

# Analytical expressions of variability in ecosystem structure and function obtained from three-dimensional stochastic vegetation modelling

Stephen P. Good, I. Rodriguez-Iturbe and K. K. Caylor

*Proc. R. Soc. A* 2013 **469**, 20130003, published 8 May 2013

---

## References

**This article cites 41 articles, 4 of which can be accessed free**  
<http://rspa.royalsocietypublishing.org/content/469/2155/20130003.full.html#ref-list-1>

## Subject collections

Articles on similar topics can be found in the following collections

[atmospheric science](#) (18 articles)

[hydrology](#) (6 articles)

[mathematical modelling](#) (88 articles)

## Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click [here](#)

## Research



CrossMark  
click for updates

**Cite this article:** Good SP, Rodriguez-Iturbe I, Caylor KK. 2013 Analytical expressions of variability in ecosystem structure and function obtained from three-dimensional stochastic vegetation modelling. *Proc R Soc A* 469: 20130003.  
<http://dx.doi.org/10.1098/rspa.2013.0003>

Received: 3 January 2013

Accepted: 8 April 2013

### Subject Areas:

hydrology, atmospheric science, mathematical modelling

### Keywords:

vegetation, structure, evapotranspiration, stochastic geometry

### Author for correspondence:

Stephen P. Good

e-mail: [sgood@princeton.edu](mailto:sgood@princeton.edu)

# Analytical expressions of variability in ecosystem structure and function obtained from three-dimensional stochastic vegetation modelling

Stephen P. Good, I. Rodriguez-Iturbe and K. K. Caylor

Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ 08544, USA

Whole ecosystem exchange of water, carbon and energy is predominately determined by complex leaf-level processes occurring at individual plants. Interaction between individuals results in a distribution of environmental conditions that drive a variety of nonlinear response functions such as transpiration and photosynthesis. The nonlinearity of biophysical processes requires higher-order statistical descriptions of micro-environment distributions in order to accurately determine the landscape-scale mean functional response. We present a mathematical framework for describing vegetation structure based on the density, dispersion, size distribution and allometry of individuals within a landscape. Using three-dimensional stochastic vegetation modelling, we develop analytic expressions of the second-order statistics of vegetation canopies, namely the mean and variance of leaf area density and leaf area index with height. These expressions also allow for the approximation of the distribution of beam penetration and sunfleck statistics through the canopy as a function of height. Finally, we demonstrate how landscape-scale fluxes are strongly affected by the variability in canopy micro-environments, and how stochastic vegetation modelling improves flux estimates relative to traditional homogeneous canopy models.

## 1. Introduction

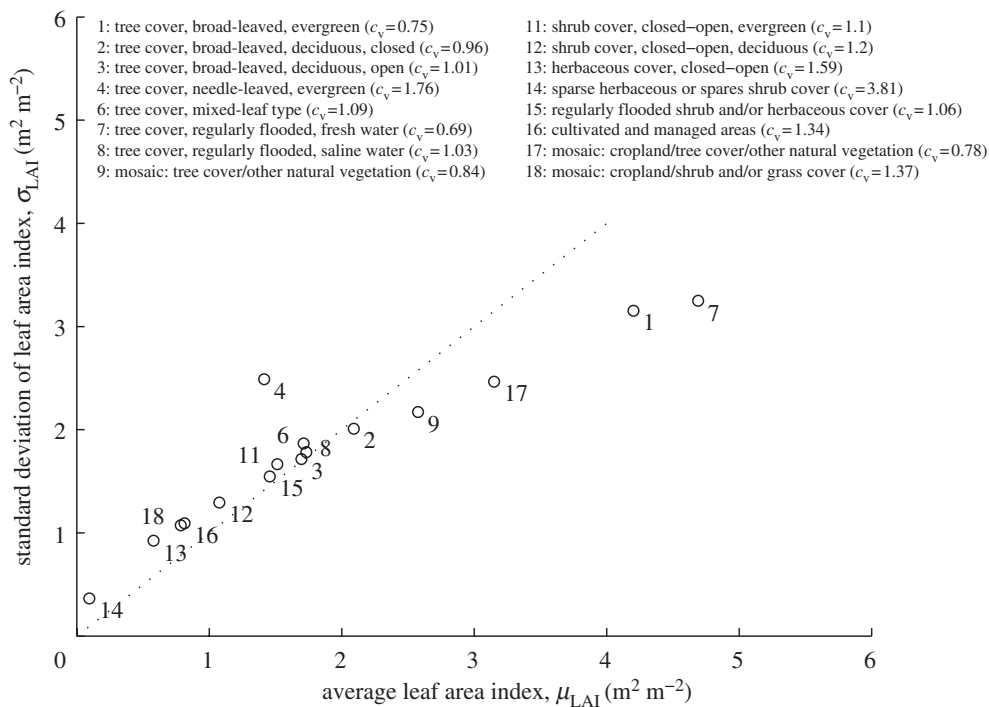
Estimates of the landscape-scale exchange of water, carbon and energy are a necessity for accurate understanding of the Earth's geological and biological

processes. The biophysical functionality of vegetated environments is therefore mathematically represented via ecosystems models. Within different ecosystem models, the spatial resolution of plant canopy architecture spans many degrees of complexity. However, it is increasingly understood that accurate description of both the vertical [1–3] and horizontal [4–7] structural dimensions is required to correctly model biophysical functionality. Accurate models of vegetation canopies that can be directly parameterized with information about the abundance and allometric properties of individuals are required to enable a rapid estimation of biophysical exchange in global climate models and at specific, intensely studied, field sites [6,8].

The description of land–atmosphere interactions is inherently influenced by the effects of scale choice. Spatial patterns in rainfall, soil properties and topography can all drive variability in ecohydrological processes, and the schematization of a surface for model representation often prescribes uniform conditions within cells [9]. In the case of heterogeneous canopies, at scales below 10 ha, variability in vegetation has been shown to significantly impact the mean field hydrological processes. Conversely, at scales larger than 100 ha, the spatial structure of climatology has been shown to control variability [10]. The distributions of soil properties and topographic relief also influence vegetation patterns and fluxes [11], and the scale at which geomorphologic features are manifest will drive variability at these spatial extents. Thus, at the field scale, where much ecohydrological field research and modelling efforts are directed, heterogeneity in vegetation structure is expected to be the dominant source of variability.

Vegetation canopies are typically modelled in two distinct manners: one highly efficient and the other highly precise. The first common approach assumes canopy structure to be horizontally homogeneous throughout a given area, and is typically referred to as ‘big-leaf’ modelling. The second approach involves explicitly specifying the location and characteristics of each individual within the area to be studied [7]. The assumption of a homogeneous field implies that the physiological response of the landscape to the average condition is equal to the average of the responses to a distribution of conditions. The classic proof by Jensen [12] demonstrated the inequality of the response to the mean and the mean of the response for nonlinear functions:  $f(\mathbb{E}[X]) \neq \mathbb{E}[f(X)]$ , when  $f(X)$  is a convex function of  $X$  (bold letters are used to represent random variables throughout the text). Ecophysiological research has shown that leaf photosynthesis and transpiration are strongly nonlinear functions of micro-environment radiation, temperature and humidity [13]. In the example case of transpiration, the phase change of liquid water to vapour in stomata is governed by the availability of a finite amount of incoming radiation; thus as leaf area at a location becomes more dense, increases in transpiration gains become smaller owing to energy limitation. In environments characterized by sparse vegetation, foliage is typically clumped into areas of varyingly dense leaf area separated by patches of bare ground [14], and the assumption of a homogeneous canopy environment with a single average leaf area will result in unreliable estimates of landscape-scale fluxes [6,7]. Natural landscapes span a wide range in both the average amount of leaf area,  $\mu_{LAI}$  ( $\text{m}^2 \text{m}^{-2}$ ), and the standard deviation of the average leaf area,  $\sigma_{LAI}$  ( $\text{m}^2 \text{m}^{-2}$ ), as shown in figure 1. More complex modelling approaches are required to accurately model surface-to-atmosphere interactions as variability in leaf area increases relative to total leaf area.

In an effort to accurately resolve landscape-scale biophysical fluxes, research has led to individual-based ecosystem models. In these models, the explicit location, size, canopy geometry and leaf physiology of each individual is simulated in a three-dimensional space. The MAESTRO model [18] and/or the coupled atmosphere–canopy Forest Light Environment Simulator (FLiES) [19] are examples of this class of model. Once the three-dimensional environment is fully specified, energy and mass transfer within the canopy are directly resolved. In the case of FLiES, a large number of individual photons ( $10^5$ – $10^8$ ) are injected into the simulated canopy, after which their path and energy status is tracked as they are reflected, transmitted and absorbed by canopy elements. The large amount of input data required, as well as the computational demands of full three-dimensional modelling, limit the applicability of these models at regional and continental scales. Another approach used to represent subgrid-scale heterogeneity is to divide areas into ‘patches’ or ‘mosaics’ that are then modelled as separate



**Figure 1.** Average and within-pixel standard deviation of leaf area index for various land cover types [15] in Africa based on moderate resolution imaging spectroradiometer (MODIS) 2006 data [16]. The value of  $\sigma_{LAI}$  is derived from the clumping index [17] (cf. §3). Land cover types with lower  $\mu_{LAI}$  values have larger variability in leaf area relative to their mean, with the coefficient of variability  $c_v$  larger than one (dotted line). Areas of higher  $\mu_{LAI}$  have less variability, with  $c_v$  less than one. Numbers reference GLC 2000 [15] land cover classifications.

entities and aggregated [20]. Finally, probability distributions of micro-climates have been used to model subgrid-scale variability [21,22]; yet, parameterizing these distributions in ecosystem models remains a difficulty.

Stochastic canopy and radiation transfer models do not require the explicit simulation of all individuals, yet are able to estimate the average conditions present in a heterogeneous system. These models, pioneered by Estonians Nilson and co-workers [23–25] and Ross and co-workers [26,27], assume that individuals are randomly distributed on a landscape. Using the basic assumption of spatial randomness, foliage area and light penetration profiles are estimated based on statistical geometry. The geometrical optical radiation transfer [28], analytical clumped two-stream [8] and canopy semi-analytical  $P_{gap}$  and radiative transfer [29,30] models are state-of-the-art radiation transfer models able to predict the light environment vertically through the canopy. Radiation transfer schemes within these models are focused on light attenuation through the canopy and predict the average probability ( $p_{gap}$ ) that light will penetrate to a given depth. These models, with the exception of those of Nilson, typically assume that individuals are randomly distributed following a spatial Poisson process on a landscape. The spatial Poisson assumption does not take into account the underdispersion of ordered systems such as crop stands and forestry plantations, nor the clumping observed in sparsely vegetated systems. The clumping of leaves into crowns is typically considered via an effective leaf area index or clumping factor,  $\Omega(z, \theta)$ , which is used to relate true leaf area to observed average light penetration [31].

The radiation transfer equations of Shabanov *et al.* [32,33] use both the probability of encountering vegetation elements and the pair-correlation distance between elements to resolve the mean field vertical radiation regime within heterogeneous canopies. This approach is based on the stochastic radiative theory of Vainikko [34,35], which was developed for addressing radiation

transfer through discontinuous clouds. The case of identically sized and shaped cylindrical vegetation elements ordered into canopies separated by gaps with a Poisson distribution in space was numerically investigated with the stochastic mixture radiative transfer model and found to significantly alter radiative fluxes relative to the classic turbid medium approach [33]. The work of Huang *et al.* [36] incorporated pair-correlation functions for both over- and underdispersed vegetation patterns. Huang *et al.* [36] clearly demonstrated that mean representations of canopy vegetation are not sufficient to accurately represent the mean landscape canopy reflectance.

