

A GEOSPATIAL FUNCTIONAL MODEL FOR OCO-2 DATA WITH APPLICATIONS ON IMPUTATION AND LAND FRACTION ESTIMATION

BY XINYUE CHANG^{*}, ZHENGYUAN ZHU^{*} AND JONATHAN HOBBS[†]

Iowa State University^{} and Jet Propulsion Laboratory, California Institute
of Technology[†]*

As NASA’s first dedicated CO₂ monitoring satellite, the Orbiting Carbon Observatory-2 (OCO-2) aims to provide a comprehensive measurement network which is essential to any carbon management strategy. The satellite data processing includes a retrieval algorithm for estimating CO₂ concentration from high-resolution spectra of reflected sunlight. However, due to a large amount of missing radiance observations, the spatial coverage of the retrieval algorithm is limited in areas of critical importance for carbon cycle science. Also, mixed land/water pixels are not used in retrieval processing. Here we propose an approach to model spectral spatial data such that radiance imputation and land fraction estimation can be tackled well. The spectral observation is modeled as functional data across geolocations in a homogeneous area of interest and can be reduced to much lower dimensions by FPCA techniques. Based on specific features of the OCO-2 instrument, the model considers separate mean functions and measurement error variance models for different spatial footprints. Then principal component scores are treated as a random field and are predicted by ordinary kriging. The proposed method is validated to impute spectral radiance with high accuracy when tested with observations over the Pacific Ocean. The unmixing approach based on our model shows the potential to obtain much more accurate land fraction estimates in the case study of Greece coastlines.

1. Introduction. Satellite remote sensing data continue to provide information on many processes in the Earth system. Geophysical quantities of interest are inferred from the radiance spectra directly observed by remote sensing instruments. A growing constellation of satellites are providing estimates of greenhouse gas concentrations globally at fine spatial resolution. Estimates of atmospheric carbon dioxide (CO₂) concentration from NASA’s Orbiting Carbon Observatory-2 (OCO-2) are providing information on the carbon cycle at global and regional scales (Eldering et al., 2017a). Several data-processing and inference stages are executed in translating the observed

Keywords and phrases: Functional Principal Component Analysis, Ordinary Kriging, Remote Sensing Data, Spectral Unmixing

satellite radiances, termed Level 1 data products, into inferences on carbon sources and sinks (Cressie, 2018). The *retrieval algorithm* implements the estimation of CO_2 concentration from Level 1 data (O'Dell et al., 2018). For OCO-2, the primary retrieval output, or Level 2 product, of interest is X_{CO_2} , which is the average concentration of carbon dioxide in a column of dry air extending from Earth's surface to the top of the atmosphere. The OCO-2 instrument observes high-resolution spectra of reflected sunlight at wavelengths (colors) with three spectrometers each focused in a narrow spectral band of the near infrared portion of the electromagnetic spectrum. The O_2 A-band covers wavelengths with substantial absorption of oxygen and the weak CO_2 and strong CO_2 bands include spectral ranges with carbon dioxide absorption. However, OCO-2 is also sensitive to other atmospheric and surface properties, including clouds and land-ocean transitions. These challenges result in a significant amount of locations with unusable data for the retrieval. Retrieval spatial coverage could improve if we are able to impute spectral observations for missing locations. For a complicated and massive data product like OCO-2, additional ancillary data is often needed and is subject to error. This includes the land fraction estimate used as an input into the retrieval. Based on the imputation algorithm proposed in this work, an unmixing approach has the potential to obtain much more accurate land fraction estimation only using measured radiance and geolocation information.

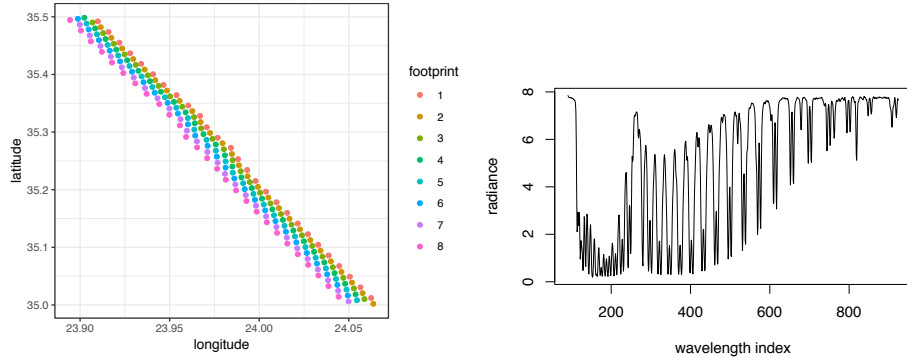
Functional data analysis (FDA) has benefited from theoretical and methodological advances in recent decades, with applications including longitudinal data, image data, geospatial problems. Functional principal component analysis (FPCA) is a widely used tool in FDA. Estimation and inference for FPCA is often characterized by how densely the function of interest is observed, i.e. a sparse scenario (Yao, Müller and Wang, 2005) as opposed to a dense scenario (Ramsay, 2004). OCO-2 represents an instance of dense functional data with each case yielding radiances at hundreds of wavelengths and many observations across space and time. The OCO-2 data present some specific challenges. Each observation location on the ground is associated with a specific location, known as a *footprint*, on the physical instrument, and measurement error properties vary with footprint. Although literature dealing with dense functional data involves measurement error estimation (Castro, Lawton and Sylvestre (1986), Yao et al. (2003)), these approaches need to be adapted to the specific case concerned about in this paper. By the nature of atmosphere and global interactions, functional observations in different locations cannot be assumed as independent, especially for locations nearby. These so-called spatially dependent functional data have gained much at-

tention recently (Liu, Ray and Hooker (2017), Li et al. (2007), Zhang et al. (2016)). For real applications, spatially dependent spectral data have not been well studied, especially for remote sensing data. More importantly, integrating different characteristics from the unique footprints together becomes crucial to the analysis of OCO-2 data. We follow the framework where principal component scores contain spatial dependence information, but the mean function is modeled with a linear structure depending on locations and footprints. This results in a unified model that allows for footprint-specific parameters and that includes both spatially dependent and spatially independent cases.

The rest of the paper is organized as follows. We introduce the structure of OCO-2 data and variables used in Section 2. Then we propose a geospatial functional model for spatial spectral data with different characteristic on different footprints in Section 3. In Section 4, we discuss the estimation and prediction based on our data model for the purpose of radiance imputation in water area and land fraction correction in mixed regions. Lastly in Section 5, the radiance imputation algorithm and land fraction estimation procedure were applied to OCO-2 Level 1 data such that spatial coverage and accuracy of retrieval algorithm can be improved.

2. OCO-2 Data. OCO-2 is part of a constellation of polar-orbiting satellites known as the A-train and completes approximately 15 orbits per day. The satellite crosses the equator in the early afternoon local time on each orbit. Orbits alternate between nadir and glint observing modes. In nadir mode, data are collected directly below the satellite, minimizing the optical path length through the atmosphere. In glint mode, the instrument points at an angle directed toward the glint spot, allowing high signal over the ocean (Eldering et al., 2017b). Our analysis focuses on glint observations, which are available over both land and ocean, including along coastlines. The OCO-2 field of view is approximately 10 km wide along an orbit track, and this spatial orientation translates to physical positions on the focal plane arrays (FPAs) for each of the spectrometers on the instrument. In order to meet bandwidth limitations for storage and downlink from the satellite, the across-track spectra are aggregated into eight discrete *footprints*, as shown in Fig 1a. Because the footprints correspond to different physical locations on the instrument, they are characterized separately in OCO-2 data processing (Crisp et al., 2017).

We use OCO-2 Level 1 products, which include latitude, longitude, orbit, footprint, and land fractions for locations and times of interest. Each unique location (corresponding to one footprint) and time defines a single



(a) spatial layout of 8 cross-track footprints from orbit 10575 (b) radiance function of O₂ band at (34.59285, 24.18569) on orbit 10575

Fig 1: OCO-2 Level 1 data

observation, or *sounding*. Level 1 data also include the wavelengths and measured radiances for each sounding. We treat the measured radiance as a function of wavelength, and wavelengths and radiances are categorized into 3 bands as introduced above. For simplicity here, we will only consider radiances in the O₂ band (Fig 1b), which is the first 1016 components of the 3048-dimensional radiance vector before filtering. The O₂ band refers to wavelengths (colors) within the 0.765 micron molecular oxygen A-band. The spectral indices, $j = 1, \dots, 1016$ correspond to physical positions on the OCO-2 spectrometer, and each position is termed as *spectral sample* in OCO-2 documentation. The wavelength corresponding to each sample varies slightly across soundings in space and time, but the within-sample variability is small compared to the monotonic change across samples with increasing j . Therefore we focus our analysis on the discrete sample index. Finally, through the following sections, measured radiance, reported in units of ($\text{photons m}^{-2}\text{sr}^{-1}\mu\text{m}^{-1}$), was divided by 10^{19} without loss of generality.

Due to the fact that eight footprints effectively constitute eight instruments, we found that 2nd order differencing estimates of measurement error (eqn. 4.4) is a function of wavelength and depends on footprints. Furthermore, it is proportional to the naive mean radiance estimate (averaging observations over locations) within corresponding footprint group, as shown in Fig 2. Besides that, as demonstrated in Fig 3, the mean radiance varies across locations and shifts among footprints.

The effect of this footprint issue not only impacts measurement error estimation in functional principal component analysis, but also results in

the footprint-dependent mean functions in functional modeling. More details will be included in the following two sections.

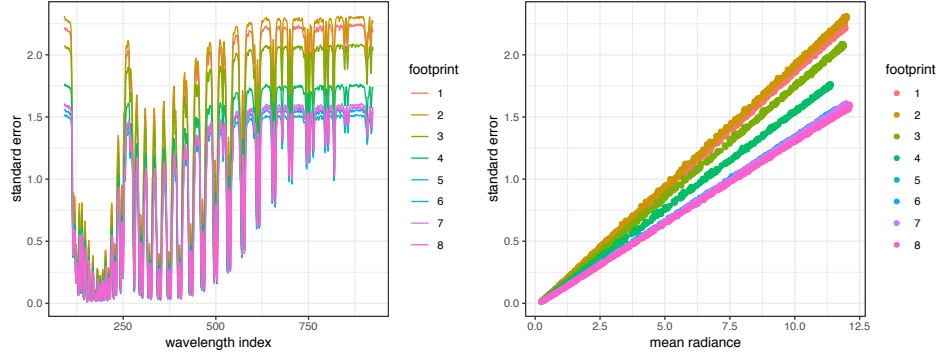


Fig 2: Measurement error estimates by footprint in region within latitude $[34.3, 34.8]$ on orbit 10575

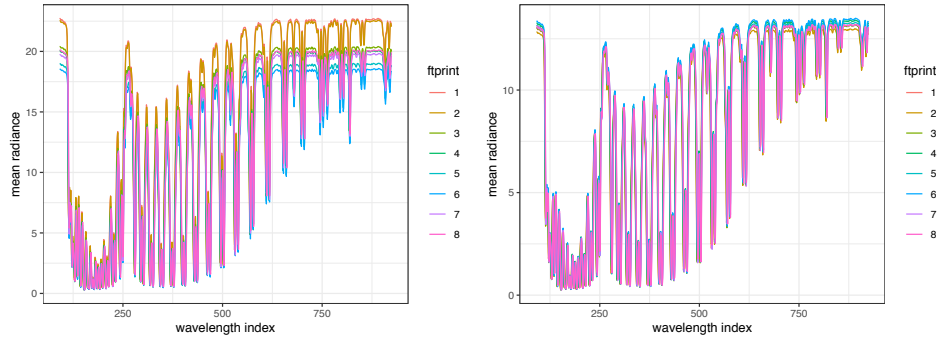


Fig 3: Mean radiance by footprint in regions within latitude $[34, 34.5]$ and $[35, 35.5]$ on orbit 10575

