Combatting seafood fraud: Application of machine learning to develop chemometric differentiation tools using metal isotope ratios and chemical contaminants in eastern oysters

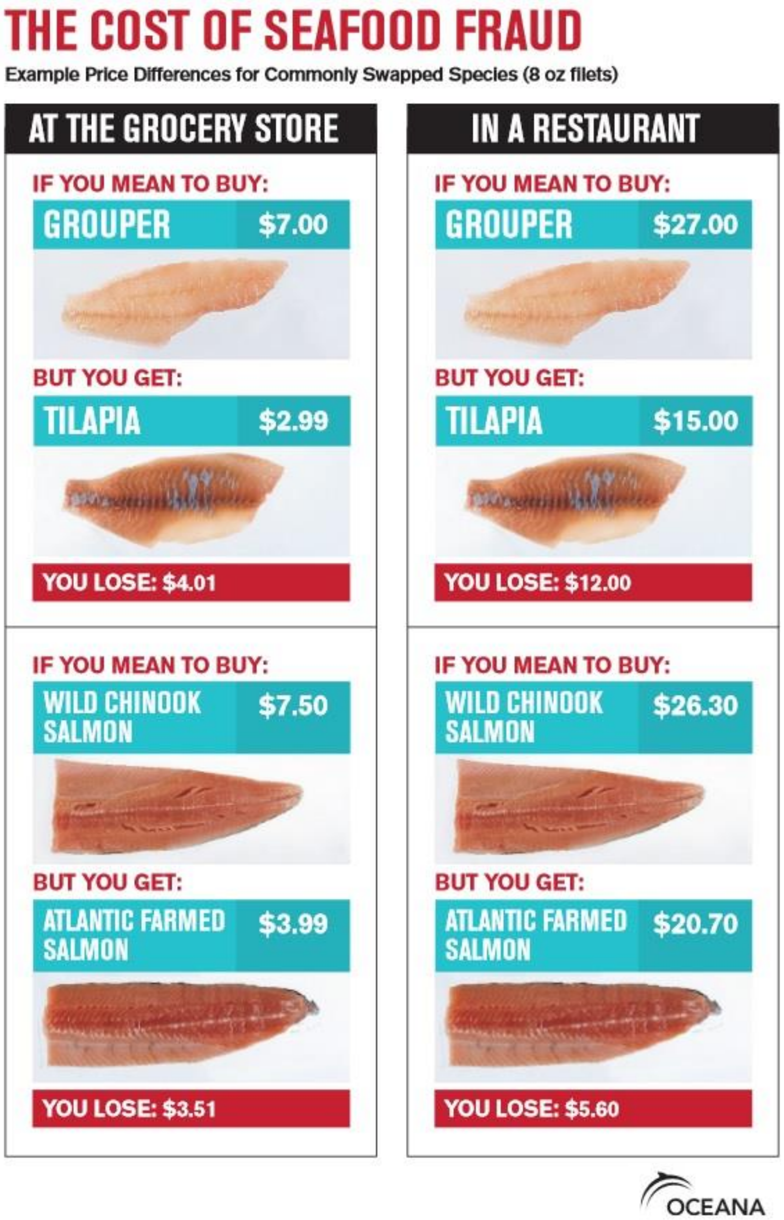
Elizabeth Crowther, Steven Christopher, Michael Winchester, Nathan Mahynski, Debra Ellisor

Collaboration Among NIST Groups

- Inorganic Chemical Metrology Group → Elizabeth, Steve, Mike
- Chemical Informatics Group → Nate
- Biospecimen Science Group → Deb

Seafood Fraud

- Examples include mislabeling of species, weight, geographic origin, wild-caught vs. farm-raised sourcing
- 36% of ~9,000 seafood products across 30 countries were found to be mislabeled [Leahy 2021]
- Loss of ~$26-50 billion per year from the global food industry [Sumaila et al. 2020]
- Interferes with consumers' decisions, poses health risks, undermines environmental conservation and management, negatively affect marine ecosystems



