

Subjective Likelihood for the Assessment of Trends in the Ocean's Mixed-Layer Depth

Ana Grohovac RAPPOLD, Michael LAVINE, and Susan LOZIER

This article describes a Bayesian statistical analysis of long-term changes in the depth of the ocean's mixed layer. The data are thermal profiles recorded by ships. For these data, there is no good sampling model and thus no obvious likelihood function. Our approach is to elicit posterior distributions for training data directly from the expert. We then infer the likelihood function and use it on large datasets.

KEY WORDS: Bayes; Climate change; Elicitation; Foundations; Likelihood; Oceanography.

1. INTRODUCTION

The typical Bayesian analysis posits data from a parametric family of sampling distributions, as in

$$\mathbf{y} \sim p(\mathbf{y} | \theta). \quad (1)$$

After \mathbf{y} has been observed, it is treated as fixed, and the likelihood function is defined to be $\ell(\theta) \equiv p(\mathbf{y} | \theta)$, a function of θ . The interpretation is that likelihood ratios $\ell(\theta_1)/\ell(\theta_2)$ quantify \mathbf{y} 's evidence for θ_1 as opposed to θ_2 .

For our dataset, there is no believable sampling model $p(\mathbf{y} | \theta)$, so we cannot assign $\ell(\theta) \equiv p(\mathbf{y} | \theta)$. We take a different approach, wherein lies the statistical novelty of this article. For several values of i (about a dozen), we show the expert \mathbf{y}_i and directly elicit his or her posterior distribution $p(\theta | \mathbf{y}_i)$. Elicitation is done under conditions in which the expert has an approximately uniform prior for θ . After elicitation, we know the prior and posterior and thus can infer the likelihood function.

After examining the dozen or so elicited posteriors and conferring with the expert, we constructed an algorithm that accepts a \mathbf{y} as input and yields a likelihood function $\ell(\theta)$ as output. After constructing the algorithm, we checked that it gave sensible results on several hundred more \mathbf{y} 's. We then applied the algorithm to our full collection of data $\{\mathbf{y}_i\}_{i=1}^T$, which, when combined with our real prior, yields the posterior that we use for inference. We call ℓ a likelihood function because it approximately summarizes the expert's weight of evidence and, when multiplied by the uniform prior, yields the posterior.

The data arise in a study of the ocean's climate. The situation is more complicated than described in this introduction because the data are a time series and our model must account for an annual cycle. Section 2 provides the scientific background and Section 3 describes the data. Section 4 describes our subjective likelihood, how it was elicited, and how it is modeled. It contains whatever statistical novelty is in this article. Section 5 describes our full model, prior, and posterior inference, accounting for the annual cycle, year-to-year variation, heteroscedasticity, and a possible secular trend. Finally, Section 6 presents a discussion.

Ana Grohovac Rappold is National Research Council Associate at US EPA, Duke University, Durham, NC 27708 (E-mail: ana@stat.duke.edu) (currently employed by the EPA). Michael Lavine is Professor, Department of Statistical Science, Duke University, Durham, NC 27708 (E-mail: michael@stat.duke.edu). Susan Lozier is Professor and Chair, Earth and Ocean Sciences, Nicholas School of the Environment and Earth Sciences, Duke University, Durham, NC 27708 (E-mail: mslozier@duke.edu). The authors gratefully acknowledge the support of the National Science Foundation's Collaboration in Mathematical Geosciences program.

2. THE OCEAN'S MIXED LAYER

Recent evidence that the world's oceans have warmed over the past 50 years (Levitus, Antonov, Boyer, and Stephens 2000) and that the attendant increase in the ocean's heat content is an order of magnitude larger than the increase in the atmospheric and cryospheric heat content (Levitus et al. 2001) have made it abundantly clear that a determination of how our global climate is changing in response to long-term natural and/or anthropogenic forcing depends on the effectiveness of the ocean as a heat reservoir. However, the effectiveness of the ocean as a reservoir is curtailed by increasing thermal stratification, which limits the extent to which surface signals can be transmitted to depth. Thus interest has focused on the upper ocean.

To a first approximation, oceanographers regard the ocean as having two layers, a mixed layer from the surface down to as much as several hundred meters, and a stratified layer beneath. The mixed layer is that part of the surface ocean that displays uniformity in such properties as temperature, salinity, and density. The mixed layer forms because the upper waters of the ocean are mixed through waves and wind and also through thermal convection when the surface waters overturn on losing heat, and thus buoyancy, to the atmosphere. Such overturning creates a mixed layer. The depth M of the mixed layer evolves through an annual cycle and depends on geographic location. Because M depends crucially on heat exchange with the atmosphere, long-term changes in heat exchange may result in long-term changes in M . Essentially, in this application we address the question as to whether or not there has been a secular trend in M in the North Atlantic subtropical gyre. Such an evaluation will increase our understanding of potential physical and biological consequences of global warming.

3. DATA

Hydrographic data such as temperature salinity and pressure, are collected from ships, sent to the National Oceanographic Data Center (NODC) where it is quality controlled, and then made publicly available (www.nodc.noaa.gov). This article reports on an analysis of NODC historical hydrographic data recorded over a small spatial region near Bermuda. We chose a sufficiently small region for our study so we can safely ignore spatial variability. We will give an analysis of data from a wide region of the North Atlantic elsewhere. A map of the data is provided in Figure 1; the data's temporal distribution is shown in Figure 2.

© 2007 American Statistical Association
Journal of the American Statistical Association
September 2007, Vol. 102, No. 479, Applications and Case Studies
DOI 10.1198/016214507000000761

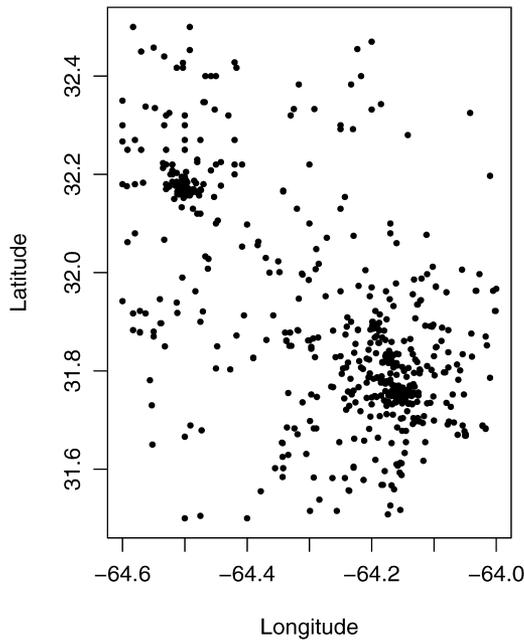


Figure 1. Data locations over the study domain, centered near Bermuda.

