RESEARCH ARTICLE

ESPL WILEY

On-site geometric calibration of RPAS mounted sensors for SfM photogrammetric geomorphological surveys

Johannes Antenor Senn[1,2] | Jon Mills[1] | Claire L. Walsh[1] | Stephen Addy[2] | Maria-Valasia Peppa[1]

[1]School of Engineering, Newcastle University, Newcastle upon Tyne, UK

[2]James Hutton Institute, Aberdeen, UK

Correspondence
Johannes Antenor Senn, School of Engineering, Newcastle University, Newcastle upon Tyne NE1 7RU, UK.
Email: j.senn@newcastle.ac.uk

Abstract

The application of structure from motion (SfM) photogrammetry for digital elevation model (DEM) and orthophoto generation from visible imagery enjoys ever-growing popularity in geomorphological research. Photogrammetry experts, however, urge that a rigorous approach is a prerequisite for reliable results—a requirement that may conflict with real-world survey. We present a method that unites the two disciplines, using the example of a challenging SfM photogrammetric survey at a Scottish river. Using simultaneous geometric pre-calibration of a multi-sensor remotely piloted aircraft system (RPAS), the method facilitates time-efficient topography mapping and the integration of other wavelengths to create orthophotos providing additional surface information. The approach utilizes an on-site 3D structure—for example, a building, as calibration object, by extracting coordinates of natural features from lidar scans and sensor imagery. We assess the workflow with specialized calibration software (VMS) and widely applied commercial SfM photogrammetric software (AM), using a DJI Phantom optical and a Workswell thermal sensor. We achieved calibration accuracies below one-third (optical) and one-quarter (thermal) of a pixel. Subsequently, we transfer the sensor parameters to pre-calibrate the SfM application and compare the results to a self-calibrated workflow. In a systematic experiment using the optical river survey dataset, we assess the effectiveness of pre-calibration, oblique imagery, scale variation and masking to mitigate systematic DEM errors. Opposing trends show between the calibration strategies. Decreasing network complexity (i.e., flying heights/view angles) improves pre-calibrated but compromises self-calibrated scenarios. Pre-calibrating (VMS) imagery from a single height (30 m nadir) yielded the best results. This finding could have implications for geomorphological surveys, in which single-scale datasets are widespread practice, despite the literature's urge towards more complex imaging networks. The self-calibrated results legitimise this insistence: The same dataset resulted in pronounced dome-shaped DEM distortion, indicating systematic errors, whereas additional flying heights and angles significantly improved the results.

KEYWORDS
geometric calibration, geomorphology, pre-calibration, remote sensing, RPAS, structure from motion photogrammetry

1 | INTRODUCTION

The rise of user-friendly photogrammetric applications (Chandler, 1999) over the last 30 years and the availability of off-the-shelf remotely piloted aircraft systems (RPAS) has led to a democratization of photogrammetry. Integration of computer vision and machine learning into photogrammetric workflows has opened this technology to a wider range of users. What used to be an expensive technology reserved for skilled experts has evolved into a tool requiring little prior knowledge (James et al.2019).

Consequently, structure from motion (SfM) photogrammetry, typically using visible imagery, has become a widely used tool in geomorphological research. Areas of application include measuring mass and soil movements (e.g., Eichel et al., 2020), coastal erosion (e.g., James & Robson, 2012), fluvial morphology (e.g., Mandlburger, 2019; Woodget et al., 2019), aeolian (e.g., Hugenholtz et al., 2013) and glacial processes (e.g., Piermattei et al., 2015), hydro- and morphodynamic modelling (e.g., Entwistle et al., 2018), flow velocimetry (e.g., Eltner et al., 2021) and erosion monitoring (e.g., Eltner et al., 2015). The reasons for its success are manifold and include cost efficiency, repeatability, image resolution and opening impassible or dangerous terrain.

Additionally, geomorphological applications (e.g., Eltner et al., 2021) often adopt several sensors with complementary spectral properties for orthophoto generation to be used for surface mapping and image classification. Thermal imagery can provide valuable information about surface properties but is less suitable for mapping topography because of its low dynamic range and image resolution (Javadnejad et al., 2020; Maes et al., 2017). The different reflective properties result in inconsistent levels of contrast between visible and thermal images: for example, on overgrown areas borders appear less pronounced, making image matching challenging. There are manifold examples for use cases of multi-sensor RPAS: Vegetation indices are used for ecosystem (e.g., Antarctic moss: Lucieer et al., 2012) or crop (e.g., vineyards, beets, forest: Maes et al., 2017; Pádua et al., 2020) monitoring. Erenoglu et al. (2017) used thermal, multispectral and RGB imagery to explore different aspects of a heritage site, and Biass et al. (2019) combined thermal and RGB sensors to monitor volume and temperature of lava flows.

Several recent publications (e.g. Eltner & Sofia, 2020; Peppa et al., 2019; Remondino et al., 2017) have pointed out that a certain level of expertise and process understanding is required to apply SfM photogrammetry appropriately. The ease of access to this technology allows results to be generated quickly and can create the illusion that the produced products are meaningful. In this context, authors such as James et al. (2019) and Remondino et al. (2017) denounce the reliability of some recent studies in the field of geomorphology. As a consequence, recent photogrammetric publications have highlighted the importance of rigorous approaches and data quality assessment. These articles propose guidelines and suggestions for good practice in the application of SfM photogrammetry in geomorphology (e.g., Eltner & Sofia, 2020; James et al., 2019). Survey design and imaging network geometry play a major role in this. Classical photogrammetry from manned aircraft applied pre-calibrated metric sensors with high geometric stability. Nadir imaging—that is, cameras facing vertically downwards, from a uniform flying height—was common practice. However, such a classical approach does not directly translate to working with RPAS. The rigorous application of non-metric, off-the-shelf imaging sensors in modern SfM photogrammetry applications is founded on the incorporation of self-calibrating bundle adjustment (also referred to as 'on-the-job self-calibration'). This process simultaneously determines the internal (lens and sensor geometry) and external (position and orientation) sensor parameters. The predominant approach to determine the external sensor parameters uses ground control points (GCPs)—that is, visible features of known coordinates. Typically, these are placed targets or permanent landmarks for which position is acquired with centimetre accuracy using survey equipment—for example, global navigation satellite system (GNSS) (real-time kinematic (RTK) or post-processed) or total stations. Sufficient number, distribution and accuracy of GCPs are key to mitigate errors in scaling, rotation and translation (Carbonneau & Dietrich, 2017). Alternative approaches apply direct georeferencing using the orientation and corrected (RTK or post-processed) GNSS position of the sensor at the time of image acquisition (Carbonneau & Dietrich, 2017). The bundle adjustment can reach its limits if the geometric networks are insufficiently rigorous—for example, exclusively parallel or nadir view directions and flat surface geometry (Griffiths & Burningham, 2019). Unsatisfactory bundle adjustment has been demonstrated to result in systematic camera calibration errors that cause dome or bowl-shaped digit5al elevation model (DEM) deformations (e.g., James et al., 2020; Sanz-Ablanedo et al., 2020).

When non-metric sensors are applied, a rigorous survey network design can mitigate systematic error. Therefore, data acquisition must be optimized for convergent image geometry by including sufficient overlap and variation in view angle and height to achieve a reliable self-calibrating bundle adjustment (Cramer et al., 2017; Przybilla et al., 2015; Wackrow & Chandler, 2008).

Previous research suggests inclusion of oblique imagery to strengthen the network (e.g., Harwin et al.2015; James & Robson, 2014); however, there are certain application scenarios in which it is not feasible. For example, nadir-only designs are applied in bathymetric through-water SfM photogrammetry to minimize light refraction angles at the water surface (e.g., Javernick et al., 2014; Slocum et al., 2020; Woodget et al., 2015). This optical offset can counteract robust bundle adjustment. The use of nadir-only surveys is ideal for orthophoto generation and has a history in classic airborne photogrammetry (Cramer et al., 2017). Time constraints can be another decisive aspect. Adding oblique imagery to surveys can multiply the flight and processing time (Meinen & Robinson, 2020), making it less economically viable for commercial providers. Some RPAS (e.g., fixed wing) feature permanent sensor mounts and may thus be restricted to nadir surveys. For these reasons, single-scale nadir surveys remain the quasi-standard for environmental mapping (Griffiths & Burningham, 2019).

An alternative strategy, if survey or terrain proves too challenging for the self-calibrating bundle adjustment, is to reduce the number of variables by decoupling interior from exterior camera parameters (Cramer et al., 2017). In such a pre-calibrated workflow, the bundle adjustment only solves external parameters while the camera model remains fixed.

Although numerous publications have highlighted its potential, there is a general lack of studies investigating methods and effectiveness of sensor pre-calibration for geomorphological applications (Oniga et al., 2018). Specialized photogrammetric pre-calibration approaches are often not suitable for geomorphological applications,

for several reasons. Most critically, the inherent geometric instability of low-cost commercial sensors can invalidate long-term calibration. Vibrations, temperature and pressure changes potentially affect the sensor geometry and may rule out transport between calibration laboratory and field site (e.g., Cramer et al., 2017; Elias et al., 2020; Sanz-Ablanedo et al., 2020). Therefore, we assume it is critical to perform the calibration on-site, immediately before or after each survey. In situ sensor calibration often applies portable calibration frames or 2D checkerboards (Griffiths & Burningham, 2019). However, when the calibration is carried out at survey scale (i.e., same distance as survey height) to maintain focus settings, as emphasized by Lichti et al. (2008), such calibration structures may not be suitable for all sensors due to the feature size. Moreover, analogous to self-calibration on flat surfaces (Sanz-Ablanedo et al., 2020), the geometry of the calibration structure is critical for building a strong image network by reducing the correlation of the camera parameters (Remondino & Fraser, 2006). Samper et al. (2013) and Oniga et al. (2018) have demonstrated that 3D calibration structures result in approximately 50% higher accuracies compared to 2D planar calibration fields.

Oniga et al. (2018) created a 3D test field on a lawn and a façade that allows calibration at scale. However, the calibration and survey sites were spatially and temporally separated. Harwin et al. (2015) set up a calibration field on-site and included some degree of three-dimensionality by placing survey targets on tripods. The downside of portable calibration structures with survey targets is the workload resulting from the required set-up and surveying for every repetition. In the case of multi-temporal surveys this can be especially time demanding.

Additional challenges arise from the application of multi-sensor systems. Few publications investigate pre-calibration of thermal sensors (Conte et al., 2018) and none of the approaches attempts simultaneous calibration of multi-sensor systems. Using the same structure for visible and thermal sensors would be the most efficient, but their dynamic properties are not necessarily compatible and hence require suitable targets (Conte et al., 2018). When investigating thermal sensors, previous studies have applied active (e.g., light bulbs) (Luhmann et al., 2013) or passive targets (e.g., holes in aluminium plate (Bison et al., 2012) or black velvet and silver heat protection foil (Westfeld et al.2015)), subsets of the survey imagery (Conte et al., 2018), 2D calibration planes (Bison et al.2012; Westfeld et al., 2015) and 3D calibration frames (Eltner et al., 2021; Luhmann et al.2013).

Most of these approaches use short sensor–object distances and calibration structures that cannot be scaled up for a calibration on survey scale. Moreover, the previous approaches do not account for (field-) time-efficiency or cannot be performed on-site. The approach presented by Senn et al. (2020) meets the criteria but has not yet been applied to pre-calibrate a survey.

1.1 | Aim and objectives

The overarching research reported in this paper is one such geomorphological case where the requirements of the survey do not readily allow for sufficient self-calibration. We conducted a multi-temporal SfM photogrammetric survey using a multi-sensor RPAS to monitor geomorphic changes induced by artificially added in-channel log jams,

designed to help restore habitat for Atlantic salmon (*Salmo salar*), on the River Gairn in Aberdeenshire within the Cairngorms National Park, Scotland. We derived topographical information from visible imagery and used orthophotos to map the surfaces of the survey site. Supplementary thermal orthophotos provide a valuable addition to aid surface classification.

A pre-calibration method aimed at geomorphological applications was described by Senn et al. (2020) and was designed as an applicable addition that can fit into restricted fieldwork schedules. It overcomes the shortcomings of previous, typically lab-based, calibration workflows, most prominently scale, workload and suitability for different sensors. The pre-calibration approach utilizes distinct features on a building present in the survey site instead of survey targets. The calibration dataset is generated by manual localization of features as 3D coordinates from terrestrial laser scans for reference, and 2D image coordinates for the calibration.

Following up on Senn et al. (2020), this paper reports on the effectiveness of sensor pre- and self-calibration, as well as other error mitigation strategies, on the accuracy of results achieved with SfM photogrammetry in a full topographic survey of wetted and dry areas in a river corridor. To this end, we conducted a systematic experiment in which we used all combinations of calibration strategy (pre-calibration in vision measurement system (VMS), pre-calibration in Agisoft MetaShape (AM) and self-calibration in AM), flight altitudes and viewing angles, as well as masking out error-prone areas. Furthermore, we evaluate a simplified scan set-up for pre-calibration reference data using a single scan instead of a registered point cloud acquired from multiple different perspectives. In addition, we demonstrate the multi-sensor applicability by creating thermal orthophotos. The objectives of the study were to:

1. assess the impact of the scan set-up and software choice on pre-calibration accuracy;
2. compare the performance of sensor pre- and self-calibration and additional error mitigation strategies in a geomorphological survey scenario;
3. evaluate the applicability of the methods with regard to a real-world survey on the River Gairn.

