DOI:10.1021/jf102046b

JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY
ARTICLE

Chemical Profiling with Modeling Differentiates Wild and Farm-Raised Salmon

Kim A. Anderson,*,†,‡ Kevin A. Hobbie,‡ and Brian W. Smith†

†Department of Environmental and Molecular Toxicology, ‡Food Safety and Environmental Stewardship Program, and Oregon State University, Corvallis, Oregon 97331, United States

Classifications of fish production methods, wild or farm-raised salmon, by elemental profiles or C and N stable isotope ratios combined with various modeling approaches were determined. Elemental analysis (As, Ba, Be, Ca, Co, Cd, Cr, Cu, Fe, K, Mg, Mn, Na, Ni, P, Pb, Sr, Ti, and Zn) of wild and farm-raised salmon samples was performed using an inductively coupled plasma atomic emission spectroscopy. Isotopic and compositional analyses of carbon and nitrogen were performed using mass spectrometry as an alternative fingerprinting technique. Each salmon (king salmon, *Oncorhynchus tshawytscha*; coho salmon, *Oncorhynchus kisutch*; Atlantic salmon, *Salmo salar*) was analyzed from two food production practices, wild and farm raised. Principal component analysis (PCA) and canonical discriminant analysis (CDA) were used for data exploration and visualization. Five classification modeling approaches were investigated: linear discriminate function, quadratic discriminant function, neural network, probabilistic neural network, and neural network bagging. Methods for evaluating model reliability included four strategies: resubstitution, cross-validation, and two very different test set scenarios. Generally speaking, the models performed well, with the percentage of samples classified correctly depending on the particular choice of model and evaluation method used.

KEYWORDS: Principal component analysis; canonical discriminant analysis; elemental analysis; stable isotope; salmon (Salmonidae); food labeling; linear discriminant function; quadratic discriminant function; neural network; genetic neural network; modeling; farmed salmon; fish production methods; stable isotopes

INTRODUCTION

The U.S. Department of Agriculture (USDA) recently announced (January 12, 2009) final regulations for the mandatory country of origin labeling (COOL) program. The rule covers, among other items, wild and farm-raised fish. As well, the European Union (EU) commission regulation (2065/2001) requires informing consumers of aquaculture methods of production, including farm-raised or wild-caught (*1*). Commodities covered under COOL must be labeled with geographic origin, and, in the case of fish, the method of production, wild or farm-raised, must be specified. Although USDA intends to use supply chain audits to ensure compliance with the rule, scientific techniques that can further support and verify the rule would be especially valuable.

Recent events such as the alleged contamination of fish from China make the determination of farm-raised versus wild and geographic origin a timely scientific inquiry (*2*). *Consumer Reports* (*3*) described "wild often isn't, if you paid extra for fresh wild salmon in late fall and winter, you may have wasted your money". The health benefits of salmon are noteworthy, and they are an excellent source of many nutrients and vitamins, such as vitamin E and omega-3 oils. Numerous studies have concluded that Americans should eat more fish. However, reports suggesting some farm-raised salmon may have a higher incidence of contaminants (*4*−*6*) and risk (*7*) may make consumers wary, especially when coupled with their lack of confidence in product labeling.

The development of chemometric methods that can confirm label indications in food commodities is therefore opportune. Development of DNA-based methods to identify fish species has proven successful (*8*); however, farm-raised and wild-caught fish will often be the same species, rendering this technique inadequate at present. Stable isotope ratio analysis of foods provides a valuable probe for both abiotic and biotic origin based on geographical and biological origin discrimination. Abiotic fractionation associated with hydrological cycles and local environments give rise to informative characteristic isotope ratio signatures for empirical geographic determination. Biota fractionation reflects characteristics of their environment through physiology uptake and metabolism of stable isotopes, such as $^{13}C/^{12}C$, $^{15}N/^{14}N$, $^{18}O/^{16}O$, and $^{2}H/^{1}H$, which form compounds in organisms. Isotope ratios have been successfully used in chemical profiling methods to determine geographic origin of biota (*9*, *10*) and seafood (*11*). Various isotope ratios of fatty acids have been successfully used to distinguish origin and production methods of Atlantic salmon (*12*−*14*). The ^{13}C nuclear magnetic resonance or gas chromatography isotope ratio mass spectrometry analysis of fatty acids, however, can be labor intensive. The extraction of fatty acids for isotope analysis

*Author to whom correspondence should be addressed [phone (541) 737-8501; (541) 737-0497; e-mail Kim.Anderson@orst.edu].

 Published on Web 10/25/2010

generally requires extraction of the fish oils, saponification, fatty acid methyl esters generation, various liquid/liquid exchanges, and solvent reductions. Bulk isotopic analysis on the fish muscle tissue would present fewer chemical manipulations, although previous studies have not used bulk isotope ratios to distinguish salmon production methods.

Although elemental profiles have been used to distinguish geographic origin of plants (*15*, *16*) and mussels (*17*), this approach has not been applied to production methods for salmon. Different feeding and living environments likely would contribute to element differences in salmon, whereas hemostasis mechanisms may reduce this affect. Previous studies have not applied elemental profiles to salmon production methods.

The aim of the present study was to demonstrate the feasibility of employing models to use either bulk stable isotopes from fish fillets or elemental profiles to independently and successfully distinguish farm-raised versus wild-caught salmon. In addition, rarely are Pacific Ocean salmon included in fish production studies, and in none of the studies to date, are both wild and farm-raised Pacific salmon included with scale. We initiated a substantial-scale analysis sampling in a total of 145 Pacific Ocean salmon with authentic production origin over a 3 year period. These salmon were then modeled using five different modeling approaches with robust evaluation of each method with four different approaches. Obtaining good modeling results independently using two different types of data sets, elements, or isotopes provides converging lines of scientific evidence for salmon production method. A continuing challenge for a study of this nature is to ensure that sample size and data used to develop the predictions or models in fact represent all of the underlying variability of the population. We examined the sample size question by limiting our training data set to only 12% of the total fish sampled. This type of testing of test set sizes has not been done before in fish profile studies; this evaluation provides insight into the modeling approach and sample size question.

