

# Robustness of prediction for extreme adaptive optics systems under various observing conditions

## An analysis using VLT/SPHERE adaptive optics data.

M. A. M. van Kooten<sup>1,\*</sup>, N. Doelman<sup>1,2</sup>, and M. Kenworthy<sup>1</sup>

<sup>1</sup> Leiden Observatory, Leiden University, Niels Bohrweg 2, 2333 CA Leiden

<sup>2</sup> TNO, Stieltjesweg 1, 2628 CK Delft

Received October xxx, 2019; accepted xxx, 2019

### ABSTRACT

**Context.** For high-contrast imaging (HCI) systems, such as VLT/SPHERE, the performance of the system at small angular separations is contaminated by the wind-driven halo in the science image. This halo is a result of the servo-lag error in the adaptive optics (AO) system due to the finite time between measuring the wavefront phase and applying the phase correction. One approach to mitigating the servo-lag error is predictive control.

**Aims.** We aim to estimate and understand the potential on-sky performance that linear data-driven prediction would provide for VLT/SPHERE under various turbulence conditions.

**Methods.** We used a linear minimum mean square error predictor and applied it to 27 different AO telemetry data sets from VLT/SPHERE taken over many nights under various turbulence conditions. We evaluated the performance of the predictor using residual wavefront phase variance as a performance metric.

**Results.** We show that prediction always results in a reduction in the temporal wavefront phase variance compared to the current VLT/SPHERE AO performance. We find an average improvement factor of 5.1 in phase variance for prediction compared to the VLT/SPHERE residuals. When comparing to an idealised VLT/SPHERE, we find an improvement factor of 2.0. Under our 27 different cases, we find the predictor results in a smaller spread of the residual temporal phase variance. Finally, we show there is no benefit to including spatial information in the predictor in contrast to what might have been expected from the frozen flow hypothesis. A purely temporal predictor is best suited for AO on VLT/SPHERE.

**Conclusions.** Linear prediction leads to a significant reduction in phase variance for VLT/SPHERE under a variety of observing conditions and reduces the servo-lag error. Furthermore, prediction improves the reliability of the AO system performance, making it less sensitive to different conditions.

**Key words.** instrumentation – adaptive optics – numerical

## 1. Introduction

In the search for new exoplanets and earth analogs, dedicated high-contrast imaging (HCI) systems have allowed the angular separation of host stars from their surroundings to reveal circumstellar disks and exoplanets. The combination of extreme adaptive optics (XAO) to provide high spatial resolution, coronagraphs to suppress the host star's light, and data reduction techniques to remove residual effects, allows HCI systems to reach post-processed contrasts of  $10^{-6}$  at spatial separations of 200 milliarcseconds (Zurlo, A. et al. 2016). VLT/SPHERE is a HCI system that has discovered two confirmed planets: HIP 65426b (Chauvin, G. et al. 2017) and PDS 70b (Keppler et al. 2018). It has also discovered a vast array of debris, protoplanetary, and circumstellar disks (e.g., Avenhaus et al. 2018; Sissa, E. et al. 2018). Operating with a tip/tilt deformable mirror (TTDM), a 41-by-41 high order deformable mirror (HODM), and a Shack-Hartmann wavefront sensor (SHWFS) sampling at 1380 Hz, the XAO system of VLT/SPHERE, SAXO delivers Strehl ratios greater than 90 % in the H band (Beuzit et al. 2019).

A major challenge with VLT/SPHERE (and other HCI systems to varying degrees) is the presence of the wind-driven halo (WDH) that dominates the wavefront error at small angular separations. The WDH is a manifestation of the servo-lag error and appears as a butterfly pattern in the coronagraphic/science images (see Cantalloube, F. et al. 2018, for details on the WDH). The servo-lag error is due to the finite time between the measurement of the incoming wavefront aberration (caused by atmospheric turbulence) and the subsequent applied correction. The resulting wavefront error is due to the outdated disturbance information and the closed-loop stability constraints. Owing to this, the halo is aligned with the dominant wind direction and severely limits the contrast at small angular separations even after post-processing techniques. The servo-lag error prevents VLT/SPHERE from achieving its optimal performance when coherence times are below 5 ms (Milli et al. 2017).

The XAO system, SAXO, has a temporal delay of approximately 2.2 SHWFS camera frames. The HODM is controlled using an integrator with modal gain optimisation (Petit et al. 2014). Within this framework, one solution to minimise the delay in SAXO itself is to run everything faster. However, this solution poses a number of hardware challenges with a new HODM that

\* For author correspondence email vkootenstrw.leidenuniv.nl

can run at the desired speed, a wavefront sensor camera with fast readout, and a more powerful real-time-computer. One alternative solution is to upgrade the controller with a control scheme that predicts the evolution of the wavefront error over the time delay. In this paper, we look at the potential of prediction to improve the performance of SAXO especially when the servo-lag error is the dominant residual wavefront error source (i.e., small coherence times).

Many different groups have worked on predictive control as a means of improving the performance of an adaptive optics (AO) system by minimising the servo-lag error. We highlight a few results from the last 15 years. Prediction, within the context of optimal control, is an ingredient in finding the optimal controller. Linear quadratic Gaussian (LQG) control has been explored by [Petit et al. \(2008\)](#) for general AO systems to perform vibration filtering with the Kalman filter. On-sky demonstrations of the LQG controller for tip-tilt/vibrational control are provided in [Sivo et al. \(2014\)](#). Laboratory work to include higher order modes for atmospheric turbulence compensation ([Le Roux et al. 2004](#)) has demonstrated a reduction in the temporal error showing predictive capabilities. The H2 optimal controller (closely related to LQG) has been tested on-sky ([Doelman et al. 2011](#)) showing a reduction in the temporal error for tip-tilt control. For multi-conjugate AO systems, a temporal aspect to the phase reconstruction for each layer has been implemented; the spatial-angular predictor is formed by exploiting frozen flow hypothesis and making use of minimum mean square error estimator (with analytical expressions for the stochastic process) ([Jackson et al. 2015](#)).

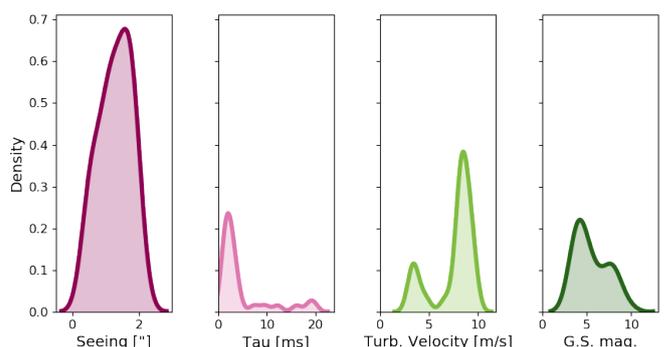
Within the HCI community, there have been many efforts to incorporate prediction into the AO control algorithm. Building on [Poyneer & Macintosh \(2006\)](#), predictive Fourier control, proposed in [Poyneer et al. \(2007\)](#), makes use of Fourier decomposition and the closed-loop power spectral density (PSD) to find components due to frozen flow. Using Kalman filtering a predictive control law is determined resulting in a reduction of the servo-lag error. Other methods in HCI have focused on splitting the prediction step from the controller. This is done by first estimating the pseudo-open loop phase (slopes, or modes), applying a prediction filter, and then controlling the HODM using the predicted phases as input into the controller. This approach allows for a system architecture that can turn prediction on and off without affecting the control loop. A similar structure has been implemented using the CACAO real-time computer ([Guyon et al. 2018](#)). Empirical orthogonal functions (EOF) is a data-driven predictor that aims to minimise the phase variance. Implemented on SCEXAO (an HCI instrument at the Subaru telescope using the CACAO real-time computer) it has been demonstrated ([Guyon et al. 2018](#)) that EOF improves the standard deviation of the point-spread-function over a set of images. However, the improvement is less than expected from initial simulations ([Guyon & Males 2017](#)). Similar methods minimise the same cost function as EOF but with a different evaluation of the necessary covariance functions (see Sec. 3.1) as reported in [van Kooten et al. \(2019\)](#) and [Jensen-Clem et al. \(2019\)](#). Current a posteriori tests, using AO telemetry ([Jensen-Clem et al. 2019](#)), show an average factor of 2.6 improvement for contrast for separations from 0 to  $10 \lambda/D$ . This approach to prediction will be implemented at the Keck telescope. One benefit to separating the prediction and the control steps is that the behaviour of the input disturbance (atmospheric induced phase fluctuations) can be studied for a given system and telescope site location, along with tests performed with AO telemetry data. We take this approach in this work, building on our earlier work ([van Kooten](#)

