Scraper original mass: a different approach and its wide range implementation

Guillermo Bustos-Pérez

# **Scraper original mass: a different approach and its wide range implementation**

Guillermo Bustos-Pérez

Department of Human Origins, Max Planck Institute for Evolutionary Anthropology, Leipzig.  
 Institut Català de Paleoecologia Humana i Evolució Social (IPHES-CERCA), Tarragona, Spain.  
 Departament d’Història i Història de l’Art, Universitat Rovira i Virgili, Tarragona, Spain.

## **Abstract**

Predicting the original mass of a retouched scraper has long been a major goal in lithic analysis. It is commonly linked to lithic technological organization of past societies along with notions of stone tool general morphology, standardization through the reduction process, use life, and site occupation patterns. In order to obtain a prediction of original stone tool mass, previous studies have focused on attributes that would remain constant or unaltered through retouch episodes. However, these approaches have provided limited success for predictions and have also remained untested in the framework of successive resharpening episodes. In the research presented here, a set of experimentally knapped flint flakes were successively resharpened as scraper types. After each resharpening episode, four attributes were recorded (scraper mass, height of retouch, maximum thickness and the GIUR index). Four machine learning models were trained using these variables in order to estimate the mass of the flake prior to any retouch. A Random Forest model provided the best results with an r2 value of 0.97 when predicting original flake mass, and a r2 value of 0.84 when predicting percentage of mass lost by retouch. The Random Forest model has been integrated into an open source and free to use Shiny app. This allows for the wide spread implementation of a highly precise machine learning model for predicting initial mass of flake blanks successively retouched into scrapers.

**Key words:** experimental archaeology; lithic reduction; flake mass; machine learning.

## **Introduction**

Scrapers are some of the most common lithic implements among archaeological lithic assemblages. They are present from the first Oldowan stone tools [[1](#ref-barsky_early_2011)–[3](#ref-semaw_26-million-year-old_2003)] through to modern ethnographic studies of hunter gatherers [[4](#ref-casamiquela_temas_1978)–[7](#ref-sillitoe_living_2003)]. The “reduction model” [[8](#ref-dibble_middle_1995),[9](#ref-dibble_interpretation_1987)] suggests that some stone tools (including scrapers) can represent different stages of reuse and modification through retouch. When considering scrapers within the reduction model, an integral concept is that of curation. The initial definition of curation included a series of behavioral patterns related to provisioning strategies [[10](#ref-binford_organization_1979),[11](#ref-renfrew_interassemblage_1973)]. However, further authors included into curation behavioral strategies such as tool transport, utilization in a wide range of tasks, anticipated production, hafting, and recycling (after the original tool had been discarded). An alternative definition of curation was proposed by Shott [[12](#ref-shott_exegesis_1996),[13](#ref-shott_tool-class_1989)], who defined curation as “the ratio of realized to potential utility”. This new approach to the definition of curation has important implications for lithic analysis and the study of lithic technological organization since it transforms curation into a continuous variable [[12](#ref-shott_exegesis_1996)]. Given this new definition, and in the framework of the reduction model, the amount of mass lost by a lithic artifact by reuse/resharpening will be equivalent to its ratio of curation. Both variables (amount in grams and percentage of mass lost by a stone tool) can be empirically calculated in experimental contexts and estimated in archaeological contexts. The presence of scrapers and their curation relates to several aspects of the organization of lithic technology of past societies [[14](#ref-clarkson_holocene_2002)–[19](#ref-shott_use_2017)]. Amount of curation affects the shape of stone tools at the moment of their discard (thus affecting the morphological variability of stone tool assemblages observed through time). Amount of curation can also relate to raw material sources, with more curated artifacts coming from longer distances. Curation also relates to the selection of technological products for more intensive retouch, or shifts in technological strategies of transport, thus informing about the cultural choices and patterns of past human groups. Finally, curation also relates to tool use and use-wear analysis.

Because of these reasons, predicting original scraper mass is a major goal in lithic analysis. Thus far, two approaches are employed to estimate the reduction and curation undergone by a retouched artifact. The first approach uses estimations derived from direct measurements on retouch. This has led to the proposal of several indexes using different measurements, such as height of retouch, length of retouched edge or projection of the original angle [[20](#ref-bustos-perez_exploring_2019)–[23](#ref-morales_measuring_2015)]. These indices usually provide good correlation values with mass lost, but they are usually conditioned by flake morphology, direction of retouch, or tool type (laterally retouched scrapers, endscrapers, bifacial products, etc.) and each one uses different scale ranges.

The second approach aims at estimating original flake mass based on remaining flake features. This approach has the advantage of not being limited by tool type, direction of retouch or flake morphology. Initial work focused on controlled experiments of flake formation using different measures of flake platform (width, depth) and exterior platform angle (EPA) to estimate flake mass [[24](#ref-dibble_platform_1997)–[28](#ref-pelcin_controlled_1996)]. Some of the reasons to select these features were that they usually remain unaltered in most retouched artefacts. These controlled experiments provided strong explanatory power for the formation of flakes, with flake mass being predicted with an r2 value above 0.8 [[25](#ref-dibble_effect_1995),[29](#ref-li_synthesis_2023)]. However, when the same variables are used to predict mass of experimentally knapped flakes, the predictive power of the model diminishes significantly, with r2 values dropping to 0.403 [[30](#ref-davis_quantifying_1998)] (0.224 for the same retouched flakes), and 0.384 [[31](#ref-shott_flake_2000)]. These results meant an important drawback since, as Dibble [[32](#ref-dibble_comment_1998)] states: “controlled experiments, in spite of their elegance, objectivity and replicability, are only useful if the results obtained from them are directly applicable to archaeological materials”.

To overcome the limitations from these results, three approaches are commonly undertaken:

1. Adding additional features as predictive variables. Commonly flake thickness is added [[19](#ref-shott_use_2017),[33](#ref-dogandzic_edge_2015)], since it is widely considered to remain unaltered through the reduction process. Other variables, such as scar count or remaining amount of cortex, seem to improve the predictive power of models [[34](#ref-bustos-perez_predicting_2021),[35](#ref-bustos-perez_multiple_2022)].
2. Applying new methods for measuring more accurately existing variables. Examples are the refinement on traditional manual measurements of platform [[36](#ref-muller_new_2016)], the use of digital photographs [[37](#ref-braun_landscape-scale_2008)], or 3D scans for measuring platform [[38](#ref-clarkson_estimating_2011),[39](#ref-maloney_experimental_2020)].
3. Applying different families of transformations in order to favor the Gaussian distribution of values of predictors and flake mass, thus increasing the predictive power of most models. These transformations usually use the cubic root [[29](#ref-li_synthesis_2023),[33](#ref-dogandzic_edge_2015),[40](#ref-dogandzic_results_2020)] or different logarithmic transformations [[30](#ref-davis_quantifying_1998),[31](#ref-shott_flake_2000),[34](#ref-bustos-perez_predicting_2021),[35](#ref-bustos-perez_multiple_2022),[38](#ref-clarkson_estimating_2011),[39](#ref-maloney_experimental_2020)].

It can be considered that these additions and improvements have provided correlation values of original flake mass on scrapers which allow for comparisons at the assemblage level. However, estimations at the individual stone artifact level remain unsatisfactory with a limited application to archaeological cases. This is due to three main reasons. First, while most research explores extensively the prediction of mass through different variables (and their interactions), it is usually not considered in the frame of continuous resharpening process, and when it is tested in this framework (continuous or single episodes of retouch), results provide lower correlation values [[32](#ref-dibble_comment_1998),[39](#ref-maloney_experimental_2020)]. Second, while most research aims at estimating original flake mass, less research provides results of estimations of percentage of mass lost against actual percentage of mass lost during retouch, which is the key component of the curated concept [[6](#ref-shott_measuring_2007),[12](#ref-shott_exegesis_1996)]. Third, most archaeological research addressing the prediction of original flake/scraper mass result in equations which might be difficult to extrapolate and practically apply other archaeological assemblages [[19](#ref-shott_use_2017),[30](#ref-davis_quantifying_1998),[41](#ref-morales_distribution_2016)]. Recently [[34](#ref-bustos-perez_predicting_2021),[35](#ref-bustos-perez_multiple_2022)], the use of machine learning has allowed the implementation of feature selection (identification of how many and which variables are better for prediction) and new algorithms. However, it has also resulted in limited improvements of the correlation coefficient [[35](#ref-bustos-perez_multiple_2022)], indicating that a possible threshold limit for this approach is being reached.

A new framework is needed to overcome the limitations of previous models (absence of being tested in sequential experimentations, higher accuracy, and easy implementation for all lithic analysts) aimed at predicting original flake mass from scraper attributes and being able to reach the individual scraper level. In the present study 134 flakes were successively retouched, providing a dataset of 694 episodes of resharpening. After each retouch episode, a series of attributes were measured and used to train four machine learning models. A Random Forest model provided the highest r2 value (0.974) when estimating scraper original mass, and the highest r2 (0.839) when estimating percentage of mass lost by retouch. The Random Forest model and all training data are implemented through a Shiny app “Original Scraper Mass Calculator v.1.0”, which allows the user to manually introduce the data from a scraper to estimate its original mass or to upload all data at once and download the results.

## **Materials and methods.**

### **Experimental sample.**

The analyzed sample consisted of 134 experimentally knapped flakes using hard hammer. The raw material of hammerstones varied widely (quartz, quartzite, sandstone, and limestone), which allowed for a diverse range of morphologies and potential active percussion areas. The experimental sample is dominated by flakes with feather terminations (n = 121; 90.3%), followed by flakes with hinge terminations (n = 10; 7.46%).

Data1 <- read.csv("Data/Data1.csv", sep = " ")  
load("Data/Data2.RData")  
library(tidyverse); library(caret)

Initial flake mass was recorded using a Sytech SY-BS502 with a precision of 0.01 grams. Average weight of the samples was 47.38 g, with 50% of flakes weighing between 18.07 and 63.16 g, and a standard deviation of 36.48. Figure 1 presents the flake mass distribution for the experimental sample, indicating a long tail of 14 flakes weighing more than 100 g.

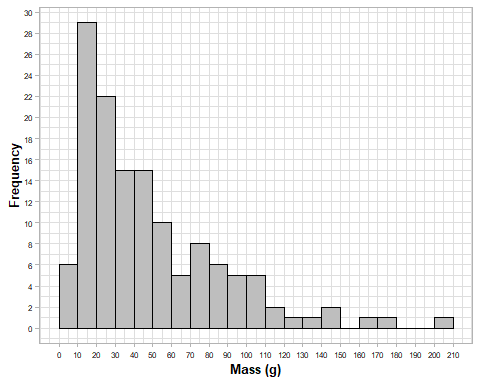
The transversal section of flakes is considered to have an important effect on estimations derived from the geometric index of unifacial reduction (GIUR) [[22](#ref-kuhn_geometric_1990)] and height of retouch. In particular, when a flake’s dorsal surface is parallel to the ventral surface, the GIUR and height of retouch will only marginally or will not at all increase after each resharpening episode, resulting in underestimations of flake mass removal [[8](#ref-dibble_middle_1995),[42](#ref-eren_kuhns_2009)]. However, the actual effect of the “flat flake problem” on the estimation of flake mass might be marginal [[43](#ref-hiscock_reality_2009),[44](#ref-hiscock_experimental_2005)]. The present study recorded flake schematic transversal section prior to retouch of each flake, with possible categories being: circular (n = 20), triangular (n = 63), triangular asymmetric (n = 29), trapezoidal (n = 13) and trapezoidal asymmetric (n = 9). The first three categories are considered to represent cases where the “flat flake problem” is not present, while the latter two are consider to represent cases were this problem is present.

