**JNCC Report**
**No. 660**

# A synthesis of best practice to ensure and evaluate the quality of LiDAR data

**Lightfoot, P. & Heeley, L.**

**July 2020**



ISSN 0963-8091

In partnership with the Scottish Government

**For further information please contact:**

Joint Nature Conservation Committee
Monkstone House
City Road
Peterborough  PE1 1JY

www.jncc.gov.uk

**This report should be cited as:**

Lightfoot, P. & Heeley, L. (2020) A synthesis of best practice to ensure and evaluate the quality of LiDAR data, JNCC Report No. 660, JNCC, Peterborough, ISSN 0963-8091.

**JNCC EQA Statement:**

This report is compliant with the JNCC Evidence Quality Assurance Policy
https://jncc.gov.uk/about-jncc/corporate-information/evidence-quality-assurance/

**Acknowledgements:**

This research was carried out with financial support from the Scottish Government.  We would like to thank our colleagues at JNCC and Scottish Government (Digital Directorate) for their support and input into the project, and to express our particular gratitude to staff at the Scottish Environment Protection Agency, Historic Environment Scotland, Natural Resources Wales and the Environment Agency for providing expert advice.

**Please Note:**

This report was previously published (July 2020) as “Review of Protocols for Quality Control of LiDAR Data”.

The content of the report has not been amended.

## Summary

The Scottish Remote Sensing Portal was developed by the Scottish Government and JNCC to provide open access to public sector LiDAR data. LiDAR (Light Detection and Ranging) uses lasers to create detailed 3D models of the earth's surface. As part of this collaboration, JNCC carried out an investigation into standard methods for ensuring and evaluating the quality of airborne LiDAR data. The investigation consisted of a desk-based literature review and consultation with experts at the Scottish Environment Protection Agency (SEPA), Historic Environment Scotland (HES), the Environment Agency (EA) and Natural Resources Wales (NRW).

The protocols for ensuring LiDAR data are fit for their intended purpose fall into three categories. The client must provide an appropriate and clear **technical specification** detailing their requirements. The contractor will carry out **quality assurance** activities before a survey, such as flight planning and instrument calibration, to ensure the data collected will meet the users' needs. After the survey, **quality control** is carried out to evaluate the quality of the LiDAR point cloud and derived surface models to ensure they meet the specifications. Some quality control activities are conducted by the contractor as part of their data processing workflow, while others are conducted by the client on receipt of deliverables.

This report outlines the factors that should be included in a technical specification and summarises standard protocols for the quality assurance and quality control of LiDAR data. It provides a recommended set of QC activities to be carried out by a client on receipt of LiDAR data from a contractor. Scripts are provided in the appendices to facilitate efficient data manipulation as part of the client's QC activities.

# Contents

# 1 Introduction

## 1.1 Scottish Remote Sensing Portal

In 2016 and 2017, Scottish Government and JNCC collaborated to develop an online data portal to host public sector LiDAR datasets. The collaboration continued between 2018 and 2020, supporting the ongoing running of the portal and delivering enhancements based on user feedback.

The Scottish Remote Sensing Portal can be accessed at:
https://remotesensingdata.gov.scot

**Figure 1.** Scottish Remote Sensing Portal home page.

The Scottish Remote Sensing Portal currently provides access to three sets of Scottish public sector airborne LiDAR data (Table 1), which can be downloaded as 10km tiles or accessed via Web Map Services (WMS). All data are shared under the Open Government Licence v3 unless otherwise stated. Information on how to contribute LiDAR or other remote sensing datasets to the portal can be found here:
https://remotesensingdata.gov.scot/contribute

**Table 1.** Summary of LiDAR datasets available through the Scottish Remote Sensing Portal.

| | Collection dates | Spatial coverage | Commissioned/ Purchased by | Collected by |
|---|---|---|---|---|
| Phase 1 | March 2011 to May 2012 | 11,845km$^2$ (10 sites) | Scottish Government, SEPA and Scottish Water | Blom on behalf of Atkins |
| Phase 2 | November 2012 to April 2014 | 3,516km$^2$ (66 sites) | Scottish Government, SEPA, SportsScotland and 13 Scottish Local Authorities | Fugro |
| Phase 3(i) | 2015 to 2016 | 11,772km$^2$ | Scottish Government (Digital Directorate) | Fugro |
| Phase 3(ii) | 2019 | | Scottish Government (Digital Directorate) | |

## 1.2 LiDAR data

LiDAR (Light Detection and Ranging) is a form of active remote sensing in which lasers are used to create detailed 3-dimensional models of the earth's topography, surface structures and vegetation. LiDAR instruments emit laser pulses and record the speed and directionality of the reflected light. This information is combined with the sensor's position and orientation, derived from its inbuilt GNSS (Global Navigation Satellite System) receiver and INS (Inertial Navigation System), to calculate the XYZ coordinates of the point from which the pulse was reflected. By emitting millions of these pulses, the LiDAR instrument generates a 'point cloud' representing the 3D structure of the surveyed area. In vegetated areas, LiDAR sensors detect the 'first return' reflected from the canopy as well as the 'last return' reflected from the ground.

Some LiDAR systems can capture the full waveform (time vs intensity) plot of the laser pulse, rather than simply capturing the discrete returns when the pulse interacts with a target such as the ground or vegetation canopy. Discrete returns are more commonly used, but the additional information provided by full waveform returns has advantages for certain applications, for example forestry or bathymetry (Fernandez-Diaz *et al*. 2014).

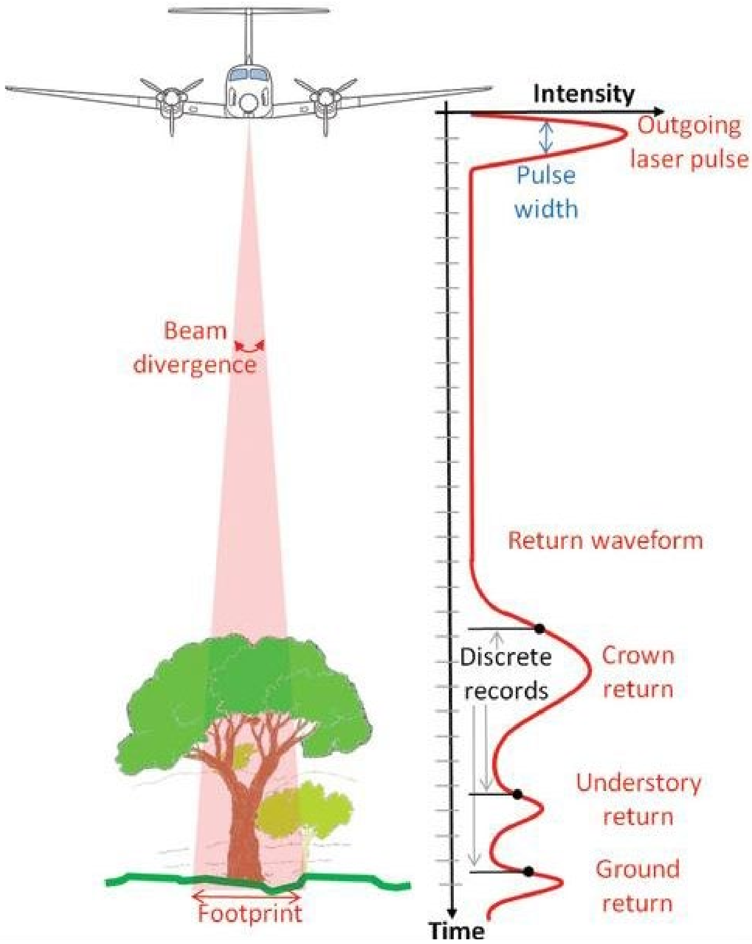


**Figure 2.** Illustration of full waveform and discrete LiDAR returns from a vegetated area. Image from Fernandez-Diaz *et al*. (2014).

LiDAR systems for terrestrial surveys typically use 1064nm near-infrared (NIR) lasers (Hopkinson *et al.* 2016). Shallow water bathymetric surveys can be carried out using LiDAR systems equipped with both NIR and green lasers, as green wavelengths can penetrate the water surface (Saylam *et al.* 2018).

