A COVARIANCE PARAMETER ESTIMATION METHOD FOR POLAR-ORBITING SATELLITE DATA

Michael T. Horrell and Michael L. Stein

University of Chicago

Supplementary Material

The following sections detail the calculation of the asymptotic standard errors given for the parameter estimates, provide a simulation study of the I-likelihood estimator, and give plots of the OMI data used in the application. Construction of the model used in the application is discussed in more detail, and a more in-depth description of the blocking and interpolation strategies used in the I-likelihood in the application is given. The intricacies of the optimizations used in the application are also described. Results comparing standard errors calculated with and without the computational shortcut presented in Section 2.4 in the paper are shown.

S1   Inference

This section discusses the substitution presented in Section 2.4 used in calculating the inverse of the Godambe information matrix. This calculation is again

$$\mathcal{G}^{-1} = \mathbb{E}_\theta(\dot{G})^{-1}\mathbb{E}_\theta\left(GG^T\right)\mathbb{E}_\theta(\dot{G})^{-1}. \tag{S1.1}$$

After $\hat{\theta}$ is plugged in for θ, the central expectation term involves calculations of the variance of several quadratic forms of the general form

$$Y^T F Y + \hat{X}^T D \hat{X}, \tag{S1.2}$$

where $\hat{X} = AY$ and A is the matrix that produces the interpolated points $\hat{X}$ from the true observations, Y. The matrices F and D are matrices obtained from taking derivatives of the Gaussian likelihood with respect to a covariance parameter. For many interpolation methods, A, D and F will depend on interpolation parameters determined by the data, but we treat these matrices as fixed in our calculations. Through simulations in Section S2, we show ignoring this dependency does not appear to substantially bias uncertainty estimates.

In calculating the variance of (S1.2), if $\hat{X}$ is treated more accurately as a linear combination of Y, the variance calculation is

$$2\mathrm{tr}(F\Sigma_Y F\Sigma_Y) + 2\mathrm{tr}(DA\Sigma_Y A^T DA\Sigma_Y A^T) + 4\mathrm{tr}(F\Sigma_Y A^T DA\Sigma_Y). \tag{S1.3}$$

Note the dependence of the covariance matrix Σ_Y on the parameter estimate, $\hat{\theta}$, is dropped in this notation. This dependence is left implicit below as well.

If $\hat{X}$ is used as pseudo-observations with covariance matrix $\Sigma_{\hat{X}}$ and cross-covariance matrix $\Sigma_{Y\hat{X}}$ with Y, the calculation is

$$2\text{tr}(F\Sigma_Y F\Sigma_Y) + 2\text{tr}(D\Sigma_{\hat{X}} D\Sigma_{\hat{X}}) + 4\text{tr}(F\Sigma_{Y\hat{X}} D\Sigma_{Y\hat{X}}^T). \tag{S1.4}$$

Thus in this expression, the covariance of $\hat{X}$ and cross-covariance of Y and $\hat{X}$ (when treating $\hat{X}$ as a linear combination of actual observations) are replaced by the approximations $\Sigma_{\hat{X}}$ and $\Sigma_{Y\hat{X}}$ respectively. We use (S1.4) to approximate $\mathcal{G}$ since this substitution creates substantial computational savings. The savings comes from fewer uses of the full covariance matrix, Σ_Y, in (S1.4) than occur in (S1.3).

S2   Simulation Study

We numerically investigated the properties of the I-likelihood by simulating data with structure similar to data collected by a polar-orbiting satellite. With these simulations, we focused on determining how different first tier and second tier adjustments to an I-likelihood affect point estimates and the measures of efficiency given in Section 2.4. Using several different simulation schemes, we showed the estimates produced using I-likelihood methods to depend predictably on the formations of the first tier and the higher level tiers of an I-likelihood. Primarily, we found two patterns: More interpolated points used in higher level tiers can lead to better estimations especially when the higher tiers are capturing important relationships thrown out in the lower tiers; I-likelihood estimations are robust to the use of incorrect models used for interpolation and moderately robust to the use of poorly chosen interpolation locations. Performance of the I-likelihood was less reliable when points were interpolated to locations outside the approximate domain of the observed data points, and more generally, we found that I-likelihood performance drops when the interpolated points are distributed severely differently from what they are assumed to be in the I-likelihood calculation.

S2.1   Simulated Data and the Model

To mirror a simplified version of polar-orbiting satellite data collected in a single latitudinal band, we simulated data on a subset of the circle × time. The data were simulated in discrete time chunks or orbits, and within each orbit 100 observations were spaced evenly across $30°$ longitude, each occurring at $0°$ latitude. The rough structure of the orbits can be seen in the plot of OMI data in Figure 1.1(b), but in this simulation there are fewer observations per orbit, and again the observations occur at a single latitude as opposed to within a $1°$ latitude band. Consecutive orbits are spaced one time unit apart; hence an orbit can also be thought of as a time unit. The longitudinal distance from the end of one orbit to the beginning of the next is $-3°$ longitude. This gives slight spatial (but not temporal) overlap of observations in consecutive orbits. More complete

spatial overlap occurs every 13.3 orbits; thus a day in these simulations may be defined by this number of orbits. For the first set of simulations, we produced 25 consecutive orbits of mean zero data, yielding data sets of size $n = 2,500$. To further investigate the calibration of standard errors, we also simulated data sets with 100 consecutive orbits, giving $n = 10,000$.

