THE EUROSDR RPAS BENCHMARK: OPEN DATASET DESCRIPTION AND SUMMARY OF KEY RESULTS

J. P. Mills *, M. V. Peppa, A. Alma'Amari, L. Davidson, J. Goodyear, N. T. Penna

School of Engineering, Newcastle University, Newcastle upon Tyne, NE1 7RU, United Kingdom
– (jon.mills, maria-valasia.peppa, A.A.S.Al-Maamari2, lesley.davidson, james.goodyear, nigel.penna)@newcastle.ac.uk

Commission I

KEY WORDS: Aerial triangulation, Benchmark, EuroSDR, Lidar, Orientation, RPAS, SfM Photogrammetry, UAV

ABSTRACT:

In 2021 EuroSDR initiated a benchmark study with the aim to evaluate the geometric quality of real-world survey data generated from state-of-practice commercial Remotely Piloted Aircraft System (RPAS) photogrammetry (including DJI P4 RTK and DJI P1) and lidar (including DJI L1 and Riegl MiniVUX). The particular benchmark focus was on achievable data quality from real-world network configurations in the absence of ground control, on-the-fly Real Time Kinematic (RTK) corrections, and/or local GNSS base station information. Successive custom datasets were released to registered benchmark participants who submitted individual outputs that were independently evaluated against reference surveys. Without the inclusion of any supporting ground information, DJI P4 RTK and DJI P1 RPAS solutions were found to deliver m- and dm-level accuracies, respectively, in both plan and height. RTK solutions were found to provide cm-level precisions and accuracies, with some outliers. The introduction of ground control points resulted in similar planimetric accuracy to the RTK solutions, but with slight improvements in height. In terms of lidar datasets, the Riegl MiniVUX solution, using corrections from a local base station, was found to provide smaller discrepancies than the DJI L1 RTK solution, when independently compared against terrestrial laser scanning surveys. This paper provides various quality statistics and demonstrates multiple ways of assessing the geometric quality of RPAS data. The EuroSDR RPAS benchmark datasets are now openly available online in order to support and facilitate further investigation by the community.

1. INTRODUCTION

1.1 Background

A desirable scenario in the application of airborne photogrammetry and/or laser scanning is the freeing of processing pipelines from the need for support in the form of ground control points (GCPs) and, ideally, local Global Navigation Satellite System (GNSS) base stations. Reports and opinions vary as to what extent it is feasible to achieve high quality, reliable surveys without such support, but it is apparent that its achievement would bring significant benefits to users that include National Mapping and Cadastral Agencies (NMCAs). Bakuła et al., (2019) note that "the topic of evaluating data, sensors and algorithms with benchmarking activities is interesting as it provides the opportunity to compare, with a unique approach, research results from independent scientists". In 2021 EuroSDR therefore initiated a benchmark project with the aim of evaluating the true geometric quality of real-world survey data generated from Remotely Piloted Aircraft System (RPAS) photogrammetry and lidar when performed under different levels of control configuration. This paper describes the benchmark dataset, which has been released as open data, and presents key results from the study. The original EuroSDR RPAS Benchmark webpage can be found at Geospatial.github (2021).

Over the last 20 years there has been a dramatic increase in the number of research studies that have applied RPAS for a plethora of remote sensing applications, as reviewed by Nex et al., (2022). Parallel to the application of RPAS in scientific research, there have also been significant advances in RPAS technology, such as rotor platforms carrying combined active and passive sensors, equipped with Real-Time Kinematic (RTK) GNSS. However, only a few recent studies have investigated the performance of such commercial RPAS (e.g. Diara et al., 2022; Kersten et al., 2022; Štroner et al., 2023; Teppati Losè et al., 2023). The EuroSDR RPAS Benchmark thereby provided the opportunity to make available to the research community standard datasets captured by the most recent RPAS technology, following best practice in Structure-from-Motion (SfM) photogrammetric capture and assessment (James et al., 2019).

1.2 Benchmark aim and objectives

The primary focus of the benchmark was to achieve a fully independent assessment of the geometric quality of RPAS survey data generated in the absence of ground control and local GNSS base station information. The study evaluated and compared resulting precisions and accuracies obtained with the popular DJI P4 RTK and latest (in August 2021) DJI P1 and DJI L1 sensors, as well as the Riegl MiniVUX lidar. Acquired data was released in stages to registered benchmark participants, with more ground support information available at each stage, who then submitted independent solutions without access to the ground truth used in evaluation. This paper describes the full dataset and summarises the independently processed solutions that were submitted to each of the three phases of the benchmark, reporting the resulting quality of the survey data produced.

2. METHODS

2.1 Study site establishment

2.1.1 Study site overview: A coordinated test field of GCPs, Check Points (CPs), test objects and profiles was established at Wards Hill quarry in Northumberland, UK, in August 2021. The design of the test field was guided by an advisory group of

* Corresponding author

NMCA and academic experts (see acknowledgements). The quarry that was used was actively producing limestone in the 1920s, but the site is now privately owned and used primarily for livestock grazing. The study site has an extent of 350 x 250 m and a ~40 m ground lowering from where limestone was quarried. The site is mainly vegetated with coarse grass, as well as a few shrubs and small trees, and also has an asphalt road running along its length (Figure 1).

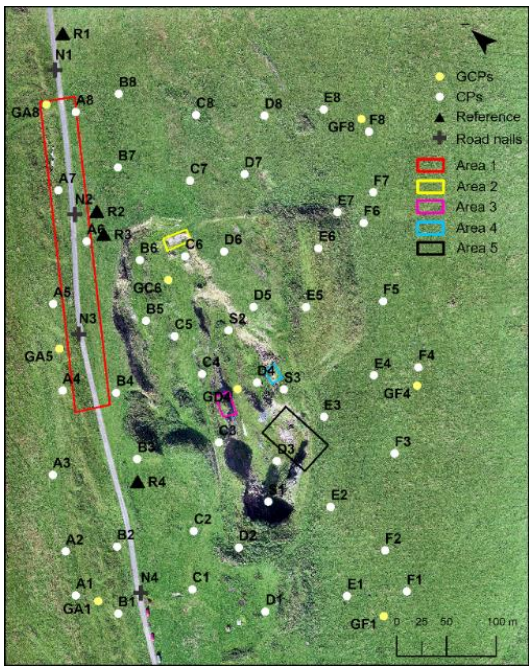

Figure 1. Test field target distribution at Wards Hill quarry, Northumberland, UK. Map data: DJI P4 RTK orthomosaic.

