

Peer Community Journal

Section: Archaeology

Research article

Published
2025-08-21

Cite as
Marcelo Cardillo and Eugenia Carranza (2025) *Applying statistical and causal modeling to interpret thermal alteration in observational lithic data*, Peer Community Journal, 5: e88.

Correspondence
marcelo.cardillo@gmail.com

Peer-review
Peer reviewed and recommended by PCI Archaeology,
<https://doi.org/10.24072/pci.archaeo.100613>



This article is licensed under the Creative Commons Attribution 4.0 License.

Applying statistical and causal modeling to interpret thermal alteration in observational lithic data

Marcelo Cardillo ,^{#,1} and Eugenia Carranza ,^{#,1}

Volume 5 (2025), article e88

<https://doi.org/10.24072/pcjournal.609>

Abstract

Wildfire events are increasingly recognized as agents of taphonomic alteration in archaeological contexts. This study applies causal and statistical modeling to assess the impact of thermal alteration on lithic assemblages in the San Matías Gulf region of Northern Patagonia, Argentina. Utilizing data derived from naturalistic sampling after recent wildfires, we developed a Directed Acyclic Graph (DAG) to identify key mediators and potential confounders affecting lithic fragmentation. A minimal adjustment set was determined to isolate the indirect effect of fire on mechanical alterations, and Random Forest classifiers were trained to evaluate the predictive power of variables such as pebble shape, fissures, and rock type. Comparison between models based on original and synthetically balanced datasets revealed significant improvements in classification accuracy, particularly for unfractured samples. The results emphasize the central role of shape in fracture likelihood, and demonstrate the utility of synthetic resampling and causal inference for interpreting complex taphonomic processes in observational archaeological data. Experimental findings corroborate the observed patterns, supporting the potential generalizability of the proposed causal model across fire-affected contexts.

¹Instituto Multidisciplinario de Historia y Ciencias Humanas, Consejo Nacional de Investigaciones Científicas y Técnicas (IMHICHU, CONICET) / Facultad de Filosofía y Letras, Universidad de Buenos Aires (FFyL, UBA) - Buenos Aires, Argentina, [#]Equal contribution

Peer Community Journal is a member of the
Centre Mersenne for Open Scientific Publishing
<http://www.centre-mersenne.org/>

e-ISSN 2804-3871



Introduction

In recent decades, causal models have gained increasing relevance in scientific research (Pearl & Mackenzie, 2018; Bulbulia, 2024; Snodgrass et al., 2024; among others). However, their adoption remains limited in the social sciences, where causal inference from observational data is commonly avoided or replaced by correlational analyses. While predictive approaches have proven useful for anticipating patterns in data, they often do so at the expense of theoretical interpretability and external validity. As Shmueli (2010) argues, explanation and prediction are distinct analytical tasks, each with its own goals and assumptions. Prediction focuses on empirical accuracy, whereas explanation requires identifying theoretically grounded causal relationships. In observational and ethnographic contexts, such as those discussed by Snodgrass et al. (2024), causal inference is key to understanding how certain practices or social processes generate observable effects (see also McElreath & Koster, 2024). Similarly, Deffner et al. (2022) contend that without a formal framework for explicitly representing causal relationships, even the mere description of phenomena across populations can lead to misinterpretations. Therefore, this study adopts an approach that complements traditional statistical techniques with causal modeling to analyze the mechanisms underlying the formation of surface lithic assemblages. One of the most powerful tools in causal modeling is the use of directed acyclic graphs (DAGs). DAGs provide a visual and mathematical framework for formalizing the causal assumptions underlying statistical analyses, thereby enhancing the transparency and reproducibility of research (McElreath & Koster, 2024). By explicitly illustrating the causal structure, DAGs help researchers identify potential sources of bias, such as confounding variables, and guide the selection of appropriate covariates for adjustment. Furthermore, they facilitate the application of causal inference techniques, such as the backdoor criterion or instrumental variables, to estimate causal effects more accurately. Their utility extends beyond theoretical modeling, as they also serve as a communication tool, enabling researchers to clearly convey complex causal relationships to diverse audiences.

In this work, we first explore the theoretical foundations and practical applications of causal inference models, highlighting their relevance to naturalistic research in archaeology. Particular emphasis is placed on the integration of these models into common research. Following this, we present a model specifically designed for experimental analyses, which is informed and validated using data derived from a naturalistic sample. Subsequently, a random forest classification tree is employed to quantify the effect of fire (not directly observed) on the probability of rock fracture throughout mediation (observed) variables, offering a nuanced understanding of the processes at play. The results are used as a foundation, alongside the causal model, to determine the minimum number of variables that need to be adjusted in order to estimate the effect of fire on pebbles in naturalistic contexts. Finally, we discuss the role of the occurrence of wildfires along the northern coast of Argentinian Patagonia in the formation of lithic landscapes, providing insights into the long-term impacts of such events on the archaeological record.

Observational data

The San Matías Gulf (Río Negro Province, Argentina) encompasses an extensive coastline stretching approximately 380 linear kilometers (Figure 1). The region is characterized by a temperate semi-arid climate, with an average annual temperature of 12°C and precipitation ranging between 100- and 350-mm. Prevailing winds originate from the western quadrant (NW, W, and SW), with maximum recorded speeds of approximately 90–100 km/h. The strongest winds typically occur between September and January. Vegetation is predominantly shrubby and corresponds to the Monte phytogeographic province (Oyarzabal et al., 2018). These environmental conditions — marked by low rainfall during summer months, elevated temperatures, and low relative humidity — create a high susceptibility to wildfires, particularly during specific months that constitute the regional “fire season” (Zacconi & Toppazzini, 2018; Sottile et al., 2018).

Within this context, a wildfire event occurred during the summers of 2019 and 2020, affecting a broad area within the study region (Figure 1). Based on local reports and satellite data (FIRMS),

the wildfire in the study area occurred naturally in a low-density shrubland without active firefighting measures such as water application or direct extinguishing in the sampled locations. Consequently, the rocks were not subjected to rapid cooling by water, which is known to cause distinct thermal shock fracturing patterns. This absence of artificial fire suppression suggests that the observed fracture patterns result primarily from natural thermal processes. Following the event, field surveys were conducted in the burned areas with the aim of making actualistic observations and conducting a preliminary assessment of the fire's potential impact on the archaeological record and surrounding landscape. Two targeted sampling campaigns were carried out, enabling an evaluation of the diversity of affected contexts, the density and distribution of archaeological materials, the variability of raw materials, and the establishment of analytical units (Cardillo et al., 2022). These efforts allowed for a preliminary assessment of the extent of fire-induced alterations to the lithic record and the development of criteria to better understand how fire processes influence the formation of surface lithic assemblages and the archaeological record as a whole.

