Supplemental Material for “Spatially-Balanced Designs that Incorporate Legacy Sites”

S.D. Foster, G.R. Hosack, E. Lawrence, R. Przeslawski, P. Hedge, M.J. Caley, N.S. Barrett, A. Williams, J. Li, T. Lynch, J.M. Dambacher, H.P.A. Sweatman, and K.R. Hayes

November 18, 2016

S.1 Default Value for The Influence of a Legacy Site

We derive a a useful default value for specifying σ as a 2-dimensional problem. The extension to other dimensions is discussed shortly. We choose σ so that the area-of-influence is equal to the average area for each sampling location (legacy sites and new sites). To this end, let A be the total area of the study region and let $A' = {}^{A}/_{n+L}$ be the average area represented by a sampling location. Choose radius r such that the area of a circle around a legacy site is equal to A', that is $r = \sqrt{{}^{A'}/_{\pi}}$ (here π is the usual constant). We allow 95% of the influence for any legacy site to be within r of the legacy site's location. That is

$$0.95 = \int_{-r}^{r} \frac{1}{\sqrt{2\pi\sigma^2}} \exp{(-\frac{x^2}{2\sigma^2})} dx.$$

Note that this approach assumes that distances in all spatial directions are commensurate. While extensions could be made, we do not as we aim to keep specification of designs simple and having non-commensurate metrics would mean specifying more parameters. We choose σ so that this relationship holds. In particular, we numerically solve

$$0 = \Phi(r; \sigma) - \Phi(-r; \sigma) - 0.95$$

for σ, where $\Phi(\cdot; \sigma)$ is the centred normal distribution function with standard deviation σ.

Extension to more than two dimensions (or one dimension) proceeds by choosing r so that it is now the volume of a (hyper-)sphere, that is for $k-$dimensions

$$r = \left(\frac{A'\Gamma\left(\frac{n}{2}+1\right)}{\pi^{\frac{k}{2}}} \right)^{\frac{1}{k}},$$

where $\Gamma(\cdot)$ is the usual gamma function (see Abramowitz & Stegun, 1965, for example). The remaining calculations follow the 2-dimensional case exactly, as the hyper-sphere is assumed symmetrical in all directions.

S.2 Invariance to Ordering

Consider the inclusion probability for new location i after it has been adjusted for legacy sites l_1 and l_2, that is:

$$\begin{aligned}\pi_{i|l_1 l_2} &= \pi_{i|l_1}\left[1-(1-\pi_{l_2})h(i,l_2)\right] \\ &= \pi_i\left[1-(1-\pi_{l_1})h(i,l_1)\right]\left[1-(1-\pi_{l_2})h(i,l_2)\right] \\ &= \pi_i\left[1-(1-\pi_{l_2})h(i,l_2)\right]\left[1-(1-\pi_{l_1})h(i,l_1)\right] \\ &= \pi_{i|l_2 l_1}.\end{aligned}$$

So, it is inconsequential which of the two legacy sites the inclusion probabilities are adjusted for first. The extension to more than two legacy sites is immediate. The result also shows that the adjustment of multiple legacy sites can be done in parallel (not sequentially).

