Supplemental Material for *Test of Weak Separability for Spatially Stationary Functional Field*

Decai Liang, Hui Huang, Yongtao Guan and Fang Yao

This supplementary material provides some discussions and additional results of spatial stationarity and Gaussian assumption in Section S.1 and S.2, some implementation issues for the choices of truncation parameter and spatial lag in Section S.3, additional simulation results in Section S.4, and some technical proofs for the propositions and lemmas in Section S.5.

S.1 Issues about the Spatial Stationarity

S.1.1 Some extensions for non-stationary data

To provide theoretical justification of the test statistics based on the lag covariance, we assume the spatial functional field $X(\boldsymbol{s}; t)$ is second-order stationary, which actually contains the following three aspects:

(i) first-order stationarity for the mean function: $\mu(\boldsymbol{s}; t) = \mu(t)$,

(ii) second-order stationarity for the covariance function: $C(\boldsymbol{s}; t_1; \boldsymbol{s}; t_2) = C(t_1; t_2)$,

(iii) second-order stationarity for the cross covariance function: $C(\boldsymbol{s}_1; t_1; \boldsymbol{s}_2; t_2) = C^{(\boldsymbol{h})}(t_1; t_2)$, where $\boldsymbol{h} = \boldsymbol{s}_1 - \boldsymbol{s}_2$.

We clarify that the spatial stationarity as specified above is a different issue from weak separability and is assumed primarily due to the lack of data replication. On the other hand, if replicates of the spatial functional field are available, one can perform the test

*Corresponding author: fyao@math.pku.edu.cn. This research is partially supported by National Natural Science Foundation of China Grants No.11931001, 11871080 and 11871485, China's National Key Research Special Program (Grant No. 2016YFC0207702), the Key Laboratory of Mathematical Economics and Quantitative Finance (Peking University), Ministry of Education. The China PM2.5 data were provided by the Institute of Atmospheric Physics, Chinese Academy of Science.

following a similar procedure as that in our paper. Suppose we observe i.i.d samples $\{X_i(\boldsymbol{s};t); i = 1,\dots,n\}$, and let $\bar{X}(s;t) = n^{-1}\sum_{i=1}^{n} X_i(\boldsymbol{s};t)$, $\tilde{X}_i(\boldsymbol{s};t) = X_i(\boldsymbol{s};t) - \bar{X}(s;t)$. Then $C(\boldsymbol{s};t_1;\boldsymbol{s};t_2)$ and $C(\boldsymbol{s}_1;t_1;\boldsymbol{s}_2;t_2)$ can be estimated by the sample covariance $\hat{C}_{\boldsymbol{s}}(t_1;t_2) = n^{-1}\sum_{i=1}^{n}\tilde{X}_i(\boldsymbol{s};t_1)\tilde{X}_i(\boldsymbol{s};t_2)$ for any $\boldsymbol{s}$, and $\hat{C}_{\boldsymbol{s}_1;\boldsymbol{s}_2}(t_1;t_2) = n^{-1}\sum_{i=1}^{n}\tilde{X}_i(\boldsymbol{s}_1;t_1)\tilde{X}_i(\boldsymbol{s}_2;t_2)$ for any $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$, without extra assumptions such as stationarity. Similarly to our paper, $\lambda_{i;r}(\boldsymbol{s})$ could be estimated by the projection of $X_i(\boldsymbol{s};t)$ onto $\hat{\psi}_r(t)$, which is the r-th eigenfunction of some empirical covariance operator, and we could propose the statistic

$$\begin{aligned} T_n(\boldsymbol{s}_1;\boldsymbol{s}_2;j;k) &= n^{1=2}\sum_{i=1}^{n}\hat{\lambda}_{ij}(\boldsymbol{s}_1)\hat{\lambda}_{ik}(\boldsymbol{s}_2) \\ &= n^{1=2}\sum_{i=1}^{n}\int\int\tilde{X}_i(\boldsymbol{s}_1;t_1)\tilde{X}_i(\boldsymbol{s}_2;t_2)\hat{\psi}_j(t_1)\hat{\psi}_k(t_2)dt_1dt_2 \\ &= \sqrt{n}\int\int\hat{C}_{\boldsymbol{s}_1;\boldsymbol{s}_2}(t_1;t_2)\hat{\psi}_j(t_1)\hat{\psi}_k(t_2)dt_1dt_2 \end{aligned} \tag{S.1}$$

for any $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$. Note that a similar degeneration would occur in (S.1) if $\{\hat{\psi}_r(t)\}$ are the eigenfunctions of $\hat{C}_{\boldsymbol{s}_1;\boldsymbol{s}_2}(t_1;t_2)$. To avoid this, $\hat{\psi}_r(t)$ could be obtained through the eigen-decompostion of the covariance $C(\boldsymbol{s};t_1;\boldsymbol{s};t_2)$ for any $\boldsymbol{s}$, or the marginal covariance (Lynch and Chen, 2018)

$$\hat{C}_S(t_1;t_2) = \frac{1}{n}\sum_{i=1}^{n}\int_S \tilde{X}_i(\boldsymbol{s};t_1)\tilde{X}_i(\boldsymbol{s};t_2)d\boldsymbol{s};$$

which aggregates the covariance information at each $\boldsymbol{s}$. A completely similar argument in our proof of Theorem 1 could be applied to derive the null distribution of $T_n(\boldsymbol{s}_1;\boldsymbol{s}_2;j;k)$. As the proposed weak separability requires that $\lambda_{i;j}(\boldsymbol{s}_1)$ and $\lambda_{i;k}(\boldsymbol{s}_2)$ are uncorrelated for any $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$, one may further consider a collection of $T_n(\boldsymbol{s}_1;\boldsymbol{s}_2;j;k)$ at different $\boldsymbol{s}_1;\boldsymbol{s}_2$ to make the test more powerful. This would be another subject to study.

