Abstract—This article presents an open source CHP toolkit designed for processing small-footprint full-waveform LiDAR data to obtain estimates of effective leaf area index (LAIe) and canopy height profiles (CHPs). The use of the toolkit is presented with a case study of LAIe estimation in discontinuous canopy fruit plantations. The experiments are carried out in two study areas, orange and almond plantations, with different percentages of canopy cover (48% and 40%, respectively). For comparison, two commonly used discrete point LAIe estimation methods are also tested. The LiDAR LAIe values are first computed for each of the sites and each method as whole, providing ‘apparent’ site-level LAIe, which disregards the discontinuity of the plantations’ canopies. Since the toolkit allows calculation of the study area LAIe at different spatial scales, between–tree-level clumping can be easily accounted for and is then used to illustrate the impact of discontinuity of canopy cover on effective leaf area index (LAIe) retrieval. The LiDAR LAIe estimates are therefore computed at smaller scales as a mean of LAIe in various grid cell sizes providing estimates of ‘actual’ site-level LAIe. Subsequently, the LiDAR LAIe results are compared to theoretical models of ‘apparent’ LAIe vs. ‘actual’ LAIe, based on known percent canopy cover in each site. The comparison of those models to LiDAR LAIe derived from the smallest grid cell sizes against the estimates of LAIe for the whole site has shown that the LAIe estimates obtained from CHP toolkit provided values closest to those of theoretical models.

Index Terms—aggregation, airborne full-waveform LiDAR, discontinuous canopy cover, effective leaf area index (LAIe), saturation, simulation, small-footprint

I. INTRODUCTION

Knowledge of three-dimensional vegetation structure and foliage amount is important for many ecological applications [1]. Such information is useful or even necessary in fire behavioral studies, flood modelling, weather forecast, environmental change [2] and carbon balance models [3, 4], as well as in forestry applications such as timber volume and biomass estimations or biodiversity studies [5, 6]. One common parameter used to describe vegetation is leaf area index (LAI).

While several definitions of this index exist in the literature, the most often cited by scientists is ‘half the total leaf area per unit ground area’ proposed by Lang et al. [7] and Chen and Black [8]. Despite the simple definition, the estimation of LAI still remains problematic [9, 10], as the only way of retrieving true value of LAI is by harvesting. This is, however, a destructive method that is not really feasible at larger scales [9] or for long term monitoring [11].

There are a number of indirect methods of estimating LAI, which rely on inferring that index from other variables [12], that have resulted in varying success in estimating LAI. They include slow, laborious ground-based methods such as litter collection or hemispherical photography, and faster airborne and space-borne remote sensing methods. Each method has its respective advantages and disadvantages. For example, remote sensing techniques typically do not distinguish between woody and foliar elements of the canopy and therefore provide plant area index (PAI) rather than leaf area index. Furthermore, almost all indirect techniques (with the exception of allometric equations) rely on the gap fraction methodology, which assumes a random distribution of foliage material within the tree canopy. This assumption is often invalid in reality, with the

Manuscript received May 27 2015; revised February 2016. This work was funded by Engineering and Physical Sciences Research Council (grant numbers: F3322708, F4011613), School of Mathematical and Physical Sciences and School of Systems Engineering of University of Reading. Field campaign data was funded by Australian Research Council projects LE0560930, DP0557543 and FS100100040. Waveform processing was funded by the National Centre for Earth Observation, UK. The authors would like to thank Mihai Tanase for the field trip to the orange orchard site.

K. D. Fieber was with University of Reading and is now with Newcastle University, School of Civil Engineering and Geosciences, Newcastle, NE1 7RU, UK.

Ian J. Davenport is with University of Reading, School of Archaeology, Geography and Environmental Sciences, Whiteknights, RG6 6DW, UK.

James M. Ferryman is with University of Reading, School of Systems Engineering, Whiteknights, RG6 6AY, UK.

Robert J. Gurney was with University of Reading and is now with Portsmouth University, School of Engineering, Anglesea Road, Portsmouth PO1 3DJ, UK.

Jeffrey P. Walker is with Monash University, Department of Civil Engineering, Clayton, VIC 3800, Australia.

Jorg Hacker is with Airborne Research Australia, School of the Environment, Flinders University, Adelaide, SA 5001, Australia.

Colour versions of the figures in this paper are available online at http://ieeexplore.ieee.org
canopy elements showing more clumpy character (especially in coniferous trees). This in turn results in underestimation of LAI. Heterogeneity in foliage distribution leads to errors in LAI derived from gap fraction measurements because of the logarithmic relationship between gap fraction and LAI – an averaged variable gap fraction usually gives rise to an underestimate in LAI. Different terms have been proposed in literature to address the above effects and their difference in relation to true LAI. For example some authors prefer to use the term ‘LAI proxy’ [13]. In this study the effective leaf area index (LAlE) term, first introduced by Black et al. [14], will be used. However, contrary to Black et al. [14] who used the term to address the violation of the randomness assumption only, in this study, following Jonckheere et al. [2], LAlE term is used to address lack of correction both for clumping effect and for woody elements of the canopy.

