Submitted to the Annals of Applied Statistics
arXiv: arXiv:0000.0000

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*[†]

As NASA's first dedicated CO_2 monitoring satellite, the Orbiting Carbon Observatory-2 (OCO-2) aims to provide a comprehensive measurement network which is essential to any carbon management strategy. And X_{CO2} retrieval algorithm is developed for estimating CO_2 concentration from high-resolution spectra of reflected sunlight at wavelengths. However, due to large amount of missing radiance functions, the spatial coverage of the retrieval algorithm is limited and needed to be improved. Also, the unstable land fraction estimates of OCO-2 data contribute to the failure of retrieval at mixed locations. So 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 OCO-2 data, the model considers footprint-separated data information through mean function and measurement error variance modeling. Then principal component scores are treated as a random field and predictive by ordinary kriging. The proposed method is validated to impute spectral radiance with high accuracy for the purpose of radiance imputation on 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 (CO_2) 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 satellite radiances, termed Level 1 data products, into inferences on carbon

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

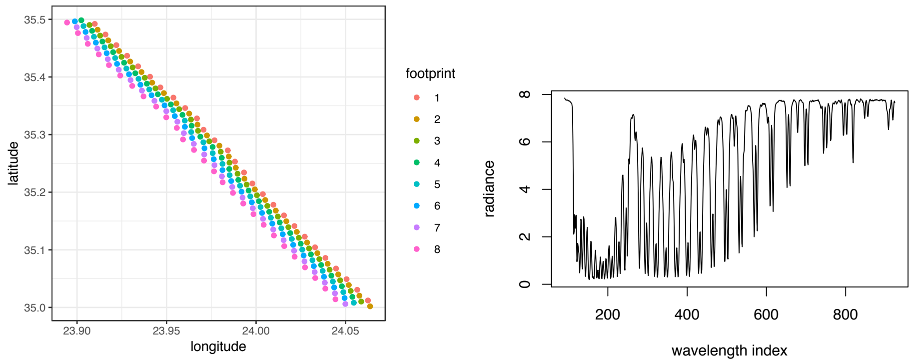
sources and sinks (Cressie, 2018). The *retrieval algorithm* implements the estimation of CO_2 concentration from Level 1 data (O'Dell et al., 2018). For OCO-2, the primary retrieval output, or Level 2 product, of interest is X_{CO2}, which is the average concentration of carbon dioxide in a column of 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.

With the development of functional data analysis in both theoretical and methodological levels in the past decades, FPCA techniques are being widely studied and applied on longitudinal data, image data, geospatial problems, etc. As one dominant field in functional data analysis, functional principal component analysis (FPCA) are usually divided into sparse scenario where only a few scatter points are observed per curve (Yao, Müller and Wang, 2005) and dense scenario where data is observed in the continuum per curve (Ramsay, 2004). Obviously, measured radiance in OCO-2 data is recognized as dense functional data in wavelength with replication on geolocations. However, as illustrated in next the section, measurement error strongly depends on the spatial layout variable called footprints and the wavelength. Although literature dealing with dense functional data involves measurement error estimation (Castro, Lawton and Sylvestre (1986), Yao et al. (2003)), these need to be generalized 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. This so-called spatially dependent functional data gained much attention recently (Liu, Ray and Hooker (2017), Li et al. (2007), Zhang et al. (2016)). But for real applications, spatially dependent spectral data has not been well studied, especially on remote sensing data. More importantly, how to integrate different characteristic from eight 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. That is an unified model considering separability resulted from footprints and covering 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 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



(a) spatial layout of 8 cross-track footprints from orbit 10575

(b) radiance function of O_2 band at (34.59285, 24.18569) on orbit 10575

Fig 1: OCO-2 Level 1 data

wavelength, and wavelengths and radiances are categorized into 3 bands as introduced above. For simplicity here, we will only consider radiance in O_2 band (Fig 1b), which is the first 1016 components of the 3048-dimensional radiance vector before filtering. The O_2 band refers to wavelengths (colors) within the 0.765 micron molecular oxygen A band. The spectral indices, $j = 1, \ldots, 1016$ 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 was divided by 10^{19} without loss of generality.

Due to the fact that 8 footprints effectively constitute 8 instruments, we found that 2nd order differencing estimates of measurement error (eqn. 4.6) 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.

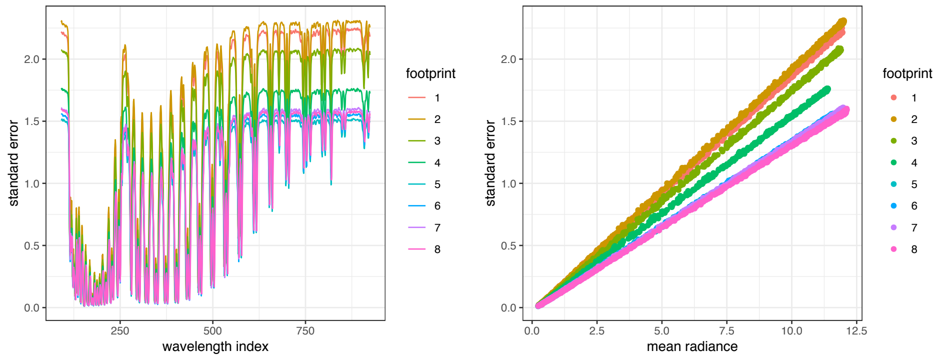


Fig 2: Measurement error estimates by footprint in region within latitude [34.3, 34.8] on orbit 10575

