

Article

Extracting Linear Structures from LiDAR Data by M_{split} Estimation and Different Segmentation Approaches

Patrycja Wyszowska  and Robert Duchnowski * 

Department of Geodesy, Institute of Geodesy and Civil Engineering, Faculty of Geoengineering, University of Warmia and Mazury in Olsztyn, 1 Oczapowskiego Street, 10-719 Olsztyn, Poland; patrycja.wyszowska@uwm.edu.pl

* Correspondence: robert.duchnowski@uwm.edu.pl

Highlights

What are the main findings?

- Processing LiDAR data of non-basic shapes demands data segmentation; this paper proposes a new solution in that context, providing models with similar accuracy but that are much faster than existing methods, such as the conventional sliding window algorithm.
- The tests show that the proposed method is especially advisable for absolute M_{split} estimation. The squared M_{split} estimation is more sensitive to the choice of segmentation method, especially for data of lower accuracy. M_{split} estimation outperforms classical estimation methods when data are contaminated with outliers.

What are the implications of the main findings?

- Extracting linear structures from LiDAR data requires segmentation when the object is complex. This paper shows that the application of the absolute M_{split} estimation and the newly proposed segmentation method provides accurate models and can be used for determining structures such as gable roofs, wall edges, or terrain profiles. The squared M_{split} estimation is recommended for modeling simpler shaped objects and using the conventional sliding window algorithm.
- The absolute M_{split} estimation together with the proposed segmentation algorithm is also recommended for processing low accuracy data or data disturbed by outliers.

Abstract

Raw LiDAR data requires processing to yield more informative, reliable, and accurate products. This paper concerns M_{split} estimation, addresses the problem of extracting linear structures from LiDAR point clouds, and examines different approaches to data segmentation. The study results show that, in practice, processing the entire observation set is usually insufficient, ineffective, and hence inadvisable. The paper considers three approaches: segmentation into disjoint intervals, a sliding window algorithm, and a new solution that combines the two. Tests based on simulated LiDAR data show that the latter approach is a good alternative to the sliding window algorithm. The new segmentation method is at least twice as fast as the sliding window algorithm, and the results are often similarly accurate (especially for narrower windows). Tests on more complex shapes or based on real data indicate that the sliding window algorithm modification is especially recommended for use with the absolute M_{split} estimation. As for the second basic M_{split} estimation variant, namely the squared M_{split} estimation, its accuracy often depends on the choice of segmentation method. The tests demonstrated that M_{split} estimation may outperform classical methods, as confirmed by processing real LiDAR data.



Academic Editors: Alfonso Farina,
Silvia Liberata Ullo and Harun
Taha Hayvaci

Received: 9 June 2026

Revised: 28 July 2026

Accepted: 28 July 2026

Published: 30 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: M_{split} estimation; LiDAR; airborne laser scanning; data segmentation

1. Introduction

Measurements from different systems are usually processed to obtain more reliable or accurate results. Methods used can be universal or dedicated to chosen datasets, processing conditions, limitations, or assumed objectives. The least squares method (LS method) is the oldest, primary, and most popular approach in such applications. While this method has many advantages, its limitations are significant in many practical applications [1–3]. Several other procedures have also often been applied or have become more advisable or popular in recent years, including M-estimation [2,4–6], R-estimation [7,8], RANSAC [9,10], global optimization or other heuristic methods [11,12], the OptD method [13,14], Kalman filtering [15,16], and M_{split} estimation [17,18]. This paper concerns the last method and its application in processing a chosen data type.

M_{split} estimation was introduced as a development of M-estimation, or more broadly, as a generalization of the maximum likelihood method [17,19]. Generally, it is meant to process data that is an unrecognized mixture of realizations of at least two random variables, where the goal is to estimate the location parameters of all variables. Usually, we assume the existence of two observation groups, which is sufficient and helpful in many practical problems. It is also noteworthy that, in addition to its main application, M_{split} estimation can be regarded as an alternative to robust procedures (e.g., [3,20]). The rapid development of M_{split} estimation has resulted in numerous variants, such as squared M_{split} estimation (SMS), absolute M_{split} estimation (AMS), Shift- M_{split} estimation, robust M_{split} estimation, and $M_{\text{split}}^{\text{CD}}$ estimation (e.g., [17–19,21,22]). The basic variants, namely SMS and AMS estimation, as well as the others mentioned, have been applied in practical problems of survey engineering, including direct outlier detection, data filtering, coordinate transformation, deformation analysis, S-transformation, linear regression analysis, marine navigation, and laser scanning data processing (e.g., [22–29]); the last application is the subject of this paper.

M_{split} estimation, in several variants, has proved its applicability for processing data from terrestrial laser scanning (TLS) (e.g., [18,30,31]) or airborne laser scanning (ALS) (e.g., [23,32,33]). The remaining questions are which variant is the best, and how can the point cloud be processed (as a single set or divided into several subsets). Of course, the answers often depend on the point cloud characteristics, such as measurement accuracy, object complexity, the expected occurrence of outliers (or their share), or the aim of data processing. The comparisons of different M_{split} estimation variants can be found in previous publications (e.g., [28,32]). Here, we focus on data segmentation and its influence on the estimation results. It is worth noting that data segmentation is often a crucial part of light detection and ranging (LiDAR) data processing (e.g., [10,34,35]). Segmentation is often applied when one is interested in a specific fragment of the object under study or wants to model or investigate different parts of a particular object. Another purpose, which is essential to the paper's scope, is to apply segmentation to complex or highly variable modeling structures over short ranges. In such cases, processing the entire point cloud may not be sufficient or advisable, as it can result in the loss of valuable information about the object's details. This paper aims to answer the question of how to divide the entire set to obtain the most reliable and accurate results when processing ALS data using M_{split} estimation, specifically for estimating linear structures such as profiles, cross-sections, or wall edges. For comparison, one can either process the entire observation set or segment the data before processing. In the latter approach, the set is divided

using different segmentation methods. For LiDAR data, for example, one can apply a sliding window algorithm, curvature-based adaptive window segmentation, region growing segmentation, RANSAC, or machine learning [10,36–45]. It is noteworthy that the presented segmentation methods are commonly used in practice to extract multiple objects from a single point cloud. Extracted points are then processed separately to model the objects in question. In the paper's study, the segmentation has a slightly different aim; it is not about extraction but about more reliable and accurate object modeling. Therefore, one considers the segmentation methods appropriate for M_{split} estimation, i.e., segmentation into disjoint parts (e.g., intervals) or the application of the sliding window algorithm. In this paper, another new approach is proposed that is, in fact, a combination of the two aforementioned methods.

The paper is organized into the following sections: a general description of the foundations of M_{split} estimation, a brief overview of the segmentation method used in M_{split} estimation, empirical analyses based on simulated observations, and numerical examples on real data from ALS.

2. Materials and Methods

2.1. Foundations of M_{split} Estimation

Generally, M_{split} estimation is a development of the maximum likelihood method and, more appropriately, M-estimation. Therefore, its objective function is created similarly; however, the main difference stems from the split functional model of observations [17]

$$\mathbf{y} = \mathbf{A}\mathbf{X} + \mathbf{v} \Rightarrow \begin{cases} \mathbf{y} = \mathbf{A}\mathbf{X}_{(1)} + \mathbf{v}_{(1)} \\ \mathbf{y} = \mathbf{A}\mathbf{X}_{(2)} + \mathbf{v}_{(2)} \end{cases} \quad (1)$$

where $\mathbf{y}$ is the observation vector, $\mathbf{A}$ is the coefficient matrix, $\mathbf{X}$ is the parameter vector, $\mathbf{X}_{(1)}$ and $\mathbf{X}_{(2)}$ are competitive versions of $\mathbf{X}$, $\mathbf{v}$ is the measurement error vector, and $\mathbf{v}_{(1)}$ and $\mathbf{v}_{(2)}$ are competitive versions of $\mathbf{v}$. Such a model reflects the fact that $\mathbf{y}$ is a mixture of two observation groups, which—from a stochastic perspective—are realizations of two random variables that differ in the expected value at least. It is noteworthy that no information is provided on how to initially divide the observation set into subsets. Each observation is assigned to either of the split functional models in an iterative process by minimizing the corresponding objective function. The division is not strict, as some observations may match both split functional models simultaneously. One should realize that the model of Equation (1) can be extended for more than two split functional models, leading to $M_{\text{split}(q)}$ estimation [46].

Referring to the general foundations of M-estimation, the objective function of M_{split} estimates has the following general form [17,46]:

$$\varphi(\mathbf{X}_{(1)}, \mathbf{X}_{(2)}) = \sum_{i=1}^n \rho(v_{i(1)}, v_{i(2)}) = \sum_{i=1}^n \rho_{(1)}(v_{i(1)}, v_{i(2)}) \rho_{(2)}(v_{i(1)}, v_{i(2)}) \quad (2)$$

where $\rho_{(1)}(v_{i(1)}, v_{i(2)}) = \rho_{(1)}(y_i; \mathbf{X}_{(1)}, \mathbf{X}_{(2)})$ and $\rho_{(2)}(v_{i(1)}, v_{i(2)}) = \rho_{(2)}(y_i; \mathbf{X}_{(1)}, \mathbf{X}_{(2)})$ are the objective function's components. Theoretically, the components can be arbitrarily chosen functions; however, they are usually supposed to be differentiable; hence, one might write the respective influence functions of M_{split} estimation [47] as

$$\begin{aligned} \psi_{(1)}(v_{i(1)}, v_{i(2)}) &= \frac{\partial \rho(v_{i(1)}, v_{i(2)})}{\partial v_{i(1)}} = \psi_{(1)}(v_{i(1)}; y_i, \mathbf{X}_{(2)}) \\ \psi_{(2)}(v_{i(1)}, v_{i(2)}) &= \frac{\partial \rho(v_{i(1)}, v_{i(2)})}{\partial v_{i(2)}} = \psi_{(2)}(v_{i(2)}; y_i, \mathbf{X}_{(1)}) \end{aligned} \quad (3)$$

and later on, the respective weight functions as

$$\begin{aligned} w_{(1)}(v_{i(1)}, v_{i(2)}) &= \frac{\psi_{(1)}(v_{i(1)}, v_{i(2)})}{2v_{i(1)}} = w_{(1)}(v_{i(1)}; y_i, \mathbf{X}_{(2)}) \\ w_{(2)}(v_{i(1)}, v_{i(2)}) &= \frac{\psi_{(2)}(v_{i(1)}, v_{i(2)})}{2v_{i(2)}} = w_{(2)}(v_{i(2)}; y_i, \mathbf{X}_{(1)}) \end{aligned} \quad (4)$$