Data0 <- read.csv("Data/Data0.csv", sep = ",")  
  
Summary\_Assem <- data.frame(  
 rbind(data.frame(data.matrix(summary(Data0$L.mm))) %>% t(),  
 data.frame(data.matrix(summary(Data0$W.mm))) %>% t(),  
 data.frame(data.matrix(summary(Data0$T.mm))) %>% t(),  
 data.frame(data.matrix(summary(Data0$Max.Thick.mm))) %>% t(),  
 data.frame(data.matrix(summary(Data0$Mean.Edge.Angle.Selected.ret))) %>% t()))  
Measure <- c("Length", "Width", "Middle Thickness", "Maximum thick",  
 "Mean angle edge prior to retouch")  
Summary\_Assem <- cbind(Measure, Summary\_Assem)  
rownames(Summary\_Assem) <- 1:nrow(Summary\_Assem)  
print(Summary\_Assem)

## Measure Min. X1st.Qu. Median Mean X3rd.Qu.  
## 1 Length 24.5 50.9500 62.600 63.23060 75.800  
## 2 Width 17.4 40.7000 47.250 49.94552 58.250  
## 3 Middle Thickness 4.2 9.8250 13.400 14.46194 18.625  
## 4 Maximum thick 6.0 12.0250 15.800 16.57313 20.000  
## 5 Mean angle edge prior to retouch 13.5 33.0825 39.835 41.27119 48.670  
## Max.  
## 1 106.70  
## 2 90.40  
## 3 34.70  
## 4 36.00  
## 5 72.33

rm(Data0)

Data1 %>%   
 filter(Episode == 0) %>%   
 ggplot(aes(Or.Weight.g)) +  
 geom\_histogram(binwidth = 10,  
 color = "black", fill = "gray",   
 boundary = 0) +  
 theme\_light() +  
 ylab("Frequency") +  
 xlab("Mass (g)") +  
 scale\_x\_continuous(breaks = seq(0, 210, 10)) +  
 scale\_y\_continuous(breaks = seq(0, 30, 2)) +  
 theme(  
 axis.text = element\_text(color = "black", size = 6),  
 axis.title = element\_text(color = "black", size = 10, face = "bold"))



All flakes were retouched until they were too small to hold while retouching (n = 4), they broke during retouch (n = 74), or the angle of retouch was too abrupt to detach additional resharpening flakes (n = 59). Retouch was done through freehand direct hard hammer on the dorsal face of flakes (direct retouch). In order to reshape the flakes one continuous retouched edge was established. After the first episode of resharpening, the retouched edge was expanded through the flakes edge in a continuous manner. This limits the potential application of this index to simple scrapers with direct retouch (which excludes double scrapers or scrapers with inverse or bifacial retouch).

Most flakes underwent between four and seven episodes of retouch (67.2%), while only seven flakes provided nine or more episodes of retouch. The experimental assemblage provided a total of 694 resharpening episodes with which to train the predictive models.

### **Feature selection.**

Based on previous research [[22](#ref-kuhn_geometric_1990),[33](#ref-dogandzic_edge_2015),[34](#ref-bustos-perez_predicting_2021),[44](#ref-hiscock_experimental_2005)] variables were recorded as predictive features. After each episode of retouch, the following variables were recorded:

* Remaining scraper mass, recorded in grams with a Sytech SY-BS502 scale and a precision of 0.01 g. This variable is selected since machine learning models will consider remaining flake mass as a baseline on minimum mass of the scraper.
* Thickness at the midpoint of the flake (measured in mm with a precision of 0.1). It is considered that thickness remains relatively unchanged as resharpening increases [[19](#ref-shott_use_2017)], and its addition tends to increase the predictive power of models [[33](#ref-dogandzic_edge_2015)].
* Maximum thickness of the flake measured in mm (with a precision of 0.1). Feature selection through all possible combination of variables indicates that the logarithmic (base 10) transformation of this variable can increase the predictive power of regression models for freehand knapped flakes. Logarithmic transformation can result in Gaussian distribution of feature values increasing the predictive power of a model [[34](#ref-bustos-perez_predicting_2021)]. It is also considered that as resharpening proceeds, the thickness at the midpoint will be displaced (since length and width will diminish), while maximum thickness will remain more stable.
* Three equidistant measures of height of retouch (*t*) and the corresponding thickness (*T*) of the flake [[44](#ref-hiscock_experimental_2005)] measured in mm (with a precision of 0.1). The average of these three points is used as a predictive feature. Here it is considered that the average height of the retouch will serve as a proxy of mass removed from the scrapper.
* The GIUR index proposed by Kuhn [[22](#ref-kuhn_geometric_1990)]. This index divides the height of retouch (*t*) by its corresponding thickness (*T*). As previously indicated, the present study records three equidistant heights of retouch (*t*), each being divided by their corresponding flake thickness (*T*). The GIUR value is calculated as the average of these three divisions. GIUR values can range from 0 (unretouched flake) to 1 (when the height of the reaches the dorsal side of a flake).  
  Variables selected for training the regression models were: scraper mass, maximum thickness (log transformed), average height of retouch (t) and value of the GIUR index.

### **Regression models and evaluation.**

Four methods were employed for regression analysis: multiple linear regression, support vector regression with a linear kernel, random forest and gradient boosting machine.

Multiple linear regression (MLR) extends the simple linear regression in such a way that it can directly accommodate multiple predictors [[45](#ref-james_introduction_2013) : 71].

Support vector machines for regression [[46](#ref-awad_support_2015)–[48](#ref-smola_tutorial_2004)] with a linear kernel (SVML) fit a linear hyperplane and a margin of error which allows for errors of points falling inside the margin. Points falling outside the margin define the support vectors. This provides a model focused on the general trend which aims to maximize the margin while minimizing the error, and which is also robust to the presence of outliers.

Random forest for regression selects random samples of the data and builds decision trees for prediction [[49](#ref-breiman_random_2001)]. As a result, each tree is built from different combinations of the data, and the average is used as prediction. This adds diversity, reduces overfit, and provides high-accuracy predictions [[50](#ref-lantz_machine_2019)].

The gradient boosting machine (GBM) is an ensemble method that builds up a final model by incrementally improving an existing one [[51](#ref-friedman_greedy_2001),[52](#ref-friedman_stochastic_2002)]. The first model uses an initial *“shallow tree”* with a constant value (average of the labels). Following this initial model, a new tree (*weak learner*) is fitted to predict the residuals of the model, contributing to the final model predictions, allowing for correction of the errors of the model. This process is repeated, allowing it to progressively identify the shortcomings of week learners on a sequence of decision trees and to reduce the errors of the ensemble predictions.

Initial models are trained to estimate original scraper mass based on the set of selected attributes. However, as previously stated, the objective is to evaluate the ability of a model to predict the curation ratio of a stone tool (percentage of mass remaining relative to its original mass). Calculating the curation ratio of a stone tool can be formulated as:

Where:  
M = mass (directly measured on the scraper).  
EOM = estimated original mass (provided by the model).

Models of both predictions (original scraper mass and curation ratio) are compared using four measures of performance: , MAE, RMSE, and MAPE. is a measure of linear correlation and of how much of the observed variation is explained by the model [[45](#ref-james_introduction_2013)]. In lithic studies, a categorization of the predictive power of indices has been proposed based on their values, where <0.1 is low, 0.1–0.25 is moderate, 0.26–0.5 is fairly large/strong, 0.51–0.8 is very large/strong and >0.8 is extremely large/strong. However, it is important to consider that different distributions of data can result in same or similar values [[53](#ref-anscombe_graphs_1973)].

Mean average error (MAE), root mean squared error (RMSE), and mean average percentual error (MAPE) provide summary values of how far predictions fall from the true value [[45](#ref-james_introduction_2013),[50](#ref-lantz_machine_2019)]. MAE measures the average magnitude of errors, regardless of signal. RMSE also provides a measure of distance between predicted and actual values, although it punishes large errors. RMSE is usually compared to the standard deviation (SD) of the variable to be predicted. If RMSE presents a lower value than the SD, this is indicative of a good model which predicts values better than taking the average value of the sample. MAPE provides a measure of distance on a proportional basis (a residual of 3 g in a 12 g flake will be much higher than in a 50 g flake). A perfect model has a MAE, RMSE, and MAPE values of 0, and in general, better models will have lower values of MAE, RMSE, and MAPE.

Collinearity of the predictors is addressed through the variance inflation factor (VIF). VIF provides a measure of correlation between predictors and their effects on the model. In the present study VIF is calculated using the R package car v.3.1.2. [[54](#ref-fox_r_2018)]. Common thresholds for VIF values [[55](#ref-marquardt_generalized_1970),[56](#ref-obrien_caution_2007)] range between 1 to 10 (considered inconsequential); 10 to 30 (cause for concern), and higher than 30 (seriously harmful). At present, the package car v.3.1.2 only allows for the calculation of the variance inflation factor for multiple linear regression. Although the different nature of the regression algorithms can result in different effects of collinearity, results from calculating the variance inflation factor in the multiple linear regression can be extrapolated to the rest of the models.

Models were evaluated using a k-fold cross validation. In k-fold cross validation, the dataset is randomly shuffled and divided into k folds. The first fold is employed as a test set, and the model is trained in the remaining folds. After this, the second fold is employed as a test set and the rest as a new training set. This process continues until all folds have served as a test set. Since the samples in each fold are determined by the initial random shuffle, it is advisable to repeat this cycle a series of times. The present work employs a 10-fold cross validation (six folds having a sample of 69 elements, and four folds having a sample of 70 elements) which is repeated 50 times with an initial random shuffling.

In addition to the above listed performance metrics, all models are graphically evaluated. A regression plot provides a scatter plot of predicted and true values along its regression line. In a good model, the regression line will pass through the center of all points, which will be evenly distributed above and below, and outliers are visible.

The distribution of residuals (difference between predicted and actual values) allows for an evaluation of bias in the model. In the present study residuals are used in combination with transversal section and different intervals of the GIUR index in order to determine possible bias or limitations of the model. Residuals plots of a good model will have the points evenly distributed around zero across the range of the independent variables, and no statistical difference will exist between the categories of flake transversal section or GIUR intervals. Evaluation of residuals is only undertaken for the considered best model.

The evaluation of models using the k-fold cross validation was performed using the complete dataset of 698 resharpening episodes from the 134 flakes. This implies that, for a prediction, previous and posterior resharpening episodes of the same flake were included in the training data set. This raises the question whether the model is overfitting from seeing previous and posterior resharpening examples of the same flake. In order to evaluate this possible source of overfitting, a random selection of 10% (n = 13) of flakes and all their resharpening episodes were removed from the training set and the remaining sample used to train the previously selected best model. This process was repeated until 100000 predictions were obtained and then the four measures of performance were calculated. In this scenario, a significant decrease in model performance suggests overfitting in the full sample that includes multiple stages of the same flake.

Additionally, two models were trained for comparison using the sample of 134 flakes. The first model uses the same set of variables from Bustos-Pérez & Baena Preysler [[35](#ref-bustos-perez_multiple_2022)] in order to predict original flake mass (log transformed). The second model was used log transformations of platform size and maximum thickness along with exterior platform angle as predictive variables in order to predict the cubic root of flake mass. Both models were trained using a multiple linear regression, and predicted values were transformed to the linear scale in order to compare performance metrics with those of the best model from the current study.

