Journal of Archaeological Science 144 (2022) 105626

Contents lists available at ScienceDirect

Journal of Archaeological Science

journal homepage: www.elsevier.com/locate/jas

Introducing SourceXplorer, an open-source statistical tool for guided lithic sourcing

Rhy McMillan [a,b,*], Nicholas Waber [c], Morgan Ritchie [a], Ellery Frahm [d]

[a] Department of Anthropology, University of British Columbia, Vancouver, British Columbia, Canada
[b] Research Unit: Analytical, Environmental, and Geo-chemistry, Department of Chemistry, Vrije Universiteit Brussel, AMGC-WE-VUB, Brussels, Belgium
[c] Terra Archaeology, Victoria, British Columbia, Canada
[d] Council on Archaeological Studies, Department of Anthropology, Yale University, New Haven, CT, USA

ARTICLE INFO

Keywords:
R Shiny
Open science
Lithic provenance
Stone tools
Trace elements
X-ray fluorescence
Sts'ailes coast Salish

ABSTRACT

Archaeologists' access to analytical infrastructure has grown exponentially over the last two decades. This is especially the case for benchtop X-ray fluorescence (XRF) and portable XRF (pXRF) instruments, which are now practically commonplace in archaeological laboratories and provide users with a non-destructive and rapid means to analyze the elemental compositions of archaeological specimens. As XRF has become more accessible, the volume of analytical measurements available in archaeological datasets as well as the number and diversity of researchers participating in data collection have inherently increased. Those researchers, who have various levels of experience with the nuances of lithic sourcing procedures, are also often the ones attempting to interpret the elemental data they produce. While standardized analytical procedures have enabled inexperienced analysts to take accurate and reproducible XRF measurements, interpreting the resulting data is more difficult to convert and standardize with the same degree of user-friendliness. To address this challenge, we have bundled a series of statistical approaches and data exploration tools into an intuitive open-source graphical user interface designed to facilitate reproducible and robust outcomes during lithic sourcing studies. Our application, SourceXplorer, permits easy access to and exploration of numeric baseline data using a map interface while facilitating a guided interpretation of source affiliations for archaeological specimens (e.g., lithics) within any natural context using multivariate statistical analyses. We demonstrate SourceXplorer's functionality in relation to a complex archaeological challenge by examining evidence for the procurement and use of lithic material from previously undocumented toolstone source locations in southwestern British Columbia, Canada. We also provide open access to SourceXplorer, including both a deployed version of the application that can be used with any Internet browser and the packaged script, which can be run locally in the open-source R statistical programming environment.

1. Introduction

Archaeologists' access to the analytical equipment required for investigating artifact chemistry has proliferated over the last two decades, expanding from instruments housed within dedicated facilities to desktop and portable instruments that are now practically commonplace in archaeological laboratories. While these advancements have resulted in a rapid increase in the volume of analytical measurements, leading to the production of large archaeological datasets, they have also increased the number and diversity of researchers who take measurements and interpret the collected data. The most prolific example – and thus, by default, the most hotly debated and sometimes bitterly contested one – derives from a geochemical technique with a long history in the archaeological sciences: X-ray fluorescence (XRF; e.g., Hall, 1958, 1960; Hammond, 1961; Hall et al., 1964). Decades after the first XRF analysers, new instruments (i.e., portable XRF or pXRF; also sometimes called handheld XRF) have become smaller as well as more sophisticated, automated, and user-friendly tools to non-destructively measure elemental concentrations in archaeological materials, typically with the goal of identifying relationships between stone artifacts and their natural sources. Using standardized methods, the expanding number and diversity of pXRF users does not pose a problem, and in some ways, it is highly desirable for innovation in archaeological science (Latour and Woolgar, 1979). Less experienced pXRF users can now consistently take

* Corresponding author. Department of Anthropology, University of British Columbia, Vancouver, British Columbia, Canada.
E-mail address: rhy.mcmillan@ubc.ca (R. McMillan).

https://doi.org/10.1016/j.jas.2022.105626
Received 6 January 2022; Received in revised form 2 May 2022; Accepted 17 May 2022
Available online 2 June 2022
0305-4403/

accurate and reproducible measurements when, for example, the instrument is mounted in a test stand rather than held in hand or when provided with suitable calibration standards and methods (Frahm, 2014, 2019a, 2019b). In contrast, interpreting the resulting data, especially when using trace element concentrations for artifact sourcing studies, has been much more challenging to adapt with the same degree of user-friendliness (see Frahm et al., 2014 for an early example of yielding an automated obsidian source identification). Despite the technical challenges involved, automating this step will not only enhance the opportunities for researchers new to pXRF and/or lithic sourcing studies to produce geochemically and archaeologically robust interpretations of the data they produce but also increase the analytical consistency between different investigators.

Prior to the emergence of pXRF instruments, elemental measurements were usually undertaken in dedicated facilities, often by the specialists who oversaw these laboratories. In the case of obsidian artifact sourcing, for example, such laboratories tended to focus on certain geographic regions (e.g., Geoarchaeological XRF Lab with the North American Southwest as in Liebmann, 2017, Northwest Research Obsidian Studies Laboratory with the Pacific Northwest of North America as in Ebert et al., 2015). As a result, those specialists commonly accumulated and applied a wealth of knowledge about obsidian and other toolstone sources in these areas, including aspects such as their physical locations and accessibility on the landscape as well as the shifting patterns in their use by different peoples through time (e.g., Rick et al., 2001; Lyons et al., 2001). Hence, the identifications made by such specialists involve not only quantitative methods (i.e., using concentrations and ratios of relevant elements, sometimes with sophisticated statistical modeling, other times with simple scatterplots) but also their understanding of a region's geological nuances, likely source locations, and documented past procurement strategies. These pieces of information facilitate sound interpretations of lithic artifact origins, which are valuable for interpreting the behaviors and agency of past peoples through their material culture. Replicating the accrued knowledge and experience of such a specialist is far more technically imposing than streamlining the measurement process by, for example, automating particular settings of pXRF detector electronics. As a result, the capacity to perform robust elemental analyses and produce source identifications for lithic artifacts—an invaluable tool in lithic research everywhere—had been effectively centralized within a handful of facilities where there was a convergence of expertise and specialized equipment.

The ability of any researcher to make accurate and reproducible interpretations of the data they generate is an essential epistemological step to fulfilling the potential of what has been termed the "pXRF revolution" (Shackley, 2011:13). Such an advancement will allow novel, valid archaeological interpretations and knowledge from more diverse perspectives (Waddell, 2018; Frahm, 2019a, 2019b), including Indigenous communities, rather than the continued siloed production of knowledge by and for a small collection of experts, expressed within their established paradigms (see, for example, Liebmann, 2017; 2018 vs. Shackley and Moore, 2018). Our field, which values a diversity of voices, can greatly benefit by creating opportunities for new researchers to generate archaeological knowledge during this revolution, through both the collection and the interpretation of data. This, however, poses a challenge: on one hand, increased data production provides valuable baseline information essential for making informed interpretations of lithic provenance at various geographic and temporal scales. On the other, these data need to be collected appropriately and—equally important—interpreted transparently and consistently, the thorough documentation of which is becoming an increasingly common requirement in academic publishing.

We have addressed this challenge by bundling a suite of statistical approaches and data exploration tools into an intuitive graphical user interface (GUI) that supports transparent and sophisticated interpretations of elemental data obtained from archaeological lithics for sourcing analyses. Our application, SourceXplorer (Fig. 1), facilitates easy and dynamic exploration of numeric baseline data by facilitating a guided interpretation of lithic provenance that can be applied to any

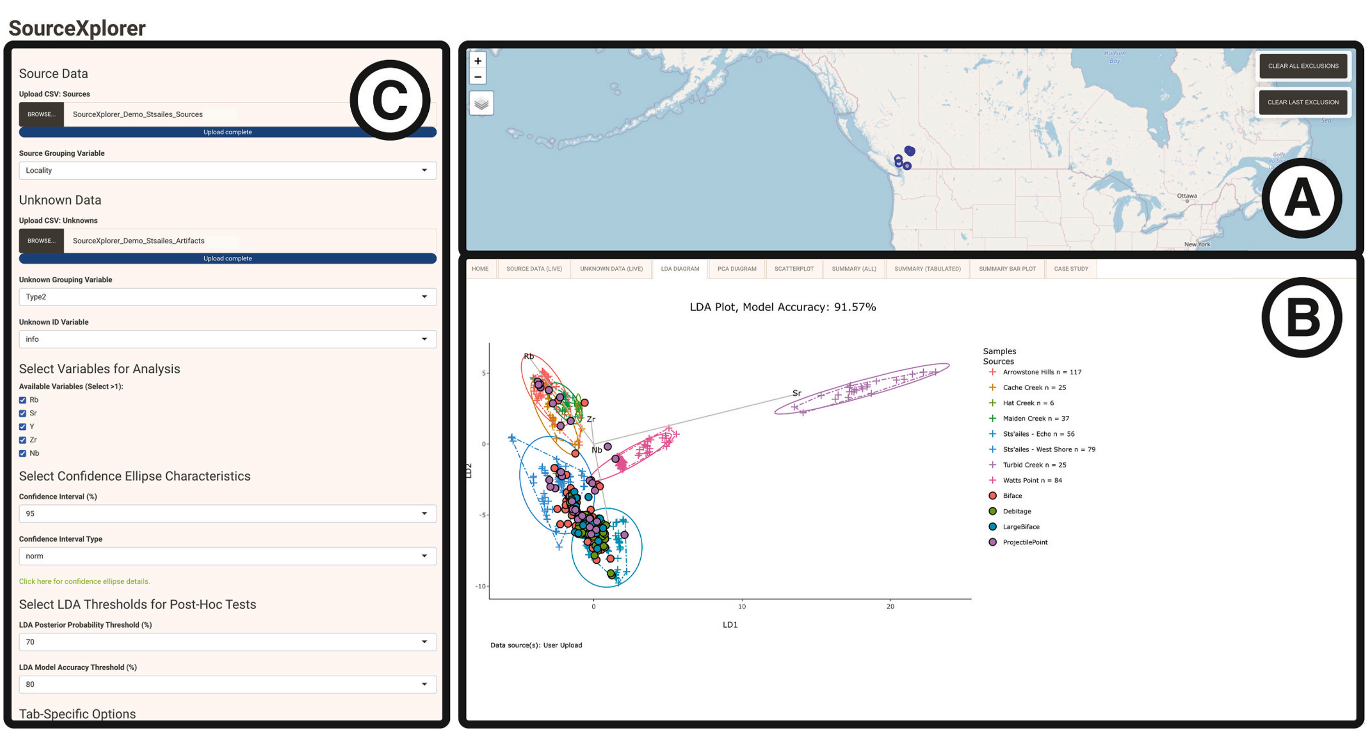


Fig. 1. Screenshot of the SourceXplorer graphical user interface and basic functions. Panel A: map interface, which allows users to control the source locations included in data visualisations and statistical models. Panel B: tab set main panel, which shows data visualisations, statistical outputs, and summary tables and diagrams. Panel C: side bar panel, which allows users to upload unknown (artifact) and source data, select grouping and identification variables from dropdown menus, control which variables are included in statistical models and data visualisations, and download plots and associated tabulated data.

material type (e.g., volcanic glass, fine-grained materials, etc.) within any natural context. The included statistical procedures predict potential source affiliations using a map-based interface and also systematically test and attempt to reject alternate hypotheses of provenance produced by predictive models. The entire application is built in the open-source R programming environment (R Core Team, 2021) using Shiny (Chang et al., 2021), a module for packaging R scripts as online apps with a GUI, which facilitates free and open access to the source code for analytical transparency and replicability. It also provides other researchers the opportunity to further improve, adapt, and advance the program. This is an advancement over other currently available Shiny GUIs (e.g., UC Berkeley's Obsidian Geochemistry Visualisation webpage: https://arf.berkeley.edu/projects/geochemistry), which are intended solely for examining low-dimensional scatterplots of two elemental concentrations and do not guide the user through the interpretation of potential sourcing outcomes.

Here, we demonstrate the functionality, scalability, and various statistical approaches included in SourceXplorer using an archaeological case study from an ancestral Sts'ailes-Coast Salish (Indigenous Salish peoples of the Pacific Northwest) settlement on the Harrison River, British Columbia, Canada. We show how SourceXplorer evaluates elemental concentrations of bifaces and associated debitage collected from the site as well as recently-documented locally-available toolstone materials. Analyzing artifacts and local source materials in tandem, we use SourceXplorer to automatically generate reproducible sourcing outcomes that address three archaeological objectives: 1) distinguishing among the Sts'ailes toolstone source locations as well as other previously-documented sources in the region; 2) determining if bifaces and debitage are composed of similar materials (documenting the manufacture of the bifaces at a possible specialist workspace site); and 3) determining whether the artifacts have compositions consistent with recently identified toolstone outcrops in Sts'ailes or other previously-documented sources in the region. We also establish a geological baseline for the Sts'ailes toolstone locations and provide compositional data to support future regional artifact sourcing studies. The reproducible results produced using SourceXplorer indicate that all of the debitage and a majority of the bifaces are composed of locally-available toolstone, revealing nuanced details about Sts'ailes lithic procurement strategies.

2. Archaeological case study: applying SourceXplorer to ancestral Sts'ailes-Coast Salish toolstone use and biface production

Stone tools and artifacts, only rarely composed of volcanic glass, are the primary artifacts recovered from most ancestral Coast Salish sites. This makes them integral to investigations of past land-use activities (e.g., Graesch, 2007; Sagarbarria, 2017), specialized production and technological change (e.g., Flenniken, 1981; McLaren and Steffen, 2008; Wilkerson, 2017), social interactions and exchange (e.g., Carlson, 1994; Morin, 2012; Rorabaugh and McNabb, 2014; Springer et al., 2018), and ceremonial spheres (e.g., Kannegaard, 2015; McLaren and Gray, 2017; Rorabaugh, 2015). Evaluating where the raw materials used to manufacture the tools originated is a critical but unfortunately infrequent component of such studies. This is because so few non-glass toolstone sources have been identified, characterized, and reported in the region (for exceptions see Reimer, 2018 and Pollock, 2018), at least in part due to the regions' complex geology, which has made sourcing any material type except for volcanic glasses challenging (e.g., Thomas, 2019; Springer et al., 2018).

