bioRxiv preprint doi: https://doi.org/10.1101/147553; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Three-dimensional plant architecture and sunlit-shaded dynamics: a stochastic model to enable high-throughput physiological analysis

Renata Retkute[a], Alexandra J. Townsend[b], Erik H. Murchie[c], Oliver E. Jensen[d], Simon P. Preston[e]

[a] *Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research, School of Life Sciences, University of Warwick, Coventry, CV4 7AL, UK*
[b] *Department of Cell and Molecular Biology, School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK*
[c] *Division of Plant and Crop Sciences, School of Biosciences, University of Nottingham, Sutton Bonington campus, Loughborough, LE12 5RD, UK*
[d] *School of Mathematics, University of Manchester, Oxford Road, Manchester, M13 9PL, UK*
[e] *School of Mathematics, University of Nottingham, Nottingham, NG7 2UH, UK*

Abstract

- *Background and Aims* Diurnal changes in solar position and intensity combined with the structural complexity of plant architecture results in highly variable and dynamic light patterns within the plant canopy. This influences productivity because photosynthesis is highly responsive to changes in light intensity. Current methods to characterise light dynamics, such as ray tracing, are able to produce data with excellent spatio-temporal resolution but are computationally intensive and the resultant data are complex and high dimensional. This necessitates development of more economical models to simulate realistic light patterns over the course of the day.

- *Methods* High-resolution reconstructions of field-grown wheat canopies were assembled in various configurations using digital imaging methods. A forward ray-tracing algorithm was employed to compute canopy light dynamics at high (1 minute) temporal resolution. The output was used to designate leaf sections as sunlit or shaded at each time interval. A stochastic model for sunlit-shaded patterns was developed and fitted using maximum likelihood estimation. Principal component analysis

bioRxiv preprint doi: https://doi.org/10.1101/147553; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

was used to study the similarities and differences in dynamic light distribution between different canopies.

- *Key Results* A two-state Markov chain model sufficiently replicates key features of the light dynamics, as achieved via ray-tracing data, provided that the rates of switching (from sunlit to shaded, and vice versa) are assumed to be distinct and to be functions of the time of day and the height within the canopy.

- *Conclusions* The Markov chain model captures the essential features of light dynamics within a canopy and enables a clear understanding of how position, time of day, and canopy characteristics affect light patterns. Furthermore, the model provides a cheap way to simulate realistic light patterns, which we anticipate being important for feeding into larger scale photosynthesis models for calculating how light dynamics affects the photosynthetic productivity of a canopy.

Keywords: 3D architecture, crop canopy, ray tracing, Markov model, emulation, photosynthesis

---

Introduction

Plant canopies are complex three-dimensional (3D) structures which influence light dynamics during a day, largely as a result of solar movement. Due to diurnal changes in solar position and occlusion caused by overlapping leaves and canopy architectural characteristics, leaves will alternate between sunlit and shaded periods. Bright periods of full sunlight which temporarily penetrate to lower layers of the canopy ('sun-flecks') can have a high degree of temporal variability. The dimensions and spatio-temporal dynamics of direct light influence many fundamental physiological functions, such as photosynthesis, photoacclimation, and photoinhibition [13, 1, 3, 19], and secondary biophysical processes such as drought tolerance, water-use efficiency [12], plant growth and crop yield [9]. This is because photosynthesis does not accurately track the fluctuations in light and there may be delays in response to a change in conditions. For example, a delay in photosynthetic induction to high light can result from the time taken to activate enzymes in the Calvin cycle, open stomatal pores and build up metabolite pool sizes [13, 1]. The speed of recovery from photoprotection in low light has recently been shown

bioRxiv preprint doi: https://doi.org/10.1101/147553; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

to determine productivity [7, 8]. This demonstrates that improvements to crop yield by optimizing photosynthesis will come from an understanding of how photosynthesis responds in naturally dynamic environments, rather than steady-state conditions. However, our understanding of how photosynthesis is integrated into canopies is hampered somewhat by the complexity and computational burden of using ray-tracing methods for analyzing light dynamics in 3D canopy reconstructions [6].

Sunlit and shaded patterns may be measured using various empirical techniques, such as hemispherical canopy photographs [10], a photosynthetically active radiation (PAR) sensor moving on a horizontal track [15], an electromagnetic 3D digitiser [17], or a near-ground imaging spectroscopy system [20]. However spatial resolution of these techniques is typically very poor. This limitation is overcome by digital reconstruction of plants and canopies [11] and digital ray tracing [18].

Current ray-tracing approaches for calculating the positions and dynamics of light in canopies can consume substantial time and computer resources. This is especially the case where large numbers of analyses are required simultaneously. To accurately and rapidly simulate sunlit-shaded patterns within complex canopies, a novel mathematical approach has been developed here. This uses a two-state Markov chain model, where states switch from sunlit to shaded and from shaded to sunlit with rates dependent on time and depth in the plant canopy. Ray-tracer outputs during the whole day were compared to the values corresponding to direct light in the absence of any shading and light patterns subdivided into a series of time points separating sunlit and shaded periods. These then were used to draw space-time diagrams of shading history over each leaf surface. The sets of time points, evaluated for canopies with different leaf area indices (LAI), were used to model the switching times as events arising from a non-homogeneous Poisson process (NHPP) [16]. We found a strong sunlit-shading switching dependence on depth and decreasing length of sunlit periods with increasing planting density. We suggest that this simplified approach can be applied to physiological studies where economical descriptions of 3-D light dynamics are required, complementing more complex ray tracing procedures.

bioRxiv preprint doi: https://doi.org/10.1101/147553; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Materials and Methods

Digital canopy reconstruction and ray tracing

To investigate differences in sunlit and shaded patterns between wheat canopies with varying LAI, and among different lines, we used field-grown wheat plants which have structures reconstructed as described in [1]. The reconstructions in [1] represent the surface of a plant with a large number, N, of small triangular patches. Figure 1A shows an example of a reconstructed wheat plant, with an individual leaf at the lower part of the plant shown in blue. A particular triangular patch indexed, say, by j is defined by the set of coordinates $\{\mathbf{x}_{1j}, \mathbf{x}_{2j}, \mathbf{x}_{3j}\}$ of its three vertices. We denote the centroid of this patch by $\bar{\mathbf{x}}_j = (1/3)\sum_{i=1}^{3} \mathbf{x}_{ij}$, and its normalised height by

$$h_j = \frac{(\bar{\mathbf{x}}_j)_3 - z_{\min}}{z_{\max} - z_{\min}},$$

in which $(\bar{\mathbf{x}}_j)_3$ denotes height from the ground of the jth patch's centroid, and $z_{\min} = \min_{ij}\left\{(\mathbf{x}_{ij})_3\right\}$, $z_{\max} = \max_{ij}\left\{(\mathbf{x}_{ij})_3\right\}$ are respectively the minimum and maximum heights amongst all the vertices in the canopy. The models developed later involve dependence on these normalised heights.

The light distribution inside a 3D canopy was computed using the fastTracer v3.0 [18] implementation of a forward ray-tracing algorithm. The software fastTracer simulates three categories of light (direct, diffuse and scattered light), and determines where individual rays of light are eventually absorbed on leaf surfaces. For this study we have considered only direct light. Figure 1B shows a configuration for the ray tracer software [18]. Rays are arranged over a grid above the plant. The direction and amplitude of each ray depends on latitude and time of day. Ray tracing is performed in a cubic domain with periodic boundary conditions on the vertical faces so that when a ray exits one boundary of the domain it re-enters on the opposite vertical face. Latitude was set to 53° (for Sutton Bonington, UK), atmospheric transmittance 0.5, light scattering 7.5%, light transmittance 7.5%, day 182 (1 July). Those parameters were chosen for ray-tracing as that was the location and day that the wheat plants were taken and imaged in accordance with Burgess et al. (2015). We calculated the direct light intercepted during the day at 1 minute resolution for every patch in the canopy. The high temporal resolution enabled us to investigate even short-term light fluctuations in the canopy. See Figure 1E for an example of the light pattern computed for a particular patch. The high amplitude envelope, shown in red in Figure 1E,

bioRxiv preprint doi: https://doi.org/10.1101/147553; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

depends on the angle between the surface of the patch and the straight line path between the patch and the sun. Periods of shading occur when this path is blocked by a leaf or stem higher in the canopy, as illustrated in Figure 1C.

