NIASRA

NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

National Institute for Applied Statistics Research Australia

University of Wollongong, Australia

Working Paper

05 -19

False Discovery Rates to Detect Signals from Incomplete Spatially Aggregated Data

Hsin-Cheng Huang, Noel Cressie,
Andrew Zammit-Mangion, and Guowen Huang

National Institute for Applied Statistics Research Australia, University of Wollongong,
Wollongong NSW 2522, Australia Phone +61 2 4221 5076, Fax +61 2 4221 4998.
Email: karink@uow.edu.au

False Discovery Rates to Detect Signals from Incomplete Spatially Aggregated Data

Hsin-Cheng Huang[a,*], Noel Cressie[b], Andrew Zammit-Mangion[b], Guowen Huang[c]

[a]*Institute of Statistical Science, Academia Sinica, Taiwan, R.O.C.*
[b]*National Institute for Applied Statistics Research Australia, University of Wollongong, Australia*
[c]*Department of Statistical Science, University of Toronto, Canada*

Abstract

There are a number of ways to test for the absence/presence of a spatial signal in a completely observed fine-resolution image. One of these is a powerful nonparametric procedure called EFDR (Enhanced False Discovery Rate). A drawback of EFDR is that it requires the data to be defined on small areas (pixels) on a rectangular spatial domain. Here, we augment EFDR to the case when data are irregular and possibly incomplete by using conditional simulation. Specifically, we condition on the available data and simulate the full rectangular image at its finest resolution many times (M, say). EFDR is applied to each of these simulations resulting in M statistically dependent p-values. We test the original null hypothesis of no signal by combining these p-values using copulas and a composite likelihood. If the null hypothesis of no signal is rejected, we then estimate the spatial signal. A simulation study and an application to temperature change in the Asia-Pacific are given to demonstrate the effectiveness of our proposed procedure.

Keywords: Conditional simulation, copula, EFDR, hypothesis testing, small area data, wavelets

*Corresponding author
Email addresses: hchuang@stat.sinica.edu.tw (Hsin-Cheng Huang), ncressie@uow.edu.au (Noel Cressie), azm@uow.edu.au (Andrew Zammit-Mangion), hgw0610209@gmail.com (Guowen Huang)

1. Introduction

Spatial statistical data have been classified as geostatistical, lattice, or point pattern (Cressie, 1993). Here, our interest is in detecting a spatial signal from irregular lattice data, sometimes called small area data, which we consider to be the result of aggregation of pixel values of a fine-resolution image where it is possible that not all these values are included in the aggregation. We emphasize that the fine-resolution values are not observed, only the irregular lattice data are. We formalize this below.

Consider a spatial Gaussian process, $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in D \subset \mathbb{R}^d\}$, defined on a regular lattice D whose locations are in d-dimensional Euclidean space $\mathbb{R}^d$ and generated by

$$Z(\boldsymbol{s}) = \mu(\boldsymbol{s}) + \delta(\boldsymbol{s}); \quad \boldsymbol{s} \in D, \tag{1}$$

where $\mu(\cdot)$ is a deterministic mean function, $\delta(\cdot)$ is a zero-mean stationary Gaussian process with a covariance function, $C(\boldsymbol{u}) \equiv \text{cov}(\delta(\boldsymbol{s}), \delta(\boldsymbol{s}+\boldsymbol{u})) = \text{cov}(Z(\boldsymbol{s}), Z(\boldsymbol{s}+\boldsymbol{u}))$. In what follows, we consider the two-dimensional Euclidean space where $d = 2$, although our approach is general and applies to any $d \in \{1, 2, 3, \dots\}$. We consider first a signal-detection problem in a standard rectangular image. That is, D is a rectangular lattice defined on $n = n_1 \times n_2$ nonoverlapping, fine-resolution areas (or pixels) $\{A_j : j = 1, \dots, n\}$ with area $\Delta_x \Delta_y$ and lattice locations at $\{\boldsymbol{s}_j : j = 1, \dots, n\} = \{(i_1\Delta_x, i_2\Delta_y) : i_1 = 1, \dots, n_1;\, i_2 = 1, \dots, n_2\}$. Then the j-th pixel value is $Z(\boldsymbol{s}_j)$, and the fine-resolution image values are denoted as $\boldsymbol{Z} \equiv (Z(\boldsymbol{s}_1), \dots, Z(\boldsymbol{s}_n))'$. If $\boldsymbol{Z}$ were observed, this would result in regular-lattice data. In a geostatistical context, where there is an underlying continuously indexed spatial process, the $\{A_j : j = 1, \dots, n\}$ have been called Basic Areal Units (BAUs); see Nguyen *et al.* (2012).

Since Gaussianity is preserved following linear operations,

$$\boldsymbol{Z} \sim \text{Gau}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{2}$$

where "Gau" here denotes an n-variate Gaussian distribution, $\boldsymbol{\mu} \equiv (\mu(\boldsymbol{s}_1), \dots, \mu(\boldsymbol{s}_n))' \equiv (\mu_1, \dots, \mu_n)'$, $\boldsymbol{\Sigma} \equiv \text{cov}(\boldsymbol{Z}) = \text{cov}(\boldsymbol{\delta})$, and the (i, j)-th element of $\boldsymbol{\Sigma}$ is $C(\boldsymbol{s}_i, \boldsymbol{s}_j)$, which is

generally non-zero. An equivalent way to write (2) is:

$$\boldsymbol{Z} = \boldsymbol{\mu} + \boldsymbol{\delta}, \tag{3}$$

where $\boldsymbol{\delta} \sim \text{Gau}(\mathbf{0}, \boldsymbol{\Sigma})$.

We are interested in detecting if there is a spatial signal in the image's mean vector $\boldsymbol{\mu}$ in the presence of the correlated noise $\delta(\cdot)$ with spatial covariance function $C(\cdot)$. For example, suppose we are comparing two noisy images and want a way to declare whether the images are different or not. This problem can be formulated in terms of a hypothesis test, where $H_0 : \boldsymbol{\mu} = \mathbf{0}$ versus $H_1 : \boldsymbol{\mu} \neq \mathbf{0}$ is tested and the regular-lattice data $\boldsymbol{Z}$ is defined to be the pixel-wise difference between the two images.

If the data vector $\boldsymbol{Z}$ at the fine resolution is completely observed, then a powerful non-parametric hypothesis-testing method based on the false discovery rate (FDR), called the enhanced FDR (EFDR) procedure (Shen *et al.*, 2002), can be applied to test H_0: $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ and, if it is rejected, to estimate the spatial signal, $\boldsymbol{\mu}$. Martinez *et al.* (2013) considered a similar testing problem with $\boldsymbol{Z}$ corresponding to a two-dimensional image of a moving-window spectrogram. They utilized a Bayesian approach and generated posterior samples to control the Bayesian FDR. Sun et al. (2015) and Risser *et al.* (2019) developed procedures to test H_0: $\mu_j = \mu_0$ (or $\mu_j \leq \mu_0$), for $j = 1, \ldots, n$, which do not require $\boldsymbol{Z}$ to be completely observed. However, the former relies on a parametric model for $\boldsymbol{\mu}$, and it is sensitive to model misspecification. The latter is also a highly parametric approach; it applies a hierarchical Bayesian model and controls the FDR in a Bayesian decision-theoretical framework. The hypothesis-testing method developed by Hering and Genton (2011) tests an average (over D) effect of $\boldsymbol{\mu} - \boldsymbol{\mu}_0$, and it does not rely directly on a Gaussian model for $\delta(\cdot)$. The trade-off taken by the authors to achieve a valid procedure is to integrate out "space." When H_0 is rejected their procedure provides no local information about where the spatial signal is. Gilleland (2013) used it to test competing weather forecasts, and provides software for it in the R package "SpatialVx". Yun *et al.* (2018) considered testing the equality of the

spatial means (or spatial covariances) between two spatio-temporal random fields. While nonparametric in nature, their approach requires p-values to be available at individual locations based on data observed at multiple time points. None of the papers reviewed above are able to address the change-of-support problem that is central to our research, and which we describe in the rest of this section.

We consider the general problem where possibly coarser-resolution irregular-lattice data are observed:

$$Z(B_k) = \frac{\sum_{\boldsymbol{s}} Z(\boldsymbol{s}) I(\boldsymbol{s} \in D \cap B_k)}{\sum_{\boldsymbol{s}} I(\boldsymbol{s} \in D \cap B_k)}; \quad k = 1, \ldots, K, \tag{4}$$

where B_k is made up of one or more pixels $\{A_j\}$, and we wish to make inference on the spatial signal $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_n)'$ at the finest resolution. In what follows, we allow for a general type of coarsening where there might be some overlap of the $\{B_k\}$, or where there is no coarsening but not every pixel in $\{A_j : j = 1, \ldots, n\}$ is included. That is, our approach can handle situations where $B_k \cap B_\ell \neq \emptyset$ for some $k \neq \ell$ and $\cup_{k=1}^{K} B_k \subsetneqq D$.

Because aggregation is a linear operator, the aggregated data vector $\tilde{\boldsymbol{Z}} \equiv (Z(B_1), \ldots, Z(B_K))'$ can be written as:

$$\tilde{\boldsymbol{Z}} = \boldsymbol{H}\boldsymbol{Z} \sim \text{Gau}(\boldsymbol{H}\boldsymbol{\mu}, \boldsymbol{H}\boldsymbol{\Sigma}\boldsymbol{H}'), \tag{5}$$

for some known $K \times n$ matrix $\boldsymbol{H}$ that represents the spatial averaging in (4). Note that K, the dimension of $\tilde{\boldsymbol{Z}}$, is usually smaller than n, but in the case of overlapping $\{B_k\}$, it could be larger than n. Importantly, the mean in (5) is $\boldsymbol{H}\boldsymbol{\mu}$, although we still wish to make inference on $\boldsymbol{\mu}$ at the finest resolution. Specifically, we wish to test H_0 and, in the case of its rejection, we wish to estimate $\boldsymbol{\mu}$ based on the available data $\tilde{\boldsymbol{Z}}$. Examples of $\tilde{\boldsymbol{Z}}$ in (5) are many: the $\{B_k\}$ that define $\tilde{\boldsymbol{Z}}$ might correspond to provinces/states in a country, or counties in a state, or blocks of an image at a coarser resolution than the image's native resolution, or where data for some areas/pixels are missing (e.g., an image of Earth's surface partially obscured by cloud).

The aggregation matrix $\boldsymbol{H}$ can also be written as $\boldsymbol{H} = \boldsymbol{\Lambda}^{-1}\boldsymbol{H}^*$, where $\boldsymbol{H}^*$ is a $K \times n$ incidence matrix that describes the aggregation relationship between $\boldsymbol{Z}$ and $\tilde{\boldsymbol{Z}}$, and $\boldsymbol{\Lambda}$ is a $K \times K$ diagonal matrix with its k-th diagonal element equal to the number of pixels in B_k, for $k = 1, \ldots, K$. If the original image $\{A_j\}$ has some pixels masked out, then those original pixels that are observed can be represented as $\{B_1, \ldots, B_K\}$ made up of $K < n$ distinct pixels from $\{A_j\}$. In that case, $\tilde{\boldsymbol{Z}}$ is a sub-vector of $\boldsymbol{Z}$, $\boldsymbol{H}$ is a sub-matrix of the n-dimensional identity matrix, and $\boldsymbol{\Lambda}$ is the $K \times K$ identity matrix.

In Section 2, we introduce our proposed signal-detection methodology that includes conditional simulation, EFDR, and combining dependent p-values to test for spatial signal. Section 3 provides simulation results that demonstrate the validity and efficiency of our solution to the problem of signal detection from incomplete spatially aggregated data. Section 4 applies our procedure to a temperature dataset over the Asia-Pacific region generated by a climate model from the National Center for Atmospheric Research (NCAR), where we consider inference for a spatial signal in the presence of aggregation and incompleteness. Finally, discussion and conclusions are given in Section 5.

2. Inferring spatial signal from data on an irregular lattice

When the full data, $\boldsymbol{Z}$, in (2) are available, we can apply the EFDR procedure (Shen *et al.*, 2002) to test H_0: $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ versus H_1: $\boldsymbol{\mu} \neq \boldsymbol{\mu}_0$. Since $\boldsymbol{\mu}_0$ is specified, without loss of generality we can assume that the null hypothesis is $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$. Then the EFDR procedure is performed in four steps.

First, we transform $\boldsymbol{Z}$ into a vector of wavelet coefficients by applying an orthogonal discrete wavelet transform,

$$\boldsymbol{\nu} = \mathcal{W}\boldsymbol{Z} = \mathcal{W}\boldsymbol{\mu} + \mathcal{W}\boldsymbol{\delta},$$

where $\mathcal{W}$ is a known $n \times n$ orthogonal discrete-wavelet-transform (DWT) matrix (see Daubechies, 1992), and $\boldsymbol{\delta}$ is given in (3). The wavelet coefficients of the noise, $\mathcal{W}\boldsymbol{\delta}$, can be written as

$$\mathcal{W}\boldsymbol{\delta} = (\boldsymbol{d}'_{-1}, \ldots, \boldsymbol{d}'_{-J}, \boldsymbol{c}'_{-J})', \tag{6}$$

where for $j = -1, \ldots, -J$, the wavelet coefficients at the j-th scale are

$$\boldsymbol{d}_j \equiv \left((\boldsymbol{d}_j^{(1)})', (\boldsymbol{d}_j^{(2)})', (\boldsymbol{d}_j^{(3)})'\right)', \quad \text{and} \quad \boldsymbol{d}_j^{(m)} \equiv \left(d_{j,k_1,k_2}^{(m)} : k_1 = 1, \ldots, n_1 2^j,\, k_2 = 1, \ldots, n_2 2^j\right)',$$

n_1 and n_2 are powers of two, and $\log_2(\min(n_1, n_2)) \geq J$. Each component, $m = 1, 2, 3$, corresponds to the horizontal, vertical, and diagonal spatial orientations, respectively; and $\boldsymbol{c}_{-J} \equiv \{c_{-J,k_1,k_2} : k_1 = 1, \ldots, n_1 2^{-J},\, k_2 = 1, \ldots, n_2 2^{-J}\}$ comprise the scaling-function coefficients. In "wavelet space," the signal can be identified more easily, since typically it has a sparse wavelet representation (i.e., only a few components of $\mathcal{W}\boldsymbol{\mu}$ are nonzero), and the error has been de-correlated.

Second, by utilizing this property that the elements of $\mathcal{W}\boldsymbol{\delta}$ tend to be uncorrelated, and that they have a homogeneous variance within each wavelet scale/orientation (Shen *et al.*, 2002), we standardize $\boldsymbol{\nu}$ by scale/orientation and assume that under $H_0 : \boldsymbol{\mu} = \mathbf{0}$, the resulting standardized coefficients are independent and identically distributed Gau$(0, 1)$ random variables. Third, we increase both statistical and computational efficiency by reducing the number of tests on the wavelet coefficients. This is achieved by ordering all the individual wavelet-coefficient hypotheses using the network structure of wavelets and selecting the number of hypotheses based on the generalized-degrees-of-freedom criterion of Ye (1998).

Fourth, we apply the false-discovery-rate (FDR) procedure of Benjamini and Hochberg (1995) to the selected wavelet coefficients to obtain a p-value for the hypothesis test of H_0 and an estimate of the spatial signal $\boldsymbol{\mu}$ through the inverse DWT. The use of p-values is sometimes controversial, as discussed in many articles in a 2019 Special Issue of The American Statistician (Wasserstein *et al.*, 2019). However, p-values have a place in statistical inference and FDR is one example of their appropriate use (Shen *et al.*, 2002; Pavlicová *et al.*, 2008).

