

Tracking and forecasting community responses to climate perturbations in the California Current Ecosystem

Mary E. Hunsicker, Eric J. Ward, Michael A. Litzow, Sean C. Anderson, Chris J. Harvey, John C. Field, Jin Gao, Michael G. Jacox, Sharon Melin, Andrew Thompson

S1 Appendix: Standardization of time series from spatially resolved datasets.

Datasets collected through the Rockfish Recruitment and Ecosystem Assessment Survey (RREAS; pelagic juvenile fish and invertebrates survey) and the California Cooperative Oceanic Fisheries Investigations (CalCOFI, ichthyoplankton survey) include spatial attributes and were standardized using Generalized Additive Models (GAM) to create a univariate time series for each species included in our analysis.

RREAS Survey

Because many species in the RREAS survey were absent from a large number of observed trawls, we modeled species occurrence and abundance separately, using a delta GAM approach with two sub-models (Hastie 1990, Guisan 2002). In the first sub-model, species occurrence (presence-absence) was modelled using a binomial GAM with a logit link. In the second model (‘positive model’), species abundance (count) conditional on the catch of at least one individual was modelled using a Poisson GAM with a log link. The formulations of the two sub-models were analogous and as follows:

$$P_{yr(\varphi,\lambda)} \text{ or } C_{yr(\varphi,\lambda)} = yr + s(\varphi, \lambda) + jday_{yr(\varphi,\lambda)} + jday_{yr(\varphi,\lambda)}^2 + \varepsilon_{yr(\varphi,\lambda)}$$

where P , the probability of species occurrence, or C , an estimate of species abundance when the species is present, is a function of year (yr , as factor), latitude (φ) and longitude (λ), and Julian day ($jday$). A two-dimensional smoothing function is denoted by s and ε indicates an error term.

Using the fitted GAMs for each species, we generated predictions of overall abundance of individual species. First, we created spatial occurrence and abundance distribution profiles by creating a grid of all combinations of model explanatory variables: geographic coordinates (latitude and longitude), Julian day, and year. We restricted the range of geographic coordinates to 20 values within the upper 0.8 and lower 0.2 quantiles of the data set, and bounded the range of Julian days to the upper 0.9 and lower 0.1 quantiles. This was done to avoid edge effects that can result in unrealistic predictions. All unique sampling years were included in the grid. Next, the probability of occurrence and estimates of abundance were predicted for each combination of the explanatory variables, and then the predictions from the two sub-models were multiplied to determine the overall abundance of individual species. Lastly, we calculated the mean standardized abundance of each species in each year from the prediction grid to generate the univariate time series of species abundance used in our study analyses.

CalCOFI survey

The CalCOFI time series of ichthyoplankton densities were standardized using a Tweedie GAM (power parameter fixed at 1.25) (Tweedie 1984, Dunn and Smyth 2002). The model formulation was as follows:

$$D_{yr(\varphi,\lambda)} = yr + season + s(\varphi, \lambda) \cdot yr + \varepsilon_{yr(\varphi,\lambda)}$$

where species density (D) is a function of year (yr), season (spring, summer), and latitude (φ) and longitude (λ). The two-dimensional smoothing function and error term are indicated by s and ε , respectively.

Similar to above, we used the fitted GAMs for each species to generate predictions of species densities. We first created a spatial abundance distribution profile by creating a grid of different combinations of the model covariates. Here we restricted the range of geographic coordinates (latitude and longitude) to those from sampling stations that were sampled ≥ 20 years. We included all sampling years in the grid and limited the season to spring only. Species densities were then predicted for each combination of the model covariates, and the univariate time series of species abundance was generated by calculating the mean standardized density of each species in each year from the prediction grid.

The delta GAMs and Tweedie GAMs were run using the ‘mgcv’ package (v1.8-34; Wood 2011, 2017) in R.

References

Dunn PK, Smyth GK. Series evaluation of Tweedie exponential dispersion model densities.

Statistics and Computing 2005; 15: 267-280.

Guisan A, Edwards TC, Hastie T. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecol. Modell. 2002; 157: 89–100.

Hastie T, Tibshirani R. Generalized additive models. London: Chapman and Hall; 1990.

Tweedie MCK. An index which distinguishes between some important exponential families.

Statistics: Applications and New Directions. Calcutta: Indian Statistical Institute.

Proceedings of the Indian Statistical Institute Golden Jubilee International Conference
(Eds. J. K. Ghosh and J. Roy) 1984; pp. 579-604.

Wood SN. Generalized Additive Models: An Introduction with R (2nd edition). Chapman and
Hall/CRC; 2017.

Wood SN. Fast stable restricted maximum likelihood and marginal likelihood estimation of
semiparametric generalized linear models. Journal of the Royal Statistical Society (B)
2011; 73:3 -36.