RMetS
Royal Meteorological Society

Bayesian modelling and ensemble reconstruction of mid-scale spatial variability in North Atlantic sea-surface temperatures for 1850–2008

Alicia R. Karspeck,[a]* Alexey Kaplan[b] and Stephan R. Sain[a]
[a]*National Center for Atmospheric Research, Boulder, CO, USA*
[b]*Lamont Doherty Earth Observatory, Columbia University, New York, NY, USA*

*Correspondence to: A. R. Karspeck, NCAR, Box 3000, Boulder, CO 80307, USA. E-mail: aliciak@ucar.edu

Existing historical records of sea-surface temperature extending back to the mid-1800s are a valuable source of information about climate variability on interannual and decadal time-scales. However, the temporal and spatial irregularity of these data make them difficult to use in climate research, where gridded and complete data fields are expected for both statistical analysis and forcing numerical models.

Infilling methods based on constraining the solution to the linear space spanned by the leading eigenvectors of the global-scale covariance, otherwise known as reduced-space methods, have proven very successful in creating gridded estimates of sea-surface temperature. These methods are especially useful for infilling the vast regions of unobserved ocean typical of the earliest segments of the data record. Regional variability, on the other hand, is not well represented by these methods, especially in data-poor regions. Here we present a method for augmenting the established large-scale reconstruction methods with a statistical model of the mid-scale variability. Using high quality sea-surface temperature data from the last 30 years including satellite-derived records, we specify a spatially non-stationary, anisotropic covariance model for the mid-scale sea-surface temperature variability. With the parameters of the covariance model estimated from the modern record, historical observations are used for conditioning the posterior distribution. Specifically, we form the expected value and correlated uncertainty of the mid-scales as well as generating samples from the posterior.

While this work focuses on a limited domain in the midlatitude North Atlantic Ocean, the method employed here can be extended to global reconstructions.

Key Words: data assimilation; uncertainty quantification; historical SST reconstruction; reduced space; optimal interpolation; multi-scale reconstruction; parametric methods; non-stationary covariance

Received 16 February 2010; Revised 28 April 2011; Accepted 12 July 2011; Published online in Wiley Online Library 25 August 2011

Citation: Karspeck AR, Kaplan A, Sain SR. 2011. Bayesian modelling and ensemble reconstruction of mid-scale spatial variability in North Atlantic sea-surface temperatures for 1850–2008. *Q. J. R. Meteorol. Soc.* **138**: 234–248. DOI:10.1002/qj.900

1. Introduction

Prior to the current era of satellite data acquisition, the main source of information on sea-surface temperatures (SST) came from the logs of ships of opportunity. These records stretch back to the mid 19th century, making them a tantalizing source of information about climate variability on interannual and decadal time-scales. However, the temporal and spatial inhomogeneity of these data make them difficult to use in standard statistical analysis procedures.

Gridded fields of SST are also needed for initialization and verification of ocean models and as time-dependent boundary conditions in atmospheric models. As a result, interpolation schemes for infilling these sparse data are tremendously important in climate research.

One popular approach to the interpolation of historical datasets is reduced-space estimation (Shriver and O'Brien, 1995; Smith *et al.*, 1996; Kaplan *et al.*, 1998, 2000; Rayner *et al.*, 2003). One of its advantages over more conventional methods such as simple kriging (which uses a stationary, localized covariance function) is its emphasis on the reconstruction of the largest and most energetic spatial scales over the entire domain of interest. This is a natural advantage for modelling climate variables such as SST, because the dynamics of the climate system often result in global-scale coherency.

It is worth considering why reduced-space estimation has been so useful in climate applications. For climate variables that possess a large spatial dimension, the relative temporal 'shortness' of the reliable observational data record that can be used for computing a sample covariance matrix often leads to rank-deficiency. Assuming a lower dimensionality of the system via truncation of the less energetic eigenvectors of the covariance matrix can circumvent this problem. Another advantage of reduced-space techniques becomes evident when the data used for reconstruction are clustered in limited areas, leaving large regions completely unobserved. Under these circumstances, inference in the interiors of the unsampled regions would be imprudent using methods that rely solely on local spatial estimation methods.

The disadvantage of using a reduced-space technique for interpolation is that there is no guarantee that the patterns of covariability that dominate within smaller subregions of the global domain will be well represented. The truncation of trailing eigenvectors necessarily excludes some structures that are better suited to local estimation techniques. Ideally, a reconstruction methodology would draw from the strengths of both types of interpolation, with the aim of representing behaviour over a range of spatial scales.

We restrict our focus to the statistical modelling and reconstruction of SST anomalies in the northern hemisphere Atlantic Ocean. We present a method to augment an existing historical SST reconstruction that uses a reduced-space Kalman smoother (Kaplan *et al.*, 1998) to capture what we will term the 'global-scale' or 'large-scale' modes of variability. The contribution of this work is to model and reconstruct what we will term 'mid-scale' variability. For the remainder of this article, we will use the terms global and mid-scale to distinguish between variability captured by the reduced-space technique and the more locally dominant variability on which we are focused.

The separation into global and mid-scales is not based on physical processes. No objective criteria for parsing between these covariance models is used, nor do we mean to imply that a given length-scale of covariability will be uniquely contained in either model. In the context of this study, mid-scales can be interpreted as the most dominant local variability not captured by the globally-based reduced space reconstruction.

Section 2 describes the historical temperature data, extending back to 1850, that are used in this reconstruction. Section 3 gives a brief description of the reduced-space Kalman smoother that was used in the published reconstruction of the large-scale SST anomalies. As we discuss, there are subjective choices that go into reduced-space techniques and our definition of mid-scales is implicitly impacted by these. Given this caveat, it is still instructive to note that the mid-scales tend to have geographic coherency of the order of 500–1300 km.

There are two main areas of emphasis in this work. They are (1) the statistical modelling of our prior knowledge of the mid-scale variability not present in the established reduced space reconstruction and (2) description of the mid-scale reconstruction in terms of the mean, covariance and samples from the posterior distribution. Section 4 outlines the statistical procedure that we use to form the posterior distribution for our mid-scale reconstruction. In section 5 we present our model for the covariance of the mid-scale variability. We employ a novel covariance parametrization developed by Paciorek and Schervish (2006) that allows for non-stationarity in the length-scales and anisotropy of the spatial correlation functions. This parametrization gives our model the flexibility to capture geographic variation in the underlying covariability of SST anomalies while still ensuring a positive-definite covariance matrix defined over the entire domain. This is a useful feature for analyses of SST in the northern Atlantic Ocean basin, where the dominant physical processes vary over the domain.

We verify the statistical model in section 6 and section 7 presents a selection of the resultant reconstructions. Because the quantification and representation of uncertainty has become an area of increased interest within the climate research community (Rayner *et al.*, 2009), we pay special attention to the uncertainty estimates implied by the posterior distribution. Specifically, we note the temporal evolution of the uncertainty due to changes in data availability through time and the spatial correlations inherent in the posterior distributions. We conclude in section 8 with a discussion of some of the broader issues relevant to this work, some of its limitations and prospects for its extension.

2. *In situ* SST observations from 1850–2008

Our reconstruction is based on the Hadley Centre Sea-Surface Temperature, version 2(HadSST2) dataset of monthly *in situ* SST anomalies from 1850–2008. HadSST2 is based on the International Comprehensive Ocean–Atmosphere Data Set (ICOADS) archive of surface marine observations collected from ships and buoys (Worley *et al.*, 2005). In HadSST2, the ICOADS SST data are subjected to quality checks, corrected for systematic bias, converted to climatological anomalies and averaged on to a $1^\circ \times 1^\circ$ grid.* Note that HadSST2 is not an interpolated product, so grid boxes where no data are present in the ICOADS database remain empty. Rayner *et al.*(2006) document the extensive work that was done to create the HadSST2 dataset, including a detailed description of the bias-correction methods.

Because data in the early part of the record were primarily collected on a volunteer basis by merchant ships, they tend to be concentrated along trade routes, leaving large portions of the ocean unobserved. While the number of records generally increases in time, sociopolitical events such as the two World Wars and the Great Depression are marked by temporary decreases in the availability of data. However,

*These data can be downloaded via the UK Met Office at http://hadobs.metoffice.com/hadsst2/data/download.html

the *in situ* coverage becomes dense in the second half of the twentieth century and by the mid 1960s observations are routinely available over nearly 90% of the North Atlantic Ocean basin. The reader is referred to Worley *et al.*(2005) and Rayner *et al.*(2006) for details on the time-evolving *in situ* data coverage.

In general, uncertainty in the monthly gridbox-average SST is dominated by instrumental error and error due to undersampling of the subgrid-scale variability (Rayner *et al.*, 2006).† In Rayner *et al.*(2009), temporally and spatially dense measurements from the Pathfinder satellite SST dataset (version 5) as well as independent estimates of instrumental error and small-scale variability (Kent and Challenor, 2006) are used to calculate the variance of this uncertainty, s^2 (see fig. 2d of Rayner *et al.*, 2009). At every month j from 1850–2008, a vector of observational errors (defined for each $1^\circ \times 1^\circ$ grid box i where observations exist) is assumed to have a multivariate normal distribution:

$$\epsilon \sim N(0, \mathcal{R}). \tag{1}$$

where $\mathcal{R}$ is a diagonal matrix with elements $s_i^2 / \min(n_{i,j}, 60)$. The temporal and spatial non-stationarity in the data error stems from both the geographical variations in s_i^2 and the irregularity of the number of samples $n_{i,j}$ collated within each grid box. The number of degrees of freedom is capped at 60 as a heuristic way to account for our understanding that not all observations in a single grid box will be statistically independent (Kaplan *et al.*, 1997).

3. Large-scale reconstruction using the reduced-space Kalman smoother

Central to all reduced-space estimation techniques is the idea that the global covariability can be adequately expressed in terms of a limited number of spatial basis functions. If these basis functions are found as the leading eigenvectors of a sample covariance matrix, these basis functions are called empirical orthogonal functions (EOFs). Typically, the number of EOFs retained in a reconstruction is far fewer than the spatial dimension of the state variable. The exact number of basis functions retained is a subjective compromise between the desire to capture a large portion of the variability in the system and the desire to drastically reduce the size of the problem. At the very least, the truncation must be severe enough to eliminate any rank-deficiency in the full-space sample covariance matrix.

In this framework, the reduced-space expression of a state vector of global SST anomalies can be written as

$$\hat{z} = \mathcal{E}\alpha. \tag{2}$$

Here $\mathcal{E}$ is an orthogonal matrix, the columns of which are the EOFs, and α is a vector of weighting coefficients. In reduced-space estimation it is generally assumed that the EOFs are fixed and α is the probabilistic variable of interest. Well-known methods of data assimilation (Kalman filtering, smoothing, variational methods, etc.) can be used to construct a posterior distribution of α.