2.1.2 Primary evaluation - 3d test field: The test field established for primary evaluation consisted of 51 CP and eight GCP targets, as shown in Figure 1. The CP test field approximates to a 6 (labelled A-F) x 8 (labelled 1-8) array, with CPs randomly placed in each grid square and identified using an alphanumeric code according to their position in the field (e.g. A3). Three "Supplementary" targets were added to densify the test field in the base of the quarry and were labelled S1, S2 and S3. The eight GCPs were labelled with a prefix of G, followed by their position in the 6 x 8 array – e.g. GA1, GA5, GA8, etc. (e.g. in Figure 2b).

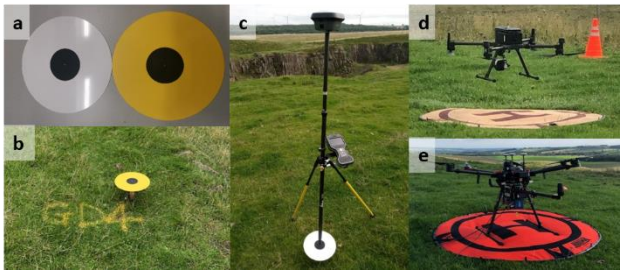

Figure 2. Circular targets, a) indoors and b) in the field; c) GNSS surveying using a bipod on a CP; d) DJI P1 mounted on a DJI Matrice 300; e) Riegl MiniVUX on a DJI Matrice 600.

All targets consisted of circular Perspex disks of 300 mm diameter with a 100 mm diameter black centre to assist with pointing. CP targets were fabricated in white and GCP targets in yellow (Figure 2a). In order to aid identification in RPAS lidar datasets, and to keep above low-lying vegetation, targets were mounted in the field on wooden stakes that were driven vertically into the ground and secured with a single screw through the centre of the target (Figure 2b). A spirit level was used to approximately level the surface of each target during establishment. The height of the target planes above ground level ranged from ~0.15 m to ~0.45 m (Figure 2b). Targets were marked with both a white A4 clipboard ID and a red / white ranging rod (which may have been vertical or horizontal at the time of data acquisition) to support target identification during processing and location of the targets in the field, respectively.

2.1.3 GNSS survey and temporal validation: Four GNSS reference stations were established at the study site (R1-4 in Figure 1) and surveyed during a field campaign that took place over a period from 22 to 26 August 2021. Stations were coordinated using static GNSS, delivering sub-cm level 3D accuracy relative to a local UK network base station in Ordnance Survey Great Britain 36 (OSGB36) and Ordnance Datum Newlyn (ODN). OSGB36 was chosen as it is the coordinate system most often utilised in surveys in Great Britain and specifically introduces "real-world" challenges of coordinate transformations. Details on the ETRS89 to OSGB36/ODN transformation can be found in Ordnance Survey (2020).

The GCPs and CPs were then coordinated with respect to the established GNSS reference stations, using 3 minutes of GNSS data per point with Leica GS18 receivers (mounted on a 1.8 m bipod) operating in "static and kinematic" mode, with 1 second data collected continuously at each point, and as the receivers moved from point to point. For GCPs/CPs with reduced sky visibility in the quarry, 5 minutes of GNSS data were collected per point. This process was undertaken four times and the data processed using Leica Infinity 3.5.0 software. The average standard deviations of all calculated targets' coordinates across the four occupations were 5.3 mm in Easting, 6.1 mm in Northing and 3.4 mm in height. GNSS heights were also independently validated by spirit levelling.

2.1.4 Secondary evaluation - areas and profiles: Five areas were defined to compare requested submission outputs (i.e. point clouds, Digital Elevation Models (DEMs) and orthomosaics), as shown in Figure 1. Area 1 covered a section of the road southwest of the quarry, and Areas 2-5 covered different segments of the quarry such as bare rock, low grass, flat and inclined slopes.

2.1.5 Terrestrial laser scanning survey: A terrestrial laser scanning (TLS) survey of the quarry was carried out using a Leica RTC360 instrument in parallel to the GNSS survey. Four CPs (B6, C3, D3 and S2) were used as control, upon which black and white Leica scanner targets were set up to allow georeferencing. 27 scans were co-registered, adopting a scan-to-scan bundle adjustment within the Leica Cyclone Register 360 software. The resultant co-registered point cloud was then georeferenced using the four targets, resulting in an absolute mean error of 9 mm with a minimum and a maximum error of 1 mm and 21 mm, respectively.

The co-registered point cloud was further processed to filter out noise (i.e. points above or below the terrain) so that it could be used as a reference dataset for assessment of the RPAS outputs over the five test areas (Figure 1). The coordinates of 11 circular targets located within the quarry were also extracted from the TLS point cloud, by precisely selecting the centre of each circle using the "pick-point" Cyclone tool. The TLS-extracted target coordinates were compared against the GNSS coordinates, with average difference standard deviations of 15 mm, 8 mm, and 12 mm in Easting, Northing and height, respectively.

2.1.6 Total station profiles: Four nails were established on the road (N1-N4 in Figure 1) and their coordinates calculated with respect to the GNSS reference stations, as explained in Section 2.1.3. A total station survey of the road was then conducted using a Leica TS16 instrument successively setup on each of the four road nails. A single road profile was generated, consisting of 245 surveying points and used to evaluate the vertical accuracy of the submitted DEMs.

2.2 RPAS sensor data capture and pre-processing

The study site was surveyed using five different RPAS mounted instruments, with operators asked to perform the survey as they would do in commercial operation. To enforce this criteria, each was limited to a single survey flight to represent "real-world" simulation, rather than any kind of "perfect" network configuration that would not be practicable in commercial practice. The RPAS flights captured three photogrammetric and three lidar datasets using the DJI P4 RTK, DJI P1, DJI L1, Routescene Velodyne Lidarpod, and Riegl MiniVUX. The lidar dataset captured with the Routescene Velodyne Lidarpod (Routescene Lidar, 2022) was not subsequently processed by any independent benchmark participant and so is excluded from results presented here. However, the corresponding dataset can be found in EuroSDR (2023).

