The dating of French gilt bronzes with ED-XRF analysis and machine learning

Arlen Heginbotham\textsuperscript{a}, Robert Erdmann\textsuperscript{b} and Lee-Ann C. Hayek\textsuperscript{c}

\textsuperscript{a}The J. Paul Getty Museum, Los Angeles, CA, USA; \textsuperscript{b}The Rijksmuseum, Amsterdam, The Netherlands; \textsuperscript{c}Smithsonian Institution, NMNH, Washington, DC, USA

\section*{ABSTRACT}
This study presents a method for estimating the date of production of Parisian cast gilt bronze artifacts from the late seventeenth to the early twenty-first centuries based on their elemental composition as measured by energy dispersive X-ray florescence spectroscopy (ED-XRF). Gilt bronzes are notoriously difficult to date; however, during this time period in France, both metallurgy and trade in metals were in a state of continual transformation. These technological and commercial transformations drove changes in the elemental composition of gilt bronzes produced by Parisian craftsmen over time. These changes in turn allow estimation of the manufacture date of gilt bronzes of unknown provenance based on their elemental composition. The composition of undated French gilt bronzes can be compared to a database of compositions derived from well-dated examples. In this paper, a database of 466 such observations collected using ED-XRF is presented as a basis for comparison. The use of an ED-XRF calibration protocol designed to deliver accurate and reproducible quantitative results (CHARMed PyMca) allows for data collected on multiple instruments to be aggregated in the database and for data collected by similarly calibrated instruments to be evaluated. A method for estimating the date of manufacture for unknown samples is presented using two machine learning techniques, support vector regression (SVR) and random forest regression (RFR). By using a simple combination of both models, an estimate of the time of manufacture may be made within ±37 years with a confidence of approximately 90%. This level of precision can be useful for distinguishing eighteenth century castings from nineteenth century copies as well as from castings made in the twentieth century.

\section*{RÉSUMÉ}
Cette étude présente une méthode permettant d’estimer la date de production d’artefacts parisiens en bronze coulé et doré datant de la fin du 17e au début du 21e siècle, à partir de leur composition élémentaire telle que mesurée par spectrométrie de fluorescente des rayons X par analyse dispersive en énergie (ED-XRF). Les bronzes dorés sont reconnus pour être difficiles à dater; cependant, au cours de cette période en France, la métallurgie et le commerce des métaux étaient en constante transformation. Ces transformations technologiques et commerciales ont entrainé des changements dans la composition des bronzes dorés produits par les artisans parisiens au fil du temps. Ces changements permettent en retour l’estimation de la date de fabrication de bronzes dorés de provenance inconnue sur la base de leur composition élémentaire. La composition de bronzes dorés français non datés peut être comparée à une base de données de compositions provenant d’exemples bien datés. Dans cet article, une base de données de 466 de ces observations recueillies en utilisant la ED-XRF est présentée comme base de comparaison. L’utilisation d’un protocole de calibration de la ED-XRF conçu pour fournir des résultats quantitatifs précis et reproductibles (CHARMed PyMca) permet aux données recueillies de plusieurs appareils d’être regroupées dans la base de données et pour les données recueillies par des appareils calibrés de façon similaire d’être évaluées. Une méthode d’estimation de la date de fabrication d’échantillons inconnus est présentée utilisant deux techniques d’apprentissage automatique soit la régression de support vectoriel (SRV) et la régression de forêts aléatoires (RFR). En utilisant une simple combinaison des deux modèles, une estimation de la date de fabrication peut être faite avec une marge d’erreur de ±37 ans et un intervalle de confiance approximatif de 90%. Ce degré de précision peut être utile pour distinguer des pièces coulées au 18e siècle de copies du 19e siècle ainsi que de moulages faits au 20e siècle. Traduit par : Isabelle Cloutier et Mireille Brulotte.

\section*{RESUMO}
Este estudio presenta un método para estimar a data de produção de artefatos parisienses de bronze fundido do final do século 17 até o início do século 21 com base em sua composição elemental medida por espectroscopia de dispersão de raios X por energia dispersiva (ED-XRF).
Os bronzes dourados são notoriamente difíceis de datar; no entanto, durante esse período de tempo na França, tanto a metalurgia quanto o comércio de metais estavam em um estado de constante transformação. Essas transformações tecnológicas e comerciais levaram a mudanças na composição elementar dos bronzes dourados produzidos pelos artesãos parisienses ao longo do tempo. As mudanças, por sua vez, permitem estimar a data da manufatura de bronzes dourados de procedência desconhecida, com base em sua composição elementar. A composição de bronzes dourados franceses não datados pode ser comparada a um banco de dados de composições derivadas de exemplos bem datados. Neste artigo, um banco de dados de 466 dessas observações coletadas usando ED-XRF é apresentada como base para comparação. O uso de um protocolo de calibração ED-XRF projetado para fornecer resultados quantitativos precisos e reproduzíveis (CHARMed PyMca) permite que os dados coletados em múltiplos instrumentos sejam agregados no banco de dados e que os dados coletados por instrumentos calibrados de forma similar sejam avaliados. Um método para estimar a data de fabricação de amostras desconhecidas é apresentado usando duas técnicas de aprendizado de máquina, a regressão vetorial de suporte (SVR) e a regressão aleatória de floresta (RFR). Usando uma combinação simples de ambos os modelos, uma estimativa do tempo de manufatura pode ser feita dentro de ± 37 anos com uma confiança de aproximadamente 90%. Este nível de precisão pode ser útil para distinguir peças fundidas do século XVIII de cópias do século XIX, bem como de peças fundidas feitas no século XX. Traduzido por Cristina Antunes e Valeria Gauz, revisado pelo Beatriz Haspo.

RESUMEN
Este estudio presenta un método para estimar la fecha de producción de artefactos parísios de bronce dorado desde finales del siglo XVII hasta principios del siglo XXI, en base a su composición elemental medida por espectroscopía de fluorescencia de rayos X de energía dispersiva (ED-XRF). Los bronzes dorados son notoriamente difíciles de fechar; sin embargo, durante este período en Francia, tanto la metalurgia como el comercio de metales se encontraban en un estado de continua transformación. Estas transformaciones tecnológicas y comerciales impulsaron cambios en la composición elemental de los bronzes dorados producidos por artesanos parísios a lo largo del tiempo. Estos cambios a su vez permiten estimar la fecha de fabricación de bronzes dorados de procedencia desconocida en función de su composición elemental. La composición de bronzes dorados franceses sin fecha se puede comparar con una base de datos de composiciones derivadas de muestras que han sido bien fechadas. En este documento, se presenta una base de datos de 466 de estas observaciones recopiladas utilizando ED-XRF como base de comparación. El uso de un protocolo de calibração ED-XRF diseñado para ofrecer resultados cuantitativos precisos y reproducibles (CHARMed PyMca) permite que los datos recopilados en múltiples instrumentos se agreguen en la base de datos y que se evalúen los datos recopilados mediante instrumentos calibrados de manera similar. Se presenta un método para estimar la fecha de fabricación de muestras desconocidas utilizando dos técnicas de aprendizaje automático, regresión vetorial de apoyo (RVS) y regresión forestal aleatoria (RFR). Al usar una combinación simple de ambos modelos, se puede hacer una estimación de la época de fabricación de ± 37 años con una confianza de aproximadamente 90%. Este nivel de precisión puede ser útil para distinguir las fundiciones del siglo XVIII de las copias del siglo XIX, así como de las piezas fundidas hechas en el siglo XX. Traducido por Amparo Rueda.

1. Introduction
Decorative and sculptural works in gilt bronze were of great importance in contemporary French decorative arts from at least the late seventeenth century through the early twentieth century, and they continue to be highly valued. A great deal has been written about the production methods of French gilt bronzes as well as the evolution of their styles and the economic and historical context of their production (Ottomeyer, Pröschel, and Augarde 1986; Verlet 1987; Chapman 1993; Considine and Jamet 2000; Mestdagh and Lécoules 2010). Recently, a number of collection and exhibition catalogs have been produced that highlight the continuing scholarly interest in French gilt bronzes (Alcouffe et al. 2004; Wilson et al. 2008; Dupuy-Baylet 2010; Vignon et al. 2016).

The dating, attribution, and authentication of French works of art in gilt bronze is a difficult matter. These objects (referred to here as “bronzes” since this is common parlance, although technically brasses) have been reproduced and restored for centuries, making attribution and authentication based on style and workmanship problematic (Kisluk-Grosheide, n.d.; de Bellaigue 1974, 1:35). Already by the fourth quarter of the eighteenth century Parisian cabinet makers were replicating and reusing gilt bronze mounts and other elements of baroque furniture from the early part of the century. Such revivals in the popularity of earlier styles have followed
each other in a complex and never-ending cycle. Leaving aside what might be called “legitimate” reproductions, produced with no intent to deceive, it is also clear that deceptive reproduction has occurred for centuries (de Bellaigue 1974, 1:36; Jacobsen 2016) and continues to this day (Vincent Noce 2016).

Further complicating matters of attribution and authentication, the traditions of material use and methods of fabrication have been passed down through the generations with remarkable consistency. A young artisan in Paris in the nineteenth century, or even today, might receive training as a founder, chaser, or gilder and learn to use tools and techniques that would be entirely familiar to his eighteenth century counterpart.

To address some of the difficulties encountered in the authentication, attribution, and dating of French gilt bronzes, a program of technical study has been developed at the J. Paul Getty Museum that was outlined in Heginbotham (2014). One specific analytical method that has proved useful is the use of energy dispersive X-ray fluorescence spectroscopy (ED-XRF) for compositional analysis of the copper alloy casting metal from which these objects are made.

This paper first introduces a database, developed at the Getty Museum, of elemental compositions of well-dated French gilt bronzes acquired using ED-XRF. We then describe how machine learning techniques can be used to build predictive models from the database. These models can be used to estimate the date of manufacture of un-provenanced works in gilt bronze. In particular, this paper focuses on the application of two machine learning methods, random forest regression (RFR) and support vector regression (SVR). While brief summaries of these techniques will be presented here, the reader is referred to the specific literature cited for full explanations of the methods.