The historical SST reconstruction that serves as the large-scale base for our mid-scale features is the reduced-space Kalman smoother described in Kaplan *et al.* (1997, 1998). Hereafter we will refer to this reconstruction as the KaplanSST. Since the details of the KaplanSST are described elsewhere in the literature and are not the focus of this work, we provide only a brief description below.

The KaplanSST is a near-global SST reconstruction wherein the EOFs are computed from the global sample covariance and the reconstructions make use of observations from all ocean basins. The version of this reconstruction used here is based on the $1^\circ \times 1^\circ$ HadSST2 dataset of *in situ* SST observations described in the previous section. The analysis is done for SST anomalies, i.e. deviations of full SST values from their monthly climatological values. The climatology is that of Smith and Reynolds (1998) for the 1961–1990 period. SST anomalies from the relatively data-rich period 1951–2007 were used for calculating the sample covariance matrix from which the orthogonal basis functions were computed and the data from 1850–2008 were used as the observations on which posterior distributions are conditioned. Consistent with Kaplan *et al.* (1998), only 80 global EOFs are retained. The $1^\circ \times 1^\circ$ HadSST2 is area averaged onto a $5^\circ \times 5^\circ$ grid prior to computing EOFs and generating the reconstruction. The resulting reconstruction is then bilinearly interpolated on to a $1^\circ \times 1^\circ$ grid. Unlike Kaplan *et al.* (1997, 1998), where only the expected value of the reduced-space analysis was computed, here the posterior covariance of the reduced space is also estimated as described in Kaplan *et al.* (2000). From the KaplanSST, then, we have a sequence of global SST anomalies (along with their corresponding error covariance) at each month from 1850–2008. By construction, the KaplanSST estimates only large-scale, globally relevant modes of variability.

As pointed out by Dommenget (2007), describing the covariability of a system in terms of a set of EOFs is purely a statistical convenience. It does not imply that the underlying covariance could not also be described using another statistical or physical model. Reduced-space techniques are useful not because they are unique models of the covariance but because they are a parsimonious way of describing (and ranking in terms of importance) the types of variability that are typically attractive to the climate science community. However, there can be coherent scales not captured by the reduced-space reconstruction that have regional importance. These scales can be modelled using localized covariance models, as described in the following sections.

4. Framework for the reconstruction of the mid-scales

Because the SST anomalies are the sum of the variability on global scales and mid-scales, the most complete reconstruction procedure considers the joint distribution of these processes. At each month, we can define a joint state as the large-scale SST anomalies in reduced space (α) appended with mid-scale SST anomalies (z') in full grid space:

$$Z = \begin{bmatrix} \alpha \\ z' \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \lambda & 0 \\ 0 & C \end{bmatrix} \right). \tag{3}$$

The matrices λ and C are the specified prior covariances on large and mid-scales (respectively). The observations can be written as

$$y = \mathcal{H}(\mathcal{E}\alpha + z') + \epsilon, \tag{4}$$

†Although we do not include them here, there are also uncertainties that stem from errors in the bias-correction algorithms.

where y is a vector of the HadSST2 observations at a given month and ϵ is the corresponding vector of measurement error (both described in section 2). The columns of the matrix $\mathcal{E}$ correspond to the basis functions of the KaplanSST that have been bilinearly interpolated on to a $1^\circ \times 1^\circ$ grid. $\mathcal{H}$ is a submatrix of the identity matrix that maps from the $1^\circ \times 1^\circ$ geographical state space of z to the observational space of y.

Although the large and mid-scales are assumed independent in the prior (3), once Z is conditioned on y they are no longer independent. We do not show this here, but it can easily be seen by forming the joint posterior $p(Z|y)$. This links the problem of reconstructing the mid-scales with the problem of large-scale reconstruction. This combination increases the effective dimension of the problem from $\mathcal{O}(10^2)$, which represents the number of EOFs retained in the KaplanSST analysis, to $\mathcal{O}(10^4)$, the number of grid points in the geographical state space. The assumed time autocorrelation in α further exacerbates this issue because smoother solutions are naturally more computationally demanding than sequential filters.

An alternative to forming the joint distribution is to recognize that the variable of interest is actually the sum of the large and mid-scales ($z = \mathcal{E}\alpha + z'$), which can be written

$$p(z|y) = \int p(z|\alpha, y)\; p(\alpha|y)\, \mathrm{d}\alpha. \tag{5}$$

The marginal posterior probability $p(\alpha|y)$ can be approximated by the reduced-space KaplanSST solution, which is multivariate normal (in time and space) with mean μ_α and covariance P_α. In forming the marginal distribution of α, the KaplanSST assumed that all variability unresolved by the leading EOFs was uncorrelated. We show in the next section that in fact there are significant correlation structures, but because the structures in z' are considerably smaller in scale than the dominant global patterns in $\hat{z}$ this is a tolerable approximation.

Using the KaplanSST distribution as a close approximation for the marginal posterior, the solution becomes

$$p(z|y) \approx \int p(z|\alpha, y)\; N(\alpha|\mu_\alpha, P_\alpha)\, \mathrm{d}\alpha. \tag{6}$$

This hierarchical form allows us to focus on forming only the first factor in the integral (with the understanding that samples can be drawn from the second). Applying Bayes' theorem, we can write

$$p(z|\alpha, y) \propto p(y|z, \alpha)\; p(z|\alpha), \tag{7}$$

where the first term on the right-hand side is a Gaussian likelihood and the second term is a Gaussian prior distribution with mean α and covariance of the mid-scales (C). We can then recognize $p(z|\alpha, y)$ as also being normally distributed, with expected value

$$\mu_{z|\alpha} = \mathcal{E}\alpha + C\mathcal{H}^{\mathrm{T}}(\mathcal{H}C\mathcal{H}^{\mathrm{T}} + \mathcal{R})^{-1}(y - \mathcal{H}\mathcal{E}\alpha) \tag{8}$$

and covariance

$$P_{z|\alpha} = C - C\mathcal{H}^{\mathrm{T}}(\mathcal{H}C\mathcal{H}^{\mathrm{T}} + \mathcal{R})^{-1}\mathcal{H}C. \tag{9}$$

The full solution given in (6) is then simply the integral over the product of two Gaussians:

$$\begin{aligned} p(z|y) &\approx \int p(z|\alpha, y)\; N(\alpha|\mu_\alpha, P_\alpha)\, \mathrm{d}\alpha \\ &= \int N(z|\mu_{z|\alpha}, P_{z|\alpha})\; N(\alpha|\mu_\alpha, P_\alpha)\, \mathrm{d}\alpha. \end{aligned} \tag{10}$$

Samples from this distribution can be formed following a simple Monte Carlo approach: given a sample from $N(\alpha|\mu_\alpha, P_\alpha)$, we can draw from $N(z|\mu_{z|\alpha}, P_{z|\alpha})$. This is a particularly attractive tactic, because only the expected value $\mu_{z|\alpha}$ depends on α and the covariance $P_{z|\alpha}$ need only be computed once regardless of the number of samples needed.

It is also possible to write the full solution by performing the integral over α. We show in Appendix A that there is a compact matrix form for this integral. However, while it is easy to write, it is not a matrix that one would like to form explicitly because of its full rank in time and state space. We draw on the result that the expected value of the full solution is

$$\mu_z = \mathcal{E}\mu_\alpha + C\mathcal{H}^{\mathrm{T}}(\mathcal{H}C\mathcal{H}^{\mathrm{T}} + \mathcal{R})^{-1}(y - \mathcal{H}\mathcal{E}\mu_\alpha). \tag{11}$$

Defining a data residual term,

$$\delta y \equiv y - \mathcal{H}\mathcal{E}\mu_\alpha, \tag{12}$$

we can write the mid-scale portion of the expected value as

$$\mu_{z'} = \mu_z - \mathcal{E}\mu_\alpha = C\mathcal{H}^{\mathrm{T}}(\mathcal{H}C\mathcal{H}^{\mathrm{T}} + \mathcal{R})^{-1}\delta y. \tag{13}$$

To demonstrate the type of correlated uncertainty associated with the mid-scales, we focus in the remainder of this article on only the conditional covariance $P_{z|\alpha}$ given by (9). It is worthwhile to keep in mind, however, that the full uncertainty, as shown in Appendix A, contains the KaplanSST posterior uncertainty (P_α) as well as terms involving the interaction between large and mid-scales.

In the following section we describe our model for the prior covariance matrix C. Once this mid-scale covariance matrix has been formed, it is straightforward (albeit computationally expensive) to generate samples from the posterior distribution defined by (9) and (13).

5. Specification of the prior covariance

5.1. A non-stationary model of the covariance

Although commonly used in spatial statistics, the Matérn class of covariance functions is less often encountered in geoscience literature. We will briefly describe its functional form and the extension of this form to a non-stationary covariance matrix.

The stationary, isotropic form of the Handcock–Stein–Wallis parametrization of the Matérn function (Handcock and Stein, 1993; Handcock and Wallis, 1994) is

$$c(d) = \sigma^2 \frac{1}{\Gamma(\nu)2^{\nu-1}} \left(2\sqrt{\nu}\frac{d}{\rho}\right)^{\nu} K_\nu \left(2\sqrt{\nu}\frac{d}{\rho}\right), \quad \rho > 0;\ \nu > 0, \tag{14}$$

where σ^2 is the variance of the process, d is a scalar distance, ρ is a spatial length-scale (also called the range parameter), Γ is the gamma function and $K_\nu(\cdot)$ is the modified Bessel function of the second kind. The shape parameter ν allows the covariance function to range from the exponential form $e^{-\sqrt{2}d/\rho}$ at $\nu = 0.5$ to the squared exponential (or Gaussian) form e^{-d^2/ρ^2} as ν approaches infinity.

In the context of optimal interpolation with geophysical data, the Matérn covariance function is attractive primarily because the parameter ν allows for flexibility in the spatial smoothness of the process. It is also the case that infinitely differentiable correlation functions, such as the Gaussian form, can result in ill-conditioned covariance matrices. Thus, from an engineering perspective, the use of a function that is less smooth than the Gaussian form can result in better numerical stability during the computation of the posterior moments.

