepancy between the observed data I_{data} and the model. We further hypothesize that the redundant entries in \mathcal{X} may contain latent information that enables the calibrator to capture relationships in the data that are not analytically tractable.

The resulting calibrated PSF model prediction can thus be expressed as

$$\tilde{I}(\mathbf{x}', \lambda) = f(\mathbf{x}', \lambda; \mathcal{Y}_s \cup \mathcal{Y}_d \oplus \tilde{\mathcal{Y}}_c), \quad \tilde{\mathcal{Y}}_c = g(\mathcal{X}; \gamma). \quad (16)$$

Here, $f(\cdot)$ is the PSF model, $g(\cdot)$ is the calibrator with trainable parameters, γ , the “ $\tilde{\cdot}$ ” notation stands for predicted quantities, $\tilde{\mathcal{Y}}_c$ is the predicted set of calibrated model inputs, $\tilde{I}(\mathbf{x}', \lambda)$ is the predicted PSF, and \oplus is the right-biased union operator, which updates overlapping entries from \mathcal{Y}_d with the values from \mathcal{Y}_c . For two arbitrary sets, A and B , it is defined as $A \oplus B = \{A \setminus B\} \cup B$. Note that the values in \mathcal{X} are standardized (zero mean and unit variance) before being input to the calibrator.

³ Throughout this work, “prediction” does not denote forecasting, but simply the inference by a trained calibrator model.

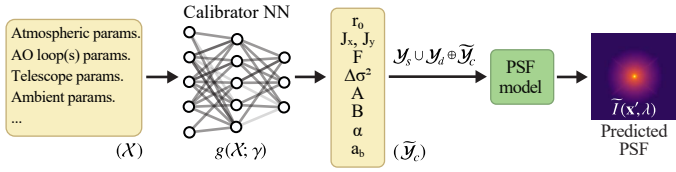


Fig. 3. Schematic of the data-calibrated PSF predictor for MUSE-NFM.

4.5. Implementation

After outlining the general concept behind the calibration approach, the actual implementation can be discussed. In this work, we propose a hybrid approach introduced in Kuznetsov et al. (2023, see our Fig. 3) that combines a data-driven calibrator, implemented as a neural network (NN), followed by a fully analytical PSF model that implements the principles outlined earlier in Sect. 3.

This hybrid design is motivated by the limited availability of independent on-sky training data. Although \mathcal{X} spans a high-dimensional parameter space, the size of the training set, \mathcal{D} , is small, necessitating strong regularization during calibrator training. This is achieved by limiting the calibrator NN to learning only the low-dimensional mappings between \mathcal{X} and $\tilde{\mathcal{Y}}_c$, while the analytical model accounts for the physics of PSF formation. It keeps the number of trainable parameters small. In addition, because the calibrator is directly coupled to a physics-driven model, only physically meaningful predictions result in realistic PSFs, which provides an additional source of regularization.

As a bonus, in this approach, the calibrator can be analyzed independently of the PSF model using feature-attribution tools, such as SHAP (Lundberg & Lee 2017), to probe the learned interactions among the reduced telemetry and model inputs. This aspect will be further explored in future work.

An NN-based solution was chosen for the calibrator due to its flexibility and ability to approximate smooth, differentiable functions efficiently⁴. The architecture is a feed-forward NN with three hidden layers of 200 neurons each, using hyperbolic tangent activation functions. The input and output dimensionalities are 61 and 27, respectively, and a dropout layer with a probability of 10% is applied before the final layer.

An obvious limitation of the proposed implementation is its restricted ability to reproduce intricate, unexplained morphological features of the PSF due to the strong parametrization assumptions imposed on the ΔW and Φ terms. However, such structures were not observed on MUSE-NFM PSFs. Moreover, as more training data accumulates in the future, additional optimizable degrees of freedom could be introduced, enabling more flexible error-absorbing components.

In the long term, a fully data-driven approach could be pursued. For example, by employing a generative model (as in Jia et al. 2021 and Long et al. 2019) to infer the PSF directly from reduced telemetry. Such a model, however, would require orders of magnitude more training data than is currently available. One possible pathway would be to pretrain such a model on synthetic data from end-to-end AO simulations and subsequently fine-tune it on on-sky observations. While this strategy would maximize flexibility by learning unexplained effects directly from data without relying on strong analytical assumptions, it would do so at the cost of interpretability and significantly higher computational demands during training.

⁴ The proposed scheme is not limited to NNs and can utilize other types of regression models, provided that they are differentiable.

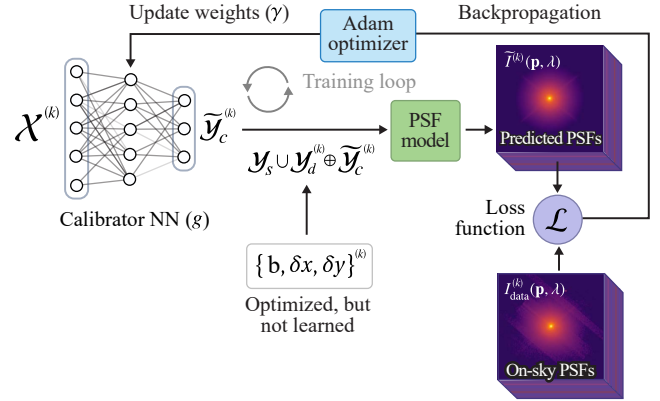


Fig. 4. Training data-driven PSF predictor.

