Consistent Predictability of the Ocean State Ocean Model (OSOM) using Information Theory and Flushing Timescales

Aakash Sane[1], Baylor Fox-Kemper[1], Dave Ullman[2], Christopher Kincaid[2], and Lewis Rothstein[2]

[1]Brown University
[2]University of Rhode Island

November 23, 2022

Abstract

The Ocean State Ocean Model OSOM is an application of the Regional Ocean Modeling System spanning the Rhode Island waterways, including Narragansett Bay, Mt. Hope Bay, larger rivers, and the Block Island Shelf circulation from Long Island to Nantucket. This paper discusses the physical aspects of the estuary (Narragansett and Mount Hope Bays and larger rivers) to evaluate physical circulation predictability. This estimate is intended to help decide if a forecast and prediction system is warranted, to prepare for coupling with biogeochemistry and fisheries models with widely disparate timescales, and to find the spin-up time needed to establish the climatological circulation of the region. Perturbed initial condition ensemble simulations are combined with metrics from information theory to quantify the predictability of the OSOM forecast system–i.e., how long anomalies from different initial conditions persist. The predictability timescale in this model agrees with readily estimable timescales such as the freshwater flushing timescale evaluated using the total exchange flow (TEF) framework, indicating that the estuarine dynamics rather than chaotic transport is the dominant model behavior limiting predictions. The predictability of the OSOM is ˜ 7 to 40 days, varying with parameters, region, and season.

Consistent Predictability of the Ocean State Ocean Model (OSOM) using Information Theory and Flushing Timescales

Aakash Sane[1], Baylor Fox-Kemper[2], Dave Ullman[3], Christopher Kincaid[3], and Lewis Rothstein[3]

[1]School of Engineering, Brown University, Providence, RI
[2]Dept. of Earth, Environmental, and Planetary Sciences (DEEPS), Brown University, Providence, RI
[3]Graduate School of Oceanography, University of Rhode Island, Narragansett, RI

Key Points:

- This paper introduces the ROMS-OSOM, a Regional Ocean Modeling System (ROMS) implementation simulating Rhode Island waterways called the Ocean State Ocean Model (OSOM).
- The predictability of the OSOM is evaluated using information theory and initial condition ensembles in summer and winter conditions.
- The flushing time scale (freshwater and salinity) of Narragansett and Mt. Hope Bays are calculated and resemble the predictability timescales, indicating that predictability is largely governed by the estuarine circulation in this model.

Corresponding author: Aakash Sane, `aakash_sane@brown.edu`

Abstract

The Ocean State Ocean Model OSOM is an application of the Regional Ocean Modeling System spanning the Rhode Island waterways, including Narragansett Bay, Mt. Hope Bay, larger rivers, and the Block Island Shelf circulation from Long Island to Nantucket. This paper discusses the physical aspects of the estuary (Narragansett and Mount Hope Bays and larger rivers) to evaluate physical circulation predictability. This estimate is intended to help decide if a forecast and prediction system is warranted, to prepare for coupling with biogeochemistry and fisheries models with widely disparate timescales, and to find the spin-up time needed to establish the climatological circulation of the region. Perturbed initial condition ensemble simulations are combined with metrics from information theory to quantify the predictability of the OSOM forecast system–i.e., how long anomalies from different initial conditions persist. The predictability timescale in this model agrees with readily estimable timescales such as the freshwater flushing timescale evaluated using the total exchange flow (TEF) framework, indicating that the estuarine dynamics rather than chaotic transport is the dominant model behavior limiting predictions. The predictability of the OSOM is $\sim$7 to 40 days, varying with parameters, region, and season.

Plain Language Summary

A new model of waterways near Rhode Island is introduced and examined. The model is intended for studying the physical circulation of this region and its ecosystem changes. This study uses a variety of metrics to assess for how long a forecast with this model might be useful (i.e., how long the model's initial state determines its behavior) and relatedly how long to run (or spin up) the model to have poorly known initial conditions not affect the result systematically.

1 Introduction

Coastal marine forecast systems are in use or development in a number of regions worldwide (e.g. Wilkin et al., 2018; Moore et al., 2011; Lellouche et al., 2018; Pinardi & Coppini, 2010; Mel & Lionello, 2014; Raboudi et al., 2019). As each region is unique, the length of forecast window and relative levels of forced to internal variability differ among these systems. The Ocean State Ocean Model (OSOM) is a new model in development, which is an extension and synthesis of past prototype models (Bergondo, 2004; Bergondo & Kincaid, 2007; Liu et al., 2016; Wertman, 2018; Ullman, 2019; McManus et al., 2020) being evaluated for potential use as a forecast system. In this evaluation, key questions are: How often should a forecast be made? How far into the future can forecasts be skillful? How long does the model take to spin up? How accurate must surface and boundary forcing be to arrive at useful forecasts, given that these data would also be predictions (e.g., from numerical weather prediction models)? Which regional societal challenges are better framed as changes to the region's climatology (i.e., projections) rather than as predictable futures that depend on the model's initial conditions (i.e., forecasts)? In this paper, a framework for addressing these questions is developed by adapting methods from information theory and ensemble-based measures of predictability, internal variability, and forced variability. The OSOM is taken as a test example of these methods and, as a coastal model in development with unique characteristics, the specific results of this study are useful for the future development of this particular model.

Forecasting hydrodynamic parameters is pertinent for an estuary as they play a vital role in controlling the physical as well as biogeochemical changes. An important aspect of forecasting is finding the predictability/forecasting timescales that limit the degree to which initial conditions govern the future behavior of the numerical model for individual parameters. These timescales quantify the persistence of anomalies and are a feature of the numerical model. Predictability is a measure of a model's ability to forecast or predict the evolution of anomalies in the future from initial conditions given prescribed external forcing. By contrast, changing forcing due to climate change (e.g., Xiu et al., 2018), altered topography via erosion or dredging (Hayward et al., 2018), changes to wastewater treatment or power plant effluent (Mustard et al., 1999), etc., are *external* factors affecting boundary conditions rather than initial conditions whose impact can be assessed using *projections* of future climatology with altered boundary conditions over a variety of plausible initial conditions. Thus, *predictability* measures a model's potential to *predict* or *forecast* a future state which is distinct from climatology, which is distinct from *projecting* the changes to climatology forced from changes to boundary conditions. The state of the system in a forecast can be only considered in a probabilistic way and hence predictability is a property involving two distributions (DelSole, 2004): predictability quantifies the departure of a forecast distribution from the climatology distribution (Shukla, 1981; Leung & North, 1990). Quantifying this departure involves measurement of uncertainty in the forecast signal. The uncertainties in the initial conditions can be thought of as anomalies which eventually are forgotten by the model, or overwhelmed by chaotic variability or the influence of boundary conditions as time proceeds until the forecast statistical distribution becomes indistinguishable from the climatology distribution. Beyond this time scale a forecast provides no additional information beyond climatology, and forecasts are then no more useful than projections of the future climatological range of possibilities.

This article has three purposes: (1) To describe the OSOM; (2) To use ensemble simulations to find predictability timescales; (3) To find estuarine flushing timescales for fresh and saline water masses and compare these to (2). The model is forced by winds, tides, river runoff, evaporation, precipitation and also forced by heat fluxes and open boundary conditions. So, unlike the numerical weather prediction models for which the information theory techniques applied here were developed, the OSOM is a forced model where much of the variability comes from external forcing that may determine the trend of the evolution of the state parameters, or alternatively internal variability (e.g., hydrodynamic instabilities and chaos) may dominate. A companion paper by the authors to this one develops a non-parametric information theory approach to quantifying the amount of internal vs. forced variability similar to the ensemble approach of (Llovel et al., 2018), and uses this metric to quantify the relative importance of different choices in boundary forcing. As the balance of sources of variability depends on forcing, resolution, classes of flow, etc., the measured forced vs. intrinsic variability depends on the specifics of the model, rather than being a general description of the waterways under study. So, too, do the predictability metrics describe the specific model being studied rather than the system. However here a comparison to traditional estuarine flushing timescales serves to illustrate that the model is governed by physical principles, so quantifying these based on the real–rather than simulated–world may nonetheless be useful in establishing physical guidelines underlying limits on predictability. Metrics from information theory provide a natural way of quantifying distances between two probability distributions (Cover & Thomas, 2012). Information theory metrics have been used in myriad ways in other fields (e.g., electronic communications, image processing, and molecular biology). Using information theory metrics for weather prediction and climate projection is well established (Leung & North, 1990; Schneider & Griffies, 1999; Roulston & Smith, 2002; Kleeman, 2002; DelSole, 2004; Haven et al., 2005), but they are not commonly used in coastal modeling. DelSole (2004) relates the requirement to quantify uncertainty with the usage of metrics from information theory. The most commonly used metrics are entropy, relative entropy, and mutual information (Shannon, 1948), although other variants are also useful (Kleeman, 2002; Leung & North, 1990). A key advantage for use of these metrics in coastal modeling is that they can be ascribed to a variety of physical or biogeochemical variables; here we examine salinity, temperature, and kinetic energy over regions and at observation locations, but in future work we will examine biogeochemical variables in the OSOM.

An important time scale for an estuary is the flushing time scale or residence time scale (Knudsen, 1900), which is defined as the average residence time of a parcel of fluid inside the estuary (e.g., Monsen et al., 2002), and thus also the average retention time of water masses in the estuary. As the numerical model represents the physical domain, there is an inherent relation between the forecasting timescales and the flushing time scale, because eventually tracer anomalies present in the initial conditions will be flushed from the estuary, and the flushing timescale is an estimate of how long this process will take (assuming the anomalies are conserved on each water parcel). Here these timescales are found for the OSOM, a model developed specifically for Narragansett Bay and connected waterways.

Narragansett Bay (NB) is a medium-sized estuary and a natural harbor. As per the classification of estuaries based on physical and hydrological attributes, NB is a class 8 estuary (a moderate area, volume, and freshwater flow estuary that is deep and salty: Engle et al., 2007). It is a prime example of a coastal plain estuary, also known as a drowned river valley, which is the most common type of estuary in temperate climates. The bay covers an area of ~ 400 km^2 (Pilson, 1985). It is 16 km wide (East-West), 32 km long (North-South), and has 412 km of shoreline. The Bay extends from the Providence and Seekonk rivers in the north to Rhode Island Sound in the South. To the east, it connects to Mount Hope Bay, fed by the Tauton River and connected by the Sakonnet River to Rhode Island Sound. The whole of the Narragansett Bay, Mount Hope Bay, associated rivers, and Rhode Island Sound is simulated in OSOM (Figure 1), but the emphasis in this paper is variables within NB and Mount Hope Bay. The average depth is 8 m and the deepest point is 60 m. The bathymetry varies with steep slopes in the Rhode Island Sound towards the open ocean and along the dredged navigation channels. The Bay provides a natural habitat for many living things and is of commercial and ecological importance to the local community. Commercial fishing and shell fishing are important economic activities and the Bay has also been used for recreational sports such as a harbor for the America's Cup and the Volvo Ocean Race sailing competitions. Recently pollution has prevented these activities; bacteria and viruses have caused beach closures, harmful blooms, and shell fishing bans, and hypoxia is frequent and sometimes induces large fish kills. OSOM will be used to simulate the physics of the Bay and predict the physical and biogeochemical conditions conducive to these events, as well as assess the impact of different management and mitigation practices. The predictability timescales studied here help reveal the utility of the model to forecast the physical conditions for harmful events.

This article has been structured as follows: Section 2 provides detail of the computational model OSOM. Section 3 describes the theory of using mutual information to find predictability timescales. Section 4 contains the ensemble simulation setup for forecasting and climatology sets. Application of mutual information to the ensembles has also been described in Section 4. Section 5 states the results for various cases and also gives the flushing timescales obtained via OSOM.

2 Ocean State Ocean Model

The Ocean State Ocean Model (OSOM) is an application of the Regional Oceanic Modelling System - ROMS (Shchepetkin & McWilliams, 2005). The curvilinear terrain-following coordinate system employed in ROMS is well suited for coastal applications since the bathymetric variations in coastal systems and estuaries are large. The model has curvilinear varying horizontal resolution as well, from $\sim$ 50m towards the North to around 200m in the south of the modelled domain. The horizontal grid consists of 1000 $\times$ 1100 grid cells and 15 terrain-following sigma levels in the vertical. The Generic Length Scale (GLS) scheme is used to represent unresolved turbulence (Umlauf & Burchard, 2003).

The offshore forcing at the open boundaries is provided by surface elevation and depth-averaged velocity using 9 tidal constituents (M2, S2, N2, K2, K1, O1, Q1, M4, M6) from the Eastcoast tidal constituent database (Mukai et al., 2002) and, at subtidal timescales, with low-pass filtered output of the hindcast version of the Northeast

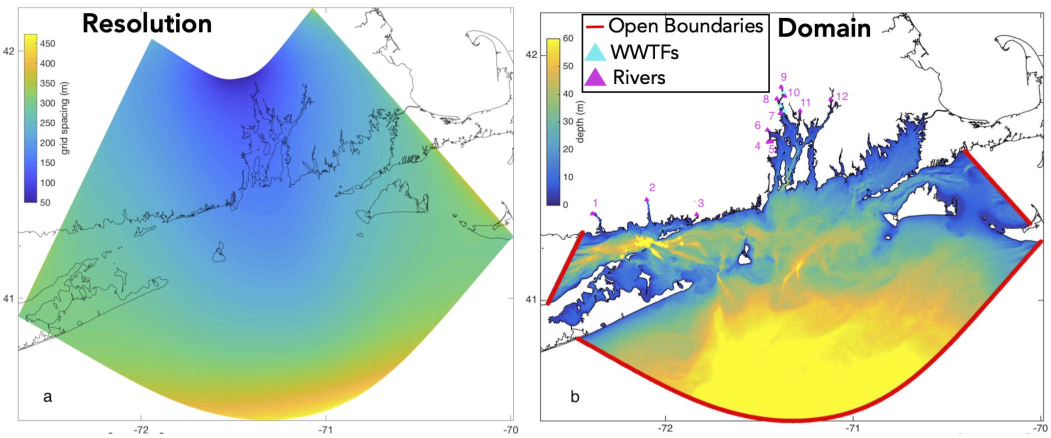

Figure 1. a. ROMS OSOM horizontal grid resolution, which is the geometric average of that in the ζ direction ($\sim$ East - West direction) and in the η direction ($\sim$ North - South direction). The finest resolution is at North where Narragansett Bay is. Resolution decreases towards the open ocean. b. Bathymetry: The Narragansett Bay and Mount Hope Bay are regions of shallow bathymetry and depth increases across the Rhode Island Sound toward open ocean. Wastewater Treatment Facilities (WWTFs) are shown in blue. Important rivers are highlighted in magenta: 1. Connecticut River, 2. Thames River, 3. Pawcatuck River, 4. Maskerchugg River, 5. Hunt River, 6. Hardig Brook, 7. Pawtuxet River, 8. Woonasquatucket and Moshassuck River, 9. Blackstone River, 10 ten Mile River, 11. Palmer River, 12. Taunton River.