The clumping is a well-known effect and has been reported to occur at several scales: between plants/trees, between branches, and between shoots [9]. Several studies have proposed methods to correct the LAlE estimates from indirect (remote sensing) measurements for clumping effects [15, 16]. Despite the fact that for discontinuous canopies such as rows of crops or trees, the underestimation of LAlE is especially pronounced [9], most of the studies only addressed between-shoot and between-branch clumping. However, few studies focused on clumping at the largest between-tree (or plant) scale. Lang and Yueqin [16] proposed a ground-based logarithmic averaging technique to account for gaps between rows of sorghum and wheat. The leaf area of crops was measured by averaging the transmission of direct sunlight linearly over a small horizontal distance and taking the logarithm of this mean. This method of LAI estimation provided better results than taking the mean of the transmission over the full distance. Ni-Meister et al. [17] used a hybrid geometric optical and radiative transfer (GORT) model to simulate the effects of discrete and heterogeneous canopies on large-footprint (SLICER) laser waveforms. The authors concluded that ignoring the clumping effect can significantly lower the estimates of foliage and biomass amount. Two further related studies by Ni-Meister et al. [18] and Yang et al [19] built on the GORT simulation model presented in Ni-Meister et al. [17] and showed the light interaction with mixed and heterogeneous plant canopies.

In recent years LiDAR technology has been found particularly well suited for describing vegetation structure [20-23], because of its ability to penetrate the tree canopy and its three-dimensional character. Much scientific discussion has been devoted to this technology and several methods of extracting LAI have been developed. Morsdorf et al. [13] computed a LiDAR LAI proxy as a fraction of the first and last returns inside the canopy in different data trap sizes (radii between 2-25m) and found the best correlation with hemispherical photography for 15m radius traps. This was due to the large range of zenith angles (0-60°) used in LAI estimation from hemispherical photographs. In contrast, fractional cover defined as vegetation hits over total laser echoes was found to provide the best results for a 2m radius data trap with fractional cover derived from hemispherical photographs at 0-10° angles. It was also found to be highly dependent on the size of the data trap adopted.

The main aim of this study is to present an open source CHP toolkit as a tool to examine the influence of discontinuity of canopy cover on LAlE retrieval from small-footprint full-waveform airborne LiDAR data. The methodology used for LiDAR processing in the CHP toolkit has previously been compared to hemispherical photography with R² of 0.83 at site-level [24]. The LAlEs estimates produced by the CHP toolkit are also compared to the performance of discrete point methods of LAlE estimation including their susceptibility to saturation. The impact of the aggregation area of LiDAR data is investigated by computing the LAlEs in decreasing grid cell sizes in two study areas with different percentages of crown coverage. Consequently the difference between the ‘apparent’ (disregarding between tree clumping, whole site) and ‘actual’ (accounting for canopy discontinuity, 2.5m cells) LAlEs of each study area was indicated. At the same time the benefits of small-footprint laser data for LAlE estimation in discontinuous canopy environments were highlighted. The results of LiDAR LAlE estimation of each method were compared to a theoretical model, and revealed the relative suitability of each method for LAlE estimation of discontinuous canopies.

II. STUDY AREA

To test the effect of the canopy discontinuity on LAlE retrieval two study areas were selected. Both study areas are located near the town of Yanco, within the Murrumbidgee catchment, New South Wales, Australia. The first one is a 100m by 50m area of orange orchard located between 393400m and 393500m (EastIng) and between 6169280m and 6169330m (Northing) (UTM, zone 55H). Ground elevation in the orange orchard ranges from 122m to 125m above mean sea level (AMSL) across the site, with the lowest elevations in the north west corner and rising towards the south. The orange trees are denser and taller in the south east while being smaller and sparser in the north west (Figure 1A). The direction of orange tree rows is south west to north east (at a bearing of about 60°) and the rows are about 7 m apart.

Figure 1. Areas of interest: A. Orange orchard aerial photography. B. Orange tree. C. Almond orchard shaded relief. D. Almond tree.
The second study site is a 100m by 50m area of almond orchard with the almond trees distributed in north-south rows with approximately 7m spacing (Figure 1C). The trees in each row are spaced about 5m from each other. The almond orchard study area is situated between 419180m and 419280m (Easting) and between 6160080m and 6160130m (Northing) (UTM, zone 55H). The ground elevation varies from 14m to 146m AMSL across the site, with the highest elevations towards the south of the site and the lowest elevations in the north west corner.