4.5.1. Training the calibrator

The data-driven calibrator, $g(\mathcal{X}^{(k)}; \gamma)$, was trained to minimize the loss function, \mathcal{L} , between the measured PSFs and the corresponding model predictions, where γ denotes the trainable weights and biases of the calibrator. In total, 314 samples from \mathcal{D} were used for training, while 78 were reserved for validation. Further details are provided in Appendix C.

During the training, as is depicted in Fig. 4, parameters $b(\lambda)$, δx , δy are optimized but are not predicted by the calibrator, since they do not vary systematically (as is briefly mentioned in Sect. 4.3). The calibrator is optimized jointly with the analytical PSF model: gradients are backpropagated through the physics-based model, which acts as a regularizer.

4.5.2. Code

Implementing the proposed hybrid approach, which combines ML with analytical modeling, required the development of dedicated software. To this end, we developed TipTorch, a novel open-source simulation tool (see the software availability section at the end of the paper).

The analytical core of TipTorch is inspired by the P3 library, now integrated into the astro-TIPTOP package (Neichel et al. 2021), which currently serves as the baseline PSF simulation tool for the ERIS, MAVIS, and HARMONI instruments. However, one of TipTorch's key features is its full differentiability. This is achieved with automatic differentiation (AD), implemented by leveraging the PyTorch framework (Paszke et al. 2019). A similar idea of differentiable PSF modeling combined with learned corrections was also previously explored in Liaudat et al. (2023), Desdoigts et al. (2023), and Desdoigts et al. (2024). In our case, using AD is essential for enforcing physics-based regularization via backpropagation through the PSF model.

In addition to AD, the PyTorch-based implementation supports highly parallelized simulations that can be flawlessly executed on both the CPU and GPU backends, enabling TipTorch to simulate dozens of polychromatic PSFs per second on desktop-level hardware⁵. The code's efficiency enables coverage of the full MUSE-NFM wavelength range and the simulation of dense stellar fields.

5. Results

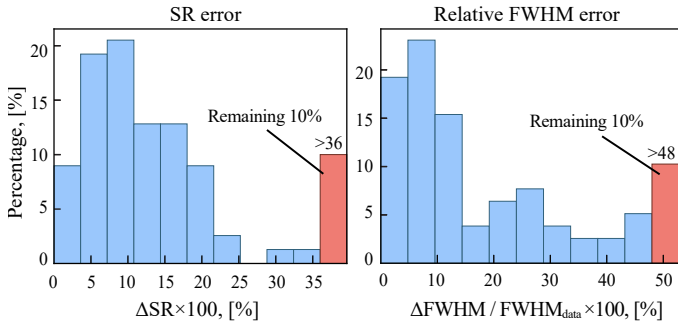
5.1. Standard stars

We assessed the performance of the proposed method on the validation subset of \mathcal{D} . Table 1 summarizes how well the calibrated

⁵ The NVIDIA RTX A4000 Laptop GPU was used for these tests.

Table 1. Median validation metrics for the calibrated model. Lower values indicate better performance.

Metric	Median value
ΔSR (%)	13.5
$\Delta\text{FWHM}_{\text{rel}}$ (%)	10.9
ΔFWHM (mas)	9.1
FVU, [%]	3.2

**Fig. 5.** Distributions of SR (left plot) and FWHM (right plot) errors for the calibrated model.

model reproduces the on-sky PSF peak, core width, and overall PSF structure.

Here, the Strehl ratio error (ΔSR) denotes the peak difference between predicted and on-sky PSFs, while the fraction of variance unexplained (FVU) was computed as in Eq. (D.1). The relative and absolute FWHM errors, $\Delta\text{FWHM}_{\text{rel}}$ and ΔFWHM , respectively, quantify the difference in PSF core width; the FWHM was derived from a 1D Moffat profile fit to radial cuts of I_{data} and \tilde{I} within 5 pixels of the PSF center. For each validation sample, metrics were first calculated per (binned) spectral slice, then averaged, and finally computed across the dataset by taking the median. More information on SR and FWHM error distributions can be found in Fig. 5.

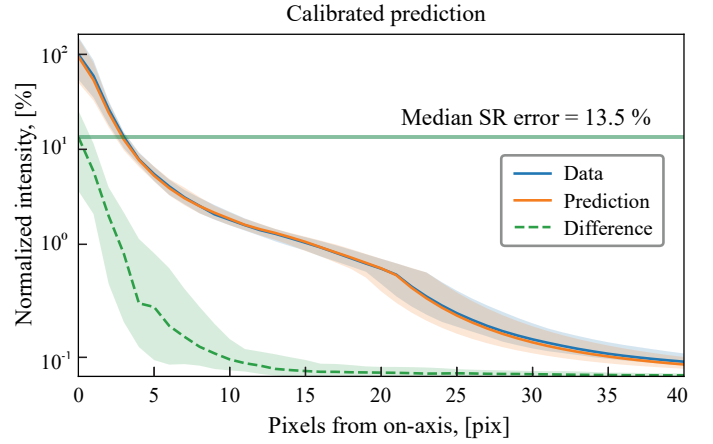
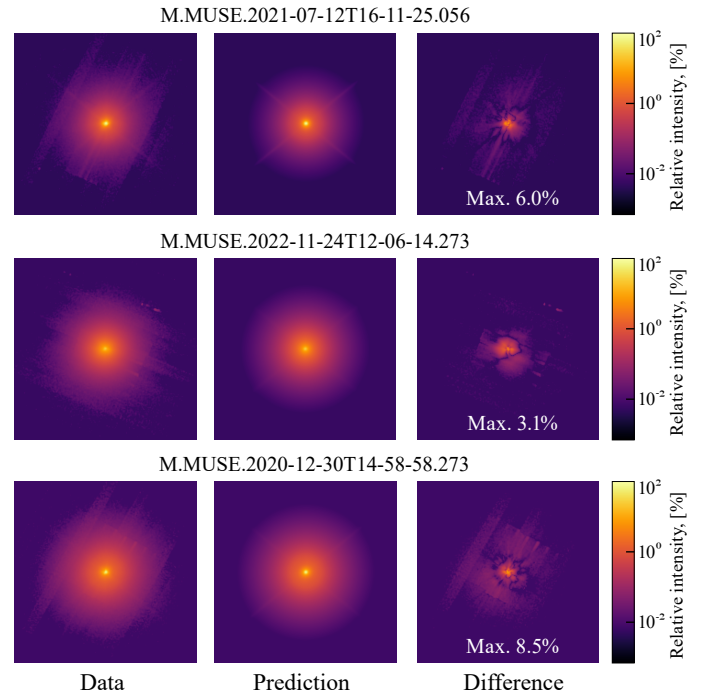
Before metrics evaluation, both observed and predicted PSFs were normalized per spectral slice to the unit flux,

