Robust and Parameter-Free Algorithm for Constructing Pit-Free Canopy Height Models

Chuanfa Chen [1,2,*] (ID)**, Yifu Wang** [3]**, Yanyan Li** [2]**, Tianxiang Yue** [3] **and Xin Wang** [2]

[1] State Key Laboratory of Mining Disaster Prevention and Control Co-Founded by Shandong Province and the Ministry of Science and Technology, Shandong University of Science and Technology, Qingdao 266590, China

[2] Shandong Provincial Key Laboratory of Geomatics and Digital Technology of Shandong Province, Shandong University of Science and Technology, Qingdao 266590, China; yylee@whu.edu.cn (Y.L.); wxin_xin@sina.cn (X.W.)

[3] State Key Laboratory of Resources and Environment Information System, Institute of Geographical Science and Natural Resources Research, Chinese Academy of Sciences, 11A, Datun Road, Beijing 100101, China; wangyf@lreis.ac.cn (Y.W.); yue@lreis.ac.cn (T.Y.)

* Correspondence: chencf@lreis.ac.cn; Tel.: +86-0532-8605-7845

Abstract: Data pits commonly appear in lidar-derived canopy height models (CHMs) owing to the penetration ability of airborne light detection and ranging (lidar) into tree crowns. They have a seriously negative effect on the quality of tree detection and subsequent biophysical measurements. In this study, we propose an algorithm based on robust locally weighted regression and robust z-scores for the construction of a pit-free CHM. A significant advantage of the new algorithm is that it is parameter free, which makes it efficient and robust for practical applications. Simulated and airborne lidar-derived data sets are employed to assess the performance of the new method for CHM construction, and its results are compared to those of three classical methods, namely the natural neighbor (NN) interpolation of the highest point method (HPM), mean filter, and median filter. The results from the simulated data set demonstrate that our algorithm is more accurate compared to the three classical methods for generating pit-free CHMs in the presence of data pits. CHM construction using the lidar-derived data set shows that, compared to the classical methods, the new method has a better ability to remove data pits as well as preserving the edges, shapes, and structures of canopy gaps and crowns. Moreover, the proposed method performs better compared to the classical methods in deriving plot-level maximum tree heights from CHMs. Thus, the new method shows high potential for pit-free CHM construction.

Keywords: canopy height model; data pit; tree crown; robust fitting

1. Introduction

In addition to characterizing ground topography, airborne light detection and ranging (lidar) data have been widely used in forestry in recent years [1,2]. Lidar-derived canopy height models (CHMs), which represent the absolute height distribution of the vegetation canopy above the ground [3,4], have a significant influence on the extraction of forest inventory information, such as individual tree crown delineation [5–7], tree height estimation [8–10], and biomass estimation [11,12].

A CHM is typically constructed by interpolating the first return lidar points and determining their heights above a digital elevation model (DEM) [13]. Data pits with heights significantly lower than those of their neighbors commonly appear in raster CHMs. According to Leckie et al. [14], data pits appear due to the combination of different flight line data sets and the penetration of the laser beam into canopy branches and foliage before its first return. Data pits may even appear

in individual flight lines at off-nadir scan angles. Ben-Arie et al. [15] identified that data pits may be caused by the combination of various factors such as data acquisition and post-processing. Vosselman [16] indicated that the planimetric error between overlapping flight strips caused by the global positioning system (GPS) and inertial measurement unit (IMU) measurements can also generate data pits, particularly for small-footprint lidar. Many algorithms have been developed to generate pit-free CHMs. The performance of the present algorithms is often subject to different parameters that should be optimally tuned. However, the process of parameter tuning is not easy in practice. Thus, parameter-free algorithms are more desirable.

In a statistical sense, data pits in a CHM can be considered outliers because the heights of data pits are significantly lower than those of their neighbors. Thus, the problems of data pits can be remedied by outlier detection techniques. We propose a robust statistical method based on this idea to detect data pits. The new method uses robust locally weighted regression to perform interpolations and robust z-scores to detect data pits. After the removal of data pits, the remaining lidar points are interpolated using the natural neighbor (NN) method to construct a pit-free CHM. NN was selected because it is parameter free, easy to apply, and highly accurate for interpolating lidar points [17]. Compared to the classical pit-free algorithms, the main novelties of the new method are as follows: (i) it is parameter-free; and (ii) it is applied on the height-normalized lidar point clouds rather than the raster CHMs, thus avoiding errors in the transformation from the randomly distributed point clouds to the rasterized CHM.

The remainder of this paper is organized as follows. Section 2 presents related work. Section 3 elaborates the principle of the new method. In Section 4, a numerical test and a real-world example are employed to assess the performance of our algorithm, and its results are compared with the results of three classical methods, that is, NN interpolation of the highest point method (HPM), mean filter, and median filter. It should be noted that the mean and median filters use a 3×3 kernel in the tests. Discussion and conclusions are presented in Section 5.

2. Related Work

It has been demonstrated that data pits have a seriously negative effect on the estimation of forest parameters. For example, Persson et al. [9] found that data pits in a CHM complicate the detection of tree crowns. Gaveau and Hill [18] demonstrated that the penetration of laser beams into a tree crown leads to the underestimation of canopy height. Khosravipour et al. [19] indicated that in the context of treetop detection, data pits in CHMs lead to large commission and omission errors. Thus, a variety of algorithms have been developed to construct pit-free CHMs. For example, Leckie et al. [14] presented an algorithm that first assigns all lidar points to 25-cm grid cells, and then employs the finite difference method to interpolate the highest point in each grid. Ben-Arie et al. [15] presented a semiautomatic pit-filling algorithm to detect and replace data pits using Laplacian and median filters. This method was further developed by Zhao et al. [20]. They used morphological crown control to recover crown coverage. Shamsoddini et al. [21] employed an adaptive mean filter with 3×3 and 5×5 kernels to detect and fill data pits. Liu and Dong [22] used a selecting and sorting scheme to select data points during CHM construction.