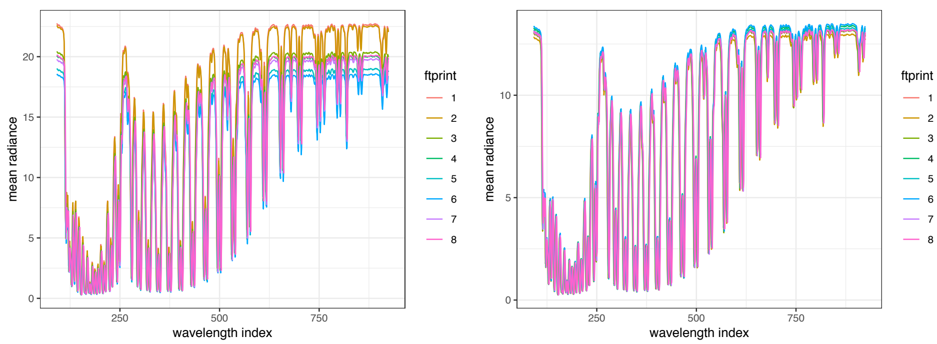


Fig 3: Mean radiance by footprint in regions within latitude [34, 34.5] and [35, 35.5] on orbit 10575

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 $\mathbf{s}_i = (L_i, l_i) = (latitude, longitude)$ is treated as a function of m_i wavelengths, i.e., function in wavelength index w_j, $1 \le w_j \le 1016$. Here $\mathbf{s}_i$ is ordered by footprint 1 to 8, sounding time, and in the area of interest $\mathcal{S} = \{\mathbf{s}_1, \cdots, \mathbf{s}_n\}$ which is not identified as a sphere. We model measured radiance as

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

where we assume error $\epsilon_{i,j}$ is independent of $f(w_j, \mathbf{s}_i)$ with mean zero and variance $\sigma^2_{p_i}(w_j)$, and p_i is the footprint where $\mathbf{s}_i$ belongs to. 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,

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

where eigenfunctions $\phi_k(\cdot)$ is an orthonormal basis in Hlibert space and the principal component score $\xi_k(\mathbf{s}_i)$ are random variables with variance λ_k. For each $\xi_k(\mathbf{s}_i)$, it depends on $\mathbf{s}_i$ and is assumed to be a mean zero, second order stationary and isotropic random field. For any two points $\mathbf{s}_i$ and $\mathbf{s}_{i'}$, we assume

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

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 principle component captures most of the variation. Thus we consider the truncated version with $K < \infty$,

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

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,

$$\mu(w_j; \mathbf{s}_i) = \beta_0^{p_i}(w_j) + \beta_1^{p_i}(w_j)(\mathbf{s}_i - \bar{\mathbf{s}}_p) \tag{3.5}$$

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

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

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 ± 0.5 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 the mixed area we are interested in, 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 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,

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

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 $\sigma^2_{p_i}(w_j)$. Note the variance function $\sigma^2_{p_i}(w_j)$ 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.

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

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

4. Estimation.

4.1. *Dense FPCA.* Let $\mathcal{W}$ be the wavelength indices available in the area of interest, usually the collection of wavelengths corresponding to radiance 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 vector of location covariate in set $\mathcal{S}_p$, i.e., $(\mathbf{s}_1 - \bar{\mathbf{s}}_p, \cdots, \mathbf{s}_n - \bar{\mathbf{s}}_p) = (\mathbf{x}_1^{\mathrm{T}}, \cdots, \mathbf{x}_8^{\mathrm{T}})$. Similarly, measured radiance at wavelength index w_j is $\mathbf{Y}_j = \{r(w_j; \mathbf{s}_1), \cdots, r(w_j; \mathbf{s}_n)\} = (\mathbf{Y}_{1j}^{\mathrm{T}}, \cdots, \mathbf{Y}_{8j}^{\mathrm{T}})$ 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 design matrix in 3.5 is constructed as following.

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

The footprint-specific coefficients $\boldsymbol{\beta}(w_j) = \{\beta_0^1(w_j), \beta_1^1(w_j), \cdots, \beta_0^8(w_j), \beta_1^8(w_j)\}$ for wavelength index w_j can be estimated by

$$\widehat{\boldsymbol{\beta}}(w_j) = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}_j \tag{4.1}$$

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

$$\widehat{\mu}(w_j;\mathbf{s}_i) = \widehat{\beta}_0^{p_i}(w_j) + \widehat{\beta}_1^{p_i}(w_j)(\mathbf{s}_i - \bar{\mathbf{s}}_p). \tag{4.2}$$

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

$$\sup_{w_j\in\mathcal{W}} ||\widehat{\boldsymbol{\beta}}(w_j) - \boldsymbol{\beta}(w_j)||_2 = O_p(n^{-\alpha\delta/2}), \tag{4.3}$$

$$\sup_{w_j\in\mathcal{W}} |\widehat{\mu}(w_j;\mathbf{s}_i) - \mu(w_j;\mathbf{s}_i)| = O_p(n^{-\alpha\delta/2}) \tag{4.4}$$

for some $0<\alpha<1$ *and* $0<\delta<1$*.*

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

$$R_w(w_j, w_{j'}) \equiv R(w_j, w_{j'}, 0) = \sum_{k=1}^{K} \lambda_k \phi_k(w_j)\phi_k(w_{j'}). \tag{4.5}$$

Then we estimate spectral covariance function as

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

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

$$\widehat{\sigma}_p^2(w_j) = \frac{1}{6(n_p-2)}\sum_{\mathbf{s}_i\in\mathcal{S}_p}\{r(w_j;\mathbf{s}_{i+2}) - 2r(w_j;\mathbf{s}_{i+1}) + r(w_j;\mathbf{s}_i)\}^2 \tag{4.6}$$

$$\widehat{\sigma}^2(w_j) = \frac{1}{n-1}\sum_{p=1}^{8} n_p\widehat{\sigma}_p^2(w_j) \tag{4.7}$$

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

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

$$\sup_{w_j, w_{j'}\in\mathcal{W}} |\widehat{R}_w(w_j, w_{j'}) - R_w(w_j, w_{j'})| = o_p(1). \tag{4.8}$$

Please see (A.18) for the derived converage rate.