LiDAR instruments are most commonly mounted on fixed wing aircraft, enabling extensive spatial coverage in a single survey, but can also be mounted on terrestrial platforms (Elsherif *et al.* 2019; Stovall *et al*. 2019) and more recently on drones *(Kellner et al. 2019; Resop et al.* 2019). Terrestrial and drone-mounted LiDAR can produce higher resolution data than airborne LiDAR, i.e. a greater number of pulses per $m^2$, but they are limited to a far smaller survey area. This report deals exclusively with assessing and ensuring the quality of airborne LiDAR; in the following sections the term ‘LiDAR’ refers to airborne LiDAR.
The accurate and detailed datasets produced by LiDAR technology support many environmental applications, including flood risk assessment and mitigation, forest monitoring, river surveys, shoreline mapping and coastal defence planning. LiDAR data can also be integrated with other types of earth observation dataset, such as freely available Sentinel-1 and Sentinel-2 data from the European Union’s Copernicus programme, to enhance accuracy of derived products (Fragoso-Campón *et al.* 2018; Sánchez Sánchez *et al.* 2018; Demir *et al*. 2019).

## 1.3 Quality Assurance and Control

It is essential that the quality of LiDAR datasets is ensured and evaluated. End-users rarely have access to the raw data gathered by the LiDAR instrument, and therefore rely on quality measures to evaluate whether the output datasets are fit for their intended purpose. In particular, users need insight into the resolution (point density) and the horizontal and vertical accuracy of the supplied data.

Quality Assurance (QA) refers to the steps taken before a survey, such as flight planning, site visits and instrument calibration, to ensure the data collected will meet the users’ needs (Habib & Van Rens 2007). Quality Control (QC) refers to the steps taken after the survey to evaluate the quality of the point cloud and derived raster products, namely digital surface models (DSM) and digital terrain models (DTM), to ensure they meet set specifications.

A number of QA and QC protocols and data specifications have been published and adopted by different organisations for operational use, but to date there is no consensus within the LiDAR community on a single agreed set of standards (Lohani *et al*. 2018; ASPRS 2018b). The development of standardised quantitative QC techniques could benefit both data providers and end users, potentially increasing productivity and efficiency as well as delivering better products (Habib 2014).

## 1.4 Project remit

The remit of this project was to investigate standard methods used to quality control LiDAR datasets and produce recommendations for the QC activities to be carried out by a client on receipt of LiDAR data from a contractor. It was not within the project remit to make recommendations on technical specifications for LiDAR surveys or on the QA and QC activities to be carried out by the contractor. However, an overview of these topics has been provided for completeness and context, and because of their influence on the client-side QC activities.

The investigation consisted of a desk-based literature review and consultation with relevant experts at Scottish Environment Protection Agency (SEPA), the Environment Agency (EA), Natural Resources Wales (NRW) and Historic Environment Scotland (HES).

# 2 LiDAR Survey Specifications

The client must provide potential contractors with an appropriate and clear set of technical specifications when issuing an invitation to tender for a LiDAR survey. This will greatly facilitate QA and QC processes and ensure data are fit for purpose.

Good examples of technical specifications are available from organisations who commission LiDAR surveys, such as the recent National LiDAR Survey for Wales (Chapman 2019) or the general technical specifications used as a minimum standard in all contracts let by the Environment Agency (Environment Agency 2013), the Regional Coastal Monitoring Programmes of England (Environment Agency 2015) or the US Government Geological Survey (Heidemann 2018).

This report does not attempt to recommend a specific set of standards to adopt, but the following sections list the factors which should be considered and included when writing a technical specification for a LiDAR survey.

## 2.1 Data Acquisition Requirements

Data acquisition requirements should be specified as clearly as possible, as these will directly inform the QA processes carried out by the contractor when planning the survey.

### 2.1.1 Survey extent

The spatial extent of the survey should be provided to the contractor as a set of coordinates, a GIS file, or a named boundary, e.g. of a country, county or designated protected area. Large survey areas may be split into zones if they have different data capture requirements, e.g. coastal zones may need to be flown at low tide.

### 2.1.2 Survey timing

The client should specify a date range for data acquisition. Suitable dates will depend on the purpose of data collection, for example the client may wish to capture data on vegetated areas in leaf-on or leaf-off conditions. Survey timing may also be specified, for example the client may wish to capture data over a coastal zone in a single low spring tide. Data collection during day or night is usually acceptable and increases opportunities for data collection under suitable conditions.

The client may specify conditions for data acquisition that will influence survey timing within the date range, for example that data should not be collected during precipitation or fog, or while there is extensive flooding or snow cover. To ensure positional accuracy, the EA specify that LiDAR data should only be collected when data is being received from at least six satellites and Geometric Dilution of Precision (GDOP) is less than 4 (Environment Agency 2013). It is considered minimum good practice to collect LiDAR data when there are at least four well-distributed satellites with elevation angles greater than 15° (Habib & Van Rens 2007).

### 2.1.3 Survey method

The client will specify the type of LiDAR survey to be carried out, i.e. topographic and/or bathymetric LiDAR and whether full waveform or discrete pulse returns are required.

### 2.1.4 Ancillary data acquisition

The client may specify that optical, infrared or thermal aerial imagery should be collected simultaneously, as this can aid interpretation of the LiDAR data. The client may also specify that a ground truth survey is carried out, but this is dealt with under Data Quality Section 2.3.

## 2.2 Deliverables

The specification should include a list of deliverables and due dates. In addition to the data files, deliverables should include survey-level and file-level metadata, one or more reports and a spreadsheet of QC issues and actions.

### 2.2.1 Data

This section covers the type and format of data outputs that could be specified. Data quality is considered separately in Section 2.3. The data products typically provided by a contractor following a LiDAR survey are a classified point cloud, digital terrain model (DTM) and digital surface model (DSM). The DSM, or 'unfiltered' raster, is created from the points classified as surface features, e.g. vegetation and buildings. The DTM, or 'filtered' raster, is created from points classified as ground features. The client may wish to specify the method to be used for the extraction of above-ground features to create the DTM, i.e. whether this process should be manual or fully automated.

Other derived datasets may include hillshade or slope rasters derived from the DTM, contour vector datasets derived from the DTM, and intensity rasters created from the intensity values of the point cloud (i.e. strength of the pulse return). Filter masks showing where above-ground features have been identified in the point cloud can also be specified as a deliverable. If aerial photography is to be collected, this will be provided as tiled orthoimagery. Flight lines may be provided as a vector file. The client may also request a vector file of the ground control points used to calibrate and process the data and those used to validate outputs.

For each spatial dataset deliverable, the client should specify:

- **Data format**. This could include LAS or LAZ format for point clouds, ASCII, GeoTiff or Cloud Optimised Geotiff for raster files, shapefile or geodatabase feature class for vector files, and PulseWaves[1] for full waveform data. Consideration should be given to whether formats are open or proprietary. Requesting LAZ for point cloud delivery will save storage space and transfer speed; files can be uncompressed to LAS format if required. The LAS/LAZ format version used should be stated in the metadata.

- **Measurement units**, e.g. elevation values to be provided in floating point in metres.

- **Planar Coordinate Reference System** (CRS) consisting of projection and datum, e.g. OSGB36 (EPSG:27700).

- **Transformation** used to convert from the CRS used for data collection to the CRS used in output products. The current (2020) transformation used to convert from WGS84 to OSGB36 is OSTN15/OSMG15 (Greaves *et al*. 2016).

- **Vertical Datum** e.g. Ordnance Datum Newlyn

[1] https://rapidlasso.com/pulsewaves/

- **Raster properties** e.g. bit depth, no data value, and whether pyramids (.OVR files) are required. Resolution is discussed in Section 2.3.2.

- **Tile size and alignment**, e.g. 1km, 5km or 10km tiles which should align perfectly with the OS grid and with each other, with no gap or overlap.

- **File naming convention** for ease of data management. For example, file names could include acquisition date, coordinates and file type. Using the same naming convention for all spatial datasets will facilitate checks of data completeness (see section 4.2). Ground truth surveys must have a unique reference number that includes the survey date.

### 2.2.2 Metadata

The project specification should state the format and structure required for survey-level and file-level metadata, including compliance with standards such as Gemini 2.3[2]. Metadata is usually provided as XML files. In particular, the specification should ensure that the metadata provided is optimised for use in the Scottish Government's spatial data discovery portal[3]. Metadata can be provided to the Scottish Government's spatial data discovery portal by uploading an XML file, duplicating and editing an existing record, or creating a new record using an online template.