Although data pits can be removed using the aforementioned image smoothing methods, the heights of all pixels in the CHMs are altered. This can cause the underestimation of tree heights and crown diameters due to the omission of treetops and the reduction of crown shoulders [19]. Thus, Khosravipour et al. [19] proposed a pit-free CHM algorithm by generating a stack of CHMs that were partially pit free—each representing a different height interval of the canopy—and finally merged into a pit-free CHM. However, the performance of the pit-free CHM algorithm is largely determined by a rasterization threshold and a series of height thresholds. This makes it difficult to apply. Thus, a parameter-free algorithm is more practical.

3. Principle of the Proposed Algorithm

The flowchart of the proposed algorithm for pit-free CHM construction is shown in Figure 1. As can be seen, the main steps are robust locally weighted regression (RLWR) and pit removal using robust z-scores. As in the case of outliers in statistics, data pits can be defined as the observations having large z-scores with respect to interpolation errors. In the presence of data pits, a robust interpolator should be used to interpolate lidar points. Here, RLWR is adopted due to its simplicity, robustness, and high accuracy with densely distributed data sets [23]. Then, observations with robust z-scores exceeding a tolerance (i.e., 2.5) are flagged as data pits.
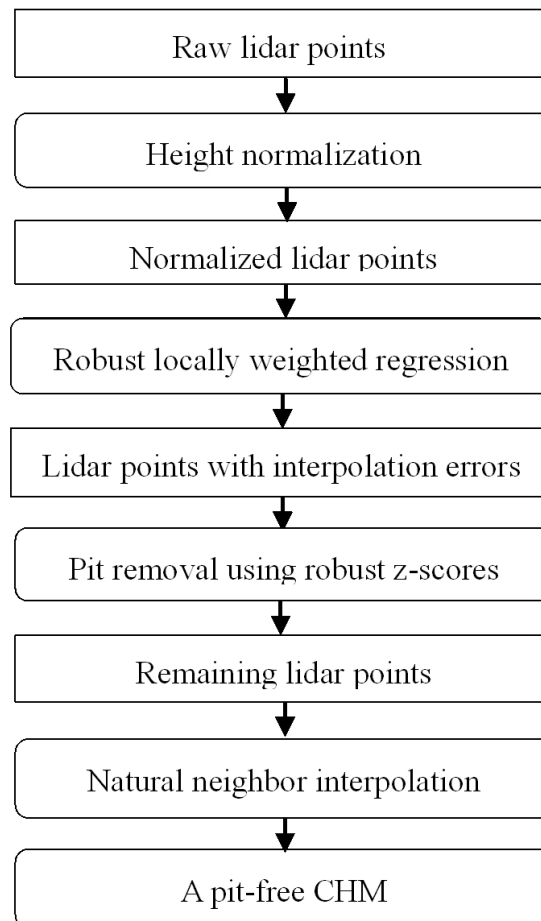

Figure 1. Flowchart of the proposed algorithm for pit-free canopy height model (CHM) construction.

3.1. RLWR

Locally weighted regression (LWR), originally proposed by Cleveland [24] and further developed by Cleveland and Devlin [25], has been widely used for function estimation. Compared to global polynomial fitting, LWR has many desirable statistical properties, such as numerical stability, minimax efficiency, and absence of boundary effects [26]. Moreover, in contrast to geostatistical interpolation methods, LWR does not require the specification of a semivariogram, which should be determined beforehand.

Given a group of 3D points $(x_i, z_i), i = 1, 2, \cdots, n, \mathrm{x} \in R^2$, the data is supposed to be fitted with the model as follows:

$$z_i = g(\mathrm{x}_i) + \varepsilon_i \tag{1}$$

where ε_i represents sample errors with zero mean and unknown variance and $g(x)$ is a smooth function to be estimated. In this study, x and z represent the horizontal coordinate and normalized height, respectively, of a lidar point.

When performing LWR, a low-degree polynomial is first fitted to the neighbors of x_i using a weighted least square algorithm, where the weight of each point is obtained from a weight function. Generally, the weight function of LWR assigns high weights to points near the point of interest and low weights to points far away. This is based on the idea that points near each other are more likely to be related than points that are further apart. Cleveland [24] suggested using the tri-cube weight function,

$$w(x_j) = \begin{cases} \left(1 - \left(\dfrac{d_{ij}}{d_{\max}}\right)^3\right)^3 & x_j \in N(x_i) \\ 0 & x_j \notin N(x_i) \end{cases} \tag{2}$$

where $N(x_i)$ is the set of the k nearest neighbors of x_i in 2D space, d_{ij} is the distance between x_i and its j-th neighbor, and $d_{\max} = \max(d_{ij}, j = 1, 2, \dots, k)$ is the maximum of $d_{ij}, j = 1, 2, \dots, k$.

The objective function of LWR is defined as

$$\min \sum_{x_j \in N(x_i)} w(x_j)\left(z_j - g(x_j)\right)^2 \tag{3}$$

Thus, the function value at x_i can be obtained by evaluating the local polynomial with the polynomial coefficients estimated from the solution of Equation (3). The commonly used polynomial is either linear or quadric. This is based on the idea that any function can be accurately and easily approximated in a small neighborhood by a low-order polynomial. High-degree polynomials are prone to overfitting the data in each subset and are numerically unstable [26].

