Empirical contrast model for high-contrast imaging

A VLT/SPHERE case study


1 European Southern Observatory, Alonso de Córdova 3107, Vitacura, Casilla 19001, Santiago, Chile
e-mail: bcourtne@eso.org
2 Research School of Astronomy and Astrophysics, Australian National University, Canberra, ACT 2611, Australia
3 Institute of Astronomy and National Astronomical Observatory, Bulgarian Academy of Sciences, 72 Tsarigradsko Shose Blvd., Sofia 1784, Bulgaria
4 Université Grenoble Alpes, CNRS, IPAG, 38000, Grenoble, France

Received May 24, 2023; accepted September 26, 2023

ABSTRACT

Context. The ability to accurately predict the contrast achieved with high-contrast imagers is important for efficient scheduling and quality control measures in modern observatories. Aims. We aim to consistently predict and measure the raw contrast achieved by SPHERE/IRDIS on a frame-by-frame basis in order to improve the efficiency of SPHERE at the Very Large Telescope (VLT) and maximise scientific yield. Methods. Contrast curves were calculated for over 5 years of archival data obtained using the most common SPHERE/IRDIS coronographic mode in the H2/H3 dual-band filter. These data consist of approximately 80,000 individual frames, which were merged and interpolated with atmospheric data to create a large database of contrast curves with associated features. An empirical power-law model for contrast —motivated by physical considerations— was then trained and finally tested on an out-of-sample test dataset. Results. At an angular separation of 300 mas, the contrast model achieved a mean (out-of-sample) test error of 0.13 magnitude with the 5th and 95th percentiles of the residuals equal to -0.23 and 0.64 magnitude respectively. The models test-set root mean square error (RMSE) between 250 and 600 mas was between 0.31 and 0.40 magnitude, which is equivalent to that of other state-of-the-art contrast models presented in the literature. In general, the model performed best for targets of between 5 and 9 G-band magnitude, with degraded performance for targets outside this range. This model is currently being incorporated into the Paranal SCUBA software for first-level quality control and real-time scheduling support. Key words. VLT/SPHERE, High, Contrast, Prediction, Model

1. Introduction

High-contrast imagers have become central tools for the discovery and understanding of exoplanets and protoplanetary disk formation, demographics, and dynamics around young stars. Instruments such as the Spectro-Polarimetric High-contrast Exoplanet REsearcher (SPHERE) at the Very Large Telescope (VLT) (Beuzit et al. 2019; Fusco et al. 2015), The Gemini Planet Imager (GPI) at the Gemini observatory (Macintosh et al. 2014), and the Subaru Coronagraphic Extreme Adaptive Optics (SCExAO) at the Subaru telescope (Sahoo et al. 2018) can achieve typical raw contrasts in the range of $10^{-4}$ to $10^{-6}$ between 0.1” and 0.5” from the central star at near-infrared (NIR) wavelengths. The instruments foreseen for the epoch of extremely large telescopes will push these boundaries even further (e.g., Brandl et al. 2021). The outstanding capabilities of these current and future instruments attached to world-class telescopes are in great demand, making the optimisation of telescope time an important task. This optimization generally requires accurate models to predict observational performance indicators (e.g., contrast for high-contrast imagers) both prior and/or early on during the observation in order to select observations that optimally exploit the atmospheric conditions. Traditionally, short-term scheduling and quality control in queue observations for high-contrast imagers such as SPHERE are carried out primarily based on atmospheric, sidereal, and/or airmass constraints. While strong correlations exist between turbulence and adaptive optics (AO) performance, there are often outliers where the measured contrast is considerably lower than what would be expected from the observed atmospheric conditions. This is typically due to local effects within the telescope or instrument. Without the ability to properly predict and measure the contrast, such observations may be scheduled and pass basic quality-control checks despite not meeting the users scientific requirements. Therefore, to optimise telescope time and the ultimate data quality provided to users, high-contrast imagers may benefit greatly from precise models to predict scientifically meaningful metrics, such as contrast or signal-to-noise ratio (S/N), which can then be measured in almost real time in order to evaluate the quality of the data. The ultimate goal is to perform short-term scheduling and quality control based on predicted and measured metrics that are of scientific significance. This, combined with significant efforts to improve quality-control software and standards (Thomas et al. 2020), as well as forecasting models at Paranal (Milli et al. 2020, 2019; Masciadri et al. 2020; Osborn & Sarazin 2018), will greatly advance real-time decision making and quality control measures. The ability to accurately predict contrast for high-contrast imagers on large telescopes is not a trivial task. While fundamental atmospheric limits are well characterized (Conan et al. 1995; Fusco & Conan 2004; Aime &
Soummer 2004; Males et al. 2021), typically nontrivial local effects can dominate achievable contrast; for example, quasi-static speckles caused by opto-mechanical imperfections and thermal drifts (Bloemhof et al. 2001; Soummer et al. 2007; Martínez et al. 2013; Vigan et al. 2022) or dome seeing (Tallis et al. 2020) and low wind effects (Milli et al. 2018; Sauvage et al. 2015).