MATERIALS AND METHODS

Reagents. Elemental stock standard solutions were purchased form Alfa Aesar Specpure (Ward Hill, MA) and J. T. Baker Plasma grade (St. Louis, MO). Optima grade concentrated nitric acid was purchased from Fisher (Pittsburgh, PA). All water used was 18 MΩ-cm water from a Barnstead EASYpureUV D7401 (Dubuque, IA). Certified reference materials (CRMs) were purchased from the the National Institute of Standards and Technology (NIST, Gaithersburg, MD) and stored according to recommendations (*18*). CRMs included NIST 1566b oyster tissue and NIST 2977 mussel tissue for elemental analysis; NIST 8542 sucrose, NIST 8550 ammonium sulfate, and acetanilide (Carlo Erba, Italy) were used for calibration and to monitor instrument performance of compositional and isotopic analysis.

Instrumentation: Elemental Analysis. Samples were analyzed for As, Ba, Be, Ca, Cd, Co, Cu, Cr, Fe, K, Mg, Mn, Na, Ni, P, Pb, Sr, Ti, and Zn using a Varian Liberty 150 ICP-AES with a V-groove nebulizer and a Varian SPS5 autosampler system (Varian, Palo Alto, CA). The following parameters were employed: 85 psi; scan integration time, 1 s (all elements); acid flexible tubing, 0.030 mm internal diameter (i.d.); replicates, two (all elements); scan window (first order), 0.120 nm; photomultiplier, tube voltage, 650 V; plasma flow, 15 L/min; auxiliary flow, 1.50 L/min; sample uptake delay, 13 s; pump rate, 15 rpm; instrument stabilization, delay, 13 s; and rinse time, 60 s. The wavelengths selected were as follows: As, 189.042; Ba, 455.403r; Be, 313.042; Ca, 422.673r; Cd, 214.441; Co, 288.615; Cr, 267.716; Cu, 324.754; Fe, 259.940; K, 766.491r; Mg, 279.078r; Mn, 293.306; Na, 589.592r; Ni, 231.604; P, 214.914; Pb 220.353; Sr, 407.771r; Ti, 334.941r; and Zn, 206.200. Additional instrumental details are as described in Anderson and Smith (*16*).

Isotopic Analysis. Carbon ($\delta^{13}C$) and nitrogen ($\delta^{15}N$) isotope ratios and bulk compositions were measured with a Carlo Erba NA1500 elemental analyzer for separation with a CosTech "Zero Blank" autosampler (CosTech Anaytical Technologies, Inc., Valencia, CA), helium dilution and reference gas verification via Finnigan/MAT CONFLO-III (Thermo Finnigan, Waltham, MA) interface, and detection via Finnigan/MAT DeltaPlusXL mass spectrometer. Instrumental conditions are described by Roy et al. (*19*) and Perez (*10*). Reference materials from NIST and Carlo Erba were used to calibrate the isotope ratios of local reference gases. The abundance of the two most abundant isotopes of carbon and nitrogen in the sample were determined by triple-collector mass spectrometer and then reported relative to separate injections of the local reference gases. Isotopic data use the isotopic notation (δ), in parts per mil (‰) relative to the universal standards Pee Dee Belemnite (PDB) for carbon and atmospheric air (^{15}N) for nitrogen. By convention, the following equation for δ was used for carbon (and an analogous equation for nitrogen):

$$\delta^{13}C = \{[(^{13}C/^{12}C_{sample}) - (^{13}C/^{12}C_{std})]/(^{13}C/^{12}C_{std})\} \times 1000$$

Sampling and Preparation. A total of 145 fish were collected from 2006 to 2008 as close to the source as possible and directly from trusted distributors. Wild and farm-raised (n = 64, and 81 respectively) included king salmon, *Oncorhynchus tshawytscha*, coho salmon, *Oncorhynchus kisutch*, and Atlantic salmon, *Salmo salar*. Each year/season both farm-raised and wild-caught were collected with similar numbers for each salmon species. Production method, fish species, producer lot number, collection data, year, and season were known for all fish. Five different aquaculture facilities were sampled. Both farm-raised and wild-caught include king and coho salmon. All king and coho salmon were from the Pacific Ocean, and all fish samples used are common salmon in the marketplace. All fish were transported on ice, stored at 4 °C, and processed within a few days of receipt. A test portion of nominally 100−200 g of skinned filets was taken from individual fish. The fish were freeze-dried and homogenized via a grinder for both elemental and isotopic analyses. Often in the marketplace consumer's purchase a fillet, and it is difficult to distinguish species beyond that it is salmon, so we explicitly did not separate out different species of salmon because this information would not be a priori known by all consumers. By including different salmon species we have in fact made the modeling and hypothesis more challenging but more realistic of consumer needs and the study objective.

Elemental Analysis. A test portion of 0.5 g was placed into a 10 mL graduated Kimax digest tube with 1 mL of nitric acid and digested at ambient temperature overnight. Two milliliters of nitric acid was added, and the sample was placed on a heating block; the temperature was ramped at approximately 15 °C/h to 120 °C. Samples were digested until NO_2 gas evolution stopped; samples were diluted to a final volume of 10 mL with water (Barnstead 18 MΩ-cm), cooled, and vortexed. Samples were decanted into luer-lock syringes and filtered through a Pall PVDF 0.45 μm filter into sample-receiving containers until ICPAES analysis.

