

RESEARCH

Open Access



# Cloth simulation-based construction of pit-free canopy height models from airborne LiDAR data

Wuming Zhang<sup>1</sup>, Shangshu Cai<sup>1,2\*</sup>, Xinlian Liang<sup>2\*</sup>, Jie Shao<sup>1</sup>, Ronghai Hu<sup>3,4</sup>, Sisi Yu<sup>5,6</sup> and Guangjian Yan<sup>1</sup>

## Abstract

**Background:** The universal occurrence of randomly distributed dark holes (i.e., data pits appearing within the tree crown) in LiDAR-derived canopy height models (CHMs) negatively affects the accuracy of extracted forest inventory parameters.

**Methods:** We develop an algorithm based on cloth simulation for constructing a pit-free CHM.

**Results:** The proposed algorithm effectively fills data pits of various sizes whilst preserving canopy details. Our pit-free CHMs derived from point clouds at different proportions of data pits are remarkably better than those constructed using other algorithms, as evidenced by the lowest average root mean square error (0.4981 m) between the reference CHMs and the constructed pit-free CHMs. Moreover, our pit-free CHMs show the best performance overall in terms of maximum tree height estimation (average bias = 0.9674 m).

**Conclusion:** The proposed algorithm can be adopted when working with different quality LiDAR data and shows high potential in forestry applications.

**Keywords:** Data pits, Tree crown, Canopy height models, Cloth simulation, Pit-free

## Background

In the last two decades, airborne light detection and ranging (LiDAR) has become a reliable remote sensing technique for forest inventory given its capability to provide precise and detailed three-dimensional (3D) information on forest structures directly (Lim et al. 2003; Popescu 2007; Hyypä et al. 2008; Yu et al. 2011; Huang and Lian 2015; Coomes et al. 2018; Stereńczak et al. 2018). The construction of a canopy height model (CHM) from LiDAR data is an effective means of increasing data usability (Mielcarek et al. 2018). The CHM represents an absolute canopy height above the terrain surface and is usually constructed by subtracting the digital terrain model

(DTM) from the digital surface model (DSM) (Khosravi-pour et al. 2014). The CHM quality significantly affects the estimation accuracy of forest inventories and subsequent biophysical parameters (Shamsoddini et al. 2013).

A common problem in the LiDAR-derived CHM is the occurrence of unnatural black holes (i.e., data pits appearing within the tree crown) (Mielcarek et al. 2018). Data pits exhibit irregular height variations, and their heights are lower than those of their neighbours in raster CHM (Khosravi-pour et al. 2014; Mielcarek et al. 2018). In a CHM with a certain resolution (< 2 m), data pits are usually single or several pixel(s) clumping together, that is, the data pit size is irregular (Zhao et al. 2013). Existing literature has indicated no specific causes of data pits. However, many researchers have inferred various possible causes. Data pits exist due to the combination of multiple factors, from data collection to post-processing. Shamsoddini et al. (2013) reported that the positioning error of the Global Positioning System (GPS) is a key cause of data pits. A significant GPS vertical error causes data pits when flight heights are < 3000 m

\* Correspondence: [sscai@mail.bnu.edu.cn](mailto:sscai@mail.bnu.edu.cn); [xinlian.liang@nlsfi](mailto:xinlian.liang@nlsfi)

<sup>1</sup>State Key Laboratory of Remote Sensing Science, Jointly Sponsored by Beijing Normal University and Institute of Remote Sensing and Digital Earth of Chinese Academy of Sciences, Beijing Engineering Research Center for Global Land Remote Sensing Products, Institute of Remote Sensing Science and Engineering, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China

<sup>2</sup>Department of Remote Sensing and Photogrammetry, Finnish Geospatial Research Institute, 02431 Masala, Finland

Full list of author information is available at the end of the article

above ground and scan angles are outside  $9^\circ$  (Latypov 2005; Goulden and Hopkinson 2010). Leckie et al. (2003) found that data pits exist because of the integration of different flight line datasets, the penetration of laser beams through canopy branches and foliage into the ground and the unevenly distributed spatial resolution of LiDAR point clouds caused by a scanning mechanism, variation in aircraft attitude and deflection of lost returns. Ben-Arie et al. (2009) inferred that, when the points are binned to a grid, many grid cells will remain empty because some information from points with similar  $x$ - $y$  coordinates but different  $z$  values are lost. After being interpolated to a raster, pits may be generated in empty cells (Axelsson 1999). Khosravipour et al. (2014) concluded that data pits may be formed during ground filtering when generating a DSM or a DTM, depending on filtering technique and LiDAR point density (Kraus and Pfeifer 1998). Previous studies have asserted that such pits have a seriously negative effect on forest parameter estimation because these pits disrupt the CHMs. Gaveau and Hill (2003) indicated that data pits cause the underestimation of tree height for LiDAR data with a mean point density of  $0.21 \text{ points} \cdot \text{m}^{-2}$ . Shamsoddini et al. (2013) found that data pits negatively affect the estimation of basal area and stand volume for LiDAR data with a mean point density of  $2 \text{ pulses} \cdot \text{m}^{-2}$ . Khosravipour et al. (2014) demonstrated that data pits lead to large omission errors (undetected trees) and commission errors (falsely detected trees) in treetop detection for LiDAR data with mean point densities of 7 and 160  $\text{points} \cdot \text{m}^{-2}$ . Thus, data pits must be filled to ensure that the uppermost layer of the forest canopy is realistically represented by the LiDAR-derived CHM.

Many algorithms have been proposed to construct pit-free CHMs. These algorithms can be divided into two categories, namely, raster- and point cloud-based methods. Raster-based methods are achieved by interpolating point clouds into raster images. The data pits are then filled by the classical image processing methods, such as mean and median filters. These filters are simple and fast, but the result is sensitive to the optimal kernel size given the pit size irregularity (Shamsoddini et al. 2013; Mielcarek et al. 2018).

Kernel size selection is critical to the success of many image processing methods. For example, when using morphological-based ground filtering algorithm, only small non-ground objects, such as trees, will be effectively removed by small kernels; however, this algorithm tends to over-remove the ground points when using large kernels (Zhang et al. 2003). Similarly, for mean and median filters, only data pits with a small size will be effectively filled by small kernels; however, these filters tend to over-smooth the CHM when using large kernels (Ben-Arie et al. 2009).

These filters negatively affect the accuracy of measurements, such as underestimation of tree heights and crown radius, deriving from the CHM. Such a phenomenon occurs because the elevation values of other pixels in CHM, except for the pitted pixels, are also changed, leading to treetop omission and crown shoulder reduction (Khosravipour et al. 2014). Accordingly, several solutions have been proposed to fill data pits without affecting other CHM values. Shamsoddini et al. (2013) detected data pits by computing a similarity index in a  $3 \times 3$  kernel and then used a mean filter with a  $5 \times 5$  kernel to fill such pits. Ben-Arie et al. (2009) used a Laplacian operator with a  $3 \times 3$  kernel to detect data pits and then utilised a median filter with a  $3 \times 3$  kernel to fill such pits. These improved algorithms only fill the pitted pixels without affecting other pixel values in CHM. However, they cannot solve the kernel size selection problem. Therefore, these algorithms may be ineffective for large pits because they use a fixed kernel size.

Another type of pit-free CHM construction methods (i.e., point cloud-based methods) directly works on height-normalised point clouds rather than raw CHMs. Khosravipour et al. (2014) constructed a stack of partial CHMs from different height intervals of the canopy; subsequently, these CHMs were combined into one CHM on the basis of the highest value across all CHMs for each  $x$  and  $y$  raster position. However, this algorithm must carefully determine a series of complicated parameters for filling pits at different heights. Chen et al. (2017) used locally weighted regression and  $z$ -scores to construct a pit-free CHM. This algorithm is highly automatic but may not work for pits in sparse density data because this algorithm requires calculation in the neighbourhood. The results were related to neighbourhood size.

The use of the aforementioned algorithms has proven to be successful. However, the prominent limitations of existing algorithms are presented as follows: (a) having applicability only to LiDAR data with regular-sized pits, (b) changing the values of not only the pitted pixels but also other CHM ones or (c) requiring numerous complicated parameters. To cope with these problems, we propose a novel algorithm based on cloth simulation for constructing a pit-free CHM. The specific objectives of this study are to (a) evaluate the applicability of the proposed algorithm to different proportions of data pits, (b) assess cloth simulation-based CHMs through visual comparison of the mean- and median-filtered CHMs, and (c) evaluate and compare the accuracy of the maximum tree height estimation using cloth simulation-based, mean-filtered and median-filtered CHMs.

