# Reducing bandwidth error by predicting wavefronts using long short-term memory network

Xuewen Liu[a], Tim Morris[a], and Chris Saunter[a]

[a]Centre for Advanced Instrumentation, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK

## ABSTRACT

Predictive control techniques have been shown to successfully bring down bandwidth error in an Adaptive Optics (AO) system by predicting future wavefronts when AO correction applies. However, most existing techniques require explicit turbulence parameter estimation and update, which adds more difficulty to real-time implementation. We thus exploit the potential of artificial neural networks (ANNs) as a rising nonlinear, model-free tool for wavefront prediction through implicit and learnt turbulence modelling. We trained and optimised an ANN model consisting mainly of LSTMs (long short-term memory) to predict Shack-Hartmann slope measurements at next time step based on a time series of past noisy measurements in a simulated $7 \times 7$ system running at 150 Hz. Residual wavefront error (WFE) within the predictor's training regime is 244.2 nm, compared with 253.9 nm with one-frame delay and 243.4 nm with the perfect delay removal. The predictor's robustness against time-variant turbulence is also preliminarily verified using synthetic wind profile that changes speed every 10 frames (15 Hz).

**Keywords:** Wavefront prediction, predictive control, artificial neural networks

## 1. INTRODUCTION

Time lag between wavefront sensing and correction due to finite integration and readout time of wavefront sensors induces the bandwidth error in Adaptive Optics (AO) systems. This error term can sometimes severely limit system performance. For example, the uncompensated bandwidth error causes the broadening of point spread function (PSF) along dominant wind directions in Extreme AO (ExAO) systems, which severely degrades imaging contrast.[1]

One way to tackle this problem is to predict future wavefronts from past wavefront measurements utilising temporal and spatial correlations within atmospheric turbulence.[2,3] One such technique is predictive control, which incorporates wavefront prediction into controller design, such as Kalman filter based Linear Quadratic Gaussian (LQG) control.[4–6] In this framework, a chosen set of state variables are used to represent the turbulence and the AO system. Temporal evolution of those variables and their links with system measurements are described using linear models such as Autoregression models. Priors from system telemetry and noise statistics are then combined to obtain the control law. LQG based predictive controllers have been deployed for AO systems on high-contrast imaging (HCI) instruments SPHERE[7] and Gemini Planet Imager[8] for both turbulence correction and vibration filtering in tip-tilt modes. Stability and robustness of a LQG controller in full-mode single-conjugate AO (SCAO) control is also verified on sky,[9] showing overall performance improvement over a standard integrator controller.

However, identification and tracking of some turbulence parameters such as number of turbulence layers and wind speeds at each layer is unavoidable for LQG predictive controllers. In this paper, we demonstrate the potential of artificial neural networks (ANNs) as a framework for wavefront prediction. ANN approach is inherently model free, giving it the potential to learn any linear and nonlinear correlations within input data sequence without user input. Here we base our ANN model on long-short term memory (LSTM) networks, which is especially designed for processing time sequences.[10,11] The memory elements within the LSTM enables it to

Further author information: (Send correspondence to X. Liu)
X. Liu: E-mail: xuewen.liu@durham.ac.uk

learn information such as wind speeds from the data and to use that information in its prediction. Furthermore, the fluid nature of the memory allows the network to adapt to changes in such information, which is especially desirable to cope with ever-evolving turbulence.

## 2. METHODOLOGY

We demonstrate the potential of an ANN predictor in numerical simulations. The predictor is trained to predict open-loop slope measurements of a Shack-Hartmann wavefront sensor (SH-WFS) at next time step based on a sequence of past noisy measurements in a single-conjugate AO (SCAO) system. The simulated $7 \times 7$ subaperture system serves two purposes: its wavefront sensing subsystem is used for generating 200,000 short sequences each of 30 time steps as training data; after training, the ANN predictor is incorporated into the system to form a predictive correction loop, which will be then compared with 1) the default one-frame delay correction, where the correction is based on WFS measurements one frame ahead to account for servo lag, and 2) delay-compensated correction, where this delay is manually removed by using current WFS measurements for correction immediately.

### 2.1 SCAO Simulation and Training Data Generation

We use Soapy for SCAO simulation.[12] Building blocks and data flow of the simulated system is shown in Fig 1. Primary simulation parameters are listed in Table 1.

