PAPER

ANTHROPOLOGY

Joseph T. Hefner,[1] *Ph.D.; M. Kate Spradley,*[2] *Ph.D.; and Bruce Anderson,*[3] *Ph.D.*

Ancestry Assessment Using Random Forest Modeling*,[†],[‡]

ABSTRACT: A skeletal assessment of ancestry relies on morphoscopic traits and skeletal measurements. Using a sample of American Black ($n = 38$), American White ($n = 39$), and Southwest Hispanics ($n = 72$), the present study investigates whether these data provide similar biological information and combines both data types into a single classification using a random forest model (RFM). Our results indicate that both data types provide similar information concerning the relationships among population groups. Also, by combining both in an RFM, the correct allocation of ancestry for an unknown cranium increases. The distribution of cross-validated grouped cases correctly classified using discriminant analyses and RFMs ranges between 75.4% (discriminant function analysis, morphoscopic data only) and 89.6% (RFM). Unlike the traditional, experience-based approach using morphoscopic traits, the inclusion of both data types in a single analysis is a quantifiable approach accounting for more variation within and between groups, reducing misclassification rates, and capturing aspects of cranial shape, size, and morphology.

KEYWORDS: forensic science, forensic anthropology, quantitative methods, random forest model, ancestry, craniometrics, morphoscopic traits

The cleverest algorithms are no substitute for human intelligence and knowledge of the data in the problem (1).

At some stage during skeletal analysis, a consulting organization (e.g., law enforcement, medical examiner's office) may ask the forensic anthropologist to assess the ancestry of a set of skeletal remains. These assessments are usually accomplished through (i) a visual assessment of morphoscopic traits and/or (ii) measurements of the cranial and postcranial skeleton. The results of these two methods may be juxtaposed, with more emphasis placed on one method over the other, or they may be treated as though they have equal weight or value. However, to our knowledge there is no empirical basis to support either assumption. The present study is intended to (i) investigate whether morphoscopic and craniometric data provide comparable information regarding biological relationships between population groups, and (ii) combine both data types into a single analytical examination incorporating a statistical method novel to ancestry assessment.

Whether an analyst is classifying sex, ancestry, or some other aspect of the biological profile, the visual assessment of character states relies on several key factors which need consideration. These include the structure of the reference data used in the analysis, some elements of decision theory, and aspects of classification mechanics for the final assessment. For example, during an analysis using *FORDISC 3.0* (FD3), the analyst should follow clear instructions for group and variable selection that are guided by statistical method and theory (2). Analyzing morphoscopic data is different in that respect; a decision on the weight, or emphasis, to assign a particular trait expression must be determined by the analyst and the analysts must decide whether that expression represents biological reality, biological admixture, individual idiosyncrasy, or "an indefinable something" (3). Thus, incorporating morphoscopic data in a final analysis is very different from metric data. As currently used, morphoscopic data analysis relies on published trait lists (see for instance Rhine [4]) with little or no empirical support. Because there are no end-user applications or published discriminant functions equivalent to *FD3* for ancestry assessment using morphoscopic traits, their analysis currently lacks scientific rigor. Hefner et al. (5–8) provide classification statistics for use with morphoscopic traits and demonstrate the utility of these methods. However, uniting craniometric and morphoscopic data into a single analysis merits further exploration.

Combining categorical and continuous data in a single analysis is not new to biological anthropology (9–19). Richtsmeier et al. (13) explore the relationship of nonmetric and metric data using the principle of functional cranial analysis and suggests "cranial metric variables [are] good predictors of nonmetric trait expression" (13:219–220). Further, Bedrick and colleagues (18) combine dental metrics and nonmetric data to investigate biological relationships among Middle Archaic cemeteries. These

[1]JPAC CIL, 310 Worchester Blvd., BLDG 45, Joint Base Pearl Harbor-Hickam, HI 96853-5530.
[2]Department of Anthropology, Texas State University, 601 University Drive, San Marcos, TX 78666.
[3]Pima County Office of the Medical Examiner, 2825 E District St, Tucson, AZ 85714.

studies utilize binary, nonmetric (i.e., epigenetic) heritable traits not typically used in forensic anthropology (7), but nevertheless, suggest a relationship worthy of inquiry.

The limitations of the classification statistics used in forensic anthropology, particularly those that include categorical data, are not always included in the broader discussion on the statistical methods used by biological anthropologists (20). While classification statistics play a vital role in decedent identification in many, if not all, forensic anthropology laboratories, the foundational assumptions behind the statistics are often left unconsidered. Feldsman (20) discusses the limitations of linear discriminant function analysis (DFA) as a classification technique, including a detailed discussion on the statistical assumptions of that method (i.e., normal distribution, homogeneity of VCVM, correlation between means and variances, independent groups, and proportional variable variance). When these assumptions are not met, he recommends classification trees as a potential substitute (20:257). Classification trees are used to predict group membership of a categorical dependent variable from measurements and observations on one or more predictor variables. These methods are some of the most common techniques used in data mining, and, as the methods are nonparametric and capable of handling very large datasets with missing data, they have great potential in anthropological research. To that end, we test the efficacy of random forest modeling (RFM), a classification tree method, using an anthropological dataset composed of cranial morphoscopic and metric data.