The i th data point \mathbf{y}_i is recorded at time t_i and consists of temperatures $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})$ measured at depths $\mathbf{d}_i = (d_{i1}, \dots, d_{in_i})$. (Other properties are also measured but are not of interest here.) The first depth is always sufficiently close to the surface so that (a) we can assume that y_{i1} is approximately equal to the surface temperature and (b) we can take $d_{i1} = 0$ without serious error. Often, but not always, the last depth is near the ocean floor. Almost always, this last depth is well below any plausible value for M ; we eliminate the few \mathbf{y}_i 's for which it is not. Temperature as a function of depth is called a temperature *profile*. Figure 3 shows one example profile. Measurements

depths are not the same from one profile to another, and, unlike the example in Figure 3, distances between successive measurements need not be constant within a profile.

Generally speaking, the upper layer of the ocean, because it is vertically mixed, should have a uniform temperature, whereas the stratified layer should have a monotonically decreasing temperature. Figure 3 exemplifies these characteristics. Locating the bottom of the mixed layer is fairly straightforward in such profiles. In Figure 3, we can say with confidence that M is somewhere near 50 meters. But not all profiles present such a clear delineation of M , for examples, see Figure 4. Each of the profiles in Figure 4 has several depths that might plausibly be the bottom of the mixed layer. A key question is: What do profiles such as those in Figure 4 tell us about M ?

If we had a physical model that predicted the spatial and temporal distribution of temperature with depth (i.e., a temperature profile) then we could fit the model to the data, and the likelihood function would quantify the information for the parameter M . However, one of the remaining fundamental problems in oceanography is a complete theoretical description of the thermocline, which expresses the horizontal and vertical change in the temperature field. In the absence of a complete description, simplifications have been offered. One such simplification of the nonlinear physics that governs the thermocline yields a prediction that the temperature should decrease exponentially from M to the ocean floor (Mellor 1996), where it is about 2°C throughout the world. If such a simplification were generally valid, then a changepoint model with three parameters—surface temperature, M , and decay rate—would fit the data well. But the data in our region of interest do not exhibit exponential decay, as illustrated in Figure 5, which shows the same profile displayed in Figure 3 along with several exponential decay curves. Overall, our understanding of the physics of the thermocline has not advanced to the point at which we could offer a model

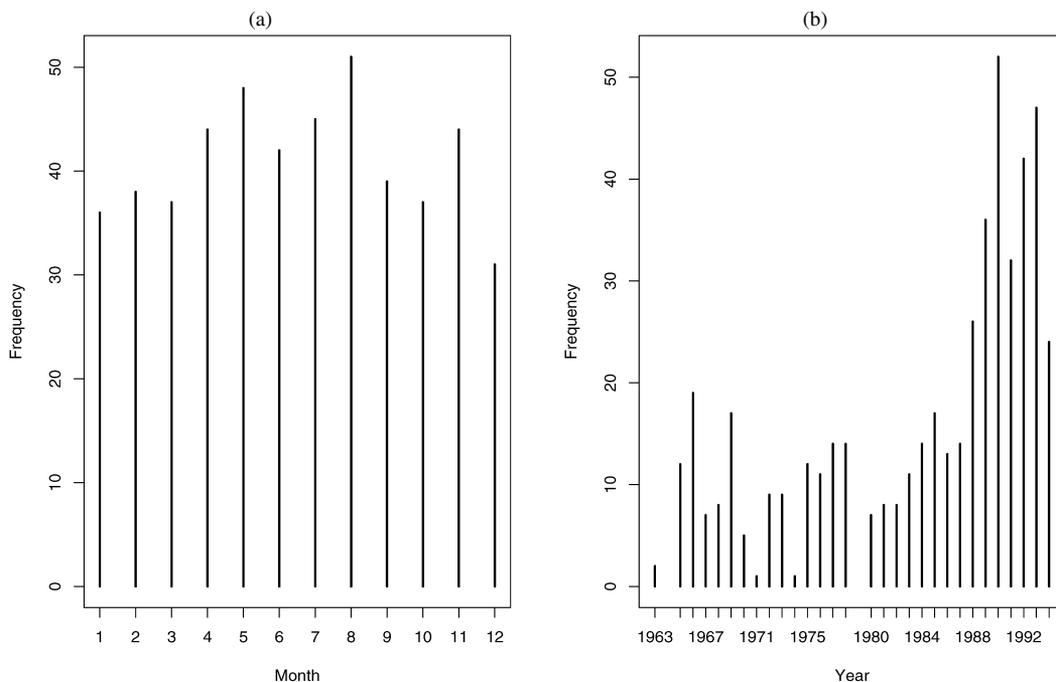


Figure 2. Temporal distribution of data used in this study: (a) monthly; (b) yearly.

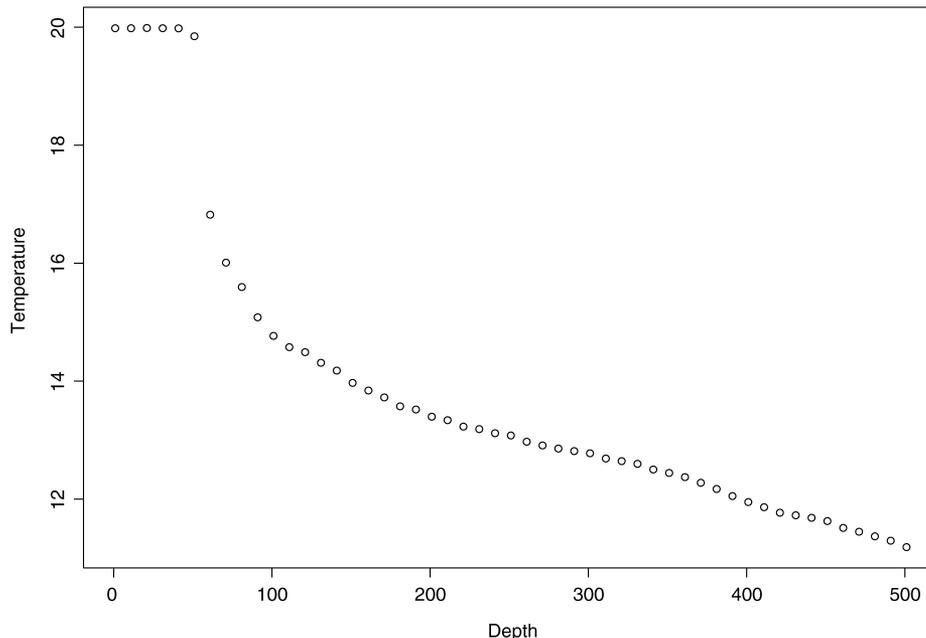


Figure 3. A sample thermal profile: Temperature as a function of depth measured at 10-m intervals. M appears to be between 40 and 60 meters.

to which to fit the data. In such a void, we instead turn to the methods described in Section 4.

4. SUBJECTIVE LIKELIHOOD

In this section we consider what can be learned from a single profile \mathbf{y} sampled at depths \mathbf{d} . The parameter of interest is M .

From either a likelihood or Bayesian perspective, what is needed is a sampling model $p(\mathbf{y} | M)$. But in Section 3 we argued that no reliable sampling model exists. So how can we describe what is learned from \mathbf{y} ? Our approach was to ask an oceanographer directly.