## Materials

### Real data

The study area is located in the Genhe Forestry Reserve ( $120^\circ 11'2''$ – $120^\circ 55'5''$ E,  $50^\circ 20'0''$ – $52^\circ 30'0''$ N), Greater Khingan

of Inner Mongolia, Northeastern China (Fig. 1a). The area size is approximately 33 km<sup>2</sup>. The elevations range from approximately 775 to 1300 m above sea level, and the slopes are less than 15° in 80% of the area. The area is characterised by monsoon-influenced subarctic climate with an annual average temperature of −5.3 °C. The forest types are mainly composed of Dahurian Larch (*Larix gmelinii*) and White Birch (*Betula platyphylla* Suk.).

The LiDAR data for the study area were collected from August to September 2012 using a Leica ALS60 system on a Yun-5 aircraft. The system wavelength was 1064 nm, and the beam divergence was 0.22 mrad. In our datasets, this system operated at a 166 kHz pulse rate at 1800 m above ground level with a field of view of approximately 30° and recorded up to four returns for each laser pulse. The scan angle statistics indicate that 97% and 99% of the pulses fall within the 0°–12° and 0°–15° zenith angles. The average point density was 8.24 points · m<sup>−2</sup> with an average spacing of 0.41 m.

Nine test plots (called L1–L9) with an area of 45 m × 45 m were established in the study area in August 2013 (Fig. 1b). The positions of the four corners of each plot were measured using a differential GPS device. The maximum tree height of each plot was measured using a hand-held laser rangefinder. The collection time between field-measured and LiDAR data was not absolutely consistent (the time interval is 1 year). Considering that the height growth rate of the tree species in the study area is relatively slow, field-measured

data could be acceptable for validating the proposed algorithm (Tang 2013).

### Simulated data

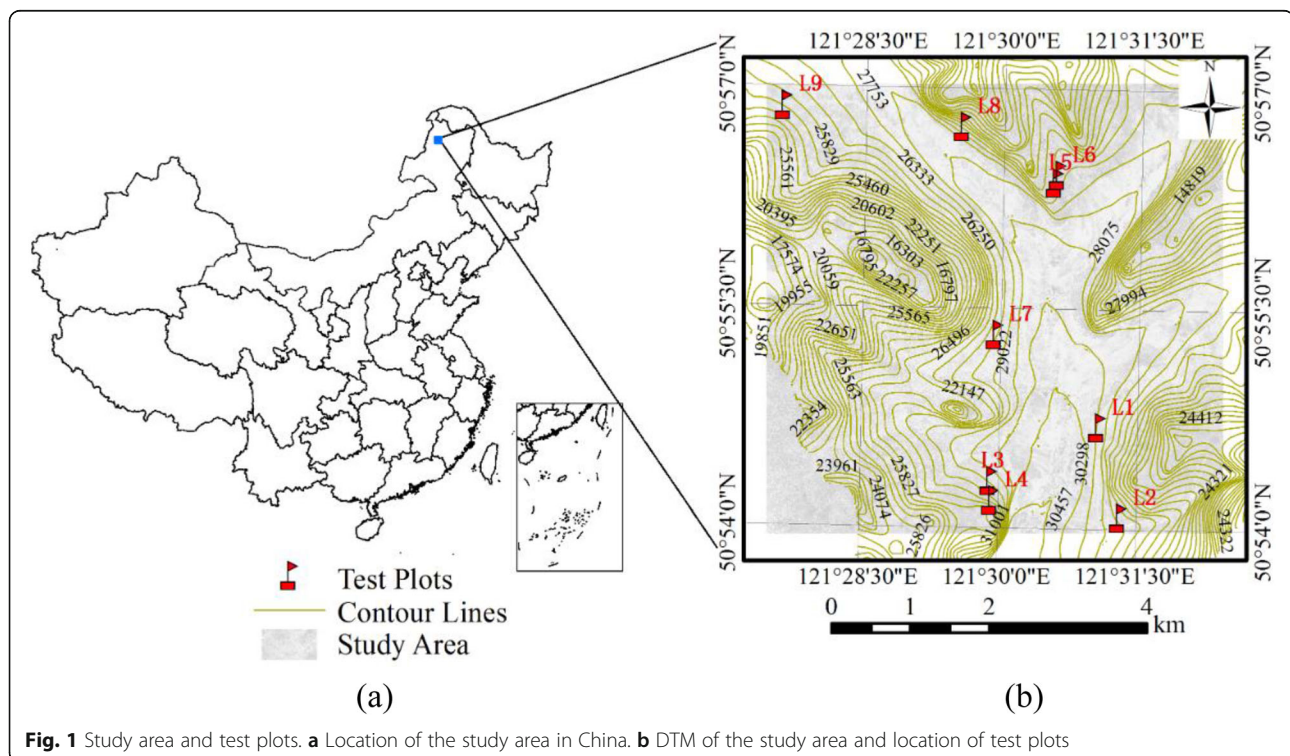
Two landscapes, namely, discrete floating hemisphere and cone canopies, were simulated. The hemisphere and cone models are expressed in accordance with the idea developed by Dong (2009).

$$\text{Hemisphere: } z = \sqrt{r^2 - x^2 - y^2}, \quad (x^2 + y^2 \leq r^2, z > 0), \quad (1)$$

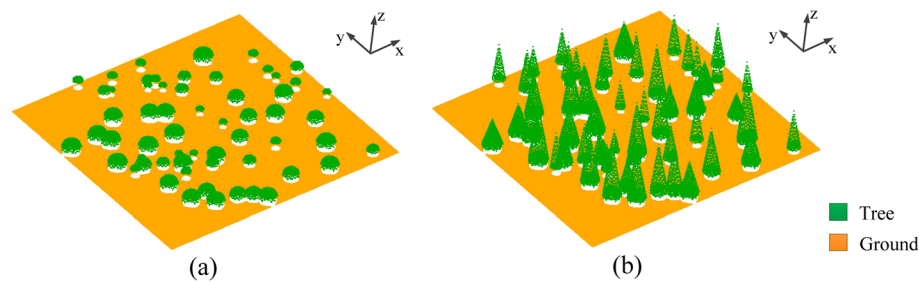
$$\text{Cone: } z = r \times \cot\theta - \frac{\sqrt{x^2 + y^2}}{\tan\theta}, \quad (x^2 + y^2 \leq r^2, 0 < \theta < 90^\circ), \quad (2)$$

where  $(x, y, z)$  represents the simulated point coordinates,  $r$  is the base radius and  $\theta$  is the angle between slant and cone heights.

Figure 2 demonstrates that the hemisphere canopy scene is a 50 m × 50 m square scene containing 60 hemisphere canopies. The radius and tree height values are set to 3–6 and 7–10 m, respectively. The cone canopy scene is a 50 m × 50 m square scene containing 60 cone canopies. The radius and tree height values are set to 3–6 and 18–55 m, correspondingly. The ground point height and point space of both simulated scene data are 0 and 0.05 m, respectively.



**Fig. 1** Study area and test plots. **a** Location of the study area in China. **b** DTM of the study area and location of test plots



**Fig. 2** Constructed 50 m × 50 m forest scene. **a** Hemisphere canopy scene. **b** Cone canopy scene

Different proportions of data pits (10%–60%) were randomly created in canopies, and their heights were randomly reduced from their original elevations. Such data pit proportions reflect the degree of irregularity in pit size. For example, high proportions of data pits in canopies typically reflect a large area of clustered dark pixels in raster CHM. Figure 3 depicts the point clouds of hemisphere and cone canopies, in which 10%, 20% and 50% of the data pits are created.

### Methodology

We present a cloth simulation-based framework to construct pit-free CHMs from point clouds and compare the results with those of other common algorithms using a comprehensive accuracy assessment procedure. The flowchart of this study is illustrated in Fig. 4. The methodology used in this study includes five main parts, namely, pre-processing, cloth simulation-based pit-free CHM construction, post-processing, comparison of different algorithms and accuracy assessment.

