

Optimising UAV topographic surveys processed with structure-from-motion: Ground control quality, quantity and bundle adjustment

James, M. R.^a, Robson, S.^b, d'Oleire-Oltmanns, S.^c and Niethammer, U.^d

^a Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YQ, U.K.

(corresponding author: m.james@lancs.ac.uk, +44 (0)1524 593571)

^b Department of Civil, Environmental and Geomatic Engineering, University College London, Gower Street, London, WC1E 6BT, U.K.

^c Department of Geoinformatics - Z_GIS, University of Salzburg, , A-5020 Salzburg, Austria.

^d Universität Stuttgart, Institut für Geophysik, Azenbergstr. 16, 70174 Stuttgart, Germany.

Abstract

Structure-from-motion (SfM) algorithms greatly facilitate the production of detailed topographic models from photographs collected using unmanned aerial vehicles (UAVs). However, the survey quality achieved in published geomorphological studies is highly variable, and sufficient processing details are never provided to understand fully the causes of variability. To address this, we show how survey quality and consistency can be improved through a deeper consideration of the underlying photogrammetric methods. We demonstrate the sensitivity of digital elevation models (DEMs) to processing settings that have not been discussed in the geomorphological literature, yet are a critical part of survey georeferencing, and are responsible for balancing the contributions of tie and control points. We provide a Monte Carlo approach to enable geomorphologists to (1) carefully consider sources of survey error and hence increase the accuracy of SfM-based DEMs and (2) minimise the associated field effort by robust determination of suitable lower-density deployments of ground control. By identifying appropriate processing settings and highlighting photogrammetric issues such as over-parameterisation during camera self-calibration, processing artefacts are reduced and the spatial variability of error minimised.

We demonstrate such DEM improvements with a commonly-used SfM-based software (PhotoScan), which we augment with semi-automated and automated identification of ground control points (GCPs) in images, and apply to two contrasting case studies – an erosion gully survey (Taroudant, Morocco) and an active landslide survey (Super-Sauze, France). In the gully survey, refined processing settings eliminated step-like artefacts of up to ~50 mm in amplitude, and overall DEM variability with GCP selection improved from 37 to 16 mm. In the much more challenging landslide case study, our processing halved planimetric error to ~0.1 m, effectively doubling the frequency at which changes in landslide velocity could be detected. In both case studies, the Monte Carlo approach provided a robust demonstration that field effort could be substantially reduced by only deploying approximately half the number of GCPs, with minimal effect on the survey quality. To reduce processing artefacts and promote confidence in SfM-based geomorphological surveys, published results should include processing details which include the image residuals for both tie points and GCPs, and ensure that these are considered appropriately within the workflow.

Keywords: UAV; ground control; structure-from-motion; bundle adjustment

1. Introduction

Unmanned aerial vehicle (UAV) surveys are being increasingly used to collect high resolution airborne imagery in a wide variety of environmental and geomorphological environments, including agricultural (Marzolff and Poesen, 2009; Marzolff et al., 2011; d'Oleire-Oltmanns et al., 2012; Eltner et al., 2015), landslide (Niethammer et al., 2012; Lucieer et al., 2014; Turner et al., 2015), coastal (Delacourt et al., 2009; Harwin and Lucieer, 2012; Goncalves and Henriques, 2015), fluvial (Lejot et al., 2007; Hervoue et al., 2011; Flener et al., 2013; Fonstad et al., 2013; Tamminga et al., 2015; Woodget et al., 2015) and glacial (Whitehead et al., 2013; Immerzeel et al., 2014) studies. Typical requirements are to derive surface change from image orthomosaics and detailed digital elevation models (DEMs) which are produced from photogrammetric processing of images.

Such processing is usually carried out with 3-D reconstruction software based on structure-from-motion (SfM) and multi-view stereo (MVS) algorithms. However, a recent review of published geomorphological studies (covering both aerial and ground-based work) demonstrated order of magnitude variations in relative survey measurement quality (Smith and Vericat, 2015). Although some of this variability is likely to result from differences in how error is assessed (Smith and Vericat, 2015), it strongly suggests that substantial improvements in survey design or processing (or both) should be possible in many cases. Furthermore, early ground-based SfM-MVS analysis of spatio-temporal erosion rates suggested that measurement precisions of $\sim 1/1000$ of the viewing distance could be achieved (James and Robson, 2012), but the ratios of root mean square error (RMSE) to viewing distances for >40 published surveys have a median value of $\sim 1/640$ (Smith and Vericat, 2015). Unfortunately, the use of user-friendly SfM-MVS software does not promote consideration of the processing parameters involved, and the lack of details means that it is not possible to understand fully the underlying sources of error in geomorphological studies. Here, we show that considering some of the photogrammetric principles involved will enable geomorphologists to truly unlock the potential of SfM-based surveys and better exploit the millimetre-to-centimetre resolution of UAV imagery. Through illustrating the sensitivity of DEMs to the values used for processing settings, we highlight the additional information that should be provided with surveys in order to increase confidence in the results.

Georeferencing forms a fundamental part of topographic surveys and, for SfM-MVS work, dense deployments of carefully-measured ground control points (GCPs) are generally used, which can represent a substantial proportion of the overall survey effort. However, it is critical that control data are incorporated suitably within the image processing in order to avoid adversely affecting DEM accuracy, and this includes weighting their contribution appropriately within the processing, and ensuring that any outliers in either survey or image measurement data are identified and eliminated. SfM-MVS software does not generally provide the detailed quality assessment diagnostics necessary for rigorous photogrammetric analysis. Thus, a comprehensive understanding of DEM accuracy and the contribution of control measurements can be difficult to achieve, hindering generic improvements in GCP deployment and processing. One solution is to use SfM to initialise processing with conventional aerial (Rosnell and Honkavaara, 2012) or oblique photogrammetric (James and Robson, 2012) software, from which detailed analyses can be obtained, or to use software that combines aspects of both SfM and photogrammetry (e.g. MicMac, Pierrot-Deseilligny and Clery, 2011). However, the use of such integrated approaches presents an additional (steep) learning curve that will deter many SfM-MVS users.

Thus, here, we develop a Monte Carlo approach in order to (1) improve DEM accuracy and reproducibility, and (2) enable reduced field survey effort through a better understanding of GCP

contributions with any of the above methods, but particularly when using SfM-MVS image processing alone. Underpinning our work is the appropriate handling of measurement error within SfM-MVS workflows; this is often poorly understood by users and somewhat hidden within ‘black box’ software. For example, to the authors’ knowledge, the values of settings used to describe the precision of image measurements have never been reported in SfM-geoscience literature, nevertheless, these are critical to obtaining accurate and repeatable results. To illustrate the importance of such settings, SfM-MVS software (PhotoScan, v.1.1.6) was used with manufacturer-recommended default values (see section 3.1) to process a case study in this work, representing a reasonably typical UAV survey. Two DEMs were produced, one processed with and one processed without independently measured check points. Within the region covered by GCPs, the DEMs showed systematic step-like differences with amplitudes of up to ~50 mm, which corresponded with changes in image overlaps (Fig. 1a, b). Similar variations were augmented by broader differences between DEMs generated with different selections of equally well distributed GCPs as control points (Fig. 1c; with RMS difference between the GCP-covered regions of the DEMs being 39 mm). In this case, the strong systematics resulted in volume differences in the western side of the survey of 173 m^3 ($112 \text{ m}^3 \text{ ha}^{-1}$), with oppositely signed volume change on the east side of 676 m^3 ($360 \text{ m}^3 \text{ ha}^{-1}$). These processing artefacts would represent important bias when comparing repeat surveys for understanding processes such as soil erosion. Such problems result from excessive influence of the relatively few ‘marker’ points used to identify GCPs positions in the images, due to the default processing settings heavily over-weighting GCP observations within the processing. Consequently, without full reporting of all processing settings values used for surveys, the likelihood of DEM shapes being substantially adversely affected by inappropriate settings cannot be discounted.

DEM accuracy assessments are often presented through calculating the RMSE on check points, but the use of such statistics alone does not expose the presence of spatially systematic (i.e. non-random) error (Kyriakidis et al., 1999). The critical importance of spatially correlated error in DEMs is well known (e.g. Weng, 2002; Oksanen and Sarjakoski, 2006; Gong-Saholiariliva et al., 2011; Chu et al., 2014) and has been shown to have strong effects on calculated erosion rates and sediment budgets (Wheaton et al., 2010; Milan et al., 2011; Lallias-Tacon et al., 2014), and on products such as slope, aspect, water content and watershed maps (Hunter and Goodchild, 1997; Oksanen and Sarjakoski, 2006; Wechsler and Kroll, 2006; Hebel and Purves, 2009; Goulden et al., 2016). Thus, an accurate understanding of error magnitude and spatial variability, along with minimisation of otherwise hidden and systematic error (e.g. Fig.1), is critical for advancing the use of SfM-MVS surveys for detailed topographic studies.

Here, our aims are to clarify the photogrammetric factors and processing settings that affect DEM accuracy and repeatability, describe the conventional quality checks that are not currently being used in geomorphological work, but can identify problems, and show how future surveys can be optimised to minimise the field effort involved. We provide software tools to enable advanced quality analysis when using SfM-MVS through validating the camera model, identifying over-parameterisation and assessing the quality and required quantity of GCPs. We demonstrate our approach by implementing it for one SfM-MVS software, but the fundamental concepts are applicable equally to any photogrammetric or SfM-MVS-based workflow. The paper is organised as follows: after providing background to the underpinning photogrammetric procedures involved, we describe our Monte Carlo-based approach and detail its application to an erosion gully survey carried out with an automated fixed-wing UAV. Finally, we demonstrate its utility in a much more challenging active landslide survey, acquired using a manually controlled quadcopter, for which the effective stability of the ground control could not be guaranteed.

2. UAV surveys, photogrammetric ground control and bundle adjustment

Over the last decade, advances in UAV hardware and firmware have rapidly enhanced image collection capabilities through improved stability and flight time of platforms, with flight planning software now enabling survey design and optimised image capture. Conventional aerial survey strategies can thus be used as a basis for UAV survey planning (Fig. 2), although platform and requirement differences usually mean that not all conventional recommendations will be appropriate and they can seldom be implemented in full. For example, UAVs are typically used over substantially smaller spatial scales (e.g. of order 10^4 – 10^6 m²) than conventional aerial surveys, which facilitates GCP deployment and field measurement. However, with a few exceptions (Turner et al., 2014; Eling et al., 2015; Mian et al., 2015), UAVs are not currently equipped with sufficient quality on-board GNSS receivers to enable camera positions to be used as useful additional control measurements for high-accuracy work so, on a per-image basis, more GCPs should be expected to be required than for modern conventional aerial surveys (i.e. similar to Fig. 2a rather than Fig. 2b). To understand the influence of GCPs on DEM quality, the underlying photogrammetric processing workflow must be considered.

Most modern photo-based topographic surveys are processed by software automatically identifying and matching corresponding features in different images, to generate a network of ‘tie point’ image observations. Photogrammetric processing of this network then determines 3-D point coordinates and camera positions through a least-squares network optimisation called a ‘bundle adjustment’ (Granshaw, 1980). ‘Self-calibrating’ bundle adjustments additionally estimate an optical model for the camera (e.g. focal length and lens distortion parameters) as part of this process. With only image-based information, the bundle adjustment results define a 3-D surface shape up to an arbitrary scale and orientation, given by the 3-D tie point positions. Incorporation of external control measurements into the adjustment has two effects; it defines the datum (i.e. the absolute orientation and scale of the DEM within the external coordinate system) and also adds constraints which, if well distributed across the project, can effectively mitigate systematic errors in the DEM shape that can accumulate across spatially extended networks (Wackrow and Chandler, 2008; James and Robson, 2014a). However, if measurement uncertainty within the control data is not properly accounted for, then adding control can introduce error into DEMs (e.g. Fig. 1).

The formal treatment of error throughout the bundle adjustment process is a fundamental feature of rigorous photogrammetry and is implemented through a stochastic error model (Mikhail et al., 2001; Luhmann et al., 2006). The error model enables each control measurement to be associated with a precision estimate (e.g. a standard deviation for each X, Y and Z coordinate, for each control point) and for these estimates, along with precision estimates for the image observations, to suitably weight the constraints within the adjustment. If control precision is overestimated (i.e. control point measurements are considered to be more precise than they actually are) then DEM shape will be distorted by overfitting to error in the control measurements (at the expense of the information provided by the tie points, e.g. Fig. 1a), and overall project accuracy will be reduced. In contrast, underestimating control precision down-weights its influence in the bundle adjustment and reduces the effectiveness of control in mitigating the systematic errors that may occur if tie points alone are used. Importantly, the appropriate weighting of tie and control point image observations can be as critical as assigning a suitable precision to the GCPs field survey measurements.

In rigorous bundle adjustment approaches, the relative weights of image observations are automatically adjusted so that the processed network precision matches an a priori estimate (e.g. Mikhail et al., 2001; Luhmann et al., 2006) and updated precision estimates are provided for all

parameters in the form of variance covariance matrices. With self-calibrating bundle adjustments, precision estimates for the recovered camera parameters can be used to ensure that camera models are not over-parameterised, by checking for statistical significance and for correlations between parameters. For example, where a recovered parameter value does not exceed its estimated precision, or where it is shown to be highly correlated with another, the parameter can be removed from the adjustment, its value fixed to zero, and the adjustment reprocessed (Granshaw, 1980). Whilst issues such as camera model over-parameterisation are well-known in the photogrammetry community (e.g. Granshaw, 1980) they are not being actively considered in SfM-MVS geoscience applications. In SfM-MVS software, automatic re-weighting may not be implemented in the bundle adjustment, and camera model parameter precisions and covariances are not generally available; consequently, users need to be aware of the potential DEM errors, and of alternative approaches that can identify problems. Thus, although UAVs offer real advantages for delivering frequent, accurate and high resolution DEMs and orthomosaics but substantial opportunities remain for optimising deployment and image processing, particularly for studies requiring high accuracy and precision, those that are carried out in difficult environments, and repeat surveys that seek to evaluate change in landform surfaces over time.