2.2.1 Photo 1 - DJI P4 RTK: Two DJI Phantom P4 RTK flights were conducted on 25 August 2021 by Newcastle University, one with RTK link disabled, used in Phase 1 of the Benchmark, and a second dataset with RTK enabled, used in Phases 2 and 3. A total of 430 nadir-looking images were collected from 11 parallel lines per flight, plus 59 and 70 oblique images with a 20º off-nadir angle for the first and second flight, respectively. The DJI P4 RTK carries a DJI FC6310R camera with a nominal 8.8 mm focal length lens and a 1" CMOS 20 megapixel sensor with nominal 2.41 x 2.41 µm pixel size. During both flights, the DJI FC6310R camera was set up with a fixed exposure time and focus set to infinity, capturing images with an 80% forward and 70% lateral overlap. The DJI P4 RTK flew at a height of 55 m above ground level with aircraft speed of 4.3 m/s.

2.2.2 Photo 2 - DJI P1: A DJI Zenmuse P1 dataset was acquired using a DJI Matrice 300 RPAS platform, as shown in Figure 2d. The DJI Zenmuse P1 consists of a 45 megapixel full-frame sensor (35.9 x 24.0 mm), with 4.4 µm nominal pixel size. The DJI Zenmuse P1 sensor is mounted on a 3-axis stabilised gimbaled system, which is fixed on the DJI Matrice 300 RTK RPAS platform. The DJI Zenmuse P1 flight was conducted by commercial company Heliguy on 23 August 2021, at a height of 50 m above the ground and aircraft speed of 5 m/s. The DJI Zenmuse P1 sensor was set up with an automatic camera exposure and a continuous focus, capturing images with 80% forward and 70% lateral overlap. During the single flight, a total of 974 nadir-looking images were collected from 24 parallel flight lines and 25 oblique images were captured with a 45º off-nadir angle, while the commercial RTK link was disabled.

2.2.3 Photo 3 / Lidar 1 - DJI L1: Two DJI L1 flights were conducted by Heliguy using a DJI Matrice 300 RPAS platform on 23 August 2021. Each flight consisted of 13 parallel flight lines with a 50% lateral overlap at a 50 m flying height above ground. The DJI L1 carries a DJI Zenmuse L1 sensor which includes a Livox Lidar scanner with GNSS and inertial navigation system (INS), together with a 1" CMOS 20 megapixel EP800 camera of nominal focal length 8.8 mm and 2.41 x 2.41 µm nominal pixel size. During the flights, the Livox Lidar module scanned the area at a 160 kHz sampling rate in a repetitive scanning pattern, supporting a maximum of 3 echo returns. A figure of eight pattern was adopted at commencement of the flight for IMU calibration. A commercial RTK link was disabled during the first flight (for use in Phase 1) and subsequently enabled during the second flight (for Phases 2 and 3). A total of 254 and 252 optical images were acquired during the two flights, alongside raw laser scanning data with corresponding calibration files and GNSS observations.

2.2.4 Lidar 3 - Riegl MiniVUX: A single Riegl MiniVUX flight was carried out with a DJI Matrice 600 RPAS platform on 24th August 2021 by Newcastle University (Figure 2e). The Riegl MiniVUX system comprised two GNSS NovAtel receivers/antennas, one IMU SPAN Airborne EKINOX by SBG Systems, and a Riegl MiniVUX-1 lidar unit. Regarding the flight mission, a 5-minute IMU initialisation was first conducted and then a 14-minute lidar data acquisition ensued. During the IMU initialisation stage, figures of eight and other manoeuvres were performed at varying altitudes of 20-120 m above the ground with aircraft speeds varying from 1 m/s to 17 m/s. During data acquisition, the Riegl MiniVUX flew at 60 m above the take-off location at 6 m/s, following five parallel lines and two cross strips, scanning with a 280 m swath width at 20% reflectance.

2.3 Benchmark procedure

2.3.1 Data release arrangements: In October 2021, an open call for expressions of interest was made via the EuroSDR RPAS Benchmark webpage (Geospatial.github, 2021). Raw data from each survey then became progressively available to registered participants in three phases, over the course of a six-month period. At the end of each phase, participants were asked to submit 3D coordinates for the CPs and a brief description of their chosen processing pipeline. The participants also had the option to submit other geospatial products from the processing pipeline, i.e.: camera and lidar orientation data, dense point clouds, DEMs and orthophotos for the five defined areas shown in Figure 1.

2.3.2 Data processing phases: Phase 1 (no ground control) referred to processing without GCPs, GNSS base station or on-the-fly RTK corrections (e.g. using commercial RTK service). Phase 1 data were released in December 2021. Phase 2 (commercial RTK or local base) referred to processing without GCPs but with a local GNSS base station and/or on-the-fly RTK corrections. Phase 2 data were released in March 2022 including the observations of the R1 GNSS base station. Phase 3 (with full ground control) referred to processing with GCPs, and/or a local GNSS base station and/or on-the-fly RTK corrections, were released in May 2022. Table 1 lists the datasets per phase, highlighting when a GNSS base station or RTK correction was included.

Phase	New dataset released	RTK corr'n	GNSS base	GCP coords
1	Photo 1 - DJI P4 RTK	-	-	-
	Photo 2 - DJI P1	-	-	-
	Photo 3 / Lidar 1 - DJI L1	-	-	-
	Lidar 2 - RS Velodyne	-	-	-
	Lidar 3 - Riegl MiniVUX	-	-	-
2	Photo 1 - DJI P4 RTK	✓	✓	-
	Photo 3 / Lidar 1 - DJI L1	✓	✓	-
	Lidar 2 - RS Velodyne	✓	✓	-
3	3D coordinates of GCPs	-	-	✓

Table 1. Phased dataset release (available via EuroSDR, 2023).

2.3.3 Host analysis procedures: Newcastle University processed all datasets for all different phases. Photo 1 - DJI P4 RTK, Photo 2 - DJI P1 and Photo 3 - DJI L1 datasets were processed using the Metashape SfM photogrammetric software package, adopting the SfM self-calibrating bundle adjustment pipeline (James et al., 2019). Newcastle also processed the Lidar 1 - DJI L1 and Lidar 3 - Riegl MiniVUX using DJI Terra and MMProcess, respectively.

Image orientation for the photogrammetric datasets at Phase 1 relied on single GNSS receiver positioning of the camera exposure stations, as there were no on-the-fly commercial RTK corrections or GNSS base station data (Table 1). In Phase 2, image orientation relied on the RTK corrected camera exposure stations for all photogrammetric datasets. In Phase 3, all photogrammetric datasets were processed using GCPs (Figure 1) to constrain the SfM self-calibrating bundle adjustment pipeline. To achieve consistency in processing during all phases, as well as achieving an overall practicable processing time, Metashape settings were standardised, as per Table 2. Detailed information about the Metashape software settings can be found in Metashape, (2023).