3. Geospatial Functional Model based on FPCA.

3.1. *Functional Data.* For the O_2 band we are interested in, the radiance $f(\cdot; \mathbf{s}_i)$ in a sounding location $\mathbf{s}_i = (L_i, l_i) = (\text{latitude}, \text{longitude})$ is treated as a function of m_i wavelengths, i.e., function in wavelength index w_j , $1 \leq w_j \leq 1016$. Here \mathbf{s}_i is ordered by footprint 1 to 8, sounding time, and in the area of interest $\mathcal{S} = \{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ which is not identified as a sphere. We model measured radiance as

$$(3.1) \quad r(w_j; \mathbf{s}_i) = f(w_j; \mathbf{s}_i) + \epsilon_{i,j}, \quad i = 1, \dots, n, \quad j = 1, \dots, m_i$$

where we assume error $\epsilon_{i,j}$ is independent of $f(w_j, \mathbf{s}_i)$ with mean zero and variance $\sigma_{p_i}^2(w_j)$, and p_i is the footprint corresponding to \mathbf{s}_i . Because of the real data characteristic, we model the mean function to be dependent on location \mathbf{s}_i . And in classic functional data analysis, we assume $f(w_j; \mathbf{s}_i)$ has a standard Karhunen-Loève expansion,

$$(3.2) \quad f(w_j; \mathbf{s}_i) = \mu(w_j; \mathbf{s}_i) + \sum_{k=1}^{\infty} \xi_k(\mathbf{s}_i) \phi_k(w_j)$$

where the eigenfunctions $\phi_k(\cdot)$ are an orthonormal basis in Hilbert space and the principal component scores $\xi_k(\mathbf{s}_i)$ are random variables with variance λ_k . The PC scores depend on \mathbf{s}_i and are assumed to be mean zero, second order stationary and isotropic random fields. For any two points \mathbf{s}_i and $\mathbf{s}_{i'}$, we assume

$$(3.3) \quad \text{Cov}\{\xi_k(\mathbf{s}_i), \xi_k(\mathbf{s}_{i'})\} = G_k(\|\mathbf{s}_i - \mathbf{s}_{i'}\|)$$

where $\|\cdot\|$ denotes the great circle distance between two geolocations. Note that we assume $\xi_k(\mathbf{s}_i)$ does not differ among footprints, which is supported by the empirical evidence. Models assuming the footprint-specific $\xi_k^p(\mathbf{s}_i)$ does not improve the prediction performance. In practice, the first few principal components capture most of the variation. Thus we consider the truncated version with $K < \infty$,

$$(3.4) \quad f(w_j; \mathbf{s}_i) = \mu(w_j; \mathbf{s}_i) + \sum_{k=1}^K \xi_k(\mathbf{s}_i) \phi_k(w_j).$$

Inspired by the empirical finding that measured radiance varies across different footprints and can be described by a linear relationship with location covariates, the mean function is modeled as a linear model with footprint specific coefficients for each fixed wavelength index,

$$(3.5) \quad \mu(w_j; \mathbf{s}_i) = \beta_0^{p_i}(w_j) + (\mathbf{s}_i - \bar{\mathbf{s}}_{p_i}) \boldsymbol{\beta}_1^{p_i}(w_j)$$

where $\bar{\mathbf{s}}_p = \frac{1}{n_p} \sum_{\mathbf{s}_i \in \mathcal{S}_p} \mathbf{s}_i$, $\mathcal{S}_p = \{\mathbf{s}_i : p_i = p\}$ and $n_p = |\mathcal{S}_p|$. And the general covariance function is given by

$$(3.6) \quad \begin{aligned} R(w_j, w_{j'}, \|\mathbf{s}_i - \mathbf{s}_{i'}\|) &\equiv \text{Cov}\{f(w_j; \mathbf{s}_i), f(w_{j'}; \mathbf{s}_{i'})\} \\ &= \sum_{k=1}^K G_k(\|\mathbf{s}_i - \mathbf{s}_{i'}\|) \phi_k(w_j) \phi_k(w_{j'}). \end{aligned}$$

So $f(w_j; \mathbf{s}_i)$ is assumed to be a realization of a stationary spectral process in our study region \mathcal{S} (basically $\pm 0.5^\circ$ in latitude). It is assumed to have non-homogeneous mean functions across spatial locations, but residuals that are well approximated by the projection on the function space spanned by the first K eigenfunctions.

3.2. Mixing Process of Water and Land. Let \mathcal{M} be a mixed area such as a coastline crossing, where land fractions are between 0 and 1. In general, physical characteristic of mixed locations could be influenced by both water and land areas. It is natural to model the observed radiance at mixed sounding location $\mathbf{s}_i \in \mathcal{M}$ as a linear combination of possible water radiance $f_w(w_j, \mathbf{s}_i)$ and land radiance $f_l(w_j, \mathbf{s}_i)$ like following,

$$(3.7) \quad r(w_j; \mathbf{s}_i) = \alpha_i f_l(w_j; \mathbf{s}_i) + (1 - \alpha_i) f_w(w_j; \mathbf{s}_i) + \epsilon_{i,j}$$

where α_i is the land fraction of the mixed location \mathbf{s}_i and $\epsilon_{i,j}$ are independent with mean zero and variance $\sigma_{p_i}^2(w_j)$. Note the variance function $\sigma_{p_i}^2(w_j)$ may not be the same as that of (3.1) since it corresponds to a mixed region \mathcal{M} .

The mixing model is actually the same as model (3.1) above except that $f(w_j; \mathbf{s}_i)$ is written as a linear combination of two underlying radiance extended from the nearby water and land processes. Then $f_l(w_j; \mathbf{s}_i)$ and $f_w(w_j; \mathbf{s}_i)$ follow the same structure with different mean and covariance functions.

$$(3.8) \quad f_l(w_j; \mathbf{s}_i) = \mu_l(w_j; \mathbf{s}_i) + \sum_{k=1}^K \xi_k^l(\mathbf{s}_i) \phi_k^l(w_j)$$

$$(3.9) \quad f_w(w_j; \mathbf{s}_i) = \mu_w(w_j; \mathbf{s}_i) + \sum_{k=1}^K \xi_k^w(\mathbf{s}_i) \phi_k^w(w_j)$$

4. Estimation.

4.1. Dense FPCA. In practice, some of the OCO-2 spectral channels, particularly for the largest and smallest wavelengths, do not produce scientifically reliable radiances. These “bad samples” are flagged and produce systematic patterns of missing radiances in the data product (Eldering et al., 2017c). Then let \mathcal{W} be the wavelength indices available in the area of interest, usually the collection of wavelengths corresponding to radiances not all missing in \mathcal{S} . We assume the trajectories of $r(w_j; \mathbf{s}_i)$ are fully observed on \mathcal{W} . Let \mathbf{x}_p be the matrix of location covariates in the footprint-specific set \mathcal{S}_p , i.e., $\mathbf{x}_p = \{(\mathbf{s}_1 - \bar{\mathbf{s}}_p)^T, \dots, (\mathbf{s}_{n_p} - \bar{\mathbf{s}}_p)^T\}^T$, and hence $\{(\mathbf{s}_1 - \bar{\mathbf{s}}_p)^T, \dots, (\mathbf{s}_n - \bar{\mathbf{s}}_p)^T\}^T = (\mathbf{x}_1^T, \dots, \mathbf{x}_8^T)^T$. Similarly, measured

radiance at wavelength index w_j is $\mathbf{Y}_j = \{r(w_j; \mathbf{s}_1), \dots, r(w_j; \mathbf{s}_n)\}^T = (\mathbf{Y}_{1j}^T, \dots, \mathbf{Y}_{8j}^T)^T$ where \mathbf{Y}_{pj} is the vector of measured radiance $r(w_j; \mathbf{s}_i)$ with $\mathbf{s}_i \in \mathcal{S}_p$. So the full design matrix \mathbf{X} corresponding to the mean function model (3.5) is constructed as

$$\mathbf{X} = \begin{bmatrix} \mathbf{1} & \mathbf{x}_1 & & & & & & \\ & & \mathbf{1} & \mathbf{x}_2 & & & & \\ & & & \ddots & \ddots & & & \\ & & & & & \mathbf{1} & \mathbf{x}_8 & \end{bmatrix}.$$

The footprint-specific coefficients $\boldsymbol{\beta}(w_j) = \{\beta_0^1(w_j), \boldsymbol{\beta}_1^1(w_j)^T, \dots, \beta_0^8(w_j), \boldsymbol{\beta}_1^8(w_j)^T\}^T$ for wavelength index w_j can be estimated by

$$(4.1) \quad \hat{\boldsymbol{\beta}}(w_j) = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}_j$$

with the consistency stated as below. Note $\|\cdot\|_2$ is the ℓ_2 norm. So we estimate the location-dependent mean function as

$$(4.2) \quad \hat{\mu}(w_j; \mathbf{s}_i) = \hat{\beta}_0^{p_i}(w_j) + (\mathbf{s}_i - \bar{\mathbf{s}}_{p_i}) \hat{\boldsymbol{\beta}}_1^{p_i}(w_j).$$

THEOREM 4.1. *Under assumptions 1, 2, 3 and 4 in specified Appendix A, for any $\mathbf{s}_i \in \mathcal{S}$,*

$$\begin{aligned} \sup_{w_j \in \mathcal{W}} \|\hat{\boldsymbol{\beta}}(w_j) - \boldsymbol{\beta}(w_j)\|_2 &= O_p(\sqrt{n^{-1} + n^{-\alpha\delta}}), \\ \sup_{w_j \in \mathcal{W}} |\hat{\mu}(w_j; \mathbf{s}_i) - \mu(w_j; \mathbf{s}_i)| &= O_p(\sqrt{n^{-1} + n^{-\alpha\delta}}) \end{aligned}$$

for some $0 < \alpha < 1/2$ and $\delta > 0$.

Based on the general covariance function (3.6) and second order stationary assumption, it follows that the spectral covariance function for any location \mathbf{s}_i is

$$(4.3) \quad R_w(w_j, w_{j'}) \equiv R(w_j, w_{j'}, 0) = \sum_{k=1}^K \lambda_k \phi_k(w_j) \phi_k(w_{j'}).$$

Then we estimate spectral covariance function as

$$\begin{aligned} &\hat{R}_w(w_j, w_{j'}) \\ &= \frac{\sum_{i=1}^n \{r(w_j; \mathbf{s}_i) - \hat{\mu}(w_j; \mathbf{s}_i)\} \{r(w_{j'}; \mathbf{s}_i) - \hat{\mu}(w_{j'}; \mathbf{s}_i)\}}{n-1} - \hat{\sigma}^2(w_j) I(j = j') \end{aligned}$$

where the measurement error variance is estimated by the average of second order differencing as following.

$$(4.4) \quad \hat{\sigma}_p^2(w_j) = \frac{1}{6(n_p - 2)} \sum_{\mathbf{s}_i \in \mathcal{S}_p} \{r(w_j; \mathbf{s}_{i+2}) - 2r(w_j; \mathbf{s}_{i+1}) + r(w_j; \mathbf{s}_i)\}^2$$

$$(4.5) \quad \hat{\sigma}^2(w_j) = \frac{1}{n-1} \sum_{p=1}^8 n_p \hat{\sigma}_p^2(w_j)$$

The next theorem shows that the $\hat{R}_w(w_j, w_{j'})$ holds the same asymptotic property as the mean function under similar regularity conditions.

THEOREM 4.2. *Under assumptions 1, 2, 3, 4 and 5 specified in Appendix A,*

$$\sup_{w_j, w_{j'} \in \mathcal{W}} |\hat{R}_w(w_j, w_{j'}) - R_w(w_j, w_{j'})| = O_p(n^{-\alpha\delta/2} + n^{2\beta(\alpha-1/2)} + n^{\alpha-1/2})$$

for some $0 < \alpha < 1/2$, $\delta > 0$ and $\beta > 0$.

