# Predictability of North Atlantic Sea Surface Temperature and Upper-Ocean Heat Content

MARTHA W. BUCKLEY AND TIM DELSOLE

*Department of Atmospheric, Oceanic, and Earth Sciences, George Mason University, Fairfax, Virginia*

M. SUSAN LOZIER AND LAIFANG LI

*Earth and Ocean Sciences, Nicholas School of the Environment, Duke University, Durham, North Carolina*

## ABSTRACT

Understanding the extent to which Atlantic sea surface temperatures (SSTs) are predictable is important due to the strong climate impacts of Atlantic SST on Atlantic hurricanes and temperature and precipitation over adjacent landmasses. However, models differ substantially on the degree of predictability of Atlantic SST and upper-ocean heat content (UOHC). In this work, a lower bound on predictability time scales for SST and UOHC in the North Atlantic is estimated purely from gridded ocean observations using a measure of the decorrelation time scale based on the local autocorrelation. Decorrelation time scales for both wintertime SST and UOHC are longest in the subpolar gyre, with maximum time scales of about 4–6 years. Wintertime SST and UOHC generally have similar decorrelation time scales, except in regions with very deep mixed layers, such as the Labrador Sea, where time scales for UOHC are much larger. Spatial variations in the wintertime climatological mixed layer depth explain 51%–73% (range for three datasets analyzed) of the regional variations in decorrelation time scales for UOHC and 26%–40% (range for three datasets analyzed) of the regional variations in decorrelation time scales for wintertime SST in the extratropical North Atlantic. These results suggest that to leading order decorrelation time scales for UOHC are determined by the thermal memory of the ocean.

## 1. Introduction

The climate of the next few decades, particularly on regional scales, will depend on natural climate variations in addition to anthropogenic forcing (e.g., Hawkins and Sutton 2009; Corti et al. 2015; DelSole 2017). Therefore, unlike centennial climate predictions, in which the climate state is largely a response to anthropogenic forcing, decadal climate predictions require an accurate specification of the initial climate state and an adequate representation of natural variability in addition to knowledge of external forcings. Given the inherently chaotic nature

of the atmosphere, prospects for decadal climate prediction lie in the interaction of the atmosphere with slower parts of the climate system: the ocean, the cryosphere, and the geosphere. The impact of ocean initialization on decadal climate predictions, termed "initial-value predictability," has been intensely studied over recent years; in particular phases 5 and 6 of the Coupled Model Intercomparison Project (CMIP5 and CMIP6) include decadal prediction experiments, which were performed using numerous participating models (Doblas-Reyes et al. 2013; Meehl et al. 2014; Boer et al. 2016; Eyring et al. 2016).

The two main approaches for studying initial-value predictability are 1) initialized predictions and 2) statistical measures of predictability. In initialized predictions, such as the CMIP5/CMIP6 decadal prediction experiments, a coupled model initialized with ocean observations is used to produce an ensemble of forecasts. Improvements in predictability are inferred by comparing the skill of the ocean initialized forecasts with the skill from uninitialized forecasts. In the initialized forecasts, predictive skill can arise from both ocean initial conditions and external forcing. In the absence of an initialized ocean,

---

$\bigcirc$ Denotes content that is immediately available upon publication as open access.

$\mathscr{P}$ Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JCLI-D-18-0509.s1.

*Corresponding author*: Martha W. Buckley, marthabuckley@gmail.com

the skill comes entirely from external forcing. Note that the skill of a forecast relative to some reference (e.g., persistence or uninitialized simulations) is a quantity evaluated at a specific lead time. The predictability time scale is a metric of the time period over which a forecast has skill.

Initialized predictions suggest that the Atlantic is a region of enhanced predictability for sea surface temperature (SST; Smith et al. 2007; Keenlyside et al. 2008; Pohlmann et al. 2009; van Oldenborgh et al. 2012; Karspeck et al. 2015; Årthun et al. 2017) and upper-ocean heat content (UOHC; Yeager et al. 2012; Hermanson et al. 2014). Using decadal prediction experiments with the Community Climate System Model version 4 (CCSM4), Karspeck et al. (2015) show that the North Atlantic is the only region where ocean initialization leads to a statistically significant increase in skill of SST forecasts (at lead times of both 2–5 and 6–9 years) compared to an uninitialized model; a similar result was found by Doblas-Reyes et al. (2013) using the multi-model mean of CMIP5 models. A number of studies have suggested that predictability of North Atlantic SSTs and UOHC is related to variations in the Atlantic meridional overturning circulation (AMOC; Keenlyside et al. 2008; Pohlmann et al. 2009; Matei et al. 2012; Yeager et al. 2012; Hermanson et al. 2014). However, as detailed in Karspeck et al. (2015), successful decadal prediction in the North Atlantic may not rely on prediction of the AMOC, but rather on adequate initialization of the temperature and salinity fields (and hence geostrophic currents, including the AMOC).

Initialized predictions have also been used to show that the recent changes in UOHC in the subpolar gyre are predictable (Yeager et al. 2012; Hermanson et al. 2014). Some studies have suggested that these decadal changes in subpolar UOHC are related to variations in the AMOC (Robson et al. 2016) while others have focused on the importance of changes in the gyre circulation related to the overlying winds (e.g., Bersch 2002; Häkkinen and Rhines 2004; Sarafanov et al. 2008; Häkkinen et al. 2011; Piecuch et al. 2017), including shifts in the Gulf Stream path (Nigam et al. 2018; Ruiz-Barradas et al. 2018).

Statistical measures of initial value predictability can be estimated from long control runs of climate models with time-invariant external conditions (Branstator et al. 2012; DelSole et al. 2013; DelSole 2017).[1] Since these methods do not require model experiments, these

techniques facilitate the comparison of predictability properties among multiple models. Branstator et al. (2012) estimate initial value predictability for UOHC (defined as average temperature over the upper 300 m) in six atmosphere–ocean general circulation models; they find that there are large regional variations in predictability time scales in the North Atlantic, and the locations of highest predictability are different for each model. Predictability time scales in the subpolar North Atlantic range from less than 3 years to more than 15 years, while time scales in the subtropical North Atlantic range from less than 2 years to 10 years.

DelSole et al. (2013) use a similar multivariate regression model trained on climate model simulations, but they apply the model to observations (with external forcing removed) as well. They find that the North Atlantic basin is the region where the regression model has the highest skill for predicting SST, with significant skill for up to 6 years for both observations and the ensemble mean of preindustrial and twentieth-century simulations. However, the skill varies considerably with ensemble member: one ensemble member has no skill at any lead time and another has skill for up to 10 years. The spread in skill is primarily due to model differences, consistent with the results of Branstator et al. (2012).

While most predictability studies are based on models, a few studies have made statistical predictions based purely on observations (Alexander et al. 2008; Zanna 2012; Ho et al. 2013; Newman 2013; Huddart et al. 2017). One method for estimating observed predictability time scales for SST is to use a linear inverse model (LIM) derived from SST observations. Predictability for decadal SST variations from LIMs is generally 3–4 years over most of the North Atlantic domain, with predictability generally being larger in subpolar regions (Zanna et al. 2012; Newman 2013; Huddart et al. 2017). Furthermore, the skill of LIMs is in many cases found to be comparable to that of initialized models, suggesting that predictions from a LIM are a useful benchmark for interannual to decadal SST predictions (Newman 2013; Huddart et al. 2017). Similar results regarding predictability in the Atlantic basin have been found using other statistical methods, such as autoregressive models and constructed analog models (Ho et al. 2013). Statistical models can also be used to help understand dynamics behind predictability. For example, statistical models demonstrate that in the Atlantic basin coupling between the tropics and extratropics plays only a small role in setting predictability time scales (Huddart et al. 2017), in contrast to the Pacific basin where coupling between the tropics and extratropics plays a much larger role (Newman 2007; Alexander et al. 2008; Newman et al. 2016; Dias et al. 2019).

---

[1] DelSole and Tippett (2018) propose a generalized framework for predictability valid when there are variations in external forcing.

In this paper, we estimate predictability time scales of Atlantic SST and UOHC purely from ocean observations. A central question in this paper is whether spatial variations in predictability time scales can be explained by spatial variations in the ocean mixed layer depth (MLD). According to the null hypothesis of SST variability, the ocean mixed layer integrates stochastic atmospheric forcing, resulting in a red SST spectrum with a dominant time scale directly proportional to the MLD (Frankignoul and Hasselmann 1977). In this framework, the fact that the North Atlantic has longer predictability time scales than other basins may simply be explained by the deep MLDs in the North Atlantic. We consider a linear dependence of predictability time scales with MLD to be a simple null hypothesis to explain spatial variations in predictability time scales; we will evaluate this null hypothesis in our study.

In section 2 we introduce the diagnostics and datasets that we will use for SST and UOHC and define a measure of the decorrelation time scale, which we use to quantify predictability of SST and UOHC. We compute decorrelation time scales for UOHC and wintertime SST in section 3 and compare them quantitatively in section 4. In section 5, we determine to what extent spatial variations in decorrelation time scales can be explained by spatial variations in the wintertime climatological MLD. In section 6 we determine to what extent decaying versus oscillatory variations play a role in setting decorrelation time scales for UOHC and wintertime SST. Finally, in section 7 we conclude and discuss how our results relate to prior studies.

## 2. Methods

In this section we describe the diagnostics that we will use for SST and UOHC. Then, we introduce a measure of the decorrelation time scale, which we use to quantify predictability of SST and UOHC.

### a. Diagnostics for SST and upper-ocean heat content

Although SST is the ocean variable that most directly interacts with the atmosphere, predictability time scales for SST may underestimate ocean memory. SST memory is lost in the summertime due to the formation of the seasonal thermocline. In winter when the mixed layers deepen, anomalies isolated beneath the seasonal thermocline can become entrained in the mixed layer. This process, often referred to as the "reemergence mechanism" (Alexander and Deser 1995), can lead to enhanced predictability for wintertime SSTs (Namias and Born 1970; de Coëtlogon and Frankignoul 2003; Deser et al. 2003).

As our goal is to estimate predictability of SST and UOHC on interannual time scales, we define measures of SST and UOHC that minimize the effects of seasonal variations, specifically the loss of ocean memory in summer. For SST, we focus on wintertime SST ($\overline{SST}^w$), where wintertime is defined as the January–March average SST. We define our measure of UOHC as the average temperature in the layer between the surface and a depth $D$:

$$H \equiv \frac{1}{D} \int_{-D}^{0} \theta \, dz, \tag{1}$$

where $\theta$ is potential temperature. We define $D$ to be the wintertime climatological MLD, which is calculated by forming a monthly climatology of MLD and taking the January–March average of this quantity at each spatial location. An alternative choice is the maximum climatological MLD; our results are unchanged for this definition of $D$. The layer from the ocean surface to $-D$ reflects the portion of the ocean that comes in contact with the atmosphere seasonally, and $H$ covaries with SST on interannual time scales (Buckley et al. 2014, 2015). Averaging over this layer implicitly accounts for reemergence because anomalies that are isolated below the seasonal thermocline in summer are included in $H$ for all months. We express the ocean heat content in units of temperature rather than Joules for convenience, as this choice allows us to more easily compare time series of SST and $H$. However, our results do not depend on this choice because 1) $D$ is constant in time and 2) our predictability measure is based on the autocorrelation function, which does not depend on the magnitude of the anomalies.

Our analysis of predictability of SST and $H$ will utilize six different gridded observational products. As our analysis requires long time series for SST and $H$, and data products have substantial uncertainties, which are generally larger for earlier periods, we choose several data products to help understand the uncertainty. [See Kennedy (2014) for a review of uncertainty in SST data products and Boyer et al. (2016) for a review of uncertainty in ocean temperature products.] We use monthly SST from Extended Reconstructed SST version 5 (ERSST v5; Smith et al. 2008), Hadley Center SST (HadISST; Rayner et al. 2003), and COBE SST2 (Hirahara et al. 2014). We calculate $H$ from monthly fields of gridded temperature from Ishii (Ishii et al. 2006), EN4.2.1 (Good et al. 2013), and the ocean temperature data product from Cheng (Cheng et al. 2017). All analyses will be restricted from 1945 to present, and the specific periods used for each product are listed in Tables 1 and 2. Our results are unchanged if we instead restrict all products to the period of 1945–2012, which is common to all data products.