Ultimately, our aim is to make an informed recommendation for sensor calibration in geomorphological research. In contrast to photogrammetry-centred research, our emphasis is to balance photogrammetric accuracy and geomorphological applicability. We appreciate that factors such as streamlined software implementation (e.g., compatible file formats), software availability and time requirement can be critical in deciding whether to implement additional processing steps to the workflow. The calibration method is specifically designed to minimize on-site time requirement, being aware that the cost–benefit consideration ultimately determines whether a method is adopted or not.

2 | METHODOLOGY AND DATASETS

The methods are decomposed into sensor pre-calibration and application to the real-world scenario of a geomorphological survey (Figure 1). Visible imagery was captured using the built-in sensor of a
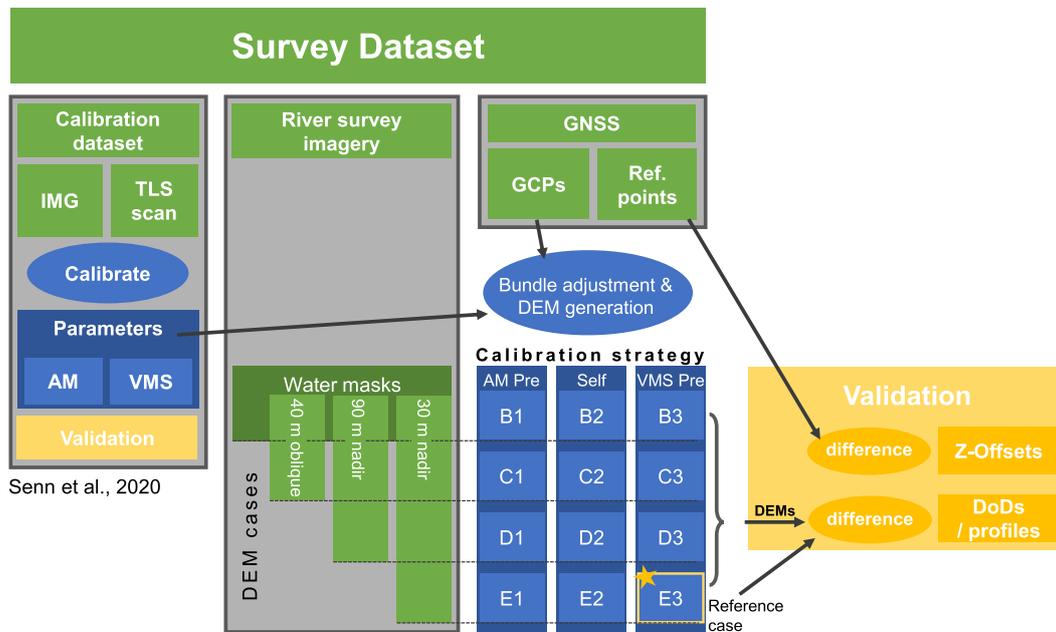
FIGURE 1 Workflow of the calibration and photogrammetric method [Color figure can be viewed at wileyonlinelibrary.com]

DJI Phantom 4 Pro (resolution 5472 × 3648 pixels, pixel size 2.41 μm, nominal focal length 8.8 mm) with a polarization filter. Thermal imagery was captured using a Workswell WIRIS Pro sensor (resolution 640 × 512 pixels, pixel size 17 μm, nominal focal length 13 mm) mounted on a DJI Matrice M600 RPAS. The calibration flights were conducted immediately prior to the survey flights with both RPAS.

The first part (Section 2) presents an updated workflow (Figure 1) of the pre-calibration method presented by Senn et al. (2020), which describes in detail aspects of the methodology and are therefore only briefly summarized. Further details are provided wherever adaptations and new procedures were applied. The key changes in the approach were to derive the calibration reference from a single terrestrial laser scan instead of a point cloud merged from multiple scans, as well as refining the criteria for target feature definition. The results in both publications are based on the same raw primary data.

The second part (Section 2) describes the survey data collection and SfM photogrammetric processing, for which we followed the guidelines provided by James et al. (2019). All datasets were collected on 12 September 2019 on the River Gairn in alternating sunny and cloudy weather conditions. The River Gairn (150 km^2) is an upland stream and tributary of the River Dee (2104 km^2). The catchment area of the study site is 44 km^2 and we measured a discharge of 1.31 m^3/s on 12 September 2019 at 12:00 GMT. The landscape is dominated by managed heather moorlands with recent tree plantations along the streams. No mature trees or taller vegetation are present along the surveyed section; it is therefore well suited for the SfM survey. The riverbanks are mostly covered by grass or heather and partly undercut. The riverbed and several exposed bank sections and gravel bars along the stream consist of coarse gravel and cobbles with one area of exposed granite bedrock at the bend next to cross-section 2 (Figure 3). The channel features riffles and pools up to approximately 1.5 m in depth and its width varies between 8 and 22 m. At the time of the survey the tree trunks for the log jam construction were already placed on the banks but not yet installed.
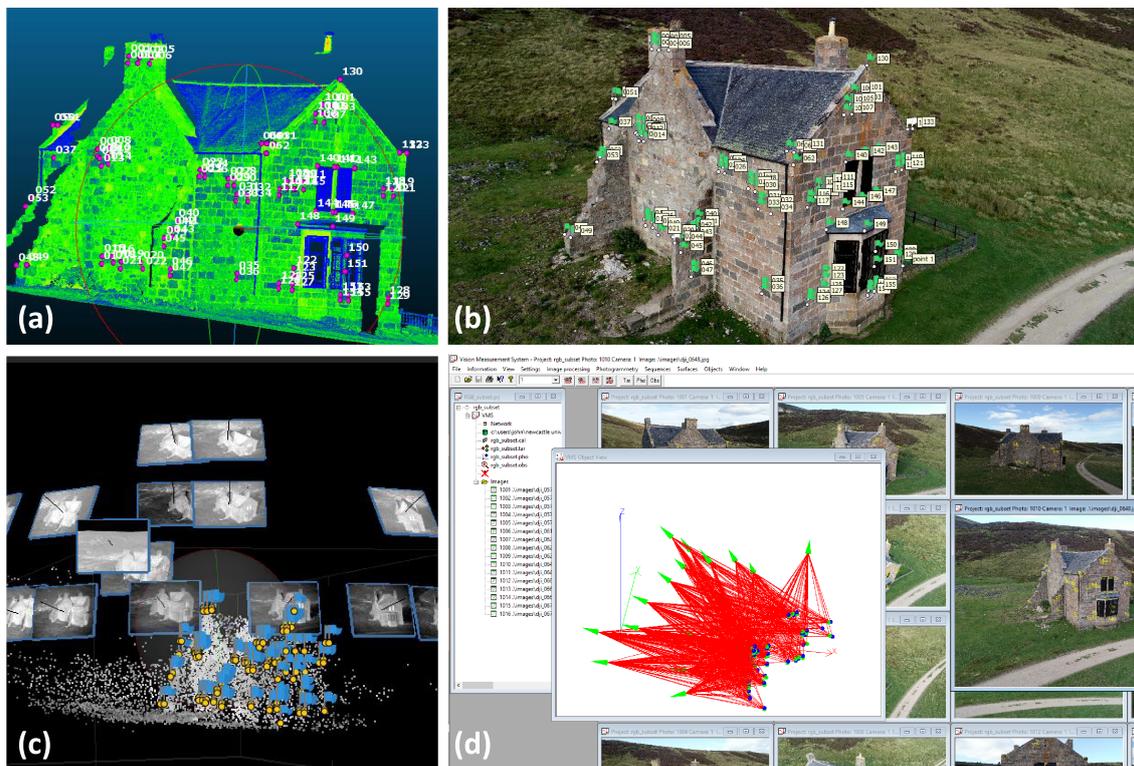
2.1 | Pre-calibration

For the sensor pre-calibration any 3D structure present on-site can be utilized as a calibration structure. This could be either natural (e.g., boulders or rock formations) or artificial (e.g., bridges, buildings or stone walls) stable structures. In our case study the sensor pre-calibration was carried out using a stone building as the calibration structure (Figure 2). Imagery from a thermal and an RGB sensor taken with varying perspectives and distances serve as the calibration dataset, and terrestrial laser scans provide the reference dataset. Conjugate features clearly visible in both imagery and point cloud were used as calibration targets. The following paragraphs provide more detail on the preparation of the reference data, preparation of the calibration data and generation of the sensor parameters.

We adopted the close-range photogrammetry software VMS (version 8.8) (Shortis & Robson, 2015) as the calibration benchmark and AM (version 1.7.2) (Agisoft LLC, 2021) as a potentially more applicable and widely adopted alternative. VMS is established as calibration software (James et al., 2020; Shortis & Luhmann, 2018) wherein the photogrammetric procedure is transparent, comprehensible and highly customizable. However, the operation requires precise knowledge and can be somewhat cumbersome, especially when utilizing large datasets.

2.1.1 | Reference dataset (3D feature coordinates)

The reference dataset was derived from a point cloud captured using a Leica ScanStation P40 terrestrial laser scanner operating from a sensor–object distance of 20 m with a resolution setting of 3.1 mm, at a distance of 10 m. The raw scans were processed using Leica Cyclone (version 9.2.1). The point cloud was not georeferenced and the calibration was performed using a local coordinate system.

The multi-scan approach described in Senn et al. (2020) covered the calibration structure from 360° to optimize network stability and

FIGURE 2 The building utilized as calibration structure. (a) Marker definition in the terrestrial laser scanner point cloud in CloudCompare; (b) marker placement on visible image in AM; (c) tie points and camera locations of the thermal sensor; (d) calibration process in the software VMS [Color figure can be viewed at wileyonlinelibrary.com]

to avoid bending from systematic errors. Here we test a single-scan set-up that uses only one scan covering two façades. In this way we intend to remove any error induced from the multi-scan point cloud registration into the calibration, as well as speeding up the field-surveying procedure. Additionally, it facilitates a stable convergent network with fewer images and thus fewer target features that require manual identification. Moreover, high noise levels in the absence of direct incident solar irradiation (e.g., north-facing façade) make identifying features in thermal imagery challenging. Therefore, it is advantageous to use façades with more favourable lighting conditions. 104 target features distributed over the south- and west-facing façades were manually extracted from the single-scan point cloud using the open-source software CloudCompare (version 2.20.2 Zephyrus; GPL Software, 2019) (Figure 2). In this context, a feature is defined as a precisely determinable individual point position—for example, the corner of a stone.

Further adaptations to the previous workflow were made in the selection of target features. To qualify as a suitable target, a feature must be visible in terrestrial laser scanning and all imaging sensors. This eliminates building outlines, corners and window openings, which are often not clearly recognizable in the thermal imagery. We found individual stones to be more clearly recognizable, albeit more challenging to precisely locate in the terrestrial laser scanner point cloud.

2.1.2 | Calibration dataset (2D feature coordinates)

The sensors were used to capture a total of 158 thermal and 101 visible images flying in circular patterns triggering at a set interval of 3 s.

The distances between features and sensors range between 8.2 and 32.4 m for the RGB sensor, and between 20.8 and 37.5 m for the thermal sensor. To ensure the reliability of the calibration by eliminating outliers, a sufficient level of redundancy in the dataset (i.e., number of images and observations) is necessary (Shortis, 2019). Adding more images does not indefinitely improve the calibration but leads to increasing computational time (Eltner et al., 2016) and manual work. All images were imported and aligned in AM to allow for selection of suitable calibration subsets. An ideal calibration dataset features a convergent network with a variety of distances and perspectives, sufficient overlap, and covers the entire sensor area (e.g., Eltner & Sofia, 2020; Kenefick, 1972; Oniga et al., 2018; Sanz-Ablanedo et al., 2020; Shortis, 2019). This minimizes the parameter correlations and ensures that the calibration accurately represents the physical model (Kenefick, 1972; Shortis, 2019). Senn et al. (2020) yielded suitable root mean square error (RMSE) values (Table 1) for subsets of 16 images using the multi-scan set-up, and Hieronymus (2012) state that a subset of 8–12 images can be sufficient. Given the increased overlap, and thus redundancy, in the single-scan set-up we concluded that a subset of 16 images provided the ideal dataset size. The feature observations were created by placing markers in AM and exported using a script for the built-in Python console. We have deliberately avoided coded targets that could be automatically recognized. Using natural features instead allows the survey to be repeated based on the same terrestrial laser scan (and feature coordinates) without having to repeat the target placement and surveying. Moreover, a higher accuracy was achieved by manually placing markers, rather than using the AM automatic `Refine Markers` tool.

TABLE 1 Overview of the datasets and resulting metrics of the sensor pre-calibration (multi-scan set-up results from Senn *et al.* 2020)

Reference set-up	Sensor	Software	Dataset	Images	Observations	RMSE (pix) Control	Check
Multi-scan	Thermal	VMS	subset	12	317	0.48	
			add6	17	404	0.53	
		AM	subset	14	331	0.81	0.54
			add6	20	432	0.88	0.73
	rgb	VMS	subset	15	609	0.88	
			add6	21	752	1.02	
		AM	subset	16	630	0.92	0.71
			add6	21	773	1.19	1.31
Single-scan	Thermal	VMS	subset	16	989	0.21	
		AM	subset	16		0.44	0.34
	rgb	VMS	subset	16	1046	0.32	
		AM	subset	16		0.54	0.54

2.1.3 | Determining calibration parameters

The reference data were split into independent check and control points for validation. The exported 3D reference feature coordinates and the 2D observation coordinates were used to determine the benchmark pre-calibration in VMS by solving the collinearity equation (Brown, 1971) as described in Senn et al. (2020). In parallel, we performed a second pre-calibration in AM by importing the 3D reference feature coordinates to generate camera parameters by self-calibrating bundle adjustment (Figure 1). As suggested by Shortis and Luhmann (2018), the parameters were initially fixed and subsequently released iteratively, beginning with the radial distortion parameters (k_1, k_2 and k_3) and the principal point offsets (x_p and y_p) to the tangential distortion (p_1, p_2) to the affinity and orthogonality terms (b_1, b_2). The AM version used does not allow a marker-based calibration; hence the photogrammetric tie points could not be excluded from the calibration, unlike the procedure presented by Senn et al. (2020). We believe a strictly marker-based calibration would facilitate the exclusion of unwanted image-based optimization in the AM software 'black-box' and at the same time allow a better comparability with VMS (Harwin et al.2015).

