Influence of force and duration on stone tool wear: results from experiments with a force-controlled robot

Authors: Johannes Pfleging¹², Radu Iovita²³, Jonas Buchli¹

Mail: pfleginj@ethz.ch, iovita@nyu.edu, buchlij@ethz.ch

Keywords: controlled experiment; micro-wear; robot; metrology; height parameters; fractal analysis

Abstract

Many studies have shown that micro-wear analysis can identify some parameters such as worked material and motion direction with varying degrees of success. However, because experiments have traditionally been carried out by un-monitored humans, we do not fully understand the role of force in wear formation. Here we compare the amount of wear produced by duration vs. applied force in a controlled experiment and using both the inspection of optical images and quantitative parameters describing surface topography. We used flint flakes attached to a force/ torque controllable robot arm to scrape standardized beech wooden planks under constant force profiles. The force profiles were obtained by previous experiments in scraping described in Pfleging et al. (2015). We varied the force level and use duration among the experiments. Worn pieces were imaged with an Alicona InfiniteFocus G4 microscope and the polished parts of the flakes were analysed using areal field parameters from metrology. The data is publicly available on the internet. Results indicate that use duration contributes more significantly to polish formation than force, confirming assumptions made in human experiments performed in the 1980s. Moreover, simple metrological height parameters appear inadequate for capturing the degree of polish. We conclude that more sophisticated quantitative methods are required to go beyond the subjective human evaluation of optical images to reconstruct past human action.

1 ETH Zürich, Institute of Robotics and Intelligent Systems, Agile & Dexterous Robotics Lab

2 New York University, Department of Anthropology, The Center for the Study of Human Origins

3 MONREPOS Research Centre and Museum, Römisch-Germanisches Zentralmuseum, Neuwied, Germany

Acknowledgements

This research has been supported by the Swiss National Science Foundation through a Professorship Award to Jonas Buchli (www.snf.ch) and by Research Grants of ETH Zürich (www.ethz.ch), ETH-36 14-1 (JB/JP). The authors thank Martin Street (Römisch-Germanisches Zentralmuseum, MONREPOS Archäologisches Forschungszentrum, Neuwied, Germany) for help with knapping the stone tools. The authors further thank Mike Lyrenmann (ITA, ETH Zürich) for supporting the set-up design, ETH Zürich carpentry for providing the workpieces, Karl and Maren of Prof. Uwe Sauers group (IMSB, ETH Zürich) for providing facilities and help with the chemical cleaning, Prof. Robert Flatt and Asel Maria Aguilar Sanchez (IfB, ETH Zürich) for supporting the microscopy, Dr. Robert Voss (IWF, ETH Zürich) and Constantin Herbst for help with the Alicona microscope, Daniel Seidel (Deutsches Institut für Luft und Raumfahrt) for providing example code of the KUKA API, Prof. James Stemp for providing image test data and advice, John D'Errico for his MATLAB interpolation function, and Francesco Bocale (KUKA Roboter Schweiz AG) for supporting the robot arm. Finally, we like to thank our colleagues from the Agile & Dexterous Robotics Lab, ETH Zürich for general suggestions and help during the project.

Introduction

Lithic use-wear experiments as methodology to generate reference data for the study of use-wear on stone tool artefacts has been applied since the first work by Semenov (1964). In this methodology stone tool replicas are used for hypothetical tasks under experimental conditions followed by macro- and microscopic analysis of the replica's surface. The observed wear pattern can contribute to functional interpretations of real artefacts showing the same or similar pattern.

Experimental activities range from single actions like hide-scraping or wood-shaving to complex, multi-step ones such as hunting and butchering of animals. Often, studies focus on single tasks, which constitute a goal-oriented activity like sawing, scraping hides of different state, cutting, reaping, incising, notching, etc.

Micro-wear analysis as special case of use-wear analysis represents a potentially rich source of information about ancient human behaviour because of the density of microscopic structures visible on an individual artefact's surface. Numerous studies have shown its potential to infer parameters of the actions performed with a prehistoric artefact, like use duration, worked material, motion direction, or force during use. Many have recently presented protocols attempting to show quantitatively the effect of use duration (Evans et al., 2014; Lerner et al., 2007), worked material (Evans and Donahue, 2008; Faulks et al., 2011; Stemp et al., 2010, 2009; Stemp and Chung, 2011; Stevens et al., 2010) and also direction of motion (Anderson et al., 2006; Vargiolu et al., 2003) on the micro-wear pattern.

More recent studies focus on the influence of force through quantitative analysis. Key (2013) recorded the force applied during cutting of tendon (125 g – 710 g) with a force sensor and analysed the resulting polish with texture analysis of SEM micrographs. He reported no statistically significant correlations between force and texture features. He speculates that this is caused by the low number of strokes executed (700). Positive results were achieved by Key et al. (2015), and Stemp et al. (2015). Using laser-scanning confocal microscopy and relative area scale algorithms they analysed basalt flakes worn in a cutting experiment. They measured the vertical force with a load cell in the range between 150 g up to 4.5 kg. Key could reliably discriminate force increments of 750 g. For smaller force differences, both authors report less reliable results of the test statistics used.

In this article we present results on the combined influence of force and duration on use-wear accrual in a very controlled approach. We used a robot arm to scrape standardized beech wood planks under varying force levels between 3 kg and 11 kg and durations between 50 to 450 strokes. The advantage of the robot compared to manual scraping is that the number of strokes and force levels can be specified in advance and the machine ensures controlled execution over the duration of the experiment. Therefore, noise due to natural variations in force of the human hand and other parameters is reduced. We quantified the wear pattern with a focus variation microscope and used topographic parameters for the analysis.

