Supplementary Material for 'A Hypothesis Test for Detecting Distance-Specific Clustering and Dispersion in Areal Data'

Stella Self, Anna Overby, Anja Zgodic, David White, Alexander C. McLain, and Caitlin Dyckman.

Web Appendix A: Derivations of PAPF Properties.

This section contains proofs of various properties of the positive area proportion function (PAPF) presented in Section 2 of the corresponding manuscript.

CLAIM 1. *Let $\mathcal{A}$ be a stationary independent process with areal units $a_1, a_2, ..., a_N$ and $P(Y_i = 1) = \lambda$. Then*

$$M_0(r) = \frac{1}{N}\sum_{i=1}^{N} M_{0i}(r).$$

PROOF. Consider

$$\begin{aligned}
M_0(r) &= E\left[\frac{1}{n_{\boldsymbol{Y}}}\sum_{i=1}^{N} M_{0i}(r)Y_i | n_{\boldsymbol{Y}} > 0\right] \\
&= \sum_{n=1}^{N} E\left[\frac{1}{n_{\boldsymbol{Y}}}\sum_{i=1}^{N} M_{0i}(r)Y_i | n_{\boldsymbol{Y}} = n, n_{\boldsymbol{Y}} > 0\right] P(n_{\boldsymbol{Y}} = n | n_{\boldsymbol{Y}} > 0) \\
&= \sum_{n=1}^{N} E\left[\frac{1}{n}\sum_{i=1}^{N} M_{0i}(r)Y_i | n_{\boldsymbol{Y}} = n\right] P(n_{\boldsymbol{Y}} = n | n_{\boldsymbol{Y}} > 0) \\
&= \sum_{n=1}^{N} \left(\frac{1}{n}\sum_{i=1}^{N} E\left[M_{0i}(r)Y_i | n_{\boldsymbol{Y}} = n\right]\right) P(n_{\boldsymbol{Y}} = n | n_{\boldsymbol{Y}} > 0) \\
&= \sum_{n=1}^{N} \left(\frac{1}{n}\sum_{i=1}^{N} E\left[M_{0i}(r)Y_i | Y_i = 1, n_{\boldsymbol{Y}} = n\right] P(Y_i = 1 | n_{\boldsymbol{Y}} = n)\right) P(n_{\boldsymbol{Y}} = n | n_{\boldsymbol{Y}} > 0) \\
&= \sum_{n=1}^{N} \left(\frac{1}{n}\sum_{i=1}^{N} E\left[M_{0i}(r) | Y_i = 1, n_{\boldsymbol{Y}} = n\right] \frac{n}{N}\right) P(n_{\boldsymbol{Y}} = n | n_{\boldsymbol{Y}} > 0)
\end{aligned}$$

Recalling that M_{0i} is a constant:

$$\begin{aligned}
&= \sum_{n=1}^{N} \left(\frac{1}{N}\sum_{i=1}^{N} M_{0i}(r)\right) P(n_{\boldsymbol{Y}} = n | n_{\boldsymbol{Y}} > 0) \\
&= \frac{1}{N}\sum_{i=1}^{N} M_{0i}(r) \sum_{n=1}^{N} P(n_{\boldsymbol{Y}} = n | n_{\boldsymbol{Y}} > 0) \\
&= \frac{1}{N}\sum_{i=1}^{N} M_{0i}(r)
\end{aligned}$$

□

CLAIM 2. *For a stationary, independent areal process $\mathcal{A}$ with areal units $a_1, a_2, ..., a_N$*

$$E[\widehat{M}(r, \boldsymbol{Y}) | n_{\boldsymbol{Y}} > 0] = M_0(r).$$

PROOF. Consider

$$E[\widehat{M}(r,\boldsymbol{Y})|n_{\boldsymbol{Y}}>0] = E\left[\frac{1}{n_{\boldsymbol{Y}}}\sum_{i=1}^{N}\widehat{M}_i(r,\boldsymbol{Y})Y_i|n_{\boldsymbol{Y}}>0\right]$$