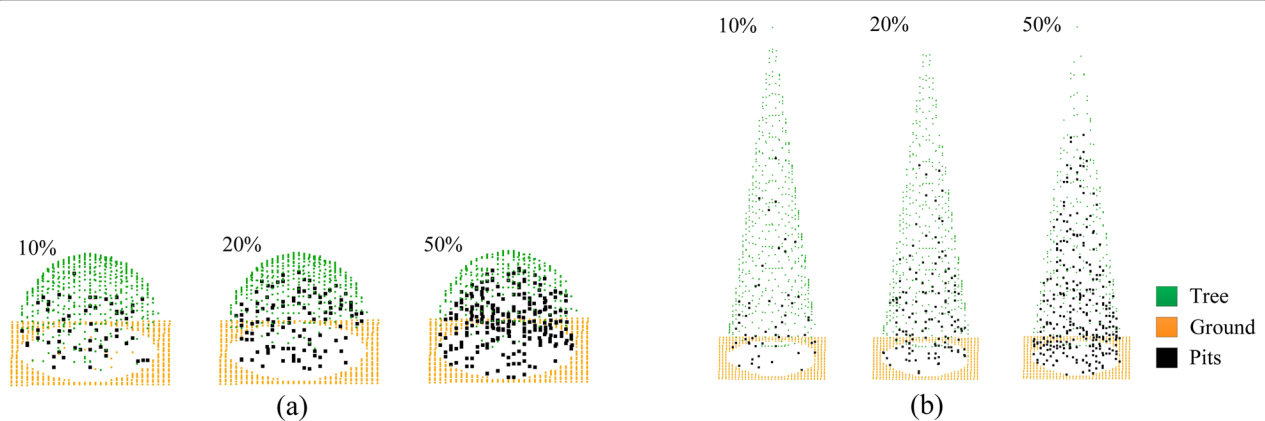
### Pre-processing

The pre-processing of LiDAR data consists of three steps, namely, noise removal, height normalisation and

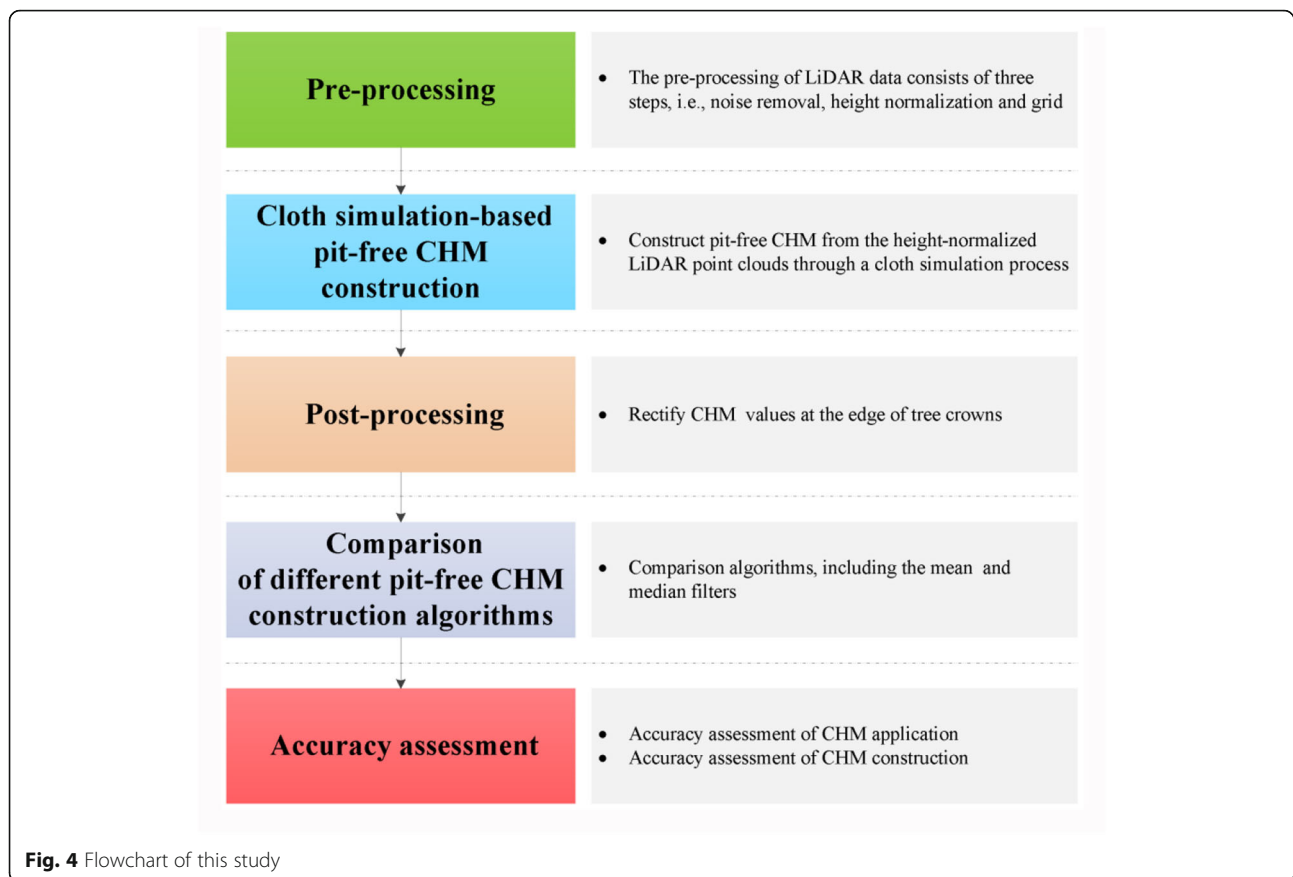
grid. In practice, the real and simulated data were pre-processed using the LAStools (point cloud processing software, Rapidlasso Company, Gilching, Germany) modules which include lasnoise, lasground and lasgrid (Isenburg 2017). The first step in the real data was to remove high or low noise points using the lasnoise module. Secondly, the lasground module was used to extract the ground points and then calculated the relative height above the ground for each point (i.e.,  $z$  coordinate). Each point was height-normalised by replacing the height of each point with its relative height above the ground. Consequently, the height of all points classified as the ground is zero. Finally, the lasgrid module was used to grid the height-normalised points, only retaining the maximum point in each grid cell. Considering that the simulated data have no noise, and the ground height is set to zero, these data were assigned to a grid, and only the maximum point in each grid cell was retained using the lasgrid module.

### Theoretical basis: cloth simulation

This algorithm is based on the physical process simulation of cloth-touching objects (Zhang et al. 2016; Zhang



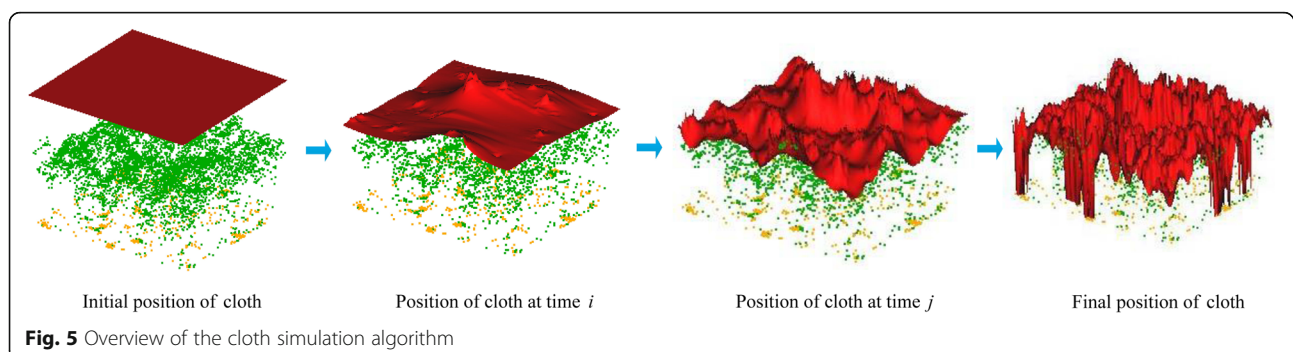
**Fig. 3** Canopy samples with 10%, 20% and 50% of data pits. **a** Hemisphere canopy. **b** Cone canopy

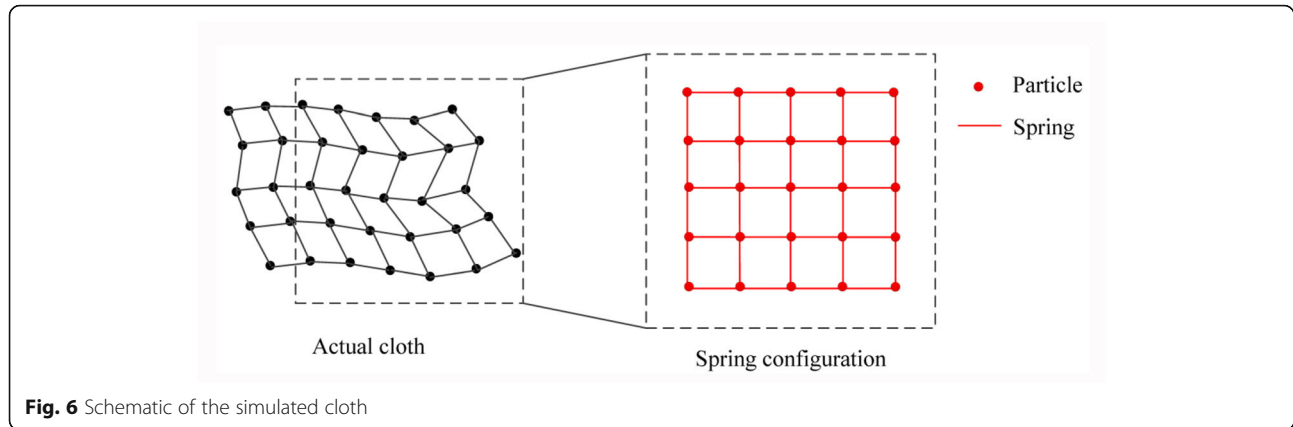


et al. 2019; Cai et al. 2019; Wan et al. 2019). Imagine that a piece of cloth is placed above a height-normalised forest landscape, and then this cloth progressively drops because of gravity. When touching the uppermost layer of the forest canopy, the corresponding particles of the cloth will stop dropping and be fixed. The movement of cloth particles near the pits will be constrained by forces produced by adjacent fixed cloth particles. The final shape of the cloth can be confirmed and regarded as a pit-free CHM (Fig. 5).