Robotic Data Generation

Many researchers have already recognized the power of mechanically controlled use-wear experiments to complement replicative experiments with human subjects (Collins, 2008; Lerner et al., 2007; Tomenchuk, 1985). The controlled task execution of an apparatus allows us to evaluate the influence of motion parameters on use-wear processes much better than tasks performed by hand. Some researchers used basic kinematic structures for controlled guidance of lithic replicas in tasks like scraping (Tomenchuk, 1985), threshing (Anderson et al., 2006; Vargiolu et al., 2003) or milling (Dubreuil et al., 2015). Others used commercial pin on disk tribometers (Lewis et al., 2011), scratch testers (Astruc et al., 2003), or custom designed machines (Collins, 2008; Tomenchuk, 1985).

Tomenchuk (1985) designed a basic electrically powered kinematic ‘wear-tester’ to perform linear scraping strokes with the specimen on exchangeable undergrounds. The machine could apply horizontal force values of 40, 50, 70, or 100 N and vertical force values up to 10 N. A mechanical counter registered each revolution of the drive wheel. However, due to safety limitations the machine could only partly be employed for Tomenchuk’s study. In his very extensive work he used the generated data to validate first principle models of use-wear processes.

Collins (2008) used a modified shaping machine (Butler 26 inch Supercrank), to carry out straight strokes for scraping and sawing with flint flakes on wood. The machine provides control over velocity, working angle, vertical pressure, stroke length, as well as the exact duration of the experiment. In her study, Collins tested the influence of working angle, edge morphology, and wood types on macroscopic wear formation of the tool and the workpiece. She took weight measurements during the experiment and imaged the tool edge using low magnification light microscopy. By comparing data of the tool and the workpiece, she was able to investigate the effectiveness of different edge morphologies.

To date, no machine can replicate the excellent craftsmanship abilities of humans. Complex tasks, like the complete butchering of an animal from skinning over gutting (evisceration) to cutting the meat, are far beyond the abilities of autonomous machines. One reason for this is that machine intelligence is too limited to deal with the complex structure of a natural environment. What remains feasible, however, is to break the task down into subtasks, each of which is carried out by a specialized machine working under well-defined and controlled environmental conditions, like in industrial manufacturing. It is the reduction of complexity which leads to a better understanding of the processes involved in the subtasks and therefore enables their automation and, consequently, an improvement in the efficiency of manufacturing.

Similar to the example of industrial manufacturing, focusing on experiments with controlled subtasks helps us to understand the effect of individual physical parameters on use-wear accrual in a complex and realistic (actualistic) setting of experimental archaeology. Focusing on a subtask in the experiment means that a motion of assessable complexity must be isolated from the actualistic, superior task. During the experiment high resolution control over the motions of the subtask must be guaranteed. Only then it is possible to study the influence of all relevant parameters on the wear process, which are beside the materials and the duration also the parameters of the motion like

force, angle, velocity, impacts, etc.

By their mechanistic nature machines execute motions more precisely and over a longer time period than humans. A test set-up based on a machine is therefore the ideal tool to create different wear patterns from a variation of individual motion parameters as well as materials and combinations thereof. Correlations between experimental parameters and the resulting wear pattern can then lead to a better understanding of the wear process and corresponding models of the subtasks. Finally, the resulting data serve as foundation to derive the broader answers related to the superior task within the archaeological setting.

An example for such a machine based set-up was established for this study. We programmed a force and impedance controlled robot arm to perform parametrized scraping strokes on a flat board made of beech wood. Everything that is required to generate one sample of an experimental series is to mount a new stone sample to the arm, type the trajectory parameters into the computer and start the program. In principle, the robot can run the task for several hours or even days without supervision, depending on the experimental question. Therefore, long-term experiments that would not be feasible with human subjects are made possible by this technology. Once the experiment is finished, the motion trajectory is saved to the computer and the stone sample can be analysed under the microscope. Trace and motion data can thus be processed and correlated later on. This automated approach also allows to repeat the experiments in case data gets lost, for statistical reasons, or if results of other researchers need to be validated.

A particular advantage of the robot arm compared to the machines presented by Tomenchuk and Collins is its flexibility. Both of the previously employed machine designs are limited to straight lines on flat workpieces under otherwise constant motion parameters. With the anthropomorphic design of our KUKA robot arm any motion profile that lies within the workspace and velocity limits of the robot can be realized. On a force controllable robot, position, velocity and forces/ torques can be controlled in the same time at any instant of time. Therefore, we can implement more complex, human-like motions, which would shift the experimental setting from the study of basic parametric influence towards capturing the effects of more goal-oriented tasks on trace development.

Materials and Methods

Experiments

Fig. 1: Experimental set-up consisting of robot arm with stone sample attached at the tip and wooden planks fixed onto the table. The vertical arrow (blue) represents the force in vertical direction, the horizontal arrow (red) shows the direction of motion, and the dashed arc indicates the working angle.

The experiments were performed with a torque controllable industrial robot arm (KUKA LWR iiwa 14 kg, Figure 1). This machine has 7 joints, a range of 0.8 m and a maximum payload of 14 kg. Its force control abilities allow the arm tip to follow specified force profiles along the surface of the workpiece. In practice, this means that we can specify the applied force at each point on the path travelled by the robot arm. The operational mode in which the robot arm was used during the experiments emulates a spring-damper-like behaviour of the robot arm (impedance control mode). This means that a difference in the specified and actual set-point of the arm tip results in a force corresponding to an imaginary, parametrized spring-damper system clamped between these two points. Impedance control is a fundamental requirement for robust interaction with the environment (Hogan, 1984).

Flakes were knapped using Baltic flint. In order to fit into the haft, they were modified with a saw to achieve dimensions of roughly 3 cm x 5 cm x 1.5 cm. Each flake was glued with epoxy into a small plastic cup, which can be attached quickly to and detached from the tip of the robot arm (Figure 2). We chose the orientation of the flakes inside the cup in such a way that only the edges created through knapping were in contact with the workpiece during the scraping. Edges created through

Fig. 2: Mechanical interface to mount the flake sample to the robot arm.

sawing usually exhibit artificial wear pattern caused by the sawing process and therefore change the natural roughness of the stone. The workpieces were mounted next to the robot arm in an aluminium frame. The frame and the robot were mounted onto a table such that the range of the robot arm completely covered the workpieces. We used standardized beech wood boards.