$$\sum_{\mathbf{p} \in \Omega} I_{\text{data}}(\mathbf{p}, \lambda) = 1, \quad \sum_{\mathbf{p} \in \Omega} \tilde{I}(\mathbf{p}, \lambda) = 1.$$

For illustration, Fig. 6 shows the median radial PSF profiles from the validation set, computed with the `photutils` Python library. Profiles were normalized to the peak of the observed PSF and scaled to percent. Table 1 and Figs. 6 and 7 demonstrate an agreement between the predicted and on-sky PSFs, indicating a good overall performance of the method (see also Appendix D.2 for a comparison with the uncalibrated case).

5.2. ω Centauri cluster

The validation dataset described above is limited in scope and serves mainly as a proof of concept. To further assess the method on scientifically relevant data, we applied it to observations of the crowded stellar field in the core of the ω Centauri cluster. The dataset, provided by Sebastian Kamann (Pechetti et al. 2024), is particularly well suited for testing because a single exposure contains multiple point sources, enabling an assessment of both prediction accuracy and field-dependent variability within the $7.5'' \times 7.5''$ field of view. Figure 8 shows the PSF prediction

**Fig. 6.** Median radial PSF profiles of the validation dataset. Error bars indicate the 1σ dispersion across the dataset. The profiles are averaged across all wavelengths. The apparent “kink” in the curves arises from the logarithmic scaling of the plot, which transitions to a linear scale just below the 1% level.**Fig. 7.** Three example PSFs from the validation dataset. Left: data. Middle: calibrated prediction. Right: residuals. All images are shown on a logarithmic scale. All PSFs are spectrally averaged.

results for the 54 brightest sources in the field. Radial profiles, computed in the same way as for Fig. 6, are shown in Fig. 9. The results demonstrate an excellent predictive performance, with a median ΔSR of less than 5%. Accuracy is higher than for the standard-star dataset, likely because the observing conditions in this dataset are well represented in the training data, whereas the standard-star validation set covers a broader range of conditions. The observation was obtained at a seeing of $0.54''$, with an NGS J-band magnitude of 13.6 and an isoplanatic angle of $1.9''$. In Fig. 9, residual “flat” discrepancies are primarily due to the faint background sources that were not included in the simulation. Although the calibrator was not explicitly trained to capture

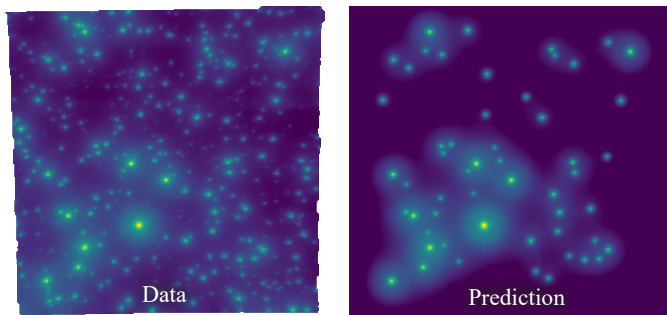


Fig. 8. Comparison between observed data (left) and predicted PSFs (right) for ω Centauri.

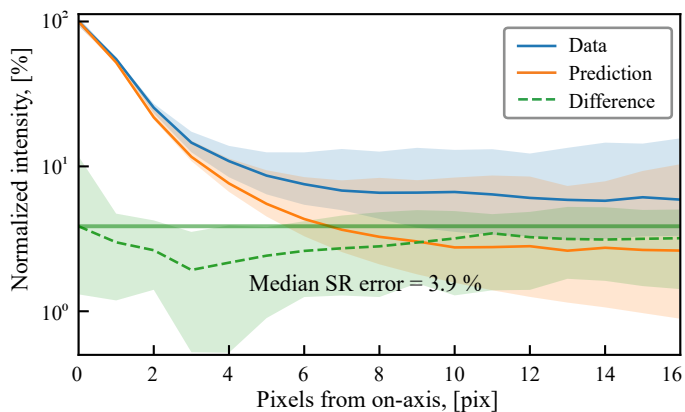


Fig. 9. Normalized radial PSF profiles for the ω Centauri field. The profiles were computed in the same way as in Fig. 6. Only astrometric shifts $\delta x, \delta y$ were fit; no other PSF parameters were adjusted.

off-axis behavior, the observed off-axis PSF variability is fully reproduced by the analytical component of the model.

