

Power of prediction: spatiotemporal Gaussian process modeling for predictive control in slope-based wavefront sensing

Jalo Nousiainen ^{a,b,*} Juha-Pekka Puska,^a Tapio Helin ^b Nuutti Hyvönen,^a
and Markus Kasper ^c

^aAalto University, Department of Mathematics and Systems Analysis, Aalto, Finland

^bLUT University, Lappeenranta, Finland

^cEuropean Southern Observatory, Garching bei München, Germany

ABSTRACT. Time delay error is a significant error source in adaptive optics (AO) systems. It arises from the latency between sensing the wavefront and applying the correction. Predictive control algorithms reduce the time delay error, providing significant performance gains, especially for high-contrast imaging. However, the predictive controller's performance depends on factors such as the wavefront sensor (WFS) type, the measurement noise level, the AO system's geometry, and the atmospheric conditions. We study the limits of prediction under different imaging conditions through spatiotemporal Gaussian process models. The method provides a predictive reconstructor that is optimal in the least-squares sense, conditioned on the fixed times series of WFS data and our knowledge of the atmospheric conditions. We demonstrate that knowledge is power in predictive AO control. With a Shack–Hartmann sensor-based extreme AO instrument, perfect knowledge of the wind and atmospheric profile and exact frozen flow evolution lead to a reduction of the residual wavefront phase variance up to a factor of 3.5 compared with a non-predictive approach. If there is uncertainty in the profile or evolution models, the gain is more modest. Still, assuming that only effective wind speed is available (without direction) led to reductions in variance by a factor of ~ 2.3 . We also study the value of data for predictive filters by computing the experimental utility for different scenarios to answer questions such as how many past telemetry frames should the prediction filter consider and whether it is always most advantageous to use the most recent data. We show that within the scenarios considered, more data provide a consistent increase in prediction accuracy. Furthermore, we demonstrate that given a computational limitation on how many past frames, we can use an optimized selection of n past frames, which leads to a 10% to 15% additional improvement in root mean square over using the n latest consecutive frames of data.

© The Authors. Published by SPIE under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JATIS.10.3.039001](https://doi.org/10.1117/1.JATIS.10.3.039001)]

Keywords: adaptive optics; Gaussian process; Bayesian inference; inverse problems

Paper 23127G received Oct. 19, 2023; revised Jun. 11, 2024; accepted Jun. 24, 2024; published Jul. 12, 2024.

1 Introduction

On ground-based telescopes, atmospheric turbulence causes variations in the optical path length of the incoming light hampering the telescope's image quality. Adaptive optics (AO) is a technique used to compensate for these variations.^{1,2} The basic principle is to use a star or multiple

*Address all correspondence to Jalo Nousiainen, jalo.nousiainen@eso.org

guide stars (artificial or natural) as a reference point to measure these variations with a wavefront sensor (WFS) and then use a deformable mirror (DM) to compensate for the variations.

Different AO techniques have been developed for different astronomical observation scenarios. Some AO systems aim to produce a good correction on a wide field of view, while others provide good corrections in multiple directions simultaneously. This paper focuses on a single conjugate AO (SCAO) that utilizes a single guide star to obtain an excellent correction on a narrow field of view close to the guide star.

Time delay error, also known as SCAO servo-lag error, is a significant error source in AO systems. It arises from the overall latency between sensing the wavefront and applying the correction. The delay occurs for various reasons, such as the time required to process the raw data, integrate the WFS frame, and calculate the control signal, as well as the DM response time. The impact of delay can be significant, especially for AO instruments dedicated to direct exoplanet imaging, such as the Gemini Planet Imager³ on the Gemini South telescope, Spectro-Polarimetric High-contrast Exoplanet REsearch⁴ (SPHERE) instrument on the European Southern Observatory's Very Large Telescope (Taltal, Chile), MagAO-X (Magellan Adaptive Optics eXtreme system⁵), and SCExAO (Subaru Coronagraphic Extreme Adaptive Optics⁶). On these instruments, the temporal delay of AO creates a halo of stellar light appearing on the science camera, hiding the much fainter (compared with the host star) exoplanet beneath it. This phenomenon, known as the wind-driven halo (WDH), occurs particularly in windy conditions when atmospheric turbulence is in rapid motion.⁷

Time delay error can be reduced with predictive control algorithms that use past telemetry data to predict incident phase aberrations at the time when the correction is applied to the DM. It is well known that under strong assumptions on Markovian dynamics and linear system response, linear quadratic (LQ) control and Kalman filtering-based prediction yield asymptotically optimal predictive control.⁸ However, the Markovian dynamics emerging from the frozen flow (FF) hypothesis with a single- or multi-layered turbulence offer only an approximate model (near Markovian) in the discretized state space.

More recently, data-driven predictive methods that use a longer time series of so-called *pseudo-open loop telemetry* data have been proposed in the literature and show promising results in numerical simulations, optical bench, or on-sky settings (see, for example, Refs. 9–11). These approaches separate the reconstruction and prediction steps, where the prediction can be made either on WFS data or reconstructed DM commands. As the data-driven methods often rely on longer time series of telemetry data than the auto-regressive models used with LQG controllers, it begs the following questions: What are the limits of prediction beyond Markovian dynamics assumptions? What are the fundamental limits of such predictive control methods in AO, and how do these connect to other errors in wavefront estimation, such as spatial aliasing?

We approach these questions by studying predictive controllers from the perspective of regression analysis beyond Markovian state space models. Specifically, we couple the wavefront reconstruction and prediction steps under a single prediction problem to achieve theoretical performance limits for predictive controllers with different levels of assumptions on the turbulence. We utilize Gaussian process (GP) regression¹² to obtain principled uncertainty quantification for the predictive estimates given a time series of past observations. GP regression, also known as kriging in spatial statistics, is a technique used for regression tasks with complex relationships between variables. It is a tempting paradigm for AO as von Kàrmàn turbulence models are GPs defined by their spatiotemporal covariance functions or power spectral density (PSD), and versatile *a priori* information regarding the turbulence flow statistics can be introduced in the inference task. As an example, we consider a spatiotemporal GP (see Fig. 1) emerging from a multilayer FF turbulence. Such a probability distribution can be easily improved by hierarchical modeling to consider the uncertainty in the estimates concerning wind speeds and the C_N^2 profile. In practice, the turbulence parameters can also be identified with external algorithms and/or devices^{13–16} such as stereo-SCIDAR.^{17,18}

This paper explores the limits of predictive accuracy in GP regression by introducing two GP prior distributions for the spatiotemporal turbulence process that capture distinct levels of information: The first (very optimistic) prior distribution uses a multilayer FF turbulence model with perfect knowledge of the dynamics (wind directions, speeds, r_0 of all layers). In contrast, the second more conservative, prior distribution represents a scenario where only isotropic temporal

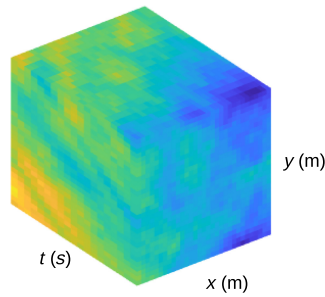


Fig. 1 Stack of spatial turbulence fields at consecutive time steps. When expanded to the temporal axis, the cumulative spatial phase aberrations can be modeled as a spatiotemporal GP. The resulting process is stationary and non-isotropic, with a covariance function that depends on spatial atmospheric parameters (fried parameter, C_N^2 profile, L_0) and temporal parameters (wind speed and direction).

correlation can be modeled, i.e., only the average wind speed is known, with no information on wind directions and C_N^2 profile. We also show how the spatiotemporal correlations enable the reconstruction of frequencies above the WFS sampling frequency, lowering the aliasing error of the DM correction. The sensitivity to high-order frequencies depends on the discretization of the WFS model; we study these discretization errors propagated to the predictive control.

Finally, we study how the number of past WFS measurements used for the prediction and reconstruction affects the prediction in different noise and atmospheric conditions. We compare different settings by computing the Bayes cost associated with different time series lengths and parametrizations of the system. As expected, when more precise prior knowledge of the dynamics is available, prediction can benefit from longer telemetry time series. Finally, we demonstrate that given a computational limitation on how many past frames we can use, it is better to use a sparse temporal sampling than a series of the last consecutive frames.

2 Related Work

This paper continues the development of predictive methods for AO control. Previous literature on the topic (see, for example, Ref. 19) demonstrated the potential for improved performance and increased stability of AO systems when predictive control techniques are used. However, further research is needed to optimize these methods for specific applications, improve their accuracy and efficiency, and analyze their effectiveness under different atmospheric conditions.

Predictive control algorithms include various types of model-based control methods such as Kalman filtering-based linear-quadratic-Gaussian controllers (LQG, e.g., Refs. 8, 20–27), as well as, closely related optimal controllers such as the \mathcal{H}_2 controller.^{28,29} In these works, the atmospheric turbulence is usually modeled with an autoregressive model in a predefined state space (either modal or zonal), and the derivation of the model sometimes includes parameter identification from past telemetry. The optimal control strategies also allow modeling dynamics beyond atmospheric turbulence, such as DM dynamics. These concepts have also been extended to tomographic systems; see, for example, Refs. 30 and 31.

More recently, neural-network-based reinforcement learning (RL) methods have been studied in various works.^{32–37} In RL, the control/system model is usually parametrized with a neural network, and the desired control is learned from the telemetry data with minimal prior assumptions on the system. However, RL is not limited to NN models, and similar strategies can be used with linear models, which leads to LQG-like derivation of the control law.^{38,39} A mathematical analysis of the overlap between RL and optimal control strategies is available in Ref. 40.

Another approach is to separate the reconstruction and prediction steps of AO control, that is, use a predictive filter that operates on an open-loop estimate (pseudo-open-loop measurement/reconstruction on the closed-loop system) of the full turbulence.¹⁰ These predictive filters aim to forecast the future wavefront based on past wavefront measurements utilizing a variety of techniques, such as Fourier analysis and Zernike polynomials, as well as machine learning. Fourier-based methods decompose the wavefront into its constituent frequency components and predict the future behavior of each component, while Zernike polynomials are a set of orthogonal

functions that can be used to represent the wavefront and predict its future behavior (e.g., Refs. 41–43). Machine learning methods can learn a predictive model of the wavefront based on historical data and make predictions based on the learned model (e.g., Refs. 10 and 44–47).