We selected suitable values for **image accuracy** based on James, Robson, et al. (2017) and Harwin et al. (2015): 0.5 pixels in both software packages and 1 pixel for the tie points in AM. Marker accuracy was set to 3 mm according to the observation provided by the terrestrial laser scans.

2.2 | Real-world application: river survey

Having successfully pre-calibrated the sensors, the subsequent step was to transfer the camera parameters to the river survey dataset. The focus of our application was on DEM generation using visible imagery. Thermal imagery, on the other hand, was utilized for the creation of orthophotos to be applied for water surface detection. The survey was conducted on a 1 km reach of the upper River Gairn, covering a total area of 0.7 km^2. The area was split into a west and an east section, with separate take-off and landing sites to avoid exceeding the legal flying distances. The optical dataset consists of a total of 922 images from three different flying heights and view angles:

- 30 m nadir, 504 images (overlap: forward 60%, lateral 60%);
- 40 m oblique, 256 images (forward 50%, lateral 40%);
- 90 m nadir, 162 images (forward 80%, lateral 60%).

The choice of rather tightly dimensioned overlaps was due to the high redundancy afforded by flying three different heights. The variation in scale and view angles was used to maximize the convergence of the geometric network (James & Robson, 2014). High redundancy was chosen to allow isolating the effect of individual error mitigation strategies while maintaining sufficient information.

The GNSS ground reference consists of a network of GCPs and additional reference point measurements. The 14 GCPs were distributed along the river corridor at varying distances from the river (Figure 3). The narrow-spaced survey design was chosen based on the requirements of the river focused survey. The GCPs were marked using custom-made circular 1 m diameter tarpaulin targets with black vinyl sectors so as to be recognizable in both the thermal and visible spectrum. During the survey they were pegged down, and the centre point was underfilled with large stones for stabilization of target elevation. In addition, 1021 reference point measurements classified into the surface classes GRS (grass and low vegetation, 158 points), GRV (gravel by the stream, 156 points), ROA (road/gravel track, 160 points), SHR (shorelines, water boundary, 245 points), SUR (water surface, 31 points) and WET (wet riverbed, 217 points), were recorded (Figure 3). In contrast to the GCPs, the point measurements are not related to visible features. Instead, they serve for assessing vertical offsets related to distorted DEMs. The GNSS reference was recorded using two Leica GS18 rovers recording in static and kinematic mode measuring each of the GCPs twice for 3 min (pole + bi-pod) and the point measurements for 3 s (hand-held pole). At the same time, a set of two Leica GS10 receivers were set up as on-site GNSS base stations.

The raw data were post-processed using Leica Infinity (version 2.4.1). In the first step, the base stations were processed with a
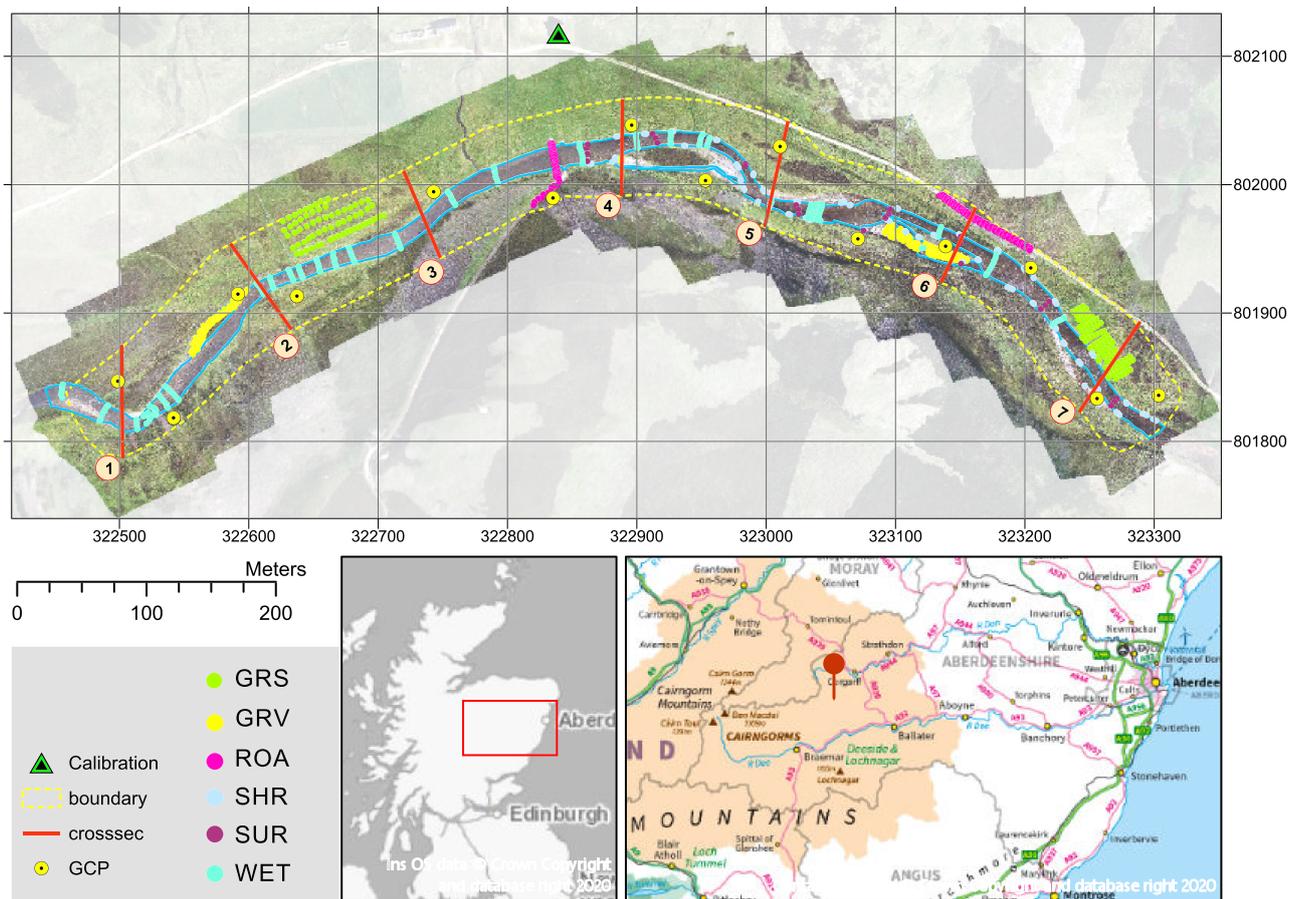
FIGURE 3 Locations of the GNSS reference points, GCPs, orthophotos (based on the presented dataset) and calibration building on the River Gairn study reach. The reference points are classified as GRS (grass and low vegetation,), GRV (gravel by the stream), ROA (road/gravel track), SHR (shorelines, water boundary), SUR (water surface) and WET (wet riverbed) [Color figure can be viewed at wileyonlinelibrary.com]

baseline from the local OSNet station in the nearby town of Braemar (BRAE), which is located 13 km away. The software has achieved a sub-millimetre accuracy for the base stations in post-processing. Subsequently, the raw data acquired by the rovers were processed with a baseline from the fixed coordinates of the local base stations. The high measurement accuracy at the base stations ensured a correspondingly high relative accuracy at the reference points (1 mm) and GCP (0.1 mm) measurements, based on metric 'CQ 3D' provided in the Leica Infinity output. However, this metric appears to be overly optimistic. Typically, the accuracy is estimated as an average over repeated measurements at the same location, which is not always feasible due to time constraints in the field. The reference data were exported from Leica Infinity and imported into R (R Core Team, 2020) for further filtering and quality assessment. Due to a faulty battery not all GNSS reference points could be successfully post-processed; therefore, the spatial distribution is not ideal.

To isolate the effect of different levels of error mitigation, the visible imagery and the GCPs were processed in AM following 12 different predefined cases, as shown in Table 2.

Cases B1, B2 and B3 use water masks during the bundle adjustment to remove the effect of light refraction at the water surface. For consistency between cases, we generated a mask for every individual image beforehand. The masks can be generated efficiently from a mesh and do not require centimetre accuracy. For this, a self-calibrating bundle adjustment was run on the full visible dataset

(`Align Photos, Accuracy: Medium`) and subsequently `Build Mesh` was run using the sparse point cloud with `Face Count` set to `high`. The mesh was displayed in colours and all vertices containing water were manually removed. Finally, the masks were created by backward projection from the survey coordinates into image coordinates using the `Import Masks` tool, selecting the method `from model`. The masks were visually inspected and exported for application in all cases that use masks.

The cases were set up in sets of a pre-calibrated pair and a complementary self-calibrated case. The cases B1, B2 and B3 use the full set of error mitigation strategies. With each subsequent set, another element of the error mitigation strategy is omitted (Figure 1). We removed the water masks from C1, C2 and C3, 40 m oblique imagery from D1, D2 and D3 and the 90 m nadir imagery that provides a second scale from E1, E2 and E3. The E cases only use the 30 m nadir imagery.

2.2.1 | DEM generation of the river survey dataset

The following workflow was applied for the DEM generation in AM. The first step was an initial rough camera alignment to aid the marker (GCP observation in AM) placement. Overexposed marker observations were omitted and blurry images were discarded. Masks were imported if required. For pre-calibrated cases the camera

TABLE 2 Processing set-up cases using different levels of error mitigation and calibration strategies

| Case | Calibration | Software | Nadir | | Oblique | Mask |
			30 m	90 m	40 m	
B1	pre	AM	x	x	x	x
B2	self		x	x	x	x
B3	pre	VMS	x	x	x	x
C1	pre	AM	x	x	x	
C2	self		x	x	x	
C3	pre	VMS	x	x	x	
D1	pre	AM	x	x		
D2	self		x	x		
D3	pre	VMS	x	x		
E1	pre	AM	x			
E2	self		x			
E3	pre	VMS	x			

parameters were imported and all parameters were set to `fixed`. The AM pre-calibrated camera parameters were imported in the proprietary xml format. For the parameters generated in VMS we used the AM import tool to convert them from mm units to pixels before saving them as xml for consistency. For all cases the bundle adjustment (`Align Photos`) was run with `Accuracy` set to `High`, with a marker accuracy of 5 mm (based on the GNSS survey precision values) and the image coordinate accuracies set to 0.5 pixels for the markers and 1 pixel for the tie points (Harwin et al., 2015; James, Robson, et al., 2017). All GCPs were used as control to maximize the network stability, while the relevant validation and detection of systematic errors was undertaken using the GNSS point reference measurements that manifest in Z-errors. For the self-calibrated cases we calibrated the sensors by iteratively releasing parameters (x_p, y_p, k_1, k_2, k_3, p_1, p_2, b_1, b_2) using the `Optimise Cameras` function. To ensure model accuracy, the sparse point clouds were filtered using the `Gradual Selection` tool, removing all tie points with fewer than three observations and reprojection errors above 0.5 pixels. The water masks were removed before building the dense cloud (`Quality: High, Depth filtering: Moderate`). Finally, the DEMs were created (default settings), resampled (pixel size 5 cm) and trimmed to a uniform resolution and boundary extent. The DEMs were then exported for analysis in R (R Core Team, 2020). In R the GNSS reference dataset was used to extract raster values from the DEMs and to illustrate the z-offsets in plots. The DEM showing the lowest errors (E3) was selected as reference case and used to calculate DEM of difference (DoD) by pairwise subtraction. Furthermore, the calibration parameters of the self-calibrated cases were exported (see Appendix Table A1) and plotted as distortion profiles alongside the precalibrated results (see Appendix Figure A1).

2.2.2 | Thermal imaging

We implemented an additional self-calibrating bundle adjustment using the thermal sensor to complement the experiment and as a proof of concept. The 'best-case scenario' with imagery from all three heights and angles was used, and no water masks were applied. The retrieved calibration parameters were exported (Table A.1) and plotted as distortion profiles alongside the pre-calibrated results (Figure A.1).

Thermal imagery is not highly suitable for DEM creation due to its low resolution and dynamic range, but it has significant potential in orthophoto creation (Maes et al., 2017). For this purpose, the SfM software (AM) requires a surface on which the imagery can be mapped. Most applications rely on DEMs from other sources—for example, lidar or SfM point clouds from visible imagery (e.g., Harvey et al., 2016; Lewis et al.2020; Maes et al., 2017).

We demonstrated the orthophoto generation workflow based on the DEM adopted from the visible dataset (case E3) in AM.

Analogously to the workflow of the visible dataset, we imported the 90 m nadir thermal imagery and fixed the camera parameters to the VMS pre-calibration parameters. The images were then aligned provisionally to aid marker placements on the GCP observations. Subsequently, the images were finally aligned using the highest accuracy setting. At this point the method deviates from the visible: Instead of building the dense cloud and deriving the DEM from thermal imagery, we imported the previously created DEM from visible imagery. Finally, we created the orthophotos using `Build Orthomosaic`.