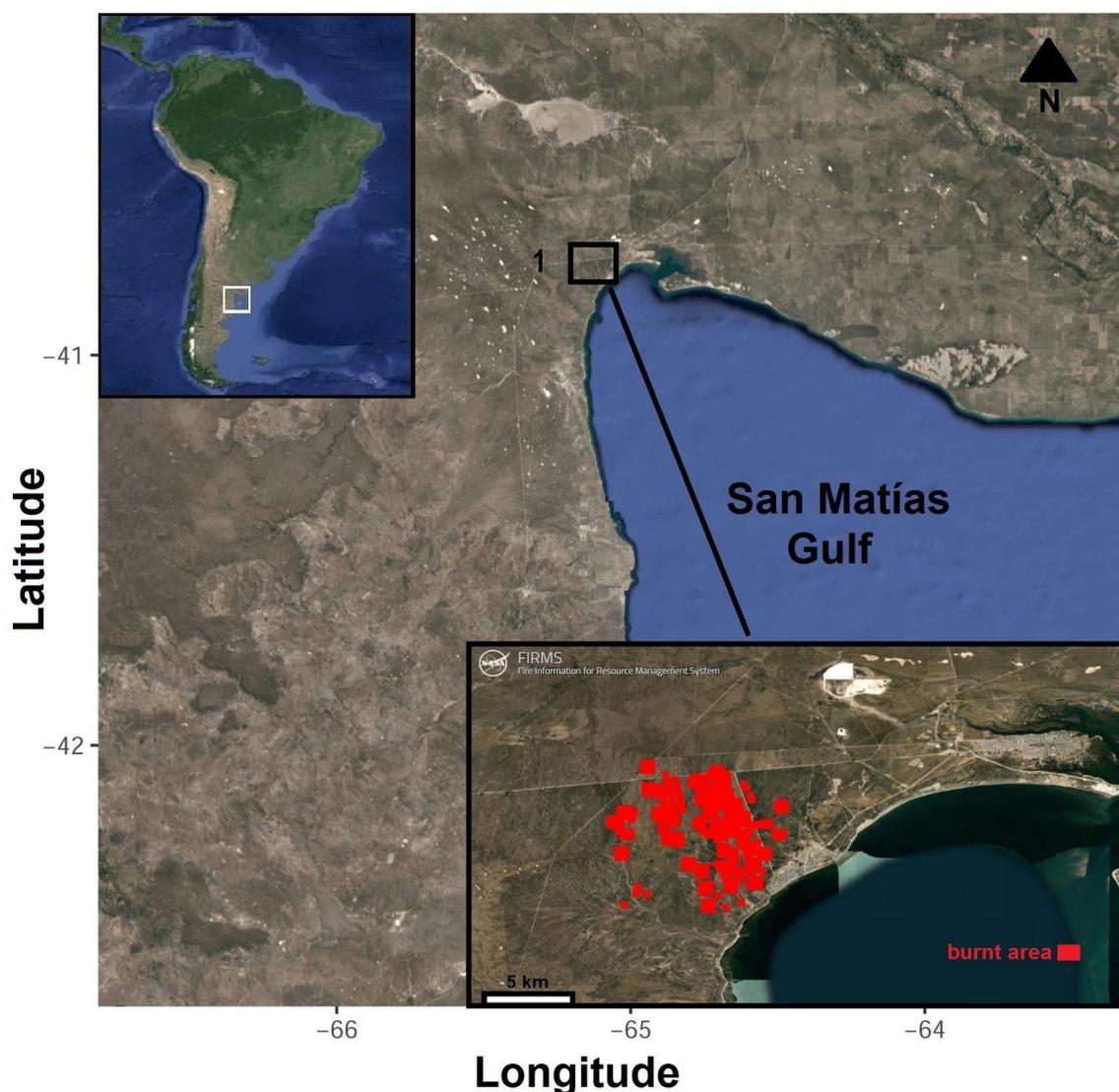


Figure 1 - Study area; 1: Satellite image of burnt area in 2021.

One of the characteristics of naturalistic observations is that there is no control over the variables affecting the phenomena of interest or, at best, such control is limited. This contrasts with controlled experiments, where exposure and treatment variables, as well as covariates, can be

regulated through randomization processes. In a randomized design, each individual has the same probability of belonging to the treatment or control group, which helps limit confounding factors. In contrast, in naturalistic samples or experiments, there is no control over the individuals included in the sample, and some of the factors under study may not be directly observable, increasing the risk of confounding factors biasing the results.

The archaeological record of the study area is characterized by surface distributions of lithic materials, associated with hunter-gatherer groups who occupied the coastal fringe and exploited marine resources over a temporal span of approximately 6,000 years (Favier Dubois, 2019). These distributions, exhibiting variable density, are linked to the occurrence of secondary pebble sources of variable knapping quality, widely distributed across the landscape as a result of Pleistocene alluvial processes (Alberti, 2016; Cardillo, 2013; Cardillo & Alberti, 2013). Early-stage reduction sequences are commonly observed within these assemblages, along with cores made on nodules of high-quality acidic and basic volcanic rocks that are readily available in the area. In contexts where evidence of fire action was identified, various degrees of thermal alteration are visible, along with the accumulation of post-depositional processes that alter archaeological materials over time (Alberti & Carranza, 2014; Alberti & Cardillo, 2018; Borella et al., 2020). Of particular interest is how these processes can generate mixed assemblages that complicate the identification of intentional human modifications, including lithic knapping or the presence of combustion features. In this context, causal modeling offers a promising approach to identifying the key variables involved in the formation of such palimpsests and to assessing the specific effects of fire on natural lithic distributions. Given the complex interplay of natural and anthropogenic factors shaping these surface scatters, disentangling the specific impact of fire remains a significant challenge. Traditional descriptive approaches are often insufficient to address this complexity, particularly in contexts where diagnostic features are lacking or obscured. Therefore, a causal modeling framework (particularly through the use of DAGs) is employed here to formalize assumptions about the relationships among variables, to identify potential sources of bias, and to guide the analytical strategy needed to isolate the effects of fire within these palimpsests.

Material and methods

Modelization

Following Grace et al. (2025), it is important to distinguish between estimating causal effects and modeling causal mechanisms (see also Shmueli, 2010). While causal effects quantify the impact of an intervention or treatment (e.g. fire) on an outcome (e.g. thermal alteration), causal mechanisms describe the internal pathways through which these effects are realized. The present study focuses primarily on modeling causal mechanisms, specifically identifying the properties of lithic materials (such as shape, fissures, and rock type; see below) that mediate the effects of fire on mechanical fragmentation. By formalizing these pathways through a causal diagram (DAG), we aim not only to estimate whether fire increases lithic fragmentation, but also to explain how and why specific lithic attributes condition the outcomes of thermal alteration.

DAGs are fundamental tools in causal inference, providing a graphical framework to represent cause-effect relationships among variables (Pearl, 1995; Pearl & Mackenzie, 2018; see Figure 2). Within this framework, causal modeling allows us to move beyond the identification of mere associations (correlations) between taphonomic processes and archaeological record formation, enabling the explicit formulation of causal hypotheses essential for understanding the complex dynamics that shape surface lithic assemblages.

By encoding causal structures through nodes and directed edges, DAGs allow researchers to formally express qualitative assumptions about dependencies and interdependencies within a system (Pearl & Mackenzie, 2018; Bulbulia, 2024). In this case, DAGs are used to model the causal effect of fire on rocks in naturalistic environments. This visual representation facilitates a systematic assessment of causal relationships, particularly in complex scenarios where experimental data are difficult to obtain.

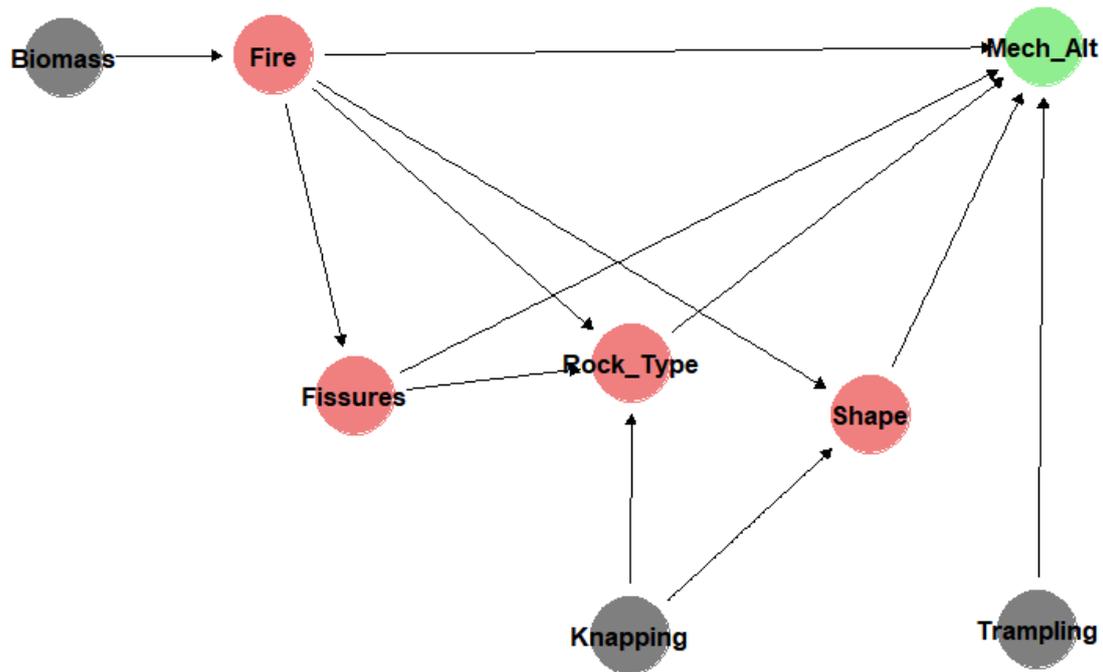


Figure 2 - Causal model for mechanical alterations of pebbles under the effect of fire. Red, direct and mediation effects (Fire -unmeasured-, Fissures, Rock_Type and Shape). Gray, latent unobserved variables (Biomass, Knapping and Trampling). Green, outcome variable (Mech_Alt).