[et al. 2019](#)), looking solely at the predictability of the pseudo open-loop slopes under various atmospheric conditions. Previous work on prediction in the AO community has focused on one or two case(s) on-sky for demonstrations of prediction, successfully showing the feasibility of predictive control. We look at how a linear minimum mean square error (LMMSE) predictor performs a posteriori on VLT/SPHERE AO telemetry data under a large set of various observing conditions (such as guide star magnitude, coherence time, and seeing conditions).

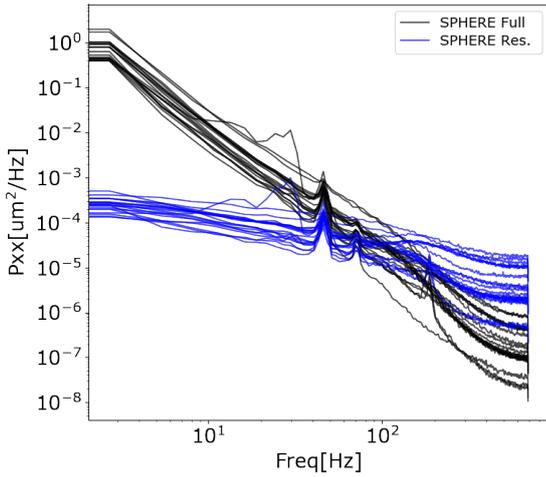
We organise this paper as follows: we introduce and summarise our SAXO data set in Sec. 2. In Sec 3 we outline our methodology, including the structure of our predictor (Sec. 3.1). We elaborate on how we apply the predictor to the VLT/SPHERE SAXO telemetry data and we present our results in Sec. 4, discussing them in Sec. 5. We look at how the predictor performs under different conditions as well as the stationarity of the turbulence. The implications of the results are discussed in Sec. 5.4, concluding in Sec. 6 including future research directions.

## 2. SAXO data

The SAXO system has the option to save the full (or partial) XAO telemetry, including HODM positions, SHWFS slopes, SHWFS intensities, interaction matrix, at the discretion of the instrument user. In this paper, we make use of 27 SAXO data sets taken between 2017 and 2019. By limiting ourselves to these years we also have estimations of the atmospheric conditions (seeing, coherence time, and turbulence velocity) from the MASS-DIMM instrument located approximately 100 meters away from UT4, which houses the VLT/SPHERE instrument. The conditions under which our data were acquired are summarised in Fig. 1, where we plot the kernel density functions estimated from the data. We note that the data set is biased toward shorter coherence times ( $\tau$ ), where the median coherence time are 2.5 milliseconds; from [Milli et al. \(2017\)](#) and [Cantalloube, F. et al. \(2018\)](#) we expect the WDH to be dominant (and thereby the servo-lag error) when the coherence time drops below 5 milliseconds. For completeness, the data set has a couple of data sets with longer coherence times. The turbulence velocity is the



**Fig. 1.** Kernel density functions for the seeing, coherence time, turbulence velocity, and guide star magnitude (r band) showing the conditions under which the VLT/SPHERE telemetry was taken. The first three panels are measurements closest to the time of the observation output by the Adaptive Optics Facility (AOF; accessed via ESO Paranal query form). The corresponding targets were found at the VLT/SPHERE ESO archive and their r-band magnitudes were found in the VizieR catalog. The kernel density functions are a nonparametric estimation of the probability functions.



**Fig. 2.** Power spectral densities, estimated using the Welch method, for all the data sets; both for the full VLT/SPHERE estimated pseudo open-loop phases and for the reconstructed VLT/SPHERE residual phases.

velocity of the characteristic turbulent layer as determined by the MASS-DIMM instrument and is associated with a characteristic altitude determine from the atmospheric profile. Therefore, it does not necessarily indicate the speed of the jet stream layer but provides a tracer atmosphere velocity. For most of the data we have bright guide star magnitudes with a mean of 5 magnitudes in the r band (the wavefront sensor bandwidth), resulting in high signal-to-noise ratios (S/N) for the SHWFS in all cases. The seeing has a Gaussian-like distribution with a mean of 1.3 arcseconds. A full summary of the entire data set (including time of observation) is provided in Tab. A.1 in the Appendix. Each of the 27 data sets differ in length and range from 10 s to 60 s, allowing us to probe different conditions while still having the opportunity to observe the behaviour of turbulence on timescales of a minute.

To study the influence of prediction on our data sets, we estimated the pseudo-open loop phases, thereby applying prediction to the zonal two-dimensional grid of phase values and not the modes. We performed the open-loop estimation using the HODM commands only because SAXO saves the full HODM voltages, not just the updates (see Appendix B). We converted to phase using the standard SAXO HODM modes. As a result, we neglected the spatial frequencies beyond the spatial bandwidth of the HODM and therefore underestimated the open-loop phase at higher spatial frequencies. We took this approach after first considering the more traditional method of using the wavefront sensor measurements and unravelling the pseudo open-loop phase from the SHWFS measurements and knowledge of the controller state. Without full knowledge of the modal gains and controller at each time step, as in our case, this method provides an inaccurate estimation.

In Fig. 2, we plot the resulting PSDs for the final pseudo open-loop phase from the VLT/SPHERE telemetry (SPHERE full) and the closed-loop phase of VLT/SPHERE residuals and identify some key features. There are peaks around 40 hertz and 60 hertz in both the open and closed-loop PSDs. Performing a modal analysis we find that the peaks appear for Zernike modes 7-11 (Noll indexing; Noll 1976) with varying amplitudes. The second feature is the increase in power at high temporal frequencies for the closed-loop PSD (i.e., the so-called waterbed effect, a result of Bode’s sensitivity integral).

### 3. Methodology

Although we are ultimately interested in the improvement in contrast using prediction, we limit ourselves in this work to studying the minimisation of the servo-lag error. From Guyon (2005), Kasper (2012), and Cantalloube, F. et al. (2018), we see that minimising the lag results in an improvement in contrast but ultimately also depends on the coronagraph of choice and how it interacts with the residual phase at small angular separations. We also do not have focal plane images taken at the same time as the data sets, making a clear claim of improvement unreliable. The metric we adopt is the temporally and spatially averaged wavefront phase variance.

