Powerful Spatial Multiple Testing via Borrowing Neighboring Information

Linsui Deng, Kejun He and Xianyang Zhang

Renmin University of China and Texas A&M University

Supplementary Material

The supplement is organized as follows. Section S.I includes some details on the simulation studies and the additional results of the ozone data analysis. Section S.II presents two main lemmas and the proof of Theorem 1. The two main lemmas are proved in Section S.III with some preliminary lemmas and discussions. Section S.IV provides detailed proofs of the preliminary lemmas. Section S.V proves the theory of the power improvement of the 2d-SMT compared to the 1d-SMT and provides a concrete example to investigate the power improvement. We discuss the estimation for the covariance of noises in Section S.VI. Section S.VII thoroughly describes the details of Algorithm 1 in Section 3 of the main paper which overcomes the computational bottleneck of a naive grid search. The R package and reproducible code of this work are accessible at https://github.com/denglinsui/TwoDSMT and https://github.com/denglinsui/2dSMT-manuscript-sourcecode, respectively.

S.I Additional Experimental Results

This section provides additional results for the simulations and the real data analysis with some implementation details. In particular, Section S.I.1 visualizes the simulation settings

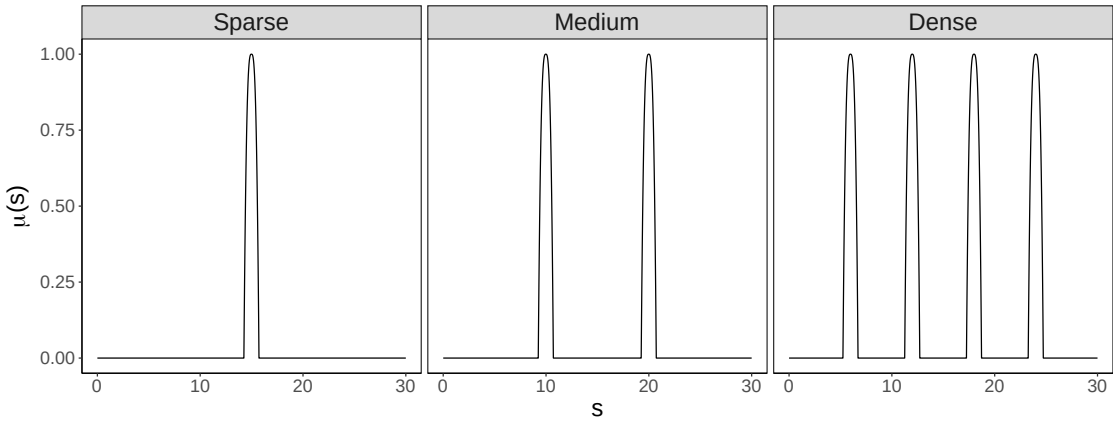
in Section 5 of the main paper. Section S.I.2 first reports the FDPs and powers under Setup II with location size $m = 900$, and then presents the results of location size $m = 2000$ under Setups I–III. Section S.I.3 describes and reports the simulations studies for a two-dimensional domain where the covariance is unknown and needs to be estimated. In Section S.I.4, we investigate the sensitivity of 2d-SMT to the number of observations in the nearest neighbors. Section S.I.5 explores a data-adaptive approach to determine the number of neighbors for each location. In Section S.I.6, we examine the integration of covariate and spatial information within the framework of the 2d-SMT method. Section S.I.7 depicts the partition of the Contiguous United States into nine regions, and provides analytical and numerical findings related to the ozone data.

S.I.1 Simulation Settings

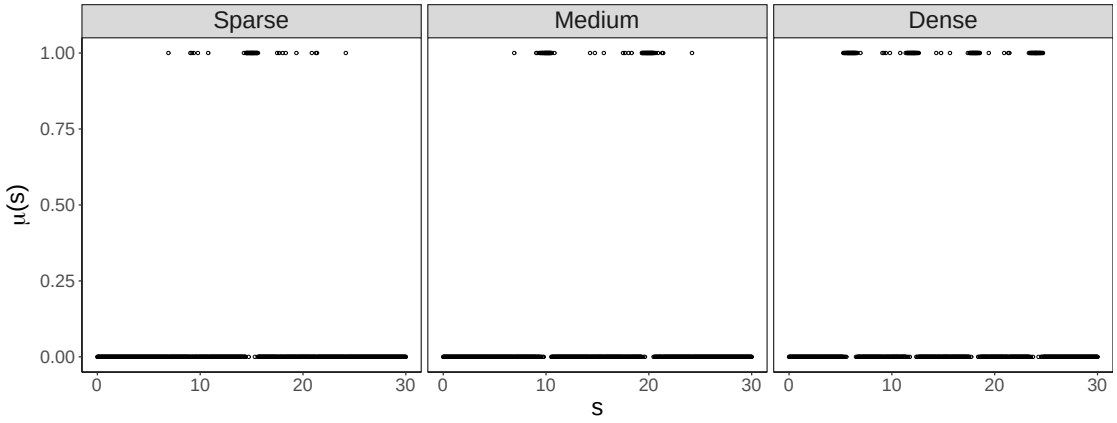
- Figure S.1 depicts the one-dimensional signal process $\mu(s)$ for Setups I–III as described in Section 5 of the main paper. The plots are depicted with magnitude $\gamma = 1$.
- Figure S.2 shows the two-dimensional signal process for three signal sparsity levels in Section S.I.3. The plots are depicted with magnitude $\gamma = 1$.
- Figure S.3 displays the spatial covariances of the noise process $\epsilon(s)$ versus the spatial distance for three dependency strengths as described in Sections 5.

S.I.2 Additional Simulation Studies for a One-Dimensional Domain

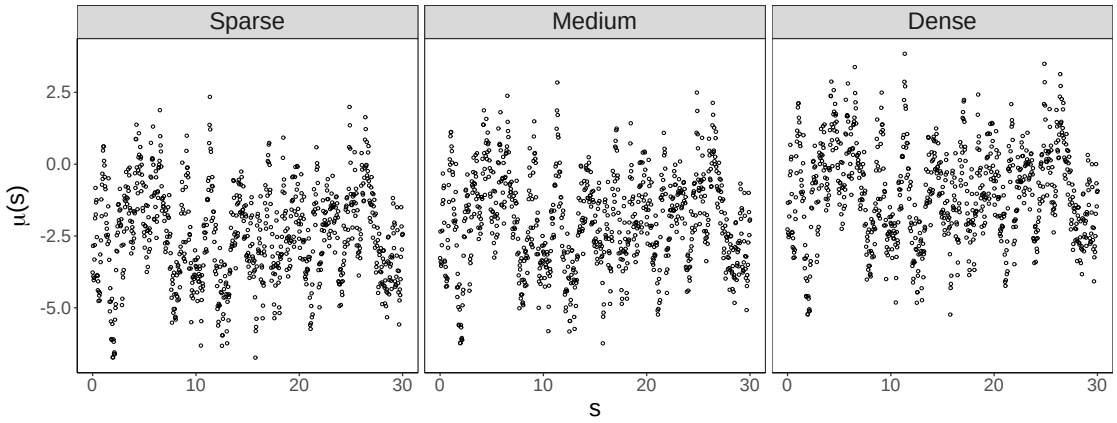
In this section, we demonstrate the performance of the 2d procedures under the Setup II (see Section 5 of the main paper) with $m = 900$ and under Setups I–III with $m = 2000$. The results for Setup II were displayed in Figure S.4 and generally similar to those in Setup I of



(a) Setup I.



(b) Setup II.



(c) Setup III.

Figure S.1: The one-dimensional signal process $\mu(s)$ ($s \in [0, 30]$) with $\gamma = 1$ in Section 5 of the main paper. The top to bottom panels correspond to Setups I–III of our simulation settings respectively.

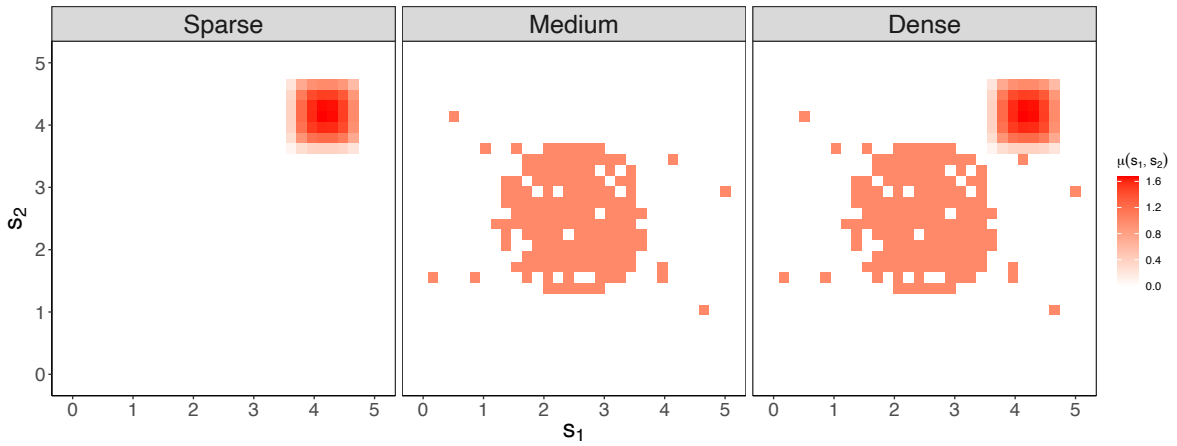


Figure S.2: The two-dimensional signal process $\mu(s_1, s_2)$ ($(s_1, s_2) \in [0, 5]^2$) with $\gamma = 1$ in Section S.I.3. The left to right panels correspond to sparse, medium, and dense signals of our simulation settings respectively.

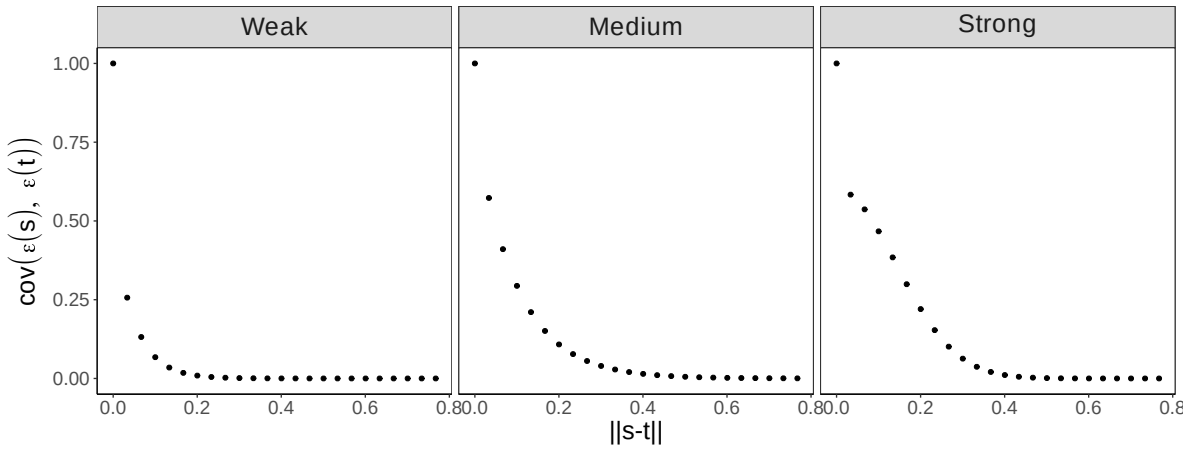


Figure S.3: The covariance structure of the noise process $\epsilon(s)$ used in our simulation studies. The x-axis is the distance between s and t, denoted as $\|s - t\|$, and y-axis is the covariance between $\epsilon(s)$ and $\epsilon(t)$, denoted as $\text{cov}\{\epsilon(s), \epsilon(t)\}$. The choices of (r, ρ_ϵ, k) for the weak, medium, and strong dependence strengths (from left to right) are: (1) $r = 0.5$, $\rho_\epsilon = 0.05$, $k = 1$ (exponential kernel), (2) $r = 0.8$, $\rho_\epsilon = 0.1$, $k = 1$ (exponential kernel), and (3) $r = 0.6$, $\rho_\epsilon = 0.2$, $k = 2$ (Gaussian kernel), respectively.

our main paper.

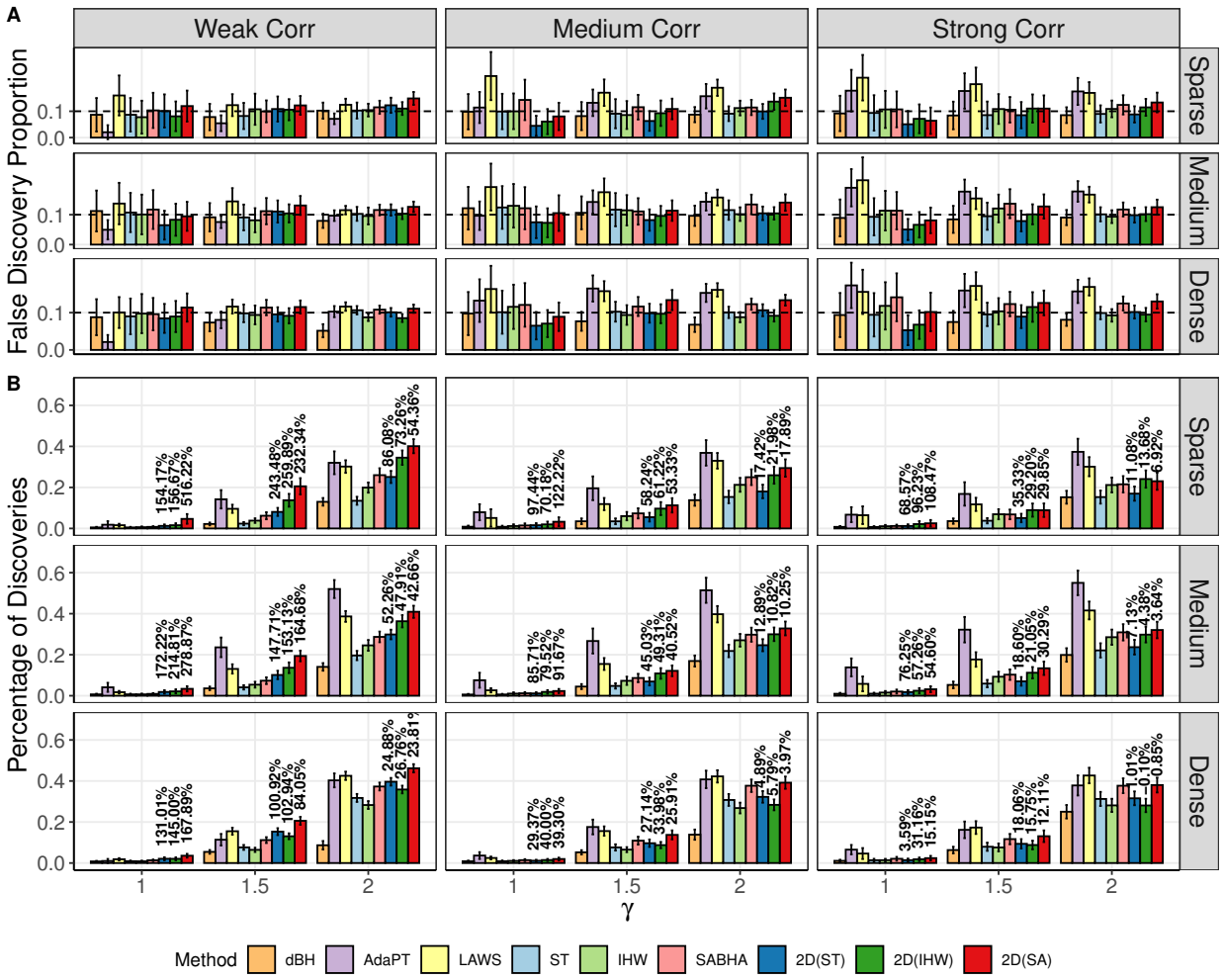


Figure S.4: The mean and (1.96 multiple of) the standard error of FDP (Panel A) and power (Panel B) under Setup II with $\gamma \in \{1, 1.5, 2\}$. The percentages on the top of bars represent the power improvement of 2d procedures compared to their 1d counterparts.

Then, we demonstrate the performance of the 2d procedures for one-dimensional domain with increased location sizes. The simulation setting is the same as Section 5 of the main paper but the process $X(s)$ was observed at $m = 2000$ locations evenly distributed over the domain $\mathcal{S} = [0, 60]$. The simulation results are reported in Figures S.5 and S.7. It can be seen that under Setups I–II, 2D (ST), 2D (IHW) and 2D (SA) satisfactorily controlled FDR

in all cases. Compared to the results of $m = 900$, the performance of LAWS and AdaPT with $m = 2000$ became better in terms of FDR control. However, they were still more likely to have FDR inflation in contrast with the proposed 2d procedures. Under Setup III, the 2d procedures were much more powerful than other methods.

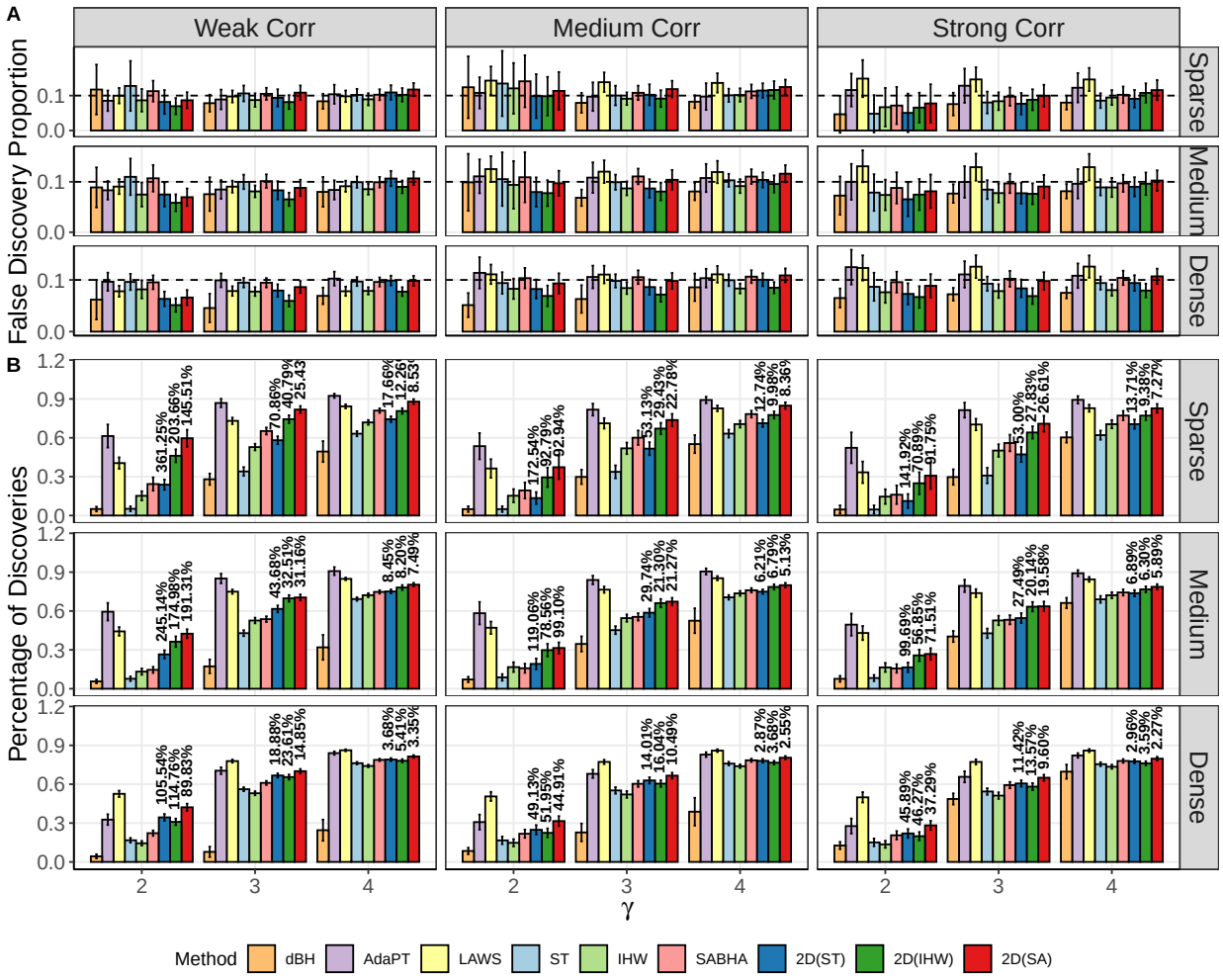


Figure S.5: The mean and (1.96 multiple of) the standard error of FDP (Panel A) and power (Panel B) under Setup I with $\gamma \in \{2, 3, 4\}$ and $m = 2000$. The percentages on the top of bars represent the power improvement of 2d procedures compared to their 1d counterparts.

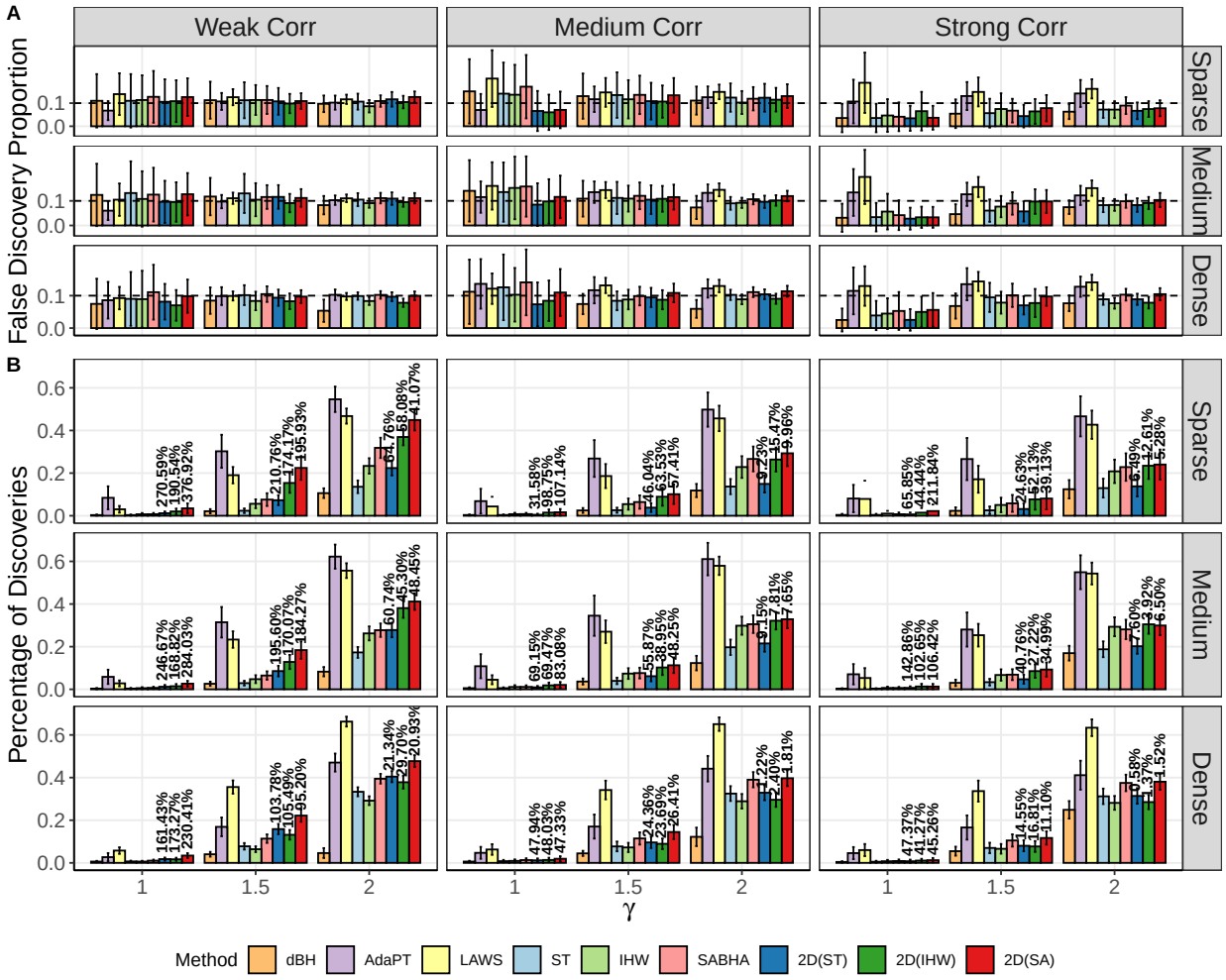


Figure S.6: The mean and (1.96 multiple of) the standard error of FDP (Panel A) and power (Panel B) under Setup II with $\gamma \in \{1, 1.5, 2\}$ and $m = 2000$. The percentages on the top of bars represent the power improvement of 2d procedures compared to their 1d counterparts.

S.I.3 Simulation Studies for a Two-Dimensional Domain with Unknown Covariance

In this section, we consider a spatial process $X(s) = \mu(s) + \epsilon(s)$ defined on the unit square $\mathcal{S} = [0, 5]^2$. We observe the process on a 30×30 lattice within the unit square. The noise $\epsilon(s)$ was generated from a mean-zero Gaussian process defined on $[0, 5]^2$ with the same

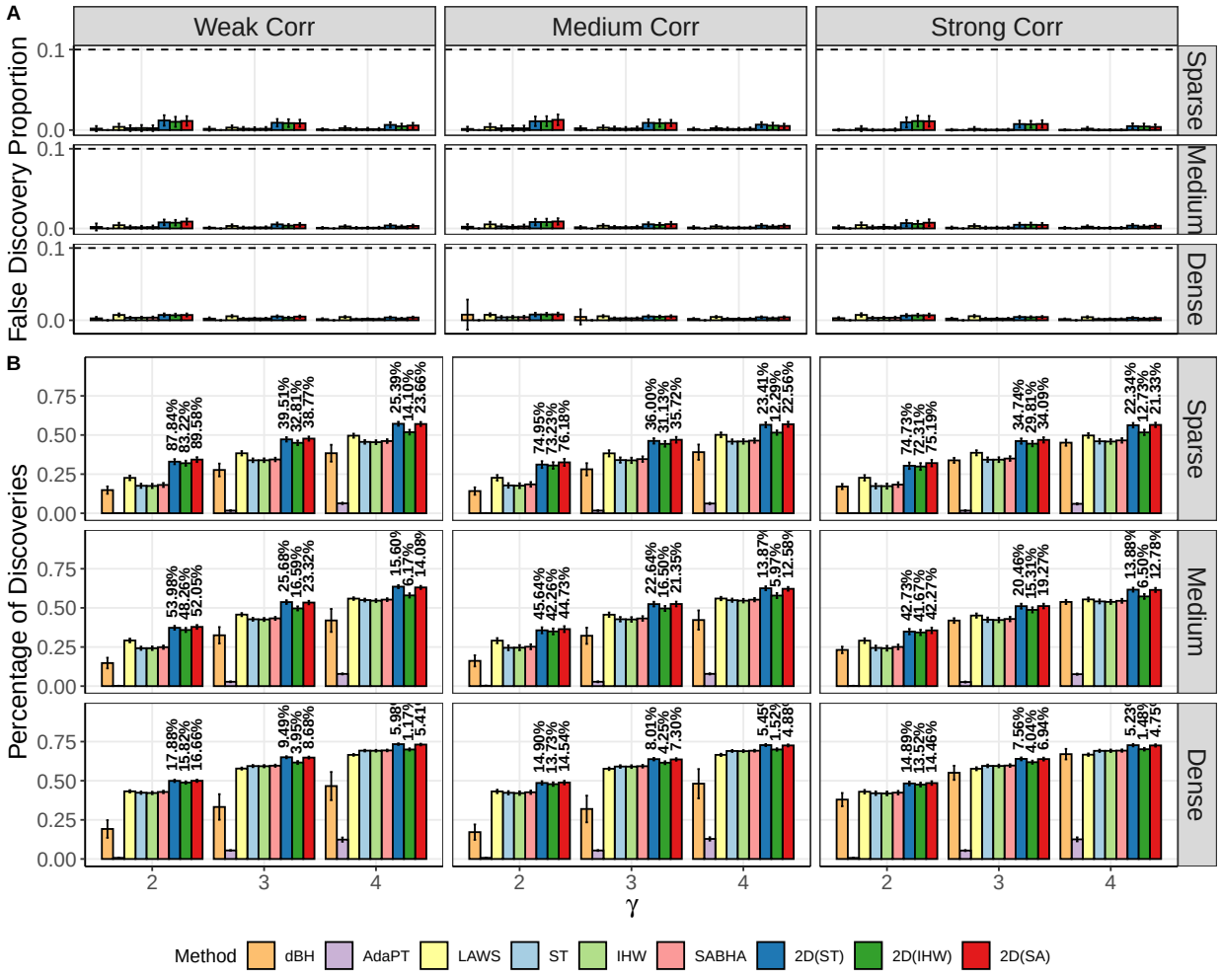


Figure S.7: The mean and (1.96 multiple of) the standard error of FDP (Panel A) and power (Panel B) under Setup III with $\gamma \in \{2, 3, 4\}$ and $m = 2000$. The percentages on the top of bars represent the power improvement of 2d procedures compared to their 1d counterparts.

covariance function as described in Section 5 of the main paper based on the distances between locations. Unlike Section 5 of the main paper, we now assume that the covariance structure is unknown, but three replications are available at each location. Given multiple realizations at each location, we could employ the maximum likelihood estimation with a pre-specified family of covariance functions to estimate the spatial covariance structure. The details are provided in Section S.VI of the supplement. We consider two structures for the signal process $\mu(s)$ with $s = (s_1, s_e)$.

- **Setup IV (smooth signal)**: $\mu_{\text{sm}}(s) = \gamma\mu_0(s)$, where $\mu_0(s) = f_1(s_1)f_2(s_2)$ and $\{f_i\}_{i=1,2}$ being generated using B-spline functions as in Setup I.
- **Setup V (clustered signal)**: $\mu_{\text{cl}}(s) = \gamma\mu_0(s)$, where $\mu_0(s) \sim \text{Bernoulli}(\bar{\pi}_0(s))$ with $\bar{\pi}_0(s) = 0.9$, if $(s_1 - 1/2)^2 + (s_2 - 1/2)^2 \leq (1/4)^2$; and 0.01 otherwise.

Three *signal sparsity levels* based on Setups IV and V were investigated: (1) sparse signal: $\mu_0(s) = \mu_{\text{cl}}(s)$; (2) medium signal: $\mu_0(s) = \mu_{\text{sm}}(s)$; and (3) dense signal: $\mu_0(s) = \mu_{\text{sm}}(s) + \mu_{\text{cl}}(s)$. The realized signal processes associated with these sparsity levels are shown in Figure S.2 of the supplement, and their percentages of the non-null locations are 5%, 17%, and 23%, respectively. We set the magnitude γ at values $\{0.5, 1, 1.5\}$ in both setups.

