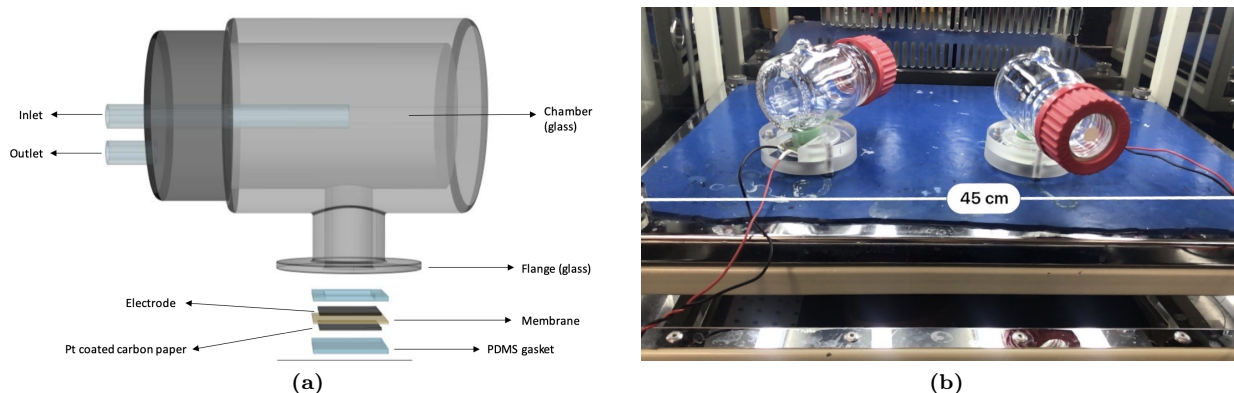# Supplementary information: Time series analysis and Long Short-Term Memory (LSTM) network prediction of BPV current density

Tonny I. Okedi and Adrian C. Fisher

List of Figures

BPV schematic and experimental set-up.

(a)

(b)

Figure S.1 – BPV schematic and experimental set-up. **(a)** The anode chamber is a 250 ml glass bottle with a glass flange extending from its side. The membrane electrode assembly is formed by a Toray carbon paper electrode (anode), a nitrocellulose membrane with 0.22 μm pores, and an Alfa Aesar Pt coated carbon paper with 3 $mg\ m^{-2}$ coating (cathode). Two polydimethyl siloxane (PDMS) gaskets were used to seal the device. Device architecture and schematic are from designs by Gonzalez-Aravena[1]. **(b)** Experimental set-up (BPV 1 and BPV 2). The anode chamber was sealed by a bottle cap (red) with a transparent top. The top of the cap had a central hole covered with a 0.22 μm nitrocellulose membrane (white) to allow air flow in to and out of the anode chamber while ensuring an axenic environment is maintained. The red and black wires are connected to titanium strips in electrical contact with the cathode and anode respectively. The *Synechococcus elongatus sp.* PCC7942 culture (5 ml) was inoculated to the neck of the glass flange (green liquid) at OD$_{750}$ = 2. All components shown in **(a)** are held together with two transparent perspex frames joined by three M5 screws (20 mm length). The screws also prop up the device above the surface to allow uninterrupted air flow to the cathode. The devices are housed in a SciQuip Incu-Shake FL18-750R incubator.

Table S.1 – Benchmark of BPV 1 and 2 performance. **NM: Not measured.**

Strain	Anode	Cathode	Light intensity ($\mu mol\ m^{-2}\ s^{-1}$)	OCP (mV)	Max power ($mW\ m^{-2}$)	Ref.
S. elongatus sp. PCCC7942	Carbon paper	Pt. coated carbon paper	21±0.3	283±7	NM	BPV 1
				241±6	NM	BPV 2
				298±7	0.016	[1]
Synechocystis sp. PCC6803	Carbon paint	Pt-containing carbon cloth	1.4	100	0.35	[2]
Paulschulzia pseudovolvox	Carbon paint	Carbon cloth with 10% Pt	400	160	6.2	[3]
Mixed photosynthetic consortium	Carbon paint w/ nanostructured polypyrrole	Pt-containing carbon cloth	1.4	250	5.9	[4]