LWR is highly sensitive to outliers due to the least squares -based objective function. Thus, if data pits exist in the neighborhood of the point of interest, LWR may lead to a biased estimation. Therefore, to reduce the influence of outliers, a robust LWR (RLWR) is adopted that assigns a robust weight to each point in the neighborhood. Cleveland [24] employed the bisquare weight function to assure the robustness of RLWR. The bisquare weight function is defined as follows

$$B(r_j) = \begin{cases} \left(1 - \left(\dfrac{r_j}{6s}\right)^2\right)^2 & \left|\dfrac{r_j}{6s}\right| < 1 \\ 0 & \left|\dfrac{r_j}{6s}\right| \geq 1 \end{cases} \tag{4}$$

where r_j is the regression residual of the j-th nearest neighbor of the point of interest, $r_j = z_j - g(x_j)$, and s is the median of $|r_i|$. Thus, the objective function of RLWR is expressed as

$$\min \sum_{x_j \in N(x_i)} B(r_j) w(x_j)\left(z_j - g(x_j)\right)^2 \tag{5}$$

Therefore, the estimation of $g(x_i)$ at x_i by RLWR involves the following steps:

(i) Finding the k nearest neighbors $N(x_i)$ of x_i from the point cloud. Here, we set $k = 12$.
(ii) Computing the weights of $x_j, j = 1, 2, \cdots, k$ based on the distance weight function, i.e., Equation (2), where $x_j \in N(x_i)$.
(iii) Estimating the polynomial coefficients by minimizing the objective function of LWR, i.e., Equation (3). Here, the linear polynomial is used.
(iv) Estimating $g(x_j)$ at $x_j, j = 1, 2, \cdots, k$ with the computed polynomial coefficients.
(v) Computing the regression residuals for $x_j, j = 1, 2, \cdots, k$, i.e., $r_j = z_j - g(x_j)$.

(vi)　Computing the robust weights of $x_j, j = 1, 2, \cdots, k$ based on the robust weight function, i.e., Equation (4).

(vii)　Estimating the polynomial coefficients by minimizing the objective function of RLWR, i.e., Equation (5).

(viii) Repeating (v)–(vii) until the polynomial coefficients are stable, or the maximum number of iteration is reached. Cleveland [24] indicated that two iterations are sufficient to obtain a good fit.

(ix)　Estimating $g(x_i)$ at x_i with the computed polynomial coefficients. The interpolation error of x_i is expressed as, $e_i = z_i - g(x_i)$.

3.2. Pit Detection Using Robust Z-Scores

As in the case of outlier detection in statistics, the well-known z-score with respect to interpolation error is employed to detect data pits in this study. Z-score is a distance-based measure that is defined as the standardized residual [27],

$$t_i = \frac{e_i - \bar{e}}{\sigma_e}, i = 1, 2, \cdots \tag{6}$$

where $\bar{e}$ and σ_e denote the mean and standard deviation, respectively, of e. Both $\bar{e}$ and σ_e have a breakdown point (BP) of zero and are sensitive to outliers. Their robust alternatives are the median and median absolute deviation, respectively, both of which have a BP of 50% [27]. The robust z-score is expressed as

$$\hat{t}_i = \frac{e_i - e_{\mathrm{median}}}{\sigma_{\mathrm{MAD}}}, i = 1, 2, \cdots \tag{7}$$

where e_{median} and σ_{MAD} represent the median and median absolute deviation, respectively, of e. In statistics, observations with $|\hat{t}_i|$ exceeding 2.5 are regarded as outliers. The threshold value of 2.5 is derived from the fact that the proportion of a random variable with a standard normal distribution exceeding 2.5 is less than 1.24%. The threshold value of 2.5 for a criterion to detect the data-pits as outliers is derived from a statistical viewpoint rather than the forest types. Therefore, this threshold is suitable for any forest type. Because the heights of data pits are always lower than those of their neighbors, we define data pits as the observations with $\hat{t}_i$ less than -2.5.

4. Experiments

We use a numerical test and a real-world example to assess the robustness of our method for CHM construction. Furthermore, plot-level maximum tree heights derived from CHMs are used to indirectly evaluate the performance of the proposed method in the real-world example.

4.1. Numerical Test

Two simple geometric models using cones and hemispheres [28] are employed to simulate tree crowns. The mathematical formulations of the two models are expressed as:

$$\mathrm{Cone} : z = -\sqrt{x^2 + y^2}/\tan 30°, x^2 + y^2 \leq 1 \tag{8}$$

$$\mathrm{Hemisphere} : z = \sqrt{1 - x^2 - y^2} - 1, x^2 + y^2 \leq 1 \tag{9}$$

The procedure for comparing the performance of our method with that of the present methods is as follows:

(i)　1000 points subject to the condition of $x^2 + y^2 \leq 1$ are randomly sampled from the model, e.g., cone. The average point space is 0.056.

(ii)　n data pits are randomly selected and their heights are artificially changed, i.e., their elevations are reduced by 0.3. Here, $n = 100$ or 200; the contaminating proportion (α) is 10% or 20%.

(iii) Our method and the three other pit-free methods (mean filter, median filter, and HPM) are used to construct CHMs with the contaminated data set.

(iv) Taking the original sample points as check points, the CHMs are assessed in terms of root mean square error (RMSE) and mean error (ME). These are expressed as:

$$\text{RMSE} = \sqrt{\frac{\sum\limits_{i=1}^{m}(z_i - \hat{z}_i)^2}{m}} \tag{10}$$

$$\text{ME} = \frac{\sum\limits_{i=1}^{m}(z_i - \hat{z}_i)}{m} \tag{11}$$

where z_i and $\hat{z}_i$ are the true and simulated values, respectively, at the i-th check point and m is the number of check points, i.e., $m = 1000$.

For mean and median filters, a raw CHM was first constructed using NN interpolation of the contaminated data set. Then, the CHM was smoothed by the two filters. For HPM, the highest point in each grid cell with a resolution of 0.056 was extracted, and then, NN was used to construct a CHM with the highest points.

The results (Table 1) indicate that, irrespective of accuracy measures, NN produces the poorest CHM with the contaminated data sets, demonstrating that data pits have a significantly negative effect on the quality of CHMs. The performance of all pit-free algorithms decreases as the contaminating proportion increases. Our method yields the best results on all simulated data sets. On average, the new method is ~2.7, 1.7, 1.8, and 2.1 times as accurate as the raw CHM, median filter, mean filter, and HPM, respectively, in terms of RMSE. In terms of ME, the corresponding figures are 31.3, 25.6, 31, and 16.6.

Table 1. Accuracy comparison between our method, raw CHM, and three other methods for removing data pits under different contaminating proportions (α) in the numerical test.