The complete workflow was developed in RStudio IDE v.2024.04.02 [[57](#ref-rstudio_team_rstudio_2019)] using the R programming language v.4.4.1 [[58](#ref-r_core_team_r_2019)]. The package tidyverse v.2.0.0 [[59](#ref-wickham_welcome_2019)] was employed for data manipulation and representation. Multiple linear regression uses the R package MASS v.7.3.60.2 [[60](#ref-venables_modern_2002)], SVM with linear kernel uses package kernlab v.0.9.32 [[61](#ref-karatzoglou_kernlab_2004)], the random forest model uses packages e1071 v.1.7.14 and ranger v.0.16.0 [[62](#ref-dimitriadou_misc_2008),[63](#ref-wright_ranger_2015)], and GBM uses package gbm v.2.2.2 [[64](#ref-ridgeway_generalized_2007)]. Training and validation of models was done using the package caret v.6.0.94 [[65](#ref-kuhn_building_2008)]. The “Original Scraper Mass Calculator” which implements the model through a user friendly interface was written using the shiny package v.1.8.1.1 [[66](#ref-chang_package_2015)–[68](#ref-wickham_mastering_2021)]. All data, code is made publicly available through a public repository organized following the structure of a research compendium [69] and using a markdown document through package bookdown v.0.39 [[69](#ref-mcnamara_dynamic_2014)–[71](#ref-xie_knitr_2014)].  
All data and the complete workflow of analysis is available as a research compendium [[72](#ref-marwick_packaging_2018)] at Github (<https://github.com/GuillermoBustosPerez/Scraper-Original-Mass/tree/main>). The complete code and files of the Original Scraper Mass calculator v.1.0.0 is also available at Github (<https://github.com/GuillermoBustosPerez/Original-Scraper-Mass-Calculator>), and the final implementation of the application can be accessed at: <https://guillermo-bustos-perez.shinyapps.io/Original-Scraper-Mass-Calculator/>

### Training of the models

The following script trains the above described models. For the SVM linear regression, random forest and GBM hyperparameter cartesian grid search is performed, and final models are trained with the best selection of hyperparameters. Final models can be accessed at <Models> folder.

source("/Scripts/11-Training-models.R")

The following line of code trains the random forest model by removing all resharpening events of 10% of flakes. This is dne in order to determine if the model is benefiting from seeing previous episodes of a same flake, thus resulting in overfit.

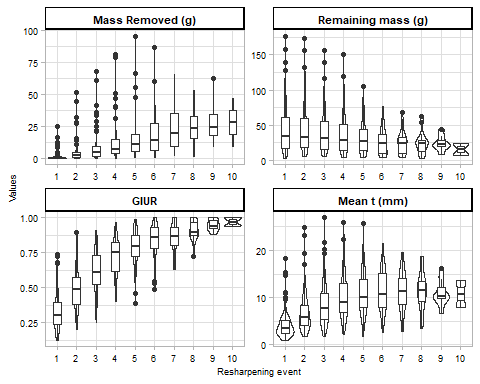
load("Models/Best-Subset-Random-Forest.RData")  
load("Models/Best-Subset-MLR.RData")  
load("Models/Best-Subset-SVM-Linear.RData")  
load("Models/Best-Subset-GBM.RData")

## **Results**

### **Resharpening effects on the experimental assemblage**

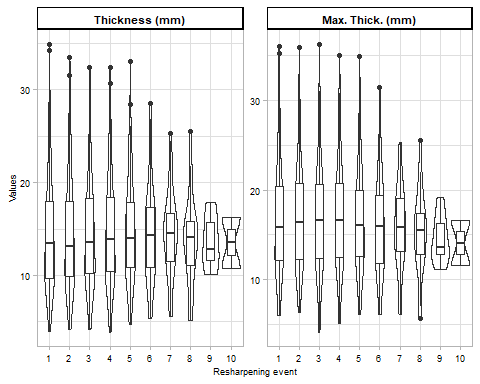
Figure 2 presents the effects of each resharpening episode on the experimental assemblage. On average, flakes from the first episode of retouch had 1.84 g removed (with the exception of an outlier flake which had 25 g removed). Maximum value of mass removed was of 95.5 g after five episodes of retouch, while flakes reaching ten episodes of retouch had an average of 28.4 g removed. As for the retouched pieces, mass of the assemblage decreased from 45.9 g after the first episode of retouch to 15.9 g after ten episodes of retouch. On average, the resharpening episodes removed 20% (with a standard deviation of 15.9%) of mass from the scrapers. One resharpening episode removed a minimum of 0.513%, and a resharpening episode removed a maximum of 67.3% of mass. 50% of the resharpening episodes removed between 5.9% and 31.3% of mass.

# Resharpenning   
Data2 %>% select(Episode, GIUR, W.Retrieved.g, Mean.t, Rem.Weight.g) %>%   
 pivot\_longer(  
 cols = c(GIUR:Rem.Weight.g),  
 values\_to = "Values",  
 names\_to = "Variables") %>%   
   
 ggplot(aes(as.factor(Episode), Values)) +  
 facet\_wrap(~ factor(Variables,   
 levels = c("W.Retrieved.g", "Rem.Weight.g", "GIUR", "Mean.t"),  
 labels = c( "Mass Removed (g)", "Remaining mass (g)", "GIUR", "Mean t (mm)")) ,   
 scales = "free") +  
 geom\_violin() +  
 geom\_boxplot(width = 0.4) +   
 theme\_light() +  
 xlab("Resharpening event") +  
 theme(  
 axis.text = element\_text(color = "black", size = 7),  
 axis.title = element\_text(color = "black", size = 7),  
 strip.text = element\_text(color = "black", face = "bold", size = 8),  
 strip.background = element\_rect(fill = "white", colour = "black", linewidth = 1))



In general, the GIUR remains fairly unidirectional, with increasing values after each episode of resharpening. The average GIUR after the first resharpening episode was 0.33. The first maximum GIUR value (1) was reached after six episodes of resharpening. Only six flakes of the experimental assemblage reached a GIUR values of 1, providing a total of eight different episodes. Mean height of retouch (t) increases during the first seven resharpening episodes, from an average value of 4.19 mm after the first episode, to a value of 11.4 mm after the seventh episodes. Average height of retouch (t) decreases progressively afterwards to a value of 10.8 mm on episode 10 of retouch. This decrease in average t is result of retouch reaching the upper dorsal side of flakes and diminishing the overall thickness.  
Both thickness at the midpoint and maximum thickness (Figure 3) remain fairly constant among all the resharpening events, although the range of values decreases with each resharpening episode (this is due to a decrease in the number of flakes due to discard). After eight resharpening events, a light decrease in both measures of thickness is observed due to retouch reaching the upper dorsal part of the flake and reducing its overall thickness.

Data2 %>% select(Episode, T.mm, Max.Thick.mm) %>%   
 pivot\_longer(  
 cols = c(T.mm:Max.Thick.mm),  
 values\_to = "Values",  
 names\_to = "Variables") %>%   
   
 ggplot(aes(as.factor(Episode), Values)) +  
 facet\_wrap(~ factor(Variables,   
 levels = c("T.mm", "Max.Thick.mm"),  
 labels = c("Thickness (mm)", "Max. Thick. (mm)")) ,   
 scales = "free") +  
 geom\_violin() +  
 geom\_boxplot(width = 0.4) +   
 theme\_light() +  
 xlab("Resharpening event") +  
 theme(  
 axis.text = element\_text(color = "black", size = 7),  
 axis.title = element\_text(color = "black", size = 7),  
 strip.text = element\_text(color = "black", face = "bold", size = 8),  
 strip.background = element\_rect(fill = "white", colour = "black", linewidth = 1))



### **Evaluation of regression models**

Table 2 presents performance metrics of each of the regression models for predicting original flake mass. All models had RMSE values lower than the standard deviation of original flake mass, indicating a good fit of the models. The random forest model presented the best values for all evaluation metrics, being able to capture 0.974 of variance, and with a MAE of 3.297 g. It also presented the lowest RMSE value (5.917), indicative of being the least affected by outliers in the predictions. The lowest MAPE value of 6.775 indicates that the random forest model is also the least affected by the size of the original blanks. The GBM model has evaluation metrics similar to those of random forest, being able to capture a similar proportion of variance (0.971) and slightly higher values of MAE (3.549) and RMSE (6.185). In general, multiple linear regression was the model with the worst performance metrics, capturing slightly more variance than the SVM (respective values of 0.963 and 0.961) and being less affected by the presence of outliers (respective values of 7.036 and 7.265).

Figure 4 presents the regression plots of all trained models. All models present regression lines with evenly distributed predictions, indicative of an absence of bias. However, it is observed for the multiple linear regression and SVM that with increasing flake mass, the range of predictions becomes more dispersed. None of the predictive variables in the multiple linear regression model presented VIF values above the threshold of 10. Average height of retouch (t) presented the highest VIF value (7.27), followed by the GIUR index (4.59). Log transformed maximum thickness and remaining scraper mass presented respective VIF values of 3.81 and 2.35.

data.frame(  
 Model = c("Mult. Linear Reg.", "SVM Linear", "Random Forest", "GBM"),  
 r2 = c(  
 round(MLR\_Model.BS$results[[3]], 4),  
 round(SVML\_Model.BS$results[[3]], 4),  
 round(RF\_Model.BS$results[[5]], 4),  
 round(GBM\_Model.BS$results[[6]], 4)),  
 MAE = c(  
 round(MLR\_Model.BS$results[[4]], 3),  
 round(SVML\_Model.BS$results[[4]], 3),  
 round(RF\_Model.BS$results[[6]], 3),  
 round(GBM\_Model.BS$results[[7]], 3)),  
 RMSE = c(  
 round(MLR\_Model.BS$results[[2]], 3),  
 round(SVML\_Model.BS$results[[2]], 3),  
 round(RF\_Model.BS$results[[4]], 3),  
 round(GBM\_Model.BS$results[[5]], 3)),  
 MAPE = c(  
 round(MLmetrics::MAPE(MLR\_Model.BS$pred$pred, MLR\_Model.BS$pred$obs)\*100, 3),  
 round(MLmetrics::MAPE(SVML\_Model.BS$pred$pred, SVML\_Model.BS$pred$obs)\*100, 3),  
 round(MLmetrics::MAPE(RF\_Model.BS$pred$pred, RF\_Model.BS$pred$obs)\*100, 3),  
 round(MLmetrics::MAPE(GBM\_Model.BS$pred$pred, GBM\_Model.BS$pred$obs)\*100, 3))  
)

## Model r2 MAE RMSE MAPE  
## 1 Mult. Linear Reg. 0.9629 4.672 7.036 12.269  
## 2 SVM Linear 0.9613 4.394 7.265 11.304  
## 3 Random Forest 0.9740 3.297 5.917 6.775  
## 4 GBM 0.9710 3.549 6.185 7.705

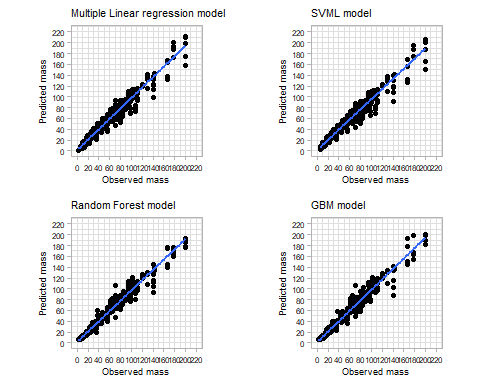
#### Put all predictions of each model ####  
  
Predicted.MLR <- as.data.frame(MLR\_Model.BS$pred) %>%   
 group\_by(rowIndex) %>%   
 summarise(  
 mean.pred = round(mean(pred),2),  
 mean.obs = round(mean(obs),2)) %>%   
 mutate(  
 Rema.Weight = Data2$Rem.Weight,  
 Real.W.Retrieved = Data2$W.Retrieved) %>%   
 mutate(  
 Pred.W.Retrieved = (mean.pred - Rema.Weight),  
 Pred.Curated = 100 - (Rema.Weight/mean.pred)\*100,  
 Real.Curated = 100 - (Rema.Weight/mean.obs )\*100,  
 Residuals = mean.obs - mean.pred,  
 Abs.W.Ret = abs(Pred.W.Retrieved),  
 Pred.Curated2 = 100 - (Rema.Weight/mean.pred)\*100)  
  
