https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021

Geoscientific Model Development Discussions

Nonparametric-based estimation method for river cross-sections with point cloud data from UAV photography

URiver-X version 1.0 -methodology development

Keywords: Nonparametric, UAV, Regression, Point Cloud, River, Cross Section

Taesam Lee[1] and Kiyoung Sung[1]

[1] Department of Civil Engineering, ERI, Gyeongsang National University,
501 Jinju-daero, Jinju, Gyeongnam, South Korea, 660-701

Corresponding Author :

Taesam Lee, Ph.D.
Gyeongsang National University, Dept. of Civil Engineering
Tel)+82-55-772-1797, Fax)+82-55-772-1799
Email) tae3lee@gnu.ac.kr

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021

Geoscientific Model Development Discussions

Abstract

Aerial surveying with unmanned aerial vehicles (UAVs) has been popularly employed in river management and flood monitoring. One of the major processes in UAV aerial surveying for river applications is to demarcate the cross-section of a river. From the photo images of aerial surveying, a point cloud dataset can be abstracted with the structure from motion (SfM) technique. To accurately demarcate the cross-section from the cloud points, an appropriate delineation technique is required to reproduce the characteristics of natural and manmade channels, including abrupt changes, bumps, and lined shapes, even though the basic shape of natural and manmade channels is a trapezoidal shape. Therefore, a nonparametric-based estimation technique, called the K-nearest neighbor local linear regression (KLR) model, was tested in the current study to demarcate the cross-section of a river with a point cloud dataset from aerial surveying. The proposed technique was tested with a simulated dataset based on trapezoidal channels and compared with the traditional polynomial regression model and another nonparametric technique, locally weighted scatterplot smoothing (LOWESS). Furthermore, the KLR model was applied to a real case study in the Migok-cheon stream, South Korea. The results indicate that the proposed KLR model can be a suitable alternative for demarcating the cross-section of a river with point cloud data from UAV aerial surveying by reproducing the critical characteristics of natural and manmade channels, including abrupt changes and small bumps, as well as the overall trapezoidal shape.

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021


Geoscientific Model Development Discussions

1. Introduction

Unmanned aerial vehicles (UAVs) have been popularly employed in recent years, especially to investigate and survey earth systems such as agriculture and coastal areas (Hugenholtz et al., 2013; Lin et al., 2018; Marfai et al., 2019; Remondino et al., 2011; Siebert and Teizer, 2014; Srivastava et al., 2020; Taddia et al., 2021; Wang et al., 2019; Watanabe and Kawahara, 2016; Yan et al., 2021). Furthermore, river management and fluvial networks have received critical attention for UAV applications (Gracchi et al., 2021; Langhammer, 2019; Lee et al., 2019; Sanhueza et al., 2019; Tomsett and Leyland, 2019). Additionally, flood monitoring and assessment are one of the major fields in which UAV aerial surveying data have been used (Anders et al., 2020; Andreadakis et al., 2020; Izumida et al., 2017; Kaewwilai, 2019; Perks et al., 2016; Zakaria et al., 2018).

For example, Andreadakis et al. (2020) employed a combination of SfM and optical granulometric techniques in estimating peak discharge and illustrated that the combined UAV technique can accurately determine peak discharge. Anders et al. (2020) tested different flying altitudes and area coverage orientations in semiarid and medium-relief areas with respect to cell size and vertical and horizontal accuracy. Perks et al. (2016) applied a novel algorithm to track features associated with free-surface velocity and to allow accurate geometric correction of velocity vectors. Those applications of UAV photographs to measure and analyze floods are based on the SfM technique.

The structure for a motion technique produces 3D information from overlapping images, where the structure refers to the relative parameters of aerial surveying, such as camera positions and focal lengths, and the relative positions of the corresponding features, while the motion refers to the movement of the camera (Javernick et al., 2014; Marteau et al., 2017; Smith et al., 2014). A

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021


Geoscientific Model Development Discussions

dense point cloud can be determined from the SfM. These point clouds are converted from an arbitrary coordinate system to a geographical coordinate system with camera position and focal length information or by associating reference points on the ground, called ground control points (GCPs), with known coordinates. A point cloud is a set of 3-dimensional points located in space. The 3D locations of a point cloud can be determined from a sensor by emitting pulses and calculating them with the position of the sensor and the pulse direction. Here, the sensor refers to a photogrammetry camera in the current study.

In UAV aerial surveying applications for river management and flood analysis, the demarcation of the cross-section of a river is the critical process. Accurate demarcation of the cross-section is mostly required to calculate the peak discharge and flow amount. However, the dense cloud point dataset obtained from UAV aerial surveying and the SfM technique mostly contains errors and does not provide direct cross-sectional information. An appropriate technique to demarcate the cross-section from the point cloud dataset is necessary to develop.

The demarcation of the cross-section in a river has been mostly made with a digital elevation model (DEM) in the literature (Gichamo et al., 2012; Petikas et al., 2020a, b; Pilotti, 2016). For example, Petikas et al. (2020b) proposed a novel method to automatically extract river cross-sections from a DEM along with a parametric cross-section extraction algorithm. However, a cross-sectional algorithm for the cloud point dataset of UAV aerial surveying has not been tested in depth since the characteristics of the point cloud dataset are far different from the DEM in that a study area for UAV aerial surveying is commonly smaller and many more points can be acquired from UAV aerial surveying.

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021


Geoscientific Model Development Discussions

89 Therefore, the current study proposes a demarcation technique for river cross-sections from
90 the point clouds of UAV aerial surveying. A cross-section of natural rivers contains abrupt changes
91 and small bumps as well as smooth variations even though it normally has a trapezoidal shape.
92 The demarcation technique must reproduce the characteristics of natural rivers as well as the abrupt
93 changes in a manmade channel. The proposed demarcation model was tested to determine whether
94 to reproduce the characteristics of natural river and manmade channels.

95

96 2. Mathematical Description

97 With the point cloud data obtained from UAV aerial surveying and postprocessing, the river
98 cross-section must be demarcated. Polynomial regression can be simply applied to the point data.
99 However, the fixed shape of the polynomial regression along with its order is limited to the highly
100 varied shape of the cross-section. Therefore, a nonparametric regression approach is adopted in
101 the current study, especially K-nearest neighbor local regression. A detailed description of
102 polynomial regression and the nonparametric regression model is shown in the following.

103 2.1. Polynomial Regression

104 A polynomial regression model can be used when the relationship between a predictor (x) and an
105 explanatory variable (y) is nonlinear or curvilinear. The kth-order polynomial regression can be
106 expressed as

107 $$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_k x^k + \epsilon = \sum_{i=0}^{k} \beta_i x^i + \epsilon = \mathbf{x\beta} + \epsilon \qquad (1)$$

108 where ϵ is considered to be random noise with zero mean.

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021

Geoscientific Model Development Discussions

2.2. KNN-based Local Linear Regression (KLR)

Assume that the current condition of the predictors x_t with the observed data pairs (x_i,y_i), for i = 1,…,n, is given for the n number of data points (i.e., the selected cloud points). The number of neighbors (k) is also assumed to be known. The predictor Y_t is estimated according to the following steps:

(a) Estimate the distances between the current and observed states of the predictors for all n observations, as follows:

$$D_j = \left(x_j - x_t\right)^2 \qquad j=1...,n \quad (2)$$

(b) Store the location indices for the k smallest distances.

(c) Fit the local linear regression to the observed dataset of the selected location indices $[x_{(p)}, y_{(p)}]$ for p = 1,…,k, where (p) indicates the p^{th} decreasing ordered location index relative to the distance measure in step (a).

(c-1) Build the weight matrix using the simple selection weight as follows:

$$\boldsymbol{W}_{KLR} = diag\left[\frac{1}{\delta}, \frac{1/2}{\delta}, \ldots, \frac{1/k}{\delta}\right] \quad (3)$$

where $\delta = \sum_{p=1}^{k} 1/p$.

$$\vec{\mathbf{X}}_t = \begin{pmatrix} 1 & x_t - x_{(1)} \\ 1 & x_t - x_{(2)} \\ 1 & \vdots \\ 1 & x_t - x_{(k)} \end{pmatrix} \quad (4)$$

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021


Geoscientific Model Development Discussions

(c-3) Estimate the parameter vector $\widehat{\boldsymbol{\beta}}_t^{KLR}$ with the weighted least square estimator from the weight matrix $\mathbf{W}_{KLR}$ in Eq. (3) as

$$\widehat{\boldsymbol{\beta}}_t^{KLR} = (\vec{\mathbf{X}}_t^{\,T} \boldsymbol{W}_{KLR} \vec{\mathbf{X}}_t)^{-1} \vec{\mathbf{X}}_t^{\,T} \boldsymbol{W}_{KLR} \mathbf{y}_{KLR} \quad (5)$$

where $\mathbf{y}_{KLR}$ is the corresponding predicted value for the ordered observations $[y_{(1)}, y_{(2)}, \ldots, y_{(k)}]^T$.

(d) Estimate the current predictor as follows:

$$y_t = \vec{\mathbf{x}}_t^T \widehat{\boldsymbol{\beta}}_t^{KLR} \quad (6)$$

where $\vec{\mathbf{x}}_t = (1 \quad x_t)$.

(e) Repeat steps (a)-(d) until the required data are simulated.

For selecting the number of neighbors *k*, a heuristic approach for estimating *k* for the KNNR model is given by $k = \sqrt{n}$ (Lall and Sharma, 1996; Lee and Ouarda, 2011; Lee et al., 2010). Therefore, Lee et al. (2017) suggested a heuristic approach for KLR in which they suggested that the multiplier be

$$k = a\sqrt{n} \quad (7)$$

where a is a multiplier and is a positive integer (i.e., 1,2,3,4).

As noted, only the partial dataset is employed for the observations rather than the whole observation dataset, unlike other regressions. For the point cloud dataset from UAV photography, this proposed approach in the current study is highly advantageous since the neighboring data point

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021


Geoscientific Model Development Discussions

is sufficient and the fitting of the target point must not be affected by the points that are far away from the target point. This advantage is further discussed in the results section.

3. Simulation Study

The performance of the KLR model in fitting the point cloud data for river cross-sections is tested with the simulated point cloud data.