We report the numerical results of the FDP and power of competing methods based on 100 simulation runs in Figure S.8. Generally speaking, the 2d procedures had the best FDR and power trade-off. LAWS showed higher power at the expense of FDR inflation. AdaPT provided reliable FDR control in all cases but their power were dominated by 2D (ST), 2D (IHW), and 2D (SA) for the sparse signal and weak correlation structures. The power improvements from the 2d procedures were most significant when the correlation was weak.

It is also worth mentioning that in the case of strong correlation, the underlying covariance function was based on the Gaussian kernel while we estimated the covariance structure using the exponential kernel. The 2d procedures appeared robust to the misspecification subject to the parametric family of covariance functions.

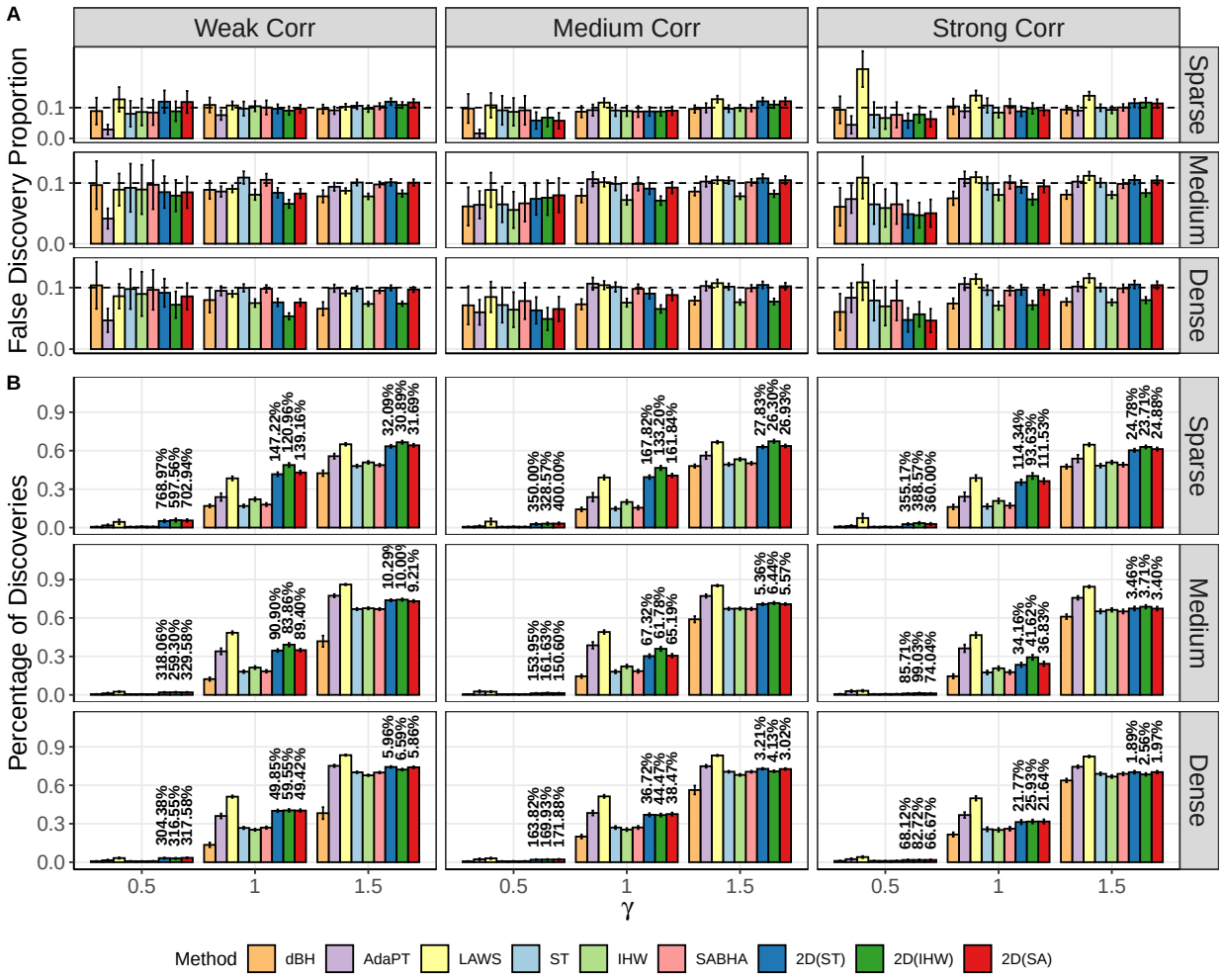


Figure S.8: The mean and (1.96 multiple of) the standard error of FDP (Panel A) and power (Panel B) on the two-dimensional domain with $\gamma \in \{0.5, 1, 1.5\}$. The percentages on the top of bars represent the power improvement of 2d procedures compared to their 1d counterparts.

S.I.4 Sensitivity to the Number of Nearest Neighbors

In this section, we conduct a simulation study to investigate the sensitivity of 2d-SMT to the number of observations in $\mathcal{N}(s)$. To be concrete, we let $\mathcal{N}(s)$ be the κ-nearest neighbors for each location s. We focus on the settings in Setups I–III and let $\kappa \in \{1, 2, 3, 4, 7, 10, 13, 16\}$. The false discovery proportions (FDPs) and powers of BH, ST, SABHA, and IHW (with the global null proportion estimate) as well as their corresponding 2d versions are summarized in Figures S.9–S.14. The difference between the 2d procedures with zero neighbor and their 1d counterparts is that the 2d procedures add a small offset q (i.e., the FDR level) to the estimate of the number of false rejections (see Section 3 of the main paper).

Figures S.9, S.11, and S.13 show that the FDP increased when incorporating the first neighbor, then gradually decreased and maintained stability afterwards as more neighbors were included for the three setups. As seen from Figures S.10 and S.12, including more neighbors improved the detection power for all the 2d procedures, which was significantly higher than that of the corresponding 1d counterpart under all setups of Setups I and II. In Figure S.14, although the detection power of the 2d procedure remained larger than that of the corresponding 1d counterpart, the different 2d procedures behaved differently for varying κ under Setup III. In particular, the powers of 2D (IHW), 2D (ST), and 2D (BH) gradually decreased when κ was larger than 7. For 2D (SABHA), when the signal was sparse, its detection power first decreased and then increased when more neighbors were included. For the medium and dense signal cases, its behavior was similar to the other 2d procedures. These findings empirically suggest that one may choose κ between 2 and 7 in 2d-SMT to

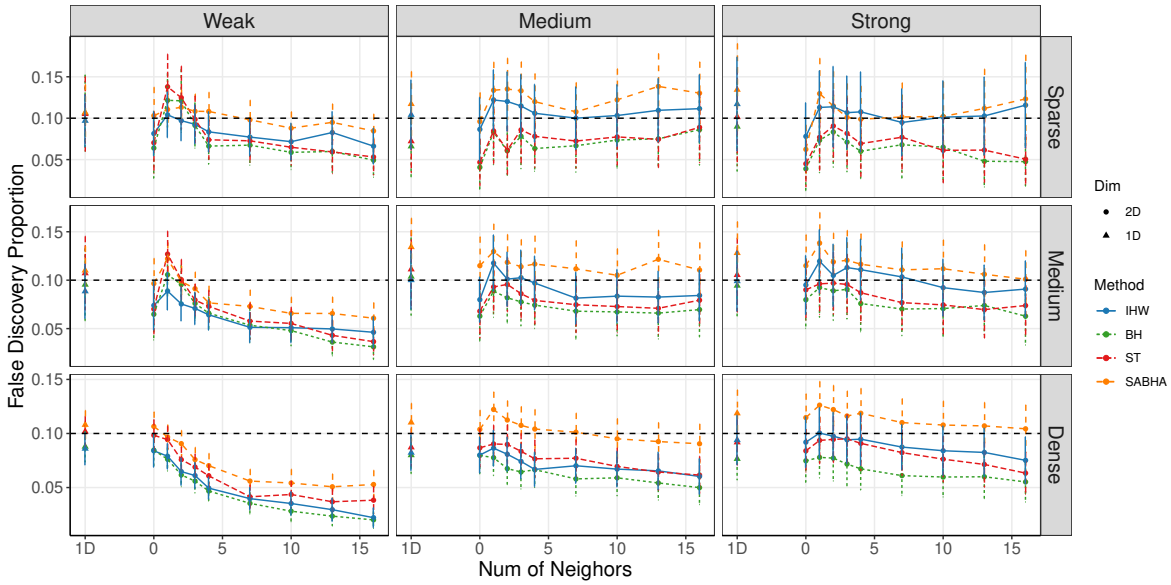


Figure S.9: The mean and (1.96 multiple of) the standard error of FDP under Setup I with $\gamma = 2$. Each color represents one particular 1d procedure (the circle at the leftmost column of each plot) and its 2d counterpart (the triangles).

control the FDP and improve the detection power.

S.I.5 Data-Adaptive Neighborhood

So far, each location has been assigned an equal number of neighbors. In this section, we present a simple strategy to determine the number of neighbors, and then assess its effectiveness through numerical simulations. The idea of this strategy is to adaptively enlarge the neighborhood size of locations that are more likely to be within a large cluster of spatial signals. We achieve this by continuously including neighbors for location s from its nearest neighbors (as candidates) until a primary statistic below zero is found from a candidate, i.e., we use the sign of the primary statistic as an initial criterion to distinguish the nulls and the alternatives in candidates. To be specific, let $\mathcal{N}_\kappa(s)$ denote the set of κ-nearest neighbors of

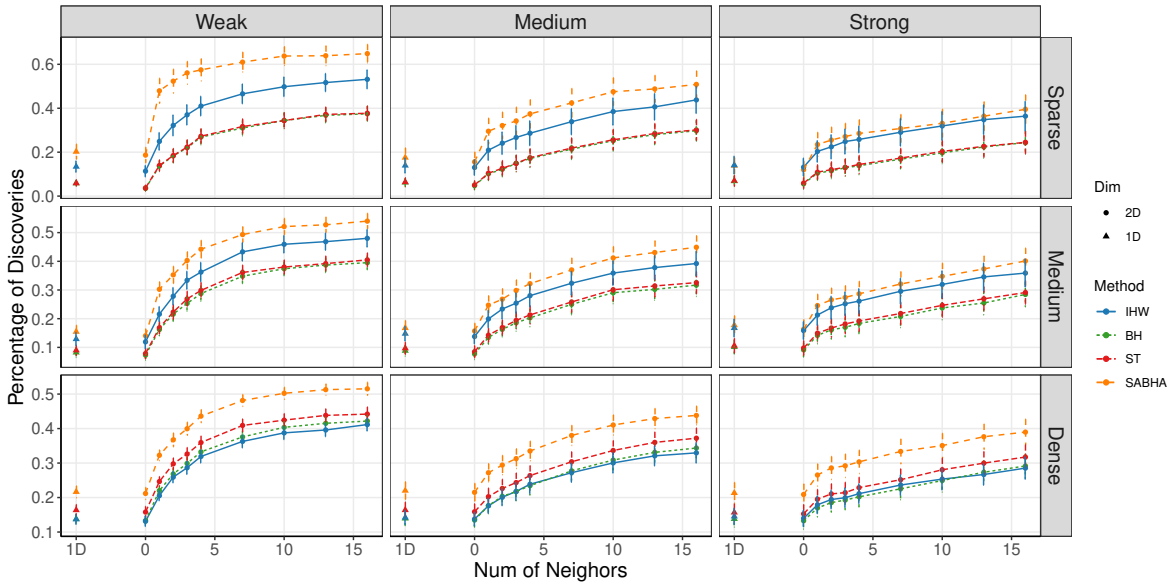


Figure S.10: The mean and (1.96 multiple of) the standard error of power under Setup I with $\gamma = 2$. Each color represents one particular 1d procedure (the circle at the leftmost column of each plot) and its 2d counterpart (the triangles).

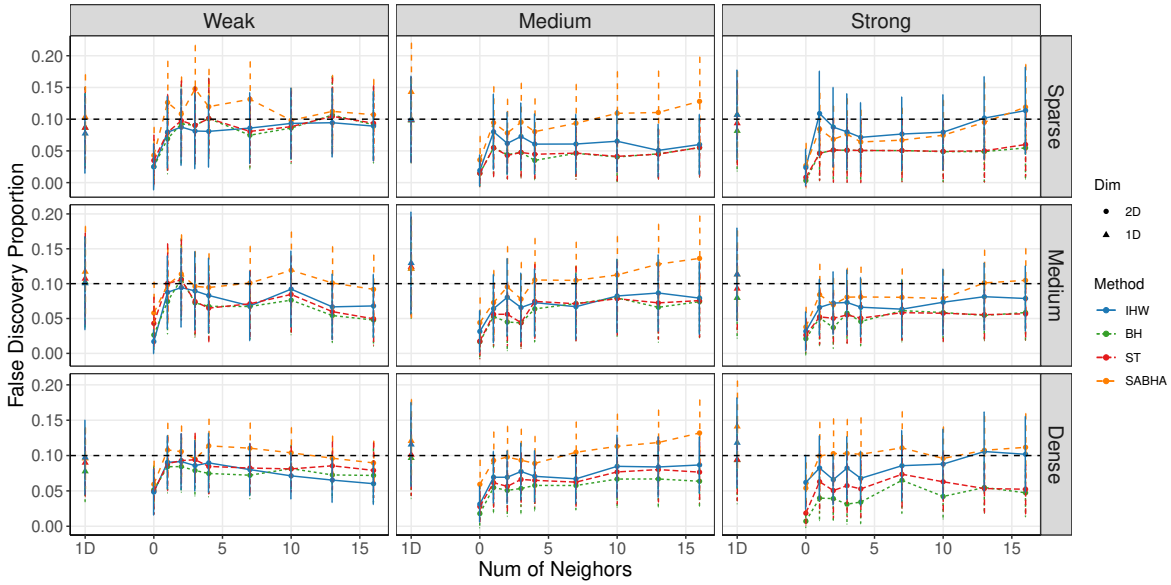


Figure S.11: The mean and (1.96 multiple of) the standard error of FDP under Setup II with $\gamma = 1$. Each color represents one particular 1d procedure (the circle at the leftmost column of each plot) and its 2d counterpart (the triangles).

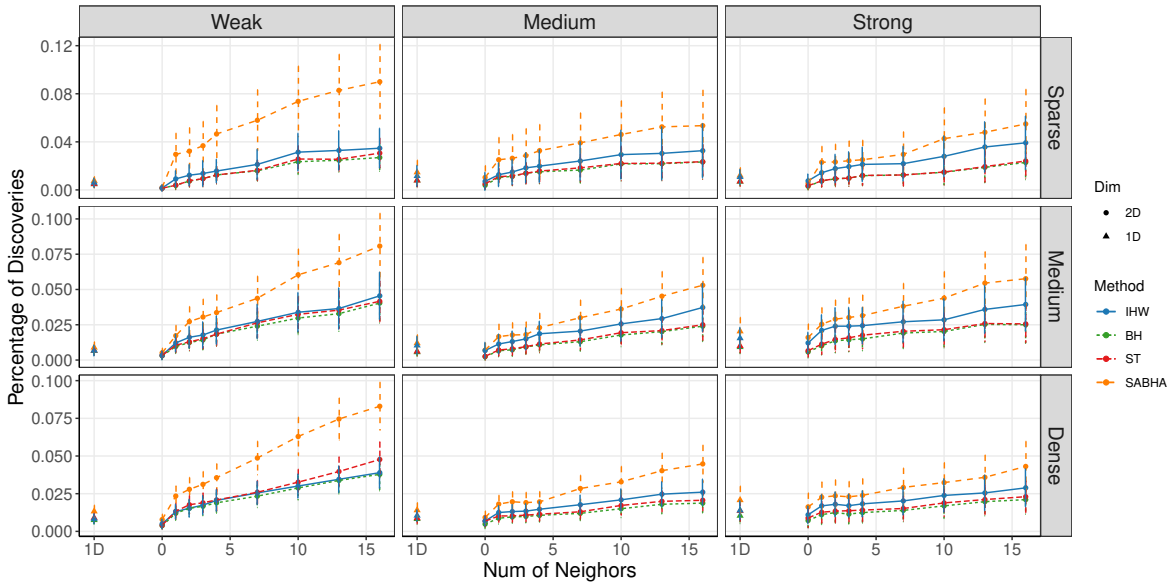


Figure S.12: The mean and (1.96 multiple of) the standard error of power under Setup II with $\gamma = 1$. Each color represents one particular 1d procedure (the circle at the leftmost column of each plot) and its 2d counterpart (the triangles).

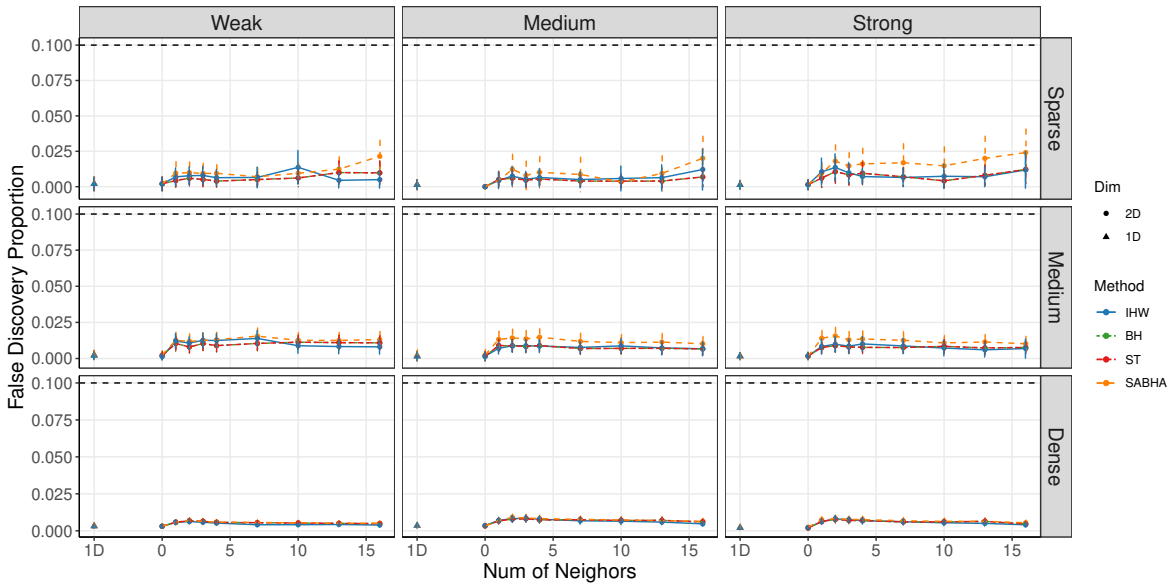


Figure S.13: The mean and (1.96 multiple of) the standard error of FDP under Setup III with $\gamma = 2$. Each color represents one particular 1d procedure (the circle at the leftmost column of each plot) and its 2d counterpart (the triangles).

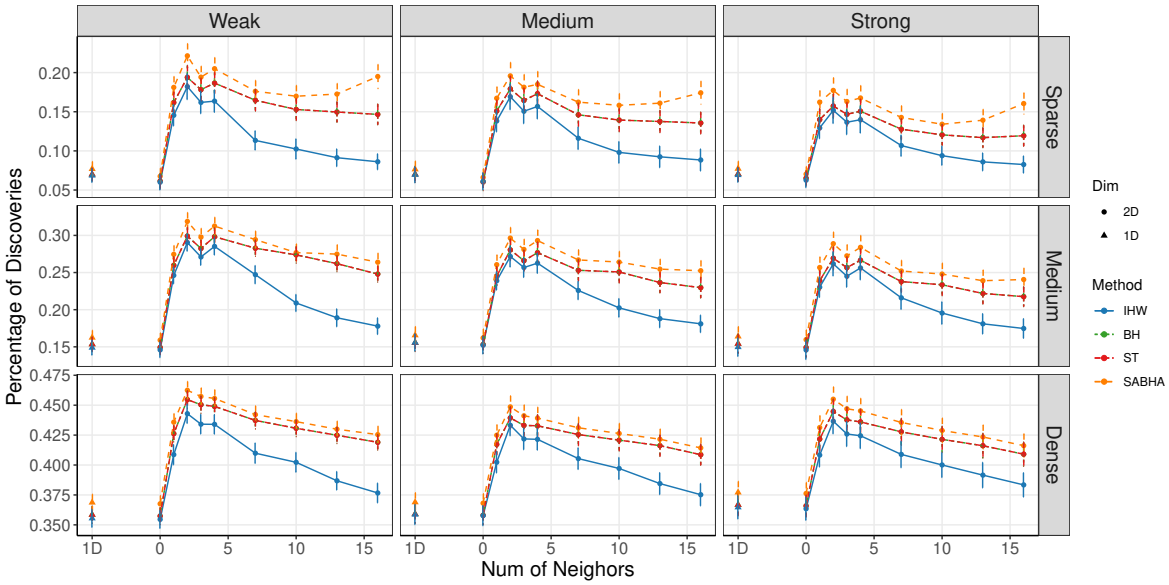


Figure S.14: The mean and (1.96 multiple of) the standard error of power under Setup III with $\gamma = 2$. Each color represents one particular 1d procedure (the circle at the leftmost column of each plot) and its 2d counterpart (the triangles).

location s. The number of neighbors for location s is determined as follows

$$\kappa_s = \begin{cases} 4, & T_2(v) < 0 \text{ for some } v \in \mathcal{N}_2(s) \\ \max\left\{2 \le \kappa \le 7 : T_2(v) > 0 \text{ for all } v \in \mathcal{N}_\kappa(s)\right\}, & \text{otherwise,} \end{cases} \tag{S.1}$$

where a value of four is used in Section 5 of the main paper and the range of κ is suggested in Section S.I.4. In this experiment, we considered two variants to implement the above strategy: 1) utilized a separate dataset specifically for the neighbor selection process, and 2) used the same dataset for both the neighbor selection process and the subsequent inference.

We generated the synthetic datasets according to Setups I–III in Section 5 of the main paper, equipped with the medium signal and weak correlation. Four different methods were compared: the standard ST and its three 2d variants. The first variant, $2\text{D}_{\text{Fix}}(\text{ST})$, assigned 4-nearest neighbors for all locations. The second variant, $2\text{D}_{\text{Ada},1}(\text{ST})$, and the third variant,

$2D_{Ada,2}(ST)$, determined the number of neighbors according to (S.1) using a separate dataset and the inference dataset, respectively.

We report the numerical results of the empirical FDP and power of the competing methods based on 100 simulation runs in Figure S.15. All 2d procedures exhibited an enhanced power relative to the standard ST procedure. Our neighbor selection strategy, as detailed in (S.1), boosted the power in Setups I and II. Notably, the $2D_{Ada,2}(ST)$ approach encountered difficulties in controlling FDR, particularly in Setup I with $\gamma = 4$. Its performance highlights the necessity for caution in data reuse, especially when no additional dataset for neighbor selection is available.

Figure S.16 displays the average number of neighbors for all locations based on 100 simulation runs. The strategy (S.1) effectively allocated a larger set of neighbors to locations under the alternative in Setups I and II, thereby improving power through adaptive neighbor selection. As illustrated in Figure S.1(c), given that locations under the alternative were dispersed into several small clusters in Setup III, determining neighbors became more complex. In conclusion, while adaptive selection of neighbors can improve detection efficiency, further investigation is required to ensure both the safety and efficacy of the neighbor selection strategy.

S.I.6 Combining Covariate and Spatial Information

Section 2.7 of the main paper introduces the concept of weighted thresholds, highlighting the potential to further exploit spatial information. When both covariate and spatial information are available, we can utilize the covariate information through weights and spatial informa-

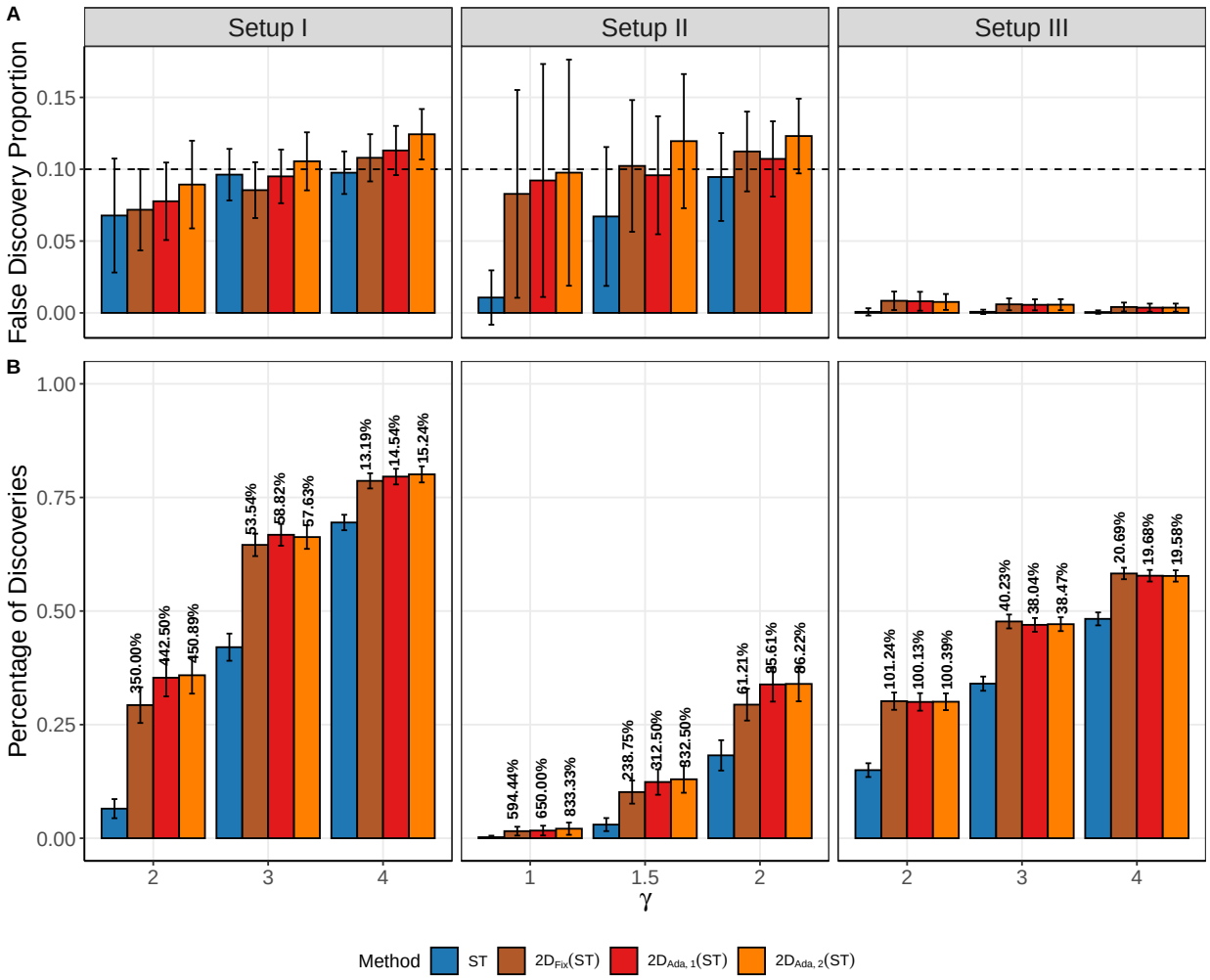


Figure S.15: The mean and (1.96 multiple of) the standard error of FDP (Panel A) and power (Panel B) under Setup I with $\gamma \in \{2, 3, 4\}$, Setup II with $\gamma \in \{1, 1.5, 2\}$, and Setup III with $\gamma \in \{2, 3, 4\}$. The percentages on the top of bars represent the power improvement of $2\text{D}_{\text{Fix}}(\text{ST})$, $2\text{D}_{\text{Ada},1}(\text{ST})$ and $2\text{D}_{\text{Ada},2}(\text{ST})$ compared to ST.

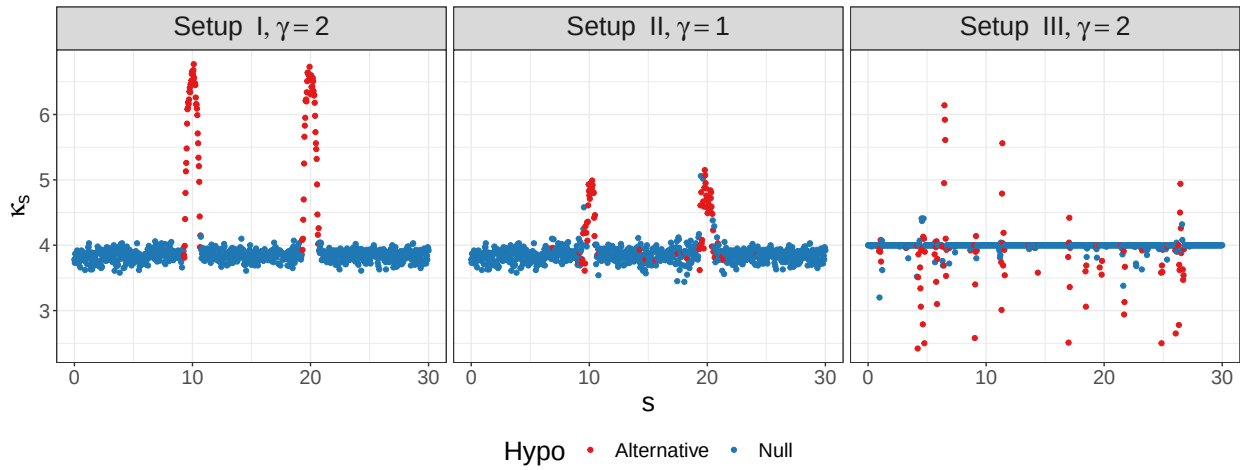


Figure S.16: The average number of neighbors for 900 locations based on 100 simulation runs.

tion through auxiliary statistics. In this section, we conduct simulations to investigate the ability of the 2d-SMT procedure to simultaneously use spatial and covariate information.

In a variant of Setup I, as illustrated in Figure S.17, we partitioned the locations into four groups: the first two groups were all under the null, whereas the third and fourth groups comprised approximately 1/2 and 2/3 of locations under the alternative, respectively. In this setup, the signal pattern exhibits both grouping designs and spatial trends. We considered the three degrees of spatial dependence as shown in Figure S.3. We evaluated four types of weights, deriving eight methods by considering both the 1d and 2d approaches. The BH and 2D (BH) used uniform weights $w(s) \equiv 1$. The FDP estimator of 2D (ST) is detailed in Section 2.6 of the main paper, while the FDP estimator of ST sets the threshold for primary statistics t_1 to be $-\infty$. To incorporate group information, we utilized the null proportion estimators as described in Section 2.5 for each group, denoted as groupwise null proportion estimates $\widehat{\pi}_0(s)$. Based on these estimates, SA_{Grp} and 2D (SA_{Grp}) assigned weights inversely proportional to $\widehat{\pi}_0(s)$, and LAWS_{Grp} and 2D (LAWS_{Grp}) assigned weights proportional to

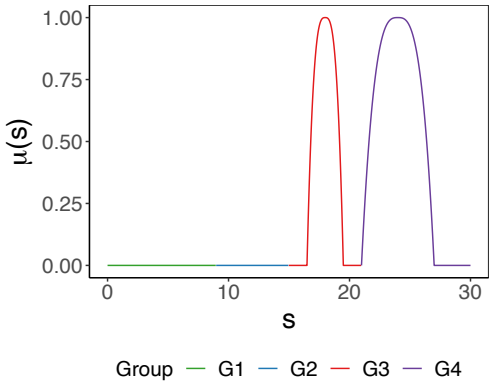


Figure S.17: The one-dimensional signal process $\mu(s)$ ($s \in [0, 30]$) with $\gamma = 1$ that possesses both spatial and group information.