S.3 Figures

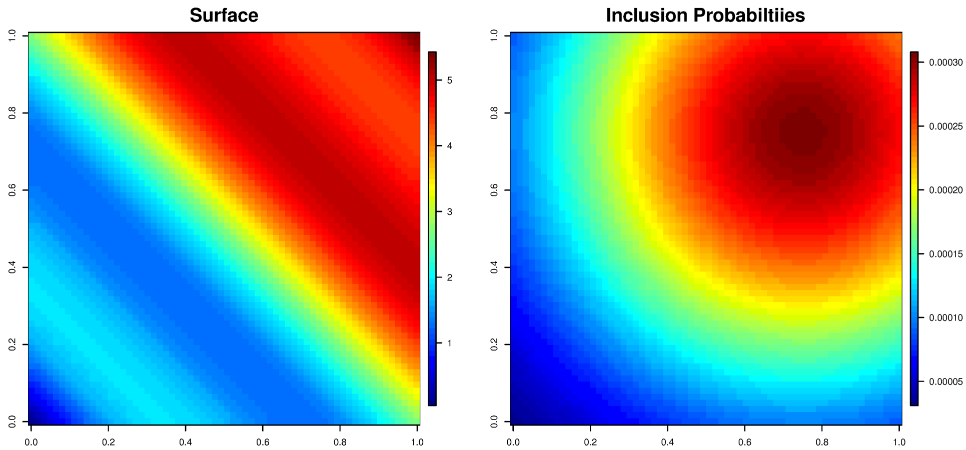

Figure S.1: Illustration of the simulation study's setup. *Left:* The surface that is to be sampled. *Right:* Inclusion probabilities for generating new sampling locations. Note that the two surfaces are similar (both have maximum in top right and minimum in bottom left) but are not proportional.

S.1 $n = 300$ Simulation Study

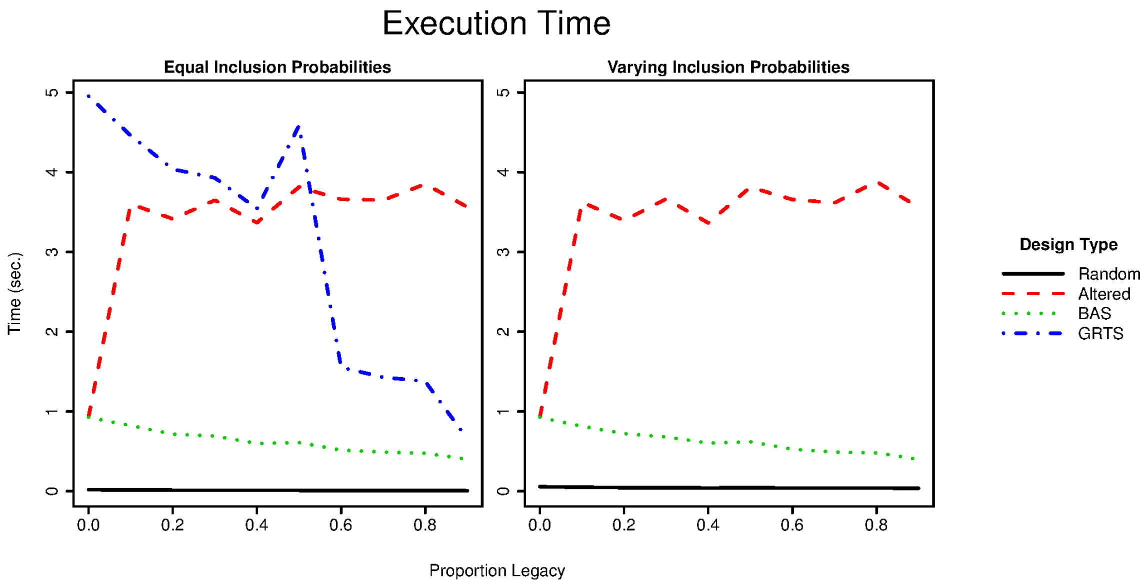

Figure S.2: **Results from Simulation** Average run times. *Left:* Equal inclusion probability scenario. *Right:* Varying inclusion probability scenario. See Section 2.2 for details on the simulation scenarios.

