Comprehensive LiDAR simulation with efficient physically-based DART-Lux model (I): Theory, novelty, and consistency validation

Xuebo Yang, Yingjie Wang, Tiangang Yin, Cheng Wang, Nicolas Lauret, Omar Regaieg, Xiaohuan Xi, Jean Philippe Gastellu-Etchegorry

Comprehensive LiDAR simulation with efficient physically-based DART-Lux model (I): Theory, novelty, and consistency validation

Xuebo Yang[a,b,c,*], Yingjie Wang[c], Tiangang Yin[d], Cheng Wang[a], Nicolas Lauret[c], Omar Regaieg[c], Xiaohuan Xi[a], Jean Philippe Gastellu-Etchegorry[c,*]

[a] Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China

[b] China National Laser Products Quality Supervising Test Center, Beijing 100094, China

[c] Centre d'Etudes Spatials de la BIOsphère (CESBIO) – Université de Toulouse, UT3, CNES, CNRS, IRD, 31401 Toulouse Cedex 9, France

[d] Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD 20740-3823, USA

* Correspondence: yangxb@aircas.ac.cn (X. Yang) jean-philippe.gastellu@cesbio.cnes.fr (J.-P. Gastellu-Etchegorry).

Abstract

Light Detection And Ranging (LiDAR) remote sensing is increasingly needed to assess the 3D architecture of Earth's surface. Physically-based LiDAR radiative transfer (RT) models are essential tools for interpreting LiDAR signals, designing LiDAR systems, and validating information retrieval methods. Discrete Anisotropic Radiative Transfer (DART) is one of the most accurate and comprehensive 3D RT models that simulate LiDAR signals of urban and natural landscapes. Its physical modeling relies on a forward Monte Carlo mode optimized by a ray-tracking technique, also called DART-RC (Ray Carlo) mode. However, DART-RC is not adapted to simulate massive LiDAR signals of large landscapes due to its constraints of

high memory demand and long computational time. Therefore, we developed a novel computationally efficient LiDAR modeling method based on a new DART modeling mode called DART-Lux. It simulates LiDAR signal by adapting the bidirectional path tracing algorithm of DART-Lux to the time and power measurements and by implementing the LiDAR instrument and multiple product outputs in DART-Lux. We verified the accuracy of DART-Lux for LiDAR modeling using DART-RC as a reference for several case studies with different LiDAR configurations (*i.e.*, single-pulse waveform, multi-pulse point cloud, multi-pulse photon counting, with and without solar signal) on realistic scenes from the RAMI experiment. Results stress that i) DART-Lux is consistent with DART-RC, for example, R^2 = 1 and rRMSE = 0.21% for the waveform of a forest simulated with a huge number of rays; ii) DART-Lux converges faster than DART-RC: its processing time is usually about half that of DART-RC, and over ten times smaller if the solar signal is simulated; iii) DART-Lux memory usage can be a hundred times less than DART-RC. Also, several sensitivity studies with various sensor configurations and solar directions illustrate the usefulness of DART-Lux for impact studies. This new DART-Lux LiDAR model opens promising perspectives for large-scale LiDAR applications with 3D modeling. It is already part of the official DART version freely available to scientists (https://dart.omp.eu).

Keywords

LiDAR, radiative transfer model, DART, bidirectional path tracing, Monte Carlo

1. Introduction

Light Detection And Ranging (LiDAR) is a well-developed active remote sensing (RS) technology that maps three-dimensional (3D) structures of Earth's landscapes by measuring time of flight of laser pulses (Wulder et al., 2012). With its advantage of 3D spatial measurement, LiDAR is widely used in many domains: topography surveying (Jaboyedoff et al., 2012), virtual smart city (Dwivedi et al., 2014), atmosphere constituent (Weitkamp, 2006), forest biomass monitoring (Zhao et al., 2018), *etc*. Driven by various application requirements and increasingly advanced devices, LiDAR system is developing towards more sophisticated instruments with higher pulse repetition frequency (PRF), higher ranging accuracy, longer battery life, *etc*. For instance, multiple spaceborne LiDAR missions were successively launched in the past twenty years, including ICESat (Ice, Cloud, and land Elevation Satellite) (Zwally et al., 2002), ICESat-2 (Markus et al., 2017), and GEDI (Global Ecosystem Dynamics Investigation) (Dubayah et al., 2020). The PRF of these LiDAR missions increases from 40 Hz for ICESat up to 242 Hz for GEDI and 10 kHz for ICESat-2. In addition to satellite platforms, the development of terrestrial (Dassot et al., 2011), mobile (Williams et al., 2013), and UAV (unmanned aerial vehicle) (Wallace et al., 2012) laser scanning systems enables rapid acquisition of high-density LiDAR data at different spatial scales. These massive LiDAR data sources drive large-area high-precision RS applications, while highlighting pressing demands to understand the capability and limitation of information retrievals to make better use of these data.

Physically-based LiDAR radiative transfer (RT) models are celebrated tools for modeling, validating, and interpreting LiDAR signals as they can simulate the laser-surface interaction mechanisms that give rise to the signals. To date, a wide range of LiDAR models have been developed with diverse complexity and efficiency. Semi-empirical models simulate the LiDAR signal as a temporal summation of Gaussian profiles that are computed by convolving the target reflectance and pulse energy distribution (Blair and Hofton, 1999; Chauve et al., 2007). Physically-based analytical models use strong simplifications on landscapes (*e.g.*, geo-optical description or turbid mediums) and ray propagation (only consider first-order scattering) (Ni-Meister et al., 2001; Sun and Ranson, 2000). Although computationally efficient, these models have an increasingly unsuitable accuracy for most potential LiDAR applications, *e.g.*, signal modeling and parameter inversions of multiple targets (Wang et al., 2021; Yang et al., 2019), LiDAR sensor design (Yang et al., 2021), *etc*. In contrast, LiDAR 3D RT models that simulate multiple scattering in LiDAR returns are much more accurate. It is the case of models that combine geo-optics with time-dependent stochastic process (Kotchenova et al., 2003), adapt radiosity (Huang and Wynne, 2013), or use Monte Carlo (MC) techniques. The LiDAR stochastic RT model solves numerically the time-dependent RT equation of LiDAR signal, but the landscapes are parameterized as geo-optical forms. The LiDAR radiosity model can handle LiDAR physical process of 3D realistic scenes with finite surfaces, but it requires huge computer load to calculate and store the view factors between any two surfaces and the radiation of any surfaces at any time interval. The MC-based models are the most commonly-used models because they are more adaptive to complex scene

structures and usually do not make simplifications on ray interactions. They are usually categorized as forward models if rays are traced from the source and backward models if rays are traced from the receiver. The MC-based LiDAR forward models include RAYTRAN (Govaerts and Verstraete, 1998; Widlowski et al., 2006), FLIES (Kobayashi and Iwabuchi, 2008), DIRSIG (Brown et al., 2005; Goodenough and Brown, 2017), and DART (Gastellu-Etchegorry et al., 2016, 2015; Yin et al., 2016). The MC-based LiDAR backward models include FLIGHT (North et al., 2010) and LIBRAT (Disney et al., 2009).

DART is one of the most comprehensive 3D RT models (Gastellu-Etchegorry et al., 2004; 2015). It uses a unique representation of landscapes and atmosphere, based on facets, turbid medium, and fluids, to simulate the 3D radiative budget as well as satellite, airborne and terrestrial RS observations (*i.e.*, LiDAR signal, reflectance and brightness temperature images) of natural and urban landscapes. For that, DART has two independent and complementary modes: i) DART-FT. It simulates passive optical images using a so-called flux tracking approach that relies on an adaptation of the Discrete Ordinate Method (DOM) (Yin et al., 2013). DART-FT also simulates the solar signal in LiDAR signals. ii) DART-RC. It simulates the signal of multi-platform multi-type LiDAR systems using the Ray Carlo approach, a forward MC-based approach that is accelerated by a “ray-tracking” technique (Gastellu-Etchegorry et al., 2016; Yin et al., 2016). DART-RC has been successfully verified by actual LiDAR measurements (*e.g.*, LVIS data) and cross-comparison against DART-FT. It also has been employed in various LiDAR works, including the development of inversion

algorithms (Grau et al., 2017; Hmida et al., 2017; Lamelas-Gracia et al., 2019; Yin et al., 2020), the design of satellite sensors (*e.g.*, NASA DESDynl, CNES LIDAR mission project) (Durrieu et al., 2013), and impact studies of canopy structures and sensor configurations on LiDAR acquisitions (Grau et al., 2017; Qin et al., 2017; Roberts et al., 2020; Wei et al., 2020), *etc.* However, DART-RC needs great processing time and computer memory to simulate accurate high-PRF LiDAR signals of large and complex landscapes represented by billions of facets. Specifically, a large number of ray samples are required to get converged results and a large memory is required to store complex 3D scene structures. The time constraint is even larger if the solar signal is considered because the laser signal and solar signal are computed separately. All these problems stress the urgent need to improve the DART efficiency of LiDAR signal modeling.

Recent years, some advanced physically-based MC render engines (*e.g.*, *LuxCoreRender*, *Mitsuba*, *Cycles*, *etc.*) in computer graphics domain greatly improved their algorithms in order to simulate images and videos of 3D scenes with much lower time cost and memory request (Iraci, 2013; Nimier-David et al., 2019; Pharr et al., 2016). The MC-based algorithms and implementation ideas of these render engines are great sources of inspiration for RS 3D radiative transfer models, such as DART, LESS (Qi et al., 2019), and ERADIATE (www.eradiate.eu/site/). In particular, DART-Lux, a new MC-based mode of DART, combines the frameworks of DART and the render engine *LuxCoreRender* (https://luxcorerender.org/) in order to reduce the computation time and the needed memory.

However, DART-Lux was initially designed for RS image rendering (Wang et al., 2022; Wang and Gastellu-Etchegorry, 2021), not for LiDAR simulation that emphasizes the measurements of power and ranging. Also, DART-Lux does not support the working mode of LiDAR instruments and the output of multiple LiDAR products. These remarks illustrate the interest to design and implement LiDAR modeling in the DART-Lux.

This paper presents the new LiDAR modeling method in DART-Lux. Section 2 details its novel theory, specifically including how to model the physical process of LiDAR time and power measurements, how to adapt the LiDAR sensor (laser source + receiver) and outputs (waveform, point cloud, photon count) into the DART-Lux model, and how to efficiently handle the LiDAR signal simulation of two sources (laser and sunlight). Section 3 presents four case studies designed to assess the accuracy and performance of DART-Lux LiDAR modeling based on realistic landscapes and different LiDAR configurations. This is followed by discussions in section 4 and finally by a conclusion. The nomenclature is shown in Appendix A.

2. DART-Lux LiDAR model

2.1. Presentation

The innovation of DART-Lux LiDAR model is to transform the laser transport problem into an integration over all possible paths in the 3D landscape that connect the laser source and the receiver. It relies on the light transport equation (Kajiya, 1986), as in Eq. (1).

$$L(r', \Omega_o) = L_e(r', \Omega_o) + \int_{\Omega} L(r', -\Omega_i) \cdot f(r', -\Omega_i, \Omega_o) \cdot |\cos\theta_i| \cdot d\Omega_i \quad (1)$$

where the exitant radiance $L(r', \Omega_o)$ at vertex r' along direction Ω_o is the sum of the emitted radiance $L_e(r', \Omega_o)$ and the scattered radiance due to incident radiance $L(r', -\Omega_i)$ along direction Ω_i (Fig. 1.a), $f(r', -\Omega_i, \Omega_o)$ is the BSDF (bidirectional scattering distribution function) of the surface at vertex r', and θ_i is the incident angle between the incident direction Ω_i and the surface normal vector.

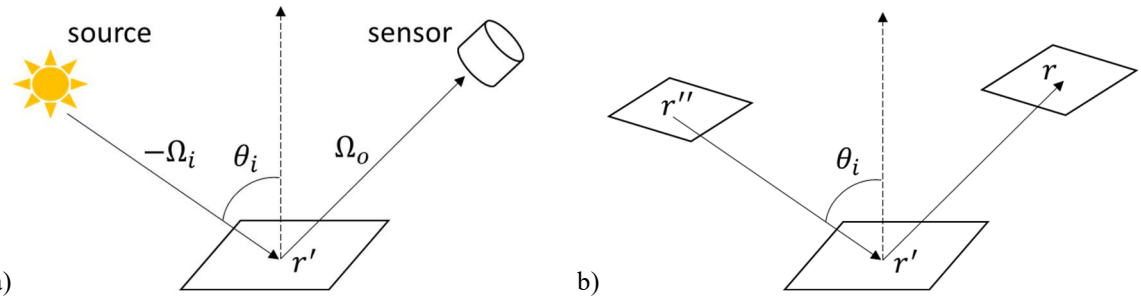


Fig. 1. Light transport equation. a) The exitant radiance along direction Ω_o is the sum of emitted radiance along direction Ω_o and all the scattered radiance to direction Ω_o due to incident radiance from direction $-\Omega_i$. b) The exitant radiance from surface r' to surface r is the sum of emitted radiance and the scattered radiance from all the scene surfaces r''.