Materials and Methods

Reference Samples

For the purpose of our analyses we use matched morphoscopic and craniometric data from three population groups frequently encountered in forensic anthropology cases in the United States: American Black, Hispanic, and American White.

Table 1 presents the sample distribution by sex and ancestry. The American Black (n = 38) and American White (n = 72) samples comprise 20th century individuals from the William M. Bass Donated Skeletal Collection housed at the University of Tennessee, Department of Anthropology, Knoxville, TN. The Hispanic data (n = 39) represent individuals predominately from Mexico obtained from the Pima County Office of the Medical Examiner (PCOME), in Tucson, Arizona. Anderson (2008) provides a thorough overview of the recognition and classification process for the unidentified border crossers (UBC) encountered in Pima County. These UBCs are identified using a "pre-established profile…that indicates the decedent was engaged in the undocumented crossing of [the U.S.] southern border" (21:12). Over 90% of all positively identified undocumented border crossers are from Mexico (21) and, as such, all of the individuals from PCOME are placed in one broad category, Southwest Hispanic (Table 2). Justification is obligatory and is provided by

TABLE 2—*Summary table of the nationalities of identified undocumented border crossers from the Pima County Medical Examiner's Office.**

Nationality	2001	2002	2003	2004	2005	2006	Total
Mexican	56	117	98	114	133	93	611
Guatemalan	0	0	7	6	6	6	25
Salvadoran	0	0	4	4	2	0	10
Brazilian	0	1	1	1	1	0	4
Honduran	0	0	1	1	1	1	4
Ecuadoran	0	0	0	2	0	1	3
Unknown	0	0	1	0	0	2	3
Colombian	0	1	0	1	0	0	2
Dominican	0	0	2	0	0	0	2
Costa Rican	0	1	0	0	0	0	1
Chilean	0	0	0	1	0	0	1
Peruvian	0	0	0	0	1	0	1
Total	56	120	114	130	144	103	667
% Mexican	100	98	86	88	92	90	92

*Modified table from Anderson (21; Table 2).

Anderson (21) who demonstrates the efficacy of the model used in that office.

Data Collection

Standard cranial measurements (Table 3) most commonly employed in forensic practice are used in subsequent analyses (22,23). The choice of these standard measurements is based on the premise that these variables are most commonly employed in ancestry analysis when using *FORDISC* (24), they reflect overall craniofacial morphology, and they maximize sample sizes in each population group for statistical analyses. All of the craniometric data were collected using a three-dimensional digitizer and 3Skull (25).

TABLE 3—*Craniometric and morphoscopic traits (with abbreviations) used in the current study.*

Craniometric Variables		Morphoscopic Trait
Max. cranial length (GOL)	Nasal height (NLH)	Anterior nasal spine (ANS)
Max. cranial breadth (XCB)	Nasal breadth (NLB)	Inferior nasal aperture (INA)
Bizygomatic diameter (ZYB)	Orbital breadth (OBB)	Interorbital breadth (IOB)
Basion–bregma Height (BBH)	Orbital height (OBH)	Malar tubercle (MT)
Cranial base length (BNL)	Biorbital breadth (EKB)	Nasal aperture shape (NAS)
Basion–prosthion length (BPL)	Interorbital breadth (DKB)	Nasal aperture width (NAW)
Maxillo-alveolar Breadth (MAB)	Frontal chord (FRC)	Nasal bone contour (NBC)
Maxillo-alveolar Length (MAL)	Parietal chord (PAC)	Nasal bone structure (NBS)
Biauricular breadth (AUB)	Occipital chord (OCC)	Nasal overgrowth (NO)
Upper facial height (UFHT)	Foramen magnum length (FOL)	Postbregmatic depression (PBD)
Minimum frontal breadth (WFB)	Foramen magnum breadth (FOB)	Posterior zygomatic tubercle (PZT)
Upper facial breadth (UFBR)	Mastoid length (MDH)	Supranasal suture (SPS)
Biasterionic breadth (ASB)		Transverse palatine suture (TPS)
Zygomaxillary breadth (ZMB)		Zygomaticomaxillary suture (ZS)

TABLE 1—*Sample composition.*

	Sex		
Ancestry	Female	Male	Total
American Black	6	32	38
American White	6	33	39
Hispanic	26	46	72
Total	38	111	149

Morphoscopic Traits

A subset of common morphoscopic traits (see Table 3) used in forensic anthropological casework are scored using the system developed by Hefner (5,26) and contained in *Osteoware* (27). *Osteoware* is a software program designed to assist in the documentation of human skeletal remains through real-time data entry into a structured query language relational database. One component of *Osteoware* is "The Macromosphoscopic Module," which incorporates Hefner's (5) definitions and illustrations in addition to several other morphoscopic traits. For each morphoscopic trait, *Osteoware* assigns an ordinal value to each character state using standardized definitions and character illustrations (line drawings). *Osteoware* and the accompanying manual are available from the Smithsonian Institution, National Museum of Natural History (https://osteoware.si.edu/).