Given the maturity of instruments such as SPHERE, data-driven analysis and empirical models are a practical way to understand these limitations and the uncertainty that random telescope and instrumental processes have on observations. Some good examples of this are the work of Martínez et al. (2013), using SPHERE data to characterize speckle temporal stability in high-Strehl regimes; that of Milli et al. (2018), who characterized the low wind effect on SPHERE; and Jones et al. (2022), who presented a data-driven analysis of SPHERE performance for faint targets. In the case of predicting contrast, various models have been explored in the literature to empirically predict the on-sky observed contrast given atmospheric and instrument conditions (Bailey et al. 2016; Courtney-Barrer et al. 2019; Xuan et al. 2018). In particular, initial work at the GPI used linear regression of AO telemetry and astronomical site monitoring data to predict contrast (Bailey et al. 2016), which was further advanced with neural networks that were able to predict the measured contrast using six input parameters available pre-observation with a contrast root mean square error (RMSE) of 0.45 magnitude at 0.25" (Savrasky et al. 2018). Correlations between measured contrast and AO error terms have also been shown in other work (e.g., Poyneer & Macintosh 2006; Poyneer et al. 2016), and in general have been shown to provide good predictive capacity of the contrast in atmospherically limited regimes (Fusco et al. 2016; Sauvage et al. 2016; Macintosh et al. 2014). In this work, we present a simple empirical model to predict the raw contrast measured by SPHERE with the goal to assist on-site quality control measures and short- to mid-term scheduling decisions. This paper begins with a brief overview of the SPHERE instrument followed by the motivation of an empirical model to which contrast data is fitted. Section 4 outlines the data preparation and pre-processing that was done before fitting the contrast model, along with the algorithms used for fitting. Section 5 presents the results along with some discussion. In Section 6, we present our conclusions and future outlook.

2. SPHERE/IRDIS

SPHERE is an extreme AO instrument installed on the Unit Telescope 3 (Melipal) at the Paranal Observatory. Its primary science goal is imaging and low-resolution spectroscopic and polarimetric characterization of extra-solar planetary systems at optical and NIR wavelengths. SPHERE consists of three science channels: the Integral Field Spectrograph (IFS) and the Infra-Red Dual-band Imager and Spectrograph (IRDIS), which both observe in the NIR, and the Zurich Imaging Polarimeter (ZIMPOL) for visible polarimetric observations. Each subinstrument has a series of coronagraphs and filters available in addition to having an extreme AO system called SAXO (Fusco et al. 2016; Sauvage et al. 2016) placed in the common path of all subinstrument channels. SAXO operates up to a frequency of 1.38 kHz on bright targets with a 40x40 spatially filtered Shack-Hartmann (SH) wavefront sensor (WFS) measuring in the optical, and a 41x41 piezoelectric high-order deformable mirror for AO actuation. SAXO also uses a dedicated differential tip/tilt sensor (Baudouz et al. 2010) in the NIR to correct for wavelength-dependent tip/tilt between the NIR and optical channels. In the present work, we tested our model on the most common SPHERE/IRDIS mode, which uses an apodised Lyot coronagraph with the H2/H3 dual-band filters centered at wavelengths 1.593μm and 1.667μm, respectively.

