AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Oceans

RESEARCH ARTICLE

10.1002/2016JC012506

Comparison of full and empirical Bayes approaches for inferring sea-level changes from tide-gauge data

Christopher G. Piecuch[1], Peter Huybers[2], and Martin P. Tingley[3]

[1]Atmospheric and Environmental Research, Inc., Lexington, Massachusetts, USA, [2]Harvard University, Cambridge, Massachusetts, USA, [3]Pennsylvania State University, University Park, Pennsylvania, USA

Key Points:
- We compare full and empirical Bayes methods for inferring sea-level changes from gappy, noisy, biased tide-gauge records
- Empirical Bayes underestimates uncertainty on sea-level changes from tide gauges whereas full Bayes gives more reliable credible intervals
- The full Bayes method is generally preferable to the empirical Bayes approach for sea-level reconstruction

Supporting Information:
- Supporting Information S1
- Figure S1
- Figure S2
- Figure S3
- Figure S4
- Figure S5
- Figure S6
- Figure S7
- Figure S8
- Figure S9
- Figure S10
- Figure S11
- Figure S12

Correspondence to:
C. G. Piecuch,
cpiecuch@aer.com

Citation:
Piecuch, C. G., P. Huybers, and M. P. Tingley (2017), Comparison of full and empirical Bayes approaches for inferring sea-level changes from tide-gauge data, *J. Geophys. Res. Oceans*, *122*, 2243–2258, doi:10.1002/2016JC012506.

Received 25 OCT 2016
Accepted 21 FEB 2017
Accepted article online 28 FEB 2017
Published online 17 MAR 2017

Abstract Tide-gauge data are one of the longest instrumental records of the ocean, but these data can be noisy, gappy, and biased. Previous studies have used empirical Bayes methods to infer the sea-level field from tide-gauge records but have not accounted for uncertainty in the estimation of model parameters. Here we compare to a fully Bayesian method that accounts for uncertainty in model parameters, and demonstrate that empirical Bayes methods underestimate the uncertainty in sea level inferred from tide-gauge records. We use a synthetic tide-gauge data set to assess the skill of the empirical and full Bayes methods. The empirical-Bayes credible intervals on the sea-level field are narrower and less reliable than the full-Bayes credible intervals: the empirical-Bayes 95% credible intervals are 42.8% narrower on average than are the full-Bayes 95% credible intervals; full-Bayes 95% credible intervals capture 95.6% of the true field values, while the empirical-Bayes 95% credible intervals capture only 77.1% of the true values, showing that parameter uncertainty has an important influence on the uncertainty of the inferred sea-level field. Most influential are uncertainties in model parameters for data biases (i.e., tide-gauge datums); letting data-bias parameters vary along with the sea-level process, but holding all other parameters fixed, the 95% credible intervals capture 92.8% of the true synthetic-field values. Results indicate that full Bayes methods are preferable for reconstructing sea-level estimates in cases where complete and accurate estimates of uncertainty are warranted.

1. Introduction

Tide-gauge sea-level data represent one of the longest instrumental records of the ocean, in some locations dating to the eighteenth century [e.g., *Ekman*, 1988; *Woodworth*, 1999; *Bogdanov et al.*, 2000; *Woodworth et al.*, 2010]. Tide-gauge data are informative for ocean circulation and climate studies but pose unique challenges from the perspective of data analysis. Affixed to land, tide gauges provide point-referenced data, and thus records are influenced by crustal motion and local hydrodynamics that can obscure regional or global ocean behavior. The records are also sparsely distributed along the coast and can feature intermittently missing values, giving an incomplete picture of the evolution of the sea-level field in space and time.

Synthesizing tide-gauge data within a coherent statistical framework has proven useful for reconstructing historical sea-level changes [e.g., *Hay et al.*, 2015]. But, correcting for, and representing the uncertainties in, the multiple potential sources of variability in the data not directly linked to ocean-volume changes (e.g., datum changes, land subsidence, and postglacial rebound) remains challenging. Some authors have used probabilistic Bayesian methods to infer sea level from tide gauges, casting sea level as a field with spatiotemporal covariance and interpreting the data as noisy, gappy, biased versions of the sea-level field [e.g., *Hay et al.*, 2013; *Kopp*, 2013; *Kopp et al.*, 2014; *Hay et al.*, 2015; *Kopp et al.*, 2016]. These methods are typically *empirically Bayesian*, in the sense that the model parameters (e.g., describing the scales of spatiotemporal covariance) are regarded as known quantities, and assigned point values [e.g., *Carlin and Louis*, 2000; *Cressie and Wikle*, 2011]. However, in reality, estimates of model parameters feature uncertainties that may not be negligible. In contrast, *fully Bayesian* approaches treat model parameters as unknowns, wherein they are assigned prior distributions, and solved for along with the process or field of interest [e.g., *Gelman et al.*, 2004; *Cressie and Wikle*, 2011]. Fully Bayesian models have been applied to various data-assimilation and inverse-modeling problems in ocean science, such as inferring surface winds over the Mediterranean Sea and equatorial Pacific [*Wikle et al.*, 2001; *Milliff et al.*, 2011; *Pinardi et al.*, 2011], describing the hydrography

of the South Atlantic Ocean [*McKeague et al*., 2005; *Herbei et al*., 2008], predicting sea-surface temperatures over the tropical Pacific [*Berliner et al*., 2000], and interpreting ocean-color retrievals [*Harmon and Challenor*, 1997]. See *Wikle et al*. [2013] for a more comprehensive review.

The relative merits of empirical and full Bayes approaches have been generally discussed in textbooks [e.g., *Cressie and Wikle*, 2011], as well as more particularly in applied studies, for example, in disease mapping [e.g., *Bernardinelli and Montomoli*, 1992; *Heisterkamp et al*., 1993; *Louis and Shen*, 1999; *Leyland and Davis*, 2005; *Ashby*, 2006] and traffic safety [e.g., *Miaou and Lord*, 2003; *Carriquiry and Pawlovich*, 2005; *Huang et al*., 2009; *Lan et al*., 2009; *Persaud et al*., 2010]. These works demonstrate that empirical Bayes methods tend to overestimate the precision (i.e., underestimate the uncertainty) on the process or field of interest, producing credible intervals that are too narrow, due to their neglect of uncertainty in the parameters. These studies reveal that fully Bayesian frameworks, which account for more sources of uncertainty, provide more representative credible intervals.

Dedicated comparisons between empirical and full Bayes approaches do not appear in the sea-level literature. Although previous results [e.g., *Cressie and Wikle*, 2011] lead to the general expectation that empirical Bayes methods will underestimate uncertainty and that full Bayes approaches will give more reliable credible intervals, the precise degree to which empirically Bayesian methods underestimate the uncertainties on the sea-level field determined from tide-gauge records is unclear, as is where exactly differences would arise in the context of a fully Bayesian framework. Given the use of both empirical and full Bayesian methods in the literature, it appears useful to make an explicit comparison of these approaches. To this end, we develop a fully Bayesian hierarchical model to infer the sea-level field from tide-gauge data and compare its output to a reduced version that is equivalent to an empirical Bayes approach.

Hierarchical models [*Gelman et al*., 2004; *Wikle and Berliner*, 2007; *Cressie and Wikle*, 2011; *Tingley et al*., 2012] are based on the notion of conditional probabilities: data are regarded as conditional upon a latent (hidden) spatiotemporal process, whose evolution is determined by various parameters, the exact numerical values of which are unknown. This conditionality is inverted using Bayes' theorem, yielding the posterior distribution of the process and the parameters given the observations. Solutions are generated using Markov chain Monte Carlo methods and constitute an ensemble of estimates of the process and the parameters—a set of possible histories, each consistent with and equally likely given the observations.

Here we assess the merits of the full Bayes approach in the context of tide-gauge records, comparing to an empirical Bayes method. We consider annual tide-gauge data along the North American northeast coast, applying our algorithm in its full hierarchical mode, iteratively solving for the process and parameters, and in a reduced empirical mode, solving only for the sea-level process based on assigned point values for model parameters. Based on a known but corrupted synthetic field, we test the ability of the two methods to infer the true process—with appropriate uncertainties—from noisy, gappy, biased data. The remainder of the paper is structured as follows: in section 2, we describe the data; in section 3, we develop the Bayesian algorithm; in section 4, we compare full and empirical Bayes implementations; in section 5, we discuss our findings. More technical and diagnostic aspects of the algorithm and its solution are presented in Supporting Information.

2. Tide-Gauge Data

We use 36 tide-gauge records on the North American northeast coast (Table 1 and Figure 1). The annual data were extracted from the Permanent Service for Mean Sea Level (PSMSL) Revised Local Reference (RLR) database [*Holgate et al*., 2013; *Permanent Service for Mean Sea Level (PSMSL)*, 2016] on 14 March 2016. Each individual data record has more than 25 years of annual values. Together they comprise 2437 gauge-years of observations over 1893–2015. As necessary context for the development of our hierarchical algorithm, we briefly explore the characteristics of sea-level changes and tide-gauge records along the coast.