2.1. Geological background

Our case study takes place in the complicated geological context of the Canadian Cordillera, which represents an amalgamation of Paleozoic to Mesozoic geological terranes—fault-bounded crustal blocks containing distinctive lithologic and stratigraphic units with unique geological histories (Schermer et al., 1984)—that were successively accreted to the North American margin since the Mesozoic. The terranes are composed of volcanic, intrusive, sedimentary, and metamorphic rocks that represent magmatic arcs, microcontinents, and ocean basins. Since the accretion of the exotic terranes, the westernmost portions of the Cordillera have been intruded by the Cretaceous-to Eocene-age igneous rocks of the Coast Plutonic Complex and overlain by relatively young (Quaternary-age) lavas (Colpron et al., 2007). Substantial glacial and glaciofluvial erosion during the Pleistocene has resulted in further geological complexity by producing secondary and tertiary sedimentary contexts that have been shown to contain toolstone material (e.g., Rorabaugh et al., 2015).

The most common and widespread toolstone-grade lithic materials in the region include fine-grained, silica-rich volcanic, volcaniclastic, and clastic sedimentary rocks (e.g., Bakewell, 1996). The toolstone-grade volcanic and volcaniclastic rocks are typically dacitic to rhyolitic lavas and tuffs, respectively, while the clastic and chemical sedimentary rocks are fine-grained, often silicified siltstones and biogenic cherts. Only a small number of these toolstone-grade sources in the region are documented and characterized adequately for inclusion in archaeological sourcing studies (e.g., Reimer, 2018).

While many terranes in the Canadian Cordillera are hundreds to thousands of square kilometers in area (Mahoney et al., 1995), some are much smaller, specifically in southernmost British Columbia. The Nooksack-Harrison terrane (NHt) is one such example and occurs on the west side of Harrison Lake in Sts'ailes territory (Fig. 2). The relatively small size of the southern terranes and their varied geological and tectonic histories has produced a unique geological context and a wide range of potential toolstone sources distributed across the region, suggesting that many toolstone materials found at archaeological sites may have been locally derived from bedrock exposures or retrieved from Quaternary sediments (e.g., Rorabaugh et al., 2015).

The NHt is extensively exposed on the west side of Harrison Lake and contains one of the most intact and well-preserved Mesozoic stratigraphic sequences in the southern Canadian Cordillera (Mahoney et al., 1995). In addition, the Harrison Lake Formation of the NHt is the most extensive Early to Middle Jurassic arc sequence preserved in the Coast Belt in the southern Canadian Cordillera (*ibid.*). The Harrison Lake Formation includes four geological members. The two basal (oldest and lowermost) members are primarily composed of clastic sediments and include conglomeratic rocks of the basal Celia Cove Member and the conformable fine-grained strata of the Francis Lake Member, which include siliceous mudstone, calcareous siltstone, volcanic lithic wacke, crystal vitric tuff, and lapilli tuff. Two volcanic suites superimpose the sedimentary units and include flows and breccias of the Weaver Lake Member and pyroclastic and epiclastic (volcaniclastic) strata of the Echo Island Member. Together, the two younger members consist of a >4 km-thick sequence of felsic (silica-rich) volcanic rocks and sediments, all of which have toolstone potential and cover a large aerial distribution on the landscape. Additionally, the subjacent units, including the Triassic Camp Cove formation, also contain minor components of possible toolstone lithologies. Despite the vast toolstone potential of the NHt, it has not been intensively investigated during archaeological sourcing studies to date.

2.2. Archaeological background

At the Sts'ailes-Coast Salish village of Yā̱çkEtEl (pronounced Yay-ss-ket-el) on the Harrison River in southwestern British Columbia (Figs. 3 and 4), we recovered a total of 164 bifaces and biface fragments, along with thousands of debitage flakes (including an atypically high proportion of biface reduction flakes) adjacent to features identified as large cedar plank houses all dating to ca. 1900–1600 cal BP (Fig. 5; Ritchie et al., 2022). The bifaces and debitage were all recovered from a small point of land protruding into the river. This is the largest collection of

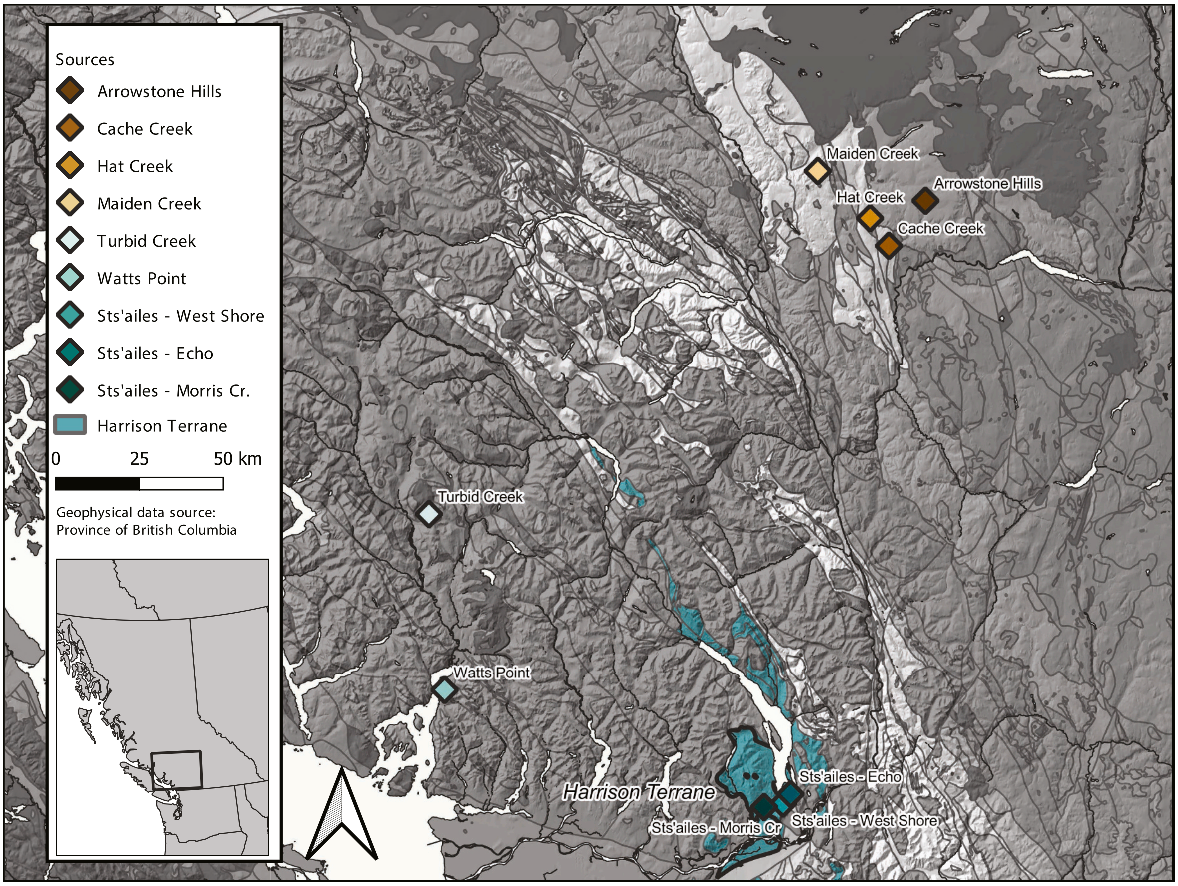


Fig. 2. Geological context of the case study region, including the Nooksack-Harrison terrane ('Harrison Terrane' in the legend) as well as the boundaries of other geological units (in greyscale) and toolstone source locations included in our study. The reader is referred to Cui et al., (2017) for more detailed information about the regional geology.

bifaces with associated debitage so far documented in the region, and it may reflect a specialist's workspace, where large lanceolate bifaces were fashioned before being distributed in ceremonial networks and included in elite burials (Ritchie et al., 2022). Coast Salish specialists were likely embedded within households that consisted of several related family groups that collectively undertook a wide range of domestic and subsistence related activities, as well as more specialized tasks like house construction, canoe-making, and weaving. These specializations brought prestige to both the individual and their household (Ames, 1995).

Yāçkʷ̓EtEl was one of at least three contemporaneous settlements on the middle Harrison River during this period, which were interconnected within a network of other settlement communities both locally and regionally (Ritchie and Lepofsky, 2020). Archaeological evidence recovered from burial mounds at the confluence of the Harrison and Fraser Rivers reveals diverse materials being exchanged between these settlements in the Salish Sea and Interior Plateau at this time, including obsidian, nephrite, abalone, and copper (Blake, 2004). Nearly absent from these other sites, however, is direct evidence of lithic acquisition and specialized production (see Morin, 2012 for exceptions), which is critical for understanding landscape use and exchange relationships between social groups across the region. Yāçkʷ̓EtEl provides a rare opportunity to examine these issues in the Salish Sea and to establish a rigorous methodological approach for tracing the movement of archaeological lithics.

3. Materials and methods

3.1. Application development and statistics

SourceXplorer was developed in the R programming environment using Shiny (Chang et al., 2021) to produce the GUI and to facilitate browser-based user access. Our application includes the ability for users to qualitatively and quantitatively compare the characteristics of unknowns (e.g., artifacts) and natural sources (regardless of the material type or geographical context) with traditional low-dimensional scatterplots as well as more sophisticated multivariate approaches, including linear discriminant analysis (LDA) and principal components analysis (PCA). Users also have access to a guided interpretation of sourcing results via a series of post-hoc tests based on canonical variables produced by both LDA and PCA.

LDA tests are performed using the 'MASS' package (Venables and Ripley, 2002). Prior to LDA analysis, scaling of variables is assessed and transformations are conducted automatically using the package 'caret' (Kuhn, 2020). The accuracy (and robustness) of each LDA model is dynamically assessed using a test/training set of 20/80% of source observations. PCA tests are conducted using 'prcomp' (Becker et al., 1988; Mardia et al., 1979; Venables and Ripley, 2002). The technique is also sensitive to the relative scaling of the original variables, so prior to PCA analysis all variables are automatically scaled by 'prcomp' so that the input data has a mean of zero and a variance of one. Sources included in the multivariate models are selected based on what is visible in the field of view (FOV) of a map interface generated by Leaflet (Cheng et al., 2021), numeric variables included in the models are user-defined in the sidebar, and the interactive data visualisations are generated in the

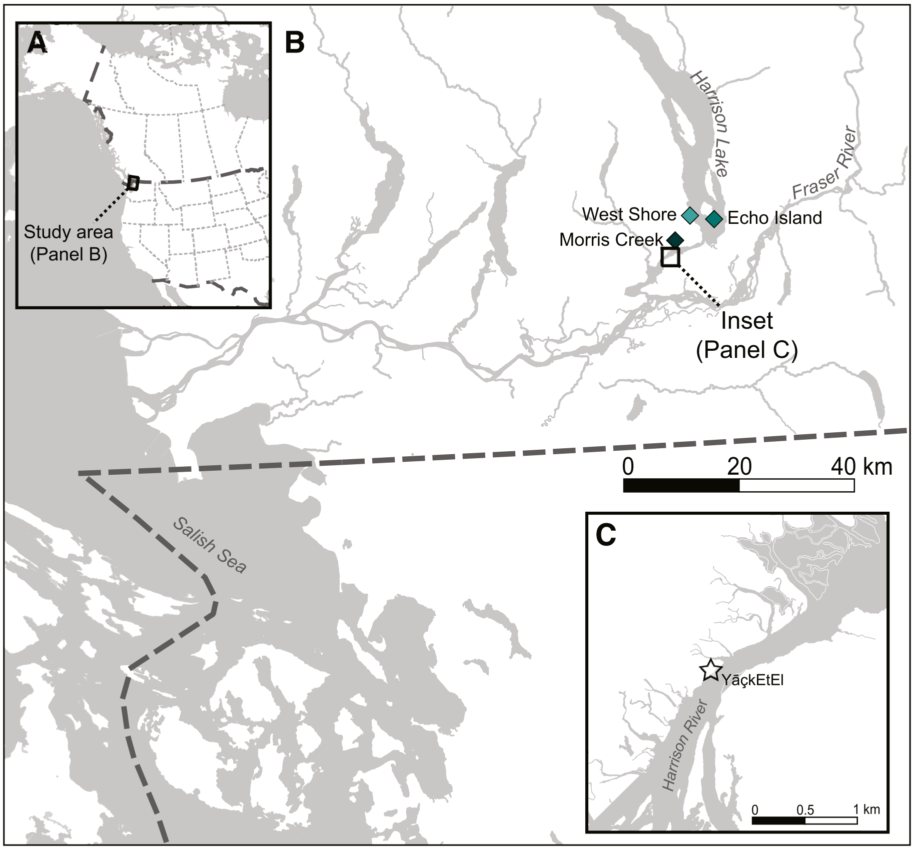


Fig. 3. Location of YäçkEtEl (white star, Panel C) and Sts'ailes toolstone source localities (coloured diamonds, Panel B) situated in western North America (Panel A).

tabset panels of the GUI by Plotly (Sievert, 2020).