3. Contrast model

We begin by outlining our motivation to build an empirical model for contrast with some statistical considerations of the measured intensity in the focal plane. It can easily be shown using Fourier optics that a phase aberration at some spatial frequency k in a pupil plane of a telescope gets mapped to a so-called speckle in the focal plane at an angular coordinate of kλ, where λ is the wavelength of the light (Roddier 2004). Such speckles are typically classified based on their temporal behavior, which ultimately determines whether or not (or how well) they can be suppressed by post-processing reduction methods. This sets the fundamental contrast limits on ground-based high-contrast imagers (Males et al. 2021). The detection of real signals (such as a planet) within the circumstellar environment of a star requires statistical knowledge of the probability of some intensity measurement in the focal plane. Various authors (e.g., Canales & Cagial (1999) and references therein) have shown —under the assumption of long exposures— that a point-wise intensity measurement (I) generally follows a modified Rician probability density function:

\[ P(I) = \frac{I}{2\sigma^2} \exp\left(\frac{-I + s^2}{\sigma^2}\right) I_0\left(\frac{2s\sqrt{I}}{\sigma^2}\right), \tag{1} \]

where \(I_0\) is the zero-order modified Bessel function of the first kind, and \(s^2\) and \(2\sigma^2\) are related to the (long-exposure) intensity of the deterministic and random speckle components of the wavefront, respectively. Soummer et al. (2007) developed this statistical framework to derive a general expression for the expected point-wise variance in a coronagraphic image as:

\[ \sigma^2 = N(I^2_{11} + NI^2_{12} + 2I_{11} + 2NI_{12} + 2I_{11} + 2I_{12}) + \sigma_p^2, \tag{2} \]

where we have kept the notation used in Soummer et al. (2007). Here, “I” generally denotes the intensity, \(\sigma_p^2\) the variance of the photon noise, and N the ratio of fast-speckle to slow-speckle lifetimes. \(I_s\) is the intensity produced by the deterministic part of the wavefront, including static aberrations, while the \(I_r\) terms correspond to the halo produced by random intensity variations; that is, atmospheric (\(I_{11}\)) and quasi-static contributions (\(I_{12}\)). In this generalized expression of the variance, several contributions can be identified by order of appearance: (1) the atmospheric halo; (2) the quasi-static halo; (3) the atmospheric pinning term, the speckle pinning of the static aberrations by the fast-evolving atmospheric speckles; (4) the speckle pinning of the static by quasi-static speckles; and finally (5) the speckle pinning of the atmospheric speckles by quasi-static speckles. Converting this to the expected contrast as a general function of radius (e.g., a typical contrast curve) requires calculation of the sum of the pixel-wise modified Rician density functions within a given annulus. No closed-form analytic solution to this exists, although there are closed-form approximations (Lopez-Salcedo 2009). Under a strong assumption of independence between pixels and, for a given spatial frequency, equal probability of the directions of aberrations (i.e. angular position of a speckle at a fixed radius), where \(\theta\) is the angular position of a speckle at a given radius, we can estimate the expected intensity variance within a thin annulus at radius r by simply scaling by the number of pixels in the
annulus, in which case we can make the proportional approximation of the 1r contrast:

\[
(C(r)) \approx \frac{(\langle r_I(r) \rangle)}{I_o},
\]

(3)

where \(I_o\) is the stellar intensity in the science channel and \(\langle . . \rangle\) is the expected value. As mentioned, this is a strong assumption that does not generally hold. For example, experience shows that there is typically anisotropy in the aberrations at a given spatial frequency, especially from biased wind directions of dominant turbulent layers. Nevertheless, this basic assumption is useful for deriving first-order contrast estimates. Analytically predicting each term in Eq. 2 prior to observation would require full knowledge of internal aberrations and wind velocity profiles, as well as the ability to reconstruct the modal distribution of the incoming phase front in order to predict AO residuals, which would difficult to achieve. Nevertheless, noting that any first-order expansion of the coronagraph PSF term (\(I_o\)) and atmospheric speckle terms (\(I_{\phi1}\)) with regard to the typical AO error budget terms would lead to various cross products of the AO error-budget terms in the pined speckles; we could make the assumption that typically one of these terms will dominate the halo at a given radius and therefore propose to model the contrast as a product of AO cross terms, each with a power law to give an appropriate weighting at a given radius, that is,

\[
C(r) \approx x(r) \prod_{i=1}^{4} \Delta_i^{\alpha_i(r)},
\]