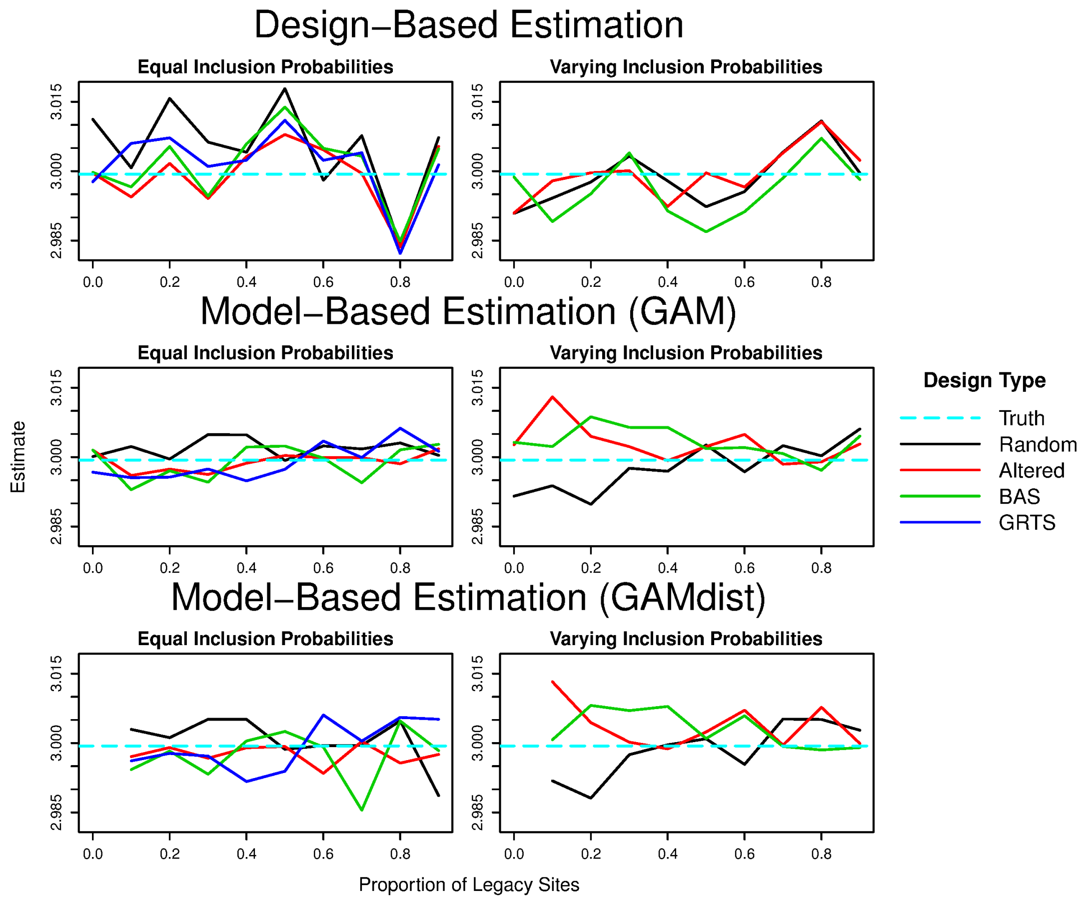

Figure S.3: **Results from Simulation** Average point estimates. *Top row:* Design-based estimator using (2). *Middle row:* Model-based estimator using (4). *Bottom Row:* Model-based estimator using (6). *Left Column:* Uniform inclusion probabilities and *Right Column:* varying inlcusion probabilities. See Section 2.2 for details on the simulation scenarios.