To construct canopies *in silico* based on the individual-plant reconstructions of [1], we exploited the periodic boundary conditions of the ray-tracer which give a natural way to "tile" individual plants to form an effective canopy. We investigated two ways to do this: (i) by putting the bounding box just outside the plant (as shown by the red rectangle in Figure 1D); or (ii) arranging plants on 3×3 square lattice a distance d apart, and putting the bounding box through the centres of boundary plants (as shown by the blue rectangle in Figure 1D). The periodic boundary conditions mean that case (i) amounts to considering an infinite square lattice of identical copies of the same plant. Case (ii) is similar, but it introduces additional heterogeneity through randomising the orientation of the different plants. We positioned the plants at distances d equal to 200mm, 150mm, 125mm and 100mm. In case (ii), we analysed the light dynamics for the plant in the centre of the 3×3 lattice. By changing configurations we constructed canopies with different LAI.

Constructing sunlit-shaded patterns

To construct sunlit-shaded patterns for each patch, we compared values of direct light computed by the ray-tracing algorithm [18] to the direct light irradiance on a unit surface in the absence of any shading [4]. Details of the latter equations are given in the Appendix; these equations are given in terms of latitude, day of year, time of day, and coordinates of three vertices; the latter are used to calculate the angle between a light ray and the normal to the trianglular patch. Figure 1E shows the direct light pattern for a particular patch in the lower part of a plant, obtained from ray-tracing software (black curve) and calculated from the equations in the Appendix (red curve). We designate a patch at a given time point as being shaded if the absolute difference between two values of the direct light was more then 10% (grey vertical lines). The shaded periods are indicated in Figure 1E by vertical grey bars. The substatial shaded period between 10 and 11 o'clock shown in Figure 1E is a consequence of the shading shown in Figure 1C. The output of the ray-tracer, used as an input to the models below, are time series with binary states (sunlit or shaded) charaterising the light pattern for every individual patch in the canopy.

bioRxiv preprint doi: https://doi.org/10.1101/147553; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

We can visualise spatiotemporal variation in the light patterns for different patches in the canopy via diagrams as shown in Figure 1F. Each row in this diagram corresponds to a particular patch in the canopy, with the patches (and hence rows) having been ordered according to the normalised heights of the patch centroids. A horizontal blue line indicates a shaded period, and white indicates sunlit. A space-time diagram of a shading for the leaf shown in blue in Figure 1A is plotted in Figure 1F, with a particular direct light pattern from Figure 1E shown in red. The diagram reveals an intricate pattern, with shadows from the upper leaves moving along the surface of the leaf as the sun changes position in the sky.

Sunlit-shaded dynamics: Model 1

We use a simplified model (Model 1) to illustrate our methodology, before implementing the full two-state model (Model 2) below.

Within some time interval of interest, $t \in (0, T)$, a particular patch within the canopy is either sunlit or shaded. In this first model, we consider only a single patch and are interested in how the rate of switching (from sunlit to shaded or vice versa) changes with time, t. At this point we do not distinguish between sunlit-to-shaded and shaded-to-sunlit events: we denote the times of such switching events by $0 < v_1 < \cdots < v_n < T$ (we introduce this distinction in Model 2 below).

The assumption of Model 1 is that switching events arise from a non-homogeneous Poisson process (NHPP). This is a stochastic process defined in terms of an intensity function $\lambda(t) \geq 0$ such that, for $0 < \delta t \ll 1$,

$$\text{Prob}\left\{\text{event in interval } (t, t+\delta t)\right\} = \lambda(t)\delta t + O(\delta t^2), \tag{1}$$

where $O(\cdot)$ is the standard big-O asymptotic notation. A property of NHPP is that the probabilities of events in distinct intervals are independent. From this independence, together with equation (1), it follows that for any interval $(t, t+u) \subseteq (0, T)$,

$$\begin{aligned}\text{Prob}\left\{\text{no events in interval } (t, t+u)\right\} &= \exp\left\{-\int_t^{t+u} \lambda(t')\mathrm{d}t'\right\} \\ &= \exp\left\{-\Lambda(t+u) + \Lambda(t)\right\},\end{aligned} \tag{2}$$

where $\Lambda(t) = \int_0^t \lambda(t')\mathrm{d}t'$. Equation (2) is useful in the following section for constructing expressions needed for fitting the model to data.

bioRxiv preprint doi: https://doi.org/10.1101/147553; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fitting Model 1

The goal is to fit the model by estimating the intensity function $\lambda(t)$ based on an observed set of switching times $v_1, \cdots, v_n$. We will do so by the standard statistical approach of maximum likelihood estimation (MLE). This involves constructing the likelihood function, which is the probability (density) function evaluated at the observed realisation of switching times, but regarded as a function of the parameter $\lambda(t)$ to be estimated. The likelihood function is

$$\begin{aligned} L(\lambda(t)) = &\Big[\exp\{-\Lambda(v_1) + \Lambda(0)\}\Big] \left[\prod_{i=2}^{n} \exp\{-\Lambda(v_i) + \Lambda(v_{i-1})\}\right] \\ &\times \Big[\exp\{-\Lambda(T) + \Lambda(v_n)\}\Big] \left[\prod_{i=1}^{n} \lambda(v_i)\right]. \end{aligned} \tag{3}$$

Expression (3) is straightforward to derive by discretising the interval $(0, T)$ with increments of size δt and writing the likelihood as a product of factors (using independence of increments) with each factor either (1) or its complement, depending on whether or not the increment contains an event, then taking the limit $\delta t \to 0$. The four factors in square brackets in (3) can be interpreted as follows: the first factor is a contribution from having no events in the interval $(0, v_1)$, the second factor from having no events in (v_i, v_{i-1}) for $i = 1, \ldots, n$, the third from having no event in the interval (v_n, T), and the fourth is the contribution from the switching events occurring at times $v_1, \ldots, v_n$. Such intuition will be helpful when constructing the likelihood functions for Model 2, but in the present case expression (3) simplifies, owing to telescoping in the exponent, to

$$L(\lambda(t)) = \exp\{-\Lambda(T)\} \prod_{i=1}^{n} \lambda(v_i). \tag{4}$$

Maximising (4) with respect to an unrestricted $\lambda(t)$ is ill-posed (since the maximising $\lambda(t)$ would be unboundedly large at the switching instants $t = v_1, \cdots, v_n$, and zero elsewhere). We will hence impose restrictions on $\lambda(t)$ by defining a set of basis functions $\mathbf{g}(t) = (g_1(t), \cdots, g_p(t))$ such that

$$\lambda(t) = \mathbf{b}^\top \mathbf{g}(t),$$

is smooth function of t. Then the fitting problem becomes one of estimating the vector of parameters $\mathbf{b} = (b_1, \cdots, b_p)^\top$ by maximising the likelihood

bioRxiv preprint doi: https://doi.org/10.1101/147553; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

function for $\mathbf{b}$,

$$L(\mathbf{b}) = \exp\left\{-\mathbf{b}^\top \mathbf{G}(T)\right\} \prod_{i=1}^{n} \mathbf{b}^\top \mathbf{g}(v_i), \tag{5}$$

where $\mathbf{G}(t) = \int_0^t \mathbf{g}(t')\mathrm{d}t'$. We may equivalently, and more conveniently, maximise the log of the likelihood, which is

$$\ell(\mathbf{b}) \equiv \log L(\mathbf{b}) = -\mathbf{b}^\top \mathbf{G}(T) + \sum_{i=1}^{n} \log\left(\mathbf{b}^\top \mathbf{g}(v_i)\right). \tag{6}$$

Although there is no simple closed-form solution for $\mathbf{b}$, equation (6) is a concave function of $\mathbf{b}$, hence its maximum is unique and is easy to determine by numerical optimisation routines; see the Appendix for details.