2.1. Space and Time Scales of Sea-Level Changes

Sea-level changes along the northeast coast can show nonrandom structure in space and time. Spatial gradients in temporal trends from long (centennial) tide-gauge records in this area have been interpreted in terms of vertical crustal motions due to postglacial rebound [e.g., *Davis and Mitrovica*, 1996; *Miller and Douglas*, 2006; *Kopp*, 2013] as well as groundwater withdrawal [*Miller et al*., 2013; *Karegar et al*., 2016]. Using

Table 1. Annual RLR PSMSL [*Holgate et al.*, 2013] Tide-Gauge Sea-Level Time Series Used in This Study[a]

Station Name	Longitude	Latitude	Timespan (Completeness)	Datum
Duck Pier outside	−75.7467	36.1833	1985–2015 (90%)	1985
Portsmouth (Norfolk Navy Yard)	−76.2933	36.8217	1936–1986 (100%)	1973
Sewells Point, Hampton Roads	−76.33	36.9467	1928–2015 (100%)	1964
Chesapeak Bay Br. Tun.	−76.1133	36.9667	1985–2015 (100%)	1986
Kiptopeke Beach	−75.9883	37.165	1952–2015 (97%)	1973
Gloucester Point	−76.5	37.2467	1951–2002 (88%)	1967
Solomon's Island (Biol. Lab.)	−76.4517	38.3167	1938–2012 (95%)	1965
Cambridge II	−76.0683	38.5733	1971–2015 (96%)	1974
Lewes (Breakwater Harbor)	−75.12	38.7817	1919–2015 (71%)	1971
Washington DC	−77.0217	38.8733	1931–2015 (98%)	1971
Cape May	−74.96	38.9683	1966–2015 (90%)	1966
Annapolis (Naval Academy)	−76.48	38.9833	1929–2015 (94%)	1974
Baltimore	−76.5783	39.2667	1903–2015 (99%)	1964
Atlantic City	−74.4183	39.355	1912–2015 (88%)	1964
Philadelphia (Pier 9N)	−75.1417	39.9333	1901–2015 (96%)	1971
Sandy Hook	−74.0083	40.4667	1933–2015 (95%)	1964
New York (The Battery)	−74.0133	40.7	1856–2015 (89%)	1964
Willets Point	−73.7817	40.7933	1932–1999 (96%)	1964
Port Jefferson	−73.0767	40.95	1958–1990 (94%)	1973
Montauk	−71.96	41.0483	1948–2015 (84%)	1964
Bridgeport	−73.1817	41.1733	1965–2014 (92%)	1972
Nantucket Island	−70.0967	41.285	1965–2015 (92%)	1973
New London	−72.09	41.36	1939–2015 (94%)	1964
Newport	−71.3267	41.505	1931–2015 (98%)	1964
Woods Hole (Ocean. Inst.)	−70.6717	41.5233	1933–2015 (90%)	1972
Providence (State Pier)	−71.4	41.8067	1939–2015 (83%)	1974
Boston	−71.0533	42.3533	1921–2015 (98%)	1964
Seavey Island	−70.7417	43.08	1927–1985 (85%)	1972
Portland (Maine)	−70.2467	43.6567	1912–2015 (99%)	1964
Yarmouth	−66.1333	43.8333	1967–2014 (73%)	1986
Bar Harbor, Frenchman Bay	−68.205	44.3917	1948–2015 (85%)	1964
Halifax	−63.5833	44.6667	1896–2012 (74%)	1988
Eastport	−66.9817	44.9033	1930–2015 (87%)	1964
Saint John, N.B.	−66.0667	45.2667	1897–2014 (63%)	1991
North Sydney	−60.25	46.2167	1970–2014 (91%)	1991
Charlottetown	−63.1167	46.2333	1912–2014 (68%)	1992

[a]The completeness is percentage of timespan for which valid annual data are available. RLR datum is the year to which the respective tide-gauge record is referenced to [*PSMSL*, 2016].

climate models, *Yin and Goddard* [2013] have argued that variable ocean dynamics might also have played a role. Some conclude that trends are higher over the mid-Atlantic states and lower along coastal New England [*Kopp*, 2013], whereas others have reasoned that distribution of the tide-gauge sea-level trends shows no clear spatial structure [*Andres et al.*, 2013].

Over shorter (interannual and decadal) time scales, coherent fluctuations in sea level are apparent along the coast, such that tide-gauge records between Cape Hatteras and Cape Breton are tightly correlated with one another [*Thompson and Mitchum*, 2014; *Woodworth et al.*, 2014; *Piecuch and Ponte*, 2015; *Piecuch et al.*, 2016]. The spatial coherence partly reflects longshore-wind-stress and atmospheric-pressure forcing [*Piecuch and Ponte*, 2015; *Piecuch et al.*, 2016].

Spectral analyses of global sea-level time series show that sophisticated stochastic models (e.g., autoregressive fractionally integrated moving average, generalized Gauss Markov) are often needed to characterize the temporal structure of *weekly* or *monthly* data records [*Hughes and Williams*, 2010; *Bos et al.*, 2014]. However, for the case of *annual* tide-gauge time series, simpler first-order autoregressive models are usually sufficient to describe the spectral nature of the sea-level records [*Bos et al.*, 2014].

2.2. Biases and Errors in Tide-Gauge Records

The imperfections inherent to tide-gauge data must also be considered in model design. Mean sea-level records are made relative to nearby benchmarks, rather than with respect to some universal reference datum [*Higginson et al.*, 2015]. Time series from the PSMSL RLR database are usually offset by values about 7 m below local mean sea level over some year during the data record—a convention adopted in the 1960s to avoid storing negative values in the computers of that time [*Intergovernmental Oceanographic*

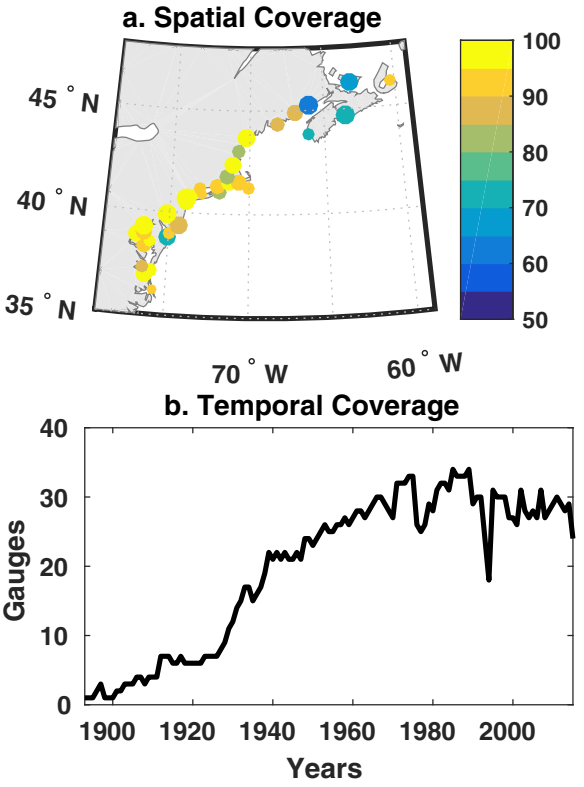

Figure 1. (a) Locations of 36 annual tide-gauge records from the PSMSL RLR database on the North American northeast coast (Table 1). Sizes of the dots correspond to the timespan of the records (larger dots denote longer records). Colors of the dots represent the completeness of the records (cf. Table 1). (b) Number of tide gauges reporting valid annual values against time.

Commission (IOC) of UNESCO, 2006]. It is often desirable to remove the time mean values and express tide-gauge data as anomalies relative to some common reference interval [*Church et al.*, 2004]. But, due to the nonuniform space-time coverage of the data (Figure 1), this can be a difficult task that cannot be accomplished simply by subtracting a value of 7 m from the various records [cf. *Tingley*, 2012].

The tide-gauge data are also subject to errors. Even though the accuracy of an individual tide-gauge reading can be as good as order 1 cm, the accuracy of monthly or annual mean sea-level data—averaged over many individual readings—is at best order 1 mm, due to possible systematic errors [*Pugh and Woodworth*, 2014]. Tide-gauge records can also be affected by "representation" error—local behavior at a location that is uncorrelated with the behavior at other locations [cf. *Ponte et al.*, 2007], which can result from, for example, local river runoff, harbor effects, or trapped circulations.

3. Model Formulation

Our model comprises three levels: a *process level*, wherein the mathematical rules that govern the evolution of the latent process are given; a *data* or *observation level*, wherein a measurement model is specified; and a *prior level*, wherein model parameters are assigned prior distributions [e.g., *Gelman et al.*, 2004; *Cressie and Wikle*, 2011].

3.1. Process Level

Consider sea level at time step $k \in \{1,\ldots,K\}$ and target location $n \in \{1,\ldots,N\}$. The evolution of the latent spatiotemporal process, $\boldsymbol{y}_k = [y_{1,k},\ldots,y_{N,k}]^{\mathsf{T}}$, is represented as a field of temporal trends superimposed on a first-order autoregressive [AR(1)] process,

$$\boldsymbol{y}_k - \boldsymbol{b}t_k = r(\boldsymbol{y}_{k-1} - \boldsymbol{b}t_{k-1}) + \boldsymbol{e}_k. \tag{1}$$

Here $\boldsymbol{b}$ is the spatial vector of temporal trends, t_k is the time at k, r is the AR(1) coefficient, and $\boldsymbol{e}_k$ is the zero-mean temporal white noise vector of spatially correlated innovations, $\boldsymbol{e}_k \sim \mathcal{N}(\boldsymbol{0}_N, \Sigma)$, such that $\boldsymbol{0}_N$ is a $N\times 1$ column vector of zeros, the symbol $\sim$ is read "is distributed as," and $\mathcal{N}(\boldsymbol{u},\mathbf{v})$ is a multivariate-normal distribution with mean $\boldsymbol{u}$ and covariance $\mathbf{v}$. Note that our use of a common value for r, which may seem overly simplistic, is justified given the data, as demonstrated by residual analyses presented in Supporting Information S1.3. Note also that we have centered the set of times $\{t_k\}$ on zero (i.e., $\sum_{k=1}^{K} t_k = 0$) so that the mean of the process (1) vanishes. The innovation vectors are temporally independent, identically distributed (IID) random variables, with exponential spatial covariance,

$$\Sigma_{ij} = \sigma^2 \exp(-\varphi|\boldsymbol{x}_i - \boldsymbol{x}_j|). \tag{2}$$