Then eigenfunctions can be obtained by discretizing the covariannce estimation (Rice and Silverman, 1991) and matrix decomposition with respect to wavelength indices,

$$(4.6) \quad \hat{R}_w(w_j, w_{j'}) = \sum_{k=1}^K \hat{\lambda}_k \hat{\phi}_k(w_j) \hat{\phi}_k(w_{j'})$$

The number of principal components K is selected as the truncation to which point certain variance can be explained (usually 99%). Although the radiance looks like discrete, it is actually assumed to be a function of only fixed wavelengths. So the corresponding $\xi_k(\mathbf{s}_i)$ can be estimated by numerical integration,

$$(4.7) \quad \hat{\xi}_k(\mathbf{s}_i) = \int \{r(w; \mathbf{s}_i) - \hat{\mu}(w; \mathbf{s}_i)\} \hat{\phi}_k(w) dw.$$

By theorem 4.2, we have the following result with respect to obtaining principal component scores through dense functional principal component analysis.

THEOREM 4.3. *Under assumptions 1, 2, 3, 4 and 5 specified in Appendix A, for any $\mathbf{s}_i \in \mathcal{S}$,*

$$\begin{aligned} \sup_{w_j \in \mathcal{W}} |\hat{\phi}_k(w_j) - \phi_k(w_j)| &= O_p(n^{-\alpha\delta/2} + n^{2\beta(\alpha-1/2)} + n^{\alpha-1/2}) \\ |\hat{\xi}_k(\mathbf{s}_i) - \xi_k(\mathbf{s}_i)| &= O_p(n^{-\alpha\delta/2} + n^{2\beta(\alpha-1/2)} + n^{\alpha-1/2}). \end{aligned}$$

for some $0 < \alpha < 1/2$, $\delta > 0$ and $\beta > 0$.

4.2. BLUP for Principal Component Scores. Suppose $\mathbf{s}_0 = (L_0, l_0)$ is a sounding location with no measured radiance observed. The variable $\xi_k(\mathbf{s}_i)$ calculated from above forms a second order stationary and isotropic random field. Let $\mathbf{\Sigma}_k = \text{Var}(\boldsymbol{\xi}_k)$ and $\boldsymbol{\sigma}_k = \text{Cov}\{\boldsymbol{\xi}_k(\mathbf{s}_0), \boldsymbol{\xi}_k\}$, which can be constructed by the covariance function $G_k(\cdot)$ estimated through fitting a covariance model to the semivariogram of $\boldsymbol{\xi}_k$ by weighted least squares. In this application, we use an exponential covariance model. Note that $\mathbf{\Sigma}_k$ is a diagonal block matrix if assuming independence among footprints, and $\boldsymbol{\sigma}_k$ are zeros except for entries at footprint p_0 .

Although $\hat{\boldsymbol{\xi}}_k$ should have zero mean based on the assumption in (3.2), we use ordinary kriging in a small sample of locations in \mathcal{S} to achieve better prediction performance. So the best linear unbiased predictor for $\xi_k(\mathbf{s}_0)$ is

$$(4.8) \quad \hat{\xi}_k(\mathbf{s}_0) = \hat{\mu} + \hat{\boldsymbol{\sigma}}_k^T \hat{\mathbf{\Sigma}}_k^{-1} (\hat{\boldsymbol{\xi}}_k - \mathbf{1}\hat{\mu})$$

where $\hat{\mu} = (\mathbf{1}^T \hat{\mathbf{\Sigma}}_k^{-1} \mathbf{1})^{-1} \mathbf{1}^T \hat{\mathbf{\Sigma}}_k^{-1} \hat{\boldsymbol{\xi}}_k$. Then the radiance function for the location \mathbf{s}_0 can be predicted as

$$(4.9) \quad \hat{f}(w_j; \mathbf{s}_0) = \hat{\mu}(w_j; \mathbf{s}_0) + \sum_{k=1}^K \hat{\xi}_k(\mathbf{s}_0) \hat{\phi}_k(w_j), \quad w_j \in \mathcal{W}$$

where K is determined by the fraction of variation explained in the eigen decomposition. To illustrate BLUP estimator $\hat{\xi}_k(\mathbf{s}_0)$ is consistent, we need to show that the parameter estimates of the semivariogram model leading to $\boldsymbol{\sigma}_k$ and $\mathbf{\Sigma}_k$ are consistent. Let $\gamma_k(h; \theta_k)$ be the semivariogram of $\xi_k(\mathbf{s}_i)$, as a function of distance h and with parameters θ_k . And $\hat{\gamma}_k(h)$ is the sample semivariogram of $\hat{\xi}_k(\mathbf{s}_i)$,

$$\hat{\gamma}_k(h_l) = \frac{1}{2N(h_l)} \sum_{\|\mathbf{s}_i - \mathbf{s}_j\| = h_l} \{\hat{\xi}_k(\mathbf{s}_i) - \hat{\xi}_k(\mathbf{s}_j)\}^2, \quad l = 1, \dots, L,$$

for a collection of L distance bins. We use weighted least squares to estimate θ_k with $V_k(\theta_k)$ as the weight matrix accordingly for the k th component. So

$$(4.10) \quad \hat{\theta}_k = \arg \min_{\theta_k} \{\hat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_k(\theta_k)\}^T V_k(\theta_k) \{\hat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_k(\theta_k)\}$$

where $\hat{\boldsymbol{\gamma}}_k = \{\hat{\gamma}_k(h_1), \dots, \hat{\gamma}_k(h_L)\}^T$ and $\boldsymbol{\gamma}_k(\theta_k) = \{\gamma_k(h_1; \theta_k), \dots, \gamma_k(h_L; \theta_k)\}^T$. So the estimator is consistent as stated in the following theorem.

THEOREM 4.4. Under conditions (C.1), (C.2), (C.3) and assumptions 1-5 specified in Appendix A,

$$\hat{\theta}_k - \theta_k \xrightarrow{p} 0 \quad \text{as } n \rightarrow \infty.$$

Remarks: We use exponential model as the choice of $\gamma_k(h; \theta_k)$, which satisfies (C.1) and (C.2) easily. For condition (C.3), a diagonal matrix using $N(h_l)/\gamma_k^2(h_l; \theta_k)$ (Cressie (1985)), $N(h_l)/h_l^2$ and $N(h_l)$ are all valid choices.

4.3. *Land Fraction Estimation.* For a mixed location \mathbf{s}_i , there exists the nearest homogeneous water area \mathcal{S}_w and land area \mathcal{S}_l . Using the imputation method above, imputed radiance extended from \mathcal{S}_w and from \mathcal{S}_l can be obtained separately, which are the underlying water and land radiance estimation $\hat{f}_w(w_j; \mathbf{s}_i)$ and $\hat{f}_l(w_j; \mathbf{s}_i)$. Assuming the mixing model 3.7, we can estimate land fraction α_i by minimizing squared loss function $\|r(w_j; \mathbf{s}_i) - \alpha \hat{f}_l(w_j; \mathbf{s}_i) - (1 - \alpha) \hat{f}_w(w_j; \mathbf{s}_i)\|_2^2$, which has the solution

$$(4.11) \quad \hat{\alpha}_i^k = \frac{\sum_{j=1}^{m_i} \{r(w_j; \mathbf{s}_i) - \hat{f}_w(w_j; \mathbf{s}_i)\} \{\hat{f}_l(w_j; \mathbf{s}_i) - \hat{f}_w(w_j; \mathbf{s}_i)\}}{\sum_{j=1}^{m_i} \{\hat{f}_l(w_j; \mathbf{s}_i) - \hat{f}_w(w_j; \mathbf{s}_i)\}^2}.$$

Although model 3.7 does not assume constant variance through different wavelengths, we use ordinary least square to estimate land fractions, because it would be impossible to estimate a suitable variance function of measurement error using limited points in mixed region. Ordinary least squares performs well in simulation and real data.

To demonstrate the value of the kriging based unmixing approach, we compare it with a simple linear interpolation method. Suppose the nearest land footprint of \mathbf{s}_i is \mathbf{s}_i^l and water footprint is \mathbf{s}_i^w , this method estimate α by minimizing $\|r(w_j; \mathbf{s}_i) - \alpha r(w_j; \mathbf{s}_i^l) - (1 - \alpha) r(w_j; \mathbf{s}_i^w)\|_2^2$, which has the solution

$$(4.12) \quad \hat{\alpha}_i^t = \frac{\sum_{j=1}^m \{r(w_j; \mathbf{s}_i) - r(w_j; \mathbf{s}_i^w)\} \{r(w_j; \mathbf{s}_i^l) - r(w_j; \mathbf{s}_i^w)\}}{\sum_{j=1}^m \{r(w_j; \mathbf{s}_i^l) - r(w_j; \mathbf{s}_i^w)\}^2}.$$

We conducted a simulation study based on OCO-2 data to illustrate the advantage of the kriging based unmixing approach over the interpolation method. For simplicity, data were only simulated on locations for a single footprint with locations equally spaced in latitude and longitude. The location in the middle (35.49881, 23.83578) is taken to be mixed with land fraction between 0 and 1. The portion south of the middle location is assumed to be water and to the north is assumed to be land. The layout similar to Fig 1a but with only one footprint and larger latitude range.

Mean functions and FPCs in water and land areas are borrowed from FPCA results of latitude between 34.7129 and 35.7523 in orbit 05216 sounded on 06-25-2015. For both land and water areas, the first principal component is assumed to be multivariate normal with covariance defined by exponential models (great circle distance h in km):

$$(4.13) \quad C_w(h) = 5 \exp(-h/10)$$

$$(4.14) \quad C_l(h) = 10 \exp(-h/7)$$

For latter principal components, we assume $\xi_2^w(\mathbf{s}_i) \stackrel{iid}{\sim} \mathcal{N}(0, 2)$, $\xi_2^l(\mathbf{s}_i) \stackrel{iid}{\sim} \mathcal{N}(0, 2)$ and $\xi_3^l(\mathbf{s}_i) \stackrel{iid}{\sim} \mathcal{N}(0, 1)$ which is consistent with what we observed in the data. As illustrated in the OCO-2 data section, the standard deviation of measurement error is proportional to the mean function, so we let relative standard deviation, the ratio of the standard deviation to the mean function, vary from 0.1 to 0.2 and did 200 simulations for each choice of measurement errors. For each simulation, a land fraction was generated according to $Uniform(0, 1)$ and the following statistics were calculated with respect to unmixing method and linear interpolation.

$$(4.15) \quad e_t = \frac{|\hat{\alpha}^t - \alpha|}{\alpha}, \quad \text{relative bias of interpolation}$$

$$(4.16) \quad e_k = \frac{|\hat{\alpha}^k - \alpha|}{\alpha}, \quad \text{relative bias of unmixing based on kriging}$$

We summarized results from the 200 simulations using 0.1 trimmed mean to compare these two methods, which are shown in Fig 4. Obviously, unmixing based on FPCA and kriging is more stable and accurate than the simple linear interpolation.