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

$$\widehat{R}_w(w_j, w_{j'}) = \sum_{k=1}^{K} \widehat{\lambda}_k \widehat{\phi}_k(w_j)\widehat{\phi}_k(w_{j'}) \tag{4.9}$$

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,

$$\widehat{\xi}_k(\mathbf{s}_i) = \int \{r(w; \mathbf{s}_i) - \widehat{\mu}(w; \mathbf{s}_i)\}\, \widehat{\phi}_k(w) dw. \tag{4.10}$$

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 $\widehat{\xi}_k(\mathbf{s}_i)$ calculated from above forms a second order stationary and isotropic random field. Let $\mathbf{\Sigma}_k = \mathrm{Var}(\boldsymbol{\xi}_k)$ and $\boldsymbol{\sigma}_k = \mathrm{Cov}\{\boldsymbol{\xi}_k(\mathbf{s}_0), \boldsymbol{\xi}_k\}$, which can be constructed by the covariance function $G_k(\cdot)$ estimated through fitting covariance function model to semivariogram of $\widehat{\boldsymbol{\xi}}_k$ by weighted least squares. In this application, we use exponential covariance function. 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 $\widehat{\boldsymbol{\xi}}_k$ should have zero mean based on the assumption in 3.2, we use ordinary kriging in a small sample regarding $\mathcal{S}$ to achieve better prediction performance. So the best linear unbiased predictor for $\xi_k(\mathbf{s}_0)$ is

$$\widehat{\xi}_k(\mathbf{s}_0) = \widehat{\mu} + \widehat{\boldsymbol{\sigma}}_k^{\mathrm{T}} \widehat{\mathbf{\Sigma}}_k^{-1} (\widehat{\boldsymbol{\xi}}_k - \mathbf{1}\widehat{\mu}) \tag{4.11}$$

where $\widehat{\mu} = \left(\mathbf{1}^{\mathrm{T}} \widehat{\mathbf{\Sigma}}_k^{-1} \mathbf{1}\right)^{-1} \mathbf{1}^{\mathrm{T}} \widehat{\mathbf{\Sigma}}_k^{-1} \widehat{\boldsymbol{\xi}}_k$.

The radiance function for the location $\mathbf{s}_0$ can be predicted as

$$\widehat{f}(w_j; \mathbf{s}_0) = \widehat{\mu}(w_j; \mathbf{s}_0) + \sum_{k=1}^{K} \widehat{\xi}_k(\mathbf{s}_0)\widehat{\phi}_k(w_j), \quad w_j \in \mathcal{W} \tag{4.12}$$

where K is determined by fraction of variation explained in eigen decomposition.

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 $\widehat{f}_w(w_j;\mathbf{s}_i)$ and $\widehat{f}_l(w_j;\mathbf{s}_i)$. Assuming the mixing model 3.7, we can estimate land fraction α_i by minimizing squared loss function $||r(w_j;\mathbf{s}_i) - \alpha\widehat{f}_l(w_j;\mathbf{s}_i) - (1-\alpha)\widehat{f}_w(w_j;\mathbf{s}_i)||_2^2$, which has the solution

$$\widehat{\alpha}_i^k = \frac{\sum_{j=1}^{m_i}\{r(w_j;\mathbf{s}_i) - \widehat{f}_w(w_j;\mathbf{s}_i)\}\{\widehat{f}_l(w_j;\mathbf{s}_i) - \widehat{f}_w(w_j;\mathbf{s}_i)\}}{\sum_{j=1}^{m_i}\{\widehat{f}_l(w_j;\mathbf{s}_i) - \widehat{f}_w(w_j;\mathbf{s}_i)\}^2}. \tag{4.13}$$

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 variance function of measurement error using limited points in mixed region. The ordinary least square 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

$$\widehat{\alpha}_i^t = \frac{\sum_{j=1}^{m}\{r(w_j;\mathbf{s}_i) - r(w_j;\mathbf{s}_i^w)\}\{r(w_j;\mathbf{s}_i^l) - r(w_j;\mathbf{s}_i^w)\}}{\sum_{j=1}^{m}\{r(w_j;\mathbf{s}_i^l) - r(w_j;\mathbf{s}_i^w)\}^2}. \tag{4.14}$$

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 of one footprint which are generated by equal space along latitude and longitude. And the location in the middle (35.49881, 23.83578) is supposed to be mixed with fraction between 0 and 1. The lower latitude part until 35 is assumed to be water and higher latitude part until 36 is land. It is similar to 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):

$$C_w(h) = 5\exp(-h/10) \tag{4.15}$$

$$C_l(h) = 10\exp(-h/7) \tag{4.16}$$

For latter principal components, we assume $\xi_2^w(\mathbf{s}_i) \overset{iid}{\sim} \mathcal{N}(0,2)$, $\xi_2^l(\mathbf{s}_i) \overset{iid}{\sim} \mathcal{N}(0,2)$ and $\xi_3^l(\mathbf{s}_i) \overset{iid}{\sim} \mathcal{N}(0,1)$ which is consistent with what we observed in the data. As illustrated in the OCO-2 data section, standard error of measurement error is proportional to mean function, so we let relative standard deviation to 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.

$$e_t = \frac{|\widehat{\alpha}^t - \alpha|}{\alpha}, \quad \text{relative bias of interpolation} \tag{4.17}$$

$$e_k = \frac{|\widehat{\alpha}^k - \alpha|}{\alpha}, \quad \text{relative bias of unmixing based on kriging} \tag{4.18}$$

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.