Cloth modelling is a term of 3D computer graphics. Typically, the cloth is modelled as a grid consisting of particles and springs. Both components control the shape of the cloth (Fig. 6).

The positions of the cloth particles in the 3D space are calculated to confirm the shape of the cloth at a specific time. The position of each particle is determined by an external force operation followed by an internal one. On the basis of Newton's second law, the relationship between position and forces is calculated using Eq. (3).





**Fig. 6** Schematic of the simulated cloth

$$m \frac{\partial Z(t)}{\partial t^2} = F_{\text{ext}}(Z, t) + F_{\text{int}}(Z, t), \quad (3)$$

where  $m$  stands for the mass of the cloth particle ( $m$  is typically set to 1),  $Z$  indicates the position of a particle at time  $t$ ,  $F_{\text{ext}}(Z, t)$  represents the external force and  $F_{\text{int}}(Z, t)$  denotes the internal force produced by inter-connections between particles. Figure 7 illustrates such a process.

For simplicity, the external force is set to zero, that is, each particle drops to measurements at a constant velocity. The displacement vector  $\overrightarrow{Z(\Delta t)}_{\text{ext}}$  of each particle after one time step  $\Delta t$  is determined only under the external force:

$$\overrightarrow{Z(\Delta t)}_{\text{ext}} = v \Delta t \cdot \vec{n}_{\text{ext}}, \quad (4)$$

where  $v$  is the velocity, and  $\vec{n}_{\text{ext}}$  represents a normalised vector that points to the movement direction,  $\vec{n}_{\text{ext}} = (0, 0, -1)^T$ . If the cloth particle collides with the measurement surface, then this particle will stop falling and be labelled unmovable. Displacement distance is the height difference (HD) between the old position of this particle and the measurement surface, that is,  $\overrightarrow{Z(\Delta t)}_{\text{ext}} = \text{HD} \cdot \vec{n}_{\text{ext}}$ .

After dropping the cloth particles, an internal force is exerted to movable particles to restrict the particles' displacement in the void areas, such as data pits, of the surface. In Fig. 8, the neighbouring particle and the current one are moved by the same distance in the opposite direction when the neighbouring particle is movable; when the neighbouring particle is unmovable, the current one is moved towards the neighbouring particle by half of the HD between the two particles. The displacement vector  $\overrightarrow{Z(\Delta t)}_{\text{int}}$  of the current particle is computed using (5).

$$\overrightarrow{Z(\Delta t)}_{\text{int}} = \sum_{i=1}^8 \frac{1}{2} b(Z_{p_i} - Z_{p_0}) \cdot \vec{n}_{\text{int}}, \quad (5)$$

where  $b$  is 0 when the current particle is unmovable, otherwise it is equal to 1,  $Z_{p_0}$  represents the height value of the current particle  $p_0$ ,  $Z_{p_i}$  ( $i = 1, 2, \dots, 8$ ) stands for the height values of the eight neighbouring particles  $p_i$  that connect with  $p_0$ , and  $\vec{n}_{\text{int}}$  represents a normalised vector that points to the movement direction,  $\vec{n}_{\text{int}} = (0, 0, 1)^T$ . The final displacement vector  $\overrightarrow{Z(\Delta t)}$  of each particle after one time step  $\Delta t$  is calculated as follows:

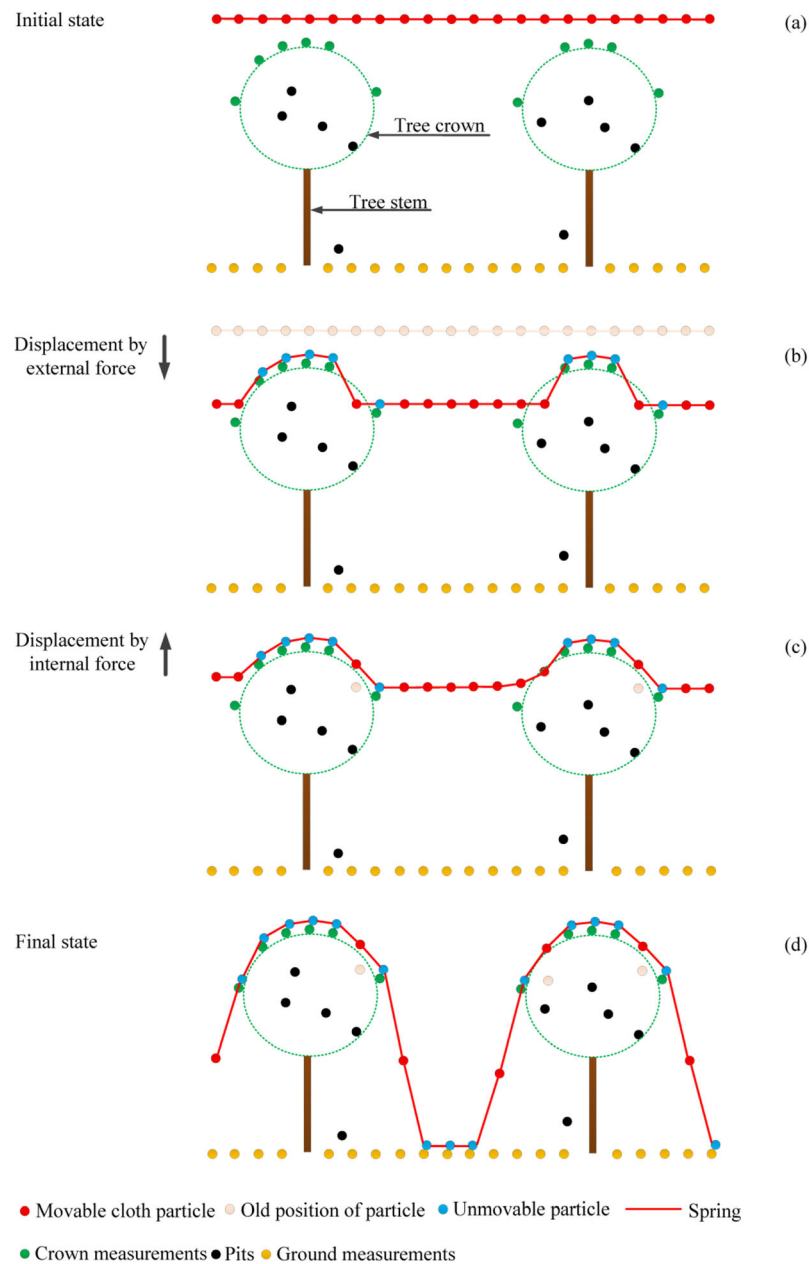
$$\overrightarrow{Z(\Delta t)} = \overrightarrow{Z(\Delta t)}_{\text{ext}} + \overrightarrow{Z(\Delta t)}_{\text{int}}, \quad (6)$$

### Post-processing

At the edge of tree crowns, this algorithm may yield relatively large errors because the cloth poorly fit with ground measurements given the internal constraints amongst particles (Fig. 9). A post-processing method is proposed to solve this problem. This method searches an unmovable particle in the eight adjacent neighbourhoods of each movable particle. If the height value of unmovable particle and that of the nearest LiDAR point for the movable particle in the  $x$ - $y$  plane equal to zero, then the movable particle is moved to the ground and set as unmovable. The example exhibited in Fig. 9 displays that  $P_2$  is the unmovable particle from the eight adjacent neighbourhoods of  $P_1$ . The height values of  $P_1$  and  $P_2$ , which are the nearest LiDAR points of  $P_1$  and  $P_2$  correspondingly, are both zero, then  $P_1$  is moved to  $P_1'$  and is set as unmovable. This procedure is repeated until all the movable particles are properly handled.