This work is closely connected to minimum variance predictive controllers, where the prediction model is constructed by utilizing the temporal or spatiotemporal covariance structure of turbulence.^{9,48,49} The cross-correlations can either be derived from the atmospheric model^{9,48,49} or by utilizing machine learning techniques.^{10,11,50} On the one hand, we expand the work of Doelman⁴⁸ from the temporal to the spatio-temporal domain, include the WFS model, and give estimates for optimal predictive control limit in all spatial locations on the telescope pupil. On the other hand, we discuss the spatio-temporal minimum variance prediction matrix (e.g., in Ref. 49) beyond deriving the minimum variance predictor for a fixed Markov state space. In particular, we derive the theoretical limits of minimum variance predictive control and study the effect of WFS modeling accuracy and the effect of the chosen past measurements on prediction accuracy under different imaging conditions and levels of prior information on the atmosphere.

Hence, we also take into account that the turbulence itself is not observed directly but through a WFS measurement, which connects our considerations to works that study priors from the perspective of wavefront reconstruction (see, e.g., Ref. 51). Moreover, methods that enable reconstruction above the sampling frequency of WFS have lately gained attention in the AO literature. Oberti et al.⁵² discussed how data from several misaligned WFS could be used to obtain super-resolution reconstructions, and Berdeu et al.⁵³ studied how the discretization of the reconstruction grid affects the aliasing error. The super-resolution has also been studied in the context of pyramid WFS. Correia et al.⁵⁴ showed how a single pyramid WFS, with facets (the four pupil images on the detector plane) that are adjusted to offset, also enables super-resolution. Contrarily, we use single slope-based WFS and utilize the FF structure of turbulence and past telemetry to gain information on the high-order spatial frequencies.

3 Atmospheric Turbulence as a Spatiotemporal GP

3.1 Spatial Modeling

The atmosphere is the main cause of distortion in the phase of the light that reaches ground-based telescopes. This distortion is caused by the random mixing of air at different temperatures, that is, the atmospheric turbulence, constantly moving due to wind. This movement causes variations in the refractive index of the air and, thus, in the optical path length of the incoming light. The turbulence is typically modeled as a series of thin, independent layers at different heights with different levels of turbulence strength. As the light propagates through such a thin layer, the resulting phase variations $\phi(x, y)$ can be modeled by a two-dimensional homogeneous GP described by the von Kàrmàn PSD

$$\hat{k}_\phi(\boldsymbol{\kappa}) \propto (|\boldsymbol{\kappa}|^2 + 1/L_0^2)^{-\frac{11}{6}}, \quad (1)$$

where $\boldsymbol{\kappa}$ is the spatial frequency, and L_0 is the so-called outer scale of the layer. The PSD expresses the amount of turbulent energy at a given spatial frequency. By the Wiener–Khinchin theorem, the corresponding covariance function $k_\phi(r)$ is given through the Fourier transform of the PSD. By computing the Fourier transformation, we get

$$k_\phi(r) = \frac{\Gamma(\frac{11}{6})}{\sqrt{2\pi^{\frac{8}{3}}}} \left(\frac{24}{5}\Gamma\left(\frac{6}{5}\right)\right)^{\frac{5}{6}} \left(\frac{L_0}{r_0}\right)^{\frac{5}{3}} \left(\frac{2\pi r}{L_0}\right)^{\frac{5}{6}} K_{\frac{5}{6}}\left(\frac{2\pi r}{L_0}\right), \quad (2)$$

where r_0 is the Fried parameter (defining the total turbulence strength), Γ is the gamma function, and $K_{\frac{5}{6}}$ is the modified Bessel function of the second kind. For a detailed derivation of the PSD and the covariance function, see, e.g., Conan.⁵⁵ The final cumulative optical path aberrations Φ is the sum of the aberrations along the line of sight

$$\Phi(x, y) = \sum_{\ell=1}^L \sqrt{\rho_\ell} \phi_\ell(x, y), \quad (3)$$

where L is the total number of layers, and ρ_ℓ and ϕ_ℓ are, respectively, the relative strength and the phase aberrations of the ℓ 'th turbulence layer. The collection of the relative strengths $[\rho_1, \dots, \rho_L]$

at the considered layers is called the discrete C_N^2 profile. The spatial turbulence statistics can be estimated, for example, from telemetry data.¹³

In what follows, we restrict our attention to the von Kàrmàn model given in Eq. (1), but let us in any case remark that deviations from the von Kàrmàn power law close to the ground^{56,57} and in the upper troposphere and stratosphere^{58–60} are well-documented. Similarly, the turbulence taking place in the telescope dome (so-called dome seeing) has different characteristics and is naturally dependent on the dome geometry.⁶¹ While there are methods to estimate these deviations, such as proposed by Helin et al.,⁶² they are based on solving an ill-posed problem over a relatively long time series of telemetry data leading to inherent uncertainty about temporal fluctuations. The non-stationarity of the turbulence is discussed more thoroughly in Ref. 63.

3.2 Spatiotemporal Model

On the millisecond scale of AO control, the changes in the turbulence pattern itself are small, and the temporal evolution is mainly driven by advection, i.e., wind at the altitudes of the layers. Hence, Taylor’s FF hypothesis provides a good approximation to the time evolution; i.e., each turbulent layer is modeled as a thin static “frozen” layer sliding over the telescope with an individual wind speed and direction.

There have been attempts to involve also other physical effects from the underpinning Navier–Stokes equation, such as kinematic diffusion⁶⁴ and intermittency.⁶⁵ To our knowledge, these dynamics models have not been studied in a predictive control context.

We now formulate two spatiotemporal distributions that will be considered in the regression below. The first distribution is based on exact knowledge of the FF model, including the correct C_N^2 profile, and it is therefore optimistic. In the second distribution, we assume that we have an estimate of the weight-averaged wind velocity over the atmosphere and, more importantly, that the spatial and temporal statistics are essentially treated independently since there is no deterministic flow originating from the spatial process.

We note that the first distribution is overly optimistic since it assumes perfect detection of FF model hyper-parameters. Poyneer et al.⁶⁶ showed that even though FF was detected most of the time from Altair and Keck AO system telemetry, it only covered 20% to 40% of the total controllable phase and originated usually from one to three layers. The latter distribution is more realistic as it neglects strong spatiotemporal correlations emerging in advection-dominated time scales. This model still assumes stable coherence time and von Kàrmàn spectrum, making it somewhat over-optimistic. Uncertainty in these hyperparameters can be added to the models similarly to uncertainty in the wind directions. The models can also be combined to simulate/predict where part of the FF cannot be identified or turbulence is only partially in FF but still follows stable coherence time.

3.2.1 FF-induced spatiotemporal GP

Let us first introduce the GP model based on FF with the exact knowledge of the layered model. We will refer to it as the FF-GP for convenience.

Consider FF on a single layer. At a given initial time instance, $t = 0$, the phase aberrations $\psi(x, y, 0) = \phi(x, y)$ follow a two-dimensional GP statistics, i.e., $\phi(x, y) \sim GP(0, k_\phi(r))$. We expand the spatial GP to the time domain according to the FF model

$$\psi(x, y, t) = \phi(x + v_x t, y + v_y t), \quad (4)$$

where $\mathbf{v} = (v_x, v_y)$ stands for the two-dimensional wind velocity vector. Consequently, the covariance function of the spatiotemporal process is obtained by

$$k_\psi((x_1, y_1, t_1), (x_2, y_2, t_2)) = k_\phi\left(\sqrt{(x_1 - x_2 + (t_1 - t_2)v_x)^2 + (y_1 - y_2 + (t_1 - t_2)v_y)^2}\right).$$

The full multi-layer phase aberrations at the entrance pupil Ψ are obtained by adding the single-layer spatiotemporal GPs together, that is

$$\Psi_{\text{FF}} \sim GP(0, k_{\Psi_{\text{FF}}}((x_1, y_1, t_1), (x_2, y_2, t_2))), \quad (5)$$

where

$$k_{\Psi_{\text{FF}}}((x_1, y_1, t_1), (x_2, y_2, t_2)) = \sum_{\ell=1}^L \rho_{\ell} k_{\psi_{\ell}}((x_1, y_1, t_1), (x_2, y_2, t_2)).$$

3.2.2 Wind-averaged FF-GP

The information about atmospheric profiles, including wind directions, can contain various uncertainties. Here, we assume that reliable information is only available regarding the weight-averaged wind velocity $|\mathbf{v}_{\text{avg}}|$ through the atmosphere. In practice, reliable information is often available regarding $|\mathbf{v}_{\text{avg}}|$, e.g., the coherence time τ_0 is directly related to this quantity via

$$\tau_0 = 6.88^{-\frac{3}{5}} \frac{r_0}{|\mathbf{v}_{\text{avg}}|}, \quad (6)$$

where r_0 is the Fried parameter. The uncertainty regarding wind direction can be rephrased as an assumption that each direction is equally probable, i.e., the wind direction is uniformly distributed $\boldsymbol{\theta} := \mathbf{v}_{\text{avg}}/|\mathbf{v}_{\text{avg}}| \sim U(\mathbb{S}^1)$, where \mathbb{S}^1 stands for the unit circle. We will refer to this model as wind-averaged FF-GP (WAFF-GP).

We formulate WAFF-GP as a zero-mean process $\Psi_{\text{WAFF}} \sim GP(0, k_{\Psi_{\text{WAFF}}}((x_1, y_1, t_1), (x_2, y_2, t_2)))$, where

$$k_{\Psi_{\text{WAFF}}}((x_1, y_1, t_1), (x_2, y_2, t_2)) = \frac{1}{|\mathbb{S}^1|} \int_{\mathbb{S}^1} k_{\phi} \left(\sqrt{(x_1 - x_2 + \Delta s \theta_x)^2 + (y_1 - y_2 + \Delta s \theta_y)^2} \right) d\boldsymbol{\theta}, \quad (7)$$

with the convention $\Delta s = (t_1 - t_2)|\mathbf{v}_{\text{avg}}|$ and $\boldsymbol{\theta} = (\theta_x, \theta_y) \in \mathbb{S}^1$. The covariance function specified by the identity in Eq. (7) is well defined as it is an average of well-defined covariances.¹²

Note that the process formulated this way coincides with the spatial von K arm an statistics for any fixed time $t_1 = t_2$, i.e.

$$k_{\Psi_{\text{WAFF}}}((x_1, y_1, t), (x_2, y_2, t)) = k_{\phi}((x_1, y_1), (x_2, y_2)).$$

For the purpose of this study, WAFF-GP presents a more pessimistic model of spatiotemporal statistics, where less information is included than is often practically available. That being said, one could relax the definition in Eq. (7) further by assuming a weighted average over $|\mathbf{v}_{\text{avg}}|$ with respect to some probability density, e.g., a uniform distribution over a confidence interval $[|\mathbf{v}_{\text{avg}}| - \epsilon, |\mathbf{v}_{\text{avg}}| + \epsilon]$ modeling the measurement accuracy. Furthermore, WAFF and FF models can be combined together to create a prior for conditions where we, for example, detect three FF layers that cover 40% of the turbulence, and movement of the other layers cannot be detected.⁶⁶

4 Reconstruction and Prediction

In this section, we describe the observational model in AO and how the Bayesian inference and prediction are carried out.