In a DAG, an arrow connecting two variables represents a path, indicating the direction of influence between variables. There are three main types of paths: forks, chains, and colliders. Forks occur when a single variable acts as a common cause for two or more variables, with arrows pointing outward (Pearl, 1995; Pearl & Mackenzie, 2018). For example, if a specific rock property influences both its heat resistance and its susceptibility to human knapping (due to its quality), this property would be the common cause, creating a fork between heat resistance and knapping patterns. Forks introduce associations because changes in the common cause affect both outcomes. Chains describe a sequence in which one variable influences a second variable, which in turn affects a third. For instance, fire (the cause) may lead to thermal alterations in rock type (an intermediary variable), which then affects flake patterns (the outcome) by altering breakage patterns. Chains propagate causal effects along the sequence, making each node dependent on its predecessors. Colliders occur when two or more variables have arrows pointing into a common effect. For example, both fire and trampling by humans or animals might independently contribute to surface alterations in lithics. Unlike forks and chains, colliders do not inherently introduce associations; however, conditioning on a collider can induce a spurious association between its causes (e.g. fire and rock properties), even if they were otherwise unrelated.

DAGs also capture key relationships crucial for causal reasoning, such as direct causation, mediation, and confounding. Causation refers to a direct influence in which changes in one variable (the cause) result in changes in another (the effect) (Pearl, 1995; Huntington-Klein, 2021). This study specifically investigates the causal role of fire in shaping the lithic landscape. However, due to the constraints of naturalistic sampling, although fire is conceptualized as the exposure variable in the causal mode (Figure 2), its effects are inferred through the mediation of observable pebble traits (e.g., shape, fissures, rock type; see below).

Mediation occurs when an intermediary variable transmits the effect of one variable on another. In this context, pebble shape or size may mediate the impact of fire on alteration patterns. Confounding arises when an external variable influences both the cause and the effect, creating a

spurious association. For instance, fire could act as a confounder if it causes mechanical alterations to rock types that were also preferentially selected for knapping, introducing bias when studying surface scatters of lithic materials.

Unlike purely statistical models that describe relationships and dependencies, DAGs enable researchers to establish causal directions based on domain knowledge and theoretical assumptions. Each directed edge in a DAG represents a hypothesized causal link. Given that many studies rely on observational data -where variable control is often impossible- DAGs provide a structured approach for inferring causal effects from non-experimental data when specific graphical conditions, such as d-separation, are met. D-separation (short for “directional separation”) is a criterion used in DAGs to determine whether two variables are conditionally independent given a set of other variables. If two variables are d-separated by a given set, adjusting for these variables effectively removes any spurious associations introduced by confounders, isolating the direct causal relationship. For example, the differential effects of fire and human activity on rocks could be d-separated by controlling for rock type (e.g. using statistical or experimental methods, see below). D-separation is thus a tool for identifying the appropriate variables for adjustment, preventing confounding in causal analysis.

Adjustment Set

In causal inference, adjustments for confounders focus on blocking backdoor paths and leveraging frontdoor paths. Backdoor paths are non-causal pathways between an independent variable *X* (e.g. fire) and an outcome *Y* (e.g. mechanical or morphological alterations) that introduce bias due to confounding. These paths contain at least one arrow pointing into *X*, and controlling for confounders along these paths ensures an unbiased estimation of the causal effect. Front door paths describe causal pathways where *X* effects *Y* through an intermediary variable *Z* (a mediator). When direct confounding between *X* and *Y* cannot be controlled, front door adjustment can still enable causal inference if *Z* satisfies specific conditions (Pearl, 1995; Pearl & Mackenzie, 2018; Huntington-Klein, 2021).

The minimum adjustment set in a causal model consists of the smallest set of variables that must be conditioned on to block all backdoor paths between exposure (fire) and outcome (mechanical alterations in our study), ensuring that observed associations reflect a true causal relationship. This set depends on the specific causal effect being estimated whether the total effect (includes all causal pathways) or the direct effect.

In this study, we focus on estimating the causal pathway through which fire affects rocks. This means assessing the extent to which the observed alterations in lithics can be attributed to fire, while controlling for relevant mediating variables. DAGs construction and analysis was made by means of *ggdag* (Barrett, 2024) and *dagitty* (Textor et al., 2016) R packages.

Naturalistic sample

Establishing the impact of natural fires on the lithic landscape through naturalistic observations presents certain challenges since not all fire-effect indicator variables can be directly observed. However, it is expected that they influence the observed relationships between the model variables. For example, even though the naturalistic sampling was conducted shortly after the fire event, some patterns were observed that may be linked to post-depositional processes, such as the horizontal displacement of artifacts, which showed varying degrees of integrity in the thermally altered assemblages (Cardillo et al., 2022). Similarly, the presence of pre-existing features, such as patinas or weathering, is not easily identifiable, as it cannot be determined with certainty whether they resulted from fire action or are merely mediating variables affecting morphological or mechanical alterations.

Sampling of altered stones involved non-systematic surveys and systematic surface collection with random sampling in areas affected by 2019–2020 wildfires, focusing on zones near Las Grutas (see Figure 1). The southern access area covered 20,340 m², and the northern access area encompassed 25,640 m² (see Figure 1). In this initial phase, a total of *n* = 203 burned lithic

fragments were collected. Clear evidence of fire impact on the landscape was recorded, including charred stems and wood, reddening (rubefaction) of surface sediments surrounding the remains of burned shrubs, and, in some instances, burned bones (Figure 3). Regarding the lithic record, pebbles composed of acidic volcanic, intermediate-basic volcanic rocks, silicified stones and sedimentary — some of which possess qualities suitable for knapping — were identified, exhibiting signs of thermal alteration. These included color changes and in situ fracturing consistent with exposure to high temperatures (Figure 3). All findings were documented through digital photography in the field, with their positions georeferenced using GPS. The collected materials were subsequently transported for laboratory analysis. The analyzed assemblage consists of 203 lithic fragments corresponding to 58 pebbles that could be refitted either totally or partially (Cardillo et al., 2022). The most frequently observed surface alterations on the rocks include color changes, followed by the development of patinas. Less commonly recorded were oxidation, carbonization, and the presence of thermally-induced flake scars. Technological attributes were also documented, allowing for the identification of artifact classes (flakes and cores) as well as distinct pseudo-artifacts (undifferentiated debris and pseudo-flakes), (see also supplementary material).



Figure 3 - Top panel: effects of fire on the landscape observed in 2021 and 2022. Bottom panel: burned surface lithic scatters showing effects such as fractures and color changes.

Regarding integrity, the degree of fragmentation into smaller particles — often rapidly dispersed — was assessed. This study focuses particularly on fracture patterns. In situ observations enabled the identification of refittable fragments, their spatial distribution, and the occurrence of breakage relative to rock type. Due to extensive fragmentation, in a number of cases ($n = 8$) it was not possible to estimate the original shape of the specimens. Therefore, for the purposes of model comparison, a subset of 50 cases was selected.