Coastal Ocean Forecast System (NECOFS), a regional model covering the northeast U. S. coastal ocean (Beardsley & Chen, 2014). The surface elevation and depth- averaged velocity forcing are implemented using the Chapman (1985) and Flather (1976) methodologies respectively. The depth-dependent velocity, temperature, and salinity at the open boundaries are forced using the Marchesiello et al. (2001) combined radiation and nudging open boundary condition using low-pass filtered NECOFS output. The nudging timescales vary with stronger nudging on inflow (timescale of 1.6h) than on outflow (timescale of 24h).

Surface heat and momentum fluxes are estimated from meteorological variables obtained from models and local observations using the updated COARE bulk formulae (Fairall et al., 2003). All meteorological forcing except for winds are assumed to be spatially uniform over the model domain. Spatially variable winds for the region were obtained from the North American Mesoscale (NAM) analyses, a data-assimilating, high resolution (12 km) meteorological simulation (`https://www.ncei.noaa.gov/data/north-american-mesoscale-model/access/historical/analysis`). Air temperature and barometric pressure were estimated by averaging the measurements at the six stations of the Narragansett Bay PORTS system (`http://www.co-ops.nos.noaa.gov/ports.html`). Precipitation and relative humidity are from observations at T. F. Green Airport, in Warwick, RI. Net shortwave and downward longwave radiative fluxes were taken from the nearest ocean gridpoint of NOAAs North American Regional Reanalysis model (`http://www.emc.ncep.noaa.gov/mmb/rreanl`). Upward longwave radiation was computed based on the ocean surface temperature in the model simulations.

Freshwater discharge from local rivers and the major waste water treatment facilities (WWTF) discharging into NB were applied as point source inflows. The discharges of many of the rivers are measured at United States Geological Survey (USGS) gauging stations (Hunt, Palmer, Moshassuck, Woonasquatucket, Blackstone, Ten Mile, Pawtuxet, Taunton, Pawcatuck, Connecticut, Quinebaug, Yantic, and Shetucket Rivers). The Moshassuck and Woonasquatucket Rivers, which discharge into the upper Providence River, were combined in the model. Likewise the gauged discharges of the Quinebaug, Yantic, and Shetucket Rivers were combined to form the model Thames River. For the small rivers entering Greenwich Bay (Maskerchugg River and Hardig Brook) which are presently not gauged, historical flow measurements were used with simultaneous measurements from the nearby Hunt River to develop a linear regression model predicting the discharge of the former from gauged measurements from the latter river. The gauging stations varied in their proximity to the locations at which the rivers discharge into the model domain. In order to account for the river discharge from the portion of the watershed downstream of the gauging station, the measured discharges were scaled up using estimates of the drainage areas upstream and downstream of the gauge under the assumption that discharge/drainage area downstream is equal to its value upstream of the gauge. Discharges from four WWTFs (Fields Point, Bucklin Point, East Providence, and East Greenwich) in the upper/mid Bay region were obtained from the plant operators.

The WWTF point sources were implemented at a single ROMS gridpoint but the discharges for the rivers are spread over 2–5 gridpoints to reduce the tendency for model instability. River forcing in ROMS requires, in addition to the river discharge discussed above, specification of the vertical profile of the river inflow transport and the concentration of tracers in the inflowing water. The vertical profile of the river inflow was specified as linearly varying with zero transport at the bottom. Salinity of the inflowing water was set to 0. In the simulations discussed here, the river water temperature was also set to 0 which eventually leads to artificially cold rivers, but experimentation versus using more realistic temperatures reveals modestly lower temperatures at the observation sites in the Bay over the integration times used (especially in winter). Setting river temperature to 0 only affected the temperatures in zone 1 and 5 for winter where rivers have more influence (Figure 4 illustrates zone boundaries). The cold bias found was about 4-6 K in zone 1 and 1-2 K in zone 5. The temperature at the grid points closest to buoys were not affected as all the observation locations shown in Figure 2 are sufficiently away from river sources. However, it is recommended for future operational simulations that time varying river water temperature be estimated using a regression equation involving air temperature as well as water temperature on the previous day.

2.1 Basic model validation

The model output has been compared with buoy data obtained from the Rhode Island Data Discovery Center (`http://ridatadiscoverycenter.org`), where a variety of regional data are accessible. In particular, the model has been compared with moored observations collected at locations shown in Figure 2. Figure 3 illustrates the best and worst matches for temperature and salinity of the model with the historical observations. Comparison of the model versus surface temperatures derived from LandSat also confirms that the patterns of heating and cooling are similar to the satellite data, although seasonality in OSOM is somewhat larger than in the satellite record (by roughly 1°C in climatological comparisons).

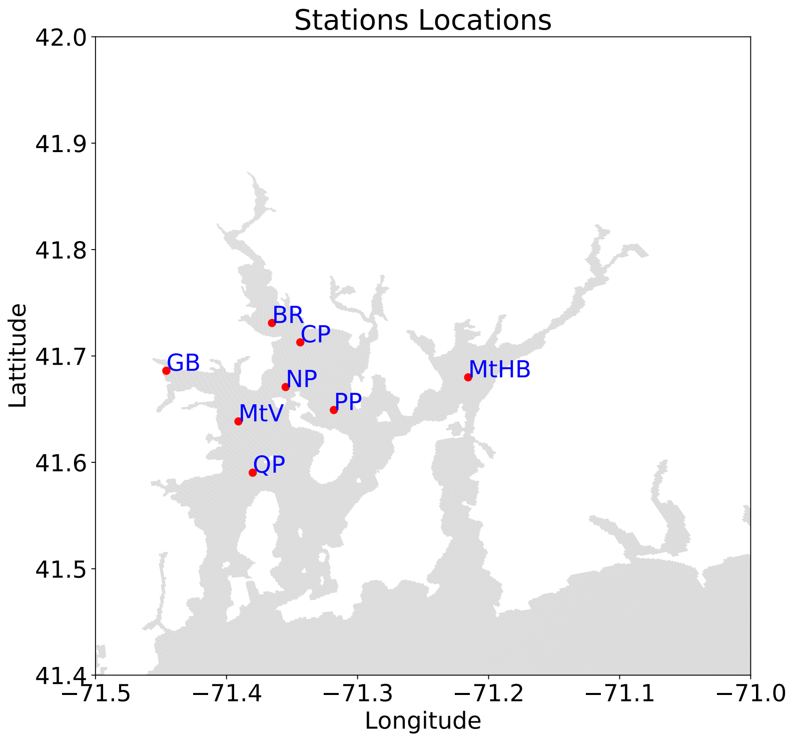

Figure 2. Stations where surface as well as bottom temperature and salinity observations are continuously collected during the months of July-August of 2006: Greenwich Bay (GB), Bullock's Reach (BR), Conimicut Point (CP), North Passage (NP), Mount Hope Bay (MtHB), Poppasquash Point (PP), Mount View (MtV), and Quonset Point (QP). Model data is compared with observations from these stations.

Figure 3 indicates that the model has skill at the high frequency variability (tides and diurnal cycle), although variability at the bottom level is underestimated. The lower frequency temperature and salinity have biases of up to 2K at the surface and

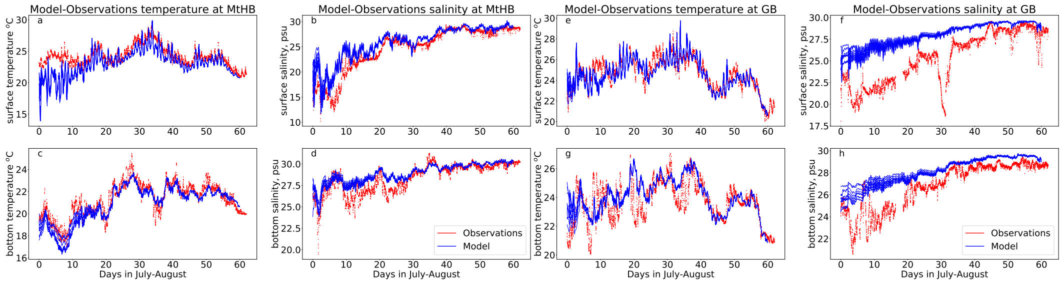

Figure 3. (a-d) Comparison of the Mount Hope Bay moored buoy observations of salinity and temperature at the surface (a, b) and maximum depth (c, d). This case is the closest match of the OSOM to the observations during the two months shown: July and August of 2006. (e, f, g, h) Comparison of the Greenwich Bay moored buoy observations of salinity and temperature at the surface (e, f) and maximum depth (g, h). This case is the poorest match of the OSOM to the observations during the two months shown: July and August of 2006. Red color represents the observed values and different colors show different ensemble members. Figures S1 to S6 in the supporting information compare the rest of the marked observation locations.

1K at the bottom, and 3 and 2 psu at the surface and bottom of MtHB. At GB, the errors at surface and bottom are up to 5 K and 6 psu and 2 K and 4 psu respectively.

The emphasis of this paper is on measuring the basic predictability of the OSOM as modeled in this version. It is not necessary for this assessment for the OSOM to be completely realistic, but these basic comparisons show that it has skill in reproducing realistic variability in temperature and salinity. Future work will address improvements in the model setup to reduce biases and errors, such as improving the assumed temperature of river inflows, parameterizations of mixing, evaluation of tides, different products for surface and offshore boundary conditions, etc.

3 Predictability using information theory

DelSole & Tippett (2007) state that the two guiding principles for measuring predictability of a variable by contrasting the forecast and a climatology distribution should be 1) separate, non-identical measures for a given prediction, and 2) the measure of predictability should be invariant to linear transformation (Schneider & Griffies, 1999; Majda et al., 2002). Measures of predictability using information theory are naturally invariant to linear transformations and will be explained in general in the following paragraphs.

Consider a signal, such as a variable or regional average of a variable modeled by the OSOM, X, having a probability distribution $p_i(x)$ when considered over a particular time or space interval. The probability distribution $p_i(x)$ is of the i^{th} event (i^{th} bin) after dividing the data into N bins. A fundamental quantity in information theory is the Shannon entropy (Shannon, 1948) defined by

$$H(X) = \sum_{i=1}^{N} p(x_i)\ log_2 \left(\frac{1}{p(x_i)} \right). \tag{1}$$

The entropy (with base 2 logarithm) is quantified in units of *bits*, because the Shannon entropy effectively measures the average amount of digital storage required to capture the information present in the variability of X.

To understand Equation 1 begin with the innermost term. Hartley (1928) first proposed using the logarithmic function $log_2(1/p(x_i))$ to quantify information or uncertainity in an event having probability $p(x_i)$. The formulation $log_2(1/p(x_i))$ implies that low probability events have higher uncertainty. Shannon (1948) completed this measure by additionally weighting the logarithm with probability giving rise to the *entropy* definition Equation 1, which resembles the thermodynamic entropy function in statistical mechanics resulting from a system that visits a set of equally probably states (e.g. Sethna et al., 2006). Shannon's entropy is formulated so that high probability events reduce uncertainty with a strong weighting because they occur often (Cover & Thomas, 2012). Shannon entropy quantifies uncertainty and the number of states needed to categorize a single probability distribution.

To compare two distributions $p(x)$ and $p(y)$ relative entropy and mutual information measures are useful comparative metrics. Kleeman (2002) recommends the relative entropy (a.k.a., Kullback-Leibler distance Cover & Thomas, 2012) for climate modelling, which is $R = \sum\limits_{i=1}^{N} p(x_i) \log_2 \frac{p(x_i)}{p(y_i)}$. Here, let X be the forecast and Y be the climatology. Recall that predictability measures the information contained in a particular forecast that is not present in the climatology, i.e., the information which stems from the forecast initial conditions. It is easy to see that if the forecast probability $p(x_i)$ equals the climatology forecast $p(y_i)$, R goes to zero indicating no distance or difference in information between the forecast and climatology. As a forecast evolves, during the time interval before R reaches zero, $p(x)$ and $p(y)$ are distinguishable (under similar levels of unpredictable noise) and after R reaches zero they are not, thus this time interval is the predictability window.

Within the predictability window, interchanging $p(x_i)$ and $p(y_i)$ changes the value of R, not just by sign from the logarithm, but also by magnitude due to the prefactor $p(x)$. Thus, the relative entropy R depends on both $p(x)$ and $p(y)$ asymmetrically and will change if they are interchanged (i.e., the metric depends on which variable is considered the climatology and which is considered the forecast). Our potential predictability will compare different ensemble members where one is taken as forecast member, and from same ensemble a different member is taken as a climatology reference (Kumar et al., 2014). As the different ensemble members should be interchangeable in this approach, the magnitude of our metric (in contrast to R) should not change by interchanging the forecast and climatology, hence a different metric is preferred: mutual information.