Illustrative example of the seasonal component concept
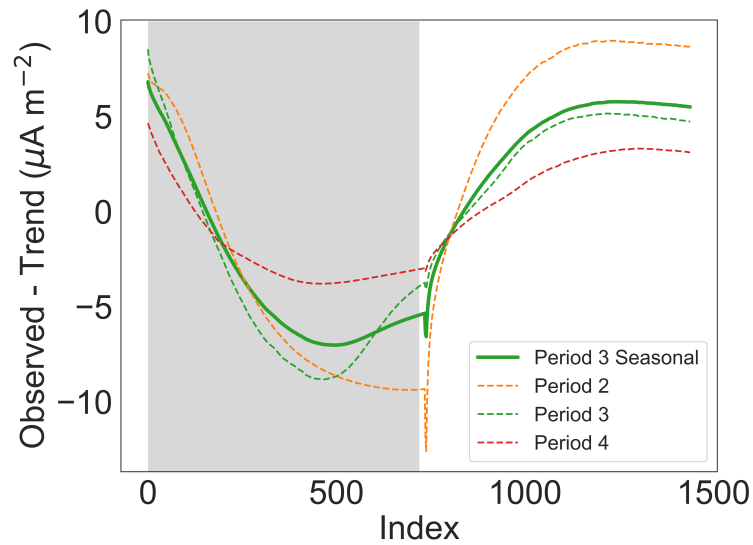
Figure S.2 – Illustrative example of the seasonal component concept. Calculation of the seasonal component for period 3 (green in Fig. 3a in main text) using a seasonal smoother of 3 consecutive periods, centred around the period. The following steps are taken: (1) for each of the periods 2, 3 and 4, the trend component is subtracted from the observed current density at each time step (orange, green and red dashed lines, respectively); (2) the seasonal component for the period is calculated by taking the mean of this difference at each time step (solid green line). It should be noted that the above is a simplified scheme of the more complex STL algorithm which is an iterative calculation whereby the trend and seasonal components are calculated in a loop. Full details of the STL algorithm may be found in the original implementation [5]. In practice, STL uses locally estimated scatterplot smoothing (LOESS), allowing both the trend and seasonal components to be estimated where there is missing data, such as at the beginning or end of the time series where there may not be enough consecutive observations either before or after the time step of interest as dictated by the size of the smoother.