Finally, three examples are given of the RFR and SVR approaches being used to assign dates to objects. These examples illustrate the usefulness of the methods in real-world applications and demonstrate the reproducibility of the dating method using ED-XRF data from different instruments that have been calibrated according to the CHARMed PyMca protocol (Heginbotham and Solé 2017).

2. Description of the data set

2.1. Acquisition of the data

The systematic study of French gilt bronzes by ED-XRF at the Getty Museum began in 2002 as part of a technical study program of the museum’s French furniture collection. Since then, many hundreds of objects have been analyzed from artworks in public and private collections in the United States and Europe. The Getty database discussed in this paper, which will also be referred to as the “training set,” represents a sub-group of the larger collection of sample data. The training set contains only analytic observations from objects that are known to have been made in Paris between the late seventeenth and early twentieth centuries, which are castings (as opposed to sheet brass stock), which are not contaminated with solder or brazing metal, and which are brasses (this excludes tin bronzes and silicon bronzes that have been used occasionally to make gilt bronzes, primarily in the mid-nineteenth and late twentieth centuries respectively). Furthermore, the training set is limited to observation from objects whose authenticity has been clearly established, and whose date of manufacture is known as precisely as possible, usually within a range of ±10 years. Whenever possible, objects whose exact provenance was well-documented were chosen for inclusion.

Many late seventeenth and eighteenth century gilt bronze mounts included in the study came from the Getty Museum’s own collections. These were chosen from pieces of furniture that had undergone detailed technical study to confirm the overall integrity of the pieces (see for example G. Wilson et al. 2008). The use of portable XRF instruments allowed for travel to collections throughout the United States and Europe, both public and private. Collections were identified that held concentrations of well-documented pieces from later periods, and thanks to the generosity of many curators, conservators, and collectors, a database of compositions was slowly built that represents a broad range of Parisian gilt bronzes from the late seventeenth to early twenty-first centuries.

The first iteration of the training set, as presented in Heginbotham (2014), included 539 observations of well-dated Parisian gilt bronzes. After careful data filtering (see below), this was reduced to 466 observations used for this research. Most of the observations that were excluded (98) had been analyzed with a Kevex secondary-target instrument between 2001 and 2005. Although the instrument produced high quality spectra and used a well-regarded fundamental parameters software, it was not possible to confirm that these early Kevex results were comparable to later spectra because insufficient standards had been analyzed before the instrument was decommissioned and dismantled. An additional 28 observations were excluded because either (1) the quality of the spectrum was found to be poor, (2) the alloy was outside of the bounds defined for the study (see: Data Filtering below), or (3) the provenance or assigned date was insufficiently secure. Compensating for some of the exclusions, an additional 53 well
Figure 1. The weight percent compositions (y-axis) plotted against known date of object (x-axis) for twelve elements as determined by ED-XRF using five different instruments to analyze 466 securely dated Parisian gilt bronze castings. These data are used as the training set to build models for the prediction of the date of manufacture for un-provenanced Parisian gilt bronzes. Note that some compositions are below zero; if the true composition of a sample is zero, a properly calibrated system should return an equal number of results above and below zero and the average of many results should closely approximate zero.
provenanced observations have been added to the database since the 2014 publication.

The majority of the data included in the Getty’s database of French gilt bronze compositions was collected between 2006 and 2012 using a Keymaster Tracer ED-XRF instrument belonging to the Getty Conservation Institute (GCI) with a rhenium-anode X-ray tube and a PIN diode detector. Some spectra have been collected with the GCI’s Bruker Artax instrument with a chromium tube and SDD detector. The Artax instrument was particularly useful in cases where castings were small and/or concave and the 1.5 mm spot size and 1 cm standoff of the instrument were beneficial. Since 2012, additional spectra have also been added to the database using three different Bruker Tracer instruments with rhodium anode tubes and SDD detectors. One of these instruments belongs to the GCI, while the others belong to the Philadelphia Museum of Art and to the Wallace Collection, London. See Experimental section below for analysis conditions used.

In all cases, spectra were collected from ungilded areas on the reverse or underside of castings. Analysis sites were chosen where the metal was flat, clean, and uncorroded, and were cleaned with a stiff natural-bristle stencil brush prior to analysis. A preliminary rapid analysis was made on each spot to ensure that no gold was present on the surface in the sample area that would interfere with quantitative analysis. If gold were detected, an alternate spot was sought. In a small percentage of instances, traces of superficial gold were removed with a glass fiber brush prior to analysis. After a satisfactory sample site was identified, a spectrum was collected for the database using a long counting time to optimize precision in the quantification of minor elements (see Heginbotham et al., Forthcoming).

Quantitative analysis of all spectra was conducted following the CHARMed PyMca protocol, designed to maximize accuracy and reproducibility of quantitative ED-XRF results regardless of the instrument used. The method relies on a recently developed set of certified copper alloy reference standards (Heginbotham et al. 2015) coupled with freely-available fundamental parameters software and a consistent calibration strategy (Heginbotham and Solé 2017). The CHARMed PyMca protocol has been shown to produce quantitative results with well characterized reproducibility and accuracy for 15 elements using a wide variety of instrument types (Heginbotham et al., Forthcoming). The use of this protocol has been critical to allowing five different instruments to be used to build and interrogate a single database of compositional data for French gilt bronzes.

The training set of French gilt bronze compositions used in this study is composed of 466 observations, each with quantitative compositional results for 15 elements (Cr, Mn, Fe, Co, Ni, Cu, Zn, As, Se, Ag, Cd, Sn, Sb, and Pb and Bi). Continued additions to the database will certainly increase its value. Figure 1 shows the concentration data for 12 elements plotted against time. Cu, as the matrix element is not included, and Cr and Se are omitted because in the training set, no results were ever recorded above the lower limit of detection (for further discussion, see “Variable Selection” below). A complete table of results is given in Appendix 1.

A useful distinction may be made between elements in the casting metal that were added intentionally by the founder and those that are present in the alloys as impurities beyond the control of the founder. For French gilt bronze, the elements normally added to copper in controlled amounts by the founder are zinc, tin and lead. Each of these elements can contribute to the working properties of the final casting. d’Arcet and de Sève (1818) document the characteristics of different alloys of these three metals with copper, as judged by the various artisans engaged in the production of gilt bronzes in the early nineteenth century. All other minor elements in the database can be considered to be unintentional elements, occurring as impurities within the copper or one of the intentionally added constituents.

2.2. Overview of patterns in the data

A cursory look at the data reveals some interesting temporal patterns. Many of these patterns can be understood in terms of (1) advances in metallurgical technology, (2) changes in trade patterns for ores and refined metal, or (3) long-term trends in foundry practice. These influences are not entirely separable because changes in metallurgy and trade both altered the cost and availability of the founders’ raw materials over time, which may in turn have influenced foundry choices.

Scores of books and articles have been published that shed light on factors influencing the concentration of specific elements in copper alloys through time (Tylecote, Ghaznavi, and Boydell 1977; Craddock and Meeks 1987; Clow 1992; Craddock 1995; Rehren 1995; Craddock and Hook 2012 to name but a very few). A useful overview of many metallurgical factors up to the mid nineteenth century may be found in Percy (1861) and more recent developments are covered in detail in various chapters of Ullmann’s encyclopedia of industrial chemistry, particularly Fabian’s chapter on copper (1985). A useful overview of factors relating to trade and ore sources may be found in Lynch (2002).
A detailed discussion of each element represented in the Getty database and all of the widely varying influences on their concentrations is beyond the scope of this paper, however some general insights may be worth mentioning. In terms of the elements intentionally added by the founders, it is clear that overall, levels of zinc trend upwards in this period while tin and lead additions trend downward. The increase in zinc and decrease in tin has been observed in other brass artifacts as well (Caple 1995) and is probably best understood as an evolution of founders’ preferred alloy type. There may be several influences at work here, but certainly a steady succession of metallurgical innovations from the early eighteenth century through the mid nineteenth century increased the availability of higher-zinc brass raw material through this period (Day 1990, 1991; Heginbotham 2014).

Declining lead content may be due to changes in intentional foundry practices, but may also be attributable to improvements in metallurgy. Historically, lead could have been added intentionally, but also may have been introduced into the brass as an unintentional impurity in copper (Percy 1861; Fennimore and Fistrovich 1996), or through the process of brass-making by cementation (Craddock 2009, 140). Both of these latter two routes would have diminished with improved copper refining methods and with the abandonment of the cementation method in the nineteenth century in favor of brass production by the addition of zinc metal (Day 1990).

Many other elements considered to be impurities (such as silver, antimony, iron, and arsenic) show overall trends downward over time. These trends are probably primarily due to the steady advances in smelting and refining technologies throughout the period, though changes in ore sources are also likely to have played a role. Other elements, such as nickel, cobalt, and manganese show moderate declines through much of the period, followed by evidence of occasional observations with elevated levels in the late nineteenth or twentieth century, likely due to either changes in ore source (e.g., for nickel in the late nineteenth century, see the discussion in Heginbotham 2014), or to intentional inclusion in modern specialty alloys (St. John 1931; Copper Development Association, Inc. 2012).

In summary, the data suggest that from 1675 to the present both metallurgy and the trade in metals were constantly evolving, resulting in changes in the compositions of the brass objects produced by Parisian craftsmen over time. These changes are not necessarily linear or continuous, but they allow for the possibility of estimating the date of production of Parisian cast gilt bronze compositions presented here offers an interesting opportunity to apply two common machine learning techniques to a challenging data set that reflects changes in metal composition arising from a complex interaction of factors related to the history of trade and the evolution of metallurgy.

3. Methods

3.1. Overview

The goal of this study was to develop a method for estimating the date of production of Parisian cast gilt bronze artifacts from the late seventeenth to the early twenty-first centuries, based on elemental composition. A great deal has been written about the analysis of compositional data from cultural heritage objects. It is well-established that no one procedure can be considered optimal for analyzing compositional data (Pollard 1986; Baxter 2001; Glascock 2016). In choosing a method of data analysis, the researcher should consider the specific research goal, the size and character of the dataset available, and the range of possible causal relationships within the research context. It is often recommended to explore datasets using several techniques and then, through comparison of results, determine the optimal modeling solution or solutions to the problem at hand.