3. Methods

To improve the generation of DEMs and orthomosaics from multi-image photogrammetric networks, we used a three-stage processing approach to integrate established photogrammetric best-practice with Monte Carlo analysis (Fig. 3). We describe the generalised procedures first, then detail their specific implementation in Section 3.4. For clarity, we use the following terminology throughout the paper:

- ‘image observations’ refer to coordinates of points identified in images (e.g. in pixels),
- ‘image residuals’ are the error estimates on these image observations determined by bundle adjustment,
- ‘GCPs’ describe any identified points for which 3-D ground coordinates are measured in the field, with their different uses within a bundle adjustment distinguished by describing them as either ‘control points’ or ‘check points’,
- ‘GCP measurements’ refer to these real-world 3-D coordinates of GCPs measured in the field and
- ‘errors on GCP’ are the differences between the GCP measurements and their equivalent coordinates derived within the photogrammetric model.

3.1. Stage I. Image observations and initial network quality control

Images were initially oriented by SfM processing and a self-calibrating bundle adjustment, to define the network of tie points and camera orientations. Self-calibration was limited to adjust camera parameters comprising principal distance (focal length), principal point, and three radial distortion terms. Although including only one or two radial terms in the parameter set should be suitable for most standard consumer cameras, this was not an option in the software used. All images acquired with the same camera and lens (or lens settings) shared the same camera model.

The resulting tie points were inspected to identify major matching errors for individual points (i.e. points that were clearly located off the surface, or had anomalously large image residuals) or images with overall greater, or strongly systematic, image residuals, which could be

removed from the network. Key tools in such a process are histogram-based analysis and the visualisation of image residuals across each image, so that any systematic patterns can be easily identified. For this work, care was taken not to remove ‘good’ tie points that displayed elevated residuals due to the camera model rather than because of poor feature matches. The resulting tie point image residuals gave a measure of the accuracy of tie point observations in the image set.

The positions of the GCPs in the images were then identified, and these observations included in the bundle adjustment without including their associated ground coordinates as control points in the adjustment. This method of processing is often termed a ‘free’ or ‘inner constraints’ bundle adjustment (Granshaw, 1980). In this way, the quality of the GCP image observations was ascertained and their measurement precision estimated, prior to them being associated with ground coordinates.

Finally, to verify the consistency of the image network with the ground control measurements, the GCP observations were then linked to their 3-D ground coordinates and a further bundle adjustment carried out, using all the GCP as control points. The discrepancies between the coordinates estimated by the bundle adjustment and the surveyed GCP measurements were then assessed for any obvious outliers, and the associated GCPs removed. At the completion of Stage I, the image network should be internally consistent with high quality observations of ground control observations, without major errors due to poorly matched tie points or misidentified GCPs.

3.2. Stage II. Self-calibration (camera model optimisation)

With a consistent set of network observations and GCP measurements, the influence of different camera model parameter sets was then assessed. Sets investigated were sequentially more complex, for example, using focal length only, then focal length with principal point and three radial parameters (Table 1). 50% of GCPs were allocated randomly as check points for error assessment, and bundle adjustments carried out for each of the four camera models, with parameters being estimated as unknowns within the adjustment. In order to generalise the results for any selection of particular GCPs used as control, the bundle adjustment was repeated in a Monte Carlo framework, with the selection of 50% of the GCP population varying for each different adjustment. The overall influence of the control within the process was also assessed by varying the prescribed 3-D measurement accuracy of the GCPs.

Table 1. Camera models and their parameters.

| Camera model | Parameters included |
|--------------|---|
| A | Focal length |
| B | Focal length, principal point, radial distortion (3 parameters) |
| C | Focal length, principal point, radial distortion (3 parameters), tangential distortion (2 parameters) |
| D | Focal length, principal point, radial distortion (4 parameters), tangential distortion (2 parameters), aspect ratio, skew |

A suitable camera model was determined by evaluating which camera models improved the overall network performance. A parameter (or parameter set) was included in the final camera

model provided that doing so decreased the RMSE magnitudes on both the check and control points, and either decreased, or did not substantially increase, the ratio between them (RMSE check points/RMSE control points). Once an appropriate parameter set was identified, parameter values were determined by a final self-calibrating bundle adjustment, using all GCP as control. It should be noted that Stages I and II can be somewhat iterative; once a suitable camera model is determined, then initial rejected observations and GCPs can be re-assessed, and included back in the network if their associated errors and residuals are suitably reduced.

3.3. Stage III. Ground control quality and quantity

With the camera model fixed, the effect of GCP density on the network accuracy and the influence of individual control points could be determined. To represent a particular control density, an appropriate number of GCPs were randomly selected to be retained as check points, and a bundle adjustment was carried out. The process was then repeated with a different random selection of GCPs as control. By carrying out this Monte Carlo analysis with different numbers of GCPs as control points, the compiled results could be used to assess variation in the network accuracy (in terms of error on control and check points) over a range of different control densities. Hence, a ‘virtual’ check point error could be estimated for the case in which all GCPs are used as control, and a minimum GCP density estimated to satisfy a particular accuracy requirement. Furthermore, by considering the error on any individual GCP over a number of Monte Carlo iterations, the general performance of individual points, when allocated as either control or check point, could be assessed.

3.4. Implementation

Processing was implemented (Fig. 3) using a combination of PhotoScan Professional (v.1.1.6, for image processing, bundle adjustment, DEM and orthoimage generation), Python (integrated into PhotoScan, for Monte Carlo scripting; the script is provided as Supplementary Material), Matlab and *sfm_georef* (v.3.0 (James and Robson, 2012), for automated GCP measurement, visualisation of image residuals, post processing and visualisation of results; freely available at <http://tinyurl.com/sfmgeoref>).

To construct the image network, images were automatically matched and oriented in PhotoScan (using the ‘align images’ function, with PhotoScan’s photo alignment settings of ‘high’ accuracy, generic pair pre-selection and the default per-image key and tie point limits of 40,000 and 1,000 respectively), resulting in a sparse cloud of tie points. GCP observations were then collected using the camera orientation data and tie points to facilitate semi-automated and automated processing using a new matching algorithm implemented in *sfm_georef*. Automated GCP identification is initiated by manually identifying the GCP position in an image, which allows the GCP to be represented by a 3-D square planar patch, textured from the image and oriented parallel to the local surface (as defined by the 3-D positions of nearby tie points). Other images are then automatically searched by projecting the patch into the images (to account for differences in viewing geometry) and carrying out a sub-pixel normalised cross-correlation search in the local area. With two or more observations of a GCP, its 3-D coordinates are then derived by ray intersection. Once the 3-D positions of three or more GCPs have been identified, an initial georeferencing transform can be derived between the SfM and the GCP coordinate systems and thus, the locations of the other GCPs can be estimated throughout the image set. If GCPs are symmetric and identical, then image texture from one representative GCP can be used to refine these positions using the 3-D patch-based cross-correlation procedure. For best use of this automated GCP identification, control targets should have a central high-contrast feature (ideally symmetrical, i.e. a circle), on a plain background sufficiently large that it exceeds the intended

patch size. For example, in our main case study, 0.24-m-diameter circular targets were used, representing image features of between ~6 and 8 pixels across (Fig. 4a, inset). The derived GCP image coordinates were then imported into the PhotoScan project for incorporation in its bundle adjustment process.

Initial network quality assessments (Fig. 3, Stage I) were carried out using a combination of point cloud editing in PhotoScan and visualisation of image residuals in *sfm_georef*. The subsequent Monte Carlo analyses (Fig. 3, Stages II & III) were carried out in PhotoScan, using a Python script (see Supplementary Material) to automate repeated bundle adjustments, varying the settings and exporting results to file at each iteration. PhotoScan does not provide any algorithm details for how it carries out ‘bundle adjustment’; the process is referred to as an ‘optimization’ which can be used to refine camera models and reduce non-linear project deformations by incorporating control data. In the ‘optimization’ process, three settings are used to describe the relative accuracies of tie, control and check points. Although strictly, these are precisions rather than accuracies, we retain the use of ‘accuracy’ here for consistency with the terminology used in the software. ‘Marker accuracy’ reflects the measurement precision of the 3-D GCP coordinates (in metres), ‘projection accuracy’ (in pixels) reflects the precision in image space that GCP observations are made to and ‘tie point accuracy’ is the equivalent measure for tie points. PhotoScan (v.1.1.6) default values for these settings are 0.1 pixels for ‘projection accuracy’, 4.0 pixels for ‘tie point accuracy’ and 5.0 mm for ‘marker accuracy’, which is also the manufacturer-advised value if GCP measurement accuracy is better than 20 mm ([http://www.agisoft.com/pdf/PS_1.1-Tutorial\(BL\)-Orthophoto,DEM\(withGCP\).pdf](http://www.agisoft.com/pdf/PS_1.1-Tutorial(BL)-Orthophoto,DEM(withGCP).pdf)).

In the work presented here, each camera model or selected set of PhotoScan metrics was assessed using 50 different random selections of GCPs as control points. Where DEMs were generated, dense point clouds were generated in PhotoScan (using settings of ‘high’ quality and ‘aggressive’ depth filtering), which were then meshed (using ‘height field’ surface type, and ‘high’ face count, with interpolation enabled). Texture for orthophotos was constructed using ‘adaptive orthophoto’ mapping and ‘mosaic’ blending modes. DEMs were exported at 0.1 m resolution, orthophotos at 0.05 m resolution.

4. Case study: Taroudant

We illustrate the application of our approach through a survey carried out ~6 km West of Taroudant, in the Souss-Massa-Drâa region of Morocco, for determining soil erosion and monitoring gully development in badland and agricultural environments (Fig. 4; d'Oleire-Oltmanns et al., 2012). Survey requirements were thus aimed at enabling detection of sub-decimetre surface change for calculating localised volumetric loss in gullies; however, change over wider areas is also of interest in such areas because gullies are often artificially infilled from surrounding material as part of land management. Imagery was acquired from a Sirius I fixed-wing UAV (MAVinci, Germany) which, following a conventional pattern of aerial photogrammetric practice, autonomously flew sets of inclined parallel flight lines, predefined using flight-planning software (Fig. 1a; Fig. 7 of d'Oleire-Oltmanns et al., 2012). Images were captured by a Panasonic Lumix GF1 digital camera (4000 × 3000 pixels) triggered at regular intervals by the flight-planning software. The GF1 used a fixed focal length 20 mm pancake lens which, from the flight height of ~100 m, projected a ground footprint of ~23 × 23 mm for each pixel. Ground control was provided through the deployment of 30 circular targets on square boards (0.3 × 0.3 m), with their coordinates measured by total station to horizontal precisions of 5 mm (*X* and *Y*) and 10 mm vertically (*Z*), (d'Oleire-Oltmanns et al., 2012). In the images, the circular targets were typically ~6–8 pixels in diameter resulting in clearly identifiable targets (Fig.

4a). Note that, in order to highlight the problems that can arise when considering areas outside the extent of the ground control, we retain these regions within the figures (e.g. Fig. 1, 4).

Imagery dominantly covered bare soil which provided good, geometrically consistent image texture. This, in combination with a high degree of overlap throughout the image block enabled PhotoScan to automatically assimilate all images into an initial model. After manually identifying three GCPs in `sfm_georef`, an estimated georeferencing transform was calculated, and the circular symmetry of the target shape enabled fully automated identification of all the remaining GCPs. Several large image residuals signalled misidentifications (e.g. circular bushes nearby a target), which were quickly corrected by adjusting the initial position of the GCP observation onto the target and re-running the correlation localisation.

4.1. *Image observations and initial network quality control*

The initial sparse cloud of photogrammetric tie points showed no visually identifiable outliers in the estimated 3-D tie point positions, but two photos had anomalously high RMS image residual magnitudes (Fig. 5a). Visualising the image residual vectors for these photos (e.g. Fig. 5b) showed that they were highly systematic, reflecting underlying errors in camera interior or exterior orientation. Thus, these two photos were removed from all subsequent processing of the network.

Re-running the bundle adjustment with the GCP observations included (but not their 3-D coordinates) indicated that the measurement quality and network geometry of the GCP image observations was in line with that of the tie points (Fig. 5c). However, when the GCP observations were linked to their ground coordinates and used as control measurements, one GCP (ID #28) showed a somewhat elevated error magnitude (Fig. 5d, e). Nevertheless, it was not sufficiently poor to classify as an outlier and reject at this stage, thus all GCP image observations and ground coordinate measurements were accepted as consistent for further processing.

To illustrate the sensitivity of the network to the PhotoScan bundle adjustment accuracy settings, 15 dispersed GCPs (50%) were selected for use as check points, and bundle adjustments carried out over ranges of values (Fig. 6). For ‘tie point accuracy’ and ‘projection accuracy’ settings, ranges of 0.1–4.0 pixels and 0.1–1.0 pixels respectively resulted in RMSE on control points of ~10–60 mm (with RMSE being greatest in either the horizontal or vertical direction, depending on the settings values), reflecting the varying weight assigned to the control observations (Fig. 6b). In contrast, RMSE on check points was relatively consistent at ~55 mm. For PhotoScan’s default settings, the ratio of check point to control point error magnitudes was a factor of ~6 (Fig. 6), reflecting a strong spatial variation in error and a likely overweighting of the GCP observations (i.e. Fig. 1b). Tie point image residuals showed standard deviations of 0.47 pixels in the x -direction, and 0.51 in y , this combination suggested a value for the ‘tie point accuracy’ setting of 0.69 pixels. The equivalent residuals analysis on GCP the observations gave a ‘projection accuracy’ of 0.50 pixels. With these values, the check to control point error ratio was reduced to ~1.3 (Fig. 6c), reflecting a better balance between tie and control point observations.