4.1 AO System and Wavefront Sensing

Single-conjugate AO systems use a single natural guide star as the reference source. The wavefront control comprises two main components: the WFS and the DM. The WFS measures the distortion caused by the atmosphere in the incoming phase along the line of sight, and the DM then compensates for these distortions by taking a shape that cancels them out. The WFS is usually set downstream from the DM, effectively measuring the deviation from a flat wavefront (closed-loop residuals). However, if the DM influence function and the control delay in the system are known, the open-loop measurement can be recovered through the so-called pseudo-open loop scheme.

This paper focuses on the open-loop setup, i.e., we assume that the WFS observes the full phase error caused by the atmosphere. We model the DM with Gaussian influence functions.

Next, let us describe the non-direct observation model. We assume that WFS measures the gradient of phase aberrations ϕ averaged over a period $\Delta\tau$ set by the AO system's frame rate on a given spatial sampling defined by, e.g., the number of lenslets in the Shack–Hartmann WFS (SHS)

$$w_i = \frac{1}{|B_i|} \int_t^{t+\Delta\tau} \iint_{B_i} \nabla\phi(x, y, t) dx dy dt \in \mathbb{R}^2, \quad (8)$$

where B_i is the sub-aperture surface indexed by $i = 1, \dots, M$. The concatenation of w at all possible locations is denoted by a measurement vector $\mathbf{w} \in \mathbb{R}^{2M}$. However, the methods discussed in Secs. 4.2 and 5 apply to any linear or linearly approximated WFS, where the mathematical model can be written out as a matrix.

Furthermore, suppose our computational resources allow utilizing the latest p subsequent data vectors with recording intervals of Δt in the prediction task. Let us denote these vectors by

$$\mathbf{w}^1, \mathbf{w}^2, \dots, \mathbf{w}^p \in \mathbb{R}^{2M}, \quad (9)$$

with the (loose) convention $\mathbf{w}^k = \mathbf{w}(k\Delta t)$ for any time-dependent entity.

4.2 Bayesian Inference and Prediction

Let us consider the incoming phase aberrations and their representation on a discrete grid. We utilize evenly spaced spatial grids following the Fried geometry or a super-sampled grid where each SHS-lenslet consists of an evenly spaced pixel grid (e.g., 4×4 or 8×8 pixels); see Fig. 2.

For any fixed time instance t , the state vector $\phi(t)$ belongs to \mathbb{R}^{N^2} (minus the inactive pixels). Here, N^2 is the number of pixels (either in the Fried or a super-resolution grid), that is, N is the number of spatial locations along a single spatial dimension. Let us write

$$\phi^1, \phi^2, \dots, \phi^p \in \mathbb{R}^{N^2},$$

for the state vectors representing the phase on the different time steps, following the notation in Eq. (9).

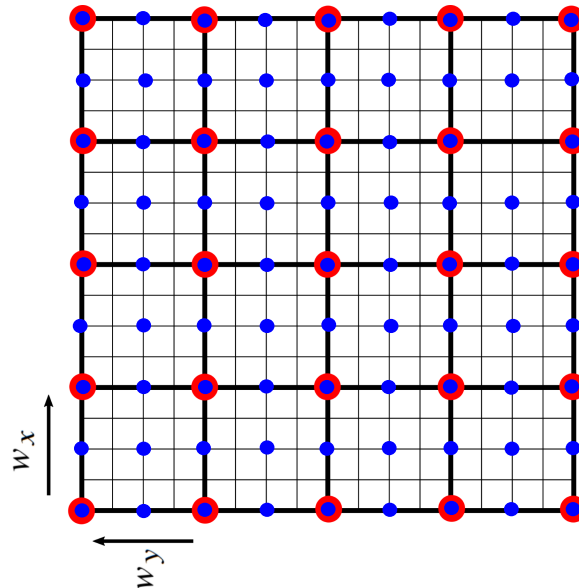


Fig. 2 Different discretization of the state and measurement domains used in this study. Thick black lines depict a 4×4 SHS sensor, the red points represent the locations of the DM actuators in the Fried geometry, and the blue points are the so-called extended Fried grid, where a set of pixels are added between the Fried grid. The SHS measures the average gradient over each lenslet. This work considers discretization grids depicted by the thin black lines to various amounts of pixels [4×4 (in the figure) to 8×8] and more coarse discretization grids defined by the Fried geometry and extended Fried grid (red dots, blue dots).

Our aim is to predict/reconstruct $\boldsymbol{\phi}^{p+2}$ as the typical time lag in next-generation telescopes is around two-time steps (expanding the method to non-integer delays is trivial). Therefore, our state vector will include $\boldsymbol{\phi}^{p+2}$ as well. Let us denote the concatenated state and data vectors as

$$\boldsymbol{\Phi} = \begin{pmatrix} \boldsymbol{\phi}^1 \\ \vdots \\ \boldsymbol{\phi}^p \\ \boldsymbol{\phi}^{p+2} \end{pmatrix} \in \mathbb{R}^{(p+2)N^2} \quad \text{and} \quad \boldsymbol{W} = \begin{pmatrix} \boldsymbol{w}^1 \\ \vdots \\ \boldsymbol{w}^p \end{pmatrix} \in \mathbb{R}^{2pM}.$$

We can formulate our prediction task as an inverse problem of solving for $\boldsymbol{\Phi}$ in

$$\boldsymbol{W} = \mathbf{A}\boldsymbol{\Phi} + \boldsymbol{\epsilon}, \quad (10)$$

where the forward operator is given as a block diagonal matrix

$$\mathbf{A} = \begin{pmatrix} A & 0 & \dots & 0 & 0 \\ 0 & A & & \vdots & \vdots \\ \vdots & & \ddots & & \\ 0 & & & A & 0 \end{pmatrix} \in \mathbb{R}^{pM \times (p+1)N}. \quad (11)$$

Each matrix A in Eq. (11) maps an incoming state $\boldsymbol{\phi}^i$, $i = 1, \dots, p$, to the corresponding data vector \boldsymbol{w}^i according to Eq. (8). We model the noise $\boldsymbol{\epsilon}$ as a zero-mean Gaussian random variable with a symmetric positive definite covariance matrix $C_{\text{noise}} \in \mathbb{R}^{2pM \times 2pM}$.

In the Bayesian paradigm, the solution to an inference problem Eq. (10) is the conditional distribution of $\boldsymbol{\Phi}$ given observational data \boldsymbol{W} , i.e., the posterior distribution. By the Bayes' formula, the Gaussian prior models crafted in Sec. 3.2 combined with the Gaussian likelihood yields a Gaussian posterior.

Here, the Gaussian prior distribution over $\boldsymbol{\Phi}$ has zero mean and the covariance matrix C_{prior} given by

$$(C_{\text{prior}})_{ij} = k_{\Psi_{\text{FF/WAFF}}}((x_i, y_i, t_i), (x_j, y_j, t_j)),$$

where $i, j = 1, \dots, (p+2)N^2$ corresponds to global indexing over the spatial and temporal variables.

It follows that the Gaussian posterior distribution is defined by the covariance matrix

$$C_{\text{post}} = (\mathbf{A}^T C_{\text{noise}}^{-1} \mathbf{A} + C_{\text{prior}}^{-1})^{-1} \quad (12)$$

and the mean vector

$$\boldsymbol{\Phi}_{\text{post}} = C_{\text{post}} \mathbf{A}^T C_{\text{noise}}^{-1} \boldsymbol{W}.$$

The sought-for prediction entails assessing the marginal posterior distribution of $\boldsymbol{\phi}^{p+2}$, i.e., the last N^2 components in $\boldsymbol{\Phi}$. For convenience, let us denote by $P: \mathbb{R}^{(p+2)N^2} \rightarrow \mathbb{R}^{N^2}$, a matrix projection that maps the full-state vector to the state at time $p+2$, i.e., $P\boldsymbol{\Phi} = \boldsymbol{\phi}^{p+2}$. It follows that the predictive posterior at time $p+2$ is given by a Gaussian distribution with the mean $P\boldsymbol{\Phi}_{\text{post}}$ and the covariance matrix $PC_{\text{post}}P^T$.

Since the conditional mean and the MAP estimate coincide for a Gaussian posterior, the natural choice for the point estimate (i.e., the prediction to be used in control) is the mean value of the posterior. The spatiotemporal AO prediction matrix $R_{\text{FF/WAFF}}$ thus takes the form

$$R_{\text{FF/WAFF}} = PC_{\text{post}} \mathbf{A}^T C_{\text{noise}}^{-1}. \quad (13)$$

Furthermore, the corresponding DM commands are calculated as standard least-squares-fit to the DM influence function. Also, the posterior covariance $P\boldsymbol{\Phi}_{\text{post}}$ can be projected to DM space using the least-squares-fitting matrix.

5 Bayesian Utility of Experimental Designs

Experimental design is the process of determining how to perform an experiment such that the informativeness of the data is maximized; therefore, the uncertainty in the estimates produced is minimized. For a general review of Bayesian optimal experimental design (OED), we refer the reader to Chaloner and Verdinelli.⁶⁷ For AO control, the design of the experiment can mean, e.g., the choice of the length of telemetry time series that are given as input for the predictive controller, as described in Sec. 6.4. It could also address other design parameters, though outside the scope of this paper, such as the number of sub-apertures to be included in the WFS or the modulation amplitude for a pyramid WFS in given conditions.