Biological Distance Comparisons

To explore the association between morphoscopic and craniometric data and the relationships between the three ancestral groups, Mahalanobis distance matrices are utilized. Mahalanobis distance provides a measure of similarity among two or more groups. A modified Mahalanobis distance statistic was obtained for the morphoscopic data using a tetrachoric correlation matrix and a FORTRAN program written by Konigsberg (28). A preliminary procedure to collapse the morphoscopic traits with more than one category (character state) into binary variables was necessary for the FORTRAN program. To dichotomize the traits, a heuristic value for "expression versus non-expression" was used as a threshold. For example, for the inferior nasal aperture morphology, scores of 3 or less were transformed to 0 (i.e., no nasal sill) and values of 4 or more were transformed to a 1 (i.e., a nasal sill is present). The dichotomization of ordinal traits is the more traditional and theoretically justified means of obtaining a distance matrix when dealing with data of this type. However, for the purposes of DFA, the original morphoscopic trait scores were used; therefore, a Mahalanobis distance matrix was also obtained for the nondichotomized (binary) morphoscopic data set for comparison purposes. A third Mahalanobis distance matrix was obtained for the craniometric data. The craniometric distance matrix and nondichotomized morphoscopic distance matrices were calculated using SAS 9.3 (29). All distance matrices were double centered and the eigenvectors computed using the EIGEN function in NTSYSpc2.2r (30). The eigenvectors were then plotted to visually assess the relationship between morphoscopic and craniometric data.

Discriminant Function Analyses

All discriminant function analyses (DFA) are conducted in SAS 9.1.3 (29). Males and females are pooled following standardization, by sex, setting the mean to 0 and the standard deviation to 1. Because the sample sizes in all population groups are small, a stepwise variable selection procedure is necessary to reduce the number of variables for the craniometric DFA. The reduction in variable numbers keeps the variable-number-to-sample-size-number appropriate. Morphoscopic traits were also subjected to DFA; however, using morphoscopic traits in a DFA likely violates some of the method assumptions; nevertheless, the classification performance can be very good, although cau-

tion must be exercised when interpreting the results (6). Full-model and stepwise procedures are used in the morphoscopic DFA. Descriptive statistics and method specific results are provided, where necessary.

Random Forest Model

Random forest modeling is a nonparametric data mining technique that falls under a suite of statistical methods known as classification trees. Similar to other knowledge discovery methods like artificial neural networks, correlation and regression trees (CARTs), cluster analysis, and anomaly detection, RFMs predict a categorical dependent variable from measurements and observations on one or more predictor variables through a series of rules, or nodes (analogous to the branches on a tree). While RFMs share commonalities with more traditional statistical methods such as linear DFA, the flexibility of RFMs makes them an attractive analytical option for exploring anthropological datasets.

Classification trees have been used extensively in the fields of probability and statistical pattern recognition (31). The utilization of these methods in applied fields for diagnosis, data structure exploration, and decision theory has been successful. Breiman (32) summarizes a variety of classification trees, including random forest models; so, an in-depth discussion of methods beyond RFMs is not presented here.

Random forest modeling comprises traditional classification trees created using a nonparametric algorithm incorporating majority voting and "bagging" to assign cases to a response class. "Bagging" (or, bootstrap aggregating) is a machine learning ensemble meta-algorithm which improves model stability and classification accuracy (33) through a training set drawn from the original data. Bagging generates multiple "new" training sets by sampling the original data (with replacement) to reduce the variance among observations and the possibility of overfitting.

Random forests can be understood using a straightforward example. Imagine sorting a collection of coins into classes using only categorical and continuous variables. The first rule (at the first node of the tree) is coin color (a categorical variable), which effectively separates the pennies (copper-colored) from the nickels, dimes, and quarters (silver-colored). At the next node, reeding, the series of grooves around a coins perimeter, separates the nickels (reeding absent) from the dimes and quarters (reeding present). Finally, coin diameter (a continuous variable) separates the dimes from the quarters using a threshold value set at diameter < 18.0 mm. This decision profile, or tree, provides an efficient method for sorting coins, but the same approach can be used to classify crania into ancestral groups using morphoscopic variables and craniometric data. Although the algorithm is more complex (and the variables are randomly selected at the beginning), the principle is nearly identical.

