Data Mining the Extreme Adaptive Optics Instrument VLT/SPHERE for Raw Contrast Prediction Pre-Observation

Benjamin Courtney-Barrer, Zahed Wahhaj [a], David Mouillet[b], and Julien Milli[a]

[a]European Southern Observatory, Alonso de Crdova 3107, Vitacura, Santiago, Chile
[b]University Grenoble Alpes, CNRS, IPAG, 38000 Grenoble, France

ABSTRACT

Data prepared by the SPHERE data center was utilised to create various candidate models to predict the raw contrast achieved by a common SPHERE observing mode pre-observation. This may be used for improving quality control, exposure time calculations and real-time observation decisions. The final model was selected for testing through a cross validation and verification process on training and verification data sets. The final model used a top level algorithm to classify data into regimes where particular physical processes dominated the achieved contrast, and then considered unique sub-level models for modeling the contrast. For these sub-level models both an empirical approach using machine learning, and a hybrid approach, mixing physical models with machine learning algorithms for parameter estimation, were considered. The final model achieved a RMSE for the predicted raw contrast of 1.31×10^{-4} (log10 RMSE = 0.25) on the test data set which outperformed both the benchmark persistence and physical models.

Keywords: SPHERE, high contrast imaging, data mining, model, machine learning

1. INTRODUCTION

SPHERE (Spectro-Polarimetric High-contrast Exoplanet REsearch[1]) is an extreme adaptive optics (AO) instrument installed on the Melipal Telescope at the Paranal Observatory. Its primary science goal is imaging, low-resolution spectroscopic, and polarimetric characterization of extra-solar planetary systems at optical and near-infrared wavelengths. To help achieve this SPHERE is equipped with a series of coronagraphs in-addition to having an extreme AO system called SAXO[2,3] which operates at a frequency up to 1.38 kHz on bright targets with a 40x40 spatially filtered Shack-Hartmann (SH) wavefront sensor (WFS) and a 41x41 piezoelectric high-order deformable mirror. To date SPHERE has discovered[4,5] and imaged various planetary systems and continues its search.

SPHERE is one among the many AO instruments at Paranal. These come in a range of flavours including Single Conjugate AO, Laser Tomography, Ground Layer AO, and Extreme AO depending on the scientific objective. Accordingly each AO system responds uniquely to changes in atmospheric parameters with respect to the instruments scientific objective. With this increased complexity and range of behavior in each AO system we can no longer assume blanket responses across systems in regards to performance as could be done with the previous generation of seeing limited instruments. Luckily, empirical data is becoming more available with the maturity of these second generation instruments. This data should be used to evolve models, not only to improve methods of quality control and real time scheduling, but also improve exposure time calculations (ETC) and gain further insight into instrument behavior. This also works towards the requirements for the ELT in regards to optimal exploitation of atmospheric conditions, flexible short-term scheduling driven by real-time decisions, and observation classification within 10 minutes.

Here we take advantage of SPHERE data mainly prepared by the SPHERE data center[6] to develop a new model framework for predicting the raw contrast achieved by SPHERE pre-observation. We take this opportunity to integrate lessons learned from the years of SPHERE's operation and propose a top level classification model for predicting when known instrumental effects (such as the low wind effect[7,8]) dominate the achieved

For further information email Benjamin Courtney-Barrer at bcourtne@eso.org

Variable Label	Description
'exptime'	science exposure time
'simbad_FLUX_XX'	target flux in R-band or H-band (XX=R or H)
'tau0_massdimm'	atmospheric coherence time
'seeing_massdimm'	atmospheric seeing
'Vo'	turbulent velocity (0.315 r_o/τ_o)
'Air Pressure Normalised [hPa]',	normalized air pressure
'Air Temperature at XX [C]'	air temperature a XX=2m,30m
'Dew Temperature at 2m [C]',	dew temperature at 2m
'Relative Humidity at 2m [%]',	relative humidity at 2m
'Wind Direction at 30m (0/360) [deg]',	wind direction at 30m
'Wind Speed at 10m',	Wind Speed at 10m
Wind Speed XX at 20m	wind speed vector components, XX = U, V, W where U,V is horizontal component at 330deg and 240deg respectively, W is vertical component
'summer_axis'	map observation date onto unit circle and project onto the summer axis
'autumn_axis'	map observation date onto unit circle and project onto the autumn axis

Table 1. variables considered for model features

contrast. For each classification class we then have class specific models where we explore both a purely empirical approach with machine learning, and also a hybrid approach of mixing physical models with machine learning algorithms for parameter estimation from empirical data. This work is one further step towards flexible and optimal short term scheduling that fully exploit the atmospheric conditions for SPHERE in-addition to improving ETC. Furthermore, this may serve as an example for the models developed to meet the ELT requirements.

2. METHOD

2.1 Data

Models were created for predicting the raw achieved contrast of a SPHERE / IRDIS mode from measured atmospheric conditions. We considered the most common IRDIS planet searching mode which achieves the deepest contrast. This mode uses the N_ALC_YJH_S coronagraph with science filter D_H23 and AO frequency = 1.38kHz with gain = 1000. Future work will extend this model to other SPHERE modes. The SPHERE /IRDIS observations considered in this work were provided by the SPHERE data center[6] and consisted of 53 observation blocks (after data cleaning) taken between December 2015 to May 2018, with raw contrast calculated pxielwise at 300, 400 and 700mas. Various physical variables were considered for the analysis, with the primary ones outlined in table 1. These were measured by the array of astronomical site monitoring instruments[9] at Paranal which include telescope sensors, the meteo station, Multi-Aperture Scintillation Sensor (MASS) and Differential Image Motion Monitor (DIMM). These variables were generally sampled every 1-5 minutes and were interpolated with respect to the IRDIS exposure readout times. Extreme outliers were dropped from the data and no interpolation was done where there were large gaps (more than 20 minutes) in the data. The data was randomly partitioned into a training, verification and test set with 60%, 20%, 20% ratios respectively.

