

Multi-level Modelling and Spatial Inference for
Large-Scale Neuroimaging Data

Supplementary Material C

Thomas J. Maullin-Sapey

Submitted in support of the thesis “Multi-level Modelling and
Spatial Inference for Large-Scale Neuroimaging Data”.

S1 Overview

This document contains supplementary results for Chapter 5 of the thesis ‘Multi-level Modelling and Spatial Inference for Large-Scale Neuroimaging Data’. The following sections provide a full account of the simulations and real data example described in Chapter 5 of the thesis document. However, unlike in Chapter 5, here we report empirical coverage simulation results for 3 significance levels ($\alpha = 0.05, 0.1$ and 0.2), 9 simulations and 2 different approaches to computing realisations of H during the Wild t -bootstrap.

In particular, as detailed in Section 5.3.2 of Chapter 5, bootstrap samples for H are obtained using the below expression;

$$\tilde{H} = \max_{\alpha \in \mathcal{P}^+(\mathcal{M})} \left(\sup_{s \in \partial^\alpha \mathcal{F}_c} \left| \min_{i \in \alpha} (\tilde{G}^i(s)) \right| \right). \quad (\text{S1})$$

The simulation results detailed in this document consider 2 approaches to evaluating the above supremum taken over $\partial^\alpha \mathcal{F}_c$. In the first approach, the location of $\partial^\alpha \mathcal{F}_c$ is evaluated using the underlying mechanism which was employed for synthetic data generation. In the second approach, the location of $\partial^\alpha \mathcal{F}_c$ is estimated from the synthetic data itself. In the following sections, the former approach is referred to as using the ‘true boundary’, whilst the latter is referred to as using the ‘estimated boundary’. The simulation results in Chapter 5 of the thesis document were generated using $\alpha = 0.05$ with suprema evaluated using $\partial \hat{\mathcal{F}}_c$ (i.e. the ‘estimated boundary’). These settings were chosen for inclusion in Chapter 5 as, in practice, $\partial \mathcal{F}_c$ (i.e. the ‘true boundary’) is unknown, and $\alpha = 0.05$ is a popular choice of tolerance level.

This document is organized as follows. Section S2 provides practical details for implementation of the method, alongside pseudocode. Next, Section S3 provides an overview of the process used to obtain the real data employed in Section 5.5 of Chapter 5. After this, Section S4 describes the 6 simulations that were not outlined in full in Chapter 5. Section S5 provides empirical coverage results obtained by using the estimated boundary to evaluate Equation (S1), with Sections S5.1, S5.2 and S5.3 providing results for $\alpha = 0.2, 0.1$ and 0.05 , respectively. Similarly, Section S6 provides empirical coverage results obtained by using the true boundary to evaluate Equation (S1) with Sections S6.1, S6.2 and S6.3 providing results for $\alpha = 0.2, 0.1$ and 0.05 , respectively. Following this, Section S7 provides an overview of the computation times observed for each of the simulations.

S2 Implementation Details

S2.1 Modelling on a Discrete Lattice

Throughout Chapter 5 of the thesis document, we assumed that the target and estimator functions are defined on a compact region of continuous space, S . In practice, however, observations are sampled on a discrete subset of S , typically a lattice. As noted by Bowring et al. [2019], great care must be exhibited when data are sampled on a lattice, as interpolation must be performed to estimate the location of the boundary $\partial\hat{\mathcal{F}}_c$.

The interpolation process begins by first identifying pairs of neighbouring gridpoints, $s_{\mathcal{I}}$ and $s_{\mathcal{O}}$, which lie inside and outside $\hat{\mathcal{F}}_c$, respectively. For each pair of neighbouring gridpoints, interpolation weights, which represent the location of $\partial\hat{\mathcal{F}}_c$, are then obtained for each study condition using the estimators $\{\hat{\mu}_n^i\}_{i \in \mathcal{M}}$. Following this, in each bootstrap instance, the interpolation weights are used to obtain interpolated bootstrap samples $\{\tilde{G}^i\}_{i \in \mathcal{M}}$ for each pair of gridpoints, $(s_{\mathcal{I}}, s_{\mathcal{O}})$. Using the interpolated bootstrap samples $\{\tilde{G}^i\}_{i \in \mathcal{M}}$, interpolated bootstrap samples of H , $\tilde{H}$, are obtained using Equation (S1). Further detail of this approach may be found in Bowring et al. [2019]. We note here that, to obtain $\tilde{H}$ using Equation (S1), it is also necessary to identify the boundary segments $\{\partial^\alpha \hat{\mathcal{F}}_c\}_{\alpha \in \mathcal{P}^+(\mathcal{M})}$. In practice, this may be achieved via careful consideration of the values of $\{\hat{\mu}^i(s_{\mathcal{O}})\}_{i \in \mathcal{M}}$ for each pair of neighbouring gridpoints, $(s_{\mathcal{I}}, s_{\mathcal{O}})$, in order to identify which of the boundaries $\{\partial\mathcal{A}_c^i\}_{i \in \mathcal{M}}$ lie between $s_{\mathcal{O}}$ and $s_{\mathcal{I}}$. This process was employed to obtain the results of the following sections.

S2.2 Pseudocode

Algorithm 1: Pseudocode to generate Confidence Regions for LM Conjunction Inference

for each study condition, $i \in \mathcal{M}$, and every gridpoint, s **do**

- Compute estimator function, $L^{i'} \hat{\beta}^i(s) = L^{i'} (X^{i'} \Sigma^i(s)^{-1} X^i)^{-1} X^{i'} \Sigma^i(s)^{-1} Y^i(s)$.
- Compute decorrelated residual vector, $R^i(s) = [\Sigma^i(s)]^{-\frac{1}{2}} (Y^i(s) - X^i \hat{\beta}^i(s))$.

end

Compute intersection set $\hat{\mathcal{F}}_c$ as the set of gridpoints where $\min_{i \in \alpha} L^{i'} \hat{\beta}^i(s) \geq c$ and identify boundary segments $\{\partial^\alpha \hat{\mathcal{F}}_c\}_{\alpha \in \mathcal{P}^+(\mathcal{M})}$.

for each bootstrap instance, instance b **do**

- Generate n Rademacher variables, $\{r_l\}_{l \in \{1, \dots, n\}}$.
- for** each boundary segment, $\partial^\alpha \hat{\mathcal{F}}_c$ **do**
 - for** every neighbouring gridpoint of $\partial^\alpha \hat{\mathcal{F}}_c$, s , and each study condition, $i \in \alpha$ **do**
 - Compute bootstrap residual mean, $\overline{R}^i(s) = \sum_{l=1}^n \frac{r_l R_l^i(s)}{n}$.
 - Compute bootstrap residual standard error, $\hat{\sigma}^{*,i}(s) = \sqrt{\frac{\sum_{l=1}^n (r_l R_l^i(s) - \overline{R}^i(s))^2}{n-1}}$.
 - Compute bootstrap instance of $G^i(s)$, $\tilde{G}^i(s) = n^{-\frac{1}{2}} \sum_{l=1}^n r_l \frac{R_l^i(s)}{\hat{\sigma}^{*,i}(s)}$.
 - end**
 - for** every neighbouring pair of gridpoints lying inside and outside of $\partial^\alpha \hat{\mathcal{F}}_c$, $(s_{\mathcal{I}}, s_{\mathcal{O}})$, and each study condition, $i \in \alpha$ **do**
 - Interpolate $\tilde{G}^i(s_{\mathcal{I}})$ and $\tilde{G}^i(s_{\mathcal{O}})$ to obtain bootstrap instance $\tilde{G}^i(s_{\mathcal{I}}, s_{\mathcal{O}})$.
 - end**
- end**
- Compute bootstrap instance of H , $\tilde{H}_b = \max_{\alpha \in \mathcal{P}^+(\mathcal{M})} (\max_{\{(s_{\mathcal{I}}, s_{\mathcal{O}})\}} |\min_{i \in \alpha} \tilde{G}^i(s_{\mathcal{I}}, s_{\mathcal{O}})|)$.

end

Estimate a as the $(1 - \alpha)^{th}$ quantile of the bootstrap instances $\{\tilde{H}_b\}$.

Construct $\hat{\mathcal{F}}_c^+$ and $\hat{\mathcal{F}}_c^-$ using the estimate of a .

S3 Real Data Methods

In this section, we detail the imaging acquisition protocol, task paradigm, preprocessing stages and first-level analysis used to generate the dataset employed in Section 5.5 of Chapter 5. In the experiment, each subject performed two tasks spread across two runs where each run contained four blocks (one for each stimuli type, e.g. faces, places, tools and body parts). During each block, the subject undertook either a 2-back memory task or a 0-back control task. The experimental design was arranged such that, in each run, two blocks were designated to the 2-back memory task, and two blocks were designated to the 0-back control task. In total, there were eight blocks across the two runs, one for every combination of task and stimuli type (e.g. ‘0-back and faces’, ‘2-back and places’,...). All image acquisitions were obtained using a 32 channel head coil on a modified 3T Siemens Skyra scanner via a gradient-echo EPI sequence with TR= 720ms, TE= 33.1ms, 2.0mm slice thickness, 72 slices, 2.0mm isotropic voxels, 208mm×180mm FOV, and a multi-band acceleration factor of 8.

Using FSL’s ‘FMRIB’s Improved Linear Model’ (FILM) toolbox (Woolrich et al. [2001]), for each of the 80 subjects, a subject-level analysis which included observations for all four stimuli types across both tasks was conducted. In each subject-level analysis, eight explanatory variables (one for each task-stimuli combination), alongside temporal derivative confound terms, were regressed onto Blood Oxygenation Level Dependent (BOLD) response. Each of the eight task-stimuli explanatory variables was convolved with FSL’s default double-gamma haemodynamic response function and temporally filtered using a high pass frequency cutoff of 200s. The functional data were subjected to identical temporal filtering, as well as prewhitening to remove autocorrelation.

Following the subject-level regression, a Contrast Parameter Estimate (CoPE) map was generated for each of the four stimuli types. Each CoPE map represented the average difference in BOLD response between a subject performing the 2–back task for a given stimuli type and the corresponding 0–back task. The 320 resultant CoPE maps (one for each of the four stimuli types across 80 subjects) were then registered to MNI space. The resultant 320 registered CoPE maps formed the dataset employed in the analysis of Section 5.5 of the thesis document.

S4 Simulation Settings

In Section 5.4.1 of Chapter 5, three simulations were described in full detail, and reference was made to a further six. Here, we describe the remaining six simulations (Simulations 4-9), the results of which are presented in Sections S5 and S6.

Simulation 4 was conducted to ensure that the method was not affected by the degree to which two sets shared a common boundary (e.g. the length of $\partial^{\{1,2\}}\mathcal{F}_c$). For this reason, Simulation 4 was similar to Simulation 1 but employed squares instead of circles for data generation. This ensured that $\mathcal{A}_c^1$ and $\mathcal{A}_c^2$ shared boundary segments at the top and bottom of $\mathcal{F}_c$ (c.f. Fig. S1 (a)). Simulation 5 aimed to assess how robust the method was when spatial structure was present in the noise. In this simulation, following the example of Bowring et al. [2019], ϵ^2 was simulated using a heterogeneous linear ramp (c.f. Fig. S1 (b)).

Simulation 6 aimed to assess how robust the empirical coverage was to the boundary segments $\partial^{\{1\}}\mathcal{F}_c$ and $\partial^{\{2\}}\mathcal{F}_c$ having differing lengths. This simulation was similar to Simulation 4, except the height and width of $\mathcal{A}_c^2$ were reduced by a factor of 2/3. Reducing the height and width of $\mathcal{A}_c^2$ in this manner ensured that $\partial^{\{2\}}\mathcal{F}_c$ increased in length as the squares were moved closer together whilst the length of $\partial^{\{1\}}\mathcal{F}_c$ remained constant. Simulation 7 aimed to assess how increasing the variance of ϵ^2 affected the empirical coverage. This simulation employed squares for generating μ^1 and μ^2 . In this simulation, the standard deviation of ϵ^1 was held fixed at 1 whilst the standard deviation of ϵ^2 varied from 1 to 3 in 0.2 increments.

The first seven simulations consider only the case in which $M = 2$. To assess the method's performance for higher values of M , Simulation 8 allowed M to vary from 2 to 5. For this simulation, the functions $\mu^1, \dots, \mu^M$ were generated using overlapping arrangements of circles similar to that employed by Simulation 1 (c.f. Fig. S1 (c)). In Simulation 8, all non-empty boundary segments, $\partial^\alpha\mathcal{F}_c$, satisfied $|\alpha| \leq 2$. To validate our method in the situation in which the boundary segments result from higher-order combinations of excursion set boundaries (i.e. $|\alpha| > 2$), Simulation 9 was run. In this simulation, M was held fixed at $M = 3$, and data was generated using two large squares with a smaller square nested perfectly within their intersection (see Fig. S1 (d)). This setup ensured that $\partial^{\{1,2,3\}}\mathcal{F}_c$ was non-empty. Unlike in the previous simulations, only the number of observations was varied for Simulation 9.

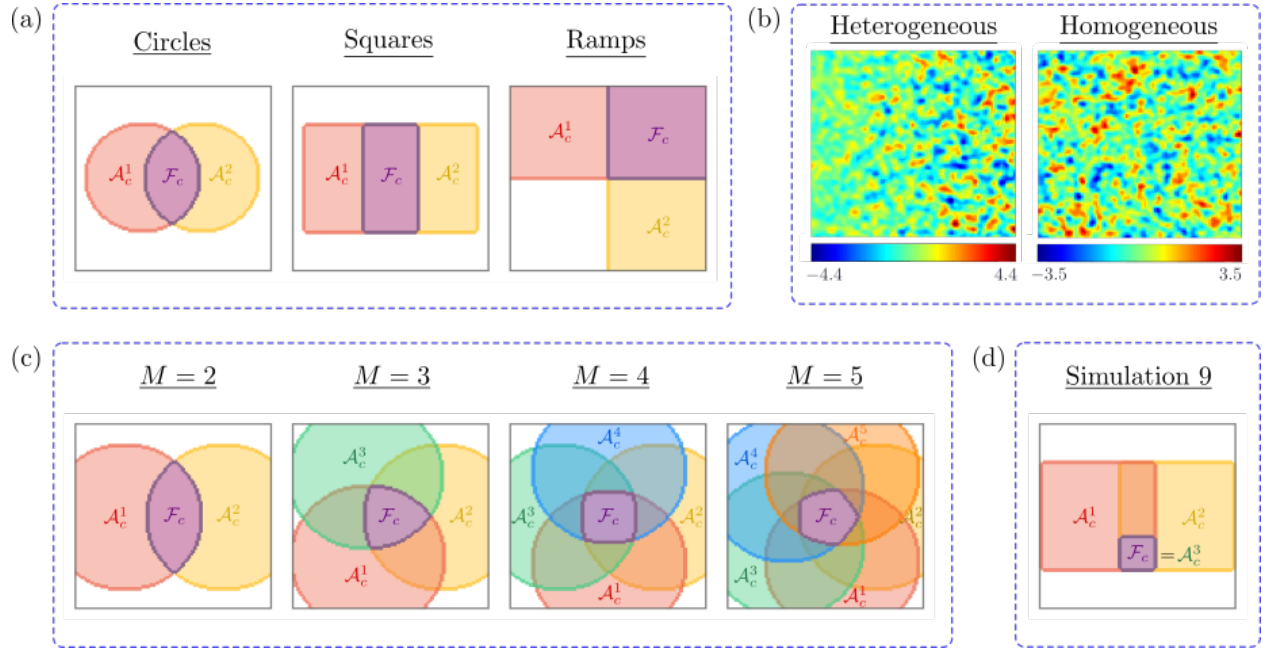


Figure S1: Binary images, alongside example noise realizations, employed for generating synthetic data. (a) The circular binary images, employed for Simulation 1, the square binary images, employed for Simulations 2, 4, 5, and 7, and the binary images that correspond to the ramp signal, employed for Simulation 3. (b) A realization of ϵ^2 generated using the heterogeneous ramp employed in Simulation 5, compared to the homogeneous noise generated for Simulations 1-4 and 6-9. (c) The circular binary images employed for varying the value of M in Simulation 8. (d) The nested square binary images employed for Simulation 9.

S5 Empirical Coverage Along the Estimated Boundary

In this section, we provide the simulation results that were obtained by using the estimated boundary, $\partial\hat{\mathcal{F}}_c$, to evaluate the suprema which appears in Equation (S1). Broadly speaking, the results in the following sections concur that, when the suprema appearing in Equation (S1) was evaluated using $\partial\hat{\mathcal{F}}_c$, the proposed method was robust to a range of secondary factors. Such factors include; signal shape, degree of overlap between excursion sets, spatial structure in the noise, correlation between the error fields, the coinciding of multiple excursion set boundaries, the intersection of excursion sets with the boundary of S , and the number of study conditions.

Notably, the simulations also showed that the method was sensitive to several factors of interest. These were; the lowering of the signals' spatial gradient, the severity of the noise in the data, the lowering of the number of observations, and the lowering of the tolerance level, α . Of these factors, only the latter was not discussed in Chapter 5 of the thesis document. Previous authors have argued that this observation may be due to violations of the inclusion statement, $\hat{\mathcal{F}}_c^+ \subseteq \mathcal{F}_c \subseteq \hat{\mathcal{F}}_c^-$, being miscounted as a result of resolution issues (c.f. Sommerfeld et al. [2018] and Bowring et al. [2019]).

It can be seen that all of the factors which negatively impacted the method's performance tended to cause conservativeness in the generation of confidence regions (e.g. the inclusion statement, $\hat{\mathcal{F}}_c^+ \subseteq \mathcal{F}_c \subseteq \hat{\mathcal{F}}_c^-$, held "too often"). This observation matches the findings of Sommerfeld et al. [2018] and Bowring et al. [2019], and is likely explained by the fact that the estimated boundary segments, $\{\partial^\alpha \hat{\mathcal{F}}_c\}_{\alpha \in \mathcal{P}^+(\mathcal{M})}$, tend to be longer than the true boundary segments, $\{\partial^\alpha \mathcal{F}_c\}_{\alpha \in \mathcal{P}^+(\mathcal{M})}$, for high thresholds. The increased length of $\{\partial^\alpha \hat{\mathcal{F}}_c\}_{\alpha \in \mathcal{P}^+(\mathcal{M})}$ can cause the estimation of $\sup_{s \in \partial^\alpha \mathcal{F}_c} |\min_{i \in \alpha}(G^i(s))|$ to be biased upwards, which in turn inflates the estimates of the quantiles of H . The consequential upward bias in the estimation of a can be seen to result in conservative estimates for the confidence regions $\hat{\mathcal{F}}_c^+$ and $\hat{\mathcal{F}}_c^-$.

This section is organized as follows. Section S5.1 provides empirical coverage results for $\alpha = 0.2$, Section S5.2 provides equivalent results for $\alpha = 0.1$ and Section S5.3 provides results for $\alpha = 0.05$. In all plots, nominal coverage is shown as a dotted gray line and a corresponding binomial confidence interval is shaded in blue. All data points displayed are averages taken across 2500 simulation instances, each using 5000 bootstrap realizations.

S5.1 Nominal Coverage: 80%

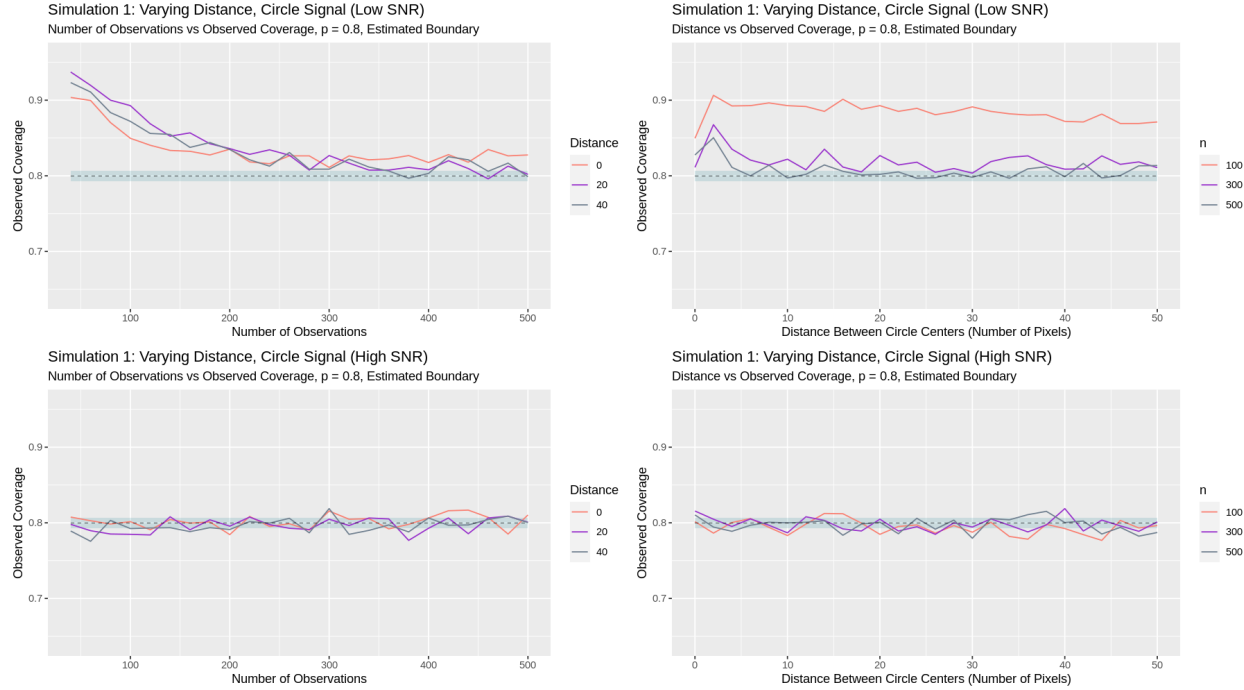


Figure S2: Empirical coverage results for Simulation 1. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for circle separations of 0, 20 and 40 pixels. Right column: Separation between circles vs observed coverage, shown for models including 100, 300 and 500 observations.

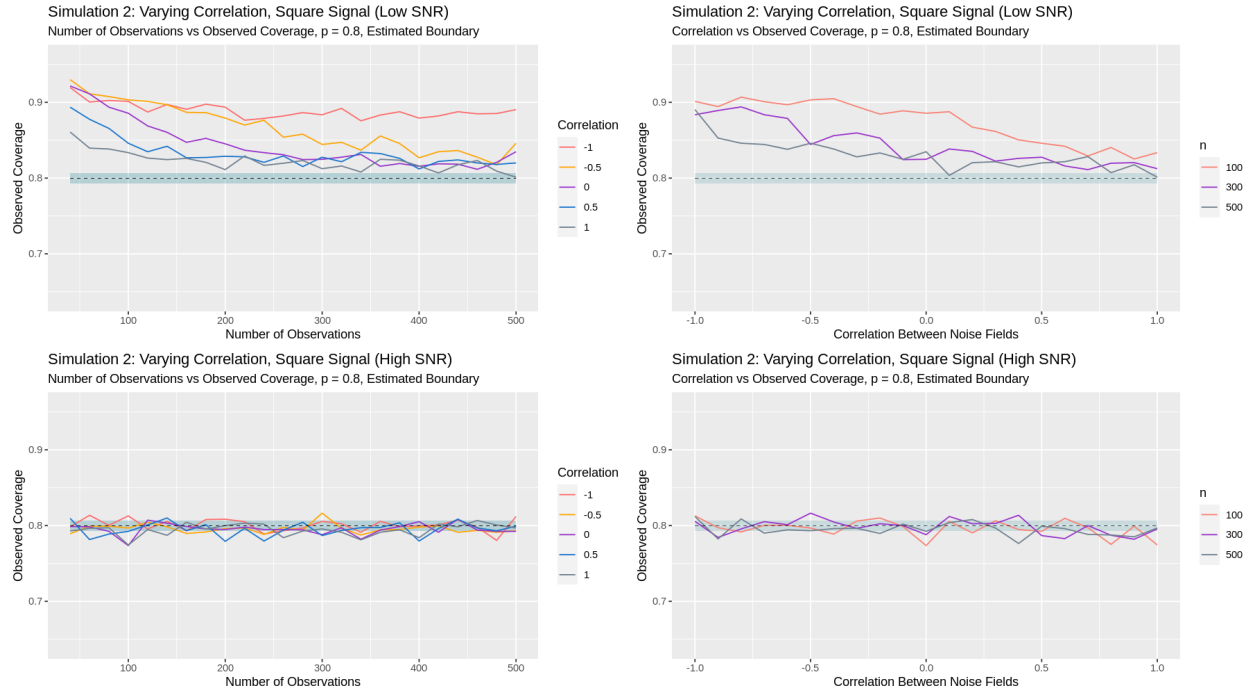


Figure S3: Empirical coverage results for Simulation 2. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for correlations of -1 , -0.5 , 0 , 0.5 and 1 . Right column: Correlation between noise fields vs observed coverage, shown for models including 100, 300 and 500 observations.