(4)

where \(x(r)\) and \(\alpha_i(r)\) are the fitted parameters for a given radius. From a basic leave-one-out analysis, the \(\Delta\) terms considered for the following model are a combination of typical (unitless) AO error-budget-like terms:

\[
\begin{align*}
\Delta_1 & \equiv \Delta_{f_{AO}} = \frac{D}{r_0}, \\
\Delta_2 & \equiv \Delta_{\text{servo}} = \frac{\tau}{\tau_0}, \\
\Delta_3 & \equiv \Delta_{\text{SNR-WFS}} = \frac{n_{\text{pix},f_{WS}}}{N_D \left( \frac{n_p^2 + \left( \frac{\lambda}{\lambda_0} \right) R_p^2}{} \right)}, \\
\Delta_4 & \equiv \Delta_{\text{SNR-SCI}} = \frac{1}{\sqrt{n_{p,\text{sci}}}},
\end{align*}
\]

(5) \(\Delta_2 \equiv \Delta_{\text{servo}} = \frac{\tau}{\tau_0},\)

(6) \(\Delta_3 \equiv \Delta_{\text{SNR-WFS}} = \frac{n_{\text{pix},f_{WS}}}{N_D \left( \frac{n_p^2 + \left( \frac{\lambda}{\lambda_0} \right) R_p^2}{} \right)},\)

(7) \(\Delta_4 \equiv \Delta_{\text{SNR-SCI}} = \frac{1}{\sqrt{n_{p,\text{sci}}}},\)

(8)

where \(D\) is the telescope diameter, \(r_0\) is the atmospheric coherence length (Fried parameter), \(\tau\) and \(\tau_0\) are the AO latency and atmospheric coherence time, respectively, \(n_p\) is the number of detected photoelectrons per defined subaperture (sum of all pixels); \(N_D\) is the number of pixels in a subaperture, \(n_{p,\text{sci}}\) is the number of detected background photoelectrons per subaperture; \(e_p\) is the read noise in electrons per pixel, and \(G\) is the gain (Robert K. Tyson 2012). The values \(n_p\) are generally inferred through fitted zero points and extinction coefficients to convert stellar magnitude (mag) to flux (see section 4.1). The residuals of such a model would therefore be due to the variance in non-AO related terms in equation 2. We also note that in a shot-noise-limited regime, the product of

\[
\Delta_{\text{red}} \equiv \Delta_{\text{red}}^{\alpha_3} = n_{\text{pix},f_{WS}}^{\alpha_4} / n_{p,\text{sci}}^{\alpha_4}
\]

(9)

can be seen as the reddening parameter. In the case of equal flux, that is, \(n_{\text{pix},f_{WS}} = n_{p,\text{sci}} = n_p\), we get \(\Delta_{\text{red}} = n_p^{(\alpha_3 - \alpha_4)/2}\) and therefore one would expect \(\alpha_4 > \alpha_3\) to maintain the situation where brighter targets generally achieve better contrast. However, in general \(n_{\text{pix},f_{WS}} \neq n_{p,\text{sci}}\) and chromatic effects of red stars may play an important role (especially with the performance of the differential tip/tilt controller). This general model confers a considerable advantage in that it can capture nonlinearities in the contrast performance and furthermore can be fitted linearly by simply considering the contrast magnitude: \(C_m = -5/2 \log_{10}(C)\), such that

\[
C_m(r) = X(r) - \sum_{i=1}^{4} \alpha_i(r) \log_{10}(\Delta_i),
\]

(10)

where \(X(r) = X(r) - C.m(r)\) is the fitted intercept. On the training dataset, we also allowed for a linear calibration of the fitted intercept \(X_f(r)\) with the model residuals given the partitioned instrumental state such that \(X(r) = X_f(r) - \Delta_r(r)\) state, where \(\Delta_r(r)\) state is the training model residual when filtering for a given observational state. The state filters considered were the sky transparency classification (e.g., thick, thin, clear, photometric), the AO gain and frequency setting, and the wavefront sensor spatial filter size (e.g., small, medium, large). This significantly improved the cross-validation performance of the model on the training dataset, while maintaining a significant sample size for the general fitting of \(\alpha\) parameters. These calibrated offsets were then used for the out-of-sample model test (without recalibration).