Mutual information, $I(X;Y)$, is symmetric in X and Y, and hence is a natural metric of distance between these variables without direction. Let two random variables X and Y have joint probability $p(x_i, y_j)$ and marginal probability $p(x_i)$ and $p(y_j)$. X and Y are divided into N bins each (they can also be divided into different bins but we have used the same number of bins for simplicity). The mutual information $I(X;Y)$ between them is (Cover & Thomas, 2012)

$$I(X;Y) = \sum_{i=1}^{N}\sum_{j=1}^{N} p(x_i, y_j)\log_2 \frac{p(x_i, y_j)}{p(x_i)p(y_j)}, \tag{2}$$

Mutual information resembles relative entropy. In fact, it measures the relative entropy between the joint distribution $p(x_i, y_j)$ and the product of the marginal distributions $(p(x_i)p(y_j))$. If X and Y are independent variables, then $p(x_i, y_j) = p(x_i)p(y_j)$ and thus $I(X;Y) = 0$. However, if they are *not independent*, so that one contains information about the other, then there is mutual information shared and $I(X,Y) > 0$. If they are totally dependent, i.e., knowing the value of X reveals the value of Y and *vice versa*, then $p(x_i, y_j) = p(x_i) = p(y_j)$ for each value of i, j and the mutual information equals the Shannon entropy: $I(X,Y) = H(X) = H(Y)$. Thus, mutual information is the metric of the information shared by X and Y versus if they were independent variables. Mutual information between X and Y is symmetric and measures a distance between the two probability distributions. It quantifies the amount of information one variable contains about the other (again in bits). It can also measure the reduction in uncertainty of one distribution given knowledge of a second distribution, or the degree to which they are not independent (Cover & Thomas, 2012): $I(X;Y)$ measures the degree of statistical constraint of X on Y and *vice versa* (Fano, 1961). Mutual information is easily extended to more than one variable leading to a multivariate predictibilty analysis (DelSole & Shukla, 2010).

Unlike relative entropy R, mutual information $I(X;Y)$ does not go to zero when $p(x)$ approaches $p(y)$, instead it approaches the Shannon entropy $H(X)$ from Eq. 1. We use the property that $I(X;Y)$ approaches $H(X)$ to delimit the predictability window, taken as when the probability distribution of the forecast and the climatology become effectively indistinguishable, taken to be the first time when $I(X;Y)$ reaches within 90% of $H(X)$. This threshold is somewhat arbitrary, as convergence is not typically monotonic or complete, so any threshold will tend to have "near misses" and later signs of potential predictability as will be illustrated in a variety of figures in the text and supplementary material. However, to compare to the flushing timescales in later sections, a threshold is a simple test, and a range of predictability timescales is then formulated by comparing to individual climatology ensemble members as well as the climatology ensemble mean to appropriately gauge the level of certainty.

DelSole & Shukla (2010) state that mutual information itself is a measure of forecast skill and provide skill scores founded on mutual information and relative entropy. The metrics in Equations 1-2 are based on the probabilities of events, not the units or dimensions of the events, so their use on various parameters and between forecasts and climatology can be compared regardless of the type of variable: physical variables, biological variables, chemical variables, or sociological variables of arbitrary units can be compared. For this reason, these information theory metrics are ideal for evaluating forecast skill in a model like OSOM where a variety of applications are intended. The metrics are also invariant under linear transformation of the signal and hence are robust to trivial changes such as changes of the units of measurement (DelSole & Tippett (2007)), unlike alternatives such as the root mean square technique for skill assessment (for example, Jin et al., 2018) which require normalization.

To find the predictability time scales of ROMS-OSOM we will compare ensembles members which differ in initial conditions. Hence our focus is on finding the potential predictability (model-model comparison) instead of actual predictability or model forecast skill (model - observation comparison, for example, Kumar et al., 2014). The climatology comes from the model simulations and is a result of past or historical forcings (hindcasts) with unperturbed initial conditions. It will be compared to forecasts with an anomaly of perturbed initial conditions that will eventually decay or be flushed out. The time it takes for the forecast to approach the climatology is the predictability time scale. In other words, the convergence between forecast member and climatology member signals the end of the predictability time period. After this period running the forecast is of no utility, and it will statistically resemble any climatological estimate without predictable consequences remaining from its initial anomaly. This decay occurs because even though an anomaly is introduced, the forcings and boundary conditions are identical between the climatology and the forecast. In a realistic forecast, the model would be initialized with observations and run with historical external forcings as future external forcings are unknown a priori. The initialization due to observations would create anomalies which are similar to perturbations we add to initial conditions in hindcasts to find potential predictability. Also, in a realistic forecast, the forecast signal will begin to diverge away from future observations and converge towards the model climatology signal–another sign marking the predictability time scale.

4 Ensemble setup

To begin, temperature and salinity were interpolated from hindcasts of the FVCOM model (Beardsley & Chen, 2014) and velocities were taken to be zero. From these conditions, the model was spun up for two months before analysis begins. Two months were estimated to be sufficient as the average flushing time in NB is about one month (Pilson, 1985), and post-analysis estimates of the predictability timescale confirm this conjecture. The initial conditions used for ensemble simulations were derived from one single spun-up simulation for each season taken from the boundary conditions for the year 2006. Simulations were performed in each of two seasons: January-February (JF) and July-August (JA). The months JA were chosen because NB faces hypoxia during those months (Codiga et al. (2009)), and JF was chosen as a contrasting alternative. For each season (JF, JA) there is a set of climatology ensemble members that were simulated consisting of 7 and 10 members respectively. The JF and JA climatology ensemble has two sets of corresponding forecast ensembles: one initialized by perturbing only temperature, and the other set initialized by perturbing only salinity.

Each climatology ensemble member is forced in the same way, but each has realistic initial conditions chosen from consecutive days selected from the spin-up run before the simulation start day (Smith et al., 2007). This method of building a climatology ensemble is perhaps unfamiliar to some readers, and differs from the typical average across multiple years of simulations (where the climatology is across varying forcing, rather than varying initial conditions). To create a larger contrast, the same initial conditions were perturbed by tripling the anomaly of each climatology ensemble member from the climatology ensemble mean. This second ensemble of enhanced initial conditions are called the "forecast ensemble", and the same number of members are in the forecast and climatology ensembles (7 in JF and 10 in JA). The forecast ensemble members by design have bigger spread in their initial conditions than the climatology ensemble. As each ensemble contains both forced and internal variability, it was not sufficient to have only one forecast represent the "climatology", but rather a mean over an ensemble of realistic initial conditions serves as a better reference climatology. Furthermore, it is potentially undesirable to compare a single climatology run versus an ensemble mean of forecasts–care is needed to compare ensemble means versus ensemble means (the approach here) and individual simulations versus individual simulations. However, comparing the individual models within the ensembles is used to formulate a range of possible predictability timescales, and comparing individual members with other individual members yields similar results to the ensemble versus ensemble comparison method used primarily here.

Model data is saved in 2 hour window time averages. The granularity is needed to capture the strong tidal variability in this region. Thus each day has 12 data points for all the variables and for all the ensemble members. Predictability analysis is performed for 3 types of data: 1) Timeseries of volume-weighted averages of variables (temperature, salinity) over the 7 zones shown in Figure 4, 2) Predictability of kinetic energy using spatial data over 7 zones, and 3) Predictability of timeseries for a grid point closest to a moored observation. Thus, the effects of predictability on different variables or different levels of averaging is illustrated.

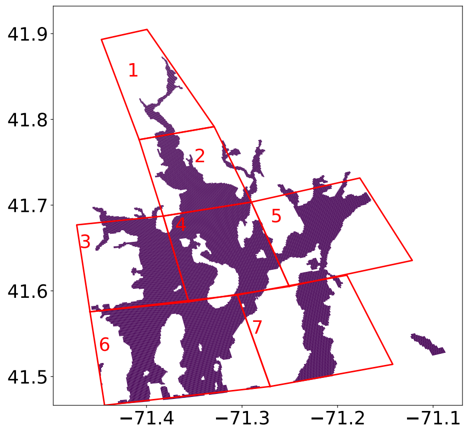

Figure 4. Narragansett Bay has been divided into 7 zones. Volume weighted temperature and salinity has been used from each zone to find predictability timescales.

The number of ensemble members is justified by deciphering whether external forcing (wind, tidal, river runoff, and evaporation/precipitation) or internal chaos (non-linearities, eddies) is setting the trend for evolution of state parameters in the ensemble mean. The methodology of Llovel et al. (2018) and Leroux et al. (2018) is used as a guide. The ratio of "noise" to signal with respect to time was found, where noise is taken as the standard deviation of the model spread and signal is the mean over the ensemble. Let σ be the standard deviation of ϕ_i^n, which is also same as the model spread. The ratio $\sigma_i/\langle\phi\rangle_i$ remains less than 0.5 within the predictability window and below 0.1 after crossing predictability time scale. Llovel et al. (2018) state that a noise to signal ratio of less than 0.5 is sufficient so that external forcing is dominant in setting the ensemble mean variability over internal chaos, indicating also that model trend is captured sufficiently with this number of ensemble members. The upcoming companion paper by the authors expands on the approach of Llovel et al. (2018) using information theory techniques to quantify forced versus internal variability even for non-Gaussian and non-independent datasets.

Let a variable in the climatology ensemble be given by $c_{t,i}^n$ where t denotes time, i denotes spatial grid-point, and n is the ensemble member. Similarly, a variable in the forecast ensemble is $f_{t,i}^n$. The information entropy metrics have been calculated between forecast and climatology using two approaches: 1) Between running time windows (probability distributions of variability in t) of spatial volume weighted averaged data (i.e., averaged over i) in a zone or at an observation location, and 2) examining the covariability of spatial grid points (probability distributions based on i) within a zone at a fixed time. The advantage of the former is that it more naturally describes the evolution of slow variations over large regions of the Bay, while the latter can be used for very rapid convergence of variables with shorter predictability timescales.

The first running window approach is primarily used for evaluating predictability of temperature and salinity. First, data is averaged (volume weighted) over each zone. Hence, $\Sigma_i \left[c^n_{t,i}\mathrm{d}V_i\right] / (\Sigma_j \mathrm{d}V_j) = \overline{c}^n_t$ and $\Sigma_i \left[f^n_{t,i}\mathrm{d}V_i\right] / (\Sigma_j \mathrm{d}V_j) = \overline{f}^n_t$ with the over-bar representing volume weighted average over a zone ($\mathrm{d}V_i$ is the volume associated with each gridpoint). Next, the ensemble mean of all climatology members was found, given by $\langle\overline{c}\rangle_t = (1/N)\sum_{n=1}^{N} \overline{c}^n_t$ where the angle brackets represent ensemble average. A running window of size τ is selected and a histogram of values is used to estimate the probability distributions of the climatology and forecasts, from which $I(f;c)^n_t$ is calculated over the time interval with climatology spanned by end members $(\langle\overline{c}\rangle_t, \langle\overline{c}\rangle_{t+\tau})$ and forecast variability $\left(\overline{f}^n_t, \overline{f}^n_{t+\tau}\right)$ according to Equation 2. Shannon entropy $H(c)^n_t$ is also calculated from these histograms for $(\langle\overline{c}\rangle_t, \langle\overline{c}\rangle_{t+\tau})$ according to Equation 1. The predictability time is taken to be when the mutual information averaged over the forecast ensemble $I(f;c)_t = (1/N)\sum_{n=1}^{N} I(f;c)^n_t$ reaches 90% of the climatology ensemble mean Shannon entropy $\langle H(c)\rangle_t$. The resulting timescales are tabulated in table 1. The uncertainty range (square brackets) for the timescale is estimated by repeating the above procedure N times replacing $\langle\overline{c}\rangle_t$ with each of the climatology ensemble members $\overline{c}^n_t$. Results for a typical zone, Zone 6, are shown in Figures 5 and 6. Predictability time scale obtained by comparing forecast ensemble members to the single unperturbed member from the climatology ensemble were similar to when compared with the mean of climatology ensemble (see Figures S26-S32 in supplemental information). Comparing climatology ensemble members with the single unperturbed climatology member also gave similar results (see Figures S33-S39 in supplemental information).

Figure 7 shows a similar method of estimating predictability at a single grid point near the Mount Hope Bay (MtHB) buoy, which follows the same algorithm but without spatial averaging. The running window method is useful when the time interval under consideration is long enough to provide a reasonable histogram approximation of the

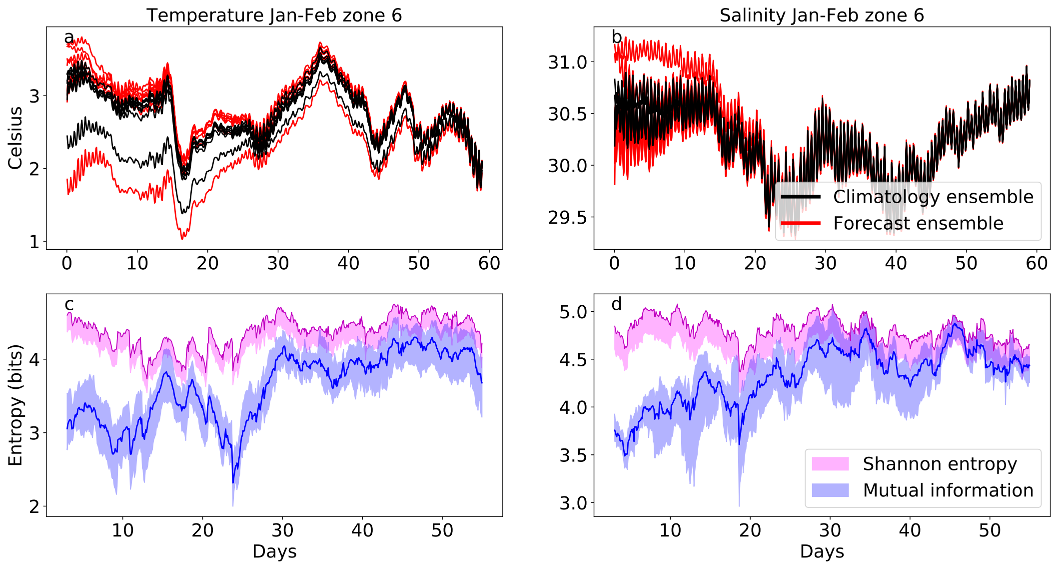

Figure 5. Predictability results for Zone 6 volume-averaged temperature (c) and salinity (d) in January to February. Top: Temperature (a) and salinity (b) timeseries from ensemble members is plotted for 7 climatology ensemble members (in black) and 7 forecast ensemble members (in red). Bottom: Information theory metrics (temperature (c) and salinity (d)) shows the convergence of mutual information (blue) with Shannon entropy (pink). The blue range indicates the forecast ensemble and the blue line is the forecast ensemble mean. The Shannon entropy of the climatological mean is shown at the top of the pink range and 90% of this value is shown as the bottom of the pink range. The mutual information converges to 90% of the Shannon entropy in 7-40 days (Table 1). Figures S14 to S19 in the supporting information show similar plots for other zones.

temporal probability distribution. The histogram intervals and bin sizes were chosen for each case such that the predictability time period is not sensitive to variations around those values (overly small or large choices show significant dependence on choices of binning and duration). The predictability timescale remains more sensitive to τ than the number of bins. While entropy and mutual information are both sensitive to data binning and duration choices, the timescale for mutual information to converge to Shannon entropy is less sensitive for the selected bin sizes and duration.