Simulating from Model 1

It follows from (1) (see [16] for details) that the distribution function for the additional time until the next event occurs given that an event occurred at time v is

$$F_v(s) = 1 - \exp\left(-\Lambda(v+s) + \Lambda(v)\right) \tag{7}$$

The distribution (7) is easy to simulate, for example using the inversion method [16].

An algorithm to simulate a sequence of event times v_1, v_2, v_3, ... is as follows. let v_1 be a simulated value from the distribution F_0. Then let v_2 equal v_1 plus a simulated value from the distribution F_{v_1}. Continue in this way, letting v_{i+1} equal v_i plus a simulated value from the distribution F_{v_i} until $v_{i+1} > T$.

Model 1: an example of simulation and model fitting

To illustrate the above ideas we use a simple model with $g = (1, t, (t-6)^2)$, for which the intensity function $\lambda(t) = 3 + 0.05t - 0.075(t-6)^2$. Here time t is measured in hours over a $T = 12$hr period starting at 6am, so that $t = t_{md} \equiv 6$hr represents noon. A realisation of the model is shown in Figure 2A. The switching rate function is shown in grey in Figure 2B. In this example, switching increases from sunrise to sunset, reaches its maximum at noon, and decreases from noon to sunset.

If we proceed with MLE based only on the single realisation in Figure 2A, we obtain the estimated intensity function $\lambda(t) = 2.2 + 0.14t - 0.061(t -$

bioRxiv preprint doi: https://doi.org/10.1101/147553; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

6)[2] (green curve in Figure 2A). The estimated intensity function matches reasonably well with the time intensity function, but not exactly because the estimate is based on a fairly small amount of data. A characteristic of the MLE is that the estimate tends to get increasingly close to the true answer as the amount of data increases. This is illustrated by the blue and red curves in Figure 2B, which are the MLE based on data from multiple realisations.

Sunlit–shaded dynamics: Model 2

The main contribution of this paper is to extend Model 1 in two ways: (i) to incorporate distinct rate functions, $\lambda^{\text{on}}(t)$ and $\lambda^{\text{off}}(t)$, for switching "on" (from shaded to sunlit) and "off" (from sunlit to shaded), respectively; and (ii) to describe multiple patches, with the rate functions for different patches depending on the normalised height, h, within the canopy (in addition to time, t, as in Model 1).

Extension (i) requires that we distinguish notationally between on-switching events, at times denoted x_i, and off-switching events, at times denoted y_i. For a given patch, since on- and off-switching events must alternate, the switching is characterised by the ordered set of times $\{x_1, y_1, x_2, y_2, \ldots, x_n, y_n\}$. The possibility that the state is initially "on" at time 0 is represented by having $x_1 < 0$, and similarly being "off" at time T is represented by $y_n > T$. The particular values of x_1 and y_n in these cases do not need to be specified. Besides these exceptions, we otherwise assume that $0 < x_i < y_i < T$ for all i. Figure 3 illustrates the notation, with the four different examples showing the four possible cases involving the different combinations of "on" and "off" states at $t = 0$ and $t = T$.

Fitting Model 2

In terms of the switching times, $\{x_1, y_1, x_2, y_2, \ldots, x_n, y_n\}$, for a given patch, the likelihood functions for $\lambda^{\text{on}}(t)$ and $\lambda^{\text{off}}(t)$ are then

$$\begin{aligned} L(\lambda^{\text{on}}) = \prod_{i=2}^{n} & \exp\left\{-\Lambda^{\text{on}}(x_i) + \Lambda^{\text{on}}(y_{i-1})\right\} \lambda^{\text{on}}(x_i) \\ & \times \left[\exp\left\{-\Lambda^{\text{on}}(x_1)\right\} \lambda^{\text{on}}(x_1)\right]^{I(x_1>0)} \\ & \times \left[\exp\left\{-\Lambda^{\text{on}}(T) + \Lambda^{\text{on}}(y_n)\right\}\right]^{I(y_n<T)} \end{aligned} \tag{8}$$

bioRxiv preprint doi: https://doi.org/10.1101/147553; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and

$$
\begin{aligned}
L(\lambda^{\text{off}}) = \prod_{i=2}^{n-1} & \exp\left\{-\lambda^{\text{off}}(y_i) + \lambda^{\text{off}}(x_i)\right\} \lambda^{\text{off}}(y_i) \\
& \times \exp\left\{-\Lambda^{\text{on}}(y_1)\right\} \lambda^{\text{off}}(y_1) \left[\exp\left\{\lambda^{\text{off}}(x_1)\right\}\right]^{I(x_1>0)} \\
& \times \exp\left\{-\Lambda^{\text{off}}(T) + \Lambda^{\text{off}}(x_n)\right\} \\
& \times \left[\exp\left\{-\Lambda^{\text{off}}(y_n) + \Lambda^{\text{off}}(T)\right\} \lambda^{\text{off}}(y_n)\right]^{I(y_n<T)}
\end{aligned} \tag{9}
$$

Here $I(\cdot)$ denotes the indicator function which equals one if its argument is true and zero otherwise. Equations (8,9) generalize Equations (3, 4) to distinguish sunlit-to-shaded from shaded-to-sunlit events.

The final step is to generalise to multiple patches, and introduce dependence of the rates on the height of each patch. Let $j = 1, \ldots, m$ index the different patches, and suppose that quantities specific to the jth patch are indicated by a suffix j. We assume $x_{1,j} < 0$ and $y_{n_j,j} > T$ if the state is "on" at the beginning and end, respectively, of the interval $(0, T)$. We let h_j denote the height of the jth patch and use subscripts on the rate functions to denote their dependence on height, i.e., the rate functions for the jth patch are $\lambda^{\text{on}}_{h_j}(t)$ and $\lambda^{\text{off}}_{h_j}(t)$.

Assuming independence of patches, the likelihoods can be constructed as a product of factors of the form (8) or (9) over index $j = 1, \ldots, m$. The log-likelihoods are then

$$
\begin{aligned}
\ell(\lambda^{\text{on}}) = \sum_{j=1}^{m} & \left\{ \sum_{i=2}^{n} -\Lambda^{\text{on}}_{h_j}(x_{i,j}) + \Lambda^{\text{on}}_{h_j}(y_{i-1,j}) + \log \lambda^{\text{on}}_{h_j}(x_{i,j}) \right\} \\
& + I(x_{1,j} > 0) \left\{ -\Lambda^{\text{on}}_{h_j}(x_{1,j}) + \log \lambda^{\text{on}}_{h_j}(x_{1,j}) \right\} \\
& + I(y_{n_j,j} < T) \left\{ -\Lambda^{\text{on}}_{h_j}(T) + \Lambda^{\text{on}}_{h_j}(y_{n_j,j}) \right\}
\end{aligned} \tag{10}
$$

and

$$
\begin{aligned}
\ell(\lambda^{\text{off}}) = \sum_{j=1}^{m} & \left\{ \sum_{i=2}^{n-1} -\Lambda^{\text{off}}_{h_j}(y_{i,j}) + \Lambda^{\text{off}}_{h_j}(x_{i,j}) + \log \lambda^{\text{off}}_{h_j}(y_i) \right\} \\
& - \Lambda^{\text{off}}_{h_j}(y_{1,j}) + \log \lambda^{\text{off}}_{h_j}(y_{1,j}) + I(x_{1,j} > 0) \Lambda^{\text{off}}_{h_j}(x_{1,j}) \\
& - \Lambda^{\text{off}}_{h_j}(T) + \Lambda^{\text{off}}_{h_j(x_{n_j,j})} \\
& + I(y_{n_j,j} < T) \left\{ -\Lambda^{\text{off}}_{h_j}(y_{n_j,j}) + \Lambda^{\text{off}}_{h_j}(T) + \log \lambda^{\text{off}}_{h_j}(y_{n_j,j}) \right\}.
\end{aligned} \tag{11}
$$

