

# Bayesian PCA for Reconstruction of Historical Sea Surface Temperatures

Alexander Ilin and Alexey Kaplan

**Abstract**—In this work, reconstructions of historical global sea surface temperatures (SST) are performed using Bayesian principal component analysis (PCA). Two PCA models are examined: a model with isotropic noise and a model which takes into account data uncertainty due to sampling errors. Inference is done by variational Bayesian learning. The methods are compared with a more traditional technique, reduced space optimal interpolation (RSOI), that is currently used in producing standard historical SST analyses. New methods were applied to the MOHSST5, an observational data set for 1856–1991 period from the United Kingdom Meteorological Office, that was used in a previously published application of the RSOI. Data uncertainty specification was also identical to the one used in that RSOI application, hence the performances of all reconstructions are directly comparable. Reconstruction results for 1982–1991 period are tested via their comparison with the NOAA monthly 1° OI (version 2) that blends in situ observations with the much better sampled satellite data. New reconstructions slightly outperform the published RSOI reconstruction in this test and suggest that further improvements are possible.

## I. INTRODUCTION

One of the most significant problems in contemporary climatology is to evaluate recent climate trends in the context of longer climate variability. Therefore, studies of centennial and decadal variability of global surface temperatures have acquired great importance. These studies are complicated by scarce coverage and often poor quality of historical climatological measurements. Relatively complete observations of global ocean temperatures became possible only after 1981, with the advent of satellite measurements. Prior to this period most of ocean surface observations were extracted from ship logs. In pre-satellite era, the period after 1950 is characterized by systematically more abundant observations and more complete spatial coverage than earlier periods. Buoy observations constitute a significant portion of the total in situ (buoy and ship) measurements taken in the post-1950 period.

Since sea surface temperature (SST) is one of the most visible variables in the climate change discussion, significant efforts have been devoted to reconstructing its historical spatial and temporal variability using statistical principles (e.g., [1], [2], [3], [4], [5], [6], [7], [8]). Reduced space

estimation ([1], [2], [4], [6]) is now a central element in all globally complete analyses of historical SST that are currently in wide use, that is United Kingdom (U.K.) Hadley Centre sea ice and SST data set, version 1 (HadISST1 [5]) and United States National Climate Data Center (NCDC) Extended Reconstructed SST data set (ERSST, version 2 [7] and version 3 [8]). HadISST1 [5] uses reduced space optimal interpolation (RSOI) [3] for their reduced space estimation step, and ERSST products ([7], [8]) use a somewhat different version of a reduced space analysis [6]. Systematic comparison of these approaches for a common set of inputs has never been performed and is in plans for near future [9].

The starting point of any version of the reduced space analysis is to find a low-dimensional vector space in which target climate fields can be accurately approximated. RSOI [2] uses empirical orthogonal functions (EOFs) found by principal component analysis (PCA). PCA is performed by eigen-decomposition of the data covariance matrix estimated from available measurements. Hereinafter term EOF is reserved for spatial patterns found in the PCA analysis, while principal component (PC) refers to an amplitude time series corresponding to an individual EOF. All PCs have variance one and are uncorrelated with each other. Since EOFs capture correlations between measurements in different spatial locations, they can be used to compute reconstructions in places of missing data.

The reduced space method developed in [2] for SST reconstruction also used the following assumptions: (1) the target variable (SST) is characterized by long-range spatial and temporal correlations; (2) even though the quality of measurements can vary considerably due to sampling errors, the variance of these errors can be roughly estimated *a priori*. In [3], the method was applied to a global historical SST data set of in situ measurements (satellite-based data products were used later for verification of these analyses). The procedure was as follows:

- 1) Estimate elements of the data covariance matrix  $\mathbf{C}$  from available measurements.
- 2) Correct the elements of  $\mathbf{C}$  by taking into account the information on data quality (sampling errors) and the spatial structure of data (by smoothing the covariances in both the longitudinal and latitudinal directions).
- 3) Compute EOFs by eigen-decomposition of  $\mathbf{C}$ .
- 4) Compute time series of principal components using a set of leading EOFs (“reduced space basis”) and the data quality information.
- 5) Rescale EOFs based on the computed principal components.

Alexander Ilin is with the Adaptive Informatics Research Center at Helsinki University of Technology, Finland (email: alexander.ilin@hut.fi). Alexey Kaplan is with the Division of Ocean and Climate Physics of Lamont-Doherty Earth Observatory of Columbia University (alexeyk@ldeo.columbia.edu).

