Assessing vertical terrain displacement from TLS data by applying M\textsubscript{split} estimation – theoretical analysis

Patrycja Wyszowska, Robert Duchnowski

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

Key words: M\textsubscript{split} estimation; terrestrial laser scanning; vertical displacement; deformation analysis; robust estimation

ABSTRACT
Terrestrial laser scanning (TLS) is a measurement technique that has become popular in the last decades. Measurement results, usually as a point cloud, contain many points measured. When the TLS technique is used to determine terrain surface (e.g., by determining terrain profiles), one should realize that some points measured do not concern the terrain surface itself, but trees, shrubs, or generally the vegetation cover. Considering terrain surface determination, they should be regarded as outliers. Some other observations can also be outliers of different origins; for example, they might be disturbed by gross errors. We should consider such observation types when the data are processed. Two leading solutions in such a context are data cleaning and the application of robust estimation methods. Robust M-estimation is the most popular for the latter approach. As an alternative, one can also consider the application of M\textsubscript{split} estimation, in which the functional model is split into two competing ones. Hence, the paper aims to analyze how M\textsubscript{split} estimation can assess vertical terrain displacement based on terrain profile determination from TLS data. We consider processing data in separate sets (two measurement epochs) or one combined set, a natural approach in M\textsubscript{split} estimation. The analyses based on simulated TLS data proved that the first solution seems better. Furthermore, the application of M\textsubscript{split} estimation can also provide more satisfactory results than the classical methods used in such a context.

I. INTRODUCTION
Laser scanning, as a very popular measurement technique, has found many applications in the last decades, e.g., in measuring vegetation, mining displacements, cultural heritage sites, urban environments, buildings, or other objects (Gordon and Lichti, 2007; Wang and Hsu, 2007; Spaete et al., 2011; Rodriguez-González et al., 2017; Crespo-Peremarch et al., 2018; Matwij et al., 2021). Depending on what platform a laser scanner is installed, one can distinguish airborne laser scanning (ALS), terrestrial laser scanning (TLS), or mobile laser scanning (MLS) (Kuzia, 2016). As a typical measurement result, a point cloud contains many points measured that are the basis for further processing. In this paper, we consider the determination of the terrain surface, by determining terrain profiles, from such a TLS result. One should realize some measured points do not concern the terrain surface itself but, for example, vegetation cover (trees, shrubs, etc.), small animals (e.g., birds), or other obstacles. If we focus on terrain surface determination, they should be regarded as so-called positive outliers (Matkan et al., 2014; Carrilho et al., 2018). Outlying observations might also have other origins, such as multipath distance ranging, leading to negative outliers. In some sense, both groups might be considered observations disturbed by gross errors. We should consider the possible occurrence of such observations when processing TLS point clouds. Two main solutions in such a context are data cleaning or application of robust estimation methods, which are well-known in surveying engineering (Hodges and Lehmann, 1963; Baarda, 1968; Pope, 1976; Gui and Zhang, 1998; Huber and Ronchetti, 2009; Duchnowski, 2013; Ge et al., 2013; Lehmann, 2013). As an alternative, one can also consider the application of M\textsubscript{split} estimation. That novel development of M-estimation assumes that the functional model is split into two or more competitive ones (Wiśniewski, 2009; 2010). Such an estimation method allows us automatically isolate observation subsets and estimate their location parameters. That feature might be advisable when processing TLS point clouds that include outliers (Wyszowska et al., 2021).

The paper’s main aim is to analyze how M\textsubscript{split} estimation can be used to assess vertical terrain displacement from estimated terrain profiles based on TLS data. We assume two measurement epochs in such a context, hence two observation sets. We compare two approaches: processing data separately (related to each measurement epoch) or in one combined set, a natural approach in M\textsubscript{split} estimation. We apply two main variants of M\textsubscript{split} estimation, i.e., the squared M\textsubscript{split} estimation, the basic variant, and the absolute M\textsubscript{split} estimation, which was introduced to be less sensitive to outliers than the first variant (Wyszowska and Duchnowski, 2019). For the sake of comparison, we also
use the following classical methods: the least-squares method, and two variants of robust M-estimation, i.e., the Huber method and the Tukey method. The example M-estimation methods differ from each other in general features of the influence functions. The Tukey influence function has rejection points in contrast to the Huber method (Gui and Zhang, 1998; Ge et al., 2013).