The majority of the earlier-mentioned models are focused on light penetration, and variability is typically considered with respect to gap probabilities, not in the distribution of leaf area itself. No models yet present analytical solutions describing the mean and variability in leaf area as a function of height based on abundance, dispersion, size distributions and allometric parameters. Representation of the statistical distribution of foliage is necessary because ecosystem fluxes are functions of both the canopy micro-environment (e.g. the profile of incident photosynthetically active radiation) and the amount of foliage experiencing these conditions. In this study, we present a framework for modelling the three-dimensional structure of vegetation that mathematically represents complex, heterogeneous canopies with minimal parameters and assumptions. We provide the analytical solution of the mean and variance of leaf area density,  $L_{AD}$  ( $\text{m}^2 \text{m}^{-3}$ ), and leaf area index,  $L_{AI}$  ( $\text{m}^2 \text{m}^{-2}$ ), as a function of landscape density, clustering level, size distribution and allometric relationships. With the canopy leaf area characterized, expressions for the landscape fractional cover, gap frequency probability density function as well as the empirical clumping factor  $\Omega$  are analytically derived. Finally, we use the distribution of  $L_{AI}$  to predict the distribution of evapotranspiration and demonstrate how spatial variability in foliage alters the average surface–atmosphere flux of water.

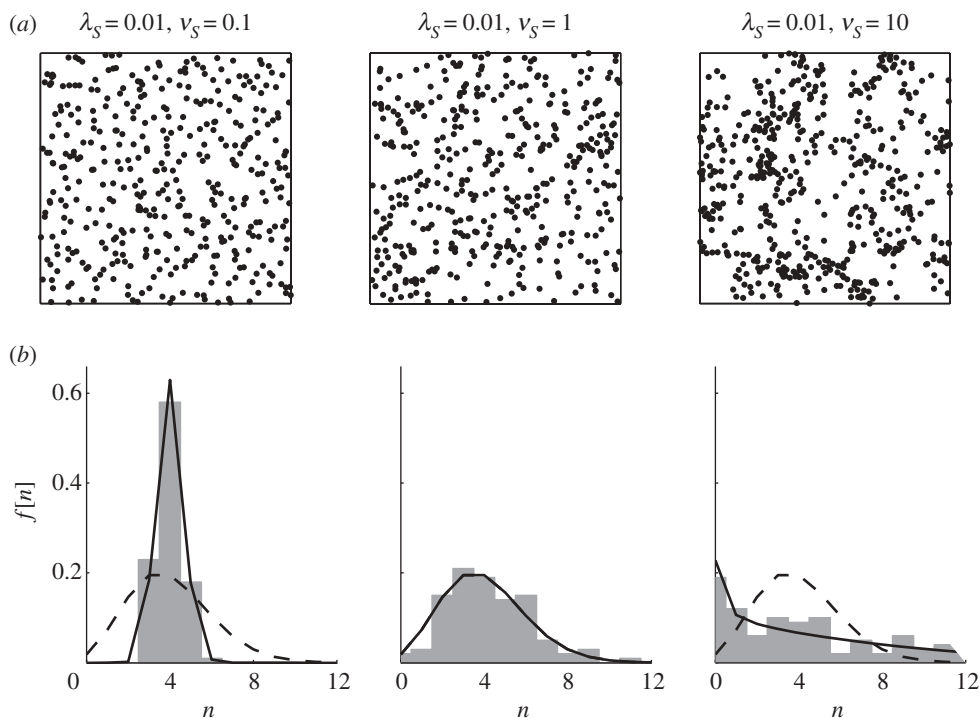
## 2. Representation of individuals on a stochastic field

The spatial distribution of individuals is described by the average number of individuals,  $\lambda_S$  ( $\text{m}^{-2}$ ), and their dispersion,  $\nu_S$  ( $\text{m}$ ), on a subplot  $S$  of the two-dimensional plane. Each individual on the landscape is assumed to possess a height randomly chosen from a known distribution characterized by a mean,  $\mu_H$  ( $\text{m}$ ), and variance,  $\sigma_H^2$  ( $\text{m}^2$ ). The canopy above each individual is assumed to conform to allometric relationships characterized by species-specific constants for canopy width ( $c_W$ ), canopy thickness ( $c_T$ ) and canopy foliage ( $c_{L_1}$  and  $c_{L_2}$ ). We directly derive the analytical expressions for the mean and variance of leaf area density,  $L_{AD}$  ( $\text{m}^2 \text{m}^{-3}$ ), and leaf area index,  $L_{AI}$  ( $\text{m}^2 \text{m}^{-2}$ ), as a function of elevation  $z$ . The appendix describes the random variables, biophysical constants, derived values and other parameters referenced in this text.

### (a) Individual density and dispersion

The number and organization of individuals of a specific species or functional type in a subplot of size  $S$  is treated as a random variable,  $N$ . Often, the spatial organization of individuals is assumed to conform to a two-dimensional Poisson process, and the distribution of  $N$  therefore follows a Poisson probability mass function [37]. The Poisson distribution is a single-parameter distribution where the average number of individuals,  $\lambda_S$ , and variance are equal. However, in more ordered systems, such as a crop fields, forestry plantation or situations where individuals strongly inhibit the growth of their neighbours, the variance in the number of individuals per area is much lower than a Poisson process. These underdispersed systems may be characterized by a binomial distribution [24,38]. Conversely, when vegetation is organized into clumped patterns, as is found in sparsely vegetated savannas [14], the variance in the number of individuals in an area  $S$  may be significantly higher than in a Poisson process. These overdispersed systems may be described by a negative binomial distribution [38].

To accurately represent the variety of landscapes observed in nature, a two-parameter discrete distribution bound at zero with adjustable variance is required. Efron [39] proposed a two-parameter double Poisson distribution,  $\sim DP(\lambda, \nu)$ , as a way to model both overdispersed (e.g.



**Figure 2.** Realizations of stochastic landscape simulations with the same landscape density ( $\lambda_S = 0.01$ ) and scale parameter ( $S = 20^2$ ), but with varying degrees of dispersion ( $\nu_S = 0.1, 1, 10$ ). A  $200 \times 200$  m region is shown in (a). (b) Normalized histograms (grey areas),  $f[n]$  of the count of individuals,  $n$ , within an area  $S$  and comparison of the simulation results to the Poisson (dashed line) and double Poisson (solid line) probability mass functions.

an ordered lattice) and underdispersed data (e.g. clumped points) [39]. The probability that the random number of individuals,  $N$ , will be equal to a number  $n$  is given by the probability mass function

$$f_{\text{DP}(S\lambda_S, \nu_S)}(n) = \frac{\nu_S^{-1/2} e^{-S\lambda_S/\nu_S}}{c_{\text{DP}}} \left( \frac{e^{-n} n^n}{n!} \right) \left( \frac{eS\lambda_S}{n} \right)^{n/\nu_S}, \quad (2.1)$$

where  $S\lambda_S$  and  $\nu_S$  are the parameters determining the location and shape of  $f_{\text{DP}(\lambda, \nu)}(n)$  [39]. The normalizing constant  $c_{\text{DP}}$  is defined such that  $\int_0^\infty f_{\text{DP}(\lambda, \nu)}(n) dn = 1$ . An approximation of  $c_{\text{DP}}$  is given [39], but  $c_{\text{DP}}$  is also easily determined numerically. The mean and variance of  $f_{\text{DP}(S\lambda_S, \nu_S)}(n)$  given over a region  $S$  are

$$\mathbb{E}[N] = \lambda_S S \quad \text{and} \quad \text{Var}[N] = \nu_S \lambda_S S. \quad (2.2)$$

The parameter  $\nu_S$  determines the amount of dispersion present in the region  $S$  and thus also the broadness of the distribution of  $f_{\text{DP}(\lambda, \nu)}(n)$ , with  $\nu_S < 1$  relating to underdispersed data,  $\nu_S = 1$  describing complete spatial randomness, and  $\nu_S > 1$  relating to overdispersed data. When  $\nu_S$  is equal to one, equation (2.1) simplifies to the Poisson distribution.  $\nu_S$  is directly analogous to the Cox index of clumping used to describe the level of dispersion in measured subplots in forests [40]. Figure 2 shows three simulations of landscapes all with the same values of  $\lambda_S$  and  $S$ , but with different values of  $\nu_S$ . Simulations were generated by dividing the area into a gridded lattice with square areas of size  $S$ . A random number of individuals drawn from the double Poisson distribution was assigned to each lattice box, and then each individual was randomly placed within the area. As shown in the bottom of figure 2, the double Poisson distribution is able to accurately represent the normalized histograms of simulated arrangements in space over a wide range of dispersion conditions.

## (b) The Poisson random case

We begin with the case of complete spatial randomness, i.e.  $\nu = 1$ , and derive expressions for the statistics of interactions of individuals. In the case of foliage density and area, interaction (i.e. overlapping) of two or more individuals is an additive process. Assuming radial symmetry of all individuals, let  $g(x; u, \mathbf{H})$  represent the influence of an individual located at  $u = (u_1, u_2)$  with the random set of characteristics  $\mathbf{H}$  at position  $x = (x_1, x_2)$  in the landscape. At this point, it is not necessary to state the form of  $g(x; u, \mathbf{H})$ , nor do we need to define the random variable  $\mathbf{H}$ . The total influence  $Y(x)$  of all elements in the landscape  $\mathbb{R}^2$  at position  $x$  is given by the sum of all contributions,

$$Y(x) = \sum_{u \in \mathbb{R}^2} g(x; u, \mathbf{H}) = \int_{\mathbb{R}^2} g(x; u, \mathbf{H}) dN(u), \quad (2.3)$$

where  $N(u)$  is a double Poisson stochastic process over the two-dimensional plane with characteristics given by (2.2). As noted by Rodriguez-Iturbe *et al.* [41], the cumulates of  $Y$  are obtained by Taylor series expansion where, subject to convergence, the  $m$ th cumulant  $\kappa_m(Y)$  is the coefficient of  $p^m/m!$  [41]; therefore,

$$\kappa_m(Y) = \int_{\mathbb{R}^2} \lambda_S \mathbb{E}[g(x; u, \mathbf{H})]^m du, \quad (2.4)$$

which is a form of Campbell's theorem [42]. The above results may also be derived directly with the fact that  $\mathbb{E}[dN(u)] = \text{Var}[dN(u)] = \lambda_S du$  for the Poisson random case [41].

The random variables  $\mathbf{H}$  and  $N$  are considered to be completely independent and identically distributed in this framework. Using the properties of  $dN(u)$ , we reformulate the results in polar coordinates, with  $r$  as the Euclidean distance from  $x$  to  $u$ . The expected value of all contributions at a random location on the stochastic landscape is

$$\mathbb{E}[Y] = \lambda_S \int_0^\infty \mathbb{E}[2\pi r g(r, \mathbf{H})] dr. \quad (2.5)$$

In a similar manner, the variance is given by

$$\text{Var}[Y] = \lambda_S \int_0^\infty \mathbb{E}[2\pi r g(r, \mathbf{H})^2] dr. \quad (2.6)$$

Equations (2.5) and (2.6) are used to estimate the first- and second-order statistics of the random process  $Y$ . Assuming that all individual elements are radially symmetric, of finite cylindrical form and homogeneous within themselves, we can complete the integration in (2.5) and (2.6) out to the element edge. This gives the element area,  $a(\mathbf{H})$ , with characteristics  $\mathbf{H}$ . Finally, we condition on the probability that a random height  $\mathbf{H}$  is equal to a value of  $h$ ,  $\mathbb{P}(h < \mathbf{H} < h + dh) \equiv f(h) dh$ . The expected value of  $Y$  is formally defined by rewriting equation (2.5) as

$$\mathbb{E}[Y] = \lambda_S \int_0^\infty a(\mathbf{H}|\mathbf{H} = h) g(\mathbf{H}|\mathbf{H} = h) f(h) dh. \quad (2.7)$$

In a similar manner, the variance of  $Y$  is formally stated by rewriting equation (2.6) as

$$\text{Var}[Y] = \lambda_S \int_0^\infty a(\mathbf{H}|\mathbf{H} = h) g(\mathbf{H}|\mathbf{H} = h)^2 f(h) dh, \quad (2.8)$$

where  $f(h)$  is a known probability distribution. Equations (2.7) and (2.8) are the core of the vegetation structure model and enable us to estimate the first and second central moments at the landscape scale of processes arising from the additive influence of a field of individuals.