4.2. *Self-calibration (camera model optimisation)*

The Monte Carlo self-calibrating bundle adjustment analyses demonstrated characteristic features (Fig. 7) with different camera model parameter sets (Table 1). Errors on control and check points decreased as radial (Fig. 7b) and then tangential parameters were included in the self-calibration (Fig. 7c). The ratio between mean RMSE on the check points and mean RMSE on the control points also decreased, indicating that the additional parameters were effective and not

overfitting the control data. However, if the most complex camera model was used (Model D in Table 1), then error ratios increased (Fig. 7d), indicative of overfitting.

For large and small values of ‘marker accuracy’, RMSE values plateau (Fig. 7). At large values (e.g. ‘marker accuracy’ >0.5 m), larger height (Z) errors were illustrative of systematic doming deformation within the dataset. This effect decreased to negligible levels when three radial and tangential distortions were optimised but, interestingly, it returned when the most complex camera parameters were optimised (Model D), as a further indicator of overfitting. Estimated camera parameters varied systematically with ‘marker accuracy’ (for example, focal length, as shown by the shading in the final plot column of Fig. 7). The fact that camera Model C, using radial and tangential distortion components (Table 1), led to the smallest range in estimated focal lengths (a variation of $\sim 0.06\%$) further supported it as the optimum parameter set. Thus, the results supported the use of camera Model C, and final parameter values were ascertained from an adjustment with all GCPs used as control.

4.3. *Ground control quality and quantity*

With the camera calibration fixed, results from bundle adjustments carried out with 15 or more GCPs used as control showed a general insensitivity to control density, with RMSE on control and check points being relatively similar (Fig. 8a). For any particular control density, the spread of RMSE values reflected variations due to the specific selection of GCPs for control or check points, and the inverse correlation between RMSE values on control and on check points (Fig. 8b) indicated that there was no generic control distribution which minimised overall error. With fewer GCPs used as control, error magnitudes on control and check points diverged; reducing the number of control points made their constraints increasingly easy to accommodate within the bundle adjustment, resulting in small RMSE values on control (Fig. 8a). However, using only a few GCP as control points (including random configurations in close spatial proximity) provided poor constraint for the wider network, so corresponding check point discrepancies were large.

Analysing the results on a per-GCP basis for adjustments carried out with >80% of GCP as control (to avoid any influence of spatially weak control distributions), provided a case for reconsidering GCP #28 as an outlier (Fig. 8c, d). Mean error magnitudes on GCP #28 were noticeably greater than on other GCPs (particularly when used as a check point), and evidence of its broader influence was provided by overall RMSE values being biased by its selection for check or control within the adjustments carried out with 15 GCPs (Fig. 8b). Furthermore, by including GCP #28 as a control point, the resulting DEM showed a systematic tilt of ~ 20 mm over its width when compared with DEMs built without GCP #28 (Fig. 8e). After verifying that this was not due to a problem in the image observations for this GCP (Fig. 8f), which had an acceptable RMS residual of 0.55 pixels, GCP #28 was excluded from all subsequent analyses. Consequently, the Monte Carlo adjustments were repeated, and produced slightly reduced RMSE values (Fig. 8g). The RMSE on GCPs when all were used as control was 41.7 mm and, by visual extrapolation of the Monte Carlo results, the corresponding value for check points would be expected to be ~ 45 mm.

For small numbers of control points, for which variability in the Monte Carlo results was dominated by the control point spatial distribution rather than the number of points, DEM accuracy was also assessed using ‘stronger’ control point spatial distributions based on conventional aerial methodologies (Fig. 2, 9). Under these conditions, RMSE on check points was approximately constant at ~ 50 mm (Table 2), and independent of the number of control points. Thus, tie points alone appeared to be highly effective at determining the camera orientations (and also the camera model, for self-calibrating adjustments) and, consequently, at defining the DEM

shape. However, comparing the associated DEMs with a benchmark DEM processed with all GCPs as control (Fig. 9) illustrated that considering the RMSE values alone concealed systematic DEM differences including linear slopes and doming deformation, which was particularly apparent if camera self-calibration was used. In some cases, such effects may be mitigated where areas of unchanging topography could be used to refine alignments between surveys, or permanent GCPs could be deployed, but otherwise their systematic nature could present substantial problems when trying to determine relatively small but widely distributed surface changes. Minimal distributions of GCPs (numbering 4 to 5 points for control purposes, with additional check points) could be considered if an accurate camera calibration was available.

Table 2. Variation in RMSE with the number of GCPs used as control points.

| Num. control points | RMSE on GCPs (mm) | | | |
|---------------------------|-------------------|-------------------|--------------|-----------|
| | Control points | | Check points | |
| | Fixed cal. | Self-cal. | Fixed cal. | Self-cal. |
| 0 ^a | 35.1 ^a | 36.7 ^a | 49.3 | 49.1 |
| 4 | 31.2 | 31.2 | 49.0 | 50.1 |
| 5 | 31.2 | 31.3 | 48.2 | 48.3 |
| 10 | 35.7 | 35.7 | 47.4 | 48.4 |
| 15 | 35.9 | 36.0 | 49.3 | 49.8 |

^a No GCPs were included within the bundle adjustment, but four were used to subsequently scale and orient the model; all other GCPs were considered as check points.

4.4. Survey improvements

Through using appropriate values for tie point and GCP image precision and robust detection of outlier GCP measurements, our improved processing removed the ~50 mm sensitivity to check points and image overlap (Fig. 1) and increased DEM repeatability; RMS differences between DEMs processed with different GCPs more than halved from 37 mm (Fig. 1c) to 16 mm (Fig. 9), which could be considered as an improvement in precision by a factor of ~2.3. Appropriate settings values resulted in the RMSE on check points decreasing from 56 mm to 47 mm; representing an accuracy improvement of ~1.2. When compared with other SfM-MVS surveys, the results are towards the upper end of expected survey quality, giving a ratio of check point RMSE to the mean viewing distance of ~1:2,100, and exceeding the ratio of ~1:640 shown to characterise a more recent review of 43 aerial and ground-based datasets (Smith and Vericat, 2015). The enhanced processing will provide valuable improvements for studying relatively rapid processes such as rill or gully erosion and, by decreasing flight height to $\lesssim 20$ m, sub-centimetre precisions should be achievable. This would enable future surveys to compete directly with laser scanning in terms of measurement quality, whilst retaining the advantages of low cost and spatial coverage, and would enable study of more difficult-to-detect processes such as inter-rill erosion (e.g. Eltner et al., 2015).

The robust GCP analysis also indicated that field effort could be substantially reduced with minimum impact on survey precision. For similar surveys, the number of GCPs deployed could be halved or, if permanent GCPs are installed at the survey extremities (e.g. Fig. 2b) and the imaging network strengthened (e.g. by integrating oblique imagery; James and Robson, 2014a; Harwin et al., 2015), then even fewer GCPs would be required (e.g. four, with some additional check points).

5. Case study: Super-Sauze

To demonstrate utility of the approach in a challenging survey, we used a study of the Super-Sauze landslide, France, associated with a relatively dense distribution of control (Fig. 10). This survey represents an interesting test due to uncertainties in the control data, internal camera stability and the variable quality and scale of the images. The Super-Sauze landslide is a persistently active slide within the Barcelonnette Basin (Southern French Alps) where the processes involved in movement are being assessed (Niethammer et al., 2010; Niethammer et al., 2012; Stumpf et al., 2013). The landslide movement is characterised by highly variable surface displacement rates which are generally $\sim 5\text{--}30\text{ mm day}^{-1}$, changing with seasonal hydrological conditions, but also varying spatially due to interactions with the stable bedrock (Benoit et al., 2015). To investigate horizontal movement, project requirements include orthomosaics with resolutions and accuracies of order centimetres to decimetre, as well as repeatable DEMs. 291 GCP targets comprising $\sim 0.4 \times 0.6\text{ m}$ rectangular sheets were deployed on the slide (Fig. 10b-d) and their coordinates measured with differential GNSS. Due to the uncertainties in locating the centre of the relatively large rectangular GCPs, and their relative movement over the survey period, a value of 0.05 m was assumed for their measurement precision.

Imagery of the whole sliding area ($\sim 850 \times 250\text{ m}$) was acquired using a remote controlled quadcopter in October 2009, with multiple flights required over a period of several days due to equipment and logistical limitations. Since the quadcopter was manually controlled, regular flight lines could not be maintained. As a result, images from the on-board Praktica Luxmedia 8213 compact camera were acquired at flying heights ranging from $\sim 20\text{ m}$ to 250 m above the surface. Imaging was carried out in a time-lapse mode (one photo every three seconds) with automated metering, the aperture fixed to give shutter speeds of $\sim 1/800\text{ s}$, and the inbuilt zoom lens set to its widest angle prior to each flight. 1486 photographs were taken, of which 405 were selected for processing. However, the photograph EXIF data showed that, of the selected images, most (359) were taken using a focal length of 6.2 mm, but others (32) were taken with a focal length of 7.6 mm, suggesting that the lens had inadvertently moved during camera mounting for one flight, and indicating a lack of internal camera stability. Consequently, two camera models were used during processing.

This case study was selected as an interesting test due to the uncertainties in the control data and the variable quality and scale of the images. Despite the challenging imagery, PhotoScan successfully assimilated all the images into a 3-D model. For determining GCP observations, variability within the GCP targets and their lack of symmetry meant that a fully automatic search in `sfm_georef` could not be used. Nevertheless, identification of each GCP in one image enabled correlation-based location in other images and substantially reduced the time taken for collection of GCP observations.

5.1. Image observations and initial network quality control

Surprisingly, the initial image network did not show any photos with anomalously elevated image residuals (Fig. 11a), although image residuals were generally greater than those of the Taroudant project. Thus, all photos were retained but $\sim 50\%$ of the tie points were then removed because they either clearly fell far from the 3-D surface or they had large image residuals (>2 pixels), which reduced the initial 82,499 tie points to 41,721. The presence of vegetation on one side of the flow is likely to have been partly responsible for large residuals, but deleted points were notably located across the entire project and were not restricted to vegetated areas. Following removal of these tie points, the overall RMS image residual magnitude was reduced from 1.28 to 0.88 pixels.

When the 291 GCPs were included as control in the bundle adjustment, the distribution of error on the control points showed an extended tail, with some discrepancy magnitudes even exceeding 0.7 m (see inset histogram, Fig. 11c). Examination of the GCPs contributing to the tail showed that they were spatially distributed across the entire project area. In order to mitigate these data a conservative threshold of 0.22 m (just greater than four times the expected GCP accuracy) was used to remove 47 GCPs, leaving 244 GCPs within the network (Fig. 11d), which had an RMS of image observation residuals of 0.86 pixels, in line with those on the tie points. The removal of so many GCPs reflected the challenging nature of the project and was probably due to areas where the ground movement was noticeably different from that of most of the GCPs.

5.2. *Self-calibration (camera model optimisation)*

For the camera model, all tested parameter sets resulted in step-shaped RMSE box plots (e.g. Fig. 12), indicating that none of the camera models could deliver good DEM shape from tie points alone. Including tangential parameters (Model C) improved results over those achieved from three radial parameters and principal point offsets of Model B (Fig. 12), but the more complex Model D did not demonstrate any further error reduction. Consequently camera Model C was used.

5.3. *Ground control quality and quantity*

When assessing the effect of reduced densities of control points, a gradual increase in error was demonstrated, mainly resulting from deterioration in the horizontal, rather than vertical direction (Fig. 13a). Median horizontal values remain <0.1 m for ~ 100 GCP or more used as control ($\sim 40\%$ in Fig. 13a). A histogram of RMSE on individual GCPs (Fig. 13b) demonstrated a skewed distribution, indicative of underlying problems in the fit to the control data. To examine the spatial distribution of this, DEMs and orthomosaics were compared between a project processed using all 244 GCPs as control and one using a random, but spatially coherent, selection of 122 GCPs (Fig. 14). Although the differences in areas within the bounds of the control show magnitudes in line with those expected (~ 0.1 m, Fig. 13a), they are strongly systematic in both horizontal and vertical components, reflecting the fact that the DEM shape still relies on the GCPs.

5.4. *Survey improvements*

Initial processing of the Super-Sauze survey gave RMSE on control points of 0.19 m in the horizontal (and 0.20 m in the vertical), suggesting that making reliable velocity measurements would require survey intervals of approximately a month or more, in order to resolve the smaller values (e.g. ~ 5 mm day⁻¹). Our enhanced processing resulted in orthomosaic and DEM products which were demonstrated as good to ~ 0.1 m (Fig. 13, 14), representing a strong increase in reliability, and doubling the useful survey frequency. Furthermore, the Monte Carlo approach indicated that this performance could be retained with almost half as many GCPs, thus delivering a substantial saving in field effort.

For comparison of the overall relative precision of the survey, the large range of viewing distances negates the relevance of 'mean viewing distance' but, using a value of 94 m determined from the average GCP observation distance, gives a ratio of $\sim 1:750$ (for an estimated GCP measurement accuracy of 125 mm). The Monte Carlo analysis demonstrated that the main limiting factor was a misfit between the control and the image data, likely due to relative movement of the GCPs over the time required for the survey. This highlights the need to understand the physical nature of the site so that GCP provision can be focussed on areas thought to be most stable. For this survey, where imagery were captured over several days, GCPs had to be located on the moving landslide. However, current UAV technology is more likely to enable full image capture over a much shorter period of time. In this case, placing GCPs on surrounding stable features,

supported by only a few check measurements made on the slide and simultaneous with image capture, would be recommended.

