University of South-Eastern Norway
—
Publication series No. 112
2023

Evaluating in-riverscapes:

Remote sensing green LiDAR efficiently provides accurate high-resolution bathymetric maps, but is limited by water penetration

Leif Kastdalen
Jan Heggenes

Leif Kastdalen and Jan Heggenes

Evaluating in-riverscapes:
Remote sensing green LiDAR efficiently provides accurate high-resolution bathymetric maps, but is limited by water penetration

University of South-Eastern Norway

Publication Series from University of South-Eastern Norway no. 112

ISSN: 2535-5325 (online)
ISBN: 978-82-7206-744-0 (online)

Foreword

Norwegian national authorities are running a research project to investigate the feasibility of using Airborne green LIDAR Bathymetry to map the topography of in-riverscapes. As part of this project, the University of South-Eastern Norway (USN) was subcontracted to analyze the accuracy and extent of LIDAR data as compared with the alternative methods of traditional stratified transect point sampling and Multi Beam bathymetry. The results are reported here. We are grateful for the close collaboration with other participants in the project, in particular Morten Stickler (Norwegian Water Resources and Energy Directorate; NVE) and Christian Malmquist (Norwegian Mapping Authority; Statens Kartverk).

Bø, 15 December 2022

Leif Kastdalen and Jan Heggenes

Abstract

This report investigates the feasibility of using Airborne green LiDAR Bathymetry to map the topography of in-riverscapes. First, the accuracy of remote sensing green LiDAR data are compared with the alternative *in situ* methods of traditional stratified transect point sampling and MultiBeam bathymetry. Secondly, the factors potentially limiting the feasibility of green LIDAR in rivers, i.e., factors explaining the loss of/lack of green LiDAR data, are explored. Thirdly, three types of green LiDAR sensors with different specifications are compared. Relatively long and variable within-river reaches (4 km or more) were studied to explore data robustness across different hydraulic patterns and water depths.

If signals were reflected, the different sensors generally provided remote sensing elevation data consistent with *in situ* elevation measurements, indicating high accuracy (±10 cm) across different hydraulic conditions. Loss of green LiDAR data was mainly a consequence of limited signal water penetration capability, i.e., the main environmental factor was water clarity. Secchi depth was a close proxy variable associated with green LiDAR penetration capability across rivers. Data loss was low until around Secchi water depth but increased rapidly thereafter. Surface water turbulence ('white water') and dark riverbed vegetation also increased green LiDAR signal loss. Point density and water penetration capabilities varied across sensors. Sensors with lower point density and thereby less spatial resolution, also had more signal strength and therefore penetrated deeper water. We conclude that green LiDAR bathymetry overall is a robust method that efficiently provides accurate elevation data across different hydraulic conditions and water depths. The main concern is signal loss, which is associated with Secchi water depth, and aggravated by surface turbulence ('white water') and dark coloured riverbed. Technically, this loss also depends on green LiDAR signal power, which vary across sensors. Therefore, signal power should be balanced with signal density (spatial resolution), depending on project objectives.

Contents

1. Introduction

Remote sensing by airborne red LiDAR is the de facto labour and cost-effective standard methodology for obtaining high-resolution 3D-Digital Elevation Model (DEM) information in terrestrial landscapes (Mandlburger 2018; Kinzel, Legleiter & Grams 2021; Islam et al. 2022). The red laser provides high accuracy, although interpreting and categorizing which object reflects the laser signal may be challenging. This is implemented using post-processing algorithm software which may influence accuracy and utility of the data (e.g., Kinzel, Legleiter & Nelson 2013; Islam et al. 2022). Unfortunately, red laser signals do not penetrate water, and are even with additional DEM modelling and information less suitable for remote sensing bathymetric measurements (Smart, Bind & Duncan 2009), especially in shallow water (< 2 m) (Pe'Eri & Philpot 2007; Pe'eri et al. 2011). Therefore, traditional *in situ* manual, time-consuming, and cost expensive methods as dGPS transect point measurements and MultiBeam Echosounder Mapping System (MBES) data collection, are still used for measuring topographic features in water covered areas, water depths and other habitat-relevant characteristics in rivers and lakes (e.g., Payne, Eggers & Parkinson 2004; Gard 2005; Jowett & Duncan 2012), with dGPS point data inevitably providing (very) few data points per area unit.

Recent developments in remote sensing monitoring and LiDAR technology are, however, highly promising. Green LiDAR have, although to variable extents, demonstrated water-penetrating capabilities and been applied in various types of remote sensing projects focusing on bathymetric measurements. Previous applications have been with relatively low resolution (one point per 1-2 m^2) and on relatively short and uniform low-gradient river reaches (Hilldale & Raff 2008; Kinzel, Legleiter & Nelson 2013). Recently developed sensors promise increased spatial resolution (point density) and water depth penetration, although resolution and likely also depth-penetrating capabilities may vary considerably among sensors (below) (Awadallah 2021). To our knowledge, the accuracy and applicability of such remote sensing green LiDAR bathymetric measurements providing high point density (Mandlburger et al. 2020), have not been systematically evaluated and tested on larger spatial scales, e.g., across a variety of hydraulic conditions on meso-habitats (< 100 m) or river scales. Neither has green LiDAR been compared with alternative *in situ* bathymetric methods in smaller rivers with more complex and longitudinally variable hydraulics, i.e., more high-gradient and relatively shallow habitats, and for longer (> 2 km) and presumably more representative river reaches. (but see Kinzel, Legleiter & Nelson 2013 for shorter river reaches; Kinzel, Legleiter & Grams 2021 for six close cross-sections in a river and tributary ; Islam et al. 2022 for a uniform, wide, low gradient river reach).

The scientific and management communities consider remote sensing Airborne green LiDAR Bathymetry (ALB) as a promising cost-effective tool for obtaining accurate data for in-riverscapes as well as in less challenging still water. Therefore, Norwegian federal agencies (The Norwegian Water and Energy Directorate, The Norwegian Mapping Authority and Norwegian Environment Agency) have established a national cross-sectoral research project with the main objective to i) evaluate the robustness of ALB technology on Norwegian rivers and shallow part of lakes with high variation of

bed slope and non-fluvial, semi-fluvial and fluvial reaches; ii) demonstrate selected application areas with respect to freshwater resource management, iii) facilitate involvement and collaboration across Norwegian management for future acquisition of green LiDAR; and iv) evaluate whether green LiDAR is a future cost-benefit tool that should be invested in in Norway.

As a collaboration partner the University of South-Eastern Norway (USN) was invited to contribute to the national project "Validation and application of Airborne Laser Bathymetry (ALB) technology for improved management and monitoring of Norwegian rivers and lakes (2021-2022)", structured into three main work packages; WP 1: Validation of ALB; WP 2: Application in River systems; and WP 3: Data acquisition and quality assurance, in which USN had the responsibility of WP 1 (Validation of ALB). The project acquired three different European green LiDAR providers/sensors that were applied on five different river systems and two different inland lakes, and then compared against "ground truth" based on dGPS and Multibeam data in pre-selected areas representing a range of water bodies.

The present report documents the study design, analyses and findings based on systematic quantitative statistical evaluations of ALB data against ground truth data (dGPS and Multibeam echosounder) in three rivers. Specific objectives have been to explore and evaluate:

1. The accuracy and robustness of airborne remote sensing green LiDAR data across a) rivers and b) sensors compared to the alternative *in situ* methods, and thereby
2. Identify factors that may a) bias the accuracy (quality) and b) lead to signal loss (quantity) of green LiDAR data.

To evaluate potential bias, we used a stratified design across pre-selected hydraulic variables, primarily water depth, water surface roughness/water column air-bubbles (turbulence, 'white water') associated with river gradient, bottom substrate (size, color), and in particular water clarity/underwater visibility (Hilldale & Raff 2008; Kinzel, Legleiter & Nelson 2013; Islam et al. 2022).

2. Methods and materials

2.1. Study areas

Three rivers located in south Norway (Figure 1) were selected for study to represent a range of waterflows, gradients, river widths and depths, and water clarity expressed as Secchi depth (Table 1) (for further river details see https://vann-nett.no/portal/).

2.2. Study design

Sampling was stratified by rivers and green LiDAR sensors (as provided by commercial providers), and all data collected in late summer 2021. To be able to compare performance of green LiDAR across different types of river morphologies and habitats, one rather long within-river reach (4.8 – 14.8 km)

with varying characteristics (waterflow, gradient, i.e., water surface smoothness and bottom substrates, water depths, underwater visibility) were selected for study in each river (Figure 1, Table 1).

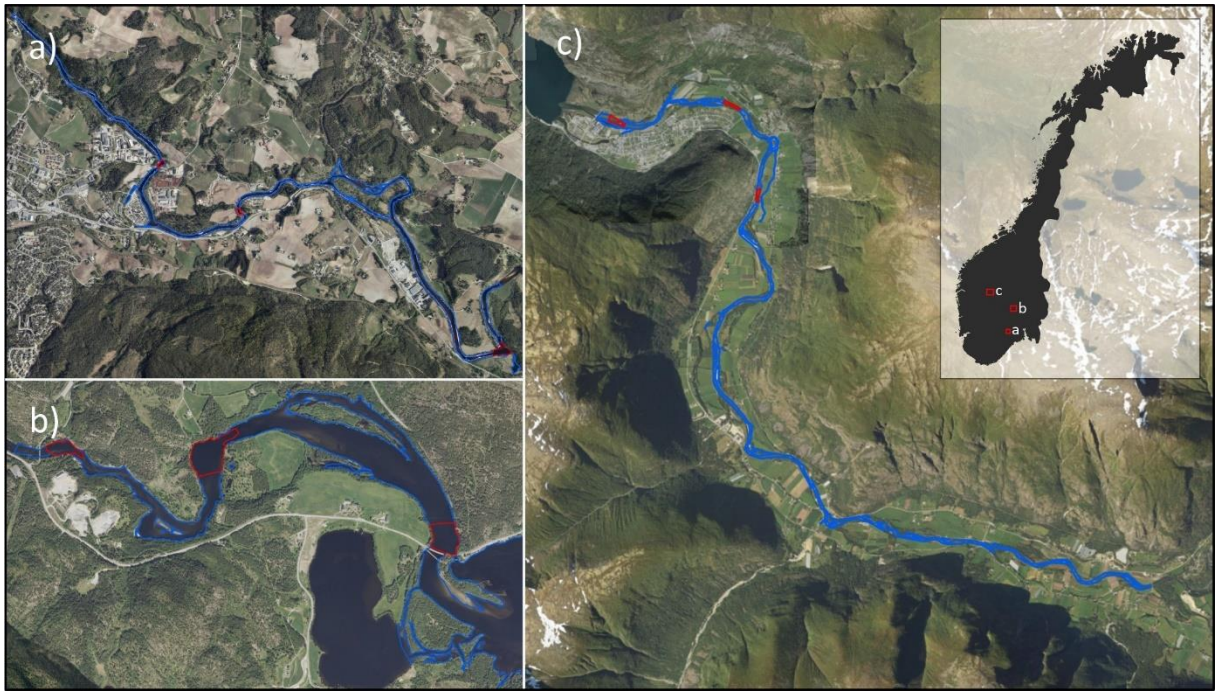


Figure 1. The reaches studied in the three rivers a) Bø (upper left; 7.7 km), b) Hallingdal (lower left; 4.8 km), and c) Lærdal (right; 14.8 km) with the three largest pools delineated (red lines). Inset: location in Norway.

Table 1. Characteristics of the sampled rivers and study reaches.

River	Mean flow (m^3/s)	Reach length (km)	Mean depth (m)	Total area (ha)	Gradient (m/km) (total, in m)	Mean bed particle size (class)	Secchi depth (m)
Bø	18.2	7.7	0.65	29.7	4.0 (31)	8.8	3
Hallingdal	108	4.8	2.15	83.0	0.3 (1.2)	4.5	5
Lærdal	36.3	14.8	0.74	58.7	4.7 (9)	8.5	>9

To compare accuracy of different green LiDAR sensors across different types of rivers, three commercial providers offering ALB services, and all using different sensors providing different power, point densities, and footprint (Figure 2, Table 2), participated in the study. The three providers fly with fixed-wing airplanes and use the sensors Riegl VQ-880, Leica Chiroptera 4x and Teledyne Optech CZMIL.

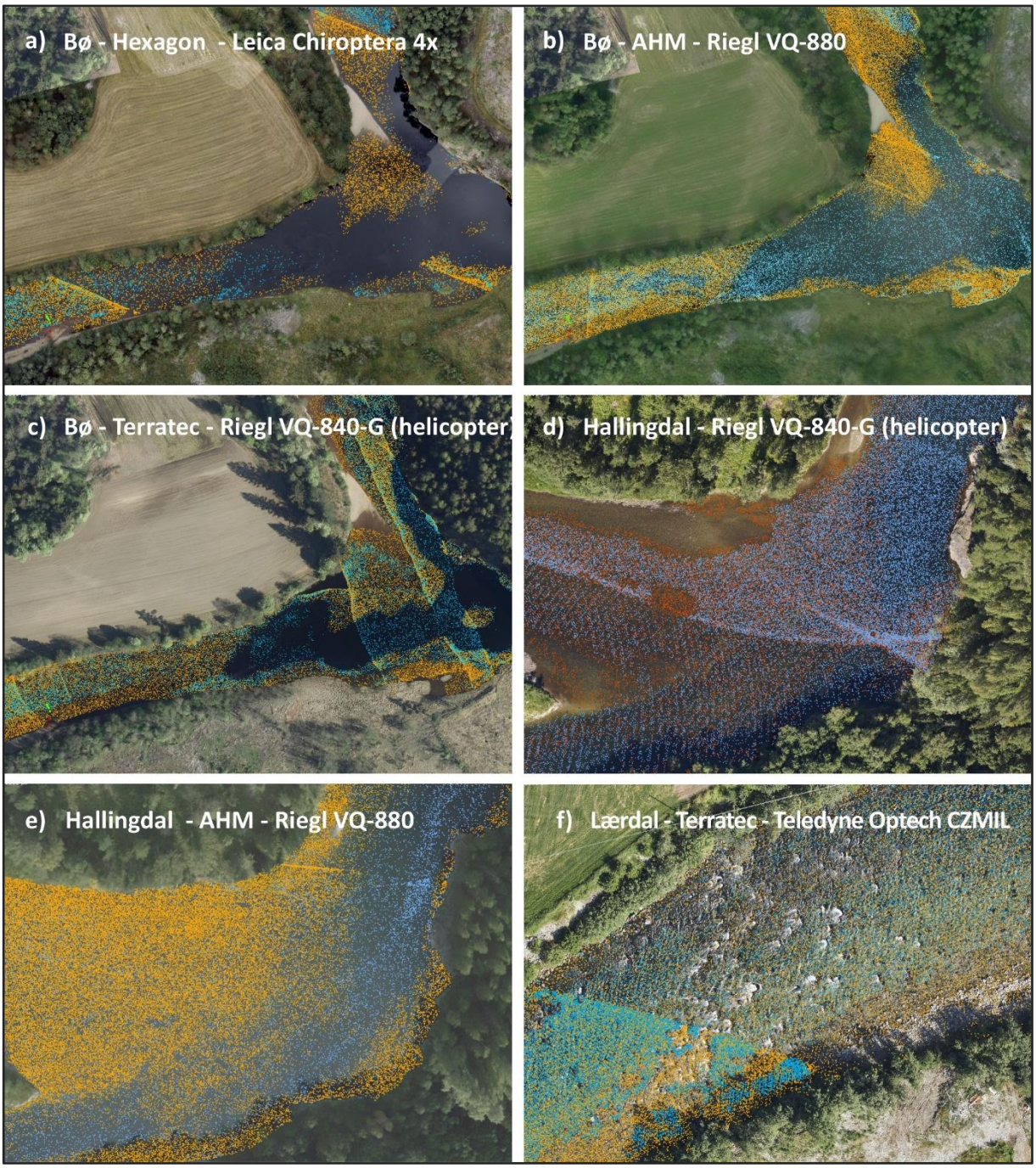


Figure 2. Examples of point density classified as water surface (blue) and river bottom (orange) across rivers and green LiDAR sensor providers (Table 2).

For unfortunate logistic reasons, all sensors could not be used for all river reaches. AHM collected data for all three rivers, Terratec flew the Hallingdal and Lærdal rivers, whereas Hexagon flew the Bø river (Table 3). Terratec flew also with a Riegl VQ-840-G sensor on a helicopter along selected parts of the study reaches in all three rivers (Bø: 4.2 km; Hallingdal: 2.6 km; Lærdal: 1.3 km).

Table 2. Characteristics of the green LiDAR sensors evaluated in the project.

Provider	Sensor	Puls energy (mJ)	Pulse length (kHz)	Laser footprint (cm)	Sampling altitude (m)	Sampling strategi	Point density* (m^2)
AHM	Riegl VQ-880-NG		-550	38 cm	400	Adaptive	67-91/939
Hexagon	Leica Chiroptera 4X	0.1	140/500	2.0 m	450	Systematic coverage	17/17
Terratec – fixed-wing	Teledyne Optech CZMIL Supernova	1.75	90-270	2.9 m	380-400	Systematic coverage	6/24
Terratec - helicopter	Riegl VQ-840-G		50-200	15-21 cm	90-140	Selected coverage	40-85/133

** Stated by the manufacturer/measured from the received data. The increase in point density is due to overlapping flight lines.*

The logistic impossibility of implementing all field surveys at the same time, did imply some variation in waterflow (Table 3, Figure 3), and therefore different elevations in measured water surface level. This required additional data post-processing and likely contributed some additional confounding noise in the data analyses (below).

For comparison, *in situ* data were collected using wading and Multibeam techniques for mapping fluvial environments. Because they are surface-operated, Multibeam echosounder mapping systems may be difficult to operate in shallow-fast water. On a cautionary note, Multibeam data may also become more 'noisy' in shallower water and for more sloping riverbeds because of the increasing beam angle (e.g., Hughes Clarke, Mayer & Wells 1996; Calder & Mayer 2003). Therefore, Multibeam was primarily used for slow/deep pool areas in our study, with a focus on investigating water-penetrating capabilities. Consequently, the three deepest pools in each river/study reach were covered by Multibeam (Figure 1; using a WIinghead i77h with tightly coupled GNSS/INS integration).

Table 3. Water flow (m^3/s) at date of data collection in 2021, by river and data provider.*

	Bø		Hallingdal		Lærdal	
LiDAR data						
AHM	21	3. Sep.	75	3.Sep.	15	26. Sep.
Hexagon	15	27. Aug.				
TerraTec fixed-wing			78	16. Jul.	20	21. Jul.
TerraTec helicopter	13	25. Sep.	51	24. Aug.	15	25. Sep.
Control data						
Multibeam (pools)	12	26. Oct.	90	10.Nov.	21	16. Nov.
Wading points (transects)	19	22.Nov.	170	3. Nov.	21	16. Nov.

**Measuring station in Bø: Hagadrag, in Hallingdal: Bergheim, and in Lærdal: Stuvane (https://sildre.nve.no/)*

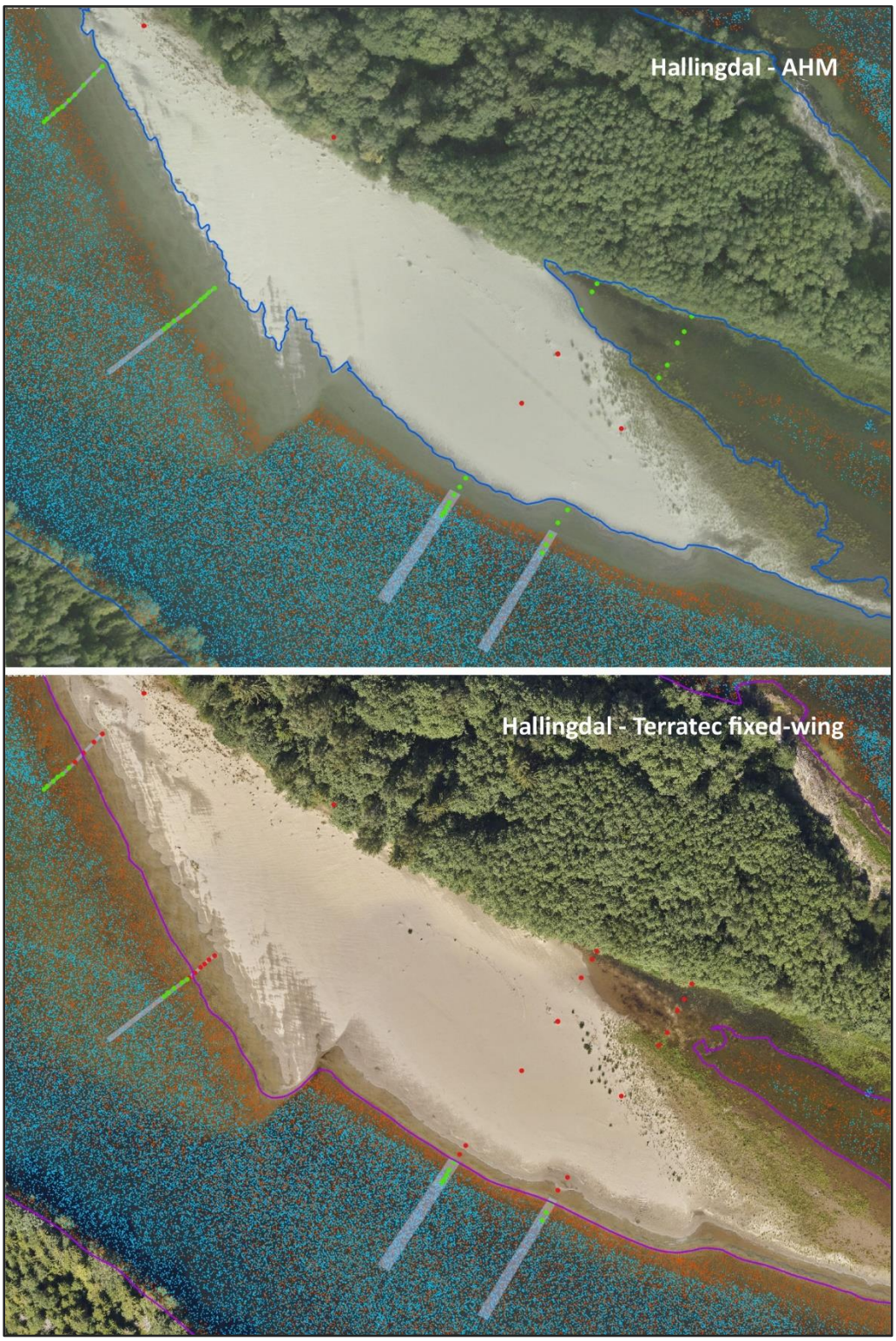


Figure 3. Due to differences in time between control (transect points; in green), and green LiDAR acquisitions in Hallingdal, the water surface changed up to 1 meter. Therefore, some transect points measured in the field were above water at the time of the LiDAR recordings (transect points in red). Rectangles indicate the areas manually extended to provide a minimum of 10 points for estimating water elevation. The same area was used for all sensors. Blue (upper panel) and red (lower panel) lines indicate riverbanks as identified by the commercial providers based on LiDAR data. LiDAR points classified to water surface in blue, and to river bottom in orange.