3 | RESULTS

3.1 | Pre-calibration results

Table 1 aggregates the results and quality metrics of the precalibrations from both scan set-ups. The image observation RMSE values of the single-scan set-up were approximately 50% lower than those found in the multi-scan set-up reported in Senn et al. (2020). The standard deviations are 0.23 (thermal) and 0.22 pixels (visible) in VMS and 0.11 (thermal) and 0.17 pixels (visible) in AM. The reduction from three to one scans reduced the errors by approximately 50%, on average. In accordance with the multi-scan set-up, the lowest values were achieved by the VMS calibration and the visible resulted in higher values than the thermal sensor. Table A1 (Appendix) contains all calibration parameters of the pre- and self-calibration cases. The

corresponding distortion profiles for radial and tangential distortions can be seen in Figure A1 (Appendix). The pre-calibrated profiles of the RGB sensor followed a similar course for both software packages. In the case of the thermal sensor, less uniform profiles were calculated. The self-calibrated RGB distortion profiles from processing the different cases of the survey dataset form a highly similar group and run parallel to the AM pre-calibration. The self-calibrated thermal camera parameters show a profile more similar to the AM pre-calibration.

3.2 | DEM error analysis of the river survey dataset

The different processing cases (Table 2) were applied in AM to produce a series of DEMs. The Z-values were extracted from the DEMs using the *xy*-coordinates of the GCPs and GNSS reference points to assess the height offsets using R (R Core Team, 2020).

Figure 4 shows boxplots of the offsets between the GCPs and the DEMs. Overall, the smallest GCP offsets were calculated using the VMS pre-calibration, the AM pre-calibration yielded the largest errors (lower for E1) and the self-calibration is in between the two (with a trend towards higher errors with less error mitigation). The largest errors were found for the D cases that use nadir imagery from two flying heights. Only for the single-scale nadir case E did the AM pre-calibration result in smaller errors at the GCPs.

The reference point classes GRV and ROA are the least susceptible to distortion or noise and are therefore used to evaluate DEM quality (Figure 5). This is reflected in the relatively small offset values and variation of these two dry surface classes compared to the other classes (Figure A4 in the Appendix). The higher levels of error in the other classes are caused by the surface characteristics—for example, turbulent flow and the refraction at the water surface (WET and SUR), obstruction by vegetation (GRS) and visually ambiguous water edges due to overhanging banks (SHR).

For self-calibrated cases, the errors show a rough trend of increasing values with the incremental omission of error mitigation strategies (Table 3 and, in Appendix, Figure A2). This trend is not only visible in the RMSE values but also in the deviation of errors, as indicated by the larger inner quartile ranges in the boxplots (Figure 5) and the standard deviations (Appendix Table A2 and Figure A3). Similarly low errors (0.2, 0.3 m) were calculated for both of the self-calibrated cases using oblique imagery with (B2) and without (C2) masking of

water bodies. The nadir-only self-calibrated cases with multi- (D2) and single-scale (E2) show the largest errors and standard deviations.

The pre-calibrated cases behave in the opposite manner, with the offsets reducing with fewer error mitigation strategies in place. The AM pre-calibrated cases show similar errors and standard deviations for the multi-scale cases (B1, C1 and D1) and slightly lower for the single-scale nadir-only case E1. Pre-calibration using VMS resulted in errors approximately half of the AM pre-calibration. However, the standard deviations reflect a similar level of precision. The smallest errors overall were found for case E3 with RMSE of 0.02 and 0.01 m. Both pre-calibrations resulted in slightly higher errors for the two-scale nadir case D.

The other surface classes, which were primarily measured for the validation of the bathymetric survey, also show interesting results and therefore warrant a brief mention here. The overhanging or opaque vegetation and riverbank structures of the classes GRS and SHR result in systematically positive DEM offset values (Appendix Figure A4). The two wet classes, SUR and WET, show opposing patterns: slightly negative offsets for SUR and positive for WET, as DEMs contain only one value per raster cell and thus represent either surface or river bed.
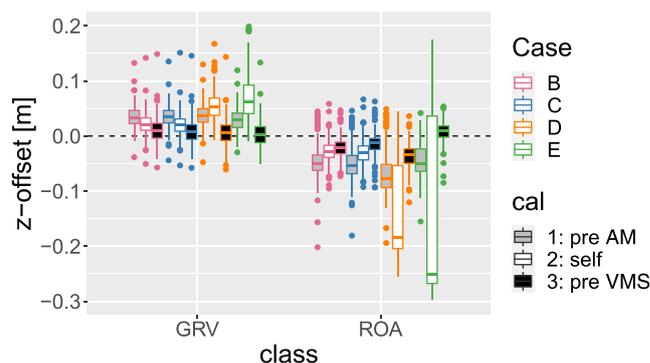
FIGURE 5 Boxplots of the z-offsets between the GNSS measurement and the calculated DEMs at the reference points of the dry classes GRV and ROA [Color figure can be viewed at wileyonlinelibrary.com]

FIGURE 4 Boxplots of the z-offsets between GNSS measurement and the calculated DEMs at the GCPs [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 3 RMSE (m) values of the z-offsets between the GNSS measurement and the calculated DEMs at the reference points. See Figure A2 (Appendix) for visualization as barplot

DEM case	GRS	GRV	ROA	SHR	SUR	WET
B1	0.14	0.04	0.05	0.19	0.10	0.20
B2	0.16	0.02	0.03	0.18	0.10	0.21
B3	0.17	0.02	0.02	0.18	0.09	0.20
C1	0.14	0.04	0.05	0.19	0.09	0.20
C2	0.16	0.02	0.03	0.18	0.10	0.19
C3	0.18	0.02	0.02	0.18	0.09	0.19
D1	0.13	0.04	0.07	0.19	0.11	0.20
D2	0.13	0.05	0.15	0.22	0.08	0.27
D3	0.16	0.02	0.04	0.18	0.11	0.19
E1	0.16	0.03	0.05	0.17	0.13	0.16
E2	0.11	0.07	0.21	0.21	0.08	0.27
E3	0.15	0.02	0.01	0.16	0.13	0.16

Based on the literature, it is to be expected that insufficient sensor calibration leads to systematic errors that result in characteristic spatial error patterns (e.g., James et al., 2020; Sanz-Ablanedo et al., 2020). To detect such DEM distortions, we plotted the z-offsets between GNSS reference points and DEMs of the relevant dry surface classes (GRV and ROA) on a map (Figure 6), as well as between the GCPs and the DEMs (Appendix Figure A5). Red shades indicate DEM elevations higher than the corresponding GNSS measurements, and blue shades correspondingly show lower values. A number of reference points could not be post-processed due to a faulty base station battery; hence the spatial coverage of the reference points is not ideal. Nevertheless, there are signs of tilting or dome-shaped distortion indicating systematic errors. These errors are particularly pronounced in the self-calibrated nadir-only cases D2 and E2 (Figure 6). To a certain extent, spatial error patterns are also visible in the AM pre-calibrated multi-scale cases B1, C1 and D1. The distribution of errors in the cases that showed larger errors in the boxplots is not random but displays systematic spatial patterns.

Because of its low errors we selected case E3 as the reference case for further analysis and applied it to calculate pairwise DoDs with all other cases (Figure 7). In addition, we extracted and plotted a set of cross-sections to visualize DEM distortions (Figure 8). A clear divide between pre- and self-calibration is evident for all case sets.

The similarities between cases are inherently greater within a calibration strategy and, consequently, using E3 as reference case creates a certain bias.

The DoDs of the AM pre-calibrated cases B1, C1 and D1 show a tilting or doming with positive offsets on the southern and negative values on the northern bank. The effect is slightly more pronounced in the two-scale nadir case D1. The nadir-only single-scale case E1 shows significantly lower offsets and less pronounced spatial pattern in the DoD. The three DoDs generated using the VMS pre-calibrated DEMs are similar, featuring low values and few systematic spatial patterns. The DoD of reference case displays the difference from itself and thus only zeros.
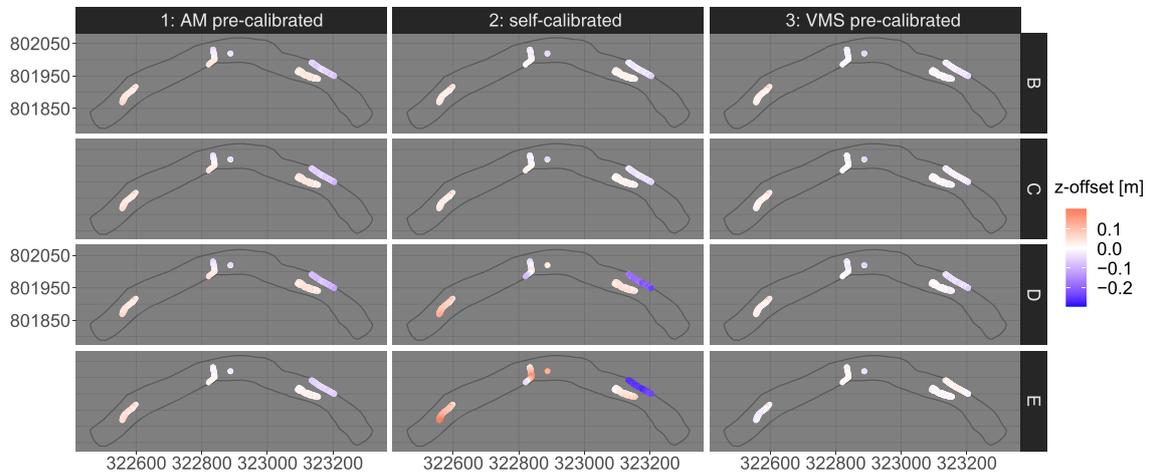
FIGURE 6 Spatial distribution of z-offsets between the DEM cases and the dry GNSS reference point classes GRV and ROA [Color figure can be viewed at wileyonlinelibrary.com]
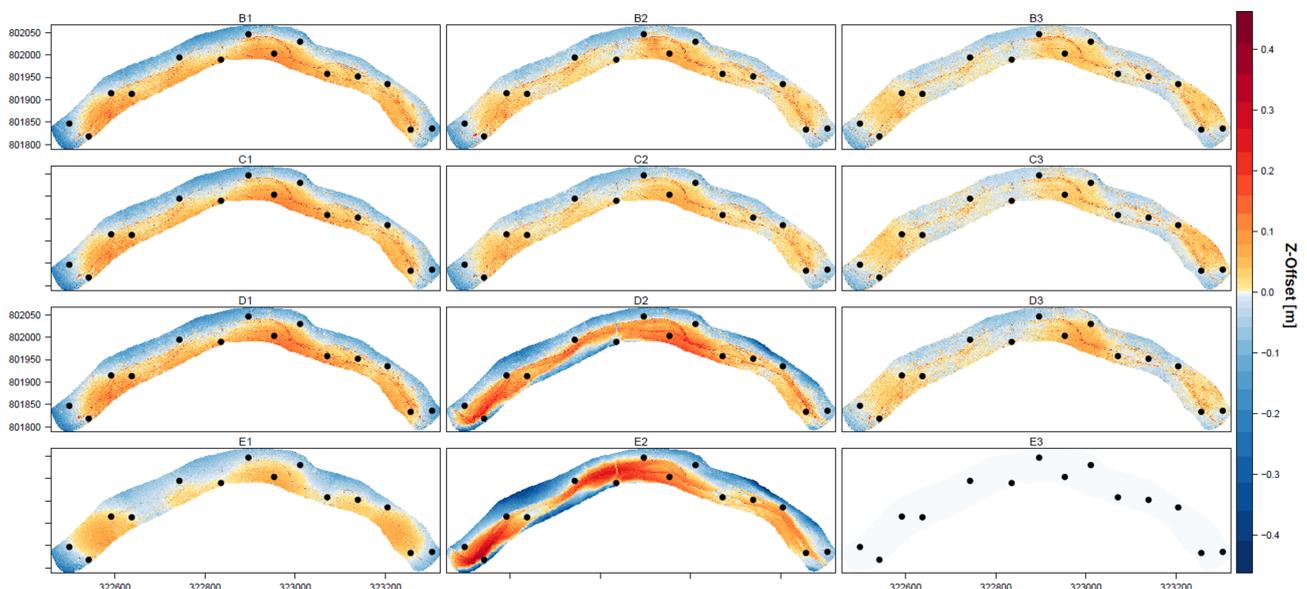
FIGURE 7 DoDs between E3 (reference case) and all other cases [Color figure can be viewed at wileyonlinelibrary.com]

An apparent divide is visible between the self-calibrated cases with and without oblique imagery. The oblique cases (B2 and C2) show lower DoD offsets and a less pronounced spatial pattern. The nadir-only cases (D2 and E2), in contrast, show a clear doming and the highest positive and negative offsets of all cases. The doming extends along the river axis and changes into a depression with increasing distance.

The DoD cross-sections in Figure 8 reveal a certain level of noise in the DEMs, which is particularly pronounced in the river sections (highlighted in grey). The doming effect observed in the DoDs of case D2 and E2 (nadir, two and one flying height) can be recognized in the run of the orange and green profiles in the centre pane. The observed tilt or slight doming in the AM pre-calibrated cases (less pronounced in E1) can be observed in an overall tilt to the right. Only in the AM pre-calibrated cases of the TR1 transect is this tilting of the DEM exceeded by the dome-shaped distortion, which is particularly pronounced towards the eastern end of the survey area.