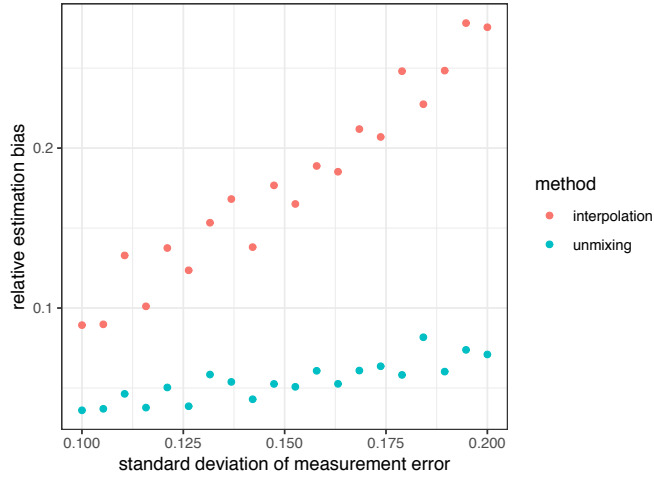
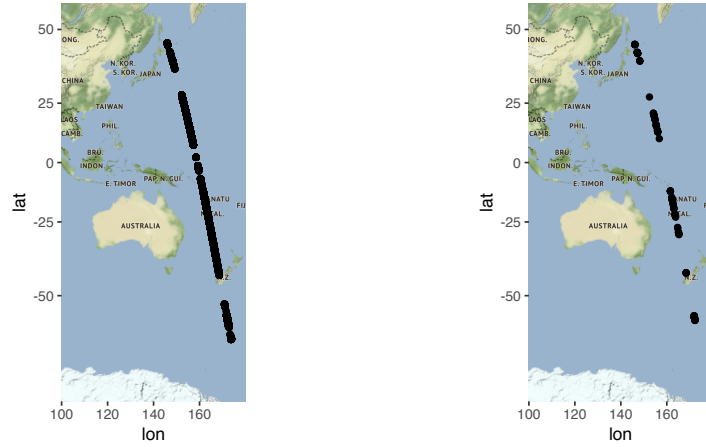


Fig 4: Relative bias against standard deviation of measurement error for both methods

5. Applications.

5.1. *Radiance Imputation over the Pacific Ocean.* OCO-2 data aim to provide a comprehensive measurement framework for CO_2 concentration and the *retrieval algorithm* implements the estimation of X_{CO_2} from Level 1 data, high-resolution spectra of reflected sunlight. As introduced in section 2, OCO-2 has a large amount of locations with completely missing radiance because of atmospheric properties including clouds and cosmic rays. So the spatial coverage of the retrieval algorithm will improve if the missing radiance can be imputed. In this section, we are interested in imputing the radiance function for wavelength indices $w_j \in \mathcal{W}$, $|\mathcal{W}| = m$ for missing sounding locations in an area \mathcal{S} . The data used as a case study in this section are downloaded from the OCO-2 Level 2 diagnostic data products (Osterman et al., 2018), available at the NASA Goddard Earth Science Data and Information Services Center (GES DISC, <https://disc.gsfc.nasa.gov/OCO-2>). We illustrate part of orbit 14793 in glint mode over the Pacific Ocean during 2017-04-13. The Level 1 variable `measured_radiance` was extracted from the dataset, along with geolocation information.



(a) orbit 14793 sounded on Pacific Ocean during 2017-04-13 (b) sampled 128 center points from orbit 14793 for experiments

Fig 5: OCO-2 data used for radiance imputation validation.

By practical experience, locations within a region spanning about 0.5° latitude can be regarded as having homogeneous covariance function $R_w(w_j, w_{j'})$. To illustrate our methods on radiance imputation in missing locations, we choose 128 locations at footprint 4 as the center points for doing the experiment described later. To guarantee that we have the right amount of data needed for FPCA and as many locations as possible for validation, the center point sampled $\mathbf{s} = (L, l)$ must satisfy the following conditions.

- region between latitude $L \pm 0.25$ has at least 164 non-missing sounding locations such that we can have at least 100 points when $T_8(\mathbf{s})$ is removed.
- $T_8(\mathbf{s})$ does not have missing locations, i.e., no gaps and total number of observations is 64.

where $T_n(\mathbf{s})$ is defined as the area consisting of closest n cross-tracks (1 to 8 footprints in a row, see Fig 1a) near \mathbf{s} . For example, $T_1(\mathbf{s})$ would be the cross-track of \mathbf{s} . If n is odd, $T_n(\mathbf{s})$ is the $T_1(\mathbf{s})$ plus $(n - 1)/2$ cross-tracks below and above \mathbf{s} . If n is even, $T_n(\mathbf{s})$ is the $T_1(\mathbf{s})$ plus $n/2$ cross-tracks observed before \mathbf{s} and $(n/2 - 1)$ cross-tracks observed after \mathbf{s} . In our validation study, n is at most 8.

Because the satellite orbit is monotonic (south-to-north) in latitude and the track is quite narrow, as shown in Fig 1a, it is sufficient to use latitude L as the covariate in model 3.5. And for simplicity, we use L_i instead of

centered latitude since it does not make a difference in estimating $\mu(w_j; \mathbf{s}_i)$ in practice. Then for each of the selected center points \mathbf{s} , we conducted the following procedure to validate our imputation algorithm.

1. Select the validation area of interest \mathcal{S} . In this study, we choose the area as latitude between $L \pm 0.25$.
2. Calculate local linear smoothing functions $\tilde{f}(w_j, \cdot)$ for each point in the 8 by 8 grid $T_8(\mathbf{s})$, which are treated as true radiance to compare with.
3. Remove 8 cross-tracks around \mathbf{s} one by one, such that region defined as $T_n(\mathbf{s}) (n = 1, \dots, 8)$ is taken away each time. Each time we remove $n (n = 1, \dots, 8)$ cross-tracks, impute for locations in the removed area $T_n(\mathbf{s})$.
4. Calculate the following two statistics for the imputed sounding locations $\{\mathbf{s}_0 : \mathbf{s}_0 \in T_n(\mathbf{s}), 1 \leq n \leq 8\}$.

$$(5.1) \quad \text{RMSE} = \sqrt{\frac{1}{m} \sum_{j=1}^m \frac{(\hat{f}(w_j; \mathbf{s}_0) - \tilde{f}(w_j; \mathbf{s}_0))^2}{\tilde{f}^2(w_j; \mathbf{s}_0)}}$$

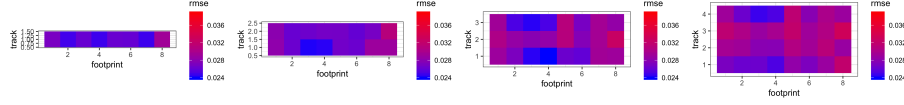
$$(5.2) \quad \text{RPMSE} = \sqrt{\frac{1}{m} \sum_{j=1}^m \left[\sum_{k=1}^K \{\xi_k(\mathbf{s}_0) - \hat{\xi}_k(\mathbf{s}_0)\} \hat{\phi}_k(w_j) \right]^2}$$

$$\text{where } \xi_k(\mathbf{s}_0) = \int \{(\tilde{f}(w; \mathbf{s}_0) - \hat{\mu}(w; \mathbf{s}_0)) \hat{\phi}_k(w) dw$$

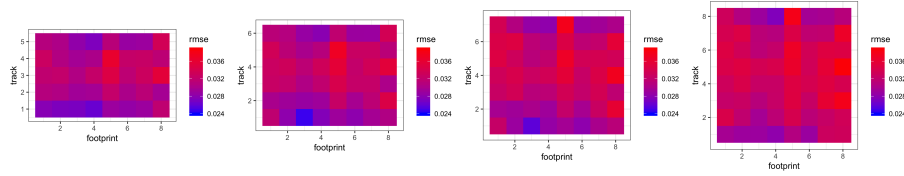
Since measured radiance can be all missing for some specific wavelength and footprint, we can either fill in missing radiance by interpolation or discard the wavelength directly such that $\beta(w_j)$ can be fully estimated. In this paper, we present the latter approach as they do not differ in terms of imputation performance. We consider two situations regarding principal component scores, with spatial dependence and without spatial dependence. Then for components which are determined to have no spatial dependence, the BLUP is reduced to $\hat{\xi}_k(\mathbf{s}_0) = \frac{1}{n} \sum_{\mathbf{s}_i \in \mathcal{S}} \xi_k(\mathbf{s}_i)$. Spatial dependence can be assessed with a permutation-based test. (Cressie and Wikle (2015)).

In Fig 6a and 6b, the 8 heatmaps represent average RMSE (5.1) in 128 imputations for removed region $T_n(\cdot)$ from $n = 1$ to $n = 8$. The imputation performance deteriorates as the size of the missing region increases. And overall, the footprints on the middle or outside, especially 5 and 8, are harder to impute than footprints next to the edge of imputed region. Fig 7 is a plot of average RMSE in the cross-track $T_1(\cdot)$ against the number of cross-tracks removed for different footprints. It is consistent with results in the heatmaps:

RMSE increases as number of cross-tracks removed increases, and footprint 5 and 8 are much worse than other footprints.



(a) average RMSE in all implementations for 1-4 cross-tracks removed



(b) average RMSE in all implementations for 5-8 cross-tracks removed

Fig 6: Removed cross-tracks colored by RMSE in radiance imputation.

The RPMSE (5.2) evaluates how well our ordinary kriging predictor for principal component score works in radiance imputation. Similarly, predicted mean square error results are summarized in Fig 8a and 8b. It is clear that prediction becomes worse as points are closer to the center and further to the edge of imputed region. Fig 9 shows a similar pattern that as removed region becomes larger, prediction error is higher for the cross-track of center point.

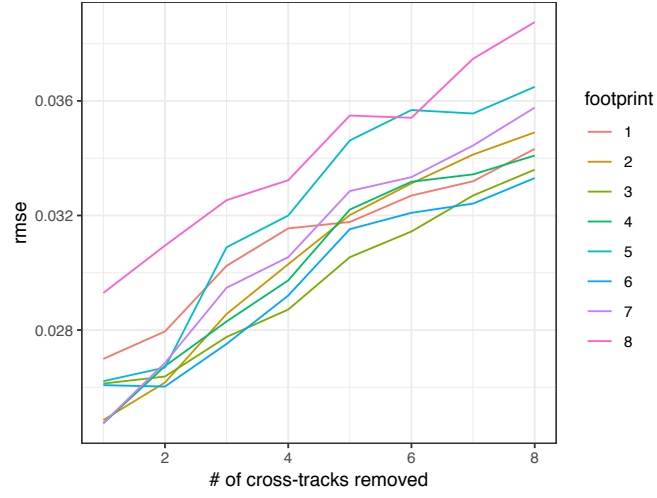
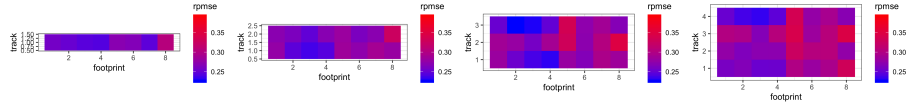
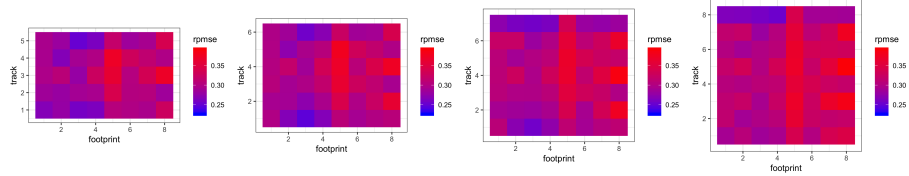


Fig 7: Average RMSE in all implementations on the cross-track of center points for different footprints



(a) average RPMSE in all implementations for 1-4 cross-tracks removed



(b) average RPMSE in all implementations for 5-8 cross-tracks removed

Fig 8: Removed cross-tracks colored by RPMSE in radiance imputation.

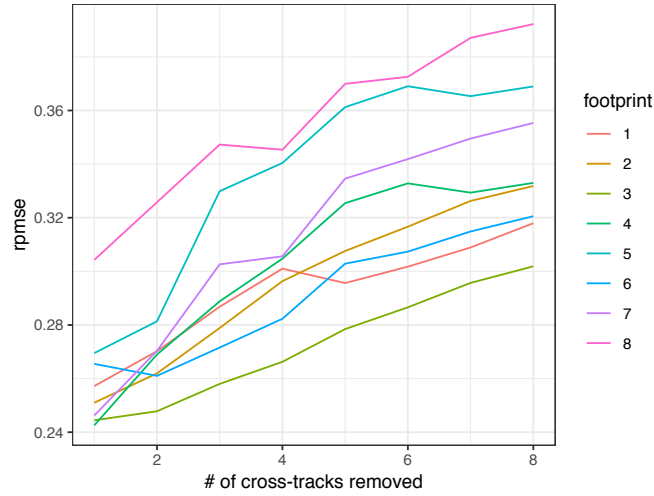


Fig 9: Average RPMSE in all implementations on the cross-track of center points for different footprints

5.2. *Land Fraction Correction around Greece.* The variable measuring land fraction in OCO-2 data is computed by mapping the OCO-2 location(longitude/latitude) to a static land/water mask. Due to the geolocation uncertainties and problematic mask in this procedure, the land fraction value provided in OCO-2 data is subject to error. Our unmixing approach can be used to provide a more accurate land fraction estimate.