Besides the above extension to replicated spatio-temporal data, the test can also be carried out by relaxing the first-order stationarity. For example, one could apply a functional regression model

$$X(\boldsymbol{s};t) = \sum_{j=1}^{\infty} Z_j(\boldsymbol{s})\,\beta_j(t) + \varepsilon(\boldsymbol{s};t);$$

where $Z_j(\boldsymbol{s})$'s are some spatial covariates and $\beta_j(t)$'s are the functional coefficients. Through the observed data $\{X(\boldsymbol{s}_i;t)\}_{i=1,\dots,N}$, this model can be then fitted using the method in Ramsay and Silverman (2005). Our test can then be performed based on the de-trended data $X(\boldsymbol{s}_i;t) - \sum_{j=1}^{p} Z_j(\boldsymbol{s})\hat{\beta}_j(t)$, i.e. the residuals $\hat{\varepsilon}(\boldsymbol{s}_i;t)_{i=1,\dots,N}$. However, to test the weak separability of $\varepsilon(\boldsymbol{s};t)$, we still need assumptions about the second-order properties (ii) and (iii).

S.1.2 Sensitivity analysis about stationarity

In practice, one could check the assumed first-order and second-order stationarity for the non-replicated spatio-temporal data through some sensitivity analysis. In the analysis of the China's $PM_{2.5}$ data, taking the NCP region for example. We compare the estimated mean functions of the six subregions of China and four inner subregions in the NCP region, respectively, as displayed in Figure S.1 and S.2. We can observe distinct mean patterns for the six subregions (evidence for non-stationarity) and a similar pattern for the four inner subregions (evidence for stationarity) in NCP. In addition, one may also perform some sensitivity analysis to check the assumed (second-order) stationarity, such as that in Section 8.3 of Zhang and Li (2020) through the FPCA approach. Specifically, we perform the FPCA procedure for the covariance and the lag covariance respectively on the whole NCP region and its two subregions. The corresponding FPC estimates agree well as shown in Figure S.3 and Figure S.4, which suggests that there is no serious violation of the (second-order) stationarity assumption. We have also done such exploratory analyses for the other five regions, as well as for the Harvard Forest data, and reach similar conclusion, which is not reported for brevity.

S.2 Discussion about the Gaussian Assumption

As stated in our paper, the Gaussian assumption in Section 3.3 is applied to simplify the estimation of the asymptotic covariance of the test statistics $T_h(\mathbb{I}_{R_N})$, which relies on the cross fourth-order moments of the FPC scores, i.e., $E(\xi_{i_1 j}\xi_{i_2 k}\xi_{i_3 j'}\xi_{i_4 k'})$, $i_1, i_2, i_3, i_4 = 1, \ldots, N$, $(j,k), (j',k') \in \mathbb{I}_{R_N}$, as shown in (28) of our manuscript. Since only one realization of the spatial functional field or the spatio-temporal data can be observed, the ordinary moment estimators for the asymptotical covariance (e.g. Aston et al., 2017; Lynch and Chen, 2018) become unfeasible. This is essentially different from the replicated or two-way functional data and gives rise to challenges for statistical inference.

S.2.1 Sensitivity analysis about the Gaussian assumption

To investigate the sensitivity of the proposed test to the Gaussian assumption, we perform a simulation study based on the χ^2 random fields (Bevilacqua et al., 2020; Ma, 2009) to investigate the sensitivity of the proposed test to the Gaussian assumption. A $\chi^2(\nu)$ random field $Y(t)$ with ν degrees of freedom is defined by $Y(t) = \sum_{i=1}^{\nu} X_i^2(t)$, where $\{X_i(t), i = 1, \ldots, \nu\}$ are a series of independent, stationary Gaussian processes with mean 0 and variance 1 (Lindgren et al., 2013). The R package "RandomFields" (Schlather et al., 2015) is used to simulate χ^2 random fields with isotropic Matern covariances. Specifically, we consider the

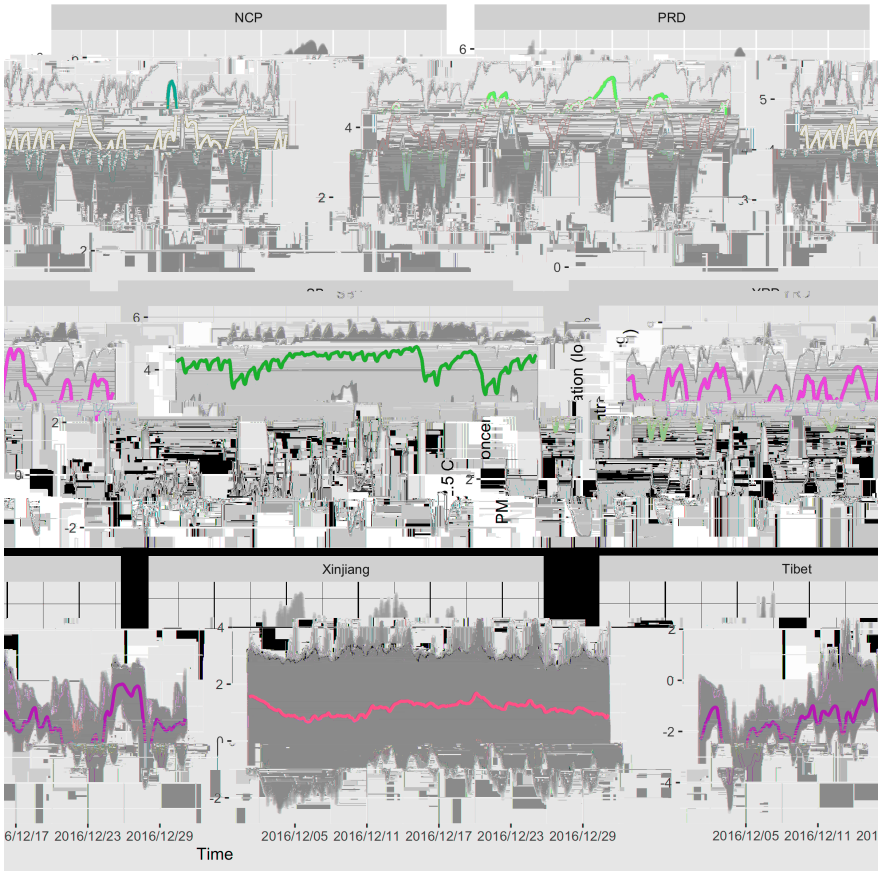


Figure S.1: Raw data (in gray) and the empirical mean functions of the six subregions in China.