The second spatial variability method evaluates entropy using all spatial grid points within a zone. Let Z be the set of all grid points in a zone. $I(f;c)_t^n$ is evaluated from Equation 2 between the spatial histograms estimating the probability distribu-

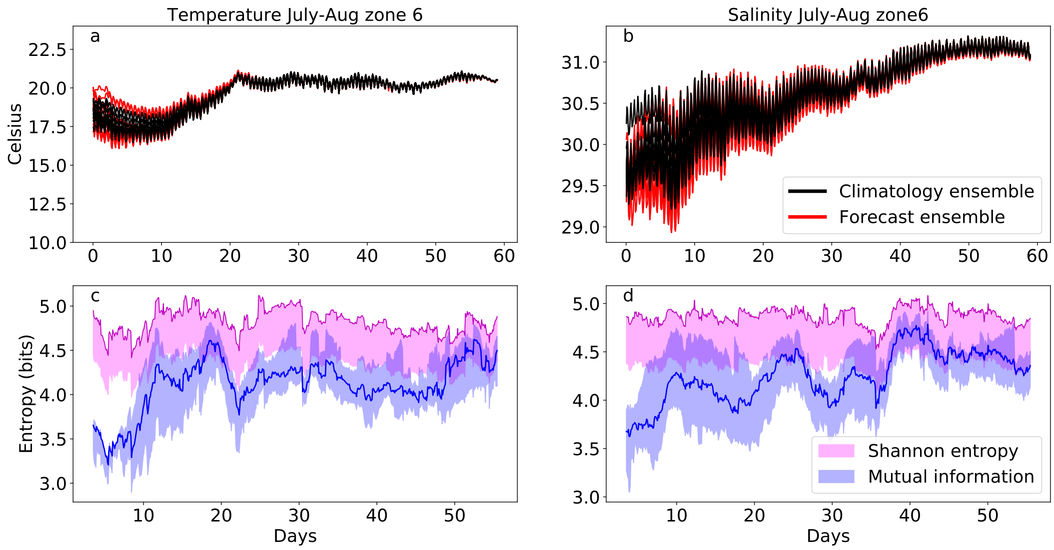

Figure 6. Predictability results for Zone 6 volume-averaged temperature (c) and salinity (d) in July to August. Top: Temperature (a) and salinity (b) timeseries from ensemble members is plotted for 10 climatology ensemble members (in black) and 10 forecast ensemble members (in red). Bottom: Information theory metrics (temperature (c) and salinity (d)) shows the convergence of mutual information (blue) with Shannon entropy (pink). The blue range indicates the forecast ensemble and the blue line is the ensemble mean. The shannon entropy of the climatological mean is shown at the top of the pink range and 90% of this value is shown as the bottom of the pink range. Figures S20 to S25 in the supporting information show similar plots for other zones.

tions of $\langle c \rangle_{t,i\in Z}$ and $f^n_{t,i\in Z}$. $H(c)^n_t$ is evaluated using Equation 1 for $\langle \overline{c} \rangle_{t,i\in Z}$. This approach eliminates the need for time windows by comparing the spatial variation between the forecast and climatology ensemble mean. This methodology has a utility when predictability is short so a running window may be longer than the predictability timescale. For example, kinetic energy has low predictability and hence this approach is used and is shown for Zone 6 in Figure 8.

Both the running window and spatial variability approaches use data without fixed references and are non-parametric. The data is not assumed to be Gaussian or any other distribution and hence our approach is robust towards all kinds of probability distributions, so long as the sampling is such that the histograms are an accurate representation of the probability distributions. Likewise, the method measures vari-

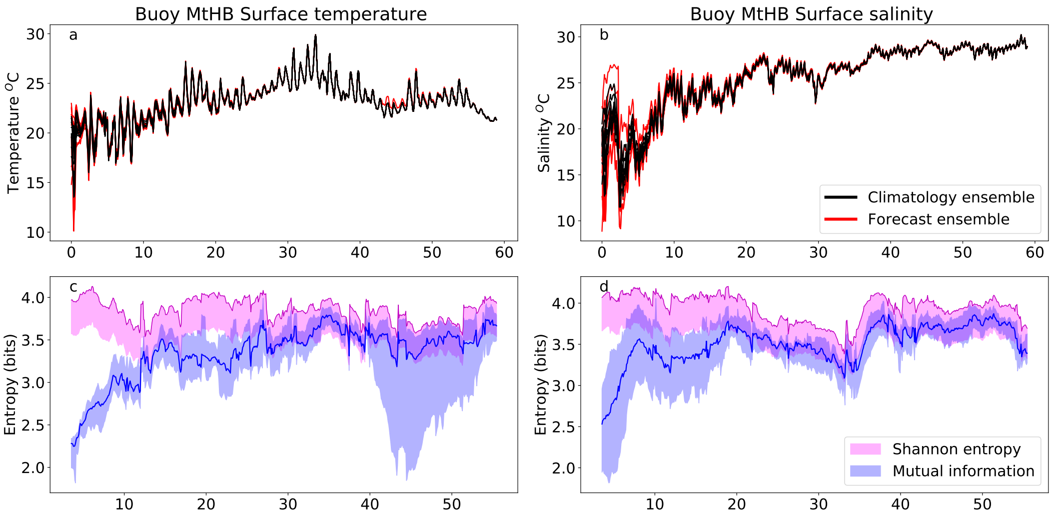

Figure 7. Surface temperature (a) and salinity (b) predictability metrics during July-August at one grid point closest to Mount Hope Bay (MtHB) buoy as shown in Figure 2. Information theory metrics for temperature and salinity are shown in c and d respectively. Surface temperature at this location is predictable for 27.4 [13.7 - 27.4] days and surface salinity is predictable for 18.5 [8.3 - 19.5] days. Figure S13 in the supporting information shows bottom temperature and salinity predictability.

ability by the same units of measure in the forecasts and climatology, so the units or standards of measurement are consistent regardless of whether physical, biological, environmental, or other metrics are chosen.

5 Results

5.1 Predictability results

Figures 5 and 6 show typical temperature and salinity results, drawn for both seasons from Zone 6. Other zones are similarly illustrated in the supplementary material. In each figure, the first row shows a timeseries comparison between the climatology ensemble (black) and forecast ensemble (red). The second row has information theory statistics, which permit a more precise time of convergence than just comparison of the timeseries in the upper row. Magenta shows $H(X)$ and the range of $H(x)^n$, the entropy of $c_{t,i}$, blue members represent $I(X;Y)^n$ and single blue line between blue shaded region is the average $I(X;Y)$ over all the $I(X;Y)^n$. Table 1 has the predictability timescales and uncertainty range. Results for each zone from 1 to 7 and combinations

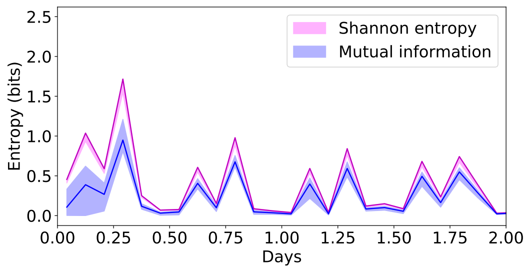

Figure 8. Kinetic energy predictability is less than 2 days for Zone 6 for July-August. In this case, the spatial variability metric was used as the predictability timescale was shorter than the running time windows. Using all the spatial grid points instead of the volume weighted time series provides enough sample points to create a probability distribution, and is also sensitive to convergence in higher-order statistics beyond the spatial mean. Alternatively, very frequent output windows in time could have been used with the time window method, but this method was chosen to illustrate the possibilities when initial condition effects are quickly lost and there is rapid convergence to climatology. Kinetic energy results for other zones is similar and are given in supporting information Figures S7 to S12.

of zones which progressively increase in volume from North to South are tabulated in Table 1. The combined zones enable us to compare the predictability time scale with flushing/turnover time scales evaluated over similar combined regions measured by distance from the northern end of the estuary to the southern end (Figure 9).

Table 1 compares the predictability timescales by region and season. The summer timescales tend to be longer, reflecting the typically drier conditions during summer of the year simulated. The timescales for salinity tend to increase as more and more of the Bay regions are included, indicating that anomalies persist somewhere within the Bay after initialization. For regions within the Bay, local circulations and patterns of mixing differ among the different regions, but few clear patterns emerge. Overall, the span of timescales is from 6.9 days to 40.5, indicating that predictions of a week or longer may potentially have skill, and that 1-2 months of spinup is necessary for initial condition effects to be lost and for forcing to become dominant.

Figure 7 shows an example of temperature and salinity predictability for a single grid point, for a location nearest to the Mount Hope Bay buoy (MtHB in Figure 2). Perhaps counter to intuition, the central predictability timescale estimates (temper-

Table 1. Predictability in days for January-February with respect to zones for temperature and salinity based on when mean mutual information between ensemble members and climatology reaches 90% of climatology's Shannon entropy for the first time. The range is estimated by the range over each of the member of the climatology ensemble.

zones	January-February		July-August	
	Temp.Pred.(days)	SalinityPred.(days)	Temp.Pred.(days)	SalinityPred.(days)
1	36.5[36.2-37.2]	7.3[6.9-7.7]	10.2[9.1-10.6]	9.4[9.1-9.9]
2	14.2[12.1-14.3]	10.5[9.4-11.0]	10.3[9.3-33.0]	27.7[26.6-29.0]
3	11.5[11.5-12.0]	18.3[18.3-19.0]	16.4[16.0-27.4]	23.8[22.1-26.3]
4	13.0[13.0-14.9]	16.9[16.7-17.0]	22.5[21.1-31.5]	31.5[31.4-32.5]
5	11.9[11.7-13.0]	16.9[16.8-17.1]	9.6[9.5-23.0]	18.5[16.6-31.2]
6	30.2[30.0-33.8]	21.9[20.1-23.0]	17.8[17.3-27.0]	23.0[22.9-24.5]
7	14.9[14.1-28.7]	25.5[19.0-26.7]	22.5[20.4-31.0]	10.0[9.0-10.3]
1to2	15.0[14.2-33.5]	9.5[9.3-9.5]	23.4[22.3-34.2]	24.8[22.1-28.1]
1to5	11.8[11.7-29.8]	17.1[17.1-17.6]	10.0[10.0-26.2]	29.4[29.4-30.6]
1to7	14.0[13.2-29.7]	17.0[17.0-18.0]	32.6[18.4-40.5]	31.4[31.4-32.6]

ature: 27.4 [13.7 - 27.4] days; salinity: 18.5 [8.3 - 19.5] days) is quite long for this one gridpoint in comparison to the predictability of the whole Zone 5 that contains it (Table 1 and Supplementary figures; zone-averaged temperature: 9.6 [9.5 - 23.0] days; zone-averaged salinity: 18.5 [16.6 - 31.2] days), but note that the estimated ranges are consistently overlapping. There are many processes which would increase the amount of internal variability at a single location, such as meandering currents, waves, and other effects of flow-topography interaction. Thus, the predictability of an individual measurement location need not agree with the predictability of the region containing it, because of this internal variability would be missing from the zone averages. However, in this case and indeed for all of the monitoring buoy locations shown in Figure 2, the buoys are deployed deliberately in locations thought to be representative of their section of the Bay rather than within a particular feature such as a regular plume or jet. Thus, the agreement in predictability timescales is perhaps not coincidental, but reflects judicious choices for observational advantage. Presenting results at this single location highlights the possibility of evaluating predictability metrics at one location, not just in regional averages, and the potential reasons why these two approaches may differ.

Likewise, predictability is not limited to temperature and salinity. The predictability of kinetic energy is shown in Figure 8 for Zone 6 and is less than 2 days. The mutual information converges towards Shannon entropy within a very short period, and the alternative method of calculating the probability distribution using spatial variability is needed. As will be shown in the next section, there is consistency between the timescales of freshwater and salinity flushing and predictability timescales, which argues that the estuarine circulation tends to dominate these tracers. However, anomalies in the kinetic energy within a region are much more quickly generated (by winds and instabilities) and dissipated (by viscous and drag parameterizations) in the OSOM, and so the predictability timescale is one to two orders of magnitude shorter for kinetic energy than for temperature and salinity. Thus, the kinetic energy example illustrates that it is important to evaluate predictability on each metric of forecast interest. The next section explores the physical implications of the predictability timescales in comparison to flushing timescales.