Method	Accuracy Measure	Cone		Hemisphere		On Average
		$\alpha = 10\%$	$\alpha = 20\%$	$\alpha = 10\%$	$\alpha = 20\%$	
Raw CHM	RMSE	0.0401	0.0566	0.0642	0.0821	0.0608
	ME	0.0292	0.0554	0.0243	0.0651	0.0435
HPM	RMSE	0.0365	0.0502	0.0461	0.0575	0.0476
	ME	0.0173	0.0351	0.0133	0.0276	0.0233
Mean	RMSE	0.0259	0.0362	0.0508	0.0471	0.0400
	ME	0.0292	0.0553	0.0265	0.0622	0.0433
Median	RMSE	0.0242	0.0354	0.0485	0.0470	0.0388
	ME	0.0269	0.0516	0.0160	0.0493	0.0360
Our method	RMSE	0.0130	0.0144	0.0303	0.0322	0.0225
	ME	−0.0015	−0.0018	−0.0015	−0.0006	−0.0014

Figure 2 shows CHMs of the cone model constructed by different methods with the original points and contaminated points under $\alpha = 10\%$. The raw CHM has many data pits randomly distributed in the image (Figure 2b). Among the pit-free algorithms, HPM yields the poorest result, because many data pits are clearly visible (Figure 2c). This is because some of the highest points still contain data pits. Even though the mean filter removes data pits, it produces smoothed square artifacts (Figure 2d). Both the median filter and our method are successful in removing data pits (Figure 2e,f). This is expected, as the two methods have high breakdown values, especially the median filter with a value of 50%. However, both filters alter the values of all CHM cells; this results in a slight distortion of the

crown shape in the right-bottom (Figure 2e). The proposed method (Figure 2f) yields the best result, which closely approximates the original (Figure 2a).
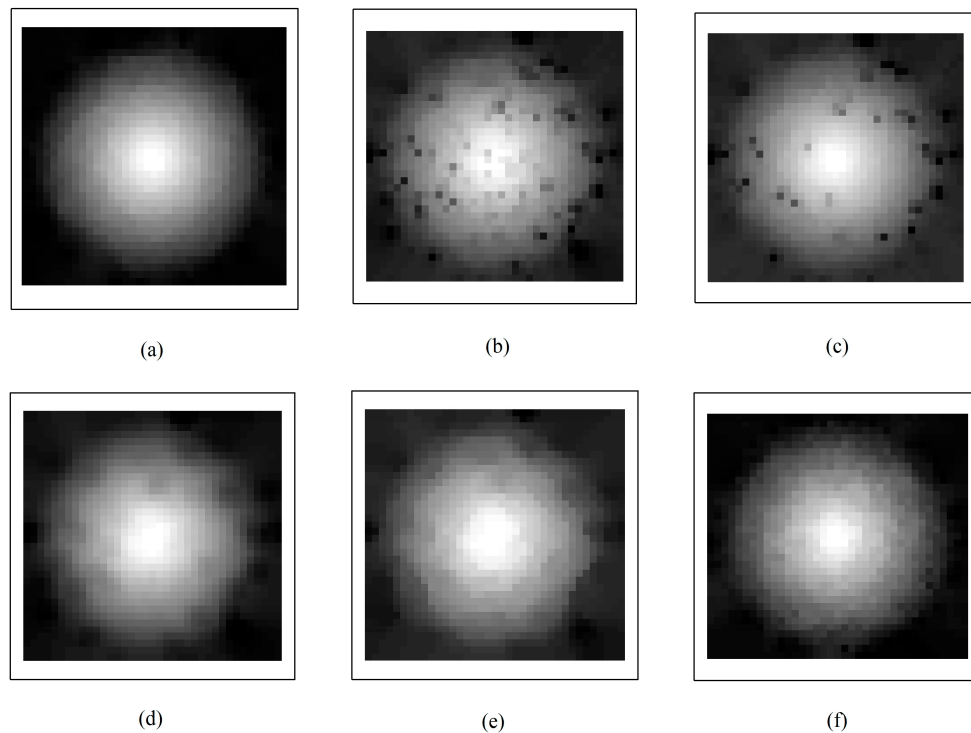

Figure 2. CHMs produced by (**a**) natural neighbor (NN) interpolation of original sample points; (**b**) NN interpolation of the contaminated data set with the contaminating proportion of 10%; (**c**) highest point method (HPM); (**d**) mean filter; (**e**) median filter; and (**f**) our method.

4.2. Real-World Example

4.2.1. Study Site and Raw Data Sets

The study site chosen was the Tianlaochi catchment (38°23′55″–38°26′57″ N, 99°53′45″–99°57′12″ E), Gansu Province, China. It is characterized by a cold semi-arid climate with annual average temperature and precipitation of 0.6 °C and 437.2 mm, respectively. The site was mainly covered by coniferous plantation forests and shrubs. The forest vegetation types included picea crassifolia and sabina przewalskii. A mean elevation of 3292 m, with a minimum of 2596 m and a maximum of 4411 m, and a mean slope of 29.6° characterized the topography of the study area.

A total of 30 rectangular plots of area 20×10 m^2 were chosen in the study site in August 2013 (Figure 3). Among the 30 plots, 26 were located in picea crassifolia stands, two in sabina przewalskii stands and two in stands of a mixture of the two species. All tree heights in each plot were measured by a handheld laser rangefinder (Haglof Vertex IV Ultrasonic Hypsometer made by HAGLOF INC), and the maximum tree height was recorded.

Raw lidar data for the study site were acquired using Leica ALS70 with a laser wavelength of 1064 nm in July 2012. During data capture, the absolute flying height was about 4800 m with a mean data density of 1 point/m^2. The lidar system recorded up to four returns for each laser pulse, depending on the ground cover. The TerraScan (http://terrasolid.fi) software, produced by Terrasolid Ltd., Helsinki, Finland, was employed to differentiate ground and non-ground points.