3.1. Model Description and Fitting

A river cross-section is generally trapezoidal due to maximum discharge and easy construction (Chow, 1959). Therefore, a trapezoidal channel was assumed with a 4 m top at both sides and a 6 m base width as well as a 1:1 side slope with a 6 m height, as shown by the thick solid blue line of Figure 1. The channel points were assumed to be measured with 0.1 m intervals, for a total of 161 points. It is assumed that these points work as cloud points that UAV cameras might capture in aerial surveying. The assumed cloud point dataset was generated based on the assumed 161 points (see the thick solid blue line in Figure 2), as follows:

$$Z = Y + \varepsilon \tag{8}$$

where Y is the assumed points, and $\varepsilon \sim N(0, \sigma_\varepsilon^2)$, i.e., normally distributed error. Note that the generated data (Z) are presented with red circles in Figure 1.

In the current study, σ_ε^2=0.2 was used following similar variability of observed data after testing several values. The magnitude of this error variance (σ_ε^2) represents the differences in the photo locations for the same cloud point of the real ground location (i.e., Y in this case). High variance indicates that extracted point clouds include high errors, and vice versa.

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021


Geoscientific Model Development Discussions

In Figure 1 and Figure 2, the simulated data are presented with red circles. The number of simulated data points was chosen to be 2 times and 10 times the assumed 161 points that were applied for the assumable measured trapezoid line (i.e., 322 and 1610 points), as shown in Figure 1 and Figure 2, respectively. Note that the recommended overlap is 70–80% frontal and 60% side in general cases. In this overlapping case, each cross-section point might be captured approximately 10 times. Therefore, the number of simulated data points is set to 10 times the number of trapezoidal channel points (a total of 1610), as shown in Figure 2. Additionally, there are some portions in which overlapping might not be achieved. Minimal overlap to be a point cloud is at least 2 times, and 2 times the channel points were also tested, as shown in Figure 1.

3.2. Simulation Results

In Figure 1, the fitted cross-section line to the KLR model is shown with the dashed black line for the generated data case with 2 times the assumed target points, while the simulated data are presented with the red circles as noted. Note that the multiplier (a) for the number of nearest neighbors as in $k= a\sqrt{n}$ in Eq. (7) was tested in this figure. As shown in Figure 1, the multiplier a= 1, 2, 3, 4 is shown in each panel. The fitted KLR line with a smaller multiplier presents more irregularity, while the line with a higher multiplier appears to be smooth. For example, the top part of the trapezoid channel with the 22 m of the y-coordinate shows that the KLR line with the multiplier a=1 (the panel(a) of Figure 1) was drawn rather coarse, but the straight shape of the original channel is preserved. At the same time, the fitted line with the high multiplier a=4 (Panel (d) of Figure 1) presents a very smooth feature and presents the original top and bottom horizontal parts, which are rather too curved.

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021


Geoscientific Model Development Discussions

185 The multipliers of *a*=2 and 3 in the fitted KLR model, as shown (Panels (*c*) and (*d*) of Figure
186 1), appear to mix the smooth and horizontal features well by fitting the top and bottom horizontal
187 lines, and the angled part of the original channel is reproduced well. This finding indicates that an
188 appropriate multiplier (*a* in Eq. (7)) is required to present the straight and angled trapezoid channel
189 better.

190 This characteristic continues to the case with the high number of captured cloud points, as
191 shown in Figure 2 (i.e., 10 times the target points, for a total of 1610 points, as shown with the red
192 circles in this figure). It is comparable to the case of 2 times the target points in Figure 1 in that all
193 of the fitted line with the KLR model with the case of 10 times presents better the original trapezoid
194 channel than the case of 2 times. It is obvious that a higher number of points can significantly
195 improve the quality of the KLR model since the nonparametric KLR model directly applies the
196 observed data and its performance highly depends on the number of data points. In other words,
197 while parametric models such as linear regression and polynomial regression estimate the
198 parameters from the data and the parameters are employed, the nonparametric KLR model
199 employs the data itself directly to estimate the cross-section. It can be appreciated that UAV aerial
200 photography usually captures a large enough number of points to produce overlapping points as
201 many as 10 times the target points.

202 The horizontal and angled trapezoid shape (i.e., the solid thick blue line in Figure 2) is
203 reproduced well by the KLR model (see the dashed black line), even though a coarse zig-zag line
204 is still observed in the case of the small multiplier (i.e., *a*=1, see Panel (*a*) of Figure 2), and the
205 angled portion is too curved in the case of the high multiplier (i.e., *a*=4, see Panel (*d*) of Figure 2).

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021


Geoscientific Model Development Discussions

206 The results of Figure 1 and Figure 2 illustrate that a value between 2 and 3 can be a good
207 selection for the multiplier. Further testing was performed to select the multiplier for the number
208 of nearest neighbors by varying the multiplier from 0.5 to 5.0 with a 0.5 interval. The root mean
209 square error was estimated as

210 $$RMSE = \sqrt{\frac{1}{\mathrm{N}}\sum_{t=1}^{N}(y_t^{KLR} - Y_t)^2} \quad (9)$$

211 where y_t^{KLR} is the KLR estimate from Eq. (7), and Y_t represents the original points with N
212 trapezoid points (here, 161 points).

213 The RMSE results of the KLR estimate with different multipliers (i.e., *a* in Eq. (7)) are shown
214 in Figure 3 for the case of 2 times (top panel) the original trapezoidal points and the case of 10
215 times (bottom panel). In the 2 and 10 times cases, the optimum multiplier (i.e., the smallest
216 multiplier *a*) can be selected to be between 1.5 and 2.5. To fully reveal the characteristics of the
217 multiplier with multiple simulations, all of the multiple simulations from 1 to 12, indicating the
218 number of overlapped photos, were tested while finding the optimum multiplier. The result in
219 Figure 4 shows that a smaller optimum multiplier is selected with a smaller number of overlapped
220 photos (or multiple simulation points) as much as 1.5–2.0, and vice versa as much as 2.0–2.5.
221 Since the number of overlapping photos might be difficult to know and each point does not have
222 the same number of points in real UAV aerial survey, the multiplier is suggested to be 2.0 in the
223 current study.

224 To compare other approaches to fit the point cloud in demarcating the cross-section of a river,
225 polynomial regression and locally weighted scatterplot smoothing (LOWESS) were also tested.
226 The result is presented in the top and bottom panels of Figure 5 and Figure 6 for the fitted line to

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021


Geoscientific Model Development Discussions

the polynomial regression (top panel) and the LOWESS model, respectively. The fitted line to the polynomial regression of 2nd degree (see the thick dash-dotted yellow line with the circle marker in the top panel of Figure 5 and Figure 6) does not reproduce the top and bottom horizontal lines of the trapezoid channel well. Better performance in the 4th degree polynomial regression model is presented (see the dotted b line with the reverse triangle marker). However, the depth of the trapezoid center is overestimated. Other degrees of polynomial regression models were also tested, but no better performance was observed.

Furthermore, the LOWESS model was additionally fitted to the simulated trapezoid channel data. Note that the LOWESS model is also a nonparametric regression model, and its detailed description is presented in the Appendix. The major difference between the LOWESS model and the KLR model is that the LOWESS model includes all of the observed data in the estimate, as shown in Eqs. (A.3) and (A.4), while the KLR model includes only the k-nearest neighbor observations, as in Eq. (4). The performance of the LOWESS and KLR models was compared in detail in Lee et al. (2017) for the heteroscedastic relation of time series data. The result in the study of Lee et al. (2017) indicated that the KLR model reproduces an abrupt change in the heteroscedastic relation. The results of the LOWESS model are presented in the bottom panels of Figure 5 and Figure 6. The results indicate that the bottom part of the trapezoidal channel is reproduced well with the LOWESS model. However, the model does not reproduce the abruptly curved area well.

Further nonparametric models such as LOWESS and other regression models such as logistic regression (Ahmad et al., 1988; Elek and Márkus, 2004; Orlowsky et al., 2010; Simonoff, 1996) can be tested. However, the simulation study with the trapezoid channel that is similar to the real river cross-section shows that the presented KLR nonparametric model is suitable for demarcating

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021


Geoscientific Model Development Discussions

the cross-section of a river. The major reason for the good performance is that the KLR model employs only k-nearest neighbor observations. This approach might not be beneficial when an overall trend is needed and not enough observations are available. However, the point cloud data taken from UAV aerial surveying often provide a large enough number of points in the data set. Furthermore, the shape of the cross-sections in a natural river is irregular, and abrupt changes can be easily observed. This feature can be captured only through fitting nearby observations. Therefore, the KLR model might be a suitable alternative to demarcating the cross-section of a river with the cloud point dataset.

4. Case Study

4.1. Study Area and Data Acquisition

4.1.1. Study Area

The study area is located in the Migok-cheon stream flowing through Hapcheon-gun, South Korea, as shown in Figure 7. The Migok-cheon stream has an 8.8 km length and 13.9 km^2 watershed area. The slope of the stream is approximately 1/50~1/400, and the study area has a slope of 1/350. This stream conjuncts to the Hwanggang River at the end of the stream, and the Hwanggang River is joined into the Nakdong River directly afterward; the Nakdong River is one of the four largest rivers in South Korea. Therefore, the Migok-cheon stream is highly affected by the water levels of the Hwanggang River and Nakdong River.

In the middle of the Hwanggang River, the Hapcheon dam is located for electric generation and water resources. The upstream Hapcheon River consists of a number of mountains, and the slope is high, producing rapid floods and short concentration times to induce floods. For example,

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021


Geoscientific Model Development Discussions

271 in August 2020, the Hapcheon dam outflowed a large amount of water downstream and induced a
272 high water level in the Hwanggang River. A number of streams joining the Hwanggang River
273 overflowed due to the high level of the Hwanggang River, including the Migok-cheon stream
274 (Seong et al., 2020). To reduce damage from floods in the area of the Migok-cheon stream, an
275 early warning system for floods is being considered. For the early-warning system for floods,
276 detailed cross-sections of the Migok-cheon stream must be requested to decide which water level
277 is appropriate for an alarm.

278 *4.1.2. Data Acquisition*

279 **Specification of Employed UAV**

280 Aerial photos over the selected Migok-cheon were obtained with the unmanned aerial vehicle (also
281 termed drone), DJI Phantom 4. This UAV is one of the most popular professional drones on the
282 market and contains an advanced stereo vision positioning system that provides precise hovering
283 even without satellite positioning support (Hamdi et al., 2019). The camera applied is FC3411 with
284 ISO-110, and the images taken from DJI Phantom 4 are 5472x3648 pixels at approximately 10 M.
285 Pix4Dcapture was employed to map the target area.

