ICES Journal of Marine Science, 2022, **79**, 1259–1273
DOI: 10.1093/icesjms/fsac050
Advance access publication date: 13 April 2022
Original Article

Ocean model-based covariates improve a marine fish stock assessment when observations are limited

Hubert du Pontavice [1,2,*], Timothy J. Miller[3], Brian C. Stock [3,4], Zhuomin Chen[5] and Vincent S. Saba[2]

[1]Atmospheric and Oceanic Sciences Program, Princeton University, 300 Forrestal Road, Sayre Hall, Princeton, NJ 08540, USA
[2]National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Northeast Fisheries Science Center, Geophysical Fluid Dynamics Laboratory, Princeton University, 201 Forrestal Road, Princeton, NJ 08540, USA
[3]National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Northeast Fisheries Science Center, 166 Water Street, Woods Hole, MA 02543, USA
[4]Institute of Marine Research, P.O. Box 1870, Nordnes, 5817 Bergen, Norway
[5]Department of Marine Sciences, University of Connecticut, Groton , CT 06340, USA

*Corresponding author: tel: +1-609-258-2617; e-mail: hubert.dupontavice@princeton.edu.

The productivity of many fish populations is influenced by the environment, but developing environment-linked stock assessments remain challenging and current management of most commercial species assumes that stock productivity is time-invariant. In the Northeast United States, previous studies suggest that the recruitment of Southern New England-Mid Atlantic yellowtail flounder is closely related to the strength of the Cold Pool, a seasonally formed cold water mass on the continental shelf. Here, we developed three new indices that enhance the characterization of Cold Pool interannual variations using bottom temperature from a regional hindcast ocean model and a global ocean data assimilated hindcast. We associated these new indices to yellowtail flounder recruitment in a state–space, age-structured stock assessment framework using the Woods Hole Assessment Model. We demonstrate that incorporating Cold Pool effects on yellowtail flounder recruitment reduces the retrospective patterns and may improve the predictive skill of recruitment and, to a lesser extent, spawning stock biomass. We also show that the performance of the assessment models that incorporated ocean model-based indices is improved compared to the model using only the observation-based index. Instead of relying on limited subsurface observations, using validated ocean model products as environmental covariates in stock assessments may both improve predictions and facilitate operationalization.

Keywords: Cold Pool index, ocean model, recruitment, retrospective prediction, Southern New England-Mid Atlantic yellowtail flounder, state–space model, stock assessment.

Introduction

In recent decades, the US northeast shelf marine ecosystem has experienced rapid ocean warming, affecting marine organisms from the surface to the bottom (Forsyth *et al.*, 2015; Pershing *et al.*, 2015; Kavanaugh *et al.*, 2017; Kleisner *et al.*, 2017). While the productivity of many fish populations is largely influenced by the environment (Vert-pre *et al.*, 2013; Szuwalski *et al.*, 2015), current management of most commercial species assumes that stock productivity is time-invariant.

In the Northeast US shelf, previous studies have identified significant relationships between stock productivity and climate variables for several groundfish including yellowtail flounder (*Limanda ferruginea*; Sissenwine, 1974; Miller *et al.*, 2016; Xu *et al.*, 2018), summer flounder (*Paralichthys dentatus*; O'Leary *et al.*, 2019), winter flounder (*Pseudopleuronectes americanus*; Bell *et al.*, 2014, 2018), and Atlantic cod (*Gadus morhua*; Miller *et al.*, 2018). When environmentally induced changes in stock productivity are not taken into account in assessments, retrospective bias can occur (Brooks and Legault, 2016; Tableau *et al.*, 2019) and affect the short-term catch advice. Therefore, it is critical to develop fishery stock assessments that integrate environmental effects as a research priority (Hare *et al.*, 2016).

The physical environment in the Southern New England (SNE) and middle Atlantic Bight (MAB) is highly dynamic due to seasonal, interannual, and decadal variability in both atmospheric and oceanographic processes. The MAB Cold Pool is a seasonally formed cold water mass that occurs from late spring to early fall and is associated with the recruitment and settlement of the Southern New England-Mid Atlantic (SNEMA) yellowtail flounder population (Sullivan *et al.*, 2000, 2005; Miller *et al.*, 2016). This stock experienced overfishing from the 1970s to the mid-1990s and, in parallel, recruitment has declined since the 1980s (NEFSC, 2012). From 1990 onwards, recruitment has remained dramatically low and the stock status is listed as overfished, with spawning stock biomass (SSB) below the management target (NEFSC, 2020). This stock, which is currently in a rebuilding plan, is assessed using a statistical catch-at-age model (ASAP, Age-Structured Assessment Program, Legault and Restrepo, 1999). One of the major sources of uncertainty in the assessment is the cause of the persistent low recruitment in recent years (NEFSC, 2012, 2020).

Recruitment of SNEMA yellowtail flounder depends on temperature conditions during the early life stages (Sullivan *et al.*, 2000, 2005). After spawning in spring and early summer, eggs are fertilized and float near the surface for about

Received: September 20, 2021. **Revised:** March 7, 2022. **Accepted:** March 8, 2022

2 months. Then, the late-stage larvae/early juveniles settle to the bottom during the summer and the early fall. Field studies suggest that the recruitment of SNEMA yellowtail flounder is closely related to ocean bottom temperature, vertical thermal structure, and maintenance of the Cold Pool (Sullivan *et al.*, 2000, 2005). Recently, modelling studies have shown that recruitment and SSB are closely related to the Cold Pool dynamics and Gulf Stream position (Miller *et al.*, 2016; Xu *et al.*, 2018).

Cold Pool effects on SNEMA yellowtail flounder recruitment have been explored by estimating its intensity using a Cold Pool Index based on *in situ* observations (hereafter, Obs_CPI; Miller *et al.*, 2016; Xu *et al.*, 2018). These bottom temperature observations are from the Northeast Fisheries Science Center (NEFSC) fall bottom trawl survey, which occurs annually during September and October (Mountain, 2003). The Obs_CPI provides a restricted perception of the potential Cold Pool impacts on recruitment because it does not cover the entire larval settlement period from summer to early fall, when temperature may play a critical role in larval survival. Another limitation is that the Obs_CPI relies on observations from the NEFSC fall survey, which can be limited or even absent during critical changes to the Cold Pool. For example, the NEFSC fall survey was severely restricted in 2017 (vessel mechanical failure) and did not take place in 2020 (COVID-19 pandemic). An alternative approach is to quantify the strength of the Cold Pool using regional ocean models. Chen and Curchitser (2020) developed a method to compute a Cold Pool Index that accounts for the Cold Pool persistence time, temperature, and volume based on a long-term high-resolution regional ocean modelling system (ROMS). While this is a valuable tool for studying the interannual variability of the Cold Pool, it is challenging to use it operationally in the SNEMA yellowtail flounder stock assessment. First, the ROMS model used in Chen and Curchitser (2020) extends only until 2007 and would, thus need to be updated every 2 years for the stock assessment. Second, the modelled bottom temperature is warm-biased during the stratified season (Chen *et al.*, 2018; Chen and Curchitser, 2020). Finally, quantifying Cold Pool persistence requires high-resolution, 3D profiles of temperature and salinity at a daily resolution.

Here, we developed three alternative Cold Pool indices that account for intensity, persistence, and the spatial extent of the Cold Pool using bottom temperature estimates from a regional ocean model and a global ocean data assimilated hindcast. Our goal was to develop tailored indices that better characterize Cold Pool interannual variations that are associated with the historical and current variability of SNEMA yellowtail flounder recruitment. We associated the new Cold Pool indices to recruitment in a state–space, age-structured stock assessment framework (Miller *et al.*, 2016; Miller and Stock, 2020; Stock and Miller, 2021). Finally, we assessed whether including ocean model based Cold Pool indices in the SNEMA yellowtail flounder stock assessment reduces retrospective patterns and improves the skill of short-term forecasts.

Material and methods

Environmental data

We used high-resolution bottom temperature ocean reanalysis and model output over the period 1972–2019 to cover the same time period as the last stock assessment (NEFSC, 2020). Instead of using observed ocean temperature data as in Miller *et al.* (2016), we combined two model-based estimates of monthly bottom temperature between 1972 and 2019.

For the period between 1972 and 1992, we used ocean bottom temperature from the long-term (1958–2007) high-resolution numerical simulation of the Northwest Atlantic Ocean in the Regional Ocean Modelling System (hereafter, called ROMS-NWA). Previous studies that focused on the ROMS-NWA-based Cold Pool highlighted strong and consistent warm bias in bottom temperature of about 1.5°C during the stratified seasons over the period of 1958–2007 (Chen *et al.*, 2018; Chen and Curchitser, 2020). In order to bias-correct bottom temperature from ROMS-NWA, we used the monthly climatologies of observed bottom temperature from the Northwest Atlantic Ocean regional climatology (NWARC) over decadal periods from 1965 to 1994. The NWARC provides high resolution (1/10° grids) of quality-controlled *in situ* ocean temperature based on a large volume of observed temperature data (Seidov *et al.*, 2016a, b; https://www.ncei.noaa.gov/products/northwest-atlantic-regional-climatology). The first step was to re-grid the ROMS-NWA to obtain bottom temperature over the same 1/10° grid as the NWARC. A monthly bias was calculated in each grid cell and for each decade (1965–1974, 1975–1984, and 1985–1994) in the MAB and in the SNE shelf:

$$BIAS_{i,\,d} = T_{i,d}^{Climatology} - \bar{T}_{i,\,d}^{ROMS-NWA}, \quad (1)$$

where $T_{i,d}^{Climatology}$ is the NWARC bottom temperature in the grid cell i for the decade d and $\bar{T}_{i,\,d}^{ROMS-NWA}$ is the average ROMS-NWA bottom temperature over the decade d in the grid cell i. Time series of the decadal bias estimates in the Cold Pool domain between ROMS-NWA and the NWA-climatology during the summer period are represented in Supplementary Material I (Figure S1.1), while monthly and interannual time series of bottom temperature between ROMS-NWA before the bias correction (ROMS-NWA) and ROMS-NWA after the bias correction (debiased ROMS-NWA) can be found in Supplementary Material I (Figure S1.2). Furthermore, maps of the mean decadal bias estimates between ROMS-NWA and the NWA-climatology during the summer period (June–September) in the Cold Pool domain are available in Supplementary Material I (Figure S1.3), and maps of the mean bottom temperature are represented for each decade in Supplementary Material I (Figure S1.4). We discussed the modelled data quality and limitations as well as the bias correction we applied to ROMS-NWA in Supplementary Material II.

For the period between 1993 and 2019, we used the bottom temperature output from the gLobal Ocean ReanalYsis and Simulation project (GLORYS12v1) ocean reanalysis, which is a global ocean, eddy-resolving, and data assimilated hindcast from Mercator Ocean (Fernandez and Lellouche, 2018; Lellouche *et al.*, 2021). This dataset has been shown to be highly representative of observational surface and bottom temperature and salinity over the Northeast US shelf (Chen *et al.*, 2021; Supplementary Material I, Figure S1.5). The comparison of the monthly and interannual bottom temperature variations highlighted that debiased ROMS-NWA and GLORYS12v1 over the Cold Pool domain (Supplementary Material I, Figure S1.2) are fully integrable.

