SPATIOTEMPORAL SATELLITE DATA IMPUTATION
USING SPARSE FUNCTIONAL DATA ANALYSIS

BY WEICHENG ZHU, ZHENGYUAN ZHU*, AND XIONGTAO DAI†

Iowa State University

Many scientific applications and signal processing algorithms require complete satellite images. However, missing data in satellite images is very common due to various reasons such as cloud cover and sensor-specific problems. This paper introduces a general spatiotemporal satellite image imputation method based on sparse functional data analytic techniques. To handle observations consisting of a few longitudinally repeated satellite images that are themselves partially observed and noise-contaminated, we propose a multi-step imputation method by following the best linear unbiased prediction principle and pooling information across all available locations and time points. Theoretical properties are established for the proposed approach under a new observation model for functional data that covers the dataset in question as a special case. Practical analysis on the Landsat data are conducted to illustrate and validate our algorithm, which also shows that the proposed method considerably outperforms existing algorithms in terms of prediction accuracy. An efficient implementation using R and Rcpp is made available in the R package stfit.

1. Introduction. Remote sensing data are collected by remote sensors mounted on satellites or aircrafts to detect the energy reflected from Earth, which are recorded in the form of images. Satellite image data is a typical spatiotemporal remote sensing data, which usually consists of repeated measures of the same location over time.

Satellite data have many important applications, such as weather forecasting, land use/cover change assessment, environmental event monitoring, etc. Many applications require spatiotemporally complete images, but the observed satellite data often have many missing values due to cloud cover or sensor-specific problems. For example, one of the well-known sensor-specific problem is the scan line corrector (SLC) failure in the Landsat 7 Enhanced Thematic Mapper Plus (ETM+) sensor. The SLC is an electro-optical mech-

*Research supported in part by CCF 1934884 and cooperative agreement 68-7482-17-009 between the USDA Natural Resources Conservation Service and Iowa State University.
†Research supported in part by NSF grant DMS-2113713.

Keywords and phrases: Satellite, Imputation, Sparse Functional data, Spatiotemporal, Gap-fill, STFIT, R
anism composed of two parallel mirrors positioned behind the primary optics to compensate for the along track motion of the spacecraft during an across track scan (NASA, 2016). Without the SLC, the resulting images, referred to as SLC-off images, have about 22% missing data in a Landsat 7 scene (Wikipedia, 2018). Cloud cover is another common reason for missing data. Research study shows that about 67% of Earth’s surface is covered by clouds (King et al., 2013). Thus, efficient imputation algorithms for the missing values in the images are highly desirable in the context of satellite data.

In recent years, many algorithms have been proposed for missing value imputation in satellite images. One of the early image imputation algorithms was the local linear histogram-matching method developed by USGS (2004) for filling the gaps in SLC-off images. This method requires one or more reference images and does pixel-wise prediction by fitting a simple linear regression using similar pixels around the target missing pixel. This method is easy to implement, but it requires high quality reference images and does not perform well in areas with heterogeneous landscape. Later, the Neighborhood Similar Pixel Interpolator (NSPI) method was developed (Chen et al., 2011) and shown to be able to accurately fill gaps in SLC-off images even in areas with heterogeneous landscape, which needs to make use of reference images. In general, these methods depend on the manual selection of reference images and thus are not applicable in large scale satellite image imputation problems.

Geostatistical methods have been applied to fill the gaps for SLC-off images (Zhang et al., 2007; Pringle et al., 2009). These methods based on spatial kriging or co-kriging techniques are generally effective when the proportion of missing values in an image is small, but they do not perform well when the proportion of missing values is large (Zhang et al., 2007; Pringle et al., 2009).

Imputation methods that utilize only temporal information are also proposed (Roerink et al., 2000; Verger et al., 2013; Moreno et al., 2014). These methods perform smoothing separately for each pixel over the temporal domain and interpolate to fill in the missing values, ignoring correlations between pixels. These methods are computationally efficient and work reasonably well, though in general, the temporal-based methods are less effective compared with the spatiotemporal-based methods (Poggio et al., 2012; Zeng et al., 2014; Gerber et al., 2018), since the latter incorporates additional spatial information.

The spatiotemporal satellite images can be viewed as observations of an underlying spatio-temporal process, and spatio-temporal kriging is a natu-
ral way to impute missing data in space and time (Cressie and Wikle, 2015; Montero et al., 2015). Typically a parametric spatio-temporal covariance model needs to be assumed (Stein, 2005; Gneiting et al., 2006) for kriging. However, a direct application of a fully parametric approach to satellite data imputation problems is difficult due to the high computational costs of inverting large spatio-temporal covariance matrices, and the non-stationary nature of most satellite data even in local areas. Many approximation approaches have been proposed, e.g., Zhang et al. (2015); Katzfuss and Cressie (2011); Ma and Kang (2020), and this remains an area of active research in both algorithm and theory.

In this paper, we propose Spatio-Temporal Functional Imputation Tools (STFIT), a general spatiotemporal satellite image imputation method based on sparse functional data analysis (FDA). The proposed method is motivated by the need for imputation in Landsat data, in which case complete observations are rarely made in either the temporal or spatial domains. First, Landsat sensors revisit the same location every 16 days, with only around 23 images available per year. With two Landsat satellites functioning at the same time (e.g., Landsat 4/5 and 7, or Landsat 7 and 8), the number of available observed images per location per year is around 45. Second, a single Landsat image acquired on a cloudy day may have many missing values. This situation is analogous to a sparsely observed subject in a typical longitudinal dataset, for whom a few noisy and irregular measurements are available.

Under our modeling approach, we assume that the satellite images are observations from a latent spatiotemporal process contaminated with measurement errors. Precisely, for each imaging site and in each year, repeated longitudinal images are available, which are however only partially observed and contaminated with noise due to, e.g., cloud coverage. The goal of the proposed algorithm is to impute the underlying yearly spatial temporal image process using these noise contaminated partial images within each year. This observation model is new, to the authors’ knowledge, and is a generalization to the sparsely observed longitudinal data (Yao et al., 2005) and the partially-observed functional data models (Kraus, 2015).

Our imputation method applies and extends the Functional Principal Analysis by Conditional Estimation (PACE) proposed by Yao et al. (2005). The underlying idea for the proposed procedure is to impute a missing pixel by borrowing information from temporally and spatially contiguous pixels based on the best linear unbiased prediction (BLUP). The proposed algorithm estimates and imputes the underlying spatiotemporal process by aggregating three components: 1. the mean function, which contains the
overall spatiotemporal information; 2. the temporal effect, informing the year-specific deviation of each location from the mean; and 3. the spatial effect, which measures the image-specific deviation from the mean in each day-of-year. Missing values in the original data are then imputed by summing up all three components.

Model components are estimated by pooling data observed at similar locations or time points and then applying local polynomial smoothers. Uniform rates of convergence for the mean and the covariance functions for the temporal and spatial imputation are derived. These convergence rates match the best known rate (Li and Hsing, 2010; Zhang and Wang, 2016; Chen and Jiang, 2017) for sparsely observed functional data, which is extended here to repeated observations on a two-dimensional spatial domain with partial observations. Predictive properties for the STFIT imputation is also presented.

The proposed imputation method is efficient and computational fast since it utilizes the most pertinent spatial and temporal information, without assuming stationarity or separability of the underlying spatiotemporal process. In the Landsat data application, STFIT compared favorably against other existing methods in terms of imputation accuracy. Although the proposed method is illustrated with the Landsat data, it is generally applicable to impute other remote sensing data, such as Moderate Resolution Imaging Spectroradiometer (MODIS) and Advanced Very High Resolution Radiometer (AVHRR) data. Software implementation in R and C++ is available in the R package \texttt{stfit} (Zhu, 2020) for the ease of use by practitioners.

The remainder of the article is structured as follows. Section 2 gives a brief introduction to the Landsat data. Section 3 describes the proposed model used for imputation of spatiotemporal remote sensing data. Section 4 details the proposed estimation procedure and imputation algorithm. Theoretical results including imputation properties of STFIT and the rates of convergence for the estimation of model components are in Section 5. Our proposed method is illustrated on real Landsat data in Section 6 and compared against existing imputation algorithms. Finally, Section 7 summarizes our contributions and points out some potential future research directions.

2. Data and background. The motivation of our study is to borrow information from Landsat images to predict the land uses at the primary sampling units (PSU) of National Resources Inventory (NRI). NRI is a longitudinal survey of land use and natural resource conditions and trends on United States (US) non-federal lands. The most common PSU of the NRI sampling design is a 160-acre square quarter-section, 0.5 miles (0.8 km)
on each side. Details of the NRI dataset and its sampling design can be found in USDA (2015); Nusser and Goebel (1997); Nusser et al. (1998). The Landsat data is a collection of globally continuous, high-resolution (30 meters), and multispectral remote sensing satellite images, which finds wide applications in diverse fields such as agriculture, forestry, land use mapping, geology, hydrology, and environmental monitoring (NASA, 2019). Landsat scene p026r031 was selected as the study area, a 170 km north-south by 183 km east-west rectangle within the state of Iowa in the US. For the purpose of our research project, we broke down the Landsat scene into 245 small sites to match the sizes and locations of PSUs in NRI.

The dataset for each site is a sequence of $31 \times 31$ pixels Landsat surface reflectance images obtained by four satellites (Landsat 4/5, 7, and 8) between 2000 and 2015. Given that there were at least two satellites capturing images at the same time (e.g., Landsat 4/5 and Landsat 7 during the period 2003-2012, and Landsat 7 and Landsat 8 during period 2012-2015), the temporal gap between two observations at each site was shorter than the theoretical 16-day cycle of a single satellite. In this paper, we focus on the surface reflectance of the blue band (wavelength between 0.45 and 0.52 micrometers), which is Band 1 for Landsat 4/5/7 and Band 2 for Landsat 8. All values used in this paper were multiplied by 10,000 for easy reporting.

Figure 1 shows an example of all observed Landsat images by Landsat 7 and 8 satellites at a chosen site in 2015 for demonstration, after removing cloud, cloud shadow, and snow areas. Of all the 44 images taken in 2015, only 17 images have observations and 13 of those images have missing values. Among the partially observed images, some, such as the ones observed in day of the year (DOY) 26 and 90, have missing strips caused by the SLC failure in Landsat 7. Some other images, such as the one for DOY 66, have many spuriously large values outside of the normal range, as compared with the rest of the images. This is most likely due to the local weather condition, which results in a reduced performance of atmospheric correction and cloud removal.

A common use of Landsat data is to serve as explanatory variable for predicting the land use of NRI, and missing values in the Landsat data will seriously affect the prediction accuracy. In this paper, our primary task is to develop a general methodology to impute missing values in the Landsat data at 30 meter resolution. The algorithm is most suitable for gap filling Landsat images on sites where a long time series of historical images is available.

3. Data Models.
3.1. Notations and Conventions. Let $a \in \mathbb{R}^p$ and $b \in \mathbb{R}^q$ be Euclidean vectors. To simplify notation, denote $[a]_i$ as the $i$th entry of $a$, and for $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ let $f(a)$ be a vector with $i$th entry being $f([a]_i)$ and $g(a, b)$ as a matrix with $(i,j)$th entry being $g([a]_i, [b]_j)$, $i = 1, \ldots, p$, $j = 1, \ldots, q$. Also let $a^r = \prod_{i=1}^p [a]_i^{[r]}$ for $r \in \mathbb{R}^p$, and $\|a\|_1 = \sum_{i=1}^p |[a]_i|$ be the $L_1$-norm. For subset $A \subset \mathbb{R}^p$, let $|A|$ and $|A|_\#$ denote the Lebesgue measure and cardinality of $A$, respectively.