The properties of each M_{split} estimation variant depend on such functions. For SMS estimation, the functions mentioned can be written as follows [17]:

$$\begin{aligned} \varphi(\mathbf{X}_{(1)}, \mathbf{X}_{(2)}) &= \sum_{i=1}^n \rho(v_{i(1)}, v_{i(2)}) = \sum_{i=1}^n v_{i(1)}^2 v_{i(2)}^2 \\ \begin{cases} \psi_{(1)}(v_{i(1)}, v_{i(2)}) = 2v_{i(1)} v_{i(2)}^2 \\ \psi_{(2)}(v_{i(1)}, v_{i(2)}) = 2v_{i(1)}^2 v_{i(2)} \\ w_{(1)}(v_{i(1)}, v_{i(2)}) = v_{i(2)}^2 \\ w_{(2)}(v_{i(1)}, v_{i(2)}) = v_{i(1)}^2 \end{cases} \end{aligned} \quad (5)$$

Note that the influence functions are unbounded; therefore, SMS estimation cannot be classified as a globally robust method [47,48].

On the other hand, AMS estimation has the following characteristic functions [19]:

$$\begin{aligned} \varphi(\mathbf{X}_{(1)}, \mathbf{X}_{(2)}) &= \sum_{i=1}^n \rho(v_{i(1)}, v_{i(2)}) = \sum_{i=1}^n |v_{i(1)}| |v_{i(2)}| \\ \begin{cases} \psi_{(1)}(v_{i(1)}, v_{i(2)}) = \begin{cases} -|v_{i(2)}| & \text{for } v_{i(1)} < 0 \\ |v_{i(2)}| & \text{for } v_{i(1)} > 0 \end{cases} \\ \psi_{(2)}(v_{i(1)}, v_{i(2)}) = \begin{cases} -|v_{i(1)}| & \text{for } v_{i(2)} < 0 \\ |v_{i(1)}| & \text{for } v_{i(2)} > 0 \end{cases} \\ w_{(1)}(v_{i(1)}, v_{i(2)}) = \begin{cases} \frac{|v_{i(2)}|}{2c^\bullet} & \text{for } |v_{i(1)}| < c^\bullet \\ \frac{|v_{i(2)}|}{2|v_{i(1)}|} & \text{for } |v_{i(1)}| \geq c^\bullet \end{cases} \\ w_{(2)}(v_{i(1)}, v_{i(2)}) = \begin{cases} \frac{|v_{i(1)}|}{2c^\bullet} & \text{for } |v_{i(2)}| < c^\bullet \\ \frac{|v_{i(1)}|}{2|v_{i(2)}|} & \text{for } |v_{i(2)}| \geq c^\bullet \end{cases} \end{cases} \end{aligned} \quad (6)$$

where $c^\bullet$ is assumed to be a small positive constant. The influence functions of AMS estimation are also unbounded; therefore, the method is not robust against outliers globally. As for AMS estimation, the values of the influence function $\psi_{(2)}(v_{i(1)}, v_{i(2)})$ for the moderate and high $|v_{i(1)}|$ are smaller than those for SMS estimation (the same concerns the other influence function). Moreover, the highest values of the weight functions for AMS estimation are obtained for minimal absolute values of $|v_{i(1)}|$ or $|v_{i(2)}|$, respectively. In other words, AMS method “favors” the observations with small measurement errors more than SMS estimation does [17,19,49]. Hence, one can claim that AMS estimation should be less sensitive to moderate or high-level measurement errors than SMS estimation, which was confirmed by prior empirical research [19,50].

M_{split} estimates can be determined by applying the Newton method. Two slightly different processes were proposed in such a context. SMS estimation (and its developments) employs the traditional iterative process, where the second parameter version depends on the first parameter version computed in the current iteration step and the second parameter version from the previous iteration step. Thus, it can be written as [9,28]

$$\begin{aligned}\mathbf{X}_{(1)}^j &= \mathbf{X}_{(1)}^{j-1} + d\mathbf{X}_{(1)}^j = \mathbf{X}_{(1)}^{j-1} - \left[\mathbf{H}_{(1)}(\mathbf{X}_{(1)}^{j-1}, \mathbf{X}_{(2)}^{j-1}) \right]^{-1} \mathbf{g}_{(1)}(\mathbf{X}_{(1)}^{j-1}, \mathbf{X}_{(2)}^{j-1}) \\ \mathbf{X}_{(2)}^j &= \mathbf{X}_{(2)}^{j-1} + d\mathbf{X}_{(2)}^j = \mathbf{X}_{(2)}^{j-1} - \left[\mathbf{H}_{(2)}(\mathbf{X}_{(1)}^j, \mathbf{X}_{(2)}^{j-1}) \right]^{-1} \mathbf{g}_{(2)}(\mathbf{X}_{(1)}^j, \mathbf{X}_{(2)}^{j-1})\end{aligned}\quad (7)$$

where $d\mathbf{X}_{(1)}$ and $d\mathbf{X}_{(2)}$ are increments to parameter vectors $\mathbf{X}_{(1)}$ and $\mathbf{X}_{(2)}$, respectively, $\mathbf{H}_{(1)}(\mathbf{X}_{(1)}, \mathbf{X}_{(2)})$ and $\mathbf{H}_{(2)}(\mathbf{X}_{(1)}, \mathbf{X}_{(2)})$ are Hessians, $\mathbf{g}_{(1)}(\mathbf{X}_{(1)}, \mathbf{X}_{(2)})$ and $\mathbf{g}_{(2)}(\mathbf{X}_{(1)}, \mathbf{X}_{(2)})$ are gradients given in the forms

$$\begin{cases} \mathbf{H}_{(1)}(\mathbf{X}_{(1)}, \mathbf{X}_{(2)}) = \frac{\partial^2 \varphi(\mathbf{y}; \mathbf{X}_{(1)}, \mathbf{X}_{(2)})}{\partial \mathbf{X}_{(1)} \partial \mathbf{X}_{(1)}^T} = 2\mathbf{A}^T \mathbf{w}_{(1)}(\mathbf{v}_{(1)}, \mathbf{v}_{(2)}) \mathbf{A} \\ \mathbf{H}_{(2)}(\mathbf{X}_{(1)}, \mathbf{X}_{(2)}) = \frac{\partial^2 \varphi(\mathbf{y}; \mathbf{X}_{(1)}, \mathbf{X}_{(2)})}{\partial \mathbf{X}_{(2)} \partial \mathbf{X}_{(2)}^T} = 2\mathbf{A}^T \mathbf{w}_{(2)}(\mathbf{v}_{(1)}, \mathbf{v}_{(2)}) \mathbf{A} \\ \mathbf{g}_{(1)}(\mathbf{X}_{(1)}, \mathbf{X}_{(2)}) = \left[\frac{\partial \varphi(\mathbf{y}; \mathbf{X}_{(1)}, \mathbf{X}_{(2)})}{\partial \mathbf{X}_{(1)}} \right]^T = -2\mathbf{A}^T \mathbf{w}_{(1)}(\mathbf{v}_{(1)}, \mathbf{v}_{(2)}) \mathbf{v}_{(1)} \\ \mathbf{g}_{(2)}(\mathbf{X}_{(1)}, \mathbf{X}_{(2)}) = \left[\frac{\partial \varphi(\mathbf{y}; \mathbf{X}_{(1)}, \mathbf{X}_{(2)})}{\partial \mathbf{X}_{(2)}} \right]^T = -2\mathbf{A}^T \mathbf{w}_{(2)}(\mathbf{v}_{(1)}, \mathbf{v}_{(2)}) \mathbf{v}_{(2)} \end{cases} \quad (8)$$

The matrices $\mathbf{w}_{(1)}(\mathbf{v}_{(1)}, \mathbf{v}_{(2)})$ and $\mathbf{w}_{(2)}(\mathbf{v}_{(1)}, \mathbf{v}_{(2)})$ are computed using the weight functions in the following way:

$$\begin{cases} \mathbf{w}_{(1)}(\mathbf{v}_{(1)}, \mathbf{v}_{(2)}) = \text{diag} \left[w_{(1)}(v_{1(1)}, v_{1(2)}), \dots, w_{(1)}(v_{n(1)}, v_{n(2)}) \right] \\ \mathbf{w}_{(2)}(\mathbf{v}_{(1)}, \mathbf{v}_{(2)}) = \text{diag} \left[w_{(2)}(v_{1(1)}, v_{1(2)}), \dots, w_{(2)}(v_{n(1)}, v_{n(2)}) \right] \end{cases} \quad (9)$$

where $\text{diag}(\circ)$ is a diagonal matrix. The traditional iterative process is possible because SMS estimation applies mutual cross-weighting, i.e., the weight function $w_{(1)}(v_{i(1)}, v_{i(2)})$ depends only on $v_{i(2)}$, and $w_{(2)}(v_{i(1)}, v_{i(2)})$ only on $v_{i(1)}$ [17]. The LS estimate is usually taken as the starting point of the traditional iterative process. The process ends for such j , for which $\hat{\mathbf{X}}_{(1)} = \mathbf{X}_{(1)}^j = \mathbf{X}_{(1)}^{j-1}$ and $\hat{\mathbf{X}}_{(2)} = \mathbf{X}_{(2)}^j = \mathbf{X}_{(2)}^{j-1}$ with the assumed tolerance ε (in the paper $\varepsilon = 10^{-6}$) [17,19].

