

# Forecasting wavefront corrections in an adaptive optics system

Rehan Hafeez<sup>①</sup>,<sup>a,b,\*</sup> Finn Archinuk,<sup>a,c</sup> Sébastien Fabbro<sup>①</sup>,<sup>a,c</sup>  
Hossen Teimoorinia,<sup>a,c</sup> and Jean-Pierre Véran<sup>①</sup>,<sup>a,c</sup>

<sup>a</sup>National Research Council Canada, Victoria, British Columbia, Canada

<sup>b</sup>University of British Columbia, Vancouver, British Columbia, Canada

<sup>c</sup>University of Victoria, Victoria, British Columbia, Canada

**Abstract.** We used telemetry data from the Gemini North ALTAIR adaptive optics system to investigate how well the commands for wavefront correction (both tip/tilt and high-order turbulence) can be forecasted to reduce lag error (due to wavefront sensor averaging and computational delays) and improve delivered image quality. We showed that a high level of reduction ( $\sim 5$  for tip-tilt and  $\sim 2$  for high-order modes) in the RMS wavefront error was achieved using a “forecasting filter” based on a linear autoregressive model with only a few coefficients ( $\sim 30$  for tip-tilt and  $\sim 5$  for high-order modes) to complement the existing integral servo-controller. Updating this filter to adapt to evolving observing conditions is computationally inexpensive and requires  $< 10$  s of telemetry data. We also used several machine learning models (long-short term memory and dilated convolutional models) to evaluate whether further improvements could be achieved with a more sophisticated nonlinear model. Our attempts showed no perceptible improvements over linear autoregressive predictions, even for large lags in which residuals from the linear models are high, suggesting that nonlinear wavefront distortions for ALTAIR at the Gemini North telescope may not be forecasted with the current setup. © *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.8.2.029003](https://doi.org/10.1117/1.JATIS.8.2.029003)]

**Keywords:** adaptive optics; machine learning; forecasting; ALTAIR; atmospheric distortion.

Paper 21149G received Nov. 22, 2021; accepted for publication Apr. 29, 2022; published online May 19, 2022.

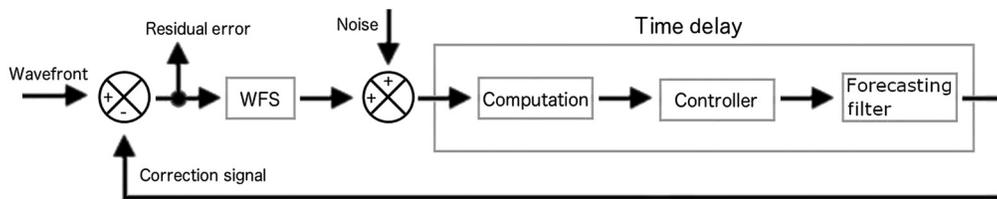
## 1 Introduction

Atmospheric turbulence causes random distortions in incoming light, putting regions of the incoming wavefronts out of phase. When an object is imaged through atmospheric turbulence, the global tip and tilt of the wavefront make the object appear to move. In contrast, higher-order wavefront distortions significantly degrade the image’s resolution. An adaptive optics (AO) system aims to reduce the effect of atmospheric turbulence by calculating these distortions and correcting them in real-time using a deformable mirror (DM), complemented by an additional mirror dedicated to stabilizing the image called a tip/tilt mirror (TTM). Image stabilization is especially critical as vibrations originating within the observatory can significantly add to atmospheric turbulence-induced tip and tilt.

AO systems have now been deployed in many ground-based astronomical observatories as the improvement in angular resolution and sensitivity afforded by AO can very significantly improve their scientific output. Many of these systems work in a closed loop, similar to the system shown in Fig. 1, with a wavefront sensor (WFS) measuring the residual wavefront after correction and a real-time controller (RTC) updating the DM and the TTM to keep the residual wavefront as small as possible. WFS measurements are obtained by continuously recording images (frames) on a high-speed camera, and the integration time for each image is typically 1 ms. The key to the performance of the AO system is the time it takes to acquire the WFS measurement and for the RTC to transfer these images and process them into a set of command signals to control the DM and the TTM. This delay, which is typically also on the order of 1 ms,

---

\*Address all correspondence to Rehan Hafeez, [rhafeez@student.ubc.ca](mailto:rhafeez@student.ubc.ca)



**Fig. 1** Block diagram of a standard AO system, including a forecasting filter to reduce lag error caused by computational delay.

causes the command signals to be slightly stale, resulting in a correction error called a lag error. If the atmospheric turbulence could be forecast within this millisecond time scale, then the lag error could be reduced. Reducing the lag error could greatly improve the imaging performance of AO systems, especially those aimed at faint imaging companions, such as exoplanets in the close vicinity of a bright parent star. For such systems, the lag error is often the main contributor to the loss of contrast due to residual atmospheric turbulence.<sup>1</sup>

### 1.1 Forecasting Filter

In most existing systems, the servo-controller is a simple integrator with a gain that can be adjusted to make the best possible compromise between increasing the rejection bandwidth (high gain) and reducing the propagation of WFS noise into the command signals (low gain).<sup>2</sup> In this architecture, no attempt is made at predicting the command signals into the future to reduce the lag error (an integrator can be seen as a zero-order predictor). However, it is reasonable to expect a significant level of correlation between realizations of the atmospheric turbulence separated by only 1 ms. Therefore, this correlation can be used to attempt to increase the accuracy of the command signals by forecasting. Several methods have already been proposed to explore this idea with predictive controllers, either based on statistical models of atmospheric turbulence or being purely data-driven. See, e.g., the recent work presented on data-driven predictive control<sup>3</sup> and the references therein.

