Title: Environmental Effects on Pelagic Fish Using Generalized Additive Models: Clupea harengus Case Study
Category: Earth and Environmental Sciences

William Hu
Research Plan

a. RATIONALE:

The Gulf of Maine (GOM) is one of the most biologically productive regions in the world, supporting 75,000 jobs and $1.4 billion in annual net revenue. The GOM Atlantic herring stock is particularly important because it supports major commercial fisheries along with coastal communities, is the main source of prey and bait for groundfish, and satisfies a crucial role within the ecosystem by connecting the microorganisms with the top predators [1, 2]. However, excessive fishing from 2006-2013 caused a significant recent decline in biomass and recruitment, forcing a fishery closure in the fall of 2018 along with significant catch reductions. The recent decline can be attributed to inadequate stock forecasts, which are based only on biomass and recruitment and are unreliable due to the high amount of unexplained variance within those variables [3]. Much of this unexplained variance can be accounted for with environmental variables. The Gulf of Maine is warming exceptionally rapidly due to a northward shift in the Gulf Stream that has allowed water temperatures to increase an average of 0.23°C/year since 2004, with an additional 3°C of warming expected by 2050, which is significantly reducing suitable habitat for herring [4, 5]. Climate change will significantly affect Atlantic herring as the growth and mortality rates, spawning and feeding patterns, and recruitment of herring are strongly dependent on environmental variables [6-9]. However, climate variables are neglected in stock forecasts because traditional regression techniques cannot account for the complexity of the relationships between environmental variables and herring biomass [10, 11].

Nevertheless, generalized additive models use piecewise polynomial functions, making them well-suited for explaining these complex relationships. Their effectiveness can be attributed to their flexibility as unlike traditional regression techniques, they can compare different types of functions with varying degrees of complexity by penalizing for the overall degree of concavity of the function. A key advantage of these models is that they allow the researcher to control the degree of complexity by adjusting the weight assigned to the penalty [12]. Generalized additive models are not widely used because they are very difficult to develop and evaluate, but they are very practical and powerful tools, as demonstrated by [13], where a generalized additive model effectively forecasted spatial distributions of river herring through 2100.

b. RESEARCH QUESTION(S)/HYPOTHESIS(ES)/ENGINEERING GOAL(S)/EXPECTED OUTCOMES:

RESEARCH QUESTION:

How can Gulf of Maine (GOM) Atlantic herring (*Clupea harengus*) spawning stock biomass (SSB) forecasts be improved and made more accurate?

HYPOTHESIS:

Lagged environmental data (including data on sea surface temperatures, bottom temperature, surface salinity, salinity at 100m of depth, salinity at 200m of depth, the Atlantic Multidecadal Oscillation, the El Niño Southern Oscillation, the East Pacific-North Pacific Oscillation, the Gulf Stream North Wall Index, the North Atlantic Oscillation, Pacific Decadal Oscillation, and the Pacific North American Index) can be used to predict fluctuations in GOM Atlantic herring SSB.

ENGINEERING GOALS:

Goal: Improve GOM Atlantic herring SSB forecasts by accounting for environmental variables.

Objective 1: Find environmental variables with the strongest relationship with SSB.

Objective 2: Develop climate-based generalized additive models (GAMs) capable of accurately forecasting GOM Atlantic herring SSB while accounting for environmental variables.

Objective 3: Use the top models to predict GOM Atlantic herring SSB through 2021.

EXPECTED OUTCOMES:

Sea surface temperature, the North Atlantic Oscillation, and the Gulf Stream North Wall index will have the most significant correlations with and explain the most variance in SSB.

c. Procedures:

Step 1: Data Collection

1. Spring and fall GOM Atlantic herring SSB, sea surface temperature (SST), surface salinity (SS), bottom temperature (BT), and bottom salinity (BS) data will be downloaded from the NEFSC bottom trawl surveys [14].
2. Monthly time series data for the Atlantic Multidecadal Oscillation (AMO), El Niño Southern Oscillation (ENSO), East Pacific-North Pacific Oscillation (EP-NP), Gulf Stream North Wall Index (GSNW), North Atlantic Oscillation (NAO), Pacific Decadal Oscillation (PDO), and Pacific North American Index (PNA) will be downloaded from NOAA for the period 1983-2018 [15-17].
3. Spring and fall seasonal averages will be calculated for SSB, SST, SS, and BS in the GOM [15-17].
4. 3-month running means (n=12) of large-scale environmental indices will be calculated for large-scale climate indices [11, 15-17].