To generate each WFS measurement sequence, a larger random phase screen following the same Von Karman statistics is generated first. Pure frozen flow is assumed, under which the screen is translated over the telescope aperture with certain velocity due to wind. At each time step, a smaller portion of the screen seen by the telescope is sampled and then passed to `SH-WFS`, from which a noisy measurement of wavefront slopes $\mathbf{S}$ (both x and y slope of each subaperture, 72 slopes in total) is output. This process continues until 30 time steps, and a measurement sequence $(\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_{30})$ is gathered as a training sample.

A new random phase screen is regenerated and a different wind velocity is randomly chosen to get another training sample in the same way described above. This process continues until 100,000 samples are gathered. We then reverse each sequence to form the other half of the training set. This corresponds to reversing the wind direction and is adopted for improving model robustness.[13]

To form a correction loop, the slope measurements are then input to a reconstructor module `Recon`, from which the deformable mirror (DM) commands and then the DM shape (phase correction) are determined. By subtracting the DM shape from the input smaller phase screen, we can obtain the root-mean-squared residual wavefront error (RMS WFE). There is typically one-frame delay between slope calculation and correction generation (the central loop in Fig. 1) to account for WFS integration and readout in real-world implementation. This delay can be compensated in simulation though by either applying slope measurements immediately for correction (the lower loop, ground truth) or feeding current measurements to an ANN predictor to predict slopes one frame ahead when the correction applies (the upper loop).

### 2.2 ANN Training and Optimisation

We use Keras for ANN training and optimisation.[14] During ANN training, the last measurement in each sample $\mathbf{S}_{30}$ is the training target while the remaining $(\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_{29})$ is the training input.

The ANN structure consists of stacked LSTM cells (the number of which is optimised later) and a final fully-connected output layer. Trainable parameters of the network are initialised using Gaussian distribution. 90% of overall samples are used for updating the network parameters (training) and the remaining 10% for validation (preventing the network from overfitting). 10% Dropout is added to each LSTM.[15] The training error is mean squared error (MSE) between the targeted output $\mathbf{S}_{30}$ and the actual network output $\tilde{\mathbf{S}}_{30}$. Adam optimisation algorithm is used for updating the network parameters in a direction that decreases the training error gradually.[16] Training is terminated after 40 epochs.

To exploit the architecture and the training procedure, we tune three hyperparameters of the network: number of LSTMs (up to 4), size of output vectors of each LSTM (an integer between 50 and 250, different for each layer) and the initial learning rate of Adam algorithm. After a set of hyperparameters is chosen, the resulting network is trained with the ANN setup stated above. The optimal architecture that achieves minimum errors on

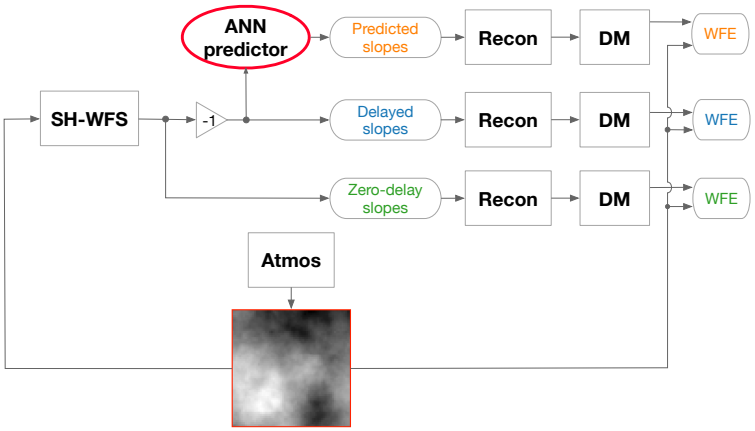


Figure 1. Diagram of the simulated SCAO system. `-1` in the triangle denotes the default one-frame delay between wavefront sensing and correction, and `1` in the triangle denotes applying the slope measurements for correction immediately to perfectly remove the bandwidth error. The ANN predictor behaviour is expected to be within those two implementations.

Table 1. Modules and corresponding parameters of the simulated SCAO system.

| **Module** | **Parameter** | **Value** |
|---|---|---|
| System | Frequency | 150 Hz |
| | Pure delay | One frame |
| Atmosphere | # of phase screens | 1 |
| | Wind speed | 10-15 m/s |
| | Wind direction | 0-360° |
| | $r_0$ @ 500nm | 0.16 m |
| | $L_0$ | 25 m |
| Telescope | Diameter | 4.2 m |
| | Central obscuration | 1.2 m |
| SH-WFS | Guide star magnitude | 10 |
| | # of subapertures | 7×7 (36 active) |
| | Photon noise | True |
| | Readout noise | 1 $e^-$ RMS |
| | Wavelength | 600 nm |
| Piezo DM | # of actuators | 8×8 (64 active) |
| SCI | Wavelength | 1650 nm |

the validation set after training terminates consists of two LSTMs, with 235 and 242 output units respectively. A learning rate of 1e-3 was used and no performance improvement was observed by modifying this rate. The resulting optimal model has 769,720 trainable parameters in total, and the corresponding computational load is $2.3\times10^8$ for a 150 Hz system.