Traditional stratified transect point sampling by wading was used for measuring water depth, bed substrates, embeddedness, and vegetation cover/type, i.e., relevant characteristics in the studied river reaches (e.g., Payne, Eggers & Parkinson 2004; Gard 2005; Jowett & Duncan 2012). For each of the study reaches, the entire reach was stratified into one of five river types according to Hauer and Pulg (2018). Considering fluvial and late glacial processes, the types were cascade, plane-bed-dimictic, plane-bed-fluvial, riffle-pool-mixed og riffle-pool-fluvial. Within each river type, ten transects were visually stratified to represent varying river gradients, as expressed by surface water roughness, using high-resolution (10 cm) aerial photos (https://www.norgeibilder.no/). Five points (at the 10, 25, 50, 75, 90 percentiles) were measured across each transect, either across the river or to as deep as it was possible to wade (155 cm).

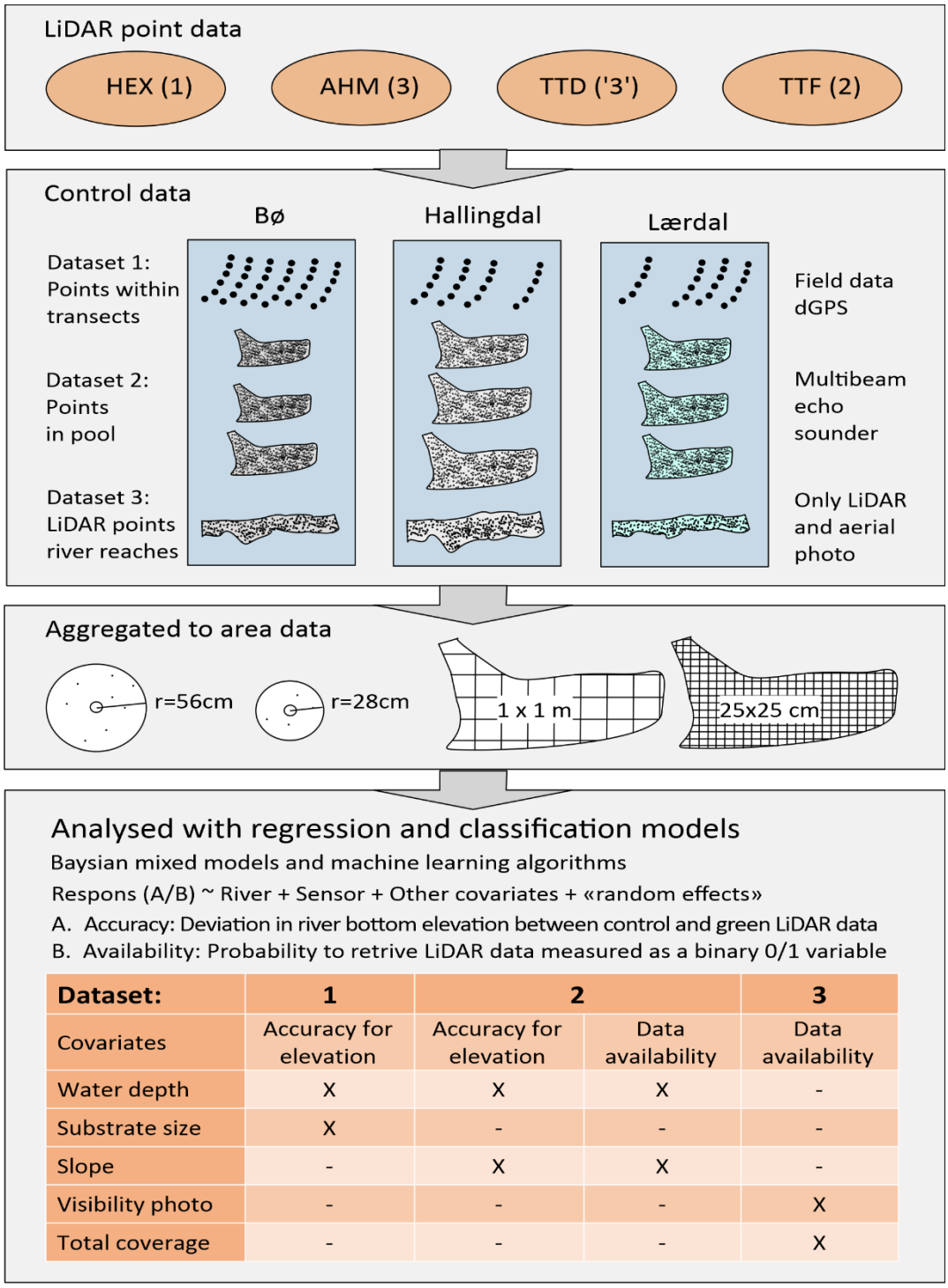
All measurement points were accurately positioned using a dGPS (Trimble R6). In each point/associated area (0.5 x 0.5 m = 0.25 m^2), water bottom height was measured, and dominant and subdominant substrate particle size classified according to a modified Wentworth scale (Heggenes 2002), embeddedness scored 1-4 (0-25%, 25-50%, 50-75%. 75-100 %), and vegetation type and cover (Braun-Blanchet scale) was noted. Water surface height was measured in one point, regularly in the first point closest to the riverbank.

2.3. Data processing

The LiDAR data were pre-processed by the commercial providers using generic algorithms and software. Points were classified using the LAS 1.4 format, i.e., especially river bottom (class 26) and water surface (class 27). AHM and Terratec also delivered a shape file for the riverbanks/wetted area (Figure 3). In addition to the LiDAR data, AHM, Hexagon and Terratec also took concomitant aerial photographs of the green LiDAR flown river reaches, using high-resolution cameras (Hasselblad H/39, Leica RCD30, Phase one iXM-RS-150F) producing RGB orthophotos. Since Hexagon (in Bø) and Terratec (helicopter) did not deliver shape files of the riverbanks, the map detection algorithm 't_concave_hull' in the R library 'lidR' was used to post-process the data to delimit the riverbank/wetted area from the points classified as 26 (bottom) and 27 (surface).

Three sub datasets were derived for further analyses, all including the processed and relevant LiDAR data (Figure 4). The first dataset also included *in situ* control data based on the wading field measurements/classifications in the dGPS located transect points/associated areas, extending from the riverbank (Dataset 1). The second dataset included the nine deeper pool areas (three pools per river), where also the *in situ* Multibeam water depth data had been collected (Dataset 2). The third dataset included green LiDAR data for the entire mapped river reaches. For dataset 1, green LiDAR data were aggregated within a buffer circle of 28 cm radius around each transect point, corresponding to the *in situ* transect point/associated area representing 0.5 x 0.5 m = 0,25m^2, and to a larger area with 56 cm radius corresponding to a spatial grid size of 1x1m, also used in the other datasets. For dataset 2 and 3 the data were aggregated and scaled into two alternative spatial grid sizes, 25x25 and 100x100 cm. Thus, potential small-spatial effects could be explored. Within each

spatial unit, we calculated number of LiDAR points together with mean, median, minimum, and standard deviation for the elevation values. The datasets were primarily analyzed using regression/classification models with measurement accuracy and LiDAR data availability as response variables (Figure 4).



Dataset:	**1**	**2**		**3**
Covariates	Accuracy for elevation	Accuracy for elevation	Data availability	Data availability
Water depth	X	X	X	-
Substrate size	X	-	-	-
Slope	-	X	X	-
Visibility photo	-	-	-	X
Total coverage	-	-	-	X

Figure 4. Overview of the data and analyses. Four different LiDAR sensors from three data providers were used in three rivers along continuous, relatively long reaches. Control data were measured in situ in 1) wading stratified transect points in the field and 2) and Multibeam echosounder in the three deepest pools of each river. The green LiDAR data were, for the transect points, aggregated within buffer circles of radius 28 and 56 cm, and for the pool data and whole reaches point data were aggregated into two alternative grid sizes (25x25 and 100x100 cm), and analysed with regression and classification models with measurement accuracy and data availability as the primary response variables. (Hex = Hexagon with Leica Chiroptera 4x, AHM with Riegl VQ-880, TTD = Terratec with Riegl 840 on a helicopter; TTF = Terratec with Teledyne Optech CZMIL).

2.4. Transect data

Time-consuming wading transect point measurements are by necessity relatively few. Moreover, it turned out that green LiDAR point coverage was uneven for some sensors and river areas, reducing the number of points available for analyses even more (Appendix 1).
Differences in measured water surface elevation were generally small across sensors and rivers, indicating similar water flows between sampling times. The exception was Hallingdal, where differences in water flows between sampling times (Table 3, Figure 3), changed measured water surface elevation, especially between lower flow green LiDAR data from Terratec in July and *in situ* wading transect points during higher flow in the fall, with a mean difference close to 1 m (Figure 5). This also reduced the number of transect points available for comparative analyses, since some transect points were 'above water' compared to the green LiDAR data at lower flows (78 of 685 observations in total).

Within the buffer areas used to analyse the transect points, many points were not covered by LiDAR data, especially for water surface elevation; of the remaining 607 points 'in water' (above, Figure 3), 216 points did not have LiDAR surface elevation data, mainly because of their uneven distribution (above, Figure 5). In addition, 12 pairs of transect points had in-between distances less than 28 cm. From these pairs, the point lacking any additional data was removed, or the point closest to the riverbank.

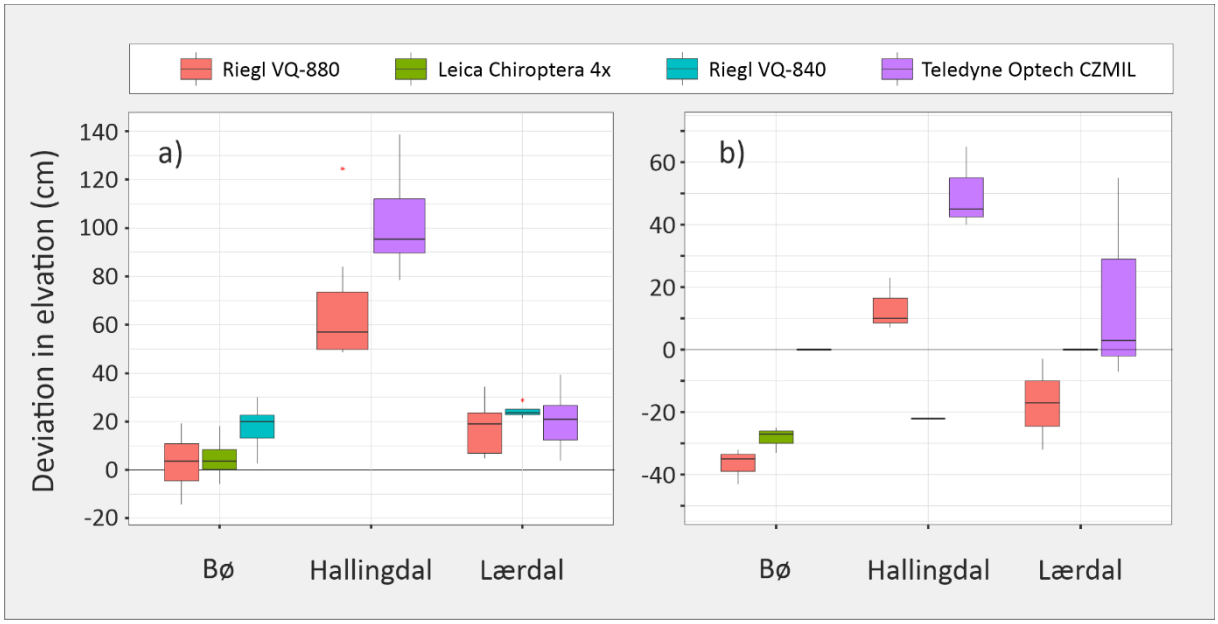


Figure 5. Differences in measured water surface elevation between control data and green LiDAR data. Left: In situ control data measured in transect points in the field with dGPS. Right: In situ control data measured with Multibeam echosounder. The elevation from green LiDAR is calculated as the peak in a frequency distribution of points classified to the surface, and for the Multibeam data as the 99% quantile of points with scan angle larger than 80° (below).

One side-channel transect in Lærdal lacked green LiDAR data. For another, registered LiDAR depths were only 0-8 cm. These data were removed.

Because most transects were in relatively slow flowing river reaches, water surface elevation was similar across all transect measurement points, and surface elevation was mostly represented by one point measurement. Therefore, we increased the sample size by adding this value (or the mean value if more than one surface elevation measurement) to all points 'in water' in the transect for which there were also LiDAR data, to estimate water depth.

Even after this extrapolation, however, elevation (in particular, surface) could not always be estimated from the LiDAR data within the 28 cm transect point radius. Therefore, data were also aggregated within a 56 cm radius (Table 4), roughly corresponding to the larger grid area of 1m² used for the later green LiDAR data analyses. In an additional step, the aggregated 56 cm circles around points along each transect were amalgamated into a 56 cm wide rectangle along a centerline covering the measurement points along the transect (Appendix 1). This extrapolation from the original *in situ* point/associated area size (0.5 x 0.5 = 0.25 m^2), may potentially introduce more noise, if it includes more differences in elevations. On the other hand, it also includes more green LiDAR points and thereby sample size. Finally, if the number of green LiDAR points within this rectangle was less than 10, the rectangle was extended to include a minimum of 10 LiDAR surface elevation points (sensor with the fewest points). These transect polygons were manually controlled against aerial photos to ensure that no river areas with higher gradient/rougher surface were included. With this extrapolation (Appendix 1), most transect measurement points could be included in the comparative analyses of water surface elevation and water depth. In total, the 28 cm and 56 cm approaches left 473 and 501 transect points for further comparative analysis (Table 4).

A comparison of river bottom elevation measurements from the 28 cm and the 56 cm areas did not indicate any significant difference in elevation (paired t-test; mean difference = 12 cm, p = 0.12), and both scenarios were analyzed further.

Table 4. Number of transect data points/associated areas (56 cm radius) available for comparative analyses. (For a radius of 28 cm, the number of comparative points/associated areas were reduced to 473). All registered wading points which also provided minimum one LiDAR hit (bottom and surface), positive depth, and classified substrate.

	dGPS		LiDAR				
River	**Transects (per km)**	**Points (n)**	**Riegl VQ-880**	**Leica Chiroptera4x**	**Teledyne Optech**	**Riegl VQ-840**	**SUM**
Bø	27 (3.5)	143	125	104		64	334
Hallingdal	11 (2.3)	74	47		22		71
Lærdal	15 (1.0)	74	62		60	17	145
SUM	53	291	234	104	82	81	501

2.5. Multibeam data

The Multibeam echosounder mapping applied on pre-selected pools in the three rivers provided high point density for river bottom elevation measurements, up to 86 000 points per m^2. In comparison, overall point density for green LiDAR data classified as river bottom varied between 7 and 142 per m^2. Thus, for later spatial grid analyses between green LiDAR and Multibeam elevations, it was primarily the lack of LiDAR points that reduced the number of observations in the data set.

Water surface elevations based on LiDAR data, were estimated as the peak in a frequency distribution of points classified to the surface (= class 27). Multibeam sensors do not directly measure water surface elevation. Because pools are primarily non-turbulent still-water surfaces, and the data used here were recorded with a scan angle up to 80°, i.e., in shallower areas, the upper 99% quantile was used as the surface water elevation reference for the Multibeam data.

Data quality control indicated that a few grid cells (less than 5 %) estimated water depth turned out negative. Visual data control showed these cells to be located along the shallow edge of the water. In these cases, riparian vegetation and protruding boulders had presumably been mis-classified as river bottom. All negative depths were excluded from further analysis

2.6. Data and analysis

2.6.1. Data and variables

For analysis of the accuracy of the remote sensing green LiDAR with regression models (below), the response variable was deviation between bottom elevation measured *in situ*, either in transect points (river reaches; Table 4) or Multibeam (selected pools; Figure 1), and the corresponding LiDAR bottom elevations.

For analyses of the availability of green LiDAR data, classification models were used with a binary response variable indicating LiDAR pulse reflection or not in the areas covered by the control data. The green LiDAR signal is weakened by the depth of the water, and at a certain depth will this prevent signal reflection. Thus, 'Water depth' was included in all models, as were 'River' and 'Sensor' (below). The studied river reaches have different gradients (Table 1). Visual interpretation of aerial photos taken simultaneously with the LiDAR data, indicated substantial variation in the amount of 'white water' areas between and within river reaches. So did the riverbed substrate colours and lightness. Therefore, in an additional 'visible lightness' analysis, the RGB values (on a lightness scale, below) from the aerial images were classified and merged to the lidar pulse data. In addition to the RGB-values derived from simultaneous aerial photos, values were also extracted from the publicly available best aerial images (https://www.norgeibilder.no/) for later modelling. Unfortunately, shadows from trees and water reflections confounded RGB values in the aerial photos available for Bø, and for the studied reach in Hallingdal only a small fraction was turbulent water. Therefore, the

effect of water 'lightness', as seen in aerial photos, was studied primarily for Lærdal (se examples of light variation in Appendix 4).

Terratec delivered RGB-values based directly on the original aerial photos with 16-bits data, whereas AHM delivered data converted to the 8-bit scale. To analyze the data on the same scale, the RGB-values were transformed to a 'lightness' index ranging from 1 (black) to 100 (white).
A thematic land cover product was first derived from the orthophotos. Classification based on Terratec orthophotos was done at the original 25 mm pixel size, while AHM data was resampled from 10 to 25 cm before classification. Main classes were dark, vegetated riverbed, lighter grey solid substrate (sand, gravel, stones not covered with vegetation), and white water surface. In addition, boulders protruding above the water and overhanging riparian vegetation were mapped (cover: less than 2% of the river) and omitted from further statistical analyses. Grid cells including more than one LiDAR point classified to ground (class 2) were also removed before analyses. The two maps were produced with a pixelwise classification using a Support Vector Machine algorithm, trained from point data collected by visual interpretation of the aerial photos (3786 and 7105 points used for Terratec and AHM, respectively).

In situ classification of dominant substrate particle size indicated finer substrates in Hallingdal, i.e., dominated by sand and gravel, as suggested by the lower gradient (Figure 6, Table 1). In Bø and Lærdal substrates were similar, and generally larger, including rocks and boulders, but also covering a range of substrates (Figure 6).

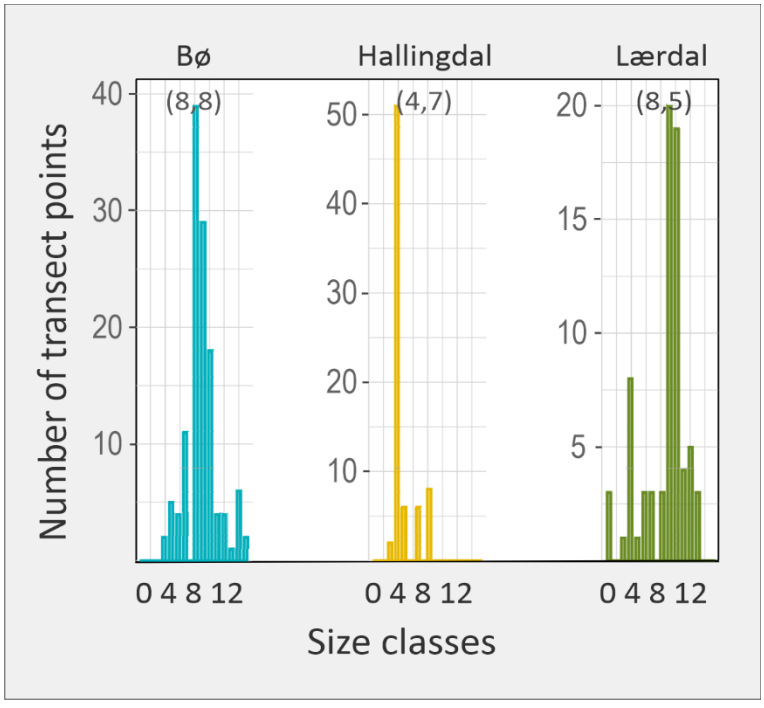


Figure 6. Riverbed particle size class distributions and variation across the study reaches in the three rivers, as classified at the transect points. Value in parenthesis on top is mean class.

2.6.2. Data challenges and limitations

Data tended to be unbalanced. Unfortunately, not all sensors were used in all rivers (Table 3, 4, Figure 4). For Dataset 1, the transect control data (Figure 3), this unbalance was aggravated by the fact that number of transect points covered also varied among sensors, because of different water flows between *in situ* and remote sensing data collection in Hallingdal (above, Table 3, Appendix 1). Also, green LiDAR points indicating riverbed elevation, varied across sensors, not only because sensors have different point densities (Table 2), but more so because water penetration capabilities between sensors, mainly effected in the pools, turned out to vary considerably (Table 2, Appendix 2).

For Dataset 2 there were major differences among pools (and rivers) in number of analysis units (grid cells), reflecting the pools' different surface areas, e.g., for grid size 1 m there were 1 734 cells in the middle pool in Bø (Figure 1), but 43 967 in the middle pool of Hallingdal (Appendix 2).

However, the large pool datasets allowed for balancing the data by randomized selection of a set number of datapoints from each pool dataset. That number was set to the smallest number of analysis units available for any pool, which still provided large datasets presumably generating little sampling error (Appendix 2). For evaluating the LiDAR data accuracy, 1000 grid cells per pool were selected, and for evaluating LiDAR data loss, 2000 grid cells were selected from each pool both for the 100 x 100 cm and 25 x 25 cm grid size. The Riegl VQ-840 (helicopter) data covered more limited pool surface areas than the fixed-wing data, and therefore may potentially be less representative, especially for grid cells 100 x 100 cm (Appendix 2).

