Forecasting Alaska Salmon Migration Timing

With Machine Learning Methods

Bryan Melchert - Purdue University Fort Wayne
Research Advisor: Dr. Yihao Deng

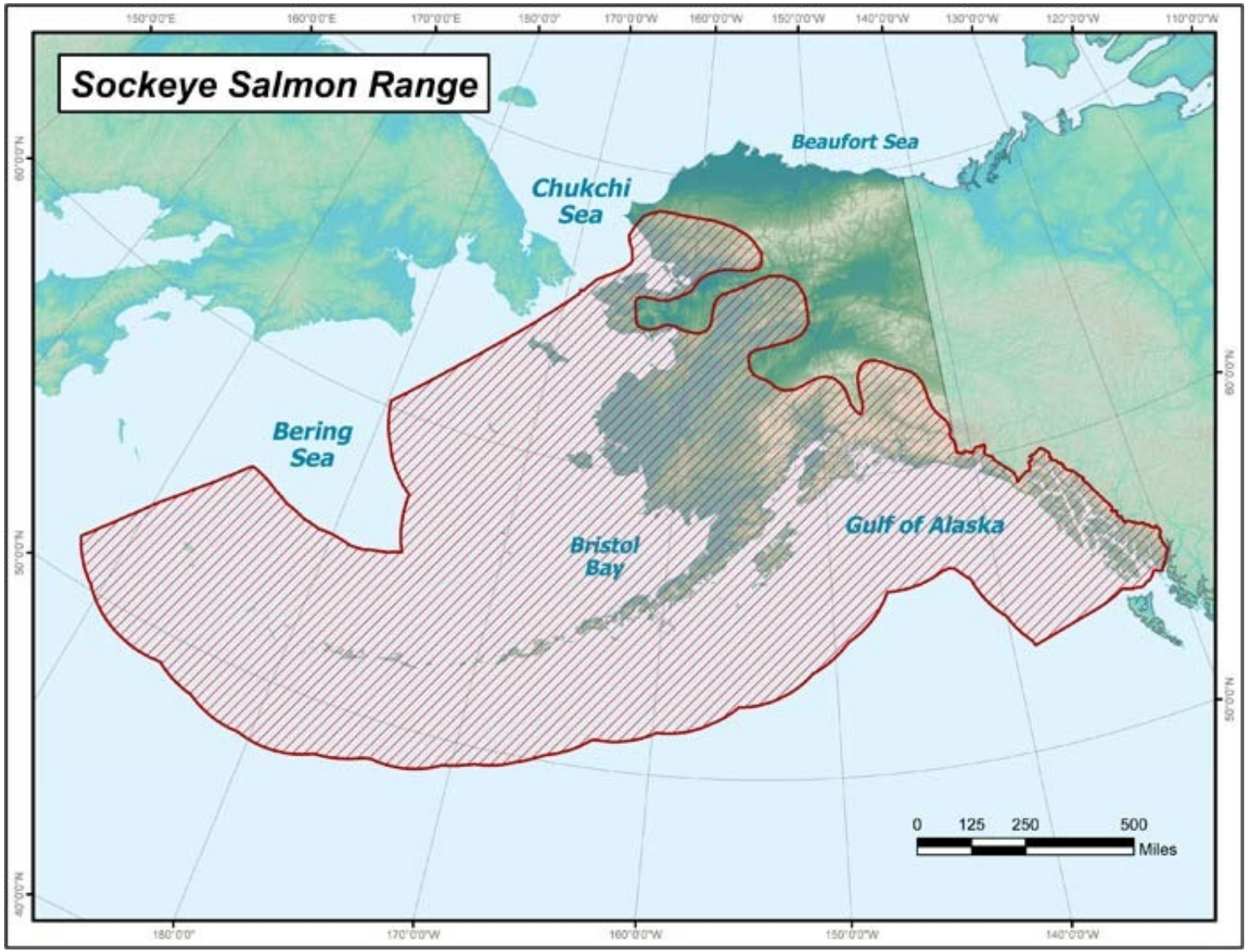
Project Abstract

The goal is to develop an accurate migration run timing forecast for the Bristol Bay, AK sockeye salmon fishery using machine learning and classical statistical models.