Eq. (1) can be transformed into an area integration form instead of solid angle integration form (Wang et al., 2022), as in Eq. (2).

$$L(r' \rightarrow r) = L_e(r' \rightarrow r) + \int_{A} L(r'' \rightarrow r') \cdot f(r'' \rightarrow r' \rightarrow r) \cdot G(r'' \leftrightarrow r') \cdot dA(r'') \quad (2)$$

where r'' and r are the previous vertex and next vertex of r' in the light transport process (Fig. 1.b), $dA(r'')$ is the area at vertex r'', and the connection function $G(r'' \leftrightarrow r')$ between

166 vertices r'' and r' is the product of $|\cos\theta_i|$ and the Jacobian term to transfer solid angle
167 integration over area integration.
168
169 In absence of intercepting medium between vertices r' and r, the exitant radiance $L(r' \to r)$
170 from r' is equal to the incident radiance at the next intersection point r. Further, since laser
171 rays might be scattered multiple times in the 3D landscape, the received laser power can be
172 represented by incrementally expanding Eq. (2) to an infinite sum of multi-dimensional
173 integral, which can be expressed as a Lebesgue integral (Veach, 1997), as in Eq. (3).

$$\Phi_{\text{lidar}} = \int_{\mathcal{D}} f(\bar{r}) \cdot d\mu(\bar{r}) \tag{3}$$

174 where Φ_{lidar} is the total laser power received by the LiDAR sensor. $\mathcal{D}$ is the set of all paths.
175 $\mathcal{D} = \bigcup_{n=1}^{\infty} \mathcal{D}_n$ with $\mathcal{D}_n$ the set of paths with *n* edges (*n* is also called path depth). $\bar{r}$ is a path of
176 the set of paths that connect the laser source and the receiver; it passes through a series of
177 vertices in the scene: $\bar{r} \in \mathcal{D} = \left\{\bar{r}_n \middle| \bar{r}_n = r_0, r_1, \dots, r_n;\ \ r_{k=0,1,\dots,n} \in A,\ n \in \mathbb{Z}^+\right\}$. $f(\bar{r})$ is the
178 power contribution of path $\bar{r}$. $d\mu(\bar{r})$ is the area-product measure of path $\bar{r}$, $d\mu(\bar{r}_n) =$
179 $dA(r_n) \cdot dA(r_{n-1}) \cdots dA(r_0)$. Note that Eq. (3) is originally defined for the image radiance in the
180 study of Veach (1997) and here we extend it to the LiDAR power.
181
182 In LiDAR modeling, the returned power and time are both recorded. The simulated temporal
183 power profile is the so-called waveform. Therefore, the LiDAR power Φ_{lidar} in Eq. (3) is not
184 an instantaneous quantity. It must be expanded to the time space, as in Eq. (4).

$$\Phi_{\text{lidar}}\left(\tau(\bar{r})\right) = \int_{\mathcal{D}} f(\bar{r}) \cdot d\mu(\bar{r}) \tag{4}$$

where $\tau(\bar{r})$ is the laser travelled time along path $\bar{r}$, and $\Phi_{\text{lidar}}\left(\tau(\bar{r})\right)$ is the returned laser power at the time $\tau(\bar{r})$.

A general-purpose method such as MC can assess the high-dimensional integral form of LiDAR waveform $\Phi_{\text{lidar}}\left(\tau(\bar{r})\right)$. Indeed, with $\bar{R} \in \mathcal{D}$ a random path sample and $p(\bar{R})$ its probability distribution density (PDF), $\frac{f(\bar{R})}{p(\bar{R})}$ is an unbiased estimate of $\Phi_{\text{lidar}}\left(\tau(\bar{r})\right)$ because the expected value $\mathbb{E}\left[\frac{f(\bar{R})}{p(\bar{R})}\right] = \int_{\mathcal{D}} \frac{f(\bar{r})}{p(\bar{r})} \cdot p(\bar{r}) \cdot d\mu(\bar{r})$ is equal to the LiDAR measurement $\Phi_{\text{lidar}}\left(\tau(\bar{r})\right)$.

Based on the above core theory, DART-Lux implements an efficient path sampling method and a specific MC integration method to solve the LiDAR RT equation (Eq. (4)) and simulate the LiDAR signal optimally. Also, DART-Lux simulates the illumination and receiving modes of LiDAR sensor and the output formats of multiple LiDAR products by abstracting the actual systems. The following sections present the detailed theory of DART-Lux LiDAR modeling:

- LiDAR instrument modeling. The illumination mode of laser source and the receiving mode of LiDAR receiver are quantitatively modeled in a formulaic manner.
- LiDAR waveform modeling. The integral form of LiDAR power profile $\Phi_{\text{lidar}}\left(\tau(\bar{r})\right)$ is computed based on a comprehensive RT framework. It involves four distinct modeling stages: efficient path generation with a bidirectional path tracing (BDPT) method, laser

travelled time computation, laser power evaluation, and waveform modeling with a multiple importance sampling (MIS) method. By inputting the modeled LiDAR instruments and 3D landscapes into the framework, the LiDAR waveform can be output.

- Solar signal modeling. A direct light sampling strategy is used to solve the modeling of "2 source (sun and laser) + 1 sensor" configuration. It simulates the solar and laser signals synchronously by maximizing the use of sampled surface vertices.

- Extension from single-pulse waveform to multi-pulse multi-platform multi-type LiDAR products.

2.2. LiDAR instrument modeling

In order to simulate LiDAR signals, the LiDAR sensor (laser source + receiver) needs to be created. To adapt the LiDAR sensor into the BDPT method of DART-Lux, the source and receiver are both designed as the start points to trace paths. Below we introduce the modeling of LiDAR sensor and the method of starting paths.

2.2.1. Laser source

The laser source is usually considered as a point source located at $P_L(x, y, H)$, emitting the energy through a conical solid angle $\Delta\Omega_{\text{fp}}$ with the central direction $\vec{\Omega}_L(\theta_L, \varphi_L)$. To start a light sub-path, the first vertex p_0 is sampled on the laser source point. The sampled PDF $p(\vec{p_0})$ is expressed as the Dirac delta function on the point P_L, as in Eq. (5).

$$p(\vec{p_0}) = \delta(p_0 - P_L) \tag{5}$$

where $\delta(p_0 - P_L)$ satisfies $\int_A \ \delta(p_0 - P_L) \cdot dA(p_0) = 1$.

Once p_0 is determined, an exitant direction Ω is sampled within the solid angle $\Delta\Omega_{\text{fp}}$ with a directional PDF $p(p_0, \Omega)$. Generally, a nadir emitted laser energy follows a 2D Gaussian distribution within the illumination region (footprint), and the energy decreases to a fraction η from footprint center to edge; usually $\eta = e^{-2}$. Based on this setting, we get:

$$p(p_0, \Omega) = \frac{\exp\left(-\frac{\tan^2\theta}{2\sigma^2}\right)}{2\pi\sigma^2 \cdot \cos^3\theta \cdot (1-\eta)} \tag{6}$$

where θ is the zenith angle of the exitant laser ray in local sensor coordinate; $\theta \leq \theta_{\text{fp}}$ (θ_{fp} is the laser beam divergence half angle). σ is the standard deviation of 2D Gaussian distribution. Appendix B details the derivation of Eq. (6).

The emitted radiance $L_e(p_0, \Omega)$ is calculated based on the laser energy and the direction of sampled laser ray. First, the emitted power $\Phi_e(p_0, \tau)$ is computed from the energy Q_e of the emitted laser pulse with duration interval $\Delta\tau$ and energy time distribution function $p(\tau)$, as in Eq. (7).

$$\Phi_e(p_0, \tau) = Q_e \cdot p(\tau) \tag{7}$$

where $p(\tau)$ satisfies $\int_0^{\Delta\tau} p(\tau) \cdot d\tau = 1$; $p(\tau)$ generally follows a 1D Gaussian distribution.

Actually, since the temporal distribution of the returned signal is the same for the power emitted at any time, we compute the signal with the instantaneous constant power $\Phi_e(p_0)$

(*i.e.*, $\Phi_e(p_0) = \Phi_e(p_0, \tau = 0) = Q_e \cdot \delta(\tau)$ with the assumption $p(\tau) = \delta(\tau)$), and convolve it with the temporal distribution. Then, the emitted radiance $L_e(p_0, \Omega)$ of the sampled ray is derived from the equality $E_{p_1} = \Phi_e(p_0) \cdot p(p_0, \Omega) \cdot \left|\cos \theta_i^{p_1}\right| = \int_\Omega L_e(p_0, \Omega) \cdot \left|\cos \theta_i^{p_1}\right| \cdot d\Omega$ with E_{p_1} incident irradiance and $\theta_i^{p_1}$ incident angle at vertex p_1. That is, as in Eq. (8).

$$L_e(p_0, \Omega) = \Phi_e(p_0) \cdot p(p_0, \Omega) \cdot \delta(p_0 - P_L) \tag{8}$$

2.2.2. LiDAR receiver

The LiDAR receiver in DART-Lux is abstracted from the actual system (Kokkalis, 2017). Specifically, an actual receiver (Fig. 2.a) collects incident rays with a convex telescope (L1 in Fig. 2.a) with radius r_{tel} and then transmits them to the photomultiplier tube detector (PMT) through focusing optics. The FOV (field of view) half angle (θ_{FOV}) of the LiDAR receiver is limited by a circular diaphragm S1 located at the focus of L1. This causes that the rays with incident angle larger than θ_{FOV} are not transmitted (*e.g.*, green rays in Fig. 2.a) and each point on LiDAR telescope has a same circular FOV. In this case, the LiDAR receiver can be treated like a pinhole camera with the lens area $A_{\text{tel}} = \pi r_{\text{tel}}^2$ (Fig. 2.b). The incident power is projected onto a virtual sensor plane at a distance F behind the receiver. Generally, F is set to 1. The sensor plane is a circular disk with radius $r_{\text{sensor}} = F \cdot \tan \theta_{\text{FOV}} + r_{\text{tel}}$. For each point on the receiver, incident power can only be projected onto the sub sensor plane with radius $r_{\text{sensor}}^{\text{sub}} = F \cdot \tan \theta_{\text{FOV}}$.

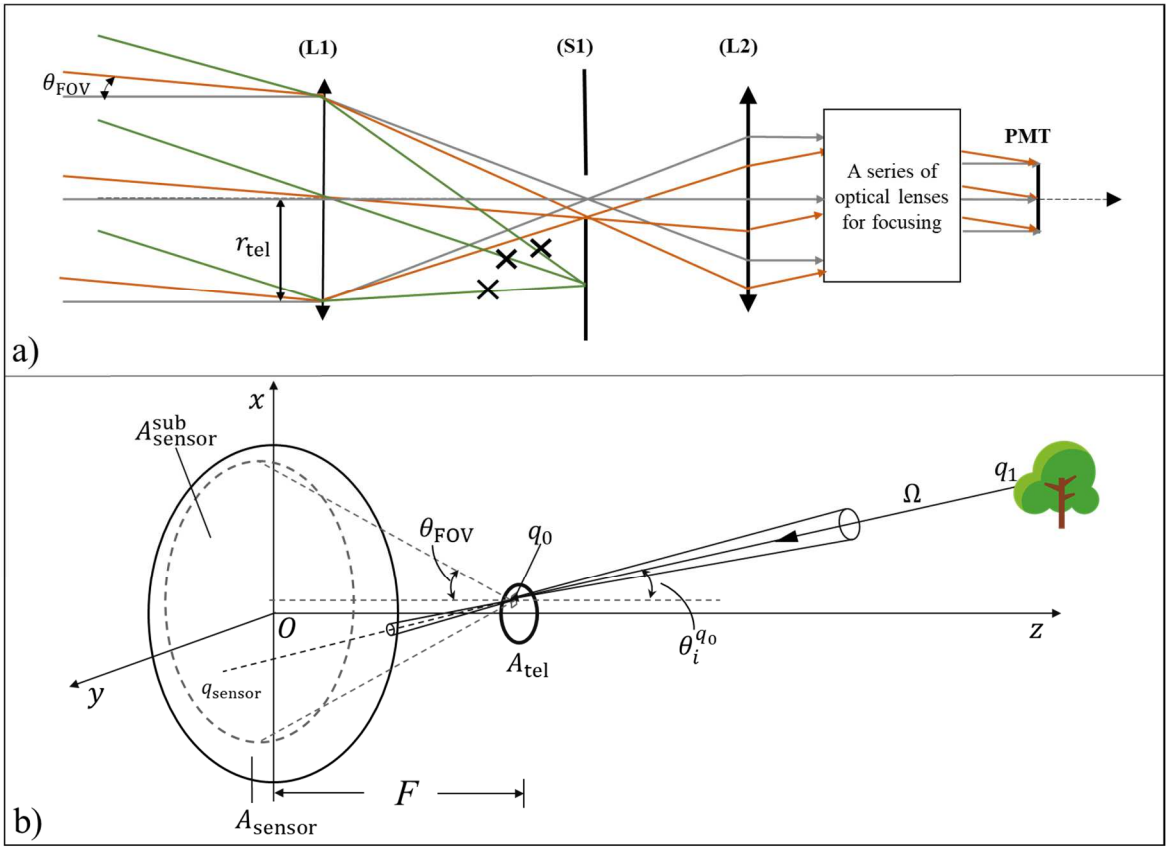


Fig. 2. a) Actual LiDAR receiving system with a telescope L1, a diaphragm S1, a collimating lens L2, a set of optical lenses, and a PMT. On-axis (gray lines) and off-axis (orange lines) rays with incident angle smaller than θ_{FOV} reach the PMT finally, whilst rays (green lines) with incident angle larger than θ_{FOV} are intercepted by S1. b) Modeled LiDAR receiver (A_{tel}) with a virtual sensor plane (A_{sensor}). Energy flux from scene surface q_1 in the scene along direction Ω is focused by the receiver vertex q_0 onto the vertex q_{sensor} on sensor plane.

To start a random walk from the receiver, the vertex q_0 is first randomly sampled on the lens with the PDF $p(q_0^{\leftarrow}) = \frac{1}{A_{\mathrm{tel}}}$. Once the receiver vertex q_0 is determined, a point q_{sensor} is sampled on the sub sensor plane of q_0 with the PDF $p(q_{\mathrm{sensor}}) = \frac{1}{A_{\mathrm{subsen}}}$. Then the exitant direction Ω of receiver ray is determined by connecting q_{sensor} and q_0, and the corresponding directional PDF $p(q_0, \Omega)$ is computed as in Eq. (9).

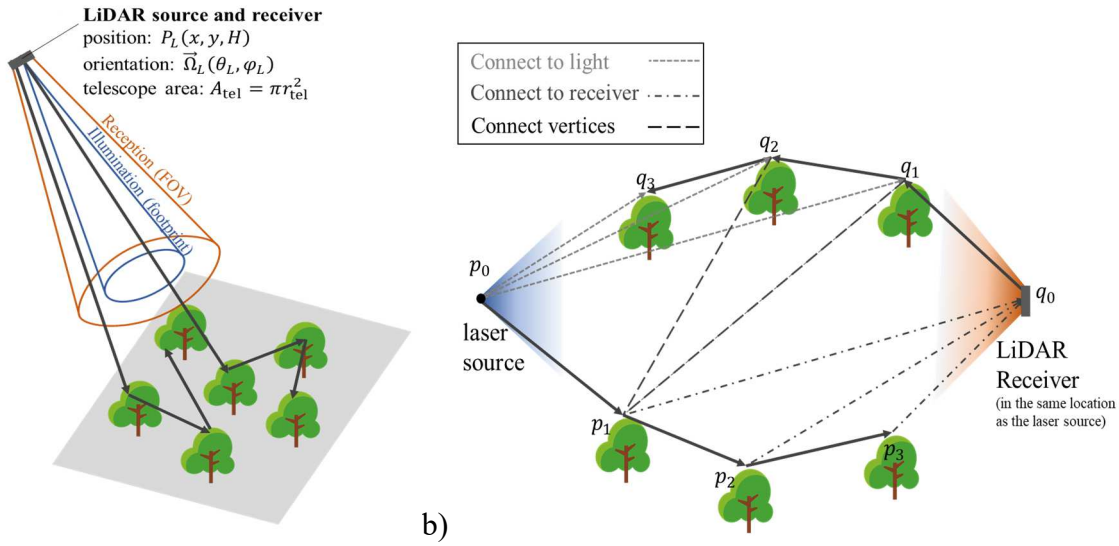
$$p(q_0, \Omega) = \frac{1}{A_{\text{subsen}} \cos^3 \theta_i^{q_0}} \quad (9)$$

with $\theta_i^{q_0}$ the incident angle at vertex q_0 between incident direction and lens normal vector (Fig. 2.b).

