Spatial Statistics and Uncertainty Quantification in Remote Sensing and Climate Science

Emily L. Kang

Division of Statistics and Data Science

Department of Mathematical Sciences

University of Cincinnati

SIAM UQ24 Satellite Event: Power of Diversity in Uncertainty Quantification (POD-UQ)

Spatial Statistics

- **Spatial Statistics** broadens the classical statistical models and methods to those **recognizing the presence and importance of spatial information**
  - It *takes spatial dependence into account*.

"After choosing the *area* we usually have no guidance beyond the widely verifiable fact that patches in close proximity are commonly more alike, as judged by the yield of crops, than those which are far apart."

– R. A. Fisher (1935) on analyzing agricultural field

A Classical Spatial Statistical Model

Suppose that we want to infer a latent univariate spatial process $\{Y(\mathrm{s}) : \mathrm{s} \in D\}$ from incomplete and noisy data at a finite of locations.

- Data process (describing observed or potentially observed data):

$$Z(\mathrm{s}) = Y(\mathrm{s}) + \varepsilon(\mathrm{s}); \mathrm{s} \in D$$

- Latent spatial process of interest: $Y(\cdot)$
- Measurement-error process: $\varepsilon(\cdot)$ white noise with $E(\varepsilon(\mathrm{s})) = 0$, $var(\varepsilon(\mathrm{s})) = \sigma_\varepsilon^2 v(\mathrm{s})$
- Data vector: $\mathrm{Z} \equiv (Z(\mathrm{s}_1), \ldots, Z(\mathrm{s}_n))'$
  - Sample size: n, which can be large.

Spatial Statistical Model, cont'd

- Model the latent process $Y(\cdot)$ with

$$Y(\mathrm{s}) = \mu(\mathrm{s}) + \nu(\mathrm{s}),$$

  - $\mu(\cdot)$ denotes the large-scale variation, referred to as trend in spatial statistics literature.
  - $\nu(\cdot)$ is a spatial process with mean zero and a covariance function (kernel):

$$E(\nu(\mathrm{s})) = 0,$$

$$cov(\nu(\mathrm{s}_i), \nu(\mathrm{s}_j)) = C(\mathrm{s}_i, \mathrm{s}_j).$$

  - $\nu(\cdot)$ and $\varepsilon(\cdot)$ are assumed to be independent.

Spatial Statistical Model, cont'd

Define Y, ε, and $\boldsymbol{\nu}$ in an analogous manner to the data vector $\mathsf{Z} \equiv (Z(\mathsf{s}_1), \ldots, Z(\mathsf{s}_n))'$.
We obtain

$$\mathsf{Z} = \mathsf{Y} + \varepsilon$$

Then

$$\begin{aligned}
\Sigma &= var(\mathsf{Z}) = var(\boldsymbol{\nu}) + var(\varepsilon) \\
&\equiv C + \sigma_\varepsilon^2 V \\
&\equiv \begin{pmatrix} C(\mathsf{s}_1, \mathsf{s}_1) & C(\mathsf{s}_1, \mathsf{s}_2) & \cdots & C(\mathsf{s}_1, \mathsf{s}_n) \\ C(\mathsf{s}_2, \mathsf{s}_1) & C(\mathsf{s}_2, \mathsf{s}_2) & \cdots & C(\mathsf{s}_2, \mathsf{s}_n) \\ \vdots & \vdots & \vdots & \vdots \\ C(\mathsf{s}_n, \mathsf{s}_1) & C(\mathsf{s}_n, \mathsf{s}_2) & \cdots & C(\mathsf{s}_n, \mathsf{s}_n) \end{pmatrix} + \sigma_\varepsilon^2 V
\end{aligned}$$

where $V \equiv diag(v(\mathsf{s}_1), \ldots, v(\mathsf{s}_n))$.

Kriging

- Kriging: Infer $Y(s_0)$ from data Z, in an optimal way
  - It is also called Spatial BLUP. Matheron (1962) used the term in honor of D. G. Krige, a South African mining engineer. It is also known as optimum interpolation in atmospheric sciences
  - Optimality: $\hat{Y}(s_0)$ minimizes the mean squared prediction error (MSPE)
    - MSPE of generic predictor $Y^*(\cdot)$: $E(Y(s_0) - Y^*(s_0))^2$
- Derivation of Kriging requires only the mean and covariance functions but we don't need to assume a joint Gaussian distribution.
  - When we add the assumption with a Gaussian process (GP), the Kriging predictor will become the predictor we use in GP regression.

Kriging, ctd

- Kriging predictor:

$$\hat{Y}(s_0) = T(s_0)'\hat{\alpha} + k(s_0)'(Z - T\hat{\alpha})$$

- Associated uncertainty, described via the Kriging standard error:

$$\begin{aligned}\sigma_k(s_0) = \quad & \{C(s_0, s_0) - k(s_0)'\Sigma k(s_0) \\ & +(T(s_0) - T'k(s_0))'(T'\Sigma^{-1}T)^{-1}(T(s_0) - T'k(s_0))\}^{1/2}\end{aligned}$$

where

$$\hat{\alpha} = (T'\Sigma^{-1}T)^{-1}T'\Sigma^{-1}Z,$$

$$k(s_0) = \Sigma^{-1}c(s_0),$$

$$c(s_0) = cov(Z, Y(s_0)) = (C(s_0, s_1), \ldots, C(s_0, s_n))',$$

and Σ denote the $n \times n$ covariance matrix of the data vector: $\Sigma_{n\times n} = var(Z)$.

- Research problems in **spatial statistics+UQ** are often driven by the need to analyze and model complex spatial data/processes, such as those from remote sensing and climate sciences:
  - Noise and missing data
  - Large data volumes
  - Complicated spatial/spatio-temporal dependence, or dependence structure related to input variables
  - Multi-resolution, multi-scale, multivariate data
  - Lack of spatially representative ground truth
  - How to perform UQ from *inferred* "data products"

I will present two case studies to demonstrate how we attempted using spatial statistical methods for uncertainty quantification in remote sensing and climate science and discuss the challenges we have found and potential research directions from different viewpoints.

1. Providing fine-scale climate projections using coarse-scale climate model output and remote sensing data, **statistical downscaling**
2. **Constructing an emulator** for a complex physics-based forward model in the Orbiting Carbon Observatory (OCO-2).

Introduction

- Climate models are crutial tools for scientists to project future climate change and to understand its potential impact.
  - Atmosphere–ocean general circulation models (GCMs) are developed to simulate the climate over the entire *globe*.
  - Due to model complexity and limitations of computational resources, GCMs are restricted to generate outputs on coarse spatial scales, typically 200–500km.
- GCM outputs are limited in describing local/regional climate phenomena that are more relevant to natural-resource management and environmental policy decisions. $\rightarrow$ **Downscaling** is needed!