Goal – Chemometric Modeling

- Eastern oysters have distinct flavors, shell shapes, & textures depending on their geographic origin – influenced by salinity and temperature
- Goal is to determine which types of chemical analyses (metals, organics, isotopes) could be used to determine the geographic origin of oysters
- Develop chemometric models using previous measurements of metal & organic contaminant concentrations, plus new measurements of metal isotope ratios within oyster tissue collected from known locations

Mussel Watch Program

- Part of NOAA's National Status and Trends (NS&T) Program

- 1986-2012, over 36,000 eastern oysters collected from 123 sites across 61 waterbodies

- ~1,700 oyster tissue composites, each analyzed for ~140 contaminants to assess long-term temporal & spatial trends

- ~174 samples frozen at the NIST Biorepository

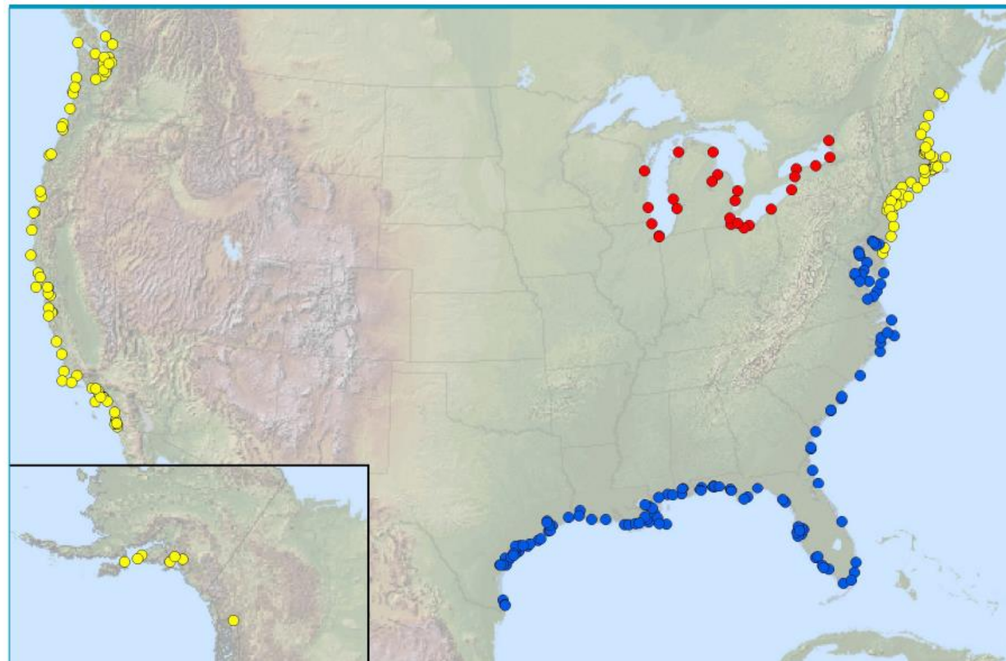
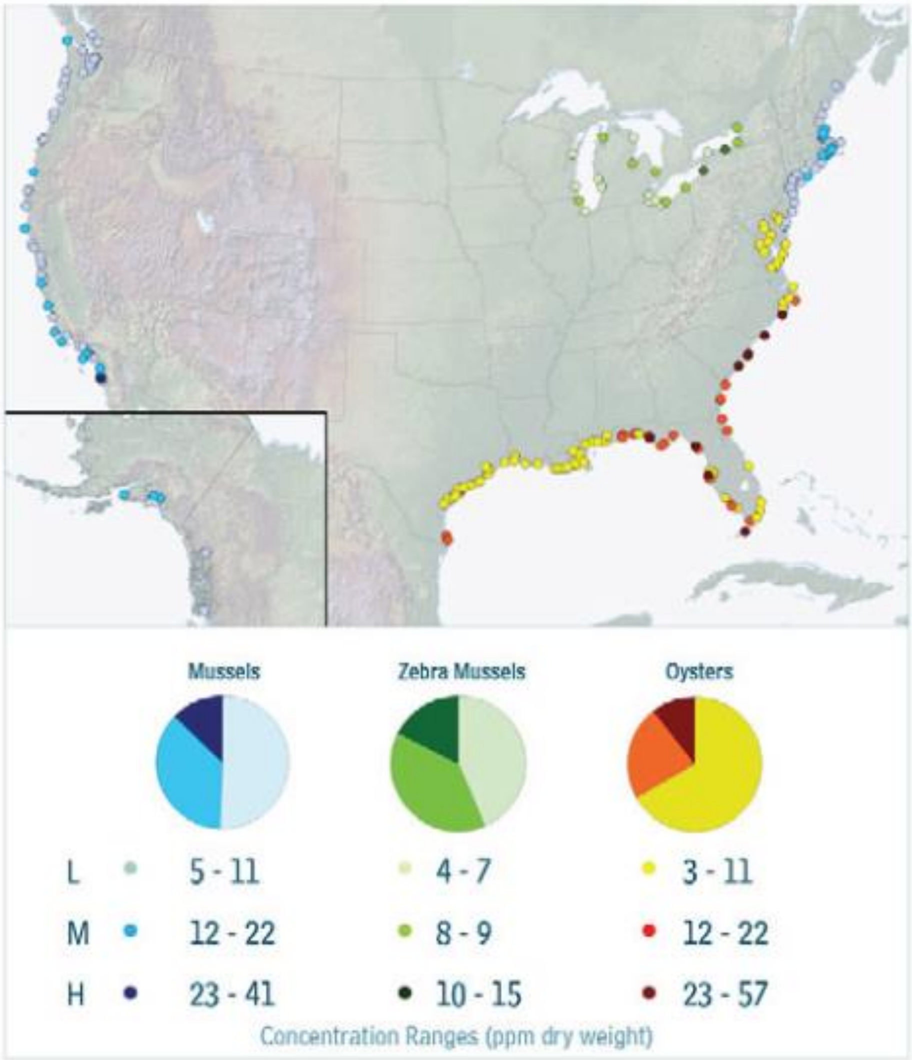