3.2. Observation Models. Let $S \subset \mathbb{R}^2$ be the common spatial support of the complete images at a single site, and $T = [0, 365]$ the time domain of days in year. We assume that the image observations have an underlying reflectance processes $X_i(s, t), s \in S, t \in T$, where $s$ is the pixel location, $t$ is the day-of-year, $i = 1, \ldots, n$ is the year, and $n$ is the total number of years. Write

$$X_i(s, t) = \mu(s, t) + R_i(s, t),$$

where $\mu(s, t) = E(X_i(s, t))$ is the mean function, and $R_i(s, t) = X_i(s, t) - \mu(s, t), i = 1, \ldots, n$ are zero-mean residual processes sharing the same covariance function

$$G((s, t), (s', t')) = \text{cov}(R_1(s, t), R_1(s', t')),$$

Fig 1: Observed Landsat 7 and 8 satellites images in year 2015 at Site B (used later as an example in Section 6), with missing values colored in black.
for location-time pairs $(s,t), (s',t') \in S \times T$. Available observations in year $i$ are made in the form of location-time-reflectance triplets

$$D_i = \{(s_k, T_{ij}, Y_{ijk}) \mid s_k \in O_{ij} \cap S_D, 1 \leq j \leq N_i, 1 \leq k \leq m\},$$

for $i = 1, \ldots, n$, where $O_{ij} = S \setminus C_{ij}$ is the $j$th random set of observed area in year $i$ where we make pixel observations, $C_{ij}$ is the random unobserved area (e.g., due to cloud cover) in which pixels are missing, $S_D := \{s_k \mid k = 1, \ldots, m\} \subset S$ is a fixed set of discrete pixel locations, $T_{ij} \in T$ is the $j$th random observation time in year $i$, and $Y_{ijk}$ is the noise-contaminated reflectance value at time $T_{ij}$ at pixel $s_k$; $N_i$ is the random number of images taken in year $i$, and $m$ is the number of pixels in each complete image. Note that the randomness in the observed pixel locations originates from the intersection of random observation area $O_{ij}$ and the non-random discrete set $S_D$ of pixel locations. The observed reflectance measure in a pixel is modeled as

$$Y_{ijk} = X_i(s_k, T_{ij}) + \epsilon_{ijk},$$

contaminated with i.i.d. noise $\epsilon_{ijk}$ with mean 0 and variance $\sigma^2$. Model components $N_i$, $T_{ij}$, $O_{ij}$, $X_i$, and $\epsilon_{ijk}$ are assumed to be all independent. Our proposed observation model consists of random longitudinal observation times $T_{ij}$ (James et al., 2000; Yao et al., 2005) and the partial spatial observations within $O_{ij}$ in each image. The spatial observation model accommodates the realistic scenario where functional data are observed on fixed discrete locations and generalizes the partially-observed functional data framework considered by Kraus (2015).

The model assumptions considered here are simple, and thus the models are particularly general and cover a wide range of data applications. Unlike in classical geostatistical analysis where stationarity in the underlying process and a parametric form for the covariance function are often required, we make no such assumptions and allow the spatio-temporal process to be nonstationary and nonparametric. The model components can be estimated fully nonparametrically here thanks to the availability of repeated image observations over time. Our fully nonparametric approach is more flexible than parametric approaches and thus is expected to yield imputation faithful to the observations.

4. Imputation procedure.
4.1. Overview. Under joint Gaussianity of the $X_i$ and $\epsilon_{ijk}$, the best prediction of $X_i(s,t)$ takes the form of

$$E[X_i(s,t) \mid D_i] = \text{cov}(X_i(s,t), Y_i \mid \tilde{s}_i, T_i) \text{cov}(Y_i, Y_i \mid \tilde{s}_i, T_i)^{-1}(Y_i - E[Y_i \mid \tilde{s}_i, T_i]),$$

where $\tilde{s}_i$, $T_i$, and $Y_i$ are vectors collecting the observation locations, days, and pixel values for year $i$, respectively, which are formed by stacking each of the three entries in an enumeration of $D_i$. Even if Gaussianity does not hold, (4) is the best linear unbiased prediction (BLUP) and thus is universally applicable without distributional assumptions. Ideally, one would like to impute $X_i(s,t)$ by

$$\tilde{X}_i(s,t) = E[X_i(s,t) \mid D_i],$$

given all the observations $D_i$ for year $i$. However, this approach is computationally burdened by the inversion of the large covariance matrix of all observed pixels within a year. Even for a small 3km $\times$ 3km region the number of all observed pixels within one year is in the order of 100,000 in the Landsat 7 data, precluding practical applications.

We propose a data- and computation-efficient imputation model, which entails the following five key steps: mean estimation, outlier detection, temporal effect estimation, spatial effect estimation, and imputation, which are detailed in Subsection 4.2–Subsection 4.6, respectively. The mean, temporal effect, and spatial effect estimation procedures involve the estimation of the model quantities, and the outlier detection step detects and removes outlier pixels and outlier images to robustify imputation results. The final imputation is performed through aggregating the mean, temporal, and spatial model components. The proposed algorithm can produce imputed values at any combination of spatial and temporal points in $S \times T$ through functional data analytic techniques. The imputation procedure is summarized in Figure 2. The dashed lines indicate that the temporal and spatial effects can be optionally performed or skipped. For example, for a completely missing image, the spatial effect cannot be calculated, and thus one can only use the mean or mean plus temporal effect to impute the image.

4.2. Mean estimation. The mean function $\mu(s,t)$ is defined on a three dimensional space $S \times T \subset \mathbb{R}^3$. For mean estimation, we propose to perform smoothing over only the temporal domain at each spatial location, obtaining estimate $\hat{\mu}(s_k, \cdot)$ that targets $\mu(s_k, \cdot)$, for each $s_k \in S_D$. This choice is data-driven and is supported by two reasons. First, smoothing directly on the 3-dimensional space is computationally intensive and might be subject to
a more severe curse of dimensionality. Second and more importantly, some observed areas displayed a certain degree of spatial inhomogeneity, and thus smoothing over the spatial domain may blur the transition boundaries.

Several smooth estimates \( \hat{\mu} \) for the mean function were implemented, such as local polynomial regression, defined for \( s_k \in S_D \), \( t \in T \) as

\[
\hat{\mu}(s_k, t) = \hat{\beta}_0,
\]

(6) \[
(\hat{\beta}_0, \ldots, \hat{\beta}_p) = \arg \min_{\beta_0, \ldots, \beta_p} \sum_{i=1}^{n} \sum_{j \in J_{ik}} [Y_{ijk} - \sum_{l=0}^{p} \beta_l(T_{ij} - t)]^2 K_{h_\mu}(T_{ij} - t),
\]

where the local polynomial coefficients \( \hat{\beta}_0, \ldots, \hat{\beta}_p \) are specific to \( (s_k, t) \), \( p \geq 0 \) is the degree of the local polynomial, \( h_\mu > 0 \) is the bandwidth, \( K_{h_\mu}(x) = K(x/h_\mu)/h_\mu \), \( K(\cdot) \) is a one-dimensional kernel function, and \( J_{ik} = \{ j \leq N_i \mid s_k \in O_{ij} \} \) is the index set of the observed time points at \( s_k \). A second option is to apply spline smoothing with the basis chosen as the B-spline (James et al., 2000) or Fourier basis (Ramsay and Silverman, 2005), for example, obtaining

\[
\hat{\mu}_H(s_k, t) = \eta_l^T (H^T H)^{-1} H^T Y_k,
\]

(7) where \( Y_k = [Y_{1k}^T, \ldots, Y_{nk}^T]^T \) is a column vector containing observed values at location \( s_k \), \( T_k \) is the concomitant time points for \( Y_k \), \( H = [\eta_{1k}, \ldots, \eta_{pk}] \) is the basis matrix, \( \eta_{lk} = \eta_l(T_k) \), and \( \eta_l, l = 0, \ldots, p \) are the basis functions. These methods had similar performance under proper selection of tuning parameters in our applications.
4.3. **Outlier detection.** The real data often contain abnormal outlying values. Outliers may seriously affect the estimated model and imputation accuracy, so it is very important to detect and remove outliers in our imputation procedure. Since the outliers beyond a normal data range in satellite data are typically removed in the preprocessing stage, the outliers detection procedure here (see Figure 2) mainly focuses on detecting the outliers that are within the normal data range. This covers the case for example, when there are light cloud cover which leads to distortion in the satellite measurements, but the outlyingness is not large enough to be detected in the preprocessing stage.

For each pixel, we first calculate the estimated residuals

\[
\hat{R}_{ijk} = Y_{ijk} - \hat{\mu}(s_k, T_{ij}),
\]

where \(\hat{\mu}\) is the estimated mean function defined in (6). A boxplot-based procedure is then applied. Pooling the estimated residuals at the same location, we remove data points beyond the boxplot whiskers, which extend 1.5 times the interquartile range from the upper and lower quartiles.

Based on exploratory analysis, an image with too many outliers is itself unreliable even if some of its pixels are not detected as outliers. We define images with outlying pixels greater than a certain threshold as outlier images and removed them from the the model estimation procedures to follow described shortly in Subsection 4.4 to Subsection 4.7. The threshold was empirically determined to be suitable for the analysis. With predetermined threshold 20%, DOY 234 in Figure 1 is an example of outlier image for which our algorithm detected 82.95% of its observed pixels are outliers; the abnormality of this image is confirmed by visual inspection. As a second example, DOY 66 had 14.85% of outlier pixels detected and removed, but the rest of the pixels were retained.

Although the mean estimation and outliers detection procedures may be performed iteratively until no outliers remain, in our experiments the iterative approach did not further remove outliers and improve imputation results. Thus, outlier detection is only performed once in the proposed procedure.

4.4. **Temporal effect estimation.** For each location, we pool all residuals from different years to imputes \(X_i(s, \cdot)\). Following the principal components analysis through conditional expectation (PACE) method (Yao et al., 2005), the temporal covariance function is first estimated, and then the temporal effect of the \(i\)th year is estimated from the observations in the \(i\)th year through functional principal component analysis (FPCA). While both the
mean and temporal effect estimation steps apply temporal smoothers, the temporal step provides year-specific estimation of the residual process for imputation, while the mean effect is common to all years.

For each year $i$ and pixel $s_k \in S_D$, recall that $J_{ik}$ is the index set of observed time points defined after (6), and denote $T_{ik} = [T_{ijk}]_{j \in J_{ik}}$ as the corresponding vector of the observed time points and $Y_{ik} = [Y_{ijk}]_{j \in J_{ik}}$ as the vector of corresponding reflectance values. At pixel $s_k \in S_D$, define the temporal covariance function and write its spectral decomposition as

$$G_{s_k}(t,t') := G((s_k, t), (s_k, t')) = \sum_{l=1}^{\infty} \lambda_{kl} \phi_{kl}(t) \phi_{kl}(t'),$$

where $(\lambda_{kl}, \phi_{kl})$ is the $l$th eigenvalue–eigenfunction pair associated with $G_{s_k}(t,t')$, $t, t' \in T$. The residual process $R_i(s_k, \cdot)$ and the original process $X_i(s_k, \cdot)$ can then be expanded in the basis spanned by the eigenfunctions through FPCA, obtaining

$$R_i(s_k, t) = \sum_{l=1}^{\infty} \xi_{ikl} \phi_{kl}(t), \quad X_i(s_k, t) = \mu(s_k, t) + \sum_{l=1}^{\infty} \xi_{ikl} \phi_{kl}(t),$$

where $\xi_{ikl} = \int_T R_i(s_k, t) \phi_{kl}(t) \, dt$ is the residual functional principal component (FPC). Given observations $(T_{ik}, Y_{ik})$ at $s_k$, the residual and the original processes are respectively predicted by their BLUPs

$$\tilde{R}_i(s_k, t) = \mathbb{E}[R_i(s_k, t) \mid s_k, T_{ik}, Y_{ik}] = \sum_{l=1}^{\infty} \tilde{\xi}_{ikl} \phi_{kl}(t),$$
$$\tilde{X}_i^T(s_k, t) = \mathbb{E}[X_i(s_k, t) \mid s_k, T_{ik}, Y_{ik}] = \mu(s_k, t) + \tilde{R}_i(s_k, t),$$

where the superscript $T$ of $\tilde{X}_i$ denotes that temporal information is incorporated in the imputation, and

$$\tilde{\xi}_{ikl} = \mathbb{E}[\xi_{ikl} \mid s_k, T_{ik}, Y_{ik}] = \lambda_{kl} \phi_{ikl}^T \Sigma_{ik}^{-1} R_{ik},$$

Here $\phi_{ikl} = \phi_{kl}(T_{ik})$, $\Sigma_{ik} = \text{cov}(Y_{ik}, Y_{ik} \mid s_k, T_{ik}) = G_{s_k}(T_{ik}, T_{ik}) + \sigma^2 I_{|J_{ik}|}$, $I_d$ is the identity matrix of size $d$; $R_{ik}$ is the residual vector corresponding to $Y_{ik}$, for which the entries contain the residuals

$$R_{ijk} = Y_{ijk} - \mu(s_k, T_{ij}).$$
To reduce computation and remove excess variation in the estimation process, we follow the PACE algorithm and approach (10)–(11) through truncated versions

\[
\tilde{R}_i(s_k, t) = \sum_{l=1}^{L_1} \tilde{\xi}_{ikl} \phi_{kl}(t),
\]

(14)

\[
\tilde{X}_i^T(s_k, t) = \mu(s_k, t) + \tilde{R}_i(s_k, t),
\]

(15)

where the number of components \(L_1\) is a positive integer practically set to be large enough to capture most of the variation in the residual process. While in practice the imputed processes (10)–(11) and (14)–(15) only need to be defined at discrete location \(s_k \in S_D\), for modeling and theory development we extend these definitions to \(s \in S\) by, for example, linearly interpolating the values of two neighboring locations at the same time \(t\).