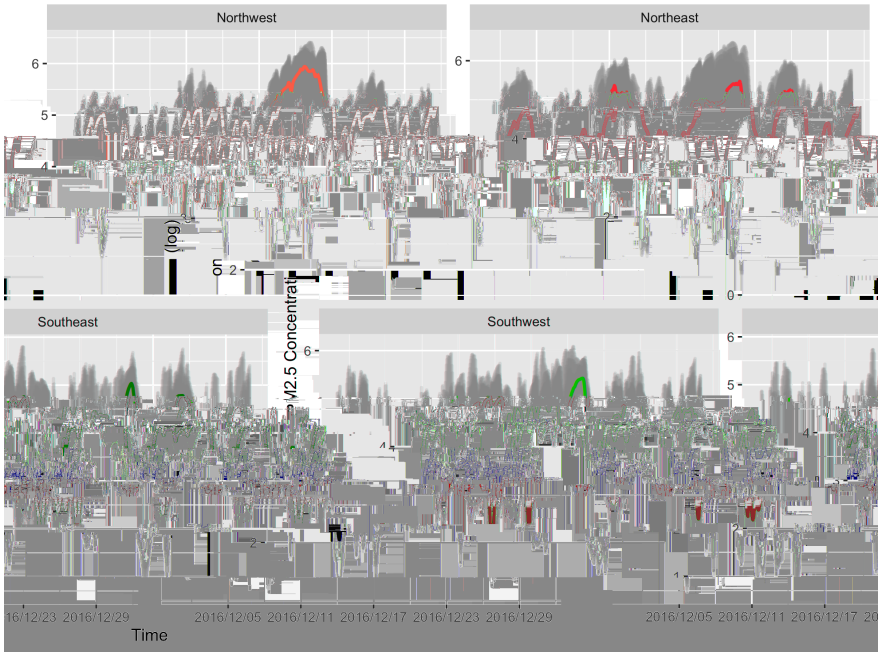
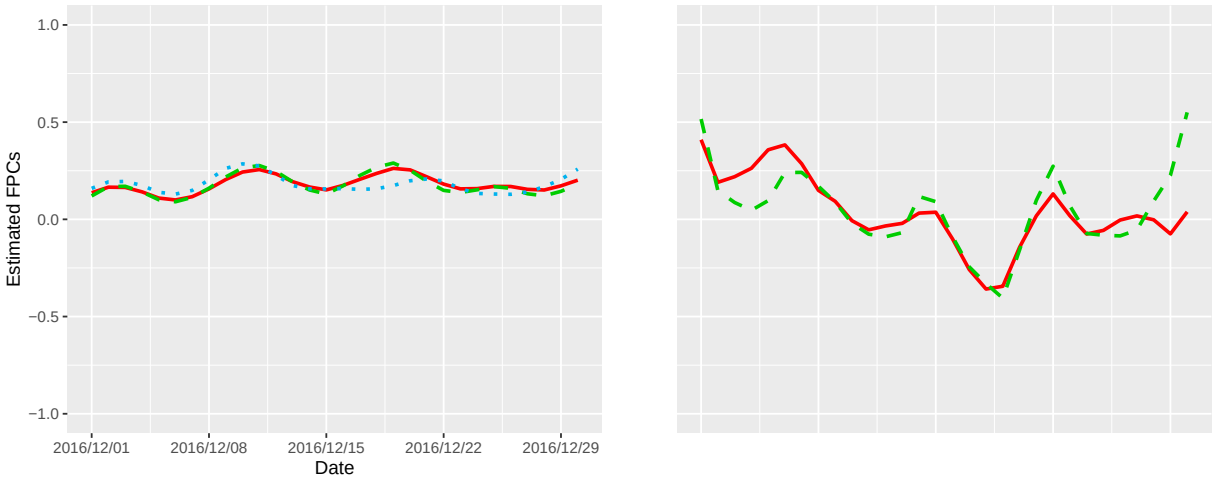
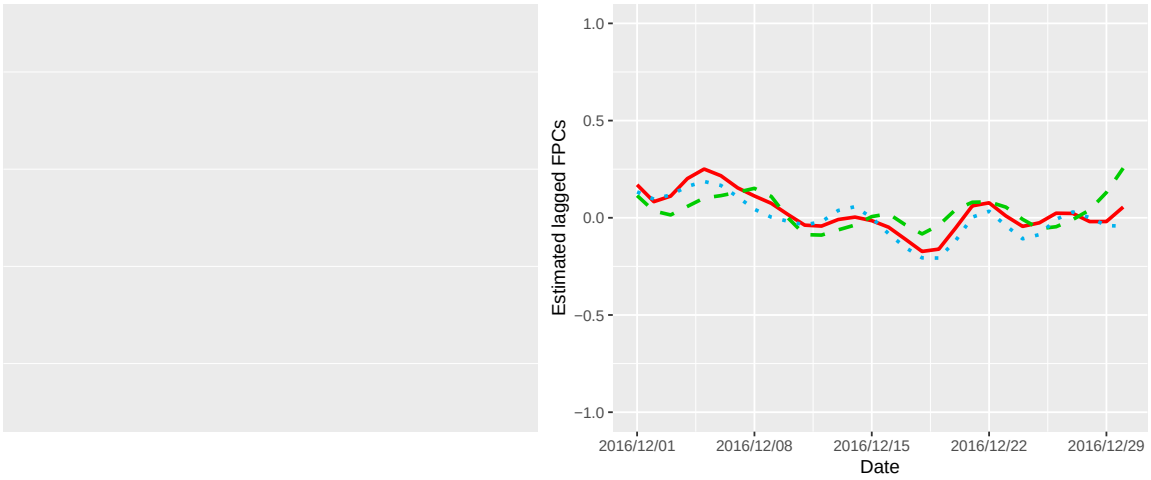


Figure S.2: Raw data (in gray) and the empirical mean functions of the four inner subregions NCP, each of which is extracted with a 20 20 grid at northwest, northeast, southwest or southeast NCP.