FIGURE 1. Distribution of oysters (*Crassostrea virginica*), mussels (*Mytilus species*), and zebra mussels (*Dreissena species*) collected and measured as part of the Mussel Watch Program.

MUSSELS (*MYTILUS SPECIES*)
OYSTERS (*CRASSOSTREA VIRGINICA*)
ZEBRA MUSSELS (*DREISSENA SPECIES*)

Kimbrough et al. 2008

Mussel Watch Program



- Variation in contaminant concentrations in oyster tissue along Gulf & eastern coasts of the US
- Spatial variation is necessary for chemical fingerprinting
- But so is temporal stability…

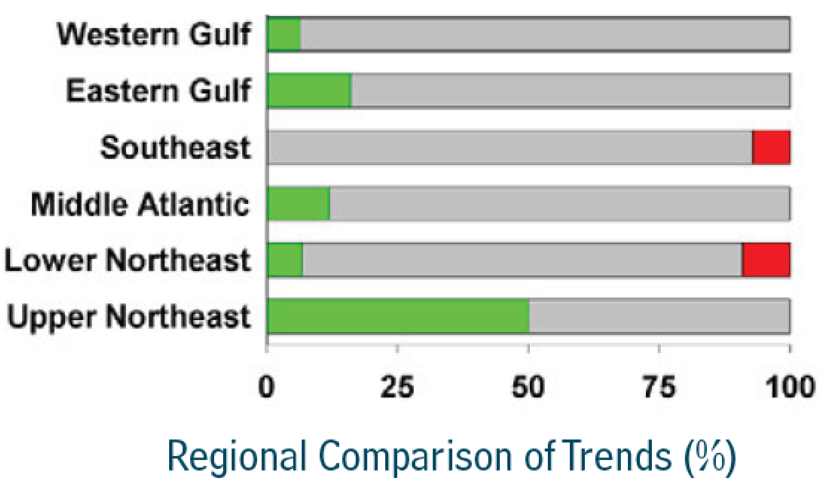
Kimbrough et al. 2008

Mussel Watch Program

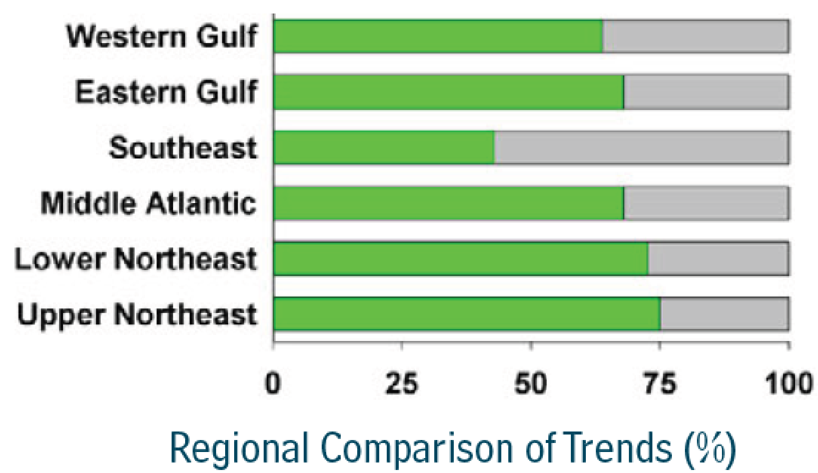
- Metal concentrations in oyster tissue were generally stable, except in localized areas where they either increased or decreased