#### 3.1. LMMSE prediction

For our predictor we chose a data-driven method called the LMMSE predictor. This approach provides a flexible framework that allows us to implement the predictor in three different ways: batch, recursive, and with an exponential forgetting factor (van Kooten et al. 2019).

A single point  $i$  of a phase screen at time  $t$  is given by  $y_i(t)$ , while  $\mathbf{u}(t)$  is a  $P^2 \times 1$  column vector containing a collection of  $P^2$  phase values on a discrete spatial grid at time  $t$ . We assume that the future value of a given phase point,  $\hat{y}_i$  at the discrete time index  $t + d$ , is a linear combination of the most recent phase values at time  $t$ . The predictor coefficients are denoted as  $\mathbf{a}_i$ . The cost function of our predictor, with  $\langle \cdot \rangle_t$  as the time average operator, is then

$$\min_{\mathbf{a}_i} \langle \|y_i(t+d) - \mathbf{a}_i^T \mathbf{w}(t)\|^2 \rangle_t, \quad (1)$$

where  $\mathbf{w}(t)$  includes a set of  $Q$  most recent measurements,

$$\mathbf{w}(t) = \left( \mathbf{u}(t)^T \quad \mathbf{u}(t-1)^T \quad \mathbf{u}(t-2)^T \quad \dots \quad \mathbf{u}(t-Q)^T \right)^T. \quad (2)$$

We allowed for both spatial and temporal regressors gathered into  $\mathbf{w}(t)$  – a  $P^2 Q \times 1$  vector. We denoted predictors of various orders by indicating the spatial order  $P$  (spatially limiting ourselves to a box of order  $P$ , symmetric around the phase point of interest, resulting in  $P^2$  spatial regressors) followed by the temporal order  $Q$ ; for example, a ‘s5t2’ predictor has  $P = 5$  and  $Q = 2$ .

Solving Eq. 1 for our zero-mean stochastic process, the solution can be written in terms of the inverse of the auto-covariance matrix and cross-covariance vector (Haykin 2002)

$$\mathbf{a}_i = \mathbf{C}_{\mathbf{w}\mathbf{w}}^+ \mathbf{c}_{\mathbf{w}y_i}, \quad (3)$$

where  $+$  denotes a pseudo inverse;  $\mathbf{C}_{\mathbf{w}\mathbf{w}}$  is the auto-covariance matrix of  $\mathbf{w}$ , the vector containing the regressors; and  $\mathbf{c}_{\mathbf{w}y_i}$  is the vector containing the cross-covariance between the true phase value,  $y_i$  and  $\mathbf{w}$ .

We can estimate the covariances in Eq. 3 directly from a training set, forming a fixed batch solution. Alternatively, we can form a recursive solution making use of the Sherman-Morrison formula (a special case of the Woodbury matrix inversion lemma). In Eq. 4 through to Eq. 6 we inserted the exponential forgetting factor in the update in the following equations, thereby forming our final LMMSE implementation. The recursive form is given in Eqs. 4 and 5 with  $\lambda = 1$ , such that all previous data is weighted equally, as follows:

$$\mathbf{c}_{\mathbf{w}y_i}(t-d) = \lambda \mathbf{c}_{\mathbf{w}y_i}(t-d-1) + \mathbf{w}(t-d)y_i(t-d) \quad (4)$$

$$\mathbf{C}_{\mathbf{w}\mathbf{w}}^+(t-d) = \lambda^{-1}\mathbf{C}_{\mathbf{w}\mathbf{w}}^+(t-d-1) - \mathbf{k}(t-d) \quad (5)$$

with

$$\mathbf{k}(t-d) = \frac{\lambda^{-2}\mathbf{C}_{\mathbf{w}\mathbf{w}}^+(t-d-1)\mathbf{w}(t-d)\mathbf{w}^T(t-d)\mathbf{C}_{\mathbf{w}\mathbf{w}}^+(t-d-1)}{1 + \lambda^{-1}\mathbf{w}^T(t-d)\mathbf{C}_{\mathbf{w}\mathbf{w}}^+(t-d-1)\mathbf{w}(t-d)}. \quad (6)$$

By updating Eqs. 4 and 5, the coefficients can be found for each time step using Eq. 3. The recursive solution goes on-line immediately with the initial auto-covariance being set to diagonal matrices with large values (as done with recursive least-squares methods) and the cross-covariance vector set to ones. By adjusting the forgetting factor, we can weigh old data by less compared to the most recent measurements, therefore allowing the tracking of slowly varying signals.

### 3.2. Comparison with EOF

Methods such as LMMSE and EOF (see Guyon & Males (2017)) and similar techniques (see Jensen-Clem et al. (2019)) all minimise the same cost function, but the evaluation of Eq. 3 is different in each case; we note that the cost function is slightly different when including exponential forgetting factor. In EOF, the solution is estimated with the inverse of the auto-covariance determined using a singular-value-decomposition that is re-estimated on minute timescales. The amount of data used to estimate the prediction filter, the numerical robustness, the noise properties of the system, and the atmospheric turbulence above the telescope all contribute to the performance of these algorithms and the final computational load. Therefore one implementation might be more suited for specific conditions than the other, but the three methods can, in ideal conditions, result in the same performance.

### 3.3. Applying prediction to SAXO telemetry

From the estimated open-loop phases (which results in a 240-by-240 phase screen using the HODM modes) we bin the data to 60-by-60 phase screens for computational memory purposes. We then performed prediction on the estimated open-loop phases assuming a 2-frame delay. For each phase point, we estimated a unique set of prediction coefficients using the equations as outlined in Sec. 3.1. Our aim is to focus on the prediction capabilities, ignoring the control aspect by assuming a perfect system – no wavefront sensor noise and a HODM that can perfectly correct all spatial frequencies predicted – and only including the delay. We note that this results in no fitting error, no spatial bandwidth limitations, and no temporal bandwidth limitations on the achievable performance.

We ran batch, recursive, and forgetting (with  $\lambda = 0.998$  as this value gives the best performance assuming  $\lambda \neq 1$ ) LMMSE predictors for each prediction order. We then subtracted the predicted phase from the pseudo open-loop phase 2 frames later resulting in the residual predictor phase. From the phases we calculated the averaged phase variance by taking the final 5 s of the data set, and therefore, all the different predictors have converged; we calculated the spatio-temporally averaged phase variance. We started with the performance of a s1t1 predictor. The s1t1 is a zero-order predictor because it only makes use of the most recent measurement for that given phase point, making it analogous to an optimised integrator (with a gain close

to unity) in our simulations; we refer to the s1t1 as the ideal VLT/SPHERE performance.

We performed simulations testing a variety of predictors with different spatial and temporal orders including s1t3, s1t10, s3t1, and s3t3. We looked at how the three different implementations of the LMMSE for each prediction order behave (see Fig 3). In Sec. 4 we present the results for the recursive s1t10 predictor, which performs the best out of all the different orders.

