

Finding suitable weather indices for novel fisheries index insurance using machine learning

Nathaniel Grimes

2024-10-07

Index insurance is a financial tool gaining traction for application in fisheries. It will cover fishers losses under extreme weather events that impact fishery productivity. This is the first assessment to determine the feasibility of such programs and whether suitable indices exist.

Table of contents

1	Introduction	1
2	Insurance Model	3
3	Data	5
3.1	Fishery Data	5
3.2	Environmental Data	5
4	Methods	7
5	Results	7
	Discussion	7

1 Introduction

Predicting fishery output from weather variables is notoriously difficult. It is widely established that climate and weather affect fishing populations (Lehodey *et al.* 2006), but most stock assessment models use little to no year to year environmental data (Privitera-Johnson and

Punt 2020). Variations in environmental conditions are now the leading cause of fishery closures and disaster relief payouts in the United States (Bellquist *et al.* 2021). Disaster declarations are becoming more frequent straining a slow, inequitable system (Holland and Leonard 2020; Jardine *et al.* 2020). Calls for new financial tools to alleviate fisher income shocks have grown (Mumford *et al.* 2009; Sethi 2010).

Index insurance has risen as a prime candidate tool to protect fishing communities during disasters (Watson *et al.* 2023). Index insurance is a financial product that pays out when an independently verified index, such as rainfall or temperature, falls below a predetermined threshold. The index is chosen to be highly correlated with the asset being insured. However, fisheries are notoriously difficult to establish clear, concise weather impacts on productivity. Fisheries are dynamic systems where weather shocks can leave residual impacts years after the initial shock (Hilborn 2003?). Individual fisheries can cover enormous areas in ocean basins. The expansive spatial coverage of fisheries makes it unclear where and how much specific weather variables have on biological abundance. In addition, if weather impacts have been observed, they are most likely highly non-linear adding further complexity. The greatest impediment to the development of fishery insurance policies is reconciling these challenges to find suitable indices that can predict fishery productivity (Watson *et al.* 2023).

Recent expansions in oceanic remote sensing has led to a wealth of new environmental indices that could be used to predict fishery productivity. Fishery data collection continues to improve with better reporting systems with longer and more detailed catch histories. This study aims to leverage these improved data sources to create suitable indices for fisheries index insurance using machine learning.

The difficulty in modelling fishery productivity with environmental indices leads to basis risk. Formally, basis risk is the probability that policyholders experience a harmful shock to their income, but the index does not trigger. Basis risk lowers demand for index insurance and remains a significant roadblock in setting up new programs (Binswanger-Mkhize 2012; Clarke 2016). Designing indices with stronger correlations to fishery losses is the most effective way to reduce basis risk (Jensen *et al.* 2019).

It is impossible to completely eliminate basis risk. In agriculture, every policy possess some level of basis risk. Well designed policies that capture up to 90% of the income variation in K

Three strategies are often used to mitigate basis risk and stimulate uptake. First, government subsidies directly mask the ineffectiveness of some triggers. The United States Risk Management Agency’s Rainfall index insurance for pasture, rangeland, and forage (RI-PRF) allows farmers to select grids of forage for cattle ranching and protect against low rainfall in 2 month intervals (e.g. Jan-Feb). Subsidies encourage farmers to buy products through reducing the premium paid by up to 60%. In Nebraska and Kansas from 2013 to 2017, the program had negative returns overall, but farmers had net positive income strictly due to the subsidies (Goodrich *et al.* 2019). Basis risk in Nebraska and Kansas introduced a 26% probability of insurance not paying out when damages were suffered (Yu *et al.* 2019). The same program

in California found basis risk reaching up to 46%, which could be driven by weak correlations ($r \in [0.071, 0.417]$) between indices and forage production (Keller and Saitone 2022).