In practice, an RFM algorithm proceeds as follows: (i) multiple bootstrapped samples are generated from the original dataset; (ii) for each bootstrapped sample, multiple decision trees are generated using a randomly sampled set of predictor variables; (iii) at each branch, a node (think of rules, or threshold values, for a given predictor variable) is used to determine how to best partition the dataset, (iv) the resulting "forest" of trees represents the final ensemble model, where each tree votes for the final classification of an individual and the majority wins (i.e., majority voting). The final step represents the "out-of-bag" (OOB) sample and, as an aggregate of the predictions, represents the
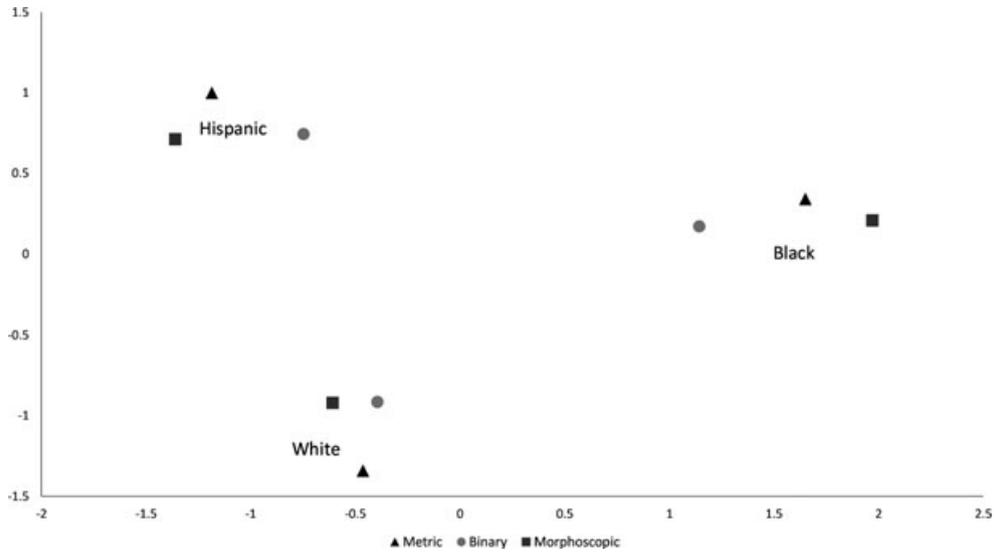
FIG. 1—*A Procrustes plot illustrating the biological relationships among American Blacks, Hispanics, and American Whites and the relationship between morphoscopic traits (squares), binary data (circles), and metric data (triangles).*

overall error rate of the model (also known as the OOB estimate of the error rate [34]). The various randomization procedures introduced during RFM analysis, while seemingly counterintuitive (32), can outperform other classifiers like DFA and kernel density estimates (32,35).

The typically large number of predictor variables and the many decision trees generated during an RFM analysis can create difficulties in the interpretation of an RFM. However, two additional outputs of an RFM greatly simplify the interpretation —variable importance and proximity measures.

Variable importance (VI) values are estimated as a function of the improvement made to a single tree by the inclusion of a variable. More specifically, the variable importance model demonstrates how the classification accuracy changes when the OOB for that variable is permuted across multiple trees, leaving all other variables unchanged. Two importance measures are used herein: the mean decrease in accuracy and the mean decrease in Gini (Gini index). The variable importance values can theoretically range from 0 to infinity. The more important a variable is, the higher the value for that variable.

Under certain conditions, the variable importance measure can be misleading, for instance, when the categorical predictor variables have a large number of character states (36). When this occurs, less important variables or categorical variables with a very large number of character states may be given more importance than is the reality. However, adding an additional layer of randomness to the analysis circumvents these problems and provides an improved importance assessment. The same methods incorporated in this study have been used successfully by other researchers. For example, Liaw and Weiner (34:20) achieve nearly the same prediction accuracy when they reduced a dataset of several thousand predictor variables down to just over twenty of the most appropriate ones drawn through the randomization process and measured using the VI values.

The proximity measure identifies observations that fall in the same terminal nodes between the various trees in a model. This essentially identifies "similar" individuals, which are intuitively in the same terminal node more often than dissimilar individuals. Such a measure is useful for finding structure in the dataset and identifying similar individuals within and between groups. Random forest models average over the predictor variables at termi-

nal nodes (i.e., the final classification is based on majority voting) and identify similar individuals in a process analogous to an adaptive nearest neighbor algorithm (33).

All RFM analyses are conducted using Rattle, a GUI-interface for the computer program R (37) beginning with 1800 trees and seven variables randomly selected and tried at each node. To explore the effects of missing data, imputation for missing variables was not employed. Instead, individuals missing a variable were not included at any node requiring that particular variable.

Results

Biological Distance Comparisons

The eigenvector plot (Fig. 1) provides a two dimensional display of the biological relationships among the three groups and the relationship between all data types. Table 4 presents the distances between each data type. The distances (see Fig. 1; morphoscopic [*squares* in Fig. 1], binary [*circles* in Fig. 1], and metric [*triangles* in Fig. 1]) by group do not overlap in each population group; however, correspondence is demonstrated. The proximity of all data types, for each group, indicates that each type provides similar information about the biological relationships among the population groups. The X axis separates the American Black group from the Hispanic and White groups. The Y axis separates the Hispanic and American Black group from the American White group. All data types, morphoscopic, binary, and metric, provide similar information regarding biological relationships of the three groups. In the American Black and Hispanic groups the morphoscopic and metric data display a close proximity to one another. In the American White group, the morphoscopic and binary data are in closer proximity. The metric (American White) and metric and morphoscopic (His-

TABLE 4—*Distance matrices from metric and morphoscopic data.**

Group	Black	Hispanic	White
Black	–	3.90	3.54
Hispanic	8.48	–	2.87
White	7.30	6.00	–

*Distances from metric data below the diagonal, morphoscopic above.