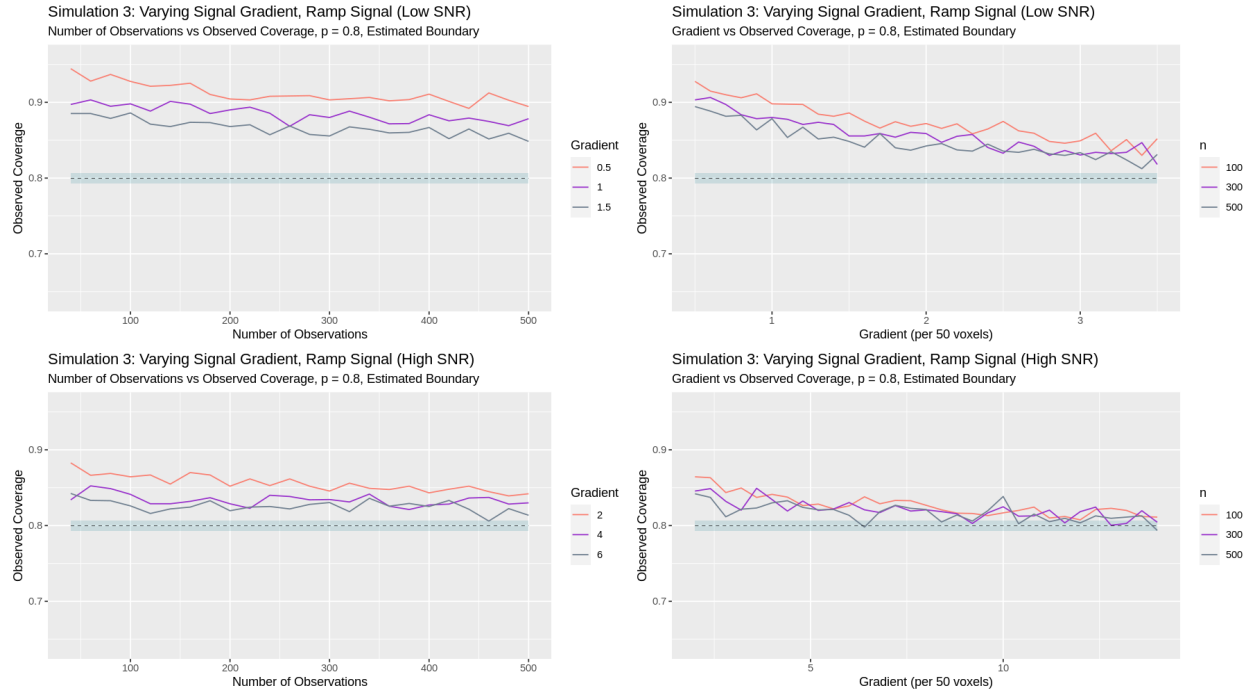


Figure S4: Empirical coverage results for Simulation 3. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for gradients of 0.5, 1 and 1.5 per 50 pixels (top), and 2, 4 and 6 per 50 pixels (bottom). Right column: Ramp gradient per 50 pixels vs observed coverage, shown for models including 100, 300 and 500 observations.

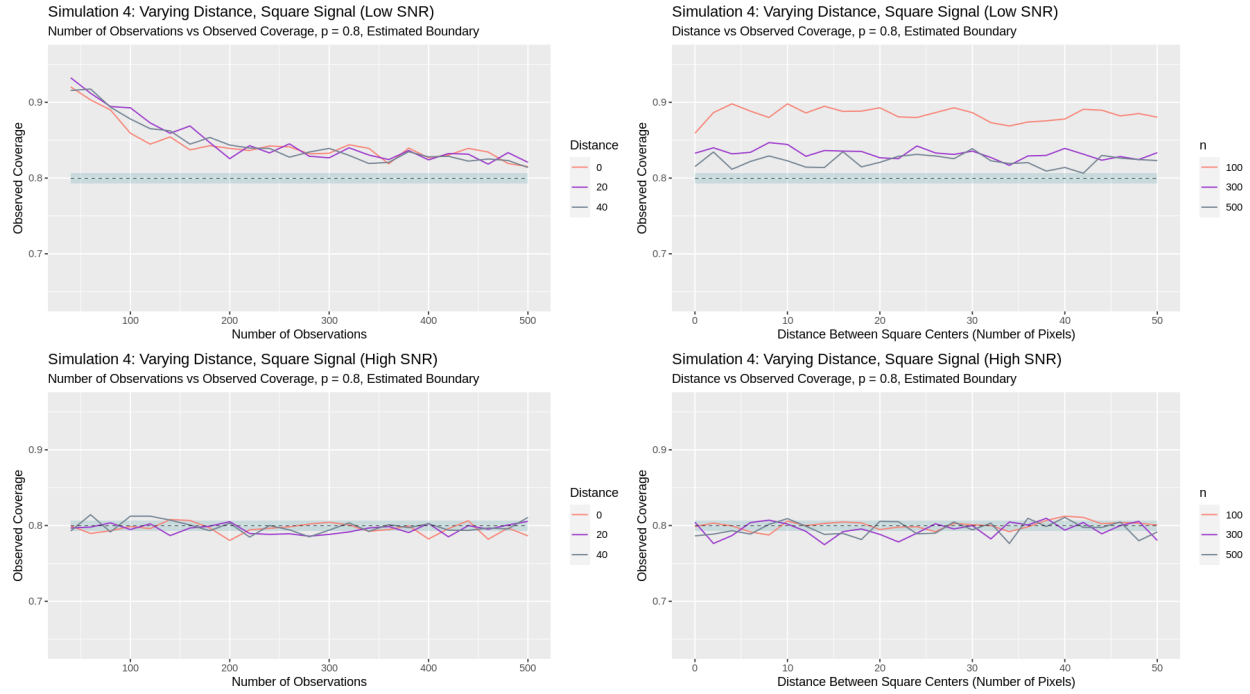


Figure S5: Empirical coverage results for Simulation 4. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for square separations of 0, 20 and 40 pixels. Right column: Separation between squares vs observed coverage, shown for models including 100, 300 and 500 observations.

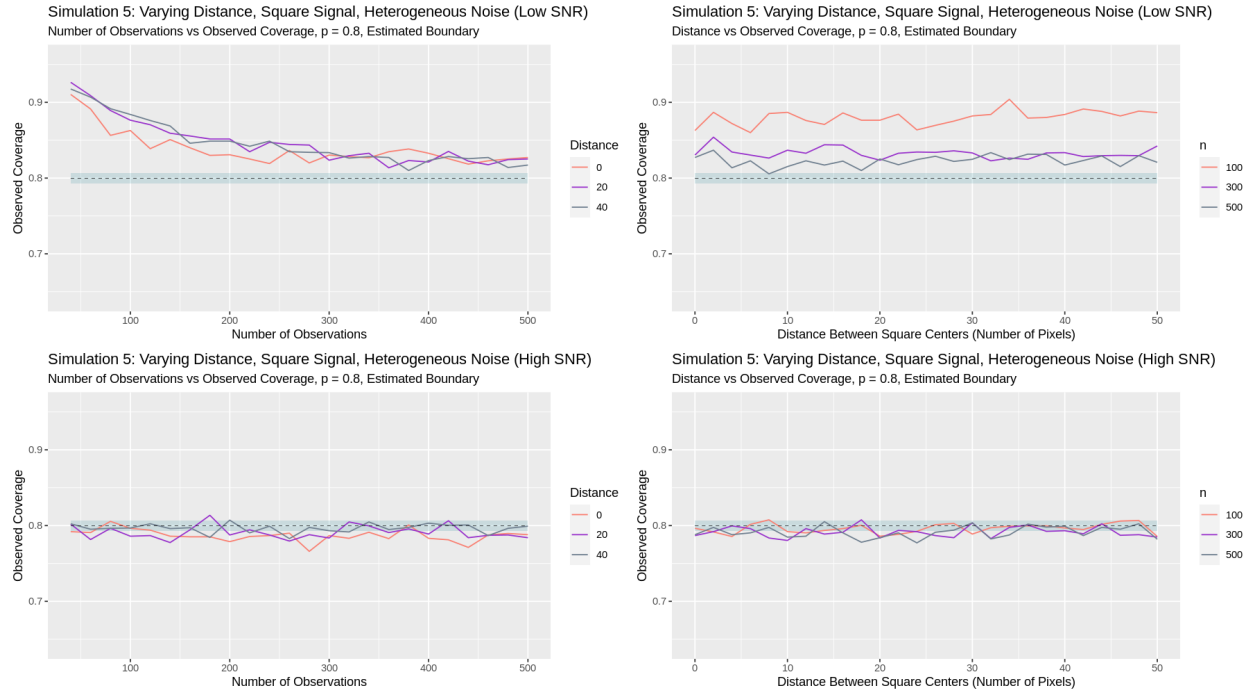


Figure S6: Empirical coverage results for Simulation 5. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for square separations of 0, 20 and 40 pixels. Right column: Separation between squares vs observed coverage, shown for models including 100, 300 and 500 observations.

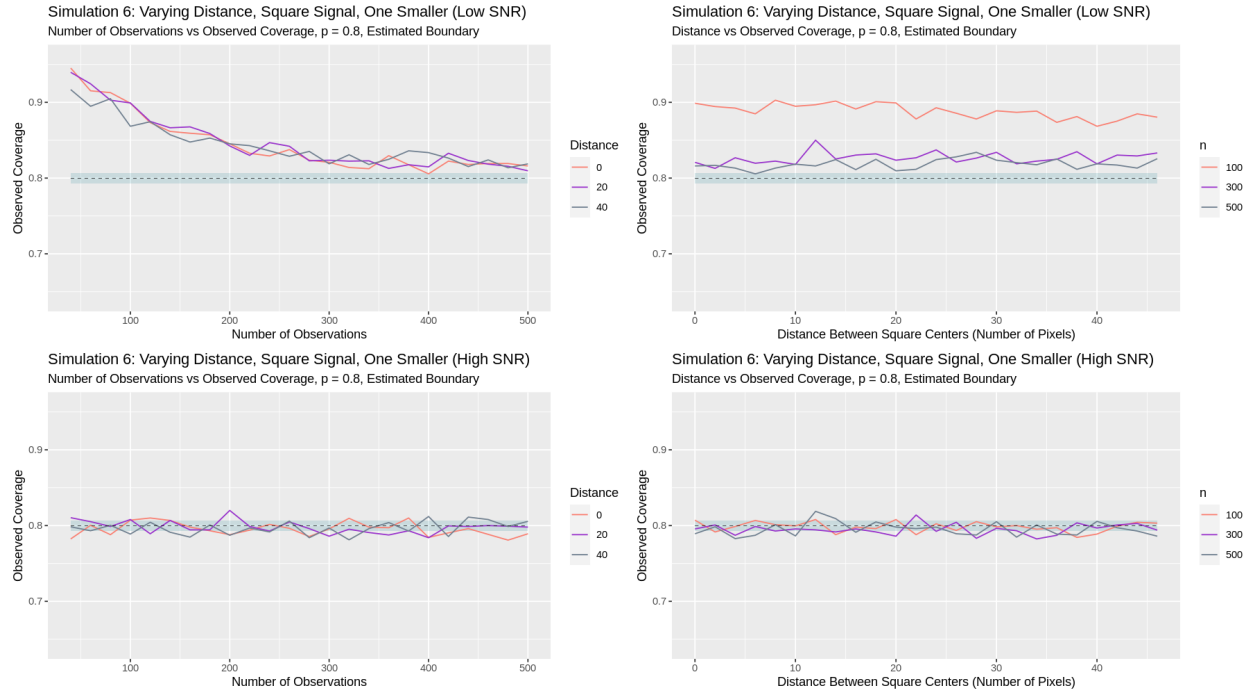


Figure S7: Empirical coverage results for Simulation 6. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for square separations of 0, 20 and 40 pixels. Right column: Separation between squares vs observed coverage, shown for models including 100, 300 and 500 observations.

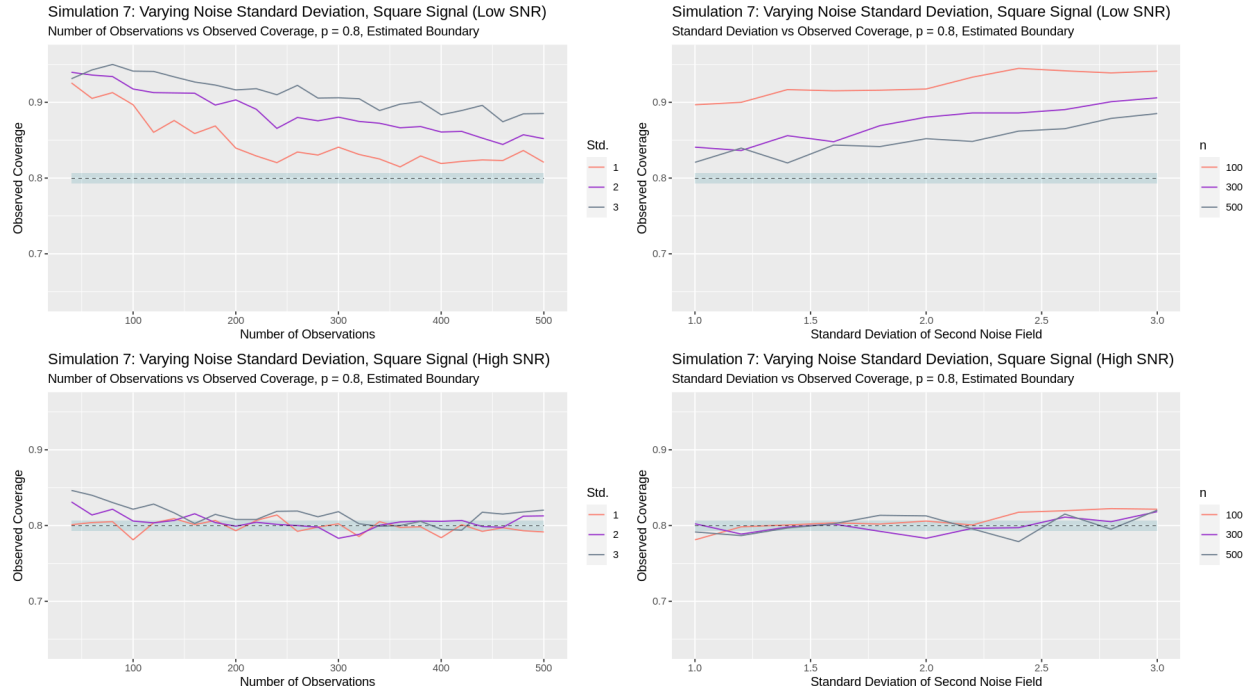


Figure S8: Empirical coverage results for Simulation 7. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for noise standard deviations of 1, 2 and 3. Right column: The standard deviation of the second noise field vs observed coverage, shown for models including 100, 300 and 500 observations.

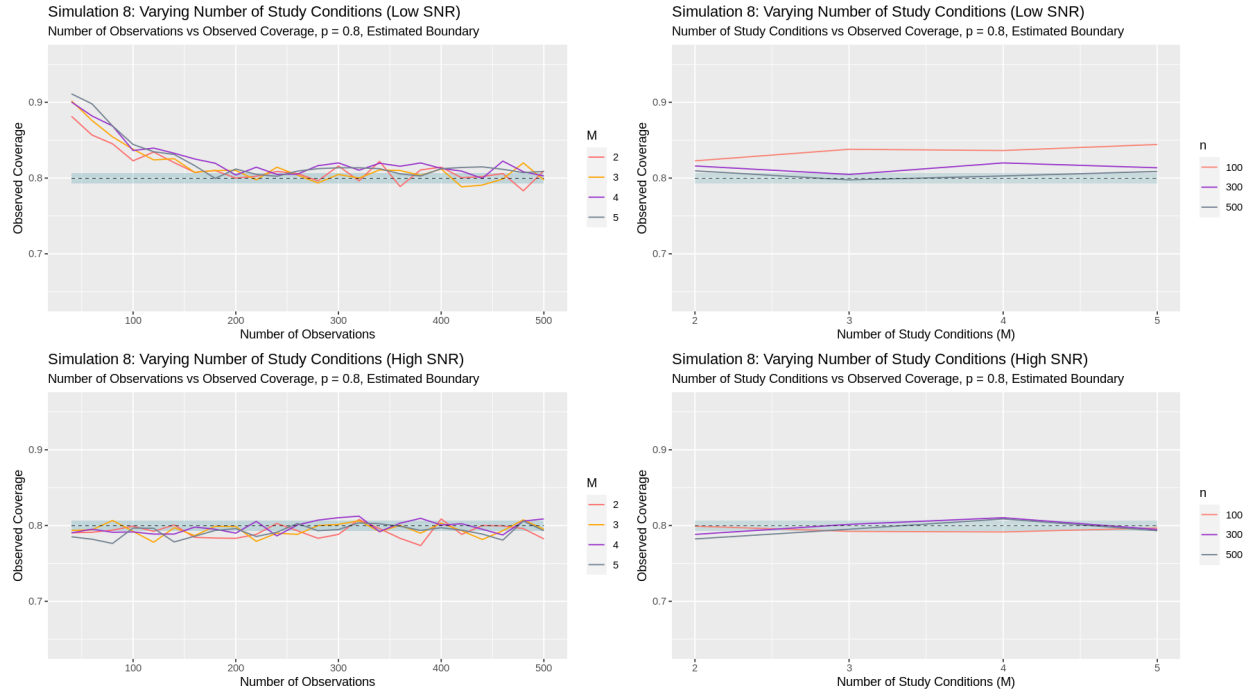


Figure S9: Empirical coverage results for Simulation 8. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for $M = 2, 3, 4$ and 5 . Right column: Number of study conditions (M) vs observed coverage, shown for models including 100, 300 and 500 observations.

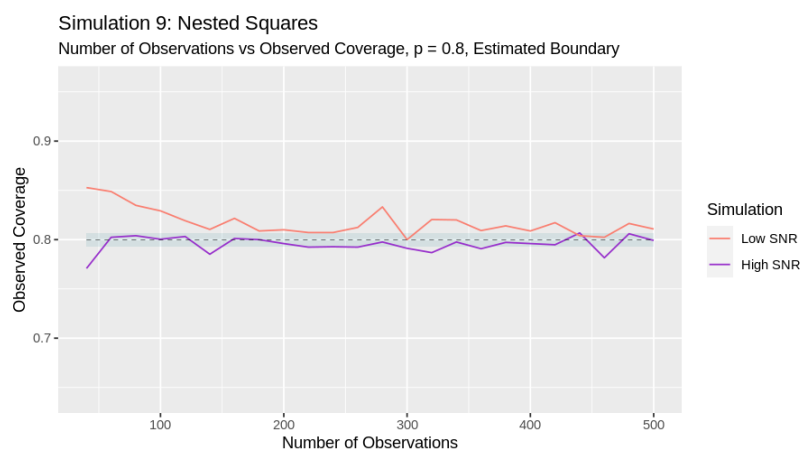


Figure S10: Empirical coverage results for Simulation 9. Results are shown for both low-SNR synthetic data and high-SNR synthetic data. Displayed is the number of observations vs observed coverage.

S5.2 Nominal Coverage: 90%

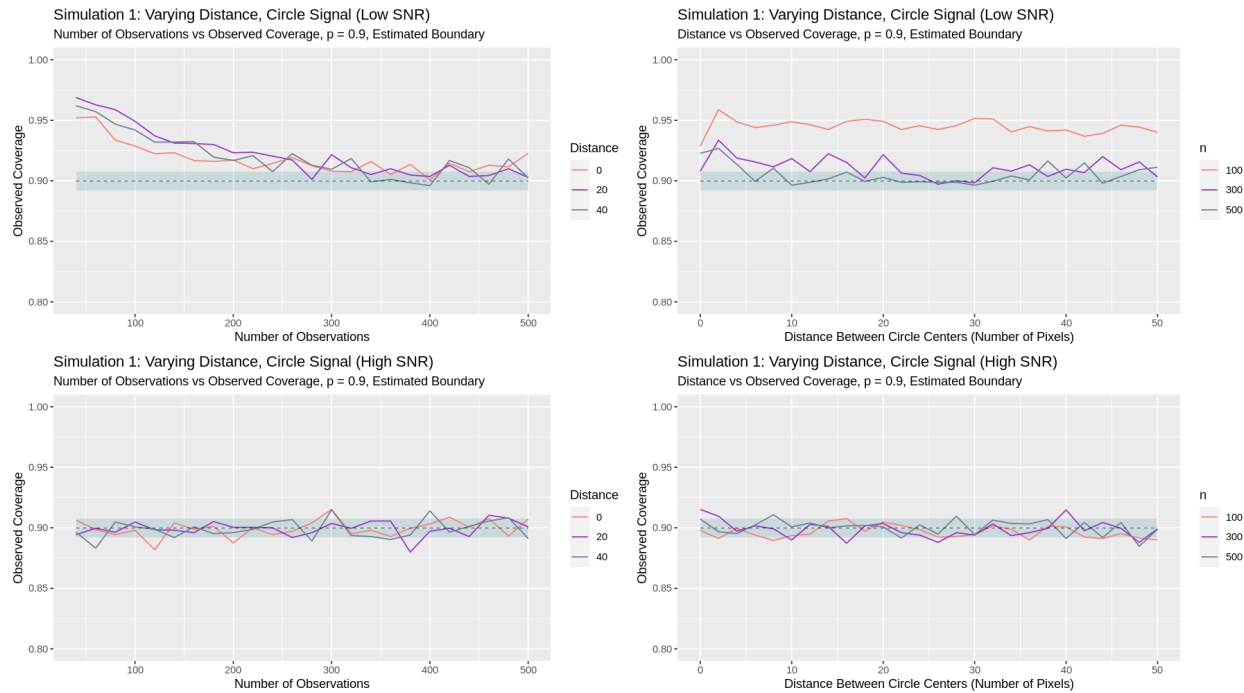


Figure S11: Empirical coverage results for Simulation 1. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for circle separations of 0, 20 and 40 pixels. Right column: Separation between circles vs observed coverage, shown for models including 100, 300 and 500 observations.

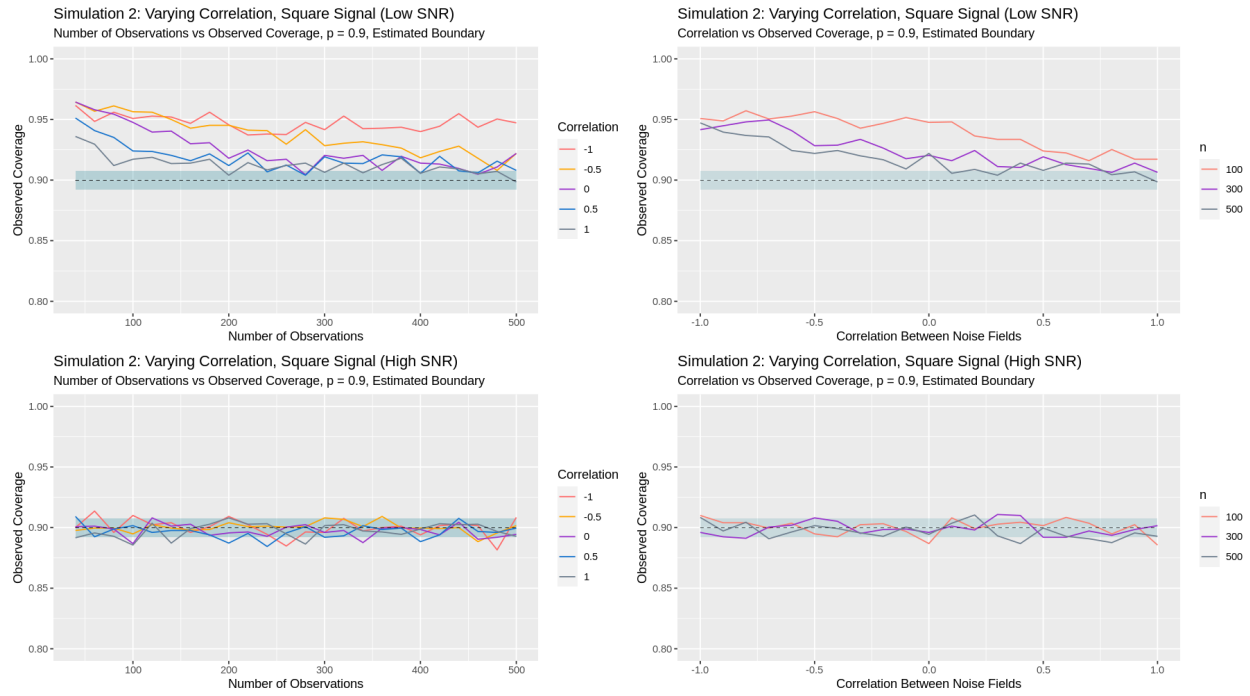


Figure S12: Empirical coverage results for Simulation 2. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for correlations of -1 , -0.5 , 0 , 0.5 and 1 . Right column: Correlation between noise fields vs observed coverage, shown for models including 100, 300 and 500 observations.