(a) $\hat{\psi}_1(t)$ (b) $\hat{\psi}_2(t)$

Figure S.3: Sensitivity analysis on the China's $PM_{2:5}$ data. The red lines are the estimated eigenfunctions of covariance $\hat{C}(t_1; t_2)$ using the whole NCP region, while the green dashed lines and blue dotted lines are the corresponding estimated eigenfunctions using two subregions (northeast and southwest) of NCP.



(a) $\hat{\psi}_1^{(h)}(t)$ (b) $\hat{\psi}_2^{(h)}(t)$

Figure S.4: Sensitivity analysis on the China's $PM_{2:5}$ data. The red lines are the estimated eigenfunctions of the lag covariance $\hat{C}^h(t_1; t_2)$ with $h = 1$ using the whole NCP region, while the green dashed lines and blue dotted lines are the corresponding estimated eigenfunctions using two subregions (northeast and southwest) of NCP.

following two scenarios:

1. $_3(\ )$ and $_4(\ )$ are generated marginally from two $^2(6)$ random elds with the same variance $!_r$ and scale parameter $_r$ as in the Gaussian case in our paper for $r = 3; 4$;

2. $_3(\ )$ and $_4(\ )$ are generated from two $^2(8)$ random elds with the same $!_r$ and $_r$ as in the Gaussian case for $r = 3; 4$;

Tables 1 and 2 report the empirical type I errors and power respectively for the di erent scenarios: (a), $_3$, $_4$ $^2(6)$; (b), $_3$, $_4$ $^2(8)$. For the alternative hypothesis, the correlation coe cient $_{12}$ for the bivariate gaussian random elds $f\ _1(\ ),\ _2(\ )g$ is set as 0.4 for cases (a) and (b). We can see that the overall performance for the non-Gaussian cases are comparable to that for the Gaussian case. These results provide evidence for the robustness of our method to the Gaussian assumption.

	Gaussian		$_3$, $_4$ $^2(6)$		$_3$, $_4$ $^2(8)$	
FVE	Para	Nonp	Para	Nonp	Para	Nonp
80%	4.6	4.2	4.0	3.4	5.4	5.0
90%	4.8	5.4	4.2	3.6	5.6	5.4
95%	4.8	6.4	6.6	5.4	5.4	6.2

Table 1: Empirical Type I errors (%) for the weak separability tests based on the lag-1 covariance under Gaussian or non-Gaussian settings, with the asymptotic covariance estimated by parametric (Para) and non-parametric (Nonp) methods.

	Gaussian		$_3$, $_4$ $^2(6)$		$_3$, $_4$ $^2(8)$	
FVE	Para	Nonp	Para	Nonp	Para	Nonp
80%	100	100	100	100	100	100
90%	95.0	94.5	100	100	100	100
95%	100	100	100	100	100	100

Table 2: Empirical power (%) for the weak separability tests based on the lag-1 covariance under Gaussian or non-Gaussian settings, with the asymptotic covariance estimated by parametric (Para) and non-parametric (Nonp) methods.

S.2.2 Additional results of the block bootstrap

We also make attempt to approximate using resampling methods. As only one realization of $X(\boldsymbol{s}; t)$ can be obtained, standard re-sampling methods for i.i.d. data such as bootstrap (Aston et al., 2017; Constantinou et al., 2017) are no longer valid. As a possible alternative, one could use block bootstrap to approximate . Speci cally, we perform block bootstrap

following the procedures in Section 10.6 of Sherman (2011). Let $T(S_N) := T_h(\mathbb{I}_{R_N})$ be the test statistic (de ned in Section 3.3 of our paper) computed from data in the overall spatial domain S_N, and $\{S^l_{M_N}; l = 1, \ldots, L\}$ be L-overlapping sub-blocks of S_N where M_N determines the size of each sub-block. Then, can be estimated by

$$\hat{b}_{BB} = \frac{1}{L} \sum_{l=1}^{L} |S^l_{M_N}| \left(T(S^l_{M_N}) - T_N \right) \left(T(S^l_{M_N}) - T_N \right)^{\mathrm{T}},$$

where $T_N = \sum_{l=1}^{L} T(S^l_{M_N})/L$. The corresponding χ^2 test can then be performed based on $\hat{b}_{BB}$.

Table 3: Empirical type I errors (%) for the block bootstrap tests with di erent M_N and ρ_1.

	$M_N = 10 \times 10$			$M_N = 8 \times 8$		
ρ_1	FVE=80%	FVE=90%	FVE=95%.	FVE=80%	FVE=90%	FVE=95%
0.2	5.8	24.0	35.6	8.4	31.2	46.0
0.1	4.4	4.4	10.8	4.0	4.0	14.4
0.05	6.4	6.4	7.4	6.8	6.8	7.8

We conduct a simulation using with the same settings as in our paper to check the performance of block bootstrap, where $N = 40 \times 40$ and the sub-block size M_N is set as 10×10 or 8×8. As it is known that the performance of block bootstrap is a ected by the strength of spatial dependence, we try di erent values for ρ_1, the range parameter of the rst FPC scores which primarily determines the spatial dependence of $X(\boldsymbol{s}, t)$. Speci cally, we set $\rho_1 = 0.2, 0.1, 0.05$, which correspond to relatively strong, moderate, or weak spatial dependece. As shown in Table 3, the type I error of the block bootstrap test is much in ated in the relatively strong spatial dependence case, especially when FVE is larger than 90%. When $\rho_1 = 0.1$ and 0.05, the block bootstrap test could achieve relatively reasonable size, except when $\rho_1 = 0.1$ and FVE=95%. It can also be seen that the block bootstrap test with the sub-block size $M_N = 10 \times 10$ performs better than that with $M_N = 8 \times 8$. These results indicate that block bootstrap may not be a good choice for the proposed weak separability test for a general spatial functional eld with strong spatial dependence.

S.3 More Implementation Issues

S.3.1 Implementation issues about the truncation parameter

In our paper, Conditions 4 and 4* provide upper bounds of R_N to allow (but not require) the truncation slowly diverges. This is in fact a major theoretical challenge in most functional

data models, as the perturbation theory becomes more intricate, which is necessary in order to re ect the genuine nonparametric nature of such models.