Isotopic Analysis. A test portion of approximately 1 mg was weighed using a Mettler Toledo UMT2 ultramicrobalance (Mettler Toledo, Columbus, OH), placed into CosTech tin capsules, and then rolled into spheres. Weighed test portions and standard reference materials were stored in CosTech 96-well plates at room temperature until analysis.

Quality Control. Quality control consisted of blanks, check standards, matrix duplicates, matrix over spikes, and certified reference materials (CRMs). Quality control samples represented about 20% of all samples analyzed. Instrument calibration consisted of three or more calibration standards; all calibration curves had $R^2 > 0.99$. Recoveries of matrix over spikes ranged from 96.5 to 118.5%, and recovery of all metals from all CRMs ranged from 92.8 to 95.2%; recoveries of individual elements used in the modeling from CRMs are presented in **Table 1**. A minimum of 10% of all fish were analyzed as duplicate for stable isotopes, with an average RPD of < 3.63% for both δ values and bulk compositions. Acetanilide was an abundant and isotopically homogeneous in-house reference material used to monitor instrumental precision; instrument runs included 5 replicates for a total of 28 replicates throughout the course of this project. Results indicated a precision of ±0.12‰ $\delta^{15}N$, ±0.28% bulk N, ±0.24‰ $\delta^{13}C$, and ±0.71% bulk C.

Statistical Analysis. Several statistical analysis methods were applied to the data. Multiple-comparison ANOVA was used in analysis by Sigma Stat for Windows, version 2.0 (Systat, Point Richmond, CA). Principal component analysis (PCA), canonical discriminant analysis (CDA), linear discriminant function analysis (LDA), and quadratic discriminant

Table 1. Average Recoveries of Elements Used in Modeling, in Standard Reference Materials and Matrix Spikes

standard reference material	n	As	Cu	Zn	Na	Ca	K	Mg	P
oyster tissue, NIST 1566b	5	93.2 ± 5.8	103.9 ± 16.2	99.6 ± 7.4	92.5 ± 2.3	98.0 ± 7.0	89.7 ± 6.6	93.3 ± 21	NC[a]
apple leaf, NIST 1515	4	BDL[b]	95.0 ± 22.6	93.6 ± 6.9	BDL	90.2 ± 5.8	89.5 ± 8.7	96.1 ± 1.9	88.3 ± 3.7
tomato leaf, NIST 1573a	2	BDL	108.9	93.7	BDL	74.3	93.0	87.2	90.3
mussel tissue, NIST 2977	4	88.5 ± 5.4	97.5 ± 21.1	93.9 ± 6.7	89.8 ± 9.4	92.3 ± 6.5	100.5 ± 6.1	NC	90.9 ± 4.3

	n	As (+ 5 ppm)	Cu (+5 ppm)	Zn (+5 ppm)	Na (+ 100 ppm)	Ca (+100 ppm)	K (+100 ppm)	Mg (+100 ppm)	P (+100 ppm)
matrix over spike	14	119.2 ± 6.2	87.8 ± 176	101.1 ± 10.4	102.3 ± 5.5	975 ± 7.6	100.8 ± 23.2	106.5 ± 6.6	106.9 ± 20.1

[a] NC, not certified. [b] BDL, below detection limit.

function analyses (QDA) were applied utilizing SAS version 9.1 (SAS Institute Inc., Cary, NC) and neural network (NN) and probabilistic neural network (PNN) analysis using NeuroShell Classifer (Ward Systems Group, Inc., V2.2, Frederick, MD). Neural network bagging (NNB) was applied using Matlab version 7.8 (TheMathWorks, Natick, MA).

Data analyses included looking at box plots of individual elements or stable isotope ratios corresponding to farm-raised and wild-caught salmon. However, better resolution for distinguishing farm-raised and wild-caught salmon was obtained by several comprehensive modeling analyses of elemental or stable isotope profiles. PCA models total data variation in terms of principal components. The first component explains maximal variation in one dimension. Subsequent components explain maximal variation unexplained by previous component combinations, and each component explains less variation than individual preceding components, tending to concentrate total variation in a few uncorrelated components. PCA does not use sample group information but is often useful for data exploration, sometimes revealing visual group clustering. Data were plotted using principal components explaining appreciable percentages of the total variation. Different pairwise combinations of principal components were explored to look for patterns tending to show differences between farmed-raised and wild-caught salmon groups.

CDA, like PCA, is a variable reduction technique. Unlike PCA, CDA includes sample group information producing components, or variables, along which differences between groups are maximized while differences within a group are minimized. CDA was used to provide data views highlighting differences between the two groups, farm-raised and wild-caught salmon.

Classification modeling methods discussed in this paper include LDA, QDA, NN, PNN, and NNB (*20*), previously described (*16*). Each trained model required fish chemical data together with sample group designation. A trained model can then be used to classify fish chemical profiles with unknown group designation. Presumably the training set contains important characteristics of larger populations. Model application and reliability were evaluated in a number of ways, discussed next.

To help gauge model reliability, up to four approaches were tried: resubstitution, cross-validation, and two test set strategies with dramatically different test set sample sizes. Resubstitution used all of the data to make models, which then classified this same data set. For cross-validation, one fish was held out during model training and then classified. This process was repeated stepwise for all samples. Applying the first test set strategy, about 88% of the data were randomly selected for modeling; the remaining 12% of the data were then classified, treated as "unknowns". Reversing set sizes, only 12% of the data were chosen for model training, when the second test set scenario was applied, whereas the other 88% formed the test set, "unknowns". Classification performances for each classification model for the production methods, farmed raised or wild caught, are discussed in the following section.

