Simulation Details

We created a simulation to compare estimation efficacy for maximum likelihood estimation (MLE) and restricted maximum likelihood estimation (RMLE). It is impossible to compare all possible permutations that might affect performance. For that reason, we created a script so the reader can decide for themselves which method to use. We give the results for two simulations, and some general recommendations, but the reader should consider their own objectives and data, and decide for themselves.

First, we simulated 100 spatial locations from a bivariate uniform distribution for the x-coordinate and y-coordinate between $(0, 1) \times (0, 1)$. Each simulated spatial point is completely independent of any other spatial points. To the simulated spatial coordinates, we added a point at $(0.5, 0.5)$ and a point at $(1, 1)$, which will not be used for estimation, but will be used for prediction. A new set of points were generated for each simulation.

We simulated a geostatistial spatial process at these 102 locations. The `R` script allows for one of three autocorrelation models to be chosen: 1) exponential, 2) spherical, or 3) Gaussian. Others can be added easily. The simulation also uses a nugget effect. Let $C(d)$ be a correlation model for some distance d; in our case, one of the 3 listed above. For example, the exponential autocorrelation function is $C(d) = \exp(-3d)$, where the 3 ensures that the effective range is reached at distance $d = 1$. Then we used the following parameterization $cov(d) = \sigma^2((1 - \pi)C(d/\rho) + \pi)$ where $\sigma^2 > 0$, $\rho > 0$, and $0 \le \pi \le 1$, and $cov(d)$ is the covariance between any two points separated by distance d. Here, σ^2 is the overall variance, ρ is the range parameter, and π is the nugget effect's proportion of the total variance. The covariance matrix $\mathbf{\Sigma}$ was constructed using $cov(d)$ for the 102 simulated points. Let $\mathbf{L}$ be the Cholesky decomposition such that $\mathbf{L}'\mathbf{L} = \mathbf{\Sigma}$. Then spatially autocorrelated random errors were simulated as $\mathbf{e} = \mathbf{L}'\mathbf{z}$, where $\mathbf{z}$ are randomly generated i.i.d. normal variates with mean zero and variance one.

We included an overall mean, β_0 to the model, and two covariates, x_1 and x_2. At all 100 spatial locations, we generated x_1 as i.i.d. from a normal distribution with zero mean and variance equal to one. For x_2, we used the easting coordinate value. Thus, x_1 was spatially unpatterned, while x_2 was highly spatially patterned, and thus is confounded with the spatial random errors. Then, the data were simulated as

$$\mathbf{y} = \beta_0\mathbf{1} + \beta_1\mathbf{x}_1 + \beta_2\mathbf{x}_2 + \mathbf{e},$$

where $\mathbf{1}$ is a vector of all ones, $\mathbf{x}_1$ is a vector of the x_1 values, and $\mathbf{x}_2$ is a vector of the x_2 values.

Once the data were simulated, we first estimated the covariance parameters using data from only the 100 randomly simulated points. The σ^2 parameter can be profiled, so we need to optimize either the MLE or RMLE equations to estimate ρ and π. Let $\rho_{\max}$ be the maximum range that we will allow, and let $\mathbf{g} = ((1, \ldots, 5)/1000, (1, \ldots, 5)/100, (1, \ldots, 9)/10, (95, \ldots, 99)/100, (995, \ldots, 999)/1000)$. Then we chose $\hat{\rho}_0$ and $\hat{\pi}_0$ as the values that maximized the loglikelihood, or restricted loglikelihood on the search grid $\{\pi \in \mathbf{g}\} \times \{\rho \in \rho_{\max}\mathbf{g}\}$. Then, using $\hat{\rho}_0$ and $\hat{\pi}_0$ as starting values, we used `optim()` in `R` to maximize the loglikelihood, yeilding $\hat{\rho}_{ml}$, $\hat{\pi}_{ml}$, and $\hat{\sigma}^2_{ml}$, or to maximize the restricted loglikelihood, yielding estimates $\hat{\rho}_{rml}$, $\hat{\pi}_{rml}$, and $\hat{\sigma}^2_{rml}$. After the covariance parameters have been estimated, then we proceeded to estimate the fixed effects and predict at the two points $(0.5, 0.5)$ and $(1, 1)$.