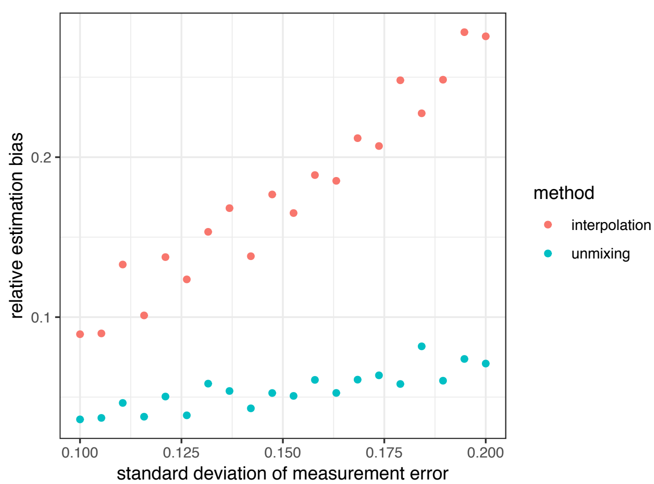
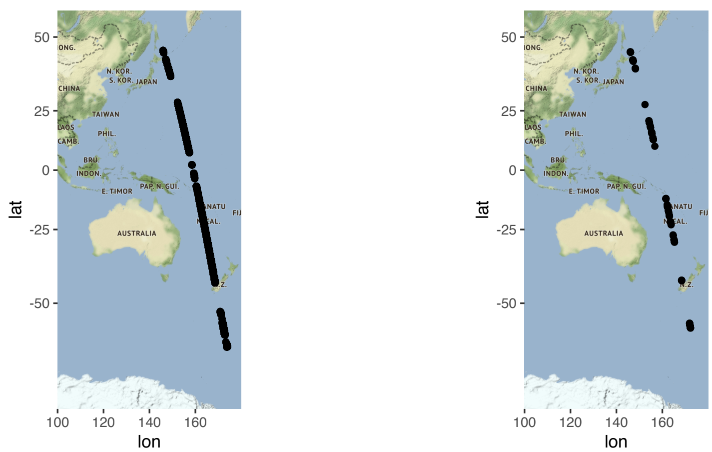


Fig 4: Relative bias against standard deviation of measurement error for both methods

5. Applications.

5.1. *Radiance Imputation on Pacific Ocean.* OCO-2 data aims to provide a comprehensive measurement framework for CO_2 concentration and the *retrieval algorithm* implements the estimation of X_{CO2} concentration from Level 1 data, high-resolution spectra of reflected sunlight at wavelengths. As introduced in section 2, OCO-2 data has large amount of locations with completely missing radiance because of atmospheric properties including clouds and cosmic-ray. So the spatial coverage of retrieval algorithm will improve if the missing radiance can be imputed. In this section, we are interested in imputing radiance function of wavelength indices $w_j \in \mathcal{W}, |\mathcal{W}| = m$ for missing sounding locations in an area $\mathcal{S}$. The data used as case study in this section is downloaded from NASA data center, part of orbit 14793 in L2DiaGL data product which are sounded on Pacific Ocean during 2017-04-13. The Level 1 variables measured radiance were extracted from the dataset.



(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, region with latitude difference less than 0.5 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 much 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 sounding goes in latitude direction 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 $\widetilde{f}(w_j,\cdot)$ 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 $T_n(\mathbf{s})(n=1,\cdots,8)$ is taken away each time. Each time we remove $n(n=1,\cdots,8)$ cross-tracks, impute for locations in the removed area $T_n(\mathbf{s})$.
4. Calculate the following two statistics for the imputed sounding locations $\{\mathbf{s}_0 : \mathbf{s}_0 \in T_n(\mathbf{s}), 1 \le n \le 8\}$.

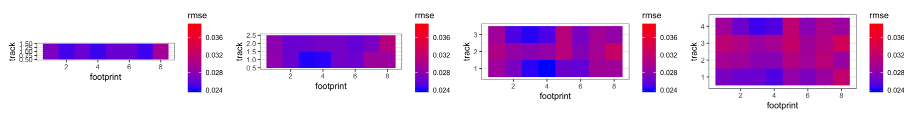
$$\text{RMSE} = \sqrt{\frac{1}{m}\sum_{j=1}^{m}\frac{(\widehat{f}(w_j;\mathbf{s}_0)-\widetilde{f}(w_j;\mathbf{s}_0))^2}{\widetilde{f}^2(w_j;\mathbf{s}_0)}} \tag{5.1}$$

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

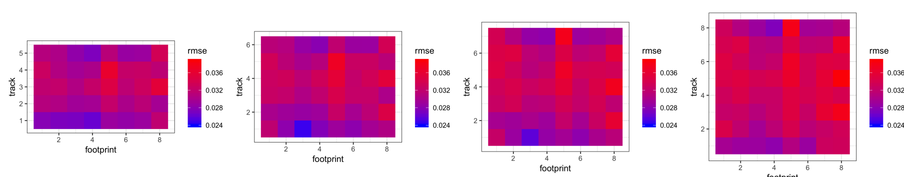
where $\xi_k(\mathbf{s}_0) = \int\{(\widetilde{f}(w;\mathbf{s}_0)-\widehat{\mu}(w;\mathbf{s}_0)\}\widehat{\phi}_k(w)dw$

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 $\boldsymbol{\beta}(w_j)$ 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 kinds of situation 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 $\widehat{\xi}_k(\mathbf{s}_0) = \frac{1}{n}\sum_{\mathbf{s}_i \in \mathcal{S}} \xi_k(\mathbf{s}_i)$. So there is a procedure determining whether the component accounts for spatial dependence in the algorithm: a permutation-based test for spatial dependence (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. The 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 work 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. And Fig 9 shows a similar pattern that as removed region becomes larger, prediction error is higher for the cross-track of center point.