2.6.3. Data structure and statistical models

The transect control data have a hierarchical (nested) structure, since each river reach has a set of transects, which again consist of 4-11 point measurements. Similarly, the Multibeam data are for each river from three within-river pools which again consists of many grid cells (Appendix 2). The analysis units are nested within rivers, and rivers are crossed with different sensors.
For a nested and crossed design like the present datasets, there are some underlying potential correlations that may not be considered in the regression and classification models: The studied river reaches, and green LiDAR sensors, may have properties that influence the green LiDAR pulses, possibly reducing within river and sensor variation.

Ordinary least squares regression assumes independence, and without pseudo-replicates (Hurlbert 1984; Hurlbert 2009). Such violations may lead to significance for factors that may be of little importance. Tentative analyses of residual plots for transect/pool data did suggest that such linear models did not address all variation. Generalized mixed regression models (GLMM) may help consider such underlying data correlations, by introducing specified random factors in addition to the ordinary fixed factors. If more information about a random factor is wanted, it may be included in the model specification as both a random and fixed factor. Therefore, the accuracy and availability of

green LiDAR were first explored by multivariate generalized linear mixed models (Figure 4) (e.g., Bolker *et al.* 2009; Harrison *et al.* 2018; Silk, Harrison & Hodgson 2020), and general additive models (GAM) (Wood 2006; Pedersen et al. 2019) for modelling more flexible regression functions and smooth functional relationships between predictor variables and the response.

Since river reaches and sensors both may have different properties, they were included as fixed and random factors in the analyses. The datasets in the present study, did generally not converge when using the frequentistic GLMM models. An alternative is to use Bayesian model approach, which tend to be flexible and more robust with unbalanced data. Therefore, the R-libraries 'INLA' (Integrated Nested Laplace Approximations) and 'bmrs' were used instead. The full Bayesian simulations in 'bmrs' are very computer intensive (Brooks *et al.* 2017; Bürkner 2017b; Bürkner 2017a), and were primarily used for control of transect data (Dataset 1, Figure 4). To balance speed and flexibility the 'INLA' modelling was preferred (Lindgren & Rue 2015; Rue *et al.* 2016; Niekerk, Bakka & Rue 2021). The 'brms' gave similar results to the INLA, and only INLA results are reported.

For model selection, the Deviance Information Criterium (DIC) was used (Ando 2007; Ward 2008; Maity, Basu & Ghosh 2021), i.e., a Bayesian parallel to the frequentistic Akaike's Information Criterion (AIC) (Akaike 1974). Model selection was based on scaled factor values, starting with a model including all relevant predictors and with interactions. To have regression coefficients directly related to the units of measurements, models are presented using the original values. However, since scaling permits direct comparison of continuous variable predictor coefficients and their importance in the model (e.g., Schielzeth 2010), the coefficients from models with scaled variables are added when considered relevant. For a frequentistic model, the importance of relevant predictors may be evaluated also by estimated p-values and confidence intervals. A Bayesian model does not provide p-values, but 95% credible intervals. The importance of a variable may therefore be judged by how far the lower or upper limit is from zero.

The categorical predictors 'River' and 'Sensor' were coded as dummy contrasts, using the omitted category as reference for comparisons, here AHM in Bø. For an alternative coding as effect contrasts, the grand mean will be the reference (here of the combined river and sensor). However, the values for Leica Chiroptera 4x sensor used in Bø drew heavily on the grand mean, making it more negative than Riegl VQ-880, thereby making it more difficult to explore potential deviations from zero.

First, a null model (without variable predictors) was investigated to see if any of the included non-variable factors (predictors) were un-informative. Then all relevant fixed and random factors were included in a full model, which thereafter was reduced stepwise by removing predictors; first random, then fixed predictors. The GLM(M) models calculate coefficients (parameters) which indicate explanatory factor direction and power, including for potential interactions. These models also output (smoothed) response curves between relevant variables. However, they also require relevant variables and interactions to be specified for a test model.

The differently structured machine learning algorithms (ML) on the other hand, are more flexible, and potential curvilinear relationships and threshold values may be identified. ML- algorithms may also be less sensitive to lack of independence in the underlying data. To make to analysis more robust, both parameterized and machine learning models are used. While the parameterized models are evaluated by an information criterium (above), the ML-approach to find the best model is evaluating sub-models continually through cross-validation. With the cross-validation technique, models can be evaluated with data not used in the development of the model. The predictive power of the ML-models was calculated by 10-fold cross-validation procedures repeated three times for both the coefficient of determination R^2 and accuracy (for classification).

The R-library 'Boruta' was used (Kursa & Rudnicki 2010; Speiser *et al.* 2019) to explore the overall importance of the relevant predictors. By using the machine learning algorithm Random Forest (Breiman 2001; Belgiu & Drăguţ 2016; Biau & Scornet 2016), Boruta adds a set of new factors which are being randomized such that the new factor values are uncorrelated to the response variable. Boruta then calculates an index for the importance of the original factors relative to the new uncorrelated factors, i.e., with importance = 0. The Boruta index will include the combined main and interaction effects for a factor.

Finally, to elucidate more in detail how factors may influence 1) accuracy of the LiDAR elevation data, and 2) the probability of receiving a reflected LiDAR signal, i.e., the water-penetrating capabilities, the algorithm Gradient Boosting Machine (GBM) (Friedman 2001; Natekin & Knoll 2013; Touzani, Granderson & Fernandes 2018) was used. From the best GBM model, Accumulated Local Effects (ALE) values are estimated (Apley & Zhu 2020; Danesh *et al.* 2022). ALE interprets how a factor (predictor) influences the results from a machine learning algorithm. The values from an ALE-analysis may be interpreted as the effect of a factor relative to the mean prediction from the data. For a given value for a factor (predictor), ALE will return how much the prediction is increased or decreased from the mean prediction. Plots of accumulated ALE-values will visualize how the different predictors, including interactions, influences the prediction.

Random forest (RF) and GBM are algorithms that use many decision trees (instead of parameterized models). RF draws (with replacement) large numbers of smaller datasets (both with respect to number of variables and data) and bases the final result on a mean of the independent smaller 'trees', i.e., datasets ('bagging'). This 'bagging' approach makes the model more robust, especially for noisy data. On the other hand, the independence of this learning structure may reduce the precision for less noisy data. GBM use a 'boosting' approach, i.e., improves the result via an iterative procedure. For each new model, the GBM attempts to improve the parts of the model that came out worst in the previous round. Thus, the GBM weighs each new model by its predictive power.

3. Results and comments

3.1. Data exploration for more detailed analysis of accuracy and robustness

3.1.1. Accuracy: precision in measurement

Overall, the remote sensing green LiDAR riverbed elevation data indicated high accuracy as compared with *in situ* control data. Median values were generally within deviations (*in situ* minus LiDAR data) less than ± 5 cm and not more than ± 10 cm for both the transect and Multibeam pool data (Figure 7 b, c). For the transect data, this may be within expected sampling errors, considering the difference in methods and especially the on average relatively large substrate particle classes/sizes in Bø and Lærdal (Figure 6). For the finer and less variable substrates in Hallingdal, estimated riverbed elevations were virtually identical between *in situ* transect point and remote sensing green LiDAR data (Figure 7 b).

The more extensive pool dataset with respect to both depth range (down to more than 10 meter) and number of data points, allowed for more confidence in the analyses of the effects of water depth on measurement accuracy, and also for analyses of data loss. The pool data lack spatially detailed information on substrate size, but in general variation in substrate size classes are rather narrow in the slow-deep pools, i.e., finer substrates. Instead, the high point density of the multibeam data allow the riverbed slope to be included in the models.

Multibeam control data from the deep pools indicated similar overall accuracy as for transect control data, but with considerably more spread and more outliers (Figure 7 a,b). This may, at least to some extent, result from the very large datasets from Multibeam and green LiDAR. More detailed analysis of this larger spread in deviations in measured riverbed elevations in pools did not indicate any consistent pattern, neither across rivers, nor sensors. Visual inspection of grid cells exhibiting deviations of more than ± 50 cm, indicated that many of these were close to the riverbanks and to riverbed with steep slopes, and to grid cells with few data points (Appendix 3). Thus, outlying observations along the riverbanks may tentatively be caused by noise, rather than sensor technical errors, i.e., disturbance/misclassification of the green LiDAR data because of (overhanging) riparian vegetation and less precise Multibeam riverbed elevation data because of the increasing angle in shallower areas. Outlying data included in a statistical model cause noise and may mask the strength of main effects. Therefore, the grid cells covered by riverbed vegetation in the aerial photos acquired during the green LiDAR recordings and the Multibeam data with recording angle more than 80° were removed before further analyses, as were deviations of more than ± 50 cm. This amounted to less than 2 % of the data.

Furthermore, deviations in riverbed elevations were not strongly or consistently related to water depth (Figure 8), neither for any sensor or river combination, nor for transect point or Multibeam

data. Still, variation in deviations appeared higher for *in situ* transect river measurements, than for Multibeam pool measurements. This may at least in part, be related to the greater substrate particle sizes in the more high-gradient river reaches with transect point measurements.

Moreover, the LiDAR data for riverbed elevation tended to vary more than the corresponding Multibeam data, i.e., in the pools. Consequently, water depth from LiDAR data were estimated as the difference between water surface elevation from green LiDAR data and the riverbed elevation based on Multibeam data.

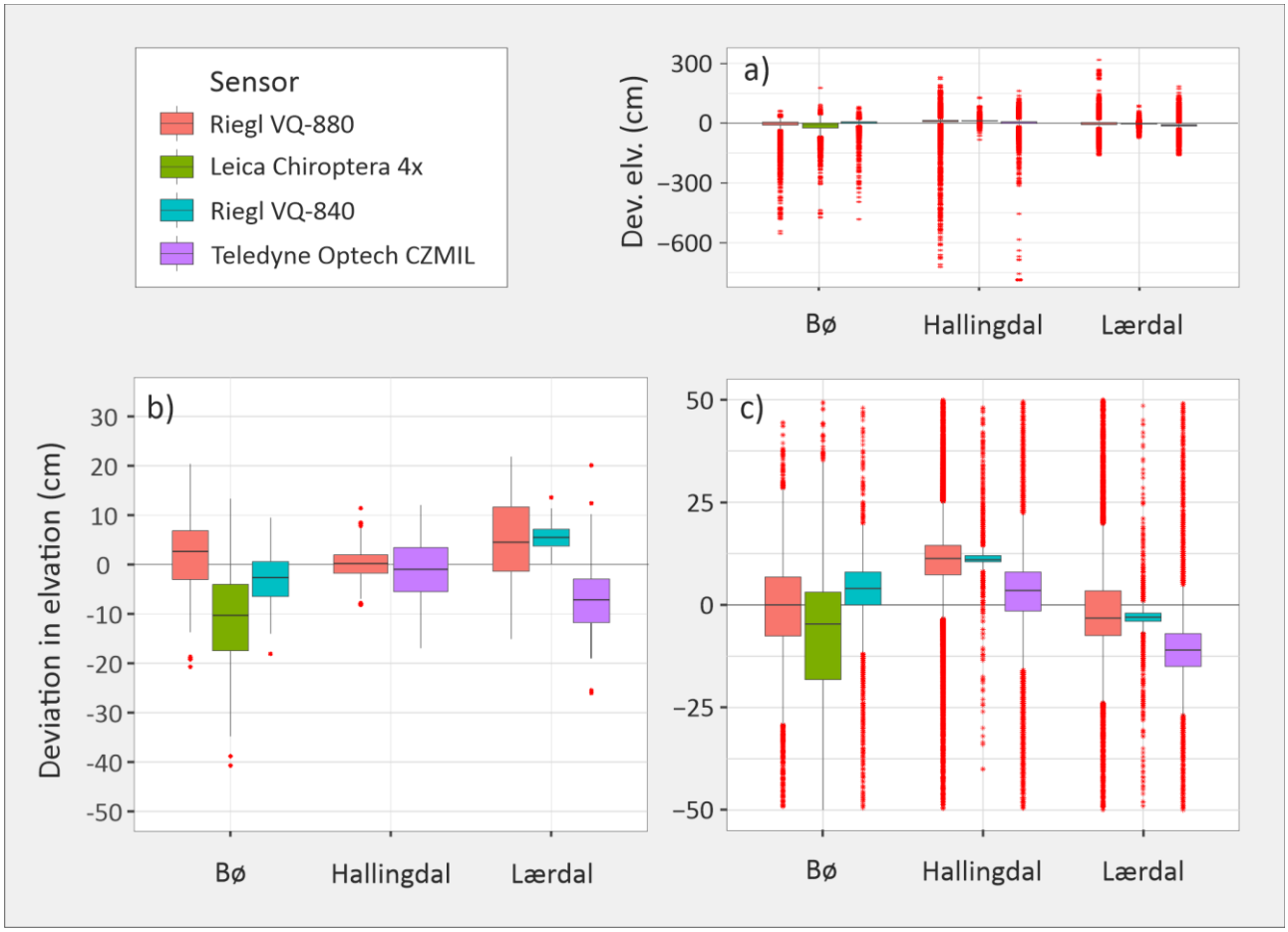


Figure 7. Deviations in estimated riverbed elevation between in situ control and remote sensing green LiDAR data. a) Control data measured at points in situ with dGPS. b) Control data measured in deep pools with Multibeam sonar, c) Control data measured in deep pools with Multibeam sonar but with data outliers larger than 50 cm removed. The LiDAR data in a) are calculated as the median value of points within a 56 cm buffer around the control point and in b) for grid cells of 100 x 100 cm.

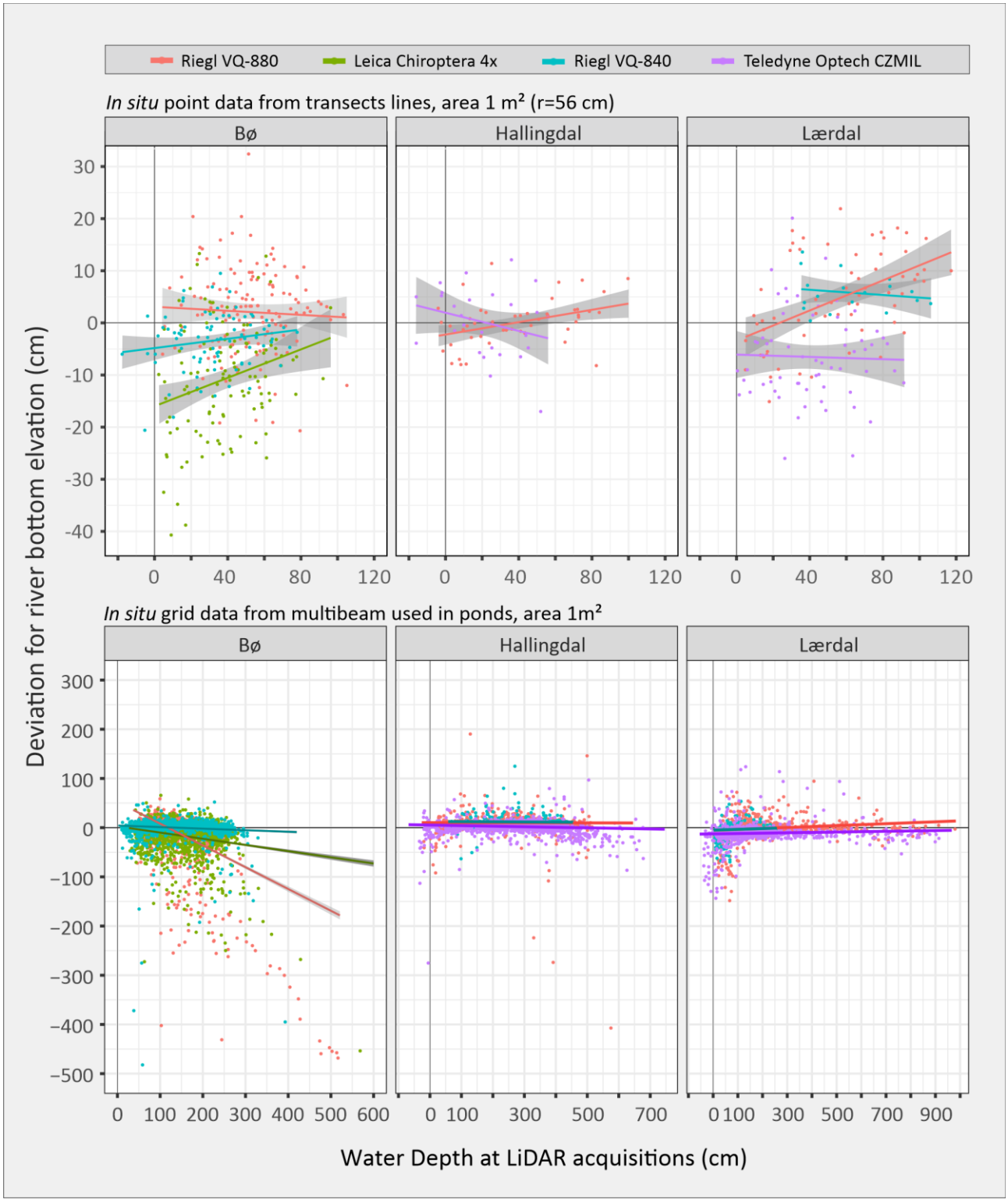


Figure 8. Deviations in riverbed elevation relative to water depth. No consistent pattern was indicated, neither across rivers, nor across sensors. LiDAR data were calculated as the mean value a) of LiDAR point elevations (within a 56 cm buffer around the in situ transect control point) compared to transect point elevation, and b) for grid of 100 x 100 cm compared to pool Multibeam data. The points in b) represent balanced random sub-samples of 1500 data points from each river and sensor combination.

3.1.2. Robustness: Green LiDAR data loss

Visual exploration of the green LiDAR data indicates a pattern where the loss is located to the deepest parts of the pools, and not to riverbanks (Appendix 3). Therefore, all grid cells with a positive water depth were used in the analysis of data availability.

Green LiDAR data loss was generally small in shallower areas but appeared to increase rapidly at a critical depth (reached in the deeper pool areas), and this pattern was the same across rivers (Appendix 3), with the greatest loss in Bøelva (up to 70%) and least in Lærdal (up to 7%), and to a lesser extent also across sensors. The difference and abruptness in data loss between rivers suggested an effect of water visibility, as indicated by Secchi depth (Table 1, Appendix 3).

Water surface turbulence ('white water') may potentially also cause loss of green LiDAR data. However, the tentative proxy variable 'Lightness index' ('white water' due to air bubbles as derived from the RGB-values in the simultaneous aerial photos) did not show any clear pattern with turbulent water, neither for the transect point nor the Multibeam data. Manual control indicated that the 'Lightness index' also turned out to be affected by especially green riverbank vegetation and incoming light/shadows along the forest lined river reaches in Bø and Hallingdal (Appendix 4). In the open landscapes along Lærdal, manual inspection indicated that the lightness proxy did reflect 'white water', i.e., surface turbulence, allowing for these analyses in Lærdal.

3.2. Accuracy in the measurements of river bottom elevation

Overall, the more detailed regression analysis corroborated the results from the data exploration. Since the main interest was to explore deviations between control and green LiDAR data, the absolute deviation values were first explored as the response. However, this would mask any directional tendency for LiDAR data (over- or underestimate riverbed elevation). For example, surface refraction might generate consistent bias. A GBM model for the transect data (dataset 1) explained 42 % of the cross-validated variation with directional deviation as response, compared to only 27% when the absolute value of the deviation was used as response. For pool data (dataset 2), the difference in explained variation was similar. The preliminary analyses indicated directional deviation to be more informative than absolute values and directional deviation was therefore used as the response in the analyses.

3.2.1. dGPS point data from transects

The Boruta approach of using uncorrelated dummy variables with the Random Forest (RF) algorithm identified Sensor as the most important variable for the accuracy of green LiDAR measured from the transect data (Figure 9). Most of the data from transect points represented water depths around 45 cm, with range down to 1.2 meter. This range might be too shallow to detect effects of water depth.

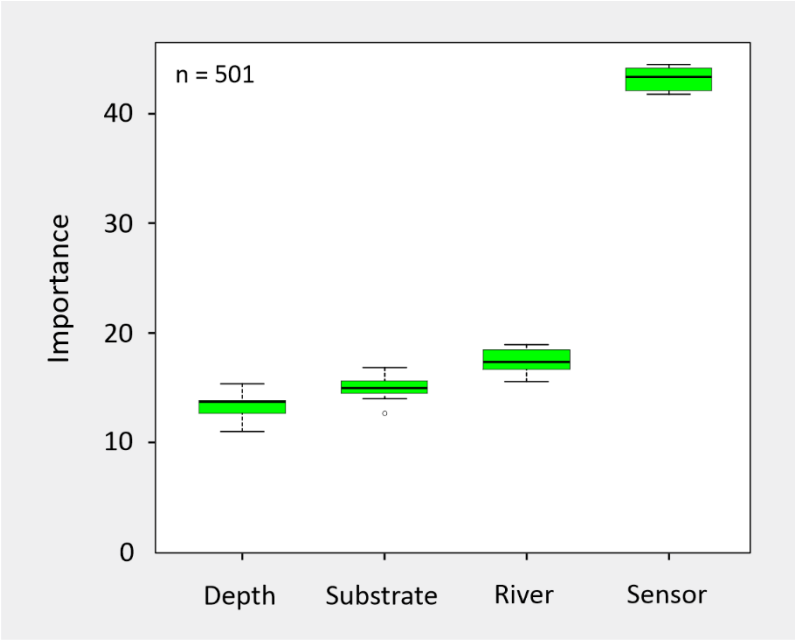


Figure 9. Importance of the predictors in a Random Forest model in explaining the response variable deviation in riverbed elevation between in situ transect points compared to remote sensing green LiDAR. The index is calculated via randomization of variables in a Random Forest model using the R-library Boruta.

The 'best' INLA mixed-regression model, derived from the DIC selection procedure, included as fixed predictors the same four variables Depth, Substrate, Sensor and River, and with River and Sensor, as well as Transects/points, as important random effects (Appendix 5). Results were similar for a 56 or a 28 cm radius around transect points, and with 6 % more data points (from 493 to 501) in the model and less variation, the 56 cm model is shown.