For performance measures, we considered bias and root-mean-squared error (RMSE) for covariance parameters, and we also considered these for the actual covariance at a short distance, a medium distance, and a long distance. That is, the covariance depends on all 3 covariance parameters, so we compared the estimated values of $\sigma^2(1-\pi)C(0.01/\rho)$, $\sigma^2(1-\pi)C(0.5/\rho)$, and $\sigma^2(1-\pi)C(2\sqrt{2}/\rho)$ to their true values. We also considered bias, RMSE, and 90% confidence interval coverage for fixed effects, and we considered bias, root-mean-squared-prediction error (RMSPE), and 90% prediction interval coverage for predicting the two points.

Each simulation study simulated 1,000 data sets for each of 5 scenarios. In scenario 1, we let $\pi = 0.99$ and $\rho = 0.01$. In this case the model is dominated by independence among the locations. It makes little sense to change ρ when $\pi = 0.99$, or to change π when $\rho = 0.01$, as near independence will always be the case. For scenario 2, we let $\pi = 0.5$ and $\rho = 0.5$, so 50% of the variance is nugget effect, and 50% of the variance is spatially autocorrelated, with a medium amount autocorrelation given by $\rho = 0.5$. For scenario 3, we again let $\pi = 0.5$ but allow for a large amount of autocorrelation by setting $\rho = 1.5$. For scenario 4, we let $\pi = 0.01$ and $\rho = 0.5$, so the covariance is dominated by spatial autocorrelation, with little nugget effect, and there is a medium amount autocorrelation given by $\rho = 0.5$. For scenario 5, we again let $\pi = 0.01$ but allow for a large amount of autocorrelation by setting $\rho = 1.5$.

For each of the 5 scenarios, we compare MLE to RMLE using the performance measures listed above: bias and RMSE for covariance parameters, bias, RMSE, and 90% confidence interval coverage for fixed effects, and bias, RMPSE, and 90% prediction interval coverage for predictions.

For the first simulation study, we used the spherical covariance function, we set σ^2 to 1, we used the 5 scenarios for π and ρ, we set $\beta_0 = 3$, $\beta_1 = 1$, and $\beta_2 = 2$. For covariance estimation, we set $\rho_{\max}$ at 2 times the maximum distance in the data set, or $2\sqrt{2}$. In a second simulation study, we used the exponential covariance function and set $\rho_{\max}$ at 4 times the maximum distance in the data set, or $4\sqrt{2}$. The results are given below, where we break out a table for each of three main kinds of inference: 1) estimating covariance parameters, 2) estimating fixed effects, and 3) making predictions.

Results

Table 1: Performance for estimating covariance parameters where range parameter is 2 times maximum spatial distance in data set with a spherical covariance function.

	$\pi = 0.99$ $\rho = 0.01$		$\pi = 0.5$ $\rho = 0.5$		$\pi = 0.5$ $\rho = 1.5$		$\pi = 0.01$ $\rho = 0.5$		$\pi = 0.01$ $\rho = 1.5$	
Measure	ML	RML	ML	RML	ML	RML	ML	RML	ML	RML
bias[a]	0.081	0.301	-0.094	0.089	-0.957	-0.314	-0.016	0.098	-0.805	-0.363
bias[b]	-0.501	-0.451	0.002	-0.037	0.106	0.035	0.011	0.008	0.026	0.017
bias[c]	-0.028	0.020	-0.103	0.048	-0.313	-0.079	-0.101	0.142	-0.567	-0.275
rmse[a]	0.170	0.707	0.215	0.508	1.030	1.064	0.148	0.350	0.868	0.920
rmse[b]	0.654	0.602	0.190	0.174	0.269	0.227	0.034	0.029	0.049	0.039
rmse[c]	0.142	0.156	0.223	0.347	0.350	0.429	0.309	0.672	0.608	0.623
bias[d]	0.231	0.249	-0.059	0.078	-0.255	-0.036	-0.105	0.138	-0.567	-0.276
bias[e]	0.000	0.011	0.009	0.066	-0.226	-0.046	0.031	0.146	-0.427	-0.186
bias[f]	0.000	0.004	0.000	0.015	-0.259	-0.186	0.000	0.025	-0.512	-0.413
rmse[d]	0.326	0.338	0.218	0.351	0.300	0.416	0.312	0.672	0.609	0.624
rmse[e]	0.002	0.048	0.030	0.232	0.235	0.348	0.096	0.524	0.448	0.524
rmse[f]	0.000	0.020	0.000	0.092	0.004	0.166	0.000	0.203	0.017	0.237