Because LDA does not consider the possibility of unknown sources in a model, summing posterior probabilities among all included sources to 100% for each unknown, the technique inherently has the potential to produce false positives. To address this, we have included two post-hoc tests that attempt to reject alternate hypotheses of provenance predicted by LDA, which are identical to the predictions that would be produced by partial least squares discriminant analyses (PLS-DA) using a full PLS model (i.e., including all latent variables; Barker and Rayens, 2003; Brereton and Lloyd, 2014). These post-hoc tests include assessing if the unknown actually falls within the convex hull and/or the confidence ellipse of the predicted source in canonical (LDA and PCA) space. A 'convex hull' is a geometric shape that can be visualised as a string or rubber band stretched around the most extreme points of a population in bivariate space. Effectively, these tests fully automate and quantify the traditional 'matching' of unknowns (e.g., artifacts) to sources using bivariate diagrams, in this case using canonical variables (PC1, PC2, LD1, LD2). Based on the relationship of an unknown to its predicted source population's convex hull and confidence ellipse in LDA and PCA space, three sourcing outcomes are possible:

1. 'Basic' - specimens must fall into either the convex hull or the confidence interval for their predicted source in either LDA or PCA space (so, at least one 'match' overall);
2. 'Standard' - specimens must fall into either the convex hull or the confidence interval for their predicted source in both LDA and PCA space (so, at least one 'match' in both LDA and PCA space); and
3. 'Robust' - unknowns must fall into both the convex hull and the confidence interval for their predicted source in both LDA and PCA space (so, matches for all distribution types in both diagrams).

Specimens that 'fail' the tests above revert to 'Unknown' for the relevant sourcing summary. Users also have the ability to select the type (multivariate-normal, multivariate-t, and Euclidean; default: multivariate-normal, used throughout this study) and confidence interval (68 or 95%; default: 95%, used throughout this study) applied to their confidence ellipses, the metadata for which is reported alongside the sourcing summaries to promote transparency. Users may also view the raw data for the most-heavily weighted variables in the LDA and PCA models in low-dimensional space to further support their results, although this diagram is not considered by SourceXplorer during the production of sourcing summaries. These tests allow a more robust and transparent application of the "provenance postulate" (attributed to Weigand et al., 1977) by taking the most likely alternate hypothesis of provenance and attempting to reject it using the location of the unknown specimen relative to the distribution of predicted source observations in canonical space. This postulate holds that all potential sources must be scrutinized and sequentially excluded ideally until one (or the null hypothesis of no known affiliation) remains (see Rapp and Hill, 1998; Neff, 2000; Wilson and Pollard, 2001; Pollard et al., 2007).

However, the convex hull and confidence ellipse tests do not address whether more variability exists within, rather than among, potential sources. If more intra- than inter-source variability exists in a model, viable source affiliations cannot be identified (Neff and Glascock, 2003). In other words, the tests do not address the possibility of overlapping sources and, thus, cases in which more than one source cannot be rejected due to similarities among populations. To address this challenge in part, SourceXplorer requires the LDA posterior probability for unknowns, a statistic indicative of how distinct the predicted source is from others in the model, to be greater than a user-selected threshold value to 'pass' (default value of 70%, a value significantly greater than 50%, which would be equivalent to equal probability in a two-source model). If the posterior probability is less than the given threshold, it will 'fail', and the sourcing outcome will revert to 'Ambiguous' (suggesting that a

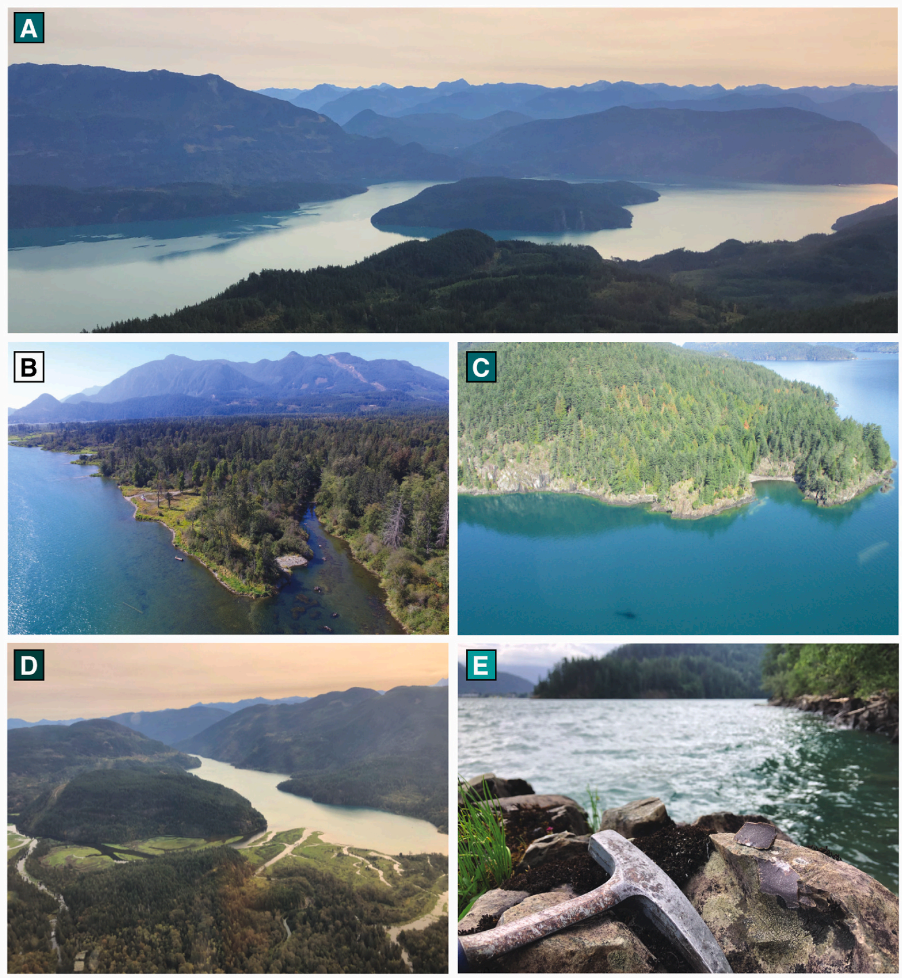


Fig. 4. Photographs of Yāçk̓EtEl and toolstone source localities in Sts'ailes territory. Panel A: aerial helicopter photograph of Echo Island looking southeast from above the west shore of Harrison Lake. Panel B: aerial drone photograph of Yāçk̓EtEl facing downstream (southwest). Panel C: aerial helicopter photograph of the south end of Echo Island (facing north), with toolstone outcrops easily accessible on the east side of the small cove. Panel D: aerial helicopter photograph of Morris Creek facing southeast. Panel E: toolstone exposures at the waterline along the West Shore facing south, with Harrison Hot Springs visible in the top left corner and the head of the Harrison River visible in the top middle of the photograph. All photographs by the authors.

specimen likely falls within the range of more than one source). Furthermore, we added another similar user-selected threshold value for the accuracy of the applied LDA models. If the model accuracy is below the threshold (default value 80%), it will fail and also revert sourcing outcomes to 'Ambiguous' for all unknowns in the model. This helps users navigate challenges associated with source sampling procedures and the robustness of applied source data libraries. The posterior probabilities, model accuracy, and user-selected thresholds are all presented in the 'Summary (All)' table to promote maximum reproducibility.

In short, the convex hull and confidence ellipse tests in SourceXplorer assess if an unknown can be considered part of a predicted source population, the probability threshold test assesses in part how distinct the predicted population is from others in the model, and the LDA accuracy threshold test assesses how well unknowns can be attributed to the 'correct' source in the model, satisfying many of the conditions and assumptions of the "provenance postulate" in an automated fashion. This series of post-hoc tests effectively helps users quickly, easily, and robustly assess the validity of their source identifications; it counteracts potential false confidence regarding source assignations that would not survive closer scrutiny, although even sourcing outcomes produced using this rigorous approach are inherently not absolute.

3.2. Application use and deployment

SourceXplorer can be accessed either through RStudio in package format (https://github.com/RhyMcMillan/SourceXplorer; see www.sourcexplorer.org for installation instructions) or directly via a web browser (https://sourcexplorer.shinyapps.io/SourceXplorer/). The most current form of the application will be permanently available at the above URLs, and older versions of the application will also be available at https://github.com/RhyMcMillan/SourceXplorer. The code and other development information are released under the open [MIT] license, permitting free use, duplication, alteration, and distribution as long as credit is given. A deployed version of the app can be accessed through any Internet browser. An Internet connection is required to produce the Leaflet base map in both situations, although the statistical functions still operate in absence of this attribute in the RStudio version.

Two types of data can be included in the statistical models and data visualisations in SourceXplorer: 1) user-uploaded source data that can include any numeric variables and must be accompanied by spatial information (latitude and longitude; SourceXplorer will automatically add 'dummy' coordinates if these columns are not present) and 2) user-uploaded unknown (e.g., artifact) data including numeric variables with the same headers as the source data. The unsupervised PCA and supervised LDA models as well as sourcing outcomes generated by SourceXplorer each include any numeric variables selected by the user that are shared between the uploaded source and unknown datasets. In the interests of data privacy, uploaded user data—both unknown and source—are not stored in the app beyond the user's immediate session, nor shared with any party. For users who would like to explore relationships among unknowns without comparing them to potential natural sources, they can be uploaded instead as source data (a source data file is required to run the app).

SourceXplorer was designed to function solely as a statistical tool that facilitates automated and guided interpretations of the relationships

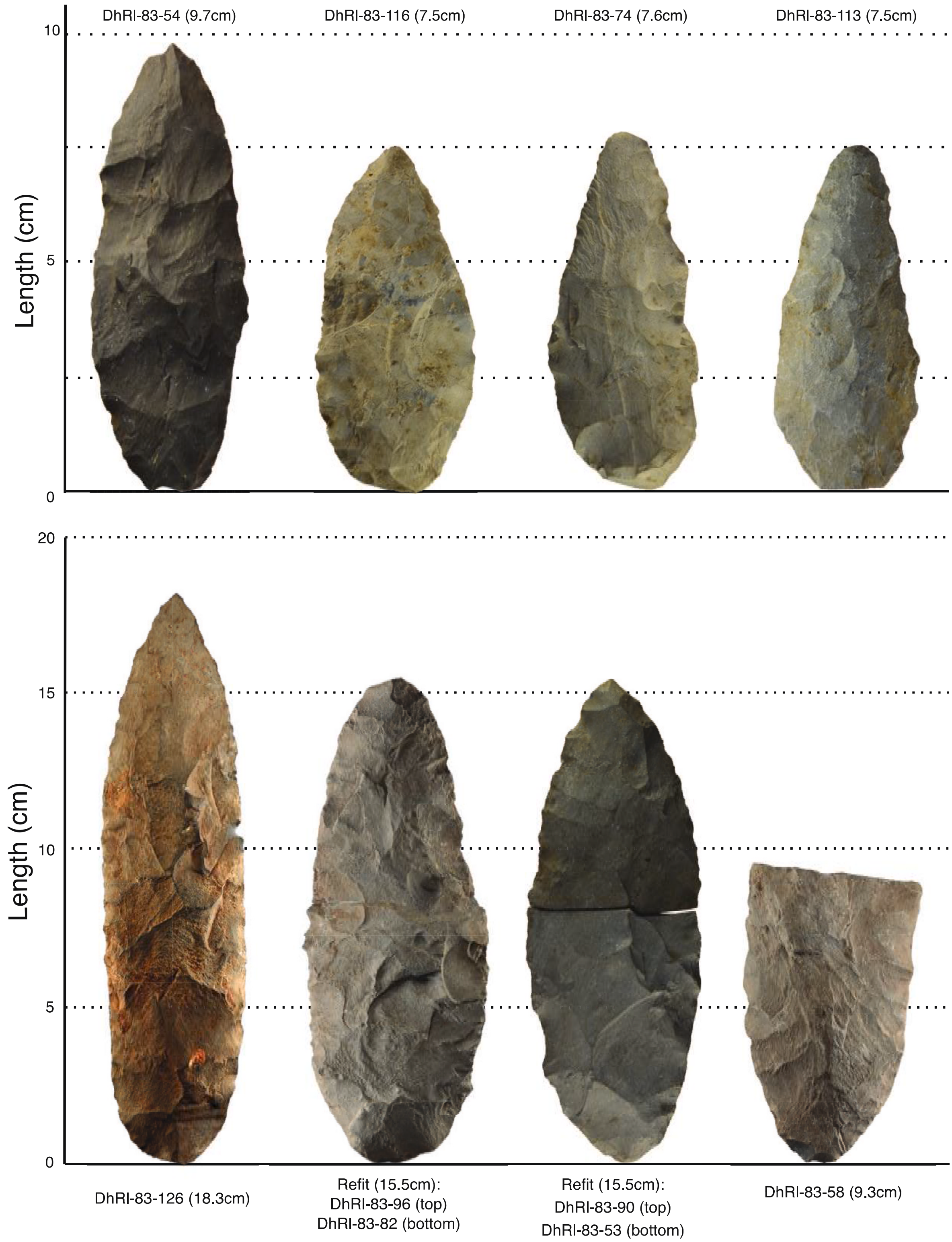


Fig. 5. Photographs of select leaf-shaped (top) and large (bottom) bifaces analysed in this study from YãçkEtEl. Note how some of the specimens exhibit sedimentary structures and sulphide mineral vacancies in their weathered surfaces (e.g., DhRl—83:116 and DhRl-83:74). Modified from Ritchie et al. (2022).

among unknowns and natural sources, not as a database. As a result, users must provide their own baseline source data for use in the application (case study data is available through a download link in the app), and we have built in automated functions that help users assess and quantify the quality of the applied multivariate models and to communicate them transparently. Importantly, SourceXplorer still relies on robust libraries of toolstone sources as much as any lithic provenance analysis does, although it helps users quantify model accuracy and provides all metadata relevant to reproducing and justifying the generated sourcing outcomes.