The more detailed analysis results from the parameterized INLA model (Appendix 5), indicated the effect of increasing Substrate class is negative, meaning that with increasing substrate particle size, the green LiDAR estimates riverbed elevation to be somewhat higher than *in situ* transect point measurements. This deviation tended to be more pronounced for the smaller 28 cm radius data (not shown). One likely explanation for this, is the use of a measuring stick for the *in situ* transect point measurements. The footprint of a stick is very small. The stick was likely placed 'on the bottom', i.e., between raised rocks, and such a measurement error would increase with increasing substrate particle size. Green LiDAR data on the other hand, measured riverbed elevation as the mean (both mean and median gave the same results) across points within the defined area (56 or 28 cm radius), and with footprints of LiDAR point to more than 20 cm, the LiDAR measured elevation would increase with increasing substrate particle size. Regardless, the effect is small, as indicated by the small coefficient.

The main effect of Depth was modelled as weakly positive (Appendix 5), and with the sensors Chiroptera 4x and Riegl VQ-840 emerging as important, and with the Riegl VQ-840 as weak. Depth is also in weak interactions with different sensors in the best INLA model. However, effects, i.e.,

coefficients, were consistently small and only with Riegl VQ-880 and Teledyne Optech CZMIL emerging as important (Appendix 5). Also, their effects were opposite, with slightly increasing deviation with depth for Riegl VQ-880 and decreasing for Teledyne Optech. This is likely explained by the fact that the sensors Riegl VQ-880 and Teledyne Optech were used in three and two rivers, respectively, and Leica Chiroptera 4x only in one. Besides, Teledyne Optech results were negative (at zero depth) relative to the Riegl VQ-880 (reference (intercept) at zero depth), which in turn was positive, i.e., increased with depth.

These depth-sensor interaction effects come in addition to the main effect, making it more difficult to see total effect. Therefore, the overall effect of water depth, as modelled with INLA and Gradient Boosting Machine (GBM) within the range of available data for river and sensor, is visualized in Figure 10, keeping substrate constant on the river mean. Again, deviations are consistently small and around 0 (generally less than 5 cm). An exception was the Leica Chiroptera 4x sensor in Bø in shallow water (< 50 cm), with green LiDAR consistently estimating the riverbed elevation about 10-20 cm higher than the *in situ* transect control measurements.

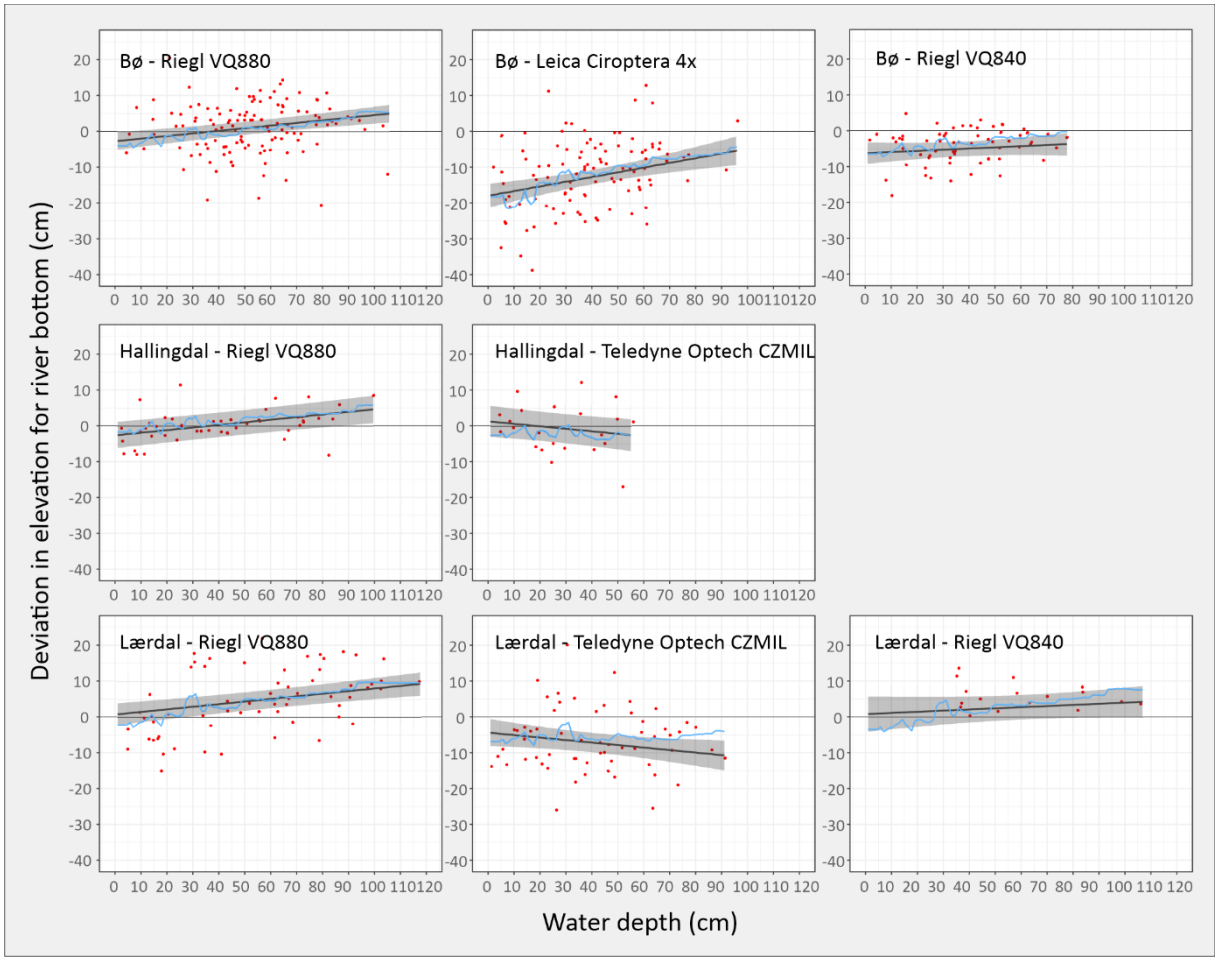


Figure 10. Deviation in riverbed elevation between in situ transect point measurements and remote sensing green LiDAR mean measurement within a 56 cm buffer area around the point. For comparison, riverbed substrate size (class) was kept constant at the mean for each river. The black line with grey credible interval indicates the INLA model result, and the blue line indicates the Gradient Boosting Machine model result. The red dots are the measured deviation for each river and sensor combination

For deeper areas, measurements were similar, but data points were few. This exception and its systematic nature may suggest some sensor calibration problem. It may also to some extent perhaps be tied in with the increased deviation with increased substrate size (above). However, it would then be expected to appear also across sensors.

The machine learning model GBM gave similar results (with R^2 = 42 %) to the INLA model, except for Teledyne Optech in Lærdal (Figure 10). Here, the INLA model suggested a weak negative trend for deviation with depth, the GBM model found no effect (even line) for the same data. However, data here were few.

As an alternative machine learning algorithm to study the influence of different predictors more in detail (not just strength), the Accumulated Local Effects (ALE) add more detail about predictor effects. Because ALE calculates the main effect of a predictor at a certain value compared to the average prediction of the data, plots of accumulated ALE-values visualize how the different predictors,

including interactions, influences the prediction. The ALE-plot for the GBM model (Figure 11) indicated again that among sensors, it was primarily Leica Chiroptera 4x, but also Teledyne Optech, that influenced the model negatively, i.e., measured the riverbed elevation higher than the *in situ* transect point control data. Moreover, as suggested by the INLA model, the ALE model also suggested a slight positive effect on deviation with increasing water depth. Assuming no bias in the *in situ* transect point measurements with increasing depth, the green LiDAR data tend to measure slightly lower riverbed elevations with increasing depth. However, this deviation may also result from the fact that *in situ* measurements become less precise with increasing water depth. Wading and accurate measurements become more difficult at greater water depths and velocities further out in the river. The ALE-analysis also suggested that the increased deviation with increased substrate particle size (above) tends to be linear (up to class 13; bedrock, also very few data point thereafter, only 4 % of the data).

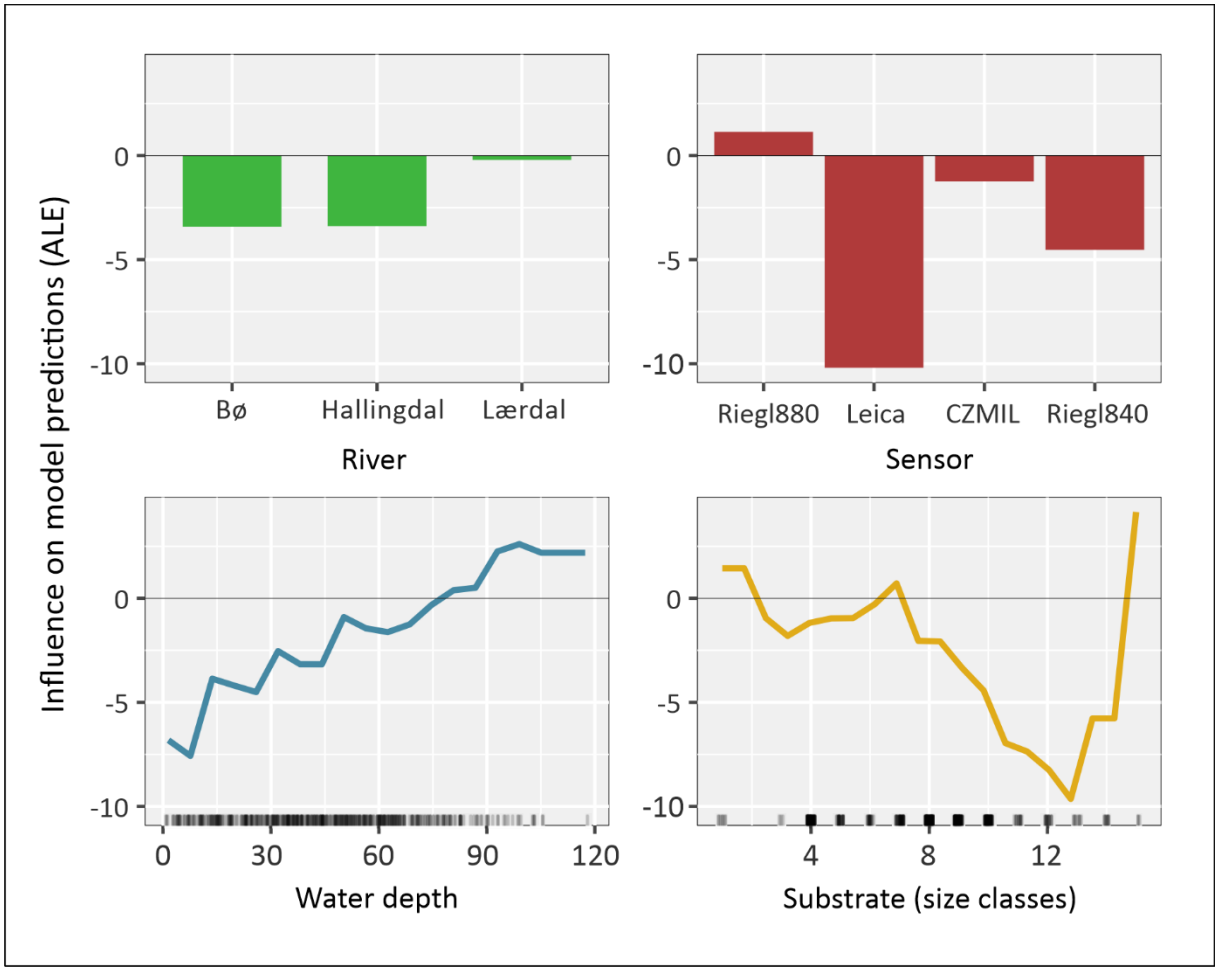


Figure 11. The Accumulated Local Effects (ALE) values visualizing the influence of the different predictors (including interactions) in the Gradient Boosting Machine model in explaining the response, i.e., the deviation in riverbed elevation between in situ transect point measurements and remote sensing green LiDAR. The main effect by the Leica Chiroptera 4x sensor (upper right panel) may be due to a calibration issue. The apparent rise in influence by the largest substrate size classes is likely an artifact due to few data points. For name abbreviations, see Table 2.

Loss of LiDAR data for riverbed elevation (56 cm radius) amounted to 21 transect points and all from Leica Chiroptera 4x data. These transect point losses were distributed across water depths and substrate size classes and reflected a clumped distribution of reflected points not associated with water depth. Thus, without data of loss linked to the water depth or substate size, data were not analyzed further.

3.2.2. Multibeam point data from pools

With the pool data, Sensor emerged again as an important variable from the Boruta analysis, but not more than River and Slope (Figure 12). Most data in the pools were from water depths around 2.5 meter and ranging down to more than 10 meters (Lærdal). The increased importance of Rivers in the pools compared to the transect data, might be due to an emerging effect of different water visibility across rivers with deeper water.

The 'best' INLA regression model for green LiDAR versus Multibeam data (1 x 1 m grid) indicated as important the fixed predictors Slope, Depth (including interactions with sensor and river), sensor Leica Chiroptera 4x, and river Lærdal. Important random effects were river, pool/grid cell, and sensor (Appendix 6). The coefficients for the fixed effects, and the variation of the random effects are very small, with differences between pools showing the highest variation among the random effects (SD = 4.3 cm). The negative association with increasing slope, means that the green LiDAR data estimated riverbed elevation to be higher than the Multibeam data.

Additional regression analyses with the smaller grid size 25 x 25 cm (not shown) indicated that the predictors River and Sensor became somewhat more important in influencing riverbed elevation deviations, whereas the influence of Depth was reduced. This was expected, as a decrease in grid size may increase within-grid differences and variation in aggregated elevation data, particularly when measurement points are few and sporadic. The green LiDAR data has many more grid cells with few data than did the Multibeam.

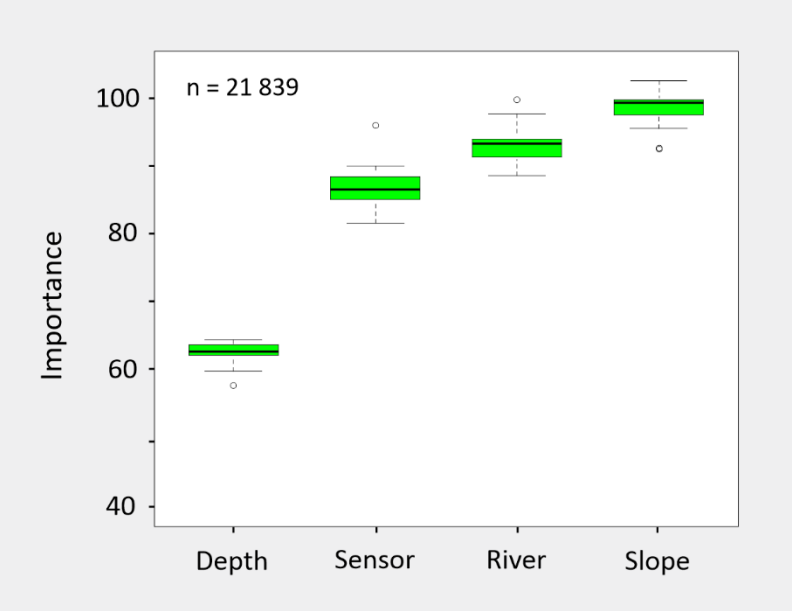


Figure 12. Importance of the predictors in a Random Forest model in explaining the response variable deviation in riverbed elevation in pools between in situ Multibeam compared to remote sensing green LiDAR. The index is calculated via randomization of variables in a Random Forest model using the R-library Boruta.

The negative effect of riverbed slope might also be amplified from the known potential noise in the Multibeam data associated high with beam angle towards the edges of the pools where also the slope is highest. Regardless, with a coefficient for the slope of -0.18, the effect of slope on the deviation between multibeam and LiDAR is small. For the maximum slope registered (70°) the model estimates an increase in deviation from a flat surface of only 12.6 cm.

The INLA regression model suggested only a weak negative effect of Depth on the deviation response (Appendix 6). The effects of water depth on sensors and rivers are, however, difficult to infer directly from the INLA model due to their interactions. By keeping the slope constant at the mean for each river, the effect of water depth (including interactions) may be seen more easily (Figure 13).

The weak effect of depth is shown by the small changes in the relationship line for most of the river-sensor combinations. The different slopes and deviations and with LiDAR data both above and below the Multibeam data, suggested that the depth effect may to some extent stem from an offset for some of the LiDAR sensor-river data, e.g., Riegl VQ-880 in Hallingdal and Teledyne Optech in Lærdal. As suggested for the offset measurement results for the *in situ* transect point data compared to remote sensing green LiDAR (Leica Chiroptera 4x) data above, this offset also for Multibeam versus green LiDAR data may stem from LiDAR calibration issues. Also, Multibeam data, typically very accurate in deeper areas, may become more 'noisy' in shallower water and for sloping riverbed elevations due to the increasing sensor/footprint angle (above) (e.g., Hughes Clarke, Mayer & Wells 1996; Calder & Mayer 2003).

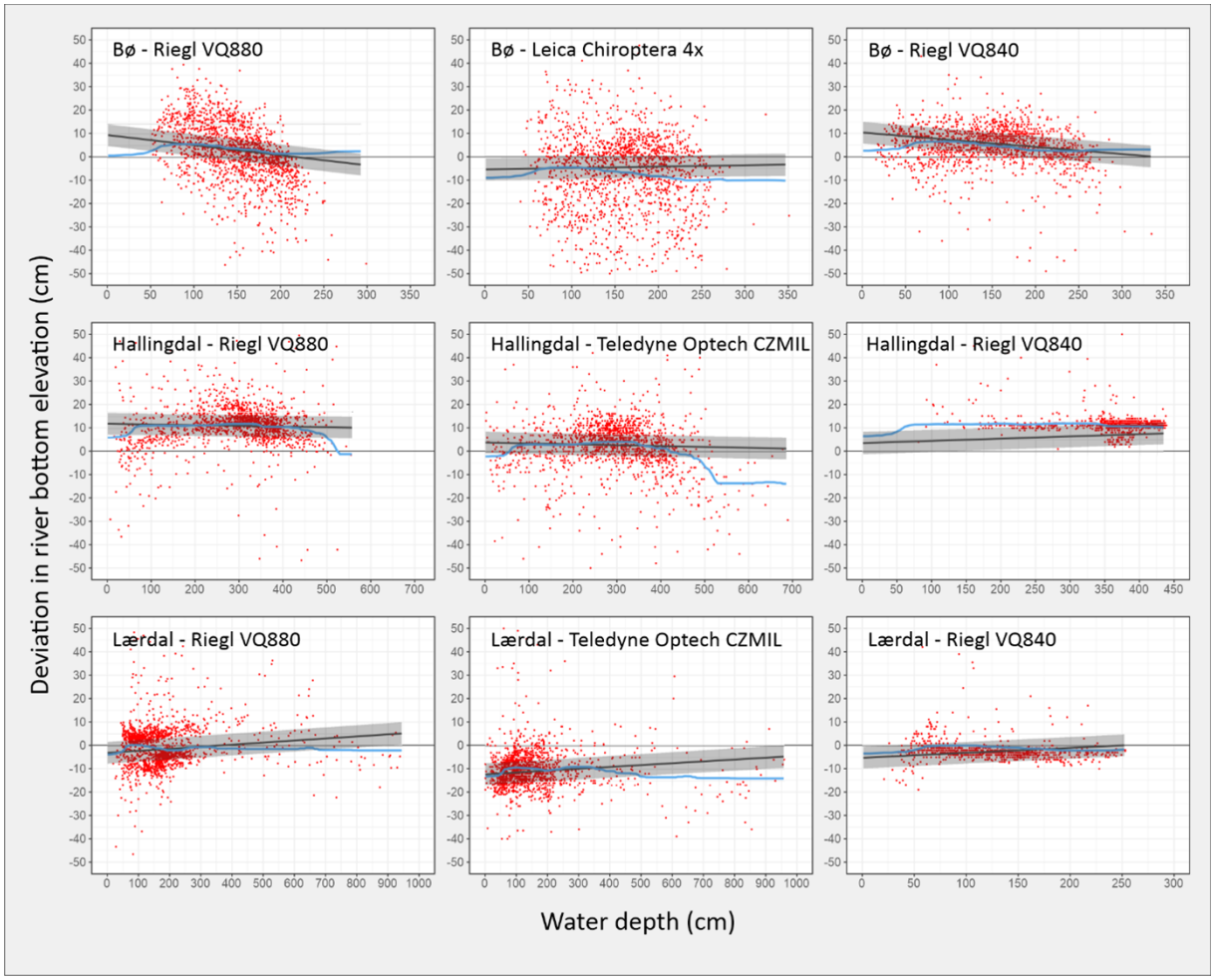


Figure 13. Deviation in riverbed elevation between in situ Multibeam measurements and remote sensing green LiDAR mean measurement within a 100 x 100 cm grid. The black line with grey credible interval indicates the INLA model result, and the blue line indicates the Gradient Boosting Machine model result. The red dots are a random sub-sample of 1500 datapoints for each river and sensor combination out of the 43 052 datapoints used in the development of the model. Outliers (deviation > 50 cm) were removed before running the model.

The result from the INLA regression model, suggested only a weak negative effect on the deviation, was confirmed by the machine learning algorithm GBM. However, the GBM model only explained up to 39 % of the deviation between Multibeam and LiDAR data. Again, the relatively low explanatory power imply that the predictors slope, depth, sensor, and river have little influence on measurement accuracy, as also seen from the small deviations with slope close to zero in Figure 13. Therefore, further analyses of measurement deviations using these procedures, were not pursued. Instead, the focus was on the increasing data loss with increased water depths. Overall, the results from the accuracy analysis indicate small measurement deviations, and when present and systematic, the cause is likely due to calibration issues.

3.3. Robustness of Green LiDAR: loss of data in deep pools

3.3.1. *In situ* Multibeam in pools

Whereas water depth tended to have little effect on measurement accuracy, and calibration issues appeared to be more prominent, water depth had a major effect on the availability of green LiDAR signals, or loss, as indicated by the Boruta analysis (Figure 14). The Boruta analysis, with data loss or not as response variable, also suggested that river was a considerably more important predictor than sensor and slope for receiving a reflected green LiDAR signal (se also % loss in Appendix 2).

The 'best' INLA binary regression model (green LiDAR signal or not), indicated a similar difference in importance between depth and slope, with depth having a coefficient more than 20-fold larger than slope in an analysis with scaled values (Appendix 7). The low variation for the two continuous variables indicated a consistent pattern. Most of the predictors, together with their associated interactions, in the 'best' model came up as important, probably due to the high number of datapoints analyzed.