6. Discussion

In this work, we have combined conventional photogrammetric quality considerations with a Monte Carlo approach to support repeatability in SfM-MVS DEMs and to assess robust optimisations to minimise field effort in repeat surveys.

6.1. Image observations and initial network quality control

The large changes in control point RMSE in the Taroudant results, as a consequence of changing input control and tie point image observation accuracy settings for processing (Fig. 6), illustrate the importance of the settings values, and erroneous software ‘accuracy’ settings produce complex systematic errors in the computed terrain (Fig. 1). Thus, appropriate observation accuracy values need to be used to avoid introducing processing artefacts into DEMs. Suitable values for both tie and control points can be determined from RMS image residual magnitudes computed from initial exploratory bundle adjustments. With appropriate values, a DEM should only be noticeably influenced by the incorporation of check point observations in the bundle adjustment if the tie points alone are very weak at constraining the DEM shape. Appropriate precision estimates for control points should connect with the expectation of the field survey method used, and it is important to recognise that there may be distinct differences between horizontal and vertical field survey capability.

The magnitude of image residuals will be affected by a wide range of factors including image texture, the camera model, stability of the camera geometry, and the accuracy of the feature localisation algorithm used in the image matching. Vegetation is usually highly challenging for accurate feature matching and, although neither project was entirely vegetation covered, both contained some, and this could also have contributed to residual magnitudes, particularly for Super-Sauze. The smaller initial RMS image residual value for the Taroudant project is consistent with the generally better image quality (crisper images) as well as the relatively constant distance between the camera and the ground surface which, by virtue of a common image to image magnification, should improve the accuracy of feature matches. The varying camera altitude within the Super-Sauze project complicates image matching and is probably a strong factor contributing to the substantially larger initial image residuals.

6.2. Camera model and self-calibration

The importance of an accurate camera model for delivering high quality DEMs is well established, particularly with relatively parallel and flat image acquisition geometries that are typical of aerial (including UAV) surveys. PhotoScan does not provide the detailed precision and covariance statistics required to fully understand the self-calibration performance, but the Monte Carlo framework provided sufficient insight to identify an appropriate selection of camera parameters and illustrate the problems introduced by over-parameterisation (Fig. 7, Model D). A limitation in the version of PhotoScan used is the inability to individually select radial distortion parameters, and three may not always be necessary. For example, laboratory calibration of DSLRs with 28 mm lenses using close-range photogrammetric software indicated that even at network precisions of 1:60,000, only principal point offsets and two radial parameters could be recovered to values greater than their precision estimates (James and Robson, 2014b).

For dense control point deployments, self-calibrating surveys are likely to outperform surveys carried out with pre-calibrated camera models (e.g. Harwin et al., 2015). If reduced GCP distributions are important, pre-calibrated cameras should be used, but calibrations need to be

better than the required network precisions, and this may be difficult to achieve with lower-end consumer cameras, due to the reduced stability of their internal geometry (Wackrow et al., 2007), as illustrated by the Super-Sauze study.

6.3. *Ground control quality and quantity*

The influence of GCPs on an image network depends on the network strength and quality, and its propensity to develop systematic deformations – i.e. how much a DEM shape would change if processed with only tie points. If processing using tie points alone (an ‘inner constraints’ solution in photogrammetry) gives a similar DEM shape to a DEM processed with control points incorporated into the bundle adjustment, then fewer GCPs will be generally necessary, as their presence is required dominantly to define the datum rather than the DEM shape. Weak or noisy image networks, particularly when made with cameras with high levels of lens distortion, will be susceptible to much greater systematic error and, consequently, will require greater control input (i.e. more GCPs). To improve weak photogrammetric surveys, the accuracy of tie point matches could be increased (e.g. through better imagery), or a more convergent image geometry could be used (Wackrow and Chandler, 2008; James and Robson, 2014a; Harwin et al., 2015). The addition of oblique photographs into a survey considerably strengthens the network geometry and reduces error in estimated camera parameters, thus minimising systematic DEM deformation, even if no control points are present (James and Robson, 2014a). Nevertheless, our results highlight the importance of ‘working from the whole to the part’ whereby GCP coverage is designed to extend to the extremities of the area of interest, in order to provide an overarching spatial framework from which DEM data are interpolated within, rather than extrapolated beyond.

The Monte Carlo analysis suggested that, if self-calibration was required, the performance of both of the projects assessed here would not be substantially degraded if ~50% fewer GCPs had been deployed. This would also provide sufficient GCP redundancy to enable use of the Monte Carlo approach for validating quality. Our approach highlights the utility of processing different GCP and check point combinations in order to demonstrate that systematic DEM errors are not present, and leading to a more robust photogrammetric solution.

6.4. *Planning practical ground control deployments*

The case studies illustrate the differing effect of GCP deployments between projects and equipment. Thus, considerations such as GCP density need to be part of the initial survey design, and underpinned by the required internal precision and overall repeatability (i.e. georeferencing accuracy) of the survey, along with the requested spatial resolutions of associated DEMs and orthomosaic products. Design considerations should include the importance of datum alignment to gravity (e.g. for modelling gradient-sensitive processes such as rainfall runoff), the presence of vegetation at the scale of the physical control targets that will obscure or obstruct lines of sight and disrupt image matching with shadow and occlusion, as well as absolute 3-D positioning.

Once survey requirements have been specified, the next step is to determine whether they can be achieved and, if so, what equipment would be needed. Required DEM and orthomosaic resolutions dictate the maximum ground pixel size, from which flight heights can be estimated for particular camera and lens specifications. Image network simulation can then be used to explore performance and determine whether precision requirements can be met (e.g. James and Robson, 2014a; Rupnik et al., 2015).

In most cases, GCP measurement precision should exceed the required survey precision. However, if GCPs are only used to scale and orient the model and not to help define its shape through being included in the bundle adjustment, a lower precision can be used if sufficient numbers of GCPs are available for averaging. With sub-centimetre resolution imagery offering the

potential for precision at millimetre scales, the highest accuracy work may actually be difficult to constrain with GNSS measurement of GCPs, and will require detailed laser or optical surveys (e.g. with a total station or theodolite).

7. Conclusions

SfM-MVS image processing greatly facilitates DEM construction from UAV and similar aerial surveys, but seldom provides the comprehensive bundle adjustment reports that are required to understand fully the errors and parameter sensitivities involved. Consequently, unexpected processing artefacts, such as DEM sensitivity to check points (Fig. 1b), may not be exposed. We demonstrate how problems can be mitigated through the use of a Monte Carlo approach to enable validation of camera parameters and to enhance error characterisation when using SfM-MVS software. The approach also helps identify minimal GCP deployments for future topographic surveys, thus offering substantially reduced field survey effort. In combination with photogrammetric best practice, our results point to the following recommendations for delivering repeatable DEMs from UAV imagery:

- 1) High quality image observations and strong network geometry (e.g. the inclusion of convergent imagery) will facilitate survey accuracy and reduce GCP requirements. Accurate tie point observations are supported by acquiring well-focussed imagery of consistent surface texture (e.g. not vegetation), from similar distances. Image networks should be checked for photos that show anomalously large or systematic residuals, which should be rejected (e.g. Fig. 5b). Tie point and GCP observations should be similarly checked for outliers that can be removed.
- 2) Following bundle adjustment, image observation residual magnitudes should be comparable to the estimated image measurement precision. If an automated comparison is not implemented within the adjustment, then appropriate image observation precision estimates can be determined from the RMS of image residual magnitudes, as a manual step for both tie points and for GCPs. In PhotoScan, these values can be used for 'tie point accuracy' and 'projection accuracy' settings respectively, and the network reprocessed. When reporting the final survey products, the values of all such processing settings used within the bundle adjustment and any precision estimates delivered by the adjustment should be documented.
- 3) For an accurate self-calibrating image network, the spatial distribution of GCPs should cover the area of interest (e.g. Fig. 2a). The needed GCP density will depend on the project accuracy required, the network geometry and the quality of image observations. As an example, to attain a survey precision of ~50 mm for the Taroudant bare-soil surface, imaged from 100 m, at a spatial resolution of ~23 mm per pixel, required 15 GCP (Fig. 8g), representing a minimum mean spacing of ~50 m. For a strong image network geometry (i.e. including oblique images), or a camera that is sufficiently accurately pre-calibrated that self-calibrating bundle adjustment is not required, GCP deployment could be substantially reduced (Fig. 9), particularly if permanent GCP could be installed.
- 4) The overall performance of the imaging network and the GCPs can be characterised by using a repeated Monte Carlo approach, in which the selection and number of GCPs used as control are varied. The results can additionally help identify poor GCP measurements and weak areas of the network, where DEM error may be greater. If necessary, final DEM processing can be carried out using all GCPs as control, with accuracy characterised by a check point error estimated from the Monte Carlo adjustments.

- 5) The current practice of reporting statistical descriptions of DEM error (e.g. RMSE on control and check points) is essential, but should be augmented by visualisation of spatial distribution (e.g. Fig. 5d, 9, 11c, 11d) to protect against hidden and potentially complex systematics (e.g. Fig. 1, 9, 14).

Our case studies demonstrated that, by careful identification of outliers, appropriate camera models and tie/control point weighting, DEM repeatability is improved and systematic processing artefacts with magnitudes of up to ~50 mm removed. For similar, repeated surveys aimed at determining geomorphological process rates, only half the number of GCPs would be required with minimal loss of precision, and further reductions could be made if permanent GCPs could be installed or more rapid UAV sorties carried out in regions of unstable terrain. Although SfM-MVS software greatly facilitates DEM generation, the underlying photogrammetric processing must be considered for its true potential to be unlocked and otherwise hard-to-detect processing artefacts be avoided. Only then will the sub-centimetre-level accuracies required for detailed process studies be possible over hectare-scale spatial extents.

Acknowledgements

For the Super-Sauze dataset, we gratefully acknowledge Sabrina Rothmund for GCP setup, Sam Tizzard for assistance with processing and all colleagues from the OMIV project (Observatoire des Instabilités de Versants) for helpful discussions and their support in the field. For the Taroudant dataset, we acknowledge all colleagues, students and friends participating in the field campaigns in Morocco supporting successful data acquisition: Meike Düspohl, Hassan Ghafrani, Christiane Giudici, Abdellatif Hanna, Johannes König, Ahmed Mellali, Maja Nägle, Werner d'Oleire-Oltmanns (alphabetical order). We thank the Editor and an anonymous reviewer for comments that helped improve the manuscript.

References

- Abdullah, Q., Bethel, J., Hussain, M., Munjy, R., 2013. Photogrammetric project and mission planning. In: J.C. McGlone (Ed.), *Manual of Photogrammetry*. American Society for Photogrammetry and Remote Sensing, pp. 1187-1220.
- Benoit, L., Briole, P., Martin, O., Thom, C., Malet, J.P., Ulrich, P., 2015. Monitoring landslide displacements with the Geocube wireless network of low-cost GPS. *Eng. Geol.* 195, 111-121.
- Chu, H.-J., Chen, R.-A., Tseng, Y.-H., Wang, C.-K., 2014. Identifying LiDAR sample uncertainty on terrain features from DEM simulation. *Geomorphology* 204, 325-333.
- d'Oleire-Oltmanns, S., Marzloff, I., Peter, K.D., Ries, J.B., 2012. Unmanned Aerial Vehicle (UAV) for Monitoring Soil Erosion in Morocco. *Rem. Sens.* 4 (11), 3390-3416.
- Delacourt, C., Allemand, P., Jaud, M., Grandjean, P., Deschamps, A., Ammann, J., Cuq, V., Suanez, S., 2009. DRELIO: An unmanned helicopter for imaging coastal areas. *J. Coastal Res. Special Issue* 56, 1489-1493.
- Eling, C., Wieland, M., Hess, C., Klingbeil, L., Kuhlmann, H., 2015. Development and evaluation of a UAV based mapping system for remote sensing and surveying applications. *Int. Arch. Photogram. Remote Sensing Spatial Info. Sci.* XL-1/W4, 233-239.
- 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 Surf. Process. Landf.* 40 (6), 741-755.
- Flener, C., Vaaja, M., Jaakkola, A., Krooks, A., Kaartinen, H., Kukko, A., Kasvi, E., Hyyppa, H., Hyyppa, J., Alho, P., 2013. Seamless mapping of river channels at high resolution using mobile LiDAR and UAV-photography. *Rem. Sens.* 5 (12), 6382-6407.
- Fonstad, M.A., Dietrich, J.T., Courville, B.C., Jensen, J.L., Carbonneau, P.E., 2013. Topographic structure from motion: a new development in photogrammetric measurement. *Earth Surf. Process. Landf.* 38 (4), 421-430.
- Goncalves, J.A., Henriques, R., 2015. UAV photogrammetry for topographic monitoring of coastal areas. *ISPRS-J. Photogramm. Remote Sens.* 104, 101-111.
- Gonga-Saholiariliva, N., Gunnell, Y., Petit, C., Mering, C., 2011. Techniques for quantifying the accuracy of gridded elevation models and for mapping uncertainty in digital terrain analysis. *Prog. Phys. Geogr.* 35 (6), 739-764.
- Goulden, T., Hopkinson, C., Jamieson, R., Sterling, S., 2016. Sensitivity of DEM, slope, aspect and watershed attributes to LiDAR measurement uncertainty. *Remote Sens. Environ.* 179, 23-35.
- Granshaw, S.I., 1980. Bundle adjustment methods in engineering photogrammetry. *Photogramm. Rec.* 10 (56), 181-207.
- Harwin, S., Lucieer, A., 2012. Assessing the accuracy of georeferenced point clouds produced via multi-view stereopsis from unmanned aerial vehicle (UAV) imagery. *Rem. Sens.* 4 (6), 1573-1599.
- 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. *Rem. Sens.* 7 (9), 11933-11953.
- Hebeler, F., Purves, R.S., 2009. The influence of elevation uncertainty on derivation of topographic indices. *Geomorphology* 111 (1-2), 4-16.
- Hervoue, A., Dunford, R., Piegay, H., Belletti, B., Tremelo, M.L., 2011. Analysis of Post-flood Recruitment Patterns in Braided-Channel Rivers at Multiple Scales Based on an Image Series Collected by Unmanned Aerial Vehicles, Ultra-light Aerial Vehicles, and Satellites. *Giscience & Remote Sensing* 48 (1), 50-73.