The model we used to simulate and interpolate data points is a simpler version of the model used in the application in Section 3. If we denote l_i and t_i as the longitude and time points of the i-th observation, the covariance function of the process Y was.

$$\text{Cov}(Y(l_i, t_i), Y(l_j, t_j)) = \sigma^2 \exp\left(-\sqrt{\frac{\text{chord}(l_i - \omega t_i, l_j - \omega t_j)^2}{\alpha^2} + \frac{(t_i - t_j)^2}{\beta^2}}\right),$$

where the function *chord* gives the chordal distance between two longitudinal points on a ring with the same radius as the earth. Note there is no nugget in this model. For some simulations we interpolated points to a non-zero latitude. To accommodate different latitudinal locations, the *chord* function is extended to give chordal distance on a sphere with the same radius as the earth. The parameters σ, α and β are standard deviation and range parameters respectively. The parameter ω is a drift parameter controlling how quickly the process rotates on the circle. General discussion of this parameter is in Section S4. For all simulations, we set the true values $\sigma = 36.69$, $\alpha = 4.24$ 1000s of Km , $\beta = 8$ orbits or 0.6 days, and $\omega = 10/14.5 = 0.69°$ longitude per orbit. These values were chosen to be similar to those estimated in the application in Section 3, but the values are not immediately comparable. In particular, we used a temporal range markedly smaller than that estimated in the application because there are a smaller number of orbits present in the simulated data. We expect parameters σ and α to be reasonably estimated using within orbit information. Parameters β and ω, however, will only be estimable using cross-orbit comparisons. We expect increased use of cross-orbit information to improve estimation of these parameters.

S2.2 Notation

For each simulation, we varied the first tier composite likelihoods, the number of interpolated observations, the interpolating models, and the interpolation locations. To consolidate this information in our notation, we represent the I-likelihoods considered in this section by letters denoting the base composite likelihood together with three subscripts.

The I-likelihoods considered have the form of either a $B_{.,.,.}$, $HB_{.,.,.}$ or an $R_{.,.,.}$. An I-likelihood with letter B has a block independent composite likelihood where the blocks are equivalent to the orbits. An I-likelihood with letters HB has a block independent composite likelihood where the blocks are made up of two half orbits. Blocks in this case are made of the last 50 points in one orbit combined with the first 50 points of the next orbit. A blocking scheme of this type has temporal variability present in the first tier composite likelihood, while the B blocking scheme does not. An I-likelihood with letter R denotes a random blocking scheme using a block independent composite likelihood.

Here, observations are partitioned via simple random sampling into 25 blocks each with 100 observations.

The subscripts control different properties of the interpolation method. All interpolations in this section are made using the general block interpolation method presented in Section 2.2 where each Y_i used to interpolate $\hat{X}_i(Y_i)$ is a vector of observations from a single orbit. The first subscript in the notation indicates the number of interpolated points at the second tier. We use values 1 and 10. The second subscript indicates the model used for interpolation. Either the true model using correct values is used, indicated by a 0, or a false model is used, indicated by a 1. The false model used has the same structure as the true model, but an incorrect parameter value is used. I-likelihoods under the false model use a simple parameter adjustment, setting $\alpha = 1$ for interpolation. Since each Y_i is data from a single orbit, only the parameter α affects interpolations. The fourth subscript indicates at what latitude the interpolated points are interpolated to. The latitudes used are 0, 1, 2, 5 and 40 degrees north. Most simulations used the proper latitude locations at $0°$.

S2.3   Blocking Type and Interpolation Number Simulation

We compared B, HB and R blocking styles for the first tier composite likelihood using 1 and 10 interpolated observations for each blocking style. For each of the simulations in this section, we used the true model for interpolation and interpolated to the $0°$ latitude. The blocks in B were used as the Y_i to generate the interpolated points for each of the estimating equations in this section.

With these six I-likelihoods, we performed 1000 independent simulations, calculated the root mean square error (RMSE) of the estimates compared to the true values and found 95% coverage probabilities using the square roots of the diagonals of $\mathcal{G}^{-1}$ in Section 2.4 as standard errors and a normal approximation. The RMSE comparisons are in Table S2.1. The coverage proportions are in Table S2.2. To further examine standard error calibration, we also compare standard errors to the sample standard deviations of the estimates obtained in these simulations in Table S2.3. Bias of the estimations is considered using density plots of the estimates using the $B_{10,0,0}$, $HB_{10,0,0}$ and $R_{10,0,0}$ estimating equations. The plot of these density estimates is in Figure S2.1.

	Simulation RMSE					
	Estimating Equation					
Parameter	$B_{1,0,0}$	$B_{10,0,0}$	$HB_{1,0,0}$	$HB_{10,0,0}$	$R_{1,0,0}$	$R_{10,0,0}$
Spatial Range	1.40	1.37	1.39	1.47	1.54	1.59
Temporal Range	172.84	5.37	17.15	3.62	6.82	6.55
SD	5.82	5.76	5.87	6.10	6.09	6.18
Drift	11.13	1.93	1.50	1.40	2.55	2.33

Table S2.1: RMSE of estimates found using simulations of 1000 independent realizations of data on the circle $\times$ time.