Again, the interpretation of the strength of the predictors is made more difficult by the interaction terms and is therefore more easily interpreted when plotted for each of the different river-sensor combinations (Figure 15). The figure clearly shows how the probability of receiving a green LiDAR signal abruptly drops at a critical depth rather specific for each river; around 3 m for Bø and 5 m for Hallingdal, i.e., corresponding to their Secchi depths (Table 1). The differences between the rivers are most easily seen for the Riegl VQ-880, which was used in all three rivers.

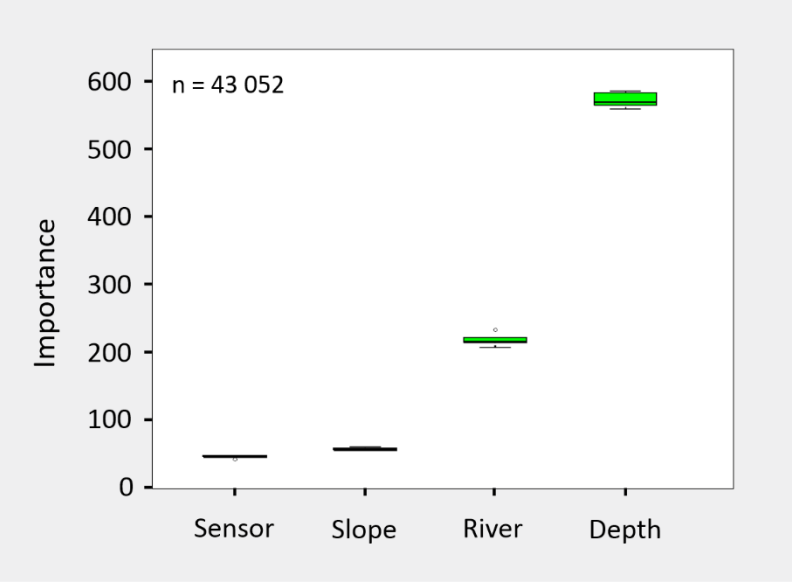


Figure 14. Importance of predictors on the binary model with response green LiDAR signal loss or not, with in situ data by Multibeam compared to remote sensing green LiDAR. The index is calculated via randomization of variables in a Random Forest model using the R-library Boruta.

Also, with data from the same river, the pattern of the sensors Leica Chiroptera 4x and Teledyne Optech may be compared with Riegl VQ-880, indicating deeper penetration. The Riegl VQ-840 only covered deeper pool areas in Bø. Here, this sensor appeared to penetrate slightly deeper that the two other sensors used. The abrupt drop in probability of receiving green LiDAR data was also dependent of sensor, with a somewhat wider 'drop' range for Teledyne Optech, which had the more powerful sensor, but with lower point density. In Hallingdal Teledyne Optech penetrated about 1 m deeper than the Riegl VQ-880. Lærdal was, however, different with no abrupt 'drop'. This reflects the fact that for the clear Lærdal, no pools were deep enough to estimate Secchi depth. Essentially, the bottom identified visible depth. For the less penetrating Riegl VQ-880 sensor, there appears to be a change in probability around 9 m depth. These results fit well with the more visual data presentation of overall data loss (Appendix 3).

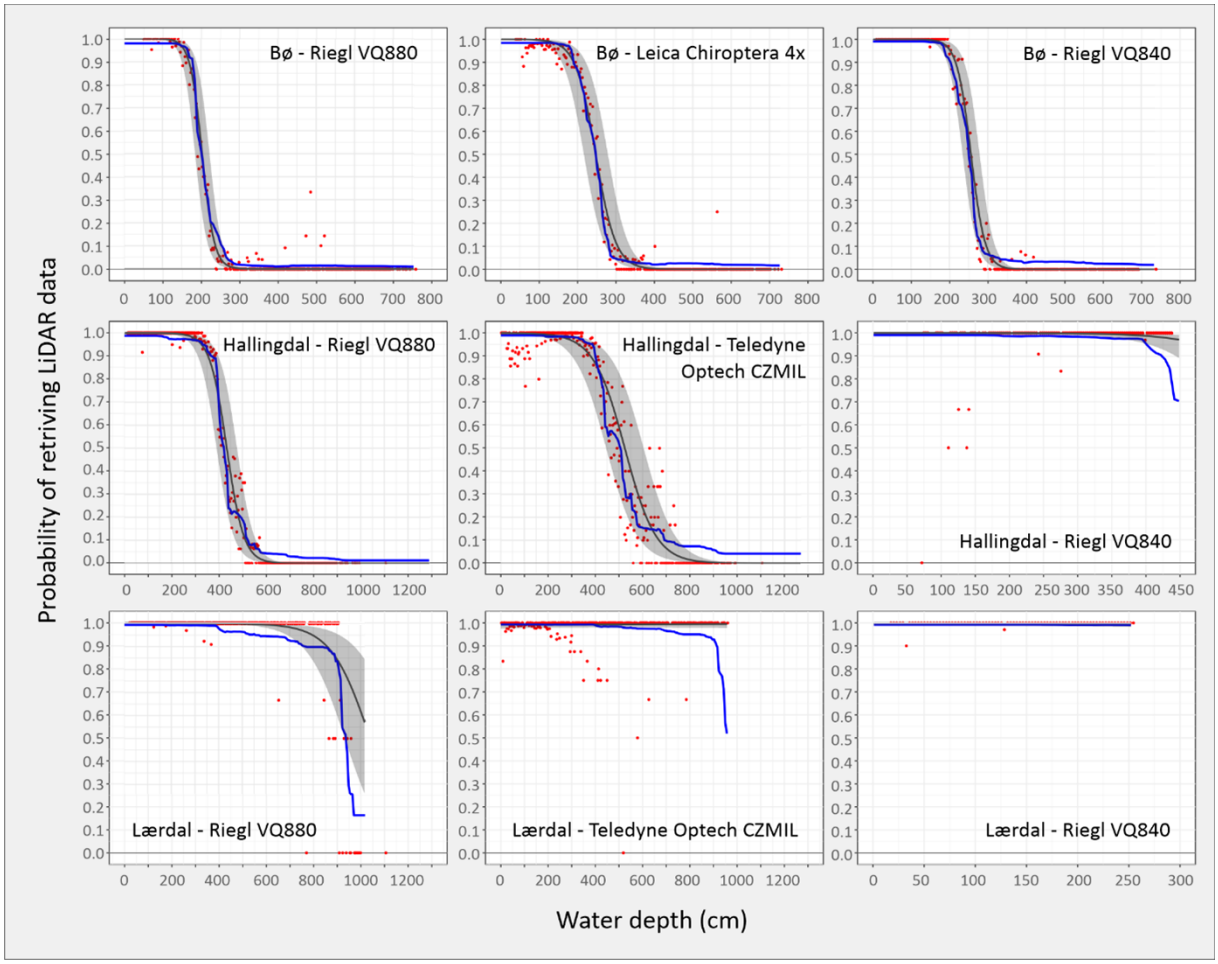


Figure 15. The probability of retrieving green LiDAR data as a function of water depth calculated from binary response models with grid size 100 x 100 cm (n = 43052). The slope was kept constant at the mean for each river. The black line with Credible Interval indicates the INLA model results, and the blue line the Gradient Boosting Machine model results. The red dots are the proportions of LiDAR data received within groups of 4 cm depth for each river and sensor combination.

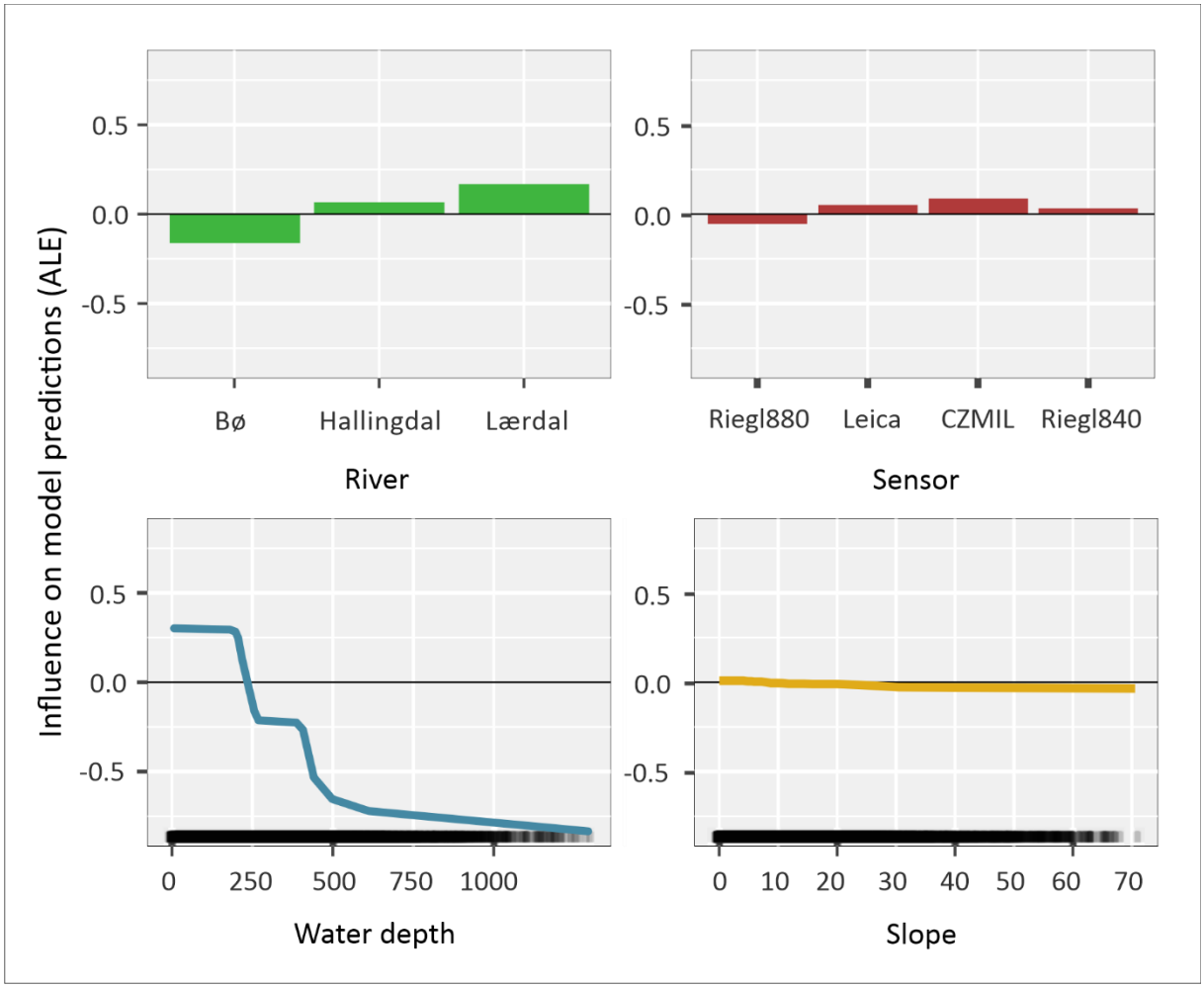


Figure 16. The influence of the predictors on the probability of receiving at green LiDAR return signal or not in the binary Gradient Boosting Machine model. Grid cell size was 100 x 100 cm, n = 44 000, and the GBM R^2 = 94.4 %. The Accumulated Local Effects (ALE) values indicate how predictors influence the predictions of a machine learning model on average, including interactions. For name abbreviations, see Table 2.

The combined effects of the three rivers and four sensors on the probability of receiving a green LiDAR return signal, summarized in the ALE-plot from the GBM algorithm (Figure 16), clearly indicate the threshold depths around 2-3 and 5-6 m for receiving a return signal (lower left panel). Slope does not seem to influence this probability much. The influence on receiving a signal is lowest in Bø and highest in Lærdal, whereas the effect of sensor is more marginal. These results indicate the importance of water visibility for the robustness of green LiDAR signals.

To explore further the potential effects of higher resolution and water depth, the analyses were also run for grid size 25 x 25 cm. To balance effort and output and focus on the deepest areas with the largest potential for green LiDAR data loss, these analyses used data from the deepest pool in each river. A 10 000 data point randomized sub-sample was extracted for each sensor.

As expected, the higher resolution and the deepest pool areas increased the probability of data loss (Table 5, Appendix 8), and especially for the Teledyne Optech sensor. With the 25 cm grid, the

probability of retrieving green LiDAR data with this sensor, was less than 50 %. However, this reduced probability was more related to riverbed elevation point density than water depth. With the 25 cm grid size, many cells did not have points at all.

The loss of green LiDAR data points and reduced point density, i.e., the reduced number of points classified as riverbed and total number of points reflected, was as expected largest in Bø and smallest in Lærdal (Table 5). The larger loss per unit area relative to number of points, was caused by the increase in 'empty' grid cells, i.e., without any LiDAR data.

To potentially mitigate the effect of data loss, interpolation algorithms were tentatively implemented and restricted to an extrapolation distance close to 1 meter from the nearest LiDAR point. Three different interpolation algorithms were explored using the R library LidR (TIN, kNN+iDW, MBA).

The increase in mapped riverbed areas using interpolation, was small for the 100 cm grid (Table 5). However, for the 25 cm grid the increase was significant, with the percent loss for the interpolated 25 cm grid, similar to the 100 cm grid (without interpolation) (Figure 15, 17 and Appendix 8). Moreover, this depended on sensor, i.e., point density, and was strongest for Teledyne Optech, i.e., the sensor with the lowest point density (Table 2). This sensor has highest power, i.e., better water penetrating capabilities. This gives a more homogeneous distribution of reflected points for river bottom, which via interpolation provides better area coverage with bathymetric data (Figure 16, 17, Appendix 9).

Table 5. Point density (per m^2) and loss of data for 100 cm grid cells compared to 25 cm grid cells. Data only from the deepest pond in each river.

	Point density (per m^2)				**% loss 100 cm**		**% loss 25 cm**	
River/Sensor	**Multi-beam**	**All reflected**	**River bottom***	**% loss**	**Median**	**Inter-polated**	**Median**	**Inter-polated**
Bø								
Riegl VQ-880	220	20.8	1.1	94.9	70	61.5	79	73
Leica Chiroptera 4x		11.8	0.2	67.2	66	51.7	88	70
Riegl VQ-840		77.0	1.5	79.0	52	43.3		
Hallingdal								
Riegl VQ-880	104	75.6	2.5	96.8	43	36.6	54	45
Teledyne Optech		22.6	0.4	81.7	5	2.0	64	6
Lærdal								
Riegl VQ-880	177	18.1	5.4	71.2	2	0.8	14	3
Teledyne Optech		1.9	0.4	74.7	2	0.0	66	5

** Classified as river bottom*

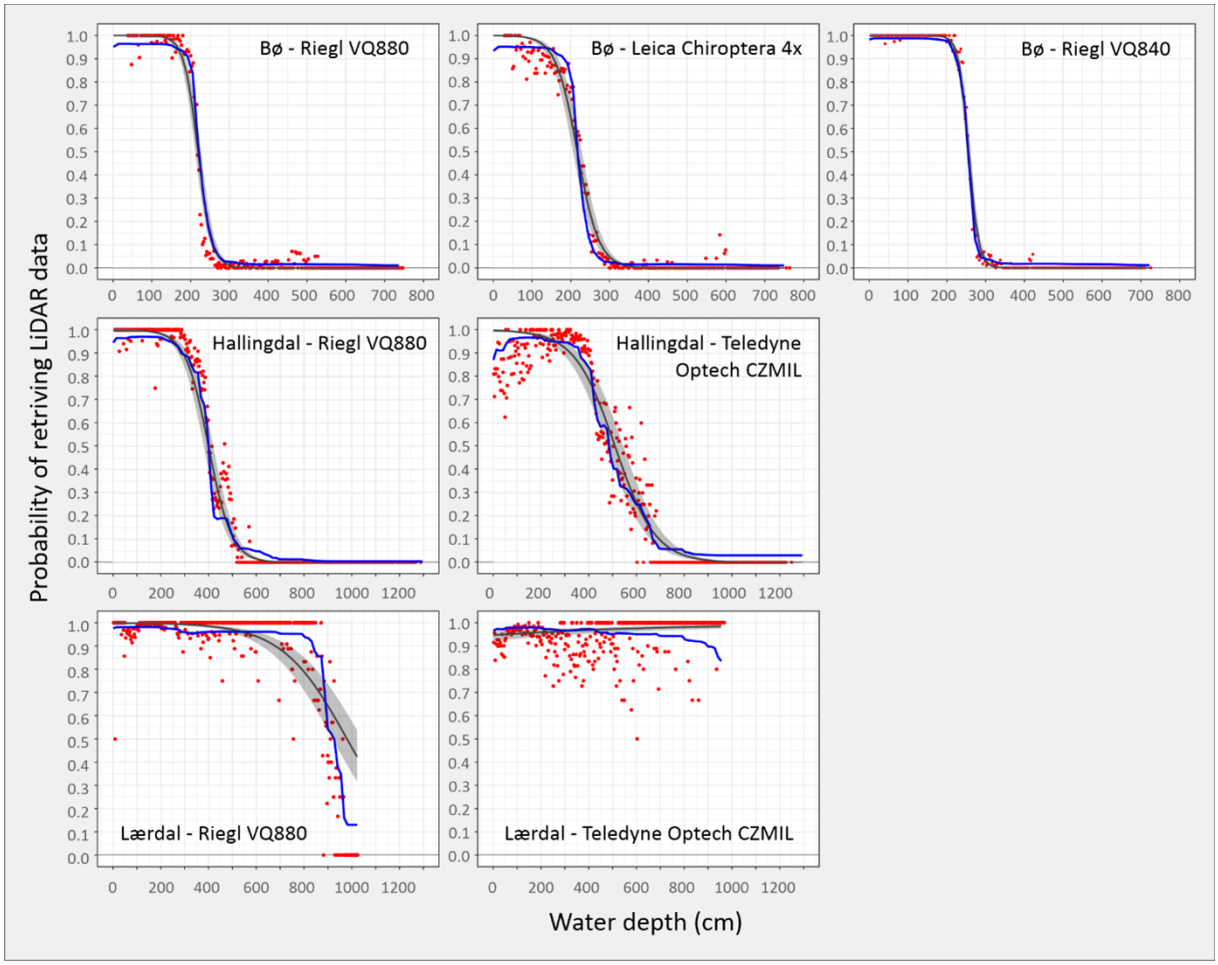


Figure 17. The probability of retrieving green LiDAR data as a function of water depth calculated from binary response models and grid of 25 x 25 cm ***with interpolation*** *(n = 70 000). The slope was kept constant at the mean for each river. The black line with Credible Interval indicates the INLA model results, and the blue line the Gradient Boosting Machine model results. The red dots are the proportions of LiDAR data received within groups of 4 cm depth for each river and sensor combination.*

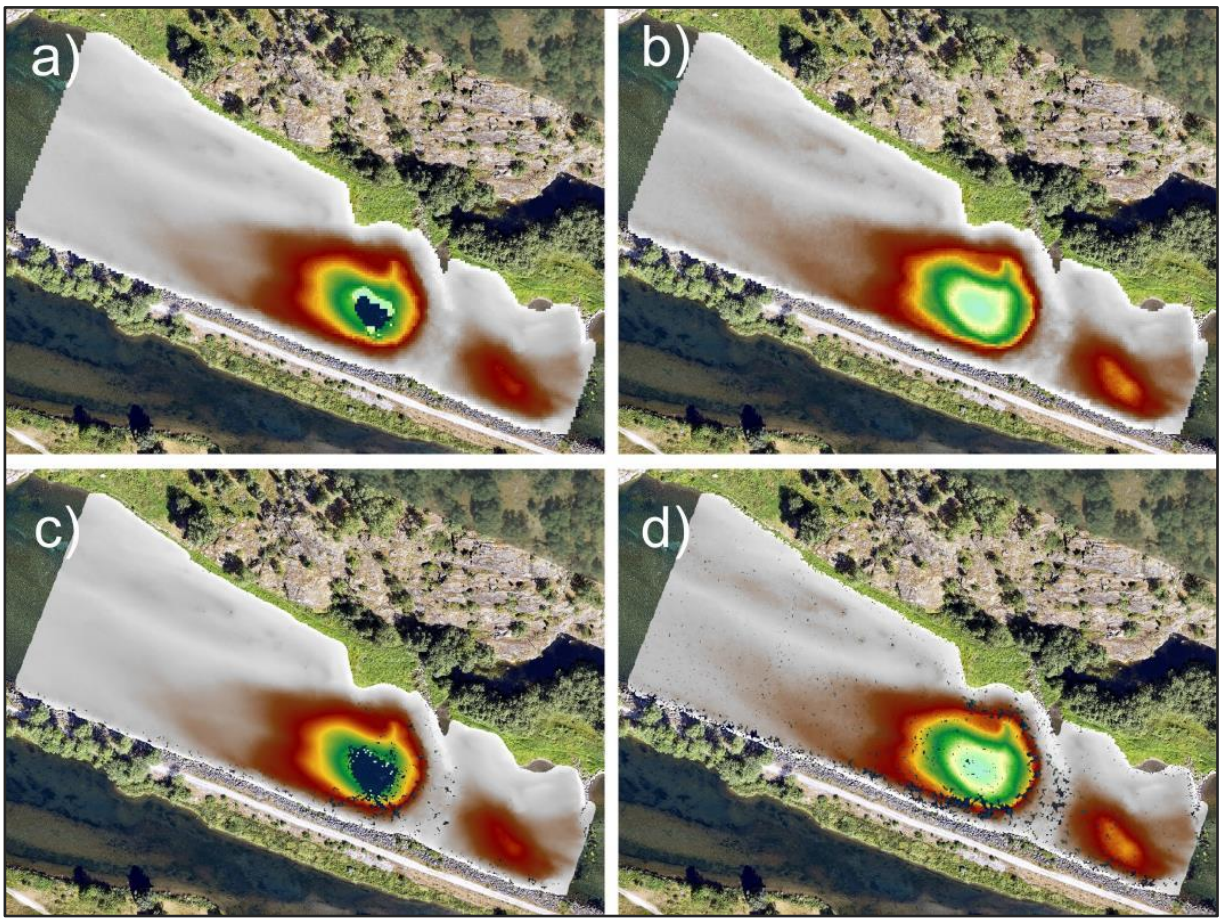


Figure 18. Bathymetric model for the deepest pool in Lærdal (about 9 m) as estimated via interpolation (limited to approximately 1 m from a LiDAR point) of the green LiDAR data with a grid resolution of 100 cm and 25 cm, and for two different sensors. a) Riegl VQ-880 and 100 cm grid, b) Teledyne Optech CSMIL and 100 cm grid, c) Riegl VQ-880 and 25 cm grid, d) Teledyne Optech CZMIL and 25 cm grid.