2.3. LiDAR waveform modeling

2.3.1. Efficient path generation

DART-Lux samples paths with a BDPT method (Veach, 1997). Fig. 3 shows the schematic diagram of the BDPT method for LiDAR simulation. Generally, the source and receiver of LiDAR sensor are considered to have the same central position $P_L(x, y, H)$ and orientation $\vec{\Omega}_L(\theta_L, \varphi_L)$ (Fig. 3.a). The paths from laser source to receiver are constructed based on the combination of two sub-paths that start from the source and the receiver, respectively. In order to illustrate the DART-Lux LiDAR model intuitively, the sensor configuration in Fig. 3.a is abstractly displayed as a "source + receiver" configuration in Fig. 3.b.



a) b)

Fig. 3. Schematic diagram of DART-Lux LiDAR model. a) A sub-path starts from the LiDAR source and a sub-path starts from the receiver (at the same location as the source). b) A series of paths that connect the LiDAR source and receiver are generated by the BDPT method.

In the BDPT method, the first random walk starts from the laser source and samples a light sub-path $\bar{p}_{N_s}=p_0, p_1, \dots, p_{N_s\text{-}1}$ with N_s vertices, of which vertex p_0 is on the laser source. Each vertex is sampled with the PDF $p(\overrightarrow{p_k})$ $(k=0,1,\dots,N_s\text{-}1)$. The second random walk starts from the receiver and samples a receiver sub-path $\bar{q}_{N_t}=q_0, q_1, \dots, q_{N_t\text{-}1}$ with N_t vertices, of which vertex q_0 is on the receiver. Each vertex is sampled with the PDF $p(\overleftarrow{q_k})$ $(k=0,1,\dots,N_t\text{-}1)$. The combination of the two sub-paths leads to the construction of a series of complete paths from source to receiver by using three sampling techniques: i) "Connect to light": the vertex on the receiver sub-path is connected to a sampled vertex on the laser source; ii) "Connect to receiver": the vertex on the light sub-path is connected to a sampled vertex on the receiver; iii) "Connect vertices": a vertex on the light sub-path is connected to a vertex on the receiver sub-path. Also, a light sub-path can randomly hit the LiDAR receiver, and a receiver sub-path can randomly hit the laser source.

Here, we illustrate the creation of a typical path: when the vertex $p_{s\text{-}1}$ on the light sub-path $\bar{p}_{N_s}$ and the vertex $q_{t\text{-}1}$ on the receiver sub-path $\bar{q}_{N_t}$ are mutually visible, a path $\bar{r}_{s,t}$ from source to receiver is generated by connecting the light sub-path $\bar{p}_s$ and the receiver sub-path $\bar{q}_t$. Eq. (10) shows the resulting path $\bar{r}_{s,t}$.

$$\bar{r}_{s,t} = \bar{p}_s, \bar{q}_t = p_0, p_1, \dots, p_{s-1}, q_{t-1}, q_{t-2}, \dots, q_0 \quad (10)$$

with $0 \leq s \leq N_s$, $0 \leq t \leq N_t$, $s + t = n - 1$, $\bar{p}_s$=$p_0, p_1, \dots, p_{s-1}$, and $\bar{q}_t$=$q_0, q_1, \dots, q_{t-1}$.

2.3.2. Laser travelled time computation

The length $l(\bar{r}_{s,t})$ of path $\bar{r}_{s,t}$ is the sum of the lengths of the light sub-path, the receiver sub-path, and the distance between the connected vertices p_{s-1} and q_{t-1}, as in Eq. (11).

$$l(\bar{r}_{s,t}) = \sum_{i=1}^{s-1} \|p_{i-1} - p_i\| + \|p_{s-1} - q_{t-1}\| + \sum_{j=1}^{t-1} \|q_{i-1} - q_i\| \quad (11)$$

The path length $l(\bar{r}_{s,t})$ is then converted to the laser travelled time $\tau(\bar{r}_{s,t})$, as in Eq. (12).

$$\tau(\bar{r}_{s,t}) = \frac{l(\bar{r}_{s,t})}{c} \quad (12)$$

with c the light speed.

2.3.3. Laser returned power evaluation

The power contribution divided by the path PDF ($\frac{f(\bar{r}_{s,t})}{p(\bar{r}_{s,t})}$) of a random path $\bar{r}_{s,t}$ can unbiasedly estimate the laser returned power Φ_{lidar}. It is computed by considering the ray propagation and interactions. Specifically, the light sub-path $\bar{p}_s$ starts from the laser source vertex p_0 that is sampled by the PDF $p(\vec{p_0})$, carries an emitted radiance $L_e(p_0 \rightarrow p_1)$, and shoots towards the surface vertex p_1. During the forward random walk, the radiance is incrementally scaled using the light transfer equation (Eq. (2)). Up to the vertex $p_{s\text{-}1}$, the estimate of the radiance from $p_{s\text{-}2}$ to $p_{s\text{-}1}$ is computed as in Eq. (13).

$$L(p_{s\text{-}2}\rightarrow p_{s\text{-}1})=\frac{L_e(p_0\rightarrow p_1)}{p(p_0^{\rightarrow})}\cdot\prod_{k=1}^{s-1}\left[\frac{f(p_{k\text{-}2}\rightarrow p_{k\text{-}1}\rightarrow p_k)\cdot G(p_{k\text{-}1}\leftrightarrow p_k)}{p(p_k^{\rightarrow})}\right] \quad (13)$$

where the virtual BSDF $f(p_{-1}\rightarrow p_0\rightarrow p_1)\equiv 1$ is introduced to reduce special cases in the mathematical formulation.

For the receiver sub-path $\bar{q}_t$ traced backwards, DART-Lux imports an importance function W_e into the receiver, which is used to convert the incident energy flux into the measured quantity (here laser power is exactly energy flux; $W_e\equiv 1$). In the process of backward tracing, W_e is treated as a quantity that exits the receiver like the emitted radiance L_e of the source. The backward importance transport starting from the receiver is treated as symmetrical to the forward light transport from the source. Hence, the estimate of importance from $q_{t\text{-}2}$ to $q_{t\text{-}1}$ is computed as in Eq. (14).

$$W(q_{t\text{-}2}\rightarrow q_{t\text{-}1})=\frac{W_e}{p(q_0^{\leftarrow})}\cdot\prod_{k=1}^{t-1}\left[\frac{f(q_k\rightarrow q_{k\text{-}1}\rightarrow q_{k\text{-}2})\cdot G(q_{k\text{-}1}\leftrightarrow q_k)}{p(q_k^{\leftarrow})}\right] \quad (14)$$

with $f(q_1\rightarrow q_0\rightarrow q_{-1})\equiv 1$.

The connection of vertices p_{s-1} and q_{t-1}, combined with Eq. (13) and Eq. (14), leads to the power estimate $\frac{f(\bar{r}_{s,t})}{p(\bar{r}_{s,t})}$ of laser power integration, as in Eq. (15).

$$\frac{f(\bar{r}_{s,t})}{p(\bar{r}_{s,t})}=L(p_{s\text{-}2}\rightarrow p_{s\text{-}1})\cdot f(p_{s\text{-}2}\rightarrow p_{s\text{-}1}\rightarrow q_{t\text{-}1})\cdot G(p_{s\text{-}1}\leftrightarrow q_{t\text{-}1})\cdot f(p_{s\text{-}1}\rightarrow q_{t\text{-}1}\rightarrow q_{t\text{-}2})\cdot W(q_{t\text{-}2}\rightarrow q_{t\text{-}1}) \quad (15)$$

2.3.4. Laser signal modeling

DART-Lux uses a robust MC integration method — MIS, to get an unbiased estimate of LiDAR power profile $\Phi_{\text{lidar}}\left(\tau(\bar{r})\right)$. The core idea of MIS is to use multiple path sampling

techniques to evaluate the integral and then combine the path contribution in a weighted manner (Veach, 1997). Eq. (16) expresses the MIS estimator of $\Phi_{\text{lidar}}\left(\tau(\bar{r})\right)$:

$$\widetilde{\Phi}_{\text{lidar}}(\tau) = \sum_{s\geq 0}\sum_{t\geq 0} w\left(\bar{r}_{s,t}\right)\cdot\frac{f\left(\bar{r}_{s,t}\right)}{p\left(\bar{r}_{s,t}\right)},\quad \text{with } l\left(\bar{r}_{s,t}\right) = c\cdot\tau \tag{16}$$

where $w\left(\bar{r}_{s,t}\right)$ is the MIS weight of the sampled path $\bar{r}_{s,t}$.

$\widetilde{\Phi}_{\text{lidar}}(\tau)$ can unbiasedly estimate the laser power profile if the MIS weight satisfies $\sum_{\substack{s\geq 0,t\geq 0\\ s+t=n-1}} w\left(\bar{r}_{s,t}\right) = 1$ whenever $f\left(\bar{r}_{s,t}\right)\neq 0$, and $w(\bar{r}_{s,t})=0$ whenever $p\left(\bar{r}_{s,t}\right)=0$. The proof is given in Eq. (17):

$$\begin{aligned}\mathbb{E}\left[\widetilde{\Phi}_{\text{lidar}}(\tau)\right] &= \sum_{s\geq 0}\sum_{t\geq 0}\int_{\mathcal{D}} \frac{w\left(\bar{r}_{s,t}\right)\cdot f\left(\bar{r}_{s,t}\right)}{p\left(\bar{r}_{s,t}\right)}\cdot p\left(\bar{r}_{s,t}\right)\cdot d\mu(\bar{r})\\ &= \sum_{n=1}^{\infty}\int_{\mathcal{D}_n}\sum_{\substack{s\geq 0,t\geq 0\\ s+t=n-1}} w\left(\bar{r}_{s,t}\right)\cdot f\left(\bar{r}_{s,t}\right)\cdot d\mu(\bar{r}) = \sum_{n=1}^{\infty}\int_{\mathcal{D}_n} f(\bar{r}_n)\cdot d\mu(\bar{r}_n)\\ &= \int_{\mathcal{D}} f(\bar{r})\cdot d\mu(\bar{r}) = \Phi_{\text{lidar}}(\tau)\end{aligned} \tag{17}$$

In practice, in order to get an accurate estimation of laser power profile, a large number of Monte Carlo experiments independent of each other are required to be implemented. Then, the result of MIS estimator would converge to the expected value $\Phi_{\text{lidar}}(\tau)$.

Since the actual LiDAR waveform measures the returned signal by sampling per fixed time interval (also called bin ξ, $\xi \in \mathbb{N}$), the simulated power profile should be treated bin by bin.

The time interval of each bin is $\frac{1}{f}$, with f the signal sampling frequency. The laser energy per bin $Q_{\text{lidar}}(\xi)$ is the sum of the returned power during the time interval $\left[\frac{\xi}{f}, \frac{\xi+1}{f}\right)$, as in Eq. (18).

$$Q_{\text{lidar}}(\xi) = \sum_{\tau \in \left[\frac{\xi}{f}, \frac{\xi+1}{f}\right)} \Phi_{\text{lidar}}(\tau) \tag{18}$$

The laser energy profile $Q_{\text{lidar}}(\xi)$ can be further converted to the number of laser photons $N_{\text{lidar}}(\xi)$ by dividing the photon energy, as in Eq. (19).

$$N_{\text{lidar}}(\xi) = \frac{Q_{\text{lidar}}(\xi)}{h\upsilon} \tag{19}$$

with h the Plank's constant, and υ the photon frequency.

The simulated laser source emits an instantaneous laser power, whereas an actual laser pulse is emitted during several nanoseconds. The temporal distribution of the emitted laser energy must be considered. Specifically, the 1D Gaussian temporal distribution of the emitted pulse $p(\tau)$ is first discretized into $p(\xi)$ in bins, and then the simulated LiDAR waveform is the convolution of the non-convolved waveform $N_{\text{lidar}}(\xi)$ and the temporal distribution $p(\xi)$.