286 **Ground Control Points**

287 Ground control points (GCPs) are the points on the ground that have measured or known
288 coordinates. To obtain GCPs, 10 specific points were measured over the target area on the ground
289 with global positioning system (GPS) surveying. The EMLID Reach RS2
290 (https://emlid.com/reachrs2/), multiband RTK GNSS receiver, with centimeter precision, was
291 employed for GPS surveying.

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021


Geoscientific Model Development Discussions

292 **Data Processing (WebODM)**

293 The aerial photos were postprocessed to build a point cloud dataset with WebODM. The WebODM
294 (https://github.com/OpenDroneMap/WebODM) is an open source tool for generating map point
295 clouds, terrain and 3D surface models from aerial images.

296 ## 4.2. Distance Measurement of the Point Cloud

297 The point cloud data from UAV photography are presented with Transverse Mercator (TM)
298 projection for x, y, and z. The TM projection is a conformal projection presented by Lambert in
299 1772. To demarcate a cross section of a river, the point cloud data must be projected to a new
300 coordinate system.

301 As an example in Figure 8, the new coordinate system can be based on the line that connects
302 N and L points presented with the thick red line in Panel (a). The extended thick red line is
303 designated as the new x-coordinate, as shown in Panel (b), and the same z-axis can be defined as
304 the original TM data. The y-coordinate can be chosen as the axis that is perpendicular to the x-
305 coordinate. Let us assume that point M, as in Panel (b), is selected among the selected point clouds
306 contained in the NL line. Note that the thick red line in Panel (a) is a group of selected points from
307 the point cloud data for defining the cross-section of the river, as shown in Panel (b).

308 All of the selected red points must be aligned according to the distance from the datum point
309 (here, N) with the new coordinate system. The new distance for the new x-coordinate can be
310 defined as *k*, as shown in Panel (c). This distance is estimated with the following equations.

311 The distances of *l*, *m*, and *n* with the TM coordinate can be estimated as follows. For example,
312 $N_{TM}(x)$ represents the x-coordinate of the TM projection for point *N*.

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021


Geoscientific Model Development Discussions

$$l = \sqrt{\left(N_{TM}(x) - M_{TM}(x)\right)^2 + \left(N_{TM}(y) - M_{TM}(y)\right)^2} \quad (10)$$

$$m = \sqrt{\left(N_{TM}(x) - L_{TM}(x)\right)^2 + \left(N_{TM}(y) - L_{TM}(y)\right)^2} \quad (11)$$

$$n = \sqrt{\left(L_{TM}(x) - M_{TM}(x)\right)^2 + \left(L_{TM}(y) - M_{TM}(y)\right)^2} \quad (12)$$

From the calculated angle of MNL (Θ) in Eq. (13), the new x-coordinate distance (k) can be calculated as in Eq. (14) with the law of cosines (i.e., $n^2 = l^2 + m^2 - 2lmcos\theta$) as the following:

$$cos\theta = \frac{l^2+m^2-n^2}{2lm} \quad (13)$$

$$k = l\,cos\theta \quad (14)$$

4.3. Results

4.3.1. Selected sites for cross-section

The two tested sites in the Migok-cheon stream are presented in Figure 9. The overall produced point cloud dataset for the UAV surveying area is presented in the left panel of Figure 9, and the picture of the left panel consists of only the collected points. Site-1 is located in the middle of the study area, while Site-2 is in the upper part of the area. Since the nearby area of Site-1 is located in the middle of the UAV surveying coverage, several images can be overleaped and captured for the same points.

Therefore, the number of points for demarcating a cross-section of the river might be sufficient to capture the detailed characteristics of the cross-section (see the top-right panel of Figure 9). In contrast, Site-2 is located at the upper part of the coverage area, and overlapping

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021

images might be limited, which indicates that the number of points to capture a target cross-section is also limited. Furthermore, a part of the cross-sectional area can be missing due to technical and environmental limitations such as waterbodies and insufficient overlapping images. For example, there are some areas in which no cloud point data exist, as on the right side of Site -1. This point is intentionally selected to verify the model performance in such a case.

4.3.2. Demarcation of the selected cross-sections

The demarcated cross-sections for the selected sites (i.e., Site-1 and Site 2) are presented in Figure 10 and Figure 11, respectively. In Figure 10, the extracted points for Site-1 are presented with red circles. As noted, a number of points are extracted from the UAV aerial photographs for Site-1 since the site is located in the middle of the coverage. The KLR fitted line shown with the dashed yellow line indicates that the fitted line reproduces the characteristics of the natural cross-section of the river well, including the overall trapezoidal shape and the natural bumps at the bottom. This line is compared with the field measurement reported in Donggwang Engr. (2004). Slight differences can be seen between the field measurement (shown with the solid blue line with the x marker) and the KLR fitted line since the field measurement took place approximately 17 years ago. However, the overall characteristics match well with each other, which indicates that the KLR fitted line fairly demarcates the cross-section of the river.

The cross-section of Site-2 is presented in Figure 11 and shows that the middle part of the river has no cloud point data. The KLR fitted line shows that the overall characteristics of the cross-section are reproduced. Even the missing part of the cross-section is also interpolated well with the KLR model by comparing the field measurements (see the solid blue line with the x marker in Figure 11). Some differences between the fitted line and the field measurement might result from the year difference. The result indicates that the KLR method can reproduce the

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021


Geoscientific Model Development Discussions

354 characteristics of the cross-section of a natural river. Furthermore, the missing part of the aerial
355 surveying can be filled up with the interpolation of the KLR method.

356 *4.3.3. Estimating cross-sectional area and wetted perimeter*

357 One of the main reasons to delineate the cross-section of a river is to model and estimate the flow
358 in the cross-section. The area and perimeter are essential to estimate for modeling the discharge of
359 the cross-section. With the KLR fitted line for the cross-sections of the river shown in Figure 10
360 and Figure 11 for Site-1 and Site-2, respectively, it is straightforward to estimate the area and the
361 perimeter. The area and perimeter can be estimated with the fixed height (H) as

362 $$A(H) = \sum_{t\in[(H-y_t)>0]} \frac{[(H-y_t)+(H-y_{t+1})]}{2} \Delta x \quad (15)$$

363 $$P(H) = \sum_{t\in[(H-y_t)>0]} \sqrt{(y_{t+1}-y_t)^2 + \Delta x^2} \quad (16)$$

364 where y_t is the fitted KLR line as in Eq. (6), and Δx is the interval of the x-coordinate in
365 estimating y_t.

366 The estimated $A(H)$ and $P(H)$ are presented in Figure 12 and Figure 13 for Site-1 and Site-
367 2, respectively. The area of Site-1 is exponentially increased, while the perimeter is increased at
368 different steps according to the heights, as shown in Figure 12. As seen in Figure 10, the width of
369 the cross-section increases as the height increases, and this feature affects the exponential increase
370 in the area as the height increases. At heights between 17.0 m and 18.5 m, the perimeter increases
371 rapidly as the height increases, as shown in the lower panel of Figure 12, since the shape of the
372 cross-section is rather flat in this range of heights, as shown in Figure 10. The other part of the
373 perimeter increases linearly except for a higher slope between 21 and 21.5 since the width becomes

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021

374 wider in this range of heights. Note that the area and perimeter outside the bank is excluded because
375 it is between 0 and approximately 10 m and over 37 m in the x-coordinate of Figure 11.

376 Similar features of the area and perimeter along with the height to Site-1 can be observed at
377 those of Site-2, as shown in Figure 13. This exponential and S-shaped increase in the area and
378 perimeter is a typical characteristic at the cross-section of natural rivers and trapezoidal channels.
379 The results of the area and perimeter show that the proposed KLR method can be a reasonable
380 alternative in demarcating the cross-section of a river obtained from a point cloud dataset.

381

382 5. Summary and Conclusions

383 The current study presents a nonparametric fitting method, the KLR, to the point cloud data
384 from UAV areal surveying to demarcate the cross-section of a river. Other than general fitting data,
385 the cross-section of a natural river generally contains sudden variation, an angled shape, and even
386 bumps as well as a linear shape even though the overall shape of a cross-section for a river is
387 trapezoidal. To accommodate all of those features of the natural cross-section, a highly flexible
388 fitting model is requested. Furthermore, the observed datae point is large enough for the point
389 cloud dataset. Therefore, the KLR model was chosen to fit the point cloud data for the cross-section.
390 The results conclude that the tested KLR model can reproduce the critical characteristics of the
391 cross-section of natural rivers with the point cloud data from UAV aerial surveying.

392 The major limitation of the point cloud data employed in the current study is that RGB
393 photographs were employed and the vegetation inside the river could generate an obscure cross-
394 sectional shape. Further optical instruments, such as hyperspectral and lidar sensors, could be

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021


Geoscientific Model Development Discussions

tested to overcome this limitation. However, a perfect solution that can remove the vegetation inside rivers has not yet been developed. To avoid this issue, points of the cross-section where little vegetation exists can be selected.

This KLR method can be easily adopted for other demarcation cases, such as buildings and structures. The proposed KLR method is a rather simple and direct approach for demarcating an area and structures. Additionally, other nonparametric techniques, such as LOWESS, can be further tested with extensive testing and adjustment. In the current study, the KLR model alone was focused on since the clustered data setting is obvious and easy to apply.

6. Code and data availability

All the employed code and excel files are available at Mendeley Data in <http://dx.doi.org/10.17632/xdw4cgnvhm.1>.

Competing interests

The author declares that they have no conflict of interest.

Author Contribution

T.L. carried out the research plan and programming as well as supervising while K.S. performed writing and collection data.

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021


Geoscientific Model Development Discussions

Acknowledgments

This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean Government (MEST).

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021

Geoscientific Model Development Discussions

References

Ahmad, M. I., Sinclair, C. D., and Werritty, A.: Log-logistic flood frequency analysis, Journal of Hydrology, 98, 205-224, 1988.

Anders, N., Smith, M., Suomalainen, J., Cammeraat, E., Valente, J., and Keesstra, S.: Impact of flight altitude and cover orientation on Digital Surface Model (DSM) accuracy for flood damage assessment in Murcia (Spain) using a fixed-wing UAV, Earth Science Informatics, 13, 391-404, 2020.