This $1.5B annual, 125 year sustainable fishery is both the largest sockeye salmon and wild salmon run in the world, supporting 14,000 jobs. Although annual abundance forecasts have been publicly released since the 1970's by the Alaska Dept. of Fish & Game, no annual run timing forecasts have been developed other than in 1997, 2004, and 2009 by Dr. Greg Ruggerone. This is despite significant demand and potential benefit to fishery managers, fish processors, and the fisherman themselves.

Domain Knowledge

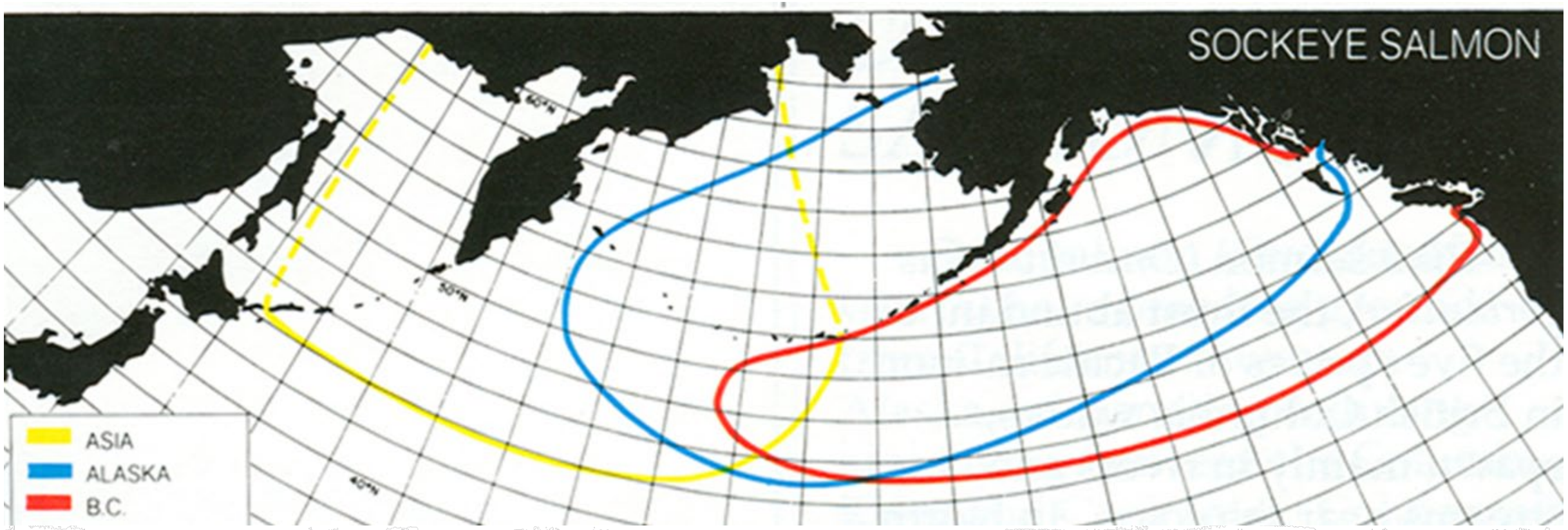
- 50+ scholarly papers read
- Sockeye spend 1-2 years in freshwater, 2-3 years adult life in the ocean
- Migration genetically/ photoperiod initiated at same time annually
- Timing variation to natal river dictated by factors experienced on return route
- Oceanographic & meteorological variables as primary determinants
- Correlations with other fishery's timing
- Historical Bristol Bay data factored in via biological sampling & annual run stats



Source: Adfg.Alaska.gov

Variables

Variable	Type	Description
Bristol Bay Fisheries	*Target*	5 natal rivers: Egegik, Naknek-Kvichak, Nushagak, Ugashik, Togiak
Other Fisheries	*Predictor*	12 salmon & herring fisheries in Alaska, Canada, Washington, & Russia
Port Moller Test Fishery	*Predictor*	Biologist sampling data from intercepted, returning salmon before fishery
Oceanographic	*Predictor*	Sea surface temp, Currents, Salinity, Evaporation, Sea pressure, Upwelling
Meteorological	*Predictor*	Air temp, Wind, Precipitation, Humidity, Lunar distance (tides), Dew Point, Cloudiness
Ocean Indices	*Predictor*	Pacific Decadal Oscillation, North Pacific Index, El Nino Southern Oscillation