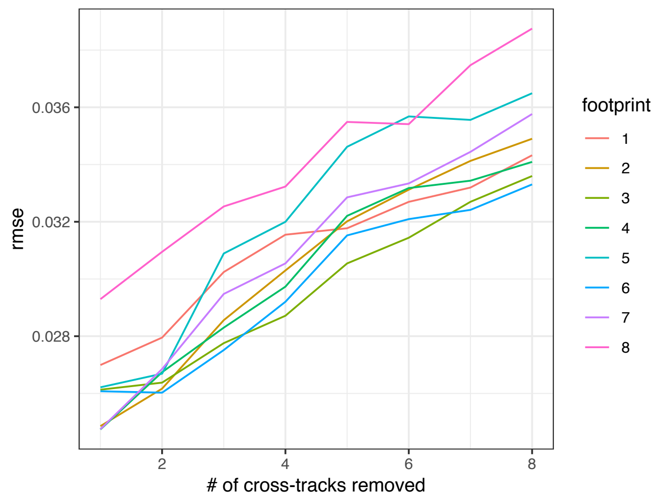
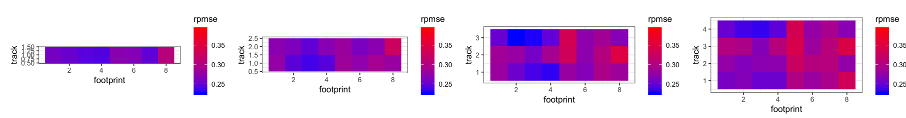
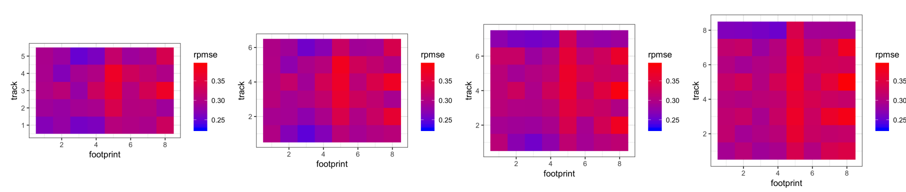


Fig 7: Average RMSE in all implementations on the cross-track of center points for different footprints



(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.

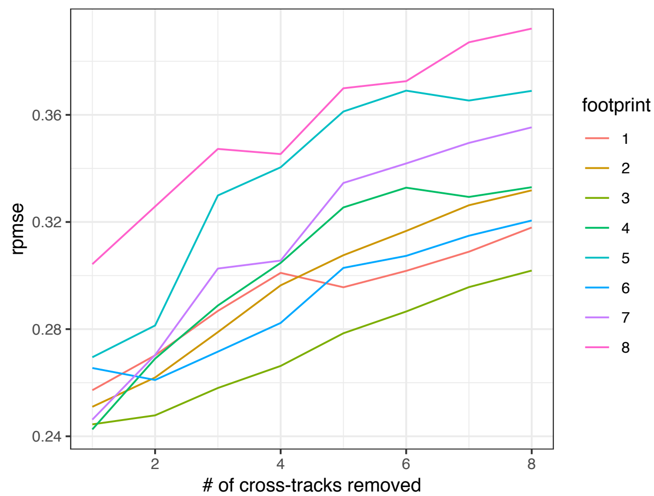
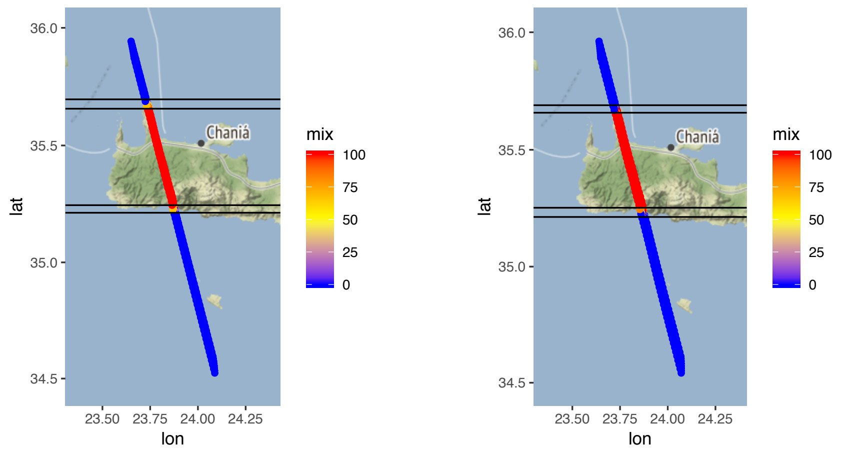


Fig 9: Average RPMSE in all implementations on the cross-track of center points for different footprints

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 not reliable. 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 by Jet Propulsion Laboratory along the coastal area of Greece, orbit 05449 on 07-11-2015 and orbit 05216 on 06-25-2016. As shown in the following pictures of sounding locations, there are two mixed regions to estimate respectively after removing missing radiance. Since the satellite came back every 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, a ground truth data regarding the selected two orbits were created manually. The coastal satellite pictures were downloaded from Google Earth and added coastline by feature editing in ArcGIS. Then each footprint's land fraction was obtained by calculating proportion of coastline polygon inside the area constructed by its four vertices.



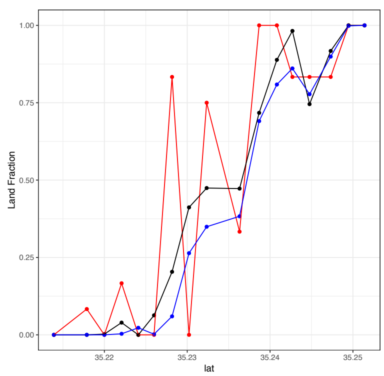
(a) Orbit 05449 (b) Orbit 05216

Fig 10: OCO-data used for land fraction estimation.

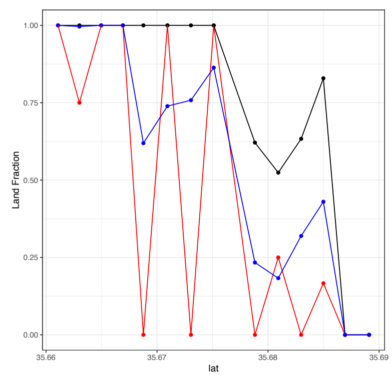
A mixed region is defined as the transition zone from either land to water or water to land. It sometimes contains pure water/land footprint, but must be mixed footprint at junctions. In our algorithm, land fractions are to be estimated in area $\mathcal{M} = \{\mathbf{s}_i : L_1 - \delta < L_i < L_2 + \delta\}$ where L_1 and L_2 are the minimum and maximum latitude of the mixed region, and δ 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 $\mathcal{A}_1 = \{\mathbf{s}_i : L_1 - \delta - 0.5 \le L_i \le L_1 - \delta\}$, the upper unmixed region is $\mathcal{A}_2 = \{\mathbf{s}_i : L_2 + \delta \le L_i \le L_2 + \delta + 0.5\}$. For a given region $\mathcal{M}$, the algorithm conduct 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 $\widehat{\xi}_{ik}$ 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 4.13, and truncated between 0 and 1 at last.