### 2.2.3 Report

The survey report is another form of metadata which provides users with more detail. Reports are usually provided in PDF format. The client should specify what must be included in the report. Critical information includes:

- Statement of project specification.

- Details of any departure from the project specification and any factors arising during survey or processing that could affect data quality.

- Data acquisition details including flight logs (aircraft details, height, heading, speed, line ID, start-stop time), flight coverage map, diagrams of satellite GDOP, LiDAR instrument make, model and calibration.

- If a ground truth survey was carried out, details should be provided including equipment used, vertical error value, surface type, height difference across the site and a photograph of the site.

- Data processing details, including software used and any systematic adjustment made using ground truth data.

- QA and QC methods and results, e.g. horizontal and vertical accuracy statistics. If adjustment was carried out to correct systematic bias, positional accuracy statistics should be quoted before and after adjustment. Key statistics should be provided in the report, while the detailed log of issues and actions can be provided as an Excel spreadsheet.

[2] https://www.agi.org.uk/agi-groups/standards-committee/uk-gemini
[3] https://www.spatialdata.gov.scot/geonetwork/srv/eng/catalog.search#/search

## 2.3 Data Quality

The main considerations in evaluating the quality of LiDAR data are horizontal and vertical positional accuracy, consistency of measurements, resolution of the data, and accurate classification of the point cloud leading to correct removal/retention of surface features in the DTM and DSM. Minimum requirements for these should be stated in the project specification.

### 2.3.1 Positional Accuracy

Absolute accuracy of LiDAR data is evaluated through comparison with ground control points (GCPs) to calculate root mean square error (RMSE). The project specification should state the maximum allowable RMSE in metres for both horizontal and vertical accuracy. In vegetated areas, the 95$^{th}$ percentile statistic (the absolute value below which 95% of observations may be found) may be a better measure of vertical accuracy because errors may not be normally distributed (ASPRS 2014). The client may specify a required confidence level for RMSE values or simply ask the contractor to report the confidence level. Mean Bias Error (MBE) may also be calculated as a measure of systematic bias.

The client should state the required vertical accuracy for the GCPs, which must be higher than the required vertical accuracy of the LiDAR data, and specify that the GCPs will be captured on a flat horizontal surface. If precise measurement of horizontal accuracy is required, the specification may state that specially designed LiDAR targets are used as GCPs (Toth *et al.* 2007). The maximum allowable length of time between the ground survey and the LiDAR survey should also be specified.

Horizontal accuracy is always lower than vertical accuracy of LiDAR data (Habib & Van Rens 2007), and this should be reflected in client specifications. Examples of positional accuracy required by UK LiDAR surveys are ≤0.10m vertical RMSE and ≤0.351 horizontal RMSE (Geijsels 2017; Chapman 2019).

### 2.3.2 Data Consistency

Relative accuracy of LiDAR data is evaluated through comparison of overlapping strips of data. Metrics of relative accuracy are less well established than those of absolute accuracy (ASPRS 2018a). A common method is to produce rasters from the point clouds for two separate survey strips and create a ‘difference raster’ for the overlapping area, in which cell values are calculated as the difference between the two input raster strips (Heidemann 2018). The area(s) used for evaluation should be non-vegetated and the data should consist only of single returns. If this method is to be used, the project specification should state the maximum allowable root mean square difference (RMSD) between overlapping rasters.

The client should also specify requirements for removal of spikes in the data, e.g. due to birds or smoke. In coastal surveys, the client may specify that seawater must be removed from the final product, and whether this should be achieved through classification of the point cloud or by clipping to a contour or supplied boundary. The specification should state that there must be no voids in the data other than those caused by water bodies and may specify interpolation methods to fill voids.

The desired % overlap between adjacent flight lines should be stated in the project specification. The larger the overlap, the greater the ability to evaluate relative accuracy, eliminate voids in data and achieve high resolution outputs. However, a large overlap increases flight times and survey costs.

### 2.3.3 Resolution

The client should specify the required minimum and average point density of the point cloud. The client and contractor should also agree on how point density will be calculated, e.g. whether all points or only last return points will be used, and how many cells within the study area need to be above a certain density threshold to pass QC. Minimum point densities required to assure the quality of surface model rasters at a given resolution (i.e. pixel size) are 1 point per square metre for the production of 1m resolution rasters and 4 points per square metre for 0.25m$^2$ rasters (Environment Agency 2015; Chapman 2019). Increasing point density increases the costs of data acquisition, storage and processing, so the point density and raster resolution should be specified based on the minimum requirements for feature detection. As a general rule, detection of features with a width of x metres requires LiDAR data with a pixel size of x/3 metres (Environment Agency 2013).

### 2.3.4 Point cloud classification and surface model generation

The classification system to be applied to the point cloud should be specified in the brief. It may be appropriate to use the American Society of Photogrammetry and Remote Sensing classification codes and flags (ASPRS 2011), or the client may wish to specify a different or project-specific classification system.

There are several interpolation methods for creating raster DSMs and DTMs from the classified point cloud, including triangulation with linear interpolation, inverse distance weighting, kriging, natural neighbour, nearest neighbour and polynomial regression, all of which will produce slightly different results (Fernandez-Diaz *et al.* 2014). The client may wish to specify the interpolation method, or simply ask that the contractor states the method used in the report.

# 3 Quality Assurance

Quality assurance is carried out by the contractor before a survey to ensure that the data collected will meet the project specifications. Detailed discussion of QA protocols is outside the remit of this project, but a brief overview is provided here for context.

The first step of quality assurance is assessment of the area to be surveyed, as this strongly influences data collection parameters (Heidemann 2018). Topography, land use and cover type, airspace control, distribution of GNSS base stations and satellite constellation distribution must all be considered in combination to determine the optimal flight parameters. For example, in forested areas it may be necessary to fly at lower altitude and slower speed with a smaller scan angle and greater overlap, increasing the number of laser pulses to ensure that enough pulses penetrate the canopy to the ground (Habib & Van Rens 2007). Scan angle has been shown to have particular influence on data quality for the generation of forestry metrics such as tree height and canopy closure (Holmgren *et al.* 2003; Qin *et al.* 2017).

Study of the survey area is also essential for selection of appropriate ground sampling locations, which should be horizontal, hard surfaces with minimal variation in elevation, free from obstructions which could block GNSS signal or reflect laser pulses (Clancy 2011). The network of GCPs should adequately represent both vegetated and non-vegetated terrain, taking account of recommendations for GCP numbers based on survey area (Table 2).

**Table 2.** Recommended number of GCPs based on survey area from Table C.1 in ASPRS (2014).

| Survey area ($km^2$) | Minimum GCPs in non-vegetated areas | Minimum GCPs in vegetated areas | Total GCPs |
|---|---|---|---|
| ≤500 | 20 | 5 | 25 |
| 501 – 750 | 20 | 10 | 30 |
| 751 – 1000 | 25 | 15 | 40 |
| 1001 – 1250 | 30 | 20 | 50 |
| 1251 – 1500 | 35 | 25 | 60 |
| 1501 – 1750 | 40 | 30 | 70 |
| 1751 - 2000 | 45 | 35 | 80 |
| 2001 – 2250 | 50 | 40 | 90 |
| 2251 - 2500 | 55 | 45 | 100 |

Control point spacing and pattern must be designed to achieve the specified data quality, taking account of practical considerations such as site access restrictions, travelling distances and staff safety. Ground sample features should cover at least 100$m^2$ on a surface that will not change between surveys to enable repeatability (Chapman 2019). Consideration of terrain, accessibility and distance to virtual reference stations will determine whether the most appropriate positioning method is real time kinematic (RTK) or post-processed kinematic (PPK) GNSS, unless this has been specified by the client.

LiDAR systems can be configured by the operator pre-survey or during flight to achieve the specified point density. The parameters which can be manipulated are scan frequency (0 - 100Hz), scan angle (0 – 30°), pulse repetition frequency (up to 150kHz) which defines number of pulses per second, beam divergence which regulates the width of the individual laser beam, and roll compensation which corrects for aircraft roll (Saylam 2009). Because of their diverse and specialised nature, LiDAR systems have their own mission planning software enabling manipulation of these parameters. Users have only limited input into quality assurance because LiDAR systems tend to be a ‘black box’ (Habib & Cheng 2006), so the quality control procedures outlined in the following sections are essential for ensuring that data meet the specification and are fit for purpose.

