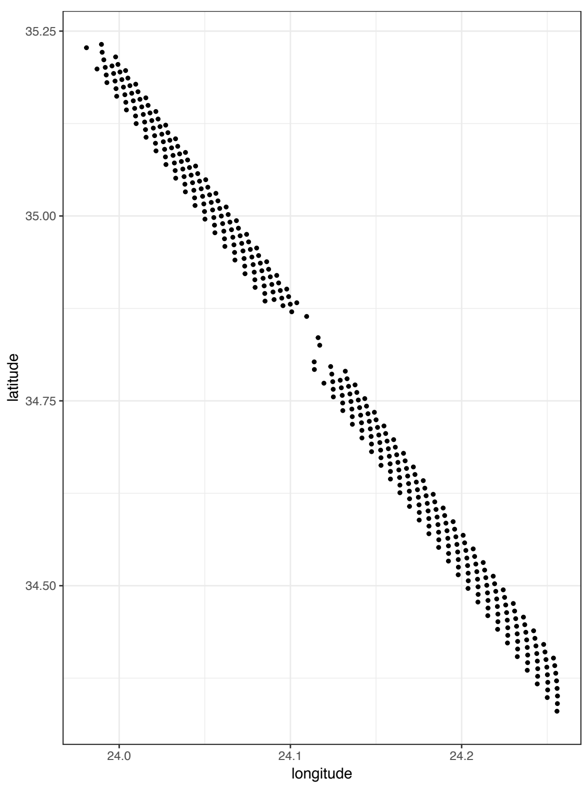
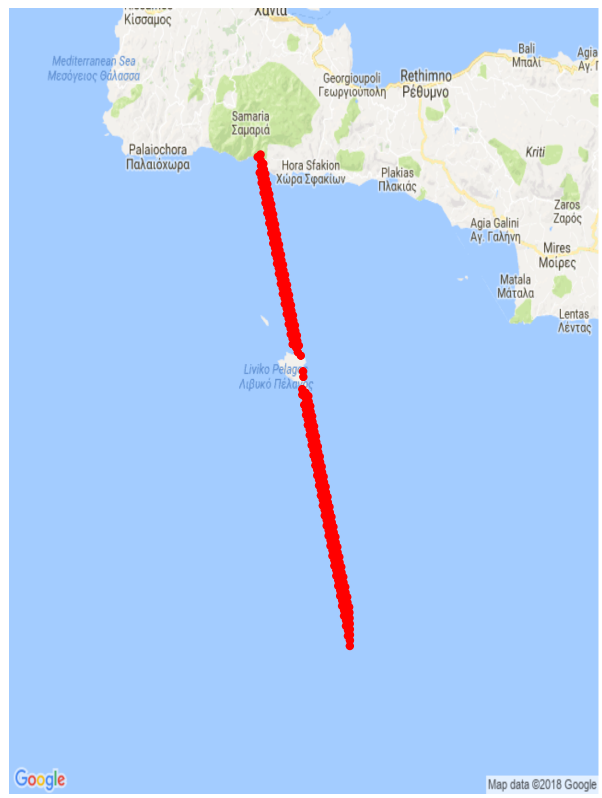
OCO2 Data Analysis

Xinyue Chang

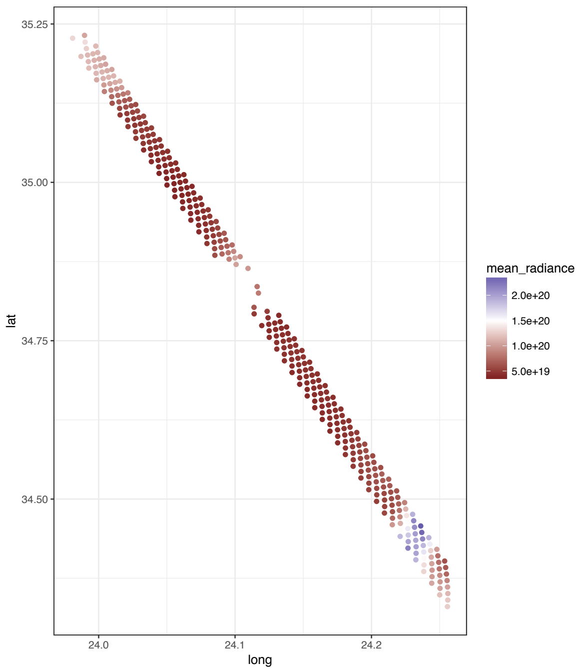
November 12, 2018

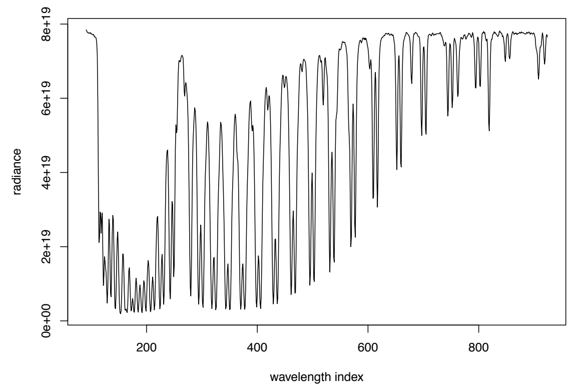
1. Data We select only water area with latitude less than 35.4 of orbit 10575, remove frequency with missing locations (since too many missing values when taking average leads to biased estimate mean function) and locations with frequency all missing, only keep the wavelength corresponding to O_2 band. Hence there are no missing values in the data after pre-processing.





In summary, there are 341 locations with 827 dimensional radiance each. Considering large variance of measurement error, there are truely some certain area where radiance values are not spatially smooth. Here is a heatmap of mean radiance and typical radiance function (O_2 band) for a spatial location.