Cold Pool indices

We explored the Cold Pool effects on the SNEMA yellowtail flounder recruitment by considering three characteristics of the Cold Pool: strength, persistence over time, and spatial extent. The first step was to define the Cold Pool domain, which is typically located within the MAB and the southern flank of Georges Bank (Houghton *et al.*, 1982; Lentz, 2017; Chen *et al.*, 2018). Here, we delineated a spatial domain covering the management area of the SNEMA yellowtail flounder comprising the MAB and in the SNE shelf between the 20 and 200 m isobaths (Chen *et al.*, 2018; Chen and Curchitser, 2020). We restricted the time period from June (to match the start of the settlement period; Sullivan *et al.*, 2005) to September (which is the average end date of the Cold Pool (calendar day 269) estimated by Chen and Curchitser (2020). The Cold Pool domain was defined as the area, wherein average bottom temperature was cooler than 10°C between June and September from 1972 to 2019. We then developed the three Cold Pool indices using bottom temperature from ocean models.

The Cold Pool Index (Model_CPI) was adapted from Miller *et al.* (2016). Residual temperature was calculated in each grid cell, i, in the Cold Pool domain as the difference between the average bottom temperature at the year y ($T_{i,y}$) and the average bottom temperature over the period 1972–2019 ($\bar{T}_{i,\,1972-2019}$) between June and September. Model_CPI was calculated as the mean residual temperature over the Cold Pool domain such that

$$Model\_CPI_y = \frac{\sum_{i=1}^{n}(T_{i,\,y} - \bar{T}_{i,\,1972-2019})}{n}, \quad (2)$$

where n is the number of grid cells over the Cold Pool domain.

The temporal component of the Cold Pool was calculated using the persistence index (Model_PI). Model_PI measures the duration of the Cold Pool and was estimated using the month when bottom temperature rises above 10°C after the Cold Pool is formed each year. We first selected the area over the Cold Pool domain in which bottom temperature falls below 10°C between June and October. We then calculated the "residual month" in each grid cell, i, in the Cold Pool domain as the difference between the month (as in month-index "1 = January," "2-February," …) when bottom temperature rises above 10°C in year y and the average of those months over the period 1972–2019. Then, Model_PI was calculated as the mean "residual month" over the Cold Pool domain:

$$Model\_PI_y = \frac{\sum_{i=1}^{n}(Month_{i,\,y} - \overline{Month}_{i,\,1972-2019})}{n}. \quad (3)$$

Finally, we developed a spatial extent index (Model_SEI) to test the hypothesis that SNEMA yellowtail flounder recruitment is dependent on the amount of Cold Pool habitat available to the larval settlers (Sullivan *et al.*, 2005). Model_SEI is estimated by the number of cells where bottom temperature remains below 10°C for at least 2 months between June and September. We also tested Obs_CPI, the Cold Pool index developed by Miller *et al.* (2016), to compare two types of indices; one calculated from the observed bottom temperature (Obs_CPI) and one calculated from the modelled bottom temperature (Model_CPI).

WHAM

We implemented the SNEMA yellowtail flounder assessment in the Woods Hole Assessment Model (WHAM), a state–space, age-structured stock assessment framework that allows explicit linking of population processes to environmental covariates (Miller and Stock, 2020; Stock and Miller, 2021). The WHAM model treated all numbers at age as independent random effects (Stock and Miller, 2021). As in Miller *et al.* (2016) and Stock and Miller (2021), age-composition data were assumed to follow a logistic-normal distribution with pooling of zero observations with adjacent ages. As in the latest SNEMA yellowtail flounder assessment, weight-at-age, natural mortality-at-age, and maturity-at-age were treated as known (see details in Supplementary Material III, S3.1; NEFSC, 2020). Selectivity of the fleet was divided into six time blocks and specific selectivity is assigned for each of the three indices of abundance coming from the spring, fall, and winter NEFSC bottom trawl surveys. Selectivity for the fleet and indices were assumed to have logistic functional form except for three blocks for which age-specific, flat-topped selectivity was used to facilitate convergence.

To conduct 4-year predictions, we fixed weight-at-age and maturity-at-age at their average values from the last 5 years of data (NEFSC, 2020). We forecasted the numbers-at-age and the environmental covariates in the prediction years by continuing the autoregressive processes.

We fitted the models using the open-source statistical software R (R Core Team, 2021) and TMB (Kristensen *et al.*, 2016), as implemented in the WHAM package (v1.0.4). The fixed effect parameters, which are estimated by maximizing the marginal likelihood within R can be found in Supplementary Material IV for the five selected models (see below, the details regarding these models). The random effects ($n = 458$) are posterior empirical Bayes estimates (Kristensen *et al.*, 2016).

Incorporating Cold Pool effects in WHAM

In WHAM, the Cold Pool Index (X_t) in year t was modelled as a first-order autoregressive, AR(1), process:

$$X_t \sim N\left(\mu_X\,(1-\phi_X) + \phi_X X_{t-1},\ \sigma_X^2\right), \quad (4)$$

with $X_1 \sim N(\mu_X,\ \frac{\sigma_X^2}{1-\phi_X^2})$ and where μ_X was the marginal mean of the process, σ_X^2 the variance of the process, and ϕ_X the autocorrelation parameter. The Cold Pool observations, y_t, were assumed to be normally distributed with mean X_t and variance $\sigma_{X_t}^2$ such that

$$y_t|X_t \ \sim\ N\left(X_t,\ \sigma_{X_t}^2\right). \quad (5)$$

The observation variances of each year t, $\sigma_{X_t}^2$, were treated as known with year-specific values for Model_CPI, Obs_CPI, and Model_PI. The observation error variance for Model_SEI was estimated from the Model_CPI and Model_PI observation variances, which were based on the same data (GLORYS12v1 and ROMS-NWA) because no direct method was identified (Supplementary Material III, S3.2).

The Cold Pool indices estimated by the models and the standard residuals are presented in Supplementary Material III, S3.3).

As in Xu *et al.* (2018), we tested different ways to incorporate the Cold Pool effect on recruitment, following Iles and Beverton (1998). The Cold Pool events were hypothesized to affect the carrying capacity by determining the amount of suitable habitat for pre-recruits, and thus its effect on recruitment was modelled as a "limiting factor" in the Beverton–Holt stock–recruit function (Miller *et al.*, 2016). We also

Table 1. Negative log-likelihood (NLL), AIC, and difference in AIC (ΔAIC) for each of the tested models.

	Obs_CPI	Model_CPI	Model_PI	Model_SEI	NLL	AIC	ΔAIC
m1	—	Controlling	—	—	−612.84	−1041.7	0
m2	—	Controlling	Limiting	—	−613.57	−1041.1	0.6
m3	—	Controlling	—	Limiting	−613.31	−1040.6	1.1
m4	—	Masking	—	—	−612.04	−1040.1	1.6
m5	—	Controlling	Masking	—	−612.94	−1039.9	1.8
m6	—	Controlling	—	Masking	−612.85	−1039.7	2
m7	—	Limiting	Masking	—	−612.51	−1039	2.7
m8	Limiting	—	—	—	−611.40	−1038.8	2.9
m9	—	Masking	—	Controlling	−612.24	−1038.5	3.2
m10	—	Masking	Limiting	—	−612.19	−1038.4	3.3
m11	—	Masking	Controlling	—	−612.09	−1038.2	3.5
m12	—	Masking	—	Limiting	−612.11	−1038.2	3.5
m13	Controlling	—	—	—	−610.96	−1037.9	3.8
m14	Masking	—	—	—	−610.74	−1037.5	4.2
m15	—	Limiting	—	—	−610.56	−1037.1	4.6
m16	—	Limiting	—	Masking	−611.53	−1037.1	4.6
m17	—	—	Controlling	—	−610.44	−1036.9	4.8
m18	—	—	Masking	—	−610.39	−1036.8	4.9
m19	—	Limiting	Controlling	—	−611.30	−1036.6	5.1
m20	—	—	Limiting	—	−609.91	−1035.8	5.9
m21	—	Limiting	—	Controlling	−610.80	−1035.6	6.1
m22	—	—	—	Controlling	−608.89	−1033.8	7.9
m23	—	—	—	Masking	−608.30	−1032.6	9.1
m24	—	—	—	Limiting	−607.96	−1031.9	9.8
m25	—	—	—	—	−600.56	−1019.1	22.6

tested the Cold Pool effect on the mortality rates for larvae and young fish *via* density-independent mortality ("controlling factor") and on pre-recruit mortality and/or growth *via* density-dependent process ("masking factor").

We tested the incorporation of multiple Cold Pool effects in WHAM. As in Miller *et al.* (2016) and Xu *et al.* (2018), we fitted a set of models including one environmental covariate. Furthermore, the WHAM model was extended to allow inclusion of two Cold Pool effects into the stock–recruit function g:

$$g\left(X_{1,t}, X_{2,t}, S_t\right) = log\left(\frac{\alpha_0 S_t e^{\alpha_1 X_{1,t-1}+\alpha_2 X_{2,t-1}}}{e^{\gamma_1 X_{1,t-1}+\gamma_2 X_{2,t-1}} + \beta_0 S_t e^{\beta_1 X_{1,t-1}+\beta_2 X_{2,t-1}}}\right), \quad (6)$$

where S_t is the SSB at time t and the two different CPIs, $X_{j,t}$, can each have a controlling (α_j), limiting (β_j), masking (γ_j), or no ($\alpha_j = \beta_j = \gamma_j = 0$) effect. The incorporation of two Cold Pool effects simultaneously allowed us to explore combined effects of the Cold Pool strength, persistence, and area. For example, a weak Cold Pool event (high CPI) that may induce an increase of the density-independence mortality, i.e. act as a "controlling factor," can have a low spatial extent (high SEI) that may affect the carrying capacity for pre-recruits as a "limiting factor" (Maunder and Thorson, 2019). The environmentally explicit stock–recruit function, g, then determines the expected number of recruits at time $t+1$, R_{t+1}, as in WHAM:

$$log\left(R_{t+1}\right) = log[(g\left(X_1, X_2, S_t\right)] + \varepsilon_{t+1}, \quad (7)$$

with $\varepsilon_{t+1} = N(-\frac{\sigma_R^2}{2}, \sigma_R^2)$, where ε_{t+1} is the stochastic recruitment deviation at time $t+1$ with variance σ_R^2 and $-\frac{\sigma_R^2}{2}$ is the bias correction term (Stock and Miller, 2021).

In order to visualize and analyze the covariate effects on the stock–recruitment relationships, we plotted the stock–recruitment relationship over the range of the different estimated Cold Pool indices (Obs_CPI, Model_CPI, Model_SEI, and Model_PI) for the four selected models that included environmental covariate effects (Supplementary Material V).