The selection of variables and the construction of the causal model were informed by empirical observations made during these field surveys (Cardillo et al., 2022). Variables included in the DAG are the outcome Condition (fractured, not_fractured); pebble shape, grouped following Waters' (1992) shape classification as a reference: oblate and bladed forms were classified as tabular, while equant and prolate forms were classified as rounded; rock type (sedimentary [Sed], acidic volcanics [AV], intermediate-basic volcanics [IBV], and siliceous rocks [Sil]), and the presence of fissures (absence = N, presence = Y). These variables were selected based on their consistent visibility, presumed relevance to thermal response (see also Cardillo et al., 2022; Carranza & Cardillo, 2024), and their potential influence on mechanical alteration patterns observed in burned lithic assemblages. Other factors, such as biomass availability, trampling intensity, or pre-existing patinas, although recognized as potentially influential, could not be systematically recorded during fieldwork and were therefore treated as unobserved variables in the model.

Methods: Decision Tree

The Random Forest approach relies on a bootstrapping procedure that generates an ensemble (or forest) of decision trees, each trained on a random subset of the data and predictors (Breiman, 2001, 2002). Compared to other classification or regression methods, the ensemble prediction of Random Forest tends to have lower variance and higher predictive performance. Also, one advantage of these methods is that they inherently account for variable interactions through hierarchical data partitioning, making them suitable for uncovering complex dependencies between variables. The performance of the model is measured using the Out-of-Bag (OOB) score, a confusion matrix, and two common feature importance metrics: Mean Decrease in Gini (MDG) and Mean Decrease in Accuracy (MDA). MDG quantifies the reduction in Gini impurity when a feature is used for node splitting in decision trees. A higher MDG value indicates that the feature plays a significant role in separating the data. MDA evaluates feature importance by measuring the drop in model accuracy when a feature is randomly permuted. Both metrics provide insights into feature importance but do not directly measure overall classification performance. Random forest classification and metrics was performed by means of randomForest R package version 4.7-1.2 (Liaw & Wiener, 2002).

However, classification models are sensitive to class imbalance in the data, as they tend to assign more weight to the most frequent categories (Menardi & Torelli, 2014). For this reason, when one class is significantly underrepresented, classifiers tend to ignore it, leading to poor model performance. This is common in observational studies, where the distribution of predictors cannot be controlled. To analyze the impact of class imbalance in the data, a second Random Forest model was fitted using oversampling with ROSE (Random Over-Sampling Examples), an R package designed to address the issue of class imbalance in binary classification (Lunardon et al., 2014). ROSE generates artificially balanced samples using a smoothed bootstrap approach, which improves both model estimation and accuracy evaluation when dealing with rare classes (Lunardon et al., 2014). Finally, both models were compared to estimate the bias in the naturalistic dataset and to assess changes in the classification model's performance in the context of more balanced samples, (see supplementary material for the complete R code and analytical workflow).

In this DAG, Fire, although not directly observed, is the treatment variable and Mechanical Alterations (Mech_Alt) is the outcome. The causal pathway from Fire to Mech_Alt is mediated by three variables: Fissures, Shape, and Rock Type, meaning that Fire influences these variables, which in turn affect Mech_Alt. For example, Fire might affect the likelihood of fracture while interacting with previous fissures in the pebble. Also, different shapes could be heated or cooled at different rates, and it is expected that different rock types could increase or reduce the effect of fire (Cattáneo et al., 1997; Mercieca & Hiscock, 2008; Halbrucker et al., 2021). Additionally, there are three unobserved variables — Biomass, Trampling, and Knapping — that could potentially influence the system. These variables are not measured or explicitly included in the model, but they may confound the relationships between Fire the mediators, and the outcome. For instance, Biomass could affect the intensity of Fire, Trampling might independently produce pseudo-artifacts and Knapping could directly cause mechanical alterations. Since these variables are unobserved,

their effects are not accounted for, which could introduce bias or uncertainty in estimating the direct and indirect effects of Fire on Mech_Alt.

In this case, five open paths were identified from Fire to Alt_Mech. Additionally, no paths were observed from a third variable linking Fire to Mech-Alt, which could act as confounders; therefore, it is not necessary to control for other variables. The minimal adjustment set includes only those variables directly contributing to the effect (Fissures, Rock_Type, and Shape) (see also supplementary material).

Results

In this sample, the percentage of fractured pebbles by shape, rock type, and presence of surface fissures suggests that rounded forms have higher percentages of fractured AV and IBV specimens, particularly among those with no visible surface fissures (27% and 30%, respectively) (Figure 4).

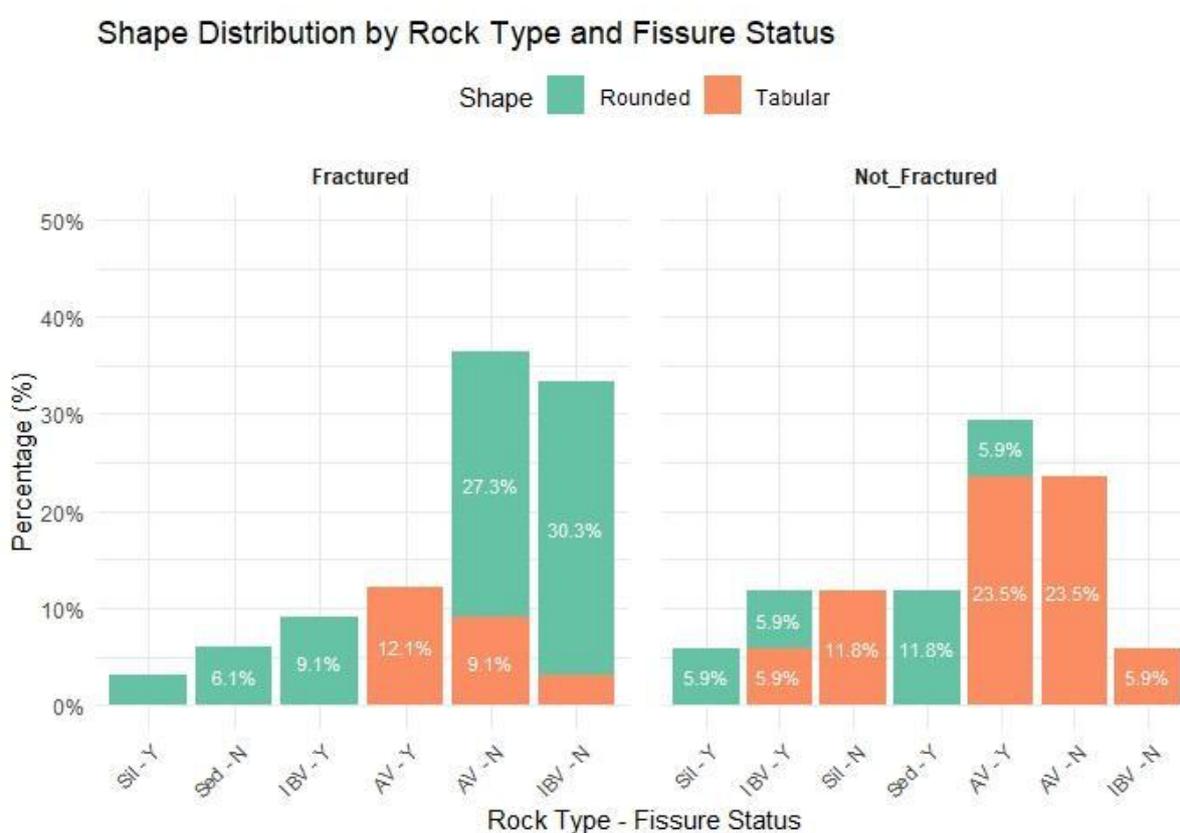


Figure 4 - Relative frequency of Fissured surface (N,Y) for each Rock_Type (Sil, Sed, IBV, AV) and Shape (Rounded, Tabular) by Condition (Fractured, Not_Fractured).