6. Discussion

Although the present study is optimized for MUSE-NFM, the approach remains general. As will be shown in forthcoming papers in the series, the analytical PSF model can simulate a wide range of correction AO regimes. In this case, instrument specificity is attributed to the data-driven calibrator, as the set of reduced-telemetry inputs is unique to each instrument. However, given sufficient data, the same training logic can be applied to any system. Forthcoming work will demonstrate the same approach applied to SPHERE (the first results were already reported in Kuznetsov et al. 2023).

Once trained, the calibrator is stable for operational use, but it can still be fine-tuned as new data becomes available. For MUSE-NFM, current training is biased toward bright on-axis stars due to the use of the standard spectro-photometric stars dataset. Thus, extending the training to faint and off-axis sources is a natural next step to explore in forthcoming papers. Therefore, regularly acquiring moderately crowded stellar fields, similar to Fig. 8, would usefully complement the current datasets to facilitate off-axis calibration. Beyond the needs of this work, such data would also aid in debugging and calibrating future PSF-modeling approaches, including ones based on classical PSF-R.

Despite the message at the beginning of the paper, Fourier modeling and PSF-R should be viewed as complementary rather

than competing approaches. Fourier models efficiently explore parameter space using readily accessible data, while PSF-R can reach speckle-level precision, which is essential for high-contrast applications. This underscores the importance of occasionally recording full AO telemetry, both to enable classical PSF-R and to validate the procedures used to derive integrated inputs for Fourier models.

7. Conclusions

We have presented a compact, physics-based, and data-calibrated framework for predicting long-exposure, AO-corrected PSFs without relying on full AO telemetry. The approach combines a Fourier-based analytical model with a lightweight neural-network calibrator trained end-to-end with the PSF model, enabling efficient use of limited on-sky training data.

When applied to MUSE-NFM, the approach successfully reproduces polychromatic, field-dependent PSFs in close agreement with on-sky observations. On a held-out validation set of spectrophotometric standard stars, the model achieves a median peak error, Δ SR, of 13.5% and a relative core FWHM error of 10.9%. In a crowded ω Centauri field, the model simultaneously predicts off-axis PSFs for 54 sources with Δ SR < 5% and a 4.6% core FWHM error, enabling clean source separation without per-star PSF fitting.

Most importantly, our results highlight the need to calibrate PSF reconstruction and prediction methods using real on-sky data to achieve the required accuracy for scientific applications. Consequently, PSF models should be conceived from the beginning with calibration capabilities in mind.

Despite the demonstration on MUSE, the framework remains instrument-agnostic: instrument specificity is concentrated in the selected set of reduced-telemetry inputs used to train the small, flexible calibrator, while the analytical core spans multiple AO regimes. This enables this method to be ported to other facilities.

Future work will (i) broaden the training set to fainter and off-axis sources and to moderately crowded fields to strengthen off-axis calibration, (ii) extend the method to other AO regimes and instruments such as SPHERE, and (iii) investigate more real scientific applications of the method. The implementation is provided in the open-source TipTorch package (see the software availability section).

Data availability

The TipTorch package is publicly available under the MIT license: <https://github.com/EjjeSynho/TipTorch>

The astro-TIPTOP package is distributed under the MIT License and can be obtained from <https://github.com/astro-tiptop/TIPTOP>

The spectrophotometric calibration standard star dataset is available through the ESO archive at https://archive.eso.org/wdb/wdb/cal/qc_products/query?mjd_obs=&instrume==MUSE&pro_catg==DATACUBE_STD&ins_setup==NFM-AO-N_Blue-IR_SCI1.0.

Acknowledgements. We want to thank Joël Vernet for his valuable insights into the MUSE instrument and for his overall support of this project. Special thanks go to Sebastian Kamann for providing the ω Centauri datasets. We also thank Martina Scialpi for testing the early version of the code. We thank Pierre Haguenaier for his help in understanding the MUSE data infrastructure. We would also like to thank Olivier Beltramo-Martin for his assistance in understanding the P3 library's source code and theoretical background. Finally, we extend our thanks to John Pritchard and Fuyan Bian for helping with the MUSE-NFM data reduction pipeline. This work benefited from

the support of the European Southern Observatory, French National Research Agency (ANR) with textitWOLF (ANR-18-CE31-0018), textitAPPLY (ANR-19-CE31-0011) and textitLabEx FOCUS (ANR-11-LABX-0013); the Programme Investissement Avenir textitF-CELT (ANR-21-ESRE-0008), the textitAction Spécifique Haute Résolution Angulaire (ASHRA) of CNRS/INSU co-funded by CNES, the textitECOS-CONYKIT France-Chile cooperation (textitC20E02), the textitORP-H2020 Framework Programme of the European Commission's (Grant number textit101004719), textitSTIC AmSud (21-STIC-09), the Région Sud and the French government under the textitFrance 2030 investment plan, as part of the textitInitiative d'Excellence d'Aix-Marseille Université A*MIDEX, program number AMX-22-RE-AB-151.