We emphasize that R_N in our test is not a tuning parameter, i.e, we do not choose a speci c R_N, but instead inspect a range of R_N in order to include enough signi cant FPCs. As stated in Conditions 4 and 4*, a reasonable truncation R_N also depends on

S.3.2 Multiple tests using di erent lags

According to the proposed de nition, the claim of weak separability in general should be a comprehensive conclusion across di erent lag h. However, it is common for spatial data that observations separated by smaller lags are more correlated (Sherman, 2011), which could result in more powerful tests using smaller lags. We also attempt the multiple tests to combine the information from di erent lags. Speci cally, we perform the Bonferroni correction through the following procedure. Assume we have $n = 4$ hypothesis tests (A_1, A_2, A_3, A_4) with each of the null hypothesis is that, $H_{0;h}$: the eigen-decomposition of lag-covariance $C^{(h)}(t_1; t_2)$ (see equation (10) in our paper) holds for any $fX(\boldsymbol{s};\ ); X(\boldsymbol{s}+h;\ )g$ for $h = 1; 2; 3; 4$. Then the global null hypothesis H_0 is that each of $fH_{0;h}g_{h=1;:::;4}$ holds, and H_0 is rejected if at least one of the $fH_{0;h}g_{h=1;:::;4}$ is rejected using the corrected signi cant level $=n$. (According to the de nition of weak separability, the global null hypothesis H_0 should be that the $H_{0;h}$ holds for any h; however, it is obviously unfeasible to perform tests for all possible h, so we just combine lag-1 to lag-4.)

We apply the Bonferroni correction in our simulation and compare its performance with the lag-1 to lag-4 covariances. The results are displayed in Table 5, where we can see that the test using the Bonferroni correction is more powerful than those using the lag-2 to lag-4 covariances, but is less powerful than that using the lag-1 covariance. Based on the new empirical evidences, we feel that the new testing procedure combining information from di erent lags through the Bonferroni correction may not lead to a more powerful test for weak separability, so we suggest using the lag-1 covariance for grid data.

Table 5: Rejection rates (%) for the weak separability tests based on the lag-1 to lag-4 covariances and Bonferroni correction.

		lag-1		lag-2		lag-3		lag-4		Bonferroni	
$_{12}$	FVE	Para	Nonp	Para	Nonp	Para	Nonp	Para	Nonp	Para	Nonp
0	80%	4.6	4.2	6.6	4.0	8.2	5.2	6.2	4.4	5.4	2.8
	90%	4.4	4.6	6.2	4.8	8.0	5.2	6.2	4.4	5.4	4.2
	95%	4.8	6.4	6.6	7.6	7.2	5.6	5.6	4.0	5.6	5.8
0.2	80%	85.0	87.5	76.5	69.5	31.5	30.5	12.0	15.0	76.0	77.0
	90%	74.5	79.5	74.0	67.5	31.0	30.5	12.0	15.5	70.5	71.5
	95%	70.5	77.5	63.5	63.0	28.5	32.0	11.0	16.5	63.5	65.0
0.4	80%	100	100	94.0	94.0	68.0	68.0	32.5	37.0	100	100
	90%	100	100	95.0	94.5	68.5	68.5	32.5	37.5	100	100
	95%	100	100	98.5	98.0	74.0	78.0	31.0	39.0	99.5	100
0.6	80%	99.0	99.5	87.5	89.0	67.5	68.0	37.0	41.0	98.5	99
	90%	100	99.5	90.0	91.0	68.0	68.0	37.0	41.0	99.5	99.5
	95%	100	100	99.5	99.5	79.5	81.5	41.5	47.5	100	100

We also perform multiple tests on the China's $PM_{2.5}$ data using the Bonferroni correction with lag-1 to lag-4 covariance. The corrected signi cant level is 0.05=4 = 0.0125. Tables 6 to 9 report the p-values of these tests across di erent R_N (at 80%, 90%, 95% FVE as in Section 5.1 of our paper). For NCP and YRD, the two regions for which the weak separability test using the lag-1 covariance was passed, the p-values across di erent lags are all larger than 0.0125, which is consistent with our previous results. In contrast, for Xinjiang and Tibet for which weak separability was rejected using the lag-1 covariance, the p-values lead to di erent conclusions, although the p-values for the lag-1 covariance are all less than 0.01. This is not surprising given that our simulation results reveal that the test based on the lag-1 covariance is more powerful than the others.

		lag-1	lag-2	lag-3	lag-4
$R_N = 2$	Para	0.136	0.313	0.405	0.442
	Nonp	0.287	0.398	0.432	0.439
$R_N = 3$	Para	0.150	0.123	0.111	0.135
	Nonp	0.212	0.103	0.072	0.079
$R_N = 6$	Para	0.759	0.707	0.768	0.811
	Nonp	0.289	0.100	0.535	0.554

Table 6: The p-values of multiple tests with lag-1 to lag-4 covariance for NCP region.

		lag-1	lag-2	lag-3	lag-4
$R_N = 3$	Para	0.925	0.868	0.794	0.653
	Nonp	0.620	0.374	0.219	0.128
$R_N = 6$	Para	0.863	0.996	0.999	0.995
	Nonp	0.399	0.791	0.676	0.477
$R_N = 9$	Para	0.997	1	0.999	0.995
	Nonp	0.119	0.424	0.430	0.587

Table 7: The p-values of multiple tests with lag-1 to lag-4 covariance for YRD region.

		lag-1	lag-2	lag-3	lag-4
$R_N = 2$	Para	0.002	0.087	0.394	0.612
	Nonp	0.000	0.003	0.133	0.518
$R_N = 5$	Para	0.000	0.073	0.309	0.000
	Nonp	0.000	0.011	0.052	0.000
$R_N = 8$	Para	0.000	0.000	0.423	0.000
	Nonp	0.000	0.000	0.058	0.000

Table 8: The p-values of multiple tests with lag-1 to lag-4 covariance for Xinjiang region.

		lag-1	lag-2	lag-3	lag-4
$R_N = 3$	Para	0.000	0.485	0.169	0.000
	Nonp	0.000	0.027	0.213	0.093
$R_N = 5$	Para	0.000	0.000	0.000	0.000
	Nonp	0.000	0.098	0.002	0.053
$R_N = 8$	Para	0.000	0.000	0.000	0.000
	Nonp	0.000	0.000	0.000	0.001

Table 9: The p-values of multiple tests with lag-1 to lag-4 covariance for Tibet region.