To evaluate our land fraction estimation method, we use data provided in the OCO-2 Level 1B product (Eldering et al., 2017c) along the coastal area of Greece, which includes orbit 05449 on 07-11-2015 and orbit 05216 on 06-25-2016. As shown in Fig 10, there are two mixed regions to estimate respectively after removing missing radiance. Since the satellite has a repeat cycle of 16 days, orbit 05449 and 05216 are actually in the same area, though they do not have exactly the same coordinates.

Since we do not know the true land fraction, ground truth data regarding the selected two orbits were created manually. The coastal satellite pictures were downloaded from Google Earth and coastlines were added by feature editing in ArcGIS. Then each footprint's land fraction was obtained by calculating the proportion of the coastline polygon inside the area constructed by its four vertices.

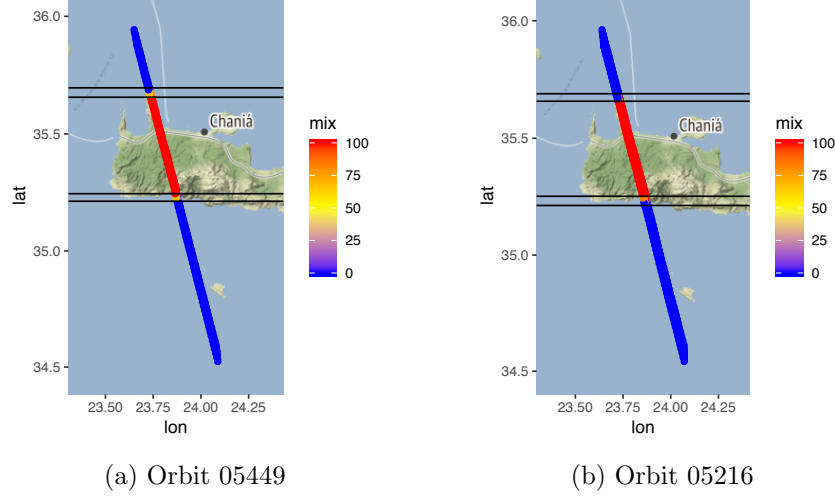


Fig 10: OCO-data used for land fraction estimation. Points are colored according to the reported land fraction for each sounding.

A mixed region is defined as the transition zone from either land to water or water to land. Typically some mixed land/water soundings are present at transition zones. In our algorithm, land fractions are to be estimated in area $\mathcal{M} = \{\mathbf{s}_i : L_1 - \delta_0 < L_i < L_2 + \delta_0\}$ where L_1 and L_2 are the minimum and maximum latitude of the mixed region, and δ_0 is the tolerance (usually set as the average latitude difference) made to account for effect of bad land fraction on the surrounding of water/land regions. Similarly, the lower unmixed region is $\mathcal{A}_1 = \{\mathbf{s}_i : L_1 - \delta_0 - 0.5 \leq L_i \leq L_1 - \delta_0\}$, the upper unmixed region is $\mathcal{A}_2 = \{\mathbf{s}_i : L_2 + \delta_0 \leq L_i \leq L_2 + \delta_0 + 0.5\}$. For a given region \mathcal{M} , the algorithm conducts the following procedure.

1. Determine the type of \mathcal{A}_1 and \mathcal{A}_2 by land fraction average: $t = 1, 2$, $\sum_{\mathbf{s}_i \in \mathcal{A}_t} \alpha_i / |\mathcal{A}_t|$. It is recognized as land if the average is more than 75%, and recognized as water if the average is less than 25%, otherwise unidentifiable. The data is qualified for our unmixing approach if both land and water are identified.
2. Do spectral imputation for locations in \mathcal{M} using radiance in area \mathcal{A}_1 and \mathcal{A}_2 as input separately. The imputation algorithm is the same as what we proposed above (section 5.1) except that a local linear smoother is applied on $\hat{\xi}_{ik}$ to reduce large variation around \mathcal{M} . The bandwidth is selected by cross validation or fixed at 0.1 if any ξ_{ik} outlier near region \mathcal{M} is detected.
3. Estimate α_i for $\mathbf{s}_i \in \mathcal{M}$ by 4.11, and truncated between 0 and 1 at

last.

We implement the unmixing algorithm on four mixed regions in Fig 10. Results in Fig 11 enable us to conclude that the unmixing approach gives a much more accurate and reliable land fraction estimate compared to the original estimates in OCO-2 data. Each plot of Fig 11 shows the change trend of land fraction estimates with respect to latitude in a mixed region for three resources (unmixing approach, OCO-2 data, ground truth). In these four plots representing four mixed regions in Fig 10, OCO-2 estimate is the most unstable and unrealistic one. And our unmixng estimate (blue line) is much more aligned with the ground truth (black line) compared to OCO-2 results (red line). Furthermore, in table 1, unmixing estimates always have a lower MSE than OCO-2 land fractions.

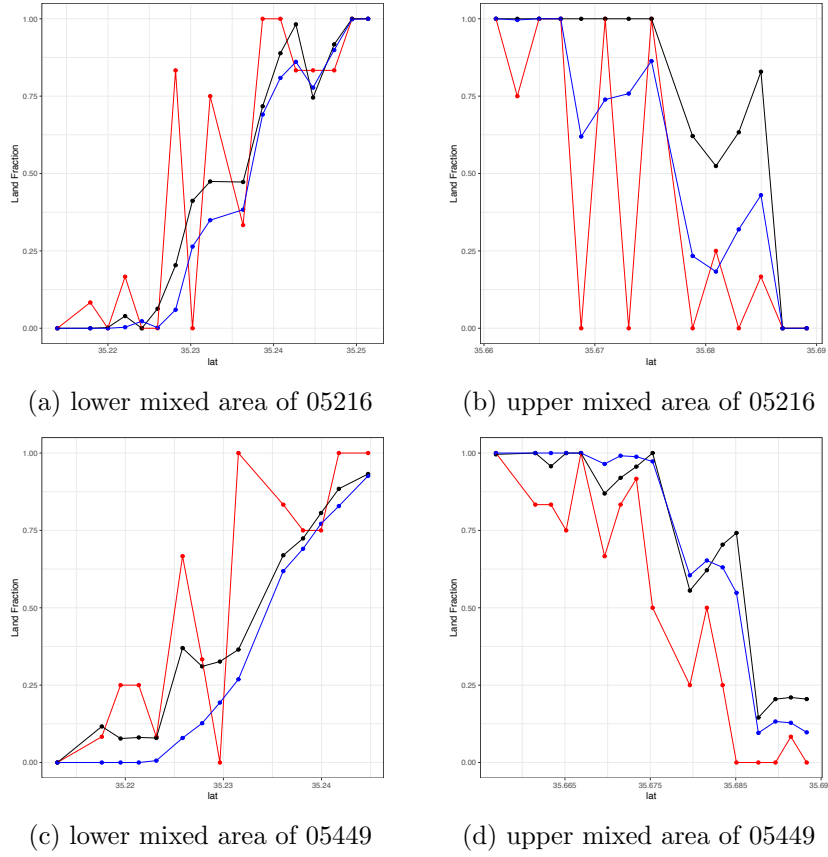


Fig 11: Land fraction from three resources against latitude in four mixed regions. *blue*: unmixing estimates, *red*: estimates in OCO-2 data, *black*: ground truth created manually.

TABLE 1
Mean Sum of squares for unmixing and OCO-2 estimates: $\sum_{s_i \in \mathcal{M}} (\hat{\alpha}_i - \alpha_i)^2 / |\mathcal{M}|$

MSE	05216 lower	05216 upper	05449 lower	05449 upper
Unmixing estimates	0.0056	0.0582	0.0132	0.0053
OCO-2	0.0592	0.0955	0.0845	0.0643

6. Discussion. In this paper, we introduced a geospatial functional model for spatial spectral data, which is inspired by imputing missing radiance for OCO-2 data. The model treated spectral radiance as a function of wavelength index and implemented footprint-specific measurement error variance and mean radiance functions. The unified framework is able to ac-

count for principal component with and without spatial dependence, which means imputation is robust and flexible to globally varying radiances. We successfully implement the algorithm and achieve acceptable high accuracy for radiance imputation at footprints over water. Furthermore, going beyond imputation, we are able to estimate land fraction in mixed footprints by unmixing water and land process, and give accurate correction for land fraction provided in OCO-2 data.

The proposed model and algorithm for doing imputation and land fraction estimation is effective over a small homogeneous area (usually within a latitude range of 0.5° in the same orbit). Due to non-stationarity, this model may not be effective for very large gaps in the data. In this case, a more complex space-time covariance model for spatial spectral data should be developed and studied to tackle imputation at a global scale.

This methodology has potential utility in data processing for OCO-2 and similar instruments. The imputation approach could allow for additional successful Level 2 retrievals when radiances can be successfully imputed. Currently the OCO-2 operational Level 2 retrieval algorithm has slightly different configurations for land and ocean soundings. Retrievals are not attempted for mixed land/ocean soundings (O'Dell et al., 2018). An accurate estimate of land fraction from the radiance data could facilitate retrievals in these mixed cases. The unmixed land fraction estimates may also have sufficient accuracy to supplement OCO-2's geolocation information.

APPENDIX A

The proof of asymptotic results for dense functional principal analysis (Theorem 4.1, 4.2 and 4.3), and consistent BLUP estimator (Theorem 4.4) for PC scores.

A.1. Mean function estimation. The mean function model 3.5 can be regarded as 8 separate linear models within each footprint group. Let $\mathbf{X}_p = [\mathbf{1} \quad \mathbf{x}_p]$ and $\boldsymbol{\beta}_p(w_j) = \{\beta_0^p(w_j), \beta_1^p(w_j)\}^T$, define total error term as

$$u(w_j; \mathbf{s}_i) = \sum_{k=1}^K \xi_k(\mathbf{s}_i) \phi_k(w_j) + \epsilon_{i,j}.$$

For any $p = 1, \dots, 8$ and any $w_j \in \mathcal{W}$,

$$(A.1) \quad \mathbf{Y}_{pj} = \mathbf{X}_p \boldsymbol{\beta}_p(w_j) + \mathbf{U}_{pj}$$

where $\mathbf{U}_{pj} = \{u(w_j; \mathbf{s}_i)\}_{\mathbf{s}_i \in \mathcal{S}_p}$ which corresponds to \mathbf{x}_p . So equivalent to (4.1), $\hat{\beta}_p(w_j) = (\mathbf{X}_p^T \mathbf{X}_p)^{-1} \mathbf{X}_p^T \mathbf{Y}_{pj}$, and

$$(A.2) \quad \hat{\beta}_p(w_j) - \beta_p(w_j) = (\mathbf{X}_p^T \mathbf{X}_p)^{-1} \mathbf{X}_p^T \mathbf{U}_{pj}$$

$$(A.3) \quad \hat{\beta}_0^p(w_j) - \beta_0^p(w_j) = \frac{1}{n_p} \sum_{\mathbf{s}_i \in \mathcal{S}_p} u(w_j; \mathbf{s}_i)$$

$$(A.4) \quad \hat{\beta}_1^p(w_j) - \beta_1^p(w_j) = (\mathbf{x}_p^T \mathbf{x}_p)^{-1} \mathbf{x}_p^T \mathbf{U}_{pj}.$$

Specifically, let $\epsilon_{pj} = \{\epsilon_{i,j}\}_{\mathbf{s}_i \in \mathcal{S}_p}$,

$$\begin{aligned} & \hat{\beta}_1^p(w_j) - \beta_1^p(w_j) \\ &= (\mathbf{x}_p^T \mathbf{x}_p)^{-1} \mathbf{x}_p^T \epsilon_{pj} + \sum_{k=1}^K (\mathbf{x}_p^T \mathbf{x}_p)^{-1} \mathbf{x}_p^T \boldsymbol{\xi}_k \phi_k(w_j) \\ (A.5) \quad &= \Delta_{pj}^1 + \Delta_{pj}^2 \end{aligned}$$

where $\Delta_{pj}^1 = (\mathbf{x}_p^T \mathbf{x}_p)^{-1} \mathbf{x}_p^T \epsilon_{pj}$ and $\Delta_{pj}^2 = \sum_{k=1}^K (\mathbf{x}_p^T \mathbf{x}_p)^{-1} \mathbf{x}_p^T \boldsymbol{\xi}_k \phi_k(w_j)$. We introduce two general assumptions, and focus on Δ_{pj}^1 first.