Estimates \(\hat{G}_{s_k}^T(t, t')\) and \(\hat{\sigma}^2\) for the temporal covariance function \(G_{s_k}^T(t, t')\) and noise variance \(\sigma^2\) are constructed by local polynomial smoothing detailed in Appendix Subsection A.1. Spectral decomposition

\[
\hat{G}_{s_k}^T(t, t') = \sum_{l=1}^{\infty} \hat{\lambda}_{kl} \hat{\phi}_{kl}(t) \hat{\phi}_{kl}(t)
\]

gives the estimated temporal eigenvalue–eigenfunction pairs \((\hat{\lambda}_{kl}, \hat{\phi}_{kl})\). Prediction \(\hat{\xi}_{ikl}\) defined in (12) is estimated by \(\hat{\xi}_{ikl} = \hat{\lambda}_{kl} \hat{\phi}_{ikl}^T \hat{\Sigma}^{-1} \hat{R}_{ik}\), where \(\hat{\phi}_{ikl} = \hat{\phi}_{kl}(T_{ik})\), \(\hat{\Sigma}_{ij} = \hat{G}_{s_k}^T(T_{ik}, T_{ik}) + \hat{\sigma}^2 I_{|J_{ik}|}\), and \(\hat{R}_{ik}\) is a vector with entries being \(Y_{ijk} - \hat{\mu}(s_k, T_{ij})\). The estimate for \(\hat{X}_i^T(s_k, t)\) is then \(\hat{X}_i^T(s_k, t) = \hat{\mu}(s_k, t) + \sum_{l=1}^{L_1} \hat{\xi}_{ikl} \hat{\phi}_{kl}(t)\).

4.5. Spatial effect estimation. Residuals after adjusting for the mean and temporal effects still encode spatial dependency and are used to estimate the spatial effect, in order to further reduce the prediction errors. The goal of this step is to impute the temporal residual process \(V_i(s, t)\) from the temporal residuals \(U_{ijk}\), defined, respectively, as

\[
V_i(s, t) = X_i(s, t) - \tilde{X}_i^T(s, t) = R_i(s, t) - \tilde{R}_i(s, t),
\]

(16)

\[
U_{ijk} = Y_{ijk} - \tilde{X}_i^T(s_k, T_{ij}),
\]

(17)

for \(i = 1, \ldots, n\), \(j = 1, \ldots, N_i\), and \(k = 1, \ldots, m\) such that \(s_k \in O_{ij}\). We define the spatial covariance function for the temporal residual process \(V_i(s, t)\) at time \(t \in T\) as \(G_{s_i}^{ST}(s, s') = E[V_i(s, t)V_i(s', t)]\) and write its spectral decomposition as

\[
G_{s_i}^{ST}(s, s') = \sum_{l=1}^{\infty} \omega_l \psi_l(s) \psi_l(s'),
\]
where \((ω_{il}, ψ_{il})\) is the \(l\)th eigenvalue–eigenfunction pair associated with \(G_{ii}^{ST}(s, s')\). The temporal residual process \(V_i(\cdot, t)\) is expanded in the basis \(\{ψ_{il}\}_{l=1}^{∞}\), obtaining

\[
V_i(s, t) = \sum_{l=1}^{∞} ζ_{ilt}ψ_{il}(s),
\]

where \(ζ_{ilt} = \int S V_i(s, t)ψ_{il}(s) ds\) is the temporal residual FPC. Since Landsat satellites do not revisit the same location in the same DOY, for imputation at each time \(t\) we need to consider all temporal residuals falling within a small time window \(T_t = [t - h_0, t + h_0]\) for some \(h_0 > 0\). For the \(i\)th year and time \(t \in T\), define index set \(K_t = \{(j, k) | T_{ij} ∈ T_t, j = 1, \ldots, N_i, k = 1, \ldots, m, s_k ∈ O_{ij}\}\) which corresponds to the observations falling within an \(h_0\)-temporal neighborhood of \(t\), and column vectors \(s_{itl} = [s_k]_{k: (j, k) ∈ K_t}\), \(T_{ilt} = [T_{ij}]_{k: (j, k) ∈ K_t}\), and \(U_{ilt} = [U_{ijk}]_{k: (j, k) ∈ K_t}\) of the concomitant locations, time points, and temporal residuals, respectively. Targeting the BLUPs

\[
\tilde{V}_i(s, t) = \mathbb{E}[V_i(s, t) | \tilde{s}_{it}, T_{ilt}, U_{ilt}] = \sum_{l=1}^{∞} \tilde{ζ}_{ilt}ψ_{il}(s),
\]

\[
\tilde{ζ}_{ilt} = \mathbb{E}[ζ_{ilt} | \tilde{s}_{it}, T_{ilt}, U_{ilt}] = ω_{iit}ψ_{ilt}^TΞ_{it}^{-1}U_{it},
\]

the temporal residual process \(V_i(s, t)\) is imputed with \(L_2 ≥ 1\) components

\[
\tilde{V}_i(s, t) = \sum_{l=1}^{L_2} \tilde{ζ}_{ilt}ψ_{il}(s),
\]

where \(ψ_{ilt} = ψ_{il}(\tilde{s}_{it})\) is a column vector of length \(|K_t|\) and \(Ξ_{it} = \text{cov}(U_{ilt}, U_{it} | \tilde{s}_{it}, T_{ilt})\), for which the form is specified in (22).

For the estimation of spatial effects, we extend the PACE approach (Yao et al., 2005) to a two-dimensional space and pool all residual images from the small time window for estimation. One has

\[
\text{cov}(U_{ijk}, U_{ij'k'}) | T_{ij}, T_{ij'}, s_k, s_{k'}) = E\{[ε_{ijk} + V_i(s_k, T_{ij})][ε_{ij'k'} + V_i(s_{k'}, T_{ij'})] | T_{ij}, T_{ij'}, s_k, s_{k'}\}
\]

\[
= σ^2 I((j, k) = (j', k')) + a_t(s_k)I(k = k') + G_{ii}^{ST}(s_k, s_{k'}) + O(h_0)
\]

as \(h_0 → 0\), where \(a_t(s_k) = 2E[ε_{ijk}V_i(s_k, t)]\) is a location-specific variance term, and the last bias term is due to Taylor expansion and the binning of \(T_t\). Constructing estimated temporal residuals \(\tilde{U}_{it}\) for which the entries are of the form \(\tilde{U}_{ijk} = \tilde{R}_{ijk} - \tilde{R}_i(s_k, T_{ij})\), we estimate the spatial effect
components $\hat{G}^ST(s, s')$, $\hat{a}_t(s)$, and $\hat{\sigma}^2_S$ by local polynomials detailed in Appendix Subsection A.1.

For the BLUP step, $(\omega_{tl}, \psi_{tl})$, $l = 1, 2, \ldots$ are estimated by the spatial eigenvalue–eigenfunction pairs $(\hat{\omega}_{tl}, \hat{\psi}_{tl})$ obtained from the eigendecomposition of $\hat{G}^ST(s, s')$. Estimators for $\hat{\zeta}_{itl}$ defined in (20) is $\hat{\zeta}_{itl} = \hat{\omega}_{tl} \hat{\psi}_T^{tl} \hat{\Xi}_{it}^{-1} \hat{U}_{it}$, where $\hat{\psi}_{tl} = \hat{\psi}_t(\hat{s}_{it})$ and $[\hat{\Xi}_{it}]_{k_1 k_2} = \hat{G}^ST([\hat{s}_{it}]_{k_1}, [\hat{s}_{it}]_{k_2}) + \hat{\xi}_t([\hat{s}_{it}]_{k_1}) I[\hat{s}_{it}]_{k_1} = [\hat{s}_{it}]_{k_2} + \hat{\sigma}^2_S I\{(\hat{s}_{it})_{k_1}, (\hat{T}_{it})_{k_1} = ([\hat{s}_{it}]_{k_2}, (\hat{T}_{it})_{k_2})\}$ by (22), for $k_1, k_2 = 1, \ldots, |K_{it}|#$. The estimator for the temporal residual process $\hat{V}_i(s, t)$ is then

$$\hat{V}_i(s, t) = \sum_{l=1}^{L_2} \hat{\zeta}_{itl} \hat{\psi}_{tl}(s).$$ (23)

4.6. Missing values imputation. The overall imputed process (resp. its truncated version) is obtained by combining the temporal and spatial predictions (10) and (19) (resp. (15) and (21)), as

$$\hat{X}^ST_i(s, t) := \mu(s, t) + \hat{R}_i(s, t) + \hat{V}_i(s, t),$$ (24)

resp. $\hat{X}^µ_i(s, t) := \mu(s, t) + \hat{R}_i(s, t) + \hat{V}_i(s, t)$.

The superscript $ST$ above indicates that information on both the spatial and temporal aspects is incorporated.

The last step of the proposed imputation procedure is to impute the missing values in a partially observed image by aggregating the mean, temporal, and spatial effects. We propose to incorporate the spatial and temporal information and impute the missing pixel at $(s_k, t)$ by

$$\hat{X}^ST_i(s_k, t) = \hat{\mu}(s_k, t) + \sum_{l=1}^{L_1} \hat{\xi}_{ikl} \hat{\phi}_{kl}(t) + \sum_{l=1}^{L_2} \hat{\zeta}_{itl} \hat{\psi}_{tl}(s).$$ (26)

For comparisons, we consider the following models which respectively incorporate the temporal and mean effects, and mean-effect only:

$$\hat{X}_i^T(s_k, t) = \hat{\mu}(s_k, t) + \sum_{l=1}^{L_1} \hat{\xi}_{ikl} \hat{\phi}_{kl}(t),$$

$$\hat{X}_i^µ(s_k, t) = \hat{\mu}(s_k, t).$$

For completeness, we also consider a spatial-only model $\hat{X}_i^S(s_k, t)$, which performs the spatial imputation step without the temporal imputation. These methods can impute images with any missing rate, even for completely missing images, and can easily be extended to impute an arbitrary pixel location $s \in S$ through interpolation; however, the imputation to a higher resolution grid is often unnecessary in application.
4.7. Inferential Properties for the Imputation. Pixel-wise prediction intervals are constructed by following Theorem 4 in Yao et al. (2005). Assuming the temporal and spatial FPCs $\xi$ and $\zeta$ are independent and follow a joint Gaussian distribution, $\hat{X}_i^{ST}(s_k, T_{ij}) - X_i(s_k, T_{ij})$ follows a zero-mean normal distribution with variance $\hat{\Phi}_{k,L_1}^T(1)^T\hat{\Phi}_{k,L_1}(1) + \hat{\Psi}_{T_{ij},L_2}(s_k)^T\hat{\Psi}_{T_{ij},L_2}(s_k)$, where $\hat{\Phi}_{k,L_1} = \hat{\Phi}_{k,L_1}(T_{ij})\hat{\Phi}_{k,L_1}(T_{ij})^T + \hat{\Psi}_{T_{ij},L_2}(s_k)^T\hat{\Psi}_{T_{ij},L_2}(s_k)$, and $\hat{\Psi}_{T_{ij},L_2}(s_k)$ is constructed as

\begin{equation}
\hat{\Psi}_{T_{ij},L_2}(s_k) = \left(\hat{\psi}_{T_{ij},1}(s_k), \ldots, \hat{\psi}_{T_{ij},L_2}(s_k)\right)^T.
\end{equation}

Estimating these quantities via the plug-in versions and obtaining $\hat{\Phi}_{k,L_1}, \hat{\Phi}_{k,L_2}, \hat{\Psi}_{T_{ij},L_2}$, and $\hat{\psi}_{T_{ij},L_2}$, the $1 - \alpha$ pointwise interval for $\hat{X}_i^{ST}(s_k, T_{ij})$ is constructed as

\begin{equation}
\hat{X}_i^{ST}(s_k, T_{ij}) \pm \Phi^{-1}(1-\alpha/2)\sqrt{\hat{\Phi}_{k,L_1}(T_{ij})^T\hat{\Phi}_{k,L_1}(T_{ij}) + \hat{\Psi}_{T_{ij},L_2}(s_k)^T\hat{\Psi}_{T_{ij},L_2}(s_k)},
\end{equation}

where $\Phi$ is the standard normal distribution function and the last square-root term estimates the standard error for the prediction.