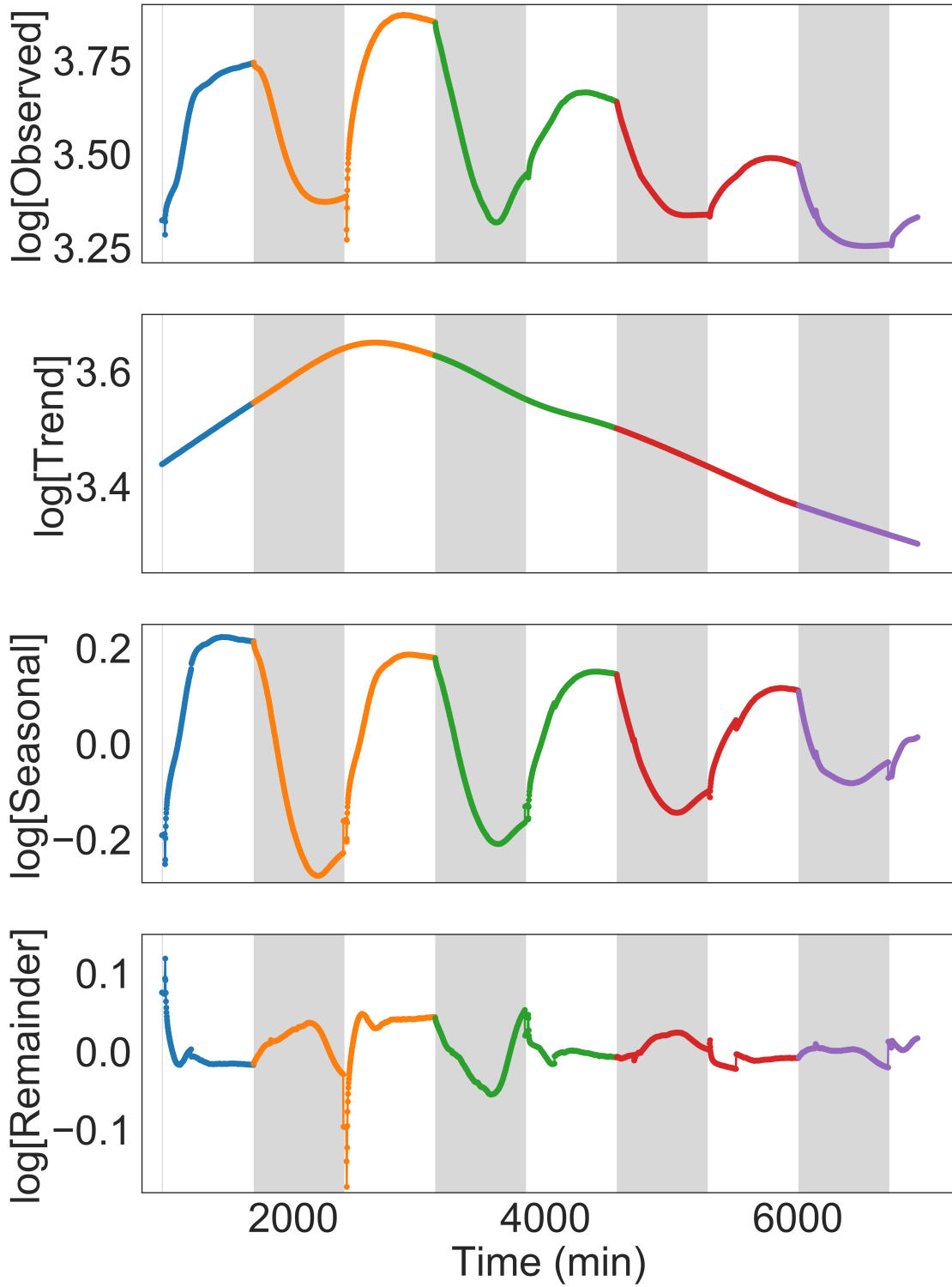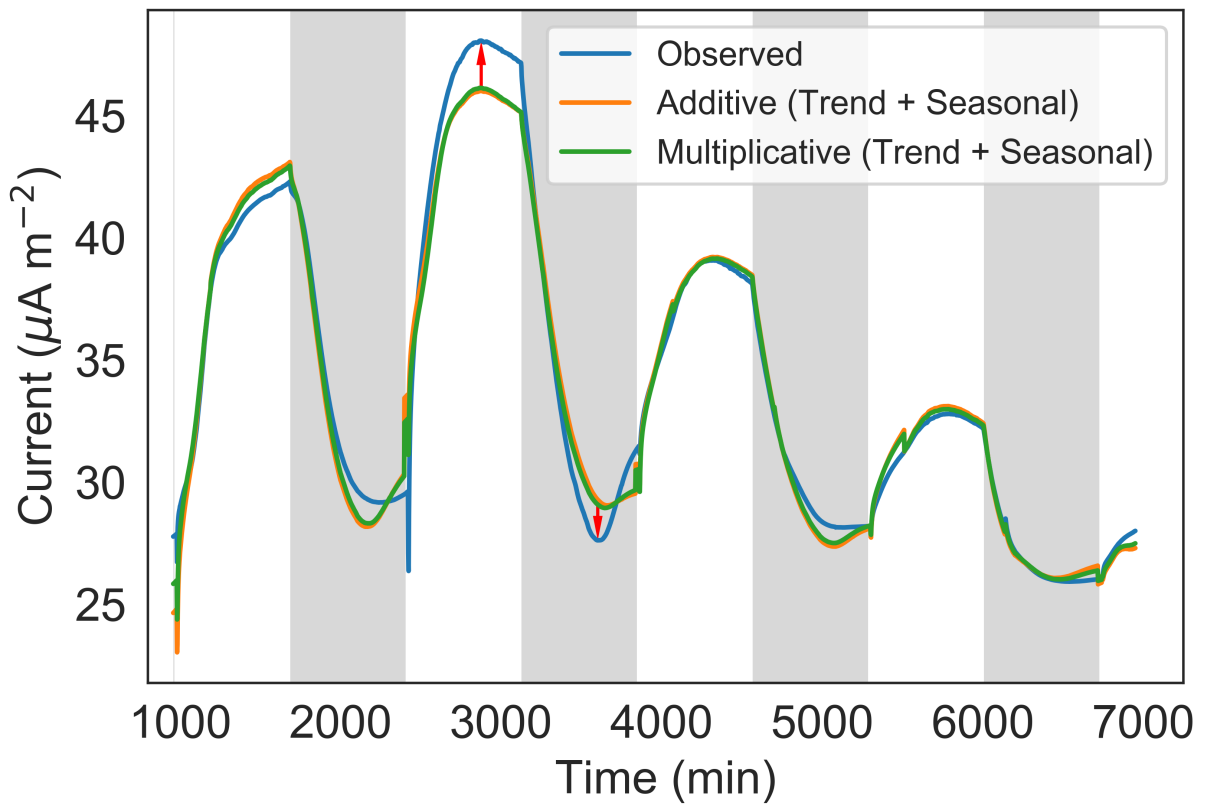# Comparison of additive and multiplicative STL Decomposition

Figure S.3 – Multiplicative decomposition of BPV 1 current density profile.