# 4 Quality Control

Quality control is carried out post-survey to ensure that the outputs, particularly the data products, meet the project specifications. The initial QC activities are carried out by the contractor as they require access to raw data and are integrated into the data processing workflow. The final set of QC activities are carried out by the client on the processed data. In both cases a log should be kept (e.g. in spreadsheet format), to track issues identified during QC and any steps taken to correct them.

## 4.1 QC activities performed by contractor

Many of the QC activities performed by the contractor are integrated into the data processing workflow, e.g. the classification of the point cloud and creation of surface models. A variety of software and tools are available to process and quality control LiDAR data, including some open source options. A list of commonly used software and tools is provided in Appendix 1.

### 4.1.1 Positional accuracy

The contractor will calculate horizontal and vertical accuracy statistics (RMSE, 95$^{th}$ Percentile, MBE) as specified in the project brief and ensure the values meet the client's specifications. The contractor will also check and confirm that the number, distribution and vertical accuracy of the GCPs meet the project specification. All accuracy statistics and confidence levels will be recorded in the spreadsheet and report.

### 4.1.2 Data consistency

The contractor will first check that the % overlap between adjacent flight lines specified in the brief has been achieved. There are several possible approaches to generating inter-swath quality metrics, i.e. evaluating the consistency of LiDAR data using the overlapping areas, but no broadly accepted method is in operational use (ASPRS 2018a). A 'difference' raster can be created as outlined in section 2.3.3, and the minimum, maximum and RMSD values reported to the client (Heidemann 2018). Other variations on this approach are to use a Triangulated Irregular Model (TIN) instead of a raster, or to calculate the statistics from randomly or manually selected points within the overlap area. These 'difference' methods do not provide a complete assessment of inter-swath consistency as they only measure vertical and not horizontal differences, systematic errors are not quantified, and the statistics may include non-valid measurements that were not made on hard surfaces (ASPRS 2018a).

An alternative approach is to use an algorithm to detect corresponding features (preferably lines or planes rather than points) in overlapping areas and calculate the transformation parameters, i.e. rotations and translations, between them (Habib & Cheng 2006; Habib *et al*. 2010). Low transformation parameter values indicate high internal accuracy, with transformation parameters of zero indicating that the surfaces match exactly. Transformation parameters can be used to describe and potentially correct systematic bias in the data. A benefit of this approach is that it is applied to raw data, eliminating the need to create rasters/TINS which is time-consuming and could produce artefacts which would skew the results.

The following quality measures can be derived from analysis of corresponding features in overlap areas (ASPRS 2018a):

- Median discrepancy angle for systematic errors. The discrepancy angle is measured from the 'centre of overlap', a line of best fit through the two sets of corresponding points.

- Mean and root mean square difference (RMSD) of horizontal errors measured on sloping surfaces.

- RMSD of vertical errors, based on measuring the perpendicular distance in metres from the centroid of one plane or line to the centroid of its corresponding feature.

### 4.1.3 Point density

The contractor should calculate minimum and mean point density of the point cloud using the agreed method to ensure these meet the client's specifications. These values should be recorded in the QC spreadsheet and report. One approach is to use the LAS Point Statistics as Raster tool in ArcMap, selecting 'pulse_count' as the method and defining the cell size of the output grid as four times the required resolution of the final output rasters (Geijsels 2017).

Although it is widely used, calculation of mean point density from a grid of square cells gives no insight into spatial variation in density. An alternative method is to generate a Voronoi diagram in which a polygon is created around each point with the boundary equidistant between the point and its nearest neighbours (Figure 3). This can be used to analyse spatial variations in the point cloud, remove or reconstruct data voids, and generate statistics on point spacing as well as point density (Rupnik *et al.* 2015).

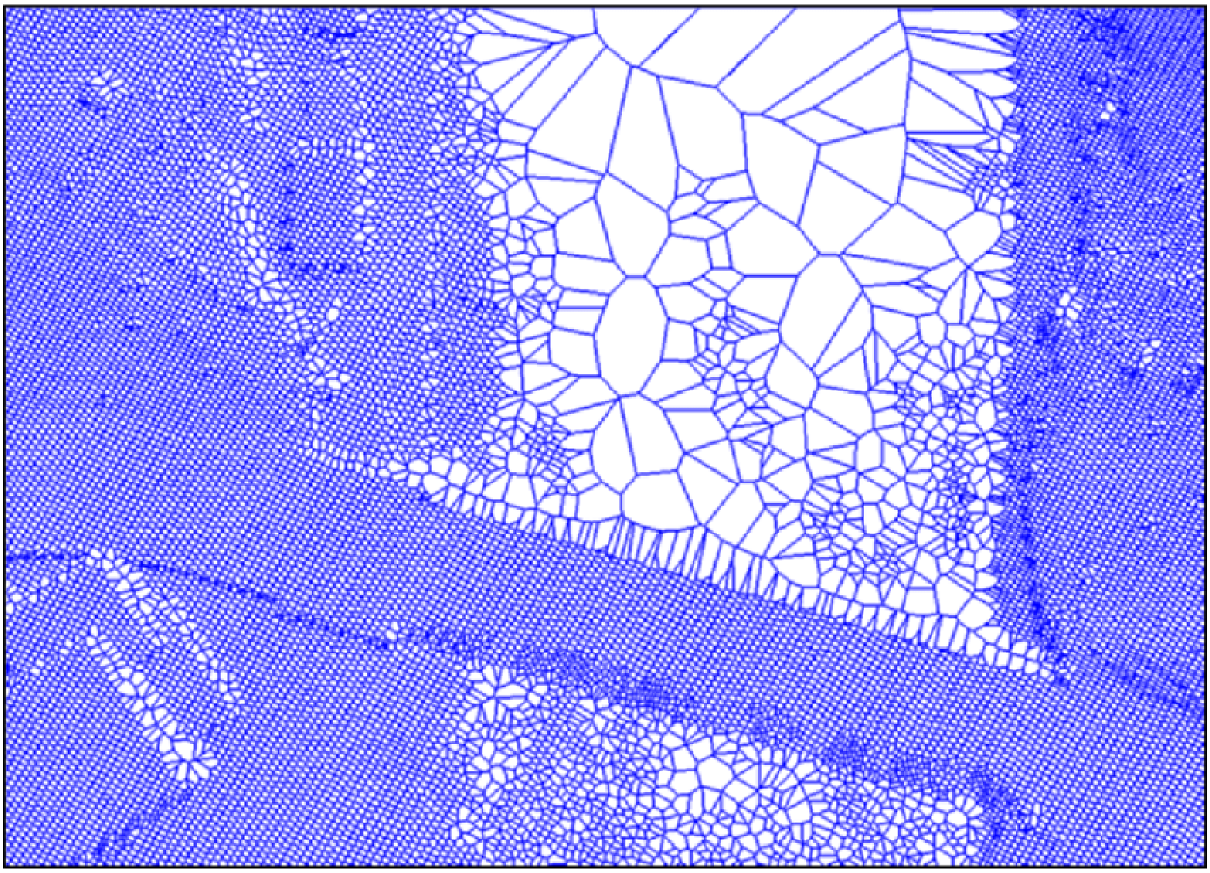

**Figure 3:** Example of a Voronoi diagram generated from a LiDAR point cloud. The very large Voronoi cells are over a river where there are voids in the data due to weak returns from water. In contrast, small Voronoi cells over the riverbanks and bridge indicate consistent high point density in those areas. There are slightly large Voronoi cells on the north-east sides of buildings caused by LiDAR shadows. Image from Rupnik *et al*. 2015.

### 4.1.4 Classified point cloud and surface models

The contractor will visually check the classification accuracy of the point cloud and ensure that no points remain assigned to class 0 (created but not processed for classification). The contractor must also check for and report on any voids in the data. Voids can be caused by shadows, e.g. from buildings or high terrain, or by weak pulse returns, e.g. from water bodies (Becker *et al.* 2009). The contractor must remove spikes from the data, e.g. due to smoke, clouds or birds, which will also create voids which may need to be filled through interpolation.