### III. Data

#### A. LiDAR

All laser scanning data were acquired by Airborne Research Australia with a full-waveform Riegl LMS-Q560 instrument [25] operating at 1550nm wavelength from a light aircraft. The orange orchard data were acquired on 3rd November 2006 as part of the National Airborne Field Experiment 2006 - NAFE’06 [26]. The flying altitude was 500 m above ground level, resulting in a 0.25 m footprint size. The average laser shot density was 4 per m² with an average point density (after a custom decomposition procedure) of 6.5 points per m². The laser altimetry data were captured along a 75 km long transect line across the Yanco site. The almond orchard data were acquired on 22nd September 2011 as part of The Third Soil Moisture Active Passive Experiment – SMAPEx-3 [27]. The flying altitude was around 400m for the almond orchard providing a beam footprint of 0.2m². Average laser shot density was 6.3 per m² and average point density (after the custom decomposition procedure) was 9.7 points per m² in the almond orchard area. In both datasets both transmitted and received waveforms were recorded and sampled with a frequency of 1GHz (1 ns corresponding to ~15cm vertical resolution).

#### B. Other data

Aerial photography was taken using an 11MegaPixel Canon EOS-1Ds digital camera fitted with a 34 mm lens, mounted on the same aircraft during the LiDAR acquisition in 2006, providing high resolution imagery over the orange orchard focus area. The ground pixel size of those images was about 15cm. The aerial image of the site was rectified for the purpose of providing ground reference data. The orange tree crowns were manually delineated on the rectified image and the percent crown cover in the study area calculated, yielding 48.2% cover. In total 206 trees (or partial trees) were delineated, with single crown area ranging from 1.4 to 20.5m². The mean crown area disregarding incomplete trees on the edges of area of interest yielded 12.5m² corresponding to a crown radius of about 2.0m (area standard deviation of 2.4 m²).

Since there was no high resolution aerial photography available for the almond orchard scene (the photography taken was of too low resolution for this purpose), the shaded relief image (Figure 1C) generated from LiDAR data was used as a base for manual delineation of the crowns. In total 160 trees (or partial trees) were delineated with single crown area ranging from 1.8 to 19.5m². The mean crown area disregarding incomplete trees on the edges of the area of interest yielded 13.1m² (which corresponds to 2m crown radius) with a standard deviation of 2.2m². The percent crown cover in this study area was 40.0%.

### IV. Methods

#### A. LAIe estimation

The techniques for Effective Leaf Area Index (LAIe) extraction from LiDAR data used in this study can be divided into two groups: (i) raw waveform methods and (ii) discrete point methods. The methodology used in the CHP toolkit is a raw waveform method. In the second group, two ways of determining gap probability from point cloud information were examined. LAIe estimates were extracted for the study areas as a whole, by processing total blocks of data (100m by 50m) at once, providing the ‘apparent’ LAIe of each site (disregarding canopy discontinuity). The computation was also performed in various decreasing grid cell sizes by dividing each study area into uniform square grid cells (50m, 25m, 10m, 5m, 2.5m, 2m, 1m), calculating LAIe for each cell and averaging across the whole site to provide one comparable site-level estimate.

Hemispherical photographs or LAI-2000 measurements were not taken at the time of LiDAR overpass at either of the sites, however, the methodology used in CHP toolkit has previously been compared with the use of hemispherical photography in a sparse forest canopy providing a high correlation of 0.83 at site-level [24]. The discrete point ratios have also been widely studied and validated [11, 13, 28-30]. In this study, a comparison to a theoretical model will be used as a way of validation.

As a result of gridding, the estimation of LAIe for a very few cells in the smallest grids saturated due to the lack of light returned from the ground (while returns from vegetation existed), making it impossible to calculate the gap fraction. This effect is well known to appear in both LiDAR and optical data based LAIe estimation [10, 31]. Richardson et al. [10] discussed possible solutions to avoid this effect in LiDAR discrete point data estimates, such as increasing point density, increasing aggregation cell size, taking into account the angle of incidence of individual pulses and configuring the detection process for lower energy reflections. One simple solution would be to set the number of ground returns to the minimum value of 1 [10], which consequently would cause some underestimation of LAIe. Furthermore, the error of underestimation would increase with increasing footprint and decreasing grid cell size. Nevertheless, this is a simple solution only in the case of discrete point methods. In waveform methodology this would mean creating an artificial ground waveform, which is not easily implemented due to energy variation between the laser shots. In this study, to enable calculation of mean LAIe for the whole area, the maximum LAIe cell value in the dataset was found prior to averaging across the site and assigned to all saturated cells in that dataset. Although this is not an ideal solution, it was the only practical option to implement for all LAIe estimation methods, and enabled the calculation of LAIe of the study areas.
1) CHP toolkit

CHP toolkit [32] is free open-source software developed at University of Reading, UK. The program uses the open-source PulseWaves [33] data format and it is designed to convert relatively small raw full-waveform small-footprint LiDAR datasets over vegetated areas into LAIe and canopy height profiles (CHPs). The waveform processing method used in CHP toolkit is based on an adaptation of the SLICER Canopy Height Profile procedure presented in Harding et al. [20] to small-footprint LiDAR data [24], where LAIe is estimated as one of the stages in the procedure. The CHP methodology is described in detail in [24] and consists of aligning the data relative to ground level, calculating the returned energy profile taking into account the vegetation-ground reflectance ratio, converting it into a canopy closure profile and subsequently to a leaf area profile, and finally turning it into a canopy height profile. The procedure does not require fully-optimized decomposition of the data; only the approximate (initial) location of the first and last pulse within the waveform and a DTM are needed.