Analysis

We tested the single effect of the four Cold Pool indices (Obs_CPI, Model_CPI, Model_PI, and Model_SEI) with three ways to incorporate each index (controlling, limiting, or masking factors). Then, we fit a set of models with the combined effects of two Cold Pool indices by testing all pairwise combinations of Cold Pool indices, excluding the Obs_CPI, with the three factor types (Table 1). Finally, we fit a model without any Cold Pool effect as a control. In order to use Akaike's Information Criterion (AIC) to compare the aforementioned Cold Pool effect scenarios, the four Cold Pool indices were included in the 25 models. Hence, the observations are included in the likelihood for the different models, but whether their effect on recruitment is estimated depends on the model in Table 1. Based on AIC, we selected five models for further analysis. These included the best fit (lowest AIC) model with one Cold Pool index, the two best models with combined Cold Pool effects (two CPI effects), the best model including effects of the Obs_CPI, and the base model without any CPI effects (denoted m1, m2, m3, m8, and m25 in Table 1). For these five selected models, we compared the estimates of recruitment, SSB, and fully selected fishing mortality (F). These estimates for the three other models with the best fit and $\Delta AIC \leq 2$ (denoted m4, m5, and m6 in Table 1) are presented in Supplementary Material VI. Furthermore, we produced two sets of 4-year predictions for years 2019–2022 under the assumption that future F is at the level that produces the MSY (F_{MSY}; Miller *et al.*, 2016; Xu *et al.*, 2018) or at the level that produces spawning biomass per recruit representing 40% of the unfished spawning biomass per recruit ($F_{40\%}$; Miller and Legault, 2017) for the five selected models.

Unlike $F_{40\%}$, F_{MSY} is a function of the stock–recruit relationship and, depending on the type of effect, also the Cold Pool indices (Miller *et al.*, 2016; Miller and Stock, 2020). Thus, we estimated F_{MSY} into the projection period over the range of estimated Cold Pool indices for the four selected models that included environmental covariate effects (Supplementary Material VII). This provided insights regarding the variations of the Cold Pool indices on the F_{MSY} estimates.

Next, we explored the impact of incorporating Cold Pool effects on the retrospective pattern generated through refitting the model to the data after removing its terminal year sequentially (Mohn, 1999). The five selected models were compared with respect to the Mohn's ρ for recruitment (ρ_R), ρSSB (ρ_{SSB}), and F (ρ_F), which were calculated from seven retrospective peels, the typical number used in Northeast US assessments (Deroba, 2014; Miller and Legault, 2017). To evaluate the models' ability to provide consistent predictions, we generated a series of retrospective predictions (Brooks and Legault, 2016; Xu *et al.*, 2018). For each model, 15 retrospective 4-year predictions between 2000 and 2014 were generated as in Xu *et al.* (2018). The models were fitted to data between 1973 and terminal years from 2000 to 2014, and we produced a series of 4-year predictions beyond the terminal year of data. First, we fitted a model to 1973–2000 and predicted recruitment and SSB from 2001 to 2004, then we fitted a model to 1973–2001 and predicted recruitment and SSB from 2002 to 2005, and so on until 2014. In the projection years, we used the F from the assessment fit using all data from 1973 to 2018. The process was repeated 15 times between 2000 and 2014. We compared the recruitment and SSB produced from the 15 retrospective predictions and from the corresponding models using the full data from 1973 to 2018.

To quantify the effects of incorporating the different Cold Pool indices in WHAM on the predictive performance, we calculated the mean absolute differences (MAD) between the retrospective recruitment (and SSB) predictions and the models using the full data for each prediction lead year (from 1 to 4 years) such that

$$MAD_t = \frac{1}{n}\sum_{t=1}^{n} |y_{t,i} - x_t|, \quad (8)$$

where n is the number of years, $y_{t,i}$ is the prediction for the year t and the lead year i, and x_t is the model estimate in year t when fit to the full data.

Results

Cold Pool indices

The time series of Obs_CPI and Model_CPI are strongly correlated ($r = 0.78$, p-value = 2.26e-10; Table 2) despite large differences in some years, e.g. 1975, 1990, and 2012 (Figure 1a and Table 2). The increasing trend of Cold Pool indices is the result of warming ocean temperatures in the MAB since 1980. However, Obs_CPI has higher observation error than Model_CPI due to the differences in data resolution (resolution of modelled data is higher than observed data). Model_PI was negatively correlated with Obs_CPI ($r = -0.76$, p-value = 1.9e-09) and Model_CPI ($r = -0.93$, p-value < 2.2e-16; Table 2), meaning that a weaker Cold Pool was associated with shorter persistence time (Figure 1b). Model_SEI was negatively correlated with Obs_CPI and Model_CPI (Table 2), although there was much more variation in the minimum spatial extent than in the maximum (Figure 1c). The intensity and

Table 2. The Pearson correlation coefficients between SNEMA yellowtail flounder recruitment (in log scale) and the Cold Pool indices (Obs_CPI and Model_CPI), the persistence index (Model_PI), and the spatial extent index (Model_SEI).

Variable1	Variable2	Correlation	p-value
Obs_CPI	Model_CPI	0.78	2.3e-10
Obs_CPI	Model_PI	−0.76	1.9e-09
Obs_CPI	Model_SEI	−0.64	1.9e-06
Model_CPI	Model_PI	−0.93	< 2.2e-16
Model_CPI	Model_SEI	−0.87	2.7e-15
Model_PI	Model_SEI	0.89	< 2.2e-16
Obs_CPI	Recruitment	−0.65	1.5e-06
Model_CPI	Recruitment	−0.66	7.4e-07
Model_PI	Recruitment	0.60	1.3e-05
Model_SEI	Recruitment	0.64	2.6e-06

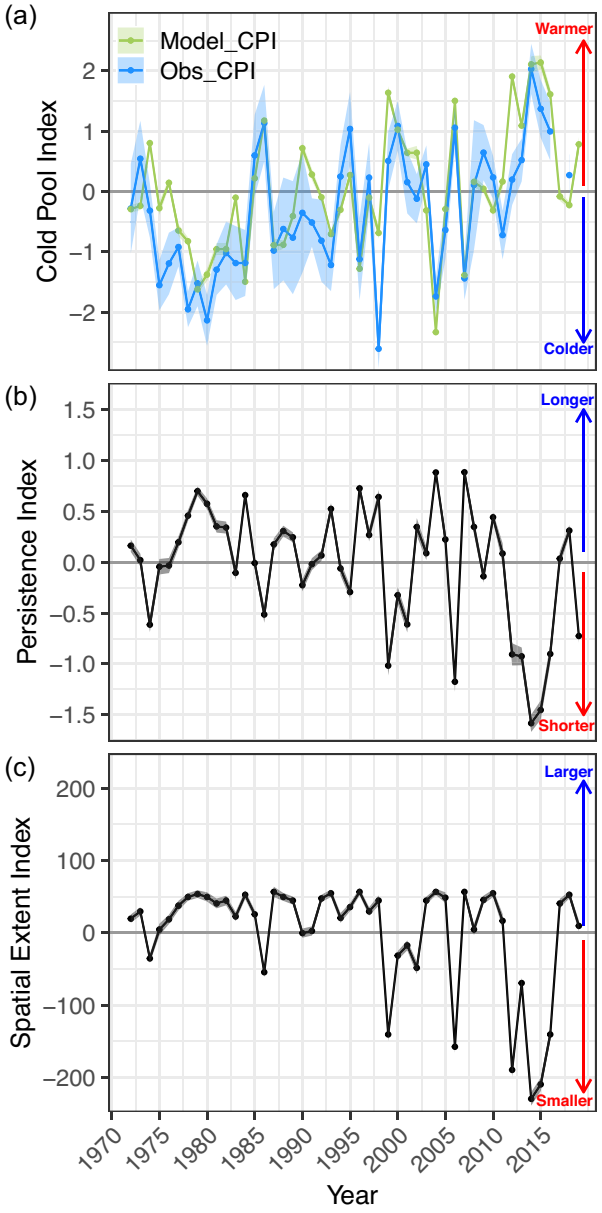

Figure 1. Cold Pool time series between 1972 and 2018. Panel (a) represents the Obs_CPI calculated using *in situ* observations as described in Miller *et al.* (2016) vs. the Model_CPI. Panel (b) shows the Model-Persistence Index. Panel (c) represents the Model-Spatial Extent Index. The polygons and the dashed lines represent the 95% *CI*s.

frequency of negative Model_SEI events have increased from 1999 onward (Figure 1c). The absolute values of correlations between the Obs_CPI and the three model-based indices were between 0.64 and −0.78, and the correlations between all four Cold Pool indices and the natural log of recruitment estimated in the last stock assessment (NEFSC, 2020) were between 0.60 and 0.66 (Table 2).

Cold Pool effects within the stock assessment model

Each of the 25 models (Table 1) converged and produced an invertible Hessian matrix (see diagnostic plots in Supplementary Material III, S3.4 and 3.5).

Based on AIC, the models including Cold Pool indices performed better than the model that did not integrate Cold Pool effects (m25, ΔAIC = 22.6). Moreover, two of the three models (m1 and m4) integrating Model_CPI had lower AIC than all the models that integrated Obs_CPI (m8, m13, and m14). The three models in which Model_CPI was included as a controlling factor (m1, m2, and m3) performed best. Among these models, the one with only one index (m1) had slightly lower AIC than those incorporating two indices (m2, ΔAIC = 0.6; m3, ΔAIC = 1.1). We then focused further analyses on the three best models (m1, m2, and m3), the best model with Obs_CPI (m8), and the model without any environmental covariates (m25).

Incorporating Cold Pool indices in the selected models (m1, m2, m3, and m8, hereafter, called g(Model_CPI), g(Model_CPI + PI), g(Model_CPI + SEI), and g(Obs_CPI), respectively) resulted in a reduced retrospective pattern of SSB, *F*, and recruitment, but was most pronounced in recruitment (Figure 2). Using model-based Cold Pool indices resulted in further reduced retrospective patterns compared to the Obs_CPI (Figure 2). While the assessment model that used Model_CPI in combination with Model_PI had only the second best fit in terms of AIC, it had the lowest retrospective pattern (reduction of $|\rho_R|$ by 0.24, $|\rho_{SSB}|$ by 0.04 and $|\rho_F|$ by 0.04 compared with the model without Cold Pool indices; Figure 2).

In the models with environmental covariates, the estimated stock–recruitment relationship varied depending on the way we linked the Cold Pool indices to recruitment ("controlling factor", "limiting factor", and "masking factor"; Supplementary Material V). The variations of the Cold Pool indices are closely associated to the large inter-annual variations in recruitment (Supplementary Material VIII) with, overall, larger recruitment associated with stronger and longer or larger Cold Pool. Moreover, in the model including two covariates, the stock–recruitment relationship is more closely associated to Model_CPI than to Model_PI and Model_SEI (Supplementary Material V).