We can generalize to the anisotropic form of the Matérn covariance by using the so-called Mahalanobis distance,

$$\tau(x, x') = \sqrt{(x - x')^{\mathrm{T}} \Sigma^{-1} (x - x')}, \tag{15}$$

in lieu of d/ρ in (14). Here x and x' are geographical locations in our spatial domain. The kernel Σ acts essentially as the square of a deformation and rotation matrix,

$$\Sigma = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} L_x^2 & 0 \\ 0 & L_y^2 \end{bmatrix} \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}^{\mathrm{T}}, \tag{16}$$

where L_x and L_y are (respectively) the zonal and meridional length-scales. The local angle of rotation relative to the zonal axis is given by θ. The kernal inverse within the Mahalanobis distance can be interpreted as the transformation of an absolute Euclidean distance to a distance measured *relative* to the rotated axis and anisotropic length-scales.

Paciorek and Schervish (2006) derive a non-stationary version of the anisotropic Matérn covariance that essentially knits together locally estimated kernels Σ to form a valid covariance matrix:

$$C(x, x') = \frac{\sigma\sigma'}{\Gamma(\nu) 2^{\nu-1}} \frac{|\Sigma|^{1/4} |\Sigma'|^{1/4}}{|\bar{\Sigma}|^{1/2}} \left(2\bar{\tau}\sqrt{\nu}\right)^{\nu} K_\nu \left(2\bar{\tau}\sqrt{\nu}\right). \tag{17}$$

Here, σ and σ' are the standard deviations at x and x' and $\bar{\tau}$ is the Mahalanobis distance based on the average of the kernels centred at x and x':

$$\bar{\tau} = \sqrt{(x - x')^{\mathrm{T}} \bar{\Sigma}^{-1} (x - x')}; \quad \bar{\Sigma} = \frac{1}{2}(\Sigma + \Sigma'). \tag{18}$$

In order to construct the full covariance matrix, we must estimate a global smoothness parameter ν and, at each geographical point, obtain the variance of the process (σ^2) and the transformation kernel Σ.

5.2. *Using modern data to obtain samples of the mid-scale variability*

In contrast to the historical *in situ* data coverage, observations of SST became abundant in the last quarter of the 20th century. The National Centers for Environmental Prediction(NCEP) optimal interpolation analysis of Reynolds and Smith (Reynolds *et al.*, 2002, hereafter NCEP OI) incorporates both *in situ* observations (from ships, moored buoys and drifting buoys) and satellite data. Satellite data coverage allows for the resolution of small-scale features (such as midlatitude eddies), while *in situ* data are used to correct for bias in the satellite-derived temperatures. Because the NCEP OI is an analyzed data product, there is a preferred correlation length-scale implicit in the SST field. As discussed in Reynolds *et al.*(2002), there is an assumed Gaussian background spatial correlation with an *e*-folding distance of approximately 700 km. However, dense satellite data coverage tends to dominate the solution, making it relatively insensitive to the imposed background length-scale. The NCEP OI is available as monthly SST anomalies on a $1^\circ \times 1^\circ$ grid from 1981 to the present. The reader is referred to Reynolds *et al.*(2002) for a full discussion of the NCEP OI data product and the observations used in their analysis.

We can assume that the NCEP OI contains both globally energetic patterns of variability and smaller scale patterns of local covariability. To determine the portion of the NCEP OI that corresponds to the global patterns of variability in the orthogonal basis functions ($\mathcal{E}$), we can write the NCEP OI data as

$$y_{\mathrm{oi}} = \mathcal{E}\alpha + \epsilon. \tag{19}$$

Exploiting the orthogonality of $\mathcal{E}$, we can write the maximum-likelihood estimate of α as

$$\alpha_{\mathrm{oi}} = \mathcal{E}^{\mathrm{T}} y_{\mathrm{oi}}. \tag{20}$$

We then define the residual of the reduced-space representation as

$$y_{\mathrm{r}} \equiv y_{\mathrm{oi}} - \mathcal{E}\alpha_{\mathrm{oi}}. \tag{21}$$

These residuals comprise a new dataset for estimating the covariance of the mid-scales.

5.3. *Pseudo-likelihood estimation of the mid-scale covariance parameters*

The NCEP OI mid-scale data can be modelled as

$$y_{\mathrm{r}} \sim N(0, C + R_{\mathrm{oi}}), \tag{22}$$

where R_{oi} is a diagonal matrix of the uncertainty in the NCEP OI mid-scale data (obtained through personal communication with Richard Reynolds) and C is the mid-scale covariance matrix parametrized by the Matérn form.

The spatial scale of this problem is such that it is impractical to estimate all the parameters of the mid-scale covariance matrix simultaneously. (In the northern hemisphere Atlantic, this would amount to a nonlinear minimization over $\approx$ 17 000 correlated variables). Instead, we independently model the kernel matrices in the vicinity of each point, knit them together to form a single, global correlation matrix (as described in section 5.1) and then estimate the variance of this process with this correlation matrix. We settle on this tactic with the knowledge that mis-specification of the correlation matrix resulting from local estimation of the kernels will result in a reduction of the variance explained by this statistical model. While suboptimal, this approach provides a tractable alternative to joint estimation of all the parameters. Appendix B describes

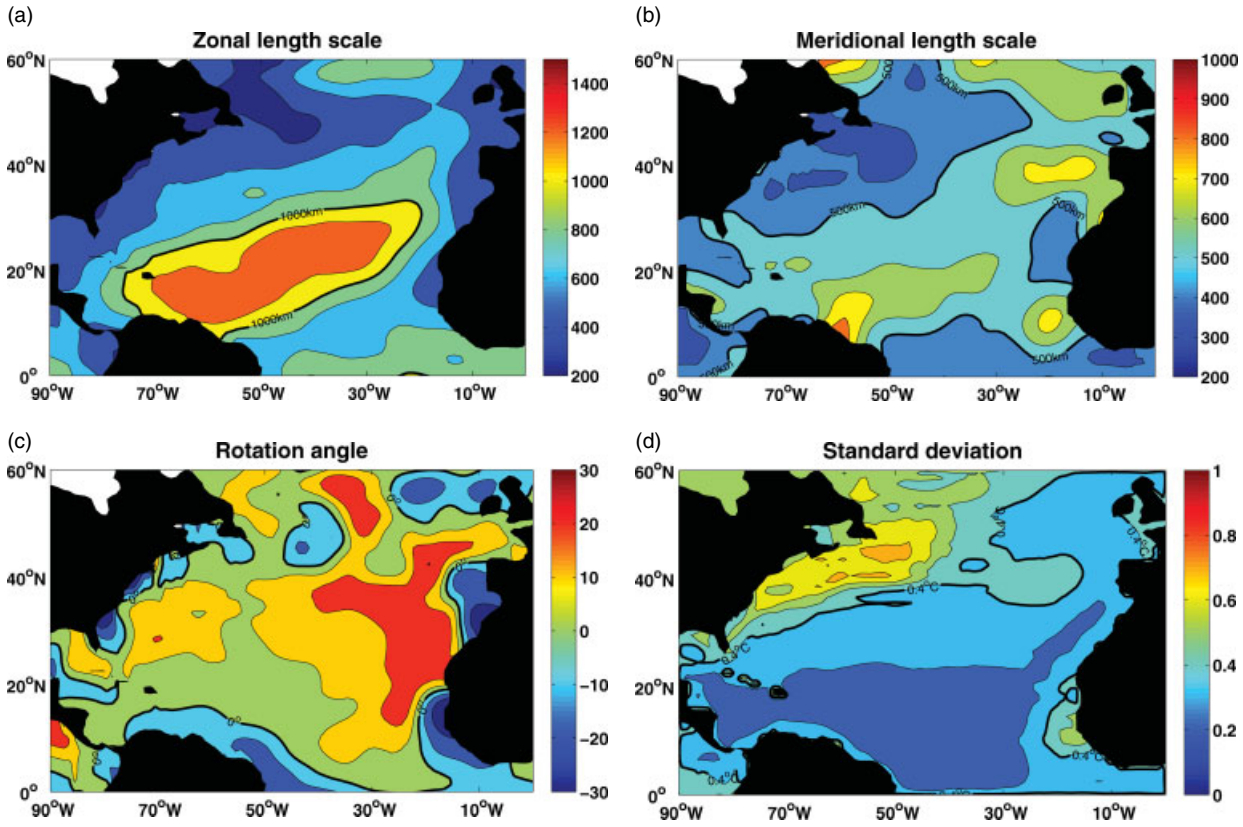

Figure 1. (a) Geographical variation in the zonal length-scale. Contour intervals are at 200 km (thick line is 1000 km). (b) Geographical variation in the meridional length-scale. Contour intervals are at 100 km (thick line is 500 km). (c) Rotation angle measured from an eastward-pointing axis. Contour intervals are at 10°, and the thick line is 0°. (d) Local standard deviation. Intervals are 0.1°C and the thick line is at 0.4°C.

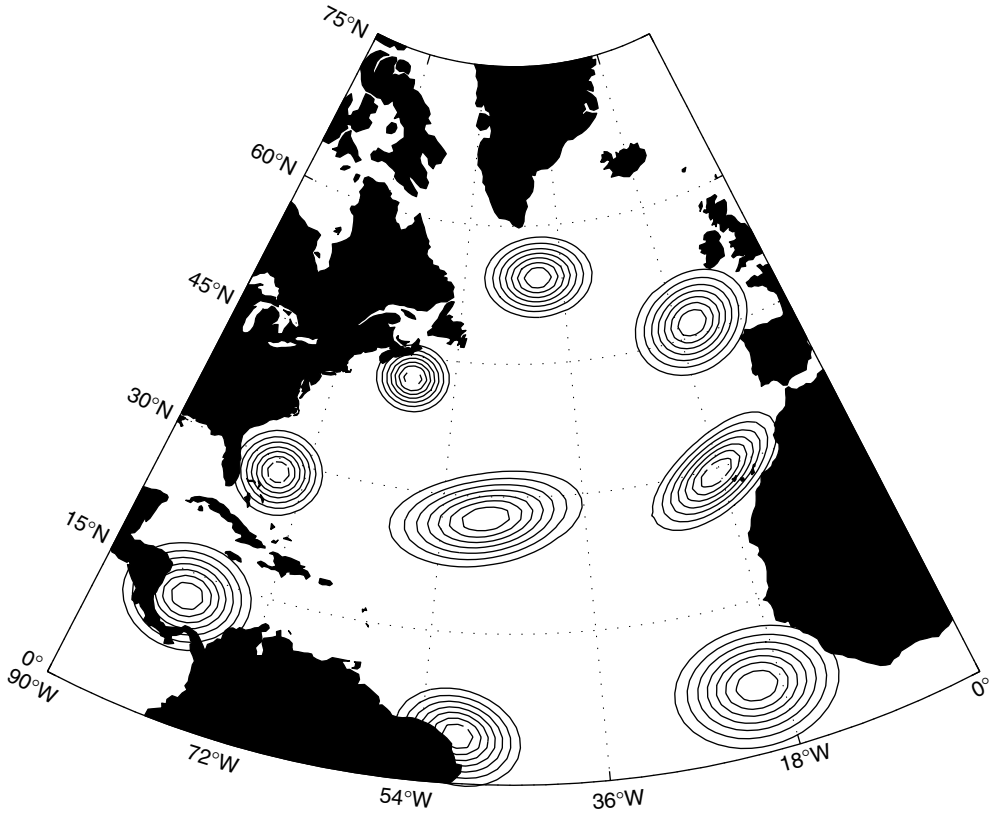

Figure 2. Contours of the correlation function at select geographical locations. This is a geometric representation of the spatial parameters from Figure 1. The smallest contour is at 0.1 and the interval of increase is 0.1.