Step 2: Variable Analysis

1. For each environmental variable, Pearson's correlations will be calculated between spring and fall SSB and lagged seasonal means of environmental variables using lags of 0-5 year.

a. To clarify, a lag of k years indicates the response variable from a given year, x, is predicted based on the explanatory variables in the year x-k.

b. The use of lags can help explain delayed effects of environmental variables on Atlantic herring and effects on different life stages.

2. For each environmental variable with a significant correlation ($p<0.05$), the lagged seasonal mean of that variables with the strongest correlation with spring SSB will be used for spring model development. The same procedure will be used to choose variables for fall models [11].

Step 3: Model Development

1. Climate-based GAMs will be developed for spring and fall SSB using past SSB and any combination of 3 environmental variables ($p<0.05$) with the Mixed GAM Computation Vehicle package in R.
2. Control models will also be formed using only past SSB to set a baseline for the accuracy of models that do not use environmental variables [11].
    a. The climate-based models can not be compared with the models currently used in stock forecasts because the stock forecasts models are proprietary.

Step 4: Model Evaluation

1. The Akaike information criterion (AIC) and restricted maximum likelihood (REML) values, among other criteria, will be used to gauge model accuracy.
2. The predicted values from the experimental models will be compared through jackknifing, hindcasting, and/or other methods.

Step 5: Short-term Forecasts

1. The top 4 climate-based models based on the model evaluation will be fed available data to predict herring SSB through 2021. This will be possible since the models will use lags.
2. Any variable that has a lag of less than 3 years will be set to its value in 2018 [19].

c. Risk and Safety:

No risks. All work was done on a Lenovo IdeaPad 100-15IBD, model 80QQ.

c. Data Analysis:

The gam.check function will ensure the validity of the models based on residual plots. The summary function will evaluate the significance of all terms in each model. The developed models will also be compared with models using a subset of its coefficients on R through the Anova and F tests.

Subject-specific Guidelines:

1. Human participants research:

N/A

2. Vertebrate animal research:

N/A

3. Potentially hazardous biological agents research:

N/A

4. Hazardous chemicals, activities & devices:

N/A

Addendum: Post Summary (changes to procedures and bibliography are bolded)

c. Procedures:

Step 1: Data Collection

1. **Spring and fall GOM Atlantic herring SSB, sea surface temperature (SST) and bottom temperature (BT) data from the spring and fall NEFSC bottom trawl surveys were requested from NOAA. The GOM region was defined by the following strata: 22-30, 34, 36-40, 56, 59-61, 64-66 [14].**
2. Spring and fall seasonal average were computed for SSB, SST, and BT.
3. **Salinity data were acquired from the Canadian Fisheries Hydrographic database for the period 1983-2018 in the GOM region, defined by the following strata: SS25, SS26, SS36-SS53, SS56, SS57 [20].**
4. **Since herring reside at depths of 1m-200m, queries were run at depths of 1m, 100m+/- 5m, and 200m +/- 5m, which corresponded to surface salinity (SS), salinity at 100 meters (S100) and salinity at 200 meters (S200) [9, 20].**
5. **Spring (March-April) and fall (September-October) averages were computed for each depth [9, 20].**
6. Data on large-scale environmental indices, including the monthly time series data for the period 1988-2018 for the Atlantic Multidecadal Oscillation (AMO), El Niño Southern Oscillation (ENSO), East Pacific-North Pacific Oscillation (EP-NP), Gulf Stream North Wall Index (GSNW), North Atlantic Oscillation (NAO), Pacific Decadal Oscillation (PDO), and Pacific North American Index (PNA) were **downloaded from the NOAA**

Earth System Research Laboratory within the Physical Sciences Division for the period 1983-2018 [15-17].