The five assessed models (and the other best models m4, m5, and m6 in Table 1) estimated very similar recruitment, SSB, and *F* between 1972 and 2018 at first glance (Figure 3a, c, and e, and Supplementary Material VI). However, the analysis of estimates relative to Base revealed inter-annual differences in recruitment, which are, on average, 10.5%, 9.7%, 9.8%, and 10.0% (2161 t, 2144 t, 1888 t, and 2092 t in absolute value) for g(Obs_CPI), g(Model_CPI), g(Model_CPI + PI), and g(Model_CPI + SEI), respectively, over the period 1972–2018 (Figure 3b). The differences were similar among the models incorporating model-based indices and reached −27% in 2013 and −41% in 2016. g(Obs_CPI) did not show the same variations in recruitment estimates compared to Base with largely higher recruitment, for example, in 1976 (+26%) and 1999 (+30%) and lower recruitment, for instance, in 1974 (−9%) and 1987 (−31%). The SSB and F estimates showed modest differences relative to Base with average differences between 1.6 and 2.4% for SSB and between 1.4 and 1.9% for *F* (Figure 3d and f). But the differences in SSB and F estimates using model-based indices are opposite of those from g(Obs_CPI). All models that included Cold Pool effects had decreased uncertainty in recruitment

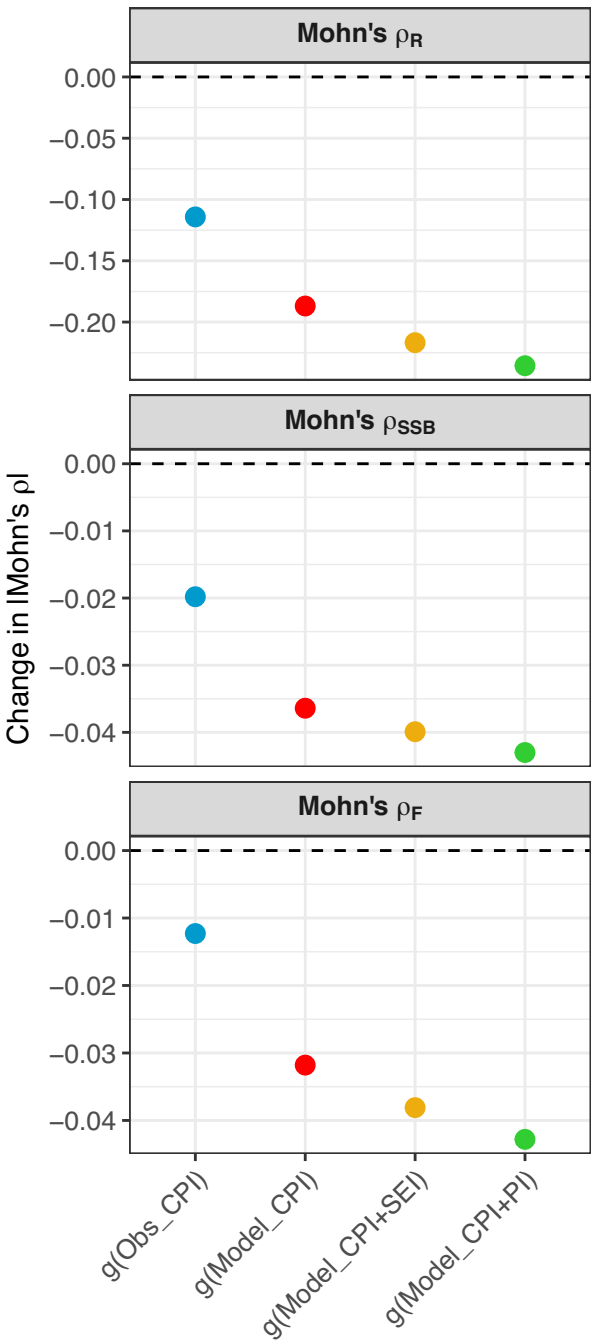

Figure 2. Change in Mohn's ρ of key quantities (*F*: "fully selected" fishing mortality rate, recruitment, and SSB: spawning stock biomass) relative to the model without environmental covariates for three models including environmental covariates.

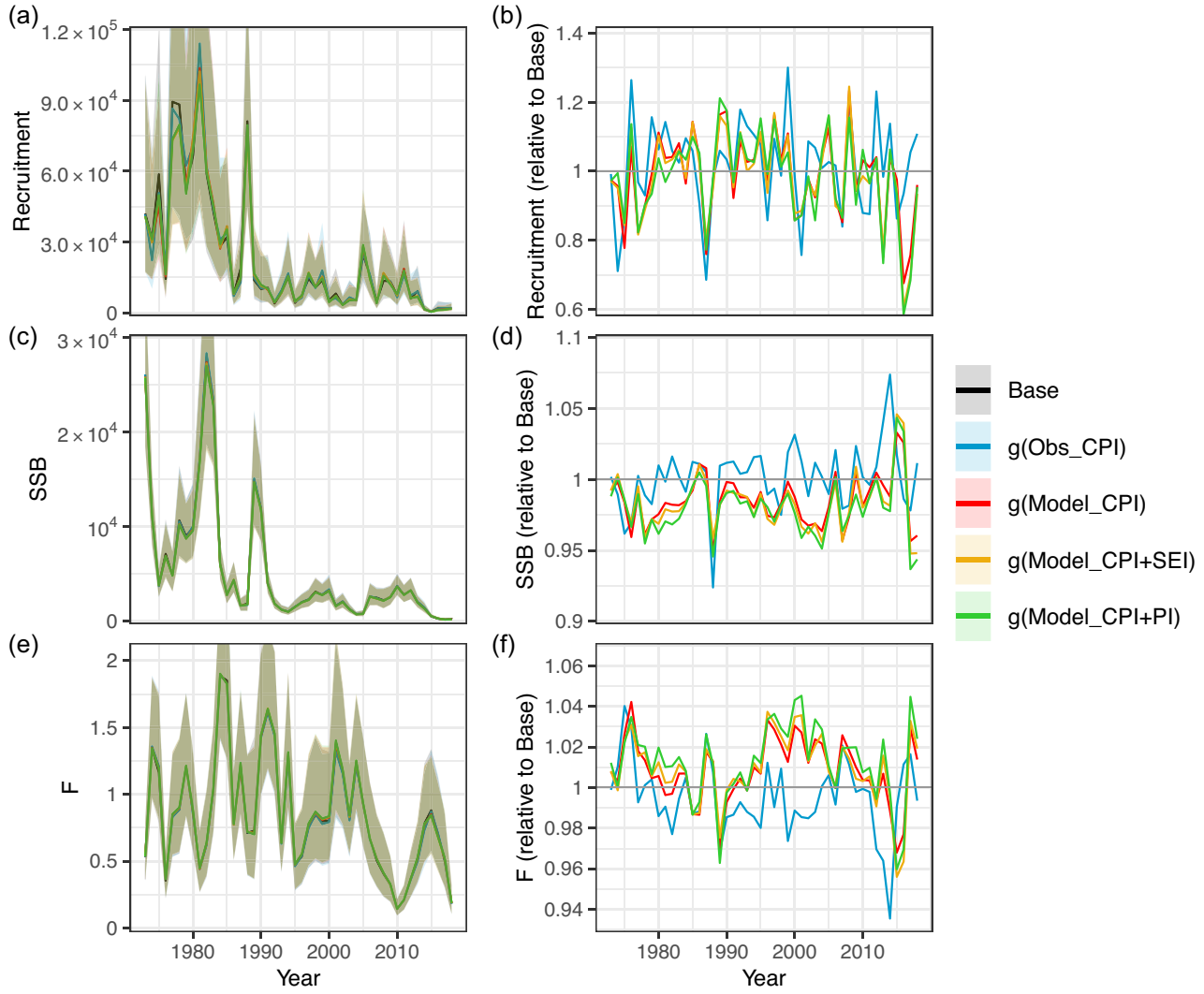

Figure 3. Annual raw (left column) and relative (to the model without environmental covariates, right column) estimates for recruitment (a and b), SSB (c and d), and fully selected fishing mortality (*F*) (e and f). Annual estimates from the four selected models are included. The coloured polygons represent the 95% *CI*s. On the left column, recruitment, SSB, and *F* from the five models are represented but overlap on each of the three panels (a), (c), and (e).

of about 5% between 1973 and 2018 compared to Base (left column, Figure 4). Including the model-based Cold Pool indices also reduced the uncertainty in SSB and *F* by a modest amount (1.5–1.9%), whereas the Obs_CPI did not (left column, Figure 4).

Cold Pool effects on stock projections

We projected the four Cold Pool indices and their errors through 2021, by continuing the autoregressive process model (Figure 5). While Model_CPI was projected to increase slightly, linked to the continued warming of the Cold Pool domain, the Obs_CPI was projected to decrease indicating improved ocean conditions for SNEMA yellowtail flounder recruitment. The Model_PI and Model_SEI were predicted to decrease slightly in the projection years. In summary, the Cold Pool was projected to be weaker, shorter, and cover less space using the model-based data products. The uncertainty of the four estimates was much higher during the projection period as expected under the AR(1) model when predicting beyond the period with observations. The uncertainty of Obs_CPI was also high in 2017 because observed data were not available this year to calculate the index and it was estimated by the AR(1) model (Figure 4a).

Overall, the short-term projections of SSB, and recruitment with $F = F_{40\%}$ were higher than stock estimates calculated with $F = F_{MSY}$ (Figure 6). When $F = F_{40\%}$, fishing mortality was the same for all models, while F_{MSY} varied with different Cold Pool-recruitment effects (Figure 6f; Supplementary Material VII). When recruitment was a function of the Cold Pool and SSB, both recruitment and SSB projections were much higher than in the model when recruitment was solely a function of SSB (Figure 6a–d). The three models with the model-based Cold Pool indices produced similar SSB and recruitment estimates in short-term projections, with 30–31% higher recruitment and 12–14% larger SSB than Base (Figure 6a–d). Using the observation-based Cold Pool projected even higher stock estimates, with 60–61% higher recruitment and 29–38% higher SSB than Base over the period 2019–2022 (Figure 6a–d).