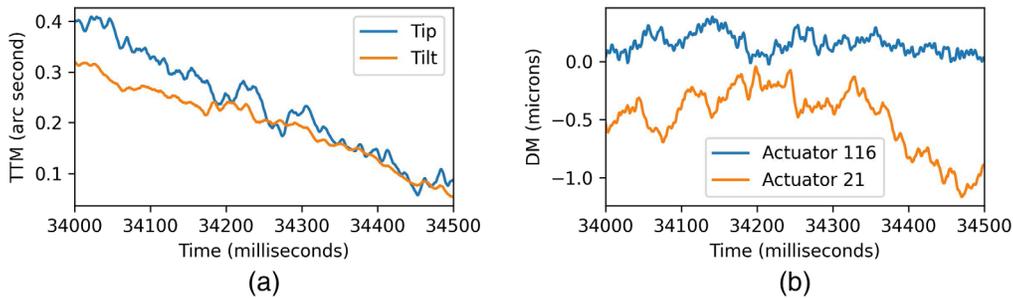
We propose adding a forecasting filter to the existing integral control that takes command signals computed from previous frames and forecasts the command for a future frame. We expect that such a forecasting filter would be easy to retrofit as an extension of conventional AO RTCs, in which the current forecasting filter is implicitly the identity operator. Figure 1 shows where in the control system the filter would be placed. An ideal forecasting filter would itself be computationally minimal to keep the additional time delay small. For example, the computational complexity of the forecasting filter could be compared with that of wavefront reconstruction.

Many previous studies on AO control rely on testing with data obtained by end-to-end AO simulation software implementing a statistical model of the atmospheric turbulence (e.g., Ref. 4) or by laboratory experiments using artificial turbulence (e.g., Ref. 3). Instead, we design and evaluate our forecasting filter using a purely data-driven approach based on on-sky AO telemetry data, as outlined in Sec. 2. AO telemetry data are usually readily available in operating AO systems and have been successfully used in a few previous studies.<sup>5–7</sup> In Sec. 3, we examine the command signal as a time series, and in Sec. 4, we outline the linear forecasting models.

In Sec. 5, we apply these models to the data and approximate parameters such as how many previous frames are required and what levels of improvement are attainable. Finally, in Sec. 6, we explore more complex, nonlinear models for forecasting using neural networks.

## 2 Data

We use AO telemetry data collected on ALTAIR at the Gemini North Telescope between August 2007 and March 2008. Each data set was obtained on a different date, at the beginning of the night, during the daily “M1 tuning” procedure, in which the AO system is locked on a bright natural guide star (NGS) near zenith, and persistent wavefront errors corrected by the AO system



**Fig. 2** Command signals as time series data sampled from an M1 tuning. (a) Tip and tilt commands for image stabilization. (b) Commands to two of the 136 active channels of the DM for high-order correction.

are offloaded to the primary mirror. The first 20 Zernike modes are continuously offloaded (tip-tilt, defocus, and coma are actually offloaded to M2) at slow speeds ( $<1$  Hz), so the offload process does not affect the millisecond time scale forecasting that we are trying to achieve here. Seeing values for these data are not available; however, it can be expected that they span a range of around 0.5 arcsec, which is the typical median seeing at this site. In all cases, the AO system operated at a frame rate of 1 kHz, which means that one WFS measurement was obtained and one command vector signal was sent every millisecond. Each telemetry data set, called “session” hereafter, contains a time series of various real-time data processed by the AO systems, including WFS images, WFS slopes (gradients), reconstructed wavefront errors, and DM/TTM commands. For this work, we only use the DM/TTM commands.

In NGS mode, ALTAIR uses a  $12 \times 12$  Shack–Hartmann WFS to drive a 177-actuator DM plus one TTM dedicated to image stabilization, at a rate of up to 1 kHz.<sup>8</sup> Only 136 actuators in the DM are actively controlled; the remaining actuators, located at the edge of the DM, are simply extrapolated. Commands to the DM actuators are recorded in a 177-element vector in units of microns of wavefront correction. TTM commands are recorded in a two-element vector in units of arcseconds of image motion on sky. Figure 2 shows samples of both types of command signals over 500 ms. For an 8-m telescope, 1 milliarcsecond of image tip or tilt is roughly equivalent to 9.7 nm RMS of wavefront error.

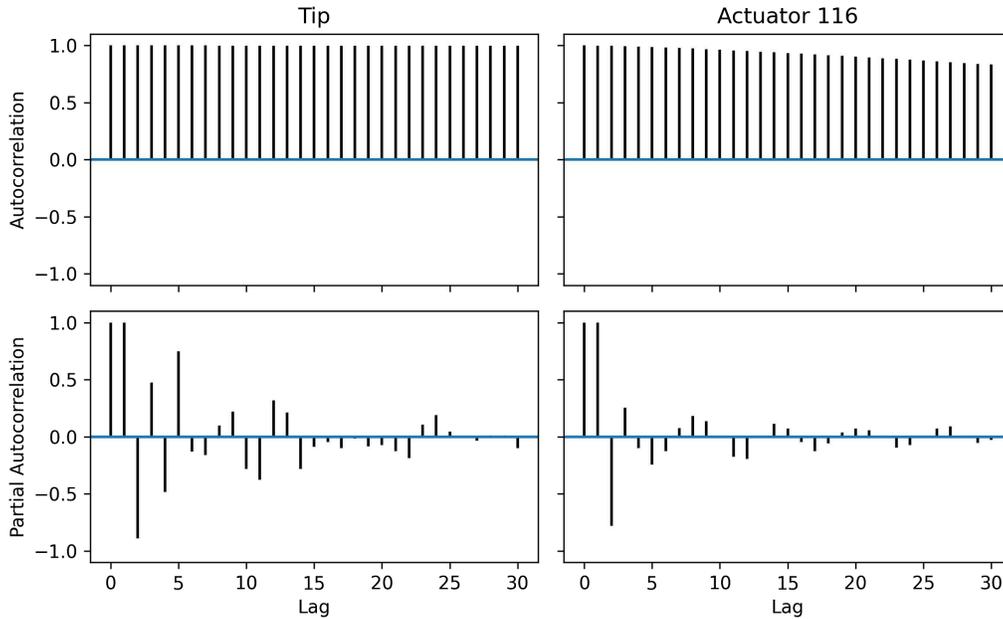
The session length was limited by the buffer capacity; therefore, each session contains  $<60$  s of data. We chose sessions with at least 50 s of data. Because these command values are recorded from a real system and analyzed after the fact, we are also limited to evaluating the predictability of these commands, that is, we were not able to test the stability of the system if the proposed forecasting filter was implemented.