- Hunter, G.J., Goodchild, M.F., 1997. Modeling the uncertainty of slope and aspect estimates derived from spatial databases. *Geographical Analysis* 29 (1), 35-49.
- Immerzeel, W.W., Kraaijenbrink, P.D.A., Shea, J.M., Shrestha, A.B., Pellicciotti, F., Bierkens, M.F.P., de Jong, S.M., 2014. High-resolution monitoring of Himalayan glacier dynamics using unmanned aerial vehicles. *Remote Sens. Environ.* 150, 93-103.
- James, M.R., How, P., Wynn, P.M., 2016. Pointcatcher software: analysis of glacial time-lapse photography and integration with multi-temporal digital elevation models. *J. Glaciol.* 62 (231), 159-169.
- James, M.R., Robson, S., 2012. Straightforward reconstruction of 3D surfaces and topography with a camera: Accuracy and geoscience application. *J. Geophys. Res.* 117, F03017.
- James, M.R., Robson, S., 2014a. Mitigating systematic error in topographic models derived from UAV and ground-based image networks. *Earth Surf. Process. Landf.* 39 (10), 1413-1420.
- James, M.R., Robson, S., 2014b. Sequential digital elevation models of active lava flows from ground-based stereo time-lapse imagery. *ISPRS-J. Photogramm. Remote Sens.* 97, 160-170.
- Kyriakidis, P.C., Shortridge, A.M., Goodchild, M.F., 1999. Geostatistics for conflation and accuracy assessment of digital elevation models. *Int. J. Geogr. Inf. Sci.* 13 (7), 677-707.
- Lallias-Tacon, S., Liebault, F., Piegay, H., 2014. Step by step error assessment in braided river sediment budget using airborne LiDAR data. *Geomorphology* 214, 307-323.
- Lejot, J., Delacourt, C., Piegay, H., Fournier, T., Tremelo, M.L., Allemand, P., 2007. Very high spatial resolution imagery for channel bathymetry and topography from an unmanned mapping controlled platform. *Earth Surf. Process. Landf.* 32 (11), 1705-1725.
- Lucieer, A., de Jong, S.M., Turner, D., 2014. Mapping landslide displacements using Structure from Motion (SfM) and image correlation of multi-temporal UAV photography. *Prog. Phys. Geogr.* 38 (1), 97-116.
- Luhmann, T., Robson, S., Kyle, S., Harley, I., 2006. *Close range photogrammetry: Principles, methods and applications*. Whittles, Caitness.
- Marzloff, I., Poesen, J., 2009. The potential of 3D gully monitoring with GIS using high-resolution aerial photography and a digital photogrammetry system. *Geomorphology* 111 (1-2), 48-60.
- Marzloff, I., Ries, J.B., Poesen, J., 2011. Short-term versus medium-term monitoring for detecting gully-erosion variability in a Mediterranean environment. *Earth Surf. Process. Landf.* 36 (12), 1604-1623.
- Mian, O., Lutes, J., Lipa, G., Hutton, J.J., Gavelle, E., Borghini, S., 2015. Direct georeferencing on small unmanned aerial platforms for improved reliability and accuracy of mapping without the need for ground control points. *Int. Arch. Photogram. Remote Sensing Spatial Info. Sci.* XL-1/W4, 397-402.
- Mikhail, E.M., Bethel, J.S., McGlone, J.C., 2001. *Introduction to modern photogrammetry*. John Wiley & Sons, Inc., New York.
- Milan, D.J., Heritage, G.L., Large, A.R.G., Fuller, I.C., 2011. Filtering spatial error from DEMs: Implications for morphological change estimation. *Geomorphology* 125 (1), 160-171.
- Niethammer, U., James, M.R., Rothmund, S., Travelletti, J., Joswig, M., 2012. UAV-based remote sensing of the Super-Sauze landslide: Evaluation and results. *Eng. Geol.* 128, 2-11.
- Niethammer, U., Rothmund, S., James, M.R., Travelletti, J., Joswig, M., 2010. UAV-based remote sensing of landslides. *Int. Arch. Photogram. Remote Sensing Spatial Info. Sci.* XXXVIII, Part 5, 496 - 501.
- Oksanen, J., Sarjakoski, T., 2006. Uncovering the statistical and spatial characteristics of fine toposcale DEM error. *Int. J. Geogr. Inf. Sci.* 20 (4), 345-369.

- Pierrot-Deseilligny, M., Clery, I., 2011. APERO, an open source bundle adjustment software for automatic calibration and orientation of set of images. *Int. Arch. Photogram. Remote Sensing XXXVIII-5/W16*, 269-276.
- Rosnell, T., Honkavaara, E., 2012. Point cloud generation from aerial image data acquired by a quadcopter type micro unmanned aerial vehicle and a digital still camera. *Sensors* 12 (1), 453-480.
- Rupnik, E., Nex, F., Toschi, I., Remondino, F., 2015. Aerial multi-camera systems: Accuracy and block triangulation issues. *ISPRS-J. Photogramm. Remote Sens.* 101, 233-246.
- Smith, M.W., Vericat, D., 2015. From experimental plots to experimental landscapes: topography, erosion and deposition in sub-humid badlands from Structure-from-Motion photogrammetry. *Earth Surf. Process. Landf.* 40 (12), 1656-1671.
- Stumpf, A., Malet, J.P., Kerle, N., Niethammer, U., Rothmund, S., 2013. Image-based mapping of surface fissures for the investigation of landslide dynamics. *Geomorphology* 186, 12-27.
- Tammaing, A., Hugenholtz, C., Eaton, B., Lapointe, M., 2015. Hyperspatial remote sensing of channel reach morphology and hydraulic fish habitat using an unmanned aerial vehicle (UAV): A first assessment in the context of river research and management. *River Research and Applications* 31 (3), 379-391.
- Turner, D., Lucieer, A., de Jong, S.M., 2015. Time series analysis of landslide dynamics using an unmanned aerial vehicle (UAV). *Rem. Sens.* 7 (2), 1736-1757.
- Turner, D., Lucieer, A., Wallace, L., 2014. Direct georeferencing of ultrahigh-resolution UAV imagery. *IEEE Trans. Geosci. Remote Sensing* 52 (5), 2738-2745.
- Wackrow, R., Chandler, J.H., 2008. A convergent image configuration for DEM extraction that minimises the systematic effects caused by an inaccurate lens model. *Photogramm. Rec.* 23 (121), 6-18.
- Wackrow, R., Chandler, J.H., Bryan, P., 2007. Geometric consistency and stability of consumer-grade digital cameras for accurate spatial measurement. *Photogramm. Rec.* 22 (118), 121-134.
- Wechsler, S.P., Kroll, C.N., 2006. Quantifying DEM uncertainty and its effect on topographic parameters. *Photogramm. Eng. Remote Sens.* 72 (9), 1081-1090.
- Weng, Q.H., 2002. Quantifying uncertainty of digital elevation models derived from topographic maps. *Advances in Spatial Data Handling*.
- Wheaton, J.M., Brasington, J., Darby, S.E., Sear, D.A., 2010. Accounting for uncertainty in DEMs from repeat topographic surveys: improved sediment budgets. *Earth Surf. Process. Landf.* 35 (2), 136-156.
- Whitehead, K., Moorman, B.J., Hugenholtz, C.H., 2013. Brief Communication: Low-cost, on-demand aerial photogrammetry for glaciological measurement. *Cryosphere* 7 (6), 1879-1884.
- 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 Surf. Process. Landf.* 40 (1), 47-64.

Figures

Fig. 1. DEM variations resulting from using PhotoScan (v.1.1.6) default or recommended processing settings in a self-calibrating bundle adjustment ('marker accuracy' = 5.0 mm, 'projection accuracy' = 0.1 pixels and 'tie point accuracy' = 4.0 pixels; see Section 3.1 for explanation of terms and Section 4 for details of the UAV survey and image set used). Unless otherwise specified, the GCP symbology used here applies throughout the figures. **(a)** Planimetric distribution of image outlines and GCP locations in the survey. **(b)** Vertical differences between two DEMs generated by processing the image set whilst using 15 GCPs as control points, but either with or without the check points present during processing. **(c)** DEM differences when different GCPs are used as control points – one DEM was generated using the GCPs identified by triangles as control points and the circles as check points, the other visa-versa. The electronic version of this figure is in colour.

Fig. 2. Recommended ground control distributions for conventional aerial surveys (e.g. Abdullah et al., 2013), **(a)** if no camera orientation data are available, and **(b)** if suitably accurate camera orientation data (derived from on-board GNSS and INS measurements) have been recorded.

Fig. 3. Workflow and software.

Fig. 4. Taroudant survey. **(a)** Example image, with a 40×40 pixel inset showing a GCP target (0.3×0.3 m). **(c)** Overview of the resulting orthomosaic from the survey, showing the gully network bracketed by roads to the north and south and increasing vegetation cover to the west (the dashed line shows the approximate outline of the image in (a); see Fig. 1a for outlines of all other the component images). The electronic versions of this, and all subsequent figures, are in colour.

Fig. 5. Assessment of network consistency. **(a)** RMS tie point image residual magnitudes (in pixels) for each image in the network. The reduced image set had images 48 and 73 removed, and the bundle adjustment included tie point and GCP observations. **(b)** Large and systematic tie point residuals in image 48, prior to its removal from the project (image represents ~90×70 m on the ground, with north to the top right, and residual vectors are magnified by ×1000). **(c)** Histograms of all image residual magnitudes for tie point and GCP observations (without the GCP observations being linked to their corresponding 3-D coordinate measurements). The results once all GCP measurements are included in the bundle adjustment, shown as **(d)** 3-D residuals on the GCPs in plan view (vectors give the horizontal residual component magnified by ×500) and **(e)** as a distribution of their full magnitudes.

Fig. 6. The effect of varying 'projection accuracy' and 'tie-point accuracy' in self-calibrating bundle adjustments for which 15 GCPs were used as control points, and 15 used as check points. **(a)** RMS of error magnitudes on control points, and **(b)** on check points. Each surface represents results for a different 'marker accuracy', which is given (in mm) by the labelled value. **(c)** The ratio of check to control point RMS errors is an indication of consistency of error distribution across the survey, with larger values potentially identifying overfitting to the control measurements.

Fig. 7. RMSE on control and check points for Monte Carlo analyses using different camera models. Within the box plots, each distribution represents the results of 50 self-calibrating

bundle adjustments carried out with a specific marker accuracy, but with different random selections of 15 GCPs used as control points. Horizontal bars indicate the median RMSE values, with boxes extending between the 25th and 75th percentiles. The whiskers denote the full range of the results not considered outliers (shown by + symbols). Processing was carried out with different camera models given in Table 1. In the plots of check error / control error, the symbols give the ratios of the mean RMSE check and control values for each marker accuracy, and are shaded by the mean focal length for the recovered cameras.

Fig. 8. The effect of control point selection and density on the RMSE on GCPs. **(a)** The box plots give the results of the Monte Carlo analyses, with each distribution representing the results of 50 bundle adjustments carried out with a specific number of different, randomly selected GCPs used as control points for each adjustment, and the remaining GCPs used as check points. **(b)** The influence of GCP #28 is illustrated by selecting the 50 bundle adjustments in which 15 GCPs were used as control and plotting the RMSE on control against check points. Solid symbols represent adjustments which included GCP #28 as a control point, for the open symbols, GCP #28 was a check point. To assess the performance of all GCPs, the RMSE on each point (separated by use as control or check), for adjustments carried out with 80% or more of the GCPs as control points, can be given as **(c)** a histogram (control are given as semi-transparent bars) **(d)** or individually for each GCP. **(e)** DEM sensitivity to GCP #28 is demonstrated by the vertical differences between a DEM derived using all GCP as control, and one derived using all GCPs except #28. **(f)** The six observations of GCP #28, extracted from the different images. The GCPs are 0.3×0.3 m across, and the vectors show the image residuals and are magnified by ×20. **(g)** The Monte Carlo analysis of **(a)**, but repeated with GCP #28 removed from the survey.

Fig. 9. Vertical differences between DEMs derived using different numbers of GCPs as control points (triangles) and a benchmark DEM (derived using all GCP as control points). In the first column, the four GCPs shown by solid circles were used to scale and orient the model, but were not included in the bundle adjustment as control points. Check points are shown as open circles. For the ‘fixed camera model’ DEMs, the camera parameter values were fixed to those used for the benchmark DEM.

Fig. 10. Super-Sauze survey. **(a)** Image distribution (illustrative outlines of only approximately one in three images are shown for clarity) and GCP locations (triangle symbols). **(b-d)** Example images, with 60×60 pixel insets showing GCP targets (~0.4×0.6 m across). **(e)** Overview of the orthomosaic produced from the survey, overlain with the GCP locations. Labelled dashed lines show the approximate outlines of the images (b-d).

Fig. 11. Assessment of network quality. **(a)** RMS tie point residual magnitudes (in pixels) for each image in the network. **(b)** Histograms of image residual magnitudes for both tie points and GCP observations. Errors on GCPs, for **(c)** the initial dataset (with vectors giving the horizontal discrepancies magnified by ×200) and **(d)** the refined dataset (with vectors magnified by ×500).