This work was supported by the Academy of Finland project ‘Novel machine learning techniques for studying climate variability and its impacts’ and the U.S. National Science Foundation grant ATM 04-17909.

To reduce the error of EOF estimation, these five steps were performed only on the time period from 1951 on, which was characterized by relatively good observational coverage. Note that in principle steps (4) and (5) could require iterating until convergence, but in practical applications ([2], [3]) the convergence occurred after the first iteration. RSOI then used resulting EOFs and the data quality information to compute reconstructions for the entire period of available historical SST observations. Another method, reduced space optimal smoother (RSOS), performs an extra step in order to further improve reconstructions:

- 6) Fit a first-order auto-regressive model to each principal component.

Thus, the RSOS reconstructions take into account the temporal information as well. Both methods produce reconstructions that are linear combinations of a set of leading EOFs.

There are certain technical problems in the described procedure that were solved ad-hoc in [2], [3]. First, the estimation quality for different elements of the covariance matrix is varying, hence the estimated “covariance” matrix has some negative eigenvalues. Second, the information about the data quality is taken into account directly when computing principal components (step 4), but when EOFs are initially calculated, it is used only in a limited aggregated way (step 2). Therefore, EOFs are later readjusted to the estimated principal components.

There are significant problems in practical applications as well. Since reduced space estimation methods ([2], [6]) are most efficient for sparsely sampled ship and buoy data sets, they are usually not applied in the same form to the data sets containing much more abundant satellite data. They also produce relatively sparse ( $4^{\circ}$ - $5^{\circ}$  spatial resolution) gridded fields, leave permanent “holes” in the locations where poor data sampling makes the existing methods of covariance and EOF estimation unfeasible, and do not explicitly account for the long-term variability (trends) in the data. Because of users’ demand, both HadISST and ERSST products incorporate a few additional steps in their procedures on top of the reduced space estimation. These steps use various heuristic approaches to blend in satellite data, to increase grid resolution, to account for long-term variability, and to interpolate remaining gaps in poorly sampled areas.

The development of new reduced space estimation methods that would outperform the RSOI application [3] and possibly solve some of aforementioned problems in a methodologically coherent way will clearly be beneficial for the future versions of HadISST and possibly ERSST as well. In particular, in the form proposed in this paper, probabilistic latent variable models provide a practical way to overcome at least some of these difficulties by avoiding a problem with negative eigenvalues, by accounting for data uncertainty during covariance estimation, and by producing reconstructions for the complete data domain.

It is possible to bring all the available knowledge (the two basic assumptions) into a single probabilistic model. The identified model would then optimally explain the data and

incorporate all types of prior information. A well-defined optimization criterion allows to estimate different parts of the model (e.g., EOFs and the principal components) in the same procedure. Other advantages include natural handling of missing values, means to reduce overfitting, possibility to include in the analysis regions even with very little data, estimation of the uncertainty of the produced results (e.g. reconstructions), criteria for automatic determination of the right number of the principal components.

In this paper, we demonstrate how to compute reconstructions of global SST using a basic probabilistic PCA model and its extension that incorporates data quality information. The performance of the new methods needs to be evaluated by comparison with a well-established existing method, e.g. RSOI [3]. While one of the current state-of-the-art analyses, HadISST1 [5], uses RSOI in the reduced space estimation step, unfortunately, neither the data set of in situ observations that is used in their analysis nor the outcome of the RSOI step alone have been released as publicly available data sets. Therefore we chose to apply the new reconstruction methods to the older observational data set which was used in [3] for their RSOI reconstruction, so that a performance of new methods can be compared to that of RSOI on the same set of inputs.

The paper is organized as follows. Section II describes the data sets used for computing historical reconstructions and for validating their results. The proposed reconstruction method is described in Section III. The results of its trial application and performance comparison versus [3] are presented in Section IV. Finally, we discuss the results and introduce directions for future research in Section V.

## II. INPUT AND VALIDATION DATA

In this paper, we compare the skill of the proposed new methodology with that of the RSOI analysis [3]. As described below, reconstructions of global SST are computed here using exactly the same observational data set and data uncertainty specifications as were used in [3], so that all reconstructions are directly comparable.