Figure S.4 – Comparison between additive and multiplicative STL decompositions of BPV 1 current density profile. No significant differences were seen between the two decompositions. The difference between the observed time series and the sum of the trend and seasonal components (as exemplified by the red arrows) is the remainder component.

Correlograms
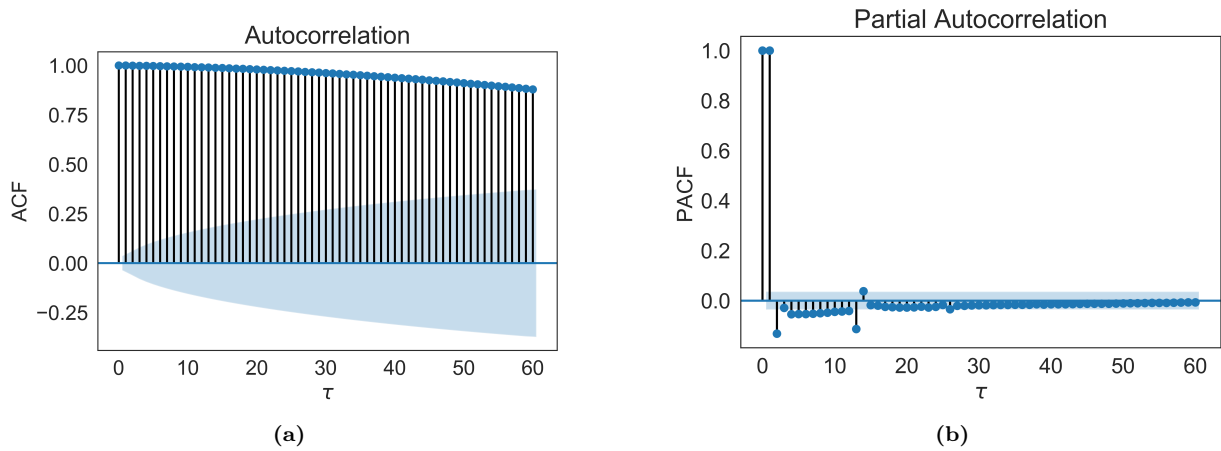
Figure S.5 – BPV 1 current density profile correlogram and partial autocorrelation plot. (a) Current density profile correlogram. **(b)** Partial autocorrelation function (PACF) plot. Both the correlogram and the PACF plot were obtained using Python's *statsmodel* module. The window size for preparing the LSTM input data was chosen to be 14 since all PACF values after lag 14 are below the 95% large-lag confidence interval. This was also the case for the seasonal component of the current density.
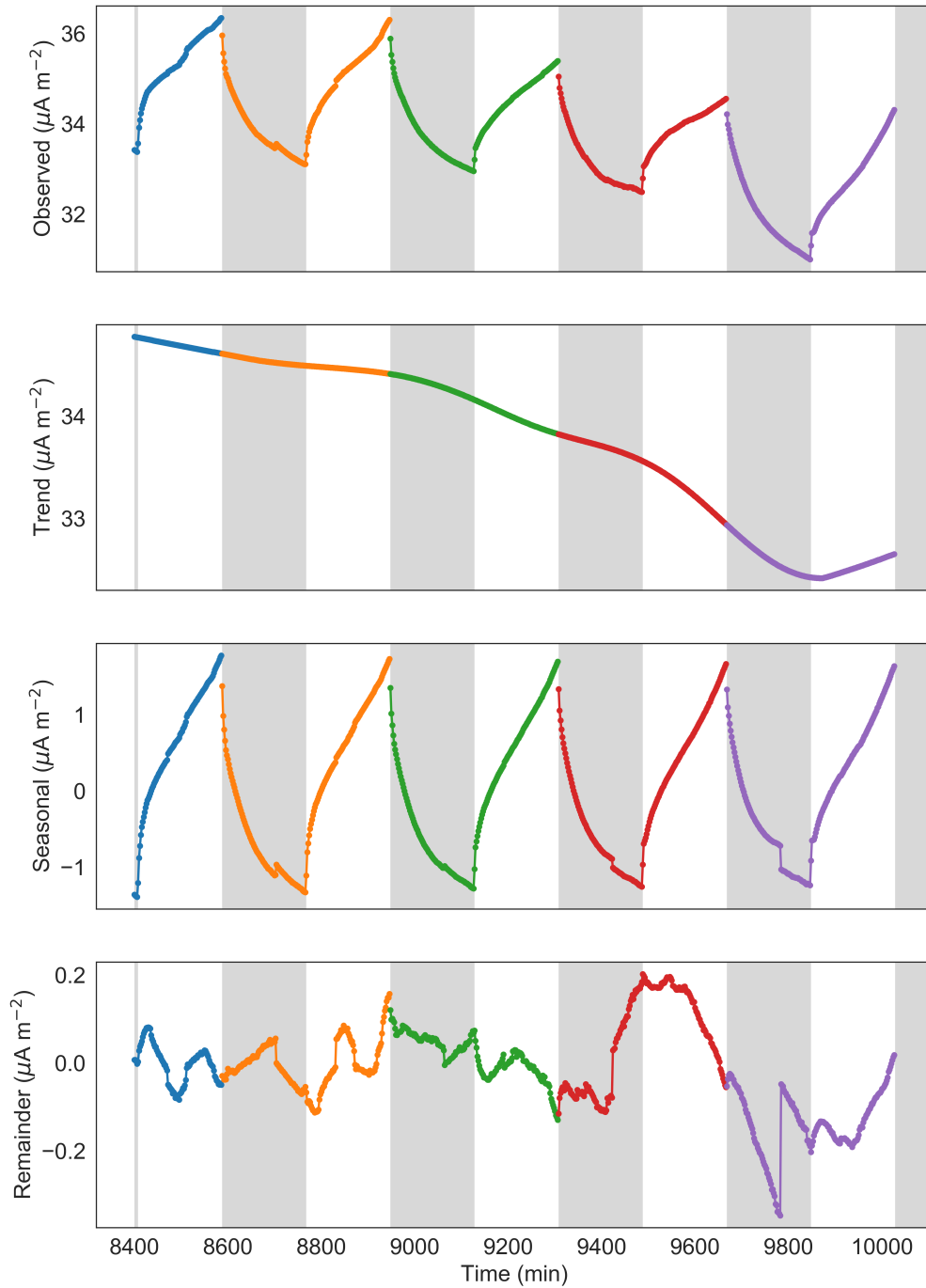
Tuning Adam learning rate