Fig. 12. RMSE on control and check points for different camera models. The box and scatter plots are as described for Fig. 7, with the focal length results being from the camera model used for most of the photographs (359 images).

Fig. 13. The effect of control point selection and density on GCP residuals. **(a)** Box plots of RMSE magnitudes on GCPs for Monte Carlo analyses with different proportions used as

control points (as described for Fig. 8). For adjustments carried out with 80% or more of the GCPs used as control points, the RMSE on each point (separated by use as control or check) can be given as **(b)** a histogram, with control as semi-transparent bars or **(c)**, individually for each GCP.

Fig. 14. Horizontal and vertical differences between results processed with all GCPs used as control points, and with 50% used as check points. **(a)** Horizontal displacements between the survey orthomosaics, identified using a normalised cross-correlation approach over a regular grid (processed in Pointcatcher software (James et al., 2016)). **(b)** Vertical differences determined by DEM subtraction.

Supplementary material

File: `Monte_Carlo_BA.py` – This is a Python script for carrying out Monte Carlo ‘optimizations’ in PhotoScan Professional (v.1.1.6), varying the GCPs used as control and the processing settings.

Figure 01

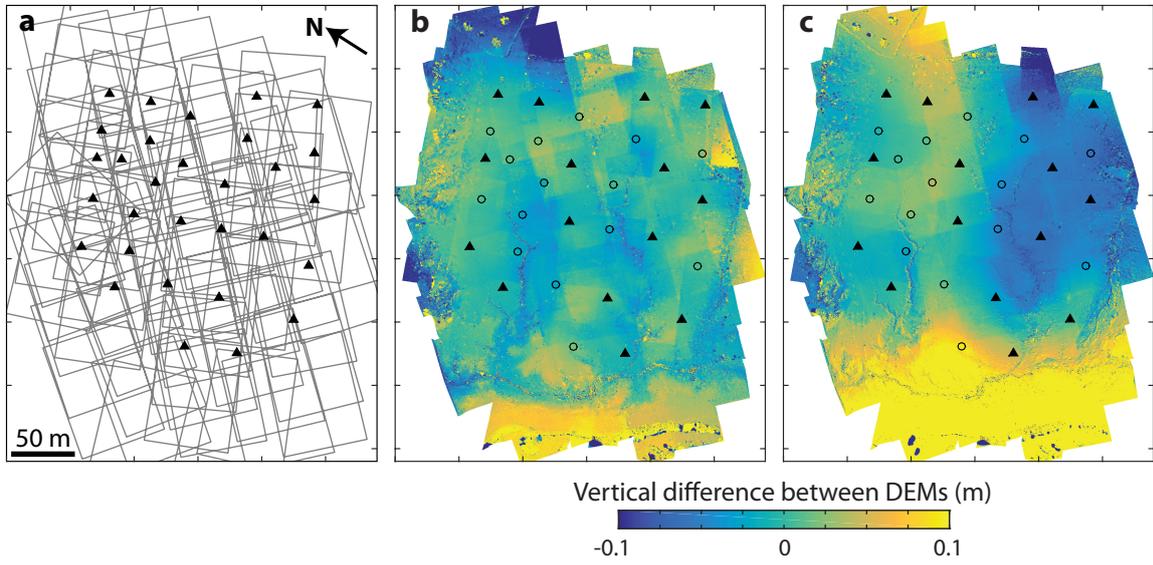
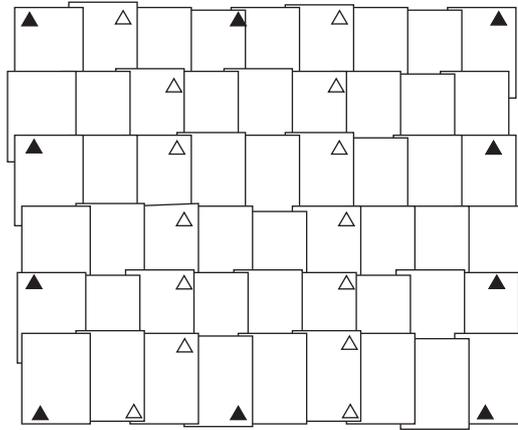
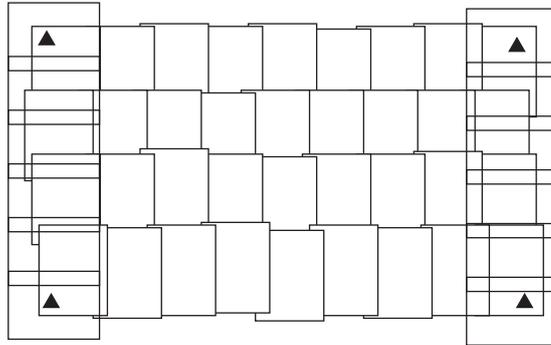