TABLE 1. Sensitivity of the decorrelation time scale $T_2$ to method and data product. Each of the three data products for $H$ is listed along the top row and left column, leading to a matrix comparing the data products. Shown is a comparison of $T_2$ for $\overline{H}$ for Ishii, EN4, and Cheng. All the data products were regridded on the same 1° grid, and $T_2$ is compared between products for all points in the ETNA. The $R^2$ values and the slopes of the linear fits are listed in the table, where the slope $m$ is calculated with the convention that the product listed along the top is $x$ and the product listed on the left is $y$. The diagonal of the matrix gives the sensitivity of each product to the use of $\overline{H}^w$ compared to the use of yearly averages $\overline{H}$. A slope of more than one indicates that $T_2$ is larger for $\overline{H}$ than $\overline{H}^w$.

| Product | Ishii | EN4 | Cheng |
|---|---|---|---|
| Ishii (1945–2012) | $R^2 = 0.84$, $m = 1.1$ | $R^2 = 0.80$, $m = 1.3$ | $R^2 = 0.65$, $m = 1.1$ |
| EN4 (1945–2017) | $R^2 = 0.80$, $m = 0.7$ | $R^2 = 0.80$, $m = 1.2$ | $R^2 = 0.68$, $m = 1.0$ |
| Cheng (1945–2016) | $R^2 = 0.65$, $m = 0.9$ | $R^2 = 0.68$, $m = 1.0$ | $R^2 = 0.78$, m = 1.1 |

The wintertime climatological MLD $D$ is an important parameter in our study, but the MLD is a diagnostic that cannot directly be measured. We use three different estimates of $D$, two based on the gridded data products (Ishii and EN4[2]) and one from the Argo MLD climatology of Holte et al. (2017), henceforth Argo MLD. In all cases the MLD is based on a threshold method in which the MLD is given by the depth at which the density exceeds a near-surface reference density by an amount $\Delta\rho$. For Ishii and EN4 (Figs. 1a,b), we use a fixed density threshold of $\Delta\rho = 0.125\,\mathrm{kg\,m^{-3}}$ applied to the monthly gridded observations (see Fig. 1 caption for details). The advantage of calculating MLDs from the gridded observational products is that $D$ is consistent with the temperature field of the gridded product. However, because the gridded products are spatially and temporally smoothed fields, we must apply a large density threshold (Toyoda et al. 2015). For the Argo MLD product (Fig. 1c) we use the MLD based on the variable density threshold of 0.2°C density equivalent applied to individual Argo profiles (see figure caption for details). The advantage of the Argo MLD product is that it is based directly on Argo profiles (Argo 2019) and can be considered a best estimate of the MLD based on data. The disadvantages are that it is restricted to the Argo period (from 2000 to the present) and the MLD field has missing points because Argo floats cannot travel in shallow regions.

In all cases $D$ is very deep in the Labrador Sea, Irminger Sea, and Iceland basin; relatively deep south of the Gulf Stream; and quite shallow elsewhere (Fig. 1). In the Labrador and Irminger Seas, $D$ is significantly larger for Ishii and EN4 (maximum about 1500 m; Figs. 1a,b) than for the Argo MLD product (maximum around 1000 m; Fig. 1c), which is consistent with the fact that a variable density threshold results in a smaller density criteria and therefore shallower MLDs in cold regions.

The value of $D$ in the Argo MLD product (Fig. 1c) is also more patchy, which is the result of calculating MLD directly from in situ profiles rather than from a highly smoothed monthly temperature field (as in Ishii and EN4).

Gridded temperature fields and $D$ are used to calculate $H$ according to Eq. (1). For Ishii and EN4, $D$ is based on the MLD calculated from the respective gridded product (e.g., Fig. 1a for Ishii and Fig. 1b for EN4); for Cheng we use $D$ based on the Argo MLD climatology (Fig. 1c). Both annual average values of $H$, henceforth $\overline{H}$, and wintertime average values of $H$, henceforth $\overline{H}^w$, are calculated from the monthly fields.

Previous work using an ocean state estimate demonstrates that SST and $H$ are highly correlated on interannual time scales (Buckley et al. 2014). We now show that a close correspondence between wintertime SST and $H$ is also seen in the Ishii gridded product. The Ishii product is chosen for this analysis since Ishii includes COBE SST2 observations, leading to a consistent product that can be used to compare SST and $H$. Figure 2 shows the squared correlation (at each grid point) between $\overline{\mathrm{SST}}^w$ and $\overline{H}^w$. The squared correlation between $\overline{\mathrm{SST}}^w$ and $\overline{H}^w$ is generally quite high; the mean value of the squared correlation over the North Atlantic domain is 0.83. Correlations are lowest in the Labrador Sea and east of Iceland. Other regions of relatively low correlations include the eastern subpolar gyre and isolated regions in the North Atlantic and South Atlantic subtropical gyres (e.g., the region south of the Grand Banks). Correlations are very high in the tropics. Low correlations between $\overline{\mathrm{SST}}^w$ and $\overline{H}^w$ generally occur where $D$ is deep.

### b. Measure of predictability time scales

To estimate the predictability time scales for wintertime SST and $H$, we use a measure of the decorrelation time scales, as defined by DelSole (2001):

$$T_2 = \sum_{k=-\infty}^{+\infty} \rho_k^2 = 1 + 2\sum_{k=1}^{\infty} \rho_k^2, \qquad (2)$$

where $\rho_k$ is the autocorrelation (for SST or $H$) at discrete lag $k$; when considering yearly/wintertime average data,

---

[2] As the Cheng product does not include salinity, we cannot calculate potential density from this product and thus cannot calculate MLDs.

TABLE 2. As in Table 1, but for a comparison of $T_2$ for $\overline{SST}^w$ between ERSST, HadISST, and COBE SST2.

| Product | ERSST | HadISST | COBE SST2 |
|---|---|---|---|
| ERSST (1945–2017) | | $R^2 = 0.61, m = 1.1$ | $R^2 = 0.79, m = 0.6$ |
| HadISST (1945–2017) | $R^2 = 0.61, m = 0.9$ | | $R^2 = 0.60, m = 0.5$ |
| COBE SST2 (1945–2016) | $R^2 = 0.79, m = 1.7$ | $R^2 = 0.60, m = 2.0$ | |

as is done here, $k$ is the number of years. For an exponentially decaying autocorrelation function, $T_2$ is equal to the familiar $e$-folding time scale. The time scale $T_2$ is chosen in contrast to other possible measures of the decorrelation time scale (e.g., Schubert et al. 1992; DelSole 2001) because it accurately estimates predictability in the presence of oscillatory variability. As shown in DelSole (2001), $T_2$ measures fundamental aspects of time series and arises naturally in statistical sampling theory.

In this paper, $T_2$ will be interpreted as a lower bound on the predictability time scale. This interpretation is motivated by the fact that autocorrelations out to lag $T_2$ tend to be significant, and therefore a linear regression model can make skillful predictions at least as long as $T_2$. It is a lower bound because it is based purely on the local autocorrelation function. Including additional predictors, such as nonlocal SST or variations in the AMOC, may improve skill (and thus correspond to longer predictability time scales). For instance, a LIM that captures covariability between multiple spatial patterns may have more skill at individual points than a regression model fitted to data at each point. Thus, our diagnostic $T_2$ is best understood as a lower bound on predictability, one which may be a useful benchmark against which to evaluate dynamical models or more sophisticated statistical measures of predictability [such a comparison is made in the framework of atmospheric predictability in Schubert et al. (1992)]. For these reasons, we henceforth refer to $T_2$ as the decorrelation time scale.

If we know the exact value of our autocorrelation function, we can integrate over all lags to find $T_2$, but for a finite length time series we can only calculate the sample autocorrelation function. For a sufficiently long time series, we can instead integrate from lag $-k_*$ to $k_*$, as long as we can chose a value of $k_*$ that is much longer than the time scale that we are trying to estimate but much shorter than the length of the time series. For the observational records that we are using, which are on the order of 70 years, it is not possible to find such a value of $k_*$. To resolve this dilemma, we fit an autoregressive (AR) model to the time series of $\overline{SST}^w$, $\overline{H}$, and $\overline{H}^w$. The AR parameters are then used to calculate the theoretical autocorrelation function, $\widetilde{\rho}_k$ at discrete lags $k$. Our estimate of $T_2$ is found by replacing $\rho_k$ with $\widetilde{\rho}_k$ in Eq. (2).

We try AR orders between 1 and 3 and find little sensitivity to AR order, particularly for AR orders greater than 1, so we choose an AR order of 2. Prior to computing the AR fits we remove a quadratic function (in time) from the time series of SST and $H$ at each spatial location. As a result of both the limited data period available and the removal of the longest time scales (due to removal of a quadratic), the resulting decorrelation time scales may underestimate predictability of SST and $H$.

## 3. Decorrelation time scales for upper-ocean heat content and SST

In this section we calculate decorrelation time scales for $H$ and wintertime SST based on gridded observational products. Decorrelation time scales for $\overline{H}$ based on Ishii, EN4, and Cheng are shown in Fig. 3 (showing the extratropical North Atlantic) and Fig. S1 (showing the entire Atlantic domain; see the online supplemental material). The largest decorrelation time scales, around 4–6 years, are seen in the subpolar gyre, in particular in the Labrador Sea, Irminger Sea, and Iceland basin. Decorrelation time scales are around two years south of the Gulf Stream path, but elsewhere in the subtropical gyres they are about a year. The black contours in Fig. 3 and in Fig. S1 show contours of $D$, and it is apparent that the largest decorrelation time scales are found where $D$ is deep.

The black dots in Fig. 3 and Fig. S1 show grid points where the decorrelation time scale is not significantly different from white noise on interannual time scales (see the appendix for details). In these regions there is no interannual predictability, although there may be predictability on shorter time scales (which could be examined by looking at monthly $H$ anomalies). One can see that the interannual predictability of $H$ is mostly restricted to the extratropics. Therefore, for the rest of our analysis we will consider only the extratropical North Atlantic (ETNA; north of 20°N).

Results shown in Fig. 3 and Fig. S1 are from yearly averages for $H$ ($\overline{H}$) but results using wintertime averages for $H$ ($\overline{H}^w$) are quite similar (Table 1). Yearly average values of $H$ include temperature anomalies in the summertime seasonal thermocline, which are likely unrelated to the deeper (and presumably more predictable) temperature
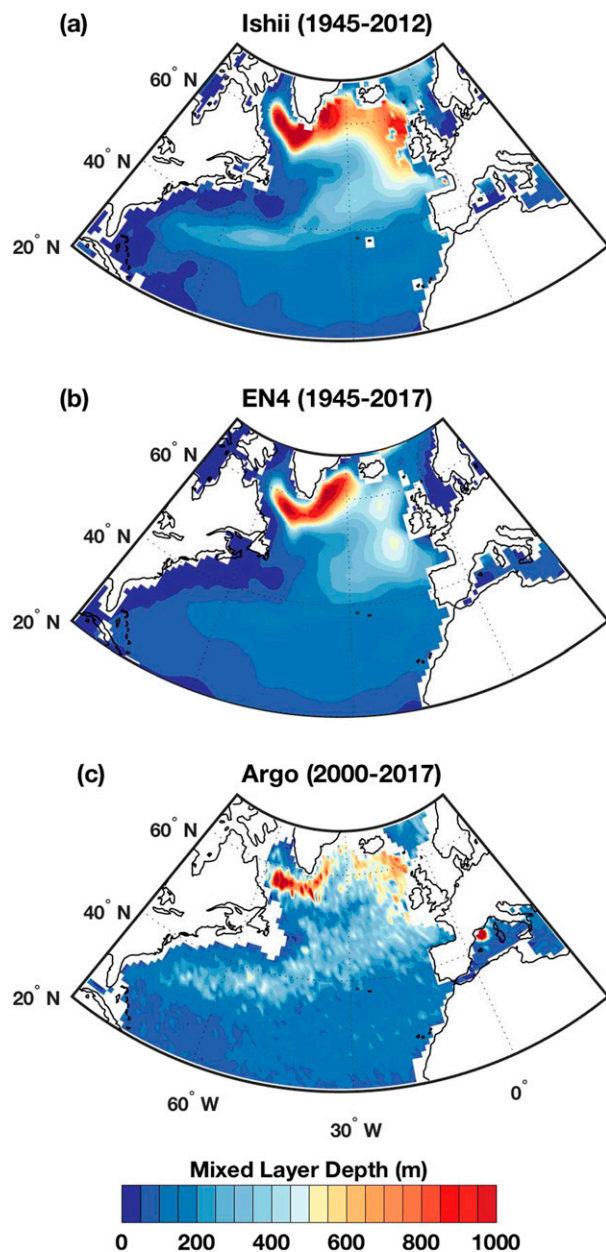
FIG. 1. The wintertime (January–March) climatological MLD $D$ from (a) Ishii, (b) EN4.2.1, and (c) the Argo MLD climatology (Holte et al. 2017). In all cases the MLD is based on a threshold method in which the MLD is given by the depth at which the density exceeds the density at a near-surface reference depth $z_o$ by an amount $\Delta\rho$. If there is no depth level where the threshold is exceeded, then the MLD is set to the ocean depth. For Ishii and EN4, we calculate the potential density from the potential temperature and absolute salinity according to the TEOS-5 algorithms. The MLDs are calculated by applying a density threshold of $\Delta\rho = 0.125\,\mathrm{kg\,m^{-3}}$ to the gridded monthly data; the reference depth $z_o$ is the cell center of the topmost grid cell (located at about 5 m). Then, a monthly climatology of MLD is calculated, and $D$ is given by the January–March average of this field at each grid point. For the Argo MLD product we use the MLD based on the variable density threshold of 0.2°C density equivalent applied to individual

anomalies formed during the winter. Therefore, consideration of only wintertime $H$ anomalies could potentially lead to increased decorrelation time scales, but in practice decorrelation time scales based on yearly average $H$ tend to be somewhat larger (see Table 1), simply as a result of suppressing noise by averaging over more months. The results from the three different data products are quite consistent. If we regrid all estimates of $T_2$ onto a common 1° grid and compute linear fits for all points in the ETNA between the various data products, we find $R^2$ values between 0.65 and 0.80 and slopes between 0.7 and 1.3 (see Table 1).