  
Predicted.SVML <- as.data.frame(SVML\_Model.BS$pred) %>%   
 group\_by(rowIndex) %>%   
 summarise(  
 mean.pred = round(mean(pred),2),  
 mean.obs = round(mean(obs),2)) %>%   
 mutate(  
 Rema.Weight = Data2$Rem.Weight,  
 Real.W.Retrieved = Data2$W.Retrieved) %>%   
 mutate(  
 Pred.W.Retrieved = (mean.pred - Rema.Weight),  
 Pred.Curated = 100 - (Rema.Weight/mean.pred)\*100,  
 Real.Curated = 100 - (Rema.Weight/mean.obs )\*100,  
 Residuals = mean.obs - mean.pred,  
 Abs.W.Ret = abs(Pred.W.Retrieved),  
 Pred.Curated2 = 100 - (Rema.Weight/mean.pred)\*100)  
  
  
Predicted.RF <- as.data.frame(RF\_Model.BS$pred) %>%   
 group\_by(rowIndex) %>%   
 summarise(  
 mean.pred = round(mean(pred),2),  
 mean.obs = round(mean(obs),2)) %>%   
 mutate(  
 Rema.Weight = Data2$Rem.Weight,  
 Real.W.Retrieved = Data2$W.Retrieved) %>%   
 mutate(  
 Pred.W.Retrieved = (mean.pred - Rema.Weight),  
 Pred.Curated = 100 - (Rema.Weight/mean.pred)\*100,  
 Real.Curated = 100 - (Rema.Weight/mean.obs )\*100,  
 Residuals = mean.obs - mean.pred,  
 Abs.W.Ret = abs(Pred.W.Retrieved),  
 Pred.Curated2 = 100 - (Rema.Weight/mean.pred)\*100)  
  
  
Predicted.GBM <- as.data.frame(GBM\_Model.BS$pred) %>%   
 group\_by(rowIndex) %>%   
 summarise(  
 mean.pred = round(mean(pred),2),  
 mean.obs = round(mean(obs),2)) %>%   
 mutate(  
 Rema.Weight = Data2$Rem.Weight,  
 Real.W.Retrieved = Data2$W.Retrieved) %>%   
 mutate(  
 Pred.W.Retrieved = (mean.pred - Rema.Weight),  
 Pred.Curated = 100 - (Rema.Weight/mean.pred)\*100,  
 Real.Curated = 100 - (Rema.Weight/mean.obs )\*100,  
 Residuals = mean.obs - mean.pred,  
 Abs.W.Ret = abs(Pred.W.Retrieved),  
 Pred.Curated2 = 100 - (Rema.Weight/mean.pred)\*100)  
  
  
#### Plots of each regression model #####  
  
ggpubr::ggarrange(  
   
 (  
 Predicted.MLR %>%   
 ggplot(aes(mean.obs, mean.pred)) +  
 geom\_point() +  
 scale\_x\_continuous(breaks = seq(0, 220, 20), lim = c(0, 220)) +  
 scale\_y\_continuous(breaks = seq(0, 220, 20), lim = c(0, 220)) +  
 ggtitle(label = "Multiple Linear regression model") +   
 geom\_smooth(method = "lm") +  
 coord\_fixed() +  
 theme\_light() +  
 ylab("Predicted mass") +  
 xlab("Observed mass") +  
 theme(  
 axis.title = element\_text(size = 7),  
 axis.text = element\_text(size = 6, color = "black"),  
 plot.title = element\_text(size = 8))  
 ),  
 (  
 Predicted.SVML %>%   
 ggplot(aes(mean.obs, mean.pred)) +  
 geom\_point() +  
 scale\_x\_continuous(breaks = seq(0, 220, 20), lim = c(0, 220)) +  
 scale\_y\_continuous(breaks = seq(0, 220, 20), lim = c(0, 220)) +  
 ggtitle(label = "SVML model") +   
 geom\_smooth(method = "lm") +  
 coord\_fixed() +  
 theme\_light() +  
 ylab("Predicted mass") +  
 xlab("Observed mass") +  
 theme(  
 axis.title = element\_text(size = 7),  
 axis.text = element\_text(size = 6, color = "black"),  
 plot.title = element\_text(size = 8))  
 ),  
   
 (  
 Predicted.RF %>%   
 ggplot(aes(mean.obs, mean.pred)) +  
 geom\_point() +  
 scale\_x\_continuous(breaks = seq(0, 220, 20), lim = c(0, 220)) +  
 scale\_y\_continuous(breaks = seq(0, 220, 20), lim = c(0, 220)) +  
 ggtitle(label = "Random Forest model") +   
 geom\_smooth(method = "lm") +  
 coord\_fixed() +  
 theme\_light() +  
 ylab("Predicted mass") +  
 xlab("Observed mass") +  
 theme(  
 axis.title = element\_text(size = 7),  
 axis.text = element\_text(size = 6, color = "black"),  
 plot.title = element\_text(size = 8))  
 ),  
 (  
 Predicted.GBM %>%   
 ggplot(aes(mean.obs, mean.pred)) +  
 geom\_point() +  
 scale\_x\_continuous(breaks = seq(0, 220, 20), lim = c(0, 220)) +  
 scale\_y\_continuous(breaks = seq(0, 220, 20), lim = c(0, 220)) +  
 ggtitle(label = "GBM model") +   
 geom\_smooth(method = "lm") +  
 coord\_fixed() +  
 theme\_light() +  
 ylab("Predicted mass") +  
 xlab("Observed mass") +  
 theme(  
 axis.title = element\_text(size = 7),  
 axis.text = element\_text(size = 6, color = "black"),  
 plot.title = element\_text(size = 8))  
 ),  
 ncol = 2,  
 nrow = 2  
)

## `geom\_smooth()` using formula = 'y ~ x'  
## `geom\_smooth()` using formula = 'y ~ x'

## Warning: Removed 1 row containing non-finite outside the scale range  
## (`stat\_smooth()`).

## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom\_point()`).

## `geom\_smooth()` using formula = 'y ~ x'  
## `geom\_smooth()` using formula = 'y ~ x'



Percentage of flake mass consumed by retouch (curation) was calculated using estimations of original flake mass for each model and each episode of resharpening. Table 3 presents performance metrics of the four models when predicting the curation ratio, and Figure 5 presents their corresponding regression plots. Important differences can be observed between models when predicting percentage of mass lost by retouch based on predictions of original mass. Of the four models, only random forest and GBM presented adequate performance metrics. Multiple linear regression presented the lowest performance values, with a linear correlation value of 0.113, while SVM with linear kernel presented a significant higher (but still unsatisfactory) r2 value of 0.399. Both models also presented RMSE values higher than the standard deviation of percentage of mass lost through retouch (SD = 15.92). Visual evaluation of the regression plots indicates that the low performance metrics from both models (multiple linear regression and SVM with linear kernel) seem to be caused by underestimations original flake mass when a small percentage of flake mass has been removed by retouch. As a result, multiple linear regression estimated an original scraper mass value lower than remaining scraper mass in 127 cases. SVM with linear kernel estimated an original scraper mass lower than remaining scraper mass in 118 cases.

ggpubr::ggarrange(  
 (  
 Predicted.MLR %>%   
 ggplot(aes(Real.Curated, Pred.Curated)) +  
 geom\_point(alpha = 0.6) +  
 geom\_smooth(method = "lm") +  
 scale\_x\_continuous(breaks = seq(-30, 70, 10), lim = c(-30, 70)) +  
 scale\_y\_continuous(breaks = seq(-30, 70, 10), lim = c(-30, 70)) +  
 ggtitle(label = "Multiple Linear Regression model") +  
 coord\_fixed() +  
 theme\_light() +  
 geom\_vline(xintercept = 0, linetype = "dashed", alpha = 0.75) +  
 geom\_hline(yintercept = 0, linetype = "dashed", alpha = 0.75) +  
 ylab("Predicted % of mass lost by retouch (curation)") +  
 xlab("Observed % of mass lost by retouch (curation)") +  
 theme(  
 axis.title = element\_text(size = 6.5),  
 axis.text = element\_text(size = 6, color = "black"),  
 plot.title = element\_text(size = 8))  
 ),  
 (  
 Predicted.SVML %>%   
 ggplot(aes(Real.Curated, Pred.Curated)) +  
 geom\_point(alpha = 0.6) +  
 geom\_smooth(method = "lm") +  
 scale\_x\_continuous(breaks = seq(-30, 70, 10), lim = c(-30, 70)) +  
 scale\_y\_continuous(breaks = seq(-30, 70, 10), lim = c(-30, 70)) +  
 ggtitle(label = "SVML model") +  
 coord\_fixed() +  
 theme\_light() +  
 geom\_vline(xintercept = 0, linetype = "dashed", alpha = 0.75) +  
 geom\_hline(yintercept = 0, linetype = "dashed", alpha = 0.75) +  
 ylab("Predicted % of mass lost by retouch (curation)") +  
 xlab("Observed % of mass lost by retouch (curation)") +  
 theme(  
 axis.title = element\_text(size = 6.5),  
 axis.text = element\_text(size = 6, color = "black"),  
 plot.title = element\_text(size = 8))  
 ),  
 (  
 Predicted.RF %>%   
 ggplot(aes(Real.Curated, Pred.Curated)) +  
 geom\_point(alpha = 0.6) +  
 geom\_smooth(method = "lm") +  
 scale\_x\_continuous(breaks = seq(-15, 70, 10), lim = c(-15, 70)) +  
 scale\_y\_continuous(breaks = seq(-15, 70, 10), lim = c(-15, 70)) +  
 ggtitle(label = "Random Forest Model") +  
 coord\_fixed() +  
 theme\_light() +  
 geom\_vline(xintercept = 0, linetype = "dashed", alpha = 0.75) +  
 geom\_hline(yintercept = 0, linetype = "dashed", alpha = 0.75) +  
 ylab("Predicted % of mass lost by retouch (curation)") +  
 xlab("Observed % of mass lost by retouch (curation)") +  
 theme(  
 axis.title = element\_text(size = 6.5),  
 axis.text = element\_text(size = 6, color = "black"),  
 plot.title = element\_text(size = 8))  
 ),  
 (  
 Predicted.GBM %>%   
 ggplot(aes(Real.Curated, Pred.Curated)) +  
 geom\_point(alpha = 0.6) +  
 geom\_smooth(method = "lm") +  
 scale\_x\_continuous(breaks = seq(-15, 70, 10), lim = c(-15, 70)) +  
 scale\_y\_continuous(breaks = seq(-15, 70, 10), lim = c(-15, 70)) +  
 ggtitle(label = "GBM model") +   
 coord\_fixed() +  
 theme\_light() +  
 geom\_vline(xintercept = 0, linetype = "dashed", alpha = 0.75) +  
 geom\_hline(yintercept = 0, linetype = "dashed", alpha = 0.75) +  
 ylab("Predicted % of mass lost by retouch (curation)") +  
 xlab("Observed % of mass lost by retouch (curation)") +  
 theme(  
 axis.title = element\_text(size = 6.5),  
 axis.text = element\_text(size = 6, color = "black"),  
 plot.title = element\_text(size = 8))  
 ),  
 ncol = 2,  
 nrow = 2)

## `geom\_smooth()` using formula = 'y ~ x'

## Warning: Removed 22 rows containing non-finite outside the scale range  
## (`stat\_smooth()`).

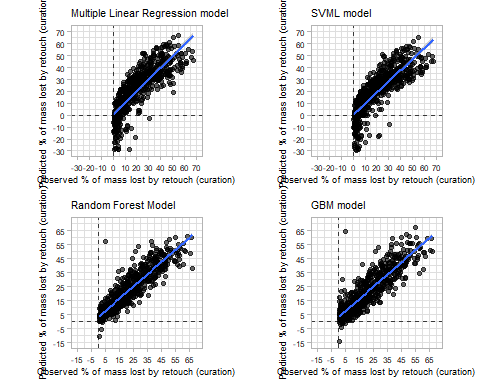
## Warning: Removed 22 rows containing missing values or values outside the scale range  
## (`geom\_point()`).