The robot was programmed to execute a scraping task with low working angle on the wooden board. The strokes followed a unidirectional, straight line of approximately 30 cm while keeping a constant force in the vertical direction (compare Figure 1). From the endpoint of the last stroke the robot arm moved the flake sample to the starting point of the next stroke without touching the workpiece. The working angle varied between 25° and 35° during scraping. These variations are caused by the vertical force applied and the natural variations in the friction force due to irregularities of the board surface.

We divided the experiments into ‘trials’, which we define as the execution of several scraping strokes with the robot using the same flake sample and keeping the vertical force level constant. Each trial was followed by a cleaning and an imaging step (see below). Depending on the experimental plan the same sample was used in a subsequent trial to generate use-wear states of multiple use durations. A summary of the motion parameters, which we specified for each trial, can be found in Online Resource 1. We used Stroke numbers between 50 and 450 and force levels⁴ between 30 N and 110 N. These values are comparable with those measured during scraping experiments involving humans (Pfleging et al., 2015). The force profiles of typical trials are illustrated in Figure 3. The wood boards were replaced after every 400 strokes. Before and after each trial, photographs of the tool tip and the boards were taken for documentation. The overall protocol can be summarized as:

4 Often, kilogram is used to express the equivalent mass necessary to generate a certain force under gravity. However we refer directly to force and therefore we use the unit, Newton (N), since it is the SI standard for that dimension. 1 kg corresponds to 9.81 N.

- 1) macroscopic documentation of the flake sample with photo camera;
- 2) specifying the number of strokes and force level in the program and starting of the trial;
- 3) macroscopic documentation of the flake sample with photo camera after the trial has finished;
- 4) sample cleaning;
- 5) analysis of the flake sample under the microscope; and
- 6) repeating the protocol starting with 1) to increase the number of strokes, or storing of the flake sample in the archive.

Fig. 3: Trajectories of the vertical force (blue, full) and desired force levels (black, dashed) of 30, 60, 90, and 110 N. The trajectories show between 29 and 31 scraping cycles. The sampling frequency is 100 Hz.

Cleaning

After each trial the flakes were cleaned following closely the cleaning procedure described in Key (2015). The flakes were washed in a grit-free detergent solution, soaked in 15% HCL solution for 15 min, washed under warm water, soaked in a 15% NaOH solution for 15 min, and finally, rinsed under warm water again.

Imaging

For the analysis of the flake surface we used an Alicona InfiniteFocus G4 microscope (Danzl et al., 2011). This 3D microscope uses focus variation and region of focus identification techniques to create

Fig. 4: Alicona Infinite Focus G4 used for imaging the flake samples. The sample is mounted on an actuated stage with rotatable axis (Real 3D).

a true-color image and micro-topographic map of the specimen surface, represented by an image of height (z) values. Depending on lighting conditions and surface properties, the z resolution of each pixel can vary. Therefore for each topographic measurement the resolution is estimated and returned in form of a third image of same size. The samples can be mounted onto a micro actuated table for remote navigation over the specimens' surface in all three axes, which helps to identify worn areas on the flake. The same microscope has been used previously in studies by Macdonald (2014) on its applicability to use wear analysis. The advantage of the Alicona over other metrological instruments like the confocal microscope are its ability to measure on tilted surfaces like the edges of flakes with ideal values for vertical resolution of up to 10 nm and lateral resolution of up to 0.44 μm according to the datasheet. However, for most of our measurements the z resolution indicated by the Alicona was in the range between 2 to 10 nm,

The samples were mounted on a sample holder with one degree of freedom in the orientation (Alicona Real 3D, compare Figure 4). This sample holder allowed us to configure the orientation of the sample with respect to the objective using the working angle from the experiment. This means we were looking at the edge of the tilted flake from top. From this perspective, we expected to capture the areas on the flake which are subject to the maximum force during the experiment.

In the next step, a low magnification objective (5x to 20x) was focused on the most upper part of the edge. From there on the objective was navigated along the edge to search for traces of wear using a 20x or 50x objective. The visually-identified spot of maximum wear was then selected for topographic measurements. Each topographic measurement was taken with the 50x (Nikon LU Plan 0.55 inf/0 EPI) and the 100x (Nikon LU Plan 0.80 inf/0 WD 3.5) objectives. The z-range was set such

Fig. 5: Image of a flake edge exhibiting well developed polish on a stripe of approximately 50 μm width stretching from the top left to the bottom center of the image. Black pixels represent missing z values of the topographic image, which do not lie in the measurement range or which could not be evaluated by the focus detection algorithm. The very light spots are quartz inclusions. The brown flakes in the lower right corner are residues of wood.

that all areas of interest on the image were covered. Parameters for light and exposure were optimised such that the whole range of the intensity histogram was exploited without exceeding the limits. We only used the coaxial light source and the SmartFlash option, which enables averaging of multiple images taken at different light setting, was turned off. Parameters for vertical and lateral resolution were set to the smallest value possible. In case the measurements had to be discarded due to an increasing number of black pixels on the image (compare Figure 5), the value for the lateral resolution was increased and the measurement was repeated. This practice ensures that the measurements are taken with the best possible trade-off between lateral resolution and vertical measurement quality. After analysis of the edge, we took one or several topographic measurements (50x and 100x) of areas ca. 1 cm away from the edge (raw areas) as reference for the natural roughness of the flake.

The average area covered by an image is 300 μm x 200 μm for the 50x objectives and 140 μm x 100 μm for the 100x objective. For a few samples we made use of the image stitching abilities of the Alicona system. This feature allows to automatically image a specified, larger region of the specimen. Taking measurements of larger parts of the flake edge provides more data for the analysis and gives

a better understanding of the edge morphology and corresponding force distribution during scraping. Depending on the size of the stitched region and the z range covered, producing such a composite scan can easily take up to 4 hours. Therefore most measurements were acquired in standard mode.