Now we present results showing the decorrelation time scales for $\overline{\mathrm{SST}}^w$ based on ERSST, HadISST, and COBE SST2 (Fig. 4). The largest decorrelation time scales, around 3–5 years, are seen in the subpolar gyre. Decorrelation time scales are around two years south of the Gulf Stream path, but elsewhere in the subtropical gyre they are about a year. The spatial pattern of decorrelation time scales is generally similar for $\overline{\mathrm{SST}}^w$ and $\overline{H}$, except for in the Labrador Sea where decorrelation time scales for $\overline{\mathrm{SST}}^w$ are much shorter than those for $\overline{H}$. The black dots in Fig. 4 show grid points where the decorrelation time scale is not significantly different from white noise on interannual time scales (see the appendix for details). In these regions there is no interannual SST predictability, although there may be predictability on shorter time scales (which could be examined by looking at monthly SST anomalies). One can see that the interannual predictability of $\overline{\mathrm{SST}}^w$ is mostly restricted to the subpolar gyre.

The results from the three different data products are quite consistent. If we regrid all estimates for $T_2$ onto a common 1° grid and compute linear fits for all points in the ETNA between the various data products, we find $R^2$ values between 0.60 and 0.79 and slopes between 0.5 and 2.0 (see Table 2). The deviation of the slopes from one is due to COBE SST2 having significantly larger decorrelation time scales than the other two products.

←

Argo profiles; the reference depth $z_o = 10\,\mathrm{m}$. In this case the density threshold changes based on the local reference temperature and salinity: the threshold is $0.03\,\mathrm{kg\,m^{-3}}$ for a reference temperature of 8°C and salinity of 35 psu, but the density threshold is smaller for colder reference temperatures. Additionally, we use the gridded product in which the monthly MLD climatology at each grid point is given by the mean of all profiles found in this month in the grid box. The value of $D$ is calculated by simply taking the maximum of the Argo-based MLD climatology at each grid point. A simple maximum is used rather than the January–March average to maximize the points where $D$ can be calculated due to data availability.
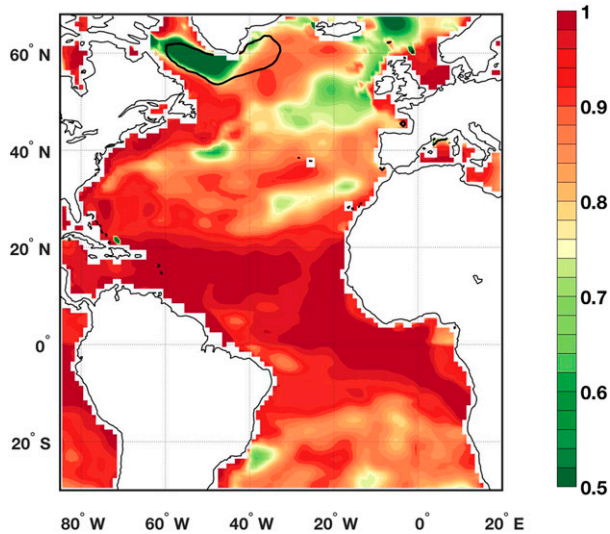
FIG. 2. Squared correlation at each grid point between $\overline{SST}^w$ and $\overline{H}^w$ for the Ishii data product. The heavy black contour (over the Labrador Sea) shows the region where the standard deviation of the wintertime (January–March average) MLD exceeds 250 m.

## 4. Comparing decorrelation time scales for upper-ocean heat content and SST

We now explore the relationship between decorrelation time scales for SST and $H$ using the Ishii data product. To present results that are most comparable, wintertime values are used for both SST and $H$, but recall that decorrelation time scales for $\overline{H}^w$ are very similar (but slightly smaller) than those for $\overline{H}$ (Table 1). Figure 5a shows a scatterplot comparing $T_2$ for $\overline{SST}^w$ to $T_2$ for $\overline{H}^w$ for all points in the ETNA. A linear fit explains 61% of the variance, and the slope of the line is one, indicating that $T_2$ for $\overline{SST}^w$ is comparable to $T_2$ for $\overline{H}^w$ for most points in the ETNA.

Now we explore the points where the linear relationship between $T_2$ for $\overline{SST}^w$ and $\overline{H}^w$ does not hold. The triangles (squares) in Fig. 5a are points that are more than two standard deviations above (below) the best-fit line. There are significantly more triangles than squares, indicating that there are many more outliers where $T_2$ is larger for $\overline{H}^w$ than for $\overline{SST}^w$. The points in Fig. 5a are shaded (grayscale) to reflect their value of $D$, and it is apparent that points where $T_2$ for $\overline{H}^w$ is significantly larger than $T_2$ for $\overline{SST}^w$ are generally located where $D$ is deep. This is what one would expect, as averaging from the surface to a depth $D$ is a more effective low-pass filter when $D$ is deep.

Figure 5b shows the spatial distribution of the outliers from the linear fit. All of the outliers are in the subpolar gyre, and the outliers where $T_2$ is larger for $\overline{H}^w$ than for $\overline{SST}^w$ are concentrated in the Labrador Sea and in the
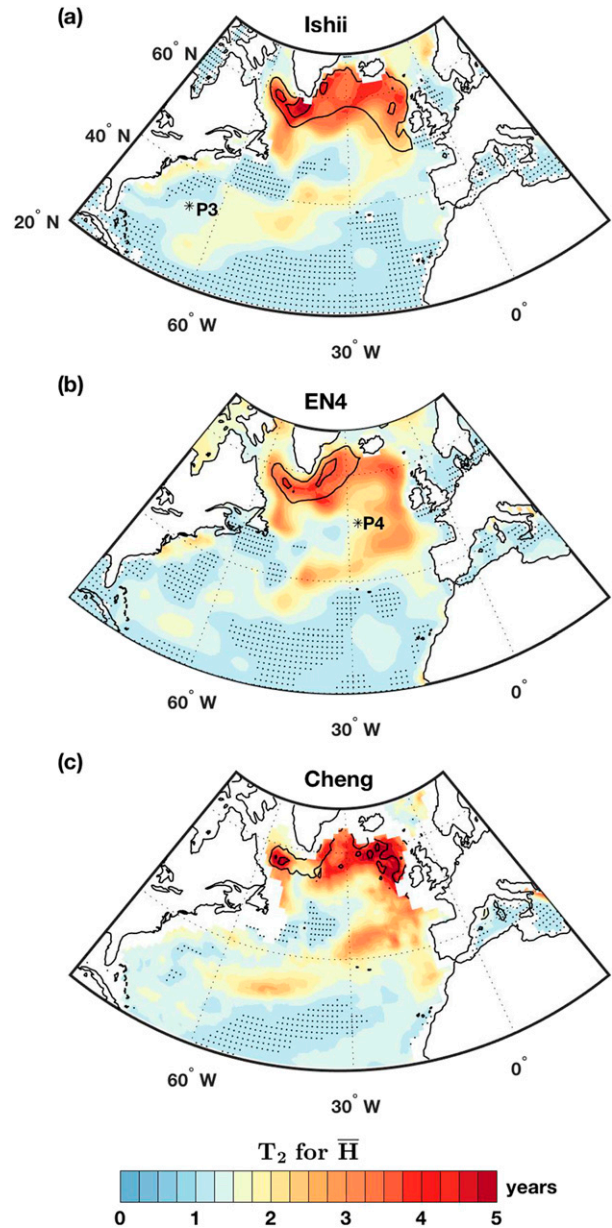


FIG. 3. The decorrelation time scale $T_2$ (colors) for $\overline{H}$ based on (a) Ishii, (b) EN4, and (c) the Cheng OHC product. The black dots show the points where the decorrelation time scale does not exceed the time scale predicted from white noise on interannual time scales at the 90% confidence level; see the appendix for details. The black contours show the wintertime climatological MLD $D$ at levels of 500 and 1000 m. For Ishii and EN4, $D$ is based on a density criterion applied to the gridded observations. For Cheng, $D$ is based on an Argo MLD climatology (Holte et al. 2017). Starred points P3 in (a) and P4 in (b) correspond to points P3 and P4 shown in Figs. 9a and 9b.

eastern subpolar gyre off the coast of the British Isles. These points, particularly those in the Labrador Sea, align with the regions where the squared correlation between $\overline{SST}^w$ and $\overline{H}^w$ is low (see Fig. 2). To
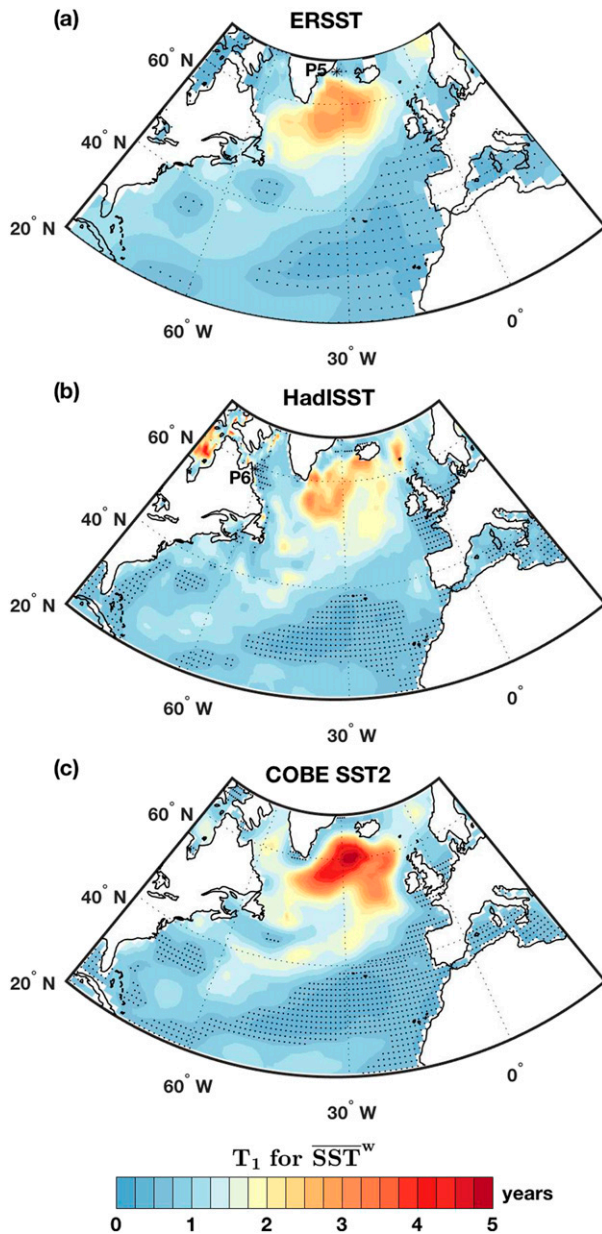
FIG. 4. The decorrelation time scale $T_2$ for $\overline{SST}^w$ based on (a) ERSST, (b) HadISST, and (c) COBE SST2. The black dots show the points where the decorrelation time scale does not exceed the time scale predicted from white noise on interannual time scales at the 90% confidence level; see the appendix for details. Starred points P5 in (a) and P6 in (b) correspond to points P5 and P6 shown in Figs. 9c and 9d.

understand why $T_2$ is so much larger for $\overline{H}^w$ than for $\overline{SST}^w$ in these regions, we choose two characteristic points, one in the Labrador Sea and one in the eastern subpolar gyre; these points are labeled on both the scatterplot (Fig. 5a) and the spatial map (Fig. 5b). The time series of $\overline{SST}^w$ and $\overline{H}^w$ at these points are shown in Figs. 5c and 5d. It is apparent that depth averaging is an

effective low-pass filter, which leads to the enhanced predictability for $\overline{H}^w$ compared to $\overline{SST}^w$ in these regions.

Although differences between decorrelation time scales for $\overline{H}^w$ and $\overline{SST}^w$ are expected in regions of deep $D$, the Labrador Sea remains a unique region in that it has such low predictability time scales for $\overline{SST}^w$, generally less than two years, despite very persistent anomalies in $H$. There are several factors that may explain this.