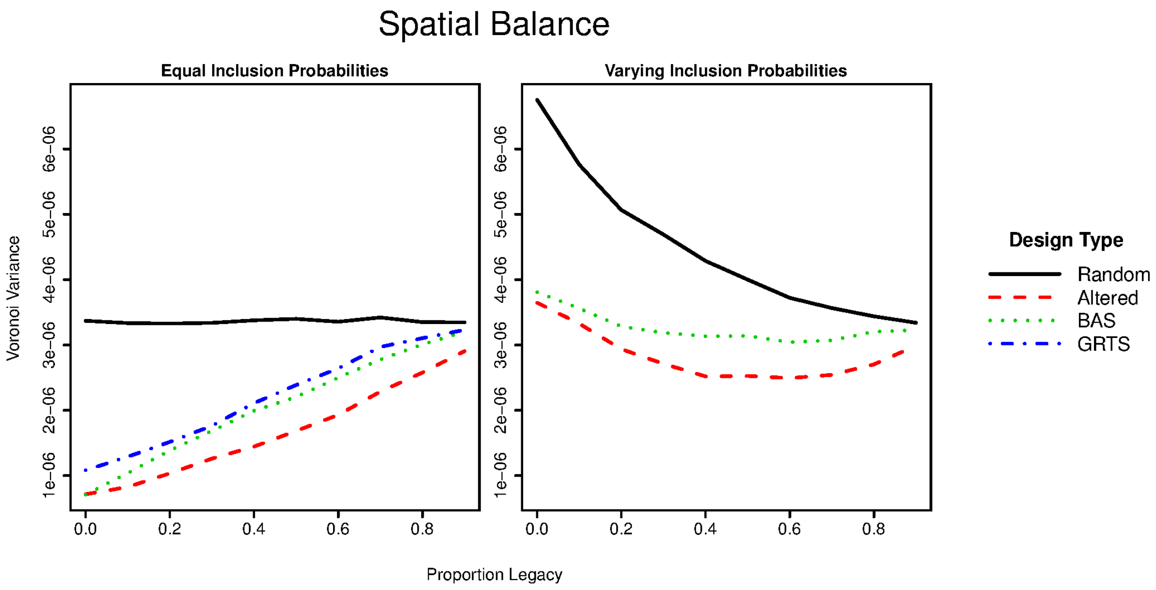

Figure S.4: **Results from** $n = 300$ **Simulation** Average spatial balance over 100 simulations for the different design types. *Left:* Equal inclusion probability scenario. *Right:* Varying inclusion probability scenario. See Section 2.2 for details on the simulation scenarios.

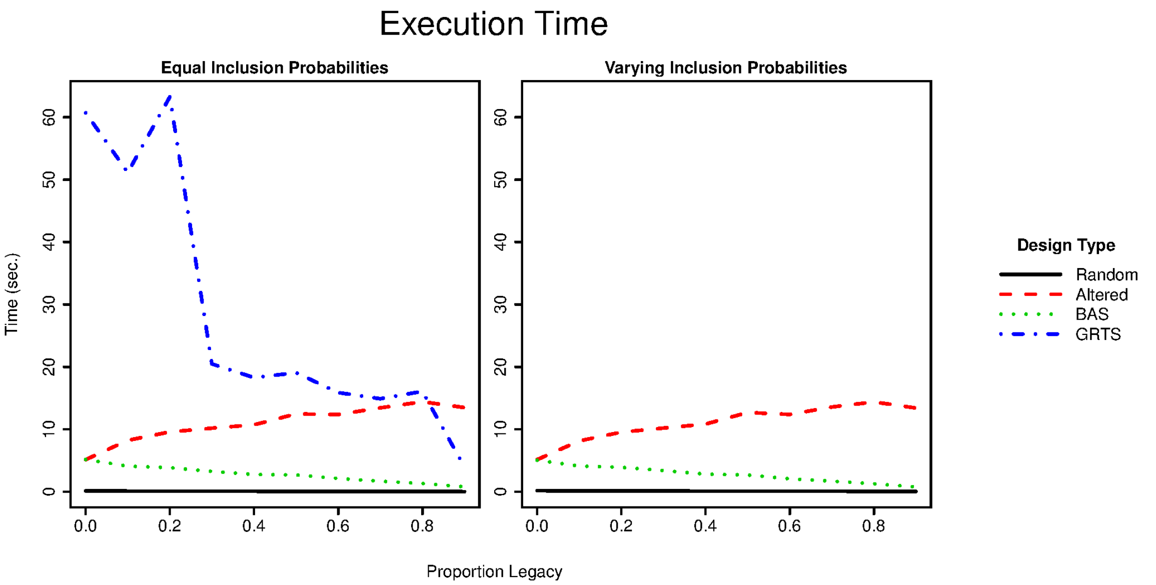

Figure S.5: **Results from** $n = 300$ **Simulation** Average execution time over 100 simulations for the different design types. *Left:* Equal inclusion probability scenario. *Right:* Varying inclusion probability scenario. See Section 2.2 for details on the simulation scenarios.