The GUI includes three main panels (Fig. 1): Panel A) the map interface showing the distribution of source specimens included in the models, Panel B) the main tab set panel, which includes a series of tabs that the user can select to show different outputs (e.g., LDA diagrams or sourcing summary tables), and Panel C) the sidebar panel, which allows users to upload their source data, control the numeric variables included in their statistical models, and select the non-numeric identifying and grouping variables for their unknowns and the grouping variable for their uploaded source data (both of which are also required to run the app), among other user-defined features. Templates are provided to the user via the app GUI into which data can be easily pasted and subsequently uploaded. Categorical and numeric data for both unknowns and source materials can be added in as additional columns in the downloadable templates. Any numeric variables shared between the unknowns and sources can be included in statistical models and resulting sourcing outcomes using a checkbox in the sidebar panel. Note that any incomplete rows in user-uploaded data will be removed before inclusion in SourceXplorer, and all remaining rows containing data passed to the statistical assessments after processing (including filtering of source data by the map FOV) are shown in the 'Source Data (Live)' and 'Unknown Data (Live)' tabs.

Once data has been uploaded, users can use dropdown and checkbox menus in the sidebar panel to select numeric and categorical variables, and the app predicts relationships among unknowns and potential sources automatically using an LDA model that includes all source locations present within the bounds of the map interface (with the option to exclude one or more specific sources by clicking on the corresponding map marker; the accuracy of the current LDA model (as assessed by an 20/80% test/training set) is shown as % correct in the plot title as well as in the 'Summary (All)' tab). Simultaneously, a PCA is conducted on the same set of observations. Using the main tab panel, users have the option to view their uploaded raw concentration data and canonical variables produced by both LDA and PCA in simple bivariate plots. Using the sidebar panel, users can download the PCA and LDA scores, weightings (loadings, scalings), and coordinates for the convex hulls and confidence ellipses presented in the diagrams for use in other data visualisation software (e.g., DataGraph; Visual DataTools, Inc., 2020). Users can also download the displayed diagrams using the sidebar panel or directly in the main tab set panel using the Plotly interface (Fig. 6 shows an example of data visualised using all three outputs).

The outcomes generated by the series of post-hoc tests described, above, are presented in the 'Summary (All)' tab, tabulated in the 'Summary (Tabulated)' tab, and visualised in a bar plot in the 'Summary Bar Plot' tab. In the 'Summary (All)' table, the sourcing summaries described above ('Basic', 'Standard', and 'Robust') and related metadata (post-hoc test results, number of sources in the models, number of source observations in the models, the variables used in the models, etc.) are displayed. In the 'Summary (Tabulated)' tab and 'Summary Bar Plot' tab users can select which sourcing summary (i.e., 'Basic', 'Standard', and 'Robust') is tabulated and plotted using a dropdown menu in the sidebar panel. Both output tables, in addition to the data used to create the PCA and LDA diagrams, can be downloaded in comma delimited format (csv) for use outside of SourceXplorer.

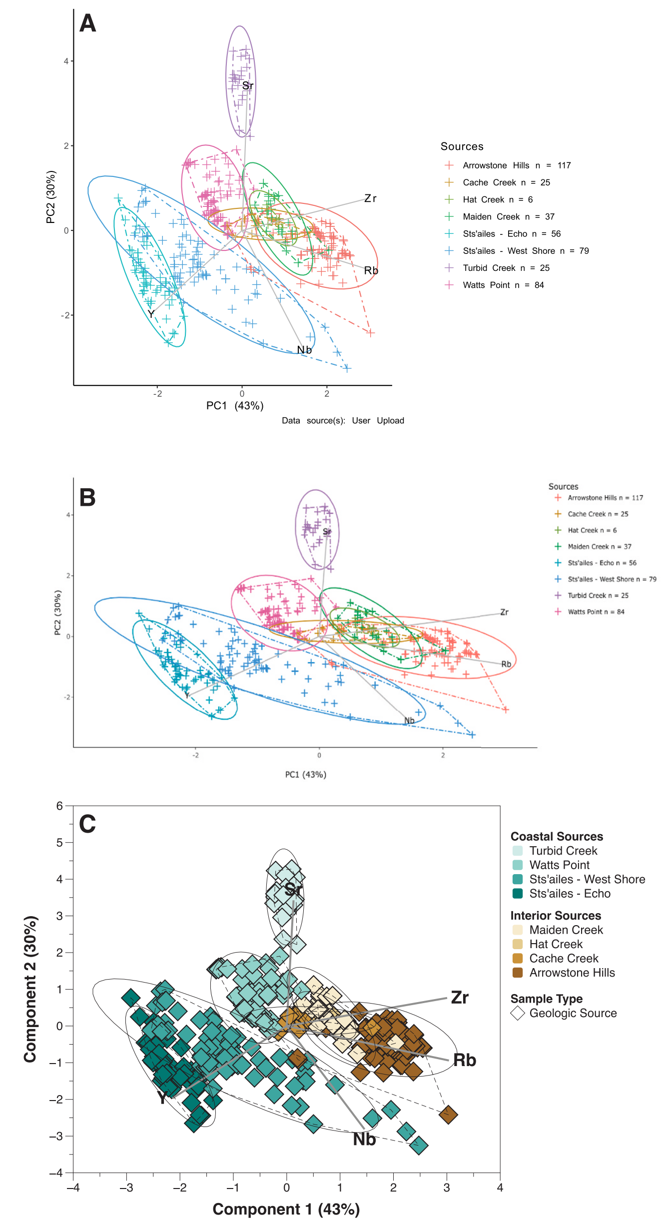


Fig. 6. Comparison of different outputs available in SourceXplorer. Panel A: raw SVG downloaded from the side panel. Panel B: raw PNG downloaded from the Plotly interface in the tab set main panel. Panel C: diagram generated in the plotting software DataGraph using the tabulated data exported from the sidebar panel.

3.3. Data collection

We conducted elemental analyses of 164 bifaces (34 of which are projectile points) and 52 bifacial reduction flakes from Yāçkʼ EtEl as well as 42 geological specimens from three toolstone localities near the site. We also analysed representative geological specimens from other documented volcanic toolstone source locations in southwestern British Columbia: Arrowstone Hills ($n = 3$), Hat Creek ($n = 2$), Maiden Creek ($n = 1$), and Watts Point ($n = 10$) and combined those with previously-documented data for the same sources presented in Pollock (2018). In total, 429 source observations from eight source localities were used in our study. Our methods for the elemental analyses are described below, and the included source locations are shown in Fig. 2.

Elemental concentrations were collected from case study source specimens and artifacts with pXRF following the procedure outlined in McMillan et al. (2019). Trace element concentrations were collected non-destructively with an Olympus Vanta C-Series pXRF instrument. Analyses were carried out using the 'GeoChem-extra' method with the factory calibration (i.e., no user factors were applied to the resulting concentrations) and fundamental parameters calibration. 'GeoChem-extra' mode varies the current and voltage of the 4-W X-ray tube (with a Rh anode) in combination with two built-in beam filters to improve the fluorescence of both lighter and heavier elements within a single analytical run. Spectra were collected for 60 s per analysis (30 s on each beam). Source specimens were each analysed three to 10 times to ensure reproducibility and to evaluate intra-specimen heterogeneity, and artifacts were analysed in at least triplicate and averaged prior to interpretation.

During the analysis of source specimens and artifacts, potential instrument drift was monitored using a certified standard from the US

National Institute for Standards and Technology (NIST): SRM 2711a, Montana Road Dust (Mackey et al., 2010). No drift was observed during any analytical sessions, and the trace element compositions of NIST 2711a were consistently within <5% (relative) of expected concentrations. In addition, the accuracy of our results was cross-checked by analysing the same materials (e.g., volcanic glasses) for trace element concentrations with ICP-MS and the pXRF used in this study (McMillan et al., 2019) as well as analysing the Peabody-Yale Reference Obsidian (PYRO) calibration and check standards (Frahm, 2019a, 2019b). The cross-validation by ICP-MS showed that most trace elements were consistently within <5% (relative) when measured by both analytical techniques. Linear regressions of the PYRO calibration standards (n = 20) showed a very good agreement among the measured and expected values (averaged formula for 5 linear regression models: y = 1.02x-3.02, R^2 = 0.98; Table 1), with an average relative difference of <10%. The PYRO check set (n = 15) both before and after calibration showed that the elements of interest in this study (Rb, Sr, Y, Zr, and Nb) were also on average within ~11% (relative) of expected values. When assessing if applying the calibration scheme to the check specimens (n = 15) significantly increased the accuracy of measured values, results showed that the calibration increased the average relative difference between measured and expected concentrations by only ~1%. Calibration also produced a slightly less ideal regression between measured and expected values for the calibrated check set compared to the uncalibrated check set (averaged formula for five linear regression models produced by calibrated check standards: y = 0.91x + 6.53, R^2 = 0.99 and uncalibrated check standards: y = 0.93x + 3.98, R^2 = 0.99; Table 1). As a result, no post-hoc calibration or adjustment of the parameters from the default settings was deemed necessary for the elements of interest in this study (as observed in Frahm, 2017).

3.4. Sts'ailes data analysis in SourceXplorer

Sts'ailes source and artifact data were uploaded into SourceXplorer via the user interface following the sequence described below (all data can be downloaded from the 'Case Study' tab in SourceXplorer or from https://github.com/RhyMcMillan/SourceXplorer/tree/master/data). We explored the data and exported PCA and LDA results for three model types: 1) only Sts'ailes sources at various resolutions/groupings; 2) Sts'ailes and regional sources at various resolutions/groupings; 3) Sts'ailes and regional sources at various resolutions/groupings and associated artifact predictions. All data for regional sources were uploaded and evaluated, then the artifact data was subsequently uploaded and explored.

With only the source data uploaded, the map interface FOV was used to exclude all sources except those from Sts'ailes (model 1; Fig. 1, panel A), and different groupings were evaluated using the dropdown menus in the side bar (Fig. 1, panel C). Results were examined in SourceXplorer using the tab set panel (Fig. 1, panel B), then exported in tabulated format for subsequent processing using the options in the sidebar (Fig. 1, panel C). Then, the map FOV was expanded to include all uploaded regional sources (model 2; Fig. 2), and the same procedure was followed as described above for the Sts'ailes sources. Once the results for the source models were exported using the sidebar panel, the artifact data were uploaded and compared to all sources in the model (model 3). The sourcing summaries and LDA and PCA data were extracted for the artifacts with source observations organised into two different resolutions/groupings: one with all Sts'ailes sources and interior sources grouped together (Sts'ailes – All, Interior-All) and one with Echo Island and West Shore/Morris Creek as well as the interior sources separated into distinct populations. The exported results were then used to create composite diagrams in the data visualisation software DataGraph (Visual DataTools, Inc., 2020).

4. Results

4.1. Yāçk̓EtEl bifaces and debitage: context

The large lithic assemblage from Yāçk̓EtEl includes 164 bifaces, 1002 artifacts indicative of a range of activities (e.g., hunting, fishing, woodworking), and 13,940 pieces of debitage. Our excavations indicate that the majority of recovered debitage and bifaces came from a discrete workspace on the edge of the river, adjoining a plank house. Radiocarbon ages of the workspace layer correspond with the earliest occupation of this house-site (ca. 1900–1600 cal. BP; Ritchie et al., 2022), supporting the inference of simultaneous occupation and use. Over time, the settlement expanded in size, and a palimpsest of occupational debris covered over the earliest plank house surfaces and the lithic workspace.

Lithic analysis of the assemblage associated with the Yāçk̓EtEl workspace indicates the lithic specialist(s) first trimmed large core-blanks at the source locale and then further reduced them at the workspace in the production of large thin bifaces (Fig. 5, bottom panel). If these broke during manufacture, they were often refurbished into smaller foliate bifaces (Fig. 5, top panel). Larger flakes removed from this process were made into bifacial knives and projectile points. Associated debitage indicates multiple reduction stages, including edging and thinning, reinforcing the inference of biface manufacturing (Ritchie et al., 2022). Using SourceXplorer to examine relationships among the trace element compositions of the different artifact types allowed us to quantitatively evaluate the likelihood of these inferences.

4.2. Sts'ailes toolstone source identification and composition

During the 2018–2021 field seasons, three toolstone source localities were identified around southern Harrison Lake and the upper Harrison River, within the traditional territory of the Coast Salish-Sts'ailes (Figs. 2–4). Located along the primary waterways of the territory and situated in close proximity to large settlement areas, these source locations were easily accessible and regularly used by the Sts'ailes. We have identified habitation areas and pictographs nearby to each of the source locations, reflecting the social significance of these places, and also their use for myriad activities including seasonal fishing and hunting. Large quantities of possible lithic reduction scattered over hundreds of meters at these source localities suggest that the Sts'ailes may have been using the toolstone material for millennia. In this context "source locality" is used to denote multiple outcroppings of similar rock within close proximity of each other. The three Sts'ailes toolstone source localities—although separated from one another by several kilometers—are

Table 1
Comparison of linear regression models produced for the trace elements of interest in this study.