The resolution of raster CHMs plays an important role in estimating individual tree attributes [29]. The CHM resolution should not be larger than half the minimum crown size, and overly smaller than

the mean point space [19,30]. Thus, a CHM resolution of 0.5 m was determined by recognizing an average crown diameter of 1.8 m and an average point space of 1 m.
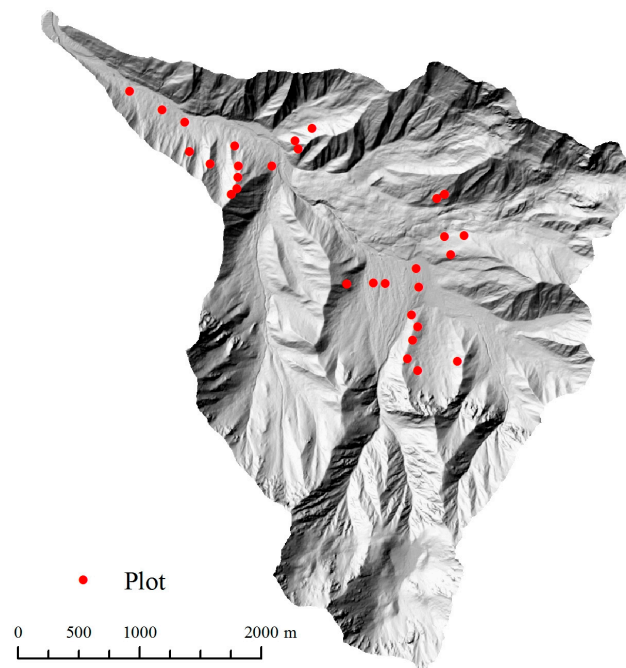

Figure 3. Distribution of plots in Tianlaochi catchment, Gansu Province, China.

4.2.2. Data Processing

The non-ground points were first height-normalized by replacing the elevation of each point with its height above the ground. Then, the four pit-free methods were employed to construct CHMs with the normalized lidar data. Since tree height underestimation is one of the major concerns for CHMs in forest areas [22], plot-level maximum tree heights were derived from the CHMs and compared with field measurements. This provides a comparison between our method and the other algorithms [3].

4.2.3. Results

Taking one plot as an example, Figure 4 shows the 0.5-m CHMs produced by different methods. The raw CHMs made by the NN interpolation are seriously influenced by data pits (Figure 4a). The CHM of the HPM appears visually similar to the original, indicating the poor performance of HPM in terms of the removal of data pits (Figure 4b). Mean and median filters (Figure 4c,d) eliminate the data pits; however, their CHMs are heavily blurred compared to the original. Our method performs better than the others for removing data pits, while preserving the edges, shapes, and structures of canopy gaps and crowns (Figure 4e).

Figure 5 shows the relationship between the measured and simulated plot-level maximum tree heights for the 30 plots. For the mean and median filters (Figure 5a,b), the linear models explain ~66.3% of the variation of plot-level maximum tree heights, whereas HPM (Figure 5c) and our method (Figure 5d) explain ~71% and 73%, respectively. The new method is more accurate compared to the classical methods, because the points of our method approximate the line $y = x$ more closely than those of the other methods.

Figure 6 shows the relationship between the measured plot-level maximum tree heights and simulated residuals, in other words, the difference between measured and simulated plot-level maximum tree heights, for the 30 plots. It can be seen that mean and median filters (Figure 6a,b) underestimate the maximum tree heights in almost all plots. HPM (Figure 6c) and our method

(Figure 6d) seem significantly better than median and mean filters. In comparison, the simulated residuals of our method are more prone to zero than those of the other methods.
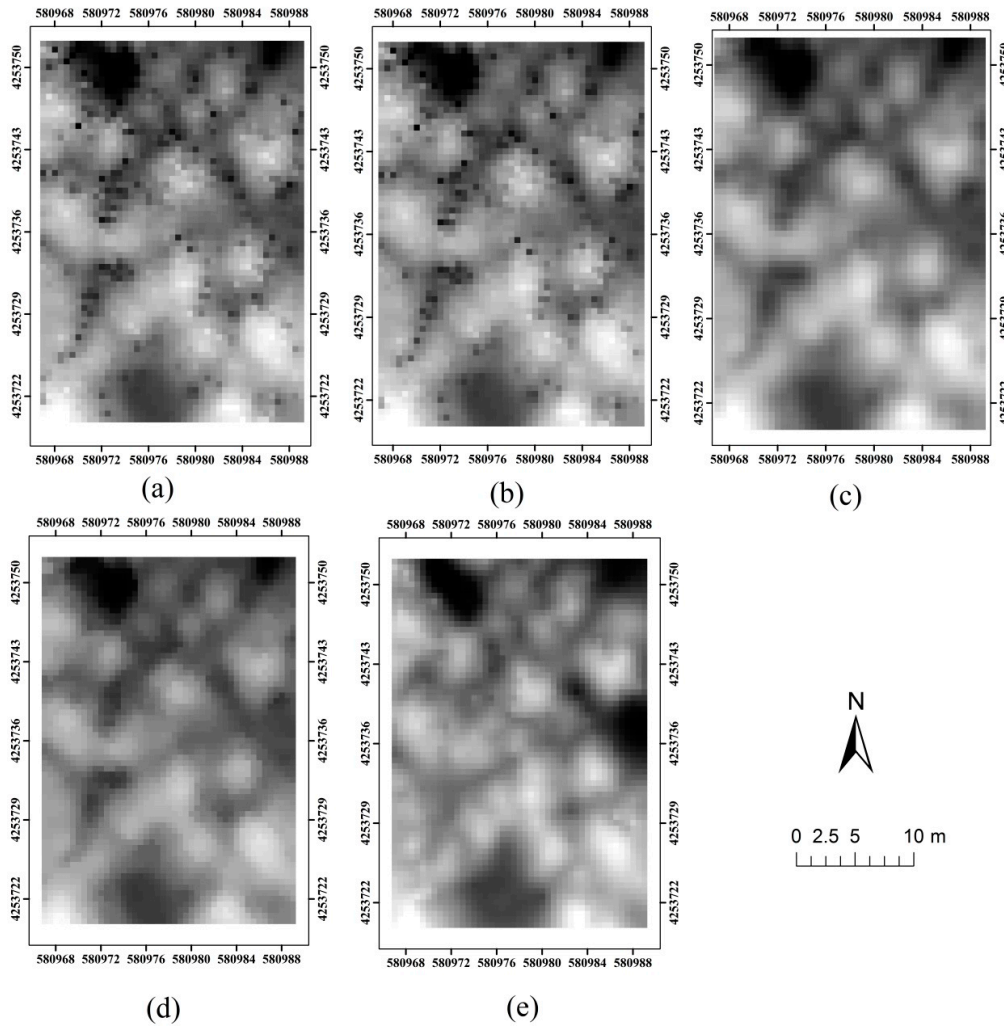