S.4 Additional Results for Simulation Study

We present the additional simulation results on the test performance with different $_1$, the range parameter of the first FPC scores which primarily determines the spatial dependence of $X(\boldsymbol{s}; t)$. The case $_1 = 0.15$, 0.2 and 0.25 can be respectively regarded as the \weak", \medium" and \strong" spatial correlation. Table 10 summarizes the empirical rejection rates from 200 simulation trials. For each $_1$, the first row $_{12} = 0$ yields the null hypothesis, and we see that both parametric and non-parametric tests have stably reasonable size across different choices of lags. We know that for data without spatial correlation (which implies $_1 = 0$), the KL expansion is always correct, thus the weak separability holds even for $_{12} > 0$. The rejection of weak separability tends to be more significant when the spatial correlation is stronger, as shown in Table 10 that the tests appear to obtain larger power for the \medium" and \strong" scenarios ($_1 = 0.2$ and 0.25). It is obvious that the lag-1 covariance outperforms the others for most scenarios except ($_1$; $_{12}$) = (0.25; 0.2) and ($_1$; $_{12}$) = (0.25; 0.4), and the test performance dramatically deteriorates as the lag number increases, especially when $_{12} = 0.4$ or 0.6.

S.5 Technical Proofs

S.5.1 Proof of Proposition 1 and Proposition 2

Proof of Proposition 1. For a strongly separable $X(\boldsymbol{s}; t)$, we see that the time covariance $C(t_1; t_2)$ implies the eigen-decomposition in time domain $C(t_1; t_2) = \sum_{r=1}^{1} !_r \ _r(t_1) \ _r(t_2)$ for some orthogonal basis system $f \ _r(t)g$. To show that $X(\boldsymbol{s}; t)$ is weakly separable, we only need to show that the projected scores $f \ _r(\boldsymbol{s})g$ on basis $f \ _r(t)g$ are mutually uncorrelated. For $j \neq k$, and any $\boldsymbol{s}_1$, $\boldsymbol{s}_2$, we have

$$\mathrm{E}f \ _j(\boldsymbol{s}_1) \ _k(\boldsymbol{s}_2)g = \mathrm{E} \int fX(\boldsymbol{s}_1; t) \quad (\boldsymbol{s}_1; t)g \ _j(t)dt \int fX(\boldsymbol{s}_2; t) \quad (\boldsymbol{s}_2; t)g \ _k(t)dt$$

Table 10: Rejection rates (%) for the parametric (Para) and non-parametric (Np) weak separability test based on lag-1 to lag-4 covariance, with di erent $_1$ and $_{12}$ values. The FVE threshold value is 95%.

$_1$	$_{12}$	lag 1		lag 2		lag 3		lag 4	
		Para	Np	Para	Np	Para	Np	Para	Np
0.15	0	6.0	3.5	7.0	7.0	4.0	5.0	3.5	3.5
	0.2	39.5	44.0	13.0	21.5	7.0	10.0	3.5	5.0
	0.4	96.0	98.0	44.0	26.0	18.5	27.0	5.0	7.5
	0.6	100	100	73.5	82.0	26.0	35.0	5.5	10.0
0.2	0	4.5	6.5	6.5	6.5	5.5	7.5	3.5	4.5
	0:2	70.5	77.5	63.5	62.5	27.5	31.5	11.0	15.0
	0:4	100	100	98.5	98.0	73.5	77.0	31.0	39.0
	0:6	100	100	99.5	99.5	79.5	81.5	41.5	47.5
0.25	0	5.0	3.5	3.5	6.5	3.0	6.0	2.5	4.5
	0.2	57.5	66.0	72.5	68.0	54.5	50.5	29.0	24.5
	0.4	98.5	98.5	100	100	88.0	87.0	54.0	56.5
	0.6	100	100	100	100	93.5	92.0	64.5	68.5

$$
\begin{aligned}
&= \mathrm{E}\int\int f X(\boldsymbol{s}_1;t_1) \quad (\boldsymbol{s}_1;t_1)g\, f X(\boldsymbol{s}_2;t_2) \quad (\boldsymbol{s}_2;t_2)g \quad _j(t_1) \quad _k(t_2)dt_1dt_2 \\
&= \int\int C(\boldsymbol{s}_1;t_1;\boldsymbol{s}_2;t_2) \quad _j(t_1) \quad _k(t_2)dt_1dt_2 \\
&= \int\int C(\boldsymbol{s}_1;\boldsymbol{s}_2)\, C(t_1;t_2) \quad _j(t_1) \quad _k(t_2)dt_1dt_2 \\
&= C(\boldsymbol{s}_1;\boldsymbol{s}_2) \int\int C(t_1;t_2) \quad _j(t_1) \quad _k(t_2)dt_1dt_2 \\
&= C(\boldsymbol{s}_1;\boldsymbol{s}_2) \int \quad !_j \quad _j(t_2) \quad _k(t_2)dt_2 = 0:
\end{aligned}
$$

The conversion between integral and double integral is allowed by the Fubini theorem. □

Proof of Proposition 2. In the proof of Proposition 1, let $j = k$ we have

$$\mathrm{E}f \ _j(\boldsymbol{s}_1) \ _j(\boldsymbol{s}_2)g = C(\boldsymbol{s}_1;\boldsymbol{s}_2) \int \quad !_j \quad _j(t_2) \quad _j(t_2)dt_2 = \ !_j C(\boldsymbol{s}_1;\boldsymbol{s}_2)$$

for each j. For a weakly separable process $X(\boldsymbol{s};t)$ with $\mathrm{E}f \ _r(\boldsymbol{s}_1) \ _r(\boldsymbol{s}_2)g = \ !_r \ (\boldsymbol{s}_1;\boldsymbol{s}_2)$, one could derive that $C(\boldsymbol{s}_1;t_1;\boldsymbol{s}_2;t_2) = \ (\boldsymbol{s}_1;\boldsymbol{s}_2) \sum_{r=1}^{1} \ !_r \ _r(t_1) \ _r(t_2) = \ (\boldsymbol{s}_1;\boldsymbol{s}_2)C(t_1;t_2)$ using equation (6), thus $X(\boldsymbol{s};t)$ is strongly separable. □