Model with two Cold Pool effects had similarly uncertain short-term recruitment projections compared to Base, while

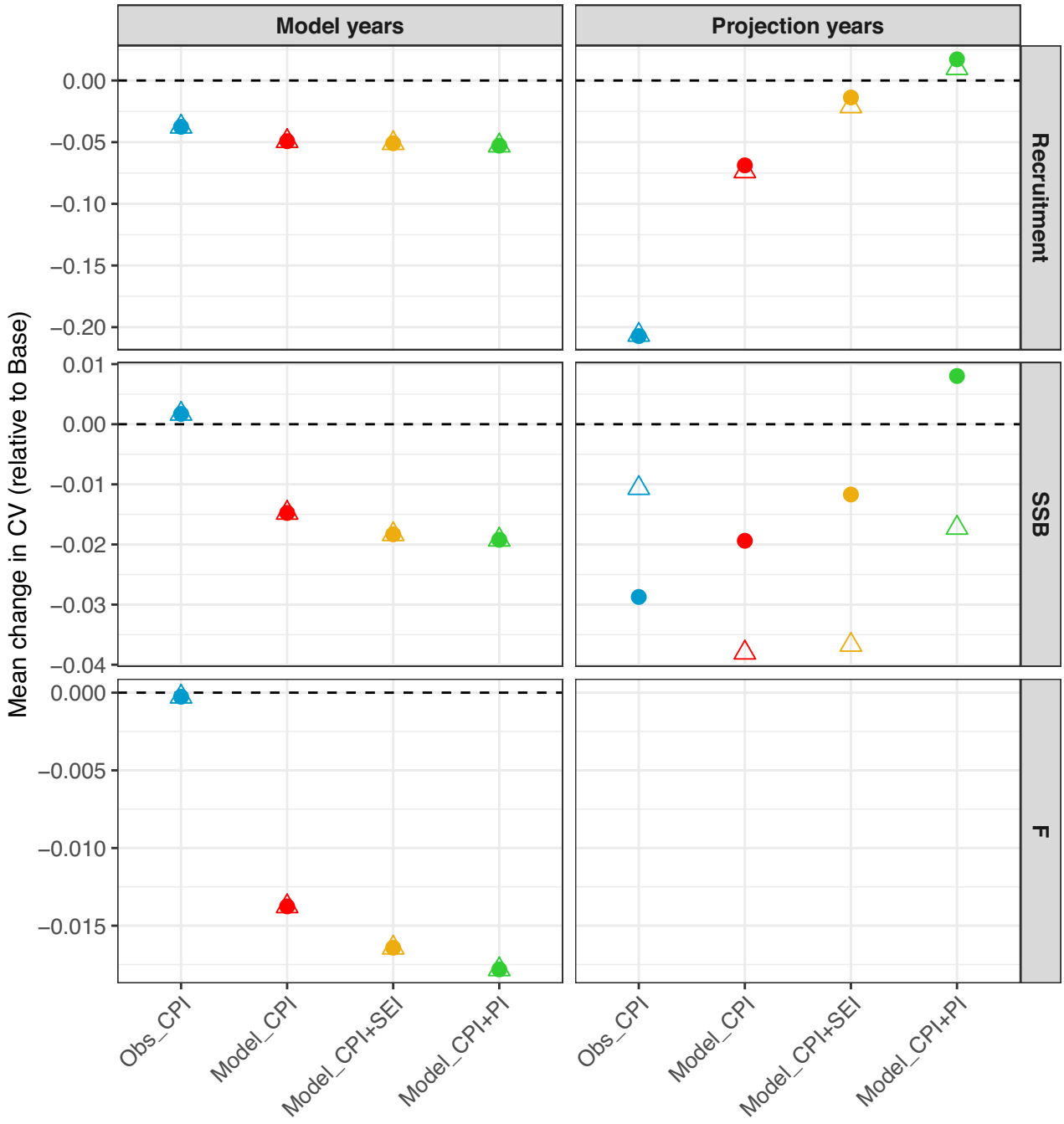

Figure 4. Mean change in coefficient of variation (*C.V.*) for key quantities (F: "fully selected" fishing mortality rate, recruitment, and SSB: spawning stock biomass) in model and projection years relative to the model without environmental covariates for three models including environmental covariates. For the projected year, the triangles and the points represent the *C.V.* for projections in which $F = F_{MSY}$ and $F = F_{40\%}$, respectively.

models including only one effect had 7% and 21% lower uncertainty (Model_CPI and Obs_CPI), respectively. The uncertainty in SSB in the projection period depended on whether F_{MSY} or $F_{40\%}$ was used, but the mean changes in the coefficient of variation compared to Base did not exceed 4% in either the projected or model years (Figure 4).

All five models predicted higher recruitment and SSB estimates (coloured lines in Figure 7 and Supplementary Material IX) than estimates when performing the assessment on full data (black lines) between 2000 and 2018. Moreover, the estimates of the models incorporating the Cold Pool effects were generally lower than Base except in a few years (Figure 7 and Supplementary Material IX).

The incorporation of Cold Pool indices substantially improved the recruitment retrospective predictions after 1 year compared to Base as suggested by the low MAD ratio (Figure 8). As prediction lead time increased, the recruitment predictions from the four models including Cold Pool indices deviated from the estimates when performing the assessment on full data (MAD ratio close to 1), and were negligibly improved in the third and fourth prediction years. In contrast, the SSB retrospective predictions were improved (relative to

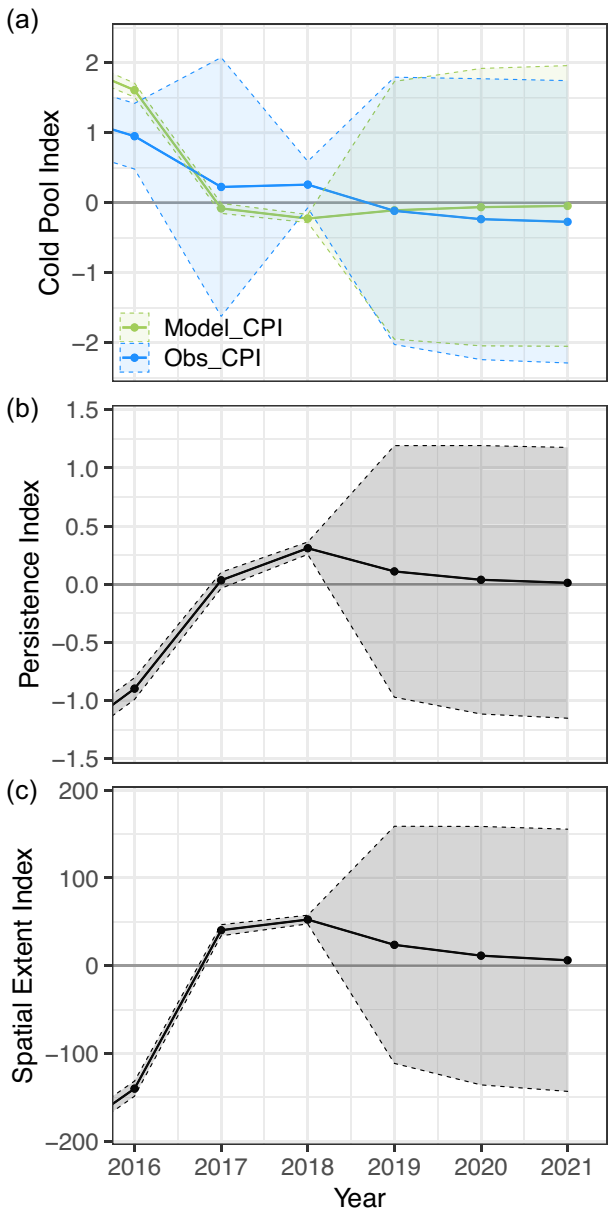

Figure 5. Cold Pool indices time series between 2016 and 2021. Panel (a) the Obs_CPI calculated using *in situ* data as described in Miller *et al.* (2016), and the Model_CPI calculated using ocean model products. Panel (b) shows the Model-persistence index. Panel (c) represents the Model-spatial extent index. The polygons represent 95% *CI*s, and the dashed line marks the terminal year in the assessment (2018).

the estimates when performing the assessment on full data) in the second and, to a lesser extent, in the third prediction years, although the improvement was less pronounced than for recruitment. For both recruitment and SSB predictions, the model that incorporated Model_CPI and Model_PI exhibited the best predictive performance (lower MAD ratio) while the model with Obs_CPI had the highest MAD ratio.

Discussion

In this study, we used environmental indices from observations, a regional ocean model hindcast, and a global ocean reanalysis and incorporated them into a state–space stock assessment framework for SNEMA yellowtail flounder. Our study demonstrates that incorporating Cold Pool effects on SNEMA yellowtail recruitment reduced the retrospective patterns and improved the predictive skills of recruitment and, to a lower extent, SSB compared to the base model without any Cold Pool effects. Furthermore, the performance of the models incorporating ocean model-based indices was improved compared to the one with the observation-based index. However, the incorporation of ocean model-based indices did not reduce the uncertainties and temporal trends in stock estimates over the period of 1972–2018. Finally, the improvement of the recruitment retrospective prediction with ocean model-based indices suggests that incorporating ocean model-based indices may benefit the predictive skills of the stock assessment models.

The mechanisms leading to environmentally induced variations in recruitment

In this study, we used a state–space age-structured assessment model to incorporate the Cold Pool effect into the Beverton–Holt stock–recruit function in the SNEMA yellowtail flounder assessment. We explored different mechanisms leading to environmentally induced variations in recruitment ("controlling factor", "limiting factor", and "masking factor") as well as the simultaneous variations of two of these processes. Our results showed that the models incorporating Obs_CPI and Model_CPI performed better than Model_SEI and Model_PI. Moreover, the inter-annual variations of the two CPIs affect the stock–recruit relationship more than the latter two (Model_SEI and Model_PI; Supplementary Material V). Our results also showed that the incorporation of the Model_SEI and Model_PI in the model as a secondary Cold Pool effect did not add substantial improvement despite a slight reduction of the retrospective patterns of recruitment (Figure 2) and very little improvement of retrospective predictions (Figure 7).

The three models that performed the best were the ones in which Model_CPI was a controlling factor, suggesting that the Cold Pool strength affected both the maximum recruits per spawning biomass (which could be due to an increase in mortality rates) and the maximum recruitment (i.e. the carrying capacity). Weak Cold Pool events may have impacted recruitment when SSB was low with decreases in recruits per spawning biomass, but also when SSB was high with reductions in the carrying capacity (Maunder and Thorson, 2019). This implies that Cold Pool variations have affected SNEMA yellowtail flounder recruitment when SSB was high before the 1990s and continued to affect recruitment after the 1990s when SSB was well-below its past level. This could be one of the reasons why SSB has remained at a very low level for over 30 years.

The best model with Obs_CPI assumed that the Cold Pool was a limiting factor for recruitment, meaning that the Cold Pool affects only the carrying capacity of the settlement area. The differences in mechanisms leading to variation in recruitment between Obs_CPI and Model_CPI may be due to the different time period considered. Obs_CPI was estimated using bottom temperature in September and October, thus it covered only the end of the settlement season. Therefore, Obs_CPI only accounts for a portion of the Cold Pool effects on recruitment.