In this paper, we study the informativeness of past WFS measurements on the predictive reconstruction quality. It is a naturally occurring question since the computational resources are scarce in AO, and one would naturally like to minimize the amount of data that is processed. Also, the GP model allows one to write the Bayesian expected loss/utility as a function of the posterior covariance matrix, as described in Eq. (14). These formulas are relatively efficient in evaluating and making computing optimal values for the decision variables feasible even for high-dimensional problems as encountered in AO.

Let us briefly review the basics of Bayesian OED in our framework. For this, assume that we have a parameter $\mathbf{d} \in \mathbb{R}^n$ that defines an experimental design for AO control, and let us account for the dependence of the forward model on the design parameter by writing the system matrix as $\mathbf{A}(\mathbf{d})$. One then chooses a loss function $u(\Phi, \mathbf{W}, \mathbf{d})$ and defines the expected loss (or negative utility) of the design as

$$U(\mathbf{d}) = \mathbb{E}_{\Phi, \mathbf{W}}[u(\Phi, \mathbf{W}, \mathbf{d})],$$

for which small values indicate informative measurements (if the choice of the loss function is adequately chosen). Comparing different experimental designs involves computing $U(\mathbf{d})$ (and possibly also its gradient) for a number of \mathbf{d} , which can be a nontrivial and computationally demanding task in general. However, in our linear and Gaussian setting, choosing the loss function as the quadratic loss

$$u(\Phi, \mathbf{W}, \mathbf{d}) = \|P(\Phi - \Phi_{\text{post}}(\mathbf{W}, \mathbf{d}))\|_2^2$$

results in the well-known A-optimality criterion. More precisely, the corresponding expected loss (or minimization target) reduces to

$$U(\mathbf{d}) = \text{tr}(PC_{\text{post}}(\mathbf{d})P^T), \quad (14)$$

where the dependence of C_{post} on \mathbf{d} is inherited from $\mathbf{A}(\mathbf{d})$ via Eq. (12). For derivations of these formulas, see, e.g., Burger et al.⁶⁸

In Sec. 6.4, we will evaluate the utility of the data on the optimality criteria for the different prior models and different noise levels. Thus, the design variable \mathbf{d} will be the choice of timesteps p used for computing the posterior. Note that including more time steps in the prediction process always decreases the uncertainty about the unknown, thus reducing the expected error. However, if a clearly diminishing return in the value of longer telemetry time series is observed, then one may conclude that the benefit from the usage of the extra data does not merit the required additional computational expense.

6 Numerical Experiments

This section demonstrates the performance of spatiotemporal GP prediction (FF-GP and WAFF-GP) through numerical experiments, where we utilize the HciPy toolbox.⁶⁹ We design the numerical experiments to demonstrate three key features of the approach: the predictive capacity, the noise reduction (see Sec. 6.2), and the ability to recover spatial frequencies above the sampling frequency of the WFS (see Sec. 6.3).

Furthermore, we demonstrate how discretization of the measurement model affects the reconstruction quality and how the Bayes loss can be used to decide the right reconstructor design for the given conditions and AO system (see Sec. 6.4). The quantity of interest in these experiments is mainly the variance of the posterior distribution that gives essentially the measure for the performance of different priors in different experiments. We compare FF-GP and

WAFF-GP models with the standard minimum variance reconstruction that only considers spatial statistics, called spatial GP (S-GP).

The three different reconstruction models are

1. **FF-GP**: the FF-GP prior that assumes perfect information about the FF and is defined by Eq. (5)
2. **WAFF-GP**: the wind-averaged FF-GP prior unknown wind direction and C_N^2 profile defined by Eq. (7)
3. **S-GP**: a spatial prior that only considers spatial information, that is, every time step is reconstructed separately. The model is non-predictive.

The Bayes formula straightforwardly gives the standard spatial reconstruction method S-GP for a single frame. It yields a Gaussian posterior distribution with the covariance matrix

$$C_{\text{post}} = (A^T C_{\text{noise}}^{-1} A + C_{\text{spatial}}^{-1})^{-1} \tag{15}$$

and the mean vector

$$\phi_{\text{post}}^p = C_{\text{post}} A^T w^p. \tag{16}$$

As mentioned above, this reconstruction technique does not consider temporal statistics, and it is, hence, always limited by the temporal delay of the AO system.

6.1 Simulation Set-Up

We simulated a 3.2-m telescope with 14% central obstruction and linear slope-based WFS (see Sec. 4.1) with 16×16 lenslets (e.g., 20-cm actuator spacing). The WFS sampling is chosen to represent XAO systems. Instead of simulating data by sliding two-dimensional FF layers and interpolating, we generate the data by constructing the spatiotemporal covariance matrix as explained in Sec. 3.2 and sampling from the model. Generating data in this manner does not include any approximations, such as interpolating the temporal movement; that is, no high-order frequencies are dampened in the temporal movement.

The parameters for the von Kàrmàn turbulence are $r_0 = 16.8$ cm and $L_0 = 20$ m. The time step between frames is set to $\Delta\tau = 2$ ms, and the atmosphere is composed of seven layers. A complete list of the parameter values employed in the simulations can be found in Tables 1 and 2.

Table 1 Simulation parameters.

| Telescope | | |
|-----------------------|-------------|------------------|
| Parameter | Value | Units |
| Telescope diameter | 3.2 | m |
| Obstruction ratio | 14 | Percent |
| Sampling frequency | 500/1000 Hz | Hz |
| Measurement noise | 100/4 | S/N % |
| WFS wavelength | 0.79 | μm |
| WFS lenslets | 16 | Across the pupil |
| Pixels | 128 | Across the pupil |
| DM actuators | 17 | Across the pupil |
| DM influence function | Gaussian | — |
| DM coupling | 40 | Percent |

Table 2 Atmospheric parameters.

| Atmosphere parameters (15 cm at 500 nm) | | | | |
|---|------------------------|------------------|---------|-----------|
| Layer | Wind direction (angle) | Wind speed (m/s) | C_n^2 | L_0 (m) |
| 1 | 80.2 | 8.5 | 0.672 | 20 |
| 2 | 90.0 | 6.55 | 0.051 | 20 |
| 3 | 95.7 | 6.6 | 0.028 | 20 |
| 4 | 101.4 | 6.7 | 0.106 | 20 |
| 5 | 177.6 | 22 | 0.08 | 20 |
| 6 | 183.3 | 9.5 | 0.052 | 20 |
| 7 | 189.1 | 5.6 | 0.01 | 20 |

6.2 Wavefront Prediction, Noise Reduction

The reconstruction for different spatiotemporal priors (for the whole time interval) is obtained with equations described in Sec. 4.2, and the prediction is the marginal distribution of ϕ^{p+2} . The spatial reconstruction (S-GP) is obtained via Eqs. (15) and (16). Since this method does not model temporal evolution, the reconstruction does not include the variance from the time delay. Hence, following the standard error budget modeling, we add the temporal variance term to the estimate to obtain the right uncertainty estimate for the reconstruction error. The variance of the change between two frames, that is, the variance of $\phi(x, y, t_1) - \phi(x, y, t_2)$, can be calculated using the spatiotemporal covariance function Eq. (5)

$$\sigma_{\text{temp}}^2 = \text{Var}(\phi(x, y, t_1) - \phi(x, y, t_2)) = 2k_{\psi}(0) - 2k_{\psi}((x, y, t_1), (x, y, t_2)). \quad (17)$$

The variance of the reconstruction error at a given location for this method is then the sum of the appropriate diagonal element in the posterior covariance Eq. (15) and the temporal variance Eq. (17).

We compare the performance of the FF-GP and WAFF-GP predictive reconstruction with the non-predictive S-GP model by examining the full posterior variance Eq. (12) and the posterior variance filtered with DM influence functions. Since the SH-WFS is a slope sensor that is not sensitive to the piston mode, the uncertainty in the piston mode dominates the posterior variance. Furthermore, the global piston mode does not affect the performance of the AO system. Hence, we filter out the piston mode in the uncertainty and the mean in all comparisons.

Figure 3 shows the piston-free predictive reconstruction accuracy of low and high noise regimes on phase and DM space, where we discretize the wavefront according to 8 pixels per WFS lenslet. The images illustrate the spatial uncertainty for different methods (FF-GP, WAFF-GP, and S-GP), i.e., the diagonal of the marginal covariance matrix of ϕ_{p+2} as an image. We observe that FF-GP delivers the smallest posterior variance, the WAFF-GP the second smallest, and the S-GP the largest for both noise levels, as expected. The posterior variances for WAFF-GP and S-GP are symmetrical since they do not consider the lateral movement in the FF hypothesis. On the other hand, the FF-GP takes into account the advection, and hence, at the edges of the telescope pupil, we see a reduction in variance downstream of the wind. Moreover, the phase reconstruction accuracy (i.e., no DM) of all the methods shows a checker-board-like pattern because the SH operator essentially gives information on the edges of lenslets. The pattern is most pronounced for S-GP and WAFF-GP. The FF can again use the FF advection to lower the uncertainty in the middle of lenslets.