Andreadakis, E., Diakakis, M., Vassilakis, E., Deligiannakis, G., Antoniadis, A., Andriopoulos, P., Spyrou, N. I., and Nikolopoulos, E. I.: Unmanned aerial systems-aided post-flood peak discharge estimation in ephemeral streams, Remote Sensing, 12, 1-27, 2020.

Chow, V. T.: Open Channel Hydraulics. , McGraw-Hill, New York, 1959.

Cleveland, W. S.: Robust locally weighted regression and smoothing scatterplots, Journal of the American Statistical Association, 74, 829-836, 1979.

Donggwang Engr.: Reports of Fundamental River Plan for Migok-cheon Stream, Hapcheon-gun, South Korea2005-326, 484 pp., 2004.

Elek, P. and Márkus, L.: A long range dependent model with nonlinear innovations for simulating daily river flows, Natural Hazards and Earth System Science, 4, 277-283, 2004.

Gichamo, T. Z., Popescu, I., Jonoski, A., and Solomatine, D.: River cross-section extraction from the ASTER global DEM for flood modeling, Environmental Modelling and Software, 31, 37-46, 2012.

Gracchi, T., Rossi, G., Stefanelli, C. T., Tanteri, L., Pozzani, R., and Moretti, S.: Tracking the evolution of riverbed morphology on the basis of uav photogrammetry, Remote Sensing, 13, 1-16, 2021.

Hamdi, D. A., Iqbal, F., Alam, S., Kazim, A., and MacDermott, A.: Drone forensics: A case study on DJI phantom 4, 2019.

Hugenholtz, C. H., Whitehead, K., Brown, O. W., Barchyn, T. E., Moorman, B. J., LeClair, A., Riddell, K., and Hamilton, T.: Geomorphological mapping with a small unmanned aircraft system (sUAS): Feature detection and accuracy assessment of a photogrammetrically-derived digital terrain model, Geomorphology, 194, 16-24, 2013.

Izumida, A., Uchiyama, S., and Sugai, T.: Application of UAV-SfM photogrammetry and aerial lidar to a disastrous flood: Repeated topographic measurement of a newly formed crevasse splay of the Kinu River, central Japan, Natural Hazards and Earth System Sciences, 17, 1505-1519, 2017.

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021

451 Javernick, L., Brasington, J., and Caruso, B.: Modeling the topography of shallow braided rivers
452 using Structure-from-Motion photogrammetry, Geomorphology, 213, 166-182, 2014.

453 Kaewwilai, A. J.: Analysis of Flood Patterns in Adams County, Pennsylvania Utilizing Drone
454 Technology and Computer Simulations Analysis of Flood Patterns in Adams County,
455 Pennsylvania Utilizing Drone, 57, 1-20, 2019.

456 Lall, U. and Sharma, A.: A nearest neighbor bootstrap for resampling hydrologic time series, Water
457 Resources Research, 32, 679-693, 1996.

458 Langhammer, J.: UAV monitoring of stream restorations, Hydrology, 6, 2019.

459 Lee, G., Choi, M., Yu, W., and Jung, K.: Creation of river terrain data using region growing method
460 based on point cloud data from UAV photography, Quaternary International, 519, 255-262, 2019.

461 Lee, T. and Ouarda, T. B. M. J.: Identification of model order and number of neighbors for k-
462 nearest neighbor resampling, Journal of Hydrology, 404, 136-145, 2011.

463 Lee, T., Ouarda, T. B. M. J., and Yoon, S.: KNN-based local linear regression for the analysis and
464 simulation of low flow extremes under climatic influence, Climate Dynamics, 49, 3493-3511,
465 2017.

466 Lee, T., Salas, J. D., and Prairie, J.: An Enhanced Nonparametric Streamflow Disaggregation
467 Model with Genetic Algorithm, Water Resources Research, 46, W08545, 2010.

468 Lin, C., Chen, S. Y., Chen, C. C., and Tai, C. H.: Detecting newly grown tree leaves from
469 unmanned-aerial-vehicle images using hyperspectral target detection techniques, ISPRS Journal
470 of Photogrammetry and Remote Sensing, 142, 174-189, 2018.

471 Marfai, M. A., Sunarto, Khakim, N., Fatchurohman, H., Cahyadi, A., Wibowo, Y. A., and Rosaji,
472 F. S. C.: Tsunami hazard mapping and loss estimation using geographic information system in
473 Drini Beach, Gunungkidul Coastal Area, Yogyakarta, Indonesia, 2019.

474 Marteau, B., Vericat, D., Gibbins, C., Batalla, R. J., and Green, D. R.: Application of Structure-
475 from-Motion photogrammetry to river restoration, Earth Surface Processes and Landforms, 42,
476 503-515, 2017.

477 Orlowsky, B., Bothe, O., Fraedrich, K., Gerstengarbe, F. W., and Zhu, X.: Future climates from
478 bias-bootstrapped weather analogs: An application to the Yangtze River basin, Journal of Climate,
479 23, 3509-3524, 2010.

480 Perks, M. T., Russell, A. J., and Large, A. R. G.: Technical note: Advances in flash flood
481 monitoring using unmanned aerial vehicles (UAVs), Hydrology and Earth System Sciences, 20,
482 4005-4015, 2016.

483 Petikas, I., Keramaris, E., and Kanakoudis, V.: Calculation of multiple critical depths in open
484 channels using an adaptive cubic polynomials algorithm, Water (Switzerland), 12, 2020a.

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021


Geoscientific Model Development Discussions

485 Petikas, I., Keramaris, E., and Kanakoudis, V.: A novel method for the automatic extraction of
486 quality non-planar river cross-sections from digital elevation models, Water (Switzerland), 12,
487 2020b.

488 Pilotti, M.: Extraction of cross sections from digital elevation model for one-dimensional dam-
489 break wave propagation in mountain valleys, Water Resources Research, 52, 52-68, 2016.

490 Remondino, F., Barazzetti, L., Nex, F., Scaioni, M., and Sarazzi, D.: UAV photogrammetry for
491 mapping and 3d modeling - Current status and future perspectives, 2011, 25-31.

492 Sanhueza, D., Picco, L., Ruiz-Villanueva, V., Iroumé, A., Ulloa, H., and Barrientos, G.:
493 Quantification of fluvial wood using UAVs and structure from motion, Geomorphology, 345, 2019.

494 Seong, K., Lee, S. O., Jung, H. J., and Lee, T.: Safety first? Lessons from the Hapcheon Dam flood
495 in 2020, Natural Hazards, In Review, 2020.

496 Siebert, S. and Teizer, J.: Mobile 3D mapping for surveying earthwork projects using an
497 Unmanned Aerial Vehicle (UAV) system, Automation in Construction, 41, 1-14, 2014.

498 Simonoff, J. S.: Smoothing Methods in Statistics, Springer-Verlag, New York, 1996.

499 Smith, M. W., Carrivick, J. L., Hooke, J., and Kirkby, M. J.: Reconstructing flash flood magnitudes
500 using 'Structure-from-Motion': A rapid assessment tool, Journal of Hydrology, 519, 1914-1927,
501 2014.

502 Srivastava, K., Pandey, P. C., and Sharma, J. K.: An approach for route optimization in
503 applications of precision agriculture using uavs, Drones, 4, 1-24, 2020.

504 Taddia, Y., Pellegrinelli, A., Corbau, C., Franchi, G., Staver, L. W., Stevenson, J. C., and Nardin,
505 W.: High-resolution monitoring of tidal systems using UAV: A case study on poplar island, MD
506 (USA), Remote Sensing, 13, 2021.

507 Tomsett, C. and Leyland, J.: Remote sensing of river corridors: A review of current trends and
508 future directions, River Research and Applications, 35, 779-803, 2019.

509 Wang, S., Ahmed, Z., Hashmi, M. Z., and Pengyu, W.: Cliff face rock slope stability analysis
510 based on unmanned arial vehicle (UAV) photogrammetry, Geomechanics and Geophysics for
511 Geo-Energy and Geo-Resources, 5, 333-344, 2019.

512 Watanabe, Y. and Kawahara, Y.: UAV Photogrammetry for Monitoring Changes in River
513 Topography and Vegetation, Procedia Engineering, 154, 317-325, 2016.

514 Yan, Y., Ma, S., Yin, S., Hu, S., Long, Y., Xie, C., and Jiang, H.: Detection and Numerical
515 Simulation of Potential Hazard in Oil Pipeline Areas Based on UAV Surveys, Frontiers in Earth
516 Science, 9, 2021.

517 Zakaria, S., Mahadi, M. R., Abdullah, A. F., and Abdan, K.: Aerial platform reliability for flood
518 monitoring under various weather conditions: A review, 2018, 591-602.

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021

Geoscientific Model Development Discussions

519

520 Appendix: LOcally WEighted Scatterplot Smoothing (LOWESS)

521 LOWESS was proposed by Cleveland (1979) as a nonparametric regression. The LOWESS
522 with one explanatory variable and one predictor variable can be defined as

523 $$y_t = m(x_t) + \varepsilon_t \quad \text{(A.1)}$$

524 where the regression curve $m(\mathbf{y}_t)$ is the conditional expectation $m(x_t) = E(Y|X = x_t)$. The
525 LOWESS estimate can be defined as

526 $$\widehat{m}_{LOWESS}(x_t) = \vec{x}_t^T \widehat{\boldsymbol{\beta}}_t^{LOWESS} \quad \text{(A.2)}$$

527 where

528 $$\widehat{\boldsymbol{\beta}}_t^{LOWESS} = (\overleftrightarrow{\mathbf{X}}_t{}^T \boldsymbol{W}_t \overleftrightarrow{\mathbf{X}}_t)^{-1} \overleftrightarrow{\mathbf{X}}_t{}^T \boldsymbol{W}_t \mathbf{y} \quad \text{(A.3)}$$

529 with

530 $$\overleftrightarrow{\mathbf{X}}_t = \begin{pmatrix} 1 & x_t^1 - x_1^1 \\ 1 & x_t^1 - x_2^1 \\ 1 & \vdots \\ 1 & x_t^1 - x_n^1 \end{pmatrix} \quad \text{(A.4)}$$

531 and

532 $$\boldsymbol{W}_t = \boldsymbol{H}^{-1} diag[K_d(\boldsymbol{H}^{-1}(\boldsymbol{x}_t - \boldsymbol{x}_1)), \cdots, K(\boldsymbol{H}^{-1}(\boldsymbol{x}_t - \boldsymbol{x}_n))] \quad \text{(A.5)}$$