Rounded pieces also show the highest occurrence of fractures among specimens with or without surface fissures. As for the unfractured pieces, the highest percentage corresponds to tabular forms, with 23.5% for AV, regardless of the presence of fissures (Figure 4). Therefore, in this assemblage, the occurrence of fractures appears to be more closely related to shape, while no clear pattern of fractures associated with the presence of surface fissures on the nodule is observed.

To explore the contribution of these three variables in relation to mechanical fractures, two Random Forest classification models were constructed: one based on the frequencies quantified

during field sampling, and another using balanced factors through synthetic resampling, in order to assess the impact of sample imbalance on the results.

The Random Forest model was used to classify rock conditions (Fractured vs. Not Fractured) based on Shape, Rock Type, and Fissures, with 1000 trees and $mtry = 1$. The out-of-bag (OOB) error rate was 48%, indicating moderate misclassification. The confusion matrix showed better performance for predicting fractured rocks (error rate 27.3%) than non-fractured ones (error rate 64.7%), (Table 1, see also supplementary material). The high misclassification rate for not fractured instances suggests that the selected variables may not be sufficiently discriminative.

The variable importance plots illustrate the contribution of each predictor to the classification performance (Figure 5). The left plot, based on the mean decrease in accuracy, shows that Shape is the most important variable, followed by Fissures and Rock Type, indicating that removing Shape would lead to the most significant drop in model accuracy (Figure 5). The right plot, which measures mean decrease in Gini, also confirms that Shape has the highest importance in determining node purity, with Rock Type playing a moderate role and Fissures being the least influential. These results suggest that Shape is the strongest predictor of rock condition, whereas Rock Type and Fissures contribute less to the classification.

Variable Importance Plot

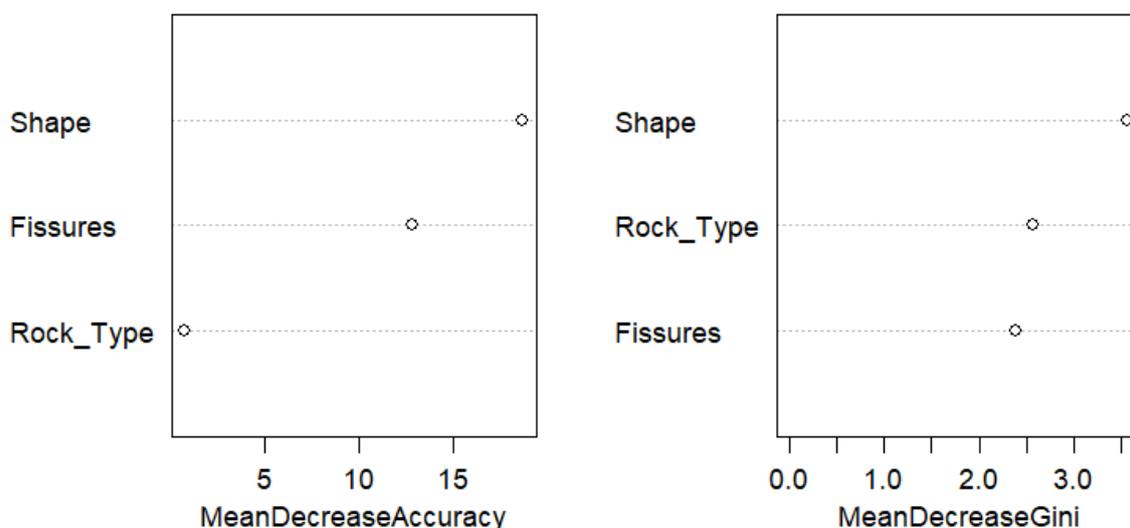


Figure 5 - Variable importance plot for naturalistic model.

The estimated probability of fracture varies across Shape and Rock Type categories, as shown in the bar plot (Figure 6). Rounded shapes exhibit a higher probability of fracture than tabular shapes, regardless of rock type. Within the fissures group, where fissures are present (Y), IBV rocks show the highest probability of fracture (~0.75), while Sed rocks exhibit the lowest (~0.3). The fracture probability decreases for tabular shapes, with IBV still showing the highest value (~0.4) and Sil and Sed the lowest (~0.25). In contrast, within the not fissured samples (N), fracture probabilities increase, particularly for Rounded shapes of AV and IBV, both reaching a probability of 1.0, indicating that all rocks of these types fractured. Sed and Sil also show elevated probabilities (~0.85 and ~0.7, respectively). For tabular shapes, fracture probabilities remain lower, with IBV and AV around 0.5 and Sil showing the lowest value (~0.2). These results suggest that rounded shapes are more prone to fracture than tabular ones, and IBV rocks consistently exhibit higher

fracture probabilities, while Sil and Sed tabular rocks appear to be the most resistant to fracturing (see also supplementary material).

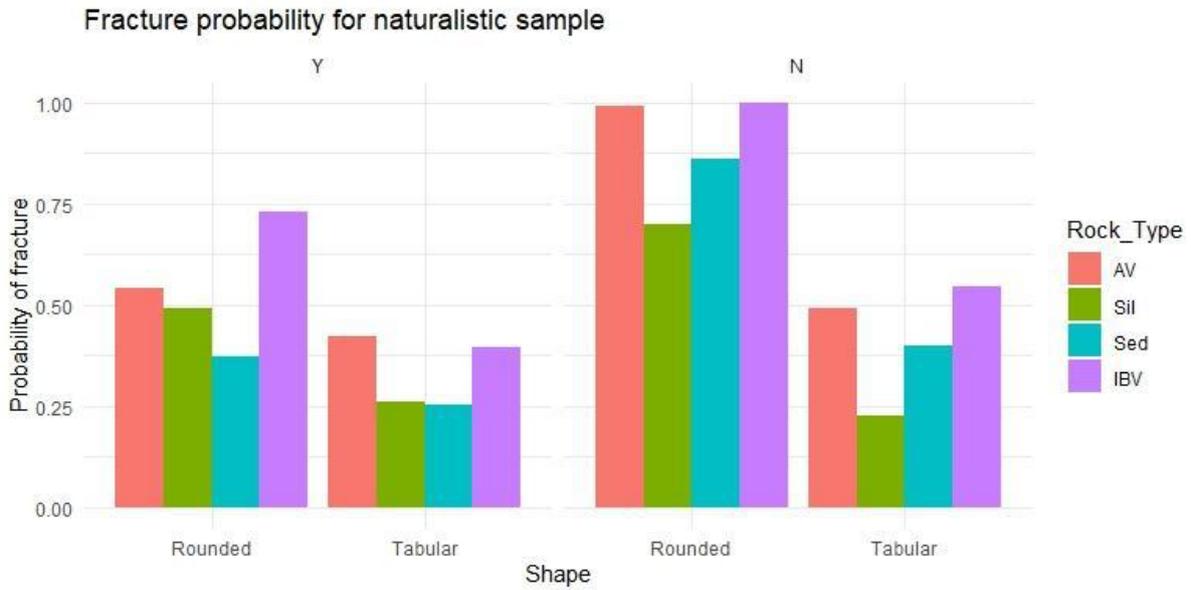


Figure 6 - Probability of fracture estimated by means of random forest trees and naturalistic samples related to different nodule shape and rock type by fissures (Y, N).

A notable reduction in misclassification error was observed in random forest performance when using a balanced dataset, decreasing the out-of-bag error rate from 48% to 16% (Figure 6). Variable importance analysis confirmed shape as the most influential predictor, followed by fissures and rock type (Figure 7).