### A. Historical Sea Surface Temperature Data

The original source of historical SST data is a collection of measurements made on ships and buoys. For use in climate research these irregularly spaced observations are usually averaged, with some quality control, in regular longitude-latitude bins. Some other intrabin statistics (number of observations, their standard deviation, quantiles of their distribution) is often calculated as well. Following [3], we use here the U.K. Meteorological Office historical SST data set (MOHSST5) [10] that contains monthly SST anomalies in the 1856-1991 period for  $5^{\circ} \times 5^{\circ}$  longitude-latitude bins. These anomalies are deviations from 1951-1980 climatological means provided by the U.K. Meteorological Office Global Ocean Surface Temperature Atlas (GOSTA) [11]. The producers of the MOHSST5 data set also applied to it the corrections for time-varying biases based on the scheme developed in [12] (“bucket corrections”). This correction

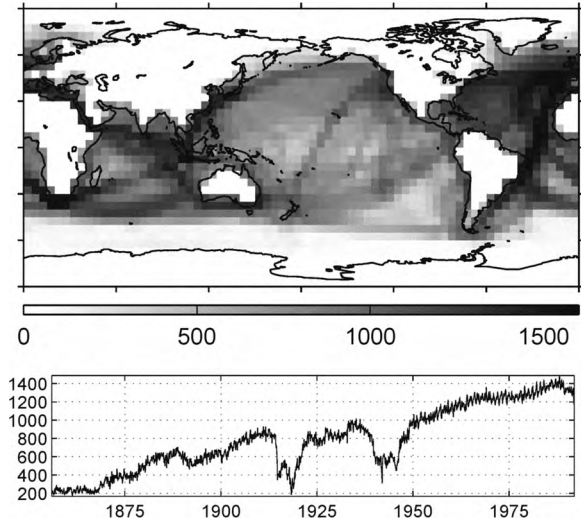


Fig. 1. The number of measurements for different spatial locations (above) and months (below) in the MOHSST5 data set.

method compensates historical SST values for spurious water temperature reductions that typically occur in a measuring bucket due to its heat exchange with the air during the time interval that the bucket stays on a ship deck before the water temperature measurement is taken. Corrections are computed as a function of year, calendar month, bucket type and ship speed for the period before 1941 (in the later period bucket measurements were mostly replaced by other techniques).

The number of available measurements in the MOHSST5 data set varies a great deal in space and time. While the data coverage is good during the last years, there are few observations before 1880 and during the two World Wars (see Fig. 1). There are naturally more measurements along the typical ship routes. In total there are 1280018 measurements in 1727 spatial locations.

### B. Data Uncertainty

The quality of the data also significantly varies in space and time. One source of effective observational error is a sampling error caused by a limited number of measurements available for estimating the SST mean over one month and a  $5^\circ \times 5^\circ$  area. In [3], the sampling error for an observation  $y_{it}$  in the  $i$ -th spatial location and month  $t$  is evaluated as

$$r_{it} = \sigma_i^2 / N_{it}, \quad (1)$$

where  $N_{it}$  is a number of individual measurements contained during month  $t$  in the  $5^\circ \times 5^\circ$  bin  $i$ , and  $\sigma_i^2$  is the intrabin variance. Values of the intrabin variances  $\sigma_i^2$  were estimated in [3] using the Comprehensive Ocean-Atmosphere Data Set (COADS), release 1 (1854-1979) [13] extended by standard release 1a (1980-1992) [14], because intrabin  $\sigma^2$  statistics was not available for MOHSST5. While newer versions of COADS data set, renamed ICOADS in 2001, include many additional observations [15], the older version is used here for a direct comparability with [3]. Also, as in [3], since

values  $N_{it}$  were not available for MOHSST5, they are taken from its closest upgrade, MOHSST6 [16].

### C. Validation Data Set

We validate reconstruction results by comparing them to the U.S. National Oceanic and Atmospheric Administration (NOAA) monthly  $1^\circ$  Optimal Interpolation (OI) data set, version 2 [17] for the period 1982–1991. This validation data set is a blend of satellite and in situ data and is considered to be the best available SST analysis for the 1982-1991 period (more recent NOAA Daily  $0.25^\circ$  OI product [18] does not cover the period before 1985). Since the original spatial resolution of validation fields is  $1^\circ$ , in order to perform our comparisons they were averaged onto the  $5^\circ$  grid that was used for reconstructions. The quality of reconstructions is assessed by computing root-mean-square differences with the validation data.

## III. RECONSTRUCTION METHOD

### A. Bayesian PCA Model

The starting point in our analysis is a simple probabilistic PCA model first discussed in [19]. The vector of observations in time  $t$  is assumed to be generated from a vector of latent variables  $\mathbf{x}_t$  using a simple transformation:

$$\mathbf{y}_t = \mathbf{W}\mathbf{x}_t + \boldsymbol{\mu} + \mathbf{n}_t,$$

where the dimension  $c$  of vector  $\mathbf{x}_t$  is smaller than the dimension  $d$  of the observed vector  $\mathbf{y}_t$ . For incomplete data, it is convenient to rewrite the model for each observed value:

$$y_{it} = \mathbf{w}_i^\top \mathbf{x}_t + \mu_i + n_{it},$$

where  $\mathbf{w}_i^\top$  denotes the  $i$ -th row of matrix  $\mathbf{W}$  and  $\mu_i, n_{it}$  are the elements of  $\boldsymbol{\mu}, \mathbf{n}_t$ , respectively.