TABLE 5—*Distribution of cross-validated grouped cases classified using various discriminant analyses and a Random Forest model.*

Group	DFA – Craniometric Only*			DFA – Morphoscopic Only[†]			DFA – Combined[‡]			RFM – Combined[§]		
	Black ($n = 33$)	Hispanic ($n = 23$)	White ($n = 60$)	Black ($n = 38$)	Hispanic ($n = 36$)	White ($n = 68$)	Black ($n = 34$)	Hispanic ($n = 28$)	White ($n = 62$)	Black ($n = 38$)	Hispanic ($n = 39$)	White ($n = 72$)
Black	81.8	9.1	9.1	84.2	10.5	5.3	79.4	14.7	5.9	88.2	2.9	8.8
Hispanic	13.0	73.9	13.0	13.9	72.2	13.9	3.6	89.3	7.1	5.9	79.4	14.7
White	10.0	11.7	78.3	7.4	20.6	72.1	1.6	11.3	87.1	4.5	0.0	95.5
% Correct	78.0			75.4			85.5			89.6		

*A discriminant function analysis using eight stepwise-selected Fordisc variables (MDH MAL OBB AUB OBH NLB FOB BPL).
[†]A discriminant function analysis using 11 morphoscopic traits (ANS, INA, IOB, NAW, NAS, NBS, NBC, NO, PBD, TPS, and ZS).
[‡]A discriminant function analysis using all of the variables in the first two models.
[§]A random Forest Model analysis using 24 craniometric variables and 11 morphoscopic traits.

panic and American Black) provide more separation among groups than the binary data. A more thorough analysis is necessary to explore this relationship; however, to our knowledge, this is the first time the relationship between these three types of (seemingly) disparate data has been quantified.

Individual Methods

Discriminant Function Analysis—Independent three-group analyses between the American Black, American White, and Hispanic datasets were used to assess the performance of craniometric and morphoscopic data in linear DFA. Missing data resulted in fluctuating sample sizes for each analysis. Morphoscopic data may not meet the statistical assumptions of a DFA (see above); however, rigorous cross-validation provides a relative sense of how well the model classifies an unknown individual, particularly since LOO does not require assumptions of normality (38). Table 5 presents the classification accuracy of all methods.

The stepwise-selected craniometric variables achieve an overall cross-validated classification rate of 78%. The American Black group has the highest classification rate of 81.8%, followed by the American White group with a 78.3% classification rate, and the Hispanic group with a rate of 73.9%. A DFA using all morphoscopic traits correctly classified 75% of the cross-validated grouped cases, however, the morphoscopic model does not perform consistently well across groups, correctly classifying only 72% of the American White sample and 72% the Hispanic sample. This model instability may result from assumption violations, sampling error, noise, or a combination of factors.

Finally, ignoring the statistical assumptions discussed above and combining the morphoscopic and craniometric variables into a single analysis, the DFA performs relatively well, classifying nearly 86% of the cross-validated grouped cases correctly. The stability of this model across groups is similar to the morphoscopic model, although the American Black sample did not perform as well as the Hispanic and American White samples.

Next, the American Black, American White, and Hispanic samples were reanalyzed with a random forest model. The initial forest included 24 forensic cranial measurements and 16 morphoscopic traits. A random 20% of the data was set aside to serve as the test set and the remaining data (the training set) was used to grow and combine 1800 trees. After the RFM was complete, individuals in the test set were classified using the new forest to generate a corresponding out-of-bag error estimate (Fig. 2). Next, one hundred bootstrap replicates were performed on the test set to obtain the majority voting classification. A final model of 500 trees provided the lowest misclassification error. Nearly 90% of the test sample was classified into the correct
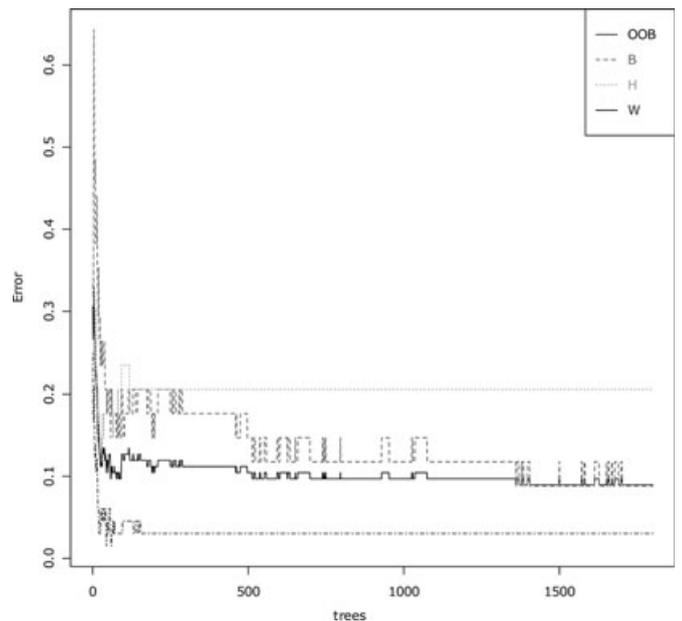