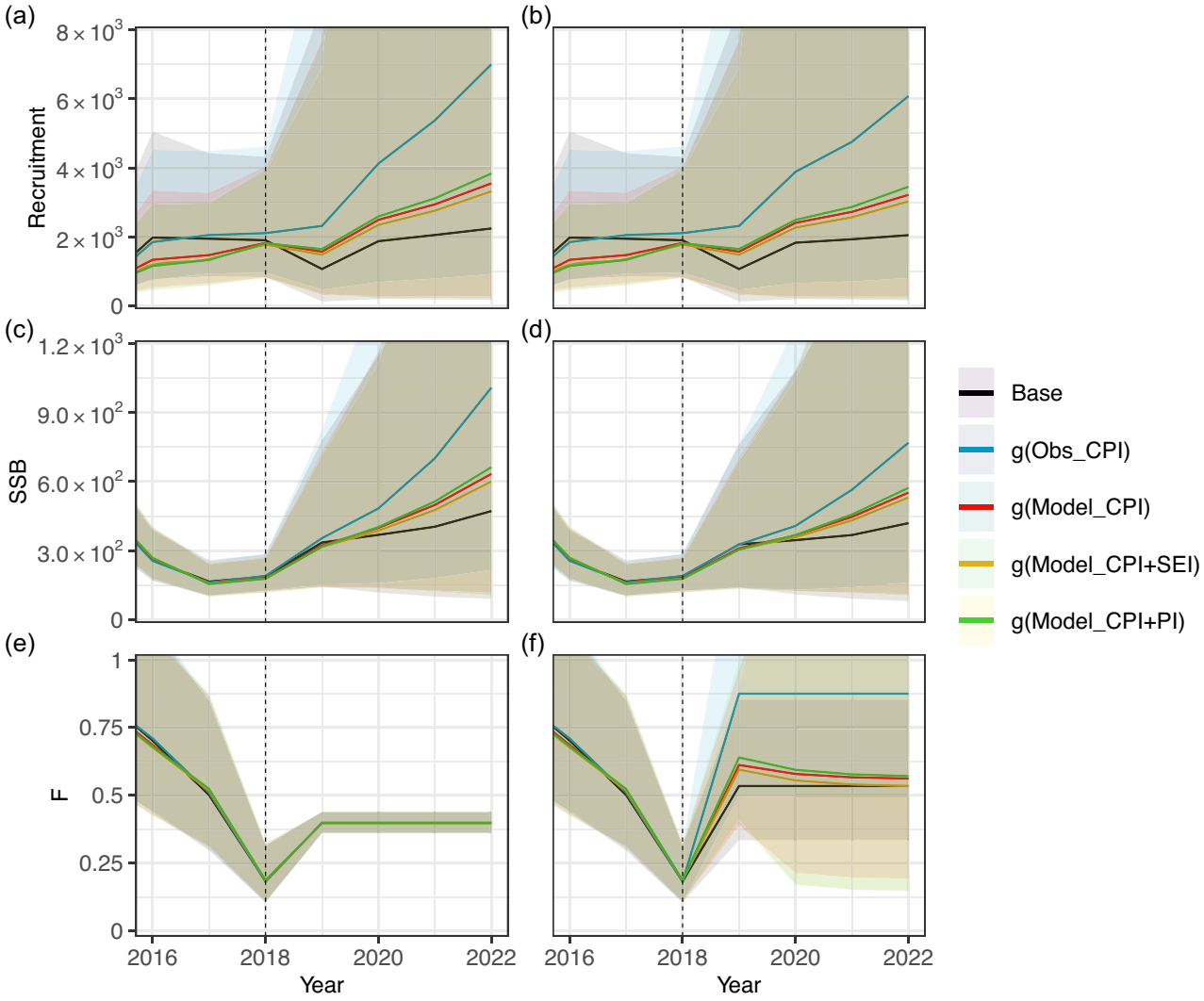

Figure 6. Estimates of projected recruitment (a) and (b), SSB (c) and (d) and fully selected fishing mortality (F) (e) and (f) for $F = F_{40\%}$ (left column, a, c, and e) and $F = F_{MSY}$ (right column, b, d, and f). The coloured polygons represent the 95% *CI*. The brown confidence intervals come from the superposition of the confidence intervals of the 5 models.

Using ocean models to integrate environmental effects

Using bottom temperature from ocean models provides a high spatial and temporal resolution of historical ocean conditions 3D, which is required to calculate the Cold Pool indices analyzed in this study. Furthermore, gaps in ocean observations, such as those in 2017 and 2020 in the Obs_CPI, are avoided using ocean models. Using bottom temperature from ocean models also allowed us to explore the effects of alternative Cold Pool indices to those developed and tested in Miller *et al.* (2016) and Xu *et al.* (2018). This historical time-series of bottom temperature in the northwest Atlantic may be valuable for other stock assessments that may benefit from a high-resolution ocean bottom temperature hindcast. Indeed, several other stocks on the Northeast US shelf have identified links between bottom temperature and population processes (Supporting Information in Chen *et al.*, 2021), such as Georges Bank and Gulf of Maine Atlantic Cod (Pershing *et al.*, 2015; Miller *et al.*, 2018), Georges Bank and Gulf of Maine Atlantic American Lobster (Tanaka *et al.*, 2019), and SNEMA Winter Flounder stock (Bell *et al.*, 2018).

A stock assessment model to account for changing ocean conditions

Over the past 45 years, the Cold Pool has become weaker, shorter in duration, and smaller in size (Supplementary Material X). The SNEMA yellowtail flounder stock may be more sensitive to these changing ocean conditions because it is the southernmost stock in the northwest Atlantic (NEFSC, 2012). Moreover, its recruitment is likely one of the most sensitive biological parameters to the environment, given that the settlement of the pre-recruits during the Cold Pool event represents a bottleneck in its life history during which the local and temporary decrease in bottom temperature impacts the survival of the settlers (Sullivan *et al.*, 2000, 2005; Haltuch *et al.*, 2019). This well-defined environmental pressure at a specific life stage is likely one of the reasons why including environment in the assessment provides good performance in terms of retrospective pattern and short-term forecast (Haltuch *et al.*, 2019). Furthermore, the structure of WHAM may allow improving the assessment framework by combining estimates of time-varying parameters as random effects and the incorporation of the interannual variations and the long-term changes

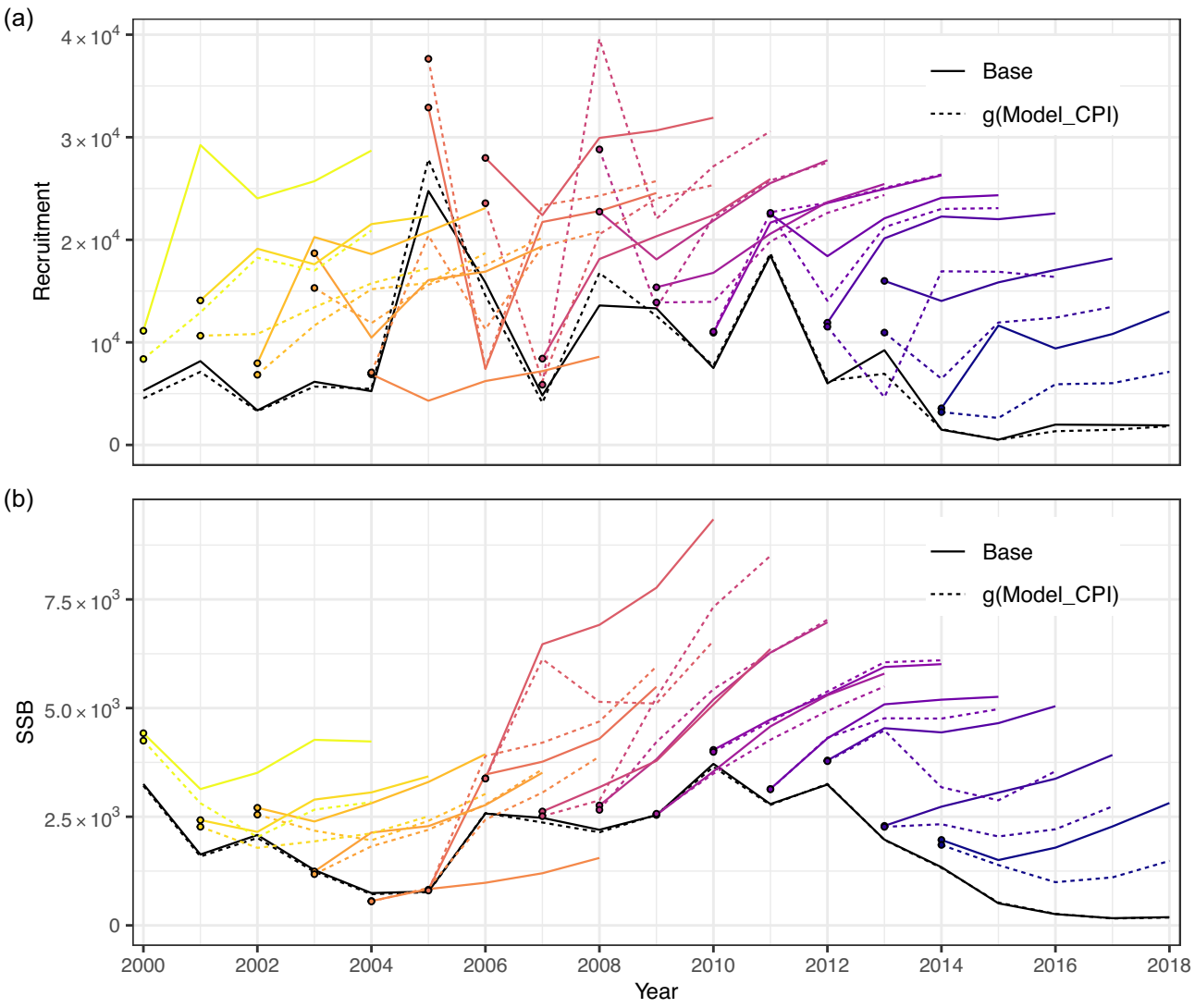

Figure 7. The retrospective prediction patterns from the model without environmental covariates (Base) and the model in which Model_CPI affects SSB and recruitment estimates (g(Model_CPI)). The two black lines represent the SSB and recruitment estimates when performing WHAM on full data. The dots are the terminal year recruitment estimates for each retrospective prediction and the coloured lines represent the 15 retrospective predictions, from yellow (2003) to dark blue (2014).

in bottom temperature directly on recruitment (Miller *et al.*, 2016; Stock and Miller, 2021). We suggest the incorporation of environmental variables in other stock assessment models using WHAM to determine if retrospective patterns can be reduced and if prediction skill can be improved.

Application to stock assessments

While reducing the retrospective patterns in the terminal year model estimate is a key step to improve confidence in the results of stock assessments, improving the model prediction skill is crucial to produce short-term (i.e. 1–3 years) population forecasts and reliable catch advice. Our results suggest that the incorporation of the Cold Pool into the Beverton–Holt stock–recruit function can reduce retrospective patterns in near-term recruitment predictions and the reduction is more pronounced when using ocean model-based indices. While the different models we tested showed similar trends in stock estimates when there were catch and survey data (between 1973 and 2018), recruitment and SSB estimates exhibited large differences during the prediction period. The large divergence of predictions highlights that care should be taken when environmental covariates are incorporated in the assessment model as this can lead to significant impacts on management decisions and short-term catch advice. The two fishing mortality scenarios used for the projections period (either $F_{40\%}$ or F_{MSY}) induced major differences in projections because F_{MSY} (instead of proxy as $F_{40\%}$) integrates the environmentally induced variations in the stock–recruits parameters, which can be valuable in a highly variable system such as the Cold Pool. However, the environmental covariates affect F_{MSY} only when they act as a "controlling factor" (Supplementary Material VII) and the method used can lead to very high F_{MSY} as we can see with g(obs_CPI) (Figure 5f).