Model	Calibration Standards			Uncalibrated Check Standards			Calibrated Check Standards		
	Slope	Intercept	R^2	Slope	Intercept	R^2	Slope	Intercept	R^2
Rb	1.04	−1.19	0.96	0.99	3.35	0.99	0.96	4.49	0.99
Sr	0.97	2.62	0.99	0.98	3.09	0.99	1.02	0.43	0.99
Y	1.06	−4.07	0.98	0.91	2.08	0.99	0.86	5.58	0.99
Zr	1.07	−6.59	0.98	0.95	11.65	0.99	0.89	17.53	0.99
Nb	0.98	−5.83	0.98	0.82	−0.28	0.98	0.84	4.63	0.98
Average	**1.02**	**−3.02**	**0.98**	**0.93**	**3.98**	**0.99**	**0.91**	**6.53**	**0.99**

on average rhyo-dacitic in composition and have comparable major, minor, and trace elemental compositions (Ritchie et al., 2022), likely because they all appear to have similar lithologies and originate from the same geological formation. This toolstone material can be characterized as waterlain and metasomatized volcanic ash (tuff; Arthur et al., 1993), which appears to have been silicified (metasomatized) into 'cherty' toolstone materials that exhibit both macro- and mesoscopic sedimentary structures, sub-mm accessory sulphide minerals (Fig. 7), as well as elemental signatures consistent with volcanically-derived materials. As a result, we compare both Sts'ailes artifacts and source materials with other documented sources of fine-grained toolstones (e.g., extrusive volcanics) in the region, including the Watts Point dacite source (Reimer, 2018) and other more trachytic extrusive volcanic sources in the interior of British Columbia (e.g., Arrowstone Hills, Maiden Creek, all part of the Kamloops Group Volcanic Suite (e.g., Ewing, 1981), hence their compositional similarity and grouping at the 'territory' resolution) in attempts to reject the null hypothesis that the materials used to manufacture the investigated artifacts were locally-derived.

4.3. *YāçkEtEl bifaces and debitage: sourcing outcomes*

4.3.1. *Relationships among Sts'ailes and other regional sources*

The two localities on the West Shore of Harrison Lake and Morris Creek can be confidently separated from the materials obtained from Echo Island when investigated using both LDA and PCA (Fig. 8; 95% confidence ellipses for each source only overlap slightly). This is likely due to outcrops at the source localities exposing different parts of the Harrison Lake Formation stratigraphic sequence or representing more proximal or distal deposits of the same units. Specimens obtained from the West Shore and Morris Creek cannot be easily differentiated from one another, however, suggesting they are composed of similar, if not the same, lithofacies or components thereof (Fig. 8). As a result, the populations from the West Shore and Morris Creek were combined for this study at the locality resolution. This allows sourcing outcomes to be assigned to the West Shore (including Morris Creek) and Echo Island

Fig. 7. Field and laboratory photographs of Sts'ailes source materials and specimens. Panel A and B: bluish cherty volcaniclastics from Echo Island. Panel C: specimen from Echo Island showing relationships among cherty and courser-grained facies exhibiting soft sediment deformation. Panels D and E: bluish black and purplish cherty volcaniclastics from the West Shore. Panel F: outcrop and float specimens of volcaniclastics from Morris Creek. Photo credit: Kaitie Purdue (panels A and B) and R. McMillan (remaining panels).

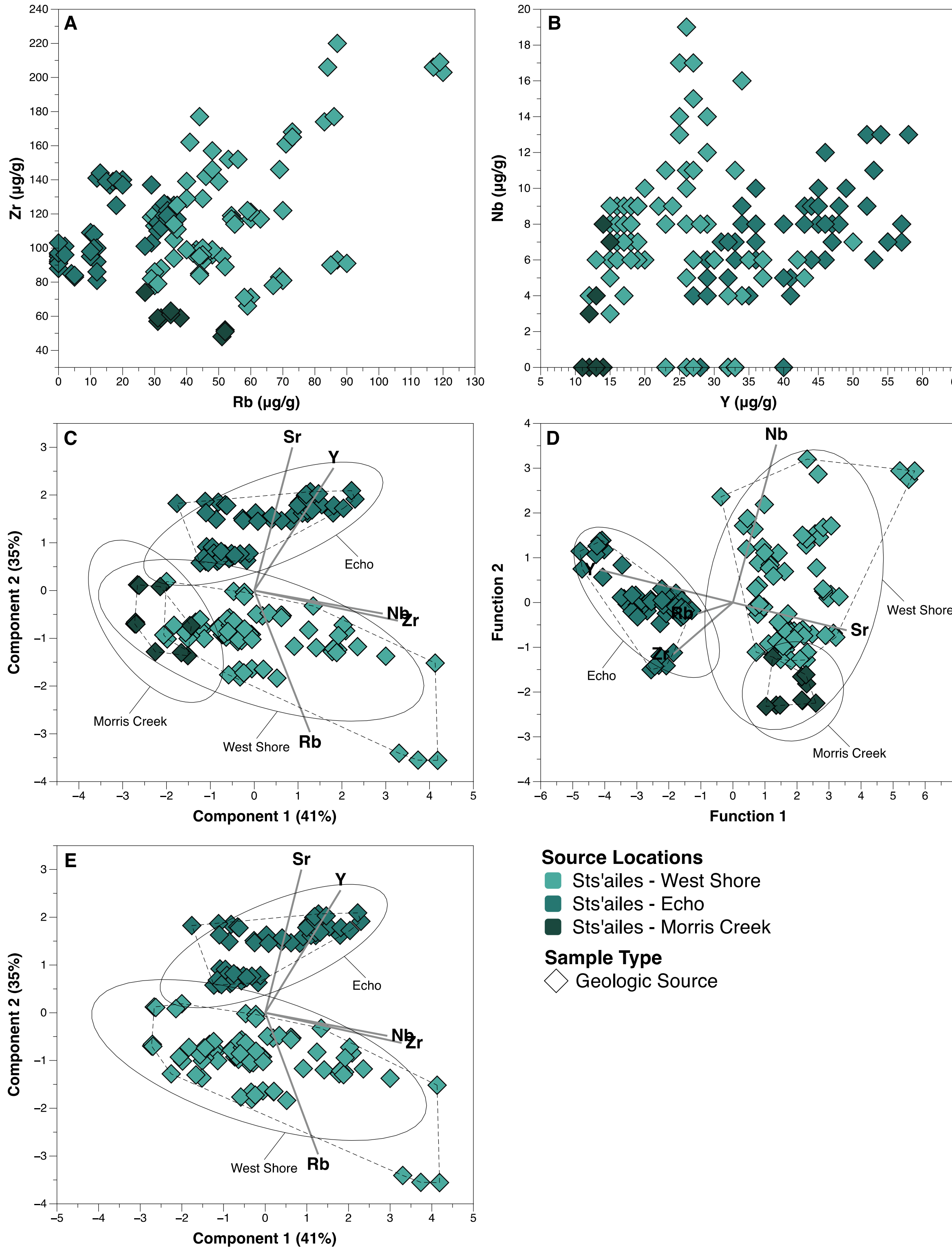


Fig. 8. *Comparison of different toolstone source localities identified in Sts'ailes, showing the compositional similarities between geological specimens collected from the West Shore and Morris Creek and the differences among those sources and the specimens obtained from Echo Island (see Ritchie et al. (2022) for discrimination diagrams showing major and minor elemental concentrations and ratios).Panels A and B: trace element concentrations. Panels C and D: PCA (Panel C) and LDA (Panel D) models with Morris Creek separated from the West Shore. Panel E: PCA model with Morris Creek grouped with West Shore. Similar results are obtained via LDA, although only one discriminant function is produced when only two groups are included in the model, which does not allow the results to be displayed in bivariate space. Grey lines = variable loadings/scalings; black lines = 95% confidence ellipses (multivariate-normal); black dashed lines = convex hulls. All diagrams were generated in DataGraph using tabulated data exported from SourceXplorer.*

(more distant from Yāçk̓EtEl) at the locality resolution, as well as to the territory resolution when all Sts'ailes source localities are combined (Sts'ailes – All). As mentioned above, due to their similarities in canonical space, we also grouped the sources from the interior of the province to avoid low posterior probabilities and sample sizes at the territory resolution. These two spatial resolutions were used in tandem to assess the potential for low posterior probabilities resulting from the compositional similarities of the two Sts'ailes source localities and interior source localities compared to the larger differences among regional sources. Additionally, Sts'ailes toolstone outcrops differ significantly from most other documented sources in the region based on trace element compositions (Fig. 9).

4.3.2. Relationships between formed tools and debitage

The sourcing results produced by SourceXplorer (Table 2, Figs. 10–12) show that the debitage and formed tools are mostly geochemically consistent with each other, indicating the debitage is likely the byproduct of the manufacture of the bifaces (including the large bifaces; Fig. 11). Our results also show that all of the analysed debitage pieces ($n = 52$; 100%) and most of the formed tools ($n = 146$; 89%) were predicted by LDA to originate from the source materials collected from Sts'ailes (at the territory resolution). Post hoc tests show that all but one of the formed tools predicted to Sts'ailes ($n = 145$) pass the standard sourcing summary and all but eleven pass the robust sourcing summary ($n = 135$) at the territory resolution, with the remainder considered 'ambiguous' or 'unknown' (Table 2). Interestingly, two of the artifacts which passed the standard but failed the robust summary, 007 and 129, are extremely low in trace element concentrations and are likely composed of a different material type than the Sts'ailes source materials, even though they still fall within the 95% confidence ellipse for Sts'ailes sources in both LDA and PCA space, documenting the value of having multiple lines of evidence (e.g., McMillan et al., 2019) and levels of post-hoc assessment scrutiny. Thus, 135 of the 164 formed tools are completely consistent in canonical space with Sts'ailes sources (robust summary) at the territory resolution. Specifically, the debitage is only consistent with the Sts'ailes sources, and all 52 debitage pieces pass both the 'basic' and 'standard' sourcing summaries at the territory resolution, with only one debitage specimen reverting to 'unknown' for the 'robust' summary (Figs. 10 and 11; Table 2). When investigated at the locality scale (separating Echo Island, West Shore/Morris Creek, and all the interior sources) and using the 'basic' sourcing summary to avoid unknowns falling between the two Sts'ailes source populations or those within the range of one locality but predicted to the other reverting to 'unknown' or 'ambiguous' (a result of the similarities between the two source localities in regional LDA models), the debitage appears to primarily originate from Echo Island ($n = 44$), with the West Shore/Morris Creek less represented ($n = 6$). With the two source localities separated in the model, only two debitage specimens are determined as "ambiguous" (basic summary) due to low posterior probability resulting from similarities between the two source localities, and none are identified as unknowns at the same level of scrutiny (basic summary). Interestingly, the opposite trend is apparent for the formed tools at the locality resolution, with most being attributed to the West Shore/Morris Creek locations ($n = 92$ basic pass, $n = 88$ standard pass) and relatively fewer to Echo Island ($n = 49$ basic pass, $n = 9$ standard pass), although most artifacts plot within the general ranges of but between the two localities, which are directly adjacent to one another in canonical space (Figs. 8–10).

4.3.3. Relationships among formed tools and Sts'ailes sources

Nearly all the formed tools ($n = 135$) are geochemically consistent with Sts'ailes toolstone sources at the territory resolution (robust pass), and the basic summary reveals that 49 of those predicted to Sts'ailes sources at the locality resolution are more consistent with the Echo Island source location (<15 km 'as-the-crow-flies' from Yāçk̓EtEl) and 92 are more consistent with the West Shore/Morris Creek source location (<10 km 'as-the-crow-flies' from Yāçk̓EtEl). Only one of the bifaces (excluding projectile points) was attributed to sources outside of Sts'ailes and passed both the standard and robust sourcing summary at the territory resolution, although that specimen did not pass the same post-hoc tests at the locality resolution.

However, the 34 projectile points recovered from the site exhibited more compositional variability: ten were attributed to the Interior sources (standard pass, territory resolution, and only nine pass the robust summary; Arrowstone Hills is ca. 200 km 'as-the-crow-flies' from Yāçk̓EtEl) and one specimen was attributed to Watts Point (ca. 100 km 'as-the-crow-flies' from Yāçk̓EtEl) at both spatial resolutions (standard and basic summaries pass, but robust summary fails; Table 2, Figs. 10–12). The remaining 25 specimens were either attributed to Sts'ailes sources ($n = 19$ for standard pass, $n = 18$ for robust pass at the territory resolution; $n = 18$ basic pass at the locality resolution) or predicted to non-local sources and determined to be ambiguous or unknowns.

5. Discussion

5.1. Benefits of the SourceXplorer statistical framework

By providing an accessible, guided method by which user data can be incorporated into dynamic supervised and unsupervised multivariate statistical models, the archaeological community and associated stakeholders, including those inexperienced in artifact provenancing, can conduct their own sourcing studies as well as assess the outcomes of lithic analyses provided by other researchers and commercial laboratories (provided the appropriate steps have been taken to ensure that the datasets are compatible). This permits the cross-validation of sourcing reports as well as the ability for non-specialists to conduct their own independent investigations. It also allows researchers and stakeholders without specialist knowledge to capitalize on the "economy of scale" afforded by the pXRF revolution. For example, with appropriate certification and a rented pXRF, one can test hundreds of artifacts for the same budget as commissioned work on a few dozen artifacts in a specialized analytical laboratory. This dramatically expands the volume of potential observations from individual artifacts to entire assemblages, and SourceXplorer provides a measure of guidance and both quantitatively robust source predictions and critical assessments of those assignations. While SourceXplorer cannot duplicate the accrued knowledge, skill, and experience of an expert, its robust statistical analyses and intuitive interface can improve the products of non-experts in a way that is at least as robust as what a 'sourcing laboratory' typically does by making a sound and transparent procedure readily accessible. Furthermore, the statistical procedure facilitated by SourceXplorer is not only applicable to archaeological challenges and elemental concentrations; it could also be readily applied to other research questions based in the natural sciences (e.g., geological provenancing) and using other numeric variables (e.g., isotopic ratios).