bioRxiv preprint doi: https://doi.org/10.1101/147553; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Once again, we need to specify particular models for the rate functions. We will consider models of the factorised forms

$$\lambda_h^{\text{on}}(t) = (\mathbf{a}^{\text{on}})^\top \mathbf{f}(h) \times (\mathbf{b}^{\text{on}})^\top \mathbf{g}(t), \tag{12}$$

$$\lambda_h^{\text{off}}(t) = (\mathbf{a}^{\text{off}})^\top \mathbf{f}(h) \times (\mathbf{b}^{\text{off}})^\top \mathbf{g}(t), \tag{13}$$

where $\mathbf{f}(h) = (f_1(h), \ldots, f_q(h))$ and $\mathbf{g}(t) = (g_1(t), \ldots, g_p(t))$ are vectors of prescribed basis functions. Fitting the model by maximum likelihood estimation then requires maximising the log-likelihood functions (10) and (11) with respect to parameter vectors $\left((\mathbf{a}^{\text{on}})^\top (\mathbf{b}^{\text{on}})^\top\right)^\top$ and $\left((\mathbf{a}^{\text{off}})^\top (\mathbf{b}^{\text{off}})^\top\right)^\top$ respectively. Some constraints on these parameter vectors are required in order that they can be uniquely identified from data: see the Appendix for discussion.

Simulating from Model 2

For the model, we make a distinction between whether at time $t = 0$ a patch is a sunlit or shaded state. In simulations, we used random starting states, distributed as

$$\text{Prob}\left\{\text{sunlit at } t = 0; h\right\} = h, \tag{14}$$

where h is the normalised height of the patch's centroid, so that patches high in the canopy tend to start off sunlit whereas those at the bottom tend to start shaded. It is easy to reproduce (14) with a different distribution, for example and empirical distribution determined from ray-tracing data.

For this model, the distribution for the time until next event depends on whether switching from sunlit to shaded or vice versa. Analogous to (7), we denote the distribution for time to next "on" event given an "off" event occurred at time x_1 by F_x^{on}; and the time to the next "off" event, given an "on"event occurred at time y_1 by F_y^{off}.

An algorithm to simulate from Model 2 is thus as follows. Simulate the initial state as either sunlit or shaded. Supposing it is sunlit, let x_1 be a simulated value from F_0^{off}. Then let y_1 equal x_1 plus a simulated value from $F_{x_1}^{\text{on}}$. Continue letting x_{i+1} equal y_i plus a simulated value from $F_{y_i}^{\text{off}}$ and y_{i+1} equal x_{i+1} plus a simulated value from $F_{x_i}^{\text{on}}$ until either $x_{i+1} > T$ or $y_{i+1} > T$.

If the initial state were instead shaded, then the only difference in the above would be in the first step, in which we would simulate y_1 from F_0^{on}, then the algorithm proceeds in the same way.

The R source code for the emulator is available on request and will be incorporated into an R package together with digital reconstructions of plants.

bioRxiv preprint doi: https://doi.org/10.1101/147553; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Results

Results and analysis of fitting Model 2 to ray-tracing data

We assembled canopies with varying LAI using reconstructions of high spatial-resolution images of wheat plants. Figure 4 shows images of the canopies we analysed: the original plant; the original plant rotated 90°, 180°, and 270°; the original plant randomly rotated and positioned at distances $d = 200$mm, 150mm, 125mm, and 100mm from a replica of a plant from the same line; and plants from two different lines.

Each of the configurations can be described by the individual profile of cumulative LAI (cLAI), which shows how leaf mass is accumulated from the top to the bottom of a canopy. Figure 4 shows the profiles of cLAI for all canopies we used for the analysis. Here canopies (A)-(D) and (I)-(J) correspond to configurations where we put the bounding box just outside the plant (as shown by red rectangle in Figure 1 D); and canopies (E)-(H) are constructed by arranging plants on a 3×3 square lattice at a distance d apart. The cLAI profiles for canopies (A)-(D) are identical, as rotation of the plant does not change the spatial distribution of leaf mass.

For the choice of basis functions in the rates (12, 13) we use $\mathbf{f}(h) = (1, h)^\top$ and $\mathbf{g}(t) = (1, (t - t_{md}^2)^\top$, where $t_{md} = 6$hr is the time at midday. In other words we assume that the switching rates have linear dependence on height, h, and parabolic dependence on t. The switching rates are then:

$$\lambda_h^{\text{on}}(t) = (1 - a^{\text{on}} h) \left(b_1^{\text{on}} + b_2^{\text{on}} (t - t_{md})^2 \right), \tag{15}$$

$$\lambda_h^{\text{off}}(t) = (1 - a^{\text{off}} h) \left(b_1^{\text{off}} + b_2^{\text{off}} (t - t_{md})^2 \right), . \tag{16}$$

We estimate the parameters $\{a^{\text{on}}, b_1^{\text{on}}, b_2^{\text{on}}, a^{\text{off}}, b_1^{\text{off}}, b_2^{\text{off}}\}$ for the ray-tracing data by maximum likelihood estimation, as described above. Table 1 shows the fitted parameters for the different canopies we considered. Also shown in the table is the LAI for each canopy computed directly from the reconstruction data. Notable from the table is that a^{on} and a^{off} are both close to 1. This indicates that the switching rates are very strongly dependent on depth within the canopy. In particular, a^{off} tends to be very close to 1, so that the off rate at the very top of the canopy is close to zero, since patches are not obstructed by other leaves and hence are permanently sunlit. Typically the parameters b_1^{off} and b_2^{off} contributing to the off rate are larger for canopies with higher LAI and the corresponding parameters b_1^{on} and b_2^{on} in the on rate are slightly smaller. This is consistent with the intuition that in dense canopies sun flecks are typically shorter and less frequent.

bioRxiv preprint doi: https://doi.org/10.1101/147553; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The information in Table 1 can be visualised in two dimensions using Principal Component Analysis (PCA). We performed PCA based on the correlation matrix of the fitted parameters. The first PC, which explains 59% of the variability, has loadings $(0.08, 0.51, 0.51, -0.50, -0.2, -0.43)$ suggesting the interpretation that it contrasts the off and on rates, or in other words that it is a measure of typical "shadedness" within the canopy. Indeed the first PC correlates very strongly with LAI (Pearson correlation coefficient of 0.984, p-value $< 10^{-7}$) as shown in Figure 4. The second PC, which explains 20% of the variability has loadings $(0.37, 0.18, -0.06, 0.26, 0.75, -0.44)$. Here b_1^{on} is dominant in this PC, and contrasts with b_2^{on} suggesting that this PC measures overall on rate and the shape of its time dependence (note, for example, that in canopies F,G, and H b_2^{on} is negative indicating the on rate is lower at midday than at the start or end of the day, in contrast to the other canopies).

The plot in Figure 4, of PC1 versus PC2, shows some very clear clustering of canopies we expect to be similar, suggesting that the PCs (and the fitted parameters from which they were computed) encode physiologically relevant information about the canopies.

Simulating realistic dynamic direct light flux density

Model 2 describes only binary sunlit-shaded dynamics. The direct light flux density during sunlit periods is given by A_{dr} defined in (22) in the Appendix. We can hence model the direct light flux density for a particular patch as

$$\text{flux density } (t) = A_{dr}(t) I(\text{patch sunlit at time t }) \tag{17}$$

in which A_{dr} depends on patch orientation, as well as latitude and day of the year, as described in the Appendix.