RESULTS AND DISCUSSION

Element Profile for Modeling Salmon Production Method. Eight elements were consistently above detection limits: arsenic, copper, calcium, magnesium, sodium, phosphorus, potassium, and zinc. Cadmium, chromium, lead, and nickel were often near or below detection limits. Box plots (**Figure 1**) are shown for both farm-raised and wild-caught salmon; the boundary of the box closest to zero indicates the 25th percentile, the solid lines within the box mark the mean and median, the boundary of the box farthest from zero indicates the 75th percentile. Whiskers above and below the box indicate the 90th and 10th percentiles, whereas the symbols represent the 5th and 95th percentiles. The *P* values are shown for each box plot. Simple elemental distribution plots showing clustering by geographic origin have been somewhat predictive in previous studies (*16*, *21*). In the present study, individual or pairwise plotting combinations of the individual elements did not significantly distinguish between groups of salmon production methods (data not shown). This is similar to previous studies with other target analytes; simple individual or pairwise combinations were not in themselves distinguishing (*12*, *22*) of fish production methods.

Multivariate modeling techniques were therefore applied. PCA results showed that the first three components explained most of the variability in the elemental data. Plotting component 1 versus component 2 showed good clustering of farmed-raised and wild-caught salmon groups (**Figure 2A**). CDA produced one canonical variable because there were two groups. The data were plotted along this variable using a frequency chart (**Figure 2B**) and display group differences while utilizing only the elemental data.

Five classification models were investigated with up to four evaluation strategies considered. Results are summarized in **Table 2**. The upper half of the table corresponds to elemental profile data and the lower half to isotope profile data. The data in **Table 2** indicate the correctly classified percentages for the appropriate modeling and evaluation approach. The four main labels across the top of the table describe a particular evaluation scenario employed. Generally, classification rates are good with one exception. Overall, neural network models appear to perform somewhat better than discriminant function models. Rates are generally higher corresponding to evaluations using larger training sets. The significance of the evaluation method is discussed further in a later subsection. Geographic authenticity methods using 8−25 elements and comparable modeling approaches have reported similar success rates (*9*, *10*, *16*). Smith and Watts recently reported using a suite of metals coupled with discriminant analysis modeling and found it was effective for predicting geographic authenticity of shrimp (*11*). Using over 150 chemical shifts combined with probabilistic neural networks, Aursand et al. (*23*) was able to correctly classify 47 of 52 (90%) wild-caught salmon correctly and 136 of 143 farm-raised salmon (95%). Although chemical analysis of fatty acids can be labor intensive, a Bayesian belief network based on fatty acids generally correctly classified with success rates of >95%, often with 100% success (*22*). However, modification of the fatty acid content of feed often occurs prior to slaughter in many aquaculture systems with the intent to more closely resemble wild fish. Changes in fatty acid feed composition may lead to changes in the fatty acid profile in farm-raised fish and raise recognized concerns about fish production methods based on fatty acids.

Isotope Profile Modeling for Salmon Production Method. Whereas the eight-element-based models were effective for visualization and classification modeling, isotope data also proved useful for model development and evaluation. Total bulk carbon