## (c) Over- and underdispersed systems

Modelling individual dispersion with a lattice-based approach provides an analytically tractable representation of complex natural phenomena. Both clustered and thinned processes may be described with respect to a measure of the variance of their intensity,  $\nu$ , and their

appropriate length scale,  $S$ . Values of  $\nu$  and  $S$  may be obtained through direct surveys at field sites. Additionally, traditional processes such as Neyman–Scott clustering models and Matérn inhibition processes may be approximated by fitting  $\nu$  and  $S$  to the theoretical correlation structure or Ripley’s  $K$  function for a particular model [43].

We define  $\mathbb{E}[Y_\nu]$  and  $\text{Var}[Y_\nu]$  as the mean and variance of the process  $Y(x)$  with a dispersion parameter  $\nu_S \neq 1$ . For all cases, the expected number of individuals and their average influence at point  $x$  does not change from the Poisson random case above, therefore  $\mathbb{E}[Y_\nu] = \mathbb{E}[Y]$ . Beginning again at (2.3), and making use of the law of total variance, we use the double Poisson properties in (2.2) and derive the second cumulant of  $Y(x)$  as

$$\begin{aligned} \kappa_2(\mathbf{Y}) = \text{Var}[Y_\nu] &= \int_S \text{Var}[g(x; u, \mathbf{H})] \mathbb{E}[dN(u)] \\ &+ \int_S \mathbb{E}[g(x; u, \mathbf{H})]^2 \text{Var}[dN(u)]. \end{aligned} \quad (2.9)$$

In the Poisson case,  $\mathbb{E}[dN(u)] = \text{Var}[dN(u)]$ , (2.9) simplifies to (2.4) through the definition of variance, i.e.  $\text{Var}[g(x; u, \mathbf{H})] = \mathbb{E}[g(x; u, \mathbf{H})^2] - \mathbb{E}[g(x; u, \mathbf{H})]^2$ . If the definition of variance is used in the first term on the right-hand side of (2.9), then this term can be separated, and the first of the resulting three terms is equal to  $\text{Var}[Y]$ . The second and third terms are multiplied by  $\mathbb{E}[dN(u)]/\mathbb{E}[dN(u)]$  and factored to form  $\mathbb{E}[Y]^2$ . Equation (2.9) is then rewritten with respect to the moments of the Poisson random case as

$$\text{Var}[Y_\nu] = \text{Var}[Y] + \frac{\nu - 1}{\lambda_S S} \mathbb{E}[Y]^2, \quad (2.10)$$

where the  $\lambda_S S$  term in the denominator arises because  $\nu_S$  is defined at a particular scale, and thus integrated over the area  $S$  only. As equation (2.10) is integrated over  $S$  and not  $\mathbb{R}^2$ , the influence of adjacent lattice cells is not considered. As  $S$  approaches the areal extent of individual elements, i.e.  $S \leq a(\mathbf{H})$ , equation (2.10) becomes a poor predictor of the  $\text{Var}[Y_\nu]$ .

We compare simulations of stochastic systems with equation (2.10) in figure 3. A  $200 \times 200$  m grid with cells of size  $S$  was created, and within each lattice cell, a random number of marks drawn from a double Poisson probability mass function were placed uniformly. Each mark was assigned a value of  $\mathbf{H}$  drawn from a gamma distribution with a mean of 10 and a variance of 25. A circular region around each mark with an area proportional to the mark value,  $a(\mathbf{H}) = 0.6\mathbf{H}$ , was allocated an intensity proportional to the mark value,  $g(x; u, \mathbf{H}) = 0.1\mathbf{H}$ . Except in the case of low values of both  $S$  and  $\nu_S$ , equation (2.10) is able to capture changes in landscape variability accurately.

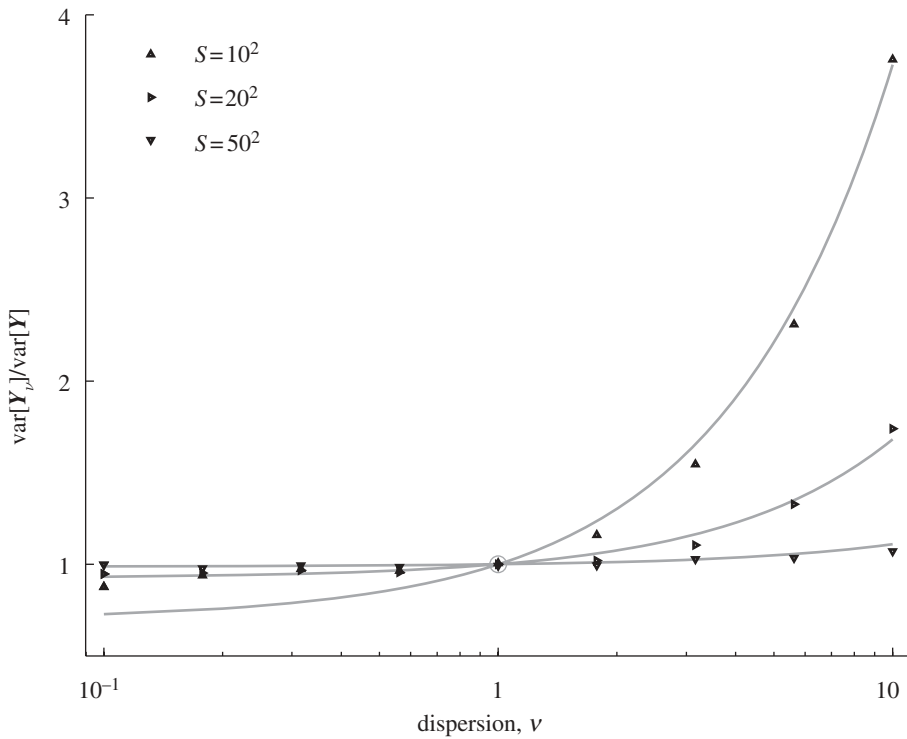
### 3. The stochastic three-dimensional canopy environment

The foliage canopies of all individuals combine to form the landscape vertical profile of leaf area, expressed as either leaf area density,  $L_{AD}$ , or leaf area index,  $L_{AI}$ . The spatial dimensions of individual canopies are characterized as functions of each individual’s height with simple allometric relationships. The stochastic double Poisson process in two-dimensional space is defined as a marked stochastic process where the value of each mark refers to the height of each individual. For representation of a three-dimensional canopy environment, we use equations (2.7) and (2.8) in the estimation of the mean and variance of leaf area in a vegetation canopy as a function of height  $z$ . Dispersion can then be addressed using equation (2.10).

#### (a) Allometric relationships

We have defined the density and dispersion of individuals on the landscape; next, we characterize the influence,  $g(\cdot)$ , that each individual exerts on the overall field. Individual canopies in the landscape take the form of vertical cylinders, though other forms, such as ellipsoids or cones, may also be constructed. These cylinders are filled uniformly with leaf area at a density specific to





**Figure 3.** Realizations of stochastic landscape simulations (black triangles) with varying degrees of dispersion compared with the theoretical result of equation (2.10) (grey lines). 100 separate realizations of  $200 \times 200$  m simulations were created to determine the variance of the process  $\mathbf{Y}$ . Individuals were simulated as in figure 2, and assigned a random mark  $\mathbf{H}$  from a gamma distribution with mean 10 and variance of 25. The area and intensity of each mark was proportional to  $\mathbf{H}$ , as  $a(\mathbf{H}) = 0.6\mathbf{H}$  and  $g(x; u, \mathbf{H}) = 0.1\mathbf{H}$ .

each height. The horizontal width,  $W$  (m), and vertical thickness,  $T$  (m), of cylindrical canopies are defined as

$$W = c_W \mathbf{H} \quad \text{and} \quad T = c_T \mathbf{H}, \quad (3.1)$$

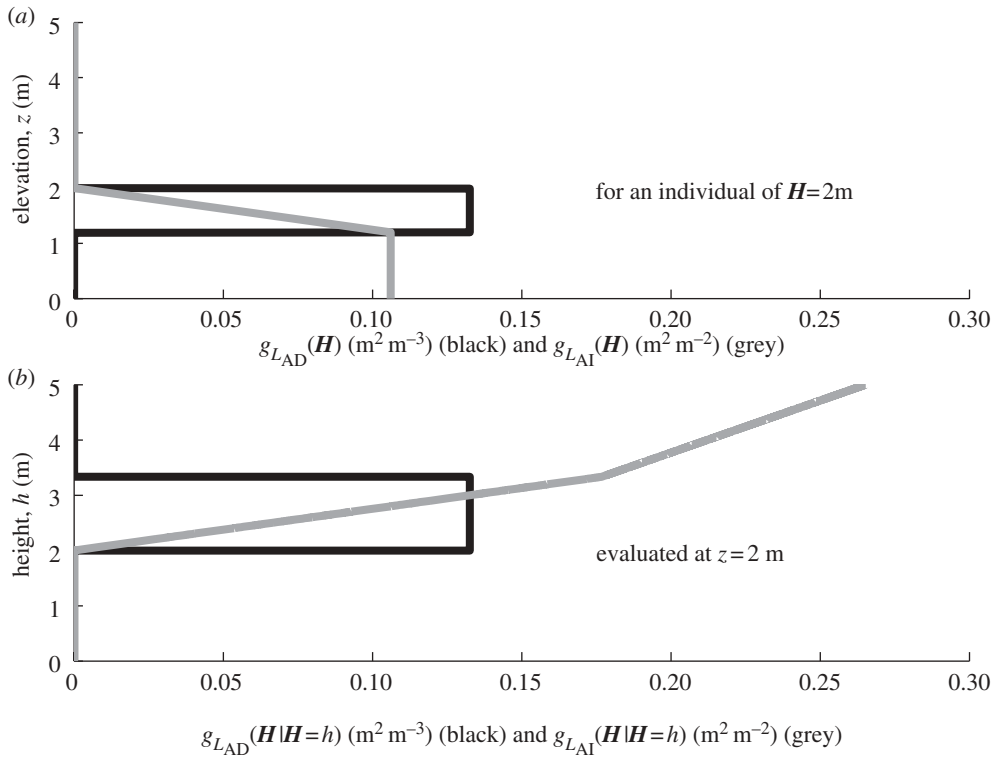
where  $c_W$  ( $\text{m m}^{-1}$ ) is a constant relating the height of an individual to its canopy width. Similarly,  $c_T$  ( $\text{m m}^{-1}$ ) is a constant relating the thickness of the canopy to the total height of the individual. The area under the canopy is  $A = (\pi/4)c_W^2 \mathbf{H}^2$ , and the volume of the cylindrical canopy is  $V = (\pi/4)c_T c_W^2 \mathbf{H}^3$ . The total foliage area present within an individual canopy,  $F$  ( $\text{m}^2$ ), is given by a simple power relationship to height as

$$F = c_{L_1} \mathbf{H}^{c_{L_2}}, \quad (3.2)$$

where  $c_{L_1}$  [ $\text{m}^{(2-c_{L_2})}$ ] and  $c_{L_2}$  [ ] are constants [27].