1) Large spatial gradients in MLDs: Regions of deep convection are characterized by weak vertical gradients and deep MLDs, but there are strong lateral gradients in properties. These lateral gradients are baroclinically unstable, leading to a rapid growth of baroclinic eddies that act to restratify the near-surface layer (Send and Marshall 1995; Jones and Marshall 1997; Marshall and Schott 1999; Frajka-Williams et al. 2014). The input of buoyant waters from boundary currents adjacent to regions of deep convection is also important in this restratification process (Straneo 2006; Schmidt and Send 2007). Time scales for restratification in the Labrador Sea are thought to be rapid (from weeks to months). While air–sea heat fluxes are quite large over the Labrador Sea and may play a role in restratification, near surface restratification occurs more rapidly than can be accounted for by air–sea heat fluxes alone, indicating the importance of lateral mixing (Send and Marshall 1995; Jones and Marshall 1997). Wintertime SST anomalies are rapidly destroyed by restratification, but since the restratification is restricted to the near-surface layers (top few hundred meters at most), anomalies in $\overline{H}^w$ remain mostly intact. Therefore, restratification processes may explain the much longer decorrelation time scales of $\overline{H}^w$ compared to that of $\overline{SST}^w$.

2) Large temporal variability in MLD: Figure 2 shows that low correlations between $\overline{SST}^w$ and $\overline{H}^w$ generally align with regions where MLDs vary strongly from one winter to the next. Temporal variability of wintertime MLDs also may lead to decorrelation time scales for $\overline{H}^w$ that are much larger than for $\overline{SST}^w$. If in a given winter the MLD is shallower than $D$, $H$ will include thermal anomalies from the previous winter, stored below this year's wintertime mixed layer. This effect will result in decorrelation time scales for $\overline{H}^w$ that overestimate the true upper ocean predictability (which would be seen if we allowed $D$ to vary from one year to the next) and are significantly larger than decorrelation time scales for $\overline{SST}^w$. We do not expect any comparable impact on decorrelation time scales for $\overline{H}^w$ during years when the wintertime MLD is deeper than $D$, as in
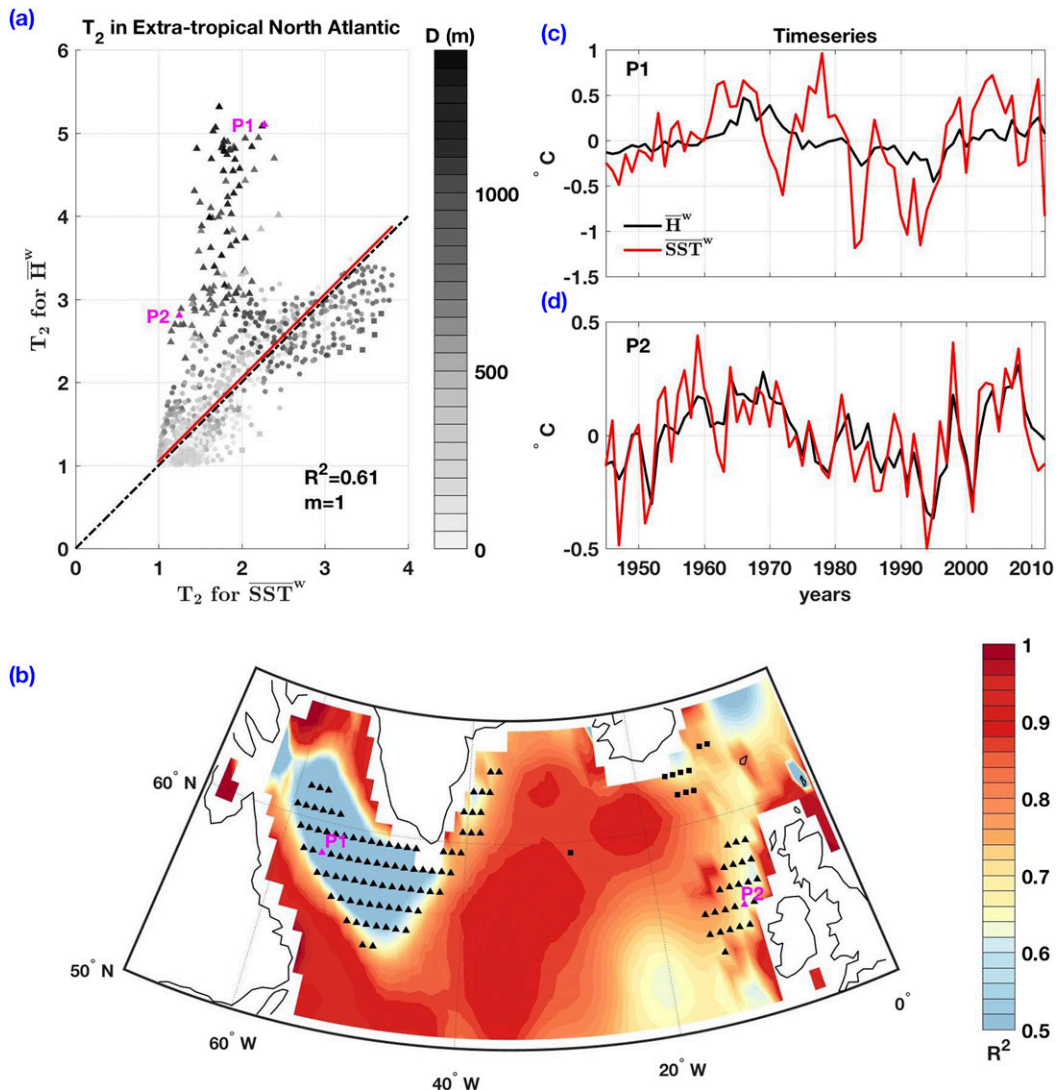
FIG. 5. A comparison of the decorrelation time scale $T_2$ for $\overline{SST}^w$ and $\overline{H}^w$ using the Ishii data product. (a) A scatterplot showing the relationship between $T_2$ for $\overline{SST}^w$ and $T_2$ for $\overline{H}^w$ for all points in the ETNA. The points are shaded (grayscale) to reflect their value of $D$. The solid red line shows the linear fit, and the black dash-dotted line is the one-to-one line. The slope and $R^2$ values for the fit are given in the bottom-right corner. The circles are points that are within the two standard deviations of the best-fit line, and the triangles (squares) are points that are more than two standard deviations above (below) the best-fit line. (b) The spatial distribution of the outliers, where black triangles (squares) indicate points that are more than two standard deviations above (below) the best-fit line. Colors show the squared correlation (at each grid point) between $\overline{H}^w$ and $\overline{SST}^w$. Two points of interest, P1 and P2, are chosen and the locations of these points are shown in both the scatterplot in (a) and the spatial map in (b). (c),(d) The time series of $\overline{SST}^w$ (red) and $\overline{H}^w$ (black) at points P1 and P2, respectively.

this case $\overline{H}^w$ anomalies will still reflect thermal anomalies in the mixed layer.

In summary, generally decorrelation time scales for wintertime SST and $H$ are similar, but in the Labrador Sea decorrelation time scales for SST are much shorter. This is likely due to large spatial and temporal variations in wintertime MLDs and modification of SST by rapid near-surface processes. However, the gridded observational datasets utilized here lack the temporal and spatial resolution to fully understand the dynamics operating in the Labrador Sea.

## 5. Can mixed layer depth variations explain spatial variations in decorrelation time scales?

In this section we explore our null hypothesis that spatial variations in decorrelation time scales are related
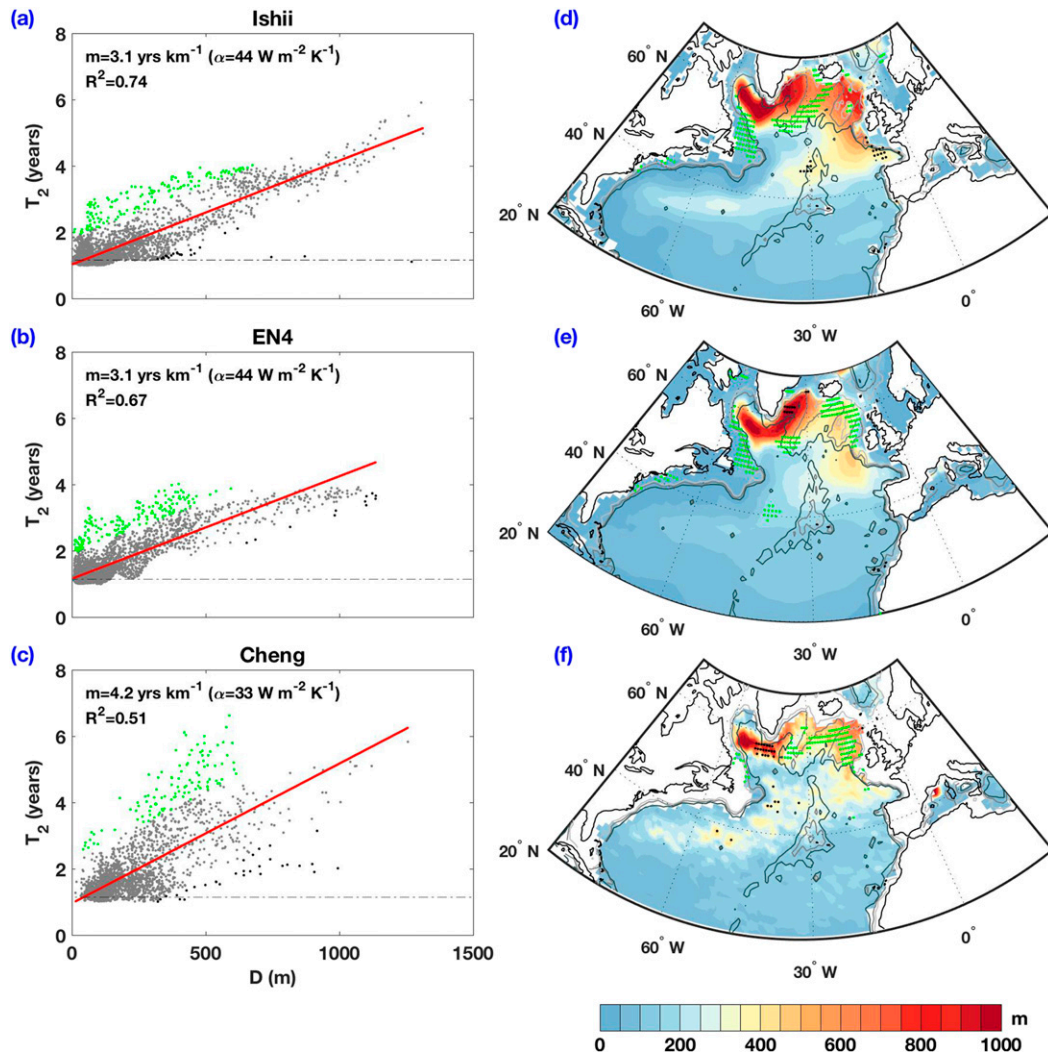
FIG. 6. How decorrelation time scales for $H$ are related to $D$: (left) Scatterplots comparing the decorrelation time scale $T_2$ for $\overline{H}$ to $D$ for all points in the ETNA for (a) Ishii, (b) EN4, and (c) Cheng. The dash-dotted gray lines show values of $T_2$ above which $T_2$ can be distinguished from white noise on interannual time scales (see the appendix). The red line is a linear fit between $T_2$ and $D$, and the values of $R^2$ and the slope $m$ and the corresponding value of the damping parameter $\alpha$ [see Eq. (3)] are given in the top-left corner of each plot. Green (black) points are outliers that are more than two standard deviations above (below) the best-fit line. (right) Spatial distribution of the outliers based on the linear fit for (d) Ishii, (e) EN4, and (f) Cheng. Green (black) points are points more than two standard deviations above (below) the best-fit line, indicating points with higher (lower) than expected predictability. Colors show a spatial map of the wintertime climatological MLD $D$, and light gray, gray, and dark gray contours show ocean bathymetry at depths of 1, 2, and 3 km, respectively.

to variations in the wintertime climatological MLD $D$. We first consider the linear relationship between the decorrelation time scale for $H$ and $D$ and then consider the relationship between the decorrelation time scale for wintertime SST and $D$. Then, we use the slopes of the linear relationships to estimate the damping parameter.