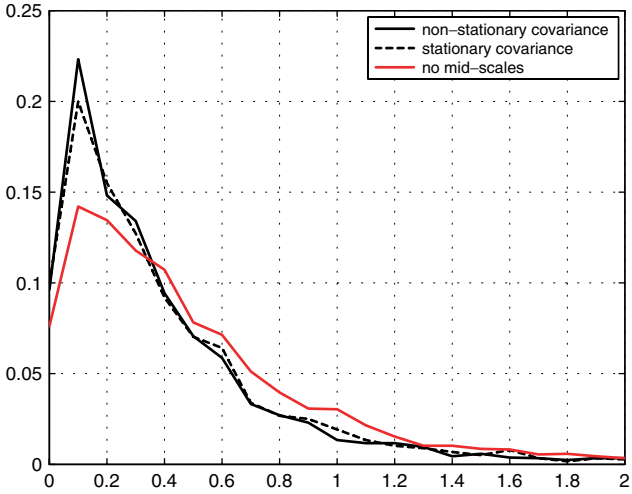

Figure 3. The empirical probability distribution of the the absolute differences between the posterior prediction and the withheld observations for the non-stationary covariance model (black solid), stationary covariance model (dashed black line) and baseline solution without the mid-scales (red line).

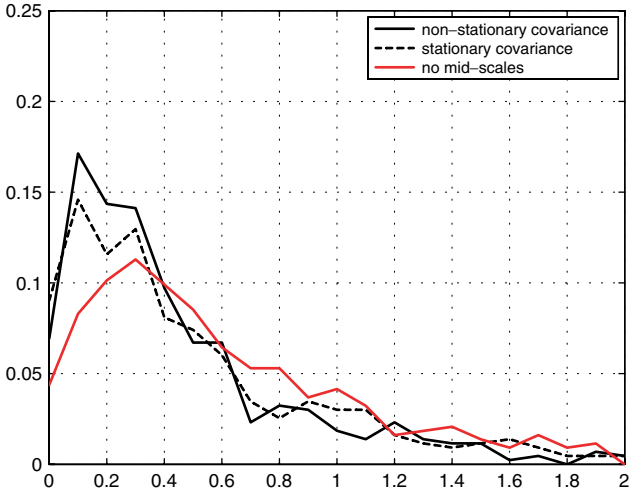

Figure 4. Same as Figure 3 except that empirical distributions were computed only for withheld observations in the Gulf Stream region.

how we determine all the parameters needed to form the mid-scale covariance matrix C.

The panels of Figure 1 show the spatial variation of the parameters of the correlation function as well as the standard deviation of the process. The ellipses in Figure 2 give a geometric representation of this same correlation model at selected geographic locations. The smallest decorrelation scales and the largest variances correspond to the path of the Gulf Stream as it separates from the east coast of the United States and crosses the basin toward Europe. The southward return-flow region of the Canary Current down the coast of north Africa is also a notable region of increased variance and decreased correlation length-scales. These are regions of high mesoscale SST variability that arise from meanders of the swift currents and eddies that shed at their boundaries. The largest scales are in the ocean interior, in the relatively quiescent centre of the subtropical gyre.

6. Verification

To verify the reconstruction, we withheld observations for comparison. Every month from 1950 to 1974 we withheld 10 randomly chosen observations from the HadSST2 dataset in the Northern Hemisphere (NH) Atlantic (3000 in total). We chose this part of the record because it is the most well observed segment not overlapping with the NCEP OI period. We then reconstructed the mid-scale SST using our non-stationary covariance model and compared results to the withheld observations. As the simplest baseline for comparison, we use the large-scale solution without any mid-scale reconstruction (i.e. $z' = 0$).

Empirical distributions of the absolute difference between the 3000 observation values and the predicted expected value are shown Figure 3. The average correction imparted by the mid-scale reconstruction is about 0.11°C and 25% of the locations used in the verification show an improvement of over 0.25°C. Furthermore, we expect that the mid-scale reconstruction is most important in eddying regions like the Gulf Stream (30°N–50°N, 75°W–45°W). Corrections to withheld observations in the Gulf Stream region tend to be larger, averaging 0.2°C with 40% of the instances showing improvement of over 0.25°C.

We also consider how well this non-stationary covariance model performs relative to a more traditional stationary statistical model for the mid-scale variability. For this test, we repeated the reconstruction with the same withheld observations but using spatially stationary zonal and meridional length-scales L_x = 1000 km and L_y = 500 km, while retaining the non-stationary prior variance.[‡] Compared with this more sophisticated baseline solution, the results are more subtle. The stationary reconstruction is shown by the black dashed lines in Figures 3 and 4. Over the entire domain, we see only a small reduction in absolute difference from using the non-stationary covariance model. In the Gulf Stream region, however, the improvement is more evident, albeit with a subtle mean improvement of 0.07°C and about 5% of the locations showing improvements greater than 0.25°C.

It is important to characterize the types of errors we expect to result from using a stationary covariance model when the underlying stochastic process is, in fact, non-stationary. Let us focus again on the Gulf Stream, because it is a high-variance region. There the stationary covariance model overestimates the length-scales. This type of mis-specification effectively reduces the resulting degrees of freedom in the system and erroneously dampens both the pointwise uncertainty in the posterior and the spatio-temporal variance of the expected value. In our experiment, for example, we had a 56% reduction in the average pointwise uncertainty variance in the Gulf Stream region when the stationary model was used and a 30% reduction in the spatio-temporal variance of the expected value. Naturally, there was also a reduction in the spatial gradients of samples drawn from the posterior. These kinds of errors are important in applications in which SST analyses are used to give boundary conditions for atmospheric models, because in frontal zones the atmosphere is responsive to the Laplacian of the SST (Minobe *et al.*, 2008).

[‡]These length-scales were chosen to be close to the grid-point average over the domain. The covariance function remained the Matérn form.

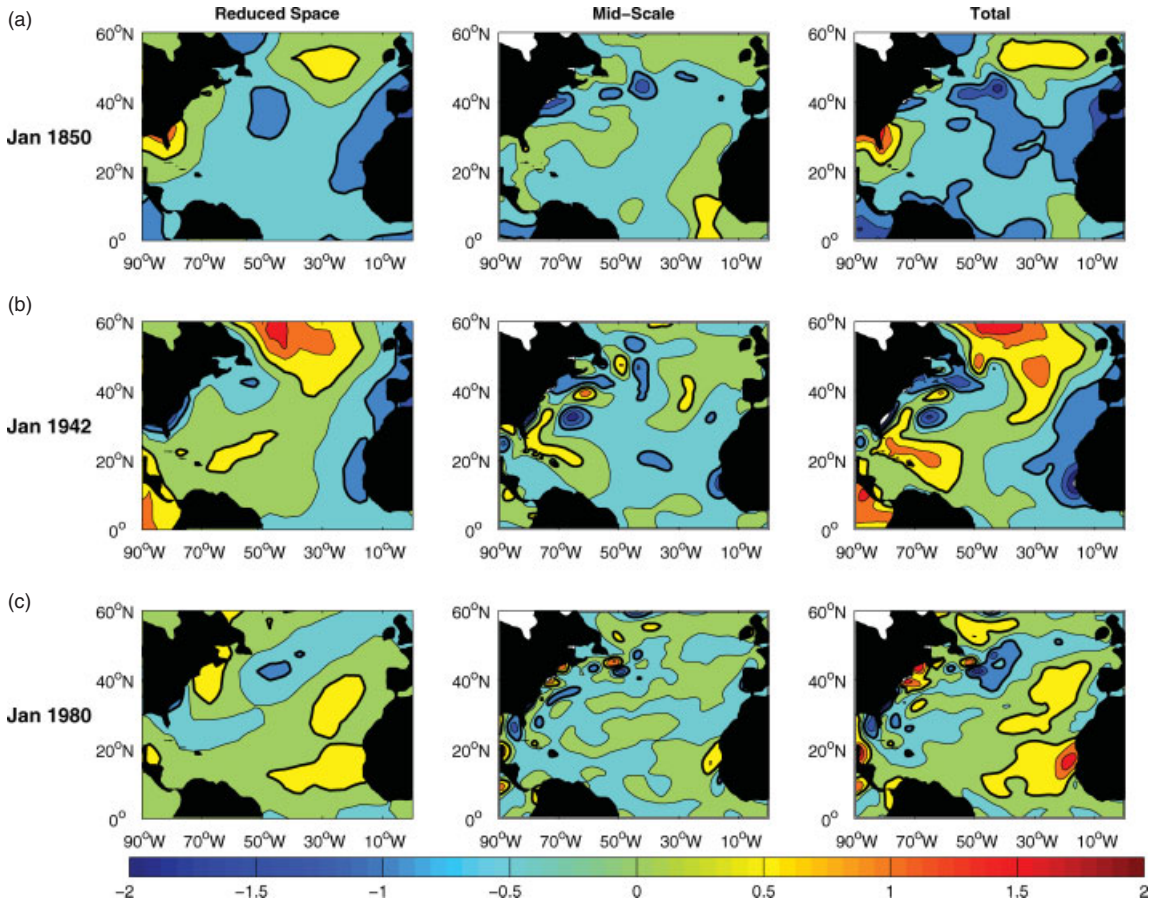

Figure 5. The first column is the expected value of the SST anomaly from the KaplanSST reduced-space optimal smoother on three dates. The second column shows the expected value of the mid-scale SST modelled in this work, and their sum is shown in the third panel. Units are in °C. Thick contour lines surround regions of absolute value greater than 0.5°C.

7. Ensemble reconstructions of the mid-scale SST

Here we present a selection of the resultant reconstructions of the non-stationary mid-scale SST. Figure 5 qualitatively illustrates the relative importance of including the mid-scales in the SST analysis. For the month of January in 1850, 1942 and 1980, the first column shows the SST anomaly from the KaplanSST reduced-space Kalman smoother. The second column shows the expected value of the mid-scale SST modelled in this work, and their sum is shown in the third panel. While the overall signal is dominated by the large-scale KaplanSST, the mid-scale reconstruction provides a significant higher resolution correction to the analysis.