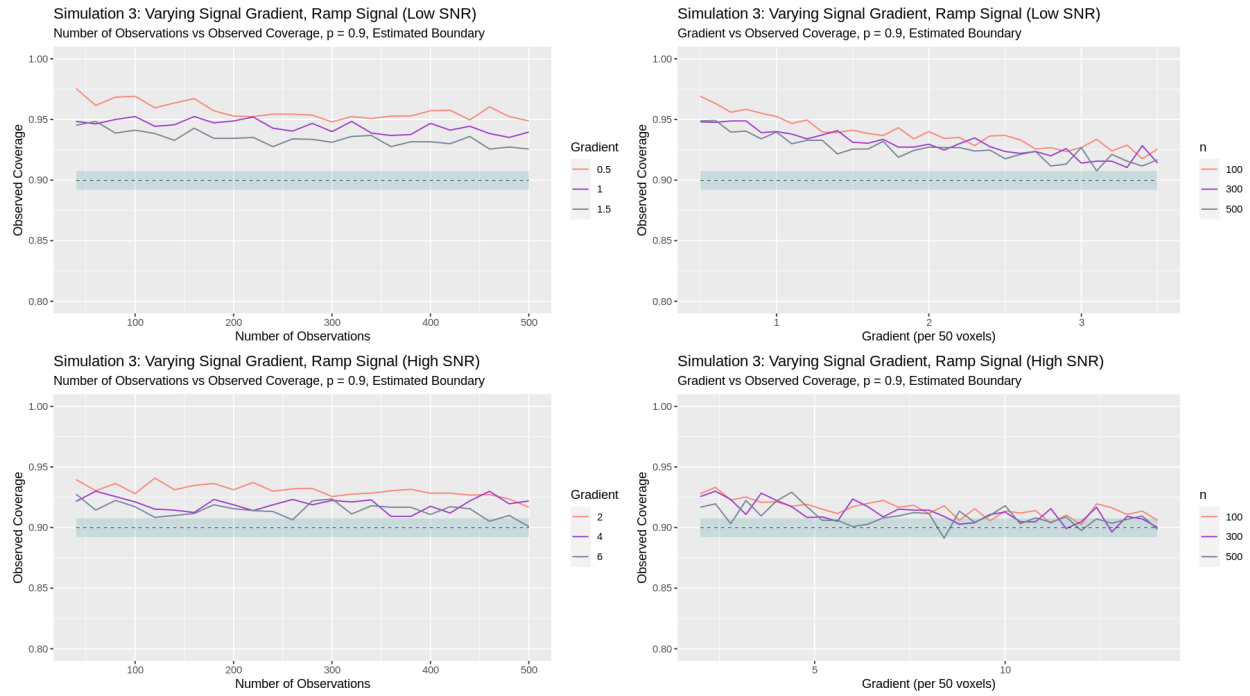


Figure S13: Empirical coverage results for Simulation 3. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for gradients of 0.5, 1 and 1.5 per 50 pixels (top), and 2, 4 and 6 per 50 pixels (bottom). Right column: Ramp gradient per 50 pixels vs observed coverage, shown for models including 100, 300 and 500 observations.

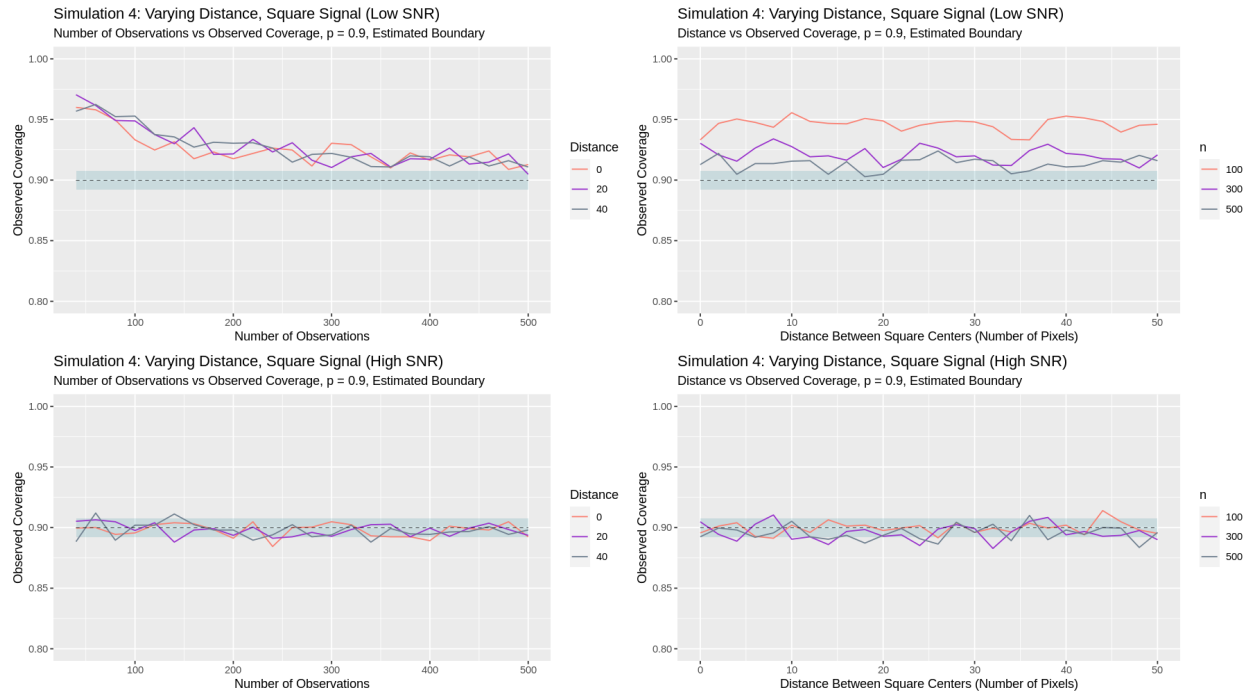


Figure S14: Empirical coverage results for Simulation 4. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for square separations of 0, 20 and 40 pixels. Right column: Separation between squares vs observed coverage, shown for models including 100, 300 and 500 observations.

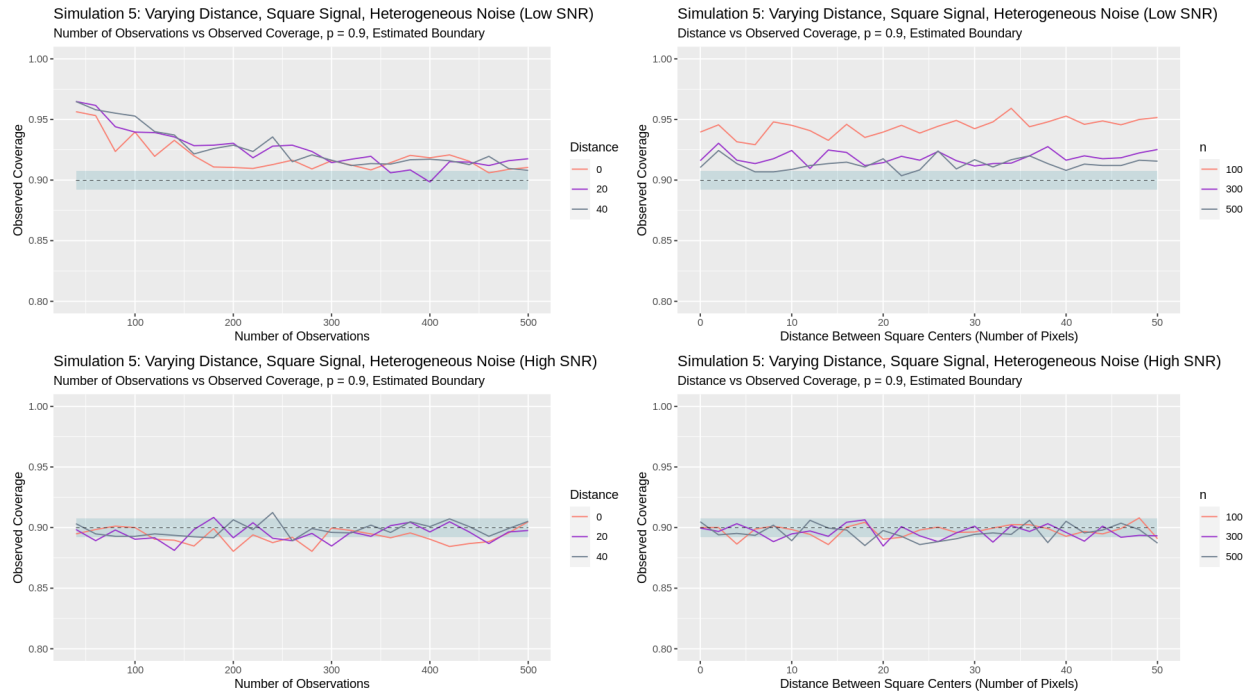


Figure S15: Empirical coverage results for Simulation 5. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for square separations of 0, 20 and 40 pixels. Right column: Separation between squares vs observed coverage, shown for models including 100, 300 and 500 observations.

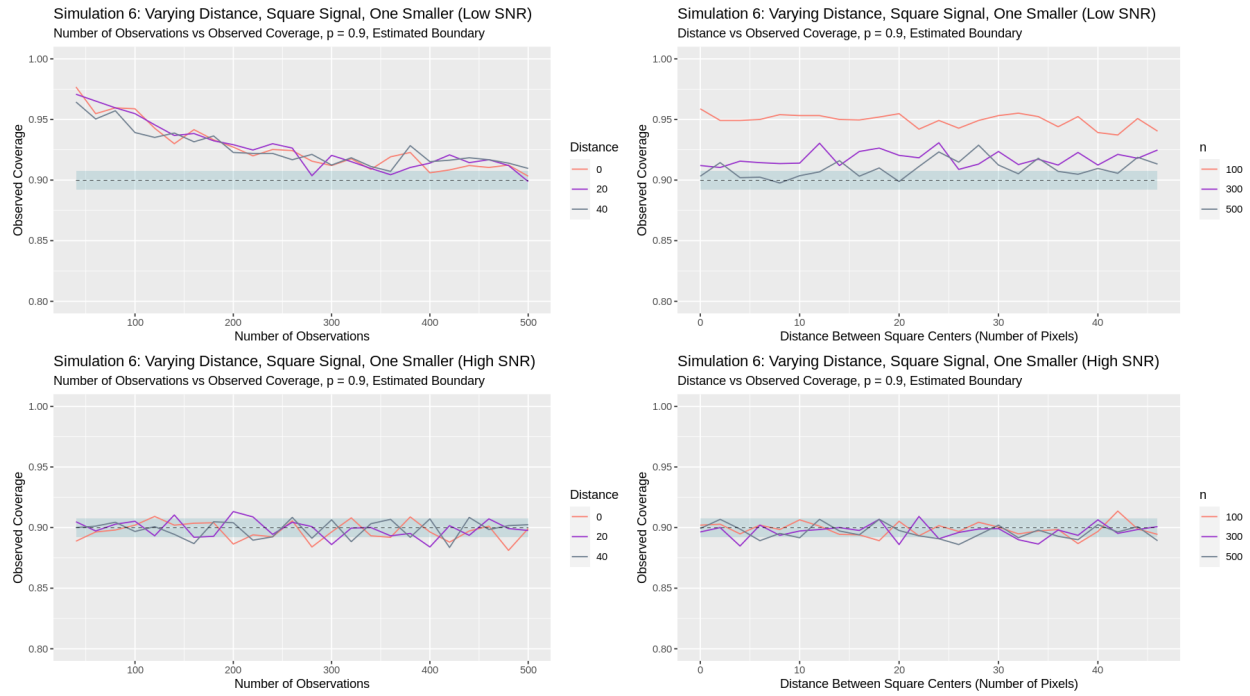


Figure S16: Empirical coverage results for Simulation 6. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for square separations of 0, 20 and 40 pixels. Right column: Separation between squares vs observed coverage, shown for models including 100, 300 and 500 observations.

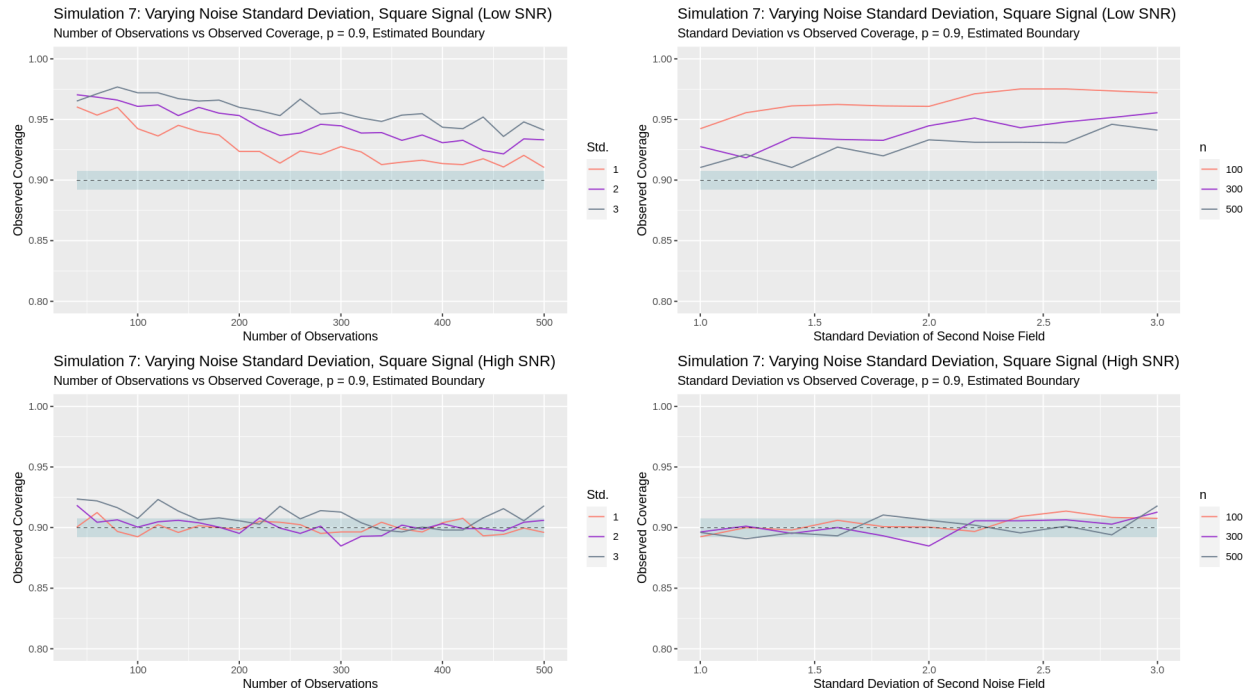


Figure S17: Empirical coverage results for Simulation 7. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for noise standard deviations of 1, 2 and 3. Right column: The standard deviation of the second noise field vs observed coverage, shown for models including 100, 300 and 500 observations.

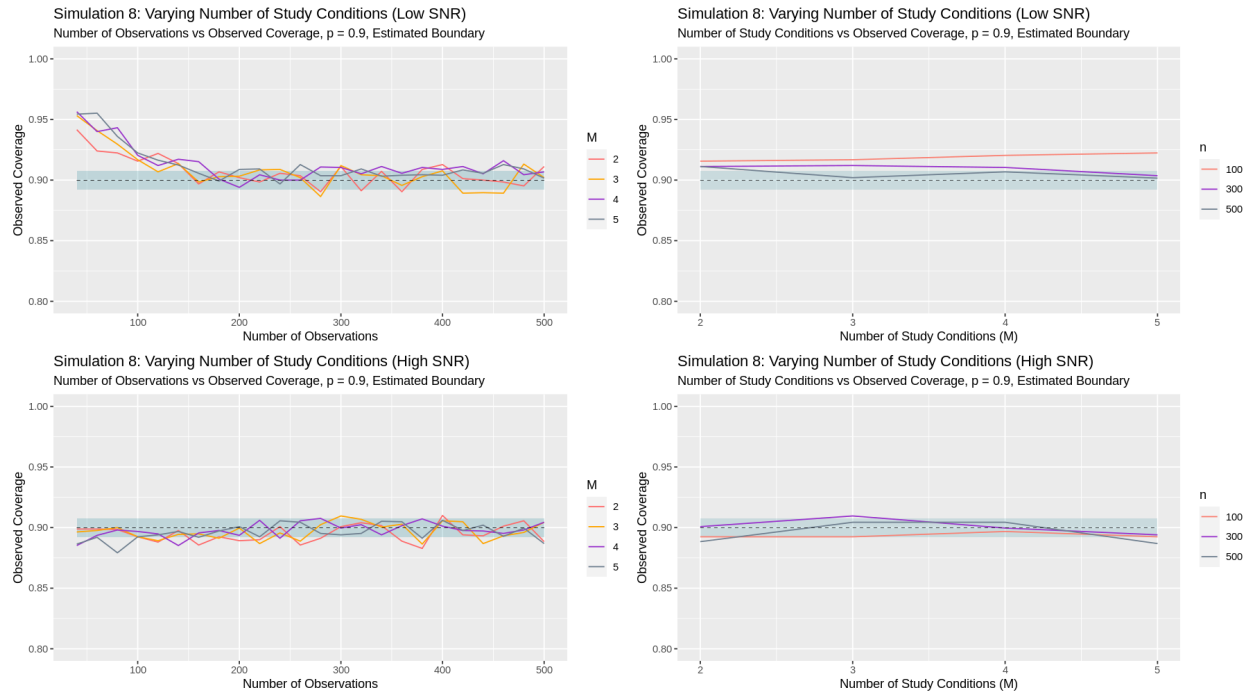


Figure S18: Empirical coverage results for Simulation 8. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for $M = 2, 3, 4$ and 5 . Right column: Number of study conditions (M) vs observed coverage, shown for models including 100, 300 and 500 observations.

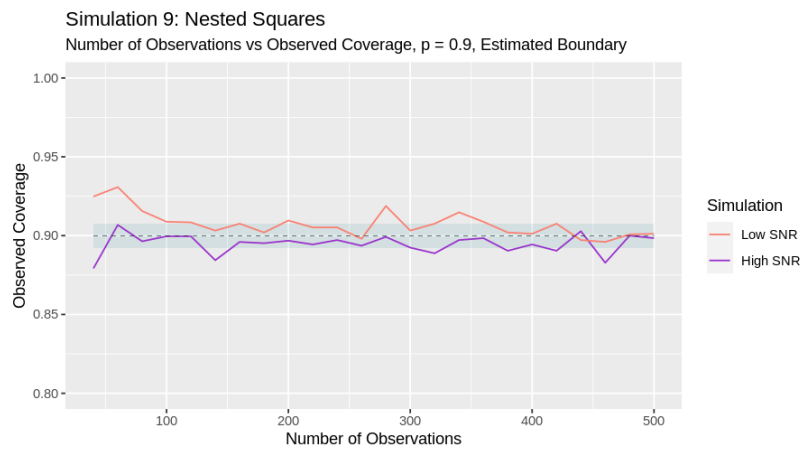


Figure S19: Empirical coverage results for Simulation 9. Results are shown for both low-SNR synthetic data and high-SNR synthetic data. Displayed is the number of observations vs observed coverage.

S5.3 Nominal Coverage: 95%

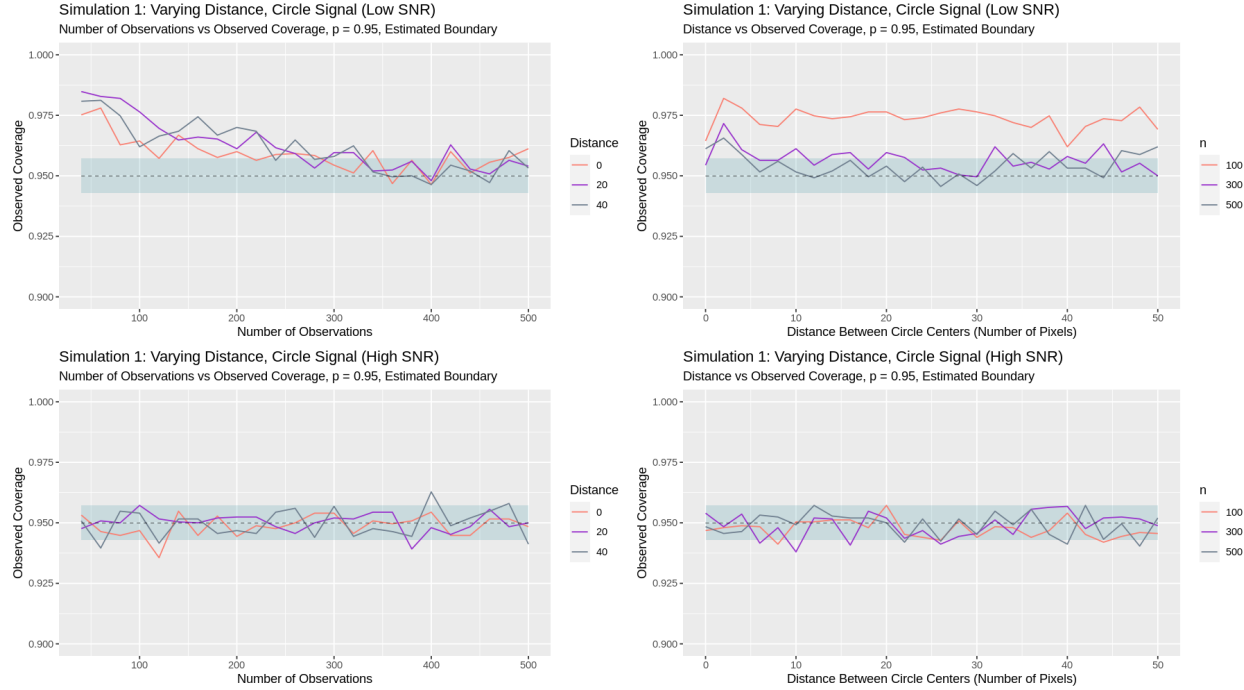


Figure S20: Empirical coverage results for Simulation 1. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for circle separations of 0, 20 and 40 pixels. Right column: Separation between circles vs observed coverage, shown for models including 100, 300 and 500 observations.

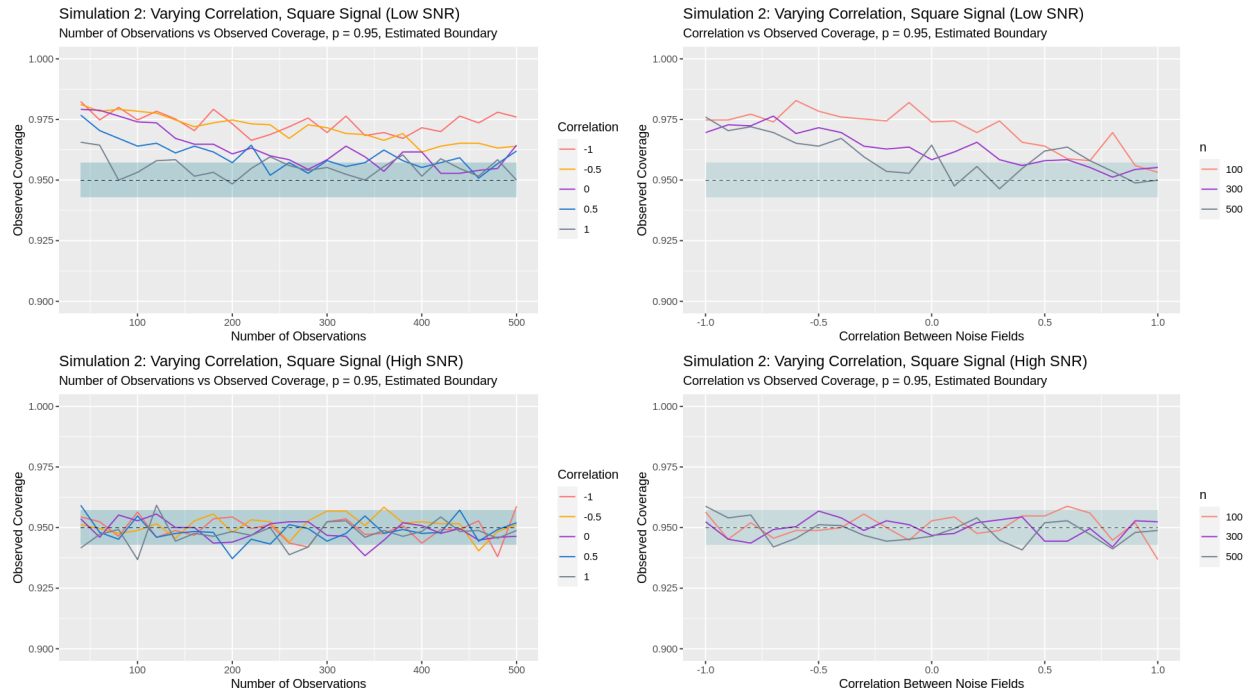


Figure S21: Empirical coverage results for Simulation 2. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for correlations of -1 , -0.5 , 0 , 0.5 and 1 . Right column: Correlation between noise fields vs observed coverage, shown for models including 100, 300 and 500 observations.