## `geom\_smooth()` using formula = 'y ~ x'

## Warning: Removed 21 rows containing non-finite outside the scale range  
## (`stat\_smooth()`).

## Warning: Removed 21 rows containing missing values or values outside the scale range  
## (`geom\_point()`).

## `geom\_smooth()` using formula = 'y ~ x'  
## `geom\_smooth()` using formula = 'y ~ x'



Linear correlation values indicate that random forest and GBM perform at least twice as well than the SVM with linear kernel, with respective r2 values of 0.839 (random forest), and 0.805 (GBM). Both models also present RMSE values lower than the SD of the sample, indicative of a good fit of the model. The random forest model presents the best performance metrics, with the highest linear correlation value (r2 = 0.839) and lowest values for MAE (4.662), RMSE (6.485) and MAPE (0.365). Random forest is closely followed by the GBM model, with a linear correlation value of 0.805. Visual evaluation of the regression plots (Figure 5) reinforces the notion of the good performance of random forest and GBM. In both cases prediction points are evenly distributed among the regression lines, and clustering of points in the lower values (due to over or underestimation of curation) are absent. GBM estimated an original scraper mass lower than the remaining scraper mass in 8 cases, while random forest made the same error in only three cases.

data.frame(  
 Model = c("MLR", "SVML", "Random Forest", "GBM"),  
 `r2 W retrieved` = c(  
 summary(lm(Real.W.Retrieved ~ Abs.W.Ret, Predicted.MLR))$r.squared %>% round(3),  
 summary(lm(Real.W.Retrieved ~ Abs.W.Ret, Predicted.SVML))$r.squared %>% round(3),  
 summary(lm(Real.W.Retrieved ~ Abs.W.Ret, Predicted.RF))$r.squared %>% round(3),  
 summary(lm(Real.W.Retrieved ~ Abs.W.Ret, Predicted.GBM))$r.squared %>% round(3)  
 ),  
 `r2 Curated` = c(  
 summary(lm(Real.Curated ~ Pred.Curated2, Predicted.MLR))$r.squared %>% round(3),  
 summary(lm(Real.Curated ~ Pred.Curated2, Predicted.SVML))$r.squared %>% round(3),  
 summary(lm(Real.Curated ~ Pred.Curated2, Predicted.RF))$r.squared %>% round(3),  
 summary(lm(Real.Curated ~ Pred.Curated2, Predicted.GBM))$r.squared %>% round(3)),  
 MAE = c(  
 round(caret::MAE(Predicted.MLR$Real.Curated, Predicted.MLR$Pred.Curated2), 3),  
 round(caret::MAE(Predicted.SVML$Real.Curated, Predicted.SVML$Pred.Curated2),3),  
 round(caret::MAE(Predicted.RF$Real.Curated, Predicted.RF$Pred.Curated2), 3),  
 round(caret::MAE(Predicted.GBM$Real.Curated, Predicted.GBM$Pred.Curated2), 3)),  
 RMSE = c(  
 round(caret::RMSE(Predicted.MLR$Real.Curated, Predicted.MLR$Pred.Curated2), 3),  
 round(caret::RMSE(Predicted.SVML$Real.Curated, Predicted.SVML$Pred.Curated2),3),  
 round(caret::RMSE(Predicted.RF$Real.Curated, Predicted.RF$Pred.Curated2), 3),  
 round(caret::RMSE(Predicted.GBM$Real.Curated, Predicted.GBM$Pred.Curated2), 3)),  
 MAPE = c(  
 round(MLmetrics::MAPE(Predicted.MLR$Real.Curated, Predicted.MLR$Pred.Curated2), 3),  
 round(MLmetrics::MAPE(Predicted.SVML$Real.Curated, Predicted.SVML$Pred.Curated2), 3),  
 round(MLmetrics::MAPE(Predicted.RF$Real.Curated, Predicted.RF$Pred.Curated2), 3),  
 round(MLmetrics::MAPE(Predicted.GBM$Real.Curated, Predicted.GBM$Pred.Curated2), 3)))

## Model r2.W.retrieved r2.Curated MAE RMSE MAPE  
## 1 MLR 0.735 0.113 11.726 54.429 1.037  
## 2 SVML 0.735 0.399 9.811 20.064 1.449  
## 3 Random Forest 0.805 0.839 4.662 6.485 0.365  
## 4 GBM 0.790 0.805 5.189 7.160 0.449

### **Bias, limitations of the best model (random forest) and comparison with other models.**

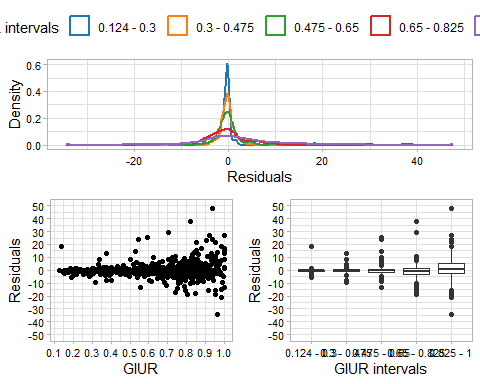
The episodes of resharpening generated 130 cases where the flat flake problem is observed and 564 cases where this problem is not observed. Students t-test comparing the residual distribution of flat and non-flat flakes shows no statistical differences (t = -1.93; p = 0.17) for all episodes of resharpening. When selecting flakes that had four or more resharpening episodes, no statistical significance is observed for the residual distribution (t = -0.79; p = 0.43). This indicates that for the given sample and the given predictive variables, random forest predictions are not affected by the flat flake problem.

Figure 6 presents the residual (when predicting original flake mass) distribution of continuous GIUR values and the same divided into five intervals. A density plot indicates that, on a general level, residuals tend to peak at the zero value. However, with increasing GIUR intervals the peak among the zero value diminishes. Scatter and box plots also indicate that, for higher GIUR values and intervals, a greater range of residual values is present. This indicates that, although residuals are evenly distributed around 0, the accuracy of predictions from the random forest diminishes among heavily retouched flakes (with GIUR values above 0.8).

Although knowing how many resharpening episodes an archaeological stone has undergone tool is unlikely, the availability of the data in a controlled experimentation allows for a better understanding of possible bias in the model. For the first two episodes of retouch, statistical differences were present between values of actual curation and estimated curation (t = 4.11; p < 0.001 and t = 2.31; p = 0.02). Statistical differences between estimated and actual curation ratios are absent for resharpening episodes three to seven. Statistically significant differences between estimated and real curated values are present for scrapers which underwent eight or more resharpening episodes (t = -3.73; p < 0.001).

Pred.RF <- RF\_Model.BS$pred %>%  
 group\_by(rowIndex) %>%  
 summarise(  
 pred = mean(pred),  
 obs = mean(obs)) %>%  
 mutate(  
 Residuals = obs - pred)  
  
  
Pred.RF <- Data2 %>%  
 select(Flake.ID, Sec.Trans, Termin, Episode,  
 W.Retrieved.g, Rem.Weight.g, Mean.t, Log.M.Thick, GIUR) %>%  
 cbind(Pred.RF) %>% mutate(  
 Cat.GIUR = cut(GIUR, 5)) %>%   
 mutate(  
 Cat.GIUR = case\_when(  
 Cat.GIUR == "(0.124,0.3]" ~ "0.124 - 0.3",  
 Cat.GIUR == "(0.3,0.475]" ~ "0.3 - 0.475",  
 Cat.GIUR == "(0.475,0.65]" ~ "0.475 - 0.65",  
 Cat.GIUR == "(0.65,0.825]" ~ "0.65 - 0.825",  
 Cat.GIUR == "(0.825,1]" ~ "0.825 - 1")  
 )  
  
ggpubr::ggarrange(  
 (  
 Pred.RF %>% ggplot(aes(Residuals, color = Cat.GIUR)) +  
 geom\_density(size = 1) +   
 theme\_light() +  
 ggsci::scale\_color\_d3() +  
 labs(color = "GIUR intervals") +  
 ylab("Density") +  
 theme(  
 legend.position = "top",  
 axis.text = element\_text(size = 8, color = "black"))   
 ),  
   
 (  
 ggpubr::ggarrange(  
 (  
 Pred.RF %>% ggplot(aes(GIUR, Residuals)) +  
 geom\_point() +  
 scale\_x\_continuous(breaks = seq(0, 1, by = 0.1)) +   
 scale\_y\_continuous(breaks = seq(-50, 50, 10), lim = c(-50, 50)) +  
 theme\_light() +  
 theme(  
 axis.text = element\_text(color = "black", size = 8)  
 )  
 ),  
 (  
 Pred.RF %>% ggplot(aes(Cat.GIUR, Residuals)) +  
 geom\_boxplot() +  
 scale\_y\_continuous(breaks = seq(-50, 50, 10), lim = c(-50, 50)) +   
 theme\_light() +  
 xlab("GIUR intervals") +  
 theme(  
 axis.text = element\_text(color = "black", size = 8)  
 )  
 ),   
 ncol = 2)  
 ),  
 nrow = 2  
 )

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.  
## ℹ Please use `linewidth` instead.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was  
## generated.



read.csv("Data/RF-results-leaving-flake-sequences.csv")

## r2 MAE RMSE MAPE  
## 1 0.96 3.989 7.478 7.929

Figure 7 presents real and estimated values of curation according to each resharpening episode. Considering all episodes of retouch, no statistically significant difference is present between values of predicted and actual curation (t = 1.12, df = 1382.7; p = 0.262). However, a statistically significant difference is present between predicted and actual values of curation for the first episode of retouch (t = 3.811, df = 224.83; p < 0.001). This indicates that for scrapers which have undergone very light retouch, the random forest model will slightly overpredict their curation ratio. Statistically significant differences are also present between values of predicted and actual curation on flakes which have undergone eight or more resharpening episodes (t = -3.731, df = 48.493; p < 0.001). This means that scrapers which have undergone multiple episodes of retouch, with more than 50% of their mass removed, will have underpredicted curated ratios (Figure 7).

Pred.RF %>%   
 mutate(  
 Real.Curated = Predicted.RF$Real.Curated,  
 Pred.Curated2 = Predicted.RF$Pred.Curated) %>%   
 select(Episode, Cat.GIUR, Real.Curated, Pred.Curated2) %>%   
 pivot\_longer(Real.Curated:Pred.Curated2,  
 names\_to = "Curation",  
 values\_to = "Values") %>%   
 mutate(Curation = factor(Curation,   
 levels = c("Real.Curated", "Pred.Curated2"),  
 labels = c("Real", "Predicted"))) %>%   
   
 ggplot(aes(factor(Episode), Values, fill = Curation)) +  
 geom\_boxplot() +  
 theme\_light() +  
 ggsci::scale\_fill\_d3() +  
 xlab("Resharpening event") +  
 ylab("Curation Index") +  
 scale\_y\_continuous( breaks = seq(-10, 70,10), limits = c(-11, 70)) +  
 theme(  
 legend.position = "bottom",  
 axis.text = element\_text(color = "black")  
 )

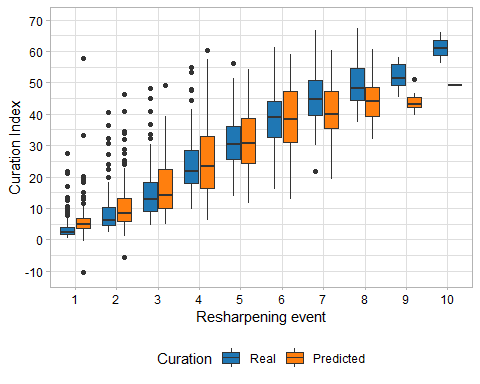


Table 4 presents the performance metrics values of the random forest model when leaving out 10% of flakes and all their resharpening episodes. Performance metrics present marginally lower values (Table 2). The Linear correlation (r2) decreases from 0.974 to 0.96, MAE increases from 3.297 to 3.989, RMSE increases from 5.917 to 7.478, and MAPE increases from 6.775 to 7.929.