II. GENERAL ASSUMPTIONS

While processing or adjusting geodetic observations, we usually apply the conventional linear functional model given as follows (Eq. 1):

\[ y = AX + v \]  \hspace{1cm} (1)

where \( y \) = vector of observations
\( A \) = known matrix of coefficients
\( X \) = vector of parameters
\( v \) = vector of measurement errors

Assuming the same accuracy for each observation, hence the weight matrix \( P \) is equal to the identity matrix, and additionally, that \( A \) is full column rank, the least-squares estimator (LS) \( \hat{X}_{\text{LS}} \) of the vector of parameters can be written as (Eq. 2):

\[ \hat{X}_{\text{LS}} = (A^\top A)^{-1} A^\top y \]  \hspace{1cm} (2)

Similar formula can be written for the M-estimate of the parameter vector when referring to the iteratively reweighted least squares (Eq. 3):

\[ \hat{X}_M = (A^\top WA)^{-1} A^\top Wy \]  \hspace{1cm} (3)

where \( W \) = diagonal matrix of weights for which \( W_{ii} = w(\hat{v}_i) \)
\( w(\hat{v}_i) \) = weight function related to particular M-estimation variant
\( \hat{v}_i \) = standardized error of \( i \)th observation
\( W_{ij} \) = \( i \)th diagonal element of matrix

The form of the weight function depends on the chosen M-estimation variant. In the Huber method it holds that (e.g., Yang, 1994; Ge et al., 2013) (Eq. 4):

\[ w_H(\hat{v}_i) = \begin{cases} 1 & \text{for} \quad |\hat{v}_i| \leq k \\ k \frac{|\hat{v}_i|}{\hat{v}_i} & \text{for} \quad |\hat{v}_i| > k \end{cases} \]  \hspace{1cm} (4)

where \( k \) = positive constant, equals to 2, 2.5 or 3

In this paper, we assume the constant \( k = 2 \). The Tukey weight function is given as (Gui and Zhang, 1998; Ge et al., 2013) (Eq. 5):

\[ w_T(\hat{v}_i) = \begin{cases} 1 - \frac{|\hat{v}_i|}{k} \quad \text{for} \quad |\hat{v}_i| \leq k \\ 0 \quad \text{for} \quad |\hat{v}_i| > k \end{cases} \]  \hspace{1cm} (5)

For the unknown variance, the constant \( k \) is usually assumed equal to 6.

M-estimates of the parameter vector are determined in the iterative process. The end of the iterative process is when the estimated vector of the parameters is not changing between the iteration steps.

Msplit estimation has a different fundamental assumption. The functional model from Equation 1 is split into two competitive ones (Wisniewski, 2009; Wyszkowska and Duchnowski, 2019) (Eq. 6):

\[ y = AX + v \Rightarrow \begin{cases} y = AX_{(1)} + v_{(1)} \\ y = AX_{(2)} + v_{(2)} \end{cases} \]  \hspace{1cm} (6)

where \( X_{(m)} \) = version of vector of parameters
\( v_{(m)} \) = version of vector of measurement errors
\( m = 1 \) or 2

The main difference between the models of Equations 2 and 6 is that we assume two versions of the parameter vector in the latter model; however, the observation vector remains the same. Thus, each observation can be described as either \( X_{(1)} \) or \( X_{(2)} \). Such a situation might stem from the theoretical origin of observations as realizations of two different random variables (which differ from each other in location parameters, here \( X_{(1)} \) and \( X_{(2)} \)). The observation division into two subsets is automatic, and it is performed during the estimation process.