A set of GBM models were used to sum up the effects of different predictors and interpolation on the power and accuracy in estimating correct fate of green LiDAR signals (Table 6). Removing variables and running new models emphasize the opportunities to substitute a variable for another, in contrast to the ALE which focus on the influence of variables in a particular model.

The achieve high predictive power for a 100 cm grid, the variable depth was included in all models. With also predictor river added, a proxy for some properties of the studied river reaches, notably water clarity, the accuracy was almost as high as for the most complex model with all four tested predictors included. Thus, adding sensor did not contribute much to the accuracy.

Table 6. Robustness of alternative GBM models in predicting the availability of green LiDAR data in grid cells of 100, 25 cm, and 25 cm interpolated grid. Models are sorted by their accuracy value in the first column. The proportions of grid cells with LiDAR data received will be the expected accuracy for a random model. The reduced 'success' in a 25 cm grid size was primarily due to fewer points classified as river bottom.

	Robustness*		
Variables	100 cm median	25 cm median	25 cm interpolated
Expected accuracy from the proportion	78%	50%	76%
River + Sensor +Depth + Slope	94 %	82 %	93 %
River + Sensor +Depth	94 %	82 %	93 %
River + Depth	93 %	72 %	93 %
Sensor + Depth	86 %	74 %	85 %
Depth	83 %	65 %	81 %
River + Sensor	79 %	72 %	77 %
River	78 %	61 %	76 %
Sensor	78 %	68 %	76 %

** Expressed as the mean of the accuracy of the GBM models estimated from 10-fold cross validation repeated 3 times.*

Interestingly, the conspicuous decline in accuracy seen with the higher 25 cm grid resolution (Figure 18, Table 6), was almost perfectly ameliorated when implementing interpolation. The accuracy of 94% achieved with four predictors is strong, even when the expected accuracy by chance alone was 78%. The Kappa value, which better reflect the accuracy when the proportion of signal/not signal is different, as here, was as high as 0.83 (not shown).

These results indicate a high probability of estimating the success of using airborne green LiDAR correctly, provided knowledge about river-specific characteristics, i.e., depths and (presumably) water clarity. To some degree sensors may also be important.

3.4. Robustness of Green LiDAR: effect of turbulence and dark river bottom

Analyses using the RGB-derived Lightness index as a proxy for surface turbidity ('white water') and the light and dark riverbeds in the clear and high-gradient Lærdal, indicated loss of green LiDAR signals associated with the lightness index (Figure 19). The probability of retrieving green LiDAR data was highest within the medium-to-low lightness range, with a gradual decline towards high lightness (60-70; i.e., white water), and a more abrupt decline towards the lowest lightness values (< 20; i.e., dark, vegetated bottom) (Figure 19). This did, however, depend strongly on green LiDAR point density, with much lower loss for the high point density LiDAR (Riegl VQ880 in Figure 19).

The dark areas, with lightness index mostly less than 30, correspond to dark vegetated riverbeds. The medium light reflection came from the lighter and more common riverbed type prevalent in Lærdal, mostly covered with rock or gravel substrates with no vegetation.

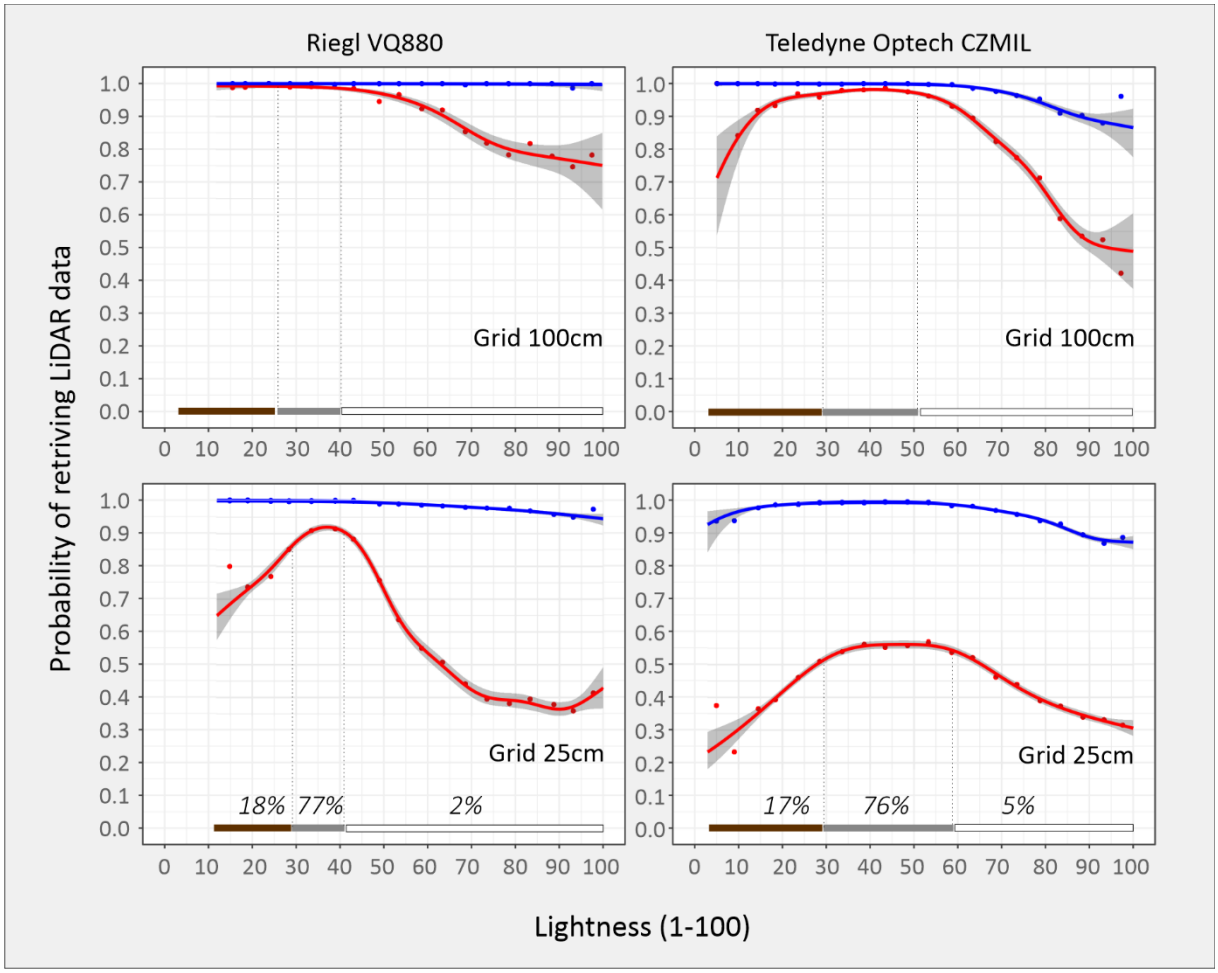


Figure 19. *Probability of retrieving green LiDAR data as a function of lightness estimated from a generalized additive model (GAM). The lightness is calculated from the green channel of the aerial photos taken simultaneously with the LiDAR data and scaled to the range 1-100. The lines represent probability with (blue) and without (red) interpolation. The interpolation is restricted to 1 meter from nearest LiDAR point. The red dots represent the percentage of grid cells with data for every 5 units interval of lightness values. The grey lines at the bottom of each figure represents the lightness range for the 10-90% percentile range of the typical three bottom conditions classified from the aerial photos. The brown lines show the lightness range of dark vegetated areas, the grey lines show the lightness range of light rocky bottom, and the white lines the lightness range of areas with turbulent surface water. The numbers above the lines are the percentage coverage of each area class in the studied river reaches (see Figure 1, Table 1).*

With lightness index values rising above 50-60, the river surface was increasingly dominated by white, turbulent water. This general pattern of the relationship between the probability of retrieving data and the lightness index (Figure 19) remained similar even when other recent orthophotos derived from https://www.norgeibilder.no/ (and not taken simultaneously with the LiDAR recordings) were used to derive the lightness index (Appendix 10, 11).

The importance of point density for retrieving green LiDAR data especially from turbulent river reaches, was reflected in the comparisons of results from both sensors and from grid cells of 25 and 100 cm (Figure 19, Appendix 11). There was a more significant drop in the probability of retrieving green LiDAR signals from the lighter riverbed type for the Teledyne Optech CZMIL compared to Riegl VQ-880 (with higher point density), and especially when the grid size was reduced from 100 to 25 cm.

On a cautionary note, the higher signal probability seen in general for the Riegl VQ-880, is not only an effect of the signal density of the sensor, but also to the adaptive flight pattern for this sensor, with repeated flightlines. The Riegl VQ-880 has approximately 13 times higher point density than the Teledyne Optech CZMIL, but this difference increased to nearly 40 times for the data, due to the adaptive, and thereby overlapping, flight strategy.

The importance of point density is also seen when LiDAR points are interpolated (i.e., restricted to approximately 1 meter from the nearest LiDAR point using a moving window of 3x3 cells for 1 meter grid and 7 x 7 for 25 cm grid). Even with interpolation across such small distances, the probability of having green LiDAR data increases dramatically, and especially for areas with high surface turbulence (Figure 19, Appendix 11). Increasing cell area provides more data points, and spots where the bottom can be 'seen' become more likely.

This association with spatial scale, reflects that 'white water' is not uniformly white due to turbulence, but a micro-mosaic of varying 'whiteness', i.e., degree of surface turbulence. Even apparently 'white water' has small cells with smoother water surface that high-density green LiDAR sensors can hit and penetrate. Data loss reflects a balance between water 'whiteness' and green LiDAR point density. An adaptive flight strategy with several overflights further enhances this 'smooth-surface-hit' probability. With repeated flightlines the angles the green LiDAR pulses hit the water surface will also differ. Moreover, turbulent surface water is dynamic, and new flight might hit a new micro-mosaic providing different smooth-surface hits.

3.5. Robustness of Green LiDAR: Total coverage of green LiDAR for the studied river reaches

The combined effects of water depth, water visibility, riverbed lightness and water surface turbulence are reflected in the total coverage of green LiDAR data for the studied river reaches (Table 7). Loss of data is generally low, primarily depending on the combined green LiDAR pulse power and river water clarity (Secchi depth), and pulse/data density.

Table 7. Loss of green LiDAR data for the studied river reaches.

Sensor	River	Area (ha)	100 cm loss		25 cm loss	
			Direct (%)	Interpolated* (%)	Direct (%)	Interpolated (%)
Riegl VQ-880	Bø	29.7	7.2	5.0	19.6	6.3
Leica Chiroptera 4x	Bø	29.7	16.1	8.3	61.2	12.5
Riegl VQ-880	Hallingdal	82.9	13.4	11.2	22.6	12.4
Teledyne Optech CZMIL	Hallingdal	83.2	11.1	7.5	71.7	9.6
Riegl VQ-880	Lærdal	58.7	0.4	0.1	8.3	0.2
Teledyne Optech CZMIL	Lærdal	58.8	4.1	0.6	66.8	2.4

* *Interpolation to a distance of approx. 1 meter from nearest LiDAR point.*

In Lærdal, with very clear water, coverage was almost complete with a grid size of 100 cm. For the high-point-density Rigel VQ-880, flown with an adaptive flightline strategy, the reduced coverage at 25 cm was relatively small, compared to the lower-point-density Teledyne Optech CZMIL, flown with only small overlap between flightlines. However, the small distance interpolation brought the CZMIL coverage close to the Riegl VQ-880.

In Hallingdal, with water visibility intermediate between Lærdal and Bø, the percentage covered of the river was less than in Lærdal. This is due to lower water visibility, but also to the combination of visibility and more deeper areas (Table 1). Here the Teledyne Optech CZMIL, with higher power pulses, covered the deeper river areas better, and therefore achieved a higher total coverage (Figure 20).

In Bø, The Riegl VQ-880 and Leica Chiroptera 4x showed the same pattern as between Riegl VQ-880 and Teldyne Optech CZMIL in the Lærdal river.

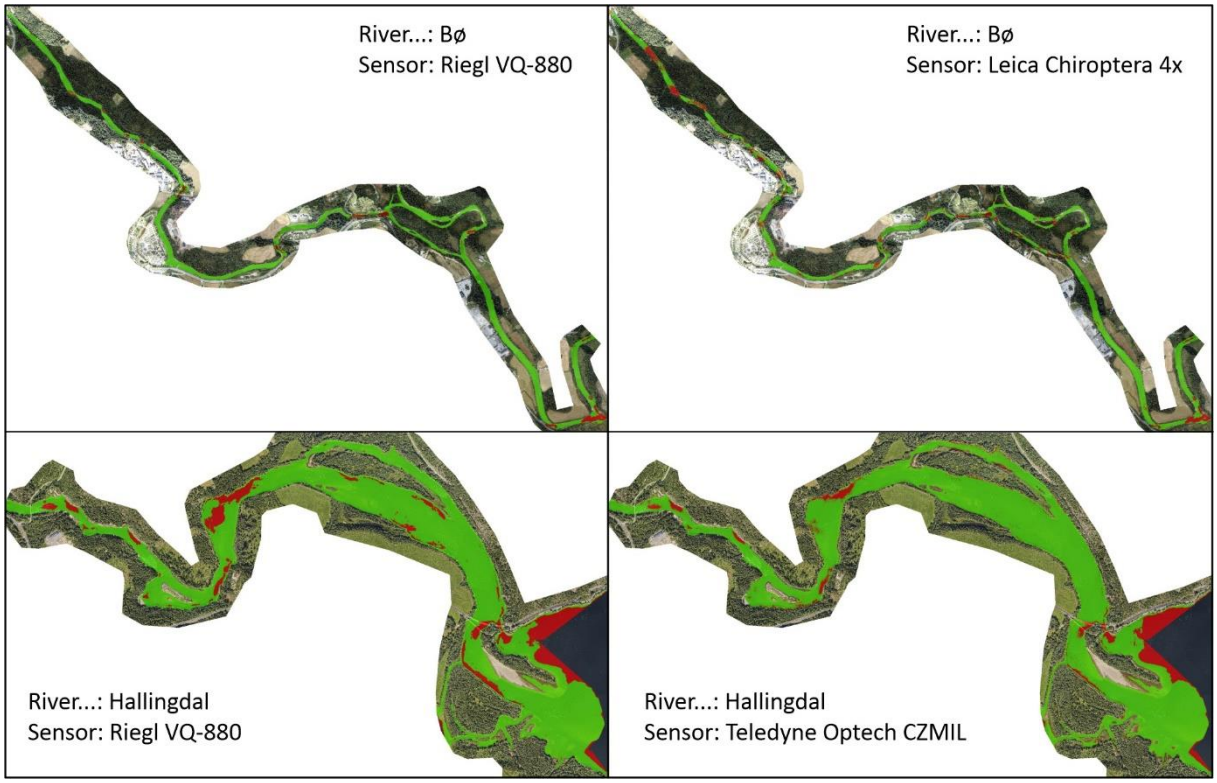


Figure 20. Total coverage for the 1 meter grid size (green) for Bø and Hallingdal rivers.

4. Discussion

The objective of the present study was to explore and evaluate the accuracy and point density (spatial resolution) of remote sensing green laser LiDAR data. To do this, data were collected and compared with *in situ* methods (transect point and Multibeam data, and across three rivers and four sensors. A second objective was to identify which factors may lead to signal loss, and thereby bias or limit remote sensing green LiDAR data acquisition. A stratified design across pre-selected hydraulic variables, primarily water depth, 'white water' associated with river gradient, bottom substrate (size, color), and water clarity/underwater visibility. These variables have been identified as potentially important in previous studies, but based on limited data acquisition from selected cross-sections or one/ few short and uniform, low gradient river reaches, i.e., limited and benign ranges of hydraulic conditions (e.g., Hilldale & Raff 2008; Kinzel, Legleiter & Nelson 2013; Mandlburger *et al.* 2020; Kinzel, Legleiter & Grams 2021; Islam *et al.* 2022). The novelty of the present study is a stratified design including 1) a variety of hydraulic conditions across three different rivers and 2) along long river reaches representing a variety of gradients and habitats, and with 3) data collection by both *in situ* (transect points, Multibeam) and remote sensing green LiDAR data, and 4) the use of recently developed green LiDAR technology with high point densities, with direct comparisons of data from four sensors.

4.1. High accuracy in reflected remote sensing green LiDAR signals

Remote sensing green LiDAR measurements of riverbed elevation were consistently close to comparative *in situ* control measurements by the alternative methods transect point measurements (about 500 points with dGPS in typical river reaches) and Multibeam (about 42 000 measurements in pools). Deviations were generally less than ± 10 cm. It is important to note none of the measurement methods represent the 'truth', meaning that deviations represent the accumulated sampling errors across methods. However, Multibeam data in deep pools may be closest to the 'truth' (e.g., Payne, Eggers & Parkinson 2004; Gard 2005; Jowett & Duncan 2012).

High numbers of outlying deviations were found for the pool data, but without any apparent pattern. These outliers may be due to more noise in the Multibeam data in shallow and sloping areas, and from riverbank vegetation causing noise in the green LiDAR data. It is known that Multibeam data tend to become more 'noisy' in shallower and more sloping riverbed areas due to increased beam angle (e.g., Hughes Clarke, Mayer & Wells 1996; Calder & Mayer 2003). Moreover, some offset measurements were probably due to sensor calibration issues. This should be avoided in the future.

If a green LiDAR signal was reflected, accuracy was consistently high across sensors and water depths. There may perhaps be a weak tendency for more deviations with increasing depths and coarser substrates, but this was somewhat inconsistent in the data. Generally high accuracy in green LiDAR data have also been reported previously, although with some variation (Kinzel, Legleiter & Nelson 2013; Mandlburger *et al.* 2020; Kinzel, Legleiter & Grams 2021; Islam *et al.* 2022).

4.2. Loss of signals with increasing depth depends on water clarity

Loss of remote sensing green LiDAR signals increased with increasing water depth, and abruptly at a 'critical' threshold depth specific to rivers. This critical depth appeared to be closely associated with water clarity, as measured with a Secchi disc. Loss of signals increased abruptly around the measured Secchi depth. This also implies that remote sensing green LiDAR acquisitions should be implemented during stable, low flows, and rising flows avoided.

The importance of water clarity for green LiDAR's water penetrating capabilities have been reported in the few relevant previous studies found in the literature (Mandlburger *et al.* 2020; Kinzel, Legleiter & Grams 2021; Islam *et al.* 2022). Water clarity/turbidity emerge as a main issue for the application of remote sensing green LiDAR in rivers (see Kinzel, Legleiter & Grams 2021 for a thorough discussion). However, reported water penetration capability relative to clarity, as expressed with Secchi depth, varies across studies. In a recent study with a high point-density green LiDAR across three small ponds and a short river reach in the European Alps, Mandlburger *et al.* (2020) reported the penetration depth to be twice the Secchi depth. Similar results were reported by Kinzel, Legleiter

and Grams (2021). Thus, green LiDAR penetration capability appear to be different across rivers, likely depending on what cause loss of clarity; light scatter due to suspended particles and/or light absorption due to dissolved materials. The results here suggest green LiDAR has reduced penetrating capability in humic-rich rivers like two of the rivers in the present study. A penetration depth to around Secchi depth, is to our knowledge the lowest reported.

Interestingly, water penetrating capabilities also varied across sensors. For technical reasons, point density is balanced by signal power for each sensor. Sensors with fewer signals but more power per signal, penetrated water somewhat deeper. In turn, the higher signal strength gives a more homogeneous distribution of reflected points, which via interpolation provided better area coverage with bathymetric data. These are important pre-project considerations. Depending on project objectives and requirements for spatial resolution, point density should be balanced by signal power. More signal power gives deeper water penetration and also more continuous and spatially homogeneous data, which may compensate for reduced point density via post-processing interpolation procedures. For example, for basic bathymetric mapping of riverscapes, coverage may be first priority, and therefore higher signal strength with higher depth penetration more important, likely in combination with interpolation algorithms for increased spatial coverage. If high resolution is first priority, e.g., to capture substrate particle size/roughness for biotic (micro)habitat assessment, higher point density will be important.

4.3. Loss of signals with increasing surface turbidity and dark river bottom

The Riegl VQ-880 and Teledyne Optech sensors provided remote sensing green LiDAR data with simultaneous aerial photos from Lærdal. The more characteristic 'white water' river reaches used for further analyses were polygonised manually.

Analyses using the RGB-derived Lightness index as a proxy for surface turbidity ('white water') in the clear and high-gradient Lærdal, indicated that any loss of green LiDAR signals in 'white water' areas was associated with sensor pulse/data density relative to the spatial scale of the analyses. On a larger 100 x 100 cm grid scale, data loss was limited. However, on a 25 x 25 cm grid scale, there was a major data loss in many cells, especially for the Teledyne Optech sensor with the lower point density (lower point density also due to few overlapping flight lines). This association with spatial scale, reflects that 'white water' is not uniformly white due to turbulence, but a micro-mosaic of varying 'whiteness', i.e., degree of surface turbulence. In apparently 'white water' have small cells with smother water surface that high-density green LiDAR sensors can penetrate. Data loss reflects a balance between water whiteness' together with bottom darkness/vegetation, and green LiDAR point density.

5. Conclusions

1. Deviations in estimated riverbed elevation/water depth across methods were generally small and around 0 (< 10 cm), indicating high accuracy and robustness of the remote sensing green LiDAR.
2. Tested predictors generally had limited predictive power (up to about 40 %), but alternative analyses/models gave consistent results with respect to indicating important predictors of deviation.
3. Deviations were not or only weakly associated with in-river characteristics (Depth, Substrate, Slope).
4. Deviations between remote sensing green LiDAR and *in situ* transect point measurement were perhaps to a small degree positively associated with water depth, resulting in slightly more deviation with increasing depth, but this trend was not consistent. Deviations were weakly positively associated with substrate size, resulting in slightly more deviation with increasing substrate size. Deviations were more strongly associated with Sensor and River, likely caused by sensor calibration issues and unbalanced study design (for logistic reasons not all sensors were tested in all rivers).
5. Deviations between remote sensing green LiDAR and *in situ* Multibeam data in pools, were weakly associated with water Depth and riverbed Slope. The stronger association with Sensor and River, may again be explained by sensor calibration issues and unbalanced study design (above).
6. Remote sensing green LiDAR data loss was strongly associated with Depth, River and Sensor, i.e., models including especially Depth and River had high predictive power (up to 93 %).
7. Green LiDAR data loss was generally small in shallower areas. However, data loss increased abruptly at a critical and river-specific water depth. This critical depth for data loss correlated with the river-specific water clarity, as expressed by Secchi depth.
8. Sensors with stronger signals (at the cost of lower point density) penetrated deeper (up to 1 m deeper across tested sensors).
9. Green LiDAR data loss increased with increasing amount of ‘white water’, i.e., surface turbulence. However, this loss was strongly reduced at higher point densities. There was also data loss associated with dark, vegetated bottom.
10. Green LiDAR data coverage was generally high, but also depended on a balance between pulse power (depth penetration) and pulse density (white water penetration) relative to river depth/water clarity and amount of river white water.