References

- Beltramo-Martin, O., Correia, C. M., Ragland, S., et al. 2019, *MNRAS*, **487**, 5450
- Beltramo-Martin, O., Marasco, A., Fusco, T., et al. 2020a, *MNRAS*, **494**, 775
- Beltramo-Martin, O., Ragland, S., Fétick, R., et al. 2020b, in *Proc. SPIE*, 11448, eds. L. Schreiber, D. Schmidt, & E. Vernet (SPIE), 114480A
- Beuzit, J. L., Vigan, A., Mouillet, D., et al. 2019, *A&A*, **631**, A155
- Bouché, N., Carfantan, H., Schroetter, I., Michel-Dansac, L., & Contini, T. 2015, *AJ*, **150**, 92
- Cibinel, A., Floc'h, E. L., Perret, V., et al. 2015, *ApJ*, **805**, 181
- Ciddor, P. E. 1996, *Appl. Opt.*, **35**, 1566
- Correia, C. M., & Teixeira, J. 2014, *JOSA A*, **31**, 2763
- Correia, C. M., Bond, C. Z., Sauvage, J.-F., et al. 2017, *JOSA A*, **34**, 1877
- Desdoigts, L., Pope, B., & Tuthill, P. 2024, in *Proc. SPIE*, 13095, eds. J. Kammerer, S. Sallum, & J. Sanchez-Bermudez (SPIE), 1309513
- Desdoigts, L., Pope, B. J. S., Dennis, J., & Tuthill, P. G. 2023, *JATIS*, **9**, 028007
- Epinat, B., Tascia, L., Amram, P., et al. 2012, *A&A*, **539**, A92
- Fétick, R. J., Jorda, L., Vernazza, P., et al. 2019a, *A&A*, **623**, A6
- Fétick, R. J. L., Fusco, T., Neichel, B., et al. 2019b, *A&A*, **628**, A99
- Fusco, T., Rousset, G., Sauvage, J. F., et al. 2006, *Opt. Express*, **14**, 7515
- Fusco, T., Bacon, R., Kamann, S., et al. 2020, *A&A*, **635**, A208
- Harder, S., & Chelli, A. 2000, *A&AS*, **142**, 119
- Jia, P., Wu, X., Li, Z., et al. 2021, *MNRAS*, **505**, 4717
- Jolissaint, L. 2010, *JEOS-RP*, **5**
- Jolissaint, L., & Kendrew, S. 2010, in *Proc. AO4ELT 1*, 05021
- Jolissaint, L., Véran, J.-P., & Conan, R. 2006, *JOSA A*, **23**, 382
- Kamann, S., Wisotzki, L., & Roth, M. M. 2013, *A&A*, **549**, A71
- Kuznetsov, A., Neichel, B., Oberti, S., & Fusco, T. 2023, in *Proc. AO4ELT 7*
- Kuznetsov, A., Oberti, S., Neichel, B., & Fusco, T. 2024, *A&A*, **687**, A221
- Lau, A., Fétick, R. J. L., Neichel, B., Beltramo-Martin, O., & Fusco, T. 2023, *A&A*, **673**, A72
- Leibundgut, B., Bacon, R., Bian, F., et al. 2019, *Messenger*, **176**, 16
- Liaudat, T., Starck, J.-L., Kilbinger, M., & Frugier, P.-A. 2023, *Inverse Probl.*, **39**, 035008
- Long, M., Soubo, Y., Weiping, N., Feng, X., & Jun, Y. 2019, *AJ*, **888**, 20
- Lundberg, S. M., & Lee, S.-I. 2017, in *Adv. Neural Inf. Process. Syst.*, **30**, 4768
- Martin, O. A., Correia, C. M., Gendron, E., et al. 2016, in *Proc. SPIE*, 9909, eds. E. Marchetti, L. M. Close, & J.-P. Véran (SPIE), 99091Q
- Massari, D., Marasco, A., Beltramo-Martin, O., et al. 2020, *A&A*, **634**, L5
- Moffat, A. F. J. 1969, *A&A*, **3**, 455
- Neichel, B. 2008, Theses, Université Paris-Diderot – Paris VII
- Neichel, B., Fusco, T., & Conan, J.-M. 2009, *JOSA A*, **26**, 219
- Neichel, B., Lu, J. R., Rigaut, F., et al. 2014, *MNRAS*, **445**, 500
- Neichel, B., Beltramo-Martin, O., Plantet, C., et al. 2021, in *Proc. SPIE*, 11448, eds. L. Schreiber, D. Schmidt, & E. Vernet (SPIE), 114482T
- Neichel, B., Agapito, G., Kuznetsov, A., et al. 2024, in *Proc. SPIE*, 13097, eds. K. J. Jackson, D. Schmidt, & E. Vernet (SPIE), 130972Y
- Oberti, S., Kolb, J., Louarn, M. L., et al. 2016, in *Proc. SPIE*, 9909, eds. E. Marchetti, L. M. Close, & J.-P. Véran (SPIE), 99091U
- Paszke, A., Gross, S., Massa, F., et al. 2019, in *Adv. Neural Inf. Process. Syst.*, **32**
- Pechetti, R., Kamann, S., Krajnović, D., et al. 2024, *MNRAS*, **528**, 4941
- Ragland, S., Jolissaint, L., Wizinowich, P., et al. 2016, in *Proc. SPIE*, 9909, eds. E. Marchetti, L. M. Close, & J.-P. Véran (SPIE), 99091P
- Rigaut, F. J., Veran, J.-P., & Lai, O. 1998, in *Proc. SPIE*, 3353, 1038
- Thatte, N. A., Melotte, D., Neichel, B., et al. 2022, in *Proc. SPIE*, 12184, eds. C. J. Evans, J. J. Bryant, & K. Motohara (SPIE), 1218420
- Véran, J.-P., Rigaut, F., Maïre, H., & Rouan, D. 1997, *JOSA A*, **14**, 3057
- Wagner, R., Saxenhuber, D., Ramlau, R., & Hubmer, S. 2022, *A&C*, **40**, 100590