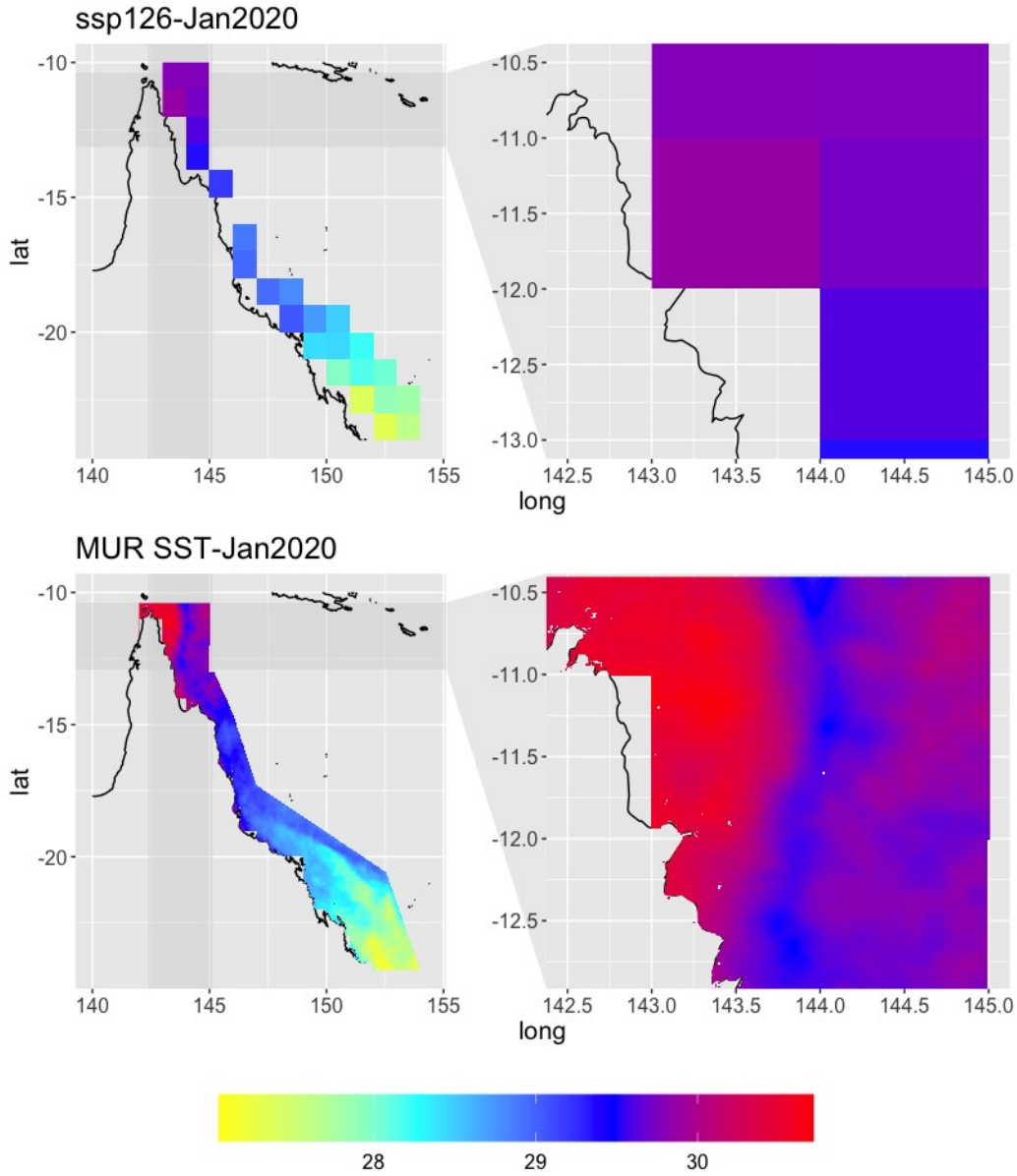
Climate Model Downscaling

- There are two main types of downscaling:
  - **Dynamical downscaling**: Using high-resolution regional simulations, also called regional climate models (RCMs), with initial and boundary conditions provided from a GCM.
  - **Statistical downscaling**: Establishing a statistical relationship to transfer coarse-resolution GCM outputs into fine-resolution outputs.
- Physics-driven vs. Data-driven
- Computation-intensive vs. Computation-efficient
- Uncertainty quantification (UQ)

Motivating Example: Downscaling Sea Surface Temperature

- Coral reefs face stressors such as rising ocean temperatures (Hughes et al., 2003; Hoegh-Guldberg et al., 2007; Gattuso et al., 2015; Masson-Delmotte et al., 2018) in addition to other local factors such as destructive fishing practices and coastal development.
  - Thermal stress index are developed by Coral Reef Watch (Liu et al, 2003, 2006) based on sea surface temperatures.
  - Downscaled sea surface temperature (SST) are required to calculate such index in order to better inform local conservation decisions.
- Van Hooidonk et al. (2016) and Dixon et al. (2022): Interpolating GCM SST outputs to the fine resolution grid and adjusting the mean based on observations.

- Our study region is the Great Barrier Reef (GBR) region.
- Data:
  - Monthly averaged NASA/JPL Multiscale Ultrahigh Resolution (MUR) satellite SST data at 1km resolution from June 2002 to December 2020; more than 300,000 pixels in the study region
  - Monthly SST outputs from a GCM model at 100km resolution in the Coupled Model Intercomparison Project Phase 6 (CMIP6).


ssp126-Jan2020
-10
-15
-20
lat
140
145
150
155
-10.5
-11.0
-11.5
-12.0
-12.5
-13.0
142.5
143.0
143.5
144.0
144.5
145.0
long
MUR SST-Jan2020
-10
-15
-20
lat
140
145
150
155
-10.5
-11.0
-11.5
-12.0
-12.5
142.5
143.0
143.5
144.0
144.5
145.0
long
28
29
30

Jointly Modeling Climate Model Output and MUR Data

- Let $Y_{t,1}(\mathrm{s})$ and $Y_{t,2}(\mathrm{s})$ denote the MUR SST and interpolated and bias corrected GCM SST at location s and time t, respectively.
- We take a **joint modeling** perspective and assume the bi-variate spatial process, $\mathrm{Y}_t(\mathrm{s}) \equiv (Y_{t,1}(\mathrm{s}), Y_{t,2}(\mathrm{s}))'$,

$$\mathrm{Y}_t(\mathrm{s}) = \boldsymbol{\mu}_t(\mathrm{s}) + \mathbf{v}_t(\mathrm{s})$$

  - The trend $\boldsymbol{\mu}_t(\mathrm{s})$ is modeled to be the 5-year moving averages of $Y_{t,2}(\mathrm{s})$ and a linear combination of large-scale basis functions.
  - We assume that the trend term captures temporal and spatio-temporal dependence and assumes that $\mathbf{v}_t(\mathrm{s})$ is independent across time t.
  - We model $\mathbf{v}_t(\mathrm{s}) = (v_{t,1}(\mathrm{s}), v_{t,2}(\mathrm{s}))'$ as a **bi-variate Gaussian process** with a basis-function representation to achieve efficient computation.

Downscaling

- Similar to what we do in GP regression, the conditional mean are used for predictions, while the conditional variance are used to describe the associated uncertainty.
- We can further generate the predictive distribution of the downscale SST field (over more than 300,000 pixels jointly) by taking advantage of the basis function representation and sampling from conditional distributions:
  1. Sample $v_{t_0,2}|\mathrm{Y}_{t_0,2}$
  2. Sample $v_{t_0,1}|v_{t_0,2}$ and the corresponding sampled $\mathrm{Y}_{t_0,1}$ is calculated as sampled $v_{t_0,1}$ plus the trend.