4. Data preparation

We developed a database of all public observations taken between 2015 and 2019, which were downloaded from the ESO SPHERE archive. In this particular study, we focused on fitting the model described above for the most commonly observed SPHERE/IRDIS mode. This mode uses an apodised Lyot coronagraph with the H2 (\(\lambda_c = 1.593\mu m\)) filter, which corresponds to the right detector in the H2/H3 dual band mode. The general FITS headers used to filter this data are displayed in Table 1. After filtering and outlier rejection (described below), train (75%) and test (25%) datasets were split to have nonoverlapping observation nights; meaning for any given sample in the train set, there did not exist a sample in the test set that was observed on the same night (and vice versa). This corresponds to 149 and 47 unique stars in the train and test sets, respectively, with only 4 stars shared between the two sets, totalling nearly 80k raw coronagraphic frames to analyze. A 75%/25% split provided a sufficient parameter space density to perform ten-fold cross validation on the training set, while allowing sufficient samples to avoid biases in the out-of-sample test. One-sigma noise levels were estimated as a function of radius in coronagraphic data cubes (DPR TYPE = OBJECT) after some basic reduction (e.g., background subtraction, flat fielding, bad pixel masking, and high-pass filtering). The standard deviation was calculated in an annulus with 4 pixels (\(\sim \lambda/D\)) from 82-1800mas, where radii that had pixels in the nonlinear regime of the IRDIS detector (\(ADU > 20k\)) were masked. Additionally, each coronagraphic (OBJECT) frame was cross correlated with a median coronagraphic image across the filtered dataset to provide an additional criterion for outlier removal. From visual inspection of individual frames, anything with a cross correlation below 0.5 seemed to correspond to frames that had obvious issues, such as bright companions in the field or where AO loops temporarily opened during an exposure. Therefore, frames that had a cross correlation with the median image of less than 0.5 were dropped from the analysis. In short
exposures, there were noticeable pinned speckles that were initially difficult to predict from atmospheric conditions. This was significantly improved by co-adding coronagraph frames that had exposure times of ≤ 64s to roughly 64s. An example of this is shown in Figure 1. A 2D Gaussian function was then fitted to the corresponding noncoronagraphic flux frames (DPR TYPE = OBJECT,FLUX) of the observation to estimate the peak flux. The contrast curve was then calculated by dividing the 1σ noise levels at a given radius by the estimated peak flux when adjusting for the different integration times and neutral density filters. To correct for changing transparency between the noncoronagraphic (flux) and coronagraphic (object) frames, the measured contrast was multiplied by the ratio of the aggregated wavefront sensor (SPARTA telemetry) flux data between the two periods. While no specific sky classifications (e.g., thin or thick clouds) were excluded from the data, we neglected data where there was significant (> 50%) variability in the wavefront sensor flux during an exposure, because this caused significant variability in the measured contrast that was unpredictable from a pre-observation perspective.

The initial contrast curve database consisted of 27135 co-added contrast frames. Each contrast curve was associated with various features, including the FITS headers from the initial files and the full available range of atmospheric parameters available from ESO MASS-DIMM and meteorological archives, which were interpolated to the mean coronagraphic frame timestamp. This included important parameters such as atmospheric seeing and coherence time. Atmospheric data prior to the last MASS-DIMM upgrade (April 2016) were neglected due to instrument biases between the old and new systems. Data were then further filtered to exclude observations that were outside of standard operational conditions and/or where feature outliers were detected. Basic filters included the following:

- All SPARTA AO loops were closed.
- SPARTA differential tip/tilt loop is closed.
- Telescope was guiding on a guidestar.
- Atmospheric seeing and coherence time measurements were in the ranges of 0.5" and 0-30ms, respectively.

<table>
<thead>
<tr>
<th>Keyword</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DPR TYPE</td>
<td>OBJECT / OBJECT,FLUX</td>
</tr>
<tr>
<td>INS COMB I COR</td>
<td>N_ALC_YJH_S</td>
</tr>
<tr>
<td>DPR TECH</td>
<td>IMAGE</td>
</tr>
<tr>
<td>INS1 FILT NAME</td>
<td>B_H</td>
</tr>
<tr>
<td>INS COMB IFLT</td>
<td>DB_H23</td>
</tr>
<tr>
<td>INS4 FILT3 NAME</td>
<td>OPEN</td>
</tr>
</tbody>
</table>