(a)

(b)
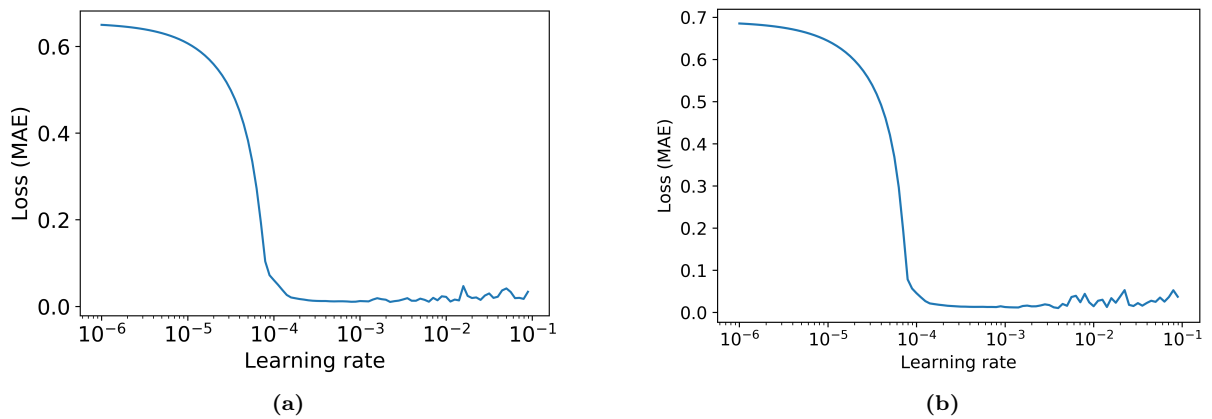
Figure S.6 – Tuning Adam optimiser learning rate for training the LSTM networks. (a) Observed current density LSTM network. **(b)** Seasonal component LSTM network. In both cases, a starting learning rate of $4 \cdot 10^{-3}$ is used since larger learning rates resulted in unstable behaviour.