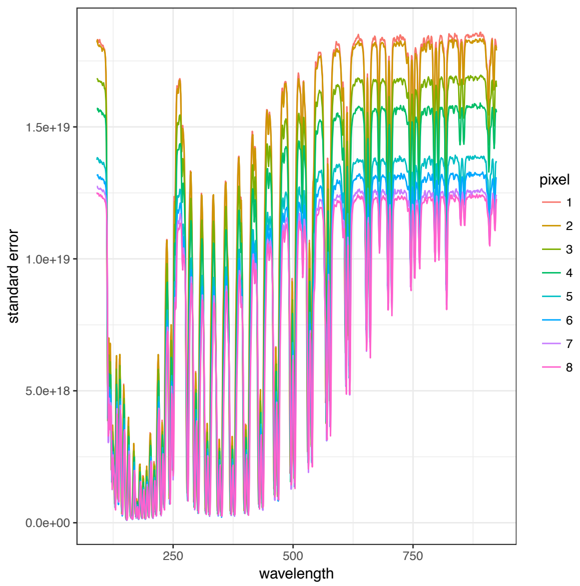


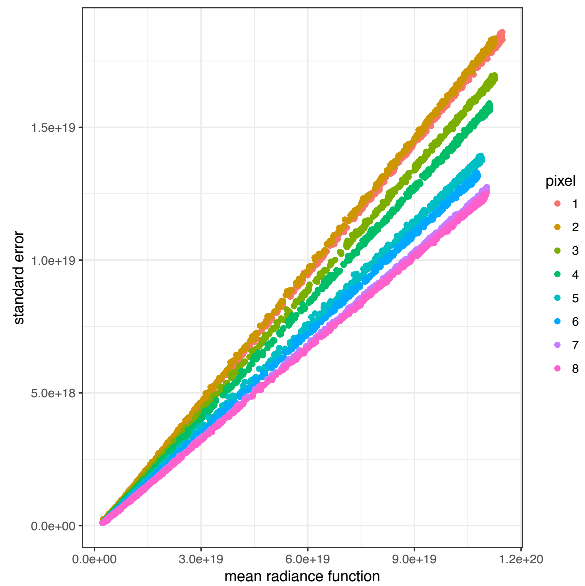
2. Basic setup Notations used in the following.

- $\mathbf{s}_i$: ith location including latitude and longitude, i.e., $(longitude, latitude)$
- w_j: jth wavelength index, $j = 1, 2, \cdots, 827$ and $1 \leq j \leq 1016$.
- $r(w_j, \mathbf{s}_i)$: observed radiance of wavelength index w_j at location $\mathbf{s}_i$.
- $f(w_j, \mathbf{s}_i)$: radiance of wavelength index w_j at location $\mathbf{s}_i$.
- $\sigma_p^2(w_j)$: variance of measurement error when wavelength is at jth index and location belongs to pixel p.
- N_p: number of subjects for pixel p.
- N: number of total subjects (for this data set: $N = 341$).
- m: number of observations for each subject (for this data set: $m = 827$).
- $\xi_k(\mathbf{s}_i)$: kth principle component score at location $\mathbf{s}_i$.

3. Measurement error It is easy to understand that spatial 2nd order difference measurement error estimate is better than frequencial 2nd order difference measurement error estimate. Since we can see that the radiance function is not smooth and change rapidly across wavelength, but it is much more reasonable to assume that radiance function does not vary too much if two locations are not far apart.

After looking into the data, we found that measurement error depends on pixels and radiance. Here are plots for illustration.





Thus, to estimate measurement error σ_2 in FPCA stage (more details below), we need to use the formula as following. For each pixel $p = 1, 2, \cdots, 8$,

$$\widehat{\sigma}_p^2(w_j) = \frac{1}{6}\frac{1}{N_p - 2}\sum_{1 \le i \le N_p} \{r(w_j, \mathbf{s}_{i+2}) - 2r(w_j, \mathbf{s}_{i+1}) + r(w_j, \mathbf{s}_i)\}^2$$

4. FPCA on radiance of wavelength The first stage is considering measured radiance as a function of wavelength. As we found in the dataset, wavelengths are increasing with index except for some extreme points and really dense across index. So it is equivalent to consider measured radiance as a function of index 1-1016.

$$r(w_j, \mathbf{s}_i) = f(w_j, \mathbf{s}_i) + \epsilon_{i,j}$$
$$f(w_j, \mathbf{s}_i) = \mu(w_j) + \sum_{k=1}^{\infty} \xi_k(\mathbf{s}_i)\phi_k(w_j)$$

$f(w_j, \mathbf{s}_i)$ is assumed to be i.i.d stochastic process across $\mathbf{s}_i$, $\epsilon_{i,j}$ is uncorrelated with mean 0 and variance $\sigma_p^2(w_j)$ where p is what $\mathbf{s}_i$'s pixel is.

For dense functional data, we can estimate mean function $\mu(w)$ by simply taking the average. That is

$$\widehat{\mu}(w_j) = \frac{1}{N}\sum_{i=1}^{N} r(w_j, \mathbf{s}_i)$$

But here to mention two concerns regarding mean function estimate. Although our mean function $\mu(w_j)$ does not depend on spatial locations, it is true that radiance function differs spatially (like temprature of higher latitude is lower). And principle components are correlated within a certain area. So we should do a local fit for FPCA model, i.e., select a smaller area instead of the whole data as our input for FPCA. Here is a roughly designed routine for local data selection.

- FPCA using all the data and approximately determine the significant range of variogram of the first principle component score, say r.

- Suppose we want to predict radiance at location $\mathbf{s}_0 = (l_1, l_2)$. Then we select data with distance of r from $\mathbf{s}_0$. This can also be simplified as latitude between $l_2 - l_0$ and $l_2 + l_0$ if distance resulted from latitude difference l_0 is r.

The other thing noticed is that direct averaging is better than inverse variance weighted average for mean function estimation. Although, we found that measurement error is different for different pixels which can be regarded as groups, inverse variance weighted average of observed radiance function is not reasonable. Please see Appendix for details.