2.1. Estimating spatial dependence in incomplete spatially aggregated data

Our goal is to make inference on $\boldsymbol{\mu}$ in (2), but we only observe $\tilde{\boldsymbol{Z}}$ in (5), not $\boldsymbol{Z}$. Our methodology is based on conditionally simulating $\boldsymbol{Z}$, conditional on $\tilde{\boldsymbol{Z}}$, which takes into account the spatial dependence given by $\boldsymbol{\Sigma} = \text{var}(\boldsymbol{Z})$ in (2). This approach is very similar

to multiple imputation that has been developed in a non-spatial context (Little and Rubin, 2002). Henceforth, we write $\boldsymbol{\Sigma}$ as $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a parameter vector to be estimated. To obtain $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, we start with the null hypothesis H_0 and a model in the wavelet domain for $\mathcal{W}\boldsymbol{\delta}$ defined by (6). As in Shen *et al.* (2002), the wavelet coefficients, $\mathcal{W}\boldsymbol{Z}$, are modeled independently as

$$\boldsymbol{d}_{-j}^{(m)} \sim \text{Gau}(\boldsymbol{0}, \theta_{3(j-1)+m}\boldsymbol{I});\ m = 1, 2, 3;\ j = 1, \ldots, J, \quad \text{and} \quad \boldsymbol{c}_{-J} \sim \text{Gau}(\boldsymbol{0}, \theta_{3J+1}\boldsymbol{I}), \tag{7}$$

and from (6), $\boldsymbol{\delta} = \mathcal{W}^{-1}(\boldsymbol{d}'_{-1}, \ldots, \boldsymbol{d}'_{-J}, \boldsymbol{c}'_{-J})'$. Hence, from (7), the vector of the variance components $\boldsymbol{\theta} \equiv (\theta_1, \ldots, \theta_{3J+1})'$ parameterizes $\text{var}(\boldsymbol{\delta}) = \boldsymbol{\Sigma}(\boldsymbol{\theta})$.

The covariance matrix in the wavelet domain, $\boldsymbol{V}(\boldsymbol{\theta}) \equiv \text{var}(\mathcal{W}\boldsymbol{\delta}) = \mathcal{W}\boldsymbol{\Sigma}(\boldsymbol{\theta})\mathcal{W}'$ is a diagonal matrix whose k-th diagonal block is given by $\theta_k\boldsymbol{I}$, for $k = 1, \ldots, (3J+1)$. We estimate $\boldsymbol{\theta}$ through

$$\hat{\boldsymbol{\theta}} \equiv \underset{\boldsymbol{\theta}\in(0,\infty)^{3J+1}}{\arg\min} \|\mathcal{W}'\boldsymbol{V}(\boldsymbol{\theta})\mathcal{W} - \hat{\boldsymbol{\Sigma}}\|_F = \underset{\boldsymbol{\theta}\in(0,\infty)^{3J+1}}{\arg\min} \|\boldsymbol{V}(\boldsymbol{\theta}) - \mathcal{W}\hat{\boldsymbol{\Sigma}}\mathcal{W}'\|_F\,, \tag{8}$$

since $\mathcal{W}$ is orthogonal. In (8), $\|\cdot\|_F$ is the Frobenius norm and $\hat{\boldsymbol{\Sigma}}$ is a smooth empirical estimator, here the maximum likelihood (ML) estimator of $\boldsymbol{\Sigma}$ under a parametric covariance model, $C(\cdot)$, and $\boldsymbol{\mu} = \boldsymbol{0}$. Note that it is generally not possible to define the usual method-of-moments estimator of $\boldsymbol{\Sigma}$ from the incomplete spatially aggregated data $\tilde{\boldsymbol{Z}}$, for which there are no replications available.

As an example, consider the Matérn covariance model for $\text{cov}(Z(\boldsymbol{s}), Z(\boldsymbol{s}^*))$ given by,

$$C(\boldsymbol{u}) = \frac{\tau^2}{2^{\nu-1}\Gamma(\nu)}\left(\frac{\sqrt{2\nu}}{\phi}\|\boldsymbol{u}\|\right)^{\nu} K_\nu\left(\frac{\sqrt{2\nu}}{\phi}\|\boldsymbol{u}\|\right); \quad \boldsymbol{u} = \boldsymbol{s}^* - \boldsymbol{s} \in \mathbb{R}^2, \tag{9}$$

where $K_\nu(\cdot)$ is the modified Bessel function of the second kind of order ν, and $\boldsymbol{\gamma} \equiv (\phi, \nu)'$ consists of a spatial-scale parameter ϕ and a smoothness parameter ν. Under H_0, the ML estimator of $\boldsymbol{\gamma}$ can be obtained from the data $\tilde{\boldsymbol{Z}}$ by minimizing the negative log profile likelihood, as follows:

$$\hat{\boldsymbol{\gamma}} \equiv \underset{\boldsymbol{\gamma}}{\arg\min}\left\{\frac{1}{2}\log|\boldsymbol{H}\boldsymbol{\Omega}(\boldsymbol{\gamma})\boldsymbol{H}'| + \frac{K}{2}\log\left\{\tilde{\boldsymbol{Z}}'(\boldsymbol{H}\boldsymbol{\Omega}(\boldsymbol{\gamma})\boldsymbol{H}')^{-1}\tilde{\boldsymbol{Z}}\right\} + \text{constant}\right\},$$

where $\boldsymbol{\Omega}(\boldsymbol{\gamma})$ is an $n \times n$ correlation matrix whose (i,j)-th entry is $C(\boldsymbol{s}_i - \boldsymbol{s}_j)/\tau^2$. Then the ML estimator $\hat{\tau}^2$ of the stationary variance τ^2 is:

$$\hat{\tau}^2 \equiv \frac{1}{K}\tilde{\boldsymbol{Z}}'(\boldsymbol{H}\boldsymbol{\Omega}(\hat{\boldsymbol{\gamma}})\boldsymbol{H}')^{-1}\tilde{\boldsymbol{Z}}.$$

Consequently, a smooth empirical estimator of $\boldsymbol{\Sigma}$ for use in (8) is given by $\hat{\boldsymbol{\Sigma}} \equiv \hat{\tau}^2\boldsymbol{\Omega}(\hat{\boldsymbol{\gamma}})$.

Let $\mathcal{W}_k$ be a sub-matrix of $\mathcal{W}$, consisting of the rows corresponding to the k-th wavelet component. Then $\mathcal{W}' = (\mathcal{W}_1', \ldots, \mathcal{W}_{3J+1}')$. It follows from (8) that $\hat{\boldsymbol{\theta}} \equiv \big(\hat{\theta}_1, \ldots, \hat{\theta}_{3J+1}\big)'$ is given by,

$$\hat{\theta}_k = \frac{\hat{\tau}^2}{n_k}\mathrm{tr}\big(\mathcal{W}_k\boldsymbol{\Omega}(\hat{\boldsymbol{\gamma}})\mathcal{W}_k'\big); \quad k = 1, \ldots, 3J+1, \tag{10}$$

where n_k is the number of rows of $\mathcal{W}_k$, and $\mathrm{tr}(\boldsymbol{A})$ denotes the trace of a square matrix $\boldsymbol{A}$.

2.2. Testing for signal in conditionally simulated images

Once the spatial covariance parameters have been estimated, we are ready to apply our new methodology, which can be summarized in the following steps. First, we simulate M times the n-dimensional vector $\boldsymbol{Z}$ *conditional on the data* $\tilde{\boldsymbol{Z}}$ (with $\hat{\boldsymbol{\theta}}$ substituted in for $\boldsymbol{\theta}$), via

$$\boldsymbol{Z}\big|\tilde{\boldsymbol{Z}} \sim \mathrm{Gau}\big(\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})\boldsymbol{H}'\big(\boldsymbol{H}\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})\boldsymbol{H}'\big)^{-1}\tilde{\boldsymbol{Z}},\ \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})\boldsymbol{H}'\big(\boldsymbol{H}\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})\boldsymbol{H}'\big)^{-1}\boldsymbol{H}\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})\big), \tag{11}$$

resulting in the M outcomes, $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M$. Then we apply the EFDR procedure to each $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M$ separately, from which we obtain corresponding p-values, $p_1, \ldots, p_M$, and estimates, $\hat{\boldsymbol{\mu}}_1, \ldots, \hat{\boldsymbol{\mu}}_M$, of $\boldsymbol{\mu}$.

Next, we combine $\{p_i : i = 1, \ldots, M\}$ into a single p-value, although doing so is not straightforward since the M p-values are statistically dependent, each being a function of $\tilde{\boldsymbol{Z}}$. Even if they were independent, the naïve approach of taking the sample mean of $\{p_i : i = 1, \ldots, M\}$ tends to produce a p-value that is too large (Brown, 1975). For independent $\{p_i\}$, Fisher (1925) proposed using the test statistic,

$$T \equiv -2\sum_{i=1}^{M}\log(p_i), \tag{12}$$

to test H_0. Brown (1975) used the same test statistic for dependent $\{p_i\}$ obtained from multiple one-sided location tests in a multivariate Gaussian setting with known covariance matrix. In the next subsection, we develop new distribution theory for T to account for the special dependence between the $\{p_i\}$ that is a consequence of the conditional simulation. A final p-value is then obtained with regard to T and its distribution, from which H_0 is tested. In the case of H_0 being rejected, an estimate of the signal $\boldsymbol{\mu}$ is given in Section 2.5.

2.3. Distribution theory for combining dependent p-values

From (12), we write $T = \sum_{i=1}^{M} t_i$, where $t_i \equiv -2 \log p_i$, and we use the flexible Gamma family of distributions to approximate the distribution of T. That is, we fit T to a $\Gamma(a, b)$ distribution whose probability density function is $f(x) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx)$, for $x \geq 0$, and 0 for $x < 0$, where our proposed methodology determines a and b. Under H_0, the marginal distribution of t_i is $\Gamma(1, 1/2)$, which is a chi-squared distribution on 2 degrees of freedom (e.g., Littell and Folks, 1971). If $\{t_i\}$ are independent, then $T \sim \Gamma(M, 1/2)$, so that $a = M$ and $b = 1/2$ (resulting in Fisher's combined probability test). In our case, $\{t_i\}$ are not independent, which leads to the need for estimates of the Gamma parameters a and b.

The dependence in $\{p_i\}$ is caused by dependence between the replicates from the conditional simulation: Each t_i depends on the original K-dimensional data vector $\tilde{\boldsymbol{Z}}$, and hence they are not independent. However, they are exchangeable (e.g., Section 3.17 of Spiegelhalter *et al.*, 2004), and hence $\text{cov}(t_i, t_j) = \sigma^2 \rho$, for $i \neq j$. We call ρ the level of exchangeability, and

$$\boldsymbol{U} \equiv \text{cov}((t_1, \ldots, t_M)') = \sigma^2 \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho & \cdots & \rho & 1 \end{pmatrix}, \tag{13}$$

which is a matrix of constant intra-class correlations. Note that $\boldsymbol{U}\mathbf{1} = \sigma^2(1 + (M-1)\rho)\mathbf{1}$, which implies that an eigenvalue of $\boldsymbol{U}$ is $\sigma^2(1 + (M-1)\rho)$. The covariance matrix $\boldsymbol{U}$ is known to have only two eigenvalues with the second one being $\sigma^2(1-\rho)$ (e.g., see Example 3.9 of Schott, 2017). Since $\boldsymbol{U}$ is nonnegative-definite, the eigenvalues must be nonnegative,

and hence $-1/(M-1) \leq \rho \leq 1$. Since ρ does not depend on M, it follows that

$$0 \leq \rho \leq 1. \tag{14}$$

Recall that $T = \sum_{i=1}^{M} t_i$. Then under H_0, the first moment is $\mathrm{E}(T) = \sum_{i=1}^{M} \mathrm{E}(t_i) = 2M$, and the second central moment is,

$$\mathrm{var}(T) \;=\; \sum_{i=1}^{M} \mathrm{var}(t_i) + \sum_{1 \leq i \neq j \leq M} \mathrm{cov}(t_i, t_j) \;=\; 4M\{1 + (M-1)\rho\},$$

since $\sigma^2 \equiv \mathrm{var}(t_i) = 4$. These relations for $\mathrm{E}(T)$ and $\mathrm{var}(T)$, together with $\mathrm{E}(V) = a/b$ and $\mathrm{var}(V) = a/b^2$, where $V \sim \Gamma(a, b)$, give the following estimating equations for the Gamma parameters a and b:

$$a = (\mathrm{E}(T))^2/\mathrm{var}(T) = M/(1 + (M-1)\rho), \tag{15}$$

$$b = \mathrm{E}(T)/(\mathrm{var}(T)) = 1/\{2(1 + (M-1)\rho)\}. \tag{16}$$

2.4. Estimation of the level of exchangeability, ρ

From (15), (16), and for ρ given, the distribution of T can be fitted to a $\Gamma(a, b)$ distribution. In this subsection, we present a method for estimating ρ based on bivariate copulas. We use copulas because we know that $t_i \geq 0$ has a marginal distribution that is exactly χ^2_2 (an exponential distribution with rate parameter 2) for all $i = 1, \ldots, M$. Therefore, the cumulative distribution function of t_i is $P(t_i \leq x) = F(x) = 1 - \exp(-x/2)$, for $x \geq 0$, and $= 0$ for $x < 0$. For $i \neq j$, we use Gaussian copulas to model the bivariate distribution of (t_i, t_j). That is, (t_i, t_j) is modeled as a bivariate exponential distribution with cumulative distribution function,

$$P(t_i \leq x_1, t_j \leq x_2) \;=\; G(F(x_1), F(x_2); r) \;\equiv\; F(x_1, x_2; r), \tag{17}$$

where for $u_1 = F(x_1)$ and $u_2 = F(x_2)$, $G(u_1, u_2; r)$ is the bivariate Gaussian copula generated by a bivariate standard Gaussian distribution with correlation $r \in [0, 1)$ (Song, 2000).

The probability density function of $G(u_1, u_2; r)$ is given by

$$g(u_1, u_2; r) = \frac{1}{\sqrt{1-r^2}} \exp\left\{-\frac{1}{2}(\Phi^{-1}(u_1), \Phi^{-1}(u_2))\left(\begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}^{-1} - \boldsymbol{I}\right)\begin{pmatrix} \Phi^{-1}(u_1) \\ \Phi^{-1}(u_2) \end{pmatrix}\right\};\ u_1, u_2 \in [0,1],$$

where Φ is the cumulative distribution function of the standard Gau$(0,1)$ distribution. It follows that $\rho = \text{corr}(t_i, t_j) = \rho(r)$ is a function of the Gaussian-based correlation r.

Let $f(x_1, x_2; r)$ be the resulting bivariate exponential probability density function obtained from (17). Since $\text{E}(t_i) = 2$ and $\text{var}(t_i) = 4$, for $i = 1, \ldots, M$, we have

$$\text{corr}(t_i, t_j) = \ \rho(r) = \frac{1}{4}\iint x_1 x_2 f(x_1, x_2; r)\text{d}x_1\text{d}x_2 - 1.$$

In our procedure, we estimate r by maximizing the composite likelihood function,

$$L(r) \ \equiv \prod_{1 \le i < j \le M} f(t_i, t_j; r),$$

where recall that $t_i = -2\log p_i;\ i = 1, \ldots, M$. Denote this estimator by $\hat{r}$, and hence the maximum composite likelihood (MCL) estimator of ρ as $\rho(\hat{r})$. We use simulation to obtain $\rho(\hat{r})$; that is, we sample from (17) with $\hat{r}$ in place of r and, from the simulations, compute the sample correlation, which we denote by $\hat{\rho}$. If $\hat{\rho}$ lies outside the interval $[0,1]$ given by (14), we put it equal to the nearest end point. Then we obtain the following estimates of the Gamma parameters, a and b:

$$\hat{a} = M/(1 + (M-1)\hat{\rho}) \quad \text{and} \quad \hat{b} = 1/\{2(1 + (M-1)\hat{\rho})\}. \tag{18}$$

2.5. A final p-value for testing H_0 and subsequent inference on $\boldsymbol{\mu}$

Recall that our goal is signal detection. From the incomplete spatially aggregated data $\tilde{\boldsymbol{Z}}$, the final p-value of our procedure for testing $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$, depends on a and b through

$$p = 1 - F_{\Gamma(a,b)}(T), \tag{19}$$

where T is given by (12) and $F_{\Gamma(a,b)}$ is the cumulative distribution function of a $\Gamma(a,b)$ random variable. Hence, once a and b are specified or estimated, p in (19) can be easily

obtained from Gamma-distribution tables. For $\hat{a}$ and $\hat{b}$ given by (18), we obtain the final p-value,

$$\hat{p} = 1 - F_{\Gamma(\hat{a},\hat{b})}(T). \tag{20}$$

We call our procedure based on (20), "CPL," which is an abbreviation of "copula."