## 3.3 | Application of the thermal dataset

We successfully applied the thermal dataset in a pre- and self-calibrated bundle adjustment (self-calibrated parameters in Appendix Table A1 and distortion profiles in Appendix Figure A1). To demonstrate a potential usage scenario, we mapped the thermal imagery of the 90 m nadir dataset onto the visible DEM (case E3). The created

thermal orthophoto is shown in Figure 9. The RMSE at the control points was 1.3 pixels.

4 | DISCUSSION

4.1 | Camera model and calibration parameters

The resulting RMSE values of the AM pre-calibration are close to half-pixel accuracy (0.44/0.34 for thermal and 0.54/0.54 for visible with standard deviations of 0.11 and 0.17 pixels) and, thus, according to
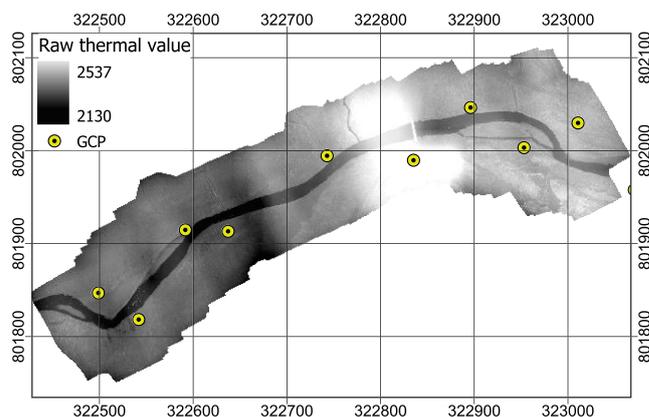
FIGURE 9 Thermal orthophoto from VMS pre-calibrated thermal imagery, projected onto a visible DEM (case E3) [Color figure can be viewed at wileyonlinelibrary.com]
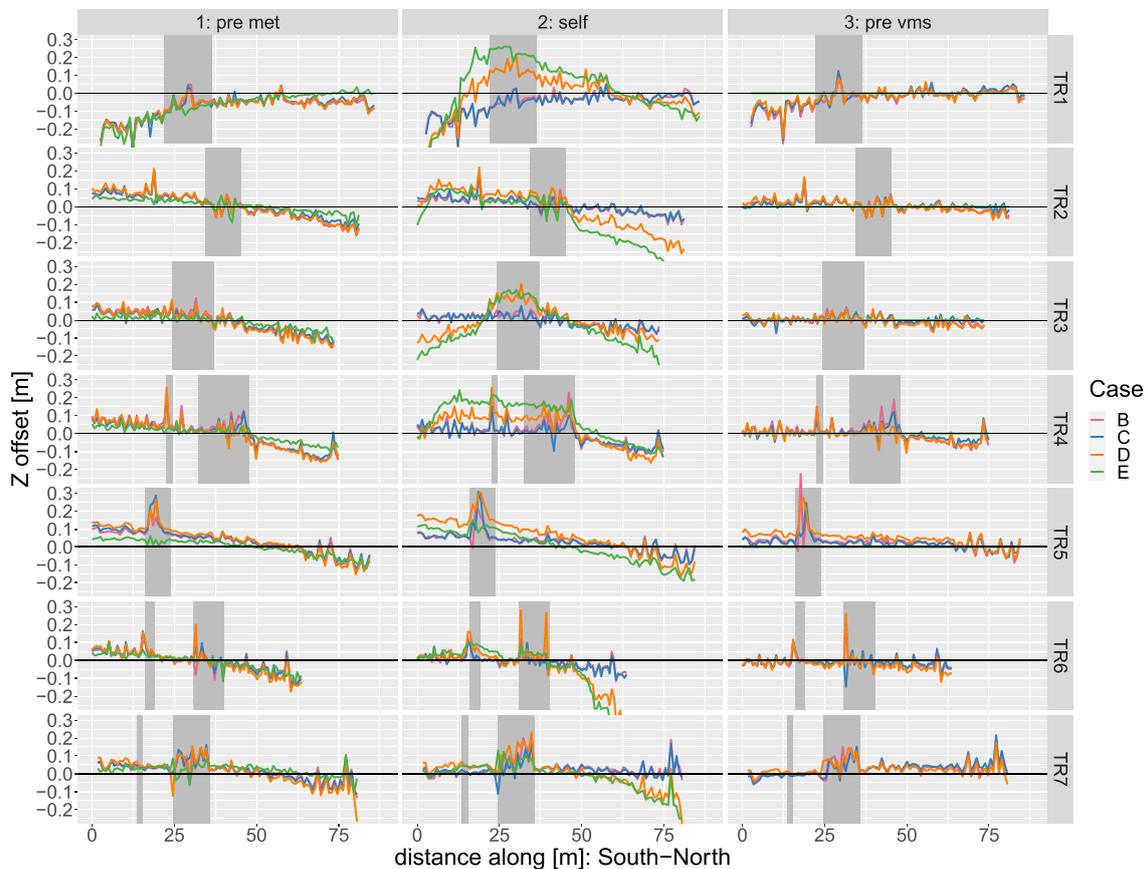
FIGURE 8 DoD cross-sections for all cases, in south–north direction (left to right). Grey sectors indicate the submerged sections. Locations indicated in Figure 3 [Color figure can be viewed at wileyonlinelibrary.com]

the literature, within the range of what can be achieved as a maximum for manual target measurements (Fraser, 2018; Geomsoft, 2008; Shortis, 2015; Shortis et al., 1995). The accuracy achieved by VMS is even better, with RMSE values of 0.21 and 0.30 (thermal, visible)—that is, below a third/quarter of a pixel and standard deviations of 0.23 and 0.22 pixels (thermal, visible). This indicates that the pre-calibrated camera calibration parameters are high quality. Furthermore, the low overall RMSE values would justify the use of values lower than one pixel for the image coordinate accuracy of the tie points in AM (James, Robson, et al., 2017).

The noticeably higher RMSE values of AM compared to the VMS benchmark, despite using the same parameters, illustrates the importance of software choice, an effect likely to be related to the tie points. Ideally, all tie points would be removed prior to determining the calibration parameters in AM to exclude the influence of its 'black-box' image-matching algorithms (Harwin et al., 2015). However, the applied software version does not allow for a marker-based calibration. To test whether the `tie point accuracy` could be exploited to decrease the weighting of the tie points in relation to the markers, we ran the AM pre-calibration with different settings (0.4, 1 and 6, based on James, Robson, et al. (2017), and 100). We found no significant effect on the RMSE, and therefore kept the initial value equal to unity. Future research should explore the potential of tie point masking, filtering or marker-to-tie-point conversion to optimize the pre-calibration capabilities with AM, as we emphasize great potential in a single-software solution.

Calibration quality is furthermore reflected in the highly similar radial and tangential distortion profiles of the RGB sensor in both pre-calibration scenarios. However, the distortion profiles of the thermal sensor deviate significantly. The principal point offsets x_p and y_p (see Table A1) are nearly identical between the two software packages for visible, while they are different in the thermal sensor. Both RGB and thermal sensors show high correlations between the tangential distortion parameters and the principal point offsets (thermal: 0.91 p_1–x_p, 0.86 p_2–y_p; and visible: 0.87, 0.72).

The typically high correlation of these parameters (Shortis, 2019) can indicate over-parametrization, and James, Robson, and Smith (2017) argue that individual parameters should be discarded if the error exceeds the parameter value. In our case the values were acceptable, and we kept the determined parameters for consistency between the two software solutions. We assume that the different solutions are related to the internal parameter weighting of the two software packages. Ultimately, the suitability for pre-calibration must be assessed by the performance in the SfM survey. For this reason, we believe a one-software solution could be beneficial. The findings of Hastedt et al. (2021) indicate that including additional parameters, such as the Fourier model, can improve the calibration validity, especially for atypical lens geometries of built-in RPAS sensors. These parameters are not implemented comparably in the applied software, so this could not be verified in this study, but should be the subject of future research.

4.2 | Calibration geometry and scan set-up

A central aim of this study was to test how the simplification of the scan set-up influences the quality of the sensor pre-calibration. We found significantly lower RMSE values using the single-scan set-up compared to the multi-scan set-up reported in Senn et al. (2020). We believe that several factors have played a role in this improvement in accuracy. Most importantly, using a single point cloud omits errors in point cloud registration. The original purpose of the multi-scan set-up was to encircle the calibration structure and thus to avoid bending and incorrect angles in the network. Previous researchers suggest the application of a 3D structure rather than a 2D calibration plane in order to create a stable geometric network (Harwin et al., 2015; Oniga et al., 2018). Our results now suggest that the single-scan of two façades provides sufficient 3D structure for a robust calibration. The major advantage of a single scan is that less façade area needs to be covered. Higher overlaps and variation of perspective and scale can be achieved with the same number of images. Consequently, it is easier to include odd angles and varying scales to optimize the convergent image network. The improved geometry has reduced the risk of parameter correlation and outliers can be eliminated more efficiently due to the higher redundancy (Shortis, 2019). At the same time, fewer target features need to be defined and manually digitized. Ultimately, processing time and manual work are limiting factors that have to be balanced with redundancy.

The single-scan set-up has some improvements that apply especially to thermal sensors. It allows exclusion of north-facing walls that are never exposed to direct solar irradiation in Scotland. This results in a low dynamic range for the imagery (a known issue of thermal sensors in SfM applications (Maes et al., 2017)) and thus makes the recognition of target features more difficult. Images acquired under direct solar irradiation provide better contrast and features can be digitized more accurately.

Changing the scan set-up also required defining and extracting a new reference dataset. We updated the conventions of feature selection based on the lessons learned in the previous approach, where the selection was mainly based on visible and terrestrial laser scanning. We found that building outlines tend to be fuzzy in the thermal imagery. Some stones that are clearly visible in thermal and terrestrial laser scanning cannot be distinguished from the surrounding mortar in the visible imagery. Overall, the selection convention evolved from corners and edges towards bricks while carefully assessing the visibility in all sensors.

The single-scan set-up not only improves the calibration accuracy but also the required workload. It cuts the time requirements for scans in the field, point cloud post-processing and registration and number of target features to manually digitize. Reducing the complexity increases the applicability and thus the potential applications of the approach.

The importance of calibrating at survey scale (i.e., sensor–object distance similar to flying height) has been emphasized previously (Griffiths & Burningham, 2019; Lichti et al., 2008; Roncella & Forlani, 2021). The sensor–object distances applied are similar to the nadir 30 m flying height. The scales of the higher flying heights are not represented in the calibration. Since the depth of field increases exponentially with increasing object distance, we assume that transferring calibration parameters generated from 30 m is more suitable than conventional pre-calibration routines using checkerboards or portable frames at short distances. Too small calibration objects would either not sufficiently cover the sensor area (Shortis, 2019) or result in blurred imagery, when flying closer with consistent focal settings (Grammatikopoulos et al.2019).

4.3 | Application in the geomorphological survey

The dry reference point classes GRV and ROA were not evenly distributed across the survey area, due to gaps in the base station data. However, we assume that the clustered distribution, with accumulations of points close to the boundaries (ROA) and the centre (GRV), is sensitive to systematic DEM distortions and thus suitable for the purpose. However, the reference points do not reflect whether the systematic error pattern shows doming or tilting deformation. This becomes clearer in our second validation dataset; the DoDs and the derived cross-sections provide better representation of spatial patterns and thus systematic errors in the camera calibration. However, the DoDs have to be assessed carefully, as they represent DEM offsets relative to a reference case (E3) and can thus be biased. Using two independent validation datasets compensates for respective weaknesses and strengthens results that point in the same direction.

The systematic design of our experiment allows us to isolate the effects of the individual mitigation measures.

4.3.1 | Water masks

The first element (only case B) of the error mitigation strategy was the water masks. However, in our scenario we did not find any significant effect of their application. None of the calibration scenarios display significant differences between cases B and C. The rationale was to mask out water bodies that are assumed to be particularly prone to errors due to the light refraction at the surface. The low point cloud confidence in the river sections of the dense point cloud, further down the processing chain, confirms this assumption. Only very few tie points are present in the water areas after the bundle adjustment without masks. This is most likely due to high tie point uncertainty after the image matching and their automatic elimination by the software. Thus, we conclude masks are not necessary in a turbulent stream, like the River Gairn, but there might be use cases where masking improves survey results—for example, on sites containing clear and calm water bodies. Other conceivable applications could be masking out moving grass or error-prone types of vegetation (e.g., Harwin et al., 2015) or snow cover in thermal imagery (e.g., Webster et al., 2018). Where the addition proves beneficial our approach can be particularly valuable because the implemented workflow allows fast and efficient generation of a global mask dataset.

4.3.2 | Oblique imagery

The second element of the error mitigation strategy was to include oblique images. The results of the scenarios using self-calibrated bundle adjustment show the biggest difference between the cases with (B2 and C2) and without (D2 and E2) oblique imagery. Relatively low RMSE values and spatial error pattern in the DoDs indicate that the best accuracy was achieved with inclusion of oblique imagery. The decline in model quality is also reflected in the increasing standard deviations by factors 4 and 8 at the dry GNSS reference points in the self-calibrated nadir cases D2 and E2. A convergent image network proves to be the necessary requirement for solving the self-calibrating bundle adjustment (James & Robson, 2014). These findings agree with

previous studies using simulation (James & Robson, 2014; James, Robson, & Smith, 2017) and applied (Nesbit & Hugenholtz, 2019) datasets that include oblique view angles. Cases D and E, in contrast, show the largest errors of the self-calibrated cases. The dome-shaped DEM distortion demonstrates systematic error due to insufficient geometry of the camera network. The self-calibrating bundle adjustment failed to simultaneously solve the external and internal camera parameters. For the pre-calibrated cases the inclusion of oblique imagery does not show significant benefit. Little variation can be found in the results of the pre-calibrated B, C and D cases. Zhou et al. (2019) even found the highest errors in scenarios combining oblique and nadir. The oblique imagery aids solving the internal camera parameters in the self-calibrating bundle adjustment, whereas the pre-calibrated bundle adjustment, when limited to solving the external camera parameters, does not seem to benefit from the convergent image network (Zhou et al.2019).