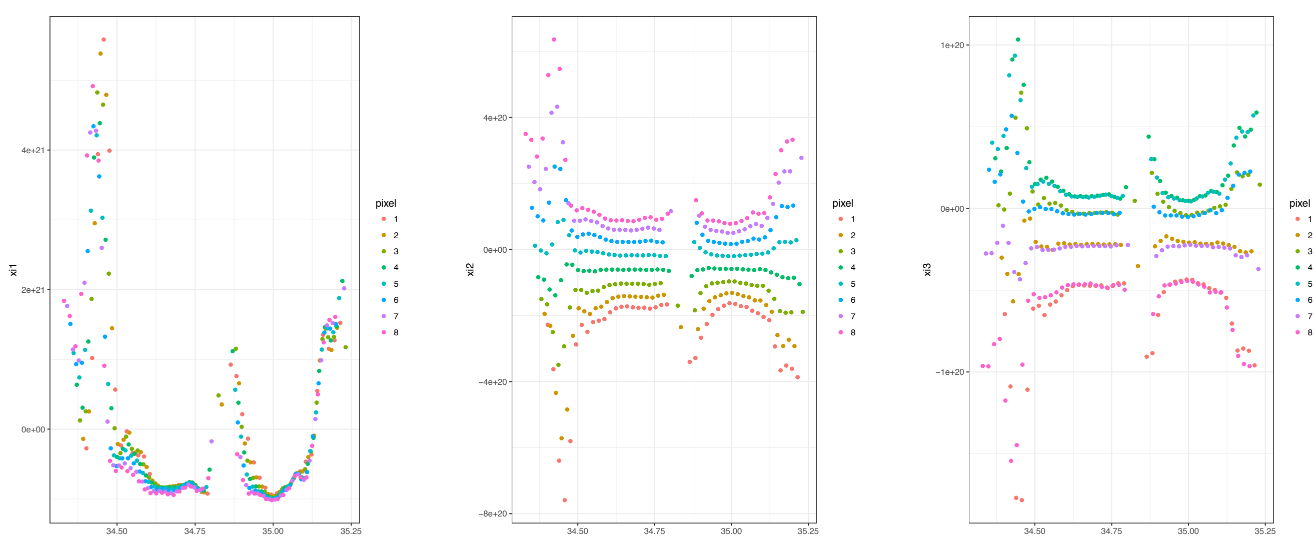
And the covariance function $G(w_1, w_2) = Cov\left\{f(w_1, \mathbf{s}_i), f(w_2, \mathbf{s}_i)\right\}$ is calculated as following (also use local data as inputs).

$$\begin{aligned}
\widehat{G} = &\left[\frac{1}{N}\sum_{i=1}^{N}\left\{r(w_j, \mathbf{s}_i) - \widehat{\mu}(w_j)\right\}\left\{r(w_l, \mathbf{s}_i) - \widehat{\mu}(w_j)\right\}\right]_{j,l=1}^{m} \\
&- diag\{\widehat{\sigma}^2(w_j) : j = 1, 2, \cdots, m\} \\
\widehat{\sigma}^2(w_j) = &\frac{1}{8}\sum_{p=1}^{8}\widehat{\sigma}_p^2(w_j)
\end{aligned}$$

Some theoretical support of total measurement error estimate $\widehat{\sigma}^2(w_j)$ can be found in Appendix.

Then we can calculate eigenfunctions and eigenvalues from the $\widehat{G}$. And K is set as the truncation to which point having variance explained by 0.9999. Also for dense functional data, we use numerical integration to estimate the corresponding FPC scores $\xi_k(\mathbf{s}_i)$, i.e., $\widehat{\xi}_k(\mathbf{s}_i) = \int\left\{r(w, \mathbf{s}_i) - \widehat{\mu}(w)\right\}\widehat{\phi}_k(w)dw$ for $k = 1, 2, \cdots, K$.

5. Spatial analysis of principle components Although we assumed spatial independence in FPCA stage, here we predict principle component score by analyzing spatial dependence among FPC scores estimated during FPCA. This contradiction will not be a critical issue if spatial dependence is weak enough under some extra condition, which will be discussed in theoretical level. Usually, not every principle component has strong spatial dependence, and these scores without much dependence shows a clear pattern across pixels.



For the principle component score $\xi_k(\mathbf{s})$ without much spatial dependence, it can be explained by a cell mean model with pixels treated as groups. Let $g(\mathbf{s})$ represents the pixel of location $\mathbf{s}$, then predictor at $\mathbf{s}_0$ can be written as $\widehat{p}(\xi_k; \mathbf{s}_0) = \sum_{p=1}^{8}\bar{\xi}_p(k)I\{g(\mathbf{s}_0) = p\}$.

For the principle component score $\xi_k(\mathbf{s})$ with strong spatial dependence, it is assumed to be a Random Field with unknown constant mean, second order stationary and isotropic. Then it holds that $C_k(h) = \text{Cov}\{\xi_k(\mathbf{s}+h), \xi_k(\mathbf{s})\}$ where h is defined as the shortest distance between two locations on earth. Estimating $C_k(h)$ is the key point for kriging in the next. First an exponential model is fitted to sample semivariogram $\widehat{\gamma}_k(h) = \dfrac{1}{2|N(h)|}\sum_{N(h)}\{\xi_k(\mathbf{s}) - \xi_k(\mathbf{s}+h)\}^2$. By *Statistical Methods for Spatial Data Analysis*, ordinary least squares and weighted least squares performed more or less equally well, and OLS indeed performed well after we tried several option for weighted matrix. So we can derive $\widehat{C}_k(h) = \gamma_k(\infty;\widehat{\boldsymbol{\theta}}) - \gamma_k(h;\widehat{\boldsymbol{\theta}})$ from fitted model $\gamma_k(h;\widehat{\boldsymbol{\theta}})$ obtained by OLS. Then the ordinary kriging predictor at location $\mathbf{s}_0$ is as following.