We compare the simulations from this model to ray-tracing data from canopy H. Figure 5A shows a comparison for individual patches between patterns generated by the ray-tracer (left column) and emulator (right column) at the top, middle and bottom of the canopy. In order to compare emulator performance over a whole canopy, we calculated the daily light intercepted per unit leaf area at each patch from light patterns produced by the ray-tracer (Figure 5B) and the emulator (Figure 5C). As can be seen from the figures, the emulator gives good results at a fraction of the computational cost. The differences can be explained by the assumption of spatial

bioRxiv preprint doi: https://doi.org/10.1101/147553; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

independence of patches for the emulator, in contrast to the ray-tracer, where neighbouring patches will be shaded at similar times.

Discussion

In this work, the effect of shading has been quantified by analysing patterns of intercepted direct light in complex canopies. A novel model was developed which provides an alternative to traditional ray-tracing that is more accessible and requires fewer resources to implement. The emulator approach, based on a two-state Markov process model, is computationally more efficient than a ray-tracing algorithm, especially when dealing with large canopies and inferring light profiles at a high temporal and special resolution. The computationally cheap way to simulate light dynamics we have proposed will be important for feeding into larger-scale photosynthesis models for calculation of the absorbed radiation and photosynthetic production by a canopy, and so useful for crop-productivity models in addition to leaf-level physiological analyses.

This is also of importance in the new field of plant and crop phenotyping where high resolution reconstructions can now be developed routinely but bottlenecks exist in terms of their analysis for physiological functioning. For example, using the reconstructions available in Pound *et al.* [11], running the ray-tracer FastTracer3 to provide data for a 24 hour period can take several days. In comparison the Markov emulator takes less then a minute to generate an individual direct light pattern without the need to run calculations for all of the canopy. The emulator can produce a comparable set of data that can be used to analyse light dynamics and compare to photosynthetic responses. This provides a convenient and high resolution approach for characterizing the canopy light environment where physical sensors cannot be used since they do not provide adequate spatial resolution and obstruct light. Such sunlit-shaded dynamics can be used in order to analyse multiple aspects of crop cultivation such as varietal selection and altered architectural characteristics, cultivation practice (such as cropping system, row spacing etc) or for photosynthetic modelling.

There are many natural extensions to the modelling we have introduced. We developed the simulation and fitting sections in terms of general basis function, but in this paper have considered only one simple special case, which amounts to imposing that the switching rate functions (15, 16) depend linearly on h and parabolically on t. Particularly with very large data sets

bioRxiv preprint doi: https://doi.org/10.1101/147553; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

that arise from ray-tracing studies, there is scope for investigating much more flexible basis functions, for example spline bases, and for considering more general forms for the rate functions than the factorised forms (12, 13) that we have so far assumed. The maximum likelihood framework in which we have developed the model fitting naturally extends to model selection, which will enable formal comparison between different candidate models (i.e., as defined through different choices of basis function) for the empirical data. Our emulator focuses specifically on direct light, with the sunlit/shaded state modelled stochastically and the amplitude during sunlit periods modelled by a deterministic light amplitude envelope function. We could straightforwardly further introduce effects of diffused and scattered light, perhaps as deterministic components contributing additively to the total incident light flux.

The heterogeneity of the light environment influences how plants respond to and exploit available resources, i.e. photosynthesis and crop-production. This has been recently demonstrated using recovery from photoprotection [7]. However, to quantify the impact of a particular photosynthetic process (Rubisco activation, stomatal opening, photoprotection) on productivity requires knowledge of the precise 'signature' of light - shade dynamics. For example longer periods in high light and low light are likely to be less productive than rapid fluctuations [14, 2]. Hence the emulator described in this paper represents a significant step towards rapid prediction of the impact on dynamic photosynthesis that a particular canopy architecture is likely to have.

Here we have quantified the mechanistic nature of light variability by analysing diurnal light patterns in canopies with different LAI. The results of the current study highlight and quantify a negative relationship between height and the expected length of sunlit periods and a positive relationship between LAI and the expected length of sunlit periods. To fully explore light environment in leaf canopy, we intend to expand the emulator to simulate dffiused and scattered light. These sources of irradiance will depend mostly on leaf aggregation, i.e. patches at the same level of cumulative LAI will be exposed to similar values of diffused and direct light [5]. A similar methodology as we used for direct light characterisation can be applied to determine the type of dependence of diffused and direct light on the height.

bioRxiv preprint doi: https://doi.org/10.1101/147553; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Acknowledgments

R.R., O.E.J., S.P.P., and E.H.M. are supported by the Biotechnology and Biological Sciences Research Council (grant BB/J003999/1).

Literature cited

[1] AJ. Burgess, R Retkute, MP. Pound, J Foulkes, SP. Preston, OE. Jensen, TP. Pridmore, and EH. Murchie. High-Resolution Three-Dimensional Structural Data Quantify the Impact of Photoinhibition on Long-Term Carbon Gain in Wheat Canopies in the Field. *Plant Physiology*, 169(2):1192, 1–12, 2015. doi: 10.1104/pp.15.00722.

[2] AJ. Burgess, R Retkute, SP. Preston, OE. Jensen, MP. Pound, TP. Pridmore, and EH. Murchie. The 4-Dimensional Plant: Effects of Wind-Induced Canopy Movement on Light Fluctuations and Photosynthesis. *Front. Plant Sci.*, 7, 2016. doi: 10.3389/fpls.2016.01392.

[3] AJ Burgess, R Retkute, T Herman, and EH Murchie. Exploring Relationships between Canopy Architecture, Light Distribution, and Photosynthesis in Contrasting Rice Genotypes Using 3D Canopy Reconstruction. *Front. Plant Sci.*, 8(734):1–15, 2017. doi: 10.3389/fpls.2017.00734.

[4] JS Cambell and JM Normal. *An Introduction to Environmental Biophysics*. Springer, New York, 1998.

[5] J Cavazzoni, T Volk, F Tubiello, and O Monje. Modelling the effect of diffuse light on canopy photosynthesis in controlled environments. In Lieth, JH and Oki, LR, editor, *Proceedings of IV International Symposium on Models for Plant Growth and Control in Greenhouses: Modeling for the 21st Century - Agronomic and Greenhouse Crop Models*, number 593 in Acta Horticulturae, pages 39–45, 2002. doi: 10.17660/ActaHortic.2002.593.4.

[6] JH Kim, JW Lee, Tae In Ahn, JH Shin, KS Park, and JE Son. Sweet Pepper (Capsicum annuum L.) Canopy Photosynthesis Modeling Using 3D Plant Architecture and Light Ray-Tracing. *Front. Plant Sci.*, 7, 2016. doi: 10.3389/fpls.2016.01321.

bioRxiv preprint doi: https://doi.org/10.1101/147553; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[7] J Kromdijk, K Glowacka, L Leonelli, ST. Gabilly, M Iwai, KK. Niyogi, and SP. Long. Improving photosynthesis and crop productivity by accelerating recovery from photoprotection. *Science*, 354(6314):857–861, 2016. doi: 10.1126/science.aai8878.

[8] T Lawson and MR. Blatt. Stomatal Size, Speed, and Responsiveness Impact on Photosynthesis and Water Use Efficiency. *Plant Physiology*, 164(4):1556–1570, 2014. doi: 10.1104/pp.114.237107.

[9] GA Maddonni, ME Otegui, B Andrieu, M Chelle, and JJ Casal. Maize leaves turn away from neighbors. *Plant Physiology*, 130(3):1181–1189, 2002. doi: 10.1104/pp.009738.

[10] RW Pearcy and WM Yang. A three-dimensional crown architecture model for assessment of light capture and carbon gain by understory plants. *Oecologia*, 108(1):1–12, 1996. doi: 10.1007/BF00333208.

[11] MP Pound, AP French, EH Murchie, and TP Pridmore. Automated recovery of 3D models of plant shoots from multiple colour images. *Plant Physiology*, 2014. doi: 10.1104/pp.114.248971.