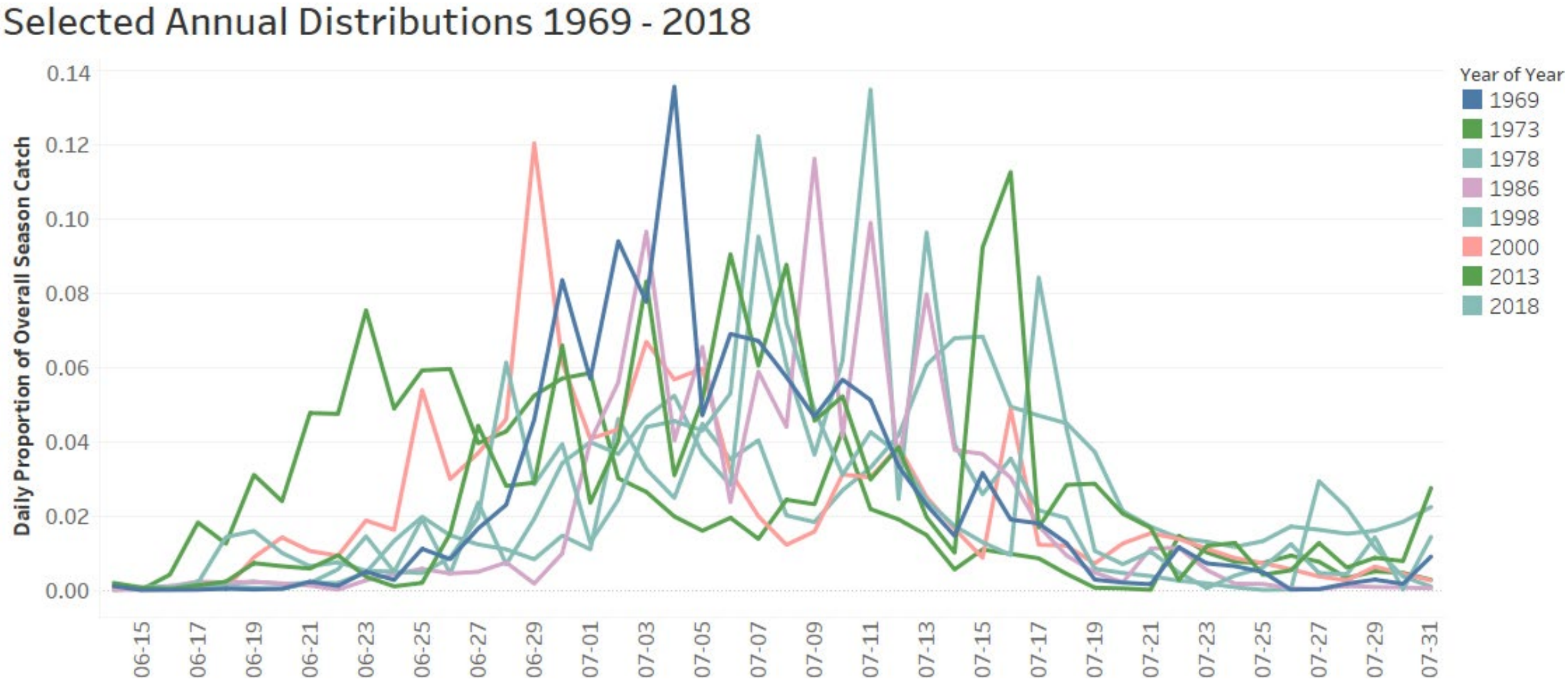


Source: pac.dfo-mpo.gc.ca

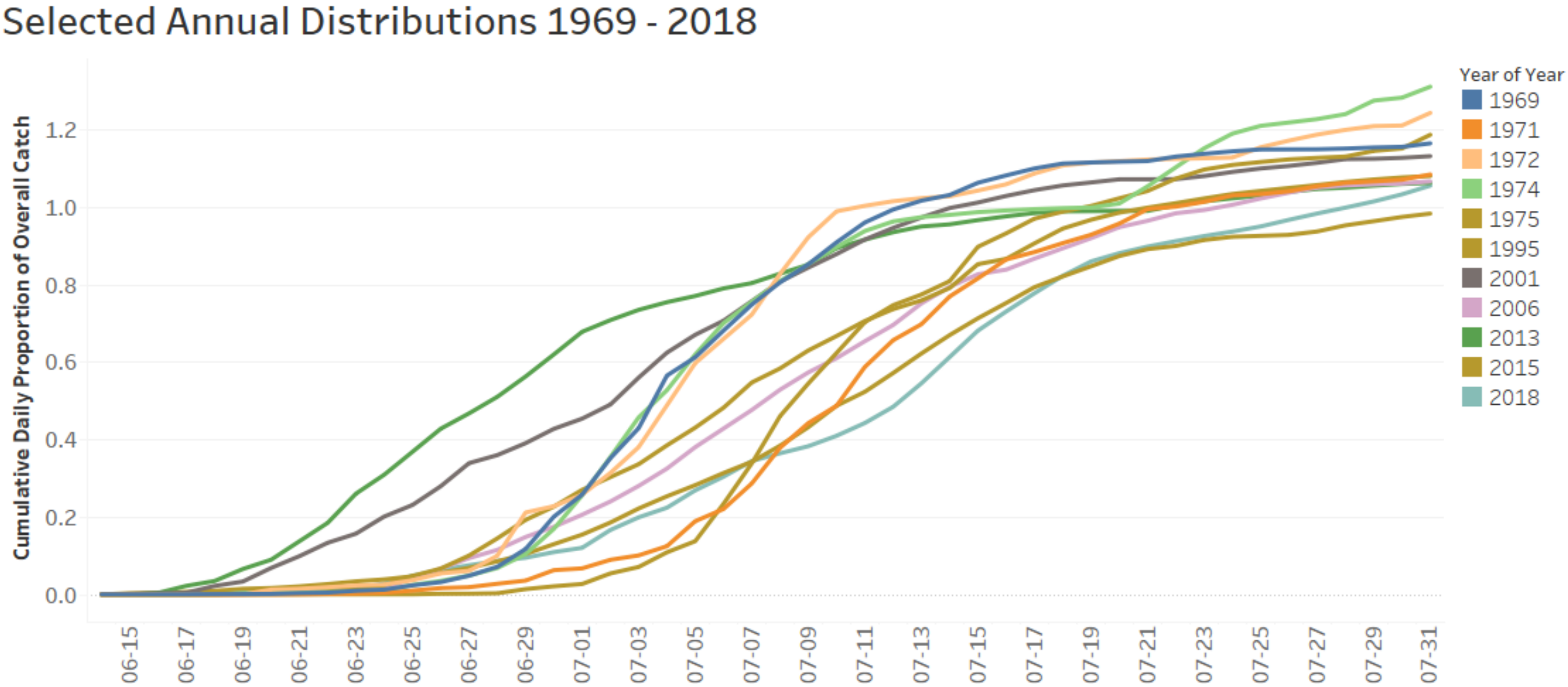
Data Collection & Munging

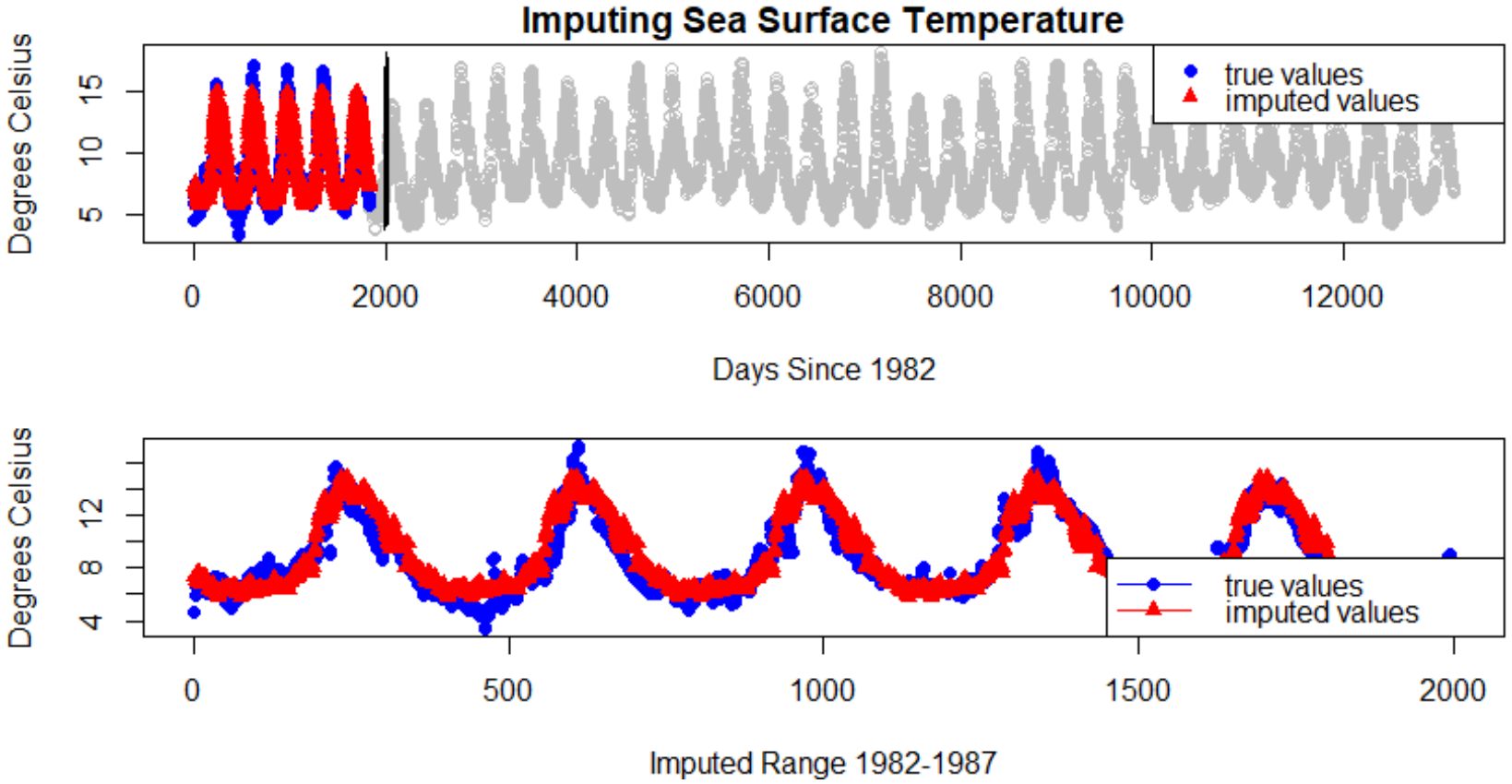
- Collection through communications with ADFG, NOAA, NASA, U of Wash., Fisheries & Oceans Canada, Russian sources
- Daily data for all variables 1968-2018, including oceanographic variables across 528 longitude/ latitude grids
- Multivariate time series imputation of missing data
- Querying multidimensional array data format w/ R and Python
- Tableau and Python for data visualization

Bristol Bay Fishery Historical Runs



Bristol Bay Fishery Historical Runs





Imputation Accuracy Measures:

	ME	RMSE	MAE	MPE	MAPE
Test set	-0.3242365	1.193511	0.9278824	-4.765717	11.02538

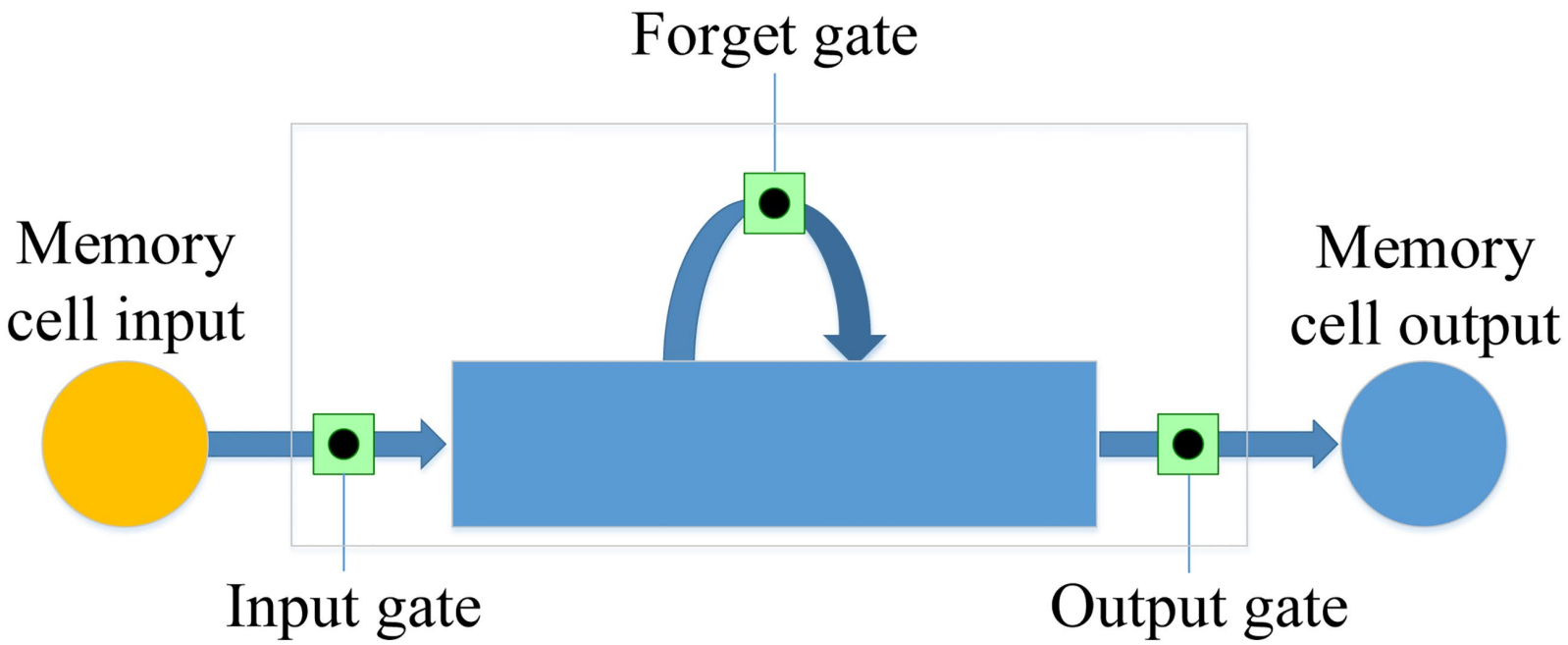
Missing Data

- Univariate time series imputing in R with ImputeTS & Forecast
  - Single time series using SARIMA and Kalman filtering
- Multivariate time series imputing in R using MTSDI & Amelia II
  - Multiple time series imputing using SARIMA and correlations between different variables

Time Series Models

Model	Description
Long Short-Term Memory	Deep learning model specialized towards time series analysis by allowing for correlations between unknown lags
Generalized Additive Model (GAM)	Estimates non-parametric functions of predictor variables, connected to the dependent variable via a link function
Generalized Auto-Regressive Conditional Heteroskedasticity (GARCH)	Conditional variance evolves according to auto-regressive procedure, with many sophisticated variations
Seasonal Auto-regressive Integrated Moving Average (SARIMA)	Combines seasonal, autoregressive & moving average temporal elements while allowing time series to become stationary
Ensemble Models	Avoids overfitting by combining multiple forecasts from less accurate models into an aggregate, boosted accuracy model

Long Short Term Memory (LSTM)



- Recurrent Neural Network (RNN) which unrolls time axes into a time series model, generates weighted pass outputs, then cycles them back into the model with new input information
- Solves short-term memory limitation of traditional RNN's, which cannot sequence long-term dependencies
- Stores useful dependency information between structure cells via Operation Gates, modifying cycle weights along sequence

Future Results Summer & Fall 2019

- Multiple model forecasts released May 2019
- JSM Data Science conference presentation July 2019
- Model accuracy assessment August 2019
- Publish in academic journal Winter 2019
- Recurring refinement of funded annual forecast 2020 & beyond

References

- Visit my GitHub account for project updates on modeling, coding, visualizations, documents, and forecasts: ***github.com/bmelchert2001***
- Visit my LinkedIn account for general personal background: ***linkedin.com/in/melcb***
- Relevant papers:

1. Ruggerone, Gregory. 2019. Sockeye Salmon Migration Timing in Bristol Bay, Alaska, Based on Oceanographic and Biological Variables
2. Elison, T., P. Salomone, T. Sands, G. Buck, K. Sechrist, and, D. Koster. 2018. 2017 Bristol Bay annual management report. Alaska Department of Fish and Game, Fishery Management Report No. 18-11, Anchorage
3. Hodgson, S. , Quinn, T. P., Hilborn, R. , Francis, R. C. and Rogers, D. E. (2006), Marine and freshwater climatic factors affecting interannual variation in the timing of return migration to fresh water of sockeye salmon (Oncorhynchus nerka)
4. A general model for salmon run reconstruction that accounts for interception and differences in availability to harvest. 2018. Curry J. Cunningham, Trevor A. Branch, Tyler H. Dann, Matt Smith, James E. Seeb, Lisa W. Seeb, Ray Hilborn