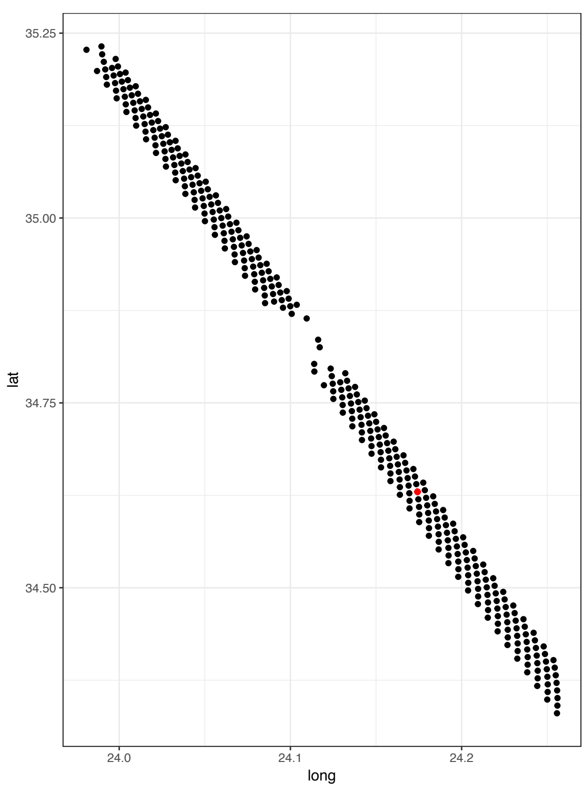
$$p(\boldsymbol{\xi}_k;\mathbf{s}_0) = \mu_k + \boldsymbol{\sigma}'_k\Sigma_k^{-1}(\boldsymbol{\xi}_k - \mathbf{1}\mu_k)$$

Based on $\widehat{C}_k(h)$ we obtained, it is able to calculate plug-in estimator $\widehat{\boldsymbol{\sigma}}_k$ and $\widehat{\Sigma}_k$. Hence,

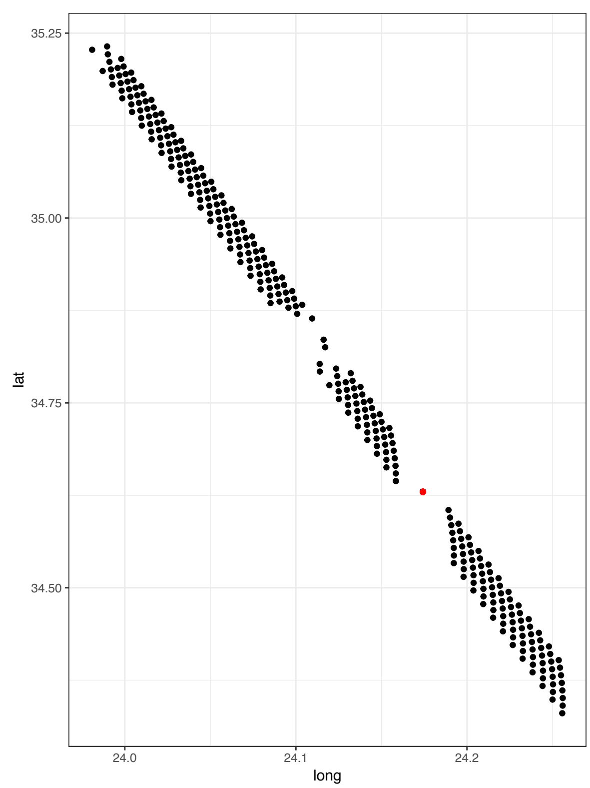
$$\widehat{p}(\boldsymbol{\xi}_k;\mathbf{s}_0) = \widehat{\mu}_k + \widehat{\boldsymbol{\sigma}}'_k\widehat{\Sigma}_k^{-1}(\boldsymbol{\xi}_k - \mathbf{1}\widehat{\mu}_k)$$

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

6. Prediction on gaps We removed one point from data and suppose to predict at this sounding which is 2016062711300034 at location $(24.1742, 34.62984)$. To look into how the method preforms under different situations, we also removed 1 to 10 tracks up and down to simulate cases with gaps absent around.



(a) The red dot is where prediction will make



(b) The case when 5 gaps are removed

Using the method talked above, we are able to predict radiance function at the red dot. Here we introduced two formula for evaluating how accurate our imputation is.

$$e_1 = \frac{1}{m} \sum_{j=1}^{m} \frac{(\widehat{f}(w_j, \mathbf{s}_0) - \widetilde{f}(w_j, \mathbf{s}_0))^2}{\widetilde{f}^2(w_j, \mathbf{s}_0)}$$

$$e_2 = \frac{1}{m} \sum_{j=1}^{m} \frac{(\widehat{f}(w_j, \mathbf{s}_0) - \widetilde{f}(w_j, \mathbf{s}_0))^2}{(\widetilde{f}(w_j, \mathbf{s}_0) - \widehat{\mu}(w_j))^2}$$

where the smoothing function $\widetilde{f}(w_j, \mathbf{s}_0)$ is supposed to be the estimate of true radiance function denoised from observed radiance. It can obtained by local polynomial, spline regression, or FPCA. The e_1 can be regarded as prediction error compared with the true radiance function. And e_2 is supposed to evaluate how good the variance component is predicted.

Here is the imputation compared with smoothing function calculated by FPCA using all data. First 6 e_1 are: 0.002861785, 0.003297388, 0.003704648, 0.003658406, 0.004395723, 0.006106047. And first 6 e_2 are: 0.004558063, 0.004815925, 0.005005032, 0.004644211, 0.006580023, 0.008675531.