We showed the oceanographer a profile (\mathbf{y}, \mathbf{d}) and asked what she could tell us about M , based on that profile. Her re-

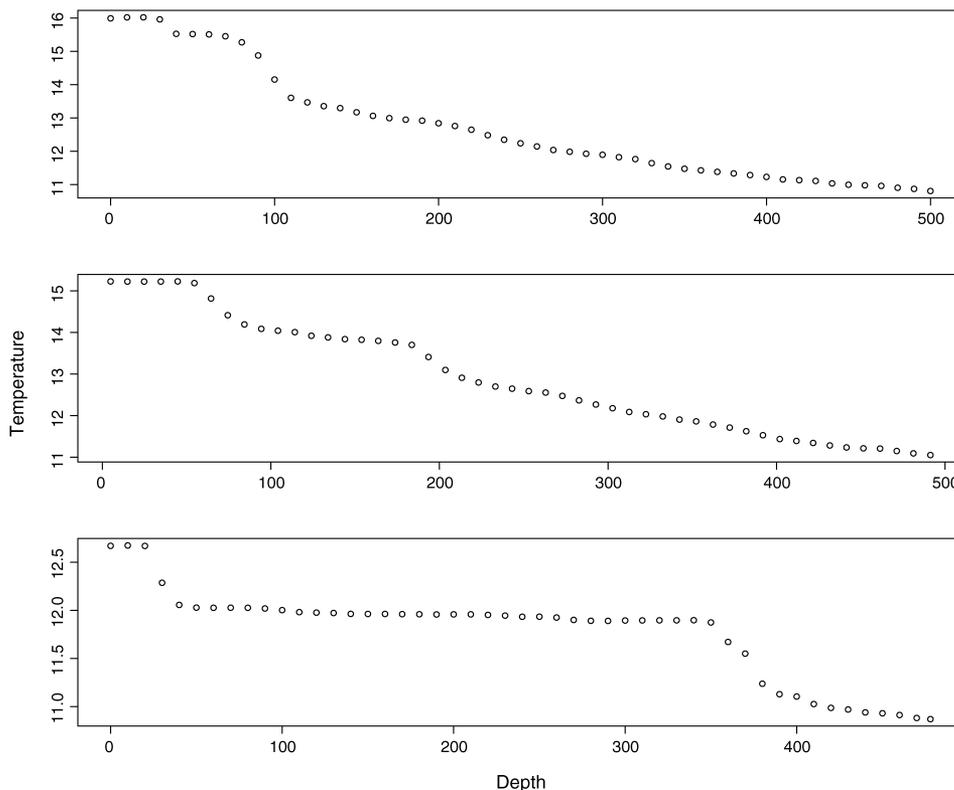


Figure 4. Three more sample profiles. Unlike the profile in Figure 3, none of these indicates M with clarity.

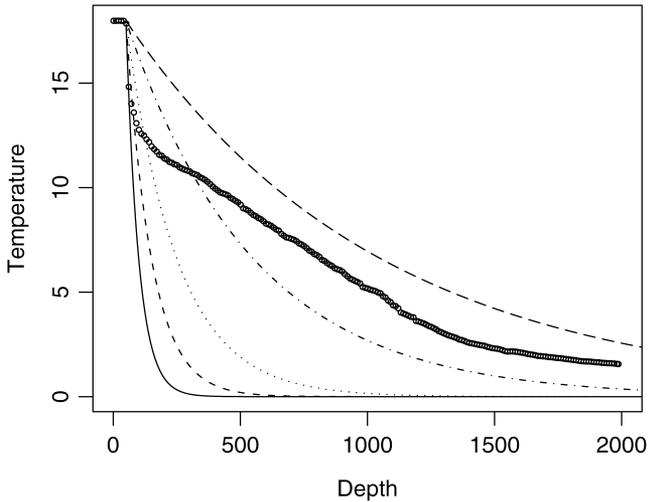


Figure 5. The same profile as in Figure 3 with several exponential decay curves. None fits the data well.

sponse was to tell us that $\Pr[M \in I_j]$ for $j = 1, \dots, n$ where I_j is the interval from d_j to d_{j+1} . (Here $d_1 = 0$ is the ocean surface; d_{n+1} is the floor, about 4,000 meters; the d_j 's need not be evenly spaced, so the I_j 's need not be equal in length.) The process was repeated for several profiles, about a dozen. Figure 3 shows an example. For this profile, the oceanographer's assessments were

$$\Pr[M \in I_6] \approx 10 \Pr[M \in I_5] \tag{2}$$

$\Pr[M \in \text{some other interval}] \approx \text{very small.}$

The next step was to ask the oceanographer how she makes her assessments. She told us that for an interval $I_j = [d_j, d_{j+1}]$, she considers two quantities: $y_1 - y_j$, the temperature drop from the surface to the top of I_j , and $(y_j - y_{j+1}) / (d_{j+1} - d_j)$, the

rate of temperature drop within I_j . Small values of $y_1 - y_j$ and large values of $(y_j - y_{j+1}) / (d_{j+1} - d_j)$ imply that I_j is likely to contain M . She also said these are the only quantities that matter; other aspects of the profile carry so little information as to be ignorable. To formalize, define

$$\Delta_1 = (\Delta_{11}, \dots, \Delta_{1n}) = (0, y_1 - y_2, \dots, y_1 - y_n)$$

and

$$\Delta_2 = (\Delta_{21}, \dots, \Delta_{2n}) = \left(\frac{y_1 - y_2}{d_2 - d_1}, \dots, \frac{y_n - y_{n-1}}{d_{n+1} - d_n} \right).$$

Then, according to the oceanographer, the probabilities $\{\Pr[M \in I_j]\}_{j=1}^n$ are some function $g(\Delta_1, \Delta_2)$.

The reasoning behind this is as follows. If interval I_j contains M , then the temperature above I_j should be roughly uniform, and $y_1 - y_j$ should be small. Thus $g(\Delta_1, \Delta_2)$ should be a decreasing function of Δ_1 . Similarly, if I_j contains M , then the rate of temperature decrease in I_j should be large, so $g(\Delta_1, \Delta_2)$ should be an increasing function of Δ_2 .

In addition, the oceanographer tells us that $g(\Delta_1, \Delta_2)$ can be written as $g_1(\Delta_1) \times g_2(\Delta_2)$, at least approximately, so our task is to find the functions g_1 and g_2 . The following paragraphs describe our elicitation and how we arrived at the particular functions g_1 and g_2 that we used in our analysis.

Our strategy is to elicit the oceanographer's posteriors for artificially constructed profiles designed to yield information about g_1 and g_2 . We illustrate this with the profile in Figure 3. In that figure there are only two intervals, I_5 and I_6 , with substantial posterior mass; they are separated by the datum (d_6, y_6) . We created a sequence of temperatures $y_1^* = y_7 < \dots < y_j^* = y_5$, showed the oceanographer profiles in which (d_6, y_6) had been replaced by (d_6, y_j^*) , and asked for her posteriors. Figure 6 illustrates this idea.