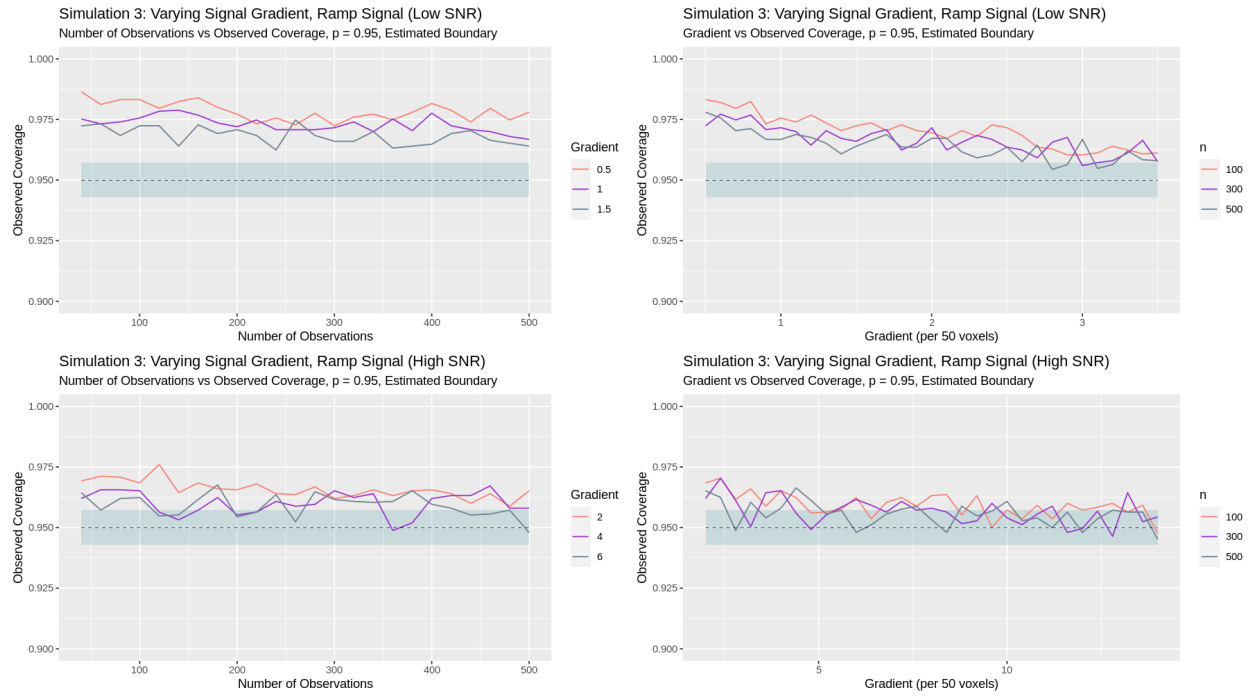


Figure S22: Empirical coverage results for Simulation 3. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for gradients of 0.5, 1 and 1.5 per 50 pixels (top), and 2, 4 and 6 per 50 pixels (bottom). Right column: Ramp gradient per 50 pixels vs observed coverage, shown for models including 100, 300 and 500 observations.

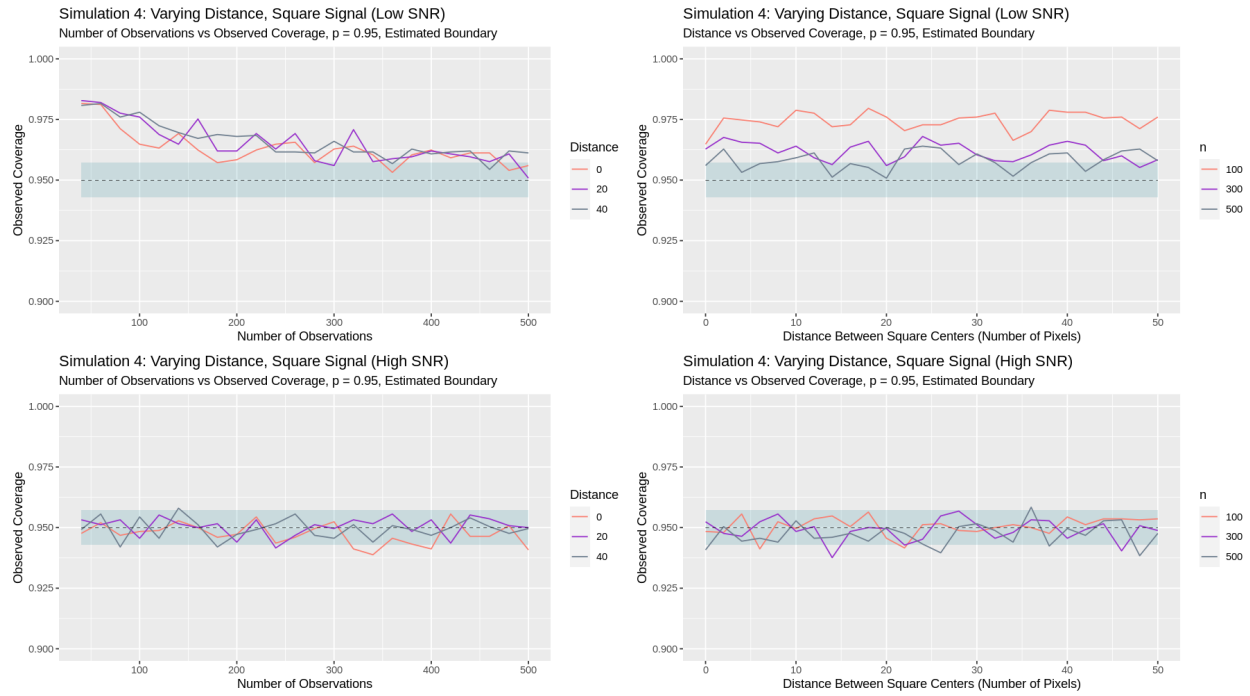


Figure S23: Empirical coverage results for Simulation 4. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for square separations of 0, 20 and 40 pixels. Right column: Separation between squares vs observed coverage, shown for models including 100, 300 and 500 observations.

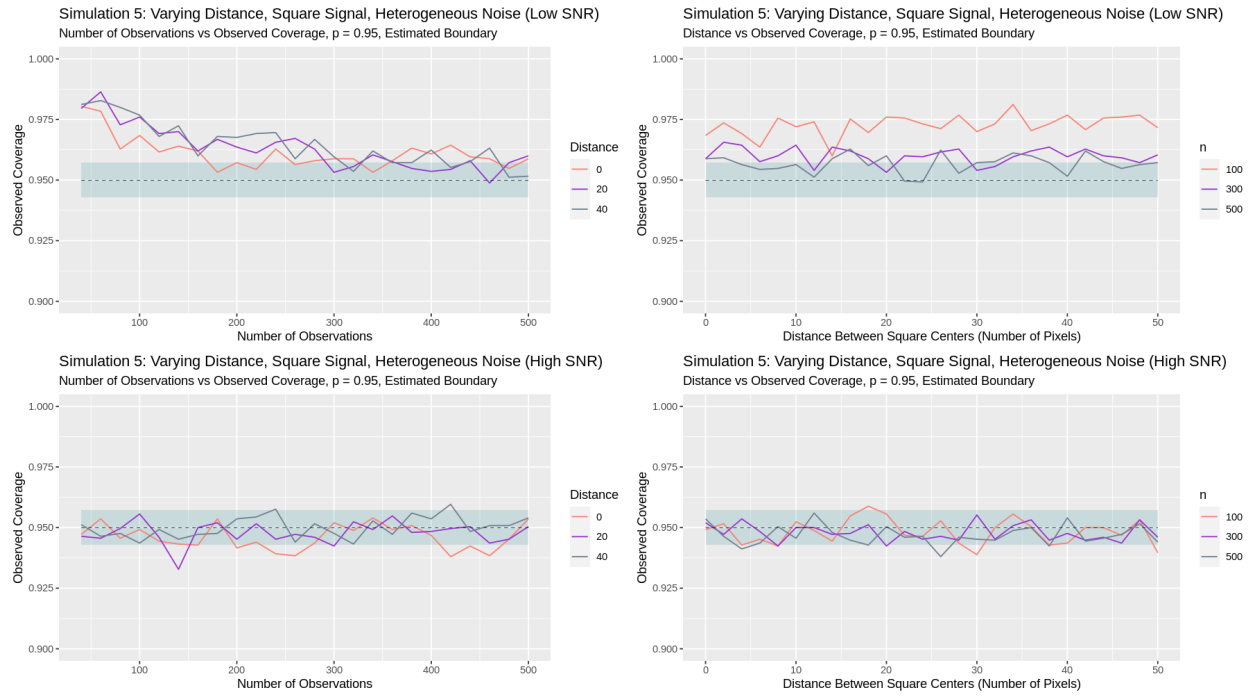


Figure S24: Empirical coverage results for Simulation 5. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for square separations of 0, 20 and 40 pixels. Right column: Separation between squares vs observed coverage, shown for models including 100, 300 and 500 observations.

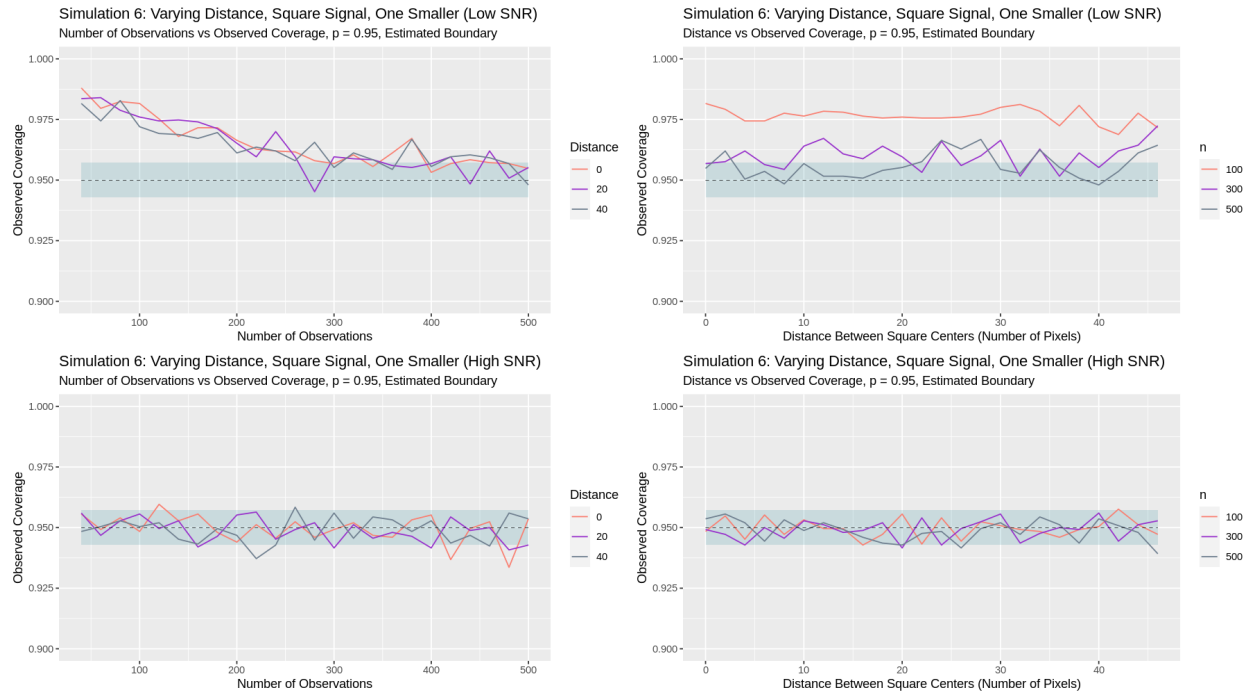


Figure S25: Empirical coverage results for Simulation 6. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for square separations of 0, 20 and 40 pixels. Right column: Separation between squares vs observed coverage, shown for models including 100, 300 and 500 observations.

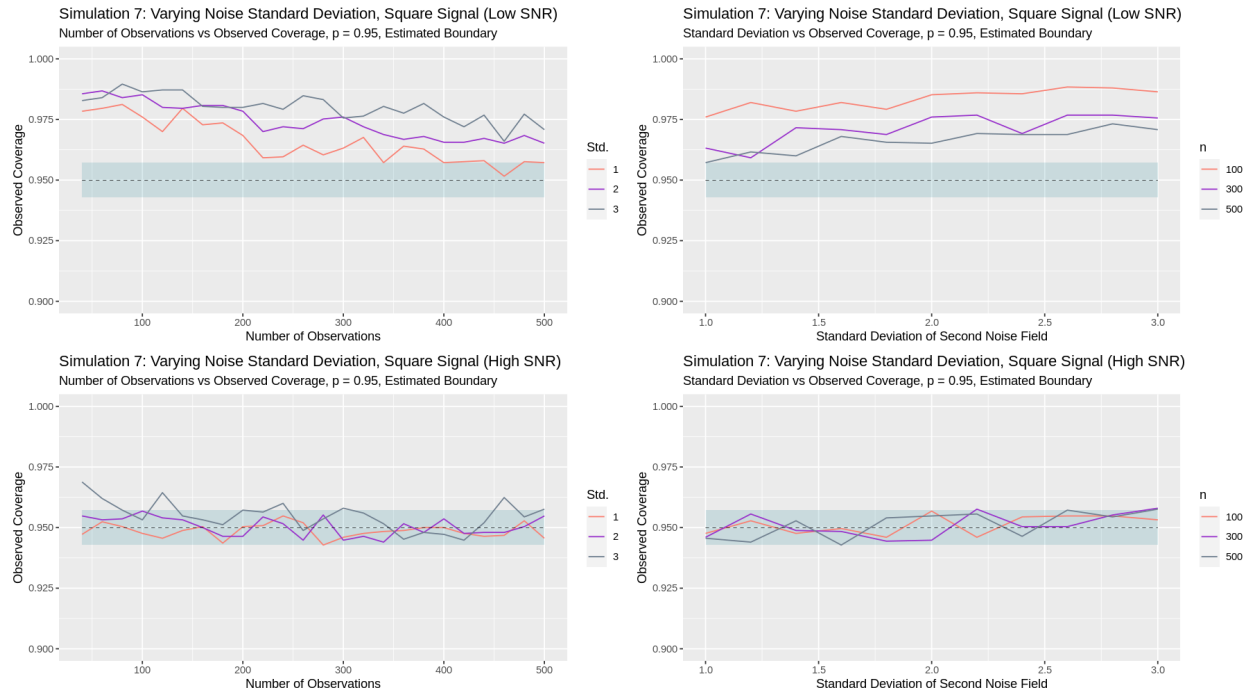


Figure S26: Empirical coverage results for Simulation 7. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for noise standard deviations of 1, 2 and 3. Right column: The standard deviation of the second noise field vs observed coverage, shown for models including 100, 300 and 500 observations.

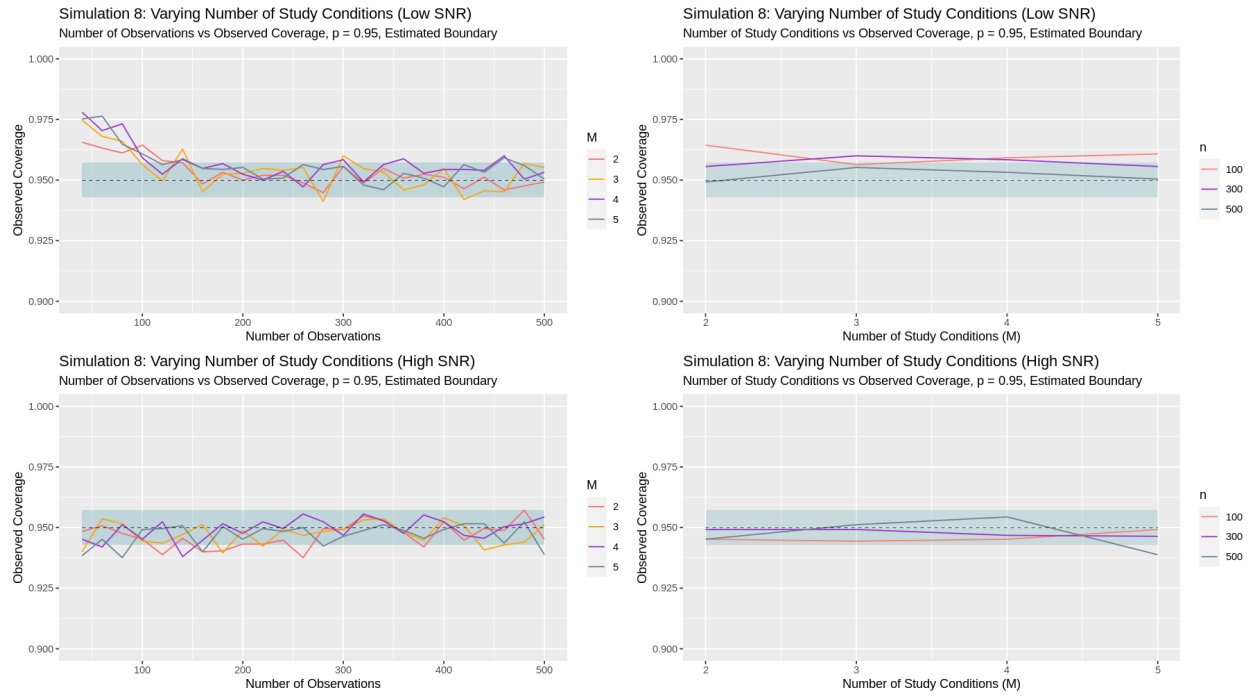


Figure S27: Empirical coverage results for Simulation 8. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for $M = 2, 3, 4$ and 5 . Right column: Number of study conditions (M) vs observed coverage, shown for models including 100, 300 and 500 observations.

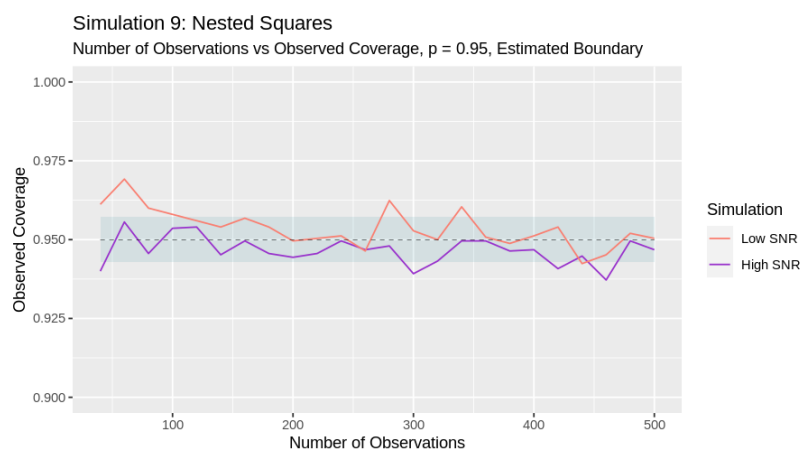


Figure S28: Empirical coverage results for Simulation 9. Results are shown for both low-SNR synthetic data and high-SNR synthetic data. Displayed is the number of observations vs observed coverage.

S6 Empirical Coverage Along the True Boundary

In this section, we provide the simulation results that were obtained by using the true boundary, $\partial\mathcal{F}_c$, to evaluate the suprema which appears in Equation (S1). In general, the results in the following sections agree with those of Section S5 and demonstrate the method’s robustness to the secondary factors previously listed.

However, unlike in the previous section, the method tended to provide liberal estimates for the confidence regions (e.g. the inclusion statement, $\hat{\mathcal{F}}_c^+ \subseteq \mathcal{F}_c \subseteq \hat{\mathcal{F}}_c^-$, held “too little”) when low-SNR data was used, and the number of observations was low. This effect was most prominent for Simulations 3 and 7, and is corroborated by the previous literature on confidence regions (c.f. Sommerfeld et al. [2018], Bowring et al. [2019]). As noted by previous authors, the apparent conflict between the method’s liberal behaviour in this section and the previous conservative results may be explained by a combination of competing factors.

When performing the method on a discrete lattice, it is expected that the suprema appearing in Equation (S1) will be underestimated (as samples cannot be drawn for the entire continuous region, $\partial\mathcal{F}_c$). This downwards bias is expected to cause an underestimation of the quantiles of H and, consequently, a . In turn, the underestimation of a results in tighter estimates for the confidence regions $\hat{\mathcal{F}}_c^+$ and $\hat{\mathcal{F}}_c^-$, thus causing the observed liberal behaviour of the method. Based on this argument, it may be natural to ask why no such liberal behaviour was observed in the results of Section S5.

We argue that the previous results were subject to the downward biases described above. However, these downward biases were countered by the upward bias caused by the uncertainty of the location of $\partial\hat{\mathcal{F}}_c$ discussed in the previous section. As the results of the previous section (and of Sommerfeld et al. [2018] and Bowring et al. [2019]) demonstrate, in practice, the upward bias tends to dominate. Consequently, we suggest that the method is more likely to produce conservative estimates than liberal estimates in practice.

This section is organized as follows. Section S6.1 provides the empirical coverage results for $\alpha = 0.2$, Section S6.2 provides equivalent results for $\alpha = 0.1$ and Section S6.3 provides results for $\alpha = 0.05$. In all plots, nominal coverage is shown as a dotted gray line and a corresponding binomial confidence interval is shaded in blue. All data points displayed are averages taken across 2500 simulation instances, each using 5000 bootstrap realizations.

S6.1 Nominal Coverage: 80%

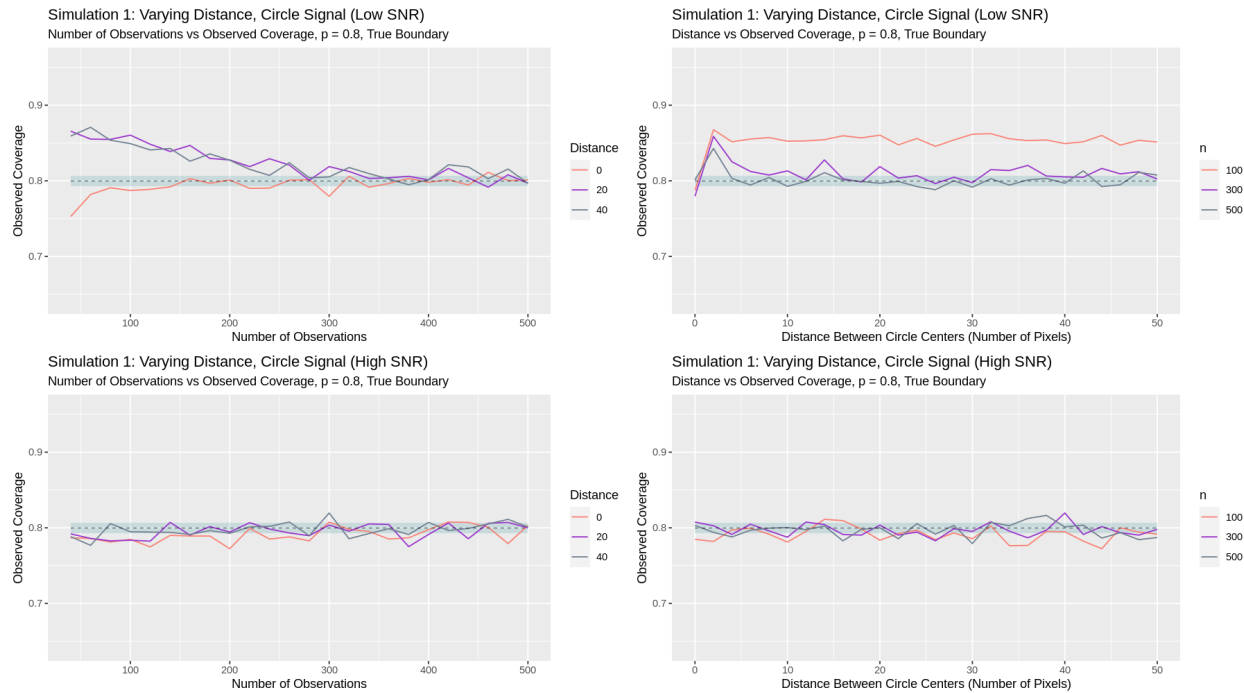


Figure S29: Empirical coverage results for Simulation 1. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for circle separations of 0, 20 and 40 pixels. Right column: Separation between circles vs observed coverage, shown for models including 100, 300 and 500 observations.