In (2), σ^2 is the sill (the limit of the semivariogram of the process [*Banerjee et al.*, 2004]), φ is the inverse range, and $|\boldsymbol{x}_i - \boldsymbol{x}_j|$ is the distance separating sites i and j. Note that the trend vector $\boldsymbol{b}$ is cast as a normal field, $\boldsymbol{b} \sim \mathcal{N}(\mu\boldsymbol{1}_N, \pi^2\mathbf{I}_N)$, with mean μ and variance π^2, such that $\boldsymbol{1}_N$ is a $N\times 1$ column vector of ones and $\mathbf{I}_N$ is the $N \times N$ identity matrix. Unlike the innovation sequences $\boldsymbol{e}_k$, the elements of the trend field $\boldsymbol{b}$ are

Table 2. Descriptions of the Model Process and Parameter Terms, Including the Functional Forms of the Priors and the Conditional Posterior Distributions[a]

Parameter	Prior Distribution	Conditional Posterior	Description
$\boldsymbol{y}_0$	MV normal	MV normal	Initial process values
$\boldsymbol{y}_{1,\ldots,K}$		MV normal	Process values
$\boldsymbol{b}$		MV normal	Spatial field of process linear trend
μ	Normal	Normal	Mean value of process linear trend
π^2	Inverse-gamma	Inverse-gamma	Spatial variance of the process linear trend
r	Uniform	Truncated normal	AR(1) coefficient of the process
σ^2	Inverse-gamma	Inverse-gamma	Sill of the process innovations
φ	Lognormal	Nonstandard	Inverse range of the process innovations
δ^2	Inverse-gamma	Inverse-gamma	Instrumental error variance
ℓ		MV normal	Spatial field of observational biases
v	Normal	Normal	Mean value of observational biases
τ^2	Inverse-gamma	Inverse-gamma	Spatial variance in observational biases

[a]See Supporting Information Text S4 and S5 for more details on the priors and conditional posterior distributions. MV stands for multivariate normal and nonstandard indicates a conditional posterior distribution that is not a standard probability distribution [*Gelman et al.*, 2004].

assumed to be spatially independent—a decision justified and discussed in more detail in Supporting Information S1.2.

3.2. Data Level

Given sea-level data from tide gauges at $M_k \leq N$ sites at time step k, we regard the data, $\boldsymbol{z}_k = \left[z_{1,k}, \ldots, z_{M_k,k}\right]^{\mathrm{T}}$, as noisy, biased, and gappy versions of the process,

$$\boldsymbol{z}_k = \mathbf{H}_k \boldsymbol{y}_k + \boldsymbol{d}_k + \mathbf{F}_k \ell. \tag{3}$$

Here $\mathbf{H}_k$ is a $M_k \times N$ selection matrix (populated with zeros and ones) that isolates the process at the measurement sites at time step k, $\boldsymbol{d}_k$ is a vector of temporally IID, spatially uncorrelated normal measurement errors, $\boldsymbol{d}_k \sim \mathcal{N}\left(\mathbf{0}_{M_k}, \delta^2 \mathbf{I}_{M_k}\right)$, ℓ is a vector of temporally IID, spatially independent data biases, $\ell \sim \mathcal{N}(v\mathbf{1}_M, \tau^2 \mathbf{I}_M)$, and $\mathbf{F}_k$ is a $M_k \times M$ selection matrix that picks out the observation biases at the various data sites. The data bias ℓ is identifiable because we have designed the process (1) to have zero mean (see Supporting Information S3). Here $\mathbf{I}_{M_k}$ and $\mathbf{I}_M$ denote the $M_k \times M_k$ and $M \times M$ identity matrices, respectively, where M is the total number of locations ($M_k \leq M \forall k$). Given the sampling distribution (3), the data errors $\boldsymbol{d}_k$ include both instrumental and representational errors.

3.3. Prior Level

To close the model, priors must be given to the model parameters and unobserved initial conditions. Our approach is to use appropriate, diffuse, weakly informative, and (mainly) conjugate forms [*Tingley and Huybers*, 2010]. Table 2 lists the prior distributions and Supporting Information S4 explains our rationale for choosing those forms.

3.4. Drawing Samples From the Posterior Distribution

We use Bayes' rule to express the posterior distribution of the process and parameters conditioned on the data in terms of the likelihood of the data given the process and the parameters and the priors of the parameters,

$$\begin{aligned} p(\boldsymbol{y}, \Theta | \boldsymbol{z}) &\propto p(\boldsymbol{z} | \boldsymbol{y}, \Theta) \cdot p(\boldsymbol{y}, \Theta), \\ &= p(\boldsymbol{z} | \boldsymbol{y}, \Theta_z) \cdot p\left(\boldsymbol{y} | \Theta_y\right) \cdot p(\Theta_z) \cdot p\left(\Theta_y\right), \\ &= p(\Theta_z) \cdot p\left(\Theta_y\right) \cdot p(\boldsymbol{y}_0) \cdot \prod_{k=1}^{K} \left[p(\boldsymbol{z}_k | \boldsymbol{y}_k, \Theta_z) \cdot p\left(\boldsymbol{y}_k | \boldsymbol{y}_{k-1}, \Theta_y\right)\right]. \end{aligned} \tag{4}$$

Here $\Theta_z \doteq \left\{\delta^2, \ell, v, \tau^2\right\}$ is the set of data parameters. We assume the breakdown,

$$p(\Theta_z) = p\left(\delta^2\right) \cdot p\left(\ell | v, \tau^2\right) \cdot p(v) \cdot p\left(\tau^2\right). \tag{5}$$

Similarly, $\Theta_y \doteq \{r, \sigma^2, \varphi, \boldsymbol{b}, \mu, \pi^2\}$ is the set of process parameters, where we assume,

$$p(\Theta_y) = p(r) \cdot p(\sigma^2) \cdot p(\varphi) \cdot p(\boldsymbol{b}|\mu, \pi^2) \cdot p(\mu) \cdot p(\pi^2). \quad (6)$$

Above, p is probability density, $|$ is conditionality, $\propto$ is proportionality, and $\Theta \doteq \{\Theta_y, \Theta_z\}$.

Markov chain Monte Carlo (MCMC) methods are used to evaluate conditional posterior distributions of the process and parameters. We use Gibbs sampling to draw values of the process and parameters with conjugate priors and a Metropolis step to draw values of the inverse range parameter (φ). See *Gelman et al*. [2004, chap. 11] for more discussion of these approaches. The full conditional posterior distributions are given in Supporting Information S5. We perform 200,000 MCMC iterations. Initial process values are set to zero, and initial parameter values are randomly drawn from the priors. The first 100,000 warm-up draws are discarded to eliminate influences from initialization transients. Further, to reduce the influence of serial correlation, sequences are thinned, with only 1 out of every 100 samples kept. The convergence of the chains is monitored by generating multiple sequences starting from dispersed initial conditions: after performing the above procedure three times (giving three separate 1000 draw sequences for each estimand), we compare the variances between and within the sequences using the convergence monitor factor ($\hat{R}$) from *Gelman et al*. [2004].

Our primary purpose is to compare empirical and full Bayes methods for inferring the sea-level field from tide-gauge data. For the interested reader, a more complete presentation and evaluation of the fully Bayesian solution is given in Supporting Information S1 and S2.

3.5. Sources of Variance in the Posterior Field Estimates

Variance in the posterior estimates of the sea-level field $\boldsymbol{y}_k$ derives from multiple sources. The noise vectors $\boldsymbol{e}_k$ and $\boldsymbol{d}_k$ in the process and data equations both contribute to posterior variance in the sea-level field estimates (equations (S5)–(S8)). We refer to this stochastic source of variance in the posterior field as *residual variability* after *Kennedy and O'Hagan* [2001]. This definition of residual variability incorporates observational error represented by $\boldsymbol{d}_k$. Were model parameters known and fixed, residual variability would be the only source of variance in the posterior field solutions. Yet model parameters are unknown and variable, thus introducing another source of variance into the posterior field estimates, which we will call *parameter uncertainty*. Note that this partitioning of the variance in the posterior distributions does not amount to a comprehensive quantification of the uncertainty in the model estimate, since we do not consider structural uncertainties with respect to the data.

As described below, residual variability and parameter uncertainty can be evaluated by running the Bayesian model in its full hierarchical mode (wherein parameters are variable and both residual variability and parameter uncertainty contribute to posterior variance in the sea-level field) as well as in reduced empirical mode (wherein the parameters are set to constants and only residual variability contributes to the posterior field variance), and examining differences between the posterior solutions. Residual variability and parameter uncertainty may not be linear and separable and could involve nonlinear interactions, but differences between the full and empirical solutions nevertheless give a sense of the influence of parameter uncertainty.

In the results that follow, we show that, in the absence of data, the variance in the posterior field estimates is mainly controlled by residual variability, whereas it is mostly governed by parameter uncertainty in the presence of observations. Also, we reveal that to generate reliable posterior solutions of the sea-level field (with meaningful error estimates), it is essential to account for uncertainty in the data-bias parameters $\{\ell, v, \tau^2\}$.