533 with the bandwidth matrix, **H**. The major characteristic of LOWESS is to employ the following
534 kernel function:

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021

Geoscientific Model Development Discussions

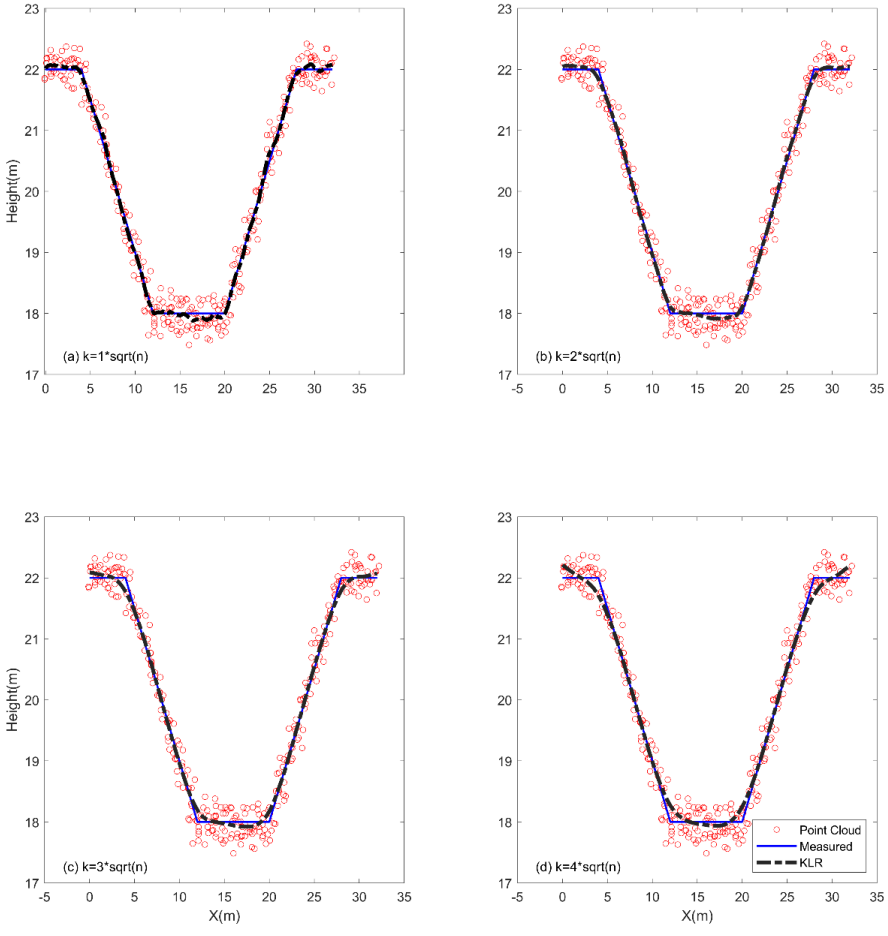
535 $$K_d(z) = \begin{cases} (1 - |z|^3)^3 & |z| < 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{(A.6)}$$

536

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021

Geoscientific Model Development Discussions

537 **Figure**



538

539 Figure 1. Assumed trapezoid channel to test the KLR model for point cloud data with different
540 portions of the number of neighbors ($k = a\sqrt{n}$, here a=1, 2, 3, and 4 at each panel). Note that (1)
541 the trapezoid sections are consistent with a 4 m top both sides and a 6 m base width as well as a
542 1:1 side slope with a 6 m height; (2) the number of points for the channel is divided at each 0.1
543 m, to total 161 points; (3) it is assumed that 10 times the divided data are collected, to total 322
544 points; and (4) the elevation of the bottom channel was assumed to be 18 m.

545

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021

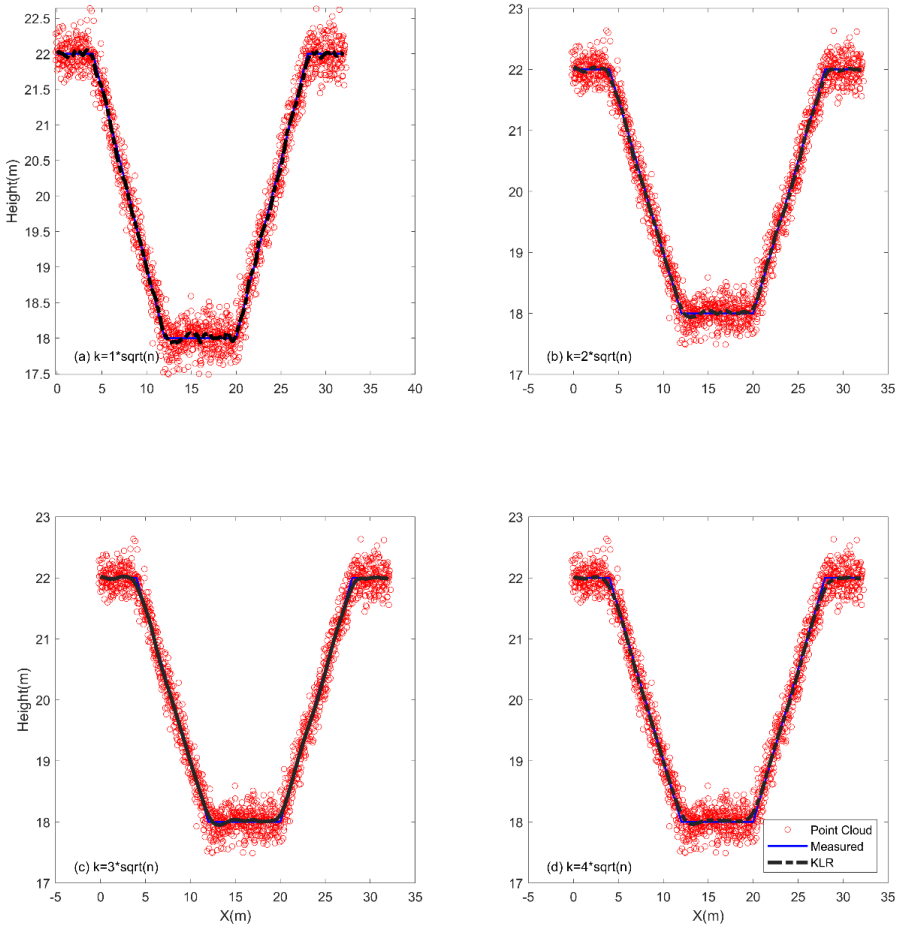


Figure 2. Assumed trapezoidal channel to test the KLR model for point cloud data with different multipliers of the number of neighbors ($k = a\sqrt{n}$). Note that (1) the trapezoidal sections are consistent with a 4 m top both sides and a 6 m base width as well as a 1:1 side slope with a 6 m height; (2) the number of points for the channel is divided at each 0.1 m, for a total of 161 points; and (3) it is assumed that 10 times the divided data are collected, for a total of 1610 points.

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021

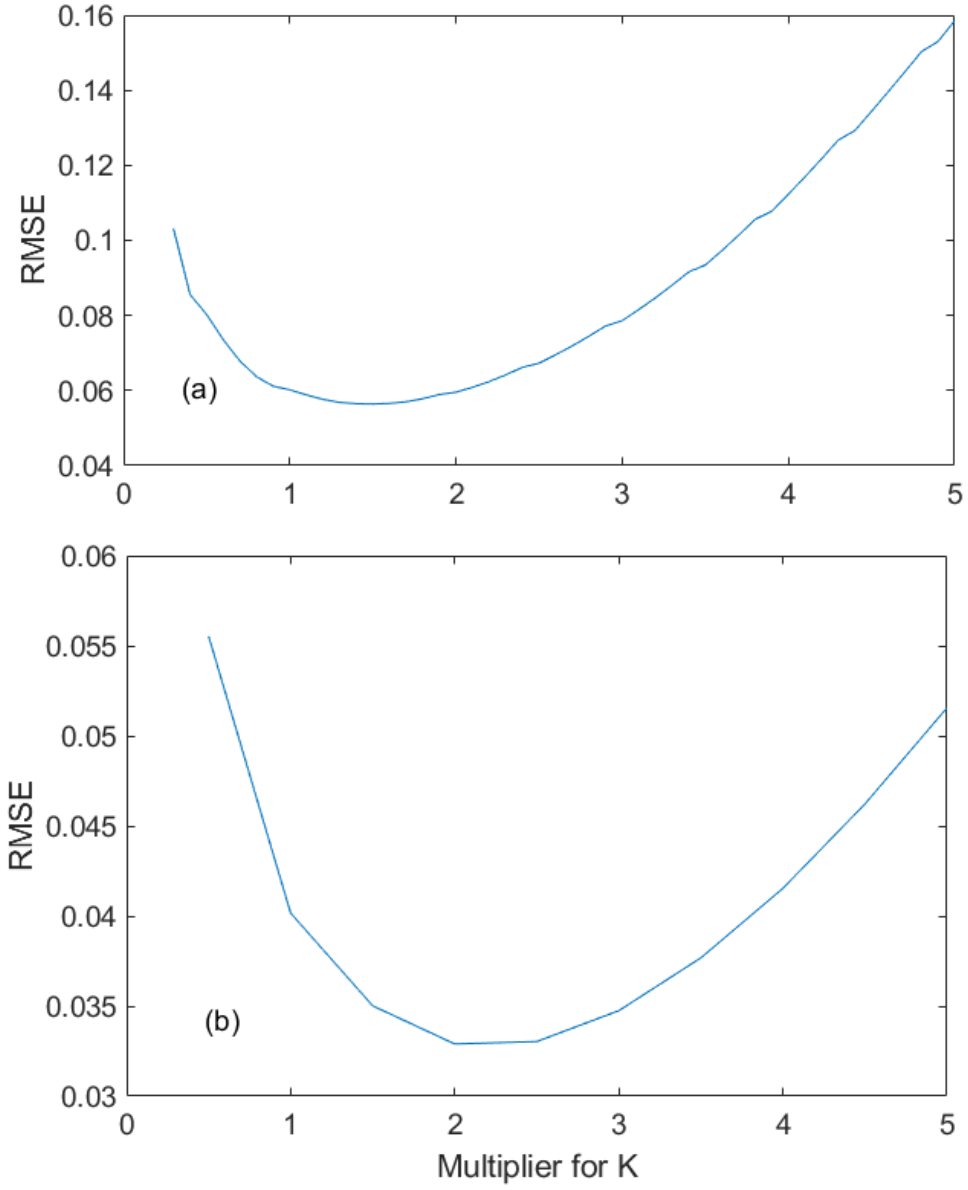
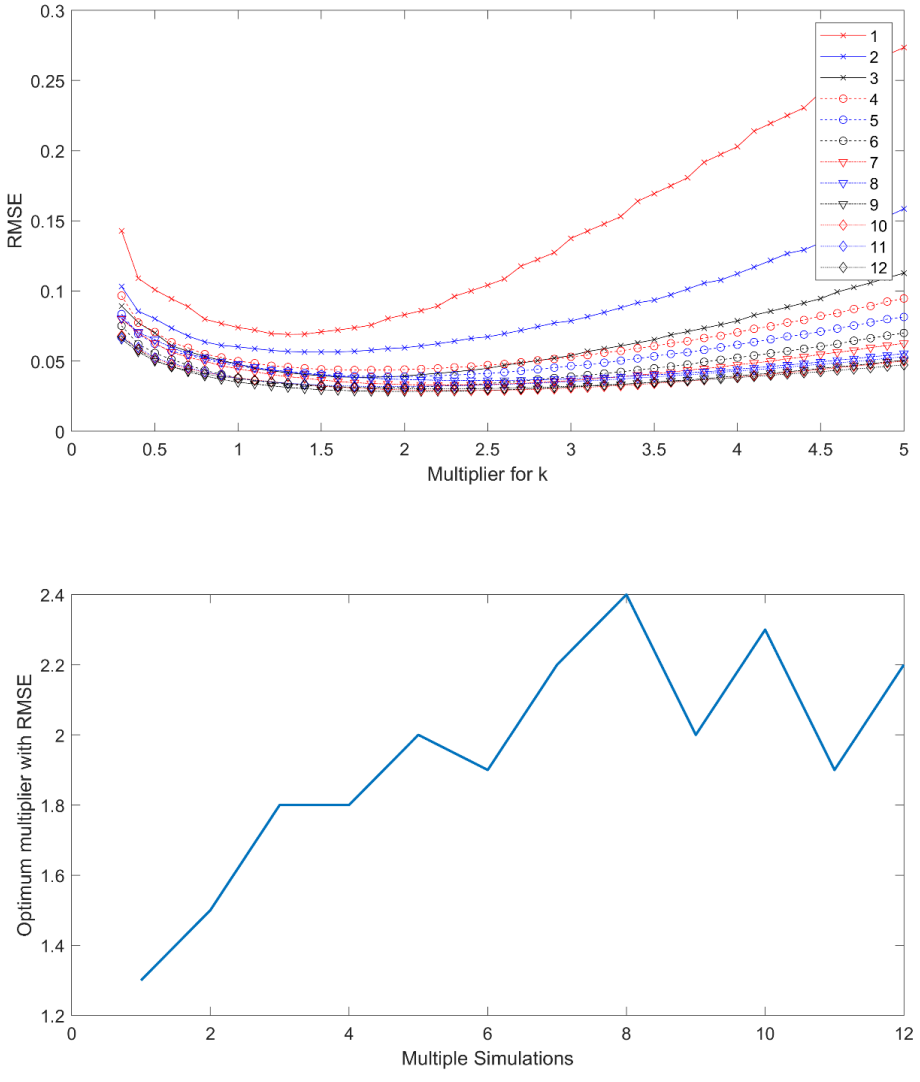