## 4. Results

We find that prediction provides a reduction in averaged phase variance when compared to the VLT/SPHERE SAXO residuals. An example of how prediction behaves spatially and temporally for a slice across the telescope aperture is shown in the bottom panel of Fig. 4. We see a reduction in the phase compared to the VLT/SPHERE residuals and see a more uniform solution in time and space. In Fig. 5 we summarise the results of running prediction on all of our data sets. As mentioned above, the averaged phase variance is found by using the last 5 s of the data. The data sets all vary in length. When we see a reduction in the residual phase variance, we would expect an improvement of performance for the XAO system for all cases independent of the guide star magnitude, the seeing, and the coherence time. We observe a reduction in the spread of residual phase variance, with prediction providing a more uniform performance for various conditions; see the kernel density function plots in the top panel of Fig. 5.

In Fig. 6 we plot the ratio of the recursive s1t10 predictor residual phase variance to the VLT/SPHERE residual phase variance defining this as the “ratio of improvement” as a function of coherence time. In the same figure, we add the ratio of improvement for the same recursive s1t10 predictor and an idealised VLT/SPHERE integrator (batch s1t1) against coherence time. We calculate the average ratio of improvement to be 5.1 and 2.0, respectively. Fig. 6 shows the relative seeing conditions indicated by the size of the marker; smaller marker sizes indicate better seeing conditions.

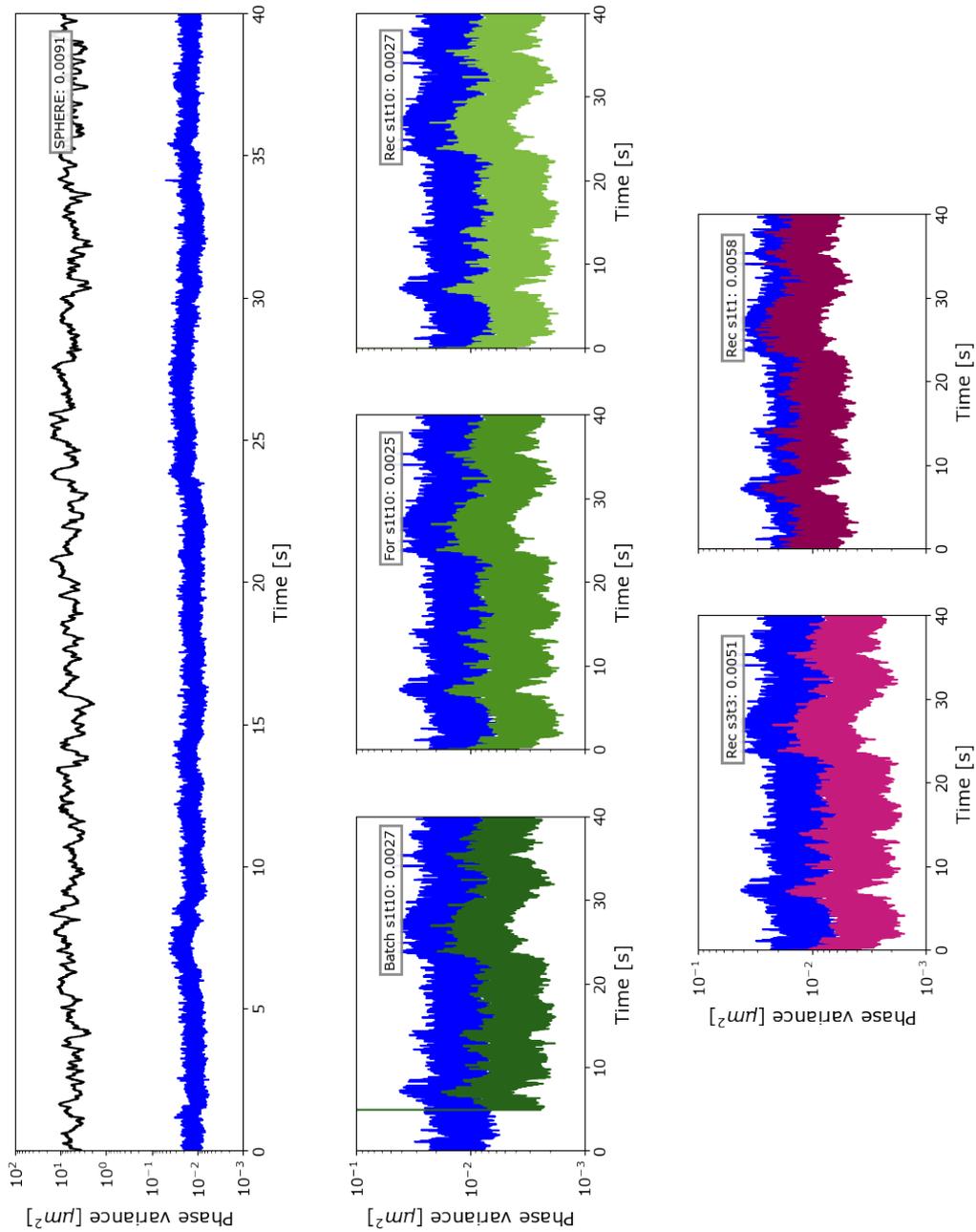
We evaluate several predictors, varying both spatial and temporal orders. We find that there is no gain in performance by adding spatial regressors and find that temporal regressors perform equally as well. These results are summarised in Tab. 1.

We see minimal evidence of nonstationary behaviour of optical turbulence. First, by looking at the coherence times in Tab. A.1, we do not see significant change in coherence time for data taken on the same night – showing that on 100 s time scales the statistics of optical turbulence do not vary significantly. Second, on shorter timescales, we do not see evidence of nonstationary turbulence. In Fig. 3, we plot the batch, recursive, and the forgetting LMMSE for a s1t10 predictor. From the average residual phase variances, we see that the batch and recursive perform the same over the full 40 s period. We do see a slight improvement for the forgetting LMMSE, implying a slight time-variant behaviour of the pseudo open-loop phase but nothing significant.