We selected R for several reasons. The system provides a powerful statistical framework with several sophisticated modules designed specifically for multivariate analysis; the programming environment facilitates nuanced design and adaptation tailored specifically to geoarchaeological applications; and the open-source nature of R serves both to eliminate financial restrictions from using our curated procedure and to permit – and indeed encourage – methodological transparency. Open-source tools, such as R, are a core component of the Open Science movement and philosophy, which emphasizes maintaining accessibility to tools and data and ensuring replicability of results by using free open-source software and open-license scripts whenever possible (for further discussion see Marwick 2016; Marwick et al., 2017, 2018). This contrasts with many commercial software options and GUIs, which often have expensive user licenses (often out of reach for researchers outside larger institutions) and opaque settings and processes (e.g., SYSTAT, Trueman et al., 2006; McMillan et al., 2017; Fordisc, Pilloud et al.,

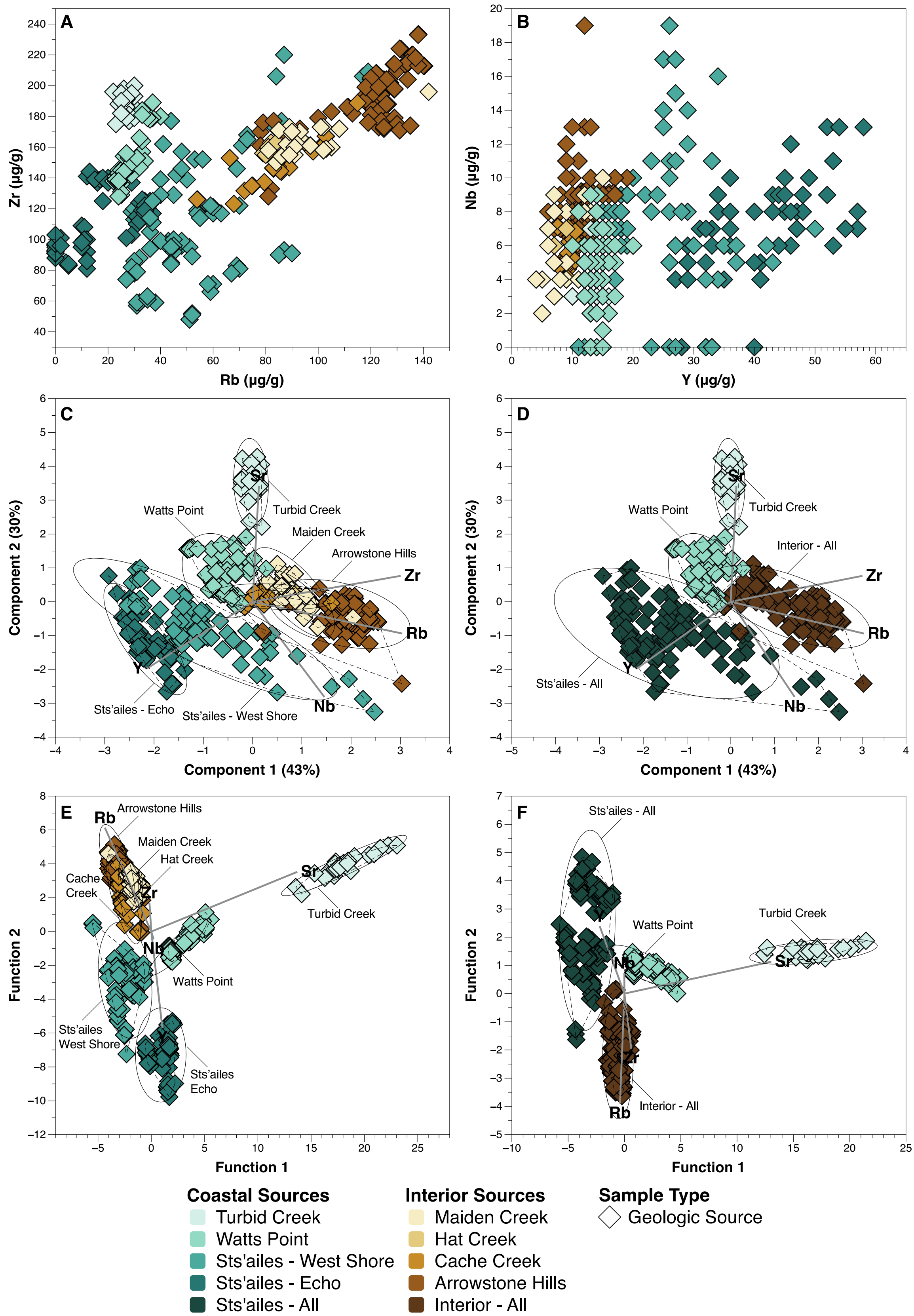
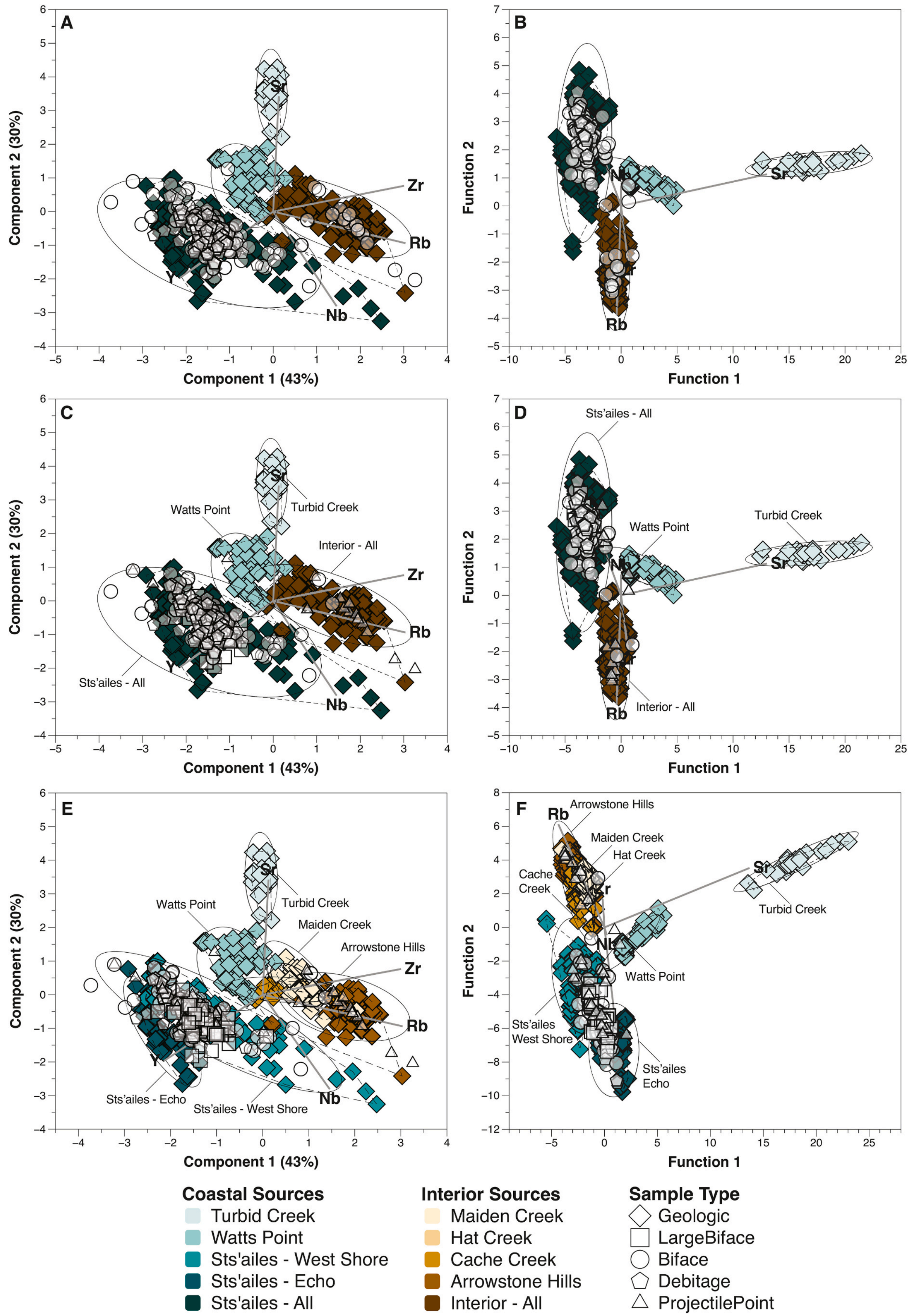


Fig. 9. *Comparison of Sts'ailes sources to other documented toolstone sources in the region. Panels A and B: trace element concentrations. Panels C and D: PCA models shown with Sts'ailes localities grouped at the locality (Panel C) and territory (Panel D) resolutions. Panels E and F: LDA models shown with Sts'ailes localities grouped at the locality (Panel E) and territory (Panel F) resolutions, both of which yield > 91% of cases correctly re-assigned using a test/training set of 20/80% of observations in the models. Grey lines = variable loadings/scalings; black lines = 95% confidence ellipses (multivariate-normal); black dashed lines = convex hulls. All diagrams were generated in DataGraph using tabulated data exported from SourceXplorer.*



(caption on next page)

Fig. 10. Comparison biface and debitage specimens (white translucent markers) to Sts'ailes sources and other documented toolstone sources in the region (coloured diamond markers). Panels A and B: PCA (Panel A) and LDA (Panel B) comparing the formed tools (bifaces) and debitage specimens to regional sources grouped at the territory resolution. Panels C and D: PCA (Panel C) and LDA (Panel D) comparing the different formed tool types and debitage specimens to regional sources grouped at the territory resolution. Panels D and E: PCA (Panel D) and LDA (Panel E) comparing the different formed tool types and debitage specimens to regional sources grouped at the locality resolution. All LDA yielded >91% of cases correctly re-assigned using a test/training set of 20/80% of source observations in the models. Grey lines = variable loadings/scalings; black lines = 95% confidence ellipses (multivariate-normal); black dashed lines = convex hulls. Diagrams were generated in DataGraph using tabulated data exported from SourceXplorer.

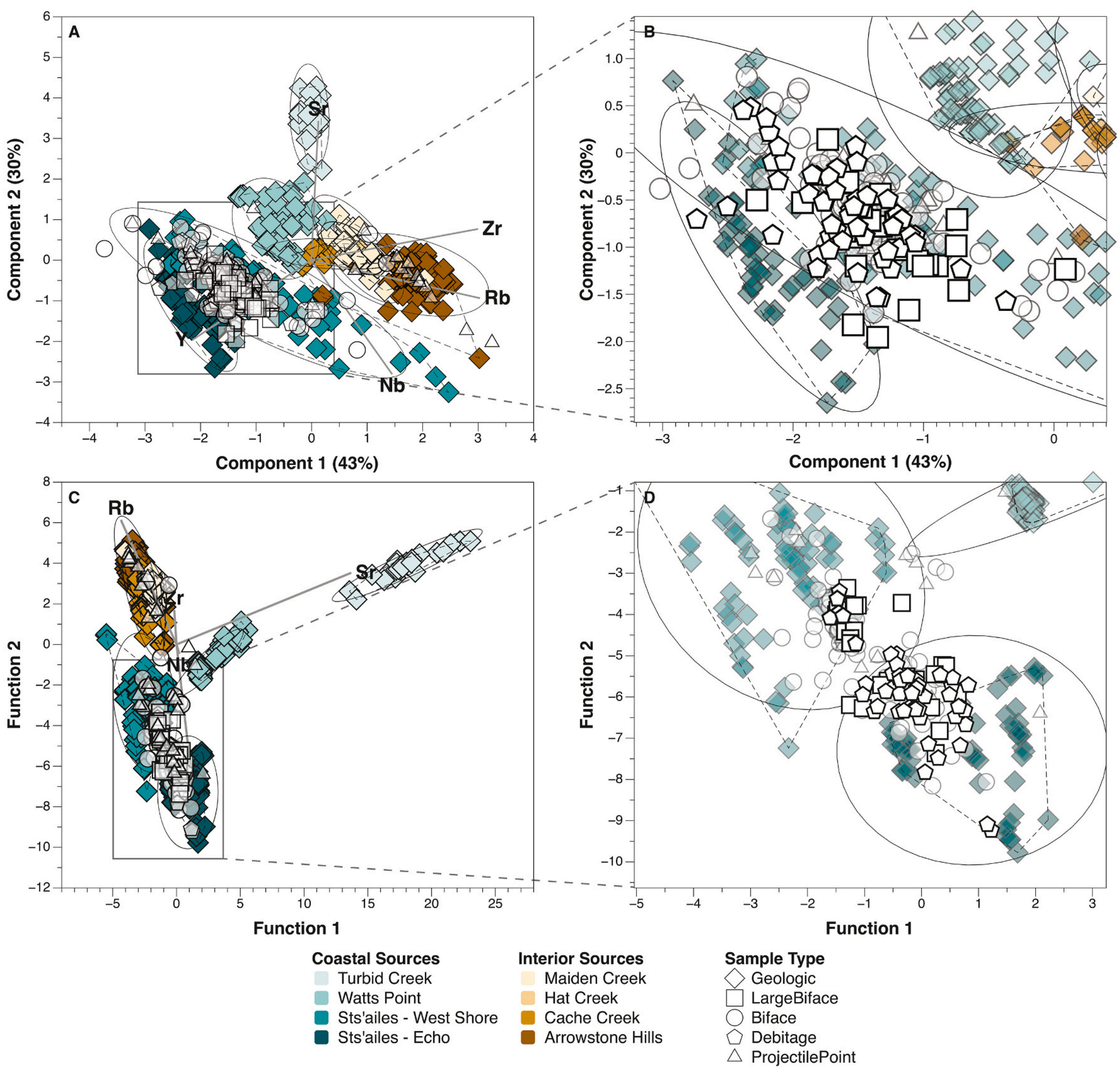


Fig. 11. Comparison biface and debitage specimens (white translucent markers) to Sts'ailes sources and other documented toolstone sources in the region (coloured diamond markers) as shown in Fig. 10 but with the relationships among the debitage and large bifaces highlighted. Panels A and B: PCA comparing the formed tools and debitage specimens to regional sources grouped at the locality resolution. Panels C and D: LDA comparing the different formed tool types and debitage specimens to regional sources grouped at the locality resolution. Grey lines = variable loadings/scalings; black lines = 95% confidence ellipses (multivariate-normal); black dashed lines = convex hulls. Diagrams were generated in DataGraph using tabulated data exported from SourceXplorer.