2.2 Methods for Dimension Reduction and Feature Selection

2.2.1 Kernel Principle Component Analysis

The data features were explored in both its raw physical interpretation (e.g. temperature, seeing etc), and also in its kernel principle component projections. Kernel principle components were used so that more complex non-linear structures in the data's distribution could be captured. The full mathematical details of kernel PCA can be found in Schlkopf's 1997 paper.[10] The key idea is that the covariance matrix used to calculate the

Kernel Name	Analytic Expression	Constraints
Linear (normal PCA)	$K(x,y) = (x \cdot y + 1)$	-
Polynomial	$K(x,y) = (x \cdot y + 1)^d$	$d > 0$
Radial Basis function	$K(x,y) = \exp\left(-\gamma \|\|x - y\|\|^2\right)$	$\gamma > 0$
Sigmoid	$K(x,y) = \tanh\left(\alpha x \cdot y + r\right)$	$\alpha > 0, \overset{\circ}{<}0$
Cosine	$K(x,y) = \frac{x \cdot y}{\|\|x\|\|\|\|y\|\|}$	-

Table 2. Kernels used for kernel principle component analysis

eigenvectors and eigenvalues (i.e. principle components) of M centered data points $\mathbf{x}$ in some given vector space R^n relies on calculating the inner products of the respective vectors. i.e. $C = 1/M \sum_{j=0}^{M} x_j x_j^T$ to solve eigenvalue equation: $\lambda v = Cv$ Where v are the eigenvectors of C, also known as the datas principle components. Using a kernel function K(x,y), Mercers Theorem guarantees the existence of a mapping to another space F through the function ϕ. i.e. $\phi : R^n \rightarrow F$, where the kernel function K(x,y) acts as an inner product in the this mapped space: $K(x,y) = \langle \phi(x), \phi(y) \rangle$. Therefore the covariance matrix in this new mapped space is: $C = 1/M \sum_{j=0}^{M} \phi(x_j)\phi(x_j)^T$. Using the kernel property described above (known as the kernel trick) this eigenvalue problem can be reformulated to find the principle components (α) in this higher dimensional space F and project down to the original space R^n without ever having to visit the space F or even explicitly know the mapping ϕ - instead just explicitly calculate the kernel function itself. The reformulated eigenvalue problem is: $M\lambda\alpha = K\alpha$ Where K is a MxM matrix with elements $K_{ij} = \langle \phi(x_i), \phi(x_j) \rangle$. When projected to the original R^n space these principle components can therefore manifest as complex non-linear functions. The kernels considered in this work were implemented with the scikit-learn package in python and are outlined in Table 2.

2.2.2 Mutual Information

Since a variety of candidate models were considered in this analysis, mutual information, which is a model independent metric, was used for feature selection. The mutual information measures the information gained (in bits) for predicting a label given knowledge about a feature. Mathematically the mutual information I(X,Y) between variables X and label Y is defined as the difference in the full and conditional entropy H i.e. $I(X,Y) = H(X) - H(X|Y)$ where $H(X) = -E[log(P(X))]$. Explicitly The mutual information can be expressed for probability distributions in X and Y:

$$I(X,Y) = \sum_{y} \sum_{x} \frac{P_{XY}(x,y) \log(P_{XY}(x,y))}{P_X(x) P_Y(y)} \tag{1}$$

Where $P_{XY}(x,y)$ is the joint distribution and $P_X(x)$, $P_Y(y)$ are the marginal probability distributions. Typically non-parametric estimations are used to calculate the mutual information for N points without model specific assumptions. For this analysis we use k-nearest neighbors (knn) estimator[11,12] due to its efficiency in multidimensional settings. This method is implemented with the scikit-learn package in python which uses the nearest neighbours to define local neighbourhoods for probability-volume elements used in estimating 1.

2.3 Machine Learning Algorithms

In this work various regression models were developed using machine learning algorithms. Therefore a variety of standard off-the-shelf candidate machine learning algorithms were considered over the training and verification process before final model selection for testing. These candidate machine learning algorithms are outlined below and were implemented with the scikit-learn python package.

- Random Forest Regressor (RF)
- Multi Layer Perceptron (MLP) - also known as a feed forward neural network
- K-Neighbors Regressor (K-NN)
- Kernel Ridge Regressor (KR)

2.4 Top Level Model Structure

The top level of the model used a classification model to classify data into regimes where distinct physical processes dominate the achieved contrast. This work only considered a thermal regime, where local thermal effects such as the low wind effect[7] and dome seeing[13] dominated the achieved contrast, and a standard regime where the instrument should be operating under "non-extreme" physical conditions. The wind-driven halo effect was not considered in this work, however it may be considered in future work. Each classified regime then had unique models that were selected through cross validation and parameter tuning of the train and verification data sets. This model structure is shown in figure 1. The classification model was developed using known institutional knowledge of telescope/instrument behavior in combination with statistical analysis of the available data.

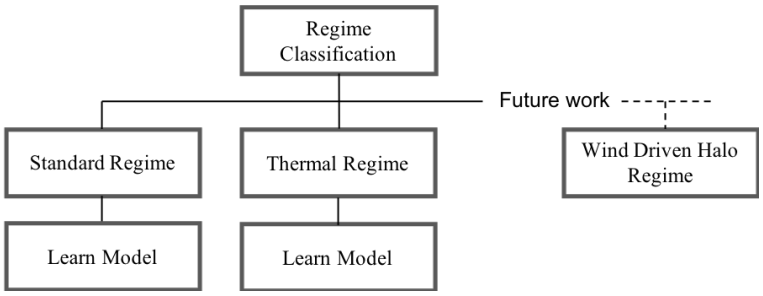

Figure 1. Top level model structure for regime classification

2.5 Standard Regime Candidate Models

For the standard regime we considered two candidate models: one empirical model learnt purely through historic data with machine learning algorithms, and a hybrid approach where additional parameters were added to a simple physical model proposed by Serabyn et al,[14] and models for the parameter were learnt via machine learning algorithms.

During testing these models were benchmarked against the raw physical model (without the additional learnt parameters) and also a persistence model with initial conditions set to the mean train-set contrast at the respective radius considered. To avoid biases, the training set was used for feature selection and reducing the number of candidate models via cross-validation performed with a course-grid search, while verification sets were then used for fine parameter tuning and final model selection for testing.