- Organic contaminant concentrations in oyster tissue tended to decrease over time

Gray = stable
Green = decreased
Red = increased
(1986-2005)
Kimbrough et al. 2008

Changes in metal concentrations

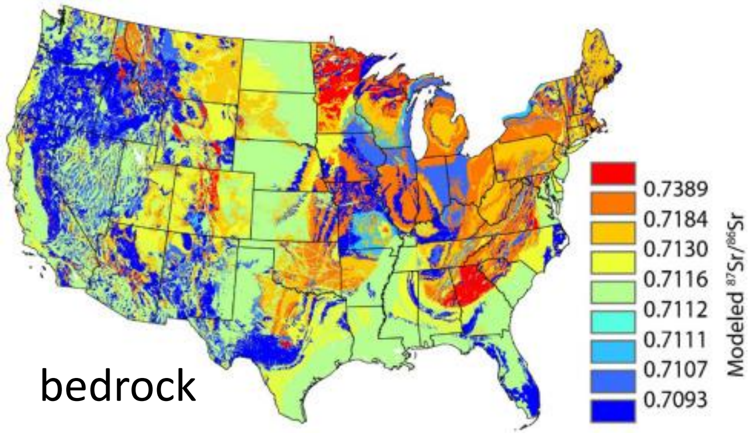


Changes in organic contaminant concentrations

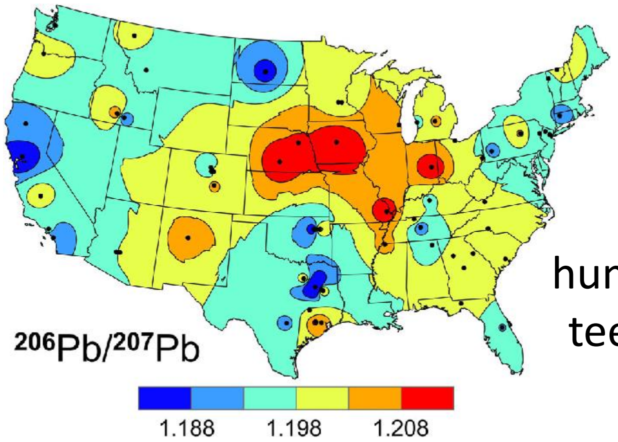


Isotopic Measurements

- Metal isotope ratios may be both spatially variable & stable over time
- Strontium (Sr) → although ocean water has a uniform Sr isotopic composition, stream water along the coast is isotopically variable due to weathering of local bedrock (Bataille & Bowen 2012)
- Lead (Pb), zinc (Zn), & copper (Cu) → isotope ratios in oysters would reflect local anthropogenic contamination, with Zn concentrations in oyster tissue being especially high due to bioaccumulation