Variable Importance for balanced model

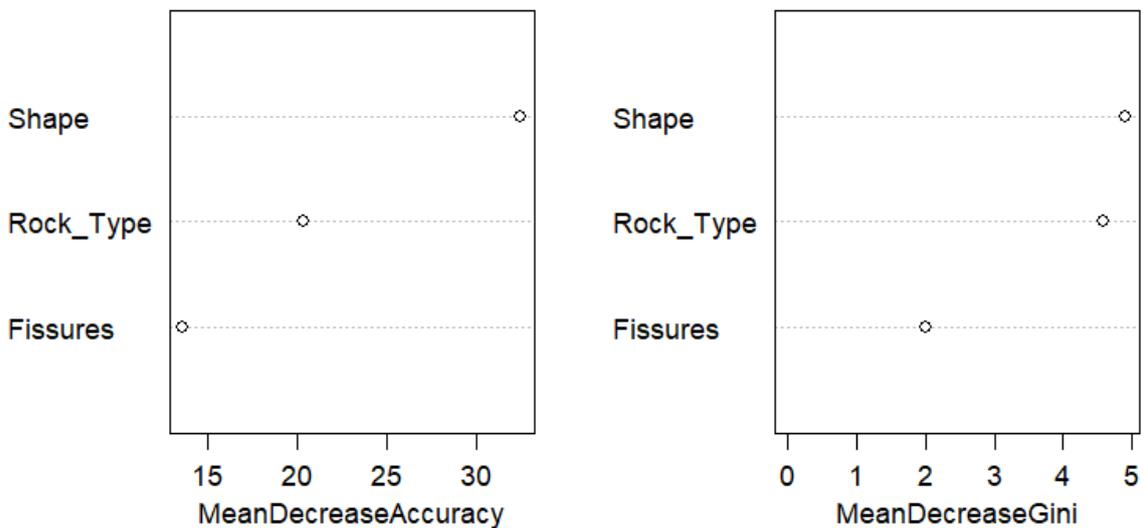


Figure 7 - Variable importance plot for balanced model.

The barplot of estimated fracture probability for each factor level, indicates IBV rocks show the highest probability of fracture (~0.75), while Sed rocks exhibit the lowest (~0.25) (Figure 8). The fracture probability decreases for tabular shapes, with IBV still showing the highest value (~0.5) and Sil and Sed the lowest (~0.25). Estimated probabilities for balanced trees also suggest that the presence of fissures (Y) could be a clear factor in explaining fracture probability, and maybe even reduce it (see discussion). These results suggest that rounded shapes are more prone to fracture than tabular ones, and IBV rocks consistently exhibit higher fracture probabilities, while Sil and Sed tabular rocks appear to be the most resistant to fracturing.

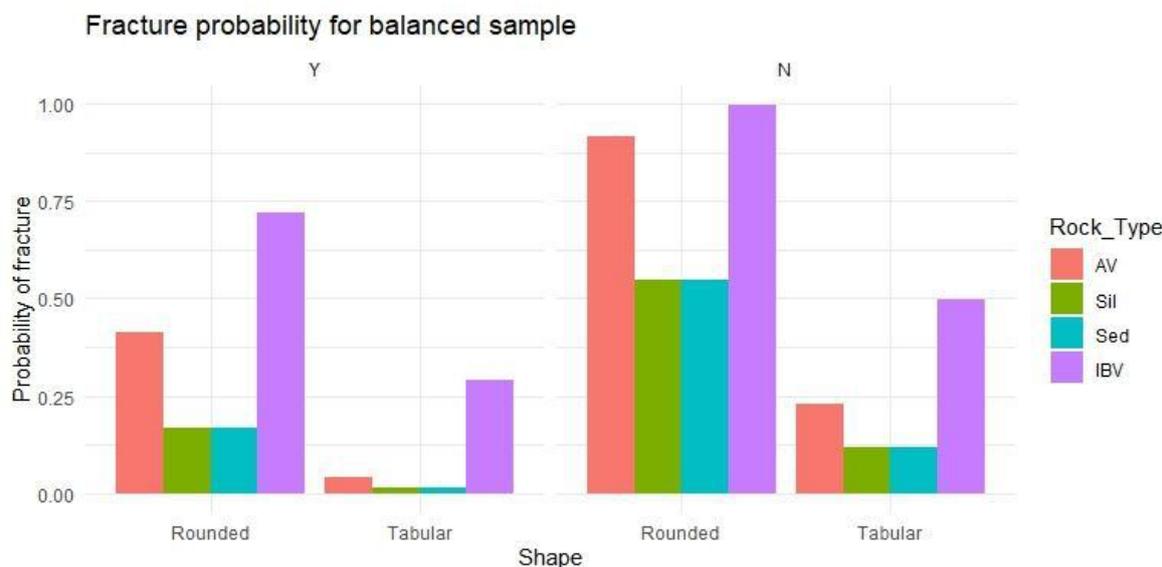


Figure 8 - Probability of fracture estimated by means of random forest trees and balanced sample related to different nodule shape and rock type by fissures (Y, N).

The comparison between two models shows that the balanced sample model significantly improved classification, reducing the overall error from 40% to 16% (Table 1). Also, the balanced sample achieved much better accuracy for “Not_Fractured” cases (92%), while the first model had higher misclassification (only 35% correct). Thus, results between unbalanced and balanced samples suggest that differences in factor levels have an impact in results, in particular in overall error and in the accuracy to predict not fractured samples.

Table 1 - Performance comparison between the two models.

Metric	Nat_sample	Bal_sample
OOB Error Rate	40%	16%
Fractured Accuracy	24/33 (72.7%)	19/25 (76%)
Not_Fractured Accuracy	6/17 (35%)	23/25 (92%)
Class Error (Fractured)	27.30%	24%
Class Error (Not_Fractured)	64.70%	8%

In general, results of the Random Forest classification align with the causal assumptions proposed in the DAG. Pebble shape, which was identified as the most influential predictor, corresponds to a key mediating variable hypothesized to transmit the effects of fire on mechanical alterations. Similarly, the contribution of rock type, although lower, are consistent with their proposed roles as mediator. The agreement between the model structure and the variable importance rankings supports the validity of the causal framework, suggesting that the mechanisms formalized in the DAG adequately capture the main pathways through which fire influences lithic fragmentation in naturalistic contexts.

Discussion

In both models, *Shape* emerged as the variable with the greatest influence on predicting fire-induced fracture, both in terms of accuracy and the Gini index. In the first model, *Fissures* ranked second in mean accuracy, followed by *Rock_Type*, whereas in the balanced model, this order was reversed. The influence of pebble shape on fracture probability has been previously discussed by various authors (Custer, 2017; Graesch et al., 2014; Pagoulatos, 1992, 2005). For instance, the association between the presence of surface fissures and fracture probability presents a counterintuitive pattern; in the adjusted model, pebbles lacking visible fissures appear more likely to be fractured. However, this may reflect a systematic underestimation of fissures in already broken specimens; it is possible that pre-existing fissures may become indistinguishable after thermal fragmentation, as they effectively constitute the fracture itself. This pattern needs to be studied further in the context of controlled experiments. As expected, the difference between models shows that class balance has a measurable impact on the relative importance of predictors. Furthermore, the balanced model demonstrated a lower average prediction error for both the presence and absence of fractures. This finding suggests that increasing the sample size or employing a stratified sampling strategy may be more effective in this case, ensuring a more even distribution across predictor variables. To this end, ongoing fieldwork is implementing random sampling within the fire-affected area using variable-sized collection grids. These grids are distributed across the landscape to account for differences in surface lithic density and to optimize data recovery for statistical analysis. Despite the usefulness of DAGs in making causal relationships explicit and formal, in naturalistic contexts one must assume the empirical completeness or sufficiency of the model, that is, that all relevant relationships are included and no unobserved confounding variables are influencing the observed outcomes. This assumption cannot be fully verified, as not all phenomena and their interactions can be observed, particularly given the complexity of natural processes compared to those in controlled, experimental settings. In this study, although fire was not directly measured, we evaluated its influence through the predictive power of its mediators — shape, fissures, and rock type — on pebble fracture. Since these variables lie along the causal pathway from fire to mechanical alteration, the model indirectly captures the structure of the fire effect, as formalized in the DAG. In this regard, the use of experimental or quasi-experimental models can serve to provide external validation criteria for the proposed models and help mitigate potential biases.