Thus, the goal of Msplit estimation is to estimate \( X_{(1)} \) and \( X_{(2)} \). It can be done by solving the optimization problem with the following general form of the objective function (Wisniewski, 2009) (Eq. 7):

\[ \phi(X_{(1)},X_{(2)}) = \sum_{i=1}^{n} \rho(\nu_{(1,i)},\nu_{(2,i)}) \]  \hspace{1cm} (7)

The component functions \( \rho(\nu_{(1,i)},\nu_{(2,i)}) \) might have different forms depending on the estimation variant. In the case of the squared Msplit estimation (SMS), it holds that \( \rho(\nu_{(1,i)},\nu_{(2,i)}) = \nu_{(1,i)}^2 \nu_{(2,i)}^2 \) as for the absolute Msplit estimation (AMS) \( \rho(\nu_{(1,i)},\nu_{(2,i)}) = \frac{1}{2} \left| \nu_{(1,i)} \right| \left| \nu_{(2,i)} \right| \). The difference between these two functions results in the general features of the Msplit estimation variants and a scheme of the iterative process. In the case of SMS estimation, the estimates can be computed by using the
modified Newton method, the so-called traditional iterative process (Wiśniewski, 2009) (Eq. 8):

\[
X_{i+1}^{(i)} = X_{i}^{(i)} + \alpha X_{i}^{(i)} = \\
= X_{i}^{(i)} - \left[ H_{m}^{(i)} \left( X_{i}^{(i)}, X_{i}^{(i)} \right) \right]^{-1} \left[ g_{m}^{(i)} \left( X_{i}^{(i)}, X_{i}^{(i)} \right) \right]
\]  

where \( \alpha \) is increment to vector of parameters

\[
H_{m}^{(i)} \left( X_{i}^{(i)}, X_{i}^{(i)} \right) = 2A^{T}w_{m}^{(i)} \left( v_{i1}, v_{i2} \right)A
\]

is Hessian

\[
g_{m}^{(i)} \left( X_{i}^{(i)}, X_{i}^{(i)} \right) = -2A^{T}w_{m}^{(i)} \left( v_{i1}, v_{i2} \right) \text{gradient}
\]

\[w_{m}^{(i)} \left( v_{i1}, v_{i2} \right) = \text{diagonal matrix}
\]

The weight functions stem from the functions \( \rho \left( v_{i1}, v_{i2}, v_{i1}, v_{i2} \right) \) by holding the following relations (Wiśniewski, 2009) (Eq. 9):

\[
\begin{align*}
\psi_{1}^{(i)} \left( v_{i1}, v_{i2} \right) & = \frac{\psi_{1}^{(i)} \left( v_{i1}, v_{i2} \right)}{2v_{i1}} \\
\psi_{2}^{(i)} \left( v_{i1}, v_{i2} \right) & = \frac{\psi_{2}^{(i)} \left( v_{i1}, v_{i2} \right)}{2v_{i2}}
\end{align*}
\]

where the influence functions \( \psi_{1}^{(i)} \left( v_{i1}, v_{i2} \right) \) and \( \psi_{2}^{(i)} \left( v_{i1}, v_{i2} \right) \) are defined as (Eq. 10):

\[
\begin{align*}
\psi_{1}^{(i)} \left( v_{i1}, v_{i2} \right) & = \frac{\partial \rho \left( v_{i1}, v_{i2} \right)}{\partial v_{i1}} \\
\psi_{2}^{(i)} \left( v_{i1}, v_{i2} \right) & = \frac{\partial \rho \left( v_{i1}, v_{i2} \right)}{\partial v_{i2}}
\end{align*}
\]

Hence, for SMS estimation, it holds that (Eq. 11):

\[
\begin{align*}
\psi_{1}^{(i)} \left( v_{i1}, v_{i2} \right) & = v_{i1}^{2} \\
\psi_{2}^{(i)} \left( v_{i1}, v_{i2} \right) & = v_{i2}^{2}
\end{align*}
\]

In the case of AMS estimation, we use the parallel iterative process conducted in the following way (Wyszkowska and Duchnowski, 2019) (Eq. 12):

\[
X_{i+1}^{(i)} = X_{i}^{(i)} + \alpha X_{i}^{(i)} = \\
= X_{i}^{(i)} - \left[ H_{m}^{(i)} \left( X_{i}^{(i)}, X_{i}^{(i)} \right) \right]^{-1} \left[ g_{m}^{(i)} \left( X_{i}^{(i)}, X_{i}^{(i)} \right) \right]
\]

The respective Hessians and gradients are computed as in Equation 7 and by applying the following weight functions resulting from Equation 9 (Wyszkowska and Duchnowski, 2019) (Eq. 13):

\[
w_{m}^{*} \left( v_{i1}, v_{i2} \right) = \begin{cases} \\
\frac{\psi_{1}^{(i)}}{2c} & \text{for } |v_{i1}| < c \\
\frac{\psi_{2}^{(i)}}{2c} & \text{for } |v_{i2}| < c
\end{cases}
\]

where \( c \) is assumed small positive constant.