The other basic M_{split} estimation variant, namely AMS estimation, cannot use the iterative process discussed. That is because its weight functions depend on both measurement errors (see Equation (6)); hence, there is no mutual cross-weighting. For that reason, a parallel iterative process was introduced [19]:

$$\begin{cases} \mathbf{X}_{(1)}^j = \mathbf{X}_{(1)}^{j-1} + d\mathbf{X}_{(1)}^j = \mathbf{X}_{(1)}^{j-1} - \left[\mathbf{H}_{(1)}(\mathbf{X}_{(1)}^{j-1}, \mathbf{X}_{(2)}^{j-1}) \right]^{-1} \mathbf{g}_{(1)}(\mathbf{X}_{(1)}^{j-1}, \mathbf{X}_{(2)}^{j-1}) \\ \mathbf{X}_{(2)}^j = \mathbf{X}_{(2)}^{j-1} + d\mathbf{X}_{(2)}^j = \mathbf{X}_{(2)}^{j-1} - \left[\mathbf{H}_{(2)}(\mathbf{X}_{(1)}^{j-1}, \mathbf{X}_{(2)}^{j-1}) \right]^{-1} \mathbf{g}_{(2)}(\mathbf{X}_{(1)}^{j-1}, \mathbf{X}_{(2)}^{j-1}) \end{cases} \quad (10)$$

where the Hessians and gradients are determined as described in Equation (8). Unlike the traditional iterative process, the second-parameter version is determined differently. It depends on both parameter versions computed in the previous iteration step. The starting points of the parallel iterative process might be chosen differently. They can be based on LS estimates, SMS estimates, or approximated using raw observations [28,51]. The condition of ending the parallel iterative process is the same as for the traditional one.

2.2. Segmentation of LiDAR Data During M_{split} Estimation

Several applications of M_{split} estimation, presented in the Introduction, demonstrate the method's applicability to LiDAR data processing. It is especially advisable when modeling objects consisting of two (or more) parts, such as two different surfaces, planes, or, in the case of linear objects, two lines (such as curve lines or a broken line) [18,31,32,52,53]. That application is natural to all variants of M_{split} estimation. The method in question may yield outstanding results (outperforming classical methods) when a point cloud is

contaminated by outliers [18,23,28,30,32]. Note that LiDAR data often include some share of outlying observations of different origin and/or magnitude [28,54]. Regardless of the applications discussed, the raw point cloud is usually segmented before modeling the object or its parts. The subsets are extracted from the entire cloud to reflect the part of the object under modeling. It is especially essential for estimating the linear representation of surfaces (3D objects), single profiles, cross-sections, or lines (2D objects). Such prepared subsets can be processed in several ways. If the surface or profile is not geometrically complex, it can be processed as a whole. Otherwise, the subset should be resegmented to accurately reflect the complexity and variability of the modeled object. In that context, two general approaches can be considered: dividing whole sets into intervals of the same width (e.g., [18,28,44]), or using the sliding window algorithm [33,42,44]. Some preliminary remarks on their application were presented in several cited papers; however, the problem of determining the most suitable approach for M_{split} estimation remains open, requiring deeper analysis. Moreover, considering the experiences and conclusions presented in the aforementioned papers, an alternative approach is proposed here. Let us consider the sliding window algorithm, where the slide is half the window width. Then, each part of the object is modeled twice based on the neighboring subsets (note that half of the observations of such subsets are common). The final model can be obtained by averaging both. This processing approach is a new method for data segmentation used in M_{split} estimation. It should provide more reliable results than processing disjoint intervals and should be much faster than the conventional sliding window algorithm. In that paper, this approach is referred to as the sliding window modification.

The general idea of this paper is to compare the new proposed segmentation method with the conventional ones discussed. Processing simulated and real data illuminates that problem and provides useful insights into the practical applications of M_{split} estimation in LiDAR data processing.

3. Results

The very first tests (see Sections 3.1–3.3) are performed on simulated observations (which were generated in Mathcad 15.0). According to the paper's objective, it concerns the modeling of 2D structures, specifically the profiles. One considers three modeling cases: a straight-line profile, a fourth-degree polynomial, or a broken line simulating a roof cross-section. Several observation sets are simulated in each test, differing in the proportion of outliers. Some simulation assumptions are common to all three profiles: the share of outliers is 0%, 10%, or 60%; all outliers are positive; and their magnitudes are either $[3\sigma; 6\sigma]$ or $[6\sigma; 12\sigma]$, where σ is the measurement standard deviation equal to 2 mm (often assumed for TLS data). Regular observations are simulated with normally distributed errors having the assumed standard deviation, whereas outliers are uniformly distributed within given intervals.

The estimation results can be compared by examining the goodness-of-fit of the estimated profiles to the simulated ones, respectively. The root-mean-square deviation (RMSD) is a measure used to assess the quantity in question [18,32]. For the paper objective, it can be computed as follows:

$$RMSD(\hat{H}) = \sqrt{\frac{\sum_{i=1}^p (\hat{H}_i - H_i)^2}{p}} \quad (11)$$

where p is the number of points chosen for the computation, and $\hat{H}_i$ and H_i are estimated and simulated values of height at the chosen profile points, respectively.

To achieve the paper's objectives, the simulated observation sets are processed in four ways: as a whole set, divided into disjoint intervals, or using a sliding window algorithm or

its modification. To assess the impact of the chosen assumptions in the last two approaches, several computational variants are examined, varying the interval range, or window width, and slide width, respectively.

3.1. Straight-Line Profile

Let us consider a profile in the shape of a straight line described by the equation:

$$H = 0.01d + 0.5 \quad (12)$$

where $d \in [0 \text{ m}, 100 \text{ m}]$, and it is the distance from the profile beginning. One thousand measurements were simulated under the conditions given in the previous section. Therefore, one considers five observation sets, including 0%, 10%, or 60% outliers of two different magnitudes, respectively. Each set is processed using the basic M_{split} estimation variants. For comparison, conventional methods are also employed, including LS estimation, the Huber method, and the Tukey method (two robust M-estimation variants). First, the sets are processed as a whole. For the sake of simplicity, that variant, as well as similar ones in the subsequent tests, are denoted as the W variant. Then, they are segmented into intervals (5 m, 10 m, or 20 m long) and processed in such parts. The variants are denoted as I5, I10, and I20, respectively. In the case of the sliding window algorithm, one also considers three variants of the window width (5 m, 10 m, and 20 m); therefore, they are noted as SW5, SW10, and SW20. In each variant, there are p windows considered, where each center of the window is used in the computation of RMSD (Equation (11)). Finally, the sliding window algorithm modification is applied, where the window width equals 5 m, 10 m, or 20 m, and the slide is 2.5 m, 5 m, or 10 m, respectively (the variants are denoted as SM5, SM10, and SM20). In all variants, the linear model is used to model the profile. In the case of M_{split} estimation, where one obtains two competitive solutions, the solution lying lower is assumed to describe the profile (the other solution describes the location of outliers in some sense; hence, it is neglected). The goodness-of-fit is described by RMSD, computed using Equation (11), with 60 evenly distributed points. All RMSDs concerning all variants are presented in Tables 1–6.

The results presented in Table 1 show that the conventional methods outperform M_{split} estimation, which is not a surprise. One might conclude that the most accurate results are usually obtained by processing the whole set. Only AMS estimation provides comparable results for almost all variants. In the case of segmentation, slightly better results are obtained with wider intervals and wider sliding windows.

Table 1. RMSDs (mm) for different segmentation variants (set without outliers).

Method	W	I5	I10	I20	SW5	SW10	SW20	SM5	SM10	SM20
LS	0.1	0.4	0.3	0.2	0.3	0.2	0.1	0.3	0.3	0.2
Huber	0.1	0.4	0.3	0.2	0.3	0.2	0.1	0.4	0.3	0.2
Tukey	0.1	0.4	0.3	0.2	0.3	0.2	0.1	0.3	0.3	0.2
SMS	1.4	2.0	2.0	2.0	1.9	1.6	1.6	1.9	1.8	1.8
AMS	1.4	1.6	1.3	1.4	1.4	1.3	1.3	1.6	1.3	1.3

Table 2. RMSDs (mm) for different segmentation variants and averaging two competitive models (set without outliers).

Method	W	I5	I10	I20	SW5	SW10	SW20	SM5	SM10	SM20
SMS _{average}	0.1	0.6	0.5	0.3	0.4	0.3	0.2	0.5	0.4	0.3
AMS _{average}	0.1	0.5	0.3	0.2	0.3	0.3	0.1	0.4	0.3	0.2

Table 3. RMSDs (mm) for different segmentation variants (set including 10% outliers—gross errors from the interval $[3\sigma-6\sigma]$).

Method	W	I5	I10	I20	SW5	SW10	SW20	SM5	SM10	SM20
LS	0.8	1.0	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
Huber	0.3	0.6	0.5	0.4	0.5	0.4	0.4	0.5	0.5	0.4
Tukey	0.6	0.8	0.7	0.7	0.7	0.7	0.7	0.8	0.7	0.7
SMS	0.8	1.1	1.1	0.9	1.1	0.9	0.8	1.1	1.0	0.9
AMS	0.6	1.3	0.8	1.0	0.9	0.7	0.6	1.0	0.6	0.7

Table 4. RMSDs (mm) for different segmentation variants (set including 10% outliers—gross errors from the interval $[6\sigma-12\sigma]$).

Method	W	I5	I10	I20	SW5	SW10	SW20	SM5	SM10	SM20
LS	1.7	2.0	1.9	1.8	1.9	1.8	1.8	1.9	1.9	1.8
Huber	0.1	0.5	0.3	0.2	0.4	0.3	0.2	0.4	0.3	0.2
Tukey	1.1	1.5	1.3	1.2	1.4	1.2	1.2	1.4	1.3	1.2
SMS	0.4	0.7	0.5	0.4	0.5	0.5	0.4	0.6	0.5	0.4
AMS	0.2	1.0	0.5	0.3	0.5	0.3	0.2	0.7	0.5	0.3

Table 5. RMSDs (mm) for different segmentation variants (set including 60% outliers—gross errors from the interval $[3\sigma-6\sigma]$).