The RMSE calculations in Table S2.1 show an increasing number of interpolated observations at the second tier of the I-likelihood adds significantly to the B and HB type estimating equations, especially in the estimation of the temporal range parameter. For the B estimating equations, this is to be expected since unlike the HB estimating equations, the first tier composite likelihoods of the B type estimating equations do not contain any temporal information. Interestingly, the changes in efficiency in estimating the temporal range from $HB_{1,0,0}$ to $HB_{10,0,0}$ show further that significant efficiency gains can be made by using a more informative second tier even when g_1 does not clearly ignore either spatial or temporal relationships. Comparing the remaining parameters for $HB_{1,0,0}$ to $HB_{10,0,0}$, we do not see these efficiency gains.

The R estimating equations show no improvement when using additional interpolated points for g_2. The random blocks formed for the R type I-likelihoods naturally contain large scale spatial and temporal relationships; hence, addition of g_2 which is meant to include large scale relationships does not significantly increase efficiency. Table S2.1 shows that the $R_{.,0,0}$ estimating equations are moderately efficient at estimating all parameters but are less efficient than the other two methods when using 10 interpolated points per orbit, especially $HB_{10,0,0}$.

95% Confidence Interval Coverage Proportion

Parameter	$B_{1,0,0}$	$B_{10,0,0}$	$HB_{1,0,0}$	$HB_{10,0,0}$	$R_{1,0,0}$	$R_{10,0,0}$
			Estimating Equation			
Spatial Range	0.891	0.879	0.883	0.871	0.863	0.881
Temporal Range	0.744	0.926	0.918	0.899	0.922	0.928
SD	0.907	0.893	0.898	0.884	0.903	0.905
Drift	0.425	0.877	0.916	0.898	0.780	0.800

Table S2.2: 95% coverage proportions on simulations of 1000 independent realizations of data on the ring $\times$ time.

The coverage proportion calculations in Table S2.2 indicate standard errors may possibly be too small; however, considering the density plots in Figure S2.1, it is also possible that our simulation is not quite large enough to properly trust the asymptotic normality assumptions entering the coverage proportion calculations. We compared the average standard error calculations to the sample standard deviations of the estimates by computing their ratios. Results are in Table S2.3. Here, we see standard errors become a better measure of the standard deviation when more interpolated locations are added to base composite likelihoods B and HB. Improvement is not seen using the R composite likelihood. In contrast to the coverage proportions in Table S2.2, standard errors for $B_{10,0,0}$, and especially $HB_{10,0,0}$, appear to be slightly *larger* than the true standard deviation. We therefore attribute much of the low coverage proportions in Table S2.2 to a non-normal distribution of $\hat{\theta}$ for the data sets of size 2,500 used in these simulations.

To explore whether or not approximate normality of $\hat{\theta}$ may better apply for larger data sets, we performed estimations on 400 independent simulations using a $B_{10,0,0}$ I-likelihood with 100 orbits or 10,000 observations total. The coverage proportions for these 400 simulations were 0.935, 0.925, 0.94 and 0.94 for the spatial range, temporal range, standard deviation, and drift parameters, respectively, and the corresponding

Average Standard Error / Standard Deviation of Estimates

Parameter	Estimating Equation					
	$B_{1,0,0}$	$B_{10,0,0}$	$HB_{1,0,0}$	$HB_{10,0,0}$	$R_{1,0,0}$	$R_{10,0,0}$
Spatial Range	1.23	1.14	1.15	1.07	1.10	1.11
Temporal Range	0.42	0.92	0.29	1.09	0.98	1.10
SD	1.23	1.14	1.14	1.08	1.10	1.01
Drift	0.84	0.96	0.94	1.02	0.93	0.86

Table S2.3: Comparison of standard error calculations to the sample standard deviation of the estimates from 1000 independent simulations.

ratios of the standard errors to the standard deviations of the estimates were 1.11, 1.13 1.12 and 0.99. These calculations show standard errors may be slightly large, but the better coverage of 95% confidence intervals indicate this $\hat{\theta}$ may be close to normally distributed when using larger data sets.

Overall $\mathcal{G}^{-1}$ appears to contain estimating efficiency information, and the square roots of its diagonals may be treated as reasonably close estimates of the component standard deviations of $\hat{\theta}$. From Figure S2.1, we also see that there is no clear bias in the estimation of parameters using these I-likelihoods.
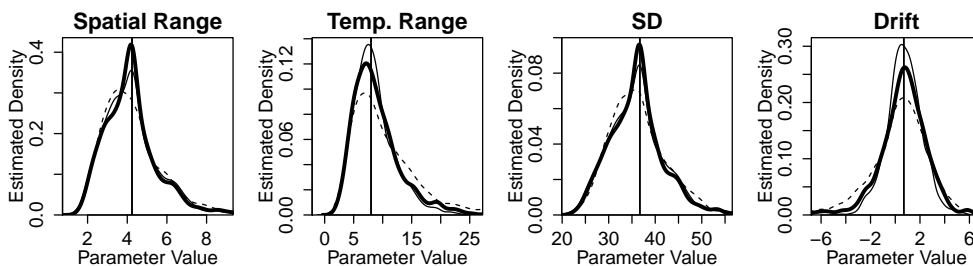

Figure S2.1: Density plots of the estimates obtained from 1000 simulations. Each of the estimating equations uses 10 interpolated values at 0° latitude for the second tier calculation. Vertical lines indicate the true value of the parameter being estimated. The bold line corresponds to $B_{10,0,0}$; the light solid line is $HB_{10,0,0}$; and the dashed line is $R_{10,0,0}$.