[12] M Qu, S Hamdani, W Li, S Wang, J Tang, Z Chen, Q Song, M Li, H Zhao, T Chang, C Chu, and X Zhu. Rapid stomatal response to fluctuating light: an under-explored mechanism to improve drought tolerance in rice. *Functional Plant Biology*, 43(8):727–738, 2016. doi: 10.1071/FP15348.

[13] R Retkute, SE. Smith-Unna, RW. Smith, AJ. Burgess, OE. Jensen, GN. Johnson, SP. Preston, and EH. Murchie. Exploiting heterogeneous environments: does photosynthetic acclimation optimize carbon gain in fluctuating light? *Journal of Experimental Botany*, 66(9):2437–2447, 2015. doi: 10.1093/jxb/erv055.

[14] JS Roden and RW Pearcy. Photosynthetic gas exchange response of poplars to steady-state and dynamic light environments. *Oecologia*, 93 (2):208–214, 1993.

[15] J Ross, M Sulev, and P Saarelaid. Statistical treatment of the PAR variability and its application to willow coppice. *Agricultural and Forest Meteorology*, 91(1-2):1–21, 1998. doi: 10.1016/S0168-1923(98)00066-5.

bioRxiv preprint doi: https://doi.org/10.1101/147553; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[16] SM. Ross. *Simulation, Fourth Edition*. Academic Press, Inc., Orlando, FL, USA, 2006. ISBN 0125980639.

[17] H Sinoquet, S Thanisawanyangkura, H Mabrouk, and P Kasemsap. Characterization of the light environment in canopies using 3D digitising and image processing. *Annals of Botany*, 82(2):203–212, 1998. doi: 10.1006/anbo.1998.0665.

[18] Q Song, G Zhang, and XG Zhu. Optimal crop canopy architecture to maximise canopy photosynthetic CO2 uptake under elevated CO2 - a theoretical study using a mechanistic model of canopy photosynthesis. *Functional Plant Biology*, 40(2, SI):109–124, 2013. doi: 10.1071/FP12056.

[19] S Vialet-Chabrand, JSA Matthews, AJ Simkin, CA Raines, and T Lawson. Importance of fluctuations in light on plant photosynthetic acclimation. *Plant. Phys.*, 2017. doi: 10.1104/pp.16.01767 .

[20] K Zhou, X Deng, X Yao, Y Tian, W Cao, Y Zhu, SL Ustin, and T Cheng. Assessing the Spectral Properties of Sunlit and Shaded Components in Rice Canopies with Near-Ground Imaging Spectroscopy Data. *Sensors*, 17(3):E578, 2017. doi: 10.3390/s17030578.

Appendix

Direct light

Solar ray direction depends on the time and location: day, d, hour, t, and latitude, l_a [4]. The sun declination angle is

$$s = -23.5 \cos\left(\frac{d+10}{365}\right). \tag{18}$$

The hour angle is

$$h_{ang} = 0.262(t - t_m), \tag{19}$$

where t_m is time of solar noon. The solar elevation angle is

$$\theta_s = \arcsin\left(\sin(s)\sin(l_a) + \cos(s)\cos(l_a)\cos(h_{ang})\right). \tag{20}$$

bioRxiv preprint doi: https://doi.org/10.1101/147553; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The solar azimuth angle is

$$\varphi_s = -\arccos\left(\frac{\sin(s)\cos(l_a) - \cos(h_{ang})\cos(s)\sin(l_a)}{\cos(s)}\right) + \pi. \tag{21}$$

Then the direction of direct light is given by a vector (x, y, z) with

$$\begin{aligned} x &= r\cos(\varphi_s), \\ y &= \pm r\sin(\varphi_s), \\ z &= -\cos(\pi - \theta_s); \\ r &= \sin(\pi - \theta_s). \end{aligned}$$

Here $y > 0$ if $t < 6$hr, and $y \geq 0$ if $t \geq 6$hr.

Direct light irradiance on a unit surface is given by

$$A_{dr} = A_s \alpha^{\frac{1}{\sin(\theta_s)}} \cos(\beta), \tag{22}$$

where α is atmosphere transmittance, A_s is a solar constant ($= 2600\mu$ mol $m^{-2}s^{-1}$), and β is the angle between the light ray and the normal to the surface.

Concavity of Model 1's log-likelihood function

Although there is no closed-form solution for b as the optimal argument that maximises (6), the optimum can nevertheless be computed easily by numerical optimisation, because (6) is a concave function. This can be seen by computing the derivatives

$$\frac{\partial \ell}{\partial \mathbf{b}} = -\mathbf{G}(T) + \sum_{i=1}^{n} \left\{\mathbf{b}^\top \mathbf{g}(v_i)\right\}^{-1} \mathbf{g}(v_i), \tag{23}$$

$$\frac{\partial^2 \ell}{\partial \mathbf{b} \partial \mathbf{b}^\top} = -\sum_{i=1}^{n} \left\{\mathbf{b}^\top \mathbf{g}(v_i)\right\}^{-2} \mathbf{g}(v_i)\mathbf{g}(v_i)^\top. \tag{24}$$

Since the Hessian (24) is negative semidefinite, the objective function (6) is concave. Together with the concave constraint $\lambda(t) = \mathbf{b}^\top \mathbf{g}(t) \geq 0$, for which $\mathbf{b}^\top \mathbf{g}(t)$ is concave (indeed affine), the numerical optimisation for estimating $\mathbf{b}$ is hence a concave maximisation. This means that the global optimum is the unique optimum, so numerical optimisation routines will not become trapped in local optima. Some numerical algorithms for such task benefit from an analytic expression for the gradient (23).

bioRxiv preprint doi: https://doi.org/10.1101/147553; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Comments on fitting Model 2

Some comments on the model and fitting are as follows. Comments focus on $\lambda_h^{\text{on}}(t)$ but apply analogously to $\lambda_h^{\text{off}}(t)$.

1. The model specified by (12) does not lead to uniquely identifiable parameter values since, for example, transformations $\mathbf{a}^{\text{on}} \to c \cdot \mathbf{a}^{\text{on}}$ and $\mathbf{b}^{\text{on}} \to c^{-1} \cdot \mathbf{b}^{\text{on}}$ lead to an unchanged $\lambda_h^{\text{on}}(t)$. To overcome this, without loss of generality we assume that $(\mathbf{a}^{\text{on}})_1 = 1$.
2. The log-likelihood function (10) is not in general a concave function of the concatenated parameter vector $\left((\mathbf{a}^{\text{on}})^\top (\mathbf{b}^{\text{on}})^\top\right)^\top$. A possible strategy for numerical maximisation is to iterate between maximising with respect to $\mathbf{a}^{\text{on}}$ with $\mathbf{b}^{\text{on}}$ fixed, and maximising with respect to $\mathbf{b}^{\text{on}}$ with $\mathbf{a}^{\text{on}}$ fixed. In this case each maximum is the concave. However, for the calculations in this paper we found that using the Nelder-Mead algorithm to optimise directly with respect to the concatenated vector reliably found the global maximum.

bioRxiv preprint doi: https://doi.org/10.1101/147553; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ID	LAI	a^{on}	b_1^{on}	b_2^{on}	a^{off}	b_1^{off}	b_2^{off}
A	0.69	0.85	3.83	0.031	1.	1.75	0.099
B	0.69	0.85	4.19	0.043	1.	1.54	0.10
C	0.69	0.85	4.40	0.017	1.	2.12	0.087
D	0.69	0.85	4.35	0.057	1.	2.01	0.099
E	1.15	0.79	3.29	0.063	0.94	1.92	0.24
F	2.11	0.62	3.32	-0.017	0.99	4.83	0.38
G	2.89	0.54	3.17	-0.021	0.99	5.58	0.43
H	4.18	0.34	2.14	-0.026	1.	8.29	0.55
I	1.23	0.78	1.73	0.047	1.	1.36	0.072
J	0.76	0.66	2.39	0.096	1.	1.053	0.19
K	1.32	0.64	1.77	0.051	0.99	1.33	0.16