In the field of technical art history, the majority of published compositional studies tend to focus attention on individual elements, binary plots of two elements, or occasionally ternary plots (Young et al. 2009; Day et al. 2010). In contrast, the archaeological literature has a stronger tradition of both descriptions of multivariable statistical methods and examples of their application to compositional data sets, covering many material types. (Baxter 2001; Baxter and Jackson 2001; Wilson and Pollard 2001; Baxter 2006; Glascock 2016).

While machine learning algorithms are widely used in scientific, social, and economic research areas, the authors are not aware of any published research that applies machine learning to the analysis of compositional data sets in the realm of cultural heritage. The data set of French gilt bronze compositions presented here offers an interesting opportunity to apply two common machine learning techniques to a challenging data set that reflects changes in metal composition arising from a complex interaction of factors related to the history of trade and the evolution of metallurgy.

Information other than chronological patterns may be present in the training set (for example, alloy preferences of individual workshops or compositional differences between large and small castings, or between sand castings and lost wax castings); however, the investigations of such patterns are beyond the scope of this paper.

3.2. Multivariable analysis

The first attempts by the authors to develop models for predicting the date of French gilt bronzes based on composition used classical multi-variable regression methods including multiple linear regression, non-linear
(quadratic), and non-parametric regressions. These techniques were applied using several data transformation methods (standardization, log transformation, log-ratio transformation) as applicable. For a full discussion of data transformation, see Baxter (2004). Subsequently two newer “machine learning” methods were considered: random forest regression (RFR) (Breiman 2001) and support vector regression (SVR) (Cortes and Vapnik 1995). These analyses were executed in Python-based jupyter notebooks using functions freely available on the internet (Pedregosa et al. 2011). The complete code used to execute the analyses can be found in Appendix II.

Classical regression methods based upon probability theory can test hypotheses, fit distributions, and estimate parameters. Models can be used to make strong inferences about larger data sets from small because they rely upon sets of assumptions. Machine learning algorithms are not useful for making causal inferences but can be used for data-driven prediction of new similar observations, based only on the training set. These algorithms are designed to predict in a series of steps, learning from the data as it continues. Thus, as expected, both SVR and RFR, the two machine learning methods we applied, performed consistently better for training set prediction and predicting manufacturing dates for unknown related observations than the classical regression methods tested.

3.3. Support vector regression

A support vector regression (SVR) is an extension of a support vector machine (SVM). It can be seen as analogous to classical discriminant analysis in that support vector methods consider multivariate data as points in multidimensional space and attempt to classify or regress observations based on their relative positions. However, SVR performs these operations by defining hyperplanes to separate classes that are distinct and not necessarily linearly separable. It optimizes boundaries using a loss function that ignores errors when a limit is set. The final prediction then will not deviate from the observation by more than this limit.

SVR also utilizes kernel functions (frequently a radial basis function or RBF) to construct new variables, increasing the dimensionality of the hyperspace in a computationally efficient manner, and allowing more effective separators to be found. When an RBF kernel is used with SVR, the relationship between the newly created variables and the original variables is mathematically complex and as a result, there is no established method to determine the relative significance or importance of the original variables in the model. For more detailed discussion of SVR and the parameters (weights) discussed below, see (Cortes and Vapnik 1995; Awad and Khanna 2015). For the complete code used to implement the SVR, see Appendix II.

Trials of the SVR model were executed using the original weight percent data we obtained from ED-XRF, as well as using standardized and log-transformed data. Both of the transformed data sets produced slightly better predictive models than the raw data and standardization was chosen as a transformation or pre-treatment for all compositional data to be analyzed by SVR. Thus the elemental values for each observation were standardized according to the formula $y = (y_i - \bar{y})/\sigma$ where $y_i$ is the standardized value, $\bar{y}$ is the initial value, $\bar{y}$ is the average of all values for that element, and $\sigma$ is the standard deviation of all values for that element. After standardization, the average value of all observations is equal to zero with standard deviation $= 1$, and individual observations are assigned positive or negative values in proportion to the number of standard deviations they fall above or below the average.

3.4. Random forest regression

Random forest regression (RFR) is a data driven, decision tree method in which classification or regression is based on a sequence of decisions that splits the multivariate training sample set into successively smaller groupings based on single-variable divisions or splits. The resulting model can be presented as a diagram resembling a tree. At each split, an algorithm (commonly a CHi-squared Automatic Interaction Detection, or CHAID algorithm) is used to divide the training set into two or more groups, selecting the single explanatory variable and the threshold value(s) that will make the split most discriminatory with regard to the dependent variable. At each subsequent split a new explanatory variable is selected and new threshold value(s) are determined. Once a decision tree is defined, unknown observations may be introduced at the “top” of the tree and processed through each split to yield a prediction of either a classification (in a classification tree) or a continuous value (in a regression tree). A detailed explanation may be found in Breiman (1984).

RFR is considered to be an ensemble learning method which generates many varied learning algorithms and aggregates the results (Breiman 2001). In this case, the method aggregates many decision trees (thus the forest),
varying each tree at random. For prediction, the method averages the results from the forest of trees and “votes” on the best predicted value. In RFR each tree is built from a randomly selected subset of the training sample set, typically comprising 60–70% of the observations. Furthermore, at each split within each tree, only a random subset of the explanatory variables is considered to find the best split. Often the square root of the number of variable is used as the maximum number to consider. The observations not used constitute an internal validation set, called by Breiman the out-of bag, or OOB samples. These are used to test the tree. Their values are predicted according to the relevant classification tree and the error rate for that tree is calculated. This strategy means that each individual tree will perform sub-optimally, but for different reasons than every other tree. In practice, when the results from this diverse “forest” of trees is aggregated, the final result is often very robust.

The RF approach has two unusual characteristics in comparison to other MVA methods. First, because RF relies on random decisions at each split, each time an analysis is run, the results will vary slightly from the previous run. With a large number of trees, the results will vary only slightly, but they will never be identical. The other unusual characteristic is that an RF analysis does not generate a distinct mathematical formula for prediction that can be examined and evaluated post facto. This lack precludes the ability to examine the results of an analysis and see how the RF generated the predictions that it did. An RF analysis can, however, generate a table of feature (or variable) importances. These values score the variables based on the decrease in accuracy that occurs in the model when that variable is removed or randomized (Breiman 2001). This metric can give the analyst information about which variables carry the most signal in making predictions.

It has been observed that the random forest approach often performs better than many other regression/classification schemes, including DA, SVM, and even neural networks (Liaw and Wiener 2002). Though not always better with all data types (Smith, Ganesh, and Liu 2013), RFR is generally considered to be very robust against overfitting (and the influence of outliers) and also to be relatively unaffected by data transformation and pre-treatment.

In this study, the number of trees in the forest ($n$ estimators) was set at 5,000, the criterion for evaluating splits was the mean squared error, the minimum number of observations per split was 2, the minimum number of observations per leaf was 1, the maximum number of variables to consider at each split was set to 4, and bootstrap samples were used in constructing trees. To simplify the procedures, it was decided to conduct the RFR analysis on standardized data as was done for the SVR; this allowed the same data files to be used for each method. Even with 5,000 trees, the processing time remained below 10 seconds on a laptop computer. For a discussion of the parameters for RFR, see (Breiman 2001). For the complete code used to implement the RFR, see Appendix II.

RFR model trials were built using the original weight percent data as determined by ED-XRF, as well as using standardized and log-transformed data. As often reported in the literature, the RFR models were not perceptibly affected by data pre-treatment and all trials produced nearly identical results.

### 3.5. Data filtering

The training set data was drawn only from well-dated French gilt bronzes and the quantitative XRF data were generated in a rigorous manner using CHARMed PyMca. Thus, any unknown observations to be analyzed through comparison to this training data must also fit this class. In particular, a series of quantitative and qualitative data filters were applied to ensure that unknowns fit the class of French gilt bronzes.

For purposes of data filtering, the type of alloy under study was defined as brasses with zinc levels above 10%, lead levels below 3.5%, and tin levels below 3%. These limits serve to exclude fundamentally different alloy types (tin bronze or silicon bronze) that have been very occasionally found in gilt bronzes from the nineteenth and twentieth centuries.

Further, unknown observations had to meet the following three criteria or else they were considered unsuitable for analysis. First, it should be established, based on stylistic and documentary evidence, that the unknown gilt bronze is highly likely to be French and have been made between 1675 and 2011. Next, the observation compositional data must have been collected using a reliable and reproducible method, preferably CHARMed PyMca, and the full suite of at least 12 elements should have been reported. Lastly, the elemental concentrations of the unknown observations should fall within the range of “normal” compositions for all elements in the training set. Unknown observations were used for analysis only if their compositions fell between the 2.5 and 97.5 percentile values of the training set for all elements.

### 3.6. Variable selection

The subject of how many and which variables to include in the analysis of compositional data is a complex topic (Baxter and Jackson 2001). In general, the use of too
Figure 2. a, b, c, d, e, f (from left to right and top to bottom). Regression plots for the RFR (top), SVR (middle) and mixed (bottom) models showing 90% confidence intervals. On the left are the regression models derived from the entire training set. On the right are the results of the respective 10-fold cross validations. The much larger confidence intervals for the cross validation plots show that both models are prone to overfitting the data. Realistic estimates of the prediction errors for each model must be derived from the cross-validation data only.
many variables can lead to overfitting (Krzanowski 2008, pp. 516–520); however, the advantage of multivariable statistical techniques as well as machine learning methods is that they aggregate weak signals from multiple elements to improve overall classification or prediction. Therefore, in this study the authors have followed the advice put forward by Pollard (1986), “The best strategy ... is to start with as many variables as possible, and only eliminate the unnecessary ones in the final stages of the study.”

