



RESEARCH ARTICLE

10.1029/2019EA000954

Key Points:

- A data-driven approach for estimating sea surface temperature in the Massachusetts and Cape Cod Bays is presented
- A new assimilation technique based on autoregressive Gaussian processes to combine satellite and in situ measurements is presented
- The multifidelity model produces continuous sea surface temperature and uncertainty maps for the Massachusetts and Cape Cod Bays

Correspondence to:

H. Babae,
h.babae@pitt.edu

Citation:

Babae, H., Bastidas, C., DeFilippo, M., Chrysostomidis, C., & Karniadakis, G. E. (2020). A multi-fidelity framework and uncertainty quantification for sea surface temperature in the Massachusetts and Cape Cod Bays. *Earth and Space Science*, 7, e2019EA000954. <https://doi.org/10.1029/2019EA000954>

Received 7 NOV 2019

Accepted 30 DEC 2019

Accepted article online 3 JAN 2020

©2020. The Authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

A Multifidelity Framework and Uncertainty Quantification for Sea Surface Temperature in the Massachusetts and Cape Cod Bays

H. Babae^{1,2}, C. Bastidas², M. DeFilippo², C. Chrysostomidis², and G. E. Karniadakis³

¹Department of Mechanical Engineering and Materials Science, University of Pittsburgh, Pittsburgh, PA, USA, ²MIT Sea Grant, Massachusetts Institute of Technology, Cambridge, MA, USA, ³Division of Applied Mathematics, Brown University, Providence, RI, USA

Abstract We present a multifidelity framework to analyze and hindcast predictions of sea surface temperature (SST) in the Massachusetts and Cape Cod Bays, which is a critical area for its ecological significance, sustaining fisheries and the blue economy of the region. Currently, there is a lack of accurate and continuous SST prediction for this region due to the high cost of collecting the samples (e.g., cost of buoys, maintenance, severe weather). In this work, we use SST data from satellite images and in situ measurements collected by the Massachusetts Water Resources Authority to develop multifidelity forecasting models. This multifidelity framework is based on autoregressive Gaussian process schemes that systematically exploit all correlations between data from multiple heterogeneous spatiotemporal sources with various degrees of fidelity. This enables us to obtain implicitly their functional relationships and, at the same time, quantify the uncertainty of the data-driven predictions. Specifically, in the current work, we develop and validate progressively more complex models, including temporal, spatial, and spatiotemporal multifidelity hindcast predictions of SST in the Massachusetts and Cape Cod Bays. Together with these predictions, we present for the first time uncertainty maps for the region.

1. Introduction

In the Gulf of Maine, Massachusetts and Cape Cod Bays encompass an area of highly productive fisheries, aquaculture facilities, and ecological resources that provide services to coastal communities. In this area, the Stellwagen Bank National Marine Sanctuary is of particular relevance by supporting strong local economies through commercial and recreational fisheries and providing more than 80% of the whale watching tourism in New England (<https://sanctuaries.noaa.gov/science/socioeconomic/factsheets/stellwagenbank.html>). The future of these benefits is uncertain as sea surface temperatures in the Gulf of Maine are increasing at one of the highest rates in the world's oceans, with clear consequences in reducing Atlantic cod recruitment that contributed to overfished stocks (Pershing et al., 2015). Similarly, in New England, the lobster industry has migrated almost entirely into Maine, being no longer supported in southern coastal states as in the past (Steneck et al., 2013), while about 40 other fish and invertebrate species are considered as highly vulnerable to climate change and its variability (Hare et al., 2016). In the summer of 2018, sea surface temperature (SST) anomalies were 1 to 3 °C above the 30-year average in the Gulf of Maine. These anomalies and ocean heat waves are becoming common in the Gulf of Maine. On 8 August 2018, the SSTs measured were the second warmest ever observed there (https://www.drought.gov/drought/sites/drought.gov.drought/files/media/reports/regional_outlooks/GOM%20Summer%202018.pdf). In view of these changes and the likely scenario they would increase in the near future, more accurate and timely predictions of sea temperature at regional and local scales might better help resource management and economic planning.

The predictability of environmental drivers is a key step in the process of linking climate and biological processes for better management of marine resources. Advances in climate prediction facilitate this management, with a particular focus on dynamical seasonal to decadal prediction systems derived from Global Climate Models (Tommasi et al., 2017). Considerable progress has been made with this modeling approach, but some limitations exist in coastal predictions at regional scales, particularly in the U.S. east coast (Stock

et al., 2015). In addition, in situ data are relatively scarce in this region from buoys, but the Massachusetts Water Resources Authority (MWRA) sustains a comprehensive monitoring program for a wide range of bay conditions (Werme et al., 2017). The MWRA measurements provide an accurate but coarse space-time measurements of various quantities of interest in the Massachusetts and Cape Cod Bays, including seawater temperature. Finally, satellite measurements have better temporal resolution but they are gappy due to cloud coverage and require calibration. Therefore, by using less conventional approaches as the one presented in this study, we aim to advance the predictability of SST in the region by using all available sources of data of variable fidelity.

Various data assimilation methods have been used to combine satellite data and in situ measurements. Among these methods, optimal interpolation (OI) is the most widely used data assimilation technique for ocean and meteorological application (see e.g., Daley, 1992). Most weather centers used OI for weather forecasts during the 1970s and 1980s. OI is also widely used for combining satellite data with in situ measurements (see e.g., ocean group for high resolution SST products; Dash et al., 2012). OI combines the background measurements \mathbf{u}^b and observations \mathbf{u}^o via the analysis equation given by:

$$\mathbf{u}^a = \mathbf{u}^b + \mathbf{P}\mathbf{H}^T(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{u}^o - \mathbf{H}\mathbf{u}^b), \quad (1)$$

where \mathbf{H} is the observation operator that interpolates from background field to observation space, and \mathbf{P} and \mathbf{R} are background and observation covariances. The interpolator \mathbf{H} is often chosen a priori, and the effect of its misspecification must be accounted for (Liu & Rabier, 2006). For sparse background measurements and observations available at different scales, determining \mathbf{P} , \mathbf{R} , and \mathbf{H} is nontrivial, and this has been the subject of intense research (see e.g., Oke & Sakov, 2008 and references therein). In addition to OI, variational methods are also used for the purpose of data assimilation. In these methods, a cost function is defined that seeks to find a smooth field while penalizing the distance between the field and the observations. In Li et al. (2015), the 4D variational data assimilation algorithm was used to combine simulation predictions with in situ measurements in Gulf of Maine with very good results. Empirical orthogonal functions have also been used for field reconstructions from noisy measurements (Yang et al., 2010) and SST data assimilation (Smith et al., 1996). In Smith et al. (1996), the empirical orthogonal function modes were computed from OI-analyzed SST fields, and in situ measurements were used to determine the time dependence of each mode. More recently, other techniques have been proposed to blend satellite data with in situ measurements and to fill in the gaps in satellite-derived SST products, including hierarchical Bayesian algorithms (Zhu et al., 2017, 2019), Bayesian maximum entropy (Li et al., 2013), and multiresolution wavelet-based algorithms (Chin et al., 2017).

Gaussian processes (GP) are nonparametric Bayesian machine learning techniques for learning functions from data (Rasmussen & Williams, 2006). GP provides a flexible prior distribution over functions, enjoys analytical tractability, and has a probabilistic workflow that yields robust posterior estimates and handles uncertainty in a principled manner. Before conditioning on the training data, a GP model is completely specified via its mean function and covariance function, which is called *kernel* and determines how the GP model extrapolates and generalizes to new data. There are many choices for kernel. For example, linear regression, splines, Kalman filters, and OI are all examples of GPs with particular kernels (Reece & Roberts, 2010). The kernel of a GP is often parameterized and the hyper-parameters are learned from the observed data, for example, using the maximum likelihood. This allows for further fine-tuning the kernel to the training data. The training also determines hyper-parameters associated with the uncertainty model. Training the kernel and the uncertainty model replaces the *guessing* work needed in OI, in which various matrices and noise models must be determined a priori.

The novelty of our approach is based on the use of a *multifidelity* framework for obtaining SST, exploiting correlations in space-time for sea surface, and estimating both mean values as well as their uncertainty as *surface maps* in the Massachusetts and Cape Code Bays by combining in situ measurements with satellite data. This framework allows us to replace the usual line error bars by space-time error uncertainty maps. Our ultimate goal is to develop a regional model that could even monitor acidification, similar to the three regional aragonite models for the U.S. West Coast (Alin et al., 2018; Davis et al., 2018; Juranek et al., 2009), but continuously updated and endowed with uncertainty quantification and forecasting capabilities that will drive future measurements.

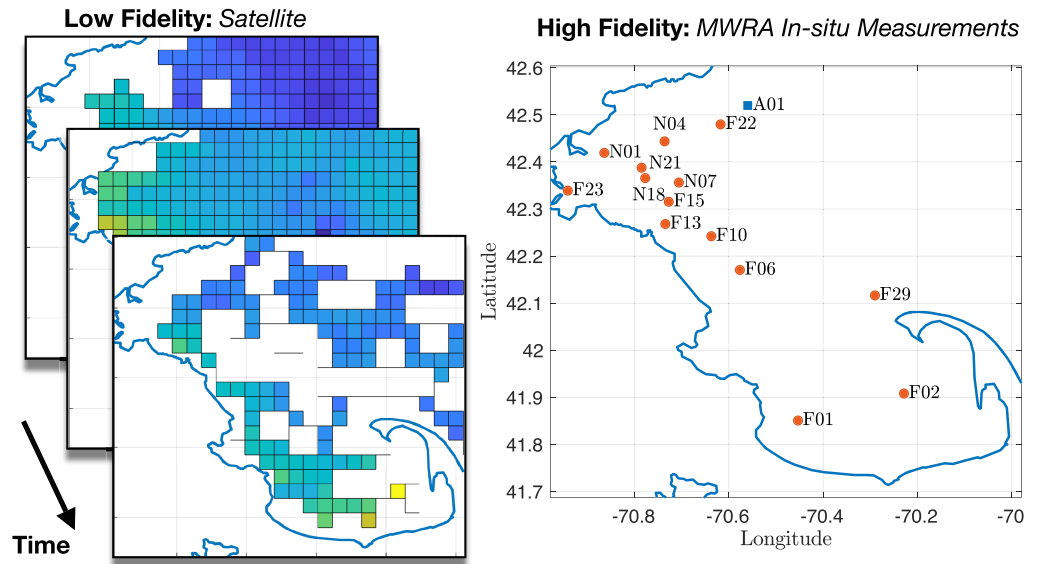


Figure 1. Multifidelity data sources: Left: The MODerate-resolution Imaging Spectroradiometer (MODIS) Terra on board NASA satellite with the spatial resolution of 0.041 degrees (latitude) \times 0.041 degrees (longitude), which is equivalent to the resolution of 4 \times 4 km (Werdell et al., 2013). Right: Massachusetts Water Resource Authority (MWRA) in situ measurements shown in red. The NERACOOS buoy A01 is shown in blue, and it is used only for validation purposes.