The CHP toolkit consists of four modules. Module 1 reads the PulseWaves format data, provides some basic information about the data and converts the data in a selected area into an ASCII file. Module 2 reads the newly created ASCII file and performs simple peak detection to locate the peaks and to estimate their amplitude and width in the returned waveform data. Module 3 uses the point cloud data from Module 2 to filter ground points from single returns and from last returns. Finally, Module 4 processes previously generated files using the CHP methodology to obtain effective leaf area index estimates (WF1) as well as canopy height profiles. The reflectance ratio used to scale ground return is estimated via the technique proposed by Armston et al. [34] unless it is not possible to calculate it. Full documentation of the software is available online [32].

2) Point methods

In order to compare the performance of raw-waveform method and discrete point ratio LAIe estimation methods, the waveforms were extracted using the GeoCodeWF commercial software, calibrated and decomposed using a custom Gaussian decomposition procedure with a trust-region-reflective algorithm. The details of the decomposition and calibration procedures can be found in [35]. The decomposition produced point clouds with information for each point including easting, northing, elevation, amplitude, pulse width, peak position within the waveform train, pulse number within the waveform, number of pulses per waveform, angle of incidence, and backscattering coefficient [36]. Subsequently, single-peak-per-waveform points, classified into ground and vegetation using the backscattering coefficient, were used to build a digital terrain model (DTM) for each of the sites. The DTM was then subtracted from the point clouds to provide each point with its elevation above ground level.

The point clouds extracted in the process of custom Gaussian decomposition (as described in [35]) were analyzed to provide a discrete point LAIe derived from gap probability. Two point methods of LAIe extraction were tested in the area of interest. Both of the methods rely on gap fraction but use different gap probability (P) ratios. In the first method (PT1) the ratio is defined as the number of single-return waveforms returned from the ground $S_g (<0.5m$ above ground) to the total number of waveforms within the grid cell $N_{WF}$ [37, 38].

$$P = \frac{S_g}{N_{WF}}$$

(1)

The second point method (PT2) uses the ratio of the number of all ground points $N_g (<0.5m$ above ground) to the sum of ground and vegetation ($N_g + N_v$) returns in the grid cell [28]. This method is also equivalent to fractional cover used by Morsdorf et al. [13], with this difference that threshold used to separate ground and vegetation returns in that study was 1.25m.

$$P = \frac{N_g}{N_g + N_v}$$

(2)

The above ratios in point methods are then used to calculate LAIe based on Aber’s [39] equation:

$$LAI_e = -\ln(P)$$

(3)

B. Theoretical model

Due to the lack of hemispherical photography or LAI-2000 measurements, the results of LiDAR LAIe were verified against a theoretical model. The theoretical model represents corrected LAI with needle-to-shoot ratio of 1 published by Chen et al. (1997) [15].

The percent canopy cover of 40% in the almond orchard and 48% in the orange orchard, estimated based on aerial photography and shaded relief, were used to simulate theoretical study area LAIe for the two cases of taking into account the discontinuity of the canopy cover (‘actual’ LAIe) and disregarding it (‘apparent’ LAIe). The simulation was performed based on a range of average single-tree leaf area index LAIe values (from 0.2 to 10 with increments of 0.1), and known vegetation fraction cover ($A_T$). The ‘actual’ LAIe (LAIe$_{ACT}$) is the product of the average single tree LAIe$_T$ and the fraction of the area covered by the tree crowns $A_T$:

$$LAIe_{ACT} = LAIe_T A_T$$

(4)

The simulated ‘apparent’ LAIe disregarding the effect of canopy discontinuity was obtained by calculating LAIe for the whole study area from summed probabilities of penetration for the tree covered area ($P_{gapT} = e^{-LAIe_T}$) and bare soil area ($P_{gapS} = 1$) according to

$$LAIe_{APP} = -\ln\left(\frac{P_{gapT} A_T + P_{gapS} A_G}{A_T + A_G}\right) = -\ln(P_{gapT} A_T + A_G)$$

(5)

where $A_G = 1 - A_T$ is the fraction of the area not covered by vegetation.
V. RESULTS AND DISCUSSION

Figure 2 presents the relationship between the integrated vegetation (R_v) and ground (R_g) returns calculated according to the methodology of Armston et al. (2013) [34] and provided by the CHP toolkit in both study sites in 1m, 2m, 2.5m and 5m cells. These plots show a linear relationship between the integrals and confirm that the assumption of constant vegetation and ground reflectance relation is valid for both study sites. A theoretical reflectance ratio between vegetation and soil at the 1550nm wavelength used by Riegl LMS-Q560 should be around 0.5, which means that soil is twice as reflective as vegetation at this wavelength. The exact value will depend on reflective properties of vegetation species and soil types. In the orange orchard this ratio was determined to be slightly above the default value (0.56), while in the almond orchard it was found to be lower (0.41). This suggests that assuming the soil properties are similar in both sites – in the Yanco area loams composed of sand, silt and clay dominate [40] – almond trees are less reflective than orange trees at 1550nm wavelength.