$\{1 - \widehat{\pi}_0(s)\}/\widehat{\pi}_0(s)$.

We report the numerical results of the empirical FDP and power of the eight methods based on 100 simulation runs in Figure S.18. All procedures controlled the FDR fairly well. The performance of the four 1d procedures indicated that incorporating null proportion estimates can significantly enhance detection power, and including groupwise estimates can provide further improvements. Furthermore, LAWS_{Grp}, which employs a weighting function that uses group information more aggressively, detected more signals. The 2d procedures outperformed their 1d counterparts by incorporating spatial information. In general, 2D (LAWS_{Grp}) achieved the highest power while controlling the FDR under the nominal level.

S.I.7 Additional Results for the Ozone Data Analysis

Figure S.19 uses distinct shapes and colors to illustrate the partition of the Contiguous United States into nine different regions as discussed in Section 6 of the main paper. The region to which each location belongs is treated as a categorical variable and is used as the covariate in IHW and the group indicator in SABHA.

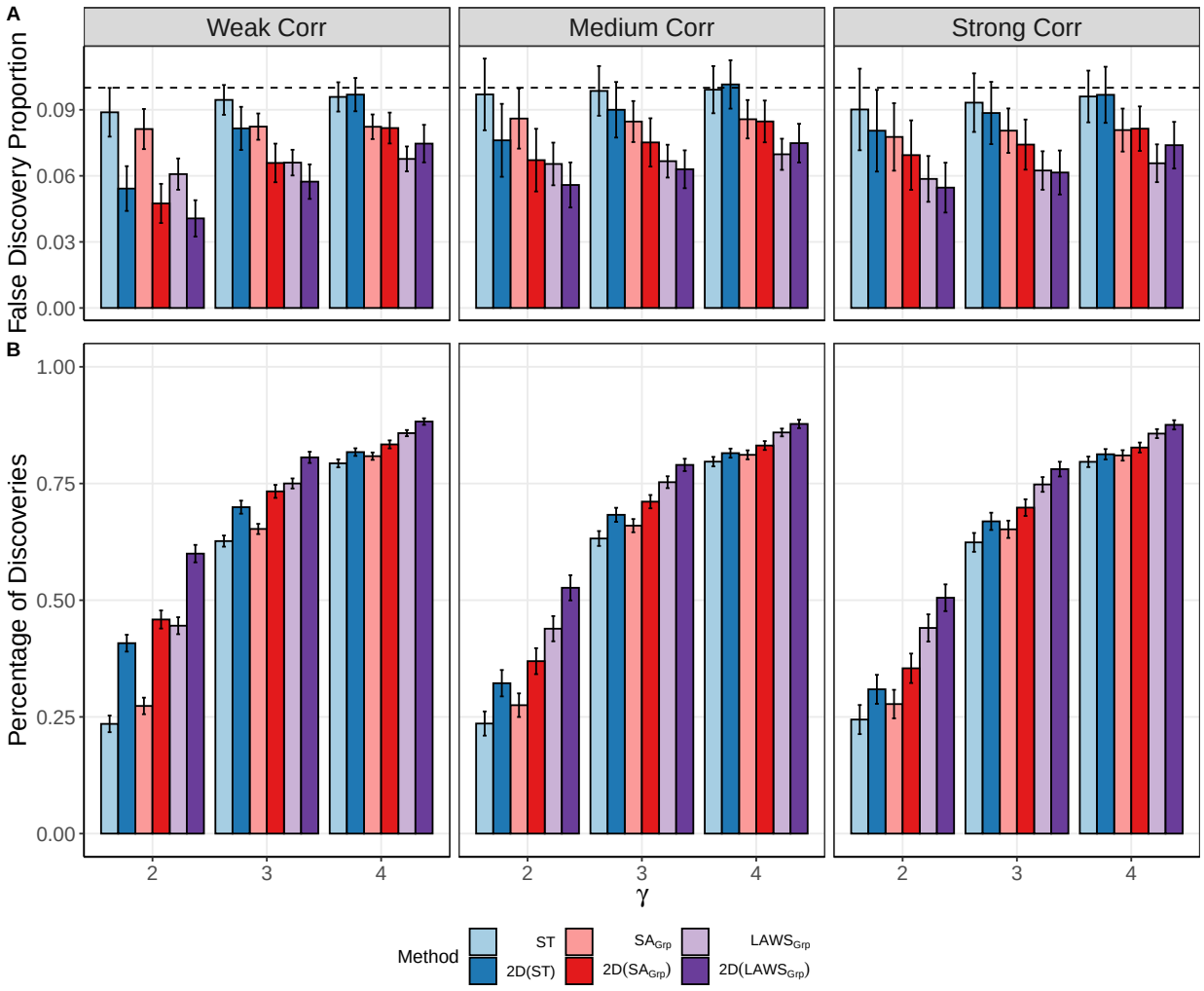


Figure S.18: The mean and (1.96 multiple of) the standard error of FDP (Panel A) and power (Panel B) under the setup that possesses both spatial pattern and group information with $\gamma \in \{2, 3, 4\}$ and $m = 900$. The percentages on the top of bars represent the power improvement of 2d procedures compared to their 1d counterparts.

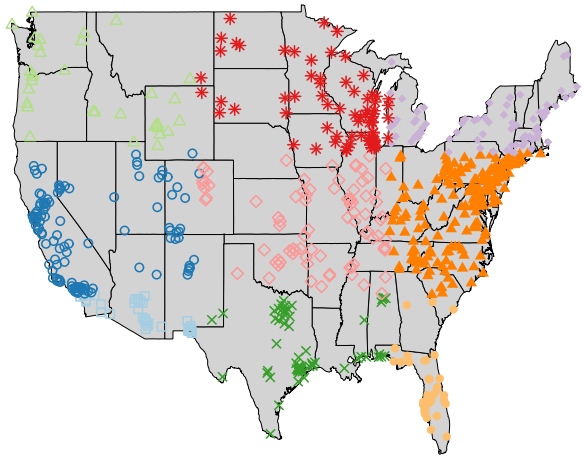


Figure S.19: The Contiguous United States is divided into nine different regions based on latitude and longitude, i.e., each station belongs to a region coded between one and nine. The stations in the same region are depicted using the same shape and color.

Table S.1 shows the locations with the most significant decline in CO or NO_2 concentration levels and whether the competing methods detected them in the study of the decline of the ozone level for various β_0. Table S.2 presents the average standardized slopes of CO and NO_2 at the locations detected by the 2d procedure or the 1d procedure, but not both. It shows that the NO_2 concentration level at the locations detected by the 2d procedures decreased faster than their 1d counterparts on average. For CO, when the null hypothesis was $\beta_0 = 0.2$ or $\beta_0 = 0.5$, the average standardized slope of its concentration level at the locations detected by the 2d procedures remained smaller than their 1d counterparts. The only exception was when $\beta_0 = 3$, the locations detected by the 1d procedure (SABHA) had a smaller average standardized slope of CO level than those detected by 2D (SA).

Motivated by Sun et al. (2015), we conducted further simulations to study the sensitivity of 2d-SMT to covariance misspecification. We considered two approaches to determine the covariance for simulating the ozone level data. The first approach fitted the residuals with

Table S.1: The locations with the most significant decline in the CO or NO_2 levels. Reported are the locations where at least one of the 2d procedures gives different decisions to their corresponding 1d counterparts. The column "Stand. Coeff." represents the value of the standardized coefficient.

Ozone precursor	β_0	Lat.	Lon.	Stand. Coeff.	Methodology					
					ST	2D (ST)	IHW	2D (IHW)	SABHA	2D (SA)
CO	0.1	40.63	-75.34	-0.56	✗	✓	✗	✗	✗	✗
	0.2	42.14	-87.80	-0.94	✗	✓	✗	✗	✗	✗
	0.3	41.00	-80.35	-2.19	✓	✓	✓	✓	✗	✓
	0.4	36.20	-95.98	-0.85	✓	✓	✓	✓	✗	✓
	0.5	41.53	-90.59	-3.46	✗	✗	✗	✗	✗	✓
NO2	0.1	40.63	-75.34	-3.49	✗	✓	✗	✗	✗	✗
	0.2	42.14	-87.80	-4.60	✗	✓	✗	✗	✗	✗
	0.3	32.87	-97.91	-43.10	✓	✓	✓	✓	✗	✓
	0.4	35.41	-94.52	-7.27	✗	✗	✗	✗	✗	✓
	0.5	41.53	-90.59	-9.16	✗	✗	✗	✗	✗	✓

Table S.2: The average standardized slopes of CO and NO_2 at the locations where at least one of the 2d procedures gives different decisions to their corresponding 1d counterparts. Each letter refers to the specific location depicted in Figure 4 of the main paper.

Ozone precursor	Methodology	β_0					
		0.2		0.3		0.5	
		Loc.	Avr.	Loc.	Avr.	Loc.	Avr.
CO	ST	—	—	—	—	—	—
	2D (ST)	d,e	-0.48	—	—	—	—
	SABHA	—	—	e, j	0.11	b	-0.96
	2D (SA)	—	—	h, i	0.56	c, e, f, h	-1.40
NO_2	ST	c	-1.30	f	-1.30	—	—
	2D (ST)	a, e, f	-2.73	a, c	-6.44	—	—
	SABHA	—	—	e, f, j	-2.81	b, d	-3.63
	2D (SA)	—	—	b, d, g	-17.17	a, c, f, g, h, i	-4.68

a Gaussian kernel, and the second utilized the empirical covariance matrix to accommodate non-stationarity. We then generated the simulated data accordingly to (6.1) of the main paper with modified covariance. The subsequent steps followed those of the ozone simulations in Section 6, which used the exponential kernel. Tables S.3 and S.4 demonstrate that the 2d procedures still achieved equal or higher power compared to their 1d counterparts while controlling FDR under 10%. To summarize, our procedure is robust to model misspecification, which confirms the reasonableness of modeling ozone data with (6.1) of the main paper.

Table S.3: Mean and standard deviation of FDPs and percentage of true discoveries (PTDs) for simulated ozone data using the Gaussian kernel. The results are based on 100 simulation runs.

Criterion	β_0	ST	IHW	SABHA	2D (ST)	2D (IHW)	2D (SA)
FDP	0.5	0.024(0.027)	0.025(0.029)	0.042(0.030)	0.025(0.028)	0.025(0.029)	0.054(0.036)
	0.4	0.022(0.018)	0.020(0.017)	0.026(0.017)	0.023(0.019)	0.020(0.017)	0.033(0.018)
	0.3	0.017(0.014)	0.013(0.013)	0.014(0.012)	0.017(0.015)	0.013(0.013)	0.015(0.012)
	0.2	0.017(0.013)	0.010(0.009)	0.012(0.010)	0.017(0.013)	0.010(0.009)	0.012(0.010)
	0.1	0.015(0.011)	0.008(0.008)	0.009(0.008)	0.015(0.012)	0.008(0.008)	0.009(0.008)
PTD	0.5	0.277(0.097)	0.280(0.095)	0.412(0.115)	0.281(0.099)	0.280(0.095)	**0.474(0.114)**
	0.4	0.401(0.115)	0.391(0.106)	0.481(0.079)	0.409(0.117)	0.391(0.106)	**0.517(0.081)**
	0.3	0.547(0.112)	0.508(0.099)	0.511(0.097)	**0.554(0.113)**	0.508(0.099)	0.521(0.096)
	0.2	0.695(0.088)	0.628(0.079)	0.633(0.081)	**0.702(0.088)**	0.628(0.079)	0.640(0.081)
	0.1	0.799(0.060)	0.716(0.061)	0.725(0.060)	**0.803(0.060)**	0.716(0.061)	0.729(0.060)

Table S.4: Mean and standard deviation of FDPs and percentage of true discoveries (PTDs) for the simulated ozone data using the empirical covariance matrix. The results are based on 100 simulation runs.

Criterion	β_0	ST	IHW	SABHA	2D (ST)	2D (IHW)	2D (SA)
FDP	0.5	0.018(0.036)	0.018(0.035)	0.027(0.040)	0.019(0.038)	0.018(0.036)	0.036(0.044)
	0.4	0.017(0.034)	0.014(0.027)	0.018(0.029)	0.018(0.034)	0.014(0.027)	0.020(0.030)
	0.3	0.016(0.025)	0.011(0.019)	0.012(0.020)	0.017(0.026)	0.011(0.019)	0.013(0.020)
	0.2	0.016(0.021)	0.009(0.015)	0.010(0.015)	0.016(0.022)	0.009(0.015)	0.010(0.016)
	0.1	0.016(0.021)	0.008(0.012)	0.008(0.012)	0.016(0.021)	0.008(0.012)	0.008(0.013)
PTD	0.5	0.258(0.189)	0.264(0.182)	0.369(0.207)	0.261(0.192)	0.265(0.183)	**0.416(0.216)**
	0.4	0.375(0.221)	0.363(0.205)	0.412(0.181)	0.380(0.223)	0.362(0.206)	**0.438(0.176)**
	0.3	0.500(0.229)	0.465(0.207)	0.480(0.190)	**0.506(0.229)**	0.464(0.207)	0.491(0.187)
	0.2	0.645(0.206)	0.582(0.189)	0.586(0.187)	**0.650(0.205)**	0.582(0.190)	0.592(0.186)
	0.1	0.761(0.152)	0.677(0.146)	0.682(0.149)	**0.765(0.151)**	0.677(0.147)	0.686(0.148)

S.II Proof of Theorem 1 of the Main Paper

We introduce two lemmas, which play key roles in the proof of Theorem 1 of the main paper.

Lemma S.1. Under Assumptions 1–9 of the main paper, for any t_1', $t_2' > 0$, we have

$$\sup_{|t_1|\leq t_1', |t_2|\leq t_2'} \left| \frac{1}{m} \sum_{s\in\mathcal{S}_m} \left[\int L\left\{t_1, t_2, x, \widehat{\rho}(s)\right\} d\widehat{G}_{\tilde{m}}(x) - \int L\left\{t_1, t_2, x, \rho(s)\right\} dG_0(x) \right] \right| = o_P(1), \tag{S.2}$$

as $m \to \infty$.

Lemma S.2. Under Assumptions 1–5, 8, and 9 of the main paper, for any t_1', $t_2' > 0$, we have

$$\sup_{|t_1|\leq t_1', |t_2|\leq t_2'} \left| \frac{1}{m_0} \widehat{V}_m\left(t_1, t_2\right) - K_0\left(t_1, t_2\right) \right| = o_P(1), \tag{S.3}$$

$$\sup_{|t_1|\leq t_1', |t_2|\leq t_2'} \left| \frac{1}{m_1} \widehat{S}_m\left(t_1, t_2\right) - K_1\left(t_1, t_2\right) \right| = o_P(1), \tag{S.4}$$

$$\left| \widehat{F}_m(\lambda) - F(\lambda) \right| = o_P(1), \tag{S.5}$$

as $m = m_0 + m_1 \to \infty$, where $\widehat{F}_m(\lambda) = m^{-1} \sum_{s\in\mathcal{S}_m} \mathbf{1}\{\widehat{T}_2(s) \leq \lambda\}$.

Lemma S.1 states that the estimated number of false discoveries in the proposed 2d-SMT procedure converges to the limiting process defined in Assumption 9 of the main paper. Lemma S.2 states the uniform convergence for the processes counting the numbers of false and true rejections and the empirical cumulative distribution function for $\widehat{T}_2(s)$. In the following derivations, we let C be a positive constant which can be different from line to line.

Proof of Theorem 1 of the main paper. We fix $t_1' = t_1^\star$ and $t_2' = t_2^\star$ as defined in Assumption

10 of the main paper. To prove Theorem 1, we first show

$$\sup_{|t_1|\leq t_1^\star,|t_2|\leq t_2^\star}\left|\widehat{\mathrm{FDP}}_{\lambda,\widetilde{\mathcal{S}}_m}(t_1,t_2)-\mathrm{FDP}_\lambda^\infty(t_1,t_2)\right|=o_P(1), \tag{S.6}$$

and

$$\sup_{|t_1|\leq t_1^\star,|t_2|\leq t_2^\star}\left|\frac{\widehat{V}_m(t_1,t_2)}{\widehat{V}_m(t_1,t_2)+\widehat{S}_m(t_1,t_2)}-\frac{\pi_0 K_0(t_1,t_2)}{K(t_1,t_2)}\right|=o_P(1). \tag{S.7}$$

To show (S.6), according to Lemma S.2 and Assumption 8 of the main paper, we have

$$\sup_{|t_1|\leq t_1^\star,|t_2|\leq t_2^\star}\left|\widehat{K}_m(t_1,t_2)-K(t_1,t_2)\right|=o_P(1),$$

where $\widehat{K}_m(t_1,t_2)=m^{-1}\{\widehat{V}_m(t_1,t_2)+\widehat{S}_m(t_1,t_2)\}$. For any $|t_1|\leq t_1^\star$, $|t_2|\leq t_2^\star$ and large enough m, we get

$$\left|\widehat{K}_m(t_1,t_2)-K(t_1,t_2)\right|\leq\frac{|K(t_1,t_2)|}{2},$$

which implies

$$\left|\widehat{K}_m(t_1,t_2)\right|\geq\frac{|K(t_1,t_2)|}{2}\geq\frac{K(t_1^\star,t_2^\star)}{2}>0 \tag{S.8}$$

because $\inf_{|t_1|\leq t_1^\star,|t_2|\leq t_2^\star}|K(t_1,t_2)|\geq K(t_1^\star,t_2^\star)>0$. For large enough m, it follows that

$$\begin{aligned}
&\widehat{\mathrm{FDP}}_{\lambda,\widetilde{\mathcal{S}}_m}(t_1,t_2)-\mathrm{FDP}_\lambda^\infty(t_1,t_2)\\
=&\frac{m^{-1}K(t_1,t_2)F_m(\lambda)\sum_{s\in\mathcal{S}_m}\int L\{t_1,t_2,x,\widehat{\rho}(s)\}\,d\widehat{G}_{\tilde{m}}(x)-\widehat{K}_m(t_1,t_2)F(\lambda)K_0(t_1,t_2)}{\Phi(\lambda)\widehat{K}_m(t_1,t_2)K(t_1,t_2)}\\
\leq&\frac{2m^{-1}K(t_1,t_2)F_m(\lambda)\sum_{s\in\mathcal{S}_m}\int L\{t_1,t_2,x,\widehat{\rho}(s)\}\,d\widehat{G}_{\tilde{m}}(x)-\widehat{K}_m(t_1,t_2)F(\lambda)K_0(t_1,t_2)}{\Phi(\lambda)K^2(t_1^\star,t_2^\star)},
\end{aligned}$$

where the last inequality holds by (S.8). Thus (S.6) follows from Lemma S.1 and Lemma S.2. Similarly, we can prove (S.7).

Next we use (S.6) and (S.7) to show $\limsup_{m\to\infty}\widetilde{\mathrm{FDR}}_m\leq q$. Due to Assumption 10 of the main paper, we have $\mathrm{FDP}_\lambda^\infty(t_1^\star,0)<q$ and $\mathrm{FDP}_\lambda^\infty(0,t_2^\star)<q$. Then, for large enough

m, we have

$$\widehat{\mathrm{FDP}}_{\lambda,\widetilde{\mathcal{S}}_m}\left(t_1^\star,0\right)<q \quad \text{and} \quad \widehat{\mathrm{FDP}}_{\lambda,\widetilde{\mathcal{S}}_m}\left(0,t_2^\star\right)<q,$$

which implies that $(\widetilde{t}_1^\star,\widetilde{t}_2^\star)$ satisfies $|\widetilde{t}_1^\star|\le t_1^\star$ and $|\widetilde{t}_2^\star|\le t_2^\star$. Thus, we get

$$\begin{aligned}&\widehat{\mathrm{FDP}}_{\lambda,\widetilde{\mathcal{S}}_m}\left(\widetilde{t}_1^\star,\widetilde{t}_2^\star\right)-\frac{\widehat{V}_m\left(\widetilde{t}_1^\star,\widetilde{t}_2^\star\right)}{\widehat{V}_m\left(\widetilde{t}_1^\star,\widetilde{t}_2^\star\right)+\widehat{S}_m\left(\widetilde{t}_1^\star,\widetilde{t}_2^\star\right)}\\&\quad\ge\inf_{|t_1|\le t_1^\star,|t_2|\le t_2^\star}\left\{\widehat{\mathrm{FDP}}_{\lambda,\widetilde{\mathcal{S}}_m}\left(t_1,t_2\right)-\frac{\widehat{V}_m\left(t_1,t_2\right)}{\widehat{V}_m\left(t_1,t_2\right)+\widehat{S}_m\left(t_1,t_2\right)}\right\}.\end{aligned}\tag{S.9}$$

Rearranging the second formula, we obtain

$$\begin{aligned}&\inf_{|t_1|\le t_1^\star,|t_2|\le t_2^\star}\left\{\widehat{\mathrm{FDP}}_{\lambda,\widetilde{\mathcal{S}}_m}\left(t_1,t_2\right)-\frac{\widehat{V}_m\left(t_1,t_2\right)}{\widehat{V}_m\left(t_1,t_2\right)+\widehat{S}_m\left(t_1,t_2\right)}\right\}\\&\quad=\inf_{|t_1|\le t_1^\star,|t_2|\le t_2^\star}\left\{\widehat{\mathrm{FDP}}_{\lambda,\widetilde{\mathcal{S}}_m}\left(t_1,t_2\right)-\mathrm{FDP}_\lambda^\infty\left(t_1,t_2\right)+\frac{\pi_0K_0\left(t_1,t_2\right)}{K\left(t_1,t_2\right)}\right.\\&\qquad\left.-\frac{\widehat{V}_m\left(t_1,t_2\right)}{\widehat{V}_m\left(t_1,t_2\right)+\widehat{S}_m\left(t_1,t_2\right)}+\mathrm{FDP}_\lambda^\infty\left(t_1,t_2\right)-\frac{\pi_0K_0\left(t_1,t_2\right)}{K\left(t_1,t_2\right)}\right\},\end{aligned}$$

which converges in probability to

$$\inf_{|t_1|\le t_1^\star,|t_2|\le t_2^\star}\left\{\mathrm{FDP}_\lambda^\infty\left(t_1,t_2\right)-\frac{\pi_0K_0\left(t_1,t_2\right)}{K\left(t_1,t_2\right)}\right\}$$

according to (S.6) and (S.7). As

$$-1\le\inf_{|t_1|\le t_1^\star,|t_2|\le t_2^\star}\left\{\widehat{\mathrm{FDP}}_{\lambda,\widetilde{\mathcal{S}}_m}\left(t_1,t_2\right)-\frac{\widehat{V}_m\left(t_1,t_2\right)}{\widehat{V}_m\left(t_1,t_2\right)+\widehat{S}_m\left(t_1,t_2\right)}\right\}\le\widehat{\mathrm{FDP}}_{\lambda,\widetilde{\mathcal{S}}_m}\left(\widetilde{t}_1^\star,\widetilde{t}_2^\star\right)\le q,$$

Lebesgues's dominated convergence theorem implies

$$\begin{aligned}&\liminf_{m\to\infty}E\left[\inf_{|t_1|\le t_1^\star,|t_2|\le t_2^\star}\left\{\widehat{\mathrm{FDP}}_{\lambda,\widetilde{\mathcal{S}}_m}\left(t_1,t_2\right)-\frac{\widehat{V}_m\left(t_1,t_2\right)}{\widehat{V}_m\left(t_1,t_2\right)+\widehat{S}_m\left(t_1,t_2\right)}\right\}\right]\\&=\inf_{|t_1|\le t_1^\star,|t_2|\le t_2^\star}\left\{\mathrm{FDP}_\lambda^\infty\left(t_1,t_2\right)-\frac{\pi_0K_0\left(t_1,t_2\right)}{K\left(t_1,t_2\right)}\right\}\ge0,\end{aligned}$$

where the last inequality stands due to the fact that $F(\lambda)/\Phi(\lambda)\ge\pi_0$ and $K_0\left(t_1,t_2\right)\le\lim_{m\to\infty}\sum_{s\in\mathcal{S}_m}\int L\{t_1,t_2,x,\rho(s)\}\,dG_0(x)/m$ implied by (4.4) in Assumption 9 of the main

paper. It then yields that

$$\begin{aligned}
&\liminf_{m\to\infty} E\left[\widehat{\mathrm{FDP}}_{\lambda,\widetilde{\mathcal{S}}_m}\left(\widetilde{t}_1^{\star},\widetilde{t}_2^{\star}\right)-\frac{\widehat{V}_m\left(\widetilde{t}_1^{\star},\widetilde{t}_2^{\star}\right)}{\widehat{V}_m\left(\widetilde{t}_1^{\star},\widetilde{t}_2^{\star}\right)+\widehat{S}_m\left(\widetilde{t}_1^{\star},\widetilde{t}_2^{\star}\right)}\right]\\
&\quad\geq \liminf_{m\to\infty} E\left[\inf_{|t_1|\leq t_1^{\star},|t_2|\leq t_2^{\star}}\left\{\widehat{\mathrm{FDP}}_{\lambda,\widetilde{\mathcal{S}}_m}\left(t_1,t_2\right)-\frac{\widehat{V}_m\left(t_1,t_2\right)}{\widehat{V}_m\left(t_1,t_2\right)+\widehat{S}_m\left(t_1,t_2\right)}\right\}\right]\\
&\quad=\inf_{|t_1|\leq t_1^{\star},|t_2|\leq t_2^{\star}}\left\{\mathrm{FDP}_{\lambda}^{\infty}\left(t_1,t_2\right)-\frac{\pi_0 K_0\left(t_1,t_2\right)}{K\left(t_1,t_2\right)}\right\}\geq 0,
\end{aligned}$$

where the first inequality holds because of (S.9) and the monotonicity of expectation. Finally, we obtain

$$\limsup_{m\to\infty}\widetilde{\mathrm{FDR}}_m\leq\liminf_{m\to\infty}E\left\{\widehat{\mathrm{FDP}}_{\lambda,\widetilde{\mathcal{S}}_m}\left(\widetilde{t}_1^{\star},\widetilde{t}_2^{\star}\right)\right\}\leq q,$$

which completes the proof. □

S.III Proofs of Lemmas S.1 and S.2

In this section, we prove Lemmas S.1 and S.2 with the help of some preliminary lemmas. In particular, Section S.III.1 presents the logic flow why the proposed estimator $\widehat{G}_{\tilde{m}}$ is able to estimate the limiting distribution of the signal process $\xi(s)$ by introducing multiple intermediate variants. Section S.III.2 states some preliminary lemmas for proving Lemma S.1. In Section S.III.3, we complete the proofs of Lemmas S.1 and S.2.

S.III.1 Notation and Proof Sketch of Lemma S.1

To better present the proof of Lemma S.1, this section briefly introduces some notation and intermediate quantities to connect the proposed general maximum likelihood estimator (GMLE) with the limiting distribution G_0. To begin with, we present a fact of GMLE that is useful in the following proof. There exists a discrete solution to (2.4) of the main paper

with no more than $|\widetilde{\mathcal{S}}|+1$ support points by the Carathéodory's theorem. Specifically, we can write the solution as

$$\widetilde{G}_{\widetilde{\mathcal{S}}}(u)=\sum_{i=1}^{\tilde{l}}\widetilde{\pi}_i\mathbf{1}\left\{\widetilde{v}_i\leq u\right\},\quad \sum_{i=1}^{\tilde{l}}\widetilde{\pi}_i=1,\quad \text{and}\quad \widetilde{\pi}_i>0, \tag{S.10}$$

where $\{\widetilde{v}_i\}_{i=1}^{\tilde{l}}$ is the set of support points and $\tilde{l}\leq|\widetilde{\mathcal{S}}|+1$. The support of $\widetilde{G}_{\widetilde{\mathcal{S}}}$ is within the range of $\{T_1(s):s\in\widetilde{\mathcal{S}}\}$ due to the monotonicity of $\phi(x-u)$ in $|x-u|$, where $\phi(x)$ is the density of standard normal distribution.

Now recall from (4.1) of the main paper that

$$T_1^*(s;r)=E\left[T_1(s)\mid\mathfrak{F}\left\{\cup_{v\in\mathcal{N}(s)}B(v;r)\right\}\right]/\zeta_r(s),$$

where $\zeta_r^2(s)=\text{var}\left(E\left[T_1(s)\mid\mathfrak{F}\left\{\cup_{v\in\mathcal{N}(s)}B(v;r)\right\}\right]\right)$. We call $\zeta_r(s)$ the normalization term since it ensures $\text{var}\{T_1^*(s;r)\}=1$. It can be directly shown that under Assumption 5 of the main paper, $\{T_1^*(s;r):s\in\widetilde{\mathcal{S}}_m\}$ are independent and normally distributed random variables with unit variance, provided that (4.2) of the main paper is satisfied for the subset $\widetilde{\mathcal{S}}_m$. This result fulfills the commonly-used independence assumption in the theory of nonparametric empirical Bayes (NPEB); see Zhang (2009); Jiang and Zhang (2009) for details. We will prove that $T_1^*(s;r)$ and $\widehat{T}_1(s)$ used in our implementation are close by using the near epoch dependency (NED, Assumption 4 of the main paper), the consistency of variance (i.e., showing $\widehat{\tau}(s)\to\tau(s)$), and the convergence of the normalization term (i.e., showing $\zeta_r(s)\to 1$). More precisely, we will show that $\widehat{T}_1(s)$ and $T_1^*(s;r)$ are close to each other through the following approximations:

$$\widehat{T}_1(s)\approx T_1(s)\approx T_1(s;r)\approx T_1^*(s;r),$$

where $T_1(s)$ is defined in (2.1) of the main paper (the auxiliary statistics with true variance); and $T_1(s;r) = T_1(s)/\zeta_r(s)$ is an intermediate variant involving the normalization term $\zeta_r(s)$ defined above. The difference between $T_1(s)$ and $\widehat{T}_1(s)$ lies oin whether the normalization term is the true standard deviation of $\sum_{v\in\mathcal{N}(s)} X(v)$ or its estimate. As for $T_1(s)$ and $T_1(s;r)$, they again differ by the normalization terms ($\tau(s)$ versus $\zeta_r(s)$). The difference between $T_1(s;r)$ and its conditional version $T_1^*(s;r)$ is controlled by NED. The following intermediate quantities are the corresponding GMLEs based on $\widehat{T}_1(s)$, $T_1(s;r)$, and $T_1^*(s;r)$:

- The GMLE based on $\{\widehat{T}_1(s), s \in \widetilde{\mathcal{S}}_m\}$ is denoted as
$$\widehat{G}_{\tilde{m}}(u) = \sum_{i=1}^{\widehat{l}} \widehat{\pi}_i \mathbf{1}\left\{\widehat{v}_i \le u\right\},$$
where $|\widehat{v}_i| \le \sup_{s\in\widetilde{\mathcal{S}}_m} |\widehat{T}_1(s)|$, $\widehat{l} \le \tilde{m}+1$, $\sum_{i=1}^{\widehat{l}} \widehat{\pi}_i = 1$, and $\widehat{\pi}_i > 0$;

- The GMLE based on $\{T_1(s;r) : s \in \widetilde{\mathcal{S}}_m\}$ is denoted as
$$\widetilde{G}_{\tilde{m},r}(u) = \sum_{i=1}^{\widetilde{l}} \widetilde{\pi}_i \mathbf{1}\left\{\widetilde{v}_i \le u\right\},$$
where $|\widetilde{v}_i| \le \sup_{s\in\widetilde{\mathcal{S}}_m} |T_1(s;r)|$, $\widetilde{l} \le \tilde{m}+1$, $\sum_{i=1}^{\widetilde{l}} \widetilde{\pi}_i = 1$, and $\widetilde{\pi}_i > 0$;

- The GMLE based on $\{T_1^*(s;r) : s \in \widetilde{\mathcal{S}}_m\}$ is denoted as
$$\widetilde{G}^*_{\tilde{m},r}(u) = \sum_{i=1}^{\widetilde{l}^*} \widetilde{\pi}^*_i \mathbf{1}\left\{\widetilde{v}^*_i \le u\right\},$$
where $|\widetilde{v}^*_i| \le \sup_{s\in\widetilde{\mathcal{S}}_m} |T_1^*(s;r)|$, $\widetilde{l}^* \le \tilde{m}+1$, $\sum_{i=1}^{\widetilde{l}^*} \widetilde{\pi}^*_i = 1$, and $\widetilde{\pi}^*_i > 0$.