The latent variables  $\mathbf{x}_t$  are assumed to be uncorrelated Gaussian variables with zero means and unit variances:

$$p(\mathbf{x}_t) = \mathcal{N}(\mathbf{0}, \mathbf{I}).$$

We use the notation  $\mathcal{N}(\mathbf{z}, \boldsymbol{\Sigma})$  for the Gaussian probability density function (pdf) with mean  $\mathbf{z}$  and covariance matrix  $\boldsymbol{\Sigma}$ . The elements  $n_{it}$  of the noise term are assumed to be independent and Gaussian:

$$p(\mathbf{n}_t) = \mathcal{N}(\mathbf{0}, \text{diag}(\mathbf{v}_t)),$$

where  $\mathbf{v}_t$  is a vector with elements  $\{v_{it}, \forall i\}$  and  $\text{diag}(\mathbf{v})$  denotes a diagonal matrix with the elements of  $\mathbf{v}$  on the main diagonal. Here, we assume individual variances  $v_{it}$  while the basic PCA model in [19] assumes equal noise variances, that is  $v_{it} = v$  for all  $i, t$ .

In order to reduce overfitting (which is possible when this simple model is applied to incomplete data [20]) and to prune out unnecessary components in  $\mathbf{x}_t$ , we introduce simple Gaussian priors for the model parameters:

$$p(\boldsymbol{\mu}) = \prod_{i=1}^d p(\mu_i), \quad p(\mu_i) = \mathcal{N}(0, \alpha_0),$$

$$p(\mathbf{W}) = \prod_{i=1}^d \prod_{k=1}^c p(w_{ik}), \quad p(w_{ik}) = \mathcal{N}(0, \alpha_k),$$

where  $w_{ik}$  denotes an element of  $\mathbf{W}$ . These priors were used for the basic PCA model in [21]. The formulas for learning this model for data sets with missing values are given in [20].

With separate noise variances  $v_{it}$  it is easy to take into account the information about the data quality. Let us assume that the observation error is a combination of two errors:  $n_{it} = n'_{it} + n''_{it}$ , where  $n'_{it}$  is the sampling error caused by a limited number of data to estimate exactly the mean over a month  $t$  in box  $i$  and  $n''_{it}$  is the noise term which models both measurement and modeling errors. By the modeling error we mean the error of reproducing the observations from a truncated number of principal components. We assume that both  $n'_{it}$  and  $n''_{it}$  are independent and Gaussian. We also assume that the variance  $r_{it}$  of  $n'_{it}$  can be estimated using (1) and the variance of  $n''_{it}$  is a constant  $v$ . In Bayesian terms, the variance  $r_{it}$  can be seen as the uncertainty of each measurement.

In the experiments, we consider two versions of the PCA model: a simple model with  $v_{it} = v$  and a model which takes into account the uncertainty information with  $v_{it} = r_{it} + v$ .

### B. Maximum A Posteriori Cost

The maximum a posteriori (MAP) estimation of the described model yields maximization of the posterior probability density of the unknown variables. This is equivalent to minimizing the cost function

$$C_{\text{MAP}} = \sum_{it \in O} \frac{1}{v_{it}} (y_{it} - \hat{y}_{it})^2 + \frac{1}{\alpha_0} \sum_{i=1}^d \mu_i^2 + \sum_{k=1}^c \frac{1}{\alpha_k} \sum_{i=1}^d w_{ik}^2 + \sum_{k=1}^c \sum_{t=1}^n x_{kt}^2 + C(v_{it}, \alpha_k), \quad (2)$$

where  $\hat{y}_{it} = \mathbf{w}_i^T \mathbf{x}_t + \mu_i$  is the reconstruction of  $y_{it}$  and  $C(v_{it}, \alpha_k)$  denote the terms which depend on the variance parameters (see the full expression for isotropic observation noise in [20]). The summation  $it \in O$  is done over the observed elements  $y_{it}$ .

The cost function (2) is simply a weighted sum-square reconstruction error plus the penalty terms corresponding to the priors. Thus, the uncertainty information is taken into account naturally by weighting correspondingly the contribution of each measurement into the optimized cost.