2.5.1 Hybrid Model

A hybrid physical/empirical model was considered where a simple physical model for coronagraphic AO systems taken from Serabyn et al[14] was adapted to include additional parameters which were fitted on the train set and then modelled by machine learning algorithms. The basic principle of the physical model is that the scattered light from a NxN actuator deformable mirror can be approximated by Fourier theory as 1-Strehl which appears in a halo of size NxN spatial resolution elements (λ/D).[14] Therefore the limiting contrast that occurs in this halo of scattered light may be approximated as $(1 - Strehl)/N^2$. The strehl ratio was approximated by the exponent of the residual phase errors (σ^2). i.e: Strehl = $exp(-\sigma^2)$, Where we consider the following explicit phase error terms of shack-hartmann (SH) fitting (σ^2_{fit}), shack-hartmann alaising (σ^2_{alias}), AO servo lag (σ^2_{servo}), and photon

shot noise (σ^2_{phot}) defined as:

$$\sigma^2_{fit} = 0.257 \left(\frac{D}{r_o}\right)^{5/3} N_{act}^{-5/6} \tag{2}$$

$$\sigma^2_{alias} = 0.3\sigma^2_{fit} \tag{3}$$

$$\sigma^2_{servo} = \left(\frac{\tau}{\tau_o}\right)^{5/3} \tag{4}$$

$$\sigma^2_{phot} = \frac{\lambda_{WFS}}{\lambda_{im}} \left(\frac{d_{subpup}}{r_{o,WFS}}\right)^2 \frac{2\pi}{N_{photons}} \tag{5}$$

Where D is telescope diameter, N_{act} is total number of deformable mirror actuators, τ is the servo loop time, τ_o is atmospheric coherence time, λ_{WFS} and λ_{im} are the wavelengths used for the wavefront sensor and coronagraphic image respectively, $r_{o,WFS}$ is the atmospheric coherence length at the wavelength of the wavefront sensor, d_{subpup} is the SH sub-pupil diameter, and $N_{photons}$ is the total number of photons received in the wavefront sensor during a single exposure.

Note that read out noise was not significant in the considered mode and therefore neglected in the noise budget along with calibration errors. Also anisoplanetism errors were irrelevant since SPHERE AO is on axis with a small FOV. Therefore the full benchmark physical model is:

$$C = \frac{1 - exp(\sigma^2_{fit} + \sigma^2_{alias} + \sigma^2_{servo} + \sigma^2_{phot})}{N_{act}} \tag{6}$$

The hybrid model was then developed by including a time dependent scaling factor ($\alpha(t)$) for the error budget terms, and an additional static noise error budget term (σ_{static}) to account for calibration and non common path aberrations (NCPA) between the science detector and wavefront sensor.

$$C = \frac{1 - exp(\alpha(t)(\sigma^2_{fit} + \sigma^2_{alias} + \sigma^2_{servo} + \sigma^2_{phot}) + \sigma_{static})}{N_{act}} \tag{7}$$

The idea was to develop models for these parameters $\alpha(t), \sigma_{static}$ using machine learning algorithms outlined in section 2.3. The parameters were first fitted to the training data and then relationships between the fitted parameters and other environmental variables were searched for to find effective features for the models. Since $\alpha(t)$ varies in time while σ_{static} is constant for a given observation run; an iterative fitting approach was taken where $\alpha(t)$ was fitted dynamically to the data within a moving box of 120 seconds, and then σ_{static} was fitted statically on the entire time series until a reasonable level of convergence was reached. The verification data set was then used for finer parameter tuning and then to compare the hybrid and empirical models.

2.5.2 Empirical Model

A purely empirical model was considered without any prior physical knowledge of the system. Here we used the machine learning algorithms introduced in section 2.3 to learn a predictive contrast model from the training data, and then used the verification data for finer parameter tuning and to compare the empirical and hybrid models.

2.6 Thermal Regime Candidate Models

Since limited data was available where thermal effects dominated the contrast, data was considered from both the train and verification sets and an empirical model was used where models were trained with the candidate machine learning algorithms outlined in section 2.3.

3. MODEL TRAINING AND SELECTION

3.1 Top Level Regime Classification

The motivation for a top level regime classification model is that many of the non-standard regimes where distinct physical processes dominate (e.g. dome seeing) do not occur often, and therefore without extensive training data it is very difficult for models to accurately learn these effects. During the early analysis before a top level model was implemented it was clear that the outliers of any trained model when tested on the verification set occurred at low wind speeds and ambient temperatures around 15C, which happens to be the maximum temperature which the M1 mirror is actively cooled. In-addition the physical processes of dome seeing and low wind effect are well understood and therefore this knowledge can be utilized to make learning the models an easier task. Analysis of the train data set clearly highlights how thermal differences within the telescope optics and surrounding environment as well as low wind effects dominate the achieved contrast. Furthermore thermal effects commonly dominated when there was excellent seeing and coherence time. Furthermore the wavefront sensor spatial filters are generally selected in real-time based on these two measurements. This issue can clearly be seen in figure 7 which shows a clear bi-modal distribution of the contrast for the small wavefront sensor spatial filters. Defining a threshold between mode 1 (good contrast) and mode 2 (bad contrast) for the small wavefront sensor filter at log10 contrast = -3.75 we can clearly discriminate these modes with the ambient temperature and wind speed indicating that mode 2 (bad contrast) is indeed in a regime where thermal effects are dominate. These results highlight why thermal effects should also be considered for real-time observation decisions and quality control.