[a] performance measure for ρ
[b] performance measure for π
[c] performance measure for σ^2
[d] performance measure for $\sigma^2(1-\pi)C(0.01/\rho)$
[e] performance measure for $\sigma^2(1-\pi)C(0.5/\rho)$
[f] performance measure for $\sigma^2(1-\pi)C(2\sqrt{2}/\rho)$
where $C(d)$ is a covariance function for distance d

Table 2: Performance for estimating covariance parameters where range parameter is 4 times maximum spatial distance in data set with an exponential covariance function.

	$\pi = 0.99$		$\pi = 0.5$		$\pi = 0.5$		$\pi = 0.01$		$\pi = 0.01$	
	$\rho = 0.01$		$\rho = 0.5$		$\rho = 1.5$		$\rho = 0.5$		$\rho = 1.5$	
Measure	ML	RML	ML	RML	ML	RML	ML	RML	ML	RML
bias[a]	0.131	0.870	0.128	0.966	-0.376	1.438	0.373	0.780	0.142	1.566
bias[b]	-0.496	-0.450	-0.079	-0.092	0.025	-0.037	0.010	0.013	0.007	0.005
bias[c]	-0.001	0.057	0.035	0.248	-0.158	0.137	0.448	0.953	-0.003	0.791
rmse[a]	0.315	1.921	0.459	1.851	1.098	2.605	0.510	1.152	0.895	2.422
rmse[b]	0.669	0.619	0.253	0.230	0.244	0.210	0.044	0.041	0.029	0.024
rmse[c]	0.153	0.181	0.231	0.480	0.265	0.473	0.641	1.344	0.494	1.354
bias[d]	0.285	0.307	0.097	0.256	-0.103	0.148	0.442	0.936	-0.004	0.785
bias[e]	0.002	0.020	0.031	0.188	-0.090	0.168	0.229	0.582	0.056	0.761
bias[f]	0.000	0.009	-0.020	0.051	-0.160	-0.000	-0.024	0.103	-0.241	0.212
rmse[d]	0.410	0.431	0.309	0.488	0.263	0.466	0.641	1.332	0.495	1.352
rmse[e]	0.007	0.056	0.081	0.367	0.156	0.418	0.329	0.937	0.395	1.233
rmse[f]	0.002	0.031	0.017	0.196	0.060	0.285	0.064	0.411	0.212	0.825

[a] performance measure for ρ
[b] performance measure for π
[c] performance measure for σ^2
[d] performance measure for $\sigma^2(1-\pi)C(0.01/\rho)$
[e] performance measure for $\sigma^2(1-\pi)C(0.5/\rho)$
[f] performance measure for $\sigma^2(1-\pi)C(2\sqrt{2}/\rho)$
where $C(d)$ is a covariance function for distance d

Table 3: Performance for estimating fixed effects where range parameter is 2 times maximum distance in dataset with a spherical covariance function.

	$\pi = 0.99$ $\rho = 0.01$		$\pi = 0.5$ $\rho = 0.5$		$\pi = 0.5$ $\rho = 1.5$		$\pi = 0.01$ $\rho = 0.5$		$\pi = 0.01$ $\rho = 1.5$	
Measure	ML	RML	ML	RML	ML	RML	ML	RML	ML	RML
bias[a]	-0.002	-0.002	0.002	0.004	-0.005	-0.007	0.020	0.019	0.025	0.023
bias[b]	0.002	0.002	-0.000	-0.001	-0.001	-0.000	0.000	0.000	0.002	0.002
bias[c]	0.011	0.011	-0.005	-0.006	0.006	0.011	-0.025	-0.024	-0.064	-0.070
rmse[a]	0.204	0.208	0.435	0.438	0.652	0.641	0.505	0.520	0.805	0.790
rmse[b]	0.101	0.101	0.083	0.083	0.076	0.076	0.041	0.041	0.026	0.026
rmse[c]	0.360	0.365	0.701	0.709	0.864	0.852	0.837	0.849	1.076	1.063
cov90[a]	0.910	0.932	0.779	0.839	0.522	0.663	0.849	0.892	0.550	0.669
cov90[b]	0.893	0.898	0.889	0.894	0.895	0.899	0.891	0.894	0.890	0.892
cov90[c]	0.910	0.927	0.780	0.847	0.628	0.726	0.843	0.878	0.622	0.711