$$= \sum_{n=1}^{N} E\left[\frac{1}{n_{\boldsymbol{Y}}}\sum_{i=1}^{N}\widehat{M}_i(r,\boldsymbol{Y})Y_i|n_{\boldsymbol{Y}}=n, n_{\boldsymbol{Y}}>0\right] P(n_{\boldsymbol{Y}}=n|n_{\boldsymbol{Y}}>0)$$

$$= \sum_{n=1}^{N}\frac{1}{n}\sum_{i=1}^{N} E\left[\widehat{M}_i(r,\boldsymbol{Y})Y_i|n_{\boldsymbol{Y}}=n\right] P(n_{\boldsymbol{Y}}=n|n_{\boldsymbol{Y}}>0)$$

$$= \sum_{n=1}^{N}\frac{1}{n}\sum_{i=1}^{N} E\left[\widehat{M}_i(r,\boldsymbol{Y})Y_i|Y_i=1, n_{\boldsymbol{Y}}=n\right] P(Y_i=1|n_{\boldsymbol{Y}}=n)P(n_{\boldsymbol{Y}}=n|n_{\boldsymbol{Y}}>0)$$

$$= \sum_{n=1}^{N}\frac{1}{n}\sum_{i=1}^{N} E\left[\widehat{M}_i(r,\boldsymbol{Y})|Y_i=1, n_{\boldsymbol{Y}}=n\right] P(Y_i=1|n_{\boldsymbol{Y}}=n)P(n_{\boldsymbol{Y}}=n|n_{\boldsymbol{Y}}>0)$$

Noting Y_i is independent from $\boldsymbol{Y}_{-i}$ and $\widehat{M}_i(r,\boldsymbol{Y})$ does not depend on Y_i :

$$= \sum_{n=1}^{N}\frac{1}{n}\sum_{i=1}^{N} E\left[\widehat{M}_i(r,\boldsymbol{Y})|n_{\boldsymbol{Y}}=n\right] P(Y_i=1|n_{\boldsymbol{Y}}=n)P(n_{\boldsymbol{Y}}=n|n_{\boldsymbol{Y}}>0)$$

$$= \sum_{n=1}^{N}\frac{1}{n}\sum_{i=1}^{N} E\left[\widehat{M}_i(r,\boldsymbol{Y})|n_{\boldsymbol{Y}}=n\right] \frac{n}{N}P(n_{\boldsymbol{Y}}=n|n_{\boldsymbol{Y}}>0)$$

$$= \frac{1}{N}\sum_{n=1}^{N}\sum_{i=1}^{N} E\left[\widehat{M}_i(r,\boldsymbol{Y})|n_{\boldsymbol{Y}}=n, n_{\boldsymbol{Y}}>0\right] P(n_{\boldsymbol{Y}}=n|n_{\boldsymbol{Y}}>0)$$

$$= \frac{1}{N}E\left[\sum_{i=1}^{N}\widehat{M}_i(r,\boldsymbol{Y})|n_{\boldsymbol{Y}}>0\right]$$

$$= \frac{1}{N}\sum_{i=1}^{N} M_{i0}(r)$$

Invoking Claim 1, we have:

$$= M_0(r)$$

□

Web Appendix B: Additional Simulation Results.

A simulation study was conducted to assess the sensitivity of the PAPF method to it's dependence on areal unit centroids. In this simulation study, an alternate version of the PAPF and the corresponding test statistics were defined, in which each areal unit centroid was replaced by a randomly selected point from each area unit. Specifically, we define

$$M_i^*(r) = E\left\{\frac{\mathscr{N}[c(\boldsymbol{\ell}_i^*,r)\cap a_i^c] + A[c(\boldsymbol{\ell}_i^*,r)\cap a_i]}{A[c(\boldsymbol{\ell}_i^*,r)\cap\mathscr{A})]}\cdot\left(\frac{\mathscr{N}(\mathscr{A})}{A(\mathscr{A})}\right)^{-1}\middle| n_{\boldsymbol{Y}}>0\right\}.$$

where $\boldsymbol{\ell}_i^*$ is any point in a_i and the expectation is taken over $\boldsymbol{Y}$ and all possible choices of $\boldsymbol{\ell}_i^*$. We then define