Data taken prior to the M2 spider re-coating carried out in November 2017 show significantly different contrast statistics for wind speeds below 3m/s due to the low wind effect, which was largely resolved by the re-coating intervention (Milli et al. 2018; Sauvage et al. 2015). Figure 2 shows the measured raw contrast at 0.3" before and after the M2 spider re-coat for low (left plot) and nominal (right plot) wind speeds. It is clear that there was a large statistical improvement in the measured contrast from the intervention in the low wind case, while there was a statistically insignificant difference when observing in wind speeds of >3m/s after the re-coating. We note that data taken with fast AO modes (1200Hz, 1380Hz) were neglected in these histograms to avoid biases, because the 1380Hz AO mode was an upgrade that was not used in earlier data, particularly before the M2 spider re-coat. This prompted us to only neglect data taken before November 2017 with wind speeds of <3m/s. After this filtering process, 8494 co-added frames remained with which to train and test the model.

4.1. Flux model
Calibrated instrumental zero points and extinction coefficients were required to estimate the photocurrent (ADU/s) received in both the wavefront sensor and science detector for a given SIMBAD magnitude and airmass prior to observation. Data were first filtered to consider SPHERE flux sequences and WFS data taken during periods classified as photometric using either the LP780 or OPEN spectral filter without restrictions on the WFS spatial...
filter size. From these filtered data, the following model was fitted:

\[
M = \beta_0 \left( 2.5 \log_{10} \left( \frac{F}{TG} \right) \right) + \beta_1 X + \beta_2 ,
\]

where \(M\) is the Simbad magnitude of the target star, \(F\) is the flux (ADU/s) measured in either the WFS (G band) —which is summed over all subapertures— or the flux template from the science detector (H band), \(T\) is a scalar to represent the relative transmission of the spectral filter, \(G\) the detector gain, \(X\) is the targets airmass, and \(\beta_0, \beta_1, \) and \(\beta_2\) are the fitted parameters corresponding to the telescope–instrument transfer function, extinction coefficient, and zero point, respectively. Fitted parameters from data taken in photometric conditions are outlined in Table 2 and are consistent with previously measured extinction coefficients at Paranal (Patat et al. 2011).

To account for sky transparency in the contrast model, sky category offsets were calibrated on the train dataset via the contrast residuals of the respective sky-category-partitioned data as described in section 3. Additionally, figure 3 displays the results when the data were partitioned into sky–transparency categories as classified by the weather officer and the above fitted photometric model was applied to each respective sky category. Figure 4 shows that the mean absolute error between the measured and predicted WFS flux given the target magnitude categories is the targets airmass, and \(\beta_0, \beta_1, \) and \(\beta_2\) are the fitted parameters corresponding to the telescope–instrument transfer function, extinction coefficient, and zero point, respectively. Fitted parameters from data taken in photometric conditions are outlined in Table 2 and are consistent with previously measured extinction coefficients at Paranal (Patat et al. 2011).

5. Results and discussion

Figure 5 shows the residuals of the contrast model as a function of separation from the central star for the training and testing datasets, which shows good generalization to the out-of-sample test. It can also be seen that mean model residuals were well centered near zero, with a mean error of 0.13 mag, and the 5th and 95th percentiles in the residuals were -0.23 and 0.64 magnitude respectively on the test set at 300mas. Figure 6 shows the general measured versus predicted raw \(\sigma_r\) contrast on the out-of-sample test set across all radii, along with the respective model RMSE at a given radial separation from the central star. Between 0.25" and 1.00", the model achieves a RMSE of between 0.17 and 0.40 magnitude on the out-of-sample test set.

We also analyze the model performance over an aggregated grid of atmospheric and star brightness categories in figure 7. The atmospheric categories considered are those currently used (ESO period p106) as user constraints for grading observations. The categories are defined from cuts in the atmospheric seeing–coherence time joint probability distribution, with T.Cat10 corresponding to the best (top 10%) atmospheric conditions, and T.Cat85 to the worst. We also simply consider three star magnitude categories of bright (\(Gmag < 5\)), mid (\(5 < Gmag < 9\)), and faint (\(Gmag > 9\)) targets. There is excellent agreement (<0.15mag residual at 0.5") in the mid category range across all atmospheric conditions; however, there is poorer performance for the bright and faint categories, particularly when in better atmospheric conditions. This would indicate that the underlying assumption that pinned atmospheric residuals dominate contrast is most valid in the mid category, while the worst performance is seen for bright and faint target where, for example, static or quasi-static pinning may become dominant. The fitted...
6. Conclusions