2017). Using R Shiny means that anyone with a personal computer and a basic grasp of the programming environment can freely apply our method with the use of an Internet connection to access the deployed online interface (R has an active and welcoming online user community to aid autodidacts in gaining competency). This is particularly relevant in the Americas and other colonial contexts, where systemic power imbalances frequently divorce Indigenous communities from the processes of data generation, interpretation, and curation, which can in part be reduced by democratizing the methods of geoarchaeological data analysis.

To further this democratizing effort, SourceXplorer application is published online and can be accessed and used for sourcing applications without any knowledge of R programming whatsoever. Shiny is an R package that allows R scripts to be published as a web application, and to be packaged with a more intuitive GUI than base R's command-line interface. Thus, users without "traditional" R competency can still freely access these tools through the web app, while those with a greater familiarity with the R environment (or an eagerness to "look under the hood") are still able to access and examine the code through the packaged script. Use, reproduction, distribution, and modification are encouraged. We believe that this is an elegant solution that achieves optimal accessibility for the most users.

Table 2
Synthesised sourcing outcomes based on two different source models (territory and locality resolutions, reported as 'robust' and 'standard' as well as 'basic' summaries, respectively) generated from the data output from the Summary (Tabulated) table.

Source	Territory Resolution (Robust Summary)				Territory Resolution (Standard Summary)				Locality Resolution (Basic Summary)			
	Debitage	Biface	Large Biface	Projectile Point	Debitage	Biface	Large Biface	Projectile Point	Debitage	Biface	Large Biface	Projectile Point
Sts'ailes (All)	51	92	25	18	52	101	25	19	–	–	–	–
Sta'ailes - Echo	–	–	–	–	–	–	–	–	44	28	15	6
Sts'ailes - West Shore	–	–	–	–	–	–	–	–	6	70	10	12
Interior (All)	0	1	0	9	0	1	0	10	–	–	–	–
Arrowstone Hills	–	–	–	–	–	–	–	–	0	0	0	9
Watts Point	0	0	0	0	0	0	0	1	0	0	0	1
Ambiguous	0	2	0	2	0	2	0	2	2	4	0	3
Unknown	1	10	0	5	0	1	0	2	0	3	0	3
Total - Sts'ailes Sources	51	92	25	18	52	101	25	19	50	98	25	18
Total - Non-Local Sources	0	1	0	9	0	1	0	11	0	0	0	10
Total - Unknown/ Ambiguous	1	12	0	7	0	3	0	4	2	7	0	6
Grand Total	**52**	**105**	**25**	**34**	**52**	**105**	**25**	**34**	**52**	**105**	**25**	**34**

5.2. Archaeological case study

This case study, which illustrates the operation and functionality of SourceXplorer, has a number of significant archaeological implications. Our identification, description, and geochemical analysis of three new toolstone source localities in the Harrison Lake and Harrison River area significantly expands the catalogue of known lithic sources in the Salish Sea region (e.g., Bakewell, 1996; Mierendorf and Baldwin, 2015; Reimer, 2018). The Sts'ailes source localities are large outcrops along the primary waterways of their territory, accessible year-round by canoe, and consequently ideal for high-volume material collection (Ames, 2002). These sources, and the objects made from the toolstone, likely have a broader distribution among regional assemblages than is currently understood. The identification of systematic variation between the Sts'ailes source locations is also significant, because it provides a means to establish which toolstone sources people from different households and settlements may have visited, providing insight into social dimensions of territoriality and landscape use at a much finer scale than was previously possible. The variability we have detected in source affiliations for debitage, bifaces, large bifaces, and projectile points similarly raises the possibility for selective raw material collection in the production of distinctive objects. This underscores the ongoing need and potential for archaeologists to continue identifying and documenting potential lithic material sources in the region at a fine spatial scale and investigating them using sophisticated statistical procedures.

SourceXplorer provides the means to readily obtain robust evidence of significant lithic production at Yāçk̓EtEl. Most of the analysed lithic material (including all of the analysed debitage) is geochemically consistent with specimens from the local toolstone sources and inconsistent with investigated non-local sources, including both the debitage and formed tools (Figs. 10 and 11). While it seems intuitive that most bifacial debitage found in association with bifaces will be a result of those bifaces manufacture, there are nevertheless exceptions. For instance, Frahm (2016), Waber (in Milward, 2019), and Milward and Waber (2019) describe situations where long curation trajectories of formal tools overlap with new tool manufacture sessions, resulting in the deposition of debitage unrelated to the tools found at the site. Bifacial debitage that differs geochemically from associated bifaces could also be a result of regular exchange between households. It could also occur if biface production occurs in various places across a single site, and individual specialists each use lithic materials from distinct sources. In the case of Yāçk̓EtEl, many of the bifaces were initially taken by an un-sanctioned collector who unsystematically excavated a portion of the lithic workspace before being stopped, and the bifaces were subsequently recovered. Using SourceXplorer, we have demonstrated that they are indeed consistent with the large assemblage of bifacial debitage that we recovered during controlled excavation.

The evidence for biface manufacture at Yāçk̓EtEl indicates a sequence with collection, preparation, and all stages of reduction taking place at a dedicated workspace associated with a house within a few kilometres of the lithic material sources. That the majority of bifaces produced at Yāçk̓EtEl were fashioned from locally occurring toolstone reveals the specialists' access to and preference for these sources. The source locations themselves were likely important places in the region's social landscape, which may have further increased the value of the toolstone. This inference is reinforced by the production of large bifaces (Fig. 11) that could be exchanged as valued prestige items across the Salish Sea. Further, the projectile points made from non-local sources indicate exchange connections over a much broader social landscape as well, potentially extending across the Cascade Mountains into the Interior Plateau. The presence of such exotic material is additional evidence of the connectivity of different widely distributed communities in the distant past (e.g., Blake 2004; Morin 2012).

6. Concluding remarks

SourceXplorer, an application built in the open-source R programming environment, facilitates holistic and step-wise multivariate investigations of the elemental relationships between archaeological lithics and potential geological sources. This application makes complex multivariate statistical assessments accessible to the wider archaeological community through an intuitive GUI, while maintaining transparency through open-source licensing and distribution of the script. Users are able to include or exclude specific sources populations in multivariate statistical models using a simple map interface. In the process, it also facilitates reproducibility and consistency when applying multivariate statistical approaches for assessing the relationships among archaeological and geological materials. Users can update multivariate models using graphical spatial information rather than cumbersome lists and source data files. Consequently, researchers in and beyond the archaeological community can circumnavigate many of the complexities of elemental analysis while standardising statistical approaches and source models, as well as eschewing the costs and requirements for specialist input that are generally required to meaningfully evaluate archaeological sourcing outcomes. Using SourceXplorer, researchers working anywhere in the world can dynamically apply sophisticated and robust multivariate statistical methods to lithic sourcing and explore

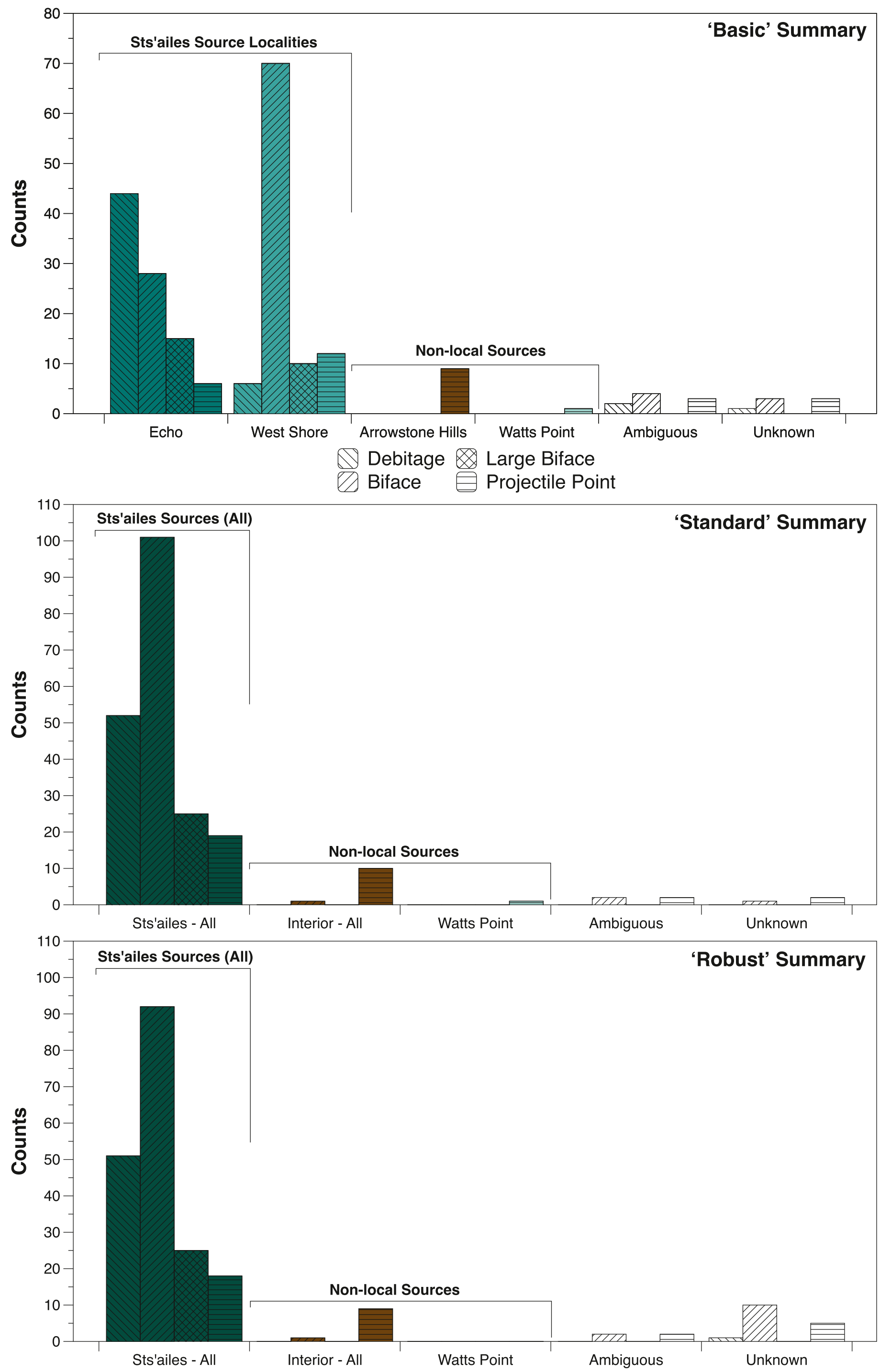


Fig. 12. Synthesised sourcing outcomes based on two different source models (territory resolution at standard pass, top, and locality resolution at basic pass, bottom) showing the distribution of artifact types among their predicted source affiliations. Data exported from the Summary (Tabulated) table and diagrams were generated in DataGraph.

questions of toolstone acquisition and distribution, including cross validation of reported sourcing outcomes.

Declaration of competing interest

None.

Acknowledgements

We are grateful for the countless R Shiny users and developers who have laid out the groundwork for many of the features we have integrated into SourceXplorer, as well as those who participate in online forums answering questions about application development and coding. We are also very grateful to Sts'ailes leadership, and Xoyethet (Boyd Peters) in particular for encouraging this research in the service of learning more about Sts'ailes history and land use. We are very appreciative for the assistance of Dr Jamie Cutts on the geological aspects of the manuscript and related fieldwork in Sts'ailes, Kaitie Purdue and Lindsey Paskulin as field technicians in Sts'ailes during the 2020 and 2021 field seasons, and Jerram Ritchie for assistance in both the laboratory and the field throughout the project. We also thank Michael Elvidge for lithic material specimen contributions and feedback on early app versions, and Mollie McDowell for extremely valuable R support. RM thanks Dr Philippe Claeys for supervisory support. RM and pXRF analyses were funded by the Social Sciences and Humanities Council of Canada (SSHRC). MR is supported by SSHRC (#6567) and the Charles and Alice Borden Fellowship in Archaeology. Much of the thinking, planning, and development of SourceXplorer took place in the traditional, ancestral, and unceded territory of the xʷməθkʷəy̓əm (Musqueam) people.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jas.2022.105626.

References

Waddell, 2018. https://www.axios.com/automating-science-2e43b5c8-cb4a-44cb-bfce-f26272b1bd4c.html.

Ames, K.M., 1995. Chiefly power and household production on the Northwest Coast. In: Foundations of Social Inequality. Springer, Boston, MA, pp. 155–187.

Ames, K.M., 2002. Going by boat: the forager-collector continuum at Sea. In: Fitzhugh, B., Habu, J. (Eds.), Beyond Foraging and Collecting. Plenum Press, pp. 17–50.

Arthur, A.J., Smith, P.L., Monger, J.W., Tipper, H.W., 1993. *Mesozoic Stratigraphy and Jurassic Paleontology West of Harrison Lake, Southwestern British Columbia*. Energy, Mines and Resources Canada. Geological Survey of Canada.

Bakewell, E.F., 1996. Petrographic and geochemical source-modeling of volcanic lithics from archaeological contexts: a case study from British Camp, San Juan Island, Washington. Geoarchaeology 11 (2), 119–140.

Barker, M., Rayens, W., 2003. Partial least squares for discrimination. J. Chemom. 17, 166–173.

Becker, R.A., Chambers, J.M., Wilks, A.R., 1988. The New S Language. Wadsworth & Brooks/Cole.

Blake, M., 2004. Fraser Valley Trade and Prestige as Seen from Scowlitz. *Complex Hunter-gatherers: Evolution and Organization of Prehistoric Communities on the Plateau of Northwestern North America*, pp. 103–112.