4.8. Efficient Implementation. Hereafter we refer to our imputation algorithm as STFIT, short for Spatio-Temporal Functional Imputation Tools. For the study of the Landsat data described in Section 6, we applied local constant estimators in smoothing the covariance for faster speed. Denote $K^{T}_{h_T}$ and $K^{S}_{h_S}$ as the kernel function used from temporal and spatial covariance estimation, respectively, where $h_T = 100$ and $h_S = 2$ are the bandwidth parameters, and $K$ is the Epanechnikov kernel, a compactly supported kernel that enjoys asymptotic optimality (Müller, 1984). The choice of kernel has a relatively small effect on statistical properties but can heavily affect computation efficiency. Employing the compactly supported Epanechnikov kernel results in a drastic speedup when dealing with image data of large volume as compared to using the non-compact Gaussian kernel.

To reduce computation, we evaluated $\hat{G}_{\theta_k}(t, t')$ at 50 evenly spaced points in time interval $T$ and linearly interpolated over the temporal domain as needed. A common time bin $T_i = T$ was selected in estimating $\hat{G}_{\theta_k}^{ST}(s, s')$ for all location pairs, while our general implementation supports varying time bins. Considering the periodic and seasonal nature of the data, spline smoothing with 11 Fourier basis was applied to estimate the mean function. An outlier threshold of 20% was used to identify outlier images. In determining the number of components $L_1$ and $L_2$, we chose $L_1$ and $L_2$ such that the fraction of variance explained (FVE) was greater than 0.99. Our algorithm
additionally supports sparse spatial covariance estimation by assuming zero correlation for pairs of pixels beyond certain distance apart, which can dramatically quicken computation when the image is large; in the case study we did not make this assumption.

We develop an R package *stfit* (Zhu, 2020) that implements the proposed imputation algorithm with other utility functions. For computation efficiency, heavy calculation steps are written in C++ through the R package *Rcpp*, and parallel computing is supported via the *foreach* (Microsoft and Weston, 2020) framework, parallelizing the pixel-wise calculation in the mean and temporal effect estimation procedures, and the image-wise principal components calculation in spatial effect estimation procedure. The code and system information to reproduce all results presented in this paper can be found at https://github.com/mingsnu/stfit/tree/master/tests/stfit_code_for_paper.

5. Theory. We state the theoretical results for the imputation algorithm and the estimation procedure and defer the technical conditions and derivation to the Appendix. For Proposition 1, recall that \( \tilde{X}_i \) defined in (5) is the BLUP given all observed data, \( \tilde{X}_i^T \) in (11) is the imputed process with temporal information, and \( \tilde{X}_i^{ST} \) in (24) is the two-step imputed process with both temporal and spatial information. Also write \( \tilde{X}_i^{RU}(s_k, t) := E[X_i(s, t) | T_{ik}, Y_{ik}, s_{it}, U_{it}] \), the BLUP given observations sharing the same location and temporal residuals sharing similar observation time. Evaluating the imputation methods under the mean squared prediction error (MSPE) criterion

\[
\text{MSPE}(\hat{A}) = E[(\hat{A} - A)^2]
\]

for a random variable \( A \) and its prediction \( \hat{A} \), we have the following ordering of the prediction errors for the BLUPs.

**Proposition 1.** For any \( s \in S_D \) and \( t \in T \),

\[
(28) \quad \text{MSPE}(\tilde{X}_i(s, t)) \leq \text{MSPE}(\tilde{X}_i^{RU}(s, t)) \leq \text{MSPE}(\tilde{X}_i^{ST}(s, t)) \leq \text{MSPE}(\tilde{X}_i^T(s, t)).
\]

The second inequality becomes an equality if \( Y_{ik} \) and \( U_{it} \) are conditionally uncorrelated given the observed time points and locations.

The additional efficiency achieved by the prediction conditional on more information is achieved at the cost of heavier computation to invert a larger covariance matrix. The uncorrelatedness in the last statement of Proposition 1 should be interpreted in an approximate sense, in that if the temporal information offered by \( Y_{ik} \) at pixel \( s_k \) and the spatial information contained
in $U_t$ near time $t$ are approximately orthogonal, then the proposed two-step prediction $\tilde{X}_{i}^{ST}(s,t)$ would perform nearly as well as the one-step prediction $\tilde{X}_{i}^{RU}(s,t)$.

The asymptotic theory of the proposed estimators are developed for the mean, temporal, and spatial smoothing steps using local linear smoothers (Li and Hsing, 2010; Zhang and Wang, 2016) with $p = 1$. Write $a_n = h_{\mu}^2 + [\log(n)/(nh_{\mu})]^{1/2}$ and $b_n = h_{T}^2 + [\log(n)/(nh_{T})]^{1/2}$.

**Theorem 1.** Suppose conditions (A1)–(A13) in the Appendix hold. For all $s_k \in S_D$, $i = 1, \ldots, n$, $k = 1, \ldots, m$, and $l = 1, 2, \ldots,$

\begin{align}
\sup_{t \in T} |\hat{\mu}(s_k, t) - \mu(s_k, t)| &= O(a_n) \text{ a.s.}, \\
\sup_{t \in T} |\hat{G}_{s_k}^{T}(t, t') - G_{s_k}^{T}(t, t')| &= O(a_n + b_n) \text{ a.s.}, \\
|\hat{\sigma}^2 - \sigma^2| &= O(a_n + b_n) \text{ a.s.}, \\
\sup_{t \in T} |\hat{\phi}_{kl}(t) - \phi_{kl}(t)| &= O(a_n + b_n) \text{ a.s.}, \\
\sup_{l \geq 1} |\hat{\lambda}_{kl} - \lambda_{kl}| &= O(a_n + b_n) \text{ a.s.}, \\
|\hat{\xi}_{ikl} - \tilde{\xi}_{ikl}| &= O(a_n + b_n) \text{ a.s.}, \\
\sup_{t \in T} |\tilde{X}_{i}^{T}(s_k, t) - \tilde{X}_{i}^{T}(s_k, t)| &= O(a_n + b_n) \text{ a.s.}
\end{align}

The imposed conditions are mild and are seen to be in accordance to the dataset we analyzed. Conditions (A1) and (A2) are moment and smoothness assumptions for the kernel function and are satisfied by the Epanechnikov kernel. According to (A3) and (A4), observations within each year are sparse but the observations pooled over all years the time points would cover any gaps in time. Observations over the spatial domain is required by (A5) to be dense and regular, and each image is assumed in (A6) to be observed at locations within a random area $O$, with the pooled observed areas covering the spatial domain. The number of observations, observation time points and areas, the underlying process, and measurement errors are required to be noninformative of each other in (A7). The temporal mean and covariance functions are assumed to be smooth in Conditions (A8) and (A9). In practice, though some image regions may display sudden changes, the lack of smoothness can be handled by separating the whole image into smooth sub regions and then applying the proposed method in each of the subregions. Appropriate bandwidths need to be set by the analyst, as per (A10) and
(A12), in a fashion dependent on the moments for the process and errors, as required by (A11) and (A13). The moment conditions are expected to be satisfied after the outlier removal step in STFIT.

For simplicity of presentation, in the next convergence theorem for the spatial imputation step we assume that the $U_{ijk}$ are being used in the spatial smoothers (36) instead of $\hat{U}_{ijk}$, so that the convergence rates from the temporal step does not manifest in the spatial imputation step below. Due to information borrowing from neighboring times points in $T_i$, the estimated temporal covariance converges to the local target $G^{ST}_{t,h_0}(s,s') := E[G^{ST}_{T,t_1}(s,s') | T_{11}, T_{11} \in T_i]$, where the expectation is taken with respect to the random observation time $T_{11}$. Let $(\omega_{t,h_0}, \psi_{t,h_0})$ be the $l$th eigenpair of $G^{ST}_{t,h_0}(s,s')$, $\psi_{itl,h_0} = \psi_{itl,0}(\hat{s}_i)$, $|\xi_{itl,h_0}|_{k_1k_2} = G^{ST}_{t,h_0}([\hat{s}_{itk_1}], [\hat{s}_{itk_2}]) + a_{itl,h_0} ([\hat{s}_{itk_1}]I([\hat{s}_{itk_1} = [\hat{s}_{itk_2}]) + \sigma^2 I([\hat{s}_{itk_1}, [T^T_{itk_2}]) = ([\hat{s}_{itk_2}], [T^T_{itk_2}])$, $\hat{\zeta}_{itl,h_0} = \omega_{t,h_0} \psi^T_{itl,h_0} \xi_{itl,h_0}^{-1} U_{it}$, $\hat{V}_{i,h_0}(s,t) = \sum_{l=1}^{L_2} \hat{\zeta}_{itl,h_0} \psi_{itl,h_0}(s)$, and $c_n = h^2_S + m^{-1} + [\log(n)/n]^{1/2}$.

**Theorem 2.** Suppose that conditions (A1)–(A7) and (A14) – (A16) in the Appendix hold. Then for any $t \in T$, $h_0 > 0$, $i = 1, \ldots, n$, $l = 1, 2, \ldots,$

\[(36) \quad \sup_{s,s' \in S} |\hat{G}^{ST}_{t,h_0}(s,s') - G^{ST}_{t,h_0}(s,s')| = O(c_n) \quad \text{a.s.},
\]

\[(37) \quad \sup_{s \in S} |\hat{\psi}_{it}(s) - \psi_{it,h_0}(s)| = O(c_n) \quad \text{a.s.},
\]

\[(38) \quad \sup_{l \geq 1} |\hat{\omega}_l - \omega_{t,h_0}| = O(c_n) \quad \text{a.s.},
\]

\[(39) \quad |\hat{\sigma}_l^2 - \sigma^2| = O(c_n) \quad \text{a.s.},
\]

\[(40) \quad \sup_{s \in S} |\hat{\zeta}_{itl} - \hat{\zeta}_{itl,h_0}| = O(c_n) \quad \text{a.s.},
\]

\[(41) \quad \sup_{s \in S} |\hat{\alpha}_{it}(s) - \alpha_{t,h_0}(s)| = O(c_n) \quad \text{a.s.},
\]

\[(42) \quad \sup_{s \in S} |\hat{V}_i(s,t) - \hat{V}_{i,h_0}(s,t)| = O(c_n) \quad \text{a.s.}
\]

Model components in the spatial imputation step are also estimated consistently, with a rate of convergence $c_n$ depending on the denseness of the locations over the spatial domain. Condition (A14) requires that the observed areas are contiguous enough; for example they are generated from blocks of cloud coverage. Conditions (A15) and (A16) are bandwidth and moment conditions for spatial smoothing.

6.1. Construction and imputation of partially observed images. To test the performance of the imputation algorithm, we selected two typical sites out of the 245 constructed sites (see Section 2) and refer to them as Site A and Site B, respectively. Site A is relatively spatially continuous and Site B has abrupt spatial changes. To show the spatial features of the two sites, a representative fully observed image from each of the sites is shown in Figure 3. Site A and Site B both have 660 images, of which 103 and 110 are completely observed, 195 and 176 are partially observed, and 362 and 374 are completely missing, respectively.

For each site and each season, we randomly selected 5 fully observed images, denoted as F1, F2, etc., except for winter of Site A for which we selected the only 4 completely observed images. The selected fully observed images and their detailed information for both sites are shown in Figure S1, Figure S2, and Table S1 in the Supplementary Material. Next, we selected 15 missing patterns with different missing rate from Site B, denoted as P1, P2, etc., as shown in Figure 4. The percentages of missing values for the images in the top, middle and bottom panels belong to three missing rate groups, below 30% (low), 30% ∼ 70% (medium) and 70% ∼ 99% (high), respectively, where the exact missing rates are included in Table S2 in the Supplementary Material.