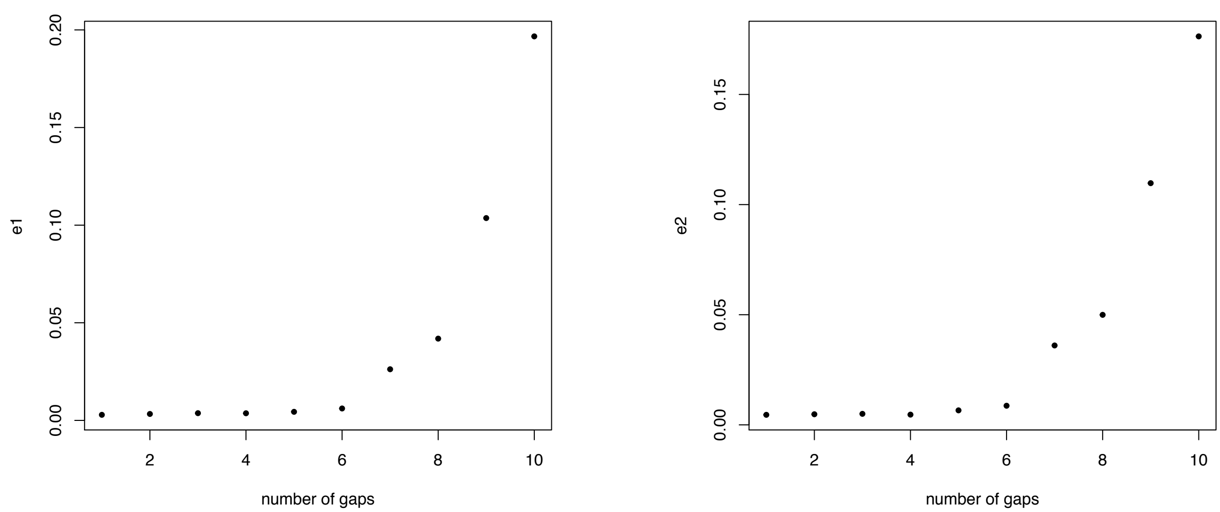


Figure 6: FPCA smoothing

And this is imputation compared with local polynomial estimator using 3 nearest neighborhoods which are assumed to have the same radiance function with prediction point. The local polynomial smoother is fitted with cross-validation bandwidth, gaussian kernel and degree = 1. First 6 e_1 are: 0.004702941, 0.005204274, 0.005672038, 0.005671203, 0.006376210, 0.008133074. And first 6 e_2 are: 0.009271686, 0.009400359, 0.009443673, 0.008781154, 0.010084702, 0.011842842.

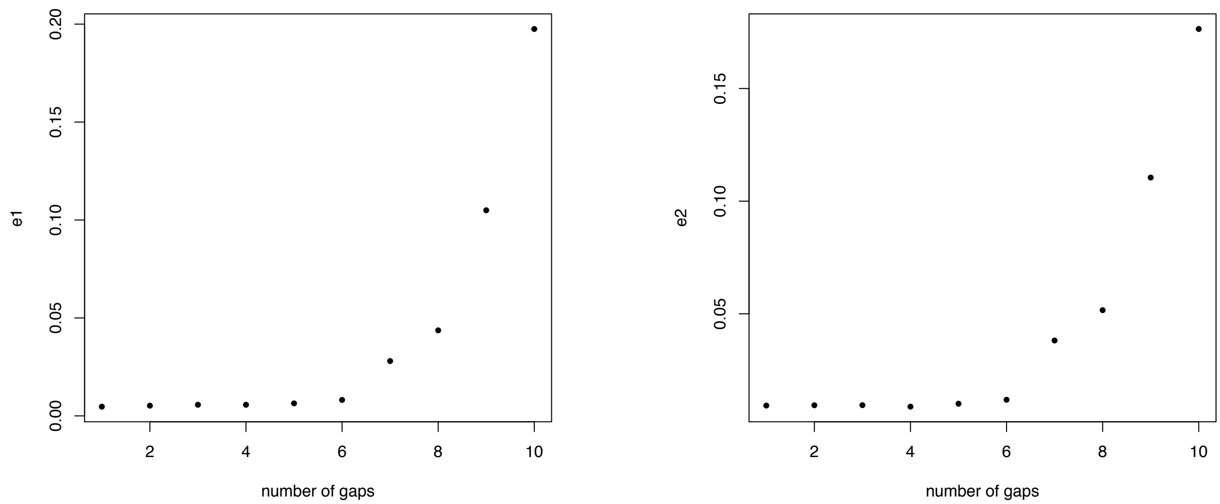
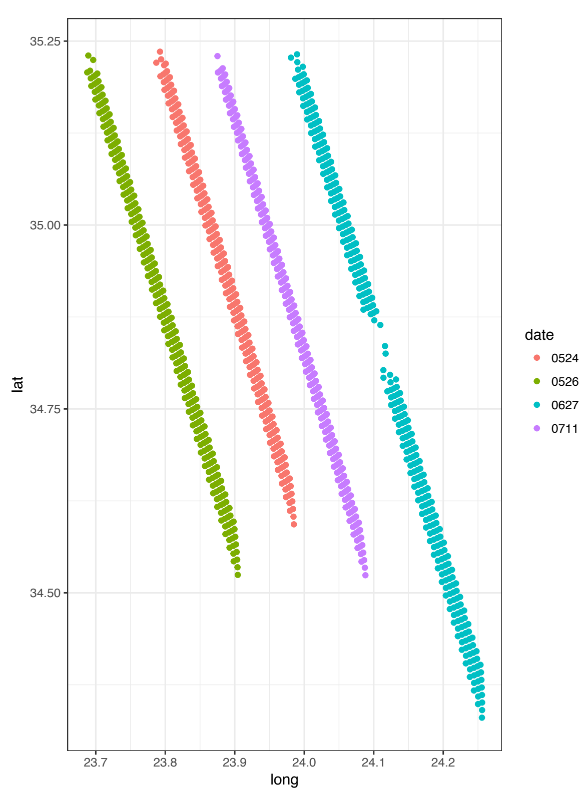
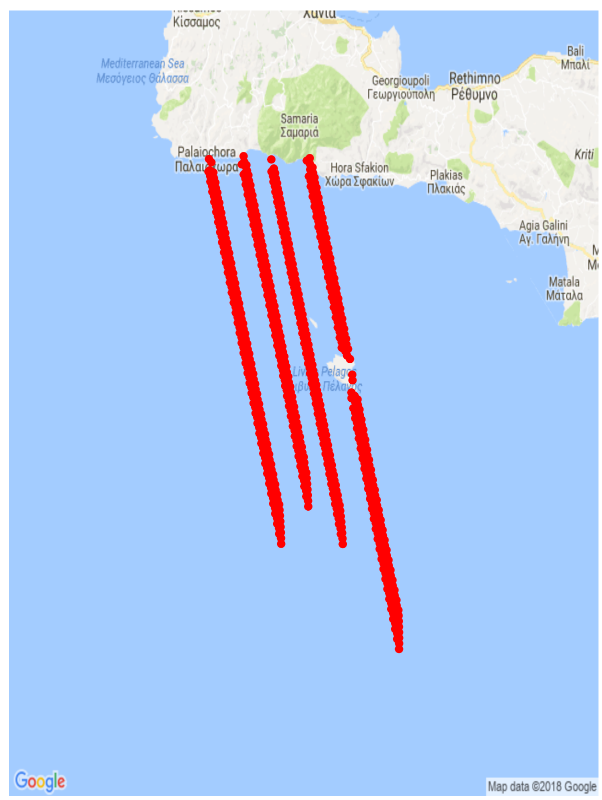