### a. Relationship between decorrelation time scale for $H$ and $D$

Figures 6a–c show scatterplots between $D$ and $T_2$ for all points in the ETNA based on $\overline{H}$ from the Ishii, EN4,

and Cheng data products. The red lines show a linear fit between $T_2$ and $D$; linear fits demonstrate that between 51% and 74% of the spatial variations in $T_2$ can be explained by spatial variations of $D$. The strong linear relationship between $D$ and $T_2$ matches the expectation from stochastic climate models in which the ocean simply thermodynamically integrates stochastic atmospheric forcing (Frankignoul and Hasselmann 1977). The slopes of the lines indicate that an increase in $D$ by 1 km results in an increase in the decorrelation time scale of 3.1–4.2 years. The precise values of $R^2$ and the
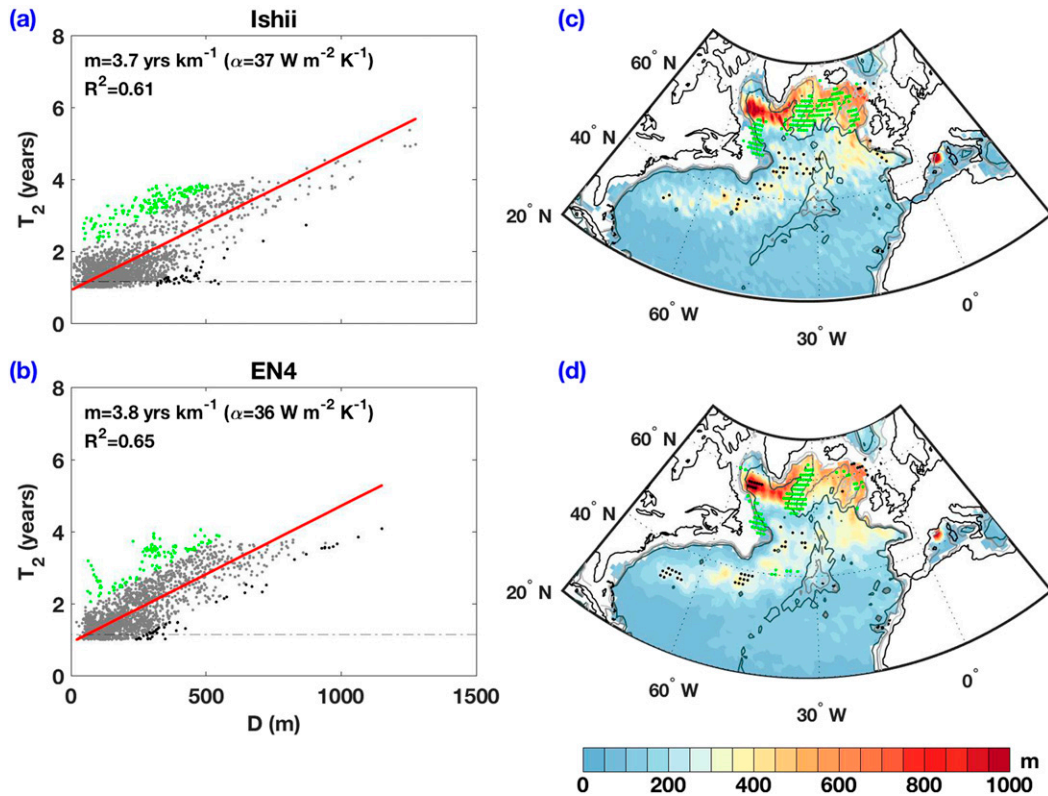
FIG. 7. Sensitivity of relationship between $T_2$ for $H$ and $D$ to the MLD product used: (a),(b) As in Figs. 6a and 6b and (c),(d) as in Figs. 6d and 6e, but $D$ is from the Argo MLD climatology (Holte et al. 2017). This value of $D$ is used for both the values of $D$ in the scatterplot and in the calculation of $H$ according to Eq. (1).

slope of the linear regression are somewhat sensitive to the MLD product. If the calculations for Ishii and EN4 are repeated using $D$ from the Argo MLD climatology (Holte et al. 2017), $R^2$ decreases (Figs. 7a,b), likely because the Argo MLD product is more patchy. Furthermore, since $D$ from the Argo MLD product is shallower than $D$ calculated from Ishii and EN4, the slope of the line increases (Figs. 7a,b) and becomes more similar to that from Cheng (which uses $D$ from the Argo MLD climatology).

In calculating the linear fits, all points in the ETNA are included, regardless of whether or not the values of $T_2$ can be distinguished from white noise on interannual time scales (dash-dotted gray lines in Figs. 6a–c, 7a, and 7b show values of $T_2$ above which $T_2$ can be distinguished from white noise on interannual time scales). All points are included since points where $D$ is shallow and there is no interannual predictability are part of the observed relationship between $T_2$ and $D$; in fact, if there were places with large $D$ where there was no interannual predictability, these points would not match the expectation from the null hypothesis and would require explanation.

Now we explore the points that do not follow the scaling of $T_2$ with $D$. The green (black) points in Figs. 6a–c, 7a, and 7b indicate outliers that are more than two standard deviations above (below) the best-fit line. There are significantly more outliers above the line than below the line, indicating that there are more points with higher than expected predictability than lower than expected predictability. Figures 6d–f show the spatial distribution of the outliers; the spatial distribution of the outliers is quite similar when the calculations for Ishii and EN4 are repeated using $D$ from the Argo MLD climatology (Figs. 7c,d). The outliers are mostly confined to the subpolar gyre, except for a few points with higher than expected predictability in the slope waters west of the Gulf Stream. Points with higher than expected predictability are almost all confined to shallow bathymetry: the North American continental shelf, the Grand Banks, and the Reykjanes Ridge, a concurrence that needs further investigation, particularly in the context of previous work suggesting that interactions of ocean currents with bathymetry are an important process in generating upper ocean temperature variability (e.g., Nigam et al. 2018). There are very few points with

lower than expected predictability, and their locations are not robust between the different data products, so it is somewhat difficult to determine their origin. There does seem to be some tendency for points with lower than expected predictability to occur in regions with large spatial and temporal gradients in $D$. This may be due to rapid restratification as a result of lateral mixing by eddies, which is expected to lead to decorrelation time scales shorter than those predicted by local values of $D$ (Send and Marshall 1995; Jones and Marshall 1997; Marshall and Schott 1999; Frajka-Williams et al. 2014).

### b. Relationship between decorrelation time scale for wintertime SST and D

We now consider the relationship between variations in $D$ and $T_2$ for wintertime SST; in considering this relationship we use $D$ based on the Argo MLD climatology for all three SST data products. Figures 8a–c show a scatterplot between $D$ and $T_2$ for $\overline{SST}^w$ for all points in the ETNA. (As before fits are for all points in the ETNA; gray dash-dotted lines show the value of $T_2$ above which $T_2$ can be distinguished from white noise on interannual time scales.) The linear fit between $D$ and $T_2$ indicates that an increase in $D$ by 1 km results in an increase in the decorrelation time scale of 1.3–2.5 years. As expected, the slope of the line is significantly shallower than that for $\overline{H}$ because predictability time scales for $\overline{SST}^w$ are generally less than those for $\overline{H}$. We find that 26%–40% of the spatial variations of $T_2$ for $\overline{SST}^w$ can be explained by spatial variations in $D$. There are several reasons for the smaller percent variance of $T_2$ explained by $D$ for $\overline{SST}^w$ compared to $\overline{H}$:

1) Predictability time scales for $H$ are expected to be proportional to $D$ according to stochastic climate models (e.g., Frankignoul and Hasselmann 1977; Deser et al. 2003; de Coëtlogon and Frankignoul 2003), but there is not the same expectation for SST, which is also influenced by surface processes, as discussed in section 4.
2) SST is a noisier field than $H$, as depth averaging acts as a low-pass filter (see section 4). Additionally, for $H$ we are able to further suppress noise by averaging over a year rather than just averaging over the wintertime months.
3) For the fit between $T_2$ and $D$ for the SST analysis, we use $D$ from the Argo MLD climatology, which is relatively patchy. Our analysis with Ishii and EN4 shows that the portion of the variance of $T_2$ explained by $D$ decreases somewhat when $D$ from the Argo MLD product is used compared to the smoother values of $D$ based on the gridded data products (cf. Figs. 6a,b and 7a,b).

4) In the Labrador Sea, predictability for $\overline{SST}^w$ is quite small, despite very large values of $D$; in contrast predictability for $\overline{H}$ is quite large, as discussed in section 4. If the Labrador Sea is excluded from our fit between $T_2$ and $D$, the percent variance explained increases by 6%–10% and the slopes increase by about $0.5\,\mathrm{yr\,km}^{-1}$ (due to the removal of low values of $T_2$ with large values of $D$).

The points above suggest that decorrelation time scales for $\overline{SST}^w$ exhibit a weaker scaling with $D$ than the time scales for $H$ due to both complex surface dynamics (particularly in the Labrador Sea) and noise (in both the SST fields and the MLD field).

Now we explore the points that do not follow the scaling of decorrelation time scale with $D$. The green (black) points in Figs. 8a–c indicate outliers that are more than two standard deviations above (below) the best fit line. There are significantly more outliers above the line than below the line, indicating that there are more points with higher than expected predictability than lower than expected predictability. Figures 8d–f show the spatial distribution of the outliers. The outliers almost all have higher than expected predictability and are located in the region with large decorrelation time scales south of Greenland. There is a small region of points with lower than expected predictability in the Labrador Sea, consistent with the fact that while $D$ is deep and decorrelation time scales for $\overline{H}$ and $\overline{H}^w$ are long, decorrelation time scales for $\overline{SST}^w$ are short in this region (see also Fig. 5).

### c. The damping parameter

The slopes of the fitted lines between $T_2$ (for $\overline{H}$ and $\overline{SST}^w$) and $D$ can be written in the units of a damping parameter $\alpha$. From Frankignoul et al. (1998), a characteristic time scale can be related to $\alpha$ as

$$\tau = \frac{\rho_o C_p D}{\alpha}, \qquad (3)$$

where $\rho_o = 1035\,\mathrm{kg\,m}^{-3}$ is a reference density and $C_p$ is the heat capacity. Inverting this relationship and setting $\tau/D = m$, where $m$ is the slope of our fit between $T_2$ and $D$, we can make a gross estimate of $\alpha$. The values of $\alpha$ calculated by this method can be understood as the best-fit damping parameter, under the assumption that the damping parameter is constant over the entire domain. Using this relation results in values of $\alpha$ ranging from 33 to $44\,\mathrm{W\,m}^{-2}\,\mathrm{K}^{-1}$ for $\overline{H}$ (Figs. 6a–c). If the calculations for Ishii and EN4 are repeated using $D$ from the Argo MLD climatology, $\alpha$ for Ishii and EN4 decreases to $36$–$37\,\mathrm{W\,m}^{-2}\,\mathrm{K}^{-1}$ (Figs. 7a,b), values more
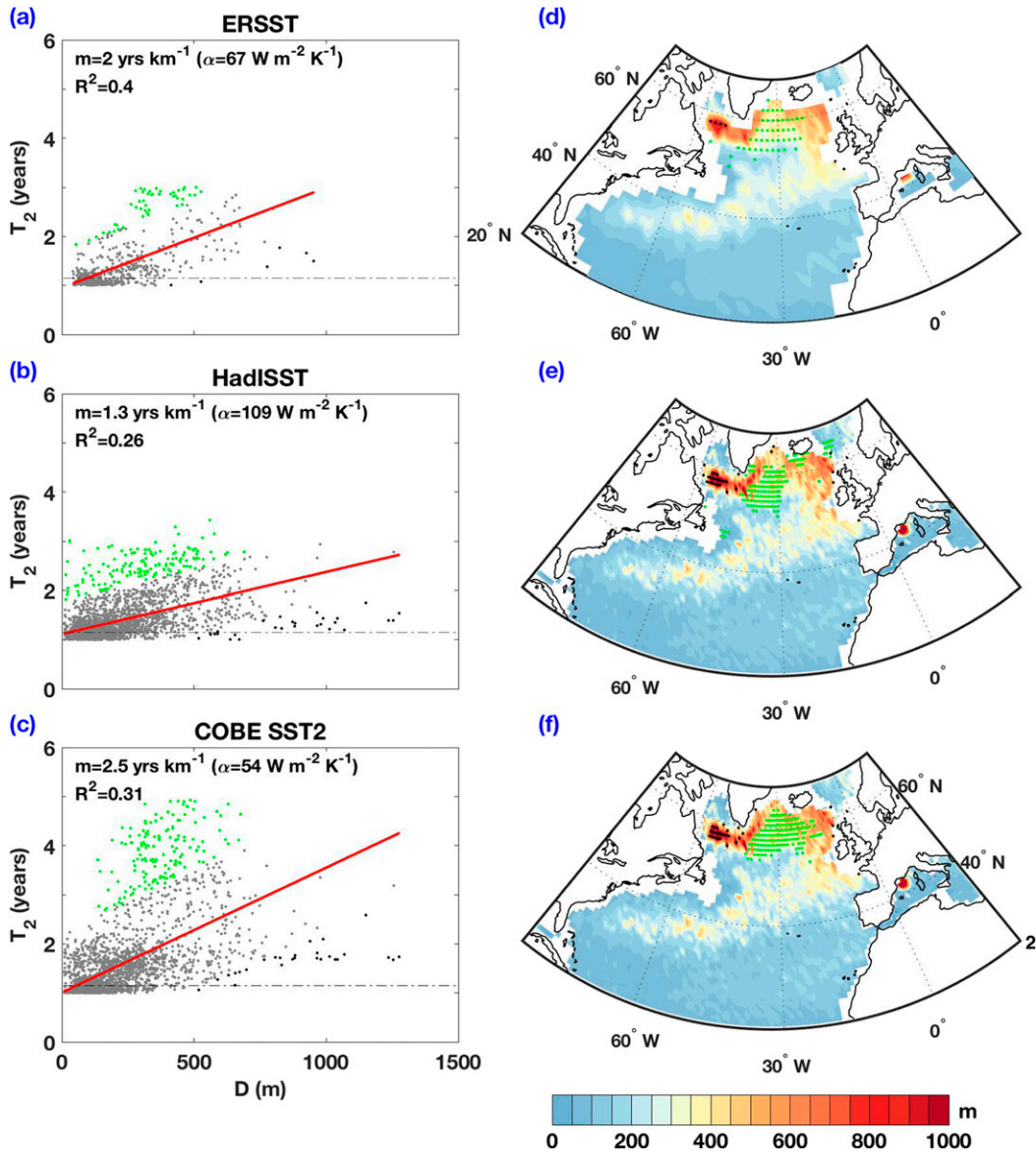
FIG. 8. How decorrelation time scales for wintertime SST are related to $D$: (left) Scatterplots comparing the decorrelation time scale $T_2$ for $\overline{SST}^w$ to $D$ (from the Argo MLD product) for all points in the ETNA for (a) ERSST, (b) HadISST, and (c) COBE SST2. The dash-dotted gray lines show values of $T_2$ above which $T_2$ can be distinguished from white noise on interannual time scales (see the appendix). The red line is a linear fit between $T_2$ and $D$, and the values of $R^2$ and the slope $m$ and the corresponding value of the damping parameter $\alpha$ [see Eq. (3)] are given in the top-left corner of each plot. Green (black) points are outliers that are more than two standard deviations above (below) the best fit line. (right) Spatial distribution of the outliers based on the linear fit for (d) ERSST, (e) HadISST, and (f) COBE SST2. Green (black) points are points more than two standard deviations above the best fit line, indicating points with higher (lower) than expected predictability. Colors show a spatial map of $D$ from the Argo MLD product (Holte et al. 2017).