Analysis

Data analysis was performed in MATLAB using custom data processing scripts. For that reason the topographic measurements were exported from the Alicona software to text files in a 3 column text format, each row corresponding to one pixel of the topographic image. Having all data accessible by a multi-purpose engineering software like MATLAB results in greater flexibility in processing, analysing and reporting.

We used two different metrological characterization methods for the analysis, areal height parameters and areal fractal parameters. Similar to the analysis protocol proposed by Evans and Donahue (2008) both methods were evaluated only on manually selected regions of interest (ROI) of the topographic image, which show the highest degree of wear. We limited the ROI size to around 1000 μm^2 for 100x images and ca. 2000 μm^2 for 50x images. On each image between 1 and 4 ROIs were selected. Additionally, each ROI was subjectively classified as either of the labels, 'raw', 'low', 'medium', or 'high' depending on the degree of wear visible on the true-color image, which was evaluated by one of the authors. These labels were later used as a measure to evaluate the performance of the metrological parameters. Statistical correlations with the experimental parameters were evaluated with a linear model and a Normal or Poisson distribution at the 5 % significance level.

For the evaluation with height parameters the images were preprocessed in 3 steps, form removal, Gaussian filtering, and outlier filtering following that order. Form removal was done by removing a fitted 2D plane from the original data. The Gaussian filter was applied with a standard deviation of 6 μm for 50x images and 4 μm for 100x images followed by subtraction of the filtered image from the original image. Both steps are essential for an unbiased surface characterization with height parameters because they separate the undesired form information from roughness. In the third step, a median filter was applied using a squared kernel with side length of 20 pixels. All of the resulting pixel values, which exceeded a threshold of 1 μm , were classified as outliers and were discarded. The choice for filter parameter values was the result of several tests on the concordance between roughness visible on the true-color image and corresponding surface roughness parameters.

From the preprocessed ROIs we extracted height parameters, arithmetic mean height (S_a , (1)), root mean square height (S_q , (2)), skewness (S_{sk} , (3)), and kurtosis (S_{ku} , (4)). Those parameters are standardized in ISO 25178 (Geometric Product Specifications (GPS) – Surface texture: areal international standard) for surface roughness analysis (Leach, 2013) and were used in previous work (Evans et al., 2014; Macdonald, 2014). They are defined as:

$$S_a = \frac{1}{A} \int \int_A |z(x,y)| dx dy \quad (1)$$

$$Sq = \frac{1}{A} \sqrt{\int \int_A z^2(x, y) dx dy} \quad (2)$$

$$Ssk = \frac{1}{Sq^3} \frac{1}{A} \int \int_A z^3(x, y) dx dy \quad (3)$$

$$Sku = \frac{1}{Sq^4} \frac{1}{A} \int \int_A z^4(x, y) dx dy \quad (4)$$

where $z(x, y)$ is the height at the x, y coordinate and A is the area covered by the ROI. We used the trapezoidal method for approximation of the integration.

In essence, height parameters are statistical moments, which describe the shape of the distribution of height values of the topographic map (see Figure 6). Sq and Sa describe the width of the distribution. A high value indicates large variations in z , which means a high level of roughness. Ssk indicates a shift of the distribution towards positive or negative direction, which speaking in geographic analogies corresponds to a plateau with few valleys or a basin with few hills respectively. Sku indicates whether values at the extreme outer tips of the distribution are present therefore being a measure for the number and extend of high peaks or steep valleys on the topographic map. Like shown in the histogram of Figure 6, after filtering this distribution of height values is usually Gaussian.

Fig. 6: Histogram of height (z) values of an example ROI after filtering. The bin size is 0.02 μm . The distribution shows a slight shift towards the negative direction. Values below -0.4 μm correspond to outliers, which could not be removed in the filtering step.

For the fractal analysis we computed the relative area-scale relationship, which has been applied for MWA before in studies by Key et. al. (2015), Scott et al. (2005), Stemp and Chung (2011), Stemp et. al. (2015, 2013), and others. From the relative area-scale relationship parameters like the fractal dimension, or smooth rough crossover (SRC) can be estimated. These parameters are useful to characterize non-structured surfaces, which cannot be described by classical geometric objects (Brown et al., 1993). The standard algorithm to determine the relative area-scale relationship uses downsampling of the image at a range of different resolutions and computes the ratio of surface area to projected area for each resampled image (Leach, 2013).

Our implementation of the algorithm resamples the ROI at scales varying between the native image resolution up to around half the smaller value between the polygon extensions in x and y. For each

scale the resampled image is triangulated such that each pixel quad splits into two triangles. Then, from the triangle normals the orientation angles are derived and the relative area is computed according to the following formula:

$$relA(s_i) = \frac{1}{N_i} \sum_{j=1}^{N_i} \frac{1}{\cos(\Theta_j)} \quad (5)$$

where $relA(s_i)$ is the relative area at scale s_i , N_i is the number of triangles at scale s_i , and Θ_j is the orientation angle of triangle j , which is equivalent to the ratio of surface area to projected area. From the relative-area scale data we computed the fractal dimension, D as

$$D = 2 - 2 \cdot m \quad (6)$$

where m is the slope at the point of steepest decent of the relative area scale plot. For decreasing roughness one expects an increase in D since the relative-area scale relationship should flatten out. Additionally, we defined the critical scale, s_c as the scale for which one has

$$relA(s_c) = m \quad (7)$$

The critical scale has a similar meaning as the SRC, since it links fractal dimension, which is intrinsically scale independent, and scale. A similar approach was used by Siegmann and Brown (1997), who defined the SRC at the intersection of the slope with the x axis. For decreasing roughness we expect a decrease of s_c since in the range of scales, that we are evaluating, polish formation should mainly remove peaks of the lower scales and leave the relative-area values of the higher scales unchanged. Figure 7 illustrates the relationship between fractal dimension and critical scale.