Alignment parameters:	
Accuracy	High
Generic preselection	No
Reference preselection	No
Key point limit	50,000
Tie point limit	0
Optimisation parameters:	
Camera distortion parameters	f, cx, cy, k1-k3, p1, p2
Adaptive camera model	No
Dense point cloud parameters:	
Quality	Medium
Depth filtering	Mild

Table 2. Metashape settings used in Newcastle University's photogrammetric processing.

After image orientation and dense point cloud reconstruction, DEMs were extracted over the five predefined areas (Figure 1). The resulting ground sampling distance (GSD) of the photogrammetric blocks for the Photo 1 - DJI P4 RTK, Photo 2 - DJI P1 and Photo 3 - DJI L1 datasets was 0.020 m, 0.007 m and 0.018 m, respectively.

Regarding lidar processing, the Lidar 1 - DJI L1 dataset was processed in DJI Terra version 3.8.0 software, which can simultaneously handle imagery and laser scanning data using RTK on-the-fly corrections. The processing pipeline adopted is explained in DJI Terra (2023). This software was also used to generate a DEM with a 0.05 m spatial resolution. This current version of DJI Terra does not allow incorporation of GCPs, hence this dataset was only analysed in Phase 2. The Lidar 3 - Riegl MiniVUX dataset was processed for Phase 2 using the Post Processing Exporter v.1.2.3791, Way Point Inertial Explorer v.8.70. and MMProcess v.17.10.0.0 software. The first software package was used to convert the raw IMU data into readable formats. The second software package was used to post-process the RPAS trajectory with respect to the local R1 GNSS base station (see Figure 1), adopting a tightly IMU-GNSS coupled differential processing pipeline. The third software package was used to produce the resulting trajectory and point cloud into the OSGB36/ODN coordinate system. Further processing, including flight strip adjustment to minimise mismatches and point cloud cleaning to remove noise, was conducted in Terrasolid software with TerraScan and TerraMatch libraries. The seven flight lines were aligned with an average magnitude of 0.045 m in the Z-axis. For more information about the processing pipeline, refer to Davidson et al. (2019). A DEM of 0.125 m spatial resolution was also generated from this dataset using the OPALS software (Pfeifer et al., 2014).

Newcastle University subsequently conducted quality assessments for the independent submissions, using the GNSS surveyed CPs, the total station road profile, and the TLS point cloud over the five predefined areas as reference datasets. Statistics, including average standard deviations and root mean square errors (RMSEs) in Easting, Northing and height were calculated between the GNSS surveyed CPs and estimated coordinates of the CPs from the various photogrammetric products. Height values were extracted from the various reconstructed DEMs (provided by the participants) along the road profile to compare with the total station generated profile for vertical accuracy evaluation. To directly compare the various submissions against the TLS product over the predefined areas, the multi-scale model-to-model cloud comparison (M3C2; Lague et al. (2013)) was applied using Cloud Compare.

Table 3 shows indicative examples of the varying point cloud densities between the five datasets, as estimated over flat terrain on asphalt in Area 1 and on bare rock in Area 5 (Figure 1) where no vegetation is present, using TerraScan in MicroStation. Note, as TLS was only acquired over the quarry, there was no point cloud available in Area 1. As ground-based observations, TLS provided the highest point cloud density, with the Lidar 3 - Riegl MiniVUX dataset providing the lowest. The Photo 2 - DJI P1, with a full frame sensor, resulted in a slightly higher point cloud resolution than the Lidar 1 - DJI L1. The resulting point cloud resolution of the Lidar 3 - Riegl MiniVUX survey proved inadequate to precisely identify the GCPs (<10 points over the 300 mm diameter of the b/w circular targets), hence this dataset was only analysed in Phase 2, using corrections from the R1 GNSS base station.

Point cloud density [pts/m^2]					
Area	TLS	Photo 1 DJI P4 RTK	Photo 2 DJI P1	Lidar 1 DJI L1	Lidar 3 Riegl MiniVUX
1	-	200	1310	1020	120
5	150k	200	1320	1120	100

Table 3. Average point cloud densities for the photogrammetric and lidar datasets processed by Newcastle University.

3. RESULTS

3.1 Summary of submissions received

In total, an encouraging 48 expressions of interest were submitted from applicants wishing to participate in the Benchmark. This comprised 25 different requests from academia and 23 from industry. Download numbers were also significant, but solutions were ultimately submitted by only six independent researchers, (four academic, two industry) plus those from Newcastle University. In total, 41 different usable solutions were received and analysed across all three benchmark phases. Figure 3 shows the number of submissions made at each phase for each dataset – only Newcastle University provided lidar solutions. Independent submissions included the use of six different commercial SfM software packages, as listed in Table 4. It should be noted that the same software package was often used multiple times, either because it was submitted by a different participant or was applied using different settings.

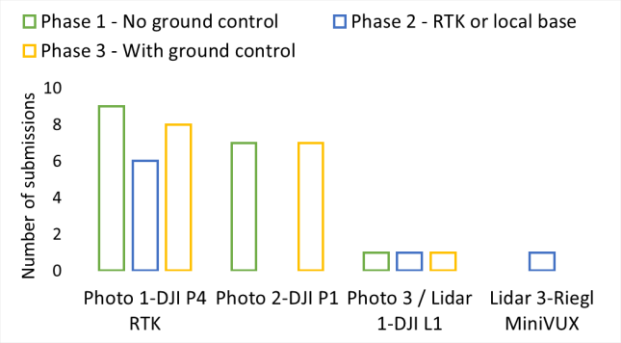

Figure 3. Independent dataset submissions at each phase.