Figure 4. CHMs produced by (**a**) NN interpolation of raw light detection and ranging (lidar) points; (**b**) HPM; (**c**) mean filter; (**d**) median filter; and (**e**) our method.
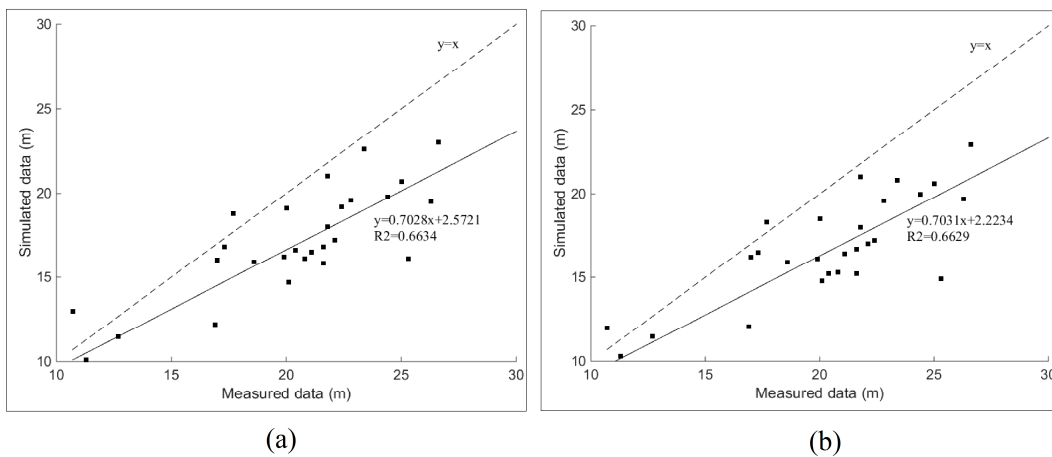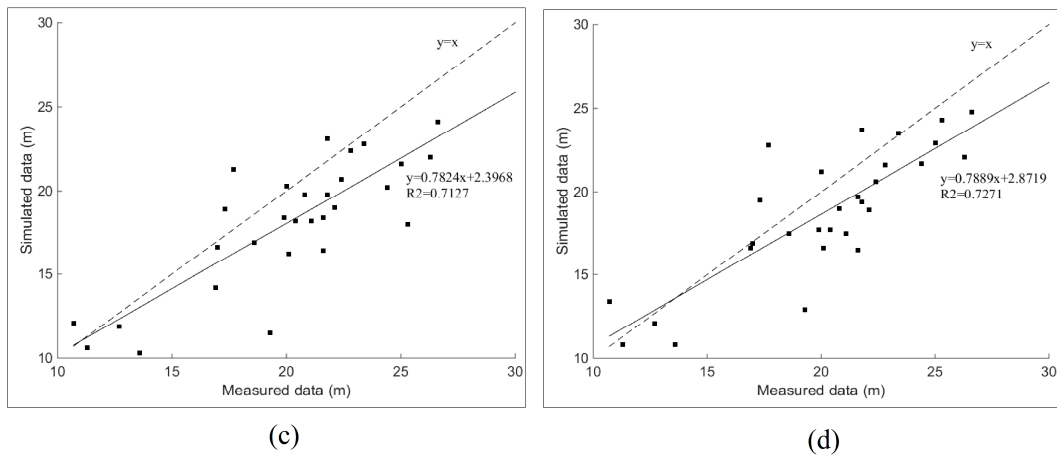