Finally, because for all these sessions the NGS is very bright, the AO loop operates at the maximum possible gain. The correction residuals are largely dominated by the lag error, and we can therefore assume that the DM and TTM command signals accurately track the incoming atmospheric turbulence with a simple delay. Therefore, for a single frame delay, the optimal command signals at time  $t + 1$  are the command signals that were computed by the system for that frame.

### 3 Time Series Analysis

We performed exploratory data analysis on a session to understand their general dynamics. The first column in Fig. 3 shows the autocorrelation and partial autocorrelation for the tip channel for an entire 60-s session. Actuator 116 is shown in the second column and was selected because it is neither in the center nor toward the edge.

As seen in the autocorrelation row in Fig. 3, there is strong correlation between commands over a long time horizon. The row of partial autocorrelations suggests that the majority of the signal can be accounted for using only the most recent commands, which is consistent with an autoregressive process. Although these partial autocorrelations become less important, they do



**Fig. 3** Time series analysis of the tip channel and actuator 116. The top row shows autocorrelations. Partial autocorrelation shows the impact of each relative frame for forecasting.

not completely disappear. This indicates that we can further extend the window of our model for marginal improvement.

In time series analysis, lag is the term used for the number of previous observations. Because lag could be mistaken for the error caused by computation time in an AO system, we introduce the term Lookback when referring to the number of previous commands the model can access.

## 4 Models

Our baseline models for both types of data (TTM and DM) are autoregressive models. Forecasts are expressed by the following equation:

$$\hat{y}_{t+1}^{AR} = \sum_{i=0}^{i=n} w_i y_{t-i} + b,$$

where the autoregressive forecast ( $\hat{y}_{t+1}^{AR}$ ) is dependent on the relative time index ( $t$ ), the Lookback number ( $n$ ), a constant bias term ( $b$ ), and fixed parameters of the model ( $w$ ).

As a metric for evaluating our models, we use the root mean square error (RMSE) between the next frame forecast  $\{\hat{y}_{t+1}\}$  and the actual observed ground truth  $\{y_{t+1}\}$ . Our baseline is the current ALTAIR system, which we call Echo, in which the absence of any forecasting returns the full lag error, that is,  $\hat{y}_{t+1}^{\text{Echo}} = y_t$ .

### 4.1 Image Stabilization

Image stabilization is controlled using TTM. Dynamic image motion is caused by atmospheric turbulence as well as by other effects within the telescope, such as wind shake and vibrations caused by the electromechanical systems.

Our aim is to make forecasts for the tip and tilt channels separately. We define two types for forecasting; models of the first type use a single channel and assume that tip and tilt are independent, and the other allows for linear interactions between both channels.

In the independent type forecasting, a single channel is used as input to forecast the command corresponding to that channel. Models using the dependent type of forecasting have access to

both channels and therefore double the number of parameters needed to define them compared with the independent type. For example, a  $\text{Single}_{\text{tip}}$  model trained with a Lookback of 50 will have 50 parameter coefficients; a  $\text{Double}_{\text{tip}}$  model with the same Lookback will have 100 parameter coefficients: 50 derived from each of the tip and tilt channels while making a forecast for the tip channel. The model parameters are found with ordinary least squares using a small amount of data from the session.

## 4.2 High-Order Turbulence Mitigation

The actuators of the DM are arranged in a square lattice with an annular shape. The uncontrolled actuators are located around the edge with one in the center of the array (caused by telescope central obscuration). For simplicity, we use separate linear models for each of the 136 controllable actuators. These 136 separate forecasts are compiled into a single forecast for the DM.

We also examine the quality of forecasts using modes of the large-scale patterns of deformation in the mirror, such as Zernike modes. Low order modes, such as defocus and astigmatisms, contain most of the variance of the atmospheric turbulence and have a higher signal-to-noise ratio than higher order modes. We see in Sec. 5.2.2 that the amount of Lookback required for a given improvement may decrease with the mode index; however, the method that we present here keeps the amount of Lookback consistent for each mode.

We use the ALTAIR system's modal basis, which spans the entire wavefront subspace spanned by the 136 controlled DM actuators. This orthonormal modal basis is constructed such that each mode fits as closely as possible the corresponding Zernike polynomial. Converting into modal space is done via a  $136 \times 136$  conversion matrix, which is available in an ALTAIR configuration file. Again, one model was fitted to each mode, meaning that the forecast requires combining many small models together before being converted back into actuator commands. The benefit of forecasting commands using modes is that a few modes can be used to forecast the majority of the variation while ignoring the noisier high order modes. This is more important where computational limits exist and implementing one forecasting filter per DM actuator is not feasible. Also, the ALTAIR modes are not rigorously statistically independent; however, the low order modes are very close to the Karhunen–Loeve modes for the Kolmogorov turbulence, making the underlying assumption of statistical independence reasonable.