5.1. Further research

A main concern is the loss of green LiDAR signals in deeper river areas, i.e., the water penetrating capability. Apart from signal strength (pulse energy), it varies with water clarity, i.e., depending on light scatter due to suspended particles and/or light absorption due to dissolved materials (above), and the LiDAR point density.

There is a tradeoff between signal strength and LiDAR point density that needs to be studied more in relation to water penetrating capability and the light absorption from turbulent water and dark, vegetated riverbed areas.

Further insight into the effect of water clarity may be gained by combining the river data with similar data for shallow parts of lakes, and thereby increase number and variety of study objects. Including relevant water clarity variables, e.g., Seechi depth, turbidity, as factors in regression models across different types of water, will help to explore how different aspects of water clarity affect the probability of receiving LiDAR data.

The potential ability of high green LiDAR point density to resolve substrate particle sizes/roughness should be explored with respect to point density relative to recording altitude and sensor.

6. Appendices

Appendix 1. Measured wading transect points are for obvious logistic reasons few (green points). In addition, the LiDAR point coverage for riverbed elevation turned out to be uneven for some sensors (orange points), and did not always overlap with transect measurements, reducing the data available for comparative analyses. Here an example from Bø river and Leica Chiroptera 4x sensor.

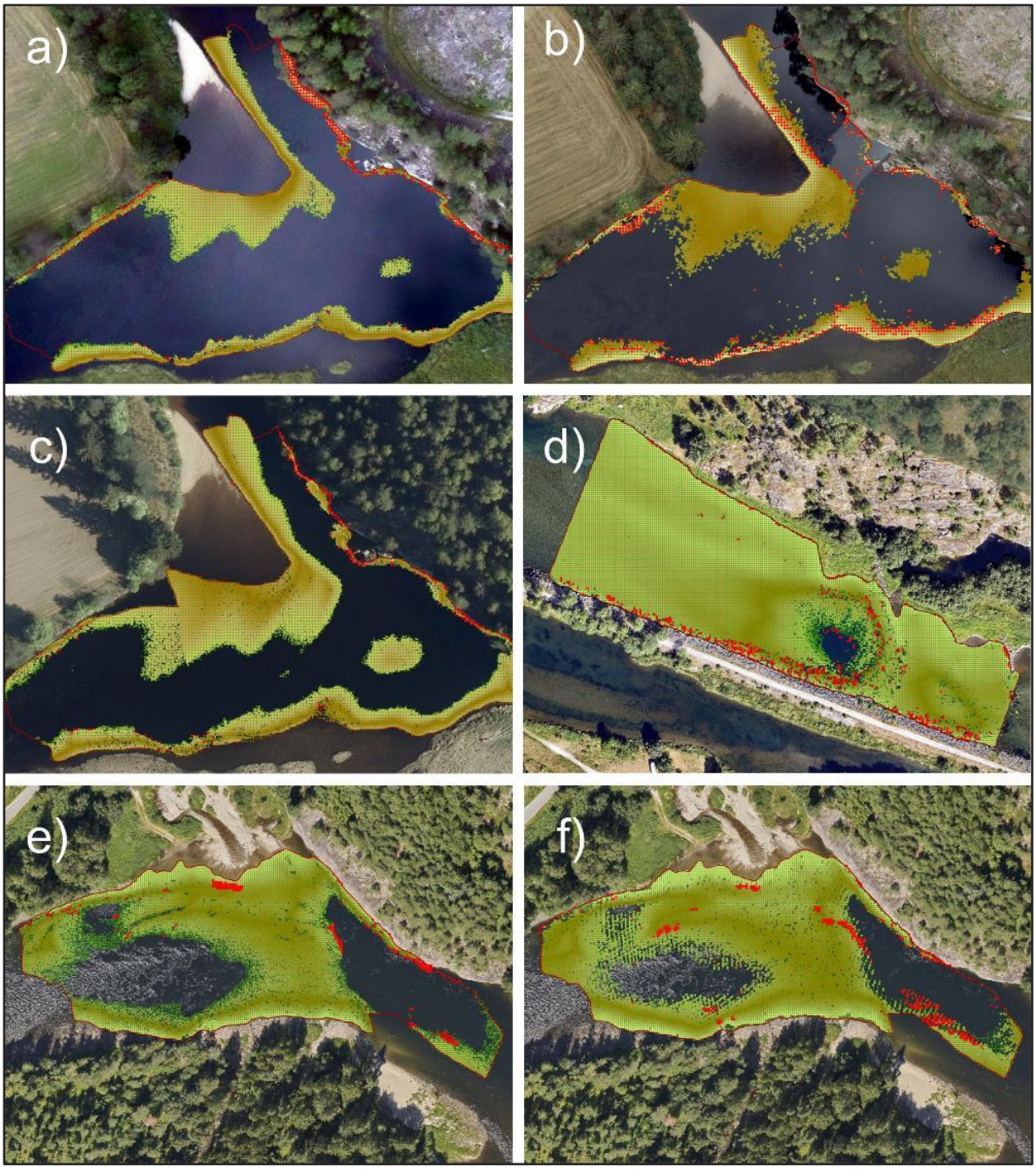
Appendix 2. *Data available[1] for analysis of pool data with 1x1 m grid size. To balance data, the same number of datapoints were drawn randomly from each pond. The number extracted was dictated by the sensor-pool combination with the fewest data points, i.e., 1000 used for analysing deviation in riverbed elevation and 2000 to analyse LiDAR data loss. Riegl VQ-840 helicopter data covered parts of the total pool area, and the percent loss is not directly comparable with the fixed-wing data.*

			Deviation in elevation				Loss of data			
Place	Sensor	All	To analyse[1]	Loss	Selection		To analyse[1]	Loss	Selection	
		N	n	%	n	%	N	%	N	%
Bø	Riegl VQ-880	14 005	4 193	70	3 000	21	12 997	7	5 526	43
Bø	L.Chiroptera 4x	14 005	5 551	60	3 000	21	12 997	7	5 526	43
Bø	Riegl VQ-840	12 094	5 901	51	2 000	17	11 326	6	4 000	35
Halllingdal	Riegl VQ-880	85 849	52 823	38	3 000	3	83 729	2	6 000	7
Halllingdal	Riegl VQ-840	6 789	5 952	12	1 000	15	6 632	2	2 000	30
Halllingdal	T. Optec CZMIL	85 849	66 310	23	3 000	3	83 075	3	6 000	7
Lærdal	Riegl VQ-880	29 492	27 392	7	3 000	10	28 068	5	6 000	21
Lærdal	Riegl VQ-840	4 148	4 034	3	1 000	24	3 875	7	2 000	52
Lærdal	T. Optec CZMIL	29 492	27 343	7	3 000	10	27 730	6	6 000	22
Sum/mean		281 723	199 499	30	22 000	14	270 429	5	43 052	29

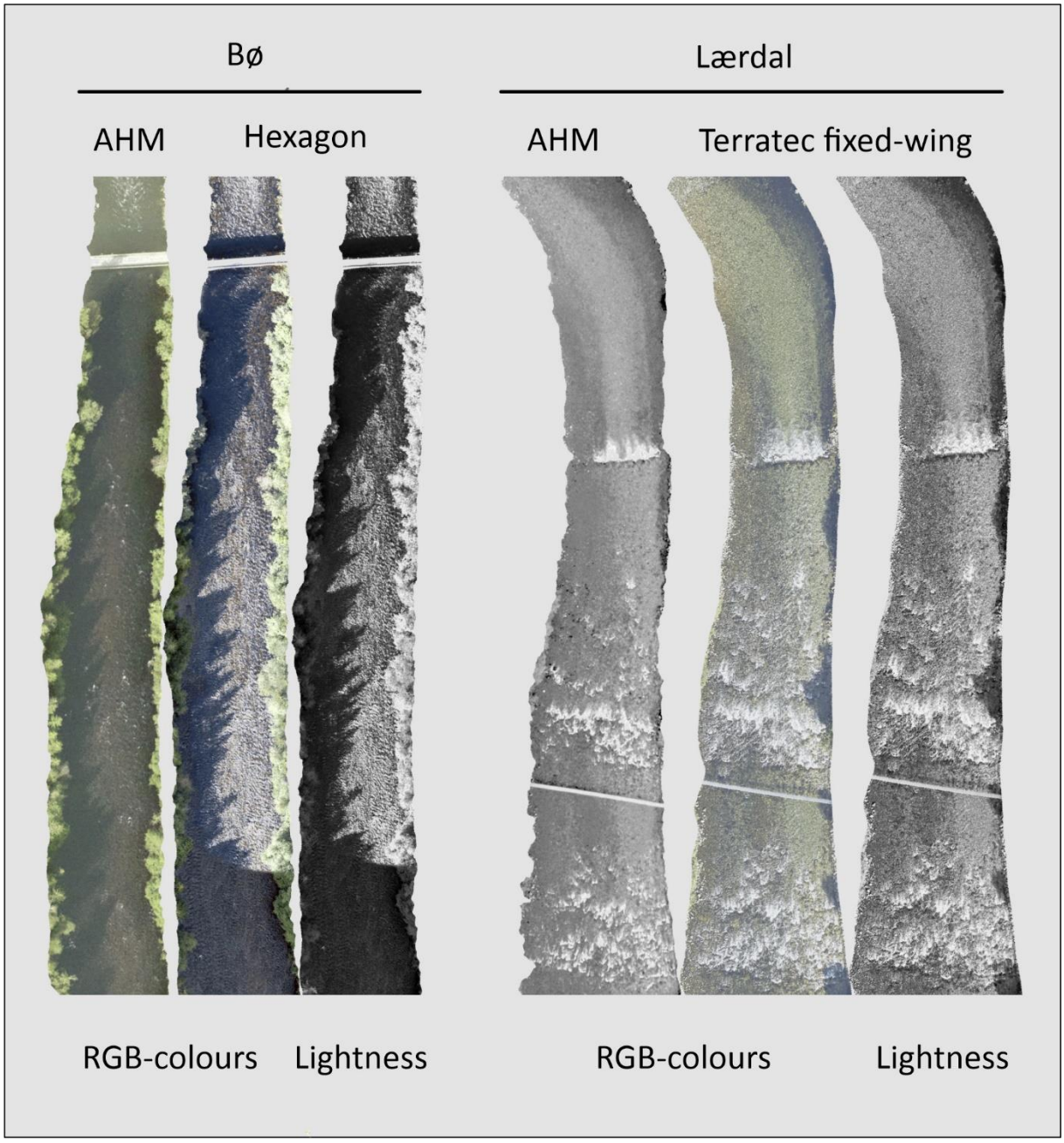
[1] Requirements for including grid cells in the analysis: 1) For deviation in river bed elevation: green LiDAR data available, positive water depth, the deviance in elevation less than 50 cm between Multibeam and LiDAR, grid cells not covered by riverbed vegetation seen in the aerial photos acquired during the lidar recordings. 2) For loss of data: positive water depth and possible to calculate slope from the Multibeam data.

[2] Total number of data points were more than 4 million with grid size 25x25 cm. However, the proportion of grid cells with LiDAR data declined, so the minimum size available after selection increased only slightly.

Appendix 3. *The majority of the outliers for the deviations (in situ control minus LiDAR data) in estimated riverbed elevation (red points) were located close to the riverbanks, to steep slopes, and in areas with low point density. Examples from the deepest pool in a) Bøelva for Riegl VQ-880, b) Leica Chiroptera 4x and c) Teledyne Optech; d) Lærdal for Riegl VQ-880; e) Hallingdal for Riegl VQ-880 and f) Teledyne Optech.*



Appendix 4. *The aerial photos showed extensive light variation across the river areas. When converted to a lightness scale, based on RGB values from aerial photos for an automated identification of turbulent water areas, green riverbank vegetation and light/shadow contrasts were scored similarly as surface white, turbulent water, as indicated by manual control. This lightness score therefore biased data along the forest lined Bø (left three panels) and Hallingdal river reaches. However, it worked well in the more open (agricultural) landscapes along Lærdal (right three panels).*



Appendix 5. *INLA regression parameters for remote sensing green LiDAR accuracy for riverbed elevation data relative to in situ measurements in transect points. The response variable (accuracy) is the deviation in measured riverbed elevation between the two methods, and within a buffer area of 56 cm around the transect points. The intercept expresses the deviation from the reference, i.e., Riegl VQ-880 sensor in Bø. The effects are visualized by river and sensor in Figure 8. *Predictors with Credible interval not including 0 are considered important.*

Fixed effects	**Mean**	**95% Credible Interval**		
Intercept	1.13	-3.03	5.29	
Leica Chiroptera 4x	-15.18	-18.62	-11.74	*
Riegl VQ-840	-3.59	-6.97	-0.21	*
Teledyne Optech CZMIL	-0.24	-36.11	35.62	
Hallingdal	-1.76	-6.30	2.79	
Lærdal	3.27	-0.31	6.84	
Substrate (1-15)	-0.45	-0.82	-0.08	*
Depth (1-120 cm)	0.08	0.04	0.11	*
T.Optech CZMIL: Halllingdal	4.39	-31.48	40.26	
Riegl VQ-840 : Lærdal	3.59	-0.29	7.46	
T.Optech CZMIL: Lærdal	-4.64	-40.51	31.23	
L.Chiroptera4x : Depth	0.06	-0.01	0.13	
Riegl VQ-840 : Depth	-0.04	-0.11	0.03	
T.Optech CZMIL: Depth	-0.14	-0.21	-0.08	*

Random effects (SD)	**Mean**	**95% Credible Interval**		
River	0.53	0.01	3.12	*
Transect	3.98	2.92	5.29	*
Points	0.38	0.01	1.80	*
Sensor	2.82	-0.27	24.45	

Appendix 6. *INLA regression parameters for remote sensing green LiDAR accuracy for riverbed elevation data relative to in situ measurements in pools by Multibeam. The accuracy is measured as the deviation of the river bottom elevation when estimated for grid cells of 1 x 1 meter with the mean point value from Multibeam and green LiDAR. *Predictors with Credible interval not including 0 are considered important.*

Fixed effects	**Mean**	**95% Credible Interval**		
Intercept	11.60	6.77	16.37	*
Leica Chiroptera 4x	-13.59	-15.10	-12.08	*
Riegl VQ-840	0.08	-0.87	1.02	
Teledyne Optech CZMIL	-5.68	-41.49	30.13	
Hallingdal	2.16	-4.54	8.90	
Lærdal	-13.26	-19.91	-6.50	*
Depth	-0.04	-0.05	-0.04	*
Slope	-0.18	-0.20	-0.16	*
Riegl VQ-840 : Halllingdal	-8.21	-9.20	-7.22	*
T.Optech CZMIL: Halllingdal	-1.92	-37.73	33.89	
Riegl VQ-840 : Lærdal	-2.54	-3.15	-1.94	*
T.Optech CZMIL : Lærdal	-3.77	-39.57	32.04	
L.Chiroptera4x : Depth	0.04	0.03	0.05	*
Riegl VQ-840 : Depth	0.01	0.01	0.02	*
T.Optech CZMIL: Depth	*0.00*	*0.00*	*0.00*	
Hallingdal : Depth	*0.04*	*0.03*	*0.04*	*
Lærdal : Depth	*0.05*	*0.05*	*0.06*	*

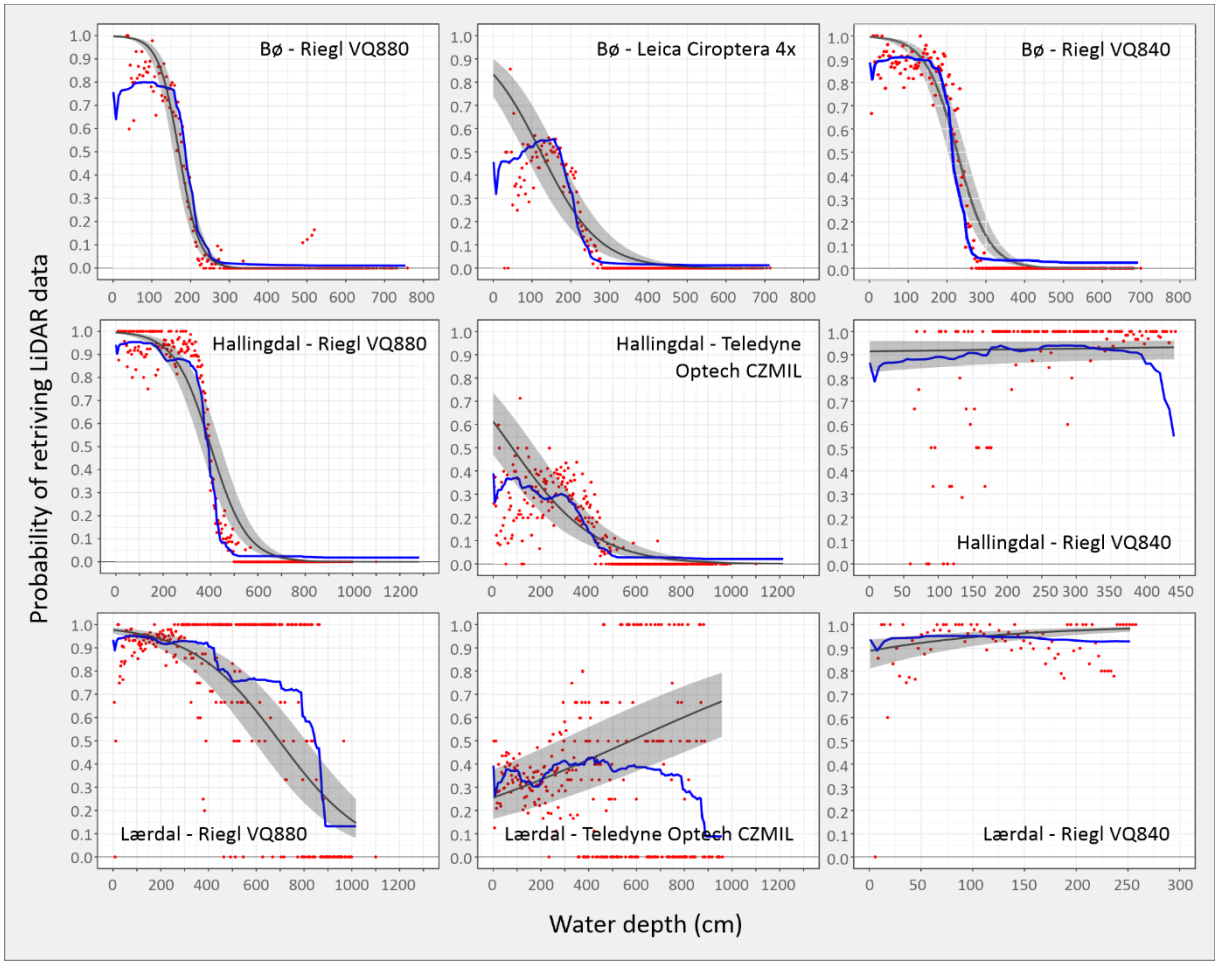
Random effects (SD)	**Mean**	**95% Credible Interval**		
River	0.19	0.03	0.66	*
Pool	4.26	2.16	8.34	*
Grid cell	0.16	0.04	0.44	*
Sensor	0.12	0.05	0.23	*

Appendix 7. *INLA regression parameters for a binary loss model in pools using grid cell size 100 x 100 cm. The numbers in brackets are the coefficients for a model with scaled values for the covariates.*

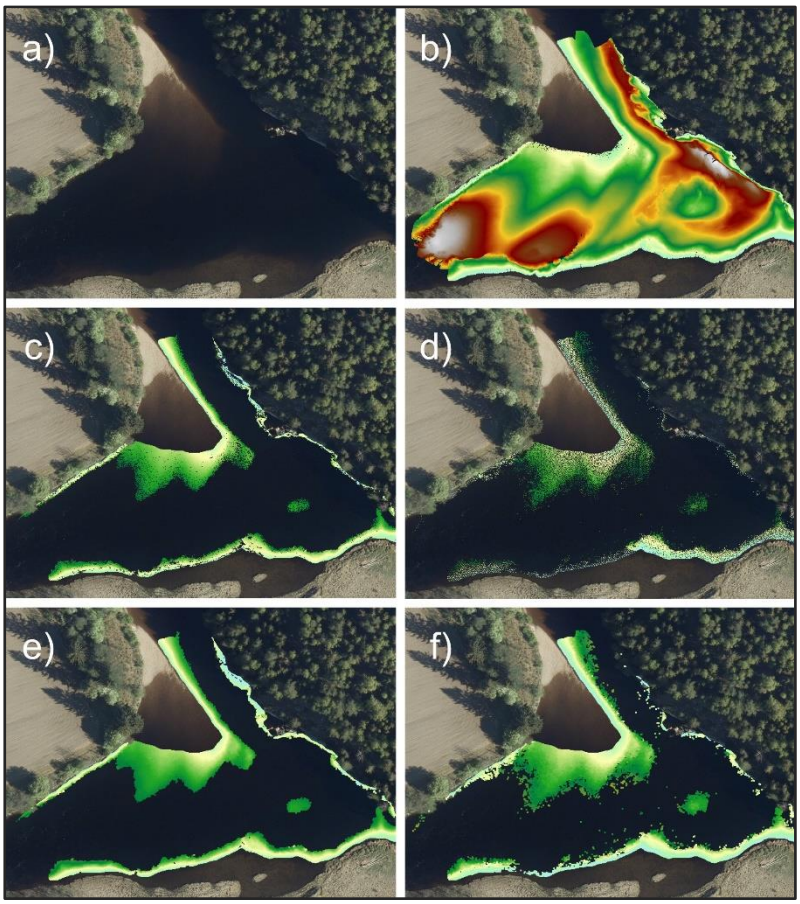
Fixed effects		**Mean**	**95% Credible Interval**		
Intercept		13.42	12.21	14.65	
Leica Chiroptera 4x		-3.51	-4.45	-2.57	*
Riegl VQ-840		1.02	-0.20	2.27	
Teledyne Optech CZMIL		-3.92	-39.74	31.89	
Hallingdal		-0.89	-2.52	0.73	
Lærdal		-0.38	-2.50	1.83	
Depth (1-1300 cm)	*(-10.2)*	-0.06	-0.07	-0.06	*
Slope (0-70 °)	(0.4)	-0.05	-0.05	-0.04	*
Riegl VQ-840 :Halllingdal		-1.23	-2.36	-0.03	*
T.Optech CZMIL:Halllingdal		-0.20	-36.01	35.6	
Riegl VQ-840 : Lærdal		-6.64	-8.62	-4.46	*
T.Optech CZMIL: Lærdal		-3.72	-39.53	32.1	
L.Chiroptera4x : Depth		0.03	0.02	0.03	*
Riegl VQ-840 : Depth		0.01	0.00	0.01	*
T.Optech CZMIL: Depth		0.01	0.01	0.01	*
Hallingdal : Depth		0.04	0.03	0.04	*
Lærdal : Depth		0.05	0.05	0.06	*