4.3.3 | Variation of scale

In the third error mitigation strategy, variation of scale clearly had an impact on the pre-calibrated cases. The results show smaller RMSE values for the cases E1 and E3 that only applied imagery from a single flying height (30 m). In addition, the DEM tilt that was present in the multi-scale cases (B1, C1 and D1) using AM does not show in the single-scale case, E1. The tilt is less prevalent in the VMS pre-calibrated cases, but the overall RMSE is likewise lowest in E3. Zhou et al. (2019) state that the focal length has the largest impact on vertical accuracy and is compensated by drifting of the camera position in the bundle adjustment. However, combining multiple flights (especially oblique ones) adds constraint to the network (Zhou et al., 2019), which may prevent it from drifting. Thus, the addition of image scales to achieve a more convergent image network does not benefit the pre-calibrated scenarios. For the self-calibrated bundle adjustment, in contrast, the variation of scale seems to be beneficial. The errors increase between cases D2 and E2 and also the DEM doming is more pronounced in E2. This agrees with the findings of Griffiths and Burningham (2019), who found that pre-calibration works better than self-calibration when applying single-scale nadir imagery, due to the lack of convergence.

4.3.4 | Effect of pre-calibration software

The pre-calibration with VMS resulted in the best results overall. This is in line with the expectation that the dedicated calibration software produces the most accurate sensor parameters. The similar level of precision as indicated by the standard deviation confirms that the larger error in the AM pre-calibration is related to systematic errors.

One aim of the present paper was to evaluate the applied workflow regarding applicability. In this context a single-software approach using exclusively AM would be preferable to provide an efficient and streamlined workflow without parameter conversion and transfer between software packages. The format conversion during the transfer from VMS to AM and the lack of consistency of internal parameter weighting have the potential to introduce error. However, the VMS benchmark dataset resulted in substantially higher DEM

quality. For the cases including oblique imagery the AM pre-calibration was even surpassed by the self-calibrated cases. In terms of efficiency a potential use case for the pre-calibration with AM is the nadir-only scenario. For this case the difference is smaller, but the pre-calibration using VMS is still superior. Future research should investigate alternative calibration software to provide a similar calibration accuracy to VMS but a more user-friendly implementation and file compatibility.

4.3.5 | Application of the thermal dataset

Additionally, we demonstrated an orthophoto generation using a pre-calibrated thermal sensor and an imported DEM. Thermal orthophotos can be valuable as an additional band for image classification or to calculate indices (e.g., Pádua et al., 2020). Within the framework of our overarching project, we see particular potential in the automatic classification of water bodies. Thermal image acquisition can be carried out on a smaller extent (e.g., less overlap, flying heights or view angles), as the bundle adjustment does not require a convergent image network if the internal sensor parameters are pre-calibrated. We assume that our findings regarding better performance of single-scale datasets with pre-calibrated sensors apply to thermal imagery too. The pre-calibrated bundle adjustment of the 30 m nadir thermal imagery failed to align a sufficient number of images. Instead, we used the 90 m (higher overlap) nadir dataset, which aligned immediately. We believe that low dynamic range of the imagery and the lack of small-scale surface structures on-site has compromised the image-matching capabilities of the software (Javadnejad et al., 2020; Maes et al.2017; Webster et al., 2018). This demonstrates a strength of pre-calibrating thermal sensors: While the optical properties on-site are often not ideal for image matching or self-calibration (Javadnejad et al., 2020), a suitable calibration object can be selected. In this context Webster et al. (2018) demonstrated that direct georeferencing can further aid the results of thermal SfM.

The orthophoto (Figure 9) shows radiometric irregularities related to changing solar irradiation conditions during the data collections. Future research should investigate the effect of more consistent lighting conditions and radiometric corrections. We believe that the application in this study can serve as a proof-of-concept application for the use of thermal sensors, but its success also suggests that the approach can be applied to other sensors—for example, multispectral and hyperspectral (e.g., Lucieer et al., 2012; Maes et al., 2017). Furthermore, we want to highlight that the method is not limited to SfM photogrammetry but can be used for other applications that require pre-calibration—for example, image rectification (Eltner et al., 2021; Grammatikopoulos et al., 2019) or image fusion to generate thermal point clouds (Javadnejad et al., 2020).

4.3.6 | Survey design and applicability in geomorphological research

Survey designs in geomorphological research are determined by the practical study requirements, and it is not always possible to adhere to the best practice guidelines for photogrammetric applications.

Time and battery life are significant constraints in RPAS-based SfM photogrammetry surveys. In this study we had to make the trade-off between length and width of the areal coverage. From a geomorphological perspective we were interested in covering the longest possible section of the river reach. The time-consuming lowest flights (30 m) determined the total area to be covered and thus the width of the GCP spacing. However, SfM surveys on linear corridors (e.g., rivers, roads, coastlines) are known to be prone to systematic errors (Ferrer-González et al., 2020; James & Robson, 2014). Therefore, the distribution of GCPs requires a certain level of width to constrain the model from rotation along the river/GCP axis in the bundle adjustment, as described by Duró et al. (2018). Moreover, the area of interest should be contained inside the GCPs to avoid errors in the unconstrained surrounding areas (Eltner & Sofia, 2020; Jaud et al., 2016; Oniga et al., 2018). The error patterns in the AM pre-calibrated cases B1, C1 and D1 and the less-pronounced tilting in the self-calibrated cases B2 and C2 could be related to the narrow survey design. The cases using VMS pre-calibration do not display spatial error patterns. Accordingly, a stable sensor pre-calibration seems to counteract such errors (James & Robson, 2014).

Furthermore, promising advances in direct georeferencing might make time-consuming ground surveys increasingly redundant (Carbonneau & Dietrich, 2017; Eltner & Sofia, 2020; Forlani et al., 2018; Hugenholtz et al., 2016; Przybilla et al., 2020; Stott et al., 2020; Turner et al., 2014). Corrected GNSS measurements allow measuring imaging position with high accuracy; however, determining sensor orientation remains challenging. Stöcker et al. (2017) emphasize that this imbalance must be met by weighting the parameters accordingly. Potential errors in the pre-calibration or direct georeferencing propagate into the final model when interior and exterior parameters are fixed in the bundle adjustment (Cramer et al., 2000). Without exterior constraint the bundle adjustment can compensate for erroneous pre-calibration by shifting camera positions (Cramer et al., 2000; Eltner & Sofia, 2020). The low errors of the single-scale pre-calibrated cases may be an example where fewer constraints (network geometry and exterior parameters) lead to better results.

An alternative could be adaptive camera calibration that solves a highly constrained bundle adjustment and subsequently removes the constraints to allow the interior parameters to readjust (Zhou et al., 2019). Further, experiments comparing calibration validity over short (landing and take-off) and long (site revisit) periods would be beneficial.

5 | CONCLUSIONS

We have demonstrated an efficient workflow for RPAS-based multi-sensor on-site pre-calibration in geomorphological research. For DEM generation from visible imagery we found the largest potential of pre-calibration in the application of single-scale nadir-only surveys. This type of survey design is particularly common in geomorphological applications and can lead to systematic errors if not handled correctly. Such a dataset (30 m nadir-only) resulted in the largest vertical offsets when applied in a self-calibrated bundle adjustment. When applied using the VMS pre-calibrated camera parameters, however, it resulted in the smallest errors overall. With regard to the application in geomorphological surveys, pre-calibrated nadir-only single-scale designs

can be more efficient in terms of time requirement or area covered. However, our results are not a comprehensive verdict to dismiss the use of self-calibration. Sufficiently accurate DEMs were created in the cases that applied a convergent image network containing varying scales and oblique imagery.

Additionally, we have demonstrated the efficiency and suitability of the approach for the simultaneous pre-calibration of multi-sensor RPAS by creating a thermal orthophoto. Additional spectral bands (e.g., thermal, multispectral, hyperspectral) can provide valuable information for mapping surface properties.

Future work will address the potential to improve AM's pre-calibration capabilities, but also broaden the view and conduct experiments with other calibration software. Our premise was the geometrical instability of consumer-grade sensors; hence we emphasized the in situ calibration. Using repeat surveys, we will test to what extent the calibration validity deteriorates over time.

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Johannes Antenor Senn 🕩 https://orcid.org/0000-0001-6627-2568
Claire L. Walsh 🕩 https://orcid.org/0000-0002-4047-1216
Stephen Addy 🕩 https://orcid.org/0000-0003-4582-0013
Maria-Valasia Peppa 🕩 https://orcid.org/0000-0001-9683-0217

REFERENCES

Agisoft LLC (2021) Agisoft metashape professional, version 1.7.2. Retrieved from http://www.agisoft.com/downloads/installer/

Biass, S., Orr, T.R., Houghton, B.F., Patrick, M.R., James, M.R. & Turner, N. (2019) Insights into pāhoehoe lava emplacement using visible and thermal structure from motion photogrammetry. *Journal of Geophysical Research: Solid Earth*, 124(6), 5678–5695. https://doi.org/10.1029/2019JB017444

Bison, P., Bortolin, A., Cadelano, G., Ferrarini, G., Furlan, K. & Grinzato, E. (2012) Geometrical correction and photogrammetric approach in thermographic inspection of buildings, Proceedings of the 2012 International Conference on Quantitative InfraRed Thermography. QIRT Council. http://qirt.org/archives/qirt2012doi/papers/QIRT-2012-285.pdf

Brown, D.C. (1971) Close-range camera calibration. *Photogrammetric Engineering*, 37(8), 855–866.

Carbonneau, P.E. & Dietrich, J.T. (2017) Cost-effective non-metric photogrammetry from consumer-grade sUAS: implications for direct georeferencing of structure from motion photogrammetry: cost-effective non-metric photogrammetry from consumer-grade sUAS. *Earth Surface Processes and Landforms*, 42(3), 473–486. https://doi.org/10.1002/esp.4012

Chandler, J. (1999) Effective application of automated digital photogrammetry for geomorphological research. *Earth Surface Processes and Landforms*, 24(1), 51–63.

Conte, P., Girelli, V.A. & Mandanici, E. (2018) Structure from motion for aerial thermal imagery at city scale: pre-processing, camera calibration, accuracy assessment. *ISPRS Journal of Photogrammetry and Remote Sensing*, 146, 320–333.

Cramer, M., Przybilla, H.-J. & Zurhorst, A. (2017) Uav cameras: overview and geometric calibration benchmark. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W6, 85–92.

Cramer, M., Stallmann, D. & Haala, N. (2000) Direct georeferencing using GPS/inertial exterior orientations for photogrammetric applications. *International Archives of Photogrammetry and Remote Sensing*, 33, 198–205.

Duró, G., Crosato, A., Kleinhans, M.G. & Uijttewaal, W.S.J. (2018) Bank erosion processes measured with UAV-SfM along complex banklines of a straight mid-sized river reach. *Earth Surface Dynamics*, 6(4), 933–953.

Eichel, J., Draebing, D., Kattenborn, T., Senn, J.A., Klingbeil, L., Wieland, M. & Heinz, E. (2020) Unmanned aerial vehicleâ based mapping of turfâ banked solifluction lobe movement and its relation to material, geomorphometric, thermal and vegetation properties. *Permafrost and Periglacial Processes*, 31(1), 97–109. https://doi.org/10.1002/ppp.2036

Elias, M., Eltner, A., Liebold, F. & Maas, H.-G. (2020) Assessing the influence of temperature changes on the geometric stability of smartphone- and Raspberry Pi cameras. *Sensors*, 20(3), 643.

Eltner, A. & Sofia, G. (2020) Structure from motion photogrammetric technique. In *Developments in Earth Surface Processes*, Vol. 23, Elsevier: Amsterdam; 1–24.

Eltner, A., Baumgart, P., Maas, H.-G. & Faust, D. (2015) Multi-temporal UAV data for automatic measurement of rill and interrill erosion on loess soil. *Earth Surface Processes and Landforms*, 40(6), 741–755. https://doi.org/10.1002/esp.3673

Eltner, A., Kaiser, A., Castillo, C., Rock, G., Neugirg, F. & Abellán, A. (2016) Image-based surface reconstruction in geomorphometry – merits, limits and developments. *Earth Surface Dynamics*, 4(2), 359–389. https://doi.org/10.5194/esurf-4-359-2016

Eltner, A., Mader, D., Szopos, N., Nagy, B., Grundmann, J. & Bertalan, L. (2021) Using thermal and Rgb Uav imagery to measure surface flow velocities of rivers. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B2-2021, 717–722. https://doi.org/10.5194/isprs-archives-xliii-b2-2021-717-2021

Entwistle, N., Heritage, G. & Milan, D. (2018) Recent remote sensing applications for hydro and morphodynamic monitoring and modelling: remote sensing for hydro and morphodynamic monitoring & modelling. *Earth Surface Processes and Landforms*, 43(10), 2283–2291. https://doi.org/10.1002/esp.4378

Erenoglu, R.C., Akcay, O. & Erenoglu, O. (2017) An UAS-assisted multi-sensor approach for 3D modeling and reconstruction of cultural heritage site. *Journal of Cultural Heritage*, 26, 79–90.