Other estimates of a and b are possible. For example, there is a simple method-of-moments that can be used and that we now present. Under H_0, $\sum_{i=1}^{M}(t_i - 2)^2/M$ is an unbiased estimator of σ^2, and $\sum_{1\le i<j\le M}(t_i - t_j)^2/\{M(M-1)/2\}$ is an unbiased estimator of $2\sigma^2(1-\rho)$. Then a method-of-moments estimator is:

$$\tilde{\rho} = 1 - \frac{\sum_{1\le i<j\le M}(t_i - t_j)^2/(M-1)}{\sum_{i=1}^{M}(t_i - 2)^2}. \tag{21}$$

Again, if $\hat{\rho}$ lies outside the interval $[0, 1]$, we put it equal to the nearest end point. Upon substituting $\tilde{\rho}$ for ρ in (15) and (16), we obtain

$$\tilde{a} \;=\; M\big/\big(1 + (M-1)\tilde{\rho}\big)\big) \quad \text{and} \quad \tilde{b} \;=\; 1\big/\big\{2(1 + (M-1)\tilde{\rho})\big\}, \tag{22}$$

and the final p-value,

$$\tilde{p} = 1 - F_{\Gamma(\tilde{a},\tilde{b})}(T). \tag{23}$$

We call our procedure based on (23), "MOM," which is an abbreviation of "method-of-moments" and is easier to implement than "CPL," although typically not as statistically efficient.

For a and b specified or estimated, the significance test of H_0 at level α is:

$$\text{reject } H_0 \text{ if } T > F^{-1}_{\Gamma(a,b)}(1-\alpha), \tag{24}$$

where T is defined by (12), and α is a pre-specified significance level between 0 and 1 (e.g., $\alpha = 0.05$). If H_0 is rejected, then we estimate $\boldsymbol{\mu}$ with

$$\hat{\boldsymbol{\mu}} \;\equiv\; \sum_{i=1}^{M} \hat{\boldsymbol{\mu}}_i / M, \tag{25}$$

where $\{\hat{\boldsymbol{\mu}}_i : i = 1, \ldots, M\}$ are given by the EFDR procedure of Shen *et al.* (2002) applied to each of the M conditional simulations.

In Section 3, we present simulation experiments for inference on the spatial signal $\boldsymbol{\mu}$ obtained from (24). Many of the results for our proposed methodology are based on CPL, where $(a, b) = (\hat{a}, \hat{b})$ given by (18), since we found it more statistically efficient (although not substantially so) than MOM given by (22). In some circumstances the computational simplicity of the estimates given by (22) may be preferred to the more involved estimates given by (18).

2.6. A summary of our proposed procedure for inferring spatial signal

For detecting and estimating pixel-scale signal from incomplete spatially aggregated data, we propose the following six steps:

1. Estimate $\boldsymbol{\theta}$ in $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ by $\hat{\boldsymbol{\theta}}$ through (10), based on the data $\tilde{\boldsymbol{Z}}$ and under H_0: $\boldsymbol{\mu} = \mathbf{0}$.
2. Using $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$, simulate M times the n-dimensional vector $\boldsymbol{Z}$ conditional on $\tilde{\boldsymbol{Z}}$ via (11), and obtain the conditional simulations $\{\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M\}$.
3. Apply EFDR to each of $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M$, from which the p-values $p_1, \ldots, p_M$, and the corresponding estimates $\hat{\boldsymbol{\mu}}_1, \ldots, \hat{\boldsymbol{\mu}}_M$ of $\boldsymbol{\mu}$, are obtained.
4. Estimate ρ, the level of exchangeability in the $\{p_i\}$, from which the Gamma parameters a and b are estimated, for example by (18) or (22).
5. Combine $\{p_1, \ldots, p_M\}$ into the final p-value using (19) and use it in (24) to test the hypothesis $H_0 : \boldsymbol{\mu} = \mathbf{0}$ at the $100\,\alpha\%$ level of significance.
6. If H_0 is not rejected, estimate $\boldsymbol{\mu}$ as $\hat{\boldsymbol{\mu}} = \mathbf{0}$ (i.e., no spatial signal). Otherwise, use (25) to obtain an estimate $\hat{\boldsymbol{\mu}}$ of the spatial signal.

3. Simulation studies

We conducted two simulation studies. In the first study, we investigated the Type-I error rates obtained using the testing rule (24) when the level of exchangeability ρ was estimated using copulas (CPL) and the method-of-moments (MOM). The test of H_0 where

the p-values are combined naïvely through their simple average (NVE), was also considered for comparison. In this initial study, the p-values $\{p_i\}$ came from a two-sided z-test for the mean of a Gaussian distribution with known variance; the purpose of the study is to assess the validity of our proposed procedure in a very simple, non-spatial setting where there is exchangeability. In the second study, we performed three experiments in a factorial design, where our proposed procedure was evaluated under three scenarios involving complete data at different scales of aggregation and incomplete data (missing in a contiguous block and missing at random) at different scales of aggregation.

3.1. Observed Type-I error rates using p-values from correlated z-tests

Let the set $\{x_1, \ldots, x_{100}\}$ be made up of elements x_i distributed independently as $\text{Gau}(\mu, 1)$, for $i = 1, \ldots, 100$. We drew $\{x_1, \ldots, x_{100}\}$ from $\text{Gau}(0, 1)$, and then randomly drew M subsamples of size $N \leq 100$ without replacement from $\{x_1, \ldots, x_{100}\}$. For $k = 1, \ldots, M$, let $\{x^*_{k,1}, \ldots, x^*_{k,N}\}$ be the k-th subsample and $z_k = (1/N) \sum_{i=1}^{N} x^*_{k,i}$ be a statistic for testing the hypotheses, $H_0 : \mu = 0$ versus $H_1 : \mu \neq 0$. Under H_0, it is easy to see that $\text{E}(z_k) = 0$ and $\text{var}(z_k) = 1/N$. However, the $\{z_k\}$ are dependent; indeed, they are exchangeable due to the sampling-without-replacement from $\{x_1, \ldots, x_{100}\}$.

The individual z-test based on the statistic z_k rejects H_0 if

$$p_k \equiv 2(1 - \Phi(\sqrt{N}|z_k|)) < \alpha, \tag{26}$$

where $0 < \alpha < 1$ is a pre-specified significance level. Because $\{z_k : k = 1, \ldots, M\}$ are exchangeable, so too are the p-values $\{p_k : k = 1, \ldots, M\}$. We are interested in knowing how well they can be combined into a single p-value using the naïve procedure of averaging (NVE), our copula-based, composite-likelihood procedure (CPL), and the method-of-moments variant (MOM). That is, we compare

1. NVE: A naïve procedure where the final p-value is $\sum_{k=1}^{M} p_k/M$.
2. CPL: The final p-value is given by (20).
3. MOM: The final p-value is given by (23).

Table 1: Empirical Type-I error rates under different values of (α, N) defined in Section 3.1 based on 50,000 simulation replicates, where values given in parentheses are the Monte Carlo standard errors.

α	N	NVE	CPL	MOM
0.01	80	0.0016 (0.0002)	0.0078 (0.0004)	0.0063 (0.0004)
	85	0.0030 (0.0002)	0.0089 (0.0004)	0.0073 (0.0004)
	90	0.0049 (0.0003)	0.0098 (0.0004)	0.0087 (0.0004)
	95	0.0065 (0.0004)	0.0098 (0.0004)	0.0090 (0.0004)
0.05	80	0.0161 (0.0006)	0.0454 (0.0009)	0.0446 (0.0009)
	85	0.0233 (0.0007)	0.0488 (0.0010)	0.0481 (0.0010)
	90	0.0305 (0.0008)	0.0482 (0.0010)	0.0478 (0.0010)
	95	0.0389 (0.0009)	0.0485 (0.0010)	0.0482 (0.0010)
0.10	80	0.0443 (0.0009)	0.0958 (0.0013)	0.1033 (0.0014)
	85	0.0574 (0.0010)	0.0968 (0.0013)	0.1044 (0.0014)
	90	0.0686 (0.0011)	0.0949 (0.0013)	0.1001 (0.0013)
	95	0.0828 (0.0012)	0.0971 (0.0013)	0.0997 (0.0013)

Note that a Type-I error occurs if, under H_0, the resulting p-value is smaller than α.

In this experimental set-up, the level of exchangeability ρ in the intra-class correlation model (13) is higher when N (i.e., the subsample size) is closer to the full sample size of 100. So we considered only $N \in \{80, 85, 90, 95\}$. In addition, we considered the three significance levels, $\alpha \in \{0.01, 0.05, 0.1\}$, commonly used in practice. The resulting empirical Type-I error rates for the three methods (NVE, CPL, MOM) under the 12 different combinations of N and α, based on 50,000 simulation replicates, are shown in Table 1.

NVE consistently gives Type-I error rates that are too low, which was expected since the sample average of the p-values results in a combined p-value that tends to be too large, causing the null hypothesis to be rejected less often than it should. The effect is less pronounced when the p-values are more correlated; that is, as the size of the subsample N increases, the Type-I error rate of NVE improves. Our proposed procedure, whether it is CPL or MOM, is adaptive to the amount of dependence, and the Type-I error rates are very close to the nominal levels for all cases.

This initial study is encouraging and indicates that our proposal is valid in the presence of dependent p-values. Next, we apply our proposed procedure in the spatial context.

3.2. Factorial experiment for spatial-signal detection

In the second study, we conducted three simulation experiments, where aggregation and "missingness" of pixels as well as signal strength were varied, to evaluate the performance of our proposed procedure summarized in Section 2.6. The "responses" used in the study are the Type-I error rate (i.e., the probability of incorrectly rejecting a true H_0), the power (i.e., the probability of correctly rejecting a false H_0), and the receiver operating characteristic (ROC) curve (i.e., a plot of the power as a function of the Type-I error rate).

In our experimental set-up, we let the finest pixel resolution be of size 64×64. That is, $\boldsymbol{Z}$ in (2) is a vector of length $n = n_1 \times n_2 = 64 \times 64 = 4{,}096$. To check the power of our proposed procedure, we generated data with a signal given by $\mu(\boldsymbol{s}) = h \times I(\boldsymbol{s} \in \Delta_r)$; $\boldsymbol{s} \in D$, where we considered four different $r \times r$ square regions $\Delta_r \subset D$ of width $r \in \{4, 6, 8, 10\}$. Here, all squares were centered at the juncture of the middle four pixels in the 64×64 region. For each Δ_r, we considered six different signal magnitudes $h \in \{0, 1, \ldots, 5\}$, where $h = 0$ corresponds to no signal and is used to compute the Type-I error rate. We generated spatially correlated errors using an exponential covariance function (i.e., the Matérn model of (9) with $\nu = 1/2$), and we checked how the power depends on the degree of spatial correlation by considering $\phi \in \{0, 5, 10\}$. For each setting of r, h, and ϕ, we simulated 400 datasets $\tilde{\boldsymbol{Z}}$, from which we obtained the empirical power and ROC curves.

Throughout the simulation, we chose two wavelet-decomposition levels (i.e., $J = 2$), resulting in seven (i.e., $3J + 1 = 7$) wavelet classes corresponding to different scales and orientations. For each of the 400 simulated datasets, we estimated $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_7)'$ through (10), where $\hat{\tau}^2$ and $\hat{\boldsymbol{\gamma}}$ are the ML estimators based on the exponential covariance model. Using the estimate $\hat{\theta}$ for a given dataset, we generated $M = 100$ conditional simulations through (11). We then used the R package "EFDR" (Zammit-Mangion and Huang, 2015) on each conditionally simulated 64×64 image, using the Daubechies least asymmetric wavelet filter of length 8 (Daubechies, 1992), and we let the number of hypotheses to be tested in the wavelet space be 100. For each dataset, these conditional simulations produced 100 p-values,

which were combined according to the statistic T in (12) and the final p-value given by (19). Then the hypothesis test for H_0 was performed using (24).

We compared the performance of our proposed procedure, CPL and MOM, with the naïve approach, NVE. In addition, we considered an ideal setting (IDL) where it is assumed that all fine-resolution pixels were observed and $\tilde{\boldsymbol{Z}}_{64\times 64} \equiv \boldsymbol{Z}$, so that the EFDR procedure can be directly applied without conditional simulation. Three experiments and an analysis of their responses are now given.

Experiment 1: Complete data at different scales of aggregation

Figure 1(a) shows randomly generated datasets with $h = 0$ (i.e., no signal) and strength of spatial dependence $\phi = 0, 5, 10$, respectively. Note that when ϕ is larger, it is more difficult to separate a signal from the spatially dependent noise, since strong spatial dependence can take on the appearance of a non-zero mean vector of spatially coherent entries. Data $\tilde{\boldsymbol{Z}}$ were generated by aggregating $\boldsymbol{Z}$ into 16×16 and 8×8 regular grid cells, and they are denoted by $\tilde{\boldsymbol{Z}}_{16\times 16}$ and $\tilde{\boldsymbol{Z}}_{8\times 8}$, respectively. Figures 1(b) and 1(c) show the data, $\tilde{\boldsymbol{Z}}_{16\times 16}$ and $\tilde{\boldsymbol{Z}}_{8\times 8}$, respectively, obtained through aggregation of the corresponding images in Figure 1(a). Figures 1(d) and 1(e) show a single conditionally simulated $\boldsymbol{Z}$, conditional on $\tilde{\boldsymbol{Z}}_{16\times 16}$ and $\tilde{\boldsymbol{Z}}_{8\times 8}$, respectively. Although we do not expect to reproduce the $\boldsymbol{Z}$ shown in column (a) exactly, the conditional simulations do produce patterns similar to $\boldsymbol{Z}$. Figure 2 shows data $\boldsymbol{Z}$ generated from (2) with spatial signal $\boldsymbol{\mu}$ corresponding to $(r, h) \in \{(4, 1), (8, 3), (10, 5)\}$ and $\phi = 5$ (i.e., $C(\boldsymbol{u}) = \exp(-\|\boldsymbol{u}\|/5)$). These 64×64 images are then aggregated, resulting in data, $\tilde{\boldsymbol{Z}}_{16\times 16}$ and $\tilde{\boldsymbol{Z}}_{8\times 8}$; an example is shown in Figure 3, in which the right panel of Figure 2 is aggregated. Full sets of plots for all combinations of the factors with $\phi = 5$ are shown in Figures S1–S3 in the Supplementary Material.