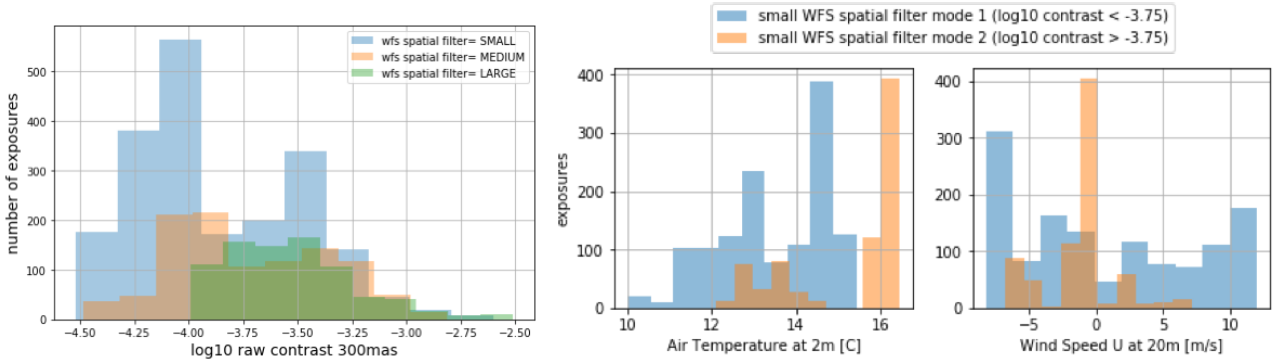

Figure 2. [Left] histogram raw contrast for different wavefront sensor (wfs) spatial filters. [Middle/Right] Splitting the bi-modal contrast distribution for the small wfs spatial filter into two modes, we plot histograms of temperature and wind speed for each mode and can clearly discriminate between them.

In the train and verification data sets only 2 observations were considered to be in a regime where thermal effects dominated the achieved contrast. This classification was done by manual inspection by considering thermal parameters in-addition to the achieved contrast relative to atmospheric conditions. Based on the best discrimination threshold for the thermal regime with respect to these parameters, the following simple threshold filter was used to classify data into the standard or thermal regime:

If wind speed magnitude at 20m<3m/s & vertical wind speed at 20m<1m/s & Temperature at 2m<13.5C

Then thermal regime

Else standard regime

Based on this classification the test data set held one observation classified in the thermal regime.

3.2 Standard Regime - Training, Verification and Model Selection

For the standard model two different approaches were experimented; a purely empirical model and a hybrid model as described above.

3.2.1 Empirical Model

For selecting features for the empirical model we consider the mutual information of the log10 contrast for the training data against both the physical parameters, and also the projected kernel principle components. Observations that were classified in the thermal regime were removed. Figure 7 plots the mutual information for given parameters. Note for the kernel PCA we only show the results from the rbf and polynomial kernel as these were the best performing of the kernel methods. Interestingly the mass-dimm coherence time and seeing parameters

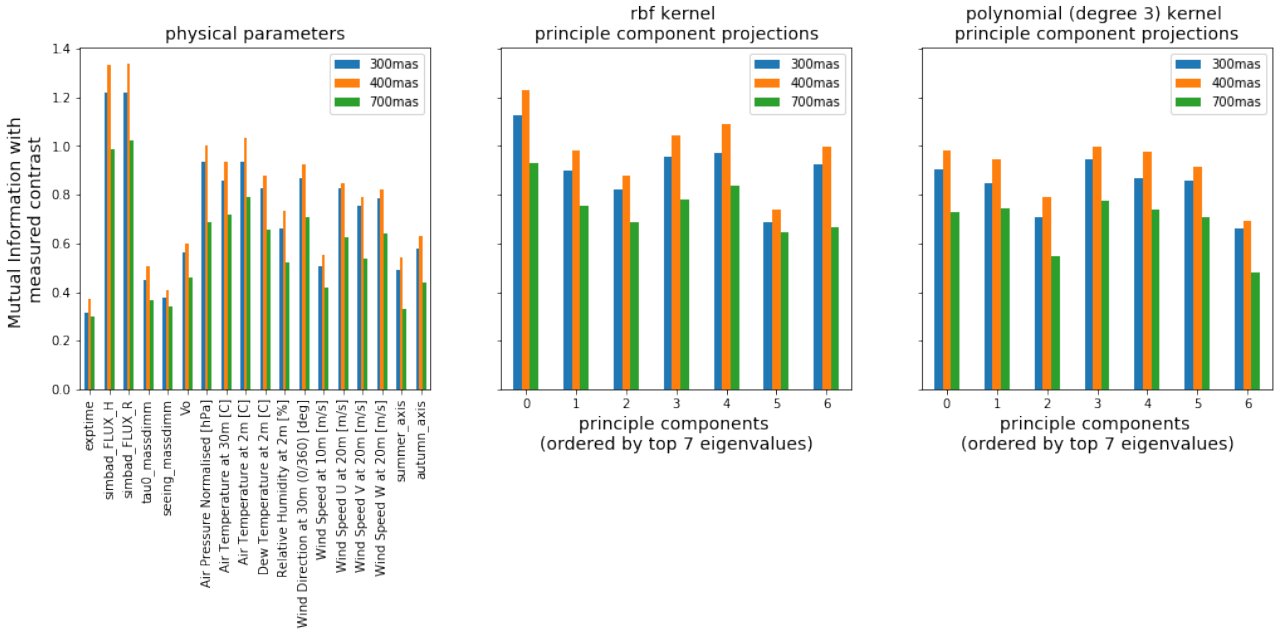

Figure 3. Mutual Information calculated between the log10 contrast and various candidate features derived from physical measurements and kernel principle component projections of the physical variables

held relatively low mutual information with the contrast while the star magnitude was the most informative parameter. The reason for this lower than expected score for τ_o and seeing is that the considered instrument mode for IRDIS was one of the most extreme modes with the fastest AO loop speed at 1380Hz (0.7ms). Therefore the dominating noise source generally came from low photon counts on the SH.