S2.4   Interpolation Using an Incorrect Model

We investigated how sensitive estimations are when interpolations are made using an incorrect model. From Table S2.1, the I-likelihood appears to be most affected by the second tier when using the B type I-likelihoods. Therefore, we compared the estimations from the $B_{10,0,0}$ estimating equations to those using this flawed interpolation strategy. The estimating equations compared in this section are $B_{10,0,0}$ and $B_{10,1,0}$. In these estimating equations, the interpolation models are different, but the locations are the same. We consider the consequences of poorly chosen interpolation locations in Section S2.5. The coverage proportions and RMSE of these incorrectly interpolated estimating

equations are in Table S2.4. The $B_{10,0,0}$ columns are the same as the $B_{10,0,0}$ columns in Tables S2.1 and S2.2.

<div align="center">

Simulation Results Comparing I-likelihood Estimations Using
Correct and Incorrect Models for Interpolation Step

</div>

	RMSE		Coverage Proportion	
Parameter	$B_{10,0,0}$	$B_{10,1,0}$	$B_{10,0,0}$	$B_{10,1,0}$
Spatial Range	1.37	1.37	0.879	0.879
Temporal Range	5.37	5.35	0.926	0.922
SD	5.76	5.77	0.893	0.897
Drift	1.93	1.93	0.877	0.877

Table S2.4: RMSE and coverage proportion comparisons on 1000 simulations and estimations when the interpolation step in the I-likelihood is made using an incorrect model. The simulation realizations are the same between both the $B_{10,0,0}$ and $B_{10,1,0}$ estimating equations.

There is strong agreement between the two estimation methods $B_{10,0,0}$ and $B_{10,1,0}$ in Table S2.4, indicating that estimations carried out using interpolation methods that are not perfect will be robust to this imperfection as long as the interpolation locations are close to the actual observation locations.

S2.5   Interpolating Outside the Domain of the Data

We considered how I-likelihood performance is affected by use of interpolation locations outside the domain of the actual observations. Until now, the interpolation locations have been at the 0° N latitude. Here we considered interpolations at 1° N, 2° N and 5° N latitudes. We compared the $B_{10,0,0}$ estimations given in Section S2.3 to $B_{10,0,1}$, $B_{10,0,2}$ and $B_{10,0,5}$. The most informative comparisons are the density plots of the estimates arising from 1000 estimation simulations using these estimating equations. The density plots are in Figure S2.2. The bold line corresponds to the estimates from
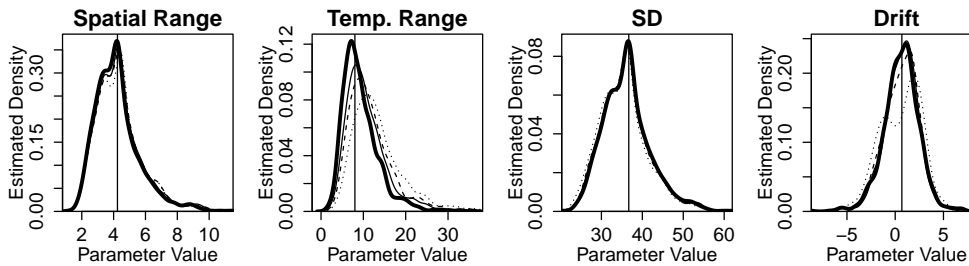

Figure S2.2: Density plots comparing estimating equations when interpolations are made outside the cloud of points formed by the actual observations. The thick line corresponds to the $B_{10,0,0}$ estimating equation. The thin solid line corresponds to the $B_{10,0,1}$ estimating equation. The dashed line is $B_{10,0,2}$ and the dotted line is $B_{10,0,5}$. Note the same data set realizations were used for each estimating equation.

the $B_{10,0,0}$ estimating equation. For both the standard deviation and spatial range parameters, we see little deviation from the bold line. In contrast, the temporal range and the drift parameters exhibit slightly higher deviations. In both the temporal range and the drift cases, the deviation gets worse as the interpolations move farther from 0° latitude; moreover, bias appears to arise but not to a great degree. Despite interpolating points to locations outside of the domain of the observations, the I-likelihood estimator exhibits robustness to these poor interpolation locations. To test this in the extreme case, we performed 1000 estimations using a $B_{10,0,40}$ estimating equation. Here, bias was a substantial problem for the temporal range and drift parameters, but the standard deviation and spatial range parameters appeared to still be estimated reasonably. A density plot of these estimations is in Figure S2.3.
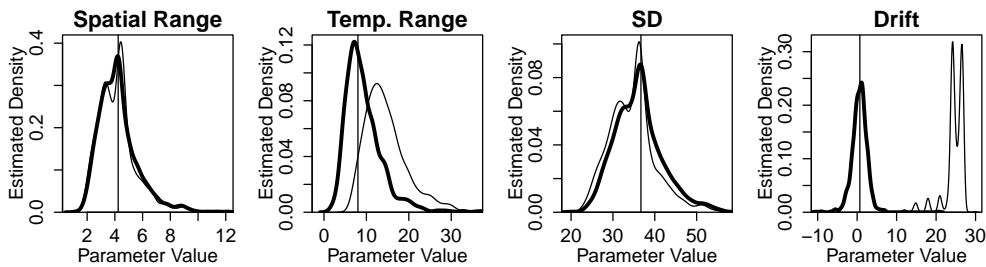

Figure S2.3: Density plots comparing distribution of estimates on 1000 simulations using the $B_{10,0,0}$ estimating equation (the thick line) and the $B_{10,0,40}$ estimating equation (the thin line).