Data sets	Software	Label numbers / Phases		
		1	2	3
DJI P4 RTK	Pix4D	1,3,5	10	
	RTKL & M'shape	2	11	16
	Metashape	6,8,9*	12	21º, 22^, 23*
	Reality Capture	4	13	19º, 20^
	Context Capture	7	14	17º, 18^
	GIQ & M'shape	-	15*	
DJI P1	Pix4D & 3DFZ	24	-	
	Pix4D	26	-	
	Metashape	27,29,30*	-	35º, 36^, 37*
	Reality Capture	25	-	33º, 34^
	Context Capture	28	-	31º, 32^
DJI L1	Metashape	38*	39*	40*
	GIQ & M'shape		41*	

Table 4. Label numbers for independent results achieved using the three photogrammetric datasets at each phase (GIQ: Grid InQuest; RTKL: RTKLib; 3DFZ: 3DF Zephyr). * Refers to a Newcastle University processed result; º Refers to exterior and interior orientation free solution; ^ Refers to exterior and interior orientation fixed solution.

For example, in Phase 3, Context Capture, Reality Capture and Metashape were each used twice by the same participant. Firstly, these software were used when both exterior and interior orientation (EO/IO) camera parameters were optimised using GCPs during the SfM self-calibrating bundle adjustment, with the EO standard error being altered to the lowest weight of EO estimates. This corresponds to a free image orientation solution and is symbolised with º in Table 4 (e.g. labelled results 17, 19, 21 in DJI P4 RTK - Phase 3). Secondly, the same software was used when only EO camera parameters were optimised using GCPs during the SfM self-calibrating bundle adjustment while IO camera parameters were kept fixed, i.e. pre-calibrated values were as per recorded EXIF image attributes. This corresponds to a fixed image orientation solution and is symbolised with ^ in Table 4 (e.g. labelled results 32, 34, 36 in DJI P1 - Phase 3).

Compared to the Newcastle processed solution, participants included only nadir-looking images, and did not always describe which exact images had been included during processing. In Phase 1, one participant used the RTKLib software package to refine the coordinates of the RPAS camera exposure stations, but no details of the processing steps were provided (labelled result 2 in Table 4 for the DJI P4 RTK). Another participant used the 3DF Zephyr software for image orientation in Phase 1 for the DJI P1 dataset (result 24 in Table 4). The same participant also utilised a 1 m spatial resolution DEM, produced by England's Environment Agency in 2020, before extracting the 3D coordinates of CPs.

3.2 Evaluation of results against independent CPs

3.2.1 Photogrammetric datasets - DJI P4 RTK: Results from 23 independent solutions (CP coordinates) for all three phases for the DJI P4 RTK dataset are shown in Figure 4.

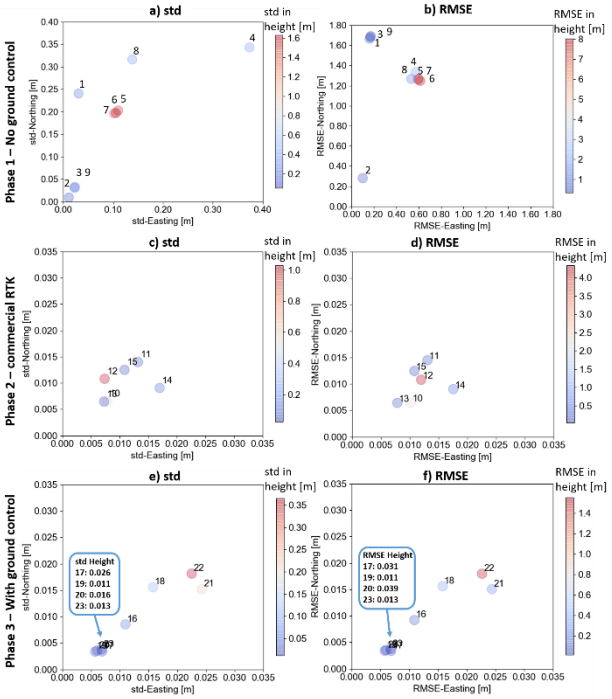

Figure 4. Photo 1 - DJI P4 RTK results. Scales in axes vary. Each point corresponds to a labelled solution, as per Table 4. The three rows refer to the three phases with the first column representing the standard deviations and the second column the RMSEs calculated against CPs. The colour scheme indicates the corresponding errors in height (different scale per graph).

Care should be taken when interpreting such global statistics, but in general, in Phase 1 (without any ground control or local base station) standard deviation values are significantly lower than RMSE figures, indicating relatively good precision but poor accuracy. The largest error observed in height (max $RMSE_H$ = 8.000 m, max std_H = 1.628 m) was from the solutions obtained in (5) Pix4D, (6) Metashape and (7) Context Capture. Solution 2, obtained using RTKLib and Metashape, provided the lowest error magnitudes in all three axes with $RMSE_E$ = 0.100 m $RMSE_N$ = 0.280 m and $RMSE_H$ = 0.053 m (nb GSD is 0.020 m). This is a remarkable result that is more comparable with those achieved in Phase 2. Due to the lack of a detailed description provided, it is impossible to fully explain the result other than to speculate it has been achieved via long baseline GNSS processing in RTKLib (similar to using RTK commercial correction). Excluding this result, RMSEs in plan are within a more typical 1.374 – 1.700 m range that might be expected when not using a GNSS base station / correction service (and/or GCPs).

RMSE magnitude from Phase 1 to Phase 2 improved from metre- to centimetre-level due to the introduction of commercial RTK on-the-fly corrections. Planimetric precisions and accuracies in Phase 2 are similar – the standard deviations (Figure 4c) and RMSEs (Figure 4d) in plan were within the range of 0.010-0.019 m for all solutions, less than 1x GSD. However in height, solution 12 from Metashape delivered m-level errors (std_H = 1.030 m and $RMSE_H$ = 4.332 m). To interpret this outcome, Phase 2 datasets were re-processed by Newcastle University

using only Metashape, which resulted in decimetre-level RMSEs when applying the coordinate system conversion in Metashape directly (0.750 m, 0.380 m and 0.815 m in Easting, Northing and height, respectively). However, by converting the RPAS camera exposure stations to OSGB36 in Grid InQuest (Ordnance Survey, 2023) before applying the SfM pipeline in Metashape (solution 15 in Table 4), cm-level RMSEs at CPs were delivered (0.011 m, 0.012 m and 0.023 m in Easting, Northing and height, respectively; Figure 4d).

In Phase 3 (Figure 4e and f), standard error and RMSE values are again very similar, the eight solutions all providing planimetric errors lower than 0.022 m – approximately equivalent to 1x GSD – but no improvement is seen in planimetry over Phase 2. However, height solutions are improved slightly in Phase 3, with the exception of an outlier solution (22) that returned a 1.550 m RMSE in height. This is attributed to the fact that IO camera parameters were kept fixed and not optimised, which is not recommended when GCPs are included in the SfM photogrammetry pipeline (James et al., 2019).