Platform level wind parameters generally had a high mutual information with contrast. This was expected as wind direction and speed are known to correlate strongly with the turbulence profiles brought to Paranal. For example subtle (5m/s) westward winds coming from the ocean correlate strongly with excellent conditions. However surprisingly a range of unexpected atmospheric parameters such as pressure and temperature held relatively high mutual information with contrast. Some mutual information with these parameters could be expected due to correlations between variables, for example, pressure is known to drive wind speed which has correlations with Paranal conditions. Also, even though data with extreme thermal effects were classified in the thermal regime, less extreme cases that weren't filtered out can still produce local turbulent effects that affect performance. These arguments at most would suggest weak to moderate mutual information with contrast, however given the relatively high score, further explanations were sought. It was found that some of these parameters had large un-physical jumps in their measurements that correlated strongly with jumps in SPHERE's contrast. The explanation for this is that the telescope pointing model was updated every three minutes with the temperature, pressure and humidity measurements to account for atmospheric differential refraction. Hence measurement errors in these variables corresponded to telescope pointing errors. Removing these data points reduced the contrast mutual information with these variables. Since these measurements were taken further upgrades and improvements have been made to Paranal's atmospheric site monitoring instruments to minimize these effects. The relatively high dependence between pressure, temperature and contrast has also been found

in data from the GPIES instrument at the Gemini Observatory, which is another high contrast imaginer similar to SPHERE[13,15]

From the above analysis two sets of features were considered; one pertaining to physical variables and the other to rbf kernel principle components. The physical features were selected due to having the highest mutual information with contrast in their group, with the exception of $\mathrm{Vo} = 0.315 r_0/tau_o$ for the physical features, which was also included based on analysis from a random forest feature importance (not included here for brevity). The physical features selected were: R band flux, turbulence speed V_o, air Temperature at 2m, normalized air pressure, and wind direction. The rbf kernel principle components were selected based on eigenvalue magnitude (captured variance). The top four eigenvectors were selected for the rbf kernel features.

For each set of features a 5-fold cross-validation on the train set was performed for each candidate model over various points in a course parameter grid search. The model performance was assessed based on the average mean absolute error (MAE) and explained variance. The 2 best models were then selected for verification. These results are shown in 4. Both the neural network (MLP) and Random Forest (RF) with the rbf kernel principle

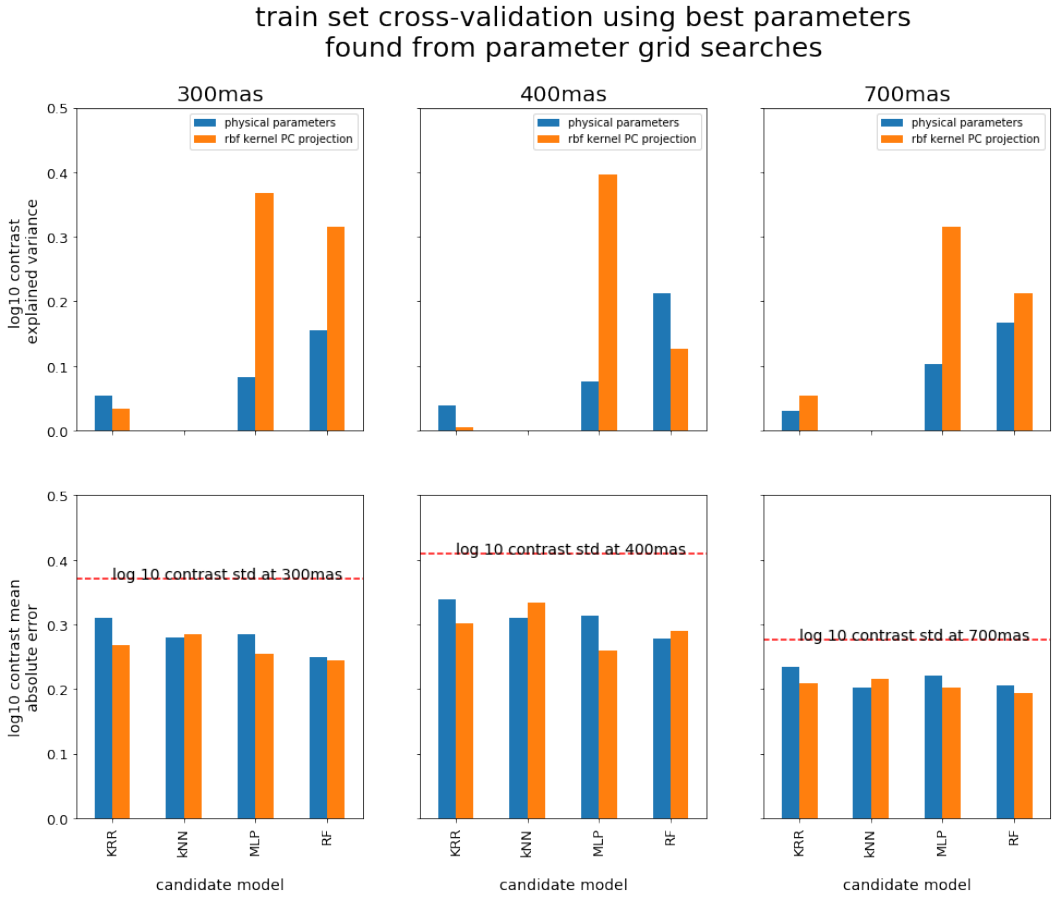

Figure 4. explained variance and mean absolute error for 5-fold cross-validation on the train data set for each candidate machine learnt model after a course grid parameter search.

components as features performed best across all radii besides RF for 400mas. The verification set was then introduced to perform a finer grid search for these two models and a final comparison was performed. It was found that the trained neural network essentially learned the train set mean with little variance and therefore did not generalize to new data. Therefore the random forest was selected as the final model with that used 10 tree estimators, with a MSE splitting criteria, no maximum tree estimator depth, minimum samples required for

split = 2, and min samples for leaf node = 1.