Contract design can mitigate basis risk through providing more options so that individuals can better select policies that protect them. Policyholders choose lower trigger levels when correlations between index and asset are low (Lichtenberg and Iglesias 2022). Lower trigger levels correspond to protection against more catastrophic shocks. Increased contract flexibility reduces basis risk by only small amounts. Yu *et al.* (2019) found that more flexible contracts could account for only 5-9% of basis risk. Farms in Kansas closer to weather stations had better predictive impacts of rain on yield (Yu *et al.* 2019).

Machine learning has exploded as a new tool to define new and better indices in agriculture index insurance Schmidt *et al.* (2022). Machine learning models excel in index insurance because indemnity contracts only need predictive relationships. Fishery stock assessments build complex models with biological foundations to accurately inform management of future fish stocks. Index insurance can look retroactively at data to uncover relationships and test out of sample predictive quality. The application of machine learning is growing in fisheries to answer a variety of data needs. Ensemble models built through combinations of random forests, boosted trees, and dynamic linear models improved Bristol Bay sockeye salmon forecasts by 15% compared to a standard lagged regression model (Ovando *et al.* 2022). Environmental variables of importance to groundfish populations in Alaska were uncovered using single index varying coefficient models regularized with LASSO (Correia2018?). Random Forests models better predict fish catch in Indonesia than traditional linear models (Rahman *et al.* 2022). The expected non-linear interactions of weather and fishery productivity merit.

This study will provide the first comprehensive examination of the necessary features of weather indices for fisheries index insurance.

We will use machine learning to uncover the most important weather variables for predicting fishery productivity. We will then use these variables to create a novel index insurance model for fisheries. We will test the model on a wide array of fisheries in California to determine the feasibility of index insurance for fisheries.

2 Insurance Model

Insurance contracts are specified by calculating payout functions ($I(\omega)$) based on independently measured weather variables. Neural networks have been used to provide non-linear payoff schedules that better reduce basis risk (Chen *et al.* 2024). While it has been shown that linear payoff functions inherently lead to basis risk and therefore lower demand (Clarke 2016; Jensen *et al.* 2016), we maintain their use to preserve measures of interpretability that are clearer for a first analysis of fishery index insurance.

Payouts will be issued when prediction models predict negative deviations from long run average value. The three prediction models ($k \in \{\text{LR, LA, RF}\}$) are a linear regression (LR), a LASSO regression (LA), and a random forest (RF).

$$I(\omega) = \max(0, (\bar{\pi} - \hat{\pi}_t^k(\omega)) \cdot l) \quad (1)$$

Where k is the prediction model, l is the level of coverage, $\hat{\pi}_t^k(w)$ is the predicted fishing variable from ω weather variables, and $\bar{\pi}_x$ the long run average of the fishing variable. The premium is calculated as the expected value of the payout function times the premium loading factor (m).

$$\rho(\omega) = \mathbb{E}[I(\omega)]m \quad (2)$$

Utility measures offer the most insightful evaluation of index insurance policies (Kenduiywo *et al.* 2021). It captures value added for policyholders, not just measures of payout frequency as other measures of basis risk. Constant absolute risk aversion allows more consistent comparison for different levels of wealth. Expected utility for a given fishery is the average utility over all years in the sample for any variable of interest π_t . Fishers are allowed to choose insurance coverage levels l to ensure feasible contracts.

$$\begin{aligned} \mathbb{E}[U_b] &= \frac{1}{n} \sum_t \frac{1 - e^{-a\pi_t}}{a} && \text{No Insurance} \\ \mathbb{E}[U_i] &= \max_l \frac{1}{n} \sum_t \frac{1 - e^{(-a(\pi_t + I(\omega, l) - \rho(w)))}}{a} && \text{Insurance} \\ U_r &= \frac{\mathbb{E}[U_i] - \mathbb{E}[U_b]}{\mathbb{E}[U_b]} \cdot 100 && \text{Percent Change in Utility} \end{aligned}$$

We will compare the percent change in fisher utility with insurance (U_i) versus without insurance (U_b) for each prediction method. The variable of interest π_t , will be fishing revenue, landings, and catch per fisher to test what measures of fishery productivity are most suitable for index insurance.