Both iterative processes end when the gradients are equal to zero, in other words when both estimated vectors of the parameters are not changing between the iteration steps. More information about iterative processes of AMS estimation can be found in (Wyszkowska and Duchnowski, 2020).

### III. Numerical Example

The methods introduced in the previous section are now applied to process simulated TLS data to determine vertical displacements of terrain.

Let us assume a terrain profile of 50 m long, which is measured at two epochs. The observation sets are simulated by taking all theoretical heights of the terrain profile at the first epoch equal to 0 m. We assume some vertical movements of the terrain, which can be described by the following heights of the terrain profile at the second epoch: 0.005 m at a distance \( D = 0 \) m, \( D = 10 \) m, \( D = 30 \) m, and \( D = 40 \) m. It allows us to describe the complete profile in the second epoch by the third-degree polynomial with the following coefficients: \( a_{0} = 5.83 \times 10^{-7} \), \( a_{2} = -3.83 \times 10^{-5} \), \( a_{1} = 6.25 \times 10^{-4} \), and \( a_{3} = 5 \times 10^{-3} \). Then, let us simulate 500 observations, TLS points, at each measurement epoch. The measurement errors are assumed to be normally distributed \( N \left( 0 \right) \). We also assume that some observations might be affected by gross errors generated from the uniform distribution in the interval from 0.005 m to 0.050 m. Thus, we assume that all outliers are positive.

In this section, we consider the following variants of the observation set:

- **Variant I** – lack of outliers in epochs I and II.
- **Variant II** – 10% outliers in epochs I and II.
- **Variant III** – 10% outliers in epoch I, 30% outliers in epoch II.
Observation sets for these variants are presented in Figures 1-3. The top panel of each figure shows observations from the measurement epoch I, the middle panel—observations from the measurement epoch II, and the bottom panel—observations from both epochs (the combined set).

First, let the polynomial coefficients be estimated in each epoch separately. Then, one can compute the estimated profile heights (every meter) at each epoch. It allows us to calculate the terrain displacements obtained by subtracting the respective point heights. Note that in the case of $M_{split}$ estimation, we always get two competitive solutions in each epoch.

Considering the terms of the case at hand (mainly, the terrain profile estimation), the polynomial in which the graph is located lower is taken as the $M_{split}$ solution. The second solution describes the location of outliers, and here is omitted.

The results are presented in Figure 4: Variant I in the top panel, Variant II in the middle panel, and Variant III in the bottom panel. All estimation methods provide very similar and correct results for the first two variants.

It is noteworthy that even LS estimation seems to show robustness against outliers. It stems from the character of outlying observations (they are spread over the whole profile randomly and uniformly). The results obtained in each epoch, which are not presented here, are disturbed by outliers significantly; however, after subtracting results from two epochs, such an effect is eliminated, resulting in the correct estimation of terrain displacements. The significant differences between different method applications are in Variant III. LS estimation, both M-estimation variants, and SMS estimation do not cope with the different shares of outlying observations in both measurement epochs. Robust M-estimation cannot deal with 30% outliers at the second measurement epoch. Only AMS estimation provides correct results in such a case. Thus, it predominates over conventional robust M-estimation for a higher percentage of outliers. AMS estimation also provides better results than SMS estimation.
results more detailedly, let us assess how the estimated displacements fit into the simulated ones. The fit accuracy can be evaluated by using the root-mean-square deviation $RMSD(\hat{H})$ in the following form (Wyszkowska et al., 2021) (Eq. 14):

$$RMSD(\hat{H}) = \sqrt{\frac{1}{n} \sum_{i=0}^{n} (\Delta H_i - \hat{\Delta H}_i)^2}$$

where $\Delta H_i$ = estimated displacements, $\hat{\Delta H}_i$ = simulated displacements, $n = 51$ = number of points for which displacements are calculated for distances $D_j = j \cdot 1 \text{ m} \ (j = 0, ..., 50)$.

Figure 3. Observation sets for Variant III.