Learning Curves

(a)

(b)
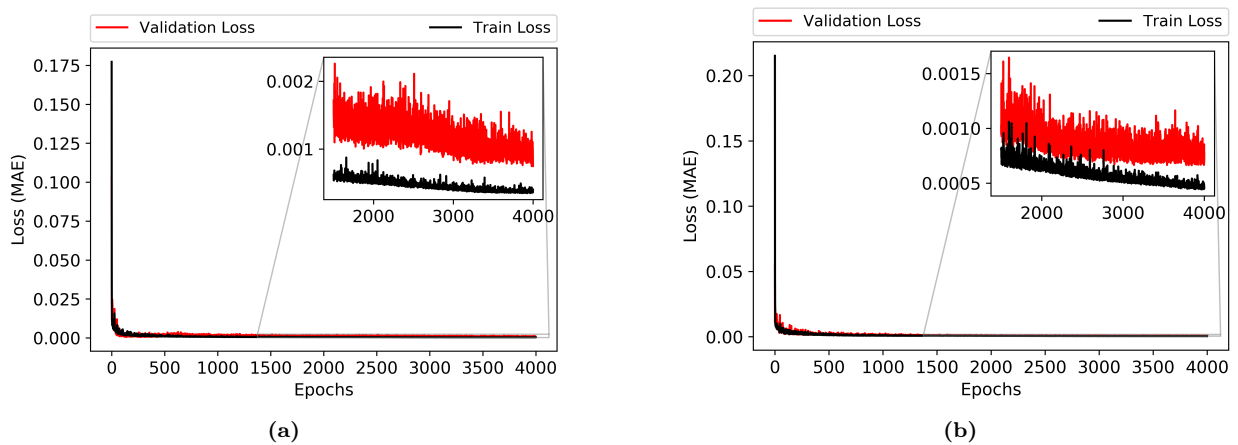
Figure S.7 – LSTM network learning curves. (a) Observed current density profile LSTM network. **(b)** Seasonal component LSTM network. Zooming in (inset graphs) shows that networks continued to learn slowly up to 4000 epochs. For the seasonal component LSTM network, early stopping at 2500 epochs was used to prevent overfitting. The noise in the learning curves is due to batched implementation during training.

Figure S.8 – Decomposition of current density profile for BPV operated in 3h:3h dark-light cycle. The operating conditions for the device were as described in the section 2.2.3 in the main text, except for the dark-light cycle. Photocurrent rise was seen to be immediate following illumination, unlike in BPV 1 and 2 operated under a 12h:12h dark-light cycle where there was a 16-18 *min* lag. In addition, there was no 'dip' in the current before the rise. The ostensibly smaller photocurrent is due to the shorter illumination interval. It was shown in Fig. 4b in the main text that it takes at least approximately 480 *min* (8 *h*) for the photocurrent to reach peak under the operating conditions applied.