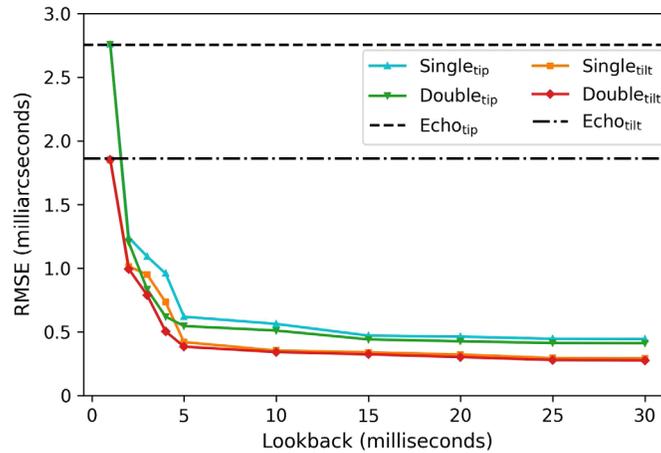
## 5 Results

Our results are separated into two parts: TTM forecasting and DM forecasting. Figures in this section use data collected on January 11, 2008 (11Jan08), unless otherwise noted. We show that this is a representative session because, although specific wavefront corrections may vary across different dates, the general trends in improvement remain consistent.

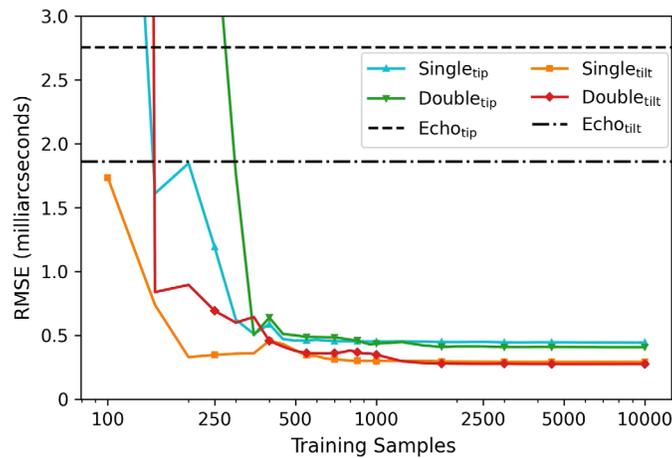
### 5.1 Tip and Tilt Forecasting

Lookback is the number of previous frames that we give a model when fitting. This AO system uses a 1-kHz frame rate, so a Lookback of 20 corresponds to 20 ms. Each model was evaluated with a 90/10 train/test split. Figure 4 shows that increasing the Lookback reduces the modeling error. Our baseline of using no forecasting filter (Echo) is shown as a dashed horizontal line for each channel. These values are stationary because the Echo forecast is equivalent to having a Lookback of 1. We see that most of the improvement is achieved for a Lookback as low as 5 and using a Lookback greater than 30 has marginal improvements. Moreover, using both channels for forecasting, a slight reduction in error can be achieved.

Given a Lookback of 30, we then test how many samples are required to fit the models. To obtain these results, the test set was held constant as the final 10% of the session, and the samples used for fitting the models immediately precede the test set. Therefore, using fewer samples does not introduce a delay between the two sets of data, so the models can be directly compared. Figure 5 shows the RMS tip-tilt error (RMSE) as a function of training data. Each point shows a model fit with the amount of training samples and evaluated on the test set. This figure shows



**Fig. 4** Forecasting error as a function of Lookback.

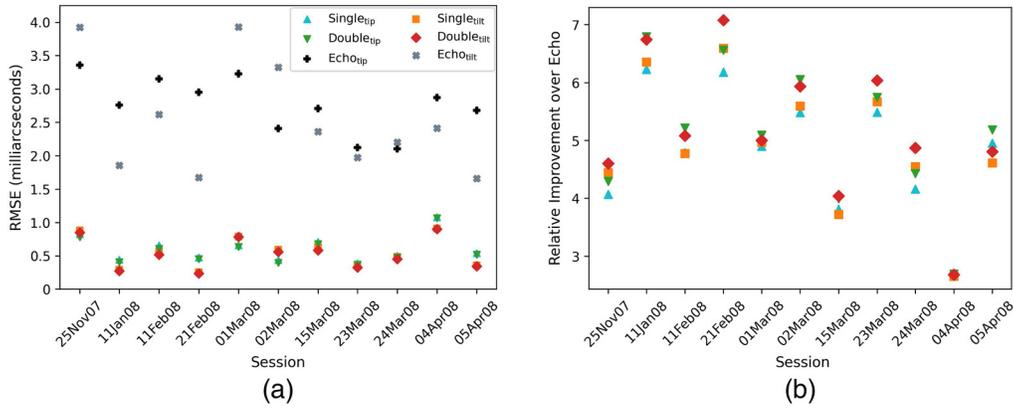


**Fig. 5** Forecasting error as function of the number of training samples.

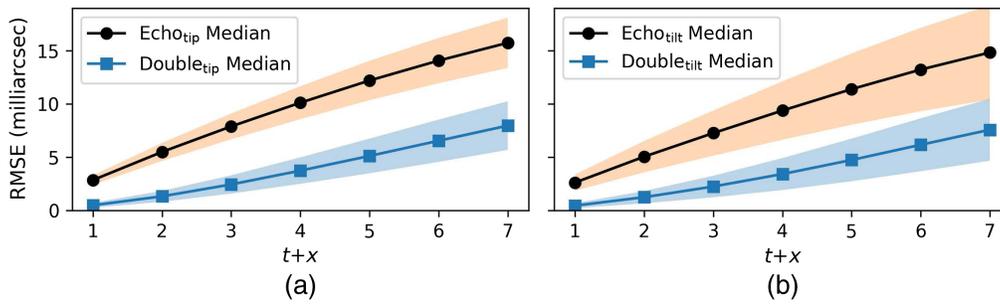
that 10,000 samples are more than sufficient, representing 10 s of data.  $\text{Double}_{\text{tip}}$  models require more samples for training as they have more degrees of freedom. Although we use 10,000 samples for training going forward, similar results can be achieved using fewer samples.  $\text{Echo}_{\text{tip}}$  and  $\text{Echo}_{\text{tilt}}$  provide a reference RMSE of these channels for this part of the session.