Future work should evaluate the true bias or relative error of the predictions of recruitment and SSB through a simulation study where the true values being estimated are known. Alternative hypotheses on the effects of the Cold Pool on recruitment could be treated as operating models and all of the es-

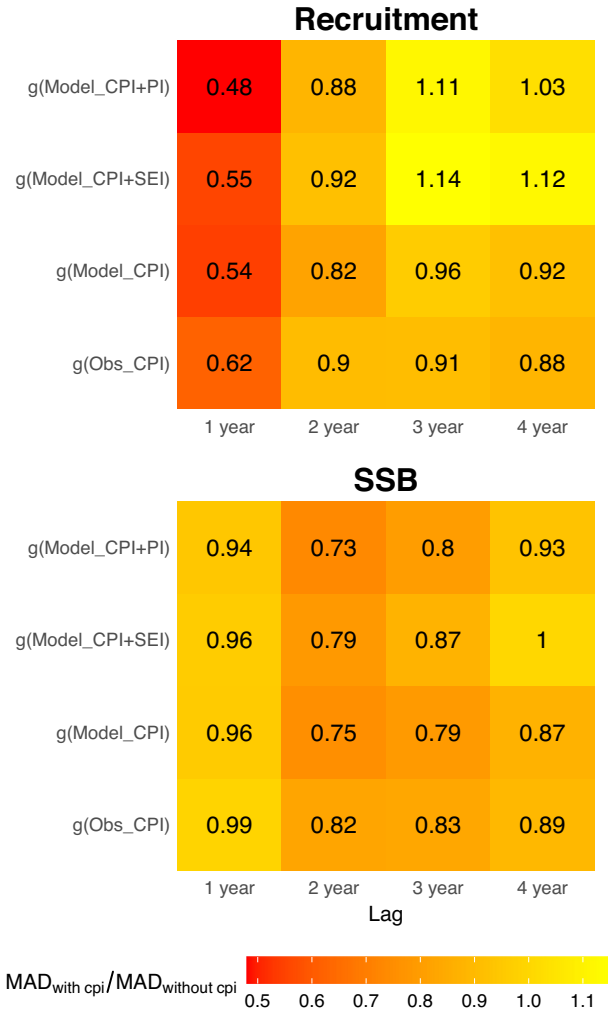

Figure 8. The MAD of retrospective predictions of recruitment and SSB for models with Cold Pool effects relative to the model without Cold Pool effects.

timating models could be fitted to each hypothetical scenario and evaluate whether any of the estimation models are robust to the hypothetical stock–recruitment relationships (Deroba *et al.*, 2015; He and Field, 2019; Perretti *et al.*, 2020). Similarly, a closed-loop management strategy evaluation simulation study with the same alternative hypothetical states of nature, but where catch advice is determined using alternative harvest control rules could determine whether any of the estimation models provide better management metrics (total yield, likelihood of being overfished, and so on) than the others (Punt, 2003; Szuwalski and Punt, 2013).

To operationalize the environment-linked assessment model, bottom temperature data would need to be available at the same time as the biological data. However, reanalysis products such as GLORYS12v1 rely on an intensive computational process and publish a new year of bottom temperature data with a lag of more than 1 year, which is a major operational limitation. A potential way around this would be to combine the model- and observation-based approaches by using the ocean model-based index and the observation-based index. Seasonal statistical models (e.g. Chen *et al.*, 2021) could also be used to produce short-term bottom temperature forecasts based on the last bottom temperature observations available. Figure 9 illustrates how the Cold Pool index could be included in the SNEMA yellowtail flounder assessment operationally..

In order to improve the incorporation of environmental indices into stock assessment models using output from ocean models, further initiatives would be necessary to get high-resolution, skillful bottom temperature over the Northeast US shelf in near-real time or with a few months lag. This ocean model output can be used to estimate the different Cold Pool indices using the method presented in this study. The product, *Operational Mercator global ocean analysis*, released by the E.U. Copernicus Marine Service Information (https://resources.marine.copernicus.eu) provides near real-time bottom temperature at a global scale and the model features are similar to those of GLORYS12v1. However, this product is not reanalyzed (in contrast to GLORYS12v1), so its skill to reproduce the spatio-temporal Cold Pool variations should be analyzed in detail before it can be used in the stock assessment model. Furthermore, longer-term hindcasts from the 1970s to present would avoid combining two ocean models as we did here with GLORYS12v1 and ROMS-NWA. Another major step to improve the stock assessment short-term projections and refine fisheries management decisions would be to introduce skillful seasonal to annual forecasts of bottom temperature over the Northeast US shelf.

Our study demonstrates that environmental covariates derived from ocean models can be incorporated in stock assessment models. However, the use of ocean models should be done with caution and a preliminary identification of model bias is essential. Observation-based climatologies developed in different regions (e.g. Johnson and Boyer, 2014; Seidov *et al.*, 2016a, 2017) may be considered as an efficient tool to bias-correct ocean models when substantial biases are identified. Moreover, the identification of the well-defined environmental effect on a specific life stage is another crucial prerequisite to incorporate robustly environmental covariates in stock assessment model (Haltuch *et al.*, 2019). Hence, skillful ocean model-based covariates may be used to incorporate environmental effects into well-defined time-varying population processes in marine fish stock assessments. Instead of relying on limited subsurface ocean observations, using ocean model as environmental covariates in stock assessments may both improve predictions and facilitate operationalization.

Supplementary data

Supplementary material is available at the ICESJMS online.

Authors' contributions

HdP implemented and ran the models, analyzed the data, interpreted the results, and drafted the manuscript. TM and BC developed WHAM, including several specific developments for this study, analyzed the data, and interpreted the results. VS initially designed the study, analyzed the data, and interpreted the results. ZC provided the ROMS-NWA output and helped interpret the ocean model outputs. All the authors were involved in the methodology and conceptualization of the study, edited the manuscript, and gave their final approval for publication.

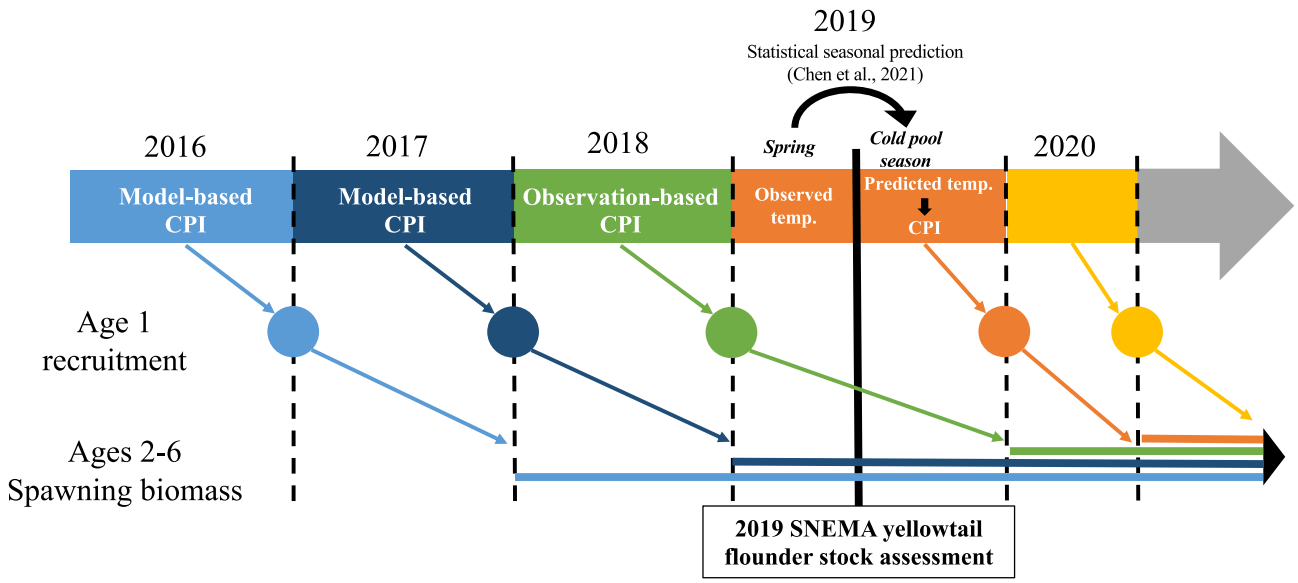

Figure 9. Conceptual design of the potential approach to use the latest available Cold Pool index (CPI) in the SNEMA yellowtail flounder stock assessment. Here, we used the example for the assessment in 2019 if the Cold Pool index had been introduced. In the 2019 SNEMA yellowtail flounder assessment conducted in summer, GLORYS12v1 could be used to calculate the Cold Pool index between 1972 and 2017 and the observation-based index could be used for 2018. Moreover, the seasonal statistical predictive model developed by Chen et al. (2021) could be applied using in situ observed bottom temperature collected by the NOAA NEFSC in the spring of 2019. This statistical model would have produced bottom temperature forecasts during the Cold Pool season in 2019 which would be used to calculate the Cold Pool index in 2019. The incorporation of the Cold Pool index could have been used to predict the recruitment in 2020.

Data availability statement

The GLORYS12v1 ocean reanalysis data set is available at the Copernicus Marine Environment Monitoring Service (CMEMS; https://resources.marine.copernicus.eu/product-detail/GLOBAL_MULTIYEAR_PHY_001_030/INFORMATION). The bottom temperature ROMS-NWA output will be shared on reasonable request to Zhuomin Chen (zhuomin.chen@uconn.edu). The Northwest Atlantic Ocean regional climatology is available at the National Center for Environmental Information (NCEI; https://www.ncei.noaa.gov/products/northwest-atlantic-regional-climatology). The code and details regarding WHAM can be found on the GitHub page: https://github.com/timjmiller/wham.

Funding

Funding for Hubert du Pontavice was provided by the NOAA NEFSC's "New England's Groundfish in a Changing Climate" program.

Competing interest statement

The authors declare no competing interests.

Acknowledgements

We thank Chris Legault for comments on an early draft of the paper.

References

Bell, R. J., Hare, J. A., Manderson, J. P., and Richardson, D. E. 2014. Externally driven changes in the abundance of summer and winter flounder. ICES Journal of Marine Science, 71: 2416–2428.

Bell, R. J., Wood, A., Hare, J., Richardson, D., Manderson, J., and Miller, T. 2018. Rebuilding in the face of climate change. Canadian Journal of Fisheries and Aquatic Sciences, 75: 1405–1414.

Brooks, E. N., and Legault, C. M. 2016. Retrospective forecasting—evaluating performance of stock projections for New England groundfish stocks. Canadian Journal of Fisheries and Aquatic Sciences, 73: 935–950.