The canopy of a single individual is filled with foliage at a leaf area density of  $g_{LAD}(\mathbf{H})$  ( $\text{m}^2 \text{m}^{-3}$ ), which is equal to the total foliage area over the total volume,  $F/V$ . The function  $g(\mathbf{H})$  is then written with respect to elevation above the surface,  $z$ , where we note the fact that the function is zero when outside the canopy as

$$g_{LAD}(\mathbf{H}) = \begin{cases} 0, & \text{for } 0 \leq z < \mathbf{H}(1 - c_T), \\ \frac{F}{V}, & \text{for } \mathbf{H}(1 - c_T) \leq z < \mathbf{H}, \\ 0, & \text{for } \mathbf{H} \leq z < \infty, \end{cases} \quad (3.3)$$



**Figure 4.** Individual leaf area density and leaf area index as a function of elevation  $z$  for an individual with a set height (*a*) and as a function of individual height  $h$  at a known elevation (*b*). In (*a*), an individual with height  $H = 2$  m is plotted with equations (3.3) and (3.5), other constants are  $c_T = 0.4$ ,  $c_W = 0.6$ ,  $c_{L_1} = 0.015$  and  $c_{L_2} = 3$ . In (*b*), the  $L_{AD}$  and  $L_{AI}$  of individuals of different heights at an elevation of  $z = 2$  m are plotted with equations (3.4) and (3.6), and the same constants are used.

where  $H(1 - c_T)$  is the elevation of the bottom of a canopy of an individual of height  $H$ . Figure 4*a* (black line) shows a single realization of equation (3.3) for an individual with  $H = 2$  m. We condition on the fact that  $H = h$ , and use equations (3.1) and (3.2) to resolve the canopy volume and total leaf area with respect to  $h$  as

$$g_{LAD}(H|H=h) = \begin{cases} 0, & \text{for } 0 \leq h < z, \\ \frac{4c_{L_1}}{\pi c_T c_W^2} h^{c_{L_2}-3}, & \text{for } z \leq h < \frac{z}{(1-c_T)}, \\ 0, & \text{for } \frac{z}{(1-c_T)} \leq h < \infty. \end{cases} \quad (3.4)$$

Equation (3.4) states that until an individual's height  $h$  reaches  $z$ , leaf area density at elevation  $z$  is zero. While the individual has a height between  $z$  and  $z/(1 - c_T)$ , the leaf area density is described by the mean leaf area density ( $F/V$ ). When an individual is so tall such that the bottom of the canopy does not reach down to elevation  $z$ , there is again zero leaf area density. Note also that as  $c_T$  approaches one, leaf area is present for any height above  $z$  and therefore the upper bound of the second inequality is infinity. Figure 4*b* (black line) shows equation (3.4) for individuals with different  $h$  values at an elevation of  $z = 2$  m.

Leaf area index above a specific point is the product of the leaf area density and the vertical length through the canopy. Leaf area index at an elevation  $z$  within a single canopy,

$g_{L_{AI}}(\mathbf{H})$  ( $\text{m}^2 \text{m}^{-2}$ ), is given by

$$g_{L_{AI}}(\mathbf{H}) = \begin{cases} \frac{(c_T \mathbf{H})F}{V}, & \text{for } 0 \leq z < \mathbf{H}(1 - c_T), \\ \frac{(\mathbf{H} - z)F}{V}, & \text{for } \mathbf{H}(1 - c_T) \leq z < \mathbf{H}, \\ 0, & \text{for } \mathbf{H} \leq z < \infty, \end{cases} \quad (3.5)$$

where the leaf area index is constant below the canopy, linearly decreases through the canopy, and is zero above an individual canopy. Figure 4a (grey line) shows a single realization of equation (3.5) for an individual with  $\mathbf{H} = 2 \text{ m}$ . We rewrite the  $g_{L_{AI}}(\mathbf{H})$  function conditioned on  $\mathbf{H} = h$  and have

$$g_{L_{AI}}(\mathbf{H}|\mathbf{H} = h) = \begin{cases} 0, & \text{for } 0 \leq h < z, \\ \frac{4c_{L1}}{\pi c_T c_W^2} h^{c_{L2}-2} \left(1 - \frac{z}{h}\right), & \text{for } z \leq h < \frac{z}{(1 - c_T)}, \\ \frac{4c_{L1}}{\pi c_W^2} h^{c_{L2}-2}, & \text{for } \frac{z}{(1 - c_T)} \leq h < \infty. \end{cases} \quad (3.6)$$

As with the case of leaf area density, the leaf area index at an elevation  $z$  is zero until an individual's height  $h$  exceeds  $z$ . As the height of an individual increases, the leaf area index at  $z$  increases slowly because the full canopy is not above  $z$ . Once the entire canopy is above the elevation  $z$ , leaf area index increases with height more rapidly as the canopy volume expands as height increases. Figure 4b (grey line) shows equation (3.6) for individuals with different  $h$  values at an elevation of  $z = 2 \text{ m}$ .

## (b) Marked heights

The size of each mark in the stochastic double Poisson process corresponds to the height  $\mathbf{H}$  of each individual. Many distributions, including the beta, normal, lognormal, gamma and Weibull, have been used to characterize the distribution of heights observed in forest stands [44]. Given the allometric relationships chosen above and the form of equations (2.7) and (2.8), the gamma distribution allows for closed-form analytical solutions to be derived for the mean and variance of  $L_{AI}$  and  $L_{AD}$ . The gamma distribution is a two-parameter continuous distribution bound at zero, of which the exponential distribution is a specific form.

The distribution of  $\mathbf{H}$  is specified as a gamma distribution,  $\sim \Gamma(k, m)$ , with a probability density function  $f_{\Gamma(k,m)}(h)$  and cumulative density function  $F_{\Gamma(k,m)}(h)$ . The probability that a random height  $\mathbf{H}$  is equal to  $h$  is defined as

$$\mathbb{P}(h < \mathbf{H} < h + dh) \equiv f_{\Gamma(k,m)}(h) dh = h^{k-1} \frac{e^{-h/m}}{\Gamma(k)m^k} dh. \quad (3.7)$$

The probability that a random height is less than  $h$  is defined as

$$\mathbb{P}(\mathbf{H} < h) \equiv F_{\Gamma(k,m)}(h) = \frac{1}{\Gamma(k)} \gamma\left(k, \frac{h}{m}\right), \quad (3.8)$$

where  $\Gamma(x)$  is the complete gamma function evaluated at  $x$ , and  $\gamma(\alpha, x)$  is the lower incomplete gamma function with parameter  $\alpha$  evaluated at  $x$ . The average height of  $\mu_H$  is equal to  $km$  and variance  $\sigma_H^2$  is equal to  $km^2$  (note,  $k = \mu_H^2/\sigma_H^2$  and  $m = \sigma_H^2/\mu_H$ ).

## (c) Landscape-level variability

We combine equations (3.4) or (3.6) with (2.7) and (2.8) to estimate the mean and variance of leaf area density, and leaf area index for a stochastic ecosystem as a function of elevation  $z$ .

**Table 1.** The functions,  $\Gamma_i^*(z)$  [ ], are defined with the cumulative distribution function  $F_{\Gamma(k_i,m)}(z_i)$  with parameters  $k_i$  and  $m$  evaluated at  $z_1 = z$  and  $z_2 = z/(1 - c_T)$ .

$\Gamma_i^*(z)$ functions	$k_i$ functions
$\Gamma_1^*(z) = \Gamma(k_1)(F_{\Gamma(k_1,m)}(z_2) - F_{\Gamma(k_1,m)}(z_1))$	$k_1 = k + c_{L_2} - 1$
$\Gamma_2^*(z) = \Gamma(k_2)(F_{\Gamma(k_2,m)}(z_2) - F_{\Gamma(k_2,m)}(z_1))$	$k_2 = k + 2c_{L_2} - 4$
$\Gamma_3^*(z) = \Gamma(k_3)((1 - c_T)F_{\Gamma(k_3,m)}(z_2) - F_{\Gamma(k_3,m)}(z_1))$	$k_3 = k + c_{L_2}$
$\Gamma_4^*(z) = \Gamma(k_4)((1 - c_T^2)F_{\Gamma(k_4,m)}(z_2) - F_{\Gamma(k_4,m)}(z_1))$	$k_4 = k + 2c_{L_2} - 2$
$\Gamma_5^*(z) = \Gamma(k_5)(F_{\Gamma(k_5,m)}(z_2) - F_{\Gamma(k_5,m)}(z_1))$	$k_5 = k + 2c_{L_2} - 3$

The mean leaf area density,  $\mu_{LAD}(z)$  ( $m^2 m^{-3}$ ), and variance of leaf area density,  $\sigma_{LAD}^2(z)$  ( $m^4 m^{-6}$ ), are described by

$$\mu_{LAD}(z) = \frac{\lambda_s c_{L_1} m^{c_{L_2} - 1}}{c_T \Gamma(k)} \Gamma_1^*(z) \quad (3.9)$$

and

$$\sigma_{LAD}^2(z) = \frac{4\lambda_s c_{L_1}^2 m^{2c_{L_2} - 4}}{\pi c_W^2 c_T^2 \Gamma(k)} \Gamma_2^*(z). \quad (3.10)$$

In a similar manner, the mean leaf area index,  $\mu_{LAI}(z)$  ( $m^2 m^{-2}$ ), and variance of leaf area index,  $\sigma_{LAI}^2(z)$  ( $m^4 m^{-4}$ ), are given by

$$\mu_{LAI}(z) = \frac{\lambda_s c_{L_1} m^{c_{L_2}} \Gamma(k + c_{L_2})}{\Gamma(k)} + \frac{\lambda_s c_{L_1} m^{c_{L_2}}}{c_T \Gamma(k)} \left( \Gamma_3^*(z) - \frac{z}{m} \Gamma_1^*(z) \right) \quad (3.11)$$

and

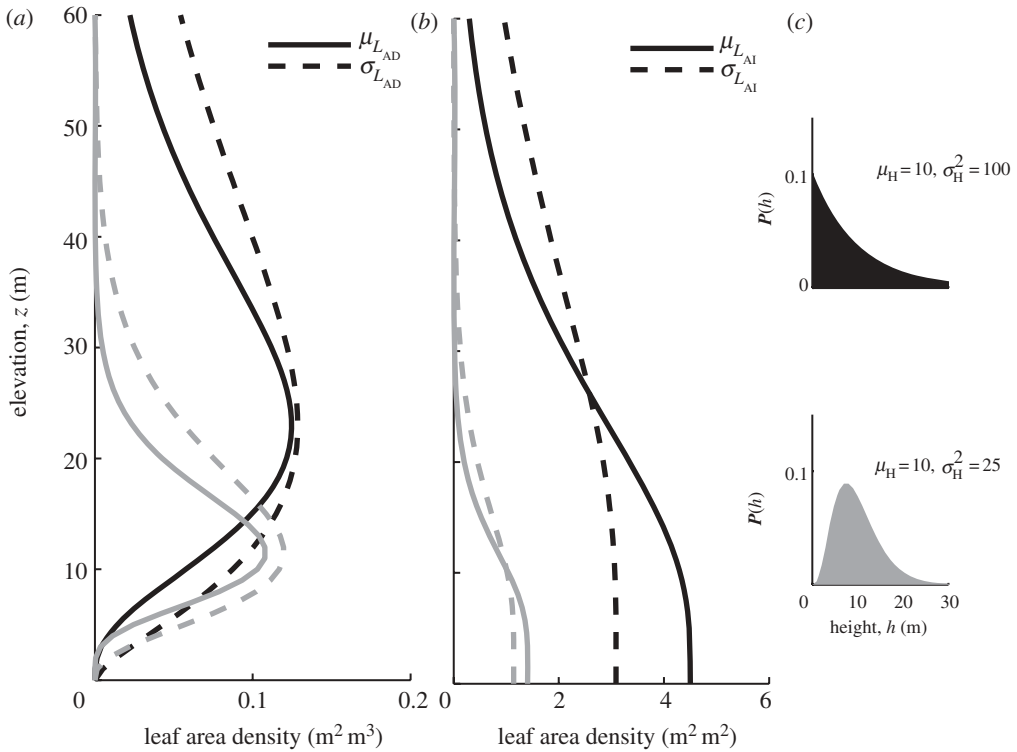
$$\sigma_{LAI}^2(z) = \frac{4\lambda_s c_{L_1}^2 m^{2c_{L_2} - 2} \Gamma(k + 2c_{L_2} - 2)}{\pi c_W^2 \Gamma(k)} + \frac{4\lambda_s c_{L_1}^2 m^{2c_{L_2} - 2}}{\pi c_W^2 c_T^2 \Gamma(k)} \left( \Gamma_4^*(z) - 2\frac{z}{m} \Gamma_5^*(z) + \frac{z^2}{m^2} \Gamma_2^*(z) \right). \quad (3.12)$$

The constants,  $\Gamma_i^*(z)$  [ ], are defined with the gamma cumulative distribution function,  $F_{\Gamma(k,m)}(x)$ , as given in table 1. The first terms of equations (3.11) and (3.12) give the values of  $\mu_{LAI}$  and  $\sigma_{LAI}^2$  at ground level ( $z = 0$ ).