2. Data Sources

For the estimation of SST in Massachusetts and Cape Cod Bays, we consider two sources of data: (1) High-Fidelity: The MWRA measurements. The MWRA routinely measures SST and various other quantities at fourteen stations in Massachusetts and Cape Cod Bays (Libby et al., 2017). These stations are shown in Figure 1 (right). The MWRA measurements are taken at various depths and on a monthly basis during nine surveys per year with the exception of winter months. In this manuscript, we use MWRA measurements taken at 1 m below the surface as an estimate for seawater surface temperature. (2) Low Fidelity: The MODerate-resolution Imaging Spectroradiometer (MODIS) Terra on board NASA satellite with the spatial resolution of 4 \times 4 km (Werdell et al., 2013). The satellite images are processed at level 3 with daily and

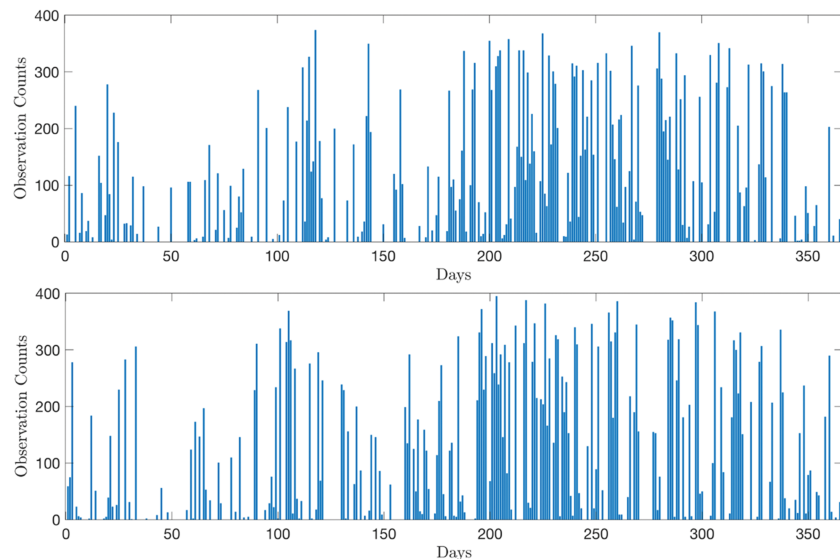


Figure 2. MODerate-resolution Imaging Spectroradiometer satellite observation counts for year of 2015 (top) and 2016 (bottom). Day 1 in each plot corresponds to 1 January of the respective year.

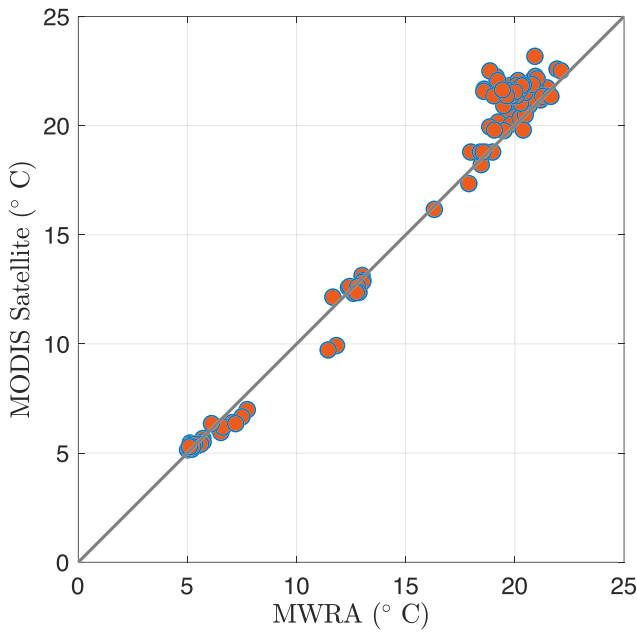


Figure 3. Correlation between temperature derived from MODerate-resolution Imaging Spectroradiometer satellite and in situ measurements at 14 Massachusetts Water Resource Authority stations during the years of 2015 and 2016.

nightly temporal resolutions and are provided through the NASA Earth Science Data System (ESDS) project. It can be accessed via Physical Oceanography Distributed Active Archive Center (PO.DAAC). Throughout this manuscript, we will use the MODIS Terra, thermal-IR SST level 3, 4 km, daily. We consider the satellite data in the region that spans the longitude of $69^{\circ} 58' W$ to $70^{\circ} 58' W$ and the latitude of $41^{\circ} 41' N$ to $42^{\circ} 36' N$. The satellite images are gappy due to cloud coverage as shown in Figure 1 (left). The observation counts of MODIS satellite for years 2015 and 2016 are shown in Figure 2. It is clear that during summer, there are significantly more satellite data available compared to late fall, entire winter, and early spring. Although the MODIS satellite data used in this study are level 3 SST products, that is, they are blended with in situ measurements when compared against MWRA measurements, the MODIS satellite measurements show up to $3^{\circ} C$ discrepancy. We expected that some of this disagreement resulted from comparing surface with subsurface measurements (MWRA data from 1 m deep), and some of it may come from the coarse satellite resolution 4×4 km. However, the correlation plot between the MWRA measurements and MODIS satellite measurements in the year 2015 and 2016 indicated that this bias is not consistent along the temperature interval (Figure 3). Therefore, we use the MODIS satellite measurements as a low-fidelity data source. Overall, the satellite data provide regional-scale patterns while the MWRA measurements are accurate but scarce in situ measurements.

3. Multifidelity Modeling

In this section, we present the details of the multifidelity modeling based on GP regression (GPR; Rasmussen & Williams, 2006), which is a non-parametric Bayesian machine learning technique that enjoys analytic tractability and provides a fully probabilistic framework for approximating functions and treating approximation uncertainties in a principled manner.

3.1. Multifidelity Scheme

For the sake of brevity we present the linear formulation for two fidelity levels (Babae et al., 2016; Kennedy & O'Hagan, 2000; Prempraneerach et al., 2017); however, the formulation can be extended to $n > 2$ levels of fidelity in a straightforward manner (see e.g., Forrester et al., 2007) and to nonlinear autoregressive schemes using deep GP (see Perdikaris et al., 2017). To this end, we consider the low-fidelity (L) and high-fidelity (H) observations to be of the form

$$y_L(x) = u_L(x) + \epsilon_L \quad \text{and} \quad y_H(x) = \rho u_L(x) + \delta(x) + \epsilon_H, \quad (2)$$

where $x = \{x_1, x_2, \dots, x_D\}$ represents the D -dimensional input space. For temporal SST multifidelity models $D = 1$, where x represents time, for spatial models $D = 2$ with x representing the pair of longitude and latitude, and finally, for spatiotemporal models $D = 3$ with x representing the triplet of longitude, latitude and time. In equation (2), $u_L(x)$ and $\delta(x)$ are assumed to be two independent random fields, each represented by a GP in the form of

$$u_L(x) \sim \mathcal{GP}(0, k(x, x'; \theta_L)) \quad \text{and} \quad \delta(x) \sim \mathcal{GP}(0, k(x, x'; \theta_\delta)),$$

and ϵ_L and ϵ_H represent the unbiased errors in the low-fidelity and high-fidelity observations, respectively. These errors may be caused by unresolved scales due to undersampling or measurement noise corruption, and they are modeled as a centered Gaussian noise $\epsilon_L \sim \mathcal{N}(0, \sigma_{n_L}^2)$ and $\epsilon_H \sim \mathcal{N}(0, \sigma_{n_H}^2)$. The primary source of noise in our high-fidelity/low-fidelity data, however, is the effect of unresolved scales, because of coarse spatial and temporal resolution of the training data. These subgrid scales are commonly modeled via isotropic Gaussian noise in ocean/atmospheric problems (see e.g., Daley, 1992, Chapter 4). This does not mean, however, that these errors are completely homogeneous. As reasoned in Daley (1992), the inhomogeneities in the noise are often on a very large scale, while the noise correlation lengths are much

smaller. Therefore, when the scale of the inhomogeneity is much larger than the scale of the correlation, the assumption of homogeneous Gaussian for the noise is not a major source of the modeling error.

The multifidelity scheme given by equation (2) tolerates and learns the systematic bias in the low-fidelity measurements. This bias is modeled by the scaling factor $1/\rho$ and the bias correction of $-\delta(x, t)/\rho$. In cases where the high-fidelity data have systematic bias, for example, due to sensor malfunction, the uncertainty has to be modeled explicitly. We use the square-exponential kernel in most cases presented in this paper (for hindcast predictions) as specified by

$$k(x, x') = \sigma^2 \exp\left(-\sum_{i=1}^D (x_i - x'_i)^2 / \theta_i^2\right), \quad (3)$$

where σ and θ_i are hyper-parameters. Therefore, the kernel at each fidelity level contains $D + 1$ hyper-parameters, which in addition to the hyper-parameters introduced by σ_{n_L} , σ_{n_H} , and ρ results in $2D + 5$ hyper-parameters in total.

3.1.1. Training

The hyper-parameters are learned from the low-fidelity and high-fidelity observations. The low-fidelity observations are denoted by $\{\mathbf{x}_L, \mathbf{y}_L\}$, where \mathbf{x}_L is a $n_L \times D$ input points and \mathbf{y}_L is a $n_L \times 1$, representing SST satellite measurements, and n_L is the number of low-fidelity measurements. Similarly, the high-fidelity observations are denoted by $\{\mathbf{x}_H, \mathbf{y}_H\}$, where \mathbf{x}_H with the size of $n_H \times D$ and \mathbf{y}_H with the size of $n_H \times 1$ are high-fidelity input points and (MWRA) measurements, respectively, and n_H is the number of high-fidelity measurements.