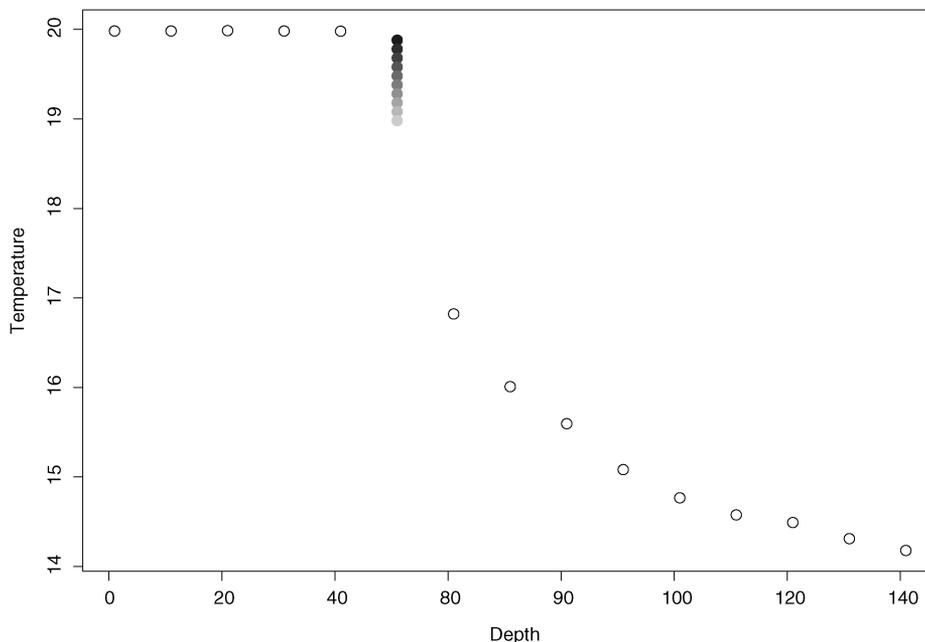


Figure 6. Artificial profiles for elicitation. The open circles are the profile from Figure 3. The shaded circles are a sequence of points meant to replace the point near (50, 19.5). We showed the oceanographer profiles containing each shaded circle in turn and elicited her posterior for each.

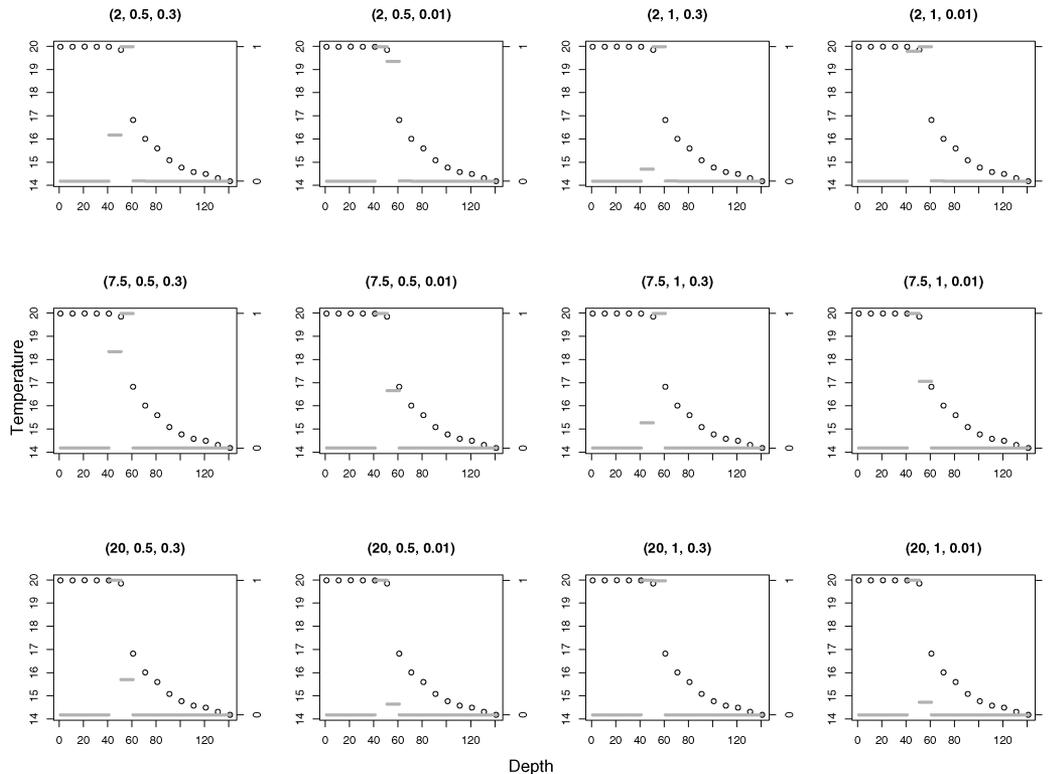


Figure 7. The profile of Figure 3 with unnormalized probabilities corresponding to three values of a , two values of b , and two values of θ from (3). Among these panels, those with $a = 2$ or $a = 7.5$, $b = 1$, and $\theta = .3$ provide the best fit to the oceanographer’s assessments in (2).

By varying only one point in the profile, we were able to learn about the shapes of g_1 and g_2 . We quickly found that the oceanographer assigns small probabilities to intervals whose Δ_1 is more than just a few tenths of a degree. Therefore, we want g_1 to decrease rapidly for small values of Δ_1 ; this consideration suggests a convex form for g_1 . After conferring with the oceanographer, we decided to try the functional form $g_1(\Delta_1) = \exp\{-a\Delta_1\}$. Similarly, we learned that g_2 should be concave and decided to try the form $g_2(\Delta_2) = [1 - \exp\{-\Delta_2/\theta\}]^b$; that is, after studying variants of the profile in Figure 6 and others, we decided to see whether for some (a, b, θ) , the functional form

$$g(\Delta_1, \Delta_2) = \exp\{-a\Delta_1\} \times [1 - \exp\{-\Delta_2/\theta\}]^b \quad (3)$$

could be made to fit the rest of the oceanographer’s assessments reasonably well. The left side of (3) is a vector with one component for each interval. The right side is also a vector; notation such as $\exp\{-a\Delta_1\}$ means elementwise multiplication and exponentiation.

No single combination (a, b, θ) fit best in all profiles. Therefore, our goal was to find a combination of (a, b, θ) that fit reasonably well in as many profiles as possible and do not make egregious errors. From our work with training data, we knew sensible ranges for the parameters. We selected several values of (a, b, θ) from their respective ranges and enlisted the oceanographer to help us choose the best combination. For each of several real profiles, we prepared a single sheet of paper with graphs of that profile overlaid with $g(\Delta_1, \Delta_2)$ for each combination of (a, b, θ) ; see Figure 7 for an example. Unnormalized posterior probabilities calculated according to (3) are plotted as

horizontal bars over their respective intervals. Looking at all of the profiles together, we asked the oceanographer to choose the best-fitting (a, b, θ) and also to tell us whether the overall fit was acceptable.