Figure 4 shows the relative gain of using the predictive reconstruction models compared with the non-predictive S-GP model on DM space. In low noise conditions, FF-GP offers a factor of 3 to 3.5 (depending on the aperture location) improvement in the posterior variance, while WAFF-GP offers a factor of 2 to 2.4 improvement [see Fig. 4(a)]. In high noise conditions, the improvement factors are 2.0 to 2.25 and 1.4 to 1.8 [see Fig. 4(b)]. The performance gain is

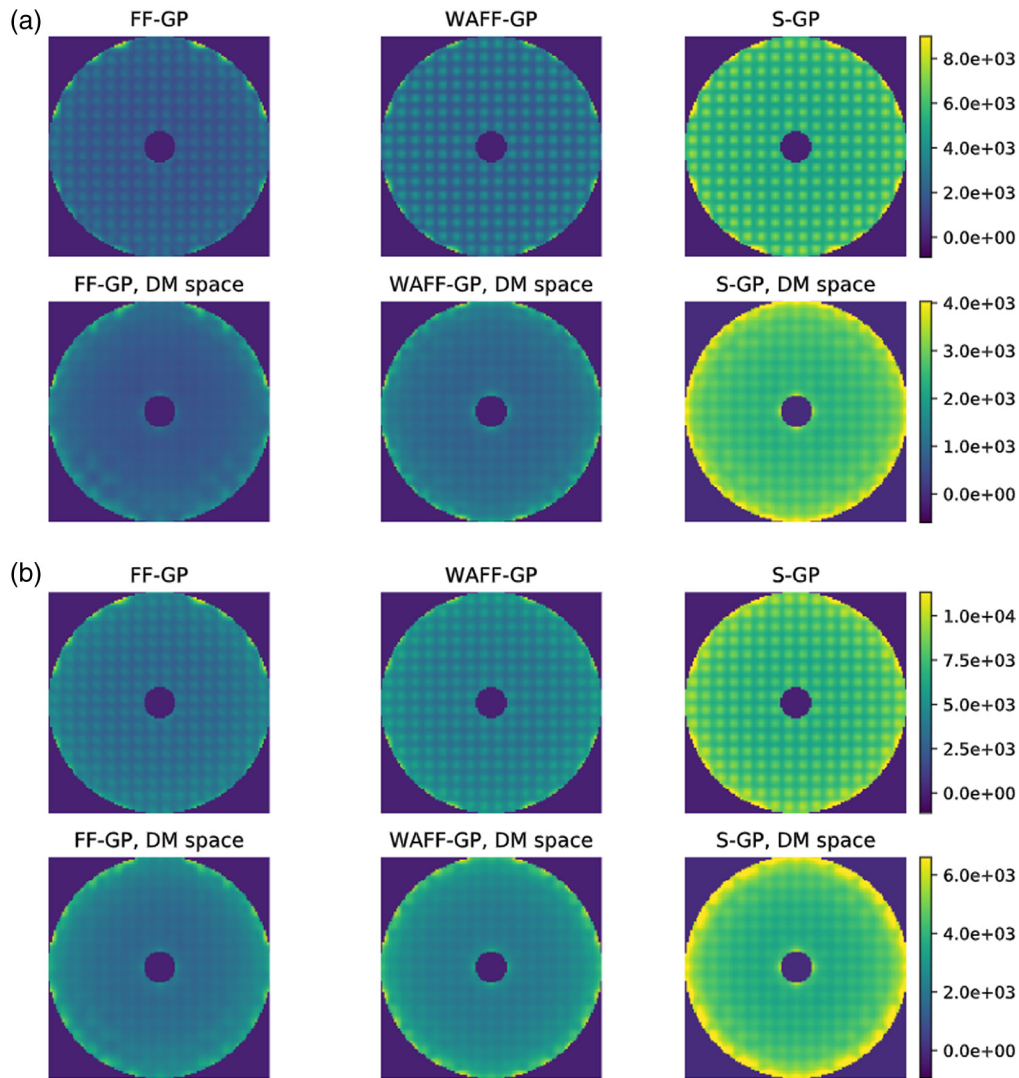


Fig. 3 Piston-free predictive reconstruction accuracy/variance (nm^2) in low [panel (a), $S/N = 100$] and high [panel (b), $S/N = 4$] noise regimes. The upper rows correspond to the full-phase estimate, and the lower row images are the least-squares fit to the DM modes. The images illustrate the spatial uncertainty, i.e., the diagonal of the marginal posterior covariance matrix ϕ^{p+2} , as an image for different methods (FF-GP, WAFF-GP, and S-GP). As S-GP does not predict, the temporal variance of two timesteps has been added to the variance estimate; see Eq. (17).

attributed to three different terms: temporal error, measurement noise, and aliasing. Prediction-wise the MAP estimate gives the minimum variance estimate of the future turbulence. Also, the usage of multiple measurements from the past allows the predictive reconstructor to average the measurement noise. Moreover, the usage of FF advection enables recovering frequencies above the cutoff frequency of the WFS, as we can see from the diminished checkerboard pattern for the FF model in Fig. 3.

6.3 Effect of Modeling Errors

The uncertainty estimates derived in the preceding subsection were obtained with the full WFS measurement model accuracy (i.e., 8×8 pixels inside each lenslet). However, using the full accuracy model comes with a cost in computational complexity in computing the posterior distribution Eq. (12) and the reconstruction matrix Eq. (13), and deriving such estimates with full accuracy becomes computationally unfeasible when bigger telescopes and longer time series are considered (see Sec. 7.1.1). Here, we examine the effect of the discretization parameter S of the WFS model that defines the discretization grid and, consequently, the accuracy of the model and the computational expense. We consider five different discretization grids: the Fried grid ($S = 1$),

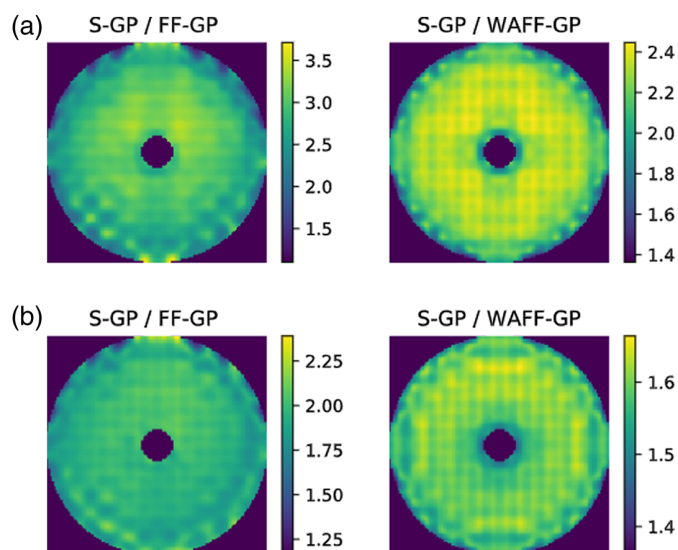


Fig. 4 Relative gain in the variance of the spatial prediction in DM space for the spatiotemporal models in low [panel (a), $S/N = 100$] and high [panel (b), $S/N = 4$] noise regimes. The images illustrate the ratio between the baseline spatial prediction (S-GP) variance and the spatiotemporal predictions (FF-GP and WAFF-GP) variances. A significant gain in the posterior variance is observed for both spatiotemporal models.

an extended Fried grid, where extra pixels are added in between DM actuators ($S = 2 + 1$), 4×4 pixels inside each lenslet ($S = 4$), 6×6 pixels inside each lenslet ($S = 6$), and the full 8×8 grid (see Fig. 2). Since our model in Eq. (12) does explicitly account for modeling errors, the accuracy of the WFS model also affects the posterior variance calculations. Hence, instead of uncertainty estimates, we ran a simple Monte Carlo simulation and compared the reconstruction accuracy of the models. The data are created with the full 8×8 grid ($S = 8$).

Figure 5 presents a comparison between the considered discretization levels. Figure 5(a) is for the FF prior, and Fig. 5(b) is for the WAFF model. The fitting error image in both figures is obtained by projecting ϕ_{p+2} to the DM influence functions. As expected, the reconstruction error is small for both models with larger S . The WFS model for $S = 1$ basically assumes that a WFS does not measure spatial frequencies above the DM spacing, making it more prone to aliasing error, while the $S = 4$ pixels grid provides performance very close to the full discretization grid for both models; hence, all the following experiments are calculated using $S = 4$.

6.4 Utility of Measurements

The numerical experiment of this section aims to investigate the utility of including more past timesteps in the prediction for both spatiotemporal models, FF-GP and WAFF-GP. Using a longer time series increases the computational demands of the prediction, and thus, beyond a certain point, additional data only provide a small benefit; thus, one can limit the number of time steps. In the experiment, the utility of the measurement is computed for time series lengths of 1 to 16 for two different noise levels. The utility function used is the square root of the trace of the posterior covariance matrix, i.e., the square root of the Bayesian A-optimality indicator, which measures the expected reconstruction error over the discretization grid projected to DM space. The expected reconstruction errors are compared with the setting where only the prediction's current measurement at $t = 0$ is used (e.g., AR model of the first order). The telescope and atmosphere parameters used are the same as in Secs. 6.2 and 6.3.

We note that since the covariance of the posterior does not depend on the data in the considered Gaussian setting, we do not need to draw data from the prior for this experiment. The results are shown in Fig. 6. For the FF-GP model, the quality of the prediction increases more with additional data, and the slope of the curves only starts leveling out near the end of the considered history interval. The improvement is more modest for the WAFF-GP model. Be that as it may, considering all data in the studied history range (and beyond) seems advantageous for both prior models as long as it is permitted by the computational constraints.