Figure 5. *Cont.*

Figure 5. Relationship between the measured and simulated plot-level maximum tree heights for the 30 plots. (**a**) Mean filter; (**b**) median filter; (**c**) HPM; and (**d**) our method.
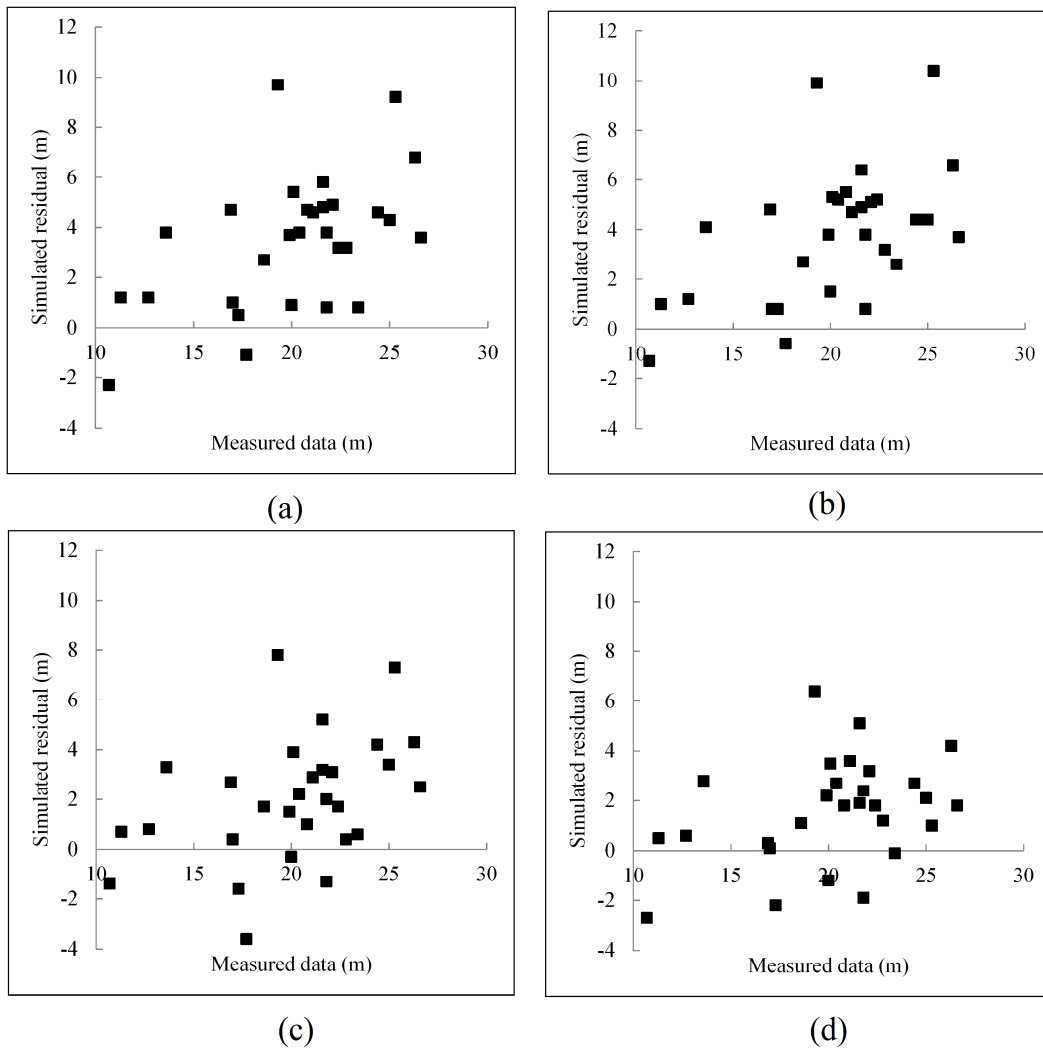

Figure 6. Relationship between the measured plot-level maximum tree heights and simulated residuals, i.e., difference between measured and simulated plot-level maximum tree height, for the 30 plots. (**a**) Mean filter; (**b**) median filter; (**c**) HPM; and (**d**) our method.

Table 2 quantitatively shows a comparison between the four methods of the accuracy in the estimation of the plot-level maximum tree heights in the 30 plots. All methods have the defect of tree height underestimation, and our method has a lower systematic error in terms of mean error (ME) than that of the classical methods. Additionally, the new method has the lowest-level variation in terms of standard deviation, and its maximum error is significantly smaller than that of the other methods.

Table 2. Accuracy comparison between the four methods (unit: m). ME, Std, and Max represent mean error, standard deviation, and maximum error, respectively, between the measured and simulated plot-level maximum tree heights in the 30 plots.

	HPM	Mean	Median	Our Method
ME	2.1	3.5	3.9	1.4
Std	2.5	2.7	2.7	2.1
Max	7.8	9.7	10.4	6.4

5. Discussion and Conclusions

5.1. Discussion

Data pits and tree height underestimation are two major problems of lidar-derived CHMs. To overcome the two problems, we proposed a robust algorithm based on locally weighted regression (LWR) and z-scores to remove data pits. Rather than removing data pits from CHM rasters, our method operated on the height-normalized lidar points in 3D space. Thus, it avoided the alteration of CHM values, and accurately preserved CHM contrast. Both the numerical test and the real-world example validated the advantages of the proposed method.

In recent years, many algorithms have been developed to generate pit-free CHMs. However, they commonly suffer from the necessity of selecting appropriate parameters. For example, the method of Khosravipour et al. [19] required a series of height thresholds and a rasterization threshold for constructing a number of partial CHMs. Liu and Dong [22] showed that a percentage threshold should be set for selecting non-ground points for CHM construction. Rizaev et al. [31] suggested choosing a proper moving window size for local minimum identification. However, it is non-trivial to tune the optimal parameters in practice. In contrast, our method is parameter free for lidar point interpolation and data pit detection, making it useful for practical applications.

5.2. Conclusions

A robust algorithm based on robust locally weighted regression and robust z-scores was proposed to reduce the negative effect of data pits on CHM derivativeness. A numerical test indicated that with the existence of data pits, the proposed method is more accurate than HPM, mean filter, and median filter methods for CHM construction. A real-world example demonstrated that the proposed method performed better than the classical methods in removing data pits and preserving the edges, structures, and shapes of canopy crowns and gaps. Moreover, the proposed method reduced the underestimation of plot-level maximum tree height. In conclusions, the newly developed method shows high potential for pit-free CHM construction.

Author Contributions: Chuanfa Chen proposed the idea of the paper. He designed the experiments and wrote the paper. Yanyan Li wrote the program codes. Yifu Wang and Tianxiang Yue collected the real-world data sets. Xin Wang analyzed the data sets.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hyyppä, J.; Yu, X.; Hyyppä, H.; Vastaranta, M.; Holopainen, M.; Kukko, A.; Kaartinen, H.; Jaakkola, A.; Vaaja, M.; Koskinen, J. Advances in forest inventory using airborne laser scanning. *Remote Sens.* **2012**, *4*, 1190–1207. [CrossRef]

2. Kato, A.; Moskal, L.M.; Schiess, P.; Swanson, M.E.; Calhoun, D.; Stuetzle, W. Capturing tree crown formation through implicit surface reconstruction using airborne lidar data. *Remote Sens. Environ.* **2009**, *113*, 1148–1162. [CrossRef]

3. Popescu, S.C.; Wynne, R.H.; Nelson, R.F. Estimating plot-level tree heights with lidar: Local filtering with a canopy-height based variable window size. *Comput. Electron. Agric.* **2002**, *37*, 71–95. [CrossRef]

4. Hyyppä, J.; Kelle, O.; Lehikoinen, M.; Inkinen, M. A segmentation-based method to retrieve stem volume estimates from 3-d tree height models produced by laser scanners. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 969–975. [CrossRef]