From the outset, Cu, as the matrix element, was excluded because its concentration can always be estimated as a function of the sum of all other elements, and therefore its inclusion was redundant. Cr and Se were excluded early in the research because all results for the training set were below the critical value (where the confidence of detection is less than 95%) and no diachronic trends were observable in the data. Following trial applications of the SVR and RFR algorithms to the training set with and without Cr and Se, the prediction performance of the models (measured as mean absolute error and root mean squared error) either stayed the same or declined with inclusion and so they were excluded. Other elements were considered for exclusion, particularly Bi, Cd, Mn, and Co. However, each of these elements did have results above the lower limit of detection for some observations and there was minor change in the distribution of results. Repeated trials of the SVR and RFR algorithms to the training set were carried out and the prediction success of the models either stayed the same or improved slightly when these elements were included, therefore they were retained. The final training set for analysis by machine learning thus included 12 elements: Mn, Fe, Co, Ni, Zn, As, Ag, Cd, Sn, Sb, Pb, and Bi.

3.7. Cross validation

Prediction models of all sorts are vulnerable to overfitting of the data, i.e., extracting some portion of random variability, or noise, from the training set and interpreting it as meaningful signal. This results in a model which is more successful at making predictions for observations in the training set than it is for new, unknown observations. Cross-validation is the technique used to estimate the accuracy of a predictive model in practice. One type of cross-validation is \( k \)-fold, where the training data set is partitioned into \( k \) subsets.

In order to test the performance of the SVR and RFR models for prediction of unknown observations on French gilt bronzes, 10-fold (\( k = 10 \)) cross validation trials were conducted on each method. The 466 observation training set was randomly divided into ten groups, or folds, of 46 or 47 observations each. Each fold was removed from the training set, one at a time, for use as a validation set while the nine remaining folds were reassembled to form a reduced training set. The reduced training set of nine folds was then used to build SVR and RFR models and then predict dates of manufacture for the known observations in the first fold or validation set. This process was repeated ten times.

Table 1. Performance metrics for the 10-fold cross validation results of RFR, SVR, and Mixed Model (weighted average) predictions based on the Getty database.

|                | Mean Absolute Error (yrs) | 90% Confidence Interval (yrs) | 90th Percentile Absolute Error (yrs) | Maximum Absolute Error (yrs) |
|----------------|---------------------------|-------------------------------|-------------------------------------|-----------------------------|
| SVR            | 20                        | 42                            | 47                                  | 178                         |
| RFR            | 16                        | 39                            | 39                                  | 114                         |
| Weighted Average | 16                       | 37                            | 41                                  | 79                          |

Figure 3. Histograms for the RFR (left), SVR (middle), and mixed (right) models showing the distribution of errors from the 10-fold cross-validation of each model. These graphs show the frequency with which errors of different magnitudes may be expected to occur for each model. A normal distribution approximation line is overlaid on each histogram.
times until each fold had been used for validation. The predicted dates of manufacture for all ten validation sets were then compared to the known dates associated with the observations and the prediction errors were evaluated.

10-fold cross validation is widely used to examine the performance of a statistical or machine learning model when conducted on unknown observations (Kohavi 1995; Arlot and Celisse 2010). The advantages of 10-fold cross validation are that every observation is used for both training and validation, and every observation is used for validation exactly once. While 10-fold cross validation can be vulnerable to problems associated with both bias and variance in the results (Arlot and Celisse 2010; Vanwinckelen and Blockeel 2012) it is generally considered to be a useful method for estimating prediction success.

Along with plots of the cross-validation data and histograms of the associated errors, several measures of predictive ability can be derived from the cross-validation data. Here, four metrics are reported that characterize different aspects of the predictive ability of the models. The first is the mean absolute error of the cross-validation results (MAE), which is the magnitude of the average error of prediction across the entire validation set of 10 folds, comprising all 466 observations.

The second metric presented is a 90% confidence interval (CI) for predictions based on the root mean square error or RMSE. The RMSE is equal to the square root of the average of each prediction error squared. If one assumes that the training set is unbiased and representative of French gilt bronzes in general, the RMSE should represent the standard deviation of future prediction errors based on similar observations. The 90% confidence interval is then given as 1.65 times the RMSE for the cross validation, based on the corresponding $t$-value for 464 degrees of freedom.

The third measure is the 90th percentile of the absolute errors of prediction from the cross validation. This magnitude of error, or less, was predicted for 90% of the cross validated observations. If the errors are normally distributed, this value should be close to the 90% CI.

The last measure used is the maximum absolute error (MaxAE) for the cross validation, which represents the largest error of prediction among all of the 466 predictions made in the course of the validation.

### Table 2. For the RFR model, silver content is the dominant variable for determining the date of manufacture of French gilt bronzes, accounting for nearly 90% of the signal.

| Ag  | Sn  | Zn  | Ni  | As  | Pb  | Fe  | Bi  | Cd  | Mn  | Sb  | Co  | sum |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 0.88| 0.03| 0.02| 0.01| 0.01| 0.01| 0.01| 0.01| 0.005| 0.005| 0.003| 1.00|     |

### Figure 4. A pair of wall lights (Unknown maker, French) from the collection of The J. Paul Getty Museum (89.df.26.1-2) thought to date to 1745–49 based on the presence of “crowned C” tax stamps. Doubts arose regarding their authenticity in the early 2000s. Gilt bronze. 72.4 × 47.6 × 26.7 cm (28 1/2 × 18 3/4 × 10 1/2 in.)
Not surprisingly, the date predictions from cross-validation are not highly precise. However, an estimate may still be made with a precision of approximately ±40 years and a confidence of approximately 90% (see more detailed metrics below). While not sufficient to differentiate castings from the first and second halves of the eighteenth century with high confidence, this level of precision can be useful for distinguishing eighteenth century castings from later nineteenth century copies as well as from castings made in recent decades.

The difference between training and validation results might be considered an unexpected result, particularly for RFR which is widely considered to be a very robust method that is particularly resistant to overfitting data (L. Breiman 2001). In fact, one author has gone as far as to say that “general cross-validation procedures are unnecessary to predict the classification performance of a given RFM [random forest model]. A cross-validation is already built-in, as each tree in the forest has its own training (bootstrap) and test (OOB) [out-of-bag] data.” (Touw et al. 2013). This study, then, should serve as a cautionary tale that regardless of a method’s reputation, cross-validation should always be undertaken to confirm the performance of the method.

The four metrics of performance (discussed in the Methods section) for the cross-validation results are presented in Table 1: the mean absolute error, the 90% confidence interval for predictions based on the RMSE, the 90th percentile of the absolute errors of prediction, and the maximum absolute error. These metrics uniformly suggest that the RFR model provides slightly better date predictions than the SVR model. The similarity between the 90% confidence interval and the 90th percentile of the absolute errors suggests that using a 90% threshold for the confidence interval is reasonable even though the distribution of errors is not normal (Figure 3).

Despite problems of overfitting, both SVR and RFR compared favorably with classical regression techniques for this dataset. In cross validation, data-driven RFR technique proved to be the most precise method for date estimation. While numerous variations of classical

### Table 3. Results of RFR and SVR analysis for the wall lights (89.DF.26)

| Specrum Name          | Instrument     | RFR prediction | SVR prediction | Weighted Average |
|-----------------------|----------------|----------------|----------------|------------------|
| No1M3brushed.csv      | Keymaster Tracer III-V; Re tube | 1895 ± 39      | 1892 ± 42      | 1893 ± 37        |
| No1M1.csv             | Keymaster Tracer III-V; Re tube | 1899 ± 39      | 1889 ± 42      | 1895 ± 37        |
| No2M1brushed.csv      | Keymaster Tracer III-V; Re tube | 1897 ± 39      | 1916 ± 42      | 1906 ± 37        |
| No2M2brushed.csv      | Keymaster Tracer III-V; Re tube | 1882 ± 39      | 1883 ± 42      | 1883 ± 37        |
| Mean Prediction       |                | 1893 ± 39      | 1895 ± 42      | 1894 ± 37        |
| Stan. Deviation of Predictions |            | 8              | 14             | 9                |
| 89.DF.26.1_S1b.spx    | Bruker Artax; Cr tube | 1905 ± 39      | 1903 ± 42      | 1904 ± 37        |
| 89.DF.26.1_S2b.spx    | Bruker Artax; Cr tube | 1904 ± 39      | 1906 ± 42      | 1905 ± 37        |
| 89.DF.26.1_S3.spx     | Bruker Artax; Cr tube | 1978 ± 39      | 1913 ± 42      | 1948 ± 37        |
| 89.DF.26.2_S1.spx     | Bruker Artax; Cr tube | 1899 ± 39      | 1884 ± 42      | 1892 ± 37        |
| 89.DF.26.2_S2.spx     | Bruker Artax; Cr tube | 1900 ± 39      | 1834 ± 42      | 1870 ± 37        |
| 89.DF.26.2_S3.spx     | Bruker Artax; Cr tube | 1910 ± 39      | 1877 ± 42      | 1895 ± 37        |
| 89.DF.26.2_S4.spx     | Bruker Artax; Cr tube | 1918 ± 39      | 1866 ± 42      | 1894 ± 37        |
| Mean Prediction       |                | 1916 ± 39      | 1883 ± 42      | 1898 ± 37        |
| St. Dev. of Predictions |            | 28             | 28             | 24               |
| 89.DF.26.1_S1.csv     | Bruker Tracer III-SD; Rh tube | 1888 ± 39      | 1933 ± 42      | 1909 ± 37        |
| 89.DF.26.1_S2.csv     | Bruker Tracer III-SD; Rh tube | 1890 ± 39      | 1894 ± 42      | 1892 ± 37        |
| 89.DF.26.1_S3.csv     | Bruker Tracer III-SD; Rh tube | 1888 ± 39      | 1922 ± 42      | 1904 ± 37        |
| 89.DF.26.2_S1.csv     | Bruker Tracer III-SD; Rh tube | 1915 ± 39      | 1929 ± 42      | 1921 ± 37        |
| 89.DF.26.2_S2.csv     | Bruker Tracer III-SD; Rh tube | 1879 ± 39      | 1907 ± 42      | 1892 ± 37        |
| 89.DF.26.2_S3.csv     | Bruker Tracer III-SD; Rh tube | 1878 ± 39      | 1885 ± 42      | 1881 ± 37        |
| 89.DF.26.2_S4.csv     | Bruker Tracer III-SD; Rh tube | 1888 ± 39      | 1911 ± 42      | 1899 ± 37        |
| Mean Prediction       |                | 1890 ± 39      | 1912 ± 42      | 1900 ± 37        |
| Stan. Deviation of Predictions |            | 12             | 18             | 13               |

Note: They probably date to the late nineteenth or early twentieth century, a period during which there was a high demand for eighteenth century style French decorative arts and a corresponding surge in the production of legitimate copies, “revival” designs, and fakes.
regression techniques and data pre-treatments were tested, the most successful classical approaches resulted in models that yielded 90% confidence intervals between ±61 and ±83 years.