The contractor will produce the surface models using the interpolation method and output resolution agreed in the project brief, followed by any agreed derived datasets such as hillshade rasters or intensity rasters. The contractor will check the rasters visually for issues such as voids or spikes in the data. The DTM and DSM will be visually compared at an appropriate resolution to ensure that above-ground features have been removed or retained correctly.

All issues detected and actions taken to correct them, e.g. removal of spikes, will be recorded in the spreadsheet and report.

## 4.2 Recommended QC workflow for client

On receipt of deliverables, the client should carry out the following checks to ensure compliance with the project specification.

### 4.2.1 Check deliverables for completeness

The client must check that:

- All deliverables have been received, namely the report(s), data files, metadata and any additional datasets such as vector files of flightlines or GCPs.
- Files are in the correct formats (and correct version of LAS/LAZ formats).
- The agreed file naming convention has been followed and there are no duplicate filenames.

Because the data are delivered as tiles, there is likely to be a very large number of files, making it difficult to check them manually. A script to count point cloud, DTM and DSM files and compare lists to check for gaps or duplications is provided in Appendix 2.

### 4.2.2 Check report and QC log

The report should include a statement of the project specifications and detail any departure from these. This will facilitate checking the report to ensure the specifications have been met. In particular, the client should check that the following criteria have been met:

- Horizontal and vertical absolute accuracy of data.
- Relative accuracy, i.e. inter-swath consistency.
- Point density.

The client should check that the accuracy values quoted in the report, e.g. vertical RMSE, match those reported in the QC spreadsheet and that they meet the minimum requirements stated in the project brief. The client should use the report and flight logs to check that the survey was carried out within the specified timeframe under the agreed conditions, e.g. weather and GNSS constellation/GDOP parameters. Finally, the client should check the report and QC spreadsheet for any specific issues highlighted by the contractor that might affect the data's fitness for purpose.

### 4.2.3 Check data quality

The first step is to check that the data provides complete coverage of the entire survey area. To facilitate this, the tiles should be merged to create a mosaic. A script for mosaicking tiles, saving the mosaic image as a GeoTiff and building pyramids is provided in Appendix 3. The mosaic can then be opened in a GIS application and visually checked to ensure that it covers the survey area. Even if the check of the list of file names shows that all files have been delivered, the coverage must be checked in GIS as some files may not open or visualise correctly (Figure 4).

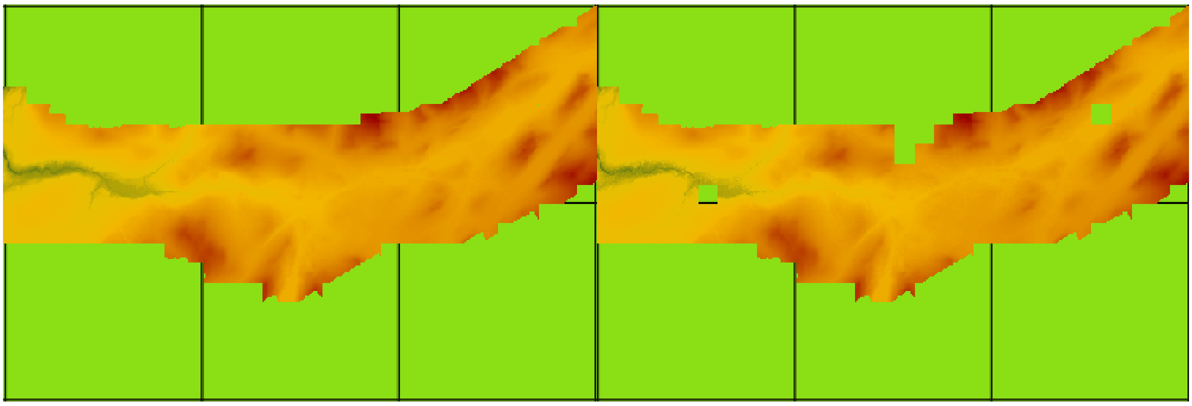

Figure 4. DTM (left) and DSM (right) derived from LiDAR data. The DSM has several 5km x 5km gaps. Although the tiles for those areas had been delivered, they were corrupt and would not visualise in GIS.

Layer Properties in GIS can be used to check that the projection, resolution, units and no data value meet the specification, and that the origin and pixel size are integer values (Figure 5).

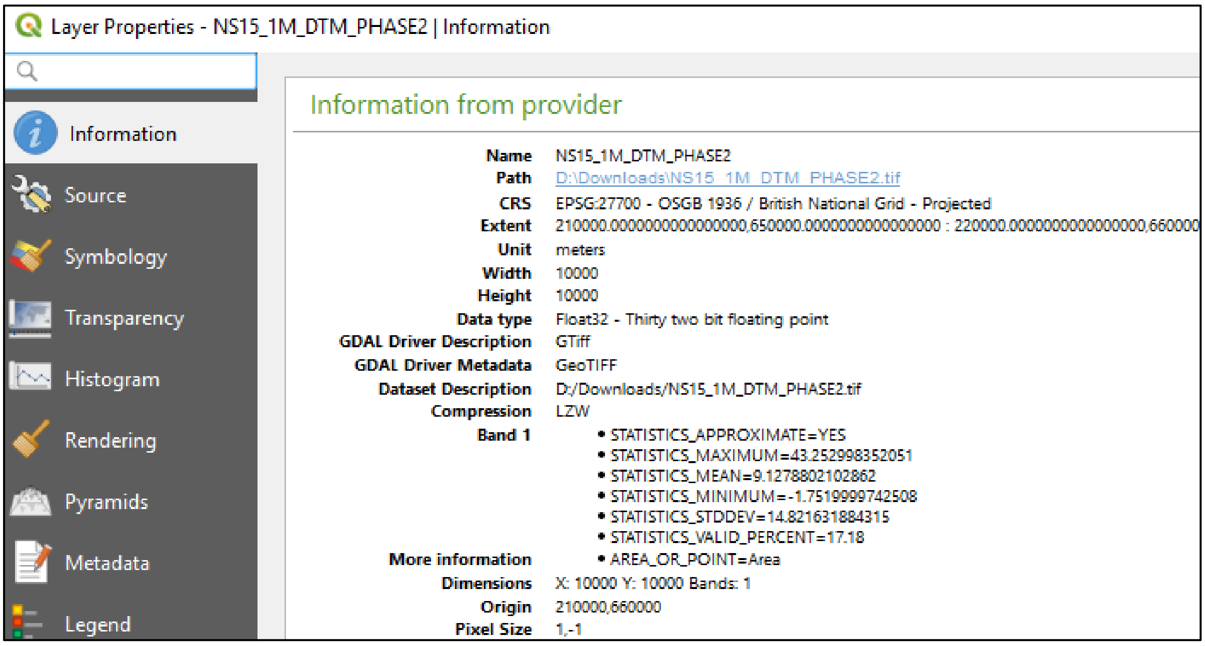


Figure 5. Layer properties window in QGIS (free open source desktop GI software[4]) showing information on a LiDAR DTM including coordinate reference system, file format and resolution.

The data can then be visually checked at the appropriate viewing scale. The recommended scale for QC of 1m resolution LiDAR data is 1:2,000 (Geijsels 2017) (Figure 6).