6 Turnover timescales

The turnover or flushing time scale is the time scale required for replenishment of a particular water mass in the estuary, based on its rate of resupply or removal. For a water mass having a volume V and volume flux rate Q, the flushing time scale is simply $\tau = V/Q$ (e.g., Monsen et al., 2002; Rayson et al., 2016). In the present study the freshwater turnover/flushing time scale and the salinity turnover time scale are calculated from the model output and compared with the predictability time scales. The approach here follows Lemagie & Lerczak (2015) in comparing estuarine timescales by standard definitions, except here the estuarine timescales are also compared with the predictability timescale.

The freshwater volume is estimated using the relation

$$V_f = \left(1 - \frac{s}{s_o}\right) V_b, \tag{3}$$

where V_f is the freshwater volume, s is the volume weighted average salinity of the Bay, s_o is the salinity of the open ocean or the salinity of the incoming volume flux in the region under consideration, and V_b is the volume of the Bay. The freshwater flushing time scale is

$$\tau_f = \frac{V_f}{Q_r}, \tag{4}$$

where Q_r is the river supply and runoff.

The salinity turnover timescale follows the isohaline procedure of MacCready (2011). The fluxes of saline water masses are calculated for each salinity class. Let $Q(s)$ be tidally averaged salinity flux corresponding to salinity s and be given by:

$$Q(s) = \left\langle\!\!\left\langle \int_{A_s} u \,\mathrm{d}A \right\rangle\!\!\right\rangle \ . \tag{5}$$

where double angled brackets denote temporal filtering over a tidal period with a Butterworth filter. A_s is the cross sectional area having salinity greater than s. $Q(s)$ is the salinity flux for the salinity belonging in the range (s, s_{max}). $Q(s)$ is evaluated laterally at a vertical cross section along the estuary, beginning at the north and proceeding south. The flux moving in, Q_{in} and moving out, Q_{out}, of the estuary is calculated using an integral over the salinity classes:

$$Q_{in,out} = \int \left.\frac{\partial Q}{\partial s}\right|_{in,out} ds \ , \tag{6}$$

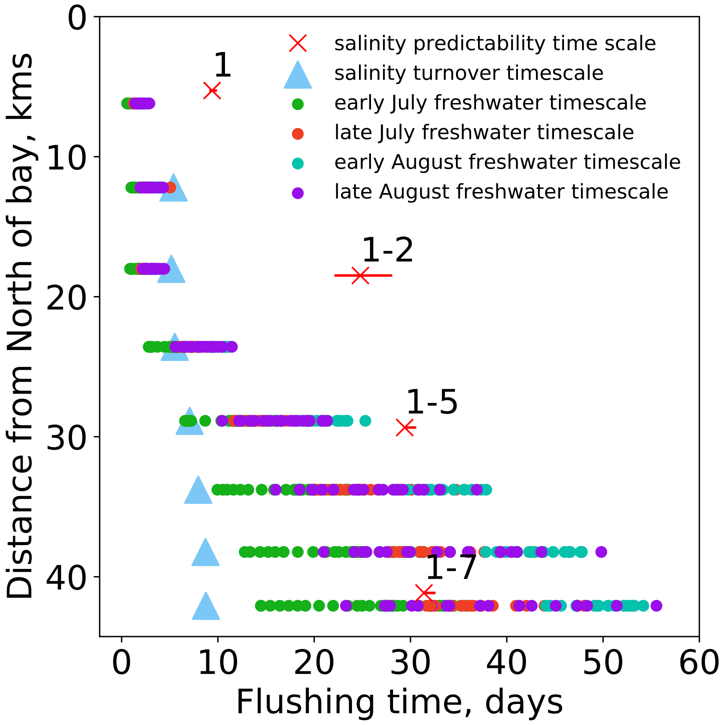

Figure 9. Freshwater flushing timescales, salinity turnover timescales, and salinity predictability timescales for July-August as a function of distance from the northernmost extent of Narragansett Bay. Blue boxes show the salinity flushing timescale (Equation 8). Circular scattered points show the freshwater flushing time estimated from freshwater volume and divided by river input (Equation 4). Different colors show averages over different periods within July - August. The salinity predictability time scale is shown by red crosses, for Zone 1 and then the combined regions (1 to 2, 1 to 5, 1 to 7) in the last three rows of Table 1.

where "in" and "out" are evaluated on the basis of the sign of the integrand. MacCready (2011) defines the fluxes as total exchange flow (TEF). The TEF relates to corresponding salt fluxes of

$$F_{in,out} = \int s \frac{\partial Q}{\partial s}\bigg|_{in,out} ds \ . \tag{7}$$

The MacCready (2011) approach results in the salinity turnover timescale of

$$\tau_s = \frac{\int s dV}{F_{in}} \ . \tag{8}$$

Using above definitions, τ_f and τ_s have been found by considering a control volume with one end fixed at the mouth of Providence river at the northernmost end of NB and the other end gradually increasing towards the open ocean. The intention is to estimate these timescales in order to check whether they agree with predictability timescales. The time scale results are displayed in Figure 9 along with predictability timescales for the corresponding regions. The y-axis is the distance from the north of the Bay to the south end of each control volume. The x-axis provides the ranges of timescales. The predictability timescales (red crosses) are consistent in magnitude with the various flushing timescales and increase as the quantity of the Bay in the control volume increases (although somewhat less rapidly with distance). Four time periods are shown by colors–early and late for July and August–illustrating that the flushing timescales vary signficantly (with the amount of precipitation, mainly).

7 Discussion

The predictability timescales measure the persistence of statistical anomalies deviating from climatology that stem from the initial conditions. These anomalies might be detected to decay, through information theory metrics, by a variety of processes: tidal or wind-driven mixing, being carried out of the Bay by advection, or becoming so well stirred by turbulent motions that they no longer persist as statistical anomalies. The consistency between the salinity and temperature predictability timescales and the salinity flushing timescales illustrates that it is likely that these anomalies are removed from the Bay primarily by the estuarine circulation whose timescale is estimated with the variety of flushing timescales shown. Even pointwise measurements tend to agree with their zone-average prediction timescale (Figure 7), which indicates that the anomalies in OSOM temperature and salinity tend to be fairly mixed over broad areas, so that regions and buoys capture much the same information. It is not clear if this is true in the real Narragansett Bay to the same degree, but the consistency in the degree of variability between the modeled buoy locations and the buoy observations (Figure 3) suggests that this may be.

The predictability timescale of kinetic energy is one to two orders of magnitude shorter than that of temperature or salinity (Figure 8). This suggests that kinetic energy in NB is not governed solely by the estuarine overturning. Indeed, NB and the OSOM are highly tidally-driven – with the majority of the kinetic energy involved in the ebb and flow. Apparently, the propagation of the tidal energy into the Bay through waves, winds, currents, dissipation and drag, and generally perturbations to the surface elevation and kinetic energy, are a rather different set of processes operating on very different timescales from the estuarine overturning that transports the salinity and temperature anomalies and their predictability.

8 Conclusions:

This study has introduced the Ocean State Ocean Model (OSOM) and measures of its intrinsic timescales. The predictability timescales range from 6.9 to 40.5 days for temperature and salinity. The predictability timescales differ for different periods of the year and the region under observation–with generally longer periods for the larger basins and under drier conditions. These relationships are consistent with the expectations of estuarine circulation dominating the flushing of anomalies in salinity and temperature, and these predictability timescales are quantitatively similar to the range of estimates of flushing timescales.

Information theory proves useful for quantifying predictability. It can also be applied to other variables such as physical, biogeochemical, and environmental metrics that are being considered for forecasting with the OSOM. Not all variables have the same timescales, as some rely on processes that operate at different speeds.

While it is important to know the predictability timescales for understanding the constraints on spinning up a model and the *potential* length of a forecast, it is important to keep in mind that the *skill* of a forecast is not simply related to the predictability. Here the model skill is adequate for the assessment of predictability (Section 2.1), but the model shows skill deficiencies in some locations, as highlighted here by comparison to observations at the Greenwich Bay buoy (Figure 3). Such biases and errors in a model may not affect the predictability timescale, but they clearly reduce the value of a forecast. Future work in tuning the model parameterizations and improved forcing will increase model skill but are not expected to change the predictability. A higher-resolution version of the model is expected to have better skill and lower biases, but the stronger chaotic transport and resolved eddying features in such a model are likely to decrease the predictability timescale (by increasing internal variability). This is one key reason why predictability metrics are not an aspect of Narragansett Bay itself, but only of this particular model: the OSOM.

In the case of temperature and salinity predictability in the OSOM, forced estuarine circulations tend to set the dominant timescales. Knowing this is useful in estimating forecast windows, spin up times, and sensitivity to forcing variability. Other systems, and perhaps the kinetic energy in this system, are dominated by internal variability rather than forced variability. A companion paper expands on this topic for coastal modeling, where a variety of different boundary forcing mechanisms can contribute.

Acknowledgments

The Rhode Island Coastal Ecology Assessment Innovation & Modeling grant (NSF 1655221) supported this work. BFK was also supported by ONR N00014-17-1-2963 and NSF 1350795. This material is based upon work conducted at a Rhode Island NSF EPSCoR research facility Center for Computation and Visualization (Brown University), supported in part by the National Science Foundation EPSCoR Cooperative Agreement #OIA-1655221. J. Benoit provided the LandSat analysis dataset and M. Brush contributed the drainage area dataset. All the data and the codes used to plot results can be downloaded via Brown University's digital archive DOI: `https://doi.org/10.26300/crbx-9784`.

References

Beardsley, R. C., & Chen, C. (2014). Northeast Coastal Ocean Forecast System (NECOFS): A multi-scale global-regional-estuarine FVCOM model. In *2014 agu fall meeting.*

Bergondo, D. (2004). *Examining the processes controlling water column variability in narragansett bay: Time series data and numerical modeling* (Doctoral dissertation, University of Rhode Island). Retrieved from `https://digitalcommons.uri.edu/cgi/viewcontent.cgi?article=1781&context=dissertations`

Bergondo, D., & Kincaid, C. (2007). *Development and calibration of a model for tracking dispersion of waters from narragansett bay commission facilities within the providence river and narragansett bay* (Tech. Rep.). Narragansett Bay Commission.

Chapman, D. C. (1985). Numerical treatment of cross-shelf open boundaries in a barotropic coastal ocean model. *Journal of Physical oceanography*, *15*(8), 1060–

1075.

Codiga, D. L., Stoffel, H. E., Deacutis, C. F., Kiernan, S., & Oviatt, C. A. (2009). Narragansett bay hypoxic event characteristics based on fixed-site monitoring network time series: intermittency, geographic distribution, spatial synchronicity, and interannual variability. *Estuaries and coasts*, *32*(4), 621–641.

Cover, T. M., & Thomas, J. A. (2012). *Elements of information theory.* John Wiley & Sons.

DelSole, T. (2004). Predictability and information theory. part i: Measures of predictability. *Journal of the atmospheric sciences*, *61*(20), 2425–2440.

DelSole, T., & Shukla, J. (2010). Model fidelity versus skill in seasonal forecasting. *Journal of Climate*, *23*(18), 4794–4806.

DelSole, T., & Tippett, M. K. (2007). Predictability: Recent insights from information theory. *Reviews of Geophysics*, *45*(4).

Engle, V. D., Kurtz, J. C., Smith, L. M., Chancy, C., & Bourgeois, P. (2007). A classification of us estuaries based on physical and hydrologic attributes. *Environmental Monitoring and Assessment*, *129*(1-3), 397–412.

Fairall, C. W., Bradley, E. F., Hare, J., Grachev, A. A., & Edson, J. B. (2003). Bulk parameterization of air–sea fluxes: Updates and verification for the coare algorithm. *Journal of climate*, *16*(4), 571–591.

Fano, T. (1961). *Transmission of information, a statistical theory of communications.* The M.I.T. Press, John Wiley & Sons.

Flather, R. A. (1976). *Practical aspects of the use of numerical models for surge prediction.* Institute of Oceanographic Sciences, Bidston Observatory.

Hartley, R. V. (1928). Transmission of information 1. *Bell System technical journal*, *7*(3), 535–563.

Haven, K., Majda, A., & Abramov, R. (2005). Quantifying predictability through information theory: small sample estimation in a non-gaussian framework. *Journal of Computational Physics*, *206*(1), 334–362.

Hayward, S., Hashemi, M. R., Torres, M., Grilli, A., Grilli, S., King, J., ... Spaulding, M. (2018). Numerical simulation of coastal erosion and its mitigation by living shoreline methods: A case study in southern rhode island. *Shore & Beach*, *86*(4), 13.

Jin, Y., Rong, X., & Liu, Z. (2018). Potential predictability and forecast skill in ensemble climate forecast: a skill-persistence rule. *Climate dynamics*, *51*(7-8), 2725–2742.

Kleeman, R. (2002). Measuring dynamical prediction utility using relative entropy. *Journal of the atmospheric sciences*, *59*(13), 2057–2072.

Knudsen, M. (1900). Ein hydrographischer lehrsatz. *Ann. Hydrogr. Marit. Meteorol.*, *28*, 316-320.

Kumar, A., Peng, P., & Chen, M. (2014). Is there a relationship between potential and actual skill? *Monthly Weather Review*, *142*(6), 2220–2227. doi: 10.1175/MWR-D-13-00287.1

Lellouche, J.-M., Greiner, E., Le Galloudec, O., Regnier, C., Benkiran, M., Testut, C.-E., ... Drillet, Y. (2018). Mercator ocean global high-resolution monitoring and forecasting system. *New Frontiers in Operational Oceanography*, 563–592. Retrieved from `https://doi.org/10.17125/gov2018.ch20`

Lemagie, E. P., & Lerczak, J. A. (2015). A comparison of bulk estuarine turnover timescales to particle tracking timescales using a model of the yaquina bay estuary. *Estuaries and coasts*, *38*(5), 1797–1814.