4.1. Model combination

The results of the SVR and RFR analysis above raise an interesting dilemma. Given two machine learning models, based on different algorithms, that show similar performance but not equivalence, is it better to select the better performing model, or are there ways to combine the models to yield a result superior to either of the two? Opinions vary dramatically (for an excellent overview see Claeskens and Hjort2010). It may be argued that if one model can be determined to be “better”, then it should be chosen; however, it may also be argued that the act of choosing one method over another introduces bias.

In this study, we have chosen to present the results of both the SVR and RFR models as well as the results of a very simple form of mixed model based on the weighted average of the two models. The predictions from each model are weighted by the inverse of the square of the model’s RMSE. Because the RMSE’s for the two models are so close, the weights are similar, yielding weights of 5.4 for RFR and 4.6 for SVR. The same cross-validation metrics can be used for this weighted model.

Table 1 shows a slight advantage to using the mixed model over the RFR alone (there are fewer outliers with very large absolute errors of prediction, resulting in a slightly reduced RMSE). However, this is balanced by a slightly increased mean absolute error and 90th percentile of absolute errors. For the examples below, the predictions of both the RFR and SVR models, along with the weighted average (combined model), are presented for interpretation along with information regarding the strengths and weaknesses of each.

4.2. Variable importances

A disadvantage of many machine learning methods is that it is often difficult or impossible to understand how an answer or prediction was obtained. Because these are iterative learning procedures, neither SVR nor RFR provide actual formulas or mathematical models for their predictions. Because neither is an inferential statistical technique there is no testing of hypotheses or equations for significance statements. However, when using RFR, an undefined quantity termed relative importance is given for the input variables. This is an indication of the relative ranking of usefulness of each of the variables for prediction purposes (see Table 2.). In this RFR model, silver content is clearly the most important variable in determining the date of manufacture of French gilt bronzes, accounting for about 89% of the overall signal.

5. Examples

The J. Paul Getty Museum has substantial holdings of French gilt bronze artworks from the late seventeenth through the late eighteenth century. Two works from the collection were subjected to study by the dating methods detailed above, a pair of wall lights acquired in 1989 (89.df.26.1-2; Figure 4), and a desk (bureau plat) acquired in 1971 (71.DA.95; Figure 5). In the early 2000s, both pairs of objects raised some questions of authenticity for the then-curator of the Museum’s Decorative Arts Department. Although it seemed highly likely based on style, technique, and provenance that the objects were of Parisian production, conventional methods of technical examination (described in Heginbotham 2014) failed to provide convincing evidence to either support or refute their authenticity as mid eighteenth-century works. In addition to these accessioned museum artifacts, a sample of casting metal from the Parisian foundry Remy Garnier, acquired by the museum in 2004, was also analyzed by ED-XRF to test the proposed date prediction models.

Further, in order to test the general proposition that the CHARMed PyMca calibration method produces ED-XRF results from different instruments that are reproducible and can be meaningfully compared, the first and third examples were analyzed using three different calibrated ED-XRF instruments and the results of the machine learning predictions for the date of production were compared.

5.1. Pair of wall lights, 1745–49 (?)

The opulent rococo wall lights were not associated with a specific maker, but were dated to 1745–49 based on the presence of tax stamps (so-called “crowned C” stamps) that were in official use in Paris only during this period. The first ED-XRF analysis of these objects was conducted in 2009 using a Keymaster Tracer handheld instrument with a Re tube. Two analyses were made on each piece, on clean, flat and ungilded surfaces. To remove traces of gold that might have interfered with analysis, three of the four sample sites were prepared with a glass bristle brush. Spectra were quantified according to the CHARMed PyMca protocol, the results were filtered and then used to generate date predictions by RFR and SVR as described above.

According to both RFR and SVR, the wall lights are very unlikely to have been made in the eighteenth
century. Rather, the results of the analysis, shown in Table 3, show that the pair are likely to have been produced in the late nineteenth or early twentieth centuries, with a weighted average prediction of 1894 ± 37 years at 90% confidence. The four individual predictions for each model are in good agreement, with standard deviation of 8 and 14 years respectively, and the RFR and SVR results are extremely similar, differing by only two years.

The late nineteenth century saw a great resurgence in the fashion for eighteenth century style French furniture and a corresponding surge in the production of legitimate copies, “revival” designs, and fakes (Mestdagh and Lécoules 2010; de Bellaigue 1974, 1:31–37). It is difficult, if not impossible, to determine the circumstances under which this pair of objects was made. It is entirely possible that they were originally made and sold as copies and that, over time, their late nineteenth or early twentieth century provenance was lost. It is also possible that they were made with the intent to deceive and to profit from the strong late nineteenth century market for eighteenth century French furniture and decorative arts.

In the context of these results, the crowned C markings on the wall lights (Figure 5) are of particular interest. The identification of these markings as tax stamps, and their link to the period 1745–49, was not widely known during the nineteenth and early twentieth centuries and this knowledge was only rediscovered and published by Henry Nocq in 1924. In the years after Nocq’s publication, crowned C’s came increasingly to be seen as an indicator of authenticity and quality (de Bellaigue 1974, 1:32–35). Therefore, regardless of the circumstance surrounding the fabrication of the wall lights, it seems likely that these pieces were fraudulently marked with crowned C stamps, with the intent to deceive, after 1924.

In early 2018, further ED-XRF analyses of the wall lights were made using two different instruments; a newer model Tracer (now manufactured by Bruker) with an Rh tube, and an Artax instrument with a Cr tube, also manufactured by Bruker. In addition to the four sample sites analyzed in 2008, one new sample site was analyzed on wall light 89.DF.26.1 and two new sites were analyzed on 89.DF.26.2. The new sites were all prepared with a glass bristle brush before analysis. Again, the spectra were quantified according to the CHARMed PyMca protocol, the results were filtered and then used to generate date predictions by RFR and SVR models and the weighted average of the two (see Table 3).

The weighted average predictions from the Artax and second Tracer analyses agree within six years with the first predictions made from the 2009 analyses, giving predictions of 1898 and 1900 (± 37 years at 90% confidence) respectively. This high level of reproducibility between instruments suggests that the Getty database of French gilt bronze compositions should be useful and provide consistent results to other laboratories with ED-XRF instruments calibrated according to the CHARMed PyMca protocol.

It is noteworthy that the seven analyses with the Artax instrument yield date predictions that vary considerably. For the seven weighted average predictions from the Artax data, the standard deviation is 24 years, substantially more than from the other instruments. The CHARMed PyMca calibration for the Artax suggests...
Table 4. The date predictions generated by RFR and SVR for the bureau plat (71.DA.95) are entirely consistent with the Getty Museum’s assigned date of 1745–49, helping to assuage concerns about the authenticity of the piece.

| Spectrum Name                        | Instrument          | RFR Prediction | SVR Prediction | Weighted Average |
|--------------------------------------|---------------------|----------------|----------------|------------------|
| 71.DA.95_m01.csv                     | Keymaster Tracer III-V; Re tube | 1743 ± 39      | 1730 ± 42      | 1737 ± 37        |
| 71.DA.95_m03.csv                     | Keymaster Tracer III-V; Re tube | 1749 ± 39      | 1753 ± 42      | 1751 ± 37        |
| 71.DA.95_m05.csv                     | Keymaster Tracer III-V; Re tube | 1753 ± 39      | 1764 ± 42      | 1758 ± 37        |
| 71.DA.95_m06.csv                     | Keymaster Tracer III-V; Re tube | 1752 ± 39      | 1764 ± 42      | 1757 ± 37        |
| 71.DA.95_m07.csv                     | Keymaster Tracer III-V; Re tube | 1762 ± 39      | 1768 ± 42      | 1765 ± 37        |
| 71.DA.95_m09b.csv                    | Keymaster Tracer III-V; Re tube | 1770 ± 39      | 1782 ± 42      | 1775 ± 37        |
| 71.DA.95_m11.csv                     | Keymaster Tracer III-V; Re tube | 1743 ± 39      | 1726 ± 42      | 1736 ± 37        |
| 71.DA.95_m13glassbrushed.csv         | Keymaster Tracer III-V; Re tube | 1765 ± 39      | 1774 ± 42      | 1769 ± 37        |
| 71.DA.95_m14b.csv                    | Keymaster Tracer III-V; Re tube | 1755 ± 39      | 1749 ± 42      | 1752 ± 37        |
| 71.DA.95_m15.csv                     | Keymaster Tracer III-V; Re tube | 1761 ± 39      | 1768 ± 42      | 1764 ± 37        |
| 71.DA.95_m16.csv                     | Keymaster Tracer III-V; Re tube | 1745 ± 39      | 1745 ± 42      | 1745 ± 37        |

Mean Prediction: 1755 ± 39
St. Deviation of Predictions: 9

Note: The mean RFR and SVR predictions are in very close agreement, even though the two models are based on very different approaches.