Here we suppress the dependence of $\widetilde{\pi}_i$, $\widetilde{v}_i$, $\widetilde{\pi}^*_i$, and $\widetilde{v}^*_i$ on r to simplify the presentation. We now describe the key idea and steps for proving Lemma S.1 as visualized in Figure S.20. The key idea is to prove that $\widehat{G}_{\tilde{m}}$ converges to G_0 in terms of the Hellinger distance when

the subset $\widetilde{\mathcal{S}}_m$ satisfies (4.2) of the main paper. The details will be shown in Lemma S.8. To arrive at this result, we need the following steps. On the one hand, (iv) and (v) in Figure S.20 state that $\widehat{G}_{\tilde{m}}$ approximates

$$G_{\tilde{m},r}(u) = \frac{1}{\tilde{m}} \sum_{s \in \widetilde{\mathcal{S}}_m} \mathbf{1}\{\xi(s)/\zeta_r(s) \leq u\}, \tag{S.11}$$

with high accuracy (see Lemmas S.5 and S.6). On the other hand, (vi) in Figure S.20 states that $G_{\tilde{m},r}(u)$ approaches

$$G_{\tilde{m}}(u) = \frac{1}{\tilde{m}} \sum_{s \in \widetilde{\mathcal{S}}_m} \mathbf{1}\{\xi(s) \leq u\}$$

by showing that $\zeta_r(s)$ tends to 1 as $r \to \infty$ (see Lemma S.3). Together with Assumption 7 of the main paper, $G_{\tilde{m},r}(u)$ converges to $G_0(u)$ as $\tilde{m} \to \infty$, which will be shown in Lemma S.7. The detailed statements of Lemmas S.3–S.8 are presented in the next subsection.

S.III.2 Some Preliminary Lemmas

Following the proof sketch in Section S.III.1, we now introduce some preliminary lemmas. These lemmas are in accordance with the strategies depicted in Figure S.20. In particular, Lemmas S.3 to S.5 focus on a series of properties derived from NED and the consistency of variance. Lemma S.6 provides a large deviation inequality that gives the convergence rate of $\widehat{G}_{\tilde{m}}$ to $G_{\tilde{m},r}$ as defined in (S.11). Lemmas S.7 and S.8 show that $d_H(f_{\widehat{G}_{\tilde{m}}}, f_{G_0}) = o_P(1)$, which is an important step in the proof of Lemma S.1. Lemma S.9 establishes the law of large numbers for the number of false/true discoveries. The relations among Theorem 1 of the main paper and Lemmas S.1–S.9 are depicted in Figure S.21.

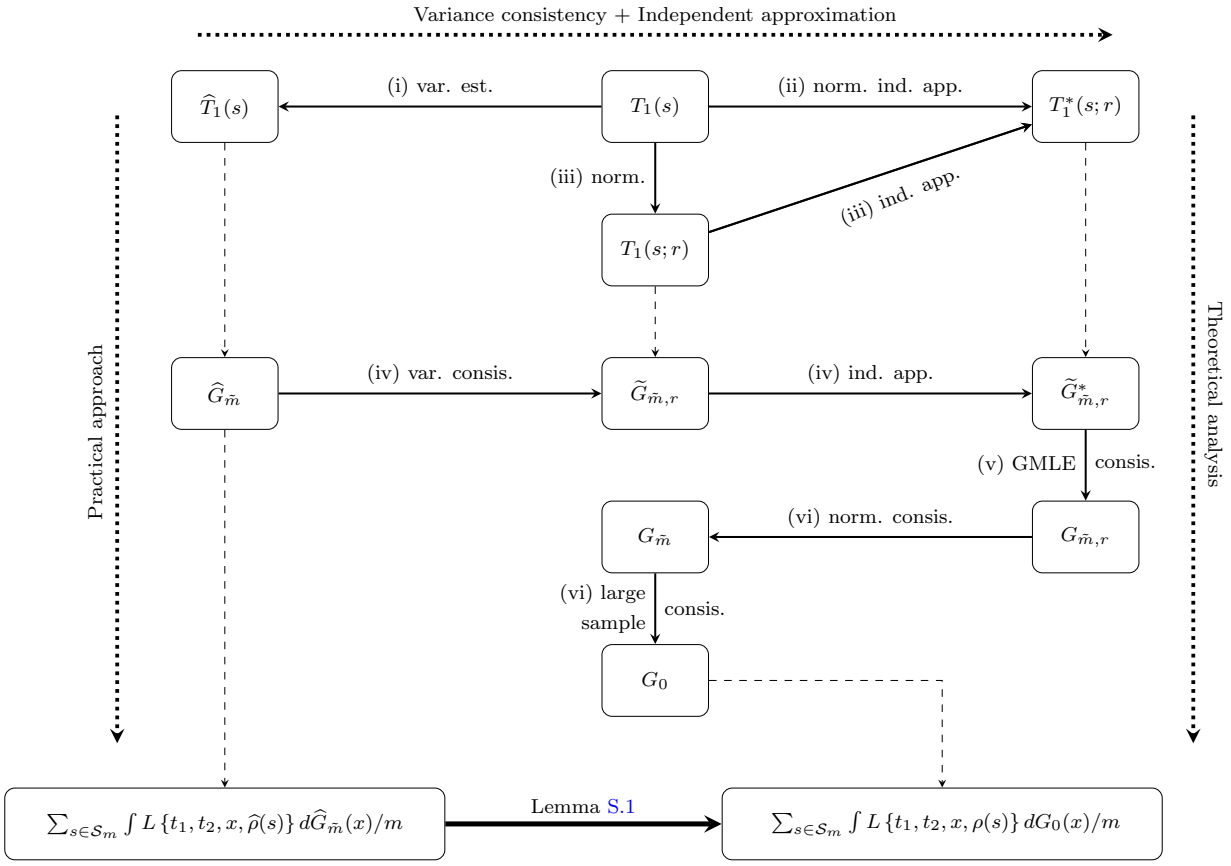


Figure S.20: The diagram of proving Lemma S.1 using the intermediate variants. The overall strategies are: (i) replacing the unknown variance with the true variance $\tau(s)$; (ii) replacing the correlated auxiliary statistics with its normalized independent approximation; (iii) analyzing the normalized independent approximation by its normalization term and independent approximation term; (iv) showing GMLE estimated from different sources are close; (v) showing GMLE $\widetilde{G}^*_{\tilde{m},r}$ estimated from $T_1^*(s;r)$ converges to $G_{\tilde{m},r}$; and (vi) proving that $G_{\tilde{m},r}$ approaches G_0 by showing $\zeta_r(s) \to 1$ and the large sample consistency. Lemma S.1 establishes the convergence using the above strategies. Here *var.*, *est.*, *norm.*, *ind.*, *app.*, and *consis.* are abbreviations for "variance", "estimation", "normalization", "independently", "approximation" and "consistency" respectively.

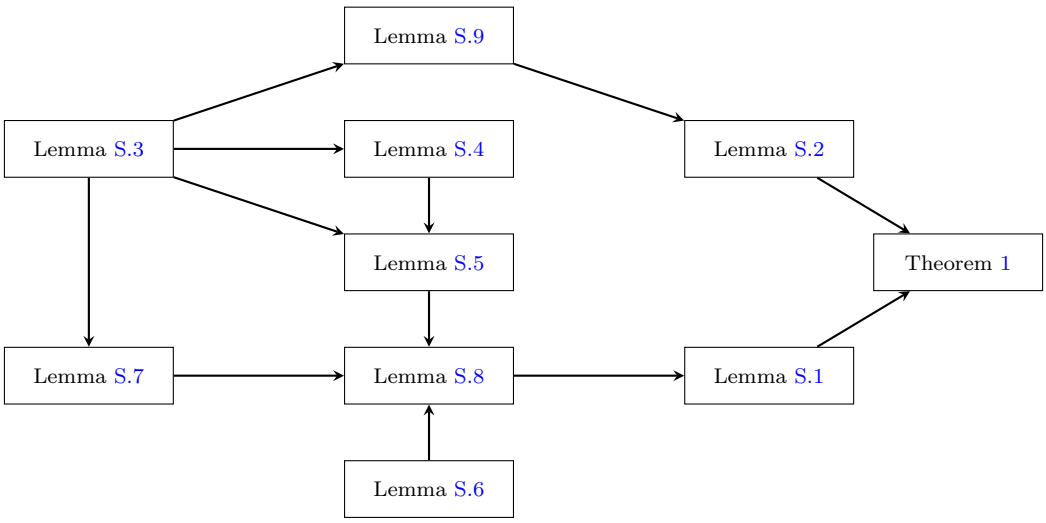


Figure S.21: The relations among Lemmas S.1–S.8 and Theorem 1.

Now, we start to present Lemmas S.3–S.9.

Lemma S.3 (Convergence of $\widehat{\tau}^2(s)$)**.** Under Assumptions 1–5 of the main paper, we have

(a) $|\zeta_r^2(s) - 1| \leq C\psi^2(r)$ and $|1/\zeta_r^2(s) - 1| \leq C\psi^2(r)$ uniformly for large enough r.

(b) $\sup_{s\in\widetilde{\mathcal{S}}_m} |\widehat{\tau}^2(s)/\tau_r^2(s) - 1| \leq C\psi^2(r) + C\tilde{m}^{-q}$ uniformly for large enough r with probability tending to one as $\tilde{m} \to \infty$.

(c) $\sup_m \sup_{s\in\mathcal{S}_m} \eta(s;r)$ is uniformly bounded for large enough r.

Remark S.1. According to the definition of NED (see Definition 1 of the main paper), Lemma S.3(a) implies that $\zeta_r(s) \to 1$ as $r \to \infty$. It further implies

$$\sup_{s\in\mathcal{S}_m} E\left\{T_1^*(s;r)\right\}^2 = 1 + \sup_{s\in\mathcal{S}_m} \left[E\{T_1^*(s;r)\}\right]^2 \leq 1 + \sup_{s\in\mathcal{S}_m} \nu_0^2/\zeta_r^2(s) \leq 1 + 2\nu_0^2 \triangleq \nu_1^2 \quad \text{(S.12)}$$

due to Assumption 7 of the main paper and the unit variance of $T_1^*(s;r)$ as defined in (4.1) of the main paper.

As for Lemma S.3(b), we note that

$$\begin{aligned}\frac{\sup_{s\in\widetilde{\mathcal{S}}_m}|\widehat{\tau}^2(s)/\tau_r^2(s)-1|}{1-\sup_{s\in\widetilde{\mathcal{S}}_m}|\widehat{\tau}^2(s)/\tau_r^2(s)-1|} &\le \{C\psi^2(r)+C\tilde{m}^{-q}\}/[1-\{C\psi^2(r)+C\tilde{m}^{-q}\}] \\ &\le C\psi^2(r)+C\tilde{m}^{-q}\end{aligned} \tag{S.13}$$

with high probability when r and $\tilde{m}$ are large. Both (S.13) and (S.12) will be useful in our theoretical analysis.

Remark S.2. A similar conclusion holds for primary statistics $T_2(s)$. Specifically, define

$$T_2^*(s;r)=E\left[T_2(s)\mid\mathfrak{F}\left\{\cup_{v\in\mathcal{N}(s)}B(v;r)\right\}\right]/\breve{\zeta}_r(s),$$

where $\breve{\zeta}_r^2(s)=\mathrm{var}\left(E\left[T_2(s)\mid\mathfrak{F}\left\{\cup_{v\in\mathcal{N}(s)}B(v;r)\right\}\right]\right)$. We have $\left|\breve{\zeta}_r^2(s)-1\right|\le C\psi^2(r)$ and $\left|1/\breve{\zeta}_r^2(s)-1\right|\le C\psi^2(r)$ uniformly and $\sup_m\sup_{s\in\mathcal{S}_m}\breve{\eta}(s;r)$ is uniformly bounded for large enough r.

Lemma S.4. Under Assumptions 1–5 and 7 of the main paper, for $\tilde{m}$ large enough and any $\delta_1,\delta_2>0$, with probability at least

$$1-\tilde{m}\exp(-\nu_1^2\delta_1^2/2)-C\tilde{m}\psi^p(r)/\nu_1^p\delta_1^p-C\psi^p(r)/\delta_2^p, \tag{S.14}$$

the following events occur simultaneously.

(a)

$$\sup_{s\in\widetilde{\mathcal{S}}_m}|T_1^*(s;r)|\le\nu_1(1+\delta_1)\quad\text{and}\quad\sup_{s\in\widetilde{\mathcal{S}}_m}|T_1(s;r)|\le\nu_1(1+2\delta_1).$$

(b)

$$\tilde{m}^{-1}\sum_{s\in\widetilde{\mathcal{S}}_m}|T_1(s;r)-T_1^*(s;r)|\le\delta_2\quad\text{and}\quad\tilde{m}^{-1}\left|\sum_{s\in\widetilde{\mathcal{S}}_m}T_1^2(s;r)-\{T_1^*(s;r)\}^2\right|\le\nu_1\delta_2(2+3\delta_1).$$

(c)

$$\tilde{m}^{-1} \sum_{s \in \widetilde{\mathcal{S}}_m} |T_1(s;r)| \leq \nu_1 (1+\delta_1) + \delta_2,$$

and

$$\tilde{m}^{-1} \sum_{s \in \widetilde{\mathcal{S}}_m} T_1^2(s;r) \leq \nu_1^2 + \nu_1 \delta_2 (2+3\delta_1) + \nu_1^2 \delta_1 (2+\delta_1).$$

Remark S.3. The three tail probabilities in (S.14) respectively correspond to the events of $\sup_{s \in \widetilde{\mathcal{S}}_m} |T_1^*(s;r) - E\{T_1^*(s;r)\}| \leq \nu_1 \delta_1$, $\sup_{s \in \widetilde{\mathcal{S}}_m} |T_1^*(s;r) - T_1(s;r)| \leq \nu_1 \delta_1$, and $\tilde{m}^{-1} \sum_{s \in \widetilde{\mathcal{S}}_m} |T_1(s;r) - T_1^*(s;r)| \leq \delta_2$. In the proof, we will show that these events imply the upper bounds in (a)–(c) of Lemma S.4.

Lemma S.5. Under Assumptions 1–5 and 7 of the main paper, with probability at least (S.14), the difference of generalized log-likelihood

$$\frac{1}{\tilde{m}} \left| \sum_{s \in \widetilde{\mathcal{S}}_m} \log \left[\frac{f_{\widetilde{G}^*_{\tilde{m},r}}\{T_1(s;r)\}}{f_{\widetilde{G}^*_{\tilde{m},r}}\{T_1^*(s;r)\}} \frac{f_{\widetilde{G}_{\tilde{m},r}}\{\widehat{T}_1(s)\}}{f_{\widetilde{G}_{\tilde{m},r}}\{T_1(s;r)\}} \frac{f_{\widehat{G}_{\tilde{m}}}\{T_1^*(s;r)\}}{f_{\widehat{G}_{\tilde{m}}}\{T_1(s;r)\}} \frac{f_{\widehat{G}_{\tilde{m}}}\{T_1(s;r)\}}{f_{\widehat{G}_{\tilde{m}}}\{\widehat{T}_1(s)\}} \right] \right| \tag{S.15}$$

is upper bounded by

$$C\nu_1^2 \left\{ \psi^2(r) + \tilde{m}^{-q} \right\} (1+\delta_1^2) + C\nu_1 \delta_2 (1+\delta_1).$$

Lemma S.6. Under Assumptions 2, 5, and 7 of the main paper, conditioning on $E|T_1^*(s;r)| \leq \nu_1$, if $\widehat{G}_{\tilde{m}}$ satisfies

$$\prod_{s \in \widetilde{\mathcal{S}}_m} \left[\frac{f_{\widehat{G}_{\tilde{m}}}\{T_1^*(s;r)\}}{f_{G_{\tilde{m},r}}\{T_1^*(s;r)\}} \right] \geq e^{-2t^2 \tilde{m} c_{\tilde{m}}^2/15}, \tag{S.16}$$

where

$$c_{\tilde{m}} = \sqrt{\frac{\log(\tilde{m})}{\tilde{m}}} \left\{ \tilde{m}^{1/b} \sqrt{\log \tilde{m}} \, (1 \vee \nu_1) \right\}^{b/(2+2b)} \tag{S.17}$$

for some $b > 0$, then there exists a universal constant t^* such that for all $t \geq t^*$ and

$\log \tilde{m} \geq 2/b$,

$$P\left\{d_H\left(f_{\widehat{G}_{\tilde{m}}}, f_{G_{\tilde{m},r}}\right) \geq t c_{\tilde{m}}\right\} \leq \exp\left\{-\frac{t^2 \tilde{m} c_{\tilde{m}}^2}{2 \log \tilde{m}}\right\} \leq e^{-t^2 \log \tilde{m}}.$$

Lemma S.6 is a direct consequence of Theorem 1 in Zhang (2009). It states that $d_H(f_{\widehat{G}_{\tilde{m}}}, f_{G_{\tilde{m},r}})$ would be small enough, as long as the generalized likelihood is nearly maximized in the sense of (S.16). This result together with Lemma S.5 allows us to replace $T_1^*(s;r)$ with $\widehat{T}_1(s)$ in estimating $G_{\tilde{m},r}$.

Lemma S.7. Under Assumptions 1–5, and 7 of the main paper, for $\tilde{m} \to \infty$ and arbitrary $\delta_3 > 0$, we have

$$d_H(f_{G_{\tilde{m},r}}, f_{G_0}) \leq C \nu_1^{1/2} \psi(r) + \delta_3,$$

with probability tending to one.

Lemma S.8. Under Assumptions 1–7 of the main paper, the GMLE based on $\{\widehat{T}_1(s), s \in \widetilde{\mathcal{S}}_m\}$ satisfies $d_H\left(f_{\widehat{G}_{\tilde{m}}}, f_{G_0}\right) = o_P(1)$ as $m \to \infty$.

Lemma S.8 is a consequence of Lemmas S.5–S.7. Its proof uses the convergence result of $d_H(f_{G_{\tilde{m},r}}, f_{G_0})$ in Lemma S.7, and shows the convergence rate in Lemma S.5 is fast enough to ensure that $\widehat{G}_{\tilde{m}}$ is an approximate GMLE of $G_{\tilde{m},r}$ and the condition in Lemma S.6 is fulfilled.

Lemma S.9. Under Assumptions 1–5, and 9 of the main paper, we have

$$\frac{\sum_{s \in \mathcal{S}_{0,m}} \mathbf{1}\left\{T_1(s) \geq t_1, T_2(s) \geq t_2\right\}}{m_0} := \frac{V_m(t_1,t_2)}{m_0} \xrightarrow{p} K_0\left(t_1,t_2\right),$$
$$\frac{\sum_{s \in \mathcal{S}_{1,m}} \mathbf{1}\left\{T_1(s) \geq t_1, T_2(s) \geq t_2\right\}}{m_1} := \frac{S_m(t_1,t_2)}{m_1} \xrightarrow{p} K_1\left(t_1,t_2\right),$$

as m_0 and m_1 goes to infinity.

Lemma S.9 addresses the challenge posed by the non-Lipschitz nature of $\mathbf{1}\{T_1(s) \geq$

$t_1, T_2(s) \geq t_2\}$, which complicates the application of the NED-based law of large numbers by a new approach.

S.III.3 Detailed Proofs of Lemmas S.1 and S.2

In this subsection, we complete the proofs of Lemmas S.1 and S.2 using Lemmas S.3–S.9.

Proof of Lemma S.1. Due to the triangular inequality, the LHS of (S.2) can be written as

$$\begin{aligned}
&\left|\frac{1}{m}\sum_{s\in\mathcal{S}_m}\left\{\int L\left\{t_1,t_2,x,\widehat{\rho}(s)\right\}d\widehat{G}_{\tilde{m}}(x)-\int L\left\{t_1,t_2,x,\rho(s)\right\}dG_0(x)\right\}\right| \\
&\quad\leq\left|\frac{1}{m}\sum_{s\in\mathcal{S}_m}\int\left[L\left\{t_1,t_2,x,\widehat{\rho}(s)\right\}-L\left\{t_1,t_2,x,\rho(s)\right\}\right]d\widehat{G}_{\tilde{m}}(x)\right| \\
&\qquad+\left|\frac{1}{m}\sum_{s\in\mathcal{S}_m}\int L\left\{t_1,t_2,x,\rho(s)\right\}\left\{d\widehat{G}_{\tilde{m}}(x)-dG_0(x)\right\}\right|.
\end{aligned}$$

Thus, we only need to prove

$$\sup_{|t_1|\leq t_1',|t_2|\leq t_2'}\left|\frac{1}{m}\sum_{s\in\mathcal{S}_m}\int\left[L\left\{t_1,t_2,x,\widehat{\rho}(s)\right\}-L\left\{t_1,t_2,x,\rho(s)\right\}\right]d\widehat{G}_{\tilde{m}}(x)\right|=o_P(1) \tag{S.18}$$

and

$$\sup_{|t_1|\leq t_1',|t_2|\leq t_2'}\left|\frac{1}{m}\sum_{s\in\mathcal{S}_m}\int L\left\{t_1,t_2,x,\rho(s)\right\}\left\{d\widehat{G}_{\tilde{m}}(x)-dG_0(x)\right\}\right|=o_P(1). \tag{S.19}$$

To simplify our presentation, we will prove (S.18) and (S.19) under the condition that $L\left(t_1,t_2,x,c\right)$ is uniformly continuous over $\mathbb{R}^3\times[-1,1]$. The justification of the uniform continuity of $L\left(t_1,t_2,x,c\right)$ is given later.

(i) Now, we prove (S.18) with the uniform continuity of $L\left(t_1,t_2,x,c\right)$. For an arbitrary $\epsilon>0$, there exists $0<\delta<2$ such that

$$\sup_{(t_1,t_2,x)\in\mathbb{R}^3}\left|L\left(t_1,t_2,x,c_1\right)-L\left(t_1,t_2,x,c_2\right)\right|<\epsilon, \tag{S.20}$$

for any $c_1, c_2 \in [-1,1]$ satisfying $|c_1 - c_2| < \delta$. Thus, we have

$$\sup_{(t_1,t_2,x)\in\mathbb{R}^3} \int \left| L\left(t_1,t_2,x,c_1\right) - L\left(t_1,t_2,x,c_2\right)\right| d\widehat{G}_{\tilde{m}}(x) < \epsilon.$$

It further implies that

$$\begin{aligned}
&P\left(\left|\frac{1}{m}\sum_{s\in\mathcal{S}_m}\int \left[L\left\{t_1,t_2,x,\widehat{\rho}(s)\right\} - L\left\{t_1,t_2,x,\rho(s)\right\}\right] d\widehat{G}_{\tilde{m}}(x)\right| > \epsilon\right) \\
&\le P\left\{\frac{1}{m}\sum_{s\in\mathcal{S}_m}\int \left|L\left\{t_1,t_2,x,\widehat{\rho}(s)\right\} - L\left\{t_1,t_2,x,\rho(s)\right\}\right| d\widehat{G}_{\tilde{m}}(x) > \epsilon; \sup_{s\in\mathcal{S}_m}\left|\rho(s)-\widehat{\rho}(s)\right| < \delta\right\} \\
&\qquad\qquad + P\left\{\sup_{s\in\mathcal{S}_m}\left|\rho(s)-\widehat{\rho}(s)\right| > \delta\right\} \\
&= P\left\{\sup_{s\in\mathcal{S}_m}\left|\rho(s)-\widehat{\rho}(s)\right| > \delta\right\} \to 0,
\end{aligned}$$

where the convergence in the last step is due to Assumptions 2 and 3(b) of the main paper.

(ii) To prove (S.19), it is sufficient to show that

$$\sup_{|t_1|\le t_1', |t_2|\le t_2'} \sup_{c\in[-1,1]} \left| \int L\left(t_1,t_2,x,c\right)\left\{d\widehat{G}_{\tilde{m}}(x) - dG_0(x)\right\}\right| = o_P(1). \tag{S.21}$$

To show (S.21), we note that the following pointwise convergence,

$$\int L\left(t_1,t_2,x,c\right)\left\{d\widehat{G}_{\tilde{m}}(x) - dG_0(x)\right\} = o_P(1), \tag{S.22}$$

can directly b obtained according to the proof of (44) in Yi et al. (2021) when (t_1,t_2,c) is fixed. For any $\epsilon > 0$, we can split $[-t_1', t_1'] \times [-t_2', t_2'] \times [-1,1]$ into B disjoint finite sets $\cup_{1\le k\le B}\mathcal{C}_k$ such that

$$\begin{aligned}
&\sup_{(t_1,t_2,c),(\tilde{t}_1,\tilde{t}_2,\tilde{c})\in\mathcal{C}_k} \left| \int \left\{L\left(t_1,t_2,x,c\right) - L\left(\tilde{t}_1,\tilde{t}_2,x,\tilde{c}\right)\right\} dG_0(x)\right| \le \epsilon/2, \\
\text{and} \quad &\sup_{(t_1,t_2,c),(\tilde{t}_1,\tilde{t}_2,\tilde{c})\in\mathcal{C}_k} \left| \int \left\{L\left(t_1,t_2,x,c\right) - L\left(\tilde{t}_1,\tilde{t}_2,x,\tilde{c}\right)\right\} d\widehat{G}_{\tilde{m}}(x)\right| \le \epsilon/2,
\end{aligned}$$

according to the uniform continuity of $L(t_1,t_2,x,c)$. Fixing $(t_1^k, t_2^k, c^k) \in \mathcal{C}_k, 1 \le k \le B$, we

have

$$\begin{aligned}
&\sup_{|t_1|\leq t_1',|t_2|\leq t_2'}\sup_{c\in[-1,1]}\left|\int L\left(t_1,t_2,,x,c\right)\left\{d\widehat{G}_{\tilde{m}}(x)-dG_0(x)\right\}\right| \\
&\quad=\sup_{(t_1,t_2,c)\in\cup_{1\leq k\leq B}\mathcal{C}_k}\left|\int L\left(t_1,t_2,x,c\right)\left\{d\widehat{G}_{\tilde{m}}(x)-dG_0(x)\right\}\right| \\
&\quad\leq\max_{1\leq k\leq B}\sup_{(t_1,t_2,c)\in\mathcal{C}_k}\left|\int L\left(t_1,t_2,x,c\right)-L\left(t_1^k,t_2^k,x,c^k\right)dG_0(x)\right| \\
&\qquad\qquad+\max_{1\leq k\leq B}\sup_{(t_1,t_2,c)\in\mathcal{C}_k}\left|\int L\left(t_1,t_2,x,c\right)-L\left(t_1^k,t_2^k,x,c^k\right)d\widehat{G}_{\tilde{m}}(x)\right| \\
&\qquad\qquad+\max_{1\leq k\leq B}\left|\int L\left(t_1^k,t_2^k,x,c^k\right)\left\{d\widehat{G}_{\tilde{m}}(x)-dG_0(x)\right\}\right| \\
&\quad\leq\epsilon+\max_{1\leq k\leq B}\left|\int L\left(t_1^k,t_2^k,x,c^k\right)\left\{d\widehat{G}_{\tilde{m}}(x)-dG_0(x)\right\}\right|.
\end{aligned}$$

Due to the pointwise convergence of (S.22) and the arbitrariness of $\epsilon>0$, the uniform convergence of (S.21) stands from the eabove displayed inequality. Finally, the definition of $K_0(t_1,t_2)$ in Assumption 9 together with (S.2) implies the conclusion of Lemma S.1.