Figure 02

a



b



△ vertical control ▲ vertical and horizontal control

Figure 03

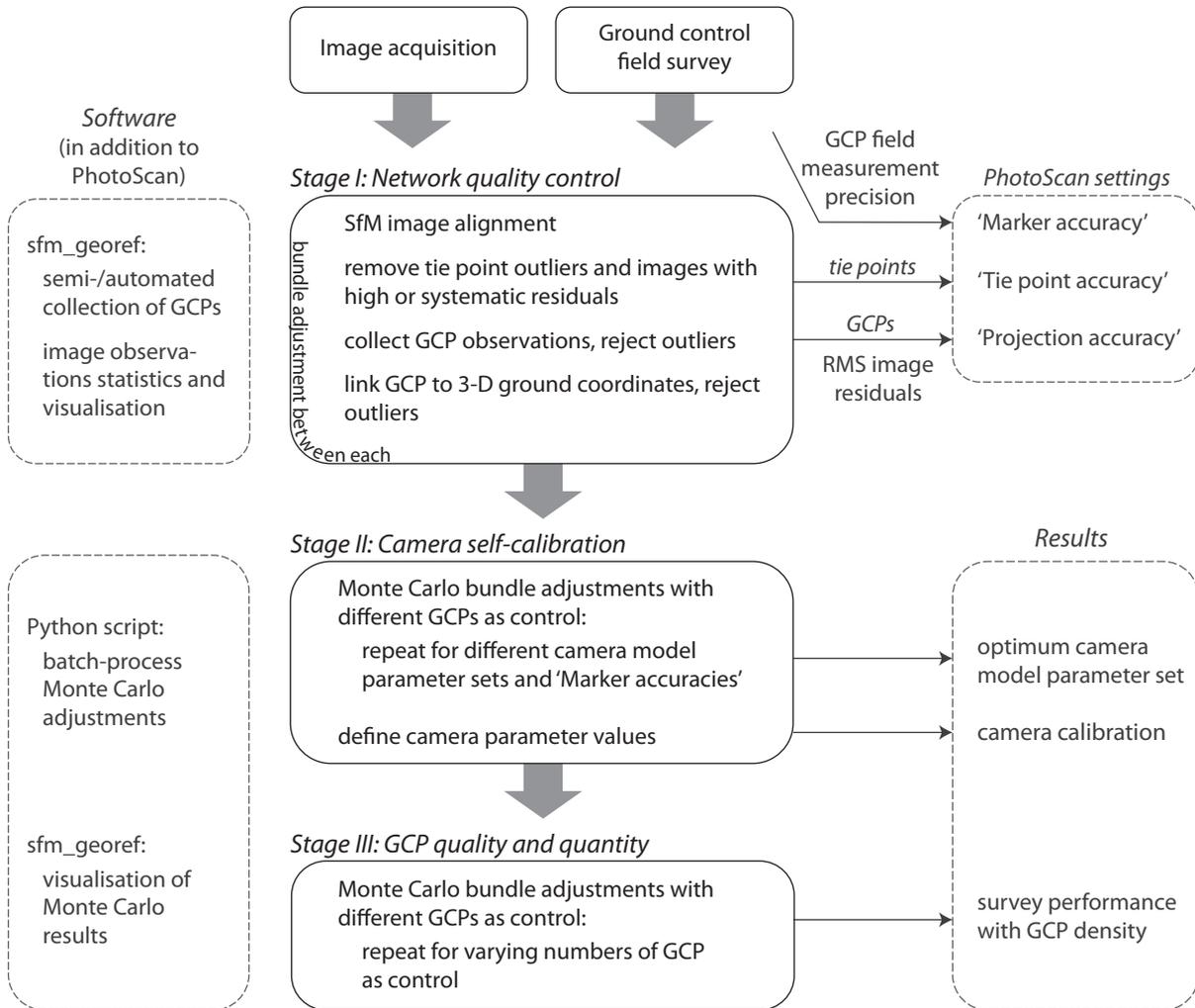


Figure 04

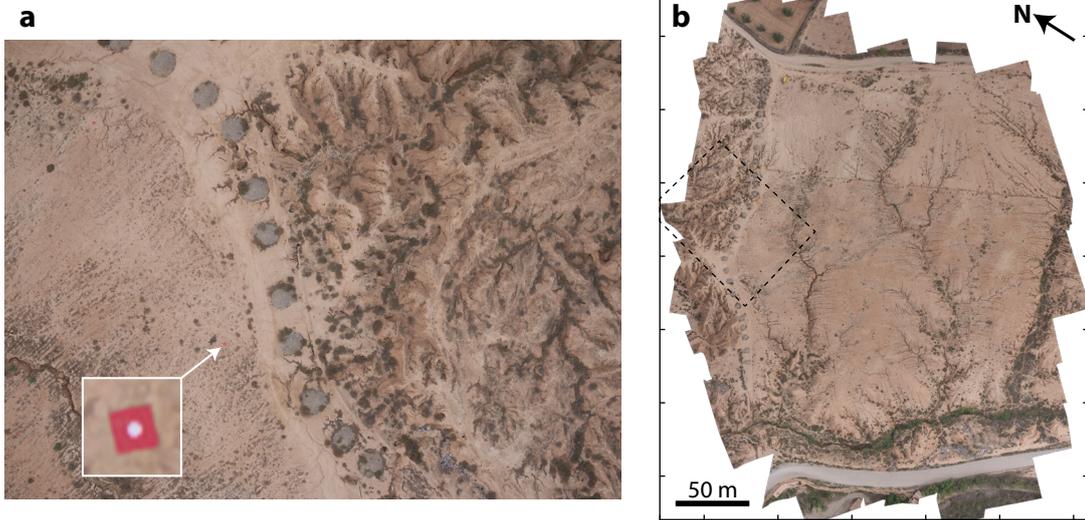


Figure 05

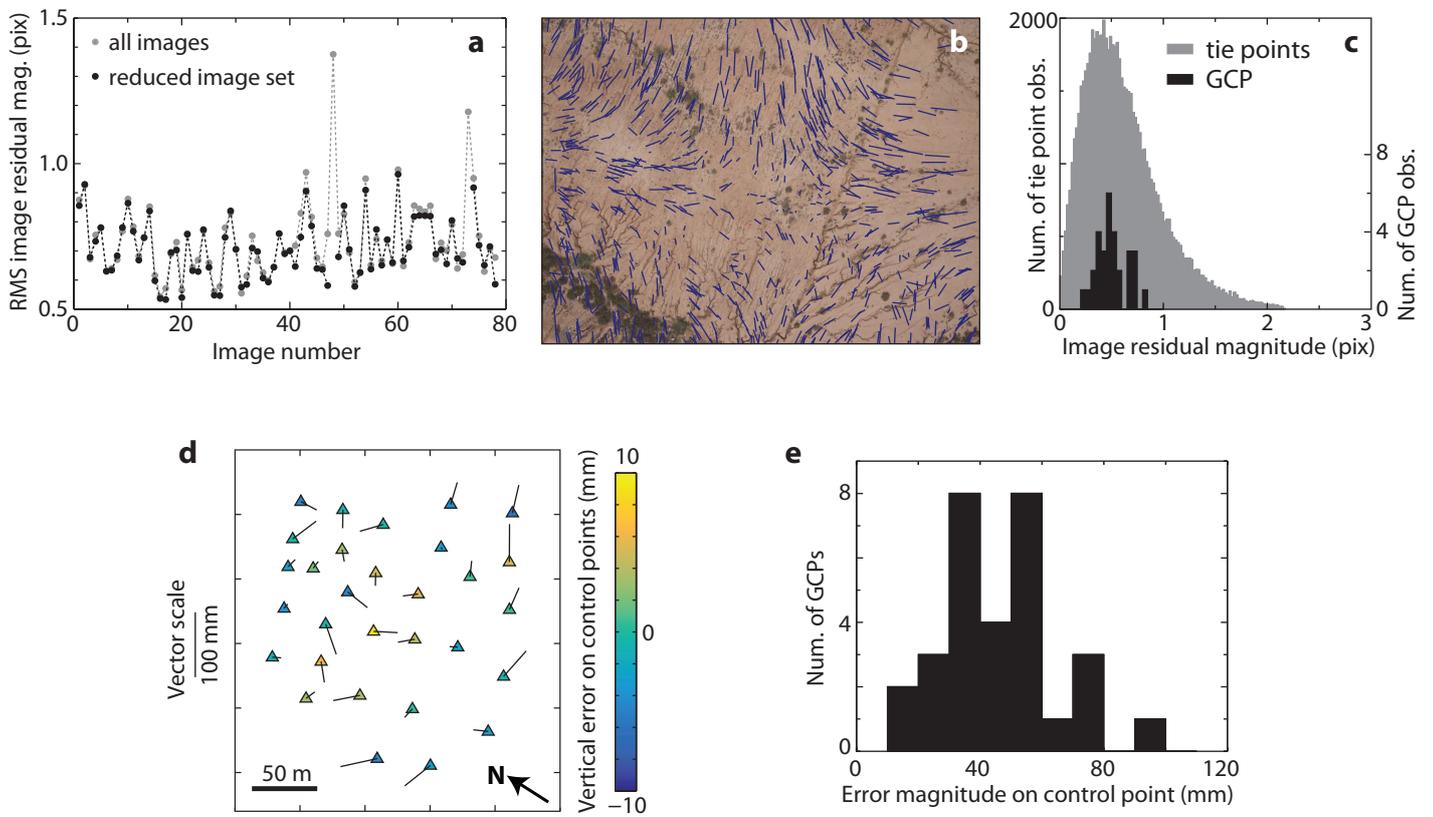


Figure 06

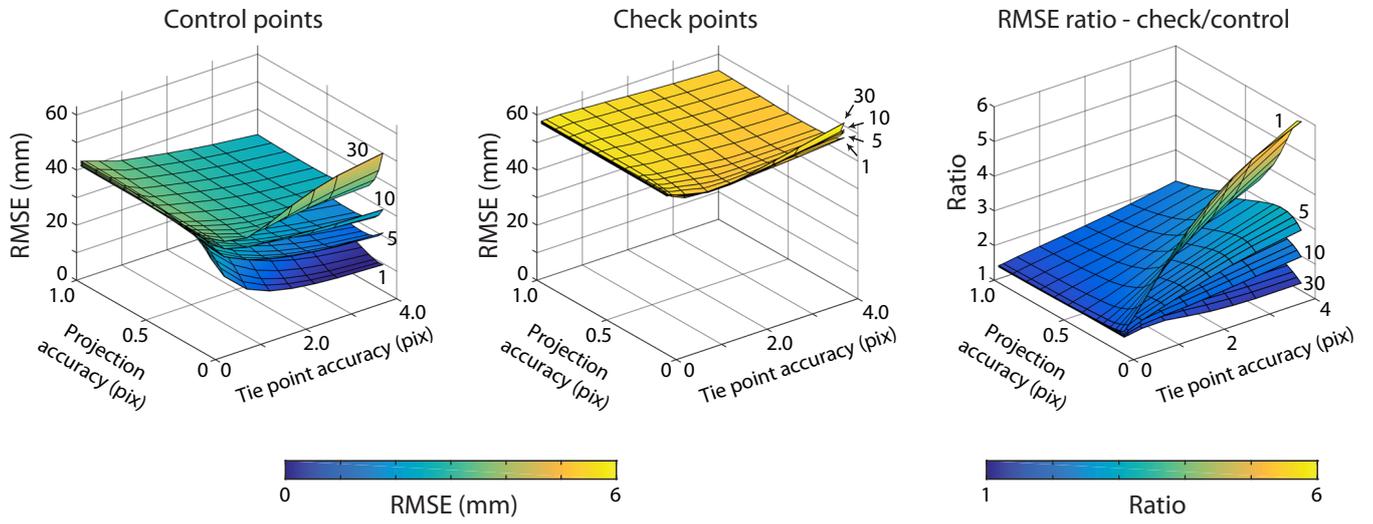


Figure 07

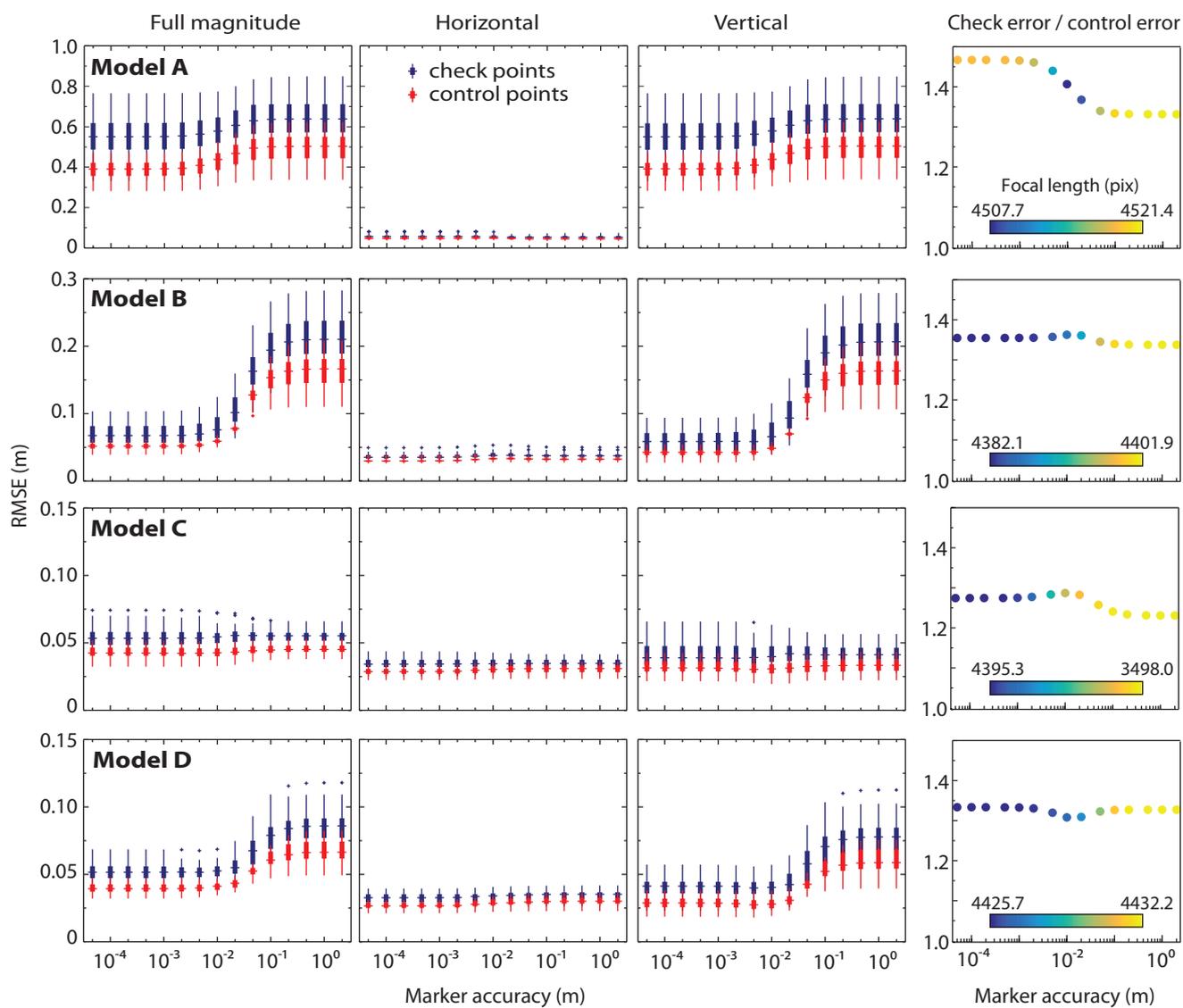