3.2.2 Photogrammetric datasets - DJI P1: Regarding assessment of the DJI P1 dataset at independent CPs for Phase 1, solutions 25 and 30 provided the highest standard deviations and RMSEs in height, respectively (std_H = 0.687 m and $RMSE_H$ = 1.860 m; Figure 5a and b). Solution 24 (Table 4) delivered a 0.043 m standard deviation and a 0.769 m RMSE in height, which was the best result achieved. It is possible that the inclusion of the Environment Agency DEM before extracting the CP coordinates improved the errors in height.

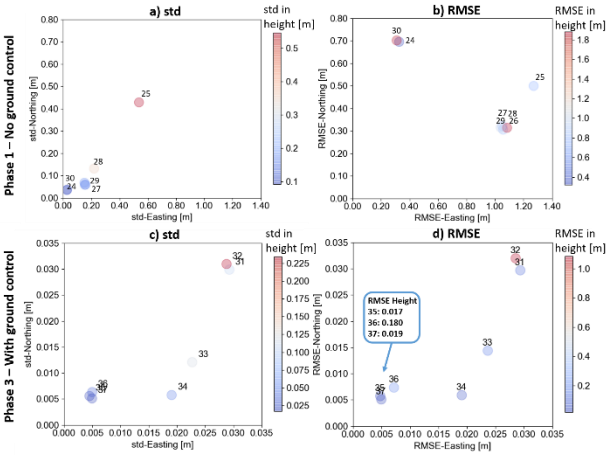

Figure 5. Photo 2 - DJI P1 a), c) standard deviations and b), d) RMSEs at CPs of labelled independent results, as listed in Table 4, for Phases 1 and 3. Scales in plan and height vary per phase.

With respect to planimetric errors, maximum discrepancies of 0.598 m and 0.641 m were observed in standard deviations and RMSEs, respectively, among all solutions. In Phase 3, all DJI P1 solutions provided standard deviations and RMSEs in plan within a range corresponding to 1-6x GSD (Figure 5c and d). The least accurate results in plan and height were obtained with Context Capture (solution 32) with IO camera parameters not being optimised during SfM bundle adjustment.

3.2.3 Photogrammetric datasets - DJI L1: DJI L1 photogrammetric results were generated only by Newcastle University (Table 5) and resulted in errors with a similar level of magnitude to those of the DJI P1 results in Phases 1 and 3 (Figure 5). As observed for the DJI P4 RTK in Phase 2 (including RTK corrections), the coordinate system transformation with Grid InQuest significantly improved the RMSEs for the DJI L1 dataset, from dm to cm-level (Table 5). This is obvious in Figures 6a and b that show the local spatial error distribution at CPs.

Phase	Standard deviation [m]			RMSE [m]		
	E	N	H	E	N	H
1	0.113	0.093	0.142	0.374	0.916	1.345
2	0.009	0.009	0.035	0.745	0.389	1.033
2*	0.009	0.009	0.035	0.005	0.123	0.008
3	0.008	0.007	0.025	0.009	0.007	0.028

Table 5. Photo 3 - DJI L1 errors estimated at CPs (* is Grid InQuest & Metashape solution; GSD is 0.018 m).

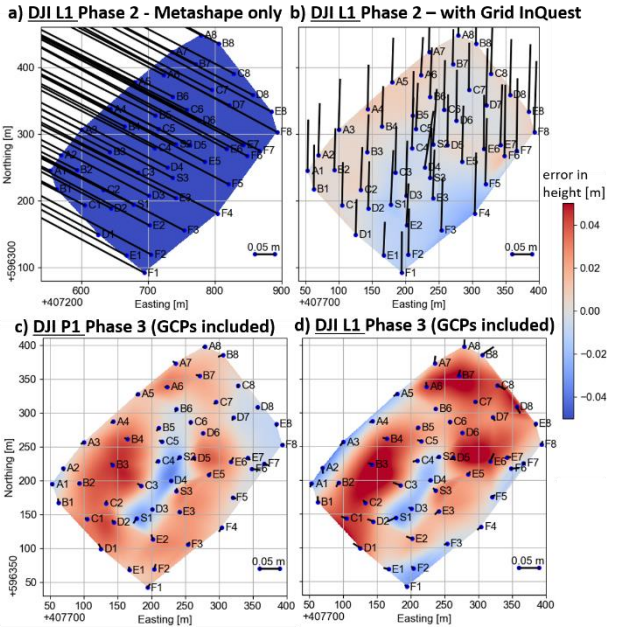

Figure 6. Spatial error distribution at CPs obtained via photogrammetry: a) and b) DJI L1 in Phase 2, c) DJI P1 and d) DJI L1 in Phase 3.

A systematic error bias with a North-South direction in Phase 2 (Figure 6b), was mostly removed with the inclusion of GCPs in Phase 3 (Figure 6d). However, Figures 6c and 6d show that even with the inclusion of GCPs there was evidence of residual systematic error in height, coincident with terrain morphology, with maximum values of 0.044 m and 0.053 m for the DJI P1 and L1, respectively. In the DJI L1 solution, the planimetric error vectors at A8 and B8 CPs (Figure 6d) are larger than those in the DJI P1 solution (Figure 6c), because only two images were captured over those peripheral targets.

3.3 Validation with TLS

Further photogrammetric evaluations were performed against the TLS survey data for those results that provided the lowest errors in the previous independent CP assessment: DJI P4 RTK result 19 and DJI P1 result 35, both from Phase 3, as per Table 4. In Phase 3 (Figure 7b), the photogrammetric outputs showed a small spread with mean / median values close to zero in all areas. The DJI P1 output showed the smaller discrepancies against the TLS data.