Tests were carried out at the usual 5% significance level, and power curves for all values of (r, ϕ) and for all methods (IDL, CPL, MOM, and NVE) are shown in Figure S4 in the Supplementary Material. These curves suggest that our proposed procedure, CPL, is slightly more competitive than MOM and IDL (although not substantially), so in Figure 4 we show

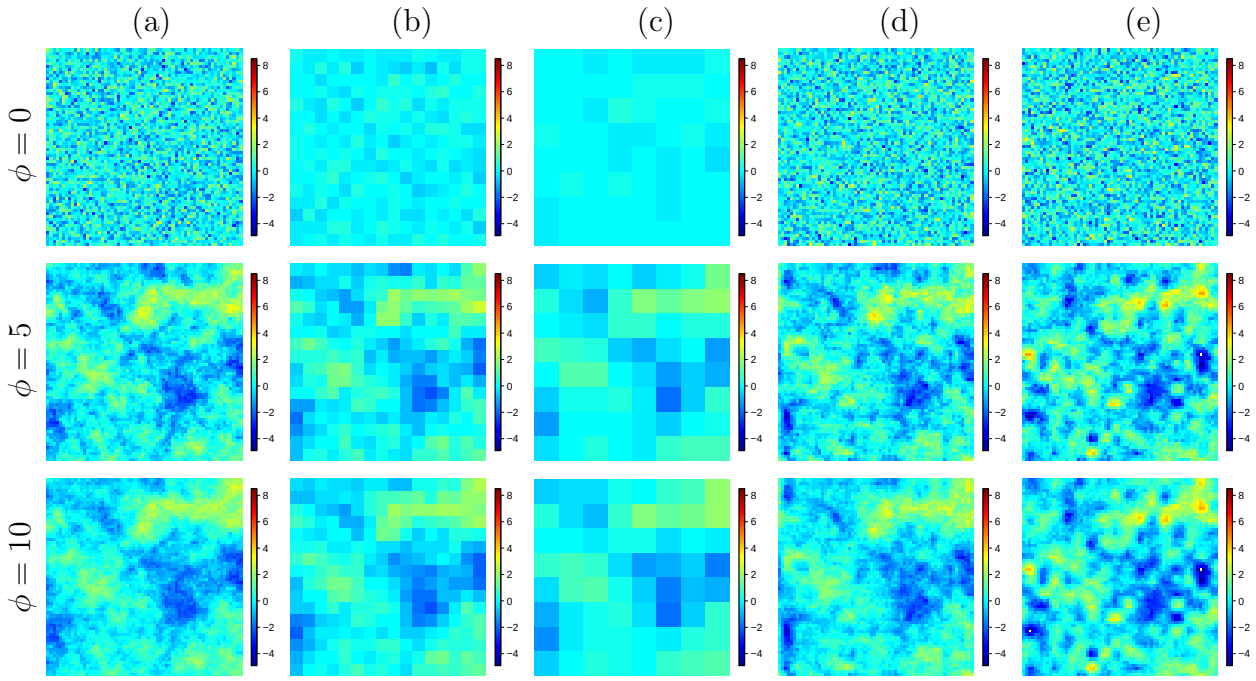

Figure 1: Column (a) Three randomly generated images of $\tilde{\boldsymbol{Z}}_{64\times64} \equiv \boldsymbol{Z}$ corresponding to $\phi \in \{0, 5, 10\}$ (respectively, from top to bottom); Column (b) The images of $\tilde{\boldsymbol{Z}}_{16\times16}$ aggregated from $\tilde{\boldsymbol{Z}}_{64\times64}$ in column (a); Column (c) The images of $\tilde{\boldsymbol{Z}}_{8\times8}$ aggregated from $\tilde{\boldsymbol{Z}}_{64\times64}$ in column (a); Column (d) A conditionally simulated image of $\boldsymbol{Z}$ conditional on $\tilde{\boldsymbol{Z}}_{16\times16}$ given in column (b); Column (e) A conditionally simulated image of $\boldsymbol{Z}$ conditional on $\tilde{\boldsymbol{Z}}_{8\times8}$ given in column (c).

only a comparison of scenarios for IDL and CPL as a function of signal magnitude h for $r \in \{6, 10\}$ and $\phi \in \{0, 5, 10\}$. Note that for a true power π, the Monte Carlo standard error of an estimated power is $\{\pi(1-\pi)/400\}^{1/2}$, which is bounded above by 0.025 (when $\pi = 0.5$). Recall that the Type-I error rate corresponds to $h = 0$ for each case.

From the power plots in Figure 4, we see that the Type-I error rates of CPL (and IDL) are roughly under control, reinforcing our conclusions from Section 3.1. For example, when $\phi = 10$, for our proposed procedure CPL, they are 0.075 for $\tilde{\boldsymbol{Z}}_{16\times16}$ and 0.060 for $\tilde{\boldsymbol{Z}}_{8\times8}$ with standard errors of 0.013 and 0.012, respectively. It is clear that the power of CPL increases with the magnitude h and the extent r of the signal, as it does for IDL. We can also see that signals can be detected much more easily for smaller ϕ (i.e., when the spatial dependence is weaker). In particular, the powers for $\phi = 0$ are considerably larger than the corresponding powers for $\phi \in \{5, 10\}$, indicating that spatial dependence makes the signal-

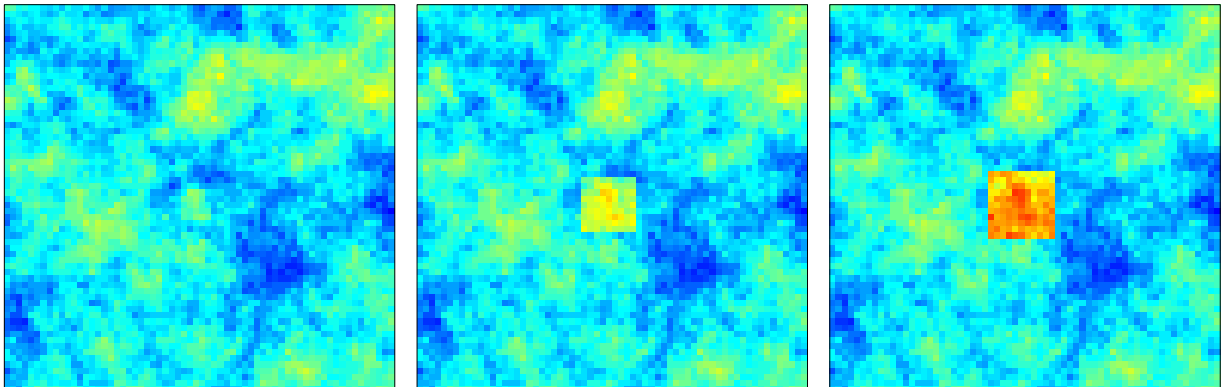

Figure 2: Images of $\boldsymbol{Z}$ with $\phi = 5$ and signals of various extents and magnitudes described in Section 3.2 and given, from left to right, by $(r, h) = (4, 1)$, $(8, 3)$, and $(10, 5)$, respectively.

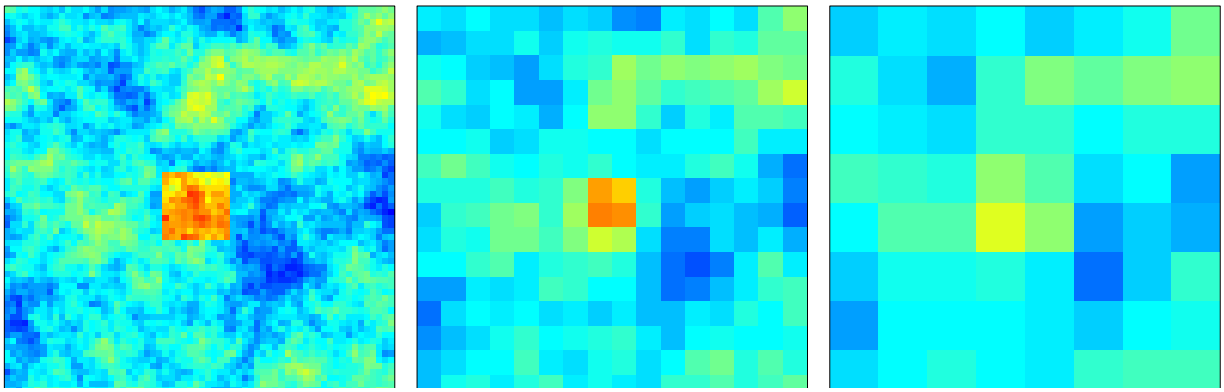

Figure 3: Aggregations of the image in Figure 2, right panel, with $(r, h) = (10, 5)$ giving, from left to right, $\tilde{\boldsymbol{Z}}_{64\times64}$, $\tilde{\boldsymbol{Z}}_{16\times16}$, $\tilde{\boldsymbol{Z}}_{8\times8}$. The left panel is identical to the right panel of Figure 2.

detection problem harder. It is also not surprising to see that our proposed procedure CPL has more power when applied to $\tilde{\boldsymbol{Z}}_{16\times16}$ than when applied to $\tilde{\boldsymbol{Z}}_{8\times8}$, although the power curve increases more slowly with h when the spatial dependence is stronger. It is encouraging that IDL's and CPL's power curves are close after one coarsening of resolution to $\tilde{\boldsymbol{Z}}_{16\times16}$. Often CPL and MOM applied to $\tilde{\boldsymbol{Z}}_{16\times16}$ outperform IDL applied to $\boldsymbol{Z}$. This is likely a consequence of basing the smoothed estimate $\hat{\boldsymbol{\Sigma}}$ on the exponential covariance function, which is of the same form as that used to generate $\boldsymbol{\delta}$ in (3). IDL makes no such assumptions when estimating the parameters $\hat{\boldsymbol{\theta}}$. In the presence of spatial dependence, another coarsening of resolution to $\tilde{\boldsymbol{Z}}_{8\times8}$ results in a substantial deterioration of the power curve of CPL, with power to find a signal arising only when $r = 10$. Again, as spatial dependence ϕ increases, the power curve

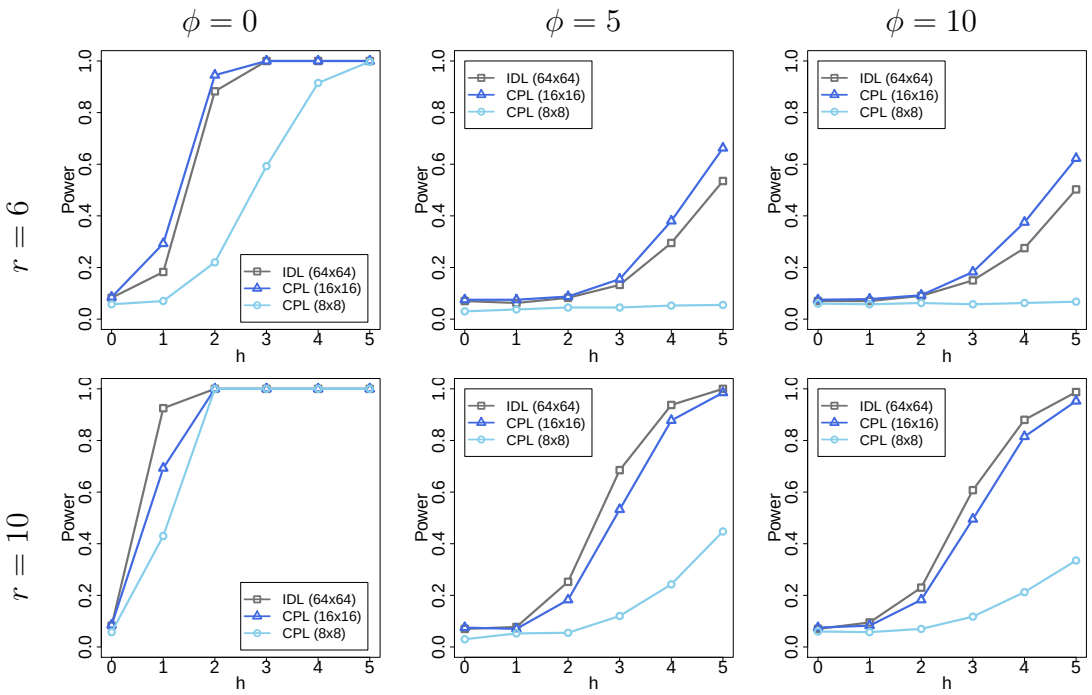

Figure 4: Empirical power curves of IDL, CPL (16×16), and CPL (8×8) in Experiment 1 described in Section 3.2, as a function of the signal's magnitude h, for $r \in \{6, 10\}$ and $\phi \in \{0, 5, 10\}$. Power curves for all values of (r, ϕ) and for all methods (IDL, CPL, MOM, and NVE) are shown in Figure S4 in the Supplementary Material.

is diminished.

We also use the receiver operating characteristic (ROC) curves (e.g., Egan, 1975) to facilitate comparison of the different methods. Figure 5 shows the ROC curves with signals of various volumes, hr^2, obtained from the 24 combinations of $r \in \{4, 6, 8, 10\}$ and $h \in \{0, 1, 2, 3, 4, 5\}$ with $\phi = 5$. Each ROC curve was computed based on 200 images of $\tilde{\boldsymbol{Z}}_{64\times 64} = \boldsymbol{Z}$, 100 of these generated under the null hypothesis of no signal and the other 100 images generated under the alternative hypothesis by adding a signal given by r and h. Each curve was traced out by varying α on the right-hand side of (24). From Figure 5, we see that the area under the curve (AUC) tends to increase with the signal volume hr^2 and decrease with the amount of aggregation, as expected. The full set of ROC curves for this experiment are shown in Figure S7 in the Supplementary Material.

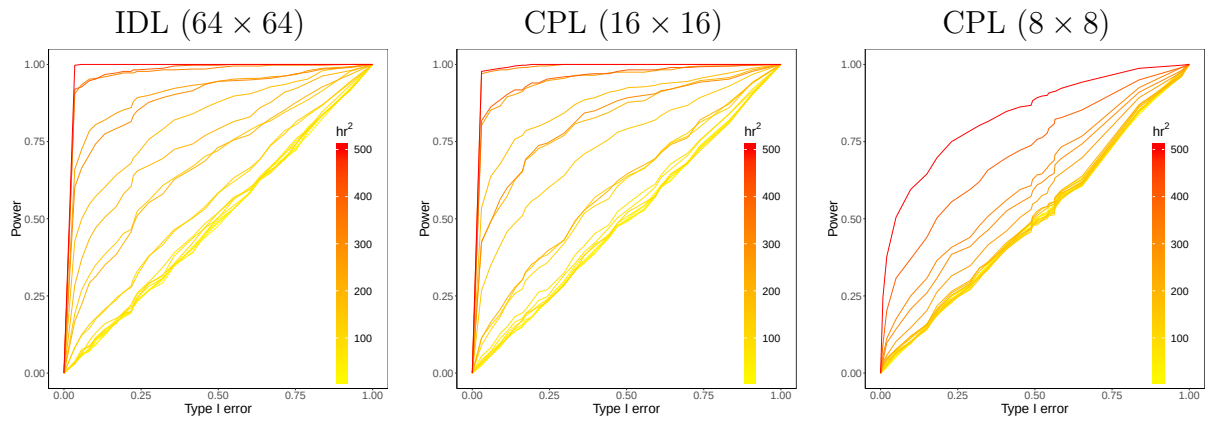

Figure 5: ROC curves of IDL, CPL (16×16), and CPL (8×8) in Experiment 1, colored according to the volume hr^2 of the signal obtained from 24 combinations of $r \in \{4, 6, 8, 10\}$ and $h \in \{0, 1, 2, 3, 4, 5\}$ for $\phi = 5$ and different scales of aggregations. The full set of ROC curves are shown in Figure S7 in the Supplementary Material.

Experiment 2: Missing data (in a contiguous block) at different scales of aggregation

Experiment 2 is similar to Experiment 1, except that here we considered missing data in the upper-right corner of $\tilde{\boldsymbol{Z}}_{64\times64}$ $(= \boldsymbol{Z})$, $\tilde{\boldsymbol{Z}}_{16\times16}$, and $\tilde{\boldsymbol{Z}}_{8\times8}$ with the fraction of missing data fixed at 9/64; see Figure 6 for an illustration. Tests were carried out at the usual 5% significance level. The power curves and the ROC curves obtained are similar to those in Figures 4 and 5, respectively; for the full set of curves, see Figures S5 and S8 in the Supplementary Material. The results demonstrate that our proposed procedure is not likely to be affected by a large contiguous block of data missing (as long as the block does not contain the true signal).

Experiment 3: Missing data (at random) at different scales of aggregation

Experiment 3 is similar to Experiment 2, except that we considered small blocks missing at random with the same fraction missing $(= 9/64)$; see Figure 7 for an illustration. These were taken from all but a few blocks that are contained in the central square region (where the signal is located), shown as a black-square outline in each panel of Figure 7. The power curves and ROC curves that we obtained were similar to those in Experiments 1 and 2 and are shown in Figures S6 and S9 in the Supplementary Material. A comparison of Experiments 2 and 3 shows that our proposed procedure, CPL, performs well irrespective of whether the

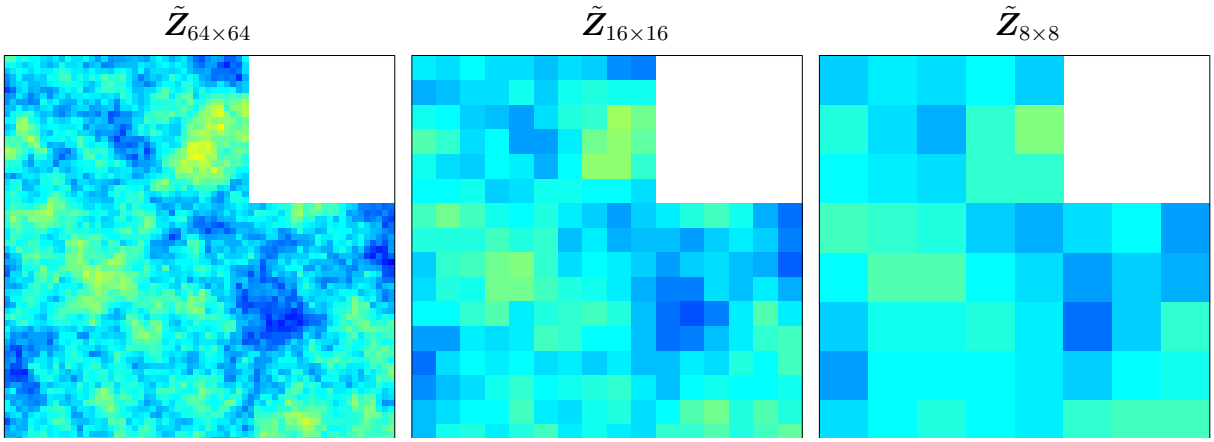

Figure 6: Missing data in Experiment 2 described in Section 3.2, at various scales of aggregation; in all images, the missing fraction is 9/64, and the spatial dependence $\phi = 5$.