Chen, Z., Curchitser, E., Chant, R., and Kang, D. 2018. Seasonal variability of the Cold Pool over the Mid-Atlantic Bight continental shelf. Journal of Geophysical Research: Oceans, 123: 8203–8226.

Chen, Z., and Curchitser, E. N. 2020. Interannual variability of the Mid-Atlantic Bight Cold Pool. Journal of Geophysical Research: Oceans, 125: 1–20.

Chen, Z., Kwon, Y., Chen, K., Fratantoni, P., Gawarkiewicz, G., Joyce, T. M., Miller, T. J. *et al.* 2021. Seasonal prediction of bottom temperature on the northeast U.S. continental shelf. Journal of Geophysical Research Oceans, 126: e2021JC017187.

Deroba, J. J. 2014. Evaluating the consequences of adjusting fish stock assessment estimates of biomass for retrospective patterns using Mohn's rho. North American Journal of Fisheries Management, 34: 380–390.

Deroba, J. J., Butterworth, D. S., Methot, R. D., De Oliveira, J. A. A., Fernandez, C., Nielsen, A., Cadrin, S. X. *et al.* 2015. Simulation testing the robustness of stock assessment models to error: some results from the ICES strategic initiative on stock assessment methods. ICES Journal of Marine Science, 72: 19–30.

Fernandez, E., and Lellouche, J. M. 2018. Product user manual for the global ocean physical reanalysis product GLORYS12V1. Copernicus Product User Manual, 4: 1–15.

Forsyth, J. S. T., Andres, M., and Gawarkiewicz, G. G. 2015. Recent accelerated warming of the continental shelf off New Jersey: observations from the CMVOleander expendable bathythermograph line. Journal of Geophysical Research Oceans, 120: 2370–2384.

Haltuch, M. A., Brooks, E. N., Brodziak, J., Devine, J. A., Johnson, K. F., Klibansky, N., Nash, R. D. M. *et al.* 2019. Unraveling the recruitment problem: a review of environmentally-informed forecasting and management strategy evaluation. Fisheries Research, 217: 198–216.

Hare, J. A., Borggaard, D. L., Friedland, K. D., Anderson, J., Burns, P., Chu, K., Clay, P. M. *et al.* 2016. Northeast regional action plan – NOAA fisheries climate science strategy (NMFS-NE-239). Northeast Fisheries Science Center, Woods Hole, MA. 100.

He, X., and Field, J. C. 2019. Effects of recruitment variability and fishing history on estimation of stock-recruitment relationships: two case studies from U.S. west coast fisheries. Fisheries Research, 217: 21–34.

Houghton, R. W., Schlitz, R., Beardsley, R. C., Butman, B., and Chamberlin, J. L. 1982. The middle Atlantic Bight Cold Pool: evolution of the temperature structure during Summer 1979. Journal of Physical Oceanography, 12: 1019–1029.

Iles, T. C., and Beverton, R. J. H. 1998. Stock, recruitment and moderating processes in flatfish. Journal of Sea Research, 39: 41–55.

Johnson, D. R., and Boyer, T. P. 2014. East Asian seas regional climatology version 2.0 from 1804 to 2014 (NCEI accession 0123300). NOAA National Centers for Environmental Information, Asheville, NC. https://www.ncei.noaa.gov/archive/accession/0123300 (last accessed 22 February 2022).

Kavanaugh, M. T., Rheuban, J. E., Luis, K. M. A., and Doney, S. C. 2017. Thirty-three years of ocean benthic warming along the U.S. northeast continental shelf and slope: patterns, drivers, and ecological consequences. Journal of Geophysical Research: Oceans, 122: 9399–9414.

Kleisner, K. M., Fogarty, M. J., McGee, S., Hare, J. A., Moret, S., Perretti, C. T., and Saba, V. S. 2017. Marine species distribution shifts on the U.S. northeast continental shelf under continued ocean warming. Progress in Oceanography, 153: 24–36.

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., and Bell, B. M. 2016. TMB: automatic differentiation and laplace approximation. Journal of Statistical Software, 70.

Legault, C. M., and Restrepo, V. R. 1999. A flexible forward age-structured assessment program. ICCAT Collective Volumes of Scientific Papers. 49: 246–253.

Lellouche, J.-M., Greiner, E., Bourdallé-Badie, R. Garric, G. Melet, A. Drévillon, M. Bricaud, C. *et al.* 2021. The Copernicus Global 1/12° Oceanic and Sea Ice GLORYS12 Reanalysis. Frontiers in Earth Science, 9: 698876.

Lentz, S. J. 2017. Seasonal warming of the middle Atlantic Bight Cold Pool. Journal of Geophysical Research: Oceans, 122: 941–954.

Maunder, M. N., and Thorson, J. T. 2019. Modeling temporal variation in recruitment in fisheries stock assessment: a review of theory and practice. Fisheries Research, 217: 71–86.

Miller, T. J., Hare, J. A., and Alade, L. A. 2016. A state-space approach to incorporating environmental effects on recruitment in an age-structured assessment model with an application to Southern New England yellowtail flounder. Canadian Journal of Fisheries and Aquatic Sciences, 73: 1261–1270.

Miller, T. J., and Legault, C. M. 2017. Statistical behavior of retrospective patterns and their effects on estimation of stock and harvest status. Fisheries Research, 186: 109–120.

Miller, T. J., O'Brien, L., and Fratantoni, P. S. 2018. Temporal and environmental variation in growth and maturity and effects on management reference points of Georges Bank Atlantic cod. Canadian Journal of Fisheries and Aquatic Sciences, 75: 2159–2171.

Miller, T. J., and Stock, B. C. 2020. The Woods Hole Assessment Model (WHAM). Version 1.0.4. https://timjmiller.github.io/wham/ (last accessed date: 9 September 2021).

Mohn, R. 1999. The retrospective problem in sequential population analysis: an investigation using cod fishery and simulated data. ICES Journal of Marine Science, 56: 473–488.

Mountain, D. G. 2003. Variability in the properties of shelf water in the middle Atlantic Bight, 1977–1999. Journal of Geophysical Research, 108: 3014.

NEFSC. 2012. Workshop of the 54th Northeast regional stock assessment (54th SAW) assessment report, Northeast Fisheries Science Center Reference Documents 12–18. United States Department of Commerce, Washington, DC. 600 pp.

NEFSC. 2020. Operational assessment of 14 northeast groundfish stocks, updated through 2018. https://s3.amazonaws.com/nefmc.org/Prepublication-NE-Grndfsh-10-3-2019.pdf (last accessed 9 June 2021).

O'Leary, C. A., Miller, T. J., Thorson, J. T., and Nye, J. A. 2019. Understanding historical summer flounder (*Paralichthys dentatus*) abundance patterns through the incorporation of oceanography-dependent vital rates in Bayesian hierarchical models. Canadian Journal of Fisheries and Aquatic Sciences, 76: 1275–1294.

Perretti, C. T., Deroba, J. J., and Legault, C. M. 2020. Simulation testing methods for estimating misreported catch in a state-space stock assessment model. ICES Journal of Marine Science, 77: 911–920.

Pershing, A. J., Alexander, M. A., Hernandez, C. M., Kerr, L. A., Le Bris, A., Mills, K. E., Nye, J. A. *et al.* 2015. Slow adaptation in the face of rapid warming leads to collapse of the gulf of maine cod fishery. Science, 350: 809–812.

Punt, A. E. 2003. Evaluating the efficacy of managing west coast groundfish resources through simulations. Fishery Bulletin – National Oceanic and Atmospheric Administration, 101: 860–873.

Core Team, R. 2021. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/. (last accessed 29 June 2021)

Seidov, D., Baranova, O. K., Johnson, D. R., Boyer, T. P., Mishonov, A. V., and Parsons, A. R. 2016a. Northwest Atlantic regional climatology (NCEI accession 0155889). NOAA National Centers for Environmental Information, Silver Spring, MD. https://www.ncei.noaa.gov/archive/accession/0155889 (last accessed 25 March 2021).

Seidov, D., Baranova, O. K., Boyer, T., Cross, S. L., Mishonov, A. V., and Parsons, A. R. 2016b. Northwest Atlantic regional ocean climatology.: 3.2 MB. U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Environmental Satellite, Data, and Information Service. National Centers for Environmental Information, Asheville, NC.

Seidov, D., Baranova, O. K., Boyer, T. P., Cross, S. L., Mishonov, A. V., and Parsons, A. R. 2017. Northeast Pacific regional climatology (NCEI accession 0163799). NOAA National Centers for Environmental Information, Asheville, NC.

Sissenwine, M. P. 1974. Variability in recruitment and equilibrium catch of the Southern New England yellowtail flounder fishery. ICES Journal of Marine Science, 36: 15–26.

Stock, B. C., and Miller, T. J. 2021. The Woods Hole Assessment Model (WHAM): a general state-space assessment framework that incorporates time- and age-varying processes via random effects and links to environmental covariates. Fisheries Research, 240: 105967.

Sullivan, M., Cowen, R., Able, K., and Fahay, M. 2000. Spatial scaling of recruitment in four continental shelf fishes. Marine Ecology Progress Series, 207: 141–154.

Sullivan, M. C., Cowen, R. K., and Steves, B. P. 2005. Evidence for atmosphere-ocean forcing of yellowtail flounder (*Limanda ferruginea*) recruitment in the middle Atlantic Bight. Fisheries Oceanography, 14: 386–399.

Szuwalski, C. S., and Punt, A. E. 2013. Fisheries management for regime-based ecosystems: a management strategy evaluation for the snow crab fishery in the eastern Bering Sea. ICES Journal of Marine Science, 70: 955–967.

Szuwalski, C. S., Vert-Pre, K. A., Punt, A. E., Branch, T. A., and Hilborn, R. 2015. Examining common assumptions about recruitment: a meta-analysis of recruitment dynamics for worldwide marine fisheries. Fish and Fisheries, 16: 633–648.

Tableau, A., Collie, J. S., Bell, R. J., and Minto, C. 2019. Decadal changes in the productivity of New England fish populations. Canadian Journal of Fisheries and Aquatic Sciences, 76: 1528–1540.

Tanaka, K. R., Cao, J., Shank, B. V., Truesdell, S. B., Mazur, M. D., Xu, L., and Chen, Y. 2019. A model-based approach to incorporate environmental variability into assessment of a commercial fishery: a case study with the American lobster fishery in the Gulf of Maine and Georges Bank. ICES Journal of Marine Science, 76: 884–896.

Vert-pre, K. A., Amoroso, R. O., Jensen, O. P., and Hilborn, R. 2013. Frequency and intensity of productivity regime shifts in marine fish stocks. Proceedings of the National Academy of Sciences, 110: 1779–1784.

Xu, H., Miller, T. J., Hameed, S., Alade, L. A., and Nye, J. A. 2018. Evaluating the utility of the gulf stream index for predicting recruitment of Southern New England-Mid Atlantic yellowtail flounder. Fisheries Oceanography, 27: 85–95.

Handling Editor: Manuel Hidalgo