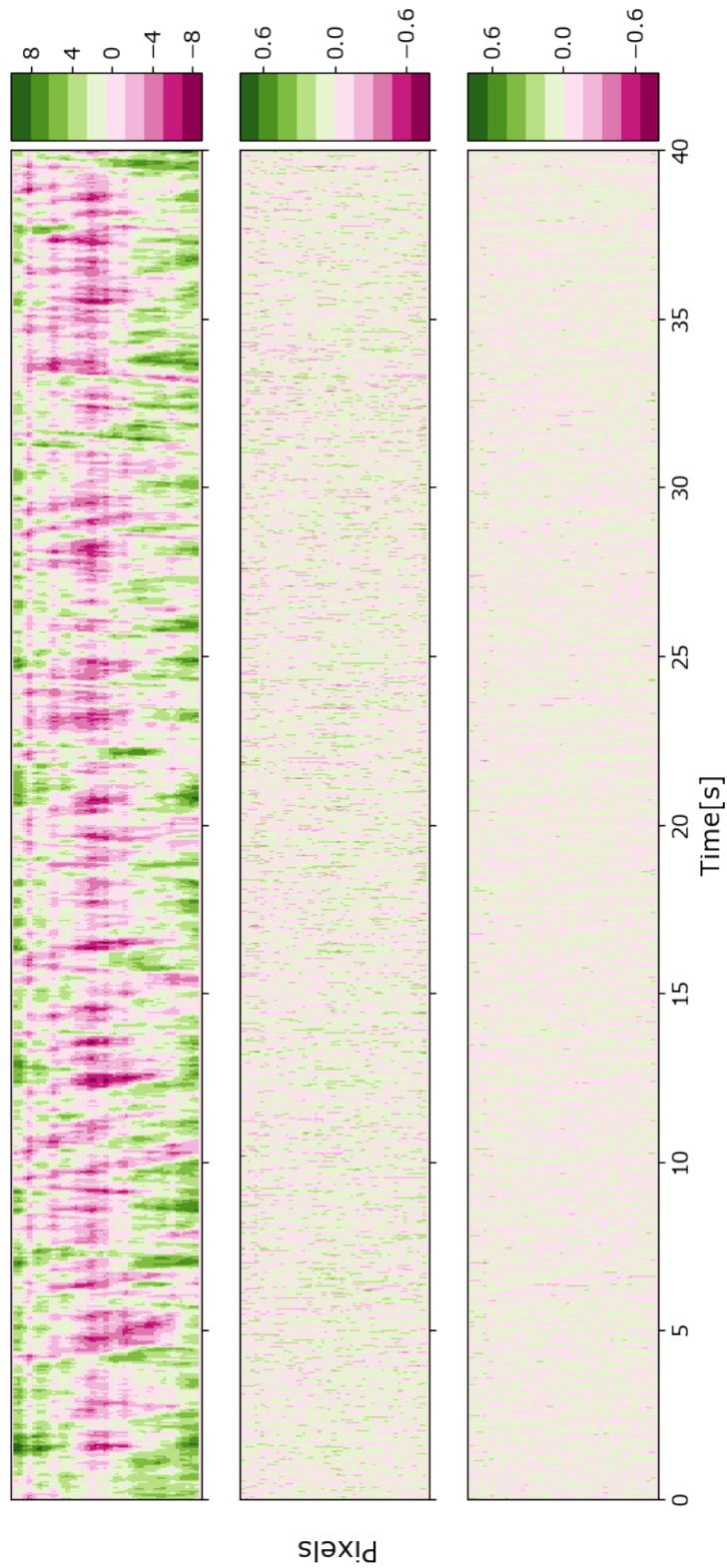
## 5. Discussion

### 5.1. Comparison of the prediction residuals to the VLT/SPHERE residuals

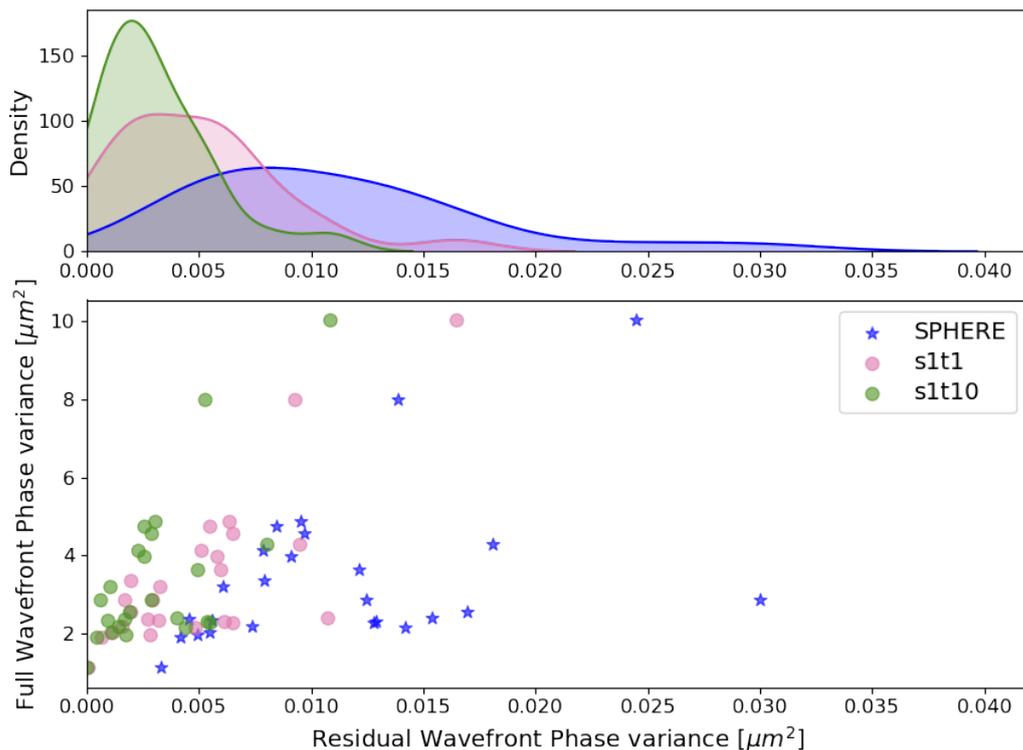
We should note the difficulties in performing a direct comparison between the predictor residuals to the real VLT/SPHERE residuals. There are a few challenges, with the first being a difference in delay. In our estimation of the open-loop phase we choose to



**Fig. 3.** In the top panel, we plot a time series showing the estimated pseudo open-loop in black, compared with the VLT/SPHERE residual in blue for a random SAXO telemetry data set. The other plots (all with the same y-scale) show various predictor residual phase variances (various green and purple lines) compared to the VLT/SPHERE residuals for the same data. The averaged phase variance, in  $\mu m^2$ , for the last 5 s is indicated in the top right corner of each plot. The forgetting s1t10 (abbrev. for) does not perform significantly better than the recursive s1t10 (abbrev. res). Secondly, the s3t3 does not perform better than the s1t10.



**Fig. 4.** Vertical slice across the telescope aperture (y-axis) showing the wavefront phase in  $\mu\text{m}$  (indicated by the colour-bars), plotted as a function of time for the pseudo open-loop phase (top), the VLT/SPHERE residual phase (middle), and slit10 predictor residual phase (bottom) for the same night as Fig. 3. Comparing the bottom two panels (colour map is the same in both), we can see the prediction residuals have a flatter and more uniform appearance compared to the real VLT/SPHERE residuals.



**Fig. 5.** Top: Kernel density function estimation for the averaged phase variances plotted in the bottom plot. The change in spread of the values is shown. Bottom: Pseudo open-loop averaged wavefront phase variance compared to residual averaged wavefront phase variance for VLT/SPHERE, a batch s1t1 (i.e., idealised integrator for VLT/SPHERE), and a recursive s1t10 predictor. The points move from right to left, indicating that a s1t1 does better than VLT/SPHERE and a high order predictor does even better than the s1t1.