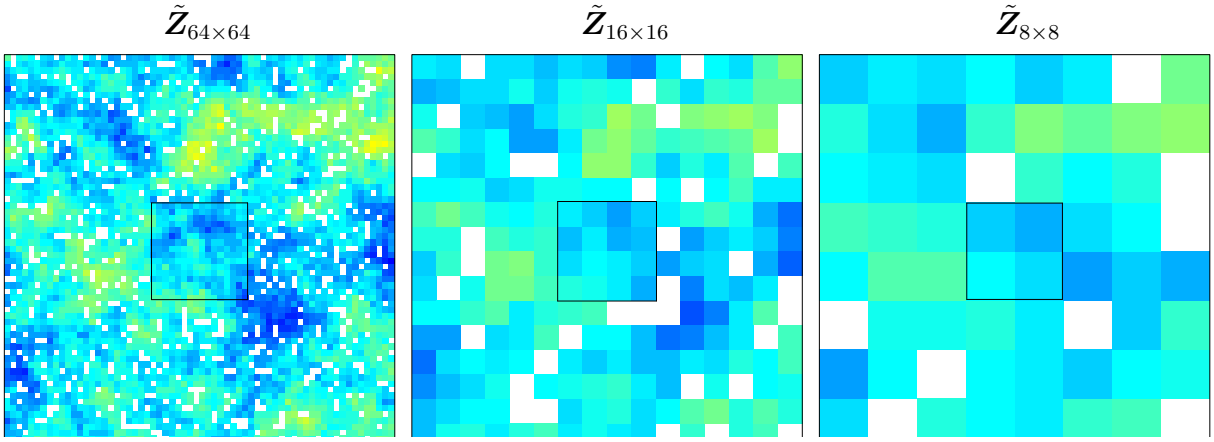

Figure 7: Missing data in Experiment 3 described in Section 3.2, at various scales of aggregation; in all images, the missing fraction is 9/64, and the spatial dependence $\phi = 5$.

data are missing at random or in a contiguous block. This is an illustration of how spatial modeling and its corresponding conditional simulation can successfully borrow strength for two very different missing-data mechanisms.

4. An application to temperature data in the Asia-Pacific

Finding signals in climate data is critically important for assessing the sustainability of Earth's ecosystems. In this section, we apply our proposed procedure to a temperature dataset obtained from the National Center for Atmospheric Research (NCAR) climate system model. The original data comprise monthly averages of 2-meter air temperatures on the

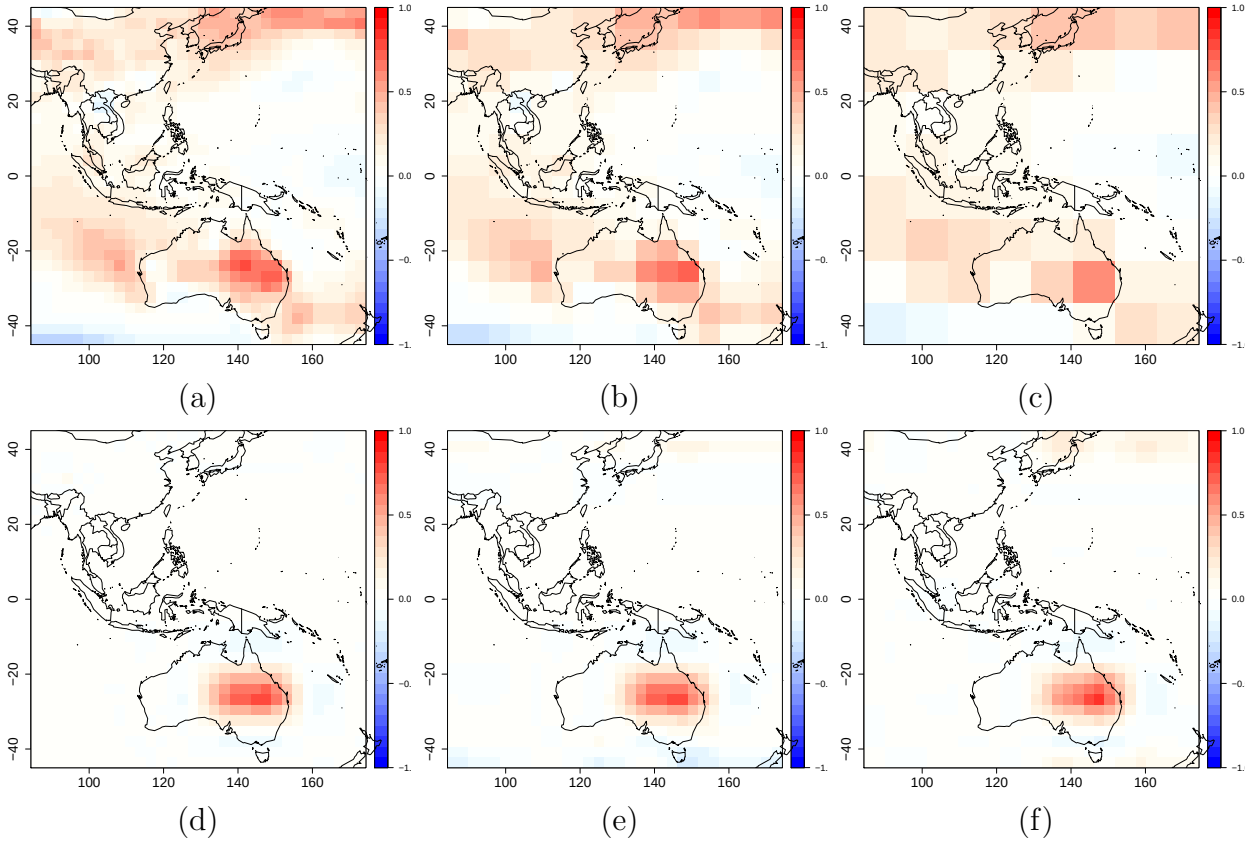

Figure 8: (a) The average 2-meter air temperature differences between the 1990s and the 1980s on the Kelvin scale at the native resolution of 32×32 pixels; (b) The image aggregated from (a) into 16×16 regular grid cells; (c) The image aggregated from (a) into 8×8 regular grid cells; (d) The 32×32 signal estimated by our proposed procedure CPL based on pixel-level data in (a); (e) The 32×32 signal estimated by our proposed procedure CPL based on the 16×16 aggregated data in (b); (f) The 32×32 signal estimated by our proposed procedure CPL based on the 8×8 aggregated data in (c).

Kelvin scale for the period 1980–1999 over the whole globe on 128×64 equiangular longitude-latitude (about $2.8^\circ \times 2.8^\circ$) grid cells or pixels. It is of interest to know if there is a decadal change in temperature (i.e., whether there is a possible signal) from the 1980s to the 1990s and, if a change has occurred, to identify the magnitudes and locations of the change. Here, we focus on a region D of 32×32 grid cells containing most of the Asia-Pacific from 45°N to 45°S and from 84°E to 174°E. We obtained the data $\boldsymbol{Z}$ by computing the average monthly temperature in the 1990s for each pixel in D, from which we subtracted the corresponding average monthly temperature in the 1980s. The resulting data, $\boldsymbol{Z} = \tilde{\boldsymbol{Z}}_{32\times 32}$, are shown in Figure 8(a).

Just as for the simulation experiments in Section 3, we considered scenarios involving complete data at different resolutions and incomplete data at different resolutions. Under the first of three scenarios, we aggregated $\boldsymbol{Z}$ into 16×16 and 8×8 regular grid cells, denoted by $\tilde{\boldsymbol{Z}}_{16\times16}$ and $\tilde{\boldsymbol{Z}}_{8\times8}$ and shown in Figure 8(b) and Figure 8(c), respectively. Under the next two scenarios, data missing in a contiguous block and data missing at random were considered: Initially, we removed a strip of data from $\tilde{\boldsymbol{Z}}_{32\times32}$, $\tilde{\boldsymbol{Z}}_{16\times16}$, and $\tilde{\boldsymbol{Z}}_{8\times8}$ to mimic a missing swath commonly seen in satellite data. The resulting data are denoted by $\tilde{\boldsymbol{Z}}^{(1)}_{32\times32}$, $\tilde{\boldsymbol{Z}}^{(1)}_{16\times16}$, and $\tilde{\boldsymbol{Z}}^{(1)}_{8\times8}$, with the same fraction of fine-resolution pixels missing ($= 1/8$), and they are shown in Figures 9(a)–(c), respectively. We then randomly removed a further 1/8 of the grid cells from the datasets $\tilde{\boldsymbol{Z}}^{(1)}_{32\times32}$, $\tilde{\boldsymbol{Z}}^{(1)}_{16\times16}$, and $\tilde{\boldsymbol{Z}}^{(1)}_{8\times8}$, respectively. The resulting data, denoted by $\tilde{\boldsymbol{Z}}^{(2)}_{32\times32}$, $\tilde{\boldsymbol{Z}}^{(2)}_{16\times16}$, and $\tilde{\boldsymbol{Z}}^{(2)}_{8\times8}$, respectively, are shown in Figures 10(a)–(c), where it is seen that the strip and a scattering of pixels leave behind irregular lattice data at different resolutions.

We applied our proposed procedure (CPL defined in Section 2) with the hypothesis test (24) to nine cases of incomplete spatially aggregated data: $\tilde{\boldsymbol{Z}}_{32\times32}$, $\tilde{\boldsymbol{Z}}_{16\times16}$, $\tilde{\boldsymbol{Z}}_{8\times8}$, $\tilde{\boldsymbol{Z}}^{(1)}_{32\times32}$, $\tilde{\boldsymbol{Z}}^{(1)}_{16\times16}$, $\tilde{\boldsymbol{Z}}^{(1)}_{8\times8}$, $\tilde{\boldsymbol{Z}}^{(2)}_{32\times32}$, $\tilde{\boldsymbol{Z}}^{(2)}_{16\times16}$, and $\tilde{\boldsymbol{Z}}^{(2)}_{8\times8}$. Similar to the numerical experiments in Section 3.2, we used the R package "EFDR" (Zammit-Mangion and Huang, 2015) with its default setting of the Daubechies least asymmetric wavelet filter of length 8 and the number of hypotheses to be tested in the wavelet space fixed at 100. As for the simulations in Section 3, we chose two wavelet scales resulting in seven wavelet classes corresponding to different scales and orientations. We estimated $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_7)'$ through (10), where $\hat{\tau}^2$ and $\hat{\boldsymbol{\gamma}}$ are the ML estimators based on the exponential covariance model.

Except for the last case, $\tilde{\boldsymbol{Z}}^{(2)}_{8\times8}$, where one of the missing cells is coincident with the potential signal, our proposed procedure rejected the null hypothesis of no decadal change in temperature at the 5% significance level. Their final p-values (on the scale $t = -2\log p$) are shown in Table 2. As expected, the values on the t-scale across rows tend to be larger for data at finer-scale resolutions. Comparison down columns supports our conclusion from

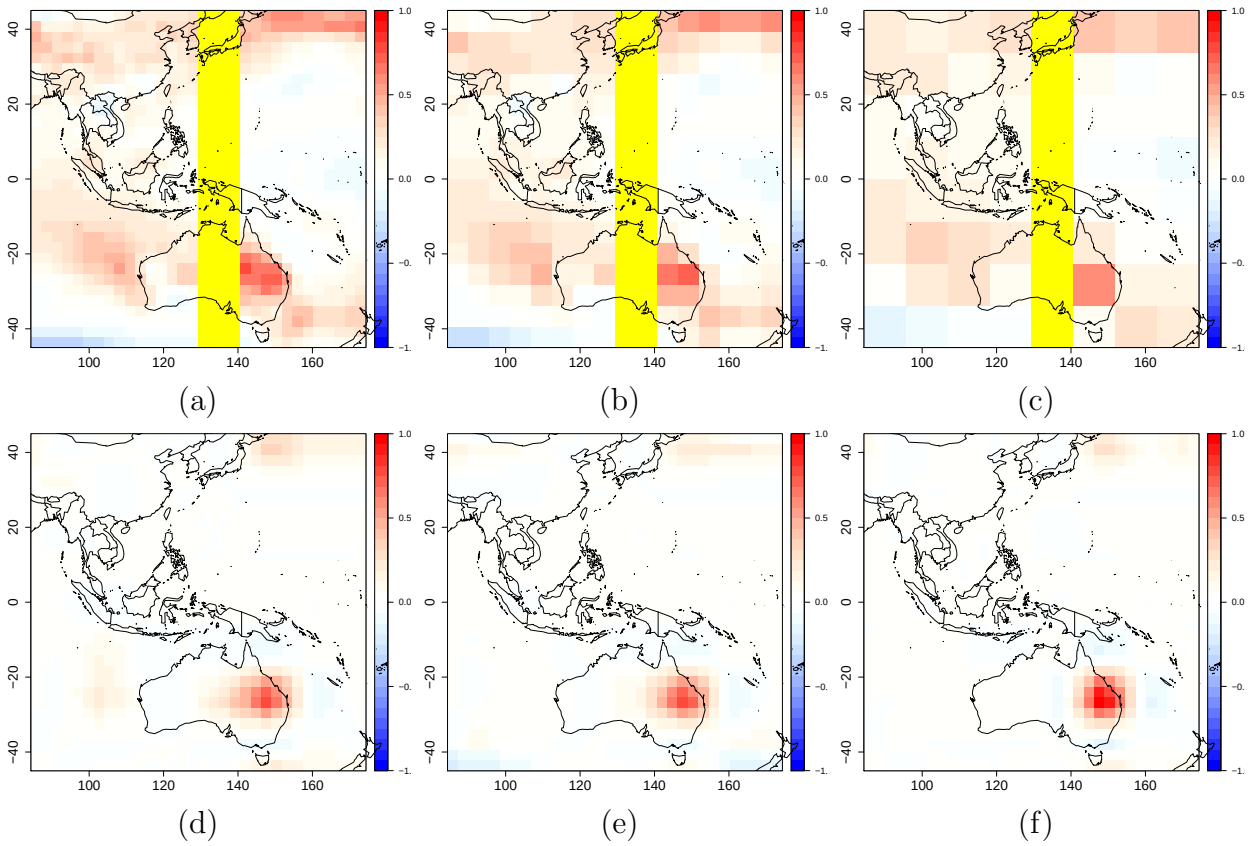

Figure 9: (a) The average 2-meter air temperature differences between the 1990s and the 1980s on the Kelvin scale at the native resolution of 32×32 pixels with a missing strip (shown in yellow); (b) The image aggregated from (a) into 16×16 regular grid cells with the missing strip superimposed; (c) The image aggregated from (a) into 8×8 regular grid cells with the missing strip superimposed; (d) The signal estimated by our proposed procedure CPL based on pixel-level data in (a); (e) The signal estimated by our proposed procedure CPL based on aggregated data in (b); (f) The signal estimated by our proposed procedure CPL based on aggregated data in (c).

the simulations in Section 3 that CPL is not greatly affected by missing data as long as the signal is observed.