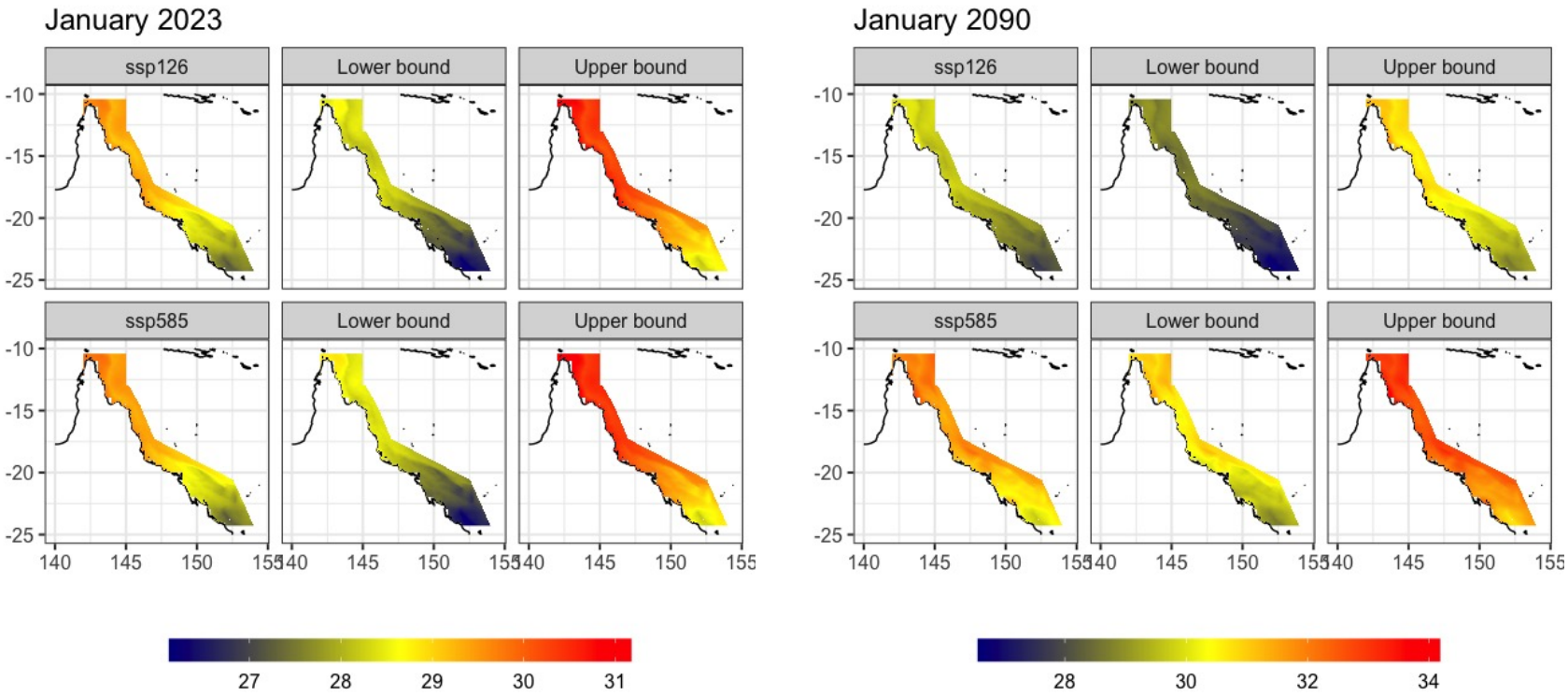


Figure:

Downscaled SSTs for January 2023 (left) and January 2090 (right) from two scenarios.

Discussion

- Let's view downscaling from the perspective of function approximation and UQ in computer experiments:
  - **Output**: SST $Y_{t,1}(\mathrm{s})$
  - **Input**: Climate model output $Y_{t,2}(\mathrm{s})$
  - Possible ways to construct and predict the response surface:
    - GP regression for $Y_{t,1}(\mathrm{s})$ on $Y_{t,2}(\mathrm{s})$?
    - GP regression for $Y_{t,1}(\mathrm{s})$ by considering $Y_{t,2}(\mathrm{s})$ and spatial location s as input together?
    - GP regression for $Y_{t,1}(\mathrm{s})$ by considering SST at $Y_{t,2}(\mathbf{u})$ with $\mathbf{u}$ within certain distance with s?
    - Allowing the GP to be nonstationary and nonseparable?
    - Constructing "deep" models and function-on-function learning?

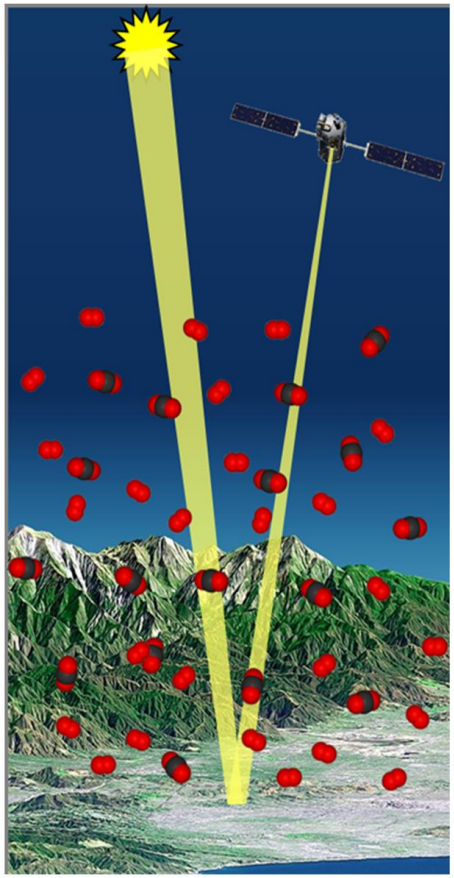
Discussion, cont'd

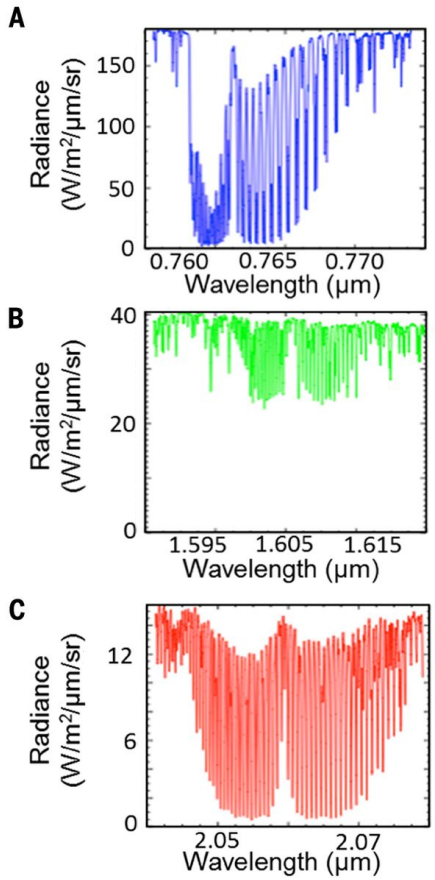
- Recall that there are two main types of downscaling
  - **Dynamical downscaling**: Using high-resolution regional simulations, also called regional climate models (RCMs), with initial and boundary conditions provided a GCM.
  - **Statistical downscaling**: Establishing a statistical relationship to transfer coarse-resolution GCM outputs into fine-resolution outputs.
  - Dynamical downscaling is physics-informed while statistical downscaling is data-driven, but dynamical downscaling is computationally intensive.