In the same plot some characteristics of the downsampling and triangulation operations can be observed. The maximum relative area appears at the smallest scale. For all increasing scales the values are smaller. This is because the original image, corresponding to the smallest scale, stores the highest amount of topographic variation. Every triangulation of images sampled at higher resolution values will be based on height values that are lower than those of the triangles of the native image (due to the interpolation) and will therefore also have lower orientation angles. Moreover, in the ideal case each relative area should also be lower than the values at all preceding scales. I.e. the data points of the relative area should be monotonically decreasing like the fitted function and should not show spikes over the whole range of scales. However, because triangulation at larger scales with equally sized triangles can usually not be done without producing a remainder of smaller sized triangles, the averaging operation in Formula (5) distorts the true value for relative area.

Fig. 7: Typical plot of the relative-area scale relationship. The blue line represents the output of the algorithm. The smooth red line is a fit to the discrete data points. The dashed vertical line marks the location of the critical scale, $s_c \approx 2.13 \mu m^2$. It corresponds to the apex (steepest decline) of the curve, which is expressed in the fractal dimension, $D \approx -2.005 \text{ } 1/\mu m^2$.

Fig. 8: Height parameters evaluated on ROIs of 50x topographic images correlated with the subjective labels. For each parameter and label the values are represented in boxplots. Red crosses mark outliers defined as points above 1.5 times the inter quartile range. Digits below the labels are the number of ROIs evaluated. Note the different scales among the parameters.

Fig. 9: Height parameters evaluated on ROIs of 100x topographic images correlated with the subjective labels. For each parameter and label the values are represented in boxplots. Red crosses mark outliers defined as points above 1.5 times the inter quartile range. Digits below the labels are the number of ROIs evaluated. Note the different scales among the parameters.

Results

Generally, both force and duration influence the degree of wear; however, this relationship is only documented when degree of wear is estimated from the optical images. Overall, we find that the 3D measurement using this instrument has a number of issues that render its use problematic.

Relationship between subjective labels and metrological parameters

Given that most use wear analysts work with optical images of polish, we classified image ROIs into “low”, “medium”, and “high” based on perceived intensity of polish. The relationship between the height parameters from ISO 25178 (Sa, Sq, Ssk, Sku) and the subjective labels for 50x and 100x images are shown in Figure 8 and Figure 9. Comparison of the two clearly shows that the height parameters obtained from the 100x images have a stronger relationship with the degree of wear observed, than those from the 50x images. For Sa and Sq of the 100x images the median is decreasing with increasing degree of wear, whereas for the 50x images the median first decreases and then increases again for “medium” and “high”. A similar comparison can be made for the Ssk values, only

Fig. 10: Fractal parameters evaluated on ROIs of 100x images correlated with the subjective labels. For each parameter and label the values are represented in boxplots. Red crosses mark outliers defined as points above 1.5 times the inter quartile range. Digits below the labels are the number of ROIs evaluated. Note the different scales among the parameters.

that for the low and medium degree of wear of the 100x images the values are slightly increasing but clearly drop for the ROIs of the fourth class. Instead, medians of the 50x images clearly increase for the first 3 classes before dropping at “high”. The Sku values are hard to compare due to the large range in which they are depicted. But upon closer inspection, both image magnifications show an increase of the Sku values with increasing degree of wear. A comparison of the quartile location at Sq and Sa shows that for 100x images the third quartile is always further away from the median than the first quartile, whereas for 50x images the arrangement is fluctuating. Similarly, the number of outliers for 100x images is generally higher for all parameters.

The relationship of the fractal parameters with labels is shown in Figure 10. In this case both parameters show a relationship with the labels, which is increasing for the fractal dimension and decreasing for the critical scale. The fractal dimension shows strong increase of the inter quartile range from 'medium' to 'high'.

Relationship between force, duration and wear

The following figures show the wear evaluated on ROIs related with the parameters of the scraping experiment (force, duration). Figure 13, and Figure Error: Reference source not found evaluate the wear based on the metrological parameters, Sa and fractal dimension, Figure 12 evaluates the wear based on the subjective labels. The correlation with critical scale is not shown here, since it did not show any improvement compared to fractal dimension. Each plot corresponds to one force level, where the x axis represents the number of scraping cycles (duration), and the y axis the degree of wear. The values for the surface roughness of the ROIs corresponding to one force-duration pair are represented by mean and variance.

The Sa values in Figure 13 lie in the range between 0.06 μm and 0.47 μm , with an outlier at 425 cycles. Values at 30, 60, 90 N and 250 cycles are missing. For the force level of 110 N regression with a linear model and a normal distribution gives a slope of -0.06 μm / 500 cycles and a p value of 0.003. For the other force levels no clear relationship could be observed.

The values for the fractal dimension in Figure Error: Reference source not found lie in the range between 2.0001 and 2.0068. For the 90 N force level regression with a linear model and a normal distribution gives a slope of 0.002 μm / 500 cycles and a p value of 0.005. Neither correlations at force levels other than 90 N nor the correlation with force show any significant relationship.

In Figure 12 labels are used for the subjective evaluation of wear. Labels were converted to natural numbers between 1 and 4 with high = 1, and raw = 4. A regression with a linear model and a Poisson distribution gives a slope of -0.85 / 500 cycles and a p value of $1.4 \cdot 10^{-11}$ for the duration and a slope of -0.55 / 500 cycles and a p value of 0.29 for the force.

Fig. 11: S_a , taken from 100x images plotted against experimental task parameters, force and duration. Each plot corresponds to one force value between 30 and 110 N. The x axes show the number of scraping cycles before data collection. For each force-duration pair the S_a values are represented by mean and variance shown as error bars. The digit next to the error bar indicates the number of ROIs evaluated for that pair. Variances are scaled for graphical reasons.