The spatial patterns of temperature changes, given by $\hat{\boldsymbol{\mu}}$ in (25) and based on $\tilde{\boldsymbol{Z}}_{32\times32}$, $\tilde{\boldsymbol{Z}}_{16\times16}$, $\tilde{\boldsymbol{Z}}_{8\times8}$, are shown in Figures 8(d)–(f), respectively. As we saw in the simulations in Section 3, our proposed procedure handles successive aggregations well, and very similar signals of temperature increase are observed over east-central Australia. For the incomplete data, $\tilde{\boldsymbol{Z}}^{(1)}_{32\times32}$, $\tilde{\boldsymbol{Z}}^{(1)}_{16\times16}$, $\tilde{\boldsymbol{Z}}^{(1)}_{8\times8}$, $\tilde{\boldsymbol{Z}}^{(2)}_{32\times32}$, and $\tilde{\boldsymbol{Z}}^{(2)}_{16\times16}$, similar signals, albeit with smaller spatial extents, were identified by our proposed procedure CPL. The results are shown in Figures 9(d)–(f) and Figures 10(d)–(e), respectively. However, in $\tilde{\boldsymbol{Z}}^{(2)}_{8\times8}$, one of the missing

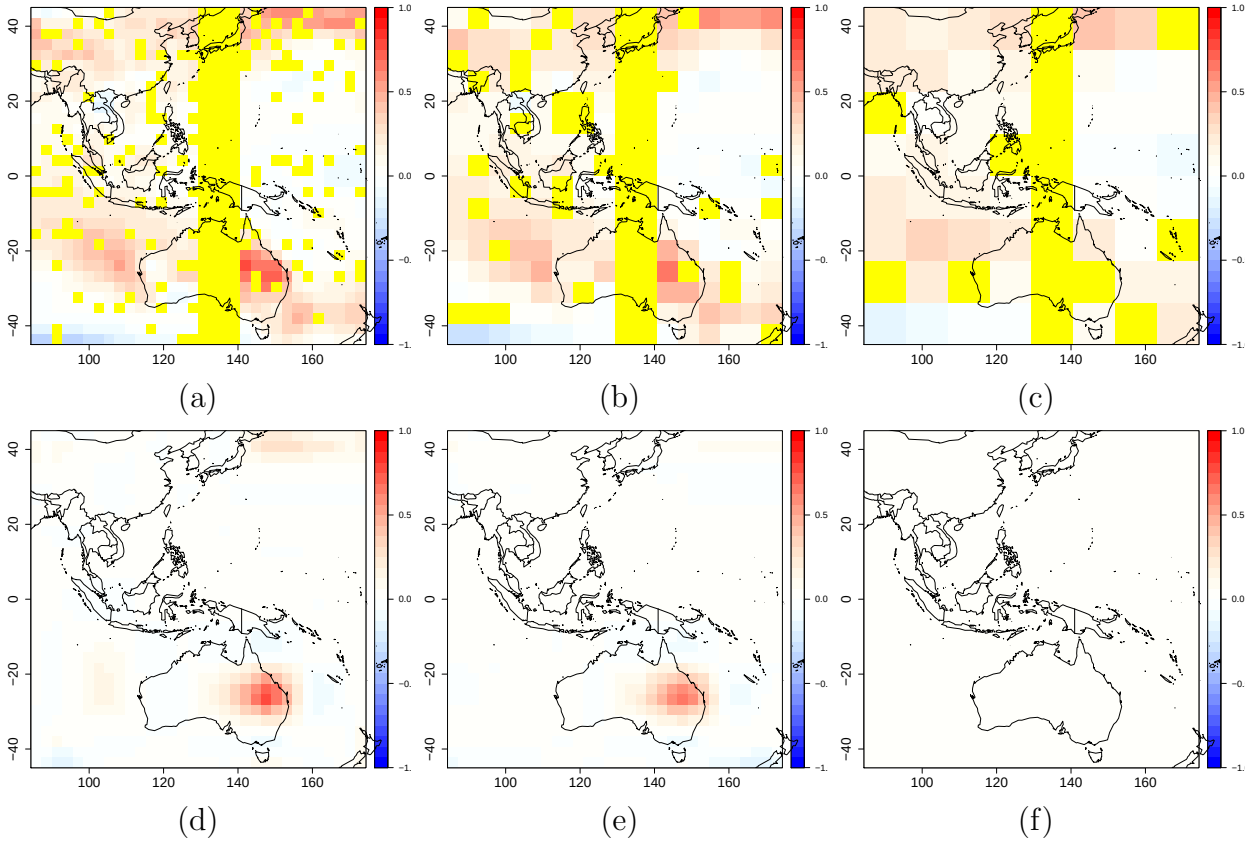

Figure 10: (a) The average 2-meter air temperature differences between the 1990s and the 1980s on the Kelvin scale at the native resolution of 32×32 pixels with a missing strip and randomly missing pixels (shown in yellow); (b) The image aggregated from Figure 8(a) into 16×16 regular grid cells with the missing strip and randomly missing grid cells superimposed; (c) The image aggregated from Figure 8(a) into 8×8 regular grid cells with the missing strip and randomly missing grid cells superimposed; (d) The signal estimated by our proposed procedure CPL based on pixel-level data in (a); (e) The signal estimated by our proposed procedure CPL based on aggregated data in (b); (f) The signal estimated by our proposed procedure CPL based on aggregated data in (c).

cells was coincident with the potential signal (see Figure 10(c)), which resulted in a failure to reject the null hypothesis and the blank image in Figure 10(f) that indicates no signal is present. When that cell was allowed to remain and a different cell removed from the 8×8 dataset, $t = -2 \log p$ went from 5.84 to 15.76, and the spatial signal seen in Figures 10(d) and 10(e) reappeared.

5. Discussion and conclusions

In this article, we have proposed a spatial hypothesis-testing procedure to detect signals from area-aggregated data, in the presence of spatially correlated noise. The procedure is

Table 2: The p-values (on the t-scale, where $t = -2\log p$) for our proposed procedure (CPL) applied to incomplete spatially aggregated temperature datasets in the Asia-Pacific under different levels of aggregation and incompleteness.

Missing Fraction	Scales of Aggregation		
	32×32	16×16	8×8
0	35.67	28.23	15.51
1/8	48.75	22.70	21.94
1/4	34.18	17.11	5.84

Note: $-2\log(0.05) = 5.99$.

based on using conditional simulations to infer the fine-scale spatial signal from incomplete spatially aggregated data.

A by-product of the research presented in this article is a novel methodology to combine exchangeable p-values into a single p-value using a bivariate Gaussian copula and a composite likelihood. In Section 3, we show that the methodology is able to properly control the Type-I error rate, even when the p-values are strongly correlated. Further, by extending n_1 and n_2 to their next powers of two, we can deal with any boundary effects caused by applying a DWT.

While we consider that the image $\boldsymbol{Z}$ at the original pixel resolution follows a multivariate Gaussian distribution, it is possible to extend the conditional-simulation approach for images generated from non-Gaussian distributions (or even discrete distributions). For example, one might use a generalized linear mixed model for non-Gaussian data, with the random effects derived from a latent Gaussian spatial process (e.g., Sengupta *et al.*, 2016; Wilson and Wakefield, 2019). This would be useful when, for example, $\boldsymbol{Z}$ are counts of events aggregated over a number of regions from a log-Gaussian Cox process. It would be straightforward to adapt the procedure proposed in this paper to these models by defining the signal $\boldsymbol{\mu}$ to be the mean of the latent Gaussian process. Then the null hypothesis, H_0: $\boldsymbol{\mu} = \boldsymbol{\mu}_0$, could be tested by conditionally simulating the hidden Gaussian process M times, conditional on the data. EFDR could then be applied to each simulated process and the resulting p-values combined into a single p-value, as we have done for Gaussian data.

Acknowledgments

Hsin-Cheng Huang's research was supported by ROC Ministry of Science and Technology grants MOST 103-2118-M-001-007-MY3 and MOST 105-2119-M-001-035. Noel Cressie's research was supported by Australian Research Council Discovery Projects DP150104576 and DP190100180, and by NSF grant SES-1132031 funded through the NSF-Census Research Network (NCRN) program. Andrew Zammit-Mangion's research was supported by an Australian Research Council Discovery Early Career Research Award (DECRA) DE180100203 and Discovery Project DP190100180. The authors would like to thank Chris Wikle, Scott Holan, Vineet Yadav, and Mike Gunson for comments on parts of this research.

Supplementary Material

The Supplementary Material contains complete figures for the simulations in Section 3. These include: $\tilde{\boldsymbol{Z}}_{16\times16}$ and $\tilde{\boldsymbol{Z}}_{8\times8}$, corresponding to $\boldsymbol{Z}$ in Figure 2; empirical power curves as a function of the signal's magnitude h, for various procedures (NVE, IDL, our proposed procedure CPL, and its variant MOM); and ROC curves for CPL, where the signal volume hr^2 is varied.

References

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing, *Journal of the Royal Statistical Society, Series B*, **57**, 289-300.

Brown, M. (1975). A method for combining non-independent, one-sided tests of significance, *Biometrics*, **31**, 987-992.

Cressie, N. (1993). Statistics for Spatial Data, rev. edn, Wiley, New York, NY.

Daubechies, I. (1992). *Ten Lectures on Wavelets*, CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, Philadelphia, PA.

Egan, J. P. (1975). *Signal Detection Theory and ROC Analysis*, Academic Press, New York, NY.

Fisher, R. A. (1925). *Statistical Methods for Research Workers*, Oliver and Boyd, Edinburgh, UK.

Gilleland, E. (2013). Testing competing precipitation forecasts accurately and efficiently: The spatial prediction comparison test, *Monthly Weather Review*, **141**, 340-355.

Hering, A. S. and Genton, M. G. (2011). Comparing spatial predictions, *Technometrics*, **53**, 414-425.

Littell, R. C. and Folks, J. L. (1971). Asymptotic optimality of Fisher's method of combining independent tests, *Journal of the American Statistical Association*, **66**, 802-806.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*, 2nd edn, Wiley, New York, NY.

Martinez, J. G., Bohn, K. M., Carroll, R. J., and Morris, J. S. (2013). A study of Mexican free-tailed bat chirp syllables: Bayesian functional mixed models for nonstationary acoustic time series, *Journal of the American Statistical Association*, **108**, 514-526.

Nguyen, H., Cressie, N., and Braverman, A. (2012). Spatial statistical data fusion for remote sensing applications, *Journal of the American Statistical Association*, **107**, 1004-1018.

Pavlicová, M., Santner, T. J., and Cressie, N. (2008). Detecting signals in FMRI data using powerful FDR procedures, *Statistics and Its Interface*, **1**, 23-33.

Risser, M. D., Paciorek, C. J., Stone, D. A. (2019). Spatially-dependent multiple testing under model misspecification, with application to detection of anthropogenic influence on extreme climate events, *Journal of the American Statistical Association*, to appear.

Schott, J. R. (2017). *Matrix Analysis for Statistics*, 3rd edn, Wiley, Hoboken, NJ.

Shen, X., Huang, H.-C., and Cressie, N. (2002). Nonparametric hypothesis testing for a spatial signal, *Journal of the American Statistical Association*, **97**, 1122-1140.

Song, P. X.-K. (2000). Multivariate dispersion models generated from Gaussian copula, *Scandinavian Journal of Statistics*, **27**, 305-320.

Sengupta, A., Cressie, N., Kahn, B. H., and Frey, R. (2016). Predictive inference for big, spatial, non-Gaussian data: MODIS cloud data and its change-of-support, *Australian and New Zealand Journal of Statistics*, **58**, 15-45.

Spiegelhalter, D. J., Abrams, K. R., and Myles, J. P. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*, Wiley, New York, NY.

Sun, W., Reich, B.J., Cai, T.T., Guindani, M., and Schwartzman, A. (2015). False discovery control in large-scale spatial multiple testing, *Journal of the Royal Statistical Society, Series B*, **77**, 59-83.

Wasserstein, R. L., Schirm, A. L., and Lazar, N. A. (2019). Moving to a world beyond "$p < 0.05$," *The American Statistician*, **73**, 1-19.

Wilson, K. and Wakefield, J. (2019). Pointless spatial modeling, *Biostatistics*, to appear.

Ye, J. (1998). On measuring and correcting the effects of data mining and model selection, *Journal of the American Statistical Association*, **93**, 120-131.

Yun, S., Zhang, X., and Li, B. (2018). Detection of local differences between two spatiotemporal random fields, Manuscript.

Zammit-Mangion, A. and Huang, H.-C. (2015). EFDR: Wavelet-based enhanced FDR for signal detection in noisy images, R package version 0.1.1. URL `https://CRAN.R-project.org/package=EFDR`.

Supplementary Material for "False Discovery Rates to Detect Signals from Incomplete Spatially Aggregated Data"

Hsin-Cheng Huang[a,*], Noel Cressie[b], Andrew Zammit-Mangion[b], Guowen Huang[c]

[a] *Institute of Statistical Science, Academia Sinica, Taiwan, R.O.C.*
[b] *National Institute for Applied Statistics Research Australia, University of Wollongong, Australia*
[c] *Department of Statistical Science, University of Toronto, Canada*

The supplementary material section contains the following nine figures:

S1 Images of $\boldsymbol{Z}$ with $\phi = 5$ and signals of various extents $r \in \{4, 6, 8, 10\}$ down rows and various magnitudes $h \in \{0, 1, 2, 3, 4, 5\}$ across columns.

S2 Images of $\boldsymbol{Z}_{16\times 16}$ obtained by aggregating $\boldsymbol{Z}$ in Figure S1 into 4×4 blocks resulting in 16×16 grid cells.

S3 Images of $\boldsymbol{Z}_{8\times 8}$ obtained by aggregating $\boldsymbol{Z}$ in Figure S1 into 8×8 blocks resulting in 8×8 grid cells.

S4–S6 Empirical power curves as a function of the signal's magnitude h, for various procedures for testing of H_0 in Experiments 1–3, respectively.

S7–S9 ROC curves for IDL and the proposed procedure, CPL, in Experiments 1–3, respectively.

*Corresponding author

Email addresses: hchuang@stat.sinica.edu.tw (Hsin-Cheng Huang), ncressie@uow.edu.au (Noel Cressie), azm@uow.edu.au (Andrew Zammit-Mangion), hgw0610209@gmail.com (Guowen Huang)

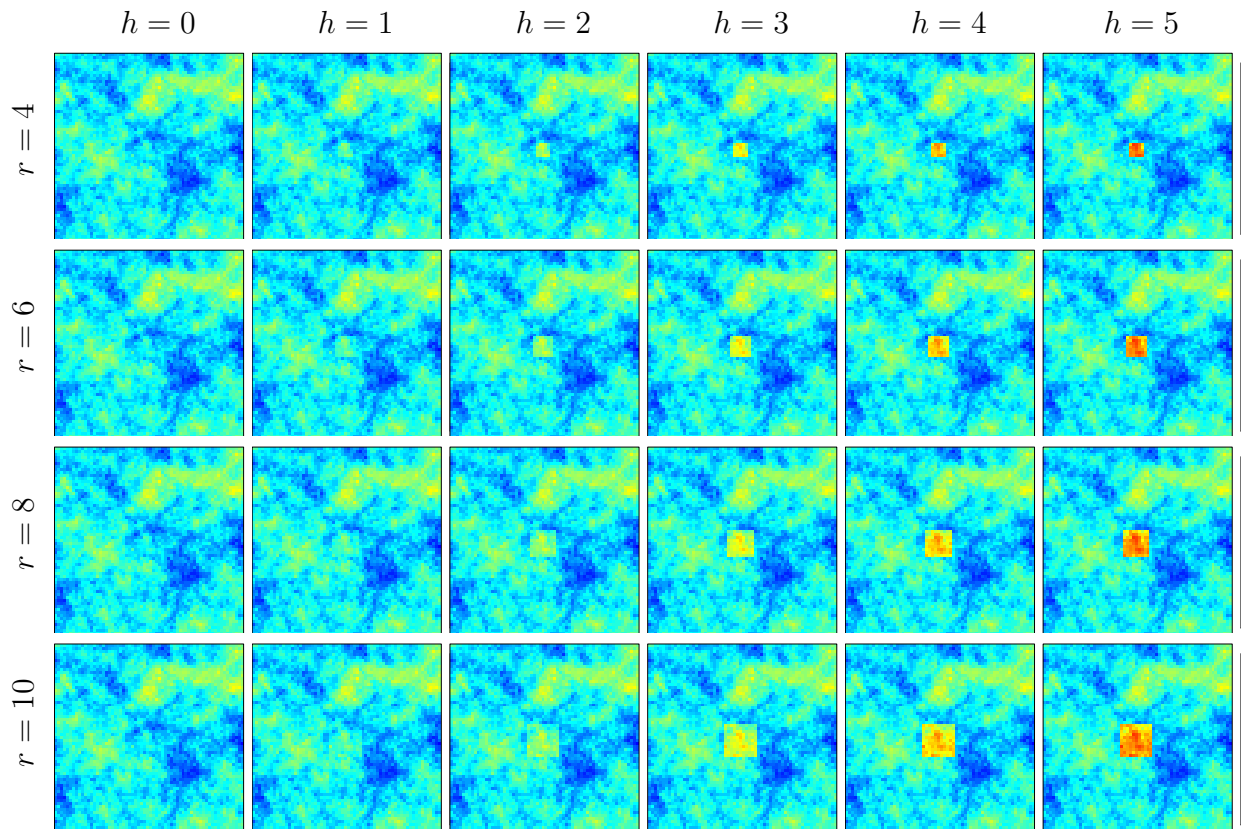

Figure S1: Images of $\boldsymbol{Z}$ with $\phi = 5$ and signals of various extents $r \in \{4, 6, 8, 10\}$ down rows and various magnitudes $h \in \{0, 1, 2, 3, 4, 5\}$ across columns.