Figure 6 shows the improvement of our models compared with Echo for multiple sessions. We fit models for each channel for each session and evaluated them on a testing set composed of the final 10% of the session. Models were fit with a Lookback of 30, with 10,000 samples. Figure 6(a) shows significant improvements over Echo. Figure 6(b) shows a factor improvement of our method above Echo and more clearly illustrates that the  $\text{Double}_{\text{tip}}$  models consistently perform better than the  $\text{Single}_{\text{tip}}$  models.

Forecasting for delays larger than one frame increases the forecasting error. However, in Fig. 7, we can see that our methodology still improves the lag error compared with the Echo baseline. This trend is observed over multiple sessions, with increased variation observed at larger time delays. The ability to forecast for longer time delays is particularly interesting as it opens the possibility of “slowing down” the AO system to increase the exposure time on the WFS, allowing the observer to use fainter sources and therefore increasing the portion of the sky available for NGS AO observations.<sup>9</sup>



**Fig. 6** Forecasting error for different sessions. (a) Full vertical range and (b) relative improvement over Echo.



**Fig. 7** TTM forecast error as a function of AO time delay, expressed in number of frames for (a) tip and (b) tilt channels. The median values are calculated from 32 sessions, and the filled area is one standard deviation.

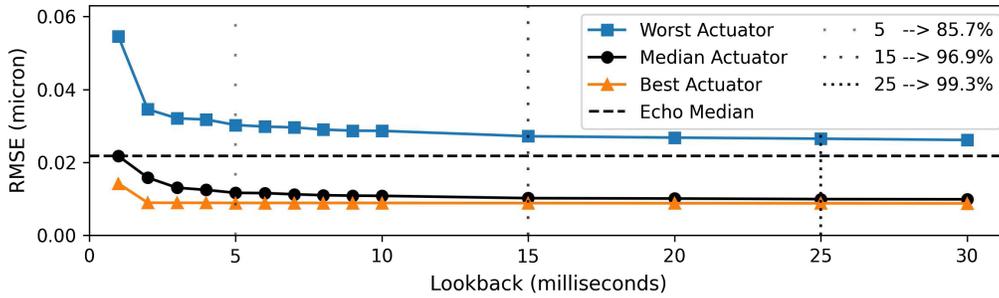
## 5.2 High-Order Turbulence Forecasting

For the high-order turbulence corrected by the DM, we test our linear forecasting model in actuator space (Sec. 5.2.1) and in modal space (Sec. 5.2.2). These models are the same linear autoregressive models outlined in Sec. 4.

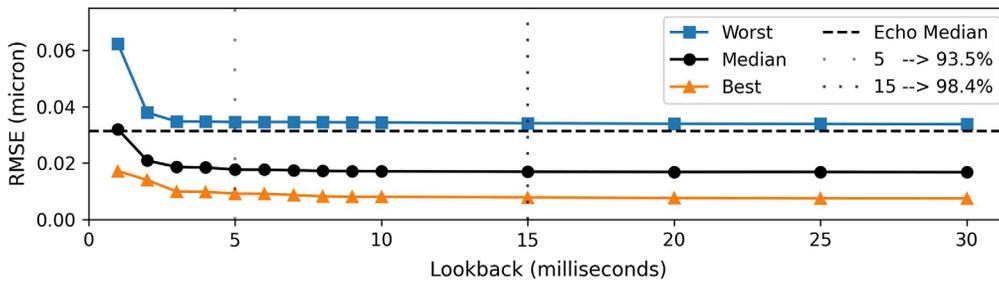
### 5.2.1 Actuator space forecasting

In actuator space, we use raw actuator commands to forecast subsequent actuator commands. We test models with a Lookback of up to 30 ms on a single session. Figure 8 shows that the forecast improvement plateaus beyond a Lookback of 20 and that over 80% of the improvement can be captured with a Lookback of 5 for this session. The distribution of actuator errors is a function of where they are located in the mirror, with central and edge actuators accounting for the highest improvements. This makes sense because these actuators are controlled by fewer and noisier WFS subapertures and therefore carry more measurement noise. Because of this variation, we show the best and worst performing actuator forecasts as well as the median. The best and worst actuators are selected by their forecasting quality at Lookback = 1. Vertical dashed lines show the improvement of the median relative to its lowest observed RMSE, which occurs at Lookback = 30.

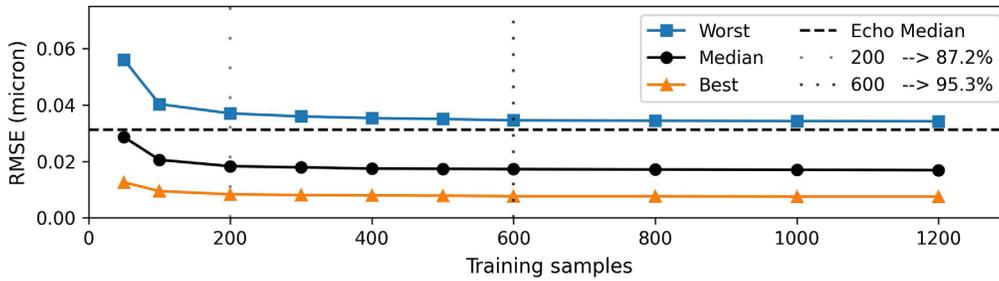
To illustrate that 11Jan08 is a representative session of the range of turbulence conditions, we show in Fig. 9 the trends for multiple sessions. Figure 9 shows the median session forecast, quality and we plotted the best, worst, and median session qualities. Although there is left-over variability within the RMSE values for each session, the trends across sessions are comparable. To help with visual clarity, we only show the median Echo.