- Following the "borrowing strength" idea from spatial statistics, how about using both **dynamical downscaling** and **statistical downscaling**?
  - Understand where statistical downscaling perform well
  - Use dynamical downscaling in certain input subspace (e.g., high errors and high uncertainty from statistical downscaling)
  - Increase synergies $\rightarrow$ physics-informed learning:
    - Fusing dynamical downscaling into statistical downscaling to improve overall downscaling results
    - Leveraging the knowledge from statistical downscaling to parameterize the regional climate model (i.e., constructing reduced-order models) to make dynamical downscaling more efficient

Motivation: Orbiting Carbon Observatory-2 (OCO-2)

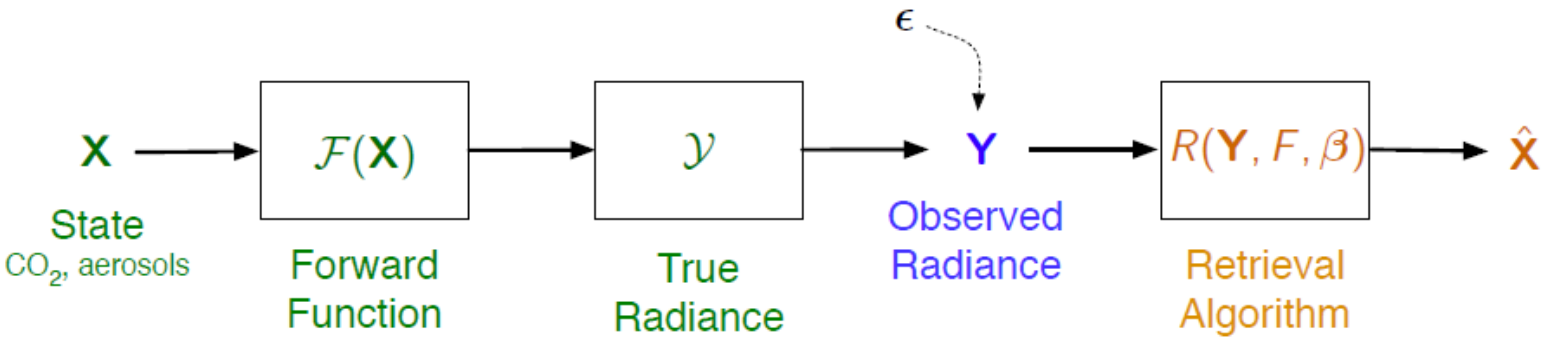
- Orbiting Carbon Observatory-2 (OCO-2) is NASA's first dedicated Earth remote sensing satellite to provide data products for atmospheric carbon dioxide.





(Credit: Eldering et al., 2017)

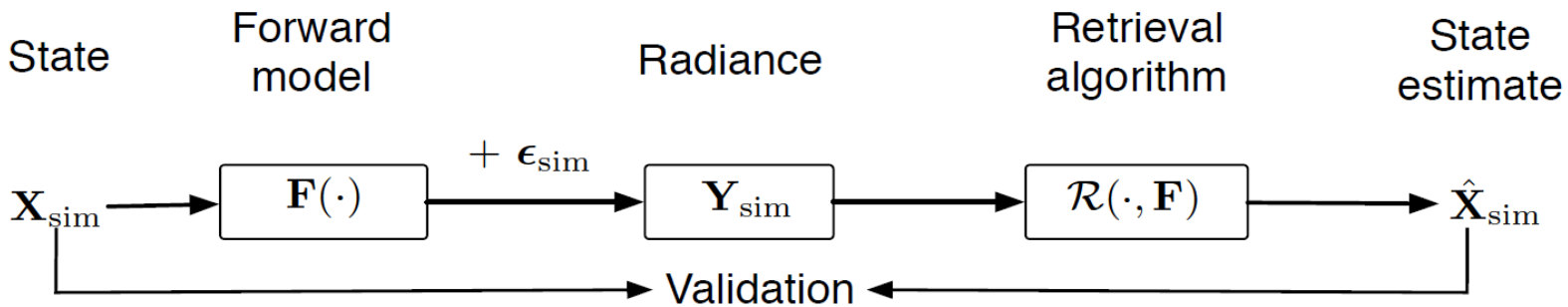
Orbiting Carbon Observatory-2 (OCO-2)



- The OCO-2 remote sensing observing system is a complex data-generating process with several key components.
  - The radiance Y is viewed as a forward function of geophysical variables, called the state vector x.
  - Instrument measures radiance, and the retrieval algorithm produces an estimate of the state vector, including CO_2 vertical profile.
- The forward model $F(x)$ is a complex simulator used to relate input variables (x), including CO_2 vertical profile, to thousands of radiances (Y) at irregular wavelengths.

Uncertainty Quantification in Remote Sensing

- Probabilistic assessment and uncertainty quantification of remote sensing retrievals are crucial to the success of using these remote sensing data to reveal valid scientific findings and to answer scientific hypotheses appropriately.
- Hobbs et al. (2017) and Turmon and Braverman et al. (2021) suggest using simulation-based experiments, known as **observing system uncertainty experiments (OSUEs)**, to quantify various sources of uncertainty in probabilistic assessment of remote sensing retrievals.
  - Such experiments are supposed to be cost-effective as they are based on simulations.



- Hobbs et al. (2017) develop an OSUE to characterize the uncertainty associated with XCO_2 under different combinations of geophysical conditions and algorithm choices. The experiment proceeds with 4 steps:
  1. simulating a large random sample of state vectors x
  2. evaluating the forward model at each of the simulated state vectors, and adding random errors to generate the radiances Y
  3. performing the retrieval algorithm to yield retrievals $\hat{x}$
  4. retrieval error analysis XCO_2 vs. $\widehat{XCO_2}$

- The computational cost of the forward model $F(\cdot)$ makes such an experiment computationally expensive:
  - The forward model $F(\cdot)$ consists of a solar model, atmospheric model, surface model, radiative transfer (RF) model, and instrument model.
  - The computation associated with the RF model is usually the rate-limiting step in many remote sensing applications, including OCO-2.
- Therefore, we would like to build an emulator $\hat{F}(\cdot)$, which does not comprise much on accuracy while enhancing computational efficiency.
  - We can employ this computationally more efficient emulator in the OSUEs to study uncertainty propagation under different combinations of geophysical conditions and algorithm choices.

Challenges

- We have $n = 10,849$ observations of radiances and state vectors based on which we will build a statistical emulator. The size of data can cause computational issues for a GP emulator. *[No longer a big issue, thanks to the previous work in this area]*
- Using GP to describe the input-output relationship?
  - The outputs of $F(\cdot)$ are radiances at hundreds of wavelengths from three bands, the O_2 band, the weak CO_2 band, and the strong CO_2 band.
  - $F(\cdot)$ is a nonlinear complex function with high-dimensional input state vector; x is m-dimensional with $m = 56$.
  - Directly modelling the relationship between high-dimensional outputs and inputs can be complicated, in terms of both computation and modeling.
    - Highly multivariate output + high-dimensional input