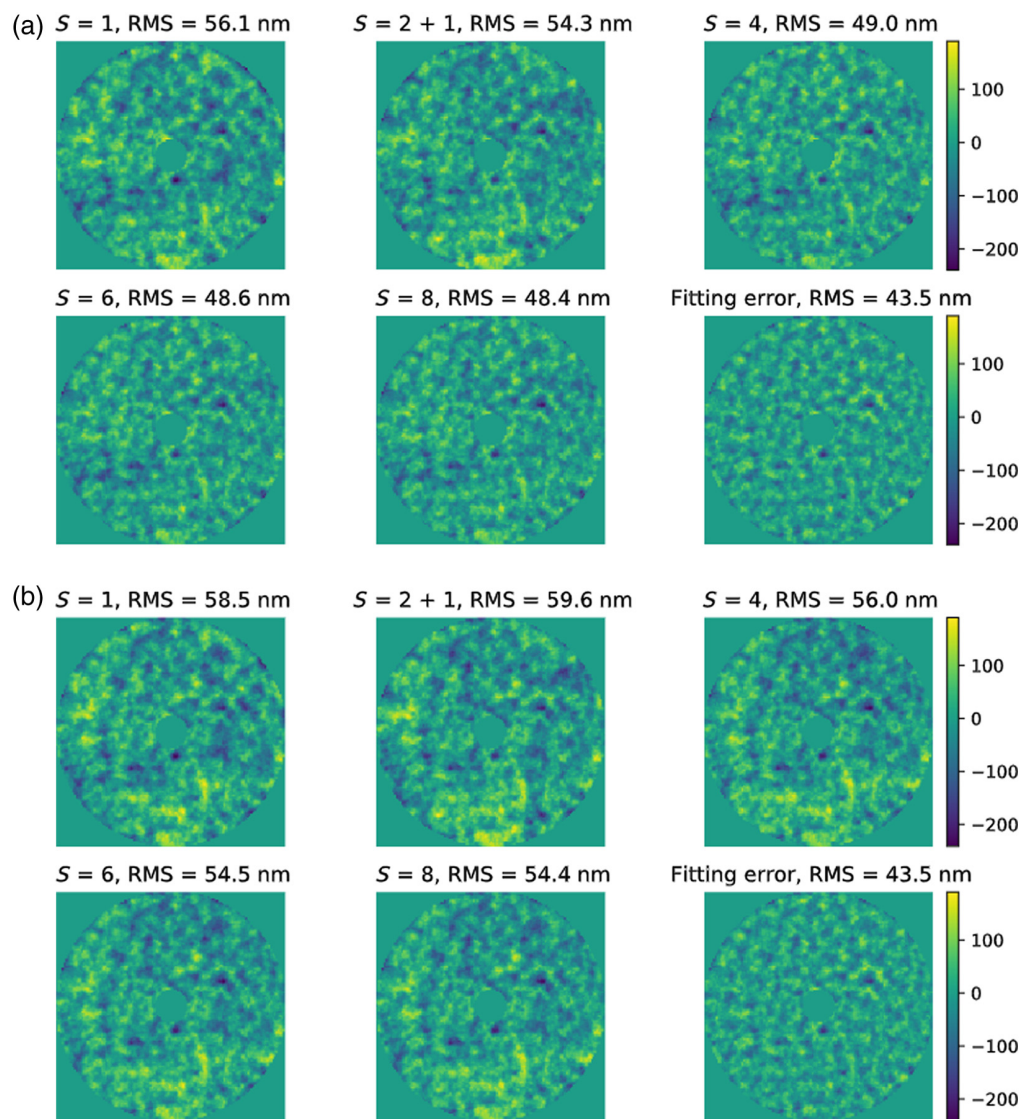


Fig. 5 Effect of the discretization accuracy on the reconstruction accuracy in Monte Carlo simulation. (a) FF-GP prior. (b) WAFF-GP prior.

6.4.1 Choice of data history

The limiting factor for using a long history of data is the matrix inversions for deriving the control matrix [Eq. (12)], not the capability to keep longer time series accessible for the controller. Hence, in addition to the question on a useful history length, another important question is which previous time steps should be used in the prediction. As the considered Gaussian priors exhibit structures of a certain size, it may not be optimal (given limitations in computational resources) to use only the most recent available data but rather choose more sparse temporal presentations from the past data.

To investigate this, we compute an optimal combination of history data to be included in the prediction by resorting to a greedy algorithm that iteratively includes data from past timesteps to minimize the associated expected reconstruction error. The posterior covariance matrix is formed and projected to the DM space, the A-optimality target is computed for each possible addition, and the choice that yields the lowest target value is added to the time series. The algorithm can be iteratively continued to include data from as many past timesteps as desired. In the numerical experiment, we compute the first five optimal choices; to be precise, the first included data always correspond to the latest timestep, $t = 0$, and the other four timesteps are chosen with the algorithm. We note that choosing the data in a greedy manner does not guarantee that the

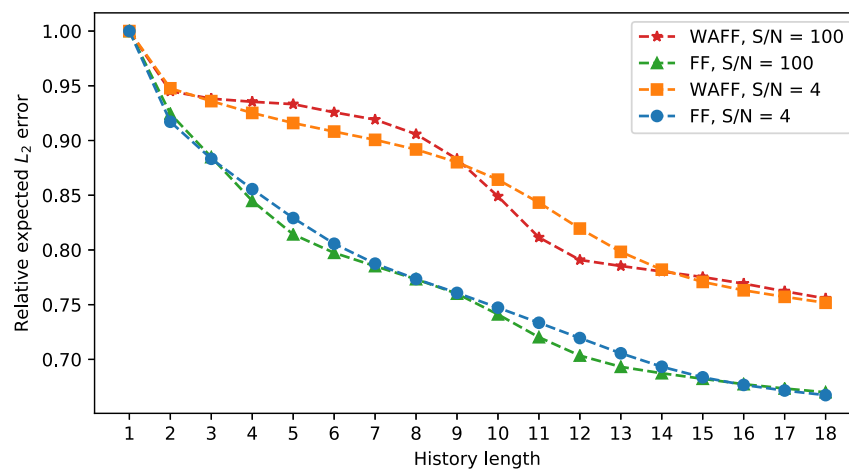


Fig. 6 Relative expected prediction errors for the FF-GP and WAFF-GP models for different lengths for the time series data in comparison to only using the current state information.

choices would be globally optimal, but such an optimization approach is adopted due to computational limitations.

Figure 7 shows the target values for the first four iterations of the greedy algorithm for both the FF-GP and the WAFF-GP models, with 2 ms frame rate and a signal-to-noise ratio of 100. As expected, choosing the latest data is not the optimal solution; instead, it seems to be better to choose the data that are temporally sparser than the frame rate at which the system operates.

The chosen timesteps, in the order that they were included by the algorithm, are $[0, 5, 13, 9, 2]$ for the FF-GP and $[0, 13, 6, 17, 3]$ for the WAFF-GP. In both cases, the final sampling of the data corresponds to intervals of roughly 6 to 8 ms (two to four frames). The optimal timesteps for the FF-GP lead to a relative expected error of 0.725, which corresponds to the same value as using the 11 to 12 latest consecutive timesteps and gives a 10% improvement compared with using just five latest consecutive steps; cf. Fig. 6. For the WAFF model, the corresponding numbers are 13 to 14 steps and 15%.

We also experimented with how the frame rate and measurement noise affect the optimal use of history data. A faster frame rate favors sparser presentation while increasing the measurement noise leads to more dense temporal sampling.

7 Discussion

We present a predictive approach based on GP modeling. In the context of the FF assumption and linear wavefront sensing, the presented FF-GP model is optimal in the least-squares sense conditioned on the fixed times series of WFS data and the specified spatiotemporal (Von Kàrmàn) prior to the turbulence. Hence, the derived improvements correspond to the limits achievable by predictive control for Von Kàrmàn turbulence. We also studied a less informative model, WAFF-GP, with only coarse assumptions about the atmosphere. These models allow a closed-form estimate of how good reconstruction and prediction can ideally be achieved with given assumptions on the telescope geometry and atmospheric conditions and our knowledge of them.

As mentioned above, the linear predictive filter/reconstructor R_{FF} is optimal in the least-squares sense conditioned on the fixed time series of WFS data and the specified spatiotemporal prior to the turbulence. Consequently, the results show that nonlinearity is only encountered in predictive control under the FF hypothesis when nonlinear WFSs are considered.

Furthermore, the results indicate that utilizing spatiotemporal correlations increases the prediction accuracy in numerical simulations, reducing variance up to a factor of 3.5 compared with a non-predictive approach, while an uncertain wind profile leads to an improvement of 2.3, which aligns well with the theoretical limits in 48 and on-sky results on Keck II.¹¹ However, as discussed in Ref. 70, it is unclear if these spatiotemporal correlations provide performance gain on real-world data. The predicted performance gain depends on various aspects, such as atmospheric conditions, WFS used, the WFS model accuracy, telescope geometry, and sampling rate.

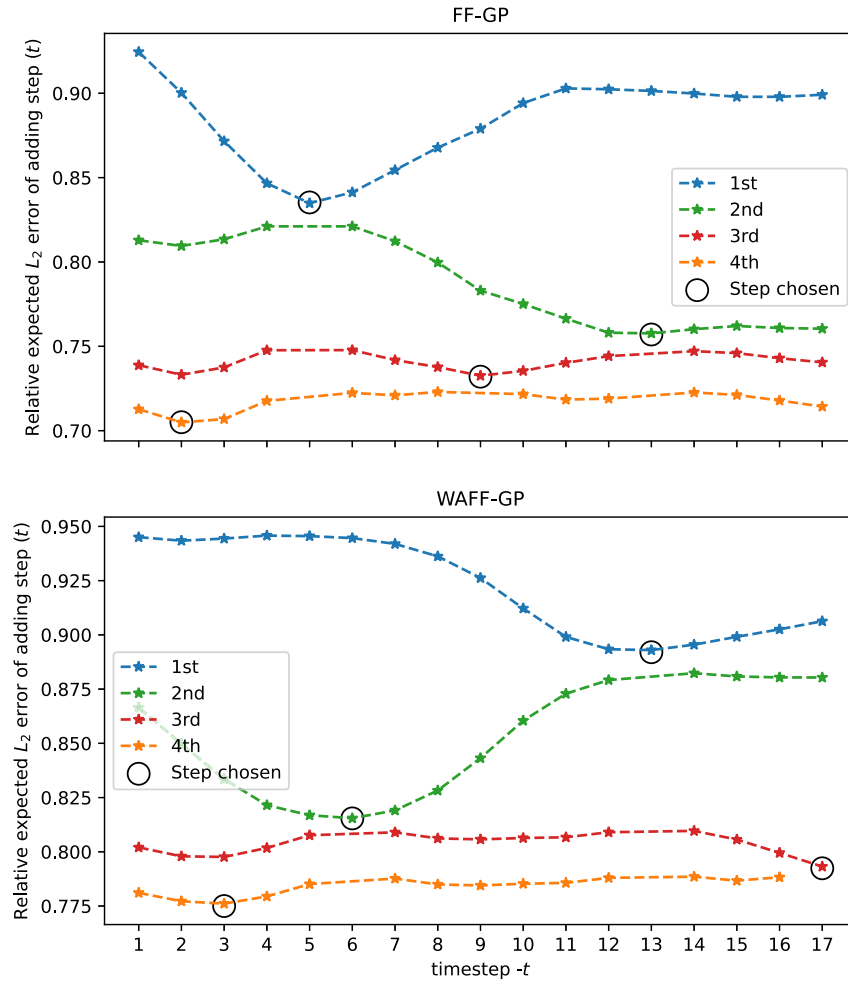


Fig. 7 Optimization process for choosing the most informative timesteps from the past. The upper plot is for the FF model, and the lower plot is for the WAFF model (frame rate 2 ms and S/N = 100). At every optimization step (first, second, third, and fourth), we plot the expected L_2 error of adding the step (t) compared with the expected L_2 error of just including step 0 (relative to the first-order prediction). The black circles indicate the chosen step, which is then omitted in subsequent calculations.