There are several problems with the MAP estimation approach. First, the MAP cost function infinitely grows when  $\alpha_k \rightarrow 0$  and therefore  $\alpha_k$  should be restricted in practice. Second, the MAP solutions are known to suffer often from the overfitting problem. Therefore, we use a more sophisticated procedure.

### C. Variational Bayesian Learning

Variational Bayesian (VB) learning [22] is known to be more resistant against the overfitting problem compared to the MAP estimation. Instead of using point estimates, VB is based on fitting an approximate pdf to the true posterior pdf of the unknown variables. Thus, the VB estimation is more sensitive to the probability mass than to the probability

density of the posterior, and therefore it often provides more robust solutions.

VB learning is usually performed by minimizing the Kullback-Leibler divergence between the posterior and the approximate pdf of a predefined tractable form. The VB cost function is

$$C_{\text{VB}} = \int q(\Theta) \log \frac{q(\Theta)}{p(\mathbf{Y}, \Theta | \xi)} d\Theta,$$

where  $\mathbf{Y}$  denotes all observed data,  $\Theta$  contains all the unknown variables and  $\xi$  are the model hyperparameters (parameters defining the priors) which are either fixed or point-estimated. Because of the tractability issues, the approximate posterior pdf is usually factorized:

$$q(\Theta) = \prod_j q(\theta_j).$$

In models with conjugate priors, one can easily find individual factors  $q(\theta_j)$  which minimize the VB cost function when the rest of the factors fixed. The update rules for the hyperparameters can also be derived for fixed  $q(\Theta)$ . Thus, learning iterates between alternate updates of individual  $q(\theta_j)$  and  $\xi$  until convergence. See more details on VB learning, for example, in [22].

### D. The Update Rules

In the Bayesian PCA model, we use point estimates for  $v$  and  $\alpha_k$  and therefore  $\Theta = \{\boldsymbol{\mu}, \mathbf{W}, \mathbf{x}_t | t\}$ . It is convenient to use the following factorization:

$$q(\boldsymbol{\mu}, \mathbf{W}, \mathbf{X}) = q(\boldsymbol{\mu}) \prod_{i=1}^d q(\mathbf{w}_i) \prod_{t=1}^n q(\mathbf{x}_t).$$

Then the marginal posterior approximations are Gaussian whose parameters can be updated as follows. For the hidden variables,  $q(\mathbf{x}_t) = \mathcal{N}(\bar{\mathbf{x}}_t, \boldsymbol{\Sigma}_{\mathbf{x}_t})$  with

$$\boldsymbol{\Sigma}_{\mathbf{x}_t} = \left[ \mathbf{I} + \sum_{i \in O_t} \frac{1}{v_{it}} \langle \mathbf{w}_i \mathbf{w}_i^T \rangle \right]^{-1}$$

$$\bar{\mathbf{x}}_t = \boldsymbol{\Sigma}_{\mathbf{x}_t} \sum_{i \in O_t} \frac{1}{v_{it}} (y_{it} - \bar{\mu}_i) \bar{\mathbf{w}}_i, \quad t = 1, \dots, n,$$

where the summation  $i \in O_t$  is done over all spatial locations for which  $y_{it}$  is observed. For the bias term,  $q(\boldsymbol{\mu}) = \prod_{i=1}^d \mathcal{N}(\bar{\mu}_i, \tilde{\mu}_i)$  with

$$\tilde{\mu}_i = \left[ \frac{1}{\alpha_0} + \sum_{t \in O_i} \frac{1}{v_{it}} \right]^{-1}$$

$$\bar{\mu}_i = \tilde{\mu}_i \sum_{t \in O_i} \frac{1}{v_{it}} (y_{it} - \bar{\mathbf{w}}_i^T \bar{\mathbf{x}}_t), \quad i = 1, \dots, d,$$

where the summation  $t \in O_i$  is done over all months for which  $y_{it}$  is observed. For matrix  $\mathbf{W}$ ,  $q(\mathbf{w}_i) = \mathcal{N}(\bar{\mathbf{w}}_i, \boldsymbol{\Sigma}_{\mathbf{w}_i})$

with

$$\Sigma_{\mathbf{w}_i} = \left[ \text{diag}(\alpha_{1..c})^{-1} + \sum_{t \in O_i} \frac{1}{v_{it}} \langle \mathbf{x}_t \mathbf{x}_t^T \rangle \right]^{-1}$$

$$\bar{\mathbf{w}}_i^T = \Sigma_{\mathbf{w}_i} \sum_{t \in O_i} \frac{1}{v_{it}} (y_{it} - \bar{\mu}_i) \bar{\mathbf{x}}_t, \quad i = 1, \dots, d.$$