Figure S2.3 indicates using interpolation locations very far from the domain of the actual observations will lead to substantial bias in an I-likelihood estimation. Though it seems quite unreasonable for a practitioner to put interpolated points so far from the true observations, it shows that bias can result if the interpolated points have a severely different distribution than they are assumed to have in the I-likelihood calculation.

More generally, bias will likely be worst when the interpolated observations have a significantly different distribution than is assumed in the I-likelihood calculations. To guard against this type of bias, comparison of the distribution of the interpolated points as linear combinations of the actual observations to the distribution of these points as pseudo-observations should be performed. One such comparison is in Section S7.

S3   Ozone Data Plots

We give plots of subsets of the data used in the application in Section 3. Figure S3.4 shows a plot of the first two consecutive days of ozone data. Note the data are plotted by longitude only, but the different latitudes present in the data can be seen in this plot. Multiple latitudes can be seen in the vertical thicknesses of certain parts of the red or black points. At a given longitude, a thicker red or black stripe roughly indicates a greater degree of latitudinal variability in the one degree latitude band in the data. In Figure S3.4, we see loosely how drift of the ozone process appears in the data. The

points from December 28 are close to an eastward rotation of the points from December 27. To avoid breaking up orbits occurring on either side of the International Date Line (IDL), we simply extend these orbits on the right side of the Figure S3.4; hence, some observations occur at longitudes larger than 180° E. Figure S3.5 gives a close up view of data from December 28 at longitudes west of 140° W without this extension. In Figure S3.5 orbital structure and the discontinuities in the observations made near the IDL are clearer.
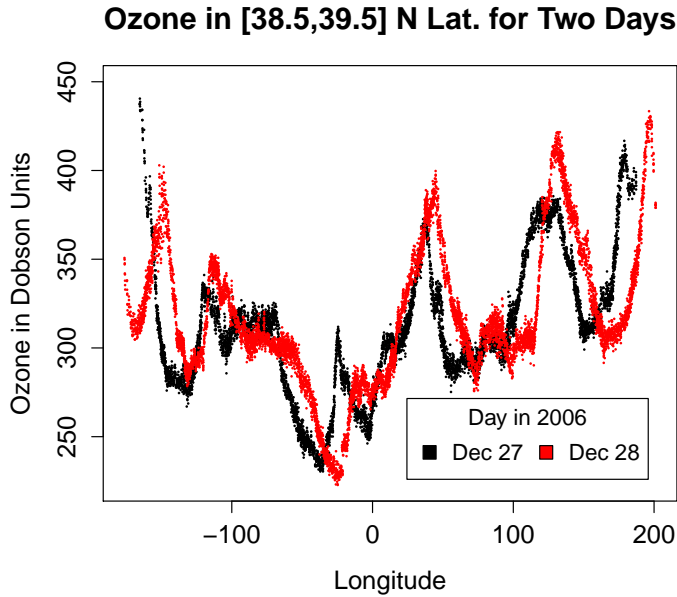
Ozone in [38.5,39.5] N Lat. for Two Days

Figure S3.4: Two days of the ozone data considered in the application in Section 3. The plot is of ozone collected between the 38.5° N and 39.5° N latitudes. Orbits of data that cross the IDL are extended on the right side of this figure (giving some observational longitudes larger than 180° E) to maintain continuity. This data is collected by the OMI on December 27, 2006 and December 28, 2006.

Figure S3.5 shows both the ozone variation that occurs across orbits and across days. Though all observations are collected in the same calendar day defined in the OMI OMDOAO3G product, the observations in orbit 15 are collected nearly 24 hours apart from the observations in orbit 29; hence, there is a clear difference in ozone levels between observations in these orbits despite being close in space. The time difference between orbits 28 and 29 is also apparent in this plot since there appears to be a lesser difference between observations in these orbits at roughly the same spatial locations. The multiple latitudes in the data can be seen more clearly in this plot as local vertical variation, indicating ozone changes across latitudes.
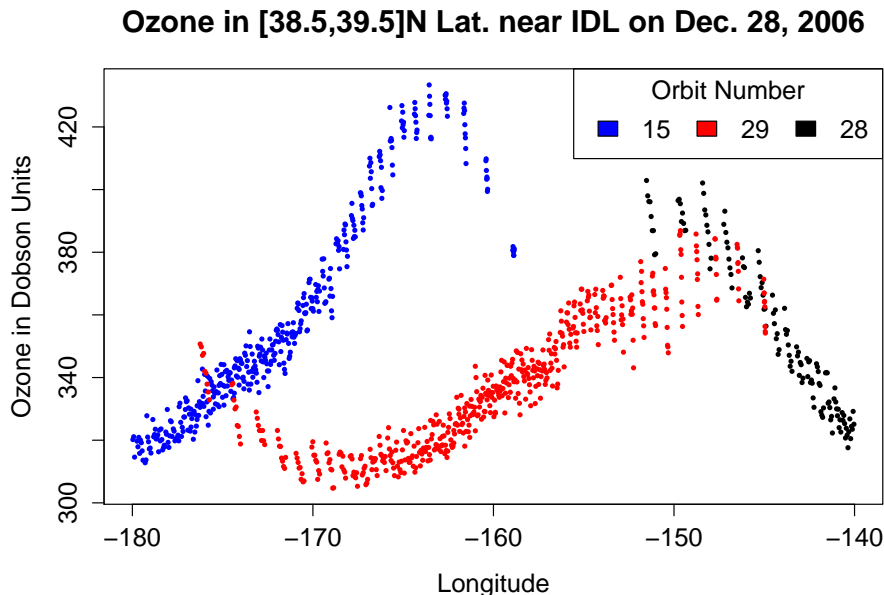
Ozone in [38.5,39.5]N Lat. near IDL on Dec. 28, 2006