In each simulation study, an artificial partially observed image is constructed by applying a missing pattern to a fully observed image, which we call the target image. Missing values in the target image are then imputed.
by applying the STFIT algorithm with all the rest of the images (either fully or partially observed) as input. A total of 285 \((19 \times 15)\) simulations for Site A and 300 \((20 \times 15)\) simulations for Site B were conducted. The root mean square prediction error (RMSPE) was used to assess the prediction accuracy of the STFIT algorithm, which is calculated for an imputed image \(\hat{X}_i(s, T_{ij})\) at time \(T_{ij}\) as

\[
\text{RMSPE} = \sqrt{\frac{|O_{ij}|}{\#} \sum_{k \in O_{ij}} [\hat{X}_i(s_k, T_{ij}) - X_i(s_k, T_{ij})]^2}
\]

where \(|O_{ij}|\) is the number of observed pixel at time \(T_{ij}\).

Fig 4: Fifteen missing patterns used in the simulation study, where white areas are observed and black areas are missing. The percentages of missing values for the images in the top, middle and bottom panels were below 30\%, 30\% \sim 70\% and 70\% \sim 99\% respectively.

6.2. Illustration of the STFIT algorithm. To illustrate the behavior of the STFIT algorithm, the imputation procedure is demonstrated on a partially observed image. The partially observed image shown in right panel of Figure 5 (the target image) was constructed by applying missing pattern P8 in Figure 4 to the fully observed image at Site B on DOY 228, 2004 as shown in the left panel of Figure 5.
We now illustrate how the missing part of the target image is recovered by the STFIT algorithm. The first step of the algorithm is mean estimation. The left panel of Figure 6 shows the estimated mean image at DOY 228 and the right panel shows the corresponding residual image. The mean image at DOY 228 already captures most of the variabilities of the target image. The residual image carries information about the target image and will be made use of by the following temporal and spatial imputation steps to improve prediction accuracy.

STFIT next performs the temporal effect estimation by pooling all residual images together. The estimated temporal effect for our target image is shown in the left panel of Figure 7, and the temporal residual image, i.e. the residual image with the temporal effect adjusted, is shown in the right panel of Figure 7. Overall, after adjusting for the temporal effect, the residuals had positive shifts compared to that before adjusting for the temporal effect, where the median of the residuals increased from -2.13 to 12.11. In general, adjusting for the temporal effect can improve the overall prediction accuracy, while it is not guaranteed to improve the prediction accuracy for each image or pixel.

To further illustrate the temporal estimation, we investigate pixel 157 (at the 6th column from the left and the 2nd row from the top). The estimated temporal covariance for pixel 157 is displayed in the left panel of Figure 8, which shows an increased covariance around the start and end of the year due to periodicity. A scatter plot of all the residuals for pixel 157 against DOY
Fig 6: Left: The estimated mean image of the target image. Right: The residual image of the target image after removing the mean

is displayed in the right panel of Figure 8, where the red and gray points represent the residuals from year 2004 and all other years, respectively, and the red line is the estimated temporal effect curve for year 2004, which is overall consistent with the temporal residuals (the red points).

We next estimated the spatial effect based on the temporal residual images. The left panel of Figure 9 shows the estimated spatial effect of the target image. From the figure we can see the spatial effect captures a lot of spatial information of the residual image. The right panel of Figure 9 is the residual image after adjusting both the temporal and spatial effects, from which we can barely see any spatial pattern left.

The missing values of the target image were imputed by summing left panel images from Figure 6, Figure 7, and Figure 9.

6.3. Imputation results. To examine the imputation results visually, we show five outputs from the imputation in Figure 10 for Site B. The first row shows the original fully observed images, F3, F7, F14 and F18, and the second row are the simulated partially observed images by applying missing patterns P6, P8, P14 and P15, respectively. The imputed images with the STFIT algorithm are given in the third row, the difference between the original images and the imputed images are given in the fourth row, and the estimated standard error images that define the halfwidth of the prediction intervals constructed in (27) are presented in the fifth row. Separate color bars are used to visualize the satellite images, error images, and stand error estimates. Visually, the imputed images are very close to the truth and most
Fig 7: Left: The estimated temporal effect of the target image. Right: The temporal residual image of the target image after removing the mean and adjusting for the temporal effect.

Fig 8: Left: The estimated covariance function for pixel 157. Right: The estimated temporal effect for pixel 157.

of the spatial details can be recovered, not only for the first two images with medium missing rates but also for the last two images with high missing rates. Notably, the high value strips near the top and right side of the images were accurately imputed and the curvy pattern around the vertical midline of the image was captured as well. No visual artefacts were produced in the imputed images. Some spatial pattern can be observed from the estimated standard error images, which can be interpreted as different textures of the
Fig 9: Left: The estimated spatial effect for the target image. Right: The spatial residual image of the target image after removing the mean and adjusting for the temporal and spatial effects

surface having different scale of variation. More than 99% of the imputed pixels fell in the 95% confidence intervals, indicating that the estimated standard errors and the prediction intervals constructed in Subsection 4.7 are not overly liberal.

To qualitatively assess the imputation results of the STFIT algorithm, the average RMSPEs by different seasons and different missing rate groups for both sites are shown in Table 1, while those for all image-missing pattern combinations are shown in Table S3 to Table S10 in the Supplementary Material. Results of both sites suggest that: (1) images with higher missing rate have higher imputation error; (2) images in the summer has lower imputation error than those in the winter. The reason for (2) is that summer exhibits less cloud cover than the winter, which could affect the overall quality of the images, and the winter has varying snow cover patterns that increase the difficulty for prediction.

Table 1
RMSPEs of the STFIT algorithm by different seasons and different missing rate groups

<table>
<thead>
<tr>
<th></th>
<th>Site A</th>
<th>Site B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Spring</td>
<td>Summer</td>
</tr>
<tr>
<td>&lt; 30</td>
<td>69.12</td>
<td>60.99</td>
</tr>
<tr>
<td>30~70</td>
<td>69.07</td>
<td>63.69</td>
</tr>
<tr>
<td>70~99</td>
<td>71.81</td>
<td>64.34</td>
</tr>
</tbody>
</table>
Fig 10: Imputation results of four simulated images at Site B. The first row shows the original fully observed images; the second row shows the simulated partially observed images; the third row shows the imputed images with the STFIT algorithm; the fourth row shows the difference between the original images and the imputed images, and the fifth row shows the estimated pixel-wise standard error images.
6.4. Contributions of the temporal and the spatial effects. Our model consists of three components, namely the mean, temporal effect, and spatial effect. To see the contribution of each component to the final imputation, we study the four methods described in the end of Subsection 4.6: (i) Mean-only ($\hat{X}_i^\mu$), (ii) Temporal-only ($\hat{X}_i^T$), (iii) Spatial-only ($\hat{X}_i^S$), and (iv) Spatio-temporal ($\hat{X}_i^{ST}$), which is the proposed STFIT full algorithm.

For each method mentioned above, we calculate the Relative RMSPEs (RRMSPE), which is defined as the ratio of the RMSPE of the method over the mean-only method $\hat{X}_i^\mu$; smaller RRMSPE indicates better performance. The average RRMSPE over all image-missing pattern combinations for Site A and B are shown in Table 2.

<table>
<thead>
<tr>
<th>Site</th>
<th>$X_i^\mu$</th>
<th>$X_i^T$</th>
<th>$X_i^S$</th>
<th>$X_i^{ST}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>0.90</td>
<td>0.73</td>
<td>0.68</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>0.98</td>
<td>0.68</td>
<td>0.69</td>
</tr>
</tbody>
</table>

Compared to the mean-only method, all of the temporal-only, spatial-only, and spatio-temporal methods had lower RMSPE for both sites, indicating that these methods are useful to extract more information from the residual images. For Site A, the spatio-temporal method had the smallest RMSPE compared to temporal-only and spatial-only methods. While for Site B, the spatial-only method had the lowest RMSPE which was close to spatio-temporal method. There is apparently less year-to-year variation in Site B than in Site A since the temporal-only method has very limited improvement on the RMSPE for Site B. Thus for Site B the temporal imputation step can potentially be skipped.

6.5. Comparison with other algorithms. Though many image imputation methods have been proposed in the literature, only a few of them have software implementations that are ready to use for practitioners. An exception is the gapfill method (Gerber et al., 2018), which is a spatio-temporal imputation method based on similar images. This method is available in the R package gapfill and outperforms the competitors Gapfill-Map (Weiss et al., 2014) and TIMESAT (Jönsson and Eklundh, 2004), becoming one of the best off-the-shelf packages for image imputation currently available. Hence, in this paper we focus on the comparison of our algorithm stfit and gapfill. We also consider the ordinary kriging method kriging implemented in the R package geoR (Ribeiro Jr and Diggle, 2016) as an additional
reference. All results reported are based on the tuning parameters found to be optimal for each method.

The gapfill method requires the input images to be on a regular grid in time. For our datasets, most of the images of the same location were observed every 8 days in a year, while the observed DOYs among different years are different. If DOY was used as the grid in time domain, there would be too many missing images in the final image array. Hence, we split the whole year into 46 evenly spaced bins, with 8 days in each bin except for the last bin which contains 5 or 6 days depending on whether the year is leap year. Imputation with the entire dataset containing a 46 days × 16 years image array using gapfill requires lengthy computation, so we instead input a 13 × 9 sub-image array centered at the target image. We have confirmed that the imputation accuracies using the sub-image arrays are similar to those of using the whole datasets through a few testing experiments.

Ordinary kriging provides the best linear unbiased prediction at an unobserved location based on observed values in the same image. We consider two imputation procedures based on ordinary kriging methods: 1) Kriging without detrending (kriging), which performed ordinary kriging assuming a constant mean over time, and 2) Kriging after detrending (kriging (detrended)), which for each image to be imputed first detrended the data by first deleting outliers and removing the mean function estimated by the STFIT algorithm, performed ordinary kriging based on the residual image, and finally obtained the imputed image as a sum of the mean and the estimated residual image. For both kriging methods the variograms were obtained by fitting an exponential variogram function model to the empirical variograms.

To compare the performances of stfit, gapfill, kriging, and kriging (detrended), we calculated the average RRMSPEs of the latter three methods relative to stfit within each missing rate group and season, and report the results in Table 3. The individual RMSPEs for image-missing pattern-algorithm combinations are shown in Table S3 to Table S10 in the Supplement. All RRMSPE for the comparison methods were greater than 1, showing that all comparison methods have lower prediction accuracy than the proposed stfit across all seasons and different missing rate groups. Among the three comparison methods, kriging performed the worst, followed by gapfill, and kriging (detrended) was second to our proposed stfit. It is expected that the kriging method does not work well as it only utilizes spatial information contained in the target image itself and it relies on a stationary assumption that is unlikely to hold in practice. However, the dramatic performance boost due to detrending is interesting, which makes
kriging (detrended) outperform gapfill. A reason is that the trend estimated from the mean estimation procedure of stfit can successfully capture majority of the variance of the image by borrowing information from nearby images in time. The gapfill method had at least 20% higher RMSPE compared to stfit method, and for most cases more than 30% higher. The largest RRMSPEs of gapfill occurred in spring at Site A, where the RMSPEs of gapfill for different missing rate groups were more than 2.5 times that of stfit. The high RRMSPEs are due to the bad prediction results of gapfill for F1 and F2 at Site A (See Table S3 in the Supplementary Materials), of which the neighboring inputs images contain abnormality. Though gapfill employs quantile regression for robustness, this procedure is less effective than stfit which incorporates a separate outlier detection step to remove low quality images. In addition, gapfill failed to impute two images at Site B, F16 and F17, due to a lack of neighboring observed images, while stfit was able to impute all images because it utilizes information across all years. The performance lift of stfit vs. kriging (detrended) for site B was smaller than that of site A, which is consistent with observations from Table 2 that the temporal effect for site B is not very strong. The performance improvement there is mainly due to the better spatial effect estimation by stfit.