Since $u_L(x)$ and $\delta(x)$ are independent GP, using the multifidelity scheme given by equation (2), it is straightforward to observe that the low-fidelity and high-fidelity observations have a joint Gaussian distribution given by $\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{K})$, where

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_L \\ \mathbf{y}_H \end{bmatrix} \quad \text{and} \quad \mathbf{K} = \begin{bmatrix} k_{LL}(\mathbf{x}_L, \mathbf{x}_L) + \sigma_{n_L}^2 \mathbf{I} & k_{LH}(\mathbf{x}_L, \mathbf{x}_H) \\ k_{HL}(\mathbf{x}_H, \mathbf{x}_L) & k_{HH}(\mathbf{x}_H, \mathbf{x}_H) + \sigma_{n_H}^2 \mathbf{I} \end{bmatrix},$$

where \mathbf{y} is an $n \times 1$ vector that contains the low-fidelity and high-fidelity observations at points $\mathbf{x} = [\mathbf{x}_L, \mathbf{x}_H]$ with the size of $n \times D$, n is the number of high-fidelity and low-fidelity data points, that is, $n = n_H + n_L$, and

$$\begin{aligned} k_{LL}(\mathbf{x}_L, \mathbf{x}_L) &= k(\mathbf{x}_L, \mathbf{x}_L; \theta_L), \\ k_{HL}(\mathbf{x}_H, \mathbf{x}_L) &= \rho k(\mathbf{x}_H, \mathbf{x}_L; \theta_L), \\ k_{HH}(\mathbf{x}_H, \mathbf{x}_H) &= \rho^2 k(\mathbf{x}_H, \mathbf{x}_H; \theta_L, \theta_\delta, \rho) + k(\mathbf{x}_H, \mathbf{x}_H; \theta_\delta), \end{aligned}$$

where $k_{LL}(\mathbf{x}_L, \mathbf{x}_L)$ is an $n_L \times n_L$ matrix that represents correlations between the low-fidelity points, $k_{LH}(\mathbf{x}_L, \mathbf{x}_H)$ is an $n_L \times n_H$ matrix, representing the correlations between the low-fidelity points and high-fidelity points, $k_{HH}(\mathbf{x}_H, \mathbf{x}_H)$ is an $n_H \times n_H$ matrix that represents the correlations between high-fidelity points, and $k_{HL}(\mathbf{x}_H, \mathbf{x}_L) = k_{LH}^T(\mathbf{x}_L, \mathbf{x}_H)$. The negative log-likelihood $L(\mathbf{y} | \theta_L, \theta_\delta, \rho, \sigma_{n_L}, \sigma_{n_H})$ is expressed by

$$-\log L(\mathbf{y} | \theta_L, \theta_\delta, \rho, \sigma_{n_L}, \sigma_{n_H}) = \frac{1}{2} \mathbf{y}^T \mathbf{K}^{-1} \mathbf{y} + \frac{1}{2} \log |\mathbf{K}| + \frac{n_L + n_H}{2} \log(2\pi).$$

The hyper-parameters $(\theta_L, \theta_\delta, \rho, \sigma_{n_L}, \sigma_{n_H})$ are obtained by minimizing the above function (i.e., maximizing the likelihood). The minimization problem is solved using the trust-region method. The kernel matrix \mathbf{K} has a size of $n \times n$ and is a full matrix; therefore, it has a memory requirement of $\mathcal{O}(n^2)$ and its inversion has a computational complexity of $\mathcal{O}(n^3)$. Although we used this approach, for large data sets, this is prohibitive. To address this issue, a recursive autoregressive scheme has been suggested (Le Gratiet & Le Gratiet, 2014), in which the data sources of the two levels of fidelity can be decoupled and equivalently reformulated in a recursive two independent Gaussian regression problems. This requires that the high-fidelity and low-fidelity data sets are nested.

3.1.2. Prediction

Once the hyper-parameters are found, the probabilistic model can be used for the prediction of the mean values at new observed locations and times along with the associated uncertainty. These predictions can be computed for *any* arbitrary point, and they are not tied to the location of the training data, for example, to

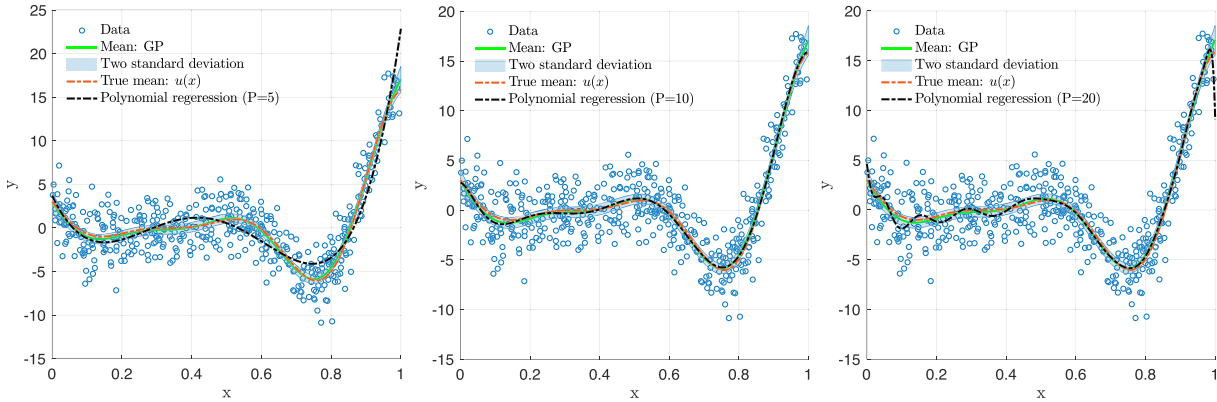


Figure 4. Comparison of Gaussian processes regression versus polynomial regression for synthetic data. The observations are generated by: $u(x) = (6x - 2)^2 \sin(12x - 4)$. GP = Gaussian processes.

satellite grids. In that sense, the predictions of the multifidelity model are continuous. The resulting posterior distribution at the high-fidelity level is taken as the prediction of the multifidelity model. The probabilistic prediction at arbitrary points \mathbf{x}_* correspond to conditioning the joint Gaussian prior distribution on the low-fidelity/high-fidelity observations (\mathbf{y}). To this end, we observe that the joint probability density function (PDF) of the new prediction points and the training points is also a Gaussian given by

$$\begin{bmatrix} u_H(\mathbf{x}_*) \\ \mathbf{y} \end{bmatrix} \sim \mathcal{N} \left(\mathbf{0}, \begin{bmatrix} k_{HH}(\mathbf{x}_*, \mathbf{x}_*) & \mathbf{q}^T \\ \mathbf{q} & \mathbf{K} \end{bmatrix} \right),$$

where $\mathbf{q}^T = [k_{HL}(\mathbf{x}_*, \mathbf{x}_L) \quad k_{HH}(\mathbf{x}_*, \mathbf{x}_H)]$. The above Gaussian relationship between the prediction points and high-fidelity/low-fidelity data can be used to find the conditional distribution to make predictions:

$$u_H(\mathbf{x}_*) | \mathbf{y} \sim \mathcal{N}(\mathbf{q}^T \mathbf{K}^{-1} \mathbf{y}, \quad k_{HH}(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{q}^T \mathbf{K}^{-1} \mathbf{q}), \quad (4)$$

where $\boldsymbol{\mu}_H = \mathbf{q}^T \mathbf{K}^{-1} \mathbf{y}$ is the posterior mean (i.e., predictions), the variance of the prediction is obtained by $\sigma_H^2 = \text{diag}(k_{HH}(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{q}^T \mathbf{K}^{-1} \mathbf{q})$, and $\pm 2\sigma_H$ represents the uncertainty of the multifidelity predictions.

3.2. Comparison with Other Methods

In this section, we compare GPR and multifidelity modeling with other existing approaches. We focus on two essential aspects of the presented multifidelity formulation, namely, regression and data assimilation.

3.2.1. Regression

First, we compare GPR with polynomial regression. GPR is a nonparametric regression technique, in which no explicit basis function is assumed, unlike the polynomial regression, which is a parametric regression technique and requires the specification of the basis (monomials). To compare GPR and polynomial regression, we consider synthetic data generated by function $u(x) = (6x - 2)^2 \sin(12x - 4)$ corrupted by noise. The size of the training set is 500. This amounts to a single-fidelity regression as there is only one data source. In Figure 4, the results of GPR and polynomial regression of different orders ($P = 5, 10$, and 20) are shown. It is clear that $P = 5$ lacks the required model complexity to accurately discover $u(x)$ and $P = 20$ is too complex. GPR, however, discovers $u(x)$ accurately (see Rasmussen & Williams, 2006 for a comprehensive discussion on GPR).

3.2.2. Data Fusion

Second, we compare the multifidelity model with OI. In OI terminology, low-fidelity and high-fidelity data are referred to as *background* and *observation*, respectively, and the mean of the posterior is referred to as *analysis*. We consider two sources of data that are synthetically generated: (1) High-fidelity by 4 samples of function $u_H(x) = (6x - 2)^2 \sin(12x - 4)$ and (2) low-fidelity by 21 samples of function $u_L(x) = 0.5u_H(x) + 10(x - 0.5) - 5$. The location of high-fidelity samples are chosen to be a subset of low-fidelity samples (same x values), which makes it unnecessary to specify the interpolator matrix \mathbf{H} in equation (1). In Figure 5a, the single-fidelity GPR with the four high-fidelity samples are shown. Obviously, four samples of $u_H(x)$ are not sufficient for an accurate regression, which is reflected in large uncertainty bands. We assume a square-exponential kernel for OI background and observation noises: $k(x, x'; \sigma, l_c) = \sigma^2 \exp(-(x - x')^2 / l_c^2)$, where σ^2 and l_c are the variance intensity and the correlation length, respectively.