S.5.2 Proof of Lemmas

Proof of Lemma 1. This lemma comes directly from Theorem 8.1.2 of Hsing and Eubank (2015) using the tightness and convergence of finite dimension projection argument as in Theorem 7.7.6 of Hsing and Eubank (2015). □

Proof of Lemma 2. The proof of Lemma 2 utilizes the central limit theorem for strongly mixing processes in Hilbert space, which is a natural extension of that in real-valued spaces (see e.g. Theorem 1.7 of Bosq, 2012). To be more specific, denote $U_i = (X_i - \bar{X}) \otimes (X_i - \bar{X})$ and $S_N = N^{-1/2} \sum_{i=1}^{N} U_i$. Under Conditions 1 and 2, the series $\sum_{i=1}^{N} \mathrm{Cov}(U_1, U_i)$ converge absolutely to an element σ_U^2 in $L^2(\mathcal{T}) \otimes L^2(\mathcal{T})$ using the Davydovs Inequality (Bosq, 2012) in Hilbert space with $q = r = v/2$ and $1/p = 1 - 2/r$, and then $\mathrm{Var}(S_N)$ converge absolutely to σ_U^2. To prove the asymptotic normality of S_N, we apply a blocking technique as in Bosq (2012); Guan et al. (2004) and denote $S_N' = N^{-1/2} \sum_{i=1}^{\mathcal{N}} V_i$, where $V_i = U_{(i-1)(p+q)+1} + \cdots + U_{ip+(i-1)q}$ with $\mathcal{N} \sim \log N$, $p \sim N/(\log N) - N^{1/4}$ and $q \sim N^{1/4}$. Using the coupling results (Lemma 1.2 of Bosq, 2012), we can construct independent random elements $W_1, \ldots, W_{\mathcal{N}}$ s.t. W_i has the same distribution as V_i, and the probability $P\left(|N^{-1/2} \sum_i (V_i - W_i)| > \epsilon\right)$ can be shown to converge to 0. The asymptotic normality of $N^{-1/2} \sum_i W_i$ follows from the i.i.d case in Lemma 1, and it can also be shown that $S_N' - S_N \xrightarrow{p} 0$. Consequently the asymptotic normality of $N^{1/2}(\widehat{\mathcal{C}} - \mathcal{C})$ holds, and similarly the asymptotic normality of $N_h^{1/2}\{\widehat{\mathcal{C}}^{(h)} - \mathcal{C}^{(h)}\}$ can be obtained. Finally the proof of joint normality follows directly by applying the Cramer-Wold device (e.g. Guan et al., 2004). □

Proof of Lemma 3. To prove Lemma 3(a), we first let $\widehat{\Delta} = \|\widehat{\mathcal{C}} - \mathcal{C}\|_{HS}$, $\widehat{\Delta}_{(h)} = \|\widehat{\mathcal{C}}^{(h)} - \mathcal{C}^{(h)}\|_{HS}$. Under Condition (17) and by the Lemma 3.3 of Hall and Hosseini-Nasab (2009), we have $\mathrm{E}(\widehat{\Delta}^C) = O(N^{-C/2})$ for any $C > 0$, and similarly $\mathrm{E}(\widehat{\Delta}_{(h)}^C) = O(N^{-C/2})$ by the proof of Lemma 2. Let $\mathcal{K}_N = \{r = 1, \ldots, \infty : \lambda_r^{(h)} - \lambda_{r+1}^{(h)} > 2\widehat{\Delta}_{(h)}\}$; that is, the index set containing $r \in \mathcal{K}_N$ for $\lambda_{r+1}^{(h)}$ for obt30(fo44(ha)e 11.9552 Tf 13.828 0 T693.85021)]TJ/F54 7.9701 Tf 6.204 6.111 [illegible]

$$+\sum_{k:k\neq j}(\lambda_j^{(h)}-\lambda_k^{(h)})^{-1}\int\phi^{(h)}C^{(h)}(\psi_j^{(h)}-\psi_j)\,\psi_k$$
$$+\sum_{k:k\neq j}\frac{\lambda_j^{(h)}-\lambda_j^{(h)}}{(\lambda_j^{(h)}-\lambda_k^{(h)})(\lambda_j^{(h)}-\lambda_k^{(h)})}\int\phi^{(h)}C^{(h)}\psi_j^{(h)}\,\psi_k:$$

Denote the last three terms by $\Delta_{1j}+\Delta_{2j}+\Delta_{3j}$, and let $M_{k;j}=(\lambda_j^{(h)}-\lambda_k^{(h)})^{-1}\int(\phi^{(h)}-C^{(h)})\psi_j\psi_k$, $\delta_j=\min_{r=1,\ldots,j}(\lambda_r^{(h)}-\lambda_{r+1}^{(h)})$. According to (6.55) of Hall and Hosseini-Nasab (2009), for each integer $b>0$,

$$E(\sum_{k:k\neq j}M_{k;j}^2)^b\le C_b(j^2N^{-1})^b \tag{S.2}$$

uniformly in $j\in\mathcal{K}_N$, under the assumption that $E(\int fX(s)g_j)^{2b}\le C_b[E\int fX(s)\psi_j g^2]^b$ which is implied by condition (18) (see e.g. (2.18) of Hall and Hosseini-Nasab, 2009). Following (6.58) of Hall and Hosseini-Nasab (2009), we have $\|\psi_j^{(h)}-\psi_j\|^2\le 32(\sum_{k:k\neq j}M_{k;j}^2)$ on $E_{N;R_N}$, which leads to the result in Lemma 3(b) that $E\|\psi_j^{(h)}-\psi_j\|^{2b}=O(j^2N^{-1})^b$. To derive the bound for $E\|\Delta_j\|^2$ in Lemma 3(c), we first note that $\|\Delta_{1j}\|=|\int(\psi_j^{(h)}-\psi_j)\psi_j|=\|\psi_j^{(h)}-\psi_j\|^2/2$, leading to $E\|\Delta_{1j}\|^2=O(j^4N^{-2})$ letting $b=2$ in (S.2). To bound Δ_{2j}, noting $\|\sum_r c_r\psi_r\|^2=\sum_r c_r^2$ due to orthonormal $\{\psi_r\}$ and using Cauchy-Schwarz inequality,