### Table 3

RRMSPEs by different seasons and different missing rate groups for different imputation algorithms gapfill (GF), kriging (KG), and kriging (detrended) (KGD) relative to the stfit (ST) method, where a relative error larger than 1 indicates worse performance than stfit; (a) Site A and (b) Site B

<table>
<thead>
<tr>
<th>Missing %</th>
<th>Spring</th>
<th>Summer</th>
<th>Fall</th>
<th>Winter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GF</td>
<td>KG</td>
<td>KGD</td>
<td>GF</td>
</tr>
<tr>
<td>&lt; 30%</td>
<td>2.52</td>
<td>3.03</td>
<td>1.37</td>
<td>4.28</td>
</tr>
<tr>
<td>30%~70%</td>
<td>2.56</td>
<td>4.13</td>
<td>1.69</td>
<td>5.36</td>
</tr>
<tr>
<td>70%~99%</td>
<td>2.60</td>
<td>3.51</td>
<td>1.33</td>
<td>4.76</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Missing %</th>
<th>Spring</th>
<th>Summer</th>
<th>Fall</th>
<th>Winter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GF</td>
<td>KG</td>
<td>KGD</td>
<td>GF</td>
</tr>
<tr>
<td>&lt; 30%</td>
<td>1.68</td>
<td>1.89</td>
<td>1.10</td>
<td>1.59</td>
</tr>
<tr>
<td>30%~70%</td>
<td>1.76</td>
<td>1.90</td>
<td>1.16</td>
<td>1.57</td>
</tr>
<tr>
<td>70%~99%</td>
<td>1.96</td>
<td>2.03</td>
<td>1.12</td>
<td>1.46</td>
</tr>
</tbody>
</table>

We performed a speed comparison for stfit and gapfill with the microbenchmark (Mersmann, 2019) R package on a MacBook Pro with 2.5 GHz Quad-Core Intel Core i7 processor and 16 GB 1600 MHz DDR3 memory, using all 8 cores
for parallel computing. The time to impute one site of the Landsat data used in this paper containing 660 31 by 31 images using \texttt{stfit} with temporal and spatial effects imputation was about 2 minutes, and for \texttt{gapfill} was about 29 minutes. The most time consuming step in the \texttt{stfit} algorithm was the spatial effect estimation, which involves the estimation of a large covariance matrix of which the computation time increases quadratically in the number of pixels. When imputing the landsat data, about 94\% of the time was spent on spatial effect estimation.

7. **Discussion.** In this paper, we introduced the STFIT algorithm for spatiotemporal satellite data imputation based on functional data analysis techniques. Compared with other existing imputation methods, our method has its advantage in many aspects. First, the STFIT algorithm is robust to outliers. The STFIT algorithm incorporates an outliers detection procedure which can greatly reduce the effect of outliers. Second, the STFIT algorithm outperforms both \texttt{gapfill} and kriging in terms of prediction accuracy, where the latter two are the top performers among existing methods. Third, the reconstructed images by STFIT preserve sharp patterns with high spatial and temporal resolution. Fourth, the STFIT algorithm guarantees a 100\% imputation rate. Fifth, the computation speed is much faster for images less than 5000 pixels than \texttt{gapfill}.

Our method also has some limitations. Because our method requires the non-parametric estimation of the spatial covariance matrix, there is a limitation on the size of the image we can handle directly. To impute large satellite images, either a divide-and-conquer or moving window approach has to be used to reduce the computation for large matrix operations. Another limitation is that our approach depends on the assumption that after temporal detrending the spatial temporal process is temporally stationary. This assumption may not hold when there are significant land cover changes on the ground in a relatively short time period, such as when urbanization and deforestation occur. To address this issue, one needs to develop a robust change detection method, to separate the time series of satellite images at the change point, and to apply our algorithm separately before and after the change point. We plan to address these issues in a future paper.
APPENDIX A: THEORETICAL RESULTS

A.1. Definition of Temporal and Spatial Smoothers. The estimation of the temporal imputation model components at each pixel $s_k \in S_D$ is defined as follows. Since $\text{cov}(R_{ij,k}, R_{ij'k} | T_{ij}, T_{ij'}) = G_{s_k}^T (T_{ij}, T_{ij'}) + \sigma^2 I(j = j')$, where $I(\cdot)$ is the indicator function, the temporal covariance (9) is estimated by local polynomial smoothing with order $p \geq 0$,

$$ G_{s_k}^T (t, t') = \hat{\beta}_{00}, $$

$$ \{ \hat{\beta}_{1, l_2} \}_{0 \leq l_1 + l_2 \leq p} = \arg \min_{\{ \beta_{1,j_2} | 0 \leq l_1 + l_2 \leq p \}} \sum_{i=1}^{n} \sum_{j \neq j' \in J_{ik}} K_{h_T}^T (T_{ij} - t) K_{h_T}^T (T_{ij'} - t') $$

$$ [\hat{R}_{ijk} \hat{R}_{ij'k} - \sum_{l_1}^{p} \beta_{1,l_2} (T_{ij} - t)^{l_1} (T_{ij'} - t')^{l_2}]^2, $$

where $h_T > 0$ is the temporal bandwidth, $K_{h_T}^T(x) = K^T(x/h_T)/h_T$, and $K^T$ is the one-dimensional kernel for temporal smoothing. The nugget effect $\sigma^2$ is estimated by

$$ \hat{\sigma}^2 = \frac{1}{|T|} \int_{t \in T} |\hat{V}_k^T (t) - \hat{G}_{s_k}^T (t, t)| \, dt, $$

$$ \hat{V}_k^T (t) = \hat{\beta}_0, \quad (\hat{\beta}_0, \ldots, \hat{\beta}_p) = \arg \min_{\beta_0, \ldots, \beta_p} \sum_{i=1}^{n} \sum_{j \in J_{ik}} \sum_{l=0}^{p} [\hat{R}_{ijk} - \sum_{l=0}^{p} \beta_l (T_{ij} - t)^{l}]^2 K_{h_T}^T (T_{ij} - t), $$

where $\hat{V}_k^T (t)$ targets $V_k^T (t) = G_{s_k}^T (t, t) + \sigma^2$. The estimation of $V_k^T (t)$ utilizes only data that contribute to the diagonal of the covariance function, following Yao et al. (2005).

We estimate the spatial covariance $G_{s}^{ST} (s, s')$ by local polynomial smoothing applied on the product residuals. The spatial covariance estimate with order $p \geq 0$ is

$$ \hat{G}_s^{ST} (s, s') = \hat{\gamma}_{00}, $$

$$ \{ \hat{\gamma}_{l_1,l_2} \}_{0 \leq \| l_1 + l_2 \| \leq p} = \arg \min_{\gamma_{l_1,l_2} | 0 \leq \| l_1 + l_2 \| \leq p} \sum_{i=1}^{n} \sum_{(j,k),(j',k') \in K_{st}} K_{h_S}^S (s_k - s) K_{h_S}^S (s_{k'} - s') $$

$$ [\hat{U}_{ijk} \hat{U}_{ij'k'} - \sum_{0 \leq \| l_1 + l_2 \| \leq p} \gamma_{l_1,l_2} (s_k - s)^{l_1} (s_{k'} - s')^{l_2}]^2, $$

where $h_S > 0$ is the bandwidth for spatial smoothing, $K_{h_S}^S (x) = K^S([x \, 1]/h_S)$ $K^S([x \, 2]/h_S)/h_S^2$, and $K^S$ is a one-dimensional kernel for spatial smoothing.
By (22), the longitudinal (respectively pixel) nugget effect is \( a_t(s) \) (respectively \( \sigma^2 \)), which are estimated from the temporal residual sharing the same time but not location (respectively time as well as location) by

\[
\hat{a}_t(s) = \hat{\gamma}_0 - \hat{G}^{ST}_t(s, s),
\]

\[
\{\hat{\gamma}_t\}_{0 \leq |t| \leq p} = \arg \min \{\gamma_t : 0 \leq |t| \leq p\} \sum_{i=1}^n \sum_{(j,k),(j',k') \in K_{it}} |\hat{U}_{ijk} \hat{U}_{ij'k'}| - \sum_{0 \leq |t| \leq p} \gamma_t(s_k - s)^2 K^{S}_{h_S}(s_k - s),
\]
and respectively

\[
\hat{\sigma}^2 = \frac{1}{|S|} \int_{s \in S} \hat{V}^{ST}_t(s) - \hat{a}_t(s) - \hat{G}^{ST}_t(s, s) \, ds, \quad \hat{V}^{ST}_t(s) = \hat{\beta}_0,
\]

\[
\{\hat{\beta}_t\}_{0 \leq |t| \leq p} = \arg \min \{\beta_t : 0 \leq |t| \leq p\} \sum_{i=1}^n \sum_{(j,k) \in K_{it}} |\hat{U}_{ijk}^2 - \sum_{0 \leq |t| \leq p} \beta_t(s_k - s)^2 K^{S}_{h_S}(s_k - s),
\]

where \( \hat{V}^{ST}_t(s) \) targets \( V^{ST}_t(s) = G^{ST}_t(s, s) + a_t(s) + \sigma^2 \). In our implementation, the longitudinal nugget effect were set to zero, and only the pixel nugget effects were estimated.

When \( p = 0 \), the solutions to (45) and (47) have the explicit formulae

\[
\hat{G}^{ST}_{h_k}(t, t') = \frac{\sum_{i=1}^n \sum_{j \neq j' \in J_k} \hat{R}_{ijk} \hat{R}_{ij'k'} K^{T}_{h_T}(T_{ij} - t) K^{T}_{h_T}(T_{ij'} - t')}{\sum_{i=1}^n \sum_{j \neq j' \in J_k} K^{T}_{h_T}(T_{ij} - t) K^{T}_{h_T}(T_{ij'} - t')},
\]

\[
\hat{G}^{ST}_t(s, s') = \frac{\sum_{i=1}^n \sum_{(j,k),(j',k') \in K_{it}, k \neq k'} \hat{U}_{ijk} \hat{U}_{ij'k'} K^{S}_{h_S}(s_k - s) K^{S}_{h_S}(s_{k'} - s')}{\sum_{i=1}^n \sum_{(j,k),(j',k') \in K_{it}, k \neq k'} K^{S}_{h_S}(s_k - s) K^{S}_{h_S}(s_{k'} - s')},
\]

### A.2. Technical Conditions.

(A1) The one-dimensional smoothing kernels \( K^{T}(\cdot) \) and \( K^{S}(\cdot) \) equal \( K(\cdot) \), where \( K(\cdot) \) is a symmetric probability density supported on \([-1,1]\) with

\[
\sigma^2_K = \int u^2 K(u) \, du < \infty, \quad \int K(u)^2 \, du < \infty.
\]

(A2) \( K(\cdot) \) is Lipschitz continuous: There exists \( 0 < M_K < \infty \) such that

\[
\sup_{-1 \leq u < v \leq 1} \frac{|K(u) - K(v)|}{|u - v|} \leq M_K, \quad \sup_{u \in [-1,1]} K(u) \leq M_K.
\]

(A3) The number of observations \( N_t \) are i.i.d. realizations of a bounded random variable with \( P(N_t \geq 2) > 0 \).
(A4) \{T_{ij} \mid i = 1, \ldots, n, j = 1, \ldots, N_i\} are i.i.d. copies of a random variable \( T \) supported on \( \mathcal{T} \). The density \( f_T(\cdot) \) of \( T \) is bounded from below and above:

\[
0 < \min_{t \in \mathcal{T}} f_T(t) \leq \max_{t \in \mathcal{T}} f_T(t) < \infty,
\]

and the second derivative \( f_T^{(2)}(\cdot) \) of \( f_T(\cdot) \) is bounded.

(A5) \( \mathcal{S} \) is a rectangular region which can be evenly divided into \( m = m(n) \) rectangular subdomains, \( m \to \infty \); \( S_D \) contains the central points of the subdomains.

(A6) The random observed sets \( O_{ij} \) are i.i.d. copies of a random set \( O \subset \mathcal{S} \).

Then \( \inf_{s \in \mathcal{S}} P(s \in O) > 0 \) and \( \inf_{s, s' \in \mathcal{S}} P(s \in O, s' \in O) > 0 \).

(A7) The underlying processes \( X_i \), the number of observations \( N_i \), the design time points \( T_{ij} \), and the observed areas \( O_{ij} \) are all independent.

The measurement errors \( \epsilon_{ijk} \) are identical and independent of the \( N_i, T_{ij}, O_{ij} \), and \( X_i \).

(A8) The second partial derivative \( \partial^2 \mu(s_k, t) / \partial t^2 \) is bounded on \( \mathcal{T} \).

(A9) The second order derivatives \( \partial^2 G_{s_k}^T(t, t') / \partial t^2 \), \( \partial^2 G_{s_k}^T(t, t') / \partial t'^2 \), and \( \partial^2 G_{s_k}^T(t, t') / \partial t \partial t' \) are all bounded on \( \mathcal{T}^2 \).

(A10) \( h_{\mu} \to 0 \) and \( \log(n) / (nh_{\mu}) \to 0 \).