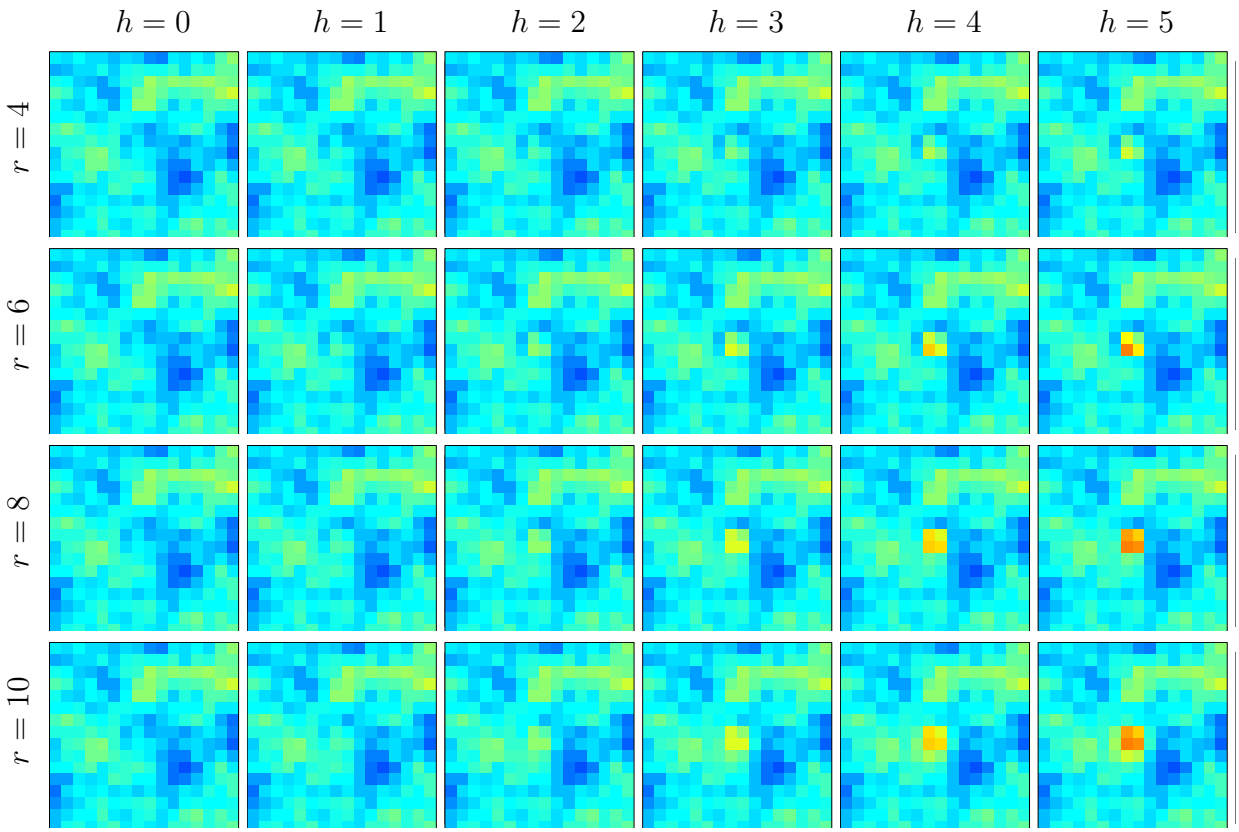

Figure S2: Images of $\tilde{\boldsymbol{Z}}_{16\times16}$ obtained by aggregating $\boldsymbol{Z}$ in Figure S1 into 4×4 blocks resulting in 16×16 grid cells.

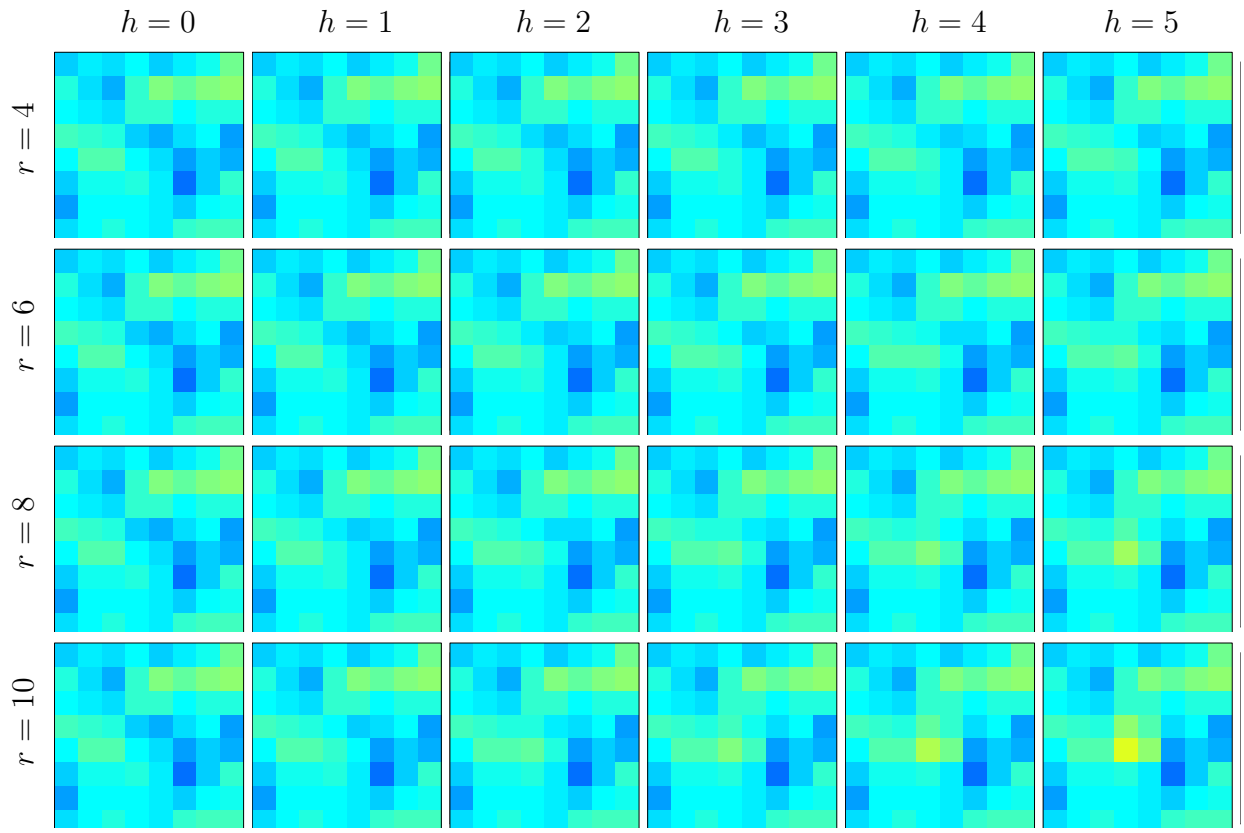

Figure S3: Images of $\tilde{\boldsymbol{Z}}_{8\times 8}$ obtained by aggregating $\boldsymbol{Z}$ in Figure S1 into 8×8 blocks resulting in 8×8 grid cells.

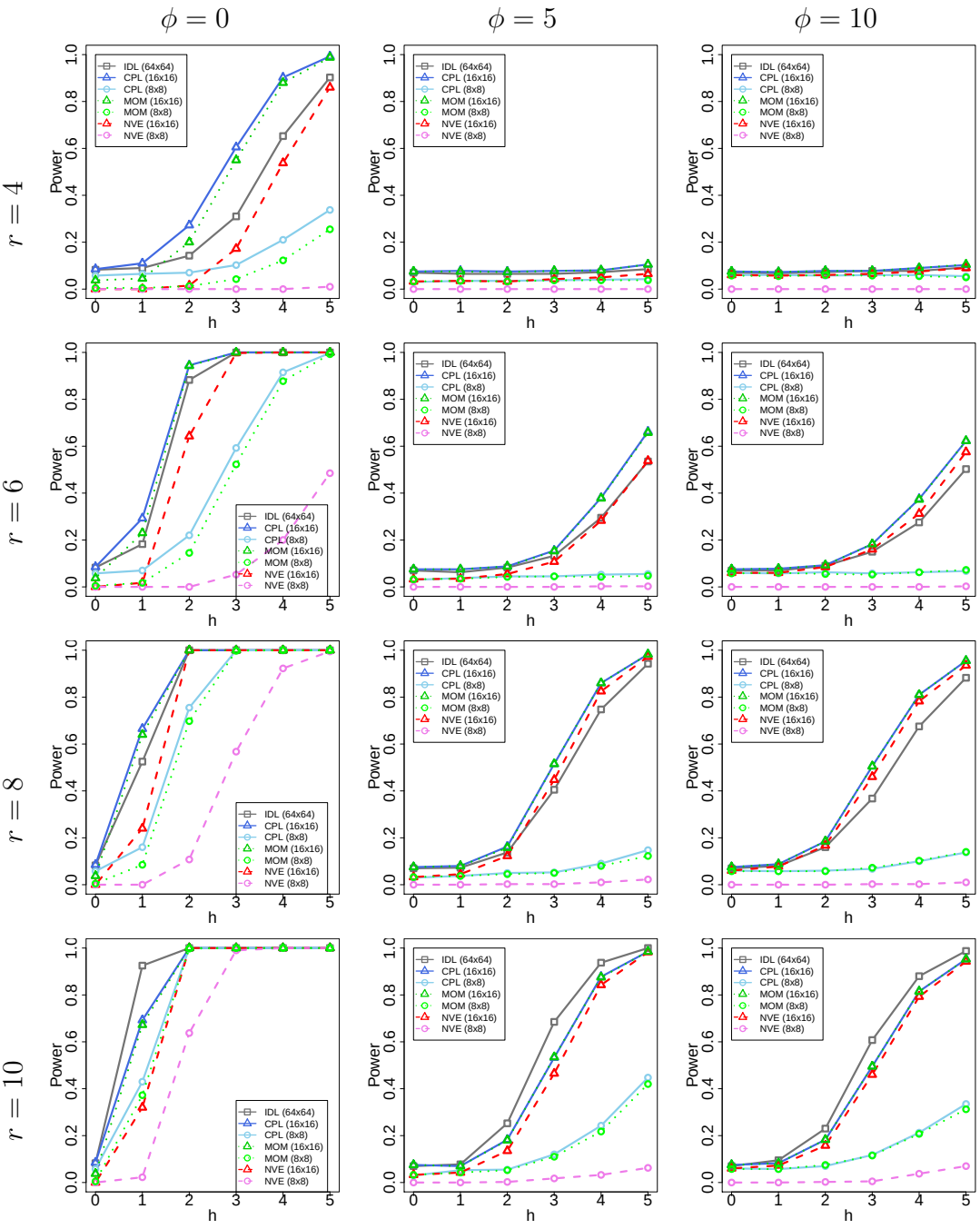

Figure S4: Empirical power curves as a function of the signal's magnitude h, for various procedures for testing H_0 in Experiment 1. Down the rows, the curves correspond to different signal extents r, while across the columns, the curves correspond to different spatial-dependence values ϕ.

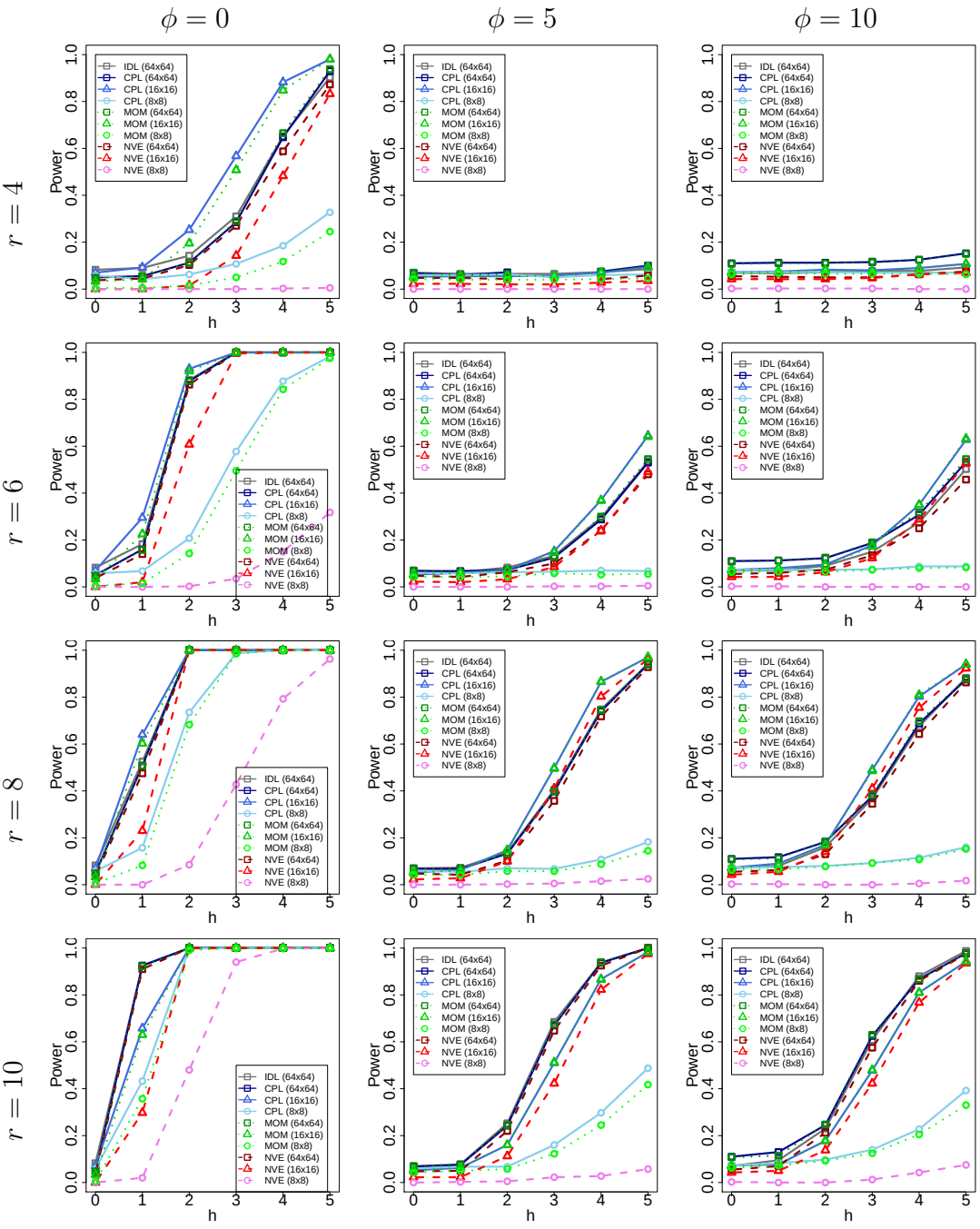

Figure S5: Empirical power curves as a function of the signal's magnitude h, for various procedures for testing H_0 in Experiment 2. Down the rows, the curves correspond to different signal extents r, while across the columns, the curves correspond to different spatial-dependence values ϕ.