Figure 3. Root mean square error (RMSE) between the KLR estimate with different multipliers (a) of the number of neighbors ($k = a\sqrt{n}$) and the original trapezoid points for the case of 2 times the original points (Panel (a)) and 10 times (Panel (b)).

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021

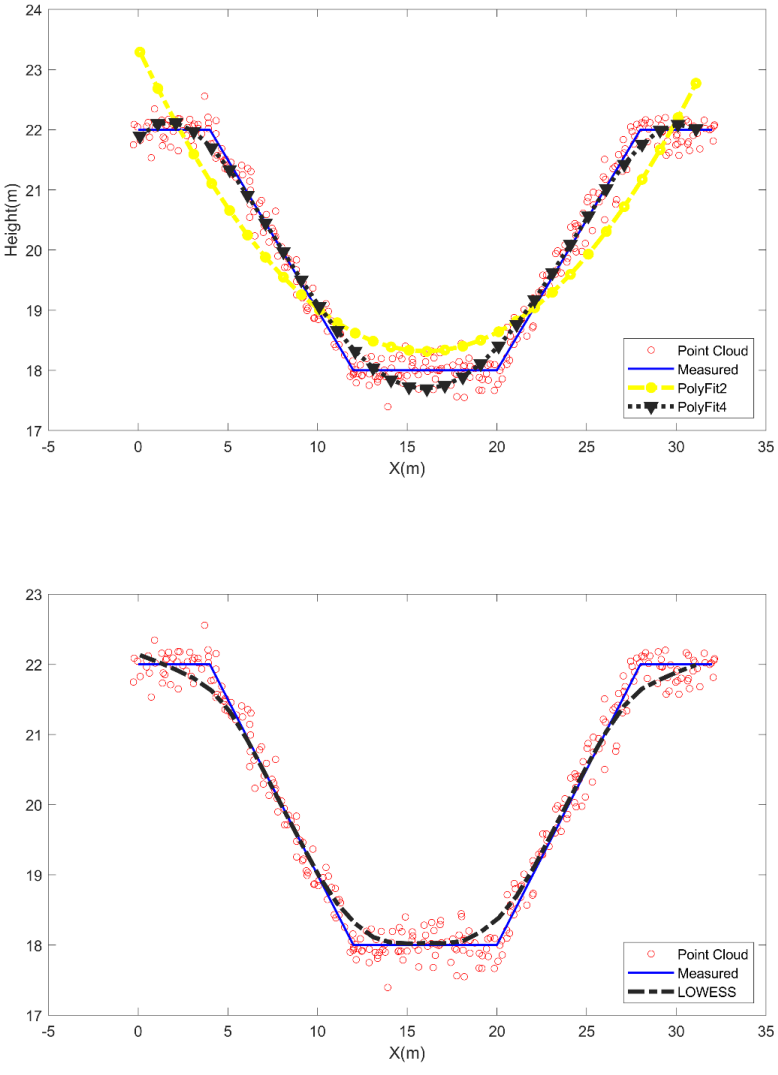
Geoscientific Model Development Discussions
Open Access
EGU



557

558 Figure 4. Root mean square error (RMSE) between the KLR estimate with different multipliers of
559 for number of neighbors ($k = a\sqrt{n}$) and the original trapezoid points for all of the cases between
560 1 and 12 times the original points (top panel) as well as the optimum multiplier with the RMSE
561 value at the top panel for each multiple simulation. Note that increasing the number of multiple
562 simulations indicates that the number of overlapped photos is increased and the cloud points are
563 multiplied.

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021



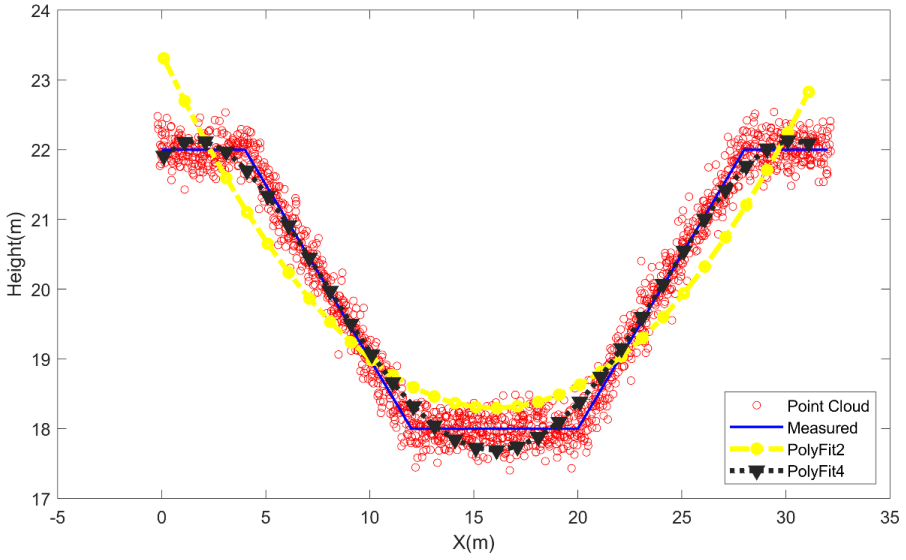
564

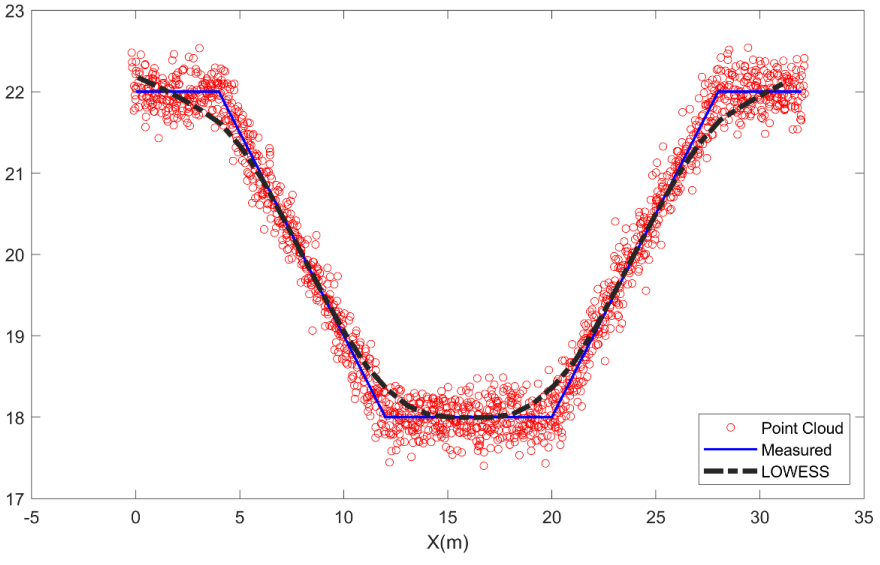
565 Figure 5. Polynomial regression (top panel) and LOWESS (bottom panel) were fitted to the
566 simulated point cloud data of 2 times the assumed cloud point (red circles) and assumed
567 trapezoidal channel (blue line).