3.6. Implementing the Algorithm in Reduced Empirical Mode

In addition to running our model in full hierarchical mode, iteratively solving for process and parameters, we also consider a "reduced" empirical mode, using fixed parameter values and solving only for the process. For the reduced mode, we set model parameters equal to their most likely values based on the posterior distributions determined from the former fully Bayesian case (see Supporting Information S1 and S2). Our choice to define model parameters in terms of their posterior modes is similar to *Kopp* [2013], who uses maximum likelihood estimation for the model parameters. Although we could have used any number of other strategies for setting the parameters (e.g., selecting the parameter vector sample that minimizes the Mahalanobis distance to the mean vector), the salient point here is not so much *how* we fix the parameters,

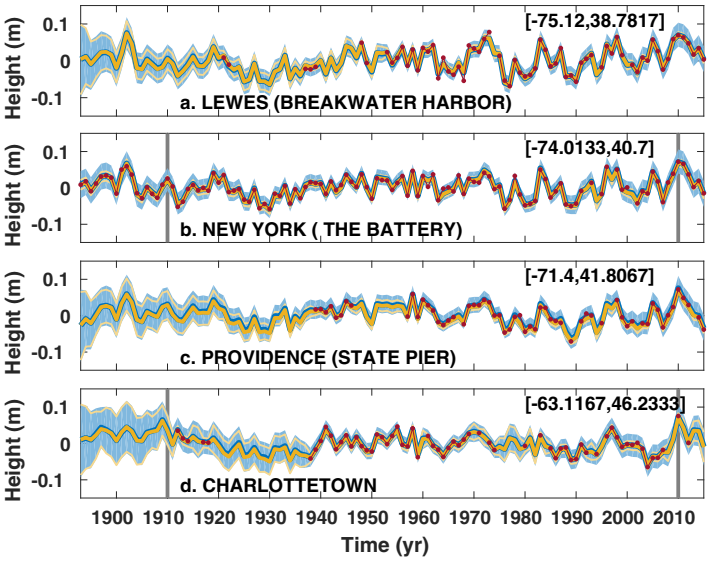

Figure 2. Detrended sea-level time series from example tide-gauge locations (indicated by black circles in Figure 3a). The raw tide-gauge records (offset by the median bias estimate and with the median trend estimate removed; Figures S1 and S2) are shown in red dash dot. The blue solid lines and light blue shading are the median estimates and 95% credible intervals of the sea-level process, respectively, from the fully Bayesian hierarchical model. The thicker and thinner yellow lines are the median estimates and 95% credible intervals of the sea-level process, respectively, from the reduced empirically Bayesian algorithm. Gray vertical lines at 1910 and 2010 in Figures 2b and 2d indicate the locations and times of the histograms shown in Figure 4.

but rather *that* we fix them at all. Through choosing sensible values for the parameters, we are able to evaluate how uncertainty in model parameters affects uncertainty in the posterior sea-level field estimates.

4. Comparison Between Empirical and Full Bayesian Methodologies

Since the empirical Bayes method neglects parameter uncertainty, the resulting credible intervals are anticipated to be too narrow [*Cressie and Wikle*, 2011]. To examine the extent of underestimation, we compare empirical and full Bayes approaches to inferring sea level from tide-gauge data. Figure 2 shows the estimated sea-level process, with uncertainties, from the Bayesian models at several example locations. The corresponding tide-gauge records at those locations are also shown for comparison. (For a better visualization of the uncertainties, linear trends have been removed from the time series.) When the data are available, they are generally in excellent agreement with the posterior solutions, signifying that the model regards the data as good estimates of the process. When the data are unavailable, the credible intervals grow, because the algorithm must use information from adjacent points in space and time to impute values into the record. The uncertainties are generally larger earlier in time, consistent with there being fewer observations earlier on in the record. For example, full Bayes 95% credible interval widths are about 80 mm in 1910 and 51 mm in 2010 on average (Figure 3). However, the uncertainties do not always decrease monotonically as time increases, on account of intermittently missing data values. For instance, 95% credible intervals grow larger at Charlottetown, Prince Edward Island, during 1975–1979 when there is a gap in that record (Figure 2d).

Median process solutions produced by the full and empirical Bayes models are essentially identical. Correlations between the median full and empirical process estimates are very high ($>$0.99 with or without trends removed), and regression coefficients between full and empirical medians are consistent with a value of unity (i.e., between 0.98 and 1.03). The credible intervals produced by the two methods show analogous *qualitative* behavior. Errors are larger earlier on in time, or in the absence of data, whereas errors are smaller later in time and in the presence of data. But, they are distinct *quantitatively* in that the credible intervals from the empirical Bayes approach are smaller than from the full Bayesian methodology. The uncertainty estimates from both methods tend to be larger in absolute terms earlier on in time, but, due partly to increasing data scarcity, discrepancies between credible intervals from the two approaches become *relatively* more marked later in the instrumental record. Such differences can be gleaned from the time series in Figure 2, but are more apparent in Figure 4, where we show example histograms of the full and empirical Bayesian estimates for detrended sea level at two tide-gauge sites (New York and Charlottetown) and two points in time (1910 and 2010). At New York in 1910 and 2010 and Charlottetown in 2010, the empirical Bayes histograms are much narrower and more strongly peaked than the full Bayes histograms, whereas at Charlottetown in 1910, both empirical and full Bayes histograms are comparably wide and diffuse. In the former instances, the tide-gauges feature observations, while in the latter instance there is a data gap (cf. Figure 2). Across all locations, the 95% credible intervals on the process from the empirical Bayes solution,

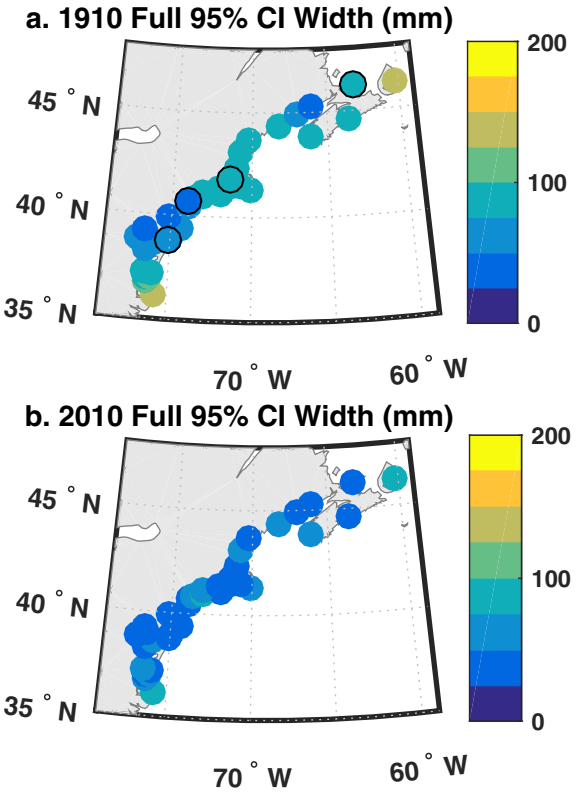

Figure 3. Widths of the 95% credible interval on the process from the full Bayesian model in (a) 1910 and (b) 2010. Black circles in Figure 3a indicate the four example tide-gauge locations shown in Figure 2.

which assumes parameters are known and uses fixed values for them, are about 23 and 56% narrower on average in 1910 and 2010, respectively, than from the full Bayes solution, which regards parameters as unknown and iteratively solves for them (Figure 5). More generally, when data are available, parameter uncertainty (captured only by the full Bayes approach) is comparatively more important than residual variability, while when data are not available, residual variability (captured by both full and empirical Bayes approaches) is relatively more important (cf. section 3.5).

Both models represent the same process and involve the same data, and so differences between their credible intervals (Figures 2, 4 and 5) must imply that one or both are unreliable in the sense that the credible interval corresponding to probability P does not contain the same fraction P of true values, being thus under or overdispersive [*Raftery et al.*, 2005]. Given results in other fields [*Bernardinelli and Montomoli*, 1992; *Miaou and Lord*, 2003], we anticipate that the empirical Bayes algorithm is underdispersive, yielding too narrow credible intervals (though the full Bayes model could also give unreliable credible intervals if the model is

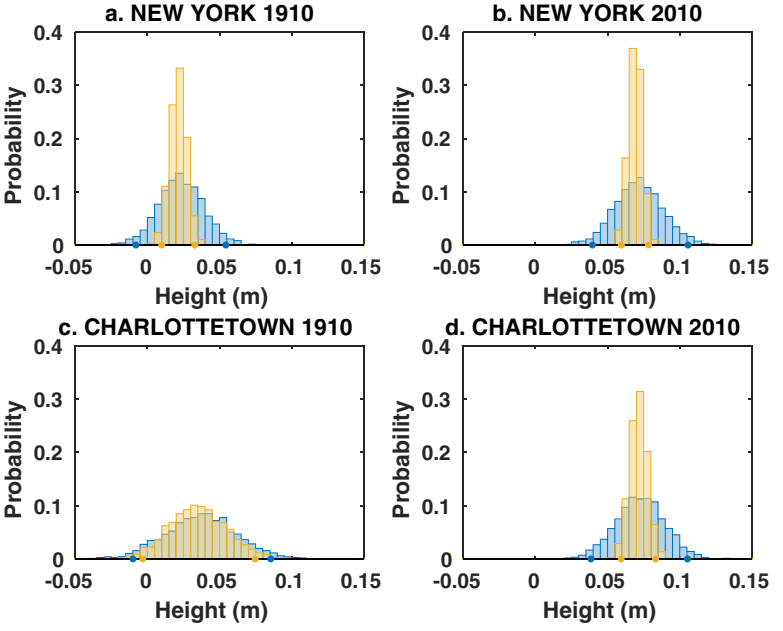

Figure 4. Histograms of estimated detrended sea level from the full (blue) and empirical Bayes model (yellow) for New York (top) and Charlottetown (bottom) in 1910 (left) and 2010 (right; cf. gray vertical lines in Figures 2b and 2d). Blue and yellow dots along the horizontal axis represent the bounds of the 95% credible intervals from the full and empirical Bayes models, respectively. Observations are available at the location and time in Figures 4a, 4b, and 4d, but not available in Figure 4c.