(A11) For each \( k = 1, \ldots, m \), there exists some \( \alpha > 2 \) such that \( E \sup_{t \in \mathcal{T}} |X(s_k, t)|^\alpha < \infty \), \( E|\epsilon_{ijk}|^\alpha < \infty \), and

\[
h_{\mu} \left[ \frac{\log(n)}{n} \right]^{2/\alpha - 1} \to \infty.
\]

(A12) \( h_T \to 0 \) and \( \log(n) / (nh_T^2) \to 0 \).

(A13) For each \( k = 1, \ldots, m \), there exists some \( \beta > 2 \) such that \( E \sup_{t \in \mathcal{T}} |X(s_k, t)|^{2\beta} < \infty \), \( E|\epsilon_{ijk}|^{2\beta} < \infty \), and

\[
h_T^2 \left[ \frac{\log(n)}{n} \right]^{2/\beta - 1} \to \infty.
\]

(A14) For any \( \epsilon, \eta > 0 \), there exists \( \delta > 0 \) such that

\[
\limsup_{n \to \infty} P\left( \sup_{\|s-s'\|_1 < \delta} \frac{1}{n} \sum_{i=1}^{n} \left| I(s \in O_{i1}) - I(s' \in O_{i1}) \right| > \epsilon \right) < \eta,
\]

\[
\limsup_{n \to \infty} P\left( \sup_{\|s_1-s'_1\|_1 < \delta} \frac{1}{n} \sum_{i=1}^{n} \left| I(s_1 \in O_{i1}, s'_1 \in O_{i1}) - I(s_2 \in O_{i1}, s'_2 \in O_{i1}) \right| > \epsilon \right) < \eta.
\]
(A15) $h_S \to 0$ and $m h_S^2$ is bounded below by 1.
(A16) For some $\nu > 2$, $E \sup_{s \in S, t \in T}|X(s, t)|^{2\nu} < \infty$, $E|\epsilon_{ijk}|^{2\nu} < \infty$, and
$$
\left( \frac{h_S^4}{m^2} + h_S^8 \right) \left[ \frac{\log(n)}{n} \right]^{2/\nu - 1} \to \infty .
$$


**Proof of Proposition 1.** Conditional on the observation time points and locations, the $U_{ijk}$ lie in the linear span of the observed $Y_{ijk}$. The first inequality follows from the property of BLUP. The second inequality is due to that $\tilde{X}_i ST(s, t)$ is linear in $\{Y_{ik}, U_{it}\}$; it becomes an equality if the predictors are uncorrelated. The last inequality is by Theorem 3.6 in Seber and Lee (2003).

**Proof of Theorem 1.** Results (29)–(30) and rates $a_n$ and $b_n$ follow from Corollary 5.1 (1) and Corollary 5.3 in Zhang and Wang (2016), respectively, and we here verify the conditions. The design time point condition (A1) in Zhang and Wang (2016) is almost surely implied by (A3)–(A7). Conditions (A1)–(A2), (B2)–(B4), (C1c)–(C3c) in Zhang and Wang (2016) are verified by (A1)–(A13) almost surely.

By (29), $\tilde{R}_{ij}^2 - R_{ij}^2 = [\mu(s_k, T_{ij}) - \hat{\mu}(s_k, T_{ij})]^2 + 2R_{ij} [\mu(s_k, T_{ij}) - \hat{\mu}(s_k, T_{ij})]$; where $sup_t |\mu(s_k, t) - \hat{\mu}(s_k, t)| = O(a_n)$ a.s. Thus $sup_{t \in T} |\tilde{V}_k(t) - V_k(t)| = O(a_n)$ a.s., where
$$
\tilde{V}_k^T(t) = \hat{\beta}_0, \quad (\hat{\beta}_0, \ldots, \hat{\beta}_p) = \arg \min_{\beta_0, \ldots, \beta_p} \sum_{i=1}^{n} \sum_{j=1}^{N_i} [R_{ijk} - \sum_{l=0}^{p} \beta_l (T_{ij} - t)]^2 K_{kp} (T_{ij} - t).
$$

Again by (29), $sup_{t \in T} |\tilde{V}_k^T(t) - V_k^T(t)| = O(a_n)$, so $sup_{t \in T} |\tilde{V}_k^T(t) - V_k^T(t)| = O(a_n)$ a.s. Then (31) follows from (46) and (30). Results (32) and (33) follow from (29) and the perturbation theorem (see, e.g. Hsing and Eubank, 2015), and (34)–(35) follow from the Slutsky’s theorem.

**Proof of Theorem 2.** We provide the proof for (36); (37)–(42) then follow by analogous arguments to the proof of Theorem 1. Let $m_{ij} = |\{k \mid (j, k) \in K_{it}\}|$ be the number of spatial observations for the $i$th year and $j$th time point, $m_i = \sum_{i=1}^{n} m_{ij}$, $N_{it} = \sum_{1 \leq j \neq j' \leq N_i} m_{ij} m'_{ij}$ be the number of raw covariances for the $i$th year, $N_{S,t} = \sum_{i=1}^{n} N_{it}$ be the total number of observations, and $T_i = \{ (j, k, j', k') \mid (j, k) \neq (j', k') \in K_{it} \}$. 

\[\]
To ease notations, all summation indices $l_{pq}$ fall in $\{0,1\}$ for $p,q = 1,2$, and all scalar quantities in the next display depend on $(s,s') \in S \times S$. Let

$$S = \begin{bmatrix}
S_{0000} & S_{1000} & S_{0100} & S_{0010} & S_{0001} \\
S_{1000} & S_{2000} & S_{1100} & S_{1010} & S_{1001} \\
S_{0100} & S_{1010} & S_{0110} & S_{0101} & S_{0101} \\
S_{0010} & S_{0110} & S_{0110} & S_{0020} & S_{0011} \\
S_{0001} & S_{1001} & S_{0101} & S_{0011} & S_{0002}
\end{bmatrix}, \quad R = \begin{bmatrix}
R_{0000} \\
R_{1000} \\
R_{0100} \\
R_{0010} \\
R_{0001}
\end{bmatrix},$$

for which the entries are defined as

$$S_{l_1l_2l_1'l_2'}(s,s') = \frac{1}{N_{s,t}} \sum_{i=1}^{n} \sum_{(j,k,j',k') \in I_i} K_{h_S}^S(s_k - s)K_{h_S}^S(s_{k'} - s)$$

$$\left(\frac{[s_k]_1 - [s]_1}{h_S}\right)^{l_{11}} \left(\frac{[s_k]_2 - [s]_2}{h_S}\right)^{l_{12}} \left(\frac{[s_{k'}]_1 - [s']_1}{h_S}\right)^{l_{21}} \left(\frac{[s_{k'}]_1 - [s']_1}{h_S}\right)^{l_{22}},$$

$$R_{l_1l_2l_1'l_2'}(s,s') = \frac{1}{N_{s,t}} \sum_{i=1}^{n} \sum_{(j,k,j',k') \in I_i} K_{h_S}^S(s_k - s)K_{h_S}^S(s_{k'} - s)U_{ijk}U_{ij'k'}$$

$$\left(\frac{[s_k]_1 - [s]_1}{h_S}\right)^{l_{11}} \left(\frac{[s_k]_2 - [s]_2}{h_S}\right)^{l_{12}} \left(\frac{[s_{k'}]_1 - [s']_1}{h_S}\right)^{l_{21}} \left(\frac{[s_{k'}]_1 - [s']_1}{h_S}\right)^{l_{22}},$$

for $s, s' \in S$. The solution to

$$\arg \min_{\gamma} \sum_{i=1}^{n} \sum_{(j,k,j',k') \in I_i} K_{h_S}^S(s_k - s)K_{h_S}^S(s_{k'} - s')$$

$$\left[U_{ijk}U_{ij'k'} - \sum_{0 \leq |\ell_1 + \ell_2| \leq p} \gamma_{\ell_1,\ell_2} (s_k - s)^{\ell_1}(s_{k'} - s')^{\ell_2}\right]^2,$$

is given by the solution $\gamma = \left[\tilde{\gamma}_{l_1l_2l_1'l_2'}\right]_{0 \leq \sum_{p,q=1}^{2} l_{pq} \leq 1}$ to the system $S\gamma = R$.

Let $S$ and $S'$ be independent uniform random variables supported on $S$ which are also independent of all other random variables. Define

$$E_D[R_{l_1l_2l_1'l_2'}(s,s')] = E[K_{h_S}^S(S - s)K_{h_S}^S(S' - s')U_{ijk}U_{ij'k'}$$

$$\left(\frac{[S]_1 - [s]_1}{h_S}\right)^{l_{11}} \left(\frac{[S]_2 - [s]_2}{h_S}\right)^{l_{12}} \left(\frac{[S']_1 - [s']_1}{h_S}\right)^{l_{21}} \left(\frac{[S']_1 - [s']_1}{h_S}\right)^{l_{22}}].$$

By (A5) and the midpoint rule for Riemann sums,

$$\sup_{s,s' \in S} |E[R_{l_1l_2l_1'l_2'}(s,s')] - E_D[R_{l_1l_2l_1'l_2'}(s,s')]| = O(m^{-1}).$$
Let \( S_{t_1l_1l_2l_2} \) denotes the matrix cofactor of the \( S_{t_1l_1l_2l_2l_2l_2} \) entry in the first column of \( S \). Suppressing the arguments \((s, s')\) in all functions,

\[
\hat{\gamma}_{0000} - G_{t,00}^{ST} \det(S) = \sum_{l_1+l_2+t_1l_2+1 \leq 1} \det(S_{t_1l_1l_2l_2l_2l_2}) \bigg[ R_{t_1l_1l_2l_2l_2l_2} - S_{t_1l_1l_2l_2l_2l_2} G_{t,00}^{ST} \bigg] \\
- \sum_{l_1+l_2+t_1l_2+1 \leq 1} \det(S_{t_1l_1l_2l_2l_2l_2}) S_{t_1l_1l_2l_2l_2l_2} \bigg[ \frac{\partial^1_{11} \partial_{12} \partial_{21} \partial_{22} \partial_{2} G_{t,00}^{ST}}{2} \bigg] \\
= \sum_{l_1+l_2+l_2 \leq 1} \det(S_{t_1l_1l_2l_2l_2l_2}) \bigg[ [R_{t_1l_1l_2l_2l_2l_2} - E(R_{t_1l_1l_2l_2l_2l_2})] \\
+ [E(R_{t_1l_1l_2l_2l_2l_2}) - E_D(R_{t_1l_1l_2l_2l_2l_2})] \bigg] + O(h_{S}^2) \\
(49) \\
= \sum_{l_1+l_2+l_2 \leq 1} \det(S_{t_1l_1l_2l_2l_2l_2}) \bigg[ R_{t_1l_1l_2l_2l_2l_2} - E(R_{t_1l_1l_2l_2l_2l_2}) \bigg] + O(m^{-1}) + O(h_{S}^2)
\]

where the big-O terms hold uniformly over \((s, s') \in \mathcal{S}\), the second summation equals 0 by the properties of the determinant, the second equality is due to Taylor expansion, and the last is due to \((48)\). Applying Lemma 2 with \( r = 0 \) and the continuous mapping theorem, \( S, S_{t_1l_1l_2l_2l_2l_2}, \) and \( \det(S) \) converge uniformly to \( E[S], \det(E[S_{t_1l_1l_2l_2l_2l_2}]), \) and \( \det(E[S]) \), respectively. Also, \( \det(E[S]) \) is bounded away from 0 for each \((s, s')\) since \( E[S(s, s')]= \) a positive Gram matrix by \((A15)\), and \( \det(E[S]) \) is uniformly bounded away from 0 by continuity. To finish, apply Lemma 2 with \( r = 1 \) to \((49)\). \( \square \)

For sequences \( x_n \) and \( y_n \), let \( x_n \asymp y_n \) denote \( 0 < \lim \inf \sup_n (x_n/y_n) \leq \lim \sup \sup_n (x_n/y_n) < \infty \).