3.2.2 Hybrid Model

Using the training data set the hybrid model parameters were fitted from equation 7. The parameter α was dynamically fitted by least squares within a moving window of 5 minutes (typically a few exposures) using a prior estimate of σ_{static}. The σ_{static} parameter was then subsequently fitted using the prior calculated $\alpha(t)$. This cycle was repeated until a reasonable convergence. It was found that the convergence was not unique and depended on the initial conditions of σ_{static}. Therefore 3 prior values (high, medium, low) were used for σ_{static} and α, σ_{static} fitting iterations were repeated 5 times (early stopping). The converged parameters α, σ_{static} from the 3 initial conditions were then averaged for the final result. A histogram of the fitted parameter values are shown in figure 5. Both the fited α and σ_{static} parameters followed a skewed lognormal distribution. Assuming

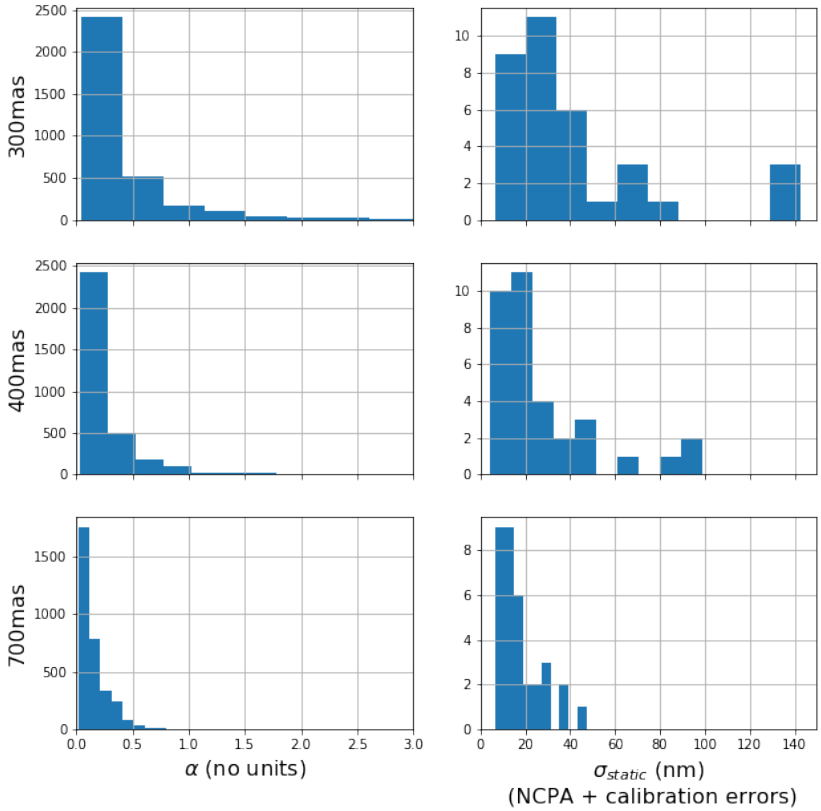

Figure 5. Histograms of fitted α, σ_{static} parameters on training data set

σ_{static} dominated by NCPA these results agree well with those measured by ZELDA which measured the NCPA in SPHERE in the range of 30 - 60nm over two nights[16] .

In general, it was found that kernel principle component projections held the most collective mutual information (rbf and 3rd degree polynomial performing best for α and σ_{static} respectively as shown in figure 6). Nevertheless both physical variable and kernel principle component features were tested in the train set 5-fold cross-validation. Features for α and σ_{static} parameters were selected via the same method as the empirical model as outlined in section 3.2.1. Tables 3 and 4 outline the train set cross validation performances (using the explained variance, and mean absolute error as metrics) of the respective machine learning algorithms applied to α and σ_{static}. The specific physical variables used are explained in the table captions. Based on these results the hybrid model verification (and comparison against the empirical model) was done with a multi-layer perceptron (feed forward neural network) for both the $\alpha(t)$ and σ_{static} parameters using the top four ranked RBF kernel

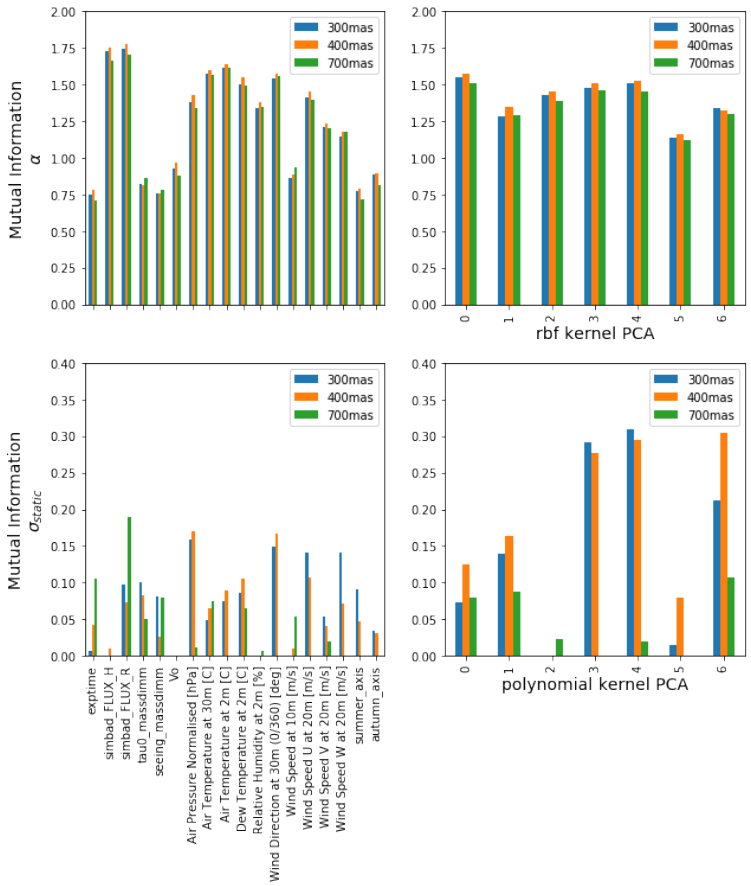

Figure 6. Mutual information of α (top row) and σ_{static} (bottom row) for both physical parameters (left column) as well as the rbf and 3rd degree polynomial kernel projections (right column) respectively.