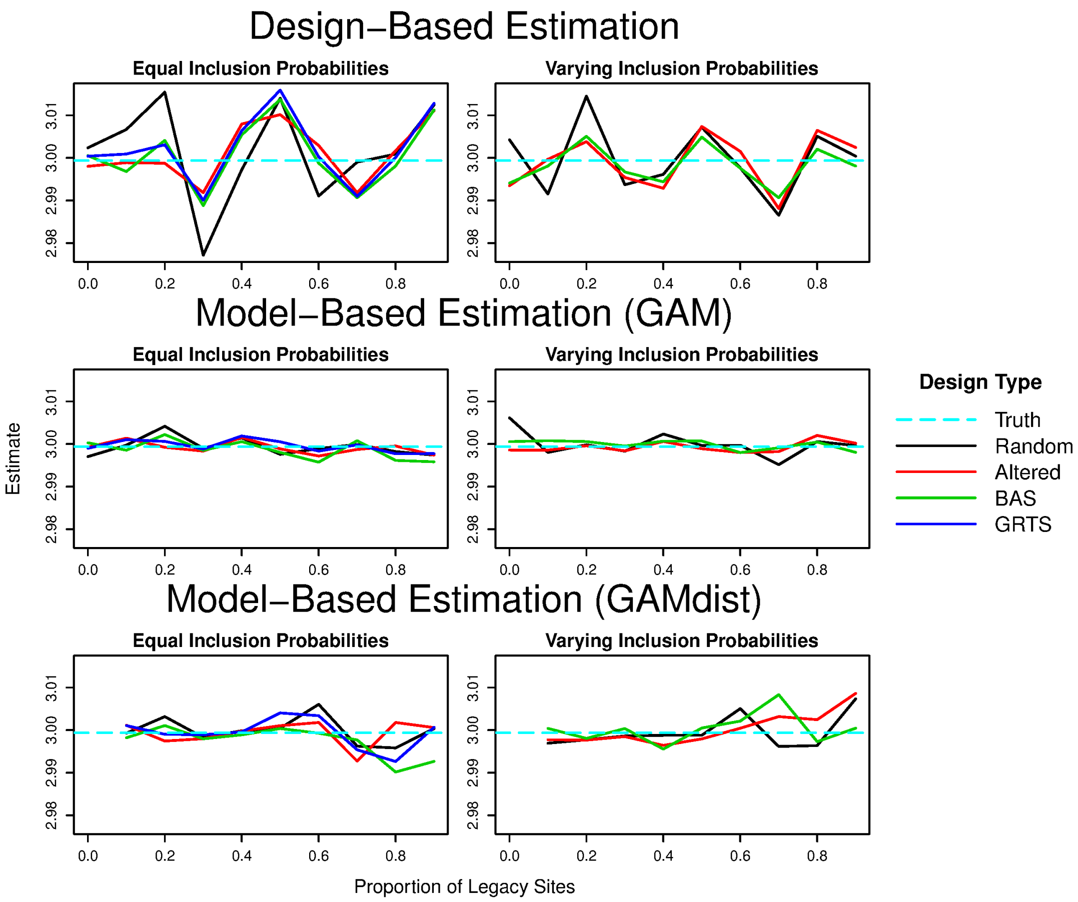

Figure S.6: **Results from** $n = 300$ **Simulation** Average point estimates. *Top row:* Design-based estimator using (2). *Middle row:* Model-based estimator using (4). *Bottom Row:* Model-based estimator using (6). *Left:* Uniform inclusion probabilities and *Right:* varying inlcusion probabilities. See Section 2.2 for details on the simulation scenarios.