### Comparison between different pit-free CHM construction algorithms

We compared our algorithm with two other pit-free CHM construction algorithms, namely, mean and median filters, to test the performance of the proposed algorithm. Firstly,

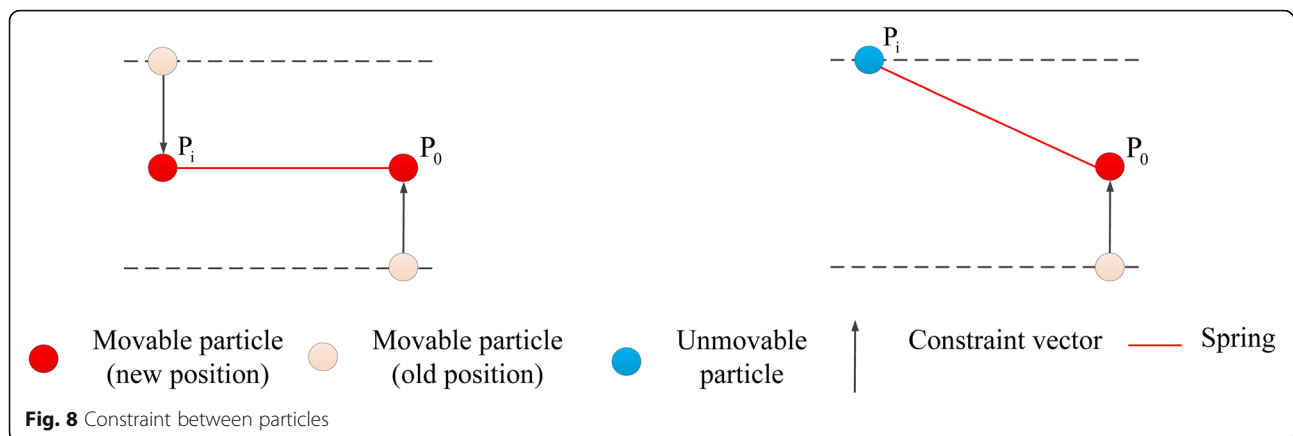


**Fig. 7** Cloth simulation process. **a** Initial state. A piece of cloth is placed above the height-normalised forest LiDAR measurements. **b** Each particle drops to the LiDAR measurements under the influence of external force, and their displacement is calculated. If the cloth particles collide with the measurements, then these particles will stop falling and be labelled unmovable. **c** Each movable particle is moved under the influence of internal force produced by the neighbouring particles. Steps (b) and (c) are repeated until the maximum height variation of all particles is sufficiently small. **d** Final state. The shape of the cloth is regarded as the pit-free CHM

the raw CHM was constructed using the las2dem modules of LAStools. The LiDAR points were triangulated into a triangulated irregular network on the basis of Delaunay triangulation and then rasterised onto a raster through standard linear interpolation. Secondly, the raw CHM was

processed using the mean and median filters for pit filling. The following derivative products were generated:

- Mean-filtered CHM: the raw CHM was filtered using a mean filter ( $3 \times 3$  kernel).



- Median-filtered CHM: the raw CHM was filtered using a median filter ( $3 \times 3$  kernel).

#### Accuracy assessment

Considering that the accuracy of the CHMs influences that of the tree height estimation in the raster-based method, the height of the tallest tree in each plot was obtained from the CHMs and compared with a field-measured height to evaluate the performance of the pit filling algorithms. This approach indirectly provides a comparison between the proposed algorithm and the other algorithms. The accuracy of the tree height measurements was assessed by computing the bias in accordance with the following mathematical formulation:

$$\text{bias} = \frac{1}{n} \sum_{i=1}^n (h_m - h_c), \quad (7)$$

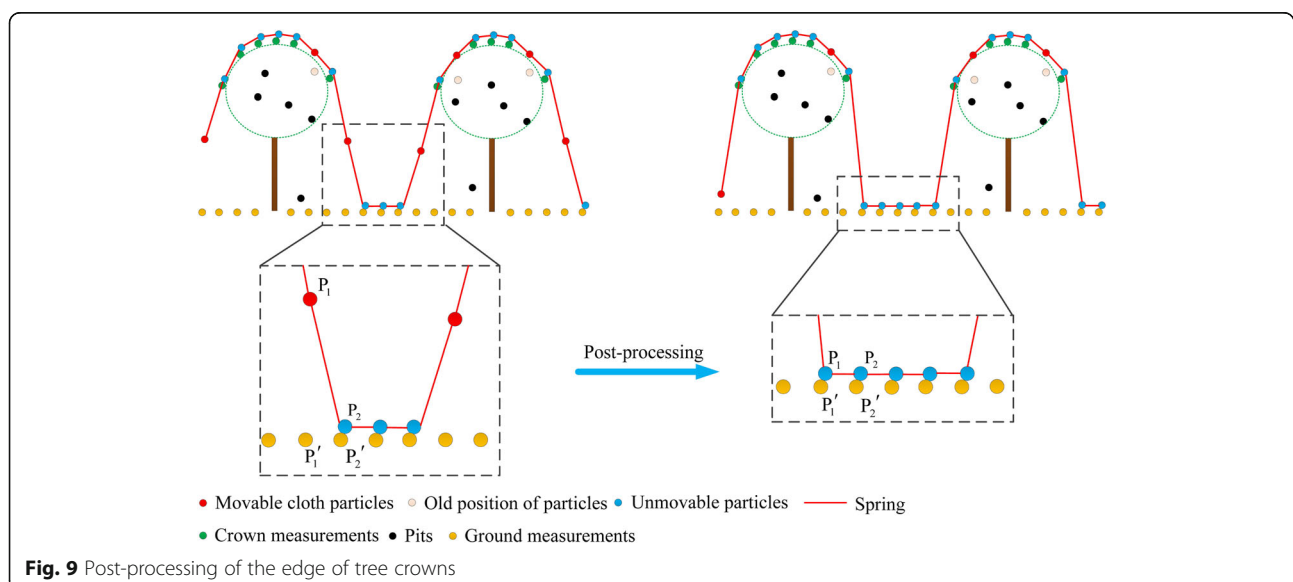
where  $n$  is the number of plots,  $h_m$  is the field-measured

plot-level maximum tree height and  $h_c$  is the plot-level maximum tree height derived from the CHM.

We constructed a reference CHM using the simulated data without data pits to validate the accuracy of the constructed pit-free CHM. Based on the index for evaluating the accuracy of DTM, the root mean square error (RMSE) was used to measure the HD between each pixel of the reference CHM and the corresponding pixel of the constructed pit-free CHM. This index is expressed as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (z_i - \hat{z}_i)^2}, \quad (8)$$

where  $z_i$  and  $\hat{z}_i$  are the true and predicted values at the  $i$ -th pixel in the CHM, respectively.  $n$  is the number of pixels in the CHM.



## Result

### Evaluation of real data

#### Visual performance

The performance of the proposed algorithm was visually compared with the result of the mean and median filters. With an average spacing of 0.41 m, the CHMs were generated with a resolution of 0.5 m. The data pits are clearly visible in the raw CHM (Fig. 10 (first row)). The mean filter fills the data pits, but the constructed CHM is heavily blurred (Fig. 10 (second row)). The median filter and our algorithm have a good ability of filling data pits (Fig. 10 (third and last rows)). The elevation values of other pixels, except the pitted ones, in the cloth simulation-based CHM are effectively maintained. The proposed algorithm performs better than the other algorithms in terms of data pit filling and appearance preservation of the uppermost layer of the forest canopy.

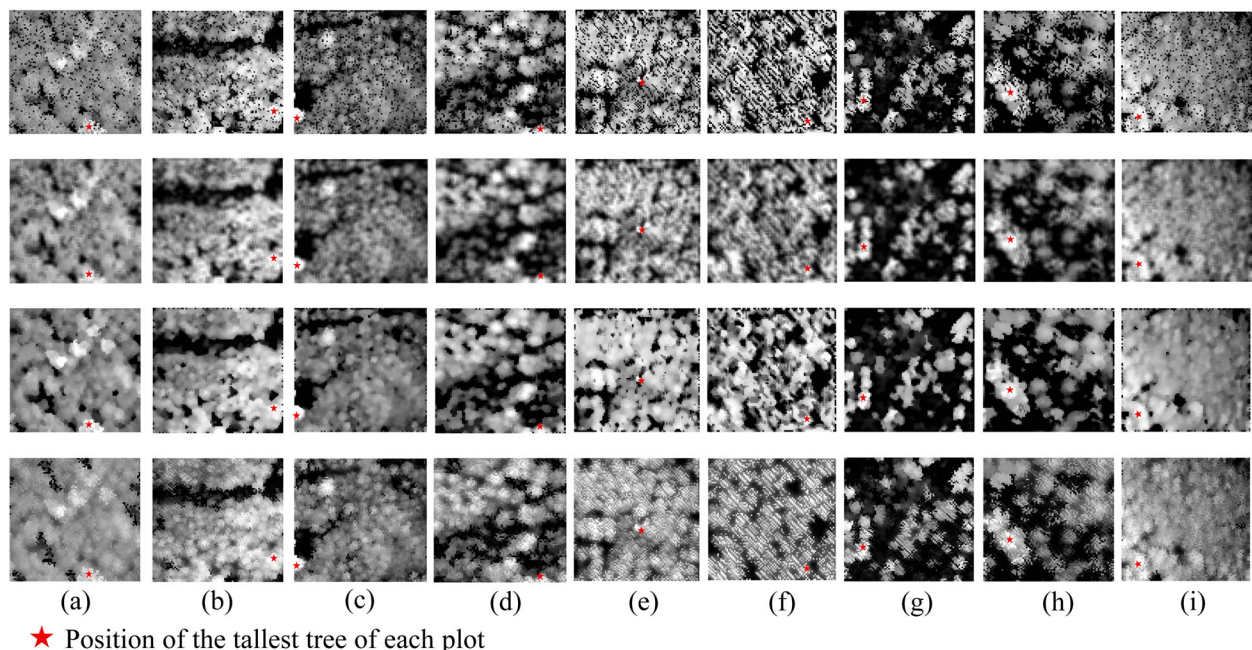
#### Assessment of the CHM application accuracy

Figure 10 and Table 1 present the position and estimated tree heights of the nine tallest trees derived from the three algorithms, respectively. Significant biases are observed between the field measurements and the maximum tree heights extracted from the CHMs constructed through different algorithms. Our algorithm is more accurate than the other algorithms. Specifically, the biases of the mean and median filters are 4.4894 and 4.4414 m, but that of the proposed algorithm is 0.9674 m (Table 2).