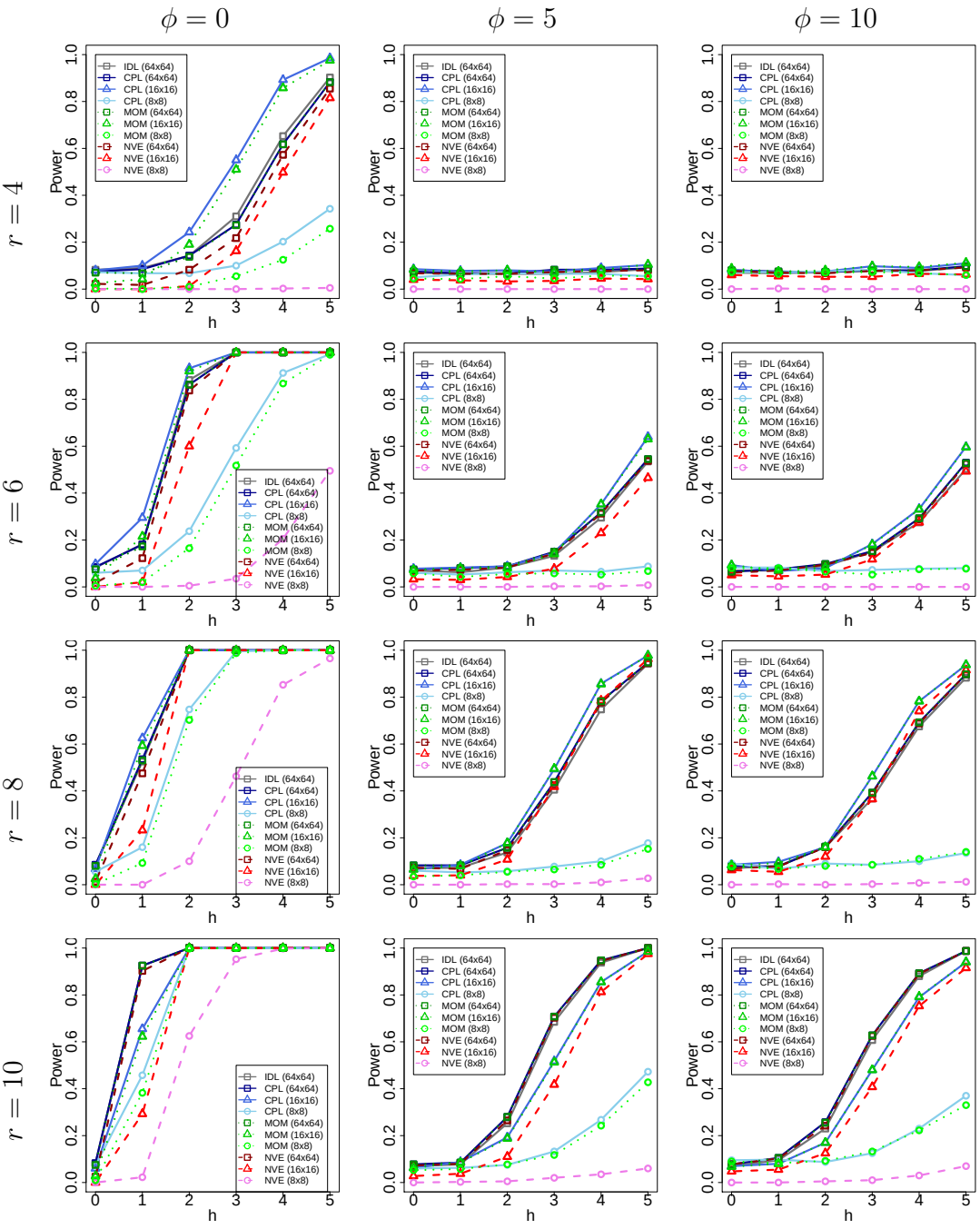

Figure S6: Empirical power curves as a function of the signal's magnitude h, for various procedures for testing H_0 in Experiment 3. Down the rows, the curves correspond to different signal extents r, while across the columns, the curves correspond to different spatial-dependence values ϕ.

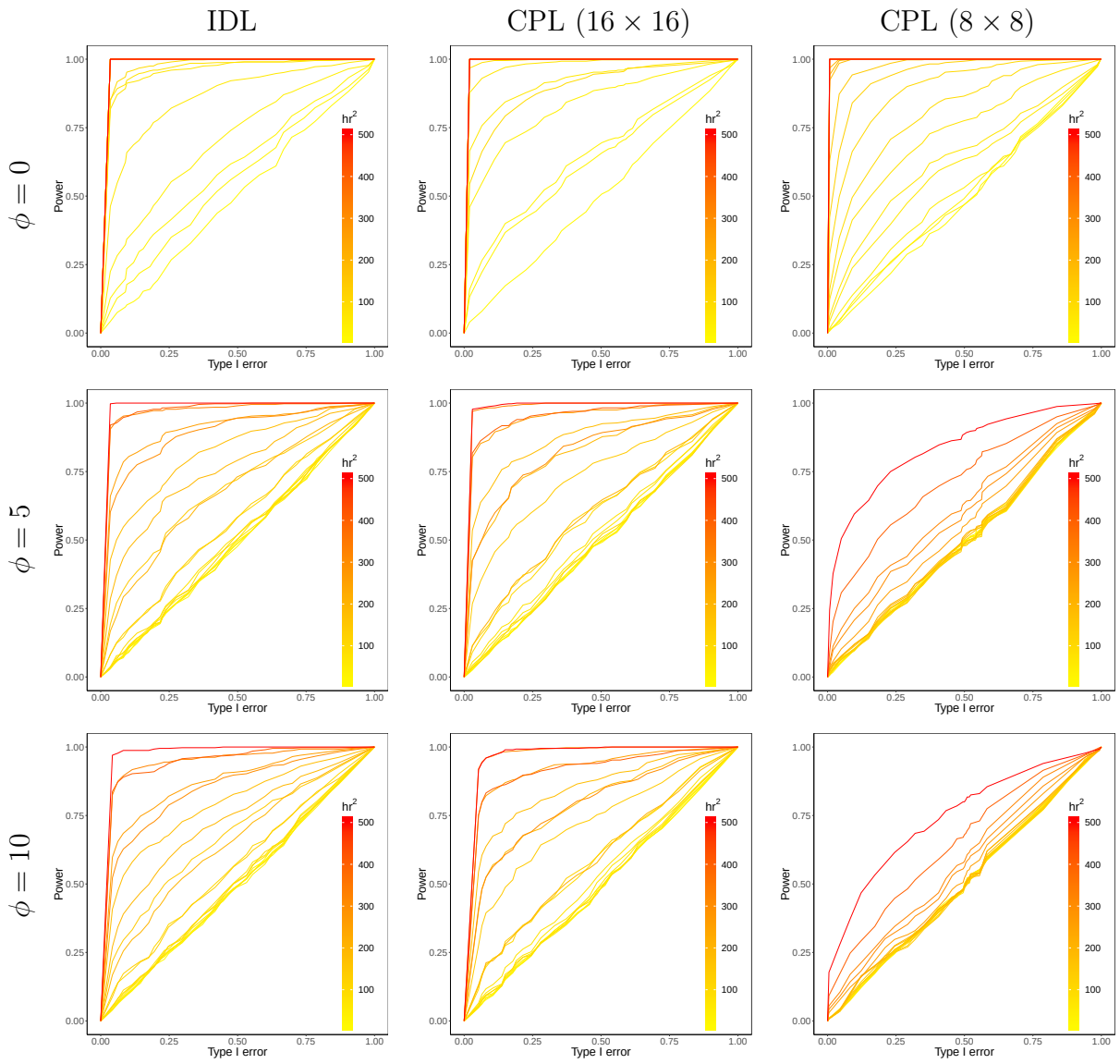

Figure S7: ROC curves for IDL, CPL (16×16), and CPL (8×8) in Experiment 1, colored according to the volume hr^2 of the signal obtained from 24 combinations of $r \in \{4, 6, 8, 10\}$ and $h \in \{0, 1, 2, 3, 4, 5\}$ in each plot. Down the rows, the curves correspond to different spatial-dependence values ϕ. Each curve was obtained by varying α on the right-hand side of (24).

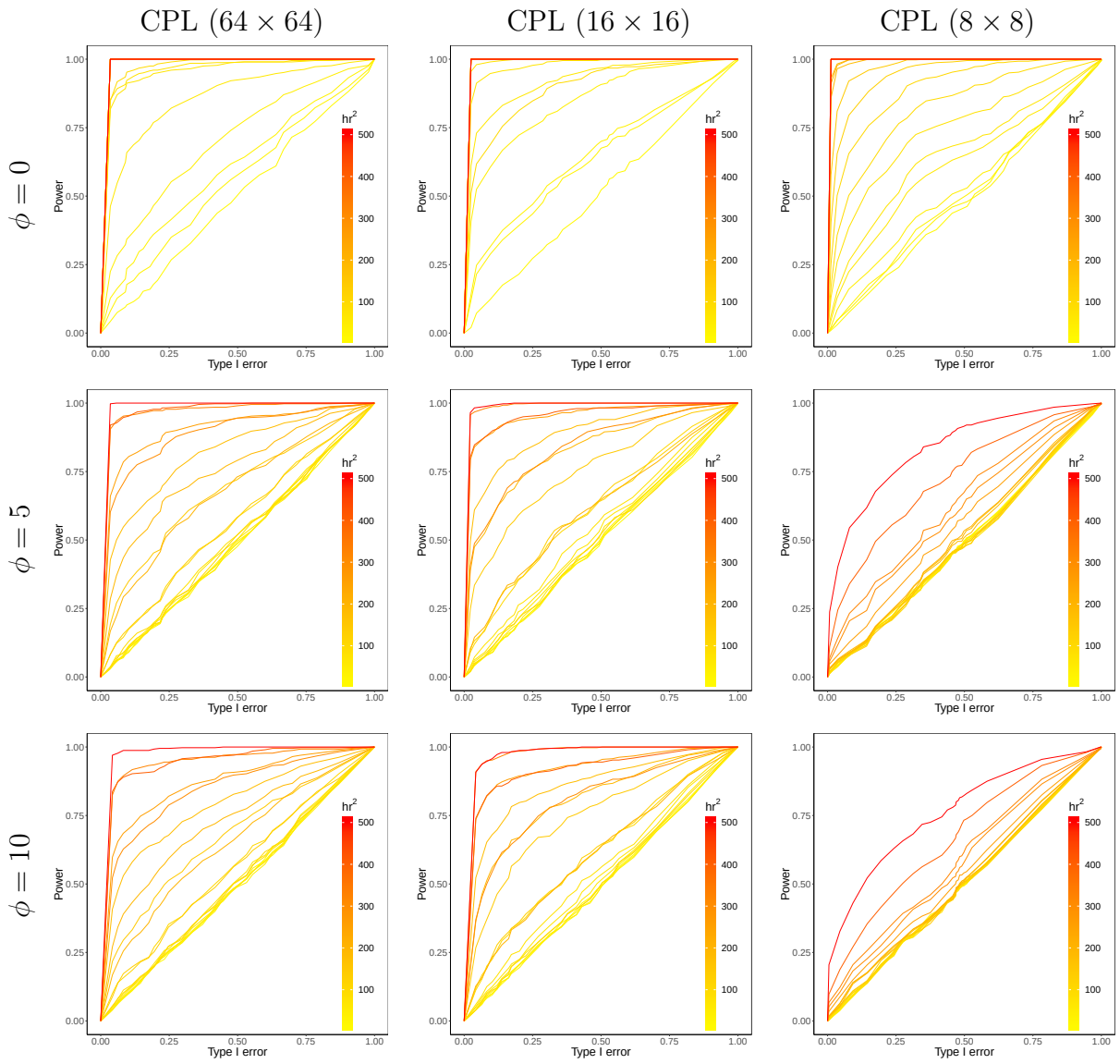

Figure S8: ROC curves for CPL (64×64), CPL (16×16), and CPL (8×8) in Experiment 2, colored according to the volume hr^2 of the signal obtained from 24 combinations of $r \in \{4, 6, 8, 10\}$ and $h \in \{0, 1, 2, 3, 4, 5\}$ in each plot. Down the rows, the curves correspond to different spatial-dependence values ϕ. Each curve was obtained by varying α on the right-hand side of (24).

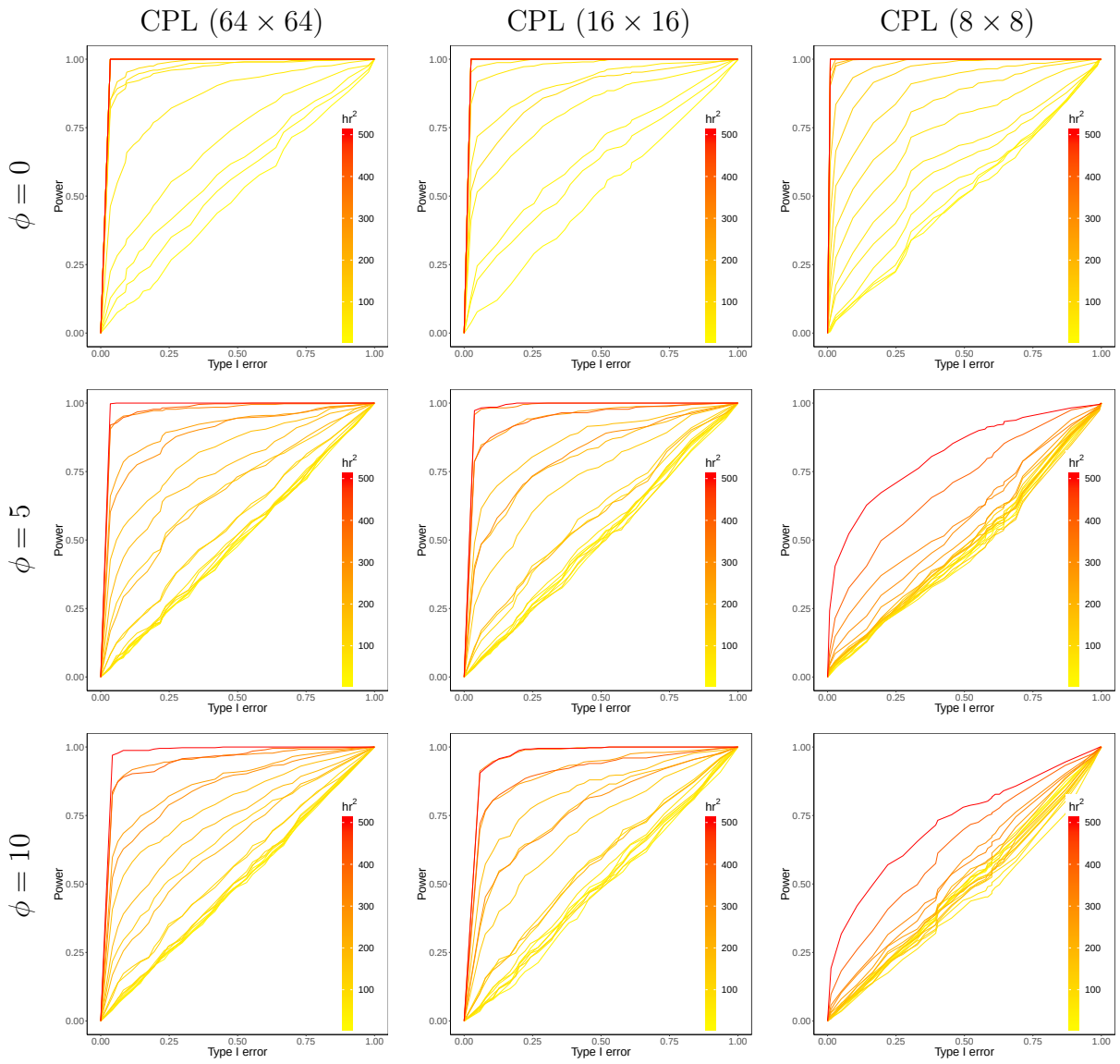

Figure S9: ROC curves for CPL (64×64), CPL (16×16), and CPL (8×8) in Experiment 3, colored according to the volume hr^2 of the signal obtained from 24 combinations of $r \in \{4, 6, 8, 10\}$ and $h \in \{0, 1, 2, 3, 4, 5\}$ in each plot. Down the rows, the curves correspond to different spatial-dependence values ϕ. Each curve was obtained by varying α on the right-hand side of (24).