radius	$\log_{10}\alpha(t)$ at 300mas (MAE/R^2)		$\log_{10}\alpha(t)$ at 400mas (MAE/R^2)		$\log_{10}\alpha(t)$ at 700mas (MAE/R^2)	
features	physical*	RBF kernel	physical*	RBF kernel	physical*	RBF kernel
KRR	0.34 / 0.04	0.31 / 0.13	0.34 / 0.02	0.32 / 0.13	0.29 / 0.02	0.26 / 0.16
kNN	0.32 / 0.00	0.35 / 0.00	0.31 / 0.00	0.37 / 0.00	0.25 / 0.08	0.32 / 0.00
MLP	0.32 / 0.09	0.30 / 0.12	0.32 / 0.08	0.31 / 0.11	0.24 / 0.24	0.26 / 0.16
RF	0.35 / 0.00	0.30 / 0.07	0.35 / 0.00	0.32 / 0.00	0.30 / 0.00	0.29 / 0.00

Table 3. Mean cross validation mean absolute error (MAE) and explained variance (R^2) for 5-fold cross validation on train set across all candidate models for the $\alpha(t)$ parameter, comparing both physical variables and kernel principle component projections as features. *Physical variables at 300,400 & 700mas were R flux, temperature at 2m , wind direction, wind speed U at 20m.

PCA and 3rd degree polynomial PCA projections as features respectively. The $\alpha(t)$ MLP used 1 hidden layers with 50 neurons and a logisitc activation function while the σ_{static} MLP used 2 hidden layers with 20 neurons and logistic activation function.

3.2.3 Verification

The verification data points classified in the standard regime were then used to select between the hybrid and empirical model for final testing. Both models had comparable results. The empirical model had a lower log10

radius	$\log_{10}\sigma_{static}$ at 300mas (MAE/R^2)		$\log_{10}\sigma_{static}$ at 400mas (MAE/R^2)		$\log_{10}\sigma_{static}$ at 700mas (MAE/R^2)	
features	physical*	Poly kernel	physical*	Poly kernel	physical*	Poly kernel
KRR	0.26 / 0.08	0.25 / 0.13	0.27 / 0.08	0.27 / 0.11	0.17 / 0.03	0.17 / 0.08
kNN	0.27 / 0.00	0.25 / 0.19	0.28 / 0.00	0.27 / 0.17	0.18 / 0.00	0.17 / 0.09
MLP	0.24 / 0.13	0.23 / 0.23	0.26 / 0.14	0.25 / 0.19	0.16 / 0.12	0.17 / 0.06
RF	0.28 / 0.00	0.28 / 0.04	0.28 / 0.08	0.28 / 0.14	0.18 / 0.00	0.19 / 0.00

Table 4. Mean cross validation mean absolute error (MAE) and explained variance (R^2) for 5-fold cross validation on train set across all candidate models for the σ_{static} parameter, comparing both physical variables and kernel principle component projections as features. *Physical variables at 300mas: were Air pressure, wind direction, wind speed U 20m, wind speed W 20m., *Physical variables at 400mas: exptime, flux R, seeing, air temperature. & *Physical variables at 700mas: exptime, flux R, seeing, air temperature

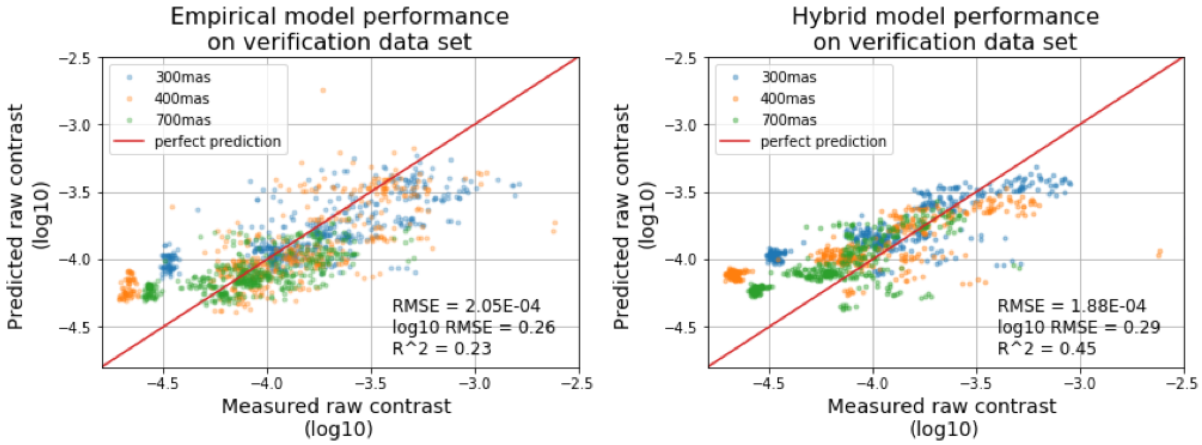

Figure 7. Predicted vs Measured raw contrast for the empirical (left) and hybrid (right) models on the verification data set.

RMSE indicating slightly better performance at lower contrasts, while the hybrid model had a lower RMSE indicating better performance at higher contrasts. However the hybrid model also had more explained variance (R^2) which ultimately led to the decision to use the hybrid model for final testing.

3.3 Model Selection - Thermal Regime

There were only 2 observations in the train and verification sets that were classified in the thermal regime. Hence the idea of cross validating various candidate models was somewhat meaningless in the hope of generalization. Therefore a more direct approach was taken where we used the simple k-nearest neighbors algorithm using uniform weighting and k=10 which was trained on the combination of train and verification data sets (80% on training, 20% test).

4. TEST RESULTS

The full model (considering both standard and thermal regime) was tested on the test data set. The performance is outlined in figure 8 and we compare the results to both the benchmark physical model (without $\alpha(t)$ or σ_{static} parameters) and also a persistence model (modelling the contrast as the train set average) in table 5.