After considering multiple profiles—some real and some artificially constructed—we settled on $a = 2$, $b = 1$, and $\theta = .3$, that is,

$$\Pr[M \in I_j | \mathbf{y}, \mathbf{d}]_1^n \propto g(\Delta_1, \Delta_2) \approx \exp\{-2\Delta_1\} \times [1 - \exp\{-\Delta_2/.3\}], \quad (4)$$

as providing sufficiently close approximations to the oceanographer’s posterior probabilities for the profiles in the training data. The third panel in the first row of Figure 7 shows the result of applying (4) to the profile of Figure 3. The main thing to note is the relative heights of the bars in that panel and how they compare with (2).

Next, we wanted to check whether (4) would fit well to other profiles. So we created many (on the order of 100) plots similar to Figure 7, all with $(a, b, \theta) = (2, 1, .3)$ but each plot for a different profile, and asked the oceanographer whether she agreed reasonably well with the probabilities assigned by g and whether there was any systematic way in which g was making errors. From the results of these approximately 100 profiles, we introduced the following three modifications to g :

1. Of all the measurements in a temperature profile, the surface measurement is the most noisy. The surface measurement can be influenced by daytime warming; changes on such a small time scale are not of interest in our study of

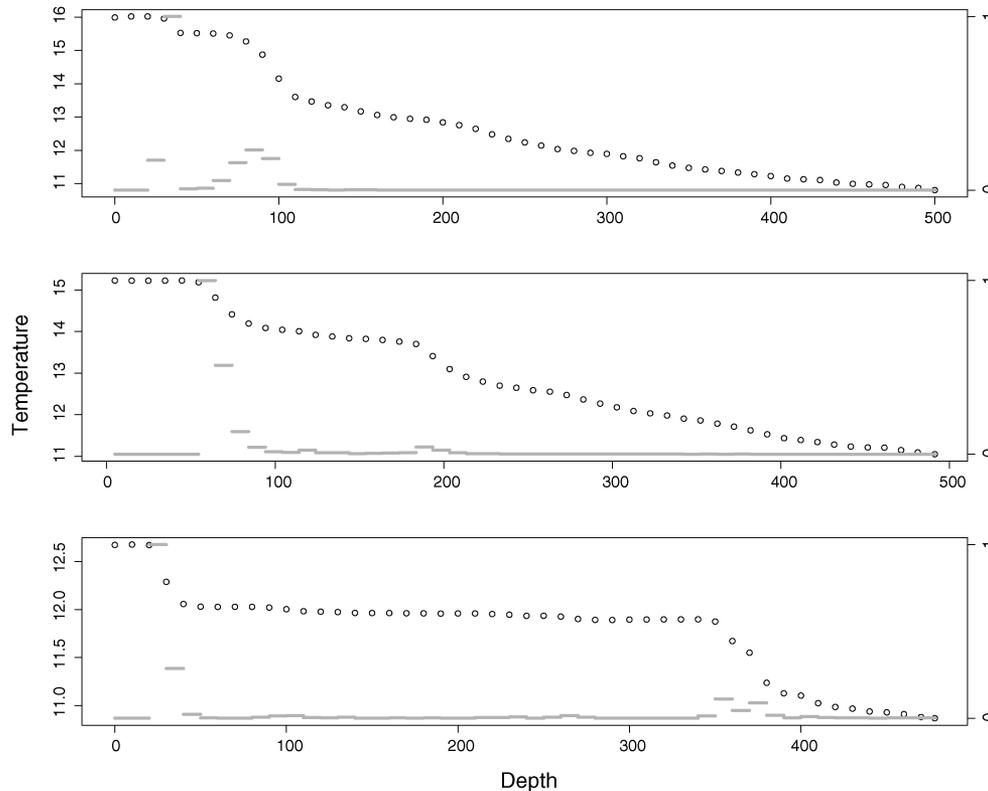


Figure 8. The three profiles of Figure 4. The horizontal bars show unnormalized posterior probabilities of intervals calculated according to (5). The oceanographer agreed that the horizontal bars approximately matched her subjective evaluation.

changes over years and decades. The key to recognizing such a situation is $y_1 > y_2 \approx y_3$.

When this happens, $\Delta_1 = (0, y_1 - y_2, \dots, y_1 - y_n)$ is not a good reflection of the information in the profile. Thus for profiles where $y_1 = \max y_i$ and $|y_2 - y_3| < .05$, we set $\Delta_1 = (0, y_2 - y_2, \dots, y_2 - y_n)$.

2. The idea behind $\Delta_{2j} = (y_j - y_{j+1}) / (d_{j+1} - d_j)$ is, obviously, that an interval containing M is likely to have a large temperature drop. But the oceanographer’s judgment is that temperature drops greater than about $.3^\circ\text{C}$ are not more indicative of M than drops of about $.3^\circ\text{C}$. Therefore, we set $\Delta_{2j} = \min((y_j - y_{j+1}), .3) / (d_{j+1} - d_j)$.
3. Equation (4) was constructed to mimic the oceanographer’s assessments of high probability intervals. For low probability intervals, it accurately captures the fact that they have low probability but might misstate their probability ratios by several orders of magnitude. In addition, it is widely believed that subjective probability assessments are often too sharp and understate the true amount of uncertainty. The result in our analysis could be oversensitivity to outliers or intervals with low probabilities. Therefore, we modify (4) and set

$$\begin{aligned} \Pr[M \in I_j \mid \mathbf{y}, \mathbf{d}]_1^2 &\propto g^*(\Delta_1, \Delta_2) \\ &\equiv \max \left[g(\Delta_1, \Delta_2), \frac{\max g(\Delta_1, \Delta_2)}{100} \right]. \end{aligned} \quad (5)$$

After again going through about 100 new profiles and the probabilities implied by (5), we were satisfied that we could

reproduce the oceanographer’s assessments to a reasonable degree of accuracy.

Figure 8 shows how (5) works for the profiles in Figure 4. The main point to note is the fact that the probability distributions in Figure 8 are bimodal.

Two more facts are needed to complete the specification of the likelihood:

1. Because we did not tell the oceanographer the time of year or physical location of the profiles, her prior for M was approximately uniform. (She agreed.)
2. Her posterior for M is approximately uniform within each interval. (She agreed.)

Equation (5) and fact 2 specify the posterior density of M . It is piecewise constant on the intervals I_j ; its value on I_j is

$$p(m \mid \mathbf{y}) = \frac{g^*(\Delta_1, \Delta_2)_j}{(d_{j+1} - d_j) \sum (g^*(\Delta_1, \Delta_2))} \quad (6)$$

for $m \in I_j$. And because the prior was uniform, (6) is also proportional to the likelihood function.

More formally, for any depth m , let $j(m)$ be the interval containing m , that is, $m \in I_{j(m)}$. Then we define the subjective likelihood function $\ell(m)$ by

$$\ell(m) \propto \frac{g^*(\Delta_1, \Delta_2)_{j(m)}}{d_{j(m)+1} - d_{j(m)}}. \quad (7)$$

Equation (7) is a rule for computing the likelihood function for any profile. In subsequent sections we apply the rule to a large collection of profiles, more than the oceanographer can

assess individually, and with a prior designed for the substantive investigation, not for elicitation.