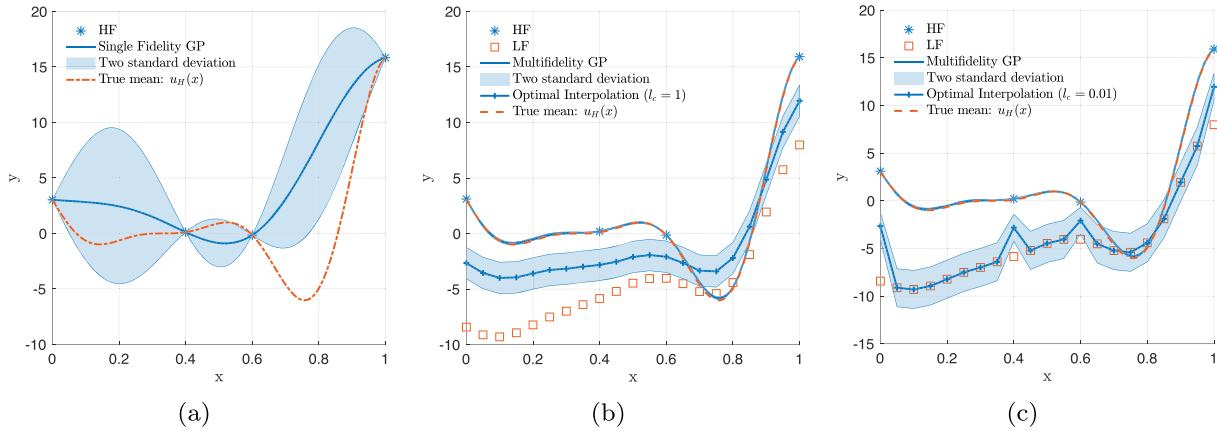


Figure 5. Comparison of the multifidelity model and optimal interpolation for synthetic data. The high- and low-fidelity observations are generated by $u_H(x) = (6x - 2)^2 \sin(12x - 4)$ and $u_L(x) = 0.5u_H(x) + 10(x - 0.5) - 5$: (a) Single-fidelity regression with Gaussian processes (GP); (b) Multifidelity GP and optimal interpolation (OI) with correlation length $l_c = 1.0$; (c) Multifidelity GP and OI with correlation length $l_c = 0.01$.

Therefore, $\mathbf{P} = k(\mathbf{x}_L, \mathbf{x}_L; \sigma_p^2, l_c^2)$ and $\mathbf{R} = k(\mathbf{x}_L, \mathbf{x}_L; \sigma_R^2, l_c^2)$. We consider $\sigma_p = \sigma_R = 1$. In Figures 5b and 5c, we compare the results of multifidelity model with OI with two different correlation lengths of $l_c = 1$ and $l_c = 0.01$. The multifidelity model recovers the true function with very good accuracy, which is reflected in small uncertainty. It is clear that the mean values of OI exhibit large discrepancy with the truth. This is mainly because in OI, both low-fidelity (background) and high-fidelity (observation) are assumed to be *unbiased*. This can be observed by subtracting the truth from both sides of equation (1) and obtain an equation for the error (for more details, see Daley, 1992). The correlation length affects the length of the region corrected by the high-fidelity observation. This is clearly demonstrated in Figures 5b and 5c, where $l_c = 1$ widens the neighborhood corrected by the high-fidelity observation, while $l_c = 0.01$ only improves the value of the assimilated data at the location of high-fidelity observation. On the other hand, the multifidelity model accounts for biased low-fidelity data and learns the bias $\delta(x)$ from the data, while the high-fidelity must be unbiased. Moreover, the multifidelity model learns the correlation length of the covariance matrices as hyper-parameters.

4. Demonstration Examples

In this section, we will demonstrate the application of the multifidelity framework in estimating SST in Massachusetts and Cape Cod Bays. In particular, we present results of three SST multifidelity models: (1) temporal, (2) spatial, and (3) spatiotemporal. For validation purposes, we split the MWRA measurements to two disjoint sets: *train* set that is used in model training and the *test* set that is used to validate the multifidelity predictions. In all of these demonstrations, we perform hindcast predictions.

4.1. Temporal Multifidelity Model of SST

With the proposed framework, we built a temporal multifidelity model for three MWRA stations: F02, F23, and N18 in the span of 2 years (2015–2016; see Figure 6d for the location of these stations in the Massachusetts Bay and Cape Cod Bay). Temporal multifidelity models are one-dimensional, and therefore, $D = 1$ with x representing time t , and the low-fidelity kernel takes the form of:

$$k(t, t') = \sigma_L^2 \exp\left(-\frac{(t - t')^2}{\theta_{L_t}^2}\right),$$

where σ_L and θ_{L_t} are the hyper-parameters. The kernel for $\delta(x)$ is defined similarly with the hyper-parameters σ_δ and θ_{δ_t} . With the addition of ρ , σ_{L_n} and σ_{H_n} , the temporal multifidelity model has seven hyper-parameters, which are learned from the satellite data and MWRA measurements. During these 2 years, 18 in situ measurements were taken at each station. In our models, we randomly chose $n_H = 11$ of those MWRA measurements as our high-fidelity training data and used the 7 remaining for validation purposes (stars and circles, respectively, as shown in Figures 6a–6c). For the low-fidelity data, we linearly interpolated the satellite measurement onto the location of the stations. Overall, we had $n_L = 125$, $n_L = 64$, and $n_L = 113$ low-fidelity points for stations F02, F23, and N18, respectively. Due to cloud coverage, we did not have

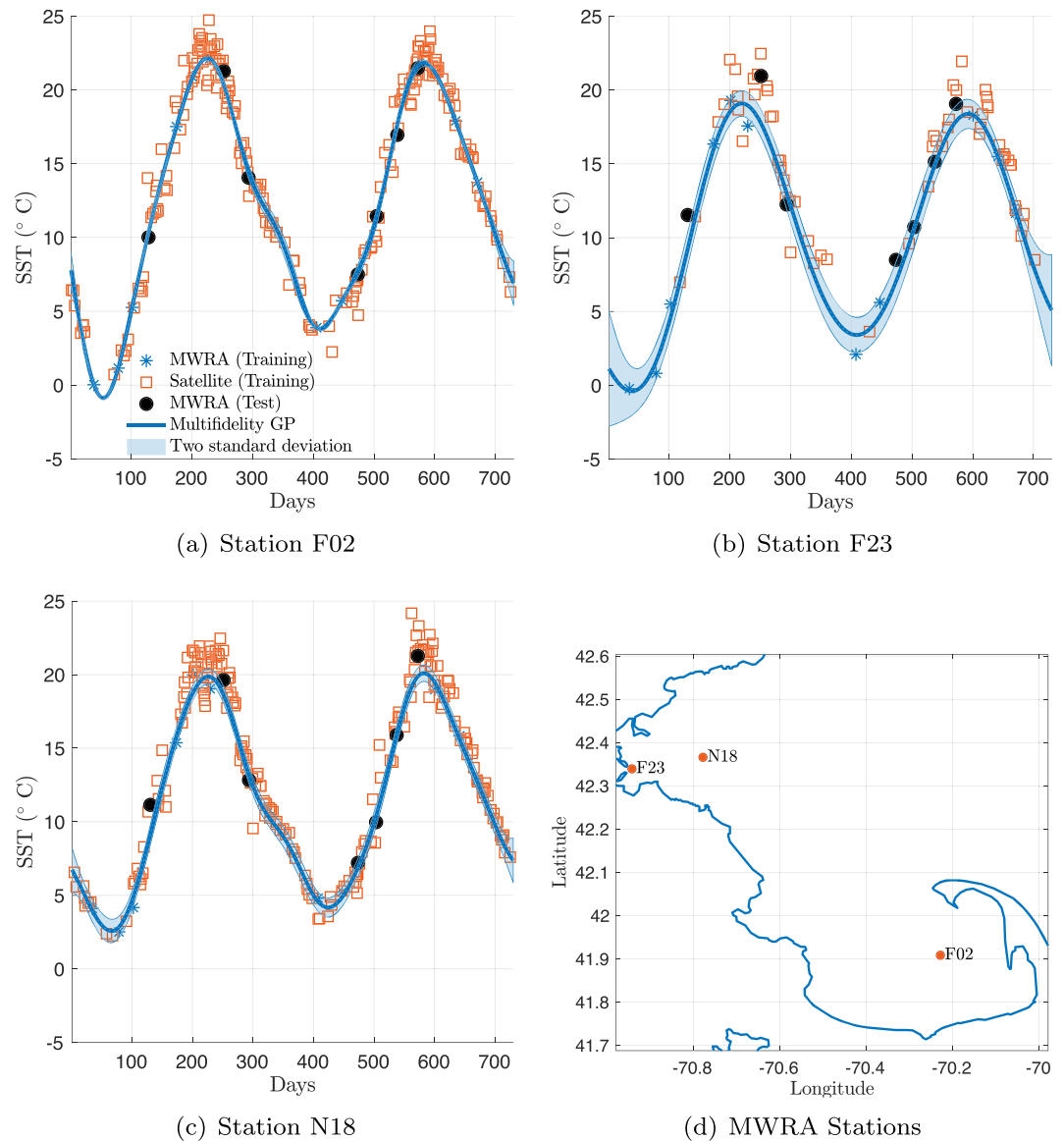


Figure 6. Validation of temporal multifidelity model of SST. Day 1 corresponds to 1 January 2015. High-fidelity data source: MWRA measurements at stations F02, F23, and N18 (star symbols), and low-fidelity data source: The MODerate-resolution Imaging Spectroradiometer satellite measurements (square symbols). The circle symbols show the MWRA measurements that are not used in training the model but are used for validation. The location of the stations is shown in panel (d). The code and data for this example are available at: <https://github.com/hbabaee/MFSST>. GP = Gaussian processes; MWRA = Massachusetts Water Resources Authority; SST = sea surface temperature.

satellite data for every day for all stations. In particular, station F02 has a low number of satellite measurements. In Figures 6a–6c, the mean values of the multifidelity model along with uncertainty band ($\pm 2\sigma$ standard deviation) are shown for stations F02, F23, and N18. We note the smaller uncertainty near the MWRA measurements, that is, high-fidelity data points. Also, station F02 has a smaller uncertainty band compared with that of stations F23 and N18. The smaller uncertainty estimate of the multifidelity model for station F02 was further justified by accurate hindcast of the MWRA test points, which are not used in training of the multifidelity model. For stations F23 and N18, the MWRA test points also lie within or near the uncertainty band.

4.2. Spatial Multifidelity Model of SST

As the second demonstration, we built space multifidelity models in the Massachusetts and Cape Cod Bays for fixed days by combining MWRA measurements with the satellite data. For the spatial

multifidelity model, $D = 2$ and $x = \{x_1, x_2\}$, where x_1 and x_2 are longitude and latitude, respectively. We use the square-exponential kernel for low-fidelity model as expressed by

$$k(x, x') = \sigma_L^2 \exp\left(-\frac{(x_1 - x'_1)^2}{\theta_{L_{x_1}}^2} - \frac{(x_2 - x'_2)^2}{\theta_{L_{x_2}}^2}\right),$$

and similarly, we employ a square-exponential kernel for $\delta(x)$. These two kernels contain six hyper-parameters: $\{\sigma_L, \theta_{L_{x_1}}, \theta_{L_{x_2}}, \sigma_\delta, \theta_{\delta_{x_1}}, \theta_{\delta_{x_2}}\}$. With the addition of ρ, σ_{n_L} and σ_{n_H} , the spatial multifidelity model has nine hyper-parameters.