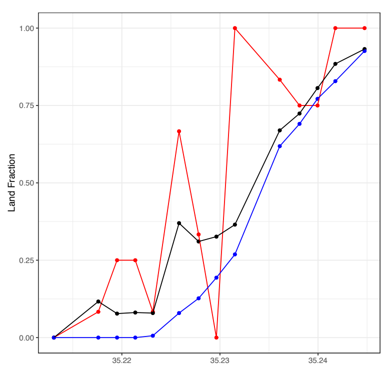
We implement the unmixing algorithm on four mixed regions in Fig 10. Results below enables us to conclude that the unmixing approach gives a much more accurate and reliable land fraction estimate compared to original measurement in OCO-2 data. Each plot of Fig 11 shows the change trend of land fraction 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.



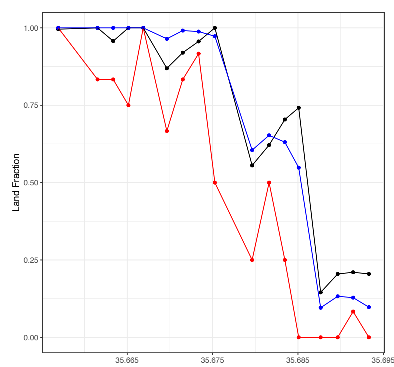
(a) lower mixed area of 05216



(b) upper mixed area of 05216



(c) lower mixed area of 05449



(d) upper mixed area of 05449

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 unmixng and OCO-2 estimates: $\sum_{\mathbf{s}_i \in \mathcal{M}} (\widehat{\alpha}_i - \alpha_i)^2 / |\mathcal{M}|$

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 treat different characteristics among footprints through the footprint specific variance of measurement error and mean function depending on footprints. The unified framework is able to account for principal component with and without spatial dependence, which means imputation is robust and flexible to globally varying radiance. We successfully implement the algorithm and achieve acceptable high accuracy for radiance imputation at footprints over water. Furthermore, go beyond imputation, we are able to estimate land fraction in mixed footprints by unmixng 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 and 4.2).

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 $\boldsymbol{\beta}_p(w_j) = \{\beta_0^p(w_j), \beta_1^p(w_j)\}$, 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 $p = 1, \cdots, 8$ and any $w_j \in \mathcal{W}$,

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

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), $\widehat{\boldsymbol{\beta}}_p(w_j) = \left(\mathbf{X}_p^{\mathrm{T}}\mathbf{X}_p\right)^{-1} \mathbf{X}_p^{\mathrm{T}}\mathbf{Y}_{pj}$, and

$$\widehat{\boldsymbol{\beta}}_p(w_j) - \boldsymbol{\beta}_p(w_j) = \left(\mathbf{X}_p^{\mathrm{T}}\mathbf{X}_p\right)^{-1} \mathbf{X}_p^{\mathrm{T}}\mathbf{U}_{pj} \tag{A.2}$$

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

$$\widehat{\beta}_1^p(w_j) - \beta_1^p(w_j) = \frac{\sum_{\mathbf{s}_i \in \mathcal{S}_p} (\mathbf{s}_i - \bar{\mathbf{s}}_p) u(w_j; \mathbf{s}_i)}{\sum_{\mathbf{s}_i \in \mathcal{S}_p} (\mathbf{s}_i - \bar{\mathbf{s}}_p)^2} \tag{A.4}$$

where $\quad \bar{\mathbf{s}}_p = \dfrac{1}{n_p} \sum_{\mathbf{s}_i \in \mathcal{S}_p} \mathbf{s}_i$. Specifically,

$$\begin{aligned} &\widehat{\beta}_1^p(w_j) - \beta_1^p(w_j) \\ =&\frac{\sum_{\mathbf{s}_i \in \mathcal{S}_p} (\mathbf{s}_i - \bar{\mathbf{s}}_p)\epsilon_{i,j}}{\sum_{\mathbf{s}_i \in \mathcal{S}_p} (\mathbf{s}_i - \bar{\mathbf{s}}_p)^2} + \frac{\sum_{\mathbf{s}_i \in \mathcal{S}_p} (\mathbf{s}_i - \bar{\mathbf{s}}_p) \sum_{k=1}^{K} \xi_k(\mathbf{s}_i)\phi_k(w_j)}{\sum_{\mathbf{s}_i \in \mathcal{S}_p} (\mathbf{s}_i - \bar{\mathbf{s}}_p)^2}. \end{aligned} \tag{A.5}$$

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* $h_{n_p} = |\mathbf{s}_{n_p-1} - \mathbf{s}_{n_p}| = O(n_p^{\alpha-1})$, $0 < \alpha < 1$. *Then location coordinates expand with* $\max\{|\mathbf{s}_i - \bar{\mathbf{s}}_p| : \mathbf{s}_i \in \mathcal{S}_p\} = O(n_p^{\alpha})$.