5. APPLICATION TO THE MIXED LAYER

Section 4 described what we learn from a single profile. This section describes how we combine information from many profiles $\mathbf{y}_1, \dots, \mathbf{y}_T$ (Fig. 1) running over a time index $t = t_1, \dots, t_T$ that spans multiple years (Fig. 2) to reach a conclusion about long-term trend in M .

Figure 9 shows the result of applying (7) to each profile separately. Each gray vertical bar is the maximum likelihood interval of some profile; its depth is indicated on the ordinate. Its abscissa is the day of year when that profile was recorded. (The curve is a posterior mean and is discussed later.) There are two features of note, an annual cycle with deeper mixed layers in the winter and greater variability in the winter. Our model accommodates both features, as well as a possible long-term trend.

5.1 Mean Annual Cycle

Let $M(t)$ be the mixed layer depth at time t . It is apparent that $M(t)$ undergoes an annual cycle. The mixed layer deepens in the fall as atmospheric temperatures decrease. Colder air means colder sea surface temperatures, which in turn means that surface waters become dense and sink. The sinking causes surface waters to mix with deeper waters. The process continues through the winter, leading to deeper mixing and increased values of M . The effect is reversed in the spring as surface waters warm. Spring heat is conveyed downward through diffusion, a much slower process than fall’s convection. Thus the annual cycle is asymmetric, and the summer mixed layer is relatively shallow and stable.

We use μ to denote the mean annual cycle. Thus, for any time t ,

$$\mu(t) = \mu(t \bmod 365).$$

We model μ with a *process convolution* (see Higdon 1998 for details) as follows. Let $0 \leq v_1 < v_2 < \dots < v_{12} \leq 365$ be an equally spaced sequence of points (mod 365), let x_1, \dots, x_{12} , be values “at” those points, and let k be a kernel. (We use a Gaussian kernel with standard deviation of 40 days.) Our model is

$$\mu(t) = \sum_{\ell=1}^{12} x_{\ell} k(t - v_{\ell}).$$

The x_{ℓ} ’s are modeled as unknowns to be fit from the data. A priori,

$$x_1, \dots, x_{12} \sim \text{iid } \mathcal{N}(0, \sigma_x^2).$$

The posterior means of the x_{ℓ} ’s yields the estimate $\hat{\mu}(t) = \sum \hat{x}_{\ell} k(t - v_{\ell})$ plotted in Figure 9.

5.2 Deviations From μ

Because atmospheric conditions vary from year to year, the actual cycle for M in a particular year may differ from μ , especially in the winter. To account for such systematic differences, we allow each winter (November–April) to have its own random effect. The random effect for the winter ending in year t is denoted by $b_{w(t)}$ and modeled as Gaussian.

To account for M ’s greater variability in winter than in summer, we use a piecewise constant variance: σ^2 in the summer and $3\sigma^2$ in the winter. Modeling the variance as piecewise constant is crude but, we believe, effective.

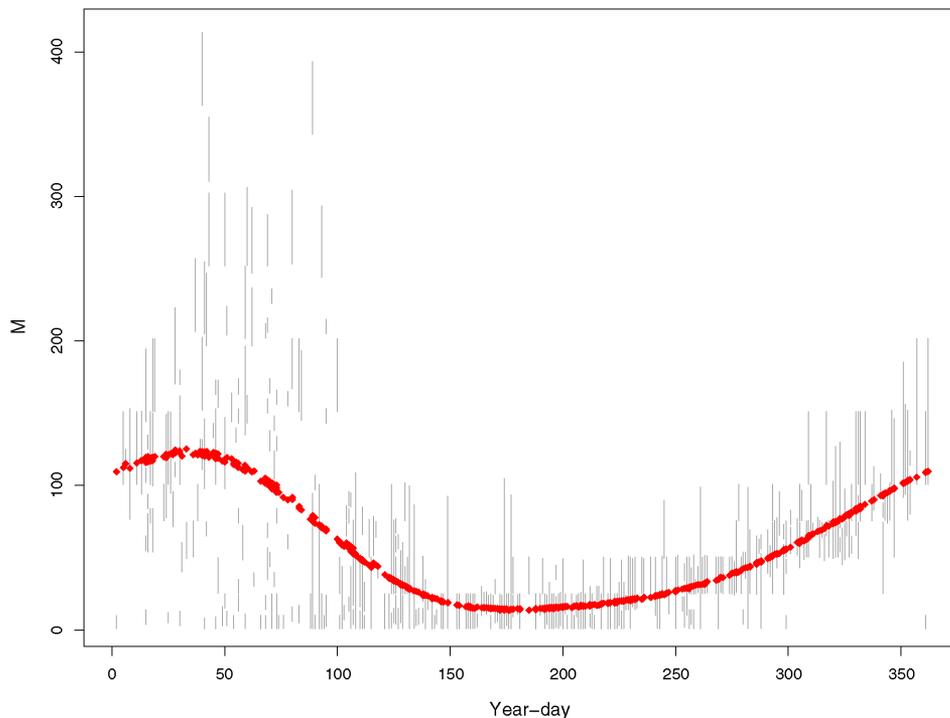


Figure 9. Maximum likelihood intervals of individual profiles (vertical bars) plotted against day of year. The curve is the posterior mean of $\mu(t)$.

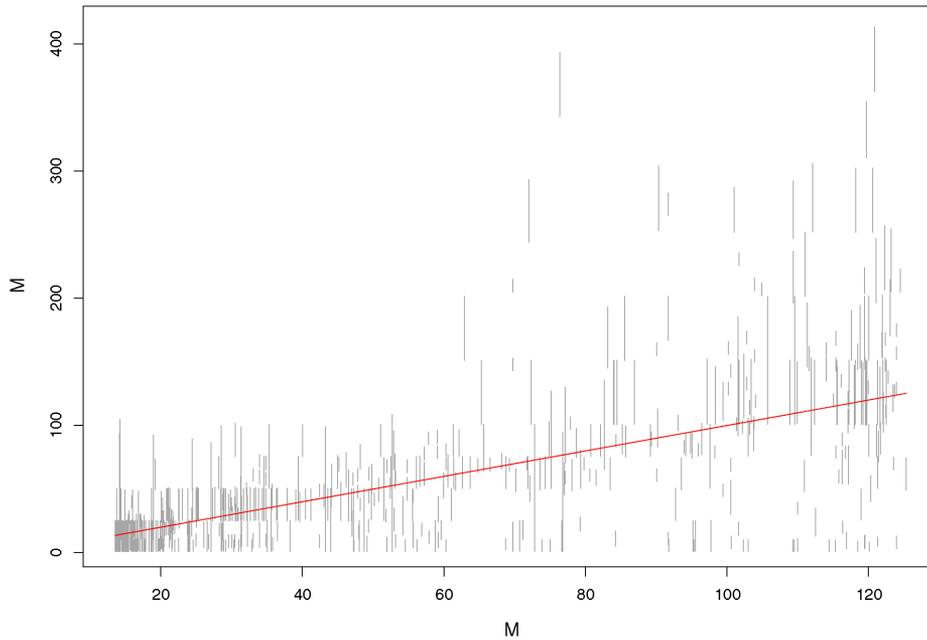


Figure 10. Maximum likelihood interval from (7) versus posterior mean of v_i .