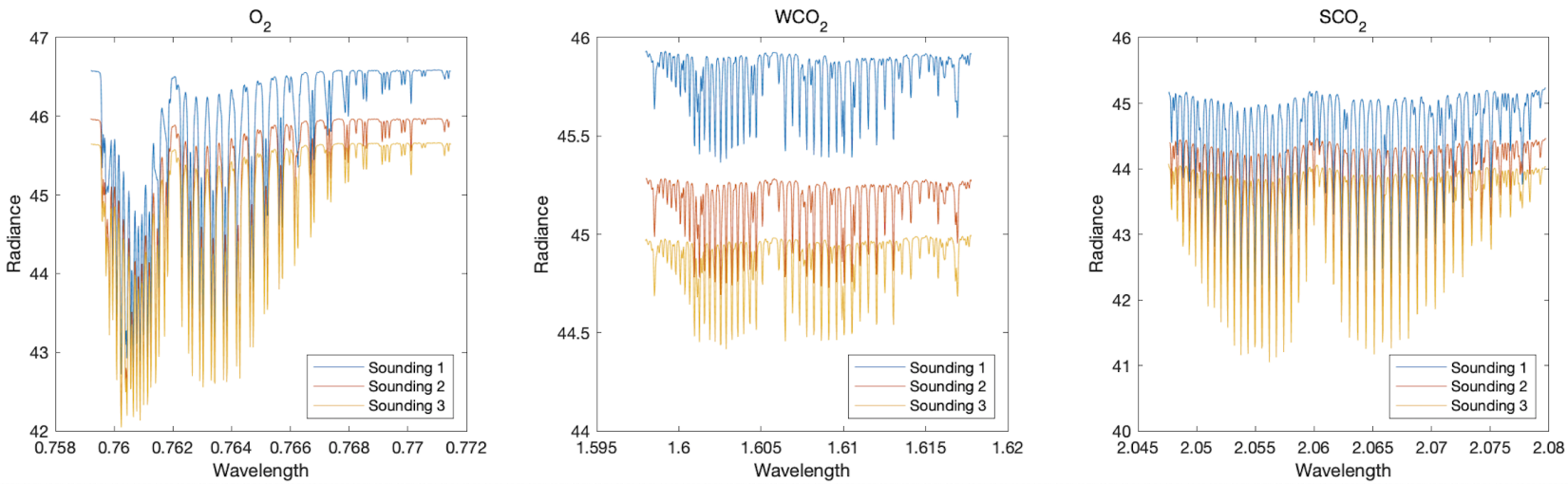
Proposed Framework

- The proposed framework consists of three steps:
  - Step 1: : Functional principal component analysis through the Principal Analysis by Conditional Estimation (PACE; Yao et al., 2005) is used to model irregularly spaced radiances and represent them via estimated low-dimensional functional principal component (FPC) scores.
  - Step 2: : The original m-dimensional input space of X is reduced to a low-dimensional space by the Gradient-based Kernel Dimension Reduction (gKDR; Fukumizu and Leng, 2014).
  - Step 3: : Construct a Gaussian process emulator using the low-dimensional pairs.

Model Set-up

- Let $Y_{ij}(w_{ij}, \mathrm{x}_i)$ be the spectrum radiance output from $F(\cdot)$ at wavelength w_{ij} from the i^{th} sounding, $i = 1, \ldots, n$; $j = 1, \ldots, n_i$. We define:

$$\mathrm{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i})'; \mathrm{x}_i = (x_{i1}, x_{i2}, \ldots, x_{im})'.$$



- Input x_i includes CO_2 profile, aerosol parameters, albedo parameters and other coefficients, altogether $m = 56$.

Functional Principal Component Analysis (FPCA)

- The radiance spectra are modeled as realizations of a random function $S(w, \mathrm{x})$ with an uncorrelated error term ϵ (mean 0 and variance σ^2).

$$Y_{ij}(w_{ij}, \mathrm{x}_i) = S(w_{ij}, \mathrm{x}_i) + \epsilon_{ij}$$

- Through Karhunen-Loéve expansion (Karhunen, 1947),

$$S(w, \mathrm{x}) = \mu(w) + \sum_{k=1}^{\infty} \xi_k(\mathrm{x})\phi_k(w)$$

- $\mu(w)$ is the mean function.
- $\xi_k(\cdot)$ is the k^{th} functional principal component (FPC) score, which are independent random variables with mean 0 and variance λ_k.
- The eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \geq \cdots \geq 0$, are corresponding to eigenfunctions $\phi_k(\omega)$.

FPCA via PACE

- The functional output $F(w, \mathrm{x})$ can be approximated by a truncated representation with leading eigenfunctions that explain the majority of variability

$$F(w, \mathrm{x}) \approx \mu(w) + \sum_{k=1}^{K} \xi_k(\mathrm{x}) \phi_k(w)$$

- We perform FPCA via the Principal Analysis by Conditional Estimation (PACE; Yao et al., 2005):
  - Pool all the sample Y_{ij}, $1 \leq i \leq n$, $1 \leq j \leq n_i$, and estimate the mean and covariance by local linear smoothing.
  - The FPC scores $\xi_k(\mathrm{x})$ are estimated through conditional expectation.
  - Asymptotic properties of the estimated eigenfunctions and FPC scores will be used to study the approximation error of the resulting GP emulator.

- We use the fraction of variance explained (FVE) to choose K.
  - With K no more than 3, we can achieve *nearly lossless representation* with at least 99.9% of variation in the radiance spectra is preserved from the three bands, respectively.
- We thus transform the original hundreds of radiances Y_i to a K-dimensional output $\boldsymbol{\xi}_i$, $i = 1, \ldots, n$.

Dimension Reduction for x

- The inputs of the forward function is an m-dimensional state vector x. In OCO-2, $m = 56$.
- An intuitive way to reduce the dimension is just to apply the principal components analysis (PCA). However, PCA only considers the correlation structure within x and ignores the role of outputs Y.
- We call $R(\mathrm{x}) \in \mathcal{R}^d$ where $d < m$ a sufficient dimension reduction if

$$p(\mathrm{Y}|\mathrm{x}) = p(\mathrm{Y}|R(\mathrm{x})),$$

where $p(\mathrm{Y}|\mathrm{x})$ and $p(\mathrm{Y}|R(\mathrm{x}))$ are conditional probability density functions with respect to x and $R(\mathrm{x})$, respectively.

Gradient-based Kernel Dimension Reduction (gKDR)

- Quite a few methods are available for dimension reduction of inputs x. We choose to use the Gradient-based Kernel Dimension Reduction (gKDR; Fukumizu and Leng, 2014)
  - The aim of gKDR is to find a projection matrix $\boldsymbol{B}$ onto a d-dimensional subspace ($d < m$) such that

$$p(\boldsymbol{\xi}|\boldsymbol{x}) = p(\boldsymbol{\xi}|\boldsymbol{B}'\boldsymbol{x})$$

  - gKDR uses positive definite kernels for nonparametric estimation of the projection matrix.
  - It does not require the gradient function of $F(\cdot)$ as in the active subspace method (**AS**; Constantine et al., 2014; Ma et al., 2021).
  - It doesn't assume any specific parametric models for $p(\boldsymbol{\xi}|\mathrm{x})$ and can be used for multivariate outputs.

gKDR, cont'd

- Choose d for dimension-reduction of x:
  - Fukumizu and Leng (2014) suggested selecting d based on the subsequent utilization of d rather than the dimension-reduction procedure when dimension reduction serves as a preprocessing step.
  - Within our proposed framework, it is intuitive to select the structural dimension d based on the predictive performance of the resulting emulator.