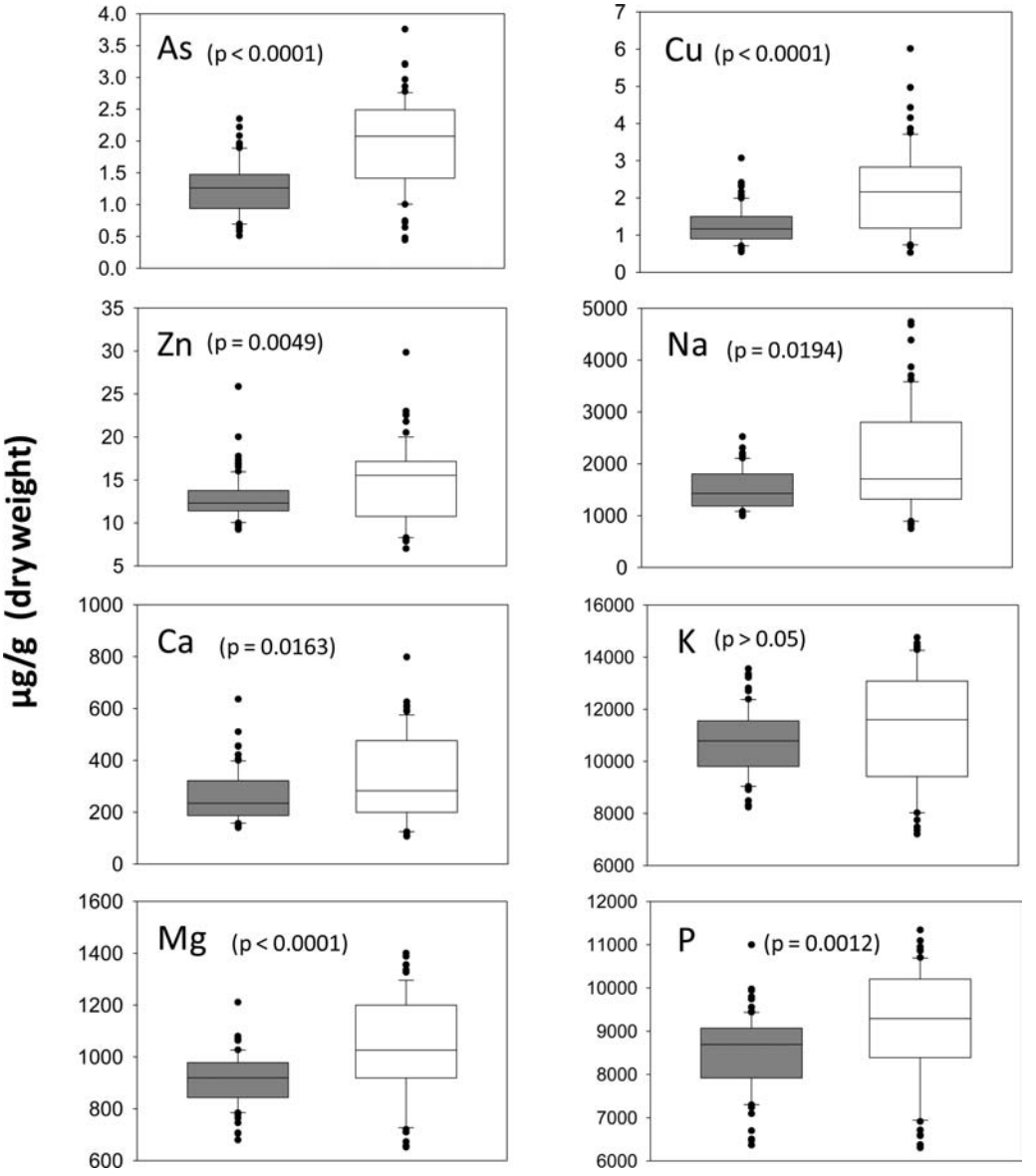

Figure 1. Element concentrations (mg/kg) and isotope ratios of farm-raised ($n = 64$) and wild-caught salmon ($n = 81$). Farm-raised salmon are indicated by the gray boxes and wild-caught are indicated by the white boxes. Significant separation was determined using a Wilcoxon signed-rank test. The boundaries of the box and whisker plots indicate the 25th and 75th (top and bottom) percentiles. The lines within the box mark the mean and the median. The whiskers above and below the box indicate the 90th and 10th percentiles. The 5th and 95th percentiles are indicated with circles.

and nitrogen and $\delta^{13}C$ and $\delta^{15}N$ values were determined. Box plots (**Figure 3**) are shown for both salmon production methods (box plot format discussed above). Stable carbon isotopes have been successfully used to distinguish the geographic origin of honey without modeling (*24*). Whereas *P* values, reported within each box plot, are < 0.05, because of overlap, classification of fish production methods by isotopes is not possible. Multivariate methods were then applied. As with the elemental data, applying PCA and CDA as variable reduction techniques provided good visualization of group distinction, as shown in **Figure 4**. Modeling the data further explored the feasibility of classifying salmon samples according to production method.

Overall, the classification modeling methods performed well, frequently correctly classifying samples with a $> 90\%$ success rate on average. Utilizing only the bulk C, N, and isotope ratios, the LDF, QDF, NN, PNN, and NNB had nominally similar success rates, ranging from 86 to 100% success. Molkentin et al. found that artificial neural network models using a combination of isotope ratios and fatty acids were effective at modeling organically farmed and wild-caught salmon for their data set (*13*). In the present study it was not necessary to combine the isotope and elemental data; independently the isotopes and elemental models classified well (**Table 2**).

Model Comparison Based on Training Set Size. Discussion of database size continues to be a critical point of food authenticity studies (*15*, *25*). Although most chemical profiling studies of fish have sample sizes similar to or smallter than our data set (*12*, *14*, *22*, *26*), sample size continues to be an essential question. As a more extreme test of our various models a smaller training set scenario was employed wherein only a fraction of the data was used to train the five models. The large training set method used nominally 125 samples (88%) of the data. In contrast, the small training set used only nominally 20 samples to make the models and then tested the fish samples. This approach represents a more rigorous measure of modeling effectiveness. The larger training set performed well with all models tested; many of the element-based models performed with 100% success, and all with $> 88\%$ success rates. In contrast, the element models constructed with the

small training set were not as successful. Although the LDA, NN, NNB, and PNN models were remarkably successful (>75%), QDA was not effective overall, with 56, 18, and 100% success for all fish, wil-caught, and farm-raised, respectively. As one would expect, the large training set models were the most successful. This general trend was also observed for the isotope-based models, although the difference between the large and small training set models was quite small. In many cases only a few percentage point difference was observed between the large and small training sets for the isotope models.

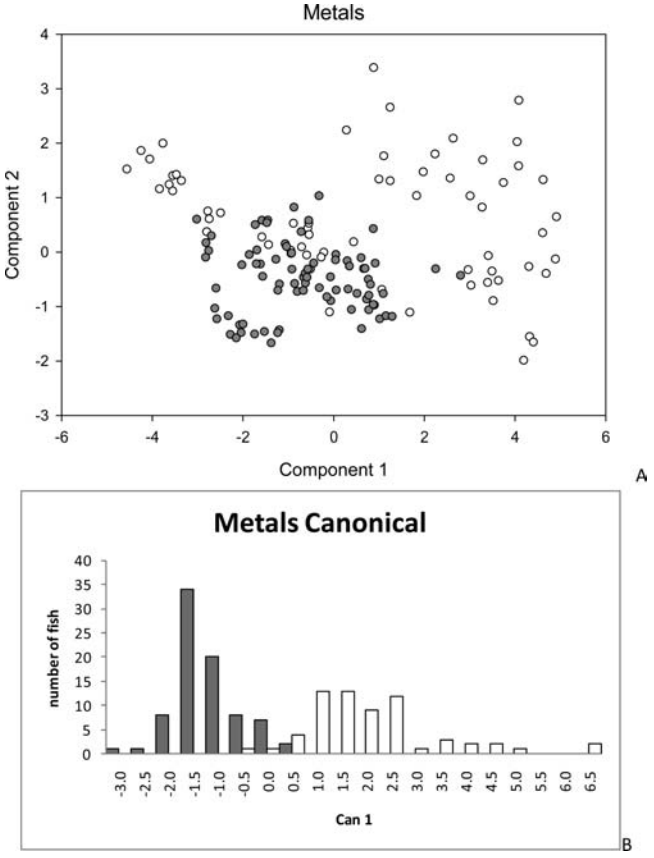

Figure 2. (**A**) Principal component 1 versus principal component 2 for the elemental chemical profile in farm-raised (gray symbols) and wild (white symbols) Pacific salmon (n = 145). (**B**) Canonical discriminant analysis (CDA) frequency histogram using CDA1 representing the two groups of farm-raised (gray bars) and wild-caught salmon (n = 145) (white bars) using all eight elements.