similar to that of Cheng ($\alpha = 33\,\text{W m}^{-2}\,\text{K}^{-1}$), which uses $D$ from the Argo MLD climatology. Values of $\alpha$ for $\overline{SST}^w$ are significantly larger, ranging from 54 to $109\,\text{W m}^{-2}\,\text{K}^{-1}$ (Figs. 8a–c).

Expressing the slope in the units of $\alpha$ is not intended to suggest that calculating $\alpha$ based on the fits between $T_2$

and $D$ is an accurate method for calculating the damping parameter; instead the goal is merely to express $m$ in units that are familiar to the reader. Namely, the canonical value of the damping by air–sea heat fluxes derived by Frankignoul et al. (1998) is $\alpha_a = 20\,\text{W m}^{-2}\,\text{K}^{-1}$, although values are somewhat higher in the subpolar

gyre and vary seasonally (values are larger in the wintertime). Differences between our calculations and that of Frankignoul et al. (1998) include:

1) Frankignoul et al. (1998) estimate only atmospheric damping $\alpha_a$, whereas we calculate the total damping rate by the atmosphere and ocean. Some studies suggest that oceanic damping is as large or larger than atmospheric damping (Hall and Manabe 1997), which explains why our calculated values of $\alpha$ are generally larger than $20\,\mathrm{W\,m^{-2}\,K^{-1}}$.

2) Our value of $\alpha$ depends linearly on the chosen value of $D$. We chose $D$ as the wintertime climatological MLD, a quantity that is quite large in the subpolar gyre, which also contributes to our relatively large values of $\alpha$.

3) Frankignoul et al. (1998) consider damping of monthly SST anomalies, whereas we consider wintertime SST and $H$. As a result of the formation of the seasonal thermocline, $\overline{H}$ and $\overline{SST}^w$ anomalies are damped to the atmosphere more slowly since the atmospheric damping is essentially zero during the summertime when anomalies are isolated below the seasonal thermocline. This process is expected to lead to a value of $\alpha_a$ smaller than that of Frankignoul et al. (1998), but this effect is relatively modest since damping is smaller in the summer (Frankignoul and Kestenare 2002).

The above differences suggest that our value of $\alpha$ should not match that of Frankignoul et al. (1998) and that it should be larger than $20\,\mathrm{W\,m^{-2}\,K^{-1}}$, as indeed it is, particularly for $\overline{SST}^w$. However, the sensitivity of our calculated value of $\alpha$ to the choice of $D$ as well as the particular data products used ($\alpha$ for $\overline{SST}^w$ varies by almost a factor of 2 between the three different data products) suggests that $\alpha$ cannot be tightly constrained from calculations of decorrelation time scales.

## 6. Are there oscillatory variations in SST or upper-ocean heat content?

The value of $T_2$ measures the decorrelation time scale of SST/$H$, but it does not distinguish between different types of variability that may be leading to the measured value of $T_2$. For example, does $T_2$ simply indicate damping with an $e$-folding time scale given by $T_2$ or does oscillatory behavior play a role? The goal of this section is to determine the extent to which decaying versus oscillatory behavior of the autocorrelation function is important in setting decorrelation time scales for $\overline{H}$ and $\overline{SST}^w$. For an AR2 model, this question can be answered analytically based on the AR coefficients (von Storch and Zwiers 2002). Figure 9 is a scatterplot showing the AR coefficients, $\phi_1$ and $\phi_2$,

from the AR2 fit to time series of $\overline{H}$ (Fig. 9a) and $\overline{SST}^w$ (Fig. 9c) for all the points in the ETNA. Points above the black parabola have exponentially decaying autocorrelation functions; the further from the parabola, the more strongly the autocorrelation function decays. The points below the parabola have (decaying) oscillatory behavior; the farther from the parabola, the stronger the oscillations are. One can see that there are numerous points in the ETNA for which time series of $\overline{H}$ and $\overline{SST}^w$ are in the regime with oscillatory behavior. Two points with the most oscillatory behavior for $\overline{H}$ are chosen: P3 from the Ishii product is located south of the Gulf Stream (labeled on Fig. 3a), and P4 from the EN4 product is located in the central subpolar gyre (labeled on Fig. 3b). The autocorrelation function of these most extreme points shows only modest oscillatory behavior, with an autocorrelation of about $-0.25$ after about 3 years (Fig. 9b). For $\overline{SST}^w$ there are very few points that are a significant distance below the parabola, and all of these points lie in regions of seasonal sea ice: P5 from ERSST is located in Denmark Strait (labeled on Fig. 4a) and P6 from HadISST is located off the coast of Labrador (labeled on Fig. 4b). The autocorrelation function for these points exhibits very weak oscillatory behavior, with an autocorrelation of about $-0.2$ after about 2–3 years (Fig. 9d). Therefore, we conclude that on the time scales resolved by our observational products, oscillatory behavior does not play a significant role for variability in either $\overline{H}$ or $\overline{SST}^w$. Instead, $T_2$ is mainly controlled by the decay of SST/$H$ anomalies, and is expected to be quite similar to the familiar $e$-folding time scale.

## 7. Conclusions

In this work, a lower bound for predictability time scales for SST and UOHC is estimated purely from ocean observations, and the degree to which spatial variations in the MLD can explain predictability time scales is quantified. Two diagnostics are used for SST/UOHC: wintertime SST and the temperature averaged over the layer from the surface to the wintertime climatological MLD (integral denoted as $H$); both diagnostics quantify the ocean memory on interannual time scales and are designed to minimize the effects of seasonal variations, specifically the loss of SST memory in the summertime due to the formation of the seasonal thermocline. Gridded observational products are used to estimate predictability for wintertime SST and $H$ in the North Atlantic using a measure of the decorrelation time scale based on the local autocorrelation function (DelSole 2001). Decorrelation time scales for both wintertime SST and $H$ are longest in the subpolar gyre, with maximum decorrelation time scales of 4–6 years.
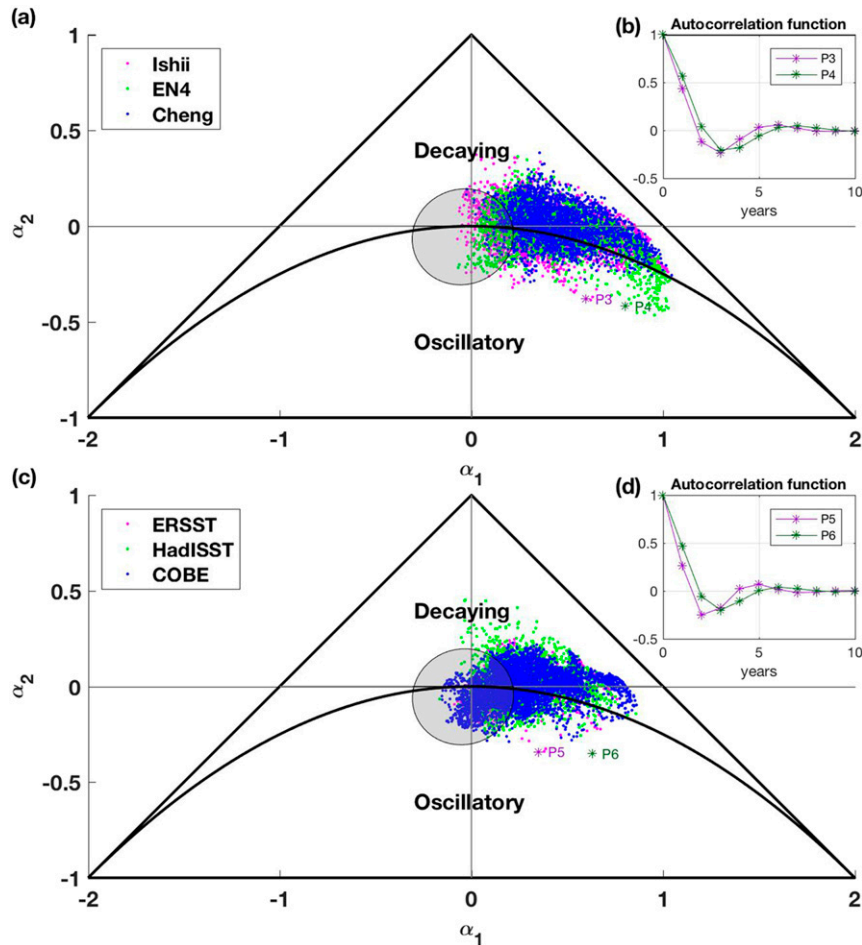
FIG. 9. Scatterplots for the AR2 coefficients, $\phi_1$ and $\phi_2$, for all points in the ETNA. (a) The AR2 coefficients for the fits to $\overline{H}$ from Ishii (magenta), EN4 (green), and Cheng (blue). (c) The AR2 coefficients for the fits to $\overline{SST}^w$ from ERSST (magenta), HadISST (green), and COBE SST2 (blue). The triangle identifies the range of coefficients for which an AR2 process is stationary; the fact that all the points lie within the triangle indicates that the time series of $\overline{H}$ and $\overline{SST}^w$ are stationary. The black parabola separates AR2 processes with exponentially decaying autocorrelation functions (above the parabola) from those with quasi-oscillatory behavior (below the parabola). The equation of the parabola is $\phi_1^2 + 4\phi_2 = 0$ (von Storch and Zwiers 2002). The shaded gray ellipse shows the AR2 coefficients that cannot be distinguished from white noise at the 90% confidence level (see the appendix and Fig. A1). For both $\overline{H}$ and $\overline{SST}^w$, we chose several points that are the farthest below the parabola and examine their autocorrelation function to see the strength of the oscillations. P3 (from Ishii, located south of the Gulf Stream, labeled in Fig. 3a) and P4 (from EN4, located in the central subpolar gyre, labeled in Fig. 3b) are labeled in (a), and their autocorrelation functions are shown (b). P5 (from ERSST, located south of Greenland, labeled in Fig. 4a) and P6 (from HadISST, located off the coast of Labrador, labeled in Fig. 4b) are labeled in (c), and their autocorrelation functions are shown in (d).

Decorrelation time scales for wintertime SST and $H$ are generally similar, except in regions with very deep mixed layers, such as the Labrador Sea, where time scales for $H$ are much larger. Oscillatory variations are found to play little role in setting predictability time scales for wintertime SST and $H$; instead decorrelation time scales are controlled mainly by the decay rate of anomalies.