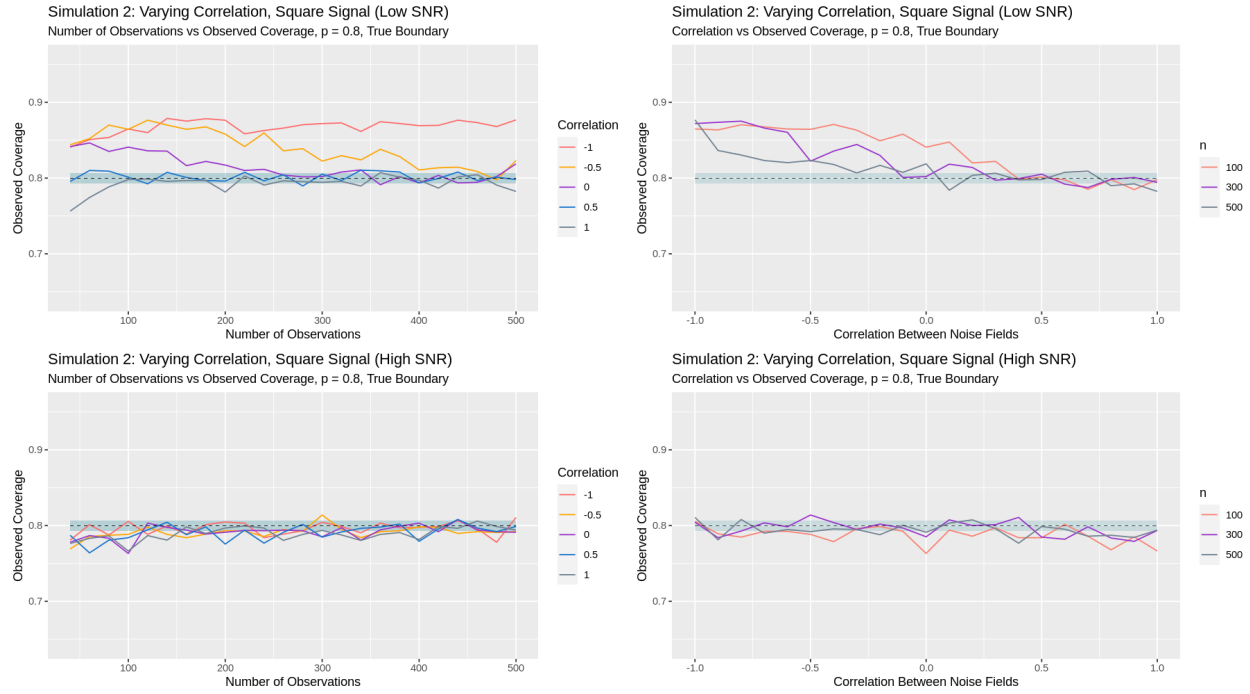


Figure S30: Empirical coverage results for Simulation 2. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for correlations of -1 , -0.5 , 0 , 0.5 and 1 . Right column: Correlation between noise fields vs observed coverage, shown for models including 100, 300 and 500 observations.

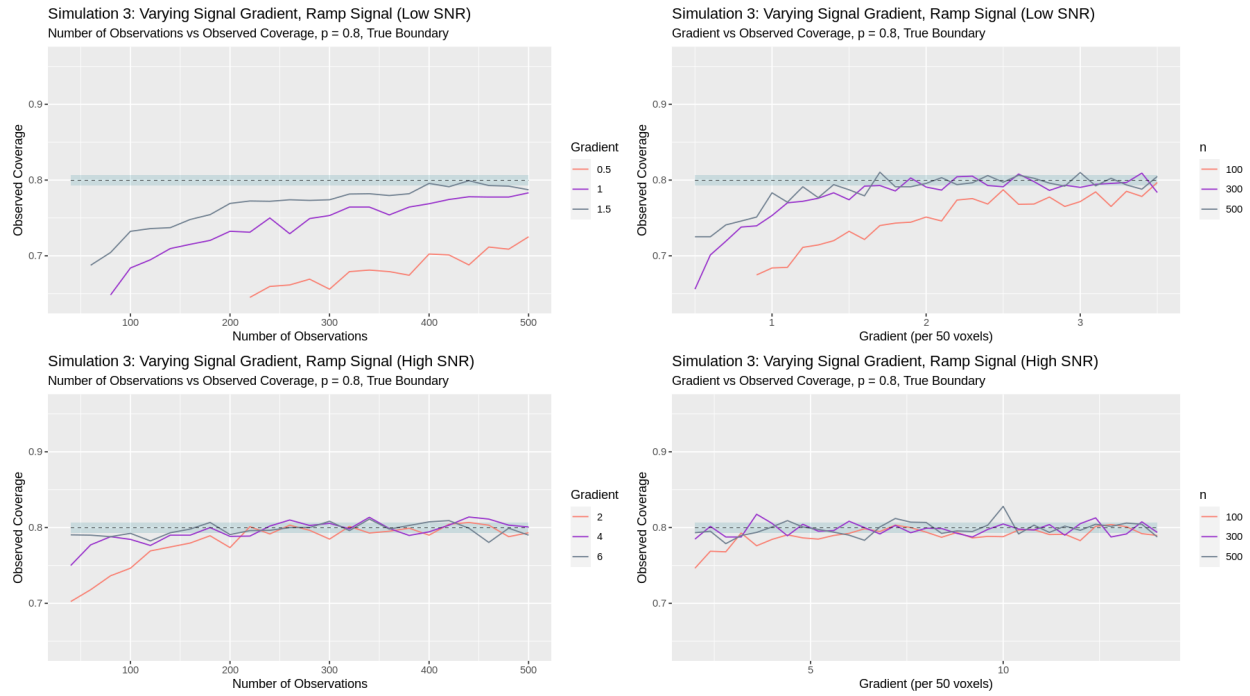


Figure S31: Empirical coverage results for Simulation 3. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for gradients of 0.5, 1 and 1.5 per 50 pixels (top), and 2, 4 and 6 per 50 pixels (bottom). Right column: Ramp gradient per 50 pixels vs observed coverage, shown for models including 100, 300 and 500 observations.

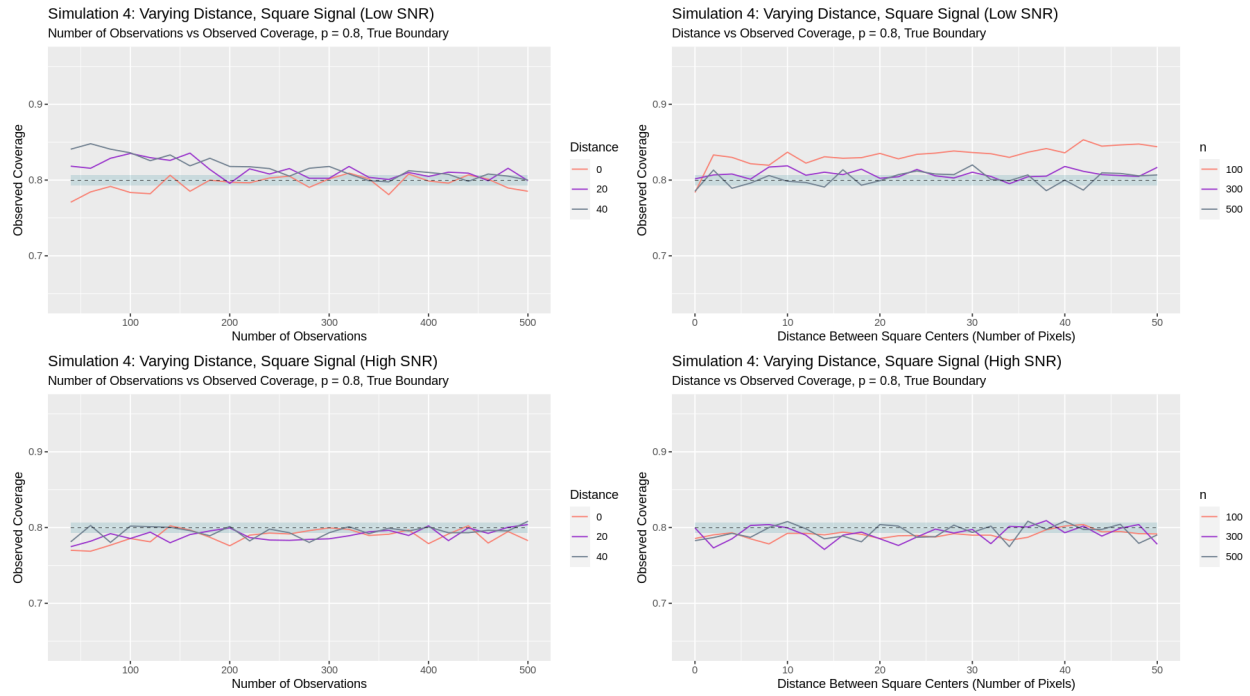


Figure S32: Empirical coverage results for Simulation 4. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for square separations of 0, 20 and 40 pixels. Right column: Separation between squares vs observed coverage, shown for models including 100, 300 and 500 observations.

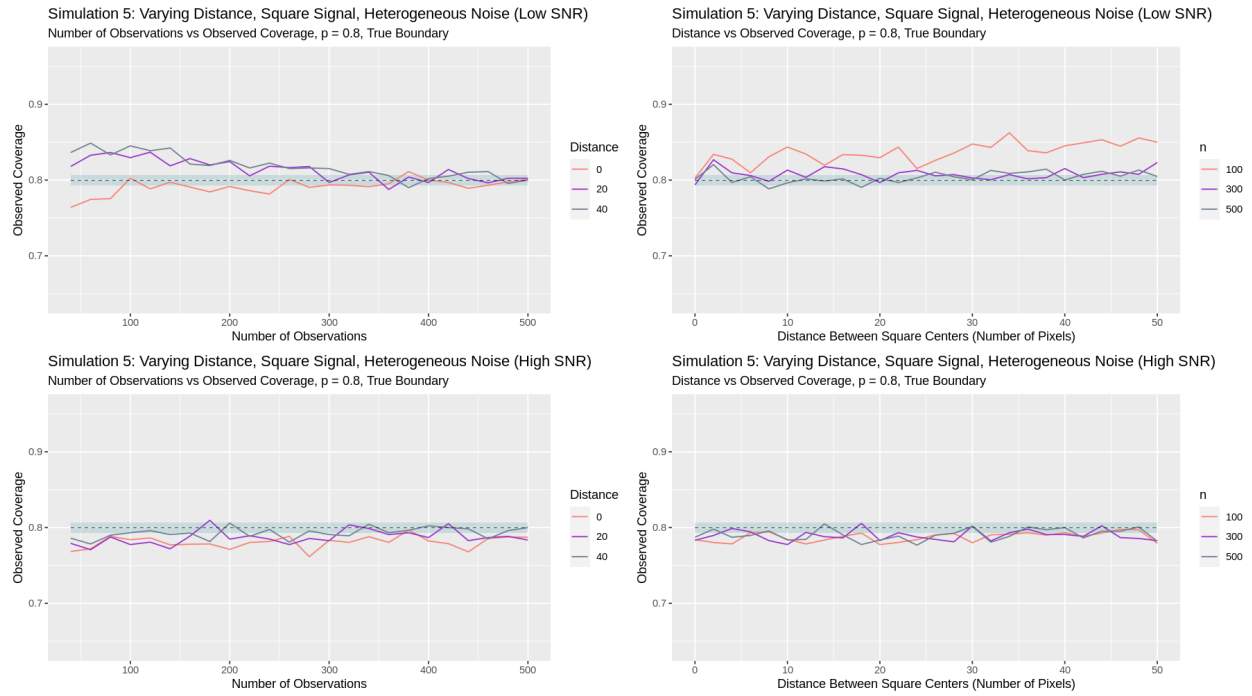


Figure S33: Empirical coverage results for Simulation 5. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for square separations of 0, 20 and 40 pixels. Right column: Separation between squares vs observed coverage, shown for models including 100, 300 and 500 observations.

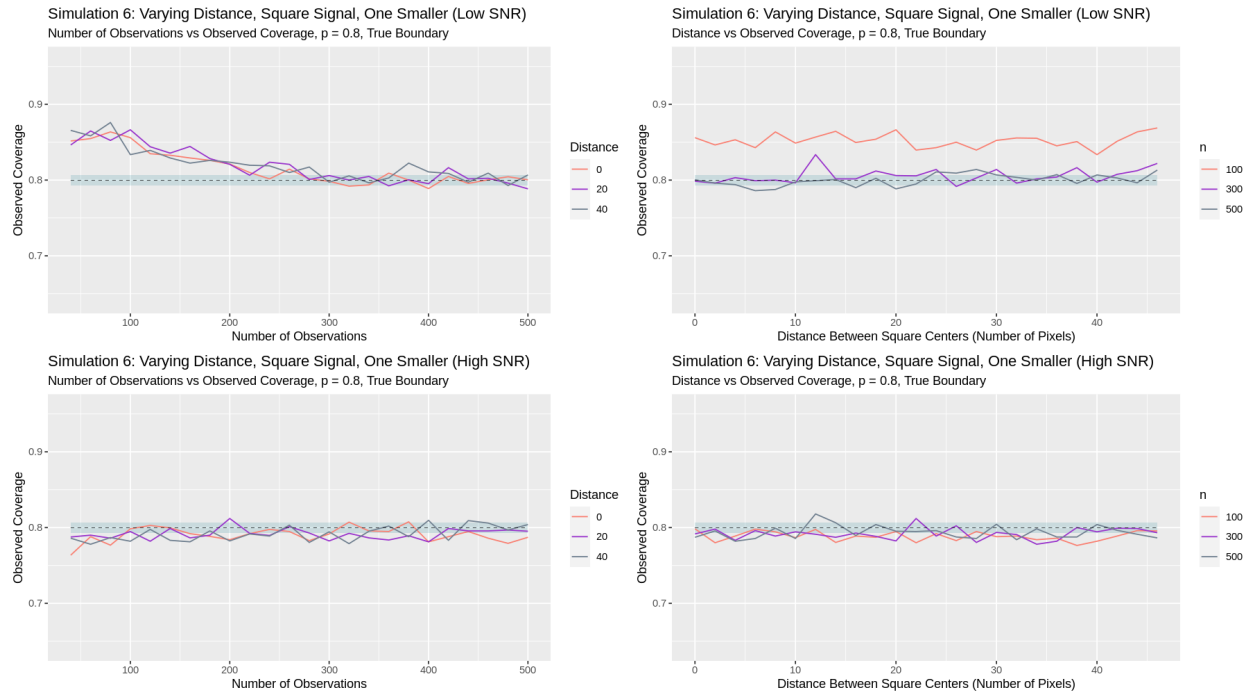


Figure S34: Empirical coverage results for Simulation 6. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for square separations of 0, 20 and 40 pixels. Right column: Separation between squares vs observed coverage, shown for models including 100, 300 and 500 observations.

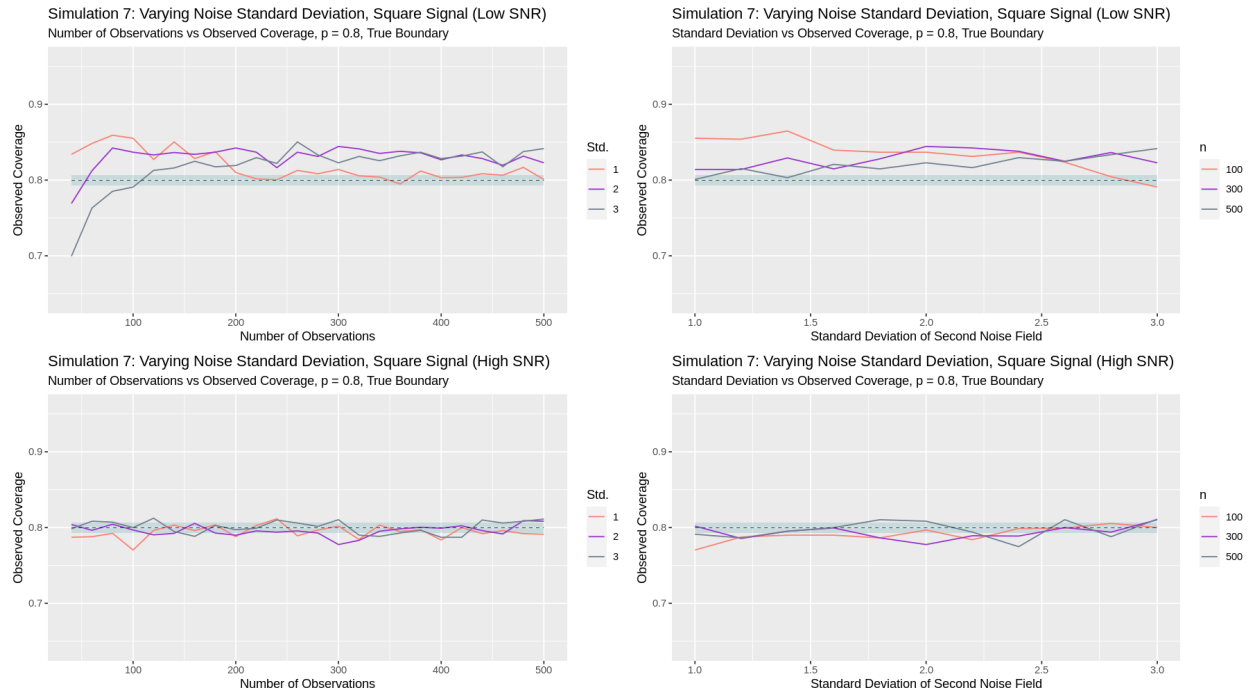


Figure S35: Empirical coverage results for Simulation 7. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for noise standard deviations of 1, 2 and 3. Right column: The standard deviation of the second noise field vs observed coverage, shown for models including 100, 300 and 500 observations.

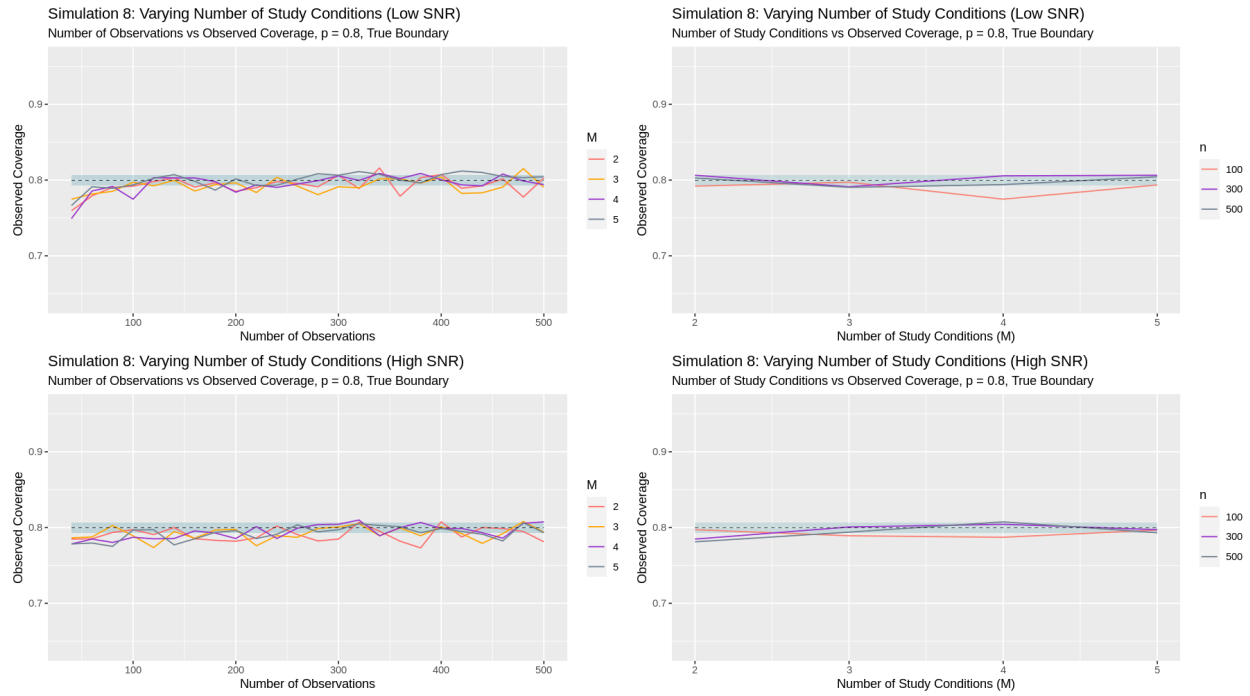


Figure S36: Empirical coverage results for Simulation 8. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for $M = 2, 3, 4$ and 5 . Right column: Number of study conditions (M) vs observed coverage, shown for models including 100, 300 and 500 observations.

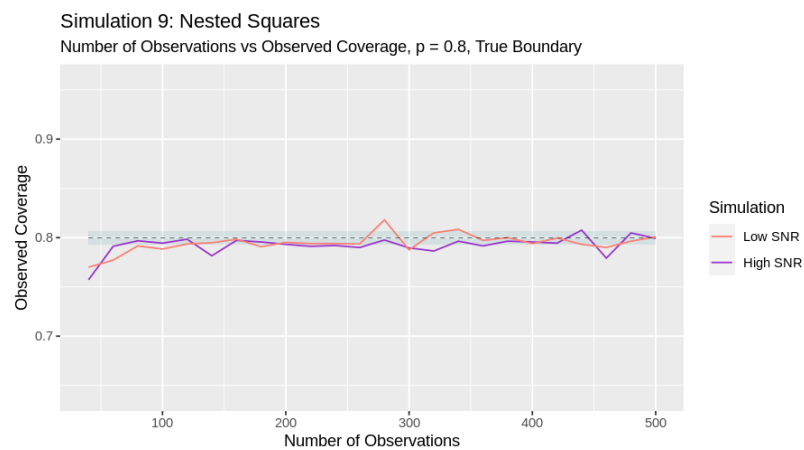


Figure S37: Empirical coverage results for Simulation 9. Results are shown for both low-SNR synthetic data and high-SNR synthetic data. Displayed is the number of observations vs observed coverage.

S6.2 Nominal Coverage: 90%

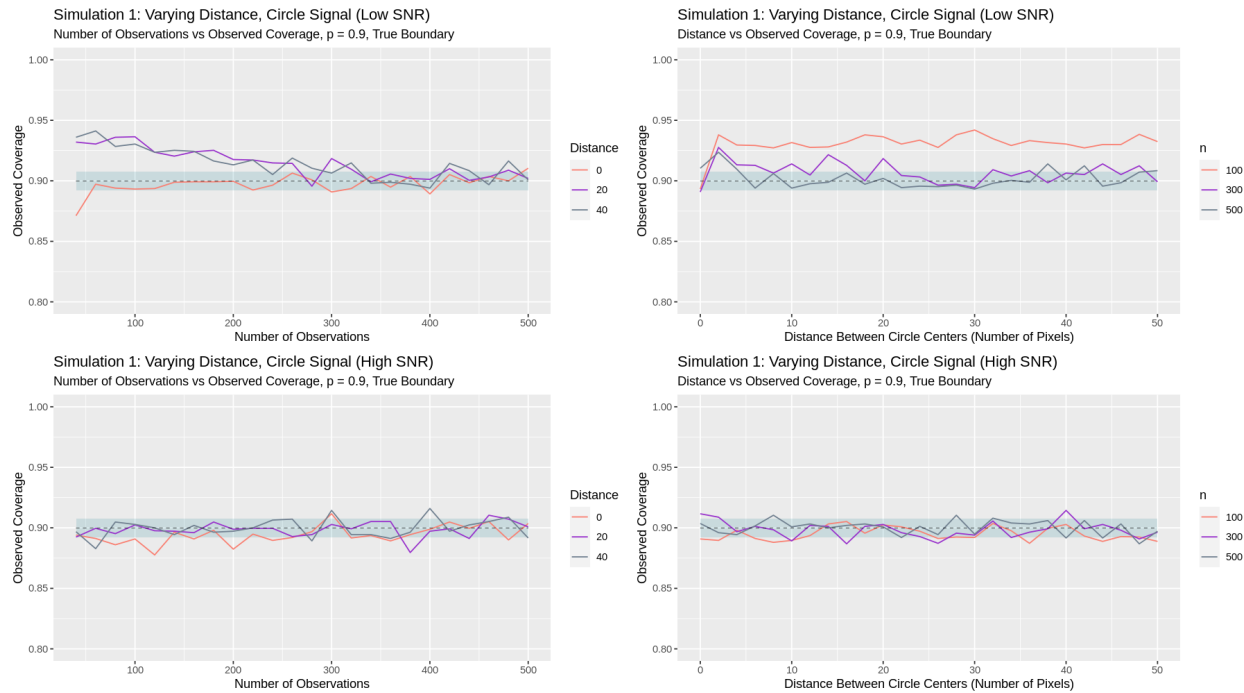


Figure S38: Empirical coverage results for Simulation 1. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for circle separations of 0, 20 and 40 pixels. Right column: Separation between circles vs observed coverage, shown for models including 100, 300 and 500 observations.