5.3 Linear Trend

In addition to annual variations, we are interested in a possible secular trend. We account for that in our model with a linear regression term βt . The focus of our inference is the posterior distribution of β .

The complete model, including the prior and the subjective likelihood, is given by

$$\begin{aligned}
 &\sigma_x^2, \sigma_b^2, \sigma_\beta^2, \sigma^2 \sim \text{iid InvGam}(.001, .001) \\
 &\text{mutually independent,} \\
 &x_1, \dots, x_{12} \mid \sigma_x^2, \sigma_b^2, \sigma_\beta^2, \sigma^2 \sim \text{iid N}(0, \sigma_x^2), \\
 &\mu(t) = \sum_{\ell=1}^{12} x_\ell k(t - v_\ell), \\
 &\{b_{w(t)}\} \mid \sigma_x^2, \sigma_b^2, \sigma_\beta^2, \sigma^2, \mathbf{x} \sim \text{iid N}(0, \sigma_b^2), \\
 &\beta \mid \sigma_x^2, \sigma_b^2, \sigma_\beta^2, \sigma^2, \mathbf{x}, \{b_{w(t)}\} \sim \text{N}(0, \sigma_\beta^2), \\
 &\nu(t) = \mu(t) + b_{w(t)} + \beta t, \\
 &\tau(t) = \begin{cases} \sigma^2 & \text{if } t \text{ is in summer} \\ 3\sigma^2 & \text{if } t \text{ is in winter,} \end{cases} \\
 &M_{t_i} \mid \sigma_x^2, \sigma_b^2, \sigma_\beta^2, \sigma^2, \mathbf{x}, \mathbf{b}, \beta \sim \text{N}(\nu(t_i), \tau(t_i)) \\
 &\text{mutually independent,} \\
 &\mathbf{y} \mid \sigma_x^2, \sigma_b^2, \sigma_\beta^2, \sigma^2, \mathbf{x}, \mathbf{b}, \beta, \mathbf{M} \sim \text{eq. (7),} \\
 &\text{mutually independent.}
 \end{aligned} \tag{8}$$

The full conditional distributions are available for all parameters except the M_{t_i} 's. Therefore, one efficient method for sampling from the posterior distribution is a Gibbs sampler with Metropolis–Hasting steps for the M_{t_i} 's.

Conditionally on current values of $v(t_i)$, we propose a new move, $M_{t_i}^* \sim \text{U}[0, \text{ocean floor}]$, and accept the move with probability

$$\min \left\{ 1, \frac{\ell(M_{t_i}^*) \exp\{-(M_{t_i}^* - v(t_i))^2 / (2\tau(t_i)^2)\}}{\ell(M_{t_i}) \exp\{-(M_{t_i} - v(t_i))^2 / (2\tau(t_i)^2)\}} \right\}.$$

The grey curve in Figure 9 shows $\hat{\mu}(t) = \sum \hat{x}_\ell k(t - v_\ell)$, the estimate of μ calculated from the posterior means \hat{x}_ℓ . Note that it is asymmetric, as expected from our understanding of the physical process. The asymmetry is partial justification for modeling μ as a process convolution rather than a summation of sinusoids. Figure 10 is another way to view the fit of the model. For each profile, it shows the posterior mean of v_i on the abscissa and the maximum likelihood interval according to (7) on the ordinate. Figure 10 indicates an overall reasonably good fit and heteroscedasticity associated with the deeper mixed layers of winter.

The posterior distribution of β is shown in Figure 11. The posterior mean corresponds to a slight shoaling (or upward) trend in M of approximately .06 meters per year, which translates to about 2.4 m over 40 years. But there is high posterior uncertainty, and our analysis does not give clear support to a secular trend in either direction of the depth of the winter mixed layer near Bermuda over the past 50 years. Further study on the response of the mixed layer depth to interannual and decadal climate variability is ongoing.

6. DISCUSSION

6.1 Foundations

We like our subjective likelihood approach in this analysis because it incorporates expert judgment in the most direct way that we can imagine. We find it an elegant and practical solution to a difficult modeling problem. Of course, it does not have the

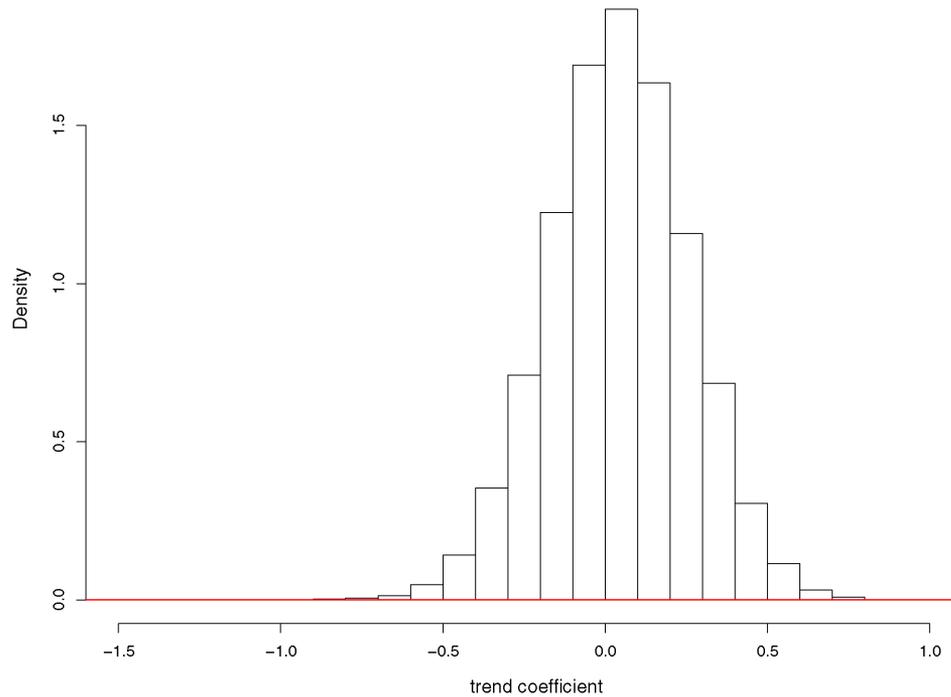


Figure 11. Posterior distribution of β : draws from the MCMC sampler.

usual philosophical underpinnings of the customary Bayesian analysis. It also raises the following question: If the oceanographer can give her posterior for M_{t_i} after seeing one profile, then why cannot she give her posterior for β after seeing them all? We think the answer is obvious. First, studying hundreds of profiles in detail (thousands when we analyze the entire Atlantic Ocean) is too difficult, and second, carefully assimilating information from all of the profiles into an opinion about β is far removed from oceanographers' experience.