$$M(r) = E\left[\frac{1}{n_{\boldsymbol{Y}}}\sum_{i=1}^{N} M_i^*(r)Y_i | n_{\boldsymbol{Y}} > 0\right]$$

and $M_{0i}^*(r,n)$ and $M_0^*(r,n)$ to be $M_i^*(r)$ and $M^*(r)$ (respectively) for a stationary, independent process conditional on $n_{\boldsymbol{Y}} = n$. Given a realization of an areal process with $\boldsymbol{Y} = \boldsymbol{y}$ and $n_{\boldsymbol{y}} > 0$ and a set of $\boldsymbol{\ell}_i$s chosen, we define the estimator

$$\widehat{M_i^*}(r,\boldsymbol{y}) = \frac{\mathscr{N}[c(\boldsymbol{\ell}_i^*, r) \cap a_i^c] + A[c(\boldsymbol{\ell}_i^*, r) \cap a_i]}{A[c(\boldsymbol{\ell}_i^*, r) \cap \mathscr{A}]} \left(\frac{\mathscr{N}(\mathscr{A})}{A(\mathscr{A})}\right)^{-1}$$

and

$$\widehat{M^*}(r,\boldsymbol{y}) = \frac{1}{n_{\boldsymbol{y}}} \sum_{i:y_i=1} \widehat{M_i^*}(r,\boldsymbol{y})$$

We then define the test statistic $T_n^*(r,\boldsymbol{y}) = \widehat{M^*}(r,\boldsymbol{y}) - M_0^*(r,n)$ and estimate its null distribution using Monte Carlo simulations.

WEB TABLE 1. *Simulation study results for study area $\mathscr{A}_1$ (the regular grid). Results displayed include the empirical rejection rate (ERR) of the positive area proportion function (PAPF), global Moran's I statistic (MI), the Getis-Ord general G statistic (GG), the spatial scan statistic method (SSS), Ripley's K-function (RK), Ripley's D-function (RD) and the average nearest neighbor method (ANN). For DGMs $M_1 - M_2$, single tailed test indicative of clustering are denoted with a C, while dispersion tests are denoted with a D. All tests were conducted at a level of $\alpha = 0.05$.*

DGM	Method	ERR	Method	ERR										
				Global	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9	r_{10}
I_2	ANN	100.0	RK	100.0	100.0	54.8	95.0	86.8	64.2	1.6	67.0	16.2	7.8	67.2
	SSS	3.0	RD	5.8	6.2	5.6	5.6	4.4	4.4	5.6	4.6	5.4	6.0	4.6
	MI	4.8	PAPF	5.8	6.6	7.6	8.2	10.0	10.6	9.4	8.2	7.4	5.8	5.4
	GG	3.6												
C_1	SSS	84.8	RD	100.0	0.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	MI	70.0	PAPF	72.4	56.2	59.6	74.4	85.6	88.2	90.6	90.8	93.2	94.2	94.6
	GG	66.6												
C_3	SSS	100.0	RD	100.0	0.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	MI	100.0	PAPF	100.0	99.0	98.8	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	GG	99.4												
C_4	SSS	100.0	RD	100.0	0.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	MI	100.0	PAPF	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	GG	100.0												
C_5	SSS	100.0	RD	0.8	0.0	4.0	4.0	1.8	1.8	0.0	2.0	1.4	1.0	1.0
	MI	100.0	PAPF	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	GG	100.0												
C_7	SSS	36.4	RD	49.2	14.8	64.4	64.4	49.8	49.8	35.4	31.2	29.6	28.2	25.4
	MI	68.4	PAPF	55. 8	57.4	61.0	67.4	62.4	54.0	46.8	32.2	29.4	27.4	26.2
	GG	68.2												
C_9	SSS	60.2	RD	93.6	36.2	99.2	99.2	95.2	95.2	83.4	73.8	62.8	60.2	55.6
	MI	99.2	PAPF	98.6	95.6	96.6	97.2	96.8	92.0	87.0	78.4	65.8	59.2	52.2
	GG	98.8												
C_{10}	SSS	95.4	RD	100.0	46.6	100.0	100.0	100.0	100.0	98.2	97.6	90.6	76.6	72.6
	MI	100.0	PAPF	100.0	100.0	100.0	100.0	100.0	100.0	99.8	99.2	96.8	91.8	82.0
	GG	100.0												
C_{11}	SSS	74.1	RD	46.4	18.0	55.2	55.2	44.8	44.8	29.2	37.6	29.8	27.4	23.0
	MI	100.0	PAPF	100.0	100.0	100.0	100.0	100.0	99.6	98.0	91.4	80.4	73.2	69.6
	GG	100.0												
D_1	MI	100.0	RD	99.6	2.4	100.0	100.0	66.8	66.8	47.2	31.6	28.2	17.0	20.8
			PAPF	13.0	96.6	99.4	99.0	86.2	55.6	38.0	29.4	19.0	19.8	15.8
D_3	MI	100.0	RD	100.0	47.8	100.0	100.0	73.0	73.0	58.6	55.6	39.6	29.6	22.8
			PAPF	51.6	100.0	100.0	100.0	99.8	83.2	86.2	68.8	60.8	58.6	59.0
M_2	MI D	0.0	RD D	0.2	52.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	MI C	19.0	RD C	100.0	0.0	25.6	25.6	100.0	100.0	99.8	100.0	99.6	88.4	75.0
	SSS C	0.0	PAPF D	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	GG C	5.8	PAPF C	100.0	21.6	30.8	81.8	99.4	100.0	100.0	100.0	100.0	99.8	97.8