ASSUMPTION 1. *Based on the location layout in Fig 1a and sounding mechanism, we assume that for each footprint group p , the minimum distance between two locations $h_{n_p} = \|\mathbf{s}_{n_p-1} - \mathbf{s}_{n_p}\| = O(n_p^{\alpha-1/2})$, $0 < \alpha < 1/2$. And locations expand with $\max\{\|\mathbf{s}_i - \bar{\mathbf{s}}_p\| : \mathbf{s}_i \in \mathcal{S}_p\} = O(n_p^\alpha)$. Moreover, n_p 's are equal and $\max_{1 \leq p \leq 8} \|\bar{\mathbf{s}}_p - \bar{\mathbf{s}}\| = O(n^\alpha)$.*

ASSUMPTION 2. *For any $\mathbf{s}_i \in \mathcal{S}$, $E \left\{ \sup_{w_j \in \mathcal{W}} |f(w_j; \mathbf{s}_i)|^{\tau_1} \right\} < \infty$ for some $\tau_1 \geq 2$. And $E(|\epsilon_{i,j}|^{\tau_2}) < \infty$ for some $\tau_2 \geq 4$.*

By multidimensional Chebyshev's inequality and $E(\Delta_{pj}^1) = 0$,

$$\{(\mathbf{x}_p^T \mathbf{x}_p)^{-1} \mathbf{x}_p^T \epsilon_{pj}\}^T V_1^{-1} (\mathbf{x}_p^T \mathbf{x}_p)^{-1} \mathbf{x}_p^T \epsilon_{pj} = O_p(1)$$

where $V_1 = \text{Var}\{(\mathbf{x}_p^T \mathbf{x}_p)^{-1} \mathbf{x}_p^T \epsilon_{pj}\} = \sigma_p^2(w_j) (\mathbf{x}_p^T \mathbf{x}_p)^{-1}$. Thus, using Cauchy-Schwarz inequality and assumption 2,

$$\begin{aligned} \|\Delta_{pj}^1\|_2^2 &= \|V_1^{1/2} V_1^{-1/2} (\mathbf{x}_p^T \mathbf{x}_p)^{-1} \mathbf{x}_p^T \epsilon_{pj}\|_2^2 \\ &\leq \|V_1^{1/2}\|_F^2 \|V_1^{-1/2} (\mathbf{x}_p^T \mathbf{x}_p)^{-1} \mathbf{x}_p^T \epsilon_{pj}\|_2^2 \\ (A.6) \quad &= O_p\{\text{tr}(V_1)\} = O_p[\text{tr}\{(\mathbf{x}_p^T \mathbf{x}_p)^{-1}\}] \end{aligned}$$

where $\|\cdot\|_F$ denotes the Frobenius norm for a matrix. Recall that $\|\cdot\|$ denotes the great circle distance. Based on assumption 1, we know that $\max\{|L_i - \bar{L}_p| : \mathbf{s}_i \in \mathcal{S}_p\} = O(n_p^\alpha)$ and $\max\{|l_i - \bar{l}_p| : \mathbf{s}_i \in \mathcal{S}_p\} = O(n_p^\alpha)$. So $\text{tr}\{(\mathbf{x}_p^\top \mathbf{x}_p)^{-1}\} = O(n_p^{-2\alpha-1})$, therefore

$$(A.7) \quad \|\Delta_{pj}^1\|_2^2 = O_p(n_p^{-2\alpha-1})$$

To prove the second term in (A.5) converges, we first need the assumption on dependence over $\xi_k(\mathbf{s}_i)$. Let $\boldsymbol{\xi}_k = \{\xi_k(\mathbf{s}_i) : \mathbf{s}_i \in \mathcal{S}\}$ be the centered, random fields of the k th principal component scores. We introduce the mixing coefficient $\alpha_{p,q}^k(h)$ of $\boldsymbol{\xi}_k$ (Guyon (1995)),

$$\alpha_{p,q}^k(h) = \sup\{|P(A \cap B) - P(A)P(B)|, A \in \mathcal{F}(\boldsymbol{\xi}_k, \Lambda_1), B \in \mathcal{F}(\boldsymbol{\xi}_k, \Lambda_2), \\ | \Lambda_1 | \leq p, | \Lambda_2 | \leq q, \text{dist}(\Lambda_1, \Lambda_2) \geq h\}$$

where $\text{dist}(\Lambda_1, \Lambda_2)$ denotes the minimal great circle distance between location sets Λ_1 and Λ_2 .

ASSUMPTION 3. Assume that any $\xi_k(\mathbf{s}_i)$ satisfies the following mixing condition: as $h \rightarrow \infty$, $\alpha_{p,q}^k(h) = O(h^{-\delta})$ with $\delta > 0$.

ASSUMPTION 4. Assume that all $\xi_k(\mathbf{s}_i)$ are bounded variables, i.e., there exists constant C such that $|\xi_k(\mathbf{s}_i)| < C$ a.s. for all $1 \leq k \leq K$.

For any location \mathbf{s}_i , follow Lemma A.1 in Li et al. (2007) and Theorem 17.2.1 in Ibragimov (1975), using assumptions 1 and 3,

$$(A.8) \quad \begin{aligned} \sum_{j \neq i}^{n_p} |\text{Cov}\{\xi_k(\mathbf{s}_i), \xi_k(\mathbf{s}_j)\}| &\leq \sum_{j \neq i}^{n_p} 4C^2 \alpha_{p,q}^k(\|\mathbf{s}_i - \mathbf{s}_j\|) \\ &= \sum_{j \neq i}^{n_p} 4C^2 O(n_p^{-\alpha\delta}) \\ &= O(n_p^{1-\alpha\delta}). \end{aligned}$$

Similarly to (A.6), define $\boldsymbol{\Sigma}_k = \text{Var}(\boldsymbol{\xi}_k)$ and

$$V_2 = \text{Var}\{(\mathbf{x}_p^\top \mathbf{x}_p)^{-1} \mathbf{x}_p^\top \boldsymbol{\xi}_k\} = (\mathbf{x}_p^\top \mathbf{x}_p)^{-1} \mathbf{x}_p^\top \boldsymbol{\Sigma}_k \mathbf{x}_p (\mathbf{x}_p^\top \mathbf{x}_p)^{-1}.$$

Hence,

$$(A.9) \quad \|(\mathbf{x}_p^\top \mathbf{x}_p)^{-1} \mathbf{x}_p^\top \boldsymbol{\xi}_k\|_2^2 = O_p\{\text{tr}(V_2)\}.$$

Because of (A.8) and assumptions 1, $V_2 = O(n_p^{-2\alpha-1} + n_p^{-2\alpha-\alpha\delta})$. Then it follows that $\|(\mathbf{x}_p^T \mathbf{x}_p)^{-1} \mathbf{x}_p^T \boldsymbol{\xi}_k\|_2^2 = O_p(n_p^{-2\alpha-1} + n_p^{-2\alpha-\alpha\delta})$. Therefore,

$$\begin{aligned} \|\Delta_{pj}^2\|_2^2 &= O\left(\sum_{k=1}^K \|(\mathbf{x}_p^T \mathbf{x}_p)^{-1} \mathbf{x}_p^T \boldsymbol{\xi}_k \phi_k(w_j)\|_2^2\right) \\ (A.10) \quad &= O_p(n_p^{-2\alpha-1} + n_p^{-2\alpha-\alpha\delta}). \end{aligned}$$

So by (A.7) and (A.10), $\|\hat{\boldsymbol{\beta}}_1^p(w_j) - \boldsymbol{\beta}_1^p(w_j)\|_2^2 = O_p(n_p^{-2\alpha-1} + n_p^{-2\alpha-\alpha\delta})$. For intercept estimate $\hat{\beta}_0^p(w_j)$, following (A.3),

$$\hat{\beta}_0^p(w_j) - \beta_0^p(w_j) = \sum_{k=1}^K \phi_k(w_j) \frac{1}{n_p} \sum_{\mathbf{s}_i \in \mathcal{S}_p} \xi_k(\mathbf{s}_i) + O_p(n_p^{-1/2}).$$

By the result on spatial weak dependence (A.8),

$$\begin{aligned} \text{Var}\left\{\sum_{\mathbf{s}_i \in \mathcal{S}_p} \xi_k(\mathbf{s}_i)\right\} &= \sum_{\mathbf{s}_i \in \mathcal{S}_p} \sum_{\mathbf{s}_j \in \mathcal{S}_p} \text{Cov}\{\xi_k(\mathbf{s}_i), \xi_k(\mathbf{s}_j)\} \\ &\leq n_p \lambda_k + \sum_{\mathbf{s}_i \in \mathcal{S}_p} \sum_{j \neq i}^{n_p} |\text{Cov}\{\xi_k(\mathbf{s}_i), \xi_k(\mathbf{s}_j)\}| \\ &= O(n_p + n_p^{2-\alpha\delta}). \end{aligned}$$

It follows that $\frac{1}{n_p} \sum_{\mathbf{s}_i \in \mathcal{S}_p} \xi_k(\mathbf{s}_i) = O_p(\sqrt{n_p^{-1} + n_p^{-\alpha\delta}})$. Then we obtain $\|\hat{\beta}_0^p(w_j) - \beta_0^p(w_j)\|_2^2 = O_p(n_p^{-1} + n_p^{-\alpha\delta})$. And the consistency of parameters is obtained,

$$\|\hat{\boldsymbol{\beta}}_p(w_j) - \boldsymbol{\beta}_p(w_j)\|_2^2 = O_p(n_p^{-1} + n_p^{-\alpha\delta})$$

Based on the spatial layout as shown in Fig 1a, we assume $n_p = n/8, p = 1, \dots, 8$, then for any $w_j \in \mathcal{W}$,

$$(A.11) \quad \|\hat{\boldsymbol{\beta}}(w_j) - \boldsymbol{\beta}(w_j)\|_2 = O_p(\sqrt{n^{-1} + n^{-\alpha\delta}})$$

For a given $\mathbf{s}_i \in \mathcal{S}$,

$$\begin{aligned} &\sup_{w_j \in \mathcal{W}} |\hat{\mu}(w_j; \mathbf{s}_i) - \mu(w_j; \mathbf{s}_i)| \\ &\leq \sup_{w_j \in \mathcal{W}} |\hat{\beta}_0^{p_i}(w_j) - \beta_0^{p_i}(w_j)| + \sup_{w_j \in \mathcal{W}} |(\mathbf{s}_i - \bar{\mathbf{s}}_p) \{\hat{\boldsymbol{\beta}}_1^{p_i}(w_j) - \boldsymbol{\beta}_1^{p_i}(w_j)\}| \\ &= O_p(\sqrt{n^{-1} + n^{-\alpha\delta}}) + O\left(\sqrt{\|\mathbf{s}_i - \bar{\mathbf{s}}_p\|_2^2 \|\hat{\boldsymbol{\beta}}_1^{p_i}(w_j) - \boldsymbol{\beta}_1^{p_i}(w_j)\|_2^2}\right) \\ (A.12) \quad &= O_p(\sqrt{n^{-1} + n^{-\alpha\delta}}) \end{aligned}$$

As shown in (A.11) and (A.12), the proof of theorem 4.1 is done.