Further assessments include the two lidar datasets processed by Newcastle University. In evaluation of M3C2 differences with respect to TLS, the DJI L1 output showed greater skewness with

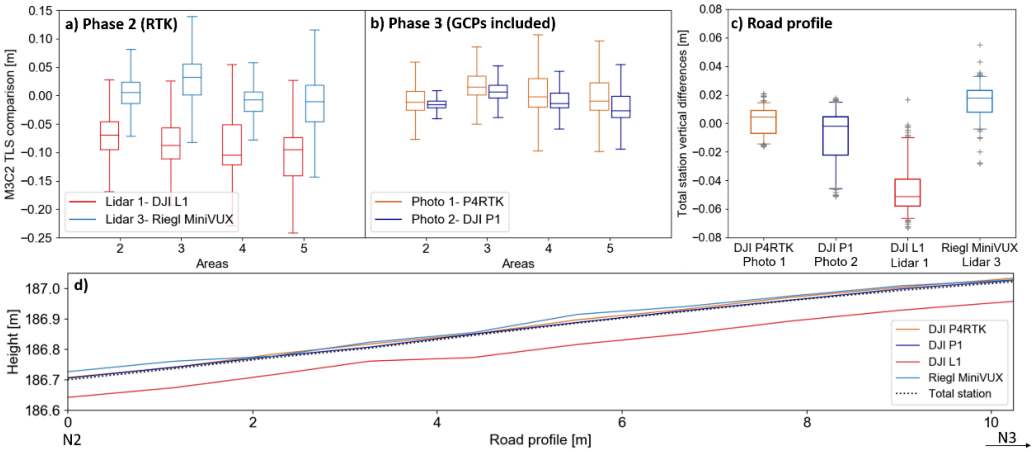

Figure 7. M3C2 TLS comparison of the a) lidar and b) photogrammetric datasets in Phase 2 and 3, respectively; c) Road profile box plots of the vertical differences from total station measurements; d) Profile of the first 10 m between road nails N2-N3 (Figure 1).

wider distributions in comparison to the Lidar 3 – Riegl MiniVUX output at the four predefined areas (see Figure 7a). Mean values ranged from -0.110 m to -0.066 m for Lidar 1, and -0.026 m to 0.005 m for the Lidar 3 dataset. Errors as large as -0.250 m were observed for Lidar 1 at Area 5. It should be noted that outliers, mainly attributed to low vegetation/grass, were excluded from the box plots in Figure 7a and b. An example of M3C2 differences is depicted in Figure 8a, b and c with the red scale colours indicating a maximum of 0.15 m due to differences over grass.

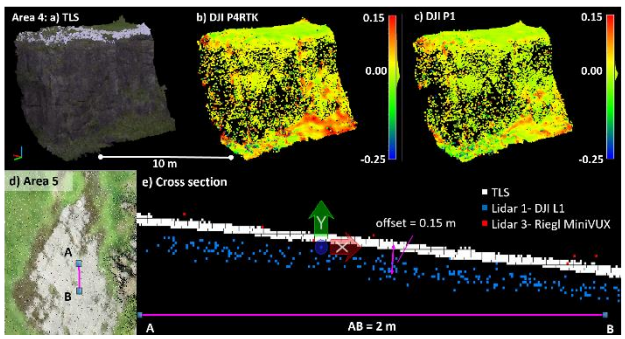

Figure 8. Area 4 M3C2 comparisons for a) TLS survey against b) DJI P4 RTK and c) DJI P1 in Phase 3; Cross-section comparison in d) Area 5 among e) all lidar datasets in Phase 2.

3.4 Validation with total station profiles

Regarding vertical validation over the road profile (between nails N1-N4 within Area 1, Figure 1), the DJI P4 RTK provided the smallest discrepancies in height against the total station observations (Figure 7c). The DJI P1 showed wider variability (ranging between -0.046 m to 0.015 m) and followed a slightly skewed distribution. Regarding the lidar datasets, mean values of -0.046 m and 0.016 m were calculated for the DJI L1 and Riegl MiniVUX, respectively. It should be noted that the outliers shown in Figure 7c were mainly observed at the borders of the submitted DEMs in Area 1. The largest discrepancies were identified in data from the DJI L1, which are also evidenced in Figure 7d. Here a systematic vertical shift below the total station height measurements is clearly observed along the length of the road profile. Similarly at a cross-section over Area 5 (Figure 8d and e), the Lidar 1 - DJI L1 point cloud was found to be 0.15 m below the TLS and the Riegl MiniVUX point clouds. This offset is also evidenced in the box plots in Figure 7a for Area 5.

4. DISCUSSION

In relation to resultant data quality, the DJI P4 RTK was found to deliver planimetric accuracy that approximated to 1x GSD (c. 0.020 m) in both Phases 2 (RTK corrections included) and 3 (GCPs included), similar to that obtained with the DJI P1 (GSD = 0.007 m). At Phase 1, the DJI P1 provided dm level accuracies in both plan and height, compared to the DJI P4 RTK m- level accuracy, with a few outliers. The largest errors in Phase 3 were observed when IO camera parameters were not optimised in the SfM bundle adjustment. Another factor that appeared to create problems was the OSGB36/ODN coordinate transformation model embedded into “black-box” SfM software packages. It is not known if such models are equivalent to the OSTN15/OSGM15 model used in Ordnance Survey’s official Grid InQuest software package.

Whilst the Riegl MiniVUX provided the lowest point density, it showed smaller errors over the predefined areas in the quarry than the Lidar 1 - DJI L1. It should be noted that the R1 GNSS base station was used to process the Riegl MiniVUX RPAS trajectory, and that may explain why it resulted in better accuracies compared to the commercial RTK solution of the Lidar-1 DJI L1. The DJI Terra software does not incorporate GCPs, hence the DJI L1 output solely relied on the RTK corrections provided by a commercial service. On the other hand, when the DJI L1 was processed solely as a photogrammetric dataset, cm-level accuracy (c. 1x GSD in plan and 1.5x GSD in height) were delivered with the inclusion of GCPs. Overall the cm range of deviations from reference ground-based datasets for both the DJI P1 and L1 were similar to those reported in the recent study Kersten et al. (2022).

5. CONCLUSIONS

This paper has described the dataset and summarised some of the key results and findings of the EuroSDR RPAS benchmark. Although responses to the benchmark were disappointingly few in number, multiple submissions from participants have allowed some interesting conclusions to be drawn. The benchmark has evaluated 41 independent submissions, generated using real-world RPAS survey data, against GNSS surveyed CPs, TLS and total station height observations, all measured on a national coordinate system. In photogrammetric processing, Phase 1, conducted without any ground control or local base station, saw submissions deliver c. m-level solutions where accuracy was

much worse than precision. One independent solution highlighted the potential of using GNSS processing techniques to improve results that otherwise would not be possible without a local GNSS base station or RTK correction (i.e. Phase 2). In Phase 2, with RTK GNSS correction available, accuracy and precision were seen to be more equal, with solutions at the cm-level, but prone to datum effects in both plan and height that, without GCPs (i.e. Phase 3) might not be evident in a real-world survey. Finally, in Phase 3, the introduction of GCPs returns the highest level of precision and accuracy (~1x GSD), albeit only slightly improved over Phase 2. Outliers occurred at all three stages, highlighting the continued importance of following an appropriate processing flow line with camera calibration, coordinate transformations, etc., particularly in the modern era of "black box" SfM photogrammetry. The inclusion of GCPs still aids in the identification of such errors, so contributing to the reliability of such surveys. Finally, in the preliminary assessment of the two RPAS lidar datasets, the Riegl MiniVUX using a local GNSS base was found to outperform the DJI L1-RTK based outputs, despite its coarser spatial resolution.