Random effects (SD)	**Mean**	**95% Credible Interval**		
River	0.04	0.02	0.07	*
Pool	0.67	0.36	1.02	*
Grid cell	0.11	0.02	0.38	*
Sensor	0.09	0.06	0.12	*

Appendix 8. *The probability of retrieving green LiDAR data as a function of water depth calculated from binary response models and grid of 25 x 25 cm (n = 44 000). The slope was kept constant at the mean for each river. The black line with Credible Interval indicates the INLA model results, and the blue line the Gradient Boosting Machine model results. The red dots are the proportions of LiDAR data received within groups of 4 cm depth for each river and sensor combination.*

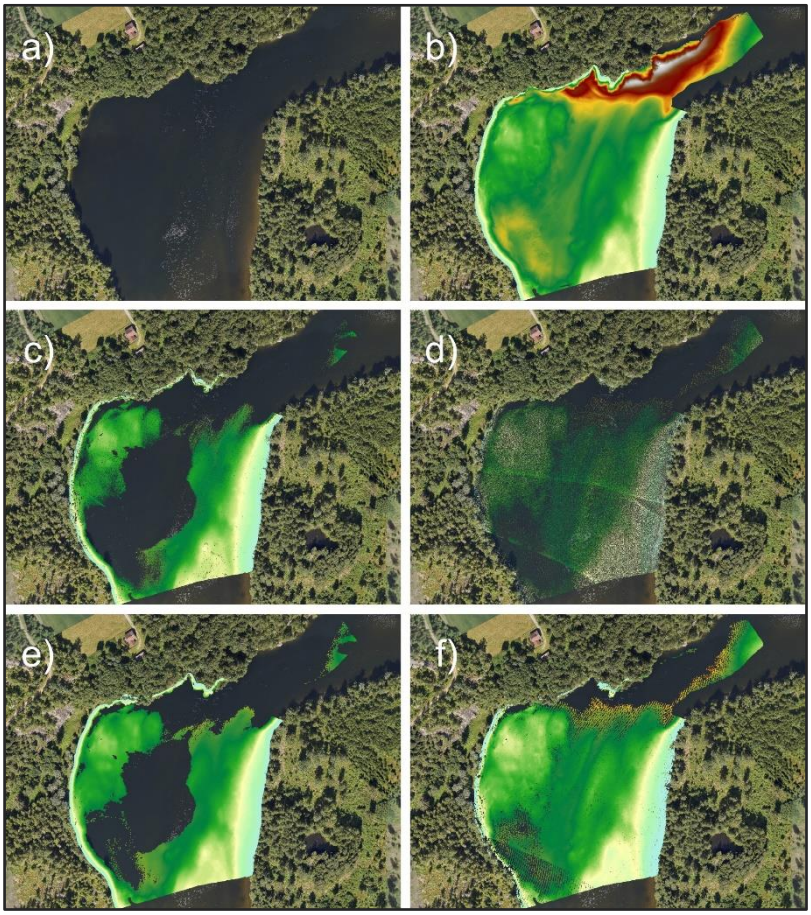


Appendix 9. *Bathymetric models.*



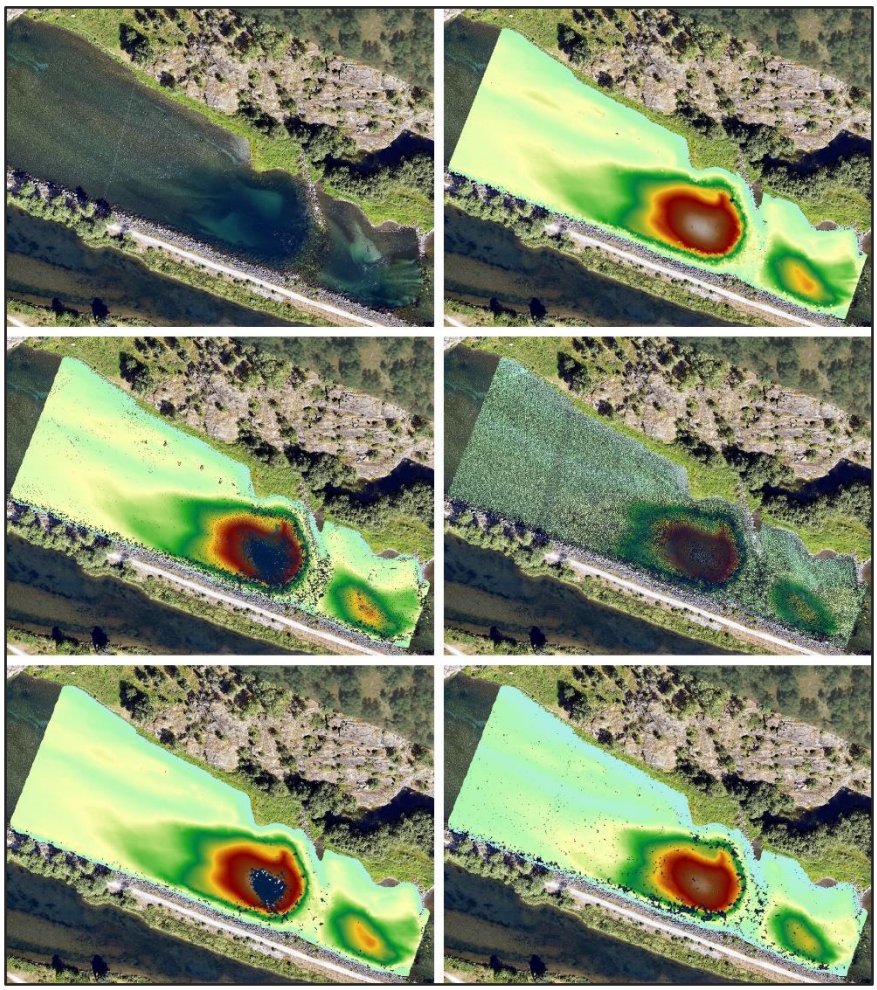
Bathymetric model for the deepest pool in Bø river estimated without and with interpolation of Multibeam/LiDAR to a grid resolution of 100 cm and 25 cm.

a) *Terratec Aerial photo 2019*
b) *Multibeam 25cm median*
c) *Riegl VQ-880 25cm median*
d) *Leica Chiroptera 4x 25cm median*
e) *Riegl VQ-880 25cm Interpolation*
f) *Leica Chiroptera 4x 25cm Interpolation*



Bathymetric model for the deepest pool in Hallingdal river (studied reach) estimated through interpolation of Multibeam/LiDAR points to a grid resolution of 25 cm.

a) Aerial photo from Terratec
b) Multibeam 25cm median
c) Riegl VQ-880 25cm median
d) Teledyne Optech CZMIL 25cm median
e) Riegl VQ-880 25cm Interpolation
f) Teledyne Optech CZMIL 25cm Interpolation



Bathymetric model for the deepest pool in Lærdal river (studied reaches) estimated through interpolation of the Multibeam/LiDAR points to a grid resolution of 25 cm.

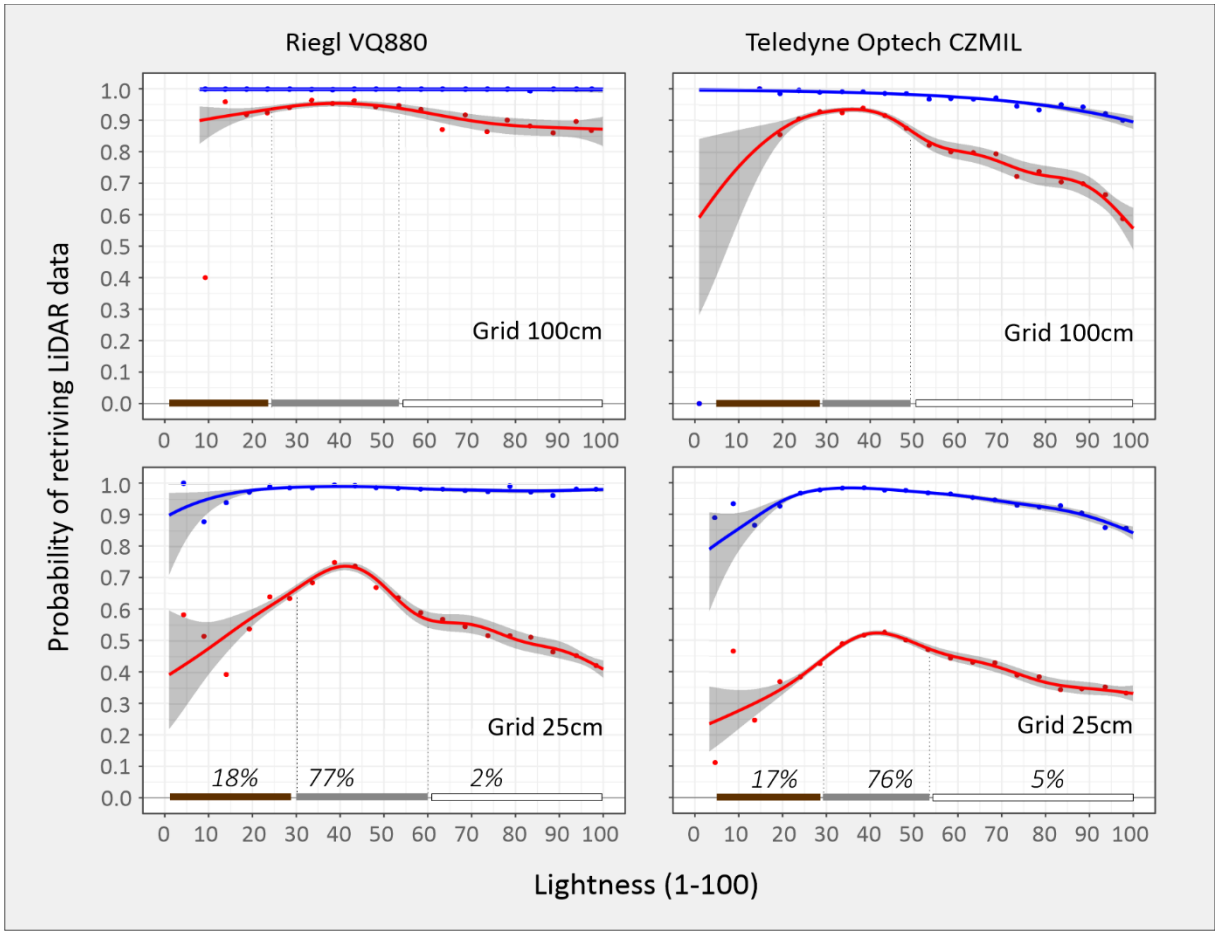
a) *Lærdal Terratec Aerial photo 2021*
b) *Lærdal Multibeam 25cm Median*
c) *Lærdal VQ-880 25cm Median*
d) *Lærdal Teledyne Optech CZMIL 25cm Median*
e) *Lærdal VQ-880 25cm Interpolation*
f) *Lærdal Teledyne Optech CZMIL 25cm Interpolation*

Appendix 10. *Percentile values for the lightness index of the three thematic classes used to map the Lærdal river. The "rgb" source represent the aerial photos recorded simultaneously with the LiDAR data, and "nib" the most recent photos from the archive norgeibilder.no.*

Classes	Source	Size	Camera	Q10	Q50	Q90
LightBottom	rgb	100	AHM	**26**	32	**39**
WhiteSurface	rgb	100	AHM	36	50	76
DarkBottum	rgb	100	AHM	20	24	31
LightBottom	rgb	100	TTF	**28**	40	**52**
WhiteSurface	rgb	100	TTF	36	58	82
DarkBottum	rgb	100	TTF	15	26	42
LightBottom	rgb	25	AHM	**28**	34	**41**
WhiteSurface	rgb	25	AHM	40	53	80
DarkBottum	rgb	25	AHM	23	26	32
LightBottom	rgb	25	TTF	**28**	41	**59**
WhiteSurface	rgb	25	TTF	37	64	90
DarkBottum	rgb	25	TTF	17	23	40

Classes	Source	Size	Camera	Q10	Q50	Q90
LightBottom	nib	100	AHM	**24**	34	**53**
WhiteSurface	nib	100	AHM	25	45	89
DarkBottum	nib	100	AHM	18	26	46
LightBottom	nib	100	TTF	**29**	37	**49**
WhiteSurface	nib	100	TTF	29	46	86
DarkBottum	nib	100	TTF	23	29	44
LightBottom	nib	25	AHM	**30**	39	**60**
WhiteSurface	nib	25	AHM	29	49	96
DarkBottum	nib	25	AHM	24	32	55
LightBottom	nib	25	TTF	**29**	39	**53**
WhiteSurface	nib	25	TTF	28	46	90
DarkBottum	nib	25	TTF	24	30	47

Appendix 11. *Probability of retrieving green LiDAR data as a function of lightness estimated from a generalized additive model (GAM). The lightness is calculated from the green channel of the latest aerial photos existing at the Norge-i-bilder.no database and scaled to the range 1-100. The lines represent probability with (blue) and without (red) interpolation. The interpolation is restricted to 1 meter from nearest LiDAR point. The grey lines at the bottom of each figure represents the lightness range for the 10-90% percentile range of the three typical bottom conditions classified from the aerial photos* (appendix 10). *The brown lines show the lightness range of dark vegetated areas, the grey lines the lightness range of light rocky bottom, and the white lines the lightness range of areas with turbulent surface water. The numbers above the lines are the percentage of each class in the studied river reaches (see Figure 1, Table 1).*



7. References

Akaike, H. (1974) A new look at the statistical model identification. *IEEE transactions on automatic control,* **19,** 716-723.

Ando, T. (2007) Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models. *Biometrika,* **94,** 443-458.

Apley, D.W. & Zhu, J. (2020) Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* **82,** 1059-1086.

Awadallah, M.O.M. (2021) Comparison between the Green and the Red LiDAR terrain models in flood inundation estimations. NTNU.

Belgiu, M. & Drăguţ, L. (2016) Random forest in remote sensing: A review of applications and future directions. *ISPRS journal of photogrammetry and remote sensing,* **114,** 24-31.

Biau, G. & Scornet, E. (2016) A random forest guided tour. *Test,* **25,** 197-227.

Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H. & White, J.-S.S. (2009) Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution,* **24,** 127-135.

Breiman, L. (2001) Random forests. *Machine learning,* **45,** 5-32.

Brooks, M.E., Kristensen, K., Van Benthem, K.J., Magnusson, A., Berg, C.W., Nielsen, A., Skaug, H.J., Machler, M. & Bolker, B.M. (2017) glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R journal,* **9,** 378-400.

Bürkner, P.-C. (2017a) Advanced Bayesian multilevel modeling with the R package brms. *arXiv preprint arXiv:1705.11123*.

Bürkner, P.-C. (2017b) brms: An R package for Bayesian multilevel models using Stan. *Journal of statistical software,* **80,** 1-28.

Calder, B.R. & Mayer, L.A. (2003) Automatic processing of high‐rate, high‐density multibeam echosounder data. *Geochemistry, Geophysics, Geosystems,* **4**.

Danesh, T., Ouaret, R., Floquet, P. & Negny, S. (2022) Interpretability of neural networks predictions using Accumulated Local Effects as a model-agnostic method. *Computer Aided Chemical Engineering*, pp. 1501-1506. Elsevier.

Friedman, J.H. (2001) Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.

Gard, M. (2005) Variability in flow–habitat relationships as a function of transect number for PHABSIM modelling. *River Research and Applications,* **21,** 1013-1019.

Harrison, X.A., Donaldson, L., Correa-Cano, M.E., Evans, J., Fisher, D.N., Goodwin, C.E., Robinson, B.S., Hodgson, D.J. & Inger, R. (2018) A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ,* **6,** e4794.

Hauer, C. & Pulg, U. (2018) The non-fluvial nature of Western Norwegian rivers and the implications for channel patterns and sediment composition. *Catena,* **171,** 83-98.

Heggenes, J. (2002) Flexible summer habitat selection by wild, allopatric brown trout in lotic environments. *Transactions of the American Fisheries Society,* **131,** 287-298.

Hilldale, R.C. & Raff, D. (2008) Assessing the ability of airborne LiDAR to map river bathymetry. *Earth Surface Processes and Landforms,* **33,** 773-783.

Hughes Clarke, J.E., Mayer, L.A. & Wells, D.E. (1996) Shallow-water imaging multibeam sonars: A new tool for investigating seafloor processes in the coastal zone and on the continental shelf. *Marine Geophysical Researches,* **18,** 607-629.

Hurlbert, S.H. (1984) Pseudoreplication and the design of ecological field experiments. *Ecological Monographs,* **54,** 187-211.

Hurlbert, S.H. (2009) The ancient black art and transdisciplinary extent of pseudoreplication. *Journal of Comparative Psychology,* **123,** 434-443.

Islam, M.T., Yoshida, K., Nishiyama, S., Sakai, K. & Tsuda, T. (2022) Characterizing vegetated rivers using novel unmanned aerial vehicle‐borne topo‐bathymetric green lidar: Seasonal applications and challenges. *River Research and Applications,* **38,** 44-58.

Jowett, I.G. & Duncan, M.J. (2012) Effectiveness of 1D and 2D hydraulic models for instream habitat analysis in a braided river. *Ecological Engineering,* **48,** 92-100.

Kinzel, P.J., Legleiter, C.J. & Grams, P.E. (2021) Field evaluation of a compact, polarizing topo‐bathymetric lidar across a range of river conditions. *River Research and Applications,* **37,** 531-543.

Kinzel, P.J., Legleiter, C.J. & Nelson, J.M. (2013) Mapping river bathymetry with a small footprint green LiDAR: applications and challenges 1. *JAWRA Journal of the American Water Resources Association,* **49,** 183-204.

Kursa, M.B. & Rudnicki, W.R. (2010) Feature selection with the Boruta package. *Journal of statistical software,* **36,** 1-13.

Lindgren, F. & Rue, H. (2015) Bayesian spatial modelling with R-INLA. *Journal of statistical software,* **63,** 1-25.

Maity, A.K., Basu, S. & Ghosh, S. (2021) Bayesian criterion‐based variable selection. *Journal of the Royal Statistical Society: Series C (Applied Statistics),* **70,** 835-857.

Mandlburger, G. (2018) A review of airborne laser bathymetry sensors. *New Challenges in Hydraulic Research and Engineering*, pp. 41-42.

Mandlburger, G., Pfennigbauer, M., Schwarz, R., Flöry, S. & Nussbaumer, L. (2020) Concept and performance evaluation of a novel UAV-borne topo-bathymetric LiDAR sensor. *Remote Sensing,* **12,** 986.

Natekin, A. & Knoll, A. (2013) Gradient boosting machines, a tutorial. *Frontiers in neurorobotics,* **7,** 21.

Niekerk, J.v., Bakka, H. & Rue, H. (2021) Competing risks joint models using R-INLA. *Statistical Modelling,* **21,** 56-71.

Payne, T.R., Eggers, S.D. & Parkinson, D.B. (2004) The number of transects required to compute a robust PHABSIM habitat index. *Hydroécologie appliquée,* **14,** 27-53.

Pe'eri, S., Morgan, L.V., Philpot, W.D. & Armstrong, A.A. (2011) Land-water interface resolved from airborne LiDAR bathymetry (ALB) waveforms. *Journal of Coastal Research*, 75-85.

Pe'Eri, S. & Philpot, W. (2007) Increasing the existence of very shallow-water LIDAR measurements using the red-channel waveforms. *IEEE Transactions on Geoscience and Remote Sensing,* **45,** 1217-1223.

Pedersen, E.J., Miller, D.L., Simpson, G.L. & Ross, N. (2019) Hierarchical generalized additive models in ecology: an introduction with mgcv. *PeerJ,* **7,** e6876.

Rue, H., Riebler, A., Sørbye, S.H., Illian, J.B., Simpson, D.P. & Lindgren, F.K. (2016) Bayesian computing with INLA: a review. *arXiv preprint arXiv:1604.00860*.

Schielzeth, H. (2010) Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution,* **1,** 103-113.

Silk, M.J., Harrison, X.A. & Hodgson, D.J. (2020) Perils and pitfalls of mixed-effects regression models in biology. *PeerJ,* **8,** e9522.

Smart, G., Bind, J. & Duncan, M. (2009) River bathymetry from conventional LiDAR using water surface returns. *18th World IMACS/MODSIM Congress*.

Speiser, J.L., Miller, M.E., Tooze, J. & Ip, E. (2019) A comparison of random forest variable selection methods for classification prediction modeling. *Expert systems with applications,* **134,** 93-101.

Touzani, S., Granderson, J. & Fernandes, S. (2018) Gradient boosting machine for modeling the energy consumption of commercial buildings. *Energy and Buildings,* **158,** 1533-1543.

Ward, E.J. (2008) A review and comparison of four commonly used Bayesian and maximum likelihood model selection tools. *Ecological Modelling,* **211,** 1-10.

Wood, S.N. (2006) *Generalized additive models: an introduction with R*. chapman and hall/CRC.

Publication series No. 112
2023

—

Evaluating in-riverscapes: Remote sensing green LIDAR efficiently provides accurate high-resolution bathymetric maps, but is limited by water penetration

—

Leif Kastdalen
Jan Heggenes

—

ISBN: 978-82-7206-744-0
ISSN: 2535-5325

usn.no