Figure 5 shows two examples of equations (3.9)–(3.12) for hypothetical distributions of heights with the same landscape density, dispersion and allometric constants. For two distributions of individual heights, one following an exponential distribution and another simulating a more even-sized distribution, we observe that the maximum mean and standard deviation of  $L_{AD}$  are similar; however, the foliage occupies a much smaller vertical band for the even-sized distribution. This translates to a smaller total leaf area index at ground level for the even-sized distribution. Variability in individual heights affects both the vertical foliage placement and total leaf area. Figure 6 depicts how the mean and variability in leaf area index relate to both landscape density ( $\lambda_s$ ) and aggregation level ( $\nu_s$ ). As expected, increasing the number of individuals present results in a direct and linear increase in landscape average leaf area index and a decrease in variability. Changes in the dispersion  $\nu_s$  do not affect mean leaf area index, but significantly affect the leaf area variability. For a given landscape density, increasing aggregation levels leads to larger variability in leaf area index, as expressed by the coefficient of variation,  $C^V(L_{AI}) \equiv \sigma_{LAI} / \mu_{LAI}$ .

#### (d) Fractional cover

The fraction of the landscape covered by canopies,  $f_{cc}$  [ ], is an important ecosystem feature, and is easily derived from the model. Because each individual is associated with a single canopy, the distribution of the number of canopies above any given location is assumed to be of the same form



**Figure 5.** (a) Landscape leaf area density and (b) leaf area index as a function of elevation  $z$  for two different distributions of heights (c). Both the means,  $\mu_{L_{AD}}$  or  $\mu_{L_{AI}}$  (solid lines) and standard deviations  $\sigma_{L_{AD}}$  or  $\sigma_{L_{AI}}$  (dashed lines) are shown. A set of individuals with a mean height of 10 m and standard deviation of 10 m conforms to an exponential distribution (black lines). A more even-sized distribution with the same mean height of 10 m but a standard deviation in height of 5 m is also plotted (grey lines). Landscape density is  $\lambda_S = 0.05$  and aggregation parameter  $\nu_S = 1$ , with  $c_T$ ,  $c_W$ ,  $c_{L_1}$  and  $c_{L_2}$  are the same as figure 4.

as that of the individuals themselves, i.e. a double Poisson distribution. The number of canopies at any given location is therefore also  $\sim DP(\lambda_{cc}, \nu_{cc})$ , where  $\lambda_{cc}$  and  $\nu_{cc}$  are rescaled parameters describing the distribution of canopy cover.

The expressions for the average number of canopies per unit area  $\lambda_{cc}$  and the dispersion in canopy cover  $\nu_{cc}$  are found with the solutions to the moments of the leaf area density. By setting  $c_{L_1}$  to  $(\pi c_W^2)/4$ ,  $c_{L_2}$  to 3, and  $c_T$  to 1, we create a canopy with a value of one that does not vary with the height of an individual. Using equation (3.9), the average number of canopies per unit area is

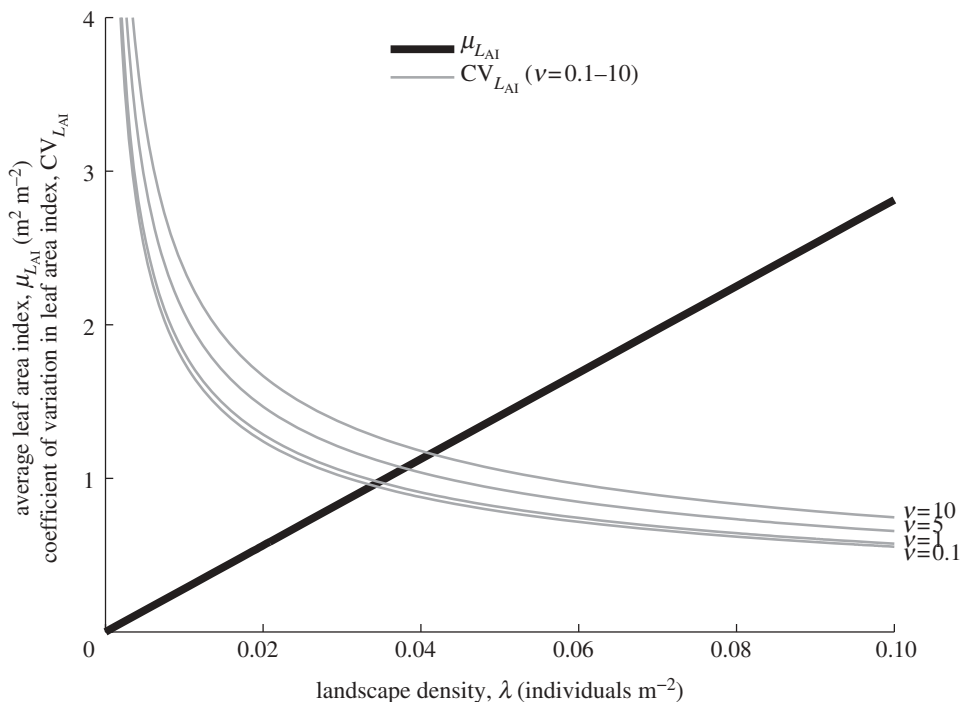
$$\lambda_{cc} = \frac{\pi \lambda_S c_W^2 m^2 \Gamma(k+2)}{4 \Gamma(k)} = \frac{1}{4} \pi \lambda_S c_W^2 (\mu_H^2 + \sigma_H^2), \quad (3.13)$$

where  $\Gamma_1^*(0) = \Gamma(k+2) = k(k+1)\Gamma(k)$  because  $z_1 = 0$  and  $z_2 = \infty$ . Similarly, equations (2.10) and (3.10) are used to find  $\nu_{cc} = 1 + \lambda_{cc}(\nu_S - 1)/(\lambda_S S)$ .

The fractional canopy cover is defined as the probability that at least one canopy is present at a given location. This probability is simply estimated as one minus the probability of zero canopies. Using the probability mass function in equation (2.1), we have

$$f_{cc} = 1 - f_{DP(\lambda_{cc}, \nu_{cc})}(0) = 1 - \frac{1}{c_{DP(\lambda_{cc}, \nu_{cc})}} \nu_S^{-1/2} e^{-(1/4)\pi \lambda_S \nu_S^{-1} c_W^2 (\mu_H^2 + \sigma_H^2)}, \quad (3.14)$$

where  $c_{DP(\lambda_{cc}, \nu_{cc})}$  is the constant of integration for the double Poisson distribution described by  $\lambda_{cc}$  and  $\nu_{cc}$ .



**Figure 6.** Average leaf area index at ground level,  $\mu_{L_{AI}}$  (black line), increases linearly with density,  $\lambda_s$ , and is independent of aggregation  $\nu_s$ . The coefficient of variability in leaf area index,  $\sigma_{L_{AI}}/\mu_{L_{AI}}$ , decreases with landscape density proportional to  $1/\sqrt{\lambda}$ . A mean height of  $\mu_H = 10$  with variance of  $\sigma_H^2 = 25$  was used, with  $c_T, c_W, c_L$ , and  $c_{L_2}$  the same as in figure 4.

### (e) Gap probabilities and radiation transfer

Characterization of the radiation regime in the vegetation canopy is necessary to drive biophysical models of photosynthesis and transpiration. Given the characterizations of the mean and variance of  $L_{AI}$  and  $L_{AD}$ , we wish to estimate the statistics of radiation transfer. We derive analytically an approximation for the fraction of light rays that pass through the canopy to elevation  $z$ . At any given location, if we consider only canopies and not stem material, then the fraction of direct beam radiation passing through a canopy with a leaf area index  $L$  is given by

$$P = e^{-(G(\theta)/\cos(\theta))L}, \quad (3.15)$$

where  $G(\theta)$  is the projection of unit foliage area on the plane perpendicular to the view direction  $\theta$  [1,24,31]. Equation (3.15) was derived based on homogeneous canopies and its application to heterogeneous environments is a simplified approximation. In the treatment here, we consider the landscape as broken up into separate homogeneous patches with different foliage densities. The leaf area index present at a given patch is considered a random variable,  $L$ , where  $L$  has mean  $\mu_{L_{AI}}$  and variance  $\sigma_{L_{AI}}^2$ . The probability that a realization of  $L$  is equal to a specific value  $l$  is denoted by  $f(l)$ ; i.e.  $\mathbb{P}[L=l] \equiv f(l)$ . Given that leaf area is a random variable in space ( $L$ ), the radiation regime is also a random process in space, denoted as  $P$ . We can then estimate the moments of light penetration as

$$\mathbb{E}[P] = \int_0^\infty e^{-(G(\theta)l/\cos(\theta))} f(l) dl \quad \text{and} \quad \mathbb{E}[P^2] = \int_0^\infty e^{-2(G(\theta)l/\cos(\theta))} f(l) dl, \quad (3.16)$$

where  $\mu_P = \mathbb{E}[P]$  and  $\sigma_P^2 = \mathbb{E}[P^2] - \mu_P^2$ .