Our decorrelation time scales are generally similar to predictability time scales calculated from observations using other statistical techniques, such as linear inverse models (LIMs). Predictability for decadal SST variations from LIMs is generally 3–4 years over most of the North Atlantic domain, with predictability generally being larger in subpolar regions (Zanna et al. 2012;

Newman 2013; Huddart et al. 2017). The similarity of our results to those from LIMs suggests that our consideration of a predictability measure based on the local autocorrelation (e.g., the decorrelation time scale) is likely not a significant limitation in the Atlantic basin. This is further supported by results from LIMs that suggest nonlocal SST variability has little influence on SST predictions in a given region of the Atlantic (Huddart et al. 2017). Our conclusion that oscillatory variations play little role in predictability of wintertime SST and $H$ is supported by results from statistical models that suggest predictability in the Atlantic, while higher than most other basins, only slightly exceeds that of a persistence forecast (DelSole et al. 2013).

While our results generally agree with other studies on the magnitude of predictability time scales in the North Atlantic, there are some differences in the detailed spatial patterns. For example, Zanna et al. (2012) find that forecast skill compared to climatology is quite low in the subpolar region south of Greenland, but we find that this region has relatively high values of $T_2$ (greater than 2 years). This result may be due to the analysis of wintertime SST (in our study) versus annual-average SST [in Zanna et al. (2012)], the different time periods studied [Zanna et al. (2012) start analysis in 1870 whereas we start in 1945], or the truncation of the SST variability using empirical orthogonal functions in the LIM. The spatial patterns of predictability time scales for SST in Huddart et al. (2017) are more similar to ours, but this study uses decadal filtering to construct the LIM, so results are not directly comparable to our study. It would be interesting to create a LIM to predict wintertime SST or wintertime/annual average $H$ (with no additional filtering) and see how these results compare to our decorrelation time scales.

Spatial variations in the wintertime climatological MLD explain 51%–74% of the regional variations in decorrelation time scales for UOHC and 26%–40% of the regional variations in decorrelation time scales for SST in the extratropical North Atlantic. These results suggest that to leading order decorrelation time scales for $H$ are determined by the thermal memory of the ocean. Interestingly, decorrelation time scales for UOHC that are significantly larger than expected based on the local mixed layer depth tend to be located in regions with shallow bathymetry. Decorrelation time scales for wintertime SST have a weaker relationship with the MLD, as a result of both noise and surface processes. In particular, in the Labrador Sea, decorrelation time scales for wintertime SST are much shorter than expected from the local MLD.

The suggestion that a large portion of the spatial variations in predictability time scales can be explained by variations in the wintertime climatological MLD raises the question of whether SST and UOHC variability in the ocean can simply be explained by stochastic atmospheric forcing. Addressing this question is the subject of ongoing work. The fact that the damping parameter calculated from our decorrelation time scales is significantly larger than the canonical atmospheric damping parameter (e.g., Frankignoul et al. 1998) suggests that oceanic damping is important, as found by Hall and Manabe (1997) and argued by Zhang (2017). Where dynamical oceanic process (e.g., ocean currents) are important in creating SST and UOHC anomalies remains to be explored in detail, but both our results and other studies suggest oceanic processes are more important in the subpolar gyre (e.g., Buckley et al. 2014, 2015; Karspeck et al. 2015; Robson et al. 2014; Menary et al. 2015; O'Reilly et al. 2016; Delworth et al. 2017; Piecuch et al. 2017; Zhang 2017).

Our observationally based results do not indicate the same degree of variation in the magnitude and spatial patterns of predictability time scales that are typically seen in CMIP5 models. Whether this is due to model deficiencies or may be related to the limited sampling period of ocean observations is currently not clear. Our results suggest that the predictability diagnostics that have been used for CMIP5 models, particularly temperature averaged over depth layers that are constant in space [e.g., top 300 m used in Branstator et al. (2012)] may not be the best measures of ocean memory. Calculating predictability time scales for wintertime SST and $H$ for CMIP5/CMIP6 control integrations and historical runs is a subject of ongoing study.

One of the most significant caveats of our study is that we only have about 70 years of observations and remove a quadratic function (in time) from our data prior to analysis. Therefore, we can only resolve variations up to decadal time scales. If longer variations play a substantial role in ocean predictability, then our results are surely an underestimate of ocean predictability.
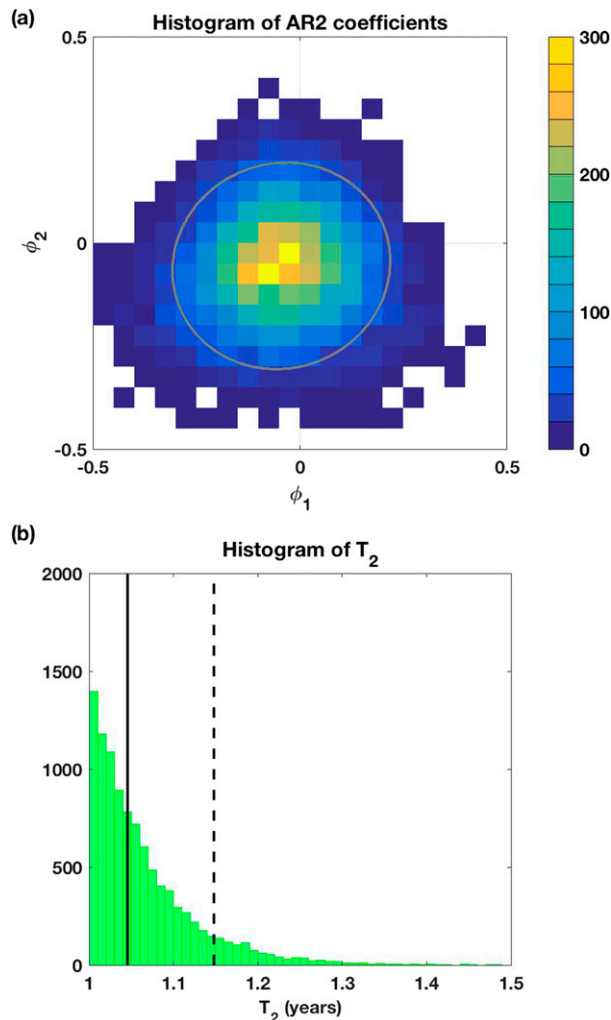
FIG. A1. Histograms of AR parameters and $T_2$ for white noise (with the same length of the observational time series). Distributions come from 10 000 repetitions (see the appendix). (a) The AR2 coefficients, $\phi_1$ and $\phi_2$, from white noise. The colors show a histogram of the coefficients and the gray ellipse shows the 90% confidence level. (b) A histogram for $T_2$. The solid black line shows the median, and the dashed black line shows the 90% quantile of the distribution.

## APPENDIX

## Determining the Level of $T_2$ That Can Be Distinguished from White Noise on Interannual Time Scales

In our analysis, we use yearly time resolution observations (wintertime average SST and yearly and wintertime average $H$), and thus we are only able to measure interannual predictability. The point of this appendix is to evaluate what value of $T_2$ is required in order to demonstrate statistically significant interannual predictability.

From a theoretical perspective, the minimum value of $T_2$ is one, and this value will occur when the autocorrelation decreases to zero at a lag of one year. However, due to finite sampling it is possible to get larger values of $T_2$ even for white noise time series, which have no interannual predictability. Thus, we generate probability distribution functions for $T_2$ based on white noise input (for time series of SST and $H$) and compute thresholds for $T_2$. Where $T_2$ calculated from the data exceeds the threshold from white noise, we conclude that there is interannual predictability. Where $T_2$ does not exceed the threshold, the variability cannot be distinguished from white noise on interannual time scales.

We first generate a time series of white noise with the same length as the number of years of the observational dataset of SST or $H$. The specific lengths of the time series of SST and $H$ for each product are given in Tables 1 and 2, but in all cases the time series are about 70 years. Then, to be consistent with the approach taken for the analysis of the data, a quadratic function (in time) is removed from the time series and AR fits are computed (using AR order of 2). We repeat this experiment 10 000 times in order to get distributions for the AR parameters resulting from white noise. Figure A1a shows a histogram of the AR2 coefficients, $\phi_1$ and $\phi_2$, based on the AR2 fit to the white noise time series. The gray line shows the 90% confidence ellipsoid for the AR parameters. Then, $T_2$ is computed from the AR parameters; the distribution of $T_2$ for an AR2 model is shown in Fig. A1b. We chose our threshold for $T_2$ to be the 90% quantile of the distribution; when $T_2$ exceeds this threshold we state that it can be distinguished from white noise.

## REFERENCES

Alexander, M. A., and C. Deser, 1995: A mechanism for the recurrence of wintertime midlatitude SST anomalies. *J. Phys. Oceanogr.*, **25**, 122–137, https://doi.org/10.1175/1520-0485(1995)025<0122:AMFTRO>2.0.CO;2.

——, L. Matrosova, C. Penland, J. D. Scott, and P. Chang, 2008: Forecasting Pacific SSTs: Linear inverse model predictions of the PDO. *J. Climate*, **21**, 385–402, https://doi.org/10.1175/2007JCLI1849.1.

Argo, 2019: Argo float data and metadata from Global Data Assembly Centre (Argo GDAC). SEANOE, accessed 2019, https://doi.org/10.17882/42182.

Årthun, M., T. Eldevik, E. Viste, H. Drange, T. Furevik, H. L. Johnson, and N. S. Keenlyside, 2017: Skillful prediction of northern climate provided by the ocean. *Nat. Commun.*, **8**, 16152, https://doi.org/10.1038/ncomms16152.

Bersch, M., 2002: North Atlantic Oscillation–induced changes of the upper layer circulation in the northern North Atlantic Ocean. *J. Geophys. Res.*, **107**, 3156, https://doi.org/10.1029/2001JC000901.

Boer, G. J., and Coauthors, 2016: The Decadal Climate Prediction Project (DCPP) contribution to CMIP6. *Geosci. Model Dev.*, **9**, 3751–3777, https://doi.org/10.5194/gmd-9-3751-2016.

Boyer, T., and Coauthors, 2016: Sensitivity of global upper-ocean heat content estimates to mapping methods, XBT bias corrections, and baseline climatologies. *J. Climate*, **29**, 4817–4842, https://doi.org/10.1175/JCLI-D-15-0801.1.

Branstator, G., H. Teng, G. A. Meehl, M. Kimoto, J. R. Knight, M. Latif, and A. Rosati, 2012: Systematic estimates of initial-value decadal predictability for six AOGCMs. *J. Climate*, **25**, 1827–1846, https://doi.org/10.1175/JCLI-D-11-00227.1.

Buckley, M. W., R. M. Ponte, G. Forget, and P. Heimbach, 2014: Low-frequency SST and upper-ocean heat content variability in the North Atlantic. *J. Climate*, **27**, 4996–5018, https://doi.org/10.1175/JCLI-D-13-00316.1.

——, ——, ——, and ——, 2015: Determining the origins of advective heat transport convergence variability in the North Atlantic. *J. Climate*, **28**, 3943–3956, https://doi.org/10.1175/JCLI-D-14-00579.1.

Cheng, L., K. E. Trenberth, J. Fasullo, T. Boyer, J. Abraham, and J. Zhu, 2017: Improved estimates of ocean heat content from 1960 to 2015. *Sci. Adv.*, **3**, e1601545, https://doi.org/10.1126/sciadv.1601545.

Corti, S., and Coauthors, 2015: Impact of initial conditions versus external forcing in decadal climate predictions: A sensitivity experiment. *J. Climate*, **28**, 4454–4470, https://doi.org/10.1175/JCLI-D-14-00671.1.

de Coëtlogon, G., and C. Frankignoul, 2003: The persistence of winter sea surface temperature in the North Atlantic. *J. Climate*, **16**, 1364–1377, https://doi.org/10.1175/1520-0442-16.9.1364.

DelSole, T., 2001: Optimally persistent patterns in time-varying fields. *J. Atmos. Sci.*, **58**, 1341–1356, https://doi.org/10.1175/1520-0469(2001)058<1341:OPPITV>2.0.CO;2.

——, 2017: Decadal prediction of temperature: Achievements and future prospects. *Curr. Climate Change Rep.*, **3**, 99–111, https://doi.org/10.1007/s40641-017-0066-x.

——, and M. K. Tippett, 2018: Predictability in a changing climate. *Climate Dyn.*, **51**, 531–545, https://doi.org/10.1007/s00382-017-3939-8.

——, L. Jia, and M. K. Tippett, 2013: Decadal prediction of observed and simulated sea surface temperatures. *Geophys. Res. Lett.*, **40**, 2773–2778, https://doi.org/10.1002/grl.50185.

Delworth, T. L., F. Zeng, L. Zhang, R. Zhang, G. A. Vecchi, and X. Yang, 2017: The central role of ocean dynamics in connecting the North Atlantic Oscillation to the extratropical component of the Atlantic multidecadal oscillation. *J. Climate*, **30**, 3789–3805, https://doi.org/10.1175/JCLI-D-16-0358.1.