WEB TABLE 2. *Simulation study results for study area $\mathscr{A}_2$ (the US counties). Results displayed include the empirical rejection rate (ERR) of the positive area proportion function (PAPF), global Moran's I statistic (MI), the Getis-Ord general G statistic (GG), the spatial scan statistic method (SSS), Ripley's K-function (RK), Ripley's D-function (RD) and the average nearest neighbor method (ANN). For DGMs $M_1 - M_2$, single tailed test indicative of clustering are denoted with a C, while dispersion tests are denoted with a D. All tests were conducted at a level of $\alpha = 0.05$.*

DGM	Method	ERR	Method	ERR										
				Global	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9	r_{10}
I_2	ANN	10.0	RK	100.0	95.4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	SSS	4.2	RD	10.8	1.8	10.0	6.2	7.2	7.2	7.2	7.8	7.0	6.6	7.0
	MI	4.0	PAPF	7.6	7.4	4.6	6.8	7.6	8.2	8.6	9.2	7.6	5.8	6.2
	GG	6.4												
C_1	SSS	100.0	RD	100.0	2.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	MI	99.0	PAPF	100.0	97.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	GG	98.8												
C_3	SSS	100.0	RD	100.0	1.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	MI	100.0	PAPF	100.0	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	GG	100.0												
C_4	SSS	100.0	RD	100.0	1.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	MI	100.0	PAPF	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	GG	100.0												
C_5	SSS	100.0	RD	100.0	0.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	MI	100.0	PAPF	100.0	0.0	20.0	92.4	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	GG	100.0												
C_7	SSS	47.4	RD	59.0	1.8	80.4	34.6	14.0	14.6	13.8	9.0	8.8	9.8	9.8
	MI	100.0	PAPF	100.0	7.6	77.0	36.0	17.4	14.8	11.2	12.0	10.2	9.6	9.2
	GG	100.0												
C_9	SSS	78.4	RD	40.4	1.6	90.4	34.4	25.6	19.4	20.4	18.6	17.0	16.4	17.6
	MI	100.0	PAPF	93.0	8.6	95.2	48.8	26.4	20.0	17.6	15.0	14.6	14.2	12.4
	GG	100.0												
C_{10}	SSS	98.6	RD	92.6	0.6	100.0	62.0	43.0	34.2	32.0	30.0	26.2	23.6	23.2
	MI	100.0	PAPF	100.0	12.4	100.0	80.4	47.8	35.6	28.0	25.6	24.2	23.4	20.6
	GG	100.0												
C_{11}	SSS	100.0	RD	17.8	0.0	13.6	16.6	13.8	13.8	14.8	14.6	14.4	14.4	14.8
	MI	100.0	PAPF	80.6	6.2	85.2	38.6	25.4	19.6	17.4	16.6	15.4	14.4	13.6
	GG	100.0												
D_1	MI	100.0	RD	80.8	0.0	93.4	30.2	11.0	4.2	2.8	2.4	2.8	2.4	2.6
			PAPF	94.6	1.2	98.6	30.4	11.8	6.0	5.8	3.4	3.0	2.4	2.6
D_3	MI	100.0	RD	91.6	0.0	96.4	22.4	2.6	1.6	1.6	0.6	0.6	0.6	0.8
			PAPF	99.8	0.0	100.0	49.2	21.0	12.2	6.6	3.2	1.4	2.4	1.4
M_1	MI D	1.8	RD D	9.0	0.0	2.0	0.0	0.0	0.0	0.6	1.4	5.4	12	27.4
	MI C	2.2	RD C	10.6	0.8	3.8	23.6	7.6	4.2	2.0	0.6	0.0	0.0	0.0
	SSS	100.0	PAPF D	62.6	75.4	4.0	0.0	0.4	1.2	4.4	8.2	13.0	16.2	23.0
	GG	22	PAPF C	36.2	0.0	17.0	51.0	28.8	21.8	15.4	11.4	7.6	4.8	1.0