7. 3-month running means (n=12) of large-scale environmental indices were calculated for large-scale climate indices [11, 15-17].

Step 2: Variable Analysis

3. For each environmental variable, Pearson's correlations were calculated between spring and fall SSB and lagged seasonal means of environmental variables using lags of 0-5 year.
   a. To clarify, a lag of k years indicates the response variable from a given year, x, is predicted based on the explanatory variables in the year x-k.
   b. The use of lags can help explain delayed effects of environmental variables on Atlantic herring and effects on different life stages.
4. For each environmental variable with a significant correlation ($p<0.05$), the lagged seasonal mean of that variables with the strongest correlation with spring SSB was used for spring model development. The same procedure was used to choose variables for fall models [11].

Step 3: Model Development

1. GAMs were developed for spring and fall SSB using past SSB and any combination of 3 environmental variables ($p<0.05$) with the Mixed GAM Computation Vehicle package in R.
2. **A total of 165 spring and 120 fall models were formed using various combinations of the chosen environmental variables.**

3. Control models were also formed using only past SSB to set a baseline for the accuracy of models that do not use environmental variables [11].
    a. The climate-based models could not be compared with the models currently used in stock forecasts because the stock forecasts models are proprietary.
4. **The k-value, or dimensionality of the smooth functions, was set to 7 for each variable based on herring lie history (herring stop spawning around age 7) [18].**
5. **Penalization was implemented with the REML-based double penalty method as it was shown by [18] to be the most effective penalization technique.**
6. **Models were ranked based on restricted maximum likelihood (REML) values [18].**

Step 4: Model Evaluation

1. **Hindcasting was performed on the 6 models for each season with the lowest restricted maximum likelihood (REML) values to test their accuracy in predicting future biomass.**
    a. **Models were refit to data from 1988-2012.**
    b. **The refitted models were used to predict SSB from 2013-2018.**
    c. **The 6 predicted values from each model were compared with the actual observed SSB [19].**
2. **Model accuracy was measured based on R^2 and root mean square error values.**

Step 5: Short-term Forecasts

1. The top 4 climate-based models based on the model evaluation were fed available data to predict herring SSB through 2021. This was possible since the models used lags.
2. Any variable that has a lag of less than 3 years will be set to its value in 2018 [19].

c. Data Analysis:

The gam.check function ensured the validity of the models based on residual plots. The summary function evaluated the significance of all terms within the models. **The amount of deviance in biomass explained was calculated in 2 different ways: The adjusted R^2 of the model fits and the R^2 values from the hindcast.**

d. BIBLIOGRAPHY:

[1] S. Lucey and J. Nye, "Shifting species assemblages in the Northeast US Continental Shelf Large Marine Ecosystem", *Marine Ecology Progress Series*, vol. 415, pp. 23-33, September 2010. [Online]. Available: https://doi.org/10.3354/meps08743

[2] A. J. Read and C. R. Brownstein, "Considering Other Consumers: Fisheries, Predators, and Atlantic Herring in the Gulf of Maine", Conservation Ecology, vol. 7, no. 1, 2003. [Online]. Available: http://www.consecol.org/vol7/iss1/art2/.

[3] National Oceanic and Atmospheric Administration, "Magnuson-Stevens Act Provisions; Fisheries of the Northeastern United States; Atlantic Herring Fishery; Adjustment to Atlantic Herring Specifications and Sub-Annual Catch Limits for 2019", *Federal Register*, vol. 84, no. 27, pp. 2760-2767, February 2019. [Online]. Available: https://www.federalregister.gov/documents/2019/02/08/2019-01658/fisheries-of-the-northeastern-united-states-magnuson-stevens-act-provisions-atlantic-herring-fishery.

[4] A. Thomas et al., "Seasonal trends and phenology shifts in sea surface temperature on the North American northeastern continental shelf", *Elem Sci Anth*, vol. 5, no. 0, pp. 48, August 2017. [Online]. Available: https://doi.org/10.1525/elementa.240.

[5] S. Smith, S. Cunniff, N. Peyronnin, and J. Kritzer, "Prioritizing coastal ecosystem stressors in the Northeast United States under increasing climate change", *Environmental Science & Policy*, vol. 78, pp. 49-57, December 2017. [Online]. Available: https;//doi.org/10.1016/j.envsci.2017.09.009.