Fig. 12: Fractal dimension of 100x images correlated with experimental parameters, force and duration. Each plot corresponds to one force value between 30 and 110 N. The x axes show the number of scraping cycles before data collection. For each force-duration pair the fractal dimension values are represented by mean and variance shown as error bars. The digit next to the error bar indicates the number of ROIs evaluated for that pair. Variances are scaled for graphical reasons.

Fig. 13: Labels assigned to ROIs from 50x and 100x images correlated with experimental parameters, force and duration. Each plot corresponds to one force value between 30 and 110 N. The x axes show the number of scraping cycles before data collection. Labels were converted to natural numbers between 1 and 4 with high = 1, and raw = 4. For each force-duration pair the label values are represented by mean and variance shown as error bars. The digit next to the error bar indicates the number of ROIs evaluated for that pair. Variances are scaled for graphical reasons.

Measurement properties

The following two figures show properties of the measurement system, which are helpful for the interpretation of the results shown above. Figure 14 shows the resolution of the topographic measurement plotted over the subjective labels. The values are estimates produced by the Alicona instrument. The best resolution is achieved on raw ROIs with ca. 1 nm, the worst resolution lies above 20 nm for regions labeled ‘high’. The boxplots clearly show a decrease in measurement quality as degree of wear increases.

The values of the raw areas are shown in Figure 15. For each image of a raw area two ROIs were selected. Values for 100x images range between 0.09 to 0.35 μm with median at 0.16 μm , values for 50x images from roughly 0.15 to 0.48 μm with median around 0.26 μm .

Fig. 14: Resolution of the topographic z values averaged for each ROI. The average values are represented in boxplots correlated with the subjective labels. Digits below the labels are the number of ROIs evaluated. Red crosses mark outliers defined as points above 1.5 times the inter quartile range.

Fig. 15: S_a values of ROIs on raw areas of the flake samples. The left box shows values for 100x images, the right box values for 50x images. Red crosses mark outliers defined as points above 1.5 times the inter quartile range.

Discussion

The results do not show a clear relationship of the metrological parameters neither with force nor with duration (Figure 13 and Figure Error: Reference source not found). The expected behaviour of decreasing roughness with increasing use duration and increasing force level could not be detected from the metrological parameters. Only the 110 and 90 N level showed a statistically significant correlation with duration for Sa and D respectively. We believe that mainly two sources of noise are responsible for the distortion of the expected relationship with force and duration: the limited area size covered by microscopic imaging compared to the overall flake surface and the limits in measurement power of the Alicona microscope on polished areas.

The first point becomes clear when comparing the average size of our ROIs ($1000\text{ }\mu\text{m}^2$), with the size of the total area in contact with the workpiece during the experiment, which can be conservatively guessed to 1 mm^2 . With a total of 4 ROIs evaluated, it means that only 0.4 % of the overall contact area is taken into account in the analysis. Further, one can assume, that the pressure distribution over the contact area is not constant when the flake edge is pushed onto the workpiece. Peaks and valleys on the edge indent the wood with varying intensity and therefore experience varying degree of mechanical pressure. Moreover, the pressure distribution is never the same among the flake samples, since microscopic and macroscopic morphology as well as overall dimensions vary naturally for each of them. This inhomogeneity together with the limited size of the microscopic images generates randomness in the surface roughness observed between the images from different flake samples as well as between the images of the same flake sample. We tried to reduce this effect as much as possible by choosing flake samples of similar size and shape and limiting the search for polished areas under the microscope to the flake edge. Only occasionally we looked at the side next to the edge in case the edge was too thin for the chosen ROI size. The controlled motions of the robot arm also helped with that respect since the working angle was always lying around 30° , therefore the contact area stayed the same for the whole time.

The second point, limits in the microscopic measurement power, is supported by the observation that the correlation of duration with labels (Figure 12) gives a better result than the correlation with metrological parameters (Figure 13 and Figure Error: Reference source not found). The subjective evaluation of wear by the human eye was based on the microscopic true-color images, whereas the metrological parameters were computed from the ROIs of the topographic image. The measurement quality of the Alicona for topographic images can vary drastically and is even getting worse the more polished the sample is. Polished areas show up as blank spots with low texture on the true-color image (compare Figure 5). Due to the low texture, the focus detection algorithm produces less accurate measurements for those areas or discards the value completely, which is then indicated by black pixels. This behaviour was already reported by Danzl et al. (2011), Evans and Macdonald (2011) and Macdonald (2014) and is reflected in the accuracy plot of Figure 14. We further suspect that the positive correlation of the Sa parameter on 50x images (Figure 8) is directly influenced by the decreasing measurement quality of z values, since the scales of the surface features observed (ca. $0.1\text{ }\mu\text{m}$) are close to the resolution limits of the microscope and might therefore mainly capture the

measurement noise.

In this study we focused on height parameters and fractal analysis for characterization of the roughness. Both methods were equally useful for quantification of wear. S_a , S_q , S_{sk} as well as S_c show a negative relationship with wear, whereas S_{ku} and D show a positive relationship. A disadvantage of the surface height parameters is clearly their sensitivity to noise and outliers in the data (Evans et al., 2014). In our case, the least sensitive results were achieved with the S_a parameter (compare Figure 9). An essential step for the correct use of height parameters is filtering. We used phase constant Gaussian filtering for separation of form from roughness as described in Leach (2013). Filter parameters of 4 μm for 100x images and 6 μm for 50x images were determined to give the best results on a test data measurement set. We also applied a median filter to remove outliers, which are mainly caused by bright spots from quartz inclusions. However, the complete removal of outliers was not possible as the histogram of Figure 6 shows. Bins at the extreme outer parts of the Gaussian curve cause a distortion of the true height parameter value. Key et al. (2015) and Stemp et al. (2015) also applied fractal analysis methods to study the influence of force on use-wear, but with positive results. However, the statistical method they used differed from our approach since they compared the Mean Square Ratio of pairs of flake samples (F-test) worn with different force values. We, in contrast, tried to correlate absolute values of metrological parameters, which is probably less robust with respect to the measurement noise (see second point above) compared to the variance analysis used by Key and Stemp. They also reported that some of their sample pairs could not significantly be discriminated and they suspect reasons for that in the natural variation of the raw surface (see first point above).