## 3. RESULTS

The optimal model is incorporated into the simulated AO system after training to form a predictive correction loop, the WFE of which will be then compared with the one-frame delay correction and the delay-compensated correction. Trainable parameters within the model are now fixed. We test the model in three different cases: 1) the AO and turbulence condition is the same as for training; 2) guide star magnitude is varied between 6 and 11, which changes signal-to-noise ratio (SNR) of wavefront sensing; 3) in addition to the increase of GS magnitude in 2), a synthetic wind speed profile is introduced where the speed changes every 10 frames (15 Hz). For each case, we generate a different test set with 1,000 slope sequences each of 100 time steps. Most parameters are the same as in Table. 1, unless stated otherwise.

Fig. 2 shows RMS WFE averaged across all 1,000 test sequences of the three correction loops in case 1). All simulation parameters are the same as during training, except that the phase screens used did not form part of the training data set and had not been observed by the network before. Wind speed is 15 m/s along horizontal direction, the magnitude of which is the upper limit of wind speed among all training sequences to exaggerate the servo lag for the purposes of the clarity within the plots presented here. At each time step during test, the prediction is based on the whole input history so far. The predictor stabilises in around 12 frames, which is less than the time it is allowed to accumulate its memory (30 time steps), indicating training using shorter LSTM sequences may be possible. Mean RMS WFE of the delayed, predicted and delay-compensated correction loop (averaged after the $12^{\text{th}}$ frame and across all sequences) is 253.9 nm, 244.2 nm and 243.4 nm respectively.

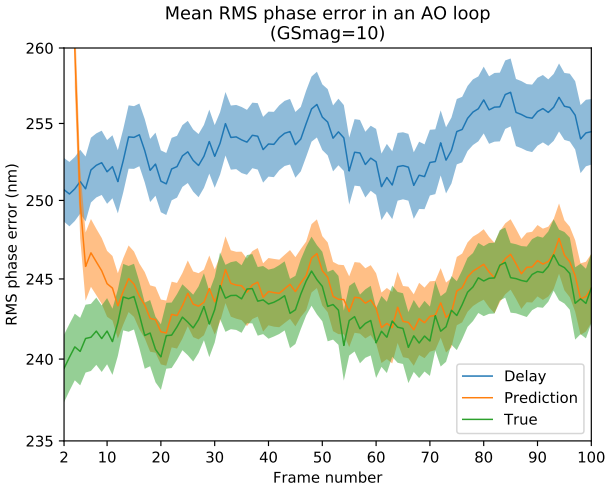


Figure 2. RMS WFE with delay (blue), after prediction (orange) and with the perfect delay compensation (green). All simulation parameters in this test are the same as during training. The predictor stabilises in 12 frames. Wind speed is 15 m/s horizontal.

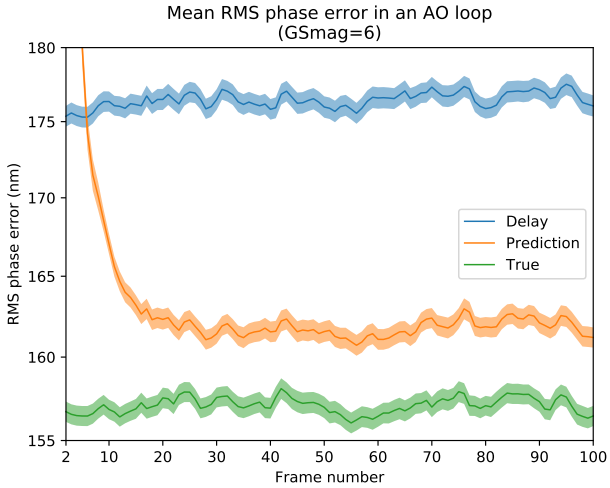


Figure 3. RMS WFE with delay (blue), after prediction (orange) and with the perfect delay compensation (green) when the guide star is 4-magnitude brighter than in training. Wind speed is 15 m/s horizontal.