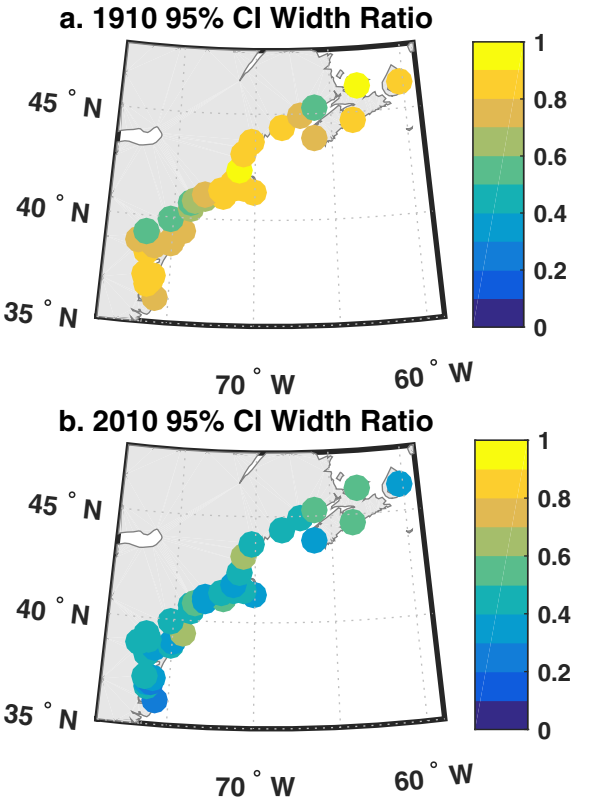

Figure 5. Widths of the 95% credible interval on the process from the empirical Bayes model divided by the corresponding widths of the 95% credible interval on the process from the full Bayes model (i.e., the values shown in Figure 3) in (a) 1910 and (b) 2010.

misspecified relative to the observations). We now turn to a numerical experiment to evaluate the reliability of the two approaches. As previously, we apply the model in its full hierarchical Bayesian and reduced empirical Bayesian implementations. The difference is that now we apply the model to corrupted data of a known synthetic process. Using parameter solutions from the full Bayes solution applied to tide gauges (Tables S1 and S2 and Figures S1 and S2), we use the process field equation (1) to create a synthetic process at the tide-gauge sites. The perfectly known synthetic fields are then corrupted by imposing noise and bias according to the measurement model (3), with gaps imparted as with the data (Figure 1). (See Supporting Information S2 for a more detailed description of the generation of the synthetic data set.) We then apply the Bayesian algorithms to the synthetic data set. Because the true process field values are known, we can evaluate the skill of the algorithms for determining the true field values from corrupted observations.

We show time series of model solutions and corrupted observations from the synthetic data analysis at a few example sites in Figure 6. Also shown are time series of the true values. The median model solutions closely follow the corrupted data and the true process. The empirical-Bayes credible intervals are narrower than the full-Bayes credible intervals, especially when observations are available. To illustrate differences in credible intervals in more detail, we show histograms of full and empirical-Bayes model solutions at a couple locations and time points, now alongside the true process values (Figure 7). In the selected examples, the known synthetic value is contained within the 95% credible interval from the full Bayes model, but is not captured by the 95% credible interval from the empirical Bayes model; in other words, the true values clearly fall outside the 95% credible interval from the empirical solution much more than 5% of the time.

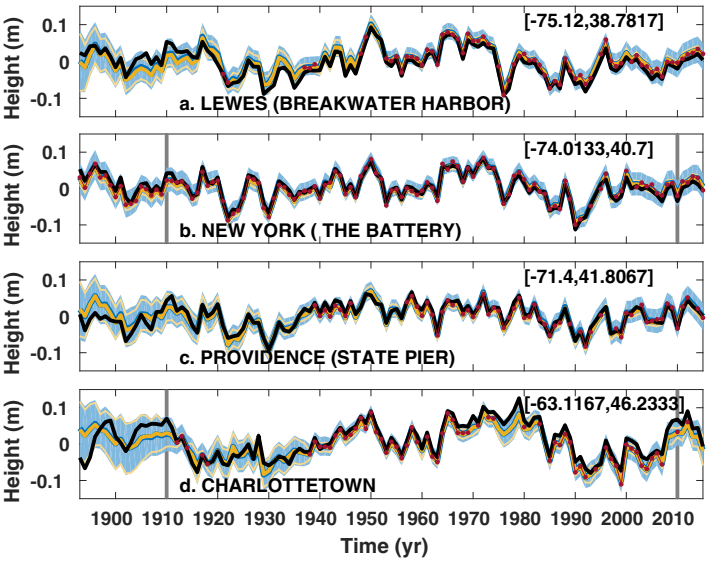

Figure 6. As in Figure 2 but for the synthetic data analysis. Black lines are the true values.

For a more systematic evaluation, we use tools from the weather-forecasting literature [*Hersbach*, 2000; *Hamill*, 2001; *Gel et al.*, 2004; *Li et al.*, 2010], namely the rank histogram, coverage rate, and continuous

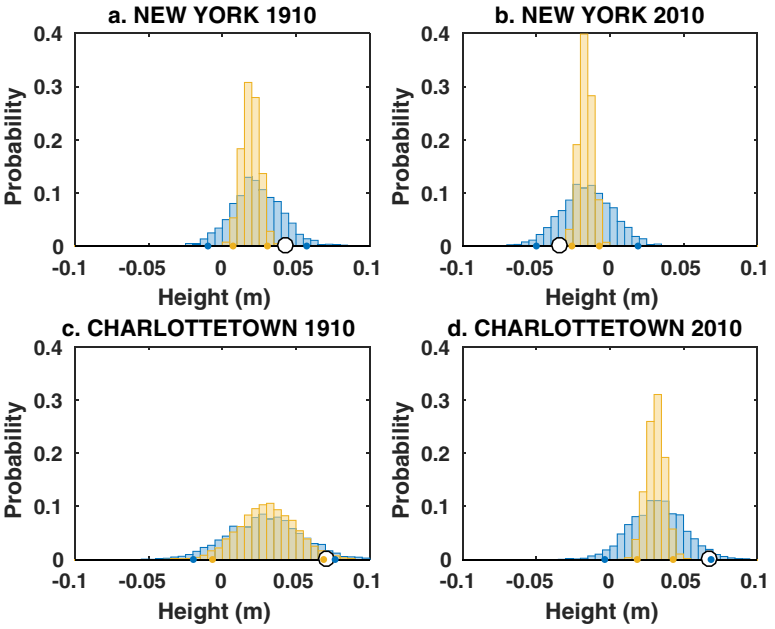

Figure 7. As in Figure 4 but for the synthetic data analysis. White dots are the true values.

ranked probability score. Given an ensemble of estimates of some field, along with the true field value (a "verifying analysis"), the rank histogram reveals whether the true value can plausibly be interpreted as a member of the estimated ensemble. It is evaluated by pooling the true and estimated values at a point, sorting the pooled values, ranking the true value, and averaging over many instances. If the true and estimated values can be described by the same probability distribution, and have the same statistical properties, then the true value is equally likely to have any rank, and so a necessary condition for reliable estimates is a uniform rank histogram. If the posterior distribution is insufficiently variable (e.g., if extreme values are observed more than expected based on the estimated ensemble), the rank histogram is concave ("U-shaped"), whereas a more bell-shaped histogram results if there is excess variability. Rank histograms are shown in Figure 8a. For the full Bayes solution, the rank histogram is more or less flat, but does show hints of overpopulation in the middle ranks, suggesting that credible intervals from the full Bayes algorithm are reliable but perhaps conservative. For the empirical Bayes solution, the rank histogram is clearly concave, signifying a lack of variability in the posterior estimates compared to the true field values.

The coverage rate of the credible intervals is the fraction of true values that fall within a given quantile, and equals that quantile when coverage is perfect. If coverage is greater than the target quantile, then the estimated credible intervals are too wide, and vice versa if the coverage is less than the target quantile. Coverage rates are shown in Figure 8b. The credible intervals from the full Bayes model are broadly reliable, but

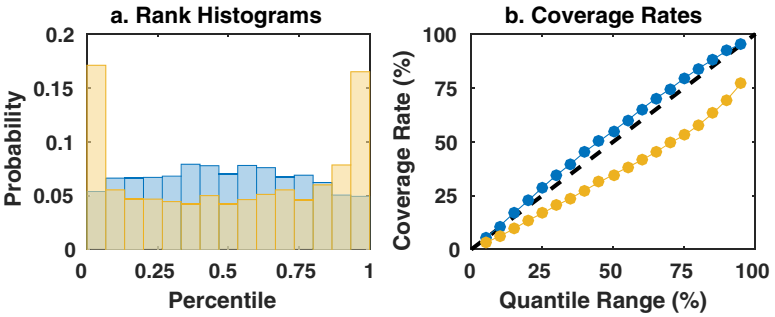

Figure 8. (a) Rank histograms and (b) coverage rates for the full (blue) and empirical (yellow) solutions. Black dashes in panel (b) mark the 1:1 line; if the coverage rates are perfect, the colored dots will fall along the black dashes.

with an excess of about 3% of true values captured by the posterior credible intervals. In contrast, the credible intervals from the empirical Bayes model are too narrow, with roughly 14% too few true values falling within the posterior credible intervals.

We also consider the continuous ranked probability score (CRPS), which measures both the accuracy and precision of an ensemble prediction [*Hersbach*, 2000]. The CRPS is a generalization of the Brier score and is based on the mean square difference between the cumulative probability distributions of the ensemble prediction and the true value. Like the Brier score, the CRPS is a proper scoring rule [*Gneiting and Raftery*, 2007]. The CRPS can be decomposed into a reliability component (Reli), which measures the average frequency with which the true value falls below some given percentile from the ensemble prediction, and a potential term ($CRPS_{pot}$) that relates to the spread of the ensemble. We assess CRPS, Reli, and $CRPS_{pot}$ values by comparing the model ensemble of posterior estimates to the true sea-level value for each individual tide-gauge location and year, and then averaging over all locations and years.