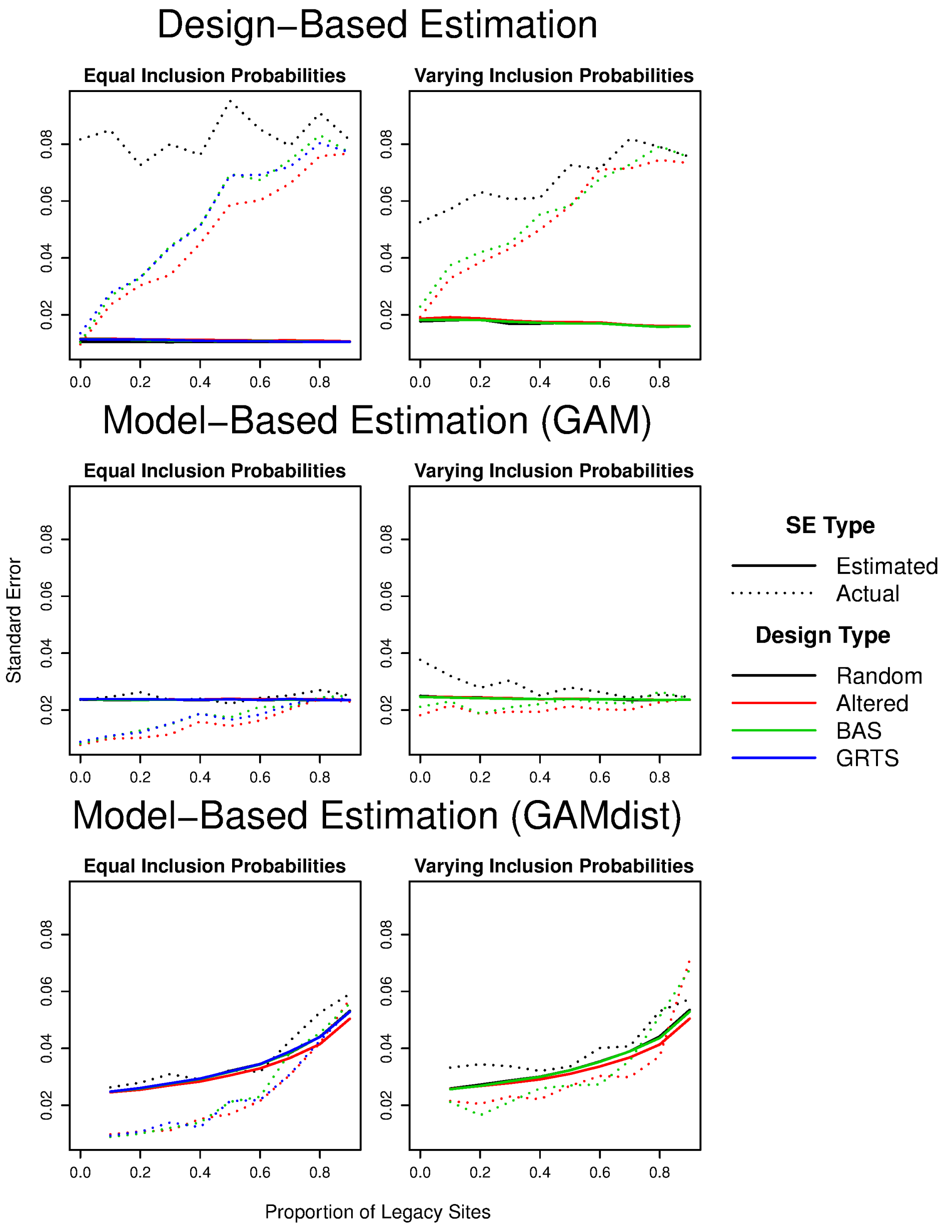

Figure S.7: **Results from** $n = 300$ **Simulation** Estimated and actual standard errors. The estimated standard error is the average, over simulations, of the estimated standard error for that simulated data set. The actual standard error is the standard deviation, over simulations, of the mean estimates. *Top row:* Design-based estimator using Stevens & Olsen (2003). *Middle row:* Model-based estimator using (4). *Bottom Row:* Model-based estimator using (6). *Left:* Uniform inclusion probabilities and *Right:* varying inlcusion probabilities. See Section 2.2 for details on the simulation scenarios.

References

Abramowitz, Milton, & Stegun, Irene A. (eds). 1965. *Handbook of mathematical functions : with formulas, graphs, and mathematical tables.* New York: Dover. Unaltered, unabridged republication of 55, National Bureau of Standards, Applied mathematics series (1964) corrected edition.–p. iv of cover.

Stevens, D.L., & Olsen, A.R. 2003. Variance estimation for spatially balanced samples of environmental resources. *Environmetrics*, **14**(6), 593–610.