Assumption 2. $\forall p \in \{1, \cdots, 8\}, \sup_{w_j \in \mathcal{W}} \sigma_p^2(w_j) < \infty$

Since $\epsilon_{i,j}$ is independent with non-constant variance $\sigma_{p_i}^2(w_j)$,

$$\mathrm{Var}\left\{\sum_{\mathbf{s}_i \in \mathcal{S}_p} (\mathbf{s}_i - \bar{\mathbf{s}}_p)\epsilon_{i,j}\right\} = \sigma_p^2(w_j) \sum_{\mathbf{s}_i \in \mathcal{S}_p} (\mathbf{s}_i - \bar{\mathbf{s}}_p)^2. \tag{A.6}$$

Under assumption 2,

$$\frac{\sum_{\mathbf{s}_i \in \mathcal{S}_p} (\mathbf{s}_i - \bar{\mathbf{s}}_p)\epsilon_{i,j}}{\sum_{\mathbf{s}_i \in \mathcal{S}_p} (\mathbf{s}_i - \bar{\mathbf{s}}_p)^2} = O_p \left\{ 1 / \sqrt{\sum_{\mathbf{s}_i \in \mathcal{S}_p} (\mathbf{s}_i - \bar{\mathbf{s}}_p)^2} \right\} = O_p(n_p^{-\alpha - 1/2}). \tag{A.7}$$

To prove the second term in (A.5) converge, 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 kth principal component scores. We introduce the mixing coefficient $\alpha_{p,q}^k(h)$ of $\boldsymbol{\xi}_k$ (Guyon (1995)),

$$\begin{aligned}\alpha_{p,q}^k(h) =& \sup\{|P(A \cap B) - P(A)P(B)|, A \in \mathcal{F}(\boldsymbol{\xi}_k, \Lambda_1), B \in \mathcal{F}(\boldsymbol{\xi}_k, \Lambda_2),\\ & |\Lambda_1| \le p, |\Lambda_2| \le q, dist(\Lambda_1, \Lambda_2) \ge h\}\end{aligned}$$

where $dist(\Lambda_1, \Lambda_2)$ 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 $h \to \infty$, $\alpha_{p,q}^k(h) = O(h^{-\delta})$ with $0 < \delta < 1$.*

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 $1 \le k \le K$.*

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

$$\sum_{j \ne i}^{n_p} |\mathrm{Cov}\{\xi_k(\mathbf{s}_i), \xi_k(\mathbf{s}_j)\}| \le \sum_{j \ne i}^{n_p} 4C^2 \alpha_{p,q}^k(h_j)$$

where $h_j = dist(\mathbf{s}_i, \mathbf{s}_j)$ and is arranged in increasing order. By assumption 1 and 3,

$$\begin{aligned}&\sum_{j \ne i}^{n_p} |\mathrm{Cov}\{\xi_k(\mathbf{s}_i), \xi_k(\mathbf{s}_j)\}|\\ =&\sum_{j \ne i}^{n_p} O(h_j^{-\delta}) = \sum_{j \ne i}^{n_p} O\{(jh_{n_p})^{-\delta}\}\\ =&O(n_p^{1-\alpha\delta})\end{aligned} \tag{A.8}$$

Hence,

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

Further, $\mathrm{E}\{\xi_k(\mathbf{s}_i)\}=0$,

$$
\sum_{\mathbf{s}_i\in\mathcal{S}_p}(\mathbf{s}_i-\bar{\mathbf{s}}_p)\xi_k(\mathbf{s}_i)=O_p(n_p^{1+\alpha-\alpha\delta/2}) \tag{A.9}
$$

And similarly, $\mathrm{Var}\left\{\sum_{\mathbf{s}_i\in\mathcal{S}_p}\xi_k(\mathbf{s}_i)\right\}=O(n_p^{2-\alpha\delta})$, so

$$
\frac{1}{n_p}\sum_{\mathbf{s}_i\in\mathcal{S}_p}\xi_k(\mathbf{s}_i)=O_p(n_p^{-\alpha\delta/2}). \tag{A.10}
$$

By (A.9), the second term of (A.5) converges as following.

$$
\begin{aligned}
&\frac{\sum_{\mathbf{s}_i\in\mathcal{S}_p}(\mathbf{s}_i-\bar{\mathbf{s}}_p)\sum_{k=1}^K\xi_k(\mathbf{s}_i)\phi_k(w_j)}{\sum_{\mathbf{s}_i\in\mathcal{S}_p}(\mathbf{s}_i-\bar{\mathbf{s}}_p)^2}\\
=&\frac{\sum_{k=1}^K\phi_k(w_j)\sum_{\mathbf{s}_i\in\mathcal{S}_p}(\mathbf{s}_i-\bar{\mathbf{s}}_p)\xi_k(\mathbf{s}_i)}{\sum_{\mathbf{s}_i\in\mathcal{S}_p}(\mathbf{s}_i-\bar{\mathbf{s}}_p)^2}\\
=&O_p(n_p^{-\alpha-\alpha\delta/2}) \qquad \text{(A.11)}
\end{aligned}
$$

Thus, $\widehat{\beta}_1^p(w_j)-\beta_1^p(w_j)=O_p(n_p^{-\alpha-\alpha\delta/2})$. For intercept estimate $\widehat{\beta}_0^p(w_j)$, by A.10

$$
\begin{aligned}
\widehat{\beta}_0^p(w_j)-\beta_0^p(w_j)&=\frac{1}{n_p}\sum_{\mathbf{s}_i\in\mathcal{S}_p}u(w_j;\mathbf{s}_i)\\
&=\sum_{k=1}^K\phi_k(w_j)\frac{1}{n_p}\sum_{\mathbf{s}_i\in\mathcal{S}_p}\xi_k(\mathbf{s}_i)+\frac{1}{n_p}\sum_{\mathbf{s}_i\in\mathcal{S}_p}\epsilon_{i,j}\\
&=O_p(n_p^{-\alpha\delta/2}).
\end{aligned}
$$

Then consistency of the parameters is obtained,

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

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

$$||\widehat{\boldsymbol{\beta}}(w_j) - \boldsymbol{\beta}(w_j)||_2 = O_p(n^{-\alpha\delta/2}) \tag{A.12}$$

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

$$\begin{aligned}
&\sup_{w_j \in \mathcal{W}} |\widehat{\mu}(w_j; \mathbf{s}_i) - \mu(w_j; \mathbf{s}_i)| \\
\leq &\sup_{w_j \in \mathcal{W}} |\widehat{\beta}_0^{p_i}(w_j) - \beta_0^{p_i}(w_j)| + \sup_{w_j \in \mathcal{W}} |\{\widehat{\beta}_1^{p_i}(w_j) - \beta_1^{p_i}(w_j)\}(\mathbf{s}_i - \bar{\mathbf{s}}_p)| \\
= &O_p(n^{-\alpha\delta/2}) + O_p(n^{-\alpha\delta/2}) = O_p(n^{-\alpha\delta/2}) \qquad \text{(A.13)}
\end{aligned}$$

The proof of theorem 4.1 is done.