**Fig. 8** Forecasting error as a function of Lookback for high order turbulence using actuator commands as inputs. Each session has 136 actuators; shown here are the two extremes and the median forecasting quality. Vertical dashed lines show the relative improvement of the median compared with the best observed improvement (i.e., at Lookback = 30).



(a)



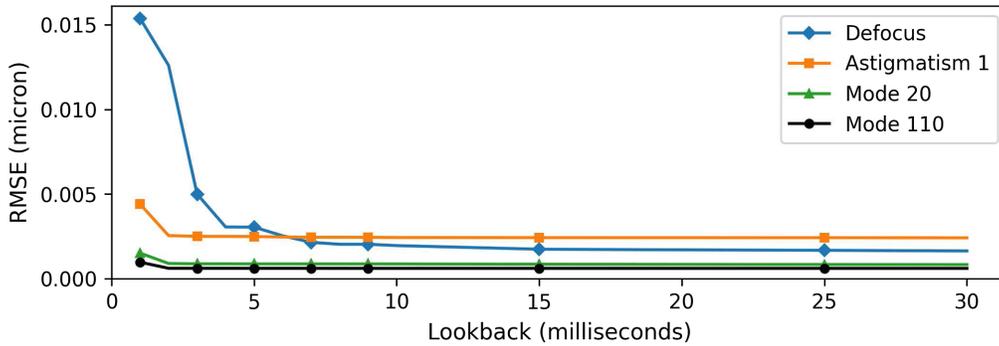
(b)

**Fig. 9** Forecasting error of multiple sessions as a function of (a) Lookback and (b) number of training samples for high-order turbulence using actuator commands as inputs. Each session is summarized by its median performance, and we plotted the best, worst, and median sessions. Vertical dashed lines show improvement relative to the best observed median improvement. (a) 10,000 samples were used for training and (b) Lookback of 20 was used.

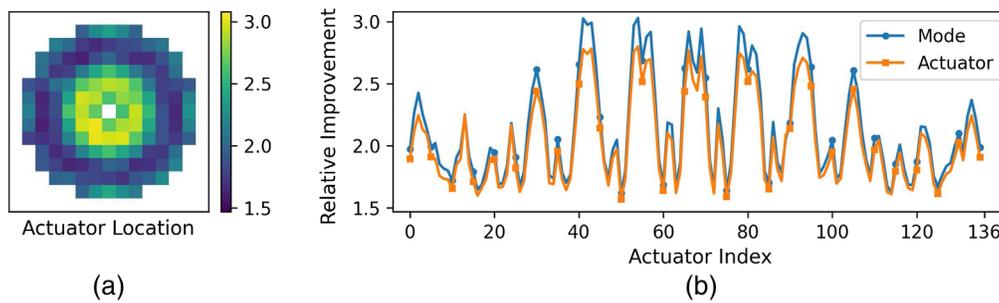
### 5.2.2 Modal space forecasting

In modal space, actuator commands are converted into modal commands with a conversion matrix. The models in this section then take modal inputs and forecast the modal representation of commands for the subsequent frame. Unlike actuator data, amplitudes vary between modes. As shown in Fig. 10, defocus (the first high-order mode) largely dominates, with amplitudes of subsequent modes decreasing as the mode number increases, as expected in a Kolmogorov turbulence. For all modes except defocus, we find that the forecasting error plateaus within a Lookback of 5. Due to its much higher amplitude, defocus benefits from a longer Lookback.

Forecasting using either modes or actuators provides an improvement over Echo. We compare each method by looking at the relative improvement over the Echo baseline. First, we convert the modal forecast into actuator space. For each actuator, we derive a single value score



**Fig. 10** Forecasting error as a function of Lookback in modal space. Four modes are selected for plotting. Defocus is mode 2, astigmatism 1 is mode 3. 10,000 samples were used for training.



**Fig. 11** Comparison of DM forecast by method. (a) Relative improvement of modal forecasts mapped to actuator locations over the Echo forecasts. The central pixel is occluded by the telescope. (b) The relative improvement over an Echo forecast at each actuator by each method.

representing the relative improvement by dividing the RMSE of the Echo forecast by the RMSE of the respective method. Figure 11(a) shows where on the actuator map these improvements occur. These values are from the modal forecast, but the actuator forecast has a nearly identical distribution. The center pixel is blank as it is obscured by the telescope. In Fig. 11(b), the relative improvement by each method is unraveled into an actuator index. Forecasting in modal space results in a lower forecasting error compared with forecasting actuators directly. Because forecasting using modes can influence the entire array, we find that ~50% of our observed improvement is from the first five modes starting with defocus. Therefore, the computational requirement could be reduced if needed by forecasting only a limited number of modes. It is worth noting that here we have restricted the system to one filter per actuator (or per mode). Grouping neighboring actuators using a spatial filter as explained in Ref. 6 might allow for taking advantage of local spatial correlations in the wavefront and thus improve the actuator space forecasting.

## 6 Residual Analysis

Although our proposed autoregressive models provide a significant improvement over Echo, we still observe an increasing error when forecasting further into the future (Fig. 7). This observation is consistent for both the tip/tilt commands and high-order modes. We expect that some nonlinear interactions might be present in the time series, which the autoregressive models lack the capacity to capture. These nonlinear interactions would be present in the residual error of our linear forecasts. In this section, we explore whether we could take advantage of these nonlinear interactions to improve our forecasts using neural network (NN) models.