RMSDs for Variants I-III are presented in Table 1. When there are no outliers, LS estimation provides the best results; however, the results of robust M-estimation variants are only a little bit worse. The results of $M_{\text{split}}$ estimation variants are also comparable (especially for AMS estimation). LS estimation provides the least accurate outcomes in Variant II, whereas other methods give similar accuracy. In the case of Variant III, the results of AMS estimation are significantly better than the results of other variants considered. Note that robust M-estimation and SMS estimation give outcomes that are less accurate than the results of LS estimation. It means that those methods have not coped with 30% outliers in the second epoch definitively.

Table 1. RMSDs of estimated displacements [mm]

<table>
<thead>
<tr>
<th>Variant</th>
<th>LS</th>
<th>Huber</th>
<th>Tukey</th>
<th>SMS</th>
<th>AMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>0.29</td>
<td>0.33</td>
<td>0.30</td>
<td>0.50</td>
<td>0.43</td>
</tr>
<tr>
<td>II</td>
<td>0.38</td>
<td>0.20</td>
<td>0.13</td>
<td>0.25</td>
<td>0.15</td>
</tr>
<tr>
<td>III</td>
<td>5.62</td>
<td>5.63</td>
<td>5.68</td>
<td>10.13</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Figure 4. Estimated vertical terrain displacement (between values from two sets) for different variants.

In many practical applications of $M_{\text{split}}$ estimation, we apply only one observation set, namely the set consisting of two (or more) observation groups. Such an approach is natural for $M_{\text{split}}$ estimation and advisable in many surveying problems (Błaszczak-Bąk et al., 2015;
Nowel, 2019; Guo et al., 2020; Wyszkowska et al., 2021. It could also be involved in deformation analysis (Zienkiewicz, 2015; 2019). Let us use that way in the present example. Thus, we apply the combined sets (presented in the bottom panels of Figures 1-3) to obtain Msplit estimates of the polynomial coefficients in one estimation process. The obtained competitive solutions refer to both epochs, respectively. One could evaluate the estimated terrain displacement using Equation 14. Table 2 presents RMSDs of the fit of the estimated displacements to the simulated ones.

Table 2. RMSDs of estimated displacements from the combined observation set [mm]

<table>
<thead>
<tr>
<th>Variant</th>
<th>SMS</th>
<th>AMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>15.17</td>
<td>10.88</td>
</tr>
<tr>
<td>II</td>
<td>26.32</td>
<td>1.82</td>
</tr>
<tr>
<td>III</td>
<td>28.87</td>
<td>21.18</td>
</tr>
</tbody>
</table>

The obtained values clearly indicate that the estimation based on the combined observation sets provides much less accurate results. It might seem that SMS and AMS estimation cannot deal with outliers. However, results obtained in Variant I are also very poor. It means that the “separation” between regular observations from the first and the second epochs, respectively, is not enough to be easily detected. Thus, the estimation process is disturbed, resulting in the low accuracy of the final outcomes. Probably, if the terrain displacements were more significant (in comparison to the measurement accuracy), the Msplit estimation would provide better results.

IV. CONCLUSIONS

Msplit estimation is a modern method dedicated to processing observation sets consisting of two (or more) observation groups. For that reason, it is applied to TLS, or more generally LiDAR, data processing. On the other hand, the method might also be used for deformation analysis. This paper answers the question is the method applicable in determining vertical terrain displacement from TLS data. From a theoretical point of view, there are no obstacles. It is shown that such an approach seems interesting and advisable, mainly when some outlying observations might occur. It is essential because outliers often concern TLS point clouds. Usually, observations of such a type do not result from gross errors but mismeasured points (measurement of different objects, obstacles, or surfaces that we are not interested in).

The presented example proves that Msplit estimation, especially AMS estimation, could be an alternative to classical methods. It can cope with a higher share of outliers than robust M-estimation can. The condition is that we should have a priori information about the location of a line or a surface under investigation. In the case of TLS data, we almost always know which estimated line or surface (from two competitive variants) we should choose as the final outcome. For example, here, we always choose the lower estimated line, which concerns the terrain profile for sure.

The paper also checks two approaches to processing TLS data from two epochs by applying Msplit estimation. There is no doubt that processing observation sets from both epochs separately is a much better approach than using the combined observation set.

V. ACKNOWLEDGEMENTS

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

References