When compared to other models using the same sample of flakes (Table 5), the random forest presents much better performance metrics. Neither of the two models presented linear correlation values (r2) above the 0.8 threshold when predicting original flake mass.

read.csv("Data/Results-Alternative-models.csv")

## Model r2 MAE RMSE MAPE  
## 1 Bustos-Perez & Baena, 2022 0.784 12.326 19.169 28.718  
## 2 EPA, Platform, Max. Thick. 0.636 15.159 23.102 38.157

### **The “Original Scraper Mass Calculator” v.1.0.0**

In order to increase the applicability of machine learning models, the first version (v.1.0.0) of the original scraper mass calculator (OSMC) has been developed (Figure 8). This app integrates the above described random forest model into a user-friendly interface, allowing one to quickly estimate original scraper mass using the same set of variables. This app is published as a free open source Shiny app, allowing for its unrestricted use by the archaeological community.

The current version of the OSMC allows the user to individually make estimations of original scraper mass or to process large amounts of data. Individual estimations can be done by manually entering data of each scraper at their corresponding spots and pressing the “Calculate original mass” button. Processing of files with multiple entries requires uploading a CSV file containing the data and with correct names for each of the column variables (an example downloadable CSV file has been made available at the app to ease this process). Column names of the CSV file must be: Rem.Weight, Mean.t, Max.thick (log10 transformation of this variable is done automatically by the app), and GIUR. Both options return an estimation of the original scraper mass and what percentage has been lost through retouch.

The app is available at:  
<https://guillermo-bustos-perez.shinyapps.io/Original-Scraper-Mass-Calculator/>

## **Discussion**

The present study was aimed at predicting original scraper mass and from this the percentage of mass lost due to retouch (curation of an artifact). A set of 134 experimental flakes were sequentially retouched, and after each episode four variables were used as predictors. The set of variables were: scraper mass, maximum thickness of the flake (log10 transformation), average height of retouch, and the GIUR [[22](#ref-kuhn_geometric_1990)]. Four models were prepared using these variables: a multiple linear regression, a supported vector machine regression, a random forest, and a GBM. Random forest had the best performance metrics, both when predicting original scraper mass (r2 = 0.974) and percentage of mass lost by retouch (r2 = 0.839). VIF values from the multiple linear regression indicate that possible collinearity between predictors is of no consequence (all features presented VIF values bellow 10). When 10% of flakes (n = 13) and all their resharpening episodes were used as test sets, the performance metrics only diminished marginally. It can be considered that this marginally lower performance is a result of a loss of information during training and not due to overfitting. The resulting model has been integrated into the first version of the “Original scraper mass calculator” in a user-friendly app, which allows random forest estimations from user data. Data can be batch processed using CSV files (columns should have same name as the ones from the model) or manually introduced.

Scrapers with direct retouch on a continuous edge are some of the most common lithic implements present from the first Oldowan stone tools [[1](#ref-barsky_early_2011)–[3](#ref-semaw_26-million-year-old_2003)] through to modern ethnographic examples [[4](#ref-casamiquela_temas_1978)–[7](#ref-sillitoe_living_2003)]. It is commonly considered that the amount of retouch that they receive is integral for understanding aspects of morphological variability and the organization of lithic technology by past societies [[8](#ref-dibble_middle_1995),[18](#ref-shott_costs_2018),[73](#ref-barton_formation_2014),[74](#ref-roth_production_1998-1)]. Thus, being able to correctly estimate the amount and percentage of mass lost by retouch is fundamental for understanding these aspects of past human groups.

Most studies aimed at predicting original flake mass use complete flakes and not in the framework of a sequential resharpening experimentation. When predicting original flake mass, most studies presented linear correlation values raging between 0.224 to 0.750 [[30](#ref-davis_quantifying_1998),[31](#ref-shott_flake_2000),[33](#ref-dogandzic_edge_2015),[38](#ref-clarkson_estimating_2011)]. Braun et al. [[37](#ref-braun_landscape-scale_2008)] obtained a linear correlation value of 0.865 for an archaeological sample of materials when predicting the log of flake mass. Dibble and Pelcin [[25](#ref-dibble_effect_1995)] obtained an of 0.815 under experimental controlled conditions. Bustos-Pérez and Baena Preysler [[35](#ref-bustos-perez_multiple_2022)] obtained an of 0.813 for free hand-held knapped flakes after predicting log transformed values of flake mass and transforming them back to the to the linear scale. However, most of these studies worked with complete flakes, making it not possible to estimate the curation ratio on a sequential resharpening experimentation.

Studies evaluating the estimated reduction percentage (ERP) do provide accuracy metrics of predictions for original volume and percentage of volume lost by retouch for an experimental assemblage of flakes [[21](#ref-eren_defining_2005),[23](#ref-morales_measuring_2015)]. It is important to consider important differences between both experimental set-ups and the present research. ERP flakes were retouched only once, with most of them having lost less than 20% of volume due to retouch (less than 15% in the case of 3D ERP). When recorded manually [[21](#ref-eren_defining_2005)], the ERP correlated very strongly with flake original volume ( = 0.823) and fairly strong with percentage of volume lost by retouch ( = 0.799 when outliers were removed). Incorporating a 3D protocol for applying the ERP [[23](#ref-morales_measuring_2015)] resulted in very strong correlation values for both estimates ( = 0.891 when estimating original volume, and = 0.812 when estimating percentage of mass lost by retouch).

The higher correlation values obtained in the present study for original mass and percentage of mass removed by retouch can be attributed to three factors. First, the present study incorporated the scraper mass and maximum thickness. Scraper mass is probably helping machine learning algorithms model the minimum mass a scraper should have. The log transformed maximum thickness has been shown to serve as good proxy for original flake mass [[34](#ref-bustos-perez_predicting_2021)]. These two variables seem to be compensating for angle and length of retouch (used in the ERP and 3D-ERP), which would account for less than three percent of the variance when estimating original scraper mass for the given sample. Second, the present study also incorporated average height of retouch which strongly serves as a proxy of the amount of mass removed (and is also included in the ERP). Third, the present study used more robust regression methods, which are able to better model the provided variables and their interactions. This is especially noticeable in the results with tree-based methods (random forest and GBM) performing significantly better than the multiple linear regression.

A series of ideal qualities have been proposed for reduction indexes [[75](#ref-hiscock_generalization_2010)]. The random forest model presented here complies with the qualities of inferential power, directionality, comprehensiveness, sensitivity, and scale-independence. The model presents an extremely large/strong inferential power (>0.8) when estimating either original scraper mass or percentage of mass lost by retouch. This has allowed the model to make inferences at the individual level for specimens. Directionality, comprehensiveness and sensitivity are also present in the proposed model, with values of curation being detected from the beginning of retouch, and increasing with each episode of retouch. However, it is observed that there will be underestimations for big flakes which have lost large amount of mass by retouch (>50%).

A common drawback for estimating original scraper mass at the individual artifact level has been the obtention of negative values of curation (which happens when the predicted mass of the original flake blank is lower than the actual mass of the retouched scraper). This error was present in only three cases of the 698 mass predictions made by the random forest model. Although this error can still happen, it seems to be infrequent in the random forest model.

The random forest model is also scale-independent, since it can be applied to blanks of different sizes and it estimates the amount of mass lost from retouch as a percentage of the original mass. However, it is also important to point out that the inferences at the individual level allow for the model to obtain predictions at the linear scale (amount of mass in grams lost by an individual scraper), which can also be considered when analyzing lithic assemblages.

The random forest model complies with some of the qualities of blank diversity. Results from the present experimentation indicate that the predictions from the model are not affected by the different transversal section of flakes. The flat flake problem was specifically evaluated, with no statistically significant differences between residual distribution of flat and non-flat flakes. A statistically significant difference was also not present between both types of flakes on more advanced episodes of retouch. However, it still remains to be tested how the proposed index will perform on endscrapers made on blades. It is considered that on this tool-type the average height of retouch and GIUR will soon reach their maximum and remain stable, despite successive episodes of retouch.

The proposed model has limited versatility, since it can be only applied to scrapers with hard hammer direct retouch among a continuous edge. Although this type of scraper is fairly predominant among lithic assemblages, other tool-types such as double scrapers, scrapers with inverse or bifacial retouch, or scrapers retouched with soft hammer cannot be analyzed with the present model.  
Several limitations of the present research framework have already been indicated. As previously indicated, results show that very lightly retouched scrapers will have their curation ratio slightly overestimated, while heavily retouched scrapers will have a curation ratio underestimated. This bias at the lower an upper range of the predictions has been previously observed for random forests when predicting original flake mass [[35](#ref-bustos-perez_multiple_2022)]. In addition, and as previously stated, the application of the current model is limited only to scrapers with direct retouch on a continuous edge (excluding endscrapers made on blades, double scrapers or bifacially retouched scrapers). The initial experimental setting also presented a series of limitations. Values of original mass of scrapers presented a skewed distribution, with a long tail for flakes with a mass above 100 g. In additional, data for flakes above 100 g is sparse, with few examples available. These two factors might be impacting accuracy of predictions for scrapers with higher values of initial mass, since the random forest and GBM have fewer examples from which to learn predictions.

Several additions can be done to expand and improve the current model in the future. As previously mentioned, the current model is limited to scrapers with direct retouch on one continuous edge. Further improvements using the current approach might benefit by adding “number of retouched edges” as a predictive variable or generating an independent new model specific for scrapers with multiple edges.

The current research tested four different machine learning algorithms (multiple linear regression, SVM regression, random forest and GBM). Random forest was the best performing algorithm although lightly retouched artifacts tend to have an overpredicted curation ratio, while extremely heavily retouched scrapers are usually underpredicted. Previous studies have shown that artificial neural networks (ANN) can improve upon random forests when predicting original flake mass [[35](#ref-bustos-perez_multiple_2022)]. Despite being computationally more expensive, further research might benefit from training ANN for predicting original scraper mass, helping to overcome the bias observed in lightly and extremely retouched artifacts.

The v.1.0.0 of the “Original Scraper Mass Calculator” does not provide indications of how reliable predictions are. Very lightly retouched artifacts tend to have an overestimated curation ratio, although predictions might be reliable. However, caution is recommended when it is inferred that a scraper has undergone multiple episodes of retouch since predictions from the current model will tend to underestimate the real amount of mass removed by retouch. Detecting cases of underestimation currently depends on the user knowledge of their lithic assemblage. An experienced lithic analyst will be able to detect cases of underestimation based on known features which determine flake mass (platform depth and size, flake thickness, etc.). Further versions of the “Original Scraper Mass Calculator” can include an indicator of the quality of predictions based on values of the GIUR index, or using an interaction of the GIUR and maximum thickness (allowing it to detect big/thick flakes which have undergone multiple episodes of resharpening).