Time of data	Pseudo open-loop [ $\mu\text{m}^2$ ]	VLT/SPHERE [ $\mu\text{m}^2$ ]	s1t3 [ $\mu\text{m}^2$ ]	s3t3 [ $\mu\text{m}^2$ ]	s3t1 [ $\mu\text{m}^2$ ]
2017-07-19T22:54:55.000	1.963	0.005	0.002	0.002	0.002
2017-07-19T23:16:39.000	2.856	0.012	0.001	0.001	0.002
2017-07-19T23:22:02.000	2.400	0.015	0.004	0.004	0.011
2017-07-19T23:50:33.000	2.148	0.014	0.004	0.004	0.005

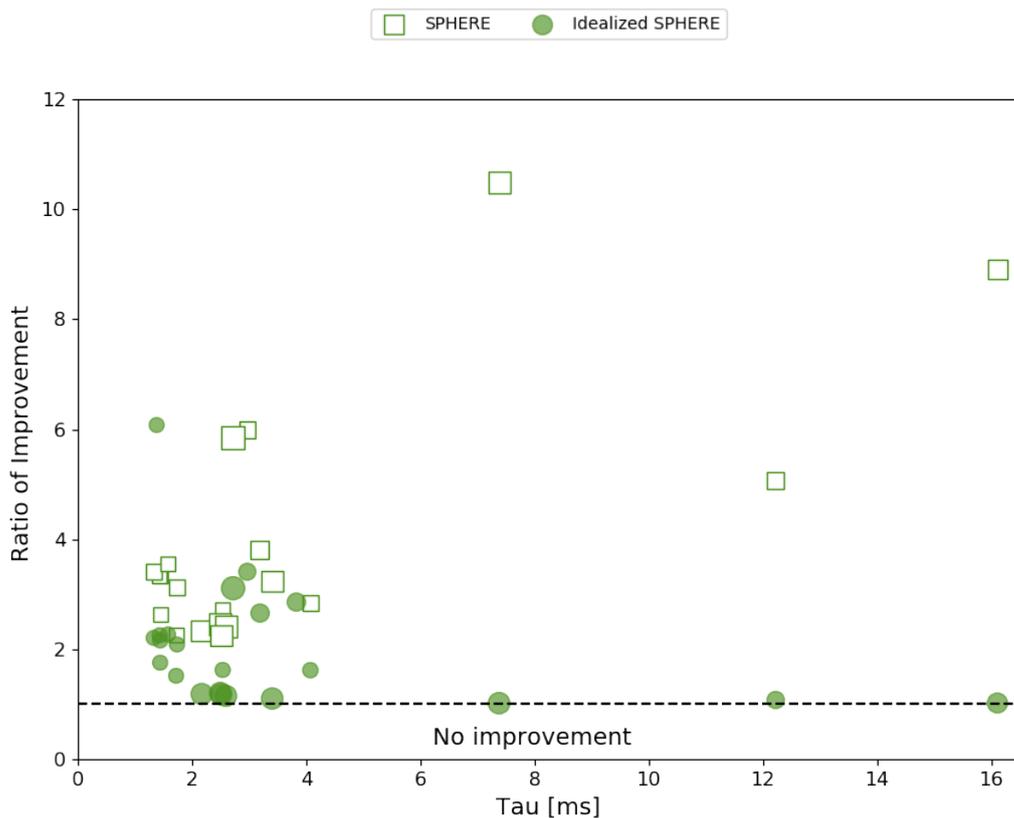
**Table 1.** Averaged phase variance for the pseudo open-loop, VLT/SPHERE residuals, s1t3 residuals, s3t3 residuals, and the s3t1 residuals. Comparing the last three columns, there is no gain by including spatial regressors to the prediction algorithm and the s1t3 does better than the s3t3.

round the frame delay to a whole frame; therefore our predictor sees a delay of 2 frames (or 1.45 ms) while the real system delay, and thereby encoded in the real VLT/SPHERE residuals, is 2.2 frames (1.59 ms). We therefore assume that the VLT/SPHERE residuals immediately have a larger phase variance compared to the case where the true delay is 2 frames. An alternative option to rounding the delay frames is to interpolate, which is a step that would have also introduced an error. We make use of the HODM commands which are saved as the total voltage on the HODM, not the update to the HODM. The SHWFS, however, is used to determine the real VLT/SPHERE residuals and therefore see higher order spatial frequencies, potentially increasing the residuals if this was not the case. Perhaps the most substantial contribution to the final performance of VLT/SPHERE results from the fact that the VLT/SPHERE performance will be limited by operational parameters. The controller will need to be stable and robust on-sky, potentially resulting in a loss of performance when compared to our idealised situation. Therefore, in Figs. 5

and 6 we plot the idealised VLT/SPHERE (batch s1t1) as well as the VLT/SPHERE residuals. The true gain in prediction will be between the idealised VLT/SPHERE and real VLT/SPHERE residuals and the ratio of improvement will fall between 5.1 and 2.0.

## 5.2. Performance under different conditions

Our data set and analysis is unique, showing that with prediction, there will be an improvement under almost all conditions even for long coherence times, and no loss in performance with prediction is observed. However, the data set does not show any clear correlations between predictor performance and observational conditions (see Figs. 6 and 7). We briefly discuss the behaviour for various observing parameters including coherence times, turbulence velocity, seeing, and finally guide star magnitudes.



**Fig. 6.** Ratio of improvement, found by taking the ratio of an idealised integrator on VLT/SPHERE to a recursive spatial-temporal predictor (slt10) phase variance as calculated from the last 5 s of data, as a function of coherence time. The size of the markers indicates the AOF seeing conditions at the time of observation. The average ratio of improvement is 5.1 when comparing the prediction to the real VLT/SPHERE residuals. When looking at the idealised VLT/SPHERE, we find an average ratio of improvement of 2.0 in wavefront variance reduction.

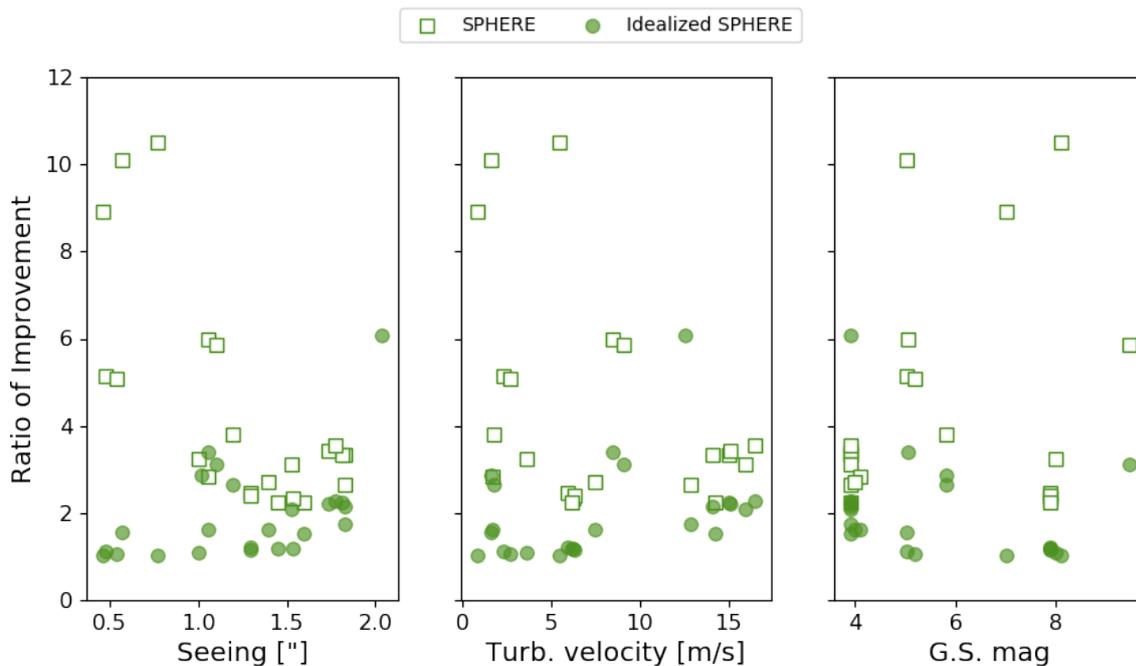
Looking closely at Fig 6 and the behaviour for various coherence times we do not find any correlation between the coherence time and performance for the true VLT/SPHERE residuals. However, when looking at the idealised VLT/SPHERE behaviour we see, at smaller coherence times, an exponential-like gain in the ratio of improvement in which the asymptote is the Nyquist sampling frequency ( $2/WFS_f = 2/1380 = 1.45ms$ ). This behaviour for the idealised case is as expected, where we have a larger improvements for shorter coherence times. We then look at the relation between the ratio of improvement and turbulence velocity (middle plot of Fig. 7). We expect a similar behaviour to that of the coherence time as they are related. We note that we also see no correlation for the ground layer wind speeds measured by the nearby meteorological tower. Studying the relation between the seeing and the ratio of improvement we notice an asymptotic behaviour where the ratio improves for better seeing conditions. We do not see any dependence between performance and S/N (or guide star magnitudes; left plot of Fig. 7). For the VLT/SPHERE residuals, the lack of correlation between performance and S/N is as expected from laboratory and on-sky validation of SAXO by Fusco et al. (2016) at these guide star magnitudes. We also note that for long coherence times VLT/SPHERE is often looking at fainter targets since the conditions are ideal. This is the case for our data as well (see Tab. A.1). In summary, we see a relation between the true VLT/SPHERE residuals and the seeing, but for the idealised VLT/SPHERE case we see a relation between the improvement and the coherence time, as expected. The lack of correlation between ratio of improvement and the other observing parameters