Brereton, R.G., Lloyd, G.R., 2014. Partial least squares discriminant analysis: taking the magic away. J. Chemometr. 28 (4), 213–225.

Carlson, R.L., 1994. Trade and exchange in prehistoric British Columbia. In: Prehistoric Exchange Systems in North America. Springer, Boston, MA, pp. 307–361.

Chang, Winston, Cheng, Joe, Allaire, J.J., Sievert, Carson, Schloerke, Barret, Xie, Yihui, Allen, Jeff, McPherson, Jonathan, Dipert, Alan, Borges, Barbara, 2021. Shiny: Web Application Framework for R. R Package Version 1.6.0. https://CRAN.R-project.org/package=shiny.

Cheng, Joe, Karambelkar, Bhaskar, Xie, Yihui, 2021. Leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library. R package version 2.0.4.1. https://CRAN.R-project.org/package=leaflet.

Colpron, M., Nelson, J.L., Murphy, D.C., 2007. Northern Cordilleran terranes and their interactions through time. GSA Today (Geol. Soc. Am.) 17 (4/5), 4.

Cui, Y., Miller, D., Schiarizza, P., Diakow, L.J., 2017. *British Columbia Digital Geology* (Data Version 2018-04-05) [Map]. British Columbia Ministry of Energy, Mines and Petroleum Resources.

DataTools, Visual, 2020. DataGraph. Inc. [Computer software], Version 4.6. https://www.visualdatatools.com/

Ebert, C.E., Dennison, M., Hirth, K.G., McClure, S.B., Kennett, D.J., 2015. Formative period obsidian exchange along the Pacific Coast of M esoamerica. Archaeometry 57, 54–73.

Ewing, T.E., 1981. Petrology and geochemistry of the Kamloops group volcanics, British Columbia. Can. J. Earth Sci. 18 (9), 1478–1491. https://doi.org/10.1139/e81-138.

Flenniken, J.J., 1981. Replicative Systems Analysis: A Model Applied to the Vein Quartz Artifacts from the Hoko River Site. Washington State University Laboratory of Archaeology.

Frahm, E., 2014. Characterizing obsidian sources with portable XRF: accuracy, reproducibility, and field relationships in a case study from Armenia. J. Archaeol. Sci. 49, 105–125.

Frahm, E., 2016. Can I get chips with that? Sourcing small obsidian artifacts down to microdebitage scales with portable XRF. J. Archaeol. Sci.: Rep. 9, 448–467.

Frahm, E., 2017. First hands-on tests of an Olympus vanta portable XRF analyzer to source Armenian obsidian artifacts. IAOS Bullet. 58, 8–23.

Frahm, E., 2019a. Beyond the Technical Revolution: Epistemological Shifts in Archaeological XRF (Or: "The World of XRF Will Never Be the Same Again"). Society for American Archaeology Annual Meeting, Albuquerque, New Mexico, pp. 10–14 (April).

Frahm, E., 2019b. Introducing the Peabody-Yale Reference Obsidians (PYRO) sets: open-source calibration and evaluation standards for quantitative X-ray fluorescence analysis. J. Archaeol. Sci.: Rep. 27, 101957.

Frahm, E., Doonan, R., Kilikoglou, V., 2014. Handheld portable X-ray fluorescence of A egean obsidians. Archaeometry 56 (2), 228–260.

Graesch, A.P., 2007. Modeling ground slate knife production and implications for the study of household labor contributions to salmon fishing on the Pacific Northwest Coast. J. Anthropol. Archaeol. 26 (4), 576–606.

Hall, E.T., 1958. Some uses of physics in archaeology. September. In: Application of Science in Examination of Works of Art-Proceedings of the Seminar, pp. 181–193. September 15-18, 1958.

Hall, E.T., 1960. X-ray fluorescent analysis applied to archaeology. Archaeometry 3 (1), 29–35.

Hall, E.T., Banks, M.S., Stern, J.M., 1964. Uses of X-ray fluorescent analysis in archaeology. Archaeometry 7, 84–89.

Hammond, P.C., 1961. The X-ray geiger-counter spectroscope as an archaeological instrument. Am. J. Archaeol. 65, 305-305.

Kannegaard, R.N., 2015. Quartz Crystal Microblade Function in the Salish Sea Region of Washington State during the Locarno Beach Phase (3500-2400 BP).

Kuhn, M., 2020. Caret: Classification and Regression Training. R package version 6.0-86. https://CRAN.R-project.org/package=caret.

Latour, B., Woolgar, S., 1979. Laboratory Life: the Construction of Scientific Facts. Princeton University Press, Princeton.

Liebmann, M., 2017. From landscapes of meaning to landscapes of signification in the American southwest. Am. Antiq. 82 (4), 642–661. https://doi.org/10.1017/aaq.2017.39.

Liebmann, M., 2018. Response to shackley and moore. Am. Antiq. 83 (4) https://doi.org/10.1017/aaq.2018.50, 756-756.

Lyons, W.H., Thomas, S.P., Skinner, C.E., 2001. Changing obsidian sources at the lost dune and McCoy Creek sites, Blitzen valley, southeast Oregon. J. Calif. Great Basin Anthropol. 273–296.

Mackey, E.A., Christopher, S.J., Lindstrom, R.M., Long, S.E., Marlow, A.F., Murphy, K.E., Paul, R.L., Popelka-Filcoff, R.S., Rabb, S.A., Sieber, J.R., Spatz, R.O., 2010. Certification of three NIST renewal soil standard reference materials for element content: SRM 2709a San Joaquin Soil, SRM 2710a Montana Soil I, and SRM 2711a Montana Soil II. NIST Special Publication 260 (172), 1–39.

Mahoney, J.B., Friedman, R.M., McKinley, S.D., 1995. Evolution of a Middle Jurassic volcanic arc: stratigraphic, isotopic, and geochemical characteristics of the Harrison Lake Formation, southwestern British Columbia. Can. J. Earth Sci. 32 (10), 1759–1776.

Mardia, K.V., Kent, J.T., Bibby, J.M., 1979. Multivariate Analysis. Academic Press, London.

Marwick, Ben, 2016. Computational reproducibility in archaeological research: basic principles and a case study of their implementation. J. Archaeol. Method Theor 1–27. https://doi.org/10.1007/s10816-015-9272-9.

Marwick, Ben, d'Alpoim Guedes, Jade, Michael Barton, C., Bates, Lynsey A., Baxter, Michael, Bevan, Andrew, Elizabeth, A., Bollwerk, et al., 2017. Open science in archaeology. SAA Archaeol. Rec. 17 (no. 4), 8–14.

Marwick, Ben, Boettiger, Carl, Mullen, Lincoln, 2018. Packaging data analytical work reproducibly using R (and friends). Am. Statistician 72 (1), 80–88. https://doi.org/10.1080/00031305.2017.1375986.

McLaren, D., Gray, B., 2017. Six Facets of Quartz Crystal Tools from the Stave River. Archaeology of the Lower Fraser River Region, pp. 113–118.

McLaren, D., Steffen, M.L., 2008. A sequence of formed bifaces from the fraser valley region of British Columbia. In: Projectile Point Sequences in Northwestern North America. SFU Archaeology Press, pp. 161–186.

McMillan, R., Weis, D., Amini, M., Bonjean, D., 2017. Identifying the reworking and stratigraphic provenance of bones by exploring multivariate geochemical relationships with the 'Perio-spot' technique. J. Archaeol. Sci. 88, 1–13.

McMillan, R., Amini, M., Weis, D., 2019. Splitting Obsidian: assessing a multiproxy approach for sourcing obsidian artifacts in British Columbia. J. Archaeol. Sci.: Rep. 28, 102040.

Mierendorf, Robert R., Baldwin, Kevin E., 2015. Toolstone geography in the northern cascades of Washington and adjacent areas. In: Ozbun, Terry, Adams, Ron (Eds.), Toolstone Geography of the Pacific Northwest. Archaeology Press Simon Fraser University, Burnaby, pp. 76–106.

Milward, R., 2019. Archaeological Impact Assessment Interim Report: Back Valley Road Upgrades and Emergency Works. HCA Permit 2015-0133. BC Archaeology Branch.

Milward, R., Waber, N.H., 2019. Final Report Volume 1: Archaeological Investigations at EeQw-166, EeQw-168, eeQw-169, EeQw-170, EeQw-174, EeQw-177, and EeQw-179 between Neskonlith 2 and Chase. HCAP: 2017-0081. BC Archaeology Branch.

Morin, J., 2012. The Political Economy of Stone Celt Exchange in Pre-contact British Columbia: the Salish Nephrite/jade Industry. Doctoral dissertation, University of British Columbia.

Neff, H., 2000. Neutron activation analysis for provenance determination in archaeology. In: Ciliberto, E., Spoto, G. (Eds.), Modern Analytical Methods in Art and Archaeology. Wiley-Interscience, pp. 81–134.

Neff, H., Glascock, M.D., 2003. Neutron activation analysis and provenance research in archaeology. Meas. Sci. Technol. 14 (9), 1516.

Pilloud, M.A., Reaux, D.J., Smith, G.M., Wiggins, K.M., 2017. Using Fordisc software to assign obsidian artifacts to geological sources: proof of concept. J. Archaeol. Sci.: Rep. 13, 428–434.

Pollard, A.M., Batt, C.M., Stern, B., Young, S.M.M., 2007. Analytical Chemistry in Archaeology. Cambridge Manuals in Archaeology. Cambridge University Press.

Pollock, C., 2018. In Pursuit of the Quarry: Exploring Lithic Exchange on the Interior Plateau of British Columbia. Doctoral dissertation, Environment: Department of Archaeology).

R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL. https://www.R-project.org/.

Rapp, G.R., Hill, C.L., 1998. Geoarchaeology: the Earth-Science Approach to Archaeological Interpretation. Yale University Press, New Haven.

Reimer, R., 2018. The Watts Point dacite source and its geological and archaeological occurrence along the shores of the Salish Sea, British Columbia Canada. J. Archaeol. Sci.: Rep. 18, 499–508.

Rick, T.C., Skinner, C.E., Erlandson, J.M., Vellanoweth, R.L., 2001. Obsidian source characterization and human exchange systems on California's Channel Islands. Pacific Coast Archaeol.

Ritchie, M., Lepofsky, D., 2020. From local to regional and back again: social transformation in a Coast Salish settlement, 1500–1000 BP. J. Anthropol. Archaeol. 60, 101210.

Ritchie, M, Morin, J, Ritchie, J, Waber, N, Blake, M, McMillan, R, 2022. Embedded Household Specialization and the Production of Bifaces at Yäçk̓EtEl, a Sts'ailes-Coast Salish Village on the Harrison River, Southwest British Columbia. J. Field Archaeol.

Rorabaugh, A.N., 2015. Investigating Restricted Knowledge in Lithic Craft Traditions Among the Pre-contact Coast Salish. Washington State University.

Rorabaugh, A.N., McNabb, C.Y., 2014. A geospatial analysis of toolstone acquisition and use: a preliminary investigation of material quality and access over 4,000 years in the Salish Sea. Canad. J. Archaeol./J. Canad. Archéol. 371–393.

Rorabaugh, A.N., Davenport, N.S., Grier, C., 2015. Characterizing crystalline volcanic rock (CVR) deposits from Galiano Island, BC, Canada: implications for lithic material procurement at the Dionisio Point locality. J. Archaeol. Sci.: Rep. 3, 591–602.

Sagarbarria, R., 2017. Pre-contact period use of sling weaponry in Pitt Polder. In: Rousseau, M.K. (Ed.), Archaeology of the Lower Fraser River Region. SFU Archaeology Press, p. 8.

Schermer, E.R., Howell, D.G., Jones, D.L., 1984. The origin of allochthonous terranes: perspectives on the growth and shaping of continents. Annu. Rev. Earth Planet Sci. 12 (1), 107–131.

Shackley, M.S., 2011. An introduction to X-ray fluorescence (XRF) analysis in archaeology. In: X-ray Fluorescence Spectrometry (XRF) in Geoarchaeology. Springer, New York, NY, pp. 7–44.

Shackley, M., Moore, J., 2018. More than just Jemez Pueblo obsidian: comment on liebmann's "… landscapes of signification in the American southwest. Am. Antiq. 83 (4), 753–755. https://doi.org/10.1017/aaq.2018.51.

Sievert, C., 2020. Interactive Web-Based Data Visualization with R, Plotly, and Shiny. Chapman and Hall/CRC Florida.

Springer, C., Lepofsky, D., Blake, M., 2018. Obsidian in the Salish Sea: an archaeological examination of ancestral Coast Salish social networks in SW British Columbia and NW Washington state. J. Anthropol. Archaeol. 51, 45–66.

Thomas, K.R., 2019. The Toolstone Formerly Known as Green Andesite: a Geochemical Characterization of Fine-Grained Lithic Materials from the Burrard Inlet Area, Vancouver, BC Canada (Master of Arts Thesis. University of British Columbia.

Trueman, C.N., Behrensmeyer, A.K., Potts, R., Tuross, N., 2006. High-resolution records of location and stratigraphic provenance from the rare earth element composition of fossil bones. Geochem. Cosmochim. Acta 70 (17), 4343–4355.

Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S, fourth ed. Springer, New York, ISBN 0-387-95457-0.

Weigand, P.C., Harbottle, G., Sayre, E.V., 1977. Turquoise sources and source analysis: mesoamerica and the southwestern USA. Exchange Syst. Prehist. 15–34.

Wilkerson, E., 2017. Cultural chronology and spatial lithic analysis at DhRp-52. In: Rousseau, M.K. (Ed.), Archaeology of the Lower Fraser River Region. SFU Archaeology Press, pp. 133–140.

Wilson, L., Pollard, A.M., 2001. The provenance hypothesis. In: Brothwell, D.R., Pollard, A.M. (Eds.), Handbook of Archaeological Sciences. Wiley, New York, USA.