**Lemma 1.** Under \((A3)-(A7)\), the following hold with probability 1:

\[
N_{S,t} \asymp nm^2, \\
\sum_{i=1}^{n} N_{it}(N_{it} - 1) \asymp nm^4, \\
\sup_n (n \max_i N_{it}/N_{S,t}) < \infty.
\]

For Lemma 2, let \( V_{ijk} = X_i(s_k, T_{ij}) - X_i^T(s_k, T_{ij}) \) and \( G_{ijj'k'} = V_{ijk} V_{ij'k'} \).
For \( l_{pq} \in \{0, 1\}, p, q = 1, 2, \) and \( r \in \{0, 1\}, \) define

\[
F_{ijkj'}(s, s') = K^S\left(\frac{[s_k]_1 - [s]_1}{h_S}\right)K^S\left(\frac{[s_k]_2 - [s]_2}{h_S}\right)K^S\left(\frac{[s_{k'}]_1 - [s']_1}{h_S}\right)K^S\left(\frac{[s_{k'}]_2 - [s']_2}{h_S}\right)
\left(\frac{[s_k]_1 - [s]_1}{h_S}\right)^{l_{11}}\left(\frac{[s_k]_2 - [s]_2}{h_S}\right)^{l_{12}}\left(\frac{[s_{k'}]_1 - [s']_1}{h_S}\right)^{l_{21}}\left(\frac{[s_{k'}]_2 - [s']_2}{h_S}\right)^{l_{22}}
\]

\[
H_{l_{11}l_{12}l_{21}l_{22}}(s, s') = \frac{1}{N_{S,t}} \sum_{i=1}^{n} \sum_{(j,k,j',k') \in T_i} F_{ijkj'}(s, s') G_{ijkj'}^r.
\]

Note that \( R_{l_{11}l_{12}l_{21}l_{22}} \) and \( S_{l_{11}l_{12}l_{21}l_{22}} \) as defined in the proof of Theorem 2 are \( H_{l_{11}l_{12}l_{21}l_{22}}/h_S^4 \) and \( H_{l_{11}l_{12}l_{21}l_{22}}/h_S^4, \) respectively.

**Lemma 2.** Under the conditions of Theorem 2, for \( H(s, s') = H_{l_{11}l_{12}l_{21}l_{22}}(s, s'), \) \( r \in \{0, 1\}, l_{pq} \in \{0, 1\} \) such that \( \sum_{p,q=1}^{2} l_{pq} = 1, \)

\[
\sup_{s,s' \in S} |H(s, s') - E(H(s, s'))| = O(h_S^4 \left[\log(n)\right]^{1/2}) \text{ a.s.}
\]

**A.4. Proof of Lemmas.**

**Proof of Lemma 1.** It suffices to show that

\[
\frac{1}{n} \sum_{i=1}^{n} N_{i1} \asymp m \text{ a.s.,}
\]

\[
\frac{1}{n} \sum_{i=1}^{n} N_{i1}^2 \asymp m^2 \text{ a.s.}
\]

By (A14) and the law of large numbers at each \( s \in S, \) classical empirical process theory (van der Vaart and Wellner, 1996) implies that

\[
\sup_{s \in S} \left| \frac{1}{n} \sum_{i=1}^{n} I(s \in O_{i1}) - P(s \in O_{11}) \right| \to 0 \text{ a.s.,}
\]

\[
\sup_{s,s' \in S} \left| \frac{1}{n} \sum_{i=1}^{n} I(s \in O_{i1}, s' \in O_{i1}) - P(s \in O_{11}, s' \in O_{11}) \right| \to 0 \text{ a.s.}
\]
Since $N_{it} = \sum_{j=1}^{N_i} \sum_{k=1}^{m} I(s_k \in O_{ij})$,

$$\frac{1}{nm} \sum_{i=1}^{n} N_{it} = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \sum_{k=1}^{m} I(s_k \in O_{ij})$$

$$= \frac{1}{m} \sum_{k=1}^{m} \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{N_i} I(s_k \in O_{ij})$$

$$= \frac{1}{m} \sum_{k=1}^{m} \left[ E[N_1] P(s_k \in O_{11}) + o(1) \right] \text{ a.s.}$$

(58)

$$\asymp 1 \text{ a.s.}$$

where the second last inequality is due to the independence of $N_i$ and $O_{ij}$ as well as (56), the $o(1)$ term is uniform over $k$, and the last is due to (A4) and (A6). Similarly,

$$\frac{1}{nm^2} \sum_{i=1}^{n} N_{it}^2 = \frac{1}{nm^2} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \left[ \sum_{k=1}^{m} I(s_k \in O_{ij}) \right]^2$$

$$= \frac{1}{m} \sum_{k,k'=1}^{m} \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{N_i} I(s_k \in O_{ij}) I(s_{k'} \in O_{ij})$$

$$= \frac{1}{m^2} \sum_{k,k'=1}^{m} \left[ E[N_1] P(s_k \in O_{11}, s_{k'} \in O_{11}) + o(1) \right] \text{ a.s.}$$

(59)

$$\asymp 1 \text{ a.s.}$$

\[\square\]

**Proof of Lemma 2.** We show only the proof for $r = 1$, since the case for $r = 0$ is then a special case when $G_{ijkj'k'} \equiv 1$.

First consider the $N_{it}$ are being conditioned on, and that the conclusions
of Lemma 1 hold. Write
\[
\sup_{s,s' \in \mathcal{S}} |H(s, s') - E(H(s, s'))| \leq \sup_{s,s' \in \mathcal{G}(\kappa)} |H(s, s') - E(H(s, s'))| \\
+ \sup_{s_1, s'_1, s_2, s'_2 \in \mathcal{S}} |H(s_1, s'_1) - H(s_2, s'_2)| \\
+ \sup_{s_1, s'_1, s_2, s'_2 \in \mathcal{S}} |E[H(s_1, s'_1)] - E[H(s_2, s'_2)]| \\
=: H_1 + H_2 + H_3,
\]
where $\mathcal{G}(\kappa) \subset \mathcal{S}$ is an equally-spaced lattice on $\mathcal{S}$ with $|\mathcal{S}|n^{2\kappa}$ points for some $\kappa > 0$. By (A2),
\[
|F_{ijk'k'}(s_1, s'_1) - F_{ijk'k'}(s_2, s'_2)| \leq C_0 \frac{||s_1 - s_2||_1 + ||s'_1 - s'_2||_1}{h_S},
\]
where $C_0$ is a universal constant, and so are $C_1, C_2, \ldots$ defined thereafter. Then
\[
\sup_{s_1, s'_1, s_2, s'_2 \in \mathcal{S}} |F_{ijk'k'}(s_1, s'_1) - F_{ijk'k'}(s_2, s'_2)| =: \frac{2C_0}{h_S} n^{-\kappa} \sup_{t \in t', s, s' \in \mathcal{S}} |V_i(t', s) V_i(t', s')|.
\]
By (62), (A16), (50) and (52) in Lemma 1, and the law of large numbers, the second and the third terms in (60) are
\[
H_2 = O(n^{-\kappa}/h_S) \text{ a.s., } H_3 = O(n^{-\kappa}/h_S).
\]
Let
\[
d_n = \{\log(n)[h_S^4/N_{S,t} + h_S^8 \sum_{i=1}^n N_{it} (N_{it} - 1)/N_{S,t}^2]^{1/2}, \quad D_n = [n/ \log(n)]d_n.
\]
The first term in (60) is then
\[
H_1 = \sup_{s,s' \in \mathcal{G}(\kappa)} |H^*(s, s') - E[H^*(s, s')]| \\
+ \sup_{s,s' \in \mathcal{G}(\kappa)} |H(s, s') - H^*(s, s')| \\
+ \sup_{s,s' \in \mathcal{G}(\kappa)} |E[H(s, s')] - E[H^*(s, s')]| \\
=: H_{11} + H_{12} + H_{13}.
\]
where

\[ H^*(s, s') = \frac{1}{N_{S,t}} \sum_{i=1}^n \sum_{(j,k,j,k') \in I_i} F_{ijjk'k'}(s, s') G_{ijjk'k'} I(G_{ijjk'k'} \leq D_n). \]

By (A16),

\[
H_{12} \leq \frac{1}{N_{S,t}} \sum_{i=1}^n \sum_{(j,k,j,k') \in I_i} M_K^4 \left( \frac{G_{ijjk'k'}}{D_n} \right)^\nu D_n I(G_{ijjk'k'} > D_n)
\]

\[
\leq \frac{1}{N_{S,t}} \sum_{i=1}^n \sum_{(j,k,j,k') \in I_i} M_K^4 \sup_{t \in T, s,s' \in S} |V_i(t, s)V_i(t, s')|^{1-\nu} D_n^{1-\nu}
\]

\[ \leq \frac{B}{n} \sum_{i=1}^n M_K^4 \sup_{t \in T, s,s' \in S} |V_i(t, s)V_i(t, s')|^{1-\nu} D_n^{1-\nu} = O(D_n^{1-\nu}) \quad \text{a.s.,} \]

and similarly

\[ H_{13} = O(D_n^{1-\nu}) \quad \text{a.s.} \]

Let \( T_i(s, s') = N_{S,t}^{-1} \sum_{(j,k,j',k') \in I_i} F_{ijjk'k'}(s, s') G_{ijjk'k'} I(G_{ijjk'k'} \leq D_n) \), so \( H^*(s, s') = \sum_{i=1}^n T_i(s, s') \), and \( T_i \) are i.i.d. To apply the Bernstein’s inequality, note that

\[
M_U = |T_i(s, s') - E[T_i(s, s')]| \leq \frac{1}{N_{S,t}} N_i t M_K^4 D_n
\]

\[ \leq C_1 D_n / n = C_1 d_n / \log(n), \]

where the first inequality is due to (A2), and the second inequality is by Lemma 1, and the last is due to (64). With \( E_{V1} = E[\sup_{t \in T, s,s' \in S} |V_1(t', s)V_1(t', s')|] < \infty \) and \( E_{V2} = E[\sup_{t \in T, s,s' \in S} |V_1(t', s_1)V_1(t', s_2)V_1(t', s_3)V_1(t', s_4)|] < \infty \) by (A16),

\[
M_{V,i} := \text{var}(T_i(s, s')) \leq E[T_i(s, s')^2]
\]

\[ \leq \frac{1}{N_{S,t}} \sum_{(j,k,j',k') \in I_i} E[F_{ijjk'k'}(s, s')^2 G_{ijjk'k'}^2]
\]

\[ + \sum_{(j_1,k_1,j_1',k_1') \in I_i} \sum_{(j_2,k_2,j_2',k_2') \in I_i} \sum_{(j_1,k_1,j_1',k_1') \neq (j_2,k_2,j_2',k_2')} E[F_{ij_1k_1j_1'k_1'}(s, s') F_{ij_2k_2j_2'k_2'}(s, s') G_{ij_1k_1j_1'k_1'} G_{ij_2k_2j_2'k_2'}]
\]

\[ \leq \frac{C_3}{N_{S,t}} \left\{ N_i t h_S^4 E_{V1} + N_i t (N_i t - 1) h_S^6 E_{V2} \right\}, \]
where the last inequality is due to the Cauchy–Schwarz inequality. Hence,

\[
\sum_{i=1}^{n} M_{V,i} \leq C_4 d_n^2 / \log(n)
\]

By the Bernstein’s inequality, (68), and (69), for any \(M > 0\),

\[
P\left( \sup_{s,s' \in \mathcal{G}(\kappa)} |H^*(s,s') - E[H^*(s,s')]| \geq M d_n \right) \leq 2 \exp \left\{ - \frac{M^2 d_n^2}{2 \sum_{i=1}^{n} M_{V,i} + M \mu M d_n / 3} \right\} |S| n^{-2\kappa}
\]

(70)

where the second inequality is by Lemma 1, and \(M^*(M) \to \infty\) as \(M \to \infty\).

Applying the Borel-Cantelli lemma to (70),

\[
\sup_{s,s' \in \mathcal{G}(\kappa)} |H^*(s,s') - E[H^*(s,s')]| = O(d_n) \quad \text{a.s.}
\]

(71)

Combining (60), (63)–(67), and (71),

\[
\sup_{s,s' \in S} |H(s,s') - E[H(s,s')]| = O(d_n + n^{-\kappa} / h_S) \quad \text{a.s.}
\]

By (50) and (51) in Lemma 1,

\[
d_n \asymp \{[\log(n) / n] [h_S^4 / m^2 + h_S^8] \}^{1/2}.
\]

(72)

Due to (A15), (A16), and (72), one can choose \(\kappa\) large enough such that \(n^{-\kappa} / h_S = o(d_n)\), so that (71) is \(O(d_n)\), and \(d_n = O(h_S^4 [\log(n) / n]^{1/2})\) by (A15). Now unconditional on the \(N_{it}\), result (53) drops out. \(\Box\)

REFERENCES