We could have analyzed the data with a nonparametric changepoint model in which temperature is roughly constant down to a depth M , then decreases monotonically but otherwise nonparametrically to the ocean floor. We tried and rejected such an approach for two reasons: (a) it yields results for single profiles that disagree with the expert's opinions, and (b) it is a more circuitous use of expert opinion than the subjective likelihood function. Details of this approach will be presented elsewhere.

One troubling aspect of subjective likelihood is the strength of its reliance on expert judgment. Although traditional subjective Bayesian analysis relies on expert judgment for the prior, the influence of that judgment diminishes as more data are collected. But with subjective likelihood, the expert's judgment enters anew with each new data point. Therefore, sensitivity analysis becomes that much more incumbent on the analyst. We note here that the posterior for β does not depend strongly on (a, b, θ) . Similarly, our agreement that $g^* = \max(g, .01g)$ ensures that the posterior will not be too heavily influenced by a small number of profiles. Nonetheless, more thorough model diagnostics are useful and will be reported elsewhere.

One pleasing aspect of subjective likelihood is the strength of its reliance on expert judgment. Many statistical analyses

treat the sampling distribution $p(\text{data} \mid \text{parameters})$ and thus the likelihood without much comment, others assess sensitivity, and still others rely on the central limit theorem to confer robustness. In contrast, subjective likelihood places subjective assessment of the likelihood front and center where it can be critiqued, analyzed, and reanalyzed. This is akin to the usual subjective Bayesian treatment of the prior, which subjective Bayesians claim to be desirable. (For another explicit treatment of Bayesian sensitivity to the sampling distribution, see Lavine 1991.)

6.2 Modeling

In modeling the trend and temporal dependencies, we made simplifying choices, believing that more sophistication would not change the analysis very much. We acknowledge that improvements are possible.

Our model treats the profiles as conditionally independent given \mathbf{M} . This is a questionable choice, however. Our expert says that the temporal decorrelation scale of the mixed layer, at least for this purpose, can be considered the time scale associated with synoptic-scale weather patterns, approximately 5 days. Profiles taken on consecutive days are likely created and influenced by almost the same physical forces and are certainly dependent. Whether they are conditionally dependent given M is another matter. In any case, there are few profiles in our data set at time intervals of less than a week, so we are willing to treat them as conditionally independent. A possible model enhancement is to use a temporal covariance function with a range on the order of about 5 days.

Our model for variance is rather crude: σ^2 in the summer and 3σ in the winter. We tried a smoother model in which σ^2

changes linearly from summer to winter and back again. The smoother model makes little difference to the posterior distribution of β . Nonetheless more sophisticated modeling is possible although we do not believe it will change the result.

Another way to improve the model is to include more covariates. Because M is related to air temperature, we might try including covariates, such as a running mean monthly temperature. But temperature covariates will be highly correlated with time of year, so the potential for improved fit might be only minor. M is also thought to be related to the NAO (North Atlantic oscillation) index, which might be another good covariate to try.

6.3 Diagnostics and Robustness

Figures 9 and 10 show overall goodness of fit. Deviations from the fitted curves are somewhat difficult to interpret, because the likelihood function is not symmetric and need not even be unimodal. For example, Figures 9 and 10 both show three profiles in which the interval of highest probability is deeper than 300 m, but yet the fitted value for one is around 80 m and those for the other two are around 120 m. The model appears not to fit those profiles well. What Figures 9 and 10 do not reveal is that each of those profiles has a secondary probability mode, according to (4), so the apparent misfit is not so severe after all. Other profiles taken around the same dates as these three have only shallower modes; thus the model favors the shallower modes in its posterior mean.

Determination of M can sometimes be complicated by daily warming of surface waters. During a hot, sunny day, surface waters can warm as much as perhaps a couple of degrees down to a depth of perhaps 20 m or so, then cool again during the night. Consequently, a profile taken in mid-afternoon might incorrectly indicate that M appears to be <20 m. Our expert oceanographer was aware of this possibility when assessing the profiles that we showed her. The coefficient of Δ_1 in (5) (i.e., $a = 2$) reflects her belief that temperature drops in the upper ocean are probably due to mixed layers, not to daily warming. But perhaps she is overly confident on this point. As a robustness check, we recalculated our posterior using a coefficient of

$a = .15$ and found that it made little difference to the inference regarding trend.

6.4 More Oceanographers

The analysis presented here relies on the judgment of a single oceanographer. ("Single" is not strictly true; we relied on two oceanographers, one of whom is the doctoral student of the other.) But of course the world contains multiple oceanographers with potentially different interpretations of thermal profiles. What are we to do with them? The answer depends on our purpose. If our purpose is to produce this oceanographer's posterior, then we ignore the others.

But if (as is actually the case) our purpose is to contribute persuasively to scientific and public discourse on ocean stratification and global warming, then we cannot ignore the others. At the very least, we should reveal this oceanographer's judgments more fully so they can be assessed by others. We hope to do this elsewhere. In addition, we could perform parallel analyses with other oceanographers' judgments to see whether conclusions about β change, or we could work with a team of oceanographers to create either a single consensus analysis or a suite of analyses to represent the range of judgments in the wider oceanographic community. In any case, these issues are well known to Bayesian statisticians who elicit priors. We do not believe that subjective likelihood raises new philosophical issues related to the existence of multiple experts.

[Received March 2006. Revised October 2006.]

REFERENCES

- Higdon, D. (1998), "A Process-Convolution Approach to Modelling Temperatures in the North Atlantic Ocean," *Environmental and Ecological Statistics*, 5, 173–190.
- Lavine, M. L. (1991), "Sensitivity in Bayesian Statistics: The Prior and the Likelihood," *Journal of the American Statistical Association*, 86, 396–399.
- Levitus, S., Antonov, J. I., Boyer, T. P., and Stephens, C. (2000), "Warming of the World Ocean," *Science*, 294, 840–842.
- Levitus, S., Antonov, J. I., Wang, J., Delworth, T. L., Dixon, K. W., and Broccoli, A. J. (2001), "Anthropogenic Warming of Earth's Climate System," *Science*, 292, 267–270.
- Mellor, G. L. (1996), *Introduction to Physical Oceanography*, Woodbury, NY: American Institute of Physics.

Comment

Robert N. MILLER

First, I commend the authors for their original approach to an important problem. I agree with their premise that diagnostic study of long-term changes in the mixed-layer depth M will increase our understanding of potential physical and biological consequences of global warming. This is very interesting work with very interesting results obtained by methods not often found in the physical oceanography literature. These methods should be applicable to other data sets gathered in oceans

and other parts of the world. This work is most welcome and deserves exposure in the physical oceanography community.

The authors point out correctly that the mixed layer is not always so clearly defined as it is in figure 3. They illustrate this with the examples in figure 4. In fact, there are examples that are even worse: temperature inversions (i.e., cases in which warmer water lies below cooler water) are not unknown, and thus tem-

Robert N. Miller is Professor, College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, OR 97331 (E-mail: miller@coas.oregonstate.edu).