LSTM prediction on observed current profile
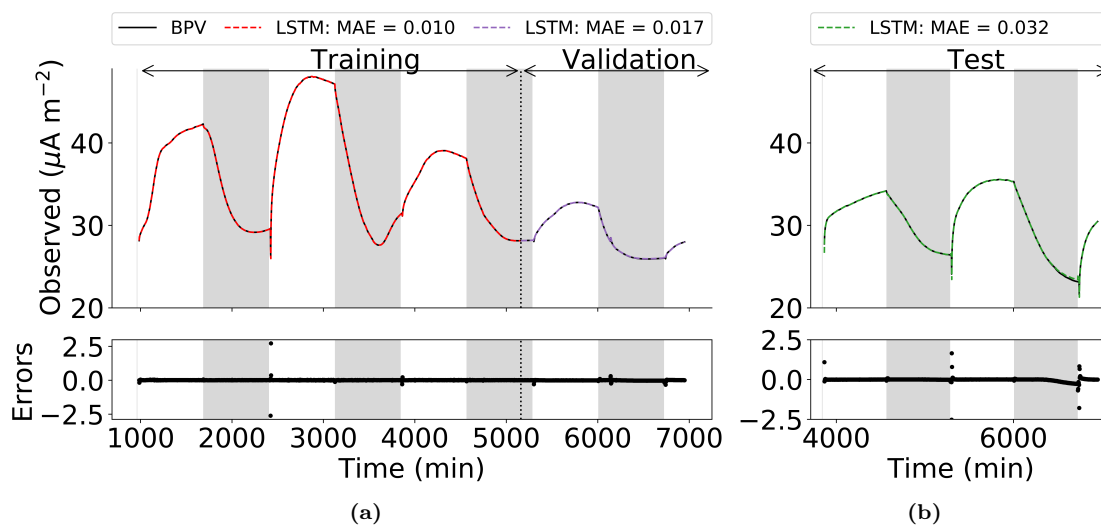
Figure S.9 – LSTM network one-step-ahead observed current density predictions for BPV 1 and 2. (a) Predicted BPV 1 observed current density (training and validation sets). **(b)** Predicted BPV 2 observed current density (test set, unseen during training).
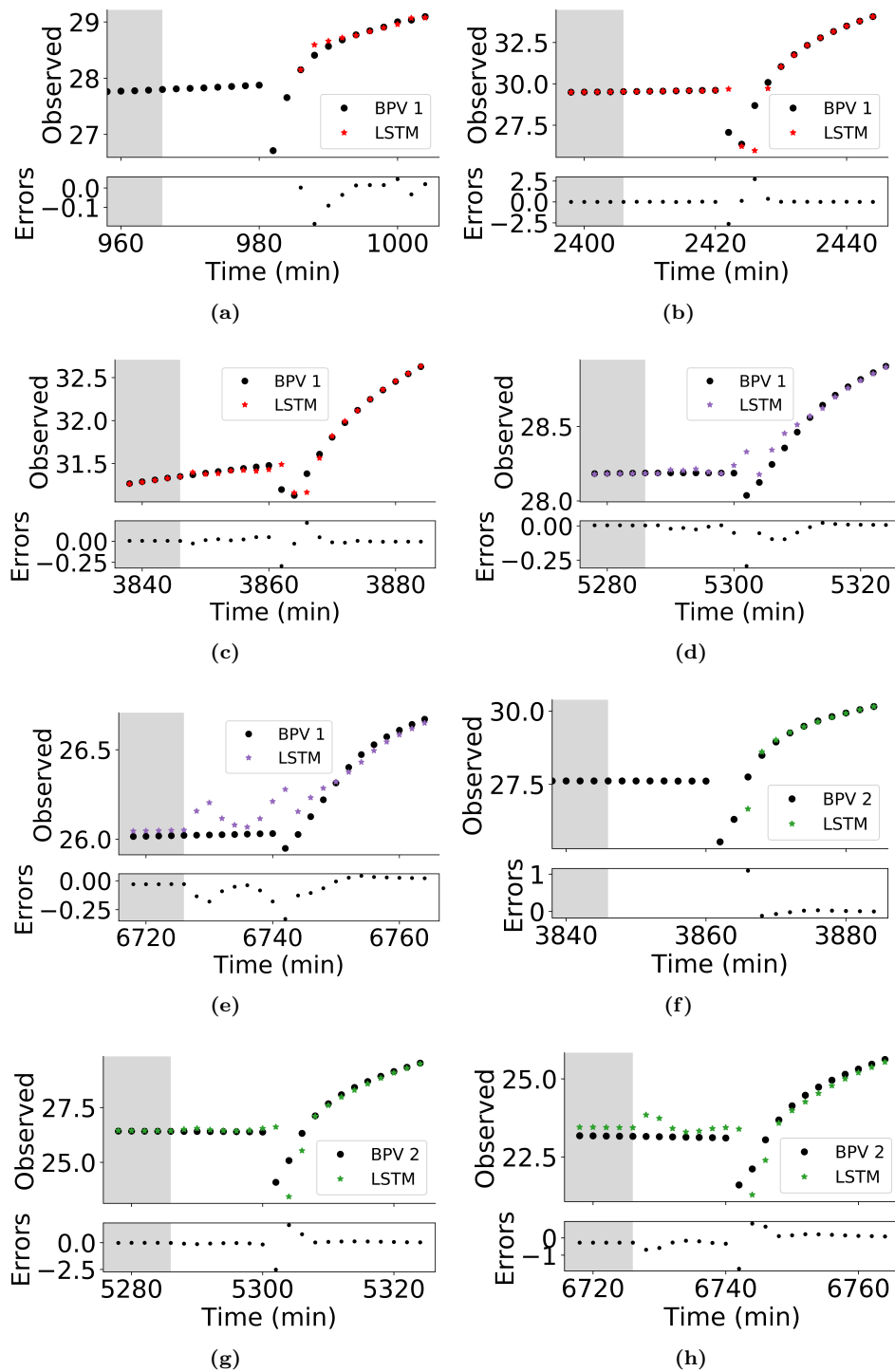
Figure S.10 – Zoomed in observed current density profiles showing the first 40 minutes post-illumination for each period. (a) – (c) BPV 1 periods 1 to 3 sequentially (training set). **(d) – (e)** BPV 1 periods 4 and 5 sequentially (validation set). **(f) – (h)** BPV 2 periods 1 to 3 sequentially (testing set). In (a) and (f), there are no LSTM predictions for the first 28 minutes post illumination (or 14 time points). This is because a window size of 14 is used to calculate the one-step-ahead value. Therefore, because these are the first periods of the BPV 1 and 2 current density profiles respectively, the first time step where there is full data to predict the current is time step 15. Y-axis units are μA m^{-2}.