The CHM resolution plays an important role in estimating forest structure attributes (Chen et al. 2006; Khosravipour et al. 2014). Some optimal resolution setting principles have been proposed. For example, Nyquist sampling theory stipulates that the resolution must not be larger than half the size of the minimum object of interest (e.g., the minimum tree crown) and overly smaller than the average pulse spacing (Nyquist 1928; Chow and Hodgson 2009; Mielcarek et al. 2018). Raster CHMs with resolutions of 0.2, 0.5, 0.8, 1 and 1.5 m are constructed from the real data. This task is performed to test the robustness of the proposed algorithm for deriving a plot-level maximum tree height from CHMs under different resolutions. Table 2 shows biases between the field measurements and the maximum tree heights extracted from the CHMs constructed by different algorithms. In the mean and median filters, the resolution significantly affects the estimation accuracy of the tree height. However, our algorithm is unaffected and achieves the maximum accuracy at all resolutions.

#### Evaluation of simulated data

On the basis of the point spacing of simulated data, the CHMs were constructed with a resolution of 0.5 m. Taking hemisphere canopy scene as an example, Fig. 11 shows the CHMs constructed by different algorithms using the data with 10% pits. The raw CHM has some data pits randomly distributed in the image, and the pit



**Fig. 10** Comparison of the visual results for the nine plots. Subfigures (a)–(i): L1–L9. From the first row to the last row: Raw, mean-filtered, median-filtered and cloth simulation-based CHMs

**Table 1** Maximum tree height estimation results for the three algorithms in the nine plots (unit: m)

|        | L1     | L2      | L3      | L4      | L5      | L6     | L7      | L8     | L9      | Avg.    |
|--------|--------|---------|---------|---------|---------|--------|---------|--------|---------|---------|
| Mean   | 26.899 | 17.4509 | 18.4181 | 15.2007 | 28.1978 | 26.066 | 23.2291 | 33.336 | 16.0481 | 22.7606 |
| Median | 27.751 | 17.451  | 18.412  | 15.2685 | 26.051  | 26.238 | 24.002  | 33.932 | 16.172  | 22.8086 |
| Ours   | 28.275 | 17.984  | 18.924  | 16.498  | 51.59   | 27.166 | 25.277  | 34.163 | 16.666  | 26.2826 |

size is small and usually a single pixel (Fig. 11b). The mean filter fills data pits, but the constructed CHM is seriously smoothed, especially for edges of tree crowns (Fig. 11c). The median filter and our algorithm exhibit optimal performance (Fig. 11d and e).

Figure 12 illustrates the CHMs constructed through different algorithms using the data with 50% pits. The pit size in the raw CHM is irregular (Fig. 12b). The mean filter yields a relatively good pit filling result than the median one, but the constructed CHM is significantly smoothed (Fig. 12c). The median filter can fill small pits, but it deteriorates large ones (Fig. 12d). In comparison with the abovementioned filters, the proposed algorithm has the optimal ability to fill data pits. The proposed algorithm only fills the pits in the raw CHM without altering the values of other pixels; accordingly, a CHM with nearly similar features to the reference CHM is generated (Fig. 12e). In summary, the proposed algorithm performs better than the other algorithms for the raw CHM exhibiting pits with irregular size variations.

Table 3 shows that the proposed algorithm generates optimal accuracy. For example, in the hemisphere canopy scene, the proposed algorithm is approximately 2.5718, 1.9461 and 1.6788 times equally accurate as the raw CHM, mean filter and median filter in terms of average RMSE, correspondingly. The accuracy of all algorithms deteriorates with the increase in the pit proportion. For example, in the hemisphere canopy scene, the RMSE of the proposed algorithm increases from 0.2031 m to 0.5209 m when the pit proportion increases from 10% to 60%. By contrast, the RMSEs of raw CHM, mean filter and median filter increase from 0.5428 m to 1.2882 m, from 0.4334 m to 1.0299 m and from 0.308 m to 1.0324 m, respectively.

**Table 2** Biases between field-measured and CHM-derived plot-level maximum tree heights for the three algorithms under different CHM resolutions (unit: m)

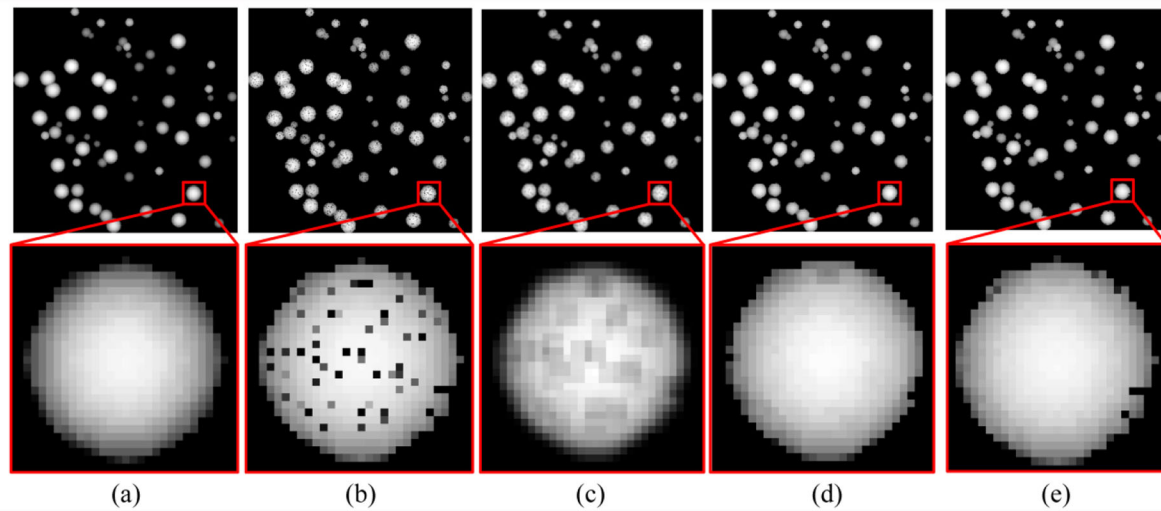
| Resolution | Mean   | Median | Ours   |
|------------|--------|--------|--------|
| 0.2        | 4.0868 | 3.4478 | 0.9674 |
| 0.5        | 4.4894 | 4.4414 | 0.9674 |
| 0.8        | 3.8001 | 3.3809 | 0.9674 |
| 1          | 4.7678 | 5.1041 | 0.9674 |
| 1.5        | 5.3144 | 5.0709 | 0.9674 |

## Discussion

One of the challenges in constructing a CHM from LiDAR point clouds is the existence of data pits. In this study, we propose a novel algorithm for constructing pit-free CHMs. In this algorithm, a pit-free raster CHM is directly constructed through a cloth simulation process instead of by the combination of various steps such as pit detection, pit filling and transformation from the randomly distributed point clouds to the raster CHM. The real and simulated data are used to validate the advantages of the proposed algorithm.

The visual comparison results show that the proposed algorithm is superior to the mean and median filters. Such filters generate new values for all pixels based on their neighbourhood values, thus affecting the original structure of the canopy surface (Ben-Arie et al. 2009; Shamsoddini et al. 2013). By contrast, our algorithm only fills data pits without changing the value of any other pixels, thereby preserving the original structure of the canopy surface. The proposed algorithm provides the possibility to minimise the distortion of the canopy's uppermost layer efficiently; this layer plays a critical role in various applications (e.g., ecological, hydrological and meteorological) sensitive to vegetation evolution at the local and regional scales (Khosravipour et al. 2014).