Method	W	I5	I10	I20	SW5	SW10	SW20	SM5	SM10	SM20
LS	5.3	5.4	5.4	5.3	5.3	5.4	5.4	5.3	5.4	5.3
Huber	5.3	5.4	5.4	5.3	5.3	5.4	5.4	5.3	5.4	5.3
Tukey	5.4	5.4	5.4	5.4	5.3	5.4	5.4	5.4	5.4	5.4
SMS	0.2	1.9	3.0	2.9	1.9	1.9	1.7	1.6	2.4	1.6
AMS	0.6	1.8	3.5	3.8	2.0	2.4	2.7	1.8	3.1	3.4

Table 6. RMSDs (mm) for different segmentation variants (set including 60% outliers—gross errors from the interval $[6\sigma-12\sigma]$).

Method	W	I5	I10	I20	SW5	SW10	SW20	SM5	SM10	SM20
LS	10.7	10.7	10.7	10.7	10.7	10.7	10.7	10.7	10.7	10.7
Huber	10.7	10.7	10.7	10.7	10.7	10.7	10.7	10.7	10.7	10.7
Tukey	10.8	10.8	10.8	10.8	10.7	10.8	10.8	10.8	10.8	10.8
SMS	0.8	5.4	4.1	1.0	3.4	2.9	1.5	4.0	4.0	0.9
AMS	0.5	7.5	5.2	7.3	5.2	5.1	4.6	5.5	6.1	6.6

The observation set is homogeneous; therefore, processing it using M_{split} estimation is pointless. However, a general property of M_{split} estimation is that the competitive solutions for processing such sets are symmetrically distributed around the true value. Hence, one can compute the average height of each point from both solutions. The respective RMSDs are presented in Table 2. In this context, the accuracy of M_{split} estimation variants is remarkably similar to that of conventional methods (presented in Table 1). Comparing the results presented in Tables 1 and 2 across different segmentation methods, one can say that for the same width of interval or window, the outcomes of all three approaches to set segmentation are comparable. The sliding window algorithm, however, seems the most accurate.

Tables 3 and 4 present results for sets contaminated with outliers. The Huber method yields the most accurate results; however, M_{split} estimation provides comparable results across many variants. The results of SMS and AMS estimations are almost always better than those of the Tukey method or LS estimation.

Processing the entire set usually provides the best results; however, the differences in accuracy are less pronounced than those shown in Table 1. Considering set segmentation, the sliding window algorithm yields the best results for lower-magnitude outliers. The proposed modification gives results of similar accuracy, and processing data in intervals is slightly less effective. All three segmentation methods yield similar results for higher-magnitude outliers.

When the outlier share is higher, at 60% (Tables 5 and 6), M_{split} estimation outperforms all conventional methods (which is not surprising, as conventional methods cannot handle such high proportions). Conventional methods yield the same unsatisfying results regardless of the variants applied. Regarding M_{split} estimation, the best results are again obtained from the variant W. Among segmentation approaches, the sliding window algorithm usually yields more accurate results. For some variants, better results are provided using the sliding window modification. For lower-magnitude outliers, both M_{split} estimation variants perform better when the intervals or sliding windows are the narrowest.

In contrast, for the larger outliers, SMS estimation provides the best results for the widest intervals and sliding windows. Regarding AMS estimation, it is hard to conclude the optimal interval width. It is better to widen the sliding windows; however, when one uses segmentation in intervals, the results are optimal for medium-width intervals. In general, processing data in intervals yields slightly worse results than when applying the sliding window algorithm or its modification.

The first test provides general information about the accuracy of modeling linear structures using different segmentation methods. The sliding window algorithm and its modification usually yield better results than segmenting into disjoint intervals. However, the best way is to process the set as a whole. It stems from the shape one deals with, meaning that one can easily take a proper functional model. For the same reason, the best results of processing segmented sets are obtained for the widest intervals or sliding windows. Therefore, in subsequent tests, more complex simulated profiles are used.

The accuracy of the results is crucial, but one should also consider the complexity and time required for the computations. Processing data in intervals is the most time-efficient approach, whereas the sliding window modification takes twice as long. The sliding window algorithm is the most time-consuming, requiring two to six times more time than its modification (depending on the window size).

3.2. Fourth-Degree Polynomial Profile

Let us consider a profile in the shape of a fourth-degree polynomial described by the equation:

$$H = -0.000000024d^4 + 0.00004752d^3 - 0.0030684d^2 + 0.066124d + 0.5 \quad (13)$$

In general, such a shape might be considered a terrain profile; note that a fourth-degree polynomial can be used to model it [55].

The regular observations and outliers are simulated under the same assumptions as in the previous test. The set without outliers is presented in Figure 1.

Considering the conclusions from the previous test, one can omit the variant with no outliers. The other variants of data processing remain the same, with one exception: in real life, one would not know the real shape of the profile; hence, the variant W is not applicable in that case. Given the paper's objectives, a second-degree polynomial can serve as a functional model when processing data in subsets derived from data segmentation [18,56].

Tables 7 and 8 present results for sets with a 10% outlier share. The accuracy of fit is usually lower than in the first test, for all methods. Moreover, for wider intervals or sliding windows, the accuracy decreases rapidly. This probably stems from the more complex shape of the profile and the "unacquaintance" with the real profile shape. The sliding

window algorithm often yields better results than its modification or processing data in segmented intervals of the same width.

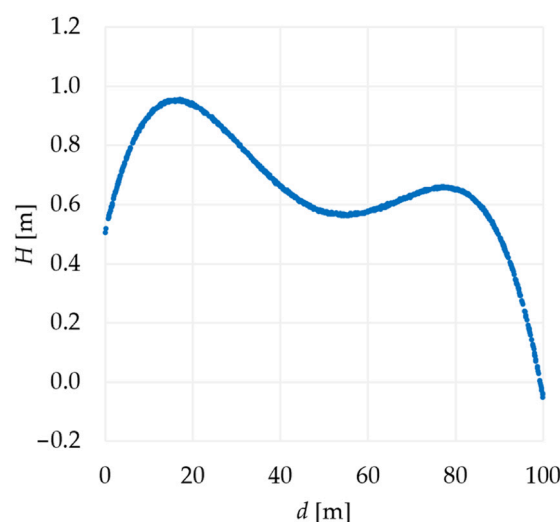


Figure 1. The simulated observation set of a fourth-degree polynomial profile without outliers.

Table 7. RMSDs (mm) for different segmentation variants (set including 10% outliers—gross errors from the interval $[3\sigma-6\sigma]$).

Method	I5	I10	I20	SW5	SW10	SW20	SM5	SM10	SM20
LS	1.2	1.1	4.1	1.2	1.0	1.3	1.1	1.1	3.3
Huber	0.9	0.7	4.0	0.8	0.6	1.2	0.8	0.7	3.2
Tukey	1.0	1.0	4.1	1.0	0.8	1.2	0.9	0.9	3.3
SMS	1.3	1.2	6.1	1.2	1.0	2.1	1.2	1.2	5.9
AMS	1.5	1.2	5.8	1.1	0.8	2.1	1.1	1.3	4.9

Table 8. RMSDs (mm) for different segmentation variants (set including 10% outliers—gross errors from the interval $[6\sigma-12\sigma]$).

Method	I5	I10	I20	SW5	SW10	SW20	SM5	SM10	SM20
LS	3.5	2.1	4.4	2.2	2.0	2.0	2.6	2.0	3.5
Huber	0.9	0.6	4.0	0.6	0.4	1.3	0.7	0.5	3.1
Tukey	2.9	1.5	4.2	1.7	1.5	1.6	2.1	1.5	3.3
SMS	0.6	0.9	8.8	0.7	0.5	1.2	0.6	0.9	7.4
AMS	3.4	1.8	6.3	0.9	0.6	1.3	2.1	1.9	6.2

Tables 9 and 10 present the goodness-of-fit for a higher share of outliers. The conventional methods yield slightly less accurate results than in the test in the previous subsection. M_{split} estimation variants yield better results for several computational variants but also produce worse or significantly worse results for others. They typically yield the most accurate results for the shortest-width intervals or sliding windows. The general tendency is that the accuracy of goodness-of-fit decreases significantly for the widest intervals or sliding windows. The sliding window algorithm typically yields much more accurate results than its modification or data processing based on interval segmentation. The sliding window modification outperforms data processing in the disjoint intervals, which is particularly evident in SMS estimation.

Table 9. RMSDs (mm) for different segmentation variants (set including 60% outliers—gross errors from the interval $[3\sigma-6\sigma]$).

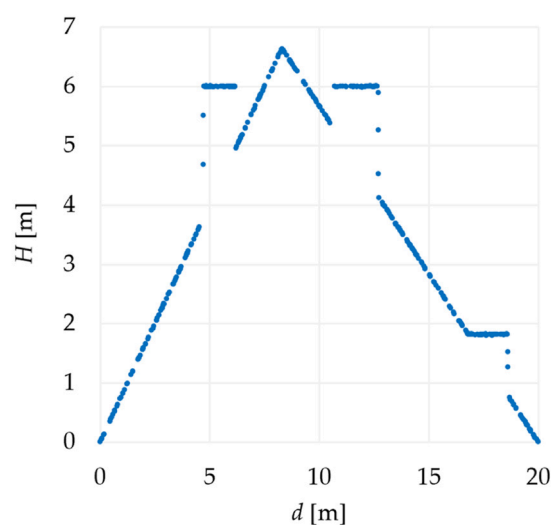
Method	I5	I10	I20	SW5	SW10	SW20	SM5	SM10	SM20
LS	5.6	5.6	6.8	5.5	5.5	5.5	5.5	5.5	6.2
Huber	5.6	5.6	6.8	5.5	5.5	5.5	5.5	5.5	6.3
Tukey	5.6	5.6	6.8	5.5	5.5	5.5	5.6	5.6	6.3
SMS	2.4	2.9	6.6	1.1	0.8	1.4	1.7	2.3	6.4
AMS	3.3	2.9	6.9	1.6	1.5	2.1	2.8	2.4	6.8

Table 10. RMSDs (mm) for different segmentation variants (set including 60% outliers—gross errors from the interval $[6\sigma-12\sigma]$).