Whereas most of the models fared quite well with the other evaluation methods, typically with >92% success rates, the smaller modeling set presented increased challenges for all modeling approaches. Simplistically, this illustrates the need to have robust databases to make the models. In this small data set case experiment, the models generally had less success, although remarkably many models still were reasonably successful when using a mere 20-sample training set. For instance, neural networks and neural network with bagging utilizing the element-based models were both reasonably successful. The isotope-generated models were also remarkably successful with successful rates of >80%, with neural network and probablisitic models the most successful.

Model Analyte Sensitivity Analysis. Previous studies have used sensitivity analysis from neural network models to determine the relative rank of chemical components used in the model. Sensitivity analysis reports some measure of the importance of the individual chemical components and has been proposed to allow for a reduced number of inputs for further calculations. Nonlinear models depend in part on all variables used, and the concept of individual variable contribution on a dependent-variable construct is inexact. In a fatty acid and isotope salmon study based on the sensitivity analysis, the authors proposed reducing to four fatty acids without reduction of classification success (*13*). However, there are limitations to this kind of ranking when using nonlinear models; the contribution of a single variable to a model composed of many variables is an imprecise concept. One might expect that the development of numerous models would lead to different solutions with different combinations of variables, different variable rankings, and yet similar successful classifications. Aursand et al. found that their Atlantic salmon models indeed frequently yielded different variable rankings (*12*). As but one example, **Table 3** shows that rankings of the elements in two models generated in the present study were indeed substantively different. In one of the large training set models Na and K were most important, whereas copper and zinc were least important. In contrast, for one of the small training set models copper and arsenic were most important. Both models, small and large training sets, were reasonably successful in classifying wild and farm-raised salmon, on average >93 and 100%, respectively, yet the variable ranking was quite different.

Table 2. Percentage of Samples Classified Correctly by Linear Discriminant Function, Quadratic Discriminant Function, Neural Network, Probabilistic Neural Network, and Neural Network Bagging, Based on the Element Concentrations or the Stable Isotope Ratio Profiling

	resubstitution			cross-validation			training set model to test set			test set model to training set		
	all fish	wild	farmed	all fish	wild	farmed	all fish	wild	farmed	all fish	wild	farmed
Elements Only												
linear discriminant function	96	94	98	95	92	98	95	88	100	84	75	91
quadratic discriminant function	98	100	96	95	100	91	95	100	91	56	18	100
neural network	99	97	100	na[a]	na	na	100	100	100	94	91	96
probablistic neural network	99	98	100	na	na	na	95	100	91	86	75	94
neural network bagging	100	100	100	na	na	na	100	100	100	92	84	99
Stable Isotopes Only												
linear discriminant function	98	98	98	96	95	96	95	88	100	92	100	86
quadratic discriminant function	99	100	99	97	100	95	95	88	100	87	88	86
neural network	100	100	100	na	na	na	95	88	95	94	100	88
probablistic neural network	99	100	99	na	na	na	100	100	100	94	95	94
neural network bagging	100	100	100	na	na	na	89	87	90	87	79	99

[a] na, not applicable.

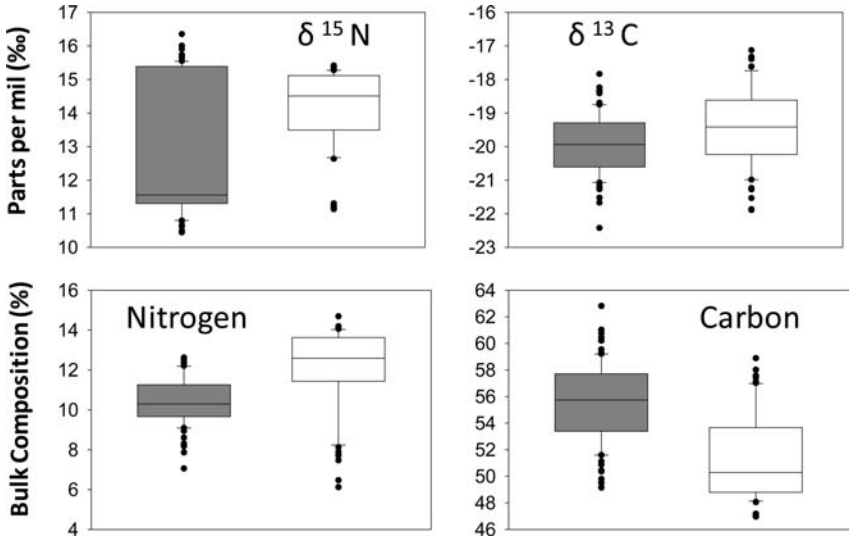

Figure 3. Isotope ratio box plots of farm-raised ($n = 64$) (white boxes) and wild-caught salmon ($n = 81$) (gray boxes). Description of box plot format provided in **Figure 1** caption.