Over the entire time period of the reconstruction, the spatio-temporal variability (in the expected value of SST anomaly) in the NH Altantic is increased by ∼60% due to the explicit modelling of the mid-scales. The variance spectrum of the SST anomaly with and without the inclusion of mid-scales helps us to quantify the relative importance of modelling these scales (Figure 6). This spectrum is calculated from the eigenvalues of the long-term covariance of NH Atlantic SST with and without mid-scales (circles versus dots, respectively). We computed the eigenvalues of this spectrum only for the NH Atlantic; they are not the eigenvalues of the global KaplanSST analysis. In general, the leading eigenvalues are associated with larger scale structures and the trailing eigenvalues are structurally smaller. We see that including the mid-scales in the analysis adds power to the trailing modes of the spectrum. This flattening of the spectrum is due to the reintroduction of spatial scales that are not captured by reduced-space reconstruction methods.

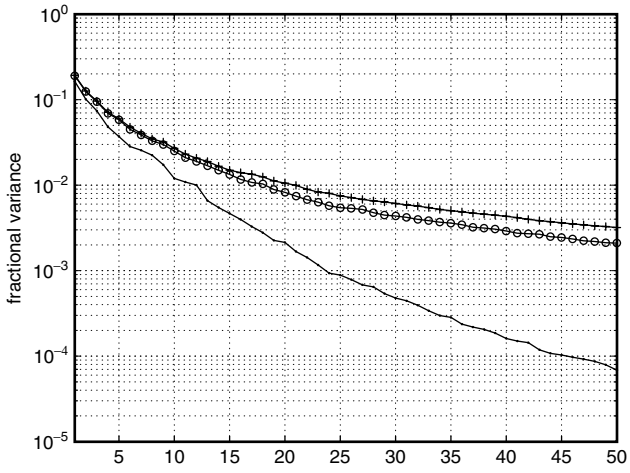

Figure 6. Spectrum of the eigenvalues of the long-term covariance of the reconstruction in the Northern Hemisphere Atlantic. Note that a log axis is used to aid in viewing the spectra. The line with small dots is the spectrum of the KaplanSST, the circles indicate the spectrum of the reconstruction expected value including the mid-scales. The line with plus symbols is the mean spectrum calculated from 20 samples of the posterior distribution, approximating the 'true' spectrum. Each spectrum is scaled by the total variance in the KaplanSST+mid-scales reconstruction.

The mid-scale reconstruction adds variance across the entire spectrum. In particular, there is a very modest, but detectible, contribution to the leading few modes. Figure 7

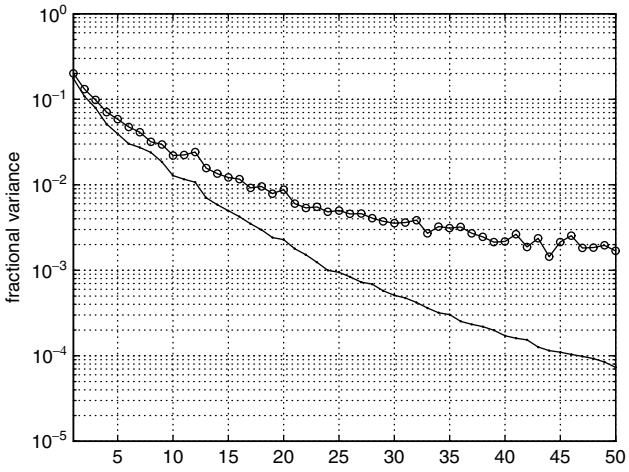

Figure 7. As in Figure 6, the line with small dots is the spectrum of the KaplanSST and the circles indicate the spectrum of the reconstruction expected value including the mid-scales. Here the total reconstructed spectrum (KaplanSST+mid-scales) has been projected on to the eigenvector basis from the KaplanSST.

shows a similar spectrum, but with the reconstructions projected on to the orthogonal structures of the KaplanSST. Here we see unequivocally that the mid-scale reconstruction has a projection on to the global solution. This is not unexpected. Independence in the prior does not guarantee independence in the posterior. It simply reinforces the idea that a joint modelling of the large and mid-scales is an important next step.

Figure 8 illustrates the long-term mean in large and mid-scale solution components. The mid-scale reconstruction introduces a cooling in the temperatures along the northern edge of the Gulf Stream region, along with a smaller warming along the southern edge. This is a course resolution of the Gulf Stream pathway. Figure 9 shows a time series of the reconstruction at 50°W and 43°N, a locally cool pivot point where the Gulf Stream turns northward to form the North Atlantic Drift. We see that this feature is temporally consistent, emerging in the late 1800s and persisting through most of the record. The grey shadow in Figure 9 is the posterior uncertainty bound (95% confidence intervals). Since the prior covariance is temporally stationary, the fluctuations in the posterior uncertainty are driven by the availability (or absence) of *in situ* HADSST2 data near this location.

A close examination of this SST anomaly time series in the context of its uncertainty reveals a slightly different story. There is little evidence, in fact, that this cool feature was ever absent. The lack of observations leads to a large uncertainty during the preceding three-quarters of a century.

Of course, pointwise uncertainties are only part of the story. The posterior uncertainties in the mid-scales are correlated in space. We illustrate these covariance structures in Figures 10–12. For January in 1850, 1942 and 1980, we have contoured the expected value and pointwise uncertainty of the mid-scale SST anomaly (top panels). In the remaining four panels we present a sample drawn from the full posterior uncertainty distribution. The black dots on the maps are points where *in situ* observations were available. As we expect, the realizations tend to be most similar (to each other and to the expected value) when the ocean is densely observed. In the early record, as well as during times of changing shipping routes (such as the early

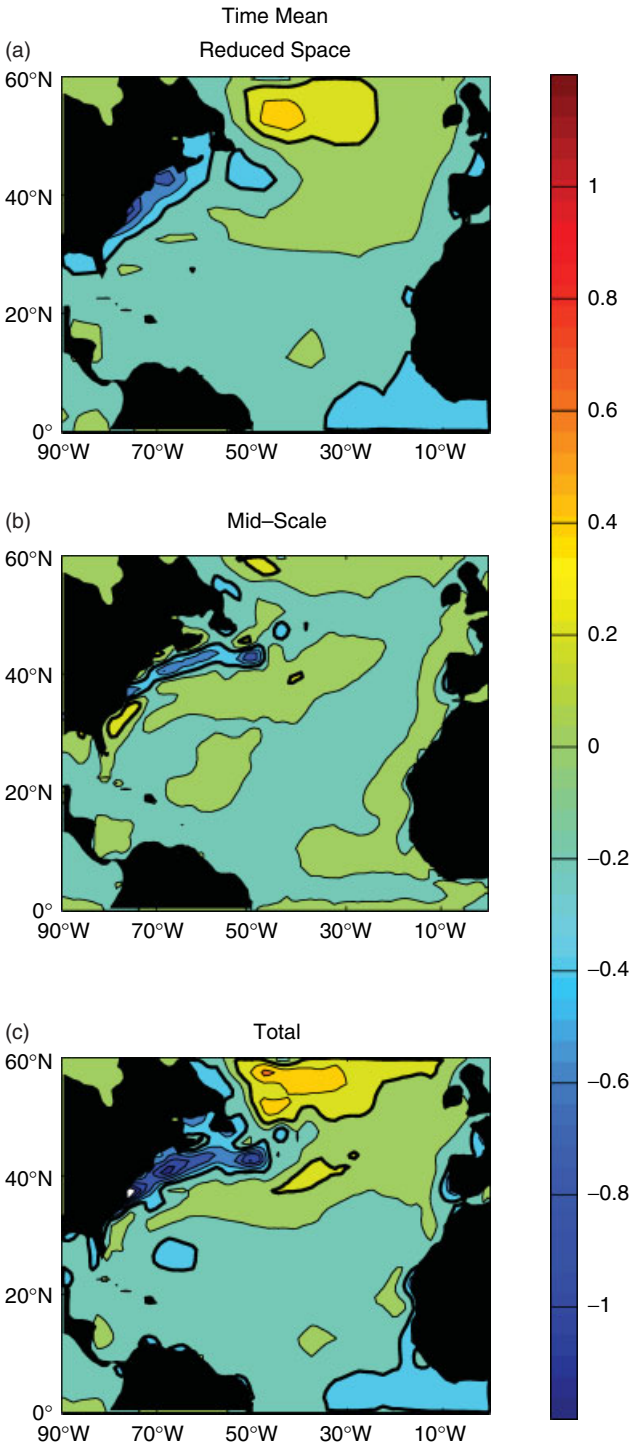

Figure 8. Top panel: the long-term mean (1850–2008) of the KaplanSST. Middle panel: long-term mean of the expected value of the reconstructed mid-scales. Bottom panel: the long-term mean of the KaplanSST plus mid-scales. Units are °C. Thick contour lines surround regions of absolute value greater than 0.2°C. Note the time mean structures in the Gulf Stream regions resulting from the mid-scale reconstruction.

1940s), we can see that the ensemble members cluster in agreement at the observation locations. We can also see the underlying covariance structures that were specified in the prior covariance are emerging in the posterior.