Figure S3.5: A close up view of the data from December 28 at latitudes west of $140°$ W longitude. Separate orbits of data are given different colors.

S4   The Model

This section provides more details and intuition about the covariance model in (3.1). We start with two independent Gaussian spatio-temporal Matérn processes in $\mathbb{R}^4$, denoted as Z_1 and Z_2, and we restrict these processes to the sphere×time. We apply a rotational transformation to both processes to create the drift component and apply a differential transformation described in Jun and Stein (2007) to Z_1 to create spatial anisotropies oriented in the same direction at different longitudes. Adding these transformed processes together along with a nugget process gives the model in the paper.

For a given spatio-temporal lag, h, the covariance model for either Z_1 or Z_2 is given by

$$Cov(h) = h^\kappa K_\kappa(h)$$

where $K_\kappa(h)$ is a modified Bessel function of the second kind (Olver et al. (2010)), and κ controls the smoothness or mean square differentiability of the resulting Gaussian random field. The process Z_1 has smoothness $\kappa = \nu + 1$ and Z_2 has smoothness $\kappa = \nu$.

We define squared distance, h^2, between two points observed at specific latitudes (L_1, L_2), longitudes $(l_1,\ l_2)$ and times $(t_1,\ t_2)$, given respectively by $Z_1(L_1, l_1, t_1)$ and

$Z_1(L_2, l_2, t_2)$ or $Z_2(L_1, l_1, t_1)$ and $Z_2(L_2, l_2, t_2)$ as follows:

$$h^2(Z_1(L_1, l_1, t_1), Z_1(L_2, l_2, t_2)) = \frac{\text{chord}(L_1, L_2, l_1, l_2)^2}{\alpha^2} + \frac{(t_1 - t_2)^2}{\beta^2} \quad \text{(S4.1)}$$

where the *chord* function gives the chordal distance between the two points on a sphere described by the two sets of latitude and longitude coordinates. Equivalently this is the $\mathbb{R}^3$ euclidean distance between these points in 3-space. The parameters α and β are the spatial and temporal range parameters, respectively.

The drift component is constructed through a rotation of the process on the sphere about earth's rotational axis. We consider drift only along longitudes because it is easily detected in the small latitudinal band of data being analyzed. To input the drift into both Z_1 and Z_2, we generate new longitude values for each observational point based on the time the observation was recorded. Let l_0 represent the original longitudinal value. The adjusted longitudinal value that enters into (S4.1) is given by

$$l_\omega = l_0 - \frac{\omega}{24} t.$$

Thus for positive ω, we have eastward drift, and for negative ω, we have westward drift. Scaling ω by $1/24$ produces an interpretation of ω as the number of degrees longitude of rotation observed per day. Drift of ozone on the globe has been considered in a similar manner by Stein, Chen, and Anitescu (2013).

Using both modified processes, $Z_1(L, l_\omega, t)$ and $Z_2(L, l_\omega, t)$ we generate a new process, Y, from which the model covariance function is derived. Repeating (4.1) this process takes the following form:

$$\begin{aligned}
Y(L, l, t | \alpha, \beta, \phi, \lambda, \omega, \sigma, \eta, \nu) &= \phi\left(\sin(\lambda)\frac{\partial}{\partial L} + \cos(\lambda)\frac{\partial}{\partial l}\right) Z_1(L, l_\omega, t) \\
&+ \sigma Z_2(L, l_\omega, t) + \eta W(L, l, t) \quad \text{(S4.2)}
\end{aligned}$$

where $\eta W(L, l, t)$ is the nugget effect. The parameters ϕ and λ control the spatial anisotropy of Y. A closed form expression for the covariance function of Y can be obtained (see Jun and Stein (2007)).

S5 The Blocking and Interpolation Step for the I-likelihood

We determined blocks for the application in Section 3 by assigning each orbit of data to a separate block. The Aura satellite makes about 14.5 orbits per day, and thus crosses the 39°N latitude on the sunlit side of the Earth 14-15 times per day. For a single latitudinal band of data, the orbits individually are computationally manageable with an average size of 506 observations. The average time difference between orbits is 99 minutes.