Moreover, if the predictive controller (e.g., empirical orthogonal functions [EOF]) is learned from simulated data, the predicted performance gain depends on the way the data are simulated. For example, the temporal interpolation of the FF turbulence screens dampens higher frequencies from the temporal spectral, leading to smoother, more predictable turbulence. Also, disentangling aliasing error and temporal error using an idealized phase sensor can lead to more optimistic performance gains.

The numerical simulations were conducted with a fairly small 16×16 system. However, since the turbulence is spatially isotropic and the WFS model operates locally, the conclusions translate (approximately) to bigger systems with similar actuator spacing, e.g., SPHERE.

Moreover, our results indicate it is always advantageous—or at least inside the preceding 18 frames and with the tested model parameters—to include more history steps into the model if permitted by the available computational resources. However, at around 16 frames, the gain from including extra timesteps starts to level out slowly. Moreover, we demonstrated that given maximum history length, a sparse sampling of past data over a longer time span leads to better predictions than a sequence of the last consecutive frames. An optimized choice of five history steps provides in the studied setting an additional 10% to 15% improvement in the RMS, indicating that optimized use of history data may also be important for machine learning-based minimum variance predictive control, such as EOF.¹⁰

7.1 Computational Complexity

7.1.1 Non-real-time computations: computing the posterior distribution

To apply the predictive reconstruction matrix in Eq. (13), one must solve for (or be able to operate with) the posterior covariance matrix in Eq. (12). The calculation of the inverse of the dense matrix in Eq. (12) is computationally expensive. For a dense symmetric positive definite matrix of size $n \times n$, a standard matrix inversion method, say, the Cholesky decomposition, has a computational complexity of $\mathcal{O}(n^3)$. Hence, operating with the inverse of a matrix of an 8-m class telescope already requires substantial memory resources, and the computational requirement would be in the order of teraFLOPs or more. As an example, 8-m-class XAO (40×40 DM) systems with $p = 5$ past telemetry require inverting a matrix of $91,524 \times 91,534$ when $S = 4$ and 7752×7752 when the $S = 1$ (Fried grid), if the size of the matrix is optimized without any approximations. Performing such computations efficiently might involve specialized hardware, parallel processing, and optimized algorithms to handle the scale of the problem efficiently, especially when degrees of freedom (DoF) is scaled up to 10^4 , in the next generation of extremely large telescopes.

Since the posterior variance does not depend on the data, the predictive reconstructor in Eq. (13) needs only be updated when the prior information (e.g., the wind direction, wind speed, or r_0) changes. The presumed rate of change in the atmospheric parameters determines the update rate for the posterior covariance.

A connection exists between the method discussed here and EOF, which uses a batch of past data to fit the spatiotemporal covariates of pseudo-open-loop telemetry. If we replace the WFS model with an ideal phase sensor (identity matrix), the EOF prediction matrix will converge to the prediction matrix derived from the FF model on the infinite data limit (stable atmospheric statistics). The EOF is machine learning-based, so it can potentially adapt to more general turbulence statistics, while ST-GP formulation proved a way to deal with aliasing error and study the theoretical limits of predictive control.

7.1.2 Real-time computations: applying the mean prediction matrix

The real-time computation needed for AO control is the mean prediction of the marginal distribution corresponding ϕ^{p+2} , i.e., a matrix multiplication between the past sequence of WFS data and the prediction matrix $R_{\text{FF}/\text{WAF}}^{\text{FF}}$ defined in Eq. (13). The reconstruction matrix shape is the $1292 \times (p \times 2400) = 1292 \times 12,000$, if $p = 5$. Again, the prediction matrix need not be recomputed as long as the atmospheric conditions stay the same. Assuming dense matrices, multiplication with the prediction matrix requires a few hundred GFLOPS of computing bandwidth, which falls comfortably within the capabilities of conventional computers available today.

7.2 Conclusion and Future Work

To conclude, this paper studies the limits of predictive control with spatiotemporal GP models. It discusses how reconstruction errors, such as temporal error, photon noise, and aliasing, can be minimized with predictive control in an optimized manner, given the computational limitation of the hardware used. It also studies how modeling errors, particularly WFS model discretization, affect the quality of the predictive controllers.

The concepts presented in this paper offer several avenues for future research. This work assumes that the knowledge of atmospheric parameters (either the full wind profile or just r_0 and τ_0) is given a prior. However, it would also be possible to incorporate their estimation into the algorithm by modeling them as random parameters and computing a *maximum likelihood* estimate at each measurement step. This is a considerable advantage of the Bayesian approach, including additional sources of uncertainty that can be done systematically. An especially interesting direction would be to see if we can recover FF parameters from on-sky data (e.g., SPHERE telemetry) and utilize them in predictive control.

Also, the application of experimental design to AO could be investigated further, and many other ways of optimizing the measurement could be considered. An interesting approach could be to discretize the data inhomogeneously along the spatial and temporal axes depending on its informativeness. Since older data can be assumed to be less important, one could consider

coarsening the discretization for older timesteps as new data are introduced. Moreover, our approach provides a systematic way to include stochastic vibrations in the model and to use Bayesian OED to ensure that the data relevant for predicting the vibrations are included in the prediction process.

This paper considers a simplified AO design with an SHS sensor operating in an open-loop setup. Theoretically, the method adapts to a closed-loop system via the pseudo-open-loop scheme, which is already provided by some real-time systems working on-sky. However, pseudo-open-loop adaption requires knowledge of hardware/software time lags, DMs' and WFS' response times, as well as calibration errors. Any bias in these reduces the method's performance and can lead to instabilities in the closed-loop control. Although this paper does not explore the effect of the mentioned error sources, it is important to consider them when implementing and deploying the technique on real hardware.

Finally, these concepts can be used to study speckle statistics under predictive control. The predictive control uncertainties could be propagated through a coronagraphic system to give the WDH intensity under optimal predictive control with the given assumption of the atmosphere. Overall, the presented method not only enhances the accuracy and resolution of reconstructions but also opens new avenues for advancing predictive control and reconstruction methodologies and provides a deeper understanding of the limits of given prediction models, e.g., those based on machine learning.

Code and Data Availability

The data and codes used in this paper are available on GitHub in Jupyter Notebook format. The data are completely numerically simulated and produced by the codes. Please visit our GitHub repository [<https://github.com/jnousi/ST-GP4AO.git>] to access the codes. The repository contains Jupyter Notebook files that outline the analysis steps taken in this study. The code is documented and annotated to help readers understand the methodology and reproduce the results.

We encourage readers to use the data and codes for their own research and to cite this paper as the source of the data. If you have any questions about the data or the codes, please do not hesitate to contact us.

Acknowledgments

The work of J.N. and T.H. was supported by the Academy of Finland (decisions 326961, 345720, and 353094). The work of N.H. and J.P. was supported by the Academy of Finland (decisions 348503, 353081, and 359181). We thank Miska Le Louarn for fruitful discussions on the project.

References

1. H. W. Babcock, "The possibility of compensating astronomical seeing," *Publ. Astron. Soc. Pac.* **65**(386), 229–236 (1953).
2. F. Roddier, *Adaptive Optics in Astronomy*, Cambridge University (1999).
3. B. Macintosh et al., "First light of the Gemini Planet Imager," *Proc. Natl. Acad. Sci. U. S. A.* **111**(35), 12661–12666 (2014).
4. J. L. Beuzit et al., "Sphere: the exoplanet imager for the very large telescope," *Astron. Astrophys.* **631**, A155 (2019).
5. J. R. Males et al., "MagAO-X: project status and first laboratory results," *Proc. SPIE* **10703**, 1070309 (2018).
6. N. Jovanovic et al., "The Subaru Coronagraphic Extreme Adaptive Optics system: enabling high-contrast imaging on solar-system scales," *Publ. Astron. Soc. Pac.* **127**(955), 890 (2015).
7. F. Cantalloube et al., "Peering through sphere images: a glance at contrast limitations," *The Messenger* **176**, 25–31 (2019).
8. C. Kulcsár et al., "Optimal control, observers and integrators in adaptive optics," *Opt. Express* **14**(17), 7464–7476 (2006).
9. M. Lloyd-Hart and P. McGuire, "Spatio-temporal prediction for adaptive optics wavefront reconstructors," in *Adapt. Opt. ESO Conf. and Workshop Proc., Proc. Top. Meeting, held October 2-6, 1995, Garching, Germany, Garching near Munich: European Southern Observatory, 1996*, edited by Martin Cullum, Vol. 54, p. 95 (1996).
10. O. Guyon and J. Males, "Adaptive optics predictive control with empirical orthogonal functions (EOFs)," arXiv:1707.00570 (2017).