568

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021

Geoscientific Model Development Discussions



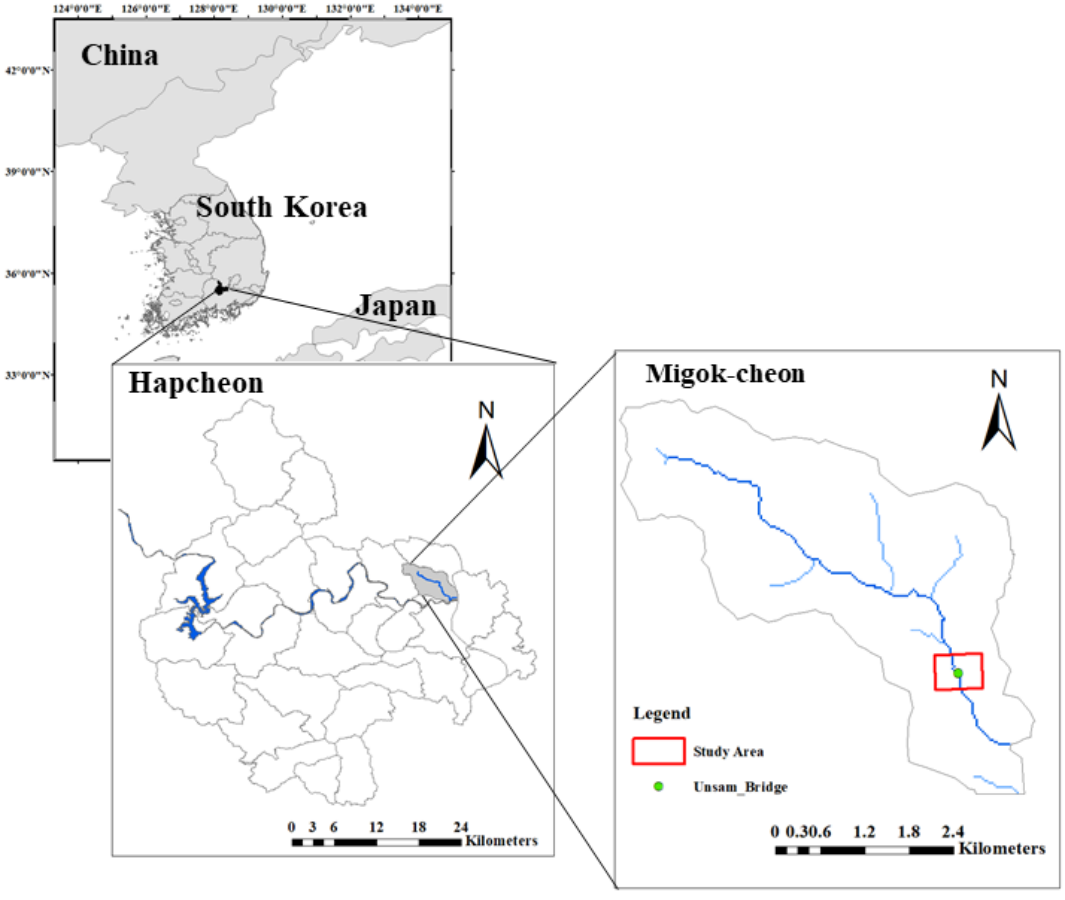


569

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021

570 Figure 6. Polynomial regression (top panel) and LOWESS (bottom panel) were fitted to the
571 simulated point cloud data of 10 times the assumed cloud point (red circles) and assumed
572 trapezoidal channel (blue line).

573
574



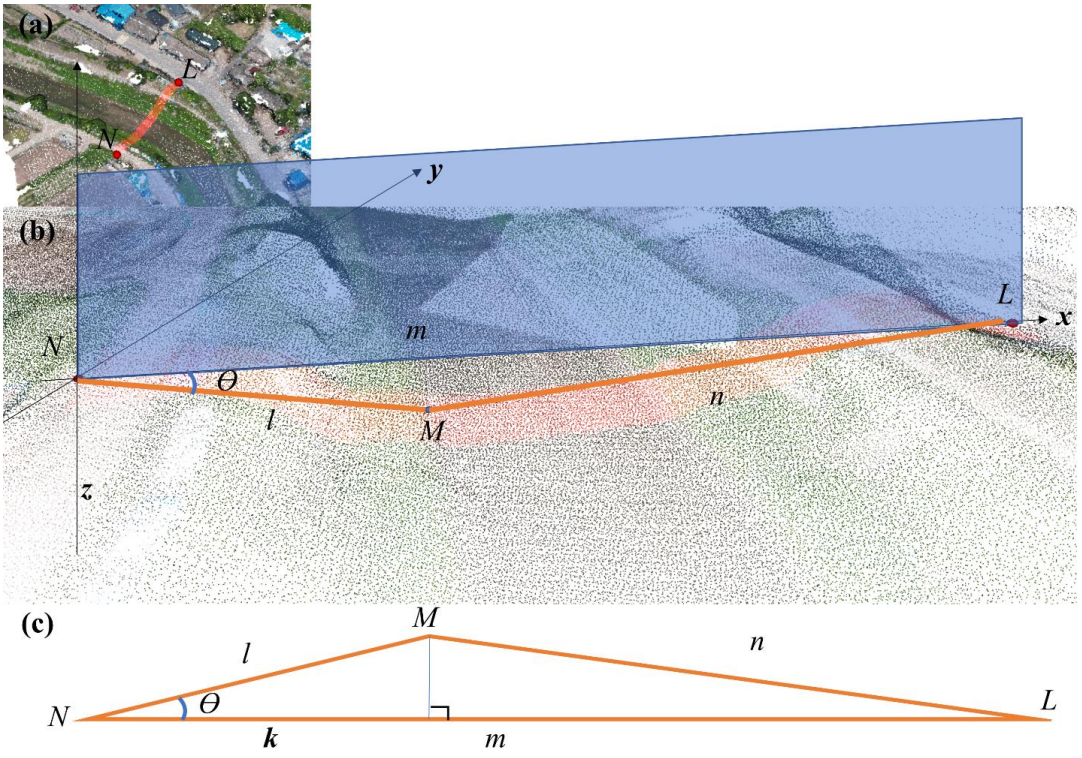
575

576 Figure 7. Study area of the applied stream, Migok-choen in South Korea, located in the province
577 of Hapchoen-gun.

578

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021

Geoscientific Model Development Discussions
Open Access
EGU



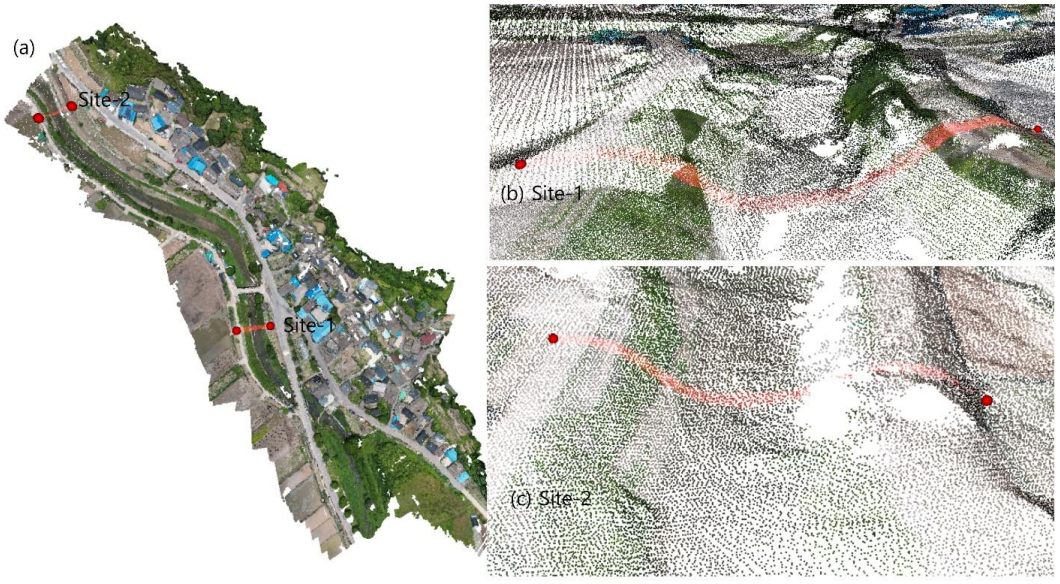
579

580 Figure 8. Example of the distance measurement: (a) aerial photo with a selected cross-section
581 (two red dots, L and N, and thick red line); (b) magnified photo of Panel (a) with assisted 3D
582 axis (x, y, and z) and the selected point (M); (c) emphasized triangle with the points of NML.
583 Note that (1) the cross-section can be defined with the x-axis by connecting points N and L with
584 the line; (2) the point M is the example point that contains the red line at Panel (a), which is a
585 group of points in reality; and (3) the actual distance of M from N in the x-axis is represented as
586 *k*, which can be designated as N to the point that meets line NL perpendicularly from M. The
587 aerial images were taken from the authors.