5. Pouliot, D.; King, D.; Bell, F.; Pitt, D. Automated tree crown detection and delineation in high-resolution digital camera imagery of coniferous forest regeneration. *Remote Sens. Environ.* **2002**, *82*, 322–334. [CrossRef]

6. Brandtberg, T.; Warner, T.A.; Landenberger, R.E.; McGraw, J.B. Detection and analysis of individual leaf-off tree crowns in small footprint, high sampling density lidar data from the eastern deciduous forest in north america. *Remote Sens. Environ.* **2003**, *85*, 290–303. [CrossRef]

7. Koch, B.; Heyder, U.; Weinacker, H. Detection of individual tree crowns in airborne lidar data. *Photogramm. Eng. Remote Sens.* **2006**, *72*, 357–363. [CrossRef]

8. Popescu, S.C.; Wynne, R.H. Seeing the trees in the forest: Using lidar and multispectral data fusion with local filtering and variable window size for estimating tree height. *Photogramm. Eng. Remote Sens.* **2004**, *70*, 589–604. [CrossRef]

9. Persson, A.; Holmgren, J.; Söderman, U. Detecting and measuring individual trees using an airborne laser scanner. *Photogramm. Eng. Remote Sens.* **2002**, *68*, 925–932.

10. Nilsson, M. Estimation of tree heights and stand volume using an airborne lidar system. *Remote Sens. Environ.* **1996**, *56*, 1–7. [CrossRef]

11. Bortolot, Z.J.; Wynne, R.H. Estimating forest biomass using small footprint lidar data: An individual tree-based approach that incorporates training data. *ISPRS J. Photogramm. Remote Sens.* **2005**, *59*, 342–360. [CrossRef]

12. Popescu, S.C. Estimating biomass of individual pine trees using airborne lidar. *Biomass Bioenergy* **2007**, *31*, 646–655. [CrossRef]

13. Naesset, E. Determination of mean tree height of forest stands using airborne laser scanner data. *ISPRS J. Photogramm. Remote Sens.* **1997**, *52*, 49–56. [CrossRef]

14. Leckie, D.; Gougeon, F.; Hill, D.; Quinn, R.; Armstrong, L.; Shreenan, R. Combined high-density lidar and multispectral imagery for individual tree crown analysis. *Can. J. Remote Sens.* **2003**, *29*, 633–649. [CrossRef]

15. Ben-Arie, J.R.; Hay, G.J.; Powers, R.P.; Castilla, G.; St-Onge, B. Development of a pit filling algorithm for lidar canopy height models. *Comput. Geosci.* **2009**, *35*, 1940–1949. [CrossRef]

16. Vosselman, G. Analysis of planimetric accuracy of airborne laser scanning surveys. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2008**, *37*, 99–104.

17. Bater, C.W.; Coops, N.C. Evaluating error associated with lidar-derived DEM interpolation. *Comput. Geosci.* **2009**, *35*, 289–300. [CrossRef]

18. Gaveau, D.L.A.; Hill, R.A. Quantifying canopy height underestimation by laser pulse penetration in small-footprint airborne laser scanning data. *Can. J. Remote Sens.* **2003**, *29*, 650–657. [CrossRef]

19. Khosravipour, A.; Skidmore, A.K.; Isenburg, M.; Wang, T.; Hussin, Y.A. Generating pit-free canopy height models from airborne lidar. *Photogramm. Eng. Remote Sens.* **2014**, *80*, 863–872. [CrossRef]

20. Zhao, D.; Pang, Y.; Li, Z.; Sun, G. Filling invalid values in a lidar-derived canopy height model with morphological crown control. *Int. J. Remote Sens.* **2013**, *34*, 4636–4654. [CrossRef]

21. Shamsoddini, A.; Turner, R.; Trinder, J. Improving lidar-based forest structure mapping with crown-level pit removal. *J. Spat. Sci.* **2013**, *58*, 29–51. [CrossRef]

22. Liu, H.; Dong, P. A new method for generating canopy height models from discrete-return lidar point clouds. *Remote Sens. Lett.* **2014**, *5*, 575–582. [CrossRef]

23.　Nurunnabi, A.; West, G.; Belton, D. Robust locally weighted regression techniques for ground surface points filtering in mobile laser scanning three dimensional point cloud data. *IEEE Trans. Geosci. Remote Sens.* **2015**, *54*, 2181–2193. [CrossRef]

24.　Cleveland, W.S. Robust locally weighted regression and smoothing scatterplots. *J. Am. Stat. Assoc.* **1979**, *74*, 829–836. [CrossRef]

25.　Cleveland, W.S.; Devlin, S.J. Locally weighted regression: An approach to regression analysis by local fitting. *J. Am. Stat. Assoc.* **1988**, *83*, 596–610. [CrossRef]

26.　Fan, J.; Gijbels, I. *Local Polynomial Modelling and Its Applications*; CRC Press: London, UK, 1996.

27.　Rousseeuw, P.J.; Hubert, M. Robust statistics for outlier detection. *WIRES. Data Min. Knowl. Discov.* **2011**, *1*, 73–79. [CrossRef]

28.　Dong, P. Characterization of individual tree crowns using three-dimensional shape signatures derived from lidar data. *Int. J. Remote Sens.* **2009**, *30*, 6621–6628. [CrossRef]

29.　Chen, Q.; Baldocchi, D.; Gong, P.; Kelly, M. Isolating individual trees in a savanna woodland using small footprint lidar data. *Photogramm. Eng. Remote Sens.* **2006**, *72*, 923–932. [CrossRef]

30.　Chow, T.E.; Hodgson, M.E. Effects of lidar post-spacing and dem resolution to mean slope estimation. *Int. J. Geogr. Inf. Sci.* **2009**, *23*, 1277–1295. [CrossRef]

31.　Rizaev, I.G.; Pogorelov, A.V.; Krivova, M.A. A technique to increase the efficiency of artefacts identification in lidar-based canopy height models. *Int. J. Remote Sens.* **2016**, *37*, 1658–1670. [CrossRef]