Another evidence from our results regards the question for the magnification factor. Figure 8 and Figure 9 show the correlation of the roughness parameters with the degree of use-wear observed. For the 50x images the S_a value is slightly increasing with increasing wear, whereas for the 100x images the relationship clearly shows the expected decrease of S_a . Therefore, we believe that the 100x objective is better suited to quantify polish on flint. However, the objectives also had a different numerical aperture (0.55 for 50x and 0.80 for 100x) and this may equally be responsible for the differences observed.

Due to the high degree of control during the robotic experiment we were able to establish a very well documented reference collection of experimental stone tool replicas. This documentation comprises the stone sample and its macroscopic and microscopic images, the workpiece, duration of use, and the motion trajectory of the experiment, which contains all kinematic and dynamic parameters like working angle, length of the stroke, velocity, and force profile in vertical and lateral direction. We published all data at an online repository (Pfleging et al., 2017), including the analysis scripts and results in MATLAB as well as a visualization of the ROIs with labels. This way, it is possible to reproduce the results shown in this article, to reconstruct working principles of the algorithms, and to test and compare new analysis methods on the data.

In the future, we plan to compare the results from the focus variation to those obtained from confocal microscopy and interferometry, both of which deal better with smooth surfaces.

Additionally, we want to advance wear quantification methods with algorithms, which combine the information of several measurements (e.g. true-color image and topographic image) to yield predictions, that are comparable with classifications by human. Finally, we plan to extend the collection of robotic tasks with more elaborate and more human-like motion pattern.

Conclusions

In this article we used a force and impedance-controlled robot to produce a large reference collection of wood-scraping using flint flakes. We studied the contribution of force and duration to increasing degree of wear on flint stone tools using both optical images and 3D topography for describing this variable. We could show a relationship with duration based on the visual classification of wear into four classes (“raw”, “low”, “medium”, and “high”) but not with force. On the other hand, wear quantification on the topographic images using the metrological features, height and fractal parameters, showed partially a relationship with duration. Our study revealed some difficulties facing advances in use-wear methodology (e.g. where to look on the edge, which instrument, which settings, limitations of the focus variation technology). In particular, we are not convinced that focus variation technology can be used for quantifying well-developed polishes due to insufficient nano texture on those areas.

The results of this study are reproducible since all images, the experimental documentation, and the analysis scripts used are accessible online (Pfleging et al., 2017). Parametrized use-wear experiments and quantitative analysis requires many datasets covering a large range of experimental variables. We believe, that machine-driven experiments and open access to use-wear datasets are a well suited approach to quickly improve inter-observer accuracy of interpretation and increase trust in the use-wear methodology among archaeologists at large.

Supplementary Material

Online Resource 1: Table of experimental parameters of each trial sorted by flake-sample ID. The columns are FlakeID: unique identifier of the flake sample; TrialID: unique identifier of the trial; ForceLevel: value of the vertical force in Newton applied during the trial; NumStrokes: number of scraping strokes executed during the trial; and NumMeasurements: number of topographic images of the flake sample taken under the microscope after the trial was finished.

Bibliography

- Anderson, P.C., Georges, J.-M., Vargiolu, R., Zahouani, H., 2006. Insights from a tribological analysis of the tribulum. *Journal of Archaeological Science* 33, 1559–1568. <https://doi.org/10.1016/j.jas.2006.02.011>
- Astruc, L., Vargiolu, R., Zahouani, H., 2003. Wear assessments of prehistoric instruments. *Wear*, 14th International Conference on Wear of Materials 255, 341–347. [https://doi.org/10.1016/S0043-1648\(03\)00173-X](https://doi.org/10.1016/S0043-1648(03)00173-X)
- Brown, C.A., Charles, P.D., Johnsen, W.A., Chesters, S., 1993. Fractal analysis of topographic data by the patchwork method. *Wear* 161, 61–67. [https://doi.org/10.1016/0043-1648\(93\)90453-S](https://doi.org/10.1016/0043-1648(93)90453-S)
- Collins, S., 2008. Experimental investigations into edge performance and its implications for stone artefact reduction modelling. *Journal of Archaeological Science* 35, 2164–2170. <https://doi.org/10.1016/j.jas.2008.01.017>
- Danzl, R., Helml, F., Scherer, S., 2011. Focus Variation – a Robust Technology for High Resolution Optical 3D Surface Metrology. *Strojniški vestnik – Journal of Mechanical Engineering* 2011, 245–256. <https://doi.org/10.5545/sv-jme.2010.175>
- Dubreuil, L., Savage, D., Delgado-Raack, S., Plisson, H., Stephenson, B., Torre, I. de la, 2015. Current Analytical Frameworks for Studies of Use–Wear on Ground Stone Tools, in: Marreiros, J.M., Bao, J.F.G., Bicho, N.F. (Eds.), *Use-Wear and Residue Analysis in Archaeology, Manuals in Archaeological Method, Theory and Technique*. Springer International Publishing, pp. 105–158. https://doi.org/10.1007/978-3-319-08257-8_7
- Evans, A.A., Donahue, R.E., 2008. Laser scanning confocal microscopy: a potential technique for the study of lithic microwear. *Journal of Archaeological Science* 35, 2223–2230. <https://doi.org/10.1016/j.jas.2008.02.006>
- Evans, A.A., Macdonald, D., 2011. Using metrology in early prehistoric stone tool research: further work and a brief instrument comparison. *Scanning* 33, 294–303. <https://doi.org/10.1002/sca.20272>
- Evans, A.A., Macdonald, D.A., Giusca, C.L., Leach, R.K., 2014. New method development in prehistoric stone tool research: Evaluating use duration and data analysis protocols. *Micron* 65, 69–75. <https://doi.org/10.1016/j.micron.2014.04.006>
- Faulks, N.R., Kimball, L.R., Hidjati, N., Coffey, T.S., 2011. Atomic force microscopy of microwear traces on Mousterian tools from Myshtylagty Lagat (Weasel Cave), Russia. *Scanning* 33, 304–315. <https://doi.org/10.1002/sca.20273>
- Hogan, N., 1984. Impedance Control: An Approach to Manipulation, in: 1984 American Control Conference. Presented at the 1984 American Control Conference, pp. 304–313.
- Key, A.J.M., 2013. Applied Force as a Determining Factor in Lithic Use-Wear Accrual: An Experimental Investigation of its Validity as a Method with which to Infer Hominin Upper Limb Biomechanics. *Lithic Technology* 38, 32–45. <https://doi.org/10.1179/0197726113Z.0000000001>
- Key, A.J.M., Stemp, W.J., Morozov, M., Proffitt, T., de la Torre, I., 2015. Is Loading a Significantly Influential Factor in the Development of Lithic Microwear? An Experimental Test Using LSCM on Basalt from Olduvai Gorge. *Journal of Archaeological Method and Theory* 22, 1193–1214. <https://doi.org/10.1007/s10816-014-9224-9>
- Leach, R. (Ed.), 2013. *Characterisation of Areal Surface Texture*. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Lerner, H., Du, X., Costopoulos, A., Ostoj-Starzewski, M., 2007. Lithic raw material physical properties and use-wear accrual. *Journal of Archaeological Science* 34, 711–722. <https://doi.org/10.1016/j.jas.2006.07.009>
- Lewis, R., Tsoraki, C., Broughton, J., Cripps, J.C., Afodun, S.A., Slatter, T., Roubos, V., 2011. Abrasive and