We now justify the uniform continuity to complete the proof of Lemma S.1. We first prove $L\left(t_1,t_2,x,c\right)$ is uniformly continuous over $\mathbb{R}^3\times[-1+\delta_\rho,1-\delta_\rho]$ for any $0<\delta_\rho<1$. Denote by $f(u,v;c)$ the bivariate normal density with mean zero, variance one and correlation $c\in(-1,1)$. Then we have

$$\begin{aligned}
L(t_1,t_2,x,c)&=\int_{t_2-x}^{\infty}\int_{t_1}^{\infty}f(u,v;c)dvdu \\
&=\int_{t_2-x}^{\infty}du\int_{t_1}^{\infty}\frac{1}{2\pi\sqrt{(1-c^2)}}\exp\left\{-\frac{u^2-2cuv+v^2}{2\left(1-c^2\right)}\right\}dv.
\end{aligned}$$

Next, we prove that given $\delta_\rho, M>0$, c_1 and c_2 such that $|c_1-c_2|<\delta$ and $0<\delta<2-2\delta_\rho$,

the following inequalities hold

$$\begin{aligned}
&\sup_{(t_1,t_2,x)\in\mathbb{R}^3;-1+\delta_\rho\le c_1,c_2\le 1-\delta_\rho} |L(t_1,t_2,x,c_1)-L(t_1,t_2,x,c_2)| \\
&\quad\le \sup_{-1+\delta_\rho\le c_1,c_2\le 1-\delta_\rho} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |f(u,v;c_1)-f(u,v;c_2)|\,dudv \\
&\quad\le 8\int_M^{\infty}\phi(u)du + \sup_{-1+\delta_\rho\le c_1,c_2\le 1-\delta_\rho} \int_{-M}^{M}\int_{-M}^{M} \left|\frac{\partial f(u,v;c)}{\partial c}|_{c=\tilde{c}(c_1,c_2)}\right| |c_1-c_2|\,dudv \\
&\quad\le 8\int_M^{\infty}\phi(u)du + \delta\int_{-M}^{M}\int_{-M}^{M} \sup_{c\in[-1+\delta_\rho,1-\delta_\rho]} \left|\frac{\partial f(u,v;c)}{\partial c}\right| dudv \\
&\quad\le 8\int_M^{\infty}\phi(u)du + \delta C(\delta_\rho,M),
\end{aligned}$$

where the second inequality is achieved by covering $(-\infty,\infty)^2$ with five regions, namely $[-M,M]^2$, $(-\infty,-M)\times(-\infty,\infty)$,$(M,\infty)\times(-\infty,\infty)$,$(-\infty,\infty)\times(M,\infty)$ and $(-\infty,\infty)\times(-\infty,-M)$ and then applying the mean value theorem to the first region and using the fact $\int_{-\infty}^{\infty} f(u,v;c)dv=\phi(u)$ (and $\int_{-\infty}^{\infty} f(u,v;c)du=\phi(v)$) for the remaining four regions, and the last inequality stands because $\partial f(u,v;c)/\partial c$ is continuous and

$$C(M,\delta_\rho)=4M^2 \sup_{(u,v)\in[-M,M]^2,c\in[-1+\delta_\rho,1-\delta_\rho]} \left|\frac{\partial f(u,v;c)}{\partial c}\right|$$

is a positive constant depending only on M and δ_ρ. For any $\epsilon>0$, we can choose $M>0$ large enough such that $8\int_M^{\infty}\phi(u)du<\epsilon/2$. We then take $\delta(M,\delta_\rho)=\min\{2-2\delta_\rho,\epsilon/2C(\delta_\rho,M)\}>0$ to fulfill $|L(t_1,t_2,x,c_1)-L(t_1,t_2,x,c_2)|\le\epsilon$ when $|c_1-c_2|<\delta(M,\delta_\rho)$. We extend the uniform continuity for the correlation parameter c in $[-1+\delta_\rho,1-\delta_\rho]$ to $[-1,1]$. To this end, it is sufficient to show that for any $\epsilon>0$, there exists $\delta_\rho>0$ such that

$$\sup_{(t_1,t_2,x)\in\mathbb{R}^3} |L(t_1,t_2,x,1-\delta)-L(t_1,t_2,x,1)|\le\epsilon,$$

$$\text{and}\quad \sup_{(t_1,t_2,x)\in\mathbb{R}^3} |L(t_1,t_2,x,-1+\delta)-L(t_1,t_2,x,-1)|\le\epsilon,$$

for any δ satisfying $0 < \delta \leq \delta_\rho$. We present the proof of the first inequality above. The proof for the second equality is similar and thus omitted. Letting $t_1 - x \leq t_2$, we have

$$\begin{aligned}
|L(t_1, t_2, x, 1-\delta) - L(t_1, t_2, x, 1)| &= |L(t_1, t_2, x, 1-\delta) - P(V_2(s) \geq t_2)| \\
&= |P\{V_1(s) + x \geq t_1, V_2(s) \geq t_2\} - P\{V_2(s) \geq t_2\}| \\
&= P\{V_1(s) < t_1 - x, V_2(s) \geq t_2\} \\
&< P\{V_1(s) < V_2(s)\},
\end{aligned}$$

where $(V_1(s), V_2(s))$ follows a bivariate normal distribution with variance one and correlation $1-\delta$. Similarly, when $t_1 - x > t_2$, we have

$$|L(t_1, t_2, x, 1-\delta) - L(t_1, t_2, x, 1)| < P\{V_1(s) > V_2(s)\},$$

which implies

$$|L(t_1, t_2, x, 1-\delta) - L(t_1, t_2, x, 1)| < P\{V_1(s) \neq V_2(s)\}.$$

As $P\{V_1(s) \neq V_2(s)\}$ as a function of the correlation δ is right continuous at $\delta = 0$, a proper $\delta_\rho > 0$ can always be selected. To sum up, we have verified the uniform continuity and the proof is thus completed. □

Proof of Lemma S.2. To show the uniform convergence of $m_0^{-1}\widehat{V}_m(t_1, t_2)$ in (S.3), we begin with the pointwise convergence. In other words, we first show $|m_0^{-1}\widehat{V}_m(t_1, t_2) - K_0(t_1, t_2)| = o_P(1)$ for any fixed (t_1, t_2). It suffices to show that

$$P\left\{m_0^{-1}\widehat{V}_m(t_1, t_2) \leq K_0(t_1, t_2) + \delta_0\right\} \to 1 \tag{S.23}$$

and

$$P\left\{m_0^{-1}\widehat{V}_m(t_1, t_2) \geq K_0(t_1, t_2) - \delta_0\right\} \to 1 \tag{S.24}$$

as $m \to \infty$.

We now focus on (S.23). Observe that

$$\begin{aligned}\widehat{V}_m(t_1,t_2) &= \sum_{s\in\mathcal{S}_{0,m}} \mathbf{1}\left\{\widehat{T}_1(s)\geq t_1, \widehat{T}_2(s)\geq t_2\right\}\\ &= \sum_{s\in\mathcal{S}_{0,m}} \mathbf{1}\left\{T_1(s)\geq \widehat{r}^{(1)}(s)t_1, T_2(s)\geq \widehat{r}^{(2)}(s)t_2\right\}\\ &:= V_m\left(\widehat{\mathbf{r}}^{(1)}t_1, \widehat{\mathbf{r}}^{(2)}t_2\right),\end{aligned}$$

where $\widehat{r}^{(1)}(s) = \widehat{\sigma}(s)/\sigma(s)$, $\widehat{r}^{(2)}(s) = \widehat{\tau}(s)/\tau(s)$ for $s \in \mathcal{S}_{0,m}$, and $\widehat{\mathbf{r}}^{(k)} = (\widehat{r}^{(k)}(s))_{s\in\mathcal{S}_{0,m}}$ for $k = 1, 2$. For an arbitrary $\delta > 0$, we define three events $\mathcal{A}_{k,\delta} = \{\sup_{s\in\mathcal{S}_{0,m}} \left|\widehat{r}^{(k)}(s) - 1\right| \leq \delta\}$ for $k = 1, 2$, and

$$\mathcal{A}_{3,\delta}(t_1,t_2) = \left\{\left|m_0^{-1}V_m(t_1,t_2) - K_0(t_1,t_2)\right| \leq \delta\right\}.$$

It can be seen that $P(\mathcal{A}_{k,\delta}) \to 1$ for $k = 1, 2$ due to Assumptions 2 and 3. Similarly, $P\{\mathcal{A}_{3,\delta}(t_1,t_2)\} \to 1$ for any fixed $(t_1,t_2) \in \mathbb{R}^2$ as $m \to \infty$, because of Assumptions 2, 8, and Lemma S.9.To verify (S.23), we notice that

$$\begin{aligned}&\left\{\frac{1}{m_0}\widehat{V}_m(t_1,t_2) \leq K_0(t_1,t_2) + \delta_0\right\}\\ &\quad\supseteq \left\{\frac{1}{m_0}V_m\left(\widehat{\mathbf{r}}^{(1)}t_1, \widehat{\mathbf{r}}^{(2)}t_2\right) \leq K_0(t_1,t_2) + \delta_0\right\} \cap (\mathcal{A}_{1,\delta}\cap\mathcal{A}_{2,\delta})\\ &\quad\supseteq \left\{\frac{1}{m_0}V_m\left(\widehat{\mathbf{r}}^{(1)}t_1, \widehat{\mathbf{r}}^{(2)}t_2\right) \leq \frac{1}{m_0}V_m((1-\delta)t_1,(1-\delta)t_2)\right\} \cap (\mathcal{A}_{1,\delta}\cap\mathcal{A}_{2,\delta})\\ &\qquad\qquad \cap \left\{\left|\frac{1}{m_0}V_m((1-\delta)t_1,(1-\delta)t_2) - K_0((1-\delta)t_1,(1-\delta)t_2)\right| \leq \delta_0/2\right\}\\ &\qquad\qquad \cap \{|K_0((1-\delta)t_1,(1-\delta)t_2) - K_0(t_1,t_2)| \leq \delta_0/2\}\\ &\quad\supseteq (\mathcal{A}_{1,\delta}\cap\mathcal{A}_{2,\delta}) \cap \mathcal{A}_{3,\delta_0/2}((1-\delta)t_1,(1-\delta)t_2)\end{aligned}$$

$$\cap \left\{ |K_0((1-\delta)t_1, (1-\delta)t_2) - K_0(t_1, t_2)| \leq \delta_0/2 \right\},$$

where the first inclusion is because of $\widehat{V}_m(t_1, t_2) = V_m\left(\widehat{\mathbf{r}}^{(1)}t_1, \widehat{\mathbf{r}}^{(2)}t_2\right)$; the second inclusion is due to the triangle inequality; as for the third inclusion, we used the fact that V_m is monotonically decreasing with respect (t_1, t_2) and

$$\left\{ \frac{1}{m_0} V_m\left(\widehat{\mathbf{r}}^{(1)}t_1, \widehat{\mathbf{r}}^{(2)}t_2\right) \leq \frac{1}{m_0} V_m\left((1-\delta)t_1, (1-\delta)t_2\right) \right\}$$

holds true under $\mathcal{A}_{1,\delta} \cap \mathcal{A}_{2,\delta}$, and the definition of $\mathcal{A}_{3,\delta}$. Thus for an arbitrary $\delta_0 > 0$, we can choose $\delta > 0$ to guarantee

$$|K_0((1-\delta)t_1, (1-\delta)t_2) - K_0(t_1, t_2)| \leq \delta_0/2,$$

because $K_0(t_1, t_2)$ is continuous, and (S.23) holds due to

$$P\left(\mathcal{A}_{1,\delta} \cap \mathcal{A}_{2,\delta} \cap \mathcal{A}_{3,\delta_0/2}\left((1-\delta)t_1, (1-\delta)t_2\right)\right) \to 1.$$

For (S.24), it can be shown similarly and thus the pointwise convergence of $m_0^{-1}\widehat{V}_m(t_1, t_2)$ holds. The uniform convergence in (S.3) can be derived similarly as in the proof of Lemma S.1 after getting the pointwise convergence of $m_0^{-1}\widehat{V}_m(t_1, t_2)$. As for (S.4) and (S.5), these two results can be proved analogously as (S.3) and thus their proofs are omitted. □

S.IV Proofs of Lemmas S.3–S.9

This section is organized as follows. Section S.IV.1 discusses how to choose a set $\widetilde{\mathcal{S}}_m$ used for NPEB such that Assumption 6 of the main paper is satisfied. Section S.IV.2 presents the detailed proofs of Lemmas S.3–S.9.

S.IV.1 Justification of Assumption 6 of the main paper

Observe that $\{T_1^*(s;r) : s \in \widetilde{\mathcal{S}}_m\}$ defined in (4.1) of the main paper are independent once (4.2) of the main paper is fulfilled. Due to Assumptions 1 and 2 of the main paper, (4.2) of the main paper holds as long as

$$\mathrm{dist}(s, s') \geq 2N_{nei}\Delta_u + 2r, \quad s, s' \in \widetilde{\mathcal{S}}_m. \tag{S.25}$$

It is because for any $w \in \cup_{v\in\mathcal{N}(s)} B(v;r)$, $\mathrm{dist}(s,w) \leq \mathrm{dist}(s,v) + \mathrm{dist}(v,w) \leq N_{nei}\Delta_u + r$, where we have used Assumption 2 of the main paper that $\mathcal{N}(s)$ is the set of nearest neighbors with uniformly bounded cardinality for each location s. With (S.25), r can be taken as large as $\widetilde{\Delta}_{l,m}/2 - N_{nei}\Delta_u$ to ensure the independence of $\{T_1^*(s;r) : s \in \widetilde{\mathcal{S}}_m\}$, where $\widetilde{\Delta}_{l,m}$ is defined as in (4.3) of the main paper. In other words, r and $\widetilde{\Delta}_{l,m}$ are of the same order. We show that $\widetilde{\mathcal{S}}_m$ can be chosen such that

$$\tilde{m} = |\widetilde{\mathcal{S}}_m| \asymp m/r^K. \tag{S.26}$$

To this end, we pick all possible $s \in \mathcal{S}_m$ into $\widetilde{\mathcal{S}}_m$ which satisfies (S.25) for any $s' \in \widetilde{\mathcal{S}}_m$ until no more locations satisfy the condition. A typical choice is

$$\widetilde{\mathcal{S}}_m = \underset{\widetilde{\mathcal{S}}_m \in \widetilde{\mathscr{S}}_m}{\arg\max} |\widetilde{\mathcal{S}}_m| \quad \text{with} \quad \widetilde{\mathscr{S}}_m = \left\{ \widetilde{\mathcal{S}}_m \subset \mathcal{S}_m : \widetilde{\mathcal{S}}_m \text{ satisfies (S.25)} \right\}, \tag{S.27}$$

which can be viewed as the largest $2N_{nei}\Delta_u + 2r$-packing, motivated by the definition of packing number (see e.g., Definition 5.4 of Wainwright, 2019). Borrowing the idea of volume comparison lemma, we construct m K-dimensional cubes centered at the locations in $\mathcal{S}_m$ with the length $\Delta_l/2$, which are non-overlapping due to Assumption 1, and $\tilde{m}$ K-dimensional cubes centered at the locations in $\widetilde{\mathcal{S}}_m$ with the length $2(2N_{nei}\Delta_u + \Delta_u + 2r)$. It is straight-

forward to verify the cubes centered at $\widetilde{\mathcal{S}}_m$ cover all cubes centered at $\mathcal{S}_m$; otherwise, (S.27) will be violated. Thus m, $\tilde{m}$, and r satisfy

$$m\Delta_l^K/2^K \leq \tilde{m}2^K(2N_{nei}\Delta_u + \Delta_u + 2r)^K,$$

where the LHS is the total volume of cubes centered at $\mathcal{S}_m$ and the RHS is the total volume of cubes centered at $\widetilde{\mathcal{S}}_m$. Hence, we obtain

$$\tilde{m} \geq m\frac{\Delta_l^K}{4^K(2N_{nei}\Delta_u + \Delta_u + 2r)^K} \propto m/r^K.$$

To satisfy the other side of (S.26), we just need to pick fewer locations into the $\widetilde{\mathcal{S}}_m$. This completes the proof of (S.26).

Next, we show that $\widetilde{\mathcal{S}}_m$ can be constructed such that $\tilde{m}^{1/(\lambda p)}\{\log(\tilde{m})\}^{-1/(2\lambda)} = o(\widetilde{\Delta}_{l,m})$ to fulfill (S.47) and $\tilde{m} \to \infty$ as $m \to \infty$, where $\lambda, p > 0$ are associated with the L_p-NED in Assumption 4. The above requirement on $\widetilde{\mathcal{S}}_m$ can be guaranteed when we take, for simplicity, $\widetilde{\Delta}_{l,m} \asymp \tilde{m}^{1/(\lambda p)}$, which implies $r \asymp \tilde{m}^{1/(\lambda p)}$. Combining with (S.26), we get

$$\tilde{m} \asymp m^{\frac{\lambda p}{K+\lambda p}},$$

which tends to infinity and $\tilde{m}^{1/(\lambda p)}\{\log(\tilde{m})\}^{-1/(2\lambda)} = o(\widetilde{\Delta}_{l,m})$ as $m \to \infty$.

S.IV.2 Detailed Proofs of Lemmas S.3–S.9

In this section, we provide the detailed proofs of Lemmas S.3–S.8 stated in Section S.III.2. We first present some results about the difference between $T_1(s;r)$ and $T_1^*(s;r)$. Suppose X is $L_p(\mathbf{d})$-NED on the random field $Y = \{Y(s), s \in \mathcal{V}_m\}$. Then the L_p norm of the difference between $T_1(s)$ and $E\left[T_1(s) \mid \mathfrak{F}\left\{\cup_{v\in\mathcal{N}(s)}B(v;r)\right\}\right]$ is controlled by

$$\left\|T_1(s) - E\left[T_1(s) \mid \mathfrak{F}\left\{\cup_{v\in\mathcal{N}(s)}B(v;r)\right\}\right]\right\|_p$$

$$\leq \sum_{v\in\mathcal{N}(s)} \frac{\left\|X(v) - E\left[X(v) \mid \mathfrak{F}\left\{\cup_{v\in\mathcal{N}(s)} B(v;r)\right\}\right]\right\|_p}{\tau(s)}$$
$$\leq \sum_{v\in\mathcal{N}(s)} \frac{d_m(v)}{\tau(s)}\psi(r) \tag{S.28}$$

due to the triangular inequality and the generalized non-decreasing property. Accordingly, the difference between the normalized variants is controlled by

$$\|T_1(s;r) - T_1^*(s;r)\|_p \leq \sum_{v\in\mathcal{N}(s)} \frac{d_m(v)}{\tau(s)\zeta_r(s)}\psi(r) := \eta(s;r)\psi(r), \tag{S.29}$$

where $\eta(s;r) = \sum_{v\in\mathcal{N}(s)} d_m(v)/\{\tau(s)\zeta_r(s)\}$. Thus, it can be seen that

$$P\left\{|T_1(s;r) - T_1^*(s;r)| > \delta\right\} \leq \frac{1}{\delta^p}\|T_1(s;r) - T_1^*(s;r)\|_p^p \leq \frac{\eta^p(s;r)\psi^p(r)}{\delta^p}$$

for any $\delta > 0$ due to the Markov inequality. Subsequently, we can establish the convergence of

$$\sup_{s\in\widetilde{\mathcal{S}}_m} |T_1(s;r) - T_1^*(s;r)| \quad \text{and} \quad \frac{1}{\tilde{m}}\sum_{s\in\widetilde{\mathcal{S}}_m} |T_1(s;r) - T_1^*(s;r)|.$$

In particular, for a fixed $\delta > 0$, the maximum difference between $T_1(s;r)$ and $T_1^*(s;r)$ is controlled by

$$\begin{aligned} P\left\{\sup_{s\in\widetilde{\mathcal{S}}_m} |T_1(s;r) - T_1^*(s;r)| \leq \delta\right\} &= P\left\{\cap_{s\in\widetilde{\mathcal{S}}_m} |T_1(s;r) - T_1^*(s;r)| \leq \delta\right\} \\ &\geq 1 - \frac{1}{\delta^p}\sum_{s\in\widetilde{\mathcal{S}}_m} \eta^p(s;r)\psi^p(r). \end{aligned} \tag{S.30}$$

The mean difference is controlled by

$$\begin{aligned} P\left\{\frac{1}{\tilde{m}}\sum_{s\in\widetilde{\mathcal{S}}_m} |T_1(s;r) - T_1^*(s;r)| > \delta\right\} &\leq \frac{1}{\tilde{m}^p\delta^p}\left\|\sum_{s\in\widetilde{\mathcal{S}}_m} |T_1(s;r) - T_1^*(s;r)|\right\|_p^p \\ &\leq \frac{1}{\tilde{m}^p\delta^p}\left\{\sum_{s\in\widetilde{\mathcal{S}}_m} \|T_1(s;r) - T_1^*(s;r)\|_p\right\}^p \end{aligned} \tag{S.31}$$

$$\leq \frac{1}{\tilde{m}^p \delta^p} \left\{ \sum_{s \in \widetilde{\mathcal{S}}_m} \eta(s;r)\psi(r) \right\}^p,$$

where the first, second, and last inequalities are due to the Markov inequality, the triangular inequality, and (S.29), respectively.

Proof of Lemma S.3. First note that X is uniformly L_p-NED on Y for $p \geq 2$ implies that X is uniformly L_2-NED on Y.

(a) Recall $\zeta_r^2(s) = \operatorname{var}\left(E\left[T_1(s) \mid \mathfrak{F}\left\{\cup_{v\in\mathcal{N}(s)} B(v;r)\right\}\right]\right)$ and $\operatorname{var}\{T_1(s)\} = 1$. Then we have

$$\zeta_r^2(s) = \operatorname{var}\left(E\left[T_1(s) \mid \mathfrak{F}\left\{\cup_{v\in\mathcal{N}(s)} B(v;r)\right\}\right]\right) \leq \operatorname{var}\{T_1(s)\} = 1 \tag{S.32}$$

according to the law of total variance. Further, setting $p = 2$ in (S.28), we obtain

$$\left\| T_1(s) - E\left[T_1(s) \mid \mathfrak{F}\left\{\cup_{v\in\mathcal{N}(s)} B(v;r)\right\}\right] \right\|_2 \leq \sum_{v\in\mathcal{N}(s)} \frac{d_m(v)}{\tau(s)} \psi(r) \leq C\psi(r)$$

where the last inequality holds by Assumptions 2, 3(*a*) and 4 of the main paper. Therefore, we get

$$\begin{aligned}
\left|\zeta_r^2(s) - 1\right| &= \left|\operatorname{var}\left(E\left[T_1(s) \mid \mathfrak{F}\left\{\cup_{v\in\mathcal{N}(s)} B(v;r)\right\}\right]\right) - \operatorname{var}\{T_1(s)\}\right| \\
&= \left| E\left(E\left[T_1(s) \mid \mathfrak{F}\left\{\cup_{v\in\mathcal{N}(s)} B(v;r)\right\}\right] - T_1(s)\right)^2 \right| \\
&= \left\| T_1(s) - E\left[T_1(s) \mid \mathfrak{F}\left\{\cup_{v\in\mathcal{N}(s)} B(v;r)\right\}\right] \right\|_2^2 \\
&\leq C\psi^2(r),
\end{aligned}$$

where the second equality stands because $E\left(E\left[T_1(s) \mid \mathfrak{F}\left\{\cup_{v\in\mathcal{N}(s)} B(v;r)\right\}\right]\right) = E\{T_1(s)\}$. For the third equality, we have

$$E\left(T_1(s) E\left[T_1(s) \mid \mathfrak{F}\left\{\cup_{v\in\mathcal{N}(s)} B(v;r)\right\}\right]\right)$$

$$= E\left\{E\left(T_1(s)E\left[T_1(s) \mid \mathfrak{F}\left\{\cup_{v\in\mathcal{N}(s)}B(v;r)\right\}\right] \mid \mathfrak{F}\left\{\cup_{v\in\mathcal{N}(s)}B(v;r)\right\}\right)\right\}$$

$$= E\left(E\left[T_1(s) \mid \mathfrak{F}\left\{\cup_{v\in\mathcal{N}(s)}B(v;r)\right\}\right]\right)^2,$$

where $T_1(s)E\left[T_1(s) \mid \mathfrak{F}\left\{\cup_{v\in\mathcal{N}(s)}B(v;r)\right\}\right] \in L_1$ is due to the Gaussianity from Assumption 5 of the main paper.

Together with (S.32), for large enough r, we have

$$\left|1/\zeta_r^2(s) - 1\right| \leq \left\{1 - C\psi^2(r)\right\}^{-1} C\psi^2(r) \leq C\psi^2(r)$$

since $|1/x^2 - 1|$ decreases as $x^2 \in (0, 1]$ increases. This completes Part (a) of this lemma.

(b) Assumptions 2 and 3 of the main paper imply $|\widehat{\tau}(s)/\tau(s) - 1| = o_P(\tilde{m}^{-q})$ and $|\widehat{\tau}^2(s)/\tau^2(s) - 1| = o_P(\tilde{m}^{-q})$. For any $0 < \delta = C\tilde{m}^{-q} < 1$, combining it with the proved conclusion in (a), we have

$$\begin{aligned}
\left|\widehat{\tau}^2(s)/\tau_r^2(s) - 1\right| &= \left|\widehat{\tau}^2(s)/\tau^2(s)\zeta_r^2(s) - 1\right| \\
&\leq \left|\widehat{\tau}^2(s)/\tau^2(s)\zeta_r^2(s) - \widehat{\tau}^2(s)/\tau^2(s)\right| + \left|\widehat{\tau}^2(s)/\tau^2(s) - 1\right| \\
&\leq \widehat{\tau}^2(s)/\tau^2(s)\left|1/\zeta_r^2(s) - 1\right| + \left|\widehat{\tau}^2(s)/\tau^2(s) - 1\right| \\
&\leq \left(1 + C\tilde{m}^{-q}\right) C\psi^2(r) + C\tilde{m}^{-q} \\
&\leq C\psi^2(r) + C\tilde{m}^{-q}
\end{aligned}$$

for large enough r with probability tending to one as $\tilde{m} \to \infty$.

For (c), we notice that

$$\sup_m \sup_{s\in\mathcal{S}_m} \eta(s;r) = \sup_m \sup_{s\in\mathcal{S}_m} \sum_{v\in\mathcal{N}(s)} \frac{d_m(v)}{\tau(s)\zeta_r(s)} \leq C \sup_m \sup_{s\in\mathcal{S}_m} \frac{1}{\zeta_r(s)},$$

where the last inequality stands due to Assumptions 2, 3(a), and 4 of the main paper. Finally, $\sup_m \sup_{s\in\widetilde{\mathcal{S}}_m} \eta(s;r)$ is bounded for large enough r according to Part (a) of this

lemma. □

Proof of Lemma S.4. Note that

$$\begin{aligned}
P\Bigg(&\sup_{s\in\widetilde{\mathcal{S}}_m}|T_1^*(s;r)-E\{T_1^*(s;r)\}|>\nu_1\delta_1\cup\sup_{s\in\widetilde{\mathcal{S}}_m}|T_1^*(s;r)-T_1(s;r)|>\nu_1\delta_1\\
&\qquad\cup\tilde{m}^{-1}\sum_{s\in\widetilde{\mathcal{S}}_m}|T_1(s;r)-T_1^*(s;r)|>\delta_2\Bigg)\\
&\le P\left[\sup_{s\in\widetilde{\mathcal{S}}_m}|T_1^*(s;r)-E\{T_1^*(s;r)\}|>\nu_1\delta_1\right]\\
&\quad+P\left\{\sup_{s\in\widetilde{\mathcal{S}}_m}|T_1^*(s;r)-T_1(s;r)|>\nu_1\delta_1\right\}\\
&\quad+P\left\{\tilde{m}^{-1}\sum_{s\in\widetilde{\mathcal{S}}_m}|T_1(s;r)-T_1^*(s;r)|>\delta_2\right\}\\
&\le\tilde{m}\exp(-\nu_1^2\delta_1^2/2)+C\tilde{m}\psi^p(r)/\nu_1^p\delta_1^p+C\psi^p(r)/\delta_2^p
\end{aligned}\tag{S.33}$$

due to (S.30), (S.31), Lemma S.3(*c*), and the observation that

$$\begin{aligned}
P\left(\sup_{s\in\widetilde{\mathcal{S}}_m}|T_1^*(s;r)-E\{T_1^*(s;r)\}|>\nu_1\delta_1\right)&=P\left(\cup_{s\in\widetilde{\mathcal{S}}_m}[T_1^*(s;r)-E\{T_1^*(s;r)\}>\nu_1\delta_1]\right)\\
&\le\tilde{m}\exp(-\nu_1^2\delta_1^2/2)
\end{aligned}$$

implied directly by Assumption 5. Thus, to obtain Parts (a)–(c) of Lemma S.4 with probability at least (S.14), it suffices to show the desired upper bounds in Lemma S.4 can be derived from the complementary event of (S.33), i.e.,

$$\sup_{s\in\widetilde{\mathcal{S}}_m}|T_1^*(s;r)-E\{T_1^*(s;r)\}|\le\nu_1\delta_1,\tag{S.34}$$

$$\sup_{s\in\widetilde{\mathcal{S}}_m}|T_1^*(s;r)-T_1(s;r)|\le\nu_1\delta_1,\tag{S.35}$$

and

$$\tilde{m}^{-1} \sum_{s \in \widetilde{\mathcal{S}}_m} |T_1(s;r) - T_1^*(s;r)| \leq \delta_2. \tag{S.36}$$

Now let us turn to the cumbersome details.

(a) First, $\sup_{s \in \widetilde{\mathcal{S}}_m} T_1^*(s;r)$ is upper bounded by $\nu_1(1+\delta_1)$ through

$$\sup_{s \in \widetilde{\mathcal{S}}_m} |T_1^*(s;r)| \leq \sup_{s \in \widetilde{\mathcal{S}}_m} |T_1^*(s;r) - E\{T_1^*(s;r)\}| + \sup_{s \in \widetilde{\mathcal{S}}_m} |E\{T_1^*(s;r)\}| \leq \nu_1(1+\delta_1), \tag{S.37}$$

where the second inequality is due to (S.12) and (S.34). For $T_1(s;r)$, we have

$$\sup_{s \in \widetilde{\mathcal{S}}_m} |T_1(s;r)| \leq \sup_{s \in \widetilde{\mathcal{S}}_m} |T_1^*(s;r)| + \sup_{s \in \widetilde{\mathcal{S}}_m} |T_1^*(s;r) - T_1(s;r)| \leq \nu_1(1+2\delta_1), \tag{S.38}$$

where the second inequality is because of (S.35) and (S.37).

(b) To measure the difference between $T_1^2(s;r)$ and $\{T_1^*(s;r)\}^2$, we have

$$\begin{aligned}
&\tilde{m}^{-1} \left| \sum_{s \in \widetilde{\mathcal{S}}_m} T_1^2(s;r) - \{T_1^*(s;r)\}^2 \right| \\
&\quad = \tilde{m}^{-1} \left| \sum_{s \in \widetilde{\mathcal{S}}_m} \{T_1(s;r) - T_1^*(s;r)\} \{T_1(s;r) + T_1^*(s;r)\} \right| \\
&\quad \leq \left\{ \tilde{m}^{-1} \left| \sum_{s \in \widetilde{\mathcal{S}}_m} T_1(s;r) - T_1^*(s;r) \right| \right\} \left\{ \sup_{s \in \widetilde{\mathcal{S}}_m} |T_1(s;r) + T_1^*(s;r)| \right\} \\
&\quad \leq \left\{ \tilde{m}^{-1} \sum_{s \in \widetilde{\mathcal{S}}_m} |T_1(s;r) - T_1^*(s;r)| \right\} \left\{ \sup_{s \in \widetilde{\mathcal{S}}_m} |T_1(s;r)| + \sup_{s \in \widetilde{\mathcal{S}}_m} |T_1^*(s;r)| \right\} \\
&\quad \leq \nu_1 \delta_2 (2 + 3\delta_1),
\end{aligned} \tag{S.39}$$

where the first inequality in the above holds by the Hölder's inequality and the last inequality is due to (S.36)–(S.38).