Table 1: Parameter estimation for the model Eq.(15)-(16). Here ID corresponds to the canopy configuration shown in Fig.4

bioRxiv preprint doi: https://doi.org/10.1101/147553; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

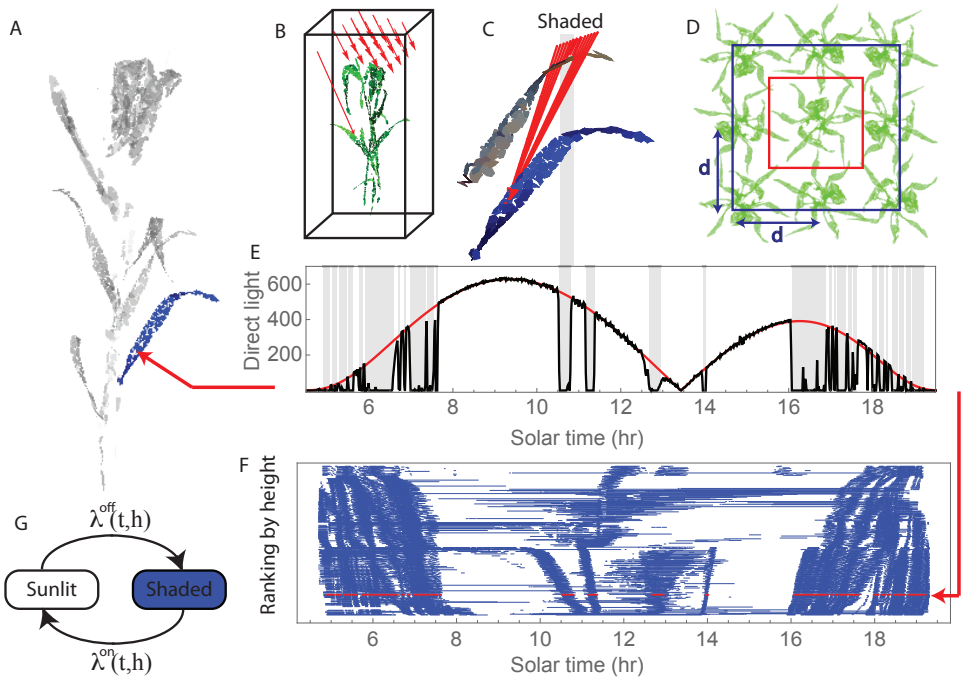

Figure 1: Quantifying sunlit and shaded dynamics. (A) Reconstructed wheat plant from [1]. (B) Set up for a ray tracer [18]. Red arrows show direct light; once a ray hits the boundary, it is moved to the opposed vertical face of the box. (C) Shading will occur when a ray is obstructed by other leaves or a stem. (D) Construction of canopies was done in two ways: putting the bounding box just outside the plant (red rectangle) or putting plants on 3×3 grid at a distance d apart and putting the bounding box through the centres of boundary plants (blue rectangle). (E) Diurnal dynamics of light at a particular triangle: ray-tracer simulation (solid black curve), light amplitude envelope (solid red curve) and inferred shading periods (horizontal grey lines). Time resolution is 1 minute. One of the shading periods is extended to (C) to indicate actual obstruction of the light ray. (F) Spatiotemporal dynamics of shaded periods on a leaf shown in blue in (A). All triangles were ranked according to their centre, each line corresponds to a shaded period. Red lines indicate shaded periods from the pattern shown in (E). (G) Two-state Markov model: for switching on (from shaded to sunlit) with rate $\lambda^{\text{on}}(t,h)$ and for switching off (from sunlit to shaded) with a rate $\lambda^{\text{off}}(t,h)$.

bioRxiv preprint doi: https://doi.org/10.1101/147553; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

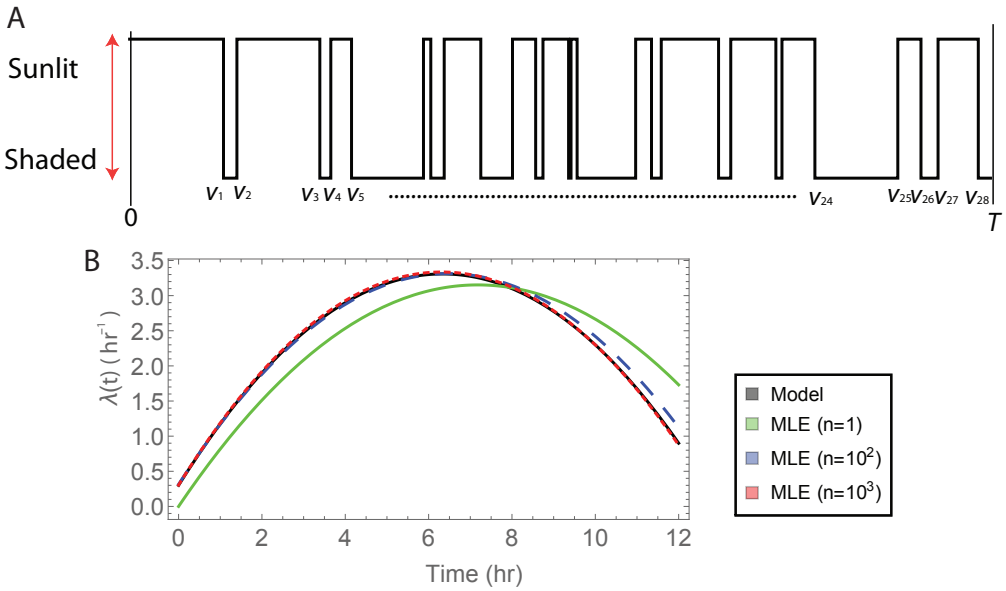

Figure 2: Maximum likelihood estimation for Model 1. (A) A realisation of a two-state Markov model with $\lambda^{\text{on}}(t) = \lambda^{\text{off}}(t) = 3 + 0.05t - 0.075(t-6)^2$. (B) The shape of switching rate, showing convergence of the MLE to the true value with an increasing number n of realisations.

bioRxiv preprint doi: https://doi.org/10.1101/147553; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