[a] performance measure for β_0
[b] performance measure for β_1
[c] performance measure for β_2

Table 4: Performance for estimating fixed effects where range parameter is 4 times maximum distance in dataset with an exponential covariance function.

	$\pi = 0.99$ $\rho = 0.01$		$\pi = 0.5$ $\rho = 0.5$		$\pi = 0.5$ $\rho = 1.5$		$\pi = 0.01$ $\rho = 0.5$		$\pi = 0.01$ $\rho = 1.5$	
Measure	ML	RML	ML	RML	ML	RML	ML	RML	ML	RML
bias^a	-0.001	-0.000	0.000	-0.000	-0.007	-0.007	0.016	0.012	0.024	0.026
bias^b	0.002	0.002	-0.001	-0.001	-0.000	-0.000	0.001	0.001	0.003	0.003
bias^c	0.010	0.009	-0.007	-0.009	0.011	0.012	-0.016	-0.018	-0.058	-0.056
rmse^a	0.206	0.211	0.408	0.418	0.582	0.582	0.500	0.529	0.726	0.739
rmse^b	0.101	0.101	0.088	0.088	0.079	0.079	0.055	0.055	0.034	0.034
rmse^c	0.363	0.369	0.649	0.657	0.784	0.781	0.809	0.830	0.979	0.982
cov90^a	0.923	0.946	0.869	0.926	0.702	0.849	0.957	0.979	0.845	0.919
cov90^b	0.894	0.903	0.879	0.890	0.891	0.902	0.883	0.891	0.889	0.896
cov90^c	0.917	0.935	0.875	0.916	0.767	0.832	0.956	0.968	0.848	0.898

^{a} performance measure for β_0
^{b} performance measure for β_1
^{c} performance measure for β_2

Table 5: Performance measures for prediction where range parameter is 2 times maximum spatial distance in data set with a spherical covariance function..

	$\pi = 0.99$ $\rho = 0.01$		$\pi = 0.5$ $\rho = 0.5$		$\pi = 0.5$ $\rho = 1.5$		$\pi = 0.01$ $\rho = 0.5$		$\pi = 0.01$ $\rho = 1.5$	
Measure	ML	RML	ML	RML	ML	RML	ML	RML	ML	RML
bias^a	-0.024	-0.025	-0.019	-0.016	-0.025	-0.024	-0.012	-0.011	-0.000	-0.000
bias^b	-0.045	-0.042	-0.027	-0.026	0.008	0.000	0.010	0.011	0.000	0.002
rmspe^a	1.018	1.022	0.838	0.829	0.737	0.732	0.421	0.421	0.257	0.255
rmpse^b	1.030	1.033	0.937	0.939	0.875	0.863	0.682	0.688	0.482	0.469
cov90^a	0.875	0.882	0.877	0.875	0.899	0.904	0.895	0.902	0.898	0.899
cov90^b	0.906	0.908	0.884	0.892	0.873	0.893	0.889	0.900	0.848	0.867

^{a} prediction at center point $(x = 0.5, y = 0.5)$
^{b} prediction at corner point $(x = 1.0, y = 1.0)$

Table 6: Performance measures for prediction where range parameter is 4 times maximum spatial distance in data set with an exponential covariance function..