FIG. 2—*The random forest model out-of-bag error estimate, across 1800 trees.*

group (see Table 5). The classification rate for each group is relatively stable, although not surprisingly, the Hispanic sample was least likely to be correctly classified. Variable importance data is presented in Fig. 3. Six morphoscopic traits and 24 craniometric variables were derived from the two RFM variable importance measures and were examined for underlying patterns to better understand their significance. All of the morphoscopic traits are centrally located in the mid-facial region of the cranium, seemingly quantifying Brues' (39) assertion that the mid-facial skeleton is the most important area to consider when estimating ancestry. However, nearly all of the morphoscopic data collected is from the midface, so the importance of this area as measured by the RFM is most likely an artifact of the data collected. The most important variable in the RFM is the nasal aperture shape (NAS). This morphoscopic trait captures both morphology and size (the position of the greatest lateral extent of the nasal opening). Along with the relative importance of INA and NBC in the final RFM model, the morphoscopic traits identified in the RFM may support Brues' (39) claim. These variables are true morphological traits, seemingly incapable of metric assessment. Unlike IOB—which is easily captured metrically using dacryon-to-dacryon—traits like NAS, INA, and NBC
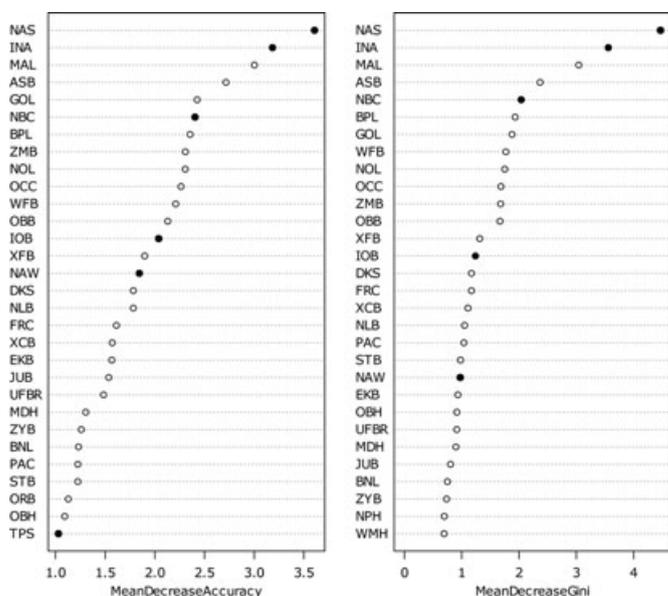
FIG. 3—*Measures of variable importance calculated during random forest modeling. Note: shaded circles represent morphoscopic traits; clear circles represent metric data.*

capture qualitative morphological data much better than any metric analysis could.

The important craniometric variables identified during the RFM also deserve a more thorough treatment. Maximum alveolar length (MAL) is the single most important craniometric variable in the analysis, with a mean decrease in accuracy of nearly 5% in the overall model when that variable is removed. Minimum frontal breadth (WFB), maximum cranial length (GOL), and basion-prosthion length were also important in model construction. These variables capture some of the unique morphological differences between American Black and White crania, and between these two groups and the Hispanic crania.

Discussion

In forensic anthropology, morphoscopic traits and craniometric data are used to classify an unknown cranium into one of several reference groups in an effort to assign a peer-perceived ancestry. Alone, these variables effectively describe shape and size (craniometric data) or morphological variability (morphoscopic traits). Together, they capture a greater amount of the variation observed within and between groups and, when used in combination in a single analysis, they more successfully classify an unknown cranium in to the appropriate reference group.

The eigenvector plot proved very interesting, for the first time quantifying the long held principle that craniometric and morphoscopic data capture similar biological information. Our results indicate that these two data types (craniometric and morphoscopic) provide equivalently the same information about population structure, so combining them in a single analysis should enhance classification power. Unfortunately, because of assumption violations, the classification statistics typically used by forensic anthropologists do not permit the inclusion of these two datasets in a single analysis. Random forest models are novel additions to the classification statistics used by forensic anthropologists. The disparity between linear DFA and RFM goes well beyond the latter's ability to include continuous and categorical data. The RFMs outperformed all of the DFAs, including

increased stability between groups in the final model. The RFMs outperformed DFAs in less obvious ways, as well. Classification trees do very well with missing data (20). Unlike a DFA, which must drop individuals missing even a single predictor variable, RFMs can (i) remove the individual from a particular tree (or in some cases at a particular node) when a variable is missing; (ii) use the class estimate median to replace the missing value; or (iii) use the proximity measure to extrapolate the missing data. We elected to use the first method to explore the effects of missing data on an RFM analysis. Averaged over multiple trees and through multiple iterations, the individual does not have to be dropped from the analysis, particularly when only a single variable is missing. In the DFAs discussed above, 19% of the morphoscopic sample, 38.3% of the craniometric sample, and 47% of the combined DFA were removed due to missing data. The RFM used 100% of the sample to derive the final models. So, not only do the RFMs outperform DFAs, but RFMs also include a larger percentage of the original reference sample.