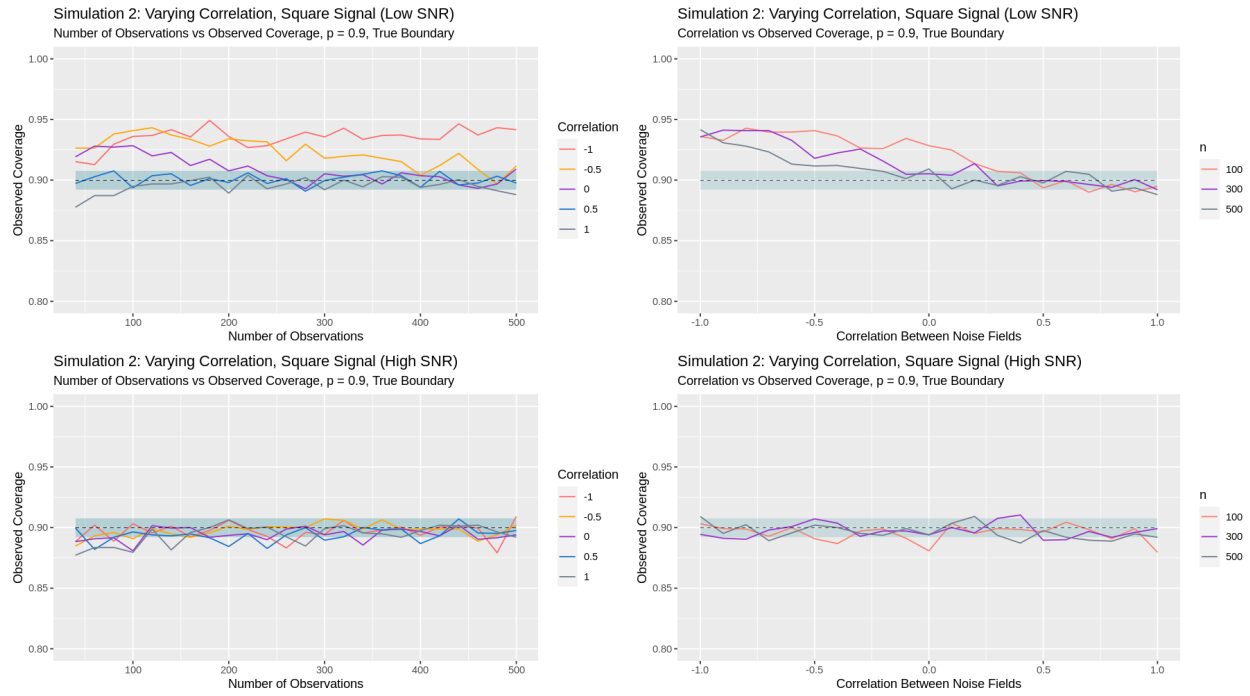


Figure S39: Empirical coverage results for Simulation 2. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for correlations of -1 , -0.5 , 0 , 0.5 and 1 . Right column: Correlation between noise fields vs observed coverage, shown for models including 100, 300 and 500 observations.

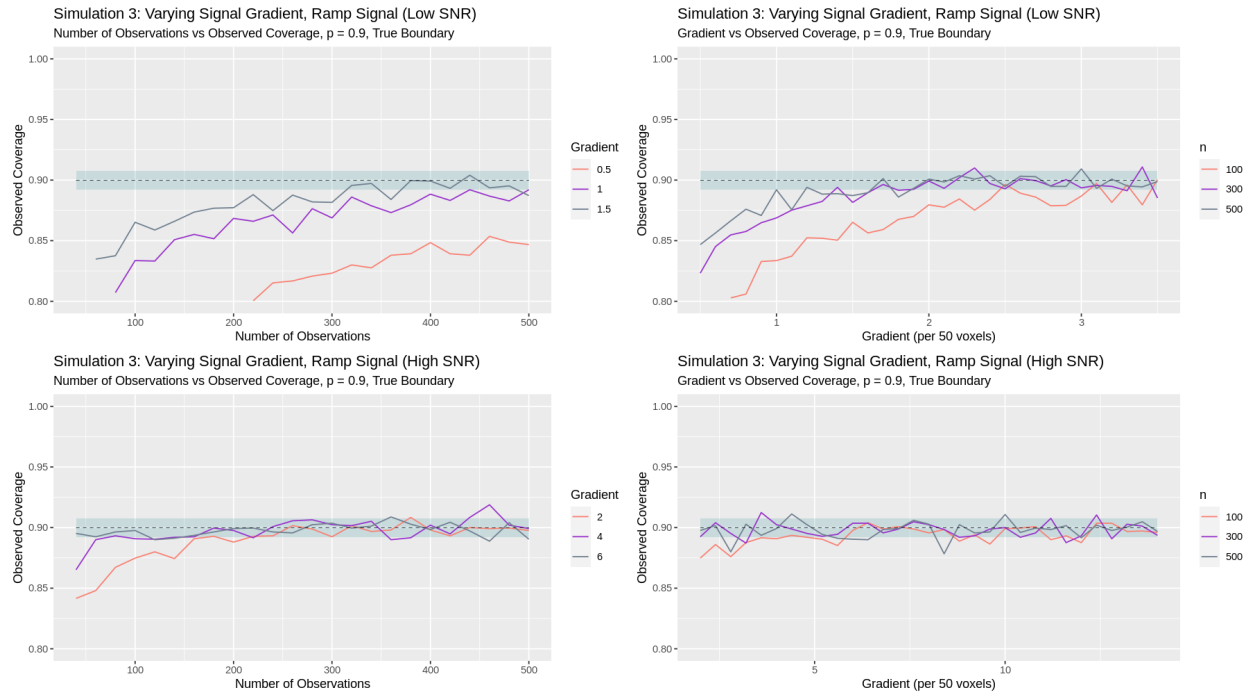


Figure S40: Empirical coverage results for Simulation 3. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for gradients of 0.5, 1 and 1.5 per 50 pixels (top), and 2, 4 and 6 per 50 pixels (bottom). Right column: Ramp gradient per 50 pixels vs observed coverage, shown for models including 100, 300 and 500 observations.

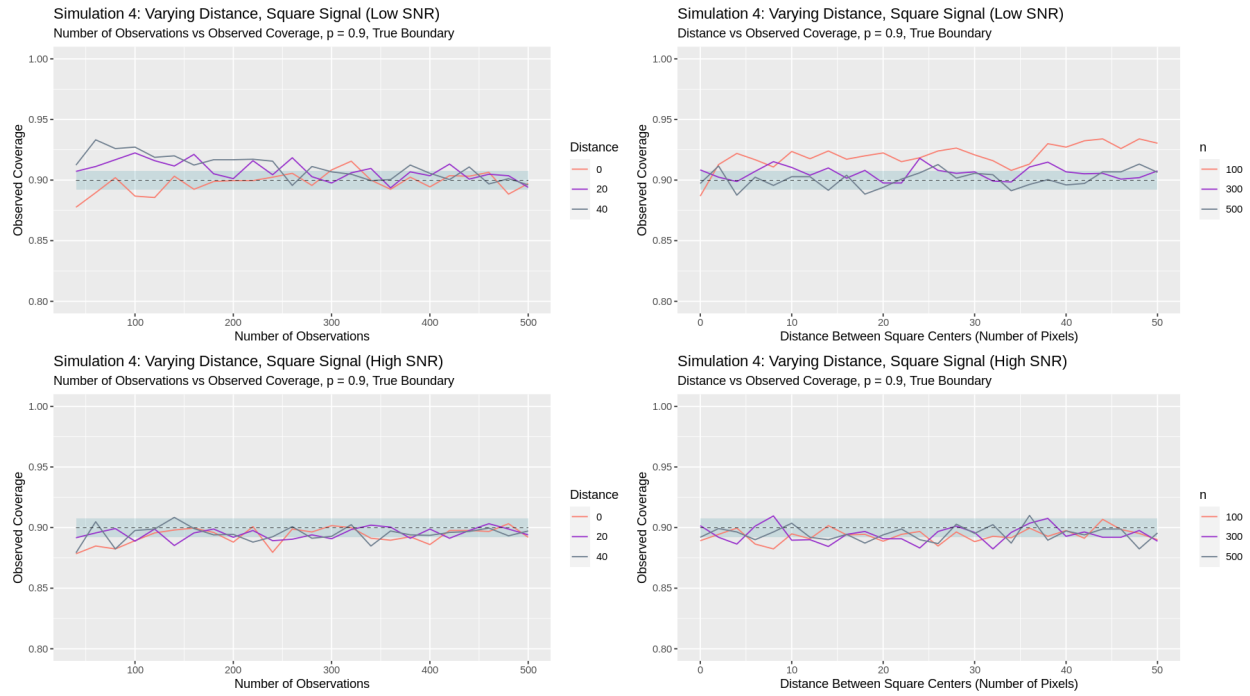


Figure S41: Empirical coverage results for Simulation 4. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for square separations of 0, 20 and 40 pixels. Right column: Separation between squares vs observed coverage, shown for models including 100, 300 and 500 observations.

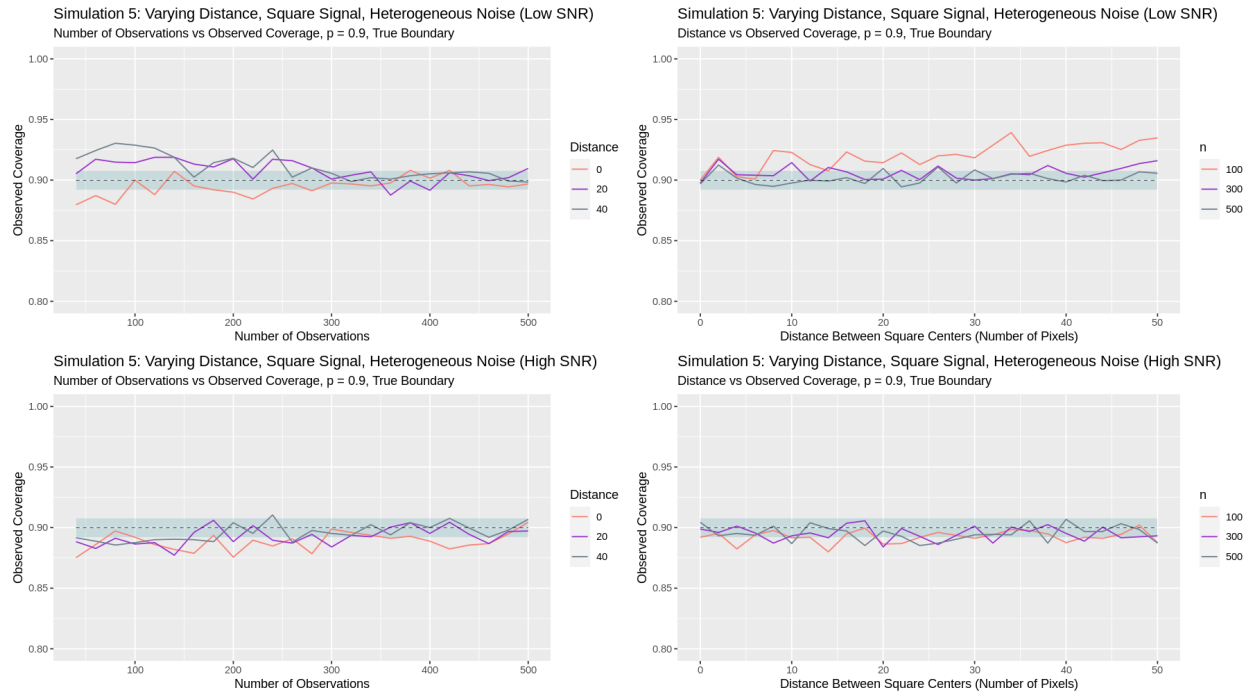


Figure S42: Empirical coverage results for Simulation 5. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for square separations of 0, 20 and 40 pixels. Right column: Separation between squares vs observed coverage, shown for models including 100, 300 and 500 observations.

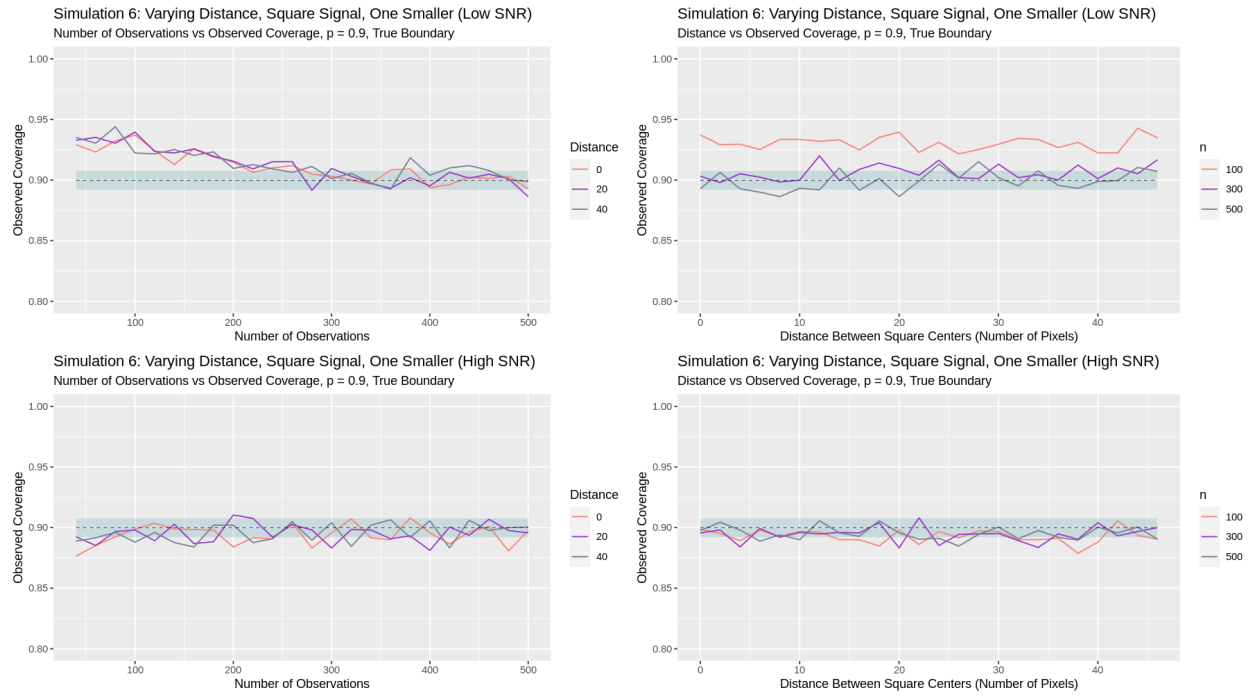


Figure S43: Empirical coverage results for Simulation 6. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for square separations of 0, 20 and 40 pixels. Right column: Separation between squares vs observed coverage, shown for models including 100, 300 and 500 observations.

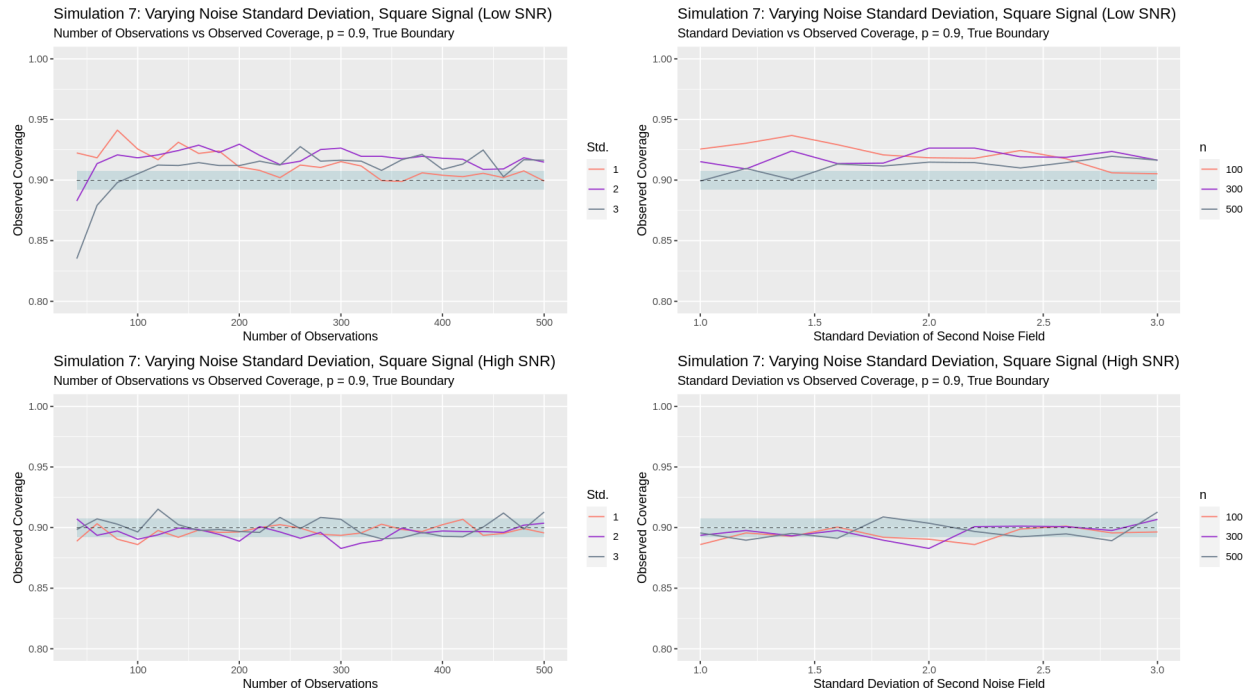


Figure S44: Empirical coverage results for Simulation 7. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for noise standard deviations of 1, 2 and 3. Right column: The standard deviation of the second noise field vs observed coverage, shown for models including 100, 300 and 500 observations.

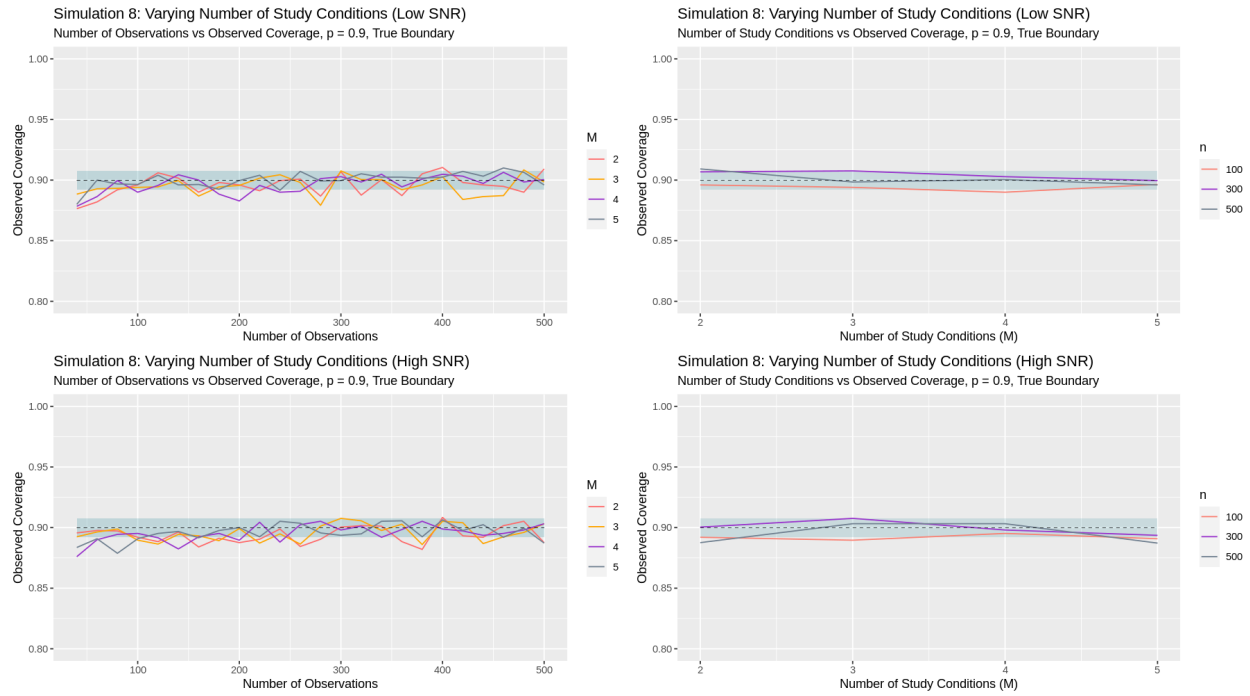


Figure S45: Empirical coverage results for Simulation 8. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for $M = 2, 3, 4$ and 5 . Right column: Number of study conditions (M) vs observed coverage, shown for models including 100, 300 and 500 observations.

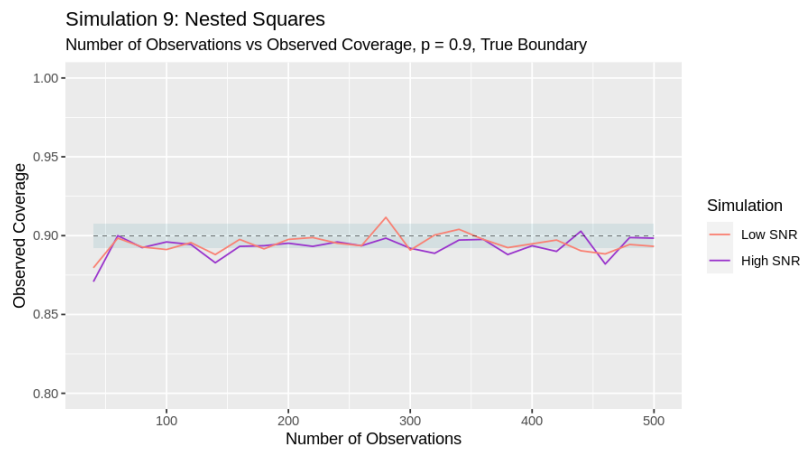


Figure S46: Empirical coverage results for Simulation 9. Results are shown for both low-SNR synthetic data and high-SNR synthetic data. Displayed is the number of observations vs observed coverage.

S6.3 Nominal Coverage: 95%

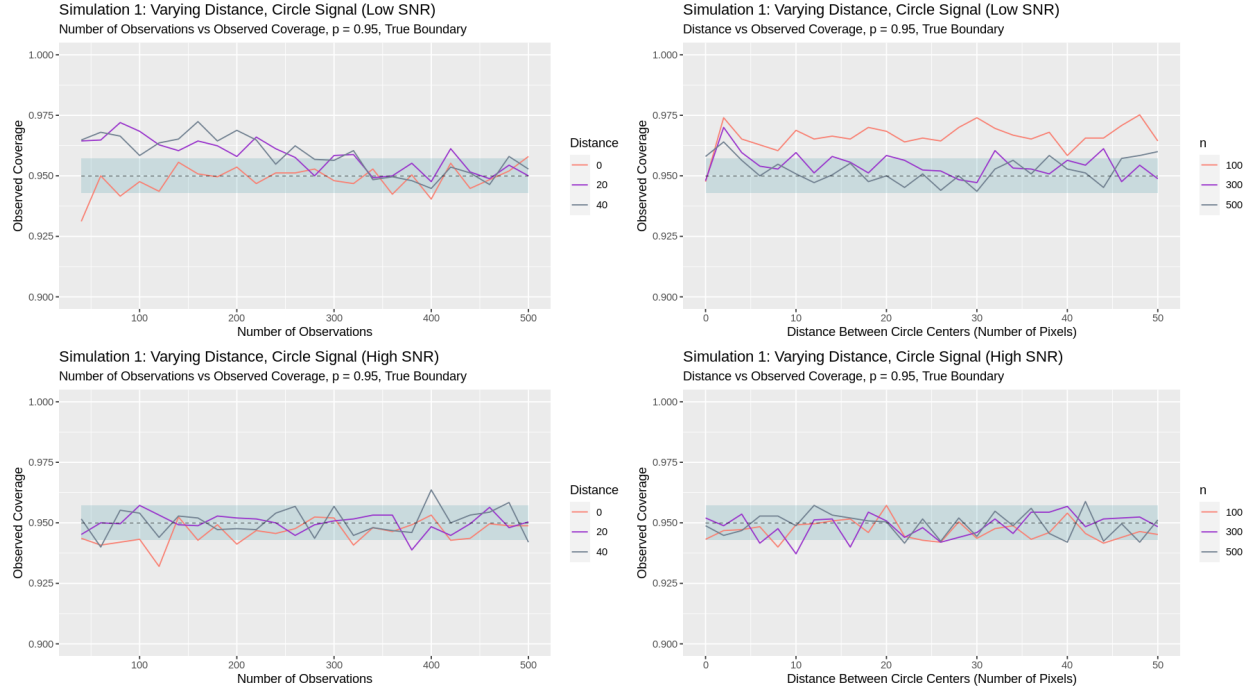


Figure S47: Empirical coverage results for Simulation 1. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for circle separations of 0, 20 and 40 pixels. Right column: Separation between circles vs observed coverage, shown for models including 100, 300 and 500 observations.