Figure 3 presents the relationship between LAIe and grid cell size for both study sites, while Table 1 and Table 2 list the percentage of saturated cells in them. From the plots in Figure 3 one can notice that LAIe estimates for cells 5m and larger are almost constant for all methods. The trees in both sites have an average radius of about 2m, which means that using a cell of 5m and larger will not resolve the heterogeneity of the studied sites as the cells will have mixtures of vegetation and bare soil within them. It seems therefore that using a 5m or 50m LiDAR aggregation area will provide similar estimates of LAIe in the studied plantations. It is expected that with the decreasing grid cell size the overall study area LAIe will increase. This is due to the fact that the smaller the cells are, the more homogenous is the area within them (either fully covered with vegetation or

![Figure 2. Relationship between integrated vegetation (R_v) and ground (R_g) return in orange orchard (A) and almond orchard (B). R_v and R_g values calculated by CHP toolkit in 1m, 2m, 2.5m and 5m grid cells.]

![Figure 3. Effective leaf area index of the site depending on the grid cell size used (logarithmic scale). A. Orange orchard site. B. Almond orchard site. The LAIe of 2m cells is considered to be ‘actual’ whereas for the larger cells ‘apparent’ (see text for further information).]
ground), resulting in a higher average estimate across the overall study area. Plots in Figure 3 confirm that for the grid cell sizes smaller than 5m the estimates start to change dramatically.

Table 1. Percent of saturated cells in Orange orchard, where LAIe is impossible to calculate.

<table>
<thead>
<tr>
<th>Grid cell size</th>
<th>WF1</th>
<th>PT1</th>
<th>PT2</th>
</tr>
</thead>
<tbody>
<tr>
<td>10m</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5m</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2.5m</td>
<td>2.3%</td>
<td>12%</td>
<td>1.6%</td>
</tr>
<tr>
<td>2m</td>
<td>6.4%</td>
<td>19%</td>
<td>5%</td>
</tr>
<tr>
<td>1m</td>
<td>23%</td>
<td>37%</td>
<td>22%</td>
</tr>
</tbody>
</table>

Table 2. Percent of saturated cells in Almond orchard where LAIe is impossible to calculate.

<table>
<thead>
<tr>
<th>Grid cell size</th>
<th>WF1</th>
<th>PT1</th>
<th>PT2</th>
</tr>
</thead>
<tbody>
<tr>
<td>10m</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5m</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2.5m</td>
<td>0.1%</td>
<td>4.6%</td>
<td>0</td>
</tr>
<tr>
<td>2m</td>
<td>0.6%</td>
<td>18%</td>
<td>0.5%</td>
</tr>
</tbody>
</table>

In the case of the waveform method with estimates from the CHP toolkit, the tendency of increasing LAIe with decreasing grid cell size is consistent. Table 1 shows, however, relatively high saturation (23%) of the 1m cell estimates, and some small saturation of 2m (6.4%) and 2.5m (2.3%) cells in the orange orchard. In almond orchard the saturation of WF1 is almost none – only 0.6% cells saturated in 1m cell estimation. The difference in saturation levels is related to the density of foliage (orange trees seem denser than almond trees) as well as to the laser point density, which was higher in the almond orchard (6.3 shots per m² as opposed to 4 shots per m² in the orange orchard). The waveform method does not rely on the number of points extracted (it uses the raw light curve) and it takes into account the radiometric properties of the target. It is therefore capable of yielding higher single-cell LAIe values further indicating that this method may be suited for LAIe retrieval in fruit plantations with discontinuous canopy cover on the condition that an appropriate cell size is chosen.

The discrete point method PT1 (using only single returns) shows the tendency to have higher LAIe values for 2.5m cells (in comparison to 5m cells and larger), while for 2m and 1m cells the estimates stabilize and even drop in the orange orchard. In the almond orchard the increase is visible throughout; its rate, however, is not very consistent. This is likely due to the highest level of saturation of that method summarized in Table 1 and 2. The PT1 method uses only waveforms containing a single ground or vegetation return meaning that the point density is decreased. This in turn makes the method more prone to saturation (reaching 37% in 1m cells in the orange orchard and 18% in 1m cells in the almond orchard). Therefore, the smaller the grid, the more likely it is that there will be no ground return present in that cell. It is clear that PT1 is the method that saturates the easiest and its estimates from the smallest cells will be biased. It is therefore not particularly suitable for LAIe estimates in fruit plantations with a discontinuous canopy, especially if the point density is not very high.