To facilitate and encourage wider use of the EuroSDR RPAS Benchmark, the full dataset has now been made open access via Zenodo (10.5281/zenodo.8309679). It is anticipated that users will find additional uses for the open RPAS datasets, e.g. for tie point extraction studies (see Peppa et al., 2022). Researchers using the dataset for their own research are kindly asked to cite this ISPRS Archives paper by way of acknowledgement.

ACKNOWLEDGEMENTS

The authors acknowledge EuroSDR for supporting and funding the reported RPAS Benchmark. Thanks are due to DJI, Heliguy and Routescene for the provision of the RPAS datasets flown for the study. Equipment used by Newcastle University in the benchmark field campaign was funded by UKCRIC - UK Collaboratorium for Research in Infrastructure & Cities: Newcastle Laboratories (EPSRC award EP/R010102/1). Craig Robson and Nathan Harrap, both of Newcastle University, contributed to the fieldwork campaign. The Advisory Group that kindly gave their time to design the study comprised: Michael Cramer (Univ. of Stuttgart, Germany); Matt Farthing (Ordnance Survey, UK); Eija Honkavaara (FGI, Finland); Michael James (Lancaster Univ., UK); Christian Malmquist (Kartverket, Norway); Gottfried Mandlburger (TU Wien, Austria); Heather Reese (Univ. of Gothenberg, Sweden); Fabio Remondino (FBK Trento, Italy). Finally, and most importantly, thanks to the independent researchers who submitted results to the EuroSDR RPAS Benchmark: Ali Abdollahi (Univ. of Tehran, Iran); Thodoris Betsas (National TU of Athens, Greece); Gustav Fiskum (Norconsult Informasjonssystemer, Norway); Jon Horgan (Ordnance Survey, UK); Valeria-Ersilia Oniga (TU Iasi, Romania); Oswaldo A. Sequera (TU Dublin, Ireland).

REFERENCES

Bakuła, K., Mills, J. P., Remondino, F., 2019. A review of benchmarking in photogrammetry and remote sensing. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 42 (Part 1/W2), 1-8.

Davidson, L., Mills, J. P., Haynes, I., Augarde, C., Bryan, P., and Douglas, M., 2019. Airborne to UAS Lidar: An analysis of UAS Lidar ground control targets. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-2/W13, 255-262.

Diara, F., Roggero, M. 2022. Quality Assessment of DJI Zenmuse L1 and P1 LiDAR and Photogrammetric Systems: Metric and Statistics Analysis with the Integration of Trimble SX10 Data. Geomatics, 2(3), 254-281.

DJI Terra, 2023. https://www.heliguy.com/blogs/posts/l1-lidar-drone-data-how-to-post-process-point-clouds-dji-terra, last accessed 11/09/23.

EuroSDR, 2023. EuroSDR RPAS benchmark datasets available in 10.5281/zenodo.8309679

Geospatial.github, 2021. https://geospatialncl.github.io/eurosdr-rpas-benchmark/, last accessed 11/09/23.

James, M. R., Chandler, J. H., Eltner, A., Fraser, C., Miller, P. E., Mills, J. P., Noble, T., Robson, S., Lane, S. N., 2019. Guidelines on the use of structure-from-motion photogrammetry in geomorphic research. Earth Surface Processes and Landforms, 44(10), 2081-2084.

Kersten, T., Wolf, J., and Lindstaedt, M., 2022. Investigations into the accuracy of the UAV system DJI Matrice 300 RTK with the sensors Zenmuse P1 and L1 in the Hamburg test field. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B1-2022, 339-346.

Lague D, Brodu N, Leroux J. 2013. Accurate 3D comparison of complex topography with terrestrial laser scanner: application to the Rangitikei canyon (N-Z). ISPRS Journal of Photogrammetry and Remote Sensing 82: 10-26.

Metashape, 2023. User manual professional edition, version 1.5, https://www.agisoft.com/pdf/metashape-pro_1_5_en.pdf, last accessed 10/01/2023.

Nex, F., Armenakis, C., Cramer, M., Cucci, D. A., Gerke, M., Honkavaara, E., Kukko, A., Persello, C., Skaloud, J., 2022. UAV in the advent of the twenties: Where we stand and what is next. ISPRS Journal of Photogrammetry and Remote Sensing, 184, 215-242.

Ordnance Survey, 2023. Grid InQuest: A software package for ETRS89 (WGS84) and OSGB36 National Grid coordinate transformations. https://www.ordnancesurvey.co.uk/business-government/tools-support/os-net/transformation, last accessed 01/05/2023.

Peppa, M. V., Morelli, L., Mills, J. P., Penna, N. T., Remondino, F., 2022. Handcrafted and learning-based tie point features – comparison using the EuroSDR RPAS benchmark dataset. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 43, 1183-1190.

Pfeifer, N, Mandlburger, G., Otepka, J., Karel W. 2014. OPALS - A framework for Airborne Laser Scanning data analysis. Computers, Environment & Urban Systems, 45 (2014), 125-136.

Routescene Lidar, 2022. https://www.routescene.com/lidar-mapping-systems/uav-lidar-systems/, last accessed 11/09/2023.

Štroner, M., Urban, R., Křemen, T., Braun, J., 2023. UAV DTM acquisition in a forested area – comparison of low-cost photogrammetry (DJI Zenmuse P1) and LiDAR solutions (DJI Zenmuse L1). European Journal of Remote Sensing, 56(1), 2179942.

Teppati Losè, L., Chiabrando, F., Maschio, P., 2023. Direct georeferencing approaches for close-range and UAV photogrammetry in the built heritage domain. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 48(M-2-2023), 1557-156.