To demonstrate the utility of representing the reconstructions via an ensemble of realizations, we can compute the long-term covariance eigenspectrum multiple times from samples taken from the posterior distribution. The line

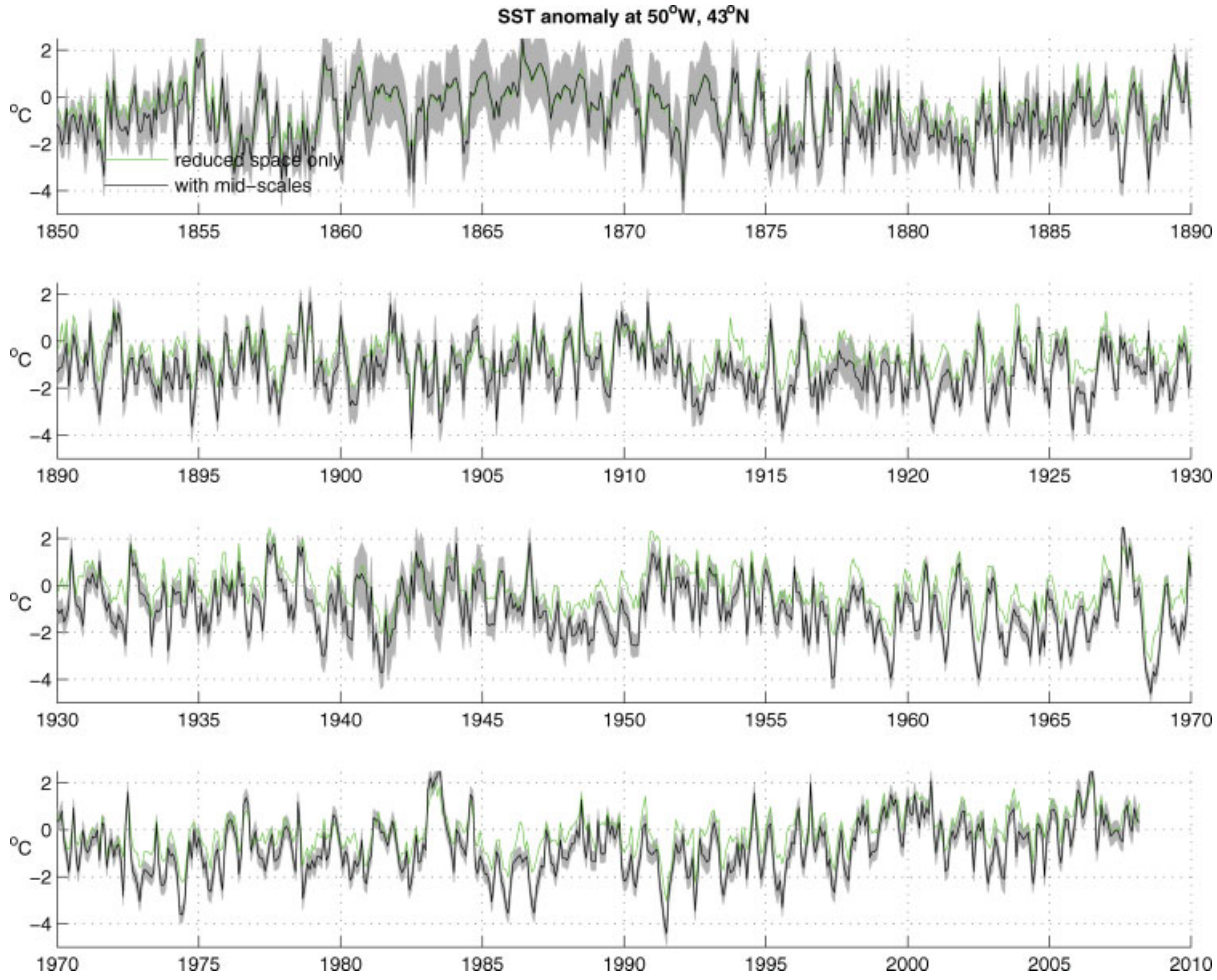

Figure 9. Time series of the SST anomaly at 50° W, 43° N (a point in the Gulf Stream region). The black line is the expected value of the KaplanSST + mid-scale reconstruction, with grey shading showing the 95% confidence intervals of the mid-scale only. The green line shows the time series from just the KaplanSST. Note the long-term cooling correction implied by the additional reconstruction of mid-scale variability in the Gulf Stream region.

marked with '+' in Figure 6 shows the average eigenspectrum for 20 realizations. This increase (∼10%) reflects the additional variability in the SST due to posterior uncertainty in the mid-scales. This additional variability will make up an even greater percentage in the early part of the record, when observations are scarce. Variability estimated from many individual realizations approximates the 'true' variability, whereas variability found from the expected value will always be biased low.

8. Discussion

This work presents the statistical modelling and reconstruction of mid-scale SST anomalies in the NH Atlantic. We use a recently developed statistical parametrization by Paciorek and Schervish (2006) for the prior covariance that allows for non-stationarity in the anisotropic correlation length-scales. The benefit of this approach is that it allows us to form a positive-definite posterior uncertainty covariance from which samples can be easily drawn. The correlated error structures in the prior naturally emerge in realizations drawn from the posterior.

Because this work is focused on the modelling and reconstruction of the mid-scale variability, we have not combined their estimates with the posterior distribution of the KaplanSST. A complete modelling of the SST anomaly would treat the global and mid-scales as a joint distribution because, while these scales can be assumed independent in the prior, they are correlated in the posterior. However, one might want to avoid joint modelling for a number of reasons. Primary among them is that the large-scale part of the solution is global in size and autoregressive in time. Dimensional reduction is thus computationally necessary for a tractable solution. In subdomains where we also want to compute the mid-scale part, such as the NH Atlantic basin considered here, it is helpful to draw on the results of the large-scale reconstruction while avoiding a global multiscale reconstruction. The hierarchical form of the full solution shown in section 4 suggests a simple Monte Carlo approach to drawing samples from the full large and mid-scale solution. This is a fruitful direction for future work.

Lorenc (1986) notes that geophysical fluids exhibit energy over a wide range of spatial scales. It is not altogether clear how these scales should be separated. We have implicitly assumed in this study that global scales of covariability can be captured via a reduced-space representation and the remaining variability can be modelled with a locally supported covariance. However, even within this framework it is clear that there are multiple scales present at this mid-level too. We can interpret our non-stationary parametrization as an attempt to select for correlation structures that are locally dominant. Even with this simplified description, mid-scale reconstructions

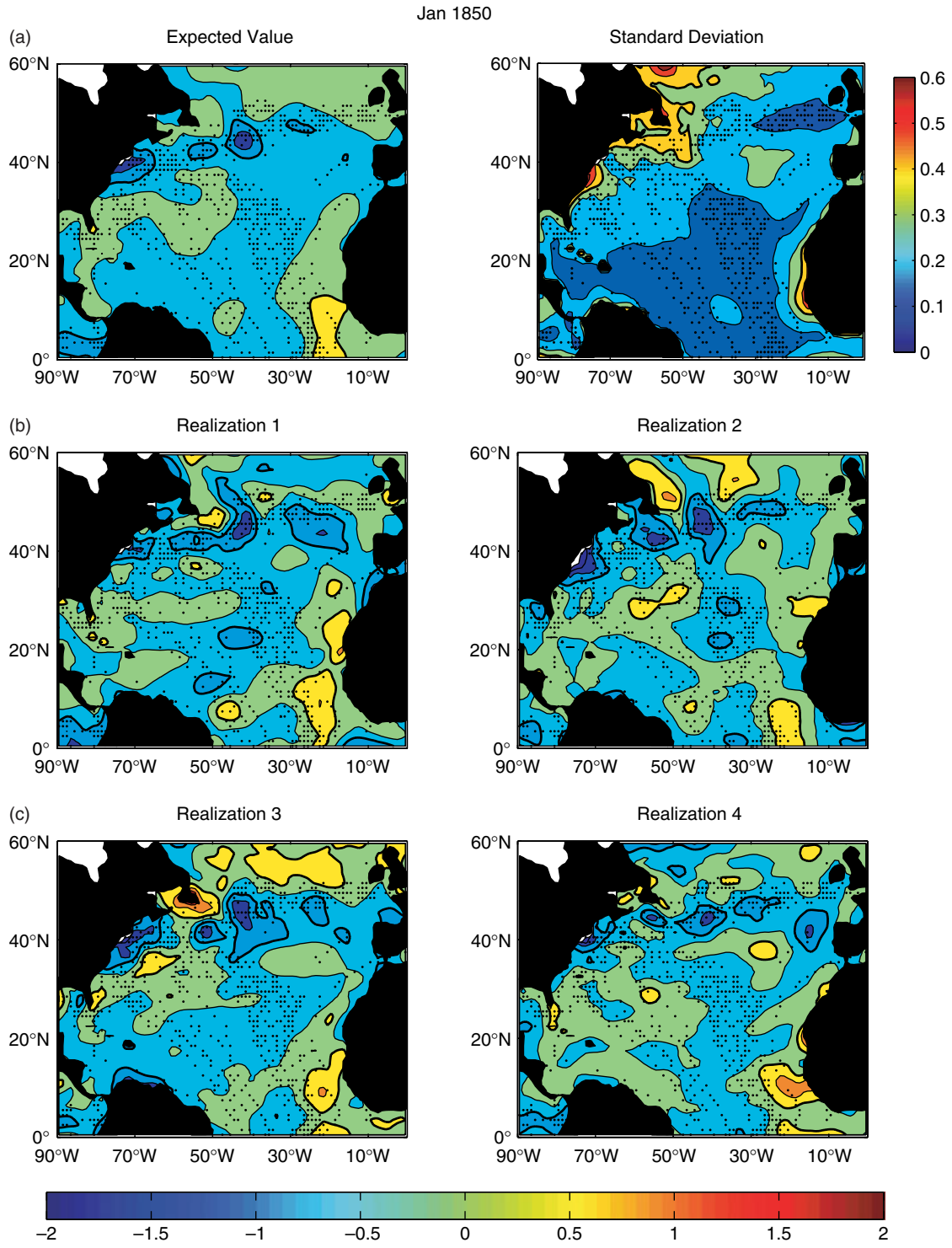

Figure 10. Expected value, standard deviation and four realizations from the posterior distribution of the mid-scale SST in January of 1850. Units are °C. Each black dot indicates the existence of at least one *in situ* observation. Thick contour lines in expected value and realization plots surround regions of absolute value of greater than 0.5°C. Thick contour lines in the standard deviation plot surround regions of greater than 0.4°C.

contribute significantly to both the long-term mean (Figures 8 and 9) and the long-term covariance spectrum (Figure 7).

For problems of the size often encountered in the climate sciences, computational constraints can limit the direct evaluation of (8) and (9). Approximations and iterative methods used to deal with this practical limitation tend to focus on the computation of the expected value (Lorenc, 1986). Lorenc (1986) and Pedder (1993) point out that methods focused on computation of the expected value tend to be best suited for problems with reasonably reliable and densely distributed observations. When data are sparse and noisy, as happens in the early part of the record presented here, a full specification of the prior covariance becomes preferable.