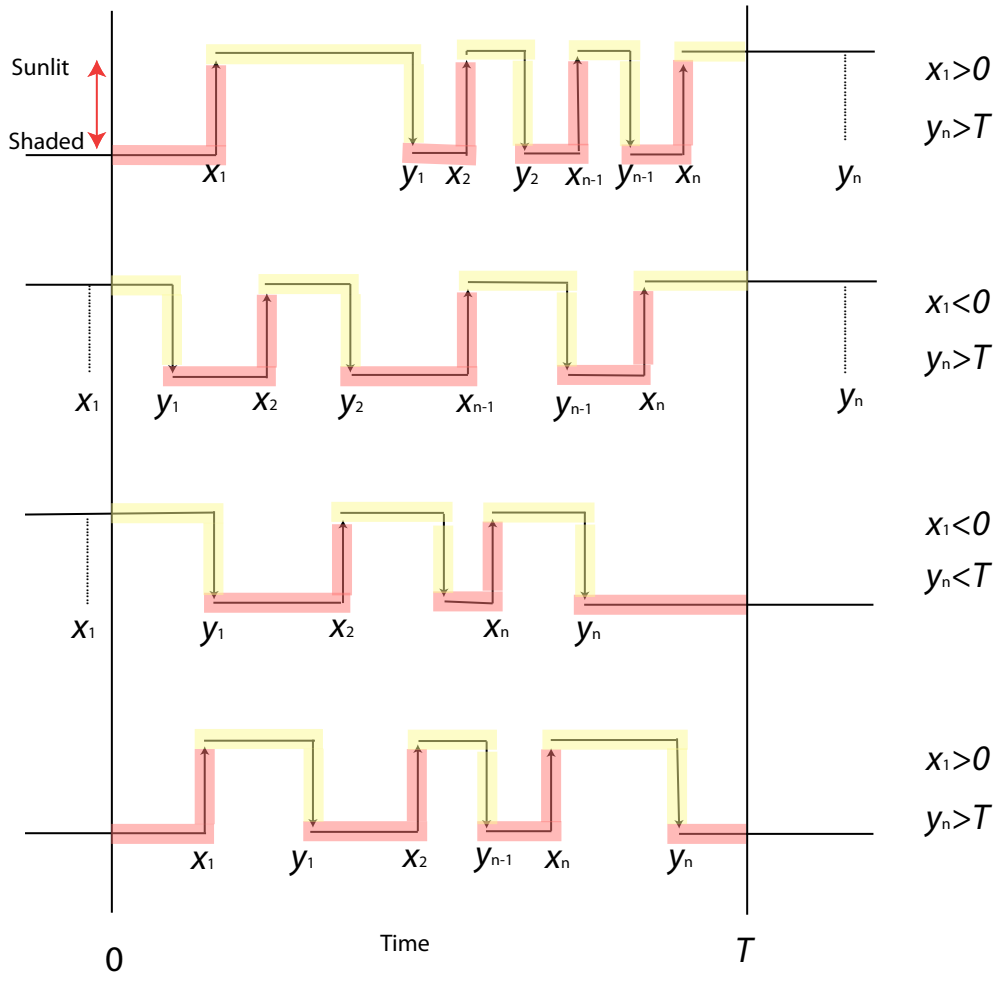

Figure 3: Illustration of the notation for Model 2. Shown are the four possible combinations of states at the beginning and end of the interval $[0, T]$. At time $t = 0$ sunlit is indicated by $x_1 < 0$ and shaded by $x_1 > 0$; time $t = T$ sunlit is indicated by $y_n > T$ and shaded by $y_n < T$. The different sections are colored to indicate how they contribute to the (log) likelihood functions: red denotes contribution to on-switching functions (7,9) and yellow to off-switching function (8,10).

bioRxiv preprint doi: https://doi.org/10.1101/147553; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

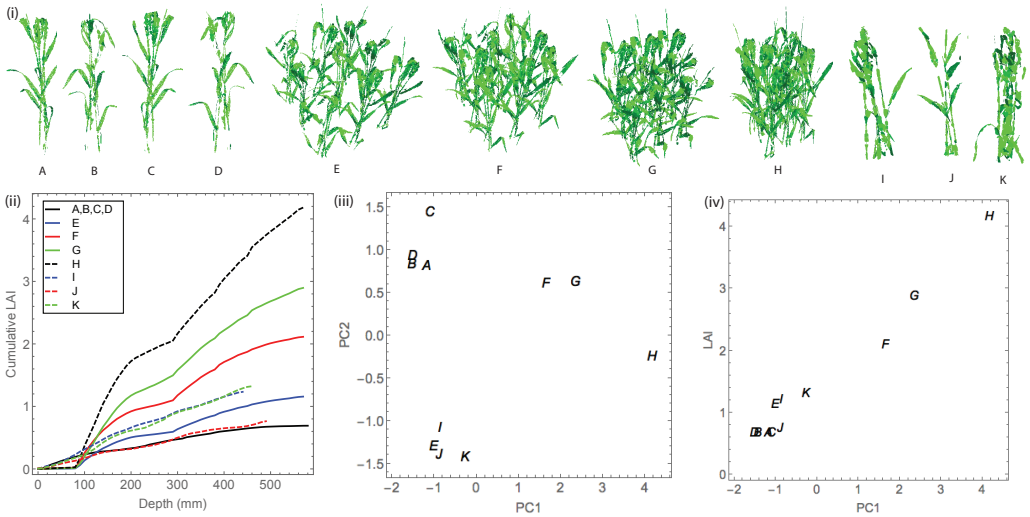

Figure 4: Relationship between the two-state Markov model and canopy physiology: (i) Reconstructed plants (A-K), (ii) cumulative leaf area index as a function of depth (bottom left), (iii) principal component analysis of fitted parameters, and (iv) relationship between the first principal component and LAI. Images of original plant (**A**); original plant rotated 90° (**B**), 180° (**C**), and 270° (**D**); original plant randomly rotated and positioned at distances 200mm (**E**), 150mm (**F**), 125mm (**G**), and 100mm (**H**); replica of a plant from the same line (**I**); plants from two different lines (**J** and **K**). For more detailed description of the lines used and information on the trial see Burgess et al. (2015).

bioRxiv preprint doi: https://doi.org/10.1101/147553; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

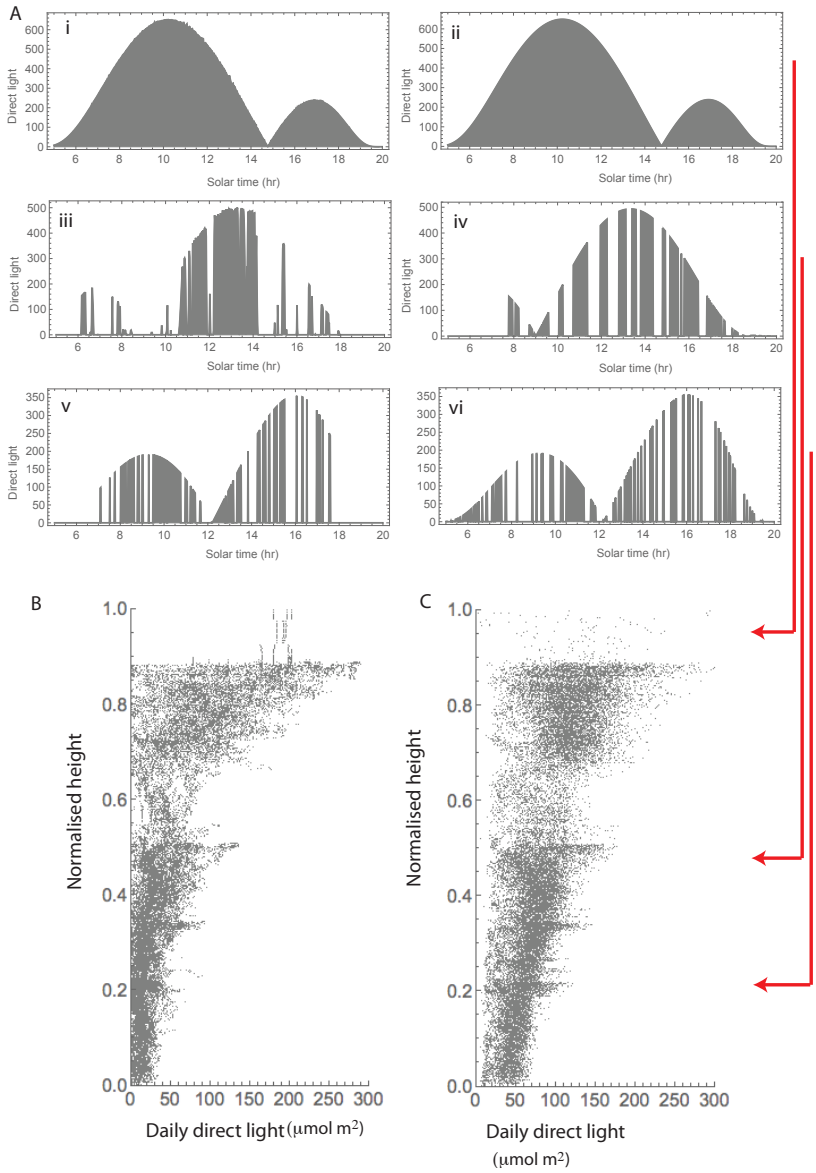

Figure 5: Results comparing ray tracer and emulator. Analysis and parameter values are for the reconstructed canopy shown in Figure 4 H. (A) Direct light pattern (left column) vs emulator (right column) for patches at top (i,ii), middle (iii, iv) and bottom (v, vi) of the canopy. Daily light intercepted per unit leaf area at each patch calculated from light patterns from obtained using the ray tracer (B) and generated by the emulator (C).