- impact wear of stone used to manufacture axes in Neolithic Greece. *Wear*, 18th International Conference on Wear of Materials 271, 2549–2560. <https://doi.org/10.1016/j.wear.2010.12.074>
- Macdonald, D.A., 2014. The application of focus variation microscopy for lithic use-wear quantification. *Journal of Archaeological Science* 48, 26–33. <https://doi.org/10.1016/j.jas.2013.10.003>
- Pfleging, J., Iovita, R., Buchli, J., 2017. Micro-wear data from robotic use-wear experiments on force. Zenodo. <https://doi.org/10.5281/zenodo.1117873>
- Pfleging, J., Stücheli, M., Iovita, R., Buchli, J., 2015. Dynamic monitoring reveals motor task characteristics in prehistoric technical gestures. *PLoS ONE* 10, e0134570. <https://doi.org/10.1371/journal.pone.0134570>
- Scott, R.S., Ungar, P.S., Bergstrom, T.S., Brown, C.A., Grine, F.E., Teaford, M.F., Walker, A., 2005. Dental microwear texture analysis shows within-species diet variability in fossil hominins. *Nature* 436, 693–695. <https://doi.org/10.1038/nature03822>
- Semenov, S.A., 1964. Prehistoric technology: an experimental study of the oldest tools and artefacts form traces of manufacture and wear, Russian ed. 1957 transl. by M. W. Thompson. ed. Cory, Adams & Mackay, London.
- Siegmann, S., Brown, C.A., 1997. Scale-Sensitive Fractal Analysis for Understanding the Influence of Substrate Toughness in Therman Spraying, in: *Thermal Spray: A United Forum for Scientific and Technological Advances*. Presented at the 1st United Therman Spray Conference, pp. 665–570.
- Stemp, W.J., Childs, B.E., Vionnet, S., 2010. Laser profilometry and length-scale analysis of stone tools: second series experiment results. *Scanning* 32, 233–243. <https://doi.org/10.1002/sca.20200>
- Stemp, W.J., Childs, B.E., Vionnet, S., Brown, C.A., 2009. Quantification and Discrimination of Lithic Use-Wear: Surface Profile Measurements and Length-Scale Fractal Analysis. *Archaeometry* 51, 366–382. <https://doi.org/10.1111/j.1475-4754.2008.00404.x>
- Stemp, W.J., Chung, S., 2011. Discrimination of surface wear on obsidian tools using LSCM and RelA: pilot study results (area-scale analysis of obsidian tool surfaces). *Scanning* 33, 279–293. <https://doi.org/10.1002/sca.20250>
- Stemp, W.J., Lerner, H.J., Kristant, E.H., 2013. Quantifying Microwear on Experimental Mistassini Quartzite Scrapers: Preliminary Results of Exploratory Research Using LSCM and Scale-Sensitive Fractal Analysis: Quantifying microwear on experimental Mistassini quartzite scrapers. *Scanning* 35, 28–39. <https://doi.org/10.1002/sca.21032>
- Stemp, W.J., Morozov, M., Key, A.J.M., 2015. Quantifying lithic microwear with load variation on experimental basalt flakes using LSCM and area-scale fractal complexity (Asfc). *Surface Topography: Metrology and Properties* 3, 034006. <https://doi.org/10.1088/2051-672X/3/3/034006>
- Stevens, N.E., Harro, D.R., Hicklin, A., 2010. Practical quantitative lithic use-wear analysis using multiple classifiers. *Journal of Archaeological Science* 37, 2671–2678. <https://doi.org/10.1016/j.jas.2010.06.004>
- Tomenchuk, J., 1985. The development of a wholly parametric use-wear methodology and its application. University of Toronto.
- Vargiolu, R., Zahouani, H., Anderson, P.C., 2003. Étude tribologique du processus d'usure des lames de silex et fonctionnement du tribulum, in: *Le Traitement Des Récoltes: Un Regard Sur La Diversité, Du Néolithique Au Présent*. Presented at the XXIIIe rencontres internationales d'archéologie et d'histoire d'Antibes, pp. 439–54.