could be a result of the different locations of VLT/SPHERE and DIMM-MASS (which measures the observing parameters) and that the values determined over minute averages do not reflect the exact values at the time of observation. Alternatively, the behaviour of the SAXO controller is limited by internal system requirements (such as vibration rejection) and not the observing conditions, resulting in a lack of correlation between observing conditions and gain in performance.

Studying Fig. 6, we can see that the ratio of improvement when using idealised VLT/SPHERE as a benchmark behaves very differently than the true VLT/SPHERE case. For a direct comparison to VLT/SPHERE, we do not see any correlations between the ratio of improvement and the coherence time. However, for the idealised case, we see that at longer coherence times we have no improvement since the SAXO XAO system can already perform well.

### 5.3. Time-invariant turbulence statistics

Observing the behaviour of the three (batch, recursive, and exponential forgetting) implementations of the LMMSE, we can comment on the stationarity of the turbulence. The LMMSE finds the optimal prediction coefficients determined from the training set. The batch is trained on the first 5 s and the recursive trains continuously. We note the residual phase variances for the last 5 s for each case; we do not see a significant difference between the batch and recursive solutions in any case, indicating the statistics of the turbulence has not changed over the mea-



**Fig. 7.** Ratio of improvement compared to seeing (left), turbulence velocity (middle), and guide star magnitude in r band (right) during the time of observation.

surement period. Studying the recursive solution, we look at the behaviour of the prediction coefficients in time across the aperture. We do not see any notable changes over the entire period once the solution has converged. We do see more fluctuations in the prediction coefficients for phase points located on the edges with a predictor using spatial information such as the s3t3. Conversely, the exponential forgetting LMMSE does show a slight improvement, however, this could be due to noise in the system that the LMMSE predictor can and does remove.

#### 5.4. Implications of results

For all conditions we see an increase of performance with prediction when compared to the VLT/SPHERE residuals, under an idealised assumption there are a few cases where the ratio of improvement is one, indicating no gain but also, notably, no loss in performance.

A more notable result is from the kernel density functions plotted in the top panel of Fig. 5. The spread of the averaged phase variance is less for the s1t10 predictor, indicating a more uniform performance in phase variance reduction for different observing conditions. From an observational point of view, having a more stable correction under different conditions is desirable, especially for the cases on surveys in which observers might be targeting similar objects and can perform reference star differential imaging from a library.

The ratio of improvement we find is 5.1, but this is probably an overestimate of the improvement we could achieve on-sky. In previous prediction work, Guyon & Males (2017) show an improvement of 7 in root-mean-square (rms) residual wavefront error while offline telemetry tests by Jensen-Clem et al. (2019) show a factor of 2.5 rms wavefront error; we note that the errors in both of these works refer to systems located on top of Mauna Kea. Although an exact direct comparison using these values is impossible, we note that we show a more modest predictive improvement compared to these studies.

When studying the prediction order we do not see a large gain from including spatial information. This is due to large sub-aperture size and the high rate of temporal sampling, which results in the turbulence only moving across a sub-aperture after many frames; for example, assuming 10 m/s and 0.2 m sub-aperture size, it takes approximately 28 frames before a cell of turbulence moves to the next sub-aperture. The turbulence is still dynamic but we sense the average variations that fall within the wavefront sensor sub-aperture. We have much finer temporal sampling. We expect a temporal-only predictor to be best suited for HCI, and by removing the spatial information, we are no longer sensitive to wind direction (spatial solution requires a symmetric choice of regressors) which therefore reduces the computational size of the prediction problem.

We do not see evidence of time-invariant turbulence, meaning that our predictor does not need to be able to track changes in turbulence behaviour on timescales less than 1-2 minutes of observations. We see a slight increase in performance from the exponential forgetting factor LMMSE solution but the difference is very slight and not substantial enough to suggest this as the best choice. From a computational point of view, the batch LMMSE is the best option and resetting it every 1 to 2 minutes (or as needed based on longer telemetry data) using 5 s of training data would be the best implementation.

## 6. Conclusions and future work

We find a reduction in the phase variance in comparison to the VLT/SPHERE residuals and determine the ratio of improvement to be 5.1 for SAXO telemetry data. When prediction is compared to an idealised VLT/SPHERE system, we find an improvement ratio of 2.0. In all cases, no matter what the observing conditions, prediction performs well with no loss in performance. Most importantly, we note that under all the 27 various observing conditions studied, we see a reliable and overall more consistent improvement of the system performance. The data set, in combination with our predictors, reveals that the optical turbulence

as seen by the telescope is time-invariant and that the temporal regressors have a larger impact on the performance of the predictor than spatial regressors. We recommend a batch (updating every few minutes as necessary) temporal-only predictor for the VLT/SPHERE to reduce the servo-lag error. In future work we will seek to investigate the effects of prediction on contrast for different types of coronagraphs and to determine whether predictive control can be used to directly optimise the raw contrast.

*Acknowledgements.* The authors would like to thank Markus Kasper and Julien Milli for providing us with the VLT/SPHERE SAXO data. The authors would also like to thank Leiden University, NOVA, METIS consortium, and TNO for funding this research.