Accurate inferences of mass lost by individual scrapers can be combined with multiple lithic analysis. Original scraper mass and the curation ratio can be combined with geometric morphometrics to determine how scraper shape changes through reduction (allometry) at a single (or multiple) sites, or at a diachronic level [[[9](#ref-dibble_interpretation_1987)]; [[16](#ref-andrefsky_construction_2008)]; [[40](#ref-dogandzic_results_2020)]; brumm\_scraper\_2011; [[76](#ref-hiscock_early_2003)]; [[77](#ref-iovita_comparing_2010)]; [[78](#ref-knell_allometry_2022)]]. Distance to raw material sources seems to have played an underlying factor of resharpening intensity among scrapers, with scrapers coming from longer distances having been resharpened more intensively [[14](#ref-clarkson_holocene_2002),[15](#ref-glauberman_late_2020),[79](#ref-byrne_dynamics_1980)–[81](#ref-kuhn_planning_1992)]. However, concerns about accuracy of indexes employed for these estimations have been raised [[23](#ref-morales_measuring_2015),[42](#ref-eren_kuhns_2009)]. Application of the proposed method can help better model the influence of raw material distance on resharpening intensity. Association between resharpening intensity, tool shape and use remain partially unexplored. Some studies have pointed to a lack of relationship between specific tool form and function [[82](#ref-borel_stone_2017)], while other studies point at potential higher functional versatility for tools coming from longer distances and higher curation ratios [[83](#ref-agam_interpreting_2020)]. Finally, the application of the proposed method can help to better model two additional aspect of lithic technological organization: changes in preferred blanks for resharpening and changes in tool gear transport. Changes in preferred initial blanks for resharpening have been observed between different Paleolithic periods but also within a same Paleolithic period. A clear example is observed among the different lithic technocomplexes of the western European Middle Paleolithic. Different lithic technocomplexes, characterized by different predominant knapping methods, show differences in the selection of initial blanks which underwent more intense resharpening [[20](#ref-bustos-perez_exploring_2019),[84](#ref-roth_production_1998),[85](#ref-faivre_late_2017)]. Additionally, strategies of tool transport also experienced changes within technocomplexes, alternating between the transport and resharpening of tools and transport of cores for immediate flake production [[17](#ref-kuhn_unpacking_1991),[86](#ref-porraz_middle_2009)].

## **Conclusions**

Predicting original scraper mass and amount of mass lost by retouch has long been a major goal in lithic analysis. An experimental sample of 134 flakes was sequentially retouched, and a new combination of variables (scraper mass, average height of retouch, maximum thickness, and value of the GIUR index) was recorded for each resharpening episode. This new set of variables in combination with more robust regression algorithms has resulted in the most accurate model to date. This higher accuracy allows for an estimation of retouch intensity at the individual scraper level. This model has been integrated into a user-friendly app in order to allow for its widespread application among the archaeological community.

## **Acknowledgments**

Guillermo Bustos-Pérez is postdoctoral researcher at the Department of Human Origins (Max Planck Institute for Evolutionary Anthropology). Several people have contributed to the improvement of the draft and testing the Original Scraper Mass Calculator. Shannon McPherron made suggestions to strength the analysis, test for overfitting and reviewed the early version of the manuscript. Etienne Nouraud helped test the first online deployment of the Original Scraper Mass Calculator and made suggestions for its improvement.

## **References**

1. Barsky D, Chapon-Sao C, Bahain J-J, Beyene Y, Cauche D, Celiberti V, et al. The Early Oldowan Stone-Tool Assemblage from Fejej FJ-1A, Ethiopia. Journal of African Archaeology. 2011;9: 207–224. doi:[10.3213/2191-5784-10196](https://doi.org/10.3213/2191-5784-10196)

2. Sahnouni M. Further research at the Oldowan site of Ain Hanech, North-eastern Algeria. Journal of Human Evolution. 2002;43: 925–937. doi:[10.1006/jhev.2002.0608](https://doi.org/10.1006/jhev.2002.0608)

3. Semaw S, Rogers MJ, Quade J, Renne PR, Butler RF, Dominguez-Rodrigo M, et al. 2.6-Million-year-old stone tools and associated bones from OGS-6 and OGS-7, Gona, Afar, Ethiopia. Journal of Human Evolution. 2003;45: 169–177. doi:[10.1016/S0047-2484(03)00093-9](https://doi.org/10.1016/S0047-2484(03)00093-9)

4. Casamiquela RM. Temas patagónicos de interes arqueológico. La talla del vidrio. Relaciones de la Sociedad Argentina de Antropología. 1978;12: 213–223.

5. Gallagher JP. Contemporary Stone Tools in Ethiopia: Implications for Archaeology. Journal of Field Archaeology. 1977;4: 407–414. doi:[10.1179/009346977791490131](https://doi.org/10.1179/009346977791490131)

6. Shott MJ, Weedman KJ. Measuring reduction in stone tools: An ethnoarchaeological study of Gamo hidescrapers from Ethiopia. Journal of Archaeological Science. 2007;34: 1016–1035. doi:[10.1016/j.jas.2006.09.009](https://doi.org/10.1016/j.jas.2006.09.009)

7. Sillitoe P, Hardy K. Living Lithics: Ethnoarchaeology in Highland Papua New Guinea. Antiquity. 2003;77: 555–566. doi:[10.1017/S0003598X00092619](https://doi.org/10.1017/S0003598X00092619)

8. Dibble HL. Middle Paleolithic Scraper Reduction: Background, Clarification, and Review of the Evidence to Date. Journal of Archaeological Method and Theory. 1995;2: 300–368.

9. Dibble HL. The interpretation of Middle Paleolithic scraper morphology. American Antiquity. 1987;52: 109–117.

10. Binford LR. Organization and Formation Processes: Looking at Curated Technologies. Journal of Anthropological Research. 1979;35: 255–273.

11. Binford LR. Interassemblage variability - the Mousterian and the ’functional ’ argument. In: Renfrew C, editor. The Explanation Of Culture Change Models in Prehistory. Gloucester: Duckworth; 1973. pp. 227–254.

12. Shott MJ. An Exegesis of the Curation Concept. Journal of Anthropological Research. 1996;52: 259–280.

13. Shott MJ. On Tool-Class Use Lives and The Formation of Archaeological Assemblages. American Antiquity. 1989;54: 9–30. doi:[10.2307/281329](https://doi.org/10.2307/281329)

14. Clarkson C. Holocene scraper reduction, technological organization and landuse at Ingaladdi Rockshelter, Northern Australia. Archaeology in Oceania. 2002;37: 79–86. doi:[10.1002/j.1834-4453.2002.tb00508.x](https://doi.org/10.1002/j.1834-4453.2002.tb00508.x)

15. Glauberman P, Gasparyan B, Wilkinson K, Frahm E, Nahapetyan S, Arakelyan D, et al. Late Middle Paleolithic Technological Organization and Behavior at the Open-Air Site of Barozh 12 (Armenia). Journal of Paleolithic Archaeology. 2020;3: 1095–1148. doi:[10.1007/s41982-020-00071-4](https://doi.org/10.1007/s41982-020-00071-4)

16. Hiscock P, Clarkson C. The construction of morphological diversity: A study of Mousterian implement retouching at Combe Grenal. In: Andrefsky W, editor. Lithic Technology: Measures of Production, Use and Curation. Cambridge: Cambridge University Press; 2008. pp. 106–135.

17. Kuhn SL. “Unpacking” Reduction: Lithic Raw Material Economy in the Mousterian of West-Central Italy. Journal of Anthropological Archaeology. 1991;10: 76–106.

18. Shott MJ. The Costs and Benefits of Technological Organization: Hunter-Gatherer Lithic Industries and Beyond. In: Robinson E, Sellet F, editors. Lithic Technological Organization and Paleoenvironmental Change: Global and Diachronic Perspectives. Cham: Springer International Publishing; 2018. pp. 321–333. doi:[10.1007/978-3-319-64407-3\_15](https://doi.org/10.1007/978-3-319-64407-3_15)

19. Shott MJ, Seeman MF. Use and Multifactorial Reconciliation of Uniface Reduction Measures: A Pilot Study at the Nobles Pond Paleoindian Site. American Antiquity. 2017;82: 723–741. doi:[10.1017/aaq.2017.40](https://doi.org/10.1017/aaq.2017.40)

20. Bustos-Pérez G, Baena J. Exploring volume lost in retouched artifacts using height of retouch and length of retouched edge. Journal of Archaeological Science: Reports. 2019;27: 101922. doi:[10.1016/j.jasrep.2019.101922](https://doi.org/10.1016/j.jasrep.2019.101922)

21. Eren MI, Domínguez-Rodrigo M, Kuhn SL, Adler DS, Le I, Bar-Yosef O. Defining and measuring reduction in unifacial stone tools. Journal of Archaeological Science. 2005;32: 1190–1201. doi:[10.1016/j.jas.2005.03.003](https://doi.org/10.1016/j.jas.2005.03.003)

22. Kuhn SL. A Geometric Index of Reduction for Unifacial Stone Tools. Journal of Archaeological Science. 1990;17: 583–593.

23. Morales JI, Lorenzo C, Vergès JM. Measuring Retouch Intensity in Lithic Tools: A New Proposal Using 3D Scan Data. Journal of Archaeological Method and Theory. 2015;22: 543–558. doi:[10.1007/s10816-013-9189-0](https://doi.org/10.1007/s10816-013-9189-0)

24. Dibble HL. Platform variability and flake morphology: A comparison of experimental and archaeological data and implications for interpreting prehistoric lithic technological strategies. Lithic Technology. 1997;22: 150–170.

25. Dibble HL, Pelcin A. The Effect of Hammer Mass and Velocity on Flake Mass. Journal of Archaeological Science. 1995;22: 429–439. doi:[10.1006/jasc.1995.0042](https://doi.org/10.1006/jasc.1995.0042)

26. Dibble HL, Whittaker JC. New Experimental Evidence on the Relation Between Percussion Flaking and Flake Variation. Journal of Archaeological Science. 1981;8: 283–296.

27. Pelcin AW. The Formation of Flakes: The Role of Platform Thickness and Exterior Platform Angle in the Production of Flake Initiations and Terminations. Journal of Archaeological Science. 1997;24: 1107–1113. doi:[10.1006/jasc.1996.0190](https://doi.org/10.1006/jasc.1996.0190)

28. Pelcin AW. Controlled experiments in the production of flake attributes. {PhD}, University of Pennsylvania. 1996.

29. Li L, Lin SC, McPherron SP, Abdolahzadeh A, Chan A, Dogandžić T, et al. A Synthesis of the Dibble et al. Controlled Experiments into the Mechanics of Lithic Production. Journal of Archaeological Method and Theory. 2023;30: 1284–1325. doi:[10.1007/s10816-022-09586-2](https://doi.org/10.1007/s10816-022-09586-2)

30. Davis ZJ, Shea JJ. Quantifying Lithic Curation: An Experimental Test of Dibble and Pelcin’s Original Flake-Tool Mass Predictor. Journal of Archaeological Science. 1998;25: 603–610. doi:[10.1006/jasc.1997.0255](https://doi.org/10.1006/jasc.1997.0255)

31. Shott MJ, Bradbury AP, Carr PJ, Odell GH. Flake Size from Platform Attributes: Predictive and Empirical Approaches. Journal of Archaeological Science. 2000;27: 877–894. doi:[10.1006/jasc.1999.0499](https://doi.org/10.1006/jasc.1999.0499)

32. Dibble HL. Comment on “Quantifying Lithic Curation: An Experimental Test of Dibble and Pelcin’s Original Flake-Tool Mass Predictor,” by Zachary J. Davis and John J. Shea. Journal of Archaeological Science. 1998;25: 611–613. doi:[10.1006/jasc.1997.0254](https://doi.org/10.1006/jasc.1997.0254)

33. Dogandžić T, Braun DR, McPherron SP. Edge Length and Surface Area of a Blank: Experimental Assessment of Measures, Size Predictions and Utility. PLoS ONE. 2015;10: e0133984. doi:[10.1371/journal.pone.0133984](https://doi.org/10.1371/journal.pone.0133984)

34. Bustos-Pérez G, Baena J. Predicting Flake Mass: A View from Machine Learning. Lithic Technology. 2021;46: 130–142. doi:[10.1080/01977261.2021.1881267](https://doi.org/10.1080/01977261.2021.1881267)