bedrock



human teeth

Fresh & Archived Samples

Fresh oysters collected in Feb 2023 by:

- Texas Parks & Wildlife Dept
- Louisiana Dept of Wildlife & Fisheries
- Choctawhatchee Basin Alliance (FL)
- Florida Fish & Wildlife Conservation Commission
- South Carolina Dept of Natural Resources
- Chincoteague Bay Field Station (VA)

Sampling sites and dates of oyster composite samples for Sr, Pb, Zn, & Cu isotope analysis.

State	General location	Specific location	NS&T site code	Oyster sample collection dates
TX	Galveston Bay	Confederate Reef	GBCR	1986, 1987, 1992, 2007, 2023
LA	Breton Sound	Bay Gardene	BSBG	1986, 1991, 2005, 2006, 2008, 2023
FL	Choctawhatchee Bay	Off Santa Rosa	CBSR	1986, 1991, 1992, 2006, 2008, 2023
FL	Matanzas River	Crescent Beach	MRCB	1988, 2005, 2007, 2023
SC	Charleston Harbor	Fort Johnson	CHFJ	1986, 2008, 2023
SC	Charleston Harbor	Shutes Folly Island	CHSF	1987, 2023
SC	Winyah Bay	Lower Bay	WBLB	1990, 2006, 2008, 2023
VA	Chincoteague Bay	Chincoteague Inlet	CBCI	1988, 2006, 2008, 2023

Lab Methods

- Oysters were shucked, and tissue was cryogenically homogenized
- <u>Next steps</u>: acid digestion → metal concentration analysis (triple quad ICP-MS) → resin column separation → recovery check (triple quad ICP-MS) → isotopic analysis (MC-ICP-MS)