The distribution of leaf area index,  $f(l)$ , is assumed to follow a gamma distribution, as defined in equation (3.7). We find  $\mu_P$  and  $\sigma_P^2$  as

$$\mu_P(z) = \left( \frac{\mu_{L_I}}{(G(\theta)/\cos(\theta))\sigma_{L_{AI}}^2 + \mu_{L_{AI}}} \right)^{\mu_{L_I}^2/\sigma_{L_I}^2} \quad (3.17)$$

and

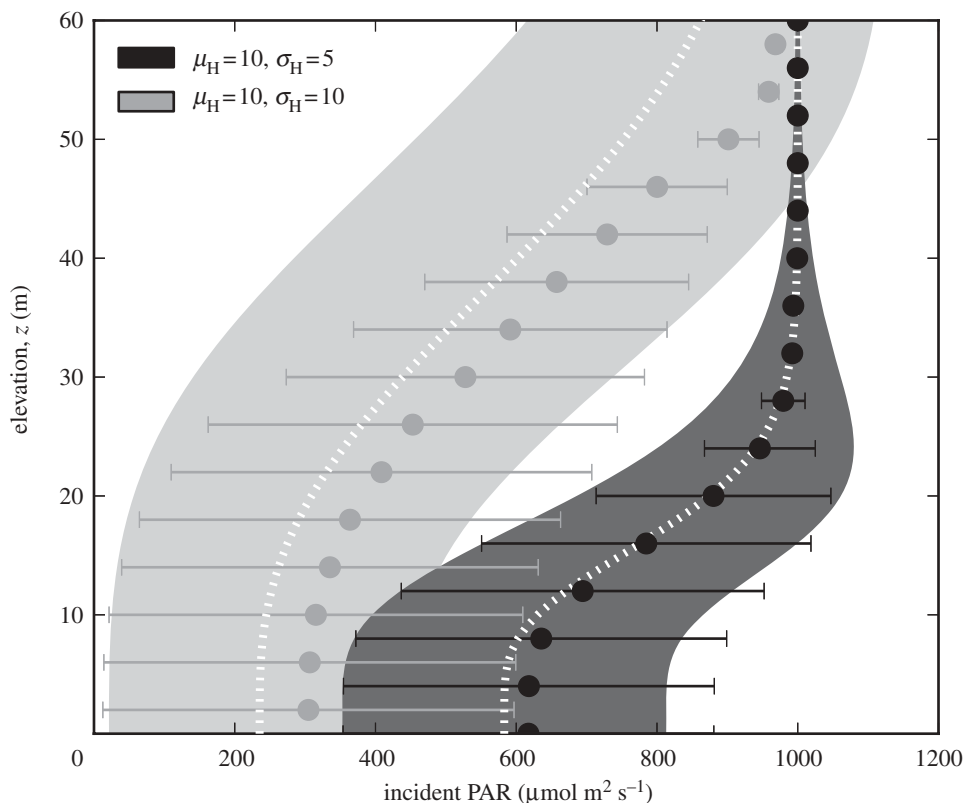
$$\sigma_P^2(z) = \left( \frac{\mu_{L_I}}{(2G(\theta)/\cos(\theta))\sigma_{L_{AI}}^2 + \mu_{L_{AI}}} \right)^{\mu_{L_I}^2/\sigma_{L_I}^2} - \left( \frac{\mu_{L_I}}{(G(\theta)/\cos(\theta))\sigma_{L_{AI}}^2 + \mu_{L_{AI}}} \right)^{2\mu_{L_I}^2/\sigma_{L_I}^2}. \quad (3.18)$$

The negative exponential transform of a gamma random variable in equation (3.15) leads to a random variable with a beta-like distribution bound between zero and one. The mean and variance of light penetration, as defined by equations (3.17) and (3.18), can then be used to model the entire distribution of light penetration for the vegetated system.

The ability of equations (3.17) and (3.18) to capture the statistics of the random radiative regime  $P$  is assessed with the three-dimensional radiative transfer model FLiES [19]. The FLiES model is a Monte Carlo ray-tracing simulator with an explicitly specified three-dimensional canopy. The model demonstrated low model-to-reference deviations in the RAMI4PILPS radiation model inter-comparison project [45] and has been coupled to energy and carbon exchange models to predict landscape-scale fluxes [7]. In FLiES, a three-dimensional vegetation configuration is specified for a  $30 \times 30$  m plot, and then a set number of individual photons are bombarded on the canopy. The energy of each photon is followed as it passes through the canopy, encounters vegetation elements and randomly scatters. After all photons have been released, absorbed photosynthetically active radiation (PAR) is calculated for each  $0.1 \times 0.1 \times 1$  m voxel by summing the change in energy of all scattering events that have occurred within the grid box.

Two separate vegetation configurations were simulated in FLiES, with 50 separate random realizations of each configuration, and  $10^6$  photons in each simulation. PAR at the top of the canopy was specified as  $1000 \mu\text{mol m}^{-2} \text{s}^{-1}$  with no diffuse light. Incident PAR at each level was calculated as the top of canopy PAR minus the absorbed PAR above. The mean and standard deviation of incident PAR was calculated for each elevation above the surface in 1 m layers. Solar zenith angle was set at 0, leaf albedo set at 0.15 and soil albedo at 0.  $G(\theta)$  was set at  $\sqrt{\alpha} \times 0.5$ , where  $\alpha$  is one albedo, and 0.5 denotes a uniform leaf angle distribution [46]. Vegetation was modelled as cylindrical canopy elements with uniform foliage density (the inner stem area of FLiES was removed by setting the constant 'bp2' to 0.0 on line 277 of the source file 'iprams.f' and recompiling the model). FLiES does not consider overlapping canopies, and when two canopies intersect, the leaf area density of this region is set at one of the two values, not their sum. This treatment of intersecting canopies diverges from the stochastic framework developed earlier, and to minimize intersecting canopies in FLiES, the parameter  $c_T$  was set to 0.1.

Vegetation canopies with an even-sized distribution of heights ( $\mu_H = 10$ ,  $\sigma_H = 5$ ) and an exponential distribution of heights ( $\mu_H = 10$ ,  $\sigma_H = 5$ ) were analysed. Figure 7 shows the incident PAR profile based on equations (3.17) and (3.18) and the FLiES model results. For both cases, the form of the mean incident PAR vertical profile obtained from the FLiES model is captured by the analytical equations derived earlier. The offset for both the ray-tracing simulations from the analytical prediction is probably caused by the different treatment of intersecting canopies between methods. This lack of an additive canopy intersection leads to a reduction in total simulated foliage and a decrease in the variability of foliage densities. Spatial variability in PAR is also captured very well by the analytical formulation when  $\sigma_H = 5$ , and to a lesser degree when  $\sigma_H$  is set to 10. The increase in variability in  $\sigma_H$  leads to more large vegetation elements and thereby more canopy intersections, and because intersecting canopies do not have an additive leaf area density in FLiES, the total variability is severely decreased. The application of equation (3.15) to a non-homogeneous field by integrating across the probability space of different leaf areas provides a fast, analytically tractable estimate



**Figure 7.** Attenuation of PAR through simulated canopies. White dashed lines and grey-shaded areas show the mean  $\pm$  1 s.d. of incident PAR as estimated from equations (3.17) and (3.18). Horizontal error bars are the mean  $\pm$  1 s.d. of incident PAR based on 50 randomly generated FLiES simulations. Parameters are the same as figure 5, except  $c_1=0.1$ .

of second-order radiation regime properties. This application is a simple case with the Sun directly overhead, no diffuse radiation and a black understorey, and therefore the application of equations (3.17) and (3.18) in their current form is limited. However, given that both the mean and variability of PAR can be solved analytically in specific cases, further work may extend this approach further.

Many models make use of an empirical clumping factor,  $\Omega$ , or an 'effective' leaf area,  $L_e$  [31]. The clumping factor scales leaf area index such that the average leaf area index inserted in equation (3.15) will give the average beam penetration (i.e.  $\mu_P = \exp[-G(\theta)\Omega(\theta, z)\mu_{LAI}/\cos(\theta)]$ ). Similarly, the 'effective' leaf area  $L_e$  is equal to  $\mu_{LAI}\Omega(\theta)$  [31]. The value of  $\Omega(\theta)$  can be derived as

$$\Omega(\theta) = -\frac{\mu_{LAI} \cos(\theta)}{\sigma_{LAI}^2 G(\theta)} \ln \left( \frac{\mu_{L_i}}{(G(\theta)/\cos(\theta))\sigma_{LAI}^2 + \mu_{LAI}} \right). \quad (3.19)$$

However, the use of a clumping factor or 'effective' leaf area presents problems when estimating biophysical processes in heterogeneous environments. Areas with varying degrees of leaf area will respond nonlinearly owing to energy limitations, such that model results using averaged leaf area or radiation transfer will overestimate ecosystem functionality.

Multi-angle observations of canopy reflectance data collected from space-based platforms have been used to estimate the clumping index for the land cover types of the Global Landcover 2000 dataset [15,17,47]. Values of  $\Omega(\theta)$  from [17] were used in conjunction with the average leaf area index obtained from NASA's moderate resolution imaging spectroradiometer (MODIS) platform to estimate the variance of leaf area index for each land cover type in figure 1. MODIS MYD15A2



data were used from the year 2006 [16] and a  $G(\theta)$  value of 0.5 was assumed. Equation (3.19) was solved numerically to obtain estimates of the variance of leaf area index for land cover types as shown in figure 1.

## 4. Variability in leaf area and biophysical processes

The leaf surface at the stomatal opening is the site at which transpiration and carbon assimilation occur. These biophysical processes are driven by a finite amount of available energy. Because of energy limitation, an area with twice as much leaf area as another will not necessarily have double the transpiration or assimilate twice the carbon. In order to accurately characterize the landscape-scale exchange of water and carbon, information about the distribution of leaf area and its micro-environment is required. By characterizing both the mean and variance of leaf area, we are able to describe the distribution of foliage area and the biophysical response at any amount of leaf area.

We present an example of how variability in the spatial organization of vegetation structure affects the total landscape flux by estimating the total landscape transpiration with the Penman–Monteith (PM) equation. The PM equation is a statement of the energy balance at the leaf scale used to estimate ecosystem transpiration rates given meteorological forcing of net radiation,  $R$ , air temperature,  $T_a$ , horizontal wind speed,  $U_x$ , relative humidity,  $h_r$ , and stomatal conductance,  $g_s$  (see [48] for a complete description of the PM equation and its terms). The evapotranspiration,  $E$  ( $\text{mm day}^{-1}$ ), from a surface is calculated as

$$E = \frac{\Delta R + \rho_a c_a g_a e_a^* (1 - h_r)}{\rho_w \lambda_v (\Delta + \gamma (1 + g_a / (g_c L)))}, \quad (4.1)$$

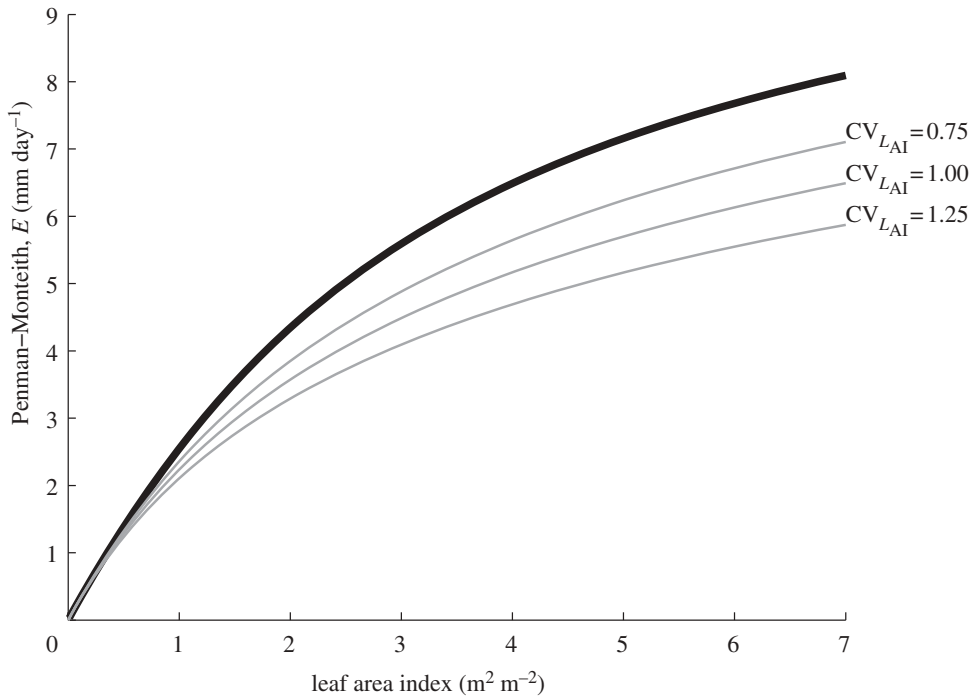
where  $L$  is a random variable describing the total leaf area index at a given location, and  $R$  is a random variable describing the net radiation present at a given location, determined as a function of  $L$ . Figure 8 (thick black line) depicts the estimated evapotranspiration for given values of leaf area index following [48]. Radiation at the top of the canopy,  $R_{\text{in}}$ , was attenuated as a function of the random leaf area index at a given location, as in equation (3.15), as  $R = R_{\text{in}} e^{-G(\theta)L}$ . Leaf area at each  $L$  was split into sunlit and shaded portions, and the total evapotranspiration for each leaf index was calculated as the sum [46].