The full Bayes CRPS (8.33 mm) is smaller than the empirical Bayes CRPS (8.75 mm), showing that the former is a better estimator (as measured by the CRPS). The full Bayes solution is also more reliable than the empirical Bayes solution; Reli for the full solution (0.02 mm) is lower than for the empirical solution (0.43 mm), consistent with the rank histograms and coverage rates (Figure 8), suggesting that the full-Bayes credible intervals are more representative than the empirical-Bayes credible intervals. The full and empirical $CRPS_{pot}$ values are nearly identical (8.31 and 8.32 mm, respectively), and close to the full Bayes CRPS. The nearly identical full and empirical $CRPS_{pot}$ values could relate to the fact that the models have identical (and, by design, perfect) structure.

To interpret the differences in process credible intervals between the full and empirical Bayes algorithms, we run an additional suite of "intermediate" synthetic data experiments. Some model parameters are now given constant values (as in the empirical Bayes method), while others are assigned weak prior constraints (as in the full Bayes approach). In one subgroup of these experiments, most parameters are variable, except for one parameter (or a small subset of parameters) that is held fixed. In the other experimental subgroup, most parameters are set to constants, but one parameter (or small subset of parameters) is allowed to vary.

Some results from the intermediate experiments are given in Table 3. If the data-bias parameters $\{\ell, v, \tau^2\}$ are fixed, while all other parameters are variable, the coverage rates and widths of the 95% process credible intervals are close to the empirical Bayes case. For all other cases of fixing one parameter (or a small subset of parameters), and allowing the remaining parameters to vary, the coverage rates and widths of the credible intervals are similar to the full Bayes case. Likewise, if data-bias parameters are variable, credible interval widths and coverage rates are close to the full Bayes case. For all other instances of letting one parameter (or a small subset) vary, results are close to the empirical Bayes case. Thus, given our model design, in order

Table 3. Example Results From the Full, Empirical, and Intermediate Bayes Experiments[a]

Experiment	Parameter(s)	Coverage Rate	NY1910	NY2010	CT1910	CT2010
Full		95.6	46.2	43.6	86.7	48.3
Empirical		77.1	23.3	18.7	75.4	24.4
SGI (fixed)	$\{\boldsymbol{b}, \mu, \pi^2\}$	92.7	40.8	38.8	82.6	42.9
	$\{\ell, v, \tau^2\}$	78.0	22.9	19.1	77.0	24.5
	φ	95.7	44.0	41.7	86.0	48.1
	δ^2	95.5	41.4	39.8	85.1	41.2
	σ^2	95.1	41.8	39.3	85.9	43.8
	r	94.8	39.3	36.7	82.5	43.7
SGII (variable)	$\{\boldsymbol{b}, \mu, \pi^2\}$	78.1	22.7	18.8	77.7	23.3
	$\{\ell, v, \tau^2\}$	92.8	40.1	38.8	84.2	44.3
	φ	76.9	22.9	18.7	77.0	24.5
	δ^2	77.0	23.1	18.5	78.0	24.0
	σ^2	76.5	22.6	18.7	77.0	24.4
	r	76.6	23.4	18.6	75.0	24.6

[a]SGI refers to the intermediate model experiments where the parameters indicated in the second column are held fixed (all other parameters are allowed to vary). SGII denotes the intermediate experiments where the parameters in the second column are allowed to vary (all other parameters are held fixed). Coverage rates in the third column are the percentages of true process values captured by the 95% credible intervals, across all locations and times, for the various Bayesian model solutions. The fourth through seventh columns show the widths (in mm) of the 95% credible intervals for the process at New York (NY) and Charlottetown (CT) in 1910 and 2010.

to achieve reliable credible intervals on the sea-level process, uncertainty in the data-bias parameters needs to be taken into account.

These results demonstrate that the fully Bayesian solution is close to optimally reliable. They also suggest that the empirical approach could potentially perform as well as the fully Bayesian approach after some postprocessing, such as removing the time mean over some standard reference interval or datum epoch. Motivated by this suggestion, and the recommendation of a reviewer, we reconsider the full and empirical Bayes solutions after removing their respective time means over the 1983–2001 National Tidal Datum Epoch (NTDE) employed by the National Ocean Service. Parallel versions of Figures 2–8, based on solutions adjusted to NTDE, are shown in Figures S6–S12.

Removing temporal means over the NTDE impacts the widths of the credible intervals from both the full and empirical Bayes solutions. For the full Bayes, the widths of the 95% credible intervals are reduced considerably for recent periods (by more than 50% in some cases), but modestly increased for the earliest periods (at most by 5%; Figure 9a). Impacts are also apparent for empirical Bayes, but are less immediately striking; widths of 95% credible intervals decrease slightly for years within the NTDE and increase modestly for periods outside the NTDE (magnitudes of change at most 8%; Figure 9b). Discrepancies between widths of the full and empirical Bayes 95% credible intervals are thus considerably reduced for the most recent years (e.g., reduced by at least one half for years after 1918), but are comparatively less affected for the earliest times (Figure 9c).

The skill of the solutions is also influenced by removing mean values over the NTDE. For the empirical Bayes solution, leveling to the NTDE results in flatter rank histograms (Figure S12), a better CRPS (Table 4), and much more accurate coverage for recent years (Figure 9d). However, although performance is generally improved, removing time means from the solutions over the NTDE does not lead to universally better results. For example, both full and empirical-Bayes credible intervals become less reliable during the early years: before leveling to the NTDE, full and empirical Bayes 95% credible intervals capture 91 and 77% of true process values, respectively, across all tide gauges over the 19 year period 1901–1919; these respective coverages drop to 84 and 71% after referencing to the NTDE (Figure 9d).

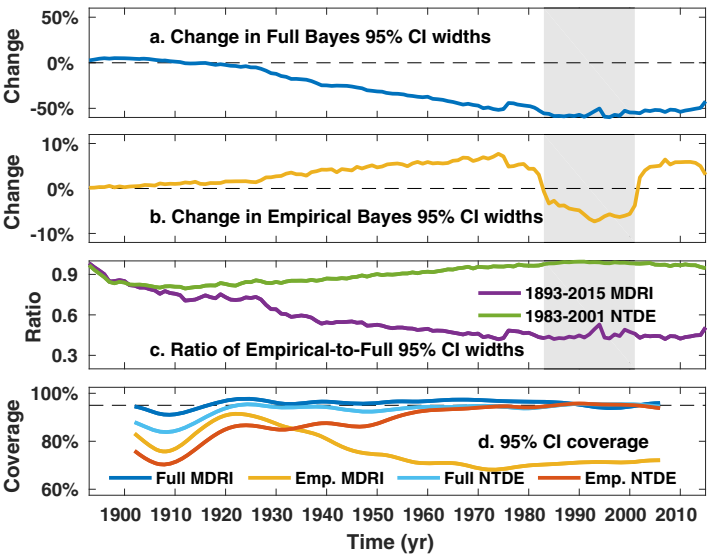

Figure 9. (a) Percent change in mean 95% process credible interval width over all 36 tide gauges for full Bayes going from the 1893–2015 model default reference interval (MDRI) to the 1983–2001 NTDE. (Negative values mean that NTDE gives narrower credible intervals than MDRI.) (b) As in Figure 9a but for empirical Bayes. (c) Ratio of average empirical-to-full Bayes 95% process credible interval width across gauges for each year referencing to MDRI or NTDE. Light gray shading in Figures 9a–9c is the NTDE. (d) Coverage of the 95% credible intervals, averaged over all sites and moving 19 year windows, of the full and empirical Bayes solutions referenced either to the 1893–2015 MDRI and 1983–2001 NTDE. Horizontal dashed dot gray line marks the target coverage.

5. Discussion

We assessed the relative merits of full and empirical Bayes approaches to infer sea-level changes from tide-gauge records. Specifically, we undertook a synthetic data analysis to evaluate the reliability of the credible intervals given by full and empirical Bayes methods. While the two approaches yielded very similar median estimates of the sea-level process, the empirical Bayes method generates credible intervals that are narrower than the full Bayes method (Figures 2–7). The differences between the full and empirical-Bayes credible intervals are more pronounced when data were available (Figures 4 and 7). Rank histograms, coverage rates, and continuous ranked probability scores in the context of the synthetic data

Table 4. CRPS, Reli, and $CRPS_{pot}$, Computed Over All Locations and Times, for the Full and Empirical Bayes Solutions Referenced to the Model Default Reference Interval 1893–2015 and the National Tidal Datum Epoch 1983–2001 (NTDE)[a]

Implementation	Reference Interval	CRPS	Reli	$CRPS_{pot}$
Full	1893–2015	8.32	0.02	8.31
Empirical	1893–2015	8.75	0.43	8.32
Full	1983–2001	7.75	0.04	7.71
Empirical	1983–2001	8.12	0.24	7.87

[a]All values have units of mm.

analysis show that the full Bayes approach is more reliable than the empirical Bayes method (Figure 8), highlighting the impact of parameter uncertainty, namely uncertainties in data-bias parameters (Table 3).

We also employed a synthetic data technique to assess the origins of differences between the full and empirical Bayes approaches. The full and empirical Bayes approaches provide nearly identical median process estimates, but with the full Bayes giving wider credible intervals. A simple analogy is illustrative for describing the origins of this result. Given an observation z drawn from a normal distribution $z|m, s^2 \sim \mathcal{N}(m, s^2)$, consider the following two cases. In the first case, the mean m is known, say $m \sim \delta(m-\mu)$, where $\delta(\cdot)$ is the Dirac delta, and so the marginal distribution is $z|s^2 \sim \mathcal{N}(\mu, s^2)$. In the second case, m is unknown, given the prior $m \sim \mathcal{N}(\mu, \tau^2)$, and so the marginal distribution is $z|s^2 \sim \mathcal{N}(\mu, s^2+\tau^2)$. The mean, as well as the median, of $p(z|s^2)$ is identical in both cases, but the variance is larger in the second case. The first case is likened to the empirical Bayes method, where model parameters are fixed constants, whereas the second is likened to the full Bayes approach, where the parameters are given prior distributions.