A simple model was trained and tested on SPHERE/IRDIS coronagraphic data to predict contrast as a function of radius in the most commonly used H-band filter. When testing on out-of-sample test data, the model achieved a mean error of 0.13 magnitude with the 5th and 95th percentiles in the residuals of -0.23 and 0.64 magnitude respectively when considering the train set median Gmag=7. Contrast degrades by roughly 20% per magnitude difference between WFS and Science channels.

To compare results to other contrast models found in the literature, we achieved a test contrast RMSE of between 0.31 and 0.40 magnitude between 250 and 600mas, or equivalently a log10 contrast RMSE of 0.13-0.16, respectively. This is comparable to the results of Savransky et al. (2018), who, when using a feedforward neural network with pre-observation data as input, achieved a test log10 contrast RMSE of 0.18 at 250mas. Such comparable results are encouraging given the relative simplicity and physical interpretability of the model presented in this work compared to more complex neural networks. Comparing the predictions of our model at 400mas to the current SPHERE exposure time calculator (ETC) offered by ESO (as of May 2023) when considering raw image predictions (EXPTIME=64s) without differential imaging (neglecting field rotation), we see in figure 9 that the ETC appears to provide very optimistic predictions for bright targets, and pessimistic predictions for faint targets relative to those of the model presented here across all turbulence categories. Our model also predicts that the contrast is less sensitive to changes in H-band flux compared to the ETC, which seems to also be reflected in the data at hand. The inclusion of these data in the ETC in short exposure time limits could be used to ultimately improve the predictive accuracy of an ETC for SPHERE users. The contrast model presented here is currently being incorporated into Paranal’s SCUBA software (Thomas et al. 2020) to be used as first-level quality control and to help improve the real-time decision process for SPHERE at Paranal. Figure 10 shows a real example of how this model could be used in operations for providing quick checks to ensure the measured raw contrast is within the expected 95% test residual range of 

\[ \text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \]

where \( y_i \) is the observed contrast, \( \hat{y}_i \) is the predicted contrast, and \( N \) is the number of observations. This is encouraging given the relative simplicity and physical interpretability of the model presented in this work compared to more complex neural networks.
600mas was between 0.31 and 0.40 magnitude, which is similar to that of other state-of-the-art contrast models presented in the literature. This model is currently being incorporated into the Paranal SCUBA software for first-level quality control and real-time scheduling support. Future work will consider fitting this model to other SPHERE instrumental modes.

Acknowledgements. This work has made use of the the SPHERE Data Centre, jointly operated by OSUG/IPAG (Grenoble), PYTHEAS/LAM/CESAM (Mar

Fig. 7. Predicted vs measured test data contrast, binned by star magnitude (faint, mid, bright), along with the atmospheric categories currently used for ranking and grading SPHERE observations in Paranal.

Fig. 8. The fitted contrast model parameters. [Top] The mean fitted alpha power indices ±2σ as a function of separation. [Bottom] The fitted intercept X(r) (in log space) as a function of separation.

Fig. 9. Comparison between the 400mas contrast predictions made by the model of this work vs the current (as of the publication of this paper) ETC on the ESO website. The ETC options were set to match the observed mode of the model developed here: DBH23 filter with coronagraph. Both the ETC and model were set with no neutral densities, exptime ~64s, DIT≤64s, without differential imaging and considered a spectral type of around GV2.

References

Fig. 10. Example of how the contrast model could be used in operations for providing quick checks to ensure the measured raw contrast (∼60s frame) is within the expected 95% range of model residuals given the target and current atmospheric conditions.

Courtney-Barrer, B., Wahhaj, Z., Moutou, D., & Milli, J. 2019, in AO4ELT6 proceedings
Poyneer, L. A. & Macintosh, B. A. 2006, Optics Express, 14, 7499
Roddier, F. 2004, Adaptive Optics in Astronomy
Sauvage, J. et al. 2015, in Adaptive Optics for Extremely Large Telescopes 4 – Conference Proceedings