Figure 7: Local polynomial smoothing

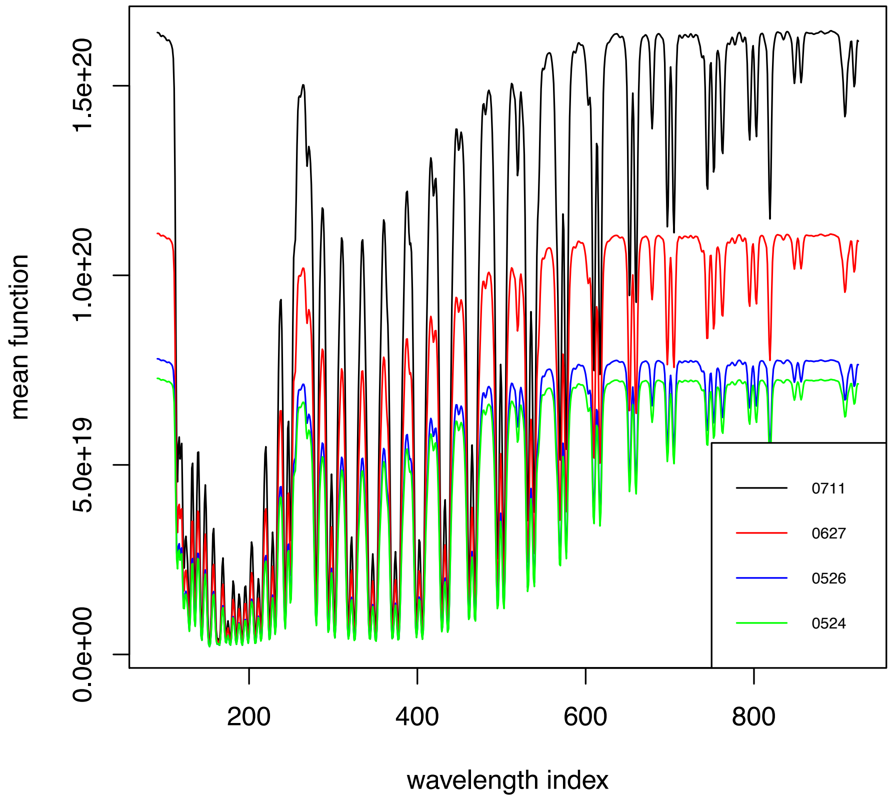
7. Time We can start by lookng into four tracks in water from different orbits (time).



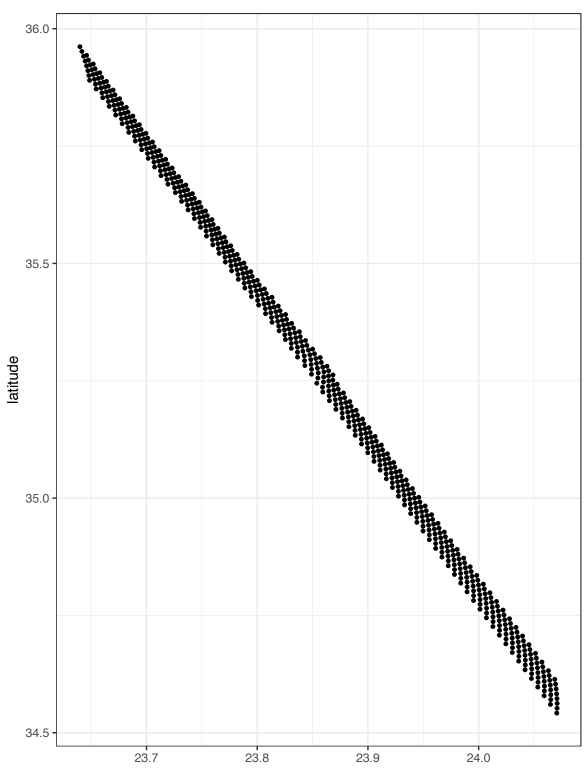


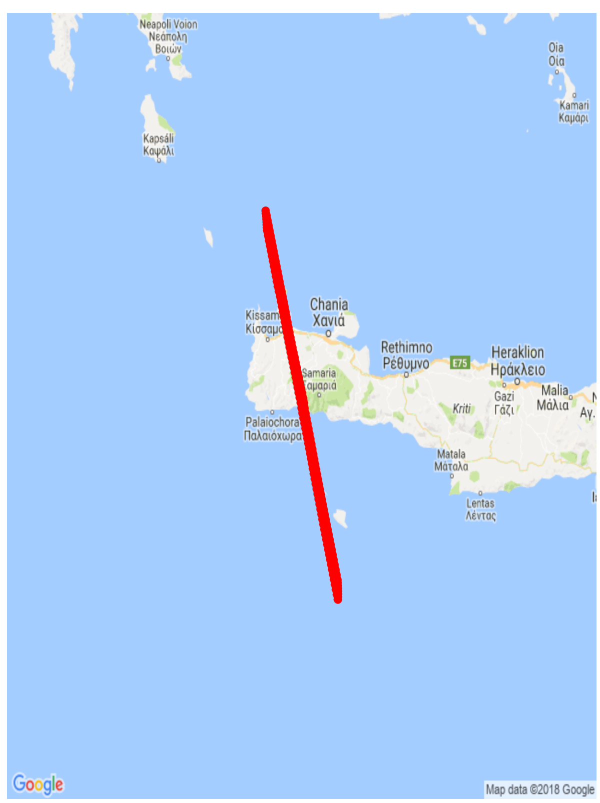
Unfortunately, they are not overlapping at all. The distance between two neighbor tracks are 7836.557, 6048.665, and 8748.324 in meters. While widths of four tracks are 1495.782, 1366.912, 1192.32, 1833.889 in meters, the gaps between gaps are not negligible.