Leroux, S., Penduff, T., Bessières, L., Molines, J.-M., Brankart, J.-M., Sérazin, G., ... Terray, L. (2018). Intrinsic and atmospherically forced variability of the amoc: insights from a large-ensemble ocean hindcast. *Journal of Climate*, *31*(3), 1183–1203.

Leung, L.-Y., & North, G. R. (1990). Information theory and climate prediction. *Journal of Climate*, *3*(1), 5–14.

Liu, Q., Rothstein, L. M., Luo, Y., Ullman, D. S., & Codiga, D. L. (2016). Dynamics of the periphery current in rhode island sound. *Ocean Modelling*, *105*, 13–24. Retrieved from `https://doi.org/10.1016/j.ocemod.2016.07.001`

Llovel, W., Penduff, T., Meyssignac, B., Molines, J.-M., Terray, L., Bessières, L., & Barnier, B. (2018). Contributions of atmospheric forcing and chaotic ocean variability to regional sea level trends over 1993–2015. *Geophysical Research Letters*, *45*(24), 13–405.

MacCready, P. (2011). Calculating estuarine exchange flow using isohaline coordinates. *Journal of Physical Oceanography*, *41*(6), 1116–1124.

Majda, A., Kleeman, R., Cai, D., et al. (2002). A mathematical framework for quantifying predictability through relative entropy. *Methods and Applications of Analy-*

sis, *9*(3), 425–444.

Marchesiello, P., McWilliams, J. C., & Shchepetkin, A. (2001). Open boundary conditions for long-term integration of regional oceanic models. *Ocean modelling*, *3*(1-2), 1–20.

McManus, M. C., Ullman, D. S., Rutherford, S. D., & Kincaid, C. (2020). Northern quahog (mercenaria mercenaria) larval transport and settlement modeled for a temperate estuary. *Limnology and Oceanography*, *65*(2), 289–303. Retrieved from https://doi.org/10.1002/lno.11297

Mel, R., & Lionello, P. (2014). Storm surge ensemble prediction for the city of venice. *Weather and forecasting*, *29*(4), 1044–1057. Retrieved from https://doi.org/10.1175/WAF-D-13-00117.1

Monsen, N. E., Cloern, J. E., Lucas, L. V., & Monismith, S. G. (2002). A comment on the use of flushing time, residence time, and age as transport time scales. *Limnology and oceanography*, *47*(5), 1545–1553.

Moore, A. M., Arango, H. G., Broquet, G., Powell, B. S., Weaver, A. T., & Zavala-Garay, J. (2011). The regional ocean modeling system (roms) 4-dimensional variational data assimilation systems: Part i–system overview and formulation. *Progress in Oceanography*, *91*(1), 34–49. Retrieved from https://doi.org/10.1016/j.pocean.2011.05.004

Mukai, A. Y., Westerink, J. J., Luettich Jr, R. A., & Mark, D. (2002). *Eastcoast 2001, a tidal constituent database for western north atlantic, gulf of mexico, and caribbean sea* (Tech. Rep.). Vicksburg, MS: Engineer Research and Development Center Coast and Hydraulics Lab.

Mustard, J., Carney, M., & Sen, A. (1999). The use of satellite data to quantify thermal effluent impacts. *Estuarine, Coastal and Shelf Science*, *49*(4), 509–524. Retrieved from https://doi.org/10.1006/ecss.1999.0517

Pilson, M. E. (1985). On the residence time of water in narragansett bay. *Estuaries*, *8*(1), 2–14.

Pinardi, N., & Coppini, G. (2010). Operational oceanography in the mediterranean sea: the second stage of development. *Ocean Sci*, *6*, 263–267. Retrieved from https://doi.org/10.5194/os-6-263-2010

Raboudi, N. F., Ait-El-Fquih, B., Dawson, C., & Hoteit, I. (2019). Combining hybrid and one-step-ahead smoothing for efficient short-range storm surge forecasting with an ensemble kalman filter. *Monthly Weather Review*, *147*(9), 3283–3300. Retrieved from https://doi.org/10.1175/MWR-D-18-0410.1

Rayson, M. D., Gross, E. S., Hetland, R. D., & Fringer, O. B. (2016). Time scales in galveston bay: An unsteady estuary. *Journal of Geophysical Research: Oceans*, *121*(4), 2268–2285.

Roulston, M. S., & Smith, L. A. (2002). Evaluating probabilistic forecasts using information theory. *Monthly Weather Review*, *130*(6), 1653–1660.

Schneider, T., & Griffies, S. M. (1999). A conceptual framework for predictability studies. *Journal of climate*, *12*(10), 3133–3155.

Sethna, J., et al. (2006). *Statistical mechanics: entropy, order parameters, and complexity* (Vol. 14). Oxford University Press.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell system technical journal*, *27*(3), 379–423.

Shchepetkin, A. F., & McWilliams, J. C. (2005). The regional oceanic modeling system (roms): a split-explicit, free-surface, topography-following-coordinate oceanic model. *Ocean modelling*, *9*(4), 347–404.

Shukla, J. (1981). Dynamical predictability of monthly means. *Journal of the Atmospheric Sciences*, *38*(12), 2547-2572. Retrieved from https://doi.org/10.1175/1520-0469(1981)038<2547:DPOMM>2.0.CO;2 doi: 10.1175/1520-0469(1981)038⟨2547:DPOMM⟩2.0.CO;2

Smith, D. M., Cusack, S., Colman, A. W., Folland, C. K., Harris, G. R., & Murphy, J. M. (2007). Improved surface temperature prediction for the coming decade from a global climate model. *science*, *317*(5839), 796–799.

Ullman, D. S. (2019). *Hydrodynamic modeling of Narragansett Bay in support of the EcoGEM ecological model* (Tech. Rep. No. GSO No. 2019-01). University of Rhode Island. Retrieved from https://digitalcommons.uri.edu/cgi/viewcontent.cgi?article=1034&context=physical_oceanography_techrpts

Umlauf, L., & Burchard, H. (2003). A generic length-scale equation for geophysical turbulence models. *Journal of Marine Research*, *61*(2), 235–265.

Wertman, C. A. (2018). *Circulation & exchange within shelf & estuarine waters driven by the atmosphere, tides and buoyancy* (Doctoral dissertation, University of Rhode Island). Retrieved from https://digitalcommons.uri.edu/oa_diss/719

Wilkin, J., Levin, J., Lopez, A., Hunter, E., Zavala-Garay, J., & Arango, H.

(2018). Coastal ocean forecast system for the us mid-atlantic bight and gulf of maine. *New Frontiers in Operational Oceanography*, 593–624. Retrieved from https://doi.org/10.17125/gov2018.ch21

Xiu, P., Chai, F., Curchitser, E. N., & Castruccio, F. S. (2018). Future changes in coastal upwelling ecosystems with global warming: The case of the california current system. *Scientific reports*, *8*(1), 2866. Retrieved from https://doi.org/10.1038/s41598-018-21247-7

Supporting Information for "Consistent Predictability of the Ocean State Ocean Model (OSOM) using Information Theory and Flushing Timescales"

Aakash Sane[1], Baylor Fox-Kemper[2], Dave Ullman[3], Christopher Kincaid[3],

and Lewis Rothstein[3]

[1]School of Engineering, Brown University, Providence, RI

[2]Dept. of Earth, Environmental, and Planetary Sciences (DEEPS), Brown University, Providence, RI

[3]Graduate School of Oceanography, University of Rhode Island, Narragansett, RI

Contents of this file

1. Figures S1 to S37.

2. Table S1

Introduction

Text S1. The supplementary information contains figures S1 to S25. All the figures have been quoted in the main text. Also, table S1 shows root mean square error between model run and observations for surface temperature and salinity as well as bottom temperature and salinity.

References

Table S1. Root mean square error between observation and a single unperturbed model run

	Temperature ^{o}C		Salinity	
	Surface	Bottom	Surface	Bottom
CP	1.55	1.43	2.69	0.91
BR	2.42	1.26	3.4	1.24
NP	1.13	0.75	2.38	0.74
MtV	1.01	1.07	1.88	0.86
MtHB	1.87	0.77	2.02	0.94
QP	1.03	2.34	2.34	0.43
PP	0.91	0.82	2.91	0.59
GB	0.89	1.21	3.28	1.7

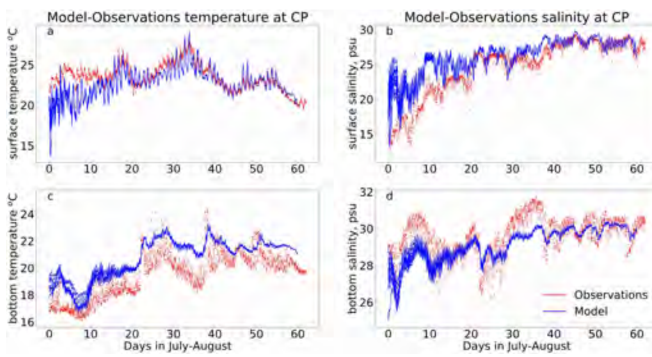

Figure S1. Comparison of model with observations collected at Conimicut Point (CP).

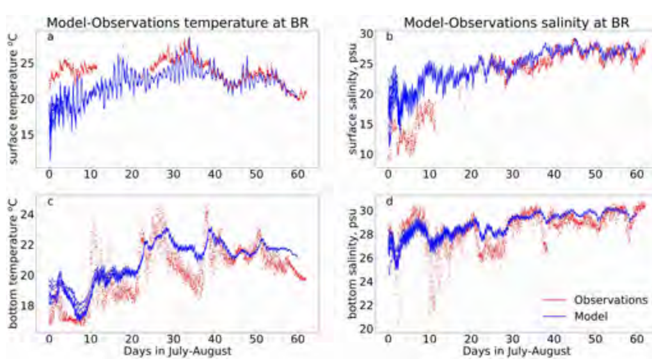

Figure S2. Comparison of model with observations collected at Bullock's Reach (BR).

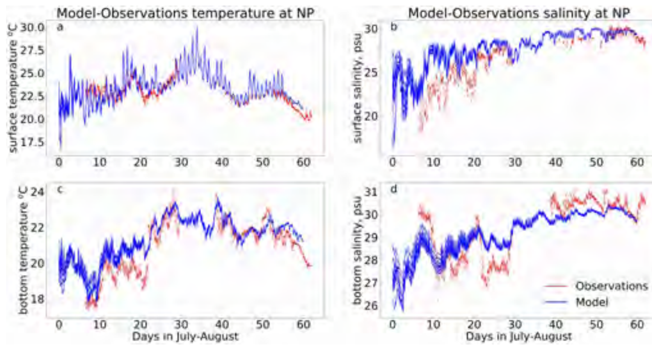

Figure S3. Comparison of model with observations collected at North Passage (NP).

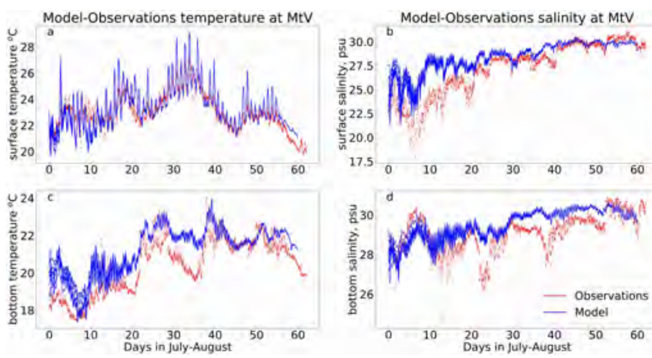

Figure S4. Comparison of model with observations collected at Mount View (MtV).

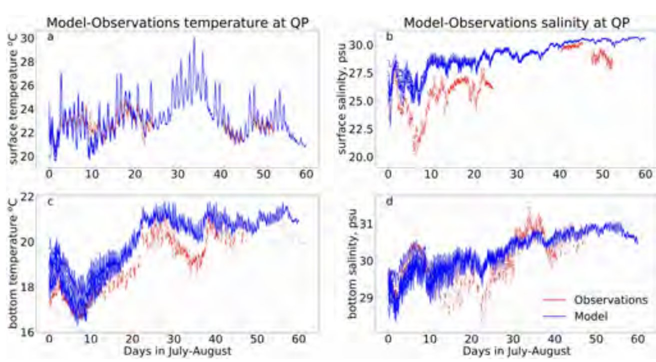

Figure S5. Comparison of model with observations collected at Quonset Point (QP).

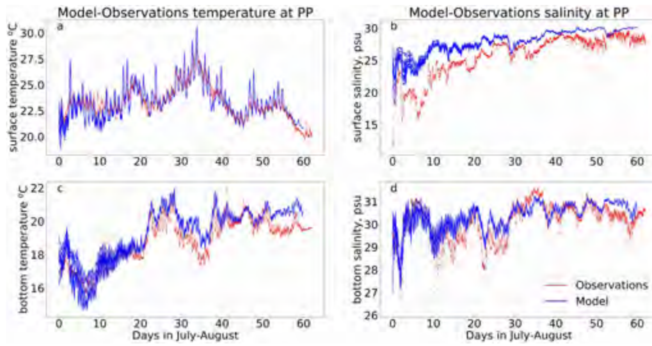

Figure S6. Comparison of model with observations collected at Poppasquash Point (PP).

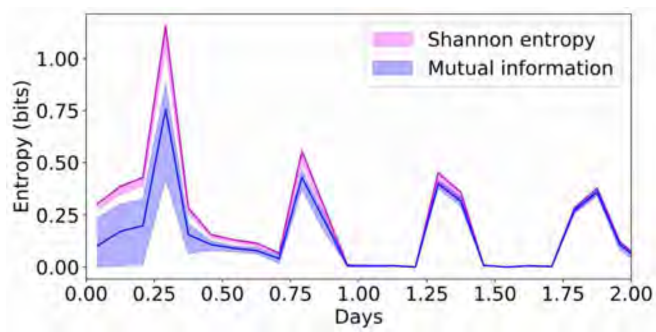

Figure S7. Mutual information between members of climatology ensemble compared with Shannon entropy of the mean of ensemble of zone 1 for the months of July-August.