Table 5. The weighted average of the date predictions generated by RFR and SVR for the Remy Garnier foundry ingot are consistent with the known date of production of 2004 for three different ED-XRF instruments, even though no other samples from this foundry are included in the Getty database.

| Spectrum Name                        | Instrument          | RFR prediction | SVR prediction | Weighted Average |
|--------------------------------------|---------------------|----------------|----------------|------------------|
| RemyGarnierCastingMetal2004.spx      | Bruker Artax; Cr tube | 2003 ± 39      | 1950 ± 42      | 1979 ± 37        |
| RemyGarnierCastingMetal2004b.spx     | Bruker Artax; Cr tube | 2003 ± 39      | 1965 ± 42      | 1981 ± 37        |
| RemyGarnierCastingMetal2004c.spx     | Bruker Artax; Cr tube | 2004 ± 39      | 1943 ± 42      | 1976 ± 37        |
| RemyGarnierCastingMetal2004d.spx     | Bruker Artax; Cr tube | 2004 ± 39      | 1954 ± 42      | 1981 ± 37        |
| RemyGarnierCastingMetal2004e.spx     | Bruker Artax; Cr tube | 2003 ± 39      | 1957 ± 42      | 1982 ± 37        |
| Mean Prediction:                     | 2003 ± 39           | 1954 ± 42      | 1981 ± 37      |
| St. Deviation of Predictions:        | 1                   | 8              | 3              |
| RemyGarnierCastingMetal2004.csv      | Keymaster Tracer III-V; Re tube | 2003 ± 39      | 1991 ± 42      | 1997 ± 37        |
| RemyGarnier read.#3 (2004).csv       | Bruker Tracer III-SD; Rh tube | 1999 ± 39      | 1999 ± 42      | 1999 ± 37        |
| Mean of 3 Predictions by Different Instruments: | 2002 | 1981 | 1992 |
| St. Dev. of Mean Predictions:        | 3                   | 24             | 10             |

Note: All RFR predictions fall within five years of the known date.

that that instrument should be expected to produce quantitative results that are comparable in precision and accuracy to the other instruments. It is possible that the instrument’s smaller analytical spot size (1.5 mm diameter vs. ~8 mm for the Tracer instruments) may cause the results to be more influenced by localized inhomogeneity in the castings’ compositions. This is merely speculation, however. Grain sizes in cast brasses can vary considerably depending on composition and cooling rate, but are mostly in the range of 50–125 µm (Haque and Khan 2008; “Resources: Standards & Properties – Copper & Copper Alloy Microstructures: Brasses”, n.d.). At this size, a 1.5 mm area of excitation should gather data from a minimum of between 144 and 900 grains, which should be enough to produce a representative result. Lead, being essentially insoluble in copper, tends to form globules that might also account for more heterogeneous measurements. While lead is not a heavily weighted variable in RFR (and would not be expected to be of large importance in SVR) silver is a very important element for prediction. Silver is preferentially soluble in lead over copper and silver-rich lead inclusions have been observed previously in copper ingots (Wang, Strekopytov, and Roberts 2018). Thus, significant inhomogeneity in lead on the millimeter scale, might also lead to inhomogeneity in silver, causing increased variance in date prediction. Further work in this area would be warranted.

5.2. Baumhauer bureau plat

The Getty’s Baumhauer bureau plat is a work of the highest quality, amply decorated with finely chased and gilded mounts. It has been proposed that the bureau plat was a gift to Empress Elizabeth of Russia from Louis XV in the late 1740s (Figure 6). As with the wall lights above, the bureau has been dated to 1745–49 based on the presence of numerous “crowned C” stamps struck on the bronzes that were in official use in Paris only during this period (Figure 7).

In the early 2000s, concerns were raised regarding the authenticity of the piece because of its extraordinary state of preservation and its elaborate internal construction, which is unusual for mid eighteenth century Parisian work. Furthermore, the Getty’s bureau plat is an example of a model that was very popular and often copied in the
nineteenth century. A very similar eighteenth century bureau plat in the collection of the Louvre (Acc. no. OA7805) is thought to have been the source for at least four copies made and stamped by prestigious cabinet-makers of the second half of the nineteenth century. In addition, several anonymous versions, also thought to be late nineteenth century, have passed though the art market over the years (Wilson and Heginbotham, Forthcoming). The earliest certain provenance of the Getty’s bureau plat dates to 1907, leaving open the possibility that it could also be a late nineteenth century copy, in spite of its reputed eighteenth century provenance.

During the course of a complete technical study of the bureau plat in the conservation laboratory of the Getty Museum in 2010, ED-XRF spectra were acquired from un-corroded, flat and ungilded surfaces of 12 different mounts, representing most mount types on the commode. The 12 spectra were quantified according to the CHARMed PyMca protocol, the results were filtered and then used to generate date predictions by RFR and SVR using a version of the database that contained no observations from this object (the observations were later included in the master database, see discussion below). Both of the machine learning models predicted dates of manufacture in the mid eighteenth century for all mounts (see Table 4). The individual results were consistent, with standard deviations of nine and 17 years for the RFR and SVR models, respectively. The mean predictions for the two models fell within three years of each other, with a weighted average of 1756 ± 37 years with 90% confidence, a result that is entirely consistent with the Museum’s presumed date of 1745–1749.

Also as part of the 2010 technical study, a thorough dendrochronological investigation of the bureau plat was carried out by Christine Locatelli and Didier Pousset. This study confirmed that the oak from which the desk was constructed grew in north-eastern France and was likely felled between 1730 and 1760 (Locatelli and Pousset 2010, on file at the J. Paul Getty Museum). This finding further confirmed that the Getty’s Baumhauer bureau plat is a beautifully preserved example of the finest eighteenth century craftsmanship and supports the dating of the desk to 1745–49 based on the crowned “C” stamps on the gilt bronzes. At the conclusion of the technical study, it was decided to include the ED-XRF results from the bureau plat’s mounts in the database of well-dated French gilt bronze compositions.

5.3. Remy Garnier casting metal

In 2004, the Getty Museum acquired a small ingot of the casting metal used to produce gilt bronzes at the Remy Garnier foundry in Paris. The material was extra molten metal that had recently been poured into an ingot mold after an active casting operation, but was never chased or gilded. The ingot was prepared for ED-XRF analysis by abrading the surface with aluminum oxide abrasive paper followed by glass fiber brushing. Analyses were made using the same three ED-XRF instruments as for the wall lights above.

RFR and SVR were used to predict a date of manufacture for the Garnier sample using the Getty database of French gilt bronze compositions (without the Garnier sample) as the training set. Separate predictions were made based on the quantitative results from each of the three instruments (see Table 5). The Getty database contains analytical results from numerous gilt bronze objects produced in the last 25 years by several contemporary Parisian foundries; however, no other observations from the Garnier foundry are included. Even without any other material from Garnier in the database, the observation was dated to within 5 years of the correct date by RFR using each of the three ED-XRF instruments.

The SVR predictions were successful based on single measurements using the two Tracer instruments, with errors of only five and 13 years; however, the SVR prediction for the Artax instrument result was quite poor, with the first analysis yielding an error of 54 years. Repeated measurements of different spots on the Garnier ingot were made with the Artax to see if small analytical spot size combined with inhomogeneity in the sample might be responsible for the large error. However, the mean of five predictions from different spots still resulted in an error of 50 years. The lack of transparency in the SVR method makes it difficult to interrogate the model and understand the reason for the error. Still, in this case the weighted average of the RFR and SVR models yielded an error of 23 years from the known date, falling well within the 37-year 90% confidence interval of the prediction.

6. Conclusion

This study introduces a database developed at the J. Paul Getty Museum of 466 elemental compositions of securely-dated French gilt bronze objects from the late seventeenth century to the early twenty-first century. The database was built over the course of a decade from ED-XRF measurements made by five different instruments that have been calibrated according to the CHARMed PyMca protocol (Heginbotham and Solé 2017). This database shows that the elemental composition of these objects evolved in complex and often non-linear ways over time, probably as a result of changes in trade patterns, metallurgical technology and foundry practice.
The database was used as a training set to generate predictive machine learning models using random forest regression and support vector regression, each based on different fundamental principles. These models (or a combination of the two) can be used to predict the date of manufacture of an undated French gilt bronze based on elemental composition with a cross-validation error of about ±40 years at a confidence level of 90%. This level of precision is substantially better than was achievable using classical inferential statistical methods, demonstrating the advantages of data-driven machine learning methods.

Given that ED-XRF is prone to poor inter-laboratory reproducibility (Heginbotham et al. 2011), the use of quantitative methods that are demonstrably reproducible is of paramount importance before results from different instruments can be aggregated. This study shows that if one uses a common set of certified reference materials and common fundamental parameters software, along with rigorous and transparent methods of calibration, ED-XRF results can be sufficiently reproducible between instruments to construct an aggregated database of compositions for up to 12 elements that is useful for multivariable analysis. More importantly, this study shows that properly calibrated ED-XRF can be sufficiently reproducible that quantitative results from the analysis of unknown observations by different instruments can yield consistent predictions of the date of manufacture using machine learning models.

7. Future work

This study does not attempt to provide a definitive study on the composition of French gilt bronzes, but rather to offer a proof of concept, highlighting the potential for the application of machine learning to the technical study of cultural heritage artifacts. Random forest models and support vector machines are two of many machine learning techniques that may find application in the analysis of compositional datasets. Another promising technique is t-distributed stochastic neighbor embedding (t-SNE; see van der Maaten and Hinton 2008) which is designed to find and graphically display structures of similarity in multivariable datasets. Preliminary work using t-SNE with the Getty gilt bronze database reveals multiple clusters of observations within the large group of mid-eighteenth century observations. As yet, the underlying meaning of these observation clusters is unclear, but further research will aim to determine if the t-SNE clustering might correlate with workshop origin, casting technique, restoration history, or some other as-yet undetermined variable.

The database of gilt bronze compositions presented here should be considered a work in progress and it is hoped that it will continue to grow and become more complete, particularly with respect to seventeenth, early nineteenth, and mid twentieth century samples. We are also actively expanding the database to include gilt bronzes made in Britain, the German-speaking regions, and the Italian peninsula. With clear procedures developed to ensure reproducibility, we hope that inter-laboratory collaboration will accelerate the growth of the database and maximize its utility to other researchers.