To see the robustness of our network in terms of different input statistics, we mainly focus on the effect of changing SNR of input slopes. GS magnitude is varied between 6 and 11. At each magnitude, corresponding 1,000 slope sequences of length 100 frames are tested. RMS WFE at magnitude 6 is shown in Fig. 3. The much smaller variance compared to Fig. 2 is mainly due to increased SNR of wavefront sensing. The average WFE after the predictor stabilises is 176.6 nm, 161.8 nm and 157.2 nm for the delayed, predicted and delay-compensated

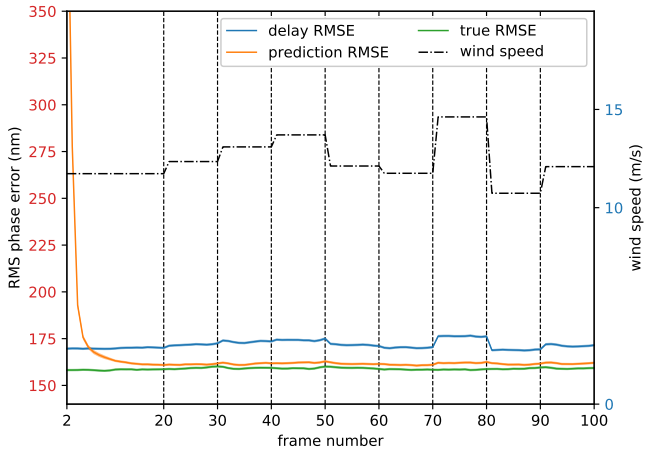


Figure 4. Robustness of the ANN predictor against wind speed fluctuations (dash-dotted line). Wind direction is constant. Guide star is 4-magnitude brighter than in training.

correction. The overall trend is similar for the remaining magnitudes, showing improved system performance in terms of WFE after prediction.

Further to increasing GS magnitude, we introduce a synthetic wind speed profile to account for time-variant turbulence statistics (dash-dotted line in Fig. 4). The wind speed is 15 m/s in the first 20 time steps to allow the predictor to build up its memory, and is then randomly changed every 10 frames (between 10 and 15 m/s). This fluctuation is reflected in the varying WFE of the delayed loop in Fig. 4, as bandwidth error scales with wind speed.[17] On the other hand, the prediction maintains relative stability mainly due to 1) the memory element within is controlled by both the forgetting and remembering mechanism, and can adjust its memory content in accordance with input statistics as the fluid nature of LSTM;[13] and 2) the scale of wind speed fluctuation (less than 5 m/s at 15 Hz) is within the adjusting capacity of our network. We have observed under certain conditions such as much more drastic wind fluctuations or a fluctuation before the predictor stabilises might deviate the predictor from its optimal performance, it is thus meaningful to further test the ANN predictor using site data.

## 4. CONCLUSIONS AND FUTURE WORK

We have demonstrated in numerical simulations of a $7 \times 7$ SCAO system operating at 150 Hz the efficacy of artificial neural networks as a nonlinear, model-free framework for wavefront prediction. Residual wavefront errors are significantly improved after prediction compared to when the servo lag is perfectly removed. The predictor is tolerant of an increase of guide star magnitude of at least 4, indicating potential robustness of an ANN predictor against different input statistics such as SNR of wavefront sensing and thus the unnecessity of re-training under such conditions. Furthermore, the predictor can self update without user tuning when wind speed changes during prediction.

Our next step will be to analyse quantitatively the reduction in bandwidth error using an ANN predictor. We will also test our approach using CANARY on-sky data to validate the model robustness in a more realistic situation.[18]

## ACKNOWLEDGMENTS

The authors would like to thank Francisco Javier de Cos Juez and Carlos González-Gutiérrez from University of Oviedo, Spain for their valuable suggestions and thought-provoking discussions with them. X. Liu is funded by Durham University through their doctoral scholarship program and China Scholarship Council (CSC).

## REFERENCES


[1] Kasper, M., "Adaptive optics for high contrast imaging," in [*Proc. of SPIE 8447*], **8447**, 84470B (2012).

[2] Wang, L., Schöck, M., and Chanan, G., "Atmospheric turbulence profiling with slodar using multiple adaptive optics wavefront sensors," *Appl. Opt.* **47**(11), 1880–1892 (2008).

[3] Poyneer, L. A., van Dam, M., and Véran, J.-P., "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).

[4] Paschall, R. N. and Anderson, D. J., "Linear quadratic gaussian control of a deformable mirror adaptive optics system with time-delayed measurements," *Appl. Opt.* **32**(31), 6347–6358 (1993).