[6] T. Brunel and M. Dickey-Collas, "Effects of temperature and population density on von Bertalanffy growth parameters in Atlantic herring: a macro-ecological analysis", *Marine*

Ecology Progress Series, vol. 405, pp. 15-28, April 2010. [Online]. Available: https://doi.org/10.3354/meps08491.

[7] V. C. Anthony and M. J. Fogarty, "Environmental Effects on Recruitment, Growth, and Vulnerability of Atlantic Herring (Clupea harengus harengus) in the Gulf of Maine Region", *Canadian Journal of Fisheries and Aquatic Sciences*, vol. 42, no. S1, pp. s158-s173, 1985. [Online]. Available: https://doi.org/10.1139/f85-271.

[8] M. Sinclair and M. Tremblay, "Timing of Spawning of Atlantic Herring (Clupea harengus harengus) Populations and the Match–Mismatch Theory", *Canadian Journal of Fisheries and Aquatic Sciences*, vol. 41, no.7, pp. 1055-1065, 1984. [Online]. Available: https://doi.org/10.1139/f84-123.

[9] US Department of Commerce, "Atlantic Herring, Clupea harengus, Life History and Habitat Characteristics", 2005.

[10] K. Drinkwater, "The response of Atlantic cod (Gadus morhua) to future climate change", *ICES Journal of Marine Science*, vol. 62, no. 7, pp. 1327-1337, January 2005. Available: https://doi.org/10.1016/j.icesjms.2005.05.015.

[11] W. Sydeman, M. García-Reyes, A. Szoboszlai, S. Thompson, and J. Thayer, "Forecasting herring biomass using environmental and population parameters", *Fisheries Research*, vol. 205, pp. 141-148, September 2018. [Online]. Available: https://doi.org/10.1016/j.fishres.2018.04.020.

[12] S. Wood, *Generalized additive models*. Boca Raton [etc.]: CRC Press - Taylor &Francis Group, 2017.

[13] P. Lynch et al., "Projected ocean warming creates a conservation challenge for river herring populations", *ICES Journal of Marine Science*, vol. 72, no. 2, pp. 374-387, 2014. Available: https://doi.org/10.1093/icesjms/fsu134.

[14] "NEFSC Ecosystems Surveys Branch: Resource Survey Reports", *Nefsc.noaa.gov*, 2019. [Online]. Available: https://nefsc.noaa.gov/femad/ecosurvey/mainpage/resource_survey_reports.html.

[15] "Climate Prediction Center - Northern Hemisphere Teleconnection Patterns", Cpc.ncep.noaa.gov, 2019. [Online]. Available: https://www.cpc.ncep.noaa.gov/data/teledoc/telecontents.shtml.

[16] P. Team, "ESRL : PSD : Climate Indices: Monthly Atmospheric and Ocean Time Series", Esrl.noaa.gov, 2019. [Online]. Available: https://www.esrl.noaa.gov/psd/data/climateindices/list/.

[17] "Climate Prediction Center - Monitoring & Data Index", Cpc.ncep.noaa.gov, 2019. [Online]. Available: https://www.cpc.ncep.noaa.gov/data/

[18] G. Marra and S. Wood, "Practical variable selection for generalized additive models", *Computational Statistics & Data Analysis*, vol. 55, no. 7, pp. 2372-2387, July 2011. [Online]. Available: https://doi.org/10.1016/j.csda.2011.02.004.

[19] H. Xu, T. Miller, S. Hameed, L. Alade and J. Nye, "Evaluating the utility of the Gulf Stream Index for predicting recruitment of Southern New England-Mid Atlantic yellowtail flounder", *Fisheries Oceanography*, vol. 27, no. 1, pp. 85-95, 2017. Available: https://doi.org/10.1111/fog.12236.

[20] "Fisheries and Oceans Canada", *Dfo-mpo.gc.ca*, 2019. [Online]. Available: https://www.dfo-mpo.gc.ca/science/hydrography-hydrographie/data-management-eng.html.