The discrete point method PT2 (using all returns) shows only a slight increase of its values with decreasing cell size in the orange orchard study site, while in the almond orchard the estimates drop altogether for all of the smallest cells (Figure 3). One possible explanation for this anomalous behavior lies in the foliage density of the almond trees. As seen in Figure 1 almond trees have less dense foliage than orange trees and therefore allow more pulses to penetrate their canopies. The nature of the PT2 method is such that it disregards the area of pulse interaction with the target and treats all the points equally. The leaves are small and sparse enough that the laser gets some ground return in each waveform. Therefore, for almond trees more pulses partially transmit through their canopy and reach the ground, increasing the number of ground points detected. As a result, the percentage of ground points relative to all detected points increases with decrease in grid cell size causing the LAIe of the cell to drop. In consequence, this particular method shows an inverse relationship (decrease of LAIe instead of increase) while the cells become more homogenous. Thus, PT2 does not seem to be suitable for LAIe estimates of discontinuous canopies with scattered foliage elements smaller than the laser footprint.

If we now consider that LiDAR estimates of LAIe for the study site as a whole correspond to the ‘apparent’ LAIe that disregards the heterogeneity of the area (expressed by Eq.5) and that LiDAR LAIe site-level estimates from 5m cells and smaller are our ‘actual’ LAIe estimates, a comparison can be made to a theoretical model of ‘actual’ vs. ‘apparent’ LAIe (Figure 4). This comparison is presented in Figure 5. The difference in the LAIe calculated using these two equations presented in section IV.B has an exponential character and increases with the increasing LAIe of an average single tree (Figure 4). The ‘apparent’ LAIe asymptotically approaches a specific value with increasing ‘actual’ LAIe for each crown coverage percentage.

The ‘apparent’ LAIe of the theoretical model in the orange orchard site (48% of canopy cover) clearly saturates around the value of 0.65 (black dotted line Fig. 5A) corresponding to the ‘actual’ mean site LAIe of around 2.2 and an average single-tree LAIe of 4.6. Both LiDAR point methods provided ‘apparent’ LiDAR LAIe estimates (100m by 50m site) either equal to (PT1) or much higher (PT2) than that value of saturation causing a clear lack of correspondence with the theoretical model in the orange orchard site for those methods. Moreover, in order to match the theoretical model, PT2’s ‘apparent’ LiDAR LAIe estimate of 0.80 in the orange orchard would have to correspond to about 80% canopy cover in the site. This implies that PT2 method is not well suited for discontinuous canopy environments. In the case of PT1 the estimates are closer to the theoretical model, however, the high level of saturation makes the ‘actual’ estimates of 2.5m cells and smaller not reliable, while estimates of larger cells disregard the heterogeneity of the site underestimating the
‘actual’ LAIe value. The waveform method provides the most plausible LiDAR LAIe values in the orange orchard. The LiDAR site-level ‘actual’ LAIe estimate of 1.15 in 2.5m grid cell size achieved the best match with the theoretical model in a site with 48% crown coverage. This corresponds to an average single-tree LAIe value of 2.4. The LiDAR LAIe WF1 values from 1m and 2m cells overestimated the ‘actual’ theoretical LAIe. This, particularly being the case for 1m cell size, is likely related to the 23% saturation of LAIe estimates and leads to bias towards higher values due to the fact that the cells with no ground return were assigned a value of the highest calculated grid-level LAIe in the dataset (LAIe of 8.3).

In the almond orchard (40% crown coverage) the theoretical model saturates at an apparent LAIe value of 0.51. This corresponds to the ‘actual’ theoretical LAIe of about 2 and an average single tree LAIe of 5. As seen in Figure 5B PT2 once again provides estimates that do not match the theoretical model. Against the forecast, especially given that there is almost no saturation present, the smallest grid LAIe estimates yield lower values than those of larger cells. This once again confirms the lack of suitability of this method for LAIe estimation in discontinuous canopies of fruit plantations. The PT1 method has provided estimates that match the theoretical model better in the almond orchard than in the orange orchard. In particular, the estimates of 2.5m cells have matched the model well. This method, however, remains the most prone to saturation with 18% of the 1m cells and almost 5% of the 2m cells saturated, while none of the cells at this size saturated in WF1 and PT2. Similarly to the orange orchard site, in the almond orchard the waveform method provided the results best matching the theoretical model, particularly for 2.5m cells (although 2m cell estimation was also very close). 1m estimates once again proved to overestimate the ‘actual’ theoretical model.

The results in both study sites showed that the spatial resolution matching the theoretical model best is sampling with 2.5m grid cell size. From the perspective of spatial uniformity, it seems that the smaller the pixel (grid cell) size, the better the estimate of the LAIe should be as each cell contains more homogenous vegetation. However, as the above results have

---

Figure 4. Theoretical relationship between simulated ‘apparent’ LAIe and ‘actual’ LAIe of a study site depending on the crown coverage of 40% or 48%.