However, radiance does vary across time. If we caculate mean function for each track, they are clearly separated by seasons.

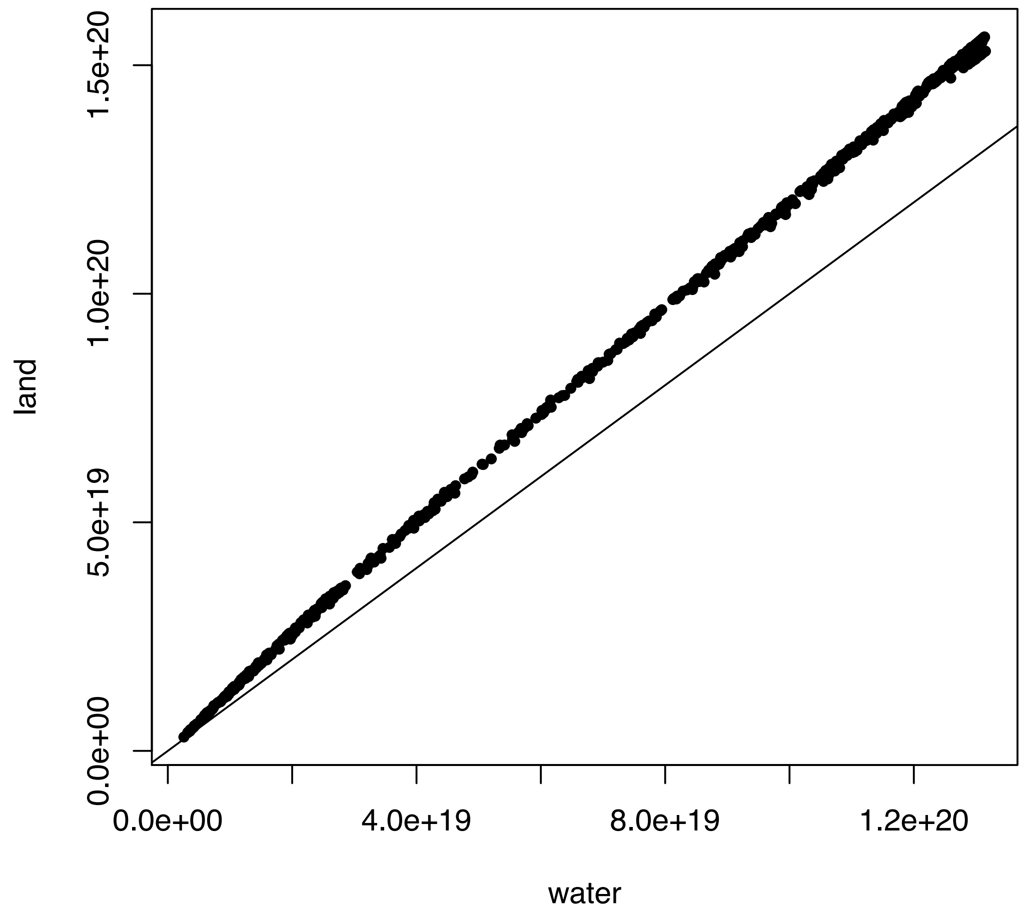


8. Water & Land Mixing We considered the 5216 orbit data which happened on 2015-6-25.

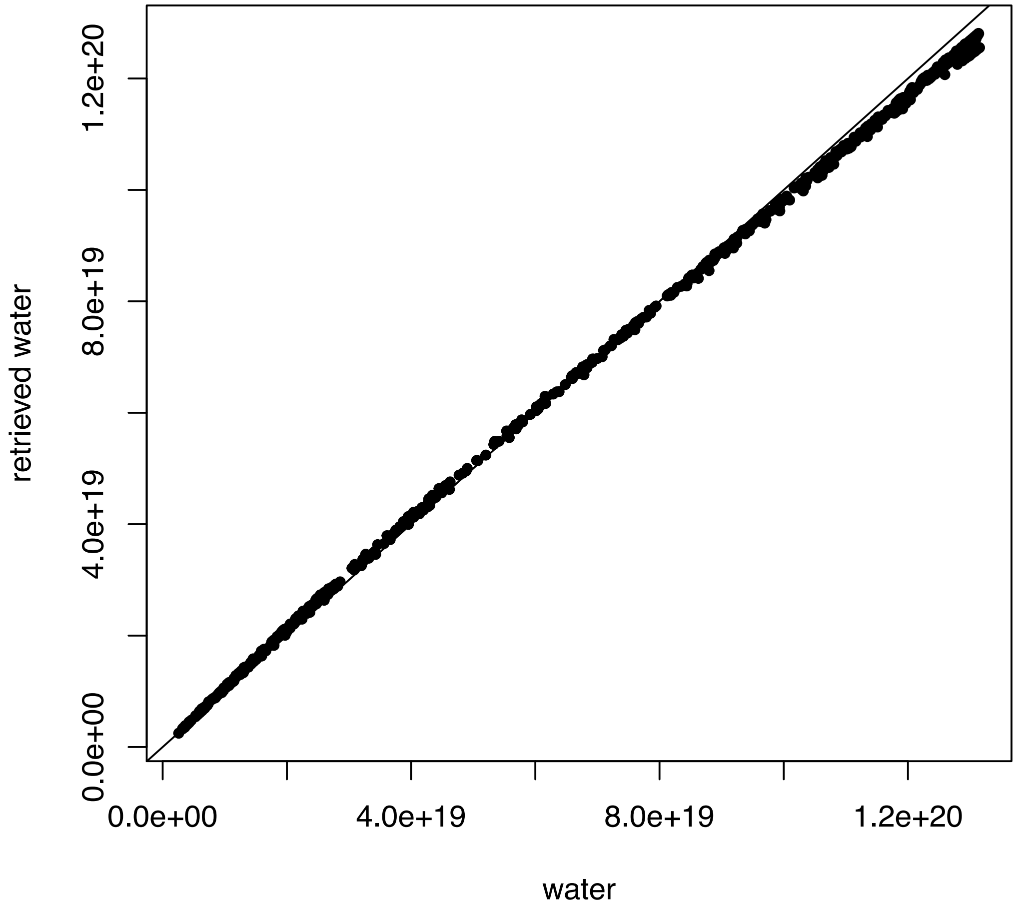




There is a relationship between land mean radiance and water mean radiance.