11. M. A. van Kooten et al., “Predictive wavefront control on Keck II adaptive optics bench: on-sky coronagraphic results,” *J. Astron. Telesc. Instrum. Syst.* **8**(2), 029006 (2022).
12. C. E. Rasmussen et al., *Gaussian Processes for Machine Learning*, Vol. 1, Springer (2006).
13. N. Morujão et al., “Integrated turbulence parameters’ estimation from Naomi adaptive optics telemetry data,” *Astron. Astrophys.* **678**, A193 (2023).
14. J. Lehtonen and T. Helin, “Real-time turbulence profiling using particle filtering,” in *Adapt. Opt. for Extremely Large Telesc.* (2019).
15. Y. H. Ono et al., “Statistics of turbulence parameters at Maunakea using the multiple wavefront sensor data of raven,” *Mon. Not. R. Astron. Soc.* **465**(4), 4931–4941 (2017).
16. B. Neichel et al., “Towards a reliability assessment of the CN2 and wind speed vertical profiles retrieved from gems,” *Proc. SPIE* **9148**, 1969–1978 (2014).
17. J. Osborn et al., “Turbulence velocity profiling for high sensitivity and vertical-resolution atmospheric characterisation with stereo-scidar,” *Mon. Not. R. Astron. Soc.* **464**, stw2685 (2016).
18. D. J. Laidlaw et al., “Automated wind velocity profiling from adaptive optics telemetry,” *Mon. Not. R. Astron. Soc.* **491**(1), 1287–1294 (2020).
19. J. Fowler and R. Landman, “Tempestas ex machina: a review of machine learning methods for wavefront control,” *Techniques and Instrumentation for Detection of Exoplanets XI* **12680**, 100–114 (2023).
20. R. N. Paschall and D. J. Anderson, “Linear quadratic Gaussian control of a deformable mirror adaptive optics system with time-delayed measurements,” *Appl. Opt.* **32**(31), 6347–6358 (1993).
21. M. Gray and B. Le Roux, “Ensemble transform Kalman filter, a nonstationary control law for complex AO systems on ELTs: theoretical aspects and first simulations results,” *Proc. SPIE* **8447**, 84471T (2012).
22. J.-M. Conan et al., “Are integral controllers adapted to the new era of ELT adaptive optics?,” in *AO4ELT* (2011).
23. C. Correia et al., “Adapting optimal LQG methods to ELT-sized AO systems,” in *1st AO4ELT Conf.-Adapt. Opt. for Extremely Large Telesc.*, EDP Sciences, p. 07003 (2010).
24. C. Correia et al., “On the optimal reconstruction and control of adaptive optical systems with mirror dynamics,” *J. Opt. Soc. Am. A* **27**(2), 333–349 (2010).
25. C. M. Correia et al., “Modeling astronomical adaptive optics performance with temporally filtered wiener reconstruction of slope data,” *J. Opt. Soc. Am. A* **34**(10), 1877–1887 (2017).
26. B. Sinquin et al., “On-sky results for adaptive optics control with data-driven models on low-order modes,” *Mon. Not. R. Astron. Soc.* **498**(3), 3228–3240 (2020).
27. P. Massioni et al., “Fast computation of an optimal controller for large-scale adaptive optics,” *J. Opt. Soc. Am. A* **28**(11), 2298–2309 (2011).
28. K. Hinnen, M. Verhaegen, and N. Doelman, “Exploiting the spatiotemporal correlation in adaptive optics using data-driven H2-optimal control,” *J. Opt. Soc. Am. A* **24**(6), 1714–1725 (2007).
29. K. Hinnen, M. Verhaegen, and N. Doelman, “A data-driven H2-optimal control approach for adaptive optics,” *IEEE Trans. Control Syst. Technol.* **16**(3), 381–395 (2008).
30. C. Petit et al., “Linear quadratic Gaussian control for adaptive optics and multiconjugate adaptive optics: experimental and numerical analysis,” *J. Opt. Soc. Am. A* **26**(6), 1307–1325 (2009).
31. B. Le Roux et al., “Optimal control law for classical and multiconjugate adaptive optics,” *J. Opt. Soc. Am. A* **21**(7), 1261–1276 (2004).
32. J. Nousiainen et al., “Adaptive optics control using model-based reinforcement learning,” *Opt. Express* **29**(10), 15327–15344 (2021).
33. J. Nousiainen et al., “Toward on-sky adaptive optics control using reinforcement learning-model-based policy optimization for adaptive optics,” *Astron. Astrophys.* **664**, A71 (2022).
34. J. Nousiainen et al., “Laboratory experiments of model-based reinforcement learning for adaptive optics control,” *J. Astron. Telesc. Instrum. Syst.* **10**(1), 019001 (2024).
35. R. Landman et al., “Self-optimizing adaptive optics control with reinforcement learning for high-contrast imaging,” *J. Astron. Telesc. Instrum. Syst.* **7**(3), 039002 (2021).
36. B. Pou et al., “Adaptive optics control with multi-agent model-free reinforcement learning,” *Opt. Express* **30**, 2991–3015 (2022).
37. B. Pou et al., “Integrating supervised and reinforcement learning for predictive control with an unmodulated pyramid wavefront sensor for adaptive optics,” arXiv:2405.13610 (2024).
38. S. Y. Haffert et al., “Data-driven subspace predictive control: lab demonstration and future outlook,” *Proc. SPIE* **11823**, 1182306 (2021).
39. S. Y. Haffert et al., “Data-driven subspace predictive control of adaptive optics for high-contrast imaging,” *J. Astron. Telesc. Instrum. Syst.* **7**(2), 029001 (2021).
40. B. Recht, “A tour of reinforcement learning: the view from continuous control,” *Annu. Rev. Control Rob. Auton. Syst.* **2**, 253–279 (2019).
41. L. A. Poyneer, B. A. Macintosh, and J.-P. Véran, “Fourier transform wavefront control with adaptive prediction of the atmosphere,” *J. Opt. Soc. Am. A* **24**(9), 2645–2660 (2007).

42. J. R. Males and O. Guyon, “Ground-based adaptive optics coronagraphic performance under closed-loop predictive control,” *J. Astron. Telesc. Instrum. Syst.* **4**(1), 019001 (2018).
43. C. Dessenne, P.-Y. Madec, and G. Rousset, “Optimization of a predictive controller for closed-loop adaptive optics,” *Appl. Opt.* **37**(21), 4623–4633 (1998).
44. R. Swanson et al., “Wavefront reconstruction and prediction with convolutional neural networks,” *Proc. SPIE* **10703**, 107031F (2018).
45. R. Swanson et al., “Closed loop predictive control of adaptive optics systems with convolutional neural networks,” *Mon. Not. R. Astron. Soc.* **503**(2), 2944–2954 (2021).
46. Z. Sun et al., “A Bayesian regularized artificial neural network for adaptive optics forecasting,” *Opt. Commun.* **382**, 519–527 (2017).
47. P. C. McGuire et al., “Adaptive optics: neural network wavefront sensing, reconstruction, and prediction,” in *Sci. Appl. of Neural Nets*, Springer, pp. 97–138 (1999).
48. N. Doelman, “The minimum of the time-delay wavefront error in adaptive optics,” *Mon. Not. R. Astron. Soc.* **491**(4), 4719–4723 (2020).
49. C. M. Correia et al., “Spatio-angular minimum-variance tomographic controller for multi-object adaptive-optics systems,” *Appl. Opt.* **54**(17), 5281–5290 (2015).
50. M. van Kooten, N. Doelman, and M. Kenworthy, “Impact of time-variant turbulence behavior on prediction for adaptive optics systems,” *J. Opt. Soc. Am. A* **36**(5), 731–740 (2019).
51. C. Béchet, M. Tallon, and É. Thiébaud, “Comparison of minimum-norm maximum likelihood and maximum a posteriori wavefront reconstructions for large adaptive optics systems,” *J. Opt. Soc. Am. A* **26**(3), 497–508 (2009).
52. S. Oberti et al., “Super-resolution wavefront reconstruction,” *Astron. Astrophys.* **667**, A48 (2022).
53. A. Berdeu et al., “Inverse problem approach in extreme adaptive optics: analytical model of the fitting error and lowering of the aliasing,” *Proc. SPIE* **12185**, 121850K (2022).
54. C. M. Correia et al., “Super-resolution wavefront reconstruction in adaptive-optics with pyramid sensors,” *Proc. SPIE* **12185**, 121850J (2022).
55. R. Conan, “Mean-square residual error of a wavefront after propagation through atmospheric turbulence and after correction with Zernike polynomials,” *J. Opt. Soc. Am. A* **25**(2), 526–536 (2008).
56. M. Bester et al., “Atmospheric fluctuations-empirical structure functions and projected performance of future instruments,” *Astrophys. J.* **392**, 357–374 (1992).
57. D. Dayton et al., “Atmospheric structure function measurements with a Shack–Hartmann wave-front sensor,” *Opt. Lett.* **17**(24), 1737–1739 (1992).
58. M. S. Belen’kii et al., “Experimental evidence of the effects of non-Kolmogorov turbulence and anisotropy of turbulence,” *Proc. SPIE* **3749**, 50–51 (1999).
59. D. T. Kyraizis et al., “Measurement of optical turbulence in the upper troposphere and lower stratosphere,” *Proc. SPIE* **2120**, 43–55 (1994).
60. B. E. Stribling, B. M. Welsh, and M. C. Roggemann, “Optical propagation in non-kolmogorov atmospheric turbulence,” *Proc. SPIE* **2471**, 181–196 (1995).
61. G. Lombardi et al., “Surface layer characterization at Paranal observatory,” *Proc. SPIE* **7733** 77334D (2010).
62. T. Helin et al., “Atmospheric turbulence profiling with unknown power spectral density,” *Inverse Prob.* **34**(4), 044002 (2018).
63. A. Tokovinin, “The elusive nature of “seeing,”” *Atmosphere* **14**(11), 1694 (2023).
64. C. R. Vogel, “Time-varying stochastic turbulence model,” *Proc. SPIE* **6272**, 982–986 (2006).
65. G. Yuan et al., “Time-varying stochastic turbulence model considering intermittency,” in *Int. Workshop on Chaos-Fractals Theor. and Appl.*, IEEE, pp. 126–129 (2009).
66. L. Poyneer, M. van Dam, and J.-P. Véran, “Experimental verification of the frozen flow atmospheric turbulence assumption with use of astronomical adaptive optics telemetry,” *J. Opt. Soc. Am. A* **26**(4), 833–846 (2009).
67. K. Chaloner and I. Verdinelli, “Bayesian experimental design: a review,” *Stat. Sci.* **10**, 273–304 (1995).
68. M. Burger et al., “Sequentially optimized projections in X-ray imaging,” *Inverse Prob.* **37**, 075006 (2021).
69. E. H. Por et al., “High Contrast Imaging for Python (HCIPy): an open-source adaptive optics and coronagraph simulator,” *Proc. SPIE* **10703**, 1070342 (2018).
70. M. Van Kooten, N. Doelman, and M. Kenworthy, “Robustness of prediction for extreme adaptive optics systems under various observing conditions—an analysis using VLT/SPHERE adaptive optics data,” *Astron. Astrophys.* **636**, A81 (2020).

Jalo Nousiainen is a postdoctoral researcher at Aalto University. His research focuses on algorithm development for high-contrast instruments dedicated to exoplanet imaging. He has a background in applied mathematics, specifically in inverse problems, Bayesian statistics, and machine learning.

Juha-Pekka Puska is a doctoral researcher at Aalto University. His research is on optimal experimental design methods and their applications to imaging problems.

Markus Kasper received his PhD from the University of Heidelberg in 1999, after which he joined ESO to handle on adaptive optics, high-contrast imaging, and astronomical instrumentation projects. He was involved in many of ESO's AO projects, such as MACAO and NACO LGS, and was ESO's project leader for VLT-SPHERE. He was the principal investigator of the ELT/EPICS phase-A study (2008–2010, predecessor of PCS) and of NEAR, the mid-IR imaging experiment to search for habitable planets in Alpha Centauri in collaboration with the Breakthrough Initiatives. He is currently starting the ELT-PCS planet imager project.

Biographies of the other authors are not available.