A description of the covariance matrix and its evolution through time is expensive to store and disseminate. A multivariate Gaussian distribution with n spatial points and m time points requires $mn(mn+1)/2$ numbers to quantify the uncertainty. Users of climate datasets may not have the computational resources or mathematical expertise to make

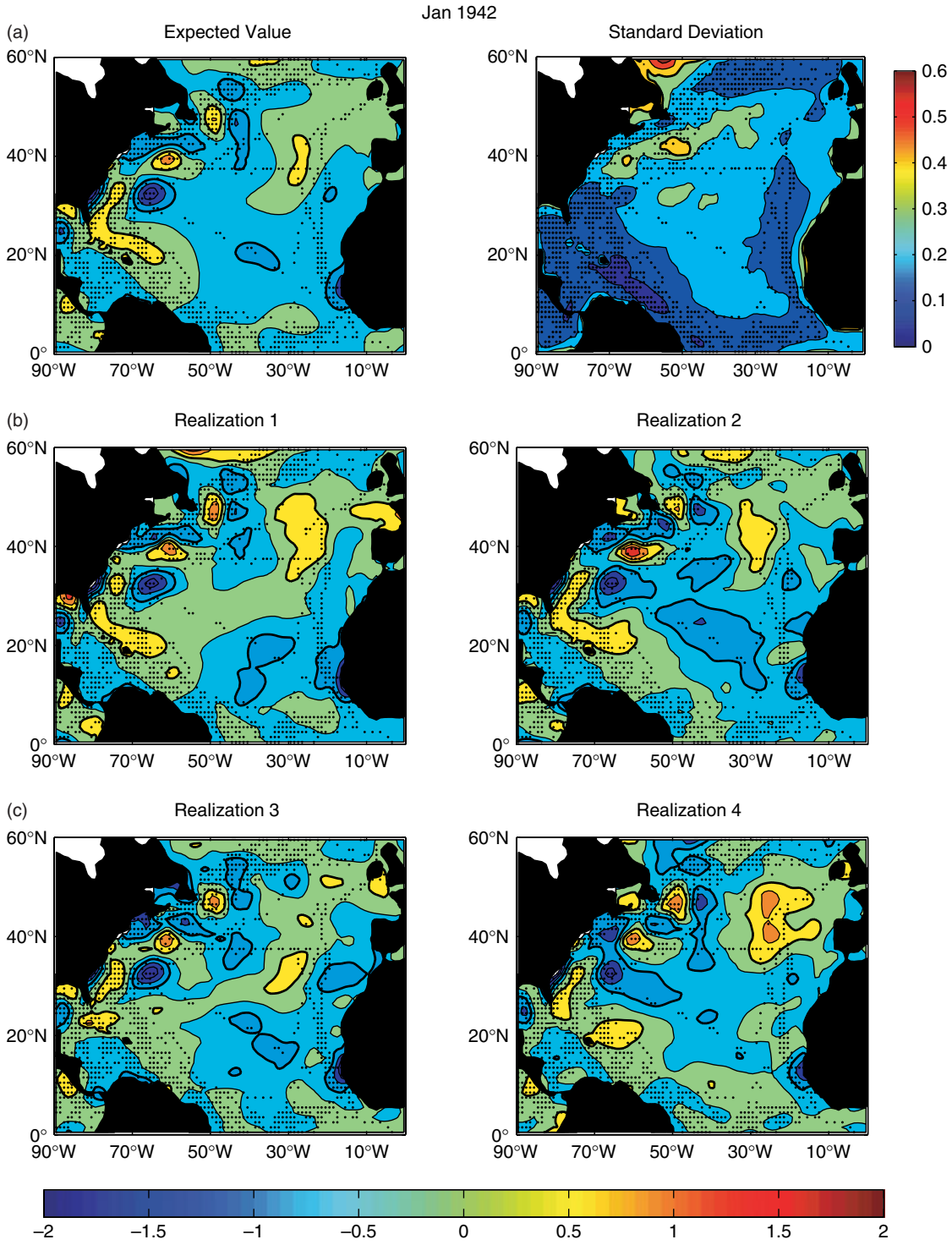

Figure 11. Same as Figure 10, but for January 1942.

use of a full covariance matrix of uncertainty information. As an alternative, samples from the posterior distribution can be distributed. To the extent that the data model is adequate, the samples form an ensemble of possible realizations of the true SST. Users can perform standard climate data analyses on the multiple realizations, building up uncertainty information in a Monte Carlo fashion.

In the specific case of generating SST datasets, another important application is the boundary forcing of atmospheric models. It is not uncommon for analyzed SST datasets to be viewed as non-probabalistic, with the uncertainty in the system assumed to stem from the internal variability of the atmosphere only. Instead, atmospheric modellers could use ensembles as a set of readily accessible and statistically rigorous possibilities with which to force their models.

The presentation of an ensemble of possible realizations of SST is especially important in data-poor regions of the ocean. It is a natural consequence of Bayesian inference that the expected value of the reconstruction in unobserved areas will relax towards the mean of the prior distribution. When considered outside the context of the full covariance information, data users can falsely interpret these locations in the data record as less energetic. A proper interpretation, in contrast, would be that there is little constraint on the

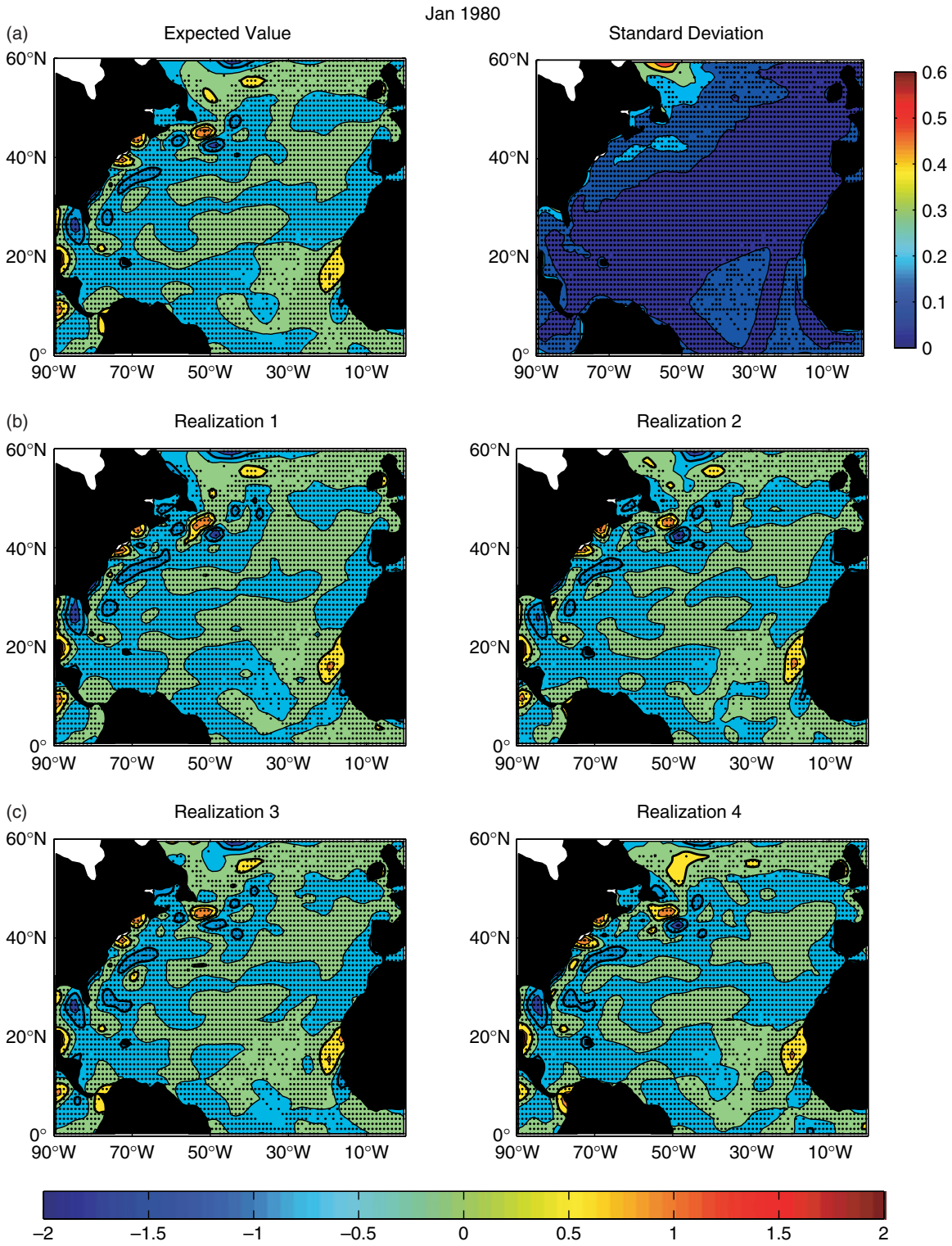

Figure 12. Same as Figure 10, but for January 1980.

possible states of the system. The sparse and irregular nature of historical data thus makes ensemble presentation an important contribution to the research community.

Finally, we suggest that with some ingenuity this method could be extended to global reconstructions. While the computational expense of covariance generation and sampling of mid-scales on a global domain is a significant obstacle, it may be possible to exploit the inherent sparsity of the correlation matrix for computational efficiency. One could also make use of the idea that mid-scale features are a relatively minor part of the total reconstruction in many regions of the ocean. It may be the case that mid-scale reconstruction in isolated domains is sufficient.

Acknowledgements

This research was supported by the National Science Foundation (NSF) grant ATM04-17909 and ATM04-1797 and National Aeronautics and Space Administration (NASA) grants NNG04GL28G and NNX09AF44G. The

formulation of this project, in its early version, was influenced by Craig Johns. The authors are grateful to Doug Nychka for discussions and encouragement that helped bring this project to its later form. Thanks also to Richard Reynolds for providing error estimates for the NCEP OI SST dataset. Comments from two anonymous reviewers resulted in significant improvements to the manuscript. The National Center for Atmospheric Research is managed by the University Corporation for Atmospheric Research under the sponsorship of the NSF.