Figure 5. Comparison of LiDAR ‘actual’ vs. ‘apparent’ LAIe to theoretical model for orange orchard (A) and almond orchard (B).
shown, there is some risk of saturation (determined mostly by the point density), and therefore results can be biased for the smallest grid cell sizes. It is thus crucial to take into account both the characteristics of the vegetation in the study area as well as the limitations in the LiDAR LAIe methods while choosing the sampling size. Since the tree crowns in both study sites are on average about 4m in diameter, it is sensible to choose a smaller grid cell size in order to depict the heterogeneity of the sites, perhaps around half the tree crown diameter. In the study sites tested this value turned out to be 2.5m (as 2m cells slightly saturated in the orange orchard). This is in agreement with the Nyquist-Shannon sampling theorem suggesting that the sampling needs to be done twice as fast as the measured function changes [41].

VI. CONCLUSIONS

This study has presented the use of an open source CHP toolkit for the estimation of LAIe in a discontinuous canopy environment of fruit plantations. The performance of the raw-waveform method of the CHP toolkit was confronted with two commonly used discrete point ratios. It was shown that disregarding the canopy discontinuity may lead to considerable underestimation of the study area LAIe, the degree of which will depend on the LAIe of the trees as well as the percent crown cover within the study area. Furthermore, by examining the mean study area LAIe at different spatial scales, the study suggests that aggregation to larger footprints (or large footprint laser scanning data) may not be suitable for LAIe estimation of areas with discontinuous canopy cover without correction to account for violation of the randomness assumption. Thus, whether the footprint is of 5m or of 50m, the LAIe estimate remains almost constant and considerably underestimated. Only aggregation to cells smaller than crown diameter reduced the effect, exposing the weakness of aggregation to larger footprints for LAIe estimation in such environments.

The methods used in the study were previously compared to LAIe estimation methods such as hemispherical photography in other studies (e.g. [24], [13], [28]). However, hemispherical photography, although commonly used, is not an ideal source of LAIe estimates as it is prone to biases and random noise related to the thresholds used to separate vegetation from sky and lack the ability to separate between–crown and within-crown gap probabilities [13]. Airborne LiDAR in comparison can provide much better horizontal sampling and, by using the gridding method proposed in this study, can take into account heterogeneity of the study area by providing between–crown and within-crown gap probabilities.

Although, the absolute verification of the results of this study was not possible due to lack of LAIe estimates from an alternative source, the comparison of the methods against a theoretical model of ‘actual’ vs. ‘apparent’ LAIe behavior addresses some points regarding LiDAR LAIe estimation in discontinuous canopies. First of all, as expected and shown by the theoretical model, disregarding the effect of canopy discontinuity can have a severe impact on the LAIe retrieval, highly underestimating it. The lower the crown percent cover in the study area, the higher the underestimation of the ‘actual’ LAIe will be. Secondly, discrete point methods used in this study are not advisable for LAIe estimation in discontinuous canopy environments. This is due to both their higher susceptibility to saturation (PT1 based on single-return data) and high overestimation and unstable behavior (i.e. PT2 based on all ground returns to the total return number). Thirdly, the raw-waveform method implementation of CHP methodology in the CHP toolkit seems to provide the most probable results of the scene LAIe, enabling an easy calculation at a desired spatial scale. The agreement of this method with the theoretical model gives some confidence in the validity of the raw-waveform approach.

Finally, this study suggests that small-footprint LiDAR data offers a unique possibility to compute the LAIe in small aggregation areas (grid cell). However, some aggregation of small footprint LiDAR data is unavoidable due to the nature of the data - the small cross-section of those systems means that not all of the shots will reach the ground, therefore making the calculation of LAIe impossible at the individual footprint level. The decision on the size of the aggregation area should be made taking into account the density of the laser data, to avoid saturation of LAIe, as well as the characteristics of the study sites, to avoid under- and over-sampling of data (tree crown sizes). The resolution should ideally be in agreement with the Nyquist-Shannon theorem suggesting sampling twice as fast as the measured function changes, and in this case it means cell size should be about half of mean crown diameter. Further work should focus on testing the toolkit in other environments including coniferous stands having different degrees of canopy percent cover and validation against other ways of LAIe estimation. Furthermore, the combined effect of the clumping effect at all levels (i.e. between-tree, between-branch and between-shoot) should be investigated.

REFERENCES


Karolina D. Fieber received the MSc degree in geoinformatics, photogrammetry and remote sensing in 2006 from AGH University of Science and Technology, Krakow, Poland. In 2015 she was awarded the PhD degree by the National Centre for Earth Observation at the University of Reading, Reading, U.K. His work has included the estimation of snow mass and soil moisture information from passive microwave observations and the extraction of land surface characteristics from airborne laser alitmetry and SAR data. He is currently working in the Department of Geography and Environmental Studies at Reading University.

Ian J. Davenport (M’04) received the Ph.D. degree in physics with astronomy from the University of London, London, U.K., in 1993. He has conducted research at the National Environmental Research Council’s Unit for Thematic Information Studies, the Environmental Systems Science Centre and the National Centre for Earth Observation at the University of Reading, Reading, U.K. His work has included the estimation of snow mass and soil moisture information from passive microwave observations and the extraction of land surface characteristics from airborne laser alitmetry and SAR data. He is currently working in the Department of Geography and Environmental Studies at Reading University.