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

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

where

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

$$\mathcal{R}_{2,n} = \frac{1}{n-1}\sum_{i=1}^{n}\left\{\sum_{k=1}^{K}\xi_k(\mathbf{s}_i)\phi_k(w_j)\right\}\left\{\sum_{k=1}^{K}\xi_k(\mathbf{s}_i)\phi_k(w_{j'})\right\} - \sum_{k=1}^{K}\lambda_k\phi_k(w_j)\phi_k(w_{j'})$$

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

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

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

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

For $\mathcal{R}_{2,n}$,

$$
\begin{aligned}
\mathcal{R}_{2,n} =& \frac{1}{n-1}\sum_{i=1}^{n}\left\{\sum_{k_1=1}^{K}\sum_{k_2=1}^{K}\xi_{k_1}(\mathbf{s}_i)\xi_{k_2}(\mathbf{s}_i)\phi_{k_1}(w_j)\phi_{k_2}(w_{j'})\right\} \\
&- \sum_{k=1}^{K}\lambda_k\phi_k(w_j)\phi_k(w_{j'}) \\
=& \sum_{k=1}^{K}\left\{\frac{1}{n-1}\sum_{i=1}^{n}\xi_k^2(\mathbf{s}_i) - \lambda_k\right\}\phi_k(w_j)\phi_k(w_{j'}) \qquad \text{(A.14)} \\
&+ \sum_{k_1=1}^{K}\sum_{k_2\neq k_1}^{K}\phi_{k_1}(w_j)\phi_{k_2}(w_{j'})\frac{1}{n-1}\sum_{i=1}^{n}\xi_{k_1}(\mathbf{s}_i)\xi_{k_2}(\mathbf{s}_i) \qquad \text{(A.15)}
\end{aligned}
$$

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

$$
\gamma_{p,q}^{k}(h) \le \alpha_{p,q}^{k}(h).
$$

Thus, we have the same result as (A.8). By assumption 4, let $\mathrm{Var}\left\{\xi_k^2(\mathbf{s}_i)\right\} = \tau_k < \infty$,

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

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

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

Then it follows (A.14) is $O_p(n^{-\alpha\delta/2})$. Since $\mathrm{E}\xi_{k_1}(\mathbf{s}_i)\xi_{k_2}(\mathbf{s}_i) = 0$, and

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

So it follows that (A.15) is $O_p(n^{-\alpha\delta/2})$ as well. Thus $\mathcal{R}_{2,n} = O_p(n^{-\alpha\delta/2})$.

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

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

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

Assumption 5. *For any kth 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,* $\mathrm{E}\{\xi_k(\mathbf{s}+h) - \xi_k(\mathbf{s})\}^2 = O(h^\beta)$ *as $h \to 0$, $\beta > 0$.*

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

$$\begin{aligned}
\Delta_i(w_j) =& r(w_j;\mathbf{s}_{i+2}) - 2r(w_j;\mathbf{s}_{i+1}) + r(w_j;\mathbf{s}_i)\\
=&\beta_1^p(w_j)(\mathbf{s}_{i+2} - 2\mathbf{s}_{i+1} + \mathbf{s}_i) + \sum_{k=1}^{K}\{\xi_k(\mathbf{s}_{i+2}) - 2\xi_k(\mathbf{s}_{i+1}) + \xi_k(\mathbf{s}_i)\}\phi_k(w_j)\\
&+ \epsilon_{i+2,j} - 2\epsilon_{i+1,j} + \epsilon_{i,j}
\end{aligned}$$

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

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

By assumption 1 and 2,

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

Thus,

$$\begin{aligned}
\widehat{\sigma}_p^2(w_j)-\sigma_p^2(w_j)=&\frac{1}{6(n_p-2)}\sum_{\mathbf{s}_i\in\mathcal{S}_p}\Delta_i^2(w_j)-\sigma_p^2(w_j)\\
=&O_p(n_p^{2\beta(\alpha-1)}+n_p^{\alpha-1}+n_p^{-1/2})
\end{aligned}\tag{A.16}$$

Assuming $n_p=n/8$, it follows that

$$\begin{aligned}
\frac{1}{n-1}\sum_{i=1}^{n}\epsilon_{i,j}^2-\widehat{\sigma}^2(w_j)=&\frac{1}{n-1}\sum_{p=1}^{8}\left[\sum_{i\in\mathcal{S}_p}\{\epsilon_{i,j}^2-\widehat{\sigma}_p^2(w_j)\}\right]\\
=&O_p(n^{2\beta(\alpha-1)}+n^{\alpha-1}+n^{-1/2})
\end{aligned}\tag{A.17}$$

Based on A.17, we can obtain $\mathcal{R}_{3,n}=O_p(n^{2\beta(\alpha-1)}+n^{\alpha-1}+n^{-1/2})$ for both $j=j'$ and $j\neq j'$. Finally,

$$\begin{aligned}
&\sup_{w_j,w_{j'}\in\mathcal{W}}|\widehat{R}_w(w_j,w_{j'})-R_w(w_j,w_{j'})|\\
=&O_p(n^{-\alpha\delta/2}+n^{2\beta(\alpha-1)}+n^{\alpha-1}+n^{-1/2})
\end{aligned}\tag{A.18}$$

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. (2018). Mission CO_2ntrol: 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., Wennberg, P. O., Crisp, D., Schimel, D. S., Gunson, M. R., Chatterjee, A., Liu, J., Schwander, 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., Wennberg, 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.

Guyon, X. (1995). *Random fields on a network: modeling, statistics, and applications*. Springer Science & Business Media.

Ibragimov, I. (1975). Independent and stationary sequences of random variables. *Wolters, Noordhoff Pub.*

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., Wennberg, 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.

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