The spatial and temporal locations we interpolated to were determined by the average longitudinal shift and average temporal shift of observations from one orbit to the
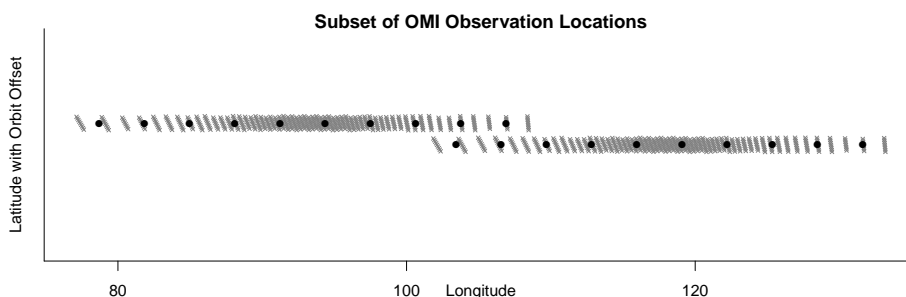
Subset of OMI Observation Locations

Figure S5.6: Plot of interpolated locations with observation locations. The locations interpolated to are restricted to each data block and are spread out within each block. An offset by block is applied to the y-axis to avoid overlap of points in different orbits.

next, and we interpolated to 39° N latitude. With this interpolation strategy and using the interpolated points as pseudo-observations, Block-Toeplitz (and Toeplitz block) structure of the covariance matrix of the interpolated points results. Figure S5.6 gives a visual of this interpolation strategy. The interpolated locations are evenly spaced within a block and in a more subjective sense, they may be seen to represent longitudinal variations at about a 3° resolution.

To generate the interpolated values within a block, we first fit a model to 10 days worth of data using a form of the I-likelihood. The model was a simplified version of (S4.2): we fixed $\nu = 0.5$, $\omega = 0$, and $\phi = \lambda = 0$ and estimated the remaining parameters. For this preliminary estimation note that interpolated points, $\hat{X}$, did not yet exist; hence, the interpolation step to form g_2 was made at each iteration of the numerical optimizer to match the parameter value being considered. The parameter estimates and resulting model obtained were then used to calculate each $\hat{X}_i(Y_i)$ for $i = 1, \ldots, 436$ via kriging. These interpolated values were used for each of the full 30-day I-likelihood estimations.

S6   Optimization

The optimizations are carried out in MATLAB using the `fminunc` function and its trust region minimizer. Analytic gradients are calculated for each parameter except for the smoothness parameter ν. We use finite differences to obtain the gradient of g with respect to ν. The optimization strategy for estimating equations (B, I), (C, N) and (C, I) is a three stage process. First, starting values for the full 30 day estimation are obtained through an estimation on 2 days of data. Second, preliminary estimates are given by fixing the spatial range parameter, α, at the 2-day optimum values and optimizing the other parameters on 30 days of data. Third, final estimates are obtained by letting all parameters vary using 30 days of data, starting at the preliminary estimates. Preliminary estimates seemed unreasonable for (B, S); therefore, the full estimates for this estimating equation are obtained from an optimization across all parameters starting at the 2-day

optimum values. Optimizing (B, N) follows the same strategy as (B, S), except here the temporal range, β, and the drift parameter, ω, are fixed.

The Block-Toeplitz structure for the second tier is utilized in both the (B, I) and (C, I) estimating equations. Akaike (1973) describes an $O(n^2)$ algorithm, similar to the Durbin-Levinson algorithm to calculate a solve and to obtain the log-determinant for Block-Toeplitz matrices. In the application, we adapt a solver written in MATLAB by Pepper (2007). Comparing (B, I) to (B, S) in Table 3.1, we show this solver produces memory savings, however, this comparison also shows there are no savings produced in terms of function computing times. Though the algorithm given by Pepper (2007) has complexity $O(n^2)$, we attribute the faster function evaluations of (B, S) to the overhead produced by using a function written in MATLAB and to the extremely well optimized matrix calculation implementations in MATLAB. For larger Block-Toeplitz matrices, we expect the Block-Toeplitz solver in (B, I) to outperform the unstructured solve used in (B, S).

To save time in finding the (C, S) estimates, optimization of this estimating equation is started at the parameter estimates for (C, I). Hence, fewer function evaluations are used to find the estimates using (C, S). The discrepancy between the memory use figures for (C, S) and (B, S) can be explained by different second tier function evaluation algorithms. The equation (B, S) uses a more memory intensive algorithm so that the Isotropic SD parameter, σ, can be found analytically given the other parameters. This algorithm was also used in the (B, I) estimating equations; hence, the memory savings made using Block-Toeplitz structure at the second tier comparing the (B, I) and (B, S) are still valid.

S7   Evaluation of the Standard Errors in Table 3.1 And a Kullback-Leibler Divergence Comparison

The assessments of the estimating equations in Table 3.1 are valid provided the stated standard errors are approximately correct. Two issues need to be considered: the non-zero expectation of the gradient of the I-likelihood and the validity of the shortcut in calculating $\mathcal{G}^{-1}$ proposed in Section 2.4. Though the I-likelihood is not an unbiased estimating equation, following the results of the simulation study in Section S2, its bias should be quite small when interpolations are performed within the domain of the actual observed data and when the number of interpolated points is small compared to the number of observations. The interpolation steps for all estimating equations used in Section 3 satisfy these criteria; hence we believe bias to be quite small in the application. As further evidence of small bias using the I-likelihood estimating equations in Table 3.1, we see agreement among the estimates in Table 3.1 that have been estimated using both the I-likelihood and the corresponding unbiased composite likelihoods.