[4] https://qgis.org/en/site/

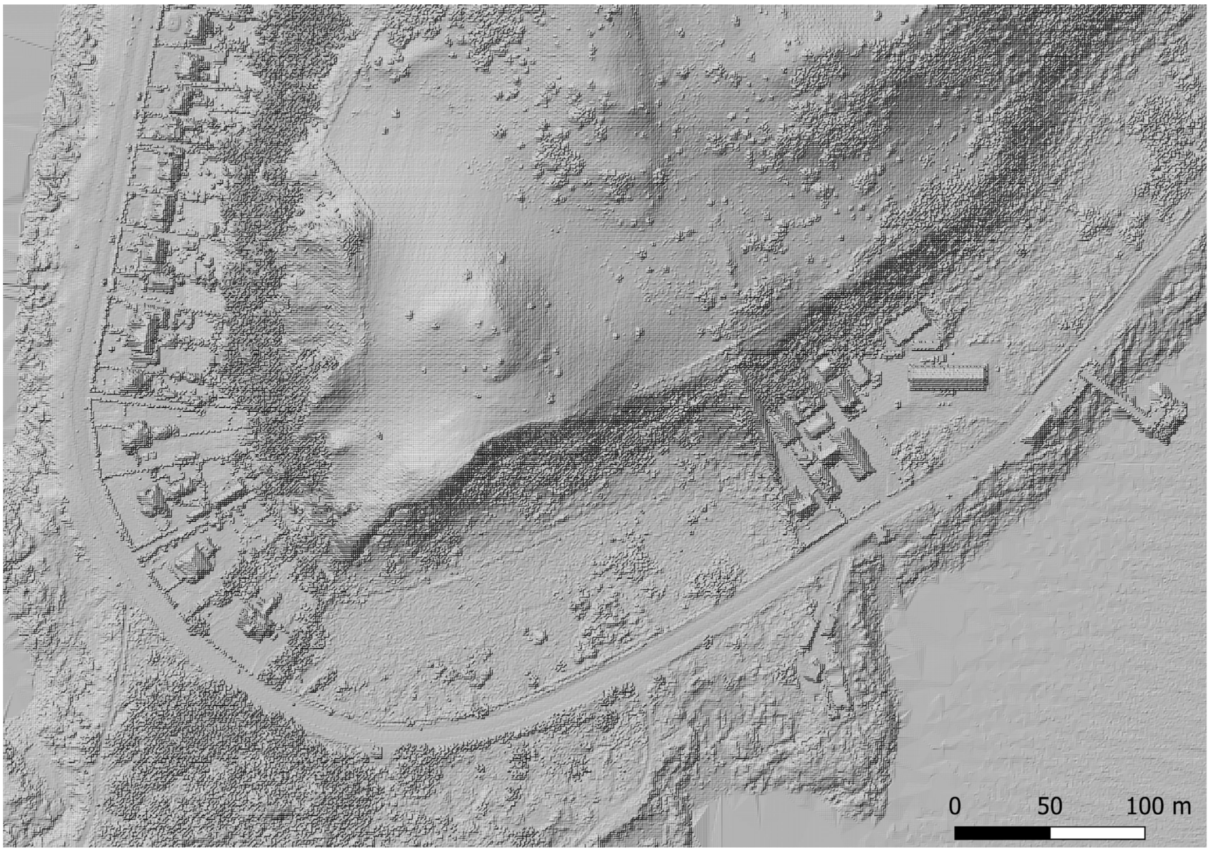


**Figure 6.** 1m resolution LiDAR DSM of Marine Parade, Isle of Cumbrae at 1:2,000 scale. Data from the Scottish Remote Sensing Portal [accessed June 2020]. A 'hillshade' visualisation has been applied in QGIS software.

As both the DTM and DSM are produced from the same point cloud, any issues such as artefacts or outliers are likely to be present in both datasets. Therefore the DSM should be checked first for issues such as incomplete coverage, voids or spikes, after which the DTM should be checked to ensure that features such as buildings and vegetation have been removed correctly (Environment Agency 2015). Layer 'swipe' tools in GI software are useful for comparing the two datasets to look for anomalies (Figure 7).


**Figure 7.** 1m resolution LiDAR DSM (left) and DTM (right) of Kirkwall, Orkney. Data from the Scottish Remote Sensing Portal [accessed June 2020]. A hillshade visualisation has been applied in QGIS.

A spectral colour ramp is best for visualising the data, as a two-colour stretch does not provide enough variation to highlight issues. Changing the thresholds between colour steps can accentuate issues in particular areas of the data. Creating a hillshade raster from the mosaic surface models is a good way to look for artefacts in the data such as abrupt edges, which can indicate issues with the sensor.

Horizontal accuracy should be checked visually through comparison with a high-resolution base map such as OS MasterMap or with high resolution aerial imagery of known horizontal accuracy. In coastal areas, the client should check that water has been removed as agreed using a consistent method (Figure 8), and that the full extent of the intertidal zone has been captured, for example with reference to OS MLWS (mean low water spring) contours.

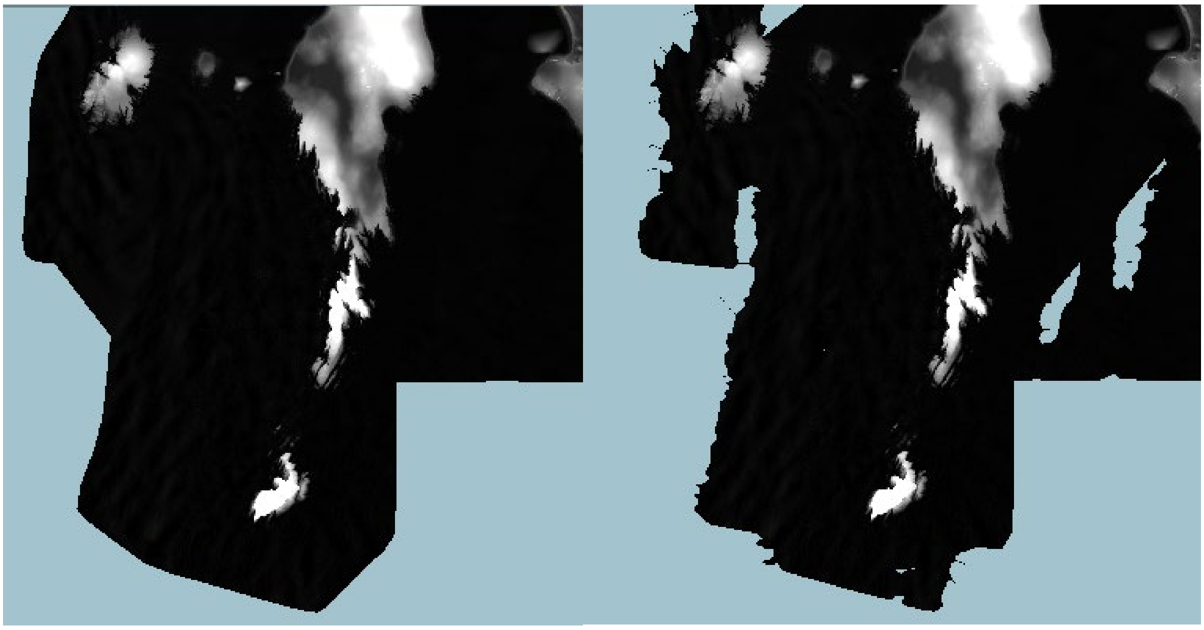

**Figure 8.** Two versions of a LiDAR DSM produced from the same point cloud. Water removal has not been carried out consistently, producing different results in the two versions.

The survey-level and file-level metadata should be checked to ensure it is in the agreed format with all fields complete and compliant with the specified standards.

As a final step, if the data are to be made available for visualisation via a website, they will need to be retiled to an appropriate size. A script to segment a mosaic imagery into tiles of a given size is provided in Appendix 3, and a script to mosaic small tiles to create larger tiles is provided in Appendix 4. After retiling, the newly created tiles should be checked to ensure they are correctly aligned with the OS grid, i.e. that their origin and pixel size are integer values.

# 5 Glossary

| | |
|---|---|
| ASPRS | American Society of Photogrammetry and Remote Sensing |
| CRS | Coordinate Reference System |
| DSM | Digital Surface Model |
| DTM | Digital Terrain Model |
| EA | Environment Agency |
| GCP | Ground Control Point |
| GDOP | Geometric Dilution of Precision |
| GIS | Geographic Information System |
| GNSS | Global Navigation Satellite System |
| GSD | Ground Sampling Distance |
| INS | Internal Navigation System |
| HES | Historic Environment Scotland |
| LiDAR | Light Detection and Ranging |
| MBE | Mean Bias Error |
| NIR | Near Infra-Red |
| NRW | Natural Resources Wales |
| PPK | Post Processed Kinematic |
| QA | Quality Assurance |
| QC | Quality Control |
| RMSD | Root Mean Square Difference |
| RMSE | Root Mean Square Error |
| RTK | Real Time Kinematic |
| SEPA | Scottish Environment Protection Agency |
| TIN | Triangulated Irregular Network |

# 6 References


ASPRS. (2018a) ASPRS Guidelines on Geometric Inter-Swath Accuracy and Quality of LiDAR Data. *Photogrammetric Engineering and Remote Sensing*. 84 (3), 117–128.