Acknowledgements

The authors would like to thank colleagues at the J. Paul Getty Museum who have supported the technical study of gilt bronzes for many years, Gillian Wilson, Charissa Bremer-David, Jeffrey Weaver, Brian Considine, Jane Bassett, Julie Wolfe and all our dedicated interns over the last 17 years.

For their invaluable assistance in arranging access to well documented objects: Adrian Alan; Benoît Jenn of the Musée des Arts Decoratifs; Jean Perfettini of Atelier du Crabe; Marie-Laure Deschamps of the Musée Carnavalet; Yannick Chastang; the staff of the Doheney Mansion; Martin Chapman and Leslie Bone of the Fine Arts Museums of San Francisco; Atelier Michel Jamet; Clara Von Englehardt and Michael Mäder of the Kunsgewerbemuseum Dresden; Christopher Payne; Jan Dorschied and Donna Corbin formerly of the Philadelphia Museum of Art; Charles Moore of the Preservation Society of Newport County; Sylvain Oudry and Delphine Elie-Lefebvre of the Institut National du Patrimoine; Jonathan Marsden and David Wheeler of the Royal Collection; Henrietta Graf, Susanne Almordian and Jens Bartoll of the Stiftung Preussische Schlösser und Gärten; Jean Claude Toulouse; Shayne Rivers, Dana Melchar and Catherine Coueignoux of the Victoria and Albert Museum; and Jurgen Huber and Helen Jacobsen of the Wallace Collection.

Great thanks also to Gareth Davies for his support, advice and encouragement, and to Robert van Langh and the staff of the Rijksmuseum, Amsterdam for hosting the first author during a sabbatical in the summer of 2016 when much of the research for this paper was conducted.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Arlan Heginbotham received his A.B. in East Asian Studies from Stanford University, his M.A. in Art Conservation from Buffalo State College, and his Ph.D. in Earth Sciences from the Vrije Universiteit Amsterdam. He is currently the Conservator of Decorative Arts and Sculpture at the J. Paul Getty Museum. Arlen’s research interests include the use of X-ray fluorescence spectroscopy as a tool for studying copper alloy artifacts, the history of metallurgy and casting
technology, the history and analysis of 17th century East Asian export lacquer, and microscopic and chemical wood identification.

Robert Erdmann, prior to earning his Ph.D. from the University of Arizona in 2006, he started a science and engineering software company and worked extensively on solidification and multiscale transport modeling at Sandia National Laboratories. He subsequently joined the faculty at the University of Arizona in the Program in Applied Mathematics and the Department of Materials Science and Engineering as Assistant Professor and then Associate Professor, where he worked on multiscale material process modeling and image processing for cultural heritage. In 2014, he moved to Amsterdam to focus full-time on combining materials science and computer science to help the world access, understand, and preserve its cultural heritage. He is currently a Senior Scientist at the Rijksmuseum, and he holds the Rijksmuseum Chair in Conservation Science as a Professor at the University of Amsterdam and the Special Chair for the Visualization of Art History at Radboud University.

Dr Lee-Aann C. Hayek is the Chief Mathematical Statistician and Senior Research Scientist at the Smithsonian Institution. She is an internationally recognized scholar and has published over 150 peer-reviewed papers in top-rated journals, plus she has written and edited multiple books. She is an elected Fellow of both the American Statistical Association and the Royal Statistical Society and has extensive experience in statistical science and computer science to help the world access, understand, and preserve its cultural heritage. He is currently a Senior Scientist at the Rijksmuseum, and he holds the Rijksmuseum Chair in Conservation Science as a Professor at the University of Amsterdam and the Special Chair for the Visualization of Art History at Radboud University.

**References**

Alcouffe, Daniel, Anne Dion-Tenenbaum, Gérard Mabille, Jean-Yves Dubois, and Nicolas Dubois. 2004. *Gilts Bronzes in the Louvre*. Dijon: Éditions Faton.

d’Arcet, Jean-Pierre-Joseph, and Jacques Eustache de Sève. 1818. *Mémoire Sur l’art de Dorer Le Bronze : Ouvrage Qui a Remporté Le Prix Fondé Par M. Ravrio et Proposé Par L’Académie Royale Des Sciences*. Paris: De l’imprimerie de Mme. Veuve Agasse.

Arlot, Sylvain, and Alain Celisse. 2010. “A Survey of Cross-Validation Procedures for Model Selection.” *Statistics Surveys* 4 (0): 40–79. doi:10.1214/09-SS054.

Awad, Mariette, and Rahul Khanna. 2015. “Support Vector Regression.” In *Efficient Learning Machines*, 67–80. Berkeley, CA: Apress. doi:10.1007/978-1-4302-5990-9_4.

Baxter, M. J. 2001. “Statistical Modelling of Artefact Compositional Data.” *Archaeometry* 43 (1): 131–147. doi:10.1111/1475-4754.00008.

Baxter, M. J. 2004. “Distance and Transformation in the Multivariate Analysis of Archaeometric Data.” *Proceedings of the International School of Physics “Enrico Fermi” = Rendiconti della Scuola internazionale di fisica “Enrico Fermi”* 154: 17–36.

Baxter, M. J. 2006. “A Review of Supervised and Unsupervised Pattern Recognition in Archaeometry.” *Archaeometry* 48 (4): 671–694. doi:10.1111/j.1475-4754.2006.00280.x.

Baxter, M. J., and C. M. Jackson. 2001. "Variable Selection in Artefact Compositional Studies." *Archaeometry* 43 (2): 253–268. doi:10.1111/1475-4754.00017.

de Bellague, Geoffrey. 1974. *Furniture, Clocks and Gilt Bronzes: The James A. de Rothschild Collection at Waddesdon Manor*. Vol. 1. 2 vols. London: National Trust.

Breiman, Leo. 1984. *Classification and Regression Trees*. Belmont, CA: Wadsworth International Group.

Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32. doi:10.1023/A:10093340324.

Caple, Christopher. 1995. “Factors in the Production of Medieval and Post-Medieval Brass Pins.” In *In Trade and Discovery: The Scientific Study of Artefacts from Post-Medieval Europe and Beyond*, edited by Duncan R. Hook, and David Gaimster. Occasional paper No. 109, 221–234. British Museum Publications Ltd.

Chapman, Martin. 1993. "Techniques of Mercury Gilding in the Eighteenth Century." In *Ancient & Historic Metals: Conservation and Scientific Research: Proceedings of a Symposium Organized by the J. Paul Getty Museum and the Getty Conservation Institute*, November 1991, edited by David A. Scott, Jerry Podany, and Brian B. Considine, 229–238. Marina del Rey, CA: Getty Conservation Institute.

Claesens, Gerda, and Nils Lid Hjort. 2010. *Model Selection and Model Averaging*. Cambridge: Cambridge University Press.

Clow, Simon. 1992. *The International Nickel Trade*. Cambridge, England: Woodhead. http://books.google.com/books?id=ZBdjvXRYQAc&dq=Orford%20copper%20smelter&pg=PP1v#v=onepage&q=Orford%20nickel%20copper%20smelter&f=false.

Considine, Brian, and Michel Jamet. 2000. “The Fabrication of Gilt Bronze Mounts for French Eighteenth-Century Furniture.” In *Gilded Metals: History, Technology and Conservation*, edited by Terry Drayman-Weisser, 283–296. London: Archetype.

Copper Development Association, Inc. 2012. “ASTM Standard Designation for Wrought and Cast Copper and Copper Alloys.” Properties of Wrought and Cast Copper Alloys. 2012. http://www.copper.org/.

Cortes, Corinna, and Vladimir Vapnik. 1995. “Support-Vector Networks.” *Machine Learning* 20 (3): 273–297. doi:10.1023/A:1022627411411.

Craddock, P. T. 1995. *Early Metal Mining and Production*. Washington, DC: Smithsonian Institution Press.

Craddock, P. T. 2009. *Scientific Investigation of Copies, Fakes and Forgeries*. 1st ed. Oxford; Burlington, MA: Elsevier / Butterworth-Heinemann.

Craddock, P. T., and D. Hook. 2012. “An Economic History of the Post-Medieval World in 50 Ingots: The British Museum Collection of Ingots from Dated Wrecks.” *The British Museum Technical Research Bulletin* 6: 55–68.

Craddock, P. T., and N. D. Meeks. 1987. “Iron in Ancient Copper.” *Archaeometry* 29 (2): 187–204. doi:10.1111/j.1475-4754.1987.tb00411.x.

Day, Joan. 1990. “Brass and Zinc in Europe from the Middle Ages to the 19th Century.” In *2000 Years of Zinc and Brass*, edited by P. T. Craddock, 123–150. Occasional Papers (British Museum). London: British Museum Publications Ltd.

Day, Joan. 1991. “Copper, Zinc and Brass Production.” In *The Industrial Revolution in Metals*, edited by Joan Day, and R. F. Tylecote, 131–199. London: Institute of Metals.
Day, Julia, Jens Stenger, Katherine Eremin, Narayan Khandekar, and Virginia Budny. 2010. “Gaston Lachaise’s Bronze Sculptures in the Fogg museum.” *Journal of the American Institute for Conservation* 49 (1): 1–26. doi:10.2307/41320430.

Dupuy-Baylet, Marie-France. 2010. *L’heure, le feu, la lumière: les bronzes du mobilier national, 1800–1870*. Dijon: Éaton.

Fabian, Harald. 1985. “Copper.” In *Ullmann’s Encyclopedia of Industrial Chemistry*, edited by Wolfgang Gerhardt, Y. Stephen Yamamoto, F. Thomas Campbell, Rudolf Pfefferkorn, James F. Rounsaville, and Fritz Ullmann, 5th, completely rev. ed. Vol. A7, 163–228. Weinheim, Federal Republic of Germany; Deerfield Beach, FL: VCH.