Appendix A: Theory

Appendix A.1: Computing OTF_{turb}

The OTF term responsible for modeling the effect of imperfect AO correction can be computed from the structure function of the AO residuals, D_φ , which in turn can be calculated using the residuals covariance matrix, B , that results from the spatial PSD of the AO residuals, W :

$$\begin{aligned} \text{OTF}_{\text{turb}}(\mathbf{v}) &= \exp(-D_\varphi(\boldsymbol{\rho}) / 2), \\ D_\varphi(\boldsymbol{\rho}) &= 2(B(0) - B(\boldsymbol{\rho})), \\ B(\boldsymbol{\rho}) &= \mathcal{F}^{-1}\{W(\mathbf{k})\}. \end{aligned} \quad (\text{A.1})$$

Here, $\boldsymbol{\rho}$ is the spatial shift vector in the pupil plane.

Appendix A.2: Jitter kernel rotation matrix

Matrix \mathbf{T} is a 2-d rotation matrix defined as

$$\mathbf{T}(\theta_J) = \begin{pmatrix} \cos \theta_J & -\sin \theta_J \\ \sin \theta_J & \cos \theta_J \end{pmatrix}. \quad (\text{A.2})$$

Appendix A.3: PSD error terms

Appendix A.3.1: Fitting error

According to the Von Kármán model, the open-loop atmospheric spatial power spectrum is defined as

$$W_\varphi(\mathbf{k}) = 0.0229 r_0^{-\frac{5}{3}} \left(\|\mathbf{k}\|^2 + L_0^{-2} \right)^{-\frac{11}{6}}, \quad (\text{A.3})$$

where r_0 is Fried's parameter, L_0 is the atmospheric outer scale. The piston filter $\tilde{\Pi}$ in Eq. 6 results from the inability of differentiating WFSs like Shack-Hartmann WFS to retrieve the average phase over the pupil. It is described by the Bessel function of the first kind defined in the spatial frequency space.

Appendix A.3.2: Wavefront sensor noise error

The WFS noise variance matrix \mathbf{C}_b is defined as

$$\mathbf{C}_b = \sigma_\eta^2 \mathbf{I}_d. \quad (\text{A.4})$$

Here, \mathbf{I}_d is the identity matrix, and the total noise variance per WFS subaperture σ_η^2 is composed of three terms:

$$\sigma_\eta^2 = \left(\frac{\lambda_{\text{atm}}}{\lambda_{\text{GS}}} \right)^2 (\sigma_{\text{det}}^2 + \sigma_{\text{ph}}^2 + \Delta\sigma^2),$$

where σ_{det}^2 , σ_{ph}^2 are detector readout and photon noise, respectively, defined in Eqs. (47) and (48) in Correia et al. (2017); the term $\Delta\sigma^2$ accounts for unexplained noise variance, which is highlighted in Sect. 4.2.3; $\lambda_{\text{atm}} = 500$ nm is the "atmospheric wavelength," at which all wavelength-dependent quantities, such as r_0 , are defined; and $(\lambda_{\text{atm}} / \lambda_{\text{GS}})^2$ term normalizes σ_η^2 from the WFS wavelength λ_{GS} to λ_{atm} .

Appendix A.3.3: Spatio-temporal error

The projection vectors \mathbf{P}_θ^L and \mathbf{P}_α^L are defined as

$$\mathbf{P}_\theta^L = \left(\left[\exp(2i\pi \mathbf{k} \cdot (\mathbf{h}[l] \boldsymbol{\theta} - \mathbf{V}[l] \Delta t)) \right]_{l,1} \right)_{l=1}^{N_L} \in \mathbb{C}^{N_L \times 1},$$

$$\mathbf{P}_\alpha^L = \left(\left[\exp(2i\pi \mathbf{h}[l] (\alpha_i \cdot \mathbf{k})) \right]_{l,i} \right)_{l=1, i=1}^{N_L, N_{\text{GS}}} \in \mathbb{C}^{N_L \times N_{\text{GS}}},$$

with

$$\begin{aligned} \mathbf{w} &= \left(\left[\mathbf{w}[l] \right]_{l,1} \right)_{l=1}^{N_L} \in \mathbb{R}^{N_L \times 1}, \\ \mathbf{V} &= \left(\left[(V_x[l], V_y[l]) \right]_{l=1, i=1} \right)_{l=1, i=1}^{N_L, 2} \in \mathbb{R}^{N_L \times 2}, \\ \mathbf{h} &= \left(\left[\mathbf{h}[l] \sec(\theta_{\text{zen}}) \right]_{l,1} \right)_{l=1}^{N_L} \in \mathbb{R}^{N_L \times 1}. \end{aligned}$$

Here, N_L is the number of atmospheric layers, N_{GS} is the number of guide stars, " \cdot " is the dot product, Δt is the WFS integration time, \mathbf{w} is the vector of the atmospheric layers' weights, \mathbf{V} is the wind vector per atmospheric layer, \mathbf{h} is the vector containing the atmospheric layers' altitudes in the observed direction, and θ_{zen} is the zenith angle. The matrix \mathbf{C}_φ weights the solution of tomographic reconstruction by providing priors on the open-loop atmospheric power spectrum per layer and is defined as $\mathbf{C}_\varphi = \mathbf{w} \mathbf{I}_d W_\varphi$. Note that \mathbf{P}_θ^L also "absorbs" the spatio-temporal shift of the atmospheric layers. Square brackets with subscripts, $[\cdot]_{l,i}$, denote individual array elements. Outer scalable parentheses with sub/superscripts, (\cdot) , indicate the assembly of such entries into a vector or matrix over the stated index ranges. Element access uses small brackets, for example $\mathbf{w}[l]$, to distinguish indexing from the parentheses used for vector/matrix construction.