The instability in classification accuracies noted in the DFA, particularly when morphoscopic traits are included, is most likely an artifact of inequality in variance in morphoscopic traits. Liaw and Weiner (34:20) looked at the rates of correct classification using a suite of statistical methods and a Monte Carlo simulation that adjusted sample size, variance, or both. Small sample size and inequality in variance were the two biggest factors accounting for fluctuating (lower or higher) classification rates. Interestingly, the authors also demonstrate that even when sample size and/or variance inequality negatively affected classification rates in methods like logistic regression and linear discriminant function, other methods, including classification and regression trees (e.g., random forest models), became more accurate by comparison (35:56). They go on to say "when variances are not homogenous, QDA [quadratic discriminant function] and CART [classification and regression trees] do a better job correctly classifying individuals than" logistic regression, linear discriminant function, or neural networks (35:56). When unequal sample sizes or inequality in variance is thought or known to occur, random forest models provide a useful alternative. The inclusion of RFM in a free-standing computer program or into a program like *Fordisc* would make RFM calculations close to instantaneous. In the future, a fully automated RFM will be available for general use.

Conclusions

Random forest modeling is a novel and effective statistical approach for ancestry estimation that can incorporate continuous and categorical variables without violating major statistical assumptions. An obvious question is whether these gains in classification accuracy are worth the added bother of a "new" statistical approach to an "old" methodological practice. The traditional, visual approach utilizing lists of extreme trait values supposedly linked to an ancestral group is no longer considered a valid alternative to, or complement of, craniometric analysis (6). However, useful alternatives that incorporate morphoscopic trait data have recently been offered (6,7). Such models close the gap between morphoscopic variables and craniometric analysis by incorporating two datasets already commonly used in forensic anthropological research.

Random forest models are a worthy addition to the traditional classification statistics already incorporated in forensic anthropological research and casework. As a nonparametric data mining technique, the data do not need to be normalized and no statisti-

cal assumptions regarding the data are necessary. Also, because each tree is built using multiple levels of randomness, each tree is essentially an independent model and, therefore, the final model is not as susceptible to overfitting as in other classification methods. The inclusion of both morphoscopic traits and craniometric variables in a single analysis is intuitively practical. As more variation is accounted for within and between groups, misclassification errors can be reduced, particularly when that variation includes aspects of shape, size, and morphology.