A potential limitation of the current study is the presence of unobserved confounders, such as trampling, kicking, and other post-depositional processes, which may influence the fragmentation and dispersion patterns of lithic materials after fire events, potentially leading to intrusion into archaeological sites (Schiffer, 1987; Petraglia, 2002; Frank & Baridón, 2021). Although these processes were recognized as plausible factors that could increase fragmentation, promote dispersion, or create clusters, they could not be systematically recorded during field surveys. To address this limitation, we are currently tracking fire-altered nodules *in situ*, in order to model post-depositional modifications over time. The possibility that this area has been subjected to multiple fire events over time raises important considerations for the cumulative effects of thermal stress on lithic materials. Repeated exposure to fire likely results in progressive fragmentation and surface modification, potentially following an exponential degradation pattern. While the present study centers on a recent, well-documented wildfire, future research could benefit from comparing surface assemblages with buried pebble deposits dated to the Pleistocene, which presumably experienced fewer fire episodes. Such comparisons would help to better understand the long-term impacts of recurrent fire on archaeological materials in this region.

A logical next step involves assessing the potential for generalizing this model. Applying a causal model to different contexts is conceptually akin to evaluating the model's external validity. Transportability and external validity (see Eren et al., 2016) both address the generalization and intervention or the results of observations across different contexts, but they differ in approach. Transportability, rooted in causal inference, assesses whether causal relationships established in one setting (e.g. an experiment) can be applied to another (e.g., a natural context) using DAGs to

evaluate differences and identify necessary adjustments (Huntington-Klein, 2021; see also Bulbulia, 2024). External validity, a broader concept in research design, concerns whether study results can be generalized to different populations, locations or conditions, without necessarily focusing on causal mechanisms. While external validity evaluates the overall applicability of findings, transportability provides a structured, causal framework to determine when and how causal effects can be transferred between settings. In this case transportability of our results to other observational contexts need to build a DAG for the target context and identify structural differences between the two DAGs — this will reveal potential bias sources that could limit transportability. A recent experimental study (Carranza & Cardillo, 2024) was conducted to validate the general patterns observed in the naturalistic context, particularly the incidence of fractures in local raw materials from the study area (acidic volcanics and intermediate to basic volcanic rocks). During the experiment, a fracture rate between 30% and 40% was recorded. These values are consistent with those observed in the naturalistic context, showing a significant association among the various alteration variables. Additionally, a range of morphological alterations typical of fire exposure — such as those previously described — were reproduced (Carranza & Cardillo, 2024). The comparison between naturalistic observations and experimental analysis will be highly valuable for assessing the potential to generalize the latter across different archaeological contexts.

Conclusions

From an archaeological perspective, the findings presented here have significant implications for the interpretation of fire-affected lithic assemblages. The central role of pebble shape, and to a lesser extent rock type and fissures, suggests that mechanical fragmentation patterns in burned contexts are not random but rather mediated by specific, observable properties of the lithic materials. The minimal adjustment set identified through the causal model, along with the low contribution of unobserved confounders, supports the notion that fire can independently generate recognizable patterns of fracture in surface lithic scatters, without necessarily requiring human agency. This insight is crucial for evaluating the integrity of archaeological assemblages and for distinguishing between intentional human activity and taphonomic processes, particularly in palimpsests where pseudo-artifacts generated by natural fires may mimic human-produced flakes. More broadly, the application of causal modeling to actualistic archaeological data offers a promising avenue for improving our understanding of site formation processes under complex natural disturbance regimes. Future research should focus on testing the generalizability of the proposed causal model across different environmental and archaeological contexts, and on integrating experimental and observational datasets to further refine the identification of fire-induced taphonomic signatures.

Acknowledgments

To the research team for their collaboration in field tasks. The work permits were granted by the National Parks Administration and the Río Negro Secretary of Culture. We acknowledge the use of data and/or imagery from NASA's Fire Information for Resource Management System (FIRMS) (<https://earthdata.nasa.gov/firms>), part of NASA's Earth Science Data and Information System (ESDIS). We are very grateful to the reviewers and editors for their time, effort, and constructive feedback throughout the evaluation process.

Preprint version 3 of this article has been peer-reviewed and recommended by PCI Archaeology (<https://doi.org/10.24072/pci.archaeo.100613>) (Arzarello, 2025).

Funding

This research was funded by the National Scientific and Technical Research Council of Argentina [PIP CONICET 112-202101-00908 CO] and National Agency for Scientific and Technological Promotion [PICT ANPCyT 2021-I-A-00436].

Conflict of interest disclosure

The authors of this preprint declare that they have no financial conflict of interest with the content of this article.

Data and supplementary information

Data or any supplementary material are available at <https://zenodo.org/records/16335811> (Cardillo & Carranza 2025).

References

- Alberti, J. (2016). Disponibilidad y explotación de materias primas líticas en la costa de Norpatagonia (Argentina): Un enfoque regional. *British Archaeological Reports, International Series 1901*, #27. Archaeopress. <https://doi.org/10.2307/j.ctv1zckxn7>
- Alberti, J., and Cardillo, M. (2018). El registro lítico en la costa del golfo San Matías (Argentina). Análisis comparativos de los materiales líticos provenientes de depósitos de superficie, enterrados y concheros de la costa rionegrina. *Revista Chilena de Antropología*, 38, 310–329.
- Alberti, J., and Carranza, E. (2014). Primera caracterización de los conjuntos líticos provenientes de depósitos de tipo conchero en la costa del golfo San Matías (Río Negro, Argentina). *La Zaranda de Ideas. Revista de Jóvenes Investigadores en Arqueología*, 10, 47–64.
- Arzarello, M. (2025). Fire as a Taphonomic Agent: Statistical and Causal Modelling of Lithic Assemblages. *Peer Community in Archaeology*, 100613. <https://doi.org/10.24072/pci.archaeo.100613>
- Barrett, M. (2024). ggdag: Analyze and Create Elegant Directed Acyclic Graphs. R package version 0.2.13.9000, commit 46d7ca800ae1f617a4a2f8c33f68dc7f644b59c6, <https://github.com/malcolmbarrett/>
- Borella, F., Cardillo, M., Alberti, J., Scartascini, F., Carranza, E., Favier Dubois, C., and Guichón-Fernández, R. (2020). Resultados preliminares de las investigaciones arqueológicas en el Área Natural Complejo Islote Lobos, costa oeste del golfo San Matías (provincia de Río Negro). *Revista del Museo de Antropología*, 13, 57–68.
- Bulbulia, J. A. (2024). Methods in causal inference. Part 1: Causal diagrams and confounding. *Evolutionary Human Sciences*, 6, e40. <https://doi.org/10.1017/ehs.2024.35>.
- Breiman, L. (2001). Random Forests, *Machine Learning*, 45(1), 5-32. <https://doi.org/10.1023/a:1010933404324>
- Breiman, L. (2002). Manual On Setting Up, Using, And Understanding Random Forests V3.1, https://www.stat.berkeley.edu/~breiman/Using_random_forests_V3.1.pdf. <https://doi.org/10.32614/cran.package.randomforest>
- Cardillo, M. (2013). Cambios en el paisaje, uso del espacio y conjuntos líticos promediados en la costa norte del golfo San Matías (Río Negro, Argentina) durante el Holoceno medio-tardío. *Comechingonia virtual*, 1, 1–26. <https://doi.org/10.4067/s0718-22442013000100013>
- Cardillo, M., and Alberti, J. (2013). Stone tool manufacture strategies and lithic raw material exploitation in coastal Patagonia, Argentina: A multivariate approach. *Journal of Archaeology*, 1–13. <https://doi.org/10.1155/2013/128470>
- Cardillo, M., Carranza, E., Alberti, J., and Borella, F. (2022). Alteraciones térmicas en guijarros costeros en la localidad de Las Grutas (Río Negro). Discutiendo sus implicancias para la