For our demonstration, we built spatial multifidelity models for the days of 20 March 2015 and 18 May 2016, when MWRA measurements were available. We note that the MWRA measurements of different stations are not taken simultaneously, but they are all taken within a 24-hr time interval. However, we assume that these measurements are taken at exactly the same time. This assumption introduces a small uncertainty that will be estimated via the noise model ϵ_H in the high-fidelity GP. In Figures 7b and 7c, the mean and the standard deviation of the multifidelity model for 20 March 2015 are shown. To build this multifidelity model, we used the satellite data shown in Figure 7a as the low-fidelity data. We observe that the uncertainty is small near the MWRA stations, and away from the stations, the uncertainty increases (Figure 7c). Comparing the satellite data and the mean of the multifidelity model reveals that the multifidelity model relies on low-fidelity data albeit with a bias correction in regions where no MWRA measurements are available.

To demonstrate the effect of MWRA measurements on the multifidelity model, we built another multifidelity model for 20 March 2015, where the measurement of the MWRA station F29 was added to the training set. The results for this case are shown in the second row of Figure 7. It is clear that the addition of station F29 expands the low-uncertainty region near north of Cape Cod, and it results in a significant update in the multifidelity prediction, specially since the region near the location of station F29 is not densely sampled by MWRA nor is it covered with satellite data for this particular day.

In the last row of Figure 7, the predictions of the multifidelity model for 18 May 2016 are shown. For this day, the satellite measurements provide better coverage in both Massachusetts Bay and Cape Cod Bay. The results show that the satellite and MWRA measurements are correlated and that the multifidelity model corrects the bias uncertainty of the satellite with the inclusion of the MWRA measurements.

4.3. Spatiotemporal Multifidelity Model of SST

In this section, we demonstrate the results of the spatiotemporal multifidelity model of SST in the Massachusetts and Cape Cod Bays. This is our most comprehensive model as it exploits both spatial and temporal correlations of the SST data in a multifidelity framework. The spatiotemporal model is three-dimensional ($D = 3$) and $x = \{x_1, x_2, x_3\}$, where x_1 represents longitude, x_2 latitude and x_3 time. There are 11 hyper-parameters for this model. The low-fidelity kernel is expressed by

$$k(x, x') = \sigma_L^2 \exp\left(-\sum_{i=1}^D \frac{(x_i - x'_i)^2}{\theta_{L_{x_i}}^2}\right),$$

and it contains four hyper-parameters $\theta_L = \{\sigma_L, \theta_{L_{x_1}}, \theta_{L_{x_2}}, \theta_{L_{x_3}}\}$. Similarly, the square-exponential kernel for $\delta(x)$ introduces four hyper-parameters $\theta_\delta = \{\sigma_\delta, \theta_{\delta_{x_1}}, \theta_{\delta_{x_2}}, \theta_{\delta_{x_3}}\}$. The multifidelity model also contains three other hyper-parameters: the scaling factor ρ and the low-fidelity and high-fidelity standard deviations expressed with $\sigma_{n_L}, \sigma_{n_H}$, respectively.

We considered 14 MWRA stations in our analysis, from which the SST measured at 11 stations in the span of years 2015–2016 are selected as the training set for high-fidelity points. The remaining three stations are used as a test set for validation of the multifidelity model. The training and test MWRA stations are shown in Figure 8f with square and circle symbols, respectively. Overall, there are $n_H = 195$ high-fidelity points available. The satellite images are used as monthly snapshots, that is, skipping every 30 days resulting in $n_L = 2,526$ low-fidelity points. Using the entire satellite set of measurements with daily temporal resolution would result in over fitting of the multifidelity model. Besides the issue of over fitting, for cases with large training data sets, the computational cost of training the multifidelity model becomes increasingly prohibitive, as the computational complexity of training the multifidelity model is $\mathcal{O}(n^3)$, where $n = n_L + n_H$. This will also be the case for example when high-resolution measurements are needed to build multifidelity model with fine space-time resolution; (see Gardner et al., 2005) for such measurements in the region. Recently, parametric GP has been developed that enables training GP with big data (Raissi et al., 2019) and

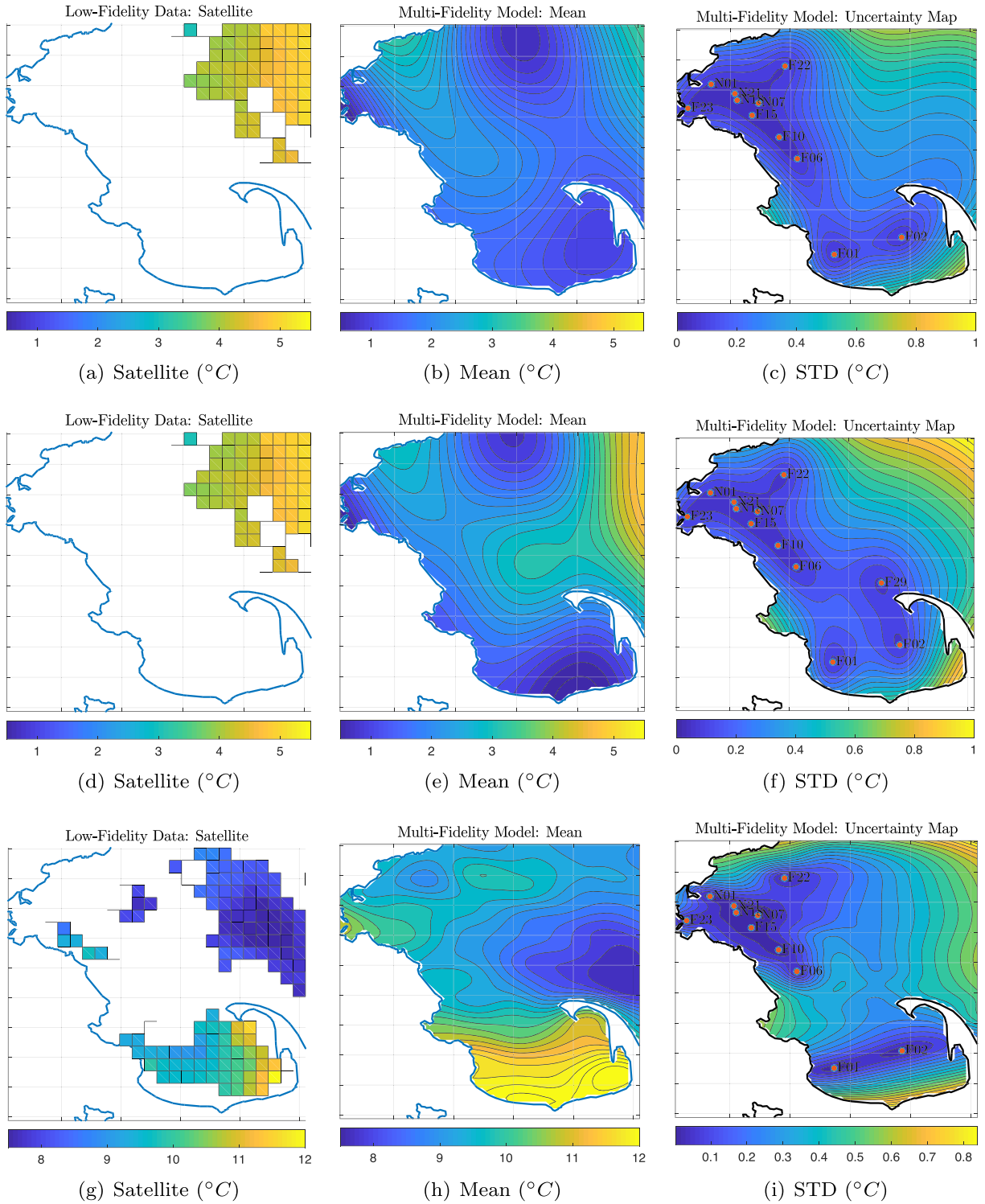


Figure 7. Spatial multifidelity model: The first two rows show results of multifidelity model for SST for 20 March 2015. The difference between the top row and the middle row is the inclusion of station F29 in the multifidelity model in the middle row. The bottom row shows results for 18 May 2016. Columns from left to right show satellite measurements, mean of multifidelity model, and standard deviation of the multifidelity model, respectively. The code and data for this example are available at: <https://github.com/hbabaee/MFSST>.