Using NN architectures to forecast the atmospheric turbulence in real time has already been explored;<sup>4,10</sup> however, as far as we know, none of these approaches have been tested on sky.

Also, these studies do not explicitly separate linear and nonlinear forecasting components, making it difficult to evaluate the real benefit of implementing a more complex modeling approach.

As not all NN architectures can take advantage of the time ordering in the data, we focus our attention on three NN architectures designed to incorporate states while processing the next inputs in the sequence. We settled on three specific recurrent and convolutional architectures, reviewed in a recent survey by Lim et al. of machine learning algorithms for time series forecasting.<sup>11</sup> We limit our analyses to the tip channel for TTM and to mode 22 for DM. We focus our analysis on residuals from forecasts at  $t + 1$  to provide a direct comparison to our linear model and at  $t + 7$  to reasonably maximize the error from nonlinear effects.

## 6.1 Methodology

We define the residual error between linear forecasts and true actuator commands at  $t + \tau$  as a new time series, where  $\tau$  represents the forecasted time. The NNs take actuator commands from previous frames and fits them to the residual of the corresponding linear forecast at  $t + \tau$ . This method extracts any potential nonlinear trends from the actuator commands that the linear model is unable to capture. The time series residual is extracted from the 11Jan08 session. For the tip channel analysis, we use forecasts made by the  $\text{Double}_{\text{tip}}$  model, which is fit with a Lookback of 30. Similarly, for mode 22, we use forecasts made with a linear model fit with a Lookback of 30. These forecasts are subtracted from the true actuator commands, resulting in a time series with  $\sim 60,000$  frames. The NNs are trained using the first 90% of the residual time series.

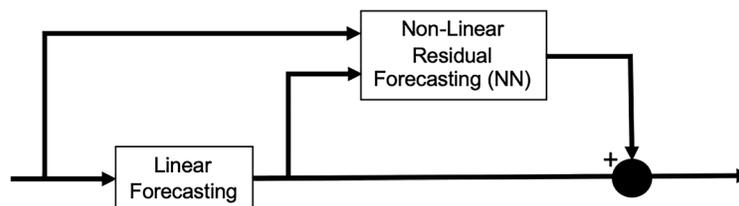
We generate a corrected actuator forecast by adding the residual forecasted by the NNs to the corresponding linear forecast. If the corrected forecast yields a significantly lower RMSE than the corresponding autoregressive forecast alone, then we can estimate the degree of nonlinearity within the time series. Such information could inform the development of more sophisticated models that could be implemented in real time. For this analysis, the new “forecasting filter” in Fig. 1 is extended to include a “nonlinear residual forecasting filter” as shown in Fig. 12.

## 6.2 Neural Network Architectures

We limit our evaluation of NN architectures to three types, following Lim et al.,<sup>11</sup> (1) recurrent NNs, implemented as a long short-term memory (LSTM) network; (2) attention mechanisms, implemented as an LSTM with an attention layer; and (3) dilated convolutional networks, implemented as the WaveNet architecture.

Both (1) and (2) rely on LSTM units, which are regularly used to model stateful sequences due to their ability to handle long-range dependencies<sup>12</sup> by incorporating a hidden state and a gated memory cell from the previous inputs. Whereas traditional sequence-to-sequence architectures based on recurrent NNs need the whole sequence, ultimately keeping only the last compressed state, the attention mechanism<sup>13</sup> adds the extracapacity to look (attend) at all previous hidden states of the sequence in the Lookback window, assigning weights to the most relevant ones for the prediction.

The WaveNet architecture was first developed as a generative model for raw audio<sup>14</sup> and is based on the convolutional NN handling the whole input sequence, as well as masking the future inputs to preserve the temporal structure from the input data. This popular network architecture has been shown to perform well not only in audio generation but also in language processing and time series forecasts.<sup>11</sup>



**Fig. 12** Block diagram of a forecasting filter with linear and residual forecasting filter.

**Table 1** Summary of NN architectures and performance compared with linear baseline.

Model	$t + 1$		$t + 7$	
	Tip RMSE (mas)	Mode 22 RMSE (nm)	Tip RMSE (mas)	Mode 22 RMSE (nm)
Simple LSTM	0.414 ± 0.002	0.830 ± 0.0003	8.681 ± 0.006	6.147 ± 0.007
Attention LSTM	0.417 ± 0.006	0.831 ± 0.0004	<b>8.672 ± 0.007</b>	<b>6.143 ± 0.002</b>
Wavenet	0.516 ± 0.137	0.830 ± 0.0001	8.699 ± 0.026	6.150 ± 0.001
Linear baseline	<b>0.411</b>	<b>0.830</b>	8.680	6.157

Note: Bold values represent results with the lowest error within each column.

Each architecture went through a process of hyperparameter optimization by training each model with a range of values, and in each case we used a 90/10 validation split. For (1) and (2), we found a stable loss plateau after 500 epochs using a batch size of 2048, with the Adam optimizer set with a learning rate of 1e-5 for the tip channel analysis and 1e-6 for mode 22. In the case of (3), the same loss plateau is achieved using a batch size of 1024, with the Adam optimizer set to a learning rate of 1e-5 after 25 epochs.

### 6.3 Results

In Table 1, we present the results of the NN correction for tip and mode 22 for both  $t + 1$  and  $t + 7$ , which we derive by adding the forecasted residual to the corresponding linear model forecast. From this, we calculate the RMSE of the corrected forecasts against the true actuator values. Due to the stochastic nature of the machine learning algorithms used, we present the mean value of the RMSE along with the standard deviations derived from 10 different seed values on the same session. For comparison, we also present the RMSE of the baseline linear model without any nonlinear correction. Because the linear baseline is deterministic and we are using the same session to isolate the NN initialization variation, the linear baseline does not have error bars.