## References

- Avenhaus, H., Quanz, S. P., Garufi, A., et al. 2018, *ApJ*, 863, 44
- Beuzit, J.-L., Vigan, A., Mouillet, D., et al. 2019, *Astronomy & Astrophysics*
- Cantalloube, F., Por, E. H., Dohlen, K., et al. 2018, *Astronomy & Astrophysics*, 620, L10
- Chauvin, G., Desidera, S., Lagrange, A.-M., et al. 2017, *A&A*, 605, L9
- Doelman, N., Fraanje, R., & den Breeje, R. 2011, 2nd Conference on Adaptive Optics for Extremely Large Telescopes, 1
- Fusco, T., Sauvage, J.-F., Mouillet, D., et al. 2016, in *Adaptive Optics Systems V*, ed. E. Marchetti, L. M. Close, & J.-P. Véran, Vol. 9909, International Society for Optics and Photonics (SPIE), 273 – 290
- Guyon, O. 2005, *The Astrophysical Journal*, 629, 592–614
- Guyon, O. & Males, J. 2017, ArXiv e-prints [arXiv:1707.00570]
- Guyon, O., Sevin, A., Gratadour, D., et al. 2018, in *Adaptive Optics Systems VI*, ed. L. M. Close, L. Schreiber, & D. Schmidt, Vol. 10703, International Society for Optics and Photonics (SPIE), 469 – 480
- Haykin, S. 2002, *Adaptive filter theory*, 4th edn. (Upper Saddle River, NJ: Prentice Hall)
- Jackson, K., Correia, C., Lardière, O., Andersen, D., & Bradley, C. 2015, *Opt. Lett.*, 40, 143
- Jensen-Clem, R., Bond, C. Z., Cetre, S., et al. 2019, in *Techniques and Instrumentation for Detection of Exoplanets IX*, ed. S. B. Shaklan, Vol. 11117, International Society for Optics and Photonics (SPIE), 275 – 284
- Kasper, M. 2012, *Proc.SPIE*, 8447, 8447
- Keppler, M., Benisty, M., Müller, A., et al. 2018, *A&A*, 617, A44
- Le Roux, B., Ragazzoni, R., Arcidiacono, C., et al. 2004, *Proc.SPIE*, 5490, 1336
- Milli, J., Mouillet, D., Fusco, T., et al. 2017, *AO4ELT5*
- Noll, R. J. 1976, *J. Opt. Soc. Am.*, 66, 207
- Petit, C., Conan, J.-M., Kulcsár, C., Raynaud, H.-F., & Fusco, T. 2008, *Optics Express*, 16, 87
- Petit, C., Sauvage, J.-F., Fusco, T., et al. 2014, in *Adaptive Optics Systems IV*, ed. E. Marchetti, L. M. Close, & J.-P. Véran, Vol. 9148, International Society for Optics and Photonics (SPIE), 214 – 230
- Poyneer, L. A. & Macintosh, B. A. 2006, *Opt. Express*, 14, 7499
- Poyneer, L. A., Macintosh, B. A., & Véran, J.-P. 2007, *J. Opt. Soc. Am. A*, 24, 2645
- Sissa, E., Olofsson, J., Vigan, A., et al. 2018, *A&A*, 613, L6
- Sivo, G., Kulcsár, C., Conan, J.-M., et al. 2014, *Opt. Express*, 22, 23565
- van Kooten, M., Doelman, N., & Kenworthy, M. 2019, *J. Opt. Soc. Am. A*, 36, 731
- Zurlo, A., Vigan, A., Galicher, R., et al. 2016, *A&A*, 587, A57

## Appendix A: Overview of SAXO data

In this Appendix we provide a full summary of our data set. The XAO telemetry data is stored on the SAXO server and a log of when AO data was taken can be found on the ESO science archive under the VLT/SPHERE instrument. The 2019 data was kindly provided to us by Markus Kasper while the other data sets were accessed by Julien Milli.

Date	Seeing[""]	Tau[ms]	Turbulence Velocity [m/s]	Guide star Magnitude r-band
2017-07-17T23:15:30.000	1.708	1.539	8.340	3.910
2017-07-17T19:07:58.000	2.037	1.381	-	3.910
2017-07-17T23:00:24.000	1.829	1.442	8.450	3.910
2017-07-17T23:02:51.000	1.827	1.443	8.450	3.910
2017-07-17T23:04:41.000	1.825	1.444	8.450	3.910
2017-07-17T23:10:31.000	1.812	1.434	8.450	3.910
2017-07-17T23:11:54.000	1.739	1.330	9.210	3.910
2017-07-17T23:29:56.000	1.776	1.579	9.280	3.910
2017-07-17T23:31:11.000	1.599	1.722	8.300	3.910
2017-07-17T23:33:29.000	1.524	1.738	7.970	3.910
2017-07-19T22:54:55.000	1.053	4.075	7.990	4.110
2017-07-19T23:16:39.000	1.014	3.829	9.120	5.819
2017-07-19T23:22:02.000	1.197	3.193	8.760	5.819
2017-07-19T23:50:33.000	1.001	3.406	9.700	8.010
2018-04-04T03:06:15.000	1.059	2.970	0.000	5.060
2018-04-04T03:12:50.000	1.099	2.721	0.000	9.480
2016-05-21T08:23:42.000	1.396	2.538	6.65	4.00
2016-05-21T09:29:07.000	1.533	2.171	8.16	7.890
2016-05-21T09:53:34.000	1.298	2.495	9.31	7.890
2016-05-21T09:55:48.000	1.292	2.596	9.03	7.890
2016-05-21T09:58:02.000	1.452	2.510	8.31	7.890
2019-01-26T01:49:48.00	0.770	7.381	3.39	8.110
2019-01-26T02:08:25.00	0.710	9.596	3.39	6.560
2019-01-26T03:34:39.00	0.460	16.113	4.65	7.017
2019-01-26T03:48:01.00	0.570	19.643	3.1	5.020
2019-01-26T04:59:53.00	0.480	18.774	3.67	5.020
2019-01-26T06:39:57.00	0.540	12.228	7.81	5.180

**Table A.1.** Summary of the VLT/SPHERE data used in this work as well as the AOF atmospheric conditions as recorded closest to the measurement time. The target r-band magnitude is provided as to provide a relative estimation of the signal-to-noise ratio for the wavefront sensor. The length of the telemetry data ranges from 10 to 60 seconds.

## Appendix B: Estimation of open-loop phase

In this Appendix, we explain how the SAXO VLT/SPHERE telemetry data is used to estimate the pseudo open-loop phase, providing an estimation of the open loop phase of the wavefront at the pupil plane due to atmospheric turbulence. Usually the pseudo-open loop phases are estimated using wavefront sensor data and controller state (deformable mirror updates, gain, and interaction matrix). Specifically, by summing the measured wavefront sensor phase and the previous deformable mirror updates, an estimation of atmospheric phase can be made. We make use of an alternative method using the HODM for which we have the full voltages applied to the mirror. Spatially, the HODM has comparable sampling to the SHWFS (41-by-41 actuators versus 40-by-40 sub-apertures) and therefore using the HODM for estimating the open-loop phase does not restrict the spatial bandwidth. Similarly, the temporal bandwidth of the system is determined by SHWFS frame rate, while the temporal bandwidth of the HODM is much higher. By using the HODM we do not lose any spatial or temporal bandwidth.

From Fig. 2, we can see that the SAXO provides a good correction for our data sets, as expected, therefore the HODM surface is representative of the full atmospheric phase. We assume the residual errors are negligible owing to the high expected Strehl ratio for our conditions (80-90% in the H band, see Fusco et al. (2016)). From Fig. 2, we also see that the closed-loop PSDs estimated from the wavefront sensor residuals are relatively flat, indicating a good correction. We can then take the HODM surface as representative of the full atmospheric phase. In an open-loop system, the full atmospheric phase is measured by the wavefront sensor and then used to determine the deformable mirror commands. The surface of the deformable mirror therefore represents the estimated open-loop phase of the wavefront in the pupil plane. It can be expressed in the form of a lifted vector as

$$y(t) = DM_{surface}(t), \quad (\text{B.1})$$

where  $DM_{surface}(t)$  is a  $41^2 \times 1$  vector.

Therefore, using a posteriori data, since we know the full voltage on the DM surface we can estimate the pseudo open-loop phase.