$$\|\Delta_{2j}\|^2\le\sum_{k:k\neq j}(\lambda_j^{(h)}-\lambda_k^{(h)})^2\int f(\beta_{(h)}^4)E(\|\psi_j^{(h)}-\psi_j\|^4)g^{1=2}=O(j^{2a+4}N^{-2})$$

letting $C=4$ in $E(\beta_{(h)}^C)=O(N^{-C=2})$ and $b=2$ in (S.2). For Δ_{3j}, noting that $\lambda_j^{(h)}-\lambda_j=\beta_{(h)}$ and $(\lambda_j^{(h)}-\lambda_k^{(h)})^{-2}\le 2(\lambda_j^{(h)}-\lambda_k^{(h)})^{-2}$, we have

$$\|\Delta_{3j}\|^2=(\lambda_j^{(h)}-\lambda_j)^2\sum_{k:k\neq j}(\lambda_j^{(h)}-\lambda_k^{(h)})^{-2}(\lambda_j^{(h)}-\lambda_k^{(h)})^{-2}\int f(\beta_{(h)}^2)\phi^{(h)}C^{(h)}\psi_j^{(h)}\psi_k g^2$$
$$\le 2\beta_{(h)}^2\sum_{k:k\neq j}(\lambda_j^{(h)}-\lambda_k^{(h)})^{-4}\int f(\beta_{(h)}^2)\phi^{(h)}C^{(h)}\psi_j\psi_k g^2$$
$$+\sum_{k:k\neq j}(\lambda_j^{(h)}-\lambda_k^{(h)})^{-4}\int f(\beta_{(h)}^2)\phi^{(h)}C^{(h)}(\psi_j^{(h)}-\psi_j)\psi_k g^2\Big\}:$$

Let $A_1=\beta_{(h)}^2\sum_{k:k\neq j}(\lambda_j^{(h)}-\lambda_k^{(h)})^{-4}\int f(\phi^{(h)}-C^{(h)})\psi_j\psi_k g^2$ and $A_2=\beta_{(h)}^2\sum_{k:k\neq j}(\lambda_j^{(h)}-\lambda_k^{(h)})^{-4}\int f(\phi^{(h)}-C^{(h)})(\psi_j^{(h)}-\psi_j)\psi_k g^2$. It then follows that $A_1\le\beta_{(h)}^2(\sum_{k:k\neq j}M_{k;j}^2)$ and

$A_2 \quad {}_j{}^4 b^4_{(h)} k^{\wedge(h)}_j \quad {}_j k^2$ (see also the proof of Lemma 1(c) of Kong et al., 2016), which leads to $EA_1 \quad {}_j{}^2 f E(b^4_{(h)}) E(\sum_{k:k6=j} M^2_{k;j})^2 g^{1=2} = O(j^{2a+4} N^{-2})$ and $EA_2 \quad {}_j{}^4 f E(b^8_{(h)}) E(k^{\wedge(h)}_j \quad {}_j k)^4 g^{1=2} = O(j^{4a+6} N^{-3})$ using the similar arguments as before. Then $EA_2 = o(j^{2a+4} N^{-2})$ by Condition 4, leading to $E k_{\ 3j} k^2 = O(j^{2a+4} N^{-2})$. Combining these results we obtain that $E k_{\ j} k^2 = O(j^{2a+4} N^{-2})$ uniformly in $j = 1;:::;R_N$. □

References

Aston, J. A., Pigoli, D., Tavakoli, S. et al. (2017), "Tests for separability in nonparametric covariance operators of random surfaces," *The Annals of Statistics*, 45(4), 1431–1461.

Bevilacqua, M., Caamaño-Carrillo, C., and Gaetan, C. (2020), "On modeling positive continuous data with spatio-temporal dependence," *Environmetrics*, p. e2632.

Bosq, D. (2012), *Nonparametric statistics for stochastic processes: estimation and prediction*, Vol. 110 Springer Science & Business Media.

Constantinou, P., Kokoszka, P., and Reimherr, M. (2017), "Testing separability of space-time functional processes," *Biometrika*, 104(2), 425–437.

Guan, Y., Sherman, M., and Calvin, J. A. (2004), "A nonparametric test for spatial isotropy using subsampling," *Journal of the American Statistical Association*, 99(467), 810–821.

Hall, P., and Horowitz, J. L. (2007), "Methodology and convergence rates for functional linear regression," *The Annals of Statistics*, 35(1), 70–91.

Hall, P., and Hosseini-Nasab, M. (2009), Theory for high-order bounds in functional principal components analysis,, in *Mathematical Proceedings of the Cambridge Philosophical Society*, Vol. 146, Cambridge University Press, pp. 225–256.

Hsing, T., and Eubank, R. (2015), *Theoretical foundations of functional data analysis, with an introduction to linear operators* John Wiley & Sons.

Kong, D., Xue, K., Yao, F., and Zhang, H. H. (2016), "Partially functional linear regression in high dimensions," *Biometrika*, 103(1), 147–159.

Lindgren, G., Rootzen, H., and Sandsten, M. (2013), *Stationary stochastic processes for scientists and engineers* CRC press.

Lynch, B., and Chen, K. (2018), "A test of weak separability for multi-way functional data, with application to brain connectivity studies," *Biometrika*, 105(4), 815–831.

Ma, C. (2009), " ^{2} Random Fields in Space and Time," *IEEE transactions on signal processing*, 58(1), 378–383.

Ramsay, J., and Silverman, B. (2005), *Functional Data Analysis*, 2nd edn Springer.

Sherman, M. (2011), *Spatial Statistics and Spatio-temporal Data: Covariance Functions and Directional Properties* John Wiley & Sons.

Zhang, H., and Li, Y. (2020), \Uni ed Principal Component Analysis for Sparse and Dense Functional Data under Spatial Dependency," *arXiv preprint arXiv:2006.13489*, .