No data on fishery insurance suppliers exists to create a market equilibrium. We iteratively vary the premium loading $m \in [1, 2]$ to create a range of coverage values fishers will be willing to pay for a given m . Then, the amount of coverage purchased times the premium loading factor approximately equals the expected revenue an insurance company would receive. Insurance companies could then examine their own administrative and legal costs to determine whether the feasible contract is profitable.

3 Data

This study attempts to cover breadth, not depth in possible indices. Each fishery has unique ecological characteristics that interact with environmental variables in different and non-linear ways. By studying a wide collection of fisheries and environmental variables we can uncover the potential feasibility of index insurance for fisheries holistically, and then further refine measures with ecologically sound models in the future.

3.1 Fishery Data

Landings revenue, and participation data comes from the West Coast Fish data package (Free *et al.* 2022). It is a reconstruction of California Fish and Wildlife Department catch data combined with PacFin receipts for Washington and Oregon. The last three years of data are updated from the CDFW Marine Fisheries Data Explorer (MFDE). Names are matched to each species within the West Coast Fish data package.

We select California fisheries with a minimum of 30 years of consecutive, non-confidential catch records at both the state and port-complex level. Unclassified catch records are dropped i.e. “Other Sharks” and similar categories. Fisheries with an average revenue greater than \$100,000 at the state and \$75,000 at port-complex level are analyzed. Twenty four fisheries at the state level and 49 fisheries at the port complex level meet these criteria. These fisheries contain the most economically important fisheries in California and their mean values are shown in `?@tbl-fish-sum`.

Fisheries have complex spatial dynamics. Agriculture has clear, quantifiable impacts of weather in grids that are well suited for index insurance. Drought on a single farm directly leads to crop loss for that farm. Whether there is sufficient spatial coverage to identify impacts down to an individual farm remains a challenge in agriculture (Dalhaus and Finger 2016; Leppert *et al.* 2021; Stigler and Lobell 2024). Fish and fishers can move thousand of miles in a given year, thus more consideration must be given to the location of weather impacts in fisheries. We spatially refine catch histories using the California CDFW fishing blocks records from the MFDE Data Explorer. Summarized catch histories of all landed fish within each block provide an average representation of effort for a given fishery. Spatial catch history is measured at both the state and port-complex level. The spatial location refines the location of environmental variables. Local weather is more likely to affect fishery productivity and catch than observations thousands of miles away.

3.2 Environmental Data

Fisheries are highly sensitive to marine heatwaves and water temperature. Sea surface temperature is a natural variable to first consider in fisheries index insurance. Sea surface temperature

data comes from the NOAA DHW data set that provides 5-km resolution of monthly temperature from 1985 to 2023. The 5-km grids are averaged within the nearest California fishing block to provide an annual time series of temperature for each fishery. Temperature is lagged from 1 to 3 years prior to account for residual impacts that carry over due to fishery biological dynamics.

Upwelling provides vital nutrients to stimulate primary productivity. The coast of California is a highly productive ecosystem due to its patterns of upwelling (Chelton *et al.* 1982; Huyer 1983). We capture upwelling through monthly observations of Coastal Upwelling Transport Index (CUTI) and Biological Effective Upwelling Transport Index (BEUTI). Both indices create measures of vertical movement in the mixed layer at 1 degree latitude intervals extending 75 km along the entire US West Coast (Jacox *et al.* 2018). The closest layer to the surface was used in this analysis as the correlation between surface index values and deeper index values are high. CUTI examines the physical measures of wind, ekman transport, and cross-shore geostrophic transport to indicate the strength of upwelling in a given month. BEUTI adds nitrate concentration in its calculation to capture more biological effects of upwelling. Fishing blocks are matched to the nearest 1 degree latitude interval to provide a monthly time series of upwelling for each fishery. Seasonal strengths of upwelling are captured by averaging CUTI and BEUTI within each quarter of the year. Spring upwelling in early March and April are especially important to a wide array of fish species. Yearly average and amplitude values (the difference between minimum observed upwelling and maximum) are also calculated. These indices are the most temporally limited datasets in this analysis, only extending from 1988 to 2023.