In most cases, the corrected forecasts made by the NN models approach do not exceed the performance of the autoregressive models. When the corrected forecasts exceed the performance of the autoregressive model, the improvements are very marginal (percent of a mas or percent of a nanometer) and come at a great computational cost. We find no evidence that additional information can be practically extracted from these models to aid in forecasting.

## 7 Conclusion and Discussion

In this paper, we have used on-sky AO telemetry data acquired on ALTAIR at Gemini North to show that a simple data-driven autoregressive forecaster is very efficient at forecasting AO correction to mitigate latency of one or more frames. This seems to be especially true for tip and tilt, possibly because these signals have sinusoidal vibration components that are easy to model. Over several different nights, we have also found that the tip/tilt RMSE could be improved by a factor of  $\sim 5$  by forecasting one frame ahead, using a Lookback of 30 frames. However, increasing Lookback time continues to marginally improve forecasting accuracy because lower temporal frequencies can be better modeled, at the cost of increasing the complexity of the forecaster (i.e., number of operations to execute in real time). We have also found that jointly forecasting tip and tilt results in slightly better results than considering the two channels independently.

For high-order modes (defocus and up), we have found that one frame ahead linear forecasting can reduce the RMSE by a factor of  $\sim 2$ . We have found that forecasting in modal space resulted in a slightly lower error. However, the real advantage of modal space forecasting is that only forecasting the first few ( $\sim 5$ ) modes is enough to achieve most of the improvement.

We note that the wavefront error improvements that we found are better than the theoretical maximum improvement (1.3) that was derived in Ref. 9 This is likely because Ref. 9 only

considered atmospheric effects, whereas our data have significant nonatmospheric components, certainly on T/T, likely on defocus and possibly other modes, mainly due to vibrations. We found a higher improvement because stationary vibrations are easier to forecast.

Our results also suggest the possibility of forecasting ahead by more than one frame, opening the possibility to decrease the AO frame rate and therefore increase sky coverage. Moreover, with a Lookback of 30 frames, 30 multiplies and add are required for each channel, which means that the computational complexity for the forecasting filter is much less than that of the wavefront computation process (which requires a number of multiplies and adds equal to the number of WFS slopes), which suggests that the forecasting filter could be easily retrofitted in the existing RTC.

Finally, we have found that using NNs to complement linear forecasting did not seem to bring any additional benefit, despite trying three different architectures. This suggests that the most efficient way to model the atmospheric turbulence to be corrected in our setup is with an autoregressive model and that residuals from such a model are random noise that cannot be forecast in any way. These residuals could include nonlinear effects in the atmospheric turbulence or in the Altair system itself.

These conclusions are only valid for the ALTAIR AO system and only for the data sets that we could access with the evaluated NN models. They would need to be verified on other systems and at other sites. However, we have highlighted the potential benefit of a simple data-driven linear forecasting model. Based on the ALTAIR data that we analyzed, we propose using a 30 coefficient filter for tip and tilt and a five coefficient filter for each of the DM actuator channels (modal coefficients are not available in real time in ALTAIR). This would result in a modest increase in computational load (i.e., compared with the wavefront reconstruction process) and therefore could potentially be implemented easily in the existing real time controller. Furthermore, updating the coefficients to reflect current observing conditions can be done by a soft real-time process looking at past telemetry data just like loop gains are currently optimized in Altair.

## Code, Data, and Materials Availability

Code and data may be made available upon request.

## References

1. O. Guyon, “Extreme adaptive optics,” *Annu. Rev. Astron. Astrophys.* **56**, 315–355 (2018).
2. E. Gendron and P. Lena, “Astronomical adaptive optics. I. Modal control optimization,” *Astron. Astrophys.* **291**, 337–347 (1994).
3. S. Y. Haffert et al., “Data-driven subspace predictive control of adaptive optics for high-contrast imaging,” *J. Astron. Telesc. Instrum. Syst.* **7**, 029001 (2021).
4. R. Swanson et al., “Closed loop predictive control of adaptive optics systems with convolutional neural networks,” *Mon. Not. R. Astron. Soc.* **503**, 2944–2954 (2021).
5. 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**, 833 (2009).
6. R. Jensen-Clem et al., “Demonstrating predictive wavefront control with the Keck II near-infrared pyramid wavefront sensor,” *Proc. SPIE* **11117**, 111170W (2019).
7. 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**, 731 (2019).
8. L. K. Saddlemyer et al., “Performance results of the reconstructor for Altair, the Gemini North AO system,” *Proc. SPIE* **4839**, 981–988 (2003).
9. N. Doelman, “The minimum of the time-delay wavefront error in adaptive optics,” *Mon. Not. R. Astron. Soc.* **491**, 4719–4723 (2020).
10. A. P. Wong et al., “Predictive control for adaptive optics using neural networks,” *J. Astron. Telesc. Instrum. Syst.* **7**, 019001 (2021).

11. B. Lim and S. Zohren, “Time-series forecasting with deep learning: a survey,” *Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci.* **379**, 20200209 (2021).
12. S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Comput.* **9**(8), 1735–1780 (1997).
13. D. Bahdanau, K. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” <https://arxiv.org/abs/1409.0473> (2015).
14. A. van den Oord et al., “Wavenet: a generative model for raw audio,” <https://arxiv.org/abs/1609.03499> (2016).

**Rehan Hafeez** is an engineering physics student at the University of British Columbia whose experience includes testing and prototyping of electromechanical systems, neural network optimization, and most recently data science and machine learning with the National Research Council of Canada. Driven by a desire to optimize how we interface with our environment, Rehan’s goal is to develop sustainable technologies to support and advance research.

Biographies of the other authors are not available.