References

1. Breiman L, Cutler A. Random forests. Berkeley, 2004; http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm (accessed February 25, 2013).
2. Jantz R, Ousley S. Fordisc 3.0 computerized forensic discriminant functions. Version 3.0. Knoxville, TN: The University of Tennessee, 2005.
3. Stewart TD. Essentials of forensic anthropology – especially as developed in the United States. Springfield, IL: Charles C Thomas, 1979.
4. Rhine S. Non-metric skull racing. In: Gill GW, Rhine S, editors. Skeletal attribution of race. Albuquerque, NM: Maxwell Museum of Anthropology, 1990;9–20.
5. Hefner J. Cranial nonmetric variation and estimating ancestry. J Forensic Sci 2009;54(5):985–95.
6. Hefner J, Ousley SD. Statistical classification methods for estimating ancestry using morphoscopic traits. J Forensic Sci (in press).
7. Hefner JT, Ousley SD, Dirkmaat DC. Morphoscopic traits and the assessment of ancestry. In: Dirkmaat DC, editor. A companion to forensic anthropology. Chichester, West Sussex: Wiley-Blackwell, 2012;287–310.
8. Hefner J. Morphoscopic traits and the assessment of American Black, American White, and Hispanic ancestry. In: Berg G, Ta'ala S, editors. Biological affinity in forensic identification of human skeletal remains: beyond black and white. Boca Raton, FL: Taylor and Francis (in press).
9. Cheverud JM, Buikstra JE. Quantitative genetics of skeletal nonmetric traits in the Rhesus Macaques of Cayo Santiago. III. Relative heritability of skeletal nonmetric and metric traits. Am J Phys Anthropol 1982;59:151–5.
10. Coppa A, Cucina A, Mancinelli D, Vargiu R, Calcagno JM. Dental anthropology of Central-Southern, Iron Age Italy: the evidence of metric versus nonmetric traits. Am J Phys Anthropol 1998;107:371–86.
11. Corruccini RS. The interaction between nonmetric and metric cranial variation. Am J Phys Anthropol 1976;44:285–93.
12. Hanihara T. Metric and nonmetric dental variations of major human populations. In: Lukacs JR, editor. Human dental development, morphology, and pathology: a tribute to Albert A Dahlberg. Eugene, OR: Department of Anthropology, University of Oregon Press, 1998;173–200.
13. Richtsmeier JT, Cheverud JM, Buikstra JE. The relationship between cranial metric and nonmetric traits in the rhesus macaques from Cayo Santiago. Am J Phys Anthropol 1984;64(3):213–22.
14. Konigsberg L, Kohn LA, Cheverud JM. Cranial deformation and nonmetric trait variation. Am J Phys Anthropol 1993;90:35–48.
15. McClelland JA. Refining the resolution of biological distance studies based on the analysis of dental morphology: Detecting subpopulations at Grasshopper Pueblo [dissertation]. Tucson, AZ: University of Arizona, 2003.
16. Sutter R, Mertz L. Nonmetric cranial trait variation and prehistoric biocultural change in the Azapa Valley, Chile. Am J Phys Anthropol 2004;123:130–45.
17. Cheverud JM, Buikstra JE, Twichell E. Relationships between non-metric skeletal traits and cranial size and shape. Am J Phys Anthropol 1979;50(2):191–8.
18. Bedrick EJ, Lapidus J, Powell JF. Estimating the mahalanobis distance from mixed continuous and discrete data. Biometrics 2000;56(2):394–401.
19. Richtsmeier JT, Cheverud JM, Buikstra JE. The relationship between cranial metric and nonmetric traits in the rhesus macaques from Cayo Santiago. Am J Phys Anthropol 1984;64(3):212–22.
20. Feldesman M. Classification trees as an alternative to linear discriminant analysis. Am J Phys Anthropol 2002;119(3):257–75.
21. Anderson BE. Identifying the dead: methods utilized by the Pima County (Arizona) Office of the Medical Examiner for undocumented border crossers: 2001-2006. J Forensic Sci 2008;53(1):8–15.
22. Buikstra JE, Ubelaker DH, editors. Standards for data collection from human skeletal remains: Proceedings of a seminar at the Field Museum of Natural History. Fayetteville, AR: Archaeological Research Series, 1994.
23. Moore-Jansen PH, Ousley SD, Jantz RL. Data collection procedures for forensic skeletal material. Knoxville, TN: Department of Anthropology, The University of Tennessee, 1994; Report No.: 48.
24. Jantz RL, Ousley SD. Fordisc 3.0: personal computer forensic discriminant functions. Knoxville, TN: University of Tennessee, 2005.
25. Ousley SD. 3skull [computer program]. Windows version 2.0.77, 2004; http://www.mercyhurst.edu/departments/applied_forensic_sciences/faculty_staff.html (accessed February 25, 2013).
26. Hefner J. The statistical determination of ancestry using cranial nonmetric traits [dissertation]. Gainsville, FL: University of Florida, 2007.
27. Osteoware [computer program]. Windows version. Washington, DC: Smithsonian Institution, 2012.
28. Konigsberg LW. Analysis of prehistoric biological variation under a model of isolation by geographic and temporal distance. Hum Biol 1990;62(1):49–70.
29. SAS 9.1.3, 8th edn. Cary, NC: SAS Institute Inc., 2002–2004.
30. NTSYSpc, editor. NTSYSpc. 2.20r. Port Jefferson, NY: Applied Biostatistics, 2005.
31. Ripley B. Pattern recognition for neural networks. Cambridge, U.K.: Cambridge University Press, 1996.
32. Breiman L. Random forests. Mach Learn 2001;45(1):5–32.
33. Breiman L. Bagging predictors: Technical Report 42. Berkley, CA: University of California, 1994.
34. Liaw A, Wiener M. Classification and regression by Random Forest. Rnews 2002;2:18–22.
35. Finch H, Schneider MK. Classification accuracy of neural networks vs. discriminant analysis, logistic regression, and classification and regression trees. Methodology: Eur J Res Methods Behav Soc Sci 2007;3(2):47–57.
36. Strobl C, Boulesteix A-L, Augustin T. Unbiased split selection for classification trees based on the Gini Index. Comput Stat Data Anal 2007;52(1):483–501.
37. R Development Core Team. A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2011.
38. Hand D. Discrimination and classification: Wiley series in probability and mathematical statistics. New York, NY: Wiley, 1981.
39. Brues A. The once and future diagnosis of race. In: Gill GW, Rhine S, editors. Skeletal attribution of race. Albuquerque, NM: University of New Mexico, 1990;1–8.

Additional information and reprint requests:
Joseph T. Hefner, Ph.D.
JPAC - Central Identification Laboratory
310 Worchester Avenue, BLDG 45
Joint Base Pearl Harbor-Hickam
HI 96853-5530
E-mail: joseph.hefner@jpac.pacom.mil