ASPRS. (2014) *ASPRS Positional Accuracy Standards for Digital Geospatial Data version 1.0*.

ASPRS. (2011) *LAS Specification version 1.4 - R6*.

ASPRS. (2018b) *Summary of Research and Development Efforts Necessary for Assuring Geometric Quality of LiDAR Data version 1.0*.

Becker, J., Stewart, C.V. & Radke, R.J. (2009) LiDAR inpainting from a single image. *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*.

Chapman, C. (2019) *National LiDAR Survey for Wales TECHNICAL SPECIFICATION (NLSW3) v.0.8.3*.

Clancy, S. (2011) The Importance of Applied Control. LiDAR Magazine 1 (2).

Demir, N., Bayram, B., Şeker, D.Z., Oy, S., İnce, A. & Bozkurt, S. (2019) Advanced Lake Shoreline Extraction Approach by Integration of SAR Image and LIDAR Data. *Marine Geodesy*. 42 (2), 166–185.

Elsherif, A., Gaulton, R., Shenkin, A., Malhi, Y. & Mills, J. (2019) Three dimensional mapping of forest canopy equivalent water thickness using dual-wavelength terrestrial laser scanning. *Agricultural and Forest Meteorology*. 276–277107627.

Environment Agency. (2013) *National Standard Contract and Specification for Surveying Services: STANDARD TECHNICAL SPECIFICATIONS version 3.2*.

Environment Agency. (2015) *Specification for LiDAR Surveys Version 1.0*.

Fernandez-Diaz, J.C., Carter, W.E., Shrestha, R.L. & Glennie, C.L. (2014) Now You See It… Now You Don't: Understanding Airborne Mapping LiDAR Collection and Data Product Generation for Archaeological Research in Mesoamerica. *Remote Sensing*. 6 (10), 9951–10001.

Fragoso-Campón, L., Quirós, E., Mora, J., Gutiérrez, J.A. & Durán-Barroso, P. (2018) Accuracy Enhancement for Land Cover Classification Using LiDAR and Multitemporal Sentinel 2 Images in a Forested Watershed. *Proceedings*. 2 (20), 1280.

Geijsels, S. (2017) *Quality Control of Lidar Data Version 1.0*.

Greaves, M., Downie, P. & Fitzpatrick, K. (2016) OSGM15 and OSTN15: updated transformations for UK and Ireland. Geomatics World p.5.

Habib, A. (2014) Optimising Quality Control for Lidar Data Processing. GIM International

Habib, A. & Cheng, R.W.T. (2006) 'Surface Matching Strategy for Quality Control of LiDAR Data', in *Innovations in 3D Geo Information Systems*. Lecture notes in geoinformation and cartography. [Online]. Springer.

Habib, A., Kersting, A.P., Bang, K.I. & Lee, D.-C. (2010) Alternative Methodologies for the Internal Quality Control of Parallel LiDAR Strips. *IEEE Transactions on Geoscience and Remote Sensing*. 48 (1), 221–236.

Habib, A. & Van Rens, J. (2007) *Quality Assurance and Quality Control of LiDAR Systems and Derived Data*.

Heidemann, H.K. (2018) 'Lidar Base Specification v.1.3', in *Book 11: Collection and Delineation of Spatial Data*. US Geological Survey Standards. [Online]. US Department of the Interior. p.

Holmgren, J., Nilsson, M. & Olsson, H. (2003) Simulating the effects of lidar scanning angle for estimation of mean tree height and canopy closure. *Canadian Journal of Remote Sensing*. 29 (5), 623–632.

Hopkinson, C., Chasmer, L., Gynan, C., Mahoney, C. & Sitar, M. (2016) Multisensor and Multispectral LiDAR Characterization and Classification of a Forest Environment. *Canadian Journal of Remote Sensing*. 42 (5), 501–520.

Kellner, J.R., Armston, J., Birrer, M., Cushman, K.C., Duncanson, L., Eck, C., Falleger, C., Imbach, B., Král, K., Krůček, M., Trochta, J., Vrška, T. & Zgraggen, C. (2019) New Opportunities for Forest Remote Sensing Through Ultra-High-Density Drone Lidar. *Surveys in Geophysics*. 40 (4), 959–977.

Lohani, B., Ghosh, S. & Dashora, A. (2018) 'A Review of Standards for Airborne LiDAR Data Acquisition, Processing, QA/QC, and Delivery', in N.L. Sarda, P.S. Acharya, & S. Sen (eds.) *Geospatial Infrastructure, Applications and Technologies: India Case Studies*. [Online]. Singapore: Springer. pp. 305–312.

Qin, H., Wang, C., Xi, X., Tian, J. & Zhou, G. (2017) Simulating the Effects of the Airborne Lidar Scanning Angle, Flying Altitude, and Pulse Density for Forest Foliage Profile Retrieval. *Applied Sciences*. 7 (7), 712.

Resop, J.P., Lehmann, L. & Hession, W.C. (2019) Drone Laser Scanning for Modeling Riverscape Topography and Vegetation: Comparison with Traditional Aerial Lidar. *Drones*. 3 (2), 35.

Rupnik, B., Mongus, D. & Zalik, B. (2015) Point Density Evaluation of Airborne LiDAR Datasets. *Journal of Universal Computer Science*. 21 (4), 587–603.

Sánchez Sánchez, Y., Martínez-Graña, A., Santos Francés, F. & Mateos Picado, M. (2018) Mapping Wildfire Ignition Probability Using Sentinel 2 and LiDAR (Jerte Valley, Cáceres, Spain). *Sensors*. 18 (3), 826.

Saylam, K. (2009) '*Quality Assurance of LiDAR Systems - Mission Planning*', in [Online]. 2009 Baltimore, Maryland: . p.

Saylam, K., Hupp, J.R., Averett, A.R., Gutelius, W.F. & Gelhar, B.W. (2018) Airborne lidar bathymetry: assessing quality assurance and quality control methods with Leica Chiroptera examples. *International Journal of Remote Sensing*. 39 (8), 2518–2542.

Stovall, A.E.L., Diamond, J.S., Slesak, R.A., McLaughlin, D.L. & Shugart, H. (2019) Quantifying wetland microtopography with terrestrial laser scanning. *Remote Sensing of Environment*. 232111271.

Toth, C.K., Csanyi, N. & Grejner-Brzezinska, D.A. (2007) 'Improving LiDAR-based Surface Reconstruction Using Ground Control', in Tregoning, P. & Rizos, C. (eds.) *Dynamic Planet: Monitoring and Understanding a Dynamic Planet with Geodetic and Oceanographic Tools IAG Symposium Cairns, Australia 22–26 August, 2005*. [Online]. Berlin, Heidelberg: Springer Berlin Heidelberg. pp. 817–824.

# Appendix 1: Software and tools for LiDAR processing and QC

This list of software and tools has been compiled from the literature and consultation with specialist staff. It is not intended as an exhaustive list or an endorsement of any particular software packages. In addition to these specialist software packages, it is assumed that staff will use a GIS package such as Esri ArcMap / ArcGIS Pro or QGIS.

| |
|---|
| **CloudCompare** (free and open source)<br>http://www.cloudcompare.org/ |
| **ENVI LiDAR**<br>L3 Harris Geospatial Solutions<br>https://www.harrisgeospatial.com/docs/using_envi_lidar_Home.html |
| **ERDAS Imagine,**<br>Hexagon Geospatial<br>https://www.hexagongeospatial.com/products/power-portfolio/erdas-imagine/ |
| **Fusion**<br>DataONE<br>https://www.dataone.org/software-tools/fusion-lidar-software |
| **GRASS GIS LiDAR tools** (free and open source)<br>https://grasswiki.osgeo.org/wiki/LIDAR |
| **LAStools**<br>rapidlasso GmbH<br>https://rapidlasso.com/lastools/ |
| **lidar2dems** (free and open source)<br>https://applied-geosolutions.github.io/lidar2dems / |
| **LiDAR360**<br>GreenValley International<br>https://greenvalleyintl.com/software/lidar360/ |
| **LP360**<br>GeoCue Group<br>https://geocue.com/products/lp-360/ |
| **Point Cloud Library** (free and open source)<br>https://pointclouds.org/ |
| **Quick Terrain Modeler**<br>Applied Imagery<br>http://appliedimagery.com/ |
| **TerraMatch**<br>Terrasolid<br>http://www.terrasolid.com/products/terramatchpage.php |
| **Terrascan**<br>Terrasolid<br>http://www.terrasolid.com/products/terrascanpage.php |

# **Appendix 2:** Script to compare lists of files and check for gaps

In Powershell, count the files delivered in LAS (point cloud) and TIF (DTM and DSM) format.

```
Write-Host @(Get-ChildItem *.las).Count

Write-Host @(Get-ChildItem *.tif).Count
```

In a terminal window you can use: (provided that the file names don't include newlines).

```
find  -maxdepth 1-name '*.las'|wc -l
```

If the number of files don't match or aren't as expected, you can compare the three product lists to find out what is missing. First, extract the list of files into an appropriate format. In Powershell, chain things together to output a basename:

```
Get-ChildItem *.las | Select Basename > /path/to/outputs/list_las.txt
```

In a terminal, use the 'find' command to get this.

```
find . -maxdepth 1 -name '*.las' > /path/to/outputs/list_las.txt
```

You will need to do some editing to remove any headers put in by the Powershell script or to remove additional file names from the eventual result, i.e. find / replace '.las' with '' in a text editor for all the lists.