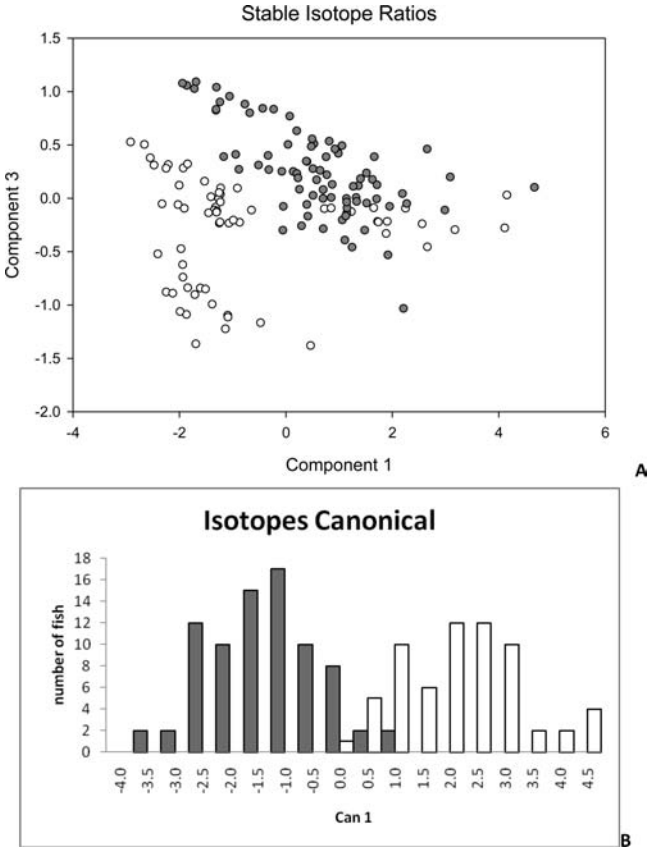

Figure 4. (**A**) Principal component 1 versus principal component 3 for the stable carbon and nitrogen isotopes ratios and bulk C/N ratio ($n = 145$). (**B**) canonical discriminant analysis (CDA) frequency histogram using CDA1 representing the two groups of farm-raised and wild-caught salmon ($n = 145$) from the stable carbon and nitrogen isotopes and bulk C/N ratios. Farm raised and wild caught denoted by gray and white, respectively.

Whereas applicable interpretation of variable ranking between models is limited, within a single model variable ranking provides some insight potentially for method simplification.

Yearly and seasonal variabilities were cautiously investigated because fish for each year or season were not always from the same farm. Multiple years and multiple seasons did display some small variations in their chemical profile (data not shown). Overall, the magnitude of the yearly or seasonal differences between farm-raised and wild salmon is small compared with the production method differences. Similar magnitude differences were reported for pistachios (*27*) and fruit (*28*); although seasonal differences were observed and could in themselves be modeled, they were small in comparison to geographic differences. We also did a cursory evaluation of fish production method by fish species. Whereas the data were encouraging, the evaluation was beyond the original study objectives, and additional salmon samples were warranted. Although year, season, and species could be parsed out into other modeling approaches, the objectives of this study were to determine a proof of concept for farming practices in consumer-based conditions. Often fish fillets are sold in markets and the salmon species is not known; also, the fish may have been frozen, so season may not be known. Therefore, our database for modeling represents a realistic marketplace, consumer-driven approach for determining farm-raised or wild salmon. By including different salmon species and different years and seasons, we have in fact made the modeling and hypothesis more challenging but more realistic of consumer needs.

Table 3. Neural Network Model: Relative Ranked Importance of Inputs Used To Classify Farm-Raised and Wild-Caught Salmon

neural model	As	Ca	Cu	K	Mg	Na	P	Zn
with large training set	3	4	7	2	5	1	6	8
with small training set	2	8	1	4	3	6	7	5

Concluding Remarks. Creating a fingerprint or unique chemical signature using trace elements or stable isotope ratios may serve as a cost-effective approach toward determining fish production methods because both approaches require minimal chemical manipulation or preparation. The two-independent-model approach based on separate data sets, elements or isotopes, provides converging lines of scientific evidence for salmon production method. This may further strengthen conclusions and potential ligation approaches. The identification of distinct elemental signature effects on wild and farm-raised salmon has not previously been described. The ease and efficiency of element and bulk stable isotope analysis make it an optimal choice for fish production determination of salmon. The databases were developed to represent a realistic consumer marketplace wherein salmon species and season may not be known. Within the framework of

this study, it appears that determining the fish production methods of salmon may be feasible through modeling of either their element profile or isotope ratio. The progression of this type of profiling study includes the addition of other geographic regions, additional seasonal variation, and additional salmon species from more locations. This information may ultimately increase food safety measures and command accountability in global food production.

ABBREVIATIONS USED

CDA, canonical discriminant analysis; COOL, country of origin labeling; LDA, linear discriminant function analysis; NN, neural network; NNB, neural network bagging; PCA, principal component analysis; PNN, probabilistic neural network; QDA, quadratic discriminant function analysis; USDA, U.S. Department of Agriculture.

LITERATURE CITED

(1) Martinez, I.; Aursand, M.; Erikson, U.; Singstad, T. E.; Veliyulin, E.; van der Zwaag, C. Destructive and non-destructive analytical techniques for authentication and composition analyses of foodstuffs. *Trends Food Sci. Technol.* **2003**, *14* (12), 489−498.

(2) Barboza, D. In China, farming fish in toxic waters. *The New York Times* 2007, p 4.

(3) The salmon scam. *Consumer Rep.* **2006**, Aug, 15.

(4) Hites, R.; Foran, J.; Carpenter, D.; Hamilton, M.; Knuth, B.; Schwager, S. Global assessment of organic contaminants in farmed salmon. *Science* **2004**, *303*, 226−229.

(5) Friesen, E. N.; Ikonomou, M. G.; Higgs, D. A.; Ang, K. P.; Dubetz, C. Use of terrestrial based lipids in aquaculture feeds and the effects on flesh organohalogen and fatty acid concentrations in farmed Atlantic salmon. *Environ. Sci. Technol.* **2008**, *42* (10), 3519−3523.

(6) Milstein, M. Fish imported with fungicide. *The Oregonian Sunday* June 5, **2005**, pp 1E, E10.

(7) Foran, J. A.; Carpenter, D. O.; Hamilton, M. C.; Knuth, B. A.; Schwager, S. J. Risk-based consumption advice for farmed Atlantic and wild Pacific salmon contaminated with dioxins and dioxin-like compounds. *Environ. Health Perspect.* **2005**, *113* (5), 552.