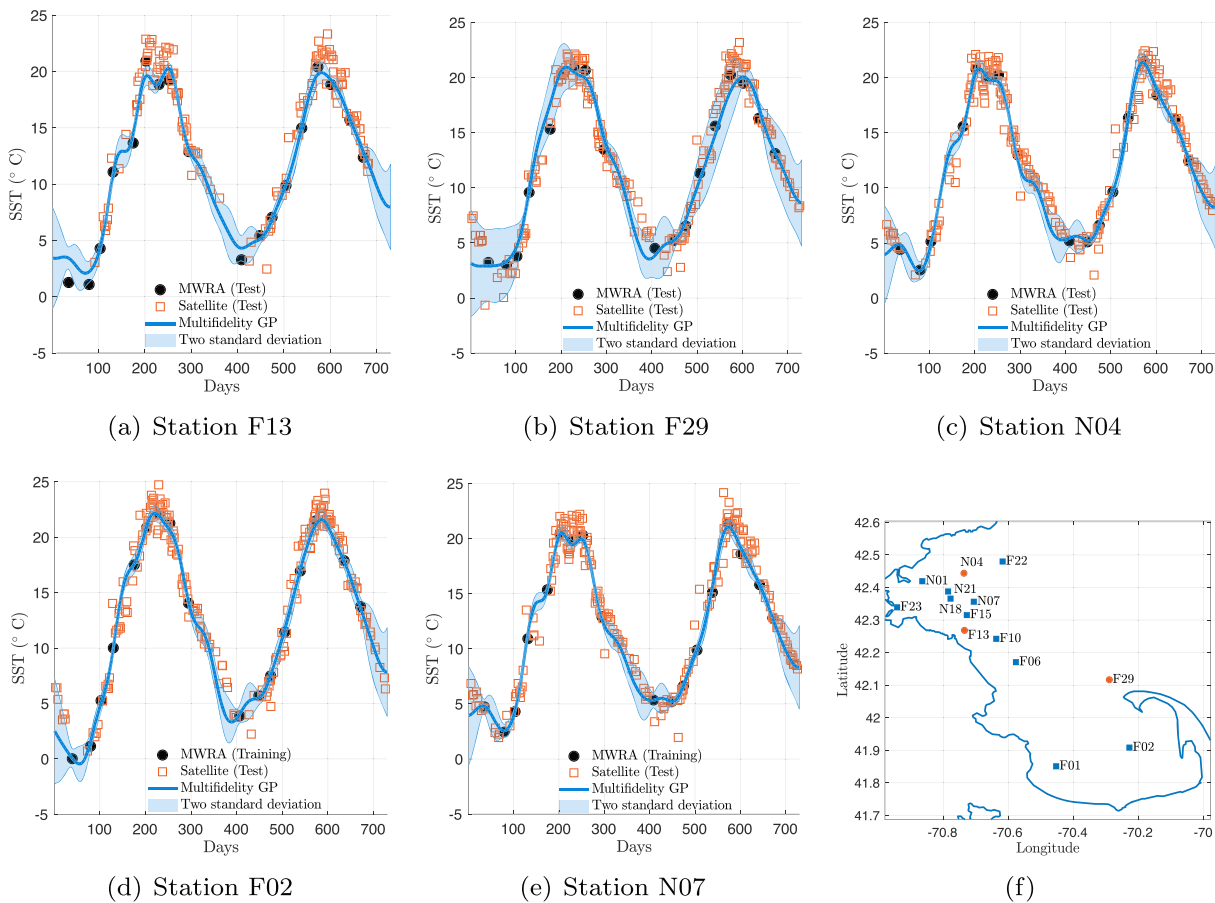


Figure 8. Validation of spatiotemporal multifidelity model of SST in years 2015–2016. Day 1 corresponds to 1 January 2015. On the top row, the multifidelity model is evaluated at three MWRA stations (F13, F29, and N04), whose measurements were not used in model training. The location of these three stations along with the rest of the stations are shown in panel (f) with circle and square symbols, respectively. The first two panels on the second row show the evaluation of the multifidelity model at MWRA stations F02 and N07, whose measurements were used in the model training. The code and data for this example are available at: <https://github.com/hbabaee/MFSST>. GP = Gaussian processes; MWRA = Massachusetts Water Resources Authority; SST = sea surface temperature.

it also alleviates the problem of over fitting. The parametric GP enables effective mini-batch training procedures with the computational complexity being reduced to $\mathcal{O}(Mn_B^3)$, where M is the number of mini-batches, and n_B is the number of training points in each mini-batch. Given that the MWRA and satellite data are provided on monthly intervals, the corresponding multifidelity model can resolve up to monthly time scales.

In Figures 8a–8c, the multifidelity model is validated with the test data set of three MWRA stations (F13, F29, and N04) in the span of 2015–2016. In all three cases, the MWRA measurements are in good agreement with the mean of the multifidelity model. These measurements lie mostly within the uncertainty bands, with a few cases nearby. Among these three stations, F29 shows the largest overall uncertainty as it is not surrounded by MWRA stations that are used in the training of the model (see Figure 8f for the location of F29). On the other hand, N04 has the lowest overall uncertainty and the multifidelity mean yields the best prediction among the three selected stations. In Figures 8d and 8e, the mean of the multifidelity model and its uncertainty are shown for stations F02 and N07. The measurements of both of these stations are used in the training of the model, and as a result, the uncertainty becomes very small near the time of the measurements. Note that in the absence of the noise model in the high-fidelity GP, represented by ϵ_H , the multifidelity mean will pass through the MWRA measurements, that is, zero uncertainty. The term ϵ_H estimates the noise in the MWRA measurements and the unresolved spatiotemporal scales in the multifidelity model.

In Figure 9a, we perform an independent validation of the spatiotemporal multifidelity model by comparing the hindcast prediction of our model with the in situ measurements of buoy A01 for the years 2015 and 2016.

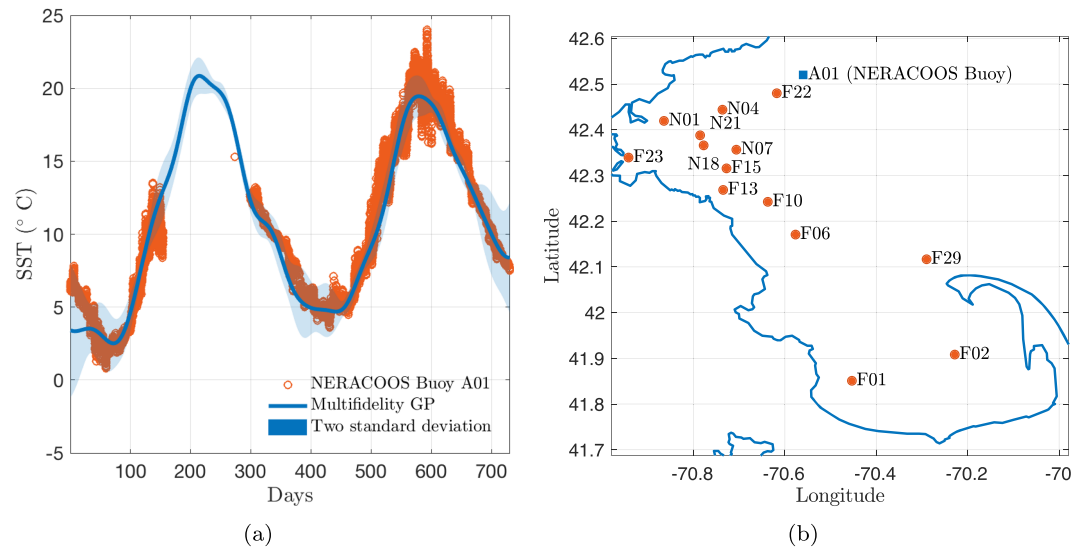


Figure 9. Independent validation of spatiotemporal multifidelity model with NERACOOS buoy A01. The code and data for this example are available at: <https://github.com/hbabaee/MFSST>. GP = Gaussian processes; SST = sea surface temperature.

The buoy A01 is part of the Northeast Coastal Ocean Observation System NERACOOS, and it is located at $42^{\circ} 31' N$ and $70^{\circ} 34' W$ as shown in Figure 9b with a square symbol. The data for this buoy can be downloaded from: <http://www.neracoos.org>. The buoy A01 measurements are taken hourly, but the satellite and MWRA training data of the spatiotemporal multifidelity model have a monthly temporal resolution. As a result, the higher frequency oscillations are not resolved in the multifidelity model and they are accounted for with the uncertainty as can be seen in Figure 9a. Overall, a good agreement between the mean of the multifidelity model and the buoy measurements is observed.

In Figure 10, we further illustrate the learning of the large-scale seasonal dynamics via the MWRA measurements in the spatiotemporal multifidelity model. In this figure, we show snapshots of the mean and the standard deviation taken every 10 days from 1 January to 21 May 2015 and the approximate days of MWRA measurements (red circles). We observe that uncertainty is consistently lower near the MWRA measurements. However, during the winter, uncertainties are larger compared to late April and May. This is due to the scarcity of satellite data and the lack of MWRA measurements during winter months. This is further confirmed by comparing the measurements of buoy A01 with the multifidelity predictions. In Figure 10, we show the prediction of the multifidelity and its uncertainty (i.e., $\pm 2\sigma$) at the location of the buoy A01 for two days. The buoy measurements are also shown for comparison. On 1 January 2015, the uncertainty is large (i.e., $\pm 4.6^{\circ} C$) due to lack of data, while on 1 April 2015, the uncertainty is lower (i.e., $\pm 1.4^{\circ} C$). In both cases, the buoy measurements lie within the uncertainty band.

5. Summary and Future Work

The presented framework goes beyond traditional predictions with line error bars to continuous space-time uncertainty maps by utilizing and developing modern machine learning techniques that enable fusion of data from all available sources with varying degrees of fidelity. In Figure 11, we compare the difference between single-fidelity Gaussian regression model with the multifidelity model for 20 March 2015 (top row) and 18 May 2015 (bottom row). In the left column, the satellite data on the corresponding days are shown, and the middle column shows the Gaussian regression of the satellite measurements, that is, the single-fidelity model. The right column shows the mean values of the multifidelity model, where we have used MWRA measurements as the source of high-fidelity data. The comparison of the mean values of the single-fidelity and multifidelity models clearly shows that these two temperature maps are correlated. However, the multifidelity model corrects the single-fidelity model up to $3^{\circ} C$ on 20 March 2015 and roughly $1-2^{\circ} C$ on 18 May 2015.