The Habitat Compression Index measures the area extent of water below average temperatures thresholds along the US West Coast (Schroeder *et al.* 2022). Habitat compression is a measure of the spatial extent of cold water habitats that are important for fish species. The index is broken down into four distinct oceanographic regions ranging from 3.5 degrees to 5.5 degrees latitude in size with coverage out to 150 km offshore. We use the cumulative habitat compression index that sums the index value in each month to provide a yearly time series of habitat compression for each fishery. The cumulative index showed stronger correlations with biological productivity measures than monthly measures (Schroeder *et al.* 2022)

The final environmental variables are the Pacific Decadal Oscillation (PDO) and the El Nino Southern Oscillation (ENSO). Both indices are well known to affect marine ecosystems and fisheries. Both indices are averaged over a given year. PDO data is taken from the PDO ERSST V5, and ENSO data is taken from the multivariate ENSO Index Version 2 (MEI.v2).

Summary statistics for the environmental data are presented in **?@tbl-env-sum**. In total, 73 fisheries with 35 years of catch data are matched to 20 spatial matched environmental variables with annual coverage from 1988 to 2023.

4 Methods

5 Results

Discussion

- Bellquist, L., Saccomanno, V., Semmens, B.X., Gleason, M. and Wilson, J. (2021) [The rise in climate change-induced federal fishery disasters in the united states](#). *PeerJ* **9**.
- Binswanger-Mkhize, H.P. (2012) [Is there too much hype about index-based agricultural insurance?](#) *Journal of Development Studies* **48**, 187–200.
- Cesarini, L., Figueiredo, R., Monteleone, B. and Martina, M.L.V. (2021) [The potential of machine learning for weather index insurance](#). *Natural Hazards and Earth System Sciences* **21**, 2379–2405.
- Chelton, D.B., Bernal, P.A. and McGowan, J.A. (1982) Large-scale interannual physical and biological interaction in the california current. *Journal of Marine Research* **40**, 1095–1125.
- Chen, Z., Lu, Y., Zhang, J. and Zhu, W. (2024) [Managing weather risk with a neural network-based index insurance](#). *Management Science* **70**, 4306–4327.
- Clarke, D.J. (2016) [A theory of rational demand for index insurance](#). *Journal: Microeconomics* **8**, 283–306.
- Dalhaus, T. and Finger, R. (2016) [Can gridded precipitation data and phenological observations reduce basis risk of weather index-based insurance?](#) *Weather, Climate, and Society* **8**, 409–419.
- Feng, P., Wang, B., Liu, D.L., Waters, C. and Yu, Q. (2019) [Incorporating machine learning with biophysical model can improve the evaluation of climate extremes impacts on wheat yield in south-eastern australia](#). *Agricultural and Forest Meteorology* **275**, 100–113.
- Free, C.M., Poulsen, C.V., Bellquist, L.F., Wassermann, S.N. and Oken, K.L. (2022) [The CALFISH database: A century of california’s non-confidential fisheries landings and participation data](#). *Ecological Informatics* **69**, 101599.
- Goodrich, B., Yu, J. and Vandever, M. (2019) [Participation patterns of the rainfall index insurance for pasture, rangeland and forage programme](#). *The Geneva Papers on Risk and Insurance - Issues and Practice* **45**, 29–51.
- Holland, D.S. and Leonard, J. (2020) [Is a delay a disaster? Economic impacts of the delay of the california dungeness crab fishery due to a harmful algal bloom](#). *Harmful Algae* **98**.
- Huyer, A. (1983) [Coastal upwelling in the california current system](#). *Progress in Oceanography* **12**, 259–284.
- Jacox, M.G., Edwards, C.A., Hazen, E.L. and Bograd, S.J. (2018) [Coastal upwelling revisited: Ekman, bakun, and improved upwelling indices for the u.s. West coast](#). *Journal of Geophysical Research: Oceans* **123**, 7332–7350.
- Jardine, S.L., Fisher, M.C., Moore, S.K. and Samhour, J.F. (2020) [Inequality in the economic impacts from climate shocks in fisheries: The case of harmful algal blooms](#). *Ecological*

Economics **176**.