Appendix A.3.4: Aliasing

In Eq. A.3.4, the substitute term $Q(\mathbf{k})$ combines the reconstruction and measurement operators (partially):

$$Q(\mathbf{k}) = 2i\pi d \mathbf{R}(\mathbf{k} + \mathbf{s}/d) \text{sinc}(k_m d) \text{sinc}(k_n d).$$

Meanwhile, the term A_φ absorbs the spatio-temporal error accumulated over the WFS integration time and AO loop lag:

$$\begin{aligned} A_\varphi(\mathbf{k}) &= \tilde{H}_1 \sum_{l=1}^{N_L} \mathbf{w}[l] \text{sinc}(k_m V_x[l] \Delta t) \text{sinc}(k_n V_y[l] \Delta t) \\ &\quad \times \exp(2i\pi t_d (\mathbf{k} + \mathbf{s}/d) \cdot \mathbf{V}). \end{aligned}$$

Both terms are defined within an SCAO framework, as elaborated earlier in Sect. 3.2.4. Here, $k_m = k_x - m/d$, $k_n = k_y - m/d$ are the spatial frequency vector components that simulate "folding" of spurious spatial frequencies into the domain of the corrected spatial frequencies, $\tilde{H}_1(\mathbf{k})$ is the weighted average of the temporal transfer function over the atmospheric layers introduced in Eq. 14 from Correia et al. (2017), \mathbf{R} is the vector that models the wavefront reconstruction, which is defined in Eqs. (24a, b) from Correia & Teixeira (2014). If the correction is applied in the middle of the integration time Δt , the total AO loop lag is $t_d = \Delta t/2 + \zeta$, with the last addend being the time lag resulting from WFS readout time and the DM command computation.

Appendix A.3.5: Chromatism error

The chromatism error arises because shorter-wavelength light undergoes stronger refraction and thus accumulates a larger OPD as they propagate through the atmosphere. As a result,

the wavefront correction computed at the WFS wavelength differs from the one required to compensate for the wavefront disturbances at the science wavelength, resulting in imperfect wavefront compensation.

Appendix A.4: Differential refraction error

The differential refraction error arises when the scientific target and GS are observed at some angular separation from the zenith. The atmosphere then acts as a prism, making wavefronts at different λ travel slightly different columns of atmospheric turbulence and thus accumulate slightly different wavefront errors.

Then, θ_λ is the apparent angular displacement between two images of the same target observed at λ and λ_{GS} , defined as in Eq. (10) of Fusco et al. (2006). In this case, the term $\Theta_{\text{DR},l}$ from Eq. (11) can be denoted as

$$\Theta_{\text{DR},l}(\lambda) := 2\pi h[l] \tan(\theta_\lambda) \mathbf{z}. \quad (\text{A.5})$$

Here, $\mathbf{z} = (1, 0)^\top$ is the zenith-aligned vector.

Appendix B: Glossary with the variable definitions

Table B.1. Glossary of the key variables used in this work.

Parameter	Description
$r_0(\lambda)$	Fried parameter λ .
L_0	Atmospheric turbulence outer scale.
τ_0	Atmospheric coherence time.
C_n^2	Refractive-index structure constant profile.
N_L	Number of discrete turbulent layers.
\mathbf{h}	Turbulent layers altitudes.
\mathbf{w}	Relative turbulence layer weights.
\mathbf{V}	Turbulence layer wind velocities.
$P(\mathbf{x})$	Telescope pupil mask.
$\delta x(\lambda), \delta y(\lambda)$	Sub-pixel astrometric shifts in the focal plane.
$J_x(\lambda), J_y(\lambda)$	FWHM of residual TT jitter.
θ_j	Rotation angle of the TT jitter reference frame.
$F(\lambda)$	Flux normalization factor.
$b(\lambda)$	Constant background flux.
θ_{zen}	Zenith angle of the observation.
θ	Angular position of the science target.
α_{pix}	On-sky pixel scale.
A, B	Amplitude and background terms of ΔW .
α_M, β_M	Width and slope parameters of ΔW .
a_b	Scaling coefficient of the static phase bump.
$\Delta\sigma^2$	Additional noise variance term.
d_{act}	DM actuator pitch.
d	WFS subaperture size.
λ_{atm}	Atmospheric phase normalization wavelength.
λ_{GS}	Guide-star WFS wavelength.
N_{GS}	Number of guide stars.
α	Angular positions of the guide stars.
σ_{det}^2	WFS detector noise variance..
D	Telescope primary mirror diameter.
Δt	WFS integration time.
ζ	Effective AO loop delay.