[5] Le Roux, B., Conan, J.-M., Kulcsár, C., Raynaud, H.-F., Mugnier, L. M., and Fusco, T., "Optimal control law for classical and multiconjugate adaptive optics," *J. Opt. Soc. Am. A* **21**(7), 1261–1276 (2004).

[6] Poyneer, L. A., Macintosh, B. A., and Véran, J.-P., "Fourier transform wavefront control with adaptive prediction of the atmosphere," *J. Opt. Soc. Am. A* **24**, 2645–2660 (Sep 2007).

[7] Petit, C., Sauvage, J.-F., Fusco, T., Sevin, A., Suarez, M., Costille, A., Vigan, A., Soenke, C., Perret, D., Rochat, S., Barrufolo, A., Salasnich, B., Beuzit, J.-L., Dohlen, K., Mouillet, D., Puget, P., Wildi, F., Kasper, M., Conan, J.-M., Kulcsár, C., and Raynaud, H.-F., "Sphere extreme ao control scheme: final performance assessment and on sky validation of the first auto-tuned lqg based operational system," in [*Proc. of SPIE 9148*], **9148**, 91480O (2014).

[8] Poyneer, L. A., Palmer, D. W., Macintosh, B., Savransky, D., Sadakuni, N., Thomas, S., Véran, J.-P., Follette, K. B., Greenbaum, A. Z., Ammons, S. M., Bailey, V. P., Bauman, B., Cardwell, A., Dillon, D., Gavel, D., Hartung, M., Hibon, P., Perrin, M. D., Rantakyrö, F. T., Sivaramakrishnan, A., and Wang, J. J., "Performance of the gemini planet imager's adaptive optics system," *Appl. Opt.* **55**(2), 323–340 (2016).

[9] Sivo, G., Kulcsár, C., Conan, J.-M., Raynaud, H.-F., Éric Gendron, Basden, A., Vidal, F., Morris, T., Meimon, S., Petit, C., Gratadour, D., Martin, O., Hubert, Z., Sevin, A., Perret, D., Chemla, F., Rousset, G., Dipper, N., Talbot, G., Younger, E., Myers, R., Henry, D., Todd, S., Atkinson, D., Dickson, C., and Longmore, A., "First on-sky scao validation of full lqg control with vibration mitigation on the canary pathfinder," *Opt. Express* **22**(19), 23565–23591 (2014).

[10] Hochreiter, S. and Schmidhuber, J., "Long short-term memory," *Neural Computation* **9**, 1735–1780 (1997).

[11] Gers, F. A., Schmidhuber, J., and Cummins, F., "Learning to forget: continual prediction with lstm," in [*1999 Ninth International Conference on Artificial Neural Networks ICANN 99. (Conf. Publ. No. 470)*], **2**, 850–855 (1999).

[12] Reeves, A., "Soapy: an adaptive optics simulation written purely in python for rapid concept development," in [*Proc. of SPIE 9909*], **9909**, 99097F, International Society for Optics and Photonics (2016).

[13] Goodfellow, I., Bengio, Y., and Courville, A., [*Deep Learning*], MIT Press (2016). http://www.deeplearningbook.org.

[14] Chollet, F. et al., "Keras." https://keras.io (2015).

[15] Gal, Y. and Ghahramani, Z., "A theoretically grounded application of dropout in recurrent neural networks," in [*Proceedings of the 30th International Conference on Neural Information Processing Systems*], *NIPS'16*, 1027–1035 (2016).

[16] Kingma, D. and Ba, J., "Adam: A method for stochastic optimization," in [*Proceedings of the 3rd International Conference on Learning Representations*], (2014).

[17] Hardy, J., [*Adaptive Optics for Astronomical Telescopes*], Oxford Series in Optical & Imaging Sciences, Oxford University Press (1998).

[18] Morris, T. J., Hubert, Z., Myers, R., Gendron, E., Longmore, A., Rousset, G., Talbot, G., Fusco, T., Dipper, N., Vidal, F., Henry, D., Gratadour, D., Butterley, T., Chemla, F., Guzman, D., Laporte, P., Younger, E., Kellerer, A., Harrison, M., Marteaud, M., Geng, D., Basden, A., Guesalaga, A., Dunlop, C., Todd, S., Robert, C., Dee, K., Dickson, C., Vedrenne, N., Greenaway, A., Stobie, B., Dalgarno, H., and Skvarc, J., "Canary: The ngs/lgs moao demonstrator for eagle," in [*1st International Conference on Adaptive Optics for Extremely Large Telescopes (AO4ELT)*], (2009).