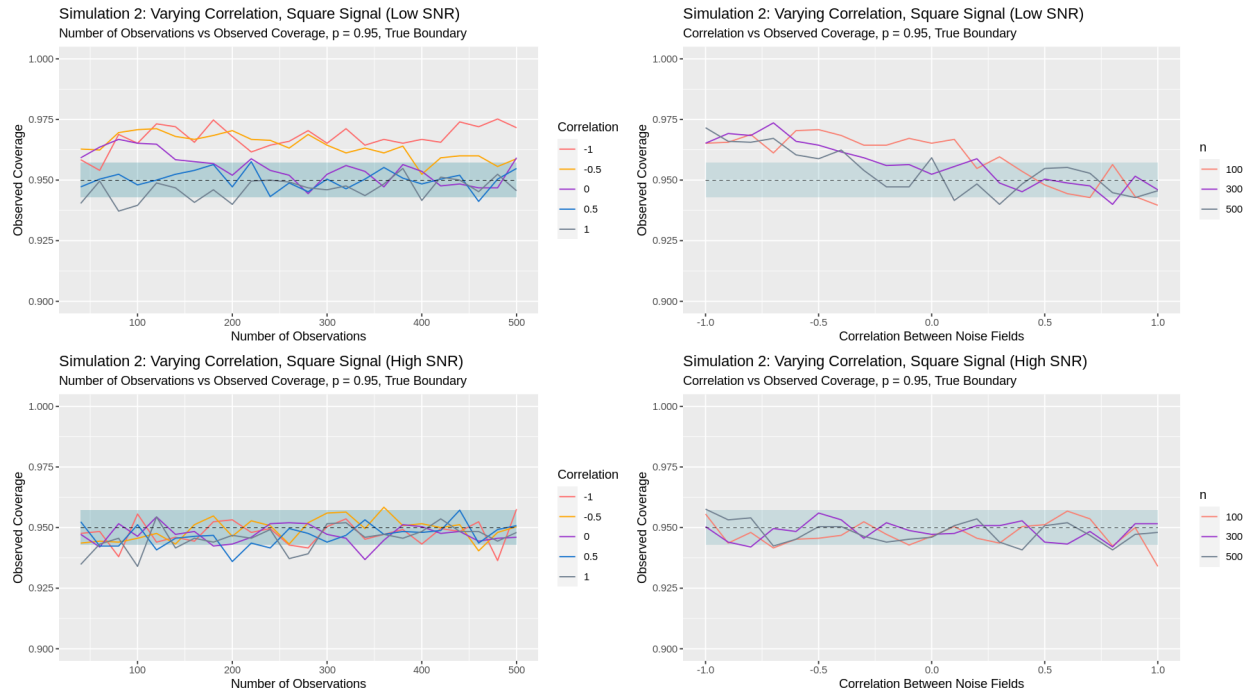


Figure S48: Empirical coverage results for Simulation 2. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for correlations of -1 , -0.5 , 0 , 0.5 and 1 . Right column: Correlation between noise fields vs observed coverage, shown for models including 100, 300 and 500 observations.

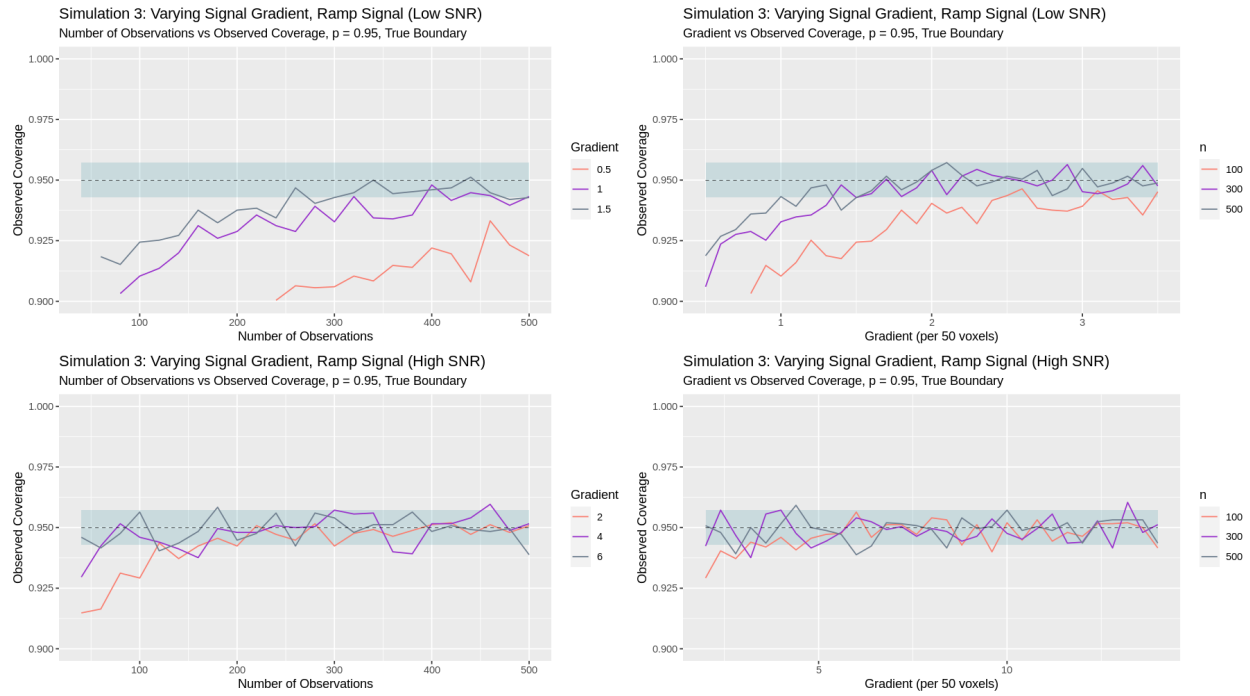


Figure S49: Empirical coverage results for Simulation 3. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for gradients of 0.5, 1 and 1.5 per 50 pixels (top), and 2, 4 and 6 per 50 pixels (bottom). Right column: Ramp gradient per 50 pixels vs observed coverage, shown for models including 100, 300 and 500 observations.

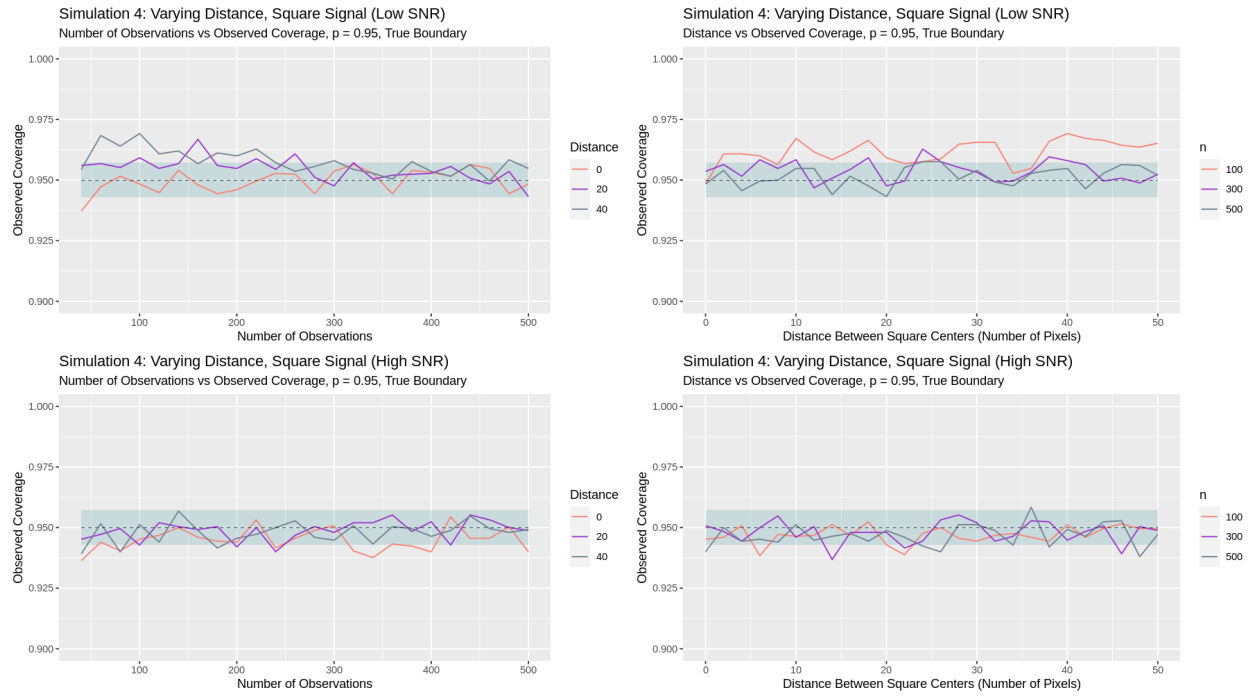


Figure S50: Empirical coverage results for Simulation 4. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for square separations of 0, 20 and 40 pixels. Right column: Separation between squares vs observed coverage, shown for models including 100, 300 and 500 observations.

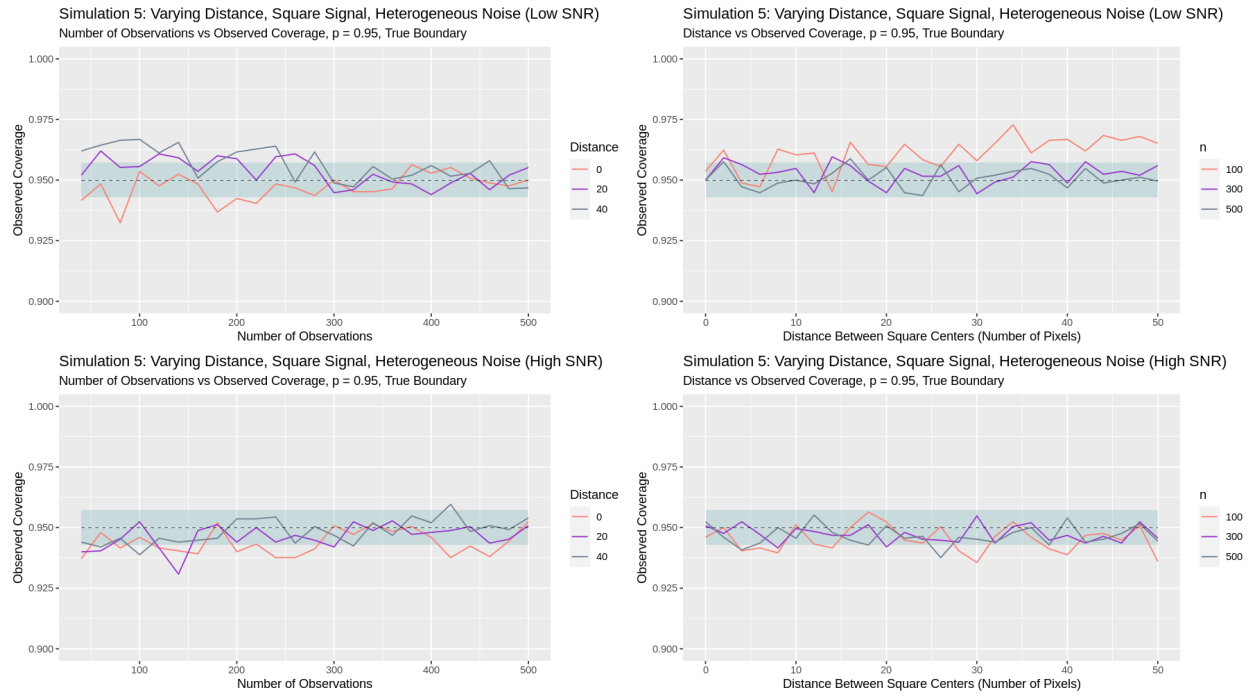


Figure S51: Empirical coverage results for Simulation 5. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for square separations of 0, 20 and 40 pixels. Right column: Separation between squares vs observed coverage, shown for models including 100, 300 and 500 observations.

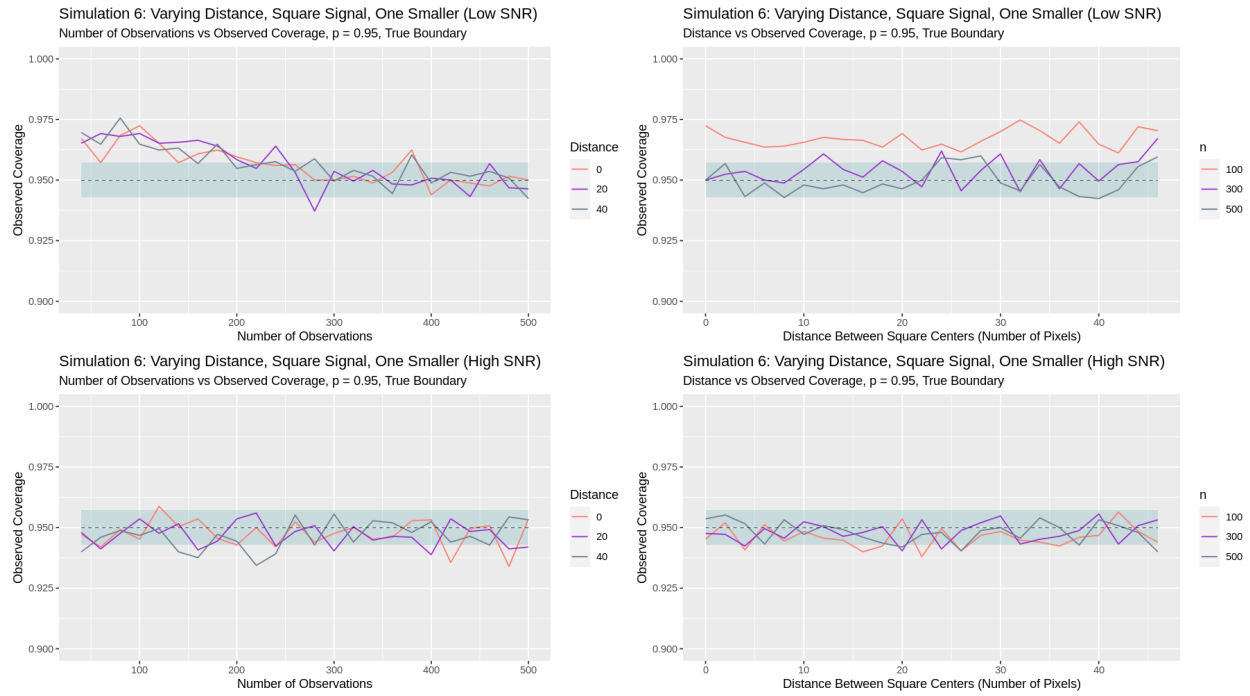


Figure S52: Empirical coverage results for Simulation 6. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for square separations of 0, 20 and 40 pixels. Right column: Separation between squares vs observed coverage, shown for models including 100, 300 and 500 observations.

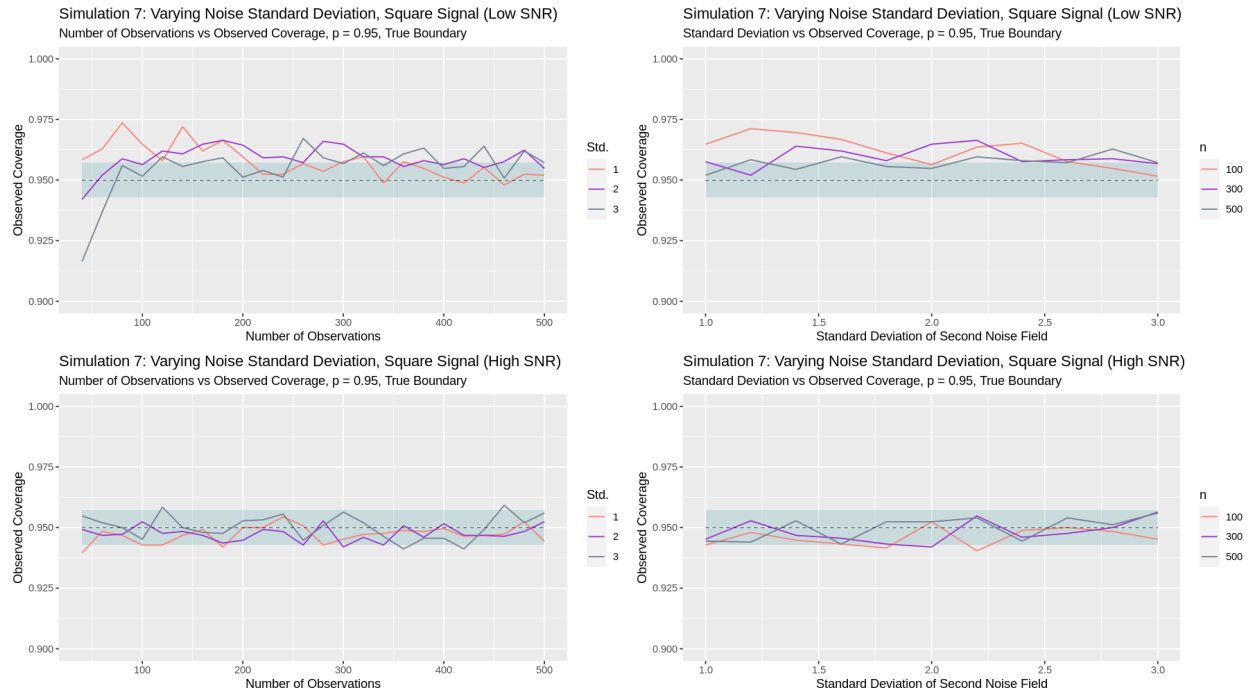


Figure S53: Empirical coverage results for Simulation 7. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for noise standard deviations of 1, 2 and 3. Right column: The standard deviation of the second noise field vs observed coverage, shown for models including 100, 300 and 500 observations.

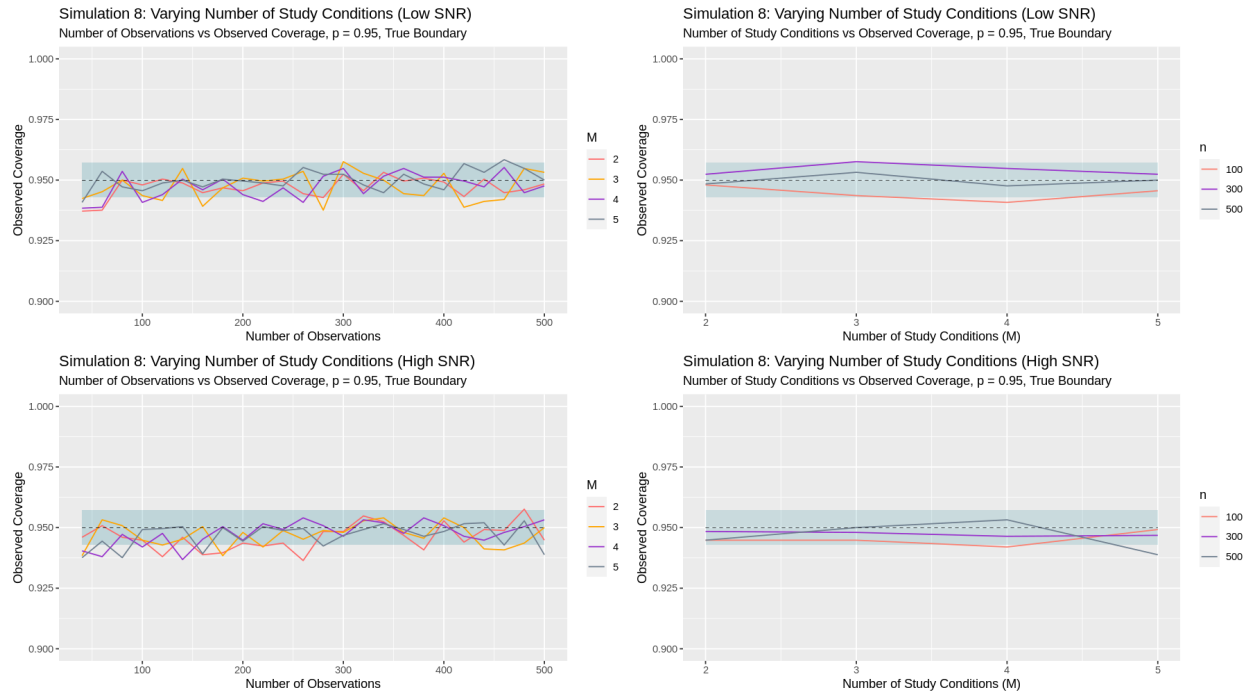


Figure S54: Empirical coverage results for Simulation 8. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs observed coverage, shown for $M = 2, 3, 4$ and 5 . Right column: Number of study conditions (M) vs observed coverage, shown for models including 100, 300 and 500 observations.

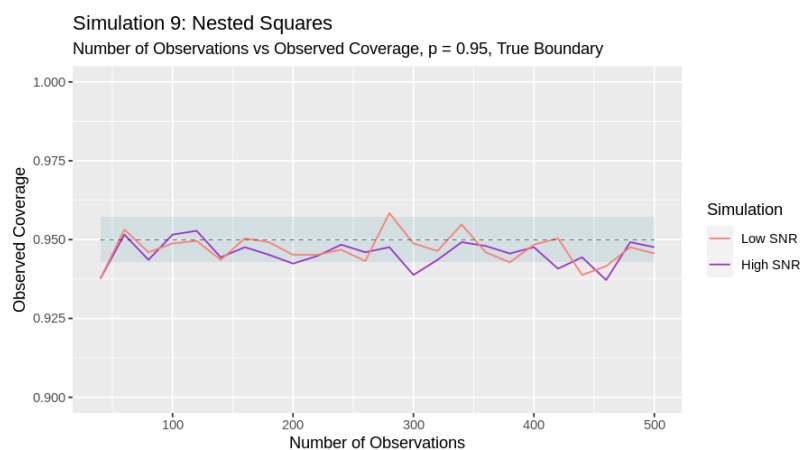


Figure S55: Empirical coverage results for Simulation 9. Results are shown for both low-SNR synthetic data and high-SNR synthetic data. Displayed is the number of observations vs observed coverage.

S7 Computation Time

In this section, we present the observed computation times for the 9 simulations described in Section S4 of this document and Section 5.4.1 of the thesis document. All results reported in this section were obtained using an Intel(R) Xeon(R) Gold 6126 2.60GHz processor with 16GB RAM and are given as averages taken over 5000 simulation instances. 2500 of the simulation instances used the estimated boundary, $\partial\hat{\mathcal{F}}_c$, to evaluate the suprema in Equation (S1), whilst the other 2500 used the true boundary, $\partial\mathcal{F}_c$. Each simulation instance was performed using 5000 bootstrap realisations. We stress here that the implementation of confidence regions provided alongside this work is far from optimal. The purpose of the inclusion of the following results is to highlight trends in computation time rather than to showcase our implementation of the method.

Across all simulations, it can be seen that computation time appeared to increase linearly with the number of observations. In addition, linear trends in computation time were observed as the overlap between excursion sets increased (c.f. Figs. S56 and S59-S61), the signal gradient was reduced (Fig. S58) and the noise standard deviation was increased (Fig. S62). Notably, varying the correlation between the noise fields appeared to have little impact on the observed computation time (c.f. Fig. S57). However, an exponential increase in computation time was observed when the number of study conditions, M , was increased (c.f. Fig. S63). This exponential increase in computation time is to be expected as the size of the power set, $\mathcal{P}^+(\mathcal{M})$, (and thus the number of boundary segments which must be considered during the bootstrap) increases exponentially with M .