A.2. Covariance function estimation. For any $j, j' \in \mathcal{W}$, by theorem 4.1,

$$\begin{aligned} & \hat{R}_w(w_j, w_{j'}) - R_w(w_j, w_{j'}) \\ &= \mathcal{R}_{1,n} + \mathcal{R}_{2,n} + \mathcal{R}_{3,n} + O_p(\sqrt{n^{-1} + n^{-\alpha\delta}}) \end{aligned}$$

where

$$\begin{aligned} \mathcal{R}_{1,n} &= \frac{1}{n-1} \sum_{i=1}^n \left\{ \sum_{k=1}^K \xi_k(\mathbf{s}_i) \phi_k(w_j) \right\} \epsilon_{i,j'} + \frac{1}{n-1} \sum_{i=1}^n \left\{ \sum_{k=1}^K \xi_k(\mathbf{s}_i) \phi_k(w_{j'}) \right\} \epsilon_{i,j} \\ \mathcal{R}_{2,n} &= \frac{1}{n-1} \sum_{i=1}^n \left\{ \sum_{k=1}^K \xi_k(\mathbf{s}_i) \phi_k(w_j) \right\} \left\{ \sum_{k=1}^K \xi_k(\mathbf{s}_i) \phi_k(w_{j'}) \right\} - \sum_{k=1}^K \lambda_k \phi_k(w_j) \phi_k(w_{j'}) \\ \mathcal{R}_{3,n} &= \frac{1}{n-1} \sum_{i=1}^n \epsilon_{i,j} \epsilon_{i,j'} - \hat{\sigma}^2(w_j) I(j = j') \end{aligned}$$

For the first term in $\mathcal{R}_{1,n}$,

$$\frac{1}{n-1} \sum_{i=1}^n \left\{ \sum_{k=1}^K \xi_k(\mathbf{s}_i) \phi_k(w_j) \right\} \epsilon_{i,j'} = \sum_{k=1}^K \phi_k(w_j) \frac{1}{n-1} \sum_{i=1}^n \xi_k(\mathbf{s}_i) \epsilon_{i,j'}$$

with $E\{\xi_k(\mathbf{s}_i) \epsilon_{i,j'}\} = 0$. And by assumption 2 and 4,

$$\begin{aligned} \text{Var} \left\{ \sum_{i=1}^n \xi_k(\mathbf{s}_i) \epsilon_{i,j'} \right\} &= \sum_{i=1}^n \sum_{i'=1}^n \text{Cov} \{ \xi_k(\mathbf{s}_i) \epsilon_{i,j'}, \xi_k(\mathbf{s}_{i'}) \epsilon_{i',j'} \} \\ &= \sum_{i=1}^n E \{ \xi_k(\mathbf{s}_i) \epsilon_{i,j'} \}^2 = O(n) \end{aligned}$$

Hence $\frac{1}{n-1} \sum_{i=1}^n \xi_k(\mathbf{s}_i) \epsilon_{i,j'} = O_p(n^{-1/2})$. So it is true that $\mathcal{R}_{1,n} = O_p(n^{-1/2})$.

For the second term $\mathcal{R}_{2,n}$,

$$\begin{aligned} \mathcal{R}_{2,n} &= \frac{1}{n-1} \sum_{i=1}^n \left\{ \sum_{k_1=1}^K \sum_{k_2=1}^K \xi_{k_1}(\mathbf{s}_i) \xi_{k_2}(\mathbf{s}_i) \phi_{k_1}(w_j) \phi_{k_2}(w_{j'}) \right\} \\ &\quad - \sum_{k=1}^K \lambda_k \phi_k(w_j) \phi_k(w_{j'}) \\ (A.13) \quad &= \sum_{k=1}^K \left\{ \frac{1}{n-1} \sum_{i=1}^n \xi_k^2(\mathbf{s}_i) - \lambda_k \right\} \phi_k(w_j) \phi_k(w_{j'}) \end{aligned}$$

$$(A.14) \quad + \sum_{k_1=1}^K \sum_{k_2 \neq k_1}^K \phi_{k_1}(w_j) \phi_{k_2}(w_{j'}) \frac{1}{n-1} \sum_{i=1}^n \xi_{k_1}(\mathbf{s}_i) \xi_{k_2}(\mathbf{s}_i)$$

For any k , let $\nu_{p,q}^k(h)$ be the mixing coefficient of $\xi_k^2 = \{\xi_k^2(\mathbf{s}_i) : \mathbf{s}_i \in \mathcal{S}\}$. By definition of the strong mixing coefficient,

$$\nu_{p,q}^k(h) \leq \alpha_{p,q}^k(h).$$

And since theorem 1 implies $\min\{\|\mathbf{s}_i - \mathbf{s}_{i'}\| : \mathbf{s}_i, \mathbf{s}_{i'} \in \mathcal{S}\} = O(n^{\alpha-1/2})$ and $\max\{\|\mathbf{s}_i - \bar{\mathbf{s}}\| : \mathbf{s}_i \in \mathcal{S}\} = O(n^\alpha)$, then we have the similar results as (A.8). By assumption 4, let $\text{Var}\{\xi_k^2(\mathbf{s}_i)\} = \tau_k < \infty$,

$$\begin{aligned} \text{Var} \left\{ \sum_{i=1}^n \xi_k^2(\mathbf{s}_i) \right\} &= \sum_{i=1}^n \text{Var} \{ \xi_k^2(\mathbf{s}_i) \} + \sum_{i=1}^n \sum_{i' \neq i}^n \text{Cov} \{ \xi_k^2(\mathbf{s}_i), \xi_k^2(\mathbf{s}_{i'}) \} \\ &= O(n + n^{2-\alpha\delta}) \end{aligned}$$

Since $E\xi_k^2(\mathbf{s}_i) = \lambda_k$, we are able to have the following, for any k ,

$$\begin{aligned} &\frac{1}{n-1} \sum_{i=1}^n \xi_k^2(\mathbf{s}_i) - \lambda_k \\ &= \frac{1}{n-1} \sum_{i=1}^n \xi_k^2(\mathbf{s}_i) - \frac{n}{n-1} \lambda_k + \frac{1}{n-1} \lambda_k \\ &= O_p \left(\sqrt{\text{Var} \left\{ \frac{1}{n-1} \sum_{i=1}^n \xi_k^2(\mathbf{s}_i) \right\}} \right) + \frac{1}{n-1} \lambda_k \\ &= O_p(\sqrt{n^{-1} + n^{-\alpha\delta}}) \end{aligned}$$

Then it follows (A.13) is $O_p(\sqrt{n^{-1} + n^{-\alpha\delta}})$. Since $E\xi_{k_1}(\mathbf{s}_i)\xi_{k_2}(\mathbf{s}_i) = 0$, and

$$\begin{aligned} & \text{Var} \left\{ \sum_{i=1}^n \xi_{k_1}(\mathbf{s}_i) \xi_{k_2}(\mathbf{s}_i) \right\} \\ &= \sum_{i=1}^n \text{Var} \{ \xi_{k_1}(\mathbf{s}_i) \xi_{k_2}(\mathbf{s}_i) \} + \sum_{i=1}^n \sum_{i' \neq i}^n \text{Cov} \{ \xi_{k_1}(\mathbf{s}_i) \xi_{k_2}(\mathbf{s}_i), \xi_{k_1}(\mathbf{s}_{i'}) \xi_{k_2}(\mathbf{s}_{i'}) \} \\ &= n\lambda_{k_1}\lambda_{k_2} + \sum_{i=1}^n \sum_{i' \neq i}^n \text{Cov} \{ \xi_{k_1}(\mathbf{s}_i), \xi_{k_1}(\mathbf{s}_{i'}) \} \text{Cov} \{ \xi_{k_2}(\mathbf{s}_i), \xi_{k_2}(\mathbf{s}_{i'}) \} \\ &= O(n + n^{2-2\alpha\delta}) \end{aligned}$$

So it follows that (A.14) is $O_p(\sqrt{n^{-1} + n^{-2\alpha\delta}})$. Thus combining (A.13) and (A.14), $\mathcal{R}_{2,n} = O_p(\sqrt{n^{-1} + n^{-\alpha\delta}})$.

For $\mathcal{R}_{3,n}$, it is easy to see that $\frac{1}{n-1} \sum_{i=1}^n \epsilon_{i,j} \epsilon_{i,j'} = O_p(n^{-1/2})$ when $j \neq j'$. And for $j = j'$, it is reduced to

$$\mathcal{R}_{3,n} = \frac{1}{n-1} \sum_{i=1}^n \epsilon_{i,j}^2 - \hat{\sigma}^2(w_j)$$

In order to illustrate for $\mathcal{R}_{3,n}$ when $j = j'$, we introduce the following assumption.

ASSUMPTION 5. *For any k th principal component score $\xi_k(\mathbf{s}_i)$ for $\mathbf{s}_i \in \mathcal{S}_p$, it is a smooth random field with mean square continuity, $E\{\xi_k(\mathbf{s} + h) - \xi_k(\mathbf{s})\}^2 = O(h^\beta)$ as $h \rightarrow 0$, $\beta > 0$.*

Given $\mathbf{s}_i \in \mathcal{S}_p$, define $\Delta_i(w_j)$ as

$$\begin{aligned} \Delta_i(w_j) &= r(w_j; \mathbf{s}_{i+2}) - 2r(w_j; \mathbf{s}_{i+1}) + r(w_j; \mathbf{s}_i) \\ &= (\mathbf{s}_{i+2} - 2\mathbf{s}_{i+1} + \mathbf{s}_i) \boldsymbol{\beta}_1^p(w_j) + \sum_{k=1}^K \{ \xi_k(\mathbf{s}_{i+2}) - 2\xi_k(\mathbf{s}_{i+1}) + \xi_k(\mathbf{s}_i) \} \phi_k(w_j) \\ &\quad + \epsilon_{i+2,j} - 2\epsilon_{i+1,j} + \epsilon_{i,j} \end{aligned}$$

Based on assumption 5, $|\xi_k(\mathbf{s} + h) - \xi_k(\mathbf{s})| = O_p(h^\beta)$. And by the result from

$R_{1,n}$, similarly we have

$$\begin{aligned} & \frac{1}{n_p - 2} \sum_{\mathbf{s}_i \in \mathcal{S}_p} \Delta_i^2(w_j) \\ &= \frac{1}{n_p - 2} \sum_{\mathbf{s}_i \in \mathcal{S}_p} (\epsilon_{i+2,j} - 2\epsilon_{i+1,j} + \epsilon_{i,j})^2 + \frac{1}{n_p - 2} \sum_{\mathbf{s}_i \in \mathcal{S}_p} O_p(h_{i+1}^{2\beta}) \\ & \quad + \frac{1}{n_p - 2} \sum_{\mathbf{s}_i \in \mathcal{S}_p} O_p\{(\mathbf{s}_{i+2} - 2\mathbf{s}_{i+1} + \mathbf{s}_i)\boldsymbol{\beta}_1^p(w_j)\} + O_p(n_p^{-1/2}) \end{aligned}$$

By assumption 1 and 2,

$$\begin{aligned} & \frac{1}{n_p - 2} \sum_{\mathbf{s}_i \in \mathcal{S}_p} \Delta_i^2(w_j) \\ &= 6\sigma_p^2(w_j) + O_p(n_p^{-1/2}) + O_p\{n_p^{2\beta(\alpha-1/2)} + \log n_p/n_p\} + O_p(n_p^{\alpha-1/2}). \end{aligned}$$

Thus,

$$\begin{aligned} \hat{\sigma}_p^2(w_j) - \sigma_p^2(w_j) &= \frac{1}{6(n_p - 2)} \sum_{\mathbf{s}_i \in \mathcal{S}_p} \Delta_i^2(w_j) - \sigma_p^2(w_j) \\ (A.15) \quad &= O_p\{n_p^{2\beta(\alpha-1/2)} + n_p^{\alpha-1/2}\}. \end{aligned}$$