A. Matrix form for the moments of the joint posterior

In section 4 of the article, we show that the posterior probability of the SST field z can be written as the integral over the product of two Gaussians:

$$p(z|y) = \int N(z|\mu_{z|\alpha}, P_{z|\alpha})\, N(\alpha|\mu_\alpha, P_\alpha)\, \mathrm{d}\alpha, \tag{A1}$$

where the first factor in the integral is the probability of the total SST field conditional on the large scale and the second term is the marginal probability of large scales. Here we obtain a compact matrix form for the results of this integral. Let us define the Kalman gain matrix K in the usual fashion:

$$K \equiv C\mathcal{H}^{\mathrm{T}}(\mathcal{H}C\mathcal{H}^{\mathrm{T}} + \mathcal{R})^{-1} \tag{A2}$$

(the reader is referred to the body of the article and equations (1), (4), (8) and (9) for a description of the variables). Using K, the expected value of the conditional probability can be written

$$\mu_{z|\alpha} = \mathcal{E}\alpha + K(y - \mathcal{H}\mathcal{E}\alpha). \tag{A3}$$

The right-hand side of (A1) can be expanded using the definition of normal distributions:

$$p(z|y) \approx \frac{1}{A} \int \mathrm{e}^{-\frac{1}{2}(z-\mu_{z|\alpha})^{\mathrm{T}} P_{z|\alpha}^{-1}(z-\mu_{z|\alpha})} \times \mathrm{e}^{-\frac{1}{2}(\alpha-\mu_\alpha)^{\mathrm{T}} P_\alpha^{-1}(\alpha-\mu_\alpha)}\, \mathrm{d}\alpha, \tag{A4}$$

where A is a normalizing coefficient. Using the definition of K and expanding $\mu_{z|\alpha}$ in terms of α and y, we can then write

$$p(z|y) \approx \frac{1}{A} \int \mathrm{e}^{-\frac{1}{2}(z-Ky-[I-K\mathcal{H}]\mathcal{E}\alpha)^{\mathrm{T}} P_{z|\alpha}^{-1}(z-Ky-[I-K\mathcal{H}]\mathcal{E}\alpha)} \times \mathrm{e}^{-\frac{1}{2}(\alpha-\mu_\alpha)^{\mathrm{T}} P_\alpha^{-1}(\alpha-\mu_\alpha)}\, \mathrm{d}\alpha. \tag{A5}$$

In this form, the integral over α is straightforward, leading to another Gaussian:

$$p(z|y) \approx N(\alpha|\mu_z, P_z) = \frac{1}{A^*} \mathrm{e}^{-\frac{1}{2}(z-\mu_z)^{\mathrm{T}} P_z^{-1}(z-\mu_z)}, \tag{A6}$$

where A^* is a new normalizing coefficient and the moments of this posterior are given by

$$\mu_z = Ky + (I - K\mathcal{H})\mathcal{E}\mu_\alpha = \mathcal{E}\mu_\alpha + K(y - \mathcal{H}\mathcal{E}\mu_\alpha) \tag{A7}$$

and

$$P_z = P_{z|\alpha} + (I - K\mathcal{H})\mathcal{E}\, P_\alpha\, \mathcal{E}^{\mathrm{T}}(I - K\mathcal{H})^{\mathrm{T}}. \tag{A8}$$

Note that (A8) can be rewritten as

$$P_z = P_{z|\alpha} + \mathcal{E}\, P_\alpha\, \mathcal{E}^{\mathrm{T}} + [K\mathcal{H}\mathcal{E}\, P_\alpha \mathcal{E}^{\mathrm{T}}\mathcal{H}^{\mathrm{T}}K^{\mathrm{T}} - K\mathcal{H}\mathcal{E}\, P_\alpha\mathcal{E}^{\mathrm{T}} - \mathcal{E}\, P_\alpha\mathcal{E}^{\mathrm{T}}\mathcal{H}^{\mathrm{T}}K^{\mathrm{T}}]. \tag{A9}$$

The first two terms on the right-hand side correspond to covariances of mid- and large-scale posterior distributions described in the text. The term in brackets involves both large- and mid-scale patterns through the factors K and $\mathcal{E}$. Its calculation requires alternating between local (high-resolution) and reduced-space (low-dimensional) procedures. An even higher computational burden must be met if the large-scale solution involves estimation both in space and time (as in the Kalman smoother). Because of these difficulties, using a Monte Carlo approach for sampling from integral (A1) is a more practical approach.

B. Parameter determination for mid-scale covariance matrix

Section 5.3 presents the parameters of the mid-scale covariance matrix that we estimate from the NCEP OI mid-scale data. Here are the details of the maximum pseudo-likelihood method used for the estimation.

To isolate the spatial correlation from the variances, we begin by standardizing the NCEP OI mid-scale data to have zero mean and variance of unity over the period 1981–2008 ($\tilde{y}_r$). Based on exploratory analysis of the NCEP OI mid-scale data, a priori we prescribe the Matérn shape parameter $\nu = 3$. Using this value of the smoothness parameter, we proceed to estimate the kernel matrix (16) centred at each grid point i. We can express the bivariate distribution of two points as

$$\begin{bmatrix} \tilde{y}_{r,i} \\ \tilde{y}_{r,j} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & C(i,j) \\ C(i,j) & 1 \end{bmatrix} \right), \tag{B1}$$

where $C(i,j) = f(\Sigma, \nu)$ is the anisotropic Matern covariance from (14)–(16) with $\sigma_i = \sigma_j = 1$. We define a likelihood function for the data centred at grid point i as the product of these bivariate normals over all points j within 20° of the location of i. We can express the maximum-likelihood estimate of the parameter Σ_i as

$$\Sigma_i = \arg\max_{\hat{\Sigma}\in\Theta} \prod_{j\neq i} p(\tilde{y}_{r,i}, \tilde{y}_{r,j}|\hat{\Sigma}, \nu) \tag{B2}$$

where Θ is the set of all 2×2 real, symmetric, positive-definite matrices. Since the multivariate normal likelihood function in (B2) is nonlinear in all three variables that comprise Σ, its maximization is performed numerically using a standard Nelder–Mead search algorithm.

Given the kernels at each point, we now combine them using (17) to form the correlation matrix $\mathcal{Q}$. The global mid-scale covariance matrix can now be defined as

$$C \equiv \mathcal{T}^{1/2}\mathcal{Q}\mathcal{T}^{1/2}, \tag{B3}$$

where $\mathcal{T}$ is a diagonal matrix containing the variance of the process at each spatial point. We can now write the multivariate normal distribution for the mid-scale data using our correlation matrix as $y_{\mathrm{r}} \sim N(0, \mathcal{T}^{1/2}\mathcal{Q}\mathcal{T}^{1/2} + \mathcal{R}_{\mathrm{oi}})$. The

maximum-likelihood estimate of the variance would then be given by

$$\mathcal{T} = \arg\max_{\hat{\mathcal{T}}} p(y_r|\hat{\mathcal{T}}^{1/2}\mathcal{Q}\hat{\mathcal{T}}^{1/2}, \mathcal{R}_{oi}). \quad \text{(B4)}$$

This too is an impractically large maximization, so we solve iteratively for $\mathcal{T}$ using a pseudo-likelihood method. Pseudo-likelihood formulations assume conditional independence of geographically distant points. Specifically, we assume conditional independence of points separated by greater than 20° of latitude or longitude.

This reduces the problem to finding the maximum pseudo-likelihood estimate within a series of smaller (20° × 20°) subregions centred on the point i. $\mathcal{T}$ is constrained at its upper bound by the pointwise sample variances of the NCEP OI mid-scale data. Although the minimization is done over the entire subregion, only the solution at grid point i is retained. As the pseudo-likelihood maximization procedure cycles through all grid points i, the solution is also constrained by any variances within the subsets that have previously been computed. This enforces continuity between neighbouring grid points. This process is iterated over the domain until convergence is achieved.

References

Anderson JL. 1996. A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J. Climate* **9**: 1518–1530.

Dommenget D. 2007. Evaluating eof modes against a stochastic null hypothesis. *Clim. Dyn.* **28**: 517–531.

Ghil M, Cobh S, Tavantzis J, Bube K, Isaakson E. 1981. *Applications of estimation theory to numerical weather prediction, in dynamic meteorology: Data assimilation methods*. Springer-Verlag: New York; pp 139–224.

Handcock MS, Wallis JR. 1994. An approach to statistical spatial-temporal modelling of meteorological fields (with discussion). *J. Amer. Statist. Assoc.* **89**: 368–378.

Handcock M, Stein M. 1993. A Bayesian analysis of kriging. *Technometrics* **35**: 403–410.

Kaplan A, Kushnir Y, Cane M, Blumenthal M. 1997. Reduced space optimal analysis for historical datasets: 136 years of Atlantic sea-surface temperatures. *J. Geophys. Res.* **102**: 27835–27860.

Kaplan A, Cane M, Kushnir Y, Clement A, Blumenthal M, Rajagopalan B. 1998. Analyses of global sea-surface temperature 1856–1991. *J. Geophys. Res.* **103**: 18567–18589.

Kaplan A, Kushnir Y, Cane M. 2000. Reduced space optimal interpolation of historical marine sea level pressure: 1854–1992. *J. Climate* **13**: 2987–3002.

Kent E, Challenor P. 2006. Toward estimating climatic trends in SST, part 2: Random errors. *J. Atmos. Oceanic Technol.* **23**: 476–486.

Lorenc AC. 1986. Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.* **112**: 1177–1194.

Minobe S, Kuwano-Yoshida A, Komori N, Xie S, Small R. 2008. Influence of the Gulf Stream on the troposphere. *Nature* **452**: 206–209.

Paciorek C, Schervish M. 2006. Spatial modelling using a new class of non-stationary covariance functions. *Environmetrics* **17**: 483–506.

Pedder MA. 1993. Interpolation and filtering of spatial observations using successive corrections and Gaussian filters. *Mon. Weather Rev.* **121**: 2889–2902.

Rayner N, Parker D, Horton E, Folland C, Alexander L, Rowell D, Kent E, Kaplan A. 2003. Globally complete analyses of sea-surface temperature, sea ice and night marine air temperature, 1871–2000. *J. Geophys. Res.* **108**: 4407.

Rayner N, Brohan P, Parker D, Folland C, Kennedy J, Vanicek M, Ansell T, Tett S. 2006. Improved analyses of changes and uncertainties in sea surface temperature measured in situ since the mid-nineteenth century: the HadSST2 dataset. *J. Climate* **19**: 446–496.

Rayner N, Kaplan A, Kent EC, Reynolds RW, Brohan P, Casey KS, Kennedy JJ, Woodruff SD, Smith TM, Donlon C, Breivik LA, Eastwood S, Ishii M, Brandon T. 2009. 'Evaluating climate variability and change from modern and historical SST observation'. In *Proceedings of OceanObs '09: Sustained ocean observations and information for society*, Vol. 2. Hall J, Harrison DE, Stammer D (eds). Noordwijk, the Netherlands Eur. Space Agency Spec. Publ. WPP-306 European Space Agency: Noordwijk, the Netherlands.

Reynolds R, Rayner N, Smith T, Stokes D, Wang W. 2002. An improved in situ and satellite SST analysis for climate. *J. Climate.* **15**: 1609–1625.

Shriver J, O'Brien J. 1995. Low-frequency variability of the equatorial Pacific Ocean using a new pseudostress dataset: 1930–1989. *J. Climate* **8**: 2762–2786.

Smith T, Reynolds R, Livezey R, Stokes D. 1996. Reconstruction of historical sea-surface temperatures using empirical orthogonal functions. *J. Climate* **9**: 1403–1420.

Smith TM, Reynolds RW. 1998. A high resolution global sea surface temperature climatology for the 1961–90 base period. *J. Climate* **11**: 3320–3323.

Worley S, Woodruff S, Reynolds R, Lubker S, Lott N. 2005. Icoads release 2.1 data and products. *Int. J. Climatol. (CLIMAR-II Special Issue)* **25**: 823–842. DOI:10.1002/joc.1166.