Deser, C., M. A. Alexander, and M. S. Timlin, 2003: Understanding the persistence of sea surface temperature anomalies in midlatitudes. *J. Climate*, **16**, 57–72, https://doi.org/10.1175/1520-0442(2003)016<0057:UTPOSS>2.0.CO;2.

Dias, D. F., A. Subramanian, L. Zanna, and A. J. Miller, 2019: Remote and local influences in forecasting Pacific SST: A linear inverse model and a multimodel ensemble study. *Climate Dyn.*, **52**, 3183–3201, https://doi.org/10.1007/s00382-018-4323-z.

Doblas-Reyes, F. J., and Coauthors, 2013: Initialized near-term regional climate change prediction. *Nat. Commun.*, **4**, 1715, https://doi.org/10.1038/ncomms2704.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 2016: Overview of the Coupled Model Intercomparison Project phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.*, **9**, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016.

Frajka-Williams, E., P. B. Rhines, and C. C. Eriksen, 2014: Horizontal stratification during deep convection in the Labrador Sea. *J. Phys. Oceanogr.*, **44**, 220–228, https://doi.org/10.1175/JPO-D-13-069.1.

Frankignoul, C., and K. Hasselmann, 1977: Stochastic climate models, Part II: Application to sea-surface temperature anomalies and thermocline variability. *Tellus*, **29**, 289–305, https://doi.org/10.3402/tellusa.v29i4.11362.

——, and E. Kestenare, 2002: The surface heat flux feedback. Part I: Estimates from observations in the Atlantic and the North Pacific. *Climate Dyn.*, **19**, 633–647, https://doi.org/10.1007/s00382-002-0252-x.

——, A. Czaja, and B. L'Heveder, 1998: Air–sea feedback in the North Atlantic and surface boundary conditions for ocean models. *J. Climate*, **11**, 2310–2324, https://doi.org/10.1175/1520-0442(1998)011<2310:ASFITN>2.0.CO;2.

Good, S. A., M. J. Martin, and N. A. Rayner, 2013: EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *J. Geophys. Res. Oceans*, **118**, 6704–6716, https://doi.org/10.1002/2013JC009067.

Häkkinen, S., and P. B. Rhines, 2004: Decline of subpolar North Atlantic circulation during the 1990s. *Science*, **304**, 555–559, https://doi.org/10.1126/science.1094917.

——, ——, and D. L. Worthen, 2011: Atmospheric blocking and Atlantic multidecadal ocean variability. *Science*, **334**, 655–659, https://doi.org/10.1126/science.1205683.

Hall, A., and S. Manabe, 1997: Can local linear stochastic theory explain sea surface temperature and salinity variability? *Climate Dyn.*, **13**, 167–180, https://doi.org/10.1007/s003820050158.

Hawkins, E., and R. Sutton, 2009: Decadal predictability of the Atlantic Ocean in a coupled GCM: Forecast skill and optimal perturbations using linear inverse modeling. *J. Climate*, **22**, 3960–3978, https://doi.org/10.1175/2009JCLI2720.1.

Hermanson, L., R. Eade, N. H. Robinson, N. J. Dunstone, M. B. Andrews, J. R. Knight, A. A. Scaife, and D. M. Smith, 2014: Forecast cooling of the Atlantic subpolar gyre and associated impacts. *Geophys. Res. Lett.*, **41**, 5167–5174, https://doi.org/10.1002/2014GL060420.

Hirahara, S., M. Ishii, and Y. Fukuda, 2014: Centennial-scale sea surface temperature analysis and its uncertainty. *J. Climate*, **27**, 57–75, https://doi.org/10.1175/JCLI-D-12-00837.1.

Ho, C. K., E. Hawkins, L. Shaffrey, and F. M. Underwood, 2013: Statistical decadal predictions for sea surface temperatures: A benchmark for dynamical GCM predictions. *Climate Dyn.*, **41**, 917–935, https://doi.org/10.1007/s00382-012-1531-9.

Holte, J., L. D. Talley, J. Gilson, and D. Roemmich, 2017: An Argo mixed layer climatology and database. *Geophys. Res. Lett.*, **44**, 5618–5626, https://doi.org/10.1002/2017GL073426.

Huddart, B., A. Subramanian, L. Zanna, and T. Palmer, 2017: Seasonal and decadal forecasts of Atlantic sea surface temperatures using a linear inverse model. *Climate Dyn.*, **49**, 1833–1845, https://doi.org/10.1007/s00382-016-3375-1.

Ishii, M., M. Kimoto, K. Sakamoto, and S.-I. Iwasaki, 2006: Steric sea level changes estimated from historical ocean subsurface temperature and salinity analyses. *J. Oceanogr.*, **62**, 155–170, https://doi.org/10.1007/s10872-006-0041-y.

Jones, H., and J. Marshall, 1997: Restratification after deep convection. *J. Phys. Oceanogr.*, **27**, 2276–2287, https://doi.org/10.1175/1520-0485(1997)027<2276:RADC>2.0.CO;2.

Karspeck, A., S. Yeager, G. Danabasoglu, and H. Teng, 2015: An evaluation of experimental decadal predictions using CCSM4. *Climate Dyn.*, **44**, 907–923, https://doi.org/10.1007/s00382-014-2212-7.

Keenlyside, N., M. Latif, J. Jungclaus, L. Kornblueh, and E. Roeckner, 2008: Advancing decadal-scale climate prediction in the North Atlantic sector. *Nature*, **453**, 84–88, https://doi.org/10.1038/nature06921.

Kennedy, J. J., 2014: A review of uncertainty in in situ measurements and data sets of sea surface temperature. *Rev. Geophys.*, **52**, 1–32, https://doi.org/10.1002/2013RG000434.

Marshall, J., and F. Schott, 1999: Open-ocean convection: Observations, theory, and models. *Rev. Geophys.*, **37**, 1–64, https://doi.org/10.1029/98RG02739.

Matei, D., H. Pohlmann, J. Jungclaus, W. Müller, H. Haak, and J. Marotzke, 2012: Two tales of initializing decadal climate prediction experiments with the ECHAM5/MPI-OM model. *J. Climate*, **25**, 8502–8523, https://doi.org/10.1175/JCLI-D-11-00633.1.

Meehl, G. A., and Coauthors, 2014: Decadal climate prediction: An update from the trenches. *Bull. Amer. Meteor. Soc.*, **95**, 243–267, https://doi.org/10.1175/BAMS-D-12-00241.1.

Menary, M. B., D. L. R. Hodson, J. I. Robson, R. T. Sutton, and R. A. Wood, 2015: A mechanism of internal decadal Atlantic Ocean variability in a high-resolution coupled climate model. *J. Climate*, **28**, 7764–7785, https://doi.org/10.1175/JCLI-D-15-0106.1.

Namias, J., and R. M. Born, 1970: Temporal coherence in North Pacific sea-surface temperature patterns. *J. Geophys. Res.*, **75**, 5952–5955, https://doi.org/10.1029/JC075i030p05952.

Newman, M., 2007: Interannual to decadal predictability of tropical and North Pacific sea surface temperatures. *J. Climate*, **20**, 2333–2356, https://doi.org/10.1175/JCLI4165.1.

——, 2013: An empirical benchmark for decadal forecasts of global surface temperature anomalies. *J. Climate*, **26**, 5260–5269, https://doi.org/10.1175/JCLI-D-12-00590.1.

——, and Coauthors, 2016: The Pacific decadal oscillation, revisited. *J. Climate*, **29**, 4399–4427, https://doi.org/10.1175/JCLI-D-15-0508.1.

Nigam, S., A. Ruiz-Barradas, and L. Chafik, 2018: Gulf Stream excursions and sectional detachments generate the decadal pulses in the Atlantic multidecadal oscillation. *J. Climate*, **31**, 2853–2870, https://doi.org/10.1175/JCLI-D-17-0010.1.

O'Reilly, C. H., M. Huber, T. Woollings, and L. Zanna, 2016: The signature of low-frequency oceanic forcing in the Atlantic multidecadal oscillation. *Geophys. Res. Lett.*, **43**, 2810–2818, https://doi.org/10.1002/2016GL067925.

Piecuch, C. G., R. M. Ponte, C. M. Little, M. W. Buckley, and I. Fukumori, 2017: Mechanisms underlying recent decadal changes in subpolar North Atlantic Ocean heat content. *J. Geophys. Res. Oceans*, **122**, 7181–7197, https://doi.org/10.1002/2017JC012845.

Pohlmann, H., J. H. Jungclaus, A. Köhl, D. Stammer, and J. Marotzke, 2009: Initializing decadal climate predictions with the GECCO oceanic synthesis: Effects on the North Atlantic. *J. Climate*, **22**, 3926–3938, https://doi.org/10.1175/2009JCLI2535.1.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407, https://doi.org/10.1029/2002JD002670.

Robson, J., R. Sutton, and D. Smith, 2014: Decadal predictions of the cooling and freshening of the North Atlantic in the 1960s and the role of ocean circulation. *Climate Dyn.*, **42**, 2353–2365, https://doi.org/10.1007/s00382-014-2115-7.

——, P. Ortega, and R. Sutton, 2016: A reversal of climatic trends in the North Atlantic since 2005. *Nat. Geosci.*, **9**, 513–517, https://doi.org/10.1038/ngeo2727.

Ruiz-Barradas, A., L. Chafik, S. Nigam, and S. Häkkinen, 2018: Recent subsurface North Atlantic cooling trend in context of Atlantic decadal-to-multidecadal variability. *Tellus*, **70A**, 1–19, https://doi.org/10.1080/16000870.2018.1481688.

Sarafanov, A., A. Falina, A. Sokov, and A. Demidov, 2008: Intense warming and salinification of intermediate waters of southern origin in the eastern subpolar North Atlantic in the 1990s to mid-2000s. *J. Geophys. Res. Oceans*, **113**, C12022, https://doi.org/10.1029/2008JC004975.

Schmidt, S., and U. Send, 2007: Origin and composition of seasonal Labrador Sea freshwater. *J. Phys. Oceanogr.*, **37**, 1445–1454, https://doi.org/10.1175/JPO3065.1.

Schubert, S. D., M. J. Suarez, and J.-K. Schemm, 1992: Persistence and predictability in a perfect model. *J. Atmos. Sci.*, **49**, 256–269, https://doi.org/10.1175/1520-0469(1992)049<0256:PAPIAP>2.0.CO;2.

Send, U., and J. Marshall, 1995: Integral effects of deep convection. *J. Phys. Oceanogr.*, **25**, 855–872, https://doi.org/10.1175/1520-0485(1995)025<0855:IEODC>2.0.CO;2.

Smith, D. M., S. Cusack, A. W. Colman, C. K. Folland, G. R. Harris, and J. M. Murphy, 2007: Improved surface temperature prediction for the coming decade from a global climate model. *Science*, **317**, 796–799, https://doi.org/10.1126/science.1139540.

Smith, T. M., R. W. Reynolds, T. C. Peterson, and J. Lawrimore, 2008: Improvements to NOAA's historical merged land–ocean surface temperature analysis (1880–2006). *J. Climate*, **21**, 2283–2296, https://doi.org/10.1175/2007JCLI2100.1.

Straneo, F., 2006: Heat and freshwater transport through the central Labrador Sea. *J. Phys. Oceanogr.*, **36**, 606–628, https://doi.org/10.1175/JPO2875.1.

Toyoda, T., and Coauthors, 2015: Intercomparison and validation of the mixed layer depth fields of global ocean syntheses. *Climate Dyn.*, **49**, 753–773, https://doi.org/10.1007/s00382-015-2637-7.

van Oldenborgh, G. J., F. J. Doblas-Reyes, B. Wouters, and W. Hazeleger, 2012: Decadal prediction skill in a multi-model ensemble. *Climate Dyn.*, **38**, 1263–1280, https://doi.org/10.1007/s00382-012-1313-4.

von Storch, H., and F. W. Zwiers, 2002: *Statistical Analysis in Climate Research*. Cambridge University Press, 484 pp.

Yeager, S., A. Karspeck, G. Danabasoglu, J. Tribbia, and H. Teng, 2012: A decadal prediction case study: Late twentieth-century North Atlantic Ocean heat content. *J. Climate*, **25**, 5173–5189, https://doi.org/10.1175/JCLI-D-11-00595.1.

Zanna, L., 2012: Forecast skill and predictability of observed Atlantic sea surface temperatures. *J. Climate*, **25**, 5047–5056, https://doi.org/10.1175/JCLI-D-11-00539.1.

——, P. Heimbach, A. Moore, and E. Tziperman, 2012: Upper ocean singular vectors of the North Atlantic climate with implications for linear predictability and variability. *Quart. J. Roy. Meteor. Soc.*, **138**, 500–513, https://doi.org/10.1002/qj.937.

Zhang, R., 2017: On the persistence and coherence of subpolar sea surface temperature and salinity anomalies associated with the Atlantic multidecadal variability. *Geophys. Res. Lett.*, **44**, 7865–7875, https://doi.org/10.1002/2017GL074342.