2.4. Solar signal modeling

DART-Lux uses a direct light sampling strategy to handle the signal simulation of "2 sources (laser and sun) + 1 sensor" configuration (Fig. 4.a) (Shirley et al., 1996). Its core idea is to sample one source from the light group in MC experiments. Specifically, the laser source and sun are selected with the equal probability $\mathrm{P}_{\text{light}} = 0.5$ when sampling a source vertex in

372 "Connect to receiver" and "Connect to light" (Fig. 4.b). If the laser source is selected, the
373 paths that connect LiDAR source and receiver are reconstructed and their power and length
374 are computed for retrieving the laser signal. If the sunlight is selected, the paths that connect
375 sunlight and receiver are reconstructed for retrieving the solar signal.
376
377 For the solar signal modeling, two assumptions are made: the Earth scene is in a sphere of
378 radius R, and parallel sunlight arises from a virtual disk at infinity with area $A_{\text{disk}}=\frac{1}{\pi R^2}$. To
379 start a random walk from the sun, a light source point p_0 is uniformly sampled on the virtual
380 disk with the PDF $p(\vec{p_0})=\frac{1}{A_{\text{disk}}}$. Since there is only one sunlight direction Ω_s, the PDF of the
381 emitted sunlight ray $p(p_0,\Omega)$ is the Dirac delta function $\delta(\Omega-\Omega_s)$. The equality
382 $E_{p_1}=E_s\cdot\cos\theta_i^{p_1}=\int_\Omega L_e(p_0,\Omega)\cdot\cos\theta_i^{p_1}\cdot d\Omega$ with E_s the solar irradiance is used to compute
383 the exitant sun radiance: $L_e(p_0,\Omega)=E_s\cdot\delta(\Omega-\Omega_s)$.
384
385 Based on the characteristics of sun and LiDAR receiver, a series of paths that connect the sun
386 and the receiver are sampled using the BDPT method. The received solar power is computed
387 and then converted to the number of solar photons per bin (N_{SN}) reaching the PMT. Since the
388 solar energy is continuous, the LiDAR waveform with solar signal is simulated by adding the
389 same number of solar photons to each bin of the convolved waveform.
390

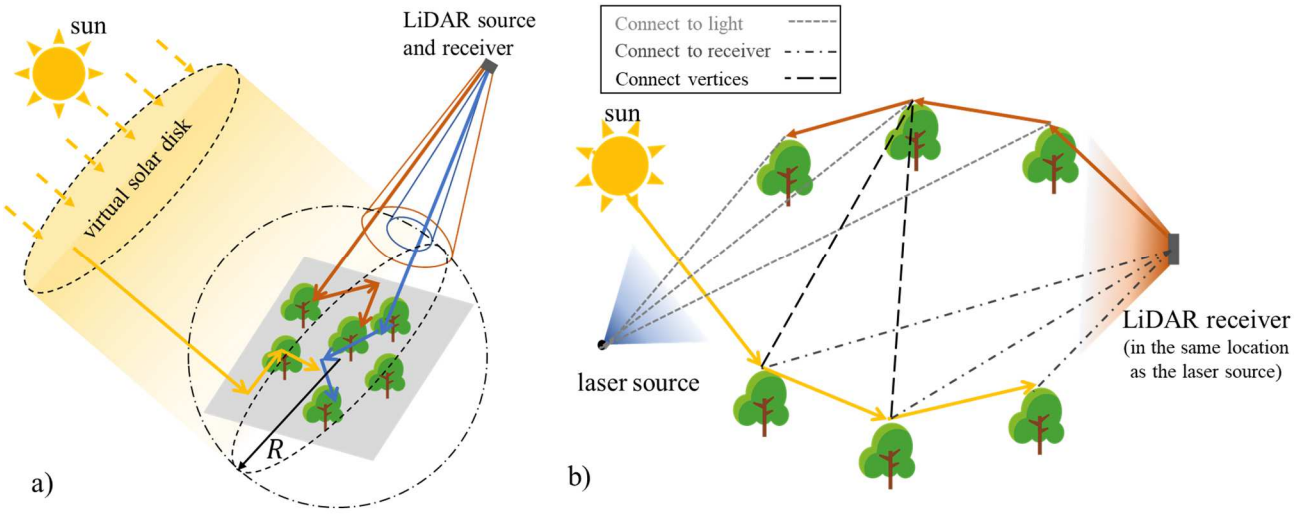


Fig. 4. Direct light sampling strategy in a "2 sources + 1 receiver" configuration. a) The light sub-path starts from the sunlight (yellow line) or laser source (blue line). The receiver sub-path starts from the LiDAR receiver (orange line). b) One example of direct light sampling: the laser source is selected in "Connect to light" and the sunlight is selected in "Connect to receiver".

2.5. Multi-platform multi-type multi-pulse LiDAR modeling

The above sections present the DART-Lux modeling of single-pulse waveform LiDAR (wLiDAR) with solar signal. Furthermore, DART-Lux extends the single-pulse-wLiDAR component to simulate multi-platform (satellite, airborne, terrestrial) multi-type (waveform, discrete return, photon counting) multi-pulse LiDAR signals using the method referred to Yin et al. (2016). The multi-pulse simulation of laser scanning systems is an iterative loop of the single-pulse modeling with specific geometry configurations and methods of detection and digitization:

- The point cloud of discrete return LiDAR (drLiDAR) is converted from the simulated

waveform using Gaussian decomposition or peak detection methods.

- The points of photon counting LiDAR (pcLiDAR) are derived per waveform using a statistical method based on the instrumental parameters of the single-photon detector.

In practice, the conversion from wLiDAR data to dr/pcLiDAR data is less time-consuming than the simulation of wLiDAR, and the time of multi-pulse simulation is almost proportional to number of pulses and time of single-pulse simulation. These remarks stress the importance of improving and validating the efficiency and accuracy of wLiDAR simulation.

3. Consistency validation of DART-Lux LiDAR model

This section assesses the performance of DART-Lux LiDAR model by using DART-RC, the standard LiDAR mode of DART, as a reference model. This work is done for two realistic landscapes (forest and citrus orchard), and for different LiDAR configurations (*i.e.*, wLiDAR, drLiDAR, wLiDAR with solar signal, and pcLiDAR with solar signal).

3.1. Simulation of single-pulse wLiDAR

The wLiDAR simulation is performed on the Jarvselja birch forest of the RAMI-IV experiment (Widlowski et al., 2015). This realistic forest scene contains 1029 trees from 7 species. Fig. 5.a shows the tree distribution map and Fig. 5.b shows the 3D representation. The LiDAR sensor mimics the spaceborne GEDI instrument (Table 1), with the footprint

426 diameter of 25 m and nadir viewing. The maximum scattering order of rays is set to a large
427 value (*i.e.*, 100) due to the strong canopy scattering in 1064 nm and large footprint region.
428
429 The consistency of DART-Lux LiDAR modeling is first verified by comparing DART-RC
430 and DART-Lux waveforms simulated with a very large number of rays (N_{rays}). Fig. 5.c
431 stresses that with $N_{rays}=10^9$, the DART-RC and DART-Lux waveforms are almost identical:
432 coefficient of determination $R^2 = 1.0$ and relative root-mean-square error rRMSE = 0.21%.
433 This result illustrates high consistency of two models for very large N_{rays}.
434
435 Using the waveform simulated with 10^9 rays as the reference, the convergence and efficiency
436 of DART-Lux is further investigated by comparing the accuracy and processing time of
437 DART-Lux and DART-RC waveforms simulated with relatively few rays. Fig. 5.d and Table
438 2 show how the accuracies of DART-Lux and DART-RC vary with the number of rays by
439 considering $N_{rays} = 10^2$, 10^3, 10^4 and 10^5. As expected, the DART-Lux and DART-RC
440 waveforms both gradually converge to the reference waveform with the increment of N_{rays}.
441 For DART-Lux, the R^2 (rRMSE) is 0.9818 (21.42%) for 10^2 rays, 0.9975 (7.28%) for 10^3
442 rays, 0.9997 (2.48%) for 10^4 rays, and 1.0000 (0.88%) for 10^5 rays. For DART-RC, the R^2
443 (rRMSE) is 0.9775 (28.09%) for 10^2 rays, 0.9958 (9.40%) for 10^3 rays, 0.9990 (4.79%) for
444 10^4 rays, 0.9999 (1.63%) for 10^5 rays. It is observed that DART-Lux converges faster than
445 DART-RC: with a same number of rays, DART-Lux is always more accurate than DART-RC,
446 as in the example of Fig. 5.e. The faster convergence of DART-Lux benefits from the better

efficiency of the BDPT method on path reconstruction. Indeed, in this experiment, DART-Lux reconstructs averagely 3.19 paths per ray whereas DART-RC only reconstructs 2.07 paths. In terms of processing time, DART-Lux is faster. It needs ~2.29×10^{-4} s per ray whereas DART-RC needs ~4.74×10^{-4} s per ray. Overall, DART-Lux gives the same accuracy as DART-RC with fewer rays and shorter processing time. For instance, to get rRMSE ≈ 5%, DART-Lux needs about 5×10^{3} rays, which takes about 1.15 s, whereas DART-RC needs about 10^{4} rays, which takes about 4.74 s.

The adaptability of DART-Lux to LiDAR systems with various footprint diameters (D) is also studied. Fig. 5.f shows the DART-Lux waveforms with D = 0.05 m, 0.1 m, 0.5 m, 1 m, 10 m, 20 m, and 30 m, which correspond to small-to-large footprints. The waveform greatly changes with the increase of D. For small D, it is essentially a single peak that corresponds to the local top canopy. For increasing D, there appears signal from other parts of the canopy, including the ground, and the waveform has multiple peaks, which reveals forest vertical distribution at a large scale. Table 2 gives the number of rays $N_{rays,min}$ that is needed to get rRMSE ≤ 5% in DART-Lux and DART-RC for various footprint diameters. For both models, $N_{rays,min}$ increases with D. Indeed, for large footprints, there are more path possibilities, and consequently a large number of rays is required to reach convergence. For a same footprint size, DART-Lux usually needs less rays than DART-RC to get the same accuracy. This advantage greatly increases with the footprint size. Indeed, the BDPT method is more advantageous if rays are frequently scattered, whereas there is little multiple scattering in

small-footprint LiDAR (Gastellu-Etchegorry et al., 2016).

Table 1. Configurations of GEDI, Riegl VUX-IUAV and MABEL LiDAR devices.

Parameters	LiDAR devices		
	GEDI	Riegl VUX-1UAV	MABEL
Type	Spaceborne wLiDAR	UAV drLiDAR	Airborne pcLiDAR
Sensor area	0.5 m^2	0.1 m^2	-
Band width	0.8 nm	-	0.03 nm / 0.4 nm
Acquisition period per bin	1 ns	0.067 ns	0.083 ns
Footprint divergence half angle	0.06 mrad	0.25 mrad	0.05 mrad
FOV divergence half angle	0.3 mrad	0.375 mrad	0.105 mrad
Pulse energy	10.5 mJ	1 mJ	6 μJ
Wavelength	1064 nm	1064 nm	532 nm / 1064 nm
Platform altitude	419 km	250 m	20 km
Along-track distance step per scan	-	0.3 m	-
Look angle step per pulse	-	0.045°	-
Maximal number of targets per pulse (drLiDAR)	-	4	-
Quantum efficiency (pcLiDAR)	-	-	12.5% / 1.5%
Detector dead time (pcLiDAR)	-	-	2.5 ns
Dark count rate (pcLiDAR)	-	-	50 Hz

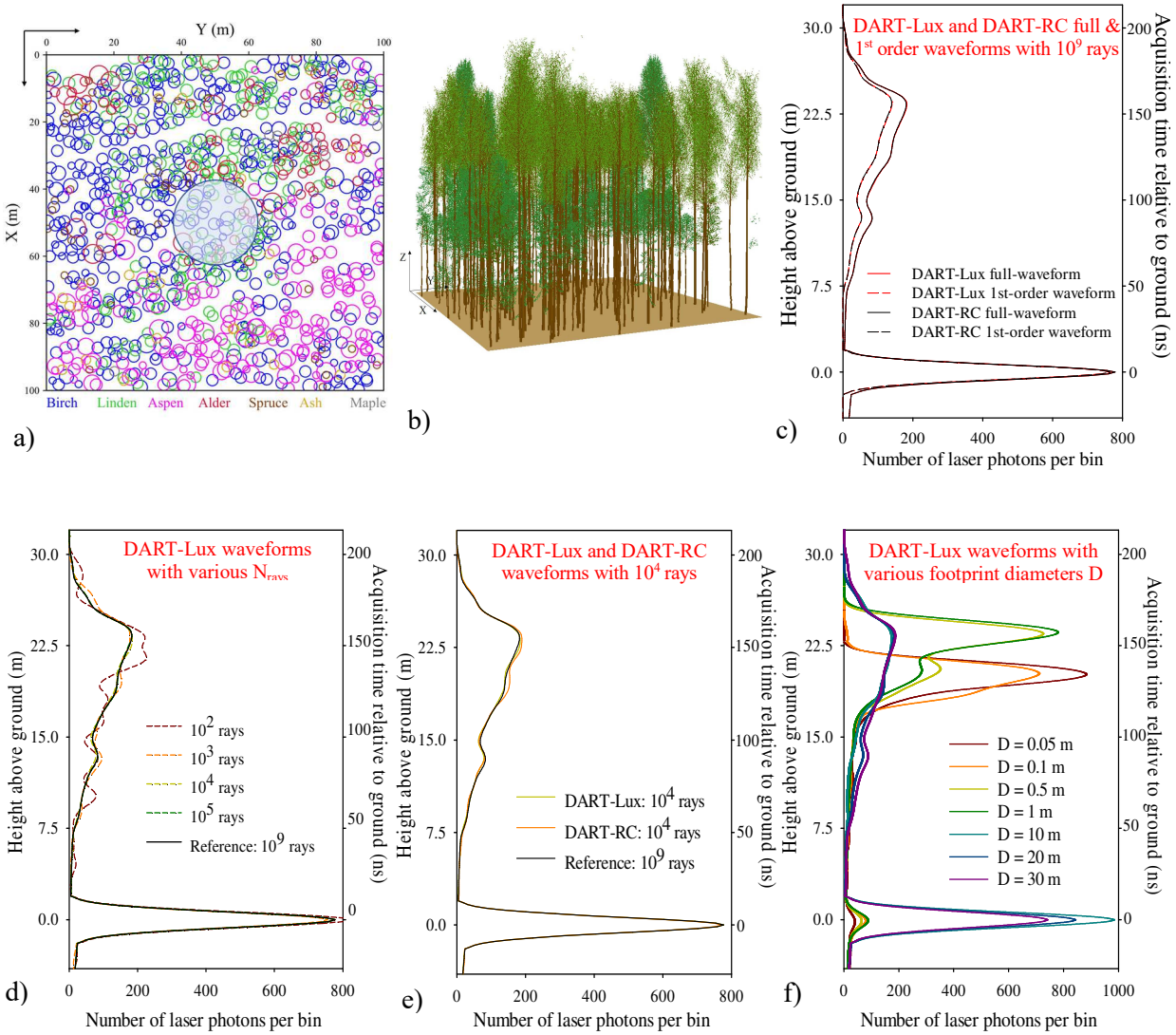


Fig. 5. Single-pulse wLiDAR in the Jarvselja birch stand. a) Tree species map with LiDAR footprint (D = 25 m) centered at (50 m, 50 m). b) 3D view of part of the scene. c) DART-Lux and DART-RC waveforms simulated with 10^9 rays. It is the reference waveform. d) Reference waveform and DART-Lux waveforms simulated with 10^2, 10^3, 10^4 and 10^5 rays. e) Reference waveform and DART-Lux and DART-RC waveforms simulated with 10^4 rays. f) DART-Lux waveforms simulated with various footprint diameters (D).