Gaussian Process (GP) Emulator

- After dimension reduction:

$$Y(w) \Rightarrow \boldsymbol{\xi} = (\xi_1, \ldots, \xi_K)'$$

$$\mathrm{x} \in \mathcal{R}^m \Rightarrow \mathbf{u} = \mathrm{B}'\mathrm{x} \in \mathcal{R}^d$$

- For each ξ_k, we consider a Gaussian process:

$$\xi_k(\mathbf{u}) \sim \mathcal{GP}\big(\mu(\mathbf{u}), C(\mathbf{u}, \mathbf{u}')\big)$$

  - The mean function is assumed to be zero.
  - The covariance function $C(\cdot, \cdot) : \mathcal{R}^d \times \mathcal{R}^d \to \mathcal{R}$ is assumed to be a separable squared exponential covariance function with parameters Θ plus a nugget effect τ^2.
- We implemented two methods; both are computational efficient but assume stationary vs. nonstationary covariance functions.
  - Nearest Neighbor Gaussian Processes (NNGP; Datta et al., 2016)
  - Local Approximate Gaussian Processes (LaGP; Gramacy et al., 2013)

Summary of Emulator-Construction Steps

- FPCA via PACE: For $i = 1, \ldots, n$,

$$Y_i(w, \mathrm{x}_i) \Rightarrow \mu(w); \{\phi_1(w), \ldots, \phi_K(w)\}; \boldsymbol{\xi}_i = (\xi_{i1}, \ldots, \xi_{iK})$$

- gKDR:

$$(\boldsymbol{\xi}_i; \mathrm{x}_i) \Rightarrow (\boldsymbol{\xi}_i; \mathbf{u}_i) \text{ where } \mathbf{u}_i = \mathrm{B}'\mathrm{x}$$

- GP emulator: For $k = 1, \ldots, K$,

$$\xi_k | \mathrm{B}'\mathrm{x} \sim \mathcal{GP}(\cdot, \cdot) \Rightarrow \xi_k^* | \mathrm{B}'\mathrm{x}^*$$

- Reconstructing the radiances:

$$\hat{F}(w, \mathrm{x}^*) = \mu(w) + \sum_{k=1}^{K} \xi_k^*(\mathrm{x}^*)\phi_k(w)$$

Approximation Error

- For the k^{th} FPC score, we obtain the overall approximation error bound of the GP emulator $\hat{\xi}_k$ to ξ_k:

$$\|\hat{\xi}_k(\mathbf{u}) - \xi_k(\mathrm{x})\| \leq \|\xi_k(\mathrm{x}) - \tilde{\xi}_k(\mathbf{u})\| + \|\tilde{\xi}_k(\mathbf{u}) - \hat{\xi}_k(\mathbf{u})\|$$

  - $\|\xi_k(\mathrm{x}) - \tilde{\xi}_k(\mathbf{u})\|$ is due to low-dimensional approximation (e.g., Liu and Guillas, 2017).
  - $\|\tilde{\xi}_k(\mathbf{u}) - \hat{\xi}_k(\mathbf{u})\|$ depends on the Gaussian process approximation (e.g., Wang et al., 2019)
- Utilizing the asymptotic properties of PACE, we are also able to provide finite-dimensional asymptotic simultaneous inference for $\{\hat{F}^K(w, \mathrm{x}) - F^K(w, \mathrm{x})\}$.

OCO-2 Data

- Input: $\{x_i\}$ are 56-dimensional state vector.
- Output: $Y_{ij}(w_{ij}, x_i)$ is the spectrum radiance output at wavelength w_{ij} with input x_i from the i^{th} sounding, $i = 1, \ldots, 10849$, $j = 1, \ldots, n_i$. Outputs from different soundings have varying wavelengths.
- We divide the data into training vs testing: training=9849, testing=1000.
- Our emulator-construction steps are implemented for the O_2, WCO_2, and SCO_2 bands, respectively.

PACE Results

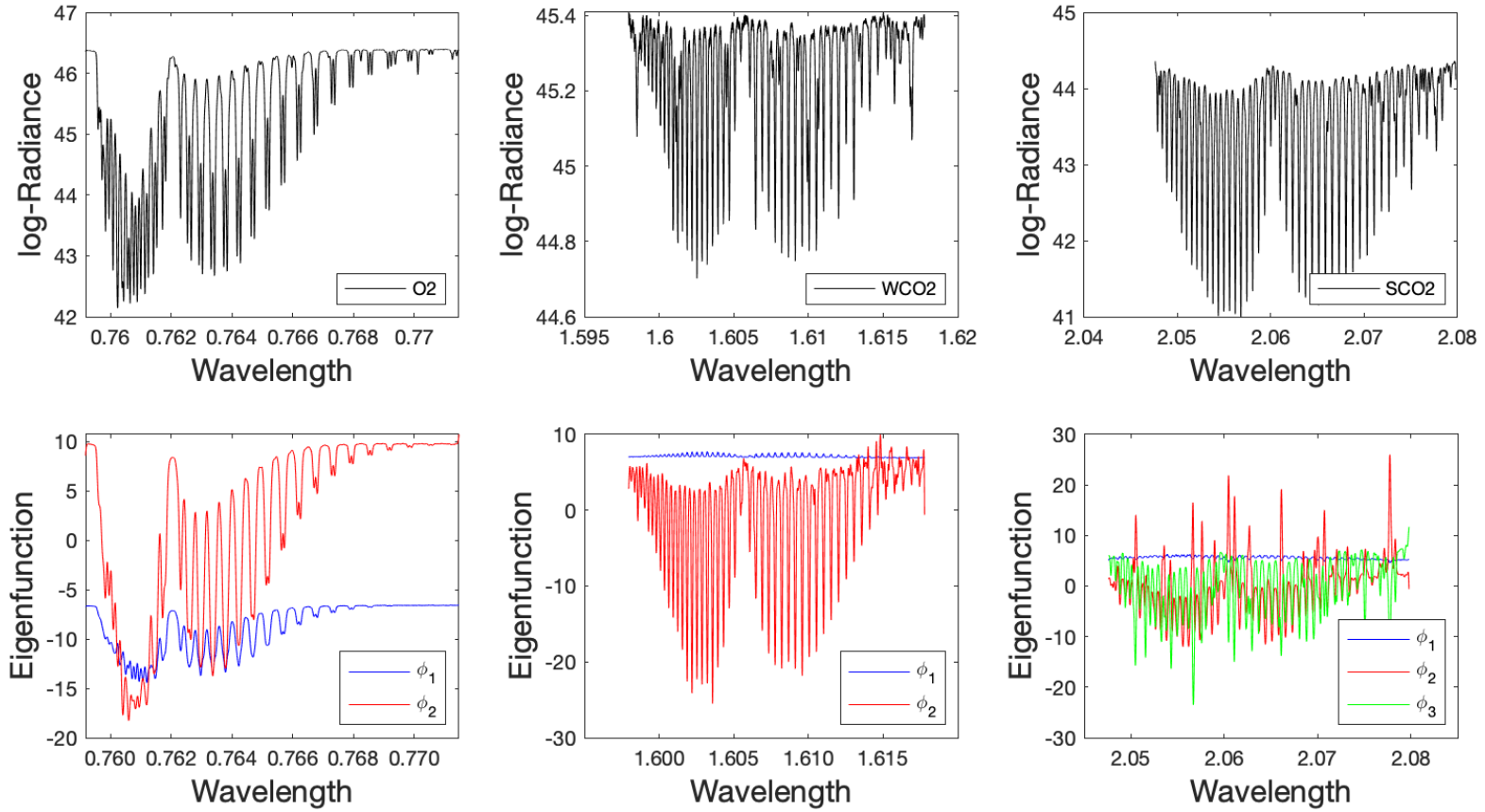


Figure: Estimated mean functions (top) and eigenfunctions (bottom) for O_2 (left), WCO_2 (middle), and SCO_2 (right), respectively.