Method	I5	I10	I20	SW5	SW10	SW20	SM5	SM10	SM20
LS	11.1	11.0	11.6	10.9	10.9	10.8	11.0	10.9	11.2
Huber	11.1	11.0	11.6	10.9	10.9	10.8	11.0	10.9	11.2
Tukey	11.2	11.1	11.7	11.0	11.0	10.9	11.0	11.0	11.2
SMS	4.6	5.1	8.5	1.4	1.2	1.6	2.9	3.3	7.5
AMS	6.4	5.3	9.4	1.6	1.7	3.1	4.4	5.3	9.4

3.3. Profile of Gable Roof

The last test based on simulated observations differs from the previous ones. The observation set is simulated as a 20 m long profile of a gable roof with two ventilation chimneys and an oriel. Data are simulated as normally distributed 400 ALS measurements with an accuracy of 5 mm [57,58]. The simulated observation set is shown in Figure 2.

**Figure 2.** The simulated observation set of the gable roof profile.

The profile's complex shape and length necessitate a change in the interval or sliding window width; hence, considering the profile's characteristics, intervals or sliding windows of width equal to 1 m seem adequate. Here, considering the roof shape, one can apply a first-degree polynomial. Processing the observation set at hand, M_{split} estimation reveals its natural objective. In many intervals (or sliding windows), there are two (or even three) lines to be modeled. In such cases, one can use competitive M_{split} estimates to model both lines simultaneously. In cases where there is only one line to be modeled, one chooses the results that describe the lower-lying line.

The accuracy of the goodness-of-fit model in representing the roof's true shape is described using the RMSDs given in Equation (11) for $p = 60$. RMSDs determined for each estimation method and the three segmentation approaches are given in Table 11.

Table 11. RMSDs (mm) for different segmentation variants (set of gable roof).

Method	Intervals	Sliding Window	Sliding Window Modification
LS	191.9	201.2	190.9
Huber	192.9	199.6	186.2
Tukey	191.8	202.2	190.1
SMS	101.5	32.0	57.9
AMS	4.3	4.2	3.5

Both M_{split} estimation variants outperform classical methods, stemming from the roof shape. For SMS estimation, the best results are achieved using the sliding window algorithm. The segmentation method proposed in the paper yields results that are less accurate than those of the sliding window algorithm but significantly more accurate than those obtained by processing data segmented into disjoint intervals. AMS estimation gives the best results, with RMSDs that are even twenty times smaller than those of SMS estimation. This probably stems from the fact that in some intervals (or windows), there are observations of three lines. AMS estimation can handle such situations, as it is less sensitive to outliers of moderate magnitude [19]. For the method in question, the sliding window modification yields the most accurate results.

One can notice that in some intervals (or windows), there is only one line to be estimated. Therefore, as in the example given in Section 3.1, one can average the results from both competitive models for such intervals or windows. RMSDs computed for such a case are presented in Table 12. The goodness-of-fit accuracy improves; however, it is more closely related to the AMS estimation.

Table 12. RMSDs (mm) for different segmentation variants and averaging two competitive models (set of the gable roof).

Method	Intervals	Sliding Window	Sliding Window Modification
SMS_{average}	101.4	31.7	57.4
AMS_{average}	2.6	2.5	2.1

3.4. Examples of Real Data Processing

Let us consider the estimation of the terrain profile located near Malbork, Poland. ALS data are downloaded from the Polish database Geoportal <https://www.geoportal.gov.pl/> (accessed on 9 July 2024). One can consider two observation sets, from 2022 and 2023. The cutouts of the original ALS point clouds related to a distance d of 100 m are shown in Figure 3.

The data shows different accuracy and the vertical shift between 2022 and 2023. It is worth noting that the first subset was obtained in April 2022, while the second one was collected in September 2023. Therefore, the first profile refers to the vegetation cover, while the second one is more or less related to the ground.

The field profiles were estimated under the following assumptions: the width of the interval (or the window) is 20 m, the modeling applies a second-degree polynomial, and finally, the competitive solution related to the lower-located line is taken. The estimated profiles using different estimation methods are presented in separate figures. For the sake of comparison, the profiles based on the raw data are also presented. Figure 4 shows the profiles obtained for segmentation into disjoint intervals (I). Figure 5 presents the profiles

determined by the sliding window algorithm (SW), while Figure 6 shows the profiles obtained by modifying the sliding window algorithm (SM).

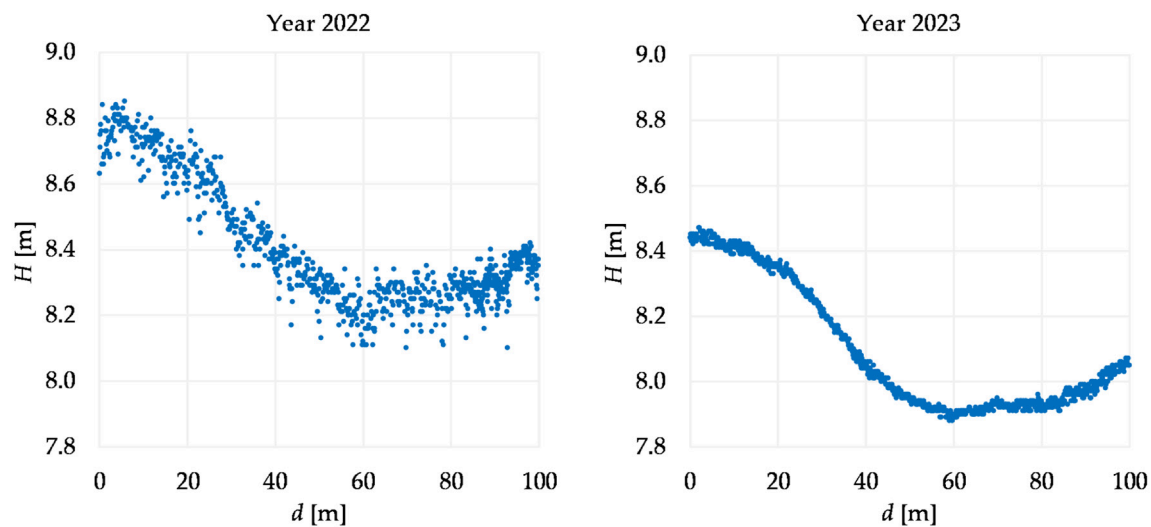


Figure 3. The observation sets of the terrain profile for the years 2022 and 2023.

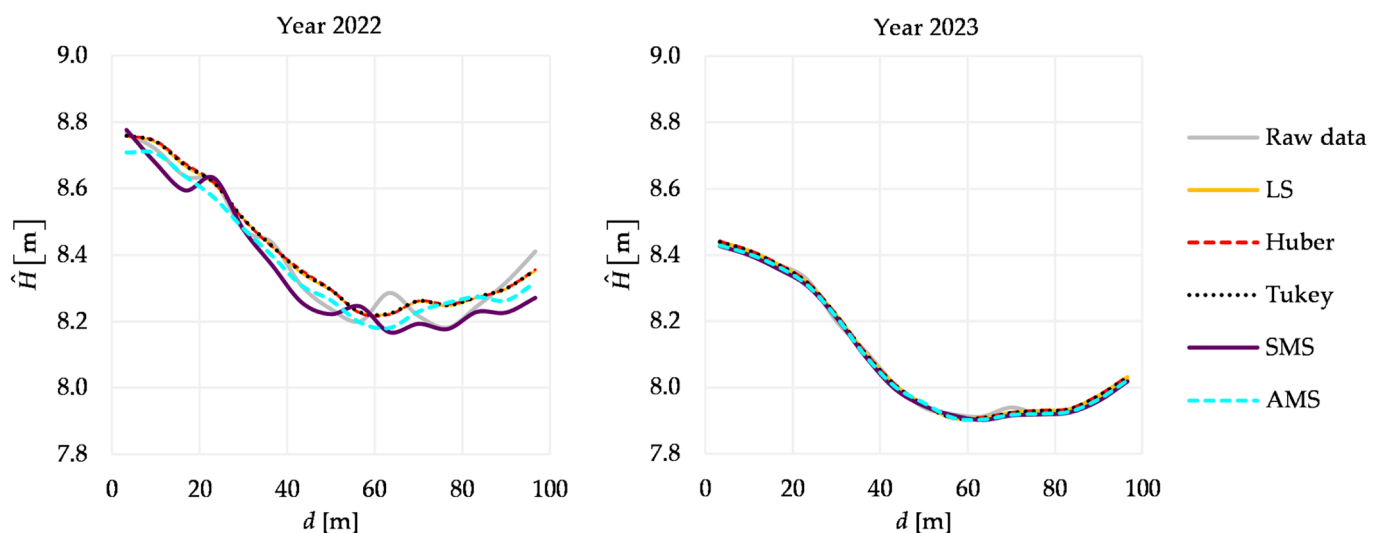


Figure 4. The estimated field profiles for the years 2022 and 2023 obtained using the segmentation into disjoint intervals (I).

The estimated profiles from 2023 are all similar to each other and to the profile obtained from the raw data, regardless of the segmentation method applied. The profiles obtained in 2022 are more diverse; however, the profiles obtained using classical methods are similar to one another. Even a simple graphical analysis reveals that the most vivid differences are between the profiles obtained by using SMS estimation. To illustrate these differences more explicitly, the discrepancies between profiles obtained using the same estimation method but different segmentation approaches are shown in Figures 7 and 8. The profiles obtained by segmentation into disjoint intervals or the sliding window algorithm modification are compared with those obtained by the sliding window algorithm. It stems from previous tests, where the sliding window algorithm typically yields the most accurate results.