Once you have the three lists of file names, you can start to compare them. Comparing three types of entry is a little difficult, but you can use tools like KDiff3 to highlight differences between three inputs, e.g. to compare the lists of grid squares.

To compare two lists you can use a simple Python script.

```
las_old = []
las_new = []

with open("D:\\p1-new-las.txt", "r") as l:
    for line in l:
        las_new.append(line.replace("\n", ""))

with open("D:\\p1-cur-laz.txt", "r") as l:
    for line in l:
        las_old.append(line.replace("\n", ""))

old_idx = 0;
new_idx = 0;

not_finished = True

with open("D:\\las-compare.csv", "w") as output:
    while not_finished:
```

```
        if (las_old[old_idx] == las_new[new_idx]):
            output.write("{0}, {1}\n".format(las_old[old_idx],
las_new[new_idx]))
            old_idx += 1
            new_idx += 1
        elif (las_old[old_idx] < las_new[new_idx]):
            print("{0}, {1}".format(las_old[old_idx],
las_new[new_idx]))
            output.write("{0}, \n".format(las_old[old_idx]))
            old_idx +=1
        elif (las_old[old_idx] > las_new[new_idx]):
            print("{0}, {1}".format(las_old[old_idx],
las_new[new_idx]))
            output.write(", {0}\n".format(las_new[new_idx]))
            new_idx +=1

        if (old_idx == len(las_old)):
            while (new_idx != len(las_new)):
                output.write(", {0}\n".format(las_new[new_idx]))
                new_idx +=1
            not_finished = False
```

This will produce a CSV with blank spaces for missing data on either side of the list. You can then make a comparison between missing data on both sides of that list, which is useful for confirming whether replacement data covers existing areas. This can be done in MS Excel using a simple formula:

=IF(AND(A2="", NOT(B2="")),TRUE,FALSE)]

You can then sum the TRUE values to get an idea of missing or extra data files.

# **Appendix 3:** Script to create mosaic using GDAL

Combine all DSM or DTM 10km grids to create a single mosaic with full coverage using gdalbuildvrt.

```
gdalbuildvrt phase-1-dtm.vrt c:\data\scotland\phase-1\*.tif
```

This produces a single virtual raster file (essentially an XML file) that can be opened in any GIS application. However, due to the size of the data, it is best to create a single TIF file using gdal_translate.

```
gdal_translate -of GTiff -co "COMPRESS=LZW" -co "TILED=YES" -co
"BIGTIFF=YES" phase-1-dtm.vrt phase-1-dtm.tif
```

- -co "COMPRES=LZW" -> Enables lossless compression on the image
- -co "TILED=YES" -> Enables inner tilling on the image
- -co "BIGTIFF=YES" -> Forces the use of the BigTIFF standard (allows > 4GB file size)

This creates a single combined GeoTIFF. The next step is to create pyramids to optimise the file for viewing so that the image can be checked as part of the QC process.

```
gdaladdo phase-1-dtm.tif 2 4 8 16 32 64 128
```

This creates an image pyramid inside the input file. You can add additional overviews if required depending on the size / complexity of the image, simply add more $2^x$ numbers to the end of the chain above.

If the imagery is to be made available for viewing on a website, it will need to be retiled. The tiles can be larger than the original input tiles but should be < 10GB in size. The following Python script creates a folder of tiles (each sized 10,240 x 10,240 pixels) from the source image. This needs to be run in an environment with gdal Python bindings.

```
import os, sys
from osgeo import gdal

dset = gdal.Open(sys.argv[1])

width = dset.RasterXSize
height = dset.RasterYSize

print(str(width) + 'x' + str(height))

tilesize = 10240

for i in range(0, width, tilesize):
    for j in range(0, height, tilesize):
        w = min(i+tilesize, width) - i
        h = min(j+tilesize, height) - j
        gdaltranString = "gdal_translate -of GTIFF -co \"TILED=YES\"
-co \"COMPRESS=LZW\" -srcwin "+str(i)+", "+str(j)+", "+str(w)+", " \
```

```
            +str(h)+" " + sys.argv[1] + "
./geoserver_output/"+str(i)+"_"+str(j)+".tif"
        os.system(gdaltranString)
        gdaladdoString = "gdaladdo
./geoserver_output/"+str(i)+"_"+str(j)+".tif 2 4 8 16 32 64 128"
        os.system(gdaladdoString)
```

# **Appendix 4:** Script to merge small tiles to create larger tiles

This script can be run in an OSGeo4W terminal on Windows to merge 1km tiles from a source folder following a format to create 10km tiles.

```
import glob
import os
import gdal_merge

dsm = glob.glob("W:\\DSM\\[A-Z][A-Z][0-9][0-9][0-9][0-9]_dsm.tif")
dsm_10k = sorted(list(set(list(map(lambda x:
"{0}{1}".format(x[7:10], x[11]), dsm)))))

dsm_10k_output_dir = "W:\\DSM\\10k"

if not os.path.exists(dsm_10k_output_dir):
    os.makedirs(dsm_10k_output_dir)

for pattern_10k in dsm_10k:
    inputs = glob.glob("W:\\DSM\\{0}[0-9]{1}[0-
9]_dsm.tif".format(pattern_10k[0:3], pattern_10k[3]))
    gdal_merge.main(['', '-o',
'{0}_1M_DSM_PHASE1.tif'.format(os.path.join(dsm_10k_output_dir,
pattern_10k)),
                     '-of', 'GTiff', '-co', 'COMPRESS=LZW', '-co',
'TILED=YES', '-init', '-9999', '-n',
                     '-9999', '-a_nodata', '-9999'] + inputs)
```

The first line:

```
dsm = glob.glob("W:\\DSM\\[A-Z][A-Z][0-9][0-9][0-9][0-9]_dsm.tif")
```

creates a list of files matching those found according to a glob pattern, in this case in the directory `W:\DSM\` the pattern being two letters, then four numbers and then `_dsm.tif`. This feeds into the next line which creates a unique sorted list of the grid squares (after they are masked to make them 10k grids):

```
dsm_10k = sorted(list(set(list(map(lambda x:
"{0}{1}".format(x[7:10], x[11]), dsm)))))
```

This takes the file name (removes the directory) and then takes the first three characters and then the fifth in the string i.e. HY2042 -> HY24. The list is converted to a set (to remove duplicates) and then sorted to produce the output.

The next ‘if’ statement ensures that the output directory exists

```
if not os.path.exists(dsm_10k_output_dir):
    os.makedirs(dsm_10k_output_dir)
```

The final loop loops through the list of 10km squares and runs gdal_merge over the relevant files to produce the output.

This creates the output file name with a full output path.

```
'{0}_1M_DSM_PHASE1.tif'.format(os.path.join(dsm_10k_output_dir,
pattern_10k))
```

This creates a list of the input files by matching it with a glob pattern against the files that exist in the input directory (i.e. HY24 matches HY2041, HY2149, *etc*...).

```
inputs = glob.glob("W:\\DSM\\{0}[0-9]{1}[0-
9]_dsm.tif".format(pattern_10k[0:3], pattern_10k[3]))
```