Evapotranspiration is a nonlinear function of leaf area, and therefore, variability in  $L$  will influence the total average landscape transpiration. For a given  $\lambda_S$  and  $\nu_S$  and a set of biophysical constants,  $\mu_H$ ,  $\sigma_H^2$ ,  $c_W$ ,  $c_T$ ,  $c_{L_1}$  and  $c_{L_2}$ , we can calculate the mean,  $\mu_{L_{\text{AI}}}$ , and standard deviation,  $\sigma_{L_{\text{AI}}}$ , of the leaf area index and estimate the distribution of  $L$ . The distribution of  $L$  is again assumed as a gamma random variable with probability density function  $f_L(l)$ . We calculate the expected landscape transpiration rate by conditioning on the probability that a particular value of leaf area is present in the canopy by the definition of expectation,

$$\mathbb{E}[E] = \int_0^\infty \frac{\Delta R(l) + \rho_a c_a g_a e_a^* (1 - h)}{\rho_w \lambda_v (\Delta + \gamma (1 + g_a / (g_c l)))} f_L(l) dl. \quad (4.2)$$

Equation (4.2) is numerically integrated to estimate the average landscape evapotranspiration.

As the landscape moves from highly regular arrangements to highly clumped arrangements, large changes in the expected transpiration rate occur. Figure 8 shows the value of  $\mathbb{E}[E]$  for landscapes with different coefficients of variability in leaf area index. Although the treatment of radiation attenuation through integration of equation (3.15) is a simplification, the presented method allows for exploration of the consequences of different vegetation patterns to be quickly investigated. In more clustered landscapes characterized by large variability in leaf area index, the landscape evapotranspiration is significantly less than the flux obtained when the average leaf area index is used to estimate evapotranspiration. Clustering leads to patches of variably dense foliage where the patch transpiration may be higher in total but lower per unit leaf area due to a limited amount of energy to drive the evaporation process.



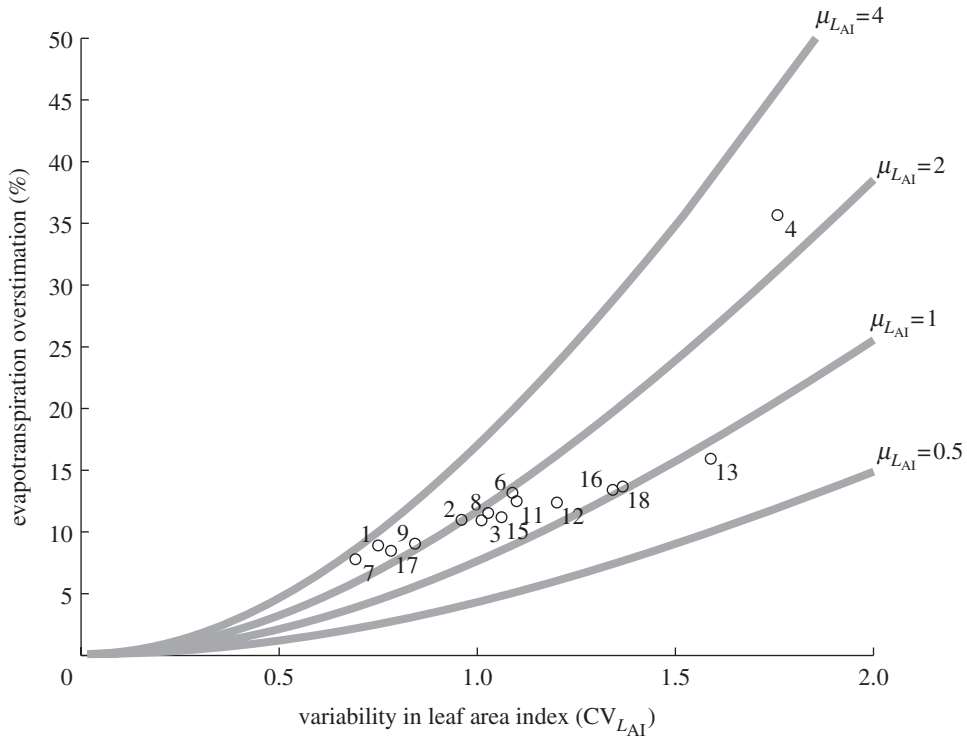
**Figure 8.** Expected landscape transpiration flux calculated with the PM equation. In the case where only an average leaf area is considered (black line),  $\mathbb{E}[ET]$  is a monotonic nonlinear function of  $\mu_{L_{AI}}$ . With the consideration of aggregation of individuals in space (grey lines), variability in canopy leaf area leads to deviations in  $\mathbb{E}[ET]$  from the simple zero variability case. Vegetation parameters are the same as figure 6 and PM parameters are  $R_{in} = 200 \text{ W m}^{-2}$ ,  $T_a = 25^\circ\text{C}$ ,  $U_x = 1 \text{ m s}^{-1}$ ,  $h_r = 0.60$ ,  $g_s = 6 \text{ mm s}^{-1}$ ,  $G(\theta) = 0.5$ , day length of 12 h.

## 5. Final discussion

Understanding variability in the structure of vegetated environments is necessary to accurately estimate the exchange of water, carbon and energy between the Earth's surface and the atmosphere. The degree to which individuals are dispersed on a landscape considerably influences structural variability, and the stochastic vegetation model presented earlier is able to incorporate diverse configurations of individuals. With the stochastic field of individuals defined as a marked double Poisson process, the earlier-mentioned methodology captures the vertical and horizontal distribution of canopy structure. Allometric relationships relate the marked two-dimensional process to a three-dimensional space where canopies occupy finite volumes and contain specified foliage amounts. The choice of cylindrical canopies and simple allometry in the  $g(\cdot)$  functions, as well as assuming the distributions of individual heights are gamma random variables, provides straightforward closed-form analytical solutions. Through stochastic spatial geometry, direct analytical expressions of the moments of the vertical distribution of foliage are available.

Descriptions of the mean and variance of leaf area density and leaf area index provide a detailed representation of vegetation structure with minimal parameterization. Analytical expression allows for fast computation of vegetation profiles that can be parameterized easily to resolve canopy structure in global ecosystem models. The amount of canopy cover, as well as the full distribution of light penetration, can be quickly modelled as functions of the vertically resolved leaf area distribution.

Furthermore, if the independence of different species or functional groups is assumed, then this methodology may be used to estimate the vertical profile of leaf area for a biologically diverse system through a superposition of separate leaf area profiles for each modelled species.



**Figure 9.** Percentage overestimation of landscape transpiration of the big-leaf model ( $CV_{L_{ij}} = 0$ ) relative to transpiration calculated with a distribution of leaf areas for different land cover types (numeric labels given in figure 1 [15]). PM parameters are the same as figure 8.

This would allow for representation of ecotones in Earth-system models, where the observed ranges of species partially overlap. However, we have assumed that leaf area distributions follow a gamma distribution, which is a bound distribution that does not permit simple combinations of multiple distributions. Other distributions, such as the normal, may be chosen such that the leaf area profiles of different species can be combined. Furthermore, by setting  $c_T$  to one, vegetation structures low to the ground, such as grass clumps, are able to be modelled. Finally, because of the generality of the model, it is possible to model the distribution of subsurface elements such as rooting networks.

Because this three-dimensional stochastic vegetation model is able to characterize both the mean and variance in leaf area distribution, differences in ecosystem processes as functions of leaf area can also be addressed. The micro-climate where leaf-level processes occur is highly dependent on the amount of leaves in the immediate area. Energy available to drive photosynthesis and transpiration is a finite quantity and, as such, increases in leaf area lead to diminishing increases in transpiration. Landscape-scale fluxes from ecosystems characterized by larger variability in their leaf area can only be estimated by consideration of the entire distribution of micro-environments. The difference between the evapotranspiration at mean leaf area index,  $E(\mu_{L_{AI}})$ , and the expected value of  $E$  given a distribution of leaf area,  $\mathbb{E}[E(L)]$ , is shown in figure 9 as the percentage of flux overestimation,  $(E(\mu_{L_{AI}}) - \mathbb{E}[E(L)]) / (\mathbb{E}[E(L)]) \times 100$  (%). The overestimation of evapotranspiration becomes large as the variability in canopy structure increases. Land cover types characterized by open vegetation cover with lower leaf area overestimate evapotranspiration to a larger degree for the modelled conditions. Other micro-environment factors such as temperature and humidity may also vary with foliage density and could possibly be incorporated into more complicated models. Finally, this methodology is formulated such that changes in the density and size distribution of individual species

can be quickly propagated into the entire vegetation structure. This straightforward analytical solution to canopy structure as a function of individual abundance is therefore suited to studies of vegetation succession, biodiversity and optimality wherein changes in external forcing and disturbances drive ecosystem dynamics.

## Appendix

**Table 2.** Random variables, biophysical constants, derived values and other parameters. CDF, cumulative distribution function.

symbol	units	description
random variables (always in bold)		
<b><math>N</math></b>	—	random number of individuals in a subplot $S$
<b><math>Y</math></b>	varies	total influence of all individuals at a random point
<b><math>H</math></b>	m	height of a random individual
<b><math>W</math></b>	m	canopy width of a random individual
<b><math>T</math></b>	m	canopy thickness of a random individual
<b><math>A</math></b>	m <sup>2</sup>	canopy area of a random individual
<b><math>V</math></b>	m <sup>3</sup>	canopy volume of a random individual
<b><math>F</math></b>	m <sup>2</sup>	canopy foliage area of a random individual
<b><math>L</math></b>	m <sup>2</sup>	leaf area index at a random location
<b><math>P</math></b>	—	penetration fraction of the Sun's direct rays
<b><math>E</math></b>	mm day <sup>-2</sup>	evapotranspiration flux
species-specific biophysical constants		
$\lambda_S$	m <sup>-2</sup>	average number of individuals per unit area in region $S$
$\nu_S$	—	dispersion in $\lambda_S$ ( $\nu_S < 1$ is more ordered, $\nu_S > 1$ clustered)
$\mu_H$	m	average height of individuals
$\sigma_H^2$	m <sup>2</sup>	variance in individual heights
$c_W$	m m <sup>-1</sup>	ratio of canopy width to canopy height
$c_T$	m m <sup>-1</sup>	ratio of canopy thickness to canopy height
$c_{L_1}$	m <sup>2-c<sub>L2</sub></sup>	total canopy foliage constant
$c_{L_2}$	—	total canopy foliage power
$G$	m <sup>2</sup> m <sup>-2</sup>	foliage projection area perpendicular to $\theta$ per unit area
derived values		
$L_{AD}$	m <sup>2</sup> m <sup>-3</sup>	leaf area density (foliage area per unit volume)
$L_{AI}$	m <sup>2</sup> m <sup>-2</sup>	leaf area index (foliage area per unit area ground)
$\mu_{L_{AD}}$	m <sup>2</sup> m <sup>-3</sup>	average landscape leaf area density at height $z$
$\sigma_{L_{AD}}^2$	m <sup>4</sup> m <sup>-6</sup>	variance of landscape leaf area density at height $z$
$\mu_{L_{AI}}$	m <sup>2</sup> m <sup>-2</sup>	average landscape leaf area index at height $z$
$\sigma_{L_{AI}}^2$	m <sup>4</sup> m <sup>-4</sup>	variance of landscape leaf area index at height $z$
$m, k$	varies	gamma distributions parameters ( $k = \mu_H^2 / \sigma_H^2$ , $m = \sigma_H^2 / \mu_H$ )
$\Gamma_{1-5}^*$	—	set of constants using gamma CDFs from table 1

(Continued.)