Using synthetic data analysis, we have shown that (given a perfect model) setting parameters equal to their most likely values is adequate for achieving a best estimate of the process, and the empirical Bayes method probably suffices in such a case. For inference on second or higher moments of the process, the empirical Bayes approach is insufficient, and the full Bayes method is preferred. An instance when it would be crucial to have good estimates of higher-order moments is characterizing long-lived extreme events on climate time scales, such as the striking increase in sea level along the northeast coast of North America between 2008 and 2010 reported on by *Goddard et al*. [2015]. However, we draw attention to the idealized nature of our synthetic data experiments. Given actual observations, some model misspecification is inevitable, and the actual most likely values will be uncertain. That the algorithm described here is *reasonable* for the case of tide-gauge sea-level data is supported by the residual analysis described in Supporting Information S1.3 and the references given in section 3. Nevertheless, our results give valuable guidance for future sea-level studies, demonstrating that full Bayes methods are safer and more transparent than empirical Bayes approaches in terms of their representation of uncertainty.

Our qualitative findings that empirical Bayes underestimates uncertainty and that full Bayes provides more reliable credible intervals are unsurprising and consistent with general expectations [e.g., *Carlin and Louis*, 2000; *Kang et al*., 2009; *Cressie and Wikle*, 2011]. It is nonetheless useful to quantify the degree to which empirical Bayes methods underestimate uncertainty on the sea-level field determined from tide-gauge records, and where exactly differences arise in the context of a fully Bayesian framework. In some cases, adjustments (e.g., bootstrapping) applied to empirical Bayes solutions have been shown to make them more reliable and account for variability in model parameters [e.g., *Laird and Louis*, 1987; *Stoffer and Wall*, 1991]. We found that the overall reliability of the empirical Bayes approach can be improved and discrepancies with respect to the full Bayes method can be reduced by subtracting mean values from the solutions during some short, recent common reference interval, for example, the 1983–2001 National Tidal Datum Epoch (Figure 9). However, such adjustments can degrade the skill of the model solutions in some instances (e.g., for earlier times), and differences between the full and empirical Bayes results remain, such that full Bayes still gives better characterization of uncertainties than empirical Bayes (e.g., Figures 9, S11, and S12 and Table 4). General statistical considerations also favor referencing to the longest possible interval; for example, choosing a shorter reference period introduces spurious *temporal* structure into the central *spatial* moments (e.g., spatial variance) of a spatiotemporal process [*Tingley*, 2012].

An additional consideration when choosing between full and empirical Bayes approaches is computational demand. In the present context the full Bayesian approach is preferred to the empirical Bayes method because there is little difference in computational cost. To run the 200,000 iterations used to generate each

1000 draw posterior solution sequence (cf. section 3.4), the full Bayes implementation takes ∼6 h, the empirical Bayes ∼4 h, on a standard desktop computer.

The hierarchical model described here is the simplest algorithm that we are comfortable applying to real sea-level data. It is possible to expand this framework for more general application. For example, a more complex spatial model may be useful, and one can imagine including any number of covariance functions associated with patterns of dynamic height, thermal expansion, and meltwater sources, with implications for changes in the gravitational field [e.g., as in *Hay et al*., 2015]. (As such development would require efforts beyond our current scope, we have chosen a simpler route for the present purposes of efficiently demonstrating the differences between full and empirical Bayes approaches.) Global Positioning System (GPS) data near tide gauges [*Wöppelmann et al*., 2009; *Santamaría-Gómez et al*., 2012] could also be incorporated, along with tide-gauge records, to distinguish sea-level changes due to vertical land motion from those due to variable ocean dynamics. Such a framework could be used, for example, to determine whether geocentric sea-level trends on the United States east coast are consistent with a weakening of the overturning circulation over the twentieth century [*Yin and Goddard*, 2013; *McCarthy et al*., 2015; *Rahmstorf et al*., 2015]. Such explorations are left to future study.

Acknowledgments

The first author was supported by NSF award 1558966 and NASA grant NNX14AJ515. The comments of two anonymous reviewers are gratefully acknowledged. Data were retrieved from the PSMSL database (http://www.psmsl.org/). The model code and solutions are available from the corresponding author upon request.

References

Andres, M., G. G. Gawarkiewicz, and J. M. Toole (2013), Interannual sea level variability in the western North Atlantic: Regional forcing and remote response, *Geophys. Res. Lett.*, *40*, 5915–5919, doi:10.1002/2013GL058013.

Ashby, D. (2006), Bayesian statistics in medicine: A 25 year review, *Stat. Med.*, *25*, 3589–3631, doi:10.1002/sim.2672.

Banerjee, S., B. P. Carlin, and A. E. Gelfand (2004), *Hierarchical Modeling and Analysis for Spatial Data*, 448 pp., Chapman and Hall, Boca Raton, Fla.

Berliner, L. M., C. K. Wikle, and N. Cressie (2000), Long-lead prediction of Pacific SSTs via Bayesian dynamic modeling, *J. Clim.*, *13*, 3953–3968.

Bernardinelli, L., and C. Montomoli (1992), Empirical Bayes versus fully Bayesian analysis of geographical variation in disease risk, *Stat. Med.*, *11*, 983–1007.

Bogdanov, V. I., et al. (2000), Mean monthly series of sea level observations (1777–1993) at the Kronstadt gauge, report, 34 pp., Finnish Geod. Inst., Kirkkonummi, Finland.

Bos, M. S., S. D. P. Williams, I. B. Araújo, and L. Bastos (2014), The effect of temporal correlated noise on the sea level rate and acceleration uncertainty, *Geophys. J. Int.*, *196*, 1423–1430, doi:10.1092/gji/ggt481.

Carlin, B. P., and T. A. Louis (2000), *Bayes and Empirical Bayes Methods for Data Analysis*, 2d ed., 440 pp., Chapman and Hall, Boca Raton, Fla.

Carriquiry, A., and M. D. Pawlovich (2005), From empirical Bayes to full Bayes: Methods for analyzing traffic safety data, October 25, 2004, White Paper, Iowa Dep. of Transp.

Church, J. A., N. J. White, R. Coleman, K. Lambeck, and J. X. Mitrovica (2004), Estimates of the regional distribution of sea level rise over the 1950–2000 period, *J. Clim.*, *17*, 2609–2625.

Cressie, N., and C. K. Wikle (2011), *Statistics for Spatio-Temporal Data*, 588 pp., John Wiley, Hoboken, N. J.

Davis, J. L., and J. X. Mitrovica (1996), Glacial isostatic adjustment and the anomalous tide gauge record of eastern North America, *Nature*, *379*, 331–333.

Ekman, M. (1988), The world's longest continuous series of sea level observations, *Pure Appl. Geophys.*, *127*, 73–77.

Gel, Y., A. E. Raftery, T. Gneiting, C. Tebaldi, D. Nychka, W. Briggs, M. S. Roulston, and V. J. Berrocal (2004), Calibrated probabilistic mesoscale weather field forecasting: The geostatistical output perturbation method, *J. Am. Stat. Assoc.*, *99*(467), 575–590.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2004), *Bayesian Data Analysis*, 2nd ed., 668 pp., Chapman and Hall, Boca Raton, Fla.

Gneiting, T., and A. E. Raftery (2007), Strictly proper scoring rules, prediction, and estimation, *J. Am. Stat. Assoc.*, *102*, 477, 359–378.

Goddard, P. B., J. Yin, S. M. Griffies, and S. Zhang (2015), An extreme event of sea-level rise along the Northeast coast of North America in 2009–2010, *Nat. Commun.*, *6*, 6346, doi:10.1038/ncomms7346.

Hamill, T. M. (2001), Interpretation of rank histograms for verifying ensemble forecasts, *Mon. Weather Rev.*, *129*, 550–560.

Harmon, R., and P. Challenor (1997), A Markov chain Monte Carlo method for estimation and assimilation into models, *Ecol. Model.*, *101*, 41–59.

Hay, C. C., E. Morrow, R. E. Kopp, and J. X. Mitrovica (2013), Estimating the sources of global sea level rise with data assimilation techniques, *Proc. Nat. Acad. Sci. U. S. A.*, *110*, 3692–3699, doi:10.1073/pnas.1117683109.

Hay, C. C., E. Morrow, R. E. Kop, and J. X. Mitrovica (2015), Probabilistic reanalysis of twentieth-century sea-level rise, *Nature*, *517*, 481–484, doi:10.1038/nature14093.

Heisterkamp, S. H., G. Doornbos, and M. Gankema (1993), Disease mapping using empirical Bayes and Bayes methods on mortality statistics in the Netherlands, *Stat. Med.*, *12*, 1895–1913.

Herbei, R., I. W. McKeague, and K. G. Speer (2008), Gyres and jets: Inversion of tracer data for ocean circulation structure, *J. Phys. Oceanogr.*, *38*, 1180–1202.

Hersbach, H. (2000), Decomposition of the continuous ranked probability score for ensemble prediction systems, *Weather Forecast.*, *15*, 559–570.

Higginson, S., K. R. Thompson, P. L. Woodworth, and C. W. Hughes (2015), The tilt of mean sea level along the east coast of North America, *Geophys. Res. Lett.*, *42*, 1471–1479, doi:10.1002/2015GL063186.

Holgate, S. J., A. Matthews, P. L. Woodworth, L. J. Rickards, M. E. Tamisiea, E. Bradshaw, P. R. Foden, K. M. Gordon, S. Jevrejeva, and J. Pugh (2013), New data systems and products at the permanent service for mean sea level, *J. Coastal Res.*, *29*(3), 493–504.

Huang, H., H. C. Chin, and M. M. Haque (2009), Empirical evaluation of alternative approaches in identifying crash hot spots, *Transp. Res. Record*, *2103*, 32–41, doi:10.3141/2103-05.