PACE Results, cont'd

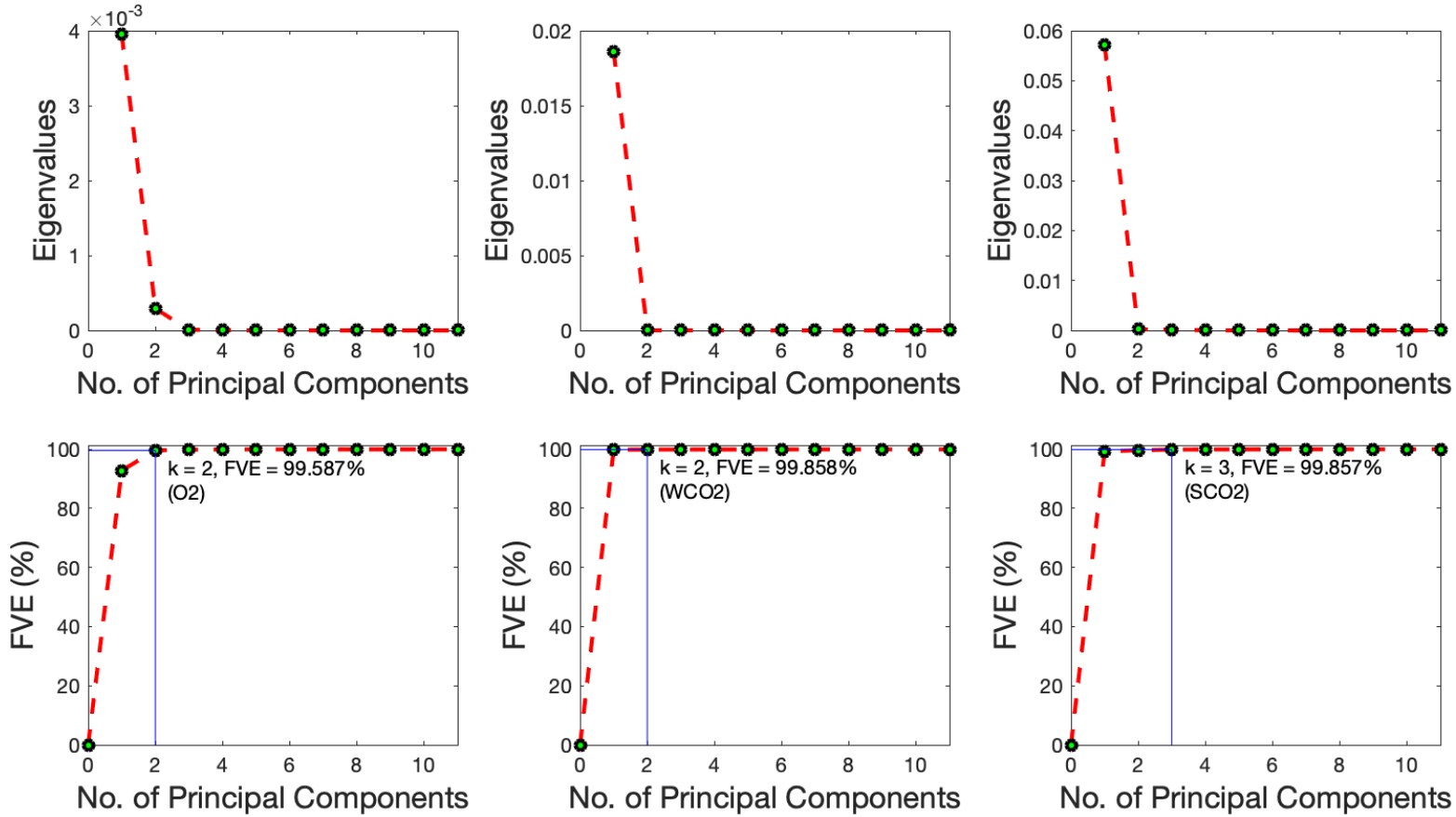


Figure: Estimated eigenvalues (top) and cumulative fraction of variance explained (FVE) (bottom) with varying K for the O_2 (left), WCO_2 (middle), and SCO_2 (right) bands, respectively.

- For $d = 2, 3, \ldots, 6$, we perform gKDR and obtain the corresponding projection matrices.
- We select d such that the resulting GP emulator gives the smallest RMSPE for FPC scores in cross validation: $d = 4, 4, 5$ for O_2, WCO_2, and SCO_2, respectively.
- We build the GP emulator with the nearest neighbor Gaussian process (NNGP) and the local approximate Gaussian process (laGP), respectively.

GP Results

- We calculate root mean squared prediction error (RMSPE), the empirical coverage probability of 95% credible/confidence interval and the continuous-rank-probability score for both NNGP and laGP for this OCO-2 data set.

		NNGP			laGP		
Band		RMSPE	$P_{CI}(95\%)$	CRPS	RMSPE	$P_{CI}(95\%)$	CRPS
O_2	FPC1	0.0050	0.994	0.0028	0.0117	0.895	0.006
	FPC2	0.0019	1	0.0018	0.0018	0.826	0.0019
WCO_2	FPC1	0.0076	0.987	0.0039	0.0114	0.922	0.0060
	FPC2	0.0054	1	0.0024	0.0083	0.934	0.0047
SCO_2	FPC1	0.0170	0.936	0.0089	0.0338	0.979	0.0195
	FPC2	0.0015	0.946	0.0008	0.0035	0.922	0.0014
	FPC3	0.0008	0.931	0.0005	0.0013	0.947	0.0009

Comparing with Active-Subspace-based Emulator (AS-E)

- We compare our emulator, referred to as gKDR-E, with the Active-Subspace-based emulator from Ma et al. (2021), referred to as AS-E.
  - The main difference is that AS-E achieves dimension reduction of inputs via the active subspace method (Constantine et al., 2014)
    - requiring the gradient function of $F(\cdot)$
    - performing dimension reduction for all the three bands together
  - In Ma et al. (2021) they show that AS-E outperforms the surrogate model in Hobbs et al. (2017), a model formulated by simplifying the relationships between geophysical variables and radiances.

- RMSPEs are calculated for each spectrum band. Our emulator outperforms AS-E.

Table: Numerical comparison of prediction performance based on AS-E and gKDR-E approaches. The predictions are obtained at $n^* = 1,000$ new input values. For each spectrum band, we compared RMSPE, $P_{CI}(95\%)$, and CRPS. The following results are reported for log-radiance.

	AS-E			gKDR-E		
	RMSPE	$P_{CI}(95\%)$	CRPS	RMSPE	$P_{CI}(95\%)$	CRPS
O_2	0.2307	0.9186	0.1255	0.0753	0.9153	0.0305
WCO_2	0.2507	0.9500	0.1332	0.0633	0.9450	0.0307
SCO_2	0.4184	0.9346	0.2254	0.1503	0.8344	0.0765

RMSPE at Different Steps

Table: Numerical comparison of prediction performance in terms of RMSPE in Step 1 (FPCA) and Steps 2-3 (Input Dimension Reduction & GP Fitting).

	AS-E			gKDR-E		
	FPCA	AS-GP	RMSPE	FPCA	gKDR-GP	RMSPE
O_2	0.0736	0.2187	0.2307	0.0570	0.0474	0.0753
WCO_2	0.0213	0.2489	0.2507	0.0215	0.0596	0.0633
SCO_2	0.1272	0.3986	0.4184	0.1246	0.0841	0.1503