(c) The upper bound for $\tilde{m}^{-1}\sum_{s\in\widetilde{\mathcal{S}}_m}|T_1(s;r)|$ is obtained by

$$\begin{aligned}
&\tilde{m}^{-1}\sum_{s\in\widetilde{\mathcal{S}}_m}|T_1(s;r)| \\
&= \tilde{m}^{-1}\sum_{s\in\widetilde{\mathcal{S}}_m}|T_1(s;r)-T_1^*(s;r)+T_1^*(s;r)-E\{T_1^*(s;r)\}+E\{T_1^*(s;r)\}| \\
&\le \tilde{m}^{-1}\sum_{s\in\widetilde{\mathcal{S}}_m}|T_1(s;r)-T_1^*(s;r)|+\sup_{s\in\widetilde{\mathcal{S}}_m}|T_1^*(s;r)-E\{T_1^*(s;r)\}|+\sup_{s\in\widetilde{\mathcal{S}}_m}|E\{T_1^*(s;r)\}| \\
&\le \nu_1(1+\delta_1)+\delta_2 \qquad \text{(S.40)}
\end{aligned}$$

where the last step in (S.40) is due to (S.12), (S.34), and (S.36). As for $\tilde{m}^{-1}\sum_{s\in\widetilde{\mathcal{S}}_m}T_1^2(s;r)$, the Hölder inequality implies

$$\begin{aligned}
&\tilde{m}^{-1}\sum_{s\in\widetilde{\mathcal{S}}_m}\left[\{T_1^*(s;r)\}^2-E\{T_1^*(s;r)\}^2\right] \\
&\quad\le\left|\tilde{m}^{-1}\sum_{s\in\widetilde{\mathcal{S}}_m}[T_1^*(s;r)-E\{T_1^*(s;r)\}]\right|\times\sup_{s\in\widetilde{\mathcal{S}}_m}|T_1^*(s;r)+E\{T_1^*(s;r)\}| \\
&\quad\le\sup_{s\in\widetilde{\mathcal{S}}_m}|T_1^*(s;r)-E\{T_1^*(s;r)\}|\times\left\{\sup_{s\in\widetilde{\mathcal{S}}_m}|T_1^*(s;r)|+\sup_{s\in\widetilde{\mathcal{S}}_m}E|T_1^*(s;r)|\right\} \\
&\quad\le\nu_1^2\delta_1(2+\delta_1), \qquad \text{(S.41)}
\end{aligned}$$

where the last line is due to (S.12), (S.34), and (S.37). We then have

$$\begin{aligned}
&\tilde{m}^{-1}\sum_{s\in\widetilde{\mathcal{S}}_m}T_1^2(s;r) \\
&\quad=\tilde{m}^{-1}\sum_{s\in\widetilde{\mathcal{S}}_m}\left[T_1^2(s;r)-\{T_1^*(s;r)\}^2\right]+\tilde{m}^{-1}\sum_{s\in\widetilde{\mathcal{S}}_m}\left[\{T_1^*(s;r)\}^2-E\{T_1^*(s;r)\}^2\right] \\
&\qquad\qquad+\tilde{m}^{-1}\sum_{s\in\widetilde{\mathcal{S}}_m}E\{T_1^*(s;r)\}^2
\end{aligned}$$

$$\leq \tilde{m}^{-1}\left|\sum_{s\in\widetilde{\mathcal{S}}_m} T_1^2(s;r) - \left\{T_1^*(s;r)\right\}^2\right| + \tilde{m}^{-1}\sum_{s\in\widetilde{\mathcal{S}}_m}\left[\left\{T_1^*(s;r)\right\}^2 - E\left\{T_1^*(s;r)\right\}^2\right]$$
$$+ \sup_{s\in\widetilde{\mathcal{S}}_m} E\left\{T_1^*(s;r)\right\}^2$$
$$\leq \nu_1^2 + \nu_1\delta_2(2+3\delta_1) + \nu_1^2\delta_1\left(2+\delta_1\right), \tag{S.42}$$

where we have used (S.12), (S.39), and (S.41) to obtain the last inequality in (S.42). The proof is thus completed. □

Proof of Lemma S.5. The proof of Lemma S.4 has shown that with probability at least (S.14), the events of (S.36)–(S.42) hold. It is thus sufficient to show that (S.15) is upper bounded by $C\nu_1^2\left\{\psi^2(r)+\tilde{m}^{-q}\right\}\left(1+\delta_1^2\right)+C\nu_1\delta_2\left(1+\delta_1\right)$ under the events of (S.36)–(S.42). To show this, we divide (S.15) into four components and examine their upper bounds under the events of (S.36)–(S.42) separately.

To begin with, it can be shown

$$\left|\frac{1}{\tilde{m}}\sum_{s\in\widetilde{\mathcal{S}}_m}\log f_{\widetilde{G}_{\tilde{m},r}}\{T_1(s;r)\} - \frac{1}{\tilde{m}}\sum_{s\in\widetilde{\mathcal{S}}_m}\log f_{\widetilde{G}_{\tilde{m},r}}\{\widehat{T}_1(s)\}\right|$$
$$= \left|\frac{1}{\tilde{m}}\sum_{s\in\widetilde{\mathcal{S}}_m}\log\left[\frac{\sum_{i=1}^{\tilde{l}}\widetilde{\pi}_i\phi\{T_1(s;r)-\widetilde{v}_i\}}{\sum_{i=1}^{\tilde{l}}\widetilde{\pi}_i\phi\{\widehat{T}_1(s)-\widetilde{v}_i\}}\right]\right|$$
$$= \left|\frac{1}{\tilde{m}}\sum_{s\in\widetilde{\mathcal{S}}_m}\log\left(e^{-\left\{T_1^2(s;r)-\widehat{T}_1^2(s)\right\}/2}\right.\right.$$
$$\left.\left.\times\frac{\sum_{i=1}^{\tilde{l}}\widetilde{\pi}_i\exp\left\{\widehat{T}_1(s)\widetilde{v}_i-\widetilde{v}_i^2/2\right\}\exp\left[\left\{T_1(s;r)-\widehat{T}_1(s)\right\}\widetilde{v}_i\right]}{\sum_{i=1}^{\tilde{l}}\widetilde{\pi}_i\exp\left\{\widehat{T}_1(s)\widetilde{v}_i-\widetilde{v}_i^2/2\right\}}\right)\right|$$
$$\leq \left|\frac{1}{\tilde{m}}\sum_{s\in\widetilde{\mathcal{S}}_m}\left\{\frac{\left|\widehat{T}_1^2(s)-T_1^2(s;r)\right|}{2}+\left|\widehat{T}_1(s)-T_1(s;r)\right|\max_{i=1,\cdots,\tilde{l}}|\widetilde{v}_i|\right\}\right|$$

$$\leq \left| \frac{1}{\tilde{m}} \sum_{s \in \widetilde{\mathcal{S}}_m} \left\{ \frac{\left| \widehat{T}_1^2(s) - T_1^2(s;r) \right|}{2} + \left| \widehat{T}_1(s) - T_1(s;r) \right| \sup_{s \in \widetilde{\mathcal{S}}_m} |T_1(s;r)| \right\} \right|$$

$$\leq \frac{\sup_{s \in \widetilde{\mathcal{S}}_m} |\widehat{\tau}^2(s)/\tau_r^2(s) - 1|}{2\left\{1 - \sup_{s \in \widetilde{\mathcal{S}}_m} |\widehat{\tau}^2(s)/\tau_r^2(s) - 1|\right\}} \left\{ \frac{1}{\tilde{m}} \sum_{s \in \widetilde{\mathcal{S}}_m} T_1^2(s;r) + \sup_{s \in \widetilde{\mathcal{S}}_m} |T_1(s;r)| \left| \frac{1}{\tilde{m}} \sum_{s \in \widetilde{\mathcal{S}}_m} T_1(s;r) \right| \right\},$$

where the second inequality follows from the fact that the support of $\widetilde{G}_{\tilde{m}}$ is within the range of $\left\{T_1(s;r) : s \in \widetilde{\mathcal{S}}_m\right\}$. Plugging in (S.38), (S.40), and (S.42), we observe that

$$\left| \frac{1}{\tilde{m}} \sum_{s \in \widetilde{\mathcal{S}}_m} T_1^2(s;r) \right| + \sup_{s \in \widetilde{\mathcal{S}}_m} |T_1(s;r)| \left| \frac{1}{\tilde{m}} \sum_{s \in \widetilde{\mathcal{S}}_m} T_1(s;r) \right|$$

$$\leq \nu_1^2 + \nu_1 \delta_2 (2 + 3\delta_1) + \nu_1^2 \delta_1 (2 + \delta_1) + \nu_1 (1 + 2\delta_1) \{\nu_1 (1 + \delta_1) + \delta_2\}$$

$$= 2\nu_1^2 + \nu_1 \delta_2 (3 + 5\delta_1) + \nu_1^2 \delta_1 (5 + 3\delta_1).$$

Combining the above with (S.13), we immediately obtain that the first component of the target difference (S.15) is bound by

$$C\nu_1 \left\{\psi^2(r) + \tilde{m}^{-q}\right\} \{2\nu_1 + \delta_2(3 + 5\delta_1) + \nu_1 \delta_1 (5 + 3\delta_1)\}.$$

Secondly, we have

$$\left| \frac{1}{\tilde{m}} \sum_{s \in \widetilde{\mathcal{S}}_m} \log f_{\widetilde{G}^*_{\tilde{m},r}} \{T_1(s;r)\} - \frac{1}{\tilde{m}} \sum_{s \in \widetilde{\mathcal{S}}_m} \log f_{\widetilde{G}^*_{\tilde{m},r}} \{T_1^*(s;r)\} \right|$$

$$= \left| \frac{1}{\tilde{m}} \sum_{s \in \widetilde{\mathcal{S}}_m} \log \left[\frac{\sum_{i=1}^{\widetilde{l}^*} \widetilde{\pi}_i^* \phi\{T_1(s;r) - \widetilde{v}_i^*\}}{\sum_{i=1}^{\widetilde{l}^*} \widetilde{\pi}_i^* \phi\{T_1^*(s;r) - \widetilde{v}_i^*\}} \right] \right|$$

$$= \left| \frac{1}{\tilde{m}} \sum_{s \in \widetilde{\mathcal{S}}_m} \log \left(e^{-\left[T_1^2(s;r) - \left\{T_1^*(s;r)\right\}^2\right]/2} \right.\right.$$

$$\left.\left. \times \frac{\sum_{i=1}^{\widetilde{l}^*} \widetilde{\pi}_i^* \exp\left\{T_1^*(s;r)\widetilde{v}_i^* - (\widetilde{v}_i^*)^2/2\right\} \exp\left[\{T_1(s;r) - T_1^*(s;r)\}\widetilde{v}_i^*\right]}{\sum_{i=1}^{\widetilde{l}^*} \widetilde{\pi}_i^* \exp\left\{T_1^*(s;r)\widetilde{v}_i^* - (\widetilde{v}_i^*)^2/2\right\}} \right) \right|$$

$$\leq \left| \frac{1}{\tilde{m}} \sum_{s \in \widetilde{\mathcal{S}}_m} \frac{\{T_1^*(s;r)\}^2 - T_1^2(s;r)}{2} \right| + \frac{1}{\tilde{m}} \sum_{s \in \widetilde{\mathcal{S}}_m} |T_1(s;r) - T_1^*(s;r)| \max_{i=1,\cdots,\widetilde{l}^*} |\widetilde{v}_i^*|$$

$$\leq \left| \frac{1}{\tilde{m}} \sum_{s \in \widetilde{\mathcal{S}}_m} \frac{\{T_1^*(s;r)\}^2 - T_1^2(s;r)}{2} \right| + \frac{1}{\tilde{m}} \sum_{s \in \widetilde{\mathcal{S}}_m} |T_1(s;r) - T_1^*(s;r)| \sup_{s \in \widetilde{\mathcal{S}}_m} |T_1^*(s;r)| ,$$

where the second inequality follows the fact that the support of $\widetilde{G}^*_{\tilde{m},r}$ is within the range of $\{T_1^*(s;r) : s \in \widetilde{\mathcal{S}}_m\}$. Using (S.36), (S.37), and (S.39), we further obtain that $\nu_1\delta_2(3+4\delta_1)$ is an upper bound for the above difference. As for the third component of the target (S.15), we have

$$\begin{aligned}
&\left| \frac{1}{\tilde{m}} \sum_{s \in \widetilde{\mathcal{S}}_m} \log f_{\widehat{G}_{\tilde{m}}}\{\widehat{T}_1(s)\} - \frac{1}{\tilde{m}} \sum_{s \in \widetilde{\mathcal{S}}_m} \log f_{\widehat{G}_{\tilde{m}}}\{T_1(s;r)\} \right| \\
&\quad \leq \frac{\sup_{s \in \widetilde{\mathcal{S}}_m} |\widehat{\tau}^2(s)/\tau_r^2(s) - 1|}{2\left\{1 - \sup_{s \in \widetilde{\mathcal{S}}_m} |\widehat{\tau}^2(s)/\tau_r^2(s) - 1|\right\}} \left\{ \left| \frac{1}{\tilde{m}} \sum_{s \in \widetilde{\mathcal{S}}_m} T_1^2(s;r) \right| + \sup_{s \in \widetilde{\mathcal{S}}_m} \left|\widehat{T}_1(s)\right| \left| \frac{1}{\tilde{m}} \sum_{s \in \widetilde{\mathcal{S}}_m} T_1(s;r) \right| \right\} \\
&\quad \leq \frac{\sup_{s \in \widetilde{\mathcal{S}}_m} |\widehat{\tau}^2(s)/\tau_r^2(s) - 1|}{2\left\{1 - \sup_{s \in \widetilde{\mathcal{S}}_m} |\widehat{\tau}^2(s)/\tau_r^2(s) - 1|\right\}} \\
&\qquad\qquad \times \left\{ \left| \frac{1}{\tilde{m}} \sum_{s \in \widetilde{\mathcal{S}}_m} T_1^2(s;r) \right| + \frac{\sup_{s \in \widetilde{\mathcal{S}}_m} |T_1(s;r)|}{1 - \sup_{s \in \widetilde{\mathcal{S}}_m} |\widehat{\tau}^2(s)/\tau_r^2(s) - 1|} \left| \frac{1}{\tilde{m}} \sum_{s \in \widetilde{\mathcal{S}}_m} T_1(s;r) \right| \right\} \\
&\quad \leq \frac{\sup_{s \in \widetilde{\mathcal{S}}_m} |\widehat{\tau}^2(s)/\tau_r^2(s) - 1|}{2\left\{1 - \sup_{s \in \widetilde{\mathcal{S}}_m} |\widehat{\tau}^2(s)/\tau_r^2(s) - 1|\right\}} \\
&\qquad\qquad \times \left\{ \left| \frac{1}{\tilde{m}} \sum_{s \in \widetilde{\mathcal{S}}_m} T_1^2(s;r) \right| + C \sup_{s \in \widetilde{\mathcal{S}}_m} |T_1(s;r)| \left| \frac{1}{\tilde{m}} \sum_{s \in \widetilde{\mathcal{S}}_m} T_1(s;r) \right| \right\},
\end{aligned}$$

where the last inequality holds for the same reason as for (S.13). Combining (S.13), (S.38), (S.40), and (S.42), we get the upper bound

$$\begin{aligned}
&C\left\{\psi^2(r) + \tilde{m}^{-q}\right\}\left\{\nu_1^2 + \nu_1\delta_2(2+3\delta_1) + \nu_1^2\delta_1(2+\delta_1) + C\nu_1(1+2\delta_1)\left\{\nu_1(1+\delta_1) + \delta_2\right\}\right\} \\
&= C\left\{\psi^2(r) + \tilde{m}^{-q}\right\}\left[\nu_1^2(1+C) + \nu_1\delta_2\left\{2 + C + \delta_1(3+2C)\right\} + \nu_1^2\delta_1\left\{2 + 3C + \delta_1(1+2C)\right\}\right] \\
&\leq C\nu_1\left\{\psi^2(r) + \tilde{m}^{-q}\right\}\left\{\nu_1 + \delta_2(1+\delta_1) + \nu_1\delta_1(1+\delta_1)\right\}.
\end{aligned}$$

For the fourth component, we obtain

$$\begin{aligned}
&\left|\frac{1}{\tilde{m}}\sum_{s\in\widetilde{\mathcal{S}}_m}\log f_{\widehat{G}_{\tilde{m}}}\{T_1^*(s;r)\}-\frac{1}{\tilde{m}}\sum_{s\in\widetilde{\mathcal{S}}_m}\log f_{\widehat{G}_{\tilde{m}}}\{T_1(s;r)\}\right| \\
&\quad\le\left|\frac{1}{\tilde{m}}\sum_{s\in\widetilde{\mathcal{S}}_m}\frac{\{T_1^*(s;r)\}^2-T_1^2(s;r)}{2}\right|+\frac{1}{\tilde{m}}\sum_{s\in\widetilde{\mathcal{S}}_m}|T_1(s;r)-T_1^*(s;r)|\sup_{s\in\widetilde{\mathcal{S}}_m}\left|\widehat{T}_1(s)\right| \\
&\quad\le\left|\frac{1}{\tilde{m}}\sum_{s\in\widetilde{\mathcal{S}}_m}\frac{\{T_1^*(s;r)\}^2-T_1^2(s;r)}{2}\right|+\frac{\sup_{s\in\widetilde{\mathcal{S}}_m}|T_1(s;r)|}{1-\sup_{s\in\widetilde{\mathcal{S}}_m}|\widehat{\tau}^2(s)/\tau_r^2(s)-1|} \\
&\qquad\qquad\times\left\{\frac{1}{\tilde{m}}\sum_{s\in\widetilde{\mathcal{S}}_m}|T_1(s;r)-T_1^*(s;r)|\right\} \\
&\quad\le\left|\frac{1}{\tilde{m}}\sum_{s\in\widetilde{\mathcal{S}}_m}\frac{\{T_1^*(s;r)\}^2-T_1^2(s;r)}{2}\right|+C\sup_{s\in\widetilde{\mathcal{S}}_m}|T_1(s;r)|\left\{\frac{1}{\tilde{m}}\sum_{s\in\widetilde{\mathcal{S}}_m}|T_1(s;r)-T_1^*(s;r)|\right\}.
\end{aligned}$$

Similarly, due to (S.36), (S.38), and (S.39), the upper bound of the above is

$$\nu_1\delta_2\left(2+3\delta_1\right)+C\nu_1\delta_2\left(1+2\delta_1\right)\le C\nu_1\delta_2\left(1+\delta_1\right).$$

Finally, combining the four components with the triangular inequality, we get the upper bound

$$\begin{aligned}
&\frac{1}{\tilde{m}}\left|\sum_{s\in\widetilde{\mathcal{S}}_m}\log\left[\frac{f_{\widetilde{G}^*_{\tilde{m},r}}\{T_1(s;r)\}}{f_{\widetilde{G}^*_{\tilde{m},r}}\{T_1^*(s;r)\}}\frac{f_{\widetilde{G}_{\tilde{m},r}}\{\widehat{T}_1(s)\}}{f_{\widetilde{G}_{\tilde{m},r}}\{T_1(s;r)\}}\frac{f_{\widehat{G}_{\tilde{m}}}\{T_1^*(s;r)\}}{f_{\widehat{G}_{\tilde{m}}}\{T_1(s;r)\}}\frac{f_{\widehat{G}_{\tilde{m}}}\{T_1(s;r)\}}{f_{\widehat{G}_{\tilde{m}}}\{\widehat{T}_1(s)\}}\right]\right| \\
&\qquad\le C\nu_1^2\left\{\psi^2(r)+\tilde{m}^{-q}\right\}\left\{1+\delta_1\left(1+\delta_1\right)\right\}+C\nu_1\delta_2\left(1+\delta_1\right) \\
&\qquad\le C\nu_1^2\left\{\psi^2(r)+\tilde{m}^{-q}\right\}\left(1+\delta_1^2\right)+C\nu_1\delta_2\left(1+\delta_1\right),
\end{aligned}$$

which finishes the proof. □

Lemma S.6 is a direct consequence of Theorem 1 in Zhang (2009). Interested readers are referred to Zhang (2009) and references therein for more details.

Proof of Lemma S.7. First, notice that $d_H^2(f_{G_{\tilde{m},r}}, f_{G_{\tilde{m}}}) \le d_{TV}(f_{G_{\tilde{m},r}}, f_{G_{\tilde{m}}})$, where d_{TV} de-

notes the total variation distance

$$d_{TV}\left(f_{G_{\tilde{m},r}}, f_{G_{\tilde{m}}}\right) = \frac{1}{2}\int \left| \int \phi(v-u)\{dG_{\tilde{m},r}(u) - dG_{\tilde{m}}(u)\} \right| dv.$$

Plugging the definitions of $G_{\tilde{m},r}(u)$ and $G_{\tilde{m}}(u)$ into the above equation, $d_{TV}\left(f_{G_{\tilde{m},r}}, f_{G_{\tilde{m}}}\right)$ can be written as

$$\begin{aligned}
&\frac{1}{2\tilde{m}}\int \left| \sum_{s\in\widetilde{\mathcal{S}}_m} \int \phi(v-u)[d\mathbf{1}\{\xi(s)/\zeta_r(s) \le u\} - d\mathbf{1}\{\xi(s)\le u\}] \right| dv \\
&\quad = \frac{1}{2\tilde{m}}\int \left| \sum_{s\in\widetilde{\mathcal{S}}_m} \phi\{v-\xi(s)/\zeta_r(s)\} - \phi\{v-\xi(s)\} \right| dv \\
&\quad \le \frac{1}{2\tilde{m}} \sum_{s\in\widetilde{\mathcal{S}}_m} \int \left|\phi\{v-\xi(s)/\zeta_r(s)\} - \phi\{v-\xi(s)\}\right| dv.
\end{aligned}$$

Note that $\phi\{v-\xi(s)/\zeta_r(s)\}$ and $\phi\{v-\xi(s)\}$ are symmetric about $x = \xi(s)\{\zeta_r(s)+1\}/2\zeta_r(s)$. The above integral is equivalent to

$$\begin{aligned}
&\frac{1}{\tilde{m}} \sum_{s\in\widetilde{\mathcal{S}}_m} \left| \int_{\xi(s)\{1+\zeta_r(s)\}/2\zeta_r(s)}^{\infty} [\phi\{v-\xi(s)/\zeta_r(s)\} - \phi\{v-\xi(s)\}]dv \right| \\
&\quad = \frac{1}{\tilde{m}} \sum_{s\in\widetilde{\mathcal{S}}_m} \left|\Phi[\xi(s)\{1-\zeta_r(s)\}/2\zeta_r(s)] - \Phi[\xi(s)\{\zeta_r(s)-1\}/2\zeta_r(s)]\right| \\
&\quad = \frac{1}{\tilde{m}} \sum_{s\in\widetilde{\mathcal{S}}_m} \left|\phi(u(s))\right| \left|\xi(s)\{1-\zeta_r(s)\}/\zeta_r(s)\right|,
\end{aligned}$$

where $u(s)$ is some value between $\xi(s)\{1-\zeta_r(s)\}/2\zeta_r(s)$ and $\xi(s)\{\zeta_r(s)-1\}/2\zeta_r(s)$ due to the mean value theorem. Finally, by Lemma S.3(a), (S.12), and the uniform boundedness of $\phi(x)$, $d_{TV}\left(f_{G_{\tilde{m},r}}, f_{G_{\tilde{m}}}\right)$ is upper bounded by $C\nu_1\psi^2(r)$. Taking the square root, we get

$$d_H(f_{G_{\tilde{m},r}}, f_{G_{\tilde{m}}}) \le d_{TV}^{1/2}\left(f_{G_{\tilde{m},r}}, f_{G_{\tilde{m}}}\right) \le C\nu_1^{1/2}\psi(r).$$

Assumption 7 states that $d_H(f_{G_{\tilde{m}}}, f_{G_0}) \le \delta_3$ for any $\delta_3 > 0$ with probability tending to one. The desired result follows from the triangular inequality. □

Proof of Lemma S.8. Due to the triangular inequality, the Hellinger distance $d_H := d_H\left(f_{\widehat{G}_{\tilde{m}}}, f_{G_0}\right)$ is bounded by the summation of the following two components,

$$d_{H,1} := d_H\left(f_{\widehat{G}_{\tilde{m}}}, f_{G_{\tilde{m},r}}\right) \quad \text{and} \quad d_{H,2} := d_H\left(f_{G_{\tilde{m},r}}, f_{G_0}\right)$$

for any $r > 0$. Thus, for an arbitrary $\epsilon > 0$, it shows that

$$\begin{aligned} P\left(d_H \geq \varepsilon\right) &\leq P\left(d_{H,1} \geq \epsilon/2\right) + P\left(d_{H,2} \geq \epsilon/2\right) \\ &\leq P\left(d_{H,1} \geq t c_{\tilde{m}}\right) + P\left\{d_{H,2} \geq C\nu_1\phi^2(r) + \epsilon/4\right\}, \end{aligned} \tag{S.43}$$

where the first inequality is because of the triangular inequality and the second inequality is achieved once (i) choosing r such that $C\nu_1\phi^2(r) \leq \epsilon/4$; (ii) $\tilde{m}$ is sufficiently large that induces arbitrarily small $c_{\tilde{m}}$ with fixed $b > 0$ defined at (S.17) in Lemma S.6. The second term of the RHS in (S.43) tends to zero by setting $\delta_3 = \epsilon/4$ in Lemma S.7. As for the first term of the RHS in (S.43), we have

$$\begin{aligned} P\left(d_{H,1} \geq t c_{\tilde{m}}\right) &= P\left(d_{H,1} \geq t c_{\tilde{m}}, \frac{1}{\tilde{m}} \sum_{s \in \widetilde{\mathcal{S}}_m} \log\left[\frac{f_{\widehat{G}_{\tilde{m}}}\{T_1^*(s;r)\}}{f_{G_{\tilde{m},r}}\{T_1^*(s;r)\}}\right] \geq -2t^2 c_{\tilde{m}}^2/15\right) \\ &\quad + P\left(\frac{1}{\tilde{m}} \sum_{s \in \widetilde{\mathcal{S}}_m} \log\left[\frac{f_{\widehat{G}_{\tilde{m}}}\{T_1^*(s;r)\}}{f_{G_{\tilde{m},r}}\{T_1^*(s;r)\}}\right] < -2t^2 c_{\tilde{m}}^2/15\right) \\ &\leq C\tilde{m}^{-2} + P\left(\frac{1}{\tilde{m}} \sum_{s \in \widetilde{\mathcal{S}}_m} \log\left[\frac{f_{\widehat{G}_{\tilde{m}}}\{T_1^*(s;r)\}}{f_{G_{\tilde{m},r}}\{T_1^*(s;r)\}}\right] < -2t^2 c_{\tilde{m}}^2/15\right) \end{aligned} \tag{S.44}$$

when t is large enough due to Lemma S.6. As for the second term of the RHS in (S.44), notice that

$$\begin{aligned} &\frac{1}{\tilde{m}} \sum_{s \in \widetilde{\mathcal{S}}_m} \log\left[\frac{f_{\widehat{G}_{\tilde{m}}}\{T_1^*(s;r)\}}{f_{G_{\tilde{m},r}}\{T_1^*(s;r)\}}\right] \\ &= \frac{1}{\tilde{m}} \sum_{s \in \widetilde{\mathcal{S}}_m} \log\left[\frac{f_{\widetilde{G}^*_{\tilde{m},r}}\{T_1^*(s;r)\}}{f_{G_{\tilde{m},r}}\{T_1^*(s;r)\}} \frac{f_{\widetilde{G}^*_{\tilde{m},r}}\{T_1(s;r)\}}{f_{\widetilde{G}^*_{\tilde{m},r}}\{T_1^*(s;r)\}} \frac{f_{\widetilde{G}_{\tilde{m},r}}\{T_1(s;r)\}}{f_{\widetilde{G}^*_{\tilde{m},r}}\{T_1(s;r)\}}\right. \end{aligned}$$

$$\times \frac{f_{\widetilde{G}_{\tilde{m},r}}\{\widehat{T}_1(s)\}}{f_{\widetilde{G}_{\tilde{m},r}}\{T_1(s;r)\}} \frac{f_{\widehat{G}_{\tilde{m}}}\{\widehat{T}_1(s)\}}{f_{\widetilde{G}_{\tilde{m},r}}\{\widehat{T}_1(s)\}} \frac{f_{\widehat{G}_{\tilde{m}}}\{T_1^*(s;r)\}}{f_{\widehat{G}_{\tilde{m}}}\{\widehat{T}_1(s)\}}\Bigg]$$

$$\geq \frac{1}{\tilde{m}} \sum_{s\in\widetilde{\mathcal{S}}_m} \log\left[\frac{f_{\widetilde{G}^*_{\tilde{m},r}}\{T_1(s;r)\}}{f_{\widetilde{G}^*_{\tilde{m},r}}\{T_1^*(s;r)\}} \frac{f_{\widetilde{G}_{\tilde{m},r}}\{\widehat{T}_1(s)\}}{f_{\widetilde{G}_{\tilde{m},r}}\{T_1(s;r)\}} \frac{f_{\widehat{G}_{\tilde{m}}}\{T_1^*(s;r)\}}{f_{\widehat{G}_{\tilde{m}}}\{T_1(s;r)\}} \frac{f_{\widehat{G}_{\tilde{m}}}\{T_1(s;r)\}}{f_{\widehat{G}_{\tilde{m}}}\{\widehat{T}_1(s)\}}\right], \quad \text{(S.45)}$$

where the last inequality is due to the definitions of $\widetilde{G}^*_{\tilde{m},r}$, $\widetilde{G}_{\tilde{m},r}$, and $\widehat{G}_{\tilde{m}}$. In addition, (S.17) implies

$$2t^2 c_{\tilde{m}}^2/15 = 2t^2 \tilde{m}^{-b/(1+b)} (\log \tilde{m})^{(2+3b)/(2+2b)} (1 \vee \nu_1)^{b/(1+b)}/15. \quad \text{(S.46)}$$

Plugging (S.45) and (S.46) into the second probability of the RHS of (S.44) yields

$$P\left(\left|\frac{1}{\tilde{m}} \sum_{s\in\widetilde{\mathcal{S}}_m} \log\left[\frac{f_{\widetilde{G}^*_{\tilde{m},r}}\{T_1(s;r)\}}{f_{\widetilde{G}^*_{\tilde{m},r}}\{T_1^*(s;r)\}} \frac{f_{\widetilde{G}_{\tilde{m},r}}\{\widehat{T}_1(s)\}}{f_{\widetilde{G}_{\tilde{m},r}}\{T_1(s;r)\}} \frac{f_{\widehat{G}_{\tilde{m}}}\{T_1^*(s;r)\}}{f_{\widehat{G}_{\tilde{m}}}\{T_1(s;r)\}} \frac{f_{\widehat{G}_{\tilde{m}}}\{T_1(s;r)\}}{f_{\widehat{G}_{\tilde{m}}}\{\widehat{T}_1(s)\}}\right]\right| > C\tilde{m}^{-a}\right),$$

for some a such that $b/(1+b) < a < \min(q, 1/p)$ when $\tilde{m}$ is large enough.