Hughes, C. W., and S. D. P. Williams (2010), The color of sea level: Importance of spatial variations in spectral shape for assessing the significance of trends, *J. Geophys. Res.*, *115*, C10048, doi:10.1029/2010JC006102.
Intergovernmental Oceanographic Commission (IOC) of UNESCO (2006), Manual on sea-level measurements and interpretation: An update to 2006, in *IOC Manuals and Guides 14*, JCOMM Tech. Rep. 31, vol. IV, 78 pp., Paris.
Kang, E. L., D. Liu, and N. Cressie (2009), Statistical analysis of small-area data based on independence, spatial, non-hierarchical, and hierarchical models, *Comput. Stat. Data Anal.*, *53*, 3016–3032.
Karegar, M. A., T. H. Dixon, and S. E. Engelhart (2016), Subsidence along the Atlantic Coast of North America: Insights from GPS and late Holocene relative sea-level data, *Geophys. Res. Lett.*, *43*, 3126–3133, doi:10.1002/2016GL068015.
Kennedy, M. C., and A. O'Hagan (2001), Bayesian calibration of computer models, *J. R. Statist. Soc. B*, *63*(3), 425–464.
Kopp, R. E. (2013), Does the mid-Atlantic United States sea level acceleration hot spot reflect ocean dynamic variability?, *Geophys. Res. Lett.*, *40*, 3981–3985, doi:10.1002/grl.50781.
Kopp, R. E., R. M. Horton, C. M. Little, J. X. Mitrovica, M. Oppenheimer, D. J. Rasmussen, B. H. Strauss, and C. Tebaldi (2014), Probabilistic 21st and 22nd century sea-level projections at a global network of tide-gauge sites, *Earth's Future*, *2*, 383–406, doi:10.1002/2014EF000239.
Kopp, R. E., C. C. Hay, C. M. Little, and J. X. Mitrovica (2016), Geographic variability of sea-level change, *Curr. Clim. Change Rep.*, *1*, 192–204, doi:10.1007/s40641-015-0015-5.
Kopp, R. E., et al. (2016), Temperature-driven global sea-level variability in the common Era, *Proc. Natl. Acad. Sci. U. S. A.*, *113*, E1434–E1441, doi:10.1073/pnas.1517056113.
Laird, N. M., and T. A. Louis (1987), Empirical Bayes confidence intervals based on bootstrap samples, *J. Am. Stat. Assoc.*, *82*, 399, 739–750.
Lan, B., B. Persaud, C. Lyon, and R. Bhim (2009), Validation of a full Bayes methodology for observational before-after road safety studies and application to evaluation of rural signal conversions, *Accid. Anal. Prev.*, *41*, 574–580.
Leyland, A. H., and C. A. Davies (2005), Empirical Bayes methods for disease mapping, *Stat. Methods Med. Res.*, *14*, 17–34.
Li, B., D. W. Nychka, and C. M. Ammann (2010), The value of multiproxy reconstruction of past climate, *J. Am. Stat. Assoc.*, *105*(491), 883–895.
Louis, T. A., and W. Shen (1999), Innovations in Bayes and empirical Bayes methods: Estimating parameters, populations and ranks, *Stat. Med.*, *18*, 2493–2505.
McCarthy, G. D., I. D. Haigh, J. J.-M. Hirschi, J. P. Grist, and D. A. Smeed (2015), Ocean impact on decadal Atlantic climate variability revealed by sea-level observations, *Nature*, *521*, 508–510, doi:10.1038/nature14491.
McKeague, I. W., G. Nicholls, K. Speer, and R. Herbei (2005), Statistical inversion of South Atlantic circulation in an abyssal neutral density layer, *J. Mar. Res.*, *63*, 683–704.
Miaou, S.-P., and D. Lord (2003), Modeling traffic crash-flow relationships for intersections, *Transp. Res. Record*, *1840*, 31–40.
Miller, K. G., R. E. Kopp, B. P. Horton, J. V. Browning, and A. C. Kemp (2013), A geological perspective on sea-level rise and its impacts along the U.S. mid-Atlantic coast, *Earth's Future*, *1*, 3–18, doi:10.1002/2013EF000135.
Miller, L., and B. C. Douglas (2006), On the rate and causes of twentieth century sea-level rise, *Philos. Trans. R. Soc. A*, *364*, 805–820, doi:10.1098/rsta.2006.1738.
Milliff, R., A. Bonazzi, C. Wikle, N. Pinardi, and L. Berliner (2011), Ocean ensemble forecasting. Part I: Ensemble Mediterranean winds from a Bayesian hierarchical model, *Q. J. R. Meteorol. Soc.*, *137*, 858–878.
Permanent Service for Mean Sea Level (PSMSL) (2016), Tide Gauge Data, Liverpool, U. K. [Available at http://www.psmsl.org/data/obtaining/.]
Persaud, B., B. Lan, C. Lyon, and R. Bhim (2010), Comparison of empirical Bayes and full Bayes approaches for before-after road safety evaluations, *Accid. Anal. Prev.*, *42*, 38–43.
Piecuch, C. G., and R. M. Ponte (2015), Inverted barometer contributions to recent sea-level changes along the northeast coast of North America, *Geophys. Res. Lett.*, *42*, 5918–5925, doi:10.1002/2015GL064580.
Piecuch, C. G., S. Dangendorf, R. M. Ponte, and M. Marcos (2016), Annual sea level changes on the North American Northeast Coast: Influence of local winds and barotropic motions, *J. Clim.*, *29*, 4801–4816.
Pinardi, N., A. Bonazzi, S. Dobricic, R. Milliff, C. Wikle, and L. Berliner (2011), Ocean ensemble forecasting. Part II: Mediterranean forecast system response, *Q. J. R. Meteorol. Soc.*, *137*, 879–893.
Ponte, R. M., C. Wunsch, D. Stammer (2007), Spatial mapping of time-variable errors in *Jason-1* and TOPEX/Poseidon sea surface height measurements, *J. Atmos. Oceanic Technol.*, *24*, 1078–1085.
Pugh, D. T., and P. L. Woodworth (2014), *Sea-Level Science: Understanding Tides, Surges Tsunamis and Mean Sea-Level Changes*, 407 pp., Cambridge Univ. Press, Cambridge, U. K.
Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski (2005), Using Bayesian model averaging to calibrate forecast ensembles, *Mon. Weather Rev.*, *133*, 1155–1174.
Rahmstorf, S., J. E. Box, G. Feulner, M. E. Mann, A. Robinson, S. Rutherford, and E. J. Schaffernicht (2015), Exceptional twentieth-century slowdown in the Atlantic Ocean overturning circulation, *Nat. Clim. Change*, *5*, 475–480, doi:10.1038/nclimate2554.
Santamaría-Gómez, A., M. Gravelle, X. Collilieux, M. Guichard, B. Martín Míguez, P. Tiphaneau, and G. Wöppelmann (2012), Mitigating the effects of vertical land motion in tide-gauge records using a state-of-the-art GPS velocity field, *Global Planet. Change*, *98–99*, 6–17.
Stoffer, D. S., and K. D. Wall (1991), Bootstrapping state-space models: Gaussian maximum likelihood estimation and the Kalman filter, *J. Am. Stat. Assoc.*, *86*(416), 1024–1033.
Thompson, P. R., and G. T. Mitchum (2014), Coherent sea level variability on the North Atlantic western boundary, *J. Geophys. Res. Oceans*, *119*, 5676–5689, doi:10.1002/2014JC009999.
Tingley, M. P. (2012), A Bayesian ANOVA scheme for calculating climate anomalies, with applications to the instrumental temperature record, *J. Clim.*, *25*, 777–791.
Tingley, M. P., and P. Huybers (2010), A Bayesian algorithm for reconstructing climate anomalies in space and time. Part I: Development and applications to paleoclimate reconstruction problems, *J. Clim.*, *23*, 2759–2781.
Tingley, M. P., P. F. Craigmile, M. Haran, B. Li, E. Mannshardt, and B. Rajaratnam (2012), Piecing together the past: Statistical insights into paleoclimatic reconstructions, *Quat. Sci. Rev.*, *35*, 1–22.
Wikle, C. K., and L. M. Berliner (2007), A Bayesian tutorial for data assimilation, *Physica D*, *230*, 1–16.
Wikle, C. K., R. F. Milliff, D. Nychka, and L. M. Berliner (2001), Spatiotemporal hierarchical Bayesian modeling: Tropical ocean surface winds, *J. Am. Stat. Assoc.*, *96*(454), 382–397.
Wikle, C. K., R. F. Milliff, R. Herbei, and W. B. Leeds (2013), Modern statistical methods in oceanography: A hierarchical perspective, *Stat. Sci.*, *28*(4), 466–486, doi:10.1214/13-STS436.
Woodworth, P. L. (1999), High waters at Liverpool since 1768: The UK's longest sea level record, *Geophys. Res. Lett.*, *26*(11), 1589–1592.

Woodworth, P. L., N. Pouvreau, and G. Wöppelmann (2010), The gyre-scale circulation of the North Atlantic and sea level at Brest, *Ocean Sci.*, *6*, 185–190.

Woodworth, P. L., M. Á. Morales Maqueda, V. M. Roussenov, R. G. Williams, and C. W. Hughes (2014), Mean sea-level variability along the northeast American Atlantic coast and the roles of the wind and the overturning circulation, *J. Geophys. Res. Oceans*, *119*, 8916–8935, doi:10.1002/2014JC010520.

Wöppelmann, G., et al. (2009), Rates of sea-level change over the past century in a geocentric reference frame, *Geophys. Res. Lett.*, *36*, L12607, doi:10.1029/2009GL038720.

Yin, J., and P. B. Goddard (2013), Oceanic control of sea level rise patterns along the East Coast of the United States, *Geophys. Res. Lett.*, *40*, 5514–5520, doi:10.1002/2013GL057992.