Appendix C: Calibrator training

As mentioned earlier, the training of the calibrator g can be formalized as the minimization of the loss function, \mathcal{L} :

$$\min_{\gamma} \frac{1}{N_{I,t}} \sum_{k=1}^{N_{I,t}} \mathcal{L} \left(I_{\text{data}}^{(k)}(\mathbf{p}, \lambda), \tilde{I}^{(k)}(\mathbf{p}, \lambda) \right),$$

$$\tilde{I}^{(k)}(\mathbf{p}, \lambda) := f(\mathbf{p}, \lambda; \mathcal{Y}_s \cup \mathcal{Y}_d^{(k)} \oplus g(\mathcal{X}^{(k)}; \gamma)).$$

Here, $N_{I,t} = 314$ is the number of training samples drawn from on-sky data. Then, for each sample k , the loss function is defined as

$$\mathcal{L}(\dots) = \frac{1}{HWN_\lambda} \sum_{\lambda} \sum_{\mathbf{p} \in \Omega} \omega_1 |\Delta^{(k)}(\mathbf{p}, \lambda)| + \omega_2 \left(\Delta^{(k)}(\mathbf{p}, \lambda) \right)^2,$$

$$\Delta^{(k)}(\mathbf{p}, \lambda) := w(\lambda) \left(\tilde{I}^{(k)}(\mathbf{p}, \lambda) - I_{\text{data}}^{(k)}(\mathbf{p}, \lambda) \right).$$

The loss combines mean-absolute (L_1) and mean-square (L_2) terms, weighted by coefficients ω_1, ω_2 tuned via hyperparameter optimization. The L_1 term emphasizes the PSF halo, while L_2 focuses on the core energy distribution. The chromatic weighting factor $w(\lambda)$ balances contributions across spectral slices: for MUSE-NFM, "blue" PSFs are more strongly blurred by poorer AO correction, while NIR PSFs are sharper and more core-concentrated.

Appendix D: PSF prediction

Appendix D.1: FVU definition

The FVU is a measure of how well a model predicts observed data. It is a normalized error metric, defined as

$$\text{FVU} = \frac{\text{var}(\tilde{\mathbf{I}} - \mathbf{I}_{\text{data}})}{\text{var}(\mathbf{I}_{\text{data}})} \times 100\%. \quad (\text{D.1})$$

A model that perfectly explains the observed data would give $\text{FVU} \rightarrow 0\%$, while $\text{FVU} > 100\%$ means that the model is worse than using the mean of the data as a predictor.

Appendix D.2: Calibrated prediction versus direct prediction

Section 4.5 describes the calibration scheme implemented via the preceding data-driven component. To quantify its benefit, we consider two simplified PSF-prediction baselines.

- Direct prediction. In this configuration, all model-related inputs available in \mathcal{X} are provided directly to the model, while $F(\lambda) = 1$ at all wavelengths and $\Delta\sigma^2 = A = B = \alpha_M = a_b = 0$. The TT jitter is assumed to be achromatic and is computed by applying Eqs. (47)–(48) of Correia et al. (2017) to the 2×2 subaperture WFS. In other words, here the PSF model is used purely analytically without any calibration or error "absorption."
- Tuned-direct prediction. Starting from the direct configuration, we apply a minimal, data-driven correction by optimizing global median values for the variables in \mathcal{Y}_c (excluding r_0 and \mathbf{w}) to minimize the discrepancy between the PSF model and the training subset of \mathcal{D} . These optimized quantities act as static offsets in model inputs, rather than functional dependencies on wavelength or field position.

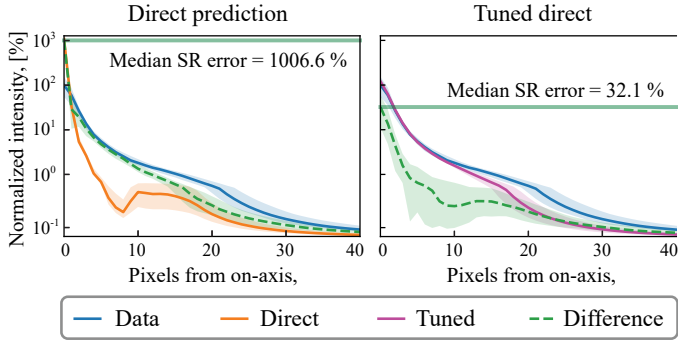


Fig. D.1. Radial profiles for the direct (left) and tuned-direct (right) baselines, computed for white-light PSFs.

Table D.1. Median validation metrics for different prediction approaches.

Metric	Direct	Tuned-direct	Calibrated
ΔSR , [%]	1006.6	32.1	13.5
$\Delta FWHM_{rel}$, [%]	80.0	20.2	10.9
FVU [%]	8610	17.4	3.2

In both cases, the photometric bias (b) and astrometric shifts (δx , δy) are obtained by image-plane fitting for each sample. Performance on the validation set for the direct and tuned-direct approaches is summarized in Fig. D.1 and Table D.1.

Relative to the calibrated prediction reported in Sect. 5.1, both simplified baselines yield significantly poorer prediction accuracy, underscoring the necessity of the data-driven calibration.

Two main factors account for the large errors obtained with direct prediction. First, MUSE PSFs are under-sampled and exhibit a steep exponential profile. Consequently, even small inaccuracies in the PSF model parameters translate into significant errors in the central pixel intensity, which in turn strongly affect the SR estimate. These peak-intensity errors also highlight the need for accurate sub-pixel astrometric fitting. Second, an adequate description of TT jitter is crucial, as it largely determines the FWHM of the PSF. In the direct approach, the TT wavefront error is derived from LO WFS noise, which underestimates its power. As a result, the predicted PSFs appear significantly sharper than observed in practice, reducing the model’s ability to reproduce the overall PSF morphology and thereby increasing the FVU error.