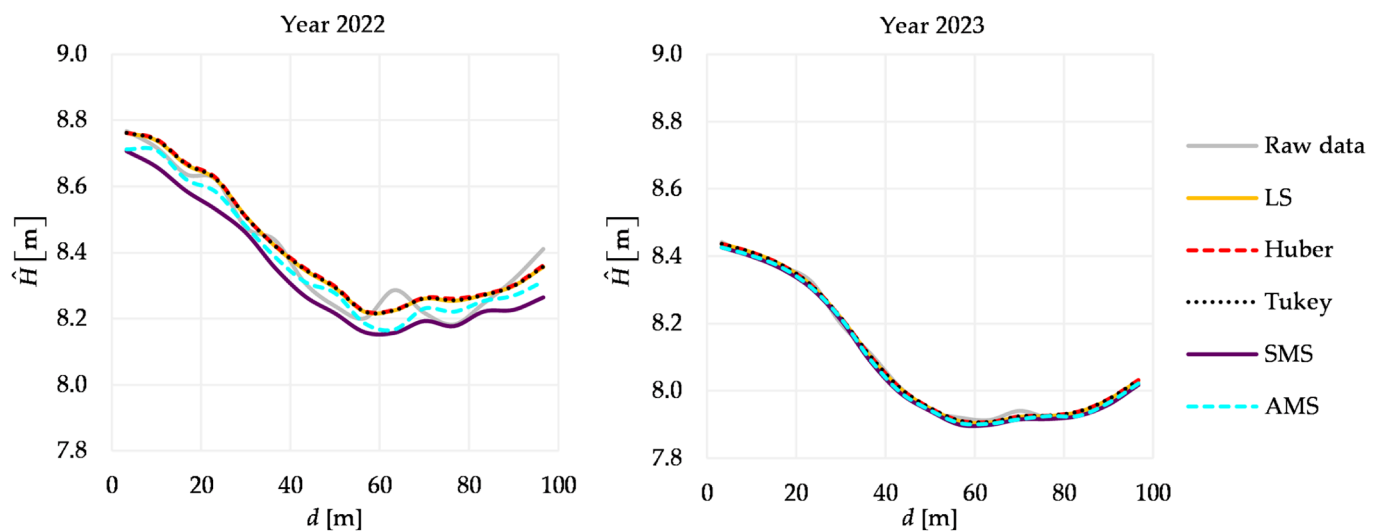


Figure 5. The estimated field profiles for the years 2022 and 2023 obtained using the sliding window algorithm (SW).

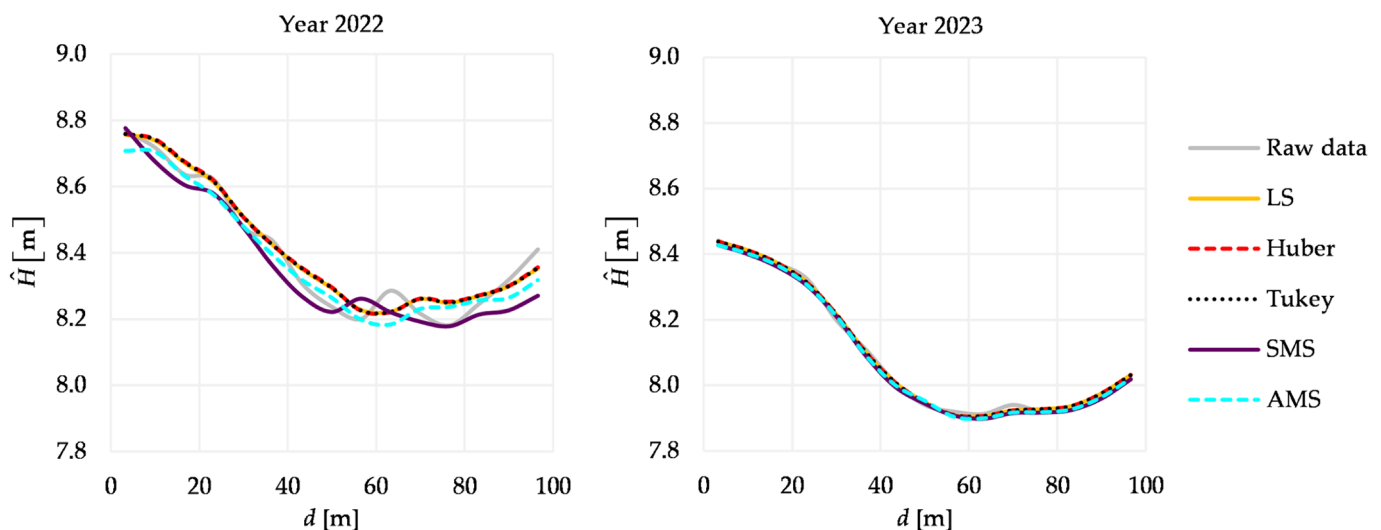


Figure 6. The estimated field profiles for the years 2022 and 2023 obtained using the sliding window algorithm modification (SM).

Figure 7 clearly shows that SMS estimation is most sensitive to the choice of segmentation approach. In contrast, the classical methods provide very similar profiles for all segmentation methods. The differences obtained for AMS estimation are of moderate magnitude. In general, one can say that the SM approach yields results more similar to those of the sliding window algorithm than segmentation into disjoint intervals does.

The differences between the estimated models obtained from 2023 (Figure 8) are significantly smaller; nonetheless, the conclusions remain very similar. SMS estimation remains the most sensitive method to the segmentation approach. For that estimation method, applying the SM approach yields results that are more comparable to those obtained with the original approach. For classical methods and AMS estimation, the differences in results obtained using different segmentation methods are often negligible.

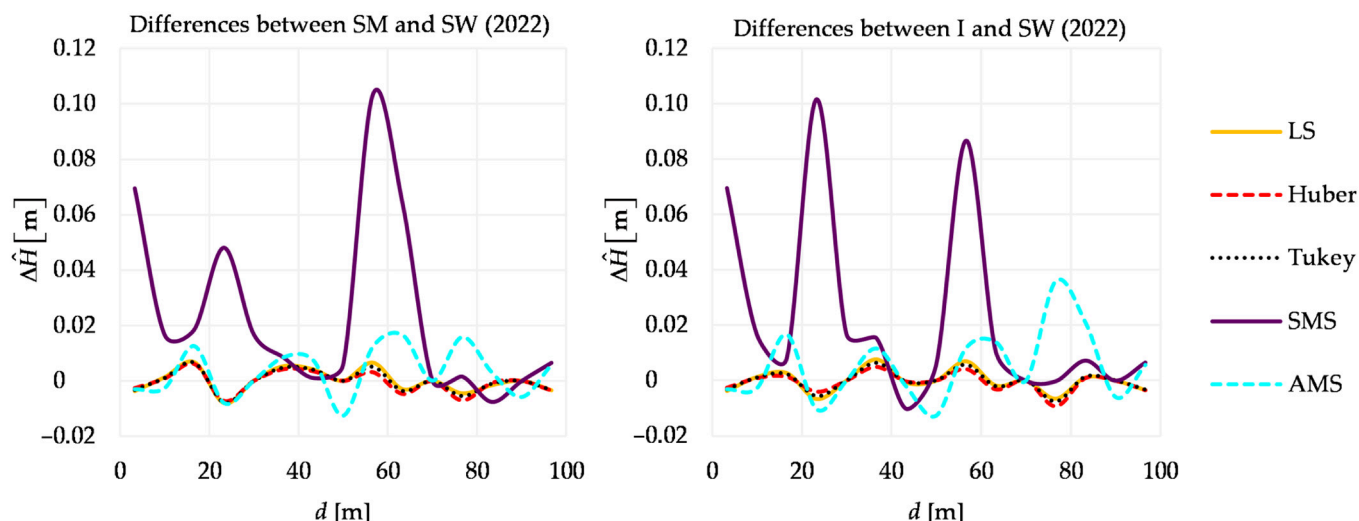


Figure 7. The differences in the point heights between models obtained using the sliding window algorithm (SW) and the sliding window algorithm modification (SM), or the segmentation into disjoint intervals (I), obtained in the year 2022.

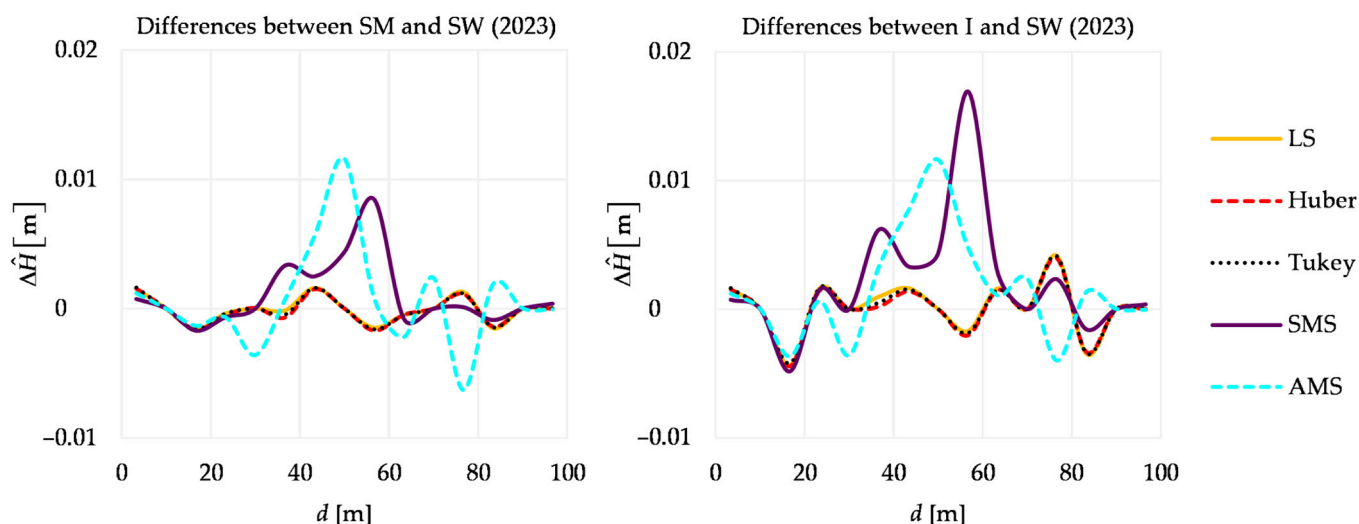


Figure 8. The differences in the point heights between models obtained using the sliding window algorithm (SW) and the sliding window algorithm modification (SM), or the segmentation into disjoint intervals (I), obtained in the year 2023.