5. DISCUSSION

The model generalized reasonably well on the test data which contained 11 observations, with 1 observation classified in the thermal regime. The test performance was slightly better then the standard regime hybrid model verification results in regards to the models RMSE. This slight improvement can be explained by the fact that the mean test contrasts were 18% lower than the mean verification contrasts. However the explained variance

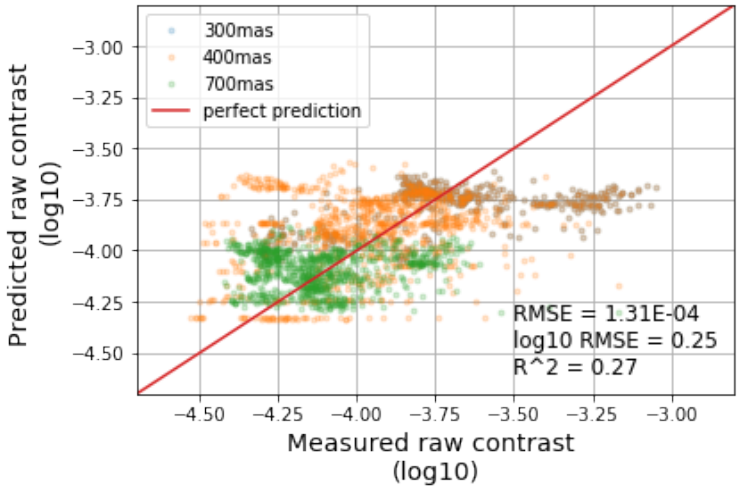

Figure 8. Predicted vs Measured raw contrast on the test data set for the full model, using a hybrid model in the standard regime and empirical model in the thermal regime.

-	hybrid model	persistence model	benchmark physical model*
RMSE	1.31×10^{-4}	1.46×10^{-4}	1.49×10^{-4}

Table 5. RMSE comparison to benchmark models, scaling factors applied to benchmark physical model at 300,400,700mas were 0.6, 0.5, 0.25 respectively

was considerably lower which was partly due to the poorer predictions of the contrast in the thermal regime due to limited data. The test results for the full model also had a considerable improvement in performance against the benchmarked persistence model and benchmark physical model as outlined in table 5. Comparing these results to literature, D.Savranskya et al[15] mined the GPIES data base and were able to predict the reduced contrast (pre-observation) with an log10 RMSE of 0.18 (RMSE = 5E-5) using a purely empirical two layer, 6 feature (prior to observations), 16 neuron MLP neural network. This result[15] is roughly a factor of 2 better than achieved here. Differences in results can be attributed that fact that they had much more data (we trained on only 35 observations) and were considering the reduced (rather than raw) contrast which averages out some of the difficult-to-predict stochastic processes present in each exposure. Considering these challenges the results reported here are encouraging, and point to the advantages of using hybrid methods where prior physical knowledge/models of the instrument are extended using statistical methods and machine learning algorithm to predict difficult parameters and modify the physical model with regard to empirical data.

Current work is being done to develop a pipeline for calculating the raw contrast in realtime as SPHERE observations are being taken. These results are being added to the The MSE DataLab[17] database which contains a huge array of telescope and instrument data ready for machine learning applications. This database will be used to test and improve this model in-addition to developing new models.

6. CONCLUSION

Various candidate models were considered to predict the raw contrast achieved by a SPHERE / IRDIS mode prior to observations were considered. A final model was selected for testing through a cross validation and verification process on train and verification data sets. The final model tested used a top level algorithm to classify data into regimes based on dominating physical processes, and then considered unique sub-level models for modeling the contrast. For these sub-level models both an empirical approach using machine learning, and hybrid approach mixing both physical models with machine learning algorithms for parameter estimation were used. The final model achieved a RMSE contrast of 1.31×10^{-4} on the test data set.

REFERENCES

[1] Beuzit, J.-L. et al., "Sphere: a planet finder instrument for the vlt," in [*Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*], *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series* **7014** (2008).

[2] Fusco, T. et al., "Saxo, the sphere extreme ao system: on-sky final performance and future improvements," in [*Adaptive Optics Systems V Series*], *Proc. SPIE* **9909** (2016).

[3] Sauvage, J.-F. et al., "Saxo: the extreme adaptive optics system of sphere (i) system overview and global laboratory performance," *Journal of Astronomical Telescopes, Instruments, and Systems* **2** (2016).

[4] Keppler, M. et al., "Discovery of a planetary-mass companion within the gap of the transition disk around pds 70," *Astronomy Astrophysics* **617** (07 2018).

[5] Chauvin, G. et al., "Discovery of a warm, dusty giant planet around hip65426," *Astronomy and Astrophysics* **605** (07 2017).

[6] Delorme, P. et al., "The sphere data center: a reference for high contrast imaging processing," (12 2017).

[7] Milli, J. et al., "Low wind effect on vlt/sphere : impact, mitigation strategy, and results," in [*Proc. SPIE 10703, Adaptive Optics Systems VI, 107032A (2018)*],

[8] Sauvage, J. et al., "Low wind effect, the main limitation of the sphere instrument," in [*Adaptive Optics for Extremely Large Telescopes 4 Conference Proceedings*], (2015).

[9] Dorigo, D. et al., "Astronomical Site Monitor Data User Manual."

[10] B, S., A, S., and KR, M., [*Kernel principal component analysis. In: Gerstner W., Germond A., Hasler M., Nicoud JD. (eds) Artificial Neural Networks*], Springer, , Berlin, Heidelberg (1997 (eleventh edition)).

[11] Kraskov, A., Stogbauer, H., and Grassberger, P., "Estimating mutual information," *Phys. Rev* **69** (2004).

[12] Ross, B. C., "Mutual information between discrete and continuous data sets," *PLoS ONE* **9(2)** (2014).

[13] Tallis, M. et al., "Air, telescope, and instrument temperature effects on the Gemini Planet Imager's image quality," *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series* **10703**, 1070356 (Jul 2018).

[14] Serabyn, E. et al., "Extreme adaptive optics imaging with a clear and well-corrected off-axis telescope subaperture," *The Astrophysical Journal* **658** (02 2007).

[15] Savransky, D. et al., "Mining the gpies database," in [*Astronomical Telescopes + Instrumentation*], (2018).

[16] Vigan, A. et al., "Calibration of quasi-static aberrations in exoplanet direct-imaging instruments with a Zernike phase-mask sensor. III. On-sky validation in VLT/SPHERE," **629**, A11 (Sep 2019).

[17] Eduardo, P. et al., "Framework to use modern big data software tools to improve operations at the paranal observatory," 93 (07 2018).