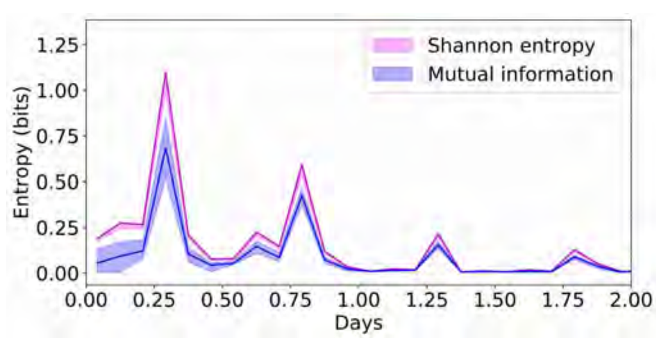

Figure S8. Mutual information between members of climatology ensemble compared with Shannon entropy of the mean of ensemble of zone 2 for the months of July-August.

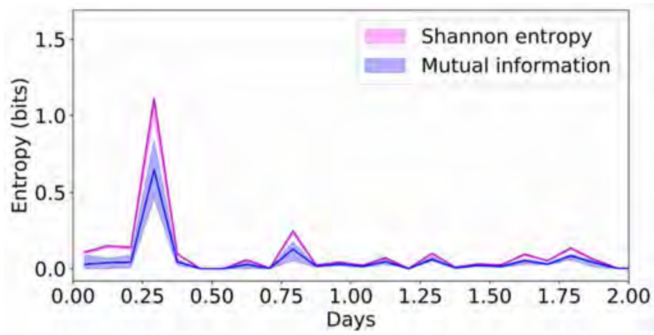

Figure S9. Mutual information between members of climatology ensemble compared with Shannon entropy of the mean of ensemble of zone 3 for the months of July-August.

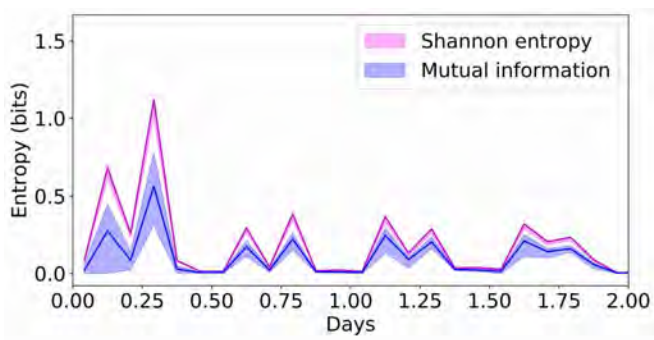

Figure S10. Mutual information between members of climatology ensemble compared with Shannon entropy of the mean of ensemble of zone 4 for the months of July-August.

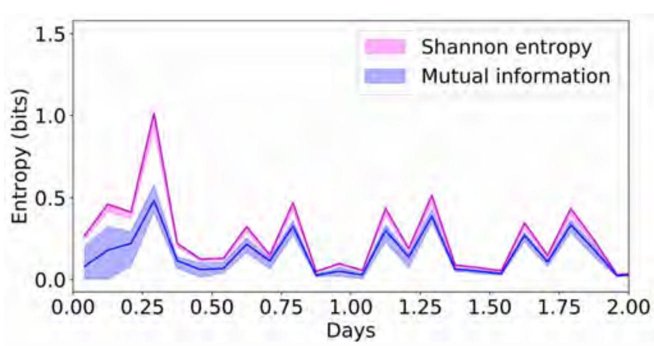

Figure S11. Mutual information between members of climatology ensemble compared with Shannon entropy of the mean of ensemble of zone 5 for the months of July-August.

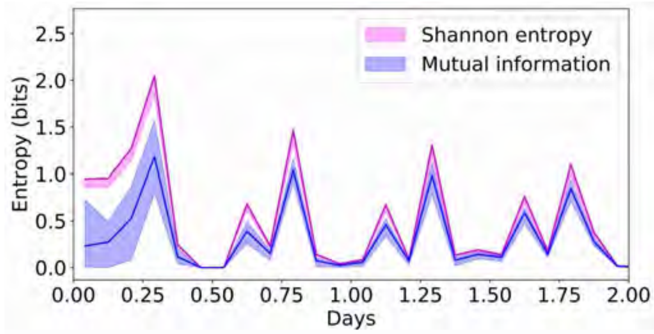

Figure S12. Figure shows predictability of kinetic energy. Mutual information between members of climatology ensemble compared with Shannon entropy of the mean of ensemble of zone 7 for the months of July-August.

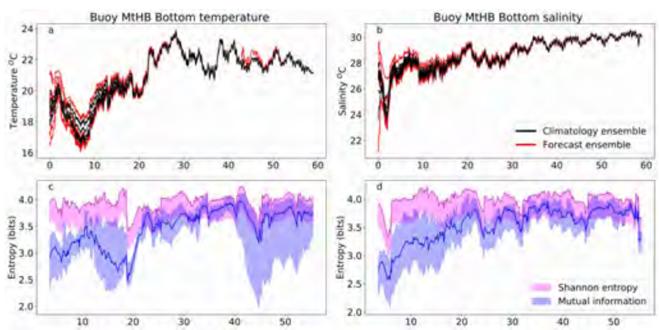

Figure S13. Bottom temperature predictability at grid point closest to MtHB buoy

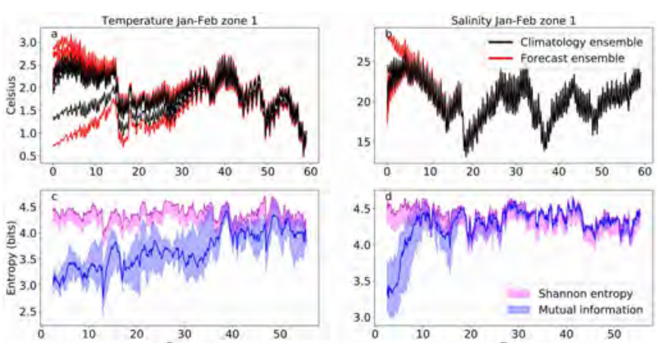

Figure S14. Results of zone 1 for January-February. Top figures shows temperature and salinity ensembles. Bottom figures show information entropy metrics applied between forecast and climatology ensembles.

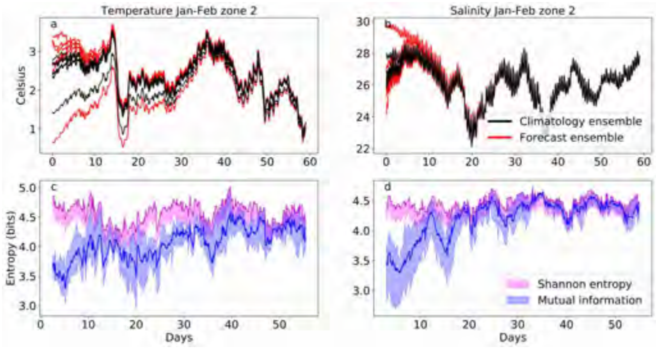

Figure S15. Results of zone 2 for January-February. Top figures shows temperature and salinity ensembles. Bottom figures show information entropy metrics applied between forecast and climatology ensembles.

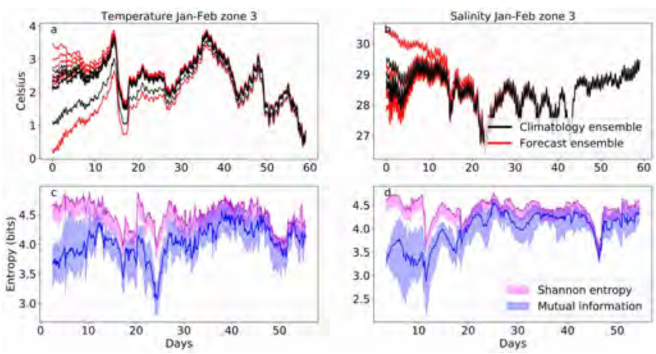

Figure S16. Results of zone 3 for January-February. Top figures shows temperature and salinity ensembles. Bottom figures show information entropy metrics applied between forecast and climatology ensembles.

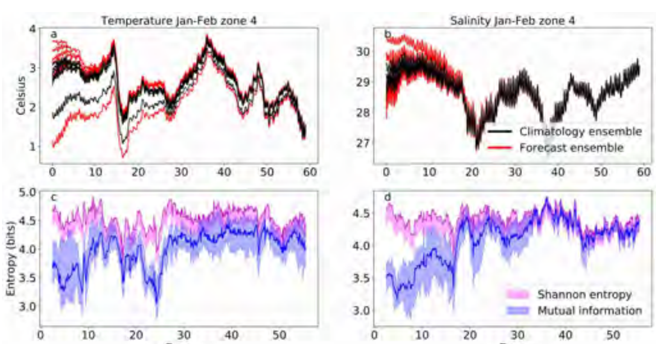

Figure S17. Results of zone 4 for January-February. Top figures shows temperature and salinity ensembles. Bottom figures show information entropy metrics applied between forecast and climatology ensembles.

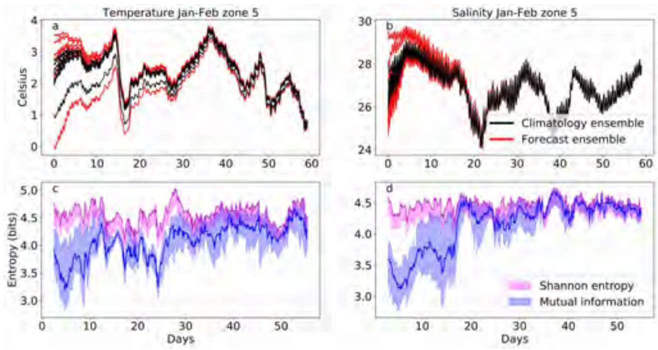

Figure S18. Results of zone 5 for January-February. Top figures shows temperature and salinity ensembles. Bottom figures show information entropy metrics applied between forecast and climatology ensembles.

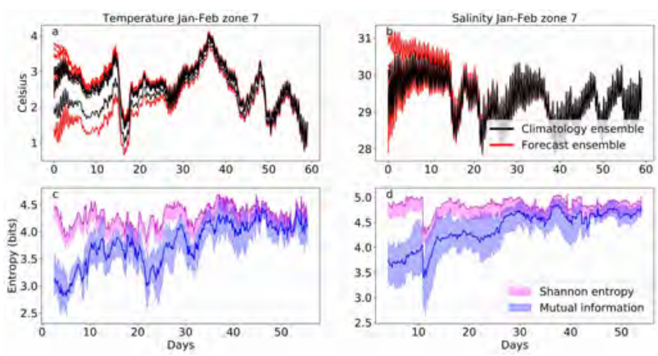

Figure S19. Results of zone 7 for January-February. Top figures shows temperature and salinity ensembles. Bottom figures show information entropy metrics applied between forecast and climatology ensembles.

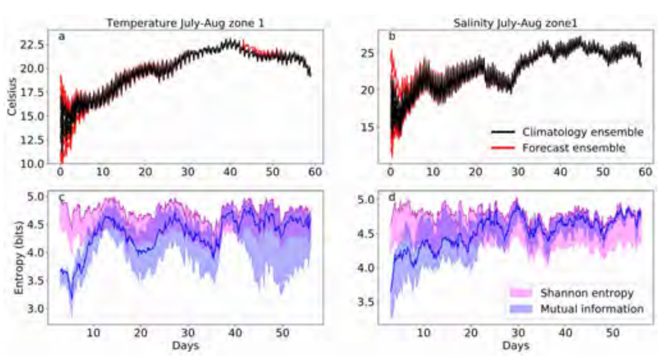

Figure S20. Results of zone 1 for July - August. Top figures shows temperature and salinity ensembles. Bottom figures show information entropy metrics applied between forecast and climatology ensembles.

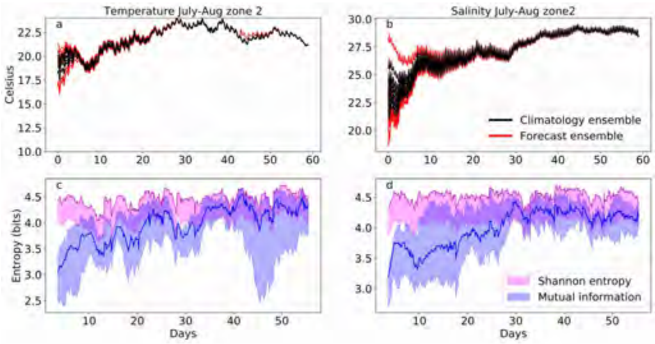

Figure S21. Results of zone 2 for July - August. Top figures shows temperature and salinity ensembles. Bottom figures show information entropy metrics applied between forecast and climatology ensembles.

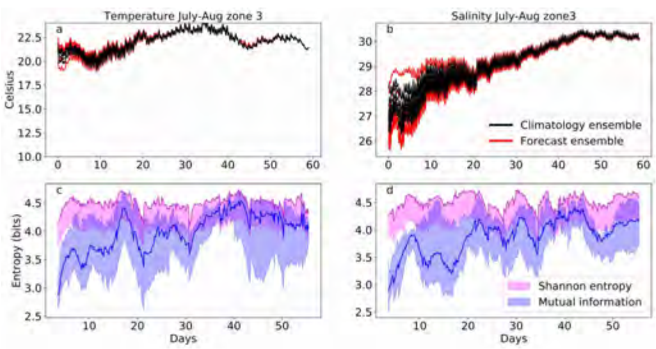

Figure S22. Results of zone 3 for July - August. Top figures shows temperature and salinity ensembles. Bottom figures show information entropy metrics applied between forecast and climatology ensembles.