Figure 08

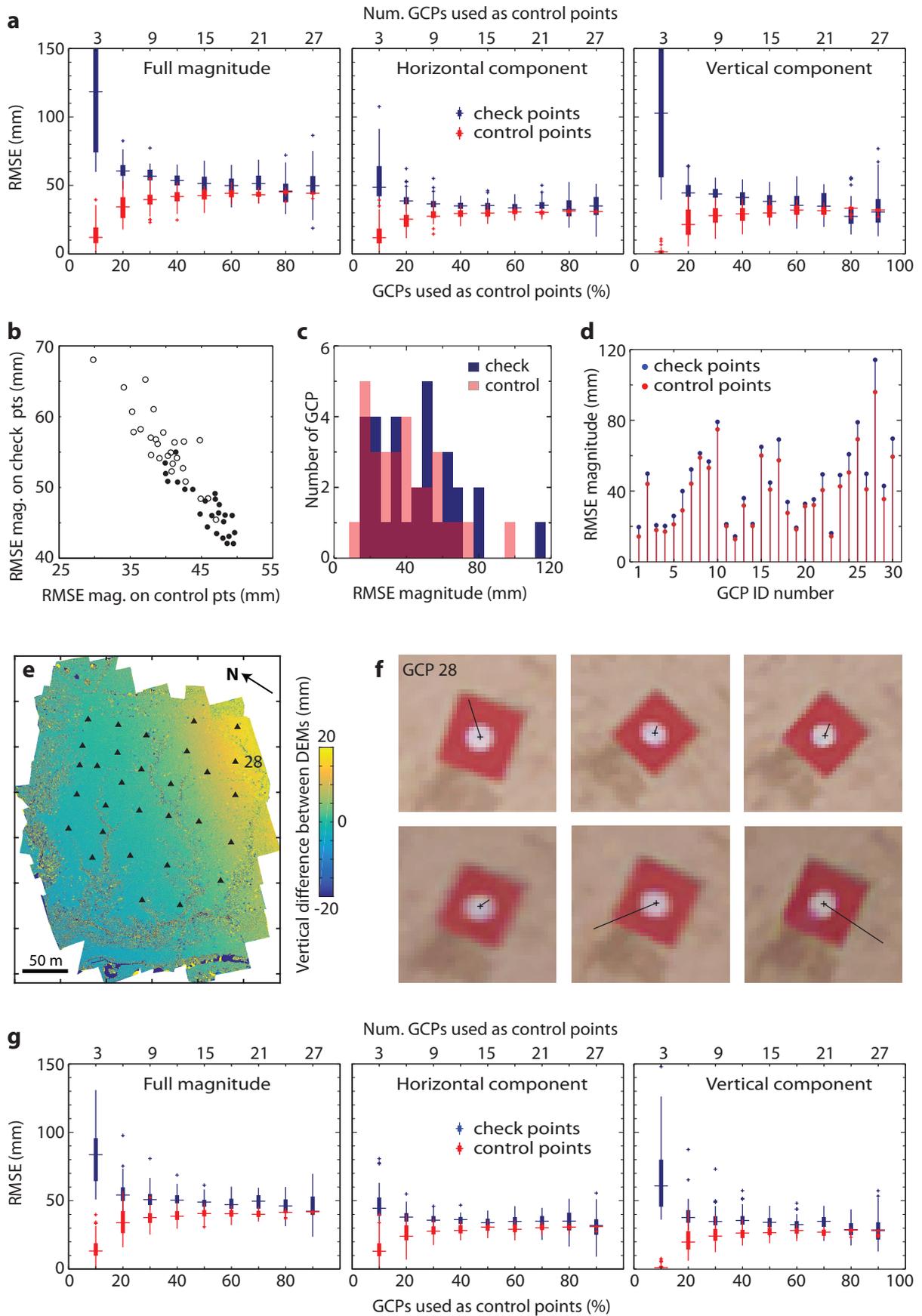


Figure 09

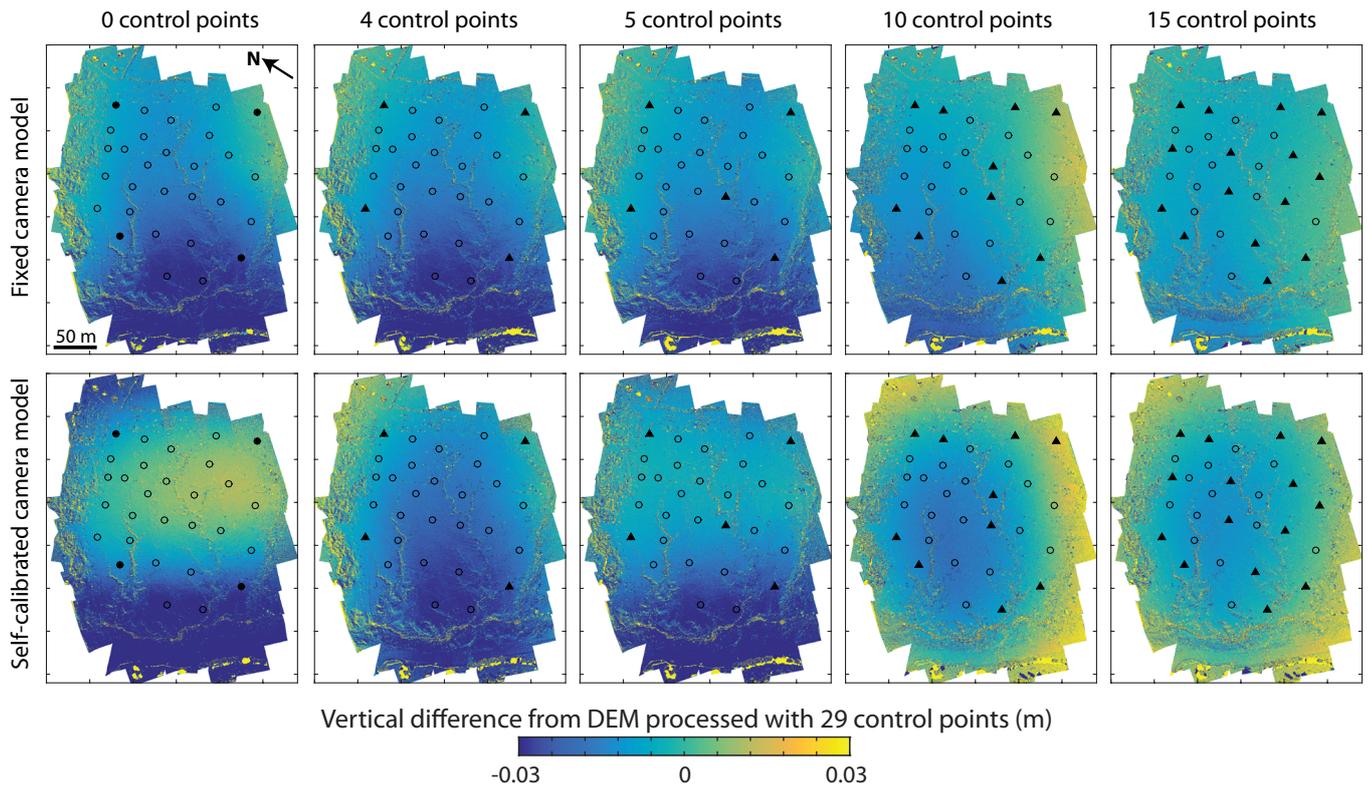


Figure 10

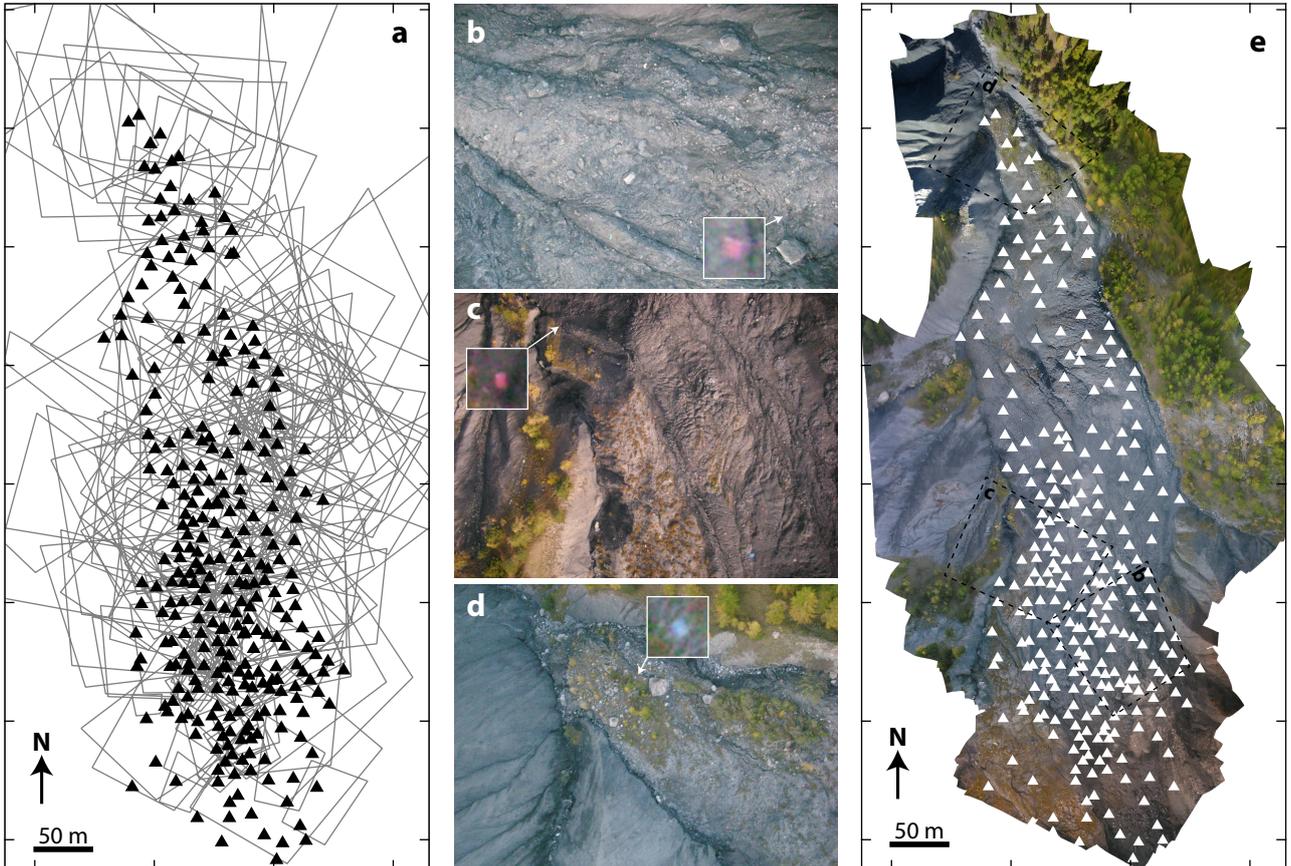


Figure 11

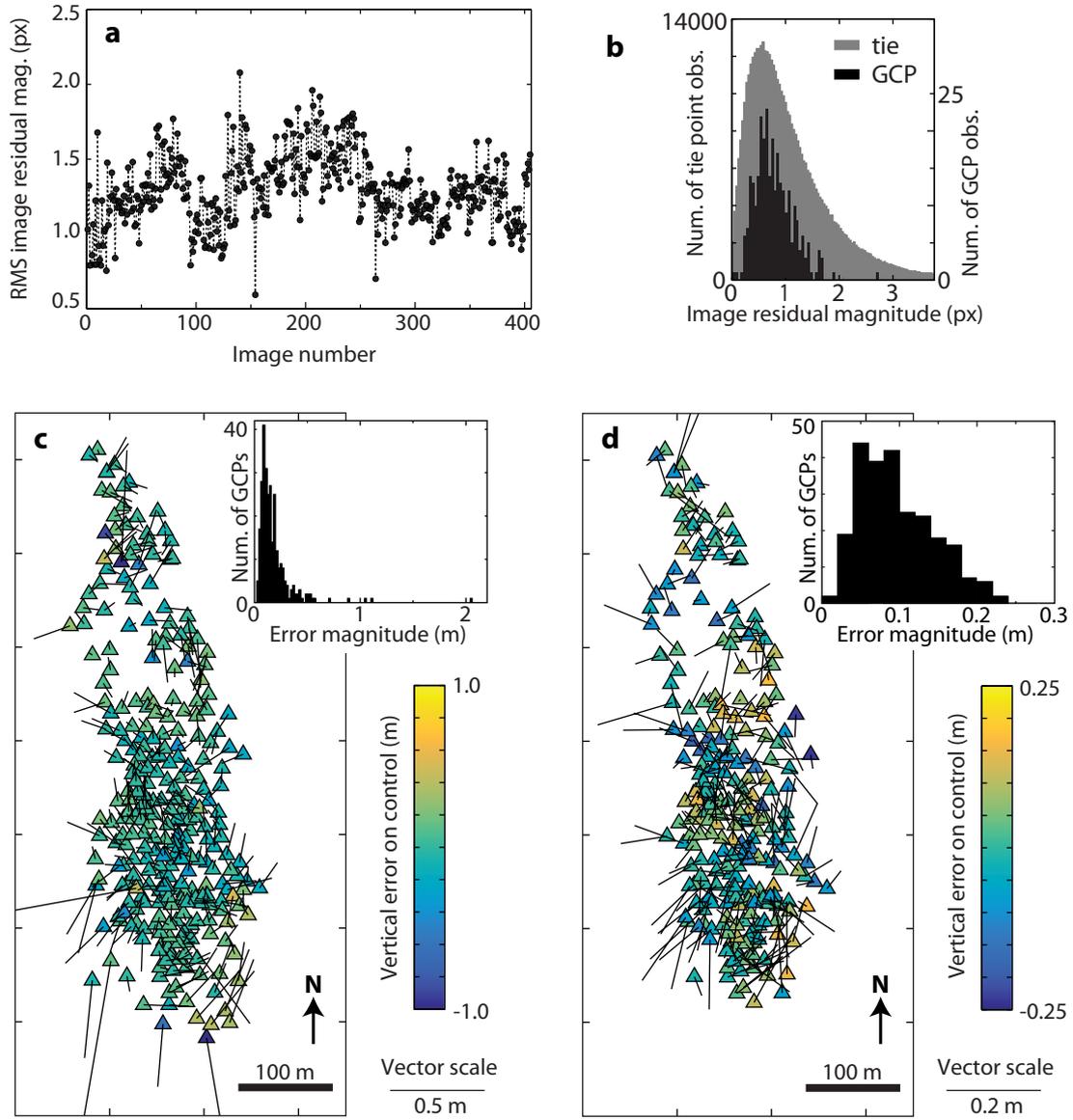


Figure 12

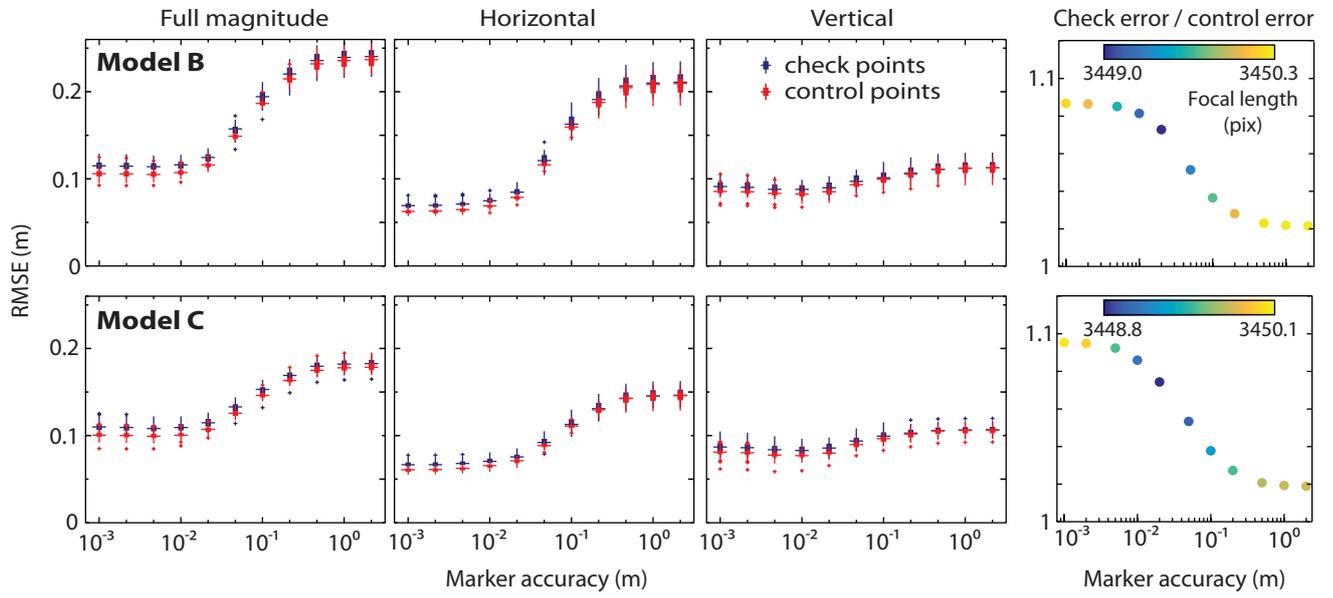


Figure 13

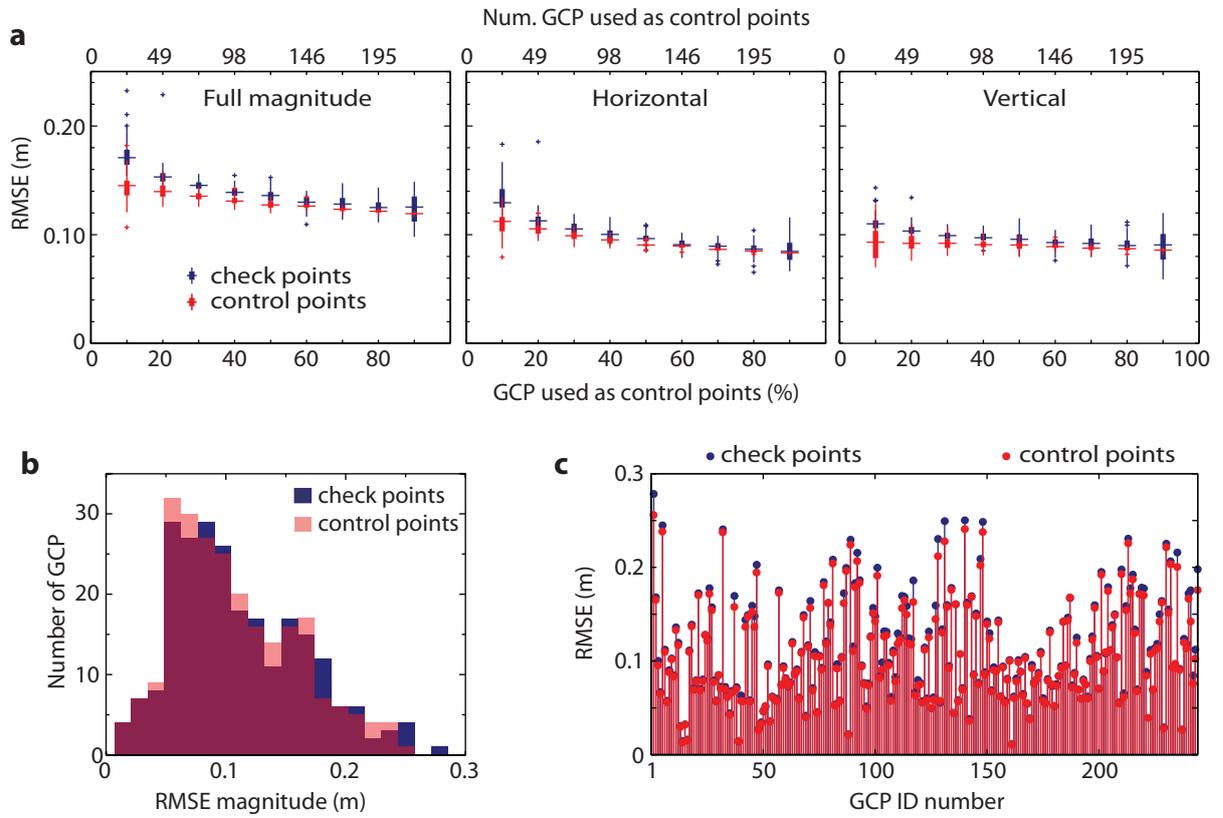


Figure 14