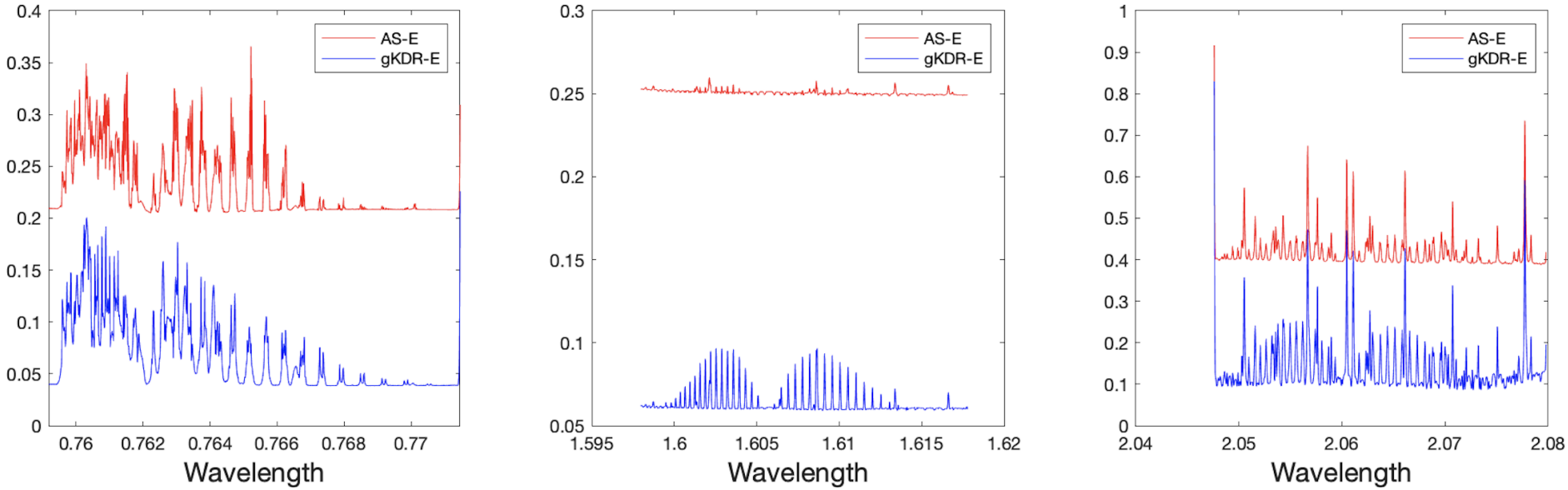


Figure: Comparison of pointwise RMSPE of predicted log-radiance based on gKDR-E and AS-E, respectively, for O_2(left), WCO_2(middle), and SCO_2(right). The x-axis represents wavelength and the y-axis represents the RMSPE averaged over $1,000$ held-out samples.

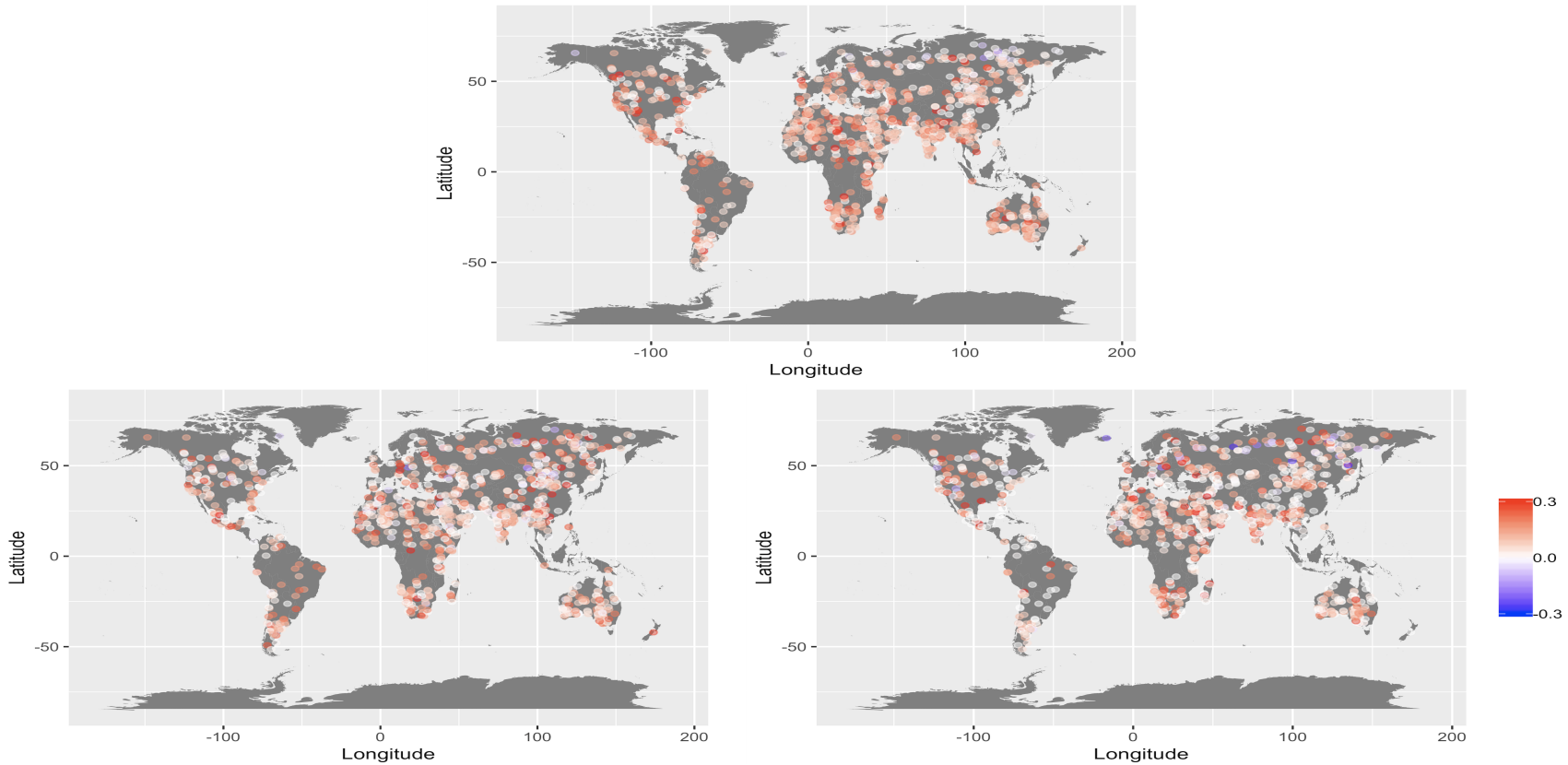


Figure: Maps of $RMSPE_{\text{AS-E}} - RMSPE_{\text{gKDR-E}}$ at the corresponding locations for the 1000 held-out samples for O_2 band (top), WCO_2 band (bottom left) and SCO_2 band (bottom right), respectively.

Discussion

- Can the emulator be further developed to improve the retrieval algorithm, i.e., solving the inverse problem?
- For example, using an emulator in the calibration-emulation-sampling (CES) approach to UQ for inverse problems:
  - Cleary et al. Calibrate, Emulate, Sample, *Journal of Computational Physics*, Vol. 424, 2021,109716.
  - Lan et al. Scaling Up Bayesian Uncertainty Quantification for Inverse Problems Using Deep Neural Networks, *SIAM/ASA J. Uncertainty Quantification*, Vol. 10, 2022, 1684-1713.
- Hybrid approaches can be used in these components:
  - Calibration/Simulation: conditional simulation, ensemble Kalman filtering, etc.
  - Emulation: GP, convolutional neural networks (CNN), etc.
  - Sampling: Metropolis-Hastings algorithms, Hamiltonian Monte Carlo, normalizing flow, etc.

Additional Remarks

- There are some common challenges:
  - curse of dimensionality: univariate $\rightarrow$ multivariate $\rightarrow$ highly-multivariate $\rightarrow$ infinite-dimensional (function)
  - scalable for large data
  - theoretical justifications
- Dive in and being open-minded: The increased diversity of viewpoints and approaches will empower UQ in scientific studies and benefit us to build an impactful community.

Acknowledgements

- This work is partially funded by, NASA/JPL, NSF (DMS-2053668), Simons Foundation's Collaboration Award (#317298 and #712755), University Research Council (URC), and TAFT research center at University of Cincinnati.

Thank you!