WEB TABLE 3. *Simulation study results for study area $\mathscr{A}_3$ (the small grid). Results displayed include the empirical rejection rate (ERR) of the positive area proportion function (PAPF), global Moran's I statistic (MI), the Getis-Ord general G statistic (GG), the spatial scan statistic method (SSS), Ripley's K-function (RK), Ripley's D-function (RD) and the average nearest neighbor method (ANN). For DGMs $M_1 - M_2$, single tailed test indicative of clustering are denoted with a C, while dispersion tests are denoted with a D. All tests were conducted at a level of $\alpha = 0.05$.*

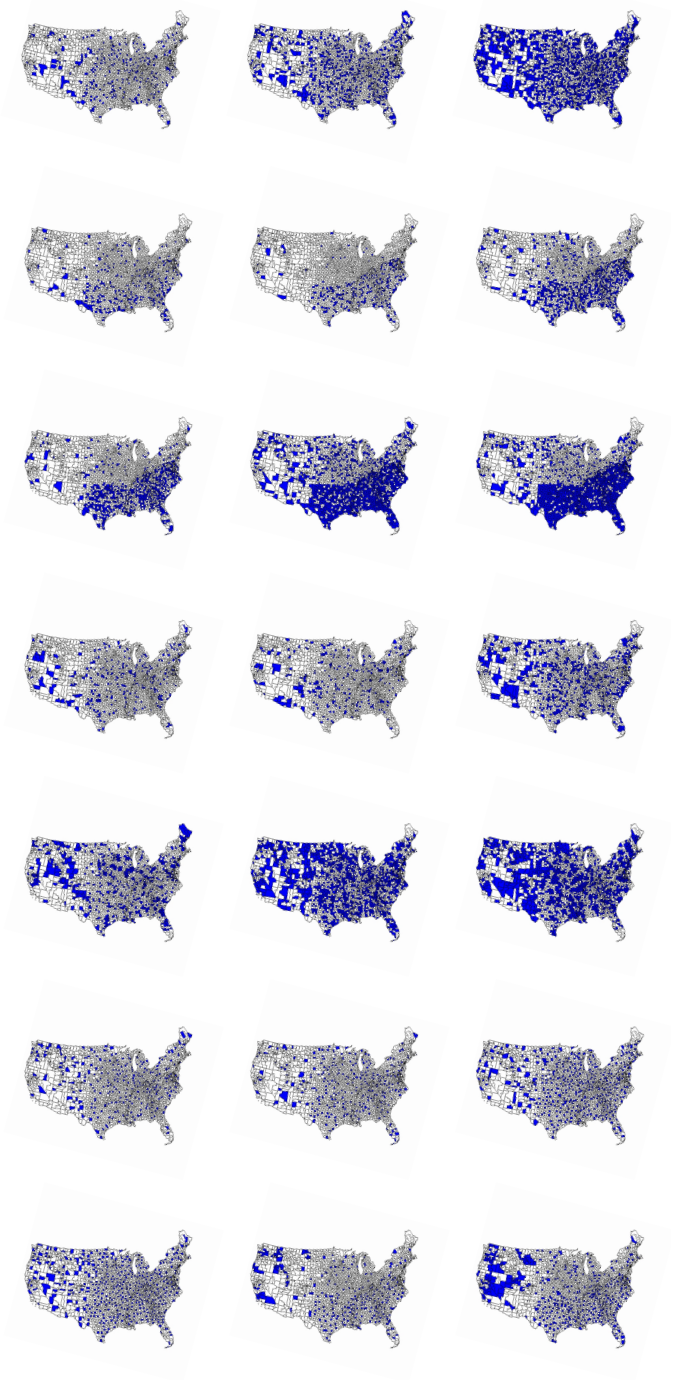
DGM	Method	ERR	Method	ERR										
				Global	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9	r_{10}
I_3	ANN	100.0	RK	37.6	37.6	0.0	0.0	2.8	0.0	0.0	0.4	0.0	4.2	0.2
	SSS	59.3	RD	8.2	5.2	5.2	6.0	9.6	9.6	9.4	9.4	9.2	8.8	8.8
	MI	4.1	PAPF	5.2	2.0	7.0	8.8	10.0	4.0	4.2	7.4	8.6	3.8	7.0
	GG	3.6												
C_5	SSS	33.5	RD	89.2	0.4	5.2	2.8	6	6	6.6	6.6	9.8	9.0	9.0
	MI	40..6	PAPF	38.2	19.4	17.8	33.2	33.4	38.2	36.6	32.8	32.8	31.8	24.6
	GG	20.2												
C_6	SSS	22.3	RD	84.0	0.0	4.4	0.8	2.8	2.8	4.2	4.2	8.6	5.4	5.4