For the model with  $v_{it} = r_{it} + v$ , the variance parameter  $v$  can be updated by minimizing the following cost function by line search methods:

$$C_{\text{VB}}(v) = \sum_{it \in O} \left[ \frac{1}{v_{it}} \langle (y_{it} - \mu_i - \mathbf{w}_i^T \mathbf{x}_t)^2 \rangle + \log v_{it} \right].$$

When  $v_{it} = v$ , the optimal  $v$  can be computed exactly:

$$v = \frac{1}{|O|} \sum_{it \in O} \langle (y_{it} - \mu_i - \mathbf{w}_i^T \mathbf{x}_t)^2 \rangle$$

$$= \frac{1}{|O|} \sum_{it \in O} [(y_{it} - \bar{\mu}_i)^T (y_{it} - \bar{\mu}_i - 2\bar{\mathbf{w}}_i^T \bar{\mathbf{x}}_t) + \Sigma_{\mu}$$

$$+ \text{tr}(\langle \mathbf{w}_i^T \mathbf{w}_i \rangle \langle \mathbf{x}_t \mathbf{x}_t^T \rangle)],$$

where  $|O|$  denotes the number of observed data.

#### IV. TRIAL APPLICATION AND PERFORMANCE ASSESSMENT

We estimated 80 principal components for the MOHSST5 data set using two versions of the PCA model: with isotropic noise  $v_{it} = v$  (we call this model VBPCA) and with individual variances  $v_{it} = r_{it} + v$  in which  $r_{it}$  were computed by (1) (we call this model VBPCA-OE). The number of principal components was chosen to be the same as in [3].

The data were preprocessed prior to PCA to account for uneven density of measurement locations on the globe: each data point was multiplied by the square root of the cosine of the latitude of the corresponding location.

The reconstructions were computed as

$$\hat{y}_{it} = \bar{\mu}_i + \bar{\mathbf{w}}_i^T \bar{\mathbf{x}}_t,$$

where the elements of  $\mathbf{w}_i$  were properly scaled to remove the effect of the data preprocessing.

The reconstruction quality was assessed by the weighted root-mean-square (rms) errors w.r.t. the validation data  $\{z_{it} | it \in V\}$ , where  $z_{it}$  is the validation SST measurement in the  $i$ -th location for month  $t$ . The overall rms error was computed as

$$\text{rms} = \left( \frac{1}{|V|} \sum_{it \in V} \omega_i^2 (z_{it} - \hat{y}_{it})^2 \right)^{\frac{1}{2}}, \quad (3)$$

where  $|V|$  denotes the number of the validation data points and  $\omega_i$  are the corresponding weights (the square root of the cosine of the latitude). The rms errors for spatial location  $i$  were computed as

$$\text{rms}_i = \left( \frac{1}{|V_i|} \sum_{t \in V_i} (z_{it} - \hat{y}_{it})^2 \right)^{\frac{1}{2}} \quad (4)$$

TABLE I  
WEIGHTED RMS ERRORS OF RECONSTRUCTIONS

	Area from [3]	Larger area
RSOI	0.4590	
VBPCA	0.4549	0.5379
VBPCA-OE	0.4510	0.5339

and similarly for month  $t$  as

$$\text{rms}_t = \left( \frac{1}{|V_t|} \sum_{i \in V_t} \omega_i^2 (z_{it} - \hat{y}_{it})^2 \right)^{\frac{1}{2}}, \quad (5)$$

where  $V_i$  is a set of months  $t$  for which  $z_{it}$  exists in the validation data set and  $V_t$  is defined likewise.

Table I presents the overall rms error (3) obtained with VBPCA and VBPCA-OE in comparison with the RSOI method used in [3]. The first column shows the errors averaged over the locations where the RSOI reconstructions are available. The second column contains rms errors for a larger area considered in the present study. Fig. 2 displays the differences of the rms errors computed for different spatial locations and months using (4) and (5), respectively, for VBPCA and VBPCA-OE in comparison with RSOI. Similar comparison results for VBPCA and VBPCA-OE are shown in Fig. 3. Note that the area covered by VBPCA and VBPCA-OE (Fig. 3) is larger than the area covered by RSOI (Fig. 2). The results indicate that VBPCA outperforms RSOI, while VBPCA-OE is the best among the three methods.