- interpretación del registro arqueológico lítico. *Revista Del Museo De Antropología*, 15(3), 273–288. <https://doi.org/10.31048/1852.4826.v15.n3.38007>
- Cardillo, M., and Carranza, E. (2025). Data and supplementary material for “Fire as a Taphonomic Agent: Statistical and Causal Modelling of Lithic Assemblages”. *Zenodo*. <https://doi.org/10.5281/zenodo.16335811>
- Carranza, E. and Cardillo, M. (2024): Burned Lithics: An Experimental Approach to Characterize the Effects of Fire on the Lithic Archaeological Record of Northern Patagonia, Argentina. *Lithic Technology*. <https://doi.org/10.1080/01977261.2024.2425876>
- Cattáneo, R., Pupio, A., Valente, M. and Barna, A.(1997). Alteración térmica en dos tipos de rocas silíceas: resultados experimentales y aporte de datos para el análisis arqueológico. *Relaciones de la Sociedad Argentina de Antropología*, XXII-XXIII: 343-361.
- Custer, J. F. (2017). Experimental analysis of fire-cracked rocks from varied use contexts: Fracture attributes. *North American Archaeologist*, 38(3), 237-291. <https://doi.org/10.1177/0197693117696533>
- Deffner, D., Rohrer, J. M., and McElreath, R. (2022). A causal framework for cross-cultural generalizability. *Advances in Methods and Practices in Psychological Science*, 5(3), 1–18. <https://doi.org/10.1177/25152459221106366>
- Eren, M. I., Lycett, S. J., Patten, R. J., Buchanan, B., Pargeter, J., and O’Brien, M. J. (2016). Test, model, and method validation: The role of experimental stone artifact replication in hypothesis-driven archaeology. *Ethnoarchaeology*, 8(2), 103–136. <https://doi.org/10.1080/19442890.2016.1213972>
- Favier Dubois, C. M. (2019). Human occupation chronologies modeled by geomorphological factors: A case study from the Atlantic Coast of Northern Patagonia (Argentina). In: Inda Ferrero, H., García Rodríguez, F. (eds) *Advances in Coastal Geoarchaeology in Latin America, The Latin American Studies Book Series*. Springer, Cham. https://doi.org/10.1007/978-3-030-17828-4_1
- Frank, A., and Baridón, J. (2021). The effect of fire in the distribution of lithic assemblages: An experimental approach. *Lithic Technology*, 47(2), 133–146. <https://doi.org/10.1080/01977261.2021.1981653>
- Graesch, A. P., DiMare, T., Schachner, G., Schaepe, D. M., and Dallen, J. (2014). Thermally modified rock: the experimental study of “fire-cracked” byproducts of hot rock cooking. *North American Archaeologist*, 35(2), 167-200. <https://doi.org/10.2190/na.35.2.c>
- Grace, J.B., Huntington-Klein, N., Schweiger, E.W., Martinez, M., Osland, M.J., Feher, L.C., Guntenspergen, G.R. and Thorne, K.M. (2025) Causal Effects Versus Causal Mechanisms: Two Traditions With Different Requirements and Contributions Towards Causal Understanding. *Ecol Lett*. 28(4):e70029. <https://doi.org/10.1111/ele.70029>
- Halbrucker, É., Fiers, G., Vandendriessche, H., De Kock, T., Cnudde, V., and Crombé, P. (2021). Burning flint: An experimental approach to study the effect of fire on flint tools. *Journal of Archaeological Science: Reports*, 36, 102854. <https://doi.org/10.1016/j.jasrep.2021.102854>
- Huntington-Klein, N. (2021). *The Effect: An Introduction to Research Design and Causality*. Taylor & Francis, 646 pp. <https://doi.org/10.1201/9781003226055>
- Liaw, A. and Wiener, M. (2002). Classification and Regression by randomForest. *R News*, 2 (3), 18-22. <https://CRAN.R-project.org/doc/Rnews>.
- Lunardon, N., Menardi, G., and Torelli, N. (2014). ROSE: a Package for Binary Imbalanced Learning. *R Journal*, 6:82–92. <https://doi.org/10.32614/rj-2014-008>
- McElreath, R. and Koster, J. (2024). The End of Human Behavioral Ecology. In: Koster J., Scelza B. and Shenk M.K. (Eds). *Human Behavioral Ecology*. Cambridge Studies in Biological and Evolutionary Anthropology: Cambridge University Press, 402-419. <https://doi.org/10.1017/9781108377911.018>
- Mercieca, A., and Hiscock, P. (2008). Experimental insights into alternative strategies of lithic heat treatment. *Journal of Archaeological Science*, 35(9), 2634-2639. <https://doi.org/10.1016/j.jas.2008.04.021>

- Menardi, G. and Torelli, N. (2014). Training and assessing classification rules with imbalanced data. *Data Mining and Knowledge Discovery*, 28:92–122. <https://doi.org/10.1007/s10618-012-0295-5>
- Oyarzabal, M., Clavijo, J., Oakley, L., Biganzoli, F., Tognetti, P., Barberis, I., Maturo, H. M., Aragón, R., Campanello, P. I., Prado, D., Oesterheld, M., and León, R. J. C. (2018). Unidades de vegetación de la Argentina. *Ecología austral*, 28(1), 40–63. <https://doi.org/10.25260/EA.18.28.1.0.399>
- Pagoulatos, P. (1992). The re-use of thermally altered stone. *North American Archaeologist*, 13(2), 115-129. <https://doi.org/10.2190/buh5-3bvn-up4k-j91g>
- Pagoulatos, P. (2005). Experimental burned rock studies on the Edwards Plateau: a view from Camp Bullis, Texas. *North American Archaeologist*, 26(3), 289-329. <https://doi.org/10.2190/9j4u-1m73-1wt4-ad9n>
- Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, 82(4), 669–710. <https://doi.org/10.2307/2337329>
- Pearl, J., and Mackenzie, D. (2018). *The Book of Why: The New Science of Cause and Effect*. Penguin Books Limited.
- Petraglia, M. D. (2002). The heated and the broken: Thermally altered stone, human behavior, and archaeological site formation. *North American Archaeologist*, 23(3), 241–269. <https://doi.org/10.2190/4d7x-h9hr-1p4j-417l>
- Snodgrass, J. G., Dengah, H. F., Sagstetter, S. I., and Zhao, K. X. (2024). Causal inference in ethnographic research: Refining explanations with abductive logic, strength of evidence assessments, and graphical models. *Plos one*, 19(5), e0302857. <https://doi.org/10.1371/journal.pone.0302857>
- Sottile, G. D., Giaché, Y. S. and Bianchi, M. M. (2018). Reconstrucción del régimen de incendios en ecosistemas templados patagónicos sobre la base de registros de carbón vegetal sedimentario (Charcoal) y polen durante el Cuaternario tardío. Tendencias metodológicas, resultados y perspectivas. In : Prieto A. R. (Ed.), *Metodologías y estrategias del análisis palinológico del Cuaternario tardío. Publicación Electrónica de la Asociación Paleontológica Argentina*, Asociación Paleontológica Argentina: Vol. 18, No. 2, pp. 102–119. <https://doi.org/10.5710/peapa.23.07.2018.262>
- Schiffer, M.B., (1987). *Formation Processes of the Archaeological Record*. University of New Mexico Press, Albuquerque.
- Shmueli, G. (2010). To explain or to predict? *Statistical Science*, 25(3), 289–310. <https://doi.org/10.1214/10-STS330>
- Textor, J., van der Zander, B., Gilthorpe, M.S., Liškiewicz, M. and Ellison, G.T. (2016). Robust causal inference using directed acyclic graphs: the R package. *Dagitty. International Journal of Epidemiology*, 45 (6), 1887-1894. <https://doi.org/10.1093/ije/dyw341>.
- Waters, M. R. (1992). *Principles of geoarchaeology: A North American perspective*. University of Arizona Press.
- Zacconi, G., and Toppazzini, M. (2018). Áreas afectadas por incendios forestales y rurales en la región pampeana y noreste de la región patagónica durante la temporada 2016- 2017. Informe Técnico N° 13. Servicio Nacional de Manejo del Fuego, Ministerio de Ambiente y Desarrollo Sustentable de la Nación.