**Table 2.** (Continued.)

symbol	units	description
$f_{cc}$	—	fraction of landscape covered by canopies
$\lambda_{cc}$	$m^{-2}$	average number of canopies above a unit area
$\nu_{cc}$	—	dispersion in the number of canopies above a unit area
$\mu_p$	—	average penetration fraction of the Sun's direct rays
$\sigma_p^2$	—	variance in penetration fraction of the Sun's direct rays
$\Omega$	—	empirical clumping index
other parameters		
$S$	$m^2$	subplot on which the stochastic dispersion is defined
$n$	—	number of individuals in the subplot $S$
$h$	m	height of a specific individual
$z$	m	vertical elevation
$l$	$m^2 m^{-3}$	a value of leaf area index
$\theta$	degrees	zenith angle of the Sun's direct rays
$x, u, r$	m	dummy variables of integration

## References

1. Monsi M, Saeki T. 2005 On the factor light in plant communities and its importance for matter production. *Ann. Bot.* **95**, 549–567. (doi:10.1093/aob/mci052)
2. De Pury D, Farquhar G. 1997 Simple scaling of photosynthesis from leaves to canopies without the errors of big-leaf models. *Plant Cell Environ.* **20**, 537–557. (doi:10.1111/j.1365-3040.1997.00094.x)
3. Drewry D, Kumar P, Long S, Bernacchi C, Liang XZ, Sivapalan M. 2010 Ecohydrological responses of dense canopies to environmental variability: 1. Interplay between vertical structure and photosynthetic pathway. *J. Geophys. Res.* **115**, G04022. (doi:10.1029/2010JG001340)
4. Li B, Avissar R. 1994 The impact of spatial variability of land-surface characteristics on land-surface heat fluxes. *J. Clim.* **7**, 527–537. (doi:10.1175/1520-0442(1994)007<0527:TIOSVO>2.0.CO;2)
5. Caylor KK, D'Odorico P, Rodriguez-Iturbe I. 2006 On the ecohydrology of structurally heterogeneous semiarid landscapes. *Water Resour. Res.* **42**, W07424. (doi:10.1029/2005WR004683)
6. Chen Q, Baldocchi D, Gong P, Dawson T. 2008 Modeling radiation and photosynthesis of a heterogeneous savanna woodland landscape with a hierarchy of model complexities. *Agric. Forest Meteorol.* **148**, 1005–1020. (doi:10.1016/j.agrformet.2008.01.020)
7. Kobayashi H, Baldocchi DD, Ryu Y, Chen Q, Ma S, Osuna JL, Ustin SL. 2012 Modeling energy and carbon fluxes in a heterogeneous oak woodland: a three-dimensional approach. *Agric. Forest Meteorol.* **152**, 83–100. (doi:10.1016/j.agrformet.2011.09.008)
8. Ni-Meister W, Yang W, Kiang NY. 2010 A clumped-foliage canopy radiative transfer model for a global dynamic terrestrial ecosystem model. I: Theory. *Agric. Forest Meteorol.* **150**, 881–894. (doi:10.1016/j.agrformet.2010.02.009)
9. Isham V, Cox D, Rodriguez-Iturbe I, Porporato A, Manfreda S. 2005 Representation of space-time variability of soil moisture. *Proc. R. Soc. A* **461**, 4035–4055. (doi:10.1098/rspa.2005.1568)
10. Rodriguez-Iturbe I, Isham V, Cox D, Manfreda S, Porporato A. 2006 Space-time modeling of soil moisture: stochastic rainfall forcing with heterogeneous vegetation. *Water Resour. Res.* **42**, W06D05. (doi:10.1029/2005WR004497)
11. Franz TE, Caylor KK, Nordbotten JM, Rodriguez-Iturbe I, Celia MA. 2010 An ecohydrological approach to predicting regional woody species distribution patterns in dryland ecosystems. *Adv. Water Resour.* **33**, 215–230. (doi:10.1016/j.advwatres.2009.12.003)

12. Jensen JLWV. 1906 Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Math.* **30**, 175–193. (doi:10.1007/BF02418571)
13. Farquhar GD, Von Caemmerer S, Berry JA. 2001 Models of photosynthesis. *Plant Physiol.* **125**, 42–45. (doi:10.1104/pp.125.1.42)
14. Scanlon TM, Caylor KK, Levin SA, Rodriguez-Iturbe I. 2007 Positive feedbacks promote power-law clustering of Kalahari vegetation. *Nature* **449**, 209–212. (doi:10.1038/nature06060)
15. European Commission Joint Research Centre. 2003 Global land cover 2000 database. See <http://bioval.jrc.ec.europa.eu/products/glc2000/glc2000.php>.
16. Land Processes Distributed Active Archive Center, NASA. 2012 MODIS MYD15A2. See <https://lpdaac.usgs.gov/>.
17. He L, Chen JM, Pisek J, Schaaf CB, Strahler AH. 2012 Global clumping index map derived from the MODIS BRDF product. *Remote Sens. Environ.* **119**, 118–130. (doi:10.1016/j.rse.2011.12.008)
18. Wang Y, Jarvis P. 1990 Description and validation of an array model: MAESTRO. *Agric. Forest Meteorol.* **51**, 257–280. (doi:10.1016/0168-1923(90)90112-J)
19. Kobayashi H, Iwabuchi H. 2008 A coupled 1-D atmosphere and 3-D canopy radiative transfer model for canopy reflectance, light environment, and photosynthesis simulation in a heterogeneous landscape. *Remote Sens. Environ.* **112**, 173–185. (doi:10.1016/j.rse.2007.04.010)
20. Anderson MC, Norman J, Kustas WP, Li F, Prueger JH, Mecikalski JR. 2005 Effects of vegetation clumping on two-source model estimates of surface energy fluxes from an agricultural landscape during SMACEX. *J. Hydrometeorol.* **6**, 892–909. (doi:10.1175/JHM465.1)
21. Avissar R, Pielke RA. 1989 A parameterization of heterogeneous land surfaces for atmospheric numerical models and its impact on regional meteorology. *Mont. Weather Rev.* **117**, 2113–2136. (doi:10.1175/1520-0493(1989)117<2113:APOHLS>2.0.CO;2)
22. Avissar R. 1992 Conceptual aspects of a statistical-dynamical approach to represent landscape subgrid-scale heterogeneities in atmospheric models. *J. Geophys. Res.* **97**, 2729–2742. (doi:10.1029/91JD01751)
23. Nilson T. 1971 A theoretical analysis of the frequency of gaps in plant stands. *Agric. Meteorol.* **8**, 25–38. (doi:10.1016/0002-1571(71)90092-6)
24. Nilson T. 1999 Inversion of gap frequency data in forest stands. *Agric. Forest Meteorol.* **98**, 437–448. (doi:10.1016/S0168-1923(99)00114-8)
25. Nilson T, Kuusk A, Lang M, Pisek J, Kodar A. 2011 Simulation of statistical characteristics of gap distribution in forest stands. *Agric. Forest Meteorol.* **151**, 895–905. (doi:10.1016/j.agrformet.2011.02.009)
26. Ross J. 1981 *The radiation regime and architecture of plant stands*. Hingham, MA: Kluwer Academic Press.
27. Ross J, Ross V, Koppel A. 2000 Estimation of leaf area and its vertical distribution during growth period. *Agric. Forest Meteorol.* **101**, 237–246. (doi:10.1016/S0168-1923(00)00089-7)
28. Li X, Strahler AH, Woodcock CE. 1995 A hybrid geometric optical-radiative transfer approach for modeling albedo and directional reflectance of discontinuous canopies. *IEEE Trans. Geosci. Remote Sens.* **33**, 466–480. (doi:10.1109/36.377947)
29. Haverd V, Lovell J, Cuntz M, Jupp D, Newnham G, Sea W. 2012 The canopy semi-analytic  $P_{\text{gap}}$  and radiative transfer (CanSPART) model: formulation and application. *Agric. Forest Meteorol.* **160**, 14–35. (doi:10.1016/j.agrformet.2012.01.018)
30. Lovell J, Haverd V, Jupp D, Newnham G. 2012 The canopy semi-analytic  $P_{\text{gap}}$  and radiative transfer (CanSPART) model: validation using ground based LIDAR. *Agric. Forest Meteorol.* **158**, 1–12. (doi:10.1016/j.agrformet.2012.01.020)
31. Ryu Y, Nilson T, Kobayashi H, Sonnentag O, Law BE, Baldocchi DD. 2010 On the correct estimation of effective leaf area index: does it reveal information on clumping effects? *Agric. Forest Meteorol.* **150**, 463–472. (doi:10.1016/j.agrformet.2010.01.009)
32. Shabanov NV, Knyazikhin Y, Baret F, Myneni RB. 2000 Stochastic modeling of radiation regime in discontinuous vegetation canopies. *Remote Sens. Environ.* **74**, 125–144. (doi:10.1016/S0034-4257(00)00128-0)
33. Shabanov NV, Huang D, Knjazikhin Y, Dickinson R, Myneni RB. 2007 Stochastic radiative transfer model for mixture of discontinuous vegetation canopies. *J. Quant. Spectrosc. Radiat. Transf.* **107**, 236–262. (doi:10.1016/j.jqsrt.2007.01.053)
34. Vainikko G. 1973 [The equation of mean radiance in broken cloudiness]. *Trudy MGK SSSR, Meteorol. Invest.* **21**, 28–37. [In Russian.]

35. Vainikko G. 1973 [Transfer approach to the mean intensity of radiation in non-continuous clouds]. *Trudy MGK SSSR, Meteorol. Invest.* **21**, 38–57. [In Russian.]
36. Huang D, Knyazikhin Y, Wang W, Deering DW, Stenberg P, Shabanov N, Tan B, Myneni RB. 2008 Stochastic transport theory for investigating the three-dimensional canopy structure from space measurements. *Remote Sens. Environ.* **112**, 35–50. (doi:10.1016/j.rse.2006.05.026)
37. Penttinen A, Stoyan D, Henttonen HM. 1992 Marked point processes in forest statistics. *Forest Sci.* **38**, 806–824.
38. Oderwald RG. 1981 Comparison of point and plot sampling basal area estimators. *Forest Sci.* **27**, 42–48.
39. Efron B. 1986 Double exponential families and their use in generalized linear regression. *J. Am. Stat. Assoc.* **81**, 709–721. (doi:10.1080/01621459.1986.10478327)
40. Neumann M, Starlinger F. 2001 The significance of different indices for stand structure and diversity in forests. *Forest Ecol. Manag.* **145**, 91–106. (doi:10.1016/S0378-1127(00)00577-6)
41. Rodriguez-Iturbe I, Cox D, Eagleson P. 1986 Spatial modelling of total storm rainfall. *Proc. R. Soc. Lond. A* **403**, 27–50. (doi:10.1098/rspa.1986.0002)
42. Cox DR, Miller HD. 1965 *The theory of stochastic processes*. Boca Raton, FL: Chapman and Hall/CRC.
43. Diggle PJ. 1983 *Statistical analysis of spatial point patterns*. London, UK: Academic Press.
44. Hafley W, Schreuder H. 1977 Statistical distributions for fitting diameter and height data in even-aged stands. *Can. J. Forest Res.* **7**, 481–487. (doi:10.1139/x77-062)
45. Widlowski J-L *et al.* 2011 RAMI4PILPS: an intercomparison of formulations for the partitioning of solar radiation in land surface models. *J. Geophys. Res.* **116**, G02019. (doi:10.1029/2010JG001511)
46. Campbell GS, Norman JM. 1998 *An introduction to environmental biophysics*. Berlin, Germany: Springer.
47. Chen J, Menges C, Leblanc S. 2005 Global mapping of foliage clumping index using multi-angular satellite data. *Remote Sens. Environ.* **97**, 447–457. (doi:10.1016/j.rse.2005.05.003)
48. Dingman SL. 2002 *Physical hydrology*. Upper Saddle River, NJ: Prentice Hall.