481 Table 2. Comparison of DART-Lux and DART-RC single-pulse wLiDAR simulations.

Evaluation indicators		DART-Lux	DART-RC
Time of tracing per ray on average[①]		2.29×10^{-4} s	4.74×10^{-4} s
Average number of effective paths generated per ray		3.19	2.07
Simulation accuracy (R^2 / rRMSE) for various N_{rays} (D = 25 m)	10^2 rays	0.9818 / 21.42%	0.9775 / 28.09%
	10^3 rays	0.9975 / 7.28%	0.9958 / 9.40%
	10^4 rays	0.9997 / 2.48%	0.9990 / 4.79%
	10^5 rays	1.0000 / 0.88%	0.9999 / 1.63%
Number of rays[②] needed to get rRMSE (accuracy) ≤ 5% for various D	D = 0.05 m	200	200
	D = 0.1 m	300	300
	D = 0.5 m	1000	1000
	D = 1 m	2000	3000
	D = 10 m	3000	5000
	D = 20 m	4000	7000
	D = 25 m	5000	10000
	D = 30 m	8000	15000

482 [①] Computing platform: Intel Xeon E5-2687 @ 3.1 GHz processor, 20 threads, 500 GB memory

483 [②] The numbers of rays here are rough estimates.

484

485 ## 3.2. Simulation of multi-pulse drLiDAR

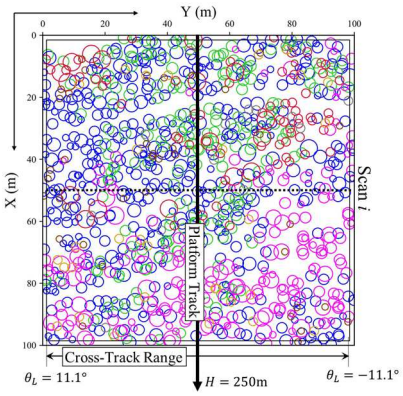
486 Here we illustrate the potential of DART-Lux to simulate point cloud with a drLiDAR
487 instrument that mimics the Riegl VUX-1UAV system (Table 1), and to quantify the errors of
488 airborne laser scanning (ALS) on deriving digital surface models (DSM). The study area is
489 the Jarvselja birch stand as in section 3.1. Fig. 6.a shows the flight configuration: swath

region of 98 m × 98 m, flight direction of +X axis, scanning direction of -Y axis, altitude of 250 m, scanning angle range of [-11.1° , 11.1°], flight speed of 16 km/h, PRF of 100 kHz, look angle step of 0.045°, and along-track step of 0.3 m. With this configuration, 160,392 pulses are simulated, *i.e.*, pulse density is ~16 pulses/m^2.

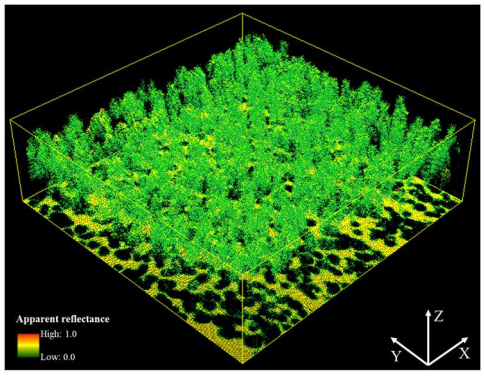
DART-Lux converts the simulated waveform of each pulse into up to 4 points per pulse with a Gaussian decomposition method. Fig. 6.b and Fig. 6.c show oblique and top views of the simulated point cloud, colored by apparent reflectance which is calculated as $\rho_i = \frac{\pi}{\Delta\Omega_i} \cdot \frac{Q_i}{Q_e}$ with Q_i decomposed pulse energy of target i, Q_e emitted pulse energy, and $\Delta\Omega_i$ solid angle of target i towards the receiver. In all, the decomposed point number is 363,202, and per pulse is ~ 2.26 points / pulse, which indicates that pulses can penetrate the canopy and reveal its internal structure. In addition, points associated to trees in the off-nadir view give blind zones in the top view (Fig. 6.c). These missing measurements would lead to inaccurate DSM derivation, so we further study how the laser scanning angle influences the accuracy of the LiDAR-derived DSM. Fig. 6.d shows the relationship of height error (*i.e.*, subtraction of the point cloud-derived DSM with 2-m resolution and the reference DSM directly derived from the 3D coordinates of scene surfaces) and the scanning angle. It demonstrates that the surface height is underestimated, and the underestimated value increases with the scanning angle. If the scanning angle exceeds 10°, the height error exceeds 0.4 m. The average DSM error of this ALS configuration is ~0.1 m.

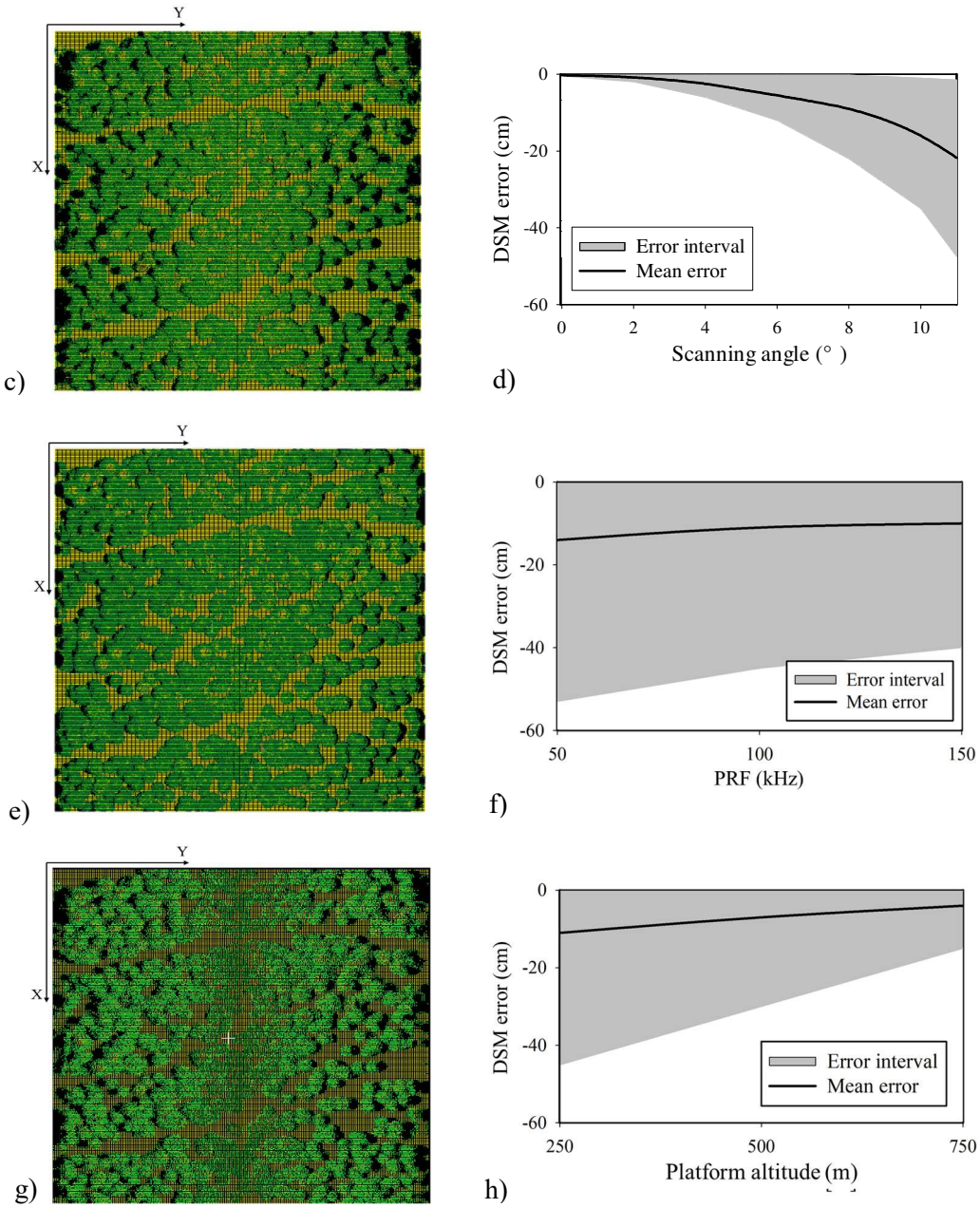
More experiments were conducted to explore the influence of the flight configurations on the DSM errors by considering three platform altitudes (H = 250 m, 500 m, and 750 m) and three PRFs (PRF = 50 kHz, 100 kHz, and 150 kHz, corresponds to pulse density ~8 pts/m^2, ~16 pts/m^2, and ~25 pts/m^2). Fig. 6.e displays the simulated point cloud for H = 500 m and PRF =100 kHz. As compared to Fig. 6.c (H = 250 m, PRF =100 kHz), a higher altitude allows the instrument to measure the same scene with smaller scanning angles, which leads to smaller height error (Fig. 6.f). For the configuration of H = 250 m and PRF =50 kHz, the simulated point cloud (Fig. 6.g) displays the similar blind region distribution but sparser points compared to Fig. 6.c. The curve in Fig. 6.h indicates that the height error has a monotonic reduction relationship with the PRF. Indeed, a larger pulse density increases the probability to detect treetops, and thus improves the DSM accuracy.

a)



b)





523 Fig. 6. Multi-pulse drLiDAR for the Jarvselja birch stand. a) Tree species map (*cf.* Fig. 5.a) and airborne LiDAR
524 configuration. b) Oblique view of simulated point cloud with H = 250 m and PRF = 100 kHz. c) Top view of b).
525 d) Relationship between the scanning angle and the height error of LiDAR-derived DSM. e) Top view of
526 simulated point cloud with H = 500 m and PRF = 100 kHz. f) Relationship between the platform altitude and the
527 DSM error. g) Top view of simulated point cloud with H = 250 m and PRF = 50 kHz. h) Relationship between
528 the PRF and the DSM error.

529

530 3.3. Simulation of wLiDAR with solar signal

531 The wLiDAR solar signal simulation of DART-Lux is compared to that of DART-RC for the
532 Wellington citrus orchard in RAMI-IV experiment. This realistic 108 m × 104 m scene
533 contains 1115 citrus trees (Fig. 7.a). The simulated wLiDAR sensor mimics the GEDI system
534 (Table 1). The sun zenith and azimuth angles are set to θ_{sun}=50° and φ_{sun}=180°,
535 respectively.

536

537 Fig. 7.b shows the DART-Lux and DART-RC waveforms with their solar signal
538 contributions. The solar signal in DART-Lux adds a constant number of photons $N_{\text{SN}} = 1.94$
539 photons/bin, a bit less than $N_{\text{SN}} = 1.96$ photons/bin in DART-RC (*cf.* enlarged view in Fig.
540 7.b). This constant photon number represents the average estimate of received solar photons
541 per bin. The estimated bias between DART-Lux and DART-RC is due to approximations
542 made by DART-RC when calculating the solar signal (Yin et al., 2016). For instance, DART-
543 RC uses a pure pinhole camera to mimic the LiDAR receiver. Also, when converting the solar
544 radiance into the solar photons, DART-RC does not consider the difference of solid angle
545 towards the sensor at different positions of the FOV. In contrast, the LiDAR receiver in
546 DART-Lux is more in line with the real instrument, which gives a more accurate estimate of
547 the solar signal.

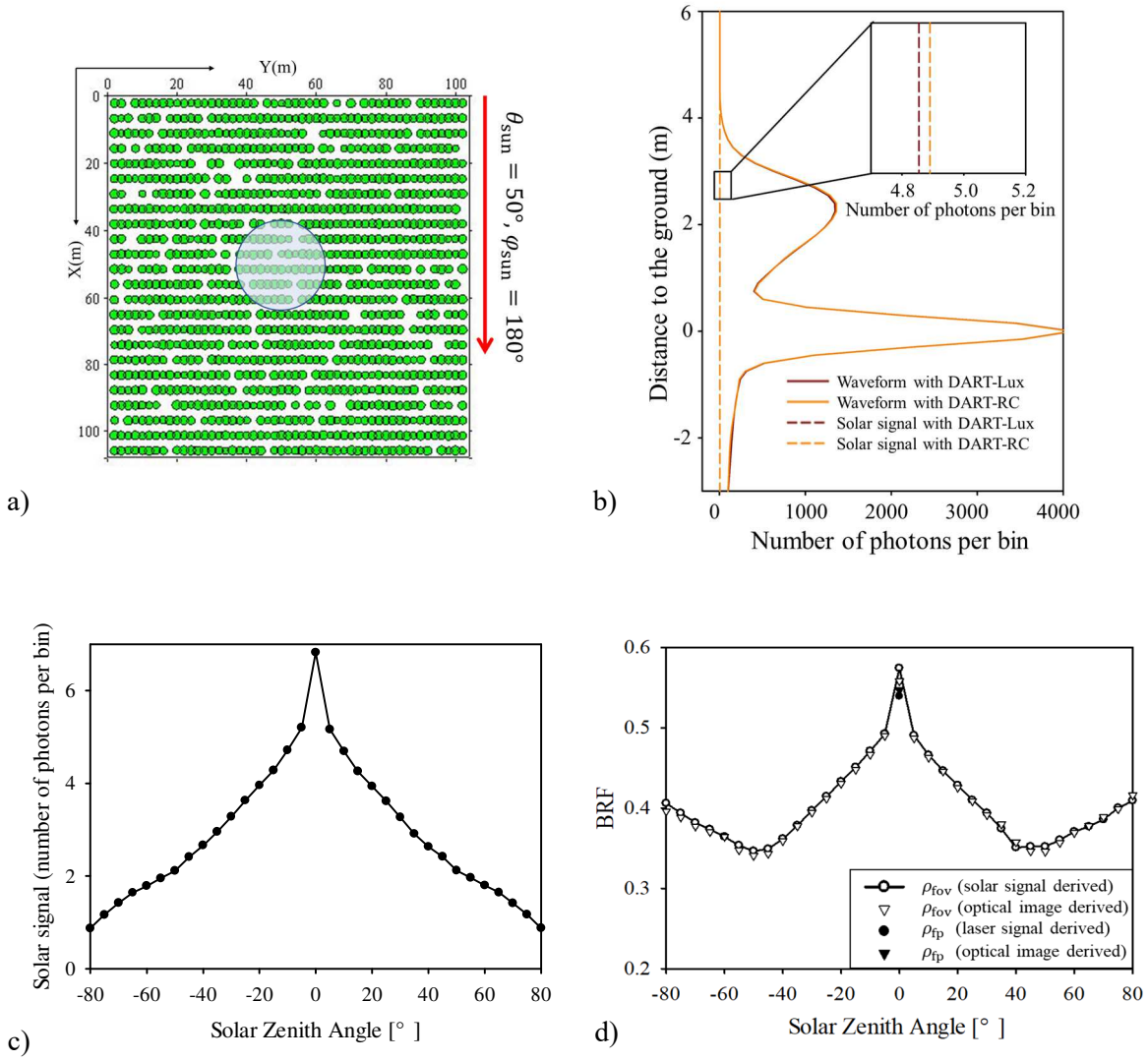
548

549 The flexibility of DART-Lux for modeling solar signal is investigated based on a set of

sensitivity experiments with various solar directions. Fig. 7.c displays the solar photons with the variation of θ_{sun} from -80° to 80° ($\theta_{\mathrm{sun}} < 0$ if $\varphi_{\mathrm{sun}} = 0$). Based on the simulated solar photon (N_{SN}) and the incident sun irradiance (E_{sun}), the bidirectional reflectance factor (BRF) of the whole scene within the FOV can be calculated as:

$$\rho_{\mathrm{fov}}(\theta_i, \varphi_i, \theta_v, \varphi_v) = \pi \cdot \frac{\bar{L}_{\mathrm{fov}}}{E_{\mathrm{sun}} \cdot \cos\theta_{\mathrm{sun}}} = \pi \cdot \frac{1}{E_{\mathrm{sun}} \cdot \cos\theta_{\mathrm{sun}}} \cdot \frac{N_{\mathrm{SN}} \cdot hv \cdot f}{\Delta\lambda \cdot A_{\mathrm{fov}} \cdot \Delta\Omega} \tag{20}$$

where $\bar{L}_{\mathrm{fov}}$ is the average exitant solar radiance within the FOV, $\Delta\lambda$ is the band width of LiDAR receiver, A_{fov} is the area of FOV, $\Delta\Omega$ is the solid angle of scene towards the receiver, (θ_i, φ_i) and (θ_v, φ_v) are the direct sun illumination direction and the LiDAR viewing direction respectively. Also, the BRF of the scene within the footprint can be calculated by total LiDAR received energy (Q_r) and emitted energy (Q_e) with the equation $\rho_{\mathrm{fp}}(0,0,0,0) = \frac{\pi}{\Delta\Omega} \cdot \frac{Q_r}{Q_e}$. In order to validate the BRFs derived from laser and solar signals of DART-Lux LiDAR mode, we simulate the BRF images with the DART-Lux image mode (Wang et al., 2022) with the same directions, and calculate the average BRFs of the FOV and footprint regions. The comparison results as in Fig. 7.d show that the BRFs derived from LiDAR (laser or solar) signals and optical images are almost consistent, demonstrating that the received solar or laser energy has potentials to reveal the scene reflectance. Additionally, the variations of solar signal and BRF in Fig. 7.c and Fig. 7.d are nearly symmetric due to the azimuthal symmetry of the scene, with maximal values when the sun and LiDAR directions are identical (θ_{sun}=0° at nadir), partly due to the hot spot effect (Kuusk, 1991). The difference of variations of solar signal and BRF stress that the solar signal received by LiDAR sensor is influenced by the coupling of scene reflectance, sun direct irradiance, and sun direction.



570 Fig. 7. wLiDAR with solar signal in Wellington Citrus Orchard. a) Citrus distribution map with LiDAR footprint
571 (D = 25 m) centered at (50 m, 50 m). b) DART-Lux and DART-RC waveforms with solar signal. c) Simulated
572 solar photons for various solar zenith angles. d) Scene BRFs within the FOV / footprint derived from simulated
573 solar / laser signals and simulated optical images.

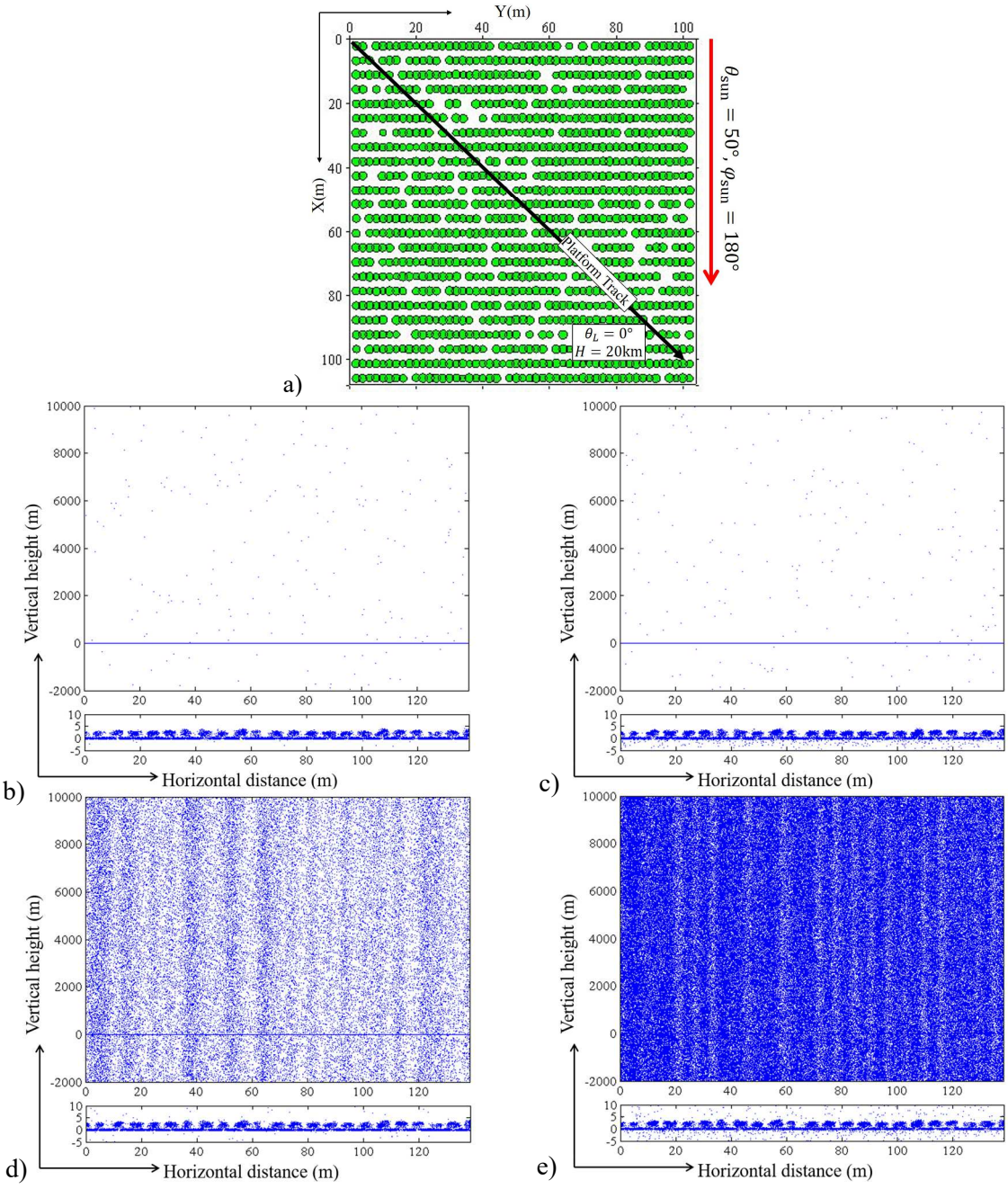
574

575 ## 3.4. Simulation of pcLiDAR with solar signal

576 The potential of DART-Lux to simulate pcLiDAR with solar signal is investigated with the
577 Wellington citrus orchard and same solar direction (θ_{sun}=50°, φ_{sun}=180°) as in section 3.3.

578 The pcLiDAR device mimics the MABEL system (Table 1) (McGill et al., 2013). The black
579 arrow in Fig. 8.a shows the flight track from (3 m, 3 m) to (101 m, 101 m). With a 0.02-m
580 along-track step, 6929 pulses are launched.
581
582 Fig. 8.b-e show the DART-Lux-simulated photon counts without and with solar signal at 532
583 nm and 1064 nm, respectively. Each figure is for two ranges: [-2000 m, 10000 m] (top) and [-
584 5 m, 10 m] (bottom). The lower sub-figures are intended to stress the LiDAR signal from the
585 scene. Without the sunlight (Fig. 8.b and c), the photons mostly come from the scene and
586 partly from the instrumental dark counts. The lower subfigures of Fig. 8.b and c indicate that
587 the density of photons from the scene is larger at 1064 nm (averagely 1.58 photons per shot)
588 than at 532 nm (averagely 1.68 photons per shot), which is due to the stronger reflectance of
589 vegetation and background at 1064 nm. With the sunlight (Fig. 8.d and e), the solar power
590 significantly increases the photon density. Again, due to the stronger reflectance of vegetation
591 and background at 1064 nm, the solar photon frequency is greatly larger at 1064 nm
592 (averagely 4.93×10^{-5} photons per bin per shot) than at 532 nm (averagely 3.23×10^{-3} photons
593 per bin per shot). Fig. 8.f and Fig. 8.g plot the number of solar photons per bin (N_{SN}) in
594 DART-Lux and DART-RC versus the pulse index at two wavelengths. They stress that the
595 solar signal fluctuation curves simulated by DART-Lux and DART-RC are almost consistent.
596 Additionally, the solar photon frequency fluctuates as the sensor moves, which is coherent
597 with the scene reflectance heterogeneity. However, the number of peaks of solar photon
598 frequency in Fig. 8.f and Fig. 8.g is less than the number of citrus rows. It can be explained

that the solar signal is influenced by the composite reflectance of light crown, shadow crown, light ground, and shadow ground due to the large footprint size (D = 2 m).



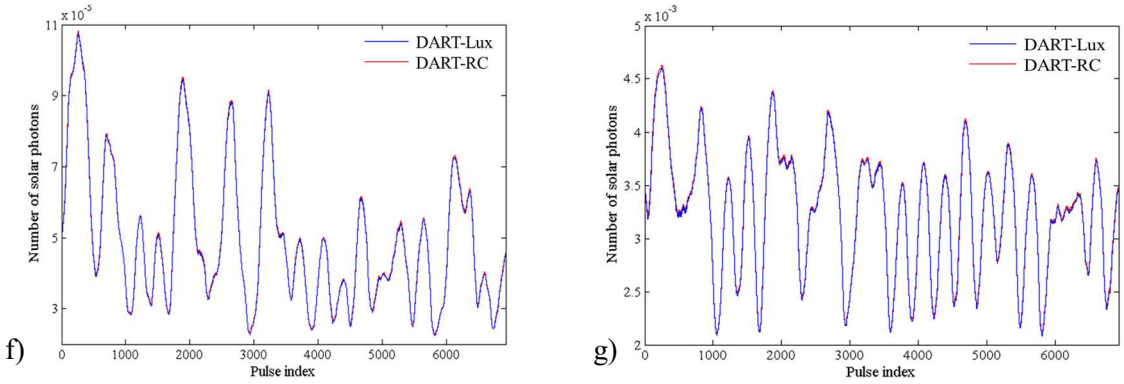


Fig. 8. pcLiDAR with solar signal in Wellington Citrus Orchard. Left: λ=532 nm. Right: λ=1064 nm. a) Citrus distribution map with sensor and sun configurations. b, c) DART-Lux pcLiDAR signal without solar signal for 2 vertical ranges (top: [-2000 m, 10000 m]; bottom: [-5 m, 10 m]). d, e) DART-Lux pcLiDAR signal with solar signal for the same ranges. f, g) Solar photons per bin simulated with DART-Lux and DART-RC against pulse index.

4. Discussion

4.1. Advantages of DART-Lux for simulating LiDAR signals

This study assessed the performance of the DART-Lux LiDAR modeling using the benchmark model DART-RC. As expected, the two independently designed models in DART converge to the same results. This not only proves the accurateness and unbiasedness of DART-Lux but also DART-RC. Additionally, DART-Lux offers significant advantages in the efficiency of simulation of LiDAR signals especially for complex scenes with billions of facets. Table 3 summarizes the processing times and memory usages of the four case studies. It shows that DART-Lux takes around half time and hundredfold less memory of DART-RC for simulating LiDAR signals. The time difference is even tenfold shorter if the solar signal is

simulated. These improvements are mostly explained by:

- BDPT algorithm of DART-Lux: it can sample more paths from source to receiver than the forward MC method of DART-RC (Table 2).

- Difference of solar signal modeling: DART-Lux simultaneously simulates the solar and laser signals reaching the PMT, whereas DART-RC simulates them in turn and then couples them together. Also, DART-RC simulates the solar signal with the DOM method, which is proven much less efficient for simulating the signal of passive sensors (LiDAR is regarded as a passive sensor when modeling solar signal) compared to DART-Lux (Wang and Gastellu-Etchegorry, 2021).

- Geometry instance technique: For the landscapes where some elements (*e.g.*, tree, maize plant) can be simulated as clones of a limited number of 3D objects, DART-Lux stores an instance of each object and clones it by the transformation matrix with scaling, rotation, and pan information. In contrast, DART-RC stores each 3D object in the memory. The order of magnitude of the gain in memory usage depends on landscapes and can be greatly larger than 100.

In addition to the above efficiency improvements, the *LuxCoreRender* used by DART-Lux supports the computing of ray-tracing on any number of CPUs and/or GPUs. It stresses that DART-Lux has potentials to improve the simulated times by an order of magnitude again. DART team is currently developing the GPU acceleration capability of DART-Lux, to

facilitate RS modeling of large-area complex scene by incorporating high-performance computing technology.

Table 3. Comparison of DART-Lux and DART-RC processing time and memory usage for the four case studies.