Fennimore, Donald L., and George J. Fistrovich. 1996. *Techniques 16th – 18th Century*, edited by Wolfgang Gerhardt, Y. Stephen Yamamoto, F. Thomas Campbell, Rudolf Pfefferkorn, James F. Rounsaville, and Fritz Ullmann, 5th, completely rev. ed. Vol. A7, 163–228. Weinheim, Federal Republic of Germany; Deerfield Beach, FL: VCH.

Fennimore, Donald L., and George J. Fistrovich. 1996. *Metalwork in Early America: Copper and Its Alloys from the Winterthur Collection*. Winterthur, DE: Henry Francis du Pont Winterthur Museum.

Glascock, Michael D. 2016. “Compositional Analysis in Archaeology.” *Oxford Handbooks Online*, April. doi:10.1093/oxfordhb/978019935413.013.8

Haque, M M, and A A Khan. 2008. “Investigation on Structure and Properties of Brass Casting.” *Journal of Materials Science & Technology* 3: 299–301.

Heginbotham, Arlen. 2014. “Bronzes Dorés: A Technical Approach to Examination and Authentication of French Gilt Bronze.” In *French Bronze Sculpture: Materials and Techniques 16th–18th Century*, edited by Bourgarit David, Bassett Jane, Bewer Francesca, Bresc-Bautier Geneviève, Malgouyres Philippe, and Scharf Guilhem, 150–165. London: Archetype.

Heginbotham, Arlen, J. Bassett, D. Bourgarit, C. Eveleigh, L. Glinisman, D. Hook, D. Smith, R. J. Speakman, A. Shugar, and R. Van Langh. 2015. “The Copper CHARMed Set: A New Set of Certified Reference Materials for the Standardization of Quantitative X-Ray Fluorescence Analysis of Heritage Copper Alloys.” *Archaeometry* 57 (5): 856–868. doi:10.1111/arch.12117.

Heginbotham, Arlen, Aniko Bezur, Michel Bouchard, Jeffrey M. Davis, Katherine Eremin, James H. Frantz, Lisha Glinisman, et al. 2011. “An Evaluation of Inter-Laboratory Reproducibility for Quantitative XRF of Historic Copper Alloys.” In *Metal 2010: Proceedings of the Interim Meeting of the ICOM-CC Metal Working Group*, October 11-15, 2010, Charleston, South Carolina, USA, edited by Paul Mandikian, Claudia Chemello, Christopher Watters, and Peter Hull, 244–255. Clemson: Clemson University Press.

Heginbotham, Arlen, David Bourgarit, Julia Day, Jan Dorschied, Joe Godla, Lynn Lee, Beth Price, Arie Pappot, and Dominique Robcis. Forthcoming. “CHARMed PyMca Part II: An Evaluation of Accuracy and Inter-Laboratory Reproducibility for ED-XRF Analysis of Copper Alloys.” *Submitted to Archaeometry*.

Heginbotham, Arlen, and V. A. Solé. 2017. “CHARMed PyMca, Part I: A Protocol for Improved Inter-Laboratory Reproducibility in the Quantitative ED-XRF Analysis of Copper Alloys.” *Archaeometry* 59 (4): 714–730. doi:10.1111/arch.12282.

Jacobsen, Helen. 2016. “Gouthiere’s Legacy: Nineteenth-Century Collectors in Britain.” In *Pierre Gouthière: Virtuoso Gilder at the French Court*, edited by Charlotte Vignon, and Christian Baulez, 139–149. New York: The Frick Collection.

Kisluk-Grosheide, Daniëlle O. n.d. “The Art, Form, and Function of Gilt Bronze in the French Interior | Essay | Heilbrunn Timeline of Art History | The Metropolitan Museum of Art.” The Met’s Heilbrunn Timeline of Art History. Accessed February 4, 2017. http://www.metmuseum.org/toah/hd/gilt/hd_gilt.htm.

Kohavi, Ron. 1995. “A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection.” In *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*, 1137–1143. IJCAI’95. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. http://dl.acm.org/citation.cfm?id=1643031.1643047.

Krzanowski, W. J. 2008. *Principles of Multivariate Analysis: A User’s Perspective*. New York: Oxford University Press.

Liaw, Andy, and Matthew Wiener. 2002. “Classification and Regression by Random Forest.” *R News* 2 (3): 18–22.

Locatelli, Christine, and Didier Pouset. 2010. “Dendrochronology Investigations on a Writing Table (71.Da.95) Attributed to the French Cabinetmaker Joseph Baumbauer.” Laboratoire D’expertise Du Bois Et De Datation Par Dendrochronologie.

Lynch, Martin. 2002. *Mining in World History*. London: Reaktion Books Ltd.

van der Maaten, Laurens, and Geoffrey Hinton. 2008. “Visualizing Data Using T-SNE.” *Journal of Machine Learning Research* 9: 2579–2605.

Mestdagh, Camille, and Pierre Lécoules. 2010. *L’ameublement d’art français: 1850–1900*. Paris: Éditions de l’Amateur.

Ottomeyer, Hans, Peter Pröschel, and Jean-Dominique Augarde. 1986. *Vergoldete Bronzen: Die Bronzebearbeiten Des Spätbarock Und Klassizismus*. 2 vols. München: Klinhardt & Biermann.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, et al. 2011. “Scikit-Learn: Machine Learning in Python.” *Journal of Machine Learning Research* 12: 2825–2830.

Percy, John. 1861. *Metallurgy; the Art of Extracting Metals from their Ores, and Adapting them to Various Purposes of Manufacture: Fuel, Fire-Clays, Copper, Zinc, Brass, Etc*. London: J. Murray.

Pollard, A. M. 1986. “Data Analysis.” In *Greek and Cypriot Pottery: A Review of Scientific Studies*, edited by R. E. Jones, 56–83. Athens: British School at Athens.

Rehren, Thilo. 1995. “Kupfer Vom Tor Zur Welt.” In *Kissipenny Und Manilla. Begleitband Zur Ausstellung Im Kultur- Und Stadthistorischen Museum Der Stadt Duisburg*, 59–64. Duisburg: Kultur und Stadthistorisches Museum.

“Resources: Standards & Properties - Copper & Copper Alloy Microstructures: Brasses”. n.d. Accessed August 12, 2018. https://www.copper.org/resources/properties/microstructure/brasses.html.

Smith, Paul F., Siva Ganesh, and Ping Liu. 2013. “A Comparison of Random Forest Regression and Multiple Linear Regression for Prediction in Neuroscience.” *Journal of Neuroscience Methods* 220 (1): 85–91. doi:10.1016/j.jneumeth.2013.08.024.

St. John, H. M. 1931. “The Influence of Impurities in Foundry Brasses and Bronzes.” *Materials Engineering* 2: 242–245.
Touw, Wouter G., Jumamurat R. Bayjanov, Lex Overmars, Lennart Backus, Jos Boekhorst, Michiel Wels, and Sacha A. F. T. van Hijum. 2013. “Data Mining in the Life Sciences with Random Forest: A Walk in the Park or Lost in the Jungle?” *Briefings in Bioinformatics* 14 (3): 315–326. https://doi.org/10.1093/bib/bbs034.

Tylecote, R. F., H. A. Ghaznavi, and P. J. Boydell. 1977. “Partitioning of Trace Elements Between the Ores, Fluxes, Slags and Metal During the Smelting of Copper.” *Journal of Archaeological Science* 4 (4): 305–333. doi:10.1016/0305-4403(77)90027-9.

Vanwinckelen, Gitte, and Hendrik Blockeel. 2012. “On Estimating Model Accuracy with Repeated Cross-Validation.” In *BeneLearn 2012: Proceedings of the 21st Belgian-Dutch Conference on Machine Learning*, 39–44.

Verlet, Pierre. 1987. *Les Bronzes Dorés Français Du XVIIIe Siècle*. Grands Manuels Picard. Paris: Picard.

Vignon, Charlotte, Christian Baulez, Anne Forray-Carlier, Joseph Godla, Helen Jacobsen, Luísa Penalva, Anna Saratowicz-Dudynska, et al. 2016. *Pierre Gouthière: Virtuoso Gilder at the French Court*.

Vincent Noce. 2016. “New Twist in Fake Antique Furniture Scandal Overshadows Opening of Biennale Des Antiquaires in Paris.” *The Art Newspaper*, September 8, 2016, sec. News. http://theartnewspaper.com/news/biennale-des-antiquaires-overshadowed-by-new-twist-in-fake-antique-furniture-scandal/.

Wang, Quanyu, Stanislav Strekopytov, and Benjamin W. Roberts. 2018. “Copper Ingots from a Probable Bronze Age Shipwreck off the Coast of Salcombe, Devon: Composition and Microstructure.” *Journal of Archaeological Science* 97 (September): 102–117. doi:10.1016/j.jas.2018.07.002.

Wilson, Gillian, Charissa Bremer-David, Jeffrey Weaver, Brian Considine, Arlen Heginbotham, Katrina Posner, and Julie Wolfe. 2008. *French Furniture and Gilt Bronzes: Baroque and Régence: Catalogue of the J. Paul Getty Museum Collection*. Los Angeles: J. Paul Getty Museum. Table of contents only http://www.loc.gov/catdir/toc/ecip0723/2007029343.html.

Wilson, Gillian, and Arlen Heginbotham. Forthcoming. *French Furniture and Gilt Bronzes: Rococo: Catalogue of the J. Paul Getty Museum Collection*. Los Angeles: J. Paul Getty Museum.

Wilson, L., and A. M. Pollard. 2001. “The Provenance Hypothesis.” In *Handbook of Archaeological Sciences*, edited by Don R Brothwell, and A. M. Pollard, 507–517. Chichester; New York: J. Wiley.

Young, Marcus L., Suzanne Schnepp, Francesca Casadio, Andrew Lins, Melissa Meighan, Joseph B. Lambert, and David C. Dunand. 2009. “Matisse to Picasso: A Compositional Study of Modern Bronze Sculptures.” *Analytical and Bioanalytical Chemistry* 395 (1): 171–184. doi:10.1007/s00216-009-2938-y.