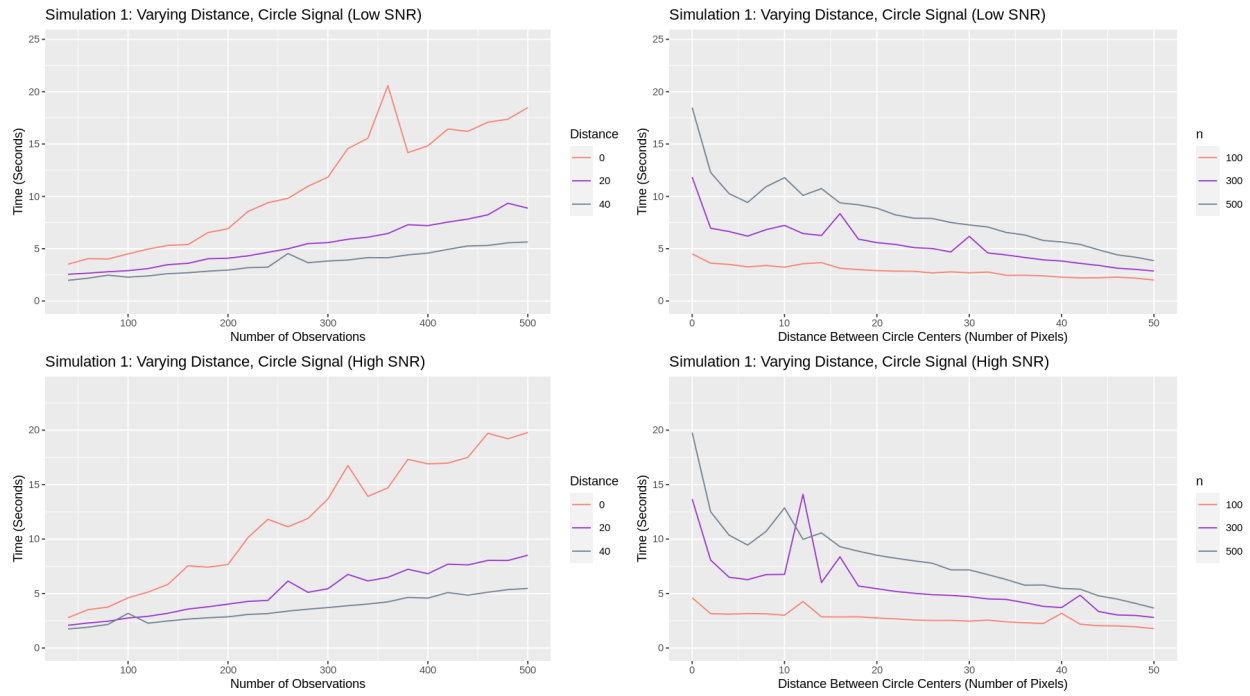


Figure S56: Bootstrap computation times for Simulation 1. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs computation time, shown for circle separations of 0, 20 and 40 pixels. Right column: Separation between circles vs computation time, shown for models including 100, 300 and 500 observations.

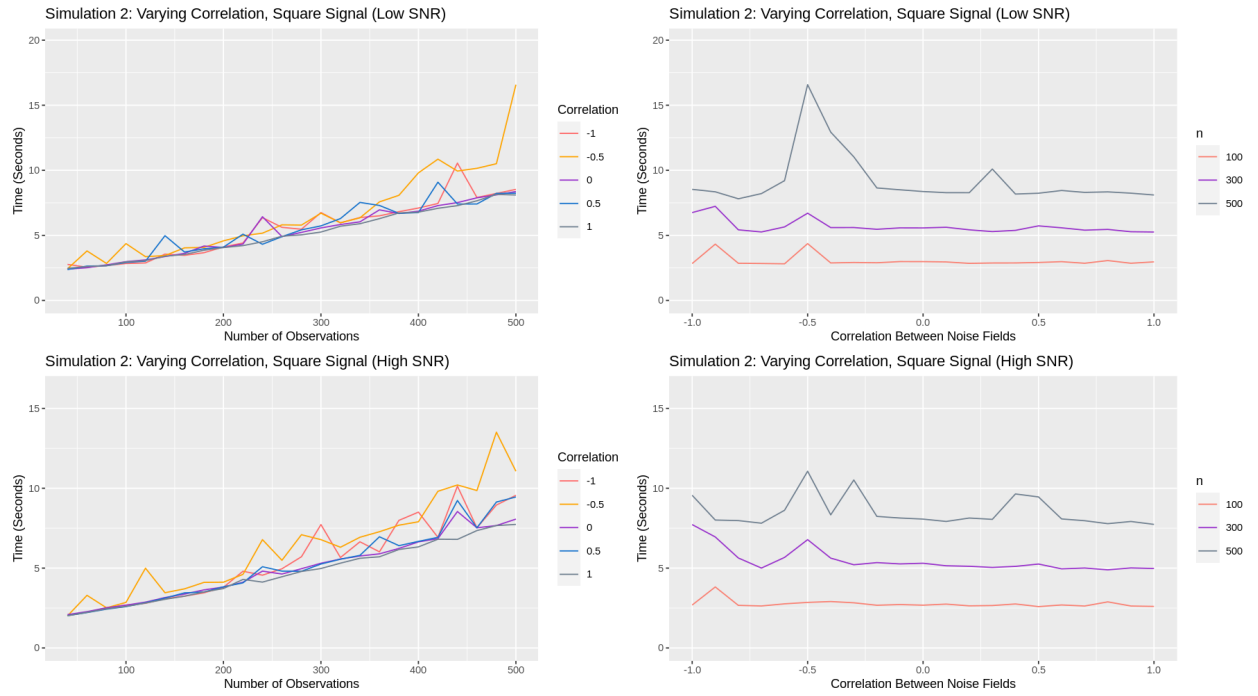


Figure S57: Bootstrap computation times for Simulation 2. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs computation time, shown for correlations of -1 , -0.5 , 0 , 0.5 and 1 . Right column: Correlation between noise fields vs computation time, shown for models including 100, 300 and 500 observations.

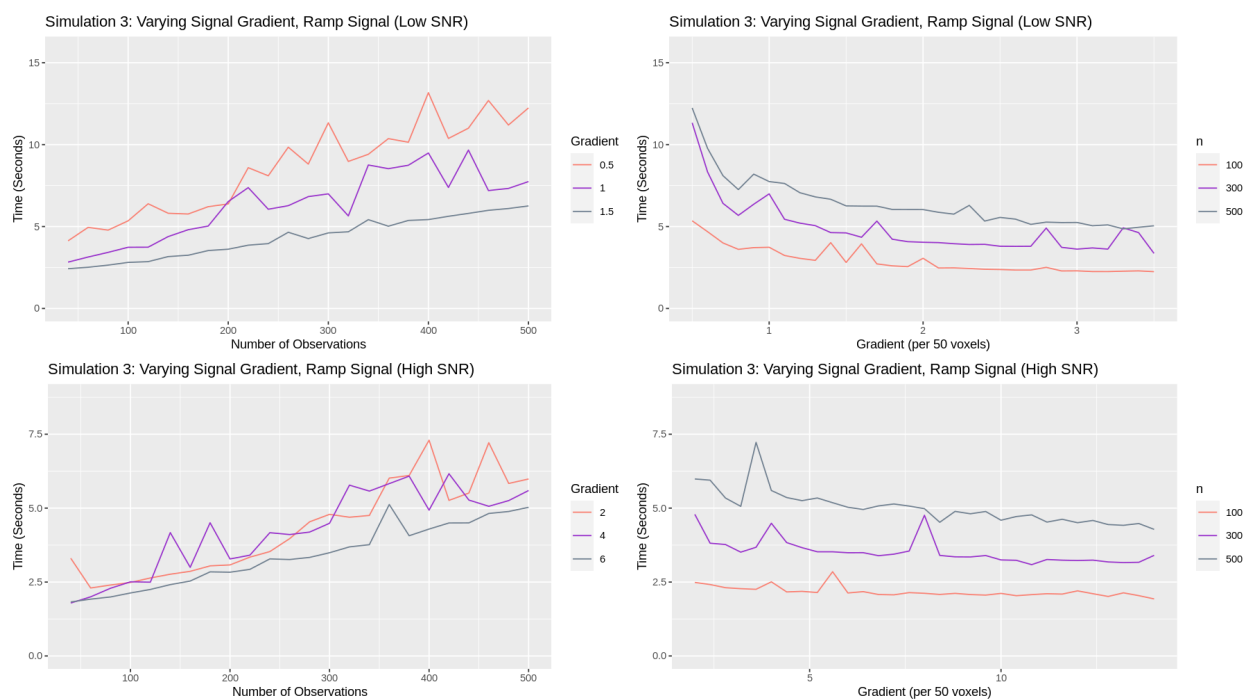


Figure S58: Bootstrap computation times for Simulation 3. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs computation time, shown for gradients of 0.5, 1 and 1.5 per 50 pixels (top), and 2, 4 and 6 per 50 pixels (bottom). Right column: Ramp gradient per 50 pixels vs computation time, shown for models including 100, 300 and 500 observations.

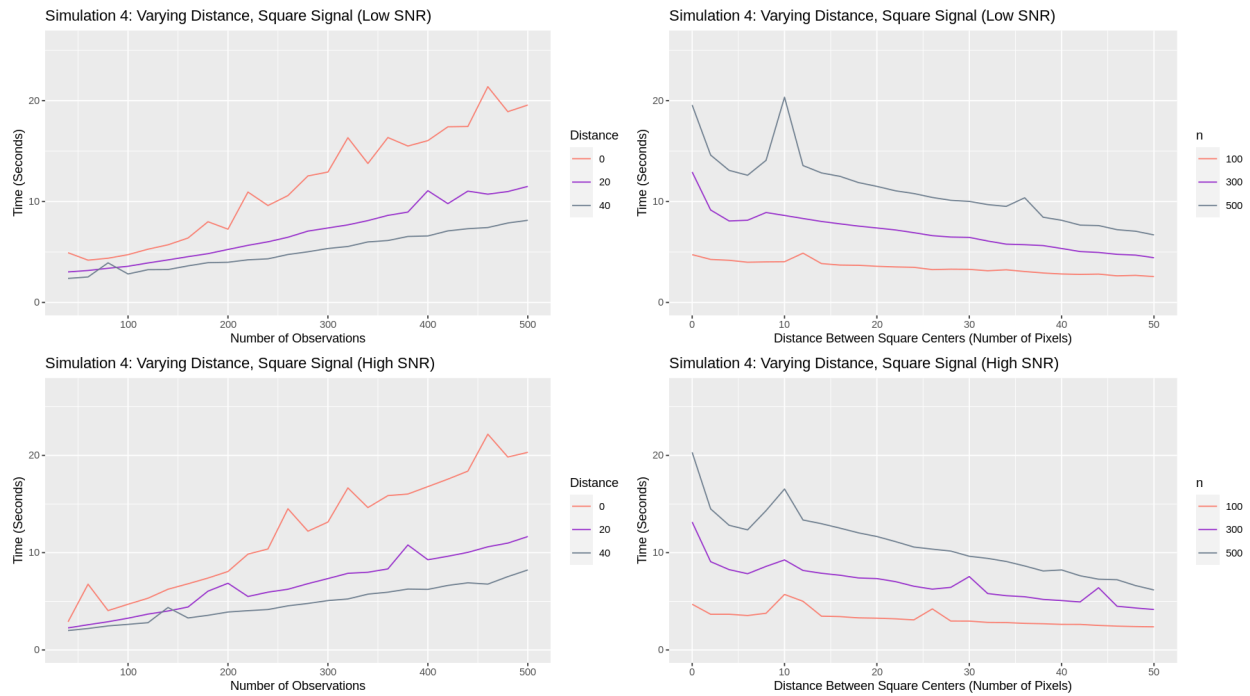


Figure S59: Bootstrap computation times for Simulation 4. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs computation time, shown for square separations of 0, 20 and 40 pixels. Right column: Separation between squares vs computation time, shown for models including 100, 300 and 500 observations.

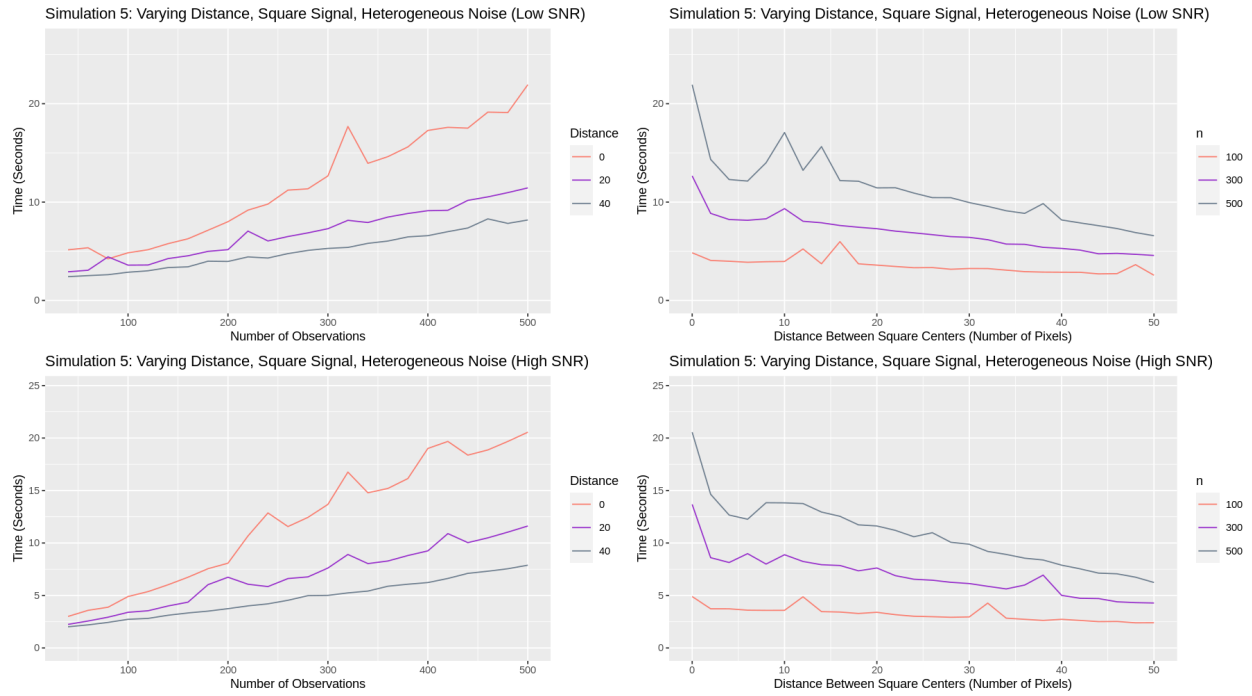


Figure S60: Bootstrap computation times for Simulation 5. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs computation time, shown for square separations of 0, 20 and 40 pixels. Right column: Separation between squares vs computation time, shown for models including 100, 300 and 500 observations.

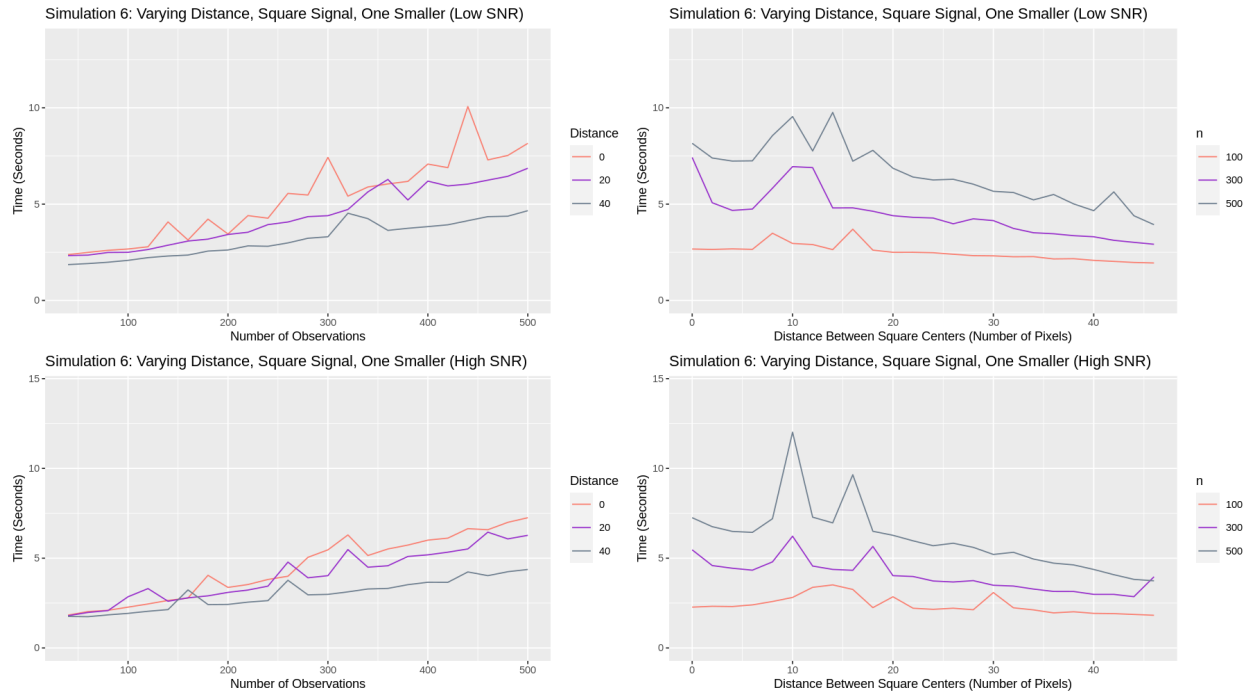


Figure S61: Bootstrap computation times for Simulation 6. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs computation time, shown for square separations of 0, 20 and 40 pixels. Right column: Separation between squares vs computation time, shown for models including 100, 300 and 500 observations.

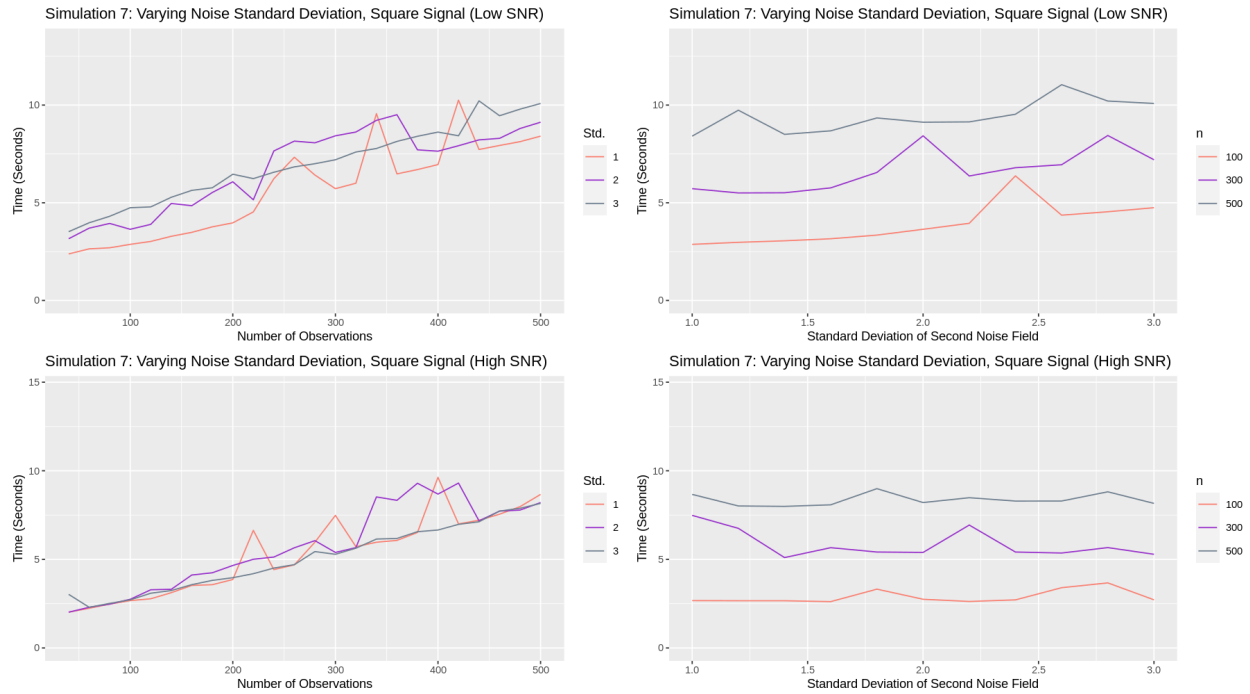


Figure S62: Bootstrap computation times for Simulation 7. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs computation time, shown for noise standard deviations of 1, 2 and 3. Right column: The standard deviation of the second noise field vs computation time, shown for models including 100, 300 and 500 observations.

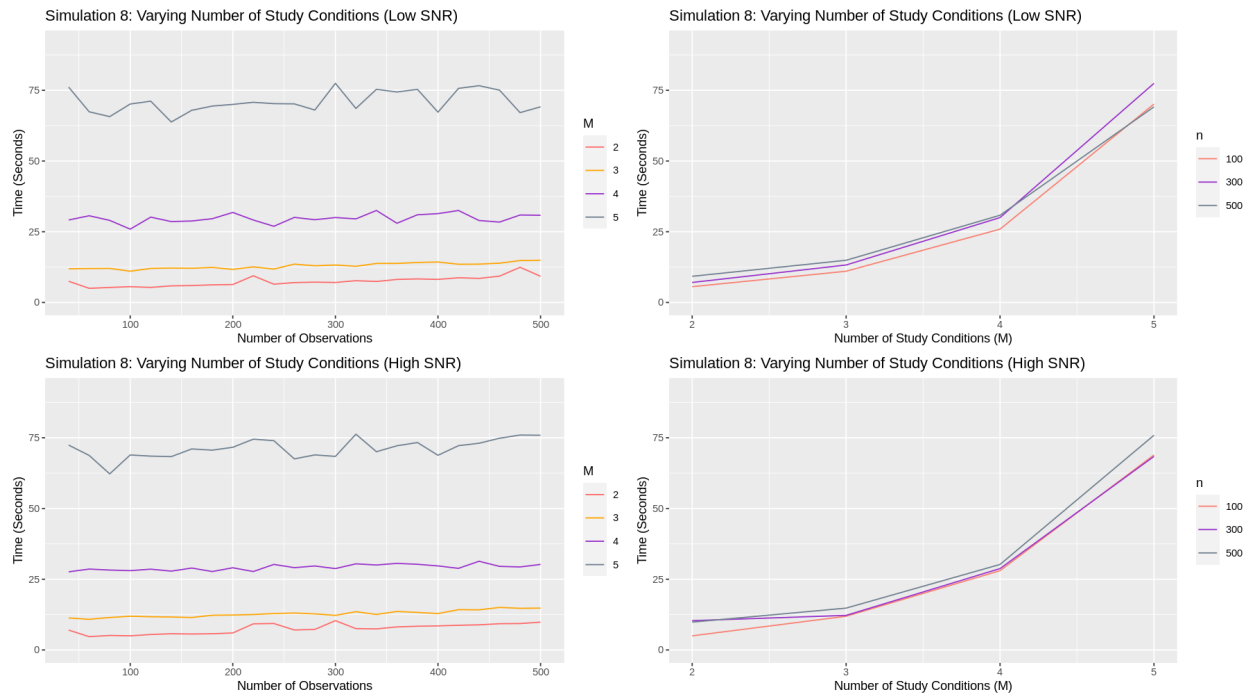


Figure S63: Bootstrap computation times for Simulation 8. Results are displayed for both low-SNR synthetic data (top row) and high-SNR synthetic data (bottom row). Left column: Number of observations vs computation time, shown for $M = 2, 3, 4$ and 5 . Right column: Number of study conditions (M) vs computation time, shown for models including 100, 300 and 500 observations.

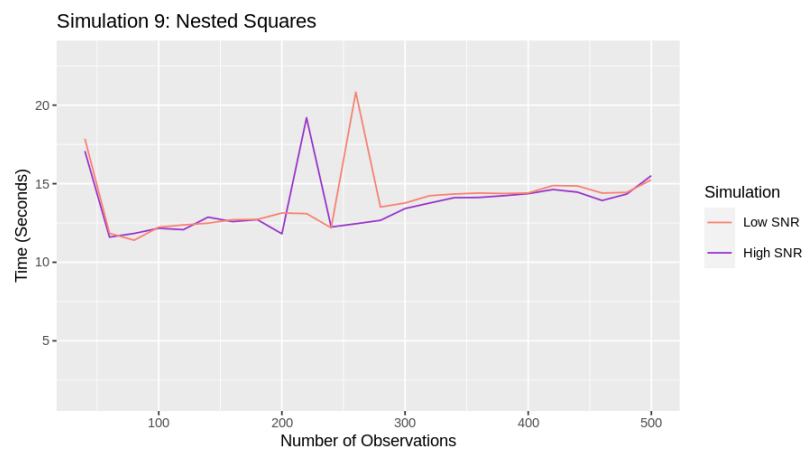


Figure S64: Bootstrap computation times for Simulation 9. Results are shown for both low-SNR synthetic data and high-SNR synthetic data. Displayed is the number of observations vs computation time.

References

- Alexander Bowring, Fabian Telschow, Armin Schwartzman, and Thomas E. Nichols. Spatial confidence sets for raw effect size images. *NeuroImage*, 203:116187, 2019. ISSN 1053-8119. doi: <https://doi.org/10.1016/j.neuroimage.2019.116187>. URL <https://www.sciencedirect.com/science/article/pii/S1053811919307785>.
- Mark W. Woolrich, Brian D. Ripley, Michael Brady, and Stephen M. Smith. Temporal autocorrelation in univariate linear modeling of fmri data. *NeuroImage*, 14(6):1370–1386, 2001. ISSN 1053-8119. doi: <https://doi.org/10.1006/nimg.2001.0931>.
- Max Sommerfeld, Stephan Sain, and Armin Schwartzman. Confidence regions for spatial excursion sets from repeated random field observations, with an application to climate. *Journal of the American Statistical Association*, 113(523):1327–1340, 2018. doi: 10.1080/01621459.2017.1341838. URL <https://doi.org/10.1080/01621459.2017.1341838>. PMID: 31452557.
-