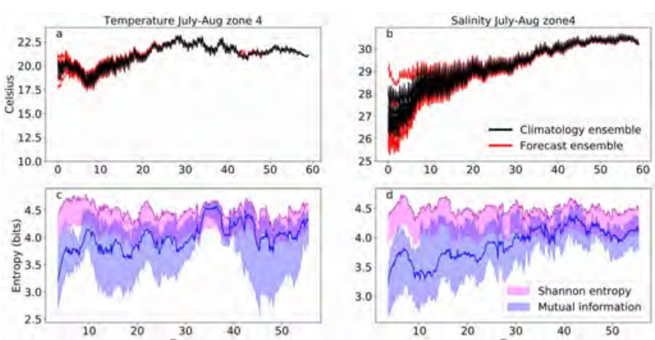

Figure S23. Results of zone 4 for July - August. Top figures shows temperature and salinity ensembles. Bottom figures show information entropy metrics applied between forecast and climatology ensembles.

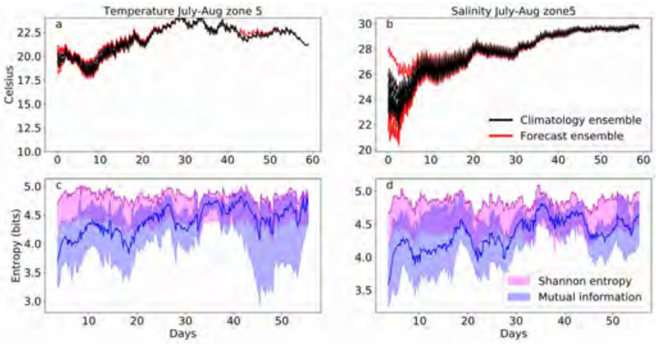

Figure S24. Results of zone 5 for July - August. Top figures shows temperature and salinity ensembles. Bottom figures show information entropy metrics applied between forecast and climatology ensembles.

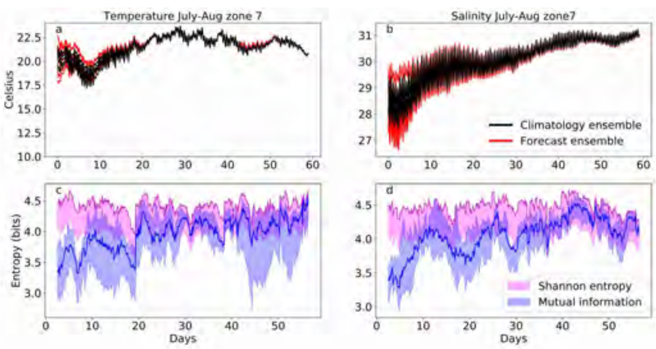

Figure S25. Results of zone 7 for July - August. Top figures shows temperature and salinity ensembles. Bottom figures show information entropy metrics applied between forecast and climatology ensembles.

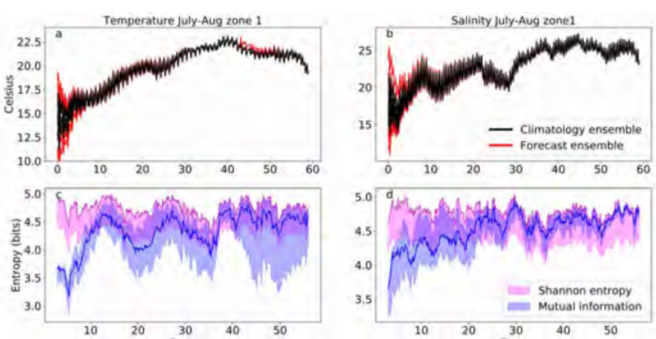

Figure S26. Results of zone 1 for July - August. Top figures shows temperature and salinity ensembles. Bottom figures show information entropy metrics applied between forecast and climatology ensembles.

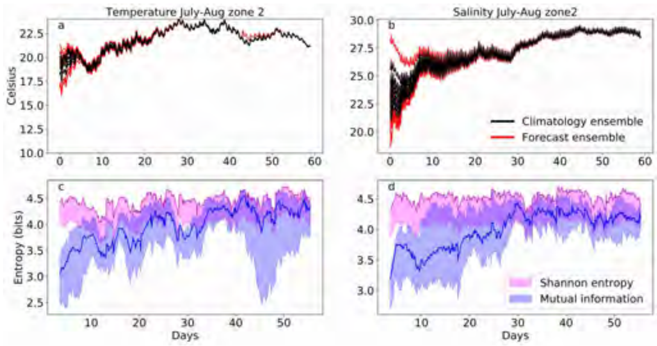

Figure S27. Results of zone 2 for July - August. Top figures shows temperature and salinity ensembles. Bottom figures show information entropy metrics applied between forecast and climatology ensembles.

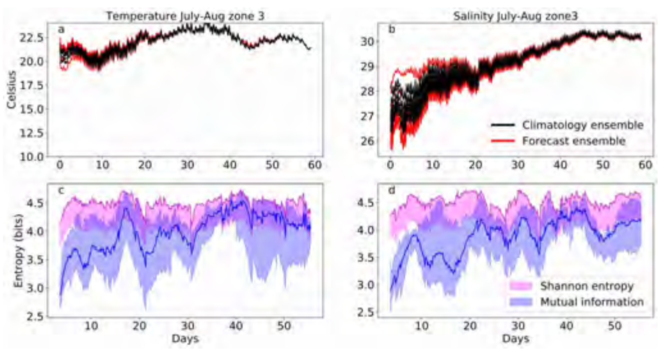

Figure S28. Results of zone 3 for July - August. Top figures shows temperature and salinity ensembles. Bottom figures show information entropy metrics applied between forecast and climatology ensembles.

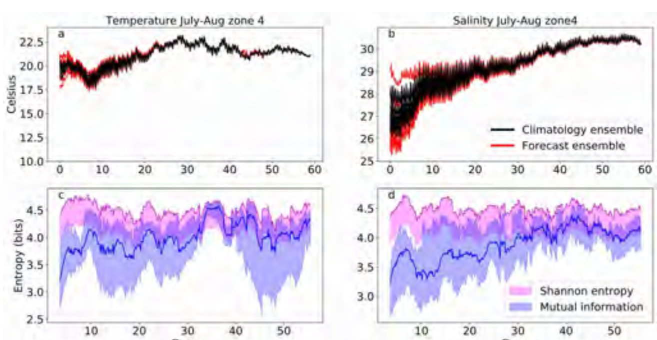

Figure S29. Results of zone 4 for July - August. Top figures shows temperature and salinity ensembles. Bottom figures show information entropy metrics applied between forecast and climatology ensembles.

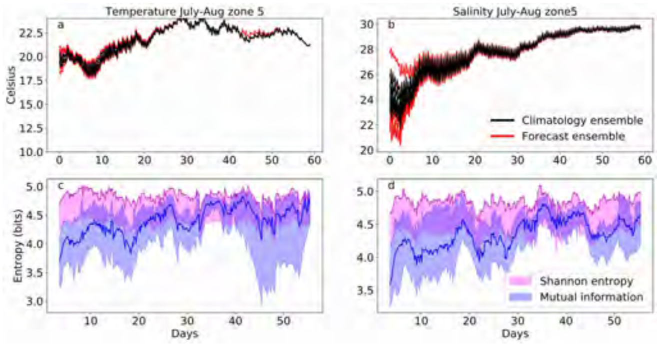

Figure S30. Results of zone 5 for July - August. Top figures shows temperature and salinity ensembles. Bottom figures show information entropy metrics applied between forecast and climatology ensembles.

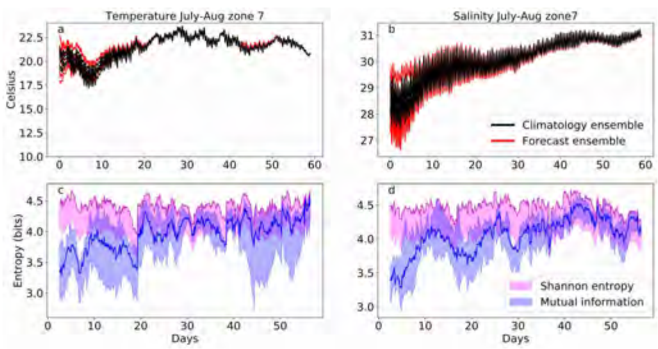

Figure S31. Results of zone 7 for July - August. Top figures shows temperature and salinity ensembles. Bottom figures show information entropy metrics applied between forecast and climatology ensembles.

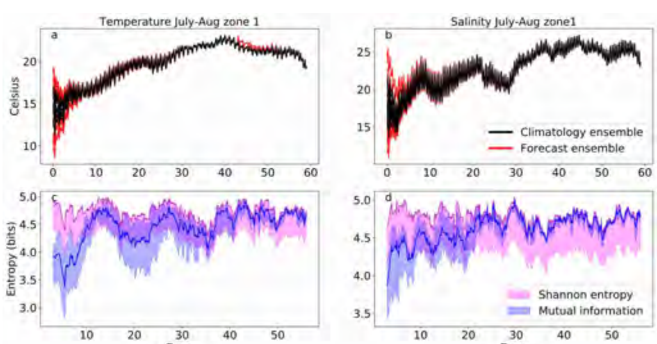

Figure S32. Results of zone 1 for July - August. Top figures shows temperature and salinity ensembles. Bottom figures show information entropy metrics applied between forecast and climatology ensembles.

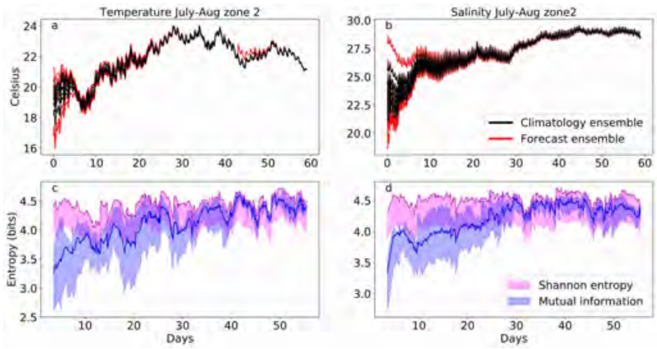

Figure S33. Results of zone 2 for July - August. Top figures shows temperature and salinity ensembles. Bottom figures show information entropy metrics applied between forecast and climatology ensembles.

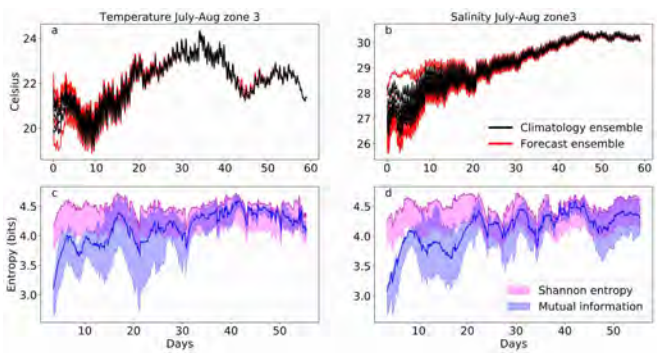

Figure S34. Results of zone 3 for July - August. Top figures shows temperature and salinity ensembles. Bottom figures show information entropy metrics applied between forecast and climatology ensembles.

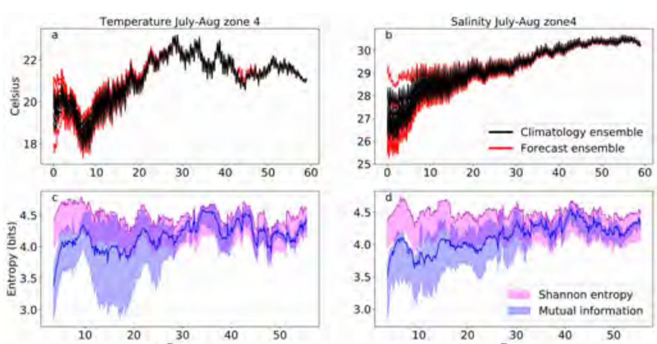

Figure S35. Results of zone 4 for July - August. Top figures shows temperature and salinity ensembles. Bottom figures show information entropy metrics applied between forecast and climatology ensembles.

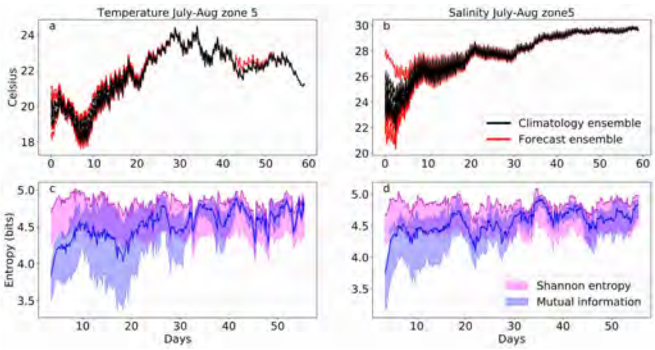

Figure S36. Results of zone 5 for July - August. Top figures shows temperature and salinity ensembles. Bottom figures show information entropy metrics applied between forecast and climatology ensembles.

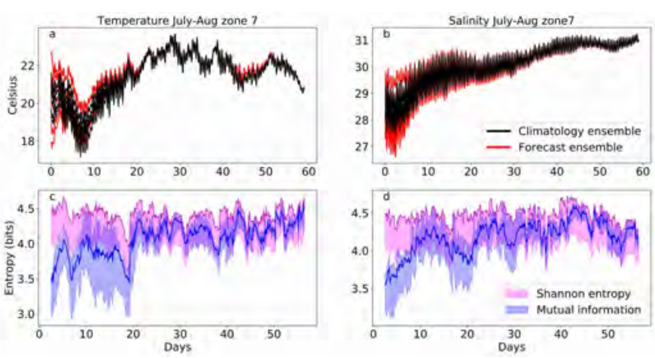

Figure S37. Results of zone 7 for July - August. Top figures shows temperature and salinity ensembles. Bottom figures show information entropy metrics applied between forecast and climatology ensembles.