(8) Gil, L. A. PCR-based methods for fish and fishery products authentication. *Trends Food Sci. Technol.* **2007**, *18* (11), 558−566.

(9) Anderson, K. A.; Smith, B. W. Effect of season and variety on the differentiation of geographic growing origin of pistachios by stable isotope profiling. *J. Agric. Food Chem.* **2006**, *54*, 1747−1752.

(10) Perez, A. L.; Smith, B. W.; Anderson, K. A. Stable isotope and trace element profiling combined with classification models to differentiate geographic growing origin for three fruits: effects of sub-region and variety. *J. Agric. Food Chem.* **2006**, *54* (13), 4506−4516.

(11) Smith, R. G.; Watts, C. A. Determination of the country of origin of farm-raised shrimp (family Penaeide) using trace metal profiling and multivariate statistics. *J. Agric. Food Chem.* **2009**, *57* (18), 8244−8249.

(12) Aursand, M.; Standal, I. B.; Praël, A.; McEvoy, L.; Irvine, J.; Axelson, D. E. ^{13}C NMR pattern recognition techniques for the classification of Atlantic salmon (*Salmo salar* L.) according to their wild, farmed, and geographical origin. *J. Agric. Food Chem.* **2009**, *57* (9), 3444−3451.

(13) Molkentin, J.; Meisel, H.; Lehmann, I.; Rehbein, H. Identification of organically farmed Atlantic salmon by analysis of stable isotopes and fatty acids. *Eur. Food Res. Technol.* **2007**, *224* (5), 535−543.

(14) Thomas, F.; Jamin, E.; Wietzerbin, K.; Guerin, R.; Lees, M.; Morvan, E.; Billault, I.; Derrien, S.; Rojas, J. M.; Serra, F.; Guillou, C.; Aursand, M.; McEvoy, L.; Praël, A.; Robins, R. J. Determination of origin of Atlantic salmon (*Salmo salar*): the use of multiprobe and multielement isotopic analyses in combination with fatty acid composition to assess wild or farmed origin. *J. Agric. Food Chem.* **2008**, *56* (3), 989−997.

(15) Anderson, K. A.; Smith, B. W. Chemical profiling to differentiate geographic growing origin of coffee. *J. Agric. Food Chem.* **2002**, *50* (7), 2068−2075.

(16) Anderson, K. A.; Smith, B. W. Use of chemical profiling to differentiate geographic growing origin of raw pistachios. *J. Agric. Food Chem.* **2005**, *53*, 410−418.

(17) Costas-Rodríguez, M.; Lavilla, I.; Bendicho, C. Classification of cultivated mussels from Galicia (northwest Spain) with European Protected Designation of Origin using trace element fingerprint and chemometric analysis. *Anal. Chim. Acta* **2010**, *664* (2), 121−128.

(18) NIST. Standard Reference Materials, Certificate of Analysis. National Institute of Standards and Technology: Gaithersburg, MD, 2009; Vol. 1649b, 1650b, 1944, 1975, 2977.

(19) Roy, D. M.; Hall, R.; Mix, A. C.; Bonnichsen, R. Using stable isotope analysis to obtain dietary profiles from old hair: a case study from Plains Indians. *Am. J. Phys. Anthropol.* **2005**, *128* (2).

(20) Anderson, K. A.; Magnuson, B. A.; Tschirgi, M. L.; Smith, B. W. Determining the geographic origin of potatoes with trace metal analysis using statistical and neural network classifiers. *J. Agric. Food Chem.* **1999**, *47* (4), 1568−1575.

(21) Kelly, S.; Heaton, K.; Hoogewerff, J. Tracing the geographical origin of food: the application of multi-element and multi-isotope analysis. *Trends Food Sci. Technol.* **2005**, *16* (12), 555−567.

(22) Axelson, D. E.; Standal, I. B.; Martinez, I.; Aursand, M. Classification of wild and farmed salmon using Bayesian belief networks and gas chromatography-derived fatty acid distributions. *J. Agric. Food Chem.* **2009**, *57* (17), 7634−7639.

(23) Aursand, M.; Mabon, F.; Martin, G. J. Characterization of farmed and wild salmon (*Salmo salar*) by a combined use of compositional and isotopic analyses. *J. Am. Oil Chem. Soc.* **2000**, *77* (6), 659−666.

(24) Whilte, J. W.; Winters, K.; Martin, P.; Rossman, A. Stable carbon isotope ratio analysis of honey: validation of internal standards procedure for worldwide application. *J. AOAC Int.* **1998**, *81* (3), 610−618.

(25) Addor, F.; Grazioli, A. Geographical indications beyond wines and spirits: A roadmap for a better protection for geographical indications in the WTO trips agreement. *J. World Intellect. Property* **2002**, *5* (6), 865−897.

(26) Foran, J. A.; Hites, R. A.; Carpenter, D. O.; Hamilton, M. C.; Mathews-Amos, A.; Schwager, S. J. A survey of metals in tissues of farmed Atlantic and wild Pacific salmon. *Environ. Toxicol. Chem.* **2004**, *23* (9), 2108−2110.

(27) Anderson, K. A.; Smith, B. W. Effect of season and variety on the differentiation of geographic growing origin of pistachios by stable isotope profiling. *J. Agric. Food Chem.* **2006**, *54* (5), 1747−1752.

(28) Perez, A. L.; Smith, B. W.; Anderson, K. A. Stable isotope and trace element profiling combined with classification models to differentiate geographic growing origin for three fruits: effects of sub-region and variety. *J. Agric. Food Chem.* **2006**, *54* (13), 4506−4516.

Received for review May 28, 2010. Revised manuscript received October 1, 2010. Accepted October 11, 2010. We thank the Oregon Agriculture Research Foundation for partial funding of this research.