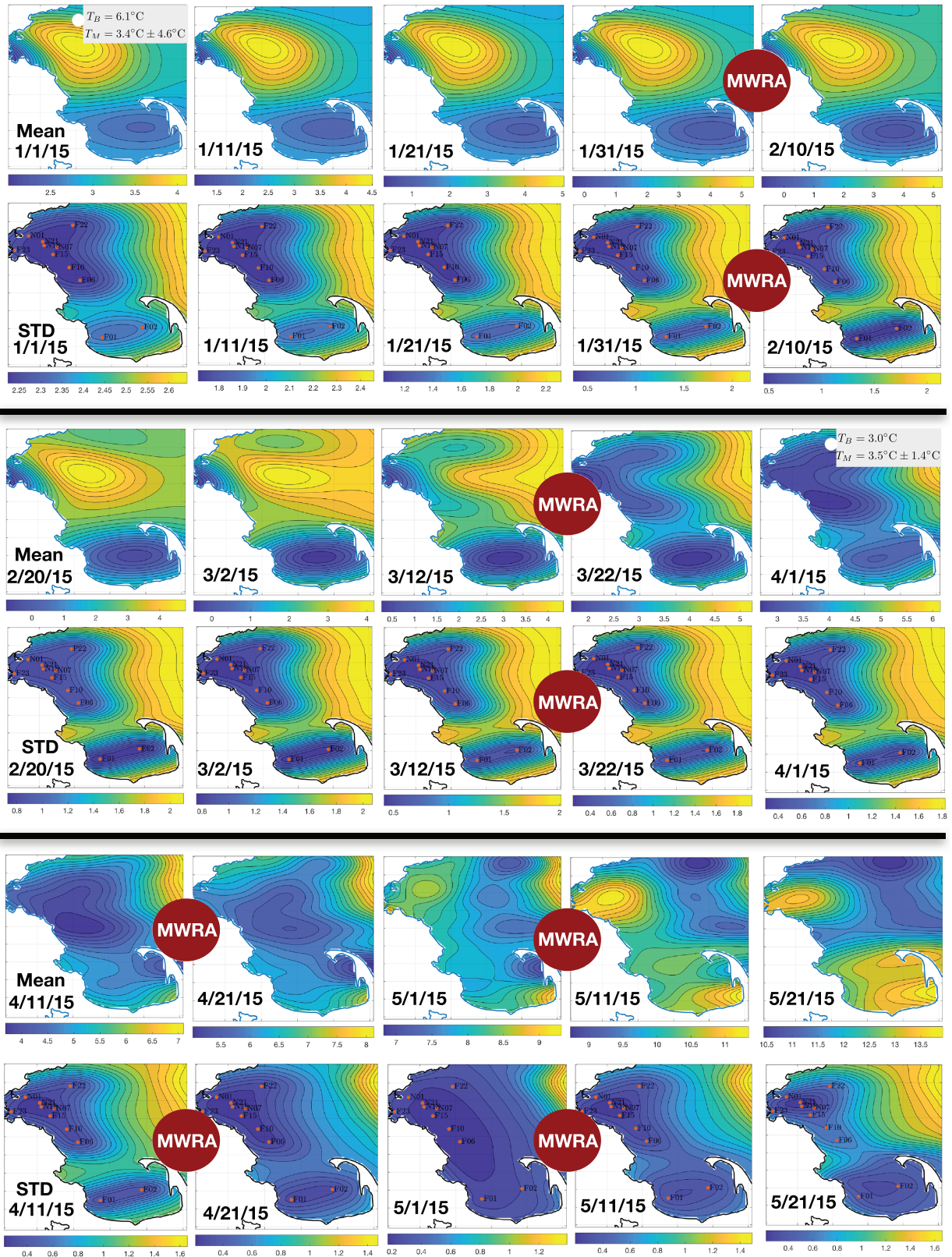


Figure 10. Spatiotemporal multifidelity model: sea surface temperature (SST) mean and the standard deviation (first and second row, respectively) starting from 1 January 2015 with 10 days skipped between subsequent snapshots. The approximate times of MWRA measurements are shown with red circles. On 1 January and 1 April, the position of buoy A01 (white dot), its SST measurement (T_B), and the prediction of the model (T_M) at that location are shown. MWRA = Massachusetts Water Resources Authority.

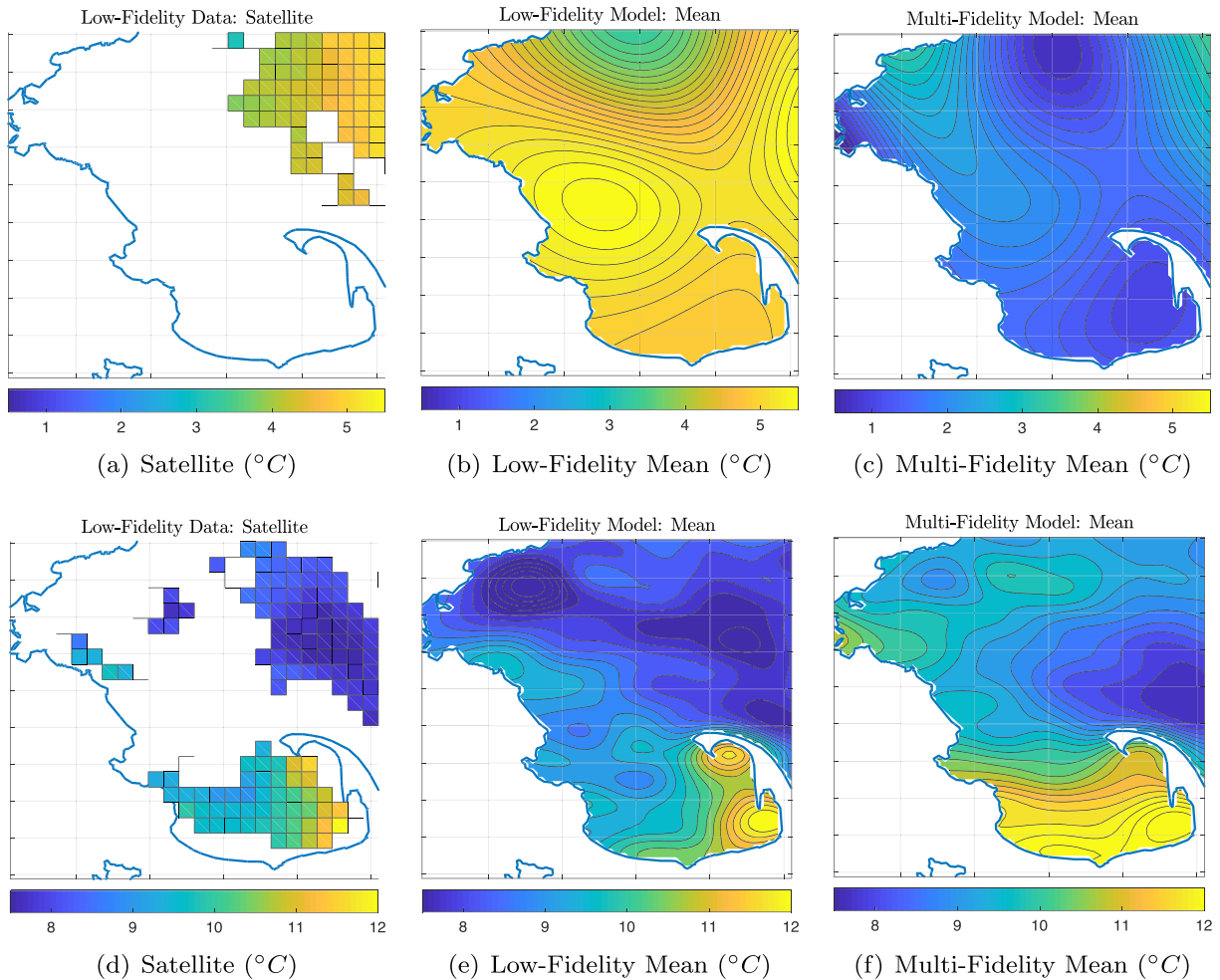


Figure 11. Comparison of spatial low-fidelity and multifidelity models of sea surface temperature for 20 March 2015 (top row) and 18 May 2015 (bottom row). Left: Satellite measurements; Middle: Mean of low-fidelity model; Right: Mean of multifidelity model.

Our long-term goal is to analyze and forecast various quantities of interest *and* their uncertainty directly related to ocean acidification, that is, salinity, temperature, dissolved inorganic carbon, pH, total alkalinity, using a flexible Bayesian data fusion framework that *leaves no data behind* —including 3D physics-based models (see e.g., Xue et al., 2014). In particular, we will study the effect of kernel on hindcast and forecast predictions. Kernels can be chosen to encode various structures in the data including smoothness, stationarity, and seasonality. In Figure 12, we show the performance of square exponential and periodic kernels

$$k_{SE}(t, t') = \sigma^2 \exp\left(\frac{-(t - t')^2}{\theta^2}\right) \quad \text{and} \quad k_{PER}(t, t') = \sigma^2 \exp\left(\frac{-2\sin^2(\pi|t - t'|/p)}{\theta^2}\right)$$

in forecasting the SST. In the above kernels σ , l , and p are hyper-parameters. The data used in Figure 12 are obtained from buoy measurements located in the Gulf of Maine at 43.02° N, 70.54° W representing SST measurements during more than 9 years — spanning from July 2006 to November 2015. It is evident that for this long time record, the square exponential kernel fails to learn the seasonality beyond the training data, and therefore, it cannot be used for reliable forecast. This is also reflected in the large uncertainty bands. On the other hand, the periodic kernel captures the seasonality and it yields accurate forecast over a long period of time. Choosing an optimal kernel can be automated via Bayesian optimization (Gardner et al., 2015; Martinez-Cantin et al., 2009), or via neural-net-induced GP (Pang et al., 2019; for a broader list of research in this area, see the Automatic Statistician Project, <https://www.automaticstatistician.com>).

Ultimately, we intend to build 3D volumetric maps of these quantities in the Massachusetts and Cape Cod Bays, including Boston Harbor and the Stellwagen Bank. This will enhance our understanding of coastal

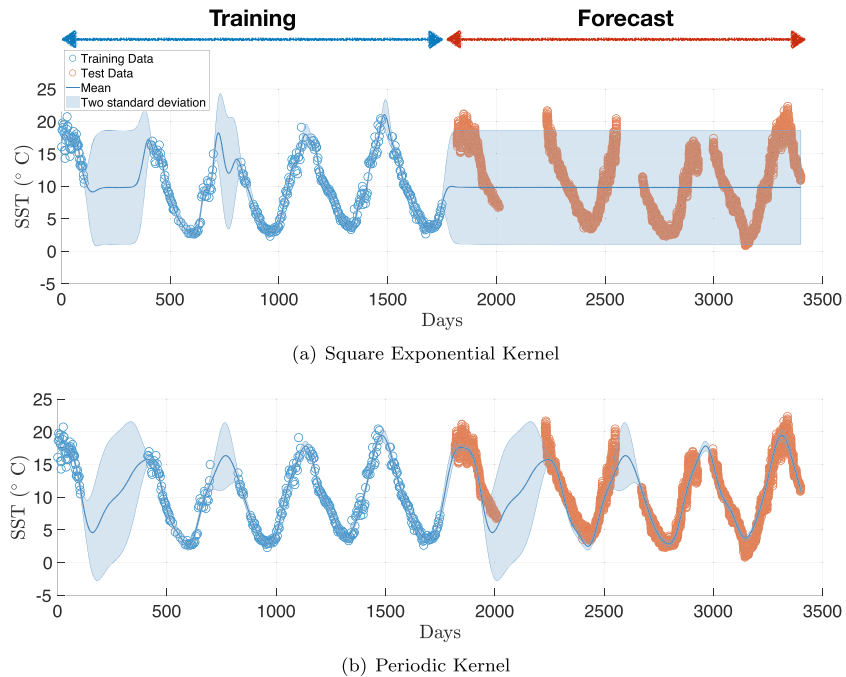


Figure 12. Effect of kernel on model performance: Forecast seasonality of sea surface temperature (SST) using Gaussian processes with (a) square exponential kernel and (b) periodic kernel. The data used are mooring measurements of a buoy in Coastal Western Gulf of Maine at 43.02° N, 70.54° W maintained by the University of New Hampshire Coastal Carbon Group. The data can be downloaded from: https://www.nodc.noaa.gov/ocads/oceans/Coastal/unh_ts.html. Day 1 corresponds to 13 July 2006. The color blue shows the data used in the training of the model (hindcast), and the data in red show the test data (forecast).

ocean acidification and our ability to focus our monitoring efforts in places of high uncertainty and providing management strategies in regions with relative low aragonite saturation means that could negatively affect valuable resources.