In calculating $\mathcal{G}^{-1}$, as mentioned in Section 2.4 and Section S1, using the shortcut of substituting $\Sigma_{\hat{X}}(\theta)$ and cross-covariance $\Sigma_{Y\hat{X}}(\theta)$ in for $A\Sigma(\theta)A^T$ and $\Sigma(\theta)A^T$ respectively may produce standard errors that are too small. To consider this problem, we

fit the model using (B, I) to one day of data to obtain $\hat{\theta}_1$ and $\Sigma_{(1)}(\hat{\theta}_1)$ and $\hat{X}_{(1)}$. The subscript (1) indicates these values correspond to the single day data set. We calculated the standard errors of $\hat{\theta}_1$ with and without these substitutions. The standard errors are given in Table S7.5. Column *SE* corresponds to the (S1.3) type calculation, and column *SE Approx.* corresponds to (S1.4). The difference between these two sets of standard

Standard Error Comparisons on 1 Day (B, I) Estimation

Parameter		Estimate	SE	SE Approx.	Percent Difference
Spatial Range	α	4.19	12.06	12.10	0.3%
Temporal Range	β	146.68	370.01	371.29	0.3%
Isotropic SD	σ	27.27	21.04	21.09	0.2%
Anisotropic SD	ϕ	17.34	63.71	63.90	0.3%
Anisotropy Angle	λ	0.099	0.080	0.080	0.003%
Nugget SD	η	2.470	0.342	0.343	0.4%
Drift term	ω	11.46	9.82	10.02	2.0%
Smoothness	ν	0.591	0.229	0.230	0.4%

Table S7.5: Standard error comparisons

errors was extremely small. All pairs of standard errors except those for $\hat{\omega}_1$ differed by less than 0.5% (in comparison to the non-substitution standard errors). The standard errors on $\hat{\omega}_1$ differed by 2.0%. Applying this type of small difference to Table 3.1, we see that the conclusions involving the efficiency of the I-likelihood methods still hold. As a curiosity, the standard errors calculated using this substitution are larger than those calculated not using the substitution. This indicates the standard errors on the I-likelihood estimates in Table 3.1 may be slightly conservative. This point is echoed in the simulation results of Section S2, but more work would be needed before one could say this pattern holds more generally.

A point of secondary interest is the estimate $\hat{\lambda} = 0.099$ on 1 day of data. This estimate conflicts with the estimates of λ given in Table 3.1. We fit the model to multiple single days of data, and find the estimate of λ changes much more than the standard error 0.080 indicates. We therefore attribute the conflicting estimates of λ to non-stationarities across time in the ozone process that will not be explicitly captured in a multiple-day estimation without use of a more complex model.

As a final overall approximation evaluation, we considered the Kullback-Leibler divergence of a Gaussian random vector that uses $\Sigma_{\hat{X}_{(1)}}(\hat{\theta})$ as a covariance matrix to one that uses the more accurate covariance matrix, $A_{(1)}\Sigma_{(1)}(\hat{\theta})A_{(1)}^T$. Using the 1 day estimates, $\hat{\theta}_1$ with $\hat{\nu}_1 = 0.59$, we found the Kullback-Leibler divergence between these two Gaussian processes to be 0.47 on 7,103 observations. By comparison, if we fit a model fixing $\nu_2 = 0.5$ (an exponential model) to the same data to obtain $\hat{\theta}_2$, the Kullback-Leibler divergence between Gaussian processes with covariance matrices $A_{(1)}\Sigma_{(1)}(\hat{\theta}_1)A_{(1)}^T$ and $A_{(1)}\Sigma_{(1)}(\hat{\theta}_2)A_{(1)}^T$ was 1.97. The approximation using covariance matrix $\Sigma_{\hat{X}}(\hat{\theta})$ is therefore a better approximation to the true process than is an exponential covariance process, an often used approximation in many spatial models and a rather small deviation

from $\hat{\nu}_1 = 0.59$.

Additional References

Akaike, H. (1973). Block Toeplitz matrix inversion. *SIAM J. Appl. Math.* **24**, 234-241.

Gelpke, V. and Künsch, H.R. (2001). Estimation of motion from sequences of images: Daily variability of Total Ozone Mapping Spectrometer ozone data. *J. Geophys. Res.* **106**, 11,825-11,834.

Jun, M. and Stein, M.L. (2007). An approach to producing space-time covariance functions on spheres. *Technometrics.* **49**, 468-479.

Olver, F.W.J., Lozier, D.W., Boisvert, R.F., and Clark, C.W. (2010). NIST Handbook of Mathematical Functions. National Institute of Standards and Technology and Cambridge University Press, Gaithersburg and Cambridge.

Pepper, K. and Fricke, T. (2011). Block Levinson solver. *MATLAB Central.* File ID: #30931

Stein, M.L., Chen, J., and Anitescu, M. (2013). Stochastic approximation of score functions for Gaussian processes *Ann. Appl. Statist.* **7**, 1162-1191.