Ferrer-González, E., Agüera-Vega, F., Carvajal-Ramírez, F. & Martínez-Carricondo, P. (2020) UAV photogrammetry accuracy assessment for corridor mapping based on the number and distribution of ground control points. *Remote Sensing*, 12(15), 2447.

Forlani, G., Dall Asta, E., Diotri, F., Cella, U.M.d., Roncella, R. & Santise, M. (2018) Quality assessment of DSMs produced from UAV flights georeferenced with on-board RTK positioning. *Remote Sensing*, 10(2), 311. https://doi.org/10.3390/rs10020311

Fraser, C. (2018) Camera calibration considerations for UAV photogrammetry. Oral presentation at ISPRS TC II Symposium Riva del Garda http://www.isprs.org/tc2-symposium2018/images/ISPRS-Invited-Fraser.pdf, 15 November 2018.

Geomsoft (2008) VMS, vision measurement system, photogrammetric software package, version 8.0, August 2008. www.geomsoft.com

G.P.L. Software (2019) CloudCompare, version 2.10.2 (zephyrus). http://www.cloudcompare.org/ (24/02/2019) http://www.cloudcompare.org/

Grammatikopoulos, L., Adam, K., Petsa, E. & Karras, G. (2019) Camera calibration using multiple unordered coplanar chessboards. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W18, 59–66. https://doi.org/10.5194/isprs-archives-xlii-2-w18-59-2019

Griffiths, D. & Burningham, H. (2019) Comparison of pre- and self-calibrated camera calibration models for UAS-derived nadir imagery for a SfM application. *Progress in Physical Geography: Earth and Environment*, 43(2), 215–235. https://doi.org/10.1177/0309133318788964

Harvey, M.C., Rowland, J.V. & Luketina, K.M. (2016) Drone with thermal infrared camera provides high resolution georeferenced imagery of the Waikite geothermal area, New Zealand. *Journal of Volcanology and Geothermal Research*, 325, 61–69. https://doi.org/10.1016/j.jvolgeores.2016.06.014

Harwin, S., Lucieer, A. & Osborn, J. (2015) The impact of the calibration method on the accuracy of point clouds derived using unmanned aerial vehicle multi-view stereopsis. *Remote Sensing*, 7(9), 11933–11953.

Hastedt, H., Luhmann, T., Przybilla, H.-J. & Rofallski, R. (2021) Evaluation of interior orientation modelling for cameras with aspheric lenses and image pre-processing with special emphasis to Sfm reconstruction. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B2-2021, 17–24. https://doi.org/10.5194/isprs-archives-xliii-b2-2021-17-2021

Hieronymus, J. (2012) Comparison of methods for geometric camera calibration. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXIX-B5, 595–599. https://doi.org/10.5194/isprsarchives-xxxix-b5-595-2012

Hugenholtz, C.H., Whitehead, K., Brown, O.W., Barchyn, T.E., Moorman, B.J., LeClair, A. et al. (2013) Geomorphological mapping with a small unmanned aircraft system (sUAS): feature detection and accuracy assessment of a photogrammetrically-derived digital terrain model. *Geomorphology* 194: 16–24. https://linkinghub.elsevier.com/retrieve/pii/S0169555X13001736

Hugenholtz, C., Brown, O., Walker, J., Barchyn, T., Nesbit, P., Kucharczyk, M. et al. (2016) Spatial accuracy of UAV-derived orthoimagery and topography: comparing photogrammetric models processed with direct geo-referencing and ground control points. *Geomatica*, 70(1), 21–30.

James, M.R. & Robson, S. (2012) Straightforward reconstruction of 3D surfaces and topography with a camera: accuracy and geoscience application. *Journal of Geophysical Research*, 117(F3). https://doi.org/10.1029/2011JF002289

James, M.R. & Robson, S. (2014) Mitigating systematic error in topographic models derived from Uav and ground-based image networks: mitigating systematic error in topographic models. *Earth Surface Processes and Landforms*, 39(10), 1413–1420. https://doi.org/10.1002/esp.3609

James, M.R., Robson, S., d'Oleire-Oltmanns, S. & Niethammer, U. (2017) Optimising UAV topographic surveys processed with structure-from-motion: ground control quality, quantity and bundle adjustment. *Geomorphology*, 280, 51–66. https://doi.org/10.1016/j.geomorph.2016.11.021

James, M.R., Robson, S. & Smith, M.W. (2017) 3-D uncertainty-based topographic change detection with structure-from-motion photogrammetry: precision maps for ground control and directly georeferenced surveys. *Earth Surface Processes and Landforms*, 42(12), 1769–1788. https://doi.org/10.1002/esp.4125

James, M.R., Chandler, J.H., Eltner, A., Fraser, C., Miller, P.E., Mills, J.P. et al. (2019) Guidelines on the use of structure from motion photogrammetry in geomorphic research. *Earth Surface Processes and Landforms*, 44(10), 2081–2084. https://doi.org/10.1002/esp.4637

James, M.R., Antoniazza, G., Robson, S. & Lane, S.N. (2020) Mitigating systematic error in topographic models for geomorphic change detection: accuracy, precision and considerations beyond off-nadir imagery. *Earth Surface Processes and Landforms*, 2020, 4878. https://doi.org/10.1002/esp.4878

Jaud, M., Passot, S., Le Bivic, R., Delacourt, C., Grandjean, P. & Le Dantec, N. (2016) Assessing the accuracy of high resolution digital surface models computed by PhotoScan® and MicMac® in sub-optimal survey conditions. *Remote Sensing*, 8(6), 465.

Javadnejad, F., Gillins, D.T., Parrish, C.E. & Slocum, R.K. (2020) A photogrammetric approach to fusing natural colour and thermal infrared UAS imagery in 3D point cloud generation. *International Journal of Remote Sensing*, 41(1), 211–237. https://doi.org/10.1080/01431161.2019.1641241

Javernick, L., Brasington, J. & Caruso, B. (2014) Modeling the topography of shallow braided rivers using structure-from-motion photogrammetry. *Geomorphology*, 213, 166–182. https://linkinghub.elsevier.com/retrieve/pii/S0169555X14000245

Kenefick, J.F. (1972) Analytical self-calibration. *Photogrammetric Engineering*, 38(11), 1117–1126.

Lewis, A., Sare, R., Lewicki, J.L. & Hilley, G.E. (2020) High-resolution imaging of hydrothermal heat flux using optical and thermal structure-from-motion photogrammetry. *Journal of Volcanology and Geothermal Research*, 393, 106818. https://doi.org/10.1016/j.jvolgeores.2020.106818

Lichti, D., Skaloud, J. & Schaer, P. (2008) On the calibration strategy of medium format cameras for direct georeferencing, International calibration and orientation workshop EuroCOW. https://infoscience.epfl.ch/record/128407

Lucieer, A., Robinson, S., Turner, D., Harwin, S. & Kelcey, J. (2012) Using a Micro-Uav for ultra-high resolution multi-sensor observations of Antarctic moss beds. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXIX-B1, 429–433. https://doi.org/10.5194/isprsarchives-xxxix-b1-429-2012

Luhmann, T., Piechel, J. & Roelfs, T. (2013) Geometric calibration of thermographic cameras. In *Thermal Infrared Remote Sensing*, Kuenzer, C., Dech, S. (eds), Vol. 17, Springer: Berlin; 27–42. https://doi.org/10.1007/978-94-007-6639-6_2

Maes, W., Huete, A. & Steppe, K. (2017) Optimizing the processing of UAV-based thermal imagery. *Remote Sensing*, 9(5), 476. http://www.mdpi.com/2072-4292/9/5/476

Mandlburger, G. (2019) Through-water dense image matching for shallow water bathymetry. *photogrammetric Engineering and Remote Sensing*, 85(6), 445–455. https://doi.org/10.14358/PERS.85.6.445

Meinen, B.U. & Robinson, D.T. (2020) Mapping erosion and deposition in an agricultural landscape: optimization of UAV image acquisition schemes for SfM-MVS. *Remote Sensing of Environment*, 239, 111666. https://linkinghub.elsevier.com/retrieve/pii/S0034425720300353

Nesbit, P. & Hugenholtz, C. (2019) Enhancing UAV SfM 3D model accuracy in high-relief landscapes by incorporating oblique images. *Remote Sensing*, 11(3), 239. https://doi.org/10.3390/rs11030239

Oniga, V.-E., Pfeifer, N. & Loghin, A.-M. (2018) 3D Calibration test-field for digital cameras mounted on unmanned aerial systems (UAS). *Remote Sensing*, 10(12), 2017. https://doi.org/10.3390/rs10122017

Pádua, L., Ado, T., Sousa, A., Peres, E. & Sousa, J.J. (2020) Individual grapevine analysis in a multi-temporal context using UAV-based multi-sensor imagery. *Remote Sensing*, 12(1), 139. https://doi.org/10.3390/rs12010139

Peppa, M.V., Hall, J., Goodyear, J. & Mills, J.P. (2019) Photogrammetric assessment and comparison of Dji Phantom 4 Pro and Phantom 4 Rtk small unmanned aircraft systems. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W13, 503–509. https://doi.org/10.5194/isprs-archives-xlii-2-w13-503-2019

Piermattei, L., Carturan, L. & Guarnieri, A. (2015) Use of terrestrial photogrammetry based on structure-from-motion for mass balance estimation of a small glacier in the Italian Alps. *Earth Surface Processes and Landforms*, 40(13), 1791–1802. https://doi.org/10.1002/esp.3756

Przybilla, H.-J., Reuber, C., Bäumker, M. & Gerke, M. (2015) Untersuchungen zur genauigkeitssteigerung von uav-bildflgen. 10.

Przybilla, H.-J., Bäumker, M., Luhmann, T., Hastedt, H. & Eilers, M. (2020) Interaction between direct georeferencing, control point configuration and camera self-calibration for Rtk-based Uav photogrammetry.

International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B1-2020, 485–492. https://doi.org/10.5194/isprs-archives-xliii-b1-2020-485-2020

R Core Team (2020) R: a language and environment for statistical computing. manual, R Foundation for Statistical Computing, https://www.R-project.org/

Remondino, F. & Fraser, C. (2006) Digital camera calibration methods: considerations and comparisons. In collaboration with H.-G. Maas and D. Schneider. https://doi.org/10.3929/ETHZ-B-000158067, http://hdl.handle.net/20.500.11850/158067

Remondino, F., Nocerino, E., Toschi, I. & Menna, F. (2017) A critical review of automated photogrammetric processing of large datasets. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W5, 591–599. https://doi.org/10.5194/isprs-archives-xlii-2-w5-591-2017

Roncella, R. & Forlani, G. (2021) UAV block geometry design and camera calibration: a simulation study. *Sensors*, 21(18), 6090.

Samper, D., Santolaria, J., Brosed, F.J., Majarena, A.C. & Aguilar, J.J. (2013) Analysis of Tsai calibration method using two- and three-dimensional calibration objects. *Machine Vision and Applications*, 24(1), 117–131. https://doi.org/10.1007/s00138-011-0398-9

Sanz-Ablanedo, E., Chandler, J.H., Ballesteros-Pérez, P. & Rodríguez-Pérez, J.R. (2020) Reducing Systematic dome errors in digital elevation models through better UAV flight design. *Earth Surface Processes and Landforms*, 45(9), 2134–2147. https://doi.org/10.1002/esp.4871

Senn, J.A., Mills, J.P., Miller, P.E., Walsh, C., Addy, S., Loerke, E., Senn, J.A., Mills, J.P., Miller, P.E., Walsh, C., Addy, S., Loerke, E. & Peppa, M.V. (2020) On-site geometric calibration of thermal and optical sensors for UAS photogrammetry. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B1-2020, 355–361. https://doi.org/10.5194/isprs-archives-xliii-b1-2020-355-2020

Shortis, M. (2015) Calibration techniques for accurate measurements by underwater camera systems. *Sensors*, 15(12), 30810–30826.

Shortis, M. (2019) Camera calibration techniques for accurate measurement underwater. In *3D Recording and Interpretation for Maritime Archaeology*, McCarthy, J. K., Benjamin, J., Winton, T., van Duivenvoorde, W. (eds), Vol. 31, Springer: Berlin; 11–27. https://doi.org/10.1007/978-3-030-03635-5_2

Shortis, M. & Luhmann, T. (2018) Camera calibration tutorial. Oral presentation at ISPRS TC II Symposium Riva del Garda https://www.isprs.org/tc2-symposium2018/, 15 November 2018.

Shortis, M. & Robson, S. (2015) VMS, vision measurement system, Photogrammetric software package, version 8.8. Geometric Software P/L, www.geomsoft.com

Shortis, M.R., Clarke, T.A. & Robson, S. (1995) Practical testing of the precision and accuracy of target image centering algorithms. In *Photonics East '95*, El-Hakim, S. F. (ed), Hakim: Philadelphia, PA; 65–76.

Slocum, R.K., Parrish, C.E. & Simpson, C.H. (2020) Combined geometric–radiometric and neural network approach to shallow bathymetric mapping with UAS imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 169, 351–363. https://doi.org/10.1016/j.isprsjprs.2020.09.002

Stöcker, C., Nex, F., Koeva, M. & Gerke, M. (2017) Quality assessment of combined IMU/GNSS data for direct georeferencing in the context of UAV-based mapping. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W6, 355–361. https://doi.org/10.5194/isprs-archives-xlii-2-w6-355-2017

Stott, E., Williams, R.D. & Hoey, T.B. (2020) Ground control point distribution for accurate kilometre-scale topographic mapping using an RTK-GNSS unmanned aerial vehicle and SfM photogrammetry. *Drones*, 4(3), 55.