Acknowledgments

The authors thank Dr. Raissi for useful comments and stimulating discussions. This research has been supported primarily by a grant from NOAA (NA18OAR4170105) and partial support from Office of Naval Research, grant number N00014-16-1-2956. The Moderate-resolution Imaging Spectroradiometer (MODIS) SST data were obtained from the NASA EOSDIS Physical Oceanography Distributed Active Archive Center (PO.DAAC) at the Jet Propulsion Laboratory, Pasadena, CA (<https://doi.org/10.5067/MODST-1D4N4>). The MWRA measurements may be downloaded from http://www.mwra.com/harbor/html/wq_data.htm.

References

- Alin, S. R., Feely, R. A., Dickson, A. G., & Hernández, J. M. (2018). Robust empirical relationships for estimating the carbonate system in the southern California current system and application to CalCOFI hydrographic cruise data (2005–2011). *Journal of Geophysical Research: Oceans*, *117*, C05033. <https://doi.org/10.1029/2011JC007511>
- Babae, H., Perdikaris, P., Chrysostomidis, C., & Karniadakis, G. E. (2016). Multi-fidelity modelling of mixed convection based on experimental correlations and numerical simulations. *Journal of Fluid Mechanics*, *809*, 895–917. <https://doi.org/10.1017/jfm.2016.718>
- Chin, T. M., Vazquez-Cuervo, J., & Armstrong, E. M. (2017). A multi-scale high-resolution analysis of global sea surface temperature. *Remote Sensing of Environment*, *200*, 154–169. <https://doi.org/10.1016/j.rse.2017.07.029>
- Daley, R. (1992). *Atmospheric data analysis*. Cambridge: Cambridge University Press.
- Dash, P., Ignatov, A., Martin, M., & Donlon, C. (2012). Group for high resolution sea surface temperature (ghrsst) analysis fields inter-comparisons—part 2: Near real time web-based level 4 sst quality monitor (l4-squam). *Deep Sea Research Part II: Topical Studies in Oceanography*, *77-80*, 31–43. <https://doi.org/10.1016/j.dsr2.2012.04.002>
- Davis, C. V., Hewett, K., Hill, T. M., Largier, J. L., Gaylord, B., & Jahncke, J. (2018). Reconstructing aragonite saturation state based on an empirical relationship for northern California. *Estuaries and Coasts*, *41*, 2056–2069. <https://doi.org/10.1007/s12237-018-0372-0>
- Forrester, A. I. J., Sobester, A., & Keane, A. J. (2007). Multi-fidelity optimization via surrogate modelling. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *463*, 3251–3269. <https://doi.org/10.1098/rspa.2007.1900>
- Gardner, G. B., Chen, R. F., & Berry, A. (2005). High-resolution measurements of chromophoric dissolved organic matter (cdom) in the neponset river estuary, boston harbor, ma. *Marine Chemistry*, *96*(1), 137–154. <https://doi.org/10.1016/j.marchem.2004.12.006>
- Gardner, J. R., Malkomes, G., Garnett, R., Weinberger, K. Q., Barbour, D., & Cunningham, J. P. (2015). Bayesian active model selection with an application to automated audiometry. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2, NIPS'15*, MIT Press, Cambridge, MA, USA, pp. 2386–2394.
- Hare, J. A., Morrison, W. E., Nelson, M. W., & Stachura, M. M. (2016). A vulnerability assessment of fish and invertebrates to climate change on the Northeast US Continental Shelf. *PLoS one*.
- Juranek, L. W., Feely, R. A., Peterson, W. T., & Alin, S. R. (2009). A novel method for determination of aragonite saturation state on the continental shelf of central oregon using multi-parameter relationships with hydrographic data. *Geophysical Research Letters*, *36*, L24601. <https://doi.org/10.1029/2009GL040778>
- Kennedy, M. C., & O'Hagan, A. (2000). Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, *87*, 1–13. <https://doi.org/10.1093/biomet/87.1.1>

- Le Gratiot, L., & Le Gratiot, J. (2014). Recursive co-kriging model for design of computer experiments with multiple levels of fidelity. *International Journal for Uncertainty Quantification*, 4(5), 365–386.
- Li, A., Bo, Y., Zhu, Y., Guo, P., Bi, J., & He, Y. (2013). Blending multi-resolution satellite sea surface temperature (sst) products using bayesian maximum entropy method. *Remote Sensing of Environment*, 135, 52–63.
- Li, Y., He, R., Chen, K., & McGillicuddy, D. J. (2015). Variational data assimilative modeling of the gulf of maine in spring and summer 2010. *Journal of Geophysical Research: Oceans*, 120, 3522–3541. <https://doi.org/10.1002/2014JC010492>
- Libby, P. S., Borkman, D. G., Geyer, W. R., & Turner, J. T. (2017). Water column monitoring results: Massachusetts Water Resources Authority.
- Liu, Z. Q., & Rabier, F. (2006). The interaction between model resolution, observation resolution and observation density in data assimilation: A one-dimensional study. *Quarterly Journal of the Royal Meteorological Society*, 128(582), 1367–1386. <https://doi.org/10.1256/003590002320373337>
- Martinez-Cantin, R., de Freitas, N., Brochu, E., Castellanos, J., & Doucet, A. (2009). A bayesian exploration-exploitation approach for optimal online sensing and planning with a visually guided mobile robot. *Autonomous Robots*, 27(2), 93–103. <https://doi.org/10.1007/s10514-009-9130-2>
- Oke, P. R., & Sakov, P. (2008). Representation error of oceanic observations for data assimilation. *Journal of Atmospheric and Oceanic Technology*, 25(6), 1004–1017. <https://doi.org/10.1175/2007JTECHO558.1>
- Pang, G., Yang, L., & Em Karniadakis, G. (2019). Neural-net-induced gaussian process regression for function approximation and pde solution. *Journal of Computational Physics*, 384, 270–288. <https://doi.org/10.1016/j.jcp.2019.01.045>
- Perdikaris, P., Raissi, M., Damianou, A., Lawrence, N. D., & Karniadakis, G. E. (2017). Nonlinear information fusion algorithms for data-efficient multi-fidelity modelling. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 473(2198), 20160751.
- Pershing, A. J., Alexander, M. A., Hernandez, C. M., & Kerr, L. A. (2015). Slow adaptation in the face of rapid warming leads to collapse of the gulf of maine cod fishery. *Science*, 350(6262), 809.
- Prempraneerach, P., Perdikaris, P., Karniadakis, G. E., & Chrysostomidis, C. (2017). Sea surface temperature estimation from satellite observations and in-situ measurements using multifidelity Gaussian process regression. In *In 2017 International Conference on Digital Arts, Media and Technology (ICDAMT)*, pp. 28–33.
- Raissi, M., Babae, H., & Em Karniadakis, G. (2019). Parametric gaussian process regression for big data. *Computational Mechanics*, 64(2), 409–416.
- Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*. Cambridge, MA: MIT Press.
- Reece, S., & Roberts, S. (2010). An introduction to Gaussian processes for the Kalman filter expert. In *In 2010 13th International Conference on Information Fusion*, pp. 1–9.
- Smith, T. M., Reynolds, R. W., Livezey, R. E., & Stokes, D. C. (1996). Reconstruction of historical sea surface temperatures using empirical orthogonal functions. *Journal of Climate*, 9(6), 1403–1420.
- Steneck, R. S., Wahle, R. A., & Sainte-Marie, B. (2013). American lobster dynamics in a brave new ocean 1. *Canadian Journal of Fisheries and Aquatic Sciences*, 70(11), 1612–1624.
- Stock, C. A., Pegion, K., Vecchi, G. A., & Alexander, M. A. (2015). Seasonal sea surface temperature anomaly prediction for coastal ecosystems. *Progress in Oceanography*, 137, 219–236.
- Tommasi, D., Stock, C. A., Hobday, A. J., & Methot, R. (2017). Managing living marine resources in a dynamic environment: The role of seasonal to decadal climate forecasts. *Progress in Oceanography*, 152, 15–49.
- Werdell, P. J., Franz, B. A., Bailey, S. W., & Feldman, G. C. (2013). Generalized ocean color inversion model for retrieving marine inherent optical properties. *Applied Optics*, 52(10), 2019–2037.
- Werme, C., Keay, K. E., Libby, P. S., & Codiga, D. L. (2017). Outfall monitoring overview: Massachusetts Water Resources Authority.
- Xue, P., Chen, C., Qi, J., & Beardsley, R. C. (2014). Mechanism studies of seasonal variability of dissolved oxygen in Mass Bay: A multi-scale FVCOM/UG-RCA application. *Journal of Marine Systems*, 131, 102–119.
- Yang, X., Venturi, D., Chen, C., Chrysostomidis, C., & Karniadakis, G. E. (2010). EOF-based constrained sensor placement and field reconstruction from noisy ocean measurements: Application to Nantucket Sound. *Journal of Geophysical Research*, 115, C12072. <https://doi.org/10.1029/2010JC006148>
- Zhu, Y., Bo, Y., Zhang, J., & Wang, Y. (2017). Fusion of multisensor ssts based on the spatiotemporal hierarchical bayesian model. *Journal of Atmospheric and Oceanic Technology*, 35(1), 91–109.
- Zhu, Y., Kang, E. L., Bo, Y., Zhang, J., Wang, Y., & Tang, Q. (2019). Hierarchical bayesian model based on robust fixed rank filter for fusing modis SST and AMSR-E SST. *Photogrammetric Engineering and Remote Sensing*, 85(2), 119–131.