Next, we shall show that the above probability goes to zero as $m \to \infty$ when the subset $\widetilde{\mathcal{S}}_m$ used for NPEB fulfills Assumption 6 of the main paper. Using Lemma S.5, we only need to show for $\delta > 0$,

$$\tilde{m} \exp(-\nu_1^2\delta_1^2/2) \leq \delta, \quad C\tilde{m}\psi^p(r)/\nu_1^p\delta_1^p \leq \delta, \quad \text{and} \quad C\psi^p(r)/\delta_2^p \leq \delta,$$

as well as

$$C\nu_1^2\left\{\psi^2(r) + \tilde{m}^{-q}\right\}\left(1+\delta_1^2\right) = O(\tilde{m}^{-a}) \quad \text{and} \quad C\nu_1\delta_2\left(1+\delta_1\right) = O(\tilde{m}^{-a})$$

for some $\delta_1, \delta_2 > 0$. Note that $\tilde{m}\exp(-\nu_1^2\delta_1^2/2) \leq \delta$ is equivalent to $\delta_1 \geq \sqrt{2\ln(\tilde{m}/\delta)}/\nu_1$. We set

$$\delta_1 = \frac{1}{\nu_1}\sqrt{2\ln(\tilde{m}/\delta)}$$

in the following calculations. For δ_2, $C\nu_1\delta_2(1+\delta_1) \leq \tilde{m}^{-a}$ implies $\delta_2 \leq C\tilde{m}^{-a}/\nu_1(1+\delta_1)$.

With the choice of δ_1, we set

$$\delta_2 = \frac{C\tilde{m}^{-a}}{\nu_1 + \sqrt{2\ln(\tilde{m}/\delta)}}.$$

Similarly, with the above choice of δ_1, the constraint $C\nu_1^2\tilde{m}^{-q}\left(1+\delta_1^2\right) \leq \tilde{m}^{-a}$ becomes

$$\tilde{m}^{-q} \leq \frac{C\tilde{m}^{-a}}{\nu_1^2 + 2\ln(\tilde{m}/\delta)},$$

which can be achieved according to the upper and lower bounds of a. As $C\tilde{m}\psi^p(r)/\nu_1^p\delta_1^p \leq \delta$, $C\psi^p(r)/\delta_2^p \leq \delta$, and $C\nu_1^2\psi^2(r)\left(1+\delta_1^2\right) \leq \tilde{m}^{-a}$, we require

$$\begin{aligned} r^{-\lambda} \asymp \psi(r) \leq \min\Bigg[& C\delta\tilde{m}^{-1/p}\sqrt{\ln(\tilde{m}/\delta)},\ C\delta^{1/p}\tilde{m}^{-a}\Big/\Big\{\nu_1 + \sqrt{2\ln(\tilde{m}/\delta)}\Big\}, \\ & \tilde{m}^{-a/2}\Big/\sqrt{\nu_1^2 + 2\ln(\tilde{m}/\delta)}\Bigg] \end{aligned} \tag{S.47}$$

after plugging in δ_1 and δ_2 with Assumption 4. Again, according to the upper bounds of a, the RHS of (S.47) is dominated by its first term when $\tilde{m}$ is large enough. Thus (S.47) is a consequence of Assumption 6 when we take $r = \widetilde{\Delta}_{l,m}/2$, which completes the proof. □

In the proof of Lemma S.8, (S.47) requires r increases as $\tilde{m}$ increases for non-degenerating $\psi(r)$. A special case is m-dependent, that is $\psi(r) = 0$, whenever $r > R$ for some $R > 0$. According to (S.47), the choice of r is free of $\tilde{m}$ and only depends on R in this special case.

Proof of Lemma S.9. We prove the first convergence and the second can be proved similarly. For arbitrary $\epsilon > 0$, we can choose m_0 large enough such that

$$\begin{aligned} & P\left(\left|\frac{V_m(t_1,t_2)}{m_0} - K_0(t_1,t_2)\right| > \epsilon\right) \\ \leq & P\left(\left|\frac{V_m(t_1,t_2)}{m_0} - E\left\{\frac{V_m(t_1,t_2)}{m_0}\right\}\right| > \epsilon/2\right) + P\left(\left|E\left\{\frac{V_m(t_1,t_2)}{m_0}\right\} - K_0(t_1,t_2)\right| > \epsilon/2\right) \\ \leq & \frac{4\mathrm{var}\left\{V_m(t_1,t_2)\right\}}{\epsilon^2 m_0^2}, \end{aligned} \tag{S.48}$$

where the first inequality is due to the triangular inequality, the first probability in line 2 is bound by Chebyshev inequality and the second probability in line 2 holds by taking large enough m_0 and applying Assumption 9 of the main paper. We just need to show that line 3 of the above formula can be arbitrarily small by taking sufficient large m_0. Note that the numerator can be expressed as follows

$$\sum_{s\in\mathcal{S}_{0,m}}\left[\sum_{\substack{s'\in\mathcal{S}_{0,m}\\ \text{dist}(s,s')\leq\widetilde{r}}}\text{cov}\left\{1\left\{T_1(s)\geq t_1,T_2(s)\geq t_2\right\},1\left\{T_1(s')\geq t_1,T_2(s')\geq t_2\right\}\right\}\right.$$
$$\left.+\sum_{\substack{s'\in\mathcal{S}_{0,m}\\ \text{dist}(s,s')>\widetilde{r}}}\text{cov}\left\{1\left\{T_1(s)\geq t_1,T_2(s)\geq t_2\right\},1\left\{T_1(s')\geq t_1,T_2(s')\geq t_2\right\}\right\}\right]. \tag{S.49}$$

The idea is to choose an appropriate $\widetilde{r}$ so that the first summation does not have many terms and the covariances in the second summation are small enough. For the first part, we adopt the same idea in Section S.IV.1. Fixing $s\in\mathcal{S}_{0,m}$, we construct K-dimensional non-overlapping cubes centered at the locations $\{s'\in\mathcal{S}_{0,m}:\text{dist}(s,s')\leq\widetilde{r}\}$ with the length $\Delta_l/2$, and one big K-dimensional cube centered at location s with the length $2\widetilde{r}+\Delta_l/2$. Since the small cubes are covered by the big cube, we have

$$m(s)\leq(4\widetilde{r}/\Delta_l+1)^K,\quad\text{because}\quad m(s)(\Delta_l/2)^K\leq(2\widetilde{r}+\Delta_l/2)^K, \tag{S.50}$$

where $m(s)=|\{s'\in\mathcal{S}_{0,m}:\text{dist}(s,s')\leq\widetilde{r}\}|$.

For the second part, we are going to replace $T_1(s)$ and $T_2(s)$ with their normalized conditional statistics $T_1^*(s;r)$ and $T_2^*(s;r)$. Similar to the discussion of $T_1^*(s;r)$, $\left(T_1^*(s;r),T_2^*(s;r)\right)$ and $\left(T_1^*(v;r),T_1^*(v;r)\right)$ are independent if $s,s'\in\mathcal{S}$ satisfy (4.2) of the. W.l.o.g, we can define

$$r = \widetilde{r}/2 - N_{nei}\Delta_u, \tag{S.51}$$

where $\widetilde{r}$ large enough so that $r > 0$. Similar to the derivation of (S.28) and (S.29), for any $\delta > 0$, we have

$$P\left\{|T_2(s) - T_2^*(s;r)| > \delta\right\} \leq \frac{1}{\delta^p}\left\|T_2(s) - T_2^*(s;r)\right\|_p^p \leq \frac{\breve{\eta}^p(s;r)\psi^p(r)}{\delta^p}, \tag{S.52}$$

where $\breve{\eta}(s;r) = d_m(s)/\{\tau(s)\breve{\zeta}_r(s)\}$. Then, we can obtain that

$$\begin{aligned}
&P\left\{T_1(s) \geq t_1, T_2(s) \geq t_2, T_1(s') \geq t_1, T_2(s') \geq t_2\right\} \\
\leq &P\left\{T_1(s) \geq t_1, T_2(s) \geq t_2, T_1(s') \geq t_1, T_2(s') \geq t_2,\right. \\
&\left.|T_1(s) - T_1^*(s;r)| \leq \delta, |T_1(s') - T_1^*(s';r)| \leq \delta, |T_2(s) - T_2^*(s;r)| \leq \delta, |T_2(s') - T_2^*(s';r)| \leq \delta\right\} \\
&+P\left\{|T_1(s) - T_1^*(s;r)| > \delta\right\} + P\left\{|T_1(s') - T_1^*(s';r)| > \delta\right\} \\
&+P\left\{|T_2(s) - T_2^*(s;r)| > \delta\right\} + P\left\{|T_2(s') - T_2^*(s';r)| > \delta\right\} \\
\leq &P\left\{T_1^*(s;r) \geq t_1 - \delta, T_2^*(s;r) \geq t_2 - \delta\right\} P\left\{T_1^*(s';r) \geq t_1 - \delta, T_2^*(s';r) \geq t_2 - \delta\right\} + C\psi^p(r)/\delta^p,
\end{aligned}$$

where the last inequality holds by (S.29), (S.52), Lemma S.3(c), and Remark S.2. Similarly, we can derive that

$$\begin{aligned}
&P\left\{T_1^*(s;r) \geq t_1 + \delta, T_2^*(s;r) \geq t_2 + \delta\right\} \\
\leq &P\left\{T_1^*(s;r) \geq t_1 + \delta, T_2^*(s;r) \geq t_2 + \delta, |T_1(s) - T_1^*(s;r)| \leq \delta, |T_2(s) - T_2^*(s;r)| \leq \delta\right\} \\
&+P\left\{|T_1(s) - T_1^*(s;r)| > \delta\right\} + P\left\{|T_2(s) - T_2^*(s;r)| > \delta\right\} \\
\leq &P\left\{T_1(s) \geq t_1, T_2(s) \geq t_2\right\} + C\psi^p(r)/\delta^p.
\end{aligned}$$

Therefore, we have

$$\operatorname{cov}\left\{1\left\{T_1(s) \geq t_1, T_2(s) \geq t_2\right\}, 1\left\{T_1(s') \geq t_1, T_2(s') \geq t_2\right\}\right\}$$

$$
\begin{aligned}
=&P\left\{T_1(s)\geq t_1, T_2(s)\geq t_2, T_1(s')\geq t_1, T_2(s')\geq t_2\right\}\\
&\qquad - P\left\{T_1(s)\geq t_1, T_2(s)\geq t_2\right\}P\left\{T_1(s')\geq t_1, T_2(s')\geq t_2\right\}\\
\leq&P\left\{T_1^*(s;r)\geq t_1-\delta, T_2^*(s;r)\geq t_2-\delta\right\}P\left\{T_1^*(s';r)\geq t_1-\delta, T_2^*(s';r)\geq t_2-\delta\right\}\\
&-P\left\{T_1^*(s;r)\geq t_1+\delta, T_2^*(s;r)\geq t_2+\delta\right\}P\left\{T_1^*(s';r)\geq t_1+\delta, T_2^*(s';r)\geq t_2+\delta\right\}+C\psi^p(r)/\delta^p\\
:=&\epsilon_\delta + C\psi^p(r)/\delta^p.
\end{aligned}
$$

Similarly, the covariance can also be lower bounded by $-\epsilon_\delta - C\psi^p(r)/\delta^p$. Combine the two parts, and we can bound (S.49) by

$$
\sum_{s\in\mathcal{S}_{0,m}}\left[\sum_{\substack{s'\in\mathcal{S}_{0,m}\\ \mathrm{dist}(s,s')\leq\widetilde{r}}}1+\sum_{\substack{s'\in\mathcal{S}_{0,m}\\ \mathrm{dist}(s,s')>\widetilde{r}}}\left\{\epsilon_\delta+C\psi^p(r)/\delta^p\right\}\right]\leq\sum_{s\in\mathcal{S}_{0,m}}m(s)+m_0^2\epsilon_\delta+Cm_0^2\psi^p(r)/\delta^p.
$$

Choose a moderate large $\widetilde{r}=(\Delta_l(m_0)^{a/K}-1)/4$ for some $0<a<1$. Then, the probability (S.48) is bounded by

$$
4\left[m_0^{-(1-a)}+\epsilon_\delta+C\left\{\Delta_l(m_0^{a/K}-1)/8-N_{nei}\Delta_u\right\}^{-\lambda p}/\delta^p\right]/\epsilon^2,
$$

which can be arbitrarily small by choosing a $\delta>0$ such that ϵ_δ be small enough and let m_0 tend to infinity. The proof is completed. □

S.V Power Analysis

In this section, we establish the power enhancement of the 2d-SMT procedure as outlined in Theorem 2 of the main paper and explore the factors contributing to power improvement.

S.V.1 Proof of Theorem 2 of the Main Paper

Proof of Theorem 2 of the main paper. (i) Recall that

$$(t_1^{2d}, t_2^{2d}) = \underset{(t_1,t_2)\in\mathcal{F}_{q,\infty}^{2d}}{\arg\max}\ K(t_1,t_2) \quad \text{and} \quad t_2^{1d} = \underset{t_2\in\mathcal{F}_{q,\infty}^{1d}}{\arg\max}\, K(-\infty,t_2),$$

where $\mathcal{F}_{q,\infty}^{2d} = \{(t_1,t_2) : \mathrm{FDP}_\lambda^\infty(t_1,t_2) \le q\}$ and $\mathcal{F}_{q,\infty}^{1d} = \{t_2 : \mathrm{FDP}_\lambda^\infty(-\infty,t_2) \le q\}$. The corresponding percentages of true discoveries in the asymptotic sense are $\mathrm{PTD}^{1d} = K_1(-\infty, t_2^{1d})$ and $\mathrm{PTD}^{2d} = K_1(t_1^{2d}, t_2^{2d})$, respectively.

According to the definition of $\mathcal{F}_{q,\infty}^{1d}$ and $\mathcal{F}_{q,\infty}^{2d}$, we have $K(t_1^{2d}, t_2^{2d}) \ge K(-\infty, t_2^{1d})$. Note that $K(t_1,t_2) = \pi_0 K_0(t_1,t_2) + (1-\pi_0)K_1(t_1,t_2)$. If $K_0(t_1^{2d}, t_2^{2d}) < K_0(-\infty, t_2^{1d})$, then we can directly conclude $\mathrm{PTD}^{2d} > \mathrm{PTD}^{1d}$. Otherwise, in the case that $K_0(t_1^{2d}, t_2^{2d}) \ge K_0(-\infty, t_2^{1d})$, we begin by dividing the numerator and denominator of $\mathrm{FDP}_\lambda^\infty(t_1,t_2)$ by $K_0(t_1,t_2)$ and expressing $K(t_1,t_2)$ in terms of $K_0(t_1,t_2)$ and $K_1(t_1,t_2)$ as follows

$$\mathrm{FDP}_\lambda^\infty(t_1,t_2) = \frac{F(\lambda)}{\Phi(\lambda)} \frac{\left(\lim_{m\to\infty} \int \sum_{s\in\mathcal{S}_m} L(t_1,t_2,x,\rho(s)) dG_0(x)/m\right)/K_0(t_1,t_2)}{\pi_0 + (1-\pi_0)K_1(t_1,t_2)/K_0(t_1,t_2)}.$$

According to Assumption 9, we have

$$\frac{\lim_{m\to\infty} \int \sum_{s\in\mathcal{S}_m} L(t_1,t_2,x,\rho(s)) dG_0(x)/m}{K_0(t_1,t_2)} \ge 1. \tag{S.53}$$

Setting the first threshold as $-\infty$, we find that

$$\lim_{m\to\infty} \int \sum_{s\in\mathcal{S}_m} L(-\infty,t_2,x,\rho(s)) dG_0(x)/m = K_0(-\infty,t_2)$$

for any $t_2 \in \mathbb{R}$. Therefore, we have

$$q = \mathrm{FDP}_\lambda^\infty\left(t_1^{2d}, t_2^{2d}\right) \ge \frac{F(\lambda)}{\Phi(\lambda)} \frac{1}{\pi_0 + (1-\pi_0)K_1(t_1^{2d}, t_2^{2d})/K_0(t_1^{2d}, t_2^{2d})} \tag{S.54}$$

and

$$q = \text{FDP}_{\lambda}^{\infty}\left(-\infty, t_2^{1d}\right) = \frac{F(\lambda)}{\Phi(\lambda)} \frac{1}{\pi_0 + (1-\pi_0)K_1(-\infty, t_2^{1d})/K_0(-\infty, t_2^{1d})}$$

because $\text{FDP}_{\lambda}^{\infty}(t_1, t_2)$ is continuous. Since $K_0(t_1^{2d}, t_2^{2d}) \geq K_0(-\infty, t_2^{1d})$, we must have $\text{PTD}^{2d} = K_1(t_1^{2d}, t_2^{2d}) \geq K_1(-\infty, t_2^{1d}) = \text{PTD}^{1d}$.

(ii) The proof is similar to that of part (i), except that (S.54) becomes strict inequality. Therefore, whether $K_0(t_1^{2d}, t_2^{2d}) < K_0(-\infty, t_2^{1d})$ or $K_0(t_1^{2d}, t_2^{2d}) \geq K_0(-\infty, t_2^{1d})$, we have $\text{PTD}^{2d} > \text{PTD}^{1d}$. □

Theorem 2 part (ii) of the main paper suggests that if the expectations of auxiliary statistics under the alternative are generally larger than those under the null, the 2d-SMT procedure is more powerful than the 1d-SMT procedure even when they have the same number of discoveries. This also means that the power improvement of the 2d-SMT procedure is not just because of making more rejections by enlarging the searching ranges of the cutoffs but also due to the auxiliary statistics' ability to differentiate between the null and alternative hypotheses.

S.V.2 An Example

In this section, we explicitly analyze the power improvement of the 2d-SMT in a specific example.

Denote $m = |\mathcal{S}_m|$, $m_0 = |\mathcal{S}_{0,m}|$, $m_1 = |\mathcal{S}_{1,m}|$, $\mathcal{S}_{0,m} = \{1, 2, \cdots, m_0\}$ and $\mathcal{S}_{1,m} = \{m_0 + 11, m_0 + 12, \cdots, m_0 + m_1 + 10\}$. Suppose that $\mu(s) = 0$ for $s \in \mathcal{S}_{0,m}$ and $\mu(s) = \mu_1 > 0$ for $s \in \mathcal{S}_{1,m}$. Since $\mathcal{N}(s)$ is the κ-nearest neighbors, we notice $\mathcal{N}(s) \subseteq \mathcal{S}_{0,m}$ for all $s \in \mathcal{S}_{0,m}$ and $\mathcal{N}(s) \subseteq \mathcal{S}_{1,m}$ for all $s \in \mathcal{S}_{1,m}$ for $0 < \kappa < 10$. Given some $\rho_X \in (0, 1)$, we let

$\sigma(s) \equiv 1$, $\text{corr}(X(s), X(v)) = \rho_X$ if $|s - v| \leq \kappa$, and $\text{corr}(X(s), X(v)) = 0$ if $|s - v| > \kappa$. In Section 2.1, we have shown that $T_1(s) = \xi(s) + V_1(s)$ and $T_2(s) = \sigma^{-1}(s)\mu(s) + V_2(s)$, where $\xi(s) = \tau^{-1}(s) \sum_{v \in \mathcal{N}(s)} \mu(v)$ and

$$\begin{pmatrix} V_1(s) \\ V_2(s) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho(s) \\ \rho(s) & 1 \end{pmatrix} \right)$$

with $\rho(s) = \{\sigma(s)\tau(s)\}^{-1} \sum_{v \in \mathcal{N}(s)} \text{corr}\{\epsilon(s), \epsilon(v)\}$. We have $\xi(s) = \sigma^{-1}(s)\mu(s) = 0$ for $s \in \mathcal{S}_{0,m}$, and $\xi(s) = \mu_1 / \sqrt{\rho_X + (1 - \rho_X)/\kappa} \triangleq \xi_1$ as well as $\sigma^{-1}(s)\mu(s) = \mu_1$ for $s \in \mathcal{S}_1$ by calculation. Furthermore, locations s not on the boundary satisfy $\tau^2(s) = \kappa + (\kappa^2 - \kappa)\rho_X \triangleq \tau$ and $\rho(s) = \rho_X / \sqrt{\rho_X + (1 - \rho_X)/\kappa} \triangleq \rho$. These conditions imply the expected proportions of true and false rejections in Assumption 9 of the main paper are

$$K_0(t_1, t_2) \approx P\{V_1(s) \geq t_1, V_2(s) \geq t_1\} \text{ and } K_1(t_1, t_2) \approx P\{V_1(s) + \xi_1 \geq t_1, V_2(s) + \mu_1 \geq t_1\}$$

because only a minority of s deviate from τ and ρ.

We can explicitly confirm that $\widehat{\text{FDP}}_{\lambda, \widetilde{\mathcal{S}}}(t_1, t_2)$ in (2.7) of the main paper, with known parameters $(\sigma(s), \tau(s), \rho(s))$, is a conservative estimator of $\text{FDP}(t_1, t_2)$. As shown in (S.7), $\text{FDP}(t_1, t_2)$ uniformly converges to $\pi_0 K_0(t_1, t_2)/K(t_1, t_2)$. In 2d-SMT, the denominator $K(t_1, t_2)$ is estimated by the proportion of discoveries. The estimation of the numerator consists of two components. The first component π_0 can be conservatively estimated by $\widehat{\pi}_0$

$$\frac{\sum_{s \in \mathcal{S}} \mathbf{1}\{T_2(s) < \lambda\}}{|\mathcal{S}|\, \Phi(\lambda)} \approx \pi_0 + (1 - \pi_0) \frac{P(V_2(s) + \mu_1 < \lambda)}{\Phi(\lambda)}$$

according to (2.6). The second component $K_0(t_1, t_2)$ is approximated by

$$\lim_{m\to\infty} \int \frac{1}{m} \sum_{s\in\mathcal{S}_m} L(t_1, t_2, x, \rho(s)) dG_0(x)$$

$$\approx \pi_0 K_0(t_1, t_2) + (1-\pi_0) P\left\{V_1(s) + \xi_1 \geq t_1, V_2(s) \geq t_1\right\},$$

which serves as a conservative estimator of $K_0(t_1, t_2)$ because $\xi_1 > 0$.

We then numerically calculated the theoretical power improvement of the 2d-SMT over the 1d-SMT, defined as $(\mathrm{PTD}^{2d} - \mathrm{PTD}^{1d})/\,\mathrm{PTD}^{1d}$, and examined how it changes with the magnitudes under the alternative μ_1, the null proportion π_0, the number of neighbors κ, and the correlation ρ_X. The theoretical power improvement in Figures S.22–S.24 aligned with our previous numerical simulations. First, Figure S.22 demonstrates that the power increased alongside a growing number of neighbors. This tendency was also observed in Setups I and II from Section S.I.4. Second, Figures S.23 and S.24 illustrate the significant power enhancement in scenarios with sparse signals and weak correlation. Third, the power improvement decreased as the PTD^{1d}, which varied by altering μ_1, increased across all configurations in this section. The second and third observations were consistent with the simulation of all setups in Section 5.

S.VI Covariance Estimation

We discuss the estimation of the covariance matrix. Suppose the error process has the decomposition

$$\epsilon(s) = \epsilon_1(s) + \epsilon_2(s),$$

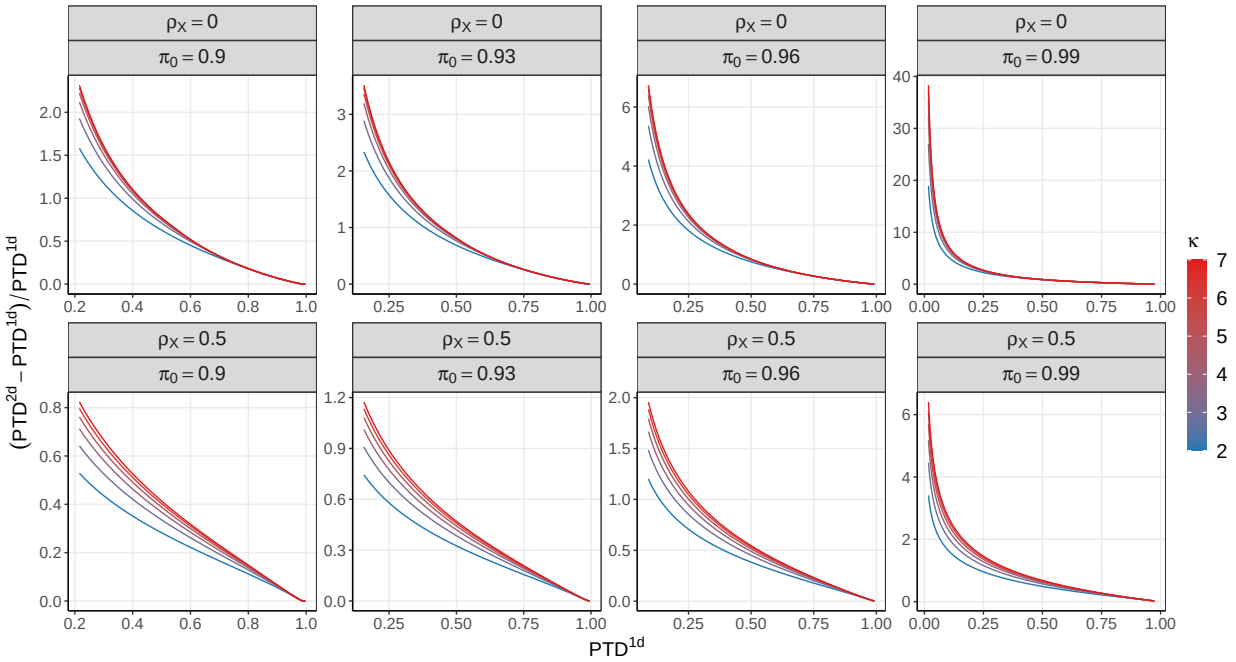


Figure S.22: The theoretical power improvement of the 2d-SMT procedure with $\kappa \in \{2, 3, \cdots, 7\}$, $\mu_1 \in [2, 5]$, $\pi_0 \in \{0.90, 0.93, 0.96, 0.99\}$, $\rho \in \{0, 0.5\}$.

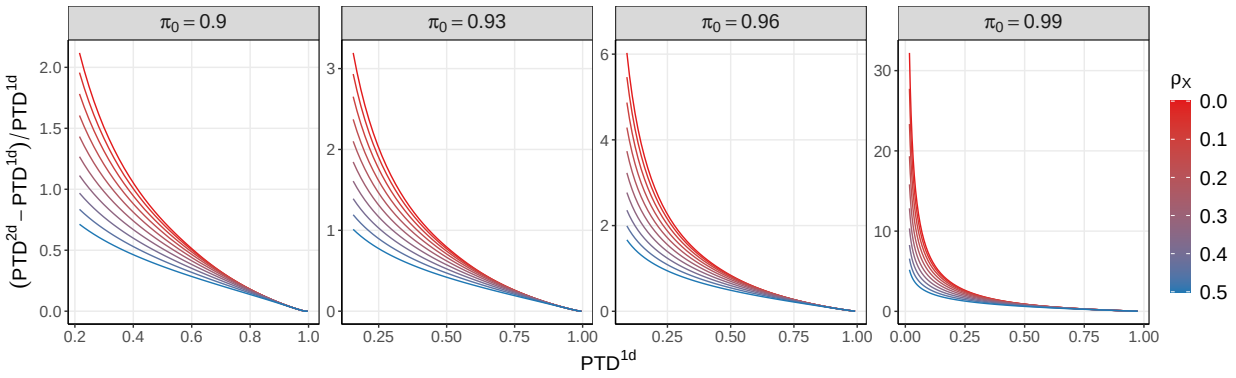


Figure S.23: The theoretical power improvement of the 2d-SMT procedure with $\kappa = 4$, $\mu_1 \in [2, 5]$, $\pi_0 \in \{0.90, 0.93, 0.96, 0.99\}$, $\rho \in [0, 0.5]$.

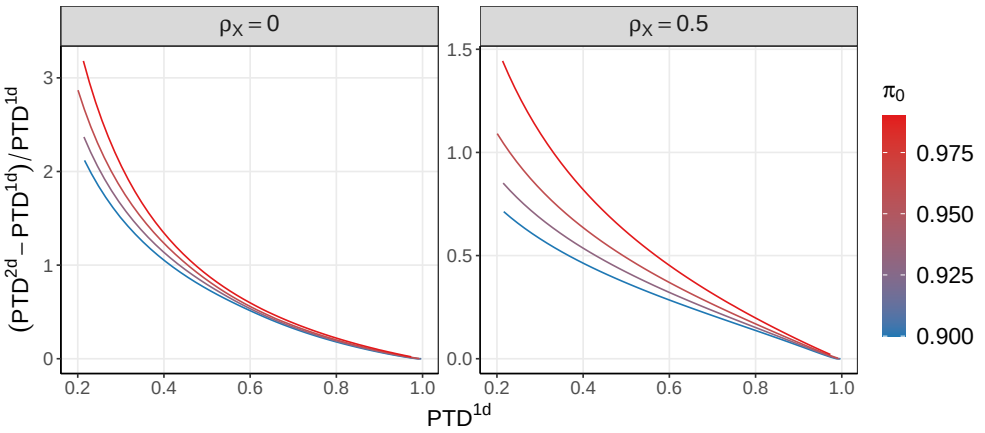


Figure S.24: The theoretical power improvement of the 2d-SMT procedure with $\kappa = 4$, $\mu_1 \in [2, 5]$, $\pi_0 \in \{0.90, 0.93, 0.96, 0.99\}$, $\rho \in \{0, 0.5\}$.

where $\epsilon_1(s)$ is a spatial process modeling the spatial correlation and $\epsilon_2(s)$ is an independent error process, known as the nugget effect, modeling measurement error. It is common to assume that (i) $\epsilon_1(s)$ is a spatial Gaussian process with the covariance function $c(\cdot, \cdot)$; (ii) $\epsilon_2(s)$ independently follows the normal distribution with mean zero and variance γ^2 at every location. For example, a popular choice of the spatial correlation function is the Matérn family of stationary correlation functions, i.e.,

$$\rho(s, s'; \nu, \phi) = \frac{1}{2^{\nu-1}\Gamma(\nu)} \left\{ \frac{2\nu^{1/2}\text{dist}(s, s')}{\phi} \right\}^{\nu} J_\nu \left(\frac{2\nu^{1/2}\text{dist}(s, s')}{\phi} \right), \quad \nu > 0, \phi > 0,$$

where Γ is the Gamma function, $J_\nu(\cdot)$ is the modified Bessel function of the second kind with order ν, and $\text{dist}(\cdot, \cdot)$ denotes the Euclidean distance. Here, ν controls the degree of smoothness and ϕ is the range parameter.