- Jensen, N.D., Barrett, C.B. and Mude, A.G. (2016) [Index insurance quality and basis risk: Evidence from northern kenya](#). *American Journal of Agricultural Economics* **98**, 1450–1469.
- Jensen, N., Stoeffler, Q., Fava, F., et al. (2019) [Does the design matter? Comparing satellite-based indices for insuring pastoralists against drought](#). *Ecological Economics* **162**, 59–73.
- Keller, J.B. and Saitone, T.L. (2022) [Basis risk in the pasture, rangeland, and forage insurance program: Evidence from california](#). *Amer. J. Agr. Econ* **104**, 1203–1223.
- Kenduiwo, B.K., Carter, M.R., Ghosh, A. and Hijmans, R.J. (2021) [Evaluating the quality of remote sensing products for agricultural index insurance](#). *PLoS ONE* **16**.
- Lehodey, P., Alheit, J., Barange, M., et al. (2006) Climate variability, fish, and fisheries.
- Leppert, D., Dalhaus, T. and Lagerkvist, C.J. (2021) [Accounting for geographic basis risk in heat index insurance: How spatial interpolation can reduce the cost of risk](#). *Weather, Climate, and Society* **13**, 273–286.
- Lichtenberg, E. and Iglesias, E. (2022) [Index insurance and basis risk: A reconsideration](#). *Journal of Development Economics* **158**.
- Mumford, J.D., Leach, A.W., Levontin, P. and Kell, L.T. (2009) [Insurance mechanisms to mediate economic risks in marine fisheries](#). *ICES Journal of Marine Science* **66**, 950–959.
- Ovando, D., Cunningham, C., Kuriyama, P., Boatright, C. and Hilborn, R. (2022) [Improving forecasts of sockeye salmon \(*oncorhynchus nerka*\) with parametric and nonparametric models](#). *Canadian Journal of Fisheries and Aquatic Sciences* **79**, 1198–1210.
- Privitera-Johnson, K.M. and Punt, A.E. (2020) [A review of approaches to quantifying uncertainty in fisheries stock assessments](#). *Fisheries Research* **226**, 105503.
- Rahman, L.F., Marufuzzaman, M., Alam, L., Bari, M.A., Sumaila, U.R. and Sidek, L.M. (2022) [Application of machine learning to investigate the impact of climatic variables on marine fish landings](#). *National Academy Science Letters* **45**, 245–248.
- Schmidt, L., Odening, M., Schlanstein, J. and Ritter, M. (2022) [Exploring the weather-yield nexus with artificial neural networks](#). *Agricultural Systems* **196**.
- Schroeder, I.D., Santora, J.A., Mantua, N., et al. (2022) [Habitat compression indices for monitoring ocean conditions and ecosystem impacts within coastal upwelling systems](#). *Ecological Indicators* **144**, 109520.
- Sethi, S.A. (2010) [Risk management for fisheries](#). *Fish and Fisheries* **11**, 341–365.
- Stigler, M. and Lobell, D. (2024) [Optimal index insurance and basis risk decomposition: An application to kenya](#). *American Journal of Agricultural Economics* **106**, 306–329.
- Watson, J.R., Spillman, C.M., Little, L.R., Hobday, A.J. and Levin, P.S. (2023) [Enhancing the resilience of blue foods to climate shocks using insurance](#). *ICES Journal of Marine Science* **80**, 2457–2469.
- Yu, J., Vandeveer, M., Volesky, J.D. and Harmonay, K. (2019) [Estimating the basis risk of rainfall index insurance for pasture, rangeland, and forage](#). *Journal of Agricultural and Resource Economics* **44**, 179–193.