Chemometric Modeling

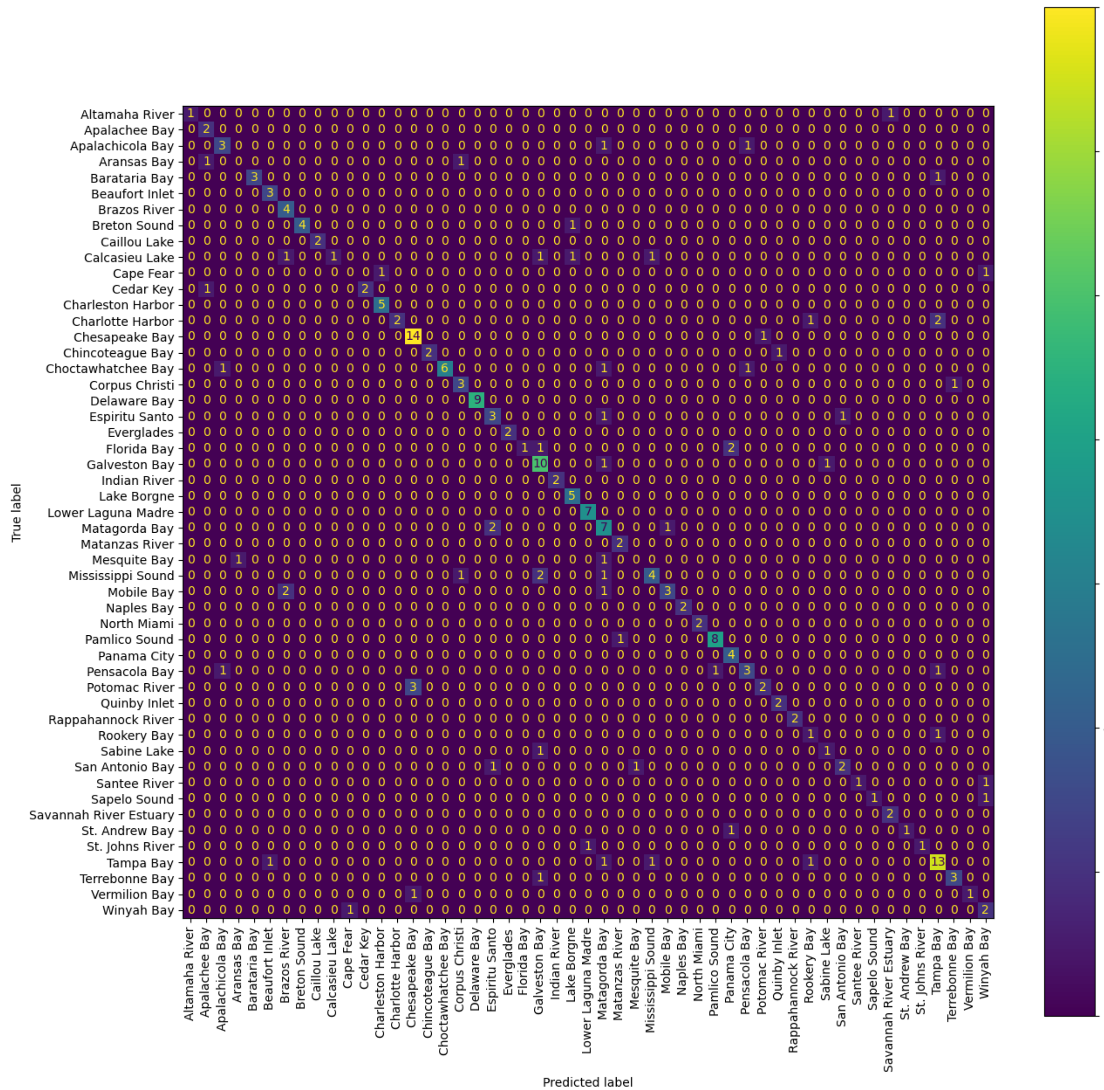
- Filter data to eliminate “holes” → choose thresholds for number of measurements per site & per year; fraction of data below detection limit; number of times particular measurements were not performed, etc.

- Option to calculate Jensen-Shannon divergence (JSD) for each analyte at each location → Higher JSD values signify analyte concentrations or isotope ratios that stand out at a particular sampling site compared to other sites (e.g., due to environmental contamination)

- Potential models to try [Hastie et al. 2009; Kamath and Liu 2021]
  - partial least-squares discriminant analysis (PLS-DA)
  - soft independent modelling by class analogy (SIMCA)
  - random forest
  - gradient-boosted tree
  - explainable boosting machine
  - support vector classifier
  - k-nearest-neighbor

Preliminary Model

- Filtered Mussel Watch Program data to eliminate “holes”
- Left with 11 metals and 2 organics
- Preliminary random forest model accurately predicted the geographic origin 72% of the time
- Next filter & test organics separately from metals, & add isotope data



Application & Significance

- Combining data from long-term monitoring efforts with new measurements on both archived and freshly collected samples
- Leverages government infrastructure and long-term investments in sample collection, analysis, and archival storage
- Applicable to the U.S. oyster industry, and would provide a framework for developing models applicable to other types of seafood
- Combatting seafood fraud and illegal, unreported, & unregulated (IUU) fishing is a top priority for the U.S. within the 2020 Maritime SAFE Act
- This Act outlines ways in which federal agencies will "expand the role of technology for combatting IUU fishing and seafood fraud" and develop methods of sharing information and data on IUU fishing and seafood fraud using "big data analytics and machine learning".

Questions?