#### V. CONCLUSIONS AND FUTURE WORK

In this work, we computed global reconstructions of historical SST using Bayesian PCA models learned by variational methods. The reconstructions produced in this study outperformed the RSOI method used in [3] by 1.7% on the weighted rms error. In addition, the new reconstructions cover a larger area than in the previous studies. The good performance of the proposed techniques can be explained by their principal way of using all available measurements and their resistance to the overfitting problem.

The results are promising because the presented approach did not make use of the temporal and spatial (geographical) structure of the data. This knowledge can be incorporated by introducing more sophisticated priors for the model parameters. The reconstruction may also be improved by increasing the number of estimated principal components because a larger area is covered in the present analysis compared to [3].

#### REFERENCES

- [1] T. M. Smith, R. W. Reynolds, R. E. Livezey, and D. C. Stokes, "Reconstruction of historical sea surface temperatures using empirical orthogonal functions," *Journal of Climate*, vol. 9, pp. 1403–1420, 1994.
- [2] A. Kaplan, Y. Kushnir, M. Cane, and M. Blumenthal, "Reduced space optimal analysis for historical datasets: 136 years of Atlantic sea surface temperatures," *Journal of Geophysical Research*, vol. 102, pp. 27 835–27 860, 1997.

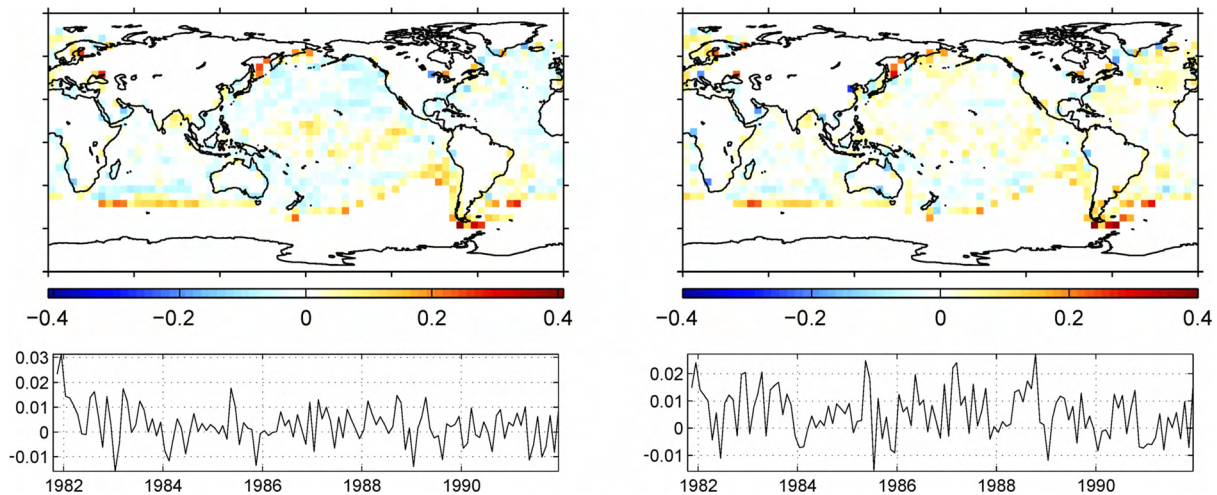


Fig. 2. Left: Difference between rms errors obtained with RSOI and VBPCA. Right: Difference between rms errors obtained with RSOI and VBPCA-OE. The upper plots display the differences of index (4) (in degrees Celsius) and the lower plots show the same for index (5). Positive values indicate superior performance of either VBPCA (l.h.s. plots) or VBPCA-OE (r.h.s. plots) while negative values indicate superior performance of RSOI.

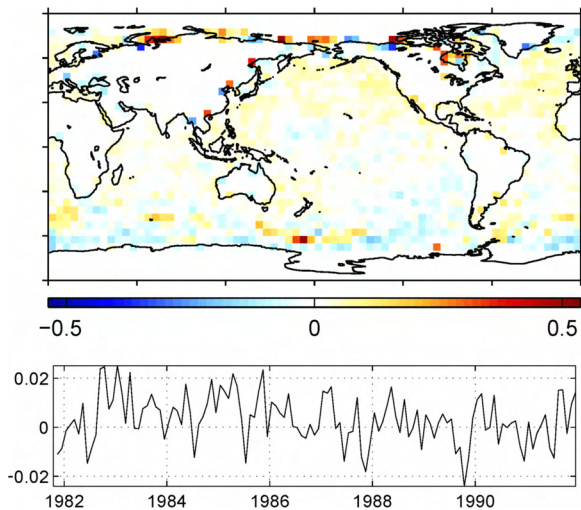


Fig. 3. Difference between rms errors obtained with VBPCA and VBPCA-OE. The upper plot displays the difference of index (4) (in degrees Celsius) and the lower plot shows the same for index (5). Positive values indicate superior performance of VBPCA-OE while negative values indicate superior performance of VBPCA.