We also evaluate the proposed algorithm by comparing the accuracy of the tallest tree height estimation using the cloth simulation-based CHM versus the mean- and median-filtered CHMs. The positive biases for all algorithms indicate that the tallest tree height may be overestimated on the ground; this phenomenon is a possible error associated with the tree height measurement in the field (Wang et al. 2019). Our algorithm is significantly accurate at tree height estimation. The present study confirms previously published statements by Chen et al. (2017); these researchers illustrated that the mean and median filters struggle with the tallest tree height estimation because such filters smoothen the treetops based on surrounding pixels. The resolution of the cloth simulation-based CHM has no effect on the estimation accuracy of the maximum tree height. However, the estimation accuracy of the tallest tree height is sensitive to the resolutions of the mean- and median-filtered CHMs. The highest pixel value within a plot may be changed because the neighbourhood pixel values may vary with the resolution.

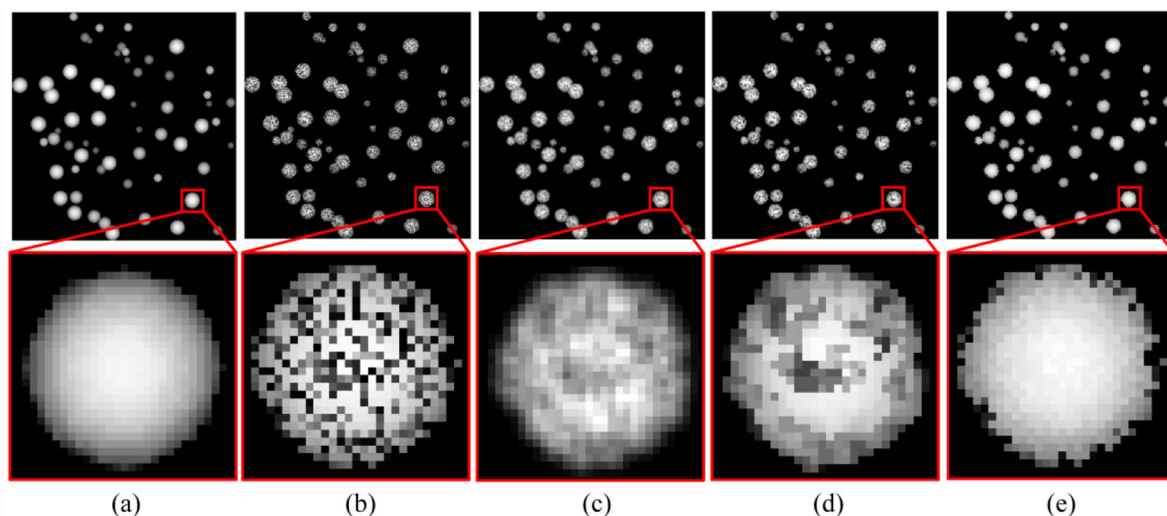


**Fig. 11** CHMs constructed through different algorithms using the data with 10% pits. **a** Reference CHM. **b** Raw CHM. **c** Mean-filtered CHM. **d** Median-filtered CHM. **e** Cloth simulation-based CHM

Furthermore, we investigate the effects of the different proportions of data pits on the accuracy of the mean and median filters and the proposed algorithm. The results show that our algorithm achieves the optimal visual performance amongst the data with different proportions of data pits. The quantitative assessment results show that the mean and median filters reduce the pit effect based on a predefined window kernel size that cannot be optimal for all different proportions of data pits. For example, the median filter can fill small pits but deteriorates large ones; this result agrees with those of

previous studies (Fig. 12d) (Shamsoddini et al. 2013). The proposed algorithm accurately represents the height of the canopy surfaces, especially for point clouds with high proportions of data pits. On this basis, our algorithm can be potentially applied to process the forest LiDAR point clouds acquired under leaf-off conditions given that most points will cluster on branches or trunks rather than canopy surfaces (i.e., high proportions of data pits).

In summary, the results demonstrate that the proposed algorithm has the following advantages: (a) The



**Fig. 12** CHMs constructed through different algorithms using the data with 50% pits. **a** Reference CHM. **b** Raw CHM. **c** Mean-filtered CHM. **d** Median-filtered CHM. **e** Cloth simulation-based CHM

**Table 3** Accuracy comparison between different algorithms for filling pits (unit: m)

| Landscapes | Pit proportion | Raw CHM | Mean   | Median | Our algorithm |
|------------|----------------|---------|--------|--------|---------------|
| Hemisphere | 10%            | 0.5428  | 0.4334 | 0.308  | <b>0.2031</b> |
|            | 20%            | 0.7562  | 0.5391 | 0.3753 | <b>0.2783</b> |
|            | 30%            | 0.9329  | 0.6652 | 0.5098 | <b>0.357</b>  |
|            | 40%            | 1.0685  | 0.787  | 0.6747 | <b>0.4248</b> |
|            | 50%            | 1.1948  | 0.9217 | 0.8752 | <b>0.4648</b> |
|            | 60%            | 1.2882  | 1.0299 | 1.0324 | <b>0.5209</b> |
|            | On average     | 0.9639  | 0.7294 | 0.6292 | <b>0.3748</b> |
| Cone       | 10%            | 1.304   | 0.7299 | 0.5077 | <b>0.4135</b> |
|            | 20%            | 1.8507  | 1.0731 | 0.7426 | <b>0.4814</b> |
|            | 30%            | 2.2675  | 1.4123 | 1.0503 | <b>0.5753</b> |
|            | 40%            | 2.6308  | 1.7823 | 1.5198 | <b>0.6553</b> |
|            | 50%            | 2.9588  | 2.1623 | 2.0756 | <b>0.7523</b> |
|            | 60%            | 3.1292  | 2.3931 | 2.4199 | <b>0.8503</b> |
|            | On average     | 2.3568  | 1.5922 | 1.386  | <b>0.6214</b> |

Note: The values in **bold** are the smallest ones in each line

proposed algorithm can effectively construct pit-free CHMs in point clouds with different proportions of data pits; (b) it can accurately fill only the pits without changing the values of all the CHM pixels, thereby preserving details of the uppermost layer of the forest canopy; and (c) it only needs to set up a CHM resolution parameter.

## Conclusion

In this study, we presented a new pit-free CHM construction algorithm based on cloth simulation. This algorithm was applied to the LiDAR data to generate raster CHM directly. The real and simulated datasets were used to compare the performance of the proposed algorithm with those of the mean and median filters. Overall, the proposed algorithm is promising in terms of robustness and high accuracy. This characteristic is useful for practical applications. Whether our pit-free CHM can improve the accuracy of extracted tree biophysical parameters, such as biomass, is interesting to investigate in the future.

## Abbreviations

3D: Three-dimensional; CHM: Canopy Height Model; DSM: Digital Surface Model; DTM: Digital Terrain Model; GPS: Global Positioning System; HD: Height Difference; LiDAR: Airborne Light Detection and Ranging; RMSE: Root Mean Square Error

## Acknowledgments

We thank Radiation Modelling and Measurement Laboratory (<http://ramm.bnu.edu.cn/>) for data provision. Jianbo Qi is thanked for his help with the figure preparation.

## Authors' contributions

WZ, SC and XL conceived the study and contributed to the manuscript writing and editing; JS coordinated the design of the program; RH processed a part of the data; SY participated the evaluation and figure preparation. GY

contributed to analyses of this study. All authors read and approved the final manuscript.

## Funding

This work was supported by the National Natural Science Foundation of China (Grant Nos. 41671414, 41971380 and 41171265) and the National Key Research and Development Program of China (No. 2016YFB0501404).

## Availability of data and materials

The data is not owned by the authors and cannot be shared.

## Ethics approval and consent to participate

The subject has no ethic risk.

## Consent for publication

All the data and relevant processing methods, including the automated generation of figures of evaluation results, belong to the host institution, namely, the Beijing Normal University.

## Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup>State Key Laboratory of Remote Sensing Science, Jointly Sponsored by Beijing Normal University and Institute of Remote Sensing and Digital Earth of Chinese Academy of Sciences, Beijing Engineering Research Center for Global Land Remote Sensing Products, Institute of Remote Sensing Science and Engineering, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China. <sup>2</sup>Department of Remote Sensing and Photogrammetry, Finnish Geospatial Research Institute, 02431 Masala, Finland. <sup>3</sup>College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China. <sup>4</sup>Cube Laboratory, UMR 7357 CNRS-University of Strasbourg, 300 bd Sebastien Brant, CS, 10413, F-67412 Illkirch Cedex, France. <sup>5</sup>Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100101, China. <sup>6</sup>University of Chinese Academy of Sciences, Beijing 100049, China.