By assuming $n_p = n/8$, it follows that

$$\begin{aligned} \frac{1}{n-1} \sum_{i=1}^n \epsilon_{i,j}^2 - \hat{\sigma}^2(w_j) &= \frac{1}{n-1} \sum_{p=1}^8 \left[\sum_{i \in \mathcal{S}_p} \{\epsilon_{i,j}^2 - \hat{\sigma}_p^2(w_j)\} \right] \\ &= \frac{1}{n-1} \sum_{p=1}^8 \sum_{i \in \mathcal{S}_p} \{\epsilon_{i,j}^2 - \sigma_p^2(w_j)\} + \frac{1}{n-1} \sum_{p=1}^8 \sum_{i \in \mathcal{S}_p} \{\sigma_p^2(w_j) - \hat{\sigma}_p^2(w_j)\} \\ (A.16) \quad &= O_p\{n^{2\beta(\alpha-1/2)} + n^{\alpha-1/2}\}. \end{aligned}$$

Based on (A.16), we can obtain $\mathcal{R}_{3,n} = O_p\{n^{2\beta(\alpha-1/2)} + n^{\alpha-1/2}\}$. Finally, combine $\mathcal{R}_{1,n}$, $\mathcal{R}_{2,n}$ and $\mathcal{R}_{3,n}$,

$$\begin{aligned} & \sup_{w_j, w_{j'} \in \mathcal{W}} |\hat{R}_w(w_j, w_{j'}) - R_w(w_j, w_{j'})| \\ (A.17) \quad &= O_p(n^{-\alpha\delta/2} + n^{2\beta(\alpha-1/2)} + n^{\alpha-1/2}). \end{aligned}$$

A.3. Principal component analysis. Similar to result (S.2) in [Li, Wang and Carroll \(2013\)](#), using (2.8) in [Hall and Hosseini-Nasab \(2006\)](#) with $K < \infty$, we have the asymptotic expansion as

$$\begin{aligned} \hat{\phi}_k(w_j) - \phi_k(w_j) = & \left\{ \sum_{\substack{k'=1 \\ k' \neq k}}^K \frac{\lambda_{k'} \phi_{k'}(t)}{(\lambda_k - \lambda_{k'}) \lambda_k} \int \int (\hat{R}_w - R_w) \phi_k \phi_{k'} \right. \\ & - \lambda_k^{-1} \phi_k(t) \int \int (\hat{R}_w - R_w) \phi_k \phi_k \\ & \left. + \lambda_k^{-1} \int (\hat{R}_w - R_w)(s, t) \phi_k(s) ds \right\} \{1 + o_p(1)\}. \end{aligned}$$

By theorem 4.2, so we have

$$(A.18) \quad \sup_{w_j \in \mathcal{W}} |\hat{\phi}_k(w_j) - \phi_k(w_j)| = O_p(n^{-\alpha\delta/2} + n^{2\beta(\alpha-1/2)} + n^{\alpha-1/2}).$$

Since we estimate principal component scores $\xi_k(\mathbf{s}_i)$ by (4.7), applying Riemann approximation on measurement error term,

$$\hat{\xi}_k(\mathbf{s}_i) = \int \{f(w; \mathbf{s}_i) - \hat{\mu}(w, \mathbf{s}_i)\} \hat{\phi}_k(w) dw + O_p(n^{-1/2}).$$

Thus,

$$\begin{aligned} & \hat{\xi}_k(\mathbf{s}_i) - \xi_k(\mathbf{s}_i) \\ = & \int \{f(w; \mathbf{s}_i) - \hat{\mu}(w, \mathbf{s}_i)\} \hat{\phi}_k(w) dw - \int \{f(w; \mathbf{s}_i) - \mu(w, \mathbf{s}_i)\} \phi_k(w) dw + O_p(n^{-1/2}) \\ = & \int \{\mu(w; \mathbf{s}_i) - \hat{\mu}(w, \mathbf{s}_i)\} \hat{\phi}_k(w) dw + \int \{f(w; \mathbf{s}_i) - \mu(w, \mathbf{s}_i)\} \{\hat{\phi}_k(w) - \phi_k(w)\} dw \\ & + O_p(n^{-1/2}) \\ = & O_p(n^{-\alpha\delta/2} + n^{2\beta(\alpha-1/2)} + n^{\alpha-1/2}). \end{aligned}$$

A.4. BLUP for principal component scores. Following [Lahiri, Lee and Cressie \(2002\)](#), we assume regularity conditions on semivariogram function and weight matrix.

- (C.1) For any $\epsilon > 0$ there exists a $\delta > 0$ such that $\inf\{\sum_{l=1}^L (\gamma_k(h_l; \theta_1) - \gamma_k(h_l; \theta_2))^2 : \|\theta_1 - \theta_2\| \geq \epsilon\} > \delta$
- (C.2) $\sup\{\gamma_k(h; \theta_k) : h \in \mathbb{R}, \theta_k \in \Theta\} < \infty$. And $\gamma_k(h; \theta_k)$ is continuous with respect to θ_k .

(C.3) $V_k(\theta_k)$ is positive definite for all $\theta_k \in \Theta$ and $\sup\{\|V_k(\theta_k)\| + \|V_k(\theta_k)^{-1}\| : \theta_k \in \Theta\} < \infty$. And $V_k(\theta_k)$ is continuous with on Θ .

Let $\tilde{\gamma}_k(h)$ be the sample semivariogram using $\xi_k(\mathbf{s}_i)$, then

$$\begin{aligned} \hat{\gamma}_k(h) - \gamma_k(h; \theta_k) &= \hat{\gamma}_k(h) - \tilde{\gamma}_k(h) + \tilde{\gamma}_k(h) - \gamma_k(h; \theta_k) \\ &= \frac{1}{2N(h)} \sum_{\|\mathbf{s}_i - \mathbf{s}_j\|=h} [\{\hat{\xi}_k(\mathbf{s}_i) - \hat{\xi}_k(\mathbf{s}_j)\}^2 - \{\xi_k(\mathbf{s}_i) - \xi_k(\mathbf{s}_j)\}^2] \\ &\quad + \tilde{\gamma}_k(h) - \gamma_k(h; \theta_k) \end{aligned}$$

By theorem 4.3, and $\tilde{\gamma}_k(h)$ is a consistent estimator under assumptions we made, then

$$(A.19) \quad \hat{\gamma}_k(h) - \gamma_k(h; \theta_k) = o_p(1)$$

By theorem 3.1 in Lahiri, Lee and Cressie (2002), we have $\hat{\theta}_k - \theta_k = o_p(1)$.

REFERENCES

- CASTRO, P. E., LAWTON, W. H. and SYLVESTRE, E. (1986). Principal modes of variation for processes with continuous sample curves. *Technometrics* **28** 329–337.
- CRESSIE, N. (1985). Fitting variogram models by weighted least squares. *Journal of the International Association for Mathematical Geology* **17** 563–586.
- CRESSIE, N. (2018). Mission CO₂ntrol: A statistical scientist’s role in remote sensing of carbon dioxide. *Journal of the American Statistical Association* **113** 152–168.
- CRESSIE, N. and WIKLE, C. K. (2015). *Statistics for spatio-temporal data*. John Wiley & Sons.
- CRISP, D., POLLOCK, H. R., ROSENBERG, R., CHAPSKY, L., LEE, R. A. M., OYAFUSO, F. A., FRANKENBERG, C., O’DELL, B. C. J. C. W., DORAN, G. B., ELDERING, A., FISHER, B. M., FU, D., GUNSON, M. R., MANDRAKE, L., OSTERMAN, G. B., SCHWANDNER, F. M., SUN, K., TAYLOR, T. E. and WUNCH, D. (2017). The on-orbit performance of the Orbiting Carbon Observatory-2 (OCO-2) instrument and its radiometrically calibrated products. *Atmospheric Measurement Techniques* **10** 59–81.
- ELDERING, A., WENNBURG, P. O., CRISP, D., SCHIMEL, D. S., GUNSON, M. R., CHATTERJEE, A., LIU, J., SCHWANDNER, F. M., SUN, Y., O’DELL, C. W., FRANKENBERG, C., TAYLOR, T., FISHER, B., OSTERMAN, G. B., WUNCH, D., HAKKARAINEN, J., TAMMINEN, J. and WEIR, B. (2017a). The Orbiting Carbon Observatory-2 early science investigations of regional carbon dioxide fluxes. *Science* **358**.
- ELDERING, A., O’DELL, C. W., WENNBURG, P. O., CRISP, D., GUNSON, M., VIATTE, C. et al. (2017b). The Orbiting Carbon Observatory-2: First 18 months of science data products. *Atmospheric Measurement Techniques* **10** 549–563.
- ELDERING, A., OSTERMAN, G., POLLOCK, R., LEE, R., ROSENBERG, R., OYAFUSO, F., CRISP, D., CHAPSKY, L. and GRANAT, R. (2017c). Orbiting Carbon Observatory (OCO-2) Level 1B Algorithm Theoretical Basis Jet Propulsion Laboratory JPL document OCO D-55206.
- GUYON, X. (1995). *Random fields on a network: modeling, statistics, and applications*. Springer Science & Business Media.

- HALL, P. and HOSSEINI-NASAB, M. (2006). On properties of functional principal components analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68** 109–126.
- IBRAGIMOV, I. (1975). Independent and stationary sequences of random variables. *Wolters, Noordhoff Pub.*
- LAHIRI, S. N., LEE, Y. and CRESSIE, N. (2002). On asymptotic distribution and asymptotic efficiency of least squares estimators of spatial variogram parameters. *Journal of Statistical Planning and Inference* **103** 65–85.
- LI, Y., WANG, N. and CARROLL, R. J. (2013). Selecting the number of principal components in functional data. *Journal of the American Statistical Association* **108** 1284–1294.
- LI, Y., WANG, N., HONG, M., TURNER, N. D., LUPTON, J. R., CARROLL, R. J. et al. (2007). Nonparametric estimation of correlation functions in longitudinal and spatial data, with application to colon carcinogenesis experiments. *The Annals of Statistics* **35** 1608–1643.
- LIU, C., RAY, S. and HOOKER, G. (2017). Functional principal component analysis of spatially correlated data. *Statistics and Computing* **27** 1639–1654.
- O'DELL, C. W., ELDERING, A., WENNBURG, P. O., CRISP, D., GUNSON, M. R., FISHER, B., FRANKENBERG, C., KIEL, M. et al. (2018). Improved Retrievals of Carbon Dioxide from the Orbiting Carbon Observatory-2 with the version 8 ACOS algorithm. *Atmospheric Measurement Techniques* **11** 6539–6576.
- OSTERMAN, G., ELDERING, A., AVIS, C., CHAFIN, B., O'DELL, C., FRANKENBERG, C., FISHER, B., MANDRAKE, L., WUNCH, D., GRANAT, R. and CRISP, D. (2018). Orbiting Carbon Observatory (OCO-2) Level 2 Data Product User's Guide, Operational L1 and L2 Data Versions 8 and Lite File Version 9 Jet Propulsion Laboratory.
- RAMSAY, J. O. (2004). Functional data analysis. *Encyclopedia of Statistical Sciences* **4**.
- RICE, J. A. and SILVERMAN, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society: Series B (Methodological)* **53** 233–243.
- YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* **100** 577–590.
- YAO, F., MÜLLER, H.-G., CLIFFORD, A. J., DUEKER, S. R., FOLLETT, J., LIN, Y., BUCHHOLZ, B. A. and VOGEL, J. S. (2003). Shrinkage estimation for functional principal component scores with application to the population kinetics of plasma folate. *Biometrics* **59** 676–685.
- ZHANG, L., BALADANDAYUTHAPANI, V., ZHU, H., BAGGERLY, K. A., MAJEWSKI, T., CZERNIAK, B. A. and MORRIS, J. S. (2016). Functional CAR models for large spatially correlated functional datasets. *Journal of the American Statistical Association* **111** 772–786.

IOWA STATE UNIVERSITY
 E-MAIL: xchang@iastate.edu
zhuz@iastate.edu

JET PROPULSION LABORATORY
 E-MAIL: Jonathan.M.Hobbs@jpl.nasa.gov