	MI	58.3	PAPF	53.2	33.8	36.8	45.2	50.4	46.8	49.6	53.0	53.0	48.0	45.2
	GG	27.6												
C_{11}	SSS	31.5	RD	94.0	5.8	7.4	5.0	13.4	13.4	13.4	13.4	15.8	15.6	15.6
	MI	31.8	PAPF	36.4	18.4	21.2	29.6	37.2	29.0	29.0	34.0	28.2	36.8	37.4
	GG	27.4												
C_{12}	SSS	22.9	RD	94.6	6.2	8.2	3.6	14.0	14.0	14.8	14.8	17.2	18.0	18.0
	MI	49.1	PAPF	52.8	35.0	37.2	44.4	42.6	47.2	44.4	48.4	50.8	52.2	47.6
	GG	38.6												
D_3	SSS	74.5	RD	15.2	0.0	3.8	10.6	16.6	16.6	22.2	22.2	10.6	10.4	10.4
	MI	13.0	PAPF	10.8	73.4	73.0	44.0	29.8	12.8	11.4	10.2	12.6	8.0	4.8
D_4	SSS	77.7	RD	15.2	0.0	4.0	8.4	17.6	17.6	30.6	30.6	11.8	10.8	10.8
	MI	18.4	PAPF	13.4	73.2	73.6	43.8	28.4	14.4	12.0	15.0	16.0	10.4	8.0

WEB TABLE 4. *Simulation study results for study area $\mathscr{A}_1$ (the regular grid) using randomly selected points rather than centroids. The table displays the empirical rejection rate (ERR) of the positive area proportion function (PAPF). For DGMs $M_1 - M_2$, single-tailed test indicative of clustering are denoted with a C, while dispersion tests are denoted with a D. All tests were conducted at a level of $\alpha = 0.05$.*