Resources		DART-Lux		DART-RC	
		Memory usage	Time	Memory usage	Time
wLiDAR①	rRMSE ⩽ 5%	1.4 GB	1.15 s	238 GB	4.74 s
	rRMSE ⩽ 2%	1.4 GB	9.20 s	238 GB	40.3 s
drLiDAR①	PPR = 50 kHz	2.4 GB	0.84 h	253 GB	1.16 h
	PPR = 100 kHz	2.4 GB	1.69 h	253 GB	2.30 h
	PPR = 150 kHz	2.4 GB	2.53 h	253 GB	3.46 h
wLiDAR with solar signal②	laser signal	3.2 GB	1 min	60 GB	2 min
	laser and solar signals	3.2 GB	2 min	60 GB	11 h
pcLiDAR with solar signal②	laser signal	4.4 GB	0.9 h	62 GB	1.2 h
	laser and solar signals	4.4 GB	1.8 h	62 GB	12.4 h

① Computing platform: Intel Xeon E5-2687 @3.1 GHz processor, 20 threads, 500 GB memory

② Computing platform: Inter Xeon E5-1620 @3.5 GHz processor, 8 thread, 64 GB memory

4.2. Comparison of bidirectional, forward, and backward path tracings for LiDAR signal modeling

By limiting the maximum number of vertices of receiver or light sub-paths as 0, the BDPT mode of DART-Lux can be degraded into the forward or backward path tracing modes. We

compare the efficiency (*i.e.*, processing time) of the three modes for modelling the laser and solar signals based on the case studies of the Jarvselija birch stand and the flat land. The results in Table 4 indicate that the best mode is influenced by various factors. If only simulating the laser signal of which the laser illumination region is smaller than the FOV region, the backward mode is always worst due to the redundant computations of rays launched outside the footprint. If the received signal is mainly from 1st-order scattering (*e.g.*, flat land), the forward mode is the best since the BDPT still traces many invalid rays outside the footprint. In contrast, if multiple scattering occurs frequently (*e.g.*, Jarvselja birch stand), BDPT has advantages over the forward and backward modes thanks to effective path generation. Note that the multiple scattering of laser pulse is influenced by many factors, *e.g.*, scene properties, laser wavelength, and footprint size (Gastellu-Etchegorry et al., 2016). The results are opposite if only simulating the solar signal of which the solar illumination region (the whole scene) is larger than the LiDAR FOV region. The forward mode converges slowest, confirming the previous study that the forward mode usually performs worse than the backward mode on the signal simulation of passive sensors (Qi et al., 2019). However, the BDPT mode is more suitable for simulating the solar signal than the backward mode when the rays are frequently scattered within the FOV. When simulating the LiDAR signal with solar signal, the optimal mode is influenced by multiple factors, including the areas of sun and laser illuminations and LiDAR observation, the contribution of high-order scatterings, and the contribution of solar signal. Although the BDPT mode cannot be optimal in all cases, its effective path generation algorithm and MIS weight mechanism make it ease the difficulty of

high-order signal modelling, especially over large-scale complicated scenarios.

Table 4. Computational time of forward, backward, and bidirectional path tracing modes for modeling the laser signal and solar signal on Jarvselja birch stand and flat land.

Computational time①		Forward mode	Backward mode	BDPT mode
Laser signal	Jarvselja birch stand	1.73 s	5.06 s	1.25 s
	flat land	0.10 s	0.25 s	0.17 s
Solar signal	Jarvselja birch stand	1.80 s	1.20 s	1.07 s
	flat land	0.63 s	0.10 s	0.37 s
LiDAR signal with solar signal	Jarvselja birch stand	3.45 s	9.82 s	2.42 s
	flat land	0.39 s	0.48 s	0.40 s

①Hough time to get rRMSE≤5% (LiDAR footprint D=30 m, FOV D_{fov}=45 m, laser wavelength λ=1064 nm).

5. Conclusion

This study developed a new computationally efficient LiDAR modeling method based on the recent DART-Lux mode of DART. This method mimics the illumination and receiving modes of LiDAR sensors, adapts the BDPT method into the LiDAR time and power measurements, introduces direct light sampling strategy to solve the modeling of solar signal in LiDAR signal, and supports the output of multi-platform multi-type LiDAR signals. The accuracy and performance of DART-Lux were assessed based on four case studies with different LiDAR configurations (wLiDAR, drLiDAR, wLiDAR with solar signal, and pcLiDAR with solar signal) on two realistic scenes (Jarvselja birch forest and Wellington Citrus Orchard), using DART-RC as a reference model. The simulation of single pulse waveform with many rays

686 stresses the consistency of DART-Lux and DART-RC: R^2 = 1 and rRMSE = 0.21%. The
687 faster convergence of DART-Lux stresses its better efficiency and accuracy: for a same
688 number of rays or a same computation time, it converges faster than DART-RC. In terms of
689 the case studies, the memory usage of DART-Lux is hundredfold less than DART-RC, and its
690 processing time is about half that of DART-RC, and tenfold less if the solar signal is
691 simulated. In the next work, the time efficiency might be improved again once DART-Lux
692 mode is developed to support GPU acceleration.
693 Additionally, a series of DART-Lux sensitivity studies were done to assess the adaptability of
694 DART-Lux to the cases of actual LiDAR configurations with various footprint sizes, scanning
695 angles, pulse densities, laser wavelengths, with or without sunlight. Their results illustrate the
696 potential of DART-Lux to provide important information for analysis and exploitation of
697 LiDAR data, such as i) the impact of scanning angle, flight altitude, and pulse density on the
698 ALS-derived DSM error; ii) the impact of sunlight on LiDAR signal, depending on the
699 landscape reflectance, laser wavelength, and directions of sun irradiation and sensor
700 observation, *etc*.
701
702 In summary, this new DART-Lux model opens promising perspectives for LiDAR
703 applications over large-area landscapes with the advantages of high accuracy, low memory
704 cost, and fast 3D modeling. Apart from the LiDAR modeling, DART-Lux is on-going
705 improved in order to become a unified and comprehensive RS 3D RT model, including
706 satellite, airborne and in-situ spectroradiometer and thermal images, bidirectional reflectance

factor, LiDAR, solar-induced chlorophyll fluorescence, and 3D radiative budget of urban and natural landscapes (Regaieg et al, 2022; Wang and Gastellu-Etchegorry, 2021; Wang et al, 2022). Currently, the DART-Lux LiDAR module is already integrated into the official DART version freely available to scientists (https://dart.omp.eu). Validation of DART-Lux with actual LiDAR measurements (*i.e.*, GEDI and ICESat-2) and atmosphere radiative transfer in DART-Lux LiDAR model are presented in a companion paper (Yang et al., 2021).

Acknowledgement

This work is supported by the National Key Research and Development Program of China (No. 2021YFF0704600), the TOSCA program of the Centre National d'Etudes Spatiales, the National Natural Science Foundation of China (No. 41871264), the Guangxi Natural Science Fund for Innovation Research Team (2019JJF50001), the Chinese Scholarship Council (No. 201904910846), and CNRS (Centre national de la recherche scientifique) through the 80|Prime program. The authors are grateful to two computer scientists of DART team (Eric Chavanon and Jordan Guilleux).

Appendix A. Nomenclature

A.1 Acronyms

ALS	Airborne laser scanning
BDPT	Bidirectional path tracing
BRF	Bidirectional reflectance factor

BSDF	Bidirectional scattering distribution function
CDF	Cumulative density function
DART	Discrete Anisotropic Radiative Transfer
DART-FT	DART model on Flux-Tracking mode
DART-RC	DART model on Ray-Carlo mode
DART-Lux	DART model on bidirectional path tracing mode
DOM	Discrete ordinates method
DSM	Digital surface model
FOV	Field of view
GEDI	Global Ecosystem Dynamics Investigation
ICESat	Ice, Cloud, and land Elevation Satellite
LiDAR	Light Detection And Ranging
wLiDAR	Waveform LiDAR
drLiDAR	Discrete return LiDAR
pcLiDAR	Photon counting LiDAR
MC	Monte Carlo
MIS	Multiple importance sampling
PDF	Probability density function
PMT	Photomultiplier tube
PRF	Pulse repetition frequency
RMSE	Root mean square error
RS	Remote sensing
RT	Radiative transfer
UAV	Unmanned aerial vehicle

A.2 Symbols

A	Area of the scene surface
A_{disk}	Area of virtual sunlight disk
A_{senor}	Area of virtual sensor plane
$A_{\text{sensor}}^{\text{sub}}$	Subarea on virtual sensor plane. It receives rays through a point on the receiving aperture
A_{tel}	Area of the receiving aperture of LiDAR telescope
A_{fov}	Area of LiDAR FOV
c	Light speed
D	Footprint diameter
$\mathcal{D}$	Set of all paths
$\mathcal{D}_n$	Set of paths with number of edges *n*
E	Irradiance, unit: W/m^2 or W/m^2/μm
E_{p_1}	Incident irradiance at vertex p_1
E_s	Solar irradiance
F	Distance of virtual sensor plane to LiDAR receiver
f	Signal sampling frequency
$f(p_k, -\Omega_i, \Omega_o)$, $f(p_{k-1} \to p_k \to p_{k+1})$	BSDF at vertex p_k with incident direction Ω_i ($p_{k-1} \to p_k$) and exitant direction Ω_o ($p_k \to p_{k+1}$)
$f(\bar{r}_{s,t})$	Power contribution in function of path $\bar{r}_{s,t}$
$G(p_{k-1} \leftrightarrow p_k)$	Connection function between vertices p_{k-1} and p_k
h	Plank's constant
L	Radiance, especially L_e is the emitted radiance of light source
$L(p_k, \Omega_o), L(p_k \to p_{k+1})$	Radiance from the vertex p_k to the direction Ω_o or the vertex p_{k+1}

$l(\bar{r}_{s,t})$	Length of path $\bar{r}_{s,t}$
$N_{\text{lidar}}(\xi)$	Number of laser photon in a function of time bin ξ
N_{SN}	Number of solar photons per time bin
N_s, N_t	Maximal number of vertices in light and receiver sub-paths.
n	Number of edges of path, also called path depth, $n \in \mathbb{Z}^+$
$P_L(x, y, H)$	LiDAR position
P_{light}	Probability to sample a source from the light group (laser + sunlight)
$p(p_k^{\rightarrow}), p(q_k^{\leftarrow})$	PDFs of sampling vertex p_k and q_k in forward and backward paths
$p(p_0, \Omega), p(q_0, \Omega)$	Directional PDFs of ray starting from p_0 and q_0 with direction Ω
$p(\bar{r}_{s,t})$	PDF of path $\bar{r}_{s,t}$
p_{s-1}	*s*-th vertex on the light sub-path (p_0 is located on the light source)
$\bar{p}_s$	Light sub-path with *s* vertices
$p(\tau), p(\xi)$	Distribution of emitted laser energy, as a function of time τ or bin ξ
Q_e	Total laser emitted energy, unit: J
$Q_{\text{lidar}}(\xi)$	The received laser energy in the function of bin ξ
q_{sensor}	Vertex located on the virtual sensor plane
q_{t-1}	*t*-th vertex on the receiver sub-path (q_0 is located on the receiver)
$\bar{q}_t$	Receiver sub-path with *t* vertices
R	Radius of virtual scene sphere
$\bar{R}$	A random path sample
r'', r', r, r_k	Vertex on the path
$\bar{r}, \bar{r}_n$	A path from source to receiver with number of edges (path depth) *n*
$\bar{r}_{s,t}$	A path from source to receiver generated by connecting the light sub-path $\bar{p}_s$ and the receiver sub-path $\bar{q}_t$, satisfying *s*+*t*=*n*-1

r_{senor}	Radius of virtual sensor plane
$r_{\text{sensor}}^{\text{sub}}$	Radius of subarea on virtual sensor plane
r_{tel}	Radius of LiDAR telescope
r_{fp}	Radius of LiDAR footprint
W	Importance function, especially W_e is the emitted importance
$w(\bar{r}_{s,t})$	MIS weight of path $\bar{r}_{s,t}$
δ	Dirac delta function
θ, φ	Zenith and azimuth angles
θ_i, $\theta_i^{p_k}$	Incident angle (at vertex p_k) between incident direction Ω_i and surface normal vector
θ_{fp}, θ_{FOV}	Half dispersion angles of LiDAR footprint and FOV
η	Fraction of laser energy at the footprint edge to the center
λ	Laser wavelength
$\Delta\lambda$	Band width of LiDAR receiver
μ	Area-product measure of path
ξ	Time bin in LiDAR waveform sequence, $\xi \in \mathbb{N}$
$\Delta\xi$	Duration bins of emitted laser pulse
σ	Standard deviation of 2D Gauss distribution of emitted pulse energy
$\tau(\bar{r}_{s,t})$	Time delay of path $\bar{r}_{s,t}$
υ	Photon frequency
Ω	Direction vector
Ω_i, Ω_o	Incident direction and exitant direction
Ω_s	Sunlight direction
$\vec{\Omega}_L(\theta_L, \varphi_L)$	LiDAR orientation

$\Delta\Omega_{\mathrm{fp}}$	Conical solid angle of LiDAR footprint
$\Phi_e(p_0,\tau)$	Emitted laser power at the time τ.
Φ_{lidar}	Total laser receiving power.
$\Phi_{\mathrm{lidar}}(\tau)$, $\Phi_{\mathrm{lidar}}(\xi)$	Laser power profile as a function of time delay τ or time bin ξ
$\widetilde{\Phi}_{\mathrm{lidar}}(\tau)$	A MIS Monte Carlo estimator for $\Phi_{\mathrm{lidar}}(\tau)$

727

728 **Appendix B. Directional PDF of rays from laser source**

729 The hit of a nadir laser pulse onto a flat plane with a unit distance from the laser source, gives
730 a circular footprint with radius $r_{\mathrm{fp}}=\tan\theta_{\mathrm{fp}}$. The fraction of laser energy at the footprint edge
731 to the center η satisfies the equality $\eta=\exp\left(-\frac{r_{\mathrm{fp}}^2}{2\sigma^2}\right)$ with σ the standard deviation of the 2D
732 Gaussian distribution of emitted energy. For any position (x,y) within the footprint, the area
733 PDF of exitant ray is:

$$p(x,y)=\frac{\frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)}{\iint_{x^2+y^2\le r_{\mathrm{fp}}^2}\frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)dxdy}=\frac{\frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)}{1-\eta},\ \text{where } x^2+y^2\le r_{\mathrm{fp}} \qquad \text{(B. 1)}$$

734 The Cartesian coordinate (x,y) can be transformed to the polar coordinate (r,φ) using
735 $x=r\cos\varphi$, $y=r\sin\varphi$ with r the distance to footprint center and φ the azimuth angle:

$$p(r,\varphi)=\frac{\frac{r}{2\pi\sigma^2}\exp\left(-\frac{r^2}{2\sigma^2}\right)}{1-\eta},\ \text{where } r\le r_{\mathrm{fp}} \qquad \text{(B. 2)}$$

736 The directional PDF of exitant ray from light source vertex p_0 is computed from the area PDF
737 by the relationship between unit solid angle and unit surface area (*i.e.*, $d\Omega=dA\cdot\frac{\left(\frac{1}{\cos\theta}\right)^2}{\cos\theta}$):

$$p(p_0,\Omega)=p(r,\varphi)\cdot\frac{\left(\frac{1}{\cos\theta}\right)^2}{r\cos\theta}=\frac{\exp\left(-\frac{\tan^2\theta}{2\sigma^2}\right)}{2\pi\sigma^2\cdot\cos^3\theta\cdot(1-\eta)},\ \text{where } \theta\le\theta_{\mathrm{fp}} \qquad \text{(B. 3)}$$

738 The random laser ray with directional PDF $p(p_0,\Omega)$ can be generated by sampling the zenith

angle θ and azimuth angle φ independently. The cumulative density functions (CDF) of θ and φ are:

$$P(\theta) = \int_0^{\theta} \frac{\sin\theta \cdot \exp\left(-\frac{\tan^2\theta}{2\sigma^2}\right)}{\sigma^2 \cdot \cos^3\theta \cdot (1-\eta)} d\theta = \frac{1-\exp\left(-\frac{\tan^2\theta}{2\sigma^2}\right)}{1-\eta}$$
$$P(\varphi) = \int_0^{\varphi} \frac{1}{2\pi} d\varphi = \frac{1}{2\pi}\varphi \quad \text{(B. 4)}$$

Based on the CDFs, the exitant direction of laser ray can be determined by inverting two random numbers $U_1, U_2 \in [0,1]$, *i.e.*, $\theta = P^{-1}(U_1)$, $\varphi = P^{-1}(U_2)$.