The slope for a simple regression is 1.171053. Considering that average of albedo is 0.21819, so we may be able to explain this relationship with the albedo. So if one land pixel was retrieved to water pixel, it can be obtained by $f_W(w_j, \mathbf{s}_i) = f_L(w_j, \mathbf{s}_i)/(1 + a(\mathbf{s}_i))$ where $a(\mathbf{s}_i)$ is the albedo of $\mathbf{s}_i$. This can also be understood like this. The albedo is 0 for water pixel, which means there exists no reflectance. But albedo is around 0.2 for land pixel, which means it would reflect about 20% radiance compared to water pixel. If we retrieve all land pixels back to water radiance by deviding (1+albedo), then the average is pretty much consistent with mean radiance from water area.



The slope from a simple regression is 0.9602408 which is close to 1.

Appendix

1. Spatial 2nd order measurement error estimation Given a pixel p, we restrict our data to be points fall into pixel p. So all the following $\mathbf{s}_i$ belongs to pixel p only and is ordered continuously along latitude.

$$\begin{aligned}
\Delta_2 &= r(w_j, \mathbf{s}_{i-1}) - 2r(w_j, \mathbf{s}_i) + r(w_j, \mathbf{s}_{i+1}) \\
&= f(w_j, \mathbf{s}_{i-1}) + \epsilon_{i-1,j} - 2f(w_j, \mathbf{s}_i) - 2\epsilon_{i,j} + f(w_j, \mathbf{s}_{i+1}) + \epsilon_{i+1,j} \\
&= \sum_{k=1}^{\infty} \{\xi_k(\mathbf{s}_{i-1}) - 2\xi_k(\mathbf{s}_i) + \xi_k(\mathbf{s}_{i+1})\} \phi_k(w_j) + \epsilon_{i-1,j} - 2\epsilon_{i,j} + \epsilon_{i+1,j}
\end{aligned}$$

By assumption, $\mathrm{E}(\Delta_2) = 0$. And $\mathrm{Var}\{\xi_k(\mathbf{s}_{i-1}) - 2\xi_k(\mathbf{s}_i) + \xi_k(\mathbf{s}_{i+1})\} \approx 0$ for all k.

$$\begin{aligned}
\mathrm{E}(\Delta_2^2) &= \mathrm{Var}(\Delta_2) \\
&\approx \sum_{k=1}^{\infty} 0 \times \phi_k^2(w_j) + \mathrm{Var}(\epsilon_{i-1,j} - 2\epsilon_{i,j} + \epsilon_{i+1,j}) \\
&= 6\sigma_p^2(w_j)
\end{aligned}$$

This gives the theoretical foundation of meausurement error estimation.

2. Mean function estimation discussion

$$\begin{aligned}
r(w_j, \mathbf{s}_i) &= \mu(w_j) + \sum_{k=1}^{\infty} \xi_k(\mathbf{s}_i)\phi_k(w_j) + \epsilon_{i,j} \\
&= \mu(w_j) + e_{i,j}
\end{aligned}$$

As we showed above, $\widehat{\sigma}_p^2(w_j)$ is supposed to be the estimate of variance of $\epsilon_{i,j}$ under our assumption. It is not directly related to $e_{i,j}$. However, to make the weighted inverse variance estimator work, $\dfrac{\sum_{p=1}^{8}\sum_{i=1}^{N_p} r(w_j, \mathbf{s}_{pi})/\widehat{\sigma}_p^2(w_j)}{\sum_{p=1}^{8} N_p/\widehat{\sigma}_p^2(w_j)}$, $\widehat{\sigma}_p^2(w_j)$ needs to be estimating variance of $e_{i,j}$. So this estimator is not reasonable and indeed does not work well (usually leads to lreally large error).

Meanwhile, the averaging of observed radiance function is much more robust and performs really well. Because unequal variance of $\epsilon_{i,j}$ won't result much difference among variance of $e_{i,j}$. And if our spatial dependence is weak, the average of observed radiance is able to converge to $\mu(w_j)$.

3. Covariance function estimation For any pixel $p = 1, 2, \cdots, 8$,

$$\begin{aligned}
\widehat{G}_p = &\left[\frac{1}{N_p}\sum_{i=1}^{N_p} \{r(w_j, \mathbf{s}_{pi}) - \widehat{\mu}(w_j)\}\{r(w_l, \mathbf{s}_{pi}) - \widehat{\mu}(w_j)\}\right]_{j,l=1}^{m} \\
&- diag\{\widehat{\sigma}_p^2(w_j) : j = 1, 2, \cdots, m\}.
\end{aligned}$$

Assume covariance function is homogeneous and N_p's are the same for all pixels,

$$\begin{aligned}
\widehat{G} = &\left[\frac{1}{N}\sum_{i=1}^{N} \{r(w_j, \mathbf{s}_i) - \widehat{\mu}(w_j)\}\{r(w_l, \mathbf{s}_i) - \widehat{\mu}(w_j)\}\right]_{j,l=1}^{m} \\
&- diag\left\{\frac{1}{8}\sum_{p=1}^{8} \widehat{\sigma}_p^2(w_j) : j = 1, 2, \cdots, m\right\}
\end{aligned}$$