DGM	ERR										
	Global	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9	r_{10}
I_1	5.6	5.8	5.4	4.4	6.4	8.2	6.4	5.6	7.4	9.0	8.0
I_3	3.0	8.0	6.4	2.8	2.6	4.0	5.4	4.4	4.4	4.2	4.4
C_2	100.0	82.2	94.8	98.8	99.8	99.8	100.0	100.0	100.0	100.0	100.0
C_6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
C_8	96.6	87.4	96.2	97.6	97.4	93.4	89.2	83.0	76.4	69.4	62.6
C_{12}	100.0	100.0	100.0	100.0	100.0	100.0	100.0	98.8	96.8	90.0	84.6
D_2	100.0	99.8	100.0	100.0	99.2	83.8	74.0	56.2	53.2	46.2	30.6

WEB TABLE 5. *Simulation study results for study area $\mathscr{A}_2$ (the US counties) using randomly selected points rather than centroids. The table displays the empirical rejection rate (ERR) of the positive area proportion function (PAPF). For DGMs $M_1 - M_2$, single-tailed test indicative of clustering are denoted with a C, while dispersion tests are denoted with a D. Due to the computational intensity of this method, the simulation study was conducted using 100 datasets rather than 500. All tests were conducted at a level of $\alpha = 0.05$.*

DGM	ERR										
	Global	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9	r_{10}
I_1	9.6	7.8	7.2	6.4	5.8	8.2	8.8	8.6	7.8	8.4	8.6
I_3	7.0	5.0	6.0	6.0	7.0	7.0	7.0	8.0	8.0	8.0	8.0
C_2	100.0	99.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
C_6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
C_8	98.0	7.8	98.2	72.4	41.2	28.2	22.8	18.6	18.4	16.2	16.8
C_{12}	96.0	9.0	96.0	62.0	28.0	22.0	20.0	15.0	14.0	14.0	16.0
D_2	96.0	2.8	98.0	50.6	16.2	7.8	6.0	4.6	2.8	2.6	2.8



WEB FIG. 1. *Examples of observed units generated under each scenario for study area $\mathscr{A}_2$. The first row displays examples of data generated under the null hypothesis of equal probability sampling without replacement (left to right: I_1, I_2, I_3). Rows 2-3 display examples of data generated with a single cluster (top left to bottom right C_1, C_2, C_3, C_4, C_5, C_6). Rows 4-5 display examples of data generated with multiple clusters (top left to bottom right C_7, C_8, C_9, C_{10}, C_{11}, C_{12}). Rows 6-7 display examples of data generated with dispersion or a mix of clustering and dispersion (top left to bottom right D_1, D_2, D_3, D_4, M_1, M_2).*

WEB TABLE 6. *Results from applying the PAPF method to conservation easements in Boulder County, Colorado. The table displays the estimated 0.025 and 0.975 quantiles of the null distribution used to define the rejection region and the observed test statistic at each considered radius.*

DGM	Radius											
	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9	r_{10}	T_{nC}	T_{nD}
$Q_{0.025}$	-16.17	-1.11	-0.95	-0.90	-0.88	-0.84	-0.82	-0.81	-0.80	-0.78		-2.49
$Q_{0.975}$	14.37	1.02	0.89	0.72	0.68	0.68	0.62	0.62	0.58	0.55	2.63	
$T(r, \boldsymbol{y})$	-10.46	1.39	0.74	0.44	0.22	0.11	0.03	-0.02	-0.05	-0.07	2.92	-1.45

WEB TABLE 7. *Results from applying the PAPF method to the US counties with high rates of childhood overweight/obesity. The table displays the estimated 0.025 and 0.975 quantiles of the null distribution used to define the rejection region and the observed test statistic at each considered radius.*

DGM	Radius											
	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9	r_{10}	T_{nC}	T_{nD}
$Q_{0.025}$	-0.27	-0.07	-0.06	-0.05	-0.05	-0.05	-0.05	-0.05	-0.05	-0.04		-2.13
$Q_{0.975}$	0.29	0.08	0.07	0.07	0.07	0.06	0.06	0.06	0.06	0.06	2.53	
$T(r, \boldsymbol{y})$	0.81	0.95	0.88	0.8	0.74	0.7	0.67	0.65	0.62	0.58	25.83	5.56