4. Discussion

LiDAR systems have become increasingly useful, accurate, and fast, making them highly popular and advantageous in various engineering and scientific applications. Raw data cannot provide satisfactory information about the object of study. To make better use of the collected data, it should be processed and analyzed appropriately. This paper addresses the problem of extracting linear structures from LiDAR point clouds using M_{split} estimation. The given examples demonstrate that methods belonging to that class may outperform classical methods for the specified problem. It especially concerns observation sets contaminated by outlying observations or objects with very complex or varying shapes.

Modeling the linear structures, such as wall edges, cross-sections, or profiles, the observation set is extracted from the entire point cloud and processed to gain the final model. The paper addresses the question of processing such created sets to obtain the most accurate and reliable model. It is proven that processing all observations in one process is admissible when the shape of the object is basic, or when one knows the exact

functional model to be applied (for example, when one knows the degree of a polynomial that accurately describes the shape of the object). Such a case is known from practice, for example, when modeling the deformation of a beam under a load. However, in many problems, complete information on the functional model remains unknown. The paper demonstrates that in such cases, it is more effective to divide the observation set into subsets that are processed separately. It considers three approaches to such processing: segmentation into disjoint intervals, the sliding window algorithm, and a modification of the latter method. The paper proves that when the object shape is very simple, all three approaches yield comparable models. For such objects, it is better to apply wider intervals (or windows). As the object's shape becomes more complex, applying narrower intervals is advisable. Such a conclusion is especially relevant to processing disjoint intervals or applying the modification of the sliding window algorithm.

Generally, the sliding window algorithm is the most accurate (but not always). In several cases, the approach proposed in the paper (modification of the sliding window algorithm) produces more accurate models. It is also noteworthy that both approaches very often yield models of similar accuracy. Processing data in disjoint intervals typically yields less accurate models than both of the mentioned methods. The disadvantage of the sliding window algorithm is time consumption. That method requires 2 to 6 times as much computation time as the other procedures presented.

The tests across different observation sets show that it is difficult to clearly indicate which M_{split} estimation variant yields more accurate models. SMS estimation works better for objects of simpler shapes, whereas AMS estimation is more efficient for more varied objects and seems more reliable for terrain profile estimation. The tests based on real data clearly demonstrate that SMS estimation results depend on the choice of segmentation method, which is particularly evident for lower-accuracy data. Therefore, one can recommend using the sliding window algorithm for that method. The models obtained using AMS estimation and different segmentations are more comparable to one another. This means that, given the model's accuracy and computational time, the sliding window modification proposed in the paper is the best option for processing ALS data using AMS estimation.

5. Conclusions

The conclusion that should be drawn concerns the window's width. The tests performed show that wider windows are recommended for objects of the simplest shapes. It not only shortens computation time but may also increase the accuracy of the models obtained. On the other hand, the more complex the object shape is, the narrower the windows one should apply. The use of overly wide windows might dramatically reduce model accuracy. Narrowing the windows allows for modeling objects in detail; moreover, it can mitigate the negative impact of failing to recognize the appropriate functional model. It is especially important in practice when one usually does not know the object's true shape.

The paper shows that the choice of segmentation method for different variants of M_{split} estimation depends on several factors, including the complexity of the object, the variability of the point cloud, the accuracy of the data, and the share of potential outliers. The presented segmentation methods are appropriate for use with M_{split} estimation; however, other segmentation methods can also be an option. Among the methods described in the Introduction, one can focus on curvature-based adaptive window segmentation or region-growing segmentation. Such approaches might be the subject of our future research.

Author Contributions: Conceptualization, R.D. and P.W.; methodology, R.D. and P.W.; software, P.W.; validation, R.D.; formal analysis, R.D.; investigation, P.W. and R.D.; resources, P.W.; data curation, P.W.; writing—original draft preparation, R.D. and P.W.; writing—review and editing, R.D. and

P.W.; supervision, R.D.; project administration, R.D.; funding acquisition, R.D.; visualization, P.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Department of Geodesy, University of Warmia and Mazury in Olsztyn, Poland, statutory research no. 29.610.001-110.

Data Availability Statement: The original data presented in the study in Sections 3.1–3.3 are openly available in the Knowledge Base of UWM (Baza Wiedzy UWM) at: <https://bazawiedzy.uwm.edu.pl/info/researchdata/UWM6d5b196d19db40498dd8563e607ef5df/> (accessed on 9 July 2024). The data presented in this study in Section 3.4 were derived from Polish database Geoportal at <https://www.geoportal.gov.pl/> (accessed on 9 July 2024).

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

Abbreviations

The following abbreviations are used in this manuscript:

ALS	airborne laser scanning
AMS	absolute M_{split} estimation
AMS _{average}	average result from two solutions of the absolute M_{split} estimation
I	segmentation into disjoint intervals
I5	variant of segmentation into disjoint intervals of 5 m long
I10	variant of segmentation into disjoint intervals of 10 m long
I20	variant of segmentation into disjoint intervals of 20 m long
LiDAR	Light Detection and Ranging
LS	least squares method
RMSD	root-mean-square deviation
SM	sliding window algorithm modification
SM5	variant of the sliding window algorithm modification with a window width of 5 m
SM10	variant of the sliding window algorithm modification with a window width of 10 m
SM20	variant of the sliding window algorithm modification with a window width of 20 m
SMS	squared M_{split} estimation
SMS _{average}	average result from two solutions of the squared M_{split} estimation
SW	sliding window algorithm
SW5	variant of the sliding window algorithm with a window width of 5 m
SW10	variant of the sliding window algorithm with a window width of 10 m
SW20	variant of the sliding window algorithm with a window width of 20 m
TLS	terrestrial laser scanning
W	variant of the whole observation set processing

References

1. Huber, P.J. Robust Estimation of a Location Parameter. *Ann. Math. Stat.* **1964**, *35*, 73–101. [\[CrossRef\]](#)
2. Xu, P. Sign-Constrained Robust Least Squares, Subjective Breakdown Point and the Effect of Weights of Observations on Robustness. *J. Geod.* **2005**, *79*, 146–159. [\[CrossRef\]](#)
3. Duchnowski, R.; Wyszowska, P. Robust Procedures in Processing Measurements in Geodesy and Surveying: A Review. *Meas. Sci. Technol.* **2024**, *35*, 052002. [\[CrossRef\]](#)
4. Yang, Y. Robust Estimation for Dependent Observations. *Manuscr. Geod.* **1994**, *19*, 10–17. [\[CrossRef\]](#)
5. Chang, X.-W.; Guo, Y. Huber's M-Estimation in Relative GPS Positioning: Computational Aspects. *J. Geod.* **2005**, *79*, 351–362. [\[CrossRef\]](#)
6. Khodabandeh, A.; Amiri-Simkooei, A.R.; Sharifi, M.A. GPS Position Time-Series Analysis Based on Asymptotic Normality of M-Estimation. *J. Geod.* **2012**, *86*, 15–33. [\[CrossRef\]](#)
7. Kargoll, B. Comparison of Some Robust Parameter Estimation Techniques for Outlier Analysis Applied to Simulated GOCE Mission Data. In Proceedings of GGSM 2004 IAG International Symposium (Gravity, Geoid and Space Missions), Porto, Portugal, 30 August–3 September 2024; pp. 77–82. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Duchnowski, R. Hodges-Lehmann Estimates in Deformation Analyses. *J. Geod.* **2013**, *87*, 873–884. [\[CrossRef\]](#)