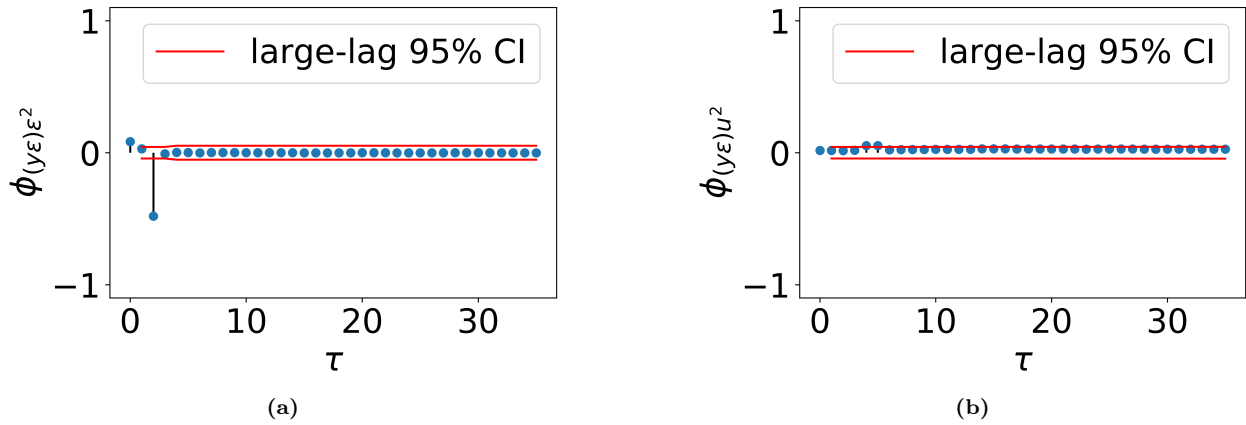
Figure S.11 – Correlation tests using inputs, residuals and outputs on the observed current density LSTM network. **(a)** Correlation between $y\epsilon$ and ϵ^2. **(b)** Correlation between $y\epsilon$ and u^2, where the input u is light status. The LSTM network fails the model validity test with a $\phi_{(y\epsilon)\epsilon^2}(\tau)$ value above the 95% large-lag confidence interval at $\tau = 2$.

References

[1] A. C. Gonzalez-Aravena, *Ph.D. Thesis*, University of Cambridge, 2017.

[2] Z. Yongjin, J. Pisciotta, R. B. Billmyre and I. V. Baskakov, *Biotechnol. Bioeng.*, 2009, **104**, 939–946.

[3] V. M. Luimstra, S. J. Kennedy, J. Güttler, S. A. Wood, D. E. Williams and M. A. Packer, *J. Appl. Phycol.*, 2014, **26**, 15–23.

[4] J. M. Pisciotta, Y. Zou and I. V. Baskakov, *PLoS One*, 2010, **5**, e10821.

[5] R. B. Cleveland, W. S. Cleveland, J. E. McRae and I. Terpenning, *J. Off. Stat.*, 1990, **6**, 3–73.