588

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021

Geoscientific Model Development Discussions
Open Access
EGU

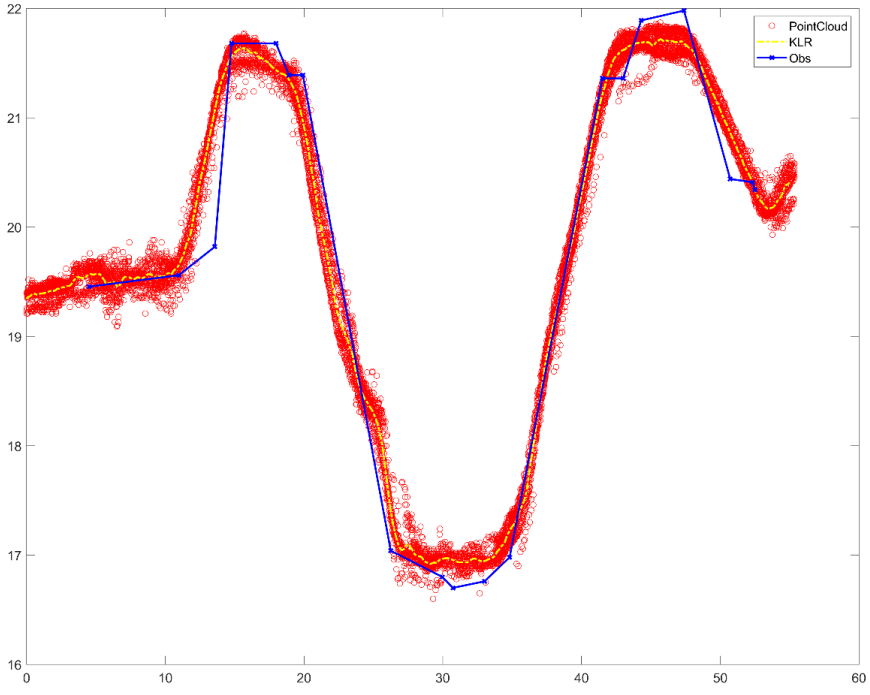


589
590 Figure 9. Two tested Sites in the Migok-cheon stream. Note that the right panels magnify the
591 tested sites by showing the point clouds of the observed data taken from the UAV photographs.
592 The aerial images were taken from the authors.
593

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021

Geoscientific Model Development Discussions

594

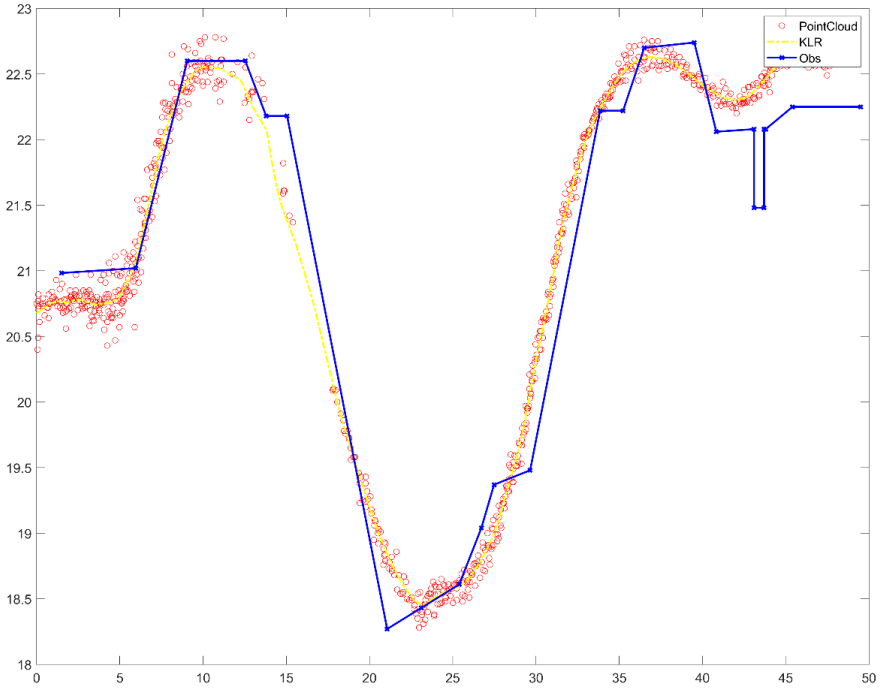


595
596 Figure 10. Point cloud data (red circles) and fitted KLR line (yellow dashed line) as well as the
597 observed surveying for Site-1. Note that the observed line was drawn from the previous
598 surveying in 2005.

599
600
601

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021

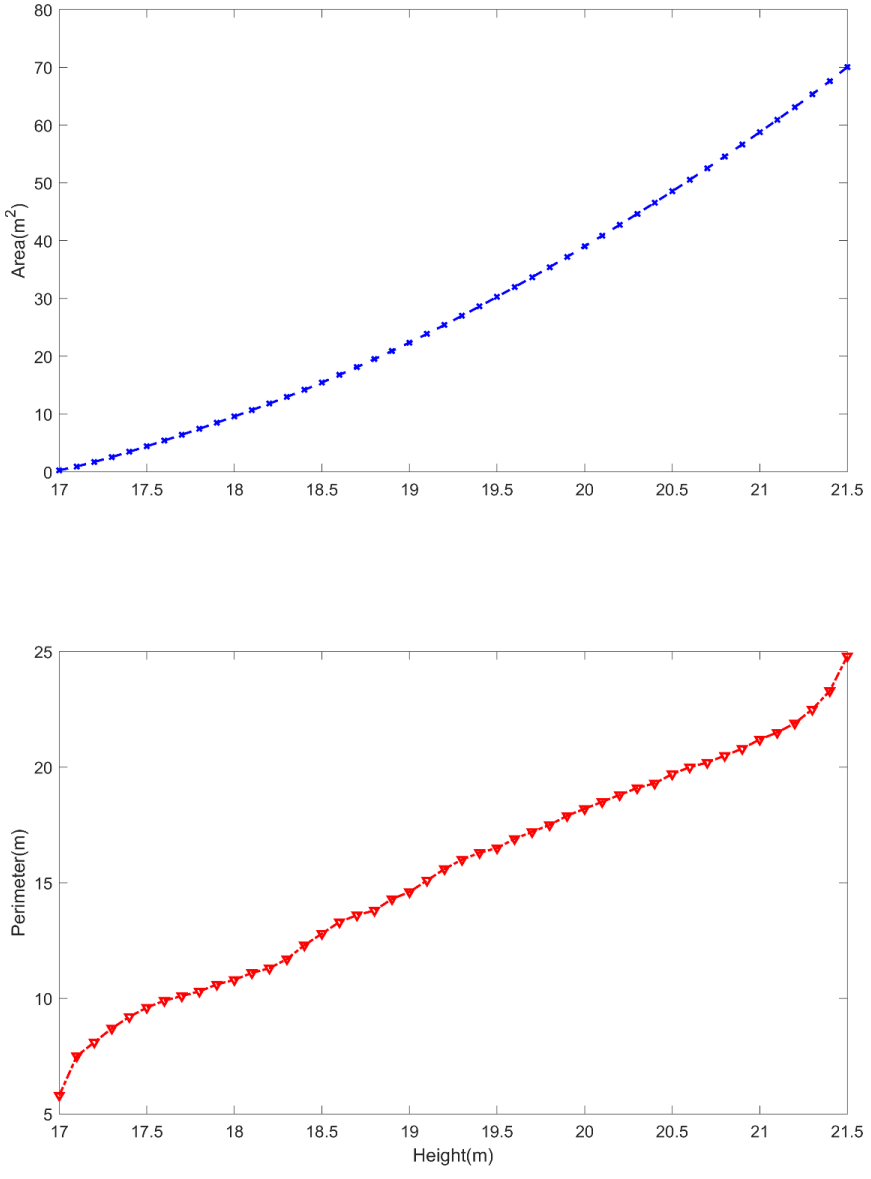
Geoscientific Model Development Discussions



602
603 Figure 11. Point cloud data (red circles) and fitted KLR line (yellow dashed line) as well as the
604 observed surveying for Site-2. Note that the observed line was drawn from the previous
605 surveying in 2005.

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021

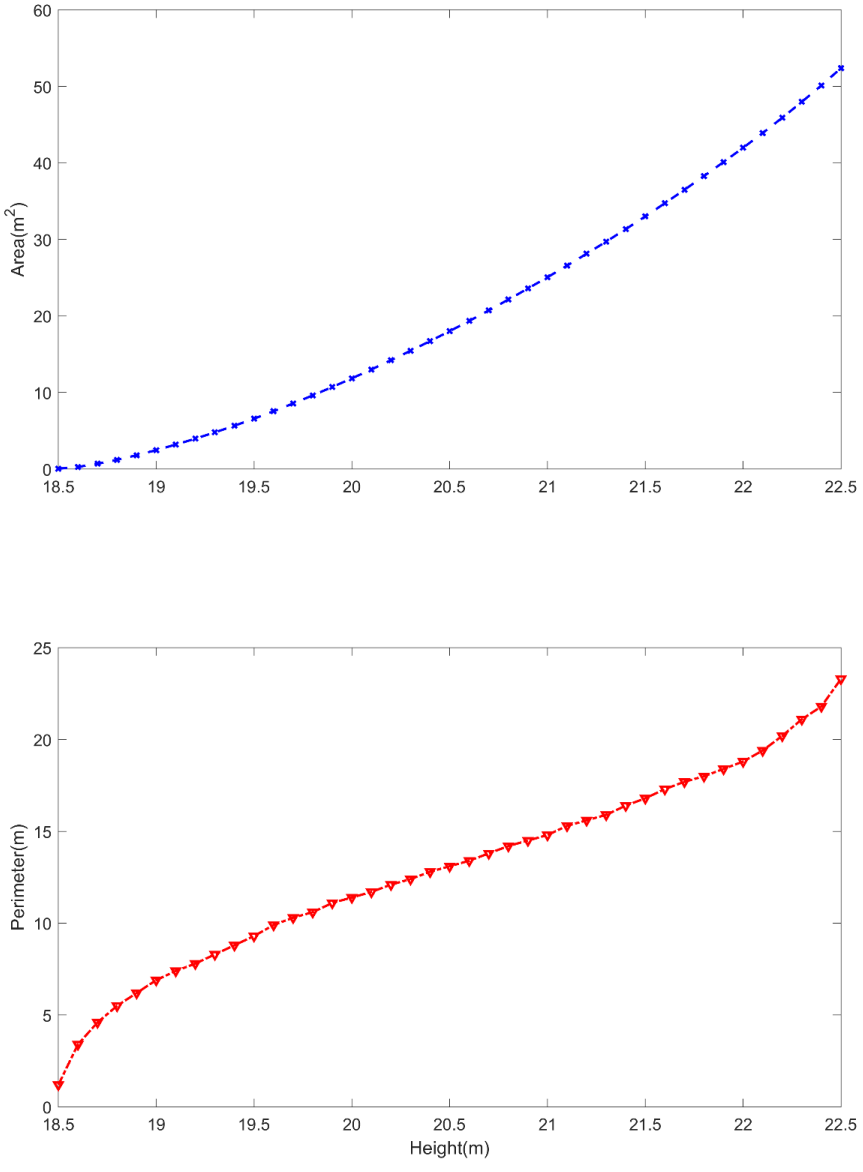
Geoscientific Model Development Discussions



606
607 Figure 12. Estimated area and perimeter with the fitted KLR line (see the dashed yellow line of
608 Figure 10) for Site-1 shown in Figure 7.
609
610

https://doi.org/10.5194/gmd-2021-309
Preprint. Discussion started: 17 November 2021

Geoscientific Model Development Discussions
Open Access
EGU



611
612 Figure 13. Estimated area and perimeter with the fitted KLR line (see the dashed yellow line of
613 Figure 10) for Site-2 shown in Figure 7.
614