35. Bustos-Pérez G, Baena Preysler J. Multiple approaches to predicting flake mass. Journal of Archaeological Science: Reports. 2022;46: 103698. doi:[10.1016/j.jasrep.2022.103698](https://doi.org/10.1016/j.jasrep.2022.103698)

36. Muller A, Clarkson C. A new method for accurately and precisely measuring flake platform area. Journal of Archaeological Science: Reports. 2016;8: 178–186. doi:[10.1016/j.jasrep.2016.06.015](https://doi.org/10.1016/j.jasrep.2016.06.015)

37. Braun DR, Rogers MJ, Harris JWK, Walker SJ. Landscape-scale variation in hominin tool use: Evidence from the Developed Oldowan. Journal of Human Evolution. 2008;55: 1053–1063. doi:[10.1016/j.jhevol.2008.05.020](https://doi.org/10.1016/j.jhevol.2008.05.020)

38. Clarkson C, Hiscock P. Estimating original flake mass from 3D scans of platform area. Journal of Archaeological Science. 2011;38: 1062–1068. doi:[10.1016/j.jas.2010.12.001](https://doi.org/10.1016/j.jas.2010.12.001)

39. Maloney TR. Experimental and archaeological testing with 3D laser scanning reveals the limits of I/TMC as a reduction index for global scraper and point studies. Journal of Archaeological Science: Reports. 2020;29: 102068. doi:[10.1016/j.jasrep.2019.102068](https://doi.org/10.1016/j.jasrep.2019.102068)

40. Dogandžić T, Abdolazadeh A, Leader G, Li L, McPherron SP, Tennie C, et al. The results of lithic experiments performed on glass cores are applicable to other raw materials. Archaeological and Anthropological Sciences. 2020;12: 44. doi:[10.1007/s12520-019-00963-9](https://doi.org/10.1007/s12520-019-00963-9)

41. Morales JI. Distribution patterns of stone-tool reduction: Establishing frames of reference to approximate occupational features and formation processes in Paleolithic societies. Journal of Anthropological Archaeology. 2016;41: 231–245. doi:[10.1016/j.jaa.2016.01.004](https://doi.org/10.1016/j.jaa.2016.01.004)

42. Eren MI, Sampson CG. Kuhn’s Geometric Index of Unifacial Stone Tool Reduction (GIUR): Does it measure missing flake mass? Journal of Archaeological Science. 2009;36: 1243–1247. doi:[10.1016/j.jas.2009.01.011](https://doi.org/10.1016/j.jas.2009.01.011)

43. Hiscock P, Clarkson C. The reality of reduction experiments and the GIUR: Reply to Eren and Sampson. Journal of Archaeological Science. 2009;36: 1576–1581. doi:[10.1016/j.jas.2009.03.019](https://doi.org/10.1016/j.jas.2009.03.019)

44. Hiscock P, Clarkson C. Experimental evaluation of Kuhn’s geometric index of reduction and the flat-flake problem. Journal of Archaeological Science. 2005;32: 1015–1022. doi:[10.1016/j.jas.2005.02.002](https://doi.org/10.1016/j.jas.2005.02.002)

45. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning with Applications in R. Second Edition. New York: springer; 2013. Available: <https://doi.org/10.1007/978-1-0716-1418-1>

46. Awad M, Khanna R. Support Vector Regression. In: Awad M, Khanna R, editors. Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers. Berkeley, CA: Apress; 2015. pp. 67–80. doi:[10.1007/978-1-4302-5990-9\_4](https://doi.org/10.1007/978-1-4302-5990-9_4)

47. Cortes C, Vapnik V. Support-vector networks. Machine learning. 1995;20: 273–297.

48. Smola AJ, Schölkopf B. A tutorial on support vector regression. Statistics and Computing. 2004;14: 199–222. doi:[10.1023/B:STCO.0000035301.49549.88](https://doi.org/10.1023/B:STCO.0000035301.49549.88)

49. Breiman L. Random Forests. Machine Learning. 2001;45: 5–32. doi:[10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324)

50. Lantz B. Machine learning with R: Expert techniques for predictive modeling. Birmingham: Packt publishing ltd; 2019.

51. Friedman JH. Greedy function approximation: A gradient boosting machine. Annals of statistics. 2001;29: 1189–1232. Available: <https://www.jstor.org/stable/2699986>

52. Friedman JH. Stochastic gradient boosting. Computational Statistics & Data Analysis. 2002;38: 367–378. doi:[10.1016/S0167-9473(01)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2)

53. Anscombe FJ. Graphs in Statistical Analysis. The American Statistician. 1973;27: 17. doi:[10.2307/2682899](https://doi.org/10.2307/2682899)

54. Fox J, Weisberg S. An R companion to applied regression. Third. Thousand Oaks: Sage publications; 2018.

55. Marquardt DW. Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation. Technometrics. 1970;12: 591–612.

56. O’brien RM. A Caution Regarding Rules of Thumb for Variance Inflation Factors. Quality & Quantity. 2007;41: 673–690. doi:[10.1007/s11135-006-9018-6](https://doi.org/10.1007/s11135-006-9018-6)

57. Team Rs. RStudio: Integrated Development for R. Boston, MA: RStudio, Inc.; 2019. Available: <http://www.rstudio.com/>

58. Team RC. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2019. Available: <https://www.R-project.org/>

59. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. Journal of Open Source Software. 2019;4: 1686. doi:[10.21105/joss.01686](https://doi.org/10.21105/joss.01686)

60. Venables WN, Ripley BD. Modern applied statistics with S. Fourth Edition. New York: Springer; 2002. Available: <https://www.stats.ox.ac.uk/pub/MASS4/>

61. Karatzoglou A, Smola A, Hornik K, Zeileis A. Kernlab - An S4 Package for Kernel Methods in R. Journal of Statistical Software. 2004;11: 1–20. doi:[10.18637/jss.v011.i09](https://doi.org/10.18637/jss.v011.i09)

62. Dimitriadou E, Hornik K, Leisch F, Meyer D, Weingessel A. Misc functions of the Department of Statistics (e1071), TU Wien. R package. 2008;1: 5–24. Available: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=3ceff42db79be884fc7b18b22fcc6ba0d8ef0faf>

63. Wright MN, Ziegler A. Ranger: A fast implementation of random forests for high dimensional data in C++ and R. arXiv preprint arXiv:150804409. 2015.

64. Ridgeway G. Generalized Boosted Models: A guide to the gbm package. R package vignette. 2007; 2007. Available: <http://CRAN.R-project.org/package=gbm>

65. Kuhn M. Building Predictive Models in R using the caret Package. Journal of Statistical Software. 2008;28: 1–26. doi:[10.18637/jss.v028.i05](https://doi.org/10.18637/jss.v028.i05)

66. Chang W, Cheng J, Allaire J, Xie Y, McPherson J. Package “shiny.” 2015 [cited 1 Oct 2024]. Available: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=b7600ac8de6a97df1bd26dcbfc43ef9fc3c8be61>

67. Kasprzak P, Mitchell L, Kravchuk O, Timmins A. Six Years of Shiny in Research-Collaborative Development of Web Tools in R. R Journal. 2020;12: 20–42. doi:[10.32614/rj-2021-004](https://doi.org/10.32614/rj-2021-004)

68. Wickham H. Mastering shiny. " O’Reilly Media, Inc."; 2021.

69. McNamara A. Dynamic Documents with R and knitr. Journal of Statistical Software. 2014;56: 1–4. Available: <https://www.jstatsoft.org/index.php/jss/article/view/v056b02/731>

70. Xie Y. Bookdown: Authoring books and technical documents with R markdown. Taylor & Francis Ltd; 2016.

71. Xie Y. Knitr: A Comprehensive Tool for Reproducible Research in R. Implementing Reproducible Research. Chapman; Hall/CRC; 2014.

72. Marwick B, Boettiger C, Mullen L. Packaging Data Analytical Work Reproducibly Using R (and Friends). The American Statistician. 2018;72: 80–88. doi:[10.1080/00031305.2017.1375986](https://doi.org/10.1080/00031305.2017.1375986)

73. Barton CM, Riel-Salvatore J. The formation of lithic assemblages. Journal of Archaeological Science. 2014;46: 334–352. doi:[10.1016/j.jas.2014.03.031](https://doi.org/10.1016/j.jas.2014.03.031)

74. Roth BJ, Dibble HL. Production and Transport of Blanks and Tools at the French Middle Paleolithic Site of Combe-Capelle Bas. American Antiquity. 1998;63: 47–62.

75. Hiscock P, Tabrett A. Generalization, inference and the quantification of lithic reduction. World Archaeology. 2010;42: 545–561. doi:[10.1080/00438243.2010.517669](https://doi.org/10.1080/00438243.2010.517669)

76. Hiscock P, Attenbrow V. Early Australian implement variation: A reduction model. Journal of Archaeological Science. 2003;30: 239–249. doi:[10.1006/jasc.2002.0830](https://doi.org/10.1006/jasc.2002.0830)

77. Ioviţă R. Comparing Stone Tool Resharpening Trajectories with the Aid of Elliptical Fourier Analysis. In: Lycett S, Chauhan P, editors. New Perspectives on Old Stones: Analytical Approaches to Paleolithic Technologies. New York, NY: Springer; 2010. pp. 235–253. doi:[10.1007/978-1-4419-6861-6\_10](https://doi.org/10.1007/978-1-4419-6861-6_10)

78. Knell EJ. Allometry of unifacial flake tools from Mojave Desert terminal pleistocene/early holocene sites: Implications for landscape knowledge, tool design, and land use. Journal of Archaeological Science: Reports. 2022;41: 103314. doi:[10.1016/j.jasrep.2021.103314](https://doi.org/10.1016/j.jasrep.2021.103314)

79. Byrne D. Dynamics of Dispersion: The Place of Silcrete in Archaeological Assemblages from the Lower Murchison, Western Australia. Archaeology & Physical Anthropology in Oceania. 1980;15: 110–119. Available: <https://www.jstor.org/stable/40386377>

80. Clarkson C. An Index of Invasiveness for the Measurement of Unifacial and Bifacial Retouch: A Theoretical, Experimental and Archaeological Verification. Journal of Archaeological Science. 2002;29: 65–75. doi:[10.1006/jasc.2001.0702](https://doi.org/10.1006/jasc.2001.0702)

81. Kuhn SL. On Planning and Curated Technologies in the Middle Paleolithic. Journal of Anthropological Research. 1992;48: 185–214.

82. Borel A, Cornette R, Baylac M. Stone Tool Forms and Functions: A Morphometric Analysis of Modern Humans’ Stone Tools From Song Terus Cave (Java, Indonesia): Stone tool forms and functions. Archaeometry. 2017;59: 455–471. doi:[10.1111/arcm.12264](https://doi.org/10.1111/arcm.12264)

83. Agam A, Zupancich A. Interpreting the Quina and demi-Quina scrapers from Acheulo-Yabrudian Qesem Cave, Israel: Results of raw materials and functional analyses. Journal of Human Evolution. 2020;144: 102798. doi:[10.1016/j.jhevol.2020.102798](https://doi.org/10.1016/j.jhevol.2020.102798)

84. Roth BJ, Dibble HL. Production and Transport of Blanks and Tools at the French Middle Paleolithic Site of Combe-Capelle Bas. American Antiquity. 1998;63: 47–62.

85. Faivre G-P, Gravina B, Bourguignon L, Discamps E, Turq A. Late Middle Palaeolithic lithic technocomplexes (MIS 5-3) in the northeastern Aquitaine Basin: Advances and challenges. Quaternary International. 2017;433: 116–131. doi:[10.1016/j.quaint.2016.02.060](https://doi.org/10.1016/j.quaint.2016.02.060)

86. Porraz G. Middle Paleolithic mobile toolkits in short-term human occupations: Two case studies. Eurasian Prehistory. 2009;6: 33–55.