Professor James M. Ferryman (M’96) received the B.S. degree in computer science in 1990 and the Ph.D. degree in computational vision in 2000, both from the University of Reading, Reading, U.K.

In 1998, he became a Lecturer in computer science with the University of Reading, in 2005, a Senior Lecturer, and in 2015 full professor. He currently leads the Computational Vision Group, School of Systems Engineering, and has previously acted as Director of Research for the School of Systems Engineering, University of Reading. His current research interests include remote sensing, automated closed-circuit television surveillance of public spaces, behavioral analysis, biometrics, robotics and autonomous systems,
novel imaging modalities, and performance evaluation. He is the author of more than 100 scientific publications. He has participated in a wide range of UK and EU funded research programmes including the EPSRC REASON Project (EP/C533402), which examined robust methods for monitoring and understanding people in public spaces, the EU Co-Friend Project (FP7-ICT-214975) on the construction of a cognitive platform for airport apron monitoring, and the EU FastPass Project (FP7-312583) on development of a harmonized, modular reference system for all European automated crossing points. He strongly supports technology transfer.

Prof. Ferryman is a member of the British Computer Society and has acted as the Director of both the British Machine Vision Association and the Security Information Technology Consortium. Since 2000, he has been a Co-Chair of the IEEE International Workshop on Performance Evaluation of Tracking and Surveillance.

Professor Robert J. Gurney received the Ph.D. degree from the University of Bristol, Bristol, U.K., in 1975. He started his research career with the Institute of Hydrology, Wallingford, U.K. for five years. He then worked at the Hydrological Sciences Branch of the NASA Goddard Space Flight Center for ten years, with five as Branch Head. He became a Professor at the University of Reading in 1990, and served concurrently as Director of the NERC Unit for Thematic Information Systems, and, from 1995, as the Director of the NERC Environmental Systems Science Centre. He stepped down from this role in 2012 and is currently Professor of Earth Observation Science in the School of Mathematical and Physical Sciences at the University of Reading. His research interests include the use of Earth observation in land surface modelling, and the use of many types of data to constrain land surface models. He has been involved in many satellite missions throughout his career, including most recently the NASA SMAP mission.

Professor Victor M. Becerra (SM’03) received his PhD in Control Engineering from City University, London, in 1994. He is currently a Professor of Power Systems Engineering at the School of Engineering, University of Portsmouth, UK. Between 1989 and 1991, he was employed as an electrical engineer at C.V.G. Edelca, Caracas, Venezuela. Between 1994 and 1999 he was a post-doctoral research fellow the Control Engineering Research Centre at City University, London. During the period between 2000 and 2015 he was an academic at the School of Systems Engineering, University of Reading, UK.

His current research interests consider a range of issues related to electrical power systems, as well as the methods and applications of automatic control. These interests include control of power systems, energy storage and its integration to power grids, computational intelligence for smart grids, computational optimal control, nonlinear control, state estimation, and robotics. During his career he has received research funding from the EPSRC, the Royal Academy of Engineering, the European Space Agency, the Technology Strategy Board, the EU, and UK industry. Professor Becerra is a Fellow of the Institution of Engineering and Technology.

Professor Jeffrey Walker received his B.E. (Civil) and B.Surveying degrees in 1995 with Hons I and University Medal from the University of Newcastle, Australia, and his Ph.D. in Water Resources Engineering from the same University in 1999. His Ph.D. thesis was among the early pioneering research on estimation of root-zone soil moisture from assimilation of remotely sensed surface soil moisture observations. He then joined NASA Goddard Space Flight Centre to implement his soil moisture work globally. In 2001 he moved to the Department of Civil and Environmental Engineering at the University of Melbourne as Lecturer, where he continued his soil moisture work, including development of the only Australian airborne capability for simulating new satellite missions for soil moisture. In 2010 he was appointed as Professor in the Department of Civil Engineering at Monash University where he is continuing this research. He is contributing to soil moisture satellite missions at NASA, ESA and JAXA, as a Science Team member for the Soil Moisture Active Passive (SMAP) mission and Cal/val Team member for the Soil Moisture and Ocean Salinity (SMOS) and Global Change Observation Mission – Water (GCOM-W) respectively.

Professor Jörg M. Hacker received the Doctorate (Dr.rer.nat.) degree in atmospheric sciences from the Meteorological Institute of the University of Bonn, Bonn, Germany, in 1980. Since 1982, he has been Flinders University of South Australia, Adelaide, South Australia, where he built up Australia’s National Research Aircraft Facility ARA—Airborne Research Australia, which today is Australia’s only National Facility engaged in using and operating aircraft for atmospheric and environmental research. Currently, he is ARA’s Director and Chief Scientist, and also Professor of Airborne Research at Flinders University. He pioneered the concept of using cost-efficient small aircraft for atmospheric research, including the research described in the current paper. He holds a pilot’s license with several special endorsements and has more than 7000 flying hours in a wide range of aircraft, including flying gliders over the Andes in South America up to 12.5 km.