[3] A. Kaplan, M. Cane, Y. Kushnir, A. Clement, M. Blumenthal, and B. Rajagopalan, "Analysis of global sea surface temperatures 1856–1991," *Journal of Geophysical Research*, vol. 103, pp. 18 567–18 589, 1998.

[4] A. Kaplan, M. A. Cane, and Y. Kushnir, *Reduced space approach to the optimal analysis interpolation of historical marine observations: Accomplishments, difficulties, and prospects*. World Meteorological Organization, 2003, pp. 199–216.

[5] N. A. Rayner, D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, "Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century," *Journal of Geophysical Research-Atmospheres*, vol. 108, 2003.

[6] T. M. Smith and R. W. Reynolds, "Extended reconstruction of global sea surface temperatures based on COADS data (1854–1997)," *Journal of Climate*, vol. 16, no. 10, pp. 1495–1510, 2003.

[7] —, "Improved extended reconstruction of SST (1854–1997)," *Journal of Climate*, vol. 17, no. 12, pp. 2466–2477, 2004.

[8] T. M. Smith, R. W. Reynolds, T. C. Peterson, and J. Lawrimore, "Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880–2006)," *Journal of Climate*, vol. 21, pp. 2283–2296, 2008.

[9] N. A. Rayner, T. Brandon, K. Casey, C. Donlon, E. Harrison, M. Ishii, A. Kaplan, E. Kent, R. W. Reynolds, and T. M. Smith, "Evaluating climate variability and change from modern and historical SST observations," in *Proceedings of OceanObs'09*, Venice, Italy, 2009, in preparation.

[10] D. E. Parker, P. D. Jones, C. K. Folland, and A. Bevan, "Interdecadal changes of surface temperature since the late nineteenth century," *Journal of Geophysical Research*, vol. 99, pp. 14 373–14 399, 1994.

[11] M. Bottomley, C. K. Folland, J. Hsiung, R. E. Newell, and D. E. Parker, *Global Ocean Surface Temperature Atlas*. Norwich, England: Her Majesty's Stn. Off., 1990.

[12] C. K. Folland and D. E. Parker, "Correction of instrumental biases in historical sea surface temperature data," *Quarterly Journal of Royal Meteorological Society*, vol. 121, pp. 319–367, 1995.

[13] S. D. Woodruff, R. J. Slutz, R. L. Jenne, and P. M. Steurer, "A comprehensive ocean-atmosphere data set," *Bulletin of the American Meteorological Society*, vol. 68, pp. 521–527, 1987.

[14] S. D. Woodruff, S. J. L. K. Wolter, S. J. Worley, and J. D. Elms, "Comprehensive ocean-atmosphere data set (COADS) release 1a: 1980–92," *Earth System Monitor*, vol. 4, no. 1, pp. 1–8, 1993.

[15] S. J. Worley, S. D. Woodruff, R. W. Reynolds, S. J. Lubker, and N. Lott, "ICOADS release 2.1 data and products," *International Journal of Climatology*, vol. 25, pp. 823–842, 2005.

[16] D. E. Parker, C. K. Folland, and M. Jackson, "Marine surface temperature: Observed variations and data requirements," *Climatic Change*, vol. 31, pp. 559–600, 1995.

[17] R. W. Reynolds, N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang, "An improved in situ and satellite SST analysis," *Journal of Climate*, vol. 15, pp. 1609–1625, 2002.

[18] R. W. Reynolds, T. M. Smith, C. Liu, D. B. Chelton, K. S. Casey, and M. G. Schlax, "Daily high-resolution-blended analyses for sea surface temperature," *Journal of Climate*, vol. 20, pp. 5473–5496, 2007.

[19] M. Tipping and C. M. Bishop, "Probabilistic principal component analysis," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 61, no. 3, pp. 611–622, 1999.

[20] A. Ilin and T. Raiko, "Practical approaches to principal component analysis in the presence of missing values," Helsinki University of Technology, Espoo, Finland, Tech. Rep. TTK-ICS-R6, 2008, available at <http://www.cis.hut.fi/alexilin/>.

[21] C. M. Bishop, "Variational principal components," in *Proceedings of the 9th International Conference on Artificial Neural Networks (ICANN99)*, 1999, pp. 509–514.

[22] C. Bishop, *Pattern Recognition and Machine Learning*. Cambridge: Springer, 2006.