References

Blair, J.B., Hofton, M.A., 1999. Modeling laser altimeter return waveforms over complex vegetation using high-resolution elevation data. Geophys. Res. Lett. 26, 2509–2512.

Brown, S.D., Blevins, D.D., Schott, J.R., 2005. Time-gated topographic LIDAR scene simulation, in: Laser Radar Technology and Applications X. International Society for Optics and Photonics, pp. 342–353.

Chauve, A., Mallet, C., Bretar, F., Durrieu, S., Deseilligny, M.P., Puech, W., 2007. Processing full-waveform lidar data: modelling raw signals. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 2007, pp. 102-107.

Dassot, M., Constant, T., Fournier, M., 2011. The use of terrestrial LiDAR technology in forest science: application fields, benefits and challenges. Ann. For. Sci. 68, 959–974.

Disney, M.I., Lewis, P.E., Bouvet, M., Prieto-Blanco, A., Hancock, S., 2009. Quantifying surface reflectivity for spaceborne lidar via two independent methods. IEEE Trans. Geosci. Remote Sens. 47, 3262–3271.

Dubayah, R., Blair, J.B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner, J., Luthcke, S., 2020. The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography. Sci. Remote Sens. 1, 100002.

Durrieu, S., Cherchali, S., Costeraste, J., Mondin, L., Debise, H., Chazette, P., Dauzat, J., Gastellu-Etchegorry, J.-P., Baghdadi, N., Pélissier, R., 2013. Preliminary studies for a vegetation ladar/lidar space mission in france, in: 2013 IEEE International Geoscience and Remote Sensing Symposium-IGARSS. IEEE, pp. 4332–4335.

Dwivedi, M., Uniyal, A., Mohan, R., 2014. New Horizons in Planning Smart Cities using LiDAR Technology. Int. J. Appl. Remote Sens. GIS 2, 40–50.

Gastellu-Etchegorry, J.-P., Yin, T., Lauret, N., Cajgfinger, T., Gregoire, T., Grau, E., Feret, J.-B., Lopes, M., Guilleux, J., Dedieu, G., 2015. Discrete anisotropic radiative transfer (DART 5) for modeling airborne and satellite spectroradiometer and LIDAR acquisitions of natural and urban landscapes. Remote Sens. 7, 1667–1701.

Gastellu-Etchegorry, J.-P., Yin, T., Lauret, N., Grau, E., Rubio, J., Cook, B.D., Morton, D.C., Sun, G., 2016. Simulation of satellite, airborne and terrestrial LiDAR with DART (I): Waveform simulation with quasi-Monte Carlo ray tracing. Remote Sens. Environ. 184, 418–435.

Gastellu-Etchegorry, J.P., Martin, E., Gascon, F., 2004. DART: a 3D model for simulating satellite images and studying surface radiation budget. Int. J. Remote Sens. 25, 73–96.

Goodenough, A.A., Brown, S.D., 2017. DIRSIG5: next-generation remote sensing data and image simulation framework. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10, 4818–4833.

Govaerts, Y.M., Verstraete, M.M., 1998. Raytran: a Monte Carlo ray-tracing model to compute light scattering

in three-dimensional heterogeneous media. IEEE Trans. Geosci. Remote Sens. 36, 0–505.

Grau, E., Durrieu, S., Fournier, R., Gastellu-Etchegorry, J.-P., Yin, T., 2017. Estimation of 3D vegetation density with Terrestrial Laser Scanning data using voxels. A sensitivity analysis of influencing parameters. Remote Sens. Environ. 191, 373–388.

Hmida, S. Ben, Kallel, A., Gastellu-Etchegorry, J.-P., Roujean, J.-L., 2017. Crop biophysical properties estimation based on LiDAR full-waveform inversion using the DART RTM. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10, 4853–4868.

Huang, H., Wynne, R.H., 2013. Simulation of lidar waveforms with a time-dependent radiosity algorithm. Can. J. Remote Sens. 39, S126–S138.

Iraci, B., 2013. Blender cycles: lighting and rendering cookbook. Packt Publishing Ltd.

Jaboyedoff, M., Oppikofer, T., Abellán, A., Derron, M.-H., Loye, A., Metzger, R., Pedrazzini, A., 2012. Use of LiDAR in landslide investigations: a review. Nat. hazards 61, 5–28.

Kajiya, J.T., 1986. The rendering equation, in: Proceedings of the 13th Annual Conference on Computer Graphics and Interactive Techniques. pp. 143–150.

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.

Kokkalis, P., 2017. Using paraxial approximation to describe the optical setup of a typical EARLINET lidar system. Atmos. Meas. Tech. 10, 3103.

Kotchenova, S.Y., Shabanov, N. V, Knyazikhin, Y., Davis, A.B., Dubayah, R., Myneni, R.B., 2003. Modeling Lidar waveforms with time-dependent stochastic radiative transfer theory for remote estimations of forest

structure. J. Geophys. Res. Atmos. 108.

Kuusk, A., 1991. The hot spot effect in plant canopy reflectance, in: Photon-Vegetation Interactions. Springer, pp. 139–159.

Lamelas-Gracia, M.T., Riaño, D., Ustin, S., 2019. A LiDAR signature library simulated from 3-dimensional Discrete Anisotropic Radiative Transfer (DART) model to classify fuel types using spectral matching algorithms. GIScience Remote Sens. 56, 988–1023.

Markus, T., Neumann, T., Martino, A., Abdalati, W., Brunt, K., Csatho, B., Farrell, S., Fricker, H., Gardner, A., Harding, D., 2017. The Ice, Cloud, and land Elevation Satellite-2 (ICESat-2): science requirements, concept, and implementation. Remote Sens. Environ. 190, 260–273.

McGill, M., Markus, T., Scott, V.S., Neumann, T., 2013. The multiple altimeter beam experimental Lidar (MABEL): An airborne simulator for the ICESat-2 mission. J. Atmos. Ocean. Technol. 30, 345–352.

Ni-Meister, W., Jupp, D.L.B., Dubayah, R., 2001. Modeling lidar waveforms in heterogeneous and discrete canopies. IEEE Trans. Geosci. Remote Sens. 39, 1943–1958.

Nimier-David, M., Vicini, D., Zeltner, T., Jakob, W., 2019. Mitsuba 2: A retargetable forward and inverse renderer. ACM Trans. Graph. 38, 1–17.

North, P.R.J., Rosette, J.A.B., Suárez, J.C., Los, S.O., 2010. A Monte Carlo radiative transfer model of satellite waveform LiDAR. Int. J. Remote Sens. 31, 1343–1358.

Pharr, M., Jakob, W., Humphreys, G., 2016. Physically based rendering: From theory to implementation. Morgan Kaufmann.

Qi, J., Xie, D., Yin, T., Yan, G., Gastellu-Etchegorry, J.-P., Li, L., Zhang, W., Mu, X., Norford, L.K., 2019. LESS: LargE-Scale remote sensing data and image simulation framework over heterogeneous 3D scenes.

820 Remote Sens. Environ. 221, 695–706.

821 Qin, H., Wang, C., Xi, X., Tian, J., Zhou, G., 2017. Simulating the effects of the airborne lidar scanning angle,

822 flying altitude, and pulse density for forest foliage profile retrieval. Appl. Sci. 7, 712.

823 Regaieg, O., Lauret, N., Wang, Y., Guilleux, J., Chavanon, E., Gastellu-Etchegorry, J.P., 2021. DART Monte-

824 Carlo based modeling of chlorophyll solar-induced fluorescence images of 3D canopies. Remote Sens.

825 Environ. (in review).

826 Roberts, O., Bunting, P., Hardy, A., McInerney, D., 2020. Sensitivity Analysis of the DART Model for Forest

827 Mensuration with Airborne Laser Scanning. Remote Sens. 12, 247.

828 Shirley, P., Wang, C., Zimmerman, K., 1996. Monte Carlo techniques for direct lighting calculations. ACM

829 Trans. Graph. 15, 1–36.

830 Sun, G., Ranson, K.J., 2000. Modeling lidar returns from forest canopies. IEEE Trans. Geosci. Remote Sens. 38,

831 2617–2626.

832 Veach, E., 1997. Robust Monte Carlo methods for light transport simulation. Stanford University PhD thesis.

833 Wallace, L., Lucieer, A., Watson, C., Turner, D., 2012. Development of a UAV-LiDAR system with application

834 to forest inventory. Remote Sens. 4, 1519–1543.

835 Wang, C., Yang, X., Xi, X., Zhang, H., Chen, S., Peng, S., Zhu, X., 2020. Evaluation of Footprint Horizontal

836 Geolocation Accuracy of Spaceborne Full-waveform LiDAR based on Digital Surface Model. IEEE J. Sel.

837 Top. Appl. Earth Obs. Remote Sens. 13, 2135-2144.

838 Wang, R., Peethambaran, J., Chen, D., 2018. LiDAR point clouds to 3-D Urban Models: A review. IEEE J. Sel.

839 Top. Appl. Earth Obs. Remote Sens. 11, 606–627.

840 Wang, Y., Kallel, A., Yang, X., Regaieg, O., Lauret, N., Gastellu-Etchegorry, J.P., 2022. DART-Lux: An

841 unbiased and rapid Mont-Carlo mode in DART for simulating remote sensing data and images. Remote
842 Sens. Environ. (in review)
843 Wang, Y., Gastellu-Etchegorry, J.P., 2021. Accurate and fast simulation of remote sensing images at top of
844 atmosphere with DART-Lux. Remote Sens. Environ. 256: 112311.
845 Wei, S., Yin, T., Dissegna, M.A., Whittle, A.J., Ow, G.L.F., Yusof, M.L.M., Lauret, N., Gastellu-Etchegorry, J.-
846 P., 2020. An assessment study of three indirect methods for estimating leaf area density and leaf area index
847 of individual trees. Agric. For. Meteorol. 292, 108101.
848 Weitkamp, C., 2006. Lidar: range-resolved optical remote sensing of the atmosphere. Springer Science &
849 Business.
850 Widlowski, J.-L., Mio, C., Disney, M., Adams, J., Andredakis, I., Atzberger, C., Brennan, J., Busetto, L., Chelle,
851 M., Ceccherini, G., 2015. The fourth phase of the radiative transfer model intercomparison (RAMI)
852 exercise: Actual canopy scenarios and conformity testing. Remote Sens. Environ. 169, 418–437.
853 Widlowski, J.L., Lavergne, T., Pinty, B., Verstraete, M., Gobron, N., 2006. Rayspread: A Virtual Laboratory for
854 Rapid BRF Simulations Over 3-D Plant Canopies. Lect. Notes Comput. Sci. Eng. 48, 211–231.
855 Williams, K., Olsen, M.J., Roe, G. V, Glennie, C., 2013. Synthesis of transportation applications of mobile
856 LiDAR. Remote Sens. 5, 4652–4692.
857 Wulder, M.A., White, J.C., Nelson, R.F., Næsset, E., Ørka, H.O., Coops, N.C., Hilker, T., Bater, C.W.,
858 Gobakken, T., 2012. Lidar sampling for large-area forest characterization: A review. Remote Sens.
859 Environ. 121, 196–209.
860 Yang, X., Wang, C., Yin, T., Wang, Y., Lauret, N., Xi, X., Li, D., Gastellu-Etchegorry, J.P., 2021.
861 Comprehensive LiDAR simulation with efficient physically based rendering DART-Lux model (II):

Validation with GEDI and ICESat-2 measurements. Remote Sens. Environ. (in review)

Yang, X., Wang, C., Xi, X., Wang, P., Lei, Z., Ma, W., Nie, S., 2019. Extraction of multiple building heights using ICESat/GLAS full-waveform data assisted by optical imagery. IEEE Geosci. Remote Sens. Lett. 16(12), 1914-1918.

Yang, X., Wang, C., Xi, X., Wang, Y., Zhang, Y., Zhou, G., 2021. Footprint size design of large-footprint full-waveform LiDAR for forest and topography applications: A theoretical study. IEEE Trans. Geosci. Remote Sens. 59(11), 9745-9757.

Yin, T., Gastellu-Etchegorry, J.P., Lauret, N., Grau, E., Rubio, J., 2013. A new approach of direction discretization and oversampling for 3D anisotropic radiative transfer modeling. Remote Sens. Environ. 135, 213–223.

Yin, T., Lauret, N., Gastellu-Etchegorry, J.-P., 2016. Simulation of satellite, airborne and terrestrial LiDAR with DART (II): ALS and TLS multi-pulse acquisitions, photon counting, and solar noise. Remote Sens. Environ. 184, 454–468.

Yin, T., Qi, J., Cook, B.D., Morton, D.C., Wei, S., Gastellu-Etchegorry, J.-P., 2020. Modeling small-footprint airborne lidar-derived estimates of gap probability and leaf area index. Remote Sens. 12, 4.

Zhao, K., Suarez, J.C., Garcia, M., Hu, T., Wang, C., Londo, A., 2018. Utility of multitemporal lidar for forest and carbon monitoring: Tree growth, biomass dynamics, and carbon flux. Remote Sens. Environ. 204, 883–897.

Zwally, H.J., Schutz, B., Abdalati, W., Abshire, J., Bentley, C., Brenner, A., Bufton, J., Dezio, J., Hancock, D., Harding, D., 2002. ICESat's laser measurements of polar ice, atmosphere, ocean, and land. J. Geodyn. 34, 405–445.