Received: 23 April 2019 Accepted: 9 December 2019

Published online: 10 January 2020

## References

- Axelsson P (1999) Processing of laser scanner data—algorithms and applications. *ISPRS J Photogramm Remote Sens* 54:138–147 <https://www.sciencedirect.com/science/article/abs/pii/S0924271699000088>. Accessed 20 Apr 2019
- Ben-Arie JR, Hay GJ, Powers RP, Castilla G, St-Onge B (2009) Development of a pit filling algorithm for LiDAR canopy height models. *Comput Geosci* 35:1940–1949 <https://www.sciencedirect.com/science/article/pii/S0098300409000624>. Accessed 20 Apr 2019
- Cai S, Zhang W, Liang X, Wan P, Qi J, Yu S, Yan G, Shao J (2019) Filtering airborne LiDAR data through complementary cloth simulation and progressive TIN densification filters. *Remote Sens* 11:1037 <https://www.mdpi.com/2072-4292/11/9/1037/htm>.
- Chen C, Wang Y, Li Y, Yue T, Wang X (2017) Robust and parameter-free algorithm for constructing pit-free canopy height models. *ISPRS Int J Geo-Inf* 6:219 <https://www.mdpi.com/2220-9964/6/7/219>. Accessed 20 Apr 2019
- Chen Q, Baldocchi D, Gong P, Kelly M (2006) Isolating individual trees in a savanna woodland using small-footprint lidar data. *Photogramm Eng Remote Sens* 72:923–932 <https://www.ingentaconnect.com/content/asprs/pers/2006/00000072/00000008/art00003>. Accessed 20 Apr 2019
- Chow TE, Hodgson ME (2009) Effects of lidar post-spacing and DEM resolution to mean slope estimation. *Int J Geogr Inf Sci* 23:1277–1295 <https://www.tandfonline.com/doi/abs/10.1080/13658810802344127>. Accessed 20 Apr 2019
- Coomes DA, Šafka D, Shepherd J, Dalponte M, Holdaway R (2018) Airborne laser scanning of natural forests in New Zealand reveals the influences of wind on forest carbon. *Forest Ecosyst* 5:10. <https://doi.org/10.1186/s40663-017-0119-6>
- Dong P (2009) Characterization of individual tree crowns using three-dimensional shape signatures derived from LiDAR data. *Int J Remote Sens* 30:6621–6628 <https://www.tandfonline.com/doi/abs/10.1080/01431160903140761>. Accessed 20 Apr 2019
- Gaveau DLA, Hill RA (2003) Quantifying canopy height underestimation by laser pulse penetration in small-footprint airborne laser scanning data. *Can J*

- Remote Sens 29:650–657 <https://www.tandfonline.com/doi/abs/10.5589/m03-023>. Accessed 20 Apr 2019
- Goulden T, Hopkinson C (2010) The forward propagation of integrated system component errors within airborne lidar data. *Photogramm Eng Remote Sens* 76:589–601 <https://www.ingentaconnect.com/content/asprs/pers/2010/00000076/00000005/art00005>. Accessed 20 Apr 2019
- Huang HG, Lian J (2015) A 3D approach to reconstruct continuous optical images using lidar and MODIS. *Forest Ecosyst* 2:20. <https://doi.org/10.1186/s40663-015-0044-5>
- Hyypä J, Hyypä H, Leckie D, Gougeon F, Yu X, Maltamo M (2008) Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. *Int J Remote Sens* 29:1339–1366 <https://www.tandfonline.com/doi/abs/10.1080/01431160701736489>. Accessed 20 Apr 2019
- Isenburg M (2017) LAStools—efficient lidar processing software. <http://rapidlasso.com/LAStools>
- Khosravipour A, Skidmore AK, Isenburg M, Wang T, Hussin YA (2014) Generating pit-free canopy height models from airborne lidar. *Photogramm Eng Remote Sens* 80:863–872 <https://www.ingentaconnect.com/content/asprs/pers/2014/00000080/00000009/art00003>. Accessed 20 Apr 2019
- Kraus K, Pfeifer N (1998) Determination of terrain models in wooded areas with airborne laser scanner data. *ISPRS J Photogramm Remote Sens* 53:193–203 <https://www.sciencedirect.com/science/article/abs/pii/S0924271698000094>. Accessed 20 Apr 2019
- Latypov D (2005) Effects of laser beam alignment tolerance on lidar accuracy. *ISPRS J Photogramm Remote Sens* 59:361–368 <https://www.sciencedirect.com/science/article/abs/pii/S0924271605000584>. Accessed 20 Apr 2019
- Leckie D, Gougeon F, Hill D, Quinn R, Armstrong L, Shreenan R (2003) Combined high-density lidar and multispectral imagery for individual tree crown analysis. *Can J Remote Sens* 29:633–649 <https://www.tandfonline.com/doi/abs/10.5589/m03-024>. Accessed 20 Apr 2019
- Lim K, Treitz P, Wulder M, St-Onge B, Flood M (2003) LiDAR remote sensing of forest. *Prog Phys Geogr* 27:88–106 <https://journals.sagepub.com/doi/abs/10.1191/0309133303pp360ra>. Accessed 20 Apr 2019
- Mielczarek M, Stereńczak K, Khosravipour A (2018) Testing and evaluating different LiDAR-derived canopy height model generation methods for tree height estimation. *Int J Appl Earth Obs Geoinf* 71:132–143 <https://www.sciencedirect.com/science/article/pii/S0303243418301478>. Accessed 20 Apr 2019
- Nyquist H (1928) Certain topics in telegraph transmission theory. *Trans Am Inst Electr Eng* 47:617–644 <https://ieeexplore.ieee.org/abstract/document/5055024>. Accessed 20 Apr 2019
- Popescu SC (2007) Estimating biomass of individual pine trees using airborne lidar. *Biomass Bioenergy* 31:646–655 <https://www.sciencedirect.com/science/article/pii/S0961953407001316>. Accessed 20 Apr 2019
- Shamsoddini A, Turner R, Trinder JC (2013) Improving lidar-based forest structure mapping with crown-level pit removal. *J Spat Sci* 58:29–51 <https://www.tandfonline.com/doi/abs/10.1080/14498596.2012.759092>. Accessed 20 Apr 2019
- Stereńczak K, Lisańczuk M, Erfanifard Y (2018) Delineation of homogeneous forest patches using combination of field measurements and LiDAR point clouds as a reliable reference for evaluation of low resolution global satellite data. *Forest Ecosyst* 5:1. <https://doi.org/10.1186/s40663-017-0128-5>
- Tang X (2013) Estimation of forest aboveground biomass by integrating ICESat/GLAS waveform and TM data. Doctoral Dissertation, University of Chinese Academy of Sciences, Beijing [http://www.wanfangdata.com.cn/details/detail.do?\\_type=degree&id=Y2440193](http://www.wanfangdata.com.cn/details/detail.do?_type=degree&id=Y2440193). Accessed 20 Apr 2019
- Wan P, Wang T, Zhang W, Liang X, Skidmore A, Yan G (2019) Quantification of occlusions influencing the tree stem curve retrieving from single-scan terrestrial laser scanning data. *Forest Ecosyst* 6:43 <https://link.springer.com/content/pdf/10.1186/s40663-019-0203-1.pdf>
- Wang Y, Lehtomäki M, Liang X, Pyörälä J, Kukko A, Jaakkola A, Liu J, Feng Z, Chen R, Hyypä J (2019) Is field-measured tree height as reliable as believed—a comparison study of tree height estimates from field measurement, airborne laser scanning and terrestrial laser scanning in a boreal forest. *ISPRS J Photogramm Remote Sens* 147:132–145 <https://www.sciencedirect.com/science/article/pii/S0924271618303046>. Accessed 20 Apr 2019
- Yu X, Hyypä J, Vastaranta M, Holopainen M, Viitala R (2011) Predicting individual tree attributes from airborne laser point clouds based on the random forests technique. *ISPRS J Photogramm Remote Sens* 66:28–37 <https://www.sciencedirect.com/science/article/abs/pii/S0924271610000651>. Accessed 20 Apr 2019
- Zhang K, Chen SC, Whitman D, Shyu ML, Yan J, Zhang C (2003) A progressive morphological filter for removing nonground measurements from airborne lidar data. *IEEE Trans Geosci Remote Sens* 41:872–882 <https://ieeexplore.ieee.org/abstract/document/1202973>. Accessed 20 Apr 2019
- Zhang W, Qi J, Wan P, Wang H, Xie D, Wang X, Yan G (2016) An easy-to-use airborne LiDAR data filtering method based on cloth simulation. *Remote Sens* 8:501 <https://www.mdpi.com/2072-4292/8/6/501>. Accessed 20 Apr 2019
- Zhang W, Wan P, Wang T, Cai S, Chen Y, Jin X, Yan G (2019) A novel approach for the detection of standing tree stems from plot-level terrestrial laser scanning data. *Remote Sens* 11:211 <https://www.mdpi.com/2072-4292/11/2/211>
- Zhao D, Pang Y, Li Z, Sun G (2013) Filling invalid values in a lidar-derived canopy height model with morphological crown control. *Int J Remote Sens* 34:4636–4654 <https://www.tandfonline.com/doi/abs/10.1080/01431161.2013.779398>. Accessed 20 Apr 2019

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)