9. Janicka, J.; Rapiński, J. Outliers Detection by RANSAC Algorithm in the Transformation of 2D Coordinate Frames. *Bol. Ciênc. Geodésicas* **2014**, *20*, 610–625. [\[CrossRef\]](#)
10. Li, L.; Yang, F.; Zhu, H.; Li, D.; Li, Y.; Tang, L. An Improved RANSAC for 3D Point Cloud Plane Segmentation Based on Normal Distribution Transformation Cells. *Remote Sens.* **2017**, *9*, 433. [\[CrossRef\]](#)
11. Baselga, S. Global Optimization Solution of Robust Estimation. *J. Surv. Eng.* **2007**, *133*, 123–128. [\[CrossRef\]](#)
12. Batilović, M.; Sušić, Z.; Kanović, Ž.; Marković, M.Z.; Vasić, D.; Bulatović, V. Increasing Efficiency of the Robust Deformation Analysis Methods Using Genetic Algorithm and Generalised Particle Swarm Optimisation. *Surv. Rev.* **2021**, *53*, 193–205. [\[CrossRef\]](#)
13. Błaszczak-Bąk, W.; Sobieraj-Żłobińska, A.; Kowalik, M. The OptD-Multi Method in LiDAR Processing. *Meas. Sci. Technol.* **2017**, *28*, 075009. [\[CrossRef\]](#)
14. Błaszczak-Bąk, W.; Suchocki, C.; Janicka, J.; Dumalski, A.; Duchnowski, R.; Sobieraj-Żłobińska, A. Automatic Threat Detection for Historic Buildings in Dark Places Based on the Modified OptD Method. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 123. [\[CrossRef\]](#)
15. Bogatin, S.; Foppe, K.; Wasmeier, P.; Wunderlich, T.A.; Schäfer, T.; Kogoj, D. Evaluation of Linear Kalman Filter Processing Geodetic Kinematic Measurements. *Measurement* **2008**, *41*, 561–578. [\[CrossRef\]](#)
16. Almagbile, A.; Wang, J.; Al-Rawabdeh, A. An Integrated Adaptive Kalman Filter for Improving the Reliability of Navigation Systems. *J. Appl. Geod.* **2023**, *17*, 295–311. [\[CrossRef\]](#)
17. Wiśniewski, Z. Estimation of Parameters in a Split Functional Model of Geodetic Observations (Msplrit Estimation). *J. Geod.* **2009**, *83*, 105–120. [\[CrossRef\]](#)
18. Wyszowska, P.; Duchnowski, R. Locally Robust Msplrit Estimation. *J. Appl. Geod.* **2025**, *19*, 145–158. [\[CrossRef\]](#)
19. Wyszowska, P.; Duchnowski, R. Msplrit Estimation Based on L1 Norm Condition. *J. Surv. Eng.* **2019**, *145*, 04019006. [\[CrossRef\]](#)
20. Duchnowski, R.; Wyszowska, P. Absolute Msplrit Estimation as an Alternative for Robust M-Estimation. *Adv. Geod. Geoinf.* **2022**, *71*, e17. [\[CrossRef\]](#)
21. Duchnowski, R.; Wiśniewski, Z. Shift-Msplrit Estimation. *Geod. Cartogr.* **2011**, *60*, 79–97. [\[CrossRef\]](#)
22. Zienkiewicz, M.H. Identification of Unstable Reference Points and Estimation of Displacements Using Squared Msplrit Estimation. *Measurement* **2022**, *195*, 111029. [\[CrossRef\]](#)
23. Błaszczak-Bąk, W.; Janowski, A.; Kamiński, W.; Rapiński, J. Application of the Msplrit Method for Filtering Airborne Laser Scanning Data-Sets to Estimate Digital Terrain Models. *Int. J. Remote Sens.* **2015**, *36*, 2421–2437. [\[CrossRef\]](#)
24. Nowel, K. Squared Msplrit(q) S-Transformation of Control Network Deformations. *J. Geod.* **2019**, *93*, 1025–1044. [\[CrossRef\]](#)
25. Janicka, J.; Rapiński, J.; Błaszczak-Bąk, W. Orthogonal Msplrit Estimation for Consequence Disaster Analysis. *Remote Sens.* **2023**, *15*, 421. [\[CrossRef\]](#)
26. Guo, Y.; Li, Z.; He, H.; Zhang, G.; Feng, Q.; Yang, H. A Squared Msplrit Similarity Transformation Method for Stable Points Selection of Deformation Monitoring Network. *Acta Geod. Cartogr. Sin.* **2020**, *49*, 1419–1429. [\[CrossRef\]](#)
27. Li, J.; Wang, A.; Xinyuan, W. Msplrit estimate the relationship between LS and its application in gross error detection. *Mine Surv.* **2013**, *2*, 57–59. [\[CrossRef\]](#)
28. Wyszowska, P.; Duchnowski, R. Processing TLS Heterogeneous Data by Applying Robust Msplrit Estimation. *Measurement* **2022**, *197*, 111298. [\[CrossRef\]](#)
29. Zienkiewicz, M.H.; Czaplewski, K. Application of Square Msplrit Estimation in Determination of Vessel Position in Coastal Shipping. *Pol. Marit. Res.* **2017**, *2*, 3–12. [\[CrossRef\]](#)
30. Duchnowski, R.; Wyszowska, P. Msplrit Estimation Approach to Modeling Vertical Terrain Displacement from TLS Data Disturbed by Outliers. *Remote Sens.* **2022**, *14*, 5620. [\[CrossRef\]](#)
31. Zienkiewicz, M.H.; Dąbrowski, P.S. Matrix Strengthening the Identification of Observations with Split Functional Models in the Squared Msplrit(q) Estimation Process. *Measurement* **2023**, *217*, 112950. [\[CrossRef\]](#)
32. Wyszowska, P.; Duchnowski, R. Msplrit Estimation with Local or Global Robustness against Outliers—Applications and Limitations in LiDAR Data Processing. *Remote Sens.* **2024**, *16*, 4512. [\[CrossRef\]](#)
33. Wyszowska, P.; Duchnowski, R. Sliding Window Algorithm Applied to Msplrit Estimation for Seasonal Change Detection from LiDAR Data. In Proceedings of the 6th Joint International Symposium on Deformation Monitoring, Karlsruhe, Germany, 7–9 April 2025; pp. 1–6. [\[CrossRef\]](#)
34. Nguyen, A.; Le, B. 3D Point Cloud Segmentation: A Survey. In Proceedings of the 2013 6th IEEE Conference on Robotics, Automation and Mechatronics (RAM), Manila, Philippines, 12–15 November 2013; pp. 225–230. [\[CrossRef\]](#)
35. Ding, Z.; Sun, Y.; Xu, S.; Pan, Y.; Peng, Y.; Mao, Z. Recent Advances and Perspectives in Deep Learning Techniques for 3D Point Cloud Data Processing. *Robotics* **2023**, *12*, 100. [\[CrossRef\]](#)
36. Filin, S.; Pfeifer, N. Segmentation of Airborne Laser Scanning Data Using a Slope Adaptive Neighborhood. *ISPRS J. Photogramm. Remote Sens.* **2006**, *60*, 71–80. [\[CrossRef\]](#)
37. Forlani, G.; Nardinocchi, C. Adaptive Filtering of Aerial Laser Scanning Data. In Proceedings of the ISPRS Workshop on Laser Scanning 2007 and SilviLaser 2007, Espoo, Finland, 12–14 September 2007; pp. 130–135.

38. Wang, J.; Shan, J. Segmentation of LiDAR Point Clouds for Building Extraction. In Proceedings of the ASPRS 2009 Annual Conference, Baltimore, MD, USA, 9–13 March 2009; pp. 1–11.
39. Douillard, B.; Underwood, J.; Kuntz, N.; Vlaskine, V.; Quadros, A.; Morton, P.; Frenkel, A. On the Segmentation of 3D LIDAR Point Clouds. In *Proceedings of the 2011 IEEE International Conference on Robotics and Automation, Shanghai, China, 9–13 May 2011*; IEEE: Piscataway, NJ, USA, 2011; pp. 2798–2805. [[CrossRef](#)]
40. Xu, B.; Jiang, W.; Shan, J.; Zhang, J.; Li, L. Investigation on the Weighted RANSAC Approaches for Building Roof Plane Segmentation from LiDAR Point Clouds. *Remote Sens.* **2016**, *8*, 5. [[CrossRef](#)]
41. He, E.; Chen, Q.; Wang, H.; Liu, X. A Curvature Based Adaptive Neighborhood For Individual Point Cloud Classification. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2017**, *XLII-2/W7*, 219–225. [[CrossRef](#)]
42. Li, J.; Xiong, B.; Biljecki, F.; Schrotter, G. A Sliding Window Method For Detecting Corners Of Openings From Terrestrial Lidar Data. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *XLII-4/W10*, 97–103. [[CrossRef](#)]
43. Dey, E.K.; Awrangjeb, M.; Tarsha Kurdi, F.; Stantic, B. Machine Learning-Based Segmentation of Aerial LiDAR Point Cloud Data on Building Roof. *Eur. J. Remote Sens.* **2023**, *56*, 2210745. [[CrossRef](#)]
44. Duchnowski, R.; Wyszowska, P. Laser Scanning Data Processed Using Msplit Estimation and Sliding Window Algorithm. *Rep. Geod. Geoinf.* **2025**, *120*, 67–74. [[CrossRef](#)]
45. Zhang, T.; Kumar, J.; Hoffman, F.M.; Ivanov, V.; Wang, J.; Sheshukov, A.Y.; Zhou, W.; Montesano, P.; Liu, D. A Region-Growing Segmentation Approach to Delineating Timberline from Satellite-Derived Tree Fractional Cover Products. *Remote Sens.* **2025**, *17*, 2002. [[CrossRef](#)]
46. Wiśniewski, Z. Msplit(q) Estimation: Estimation of Parameters in a Multi Split Functional Model of Geodetic Observations. *J. Geod.* **2010**, *84*, 355–372. [[CrossRef](#)]
47. Duchnowski, R.; Wiśniewski, Z. Robustness of Msplit(q) Estimation: A Theoretical Approach. *Stud. Geophys. Geod.* **2019**, *63*, 390–417. [[CrossRef](#)]
48. Wiśniewski, Z.; Zienkiewicz, M.H. Empirical Analyses of Robustness of the Square Msplit Estimation. *J. Appl. Geod.* **2021**, *15*, 87–104. [[CrossRef](#)]
49. Wyszowska, P.; Duchnowski, R. Two Variants of Msplit Estimation—Similarities and Differences. *Adv. Geod. Geoinf.* **2022**, *71*, e22. [[CrossRef](#)]
50. Wyszowska, P.; Duchnowski, R. Performance of Msplit Estimates in the Context of Vertical Displacement Analysis. *J. Appl. Geod.* **2020**, *14*, 149–158. [[CrossRef](#)]
51. Wyszowska, P.; Duchnowski, R. Iterative Process of Msplit(q) Estimation. *J. Surv. Eng.* **2020**, *146*, 06020002. [[CrossRef](#)]
52. Janicka, J.; Rapiński, J.; Błaszczak-Bąk, W.; Suchocki, C. Application of the Msplit Estimation Method in the Detection and Dimensioning of the Displacement of Adjacent Planes. *Remote Sens.* **2020**, *12*, 3203. [[CrossRef](#)]
53. Janowski, A. The Circle Object Detection with the Use of Msplit Estimation. *E3S Web Conf.* **2018**, *26*, 00014. [[CrossRef](#)]
54. Carrilho, A.C.; Galo, M.; Santos, R.C. Statistical Outlier Detection Method For Airborne Lidar Data. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *XLII-1*, 87–92. [[CrossRef](#)]
55. Wyszowska, P.; Duchnowski, R.; Dumalski, A. Determination of Terrain Profile from TLS Data by Applying Msplit Estimation. *Remote Sens.* **2021**, *13*, 31. [[CrossRef](#)]
56. Schmidt, J.; Evans, I.S.; Brinkmann, J. Comparison of Polynomial Models for Land Surface Curvature Calculation. *Int. J. Geogr. Inf. Sci.* **2003**, *17*, 797–814. [[CrossRef](#)]
57. Glennie, C. Rigorous 3D Error Analysis of Kinematic Scanning LIDAR Systems. *J. Appl. Geod.* **2007**, *1*, 147–157. [[CrossRef](#)]
58. Crespo-Peremarch, P.; Tompalski, P.; Coops, N.C.; Ruiz, L.Á. Characterizing Understory Vegetation in Mediterranean Forests Using Full-Waveform Airborne Laser Scanning Data. *Remote Sens. Environ.* **2018**, *217*, 400–413. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.