Generally, we can specify $c(\cdot, \cdot) = \sigma^2 \rho(\cdot, \cdot; \theta)$ for a correlation function ρ parameterized by θ, e.g., $\theta = (\nu, \phi)$ for the Matérn family. Recall that $\mathcal{S}_m = \{s_1, \ldots, s_m\}$ and let $C_m(\theta, \sigma^2) =$

$(\sigma^2\rho(s_i, s_j; \theta))_{i,j=1}^m$. The log-likelihood function of $\mathbf{X} = (X(s_1), \ldots, X(s_m))^\top$ is

$$l_m(\boldsymbol{\mu}, \theta, \sigma^2, \gamma^2) = -\frac{m}{2}\log(2\pi) - \frac{1}{2}\log\det\{C_m(\theta, \sigma^2) + \gamma^2 I_m\} \\ - \frac{1}{2}(\mathbf{X} - \boldsymbol{\mu})^\top\{C_m(\theta, \sigma^2) + \gamma^2 I_m\}^{-1}(\mathbf{X} - \boldsymbol{\mu}),$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)^\top$. Jointly estimating $(\boldsymbol{\mu}, \theta, \sigma^2, \gamma^2)$ without special structure for $\boldsymbol{\mu}$ is challenging. Some constraints are typically needed to regularize the form of $\boldsymbol{\mu}$. For example, when external covariates $\mathbf{Z} = (\mathbf{Z}(s_1), \ldots, \mathbf{Z}(s_m))^\top$ with $\mathbf{Z}(s_j) = (Z_1(s_j), \ldots, Z_p(s_j))^\top$ is available, we can model the signal by $\boldsymbol{\mu} = \mathbf{Z}\boldsymbol{\beta}$ for $\boldsymbol{\beta} \in \mathbb{R}^p$. Then, the parameters are estimated by

$$\underset{\boldsymbol{\beta}, \theta, \sigma^2, \gamma^2}{\arg\min}\ \log\det\{C_m(\theta, \sigma^2) + \gamma^2 I_m\} + (\mathbf{X} - \mathbf{Z}\boldsymbol{\beta})^\top\{C_m(\theta, \sigma^2) + \gamma^2 I_m\}^{-1}(\mathbf{X} - \mathbf{Z}\boldsymbol{\beta}).$$

Profiling out $\boldsymbol{\beta}$, we can obtain the estimates for the target parameters $(\theta, \sigma^2, \gamma^2)$ by solving the following problem

$$\underset{\theta, \sigma^2, \gamma^2}{\arg\min}\ \log\det\{C_m(\theta, \sigma^2) + \gamma^2 I_m\} + \{\mathbf{X} - \mathbf{Z}\widehat{\boldsymbol{\beta}}(\theta, \sigma^2, \gamma^2)\}^\top \\ \times \{C_m(\theta, \sigma^2) + \gamma^2 I_m\}^{-1}\{\mathbf{X} - \mathbf{Z}\widehat{\boldsymbol{\beta}}(\theta, \sigma^2, \gamma^2)\},$$

where $\widehat{\boldsymbol{\beta}}(\theta, \sigma^2, \gamma^2) = \arg\min_{\boldsymbol{\beta}}(\mathbf{X} - \mathbf{Z}\boldsymbol{\beta})^\top\{C_m(\theta, \sigma^2) + \gamma^2 I_m\}^{-1}(\mathbf{X} - \mathbf{Z}\boldsymbol{\beta})$.

Remark S.4. In some applications, we have multiple observations from the spatial random field

$$X_i(s) = \mu(s) + \epsilon_i(s), \quad i = 1, 2, \ldots, n.$$

The joint log-likelihood function is given by

$$l_{n,m}(\boldsymbol{\mu}, \theta, \sigma^2, \gamma^2) = -\frac{mn}{2}\log(2\pi) - \frac{n}{2}\log\det\{C_m(\theta, \sigma^2) + \gamma^2 I_m\}$$

$$-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{X}_i-\boldsymbol{\mu})^\top\{C_m(\theta,\sigma^2)+\gamma^2 I_m\}^{-1}(\mathbf{X}_i-\boldsymbol{\mu}),$$

with $\mathbf{X}_i=(X_i(s_1),\ldots,X_i(s_m))^\top$. In this case, we can estimate the covariance parameters by solving the problem

$$\arg\min_{\theta,\sigma^2,\gamma^2}\ n\log\det\{C_m(\theta,\sigma^2)+\gamma^2 I_m\}+\sum_{i=1}^{n}(\mathbf{X}_i-\bar{\mathbf{X}})^\top\{C_m(\theta,\sigma^2)+\gamma^2 I_m\}^{-1}(\mathbf{X}_i-\bar{\mathbf{X}})$$

with $\bar{\mathbf{X}}=n^{-1}\sum_{i=1}^{n}\mathbf{X}_i$.

S.VII Searching Algorithm

This section provides more details about Algorithm 1 in Section 3 of the main paper, and examines its computational complexity through numerical studies. Finding the optimal thresholds requires solving the constrained optimization problem (2.8) of the main paper. Due to the discrete nature of the problem, the solution can be obtained if we replace $\mathcal{F}_q$ by

$$\{(t_1,t_2)\in\mathcal{T}:\widehat{\mathrm{FDP}}_{\lambda,\widetilde{\mathcal{S}}}(t_1,t_2)\le q\},$$

where $\mathcal{T}=\{(\widehat{T}_1(s),\widehat{T}_2(s')):s,s'\in\mathcal{S}\}$ is the set of all candidate cutoff values. A naive grid search algorithm would require evaluating $\widehat{\mathrm{FDP}}_{\lambda,\widetilde{\mathcal{S}}}$ at $|\mathcal{S}|^2$ different values, which is computationally prohibitive for a large number of spatial locations. Interestingly, we show that there exists a faster algorithm that retains an exact maximization of (2.8) shown in the main paper. We derive our algorithm in three steps utilizing the specific structure of the optimization problem. For ease of presentation, we assume that there is no tie among $\{\widehat{T}_j(s):s\in\mathcal{S}\}$ for $j=1,2$.

Step 1. We partition the candidate set $\mathcal{T}$ into I subsets (say $\{\mathscr{S}_i\}_{i=1}^{I}$) such that the rejection set remains unchanged using the cutoffs within the same subset. For example, let

$\mathcal{D}(t_1, t_2) = \{s \in \mathcal{S} : \widehat{T}_1(s) \geq t_1, \widehat{T}_2(s) \geq t_2\}$ be the set of locations rejected using the cutoff (t_1, t_2). Then we have $\mathcal{D}(t_1, t_2) = \mathcal{D}(t_1', t_2')$ for $(t_1, t_2), (t_1', t_2') \in \mathscr{S}_i$. As $L\left(t_1, t_2, x, \widehat{\rho}(s)\right)$ is a non-increasing function of (t_1, t_2), we know that $\widehat{\mathrm{FDP}}_{\lambda,\widetilde{\mathcal{S}}}(t_1, t_2)$ is a non-increasing function within each $\mathscr{S}_i$ (note that $\widehat{R}(t_1, t_2)$ is a constant over $\mathscr{S}_i$). For $(t_1, t_2) \in \mathscr{S}_i$, $\widehat{\mathrm{FDP}}_{\lambda,\widetilde{\mathcal{S}}}(t_1, t_2)$ achieves its minimum value at $(\max_{(t_1,t_2)\in\mathscr{S}_i} t_1, \max_{(t_1,t_2)\in\mathscr{S}_i} t_2)$. Thus we can reduce the candidate set from $\mathcal{T}$ to

$$\mathcal{T}' = \left\{ \left(\max_{(t_1,t_2)\in\mathscr{S}_i} t_1, \max_{(t_1,t_2)\in\mathscr{S}_i} t_2 \right) : 1 \leq i \leq I \right\}, \tag{S.55}$$

where the maximum is defined to be infinity when $\mathscr{S}_i = \emptyset$. For each $\mathscr{S}_i$, we let $\mathcal{D}_i \subseteq \mathcal{S}$ be the set of locations associated with the hypotheses being rejected. It is not hard to verify that

$$\left(\max_{(t_1,t_2)\in\mathscr{S}_i} t_1, \max_{(t_1,t_2)\in\mathscr{S}_i} t_2 \right) = \left(\min_{s\in\mathcal{D}_i} \widehat{T}_1(s), \min_{s\in\mathcal{D}_i} \widehat{T}_2(s) \right). \tag{S.56}$$

Below we derive an alternative expression for $\mathcal{T}'$ which facilitates the implementation of our fast algorithm. Without loss of generality, let us assume that $\mathcal{S} = \{s_1, s_2, \cdots, s_m\}$ and

$$\widehat{T}_2(s_1) > \widehat{T}_2(s_2) > \cdots > \widehat{T}_2(s_m). \tag{S.57}$$

We claim that

$$\mathcal{T}' = \left\{ (\widehat{T}_1(s_l), \widehat{T}_2(s_k)) : \widehat{T}_1(s_l) \leq \widehat{T}_1(s_k) \text{ and } l \leq k,\ k = 1, 2, \ldots, m \right\} \cup \{(\infty, \infty)\}. \tag{S.58}$$

To prove the above result, let us consider any cutoff $(\widehat{T}_1(s_l), \widehat{T}_2(s_k))$ from the set defined in the RHS of (S.58). There exists a $1 \leq i \leq I$ such that the corresponding set of rejected

locations is given by $\mathcal{D}_i$. Then we have

$$\min_{s\in\mathcal{D}_i}\widehat{T}_1(s)\geq\widehat{T}_1(s_l)\quad\text{and}\quad\min_{s\in\mathcal{D}_i}\widehat{T}_2(s)\geq\widehat{T}_2(s_k). \tag{S.59}$$

Also note that $\mathcal{H}_{s_l}$ and $\mathcal{H}_{s_k}$ are both rejected as $\widehat{T}_2(s_l)\geq\widehat{T}_2(s_k)$ for $l\leq k$ and $\widehat{T}_1(s_l)\leq\widehat{T}_1(s_k)$ by the requirement in (S.58). Hence, both inequalities in (S.59) become equalities, i.e., $\min_{s\in\mathcal{D}_i}\widehat{T}_1(s)=\widehat{T}_1(s_l)$ and $\min_{s\in\mathcal{D}_i}\widehat{T}_2(s)=\widehat{T}_2(s_k)$, which shows the RHS of (S.58) belongs to $\mathcal{T}'$. To show the other direction, we note that for any $1\leq i\leq I$, there exist $1\leq k,l\leq m$ such that $s_k,s_l\in\mathcal{D}_i$, $\widehat{T}_1(s_l)=\min_{s\in\mathcal{D}_i}\widehat{T}_1(s)$ and $\widehat{T}_2(s_k)=\min_{s\in\mathcal{D}_i}\widehat{T}_2(s)$. Therefore, we get $\widehat{T}_1(s_l)\leq\widehat{T}_1(s_k)$ and $\widehat{T}_2(s_k)\leq\widehat{T}_2(s_l)$. In view of (S.57), we must have $l\leq k$ and thus $(\widehat{T}_1(s_l),\widehat{T}_2(s_k))$ is an element of the set in the RHS of (S.58).

To understand the complexity of $\mathcal{T}'$, we point out two extreme cases. In the first case, we assume that $\widehat{T}_1(s_1)>\widehat{T}_1(s_2)>\cdots>\widehat{T}_1(s_m)$. Then we have $\mathcal{T}'=\{(\widehat{T}_1(s_k),\widehat{T}_2(s_k)):k=1,2,\ldots,m\}\cup\{(\infty,\infty)\}$ and hence $|\mathcal{T}'|=m+1$. In the second case, suppose $\widehat{T}_1(s_1)<\widehat{T}_1(s_2)<\cdots<\widehat{T}_1(s_m)$. Then $\mathcal{T}'=\Big\{(\widehat{T}_1(s_l),\widehat{T}_2(s_k)):l\leq k,\,k=1,2,\ldots,m\Big\}\cup\{(\infty,\infty)\}$ and $|\mathcal{T}'|=m(m+1)/2+1$. In general, the cardinality of $\mathcal{T}'$ is between these extreme cases. See Figure S.25 for an illustration about $\mathcal{T}'$ in the intermediate case.

Step 2. Suppose that we have found some cutoff $(t_1,t_2)\in\mathbb{R}^2$ such that $\widehat{\mathrm{FDP}}_{\lambda,\widetilde{\mathcal{S}}}(t_1,t_2)\leq q$. Then we can reduce the candidate set by removing those cutoffs whose rejection numbers are less than $\widehat{R}(t_1,t_2)$. For example, let

$$\widetilde{t}_2^{\#}=\min\operatorname*{arg\,max}_{t_2:\widehat{\mathrm{FDP}}_{\lambda,\widetilde{\mathcal{S}}}(-\infty,t_2)\leq q}\widehat{R}(-\infty,t_2),$$

which can be efficiently computed using the BH procedure. Clearly, $\widehat{R}(t_1,t_2)\leq\widehat{R}(-\infty,\widetilde{t}_2^{\#})$ for $t_2>\widetilde{t}_2^{\#}$. Thus we can further reduce $\mathcal{T}'$ to $\mathcal{T}''=\big\{(t_1,t_2)\in\mathcal{T}':t_2\leq\widetilde{t}_2^{\#}\big\}$.

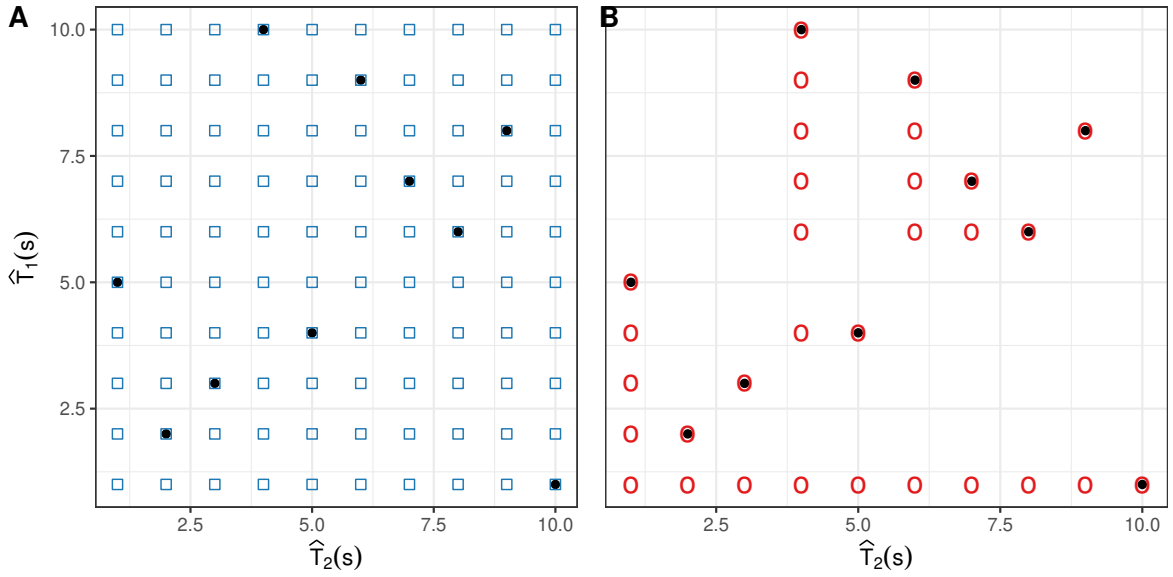


Figure S.25: The black points denote the values of the test statistics. The blue squares in Panel A denote all the candidate cutoffs in $\mathcal{T}$. The red circles in Panel B correspond to the cutoffs in $\mathcal{T}' \setminus \{(\infty, \infty)\}$.

Step 3. Searching over $\mathcal{T}''$ can still be computationally expensive. Here we show that pruning can be used to increase the computational efficiency whilst still ensuring that the method finds a global optimum of (2.8) of the main paper. The essence of pruning in this context is to remove those cutoffs that can never deliver the best number of rejections. To introduce pruning, we denote the elements in $\mathcal{T}''$ by $(t_{1,i,j}, t_{2,i})$ for $i = 1, 2 \ldots, \breve{m}$ and $1 \leq j \leq m_i$, where $\breve{m}$ is the number of $\widehat{T}_2(s)$'s that are no larger than $\widetilde{t}_2^{\#}$. Suppose the points are sorted in the following way: (1) $t_{2,1} > t_{2,2} > \cdots > t_{2,\breve{m}}$; (2) $t_{1,i,1} > t_{1,i,2} > \cdots > t_{1,i,m_i}$ for all $1 \leq i \leq \breve{m}$. We then examine the candidate cutoffs in a sequential way. In the outer loop of Algorithm 1 of the main paper below, we consider $t_{2,i}$ for i running from 1 to $\breve{m}$. In the inner loop, for a given i, we consider $t_{1,i,j}$ for j running from 1 to m_i. Let (t_1^C, t_2^C) be the best cutoff value we have found so far that delivers the largest number of rejections while controlling $\widehat{\text{FDP}}_{\lambda,\widetilde{\mathcal{S}}}$ at the desired level q. We skip all the remaining cutoffs whose rejection

numbers are less than $\widehat{R}(t_1^C, t_2^C)$. We replace (t_1^C, t_2^C) by $(t_{1,i,j}, t_{2,i})$ if one of the following two conditions is satisfied:

1. $\widehat{R}(t_{1,i,j}, t_{2,i}) = \widehat{R}(t_1^C, t_2^C)$ and $\widehat{\mathrm{FDP}}_{\lambda,\widetilde{\mathcal{S}}}(t_{1,i,j}, t_{2,i}) < \widehat{\mathrm{FDP}}_{\lambda,\widetilde{\mathcal{S}}}(t_1^C, t_2^C) \leq q$;

2. $\widehat{R}(t_{1,i,j}, t_{2,i}) > \widehat{R}(t_1^C, t_2^C)$ and $\widehat{\mathrm{FDP}}_{\lambda,\widetilde{\mathcal{S}}}(t_{1,i,j}, t_{2,i}) \leq q$.

When the cutoff $(t_{1,i,j}, t_{2,i})$ being examined satisfies neither condition, instead of moving directly to the next cutoff on the list, we can use the statistics we have computed in the current step to decide the minimum number of rejections required for the next cutoff. Specifically, note that

$$\frac{\int L\left(t_{1,i,j}, t_{2,i}, x, \widehat{\rho}(s)\right) d\widehat{G}_{\widetilde{\mathcal{S}}}(x)}{\widehat{R}(t_{1,i,j'}, t_{2,i})} \leq \frac{\int L\left(t_{1,i,j'}, t_{2,i}, x, \widehat{\rho}(s)\right) d\widehat{G}_{\widetilde{\mathcal{S}}}(x)}{\widehat{R}(t_{1,i,j'}, t_{2,i})} = \widehat{\mathrm{FDP}}_{\lambda,\widetilde{\mathcal{S}}}(t_{1,i,j'}, t_{2,i}).$$

For a cutoff to be valid, we need $\widehat{\mathrm{FDP}}(t_{1,i,j'}, t_{2,i}) \leq q$, which implies that

$$\begin{aligned}
\widehat{R}(t_{1,i,j'}, t_{2,i}) &\geq \left\lceil \frac{1}{q} \int L\left(t_{1,i,j}, t_{2,i}, x, \widehat{\rho}(s)\right) d\widehat{G}_{\widetilde{\mathcal{S}}}(x) \right\rceil \\
&= \left\lceil \frac{1}{q} \widehat{\mathrm{FDP}}_{\lambda,\widetilde{\mathcal{S}}}(t_{1,i,j}, t_{2,i}) \widehat{R}(t_{1,i,j}, t_{2,i}) \right\rceil \\
&:= R_{\mathrm{req}},
\end{aligned}$$

where $\lceil \cdot \rceil$ denotes the ceiling function. When there is no tie, increasing j by k brings k more rejections. Thus we must have

$$j' - j \geq \max\left\{1, R_{\mathrm{req}} - \widehat{R}(t_{1,i,j}, t_{2,i})\right\}. \tag{S.60}$$

This observation is used to prune the candidate set, which improves efficiency. Combining the insights from the above discussions, we propose Algorithm 1 of the main paper, the fast searching algorithm.

Step 4.(Optional) We notice that iterating i from 1 to $\breve{m}$ in the outer loop is overly

time-consuming. To address this, we set $m_{\text{stop}} \geq 0$ and terminate our searching procedure if we encounter an $i^\star$ such that $\widehat{\text{FDP}}_{\lambda,\tilde{\mathcal{S}}}(t_{1,i^\star,j}, t_{2,i^\star}) \leq q$ for some $j \in [m_i^\star]$ and $\widehat{\text{FDP}}_{\lambda,\tilde{\mathcal{S}}}(t_{1,i^\star+i,j}, t_{2,i^\star+i}) > q$ for all $j \in [m_{i^\star+i}]$ and $i = 1, \cdots, m_{\text{stop}}$. Although this approach limits iterations in the outer loop and lacks a theoretical guarantee of identifying the optimal cutoff, forthcoming numerical results illustrate that the cutoffs identified after Step 3 and Step 4 are always the same.

Remark S.5. Algorithm 1 of the main paper can be modified to find the optimal cutoffs for the 2d procedure coupled with various weighted BH procedures (wBH) as discussed in Section 2.7 of the main paper. Recall that the rejection rule for the hypothesis at location s is given by $p_1(s) \leq \min\{\tau, w(s)t_1\}$ and $p_2(s) \leq \min\{\tau, w(s)t_2\}$ (or equivalently $p_1(s) \leq \tau$, $p_2(s) \leq \tau$, $p_1(s)/w(s) \leq t_1$, and $p_2(s)/w(s) \leq t_2$), where $p_j(s) = 1 - \Phi(\widehat{T}_j(s))$ for $j = 1, 2$. Thus the set of all candidate cutoff values is

$$\Big\{\big(p_1(s)/w(s), p_2(s)/w(s)\big) : p_1(s) \leq \tau, p_2(s) \leq \tau, s \in \mathcal{S}\Big\}.$$

Algorithm 1 of the main paper can be modified to find the optimal thresholds (t_1, t_2) for the weighted p-values.

Remark S.6. When there exist ties among $\{\widehat{T}_j(s) : s \in \mathcal{S}\}$ for $j = 1, 2$, Algorithm 1 of the main paper still manages to find the target threshold. The key here is to argue that line 11 of Algorithm 1 of the main paper does not skip any cutoff whose rejection number is no less than R_{req}. To see this, suppose

$$t_{1,i,1} = \cdots = t_{1,i,j_1} > t_{1,i,j_1+1} = \cdots = t_{1,i,j_2} > \cdots > t_{1,i,j_{m_i'-1}+1} = \cdots = t_{1,i,j_{m_i'}},$$

where $j_0 = 0$ and $j_{m_i'} = m_i$. Obviously, the rejection numbers of the nearby cutoffs satisfy

that $\widehat{R}(t_{1,i,j}, t_{2,i}) = \widehat{R}(t_{1,i,j_k}, t_{2,i})$ for $j_{k-1} < j \leq j_k$ and $\widehat{R}(t_{1,i,j_k}, t_{2,i}) - \widehat{R}(t_{1,i,j_{k-1}}, t_{2,i}) = j_k - j_{k-1}$ for all $0 < k \leq m'_i$. For non-nearby cutoffs, suppose they locate between $j_{k-1} < j = j_k - l \leq j_k$ and $j_{k'-1} < j' = j_{k'} - l' \leq j_{k'}$ for some $j < j'$ and $0 < k, k' \leq m'_i$. Then,

$$\begin{aligned}\widehat{R}(t_{1,i,j'}, t_{2,i}) - \widehat{R}(t_{1,i,j}, t_{2,i}) &= \widehat{R}(t_{1,i,j_{k'}}, t_{2,i}) - \widehat{R}(t_{1,i,j_k}, t_{2,i}) \\ &= \sum_{l=k}^{k'-1} (j_{l+1} - j_l) \\ &= j_{k'} - j_k.\end{aligned}$$

We now show that the rejection numbers of the skipped cutoffs are no more than R_{req}. Start with an arbitrary cutoff $(t_{1,i,j}, t_{2,i})$ whose rejection number is less than R_{req}, i.e., $\widehat{R}(t_{1,i,j}, t_{2,i}) < R_{\text{req}}$. Line 11 in Algorithm 1 suggests $j' = j + R_{\text{req}} - \widehat{R}(t_{1,i,j}, t_{2,i})$. As $j = j_k - l$, $j' = j_{k'} - l'$ and $\widehat{R}(t_{1,i,j'}, t_{2,i}) - \widehat{R}(t_{1,i,j}, t_{2,i}) = j_{k'} - j_k$, we have

$$\widehat{R}(t_{1,i,j'}, t_{2,i}) - \widehat{R}(t_{1,i,j}, t_{2,i}) = j_{k'} - j_k = l' - l + R_{\text{req}} - \widehat{R}(t_{1,i,j}, t_{2,i}),$$

which implies that

$$\widehat{R}(t_{1,i,j'}, t_{2,i}) = R_{\text{req}} + l' - l.$$

We thus only need to verify that the rejection number of $(t_{1,i,j_{k'-1}}, t_{2,i})$ is less than R_{req}. Indeed, when $l' \leq l$,

$$\widehat{R}(t_{1,i,j_{k'-1}}, t_{2,i}) < \widehat{R}(t_{1,i,j'}, t_{2,i}) \leq R_{\text{req}}.$$

When $l < l'$,

$$\widehat{R}(t_{1,i,j_{k'-1}}, t_{2,i}) = R_{\text{req}} + l' - l - (j_{k'} - j_{k'-1}) < R_{\text{req}} - l \leq R_{\text{req}}.$$

In both cases, the skipped cutoffs induce no more than R_{req} rejections.

We briefly analyze the computational complexity of our fast searching algorithm. The computational complexity of estimating the FDP in (2.7) of the main paper for a single cutoff is $O(m)$ because the number of supporting points of $\widehat{G}_{\tilde{\mathcal{S}}}(x)$ is fixed in practice. The total computational complexity primarily depends on the number of candidate cutoffs required for the FDP estimation. Considering the stochastic nature of primary and auxiliary statistics, we numerically investigated the relationship between the number of candidate cutoffs and the number of locations. This analysis was conducted within the framework of Setup I in Section 5 of the main paper, focusing on scenarios with medium signal strength, medium correlation, and $\gamma = 2$. We varied the location size m from 100 to 2000 and defined the associated domain as $\mathcal{S} = [0, m/30]$.

Figure S.26 displays the number of candidate cutoffs for different searching strategies: the naive grid search (S0, m^2), after Step 1 (S1, $|\mathcal{T}'|$); after Step 2 (S2, $|\mathcal{T}''|$); after Step 3 (S3), and after Step 4 (S4). The number of candidate cutoffs for strategies S0, S1, and S2 appeared to scale quadratically with the number of locations. The difference between S0 and S1 highlighted the efficiency gains from implementing Step 1. In contrast, the nearly identical performances of S1 and S2 suggested that Step 2 did not significantly enhance computational speed. However, for strategies S3 and S4, the number of candidate cutoffs increased linearly with the number of locations, indicating substantial computational acceleration due to Step 3. Finally, the cutoffs determined after executing Steps 1–3 were identical to those obtained after completing Steps 1–4, which showcases that Step 4 could accelerate searching without sacrificing accuracy. To sum up, Step 3 in our fast searching algorithm can drastically reduce the computational complexity and be further improved by Step 4.

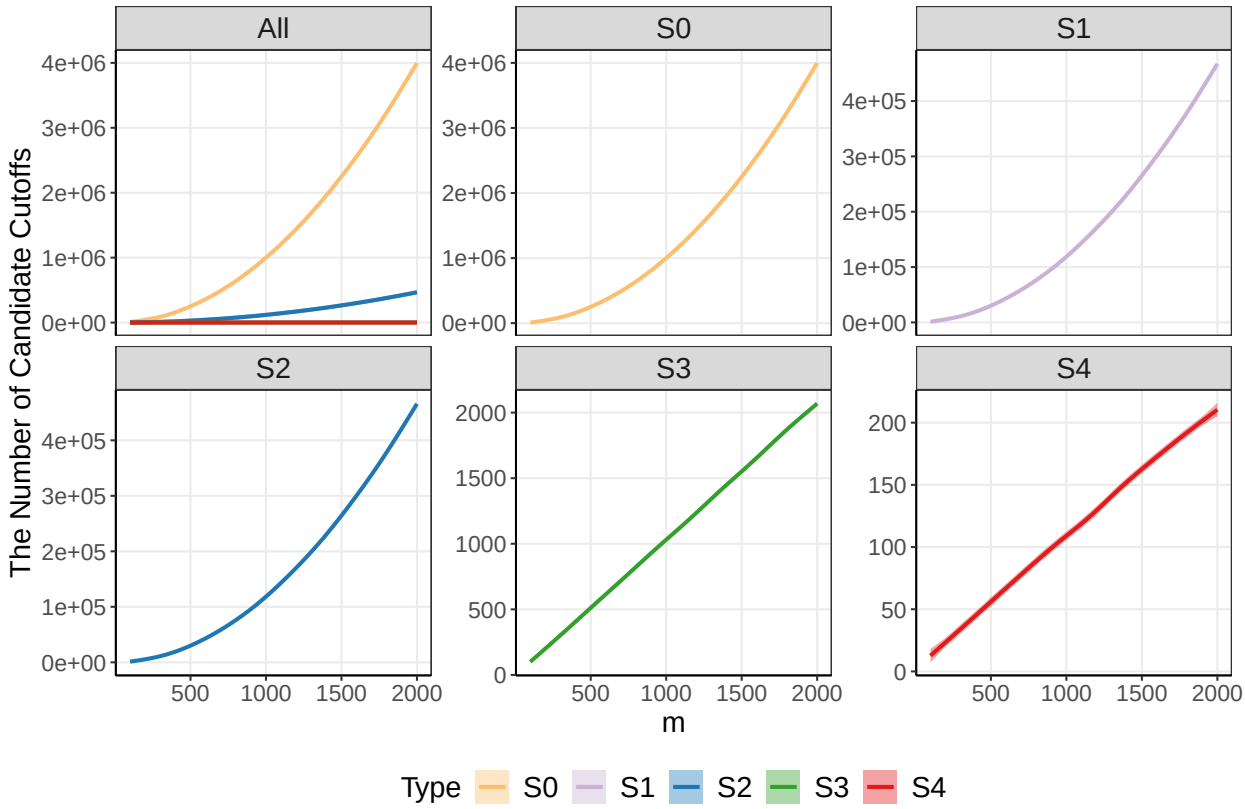


Figure S.26: The number of candidate cutoffs for Setup I with the medium signal, medium dependency, $\gamma = 2$, and m locations. Different colors represent the number of candidate cutoffs for the naive grid search (S0, m^2), after Step 1 (S1, $|\mathcal{T}'|$); after Step 2 (S2, $|\mathcal{T}''|$); after Step 3 (S3), and after Step 4 (S4).

References

Jiang, W. and Zhang, C.-H. (2009), "General maximum likelihood empirical Bayes estimation of normal means," *Ann. Stat.*, 37, 1647–1684.

Sun, W., Reich, B. J., Cai, T. T., Guindani, M., and Schwartzman, A. (2015), "False discovery control in large-scale spatial multiple testing," *J. R. Stat. Soc., Ser. B, Stat. Methodol.*, 77, 59–83.

Wainwright, M. J. (2019), *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge: Cambridge University Press.

Yi, S., Zhang, X., Yang, L., Huang, J., Liu, Y., Wang, C., Schaid, D. J., and Chen, J. (2021), "2dFDR: A new approach to confounder adjustment substantially increases detection power in omics association studies," *Genome Biol.*, 22, 208.

Zhang, C.-H. (2009), "Generalized maximum likelihood estimation of normal mixture densities," *Stat. Sin.*, 19, 1297–1318.