Turner, D., Lucieer, A., Malenovský, Z., King, D. & Robinson, S. (2014) Spatial Co-registration of ultra-high resolution visible, multispectral and thermal images acquired with a micro-UAV over Antarctic moss beds. *Remote Sensing*, 6(5), 4003–4024. https://doi.org/10.3390/rs6054003

Wackrow, R. & Chandler, J.H. (2008) A convergent image configuration for DEM extraction that minimises the systematic effects caused by an inaccurate lens model. *Photogrammetric Record*, 23(121), 6–18. https://doi.org/10.1111/j.1477-9730.2008.00467.x

Webster, C., Westoby, M., Rutter, N. & Jonas, T. (2018) Three-dimensional thermal characterization of forest canopies using UAV photogrammetry. *Remote Sensing of Environment*, 209, 835–847.

Westfeld, P., Mader, D. & Maas, H.-G. (2015) Generation of TIR-attributed 3D point clouds from UAV-based thermal imagery. *PFG*, 2015(5), 381–393.

Woodget, A.S., Carbonneau, P.E., Visser, F. & Maddock, I.P. (2015) Quantifying submerged fluvial topography using hyperspatial resolution UAS imagery and structure from motion photogrammetry. *Earth Surface Processes and Landforms*, 40(1), 47–64. https://doi.org/10.1002/esp.3613

Woodget, A.S., Dietrich, J.T. & Wilson, R.T. (2019) Quantifying below-water fluvial geomorphic change: the implications of refraction correction, water surface elevations, and spatially variable error. *Remote Sensing*, 11(20), 2415.

Zhou, Y., Rupnik, E., Meynard, C., Thom, C. & Pierrot-Deseilligny, M. (2019) Simulation and analysis of photogrammetric UAV image blocks influence of camera calibration error. *Remote Sensing*, 12(1), 22. https://doi.org/10.3390/rs12010022

APPENDIX
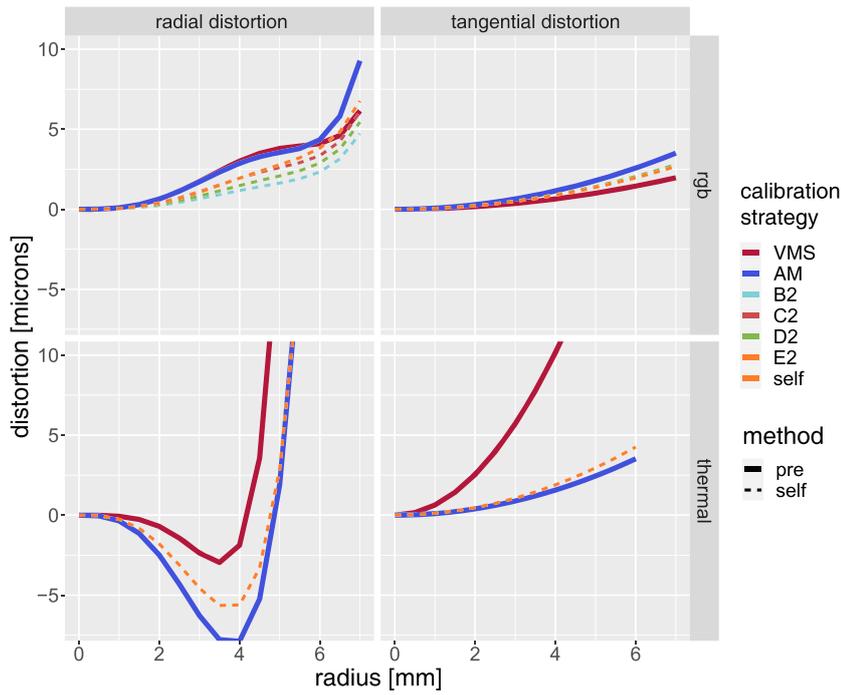
FIGURE A1 Lens distortion profiles of the DJI Matrice M600 RGB sensor and the Workswell WIRIS pro thermal sensor for all pre- and self-calibrated cases [Color figure can be viewed at wileyonlinelibrary.com]
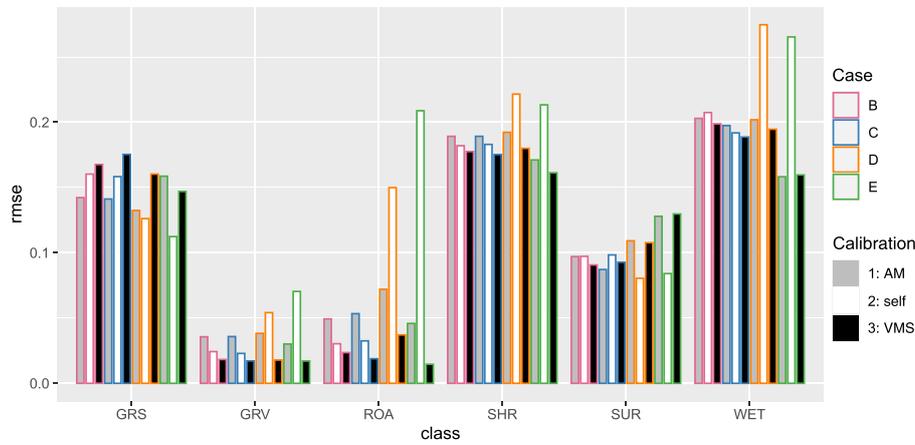

FIGURE A2 Bar chart of the RMSE values of the z-offsets between the GNSS measurement and the calculated DEMs at the reference points. See Table 3 for data [Color figure can be viewed at wileyonlinelibrary.com]
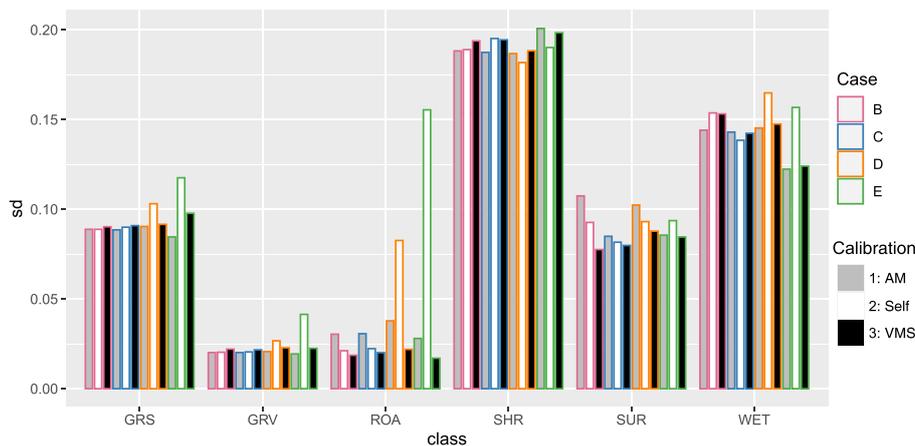
FIGURE A3 Bar chart of the standard deviation in the z-offsets between DEM and the GNSS reference points. See Table A2 for raw data [Color figure can be viewed at wileyonlinelibrary.com]

FIGURE A4 z-offsets between the DEM cases and the GNSS reference points (GRS, GRV, ROA, SHR, SUR and WET) [Color figure can be viewed at wileyonlinelibrary.com]
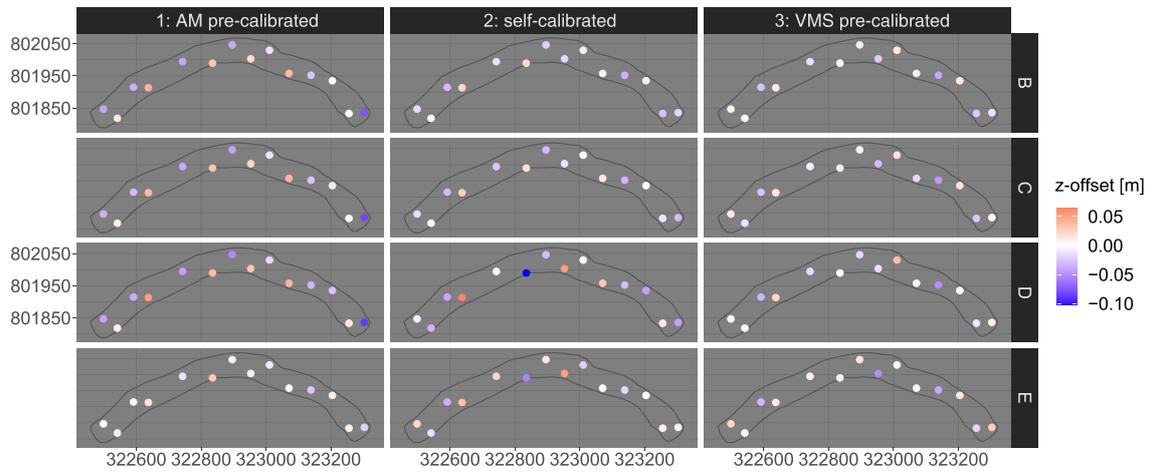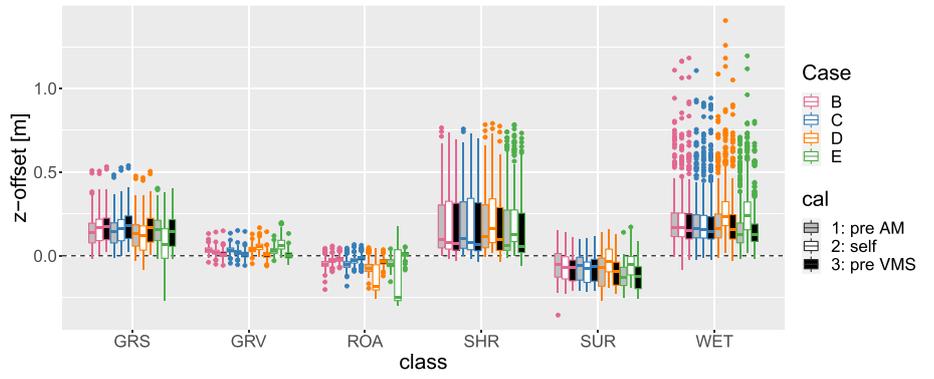
FIGURE A5 Spatial distribution of z-offsets between the DEM cases and the GCPs [Color figure can be viewed at wileyonlinelibrary.com]

TABLE A1 The calibration parameters

	Sensor	ID	f	x_p	y_p	k_1	k_2	k_3	p_1	p_2
Pre-cal	Thermal	VMS	13.055	2.43×10^{-1}	1.50×10^{-1}	-6.25×10^{-5}	-9.03×10^{-6}	6.94×10^{-7}	5.24×10^{-4}	3.58×10^{-4}
		AM	13.084	-7.87×10^{-2}	3.87×10^{-2}	-3.75×10^{-4}	1.59×10^{-5}	-1.30×10^{-8}	-4.59×10^{-5}	-8.64×10^{-5}
	rgb	VMS	8.831	8.90×10^{-3}	-2.33×10^{-2}	9.23×10^{-5}	-3.48×10^{-6}	3.99×10^{-8}	-1.71×10^{-5}	-3.64×10^{-5}
		AM	8.844	8.80×10^{-3}	-2.32×10^{-2}	9.54×10^{-5}	-4.02×10^{-6}	5.36×10^{-8}	2.58×10^{-5}	6.67×10^{-5}
Self-cal	rgb	B2	8.846	1.12×10^{-2}	-2.42×10^{-2}	-3.48×10^{-5}	1.32×10^{-6}	-1.83×10^{-8}	9.17×10^{-6}	5.58×10^{-5}
		C2	8.845	1.06×10^{-2}	-2.47×10^{-2}	-5.81×10^{-5}	2.18×10^{-6}	-2.78×10^{-8}	1.08×10^{-5}	5.51×10^{-5}
		D2	8.521	-6.00×10^{-4}	-2.13×10^{-2}	-4.28×10^{-5}	1.56×10^{-6}	-2.06×10^{-8}	1.20×10^{-5}	5.57×10^{-5}
		E2	8.330	-3.00×10^{-3}	-1.83×10^{-2}	-5.43×10^{-5}	1.88×10^{-6}	-2.39×10^{-8}	1.17×10^{-5}	5.33×10^{-5}
	Thermal	T	13.163	-1.26×10^{-1}	3.47×10^{-2}	-2.66×10^{-4}	1.04×10^{-5}	4.62×10^{-8}	1.10×10^{-4}	-4.32×10^{-5}

TABLE A2 Standard deviations of the z-offsets between DEM and the GNSS reference points. See Figure A3 for visualization as bar chart

DEM case	GRS	GRV	ROA	SHR	SUR	WET
B1	0.09	0.02	0.03	0.19	0.11	0.14
B2	0.09	0.02	0.02	0.19	0.09	0.15
B3	0.09	0.02	0.02	0.19	0.08	0.15
C1	0.09	0.02	0.03	0.19	0.08	0.14
C2	0.09	0.02	0.02	0.20	0.08	0.14
C3	0.09	0.02	0.02	0.19	0.08	0.14
D1	0.09	0.02	0.04	0.19	0.10	0.15
D2	0.10	0.03	0.08	0.18	0.09	0.16
D3	0.09	0.02	0.02	0.19	0.09	0.15
E1	0.08	0.02	0.03	0.20	0.09	0.12
E2	0.12	0.04	0.16	0.19	0.09	0.16
E3	0.10	0.02	0.02	0.20	0.08	0.12