	$\pi = 0.99$ $\rho = 0.01$		$\pi = 0.5$ $\rho = 0.5$		$\pi = 0.5$ $\rho = 1.5$		$\pi = 0.01$ $\rho = 0.5$		$\pi= 0.01$ $\rho= 1.5$	
Measure	ML	RML	ML	RML	ML	RML	ML	RML	ML	RML
bias[a]	-0.024	-0.024	-0.010	-0.010	-0.030	-0.031	-0.013	-0.013	-0.002	-0.002
bias[b]	-0.046	-0.044	-0.023	-0.024	0.007	0.005	0.000	0.002	-0.001	-0.001
rmspe[a]	1.025	1.025	0.876	0.879	0.763	0.760	0.568	0.569	0.340	0.340
rmspe[b]	1.035	1.036	0.963	0.963	0.897	0.890	0.801	0.811	0.590	0.590
cov90[a]	0.878	0.883	0.876	0.881	0.904	0.910	0.892	0.898	0.898	0.905
cov90[b]	0.905	0.907	0.900	0.900	0.886	0.891	0.898	0.904	0.875	0.887

[a] prediction at center point $(x = 0.5, y = 0.5)$
[b] prediction at corner point $(x = 1.0, y = 1.0)$

Discusion of Results

My impressions from these simulations is that, like you Dale, it seems that MLE actually estimates covariance parameters better than RMLE. However, neither seems to actually do a great job! It does appear that, for estimating σ^2 in particular, RMLE is always larger than MLE (more positive bias) in Tables 1 and 2, but in general, MLE is closer (according to RMSE). RMLE clearly overestimates ρ if the max range is set higher (Table 2 versus Table 1). There is not much information on estimating π and ρ in scenario 1 because they are confounded. However, both MLE and RMLE do a pretty decent job when $\pi = 0.01$, or the covariance is dominated by spatial autocorrelation, and here RMLE has slightly smaller RMSE.

For estimating the actual covariances at short, medium, and long distances, it is interesting that both MLE and RMLE underestimate at maximum autocorrelation, scenario 5, for all distances in Table 1, but much of that is corrected in Table 2. RMLE in particular seems to overestimate covariance at all distances, and especially in comparison to MLE. Once again, as far as RMSE is concerned, MLE is generally better.

One general conclusion that I have made is that estimating these parameters is not all that great in their own right. Also, I have always drunk the Koolaid that RMLE are unbiased estimating equations. That is based on some theory that I have not tried to understand. However, I would have a hard time making that case, at least for these sample sizes. Maybe in the limit, but that limit must be very large and likely not practical.

On to fixed effects, Tables 3 and 4, we can dispense with bias, as we know they are unbiased. Also, we can pretty much dispense with RMSE, at least for comparison purposes, as MLE and RMLE are very similar – neither has the edge. However, when it comes to coverage, it is clear that RMLE always has wider intervals than MLE, and, because they often undercover, and undercover is probably the worse type of error, RMLE should be preferred (at least to me). In comparing the RMSE between Table 4 and 5, it is also apparent that using the max range of 2 times maximum distance when estimating ρ is not as good as using 4 times maximum distance when estimating ρ. This is very apparent when estimating β_3, as the coverage is very poor in Table 3, but not nearly as bad in Table 4.

Finally, for prediction, Tables 5 and 6, both MLE and RMLE perform very well. Again, we can see that in general RMLE has wider intervals then MLE, but sometimes that leads to overcoverage. Especially in Table 6, neither method seems to be advantageous across the board.

My final comments concern setting the maximum range parameter during optimization. It clearly has an effect in these Tables, although I changed both the autocovariance function (spherical to exponential) and the maximum range (2 to 4 times max distance) when doing simulation study 2. For some reason, I have always used 4 time max distance. That is what is coded into the stream network R package that I developed and maintain. I don't recall how I arrived at that. Perhaps you don't think we should cap it at all? But I think that I ran into optimization stability issues if completely unconstrained. Now that I have the script set up, and it uses `xtable` package to directly create Latex tables that can be pasted into Latex, I will run a few more simulations to see if I can learn anything about setting max range. This can just run in the background while I move forward. It takes between 2 to 3 hours to do a whole simulation study, with 1000 simulations for each of the 5 scenarios.

I actually think that these simulations are fairly comprehensive. I think the 5 scenarios cover most of the important space for the covariance parameters. We included an unconfounded and spatially confounded covariate. Probably the main things left to investigate, then, are the effect of the covariance function, setting max range during estimation, and sample size/configuration. Not that I am going to do that. I think the best thing we can do for our readers is encourage them to use the R script to learn about these themselves.