The Nubian spectrum: 3D geometric morphometric perspectives on Levallois core reduction at Tweefontein, South Africa

SI1. Extended methods

1. Field methods

1.1 Permissions

We verified with the permitting authority (South African Heritage Resources Agency, SAHRA) that no collection permit was required since artefacts temporarily removed for analysis were returned to their original positions. Permission to access the site on the property Tweefontein was arranged through Francois and Nicolette van der Merwe of Brakfontein farm.

1.2 Field sampling

Two sampling methodologies were applied:

Transect sample: The main sample area was formed of 8 1-m-wide parallel transects where all complete cores (and points) were marked, numbered, and temporarily removed for analysis. These were then returned to their marked location. The numbering system is arbitrary and derives from material availability: red and green flags (denoting R or G prefixes), and sheets of numbered stickers.

External sample: In the area surrounding the grid, additional complete preferential cores were marked and collected following the same procedure in order to boost the sample of cores suitable for GM analysis.

1.3 Spatial data collection

Artefact positions were recorded using Emlid dGPS base and rover units; however, issues with satellite geometry and atmospheric conditions rendered GPS data from several transects unusable for spatial analysis. This did not compromise the return of artefacts to their original marked positions. The partial spatial data are not presented here.

1.4 Artefact recording

Artefacts were 3D scanned by EH, MS and CS using an Einscan Pro 2 × Plus structured light scanner. Each specimen was captured in multiple orientations to ensure complete coverage and these scans were automatically aligned in the associated Shining 3D software. Lithic attributes were recorded by EH and MS.

Following fieldwork, in the data processing phase, scan data were not found for artefacts R128, R130, G299 and G471, and metric attribute data were not found for artefacts G244, G261, G267, G268, G275, G288 and G291. (These samples must have been accidentally overlooked and returned to the site prior to data capture). These were excluded from specific GM and metric analyses but included in the overall sample since core type and raw material data were available.

2. Geometric morphometric and other analytical methods

2.1 3D data processing

Mesh models were generated using Shining 3D software at medium resolution, exported in PLY format, and simplified using the quadric edge decimation resampling tool in Meshlab (quality threshold 0.3).

2.2 Landmark placement

This followed the protocol set out in Hallinan & Cascalheira (2025), established specifically in relation to Nubian cores. This is directly transferrable to cores with a morphology structured as two hemispheres separated by a plane of intersection, where the final platform and flake removal surface can be identified. Where cores could not be oriented according to a clear upper or lower surface, or the platform for a final or main removal identified for the fixed landmark points, these were excluded. These models are available in the folder "nonGM_cores" for reference. A full inventory detailing these artefacts and the reason why specimens were unsuitable for GM is provided in the data sheet "Tweefontein_cores".

2.3 Surface area and volume

R script was used to extract surface area and volume data from PLY models. However, not all were successful since the model decimation stage introduced holes in some meshes. Where the automated process failed, model PLY files were opened in Meshlab, repaired and cleaned using "Repair non Manifold Edges" using "Close holes" functions. Volume and surface area could then be obtained using the "Compute Geometric Measures" function. These were then input into the appropriate data spreadsheets manually for both the whole grid sample (including GM and non-GM cores). The surface area and volume data for the GM core sample were also added to the "Core_attributes" data sheet.

Note that surface area and volume data were not obtained for R180 because this artefact comprises two conjoining pieces but part of the core back is missing. No scans were found for G299 and G471.

2.4 Scar density analysis

Scars were counted from 3D models by OS. Models were visualised in Meshlab and removal scars larger than 1 cm included in the scar count, using the z-painting tool to mark each counted removal. Scar Density Index (SDI) for each core was calculated by dividing the total number of scars by surface area.

2.5 Core hemisphere volume estimation

Core models were split into upper and lower hemispheres using the "cutMeshPlane" function from the Morpho package. This divides the vertices of the mesh on either side of a plane which is defined by a normal vector passing through the centroid. The vertices of each mesh half were classified as positive or negative to determine which were above or below the plane – i.e. upper

or lower core surface. Since these split mesh halves are open shapes, convex hulls were generated using the "convhulln" function from the geometry package to close each mesh in order to estimate the volume. This is not a precise measure of each surface volume, tending to over-estimate given that scar surfaces are concave and are filled by the bounding hull. However, the sum of both halves was in most cases within ± 20% of the actual core volume (79% of the sample); specimens with a higher than 20% difference had a prominent concave preferential scar on the upper surface.

3. Accompanying data

The folder ***Tweefontein data***, available from the Open Science Framework repository https://osf.io/JQU2K/, includes the following R script and files required to replicate all analyses in full. Its contents should be downloaded and used as the root directory for analyses with the R script:

3.1 R files (.R, .Rmd and .Rdata)

Semilandmarking_SA: Code for preparing the landmark data for subsequent analyses. For a full explanation of the data preparation protocol, see Hallinan & Cascalheira, 2025 and the accompanying compendium https://osf.io/sj8zv/

Analyses_SA: Code for performing all analyses. This script is explained in full in the associated R Markdown and .HTML file ***The_Nubian_Spectrum_Analyses***.

The .Rdata files ***outlineSA***, ***patchedSA*** and ***prefSA*** are also provided so that analyses can be replicated without running the computationally-heavy semilandmark processing script.

3.2 3D mesh models (.ply) and landmarks (.pts)

These are the raw data, organised into four folders:

GM_cores1 and ***GM_cores2*** contain 3D models required for the semilandmark processing and patch placement steps in the R File ***Semilandmarking_SA***. The corresponding landmark co-ordinates are also in this folder as PTS files. The reason for dividing these specimens between two folders is because 37 cores experienced an inversion of the upper and lower surface patches during the patch placement stage. To rectify this, two additional surface landmarks were added to each face at the centroid of the specimens in ***GM_cores2*** which ensured the patches were oriented correctly. These orientation landmarks were then removed and the samples recombined prior to subsequent analysis.

Pref_scar contains 3D models corresponding to cores with preferential scars suitable for shape analysis, with corresponding .PTS landmark configurations (89 specimens). This duplicates files in the other GM core folders but are grouped separately for ease of access.

nonGM_cores contains all other 3D scanned cores from the transect sample, included in the metric analyses (125 specimens). There are no landmark files.

An additional .PTS landmark co-ordinate file ***R146_core_template*** is required for placing the template patch during semilandmark processing.

3.3 Comma separated value spreadsheets (.csv)

Transect_sample: Full attribute data for all 261 cores in the transect sample. This includes cores used in the GM sample, and in the non-GM sample. These were the basis for the metric analysis of the overall assemblage.

GM_sample: Full attribute data for the 179 cores in the GM sample (including cores both from the transect and external supplementary samples). These were the subject of GM and other analyses using 3D models (SDI, volumes etc.)

Core_links, ***Pref_links***: Links between landmarks for generating the mean shape outline for cores and preferential scars

SA_sites_cores: Contains the summary data for generating histograms comparing core frequencies and terminologies from comparative South African MSA assemblages.

3.4 Excel Worksheet (.xlsx)

Tweefontein_cores: This spreadsheet contains the full data for all analysed cores and is supplementary to the attribute information contained in the other .csv files. Additional information includes reasoning why specimens were not included in the GM sample, and descriptive comments regarding the interpretation of each specimen.